INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

SMR.780 - 42

FOURTH AUTUMN COURSE ON MATHEMATICAL ECOLOGY

(24 October - 11 November 1994)

"Parameter Estimation in Nonclosed Social Networks Related to Dynamics of Sexually Transmitted Diseases"

Carlos Castillo-Chavez
Biometrics Unit
Cornell University
Ithaca, NY 14853-7801
U.S.A.

These are preliminary lecture notes, intended only for distribution to participants.

Modeling the AIDS Epidemic: Planning, Policy, and Prediction, edited by E. H. Kaplan and M. L. Brandeau. Raven Press, Ltd., New York © 1994

28

Parameter Estimation in Nonclosed Social Networks Related to Dynamics of Sexually Transmitted Diseases

Shu-Fang Hsu Schmitz and Carlos Castillo-Chavez*

*Institut für Mathematische Statistik, Universität Bern CH-3012 Bern, Switzerland;
Biometrics Unit, Cornell University, Ithaca, New York 14853

INTRODUCTION

The process of defining, modeling, and estimating parameters useful for the study of dynamic contact structures is of great importance to such fields as social dynamics, epidemiology, population genetics, cultural anthropology, demography, evolutionary biology, ecology, and immunology. For example, age-dependent contact structures have been used in the study of the dynamics of communicable diseases (CDs) and mathematical epidemiology since 1974 (1–7). CDs such as measles, chicken pox, influenza, and colds are transmitted mostly through casual contacts. Mathematical models help one understand and quantify the effects that age-dependent contact structures have on the transmission dynamics of CDs. Outbreaks usually begin in schools where the rate of casual contacts is higher than in other social settings. The high level of contacts between children has been used to explain primary and secondary outbreaks of some CDs (4).

Casual contacts, the main mode of transmission of CDs, are modeled adequately through the use of proportionate mixing, in which all individuals are assigned age-dependent activity levels and where contacts are assumed to occur in proportion to age-dependent activity levels weighted by their corresponding density (8,9). On average, children may have more contacts because they are more active and/or because they represent a larger proportion of the age-structured population.

The contact structure is not the only feature of importance in the study of the transmission dynamics of CDs; for example, time scales may be quite relevant (5,6). In many instances, there is a significant difference between a host's life expectancy and the average length of the disease's infectious period. To study

single epizootic events it is common to ignore demographic effects by assuming that the population under consideration has reached a stable age distribution. This last assumption also may be useful in the study of the long-term dynamics of CDs where disease-induced mortality is not a factor and where the rate of population growth is not significant. The use of proportionate mixing and the assumption that a population has reached a stable age distribution have been quite useful in the study of disease persistence (endemicity), in the evaluation of disease control strategies, and in the study of the effectiveness of vaccination programs (3,7,10). The usefulness of these assumptions in today's world is becoming limited for modeling of treatable and untreatable sexually transmitted diseases (STDs) because of large increases in migration and travel rates within and between populations.

Proportionate mixing provides an appropriate model of population heterogeneity in the context of CD dynamics but does not provide an all-purpose model. Epidemiologic data, biological and sociological realism, and important demographic considerations did not play an important role (with some exceptions, see refs. 11–16) in the development of mathematical and theoretical epidemiology until the dramatic rise of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS). The spread of HIV/AIDS, particularly in industrialized nations, forced theoreticians to examine potential mechanisms for the spread of HIV using more plausible scenarios. Realistic models incorporating the role of long and variable incubation periods, age-of-infection infectivity, and social dynamics have been developed by many (17–27). The importance of social dynamics, the main topic of this chapter, emerged with the generation of models that incorporate relevant sociological/epidemiological factors including varying degrees of sexual activity, alternate modes of transmission (needle sharing, anal sex, etc.), sexual preference (bi-, hetero-, and homosexual activity), and heterogeneity in pairing/contact structures (8,9,28–50).

Research on statistical and mathematical approaches to HIV dynamics has been extensive over the last 7 years. Several volumes devoted to issues of importance to HIV/AIDS dynamics including parameter estimation, short-term predictions, forecasting, social dynamics, and immunology have appeared over the last several years (19,51,52). The recent book by Hethcote and Van Ark (53) provides a detailed data-driven study of HIV dynamics, and the encyclopedic book by Anderson and May (7) gives a panoramic view of the growing field of theoretical epidemiology with emphasis on the extensive contributions of Anderson, May, and their collaborators.

Although the contact structure of a population is one of the main factors influencing the incidence of sexually transmitted diseases, it has proved difficult to determine population mixing patterns from observable data. Even in situations where reasonable samples have been drawn from selected target populations (such as college students, bar patrons, or participants in drug treatment programs), members of the target group interact significantly with members outside the target group. This latter situation is problematic for existing models that implicitly assume that the populations are closed, that is, all social or sexual contacts occur within the groups specified in the model. To enable such models to be employed, a procedure is needed to "close" the population on the basis of incomplete observations.

This chapter proposes a new approach to the estimation of nonrandom contact patterns that explicitly recognizes the interaction between members of a target population (who can be sampled) and individuals in nontarget populations (who cannot be sampled). Included in our contribution is a method for estimating the size of the nontarget population that is interacting with the target population sampled. This allows us to construct a pattern of interactions among target and nontarget populations that is consistent with known axioms of population mixing. Our technical work developed alongside our empirical study of dating and sexual activity among college students (54–57). This survey reveals that random mating is not descriptively accurate, and highlights convincingly the strength of the social or sexual interaction between the target and nontarget populations. Although our study may not be representative of all U.S. college students, the features we infer from our sample via our modeling approach are in line with our daily unscientific observations (e.g., strong within-class mixing, women prefer to mix with older men). The development of a methodology that incorporates these features into models of STD transmission dynamics is the goal of this chapter.

We proceed as follows: First we describe a general axiomatic approach for modeling contact processes in heterogeneous mixing populations. The next section employs this framework to model dating, sexual mixing, and pair formation in the context of heterosexually active populations. The data structure used to illustrate our approach to constructing mixing matrices is described next. Then, we present a mark-recapture model for estimating the size of the nontarget population that interacts with our sampled target population. We next reduce the problem of completely specifying mixing matrices to that of estimating a single parameter, and illustrate our completion algorithm using the data from our survey of college undergraduates. Finally we summarize our results and discuss potential applications of the algorithm.

MIXING BETWEEN I INTERACTING SUBPOPULATIONS

Busenberg and Castillo-Chavez (8,9) have shown that all mixing structures in which individuals interact with members of all subpopulations can be expressed as a multiplicative perturbation of proportionate mixing. In this section we briefly summarize their result using a population comprising I distinct types or groups. The ith group has $T_i(t)$ individuals at time t and an average number of C_i partners per person per unit time. The social/sexual contact structure of the population is modeled by an $I \times I$ matrix of probabilities $P(t)$, where $P_{ij}(t)$ gives the probability that a partner selected by a sexually active individual at time t in group i is a member of group j. The matrix $P(t)$ must satisfy the following constraints or mixing axioms:

(A1) $P_{ij}(t) \geq 0$ for $1 \leq i, j \leq I$ and all t.

(A2) $\sum_{j=1}^{I} P_{ij}(t) = 1$ for $1 \leq i \leq I$ and all t.

(A3) $C_i T_i(t) P_{ij}(t) = C_j T_j(t) P_{ji}(t)$ for $1 \leq i, j \leq I$ and all t

Constraints (A1) and (A2) make P a stochastic matrix, and (A3) guarantees conservation in the number of new pairings/contacts per unit time between types. Busenberg and Castillo-Chavez's representation theorem states that any P that satisfies the constraints (A1)–(A3) may be written in the form:

$$P_{ij} = \bar{P}_j \left[\frac{Q_i Q_j}{V} + \phi_{ij} \right] \text{ for } 1 \leq i, j \leq I \qquad [1]$$

where

$$\bar{P}_j = \frac{C_j T_j}{\sum_{k=1}^{I} C_k T_k} \text{ for } 1 \leq j \leq I \qquad [2]$$

represents random or proportionate mixing between groups,

$$Q_i = 1 - \sum_{k=1}^{I} \bar{P}_k \phi_{ik} \text{ for } 1 \leq i \leq I, \qquad [3]$$

$$V = \sum_{k=1}^{I} \bar{P}_k Q_k, \qquad [4]$$

and $\phi = \{\phi_{ij}\}$ is an $I \times I$ symmetric matrix. The matrix ϕ is a measure of mutual preference or affinity for sexual partners between pairs of groups (33,36,58–60). Specific preference structures are determined by the elements of the ϕ matrix. For example, following Blythe and Castillo-Chavez (61), we may parametrize ϕ as follows:

1. Each ϕ_{ij} can take one of only two values, a or b, where $0 \leq b \leq a < 1$.
2. All the elements in each diagonal or off-diagonal of the ϕ matrix are the same; for example, for $I = 4$, the ϕ matrices may look like:

$$\phi = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix} \text{ or } \phi = \begin{bmatrix} a & a & b & b \\ a & a & a & b \\ b & a & a & a \\ b & b & a & a \end{bmatrix}$$

This restriction on the mixing parameters $\{\phi_{ij}\}$ gives us a mixing framework (the function on the right-hand side of Equation [1]) that is fairly simple (only two values are used to describe the ϕ matrix) and capable of considerable flexibility. We note that if $a = b$ we recover proportionate mixing, whereas a flexible form of like-with-like mixing is obtained with the parametrization $\phi_{ij} = a$ if $i = j$, $\phi_{ij} = b$ otherwise.

