School-Level Intraclass Correlation for Physical Activity in Sixth Grade Girls

DAVID M. MURRAY[1], **JUNE STEVENS**[2], **PETER J. HANNAN**[3], **DIANE J. CATELLIER**[4], **KATHRYN H. SCHMITZ**[3], **MARSHA DOWDA**[5], **TERRY L. CONWAY**[6], **JANET C. RICE**[7], and **SONG YANG**[8]

[1]*Division of Epidemiology, School of Public Health, The Ohio State University, Columbus, OH*

[2]*Departments of Nutrition and Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC*

[3]*Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN*

[4]*Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC*

[5]*Department of Exercise Science, University of South Carolina, Columbia, SC*

[6]*Graduate School of Public Health, San Diego State University, San Diego, CA*

[7]*Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA*

[8]*Division of Epidemiology and Clinical Applications, National Heart Lung and Blood Institute, Bethesda, MD*

Abstract

Purpose—The Trial for Activity in Adolescent Girls (TAAG) is a group-randomized trial (GRT) to reduce the usual decline in moderate to vigorous physical activity (MVPA) among middle school girls. We report the school-level intraclass correlation (ICC) for MVPA from the TAAG baseline survey of sixth grade girls and describe the relationship between the schedule of data collection and the ICC.

Methods—Each of six sites recruited six schools and randomly selected 60 sixth grade girls from each school; 74.2% participated. Girls were grouped in waves defined by the date measurements began and asked to wear an Actigraph accelerometer for 6 d. Occasional missing data were replaced by imputation, and counts above 1500 per 30 s were treated as MVPA, converted into metabolic equivalents (METs), and summed over 6 a.m.-midnight to provide MET-minutes per 18-h day. Mixed-model regression was used to estimate ICC.

Results—The school-level ICC were higher when estimated from a single wave compared with three waves (e.g., 0.057 vs 0.022) and across weekdays compared with weekend days (e.g., 0.024 vs 0.012). Power in a new trial would be greater with some schedules (e.g., 88% given three waves and 6 d) than with others (e.g., 23% given one wave and Tuesday only).

Conclusions—The schedule of data collection can have a dramatic effect on the ICC for MVPA. In turn, this can have a dramatic effect on the standard error for an intervention effect and on power.

Address for correspondence: David M. Murray, Ph.D., Chair, Division of Epidemiology, School of Public Health, The Ohio State University, B222 Starling Loving Hall, 320 West 10th Avenue, Columbus, Ohio 43210; E-mail: dmurray@sph.osu.edu

Investigators will need to consider the expected magnitude of the ICC and the validity of the MVPA estimates associated with their data collection schedule in planning a new study.

Decreases in the levels of physical activity in youth have led to intense interest among investigators in the creation and testing of interventions to promote increased physical activity in children. Accelerometry offers an objective measure that can be used to evaluate these interventions. Accelerometers record counts that represent movements summed over a specified time interval. Accelerometry counts have been calibrated against oxygen consumption and can be used to predict energy expenditure (16). In addition, counts have been calibrated against known intensity levels of activity, and thus can provide estimates of time spent performing sedentary, light, moderate, and vigorous activity (21). Coupled with these attributes, the relatively low cost of accelerometry data and its application to free-living situations has led to its use in studies evaluating physical activity interventions in children (1,6,12,14,18,20).

Where the intervention manipulates the social or physical environment, involves group processes, or cannot be delivered to individuals, these trials usually employ a group-randomized trial (GRT) design. In a GRT, identifiable social groups are randomly assigned to study conditions, and observations are made on the members of those groups (8, p. 3). Unfortunately, GRT usually have less statistical power than trials in which individuals are randomized because there is extra variation among the groups beyond the variation among the members and because the degrees of freedom available to estimate that extra variation are often limited. The magnitude of the extra variation is indexed by the group-level intraclass correlation (ICC); in order to plan a trial with sufficient power, it is critical for investigators to have a good estimate of the expected ICC and to be aware of other issues in study design and analysis that can impact the power of a GRT(8, pp.6-9).

We know from previous studies that school-level ICC for a variety of variables can vary by grade, time of year, and other factors (10). In the Pathways study moderate to vigorous physical activity (MVPA) was assessed by accelerometry in third grade American Indian children with all children within a school measured in a single 24-h period; the school-level ICC for MVPA was very high at 0.2 (19). In contrast, the estimate of the school-level ICC for physical activity assessed by accelerometry was zero in a pilot study of eighth grade girls for the Trial of Activity in Adolescent Girls (TAAG) (9). In that study MVPA was measured over a 7-d period and girls within schools were measured in three different waves defined by the date on which measurements began. These large differences in the ICC estimates could be due to differences in study design, data collection or analytic methods, the ages of the children, the study populations, or other factors.

In this paper, we use accelerometry data from sixth grade girls who participated in the TAAG baseline survey to investigate the effect of selected design and analytic methods on the school-level ICC for MVPA. For both weighted and unweighted measures of MVPA we compare the impact of day of the week, number of days of data collection, and number of waves of data collection. We hypothesized that the ICC would be larger when data from girls within schools were collected 1) on weekdays compared with weekend days, 2) over a shorter (1 d) as compared with a longer (6 d) time interval, 3) in the same week for all participants in a school compared with different weeks. Each of these factors should serve to increase the homogeneity of activity and thus increase the ICC. Though we did not have additional *a priori* hypotheses,

we were also interested in looking at the ICC calculated for individual days and for the 4-d period Thursday through Sunday. Some studies might have resources only for a single day's measurements; others might have resources for several days, but not for a full week. These investigators might benefit from having ICC estimates based on a single day or on several days. After consultation with several physical activity specialists associated with TAAG, we determined that Thursday through Sunday was the period that would likely be used for measurements if a full week was not possible, both because this period included weekend days and weekdays and because it would allow staff to fit the girls with the devices on a school day (Thursday morning) and recover them on another school day (Monday morning). Those specialists could not agree on which single day would be used, so we examined all of them. We had no *a priori* expectations about what values we might obtain, but we did have an *a priori* reason for reporting the ICC for individual days and for the period Thursday through Sunday. We demonstrate the impact of these design features on statistical power and make recommendations to enhance efficiency.

METHODS

TAAG

TAAG is a multicenter GRT designed to test an intervention to reduce by half the age-related decline in MVPA in middle school girls (18). TAAG has six sites (at the Universities of Arizona, Maryland, Minnesota, and South Carolina; San Diego State University; and Tulane University), a coordinating center (at the University of North Carolina, Chapel Hill), and a project office at the National Heart Lung and Blood Institute. Each site recruited six schools according to standard eligibility criteria. Schools were randomly allocated within site and school district to either an intervention or a control condition. Baseline measures were collected from a cross-sectional sample of sixth grade girls in each school in spring 2003, and the intervention began soon thereafter. Follow-up measures were collected in spring 2005 from a cross-sectional sample of eighth grade girls in the same schools. The intervention effect will be estimated as the mean difference in MVPA between intervention and control schools in the eighth grade, controlling for sixth grade levels. Activity was measured using an Actigraph accelerometer (Actigraph, Manufacturing Technologies, Inc., http://www.mtiactigraph.com/).

