Potential bias in ophthalmic pharmaceutical clinical trials

Paul Varner
John J Pershing Veterans’ Administration Medical Center, Poplar Bluff, Missouri, USA

Abstract: To make clinicians aware of potential sources of error in ophthalmic pharmaceutical clinical trials that can lead to erroneous interpretation of results, a critical review of the study design of various pharmaceutical ophthalmic clinical trials was completed. Discrepancies as a result of study shortcomings may explain observed differences between reported ophthalmic trial data and observed clinical results.

Keywords: evidence-based medicine, validity, bias, gold standard

Introduction

Although not necessarily a new concept, since its designation by Sackett et al in 1996, evidence-based medicine (EBM) has become a watchword for the medical community. EBM is the term applied to the application of the scientific method to medical practice and involves the direct application of the best-available medical research data to clinical patient care. It requires the integration of available research evidence with clinical experience and may challenge long-standing medical traditions that have not yet been subjected to appropriate scientific scrutiny. Among clinical providers from various medical specialties there is a growing awareness of the importance of applying EBM in health care settings (Hagdrup et al 1998; Stefanski et al 2004; Luehr 2006). Whether providers are making the transition from realization to application, however, remains equivocal (Benatar 2005) or unknown (Robinson et al 2004).

The emerging appreciation of EBM is also recognized within the ophthalmic literature (Sharma 2001; Straatsma 2003) as well as by the International Council of Ophthalmology (2007). The appearance of various letters to ophthalmic journal editors also suggests that clinicians realize the shortcomings inherent in the reported findings for diverse ophthalmic clinical trials involving “normal-tension” glaucoma (Caprioli and Maguire 2001), neovascular age-related macular degeneration (Spaide 2006), and dry eye syndrome (Brown 2007), just to name a few. Additionally, ophthalmic investigators and providers are aware of differences that arise between analogous trials – eg, the Ocular Hypertension Treatment Study (OHTS) (Kass et al 2002) and the European Glaucoma Prevention Study (EGPS) (EGPS Group 2005) – and are actively seeking to understand these differences.

However, most medical providers continue to offer clinical recommendations to patients based on didactic instruction, clinical lore, or habit because of the continuing frustration faced when trying to reconcile the necessity to review an overwhelming amount of published medical information within the time limitations of daily practice (Barton 2001). As succinctly stated by Lee, when it comes to medical knowledge “there is simply too much to know” (Lee 2005).

With respect to medical literature, there is a common fallacy that faces reviewers: that if a recommendation appears in print, it must be valid (ie, a form of assumption
bias on the part of the reader). This form of bias occurs when journal reviewers under time constraints quickly scan abstracts and assume that the statements provided in the “conclusion” section must be accurate and backed up by the following pages of the article and the accompanying references (ie, written word bias). Another form of assumption bias may be found in the belief that the contents of expert speaker lectures or recommendations from pharmaceutical company representatives are entirely objective and lacking bias (authority bias).

Unfortunately, these assumptions are not always true. Although not explicitly stated during professional training, all healthcare providers are expected to provide their own self-education once they leave their institutions of graduate learning. This means continually reviewing pertinent journals, and then completing the Herculean task of critically evaluating the literature, dismissing incomplete data and applying new information directly to patient encounters. Through this process the implicit, underlying belief is that clinicians will continue to grow in knowledge, thereby becoming better, more-informed clinicians as they accumulate more clinical experience based on EBM. At no time would a medical professional be expected to be the same provider that he or she was a year previously. This would seem to be the expectations of patients as well (Lee 2005) and enters into the controversial side issues of ongoing provider competency (American College of Physicians 2004).

Fortunately, there are now online open-access resources available to disseminate EBM: The Cochrane Collaboration (www.cochrane.org/index0.htm) and Bandolier (www.jr2.ox.ac.uk/bandolier/), among others. Also, clinical trial registry is now required prior to publication of results of trials in some journals (http://clinicaltrials.gov/), which may help to prevent non-publishing of negative results of some clinical trials.