Multigroup models for STDs have been studied by a variety of groups including Lajmanovich and Yorke (62), Jacquez et al. (45), Castillo-Chavez et al. (22), Huang et al. (25), and Huang (63). Many researchers have been satisfied with the use of proportionate mixing (Equation [2]) or preferred mixing because of their

mathematical simplicity (but see refs. 22,25,63). Since our main objective in this chapter is to present our method for determining the shape of the mixing matrix, it is important to keep in mind as reference models the shapes of the proportionate mixing and of the preferred mixing matrix. The latter matrix is given by

$$P_{ij} = h_i \delta_{ij} + (1 - h_i) \frac{(1 - h_j)\,\overline{P}_j}{\sum_{k=1}^{I} (1 - h_k)\,\overline{P}_k}.$$

where the h_i's are nonnegative constants between 0 and 1. These constants represent the proportions of group contacts/partnerships that are "reserved" for within-group mixing. The term δ_{ij} equals 1 if $i = j$ and 0 otherwise. Consequently, those partnerships that are not reserved for within-group mixing are assumed to follow proportionate mixing. Figure 1 illustrates the shape of a random or proportionate mixing matrix ($h_i = 0$ for all i) obtained from the aggregated data presented below in the Example section with corresponding pair-formation parameter equal to 2. Figure 2 shows a preferred mixing matrix with $h_i = 0.2$ for all i, a "diagonal" perturbation from Fig. 1. Another perturbation is presented in Fig. 3, using $h_1 = h_5 = 0.4$, $h_2 = h_4 = 0.3$, and $h_3 = 0.2$. Preferred mixing, as shown by Blythe and Castillo-Chavez (64), corresponds to the (frequency dependent) preference function $\phi_{ij} = h_i \delta_{ij} / \overline{P}_i$. Thus, to maintain a fixed proportion of contacts with one's group regardless of the population dynamics, individuals must continually adjust their preference. This again highlights the deficiencies of this model.

Finally, because the data collected come from two-sex sexual or dating interactions, we are forced to modify the framework of this section to include this added social structure. This is the topic of the next section in this chapter.

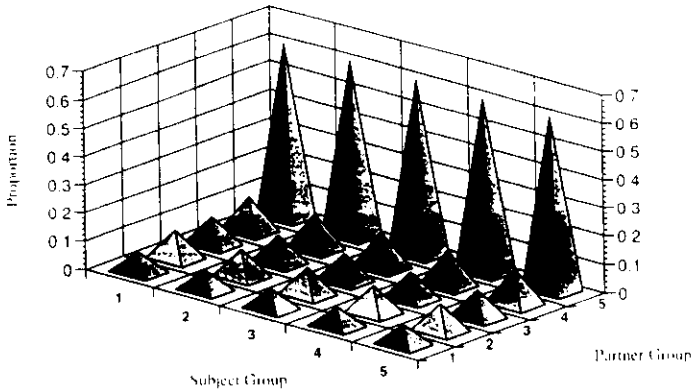

FIG. 1. Example graph of one-sex random mixing.

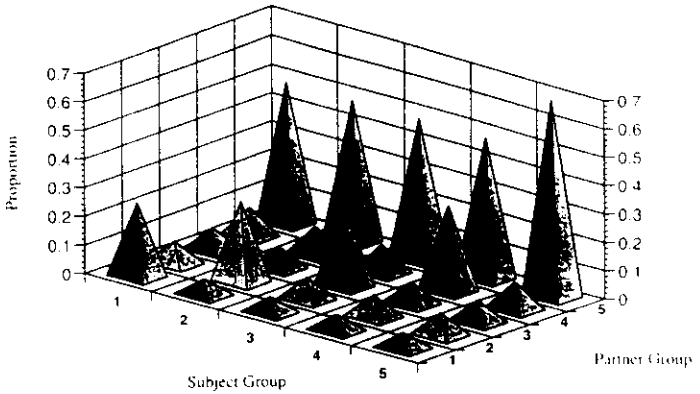

FIG. 2. Example graph 1 of one-sex preferred mixing.

TWO-SEX MIXING STRUCTURES

In this section, we introduce two-sex mixing structures in a heterosexually active population with $(I + J)$ groups. The notation is similar to that above, except that we use superscripts m and f and subscripts i and j for men and women, respectively. This population is divided into groups or subpopulations that are defined by gender and possibly, race, socioeconomic background, average degree of sexual activity,

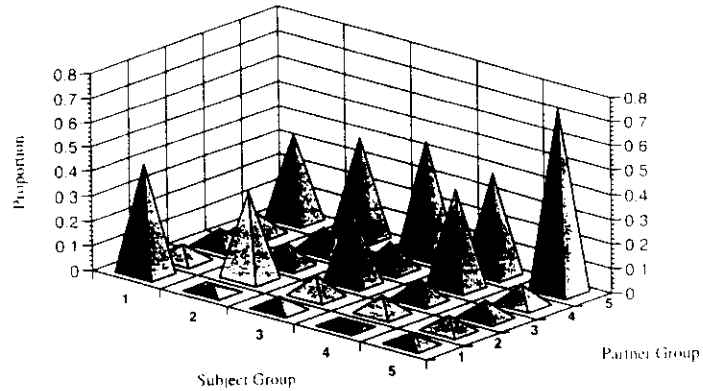

FIG. 3. Example graph 2 of one-sex preferred mixing.

and so forth. We consider I sexually active groups of men and J sexually active groups of women. The following definitions are needed:

$P_{ij}^m(t)$ = fraction of partnerships of men in group i with women in group j at time t

$P_{ji}^f(t)$ = fraction of partnerships of women in group j with men in group i at time t

C_i^m = average (constant) number of female partners per men in group i per unit time, or the pair-formation rate of ith male group

C_j^f = average (constant) number of male partners per women in group j per unit time, or the pair-formation rate of jth female group.

The set of mixing probabilities $\{P_{ij}^m(t)$ and $P_{ji}^f(t) : i = 1, \ldots, I$ and $j = 1, \ldots, J\}$ establishes the mixing/pair-formation structure in heterosexually active populations provided they satisfy the following definition:

Definition: $\{P_{ij}^m(t), P_{ji}^f(t)\}$ is called a mixing/pair-formation matrix if and only if it satisfies the following properties at all times:

(B1) $0 \le P_{ij}^m(t) \le 1$ and $0 \le P_{ji}^f(t) \le 1$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and all t.

(B2) $\sum_{j=1}^{J} P_{ij}^m(t) = 1$ for $i = 1, \ldots, I$ and all t; $\sum_{i=1}^{I} P_{ji}^f(t) = 1$ for $j = 1, \ldots, J$ and all t.

(B3) $C_i^m T_i^m(t) P_{ij}^m(t) = C_j^f T_j^f(t) P_{ji}^f(t)$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and all t.

Property (B3) can be interpreted as a conservation of partnership-formation rates between two groups. A useful particular solution is the Ross solution, which corresponds to proportionate mixing in the context of heterosexually active populations. The Ross solution is denoted by $\{\bar{P}_i^m, \bar{P}_j^f\}$, where

$$\bar{P}_j^m = \frac{C_j^f T_j^f}{\sum_{i=1}^{I} C_i^m T_i^m} \quad \text{and} \quad \bar{P}_i^f = \frac{C_i^m T_i^m}{\sum_{j=1}^{J} C_j^f T_j^f} \tag{6}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, I$. Note that $\sum_{i=1}^{I} C_i^m T_i^m(t) = \sum_{j=1}^{J} C_j^f T_j^f(t)$ at all times by (B3). All solutions to axioms (B1)–(B3) can be generated as multiplicative perturbations of the Ross solution. Figures 4 and 5 illustrate the shape of feasible male and female random mixing matrices generated from our survey data under the assumption of heterosexual random mixing. The real mixing matrices for the first four groups, using the same pair-formation parameter, are sketched in Figs. 6 and 7. It is clear that this sample from the target population does not mix at random. To describe nonrandom mixing in mathematical terms, that is, all perturbations of the Ross solution satisfying (B1)–(B3), we need the following definitions:

ϕ_{ij}^m = measure of preference that group i men have for group j women, $i = 1, \ldots, I$ and $j = 1, \ldots, J$

$l_i^m = \sum_{k=1}^{J} \bar{P}_k^m \phi_{ik}^m$ = weighted average preference of group i men, $i = 1, \ldots, I$

$Q_i^m = 1 - l_i^m$, $i = 1, \ldots, I$.

