

## Intraclass Correlation Estimates in a School-based Smoking Prevention Study

### Outcome and Mediating Variables, by Sex and Ethnicity

Ohidul Siddiqui, Donald Hedeker, Brian R. Flay, and Frank B. Hu

Most school-based smoking prevention studies employ designs in which schools or classrooms are assigned to different treatment conditions while observations are made on individual students. This design requires that the treatment effect be assessed against the between-school variance. However, the between-school variance is usually larger than the variance that would be obtained if students were individually randomized to different conditions. Consequently, the power of the test for a treatment effect is reduced, and it becomes difficult to detect important treatment effects. To assess the potential loss of power or to calculate appropriate sample sizes, investigators need good estimates of the intraclass correlations for the variables of interest. The authors calculated intraclass correlations for some common outcome variables in a school-based smoking prevention study, using a three-level model—i.e., students nested within classrooms and classrooms nested within schools. The authors present the intraclass correlation estimates for the entire data set, as well as separately by sex and ethnicity. They also illustrate the use of these estimates in the planning of future studies. *Am J Epidemiol* 1996;144:425–33.

adolescence; cluster analysis; epidemiologic methods; models, statistical; research design; smoking

In school-based smoking intervention studies, randomization of schools rather than students to different treatment conditions is the usual approach to sampling (1, 2). In this situation, the treatment effect must be evaluated on the basis of the between-school variance. Cornfield (3) pointed out that in randomization of schools to different treatment conditions, students contribute less information than they would if they were individually randomized to different treatment conditions. This occurs because the students within a school cannot be regarded as independent. As a result, the power of the test of the treatment effect is reduced (4–6).

If the number of students per school is fixed, the loss of information due to the randomization of schools can be compensated for by increasing the number of schools. However, this has cost implications and is not always feasible. Thus, it is important to efficiently determine the appropriate number of schools needed to achieve the study's goals. This means that an accurate

sample size calculation is needed in order to obtain the precise variance estimate and the power of the test of the treatment effect at a given level. An important aspect of sample size estimation is obtaining precise estimates of the intraclass correlations (the interdependence of the students within schools) for the outcome variables under study (7).

The intraclass correlation coefficient (ICC) measures the relatedness of the students within a group, such as a school or classroom. It is the ratio of the variance component due to schools or classrooms to the total variance for individual students. If the total variance is denoted

$$\sigma^2 = \sigma^2_{sch} + \sigma^2_{class} + \sigma^2_{student},$$

where  $\sigma^2_{sch}$  is the component of variance between schools,  $\sigma^2_{class}$  is the variance between classrooms within schools, and  $\sigma^2_{student}$  is the variance associated with students within classrooms and schools, then the ICC for the students within schools is conventionally defined (1, 8) as

$$ICC = \sigma^2_{sch} / (\sigma^2_{sch} + \sigma^2_{class} + \sigma^2_{student}).$$

The ICC for the students within classrooms is defined as

$$ICC = (\sigma^2_{sch} + \sigma^2_{class}) / (\sigma^2_{sch} + \sigma^2_{class} + \sigma^2_{student}).$$

Received for publication July 3, 1995, and in final form February 23, 1996.

Abbreviations: DEFF, design effect; ICC, intraclass correlation coefficient; IF, inflation factor; NIH, National Institutes of Health; S/R, social influences/resistance skills.

From the Prevention Research Center (M/C 275), School of Public Health, University of Illinois at Chicago, 850 West Jackson Boulevard, Suite 400, Chicago, IL 60607–3025.

Since the students within a classroom are more likely to be similar than the students within a school, the ICCs at the classroom level are always higher than or equal to (if  $\sigma^2_{\text{class}} = 0$ ) the ICCs at the school level.

For any specific outcome variable, the design effect (DEFF) is the ratio of the variance of an overall sample mean estimated from school means to the variance of an overall sample mean estimated from the students. Kish (9) defined the relation between the DEFF and the ICC as  $\text{DEFF} \approx [1 + (n - 1) \times \text{ICC}]$ , where  $n$  is the sample size per school. Donner et al. (10) defined DEFF as the inflation factor (IF). Since the DEFF is a function of  $n$  and the ICC, a zero ICC value is expected when  $\text{DEFF} = 1$ . While it is true that a positive ICC does not change the estimate of the overall mean for any outcome variable, the variance of the mean is changed (since the variance of the mean =  $(\sigma^2_{\text{student}}/mn) \times \text{IF}$ , where  $m$  is the number of schools and  $n$  is the number of students in each school). In such a situation, the extra variance of the outcome variable can reduce statistical power substantially. Therefore, to assure a test that has a power of  $1 - \beta$  with  $\alpha$  level of significance, the number of schools needed per condition depends on the estimate of the ICC for the outcome variable. Accurate sample size estimation thus requires the availability of precise ICC estimates for the outcome variables under study.

