



Published in final edited form as:

Sex Transm Infect. 2014 June ; 90(4): 332–336. doi:10.1136/sextrans-2013-051282.

Assortativity coefficient-based estimation of population patterns of sexual mixing when cluster size is informative

Siobhan K Young¹, Robert H Lyles², Lawrence L Kupper¹, Jessica R Keys¹, Sandra L Martin¹, Elizabeth C Costenbader³

¹University of North Carolina, Chapel Hill, North Carolina, USA

²Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

³FHI 360, Durham, North Carolina, USA

Abstract

Objectives—Population sexual mixing patterns can be quantified using Newman’s assortativity coefficient (r). Suggested methods for estimating the SE for r may lead to inappropriate statistical conclusions in situations where intracluster correlation is ignored and/or when cluster size is predictive of the response. We describe a computer-intensive, but highly accessible, *within-cluster resampling* approach for providing a valid large-sample estimated SE for r and an associated 95% CI.

Methods—We introduce needed statistical notation and describe the *within-cluster resampling* approach. Sexual network data and a simulation study were employed to compare *within-cluster resampling* with standard methods when cluster size is informative.

Results—For the analysis of network data when cluster size is informative, the simulation study demonstrates that within-cluster resampling produces valid statistical inferences about Newman’s assortativity coefficient, a popular statistic used to quantify the strength of mixing patterns. In contrast, commonly used methods are biased with attendant extremely poor CI coverage. Within-cluster resampling is recommended when cluster size is informative and/or when there is within-cluster response correlation.

Conclusions—Within-cluster resampling is recommended for providing valid statistical inferences when applying Newman’s assortativity coefficient r to network data.

Sexually transmitted infection (STI) epidemic trajectory is largely determined by patterns of sexual contact within and between population groups (commonly referred to as sexual mixing).¹ By convention, sexual mixing (*mixing* for simplicity) is expressed on a continuum ranging from perfectly ‘assortative’ to perfectly ‘disassortative’.^{2,3} *Assortative* mixing occurs when people select sexual partners from within their own group (ie, characteristically

Correspondence to: Dr Siobhan K Young, Department of Maternal and Child Health, University of North Carolina at Chapel Hill, 401 Rosenau Hall, CB# 7445, Chapel Hill, NC 27599-7445, USA; siobhankyong@gmail.com.

Contributors SKY, LLK, JRK, RHL and SLM collaborated in the writing of the manuscript. SKY, JRK and RHL performed the statistical analyses. LLK chose the main directions for the analyses and participated in the interpretation of results. ECC provided data from the HCMC Sex Network Survey. All authors revised the manuscript before submission.

Competing interests None.

or behaviourally similar to themselves).⁴ In contrast, *disassortative* mixing occurs when people choose sexual partners from outside their own group (ie, unlike themselves).⁵ If people selected sexual partners without regard to group, mixing would be considered random.⁶

Evidence suggests that sexual partnering patterns (and hence the networks generated by these patterns) are in fact non-random, and this non-randomness has substantial implications for both the rate and degree to which STIs, including HIV, may spread throughout a population.^{2 6–8} Assortative mixing implies that people tend to mix within closed groups, say, young men with young men, injection drug users (IDUs) with other IDUs, etc. Under such conditions, infection generally persists within the closed groups into which it was introduced; this typically results in multiple, distinct and quickly evolving epidemics within a population.^{9 10} Under disassortative mixing conditions, such as when people from a low STI prevalence group mix with people from a high STI prevalence group (eg, young girls with older promiscuous men, non-IDUs with IDUs, etc.), the result is typically a sustained epidemic that progresses slowly from high prevalence (ie, high-risk) groups to the general population.^{9 11}

Given that an enhanced understanding of mixing patterns can guide the development of targeted intervention programmes to reduce STIs among those groups most at risk, a number of studies have attempted to quantify population-level mixing patterns using both sociometric^{12 13} and egocentric^{14–16} network data to construct a mixing matrix.³ A mixing matrix is a cross-tabulation of the value of a characteristic (eg, HIV status, race, age group) of an index participant (row entry) with the corresponding value of the characteristic of a named partner (column entry). More specifically, the (i,j)th cell of a mixing matrix contains the observed proportion of dyads (ie, ‘edges’ in network terminology) in which index participants have level i and named partners have level j of the characteristic under study.

