Published in final edited form as:

Contemp Clin Trials. 2011 March ; 32(2): 225–232. doi:10.1016/j.cct.2010.11.001.

Comparison of intraclass correlation coefficient estimates and standard errors between using cross-sectional and repeated measurement data: The Safety Check Cluster Randomized Trial

Edward H. Ip, Ph.D.^{1,*}, Richard Wasserman, M.D., MPH², and Shari Barkin, M.D.³

¹Department of Biostatistical Sciences and Department of Social Sciences and Health Policy, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC ²Pediatric Research in Office Settings (PROS), Department of Research; Center for Child Health Research, American Academy of Pediatrics, Elk Grove Village, IL; Department of Pediatrics, University of Vermont, Burlington, VT. Richard.Wasserman@uvm.edu ³Department of Pediatrics, Division of General Pediatrics, Monroe Carell, Jr. Children's Hospital at Vanderbilt University, Nashville, TN. shari.barkin@Vanderbilt.Edu

1. Introduction

Cluster randomized trials have become increasingly common in studies that aim to evaluate non-pharmaceutical, behavioral, and quality-improvement intervention [1,2]. Examples of such interventions include community-improvement projects [3], lifestyle interventions such as cessation of smoking and increasing physical activity [4], education intervention trials [5], and office-based child-care intervention programs [6]. The cluster design is often used to minimize contamination between intervention and control groups.

Under a cluster design, individuals within a cluster may be more alike than individuals across clusters. A universally used parameter for measuring the extent to which individuals within a cluster are similar to one another is the intraclass correlation (ICC). Suppose the ICC for patients within the same clinic was 0.04, which is not an uncommon level of correlation [4], and that there were 100 patients in a clinic. The sample size for a study based on the clinic as a unit of randomization requires five times as many patients as a study using the patient as a unit of randomization. This demonstrates how varying ICC estimates can dramatically change the sample size that would be required to test a hypothesis, and thus its estimation has important implications for study design, study cost, and implementation [2,7].

Consequently, to conduct a power analysis for a cluster randomized trial, reliable estimates of the magnitude of ICC are often required. At least two challenges exist for accomplishing power analysis. First, existing ICC estimates usually come from small pilot studies that use cluster sampling or other intervention studies of a similar kind. Unfortunately, ICC values from pilot studies can be rather unstable, and published ICC values from large studies are

© 2010 Elsevier Inc. All rights reserved.

*Corresponding author: Address: Department of Biostatistical Sciences, Medical Center Blvd., WC23, Winston-Salem, NC27012, USA, Telephone number: (336)716-9833, Fax number: (336)716-6427, eip@wfubmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

rarely available. Murray et al. [8] reviewed a large number of cluster randomized trials in terms of study design and data analysis and concluded that the lack of published estimated ICCs may lead to inaccurate power calculations and compromise the ability to test the hypothesis articulated. This is not a new problem [5,9,10]. Second, even for published values of ICC, their standard errors are often too large for practical use. For example, an ICC value of 0.02 that has a standard error of 0.01 implies that the 95 percent confidence bounds are 0.00 and 0.04. For the above clinical example, the sample sizes corresponding to these bounds would range from one third to 1.7 times that of the estimate of 0.02.

This paper has two objectives. First, we report ICC numbers from a national study in the United States—the Randomized Controlled Trial to Prevent Child Violence (Safety Check, NIH-R01HD042260), a multi-level, cluster randomized trial that used the clinical office to deliver the intervention [6]. In addition to reporting ICC values, we show how data collected from repeated measures in cluster randomized trials can be pooled together to improve the precision of ICC estimates. Because studies that have repeated measures over participants are becoming more common, the method of using repeated measurements for estimating ICC has important implications for future design in randomized trials and the analysis of clustered data.