Recently, analysis of mediating variables in smoking prevention studies has drawn attention. For example, Mackinnon et al. (11) demonstrated that changes in normative social beliefs mediated the effects of an intervention program on adolescent cigarette use. If we are to learn more about how prevention programs might work, future smoking prevention studies must be designed to focus on mediating variables as well as the outcome variables of interest. Consequently, it will be important in planning future studies to take into account the ICCs for mediating variables in order to assure adequate power for tests of interest.

The National Institutes of Health (NIH) recently instituted a policy that sex and racial/ethnic issues must be considered when developing research designs and sample sizes in studies that are supported by NIH grants (12). This policy corresponds to the recent focus on assessing the efficacy of theory-based smoking prevention programs in culturally diverse populations. According to the NIH, a primary aim of intervention research is to determine whether or not the intervention being studied affects the sexes or members of various ethnic groups differently. The availability of ICCs for smoking-related outcome variables by sex and race will be necessary for designing future smoking prevention studies. While a few ICC estimates are available in the literature on school-based

smoking prevention studies (1), no one has yet attempted to estimate ICCs for smoking-related behavioral and mediating variables by sex and ethnic group. Consequently, the purposes of this paper are to provide estimates of ICCs (with standard errors) for some important smoking-related behavioral and mediating variables (henceforth called simply "outcome variables") in a school-based smoking prevention study, and to provide the estimates with respect to different sex and ethnic groups. We also show how to use these estimates to calculate sample sizes for future school-based smoking prevention studies.

## METHODS

### Subjects

The data for this study were collected as part of the Television, School, and Family Project, a longitudinal research project concerned with adolescent smoking prevention in Los Angeles and San Diego, California (13, 14). This project employed the most commonly used design in school-based health promotion studies, the nested cohort design. Randomization to various treatment conditions was made at the school level, while much of the intervention was delivered to students within classrooms. A cohort of 6,695 seventh-grade students from 287 classrooms in 47 public schools in Los Angeles and San Diego were surveyed at four time points. The students were pretested in January 1986 (wave A). Students who took the pretest completed an immediate postintervention questionnaire in April 1986 (wave B), a 1-year follow-up questionnaire in April 1987 (wave C), and a 2-year follow-up questionnaire in April 1988 (wave D). Of the original cohort of 6,695, 5,475 (81.8 percent), 4,854 (72.5 percent), and 3,729 (55.7 percent) were recontacted at waves B, C, and D, respectively. In the cohort at baseline, 49.6 percent of the students were female; 15.5 percent were black, 32.5 percent were white, 35.5 percent were Hispanic, and 16.5 percent were members of other ethnic groups (Japanese, Chinese, etc.—henceforth called Asians).

Attrition analysis indicated that smokers and blacks were more likely to be lost to follow-up than the other groups (13, 15). To deal with this issue of attrition for a given measure and wave, we first concentrated on students for whom we had available data on the measure at that wave (but not necessarily across waves or even across measures within the given wave). As a result, the number of students differed from measure to measure and from wave to wave. In a second approach, we also calculated the intraclass correlations for some of the measures considering only those students on whom we had available data for a given

measure at all waves. The two approaches yielded very similar intraclass correlations; in particular, the correlations had the same pattern over time. Thus, we report here results based on the first approach (i.e., based on students with available data for the particular measure at the given time point).

## Measures

**Tobacco and health knowledge.** Seven questionnaire items were used to assess students' knowledge of tobacco and health (tobacco knowledge). A student's score on the tobacco knowledge scale was the number of items the student answered correctly.

**Social influences/resistance skills knowledge.** Eight questionnaire items were used to assess students' knowledge of information taught about social influences/resistance skills (S/R knowledge). A student's score on the S/R knowledge scale was the number of items the student answered correctly.

**Refusal self-efficacy.** Two items assessed a student's confidence in refusing to smoke cigarettes. The refusal self-efficacy (refusal skill) scale score was the sum of the ratings for these two items.

**Coping effort.** Three items assessed a student's efforts to resist trying cigarettes. A student's coping effort scale score was the sum of the student's ratings for the three items.

**Intention.** Scores on two items designed to assess the intention to use cigarettes in the future were summed.

**Prevalence of student smoking.** Scores on two items designed to assess the prevalence of student smoking were summed.

**Current smoking status.** One questionnaire item was used to assess the current smoking status of the students.

The various scales and their reliabilities are described more fully elsewhere (13).

## Analysis

For each of the outcome variables except current smoking status, the ICCs were calculated at both school and classroom levels for each of the four waves. Since the treatment conditions had significant effects on mediating variables (13) such as knowledge, prevalence estimates, and coping efforts, we controlled for the treatment condition effects by including dummy-coded treatment condition effects as exogenous variables in the regression models used to calculate the variance components  $\sigma^2_{sch}$ ,  $\sigma^2_{class}$ , and  $\sigma^2_{student}$  for the outcome variables. The variance components were also computed separately for different sex and ethnic groups. Since our data were multilevel

(students nested within classrooms nested within schools) in an unbalanced design (numbers of students within classrooms and numbers of classrooms within schools were not constant), the ML3 multilevel analysis program (16) that utilizes iterative generalized least-squares estimation was used to estimate the variance components for the outcome variables. Following the estimation of the variance components, the ICCs for the outcome variables at the school and classroom level were estimated using the formula given above.

