

Health Education & Behavior

<http://heb.sagepub.com>

Estimates of Intraclass Correlation for Variables Related to Behavioral HIV/STD Prevention in a Predominantly African American and Hispanic Sample of Young Women

Sherri L. Pals, Brenda L. Beaty, Samuel F. Posner and Sheana S. Bull
Health Educ Behav 2009; 36; 182
DOI: 10.1177/1090198108327731

The online version of this article can be found at:
<http://heb.sagepub.com/cgi/content/abstract/36/1/182>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Society for Public Health Education](#)

Additional services and information for *Health Education & Behavior* can be found at:

Email Alerts: <http://heb.sagepub.com/cgi/alerts>

Subscriptions: <http://heb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://heb.sagepub.com/cgi/content/refs/36/1/182>

Estimates of Intraclass Correlation for Variables Related to Behavioral HIV/STD Prevention in a Predominantly African American and Hispanic Sample of Young Women

Sherri L. Pals, PhD
Brenda L. Beaty, MSPH
Samuel F. Posner, PhD
Sheana S. Bull, PhD

Studies designed to evaluate HIV and STD prevention interventions often involve random assignment of groups such as neighborhoods or communities to study conditions (e.g., to intervention or control). Investigators who design group-randomized trials (GRTs) must take the expected intraclass correlation coefficient (ICC) into account in sample size estimation to have adequate power; however, few published ICC estimates exist for outcome variables related to HIV and STD prevention. The Prevention Options for Women Equal Rights (POWER) study was a GRT designed to evaluate a campaign to increase awareness and use of condoms among young African American and Hispanic women. The authors used precampaign and postcampaign data from the POWER study to estimate ICCs (unadjusted and adjusted for covariates) for a variety of sexual behavior and other variables. To illustrate the impact of ICCs on power, the authors present sample-size calculations and demonstrate how ICCs of differing magnitude will affect estimates of required sample size.

Keywords: *intraclass correlation; group-randomized trial; HIV/STD prevention*

Community-based interventions to reduce risk behaviors for human immunodeficiency virus (HIV) and other sexually transmitted diseases (STDs) often involve random assignment of intact groups of participants, such as the residents of cities or neighborhoods, to study conditions (e.g., to control or intervention; Pedlow & Carey, 2003; Robin et al., 2004; Semaan et al., 2002). This design may be advantageous when

Sherri L. Pals, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia. Brenda L. Beaty, Colorado Health Outcomes Program, University of Colorado at Denver and Health Sciences Center. Samuel F. Posner, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia. Sheana S. Bull, Colorado Health Outcomes Program, University of Colorado at Denver and Health Sciences Center.

Address correspondence to Sherri L. Pals, Division of HIV/AIDS Prevention, National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E45, Atlanta, GA 30333; phone: (404) 639-6174; fax: (404) 639-6127; e-mail: sfv3@cdc.gov.

Health Education & Behavior, Vol. 36 (1): 182-194 (February 2009)

DOI: 10.1177/1090198108327731

© 2009 by SOPHE

182

the intervention is designed to operate at a group level, randomization at an individual level is not feasible, or investigators are concerned about contamination of study conditions. Recent reviews of HIV/STD intervention studies have identified trials in which schools, classrooms, communities, clinics, and other social groups were the units of randomization and outcome variables were measured at the individual level (Pedlow & Carey, 2003; Robin et al., 2004; Semaan et al., 2002). These studies are generally referred to as cluster-randomized or group-randomized trials (GRTs); here, we use the term *group* to represent the units of randomization and *study condition* to refer to the treatment to which groups are assigned.

GRTs require different analytic methods than studies in which individual participants are randomly assigned to study conditions (Donner & Klar, 2000; Murray, 1998). Because members of social groups commonly share characteristics and experiences, observations from members of these groups tend to be correlated, whereas correlation among participants in individually randomized trials is not expected unless the participants interact with each other. This within-group correlation in GRTs means that GRTs have an additional between-group component of variation that individually randomized trials do not, and they require the use of analytic methods that account for this between-group variation. If GRTs are conducted with traditional methods for randomized clinical trials, which do not account for this variation, the Type I error rate may be inflated (Zucker, 1990). Two books (Donner & Klar, 2000; Murray, 1998) and numerous journal articles (e.g., Feng, Diehr, Peterson, & McLerran, 2001; Klar & Donner, 2001; Murray, Varnell, & Blitstein, 2004) have described appropriate design and analytic methods for GRTs, and a recent review demonstrated increasing awareness of these methods among investigators publishing GRT results (Varnell, Murray, Janega, & Blitstein, 2004).

