

re /@

Ophthalmic pharmaceutical clinical trials: interpretation

Interpretation of ophthalmic pharmaceutical clinical trials involves basic – but not in-depth – understanding of terminology and methods of statistics. Key concepts underlying statistical methodology and factors to consider regarding interpretation of clinical reports are reviewed.

Keywords: association • causation • confounder • correlation • generalizability • *post hoc* analysis • power • sensitivity • significant figures • specificity

Once reviewers of an ophthalmic clinical trial have considered the underlying factors and various aspects of a research study design and found them to be satisfactory (see accompanying article), then attention can then be turned to an interpretation of the trial results. Ultimately, a decision whether to integrate the findings into daily clinical practice must be made. In order to make this determination, a basic understanding of statistical methodology and the mathematical concepts behind these techniques is important in order to interpret the findings of clinical research.

Statistical considerations provide a context for outcomes

If the adage 'what gets counted, counts' is accurate, then the masses of data compiled by clinical investigators each year attest to the importance of medical research. Making sense of the mountains of numbers is only possible through statistical methodology, and debate has opened about the extent of statistical knowledge that ophthalmic providers should have [1–4].

The world of statistics

Broadly, there are two major types of statistics: descriptive and inferential. **Descriptive statistics** are used to 'describe' and include measures of central tendency (mean, median, and mode) and dispersion (range, variance, and standard deviation). These methods are typically taught in introductory statistics courses and do not seem to foreign to clinicians. Numbers derived from descriptive calculations can only relate *how things are right now*.

Inferential Statistics seek to draw conclusions (i.e., make 'inferences') from the numbers. These methods test significance and hypotheses in a sample group and try to reach conclusions beyond the immediate data, for application to the larger population. In other words, numbers derived from inferential methods project **how things might be**. These mathematics allow research to expand into the realm of probability and p values, as such statistics are often based on the normal distribution (the 'Gaussian Distribution' or 'Bell Curve') of clinical findings for biological research.

Contrary to descriptive statistics, inferential methods involve complex mathematical computations. Some of these techniques may be familiar to healthcare providers, but many are not. It should also be emphasized that most inferential statistics require access to the raw data and computer software packages for manipulation.

Randomized controlled trials (RCTs) are generally perceived to provide the highest level of biological evidence in medical

Paul Varner

John J Pershing VAMC, 1500 N Westwood Blvd, Poplar Bluff, MO 63901, USA Tel.: +1 573 448 4303 paul.varner@va.gov



research, and they use inferential statistics as a mathematical basis for testing hypotheses. The data generated by RCTs provide probability of future morbidity and this is only possible with the statistical application of probability. Thus, RCTs and inferential statistics are highly interdependent.

What should clinicians understand in order to be informed reviewers of the medical literature? Significance, random, bias, power, sensitivity, confounding and so on carry different meanings in the mathematical world than in the clinical one, and a few points are worthy of consideration.

Statistical testing

The choice of mathematical methods used in hypothesis-testing inferential statistics depends on the parameter being investigated. Studies can focus on **categorical variables** (qualitative, nonscalar, a few possibilities) and **continuous variables** (quantitative, ranked on scales, many possible values). Detailed examination of specific statistical testing is beyond this basic treatment of the subject, although description of the exact tests used is expected in a 'statistical methods' section of a clinical trial.

Categorical ophthalmic variables include gender, 'mild/moderate/severe' determinations, contact lens exchange schedules (daily, weekly, monthly) and the like. The latter descriptors are easier for patients to understand, but always remain somewhat arbitrary for clinicians (i.e., 'a patient has moderate glaucoma').

Continuous variables, by contrast, more accurately describe a disease. These data include age, intraocular pressure (IOP), refractive error, amount of visual field loss, visual acuity, contrast sensitivity, etc., but are harder for study participants to grasp (i.e., 'glaucoma in this patient is progressing at a rate of 2 dB per year'). It is unfortunate that such specific descriptions have not found wide-spread usage in the vernacular of clinical care – providers and patients alike seem to prefer the former analog explanations versus the latter digital descriptors necessary to statistical analysis.

At times, continuous data are converted to categorical variables ('dichotomization') in order to assign patients to one of two groups. Such ophthalmic binary outcomes involve yes/no distinctions, including whether a person has optic neuropathy or is considered legally blind, and are useful for clinical reporting. Problems associated with dichotomization of data include loss of information regarding the study population, loss of statistical power and the introduction of bias for associations [5].

Specific statistical techniques are applied to both categorical and continuous data based on whether the data are normally distributed (termed 'parametric') or not ('nonparametric'). Normally distributed ophthalmic characteristics such as IOP, cup-to-disk ratios and corneal thickness require parametric tests of the null hypothesis (to be discussed). Chi-squared (χ^2) and t-tests are parametric tests commonly used for ophthalmic clinical trials.

Nonparametric tests of the null hypothesis are used for data that do not have a normal distribution. Degree of diabetic retinopathy, age at time of central retinal venous occlusion and retinal blood flow are examples of ophthalmic variables that are not normally distributed within the population. Wilcoxon and Mann–Whitney U tests are examples of statistical tests that are used to rank data (highest to lowest) prior to mathematical manipulation.

In the final analysis, reviewers of ophthalmic clinical trials can suspect that a t-test is not the best choice in a study of grading diabetic retinopathy, for example. However, beyond that initial suspicion – and even if the 'best choice' statistical method is used – without access to raw data and the computer software package required in order to repeat calculations for themselves, there remains a leap of faith on the part of those not involved in the research that the calculated results reported in published papers accurately reflect the original data.

There has been some assessment of the use of statistical methods in ophthalmic research, suggesting variability of application of the use of statistical techniques in ophthalmic clinical trials [6.7], so some questions of optimal statistical methodology remain to be decided. Since the primary focus of clinicians is to provide patient care, it seems unlikely that they also be expected to understand the full nuances of statistical applications required in order to resolve these issues. Rather, use of full-time statisticians in the peer review process – parallel to clinical review – may help to address these considerations in the years to come.

Null hypothesis

What initially appears to be an easy concept to grasp (i.e., there is no difference between two entities) can quickly become muddled in double-talk explanations of 'accepting/rejecting' or 'proving/disproving' the null hypothesis – or whether it is true/false. Verification of the null hypothesis can even include its own significance testing. Perhaps the complexity deepened as medicine applied a mathematical construct to the clinical environment. The difficulty is intensified for clinicians as the null hypothesis is not always explicitly stated in reports for reviewers to consider.

The null hypothesis is indeed the theoretical basis of statistical methodology in clinical research – without

it there is no way to know what is 'significant' (in the statistical sense). For the purposes of pharmaceutical clinical trials then perhaps the most straightforward understanding of it is this: the null hypothesis is the basic statistical understanding that there is no mathematical difference between study groups. The clinical corollary is that there is no difference between two (or more) treatments.