The TAAG baseline measurement of MVPA

Each of the six sites obtained a complete roster of the sixth grade girls enrolled in each participating school in the winter of 2003. The coordinating center selected 60 girls at random from each school for invitation to the measurement of MVPA by accelerometry. Inclusion criteria were: 1) informed consent signed by a parent or guardian, and 2) informed assent from the girl. Exclusion criteria were: 1) medical condition for which physical activity is contraindicated, and 2) inability to complete a brief questionnaire. Girls who were ineligible or who transferred out of the participating schools before the data were collected were replaced by other girls selected at random from the school rosters.

At each school, data were collected over a period of two to five nonconsecutive weeks. Most girls were measured in the first 2 wk, and the remaining girls were measured as they could be scheduled. The goal at each school was to measure 80% of the 60 girls selected to represent their school, so TAAG staff at each site made repeat visits to the participating schools to measure as many of the selected girls as possible. As in the earlier pilot study, this measurement schedule permitted examination of week-to-week variation in physical activity due to factors such as weather, field trips, etc. We refer to the first and second weeks of data collection in each school as the first and second waves because the specific weeks involved varied from school to school; because most girls were measured in the first two waves, subsequent weeks were pooled and treated as a third wave.

Human subjects approval

The TAAG baseline data collection protocol was approved by the institutional review board of each participating site and collaborating institution. It was also approved by the appropriate school and school district officials.

Variables of interest and their measures

The primary outcome for TAAG is intensity-weighted minutes (i.e., MET-minutes) of total MVPA, measured across both school time and nonschool time and across the entire week. Other measures include unweighted minutes of MVPA as well as demographic and anthropometric variables (ethnicity, age, weight, height, and body mass index). All measurement staff completed a measurement training and certification program.

The measurement of MVPA required six measurement days and two school visits by TAAG staff for each girl. During the first school visit, staff confirmed eligibility, consent and assent, obtained baseline information including date of birth, height, weight, and ethnicity, and gave instructions. Using a standardized protocol, staff measured each girl's weight twice to the nearest 0.1 kg on an electronic scale (Seca, Model 770, Hamburg, Germany). Staff measured height twice to the nearest 0.1 cm using a portable stadiometer (Shorr Height Measuring Board, Olney, MD). Each girl was asked to wear the accelerometer for six complete days, except at night while sleeping, during any activity that might get the accelerometer wet (e.g., bathing, swimming), or during any activity for which the wearing of the accelerometer was prohibited (e.g. competitive team sport activities). Girls were instructed to wear the accelerometer on their right hip, attached to a belt provided by the study and standardized across sites; girls returned the accelerometer 6 d later at the second school visit. We avoided Fridays and Mondays as starting days because absenteeism is usually higher on those days.

Data processing and analysis

Accelerometer readings were processed using methods similar to those reported by Puyau et al. (13). Readings above 1500 counts per half minute were treated as MVPA, while readings below that threshold were ignored; we reported previously that this threshold had the optimal sensitivity and specificity for discriminating brisk walking from less vigorous activities in eighth grade girls (21). Half-minute counts were used instead of full-minute counts based on the expectation that they would be more sensitive to fluctuations in activity levels.

Occasional missing accelerometry data within a girl's 6-d record were replaced via imputation based on the expectation maximization (EM) algorithm; details are reported elsewhere (4) and we provide only a brief summary here. The monitor records the slightest motion as a nonzero count, so we judged a sustained (20 min) period of zero counts as a time when the monitor was not being worn. There was considerable between- and within-girl variation in the amount of time they wore the monitors; for example, the girls wore the devices for an average of 9.5 h·d^{-1} on weekends and 13.2 h·d^{-1} on weekdays. We judged girls to be compliant with the protocol if they wore the monitor 80% of the time available in a given block of time; blocks represented before school, during school, after school, early evening, and evening. If the girl was compliant for a block, we used the data provided; if not, we used imputation to fill in the missing data for that block, with at least one compliant day required for each girl. The result was a set of six 18-h days of data for each girl, 6 a.m. to midnight. Evaluation of our imputation procedure indicated that it provided valid results, even when data were not missing at random (4).

Counts above 1500 per half minute were converted into METs using a regression equation developed from a second pilot study for TAAG (16); the METs were summed over the 6 a.m.-

midnight day to provide MET-minutes per day of MVPA, where 1 MET·min represents the metabolic equivalent of energy expended sitting at rest for one minute.

The goal of the analysis was to estimate the variability due to sites, schools (within sites), wave of data collection (within school and site), and girls (within wave, school, and site) to allow calculation of the school-level and wave-level ICC. To do so, a general linear mixed model (7) was applied to predict the intensity-weighted MVPA score and separately to predict the unweighted MVPA score. The general linear mixed model is appropriate for correlated data with normally distributed errors. The model was fit using SAS PROC MIXED, Version 8.2 (15). Site, school, wave, and girl were modeled as random effects. Because our previous study had shown that the school-level ICC was reduced with regression adjustment for ethnicity (9), we evaluated ethnicity as a fixed effect. Confidence intervals for the school-level ICC were calculated based on the F-distribution with 30 df for schools (6 sites × (6 schools/site - 1)), 72 df for wave (6 sites × 6 schools/site × (3 waves/school - 1)), and 1495 df for girls (6 sites × 6 schools/site × 3 waves/school × (14.84 girls/school - 1)) (17, pp. 245-246).

We conducted this set of analyses separately for MET-minutes and for minutes of MVPA. We also repeated the analyses after omitting wave as a random effect to gauge the benefit of having that term in the model. We repeated these analyses separately for all days combined, all weekdays combined, both weekend days combined, Thursday through Sunday combined, and each day of the week separately to estimate ICC as they might be available to investigators who could not collect data for six consecutive days. Finally, we conducted all analyses initially for waves 1-3 combined and then for wave 1 only to gauge the benefit of spreading data collection in a school over time versus concentrating it in a single week.

