Second, the authors did not stratify according to age, whereas it is known that the essence and rates of language development in infants are different from toddlers and school children. Speech and language develop more rapidly during the first 2 years of life, than in subsequent years. Furthermore, it has been argued that (intertemporal) hearing loss caused by OME has different effects on early and later language development because of 1-to-1 interaction typically for early childhood contrasted with communication in noise at later ages.

Maroeska M. Rovers, PhD
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
3508 AB Utrecht, Netherlands

Gerhard A. Zielhuis, PhD
Department of Epidemiology and Biostatistics
University Medical Center Nijmegen
6500 HB Nijmegen, Netherlands

REFERENCES


In Reply.—

We appreciate the opportunity to respond to the letter of Drs Rovers and Zielhuis regarding our article published in the March 2004 issue of Pediatrics.1 Drs Rovers and Zielhuis describe how the conclusion of the study is in agreement with the available evidence but cite 2 areas in which they feel that the methods used in the meta-analysis are incorrect.

First, Rovers and Zielhuis question our pooling of data from 3 randomized, clinical trials examining the effects of treatment with ventilation tubes on speech and language in children with persistent otitis media with effusion (OME) and 4 observational studies of children with varying histories of OME. As noted in our methods section, group studies (which include both of these study types) “compared outcomes in ≥2 independent and parallel groups of children with differing levels of OM based on historical experience or randomization to tympanostomy tubes versus watchful waiting.” Whereas we agree with Rovers and Zielhuis that the study populations vary, we considered them similar enough for statistical pooling. As noted in our results section, analysis of variance revealed no difference in outcomes for all meta-analyses when data were stratified by randomized versus observational design. The random-effects model was used to provide conservative confidence limits on effect size, not to justify pooling studies that should not be combined. We feel that we have provided readers with enough detail regarding methods and results to form their own conclusions about overall validity of the analyses.

Second, Rovers and Zielhuis state that we did not stratify our samples by age and note that language development as well as the potential effect on hearing loss caused by OME differ in infants, toddlers, and school-aged children. In the reported analyses in this article, we did stratify our samples by age. We grouped studies in our review by the age of the speech or language outcome assessment: infancy (1–2 years), preschool (2–5 years), or school age (5–8 years). Because data had to be available from ≥3 studies to justify statistical pooling for each meta-analysis, we were able to examine only the following age groupings: 1) hearing loss and receptive language and expressive language at 1–2 years; 2) OME and receptive language and expressive language at 2–5 years; 3) OME and receptive language, expressive language, receptive vocabulary, and speech at 3 years old; and 4) OME and expressive vocabulary and expressive syntax at 3–5 years. We are very aware of the differences in language development during early childhood, and that is why we stratified our samples.

We appreciate the comments from Rovers and Zielhuis about our work.

Joanne E. Roberts, PhD
Frank Porter Graham Child Development Institute and Departments of Pediatrics and Speech and Hearing Sciences
University of North Carolina Chapel Hill, NC 27599-8180

Richard Rosenfeld, MD, MPH
Department of Otolaryngology
State University of New York Downstate Medical Center and Long Island College Hospital
Brooklyn, NY 11201

Susan A. Zeisel, EdD
Frank Porter Graham Child Development Institute
University of North Carolina
Chapel Hill, NC 27599-8180

REFERENCE


The Review Process Fails to Require Appropriate Statistical Analysis of a Group-Randomized Trial

To the Editor.—

We are surprised that the journal’s review process of the school-randomized trial reported by MacKelvie et al1 did not insist on an analysis appropriate to the group-randomized design or at least require stronger justification of the assumptions involved in ignoring the randomization design in the analysis.

Randomizing intact social groups is a common approach outside the clinic because it is often easier, and possibly only feasible, to intervene with a whole class, troop, church, or community rather than to work with individuals. Members within intact social groups tend to be more like each other than they are like members in other groups, making for some redundancy of information and increased variance compared with the same number of subjects individually randomized. As a general rule, group-randomized trials that are analyzed by using methods appropriate for individual-level trials will overestimate the significance of the effects.