To the dismay of researchers, mathematicians and clinicians alike, honest mistakes sometimes occur in inferential testing of the null hypothesis. This can be the result of a sampling error, when the sample under study is not representative of the general population (i.e., 'outliers' on the normal distribution), and researchers must account for the possibility of chance findings. There is a litany of statistical procedures and methods available to address the chance of experimental errors.

Significance

If there is no difference between two treatments, but the mathematics indicates that there is (an intervention is falsely claimed to have a positive outcome, but does not – a 'false positive'), then statisticians refer to this as a Type I error (designated as α , *p* value or 'significance'). This leads to claims of a treatment effect that is not there – an unjustified claim. In order to achieve veritable statistical significance (in its mathematical sense), the likelihood of making this kind of error must be quantified. The probability (also in the numerical connotation) of making a Type I error is commonly calculated to be 5% in medical research (*p* = 0.05). This means that, statistically-speaking, one out of every twenty hypothesis tests performed for this level of significance will be in error.

Thus, in medical reporting 'significance' defined as p = 0.05 means 'less than a 5% chance of making a false claim'; it is not merely 'importance,' as in colloquial usage. Type I errors are not the only possible mistakes encountered in statistical research.

Conversely, if there is a difference between two treatments, but the mathematics fails to demonstrate that there is (an intervention falsely claimed to have no effect, but really does – a 'false negative'), then a Type II error (β) occurs. This is a failure to detect a treatment effect. Perhaps less common – or at least less detrimental to the public – than a Type I error, the maximum probability of generating this kind of mistake is set at 20% in medical research (β = 0.20), indicating that there is less than a one in five probability of missing a correct outcome. Type II errors are not directly addressed in statistics of medical research. Rather this is assessed through a mathematical corollary of β , namely power.

Power

Mathematically speaking, power $(1-\beta)$ is inversely related to Type II errors (β) and is the statistical ability to detect the other 80% of results (finding a difference that is really there). In medical research, the power of a study is usually set at 0.80, meaning that there is an 80% chance of detecting the specified treatment effect. It is important to note that the treatment effect is projected by the study investigators.

'Powering a study' means that enough participants (sample size) are included to provide meaningful results. In medical research, power increases with larger sample sizes (indicating that sample results are more likely to reflect the greater, general population) and with larger treatment effects (easier to detect changes mathematically).

Typically, equal treatment groups will maximize statistical power ('balanced randomization'); however, alternative allocation ratios (e.g., 1:2, 1:3, etc.) are often used for subgroup analysis, to improve patient motivation for trial participation (patient motivation to obtain novel therapeutic treatment, although this violates the principle of equipoise on the part of trial participants), or at the request of regulatory boards [8].

So, 'power' in the medical literature means '80% chance of finding a correct outcome' (20% chance of committing a Type II error), and not 'strength' as in common parlance.

It is important to re-emphasize that significance levels (α) and mathematical power $(1-\beta)$ determinations are arbitrary. In medical research, these values have generally been agreed upon as 0.05 and 0.80, respectively. Adherence to these guidelines (or better) – in the absence of an excessive allocation ratio--indicates consensus with the statistical methodology of medical research.

Sensitivity/specificity

Possibly no statistical terms are used more indiscriminately than sensitivity and specificity. Most germane to laboratory testing, these terms have infiltrated the rest of medical literature; however, they are rates (i.e., have numerators and denominators) and remain mathematical constructs of validity [9]. Briefly, 'sensitivity' correctly identifies those with a condition or an agent that really does work (a measure of true positives); whereas 'specificity' properly excludes those who do not have the condition, or identifies an ineffective agent that really does not work (a measure of true negatives).

An ideal outcome is 100% sensitive (identifies all true outcomes) and 100% specific (rules out all negative outcomes). Sometimes ophthalmic reports yield numbers of specificity and sensitivity (Table 1). If not referencing a diagnostic test, then what do these values mean? In such cases, sensitivity is used as a measure of

Clinical Trial Perspective Varner

Table 1. Sensitivity and specificity in selected recent ophthalmic reports [†] .						
	Sensitivity	Specificity				
Digital Imaging in screening diabetic retinopathy [10]						
Observer 1	58.8	69.1				
Observer 2	57.3	68.3				
Frequency doubling technology for glaucomatous patterns [11]						
Cluster N	93.1	82.8				
IFN-γ release assay for tuberculous uveitis [12]						
QuantiFERON-TB gold In-tube	64	99				
• T-SPOT.TB	67	91				
OCT and digital photography for diabetic macular oedema [13]						
England's scheme	72.6	66.8				
Scotland's scheme	59.5	79				
OCT angiography of optic disk perfusion [14]	100	100				
OCT in detecting polypoidal choroidal vasculopathy [15]	94.6	92.9				
Pediatric vision screening [16]						
PlusoptiX photoscreener	98	88				
SureSight autorefractor	95	65				
Pinhole detection of refractive error [17]						
Definition 1	83.9	98.8				
Definition 2	89.7	88.9				
Definition 3	75.9	97.8				
Scheimpflug tomography for subclinical keratoconus [18]						
Range for 11 Indices	62.2–92.7	73.7–90.9				
Self-Triage for ophthalmic emergencies [19]	94.3	76.4				
Vertical topographic thickness Map for keratoconus [20]						
Range of 5 indices	60–100	93–100				
[†] All values reported as percentages. OCT: Optical coherence tomography.						

the rate of true positives (loosely 'presence'); whereas specificity is an indicator of the rate of true negatives (roughly 'accuracy').

Is there a standard of 'acceptable' sensitivity or specificity? Since sensitivity is related to power $(1-\beta)$, then it might be expected that 80% sensitivity would be considered an adequate benchmark. Also, considering specificity as an indicator of true negatives that is related to confidence intervals $(1-\alpha)$, then an acceptable criterion might be expected to be 95%. Indeed these values of sensitivity and specificity have been used as cutoff values in screening programs for diabetic retinopathy [21], and – until further study – perhaps serve as points of departure in determining acceptable values for studies reporting sensitivity and specificity.

The relationships of significance, power, errors, negative/positives, sensitivity/specificity and trial conclusions are summarized in Table 2.

Missing data

100% collection of all data throughout the course of a trial rarely occurs and is the result of voluntary patient withdrawal, mortality, loss to follow-up and other factors – most of which are beyond control of researchers. Nearly all RCTs have missing data with up to 20% of participants having incomplete outcomes [22]. Failure to account for all patients and data can bias results, as lost values might not be missing at random – as is possible if study withdrawals are occurring in systematic ways [23,24]. A common example is when there are excessive side effects from a study medication: more subjects may withdraw from the active treatment arm than from the control arm due to unforeseen, yet veritable adverse reactions.