We use the general notation of Murray (8, pp. 131-132) in which m members (here girls) are nested within each of s subgroups (here waves), which are nested within each of g groups (here schools), so that there are ms girls per school. We use σ_m^2, σ_s^2, and σ_g^2 to represent member (here girl), subgroup (here wave), and group (here school) components of variance, respectively. Excluding site, which plays no role in the calculation of the school and wave ICC, the three components of variance sum to the total random variation in the dependent variable, $\sigma_y^2 = \sigma_m^2 + \sigma_s^2 + \sigma_g^2$. The wave- and school-level ICC are defined as:

$$\text{wave−level ICC}: \quad \text{ICC}_s = \frac{\sigma_s^2}{\sigma_m^2 + \sigma_s^2 + \sigma_g^2}$$

$$\text{school−level ICC}: \quad \text{ICC}_g = \frac{\sigma_g^2}{\sigma_m^2 + \sigma_s^2 + \sigma_g^2}$$

[1]

In the models that ignored wave, there is no estimate of σ_s^2 and thus no wave-level ICC; the school-level ICC is then calculated as shown above but omitting σ_s^2, recognizing that the estimates for σ_m^2 and σ_g^2 will be different in models that do and do not include wave.

We applied these estimates of the components of variance at the school, wave, and girl levels to a hypothetical new study to examine the impact of the study design on detectable difference and power given a specified number of schools per condition. We also examined the impact of the study design on the number of schools per condition required for a specified detectable difference and level of power.

ICC: RESULTS AND DISCUSSION

Across the six sites, 1603 girls provided useable data for MVPA, representing 74.2% of those invited. The number of girls per site ranged from 232 to 291; 44.5% were white, 22.0% were African-American, 21.8% were Hispanic, 3.8% were Asian, 0.8% were Native American, and 7.1% were multiethnic. Most of the observations were collected in wave 1 (52.2%) and wave 2 (35.8%), with many fewer in wave 3 (12.0%).

Table 1 provides descriptive statistics for age, BMI, weight, minutes of MVPA, and MET-minutes of MVPA. Table 1 also presents descriptive statistics for MVPA and MET-minutes of MVPA by wave; the average and median values on both measures increased over the sequence of waves, though their variability did not. Table 2 presents the variance components and ICC for MET-minutes of MVPA as a function of the days and waves included in the analysis and whether wave was modeled or ignored. All values are adjusted for ethnicity; in all cases, the adjusted variance components and ICC were smaller than the unadjusted values, which are not shown. Table 3 presents the parallel information for minutes of MVPA. The 95% confidence bounds are provided to indicate the precision of the measurement of the ICC and should not be used as the basis for ignoring the ICC if the bound includes zero. (Standard errors for very small components of variance are not well estimated, and power for the test that ICC = 0 is often poor. The prudent course is to employ an analysis that reflects fully the nested design (5,8, p. 232)).

Several patterns are apparent from the results presented in Tables 2 and 3. First, the patterns are the same for MET-minutes of MVPA and minutes of MVPA, with only minor discrepancies, though the absolute values differ for the two measures.

Second, the school-level ICC were much lower when estimated from a model that included wave than when estimated from a model that ignored wave. When wave is included in the model, some of the variance otherwise attributable to school is attributed instead to wave, so it is not surprising that the school-level ICC were lower when estimated from a model that included wave. As presented later, the variance of the intervention effect is defined differently in analyses that include or exclude wave so that the benefit of the smaller school-level ICC is not as great as it might first appear.

Third, consistent with hypothesis 1, the school-level ICC were uniformly lower when calculated across weekend days rather than weekdays. For example, in Table 2, the school-level ICC when wave was ignored was 0.024 for weekdays but 0.012 for weekends. This suggests that physical activity patterns are more idiosyncratic on weekend days than on weekdays, perhaps because girls get a higher proportion of their activity in school-related group programs such as physical education classes on weekdays compared with weekends. As noted above, compliance for wearing the monitors was lower on weekends than weekdays. This difference in compliance could be responsible for some of the differences observed in the ICC, but we remind the reader that our imputation scheme generated a complete 18-h day for each girl, both for weekend days and for weekdays, so that compliance differences among the days of the week did not result in differences in the amount of data available for analysis.

Fourth, the school-level ICC were smaller when estimated from data collected on Thursday through Sunday than when collected on all 6 d, across weekdays, or across weekends. For example, in Table 2, the school-level ICC when wave was ignored was 0.004 for Thursday through Sunday, but 0.022 for all 6 d, 0.024 for weekdays, and 0.012 for weekends. This is due in part to the inclusion of two weekend days and just two weekdays, since the weekend days had lower ICC than the weekdays. But the analyses across the weekend days resulted in higher school-level ICC than those obtained for the data collected on Thursday through Sunday, so that the inclusion of two weekend days is not the only explanation for this pattern. It may

be that the factors responsible for shared variation attributable to school are different during the week and on weekends so that when those periods are combined, the net shared variation attributable to school is lower than in either period alone.

Fifth, the results with regard to hypothesis 2 were mixed. We did observe the highest ICC in the analyses of data collected on a single day; the highest school-level ICC were often associated with Wednesdays, though other weekdays sometimes had higher values. At the same time, many days had ICC that were appreciably lower than when estimated from all 6 d; for example, Sunday was associated with the lowest school-level ICC in all cases. Thus we cannot conclude that the ICC are always higher when based on a shorter time interval, and instead must conclude that the nature of the time interval plays an important role. The pattern observed among the day-specific ICC is not likely due to chance, as the value for a given day of the week represents a weighted average over many Tuesdays, or many Thursdays, etc., with at least 3 d from each field center, often measured in different months. If the observed pattern were due entirely to chance variation from day to day, we would expect these variations to average out over a series of Tuesdays over several months, reducing the variation in the day-specific ICC. Instead, we see appreciable differences among the day-specific ICC, with much lower values on some days and higher values on other days. The confidence intervals are of course wider than the CI reported for the ICC based on all days, as the day-specific ICC were based on a single day instead of 6 d. As a result, few of the day-specific values are significantly different from one another. Even so, the day-specific differences may be real, and it will be important for other investigators to look for such patterns in their data.

Sixth, consistent with hypothesis 3, the ICC based on 6 d of measurement were higher when estimated from the first wave than when estimated from all three waves combined. This comparison is made using the all days rows for all waves, wave ignored versus wave 1, wave not applicable. For weighted and unweighted MVPA, the school-level ICC were twice as large when based on one wave than when based on three waves combined.

These patterns provide additional evidence that school-level ICC might be dependent on the schedule of data collection. Previously, we have seen evidence that school-level ICC for cigarette smoking are higher when measured in the spring than when measured in the fall, at least among junior and senior high school youth (10). Here, we see evidence that school-level ICC for physical activity are higher when measured during a single week in a school compared with measurements taken over several weeks. We see evidence that school-level ICC are higher when measured on weekdays than on weekend days. And we see considerable variation among measures taken on a single day, with the lowest values on Sundays and the highest values during the week.

