Fatal flaws in the design of pediatric ophthalmology and strabismus studies

Scott K. McClatchey, MD

When I started my fellowship, my 2-year-old son had a cataract in one eye that was slowly worsening. Because of this, I sought to understand the myopic shift after cataract surgery and intraocular lens implantation in young children, based on the aphakic patients of Dr. Marshall Parks. Parks said that in his experience, the type of cataract mattered: those with nuclear cataract had rapid and large postoperative myopic shift, whereas those with lamellar cataract had a much smaller myopic shift. When I plotted the refractions of his patients versus the time after surgery, this was indeed the case! But I had asked the wrong question. Eyes with nuclear cataract had surgery in infancy; those with lamellar cataract often had surgery after 2 years of age. When the data were transposed to aphakic refraction versus age, a completely different answer was clear: all cataract types followed the same semilogarithmic curve. Although the correlation between myopic shift and cataract type was real, the inference that the type of cataract caused varied myopic shift was false. Asking the right question of carefully gathered data led to a predictive model (rate of refractive growth) that applies equally to aphakic, pseudophakic, and normal eyes.[1]

The task of a researcher in pediatric ophthalmology appears simple at first. Children may have one underlying eye problem rather than the accumulated pathology that occurs in the elderly, but this simplicity is deceiving. A problem such as anisometropic hyperopia can lead to a cascade of sequellae, such as amblyopia, esotropia, and excyclotorsion. In a group of subjects, the effects of varied refractive errors, follow-up period, age, variable treatment history, and visual development compound this complexity. The researcher who attempts to study a seemingly simple subject, such as patching for amblyopia, must take this into account and carefully frame the research question, or the heterogeneity of subjects will bias and confound any results. High-quality research in pediatric ophthalmology is difficult to perform because of many factors, including the challenges in obtaining precise data in uncooperative children and the need for third-party consent. As a result, most published studies in our field are retrospective, small, or potentially confounded by the many factors we deal with clinically on a daily basis.

The *Journal of the American Academy for Pediatric Ophthalmology and Strabismus* receives hundreds of manuscripts each year. Many have problems with writing, data presentation, and overreaching conclusions. These problems are fixable, based on feedback from peer reviewers and editors. However, for a study to be worth publishing in *J AAPOS*, it must have a clinically relevant finding that is supported by correct analysis of the data and is of interest to our readers. If the research question is not carefully considered, the design is haphazard, or the data are of poor quality, there is no point in putting the manuscript format and discussion in order: such flaws can be fatal. In this essay, I wish to share my observations on the kinds of research errors I have encountered in manuscripts submitted to *J AAPOS*.

Carefully Consider and Frame the Question

The researcher pondering a study should seek to ask one question. For example, what is the rate of nystagmus at age 2 years in children with an estimated gestational age of <31 weeks? The question should be fleshed out with detail and precision; for example, including definitions of conditions, inclusion and exclusion criteria, and minimum length of follow-up. This should be done before the study is performed rather than post hoc. Secondary outcome metrics can be studied, but the study should be powered for the primary outcome: secondary outcome results cannot be considered to have the same significance as the metric the study was designed to assess. Many manuscripts fail to define what they are studying or include a host of outcomes without care; these naturally lead to confusing or invalid results.

The question should also be based on the correct model, or an erroneous conclusion will result. If one studies axial length per year after cataract surgery, the logarithmic nature of the growth of the eye[2] will confound any results: in this example, the researcher should convert the growth curve to a linear model (axial length vs log of age) when age at surgery and length of follow-up no longer skew the results.[1] The same

Author affiliations: Department of Ophthalmology, Bob Wilson Medical Center, San Diego, California

Submitted March 6, 2019.
Revision accepted May 14, 2019.
Published online July 6, 2019.

Correspondence: Scott K. McClatchey, MD, Bob Wilson Medical Center, Ophthalmology Suite 202, 34520 Bob Wilson Drive, San Diego, CA 92134 (email: Scott.K.McClatchey.civ@mail.mil).

