Measurement of Corneal Epithelial Permeability to Fluorescein
A Repeatability Study

Nancy A. McNamara,* Robert E. Fusaro,*† Richard J. Brand,*† Kenneth A. Polse,* and S. P. Srinivas*

Purpose. Permeability (PdC) to sodium fluorescein (F) is a characteristic of the barrier function of the corneal epithelium. The repeatability of several in vivo fluorophotometric methods used to measure permeability in humans remains largely undocumented. This study examines the repeatability of a method based on topical instillation of a single drop of F for the quantitative assessment of PdC.

Methods. Nine healthy subjects with no history of ocular disease provided 1 (n = 1), 2 (n = 1), or 3 (n = 7) repeated measurements of each eye at successive visits. After making 3 baseline fluorescence scans centrally through the tear film and cornea, 2 µl of 0.35% F were instilled and 10 fluorescence scans were obtained at approximately 2-minute intervals immediately after instillation. Subsequently, the eyes were rinsed three times with nonpreserved saline and four additional scans were performed.

Results. PdC was calculated by dividing the baseline-corrected postrinse stromal fluorescence by the time integral of the tear film fluorescence calculated over the 20-minute exposure period. After applying a logarithmic transformation to the PdC estimates, a mixed-model analysis was used to assess measurement repeatability. On the PdC scale, there is an estimated 95% chance that a second measurement could be as much as 2.88 times higher or 0.35 times lower than a first measurement.

Conclusions. This substantial variability between repeated measurements indicates that the single-drop procedure is unreliable for monitoring individual patient changes. However, with careful sample size planning, this technique can be used in population-based research to compare differences in treatment effects between groups of subjects. Invest Ophthalmol Vis Sci. 1997;38:1830-1839.

The corneal epithelial cells with the associated tight junctions provide ocular protection by forming a resistant barrier to the passage of hydrophilic substances, macromolecules, and pathogens. Several investigators have attempted to quantify this barrier function using fluorophotometry to measure the rate at which a topically applied ophthalmic dye, fluorescein (F), penetrates the cornea. This procedure not only may facilitate our understanding of epithelial barrier function but also may be useful clinically because an increased corneal uptake of F reveals subtle damage to the epithelium.

Adler et al1 were the first to report fluorophotometric studies of fluorescein penetration into the human eye. Over the years, other investigators developed strategies to assess corneal epithelial permeability by applying a single drop of F to the ocular surface and estimating the amount of dye remaining in the eye at a specified time after the application.2-8 Although simple to perform, early studies using this single-drop approach did not consider the turnover of the tear film that continuously reduces the concentration of F available for transport across the epithelium.
To address this problem, deKruijf and coworkers first reported studies using an eyebath to deliver F to the epithelial surface. This method maintains a known concentration of F at the corneal surface; and epithelial permeability ($P_{dc}$) is estimated by dividing the resultant increase in corneal fluorescence by the product of the bath concentration and bathing time. Using this technique, deKruijf estimated the mean ± SD permeability in normal subjects to be 0.038 ± 0.017 nm/second ($n = 86$ eyes). Although the eyebath technique provides an easy method to estimate human in vivo $P_{dc}$, its clinical use is limited because the method of F application is difficult for many subjects to tolerate and may lead to undesirable effects such as fluorescein staining of the skin.

Recognizing the clinical limitations of the eyebath method, Gobbels and coworkers attempted to improve the single-drop technique by accounting for individual tear turnover rates and changing concentration gradients. Using the Fluorotron Master (OcuMetrics, Mountain View, CA), they estimated permeability by dividing the increase in stromal fluorescence from baseline autofluorescence levels 45 minutes after F instillation by the tear film fluorescein concentration integrated during the examination time of 45 minutes. Although this method was an improvement over the earlier single-drop strategy, they failed to rinse the eyes before measuring stromal fluorescence. Trapped F in the conjunctival cul-de-sac can be expected to gradually mix with freshly secreted tears, which leads to artificially elevated estimates of stromal fluorescence. In addition, the retention of F may vary considerably between subjects, between visits, and may be related to the integrity of the anterior ocular surface.

