

A Guide to Clinical Trials

CBC (INCLUDES DIFF/PLT)	
WHITE BLOOD CELL COUNT	4.8-10.8
RED BLOOD CELL COUNT	4.2-5.9
HEMOGLOBIN	12-16
HEMATOCRIT	37-47
MCV	86-106
MCH	31-36

THOUS/MCL	3.8-10.8
BILL/MCL	4.20-5.80
	13.2-17.1
	38.5-50.0
FL	80.0-100.0
	9-33.0

PART II: INTERPRETING MEDICAL RESEARCH

Part I of this two-part article, which appeared in the Summer 2005 issue of BETA, provided an overview of the clinical trial process. Part II covers features of clinical trials and interpretation of study results.

Clinical trials provide the foundation for evidence-based medicine, or medical decision-making guided by data from formal research. Medical professionals keep up with the latest information by reading peer-reviewed medical journals and attending conferences. Likewise, HIV positive people can keep abreast of the state of the art by following the medical literature and community publications like *BETA*.

Trials offer important information about a therapy's benefits and risks in a population, but they cannot predict how well a given treatment will work for a specific person. Healthcare providers, therefore, must still rely heavily on clinical experience, intuition, and a careful evaluation of the various factors unique to each individual case—the practice of medicine remains an art as well as a science.



13.4
183
2129
1236
456
51
27
54
31
1
0

32.0-36.0
11.0-15.0
140-400
1500-7800
850-3900
200-950
10-300
0-20



HIV-1 RNA, Q
hdNA, 3RD
COPIES/ML

Characteristics of Medical Research

In the hierarchy of medical research, some types of studies are regarded as more credible than others. Research is considered most valid when it focuses on events in progress rather than those that have already occurred, includes enough participants observed over a long enough period so that the results are **statistically significant** (not likely to be due to chance alone), and takes steps to reduce the influence of confounding factors and minimize bias on the part of investigators and subjects.

Retrospective vs Prospective Studies

Retrospective studies look back at events that happened in the past, often using medical records. In **prospective studies**, a group of subjects is selected and followed forward in time. Retrospective studies are considered less reliable because it is more difficult to control (or even recognize) potential confounding factors when looking at past events. For example, it would not be very useful to compare the results from a recent study of atazanavir (Reyataz) with data from an early-1990s trial of a first-generation protease inhibitor such as indinavir (Crixivan), because both the nature of HIV disease and the standard of care have changed so much in the intervening decade.

In addition, important pieces of information may be unavailable when looking back over time. For instance, medical records dating back to the early years of the epidemic would not include HIV viral load measurements, since this test was not widely used until the mid-1990s.

Study Size and Length

Other factors being equal, longer trials with larger **sample sizes**—that is, more participants—are considered more reliable than shorter studies with fewer subjects. Longer and larger trials produce more data, making it less likely that the observed outcome is simply due to chance. The ability of a

study to produce statistically significant results is known as its **power**.

A report of the natural history of a disease and its treatment in a specific individual or a small group of patients is called a **case report** or a **case series**, respectively. Case reports often describe exceptional or unusual events and have the benefit of speed; as such, they may uncover uncommon side effects (such as heart problems associated with protease inhibitors) before they are revealed in clinical trials. This type of **anecdotal evidence** may be interesting, but is not considered conclusive because it is impossible to know how individual factors may have influenced the observed events.

Case-control studies provide an additional level of reliability. In these studies, each person with the variable under study (a case) is matched with one or more individuals with otherwise similar characteristics (a control). This matching makes it easier to discern the effect of a particular variable by ensuring that cases and controls are alike in other respects.

In a **cohort study**, a group of individuals with shared characteristics is selected and followed forward in time, typically for many years. The HIV Outpatient Study (HOPS), the Multicenter AIDS Cohort Study (MACS), the Women's Interagency HIV Study (WIHS), and the Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) study are all examples of cohort studies looking at various populations with HIV/AIDS. In this type of study, researchers do not perform a specific intervention (such as administering a particular drug), but rather observe the effect of various factors (e.g., demographic characteristics, type of therapy, hepatitis C coinfection) on the natural history of the disease over time.