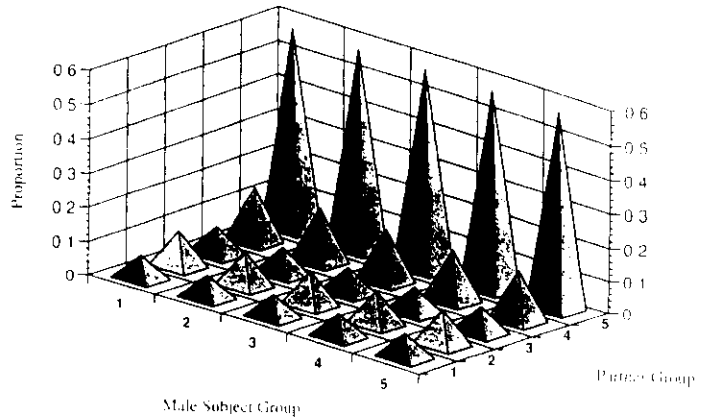

FIG. 4. Example graph of two-sex random mixing for men

We require at all times that $0 \le Q_i^m \le 1$ and that

$$\sum_{i=1}^{I} l_i^m \bar{P}_i^f = \sum_{i=1}^{I} \sum_{k=1}^{J} \bar{P}_k^m \phi_{ik}^m \bar{P}_i^f < 1. \tag{7}$$

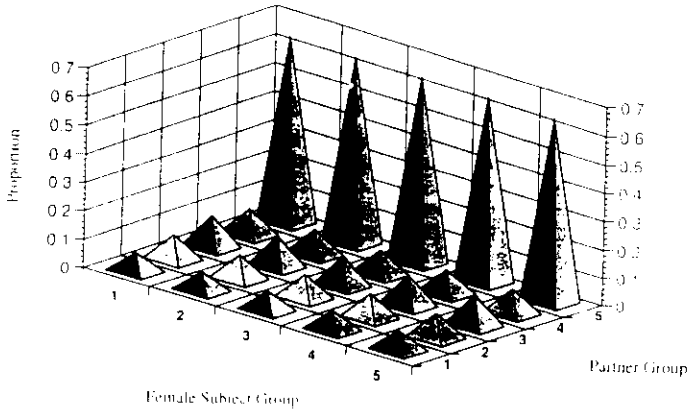

FIG. 5. Example graph of two-sex random mixing for women.

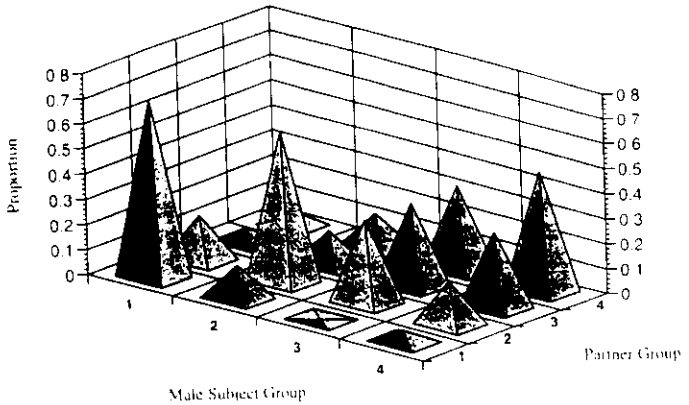

FIG. 6. Example graph of two-sex mixing for men.

Similarly, we let

ϕ_{ji}^{f} = measure of preference that group j women have for group i men, $j = 1, \ldots, J$ and $i = 1, \ldots, I$

$l_j^f = \sum_{i=1}^{I} \bar{P}_i^f \phi_{ji}^f$ = weighted average preference of group j women, $j = 1, \ldots, J$

$Q_j^f = 1 - l_j^f$, $j = 1, \ldots, J$.

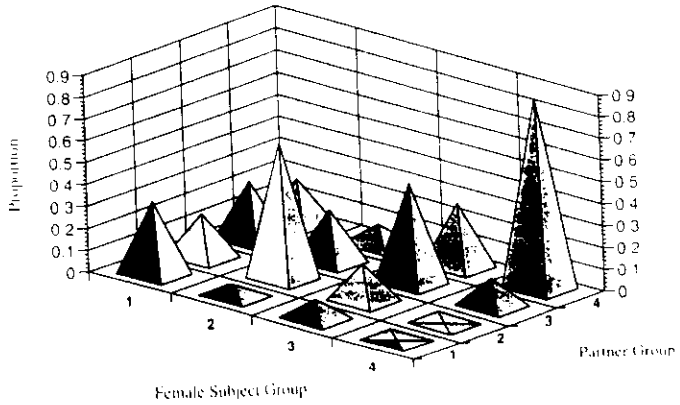

FIG. 7. Example graph of two-sex mixing for women.

Again, we require at all times that $0 \le Q_j^f \le 1$ and that

$$\sum_{j=1}^{J} l_j^f \bar{P}_i^m = \sum_{j=1}^{J} \sum_{i=1}^{I} \bar{P}_i^f \phi_{ji}^f \bar{P}_i^m < 1. \qquad [8]$$

Using the above notation, Castillo-Chavez and Busenberg (35) have shown that all solutions to axioms (B1)–(B3) are given by the following multiplicative perturbations to the Ross solution $\{\bar{P}_j^m, \bar{P}_i^f\}$:

$$P_{ji}^m = \bar{P}_i^m \left[\frac{Q_j^f Q_i^m}{\sum_{l=1}^{I} \bar{P}_l^f Q_l^m} + \phi_{ji}^m \right] \text{ and } P_{ji}^f = \bar{P}_i^f \left[\frac{Q_i^m Q_j^f}{\sum_{k=1}^{J} \bar{P}_k^m Q_k^f} = \phi_{ji}^f \right] \qquad [9]$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$. Their theorem explicitly states:

Theorem: Let $\{\phi_{ji}^m\}$ and $\{\phi_{ji}^f\}$ be two nonnegative matrices. Let $l_i^m = \sum_{k=1}^{J} \bar{P}_k^m \phi_{ik}^m$ and $l_j^f = \sum_{l=1}^{I} \bar{P}_l^f \phi_{jl}^f$, where $\{\bar{P}_j^m, \bar{P}_i^f; j = 1, \ldots, J \text{ and } i = 1, \ldots, I\}$ denotes the Ross solution. Let $Q_i^m = 1 - l_i^m$, $i = 1, \ldots, I$ and $Q_j^f = 1 - l_j^f$, $j = 1, \ldots, J$. If ϕ_{ji}^m and ϕ_{ji}^f are chosen in such a way that $0 \le Q_i^m \le 1$, $0 \le Q_j^f \le 1$, $\sum_{i=1}^{I} l_i^m \bar{P}_i^f < 1$, and $\sum_{j=1}^{J} l_j^f \bar{P}_j^m < 1$,

then

$$\phi_{ji}^m = \phi_{ji}^f + Q_i^m Q_j^f \left[\frac{\sum_{k=1}^{J} \bar{P}_k^m Q_k^f - \sum_{l=1}^{I} \bar{P}_l^f Q_l^m}{\left(\sum_{l=1}^{I} \bar{P}_l^f Q_l^m \right) \left(\sum_{k=1}^{J} \bar{P}_k^m Q_k^f \right)} \right] \qquad [10]$$

if and only if all solutions to axioms (B1)–(B3) are given by Equation [9].

Although the above representation theorem looks complicated, we can easily use it to generate many solutions with only one or two parameters. It is possible to generate the type of mixing observed in the data used below to test our algorithm because Hsu Schmitz et al. (65) have shown that all parametrizations for $\{\phi_{ji}^m\} = \{\phi_{ji}^f\}^T$ are legitimate (i.e., they satisfy all the conditions of the above theorem including Equation [10]). This result immediately allows the generation of a rich and flexible class of parametric solutions. However, we will not pursue this direction in this chapter. In general the assumption that C_j^m and C_i^f are constant is inconsistent and (B3) must be modified. This is easily done (see ref. 65).

DATA STRUCTURE OF NONCLOSED NETWORKS

The mixing structures discussed earlier are applicable to closed populations by the implicit assumption that all population groups are captured in the model. For data collected from the real world, the population covered probably is not closed. Usually the data cover not only the target population but also the nontarget population. If the nontarget population plays a considerable role in the network, then we

should not ignore it. Without direct information on the nontarget population, the mixing matrices are not complete, and demographic dynamics and disease transmission cannot be predicted correctly. Therefore, the issue of how to obtain indirect information on the nontarget populations and their effect on network interactions must be addressed before further study. In this section, we describe the potential data structure of nonclosed two-sex mixing populations. Then we conditionally "close" the network and complete the mixing matrices. An illustrative example is provided last.

Following the notation presented earlier, we let the Ith male group and the Jth female group consist of individuals from the nontarget populations (i.e., they are members of an unobservable subpopulation). The first $I-1$ male groups and the first $J-1$ female groups are composed of men and women from the target populations, respectively. Suppose we are interested in the heterosexual contact structure of a given target population at a given time and we know the sizes of the target male groups, $R_i^m (i = 1, \ldots, I-1)$, and of the target female groups, $R_j^f (j = 1, \ldots, J-1)$. To gather data for this study, we do stratified sampling at a given time to randomly select respondents from those $I-1$ male groups and those $J-1$ female groups to our questionnaire. The questions concerning a given time period (our time unit) are: if they were sexually active or not; if yes, how many distinct partners they had; and how many of those partners belonged to different target and nontarget groups. In our data the term "sexually active" means having sexual contacts during the time period. The data are represented by the following notation:

S_i^m: sample size of ith target male group, $i = 1, \ldots, I-1$

S_j^f: sample size of jth target female group, $j = 1, \ldots, J-1$

A_i^m: number of sexually active individuals among S_i^m

A_j^f: number of sexually active individuals among S_j^f

Y_{ik}^m: number of distinct female partners of individual k in A_i^m

Y_{jr}^f: number of distinct male partners of individual r in A_j^f

X_{ik}^m: among Y_{ik}^m, number of distinct female partners from the target population

X_{jr}^f: among Y_{jr}^f, number of distinct male partners from the target population

U_{ijk}^m: among Y_{ik}^m, number of distinct female partners from group j

U_{jir}^f: among Y_{jr}^f, number of distinct male partners from group i.

