A Simple Method for Evaluating the Clinical Literature

Robert J. Flaherty, MD

The "PP-ICONS" approach will help you separate the clinical wheat from the chaff in mere minutes.

Keeping up with the latest advances in diagnosis and treatment is a challenge we all face as phycians. We need information that is both valid (that is, accurate and correct) and relevant to our patients and practices. While we have many sources of clinical information, such as CME lectures, textbooks, pharmaceutical advertising, pharmaceutical representatives and colleagues, we often turn to journal articles for the most current clinical information.

Unfortunately, a great deal of research reported in journal articles is poorly done, poorly analyzed or both, and thus is not valid. A great deal of research is also irrelevant to our patients and practices. Separating the clinical wheat from the chaff can take skills that many of us never were taught.

The article "Making Evidence-Based Medicine Doable in Everyday Practice" in the February 2004 issue of *FPM* describes several organizations that can help us. These organizations, such as the Cochrane Library, Bandolier and Clinical Evidence, develop clinical questions and then review one or more journal articles to identify the best available evidence that answers the question, with a focus on the quality of the study, the validity of the results and the relevance of the findings to everyday practice. These organizations provide a very valuable service, and the number of important clinical questions that they have studied has grown steadily over the past five years. (See "Four steps to an evidence-based answer" on page 48.)

If you find a systematic review or meta-analysis done by one of these organizations, you can feel confident that you have found the current best evidence. However, these organizations have not asked all of the common clinical questions yet, and you will frequently be faced with finding the pertinent articles and determining for yourself whether they are valuable. This is where the PP-ICONS approach can help.

What is PP-ICONS?

When you find a systematic review, meta-analysis or randomized controlled trial while reading your clinical journals or searching PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), you need to determine whether it is valid and relevant. There are many different ways to analyze an abstract or journal article, some more rigorous than others.[1,2] I have found a simple but effective way to identify a valid or relevant article within a couple of minutes, ensuring that I can use or discard the conclusions with confidence. This approach works well on articles regarding treatment and prevention, and can also be used with articles on diagnosis and screening. ➤

Dr. Flaherty is a family physician with the Student Health Service and WWAMI Medical Program, Montana State University-Bozeman, and a clinical professor in the Department of Family Medicine, University of Washington. Conflicts of interest: none reported.

ILLUSTRATION BY GREG CLARKE

◄ ➤

Physicians often turn to journal articles to find the most current clinical information; however, much research is not valid or relevant.

◄ ➤

The PP-ICONS approach can help physicians quickly determine whether a particular study is valuable.

◄ ➤

PP-ICONS stands for problem, patient or population, intervention, comparison, outcome, number of subjects and statistics.

◄ ➤

Writing down your clinical question can help you clarify exactly what you are looking for in the clinical literature.

The most important information to look for when reviewing an article can be summarized by the acronym "PP-ICONS," which stands for the following:
- Problem,
- Patient or population,
- Intervention,
- Comparison,
- Outcome,
- Number of subjects,
- Statistics.

For example, imagine that you just saw a nine-year-old patient in the office with common warts on her hands, an ideal candidate for your usual cryotherapy. Her mother had heard about treating warts with duct tape and wondered if you would recommend this treatment. You promised to call Mom back after you had a chance to investigate this rather odd treatment.

When you get a free moment, you write down your clinical question: "Is duct tape an effective treatment for warts in children?" Writing down your clinical question is useful, as it can help you clarify exactly what you are looking for. Use the PPICO parts of the acronym to help you write your clinical question; this is actually how many researchers develop their research questions.

You search Cochrane and Bandolier without success, so now you search PubMed, which returns an abstract for the following article: "Focht DR 3rd, Spicer C, Fairchok MP. The efficacy of duct tape vs cryotherapy

KEY POINTS

- Reading the abstract is often sufficient when evaluating an article using the PP-ICONS approach.

