# Choice of Analytic Approach for Eye-Specific Outcomes: One Eye or Two?

ANNA KARAKOSTA, MARIA VASSILAKI, SOTIRIS PLAINIS, NAZIK HAG ELFADL, MILTIADIS TSILIMBARIS, AND JOANNA MOSCHANDREAS

• PURPOSE: **To investigate the use of analytic approaches for eye-specific outcomes in ophthalmology publications.**
• DESIGN: **A review of analytic approaches used in original research articles published in ophthalmology journals.**
• METHODS: **All 161 research articles published in 5 ophthalmology journals in the first 2 months of 2008 were considered. Publications were categorized according to analytic approach: 1 eye selected, both eyes contribute, or per-individual outcome. Studies were considered suboptimal when criteria for eye selection were not provided or when measurements from both eyes were included without interocular correlation being considered. Visual impairment prevalence data were used to illustrate analytic approach choices.**
• RESULTS: **Measurements from both eyes were included in 38% of the 112 studies that used statistical inferential techniques. In 31 (74%), there was no mention of possible correlation. Only 7% used statistical methods appropriate for correlated outcomes. In 35 studies (31%), measurements from 1 eye were selected; 31% of these did not provide selection criteria. In 67%, only univariate tests were used. A review of 47 articles published in 2011 produced similar findings. Characteristics of studies were not found to differ according whether the studies were suboptimal. Using a test appropriate for correlated outcomes resulted in a *P* value 3.5 times that obtained ignoring the correlation.**
• CONCLUSIONS: **Between-eye correlation seems not to be assessed commonly in ophthalmology publications, although its knowledge aids the choice of analytic approach when eye-specific variables are of interest. Statistical methods appropriate for correlated ocular outcome data are not being applied widely. (Am J Ophthalmol 2012;153:571–579. © 2012 by Elsevier Inc. All rights reserved.)**

AJO.com Supplemental Material available at AJO.com.
Accepted for publication Aug 16, 2011.
From the Department of Social Medicine, Faculty of Medicine, University of Crete, Heraklion, Crete, Greece (A.K., M.V., N.H.E., J.M.); the Institute of Vision and Optics (IVO), School of Health Sciences, University of Crete, Heraklion, Crete, Greece (S.P., M.T.); and the Department of Ophthalmology, Faculty of Medicine, University of Crete, Heraklion, Crete, Greece (M.T.).
Inquiries to Joanna Moschandreas, Department of Social Medicine, Faculty of Medicine, University of Crete, P. O. Box 2008, Heraklion 71003, Crete, Greece; e-mail: j.moschandreas@med.uoc.gr

MEASUREMENTS OBTAINED FROM BOTH EYES OF an individual often are correlated, that is, measurements obtained in one eye are more likely to be similar to those of the other eye than to ocular measurements from an unrelated person. Standard statistical inferential techniques such as *t* tests, analyses of variance, confidence intervals, and linear regression are valid, however, only under the assumption that observations are independent.[1] Simulation studies have demonstrated that including measurements from both eyes without adjusting for the correlated nature of the data may have a substantial effect on the results.[2,3] Inclusion of measurements from fellow eyes without consideration of their possible correlation usually results in underestimated standard errors, and thus falsely small *P* values and falsely precise confidence intervals, with the magnitude of the problem increasing as the correlation increases.[1,2,4] For many ocular variables, the correlation between fellow eyes has been reported to be high. Correlations of approximately 0.8 have been reported for intraocular pressure (IOP), cup-to-disc ratio, threshold sensitivity, and short-wave length automated perimetry parameter pattern standard deviation data.[5,6]

Between-eye correlation sometimes can be exploited in a paired-eyes study design. In both the United States Diabetic Retinopathy Study[7] and the Glaucoma Laser Trial,[8] for example, one eye was randomized to receive treatment, whereas the fellow eye served as the control. Other analytic approaches include the use of measurements from only 1 eye (selected using defined criteria) or averaging measurements over the 2 eyes.[9,10] Both of these approaches are statistically valid, but are likely to be inefficient, resulting in lower power and less precise estimates than when all available measurements are incorporated. The extent of the loss of statistical information when averaging is likely to be smaller than when only 1 eye is selected. With both approaches, the loss of information is greater when the correlation is low. Correlated ophthalmic data form a subset of clustered data, having maximum cluster size of 2. Family, school, periodontal, and otolaryngologic data also are examples of clustered data. Methods for the modelling of clustered data have been developed in recent years, initially in the field of social and educational statistics,[11] but also in medicine.[12,13] Both univariate and multivariate statistical methods that ac-

0002-9394/$36.00
doi:10.1016/j.ajo.2011.08.032

count for the correlations between fellow eyes have been developed over the past 3 decades.[4,6,14–20]

A 1998 review of 79 *British Journal of Ophthalmology* publications indicated that many studies failed to use all the available data and that a substantial proportion used inappropriate statistical methods.[10] To our knowledge, no review of analytic approaches used in ophthalmology journals has been published since then. The main aim of the present study was to summarize the approaches currently in use to account for correlated measurements between fellow eyes by reviewing publications in 5 ophthalmology journals. A secondary aim is to illustrate the application of 4 different valid analytic approaches for correlated binary outcome data.