We can summarize the data by

$$Y_i^m = \sum_{k=1}^{A_i^m} Y_{ik}^m = \text{total number of female partners of individuals in } A_i^m$$

$$Y_j^f = \sum_{r=1}^{A_j^f} Y_{jr}^f = \text{total number of male partners of individuals in } A_j^f$$

$$X_i^m = \sum_{k=1}^{A_i^m} X_{ik}^m = \text{among } Y_i^m, \text{ total number of female partners from the target population}$$

$$X_j^f = \sum_{r=1}^{A_j^f} X_{jr}^f = \text{among } Y_j^f, \text{ total number of male partners from the target population}$$

$$Y_+^m = \sum_{i=1}^{I-1} Y_i^m = \text{total number of female partners of all sampled sexually active men}$$

$$Y_+^f = \sum_{j=1}^{J-1} Y_j^f = \text{total number of male partners of all sampled sexually active women}$$

$$X_+^m = \sum_{i=1}^{I-1} X_i^m = \text{among } Y_+^m, \text{ total number of female partners from the target population}$$

$$X_+^f = \sum_{j=1}^{J-1} X_j^f = \text{among } Y_+^f, \text{ total number of male partners from the target population,}$$

$$U_{ij}^m = \sum_{k=1}^{A_i^m} U_{ijk}^m = \text{among } Y_i^m, \text{ total number of female partners from group } j.$$

$$U_{ji}^f = \sum_{r=1}^{A_j^f} U_{jir}^f = \text{among } Y_j^f, \text{ total number of male partners from group } i.$$

We can obtain point estimates of the average number of partners per person per unit time and of the entries in the mixing matrix as follows:

$C_i^m = Y_i^m / A_i^m$ = average number of female partners per sexually active man in group i per unit time, $i = 1, \ldots, I-1$

$C_j^f = Y_j^f / A_j^f$ = average number of male partners per sexually active woman in group j per unit time, $j = 1, \ldots, J-1$

$P_{ij}^m = U_{ij}^m / Y_i^m$ = fraction of sexual contacts of men in group i with women in group j at the given time, $i = 1, \ldots, I-1$ and $j = 1, \ldots, J$

$P_{ji}^f = U_{ji}^f / Y_j^f$ = fraction of sexual contacts of women in group j with men in group i at the given time, $j = 1, \ldots, J-1$ and $i = 1, \ldots, I$.

The matrix $\{P_{ij}^m, P_{ji}^f\}$ from the above data structure is not complete because we do not have a closed network: individuals in the nontarget male and female populations were not surveyed, so the rows $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$ are missing. Below we show how to conditionally "close" the network and complete this matrix but do not guarantee that $U_{ij}^m = U_{ji}^f$ as required by theory. The problem arises from the fact that we are dealing with a sample and not a census (this is evident in Tables 2 and 3). Usually data satisfy axioms (B1) and possibly (B2) but not axiom (B3). The same is true for estimates of $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$. In general (B3) is formulated with C_i^m and C_j^f functions of the number or proportions of sexually active individuals (see ref. 65).

MARK-RECAPTURE METHODOLOGY FOR ESTIMATING NONTARGET POPULATION SIZES

We assume that the group sizes in the target population are known. However, the sizes of the sexually active subgroups, T^g_ν ($g = m$ or f, $\nu = 1, \ldots, I-1$ or $J-1$), are not known. The assumption that all individuals in these groups were sexually active is certainly not realistic. A natural way to estimate the sexually active group sizes is given by the following formulas:

$$\hat{T}^m_i = R^m_i \times (A^m_i/S^m_i) \text{ and } \hat{T}^f_j = R^f_j \times (A^f_j/S^f_j), \quad [11]$$

where R^m_i and R^f_j denote the known target group sizes, $i = 1, \ldots, I-1$ and $j = 1, \ldots, J-1$. In fact, these estimators are the maximum likelihood estimators (see ref. 55). Because the survey's definition of sexually active is tied up with a specific time period (a narrow definition), individuals in the target population who are sexually active but did not have sexual contacts during that specific period do not contribute to the above estimates. Since we do not have direct information on the nontarget male and female populations, the size of their sexually active subsets have to be estimated by other methods. Rubin et al. (55) introduced modified mark-recapture methods to obtain conditional estimates of the size of these subsets. The general procedure is summarized in two steps below.

First mark a random sample of size n_1 from a population of size N (unknown) and release them. After a certain period of time, the second step that collects a random sample of size n_2 from the same population is enforced. The number of marked individuals in this second sample is denoted by m_2. Bailey (66) introduced the binomial model as a useful approximation to the classic hypergeometric model that arises when only a single capture is possible after marking. His model is given by the expression

$$\mathbf{P}(m_2|n_1, n_2) \approx \binom{n_2}{m_2} \left(\frac{n_1}{N}\right)^{m_2} \left(1 - \frac{n_1}{N}\right)^{n_2 - m_2}. \quad [12]$$

If individuals are sighted by observers, instead of physically captured, then since different observers may sight the same individuals, individuals in the population may be sampled with replacement. If sampling is done with replacement, then the binomial model holds exactly (see ref. 67). Because the maximum likelihood estimator for N, namely $N^* = n_1 n_2/m_2$ (the Lincoln-Petersen estimator), is biased, Bailey (66) suggested the following estimator for N, $\hat{N}$, and for its variance, $\hat{v}(\hat{N})$:

$$\hat{N} = \frac{n_1 (n_2 + 1)}{m_2 + 1}, \quad [13]$$

$$\hat{v}(\hat{N}) = \frac{(n_1)^2 (n_2 + 1)(n_2 - m_2)}{(m_2 + 1)^2 (m_2 + 2)}. \quad [14]$$

These estimators are less biased with proportional biases of order $\exp(-n_1 n_2/N)$ and $(n_1 n_2/N)^2 \exp(-n_1 n_2/N)$, respectively. For the data structure described earlier, we assume that all sexually active individuals of a given gender g in the target population are marked and those in the sample constitute the first sample of size T^g_+. Individuals sampled who were sexually active serve as observers who "sight" their partners of the other gender by sexual/social contact. Thus, those partners from the target (marked) and nontarget (unmarked) populations constitute the second sample. Because different observers may have the same sexual partners, the second sampling procedure for partners must be done with replacement. Bailey's binomial model is exact in this case and hence more appropriate for our data. By Equation [13], the estimates of the total number of sexually active individuals in the target and nontarget populations for both genders are:

$$\hat{N}^m = \frac{T^m_+ (Y^f_+ + 1)}{X^f_+ + 1} \text{ and } \hat{N}^f = \frac{T^f_+ (Y^m_+ + 1)}{X^m_+ + 1}. \quad [15]$$

Note that the information in the second sample is from observers of the other gender. The estimated variances of $\hat{N}^m$ and $\hat{N}^f$ are analogous to those in Equation [14]. However, T^m_+ and T^f_+ are not known in our case. We estimate them as

$$\hat{T}^m_+ = \sum_{i=1}^{I-1} \hat{T}^m_i \text{ and } \hat{T}^f_+ = \sum_{j=1}^{J-1} \hat{T}^f_j \quad [16]$$

and use these estimates to obtain the maximum likelihood estimates of the total number of sexually active individuals, $\tilde{N}^g$:

$$\tilde{N}^m = \frac{\hat{T}^m_+ (Y^f_+ + 1)}{X^f_+ + 1} \text{ and } \tilde{N}^f = \frac{\hat{T}^f_+ (Y^m_+ + 1)}{X^m_+ + 1}. \quad [17]$$

The estimated variances of $\tilde{N}^m$ and $\tilde{N}^f$ are provided by Rubin et al. (55). These variances incorporate the additional variation due to $\hat{T}^m_+$ and $\hat{T}^f_+$. Since $\tilde{N}^m$ and $\tilde{N}^f$ include sexually active individuals from the target and nontarget populations, that is, $\tilde{N}^m = \hat{T}^m_+ + \hat{T}^m_i$ and $\tilde{N}^f = \hat{T}^f_+ + \hat{T}^f_j$, the estimated sizes of sexually active nontarget populations are

$$\hat{T}^m_i = \tilde{N}^m - \hat{T}^m_+ = \frac{\hat{T}^m_+ (Y^f_+ - X^f_+)}{X^f_+ + 1} \text{ and } \hat{T}^f_j = \tilde{N}^f - \hat{T}^f_+ = \frac{\hat{T}^f_+ (Y^m_+ - X^m_+)}{X^m_+ + 1}. \quad [18]$$

However, sexually active individuals in the nontarget population will not be "sighted" if they did not engage in sexual activity with individuals from the target population. Therefore, $\hat{T}^m_i$ and $\hat{T}^f_j$ are conditional estimates that count only those individuals in the nontarget population who had at least one sexual contact with partners from the target population of the opposite gender during the surveyed period. This is one of the first data sets of this type, and its limitations may be perceived as too strong. However, this data structure brings to the forefront the even stronger limitations that are implicit in current mathematical and statistical models.