2. Data and methods

2.1 Data

Data from the Safety Check study [6] were derived from families seeing providers in the Pediatric Research in Office Settings (PROS) Network, the practice-based research network of the American Academy of Pediatrics (AAP). All 677 practices belonging to PROS were invited to participate. In this sample, 137 pediatric practices across the country that were randomized were included. The unit of randomization was the pediatric office, and each unit was assigned to an office-based violence-prevention intervention or a control group. In the intervention group, practitioners were trained to (1) review the parent pre-visit survey including patient-family behavior, parental concern about media use, discipline strategies, and children's exposure to firearms; (2) counsel using brief principles of motivational interviewing; (3) identify and provide local agency resources for anger and behavior management when indicated; and (4) instruct patient-families to use tangible tools (minute timers to monitor media time/time outs; firearm cable locks to store firearms more safely where children live or play). The control group received usual care enhanced by a literacy promotion attention placebo handout. Randomization was performed at the practice level with computerized random numbers by the study biostatistician who was blinded to the identity of the practices. After allocation, practices were alerted to their group assignment via a letter included with the training materials. Practices were blinded as to the study hypotheses. Each practice had 1–10 providers, with a harmonic mean of 2 providers per practice. Because most practices had either one or two providers, the design at the practice level was not balanced. Each provider recruited 30 consecutive families that presented for a well-child visit for children ages 2–11 years. As a result of the cluster design, patients were clustered within provider, and providers were clustered within practice. Parents were surveyed at baseline, 1 month, and 6 months. The main outcomes were change over time in self-reported media use < 120 minutes per day, use of timeouts for a disciplinary purpose, and the use of firearm cable locks. The study was approved by the Wake Forest University School of Medicine IRB, and the work was carried out in accordance with The Code of Ethics of the World Medical Association for experiments involving humans.

A subset of the entire sample that included all families recruited between 2002 and 2006 was used to calculate the ICC. There were 3,294 families in the selected sample. While the protocol required each provider to recruit 30 families, there were some providers who stalled

and recruited a varying number of families. To prevent including providers who might have recruited a non-representative sample of families, we only included providers who recruited more than 25 families. Follow-up surveys were conducted on the phone by a professional survey center 1 and 6 months after the baseline survey, and we only included families who responded to the follow-up surveys in the final sample. The total number of families included in the final sample was 2,649 (80.4% of the baseline sample), the number of providers in the final sample was 90, and the number of practices was 68.

2.2 Outcome measures

Parents/legal guardians completed surveys, gathering such data as: demographics (e.g., age of child, number of children in the home, race/ethnicity of the child, parental home structure, maternal education), reported parental behaviors (e.g., media use, and discipline practices), and parental attitudes toward these same topics. We first examined both the content of the survey questions and the distributions of the responses. We then selected a total of 10 questions from parental behaviors and attitudes for reporting ICCs. These questions all showed bell-shaped distributions in response patterns, and they also reflected the general scope of the survey questions.

Media-related behavior questions included: "When this child is at home, how many hours per day do they watch TV/videos or play computer games/gameboy?" This was broken down by hours on an average weekday and average weekend day. For the purposes of our analysis, we collapsed these data to reflect average media use time per day (including television, video, computer games, and electronic handheld devices). Our survey presented three survey items assessing parental strategies for their child's media use. These questions asked how often, in the past month, the parent restricted use, explained content, or allowed unlimited media use. The response scale for each question was from "never" to "always" on a 4-point scale. Three media-related items assessed parental awareness about the potential negative outcomes of watching violent media: "watching violent TV programs makes children more afraid," "children's behavior is not influenced by what they see on TV," and "children who watch violent TV think real-life violence is normal behavior." These responses were rated on a 5-point scale from "strongly disagree" to "strongly agree." We developed an awareness scale (range: 3 to 15; Cronbach's alpha = 0.6) that was derived from the above three items. A higher value on the awareness scale indicated better understanding of the effects of exposure to violent media.

Questions about current discipline practices included: In the past month, "How often did you use time-outs or cool-down periods?" "How often have you yelled at this child?" "How often did you take away privileges (something this child enjoys)?" and "How often did you spank this child?" Response categories varied on a 4-point scale from "never" to "always."

2.3 Baseline-only ICC

The intraclass correlation is mathematically defined as the ratio of the between-cluster variance to the total variance for an individual family. When the sample is relatively homogeneous and there is little clustering effect, the between-cluster variance tends to be small, leading to a small ICC. At the practice level, which is the unit of randomization, ICC is defined as:

$$\rho_{pract} = \sigma_{pract}^2 / (\sigma_{pract}^2 + \sigma_{prov}^2 + \sigma_{family}^2), \quad (1)$$

where σ_{pract}^2 is the component of variance between practices, σ_{prov}^2 is the component of variance between providers within the same practice, and σ_{family}^2 is the variance between families within the same provider. Equation 1 represents the ICC for individuals within practices but belonging to different providers and different families. The standard error of the ICC estimate, which indicates accuracy of the estimate $\hat{\rho}_{pract}$ of the quantity in Eq. 1, is given by the following equation, originally formulated by Fisher [11] and discussed in [12]:

$$SE(\hat{\rho}_{pract}) = \sqrt{\frac{2(1 - \hat{\rho}_{pract})^2 [1 + (m - 1)\hat{\rho}_{pract}]^2}{m(m - 1)k}}, \quad (2)$$

where m is the number of providers per practice, and k is the number of families per provider. When m and k are not constant, we use the harmonic mean values for the equation. Because only baseline values are used here, this approach to producing ICC estimates will be called baseline only.