A key part of designing GRTs involves anticipating the intraclass correlation (ICC), the magnitude of the correlation of observations within groups. Also interpreted as the proportion of the variation in an outcome variable due to the group, the ICC has a substantial impact on the sample size needed in a GRT (Donner, Birkett, & Buck, 1981). The variance of the difference between study condition means or proportions in a GRT is $[1 + (m - 1)ICC]$ times as great as the variance expected in an individually randomized trial, where m is the number of members per group; this quantity is referred to as the variance inflation factor (VIF; Donner et al., 1981) or design effect (DEFF; Kish, 1965).

Estimates of the sample size necessary for GRTs must take the VIF into account to have adequate power to detect an intervention effect. To determine the sample size needed, investigators who are planning a GRT need to have estimates of the ICC expected in the planned trial. Because ICCs vary by outcome variable, group type and size, and study design (Murray & Blitstein, 2003), investigators should use ICCs from trials similar to the one they are planning. In addition, because ICC estimates may vary due to sampling error, investigators planning a GRT should have access to ICC estimates from multiple studies. However, in only a few substantive areas, such as smoking prevention and physical activity (Murray & Blitstein, 2003), are there enough published reports of ICCs to allow investigators to select those that are most applicable to a planned trial.

ICCs can also vary due to the inclusion of covariates in the analytic model. The variation between groups in a GRT, which forms the numerator of the ICC, can be due to either individual or group-level characteristics. To the extent that these variations can be identified, they may be used to explain between-group variation and thus reduce the ICC (Murray & Blitstein, 2003). A reduction in the ICC, with all other factors held constant, will result in a smaller VIF and greater power to detect an intervention effect in a GRT. Group-level covariates can reduce the ICC by explaining between-group variation, but

they also reduce the denominator degrees of freedom for the F statistic in a GRT analysis testing for intervention effects. If all other quantities remain the same, a reduction in denominator degrees of freedom will decrease the size of the F statistic and increase the critical value for determining statistical significance. If the decrease in the ICC due to the inclusion of group-level covariates is not sufficient to offset the loss of degrees of freedom, group-level covariates can actually reduce power to detect an intervention effect in a GRT. Individual-level covariates do not reduce degrees of freedom, so any reduction in the ICC due to the inclusion of such covariates in the model increases power. Given the small number of groups included in most GRTs, it is advantageous to explore reducing the ICC through the use of individual-level covariates. It is important to note that covariate inclusion has the potential to increase or decrease the intervention effect size. If including a covariate reduces the difference between study conditions by explaining some of the between-condition variability, the intervention effect size may be reduced. If, however, the difference between conditions remains the same while the ICC is reduced, the intervention effect size will increase.

Although a substantial number of domestic and international HIV/STD prevention trials have involved random assignment of groups, to our knowledge only two papers have published ICC estimates from group-randomized HIV/STD prevention trials (Feldblum et al., 1999; Hughes, 2005). Feldblum et al. (1999) reported ICCs for the prevalence of STDs from a community-randomized trial in rural Kenya, and Hughes (2005) reported ICCs for the prevalence of STDs from a city-randomized trial in Peru. Although these ICCs may be useful to investigators planning similar trials, they are not likely to be useful in domestic trials or trials with other types of outcome variables, such as sexual risk behaviors.

Our goals in conducting this analysis were to (a) estimate ICCs for a variety of sexual behavior and other variables using data from a GRT targeting young African American and Latina women, (b) compare unadjusted ICCs with ICCs adjusted for individual-level covariates, (c) make these estimates available to investigators designing future trials, and (d) demonstrate sample size estimation for GRTs.

METHOD

Study Design and Sampling

The data we used to estimate ICCs were from the Prevention Options for Women Equal Rights (POWER) study (Bull et al., 2008), a GRT designed to evaluate the effectiveness of a campaign promoting condom awareness and use among 15- to 25-year-old African American and Latina women. In the POWER study, data were collected from women before and after the condom awareness campaign; thus, the design was a pretest–posttest serial cross-sectional GRT. Twelve noncontiguous neighborhoods in four western U.S. cities (Los Angeles/Long Beach, Oakland/San Francisco, Las Vegas, and San Diego) were selected for the study, and neighborhoods within each city were then randomly assigned (a priori stratification) to either intervention or control status. At the precampaign and postcampaign measurements, participants were recruited via a time–space sampling approach (Muhib et al., 2001). After identifying venues within the neighborhoods at which members of the target population were likely to be found, POWER study researchers visited each venue and, through observation and interviews with members of the target population and key informants, identified intervals called “venue-day-time” units (VDTs) during which members of the target population were

expected to gather. They then eliminated from the sampling frame VDTs that were determined to be unsafe or for which they were unable to obtain a venue owner's permission to collect data. From the remaining set of VDTs, they randomly selected VDTs during which women were screened for eligibility, and those found to be eligible were asked to complete the questionnaire.