The article selection process, the exclusion criteria, and a description of the analytic approaches are presented in the Methods section below, together with the exploratory statistical methods used. A univariate test appropriate for the comparison of correlated binary outcomes and estimation of the intraclass correlation coefficient are also described in this section. The Results section presents the results of the review and the comparison of characteristics according to whether the articles have a suboptimal design. The Results section also contains an illustration of 4 valid analytic approaches for the comparison of correlated binary outcomes (the presence of mild visual impairment) and a commonly used incorrect approach. In the Discussion section, the results are elucidated and discussed in the context of other studies.

## METHODS

**• SELECTION AND CATEGORIZATION OF ARTICLES:** An ophthalmologist (M.T.) selected 2 general and 3 subspecialty ophthalmology journals from the top 50% of the list of 45 ophthalmology journals on the ISI Web of Knowledge (ranked according to their 2007 impact factor). Both general and subspecialty ophthalmology journals were selected to reflect the broad spectrum and the distribution of analytic approaches in these published ophthalmologic articles. The journals selected were *Acta Ophthalmologica*, the *American Journal of Ophthalmology*, the *Journal of Cataract & Refractive Surgery*, *Retina*, and the *Journal of Glaucoma*. All original articles published in the first 2 months of circulation in 2008 were reviewed. *Acta Ophthalmologica* had February as its first month of circulation.

A questionnaire consisting of 30 items was developed, tested, and completed for each published original article. Questions related to study design, study aims, numbers and types of variables considered, units of measurement and analysis, statistical methods used (univariate or multivariate), sample size, and the proportion of missing values were considered. Study design was ascertained according to the *American Journal of Ophthalmology* guidelines. The questionnaires were completed by a doctoral student (A.K.) and a postgraduate physician (N.H.E.) and were checked by a medical statistician (J.M.). All questions related to study design were reviewed by an epidemiologist (M.V.).

Articles in which there was no use of statistical inferential techniques were excluded from further analysis. Subsequently, articles were classified according to the type of study undertaken. Animal studies and laboratory experiments were excluded. Using a scheme similar to that presented by Murdoch and associates, the remaining studies were broadly categorized into the following groups:[10]

1. Studies with outcomes measured at the ocular level in which both eyes are eligible (for at least some subjects), but measurements from 1 eye are chosen for inclusion in the statistical analysis, for example, right or left eye, random selection, dominant eye, better or worse eye, or the first eye with the condition.
2. Studies in which only 1 eye from each subject is eligible for inclusion, for example, the eye that was operated on or the single eye with the condition of interest (rare disease).
3. Studies in which some or all individuals contribute measurements on both eyes in the statistical analysis. This may be because a paired design is used at the ocular level, for example, eyes are randomized so that one receives local treatment and the other does not. However, it may be because information from both eyes is used within each treatment group, either with or without adjustment for the possible correlation.
4. Studies in which ocular outcomes are summarized per individual before analysis, resulting in statistical analysis at the subject level. For example, the average of the separate measurements in each eye is calculated or the results are pooled. This category includes investigations in which information from each eye separately is not of interest. For example, certain conditions are diagnosed at the subject level, but use measurements from both eyes.

It was expected that some studies would have both eye-specific and per-individual outcomes of interest (eg, best-corrected visual acuity). Any study in which at least 1 main outcome was eye specific was classed as belonging to group 1, 2, or 3 as appropriate.

Univariate tests were used to compare the characteristics of studies classed as being methodologically suboptimal with those of the other studies. Methodologically suboptimal studies were those that either (1) included measurements from both eyes without mention, or assessment, of possible interocular correlation, (2) did not provide the number of participants, or (3) did not describe the method used to select the eye chosen for inclusion in the study. Qualitative variables were compared using the chi-square test of independence or the Fisher exact test as appropriate. Sample sizes were compared using the nonparametric Mann–Whitney *U* test. A 5% significance level was chosen.

Subsequently, a review of 47 original articles published in the same 5 journals in February 2011 was undertaken (50% of the total number of articles published in each journal) using exactly the same procedure as for the 2008 publications. Differences in proportions (in 2011 and 2008) were compared using Z tests and Newcombe-Wilson hybrid score confidence intervals.

• **A HYPOTHESIS TEST APPROPRIATE FOR THE COMPARISON OF PROPORTIONS WITH CORRELATED OUTCOMES:** As described in Fleiss and associates, the proportions of eyes with a characteristic can be compared between 2 samples accounting for the possible correlation, using an asymptotic approach with variance inflation factors applied to adjust the variance of the difference in proportions and to calculate an asymptotically normally distributed Z statistic.[21]

The test statistic is:

$$z = \frac{p_1 - p_2}{\sqrt{\mathrm{Var}(p_1 - p_2)}}$$

with

$$\mathrm{Var}(p_1 - p_2) = \frac{p_1 q_1 f_1}{g_1} + \frac{p_2 q_2 f_2}{g_2}$$

where $g_1$ is the number of eyes in group 1, $g_2$ is the number of eyes in group 2, and the variance inflation factor for group 1 is $f_1$, where

$$f_1 = 1 + \left[\left(s_1^2/\overline{n}_1\right) + \left(\overline{n}_1 - 1\right)\right] r_1,$$

where $r_1$ is the intraclass correlation coefficient for group 1 and similarly for group 2.