- The most relevant studies will involve outcomes that matter to patients (e.g., morbidity, mortality and cost) versus outcomes that matter to physiologists (e.g., blood pressure, blood sugar or cholesterol levels).

- Ignore the relative risk reduction, as it overstates research findings and will mislead you.

in the treatment of verruca vulgaris (the common wart). *Arch Pediatr Adolesc Med.* 2002 Oct;156(10):971-974."

You decide to apply PP-ICONS to this abstract (shown on page 49) to determine if the information is both valid and relevant.

Problem. The first P in PP-ICONS is for "problem," which refers to the clinical condition that was studied. From the abstract, it is clear that the researchers studied the same problem you are interested in, which is important since flat warts or genital warts may have responded differently. Obviously, if the problem studied were not sufficiently similar to your clinical problem, the results would not be relevant.

Patient or population. Next, consider the patient or population. Is the study group similar to your patient or practice? Are they primary care patients, for example, or are they patients who have been referred to a

FOUR STEPS TO AN EVIDENCE-BASED ANSWER

When faced with a clinical question, follow these steps to find an evidence-based answer:

1. Search the Web site of one of the evidence review organizations, such as Cochrane (http://www.cochrane.org/cochrane/revabstr/mainindex.htm), Bandolier (http://www.jr2.ox.ac.uk/bandolier) or Clinical Evidence (http://www.clinicalevidence.com), described in "Making Evidence-Based Medicine Doable in Everyday Practice," *FPM*, February 2004, page 51. You can also search the TRIP+ Web site (http://www.tripdatabase.com), which simultaneously searches the databases of many of the review organizations. If you find a systematic review or meta-analysis by one of these organizations, you can be confident that you've found the best evidence available.

2. If you don't find the information you need through step 1, search for meta-analyses and systematic reviews using the PubMed Web site (see the tutorial at http://www.nlm.nih.gov/bsd/pubmed_tutorial/m1001.html). Most of the recent abstracts found on PubMed provide enough information for you to determine the validity and relevance of the findings. If needed, you can get a copy of the full article through your hospital library or the journal's Web site.

3. If you cannot find a systematic review or meta-analysis on PubMed, look for a randomized controlled trial (RCT). The RCT is the "gold standard" in medical research. Case reports, cohort studies and other research methods simply are not good enough to use for making patient care decisions.

4. Once you find the article you need, use the PP-ICONS approach to evaluate its usefulness to your patient.

ABSTRACT FROM PUBMED

Using the PP-ICONS approach, physicians can evaluate the validity and relevance of clinical articles in minutes using only the abstract, such as this one, obtained free online from PubMed, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi. The author uses this abstract to evaluate the use of duct tape to treat common warts.

OBJECTIVE: To determine if application of duct tape is as effective as cryotherapy in the treatment of common warts. DESIGN: A prospective, randomized controlled trial with 2 treatment arms for warts in children. SETTING: The general pediatric and adolescent clinics at a military medical center. PATIENTS: A total of 61 patients (age range, 3-22 years) were enrolled in the study from October 31, 2000, to July 25, 2001. 51 patients completed the study and were available for analysis. INTERVENTION: Patients were randomized using computer-generated codes to receive either cryotherapy (liquid nitrogen applied to each wart for 10 seconds every 2-3 weeks) for a maximum of 6 treatments or duct tape occlusion (applied directly to the wart) for a maximum of 2 months. Patients had their warts measured at baseline and with return visits. MAIN OUTCOME MEASURE: Complete resolution of the wart being studied. RESULTS: Of the 51 patients completing the study, 26 (51%) were treated with duct tape, and 25 (49%) were treated with cryotherapy. Twenty-two patients (85%) in the duct tape arm vs 15 patients (60%) enrolled in the cryotherapy arm had complete resolution of their warts (P = .05 by chi2 analysis). The majority of warts that responded to either therapy did so within the first month of treatment. CONCLUSION: Duct tape occlusion therapy was significantly more effective than cryotherapy for treatment of the common wart.

tertiary care center? Are they of a similar age and gender? In this case, the researchers studied children and young adults in outpatient clinics, which is similar to your patient population. If the patients in the study are not similar to your patient, for example if they are sicker, older, a different gender or more clinically complicated, the results might not be relevant.