Measures

Awareness of Condoms. Participants' awareness of condoms was assessed by questions such as, "Have you ever heard of a male condom?" "Have you ever seen any information or ads about male condoms?" "Have you ever read any information about male condoms?" and "Have you ever seen a male condom?"

Outcome Expectancies, Norms, and Intentions Related to Condom Use. Negative and positive outcome expectancies were measured with two 3-item scales, with items such as, "How likely is it that your partner would think you were having sex with another person if you said you had to use a condom?" and "How likely is it that your partner would be happier if you used a condom?" The condom norms scale included 5 items asking women how likely women like them would be to get and use condoms. Participants' intentions to use condoms were measured on the basis of their responses, on a 3-item scale, to questions such as, "During the next month, how likely is it that you will try to get condoms?" Response formats for all four scales were 5-point Likert-type scales ranging from *not at all likely* to *extremely likely*. Reliability for the four scales was adequate, with Cronbach's alpha coefficients ranging from .76 to .93.

Sexual Behavior. Participants were asked several questions about their sexual behavior, including whether they had ever had sex, the frequency with which they had had sex and used condoms during the previous 90 days, and whether they had used a condom the last time they had sex. Participants were also asked whether they had had an STD during the previous year and were given the following response choices: "No"; "Yes"; "I think so, but I'm not sure"; and "I have no idea." Responses of "I think so, but I'm not sure" were recoded as "Yes" and responses of "I have no idea" were treated as missing to create a dichotomous (yes-no) variable.

We treated two variables as covariates: age (15-17, 18-19, and 20-25) and age-appropriate education (yes-no). Participants were asked their age in years and the number of years of school they had completed. Those who were aged 18 or older who had 12 years of education were considered to have age-appropriate education. We considered respondents younger than 18 to have age-appropriate education if their age in years was no more than 6 more than the number of years of education they reported having had.

Statistical Methods

For each outcome variable, we fit two separate regression models to data collected before the intervention and to data collected after the intervention. Model 1 for pre-campaign data included only the neighborhood as a random effect, and Model 2 at pre-campaign included neighborhood as a random effect and age and age-appropriate education as fixed effects. Models 1 and 2 for postcampaign data were similar to those for precampaign data, except that they both also included the intervention condition as a fixed effect (Giraudeau, 2006). All models estimated both a variance component for neighborhood and a residual variance. For dichotomous variables, we estimated

variance components using the SAS GLIMMIX macro in SAS Software Version 9.1 (SAS Institute, 2003). GLIMMIX fits generalized linear mixed models using restricted pseudo-likelihood estimation (Wolfinger & O'Connell, 1993). We used a logit link to perform a mixed-model logistic regression analysis and converted variance components from the logit scale to the linear scale prior to computing the ICC (Murray, 1998). We estimated variance components for continuous variables using SAS PROC MIXED, also in SAS Software Version 9.1 (SAS Institute, 2003). ICCs were computed as follows:

$$\frac{\sigma_g^2}{\sigma_g^2 + \sigma_m^2},$$

where σ_g^2 is the group component of variance and σ_m^2 is the member component of variance, or residual. We estimated confidence intervals using Searle's (1971) method. The percentage change in the ICC from Model 1 to Model 2 was computed as $(\rho_2 - \rho_1)/\rho_1$ multiplied by 100, where ρ_1 is the ICC from Model 1 and ρ_2 is the ICC from Model 2.

RESULTS

Table 1 presents the frequencies of several demographic and sexual behavior variables among study participants. Across the 12 neighborhoods, 3,407 women participated in the precampaign survey, and 3,003 participated in the postcampaign survey; we excluded data for 195 (5.7%) of the participants in the precampaign survey and 145 (4.8%) of the participants in the postcampaign survey because they were missing one or both of the covariates included in the adjusted models (age and age-appropriate education). Most of the participants in both surveys were under 20 years of age and had achieved an age-appropriate level of education. Of those who responded to the race/ethnicity question, more than 70% identified themselves as African American or Latina. Most of the women who reported their monthly income earned less than \$1,000 per month, with a substantial minority (about 40%) earning less than \$250 per month. Most of the women in both samples had had sex at least once during their lifetime, with 15% to 20% reporting having had more than five lifetime sexual partners.