We also observe that individuals with high contact rates are more likely to be "sighted." It is nearly impossible to modify the sampling procedure to take into account this effect. An alternative approach is to modify these equations to incorpo-

rate the effects of the biases. Modifications should be closely connected to data and, consequently, to survey design. Our data, used in the Example section, do not seem to be seriously affected by this source of bias as the "average" contact rates of the interacting subpopulations do not vary that much.

COMPLETION OF THE MIXING MATRIX

We assume that the combination of our target and nontarget sexually active populations of both genders constitute a closed mixing network and hence its associated mixing matrix satisfies axioms (B1)–(B3). Using the mark-recapture methodology described above, we obtain conditional estimates for T_i^m and T_j^f. However, the rows $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$ and the averages C_i^m and C_j^f are not yet known. We can obtain point estimates of these unknown parameters by assuming that the data are consistent with the above two-sex mixing framework, which reduces the estimation of all the unknown parameters to that of estimating a single pair-formation parameter. Then the shape of the complete mixing matrix can be calculated from the data. The procedure described in this section is an alternative version of that in ref. 57.

First we sum over j on both sides of the equation in axiom (B3) and obtain

$$C_i^m \hat{T}_i^m = K_i^f + C_J^f \hat{T}_J^f P_{Ji}^f, \tag{19}$$

where $K_i^f = \sum_{j=1}^{J-1} C_j^f \hat{T}_j^f P_{ji}^f$ is nonnegative (because C_j^f and $\hat{T}_j^f$ are positive, and P_{ji}^f are nonnegative) and can be computed from the data (it is therefore known). Rearranging Equation [19] for $i = I$ yields

$$C_J^f \hat{T}_J^f P_{JI}^f = C_I^m \hat{T}_I^m - K_I^f \geq 0, \tag{20}$$

which leads to a lower bound for C_I^m:

$$C_I^m \geq K_I^f / \hat{T}_I^m. \tag{21}$$

Similarly, we sum over i on both sides of the equation in axiom (B3) and obtain

$$C_j^f \hat{T}_j^f = K_J^m + C_I^m \hat{T}_I^m P_{Ij}^m, \tag{22}$$

where $K_J^m = \sum_{i=1}^{I-1} C_i^m \hat{T}_i^m P_{ij}^m$ also is known and nonnegative. Rearranging Equation [22] for $j = J$ yields

$$C_I^m \hat{T}_I^m P_{IJ}^m = C_J^f \hat{T}_J^f - K_J^m \geq 0, \tag{23}$$

which gives a lower bound for C_J^f:

$$C_J^f \geq K_J^m / \hat{T}_J^f. \tag{24}$$

Since only those sexually active individuals who had at least one sexual contact during the surveyed time period are under consideration, C_I^m and C_J^f must be greater than or equal to one. Therefore, we can use the following refined lower bounds

$$C_I^m \geq \max(K_I^f / \hat{T}_I^m, 1) \text{ and } C_J^f \geq \max(K_J^m / \hat{T}_J^f, 1). \tag{25}$$

To find the relationship between C_I^m and C_J^f, we sum over i and j on both sides of the equation in axiom (B3) and obtain

$$\sum_{i=1}^{I} C_i^m \hat{T}_i^m = \sum_{j=1}^{J} C_j^f \hat{T}_j^f, \tag{26}$$

or equivalently

$$C_I^m \hat{T}_I^m - C_J^f \hat{T}_J^f = G^f - G^m, \tag{27}$$

where $G^m = \sum_{i=1}^{I-1} C_i^m \hat{T}_i^m$ and $G^f = \sum_{j=1}^{J-1} C_j^f \hat{T}_j^f$. Both can be computed from data.

Because of insufficient information in the data, there is no way of estimating unique values for C_I^m and C_J^f. But if one of these two parameters is known, then the other one can be uniquely obtained through Equation [27]. Without independent estimators for C_I^m or C_J^f, estimates of the rows $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$ are not possible. Estimation of all the unknown parameters must be conditioned on the assumption that either C_I^m or C_J^f is known. If we assume that $\tilde{C}_I^m$ is an appropriate value for C_I^m, the pair-formation parameter, then

$$\tilde{C}_J^f = (\tilde{C}_I^m \hat{T}_I^m - G^f + G^m)/\hat{T}_J^f. \tag{28}$$

Plugging $\tilde{C}_I^m$ and $\tilde{C}_J^f$ into Equations [19] and [22] specifies the values of P_{Ji}^f and P_{Ij}^m, respectively:

$$\tilde{P}_{Ji}^f = \begin{cases} (C_i^m \hat{T}_i^m - K_i^f) / (\tilde{C}_J^f \hat{T}_J^f) \text{ for } i = 1, \ldots, I-1 \\ (\tilde{C}_I^m \hat{T}_I^m - K_I^f) / (\tilde{C}_J^f \hat{T}_J^f) \text{ for } i = I \end{cases} \tag{29}$$

$$\tilde{P}_{Ij}^m = \begin{cases} (C_j^f \hat{T}_j^f - K_j^m) / (\tilde{C}_I^m \hat{T}_I^m) \text{ for } j = 1, \ldots, J-1 \\ (\tilde{C}_J^f \hat{T}_J^f - K_J^m) / (\tilde{C}_I^m \hat{T}_I^m) \text{ for } j = J. \end{cases} \tag{30}$$

By Equations [28], [29], and [30], the first derivatives of $\tilde{P}_{IJ}^m$ and $\tilde{P}_{JI}^f$ with respect to $\tilde{C}_I^m$ are:

$$\frac{\partial}{\partial \tilde{C}_I^m} \tilde{P}_{IJ}^m = \frac{G^f - G^m + K_J^m}{(\tilde{C}_I^m)^2 \hat{T}_I^m}, \tag{31}$$

$$\frac{\partial}{\partial \tilde{C}_I^m} \tilde{P}_{JI}^f = \frac{\hat{T}_I^m (G^m - G^f + K_I^f)}{(\tilde{C}_I^m \hat{T}_I^m - G^f + G^m)^2}. \tag{32}$$

Clearly the sign of $G^f - G^m + K_J^m$ determines if $\tilde{P}_{IJ}^m$ increases or decreases with $\tilde{C}_I^m$, and the sign of $G^m - G^f + K_I^f$ determines the behavior of $\tilde{P}_{JI}^f$. If $\tilde{P}_{IJ}^m$ increases with $\tilde{C}_I^m$, then some $\tilde{P}_{Ij}^m$ $(j = 1, \ldots, J-1)$ must decrease by axiom (B2). Similarly, if $\tilde{P}_{JI}^f$ increases with $\tilde{C}_I^m$, then some $\tilde{P}_{Ji}^f$ $(i = 1, \ldots, I-1)$ must decrease.

Thus, once we know the pair-formation parameter $\tilde{C}_I^m$, that is, the average number of partners per male in the nontarget population, we can obtain the average number of partners per female in the nontarget population, $\tilde{C}_J^f$, and the mixing proportions for nontarget populations, $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$. The mixing matrix is now completed under the condition that all the sexually active individuals in the non-

target populations have at least one sexual contact with individuals in the target populations. In general Equation [26] does not hold and in fact is not necessary (see ref. 65).

Recall that we assume the combination of our target and nontarget sexually active populations of both genders constitutes a closed network with a mixing matrix satisfying axioms (B1)–(B3). However, data collected from the real world may violate axiom (B3), and may result in estimated values of $\{P_{ij}^m\}$ and $\{P_{ji}^f\}$ that do not satisfy axioms (B1) or (B3). Since our main objective is to roughly determine the shape of the mixing matrix for a real population from a single sample, this violation is tolerated until a better method is developed.

EXAMPLE

This example deals with the surveyed sexual behavior of college students as reported in refs. 49 to 52. The target populations are male and female college students in a given university. Students of each gender are categorized by school year into four groups: 1 (freshman), 2 (sophomore), 3 (junior), and 4 (senior). In addition to these four groups, one more group, here referred to as "other," accounts for their partners who do not belong to the target population. The sizes of groups 1 through 4 for both genders are known because they are available from the university registrar's office.

Table 1 lists the group sizes (R), sample sizes (S), sexually active subsample sizes (A), sexually active proportions in the samples ($A \div S$), and estimated sexually active subgroup sizes ($\hat{T}$), all rounded to integers. Table 1 also includes the sums of the four groups in the target population. The observed overall sexually active proportion for male students is 34.1%, which is significantly smaller than the observed overall sexually active proportion of 43.5% for female students (one-sided $p = .015$). The sexual partnership distribution (U), the mixing proportions (P), and the

TABLE 1. *Population sizes and sample sizes for men[a] and women[b]*

Group i/j	Population size R	Sample size S	Sexually active subsample size A	Sexually active proportion $A \div S$	Estimated sexually active subpopulation size $\hat{T}$
1	1,673	79	16	0.203	339
	1,278	68	20	0.294	376
2	1,589	60	24	0.400	636
	1,308	68	26	0.382	500
3	1,591	63	20	0.317	505
	1,277	61	36	0.590	754
4	1,686	47	25	0.532	897
	1,348	56	28	0.500	674
Total	6,539	249	85	0.341	2,377 ($\hat{T}_+$)
	5,211	253	110	0.435	2,304

[a]Upper line.
[b]Lower line.