$s_1^2$ and $s_2^2$ represent the variance of the cluster sizes in groups 1 and 2, respectively, and $\overline{n}_1$ and $\overline{n}_2$ represent the arithmetic mean cluster size. When all clusters are the same size (ie, all individuals contribute 2 eyes), $s_1^2 = s_2^2 = 0$ and $\overline{n}_1 = \overline{n}_2 = 2$ and the variance inflation factor simplifies to $f_1 = 1 + r_1$.

The test statistic Z has a standard normal distribution in large samples, under the null hypothesis $H_0: p_1 = p_2$. If the samples are small, the type I error rates are likely to be inflated.

In the case of binary outcomes, the intraclass correlation coefficient can be estimated using the formula:

$$r = \frac{\sum_{i=1}^{k} \left\{Y_{i+}(Y_{i+} - 1) - 2p(n_i - 1)Y_{i+} + n_i(n_i - 1)p^2\right\}}{\sum_{i=1}^{k} n_i(n_i - 1)p(1 - p)}$$

as given in Fleiss and associates (Equation 15.4, page 443), where $Y_{i+}$ represents the number of eyes with the outcome (and so takes values of 0, 1, or 2) for subject $i = 1, \ldots, K$, whereas p is the total number of eyes with the outcome divided by the total number of eyes and $n_i$ is the cluster size.[21] If all subjects contribute measurements from both eyes, there is a constant cluster size of 2, that is, $n_i = 2$ for all subjects. In this case, the formula reduces to the formula:

$$\frac{P_{++} - p^2}{p(1 - p)}$$

where $P_{++}$ is the proportion of individuals with the finding in both eyes, as given in Murdoch and associates and Thompson.[10,20]

**TABLE 1.** Distribution of the Frequency of Usage of Analytic Approaches in Clinical Investigations with Eye-Specific Outcomes (n = 112)[a]

| Analytic Approach | Frequency (%) |
|---|---|
| 1. One eye from each subject selected for inclusion | 35 (31.2%) |
| Right eye | 3 (2.7%) |
| Random selection | 6 (5.4%) |
| Distance-dominant/non–distance-dominant eye | 1 (0.9%) |
| Clinical criteria: | |
| "Better" or "worse" eye | 4 (3.6%) |
| The first eye to develop the condition | 1 (0.9%) |
| The eye with the condition when present in both eyes, only 1 was selected | 9 (8.0%) |
| Selection criteria not provided | 11 (9.8%) |
| 2. One eye from each subject eligible | 23 (20.5%) |
| The eye that was operated on | 18 (16.1%) |
| Only one eye had the condition, eg, rare disease | 5 (4.5%) |
| 3. Analysis at ocular level, with some or all subjects contributing measurements from both eyes[b] | 42 (37.5%) |
| Paired data (treatment A applied in right eye, treatment B in left eye) | 6 (5.4%) |
| All subjects contribute from both eyes, nonpaired | 3 (2.7%) |
| Some subjects contribute only from 1 eye | 33 (29.5%) |
| 4. Summary of ocular findings per individual | 12 (10.7%) |
| Average taken of measurements in the 2 eyes of each subject | 1 (0.9%) |
| "Pooling" of results of each eye, eg, AMD in at least 1 eye | 11 (9.8%) |

AMD = age-related macular degeneration.

[a]In 5 (4.3%) of 117 studies, there was no information on whether each subject contributed more than 1 ocular measurement, because the number of participants was not stated. The analytic approach taken in these 5 studies could not be ascertained.

[b]The possibility of between-eye correlation was accounted for in 11 of the 42 studies (details are provided in the text).

**TABLE 2.** Characteristics of the 117 Clinical Research Studies According Whether Methodologic Shortcomings Were Apparent

| | Studies Classified as Methodologically Suboptimal (n = 47)[a] | Studies without Apparent Methodologic Limitations (n = 70) | *P* Value |
|---|---|---|---|
| Median no. of subjects (minimum, maximum) | 60 (10, 3654) | 66 (6, 7682) | .335 |
| No. of variables described in the text | | | |
| 1 to 5 | 5 (11%) | 3 (4%) | .265 |
| 6+ | 42 (89%) | 67 (96%) | |
| Are all the variables described in the text eye specific? | | | |
| Yes | 5 (11%) | 2 (3%) | .115 |
| No | 42 (89%) | 68 (97%) | |
| Type of outcomes considered in study | | | |
| Only quantitative | 20 (43%) | 46 (66%) | .022 |
| Only qualitative | 7 (15%) | 10 (14%) | |
| Both | 20 (43%) | 14 (20%) | |
| Study design | | | |
| Prospective | 19 (40%) | 36 (51%) | .500 |
| Retrospective | 21 (45%) | 26 (37%) | |
| Uncertain/neither | 7 (15%) | 8 (11%) | |
| Are there missing values?[b] | | | |
| No | 31 (74%) | 46 (70%) | .645 |
| Yes | 11 (26%) | 20 (30%) | |
| Statistical tests/models applied | | | |
| Only univariate | 34 (72%) | 44 (63%) | .286 |
| Multivariable | 13 (28%) | 26 (37%) | |

[a]The 47 studies classed as being methodologically suboptimal consisted of 31 studies in which the number of eyes included was more than the number of subjects, but there was no assessment of possible interocular correlation, 5 studies in which the number of subjects was not stated, and 11 studies in which 1 eye was selected, but the selection method was not described.