a. Repeated-measurements ICC—When repeated measurements of an outcome variable are available, the variance between families within the same provider can further be decomposed into two separate components—the variance component of families within provider, and the variance component from repeated measurements over the same family. With repeated measurements, ICC is defined as:

$$\rho_{pract}^* = \sigma_{pract}^2 / (\sigma_{pract}^2 + \sigma_{prov}^2 + \sigma_{family}^2 + \sigma_{rep}^2), \quad (3)$$

with standard error associated with the estimate given by:

$$SE(\hat{\rho}_{pract}^*) = \sqrt{\frac{2(1 - \hat{\rho}_{pract}^*)^2 [1 + (m - 1)\hat{\rho}_{pract}^*]^2}{m(m - 1)kr}}, \quad (4)$$

where r is the number of occasions for which a family is measured. In Equation 3, σ_{rep}^2 denotes the component of variance across repeated measurements. In the language of multi-level analysis, Eq. 3 contains four levels of data—practice, provider, family, and repeated measurements over family.

The ICCs for various levels of clustering are defined similarly. For example, with a single occasion of measurement for each family, the ICC at the provider level is given by:

$$\rho_{prov} = (\sigma_{pract}^2 + \sigma_{prov}^2) / (\sigma_{pract}^2 + \sigma_{prov}^2 + \sigma_{family}^2). \quad (5)$$

b. Statistical evaluation—The evaluations of variance components for both baseline-only and repeated-measurement ICCs were conducted using estimation methods for multi-level linear models, and they were implemented through the program PROC MIXED in SAS (Statistical Analysis Systems Institute, Inc.). The program codes for both the baseline-only and repeated-measurement approaches are included in Appendix A for illustrating the multi-level linear model. In order to accommodate a possible difference in outcome measures over time, the estimation of the variance component for repeated measurement needs to be adjusted. The adjustment was made through adding time as a fixed effect in the procedure

PROC MIXED. Accordingly, the adjustment in effect “de-means” observations across time points to derive proper variance components. The ICCs can therefore be viewed as measures of cluster strength defined within a framework of a hierarchical model with covariates [17]. Although computationally PROC MIXED is not as efficient as PROC NESTED [18]—another SAS multi-level sampling procedure for estimating variance components—it has the advantage of being able to handle fixed effects and unbalanced designs [19]. PROC MIXED also subsumes the variance component estimation program PROC VARCOMP, which only estimates simple random effects. We used PROC MIXED for both three-level (practice, provider, and family) and four-level (practice, provider, family, and repeated measurements) variance-component analysis such as the one entailed by Eq. 1.

c. Bootstrap method—The commonly used confidence interval derived from using Eq. 4 is based on several statistical assumptions: that the sample is sufficiently large, that ICC is not close to one, and that the cluster randomized trial has a balanced design. In order to validate Eq. 4 and its counterpart Eq. 3 against these assumptions, we used Monte Carlo simulation to report bias-corrected bootstrap confidence intervals [13] in reporting ICCs. The bias-corrected bootstrap method first resamples multiple times from the data set, with replacement, to create a bootstrap sample. Then each bootstrap sample was analyzed exactly the same way as one would analyze the original data set. The same analysis was applied to 200 bootstrap samples to generate an empirical distribution of the estimates. The difference between the 5th and 50th percentiles of the bootstrap estimates, and between the 51st and 95th percentiles, were respectively used to estimate the left and right 95-percent confidence intervals. Because of the unbalanced nature of the data, a bootstrap sampling scheme that sampled according to the nested hierarchy (e.g., sample first from practice, then from provider within practice) will produce samples of different sizes. Therefore, it was decided that a fixed sample size of $N = 2,649$ would be directly drawn for each of the 200 bootstrap samples. In simulation experiments for studying ICC [14], it was reported that the two methods did produce similar results when ICC values were not too high, which is the case in the present study. Besides providing nonparametric confidence intervals, the bootstrap method also served as a descriptive tool for assessing the stability and replicability of the results across repeated samplings [15]. Prior work on applications of bootstrap methods to ICC analysis suggest that the nonparametric method is reliable provided the sample size is not too small [16].