For current smoking status, the possible responses were "yes" and "no." For this dichotomous outcome, recently developed random-effects models for ordinal outcome data (17) were used to calculate the variance components after controlling for the treatment condition effects as before. Since, at present, this procedure has only been developed for two-level data (18), our strategy was to fit two-level models of both students within schools (treating schools as a random effect) and students within classrooms (treating classrooms as a random effect). For current smoking status, it was always the case that once dummy variables for treatment were included in the model, the school variance component for the students within-school was very close to zero. We thus based our results for current smoking status on the students-within-classrooms model. For both the continuous outcomes and the dichotomous outcomes, the standard errors for each of the ICC estimates were calculated using the variance formula for the ICC originally derived by Fisher (19) and more recently described by Donner and Koval (20):

$$\text{Var}(\hat{\rho}) = 2(1 - \hat{\rho})^2 [1 + (n - 1)\hat{\rho}]^2 / n(n - 1)m,$$

where  $n$  is the average (harmonic mean) number of students per school/classroom,  $m$  is the number of schools/classrooms, and  $\hat{\rho}$  is the ICC estimate of the outcome variable.

## RESULTS

Table 1 shows the estimated ICCs for each of the outcome variables at the school and classroom levels. As expected, ICCs at the classroom level were much higher than those at the school level. For example, the ICCs for tobacco knowledge at waves A, B, C, and D were 0.153, 0.131, 0.151, and 0.138, respectively, at the classroom level, and the corresponding ICCs at the school level were 0.098, 0.082, 0.096, and 0.078. Over time, the ICCs for the outcome variables at the school level did not seem to follow any systematic pattern, instead remaining fairly consistent.

At the classroom level, the ICCs for the outcome variables generally decreased over time—e.g., the

**TABLE 1. Estimates of Intraclass correlations ( $\hat{\rho}$ ) for outcome variables at the school and class level: Television, School, and Family Project, 1986–1988**

Outcome variable		Wave A (n = 5,113–6,586)*		Wave B (n = 4,763–5,421)		Wave C (n = 4,031–4,838)		Wave D (n = 2,928–3,724)	
		$\hat{\rho}$	SE†	$\hat{\rho}$	SE	$\hat{\rho}$	SE	$\hat{\rho}$	SE
Tobacco knowledge	School	0.098	0.009	0.082	0.011	0.096	0.021	0.078	0.019
	Class	0.153	0.010	0.131	0.014	0.151	0.017	0.138	0.019
S/R† knowledge	School	0.016	0.006	0.014	0.005	0.014	0.006	0.013	0.007
	Class	0.170	0.010	0.145	0.013	0.104	0.013	0.092	0.017
Coping effort	School	0.008	0.004	0.008	0.004	0.004	0.004	0.019	0.008
	Class	0.044	0.008	0.031	0.009	0.023	0.008	0.023	0.011
Intention to smoke	School	0.005	0.003	0.007	0.004	0.010	0.005	0.010	0.006
	Class	0.032	0.007	0.024	0.008	0.024	0.008	0.022	0.011
Refusal skills	School	0.008	0.004	0.009	0.004	0.019	0.007	0.011	0.006
	Class	0.015	0.005	0.013	0.007	0.031	0.009	0.017	0.011
Prevalence of student smoking	School	0.013	0.006	0.033	0.009	0.025	0.009	0.014	0.004
	Class	0.024	0.008	0.056	0.010	0.037	0.010	0.029	0.014
Current smoking status‡	Class	0.090	0.011	0.074	0.006	0.071	0.006	0.044	0.004

\* Since the sample size ( $n$ ) was different for different scales within each wave, the range of  $n$ 's for the different scales within each wave is given.

† SE, standard error; S/R, social influences/resistance skills.

‡ At the school level, the program was not able to produce a nonzero variance estimate; thus, the intraschool correlation is not estimated as being different from zero in this case.

ICCs for current smoking status were 0.090, 0.074, 0.071, and 0.044 at waves A, B, C, and D, respectively. The ICCs for the tobacco knowledge, S/R knowledge, and current smoking status (only at the classroom level) scales were always higher than the ICCs for the coping effort, intention, refusal skill, and smoking prevalence scales. This means that students within the schools and classrooms were more clustered with respect to their knowledge and behavior than they were with respect to the mediating variables.

Table 2 shows the ICC estimates for the scales by ethnic group. There was no obvious pattern in the changes in the ICCs over time with respect to race. To examine more formally whether detectable differences were present, we performed analysis of variance using the ICCs given in table 2 as the dependent variable and race, time, and the outcome scales as independent factors. The analysis of variance indicated that whites had, on average, the highest ICCs for all of the scales, blacks had the lowest ICCs for all of the scales, and Hispanics and Asians were intermediate. These results were maintained for both school- and class-level analyses.