Absent values can be accounted for by averaging adjacent data, carrying forward the previous value, or by inserting a conservative or computerized value [25]. In

Table 2. Relationship of statistical terms.								
	With respect to null hypothesis	Investigators conclude that the intervention	While the intervention really does	Measured by	Statistically by	Mathematically expressed as		
True positive	Correctly accepted (no difference)	Works	Work	Sensitivity (rate of true positives)	Power	1-β		
False positive	Incorrectly accepted (+ difference)	Works	Not Work	Type I errors, <i>p</i> value	Significance	α		
True negative	Correctly rejected (no difference)	Does not work	Not Work	Specificity (rate of true negatives)	Confidence interval	1-α		
False negative	Incorrectly rejected (+ difference)	Does not work	Work	Type II errors		β		

Notes: Accept null hypothesis (no difference) as true, then there really is no difference between groups. Reject null hypothesis (there is a difference) as true, then there is a difference. Reminder: Rates include numerators and denominators.

order to handle missing data points statistically, researchers employ mathematical constructs to account for missing data points. Box 1 summarizes commonly used techniques for handling missing data techniques. All trials are subject to missing data and no single method is valid in all clinical trial situations [26]. The particular statistical technique employed by the researchers should be clearly stated in the discussion of statistics in the clinical report. As an interesting footnote, increasing a study's sample size does not appear to affect the outcomes of some of these methods for managing missing data [27].

Withdrawal rate

A common source of missing data in clinical research arises from subjects who no longer wish to remain under study. High numbers of study withdrawals (i.e., the 'dropout rate' or attrition rate) can bias results, especially if losses are dissimilar between groups. As a result, clinicians reviewing medical reports need to know: is there an 'acceptable' dropout rate below which study results are not felt to be invalid?

Although much higher for psychiatric medicine (up to 50%) [28], dropout rates across other medical specialties have been reported at 10% or less for the vast majority of RCTs in China [29]. Similar studies in the west have not been undertaken.

Dropout rates have not been systematically studied in ophthalmic research; however, Stewart incidentally [27] noted that 70% of studies involving ocular hypotensive medications had dropout rates of 10% or lower. The Ocular Hypertension Treatment Study and European Glaucoma Prevention Study lost 10% and 25–36% of participants, respectively [30,31]. 10–27% of patients are unable to complete intravitreal anti-VEGF injections studies in neovascular age-related macular degeneration (AMD) [32–35].

Box 1. Statistical techniques to handle missing data.

- As Treated analysis applies patient information based on the treatment received, which may or may not have been the group to which the patient was assigned.
- Intention to Treat (ITT) is the opposite of 'As Treated' analysis. The ITT approach is usually characterized as 'once randomized, always analyzed.' In other words, any deviations from study protocol (noncompliance, withdrawal, receipt of incorrect study medication) are not considered for ITT.
- Per Protocol analysis involves data from all trial participants who followed the trial protocol and completed the
 prescribed time of participation. By default, the per protocol population is smaller than the ITT population,
 and could be liable to biased study results.
- Last Observation Carried Forward substitutes interim data points for 'final' observations for trial participants who withdraw/dropout of a clinical trial. The underlying assumption is that subjects tend to improve throughout the course of the study; therefore, interim data will give a conservative assessment of the treatment effect. This is a 'simple-imputation' method for supplying missing data.
- **Bootstrapping** is a 'multiple-imputation' method that uses resampling techniques to provide more reliability. Used when assumptions related to data are in doubt regarding normal distribution. This method requires the collection of more data believed to be associated with study withdrawals.
- Generalized Estimating Equations are used to analyze correlated outcomes in data that are collected over time (longitudinal). This is especially germane to ophthalmic research where between-eye comparisons are made.

Based on informal review of ophthalmic studies, it might be suspected that 10% would be a ceiling threshold, with dropout rates >20% (loss of one in five participants) likely to be unacceptable. This information should be explicitly stated by authors, and inclusion into a CONSORT diagram is a graphical method to achieve this end [36].

Statistical versus clinical significance

At present, the ophthalmic field lacks consensus standards for what constitutes 'clinically significant' findings for most eye diseases – statistical significance is an attempt to fill that void.

At the conclusion of a report with statistically significant findings, readers are left to decide for themselves whether to 'accept/reject' the findings and apply the findings to their own clinical practices. For positive results, a determination must be made whether the statistically significant conclusions are also clinically significant. As Thompson reminds us, 'We should not accept as fact that a new treatment is superior to another based on statistical analysis alone' [4].

In a sense, eye providers have an important paradigm in this regard: macular edema associated with diabetic retinopathy. Diabetic macular edema has a variety of clinical presentations, but some characteristics have higher risk for visual morbidity than others. Ophthalmic researchers have specifically defined clinically significant macular edema in order to facilitate the study of this condition [37]. Medical researchers often specify 'significant' results by *p* values; however, all clinicians can verify that reported probabilities do not always correlate to the biological field of human health. Thus, differences exist between the significance of mathematics and clinical practice.

The Macular Photocoagulation Study [38] demonstrated a mathematical superiority of focal laser for well-defined, subfoveal choroidal neovascularization in age-related macular degeneration causes (statistical significance); however, the tradeoff of immediate loss of central visual acuity versus gradual long-term vision loss indicated the need for improved treatment modalities (clinical significance) [39-41].

One question that arises is a parallel to the goal of a study. Clinical trial 'goals' or 'purposes' are generally stated in article abstracts, but a helpful question for reviewers to answer for a clinical report is: why is this important clinically? The answer to this initial question may provide a valuable context in which to frame the answer to the question of clinical versus statistical significance.

There is no easy resolution to the problem of statistical versus clinical significance; however, in equivocal cases, whether a provider's own clinical observations match reported findings or not may provide the answer. This is another use of the keen observational skills of astute clinicians.

Significant figures

All scientific measurements are obtained with an understanding of significant figures, and the importance is emphasized to all undergraduates. Rigorous scientific research precludes the reporting of excessive significant figures, which can give a false impression of validity. Briefly, the last significant figure of any measurement is the first digit with uncertainty (the digit that is estimated). Few medical journals specify use of significant figures [42.43], and ophthalmic trials are notorious for indiscriminate reporting of findings that do not follow scientific guidelines for significant figures. The last significant figures for IOP readings, pachymetry and letters gained on visual acuity charting are in the ones column, yet ophthalmic reports frequently include reports of tenths and hundredths.