Quite simply, the consequences of failing to review the literature critically are untoward morbidity and mortality for patients. The literature is replete with examples of failed medical therapies once heralded as being definitive: intracranial-extracranial arterial bypass to reduce the risk of ischemic stroke (EC-IC Bypass Study Group 1985), anti-inflammatory therapy with rofecoxib (Vioxx®, Merck and Co Inc, Whitehouse Station, NJ, USA) therapy (Bresalier et al 2005), and use of oral prednisone as monotherapy in patients with optic neuritis (Beck et al 1992). Chalmers statement that “One only has to review the graveyard of discarded therapies to discover how many patients have benefited from being randomly assigned to a control group” (1968) reminds us that all providers should be even more diligent about experimental therapies and critical of the results of the reports. Finally, providers cannot rely on the existing mechanisms of public oversight groups to ensure that reported clinical trial results are wholly trustworthy (Topol 2004).

The unwritten corollary of these expectations is that clinicians will provide the ultimate control over new medical practices, procedures, and pharmaceutical products entering the health care marketplace. In other words, sceptical review of medical literature by practitioners with the null hypothesis firmly in mind (ie, there is no difference between treatment groups) and assessing the consistency of results with personal observations would satisfy the demands of EBM and avoid introduction of sub-optimal treatments and agents into daily practice. However, practitioners continue to make clinical recommendations based on peer consultation, expert opinion provided at continuing medical education conferences, or (as can be the case with pharmaceutical agents) based on data provided by representatives of pharmaceutical corporations.

The business of pharmaceutical clinical trials

To understand the impetus behind this process, practitioners must realize the financial implications of bringing a new pharmaceutical agent to the marketplace. The estimated cost for the discovery, development, experimental trial and introduction of a single, new, successful drug to the marketplace ranges from approximately €277 million ($390 million; Kettler 1999 [estimates updated to 2007 values via inflation calculator]) to €568 million ($800 million; Pharmaceutical Research and Manufacturers of America 2007). Of this total, slightly less than 50% is incurred in the execution of clinical pharmaceutical trials (Kettler 1999).

In terms of global expenditures for pharmaceutical research and development, the top fifty pharmaceutical companies on the planet spent an estimated €62 billion in 2006 alone (MedAdNews 2007). In order to provide a broader frame of reference, bear in mind that this sum for research and development expenditures would rank the pharmaceutical industry number 56 out of 180 in a world ranking of countries by Gross Domestic Product (International Monetary Fund 2007).

By contrast, publicly-funded (government) spending on pharmaceutical research lags considerably behind private sector spending. The US government invested approximately €7 billion ($10 billion) per year in pharmaceutical research and development in the 1990s (Congressional Budget Office 2006), and the EU estimates for pharmaceutical research funding was expected to be approximately €0.5 billion
per year between 2002 and 2006, which represented a
doubling of expenditures over the previous four years
(European Commission 2002). Contrast these figures to the
€21 billion ($29.9 billion) spent annually on pharmaceutical
promotion to the public via direct-to-consumer campaigns
(Donohue 2007). In summary, an enormous amount of money
is spent each year on pharmaceutical research and private
industry is funding the bulk of it.

**Literature review**

Within this environment of colossal financial pressure to
realize returns on investments, the potential for bias in phar-
aceutical trial conduct and reporting remains real and is
recognized in the medical literature (Als-Nielsen et al 2003;
Liesegang et al 2005). Thus, the pertinent question remains:
how can clinicians obtain objective knowledge regarding
novel pharmaceutical agents? The answer, of course, is by
developing an awareness of this problem and addressing it
through continuous, careful review of medical literature.
Again, formal training for healthcare providers offers very
little in the way of teaching methods for critical literature
review, which is necessary to achieve this end. Furthermore,
with the explosion of medical research publication that was
taking place by the 1980s, it became clear that providers
were unable to keep up with the deluge of new information
that they were facing. Across all medical disciplines over 11
million randomized controlled trials had been recorded via
MEDLINE by 2001 (Tsay and Yang 2005).

More germane to eye care, according to MedBioWorld
there are more than 80 journals dedicated exclusively to eye
and vision research (2007). In 2006 the Cochrane Eyes and
Vision Group tallied more than 5,000 ophthalmic clinical
trials in just 7 different ophthalmic journals accounting
for 125 years of cumulative publications; while in 2004
alone, there were 1,919 articles published in 4 leading
ophthalmology journals (Lai et al 2006).

In order to provide some structure to clinical trial
review, various medical journals have provided methods
for systematic literature reviews with the goal of teaching
these techniques to their readers (Evidence-Based Medicine
Working Group 1993–2000; Sackett et al 1997; Glasser
and Howard 2006; Ives 2006). These reviews provide good
points of departure when learning or teaching critical litera-
ture review.