Recently, Joshi and coworkers improved Gobbel's procedure by adding a saline rinse 20 minutes after instilling the initial drop of F. Nonetheless, the average permeability values reported for healthy subjects using this "single-drop with rinse" technique are almost 6 times greater than those found using the eyebath procedure. Assuming the bath technique is accurate, this indicates that some methodological problem may exist with the single-drop method that leads to a substantial overestimate of $P_{dc}$.

Both overestimation of stromal fluorescence and underestimation of tear film fluorescence could inflate permeability estimates and therefore may account for the discrepancy between the eyebath and single-drop results. Overestimation of stromal fluorescence occurs when rinsing is insufficient; underestimation of the tear film fluorescence occurs if the initial dye concentration immediately after topical application of 0.75% F is too high to be accurately measured by the Fluorotron Master. At increased F concentrations, the probability of light being absorbed by fluorophores and converted to thermal energy increases (i.e., quenching) and light absorbed by the outer layers of the solution reduces illumination of the deeper layer (i.e., self-absorption); both mechanisms decrease the net fluorescence emission. Accounting for this decrease in fluorescence emission reduced the discrepancy between the eyebath and single-drop procedures; however, the corrected $P_{dc}$ values were still approximately four times greater than those reported with the bath method.

Although the basic single-drop technique described by Joshi and colleagues has potential clinical applicability, it cannot be used with confidence until certain issues are resolved. For example, can the single-drop technique be further refined to provide permeability values closer to those obtained using the bath technique? If so, what is the within-subject and between-subject variability of permeability estimates obtained on a group of untrained subjects using the single-drop paradigm? Having addressed the issues of validity and repeatability, is it possible to detect subtle alterations in epithelial integrity that would not be apparent during routine slit-lamp examination of the cornea? To approach this final question, the sample size required to detect various changes in $P_{dc}$ must be calculated.

In this study, we used a modified single-drop protocol to obtain repeated measurements of permeability on human volunteers. These data were then used to assess the repeatability of the measurement procedure. Finally, we estimated sample sizes required to detect treatment differences between paired comparison groups of fellow eyes.

**METHODS**

**Solution Preparation**

Previous experiments with the Fluorotron Master fluorophotometer indicated that measured values greater than 2000 ng/ml would be suspect because of the effects of quenching and self-absorption that cause a downward bias in measured fluorescence. By examining the fluorescence concentration values measured shortly after topical administration of the single drop when F concentration is highest, we found that 2 µl of 0.35% F best minimized quenching and self-absorption effects while maintaining a sufficient concentration gradient to provide enough stromal uptake of F for permeability calculations. To prepare a 0.35% isotonic sodium fluorescein solution, we first diluted 5 ml of 10% F (MWt, 376.3 g) to a concentration of 5.65% with 3.85 ml of distilled water, and the resulting solution was then diluted to 0.35% with nonpreserved Unisol saline (Alcon Laboratories, Fort Worth, TX). The final osmolarity of this solution was 303 millios-
mole per kg and pH was 7.4. Given the small drop size, the inherent buffering capacity of tears, and the results of a previous study suggesting that use of phosphate buffers might alter epithelial permeability, no additional buffers were used in the preparation of the solutions.

Instrumentation
We used the commercially available Flurotron Master automated scanning fluorophotometer to perform all the fluorescence scans. By passing tungsten halogen light through two successive blue excitation filters, the device produced a 100 nm bandwidth beam of blue light that coincided with the absorption spectrum of F. The intersection of the excitation beam with the emission beam at 28° forms a 0.8 mm high column with a diamond shaped cross-section of dimensions 50 μm wide × 0.50 mm deep in the horizontal plane. Whereas the maximum depth of the diamond (anterior to posterior) extends approximately 0.50 mm, the bulk of the detected signal originates from a region with a depth of approximately 0.125 mm. This instrument is more fully described elsewhere. Each measurement consisted of a 5- to 8-second scan along the optical axis of the eye, beginning at the tear film and passing through the cornea into the anterior chamber. A computer-driven stepper motor performed 8 steps/mm for a total of 50 measurements, each lasting 100 msec. The instrument counted photons of excited fluorescent light at each step. Because of the limited instrument spatial resolution, the tear film and corneal fluorescence profiles could not be distinguished. Barring problems of quenching and self-absorption, the area under the fluorescence profile obtained from a scan is proportional to the fluorescein mass encountered along the scan path.