3. Results

Table 1 shows the characteristics of the sample by intervention status. Tables 2 and 3 show the ICCs, the asymptotic confidence intervals, and the bias-corrected bootstrap confidence intervals for repeated measurements for the practice and provider levels, respectively. ICC values have the ranges 0.012–0.110 and 0.018–0.110 at the practice and provider levels, respectively. Tables 4 and 5 show the ICCs, the asymptotic confidence intervals, and the bias-corrected bootstrap confidence intervals for baseline-only data for the practice and provider levels, respectively. For the baseline-only approach, ICC values have the ranges 0.0007–0.141 and 0.017–0.160 at the practice and provider levels, respectively.

4. Discussion

We compared both the ICC estimates and their associated standard error estimates across the items for the repeated-measurement approach and the baseline-only approach, examining them at three levels: patient-family, provider, and practice. Because we pooled together intervention and control groups for estimating ICCs, it is important to verify that the characteristics of the two groups are similar. Table 1 shows that the summary statistics are

comparable between the intervention and control groups, even though some variables have significant *p*-values, which are due to the relatively large sample sizes.

Except in the case of media use, the ICC values for the repeated-measurement and baseline-only approaches generally agreed both at the practice level (Tables 2 and 4) and at the provider level (Tables 3 and 5). We noticed that the variance component from repeated measurements for media is quite large (12.2 vs. 0.53 from practice) because of a substantial amount of variation in reported media use within the family over the 6-month period of the study. Also, with the exception of media use, all of the repeated-measurement ICC estimates—provider and practice level included—fall within the 95% confidence intervals calculated from the baseline-only approach using Eq. 2. If we assume that the repeated-measurement ICC values are close to the population ICC values, then this would indicate that Eq. 2 is a good approximation of the true standard error.

Compared with the baseline-only approach, the confidence intervals for ICCs were narrower, indicating precision, when repeated measurements were used. At the practice level, the mean width of the ICC confidence interval using repeated measurement is 0.020 (Table 2), while that from using baseline-only data is 0.048 (Table 4). The reduction in the width of the confidence interval is 58%. At the provider level, the reduction is 42% (mean widths are, respectively, 0.027 and 0.048 from Tables 3 and 5). In terms of the ratio of standard error to ICC value, which indicates the size of the relative margin of error (smaller implies higher precision), the values (SD) for repeated measurements and baseline only are, respectively, 0.189 (0.03) and 0.260 (0.09) for provider-level ICCs. For practice-level ICCs, the values are, respectively, 0.223 (0.178) and 0.328 (0.147).

The results from bootstrap 95% confidence intervals lend some support to the validity of Eqs. 2 and 4 for calculating standard errors. For example, for repeated-measurement practice-level ICCs, the average absolute difference (SD) between the lower confidence intervals for the asymptotic and bootstrap methods is 0.0046 (0.0027). For the upper confidence intervals, the average absolute difference is 0.0044 (0.0030). Generally, the bootstrap method produces slightly wider confidence intervals. The bootstrap method also shows that, with the exception of media use, the results from using repeated-measurement and baseline-only are both rather stable. While the bootstrap confidence intervals agree reasonably well with the asymptotic calculation, especially for the repeated-measurement ICCs, the bootstrap methods tend to produce asymmetric confidence intervals. Figure 1 shows the asymmetry of the width of the left and right bootstrap confidence intervals for practice-level ICCs from repeated measurements. The corresponding bootstrap confidence intervals for the baseline-only method are shown in Fig. 2. Not surprisingly, the widths of the right confidence intervals tend to be longer because ICCs are bounded by zero on the left. However, it is unexpected that the symmetrical baseline-only asymptotic confidence intervals contain more repeated-measurement ICCs than the asymmetrical bootstrap confidence interval method does. The bootstrap confidence intervals only contain 8 and 6 repeated measurements for practice and provider ICCs, respectively, while the asymptotic confidence intervals contain 9 and 9, out of 10 and 10 confidence intervals, respectively.

Because of attrition, follow-up measurements may be missing. Furthermore, nonresponse to individual items was also present. In order to produce comparable estimates for the baseline-only and repeated-measurement approaches, in this paper we only reported results of families that had data on all three time points. Within this “complete case” sample, nonresponse rates varied across different items, and the average rate was approximately 8%. The computational procedure in PROC MIXED does not delete subjects that contain partially missing values [20], and it handles missing values by making the assumption that missing data are missing at random [21]. This flexibility allows the preservation of partially

missing data cases across repeated measurements and increases the power of the proposed procedure. For data within which a non-missing-at-random mechanism might be potentially at work, the estimation of ICC will require further investigation.