Clinical trials are carefully planned studies looking at particular therapeutic interventions. The process proceeds in phases, with each successive stage lasting longer and including

more subjects (this is covered in “Part I: Understanding Clinical Studies”). This is done to achieve a trade-off of safety and credibility. Only small numbers of participants are exposed to potentially risky new agents during Phase I trials. After a drug is shown to be generally safe, large numbers of subjects are included in Phase III trials to obtain more reliable data on **efficacy** (how well it works).

Once several studies have been done looking at a particular therapy, researchers may conduct a **systematic review** (comprehensive overview of related studies) or a **meta-analysis** (mathematical analysis that incorporates data from multiple studies). These secondary studies provide a “big picture” summary of information amassed so far. If several well-designed trials produce similar results, confidence in the outcome is enhanced.

Controlled Studies

Clinical trials with large sample sizes and long follow-up periods provide stronger evidence than single case reports or case-control studies, but may still leave room for **bias** (“favoritism” or “prejudice” that skews an outcome in a systematic way) and **confounding factors** (extraneous variables that can distort a trial's outcome).

Various strategies are employed to minimize conscious or unconscious influences that could unfairly affect a trial's results. The “gold standard” for research on medical interventions is the prospective, **double-blind**, randomized, controlled trial. Briefly, *double-blind* means that neither the investigators nor the subjects know who is receiving the experimental agent. **Randomization** refers to the process of assigning subjects by chance to the various treatment arms. This is done to help ensure that at the outset of the trial the subjects in the various arms are comparable, or as similar as possible in every respect except for the type of intervention they are receiving. A **controlled trial** is one in which the experimental

agent is compared against something else, either a placebo (inactive or mock therapy) or an existing effective treatment (these characteristics are described in more detail in Part I).

Statistics 101

Investigators are not always able to design and implement randomized controlled trials to test every hypothesis. For instance, it would be unethical to randomly assign HIV positive pregnant women either to give birth vaginally or undergo a cesarean section (c-section) to see which method results in a lower rate of mother-to-child HIV transmission. The best researchers can do is compare the HIV status of infants who happened to be born vaginally or through c-section, but these groups might differ in other ways (e.g., perhaps women who had c-sections were less likely to have received prenatal care).

Fortunately, researchers can use various statistical methods to make adjustments for systematic (consistent and predictable) differences between groups of subjects. For example, it is a common finding that individuals coinfect with HIV and hepatitis C virus (HCV) are more likely to be injection drug users and tend to be younger than people with HIV alone. It is also known that HCV-related liver damage increases with age and that older individuals tend to respond less well to interferon-based therapy. Therefore, investigators must adjust for the subjects' age if they are attempting to determine whether coinfection is associated with liver disease progression or response to anti-HCV treatment. Investigators can also **stratify** their data to look separately at subgroups with different confounding characteristics.

Another statistical concern related to clinical studies—especially those that include representative “real world” populations—is that raw data are rarely “clean,” or free of potentially confounding influences. Investigators often must take multiple coexisting factors into account. Looking again at hepatitis C, it is known that, along

with older individuals, men tend to respond less well to interferon than women, and African-Americans respond less well than whites. Thus, researchers looking at the relative benefits of two different interferon-based regimens would need to use mathematical models that account for how all these variables interact to influence the observed outcome. It is not uncommon that a factor that initially seems important in a **univariate** analysis that looks at a single variable alone will no longer appear relevant when a **multivariate** analysis is performed to account for multiple interacting variables.

Statistical Significance

As noted above, study results are considered statistically significant if there is little likelihood that the observed outcome was due to chance alone. When looking at data from different arms of an interventional clinical trial, researchers attempt to determine whether the **null hypothesis**—the assumption that the various interventions are equally effective—is true or false.