TABLE 2. *Sexual partnership distribution of male students by counts[a] and proportions[b]*

Male group i	Female partner group j							
	1	2	3	4	Subtotal X_i^m	5 (Other)	Total Y_i^m	Average C_i^m
1	12	3	1	1	17	10	27	1.69
	0.444	0.111	0.037	0.037		0.370		
2	2	9	2	2	15	17	32	1.33
	0.063	0.281	0.063	0.063		0.531		
3	0	4	4	4	12	14	26	1.30
	0.000	0.154	0.154	0.154		0.538		
4	1	4	7	11	23	12	35	1.40
	0.029	0.114	0.200	0.314		0.343		
Total	15	20	14	18	67(X_+^m)	53	120(Y_+^m)	1.41
	0.125	0.167	0.117	0.150		0.442		

[a]U_{ij}^m, upper line.
[b]P_{ij}^m, lower line.

total (Y) and average (C) number of distinct partners for sexually active male and female students are presented in Tables 2 and 3, respectively. Male students have a higher overall average number of distinct sexual partners (1.41) than female students (1.24); however, the difference is not significant (two-sided $p = .147$). The overall proportion of sexual relationships with partners of group 5 (other) is 44.2% for men and 50.0% for women. Hence, the interactions with members of group 5 should not be ignored in the study of the effects of mixing patterns on the dynamics of sexually transmitted diseases.

To quantify the potential effect that individuals in group 5 may have on disease transmission, we need to estimate the elements of the last rows of the mixing matrix. The incomplete male and female mixing matrices are plotted in Figs. 8 and 9, respectively. Despite the fact that these figures ignore the effects of group 5, they still show strong evidence of like-with-like mixing between members of the first four groups and a tendency for older men to interact with younger women. Obvi-

TABLE 3. *Sexual partnership distribution of female students by counts[a] and proportions[b]*

Female group j	Male partner group i							
	1	2	3	4	Subtotal X_j^f	5 (Other)	Total Y_j^f	Average C_j^f
1	5	3	4	3	15	14	29	1.45
	0.172	0.103	0.138	0.103		0.483		
2	1	13	5	2	21	15	36	1.38
	0.028	0.361	0.139	0.056		0.417		
3	2	4	11	7	24	18	42	1.17
	0.048	0.095	0.262	0.167		0.429		
4	0	0	1	7	8	21	29	1.04
	0.000	0.000	0.034	0.241		0.724		
Total	8	20	21	19	68(X_+^f)	68	136(Y_+^f)	1.24
	0.059	0.147	0.154	0.140		0.500		

[a]U_{ij}^f, upper line.
[b]P_{ji}^f, lower line.

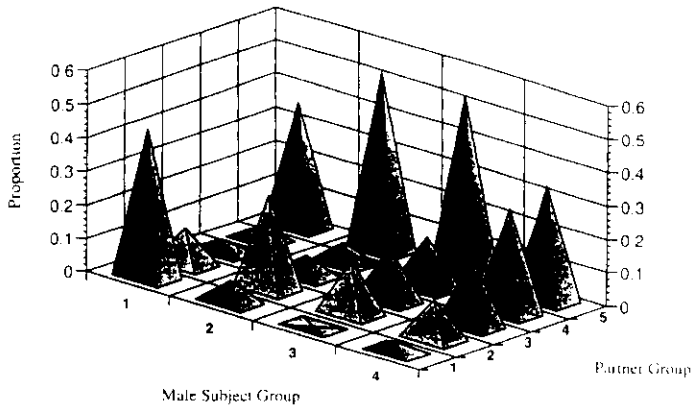

FIG. 8. Male incomplete mixing matrix from data.

ously the usual assumption of random or proportionate mixing used in the mathematical modeling of STDs does not fit here.

To use the mark-recapture methodology, we assume that sexually active college students are marked and sexually active individuals of group 5 are unmarked. The number of groups for men and women are the same, namely $I = J = 5$. Our observers are the surveyed sexually active students. From Equation [18] we estimate the sizes of sexually active subgroups in group 5 of both genders:

$$\hat{T}_5^m = \frac{2377\,(136-68)}{68+1} = 2343 \text{ and } \hat{T}_5^f = \frac{2304\,(120-67)}{67+1} = 1796.$$

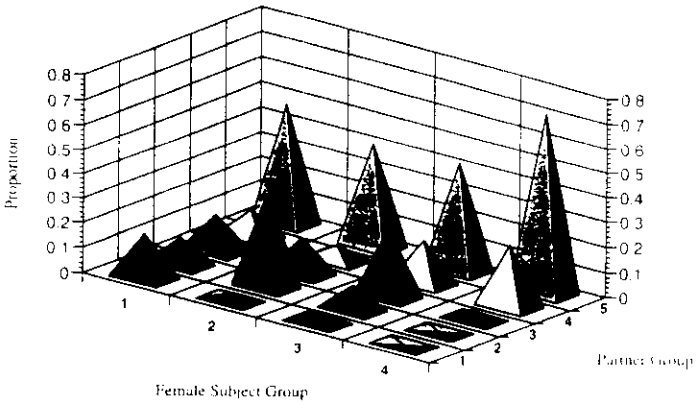

FIG. 9. Female incomplete mixing matrix from data.

Since $K_5^m = 1445.129$ and $K_5^f = 1437.012$, the lower bounds for the average number of partners for individuals in group 5 are

$$C_5^m \geq \max(1437.012/2343,\ 1) = \max(0.613,\ 1) = 1$$

$$C_5^f \geq \max(1445.129/1796,\ 1) = \max(0.805,\ 1) = 1.$$

In addition, Equation [28] implies that

$$\bar{C}_5^f = (2343/1796)\,\bar{C}_5^m + (3331.09 - 2818.34)/1796 = 1.305\,\bar{C}_5^m + 0.285 \geq \bar{C}_5^m.$$

That is, the average number of partners per woman in group 5 is greater than that for men, whereas the situation is reversed among individuals in groups 1, 2, and 4 (the values are very close in group 3). If we assume that $\bar{C}_5^m = 1$, then $\bar{C}_5^f = 1.590$; and from Equations [29] and [30] we obtain

$$\bar{P}_{51}^m = 0.086,\ \bar{P}_{52}^m = 0.062,\ \bar{P}_{53}^m = 0.194,\ \bar{P}_{54}^m = 0.056,\ \bar{P}_{55}^m = 0.683,$$

$$\bar{P}_{51}^f = 0.146,\ \bar{P}_{52}^f = 0.160,\ \bar{P}_{53}^f = 0.081,\ \bar{P}_{54}^f = 0.296,\ \bar{P}_{55}^f = 0.317.$$

Because of rounding, $\sum_{j=1}^{5} \bar{P}_{5j}^m$ is not exactly equal to 1. The above calculation is used only to demonstrate the estimation procedure. Table 4 lists $\bar{C}_5^f$, $\{\bar{P}_{5j}^m\}$, and $\{\bar{P}_{5i}^f\}$ calculated with double precision for different values of $\bar{C}_5^m$. It is clear that $\bar{C}_5^f$, $\bar{P}_{55}^m$, and $\bar{P}_{55}^f$ increase with $\bar{C}_5^m$, whereas, for other values of i and j, $\bar{P}_{5j}^m$ and $\bar{P}_{5i}^f$ decrease with $\bar{C}_5^m$. Figures 10 through 13 illustrate the shape of the completed matrices with different values for the pair-formation parameter $\bar{C}_5^m$. A second example that uses dating data from the same college population exhibits similar results (see ref. 57).

In this example, the data satisfy axioms (B1) and (B2) but not axiom (B3). The same is true for the estimated $\{\bar{P}_{5j}^m\}$ and $\{\bar{P}_{5i}^f\}$. Axiom (B3) is violated because we could not survey all individuals in the population.

TABLE 4. *Mixing proportions of men[a] and women[b] in group 5 for different average numbers of partners ($\bar{C}_5^m$)*

		$\bar{P}_{5j}^{m[a]}$ or $\bar{P}_{5i}^{f[b]}$				
				j or *i*		
$\bar{C}_5^m$	$\bar{C}_5^f$	1	2	3	4	5
1.000	1.592	0.086	0.062	0.193	0.055	0.603
		0.146	0.160	0.081	0.296	0.318
1.500	2.245	0.057	0.041	0.129	0.036	0.736
		0.103	0.114	0.057	0.210	0.516
2.000	2.897	0.043	0.031	0.097	0.027	0.802
		0.080	0.088	0.044	0.163	0.625
2.500	3.549	0.034	0.025	0.077	0.022	0.841
		0.065	0.072	0.036	0.133	0.694
3.000	4.202	0.029	0.021	0.064	0.018	0.868
		0.055	0.061	0.031	0.112	0.741
3.500	4.854	0.025	0.018	0.055	0.016	0.887
		0.048	0.053	0.026	0.097	0.776

[a] $\bar{P}_{5j}^m$, upper line.
[b] $\bar{P}_{5i}^f$, lower line.