Subjects
Nine subjects, aged 23 to 45 years (mean, 30 years) with no history of ocular disease were recruited from the University of California, Berkeley campus. All subjects were free of corneal abnormalities, lid disease and had a normal tear meniscus when examined with white light. Potential subjects taking systemic medications known to affect tear quantity were excluded as were persons with seasonal allergies. Although two of the nine subjects reported daily wear of contact lenses (one hydrogel and one rigid gas permeable), lens wear was discontinued a minimum of 24 hours before permeability measurements and neither subject had experienced an ocular complication within the previous 6 months. Informed consent was obtained after a full description of the procedure. This study observed the tenets of the Declaration of Helsinki and was approved by the University of California, Berkeley, Committee for Protection of Human Subjects.

We obtained three separate measurements of permeability on each eye of seven subjects usually during a 1- to 2-week period. We collected only one measurement on each eye of another subject who was unable to return for additional visits. For an additional subject, the first visit measurements (right and left eye) were excluded from the analysis because of insufficient F instillation as evidenced by examination of the tear film profiles. Therefore, including all right and left eye data, we have 46 permeability measurements.

Measurement Technique
Before F instillation at each visit, three baseline fluorescence scans were taken and averaged for each eye to determine the corneal autofluorescence at the excitation and emission wavelengths of F. A micropipette was used to instill 2 ± 0.001 μl of 0.35% F into the central, lower conjunctival cul-de-sac of the right eye, and the subject was instructed to close and roll the eye to evenly distribute the F to the tear film. A 20-minute timer was started and the right eye was scanned within 45 seconds after dye instillation. The same procedure was then immediately repeated on the left eye. During the remainder of the 20-minute period, alternate measurements between the right and left eye were made approximately every 2 minutes for a total of 10 scans per eye. Actual time of each measurement was recorded by the fluorophotometer. Then both eyes were thoroughly rinsed by gently spraying the entire ocular surface including the superior and inferior conjunctival cul-de-sacs for approximately 1 minute with a steady stream of nonpreserved sterile saline. We repeated this rinsing procedure a total of 3 times on each eye and then measured stromal fluorescein levels 4 times on each eye for the next 10 minutes. Previous laboratory experiments suggested that several thorough rinses were necessary to remove residual F from the tear film and conjunctiva that would otherwise artificially elevate the measured stromal fluorescence values. All measurements for eight subjects were made by one observer (NAM) who also served as the ninth subject and was, therefore, measured by a technician trained in fluorophotometry.

Permeability Estimates
We estimated permeability after topical application of a single drop of F using a simple mathematical model similar to that described in an earlier study. Both the prerinse tear film fluorescence and the resultant increase in stromal fluorescence were measured using fluorophotometry, and the permeability to F was determined using the following model as previously described.
AUC(t) versus time. We then cleaned each plot by averaging the four postrinse AUCs and then subtracting the baseline AUC. Accurately determining the corneal epithelial thickness, \( Q_{dc} \), is difficult and recent values reported in the literature differ by as much as a factor of 4 from what was previously considered to be a normal tear film thickness. Concurrent determination of \( Q_{dc} \) along with \( P_{dc} \) (using presently available methods) would likely interfere with the permeability assessment. We therefore adopted Joshi’s method and assumed \( Q_{dc} \) to be a constant 8 \( \mu \)m for all subjects, as previously reported in the literature.

We estimated \( \int_0^T AUC_d(t) \) dt by first plotting AUC(t) versus time. We then cleaned each plot by deleting the occasional outlying AUC(t) value that exceeded its immediate predecessor or successor by more than 20,000 \([(ng/ml) \times mm]\). This criterion led to the deletion of only 5 of 500 AUC(t) values. For each of the cleaned curves, we used a penalized likelihood approach to fit a cubic B-spline to the 10 prerinse, baseline-corrected AUC(t) values plotted against time of measurement. Then, we extrapolated the left end of the fitted curve to the time of instillation, and numerically integrated the curve to estimate the total F dose applied to the epithelium for the 20-minute exposure period. These calculations were done using functions available in S-Plus version 3.0.

### Statistical Methods

#### Sources of Variation

The repeatability of a measurement procedure can be quantified in various ways depending partly on the intended setting of use. A key ingredient of all repeatability indices is the variability of measurements taken repeatedly on the same entities. We will first describe the potentially important sources of variation that we can estimate given the study design, and then describe the statistical modeling techniques used to estimate them.