Table 2 presents the estimated ICCs and 95% confidence intervals for condom awareness variables. Most of the ICCs for condom awareness variables were larger than .05, indicating that more than 5% of the variance in these variables was due to the neighborhood. Postcampaign ICCs were consistently larger than precampaign ICCs, with differences of .05 or greater for several variables. For example, in the precampaign survey, the unadjusted ICC from Model 1 for "ever seen a male condom" was .0190, and in the postcampaign survey, the same ICC from Model 1 was .0254. When we adjusted for participants' age and age-appropriate education, most of the ICCs for condom awareness variables increased; however, the increases were small (generally less than 5%), and some of the decreases were substantial.

Table 3 presents ICCs for participants' outcome expectancies, norms, and intentions related to condom use. Adjustment for covariates resulted in decreases in the ICC for all of these variables; most were small decreases, but for intention to use condoms, the difference at both time points was greater than 20%. ICCs for intention to use condoms were very small at both time points, particularly for adjusted models. As with the awareness variables, ICCs were usually higher in the postcampaign survey than in the

Table 1. Characteristics of Prevention Options for Women Equal Rights (POWER) Participants Included in Analysis Estimating Intraclass Correlation Coefficients (ICCs)

Variable	Precampaign (<i>n</i> = 3,212): Number (%)	Postcampaign (<i>n</i> = 2,858): Number (%)
Intervention condition		
Control	1,592 (49.6)	1,410 (49.3)
Intervention	1,620 (50.4)	1,448 (50.7)
Age, years		
15-17	1,359 (42.3)	1,207 (42.2)
18-19	628 (19.6)	530 (18.5)
20-25	1,225 (38.1)	1,121 (39.2)
Age-appropriate education		
Yes	2,358 (73.4)	2,243 (78.5)
No	854 (26.6)	615 (21.5)
Race/ethnicity		
African American	1,078 (34.9)	919 (35.0)
Latina/Hispanic	1,378 (44.6)	1,036 (39.4)
Other	636 (20.6)	673 (25.6)
Monthly income		
\$0-\$250	1,102 (39.5)	898 (38.4)
\$251-\$500	475 (17.0)	314 (13.4)
\$501-\$1,000	612 (21.9)	523 (22.3)
>\$1,000	601 (21.5)	606 (25.9)
Ever had sex		
Yes	2,219 (69.6)	1,923 (67.4)
No	969 (30.4)	929 (32.6)
Lifetime number of sex partners		
0	969 (31.3)	929 (32.6)
1-5	1,574 (50.9)	1,221 (44.9)
>5	548 (17.7)	569 (20.9)

NOTE: Numbers may not total 100% or full number of participants due to rounding error or missing data.

precampaign survey, though the ICCs for condom norms were substantially smaller in the postcampaign survey.

ICCs for sexual behavior variables, presented in Table 4, were generally smaller than those for awareness of condoms variables, with most between .01 and .03. As with the awareness variables, the differences in ICCs between Models 1 and 2 were inconsistent, but for some sexual behavior variables, adjustment resulted in large increases in ICCs. For the two sexual behavior variables calculated both for the entire sample and for only those who reported ever having had sex, the adjusted ICCs were larger when the sample was limited to women who reported having had sex. As with the other variables, ICCs were usually higher for postcampaign survey data than for precampaign survey data.

DISCUSSION

This study presents the first ICC estimates that we are aware of from a behavioral intervention trial focused on reducing risky sexual behavior. Although these estimates

Table 2. Intraclass Correlation Coefficients (ICCs) and 95% Confidence Intervals (CIs) for Awareness of Condoms Variables Among Participants in the Prevention Options for Women Equal Rights (POWER) Study

Variable	Model 1: ^a ICC (95% CI)	Model 2: ^b ICC (95% CI)	Model 2 Versus Model 1 (%)
Ever heard of a male condom			
Precampaign rate: 93.9%	.0100 (.0032-.0350)	.0063 (.0013-.0247)	-37.39
Postcampaign rate: 94.4%	.0185 (.0073-.0587)	.0186 (.0074-.0590)	0.59
Seen any info on a male condom			
Precampaign rate: 84.7%	.0264 (.0116-.0787)	.0221 (.0093-.0674)	-16.50
Postcampaign rate: 83.0%	.0723 (.0356-.1888)	.0729 (.0359-.1902)	0.83
Read any info on a male condom			
Precampaign rate: 77.5%	.0423 (.0198-.1186)	.0405 (.0189-.1141)	-4.30
Postcampaign rate: 75.4%	.0772 (.0383-.1997)	.0818 (.0408-.2094)	5.88
Ever seen a male condom			
Precampaign rate: 89.7%	.0190 (.0078-.0594)	.0166 (.0065-.0529)	-12.87
Postcampaign rate: 88.2%	.0254 (.0108-.0768)	.0268 (.0116-.0805)	5.67

a. Model 1 for precampaign data included only the neighborhood as a random effect; Model 1 at postcampaign also included intervention condition as a fixed effect.

