# Appendix G
# Common Assumptions for Statistical Models of Parental Reports of Children's Fruit and Vegetable Consumption in a Clustered, Quasi-Experimental Design

# Common Assumptions for Statistical Models of Parental Reports of Children's Fruit and Vegetable Consumption in a Clustered, Quasi-Experimental Design

Sample size estimation procedures are used to quantify a researcher's level of confidence regarding their ability to accurately reject the null hypothesis when empirical differences are statistically significant. To achieve this end, a number of assumptions may be necessary. When information from previous studies or pilot studies similar to the study being planned is available, the potential validity of sample size estimation is improved. Without this information, researchers must rely on their best judgment and leverage their experience to identify reasonable values to justify sample sizes.

Our main outcome measure and the focus of sample size estimation is the change in consumption of servings of fruits and vegetables by children participating in the program intervention (Eagle Adventure, All 4 Kids, or EWPHCCS) as reported by their parents or caregivers. We begin with mean and standard deviation estimates from a trial in Chicago in which parents reported their children's fruit and vegetable consumption. The study included six lower socioeconomic status communities and collected data from 516 parents on their young children's dietary intake. In this study population, mean fruit and vegetable consumption was 3.83 servings per day, with a standard deviation of 2.04 servings (Evans et al., 2007).

Perhaps the most crucial component of sample size estimation is determination of program impact, often referred to as the effect size or the minimum detectable effect. This number describes the anticipated change in observed outcomes among participants as a result of participating in the intervention. For our purposes, we aim to identify a change of 0.30 standard deviation units or greater. Based on the findings from the Chicago study, this suggests a realized change of 0.61 servings of fruit and vegetables from baseline values. This expectation is consistent with findings reported in a recent meta-analysis by Knai and colleagues (2006) who found that across a range of dietary interventions, children's fruit and vegetable consumption increased by 0.3 to 0.99 servings per day.

Our sample size estimation procedures follow the convention of estimating sample size allowing for a Type II error rate of 0.20 (yielding 80 percent statistical power) and a Type I error rate of 0.05, with a two-tailed test.

## Assumptions of the Model

One of the key assumptions of sample size determination, in a study where schools/centers are randomly allocated and students (or parents) within schools/centers are the primary unit of observation, is an accurate estimate of the intraclass correlation coefficient (ICC). The ICC is the proportion of variation attributable to the cluster (e.g., school or center) over and above the variation attributable to the individual. As such, the

ICC has a multiplier effect that can substantially increase variation. We are unaware of any study that has published ICC estimates on parent's reports of children's dietary intake. However, a study of middle school youth reported an ICC of 0.034 for self-reported fruit and vegetable consumption (Murray, Phillips et al. 2001). Using this study as a starting point, and recognizing the differences between the participants in Murray et al. and our study, we will employ an ICC 0.05 for our calculations.

Another assumption is the form of the statistical model. Our calculations are appropriate for a mixed model regression model that includes baseline and follow-up measures of the outcome of interest (i.e., pre-/post-test model) and allows for the inclusion of covariates associated with the outcome variable, but independent of the intervention. This model allows for two sources of reduction to the variance of the outcome. First, the use of a pre-test/post-test model implies that baseline differences and potential confounding influences that may be associated the selection will be minimized. Second, the inclusion of covariates associated with the outcome of interest, but independent of the intervention, can further reduce unwanted variation in the outcome and improve statistical power. The decision of which (if any) variables to include in the model will be determined through examination of the baseline data. Demographic variables such as age, sex, and race/ethnicity are typically included.

## Are These Assumptions Realistic for the Eagle Adventure, All 4 Kids, and EWPHCCS Programs?

Several factors must be considered to assess whether the assumptions described above are realistic for the Eagle Adventure, All 4 Kids, and EWPHCCS Programs. First, consider the assumed ICC. The estimated ICC of 0.05 is conservative relative to the one identified in the literature. Additionally, factors associated with larger ICC values are not present in the current study. The ICC is a function of several factors including the size and inherent cohesiveness of the group. That is to say, smaller groups with greater interaction tend to be more similar and have larger ICCs. Further, the dissimilarity among the groups allocated to condition tends to make the member of a given group more alike to each other relative to members of another group; this also tends to increase the magnitude of the ICC. Neither of these issues is of great concern for the current study. For the Eagle Adventure and All 4 Kids Programs, the schools/centers are likely to be quite similar, are located within a common setting, and there is little reason to anticipate high levels of parental interaction. For the EWPHCCS Program, there is a little concern that centers in the New York City area could be different from the centers in the remaining areas of New York State and this might increase the ICC. However, the level of the ICC could be twice as large as anticipated and the statistical model would still have better than 95 percent probability of identifying program impacts of 0.30 standard deviation units or greater as statistically significant, ceteris paribus.

Next consider the adjustments that will attenuate the standard error of the intervention effect. The expectation for overtime correlation is reasonable. While we are not aware of any reported estimate of the long-term reliability of parental reports of their children's dietary intake, Resnicow and colleagues have reported over-time correlation greater than 0.30 for both boys and girls over 1- and 2-year periods (Resnicow, Smith et al. 1998). For

the current study, the pre-to-post-test period is relatively short, so parents' reports of their children's dietary behavior are quite likely to be consistent over time. The magnitude of adjustment for other covariates is similarly reasonable. Consider, for example, that in Murray et al. (2001), the inclusion of demographic covariates reduced the ICC for consumption of fruits and vegetables from 0.34 to 0.11 -- a reduction close to 0.66 in the component of variation associated with the standard error of the intervention effect.

## Assessing the Likelihood of Identifying Statistically Significant Effects

Our aim is to present a model that provides an 80 percent probability that differences between the intervention and control groups that truly exist will be identified. Based on the characteristics of the study outlined above, our model will be sufficiently powered if the following assumptions hold:

The ICC is 0.05 or smaller.

The magnitude of adjustment for the correlation between pre- and post-test measures on the same participants is approximately 75 percent at both the individual and group levels.

The magnitude of adjustment for covariates is approximately 30 percent at both the individual and group levels.

## References

Evans, W.D., Necheles, J., Longjohn, M., & Christoffel, K.K. (2007). The 5-4-3-2-1 Go! intervention: social marketing strategies for nutrition. *Journal of Nutrition Education and Behavior, 39*(2, supplement 1), s55–s59.

Knai, C., Pomerleau, J., Lock, K., & McKee, M. (2006). Getting children to eat more fruit and vegetables: A systematic review. *Preventive Medicine, 42*(2), 85–95.

Murray, D.M., Phillips, G.A., Birnbaum, A.S., & Lytle, L.A. (2001). Intraclass correlation for measures from a middle school nutrition intervention study: Estimates, correlates, and applications. *Health, Education, and Behavior, 28*(6), 666–679.

Resnicow, K.E.N., Smith, M., Baranowski, T.O.M., Baranowski, J., Vaughan, R., & Davis, M. (1998). Two-year tracking of children's fruit and vegetable intake. *Journal of the American Dietetic Association, 98*(7), 785–789.