Statistical discussions contained in clinical trial reports
can also appear overwhelming. Not all clinicians can hold
advanced mathematics degrees in statistics in order to
fully interpret the significance of applied methods for Cox
regression models, paired-eye correlation techniques, and
non-parametric bootstrap procedures (examples from only
the Advanced Glaucoma Intervention Study trial 2004);
however, distilling the presented trial data into useful clinical
information is an EBM skill incumbent upon all providers.
Number-needed-to-treat analysis is a relatively straightfor-
ward mathematical application to help clinicians who are
non-statisticians (McQuay and Moore 1997).

A general framework for critically appraising any clinical
trial can be organized under four general headings: (1) has the
study been appropriately designed? (2) Have the study results
been correctly analyzed? (3) Have the study results been
interpreted correctly? (4) Can the study results be applied
to the patients in the reviewer’s practice? (Ives 2006). A
more detailed summary of questions to ask when reviewing
pharmaceutical (or other clinical) trial data is provided in
Table 1. For further in-depth discussions of the specifics of
clinical trial design (ie, randomization schemes, determina-
tion of n values, alpha calculations, statistical power, etc)
please consult appropriate references.

**Clinical trial validity and bias**

When assessing EBM, reviewers of clinical trial data must
consider internal and external validity of the clinical trial
results. When *internal validity* is met, readers can be assured
that the study has truly examined what it set out to measure.
The majority of the questions in Table 1 address this com-
ponent. *External validity* is the ability to generalize findings
from the study cohort to the general population and will be
discussed in turn.

In order to determine study validity, readers must first
consider potential sources of error in the study. Essentially
all medical studies contain varying degrees of *bias* (Grimes
and Schulz 2002), which is considered to be: “any trend in
the collection, analysis, interpretation, publication or review
of data that can lead to conclusions that are systematically
different from the truth” (US Centers for Disease Control
and Prevention 2006). The presence of bias is controlled
via careful study design and statistical analysis; however the
complete elimination of all forms of bias may still be unat-
tainable. The presence of bias does not automatically imply
that the study results should be disregarded, however the
final acceptance of study results ultimately resides with the
reader, who must balance potential sources of bias against
their possible impact on the recommendations and conclu-
sions of the authors, and secondary interpretations by experts.
Once more, the onus for change in clinical practices based
on EBM lies with the provider.
There are over one hundred recognized types of bias in the medical literature (Armitage and Colton 2005). Although there is no consensus regarding a set number of “major” sources of bias, experimental errors can be loosely considered to occur during three periods: before the study begins (selection bias), while the data is being collected during the study (information bias), and after the conclusion of the study period when data interpretation occurs (interpretation bias). (For further discussion regarding sources of research bias see also: Rothman and Greenland 1998; Grimes and Schultz 2002; Chow and Liu 2003, among many others.) For ophthalmic pharmaceutical trials, careful attention must be paid to the effects of selection bias and will be the focus of the following discussion. Specifically, a review of patient selection criteria, standards of comparison, and meeting well-defined endpoints are areas of potential selection bias in some ophthalmic clinical trials.

Patient selection

Based on a review of existing literature, the issue of bias in patient recruitment for inclusion into ophthalmic studies has not been directly addressed. Nor have the effects of volunteer bias (study volunteers tend to be healthier than non-participants and may exhibit outcomes unlike other subjects) or study remuneration (another form of volunteer bias) for participation in ophthalmic studies.

Patient recruitment

Bias in the recruitment of controls for comparison to affected study participants would not seem to be an issue for many advanced-stage ophthalmic pharmaceutical trials. Presumably because at this point in the process (ie, Phase 3 trials in the US), investigational medications are typically used only on those patients who already manifest a given condition for which there is no definitive treatment (ie, healthy human subjects in a control group are not given experimentally-induced choroidal neovascular membranes or intraocular pressure increases in order to test efficacy of experimental pharmaceutical agents). In general, these trial subjects are recruited directly from patient encounters in participating study sites as they are diagnosed clinically, and appropriate randomization schemes are used to prevent unexpected differences between cohorts. However, reviewers should be aware that there are many industry-sponsored pharmaceutical trials that actively recruit patients and remunerate for their participation (Center Watch 2007). The reader is left to attend to study sponsorship and to discern whether volunteer bias could account for small discrepancies in reported treatment effects.