We would offer a strong cautionary note against choosing a data collection schedule based only on the magnitude of the ICC or the standard error for the intervention effect and their implications for power. Our results suggest, for example, that the most efficient plan for an intervention effect in MET-minutes of MVPA would be to collect data on Thursday through Sunday over three waves and to ignore wave in the analysis. But reliability was reported at only 0.42 when based on a single day of measurement, increased to 0.75 when based on 4 d, and increased to 0.82 when based on 6 d (9); subsequent analysis suggested that reliability was only 0.62 when based on Thursday through Sunday. Certainly, investigators should recognize the value of trading power for reliability and validity, and in this case, the higher reliability and validity available given 6 d of measurements would seem to justify that approach, given adequate resources, even if that required a somewhat larger study. The estimate of the treatment effect may also depend on the days used for measurement; for example, if students are more likely to follow the intervention during weekdays than weekends, it would be important to measure as many weekdays as possible.

APPLICATION: POWER, DETECTABLE DIFFERENCE, AND SAMPLE SIZE IN A NEW TRIAL

We will demonstrate the application of these results and implications for study planning through power, detectable difference, and sample size calculations for a hypothetical new trial. We assume that the goal of the new trial will be to evaluate a 2-yr intervention among fifth and sixth graders designed to increase their MVPA. Baseline data will be collected in late fourth grade or early fifth grade, and follow-up data will be collected in late sixth grade. We can compare trials that collect data over six consecutive days per girl with trials that collect data over a long weekend or on a single day. We can compare trials that collect data over several weeks in each school with trials that collect data during a single week in each school. We will assume that these trials will be single-center trials, that is, no need to represent site in the design or in the analysis. We will also assume that the trials will have a nested cross-sectional design, that is, independent cross-sectional surveys at baseline and follow-up.

Sample size calculation for any trial requires explication of the primary analysis plan for the trial. We will use the approach adopted for TAAG, where the analysis will be conducted in two stages as if there were no overlap among girls measured at baseline and follow-up. This two-stage approach avoids many of the complexities inherent in a single-stage mixed-model regression analysis and mimics an analysis of covariance performed on follow-up data with regression adjustment for baseline values on the primary endpoint (9,18). Importantly, our developmental work in TAAG showed that this analytic approach had better power in TAAG than the alternatives considered, including a traditional cohort design in which the same girls would be followed over time and analyzed using repeated-measures techniques (18); similar findings have been reported in another study that employed a nested cross-sectional design (11). Certainly, there will be overlap among the independent cross-sectional samples of sixth and eighth graders, but we will ignore it in the primary analysis, as the expected benefit from using it (variance reduction due to correlation over time) is quite small relative to the expected price for using it (diluted intervention effect from imputing data for girls lost to follow-up). We will assume ethnicity is included as a covariate in the first stage; we will compare models that include wave as a nested random effect with models that ignore wave.

In the first stage of the adjusted analysis, the girls' MET-minutes or minutes of MVPA would be regressed on school, time (baseline, follow-up), their interaction, wave, and ethnicity; study condition would not be included in that model.

$$MVPA = intercept + school + time + time \times school + ethnicity + wave + girl \qquad [2]$$

Here, all terms would be modeled as fixed effects, except for wave and girl. Of interest is the estimation of ethnicity-adjusted school means for MET-minutes of MVPA, estimated for each school at baseline and at follow-up. By performing a pooled analysis of both surveys in the first stage, we would standardize the results for the two surveys against the same reference distribution for ethnicity, here the average ethnicity distribution over time. The result of the first stage would be two adjusted mean MVPA values (baseline, follow-up) for each participating school. The version of equation 2 without wave would differ only in the omission of wave.

The second stage of analysis would be conducted on the adjusted or unadjusted means from the first stage. We would regress the follow-up school mean for MET-minutes or minutes of MVPA on condition, adjusting for the baseline school mean:

$$MVPA_{followup} = intercept + condition + MVPA_{baseline} + school \qquad [3]$$

Here, intercept, condition, and baseline MVPA would be fixed effects while school would be a random effect; in this model, there is no residual error beyond school. Given a proper

randomization and a well-executed study, this model provides an unbiased test of the intervention effect. It also provides the statistical basis for inferences to schools like those included in the study.

Assumptions

The means and standard deviations for MET-minutes and minutes of MVPA for sixth graders were taken from the results reported in Table 1. The test of the intervention effect will be two-tailed, with a Type I error rate of 5%, and based on the two-stage analysis used in TAAG. To be conservative, the sample size calculations ignore any correlation over time at the level of the girl or school. The calculations assume an attrition rate of 36% over 2 yr, which is the same assumption that was made for TAAG; power will, of course, be better for studies whose investigators expect a lower attrition rate. The calculations also assume that occasional missing accelerometry data for a single girl will be replaced via imputation using the EM algorithm (4). The calculations assume that 20% of girls will refuse the accelerometry measurements at the baseline and follow-up surveys; we will not impute data for refusals based on our earlier observation that doing so is likely to have little impact in terms of bias or power (9).

Consistent with the expectations for TAAG, the sample size calculations assume that girls enrolled in the intervention schools during the 2-yr intervention will display an intervention effect of 10% more MET-minutes or minutes of MVPA than girls in the control schools. Also consistent with the expectations for TAAG, the calculations assume that girls enrolled in the intervention schools only in the second year of the intervention period will display an intervention effect of 6% more MET-minutes or minutes of MVPA. Combined with expectations for attrition and refusals, we anticipate that the average intervention effect will be 9.2% of the mean, or 13.4 MET·min of MVPA and 2.18 min of MVPA.

Methods

The test for the intervention effect is given by the *F*-test for condition from the second stage of the analysis. Because that test involves a 1-*df* contrast, it is convenient to use the equivalent *t*-test for the power analysis, written generically as $t = \Delta/\sigma_\Delta$, where Δ is the intervention effect and σ_Δ is the standard error of the intervention effect. We calculated the approximate power using equation 9.7 from Murray (8, p. 358):

$$\text{power} = \text{prob}\left(t \le t_\beta\right), \text{ where } t_\beta = \frac{\Delta}{\sigma_\Delta} - t_{\text{critical}:\alpha/2} \quad [4]$$

The $t_{\alpha/2}$ and t_β are the critical values for the *t*-distribution reflecting the Type I and II error rates and the *df* available for the test for the intervention effect. Those *df* are calculated as c(g - 1) - 1, where there are c conditions and g schools per condition and there is one school-level covariate, the baseline school mean for MVPA.

We calculated the absolute detectable difference using equation 9.11 from Murray (8, p. 360):

$$\Delta = \left[\sigma_\Delta^2 \left(t_{\text{critical}:\alpha/2} + t_{\text{critical}:\beta} \right)^2 \right]^{1/2} \quad [5]$$

The relative detectable difference was calculated as $(\Delta/\mu) \times 100$, where μ is the mean.