To adapt Levy's perspicacious observation regarding significant figures [44] to ophthalmic practice, should a reader be concerned that a pachymetry reading of 568.5 μ m is more than 568.4 μ m and less than 568.6 μ m? How does a reported average IOP of 23.5 mmHg compare to 23.4 or 23.6 mmHg? What do 12.3 letters of improvement on an Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity chart or 5.7 intravitreal injections really mean? Finally, if percentage – by definition – indicates *per centum* ('per hundred') then what can we say about a value of 69.7%?

More decimal places does not equate to greater significance of findings. Readers should be aware that the addition of insignificant figures does not improve validity of trial results and calls into question statistical versus clinical significance mentioned above.

Ad hoc/post hoc analysis

In order to ensure study validity, all primary and secondary study outcomes should be clearly specified in design protocol prior to enrollment of the first participant. Altering research parameters during the study represents a failure to follow the scientific method and can give the impression of altering results to fit the study hypothesis. *Ad hoc* and *post hoc* reports occasionally appear in the ophthalmic literature.

Ad hoc (Latin: for this) analysis is a one-time event (statistical test, data summary, analytical process), intended to answer a specific question. More common in the business world or for committee formation, this reporting is rarely used in medicine. Ongoing safety monitoring typically occurs throughout clinical research studies and is not conducted on an *ad hoc* basis (e.g., safety monitoring during the Ischemic Optic Neuropathy Decompression Trial discovered possible harm during ongoing review; this was not an *ad hoc* analysis [45]).

Post hoc (Latin: after this) analysis occurs after the study has concluded and all data have been collected. In its true form *post hoc* analysis is used to study data trends that were not foreseeable *a priori*. However, depending on the content of the *post hoc* report, these findings can take the form of 'data dredging' (manipulation of data to prove a point [46]) or permutations intended to make findings more robust.

Occasionally, researchers discover incidental, interesting clinical findings within a study cohort that are unrelated to the initial study protocols. Subsequent reports can occur years after the original study aims are achieved but are believed to provide additional clinical guidance. Patient accrual for the ETDRS began in 1979 and was concluded by 1985 with follow-up through 1989 [47]; however, the final ETDRS numbered report (#25) discussing visual acuity results after cataract surgery in patients with diabetic retinopathy was not published until 1999 [48]. Although ETDRS was piloted 'to evaluate argon laser photocoagulation and aspirin treatment in the management of patients with nonproliferative or early proliferative diabetic retinopathy' [47], adjunct findings were reported to help further the clinical management of patients.

Extensive *post hoc* publications for ophthalmic trials are now common. The Collaborative Ocular Melanoma Study produced 28 numbered and at least 43 other papers between 1985 and 2006 [49]. The Collaborative Longitudinal Evaluation of Keratoconus study produced 39 papers between 1996 and 2011 [50]. Such extensive publication lists have not been systematically studied; however, due the violation of the scientific method, *post hoc* findings probably require independent, prospective confirmation and – in the absence of highly-compelling evidence – should not be used to alter clinical practice.

Interpretation

Once readers sense that a clinical trial has been performed under sound conditions of biological study, and that the reported results truly reflect a study's purposes, then interpretation with respect to clinical application must follow. Put simply: should a clinician integrate these results into daily practice? To answer this question, other factors must be considered.

Association (correlation)

An association is a situation 'in which two attributes occur together either more or less often than expected by chance' [51]. One-to-one associations are infrequent in medicine [52], and – based on current medial knowledge – are more likely to be discovered in acute diseases (consider Koch's Postulates), where the exposure to disease pathway is easier to recognize.

Although sometimes used interchangeably with association, correlation is a more specific statistical comparison of two variables and indicates how well one characteristic can be used to predict another. Thus two entities can be associated in time or place by observation, but only correlated mathematically. Correlation coefficients are often used as a mathematical measure of this indicator. Clinically, there is a high degree of correlation among glaucomatous optic neuropathy, visual field defects and nerve fiber layer defects, although the exact relationship varies among research groups. (This is important to remember in ophthalmic clinical trials of glaucoma.) Results from the study of chronic diseases must use correlation mathematics to tease out multifactorial characteristics.

Astute scientific observations by providers lead to the consideration of associations, which in turn guide medical research. The clinical observations that more ophthalmic complications were observed in soft contact lens extended wearers than other materials/wearing schedules led to the reporting of important preliminary data that quantified clinical observations [53]. These initial findings were confirmed and this association is now widely acknowledged. Reported associations should not be ignored; neither should they be immediately integrated into 'clinical experience,' as the strength of coincidence can alter clinical perceptions.

Coincidence

A coincidence is 'a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection' [54]. What observers subjectively classify as 'meaningful coincidences' occur more often than chance allows [55] and are powerful phenomena.

Ophthalmic providers are sometimes amazed to see back-to-back, nearly identical cases of end-stage glaucoma, three central retinal venous occlusions within a short period of time, or – in the case of an American veterans' medical center – three consecutive female patients. Yet, over the course of long-term clinic care, these are normal variations in the statistics of chance in time and place, brought to our attention by unpredictable events or motivations.

The power of coincidence has been poorly studied in medical and ophthalmic research; however, Carel *et al.* reported spurious associations between IOP and other health parameters [56], and there is a single report of optic neuritis in which an initial association was later felt to be coincidental [57].



Statisticians also recognize the need to study coincidence mathematically [54]. This is especially important as coincidences can often be associated and correlated, yet be totally meaningless – the so-called 'spurious correlations.' Although US spending on science, space and technology may highly correlate (in the fully statistical sense) with the suicide rate by hanging, strangulation and suffocation in recent times [58], few would argue that these wholly unrelated characteristics could be used to predict one another over longer periods of time.

A final note about coincidence: it increases with *n*. As the sample size increases, the margin of error decreases, but coincidences become more likely (the *Law of Very Large Numbers* of probability theory).

Confounders

A medical confounder is an epidemiological case of mistaken identity. Without careful study design a confounder can be assigned causation instead of the veritable antecedent(s). Identification of a variable as a true confounder must satisfy three well-established criteria: association with the exposure, association with the outcome and not be in the causal pathway between the two (exposure and outcome) [59].

While this classic epidemiological model may work well in acute infectious diseases with single-antecedent diseases and an isolated confounder, chronic diseases do not appear to follow this pathway. Perhaps a multifactorial model for chronic diseases may be considered. Figure 1 depicts such a non-linear construct with multiple disease risk factors and the possibility of several confounders.

Unfortunately, confounding is often reported in terms commensurate with 'confusion' than with the path of causation. Results that can cloud results and lead to possible misinterpretation of clinical results have been reported for a few ophthalmic conditions [60–63]; however, these seem to be suspicions of clinical confusion versus actual confounding in the epidemiological sense of the term.