If the entities being measured are stable during the period when repeated measurements are obtained, the observed variability all comes from the inherent imprecision in the measurement process. We call this “measurement variation.” The underlying characteristics of many biological phenomena typically also fluctuate. This variation, which is often referred to as “lability,” may also contribute to the observed differences between repeated measurements. With this design, only their combined influence can be estimated because the \( P_{dc} \) measurements cannot be repeated on the same subject in close enough succession to exclude the effect of lability.

In this study, we repeatedly measured both eyes of each subject. Therefore, there are potentially at least two additional sources of random variation in the observed data. Generally, there is some between-subject variation reflecting persistent, unique features of each subject that are consistently expressed in all measurements on both eyes. There may also be persistent differences between eyes of a subject, i.e., between-eye variation.

In studies with repeated measurements on both eyes, measurement variation and lability could potentially occur at the subject level, the eye level, or both. Any measurement variation and lability that operate at the subject level from internal and external time-changing influences that apply equally to both eyes in a measurement session can be called “visit-to-visit variation.” At the eye-specific level, both measure-
ment variation and lability may combine to produce ‘‘within-eye variability’’; for example, if the permeability of fellow eyes varied differently with time, each eye would display differing amounts of within-eye variability. It is within-eye variability that is relevant for paired-eye comparison studies because variability operating at the subject level (i.e., affecting both eyes equally) cancels. The model selection strategies outlined below suggested that the variation observed in the repeated measurements of fellow eyes could be adequately described by between-subject and within-eye components of variance.

In addition to the random variation in the observed \( P_{dc} \) data, there could be systematic differences, on average, between right and left eye measurements or between measurements obtained at the three different visits. Potentially, \( P_{dc} \) measurements could have a different mean for each of the six eye-visit combinations.

**Statistical Modeling.** While accounting for potential eye and visit effects, we estimated the variance components necessary to assess this procedure’s repeatability by fitting mixed-effects analysis of variance models.\(^{29}\)

The reported estimates were obtained using standard maximum likelihood techniques implemented in BMDP Program 5V\(^{30}\); restricted maximum likelihood methods,\(^{31}\) also implemented in Program 5V, provided very similar estimates. We confirmed these computations using SAS PROC MIXED version 6.08 (SAS Institute, Cary, NC).\(^{32}\)

Examing residual plots suggested that a natural log (ln) transformation of the original \( P_{dc} \) data stabilized the within-subject variability and induced greater symmetry. Hence, we estimated the necessary fixed effects and variance components using the ln(\( P_{dc} \)) measurements. For a symmetric distribution, the median and mean coincide and, because median[ln(\( P_{dc} \))] = ln[median(\( P_{dc} \))], an estimate of a mean on the ln(\( P_{dc} \)) scale can be backtransformed through exponentiation to give an estimate of the median \( P_{dc} \).

A mixed-effects model may be specified through its fixed effects and covariance structure.\(^{29}\) We fit a series of mixed models that included the following options for the fixed, or systematic differences: none, eye, visit, eye and visit additively, eye and visit with the possibility of interaction. For each set of fixed effects, we sequentially fit models with one of several covariance structures that each represented plausible patterns of dependency among the repeated measurements made on subjects over eyes and visits. For example, a compound symmetry covariance matrix assumes the following simple structure: (1) all measurements of a single subject’s eyes are correlated through a between-subject component of variance (common to both eyes); and (2) all measurements of the same individual’s eyes are conditionally independent and reflect a single source of within-subject variation. More complex covariance models, in addition, include parameters reflecting potential visit-to-visit and between-eye variance components. We evaluated a series of mixed-models, comprised of the various combinations of fixed effects and covariance structures, by strategies outlined immediately below.

We used two strategies for selecting among the candidate models. The first, more subjective approach, sought the simplest model that adequately represented the data in light of our primary objective, i.e., to estimate within-eye variability. For example, our estimates of visit-to-visit and between-eye variance components were negligible. Thus, we chose the compound symmetry covariance structure over more complex alternatives because it provided comparable estimates of within-eye variability. The second strategy used a more automated model selection procedure based on Akaike’s Information Criterion (AIC),\(^{33}\) which permits comparison of the relative goodness-of-fit of nonnested models. The AIC approach penalizes the likelihood function evaluated at the maximum likelihood estimates of the model parameters by the number of parameters in the model. Both the subjective and automated model selection procedures suggested that a relatively simple model including a fixed left versus right eye effect and a compound symmetry covariance structure was the most appropriate model to use for this analysis.