A limitation of this study is that some of the questions reported in this paper contain ordinal, not continuous, response categories. Although we have included only questions for which distributions are bell-shaped, the procedure should only be viewed as an approximation estimating method. For rigorously treating ordinal data, one possible solution is to develop a generalized nonlinear model framework and utilize programs such as PROC NLMIXED. For ICC estimation for binary data, alternative procedures are available for cross-sectional data [22] and could be further extended to repeated measurement.

With the growth of cluster randomized trials conducted in research-based networks, the estimation and reporting of intraclass correlation will play an increasingly important role in study design and data analysis. From an evidence-based approach that aims to improve the quality of reporting, the Consolidated Standards of Reporting Trials (CONSORT) statement recommends standard reporting protocols. It is clear that ICC must be a part of this trial report. Even more importantly, to design a group-randomized, controlled trial with sufficient power to test hypotheses, a reasonable estimate of the ICC must be utilized [8,23]. We report here on ICCs generated from national data derived from an office-based clinical trial in the United States. Previously published ICC estimates for pediatric office-based visits, to our knowledge, are rare. In addition, except in a limited setting [24,25], it appears that the use of repeated measurements to improve ICC estimates also has not been fully studied. Based on the analysis of a U.S. national sample of relatively large size with repeated measurements, we expect that both the ICC estimates and the methods for improving their precision reported in this paper are generalizable.

5. Conclusion

In conclusion, the repeated measurement approach provides ICC estimates with substantially narrower confidence intervals. The asymptotic standard error estimates for both repeated measurement and baseline-only tend to be slightly smaller, as evaluated against the nonparametric bootstrap procedure. The ICC values derived from baseline-only and repeated measurement estimation methods generally agree well. For variables that have a large variation over repeated measurements, the ICC estimates using repeated-measurement and baseline-only data can be very different.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by a grant from the National Institute of Child Health and Human Development (NICHD) (R01 HD 42260), the Agency for Healthcare Research and Quality (AHRQ), The Robert Wood Johnson Generalist Faculty Scholars Program (RWJGFS), and the American Academy of Pediatrics' Friends of Children Fund.

We especially appreciate the efforts of the PROS practices and practitioners. The pediatric practices or individual practitioners who enrolled participants in this study are listed here by AAP Chapter: *Alabama*: Pediatric Care Group (Montgomery); *Alaska*: Anchorage Pediatric Group, LLC (Anchorage), Practice of Joy Neyhart, MD (Juneau); *Arizona*: Orange Grove Pediatrics (Tucson), Tanque Verde Pediatrics (Tucson); *California-1*: Practice of Arthur S Dover, MD (Freedom), Sierra Park Pediatrics (Mammoth Lakes), Palo Alto Medical Foundation (Palo Alto), Pediatric & Adolescent Medical Associates of the Pacific Coast, Inc (Salinas); *California-3*: East County Community Clinic – Lakeside (Lakeside), Pediatric Medical Associates of Tri-City, Inc (Vista), La Jolla Pediatrics (La Jolla); *California-4*: Edinger Medical Group & Research Center, Inc (Fountain Valley); *Colorado*: Community Health Services (Commerce City), Lamar Pediatrics (Lamar), Rocky Mountain Health Centers, North (Denver);