Table 3 shows ICC estimates for the outcome variables by sex. The ICCs for all of the outcome variables except current smoking status did not follow any systematic pattern (higher/lower) by sex at the school and classroom levels. Analysis of variance was carried out considering the ICC estimates given in table 3 as the dependent variable and sex, time, and the outcome

scales as independent factors. The analysis of variance indicated that the ICC estimates for most of the scales were slightly higher for females than for males. In particular, the ICCs for current smoking status were higher for females than for males at waves A, B, and C, but at wave D, they were similar. Thus, in terms of smoking status, females were more highly clustered than males at the early waves, though the degree of clustering decreased over time.

## DISCUSSION

Most of the ICCs were clearly different from zero for all of the smoking-related outcome variables. Over time, the ICCs at the classroom level generally decreased. One possible explanation for the decrease in the ICCs over time might be that classroom level was defined in the seventh grade, and not all of the students stayed in the same classroom over time. The ICCs for the tobacco knowledge and S/R knowledge scales were higher than those for the smoking behavioral scales. The higher values of the ICCs suggest that students were more homogeneous in their knowledge of smoking consequences and social resistance than in their smoking behavior.

The ICC estimates by ethnicity indicated that whites were different from the other ethnic groups with respect to each of the knowledge and behavior scales related to smoking at early ages, but the difference diminished by later ages. One possible explanation is

TABLE 2. Estimates of intraclass correlations ( $\hat{\rho}$ ) for outcome variables with respect to ethnicity at the school and class level: Television, School, and Family Project, 1986-1988\*

Outcome variable			Wave A (n = 684-2,337)†		Wave B (n = 677-1,915)		Wave C (n = 402-1,707)		Wave D (n = 278-1,359)	
			$\hat{\rho}$	SE‡	$\hat{\rho}$	SE	$\hat{\rho}$	SE	$\hat{\rho}$	SE
			Tobacco knowledge	School	Black	0.000	0.093	0.000	0.093	0.035
		White	0.001	0.022	0.009	0.016	0.053	0.027	0.012	0.029
		Hispanic	0.009	0.023	0.040	0.028	0.017	0.037	0.059	0.041
		Asian	0.036	0.043	0.014	0.034	0.075	0.048	0.017	0.062
	Class	Black	0.000	0.080	0.041	0.080	0.035	0.080	0.000	0.080
		White	0.039	0.028	0.095	0.030	0.117	0.041	0.101	0.040
		Hispanic	0.038	0.027	0.115	0.030	0.105	0.029	0.105	0.038
		Asian	0.036	0.040	0.077	0.041	0.129	0.042	0.058	0.067
S/R‡ knowledge	School	Black	0.000	0.085	0.000	0.085	0.000	0.085	0.000	0.085
		White	0.025	0.024	0.018	0.016	0.013	0.020	0.013	0.030
		Hispanic	0.016	0.024	0.022	0.025	0.017	0.030	0.016	0.034
		Asian	0.000	0.032	0.000	0.032	0.000	0.032	0.000	0.032
	Class	Black	0.000	0.080	0.066	0.079	0.000	0.080	0.054	0.079
		White	0.065	0.029	0.130	0.031	0.099	0.040	0.065	0.039
		Hispanic	0.030	0.026	0.107	0.029	0.079	0.028	0.073	0.038
		Asian	0.058	0.040	0.091	0.041	0.128	0.042	0.103	0.066
Coping effort	School	Black	0.000	0.085	0.000	0.085	0.000	0.085	0.013	0.088
		White	0.019	0.013	0.029	0.018	0.007	0.018	0.008	0.023
		Hispanic	0.000	0.019	0.006	0.023	0.002	0.028	0.031	0.037
		Asian	0.000	0.032	0.000	0.032	0.008	0.047	0.000	0.060
	Class	Black	0.031	0.078	0.001	0.078	0.016	0.078	0.012	0.078
		White	0.065	0.028	0.043	0.027	0.026	0.062	0.008	0.036
		Hispanic	0.064	0.023	0.036	0.027	0.003	0.025	0.036	0.037
		Asian	0.044	0.039	0.055	0.039	0.021	0.038	0.000	0.037
Intention to smoke	School	Black	0.000	0.085	0.000	0.085	0.000	0.085	0.000	0.085
		White	0.015	0.012	0.000	0.013	0.000	0.014	0.013	0.022
		Hispanic	0.000	0.018	0.019	0.023	0.009	0.029	0.018	0.035
		Asian	0.001	0.028	0.000	0.027	0.000	0.046	0.000	0.060
	Class	Black	0.032	0.078	0.003	0.078	0.017	0.078	0.000	0.078
		White	0.034	0.027	0.025	0.026	0.027	0.026	0.032	0.037
		Hispanic	0.017	0.020	0.028	0.026	0.013	0.025	0.018	0.036
		Asian	0.039	0.038	0.016	0.038	0.001	0.037	0.000	0.064
Refusal skills	School	Black	0.000	0.085	0.000	0.085	0.000	0.085	0.000	0.085
		White	0.006	0.010	0.014	0.015	0.069	0.026	0.000	0.021
		Hispanic	0.002	0.018	0.003	0.020	0.016	0.030	0.017	0.030
		Asian	0.000	0.027	0.000	0.032	0.000	0.046	0.000	0.060
	Class	Black	0.000	0.078	0.000	0.078	0.006	0.078	0.000	0.078
		White	0.032	0.027	0.023	0.026	0.069	0.026	0.017	0.036
		Hispanic	0.011	0.020	0.006	0.025	0.025	0.036	0.033	0.036
		Asian	0.016	0.038	0.032	0.038	0.044	0.039	0.000	0.037
Prevalence of student smoking	School	Black	0.000	0.085	0.000	0.085	0.000	0.085	0.016	0.147
		White	0.034	0.024	0.044	0.022	0.017	0.021	0.023	0.035
		Hispanic	0.017	0.030	0.034	0.030	0.040	0.038	0.003	0.032
		Asian	0.031	0.042	0.026	0.036	0.036	0.051	0.000	0.060
	Class	Black	0.000	0.081	0.073	0.080	0.015	0.081	0.052	0.081
		White	0.044	0.028	0.065	0.029	0.040	0.038	0.034	0.038
		Hispanic	0.021	0.026	0.090	0.029	0.040	0.027	0.016	0.036
		Asian	0.054	0.041	0.026	0.040	0.037	0.040	0.013	0.067
Current smoking status§	Class	Black	0.120	0.076	0.040	0.077	0.001	0.077	0.001	0.077
		White	0.080	0.025	0.053	0.027	0.051	0.027	0.058	0.028
		Hispanic	0.091	0.024	0.058	0.022	0.025	0.020	0.001	0.019
		Asian	0.097	0.040	0.090	0.040	0.001	0.037	0.001	0.037