Intervention. The intervention could be a diagnostic test or a treatment. Make sure the intervention is the same as what you are looking for. The patient's mother was asking about duct tape for warts, so this is a relevant study.

Comparison. The comparison is what the intervention is tested against. It could be a different diagnostic test or another therapy, such as cryotherapy in this wart study. It could even be placebo or no treatment. Make sure the comparison fits your question. You usually use cryotherapy for common warts, so this is a relevant comparison.

> We have long assumed that improving the physiologic parameters of a disease will result in a better disease outcome, but that is not necessarily true.

Outcome. The outcome is particularly important. Many outcomes are "disease-oriented outcomes," which are based on "disease-oriented evidence" (DOEs). DOEs usually reflect changes in physiologic parameters, such as blood pressure, blood sugar, cholesterol, etc. We have long assumed that improving the physiologic parameters of a disease will result in a better disease outcome, but that is not necessarily true. For instance, finasteride can improve urinary flow rate in prostatic hypertrophy, but it does not significantly change symptom scores.[3]

DOEs look at the kinds of outcomes that physiologists care about. More relevant are outcomes that patients care about, often called "patient-oriented outcomes." These are based on "patient-oriented evidence that matters" (POEMs) and look at outcomes such as morbidity, mortality and cost. Thus, when looking at a journal article, DOEs are interesting but of questionable relevance,

whereas POEMs are very interesting and very relevant. In the study on the previous page, the outcome is complete resolution of the wart, which is something your patient is interested in.

Number. The number of subjects is crucial to whether accurate statistics can be generated from the data. Too few patients in a research study may not be enough to show that a difference actually exists between the intervention and comparison groups (known as the "power" of a study). Many studies are published with less than 100 subjects, which is usually inadequate to provide reliable statistics. A good rule of thumb is 400 subjects.[4] Fifty-one patients completed the wart study, which is a pretty small number to generate good statistics.

Statistics. The statistics you are interested in are few in number and easy to understand. Since statistics are frequently misused in journal articles, it is worth a few minutes to learn which to believe and which to ignore.

Relative risk reduction. It is not unusual to find a summary statement in a journal article similar to this one from an article titled "Long-Term Effects of Mammography Screening: Updated Overview of the Swedish Randomised Trials":[5]

"There were 511 breast cancer deaths in 1,864,770 women-years in the invited groups and 584 breast cancer deaths in 1,688,440 women-years in the control groups, a significant 21 percent reduction in breast cancer mortality."

This 21-percent statistic is the relative risk reduction (RRR), which is the percent reduction in the measured outcome between the experimental and control groups. (See "Some important statistics," below, for more information on calculating the RRR and other statistics.) The RRR is not a good way to compare outcomes. It amplifies small differences and makes insignificant findings appear significant, and it doesn't reflect the baseline risk of the outcome event. Nevertheless, the RRR is very popular and will be reported in nearly every journal article, perhaps because it makes weak results look good. Think of the RRR as the "reputation reviving ratio"

> The relative risk reduction is very popular and will be reported in nearly every journal article, perhaps because it makes weak results look good.

SOME IMPORTANT STATISTICS

Absolute risk reduction (ARR): The difference between the control group's event rate (CER) and the experimental group's event rate (EER).

Control event rate (CER): The proportion of patients responding to placebo or other control treatment. For example, if 25 patients are in a control group and the event being studied is observed in 15 of those patients, the control event rate would be 15/25 = 0.60.

Experimental event rate (EER): The proportion of patients responding to the experimental treatment or intervention. For example, if 26 patients are in an experimental group and the event being studied is observed in 22 of those patients, the experimental event rate would be 22/26 = 0.85.