Researchers use the **P value** to indicate the probability that an observed result is true and not just due to happenstance (for example, that an experimental agent really works, not just that more of the subjects who took it had the good luck to improve). While studies may use different cut-off values, a P value below 0.05 ($p < 0.05$) is traditionally accepted as an indication of statistical significance. This means that the likelihood is less than 5%, or 1 in 20, that the observed difference between study arms was simply due to chance. Smaller P values indicate even greater certainty. A P value below 0.01 ($p < 0.01$)—considered “highly statistically significant”—means that there is less than a 1% probability that the observed outcome would have occurred by chance alone.

While the P value provides a single cut-off for statistical significance, the **confidence interval** (CI) provides a range within which the true result is

likely to fall. Researchers traditionally use a 95% CI, meaning that there is a 95% likelihood that the actual difference lies within this range. Studies with higher power (e.g., more subjects) typically produce narrower confidence intervals, meaning researchers can feel more certain about the accuracy of their results. The actual values included in a CI also convey useful information. In an interventional trial, if the null hypothesis were true, the difference between two treatments under study (or treatment and placebo) would be zero. Thus, if a CI includes zero, researchers cannot rule out the possibility that the interventions were equally effective. In trials looking at risk factors for a condition, a **relative risk** or **odds ratio** (OR) of 1 means that a factor had no effect; in this case, if a CI includes 1, researchers cannot rule out the possibility that the risk factor had no impact on the condition's occurrence.

Interpreting Significance

All studies yield a certain level of **false positive** and **false negative** data. For instance, an experimental drug may seem to work for a particular subject even though it is, in fact, ineffective overall; conversely, an agent may not help a specific subject even though it is effective overall. The goal in a well-designed study is for these types of subject-specific variability to cancel each other out, so that any actual benefit of an intervention will become apparent. Failure to detect a true difference between interventions is known as a **type II** error, while erroneously finding a difference between two interventions that are in fact equally effective is called a **type I** error.

If the difference between study arms is statistically significant—that is, the P value is larger than the chosen cut-off value and/or the CI does not include zero—the investigators can be reasonably confident that the null hypothesis is false and that one intervention really is superior to another. In real world terms, if the

The “gold standard” for the presentation of medical research is publication in a peer-reviewed professional journal.

observed difference in HIV viral load suppression between two study arms receiving two different drug regimens is statistically significant, this suggests that one regimen really does work better.

If the observed difference is not statistically significant, it could be that the two regimens have about the same efficacy (or lack thereof). But it could also mean that the study was **underpowered** or too small to demonstrate an effect. Larger and longer-lasting studies—those with higher power—are more likely to produce significant results. Studies with low power produce wide CIs, meaning the true result could lie within a broad range. Statistical tools are available to help investigators determine in advance how large a sample size they will need to detect a true difference between study groups.

Reporting Study Results

After a clinical trial is completed, investigators typically present their research results to their colleagues. The two main venues for disseminating data from medical research are scientific meetings and professional journals.

Scientific Conferences

Often researchers first publicly present their findings at conferences devoted to their fields of study. Important HIV/AIDS meetings include the Conference on Retroviruses and Opportunistic Infections held each winter, the Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) held each fall, the biannual International AIDS Conference, and various smaller gatherings—sometimes devoted to specific topics

like drug resistance or complications of therapy—organized by groups such as the International AIDS Society, the British HIV Association, and the European AIDS Clinical Society. In addition, pharmaceutical companies commonly sponsor meetings to present the latest research on their experimental drugs.

The most interesting or groundbreaking studies are usually presented orally by one of the authors, often accompanied by slides. While study abstracts are typically submitted months in advance, important last-minute results are sometimes included as “late-breakers.” Research that is not selected for oral sessions may be presented on posters. Abstracts from both oral and poster presentations are typically published in a catalog and may also be made available on the web.

Medical Journals

The “gold standard” for the presentation of medical research is publication in a peer-reviewed professional journal. Journal editors send out submitted articles for review, usually by 1–4 selected colleagues who work in the same field, to ensure that the study appears well designed, the methods sound, and the data plausible.