\* For some outcome variables, the program was not able to produce nonzero variance estimates; thus, the intraclass correlations are not estimated as being different from zero in these cases.

† Since the sample size ( $n$ ) was different for different scales within each wave, the range of  $n$ 's for the different scales within each wave is given.

‡ SE, standard error; S/R, social influences/resistance skills.

§ At the school level, the program was not able to produce a nonzero variance estimate; thus, the intraschool correlation is not estimated as being different from zero in this case.

that parental influences might be more salient for whites than for the other ethnic groups at early ages, while at later ages, the influence of friends is more salient than parental influence.

Regarding sex, the ICC estimates for all of the variables except current smoking status were very similar, and they were also similar to the estimates in the combined data set. This indicates that for most of

**TABLE 3. Estimates of intraclass correlations ( $\hat{\rho}$ ) for outcome variables with respect to sex at the school and class level: Television, School, and Family Project, 1986-1988**

Outcome variable			Wave A (n = 2,387-3,313)*		Wave B (n = 2,290-2,785)		Wave C (n = 1,915-2,461)		Wave D (n = 1,368-1,923)	
			$\hat{\rho}$	SE†	$\hat{\rho}$	SE	$\hat{\rho}$	SE	$\hat{\rho}$	SE
Tobacco knowledge	School	Male	0.027	0.011	0.052	0.016	0.075	0.021	0.064	0.021
		Female	0.019	0.010	0.028	0.011	0.102	0.027	0.089	0.025
	Class	Male	0.039	0.015	0.131	0.022	0.135	0.022	0.118	0.029
		Female	0.048	0.016	0.094	0.018	0.159	0.023	0.153	0.026
S/R‡ knowledge	School	Male	0.012	0.008	0.006	0.007	0.008	0.009	0.017	0.013
		Female	0.007	0.008	0.017	0.009	0.021	0.011	0.013	0.012
	Class	Male	0.049	0.016	0.098	0.019	0.089	0.020	0.086	0.028
		Female	0.076	0.017	0.121	0.020	0.093	0.020	0.122	0.025
Coping effort	School	Male	0.012	0.007	0.006	0.006	0.004	0.007	0.015	0.011
		Female	0.001	0.005	0.014	0.008	0.005	0.008	0.021	0.013
	Class	Male	0.025	0.010	0.029	0.015	0.036	0.015	0.032	0.020
		Female	0.076	0.014	0.042	0.014	0.025	0.014	0.026	0.020
Intention to smoke	School	Male	0.004	0.005	0.001	0.005	0.003	0.007	0.006	0.009
		Female	0.012	0.007	0.027	0.010	0.006	0.008	0.007	0.010
	Class	Male	0.030	0.011	0.017	0.014	0.002	0.013	0.012	0.019
		Female	0.050	0.013	0.045	0.014	0.059	0.016	0.022	0.020
Refusal skills	School	Male	0.009	0.006	0.002	0.005	0.021	0.010	0.006	0.010
		Female	0.004	0.005	0.013	0.008	0.017	0.011	0.019	0.012
	Class	Male	0.021	0.010	0.010	0.013	0.035	0.015	0.015	0.020
		Female	0.013	0.010	0.031	0.013	0.023	0.014	0.018	0.020
Prevalence of student smoking	School	Male	0.008	0.008	0.030	0.012	0.023	0.012	0.016	0.015
		Female	0.024	0.011	0.036	0.013	0.013	0.010	0.009	0.013
	Class	Male	0.009	0.013	0.051	0.016	0.033	0.021	0.016	0.025
		Female	0.042	0.015	0.077	0.017	0.020	0.016	0.060	0.027
Current smoking status‡	Class	Male	0.074	0.013	0.050	0.012	0.030	0.011	0.048	0.012
		Female	0.141	0.018	0.144	0.018	0.114	0.016	0.055	0.013