Connecticut: Mauks Koepke Medical, LLC (Danbury), Jeff Cersonsky, MD (Southbury), Pediatric Associates of Connecticut, PC (Waterbury); *Florida*: Atlantic Coast Pediatrics (Merritt Island), Family Health Center East & Oviedo Children's Health Center (Orlando), Heartland Pediatrics of Lake Placid (Lake Placid); *Georgia*: The Pediatric Center (Stone Mountain), Practice of Victor Lui, MD (Chamblee), Practice of Nandlal Chainani, MD (Ocilla), Snapfinger Woods Pediatric Associates, PC (Decatur), Gwinnett Pediatrics & Adolescent Medicine (Lawrenceville); *Hawaii*: Island Youth Heart and Health Center (Hilo), Children's Medical Association Inc (Aiea), Medicine Pediatrics Associates (Honolulu); *Iowa*: Children's Hospital Physicians (Des Moines); *Illinois*: Yacktmann Children's Pavillion (Park Ridge), S.W. Pediatrics (Orland Park), Stroger Hospital of Cook County (Chicago), Macomb Pediatrics, SC (Macomb); *Indiana*: Georgetown Pediatrics (Indianapolis), Jeffersonville Pediatrics (Jeffersonville); *Kansas*: Ashley Clinic (Chanute); *Louisiana*: Carousel Pediatrics (Metairie), Shalom Clinic for Children (Natchitoches), The Baton Rouge Clinic, AMC (Baton Rouge); *Massachusetts*: Burlington Pediatrics (Burlington), Holyoke Pediatric Associates (Holyoke), Pediatric Associates of Norwood (Franklin), Mary Lane Pediatric Associates (Ware); *Maryland*: Practice of Steven E Caplan, MD, PA (Baltimore), Dundalk Pediatric Associates (Baltimore), Practice of Ralph Brown, MD (Baltimore); *Maine*: Maine Coast Memorial Hospital (Ellsworth); *Michigan*: Pediatric & Adolescent Medicine (Bay City), Pediatric Health Care (Sterling Heights); *Minnesota*: Brainerd Medical Center, PA (Brainerd), Lakeview Clinic - Watertown Pediatrics (Watertown); *Missouri*: Children's Mercy Hospital Pediatric Care Center (Kansas City), Tenney Pediatric and Adolescent LLC (Kansas City); *North Carolina*: Guilford Child Health, Inc - Greensboro (Greensboro), Goldsboro Pediatrics, PA (Goldsboro), Aegis Family Health Center - Winston East Pediatrics (Winston-Salem), Guilford Child Health, Inc - High Point (High Point); *North Dakota*: Altru Clinic (Grand Forks); *New Hampshire*: Foundation Pediatrics (Nashua); *New Jersey*: Lourdes Pediatric Associates (Camden), Chestnut Ridge Pediatric Associates (Woodcliff Lake); *New York-1*: Elmwood Pediatric Group (Rochester), Lewis Pediatrics (Rochester), United Medical Associates Pediatrics (Binghamton), Wayne Medical Group (Williamson); *New York-3*: Pediatric Primary Care-Montefiore Medical Center (Bronx), Cardinal Mc Closkey Services (Bronx), Pediatric Practice Bronx-Lebanon Hospital (Bronx), Westchester Avenue Medical and Dental Center (Bronx), Bronx Lebanon Pediatric Clinic - Third Avenue (Bronx); *New Mexico*: Presbyterian Family Healthcare - Rio Bravo (Albuquerque), Santa Fe Pediatric Associate, PC (Santa Fe); *Ohio*: Oxford Pediatrics & Adolescents (Oxford), Pediatric Associates of Lancaster (Lancaster); *Oklahoma*: Pediatric & Adolescent Care, LLP (Tulsa), OK State University (OSU) - Center for Health Sciences (Tulsa); *Oregon*: NBMC (Coos Bay); *Ontario*: Richard J MacDonald, MD (Oakville, Ontario); *Pennsylvania*: Pennridge Pediatric Associates (Sellersville), Buckingham Pediatrics (Buckingham), Pediatric Practices of Northeastern Pennsylvania (Honesdale), Laurel Health Center - Blossburg (Blossburg); *Quebec*: Clinique Enfant-Medic (Dollard des Ormeaux); *Puerto Rico*: Practice of Dra Ethel Lamela, MD (Isabela). *Rhode Island*: Northstar Pediatrics (Providence); *South Carolina*: Edisto Pediatrics (Walterboro), Oakbrook Pediatrics (Summerville), Palmetto Pediatrics & Adolescent Clinic, PA (Columbia), Barnwell Pediatrics, PA (Barnwell); *Tennessee*: ETSU Physicians & Associates (Johnson City), Pediatric Consultant, PC (Memphis), Memphis and Shelby County Pediatric Group (Memphis); *Texas*: The Pediatric Clinic (Greenville), Winnsboro Pediatrics (Winnsboro), Su Clinica Familiar (Harlingen), Parkland Health & Hospital System (Dallas), Child Wellness Center (Horizon City), Danette Elliott-Mullens, DO, PA (New Braunfels); *Utah*: University of Utah Hospitals & Clinics (Park City), Utah Valley Pediatrics, LC (American Fork), University of Utah Health Sciences Center (Salt Lake City), Willow Creek Pediatrics - Draper (Draper), IHC Health Center - Memorial (Salt Lake City); *Virginia*: Tidewater Pediatric Consultants, PC (Virginia Beach), Hampton Roads Pediatrics/CMG (Hampton), Alexandria Lake Ridge Pediatrics (Alexandria), Pediatrics of Arlington, PLC (Arlington), Fishing Bay Family Practice (Deltaville); *Vermont*: University Pediatrics, UHC Campus (Burlington), Pediatric Medicine (South Burlington), Brattleboro Primary Care (Brattleboro), University Pediatrics (Williston), Practice of Rebecca Collman, MD (Colchester), Springfield Pediatric Network (Springfield); *Washington*: Harbor Pediatrics (Gig Harbor), Central Washington Family Medicine (Yakima); *Wisconsin*: 16th Street Community Health Center (Milwaukee), Beloit Clinic SC (Beloit), Gundersen Clinic - Whitehall (Whitehall), Ministry Medical Group - Woodruff (Woodruff); *West Virginia*: Grant Memorial Pediatrics (Petersburg); *Wyoming*: Jackson Pediatrics, PC (Jackson).