Twenty-five years ago, Cornfield2 warned clearly that “randomization by cluster accompanied by an analysis appropriate to randomization by individual is an exercise in self-deception and should be discouraged.” Methodological reviews (eg, Donner et al,3 Simpson et al,4 and Smith et al5) show that Cornfield’s message is not well heeded and point to neglect in the review process for insisting on appropriate attention to the analytic issues incurred by the choice of a group-randomized trial. When randomization is by group but analyzed by individual, chance differences between the schools randomized to the control versus the intervention condition are confounded with the intervention effect. If these chance differences are not separated out as a school-random effect, 2 consequences occur: 1) the realized difference is attributed to the intervention, and 2) the variation against which to assess the intervention effect is underestimated,6 making for an overly sensitive analysis that increases the probability of false-positive inference. The reviewers allowed a curious statement: “…we analyzed and adjusted for these differences by subject so as not to bias results by comparing school means.”7 Adjustment for individual covariates is possible in a group-randomized trial.7 In the limitations the authors say: “Randomization by school, as opposed to individual children, introduced bias and influenced the generalizability of results. However, this design was most feasible…”8 in avoiding contamination, etc. Yes, there are good reasons for randomizing intact social groups, but the statistical and inferential implications cannot then be ignored.

Because Pediatrics is a prominent vehicle for disseminating important pediatric research, we urge that the editorial and review process fails to require appropriate statistical analysis of group-randomized trials.
process incorporate more careful statistical review of submissions in which the study design is group-randomized.

Peter J. Hannan, MStat
Simone A. French, PhD
John H. Himes, PhD, MPH
Jayne A. Fulkerson, PhD
Mary Story, PhD
Division of Epidemiology
School of Public Health
University of Minnesota
Minneapolis, MN 55454-1015

REFERENCES


In Reply.—

We thank Mr Hannan et al for their comments with respect to our school-based intervention.1 We appreciate the opportunity to discuss the cluster randomization model and address some of the challenges confronting those who conduct “real-world” pediatric research trials. Hannan et al highlight the need for appropriate methods of statistical analysis when cluster randomization is used in school or community trials. School-based studies have a nested structure whereby schools reside within a community, classes within schools, and students within classes.

In the context of our intervention study, randomization of schools was the only means to deliver an exercise program to children without contaminating the control group. Because the intervention frequently took place outside of closed classrooms (ie, on the playground), changes in physical activity patterns within schools would easily be observed by controls if children were randomized at an individual level within classes or if classes (within schools) were randomized as clusters. Hannan et al raise the issue that responses observed on students within the same class or within the same school may be correlated. This correlation, if it exists, should then be accounted for in the data analysis.

One method of analyzing clustered data is to take the school as the unit of analysis and to compare the means of the responses observed at each school. Because these derived data (school means) can be treated as independent observations, this approach allows us to circumvent the correlation issue altogether. However, we believe that this method is inappropriate for 2 reasons. First, although teachers delivered the intervention to all students in their class, we assessed only those children who provided consent. Thus, the numbers of participants ranged from 1 to 12 per school at the end of the 20-month trial. The wide range in number of participants per school implies that the variability associated with the timing and magnitude of linear growth, which is highly individualized, represents the greatest source of variability in pediatric bone studies.2,3 Adjustment for confounders on an individual level, which is not possible when school is the unit of analysis, is therefore desirable. Simpson et al4 recommend using the individual as the unit of analysis while accounting for relationships between responses of individuals in the same cluster.