There are several “tiers” of journals that publish medical research, from general science magazines like *Science* and *Nature*; to broad medical publications such as *The Lancet*, *Journal of the American Medical Association*, and *New England Journal of Medicine*; to specialized journals such as *AIDS*, *Clinical Infectious Diseases*, and *Journal of Virology*. Medical journal articles adhere to a basic standard format and usually include the following elements:

- **Abstract/Summary.** A short synopsis of a research article laying out the objective or goal of the study, the trial design and methods, a summary of the results obtained (and usually their statistical significance), and the authors’ conclusion or “take home” message.
- **Introduction/Background.** This usually includes a statement summarizing the problem or issue to be investigated, a brief review of what is known to date (with references to key literature), the rationale for the study (why was it done?), and the hypothesis (what did the authors hope to show?).
- **Design and Methods.** These sections (which may be combined) provide in-depth information about how the study was designed and carried out, including a detailed description of the study population, which treatment(s) were used, which tests were performed, and how data was collected and analyzed.
- **Results.** This section gives a detailed description of the data collected by the researchers and the results of their statistical analyses, often including tables, charts, and graphs.
- **Discussion.** In this section the authors interpret their results, draw their conclusions, and discuss what their findings mean—for example, whether the initial hypothesis was confirmed, how the results might affect clinical practice, potential limitations of the study, and suggestions for further research.