For a case of possible, true ophthalmic confounding, consider IOP and glaucoma. Given lack of definitive causation for IOP and Primary Open-Angle Glaucoma (POAG) to date, should IOP instead be considered a confounder? Following the model in Figure 1, the lack of biological plausibility discounts the likelihood of a relationship between environmental and lifestyle factors and this optic neuropathy. Population studies certainly support the association of age with POAG, and perhaps family history could be reinvestigated to control for cases of ocular hypertension; however, the exact exposures/factors that result in the optic neuropathy of POAG remain unknown. So the association of IOP with other exposure(s) or factors remains indeterminate. The outcome of POAG is optic neuropathy, which is correlated with visual field loss and central visual acuity loss in some cases (the natural history of the disease does not indicate eventual acuity loss for all patients with POAG), but that effect currently remains without cause.

Lack of definitive proof of causation suggests that IOP might not be in the chain of causation for POAG (secondary forms notwithstanding). Perhaps other, heretofore unrecognized clinical characteristics indicating subtypes are the true confounders. However, this is strictly conjectural and remains to be determined.

Reduction of confounding is achieved via study design (matching or randomization) and statistical methodology, especially multivariable analysis [64], and has been recognized in the ophthalmic literature [65,66]. Use of these study designs and statistical techniques should provide a measure of control for confounding.

Regression to the mean

Regression to the mean can act as a confounder. In this case the confounder is a variation in measurement, the results of which can be misinterpreted to represent the results of an intervention.

This commonly used phrase is another statistical phenomenon used to identify outlying data. Measurement of any variable at a single point in time can fall anywhere along the normal distribution for that characteristic. A single measurement – by chance – may reflect an extreme value; whereas the next measurement of that same variable may give a different result – one closer to the 'mean' for that characteristic.

Regression to the mean is a purely statistical term and not some type of physiological process [67,68]. It is an artifact that can affect the results of clinical results. This phenomenon is widely recognized in ophthalmic studies of uniocular trials for ocular hypotensive pharmaceutical agents, although its physiological nature remains unknown [67]. Use of single visual acuity measurements can make surgical outcomes appear better than they are due to the effects of regression to the mean [69]. Likewise, the somewhat subjective results obtained for automated perimetry have been recognized to show similar trends [70].

It should be pointed out that regression to the mean does *not* predict future events – in other words, a statistically high measurement is not always followed by a statistically low datum to somehow 'even out' the data. This is a form of the *Gambler's Fallacy*, a distortion of the *Law of Averages*.

Regression to the mean is an inherent problem for study inclusion when patients self-report symptomatology – if patients with worse symptoms are included for study during exacerbation of findings (i.e., when they

Ophthalmic pharmaceutical clinical trials: interpretation Clinical Trial Perspective

are 'at their worst'). The natural course of waxing and waning conditions may parallel treatment, and natural remission of symptoms can be misinterpreted as treatment effect, and even calculated as such, thus leading to 'improvement' attributed to the study medication and not regression to the mean.

Regression to the mean is typically controlled in clinical studies by multiple baseline measurements prior to trial inclusion and/or statistical measures to reduce its effect [71].

Generalizability

This is the concept that the results obtained from a study sample (test group) can be extended to represent the broader population (other patients). Also known as 'external validity,' generalizability represents clinical application of research trial results to daily practice. This is especially important where geographic or ethnic variations of populations must be considered. In other words, clinicians wish to know if the results from a trial on one continent can be applied to patients in other areas of the globe.

Occasionally, this is indirectly provided, as in the case of intravitreal injections for CNVM for AMD. The results of the European IVAN trial [72] paralleled those of the American CATT study [73]; therefore, the results of those studies confirm one another and have broader credibility for clinicians.

Conversely, the results of the European Glaucoma Prevention Study [31] did *not* confirm those reported for the Ocular Hypertension Treatment Study conducted in the USA [30]. In the face of conflicting data, providers are faced with two considerations: were the study samples disparate or were the results of one trial subject to misinterpretation (i.e., clouded by the factors discussed in this paper)? Fitting a line to two data points is often problematic – perhaps a third trial would provide definitive clinical information.

Sometimes research trial results seem at odds with observed clinical cases. Non-generalizability has not been systematically studied in medicine or ophthalmology; however, multi-centered studies with larger sample size, use of one eye per subject and inclusion criteria that are not too exclusive (see accompanying paper for further discussion of these important influences) are factors that favor generalizability of clinical trial findings to other practice populations.

Occasional disparities in generalizability might be due in part to Hawthorne Effects – named for the site of the investigations, not a specific investigator.

Hawthorne effect

Now more than a century after the original studies at the Hawthorne plant of the Western Electric Company in



Figure 1. Confounders in chronic disease.

Chicago, this phrase is used arbitrarily [74]. It is not synonymous with Placebo Effect, but rather most accurately describes the phenomenon whereby improved performance (productivity at the plant in the case of the original studies) is the result of attention of observation (special attention and privileges bestowed upon the workers).

Clinical research in medicine views the Hawthorne Effect as 'the mere awareness [that] being under observation can alter the way in which a person behaves' [75]. It might be expected that participants in medical research would be more likely to take study medications just because they are under observation and being instructed to do so; yet the importance of this in ophthalmic research remains unclear as demonstrated by conflicting evidence of compliance with ocular hypotensive therapy during clinical study [76].

In the effort to prevent loss of data during clinical trials, study monitors encourage participants to keep follow-up appointments ('Intent to Attend' assessments) and to comply with instructions or treatments throughout the course of the trial. As a result, there are more opportunities for patient education and behavior reinforcement within clinical trials than are available in typical patient care. So, though Hawthorne Effects may not be directly observed during clinical trials, they may provide an explanation for disparities between reported results and clinical observations of the same treatments.

Causation

When associations lead to statistically relevant correlations, which are felt to be free of coincidence and the effects of confounding, and when published results can be externally validated, then the possible links of causation may be explored. Causality is established when one event consistently leads to the occurrence of a second event. The mantras 'association is not causation' and 'correlation does not imply causation' are common in statistical discussions. Bradford Hill's seminal paper [52] on the topic lists ten factors that must all be satisfied prior to assigning causation to an association. Box 2 briefly reviews these criteria.

As 'an extraordinary claim requires extraordinary proof' [77], so must biological evidence provide a solid medical foundation on which to frame causation. Hence, the ten criteria to satisfy medical causation. Perhaps part of the explanation for such lengthy times for research innovations to find widespread acceptance in clinical practice [78] is that providers wish to see things for themselves.

It seems that our perceptions of causation have been tempered by acute diseases – especially infectious diseases – which are mostly treatable in short dosing regimens (or more often, the natural courses are self-limited). In these cases, cause and effect are far simpler to demonstrate as are responses to treatment. Acute events – by definition – are short in duration and also have 'cures.' Contrast these features to pathways for chronic diseases.