In our original analysis, we did not specify either class or school as clusters. With respect to schools, we agree with Hannan et al that “chance” differences may exist among schools and that this warrants both biological and statistical consideration. However, our target schools in the Richmond school district are located in close geographical proximity and are fairly homogenous in their racial and socioeconomic mix. We had no reason to believe that between-school differences would be greater than within-school differences. With respect to classes, we would expect differences in the way that the physical activity intervention is delivered by classroom teachers and therefore that responses observed on children in the same class may be correlated. However, students received the intervention from different classroom teachers as they progressed through grade levels, and classes do not remain as intact groups over school years (thus, “class” is not a well defined cluster in our study). To compensate for this limitation of the study, we conducted classroom visits and teacher training to standardize the intervention across classes and schools to equalize teacher influence across cases as far as possible. Formally, we may assess the effect of school on the change in bone mineral content at the femoral neck and the lumbar spine using a linear mixed model, with the school effect designated as random. We incorporated baseline bone mineral content, height, height change, maturity, and physical activity outside of school as covariates. In both analyses, the variability across schools (femoral neck: SD = 0.00001; lumbar spine: SD = 0.13) was far less than the variability among students within each school (femoral neck: SD = 0.20; lumbar spine: SD = 2.70). Furthermore, the estimated effect of the intervention and its standard error were similar to those in our original analysis (when school was not designated as random), which also suggests a negligible school effect.

We welcome the debate on the utility of schools as the unit of randomization and analysis in our work with children and look forward to ongoing discussions of these methods through our work and that of our colleagues.

Kerry MacKelvie O’Brien, PhD
Endocrinology and Diabetes Unit
BC Children’s Hospital
Food, Nutrition and Health
University of British Columbia
Vancouver, BC, Canada V6H 3V4

Heather McKay, PhD
Department of Orthopaedics/Family Practice
Division of Orthopaedic Engineering Research
University of British Columbia
Vancouver, BC, Canada V5Z 1L8

Rachel M. Altman, PhD
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada V5A 1S6

Patricia A. Janssen, PhD
Department of Health Care and Epidemiology
University of British Columbia
St Paul’s Hospital
Vancouver, BC, Canada V6Z 1Y6

Karim Khan, MD, PhD
Department of Family Practice
University of British Columbia
Vancouver, BC, Canada V6T 1V6

REFERENCES

2. Bailey DA, McKay HA, Mirwald RL, Cracker PRE, Faulkner RA. The University of Saskatchewan Bone Mineral Accrual Study: a six year

Television Viewing and Attention Deficits in Children

To the Editor.—

The report by Christakis et al.1 suggests that early-age television viewing is associated with attention deficits later in childhood. It has been argued that television promotes inactivity, obesity, promiscuity, and possibly aggression in adults, and thus the impact of television viewing by children is certainly a concern. The message resonates in a society seemingly obsessed with public health villains. Although television eventually may earn a place among contemporary threats to our well-being, Christakis et al do not provide a convincing case for this conclusion.

The use of 1.2 standard deviations above the mean on a hyperactivity scale as the threshold for attention disorders is problematic. The authors defend the arbitrary threshold by noting that it corresponds to previous reports of attention-deficit/hyperactivity disorder (ADHD) prevalence in age-matched community samples. However, this misapplication of statistical data seriously compromises the very foundation of the authors’ conclusions.

In the first place, the choice of 1.2 standard deviations above the mean as the threshold for attention disorders is specious and begs the question by assuming the conclusion before advancing the argument. According to the American Diabetes Association, 6% of the US population has diabetes, but this does not imply that everyone with blood glucose >94th percentile has diabetes, and measurement of blood glucose is a far more reliable test than subjective responses to behavioral surveys. Second, the 10% incidence of ADHD cited by the authors is not entirely consistent with recent estimates. Previous reports state that ADHD affects anywhere from 4% to 12% or from 3% to 10% of children in the United States.3 Finally, the authors admit that the surveys used in this study are not equivalent to a clinical diagnosis of ADHD, yet they are comfortable with choosing a cutoff that fits the bona fide incidence of ADHD.