In one of the first efficacy studies for olopatadine hydrochloride (Patanol®), patients with a prior history of active allergic conjunctivitis were required to have positive dermatological (skin prick), serological (radioallergosorbent test [RAST]), or conjunctival test for enrolment into the study (Abelson and Spitalny 1998). To remain enrolled in the study, participants were required to have a positive conjunctival allergen challenge (CAC) at subsequent visits. CAC testing requires provocative inoculation of eyes with increasing allergenic inoculations, assessing for allergic responses, and then evaluating ocular responses to trial medications (Abelson et al 1990). CAC testing provides highly reproducible results and has been used frequently when studying groups of human ocular allergies (Friedlaender 2004), although questions remain regarding individual responses (Aichane et al 1993) and use of rescue medications and quality of life issues (Radcliff et al 2006). Few clinicians have direct clinical experience with CAC testing and may be left with lingering questions regarding the generalised applicability of CAC testing and the results of such trials. The result may be reliance on the opinions of others. However, clinicians
are reminded that questions regarding patient recruitment, olopatadine and EBM have not been definitively answered in the literature (especially with respect to a gold standard for allergic conjunctivitis discussed below).

Eligibility (inclusion and exclusion) criteria

Ophthalmic providers must recognize that certain inclusion and exclusion criteria for clinical trials may create other forms of internal bias that can affect external validity of reported results. This in turn may explain differences between the responses of patients reported in studies to those observed in non-study site clinics (ie, a failure to demonstrate external validity). Unfortunately, there are many ophthalmic clinical trials that leave unanswered questions for the reader to ponder.

For example when considering a study of ocular hypertension or open-angle glaucoma (OAG), there is a question as to whether the inclusion of patients with secondary open-angle glaucoma forms (eg, pseudoexfoliation [PXG] or pigmentary dispersion glaucoma [PDG]) affect the outcomes. When exact etiologies remain unknown, reviewers cannot be certain that it will not. Furthermore, results from one racial population may not necessarily translate to similar findings in other racial groups.

In neovascular age-related macular degeneration (AMD), questions remain as to whether occult, classic, and mixed forms of choroidal neovascular membranes (CNVM) follow the same natural progression. Conversion between types is known to occur. Conversely, do patients with recurrent choroidal revascularization respond differently to an intervention than those with primary lesions? Clinical experience suggests that they do in some cases.

For ophthalmic complications of diabetes mellitus, the question arises as to whether patients in a study of clinically-significant macular oedema (CSME) respond differently than those with non-CSME. Perhaps more to the point, why treat those patients previously described as being “non-clinically significant”? In addition, it is not clear whether subjects will have equivalent responses to a treatment if there is an ischemic macula versus a normally-perfused macula, and what if those patients retain good visual acuity at the outset? Does the duration of the macular oedema affect the outcomes?

Finally, can the results of a study targeting primarily female patients be applied to a clinical population which also includes males, or even predominantly males as in the case of an American Veterans’ Health Administration eye clinic?

Specific examples of these types of potential selection bias regarding eligibility criteria observed in ophthalmic studies are presented in Table 2. Certainly other questions that arise from clinical trial review exist as well. It is also important to bear in mind that most patients in these studies were excluded from having other ocular co-morbidities such as cataracts, subnormal vision, prior ophthalmic surgeries, and the like. Patients with co-morbid conditions may or may not respond to treatments in the same way, further questioning the issue of external validity of the studies. While generalization from single morbidity patients to co-morbid ones has been discussed in other medical disciplines (eg, primary care) (Fortin et al 2006), to date it has not been directly addressed for ophthalmic patients.

If the natural courses of the disease for subtypes of patients are etiologically identical, then it is acceptable to combine sub-groups for statistical analysis. However, the authors should point out these differences (if they exist) and, likewise, the

<table>
<thead>
<tr>
<th>Table 2 Clinical studies with uncertain inclusion/exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pharmaceutical intervention</strong></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Hypotensive eye drops vs surgery</td>
</tr>
<tr>
<td>Laser/eye drops vs observation</td>
</tr>
<tr>
<td>Ranibizumab for AMD</td>
</tr>
<tr>
<td>Anecortave acetate for AMD</td>
</tr>
<tr>
<td>Pegaptanib sodium for DME</td>
</tr>
<tr>
<td>Bevacizumab for DME</td>
</tr>
<tr>
<td>Triamcinolone acetonide for DME</td>
</tr>
<tr>
<td>Hypotensive eye drops vs observation</td>
</tr>
<tr>
<td>Triamcinolone acetonide for DME</td>
</tr>
<tr>
<td>Cyclosporine for dry eye syndrome</td>
</tr>
</tbody>
</table>