b. Model 2 at precampaign included neighborhood as a random effect and age and age-appropriate education as fixed effects; Model 2 at postcampaign also included intervention condition as a fixed effect.

are specific to the POWER study design, the outcome variables measured, and the sampling method used, they may be useful to investigators planning similar trials. ICC estimates for the POWER variables of interest ranged from about .003, a small ICC, to about .16, a very large ICC. Investigators should be aware, however, that the impact of the ICC on the sample size needed for a GRT also depends on the number of members per group. For example, with 20 members per group and an ICC of .003, the VIF would be 1.057, only a slight increase in the variance due to within-group correlation. However, with 200 members per group and an ICC of .003, the VIF would be 1.597, indicating an increase in the variance of more than 50% attributable to the group-randomized design. One author commented that “a smaller intraclass correlation coefficient indicates greater power and will not require significant adjustment to the sample size” (Parker, Evangelou, & Eaton, 2005, p. 261). This statement reflects a common misunderstanding concerning ICCs that can lead to a substantially underpowered study, particularly in trials with large groups.

ICCs were generally larger at postcampaign than precampaign; this probably resulted from the contamination of study conditions that was discovered at the postcampaign survey (Bull et al., 2008). If the intervention affected outcomes in neighborhoods in a way that could not be controlled for by the variable representing the condition (intervention or control) to which a neighborhood was assigned, this could increase the variability between neighborhoods and thus increase the ICC. Investigators designing trials with neighborhood as the unit of assignment should take great care to avoid such contamination for two reasons: (a) to avoid attenuating the intervention effect by delivering the intervention to control communities, and (b) to avoid increasing the ICC and thereby decreasing their study’s power to detect an intervention effect.

Table 3. Intraclass Correlation Coefficients (ICCs) and 95% Confidence Intervals (CIs) for Outcome Expectancies, Norms, and Intentions to Use Condoms Among Participants in the Prevention Options for Women Equal Rights (POWER) Study

Variable	Model 1: ^a ICC (95% CI)	Model 2: ^b ICC (95% CI)	Model 2 Versus Model 1 (%)
Intention to use condoms scale			
Precampaign mean (SD): 2.98 (1.44)	.0044 (.0001-.0202)	.0033 (.0005-.0170)	-26.30
Postcampaign mean (SD): 2.94 (1.49)	.0059 (.0007-.0253)	.0046 (.0000-.0217)	-21.75
Condom norms scale			
Precampaign mean (SD): 3.09 (1.25)	.0337 (.0152-.0978)	.0321 (.0143-.0936)	-5.01
Postcampaign mean (SD): 2.95 (1.37)	.0103 (.0029-.0373)	.0094 (.0024-.0348)	-8.74
Positive outcome expectations scale			
Precampaign mean (SD): 3.21 (1.21)	.0264 (.0113-.0792)	.0252 (.0108-.0762)	-4.36
Postcampaign mean (SD): 3.18 (1.29)	.0268 (.0113-.0811)	.0259 (.0109-.0788)	-3.28
Negative outcome expectations scale			
Precampaign mean (SD): 1.82 (1.02)	.0116 (.0038-.0400)	.0113 (.0036-.0392)	-2.68
Postcampaign mean (SD): 1.93 (1.15)	.0429 (.0197-.1214)	.0417 (.0191-.1186)	-2.70

a. Model 1 for precampaign data included only the neighborhood as a random effect; Model 1 at postcampaign also included intervention condition as a fixed effect.

b. Model 2 at precampaign included neighborhood as a random effect and age and age-appropriate education as fixed effects; Model 2 at postcampaign also included intervention condition as a fixed effect.

Even for the precampaign survey data, ICCs for the condom awareness variables were generally larger than for the sexual behavior and attitudes, norms, and intentions variables. This may be due to local prevention efforts within neighborhoods that exposed women to condom-related materials or to differences between communities in the prevalence of other variables that predict condom awareness that were not included in our analysis.