[b]In 5 (11%) of the 47 suboptimal studies and 4 (6%) of the 70 other studies, it was not possible to ascertain whether there were missing data.

## RESULTS

• **REVIEW OF ORIGINAL ARTICLES:** In total, 161 original articles published in 2008 were reviewed. Of these, 31 articles were excluded from further analysis because no statistical inferential techniques were used (19.3%). One hundred eighteen (73.3%) of the remaining 130 articles that described the results of clinical or epidemiologic human investigations were considered further. The other 12 studies comprised 7 animal experiments, 3 laboratory studies, an article assessing the flow rate of vitreous cutters in different viscosity environments, and a study assessing ways to deal with the problem of missing values. One of the 118 investigations was a cross-sectional study examining the attitudes and practices of ophthalmologists in assessing primary angle-closure suspects. This study did not involve ocular measurements, so the article was not considered further. Of the remaining 117 original publications, 63 (54%) were judged to have an interventional design, 52 (44%) were judged to have an observational design, and 2 (2%) had both an interventional and an observational component. Seventeen studies (15%) were randomized controlled trials, and 36 were considered to be interventional case series (31%). There were 16 observational case series (14%), 6 case-control studies (5%), 6 cross-sectional studies (5%), and 2 cohort studies (2%).

The analytical approaches selected in the articles reviewed are presented in Table 1. In 5 (4%) of the 117 articles, the number of participants was not stated, so it could not be ascertained whether these studies used information from both eyes. For example, in 1 study, it was only reported that the measurements were from "2000 consecutive eyes undergoing surgery." These studies could not be classified (in Table 1). Measurements were selected from 1 eye from each subject in 35 (31%) of 112 articles. Clinical criteria were used for the choice of eye in 14 (58%) of the 24 articles in which the selection criteria were reported. Random selection was used in 6 investigations (25%), the right eye was selected in 3 investigations (13%), and the distance-dominant eye was used in 1 study (4%), but in only 2 of these studies was it mentioned that the analyses were repeated using the other eye. The method of eye selection was not reported in 11 (31%) of the 35 studies. For example, the sentence "Only one eye per patient was allowed to enter the study" appears in the methods section of 1 publication, without any further details. The number of eyes simultaneously included in the analyses was greater than the number of subjects in 42 (38%) of 112 studies. Six of these studies had a paired design (14%), 2 used modelling appropriate for correlated data (generalized estimating equations), a between-eye correlation coefficient was calculated in 2 studies (5%), and in 1 of the 42 studies, the statistical analysis was repeated for each eye separately. In the 31 remaining studies, measurements made on both eyes were included in the statistical analyses, without any adjustment for the possible lack of independence of the observations (74% of the 42 articles). The median percentage of patients contributing measurements from both eyes in these 31 studies was 39% (minimum, 2%; maximum, 100%).

In 94% of the articles, inferential statistical techniques were applied for more than 1 outcome (110 of 117 publications). In 1 study, for example, main outcome variables of interest included both total foveal thickness (in micrometers) and visual acuity (logarithm of the minimal angle of resolution units), measured in the same eye. Any study in which at least 1 outcome was eye-specific was classed as having analytic approach 1, 2, or 3 in Table 1, as appropriate. The median number of participants per study was 62 (25th percentile, 30; 75th percentile, 162). Only in 39 investigations (33%) were more than 1 predictor or explanatory variable considered in the statistical analyses; in the remaining 78 (67%), only univariate methods were applied. The characteristics of the 47 studies classed as methodologically suboptimal were not found to differ to any major extent from the other 70 studies (Table 2). The only statistically significant difference (unadjusted for multiple comparisons) was in the type of outcomes considered: the suboptimal studies were more likely to have assessed both quantitative and qualitative outcomes than those using appropriate selection methods ($P$ = 0.022).

Grouping the articles according to their analytic approach proved challenging in 2 cases. In the first, a population-based prospective study, in one main analysis, the outcome variable indicated the appearance of reticular drusen in either eye (so in effect the statistical analysis was at the subject level), and in a separate analysis, the relationship between age and reticular drusen was assessed using methods appropriate for correlated data to investigate eye-specific risk factors influencing the presence of disease.[22] The study was classified in the "All subjects contribute from both eyes, nonpaired" subcategory. In the second publication, different treatments were applied to distance dominant and nondominant eyes, and subsequently outcomes were measured and analyses were undertaken separately in distance-dominant and non–distance-dominant eyes. Binocular vision, patient satisfaction, and spectacle use also were outcomes of interest.[23] The study was classed in the category "One eye from each subject selected for inclusion," subcategory "Distance dominant eye/non–distance-dominant eye selected" in Table 1.