**RESULTS**

The primary objective was to quantify the variability of repeated \( P_{dc} \) measurements of the same eye. This will enable us to evaluate the appropriateness of our measurement procedure for clinical practice or research. It is important, however, when estimating within-eye variation from data collected repeatedly from fellow eyes, to account for the potential correlation between the fellow eyes and among repeated measurements of the same eye. In addition, a thorough examination of measurement reliability requires exploration of potential systematic differences between left and right eyes, or between visits. In this section, we summarize the principle findings, which we derived through a comprehensive examination of potential sources of random and systematic variation using mixed-effects models as described in the Methods section. First we examine the repeated measurements graphically.

The 46 permeability measurements, after transformation by computing their natural logarithms, are displayed in Figure 1. The repeated measurements plotted by visit, for each right eye in Figure 1a and each left eye in Figure 1b, are connected to indicate which measurements relate to the same eye. Each subject is assigned a unique plotting symbol that is consistently used in both parts of Figure 1. Two subjects with in-
Repeatability of Epithelial Permeability Measurements

FIGURE 1. Repeated measurements of epithelial permeability on the natural log (ln) scale plotted by visit for each right eye (A) and left eye (B). Measurements of the same eye are connected by a solid line, and each subject is assigned a unique plotting symbol that is used consistently in both parts of Figure 1.

complete data are included in the plot: one who was only measured at visit 1 and another whose data from visit 1 were not usable because, as was clear from examination of the data, insufficient F was instilled. The transformation to the natural log scale was used to stabilize the within-eye variability and to make the data more symmetrically distributed.

The lines connecting the three repeated measurements of the same eyes for seven subjects are roughly parallel in both sides of Figure 1. The separation between these lines represents systematic differences between eyes of different subjects (leading to between-eye variation in $P_{dc}$ measurements), which seem to be expressed fairly consistently in fellow eyes. The fluctuations of measurements of the same eye about the connecting line reflect within-eye variability. There does not appear to be a systematic trend in the average $P_{dc}$ across visits, but the right eye $P_{dc}$ appears somewhat lower than the corresponding left eye values, particularly at the third visit. These latter qualitative impressions are quantified in Table 1.

Systematic Differences

In Table 1, the mean ln($P_{dc}$) measurements and their standard errors calculated directly from the data displayed in Figure 1 are presented for each eye and visit. To facilitate comparison with previous studies, we also provided estimates of median $P_{dc}$ on the original scale derived by exponentiating the mean estimates on the ln($P_{dc}$) scale. We confirmed all of the empirical estimates in Table 1 by fitting mixed models that allow for a different mean response at each visit; this modeling approach accounts for the imbalance in the data resulting from the missing data for two of the nine subjects. These model-based estimates were virtually identical to the empirical estimates in Table 1 and, therefore, are not reported.

At visits 1 and 2, the mean ± SE differences in ln($P_{dc}$) between fellow eyes (right–left) were $-0.12 \pm 0.15$ and $-0.15 \pm 0.10$, respectively, compared with $-0.71 \pm 0.24$ at visit 3. This confirms the visual impression of Figures 1A and B that the right $P_{dc}$ was consistently lower than the left, particularly at visit 3. We found no evidence of secular trends in either the $P_{dc}$ measurements or in the within-eye variability that might explain this phenomenon.