Adjustment for age and age-appropriate education yielded consistent reductions in the ICC only for the sets of variables in the outcome expectancies, norms, and intention to use condoms categories. For some variables, covariate adjustment resulted in large percentage increases in the ICC. Murray and Blitstein (2003) pointed out that this can happen when covariate imbalance across groups makes groups seem more alike on a variable of interest than they really are. The impact of covariates on ICCs for sexual risk behavior variables should be examined further in other GRTs. If covariates that consistently explain neighborhood-to-neighborhood variability in sexual risk outcome variables can be identified, they can be used in the planning of future trials. If specified a priori, such covariates may be used to reduce the ICC and increase the power of a study

Table 4. Intraclass Correlation Coefficients (ICCs) and 95% Confidence Intervals (CIs) for Sexual Behavior Variables Among Participants in the Prevention Options for Women Equal Rights (POWER) Study

Variable	Model 1: ^a ICC (95% CI)	Model 2: ^b ICC (95% CI)	Model 2 Versus Model 1 (%)
Had sex during past 90 days			
Precampaign rate: 57.5%	.0102 (.0032-.0358)	.0063 (.0013-.0251)	-38.03
Postcampaign rate: 52.8%	.0279 (.0120-.0834)	.0286 (.0124-.0852)	2.62
Used a condom at last sex			
Precampaign rate: 56.8%	.0177 (.0063-.0585)	.0163 (.0056-.0547)	-8.02
Postcampaign rate: 56.4%	.0237 (.0089-.0755)	.0256 (.0099-.0805)	8.16
Used a male condom at last sex			
Precampaign rate: 56.5%	.0159 (.0053-.0538)	.0148 (.0048-.0508)	-7.04
Postcampaign rate: 56.5%	.0237 (.0089-.0760)	.0265 (.0103-.0831)	11.67
Had unprotected sex during past 90 days (entire sample)			
Precampaign rate: 33.4%	.0100 (.0032-.0351)	.0109 (.0036-.0374)	8.27
Postcampaign rate: 29.8%	.0371 (.0168-.1065)	.0426 (.0197-.1201)	14.91
Had unprotected sex during past 90 days (those who ever had sex)			
Precampaign rate: 48.0%	.0126 (.0037-.0451)	.0143 (.0045-.0494)	12.83
Postcampaign rate: 44.7%	.0606 (.0282-.1655)	.0628 (.0294-.1706)	3.73
Had an STD in past year (self-report, entire sample)			
Precampaign rate: 8.1%	.0081 (.0022-.0297)	.0103 (.0034-.0358)	27.72
Postcampaign rate: 8.3%	.0144 (.0052-.0477)	.0130 (.0045-.0440)	-9.58
Had an STD in past year (self-report, those who ever had sex)			
Precampaign rate: 11.6%	.0176 (.0063-.0582)	.0200 (.0075-.0646)	13.80
Postcampaign rate: 12.3%	.0182 (.0061-.0612)	.0197 (.0069-.0652)	8.42

a. Model 1 for precampaign data included only the neighborhood as a random effect; Model 1 at postcampaign also included intervention condition as a fixed effect.

b. Model 2 at precampaign included neighborhood as a random effect and age and age-appropriate education as fixed effects; Model 2 at postcampaign also included intervention condition as a fixed effect.

for primary outcome analyses; however, this use of covariates as eligibility criteria may also reduce the generalizability of a study's findings.

For the two variables that we computed both with the entire sample and then with only those women who reported ever having had sex, the ICCs were always smaller when the entire sample was used. This is probably because the same community characteristics that influence whether a woman becomes sexually active may also influence whether a sexually active woman uses protection. When women who report never having had sex are excluded from such an analysis, communities become more variable because the denominator used in calculating the rate of unprotected sex in a particular community is a function of the percentage of community residents reporting ever having had sex. Investigators should keep this effect in mind when choosing outcome variables for intervention trials. They should also remember that an effective sexual

behavior intervention will decrease the percentage of women who report having had sex during a particular time period, so excluding women on the basis of their reported sexual behavior could obscure an intervention effect even if the ICC did not increase.

As previously noted, one of our purposes in publishing these ICCs is to aid investigators in planning future trials. Recent reviews have found that many GRTs had few and sometimes only one group per study condition (Simpson, Klar, & Donner, 1995; Varnell, Murray, & Baker, 2001). The one group per condition design is invalid for GRTs because between-group variation cannot be separated from between-condition variation. Studies with few groups assigned to each condition may not be much better because randomization may not be able to distribute sources of bias evenly across conditions, and the power of such studies to detect an intervention effect may be very low.