An example of a mixing matrix is given in table 1, which provides real dyadic data on HIV status from 253 male indexes at high risk for HIV and their up to five nominated sexual partners. These data were collected as part of an FHI 360 sexual network study, *Sexual Behavioral Relationships and HIV Infection in Ho Chi Minh City, Vietnam*. Partner HIV status in table 1 is as reported by the index subjects and was set to ‘negative’ in cases where it was unknown. Although this induces some potential misclassification into the observed negative partner status indicators (see ‘Discussion’), we treat the resulting HIV status data at face value for the illustrative purposes targeted in this paper.

Newman³ provides an excellent discussion of the pros and cons of suggested methods for analysing mixing patterns using data cast in the form of a mixing matrix. In his well-cited paper, Newman introduces an *assortativity coefficient* r , discusses its properties and interpretation, and suggests two methods for estimating a SE for r . Newman’s r and his SE estimation methods have been used by numerous authors.^{15–17}

Newman’s suggested methods for estimating a SE for r are based on the assumption that the edges (or dyads) in a mixing matrix are mutually independent. However, this assumption is often not appropriate since, for network data, a particular index subject typically is

a member of more than one dyad. More specifically, dyads sharing the same index subject form a cluster of potentially correlated observations, and ignoring such intracluster correlation can lead to the use of incorrect SE estimates and hence to inappropriate statistical conclusions.¹⁸ Bohl *et al*¹⁷ recognised this non-independent dyad issue, but offered only an ad hoc remedy without rigorous statistical justification for its use.

An even more problematic issue that can invalidate the standard application of Newman's assortativity coefficient to network data is *informative cluster size* (ICS).¹⁹ ICS refers to the phenomenon where cluster size is predictive of (ie, is correlated with) the response under consideration. Table 2 illustrates the ICS phenomenon using our example of high-risk partnership data from HCMC, Vietnam. Here, we can see that the highest proportion of HIV-positive male index subjects resides in a cluster size of 2. This likely reflects the fact that 81% of HIV-positive male indexes were injecting drug users (IDUs) and of these, more than half (65%) reported a single female primary partner. In contrast, the cluster size distribution tends to favour larger clusters for HIV-negative male index subjects ($p < 0.001$). Despite their own low-risk behaviour, many monogamous women in Vietnam are at risk for infection resulting from marriages or partnerships with high-risk primary partners, including IDUs.^{20–22} Thus, one consequence of interdependence is that the behaviours of HIV-positive male indexes may influence the HIV status of their female partners. These considerations motivate the need to estimate Newman's r in the presence of ICS. The ICS issue is especially troubling because, as we will see, ICS can introduce bias not only into SEs but into the estimate of r itself.

In this paper, we describe and demonstrate a computer-intensive approach for providing a valid estimate and CI for Newman's assortativity coefficient. This approach, called within-cluster resampling (WCR),¹⁹ is easily understood, has strong theoretical justification, is readily implemented and is valid whether or not cluster size is informative. In section 'Methods', we introduce needed statistical notation and describe the WCR approach. In section 'Results', WCR is compared with the so-called 'naive' method, which assumes that dyads are mutually independent. Both methods are used to compute Newman's r and to construct an associated 95% CI for the mixing matrix in table 1. We also provide simulation results that compare the statistical properties of these two methods when cluster size is informative. Section 'Discussion' contains a discussion of our findings and recommendations.

METHODS

The notation used in this paper is the same as that used by Newman.³ In particular, for a ($k \times k$) mixing matrix, e_{ij} is the proportion of all dyads associated with level i for an index participant and level j for a named partner. Also, $a_i = \sum_{j=1}^k e_{ij}$ is the sum of the proportions in row i , and $b_j = \sum_{i=1}^k e_{ij}$ is the sum of the proportions in column j ; clearly, $\sum_{i=1}^k a_i = \sum_{j=1}^k b_j = 1$. Newman's assortativity coefficient r is defined as

$$r = \frac{\sum_{i=1}^k e_{ii} - \sum_{i=1}^k a_i b_i}{1 - \sum_{i=1}^k a_i b_i},$$