Number needed to treat (NNT): The number of patients that must be treated to prevent one adverse outcome or for one patient to benefit. The NNT is the inverse of the ARR; NNT = 1/ARR.

Relative risk reduction (RRR): The percent reduction in events in the treated group compared to the control group event rate.

	When the experimental treatment reduces the risk of a bad event:	Example: Beta-blockers to prevent deaths in high-risk patients with recent myocardial infarction:	When the experimental treatment increases the probability of a good event:	Example: Duct tape to eliminate common warts.
Relative risk reduction (RRR)	CER-EER/CER	(.66 -. 50)/.66 = .24 or 24 percent	EER-CER/CER	(.85-.60)/.60 = .42 or 42 percent
Absolute risk reduction (ARR):	CER-EER	(.66 - .50) = .16 or 16 percent	EER-CER	.85-.60 = .25 or 25 percent
Number needed to treat (NNT)	1/ARR	1/.16 = 6	1/ARR	1/.25 = 4

or the "reporter's reason for 'riting." Ignore the RRR. It will mislead you. In our wart treatment example, the RRR would be (85 percent - 60 percent)/60 percent x 100 = 42 percent. The RRR could thus be interpreted as showing that duct tape is 42 percent more effective than cryotherapy in treating warts.

Absolute risk reduction. A better statistic is the absolute risk reduction (ARR), which is the difference in the outcome event rate between the control group and the experimental treated group. Thus, in our wart treatment example, the ARR is the outcome event rate (complete resolution of warts) for duct tape (85 percent) minus the outcome event rate for cryotherapy (60 percent) = 25 percent. Unlike the RRR, the ARR does not amplify small differences but shows the true difference between the experimental and control interventions. Using the ARR, it would be accurate to say that duct tape is 25-percent more effective than cryotherapy in treating warts.

Number needed to treat. The single most clinically useful statistic is the number needed to treat (NNT). The NNT is the number of patients who must be treated to prevent one adverse outcome. To think about it another way, the NNT is the number of patients who must be treated for one patient to benefit. (The rest who were treated obtained no benefit, although they still suffered the risks and costs of treatment.) In our wart therapy article, the NNT would tell us how many patients must be treated with the experimental treatment for one to benefit more than if he or she had been treated with the standard treatment.

Now this is a statistic that physicians and their patients can really appreciate! Furthermore, the NNT is easy to calculate, as it is simply the inverse of the ARR. For our wart treatment study, the NNT is

EXAMPLES OF NNTS

The number needed to treat (NNT) is one of the most useful statistics for physicians and patients. It calculates the number of patients that must be treated to prevent one adverse event or for one patient to benefit. Note that NNTs for preventive interventions will usually be higher than NNTs for treatment interventions. The lower the NNT, the better.

The following examples of NNTs are borrowed from an excellent list available through the Bandolier Web site at http://www.jr2.ox.ac.uk/bandolier/band50/b50-8.html.

Therapy	NNT
Triple antibiotic therapy to eradicate *H. pylori*	1.1
Isosorbide dinitrate for prevention of exercise-induced angina	5
Short course of antibiotics for otitis media in children	7
Statins for secondary prevention of adverse cardiovascular outcomes	11
Statins for primary prevention of adverse cardiovascular outcomes	35
Finasteride to prevent one operation for benign prostatic hyperplasia	39
Misoprostol to prevent any gastrointestinal complication in nonsteroidal anti-inflammatory drug users	166

1/25 percent =1/0.25 = 4, meaning that 4 patients need to be treated with duct tape for one to benefit more than if treated by cryotherapy.

Wrapped up in this simple little statistic are some very important concepts. The NNT provides you with the likelihood that the test or treatment will benefit any individual patient, an impression of the baseline risk of the adverse event, and a sense of the cost to society. Thus, it gives perspective and hints at the "reasonableness" of a treatment. The value of this statistic has become appreciated in the last five years, and more journal articles are reporting it.