*Abbreviations: AMD, age-related macular degeneration; CNVM, choroidal neovascular membranes; CSME, clinically significant macular oedema; DME, diabetic macular oedema; MDRSG, Macugen Diabetic Retinopathy Study Group; DRCRN, Diabetic Retinopathy Clinical Research Network; EGPS, European Glaucoma Prevention Study; OAG, open-angle glaucoma; OCT, optical coherence tomography; PDG, pigmentary dispersion glaucoma; PSG, pseudoexfoliation.*

**Pharmaceuticals:** ranibizumab (Lucentis®, Genentech Inc, South San Francisco, CA, USA); anecortave acetate (Retaane®, Alcon Research LTD, Fort Worth, Texas, USA); pegaptanib sodium (Macugen®, [OSI] Eyetech Pharmaceuticals Inc, New York NY, USA); bevacizumab (Avastin®, Genentech Inc, South San Francisco, CA, USA); cyclosporine (Restasis®, Allergan, Irvine, CA, USA).
must not compare metaphorical “apples” to “pears”.

Additionally, trials assessing the same outcomes may include clinical endpoint are identically designed, so direct com-
of study subjects. Insuffi cient patient numbers lead to weak
Strength of clinical trial results often depends on the number
is also worthy of note that not all trials that defi ne the same
investigators. Reviewers of clinical trials should be aware
When considering the performance of a pharmaceutical
standards of comparison
Standards of comparison
When considering the performance of a pharmaceutical
agent, a proper frame of reference must be utilized by study
investigators. Reviewers of clinical trials should be aware
when studies lack solid standards of comparison or when
other factors cloud direct comparisons between groups. It
is also worthy of note that not all trials that defi ne the same
clinical endpoint are identically designed, so direct com-
parison of resultant findings can be potentially misleading.
Additionally, trials assessing the same outcomes may include
patients at different stages of the disease. In sum, the study
must not compare metaphorical “apples” to “pears”.

Insufficient patient numbers
Strength of clinical trial results often depends on the number
of study subjects. Insuffi cient patient numbers lead to weak
associations; too many patients no longer adds to statisti-
cal power. In general, the number of patients enrolled in a
study is based upon the suspected treatment effect (ie, if
investigators expect a 20% difference between groups, then
x number of patients are recruited; if the expectation is 30%,
then y number of patients are enrolled, and so on). In gen-
eral, the narrower the expected treatment margin, the larger
the number of patients required to statistically demonstrate
that difference (Beck 2006). Because of the variance in dis-
ease incidences, there is no accepted minimum number of
patients required in order to statistically power a study. For
pharmaceutical trials, the US Food and Drug Administration
(FDA) recommends that Phase 3 trials (ie, those designed
to ensure safety and efficacy before pharmaceutical release
for use by the general population) enrol “several hundred to
several thousand” patients (FDA 2007) to ensure validation
of clinical trial data. Because of the strict controls involved
in these trials, Phase 3 trials results should be rigorously
scrutinized, and their outcomes seriously considered, prior
to the integration of a new agent into clinical practice.
Conversely, readers should be aware that Phase 3 trials are
also industry-sponsored and that some Phase 3 data remain
unpublished and tightly guarded due to proprietary interests
(yet another source of potential bias).

When studies fail to enrol a signifi cant number of patients,
then small differences in study groups can lead reviewers
to question slim margins of successful treatment. There are
many pharmaceutical studies with low numbers of partici-
pants. The Phase 3 study for Patanol® enrolled 169 patients
in three arms (Abelson and Spitalny 1998), whereas the
Verteporfin in Photodynamic Therapy (VIP Study Group
2001) examined 120 patients in two arms and the Collabora-
tive Normal-Tension Glaucome Study (1998, not a Phase 3
pharmaceutical study) enrolled 140 patients in 2 trial arms.
Readers should bear in mind that the n value represents the
total number of study participants (or eyes, depending on the
study), and that the final number of patients in each arm of
the study will be reduced accordingly.