a quantity that is identical in structure to the well-known kappa statistic.²³ The maximum value of r is equal to +1, a value that indicates perfect assortativity; this maximum value of +1 occurs when the observed proportion of agreement $\sum_{i=1}^k e_{ii}$ is equal to 1 so that all off-diagonal elements of the mixing matrix are equal to 0. When there is essentially only random mixing, so that $\sum_{i=1}^k e_{ii}$ is close in value to the expected proportion of chance agreement $\sum_{i=1}^k a_i b_i$, then $r \approx 0$. Newman’s r can also take negative values, suggesting the presence of disassortativity. The minimum possible value of r is equal to $-\sum_{i=1}^k a_i b_i / (1 - \sum_{i=1}^k a_i b_i)$, a quantity that depends on the marginal totals of the mixing matrix. Only in the limit, as k decreases to 2, as the correlation between the row and column totals approaches +1, and as the variances of the marginal totals increase, does r approach the value -1.

In the description of WCR to follow, let us assume that we have C index subjects (or clusters), and let n_i (> 1) be the number of partners associated with the i th index subject, $i = 1, 2, \dots, C$. So, there are n_i presumably non-independent dyads (or edges) in the i th cluster.

To implement WCR, we randomly sample *with replacement* one dyad (ie, partner) from each of the C clusters. This resampled data set would then involve C mutually independent dyads (assuming no study subject appears in more than one cluster), and so we can validly analyse such a data set using standard methods for computing a kappa-type statistic (like Newman’s r) and its estimated variance based on a multinomial distribution assumption for a $(k \times k)$ table.²⁴ We then generate Q of these resampled data sets (each of size C), where Q is large (in our experience, $Q=200$ seems sufficiently large to guarantee reliable results).

Now, for the q th resampled data set, $q = 1, 2, \dots, Q$, suppose that we obtain k^2 cell entries $\{e_{qij}\}$, the k row totals $\{a_{qi}\}$ and the k column totals $\{b_{qj}\}$. Using these data, we compute Newman’s assortativity coefficient as

$$r_q = \frac{\sum_{i=1}^k e_{qii} - \sum_{i=1}^k a_{qi} b_{qi}}{1 - \sum_{i=1}^k a_{qi} b_{qi}}$$

and its estimated kappa-type variance as

$$\widehat{V}(r_q) = \frac{1}{C(1 - \sum_{i=1}^k a_{qi}b_{qi})^2} \left\{ \sum_{i=1}^k e_{qii} [1 - (a_{qi} + b_{qi})(1 - r_q)]^2 + (1 - r_q)^2 \times \sum_{\text{all } i \neq j} e_{qij} (b_{qi} + a_{qj})^2 - \left[r_q - \left(\sum_{i=1}^k a_{qi}b_{qi} \right) (1 - r_q) \right]^2 \right\}.$$

When $k=2$, so that we have a 2×2 table, the variance estimator accompanying Newman's r (ie, the unweighted kappa statistic) is seen to simplify as follows:

$$\widehat{V}(r_q) = \frac{1}{C[1 - (a_{q1}b_{q1} + a_{q2}b_{q2})]^2} \left\{ e_{q11} [1 - (a_{q1} + b_{q1})(1 - r_q)]^2 + e_{q22} [1 - (a_{q2} + b_{q2})(1 - r_q)]^2 + (1 - r_q)^2 [e_{q12} (b_{q1} + a_{q2})^2 + e_{q21} (b_{q2} + a_{q1})^2] - [r_q - (a_{q1}b_{q1} + a_{q2}b_{q2})(1 - r_q)]^2 \right\}.$$

Then, the WCR estimator of the true population value of Newman's assortativity coefficient is simply the mean of these Q r_q values, namely,

$$\bar{r} = \frac{1}{Q} \sum_{q=1}^Q r_q.$$

Because the Q resampled data sets contain overlapping observations, the estimated variance $\widehat{V}(\bar{r})$ of \bar{r} must be appropriately determined. Hoffman *et al*¹⁹ provide a formula for a more general setting; in our particular situation, the appropriate variance expression is

$$\widehat{V}(\bar{r}) = \frac{1}{Q} \sum_{q=1}^Q \widehat{V}(r_q) - \left(\frac{Q-1}{Q} \right) S_r^2,$$

where

$$S_r^2 = \frac{1}{(Q-1)} \sum_{q=1}^Q (r_q - \bar{r})^2$$

is simply the sample variance of the Q r_q values.