A summation of statistics for the nonstatistician

Taking all of the above points into consideration, it seems counterproductive to expect clinicians to be able to replicate the nuances of statistical methodology. Rather, in order to successfully review ophthalmic clinical trial design and to interpret the results from those trials, the role of non-researchers remains to ensure that standard protocols are followed. After all, they are the ones instituting findings at the grass roots level.

The ability to review the ophthalmic literature requires transparency of reporting procedures and findings. This is not a novel concept in medicine [25,79] or in ophthalmic research [80], and goes a long way toward validating reports. Readers should be able to easily follow the entire experimental process involved in a clinical trial: rationale for conducting the trial, the study's stated purpose, criteria for subject involvement (projection of participant numbers, inclusion/ exclusion criteria, randomization schemes), primary/ secondary end points, full accountability for subjects and data throughout the duration of the trial, statistical methods regarding how the raw data were manipulated, how conclusions were drawn, and whether author recommendations support the conclusions.

Reports missing key information leave unanswered question, which only reviewers of the literature can ask. Follow-up letters to journal editors sometimes provide important opportunities for authors to clarify ambiguous points and are part of the process of medical reporting. In fine, clinicians remain the check and balance on the reporting portion of the experimental process.

Conclusion

To successfully review ophthalmic literature, clinicians must have some familiarity with the basics of terminology and statistical methodology, albeit not extensively. Misunderstanding of the terms associated with statistics seems to have created a chasm between the clinicians and the mathematicians, but these can be readily ameliorated.

Clinicians need not be statisticians in order to review medical literature, yet critical interpretation of published findings is required to maintain the integrity of the scientific method clinical research. By understanding the factors providing the context of this process and applying a few concepts to the interpretation of published reports – perhaps even following a standardized checklist for determination of completeness – readers can best evaluate the data and determine whether or not the stated conclusions will benefit patients and represent a change in medical practice patterns.

Future perspective

Increased awareness by clinicians of the importance of hypothesis testing via inferential statistics will continue

Box 2. The case for causation.

- Strength Greater numbers have greater plausibility.
- Consistency Has the association been observed by different researchers in different places in different times?
- Specificity One-to-one relationships are useful in acute settings; a cluster of findings may be appropriate in chronic diseases.
- Temporality Antecedents must precede outcomes in time.
- Dose-response curve Does the rate of the disease increase with more exposure to the associated factor?
- Biological Plausibility Does the association make logical sense in the chain of causation? Understanding may change, based on new research findings over time.
- Coherence Does interpretation of the data follow a rational sequence of events?
- Experiment Can the association be put to the test and withstand scrutiny under new conditions?
- Analogy Do other chains of causation exist from which we can make scientific comparisons?
- Tests of significance Do the numbers back up all of the above factors?

Information taken from [52].

Ophthalmic pharmaceutical clinical trials: interpretation Clinical Trial Perspective

to develop in parallel with ongoing improvement in clinical trial design. It is expected that statistical methodology will continue to mature during the years to come as 'best practices' emerge for the handling of data. The standardized reporting of statistical methodology in the medical literature will also improve interpretation of clinical reports as transparency of clinical research reporting advances.

Acknowledgement

The author would like to thank A Varner for her assistance in the preparation of **Figure 1**.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

Disclaimer

Opinions expressed in this article are those of the author alone, and not of the US Department of Veterans Affairs.

Executive summary

General considerations for statistical review

- Clinicians are not expected to be statisticians.
- 'Statistical methods' sections in clinical reports should be specific regarding which tests are used.
- Acceptable significance level (α) is 0.05 or less.
- Acceptable power (1 β) is 0.80 or greater.
- No excessive allocation ratios.
- Benchmark for sensitivity: 80%.
- Benchmark for specificity: 95%.
- Technique for handling missing data points should be specified.
- Threshold dropout rate $\leq 10\%$; >20% likely unacceptable.
- Statistical significance should be verified by clinical experience. Is this result important clinically?
- Findings should be reported only in significant figures.
- Most post hoc findings require independent, prospective confirmation.

Interpretation

- An association is made by observation in time and place. Possible associations should not be ignored, but not allowed to fit conclusions to hypotheses.
- Correlation is a statistical measure of how well two variables can be used to predict another.
- Coincidence is a powerful phenomenon. Some associations and correlations can be spurious.
- Confounding does not mean 'confusion.' Chronic diseases may have multiple confounders.
- Study design (matching or randomization) and statistical methodology (multivariable analysis) are used to control for confounders.
- Regression to the mean is a statistical term, not a physiological process, and can be controlled for by multiple baseline measurements.
- Generalizability (external validity) is improved with multicentered studies, larger sample sizes, use of one eye
 per subject and appropriate inclusion/exclusion criteria.
- The Hawthorne Effect (improved performance is the result of attention during observation, not other factors) may also play a part in nongeneralizability of research trial results to daily clinical practice.
- Causation must satisfy many criteria: strength, consistency, specificity, temporality, dose-response curve, biological plausibility, coherence, experiment, analogy and mathematical tests of significance.
- The entire experimental process should be transparent and easy to follow.

References

Papers of special note have been highlighted as:

- of interest; •• of considerable interest.
- Morse AR, Albert DM. Effectively implementing evidencebased medicine. *Ophthalmology* 120, 2163–2165 (2013).
- Czyz CN. Re: Morse *et al.* effectively implemented evidencebased medicine. *Ophthalmology* 121, e47–e48 (2014).
- 3 Lisboa R, Meira-Freitas D, Tatham AJ *et al.* Use of statistical analyses in the ophthalmic literature. *Ophthalmology* 121, 1317–1321 (2014).
- 4 Thompson JT. Three kinds of lies. *Ophthalmology* 121, 1315–1316 (2014).
- Balanced overview of issue of statistical knowledge for clinicians.
- 5 Cumberland PH, Czanner G, Bunce C *et al.* Ophthalmic statistics note: the perils of dichotomizing continuous variables. *Br. J. Ophthalmol.* 98, 841–843 (2014).
- 6 Murdoch IE, Morris SS, Cousens SN. People and eyes: statistical approaches in ophthalmology. *Br. J. Ophthalmol.* 82, 971–973 (1998).