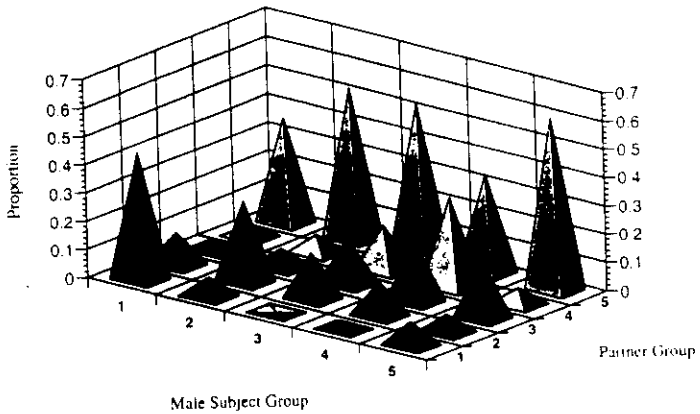

FIG. 10. Completed male mixing matrix from data with pair-formation parameter = 1.

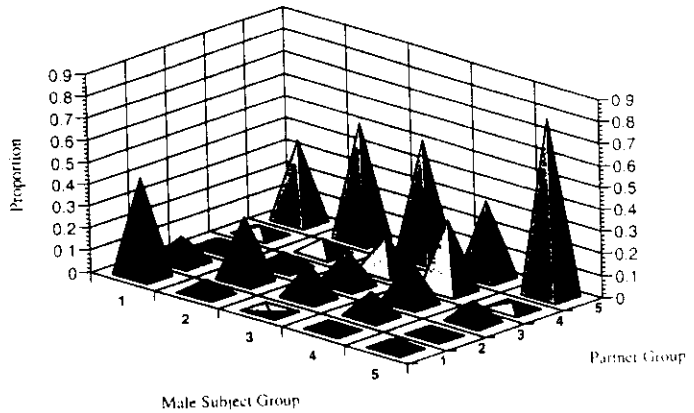

FIG. 11. Completed male mixing matrix from data with pair-formation parameter = 2.

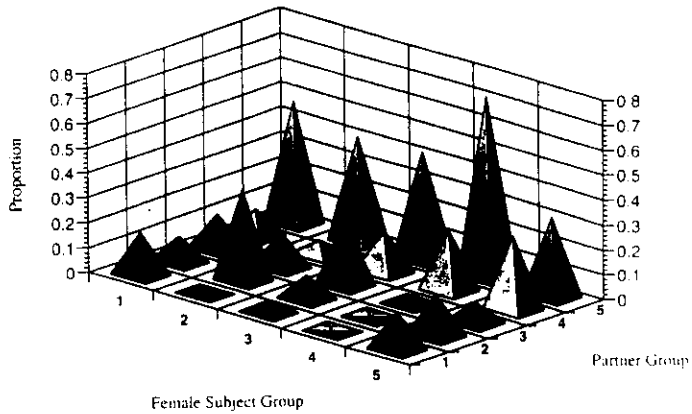

FIG. 12. Completed female mixing matrix from data with pair-formation parameter = 1.

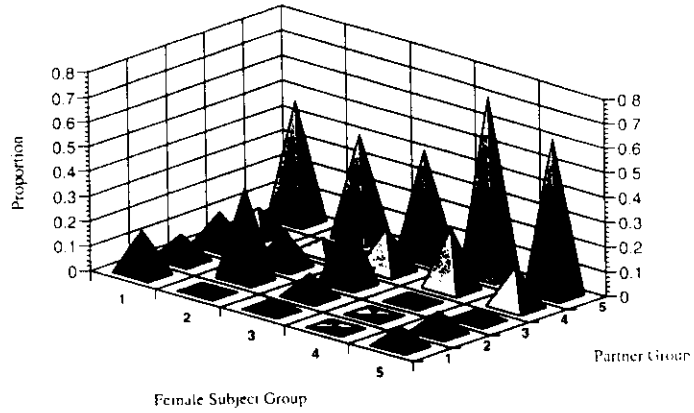

FIG. 13. Completed female mixing matrix from data with pair-formation parameter = 2.

CONCLUSIONS

Models for the dynamics of STD transmission have implicitly assumed that the mixing network is closed. Sociologists, epidemiologists, and theoreticians interested in connecting their models to data have not only found it difficult to collect these data but also have been forced by the modeling structure to construct data that assume that the mixing network is closed. For example, some researchers have used racial data on marriages and assumed that the social/sexual mixing structure of a population is proportionally reflected in these data. This assumption not only imposes a like-with-like mixing structure but also may impose a like-with-like mixing structure that is independent of population dynamics such as preferred mixing. The danger of these assumptions may become more evident when we observe that the data presented here also took the existing superimposed social structure used in U.S. universities (first year, second year, etc.).

In this chapter we have presented a mechanism for estimating the shape of a mixing matrix from a single survey. The data structure section helps identify the parameters needed for this estimation. We hope that this may be useful to researchers planning to construct survey instruments to identify the social structure of the population.

The role of nontarget populations was highlighted because it played such a prominent role in our example. With data collected from a survey that asks specific questions about sexual behavior, the size of the sexually active nontarget population can be conditionally estimated by using mark-recapture methodology. The condition is that all individuals in this sexually active nontarget population have at least one sexual contact with individuals from the target population. We also have assumed that the average contact rates between target and nontarget populations are similar—an assumption that could be relaxed if more data were available. Even after the estimation of the size of the sexually active nontarget population was completed, one row was still missing in our mixing matrix for each gender. Point estimates of the elements of this row were carried out by assuming that the elements are consistent with the two-sex mixing axioms, which reduces the computation to that of estimating a single pair-formation parameter, namely the average number of partners for men (or women) in the nontarget population. Lacking an independent estimate of this parameter left us no alternative but that of declaring it a free parameter. The larger the free parameter, the larger the mixing proportion from the nontarget population of a given gender to the nontarget population of the other gender, and the smaller the corresponding mixing proportions to the target population.

The example of sexual behavior of college students reveals that the proportion of relationships with individuals from the nontarget population is high (44.2% for men and 50.0% for women). Mixing matrices that exclude the nontarget population may not provide a complete picture of the social network and may lead to erroneous conclusions. The example shows that random mixing is unlikely for this college population with the university's superimposed classification. There is some evi-

dence of like-with-like preference and of pairing between older men and younger women within the groups in the target population.

If we use different criteria to categorize individuals (e.g., sexual activity) and consider members of the nontarget population to be prostitutes, injecting drug users, or bisexuals who may not be willing to respond to a survey and who may be at high risk of HIV/AIDS, what picture do we get? Long-term forecasting of HIV/AIDS is being carried out without estimates of the mixing matrices that model realistic, inconvenient social structures. Even the standard classifications used by the Centers for Disease Control and Prevention lead to conclusions that may not hold up under a different classifying system. The fact that social and disease dynamics have not been studied systematically provides one more significant example of the importance of interdisciplinary research to understand better the spread of CDs.

ACKNOWLEDGMENTS

This research was partially supported by NSF grant DEB-925370 to Carlos Castillo-Chavez and by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University (contract DAAL03-91-C-0027). It was partially completed while Carlos Castillo-Chavez was a member of the Isaac Newton Institute in Cambridge, England. The authors would like to thank Ed Kaplan for his suggestions and detailed evaluation of each page of this chapter. His contributions were invaluable to this chapter and our future research; of course, the responsibility for omissions and errors lies entirely with us. This chapter also benefited from the comments of an anonymous referee.

REFERENCES

1. Hoppensteadt F. An age dependent epidemic model. *J Franklin Instit* 1974;297:325–33.
2. Dietz K. Transmission and control of arbovirus diseases. In: Cooke KL, ed. *Epidemiology*. Philadelphia: Society for Industrial and Applied Mathematics; 1975:104–21.
3. Anderson RM, May RM. Spatial, temporal and genetic heterogeneity in host populations and the design of immunization programmes. *IMA J Math Appl Med Biol* 1984;1:233–66.
4. Dietz K, Schenzle D. Proportionate mixing models for age-dependent infection transmission. *J Math Biol* 1985;22:117–20.
5. Castillo-Chavez C, Hethcote H, Andreasen V, Levin SA, Liu WM. Cross-immunity in the dynamics of homogeneous and heterogeneous populations. In: Hallam TG, Gross LG, Levin SA, eds. *Mathematical ecology*. Singapore: World Scientific Publishing Co; 1988:303–16.
6. Castillo-Chavez C, Hethcote H, Andreasen V, Levin SA, Liu WM. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *J Math Biol* 1989;27(3):233–58.
7. Anderson RM, May RM. *Infectious diseases of humans*. Oxford, England: Oxford Science Publications; 1991.
8. Busenberg S, Castillo-Chavez C. Interaction, pair formation and force of infection terms in sexually-transmitted diseases. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag 1989:289–300.
9. Busenberg S, Castillo-Chavez C. A general solution of the problem of mixing subpopulations, and its application to risk- and age-structured epidemic models for the spread of AIDS. *IMA J Math Appl Med Biol* 1991;8:1–29.