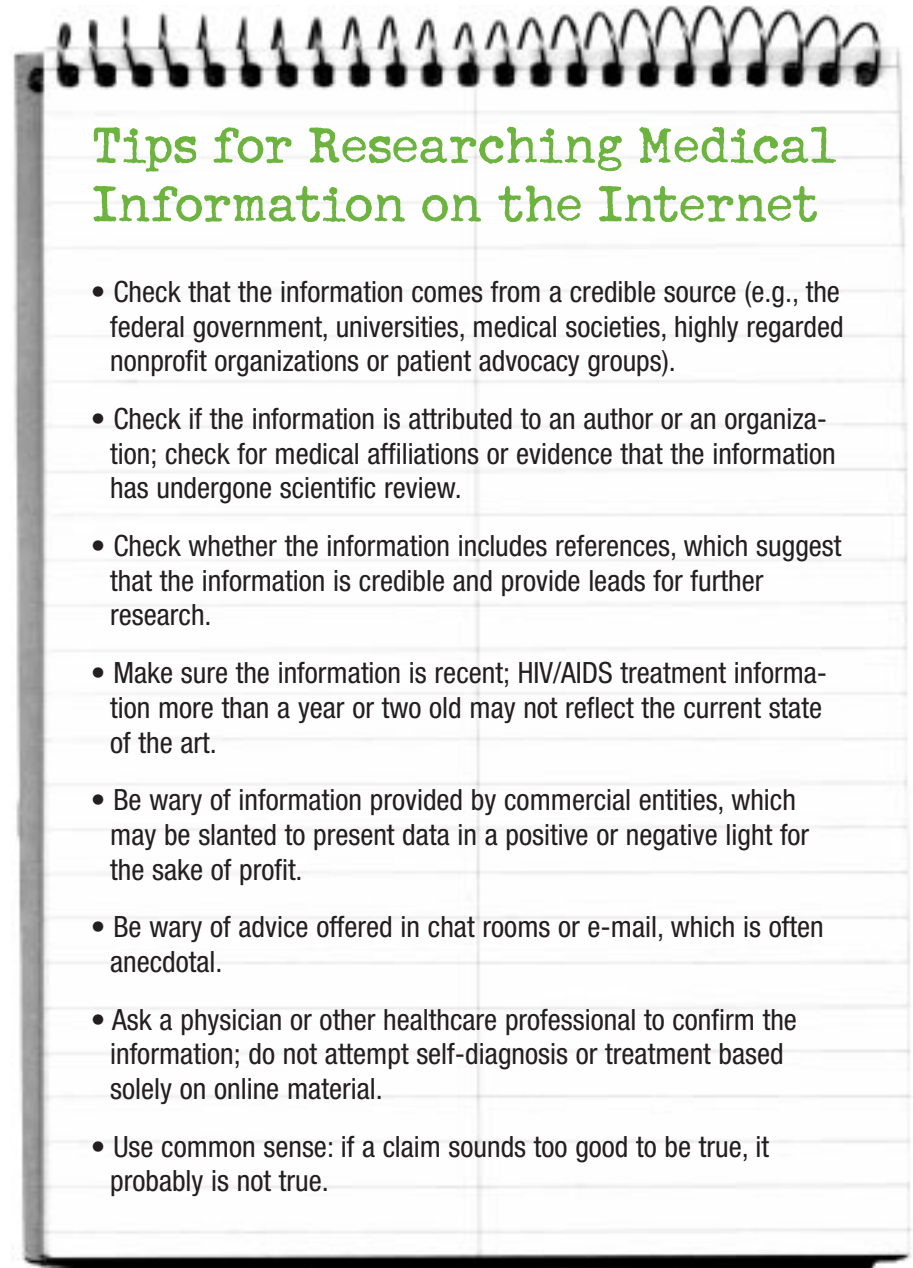
Finding Useful Research Results

With improvements in information technology—and a shift away from the notion that medical professionals are unquestionable authorities—a growing number of people have taken an interest in exploring medical research for themselves. But just because a great deal of medical information is available on the Internet and elsewhere does not suggest that all of it is credible.

The most comprehensive source of peer-reviewed medical literature is MEDLINE, a database of more than 10 million references. MEDLINE can be accessed online using a variety of front-end tools including PubMed, a search service provided by the U.S. National Library of Medicine. With so much information available, entering a broad search term like “HIV” can feel like taking a drink from a fire hose. Users will obtain more useful and relevant results using narrower search criteria, for example a specific drug name or side effect.

MEDLINE provides free access to research abstracts, but users often must dig up actual medical journals to obtain the full text of articles. University and medical center libraries carry the most popular, reputable medical journals, and usually a selection of smaller, more specialized ones as well. Although intended for use by students and staff, some university libraries allow members of the public to access their collections. Most medical journals are available on the web, but generally offer only abstracts for free. Some provide immediate free full-text access to studies deemed particularly important or groundbreaking, and others offer full access to issues that are more than six months or a year old.

Other good online sources of medical information include sites sponsored by the federal government (e.g., National Institutes of Health, Centers for Disease Control and Prevention), universities (e.g., HIV InSite, Johns Hopkins AIDS Service), medical societies (e.g., International AIDS Society, American Society for Microbiology), nonprofit organizations (e.g., American Liver Foundation, American Heart Association), and patient advocacy and support groups (e.g., San Francisco AIDS Foundation, Project Inform, Hepatitis C Support Project). There are several independent sources of high-quality HIV/AIDS information supported by pharmaceutical companies, including AIDSmeds.com, The Body, and HIVandHepatitis.com.



Tips for Researching Medical Information on the Internet

- Check that the information comes from a credible source (e.g., the federal government, universities, medical societies, highly regarded nonprofit organizations or patient advocacy groups).
- Check if the information is attributed to an author or an organization; check for medical affiliations or evidence that the information has undergone scientific review.
- Check whether the information includes references, which suggest that the information is credible and provide leads for further research.
- Make sure the information is recent; HIV/AIDS treatment information more than a year or two old may not reflect the current state of the art.
- Be wary of information provided by commercial entities, which may be slanted to present data in a positive or negative light for the sake of profit.
- Be wary of advice offered in chat rooms or e-mail, which is often anecdotal.
- Ask a physician or other healthcare professional to confirm the information; do not attempt self-diagnosis or treatment based solely on online material.
- Use common sense: if a claim sounds too good to be true, it probably is not true.

Pharmaceutical company web sites can provide useful information (in particular, full prescribing information for specific drugs) but beware of bias. To address concerns that unfavorable study data about experimental drugs have not been widely available, the industry trade group Pharmaceutical Research and Manufacturers of America (PhRMA) recently launched an online repository of published and unpublished clinical trial results at www.clinicalstudyresults.org. (See the sidebar on this page for tips on

locating credible medical information on the Internet).