What is a reasonable NNT? In a perfect world, a treatment would have an NNT of 1, meaning that every patient would benefit from the treatment. Real life is not so kind (see "Examples of NNTs," above). Clearly, an NNT of 1 is great and an NNT of 1,000

Now this is a statistic that physicians and their patients can really appreciate!

SPEEDBAR®

◄ ➤

The absolute risk reduction shows the true difference between the experimental and control interventions.

◄ ➤

The most clinically useful statistic is the number needed to treat (NNT), which tells you the number of patients who must be treated to prevent one adverse outcome or for one patient to benefit.

◄ ➤

The NNT is easy to calculate, as it is simply the inverse of the absolute risk reduction.

◄ ➤

In a perfect world, a treatment would have an NNT of 1, but real life is not so kind.

is terrible. Although it is hard to come up with firm guidelines, for primary therapies I am satisfied with an NNT of 10 or less and very pleased with an NNT less than 5. Our duct tape NNT of 4 is good, particularly since the treatment is cheap, easy and painless.

Note that NNTs for preventive interventions (e.g., the use of aspirin to prevent cardiac problems) will usually be higher than NNTs for treatment interventions (e.g., the use of duct tape to cure warts). Prevention groups contain both higher-risk and lower-risk individuals, so they produce bigger denominators, whereas treatment groups only contain diseased patients. Thus, an NNT for prevention of less than 20 might be particularly good.

When discussing a particular therapy, I explain the NNT to my patient. Since this statistical concept is easy to understand, it can help the patient be a more informed partner in therapeutic decisions.

You will soon start to see a similar statistic, the number needed to screen (NNS), which is the number of patients needed to screen for a particular disease for a given duration for one patient to benefit.[6] Although few NNSs have been calculated, they are likely to involve higher numbers, since the screening population consists of patients with and without the disease. For example, in the article on mammography screening mentioned above, the NNS was 961 for 16 years. In other words, you would need to screen 961 women for 16 years to prevent one breast cancer death.

The good news and the bad

Using PP-ICONS to assess the wart study, the problem, the patient/population, the intervention, the comparison and the outcome are all relevant to your patient. The number of subjects is on the small side, making you a little wary, but the intervention is cheap and low-risk. The statistics, particularly the NNT, are reasonable. On balance, this looks like a fair approach, so you call the patient's mother and discuss it with her.

The PP-ICONS approach is an easy

> **You can apply PP-ICONS even when reading news stories describing medical breakthroughs.**

way to screen an article for validity and relevance, and the abstract often contains all of the information you need. Even the statistics can be done quickly in your head. You can apply PP-ICONS when searching for a particular article, when you come across an article in your reading, when data are presented at lectures, when a pharmaceutical representative hands you an article to support his or her pitch, and even when reading news stories describing medical breakthroughs.

Don't be discouraged if you find that high-quality articles are rare, even in the most prestigious journals. This seems to be changing for the better, although many careers are still being built on questionable research. Nevertheless, screening articles will help you find the truth that is out there and will help you practice the best medicine. And as we become more discerning end-users of research, we might just stimulate improvements in clinical research in the process. **FPM**

Send comments to fpmedit@aafp.org.

1. Miser WF. Critical appraisal of the literature. *J Amer Board Fam Pract.* 1999;12(4):315-333.

2. Guyatt GH, et al. Users' guides to the medical literature. How to use an article about therapy or prevention. Are the results of the study valid? *JAMA.* 1993;270(21):2598-2601.

3. Lepor H, et al. The efficacy of terazosin, finasteride or both in benign prostatic hyperplasia. Veterans Affairs Cooperative Studies Benign Prostatic Hyperplasia Study Group. *N Engl J Med.* 1996;335(8):533-539.

4. Krejcie RV, Morgan DW. Determining sample size for research activities. *Educational and Psychological Measurement.* 1970;30:607-610.

5. Nystrom L, et al, Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet.* 2002;359(9310): 909-919.

6. Rembold CM. Number needed to screen: development of a statistic for disease screening. *BMJ.* 1998;317:307-312.