A valid large-sample (eg, $C = 50$) 95% CI for the true population value of Newman's assortativity coefficient is then

$$\bar{r} \pm 1.96\sqrt{\widehat{V}(\bar{r})}.$$

RESULTS

For the HCMC partnership mixing matrix (253 clusters and 594 dyads), the naive method gave an r value of 0.465, an estimated SE of 0.062 and a 95% CI of 0.342 to 0.587 (see table 1). In contrast, WCR (again, with $Q=200$) produced an r value of 0.539, an estimated SE of 0.068 and a 95% CI of 0.407 to 0.671. Thus, for the mixing matrix in table 1, the naive and WCR methods produce decidedly different r values in the presence of the ICS demonstrated in table 2. We note that the SE estimates for the two methods are not very different, although the naive method would be expected to produce somewhat lower (and potentially biased) estimated SEs as compared with those produced by WCR when there is intracluster response correlation (see ‘Discussion’).

To investigate these patterns further, we conducted a simulation study designed to reflect the observed data (table 1) and cluster size distribution (table 2) for the male index case scenario. We first generated an indicator of index status using a Bernoulli random number generator, according to the observed proportion of index subjects that were HIV-positive. A partner status indicator was then generated using a logistic regression model with index case status as a binary predictor, and with true coefficients equal to the estimates obtained from fitting such a model to the observed data. Partner indicators were thus correlated (ie, clustered), although they were conditionally independent given the index subject’s status (see ‘Discussion’). We simulated 500 data sets independently under these specifications, with cluster sizes for ‘positive’ and ‘negative’ index subject clusters generated according to the distributions shown in table 2. These conditions closely mimicked the example data scenario, dictating a true (‘population’) value of Newman’s assortativity coefficient of 0.521 and ensuring an average of approximately 594 dyads for each simulated sample. For each such data set, $Q=200$ resamples were obtained for the WCR method.

As shown in table 3, the naive method gave a mean r value of 0.459, a mean estimated SE of 0.062 and the percentage of 95% CIs that covered the population value of 0.521 as 84.2%. The WCR method, in comparison, gave a mean r value of 0.515, a mean estimated SE of 0.067 and a 95% CI coverage percentage of 95.0%.

This simulation study based on the HCMC partnership data clearly illustrates the fact that, when cluster size is informative, the naive method is apt to produce a biased estimator of the population value of r with correspondingly poor CI coverage. In contrast, WCR produced an essentially unbiased estimator of the population value of r , along with nearly nominal CI coverage. Thus, for network data with ICS, WCR is recommended as the analysis method of choice.

DISCUSSION

The crucial issue highlighted in this paper is the fact that a standard application of Newman’s assortativity coefficient r (equivalent to Cohen’s kappa statistic) is subject to bias and potentially invalid conclusions under many realistic scenarios in studies of sexual mixing. In particular, the usual estimator for r is generally biased whenever cluster size is informative (ie, associated with index subject status), as is the case in our motivating

study of HIV positivity. In our simulation study (table 3) designed to mimic the example conditions, we clearly demonstrate this bias and show that WCR provides an excellent solution to the problem.

As indicated in section ‘Methods’, the male index partnership data were used for the primary purpose of illustrating the fallibility of the standard estimator of Newman’s r in the presence of ICS, and the benefits of the WCR approach. Thus, for our current purposes, we ignore possible misclassification of negative status indicators in cases where the index subject reported a partner’s HIV status as unknown. Future work in this area could involve the development of methodology to correct a WCR-based assortativity coefficient estimator for such misclassification either via sensitivity analysis based on putative misclassification rates or via the direct incorporation of validation data.

A clever alternative methodology for validly analysing clustered data when cluster size is informative is a generalised estimating equations (GEE) approach described in the interesting paper by Williamson *et al.*²⁵ In this paper, the authors demonstrate theoretically and by simulation using 500 clusters that their GEE method is equivalent to WCR for large samples. And, based on a limited small-sample simulation study involving 50 clusters, they claim that their method is slightly better than WCR. However, their simulation and numerical example results indicate that both methods lead to identical statistical inferences when using a Z statistic (computed as a parameter estimate divided by its estimated SE). Although the Williamson *et al.*²⁵ GEE method is a valid statistical approach, the WCR method is equally justifiable and we feel that it is more intuitively appealing in the setting of this paper. In addition, given the widespread availability of high-speed computing, implementation of the resampling necessary for the WCR method is straightforward, and appropriate SAS code for implementing our WCR calculations can be provided by the study authors upon request. Finally, since each resampled data set in the WCR approach contains mutually independent observations, WCR does not require specification of a within-cluster correlation structure and estimation of a variance–covariance matrix as needed with the GEE methodology.