Cautions to Keep in Mind

When reading medical literature, there are several potential pitfalls to keep in mind. Researchers understandably wish to produce interesting and groundbreaking results, and may have a conscious or unconscious tendency to make their findings appear more promising or more conclusive than they actually are. Likewise, journal editors want to publish studies

that attract readers and advance the state of the science.

These motivations contribute to a phenomenon known as **publication bias**, whereby studies are more likely to be published if they produce positive results (not necessarily “good” results, but those that confirm the investigators’ initial hypotheses). Studies that produce negative results—for example, that an investigational agent works no better than existing therapies—are less likely to see the light of day.

A potentially more serious concern is the desire of pharmaceutical companies to make their experimental therapies look as good as possible in the hopes of obtaining FDA approval and the large profits a successful anti-HIV drug could bring. This concern extends to researchers who have financial ties to the pharmaceutical industry or a personal financial stake in the outcome of a clinical trial (for example, owning stock in a drug company or holding a patent on an experimental agent).

Post-Hoc Manipulations

Researchers may attempt to make their findings appear more positive by essentially changing the rules after the game has started (*post hoc* means “after the fact”). That is, they may fail to follow the procedures and methods set forth in their original protocol. Sometimes this is unavoidable; for example, it is not uncommon for researchers to broaden inclusion criteria, remove exclusion criteria, or settle for a smaller sample size because they have trouble recruiting enough suitable subjects.

One common way researchers may deviate from their original protocol is to analyze subjects based on the treatment they actually received rather than the one they were initially assigned. This is the difference between an **as-treated** and an **intent-to-treat** analysis. An intent-to-treat analysis uses data from all participants who were initially enrolled in a given study arm, whether or not they

stayed on the assigned treatment. An as-treated analysis excludes subjects who ended up not receiving the originally assigned intervention for the intended length of time, often because they withdrew from the study prematurely (for example, due to intolerable side effects) or because the treatment was not working. Using an as-treated analysis presents problems because subjects who do not receive their assigned treatment for the full, specified period tend to differ in systematic ways from those who remain on their assigned therapy the whole time (known as **exclusion** or **attrition** bias).

For example, say a study includes two arms, each with 100 subjects, randomly assigned to receive two different drug regimens for 48 weeks. In arm A, 90 subjects remain on their assigned regimen for the initially specified period and 60 of these achieve an undetectable HIV viral load (a response rate of about 66%). In arm B, 50 subjects remain on their assigned regimen for the whole time and 40 of these achieve an undetectable viral load (a response rate of 80%). But the other 50 participants drop out of the study early because they are unable to tolerate the side effects of the experimental drugs. In this case, it would not be fair to conclude that regimen B is superior to regimen A, because its usefulness is limited by a high rate of toxicity.

Researchers may also be tempted to exclude from their analysis subjects who fail to achieve good adherence. Perhaps regimen B appears more effective in those who actually take it as directed, but it is much less convenient (e.g., more pills per day or stricter food requirements), resulting in poor adherence.

To avoid this pitfall, researchers should account for all subjects who were initially assigned to a given study arm, whether or not they continued to receive the intended intervention for the entire period. At the very least, if the investigators analyze only those participants who successfully com-

pleted the initially specified course of therapy, they should clearly state that they performed an as-treated analysis. In practice, it is not uncommon for researchers to provide both intent-to-treat and as-treated results, especially if they differ considerably; typically, as-treated data make an intervention look more promising than the corresponding intent-to-treat results.

Another way researchers may attempt to find a silver lining in negative results is to perform various unplanned subgroup analyses. For instance, they may “mine” or “dredge” their data to see if anything promising turns up. This is a problem because in any study, some positive association for some subgroup of subjects is likely to occur by chance alone. A possible example of inappropriate subgroup analysis came to light in February 2003 when VaxGen announced that its AIDS VAX vaccine appeared to protect African-American subjects from contracting HIV, although this effect was not seen in the study population as a whole. Critics argued that the racial subgroup was too small (314 blacks out of 5,009 “high-risk” volunteers) and that the researchers failed to make appropriate adjustments in their analysis, thus overestimating the statistical significance of the results. To circumvent such concerns, investigators should specify at the outset what they are looking for, what types of analyses they plan to conduct, and how they will stratify their subjects.

Finally, it is not unknown for researchers to perform a variety of unusual and esoteric statistical tests in an attempt to turn up something worth reporting. If the authors of a study include statistical manipulations that are not the norm for the type of trial they conducted, they should clearly explain why.

Living in the Real World

When exploring medical literature, it is important to bear in mind

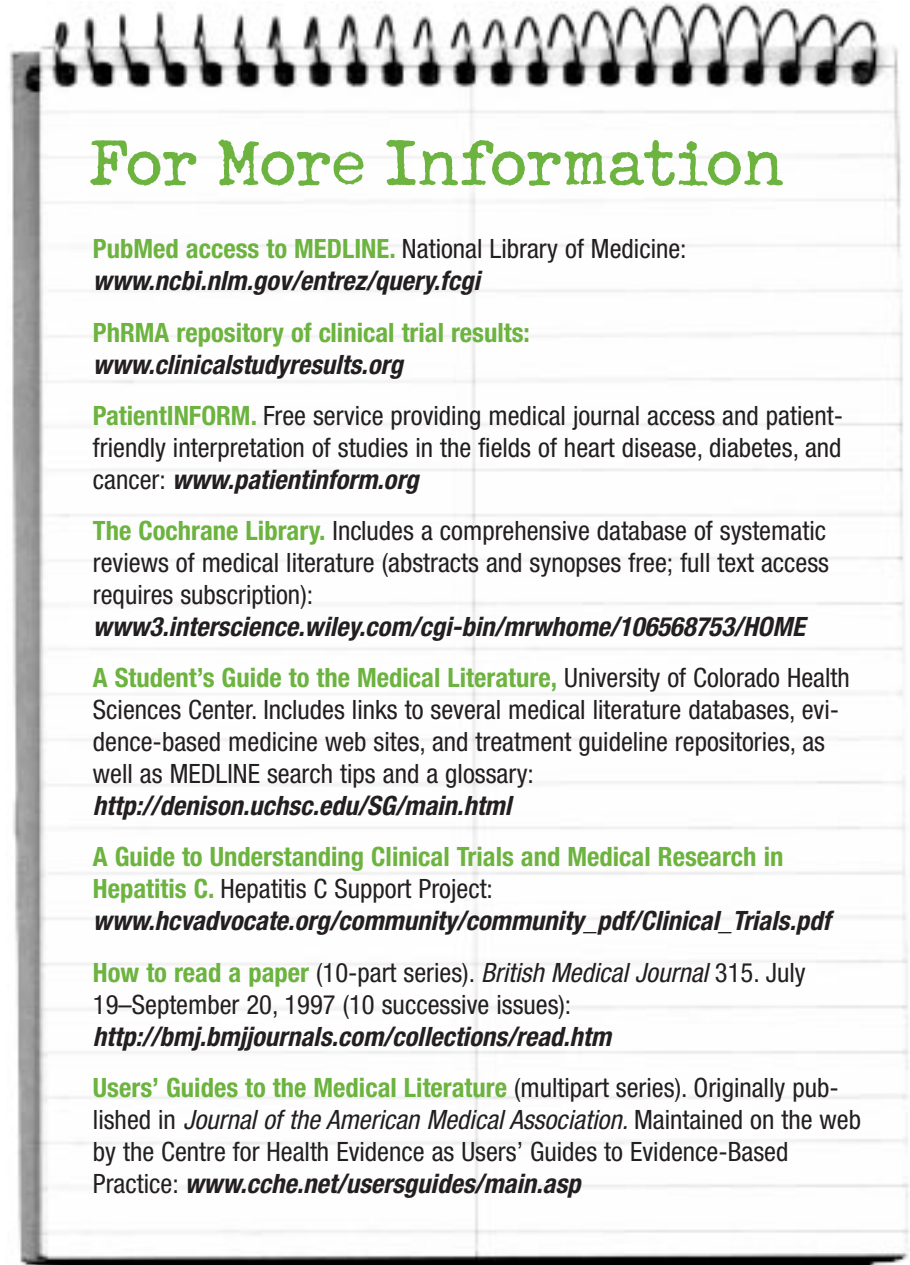
that statistical significance does not always imply clinical significance. A study might find, for example, that drug A increases subjects' CD4 cell counts by 5 cells/mm³ while drug B increases CD4 counts by 10 cells/mm³. If the study was large enough, this difference might prove statistically significant, but it may still be essentially meaningless in terms of the actual clinical benefit it offers to people with HIV.