\* Since the sample size ( $n$ ) was different for different scales within each wave, the range of  $n$ 's for the different scales within each wave is given.

† SE, standard error; S/R, social influences/resistance skills.

‡ At the school level, the program was not able to produce a nonzero variance estimate; thus, the intraschool correlation is not estimated as being different from zero in this case.

the variables there was no differential clustering by sex at the school or classroom level. However, for current smoking status, the ICC estimates for females were higher than the estimates for males at waves A, B, and C, while at wave D, the estimates for males and females were similar. This implies that at earlier ages the females were more highly clustered in their smoking behavior. However, at later ages, the clustering within-classroom decreased, and the prevalence of current smoking among the female students did not vary greatly from classroom to classroom.

### Implications

Several authors have discussed how to use external estimates of intraclass correlations in calculating the

number of schools needed per treatment condition to detect a treatment effect (21, 22). We will illustrate below the use of the ICC estimates for sample size calculation by assuming fixed Type I and II errors. We will also show that even very small ICC values can have dramatic effects on sample size estimation. The usual formula (22) used for sample size estimation when there are two conditions is

$$n = (t_{\alpha/2} + t_{1-\beta})^2 2 \sigma^2 / \delta^2,$$

where  $n$  is the number of students required per treatment condition,  $t_{\alpha/2}$  is the value of the  $t$  distribution (for large  $n$ 's, it is equivalent to the normal distribution) representing the two-tailed Type I error rate,  $t_{1-\beta}$  is the value of the  $t$  distribution representing the de-

sired power level  $1 - \beta$ , 2 represents the number of treatment conditions to be compared,  $\sigma^2$  is the variance of the outcome variable, and  $\delta$  is the hypothesized difference between the two treatment condition means. If we want to estimate the number of schools  $m$  (assuming  $n$  students from each school) required for each treatment condition, the modified formula is

$$m = (t_{\alpha/2} + t_{1-\beta})^2 2 \sigma^2 / n \delta^2,$$

where  $\sigma^2/n$  is the theoretical variance of the school mean. To include the expected intraschool correlation in the data,  $\sigma^2/n$  must be multiplied by the IF (22):

$$m = (t_{\alpha/2} + t_{1-\beta})^2 2 \sigma^2 \text{IF} / n \delta^2,$$

where  $\text{IF} = [1 + (n - 1)\text{ICC}]$ . Assume that the Type I and II error rates are 0.05 and 0.20, respectively, and that the expected mean differences between two treatments (effect size) are 0.2 and 0.5, with variance  $\sigma^2 = 1.0$ . Cohen (23) conventionally defined a small effect size as 0.2, a medium effect size as 0.5, and a large effect size as 0.8; here, we are considering only small and medium effect sizes in order to demonstrate the effect of intraclass correlation on sample size estimation. We used SSIZE, a program written by Hsieh (21) for sample size calculation with clustered randomization, for calculating the number of schools,  $m$ , at different levels of intraclass correlations, varying the number of students,  $n$ , from 30 to 150 in each school. From table 4 and figure 1, it is clear that the number of schools needed per treatment condition increases as the ICC increases.

Next we will illustrate the implications of the NIH's policy of considering sex and racial/ethnic issues in sample size estimation. Let us consider the ICC estimates for a particular outcome variable, e.g., current smoking status. The estimate for the entire data set at the classroom level is 0.090 at wave A. The corre-

sponding estimates for males and females are 0.074 and 0.141. Assume that the Type I and II error rates are 0.05 and 0.20, respectively, and that the expected mean difference between two treatments (the effect size) is 0.5, with variance  $\sigma^2 = 1.0$ . The estimated numbers of schools required at average classroom sizes of 30, 90, and 150 are eight, seven, and seven, respectively. We assume that the sex ratio at the classroom level is 0.5. The estimated numbers of schools for males at average classroom sizes of 15, 45, and 75 (50 percent of classroom size) are nine, six, and six, and the estimated numbers of schools for females are 13, 10, and 10 (shown in figure 1). Any statistical test for females based on the estimated number of schools using the ICC estimate for the entire data set will have less power than it would if sex were ignored. However, any test for females or males based on the estimated number of schools required using the ICC for the females (since the ICC estimate for the females is larger) will have at least the expected power or more. Therefore, the availability of ICC estimates by sex is important for estimating appropriate sample sizes for future smoking research, and consequently research findings based on the appropriate sample size will be of benefit to all subjects under study. In the same way, we can also demonstrate the importance of considering ethnicity in estimating the appropriate sample size.