Murray (1998) presented the following formula for estimating the number of groups needed in a GRT (posttest only, difference between two study conditions):

$$g = \frac{2\hat{\sigma}_y^2[1 + (m - 1)ICC_{m:g:c}](t_{critical:\alpha/2} + t_{critical:\beta})^2}{m\hat{\Delta}^2},$$

where g is the number of groups per condition, $\hat{\sigma}_y^2$ is the residual variance of the dependent variable, m is the number of members per group, $ICC_{m:g:c}$ is the ICC measuring the correlation of members nested within groups, nested within study conditions, $t_{critical:\alpha/2}$ and $t_{critical:\beta}$ are critical values from the t distribution, and $\hat{\Delta}^2$ is the squared difference between study condition means. We focus on the number of groups per condition rather than the total sample size here because adding extra groups has a greater impact on GRT power than increasing the number of members per group. This is because increasing group size increases the variance inflation factor, whereas adding groups increases the overall sample size and thus increases the study's power without affecting the variance inflation factor. The same formula, adapted for studies with a binary primary outcome variable, is

$$g = \frac{[\hat{P}_1(1 - \hat{P}_1) + \hat{P}_2(1 - \hat{P}_2)][1 + (m - 1)ICC_{m:g:c}](t_{critical:\alpha/2} + t_{critical:\beta})^2}{m(\hat{P}_1 + \hat{P}_2)^2},$$

where \hat{P}_1 and \hat{P}_2 are the event rates in the two study conditions.

If we use the unadjusted ICC from the POWER study precampaign data for unprotected sex during the past 90 days (.01004), and assume that there are 200 members per group and that 35% of those in the control condition and 25% of those in the intervention condition will report having had unprotected sex, the number of groups needed for each condition (with an alpha level of .05 and power of 80%) is

$$6.006 = \frac{[.35(1 - .35) + .25(1 - .25)][1 + (200 - 1).01004](-2.23 - 0.88)^2}{200(.35 - .25)^2},$$

or about 6 groups per condition. The reader should note that this formula often requires iteration, changing the values for the t distribution as the number of groups changes.

To demonstrate the impact of a much larger ICC, we repeat the calculation, assuming an ICC of .05. This would increase the necessary sample size as shown:

$$18.844 = \frac{[.35(1 - .35) + .25(1 - .25)][1 + (200 - 1).05](-2.03 - 0.85)^2}{200(.35 - .25)^2},$$

or about 19 groups per condition. Thus, the importance of obtaining ICC estimates relevant to a planned trial and computing the sample size needed using those estimates cannot be overstated. Murray (1998) discussed the impact of ICC, the number of groups per condition, and the number of members per group in more detail and presented several charts showing the detectable difference (between study conditions) as a function of these parameters. Investigators should also always take into account the variability in ICC estimates from prior studies and perform a sensitivity analysis when planning a trial (Feng & Grizzle, 1992).

Sample-size formulas for GRTs can be found in a variety of references, including two recent comprehensive texts (Donner & Klar, 2004; Murray, 1998). Murray (1998) presented formulas that account for multiple measurements taken over time and covariate adjustment and can be used to determine the sample size needed in a proposed study or the detectable difference between study conditions. Donner and Klar (2004) presented formulas for a variety of outcome variable distributions and also discussed several strategies for increasing power in a GRT.

The ICC estimates we present are specific to the design of the POWER study (stratified serial cross-sectional), the sampling method (venue-based) used, the type of groups randomized (neighborhoods), the outcome variables chosen, and the participants recruited (15- to 25-year-old mostly African American and Latina women) and thus cannot be generalized to all HIV/STD prevention studies. In addition, because the trial involved only 12 neighborhoods, our ICC estimates are less precise than ICC estimates for a GRT with many more groups (i.e., confidence intervals around some of the ICC estimates are wide). However, currently there are no published estimates of ICCs for behavioral variables in HIV/STD prevention, so this study contributes valuable information by providing investigators with estimates to use as a starting point for study planning. Also, it is important to recognize that sample size estimation for a GRT should make use of multiple ICC estimates from different studies rather than an estimate from a single study. Currently, this is not possible for behavioral HIV/STD prevention trials because of the lack of published estimates. More work estimating ICCs for behavioral variables in HIV/STD prevention and exploring the impact of covariates on these ICCs is critically needed to strengthen conclusions drawn about the magnitude of ICCs for behavioral HIV/STD prevention outcome variables.

IMPLICATIONS FOR PRACTICE

Public health practitioners should consider using GRT designs for evaluation of programs in HIV/STD prevention. Often, behavioral HIV/STD interventions are initially evaluated for efficacy under carefully controlled experimental conditions and then adapted to fit the needs of community-based organizations (CBOs). Evaluation of these adaptations and other programs developed by CBOs could be done using a GRT design, with sample size based on our estimates and similar estimates from other studies. The GRT design is a much stronger design than others often used for this type of evaluation, such as pretest–posttest no-control-group designs, because it allows causal inference, where nonrandomized studies do not.

Our estimates could also be used to design large surveys (e.g., needs assessment surveys) in CBOs or other organizations working in HIV/STD prevention. ICCs and VIFs (referred to as “design effects” in survey sampling) can be used to determine how many organizations and survey respondents within organizations are needed for a given confidence interval width around estimates of interest.