References

1. Donner, A.; Klar, N. Design and analysis of cluster randomization trials in health research. Arnold; 2000.
2. Eldrige, et al. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care? Clin Trials 2004;1:80-90. [PubMed: 16281464]
3. Murray DM, Short B. Intraclass correlation among measures related to alcohol use by young adults: estimates, correlates and applications in intervention studies. J.Stud.Alcohol 1995;56:681-694. [PubMed: 8558900]
4. Hannan PJ, 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. [PubMed: 8117787]

5. Parker DR, Evangelou E, Eaton CB. Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART). *Contemp Clin Trials* 2005;26:260–267. [PubMed: 15837446]
6. Barkin, et al. Is office-based counseling about media use, timeouts, and firearm storage effective? Results from a cluster-randomized, controlled trial. *Pediatrics* 2008;122:15–25.
7. Baskerville NB, Hogg W, Lemelin J. The effect of cluster randomization on sample size in prevention research. *J Fam Pract* 2001;50:W241–W246. [PubMed: 11252223]
8. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am J Public Health* 2004;94:423–432. [PubMed: 14998806]
9. Adams G, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785–794. [PubMed: 15485730]
10. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;20:391–399. [PubMed: 11180309]
11. Fisher, R. Statistical methods for research workers. Oliver and Boyd; 1925.
12. Murray DM, Clark MH, Wagenaar AC. Intraclass correlations from a community-based alcohol prevention study: the effect of repeated observations on the same communities. *J Stud Alcohol* 2000 November;;881–890. [PubMed: 11188494]
13. Efron, B.; Tibshirani, R. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
14. Roberts J, Fan X. Bootstrapping within the multilevel/hierarchical linear modeling framework: a primer for use with SAS and Splus. *Multiple Linear Regression Viewpoints* 2004;30:23–34.
15. Thompson B. Exploring the replicability of a studys results—bootstrap statistics for the multivariate case. *Educ Psychol Meas* 1995;55:84–94.
16. Ukoumunne OC, et al. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Stat Med* 2003;22:3805–3821. [PubMed: 14673940]
17. Eldridge SM, Ukoumunne OC, Carlin JB. The intraclass correlation coefficient in cluster randomized trials: a review of definitions. *Int Stat Rev* 2009;77:378–394.
18. Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *Am J Epidemiol* 1999;149:876–883. [PubMed: 10221325]
19. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Stat* 1998;23:323–355.
20. Wolfinger RD. An example of using mixed models and PROC MIXED for longitudinal data. *J Biopharm Stat* 1997;7:481–500. [PubMed: 9358325]
21. Little, R.; Rubin, D. Statistical analysis with missing data. Wiley; 2002.
22. Gulliford MC, et al. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005;58:246–251. [PubMed: 15718113]
23. Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. *Clin Trials* 2005;2:99–107. [PubMed: 16279131]
24. Rosner B. On the estimation and testing of interclass correlations: the general case of multiple replicates for each variable. *Am J Epidemiol* 1982;116:722–730. [PubMed: 7137158]
25. Murray DM, et al. Analysis strategies for a community trial t reduce adolescent ATOD use: a comparison of random coefficient and ANOVA/ANCOVA models. *Contemp Clin Trials* 2006;27:188–206. [PubMed: 16324889]

FIGURE 1.
Widths of left and right confidence intervals for ICC estimates using bias-corrected bootstrap methods for repeated measurement approach.

FIGURE 2.
Widths of left and right confidence intervals for ICC estimates using bias-corrected bootstrap methods for baseline-only approach.