Clinical Trial Perspective Varner

- 7 Lee CF, Cheng ACO, Fong DYT. Eyes or subjects: are ophthalmic randomized controlled trials properly designed and analysed? *Ophthalmology* 119, 869–872 (2012).
- 8 Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. *J. Med. Ethics.* 24, 401–408 (1998).
- 9 Parikh R, Mathai N, Parikh S et al. Understanding and using sensitivity, specificity and predictive values. Indian J. Ophthalmol. 56, 45–50 (2008).
- 10 Gupta V, Bansal R, Gupta A, Bhansali A. Sensitivity and specificity of nonmydriatic digital imaging in screening diabetic retinopathy in Indian eyes. *Indian J. Ophthalmol.* 62, 851–856 (2014).
- 11 Bowd C, Weinreb RN, Balasubramanian M *et al.* Glaucomatous patterns in Frequency Doubling Technology (FDT) perimetry data identified by unsupervised machine learning classifiers. *PLoS ONE* 9, e85941 (2014).
- 12 Ang M, Wong WL, Kiew SY *et al.* Prospective head-tohead study comparing 2 commercial interferon gamma release assays for the diagnosis of tuberculous uveitis. *Am. J. Ophthalmol.* 157, 1306–1314 (2014).
- 13 Prescott G, Sharp P, Goatman K et al. Improving the costeffectiveness of photographic screening for diabetic macular oedema: a prospective, multi-centre, UK study. Br. J. Ophthalmol. 98, 1042–1049 (2014).
- 14 Jia Y, Wei I, Wang X *et al.* Optical coherence tomography angiography of optic disc perfusion in glaucoma. *Ophthalmology* 121, 1322–1332 (2014).
- 15 De Salvo G, Vaz-Pereira S, Keane PA *et al.* Sensitivity and specificity of spectral-domain optical coherence tomography in detecting idiopathic polypoidal choroidal vasculopathy. *Am. J. Ophthalmol.* doi:10.1016/j.ajo.2014.08.025 (2014) (Epub ahead of print).
- 16 Silbert D, Matta N, Tian J, Singman E. Comparing the SureSight autorefractor and the plusoptiX photoscreener for pediatric vision screening. *Strabismus* 22, 64–67 (2014).
- 17 Marmamula S, Keeffe JE, Narsaiah S *et al.* Population-based assessment of sensitivity and specificity of a pinhole for detection of significant refractive errors in the community. *Clin. Exp. Optom.* doi:10.1111/cxo.12172 (2014) (Epub ahead of print).
- 18 Ruiseñor Vázquez PR, Galletti JD, Minguez N et al. Pentacam Scheimpflug tomography findings in topographically normal patients and subclinical keratoconus cases. Am. J. Ophthalmol. 158, 32–40 (2014).
- 19 Eijk Es, Busschbach JJ, Monteban H *et al.* Towards patient self-triage in the ophthalmic emergency department: sensitivity and specificity of a self-triage instrument. *Acta Ophthalmol.* doi:10.1111/aos.12342 (2014) (Epub ahead of print).
- 20 Abou Shousha M, Perez VL, Fraga Santini Canto AP *et al.* The use of Bowman's layer vertical topographic thickness map in the diagnosis of keratoconus. *Ophthalmology* 121, 988–993 (2014).
- 21 Saari JM, Summanen P, Kivelä T, Saari KM. Sensitivity and specificity of digital retinal images in grading diabetic retinopathy. *Acta Ophthalmol. Scand.* 82, 126–130 (2004).

- 22 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin. Trials* 1, 368–376 (2004).
- Important discussion of problem of missing data.
- 23 Buehl W, Heinzl H, Mittleboeck M, Findl O. Statistical problems caused by missing data resulting from neodymium: YAG laser capsulotomies in long-term posterior capsule opacification studies: problem identification and possible solutions. J. Cataract Refract. Surg. 34, 268–273 (2008).
- 24 Belin TR. Missing data: what a little can do, and what researchers can do in response. *Am. J. Ophthalmol.* 148, 820–822 (2009).
- 25 Stanley K. Evaluation of randomized clinical trials. *Circulation* 115, 1819–1822 (2007).
- 26 Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Stat. Med.* 20, 3931–3946 (2001).
- 27 Stewart WC, Jackson AL, Jenkins JN. Dropout rates for intent-to-treat and per protocol analysis. *Am. J. Ophthalmol.* 137, 639–645 (2004).
- 28 Santana L, Fontenelle LF. A review of studies concerning treatment adherence of patients with anxiety disorders. *Patient Prefer. Adherence* 5, 427–439 (2011).
- 29 Zhang D, Yin P, Freemantle N *et al.* As assessment of the quality of randomised controlled trials conducted in China. *Trials* 9, 22 (2008).
- 30 Kass MA, Heuer DK, Higginbotham EJ *et al.* The ocular hypertension treatment study. [OHTS] *Arch. Ophthalmol.* 120, 701–713 (2002).
- 31 The European Glaucoma Prevention Study (EGPS) Group. Results of the European glaucoma prevention study. *Ophthalmology* 112, 366–375 (2005).
- 32 Rosenfeld PH, Brown DM, Heier JS *et al.* Ranibizumab for neovascular age-related macular degeneration. [MARINA Study]. *N. Engl. J. Med.* 355, 1419–1431 (2006).
- 33 Heier JS, Boyer DS, Ciulla TA *et al.* Ranibizumab combined with verteporfin photodynamic therapy in nevoascular agerelated macular degeneration. Year 1 results of the FOCUS study. *Arch. Ophthalmol.* 124, 1532–1542 (2006).
- 34 Brown DM, Michels M, Kaiser PK. Ranibizumab versus verteporfin photodynamic therapy for neovacsular age-related macular degeneration: two-year results of the ANCHOR study. *Ophthalmology* 116, 57–65 (2009).
- 35 Abraham P, Yue H, Wilson L. Randomized, double-masked, sham-controlled trial of ranibizumab for neovascular agerelated macular degeneration: PIER study year 2. *Am. J. Ophthalmol.* 150, 315–324 (2010).
- 36 Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ*. 332, 969–971 (2006).
- 37 Early Treatment Diabetic Retinopathy Study Research Group. Early photocoagulation for diabetic retinopathy: ETDRS report number 9. *Ophthalmology* 98, 766–785 (1991).
- 38 Macular Photocoagulation Study Group. Subfoveal neovascular lesion in age-related macular degeneration:

Ophthalmic pharmaceutical clinical trials: interpretation Clinical Trial Perspective

guidelines for evaluation and treatment in the Macular Photocoagulation Study. *Arch. Ophthalmol.* 109, 1242–1257 (1991).