Although we focus primary attention on the bias introduced by ICS, it is important to note that the naive application of Newman’s r is also problematic in any case where status indicators for non-index subjects within clusters remain correlated conditional on the status of the index subject (even in the absence of ICS). In that case, the usual SE estimator that accompanies r will typically be biased downward, leading to overly optimistic CIs. Although our simulation study (table 3) was conducted under ICS and an assumption of independent partner status indicators conditional on index status, other simulations (not shown) demonstrate this fallibility of the usual SE estimator. The key conclusion is that the standard application of Newman’s r cannot be recommended in dyadic studies similar to our motivating example due to the likelihood of ICS, residual within-cluster correlation, or both. The WCR approach provides an accessible alternative that remains statistically sound, protecting against bias and invalid conclusions.

Acknowledgements

The participant data used in our motivating example were from the Sexual Behavioral Relationships and HIV Infection in Ho Chi Minh City, Vietnam (FHI 360 study# 10027). Financial assistance for FHI 360 study# 10027 was provided by the US Agency for International Development (USAID) under the terms of Cooperative Agreement No. GPO-A-00-05-00022-00, the Contraceptive and Reproductive Health Technologies Research and Utilization (CRTU) Program.

REFERENCES

1. Aral SO. Sexual network patterns as determinants of STD rates: paradigm shift in the behavioral epidemiology of STDs made visible. *Sex Transm Dis* 1999;26:262–4. [PubMed: 1033278]
2. Doherty IA, Padian NS, Marlow C, et al. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J Infect Dis* 2005;191(Suppl 1):S42–54. [PubMed: 15627230]
3. Newman MEJ. Mixing patterns in networks. *Phys Rev. E Stat Nonlin Soft Matter Phys* 2003;67(2 Pt 2):026126. [PubMed: 12636767]
4. Gorbach PM, Murphy R, Weiss RE, et al. Bridging sexual boundaries: men who have sex with men and women in a street-based sample in Los Angeles. *J Urb Health* 2009;86(Suppl 1):63–76.
5. Hertog S Heterosexual behavior patterns and the spread of HIV/AIDS: the interacting effects of rate of partner change and sexual mixing. *Sex Trans Dis* 2007;34:820–8.
6. Morris M Sexual networks and HIV. *AIDS* 1997;11(Suppl A):S209–16. [PubMed: 9451987]
7. Laumann EO, Gagnon JH, Michael RT, et al. *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press, 1994.
8. Anderson RM. Mathematical models of the potential demographic impact of AIDS in Africa. *AIDS* 1991;5(Suppl 1):S37–44. [PubMed: 1669923]
9. Williams ML, Bowen AM, Timpson S, et al. Drug injection and sexual mixing patterns of drug-using male sex workers. *Sex Transm Dis* 2003;30:571–4. [PubMed: 12838086]
10. Ford K, Sohn W, Lepkowski J. American adolescents: sexual mixing patterns, bridge partners, and concurrency. *Sex Transm Dis* 2002;29:13–19. [PubMed: 11773873]
11. Rothenberg R How a net works: implications of network structure for the persistence and control of sexually transmitted diseases and HIV. *Sex Transm Dis* 2001;28:63–8. [PubMed: 11234787]
12. Helleringer S, Kohler HP. Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. *AIDS* 2007;21:2323–32. [PubMed: 18090281]
13. Stoner BP, Whittington WL, Hughes JP, et al. Comparative epidemiology of heterosexual gonococcal and chlamydial networks: implications for transmission patterns. *Sex Transm Dis* 2000;27:215–23. [PubMed: 10782743]
14. Garnett GP, Hughes JP, Anderson RM, et al. Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex Transm Dis* 1996;23:248–57. [PubMed: 8724517]
15. Doherty IA, Adimora AA, Muth SQ, et al. Comparison of sexual mixing patterns for syphilis in endemic and outbreak settings. *Sex Transm Dis* 2011;38:378–84. [PubMed: 21217418]
16. Doherty IA, Schoenbach VJ, Adimora AA. Sexual mixing patterns and heterosexual HIV transmission among African Americans in the southeastern United States. *J Acquir Immune Defic Syndr* 2009;52:114–20. [PubMed: 19506485]
17. Bohl DD, McFarland W, Raymond HF. Improved measures of racial mixing among men who have sex with men using Newman's assortativity coefficient. *Sex Transm Infect* 2011;87:616–20. [PubMed: 21983853]
18. Liang KY, Zeger SL. Regression analysis for correlated data. *Ann Rev Pub Health* 1993;14:43–68. [PubMed: 8323597]
19. Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. *Biometrika* 2001;88:1121–34.
20. Nguyen TH, Nguyen TL, Trinh QH. HIV/AIDS epidemics in Vietnam: evolution and responses. *AIDS Educ Preven* 2004;16(3 Suppl A):137–54.
21. Nguyen TA, Oosterhoff P, Hardon A, et al. A hidden HIV epidemic among women in Vietnam. *BMC Pub Health* 2008;8:37. [PubMed: 18221565]