Variance Component Estimation

Using approaches outlined in the Statistical Methods section, we chose a mixed model with a systematic

<table>
<thead>
<tr>
<th>Table 1. Estimates of the Mean ± SE ln($P_{dc}$) and Back-Transformed Median $P_{dc}$ Calculated for Right and Left Eyes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visit 1</strong></td>
</tr>
<tr>
<td>Mean ln($P_{dc}$) ± SE (nm/second)</td>
</tr>
<tr>
<td>OD</td>
</tr>
<tr>
<td>OS</td>
</tr>
</tbody>
</table>
with a systematic left versus right eye effect over visits, individual patient. However, in population-based clinical
between two repeated measurements with probability
sated for the limited precision of the individual mea-
The primary goal was to estimate the repeatability of
variance = 0.2223 ± 0.1185.
The model-based estimate of the mean amount by which the left eye exceeded the right eye, which essentially smooths the eye effect over visits, was 0.327 ± 0.108 (two-sided \( P = 0.002 \)) on the ln(\( P_{dc} \)) scale. There are two reasons we do not model the eye \( \times \) visit interaction suggested above by Table 1. First, there are no obvious experimental design strategies implied by this phenomenon because the study periods for subjects were staggered. By not modeling this rather implausible interaction between eye and visit effects, our estimate of the within-eye variance is slightly inflated. However, because the primary objective was to assess repeatability, this is an appropriately conservative approach. Second, as described in the Statistical Methods section, we found that the simpler model, with a systematic left versus right eye effect over visits, fit the data slightly better than models including interactions as evidenced by a larger AIC value.

**Repeatability**
The primary goal was to estimate the repeatability of the \( P_{dc} \) procedure using the variance component estimates. Altman and Bland\(^*\) described an intuitive index of repeatability, which they termed the 95% "limits of agreement" (LoA). Essentially, this represents the maximum difference that is expected to occur between two repeated measurements with probability of 0.95, absent a fundamental change in the entity measured. The LoA can be readily calculated, assuming the differences are approximately normally distributed, as \( \pm 2.0 \times \sqrt{2} \sigma_{we} \), where \( \sigma_{we} \) is the estimated within-eye variance. Using the within-eye variance estimated above as 0.1398, this provides 95% LoA = \( \pm 1.0575 \ln \) units. Translated to the \( P_{dc} \) scale this implies, with a 95% probability, that a second measurement could range from \( \exp^{1.0575} = 2.88 \) to \( \exp^{-1.0575} = 0.35 \) times the first as a result of within-eye variation. This indicates that repeated measurements of an individual eye can be expected to differ substantially without an underlying change in an eye’s true \( P_{dc} \).

The wide LoA suggests that the single-drop procedure is of limited value in assessing the \( P_{dc} \) of an individual patient. However, in population-based clinical research, the objective is typically to compare group means or in oculc research, the differences between paired-eyes in a group of subjects. This procedure may be useful when enough study participants can be recruited to enhance the precision of group mean or paired-eye difference estimates and thereby compensate for the limited precision of the individual measurements. To explore whether group studies using the single-drop technique may be feasible, we have estimated the sample sizes required to detect various percentage changes in mean epithelial permeability.

The estimated sample size required for a paired-eye comparison is shown in Figure 2 where a treatment is randomly assigned to one eye of each subject, whereas the fellow eye serves as the control. The vertical axis represents the number of subjects necessary to detect various average percentage changes in permeability with a power of 0.9 and type I error of 0.05 using a two-sided \( z \)-test, assuming negligible between-subject variation in treatment effect.

The wide LoA suggests that the single-drop procedure is of limited value in assessing the \( P_{dc} \) of an individual patient. However, in population-based clinical research, the objective is typically to compare group means or in ocularesearch, the differences between paired-eyes in a group of subjects. This procedure may be useful when enough study participants can be recruited to enhance the precision of group mean or paired-eye difference estimates and thereby compensate for the limited precision of the individual mean