We also suspect that the authors selectively indite television viewing because it fits their hypothesis, overlooking other possibilities apparent in their results. Inspection of the data in their Table 2 reveals that maternal self-esteem is more highly correlated with the outcome measure of attention deficit than the hours of television watched per day, a result never mentioned in the discussion. Proving a causal relationship between television viewing and behavioral traits has always been difficult. The authors of this study acknowledge that content is quite likely a critical determinant in television-viewing influence, but their conclusion does not take this important factor into account. Children raised in an environment that promotes attention-seeking behavior may have greater exposure to television because of factors that limit more wholesome activities. On the other hand, it makes just as much (or more) sense to hypothesize that poorly behaved children are exposed to more television by parental choice.

Thus, particular care must be taken when interpreting statistical arguments built on such precarious footing. When reviewers and readers evaluate conclusions based on statistical analysis of mined data, would be wise to consider the likelihood that the statistics are being used, in the words of Andrew Lang, “...as a drunken man uses lampposts—for support rather than illumination.”

ROGER L. BERTHOLF, PhD
STEVE GOODISON, PhD
Department of Pathology
University of Florida Health Science Center
Jacksonville, FL 32209

REFERENCES


In Reply.—

We appreciate the opportunity to reply to Drs Bertholf and Goodison. Their first concern is with our cutoff for attentional problems. The use of standard deviations from the mean to define an “abnormal” level is in fact a common approach. Although it is by no means diagnostic of a problem (and we never make a claim that it is), it suggests that the reported value is far from the mean and at minimum would represent worrisome (and attentional) problems. Although one could set the threshold either higher or lower, we chose that point in part because it is consistent with the recent population-based estimates of attention-deficit/hyperactivity disorder (ADHD), a clinical diagnosis that, by definition, represents abnormal ability to attend. In one multicenter study, 9% of children presenting for nonacute primary care visits had an attention disorder,1 whereas the percentage of elementary school children in a general population sample who meet Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria for ADHD was estimated recently at 11.4%.2 Furthermore, we believe that a cutoff of the 10th percentile has face validity. The readers are free to decide on their own if the associations we found between our primary exposure (television) and being at or below that percentile for attention are clinically meaningful.

They also complain about the problem of measurement error because of the subjectivity of parental surveys, but they are incorrect in their assessment of what the likely effects of measurement error would be on our analysis. Generally, random measurement error would reduce the ability to detect an effect in data by biasing toward the null,2 so our significant finding despite the existence of measurement error suggests that, if anything, the true effect may be stronger than what we were able to identify.

As to the question of the relative contribution of the different correlates, we believe that their concern is misguided. Following standard scientific methods, we conducted this study to explicitly test an a priori specific hypothesis that has existed in the literature for a long time, in support of which there has been sound, if only circumstantial, evidence to date. The other variables in our model were control variables that might plausibly confound our primary relationship of interest but about which we did not in fact have hypotheses articulated ex ante. Although we did not discuss them in the article, their directionality is in fact plausible, and they themselves may warrant additional investigation. Finally, the fact that they are controlled for in our multivariate model suggests that television viewing is associated with attentional problems even when these other independent predictors are adjusted for, which is a strength of our analysis.

For what it may be worth, the letter writers are also incorrect in their assessment of the relative effects of early television exposure and mother’s self-esteem. It is true that the odds ratio presented in our table is greater for self-esteem than for hours of television watched. However, the variables are not scaled the same: an inspection of Table 1 in the article will reveal that the standard deviation of mother’s self-esteem is 1.0, whereas the standard deviation of television hours watched per day is 2.9. Adjusting for the difference in the scaling of the variables, the magnitude of the effects of a 1-standard-deviation change in mother’s self-esteem and television hours watched per day are similar.

Bertholf and Goodison rightly argue that our analysis cannot be seen as the definitive word on the relationship between television and attentional problems. We acknowledged the limitations of the data in our article and couched our conclusions with the appropriate caveats. Nevertheless, we believe that our findings are important and innovative and have moved the field forward, most positively and some defensively. As with all new findings, however, only comprehensive subsequent research effort