22. Phinney HM. "Rice is essential but tiresome; you should get some noodles": Doi Moi and the political economy of men's extramarital sexual relations and marital HIV risk in Hanoi, Vietnam. *Am J Pub Health* 2008;98:650–60. [PubMed: 18309136]
23. Cohen J A coefficient of agreement for nominal scales. *Educ Psych Meas* 1960;20:37–46.
24. Fleiss JL, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psych Bull* 1969;72:323–27.
25. Williamson JM, Datta S, Satten GA. Marginal analyses of clustered data when cluster size is informative. *Biometrics* 2003;59:36–42. [PubMed: 12762439]

Key messages

- Once a sexually transmitted infection enters a population, epidemic trajectory is largely determined by patterns of sexual contact between partners (referred to as sexual mixing).
- Using partnership data, population-level sexual mixing patterns can be quantified by calculating the assortativity coefficient (r).
- Standard application of r assumes that partnerships are mutually independent, leading to bias and invalid statistical conclusions under many scenarios of sexual mixing.
- A within-cluster resampling approach provides valid estimation of r and accurate CI coverage in the presence of informative cluster size, residual within-cluster correlation, or both.

Table 1Male indexes' HIV status cross-classified by female partners' HIV status^{*}

Male index HIV status	Female partner HIV status			Newman's κ estimate (SE) (95% CI) [†]	
	Negative	Positive	Total	Naive	WCR
Negative	513 (86%)	27 (5%)	540	0.465 (0.062)	0.539 (0.068)
Positive	26 (4%)	28 (5%)	54	(0.342 to 0.587)	(0.407 to 0.671)
Total	539	55	594		

^{*} Male index participants (n=253) in the Ho Chi Minh City Sex Network study (2009) provided data on their HIV status as well as the HIV status of up to five of their recent sexual partners (n=594). Male indexes' HIV status was obtained via a combination of biological testing and self-report; partner HIV status was self-reported by male index participants.

[†] Newman's κ is equivalent to the standard unweighted kappa statistic.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Size distributions for clusters with male index subjects (253 clusters; 594 dyads)*

Cluster size [†]	HIV-positive indexes # of clusters (%)	HIV-negative indexes # of clusters (%)
2	41 (89%)	53 (25%)
3	3 (7%)	50 (24%)
4	1 (2%)	47 (23%)
5	1 (2%)	39 (19%)
6	0 (0%)	18 (9%)

* Fisher's exact p value<0.0001 for association between cluster size and index status.

[†] Cluster size includes index subject.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Simulation results assessing naive and WCR* estimators under conditions mimicking informative cluster sizes observed in HIV status data

	Naive	WCR
True assortativity coefficient=0.521		
Mean r estimate [†]	0.459	0.515
Empirical SD across 500 simulations	0.065	0.067
Mean estimated SE	0.062	0.067
95% CI coverage	84.2%	95.0%

* 500 simulations in each case; WCR based on Q=200 resamples from each simulated data set.

[†]Newman's r is equivalent to the standard unweighted kappa statistic.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript