

Evidence-Guided Ophthalmology

Donald S. Fong, MD, MPH; Frederick L. Ferris III, MD

One challenge for ophthalmologists entering the 21st century will be to make clinical decisions based on valid information or evidence rather than intuition, hearsay, or peer practice. The practice of ophthalmology has been information based since its inception. However, before the development of modern clinical study design and statistics, ophthalmologists had been applying information from all sources to their practice, without regard to its quality. Subconsciously, many practitioners may give equal weight to the printed word, no matter the source. Because of an information explosion, ophthalmologists must now pick and choose evidence from a mountain of published information. How can this be done efficiently?

Reading peer-reviewed journals is an important first step. Information gathered from this activity raises awareness on developments in the field. Knowledge about new techniques and approaches provides the framework for evidence-guided ophthalmology (EGO), that can eventually be incorporated into practice. However, practicing EGO is not just critically reading the literature. Evidence-guided ophthalmology is a directed approach to taking new information and incorporating it into clinical practice. This report provides a description of this approach.

PRACTICING EGO

Asking the Question

The first step toward practicing EGO is directing the reader's inquiry with a well-phrased question. Knowledge of the literature is important so that the question can be answered if an answer exists. Without knowing the contents of the published literature, a well-formed question cannot be made. The question should be organized into 3 elements: exposure, out-

come, and setting. Exposure is a term from epidemiology, which describes what the patients were exposed to. The exposure might be a treatment for an existing disease or a risk factor that might increase or decrease the risk of developing the disease. The outcome is the precise end point of interest. The more precisely this is specified, the more specific the answer can be. Outcomes do not have to be desirable end points such as improvement in vision; they may be undesirable end points such as adverse effects. The specific setting can be very important in narrowing the search. For example, let's say the reader is interested in knowing whether to stop aspirin therapy in patients with diabetic retinopathy. The first step is focusing the question: Does oral aspirin treatment (exposure) affect vitreous hemorrhage (outcome) in patients with diabetes mellitus (setting)?

Getting the Evidence

The second step is finding the evidence. There are many ways of finding information electronically. CD-ROMs can be helpful, but the best sources of information are the databases maintained by the National Library of Medicine, such as MEDLINE. Prior to June 1997, ophthalmologists had to search MEDLINE at a library with a di-

From the Departments of Ophthalmology, Southern California Permanente Medical Group, Baldwin Park, and University of California at Los Angeles School of Medicine (Dr Fong); and the Division of Biometry and Epidemiology, National Eye Institute, National Institutes of Health, Bethesda, Md (Dr Ferris).

National Library of Medicine: Internet Grateful Med Search Screen

Internet Grateful Med is currently set to search file MEDLINE

Tip: bookmark the [IGM front screen](#), not this page. [Here's why.](#)

Enter Query Terms:

Search for
 vitreous hemorrhage as Subject

AND search for
 aspirin as Subject

_____ as Subject

Apply Limits:

Languages: Publ Types:

Study Groups: Gender:

Age Groups: Journals:

Year range: Begin year through End year

Internet Grateful Med is currently set to search file MEDLINE

Image of the National Library of Medicine's search screen.

rect connection to MEDLINE via password. Now, through 2 web-based interfaces, MEDLINE can be accessed through either PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) or Internet Grateful Med (<http://igm.nlm.nih.gov/>) without charge or password. In the past, searching with either Index Medicus or MEDLINE required the use of Medical Subject Headings (MeSH), but use of MeSH is no longer necessary. The "regular" language of medicine can now be directly entered because the current search interfaces have built-in thesauruses for interpreting user entries. The PubMed interface is easier to use, but it has less powerful search options. A useful PubMed tool is "Clinical Queries" with built-in sensitivity and specificity concepts related to therapy, diagnosis, etiology, and prognosis, which automatically filter the literature. The Grateful Med interface is more powerful, and search limits such as English, human studies, age groups, clinical trials, and gender can be easily added.

After getting a list of references from the search interfaces,

the full text of the articles is often necessary to learn the results of a study. Depending on the size of the reader's hospital or clinic, the library may have the relevant journal. If not, the librarian can usually get a copy of the article via DOCLINE or arrange for an interlibrary loan of the journal. Loan-some Doc is service provided by the National Library of Medicine that allows registered users to have articles mailed or faxed to them. For some journals, many of the articles are on the World Wide Web. These journals include: *Archives of Ophthalmology* (<http://www.ama-assn.org/public/journals/opht/opthome.htm>), the *American Journal of Ophthalmology* (<http://www.ajo.com>), and others.

Using the Internet Grateful Med interface, we enter "vitreous hemorrhage" and "aspirin" in the query terms, and we apply "clinical trial" to limit the publication types (**Figure**). Five articles are found, and after reviewing the abstracts, we choose to read "Effects of Aspirin Treatment on Diabetic Retinopathy."¹

EVALUATING THE EVIDENCE

Although, the quality of reports in peer-reviewed journals has been steadily increasing, the level of evidence varies from one report to the next. Some articles are published not because they provide good evidence, but because they are speculative and stimulate additional inquiries. Careful review requires the separation of conjecture from evidence.

Fortunately, neither a degree in epidemiology nor biostatistics is necessary to appraise the evidence in the peer-reviewed literature. To illustrate how judging articles can be easy, this article will focus on evaluating articles on treatment. As seen in the **Table**, the Evidence-Based Medicine Working Group attempts to judge the evidence by distilling it to the following 3 specific questions²: (1) Are the results valid? (2) What are the results? (3) Can the results be applied to my patients? We will address how these 3 value judgment questions relate to the hypothetical clinical question about the effects of aspirin treatment on diabetic retinopathy.

Are the Results Valid?

To appropriately interpret the results of a study, we have to assess how the methods used in the study would affect the results and conclusions. Flaws in study design can leave the results uninterpretable. Primary and secondary criteria can be developed to assist in the assessment of the literature. If an article fails the primary criteria, the information should not be considered evidence even if the article is the only report on the subject or question. Secondary criteria are additional requirements that help identify potential problems in study validity. Problems with secondary criteria may weaken the apparent validity of the results, but depending on the magnitude of the deficiency, the article can still be used as evidence.

The 2 primary criteria for assessing a treatment study are: (1) Were patients assigned to treatment and controls by randomization? Past experience has demonstrated that investigator enthusiasm can favor enrollment of a potential study patient into one treatment group vs another in subtle (often subconscious) but powerful ways. Investigators may subconsciously choose to only enroll younger or apparently more healthy patients into the treatment arm because they believe that these patients will be more likely to comply with the treatment and complete follow-up, or might be less likely to suffer adverse experiences from the treatment. On the surface, this may not seem to be a problem. However, if the disease is less severe in these patients, or if such patients are more likely to spontaneously recover, the results would then be biased. The results may only reflect the distribution of disease; the treatment group may seem to have better results because the patients in that group had less severe disease. Without randomization, there is no way to really know how potential differences in study groups will affect the results. Randomization tends to balance risk factors, both known and unknown, in the study groups. Larger study groups increase the likelihood that the risk factors will

Evaluating the Evidence

- What were the results?
 - What is the magnitude (size) of the treatment effect?
 - Does the treatment effect increase with higher doses?
 - Is there a biological mechanism for the treatment effect?
- Are the results valid?
 - Primary criteria
 - Were patients assigned to treatment and control by randomization?
 - Were all patients entered in the trial accounted for at the end of the study?
 - Secondary criteria
 - Was masking used?
 - Were the groups similar at the start of the trial?
 - Were the groups treated equally?
- What were the results?
 - To what degree could these results be due to bias, confounding, or chance?
- Can the results be applied to my patients?
 - Are the study patients similar to mine?
 - Were all important outcomes considered?
 - Are the treatment benefits greater than the risks?

be balanced. Nonrandomized studies can provide evidence, but the evidence should generally be considered weaker than that based on clinical trial results. In the hypothetical question on the effect of aspirin on vitreous hemorrhage, the article that was selected from our literature search reports the results of a large randomized clinical trial in which patients with diabetic retinopathy were randomly assigned to either aspirin at 650 mg per day, or placebo.

(2) Were all patients who entered the trial properly accounted for at the end of the study? Every patient who enters a clinical trial should be accounted for at its end. The greater the percentage of missing information at the end of the study, the more suspect the results. This is because patients who do not finish the study may have developed problems. For example, patients may have suffered an adverse event related to the treatment and decided to go elsewhere. Such patients might have an adverse experience that cannot be assessed because the information is missing. This is a particular problem because both the adverse experience and the missing information are related to the treatment. One conservative approach to assessing missing data is to attribute an adverse experience to all patients who had missing information at the end of the study. One might also consider how the results would be affected if the reason that persons in the treated arm did not return for follow-up was be-

cause of poor results, while the reason the control group had missing patients was because they had good results and decided that returning for study visits was not necessary. In Table 2 of "Effects of Aspirin Treatment on Diabetic Retinopathy," the investigators reported that 93% of all patients were accounted for.¹ The largest effect this missing information could have on a treatment difference is therefore 7%.

A truly randomized comparison must include all of the randomized individuals in the outcome assessment. Omitting subgroups because of missing information or failure to comply with treatment creates a nonrandomized subgroup analysis. It is often tempting to eliminate all patients who did not comply with the study treatment. However, even if the treatment was not taken, or if it was the opposite of the original assignment, the main analysis should be done according to the original treatment assignment. This is called an "intention to treat" analysis.

An example of the problems with these nonrandomized comparisons can be seen using data from the Coronary Drug Project.³ This study was designed to assess the safety and efficacy of several lipid-lowering drugs in patients with coronary heart disease. One of the drugs studied was clofibrate. The 5-year mortality rates in the 1103 patients assigned to clofibrate and in the 2789 patients assigned to placebo were 20% and 20.9%, respectively ($P = .55$). However, only about two thirds of the

clofibrate group were considered to be good adherers (taking 80% or more of the study drug) throughout the 5-year study period, and in this group, the 5-year mortality rate was 15%. This was substantially lower than the 24.6% 5-year mortality rate in the group that was not taking study medication ($P=.00011$). Based on these data, one could conclude that clofibrate markedly lowered the mortality rate. Interestingly, about two thirds of the placebo group were also considered to be good adherers to their study medication. In the placebo group, the good adherers also had a much lower 5-year mortality rate than the poor adherers (15.1% vs 28.3%, respectively [$P=4.7 \times 10^{-16}$]).

This demonstrates the danger of assessing the treatment effect in subgroups of patients. In this case, the lower mortality seen in the group adhering to the study medication was not a result of the medication, but rather associated with patient behavior. One can also easily see the problem in comparing outcomes in those who regularly took clofibrate with outcomes in the entire placebo group. Even the comparison between adherers in the treated and control groups is problematic, because there may be different motivations to adherence in the 2 groups that could bias the results. It is only the overall comparison that is truly a randomized comparison. This primary analysis is considered hypothesis testing. Other subgroup analyses may be interesting, but they are considered hypothesis generation.

Secondary criteria include the following: (1) Was masking used? Everyone who is involved with a study is likely to have an opinion, conscious or unconscious, as to what the results will show. Patients who know they are in the "treatment group" want the treatment to work, may complain less about their adverse effects, and may try harder when reading the eye chart. Study personnel who want the treatment to work may try harder to measure an improvement in the outcome variables with patients who got the treatment. One way to avoid this source of bias is to let neither the patient nor study personnel know which treatment was given to the pa-

tient. The reader should evaluate how well the study investigators tried to minimize this source of bias. Because a matched placebo tablet was used in the aspirin study, both the patients and the investigators were likely to be unaware of who was assigned to take aspirin.

(2) Were the groups similar at the start of the trial? We discussed earlier that imbalances in the distribution of prognosis affecting risk factors might affect the results. The reader should look at the imbalances and the size of the imbalances. Obviously, if the imbalance is large and the risk factor strongly affects the results, the reader has to be careful in interpreting the results. Although randomization increases that likelihood that factors will be balanced in the study groups, imbalances can occur. If imbalances occur despite randomization, the imbalances can be at least partially accounted for by performing an analysis that adjusts for the risk factor(s) that are not balanced. If both the adjusted and unadjusted analyses show the same results, then the reader can be more certain that the results are valid. The article under discussion did not report the baseline characteristics, but Table 5 of the accompanying article⁴ did compare age, duration of diabetes, type of diabetes, race, blood pressure, levels for serum lipids and hemoglobin A_{1c}, body weight, visual acuity, and level of retinopathy. There were no important or clinically significant differences ($P<.01$) between the aspirin and the placebo groups.

(3) Were the groups treated equally? Sometimes the control and treatment groups may be treated differently. If for example, there was reason to worry particularly about the control group (because they were not getting treatment) or about the treated group (because there may be some adverse effects from the treatment), the investigator may choose to follow that group more carefully. Although seemingly harmless, a group that is being observed more frequently may have more adverse events recorded, or they might be receiving better medical treatment. This ascertainment bias could have an important effect in assess-

ing the study results. The Early Treatment Diabetic Retinopathy Study article specified that all patients were treated similarly.⁴

What Were the Results?

If the methods of the article are valid, then it is appropriate to assess the results. The results should be examined for the magnitude of the treatment effect. A treatment effect that is dose dependent would confirm that the effect is related to the treatment. In addition, if the treatment has biological plausibility, the reader then can be further assured of the validity of the treatment effect. If there is no plausible biological mechanism for its actions, then the treatment effect might be questioned. Finally, confirmation of the results in other studies provides good evidence that the results are valid.

In the study, "Effects of Aspirin Treatment on Diabetic Retinopathy,"¹ the investigators were not able to find an effect. They reported a relative risk for development of vitreous hemorrhage (aspirin to placebo) of 1.05 (99% confidence interval, 0.81-1.36). The relative risk is the ratio of the risk in the intervention group divided by the risk in the control group. When the relative risk is 1.0, there is no difference between the risk of reaching the end point for patients assigned to the aspirin, and that of patients assigned to placebo. A relative risk substantially less than 1.0 indicates a reduced risk (in this case for the aspirin-treated group), while a relative risk substantially greater than 1.0 indicates an increased risk. A confidence interval that includes 1.0 indicates that the observed data are consistent with no difference between the 2 treatment groups. In this case, the relative risk of developing vitreous hemorrhage compared with placebo is 1.05, but the confidence interval includes 1.0. This suggests that aspirin has little or no effect. Clinical trials cannot assess whether or not 2 treatments are identical. There is always some uncertainty of the results. The tightness of the confidence interval (0.81-1.36) identifies the magnitude of this uncertainty. The "true" effect of aspirin on vitreous hemorrhage is likely to lie

between a 19% beneficial effect for aspirin and a 36% harmful effect on this particular outcome.

Can the Results Be Applied to My Patients?

One way to examine this issue is to assess whether patients similar to the reader's patients were well represented in the study. If they were, then the results are likely to apply. However, if there are differences, then clinical judgment is required to determine whether the differences are significant. If the differences are minor, then the results are also likely to apply. If substantive differences are present, then the reader should determine how the differences might affect the results. In "Effects of Aspirin Treatment on Diabetic Retinopathy,"¹ the inclusion criteria are broad, so the results should be broadly applicable.

After deciding about the types of patients covered by the study, the reader has to determine whether all clinically important outcomes were studied. In ophthalmology, visual acuity is an important outcome variable. If the visual acuity improves in the treatment group vs the control group, it is likely that the treatment is effective. Surrogate measures such as intraocular pressure can also be used if the surrogate measure has been previously shown to correlate with an outcome of interest such as visual field measurements or visual acuity. A recent clinical trial showing the risk of using surrogate end points is a clinical trial on vitamin A and retinitis pigmentosa.⁵ In that study, the authors conducted a well-designed and well-executed clinical trial. However, their outcome assessment was based on changes in electroretinograms, which some clinicians did not accept as a standard clinical measure of effectiveness. As a result, the study results have remained controversial and have not had the desired effect on clinical practice, despite the fact that the study investigators have offered further evidence that their results also applied to visual field measurements.⁶

The reader should also look for harmful effects. If a treatment shows efficacy but has significant adverse effects, the reader may be less likely to prescribe it. The only way to evaluate the adverse effects is to perform

an analysis of all the adverse experiences in the study. If the study fails to address adverse events or "quality-of-life" outcomes, the reader should be cautious about broadly applying the results. The assessment of the effect of macular hole surgery provides a good example of the need to consider both the benefits and the potential risks. Clinical trials have shown that the surgery is effective, but few studies have addressed the effect of the required postoperative positioning. After macular hole surgery, patients are asked to remain in the face-down position for at least 1 to 2 weeks. Although gain in visual acuity following surgery does occur, it does come at a cost. Face-down positioning requires a caretaker for meal preparation and household chores. In addition, socializing, watching television, and other activities have to be curtailed during this period. This reduction in quality of life is probably most difficult for older patients (the group most likely to be offered surgery), but its effect is rarely included in studies. The practitioner should consider and discuss both benefits and potential adverse effects with each patient.

Assessment

We asked the question, "does oral aspirin treatment affect vitreous hemorrhage in patients with diabetes mellitus?" After finding a list of articles, we decided to evaluate the article "Effects of Aspirin Treatment on Diabetic Retinopathy."¹ We decided that the methods were valid, that aspirin does not have an appreciable effect on the development of vitreous hemorrhage, and that the inclusion criteria of the study were sufficiently broad to make the study results generalizable. Based on this, we decided that the article provided good evidence that can be applied to answer our clinical question.

Using an approach similar to that described in this article, the American Academy of Ophthalmology's Ophthalmic Technology Assessment Committee (OTAC) reports the evidence on new and emerging procedures. Publications from OTAC often provide a solid first step at determining whether one should adopt a new procedure.

CONCLUSION

The practice of ophthalmology in the next century demands the inclusion of scientifically obtained evidence in management decisions. Patients, insurers, and fellow practitioners will demand such evidence for making treatment decisions. This review outlined the methods used to incorporate scientific evidence into an EGO practice. Practicing EGO is not simply using the published literature. Rather, the published literature is combed for evidence, and only articles that are of high quality are used to answer specific clinical questions.

To practice EGO, the first step is phrasing a question that can be answered by the available literature. After conducting a literature search, the retrieved articles are evaluated critically. The study design is considered for validity. The methods are evaluated to assess to what degree bias, confounding, or chance could have affected the results. The study results are examined, and the applicability of the results to the practitioner and the patient is assessed. Finally, all the information is synthesized and an assessment based on the benefits and risks is considered. This approach will help to maximize the chance for good patient outcomes.

Accepted for publication August 22, 2000.

Corresponding author and reprints: Donald S. Fong, MD, MPH, 1011 Baldwin Park Blvd, Baldwin Park, CA 91706 (e-mail: donald.s.fong@kpp.org).

REFERENCES

1. Early Treatment Diabetic Retinopathy Study Research Group. Effects of aspirin treatment on diabetic retinopathy: ETDRS report No. 8. *Ophthalmology*. 1991;98:757-765.
2. Guyatt GH, Sackett DL, Cook DJ, Evidence-Based Medicine Working Group, for the users' guides to the medical literature. II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA*. 1993;270:2598-2601.
3. The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med*. 1980;303:1038-1041.
4. Early Treatment Diabetic Retinopathy Study Research Group. Early treatment diabetic retinopathy study design and baseline characteristics: ETDRS report No. 7. *Ophthalmology*. 1991;98:741-756.
5. Berson EL, Rosner B, Sandberg MA, et al. A randomized trial of vitamin A and vitamin E supplementation for retinitis pigmentosa. *Arch Ophthalmol*. 1993;111:761-772.
6. Massof RW, Finkelstein D. Supplemental vitamin A retards loss of ERG amplitude in retinitis pigmentosa. *Arch Ophthalmol*. 1993;111:751-754.