Of the 47 original 2011 articles reviewed, 9 did not involve any statistical inferential techniques and 1 study used rodents. Three (8%) of the 37 remaining original articles could not be considered further because the number of participants was not stated. Measurements were selected from 1 eye from each subject in 14 (41%) of 34 articles. In only 1 study was it mentioned that the analyses were repeated using the other eye. The method of eye selection was not reported in 8 (57%) of the 14 studies. The number of eyes simultaneously included in the analyses was greater than the number of subjects in 15 (44%) of 34 studies. One of these studies had a paired design and 1 used modelling appropriate for correlated data (generalized estimating equations). In the 13 remaining studies, measurements obtained from both eyes were included in the statistical analyses, without any adjustment for the possible lack of independence of the observations (87% of the 15 articles). Further details are provided in the Supplemental Table. The median percentage of patients contributing measurements from both eyes in these 13 studies was 28% (minimum, 2.5%; maximum, 60%). The observed proportion of studies with both eyes contributing measurements was slightly higher in the 2011 articles (44% in 2011, 38% in 2008; $P$ = .62, with the 95% confidence interval indicating a proportion between 11 percentage units lower and 25 percentage units higher in 2011), but the proportion of these that do not use methods appropriate for correlated data also seems to be somewhat higher, although not to a statistically significant extent (87% in 2011, 74% in 2008; $P$ = .51, with the 95% confidence interval indicating a proportion between 14 percentage units lower and 31 percentage units higher in 2011).

No major differences were found in the analytic approaches chosen in 2008 between general and subspecialty ophthalmology journals. The percentage of articles in which correlation was not accounted for was 32% (18 of 57) and 24% (13 of 55) in the general and subspecialty journals, respectively (95% confidence interval [CI] for difference, −9% to 24%; $P$ = .47). When the 2 journal groups were compared using the 2011 data, however, the following differences were apparent: in the subspecialty journals, the number of eyes included in the analysis was greater than the number of subjects in 55% of articles, whereas the corresponding percentage was 25% in the 2 general ophthalmology journals (95% CI for difference, −5% to 54%; $P$ = .19). In contrast, a summary measure (pooling or averaging) was used in 25% of the articles from general journals, but in none of the subspecialty journals (95% CI for difference, 3% to 53%; $P$ = .037, Fisher exact test). The percentage of 2011 articles in which correlation was not accounted for was 25% (3 of 12) and 46% (10 of

**TABLE 3.** An Illustration of the Use of 4 Different Analytic Approaches for the 2-Group Comparison of Correlated Binary Outcome Data: A Comparison of the Prevalence of Mild Visual Impairment in School Children from Stara Zagora, Bulgaria, and Crete, Greece

| Region | Total No. of Eyes Tested | No. of Eyes with VA < 0.5 (%) | Test Statistic | *P* Value |
|---|---|---|---|---|
| Approach 1: inclusion of both eyes and application of a univariate test appropriate for correlated data | | | | |
| Stara Zagora | 322 | 15 (4.7%) | Z = 2.314 | .021 |
| Crete | 602 | 59 (9.8%) | | |
| Approach 2: selection of right eye measurements only and application of a chi-square test | | | | |
| Stara Zagora | 161 | 7 (4.3%) | Chi-square = 5.805 | .016 |
| Crete | 301 | 33 (11.0%) | | |
| Approach 3: selection of left eye measurements only and application of a chi-square test | | | | |
| Stara Zagora | 161 | 8 (5.0%) | Chi-square = 2.071 | .151 |
| Crete | 301 | 26 (8.6%) | | |
| Approach 4: use of clinical critera to define the outcome at the subject level and application of a chi-square test to compare the proportions of children with VA < 0.5 in at least 1 eye[a] | | | | |
| Stara Zagora | 161 | 9 (5.6%) | Chi-square = 6.114 | .013 |
| Crete | 301 | 39 (13.0%) | | |

VA = visual acuity.

[a]In Plainis and associates, 28 children from Crete with missing values for other measurements were excluded from the analysis, so the frequencies differ slightly from those presented here.[23]

22) in general and subspecialty journals, respectively (95% CI for difference, −46% to 13%; *P* = .29, Fisher exact test). Because these proportions are based on small frequencies, however, any inferences are limited.

• **ILLUSTRATION OF ANALYTIC APPROACHES:** Correlated binary outcome data from two independent samples were compared using different analytic approaches. The outcome of interest in the present example was the presence of mild visual impairment from uncorrected refractive error defined as presenting decimal visual acuity less than 0.5 (ie, logarithm of the minimal angle of resolution acuity, > 0.3) with refractive correction, if worn. The aim was to compare the prevalence of mild visual impairment between schoolchildren in 2 geographical areas: Crete, Greece, and Stara Zagora, Bulgaria. Both eyes were examined in all participants. Further details of the study design can be found in Plainis and associates.[24] The intraclass correlation coefficient was estimated for each region separately and was found to be relatively high in both groups: 0.68 in the Greek children and 0.79 in the Bulgarian children.

When both eyes were included in the analysis and the correlation adjusted for appropriately using the Z test described above, the difference in the proportions with mild visual impairment was statistically significant (*P* = .021). The results obtained using this analytic approach and also 3 others (right eye selected, left eye selected, analysis at the level of the individual) are presented in Table 3. In all 4 approaches, a higher proportion of children from Crete had mild visual impairment. The choice of either right or left eye resulted in rejection of the null hypothesis in one case, but not in the other: right eye, *P* = .016; left eye, *P* = .151. The approach taken previously with these data had the subject as the unit of analysis, with a definition "[visual acuity]<0.5 in at least one eye"[24]: the corresponding *P* value was .013 (Table 3). A comparison of the 2 proportions under the incorrect assumption of independence led to a *P* value of .006.