10. Anderson RM. *Population dynamics of infectious diseases: theory and applications*. London, New York: Chapman and Hall; 1982.
11. Nold A. Heterogeneity in disease-transmission modeling. *Math Biosci* 1980;52:227–40.
12. Hethcote HW, Yorke JA. *Gonorrhea, transmission dynamics, and control*. Lecture Notes in Biomathematics, Vol. 56. Berlin: Springer-Verlag;1984.
13. Sattenspiel L. Population structure and the spread of disease. *Hum Biol* 1987;59:411–38.
14. Sattenspiel L. Epidemics in nonrandomly mixing populations: a simulation. *Am J Phys Anthropol* 1987;73:251–65.
15. Sattenspiel L, Simon CP. The spread and persistence of infectious diseases in structured populations. *Math Biosci* 1988;90:341–66.
16. Sattenspiel L, Castillo-Chavez C. Evironmental context, social interactions, and the spread of HIV. *Am J Hum Biol* 1990;2:397–417.
17. Blythe SP, Anderson RM. Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). *IMA J Math Appl Med Biol* 1988; 5:1–19.
18. Blythe SP, Anderson RM. Variable infectiousness in HIV transmission models. *IMA J Math Appl Med Biol* 1988,5:181–200.
19. Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989.
20. Castillo-Chavez C, Cooke K, Huang W, Levin SA. The role of infectious periods in the dynamics of acquired immunodeficiency syndrome (AIDS). In: Castillo-Chavez C, Levin SA, Shoemaker C, eds. *Mathematical approaches to ecological and environmental problem solving*. Lecture notes in Biomathematics, Vol. 81). Berlin: Springer-Verlag; 1989:177–89.
21. Castillo-Chavez C, Cooke K, Huang W, Levin SA. The role of long incubation periods in the dynamics of HIV/AIDS. Part 1: Single population models. *J Math Biol* 1989;27:373–98.
22. Castillo-Chavez C, Cooke K, Huang W, Levin SA. Results on the dynamics for models for the sexual transmission of the human immunodeficiency virus. *Appl Math Lett* 1989;2(4):327–31.
23. Castillo-Chavez C, Cooke K, Huang W, Levin SA. On the role of long incubation periods in the dynamics of HIV/AIDS. Part 2: Multiple group models. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989:200–17.
24. Castillo-Chavez C, Fridman S, Luo X. Stochastic and deterministic models in epidemiology. In: *Proceedings of the First World Congress of Nonlinear Analysts*. Berlin: Walter de Druyter & Co. [in press].
25. Huang W, Cooke K, Castillo-Chavez C. Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS transmission. *SIAM J Appl Math* 1992;52(3):835–54.
26. Thieme HR, Castillo-Chavez C. On the role of variable infectivity in the dynamics of the human immunodeficiency virus epidemic. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989:157–76.
27. Thieme HR, Castillo-Chavez C. How may infection-age dependent infectivity affect the dynamics of HIV/AIDS? *SIAM J Appl Math* 1993;5:1447–79.
28. Anderson RM. The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *J Roy Stat Soc A* 1988;151:66–93.
29. Anderson RM. The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS. *J AIDS* 1988;1:241–56.
30. Anderson RM, May RM. Transmission dynamics of HIV infection. *Nature* 1987;326:137–42.
31. Anderson RM, May RM, Medley GF, Johnson A. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J Math Appl Med Biol* 1986;3:229–63.
32. Anderson RM, Blythe SP, Gupta S, Konings E. The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behavior. *Phil Trans Roy Soc Lond B* 1989;325:45–89.
33. Blythe SP, Castillo-Chavez C. Like-with-like preference and sexual mixing models. *Math Biosci* 1989;96:221–38.
34. Blythe SP, Castillo-Chavez C, Casella G. Empirical methods for the estimation of the mixing probabilities for socially-structured populations from a single survey sample. *Math Pop Stud* 1992; 3(3):199–225.

35. Castillo-Chavez C, Busenberg S. On the solution of the two-sex mixing problem. In: Busenberg S, Martelli M, eds. *Proceedings of the International Conference on Differential Equations and Applications to Biology and Population Dynamics*. Lecture Notes in Biomathematics, Vol. 92. Berlin: Springer-Verlag 1991;80–98.
36. Castillo-Chavez C, Busenberg S, Gerow K. Pair formation in structured populations. In: Goldstein J, Kappel F, Schappacher W, eds. *Differential equations with applications in biology, physics and engineering*. New York: Marcel Dekker; 1991:47–65.
37. Dietz K. On the transmission dynamics of HIV. *Math Biosci* 1988;90:397–414.
38. Dietz K, Hadeler KP. Epidemiological models for sexually transmitted diseases. *J Math Biol* 1988; 26:1–25.
39. Gupta S, Anderson RM, May RM. Network of sexual contacts: implications for the pattern of spread of HIV. *AIDS* 1989;3:1–11.
40. Hadeler KP. Pair formation in age-structured populations. *Acta Appl Math* 1989;14:91–102.
41. Hadeler KP. Modeling AIDS in structured populations. *47th Session of the International Statistical Institute, Paris, August/September. Conf. Proc.* 1989;C1-2.1:83–99.
42. Hadeler KP, Nagoma K. Homogeneous models for sexually-transmitted diseases. *Rocky Mountain J Math* 1990;20:967–86.
43. Hyman JM, Stanley EA. Using mathematical models to understand the AIDS epidemic. *Math Biosci* 1988;90:415–73.
44. Hyman JM, Stanley EA. The effect of social mixing patterns on the spread of AIDS. In: Castillo-Chavez C, Levin SA, Shoemaker C, eds. *Mathematical approaches to problems in resource management and epidemiology*. Lecture Notes in Biomathematics, Vol. 81. Berlin: Springer-Verlag; 1989;119–90.
45. Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T. Modeling and analyzing HIV transmission: the effects of contact patterns. *Math Biosci* 1989;92:119–99.
46. Jacquez JA, Simon CP, Koopman J. Structured mixing: heterogeneous mixing by the definition of mixing groups. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989:301–15.
47. Kaplan EH, Lee YS. How bad can it get? Bounding worst case endemic heterogeneous mixing models of HIV/AIDS. *Math Biosci* 1990;99:157–80.
48. Koopman JS, Simon CP, Jacquez JA, Park TS. Selective contact within structured mixing with an application to HIV transmission risk from oral and anal sex. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989:316–48.
49. May RM, Anderson RM. The transmission dynamics of human immunodeficiency virus (HIV). *Phil Trans Roy Soc Lond B* 1989;321:565–607.
50. Waldstätter R. Pair formation in sexually transmitted diseases. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989: 260–74.
51. Gabriel JP, Lefevre C, Picard P. *Stochastic processes in epidemic theory*. Lecture Notes in Biomathematics, Vol. 90. Berlin: Springer-Verlag; 1990.
52. Jewell NP, Dietz K, Farewell VT, eds. *AIDS epidemiology: methodological issues*. Berlin: Birkhäuser; 1992.
53. Hethcote HW, Van Ark JW. *Modeling HIV transmission and AIDS in the United States*. Lecture Notes in Biomathematics, Vol. 95. New York: Springer-Verlag; 1992.
54. Crawford CM, Schwager SJ, Castillo-Chavez C. A methodology for asking sensitive questions among college undergraduates. *Biometrics Unit Technical Report*. Ithaca, NY: Cornell University; 1990:BU-1105-M.
55. Rubin G, Umbach D, Shyu SF, Castillo-Chavez C. Application of capture-recapture methodology to estimation of size of population at risk of AIDS and/or other sexually-transmitted diseases. *Stat Med* 1992;11:1533–49.
56. Castillo-Chavez C, Shyu SF, Rubin G, Umbach D. On the estimation problem of mixing/pair formation matrices with applications to models for sexually-transmitted diseases. In: Dietz K, Farewell VT, Jewell NP, eds. *AIDS epidemiology: methodological issues*. Berlin: Birkhauser; 1992:384–402.
57. Hsu Schmitz SF, Castillo-Chavez C. Completion of mixing matrices for nonclosed social networks. In: *Proceedings of the First World Congress of Nonlinear Analysts*. Berlin: Walter de Druyter & Co. [in press].
58. Castillo-Chavez C, Blythe SP. Mixing framework for social/sexual behavior. In: Castillo-Chavez C,

ed. *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics, Vol. 83. Berlin: Springer-Verlag; 1989:275–88.
59. Blythe SP, Castillo-Chavez C, Palmer J, Cheng M. Towards unified theory of mixing and pair formation. *Math Biosci* 1991;107:379–405.
60. Blythe SP, Castillo-Chavez C, Busenberg S. Affinity and paired-event probability. *Biometrics Unit Technical Report*. Ithaca, NY: Cornell University; 1993;BU-1084-M.
61. Blythe SP, Castillo-Chavez C. Like-with-like mixing and sexually transmitted-disease epidemics in one-sex populations. *Biometrics Unit Technical Report*. Ithaca, NY: Cornell University; 1990; BU-078-M.
62. Lajmanovich A, Yorke JA. A deterministic model for gonorrhea in a nonhomogeneous population. *Math Biosci* 1976;28:221–36.
63. Huang W. *Studies in differential equations and applications*. Ph.D. dissertation, Claremont, California, 1989.
64. Blythe SP, Castillo-Chavez C. Is there a marriage function yet? *Biometrics Unit Technical Report*. Ithaca, NY: Cornell University; 1991;BU-1135-M.
65. Hsu Schmitz SF, Castillo-Chavez C. On the evolution of marriage functions: it takes two to tango. *Biometrics Unit Technical Report*. Ithaca, NY: Cornell University; 1993;BU-1210-M.
66. Bailey NTJ. On estimating the size of mobile populations from recapture data. *Biometrika* 1951;38:293–306.
67. Seber GAF. *The estimation of animal abundance and related parameters*. New York: Macmillan; 1982.