We strongly urge those planning GRTs in HIV/STD prevention to consider issues of sample size carefully. Studies with few groups per condition risk producing a biased estimate of the intervention effect or not having enough power to detect an intervention effect. GRTs are often very expensive studies, and it is unfortunate that many are designed with too little power to detect even a moderate or large intervention effect.

References

- Bull, S. S., Posner, S. F., Ortiz, C., Beaty, B., Benton, K., Lin, L., et al. (in press). A social marketing campaign promoting female and male condoms. *Journal of Adolescent Health*, 43(1), 71-78.
- Donner, A., Birkett, N., & Buck, C. (1981). Randomization by cluster: Sample size requirements and analysis. *American Journal of Epidemiology*, 114(6), 906-914.
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.
- Donner, A., & Klar, N. (2004). Pitfalls of and controversies in cluster randomization trials. *American Journal of Public Health*, 94(3), 416-422.
- Feldblum, P. J., Chen-Mok, M., Bwayo, J. J., Omari, M., Kuyoh, M., & Ryan, K. A. (1999). Intraclass correlation of STD prevalence in a community intervention trial in Kenya. *The Lancet*, 354(9187), 1356-1357.
- Feng, Z., Diehr, P., Peterson, A., & McLerran, D. (2001). Selected statistical issues in group randomized trials. *Annual Review of Public Health*, 22, 167-187.
- Feng, Z., & Grizzle, J. E. (1992). Correlated binomial variates: Properties of estimator intraclass correlation and its effect on sample size calculation. *Statistics in Medicine*, 11(12), 1607-1614.
- Giraudeau, B. (2006). Model mis-specification and overestimation of the intraclass correlation coefficient in cluster randomized trials. *Statistics in Medicine*, 25(6), 957-964.
- Hughes, J. P. (2005). Using baseline data to design a group randomized trial. *Statistics in Medicine*, 24(13), 1983-1994.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley.
- Klar, N., & Donner, A. (2001). Current and future challenges in the design and analysis of cluster randomization trials. *Statistics in Medicine*, 20(24), 3729-3740.
- Muhib, F. B., Lin, L. S., Steuve, A., Miller, R. L., Ford, W. L., Johnson, W. D., et al. (2001). A venue-based method for sampling hard-to-reach populations. *Public Health Reports*, 116(Suppl. 1), 216-222.
- Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.
- Murray, D. M., & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review*, 27(1), 79-103.
- Murray, D. M., Varnell, S. P., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health*, 94(3), 423-432.
- Parker, D. R., Evangelou, E., & Eaton, C. B. (2005). Intraclass correlation coefficients for cluster randomized trials in primary care: The Cholesterol Education and Research Trial (CEART). *Controlled Clinical Trials*, 26(2), 260-267.

- Pedlow, C. T., & Carey, M. P. (2003). HIV sexual risk-reduction interventions for youth: A review and methodological critique of randomized controlled trials. *Behavior Modification, 27*(2), 135-190.
- Robin, L., Dittus, P., Whitaker, D., Crosby, R., Ethier, K., Mezooff, J., et al. (2004). Behavioral interventions to reduce incidence of HIV, STD, and pregnancy among adolescents: A decade in review. *Journal of Adolescent Health, 34*(1), 3-26.
- SAS Institute. (2003). *SAS/STAT Software for PC* (Version 9.1). Cary, NC: Author.
- Searle, S. R. (1971). *Linear models*. New York: Wiley.
- Semaan, S., Kay, L., Strouse, D., Sogolow, E., Mullen, P. D., Neumann, M. S., et al. (2002). A profile of U.S.-based trials of behavioral and social interventions for HIV risk reduction. *Journal of Acquired Immune Deficiency Syndromes, 30*(Suppl. 1), S30-S50.
- Simpson, J. M., Klar, N., & Donner, A. (1995). Accounting for cluster randomization: A review of primary prevention trials, 1990 through 1993. *American Journal of Public Health, 85*(10), 1378-1383.
- Varnell, S. P., Murray, D. M., & Baker, W. L. (2001). An evaluation of analysis options for the one-group-per-condition design: Can any of the alternatives overcome the problems inherent in this design? *Evaluation Review, 25*(4), 440-453.
- Varnell, S. P., Murray, D. M., Janega, J. B., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health, 94*(3), 393-399.
- Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation, 48*, 233-243.
- Zucker, D. M. (1990). An analysis of variance pitfall: The fixed effects analysis in a nested design. *Educational and Psychological Measurement, 50*, 731-738.