In our earlier paper, we provided a formula for σ_Δ^2 based on equation 9.25 from Murray (8, p. 367). That formula was based on the results of the analysis of eighth grade data, where there was no evidence of any variation attributable to wave:

$$\sigma_\Delta^2 = \frac{2\left(\sigma_m^2 + \text{ms}\sigma_g^2\right)}{\text{msg}} \quad [6]$$

Equation 6 is appropriate for the estimates derived from the analytic model that ignores wave, as the only components of variance estimated were for girls and schools. However, if wave is included as a nested factor and the wave component of variance is greater than zero, equation 6 will underestimate the variance of the intervention effect, andthiscould lead to anundersized study. A more general form of equation 6 that would allow for a nonzero component of variance for wave is:

$$\sigma_\Delta^2 = \frac{2\left(\sigma_m^2 + ms_g^2 + m_g^2\right)}{msg} \quad [7]$$

We calculated the number of schools (g) required based on equation 9.22 from Murray (8, p. 364). Working from equation 6:

$$g = \frac{2\left(\sigma_m^2 + ms\sigma_g^2\right)\left(t_{critical:\alpha/2} + t_{critical:\beta}\right)^2}{ms\Delta^2} \quad [8]$$

Working from equation 7:

$$g = \frac{2\left(\sigma_m^2 + ms\sigma_2^2 + ms\sigma_g^2\right)\left(t_{critical:\alpha/2} + t_{critical:\beta}\right)^2}{ms\Delta^2} \quad [9]$$

The *df* upon which the values for the *t*-values depend will change as the value calculated for g changes; convergence may require two or three iterations.

Investigators may not have access to variance component estimates; instead, they may have only estimates of ICC and σ_y^2. Equations 6 and 7 can be rewritten in terms of these parameters working from the expressions provided in equation 1.

Example

Consider the adjusted components of variance for MET-minutes of MVPA in sixth grade girls measured over all days in all three waves, as reported in the first row of Table 2. Assume 30 girls are to be measured in each of three waves in each of 20 schools in each condition in the new trial. Given c = 2, g = 20, and *df* = c(g - 1) - 1 = 37, the critical *t*-values are $t_{\alpha/2} = 2.026$ and $t_\beta = 0.851$. Because we included wave in the model, equation 7 is appropriate.

$$\begin{aligned}
\sigma_\Delta^2 &= \frac{2\left(\sigma_m^2 + m\sigma_s^2 + ms\sigma_g^2\right)}{msg} \\
&= \frac{2(5728 + 30 \times 305 + 30 \times 3 \times 9.1)}{30 \times 3 \times 20} \\
&= 17.44
\end{aligned}$$

The power for a 10% relative difference in MET-minutes of MVPA, 14.5 MET·min corrected for attrition and refusals to be 13.4 MET·min, is calculated using equation 4:

$$\begin{aligned}
power &= prob\left(t \le t_\beta\right), \text{ where } t_\beta = \frac{\Delta}{\sigma_\Delta} - t_{\alpha/2} \\
&= prob(t \le 1.18), \text{ where } t_\beta = \frac{13.4}{4.18} - 2.026 = 1.18 \\
&= 0.88
\end{aligned}$$

This means that there would be 88% power to detect an intervention effect equal to 10% of the mean, given the assumptions described above.

The detectable difference in MET-minutes of MVPA is calculated using equation 5:

$$
\begin{aligned}
\Delta \quad &= \left[\sigma_\Delta^2 \big(t_{\text{critical}:\alpha/2} + t_{\text{critical}:\beta} \big)^2 \right]^{1/2} \\
&= \left[17.44(2.206 + 0.851)^2 \right]^{1/2} \\
&= 12.0
\end{aligned}
$$

So with 20 schools per condition and 80% power, the study would be able to detect an absolute difference between the two conditions of 12.0 MET·min, after correction for refusals and attrition, or 13.0 MET·min for students exposed to the full intervention, equal to a relative effect of $13.0/145.6 = 8.9\%$.

The investigators may choose to reduce the size and cost of the study by determining the number of schools required to achieve 80% power to detect a 10% relative difference, or 13.4 MET·min after correction for attrition and refusals. Using equation 9:

$$
g \quad = \frac{2\big(\sigma_m^2 + ms\sigma_s^2 + ms\sigma_g^2\big)\big(t_{\alpha/2} + t_\beta\big)^2}{ms\Delta^2}
$$

$$
\frac{2(5728 + 30 \times 305 + 30 \times 3 \times 9.1)(2.026 + 0.851)^2}{30 \times 3 \times 13.4^2}
$$

$$
= 16.1
$$

Rounding up, this initial calculation suggests that only 17 schools may be required per condition for 80% power for an absolute difference of 13.4 MET·min. But this calculation used *t*-values based on 20 schools per condition and $2(20 - 1) - 1 = 37$ *df*, and it is important to repeat the calculation with *df* based on 17 schools per condition:

$$
g \quad = \frac{2\big(\sigma_m^2 + ms\sigma_s^2 + ms\sigma_g^2\big)\big(t_{\alpha/2} + t_\beta\big)^2}{ms\Delta^2}
$$

$$
\frac{2(5728 + 30 \times 305 + 30 \times 3 \times 9.1)(2.040 + 0.853)^2}{30 \times 3 \times 12.0^2}
$$

$$
= 16.3
$$

Because this result also rounds up to 17, the iteration process has converged and we can stop. The result indicates that at least 80% power would be available for a relative difference of 10%, or 13.4 MET·min after correction for refusals and attrition, with 17 schools per condition.

SAMPLE SIZE CALCULATIONS: RESULTS AND DISCUSSION

Tables 4 and 5 summarize the results for power and detectable difference (absolute and relative) given 20 schools per condition for a new trial separately for each set of variance components reported in Tables 2 and 3, respectively. Tables 4 and 5 also summarize the results for schools per condition given 80% power for an intervention effect equal to 10% of the mean, after correction for attrition and refusals. In this new trial, schools would be randomized to conditions, independent cross-sectional samples of girls would be measured at baseline and 2 yr later, an intervention to promote physical activity would be introduced in the intervention schools for 2 yr, and assumptions would be made as noted earlier. Each trial would be based on two conditions, and 90 girls would be measured per school, regardless of the number of schools or the schedule of those measurements. Several patterns emerge from Tables 4 and 5. First, the patterns in the two tables are generally similar in relative terms, though the absolute values in some columns are quite different.