\[ 2 \text{ is the estimated} \]
\[ \sigma_{we} \text{ of agreement} \] (LoA). Essentially, this represents
\[ \text{of 0.95, absent a fundamental change in the entity}
\[ \text{maximum difference that is expected to occur}
\[ \pm 2.0 \times \sqrt{2} \sigma_{we} \text{, where } \sigma_{we} \text{ is the estimated}
\[ \exp^{1.0575} = 2.88 \text{ to } \exp^{-1.0575} = 0.35 \text{ times the first as a result of within-eye variation.}
\[ \text{indicates that repeated measurements of an individual}
\[ \pm 1.0575 \ln \text{ units. Translated to the } P_{dc} \text{ scale this implies,}
\[ \text{95% LoA} = \pm 1.0575 \ln \text{ units. Translated to the } P_{dc} \text{ scale this implies,}
\[ \text{sample size estimates are based on our estimate of within-}
\[ \text{738 Investigative Ophthalmology & Visual Science, August 1997, Vol. 38, No. 9}
\[ \text{1836 Investigative Ophthalmology & Visual Science, August 1997, Vol. 38, No. 9}
\[ \text{sample size estimates are based on our estimate of within-}
\[ \text{eyepaired-eyes in a group of subjects. This procedure may be useful when enough study participants can be recruited}
\[ \text{single-drop technique may be feasible, we have estimated the sample sizes required to detect various percentage changes in mean epithelial permeability.}
\[ \text{95% pointwise confidence intervals for a paired-eye comparison where a treatment is randomly assigned to one eye of each subject, whereas the fellow eye serves as the control. The vertical axis represents the number of subjects necessary to detect various average percentage changes in permeability with a power of 0.9 and type I error of 0.05 using a two-sided } z \text{-test, assuming negligible between-subject variation in treatment effect.}
\[ \text{sample size estimates are based on our estimate of within-eye variance and its standard error (0.1398 ± 0.0317); we assume that by comparison the between-subject variability in treatment effect is negligible, i.e., that subjects tend to respond similarly to treatment. The vertical axis in Figure 2 represents the number of subjects necessary to detect various average percentage changes in permeability (represented by the horizontal axis) with power of 0.9 and Type I error of 0.05 using a two-sided } z \text{-test. The results in Figure 2 are conservative in that they contemplate investigating a treatment effect that could either be higher or lower than the comparison intervention’s effect. When a one-sided test of the null hypothesis is appropriate,}
\[ \text{95% pointwise confidence intervals for a paired-eye comparison where a treatment is randomly assigned to one eye of each subject, whereas the fellow eye serves as the control. The vertical axis represents the number of subjects necessary to detect various average percentage changes in permeability with a power of 0.9 and type I error of 0.05 using a two-sided } z \text{-test, assuming negligible between-subject variation in treatment effect.}
\[ \text{sample size estimates are based on our estimate of within-eye variance and its standard error (0.1398 ± 0.0317); we assume that by comparison the between-subject variability in treatment effect is negligible, i.e., that subjects tend to respond similarly to treatment. The vertical axis in Figure 2 represents the number of subjects necessary to detect various average percentage changes in permeability (represented by the horizontal axis) with power of 0.9 and Type I error of 0.05 using a two-sided } z \text{-test. The results in Figure 2 are conservative in that they contemplate investigating a treatment effect that could either be higher or lower than the comparison intervention’s effect. When a one-sided test of the null hypothesis is appropriate,}
the curves in Figure 2 would be 0.815 times as high as the ones shown for the two-sided situation.

DISCUSSION

The results indicate that the single-drop method for measuring epithelial permeability lacks sufficient precision for monitoring the \( P_{dc} \) of individual patients in clinical practice because of substantial within-eye variability. One potential source of variability in the estimate of \( \int_0^T AUC_d(t) \) dt may be poor mixing of fluorescein with the tears. Incomplete \( F \) mixing may prevent accurate determination of tear film concentrations during the first 5 to 10 minutes after dye application. This potential source of variability is unique to the single-drop method because the eyebath maintains a known concentration gradient of \( F \) at the corneal surface. If we assume negligible lateral diffusion of \( F \) on penetration of the cornea, however, then poor mixing of tear film \( F \) does not affect the estimate of permeability using the single-drop procedure. This is because the measurements of tear film fluorescence and stromal uptake of \( F \) from the tear film are made at the same corneal position; thus, the mass of \( F \) taken up by the cornea depends only on the mass of \( F \) in the tear film overlying this corneal position during the 20-minute exposure period and not the mass of \( F \) across the entire corneal surface.

In addition to random within-eye variability, we observed that for the same visit, left eye \( P_{dc} \) measurements were systematically greater than right eye measurements. In our protocol, the alternating measurement sequence always started with the right eye, and alignment of the left eye was more difficult for the examiner because the fixation distance from the alignment window of the fluorophotometer to the left eye was greater compared with the right eye. One or both of these factors could account for the systematic between-eye difference because it is unlikely that the right and left eyes truly have different permeabilities.