Failure to compare against the “gold standard”
If a new therapeutic agent is recommended for integration
into current disease management plans, how can clinicians
be assured that the recommendation is valid if the agent in
question has not been compared against the gold standard
treatment for that condition? Originally, the term gold
standard applied to a diagnostic test that was considered to
be defi nitive in its determination of a condition. However, in
more recent times, the phrase has come to represent any
medical benchmark (clinical outcome, diagnostic test, biopsy,
chronic disease index, medication, autopsy, etc) against
which new procedures and tests can be measured. There are
many ophthalmic gold standards ranging from clinical test-
ing for malingering (Gundogan et al 2007) and automated
perimetry for glaucoma (Thomas and George 2001) to com-
puterized tomography for suspected orbital fractures (Cruz
and Eichenberger 2004) and reference stereo photographs
for grading diabetic retinopathy (Williams et al 2004). In
essence, ophthalmic providers are well aware of certain clini-
cal practices against which others are compared.

This being the case, a gold standard retains its place of prominence until a new test, medication, or procedure
challenges its status and there is widespread confirmation via
research data that there is a new gold standard. Unfortunately, there is no governing medical body to oversee these transitions and the metamorphosis may take years for providers (consider the apparent low impact of the results of the optic neuritis treatment trial [Trobe et al 1999]). With regard to pharmaceutical agents, in the absence of a widely-accepted gold standard, researchers may need to compare an intervention to nothing (ie, the natural course of the disease) or a placebo, which will be discussed below. Table 3 lists a few studies that have abandoned important ophthalmic gold standards.

Readers should also be aware of the importance of unexpected findings when head-to-head comparisons with gold standards do periodically occur. In the only randomized clinical trial of an ophthalmic fourth-generation fluoroquinolone since US approval in 2003, moxifloxacin (Vigamox®, Alcon Research LTD, Fort Worth, TX, USA) was found to be an equivalent – but not superior – therapy for bacterial corneal ulcers when compared to the anti-bacterial gold standard: fortified topical anti-bacterial solutions or an older amino-glycoside (Constantinou et al 2007).

The results of the only study comparing olopatadine hydrochloride to a mild ophthalmic steroid showed superiority of olopatadine; however, the comparison was not made to a classic corticosteroid like dexamethasone, fluorometholone or prednisolone, but rather to a newer, ophthalmic corticosteroid, loteprednol etabonate (Lotemax®, Bausch and Lomb, Tampa, FL, USA; Berdy et al 2002). Prior to the introduction of olopatadine, gold standard treatment for allergic conjunctivitis was single-action ophthalmic antihistamine or steroid eye drops (Spalton 1984).

In the case of CNVM related to AMD, neither verteporfin, pegaptanib, ranibizumab or bevacizumab have been compared to thermal laser photocoagulation, the previous gold standard for these cases. Interestingly, when submacular surgery was compared to laser photocoagulation no advantage to incisional surgery was found, the negative results were reported (Submacular Surgery Trials pilot Study Investigators 2000), and interest in the procedure waned as should be expected.

However, reviewers are left to wonder why the anti-bacterial comparison to the gold standard was so slow in coming and why the anti-allergy trial did not compare the study medication to the older gold standard. Neither of these questions remains answered. Even bigger questions remain regarding why ophthalmic practitioners so quickly converted prescribing patterns in favor of the newer medications.

**Failure to compare to placebo**

The placebo effect on medical studies has been well documented for many years, yet it remains poorly understood and its overall effect varies among trials. However, the inherent problems in comparing a pharmaceutical agent to baseline performance (ie, nothing) are also widely recognized and the failure to include placebo control may introduce a significant source of bias to study outcomes. An excellent example of the failure to include placebo control is the Ocular Hypertension Treatment Study (OHTS, Kass 2002). OHTS investigators were attempting to compare the treatment of ocular hypertension to controls who received no placebo treatment. Because the results of this study were not unequivocal, a reviewer must ask if the results had been different (or the treatment effect even narrower than what was reported) had the control group received placebo treatment. The lack of placebo control leaves room for doubt in the results of a clinical trial that has generated so much attention since its publication.