Second, the estimates for σ_Δ were often lower when estimated based on an analysis that modeled rather than ignored wave as a nested random effect. This was true in 6 of 11 analyses for MET-minutes of MVPA, and in 7 of 11 analyses for minutes of MVPA. One previous paper

has reported almost uniformly higher estimates of σ_Δ when subgroup was modeled rather than ignored (11); that paper reported findings for alcohol, tobacco, and other drug use among adolescents. The explanation for these divergent findings lies in how much of the variability due to the subgroup (here wave) is random versus systematically associated with the group (here school). When variability is largely random, ignoring the subgroup will cause the subgroup variation to drop to the residual error where it will be well controlled by the msg in the denominator of the standard error formula; when variation is systematically associated with the group (here school), ignoring the subgroup will cause the subgroup variation to rise to the group where it will not be as well controlled because the denominator is reduced to g. In these data from TAAG, the variation attributable to the wave appears to be at least in part systematically associated with school, perhaps due to the weather, school functions, and other factors that would vary by wave within school.

Third, the estimates for σ_Δ were uniformly lower when calculated across weekend days rather than weekdays. This translated into substantial differences in the number of schools required per condition. This is likely due to the fact that physical education classes and other activities affect girls in groups during the week, whereas girls are more idiosyncratic in their pursuits on weekends.

Fourth, the estimates for σ_Δ were uniformly smaller when estimated from data collected on Thursday through Sunday than when collected on all 6 d, across weekdays, or across weekends. Here, too, this translated into substantial differences in the number of schools required per condition.

Fifth, the largest estimates for σ_Δ were obtained when estimated from data for a single day, always a weekday. The lowest estimate for σ_Δ from a single day was obtained when estimated from data for Sunday.

Sixth, the estimates for σ_Δ based on 6 d of measurement were higher when estimated from the first wave than when estimated from all three waves combined. Because girls were not randomized to waves or days, it is not possible to conclude based on this evidence that spreading the survey out over several weeks caused the lower standard error. As noted earlier, it may be that the girls who were surveyed in wave 1 were somehow different than those surveyed in later waves and that those differences may account for this pattern. Even so, these results are consistent with the interpretation that spreading data collection out over time may help reduce school-level ICC and so reduce σ_Δ in a school-based GRT.

SUMMARY AND COMMENT

These results affirm the importance of developing accurate estimates of the ICC expected to apply in the primary analysis as part of the planning for any new GRT. Here we report estimates for total MVPA including school and nonschool time. The previous estimates reported for eighth grade girls were considerably smaller than for sixth grade girls, so that a much larger study would be required for a comparable intervention effect in sixth grade girls. We could not have known that without having estimates for sixth graders. As important, the ICC measured using the same protocol were appreciably different as a function of the days and waves included in the analysis and whether wave was modeled or ignored. Those differences translated into often dramatic differences in the estimates of σ_Δ, power, detectable difference, and sample size. The implications for cost and logistics are obvious.

These results also affirm the importance of publishing variance components and ICC for physical activity from other datasets, so that the population of available estimates can grow. It would be of interest to see whether investigators would replicate the patterns we have observed associated with the sample analyzed, whether wave was modeled or ignored, and which days

were included in the analysis, for example. Once a sufficient number of estimates is available, it may be possible to take advantage of those estimates in the analysis of existing studies and in planning new studies (2,3). In the meanwhile, the estimates and methods presented here, and in our earlier paper, can be used to calculate the size required for a new study.

REFERENCES

1. Beech BM, Klesges RC, Kumanyika S, et al. Child- and Parent-targeted interventions: the Memphis GEMS Pilot Study. Ethn. Dis 2003;13:S40–S53. [PubMed: 12713210]

2. Blitstein JL, Hannan PJ, Murray DM, Shadish WR. Increasing the degrees of freedom in existing group randomized trials through the use of external estimates of intraclass correlation: the DF* approach. Eval. Rev 2005;29:241–267. [PubMed: 15860765]

3. Blitstein JL, Murray DM, Hannan PJ, Shadish WR. Increasing the degrees of freedom in future group randomized trials: the df* approach. Eval. Rev 2005;29:268–286. [PubMed: 15860766]

4. Catellier DJ, Hannan PJ, Murray DM, et al. Imputation of missing data when measuring physical activity by accelerometry. Med. Sci. Sports Exerc 2005;37(11 Suppl):S555–S562. [PubMed: 16294118]

5. Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold; London: 2000. p. 10

6. Going S, Thompson J, Cano S, et al. The effects of the Pathways Obesity Prevention Program on physical activity in American Indian children. Prev. Med 2003;37:S62–S69. [PubMed: 14636810]

7. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 1977;72:320–338.

8. Murray, DM. Design and Analysis of Group-Randomized Trials. Oxford University Press; New York: 1998.

9. Murray DM, Catellier DJ, Hannan PJ, et al. School-level intraclass correlation for physical activity in adolescent girls. Med. Sci. Sports Exerc 2004;36:876–882. [PubMed: 15126724]

10. 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;140:1038–1050. [PubMed: 7985652]

11. Murray DM, Van Horn ML, Hawkins JD, Arthur MW. Analysis strategies for a community trial to reduce adolescent ATOD use: A comparison of random coefficient and ANOVA/ANCOVA models. Contemporary Clinical Trials 2006;27:188–206. [PubMed: 16324889]

12. Prochaska JJ, Sallis JF. A randomized controlled trial of single versus multiple health behavior change: promoting physical activity and nutrition among adolescents. Health Psychol 2004;23:314–318. [PubMed: 15099173]

13. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. Obes. Res 2002;10:150–157. [PubMed: 11886937]

14. Roemmich JN, Gurgol CM, Epstein LH. Open-loop feedback increases physical activity of youth. Med. Sci. Sports Exerc 2004;36:668–673. [PubMed: 15064595]

15. SAS Institute. SAS/STAT Software: Changes and Enhancements, Release 8.2. SAS Institute, Inc; Cary, NC: 2001. 2000. p. 2085-2286.

16. Schmitz KH, Treuth M, Hannan PJ, et al. Predicting energy expenditure from accelerometry counts in adolescent girls. Med. Sci. Sports Exerc 2005;37:155–161. [PubMed: 15632682]

17. Snedecor, GW.; Cochran, WM. Statistical Methods. 8th ed.. The Iowa State University Press; Ames, IA: 1989. p. 244-245.

18. Stevens J, Murray DM, Catellier DJ, et al. Design of the Trial of Activity in Adolescent Girls (TAAG). Control. Clin. Trials 2005;26:223–233.