## DISCUSSION

THE FINDINGS OF THE PRESENT REVIEW INDICATE THAT interocular correlation is not assessed frequently or exploited in clinical ophthalmologic studies and that violation of the statistical assumption of independence remains fairly common practice. Although the issue of correlated data in ophthalmic research repeatedly has been raised in the literature, in the present study, it was found that interocular correlation was mentioned, assessed, or adjusted for in a small proportion of the articles in which the statistical analysis included measurements from both eyes simultaneously. The results obtained using the 2011 publications do not indicate any trend toward more appropriate statistical analyses in the past 3 years. In fact, the extent of inappropriate use of statistical methods in recent ophthalmology publications seems to remain similar to that reported in 1998.[10] These findings are not unique to the ophthalmologic literature: in a 2006 review of orthopedic articles, it was found that 42% of the clinical studies considered involved the use of 2 limbs or multiple joints from single individuals without using an appropriate analytic approach or study design (60 of 143 articles reviewed).[25]

A finding of note in the present review is the high proportion of articles in which only univariate techniques were used, commonly *t* tests and 1-way analyses of variance or their nonparametric equivalents, chi-square tests, or tests of correlation between variables. In none of the studies that used only univariate techniques did the analysis account for the possible correlations between fellow eyes. A univariate method for comparing binary outcomes with correlated data was illustrated in the present study. Alternative procedures also have been proposed.[26,27] Univariate nonparametric tests for correlated quantitative outcomes have been described including tests equivalent to the Mann–Whitney *U* test[6,19,28] and the Wilcoxon signed rank test.[29] The between-eye correlation was estimated in only 2 articles; the Pearson correlation coefficient was calculated in both studies. The intraclass correlation coefficient is the method of choice with clustered data, that is, when it is not the relationship between 2 different variables that is of interest, but between pairs of measures (variables) from the same class.[1,21] Various nonparametric and parametric methods for estimating the intraclass correlation coefficient have been published.[30–33] For quantitative clustered data, the usual approach is to calculate the intraclass correlation coefficient using within-subject and between-subject components of variance in a random-effects 1-way analysis of variance framework.[3,31,34] If there is reason to believe that the 2 observations on each subject differ in some systematic way, then a 2-way model is preferred.[31] This could be the case if, for example, measurements from the dominant eye are expected to be higher than measurements from the nondominant eye. A method for calculating the intraclass correlation coefficient for binary data is provided by Fleiss and associates.[29] A statistical model for the intraclass correlation coefficient also has been presented in the context of analysis of clustered neuroscience data.[35]

When measurements are at the ocular level, the extent to which selecting only 1 eye (as opposed to using both simultaneously) affects the results depends on the between-eye correlation, and thus on the variable being assessed. If the intraclass correlation coefficient is close to one, then choosing only 1 eye, or taking the average of both eyes, may not result in a loss of efficiency and may be logical choices, as long as eye-specific predictors are not of interest. It also should be noted that the degree of correlation for a particular characteristic differs according to other characteristics of the population being studied. In a recent study, for example, the agreement between eyes in local response amplitude measurements was weaker for patients with autosomal recessive Stargardt disease (intraclass correlation coefficient, 0.52) than subjects without retinal pathologic features (intraclass correlation coefficient, 0.79).[36]

The analytic approaches presented in Table 3 illustrate that similar results regarding statistical significance are not guaranteed by the use of valid, but different, eye selection and analytic approaches. It should be noted that the low prevalence of mild visual impairment in Stara Zagora is likely to have played a role in the comparisons. For example, a difference of 2 children with mild visual impairment in the left eye (6 instead of 8) would have resulted in a statistically significant result. When analyses are performed using only 1 eye, it is recommended that they are repeated using the other eye. Ideally, similar results are obtained. If not, then there is some doubt about their generalizability. The choice of analytic approach depends on the clinical definition of the outcome of interest. In the mild visual impairment prevalence comparison, for example, both an eye-specific analysis and analysis at the subject level are valid. In the related publication, the analysis was at the subject level, based on the definition of a subject having mild visual impairment if they had decimal visual acuity less than 0.5 in at least 1 eye, according to World Health Organization standards.[24]

There may be cases in which it is not of interest to include both eyes in the statistical analysis, and 1 eye is selected based on clinical criteria (such as those in Table 1). It has been reported in the literature that the use of clinical criteria for the selection of 1 of 2 fellow eyes is likely to result in bias.[10] However, this will not always be the case. For example, Rubin and associates showed that visual acuity in the better-seeing eye can be used as a proxy for binocular acuity when the latter is not available.[37] Thus, within the category of studies with measurements at the ocular level in which 1 eye is chosen for inclusion, particular attention needs to be paid to the research question. Bias may arise with clinical criteria if the results are to be generalized to all eyes and the condition under study in 1 eye is not independent of the other eye. If single eye measurements represent a subject-specific diagnosis, however, as in the binocular acuity proxy and mild visual impairment examples described above, then the aforementioned bias will not be an issue. In the same category of studies (measurements at the ocular level in which 1 eye is chosen for inclusion), the selection of right (or left) eye is considered statistically valid under the assumption that the consideration in question does not favor a particular eye, so the results can be generalized. For example, if high IOP, or conditions associated with high IOP, favor the right eye, it would not be valid to take IOP measurements in the right eye of middle-aged men to estimate average IOP in this age group. Some instances of conditions being more prevalent in either the left or right eye have been reported, for example, early glaucomatous defect has been reported to favor the right eye in Turkish migraine sufferers.[38]