In addition, it is important to think about whether the results of a study can be **generalized** to other people with the disease. For example, if an HIV trial had strict exclusion criteria (not admitting individuals with HCV coinfection or a history of injection drug use or psychiatric conditions, say), or was unattractive to a certain subset of potential subjects (such as requiring frequent clinic visits to a remote location, such that low-income women with children found it impossible to participate), its outcome—no matter how promising—might be essentially irrelevant to a large proportion of the HIV positive population.

Conclusion: Making the Most of Medical Research

Medical research is typically written for professionals, and can be difficult for nonspecialists to comprehend. However, with a basic grasp of clinical trial design and statistics, a good medical dictionary or glossary (available online), awareness of a few common pitfalls, and a bit of practice, most people should be able to understand the language of medical literature.

Recent events—including the soaring cost of prescription medications and controversy over the safety of certain classes of FDA-approved drugs—have focused unprecedented attention on medical research in the past year. The FDA has been accused of being too lax in demanding that drug companies conduct the required post-marketing studies to ensure that their products are safe over the long-term. The medical publishing field has been criticized for its propensity



For More Information

PubMed access to MEDLINE. National Library of Medicine:
www.ncbi.nlm.gov/entrez/query.fcgi

PhRMA repository of clinical trial results:
www.clinicalstudyresults.org

PatientINFORM. Free service providing medical journal access and patient-friendly interpretation of studies in the fields of heart disease, diabetes, and cancer: www.patientinform.org

The Cochrane Library. Includes a comprehensive database of systematic reviews of medical literature (abstracts and synopses free; full text access requires subscription):
www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

A Student's Guide to the Medical Literature, University of Colorado Health Sciences Center. Includes links to several medical literature databases, evidence-based medicine web sites, and treatment guideline repositories, as well as MEDLINE search tips and a glossary:
<http://denison.uchsc.edu/SG/main.html>

A Guide to Understanding Clinical Trials and Medical Research in Hepatitis C. Hepatitis C Support Project:
www.hcvadvocate.org/community/community_pdf/Clinical_Trials.pdf

How to read a paper (10-part series). *British Medical Journal* 315. July 19–September 20, 1997 (10 successive issues):
<http://bmj.bmjournals.com/collections/read.htm>

Users' Guides to the Medical Literature (multipart series). Originally published in *Journal of the American Medical Association*. Maintained on the web by the Centre for Health Evidence as Users' Guides to Evidence-Based Practice: www.cche.net/usersguides/main.asp

to publish mostly positive studies, and proposals have been put forth to make all clinical trial results available for free over the Internet. There has been increased scrutiny of conflicts of interest within the medical research and regulatory establishments, leading to stricter rules about consulting arrangements and stock ownership. Finally, concern has arisen about how the pharmaceutical industry shapes the overall research agenda by focusing on “lifestyle” and “me too” drugs that stand to produce large profits,

rather than therapies that will provide the most good for the most people.

As researchers, clinicians, regulators, and politicians sort out these difficult issues, people affected by HIV and other disease can protect themselves by arming themselves with knowledge.

A Guide to Clinical Trials, Part I and II, were prepared for the San Francisco AIDS Foundation by Liz Highleyman.