There is cause for concern when pharmaceutical studies for bacterial keratitis are not placebo-controlled. Although terminology is likely a barrier to understanding in this instance (ie, superficial keratitis versus corneal ulceration, or the use of the term “microbial” keratitis, requiring readers to determine which microscopic organism is being discussed), it is generally conceded that non-ulcerative bacterial keratitis

---

**Table 3 Clinical studies not addressing established ophthalmic gold standards**

<table>
<thead>
<tr>
<th>Pharmaceutical intervention</th>
<th>Reference</th>
<th>Gold standard</th>
<th>Gold standard reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verteporfin for AMD</td>
<td>TAP 1999</td>
<td>Thermal laser photocoagulation</td>
<td>MPS 1991</td>
</tr>
<tr>
<td></td>
<td>TAP 1999</td>
<td>Geographical Criteria for CNVM</td>
<td>MPS 1991</td>
</tr>
<tr>
<td>Triamcinolone acetonide for DME</td>
<td>DRCRN 2007</td>
<td>OCT in lieu of CSME</td>
<td>ETDRS 1991</td>
</tr>
<tr>
<td>Olopatadine for ocular allergies</td>
<td>Abelson 1998</td>
<td>Mild corticosteroid</td>
<td>Spalton 1984</td>
</tr>
</tbody>
</table>

*iie, subfoveal, juxtafoveal and extrafoveal.

**Pharmaceuticals:** verteporfin (Visudyne®; Novartis AG, Basel, Switzerland).

**Abbreviations:** AMD, age-related macular degeneration; CNVM, choroidal neovascular membranes; TAP, Treatment of Age-related Macular Degeneration with Photodynamic Therapy Study Group; MPS, Macular Photocoagulation Study Group; ETDRS, Early Treatment Diabetic Retinopathy Study Research Group; DRCRN, Diabetic Retinopathy Clinical Research Network; CSME, clinically significant macular oedema; OCT, optical coherence tomography.
(ie, a mild superficial keratitis) is a self-limiting condition (Chung and Cohen 2000; Sheikh and Hurwitz 2001; Epling and Smucny 2006). Readers should be wary of lack of placebo control in some of these studies when reviewing head-to-head comparisons of anti-bacterial agents. Unfortunately, Phase 3 clinical trial data for fourth generation fluoroquinolones are not available for review in the public domain (per personal correspondence to both Allergan and Alcon Research LTC).

In anti-bacterial studies that are placebo-controlled, reviewers must also weigh slight reductions in duration of ocular symptoms against contributions to increasing bacterial resistance to anti-bacterial therapy. Finally, when comparing against gold standards, readers should be aware that MIC$_{90}$ (minimum inhibitory concentration required to inhibit 90% of bacterial isolates) values do not significantly differ among ophthalmic fluoroquinolones and are far exceeded in current drug delivery methods (see Kowalski et al 2005, for further discussion of this important issue).

**Confounders**

Reviewers of studies must be wary of potential confounders that may influence published results. Confounders are sources of error that occur when a factor other than the exposure of interest affects the result of the study (Hatch 1998). In other words, these are factors that have not been controlled for via study design and which may lead to faulty causation. Readers cannot be sure if the reported effect was a direct result of the study intervention or an indirect result of the confounder, especially when reported treatment effects are narrow.

Examples of potential confounders in ophthalmic studies include the concurrent use of Centrum® (Whitehall-Robins Healthcare, Madison, NJ, USA) multivitamin and mineral supplements with the study formulation in the Age-Related Eye Disease Study (AREDS 2001), allowance of verteporfin therapy before and during the pegaptanib Phase 3 trial for neovascular AMD (Gragoudas et al 2004), or the use of verteporfin therapy during the Phase 3 MARINA study of ranibizumab (Rosenfeld et al 2006).

In the case of the AREDS study, how can the reviewer be certain that the Centrum supplements were not responsible for the mild, reported treatment effect and not the study medication? In the cases of pegaptanib or ranibizumab can we be certain that the observed effect was due to the investigational medication or a delayed response to verteporfin (or, if the study medications are more effective, then what becomes the new gold standard?). If these potential confounders are not sufficiently explained by the investigators, then readers are left to question the veracity of the reported conclusions found in the abstracts.

**Study endpoints**

A crucial factor to consider in all medical research is the definition of study endpoint. Since very few ophthalmic conditions are reported in terms of “cure rates” or “prevention rates,” investigators must pre-determine endpoints for research with finite temporal limits. From the standpoint of study design, research endpoints should be succinctly defined, reproducible, and appropriate to the hypothesis. Literature reviewers must ask themselves “did the endpoint(s) answer the original question?” Failure to report results in terms of primary outcomes represents a failure of a study to provide meaningful clinical information, and calls into question the original study hypothesis. It is also helpful to see outcomes written in terms of null hypotheses, but often readers are left to deduce this for themselves.