J AAPOS 2019;23:191-192.

1091-8531/$36.00
https://doi.org/10.1016/j.jaapos.2019.05.009

need to use the correct model applies to visual acuity (using logMAR for calculations instead of Snellen[3]), astigmatism (using power vectors [M, J0, J45[4]]), and other correlations where a nonlinear condition exists.

Sometimes it is appropriate to mine data for correlations or as a pilot trial to determine the variance in a particular metric, as a basis for a future study. If data is mined in this fashion, if analysis is post hoc, or if the study is a pilot trial, say so, and limit your conclusions accordingly.

Design the Study Well

A carefully framed hypothesis will usually guide the design of the study. This hypothesis should be structured to answer the primary question; all other outcomes are secondary. When multiple outcomes are statistically tested, be sure to correct for multiple comparisons. Many times in ophthalmology, we study an outcome that can happen in two eyes (such as the presence of glaucoma in one eye of a child with bilateral aphakia); this is clustered data. For conditions that can affect both eyes (eg, glaucoma), study one eye (eg, arbitrarily choose the left eye). The design should correct for confounding factors; ideally this is done with randomization and double masking. Multiple regression analysis can also help identify confounding factors in studies that cannot be done prospectively.

For most things we study, follow-up really matters. Patients lost to follow-up are often different from the ones who return—the parent may know something is in need of care, resulting in unsuspected selection bias. Another common problem is including patients with short follow-up when studying a long-term problem; there may simply not have been enough time for these patients to manifest the condition.

Too often authors give excessive credence to "$P = 0.05$" being "statistically significant." This would seem to be valid if all assumptions are correct and there is no bias. However, even prospective randomized clinical trials that get this result ($P = 0.05$) will turn out to be correct only 70% of the time at best.[5] This is a particular problem for retrospective studies, where there is often little uniformity in data collection. For these reasons, and to adequately power the study, the sample size should be set high in order to exceed the minimum required for type I error = 0.05, especially in retrospective studies.

Obtain Data Ethically, Uniformly, according to the Study Design, and with Adequate Follow-up

Studies should follow ethical guidelines, keeping the safety of—and benefit to—the patient foremost. When possible, prospective studies should obtain assent from minor subjects. Institutional review board (IRB) approval should be obtained when necessary; if in doubt, ask! This approval must be obtained before building a cohort: on occasion, a study is labeled as retrospective, but patients were given different interventions for the sake of research prior to obtaining IRB approval.

The data must be carefully and consistently gathered. This is especially important for retrospective studies, where inadequate follow-up can compound the underlying biases. Correct techniques should be used: poor fixation targets, nonstandardized measurements, or ineffective cycloplegia can make data useless. The study should take into account that children change rapidly as they age, performing better on vision and alignment measurements.

The rate of refractive growth (RRG) model has proven useful in predicting future refractions of pseudophakic children (within the limitations of the large variance in RRG). My son now has one eye that is pseudophakic. His refractions of this eye followed the upper standard deviation curve of the RRG model, resulting in a refraction of −8.0 D at age 20; he had PRK and sees well without glasses to this day.

Research based on a carefully considered question, proper design, and high-quality data are worth publishing. When such a study is written for publication, the authors should take care to write it in a way that makes the purpose, design, and results clear to the reviewers, including a descriptive title, a cover letter that lets the editors know the significance of the study, and a manuscript written with good English grammar and structure. The results of good research further our goal of improving the lives of our patients.

References

1. McClatchey SK, Hofmeister EM. The optics of aphakic and pseudophakic eyes in childhood. Surv Ophthalmol 2010;55:174-82.
2. Sampaolesi R, Caruso R. Ocular echometry in the diagnosis of congenital glaucoma. Arch Ophthalmol 1982;100:574-7.
3. Holladay JT. Proper method for calculating average visual acuity. J Refract Surg 1997;13:388-91.
4. Miller JM. Clinical applications of power vectors. Optom Vis Sci 2009; 86:599-602.
5. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values. R Soc Open Sci 2014;1:140216.