- 39 Moisseiev J, Alhalel A, Masuri R, Treister G. The impact of the macular photocoagulation study results on the treatment of exudative age-related macular degeneration. *Arch. Ophthalmol.* 113, 185–189 (1995).
- 40 Gelfand YA, Linn S, Miller B. The application of the macular photocoagulation study eligibility for laser treatment in age-related macular degeneration. *Ophthalmic Surg. Lasers* 28, 823–827 (1997).
- 41 Beatty S, Au Eong KG, McLeod D, Bishop PN. Photocoagulation of subfoveal choroidal neovascular membranes in age related macular degeneration: the impact of the macular photocoagulation study in the United Kingdom and Republic of Ireland. *Br. J. Ophthalmol.* 83, 1103–1104 (1999).
- 42 Manuscript submission guidelines. J. R. Soc. Med. www.uk.sagepub.com/msg/jrs.htm
- 43 Guide for authors. Air Med. J. www.elsevier.com/journals/air-medical-journal/1067–991X/ guide-for-authors
- 44 Levy G. Significant figures or significant nonsense? Clin. Pharmacol. Ther. 59, 363 (1996).
- •• Brilliant observation regarding significant figures in medicine.
- 45 The Ischemic Optic Neuropathy Decompression Trial Research Group. Optic nerve decompression surgery for nonarteritic anterior ischemic optic neuropathy (NAION) is not effective and may be harmful. *JAMA* 273, 625–632 (1995).
- 46 Siddhartha Kalla. Data Dredging. https://explorable.com/data-dredging
- 47 National Eye Institute. Early Treatment Diabetic Retinopathy Study. www.nei.nih.gov/neitrials/static/study53.asp
- 48 Chew EY, Benson WE, REmaley NA *et al.* Results after lens extraction in patients with diabetic retinopathy: early treatment diabetic retinopathy study report number 25. *Arch. Ophthalmol.* 117, 1600–1606 (1999).
- 49 The Collaborative Ocular Melanoma Study. Publications of the COMS. http://pages.jh.edu/wctb/coms/general/publicat/ pubs.htm#ocrai.
- 50 Collaborative Longitudinal Evaluation of Keratoconus. Peer-Reviewed Publications.
 - https://vrcc.wustl.edu/clekarchive/clekbiblio.html.
- 51 Dorland's Medical Dictionary for Health Consumers. Definition of Association. Saunders, Elsevier (2007). http://medical-dictionary.thefreedictionary.com/Association +(disambiguation).
- 52 Bradford Hill A. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58, 295–300 (1965).
- •• Classic paper on causation.
- 53 Chalupa E, Swarbrick HA, Holden BA, Sjöstrand J. Severe corneal infections associated with contact lens wear. *Ophthalmology* 94, 17–22 (1987).

- 54 Diaconis P, Mosteller F. Methods for studying coincidences. J. Am. Stat. Assoc. 84, 853–861 (1989).
- 55 Jung CG. Synchronicity: An Acausal Connecting Principle. Princeton University Press, Princeton, NJ, USA (1973).
- 56 Carel RS, Korczyn AD, Rock M, Goya I. Association between ocular pressure and certain health parameters. *Ophthalmology* 91, 311–314 (1984).
- 57 Sudarshan S, Huang EHZ, Lim PL *et al.* A case of postvaccination optic neuritis: coincidence or causative? *Eye* 26, 1498 (2012).
- 58 Spurious Correlations. www.tylervigen.com/.
- 59 van Stralen KJ, Dekker FW, Zoccali C, Jager KJ. Confounding. *Nephron Clin. Pract.* 116, c143–c147 (2010).
- 60 Karlowicz MG, Giannone PJ, Pestian J et al. Does candidemia predict threshold retinopathy of prematurity in extremely low birth weight (</=1000g) neonates? *Pediatrics* 105, 1036–1040 (2000).
- 61 Realini T, Barber L, Burton D. Frequency of asymmetric intraocular pressure fluctuations among patients with and without glaucoma. *Ophthalmology* 109, 1367–1371 (2002).
- 62 Wood JP, Osborne NN. Zinc and energy requirements in induction of oxidative stress to retinal pigmented epithelial cells. *Neurochem. Res.* 28, 1525–1533 (2003).
- 63 Arora S, Rayat J, Damji KF. Optic nerve gray crescent can confound neuroretinal rim interpretation: review of the literature. *Can. J. Ophthalmol.* 49, 238–242 (2014).
- 64 Jepsen P, Johnsen SP, Gillman MW, SØrensen HT. Interpretation of observational studies. *Heart* 90, 956–960 (2004).
- 65 Kuper H, Gilbert C. Epidemiology for ophthalmologists: an introduction to concepts, study designs, and interpreting findings. *Br. J. Ophthalmol.* 89, 378–384 (2005).
- 66 Mousa A. Detection and management of confounding variables in prevention of blindness research: a methodological approach with age and sex as models. *Austin J. Clin. Ophthalmol.* 1, 5 (2014).
- 67 Realini T, Fechtner RD, Atreides S-P, Gollance S. The uniocular drug trial and second-eye response to glaucoma medications. *Ophthalmology* 111, 421–426 (2004).
- 68 Novack GD, Crockett RS. Regression to the mean. Ocul. Surf. 7, 163–165 (2009).
- 69 Aaron MT, Applegate RA, Porter J *et al*. Why preoperative acuity predicts postoperative acuity in wavefront guided LASIK. *Optom. Vis. Sci.* 87, 861–866 (2010).
- 70 Nouri-Mahdavi K, Caprioli J, Coleman AL et al. Pointwise linear regression for evaluation of visual field outcomes with the advanced glaucoma intervention study methods. Arch. Ophthalmol. 123, 193–199 (2005).
- 71 Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am. J. Epidemiol.* 104, 493–498 (1976).
- 72 The IVAN Study Investigators. Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration. One-year findings from the IVAN randomized trial. *Ophthalmology* 119, 1399–1411 (2012).

Clinical Trial Perspective Varner

- 73 CATT Research Group. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med.* 364, 1897–1908 (2011).
- 74 Wickström G, Bendix T. The "Hawthorne effect"—what did the original Hawthorne studies actually show? *Scand. J. Work Environ. Health* 26, 363–367 (2000).
- 75 De Amici D, Klersy C, Ramajoli F *et al.* Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control. Clin. Trials* 21, 103–114 (2000).
- 76 Waterman H, Evans JR, Gray TA *et al.* Interventions for improving adherence to ocular hypotensive therapy. *Cochrane Database Syst. Rev.* 4, CD006132 (2013).
- 77 Truzzi M. On the extraordinary: an attempt at clarification. Zetetic Scholar 1, 11 (1978) [Journal later known as Skeptical Inquirer].

- Good reminder for all scientists.
- 78 Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J. R. Soc. Med. 104, 510–520 (2011).
- An astounding paper with wide-reaching implications for all medical disciplines.
- 79 Bonita RE, Adams S, Whellan DJ. Reporting of clinical trials: publication, authorship, and trial registration. *Heart Fail. Clin.* 7, 561–567 (2011).
- 80 Nayak BK. Marching ahead with clinical trial registration. Indian J. Ophthalmol. 56, 175–176 (2008).