19. Stevens J, Suchindran C, Ring K, et al. Physical activity as a predictor of body composition in American Indian children. Obes. Res 2004;12:1974–1980. [PubMed: 15687399]

20. Story M, Sherwood NE, Himes JH, et al. An after-school obesity prevention program for African-American girls: the Minnesota GEMS pilot study. Ethn. Dis 2003;13:S54–S64. [PubMed: 12713211]

21. Treuth MS, Almeida J, Catellier DJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. Med. Sci. Sports Exerc 2004;36:1259–1266. [PubMed: 15235335]

TABLE 1

Descriptive statistics for age, BMI, weight, minutes of MVPA, and MET-minutes of MVPA.

| | N | Mean | SD | Median | Site Mean | | |
					Minimum	Maximum
Age (yr)	1603	12.0	0.50	11.9	11.7	12.3
BMI (kg·m⁻²)	1603	20.9	4.95	19.7	19.9	21.8
Weight (kg)	1603	49.1	14.1	46.5	46.6	52.1
Minutes of MVPA	1603	23.7	11.7	21.3	20.9	27.7
Wave 1	837	23.0	11.3	20.7		
Wave 2	573	24.1	12.4	21.4		
Wave 3	193	25.1	11.0	22.9		
MET-minutes of MVPA	1603	145.6	79.8	127.6	130.6	173.4
Wave 1	837	140.9	75.9	123.8		
Wave 2	573	149.8	86.8	130.1		
Wave 3	193	153.8	73.8	140.3		

SD, standard deviation; BMI, body mass index; MVPA, moderate to vigorous physical activity; MET, metabolic equivalent.

Site mean minima and maxima were identified from the six site means available for each measure.

TABLE 2

Variance component estimates and intraclass correlations for MET-minutes of MVPA as a function of the days and waves included in the analysis and whether wave was modeled or ignored. All values are adjusted for ethnicity.

Data	Wave	Days	N	Component of Variance				School			Wave		
				School	Wave	Girl	Total	ICC	LCB	UCB	ICC	LCB	UCB
All waves	Modeled	All days	1603	9.1	305	5728	6042	0.002	-0.014	0.034	0.050	0.021	0.096
		Weekdays	1603	70.3	290	7466	7826	0.009	-0.007	0.043	0.037	0.010	0.078
		Weekends	1603	0.0	393	9557	9951	0.000	-0.013	0.028	0.040	0.012	0.081
		Thursday through Sunday	1232	0.0	295	6322	6618	0.000	-0.017	0.034	0.045	0.010	0.096
		Monday	1401	0.0	768	10,365	11,133	0.000	-0.019	0.038	0.069	0.032	0.123
		Tuesday	1024	455.5	0	28,564	29,020	0.016	-0.003	0.054	0.000	-0.029	0.045
		Wednesday	1151	291.8	292	10,606	11,189	0.026	0.003	0.075	0.026	-0.005	0.075
		Thursday	1324	75.9	647	9919	10,641	0.007	-0.013	0.050	0.061	0.025	0.115
		Friday	1512	88.4	331	12,515	12,935	0.007	-0.008	0.037	0.026	0.000	0.065
		Saturday	1603	64.3	573	16,550	17,188	0.004	-0.010	0.033	0.033	0.007	0.073
		Sunday	1603	0.0	225	10,554	10,779	0.000	-0.011	0.023	0.021	-0.002	0.057
	Ignored	All days	1603	136		5897	6033	0.022	0.006	0.056			
		Weekdays	1603	191		7627	7818	0.024	0.008	0.059			
		Weekends	1603	123		9805	9927	0.012	0.000	0.039			
		Thursday through Sunday	1232	29		6545	6574	0.004	-0.008	0.030			
		Monday	1401	124		10,938	11,062	0.011	-0.002	0.039			
		Tuesday	1024	455		28,564	29,019	0.016	-0.003	0.054			
		Wednesday	1151	480		10,724	11,204	0.043	0.017	0.095			
		Thursday	1324	383		10,231	10,614	0.036	0.014	0.082			
		Friday	1512	215		12,702	12,917	0.017	0.002	0.047			
		Saturday	1603	259		16,906	17,165	0.015	0.002	0.043			
		Sunday	1603	30		10,730	10,760	0.003	-0.006	0.022			
Wave 1	Not applicable	All days	837	300		4996	5296	0.057	0.022	0.125			
		Weekdays	837	384		5793	6177	0.062	0.025	0.133			
		Weekends	837	214		8357	8571	0.025	0.000	0.074			
		Thursday through Sunday	685	49		5360	5409	0.009	-0.014	0.055			
		Monday	714	513		7738	8251	0.062	0.022	0.138			
		Tuesday	497	966		12,943	13,909	0.069	0.019	0.163			
		Wednesday	615	886		8604	9490	0.093	0.041	0.189			
		Thursday	727	481		9224	9704	0.050	0.014	0.118			
		Friday	795	476		11,946	12,422	0.038	0.008	0.098			
		Saturday	837	615		14,869	15,484	0.040	0.010	0.098			
		Sunday	837	41		8782	8823	0.005	-0.013	0.041			

MET, metabolic equivalent; MVPA, moderate to vigorous physical activity; ICC, intraclass correlation; LCB, lower 95% confidence bound; UCB, upper 95% confidence bound.

TABLE 3

Variance component estimates and intraclass correlations for minutes of MVPA as a function of the days and waves included in the analysis and whether wave was modeled or ignored. All values are adjusted for ethnicity.