Although describing the study design of each investigation was not a primary focus of the present study, we tried to ascertain the design in each original work, but found that study design was not always described accurately. In addition, the handling of missing data often was not addressed, for example, were all subjects with any missing data excluded? Were imputation techniques used? However, this review focused on the appropriate choice of analytic technique with correlated ophthalmologic data, and not on other statistical issues also relevant to the interpretation of study findings.

A limitation of the present study is that it deals with relatively simple analytic issues, focusing only on outcome measures. It is possible that the exposure may differ within each cluster, that is, between eyes belonging to the same individual. For the binary outcome hypothesis test illustrated above, Fleiss and associates provide details of how to estimate the variance and intraclass correlation coefficient when the exposure varies within clusters (pages 446 to 447).[21] An example of possibly varying exposure within clusters is provided in the design of the randomized, controlled Study for Medical Testing in Cataract Surgery, where eye surgeries were randomized so that the patient either received routine testing or did not before surgery.[39]

In conclusion, the choice of analytic approach in a study should be made on the basis of the research question and the extent of interocular correlation. If eye-specific variables are not of interest, then the most appropriate statistical analysis may be at the level of the individual. In this case, whether fellow-eye data are correlated will not be a concern. For studies involving eye-specific outcomes, the findings of the present review indicate that the use of measurements from only 1 eye remains a common analytic approach. This approach is statistically valid, but is likely to be efficient only when the interocular correlation is very high. The correlation seldom was assessed in the articles reviewed. In addition, a substantial proportion of the studies incorporated measurements from both eyes of some or all subjects without accounting for any possible correlation. Surprisingly, most of the articles reviewed presented results from only univariate analyses. Both parametric and nonparametric univariate methods appropriate for correlated ocular data have been proposed in the literature, although their applicability for each particular study requires careful consideration. To summarize, estimation of the correlation between fellow eyes seems not to be undertaken frequently in studies, even though knowledge of the extent to which the data are correlated aids the decision regarding the most appropriate analytic approach when eye-specific outcomes are of interest.

THE AUTHORS INDICATE NO FINANCIAL SUPPORT OR FINANCIAL CONFLICT OF INTEREST. INVOLVED IN DESIGN OF STUDY (J.M., M.T.); Collection, management, analysis, and interpretation of data (A.K., J.M, M.V., N.H.E., S.P., M.T.); Preparation of manuscript (J.M., A.K., M.V.); and Review and approval of manuscript (J.M., A.K., M.V., S.P., N.H.E., M.T.).

## REFERENCES

1. Sainani K. The importance of accounting for correlated observations. PM R 2010;2(9):858–861.
2. Rosner B. Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. Biometrics 1982;38(1):105–114.
3. Rosner B, Glynn RJ, Lee ML. Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach. Biometrics 2003;59(4):1089–1098.
4. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level modelling. Stat Med 1998;17(11):1261–1291.
5. Katz J, Zeger S, Liang KY. Appropriate statistical methods to account for similarities in binary outcomes between fellow eyes. Invest Ophthalmol Vis Sci 1994;35(5):2461–2465.
6. Harle DE, Evans BJW. Perimetry and migraine-deficits may not implicate glaucoma. Graefes Arch Clin Exp Ophthalmol 2006;244(10):1377–1378.
7. The Diabetic Retinopathy Study Research Group. Photocoagulation treatment of proliferative diabetic retinopathy: the second report from the Diabetic Retinopathy Study. Arch Ophthalmol 1978;85(1):82–106.
8. Beckman H, Maguire M, Wilensky J, Ritch R. The Glaucoma Laser Trial: I. Acute effects of argon laser trabeculectomy on intraocular pressure. Arch Ophthalmol 1989; 107(8):1135–1142.
9. Newcombe EG, Duff GR. Eyes or patients? Traps for the unwary in the statistical analysis of ophthalmological studies. Br J Ophthalmol 1987;71(9):645–646.
10. Murdoch IE, Morris SS, Cousens SN. People and eyes: statistical approaches in ophthalmology. Br J Ophthalmol 1998;82(8):971–973.
11. Goldstein H. Multilevel Statistical Models. 2nd ed. London: Arnold; 1995:1–2.
12. Brown H, Prescott R. Applied Mixed Models in Medicine. 2nd ed. Chichester: John Wiley; 2006:365–367.
13. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomized trials. Br Med J 2004;328(7441):702–708.
14. Donner A. Statistical methods in ophthalmology: an adjusted chi-square approach. Biometrics 1989;45(2):605–611.
15. Glynn RJ, Rosner B. Comparison of alternative regression models for paired binary data. Stat Med 1994;13(10):1023–1036.
16. Glynn RJ, Rosner B. Accounting for the correlation between fellow eyes in regression analysis. Arch Ophthalmol 1992; 110(3):381–387.
17. Glynn RJ, Rosner B. Methods to quantify the relation between disease progression in paired eyes. Am J Epidemiol 2000;151(10):965–974.
18. Glynn RJ, Rosner B. Multiple imputation to estimate the association between eyes in disease progression with interval-censored data. Stat Med 2004;23(21):3307–3318.
19. Rosner B, Glynn RJ, Lee ML. A nonparametric test of observational non-normally distributed ophthalmic data