TABLE 1

Demographic Characteristics of the U.S. National Sample of Study Participants at Baseline in Year 2005 (N = 3,294)

Domain	Intervention N (%)	Control N (%)
Age		
2–5 yrs	927 (51.9)*	774 (55.2)
6–11 yrs	860 (48.1)	628 (44.8)
Adults in the home		
2 parents	1,330 (74.2)	1,035 (72.5)
1 parent	459 (25.6)	384 (26.9)
Other	3 (0.2)	9 (0.6)
Income		
< \$40,000/yr	635 (38.7)	562 (42.6)
\$41–80,000/yr	523 (31.8)	462 (35.1)
> \$80,000/yr	485 (29.5)	294 (22.3)**
Race/Ethnicity		
African American	237 (14.2)	137 (10.3)
Latino	276 (16.6)	290 (21.7)
White	1,043 (62.5)	809 (60.5)
Other	112 (6.7)	101 (7.5)**
Maternal Education		
< High School	172 (9.6)	137 (9.7)
> HS < College	957 (53.5)	820 (58.2)
> = College	658 (36.8)	453 (32.1)*
Respondent		
Mother	1,651 (90.3)	1,288 (89.9)
Father	130 (7.2)	111 (7.8)
Other	44 (2.5)	46 (3.3)

* p<0.05,

** p<0.01

Variance Components, ICCS, Asymptotic Confidence Intervals, and Bias-Corrected Confidence Intervals at Practice Level With Repeated Measurements

TABLE 2

Variable	ICC	LB	UB	BSLB	BSUB
Media use	0.040	0.029	0.052	0.026	0.055
Let child watch TV alone	0.061	0.044	0.077	0.046	0.076
Restrict TV	0.023	0.015	0.030	0.010	0.035
Explain TV	0.012	0.007	0.016	0.008	0.015
Use time out	0.021	0.014	0.028	0.010	0.032
Use yelling	0.019	0.010	0.028	0.008	0.030
Use taking away privilege	0.020	0.011	0.028	0.008	0.031
Use spanking	0.110	0.083	0.137	0.093	0.126
Watching violent TV makes child more afraid	0.022	0.013	0.030	0.009	0.034
Watching violent TV makes child think violence is normal	0.024	0.016	0.031	0.011	0.036

* The columns (from left to right) respectively represent intraclass correlation, lower and upper bounds of confidence limits, and bias-corrected bootstrap lower and upper bounds.

TABLE 3
ICCs, Asymptotic Confidence Intervals, and Bias-Corrected Confidence Intervals at Provider Level With Repeated Measurements

Variable	ICC	LB	UB	BSLB	BSUB
Media use	0.055	0.037	0.072	0.044	0.066
Let child watch TV alone	0.072	0.051	0.094	0.061	0.087
Restrict TV	0.024	0.014	0.033	0.016	0.033
Explain TV	0.018	0.010	0.026	0.009	0.028
Use time out	0.021	0.012	0.030	0.014	0.029
Use yelling	0.019	0.011	0.027	0.011	0.030
Use taking away privilege	0.020	0.011	0.028	0.011	0.028
Use spanking	0.0110	0.080	0.139	0.096	0.126
Watching violent TV makes child more afraid	0.038	0.025	0.051	0.030	0.048
Watching violent TV makes child think violence is normal	0.024	0.014	0.033	0.014	0.032

TABLE 4

Variance Components, ICCS, Asymptotic Confidence Intervals, and Bias-Corrected Confidence Intervals at Practice Level for Baseline-Only Measurement

Variable	ICC	LB	UB	BSLB	BSUB
Media use	0.114	0.069	0.160	0.070	0.157
Let child watch TV alone	0.091	0.053	0.129	0.064	0.122
Restrict TV	0.023	0.008	0.038	0.001	0.053
Explain TV	0.007	0.000	0.016	0.007	0.026
Use time out	0.027	0.011	0.044	0.003	0.053
Use yelling	0.020	0.006	0.034	0.000	0.047
Use taking away privilege	0.030	0.012	0.048	0.006	0.062
Use spanking	0.141	0.088	0.193	0.110	0.172
Watching violent TV makes child more afraid	0.020	0.006	0.034	0.009	0.053
Watching violent TV makes child think violence is normal	0.031	0.013	0.049	0.006	0.056

TABLE 5
 ICCs, Asymptotic Confidence Intervals, and Bias-Corrected Confidence Intervals at Provider Level for Baseline-Only Measurement

Variable	ICC	LB	UB	BSLB	BSUB
Media use	0.160	0.115	0.204	0.128	0.192
Let child watch TV alone	0.095	0.063	0.127	0.075	0.119
Restrict TV	0.043	0.023	0.063	0.026	0.063
Explain TV	0.017	0.003	0.031	0.000	0.036
Use time out	0.027	0.011	0.044	0.010	0.046
Use yelling	0.020	0.005	0.034	0.001	0.042
Use taking away privilege	0.030	0.013	0.047	0.012	0.051
Use spanking	0.150	0.107	0.192	0.122	0.172
Watching violent TV makes child more afraid	0.038	0.019	0.057	0.021	0.060
Watching violent TV makes child think violence is normal	0.039	0.020	0.058	0.016	0.058