Data	Wave	Days	N	Component of Variance				School			Wave		
				School	Wave	Girl	Total	ICC	LCB	UCB	ICC	LCB	UCB
All waves	Modeled	All days	1603	0.76	6.21	122	129	0.006	-0.011	0.041	0.048	0.019	0.093
		Weekdays	1603	1.97	7.02	150	159	0.012	-0.006	0.051	0.044	0.016	0.088
		Weekends	1603	0.44	6.30	184	191	0.002	-0.011	0.030	0.033	0.007	0.073
		Thursday through Sunday	1232	0.00	5.21	128	133	0.000	-0.016	0.033	0.039	0.006	0.089
		Monday	1401	0.00	17.57	193	210	0.000	-0.021	0.043	0.084	0.044	0.142
		Tuesday	1024	9.26	2.66	303	315	0.029	0.005	0.079	0.008	-0.023	0.056
		Wednesday	1151	6.71	6.89	222	236	0.028	0.004	0.080	0.029	-0.003	0.079
		Thursday	1324	1.72	14.83	228	244	0.007	-0.013	0.050	0.061	0.025	0.115
		Friday	1512	2.10	9.67	253	265	0.008	-0.008	0.042	0.037	0.009	0.079
		Saturday	1603	2.19	8.30	309	320	0.007	-0.007	0.036	0.026	0.002	0.064
		Sunday	1603	0.00	3.64	193	197	0.000	-0.011	0.022	0.018	-0.004	0.054
	Ignored	All days	1603	3.51		125	128	0.027	0.010	0.064			
		Weekdays	1603	5.00		154	159	0.032	0.012	0.071			
		Weekends	1603	2.62		188	190	0.014	0.001	0.041			
		Thursday through Sunday	1232	0.62		131	132	0.005	-0.008	0.031			
		Monday	1401	4.13		205	209	0.020	0.003	0.054			
		Tuesday	1024	11.31		304	315	0.036	0.010	0.087			
		Wednesday	1151	11.06		225	236	0.047	0.019	0.102			
		Thursday	1324	8.74		235	244	0.036	0.013	0.081			
		Friday	1512	5.89		258	264	0.022	0.006	0.056			
		Saturday	1603	5.01		315	320	0.016	0.002	0.044			
		Sunday	1603	0.38		196	197	0.002	-0.007	0.021			
Wave 1	Not applicable	All days	837	6.02		109	115	0.052	0.019	0.118			
		Weekdays	837	8.49		126	135	0.063	0.026	0.135			
		Weekends	837	3.89		176	180	0.022	-0.002	0.069			
		Thursday through Sunday	685	0.36		119	119	0.003	-0.018	0.045			
		Monday	714	12.41		143	155	0.080	0.035	0.164			
		Tuesday	497	24.42		240	265	0.092	0.035	0.195			
		Wednesday	615	19.13		187	206	0.093	0.041	0.189			
		Thursday	727	9.81		215	225	0.044	0.010	0.109			
		Friday	795	12.75		245	258	0.049	0.016	0.115			
		Saturday	837	11.91		292	304	0.039	0.010	0.097			
		Sunday	837	0.00		186	186	0.000	-0.016	0.033			

MVPA, moderate to vigorous physical activity; ICC, intraclass correlation; LCB, lower 95% confidence bound; UCB, upper 95% confidence bound.

TABLE 4

Standard errors, power, detectable differences (absolute and relative), and schools per condition for each set of variance component estimates reported in Table 3 for MET-minutes of MVPA.

Data	Wave	Days	Given 20 Schools per Condition				Schools per Condition for 80% Power
			Standard Error	Power	Detectable Difference	Relative Detectable Difference	
All waves	Modeled	All days	4.18	0.88	12.0	0.08	17
		Weekdays	5.00	0.74	14.4	0.10	23
		Weekends	4.87	0.76	14.0	0.10	22
		Thursday through Sunday	4.11	0.89	11.8	0.08	16
		Monday	6.09	0.57	17.5	0.12	34
		Tuesday	8.79	0.31	25.3	0.17	69
		Wednesday	7.12	0.44	20.5	0.14	46
		Thursday	6.34	0.54	18.2	0.13	37
		Friday	5.81	0.61	16.7	0.12	31
		Saturday	6.63	0.50	19.1	0.13	40
		Sunday	4.39	0.85	12.6	0.09	18
	Ignored	All days	4.48	0.83	12.9	0.09	19
		Weekdays	5.25	0.70	15.1	0.10	26
		Weekends	4.82	0.77	13.8	0.10	22
		Thursday through Sunday	3.19	0.98	9.2	0.06	10
		Monday	4.95	0.75	14.2	0.10	23
		Tuesday	8.79	0.31	25.3	0.17	69
		Wednesday	7.74	0.39	22.3	0.15	54
		Thursday	7.05	0.45	20.3	0.14	45
		Friday	5.97	0.59	17.2	0.12	33
		Saturday	6.68	0.49	19.2	0.13	40
		Sunday	3.86	0.92	11.1	0.08	15
Wave 1	Not applicable	All days	5.96	0.59	17.1	0.12	32
		Weekdays	6.70	0.49	19.3	0.13	41
		Weekends	5.54	0.65	15.9	0.11	28
		Thursday through Sunday	3.29	0.98	9.5	0.07	11
		Monday	7.74	0.39	22.3	0.15	54
		Tuesday	10.54	0.23	30.3	0.21	98
		Wednesday	9.91	0.25	28.5	0.20	87
		Thursday	7.64	0.39	22.0	0.15	52
		Friday	7.81	0.38	22.4	0.15	55
		Saturday	8.83	0.31	25.4	0.18	70
		Sunday	3.73	0.94	10.7	0.07	14

MET, metabolic equivalent; MVPA, moderate to vigorous physical activity.

TABLE 5

Standard errors, power, detectable differences (absolute and relative), and schools per condition for each set of variance component estimates reported in Table 4 for minutes of MVPA.

Data	Wave	Days	Standard Error	Given 20 Schools per Condition			Schools per Condition for 80% Power
				Power	Detectable Difference	Relative Detectable Difference	
All waves	Modeled	All days	0.65	0.91	1.9	0.01	15
		Weekdays	0.77	0.78	2.2	0.02	21
		Weekends	0.68	0.88	1.9	0.01	17
		Thursday through Sunday	0.56	0.96	1.6	0.01	12
		Monday	0.89	0.66	2.6	0.02	28
		Tuesday	1.16	0.44	3.3	0.02	46
		Wednesday	1.07	0.50	3.1	0.02	39
		Thursday	0.96	0.60	2.8	0.02	32
		Friday	0.90	0.65	2.6	0.02	28
		Saturday	0.92	0.64	2.6	0.02	29
		Sunday	0.58	0.95	1.7	0.01	13
	Ignored	All days	0.70	0.86	2.0	0.01	18
		Weekdays	0.82	0.74	2.4	0.02	24
		Weekends	0.69	0.87	2.0	0.01	17
		Thursday through Sunday	0.46	1.00	1.3	0.01	8
		Monday	0.80	0.76	2.3	0.02	23
		Tuesday	1.21	0.41	3.5	0.02	50
		Wednesday	1.16	0.44	3.3	0.02	46
		Thursday	1.07	0.51	3.1	0.02	39
		Friday	0.94	0.62	2.7	0.02	30
		Saturday	0.92	0.63	2.7	0.02	30
		Sunday	0.51	0.99	1.5	0.01	10
Wave 1	Not applicable	All days	0.85	0.70	2.4	0.02	25
		Weekdays	0.99	0.57	2.9	0.02	34
		Weekends	0.76	0.79	2.2	0.02	20
		Thursday through Sunday	0.41	1.00	1.2	0.01	7
		Monday	1.18	0.43	3.4	0.02	48
		Tuesday	1.65	0.24	4.7	0.03	91
		Wednesday	1.46	0.30	4.2	0.03	71
		Thursday	1.10	0.48	3.2	0.03	42
		Friday	1.24	0.39	3.6	0.02	53
		Saturday	1.23	0.40	3.5	0.02	51
		Sunday	0.45	1.00	1.3	0.01	8

MET, metabolic equivalent; MVPA, moderate to vigorous physical activity.