with eye-specific exposures and outcomes. Ophthalmic Epidemiol 2007;14(4):243–250.
20. Thompson JR. The $\chi^2$ test for data collected on eyes. Br J Ophthalmol 1993;77(2):115–117.
21. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. 3rd ed. New York: John Wiley; 2003: 440–461.
22. Klein R, Meuer SM, Knudtson MD, Ivengar SK, Klein BEK. The epidemiology of retinal reticular drusen. Am J Ophthalmol 2008;145(2):317–326.
23. Assil KK, Chang SH, Bhandarkar SG, Sturm JM, Christian WK. Photopic pupillometry-guided laser in situ keratomileusis of hyperopic presbyopia. J Cataract Refract Surg 2008; 34(2):205–210.
24. Plainis S, Moschandreas J, Nikolitsa P, et al. Myopia and visual acuity impairment: a comparative study of Greek and Bulgarian school children. Ophthalmic Physiol Opt 2009; 29(3):312–320.
25. Bryant D, Havey TC, Roberts R, Guyatt G. How many patients? How many limbs? Analysis of patients or limbs in the orthopaedic literature: a systematic review. J Bone Joint Surg Am 2006;88(1):41–45.
26. Tang ML, Tang NS, Rosner B. Statistical inference for correlated data in ophthalmologic studies. Stat Med 2006; 25(16):2771–2783.
27. Tang NS, Tang Ml, Qiu SF. Testing the equality of proportions for correlated otolaryngologic data. Comput Stat Data Anal 2008;52(2):3719–3729.
28. Lee MLT, Dehling, HG. Generalized two-sample U-statistics for clustered data. Statistica Neerlandica 2005;59(3): 313–323.
29. Rosner B, Glynn RJ, Lee ML T. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006;62(1):185–192.
30. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effect model. Int Stat Rev 1986;54(1):67–82.
31. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1(1):30–46.
32. Müller R. Büttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13(23–24):2465–2476.
33. Ridout MS, Demetrio CGB, Firth D. Estimating intraclass correlation for binary data. Biometrics 1999;55(1):137–148.
34. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Oxford: Blackwell Science; 2002: 704–707.
35. Galbraith S, Daniel JA, Vissel B. A study of clustered data and approaches to its analysis. J Neurosci 2010;30(32): 10601–10608.
36. Tosha C, Gorin MB, Nusinowitz S. Test-retest reliability and inter-ocular symmetry of multi-focal electroretinography in Stargardt disease. Curr Eye Res 2010;35(1): 63–72.
37. Rubin GS, Munoz B, Bandeen-Roche K, West SK. Monocular versus binocular visual acuity as measures of vision impairment and predictors of visual disability. Invest Ophthalmol Vis Sci 2000;41(11):3327–3334.
38. Yenice O, Temel A, Incili B, Tuncer N. Short-wavelength automated perimetry in patients with migraine. Graefes Arch Clin Exp Ophthalmol 2006;244(5):589–595.
39. Katz J, Schein OD, Tielsch JM, et al. Effects of randomizing second eyes in a trial to evaluate preoperative medical testing for cataract surgery. Ophthalmic Epidemiol 1997;4(2): 101–105.

**SUPPLEMENTAL TABLE.** Distribution of the Frequency of Usage of Analytic Approaches in 34 Original Articles Published in 5 Journals in February 2011[a]

| Analytic Approach | Frequency (%) |
|---|---|
| 1. One eye from each subject selected for inclusion | 14 (41%) |
| Right eye selected | 1 (3%) |
| Random selection of eye | 2 (6%) |
| Clinical criteria: | |
| Better or worse eye | 1 (3%) |
| The first eye to develop the condition | 1 (3%) |
| The eye with the condition was selected; when both eyes had the condition, only 1 eye was included | 1 (3%) |
| Selection criteria not provided | 8 (24%) |
| 2. One eye from each subject eligible | 2 (6%) |
| The eye that was operated on | 1 (3%) |
| Only 1 eye had the condition, eg, rare disease | 1 (3%) |
| 3. Analysis at ocular level, with some or all subjects contributing measurements from both eyes[b] | 15 (44%) |
| Paired data (treatment A applied in right eye, treatment B applied in left eye) | 1 (3%) |
| All subjects contribute from both eyes, nonpaired | 0 |
| Some subjects contribute only from 1 eye | 14 (41%) |
| 4. Summary of ocular findings per individual | 3 (9%) |
| Average taken of measurements in the 2 eyes of each subject | 0 |
| Pooling of results of each eye | 3 (9%) |

[a]Three (8%) of the 37 eligible studies were excluded because there was no information on the number of measurements contributed by each subject, because the number of participants was not stated.

[b]The possibility of between-eye correlation was accounted for in 2 of the 15 studies: one study had a paired design, the other used a method appropriate for correlated data.