More to the point, however, reviewers must ask themselves a more clinically-relevant question regarding a study, namely: does the study ask a clinically-useful question? If a pharmaceutical trial is addressing a clinical entity like bacterial conjunctivitis, which is described as a self-limiting condition, and which is difficult to definitively diagnose without point-of-care testing, then perhaps it is more pertinent to question whether the self-limiting condition needs any active pharmaceutical agent at all, rather than assessing whether the medication resolves the condition. Even more basic to this discussion is the questions of whether there is any such thing as a ‘garden variety’ bacterial conjunctivitis. Although usually explicitly stated, there is another germane clinical question for readers to consider for an antibacterial ophthalmic medication: is in vitro data applicable to in vivo clinical settings? This is another test of external validity. These ideas are rarely discussed in ophthalmic anti-bacterial pharmaceutical trials, leaving the readers to wonder why.

Readers must also consider financial implications of pharmaceutical trials on their patients and society as a whole. Are the findings significant enough to justify the cost of switching to a new gold standard? As it remains incumbent upon all medical providers to consider medical costs for interventions, literature reviewers should be aware of the phenomenon of “latest-greatest” drugs. The assumption bias here is that the newest pharmaceutical agent on the market must somehow be the best. Clinically, readers need to ask more challenging questions of study reports, including whether the drug better than existing treatments and gold standards, or placebo.
This is especially important when the more inexpensive alternative may also be more efficacious. Consider the possibility of mild ophthalmic corticosteroids for the palliative relief of the symptoms of allergic conjunctivitis. Olopatadine hydrochloride costs approximately £57 ($80) per 5 mL versus £25 ($35) per 5 mL for generic olopatadine (available in the UK as Optanol®, but not yet available in the US). Then compare both of these to €14 ($20) per 5 mL for generic fluorometholone. Two to four times as many patients can be treated with the same clinical outcome with the generic corticosteroid eye drop, however, to date, a direct comparison of these ophthalmic medications has not occurred.

Finally, medical providers should always think in terms of functional outcomes. Functional outcomes are those endpoints which address mortality and morbidity. Fortunately, few ophthalmic clinical trials need address mortality issues, but many ophthalmic conditions impinge on important issues of morbidity: vision loss and quality of life (Vu et al 2005). Specifically, can a patient undergoing the treatment perceive an “improvement” following the intervention? Although subjective in origin, ophthalmic patients typically wish for their vision to improve back to reading/driving levels (in the cases of vitreo-retinal, lenticular, or corneal interventions, for example) or for their eyes to feel better (in the cases of conjunctivitides, uveitides, for example). Readers would always hope that study endpoints would be written to allow for functional improvements that patients can appreciate.

A shortcoming of many retinal studies is that study endpoints are defined in terms of improvement via halving of the visual angle. While statistically significant, there is considerable room for claims regarding clinical significance. Few patients seem to rave about improvements from 20/200 to 20/100 (three lines) on the Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity charts. While a statistical victory, such patients are still unable to drive or to read their mail without low vision magnification, and this calls into question the issues of both internal and external validations of study results. Pharmaceutical studies in primary open-angle glaucoma are fraught with the additional problem of a lack of early patient symptoms, making the realization of functional improvement even more difficult for study investigators to address. This may continue to be a problem until the ophthalmic community can begin reporting findings in terms of “cure rates” or “prevention rates.”

Conclusion
Clinical trial results that can withstand careful scrutiny by many readers may stand the test of time; however, immediate acceptance of conclusions found in abstracts may lead to erroneous generalization of study results to clinical populations. Only individual practitioners can evaluate the evidence for themselves, and the responsibility to do so remains incumbent upon them. When reviewing ophthalmic pharmaceutical trials, readers must consider the internal validity issues of patient selection, standards of comparison and study endpoints, and then determine if the study results are externally valid and can be applied to their own patient populations. The final determination regarding a switch in clinical practice to a new gold standard should be based on individual review and not on other sources. It is the intent of this article to assist practitioners in their ongoing quest to provide the best clinical care for their patients.

Acknowledgments
The views expressed in this article are those of the author and do not necessarily represent the positions of the Department of Veterans Affairs. The author has no financial interests to disclose.

References

Beck RW. 2006. Sample size for a clinical trial: why do some trials need only 100 patients and others 1000 patients or more? Ophthalmol, 113:721–2.


Potential bias in ophthalmic pharmaceutical clinical trials