### Summary and conclusions

In school-based smoking prevention studies, analytic methods must account for the school as the unit of assignment to treatment conditions, and the adjustment is based on the intraclass correlation. In the absence of precise estimates of the intraclass correlation for outcome variables, sample size estimation will

**TABLE 4. Estimated number of schools ( $m$ ) at different intraschool correlations and at different numbers of students ( $n$ ) per school that provides at least 80% power to detect effect sizes of 0.2 and 0.5 between two treatment conditions: Television, School, and Family Project, 1986-1988\***

$n$	Intraschool correlation											
	Small effect size (0.2)						Medium effect size (0.5)					
	0.0	0.005	0.01	0.025	0.05	0.10	0.0	0.005	0.01	0.025	0.05	0.10
30	13	15	17	23	32	50	2	3	3	4	6	9
60	7	9	11	17	26	45	1	2	2	3	5	8
90	5	7	9	14	24	44	1	1	2	3	4	7
120	4	6	8	13	23	43	1	1	2	3	4	7
150	3	5	7	13	22	42	1	1	2	2	4	7

\* Assuming that  $\sigma^2 = 1.0$  and the Type I error  $\alpha = 0.05$ .

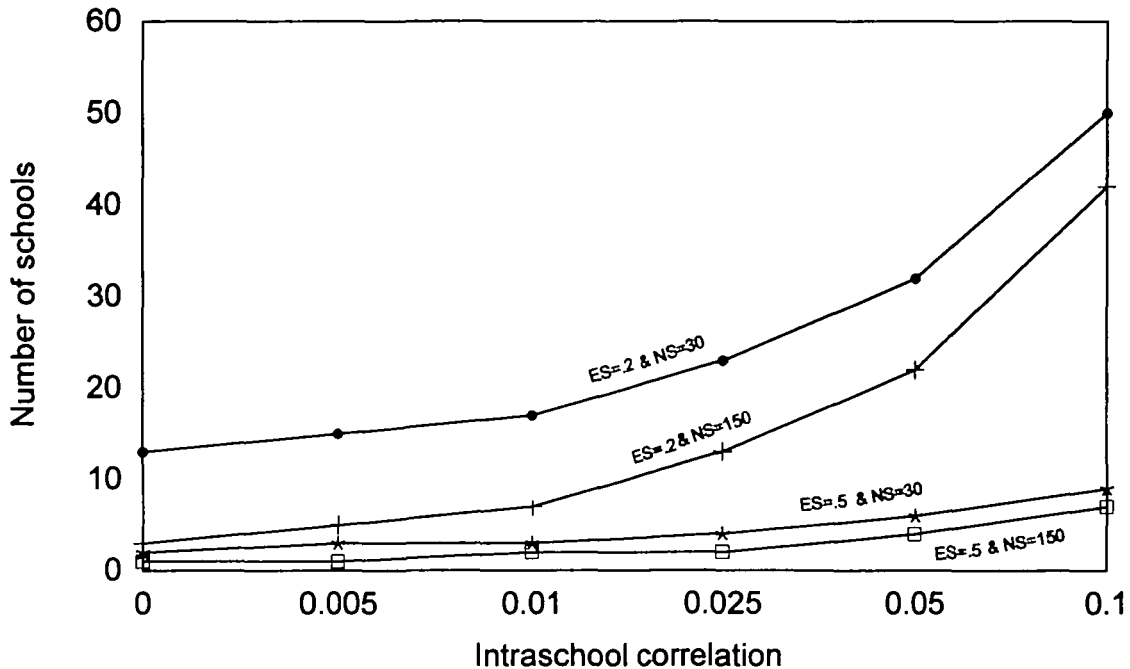


FIGURE 1. Estimated numbers of schools required to reach a power of 0.8 at different Intraschool correlations for average number of students (NS) (30 and 150) and effect size (ES) (0.2 and 0.5): Television, School, and Family Project, 1986–1988.

be incorrect. Several authors have published estimates of intraclass correlations for cigarette smoking and drug use. However, published estimates of intraclass correlations have been lacking for important outcome variables such as students' current smoking status, intentions to smoke, efforts to resist trying cigarettes, confidence in refusing to smoke cigarettes, tobacco and health knowledge, social influences/resistance skills knowledge, and smoking prevalence. In our study, the intraclass correlations for the outcome variables generally decreased over time. In this situation, the investigator may conservatively use the highest intraclass correlation in sample size estimation. The ICC estimates by sex and ethnicity were different from the estimates for the entire data set, and this paper demonstrates the implications of including sex and ethnicity in calculations of sample size.