One participant (represented by the solid octagon in Figures 1a and 1b) showed the highest \( P_{dc} \) measurements at nearly every visit. This participant wore rigid gas permeable contact lenses and other studies involving this participant indicated that the corneal endothelium is also highly permeable to fluorescein. If this subject is atypical, then inclusion unrealistically inflates the estimate of between-subject variability. Re-fitting the same model without this subject's data reduces the between-subject variance estimate (± SE) from 0.2223 ± 0.1185 to 0.1044 ± 0.0659 but has relatively little impact on the estimated within-eye variability, which only decreases from 0.1998 ± 0.0517 to 0.1226 ± 0.0297. With this one high \( P_{dc} \) subject excluded, the estimated between-subject variability is of comparable magnitude to the corresponding estimate of within-eye variability, which suggests that the later may be sufficient to obscure important differences among more typical subjects.

In addition to substantial variability, we note that there is some discrepancy between \( P_{dc} \) values obtained using the single-drop and eyebath procedures. Joshi and coworkers reported \( P_{dc} \) values almost 4 times higher than those obtained using the eyebath method even after correcting for the underestimation of tear film fluorescence that may have occurred as a result of fluorescent quenching and self-absorption. To minimize the problem with fluorescent quenching, we reduced the concentration of fluorescein from 0.75% used in the earlier study to 0.35% and also included a vigorous rinse to more effectively remove any \( F \) trapped in the conjunctival folds and fornices. These modifications resulted in a median \( P_{dc} \) estimate that was less than twice the mean reported by the eyebath studies.

One possible explanation for this remaining discrepancy may be that the epithelial barrier operates differently in the eyebath environment compared with the more natural tear film environment. Also, our working assumption that all subjects have the same \( s \) working assumption that all subjects have the same \( s \) was less than twice the mean reported by the eyebath studies. Although the \( P_{dc} \) values that we report are closer to those obtained using the eyebath method, we cannot conclude that 0.35% \( F \) is the optimal concentration of dye for use in single-drop permeability assessments. In fact, on two subjects, Joshi and coworkers found substantially less variability in \( P_{dc} \) measurements using a 0.75% concentration of \( F \). We note that when tear film fluorescence values obtained immediately after instillation of 0.35% \( F \) dye are low, minimal \( F \) is available for corneal uptake and postrinse values of stromal fluorescence are generally only slightly greater than those measured at baseline, which is likely to degrade the precision of the \( P_{dc} \) measurements. The 0.75% \( F \) concentration used by Joshi and coworkers would be expected to give a higher postrinse signal and this could reduce variability; however, the single
rince protocol may have lead to postrinse fluorescence measurements that were artificially elevated because of residual F in the tear film. Additional studies using higher F concentrations combined with a thorough rinse may be useful to determine the optimal strategy for the single-drop method.

Although additional modification will be necessary before the single-drop technique provides a reliable method for monitoring individual Pdc measurements in clinical practice, the current technique can give useful information for detecting treatment differences between groups of subjects or eyes in population-based research. Figure 2 provides useful guidelines for determining the sample size necessary to detect differences in Pdc between groups of subjects or eyes using the current single-drop protocol. For example, Figure 2 suggests that 60 subjects (1 treatment and 1 control eye per subject) would be necessary to reject the null hypothesis of no difference between the treatment and control with a probability of 0.90, given that the treatment in fact induces a 25% difference in Pdc on the average compared with the control intervention. As the true treatment effect increases, the number of subjects is correspondingly reduced; thus, before estimating the required sample size, one must decide the smallest effect that is clinically important to detect. It was observed in this laboratory that Pdc measurements obtained on six corneas exhibiting submundo amounts of central corneal staining on slit-lamp examination typically were increased approximately 400 to 500% from the average reported in this article. Therefore, this sample size plot suggests that a manageable number of subjects would be required to detect what may be rather modest changes in epithelial barrier function, which would probably not be detectable by slit-lamp examination. If the single-drop assessment of Pdc can detect changes in epithelial integrity that are not revealed by currently available methods (e.g., slit-lamp examination, pachometry), then it may prove to be a useful tool for evaluating epithelial barrier function.

Key Words
corneal epithelial permeability, epithelial barrier function, fluorophotometry, repeatability, sample size

Acknowledgments

The authors thank David Maurice, Department of Ophthalmology, Columbia University, New York, for his scientific advice, and John Fiorillo for his help in preparing the manuscript.

References