The availability of ICC estimates for smoking outcome variables in school-based smoking prevention studies is limited. Murray and Hannan (1) published the first ICC estimates for adolescent tobacco and drug use measures. Murray et al. (6) published ICC estimates for weekly smoking prevalence and number of cigarettes smoked in the past week. In both studies, the range of ICC estimates for smoking outcome variables was 0–0.07. This range coincides with the range of ICC estimates for prevalence of smoking and current smoking in this study (0–0.074), suggesting that ICC estimates from this study for prevalence of smoking

and current smoking are generalizable. To our knowledge, no one has previously published ICC estimates of smoking prevalence or current smoking by sex and race. Since the data set we used had good representation of different sex and racial groups, we assume that our ICC estimates for the smoking outcome variables regarding sex and race are also generalizable. It also appears that no one has yet published ICC estimates for the mediating variables in a smoking prevention study. Since the sample size of our data set was relatively large and there was good representation of sex and race, we also expect the ICC estimates for the mediating variables to be generalizable.

The availability of these external estimates of the ICCs for smoking-related variables should help investigators in designing future school-based smoking prevention studies. As more ICC estimates become available, the pooling of estimates across similar studies and outcome variables will allow researchers to estimate more precisely the required sample sizes for future smoking studies. The degree to which these ICC estimates can reasonably be used in studies of other health-related variables (besides smoking) is an important issue, one that we are currently examining. Hopefully, as more is learned about the degree of clustering of many types of health-related outcomes, investigators will be able to more efficiently design studies based on specific outcomes and specific sampled populations.



## ACKNOWLEDGMENTS

Collection of data for this research was supported by grant R01-DA0348 from the National Institute of Drug Abuse to Drs. B. R. Flay, W. B. Hansen, and C. A. Johnson. The analyses reported here were completed with support from grant R01-DA06307 to Drs. B. R. Flay and J. Richardson.\*

## REFERENCES

- Murray DM, Hannan PJ. Planning for the appropriate analysis in school-based drug-use prevention studies. *J Consult Clin Psychol* 1990;58:458-68.
- Donner A. A regression approach to the analysis of data arising from cluster randomization. *Int J Epidemiol* 1985;14:322-6.
- Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978;108:100-2.
- Farquhar JW, Fortmann SP, Maccoby N, et al. The Stanford Five-City Project: design and methods. *Am J Epidemiol* 1985;122:323-34.
- Jacobs DR Jr, Luepker RV, Mittelmark MB, et al. Community-wide prevention strategies: evaluation design of the Minnesota Heart Health Program. *J Chronic Dis* 1986;39:775-88.
- Murray DM, Rooney BL, Hannan PJ, et al. Intraclass correlation among common measures of adolescent smoking: estimates, correlates, and applications in smoking prevention studies. *Am J Epidemiol* 1994;140:1038-50.
- Hannan PJ, Murray DM, Jacobs DR Jr, et al. Parameters to aid in the design and analysis of community trials: intraclass correlations from the Minnesota Heart Health Program. *Epidemiology* 1994;5:88-95.
- Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. *Stat Med* 1994;13:61-78.
- Kish L. Survey sampling. New York, NY: John Wiley and Sons, Inc, 1965.
- Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol* 1981;114:906-14.
- MacKinnon DP, Johnson CA, Pentz MA, et al. Mediating mechanisms in a school-based drug prevention program: first-year effects of the Midwestern Prevention Project. *Health Psychol* 1991;10:164-72.
- National Institutes of Health. NIH guide for grants and contracts. Vol 23, no. 11. Bethesda, MD: National Institutes of Health, 1994.
- Flay BR, Miller TQ, Hedeker D, et al. The Television, School, and Family smoking prevention and cessation project. VIII. Student outcomes and mediating variables. *Prev Med* 1995;24:29-40.
- Flay BR, Brannon BR, Johnson CA, et al. The Television, School, and Family smoking prevention and cessation project. I. Theoretical basis and program development. *Prev Med* 1988;17:585-607.
- Siddiqui O, Flay BR, Hu FB. Factors affecting attrition in a longitudinal prevention study. *Prev Med* (in press).
- Prosser R, Rasbash J, Goldstein H. ML3 software for three-level analysis: users' guide for vol 2. London, England: Institute of Education, University of London, 1991.
- Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994;50:933-44.
- Hedeker D. MIXOR: a Fortran program for mixed-effects ordinal probit and logistic regression. (Prevention Research Center technical report). Chicago, IL: University of Illinois, 1993.
- Fisher RA. Statistical methods for research workers. Edinburgh, Scotland: Oliver and Boyd, 1925.
- Donner A, Koval JJ. Design considerations in the estimation of intraclass correlation. *Ann Hum Genet* 1982;46:271-7.
- Hsieh FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med* 1988;7:1195-201.
- Snedecor GW, Cochran WG. Statistical methods. 8th ed. Ames, IA: The Iowa State University Press, 1989.
- Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: L Erlbaum Associates, 1988.