

Proof and Consequences

Before you undergo treatment, look closely at the evidence.

BY TOM VALEO

Suppose you've had a minor stroke. A small blood clot has blocked blood flow to a part of your brain. Doctors at the emergency room managed to dissolve the blockage, and now you're fine. But you'll need ongoing treatment to prevent a recurrence. You expect your doctor to rely on the best possible evidence, right?

That's why the American Academy of Neurology (AAN) analyzes all available evidence for a particular treatment and then classifies that data into one of four categories, based on the quality of the studies that produced the evidence.

Doctors aren't the only people who should know how to judge the quality of proof. Not all studies are created equal, and knowing how to evaluate evidence will make you a smarter—and maybe healthier—patient or caregiver.

Not all studies are created equal.

THE BEST EVIDENCE

Let's get back to our example above. Several follow-up treatments exist for people who have had a minor stroke. Low doses of aspirin inhibit blood clotting, making another stroke less likely, but so does a drug called warfarin. Which is better?

The best way to find out would be to conduct a study involving people who have had a minor stroke. Half would receive warfarin; the other half aspirin. At the end of the study, if the group that took warfarin had fewer strokes, the drug would be deemed more effective than aspirin. If there was no difference in the number of strokes in each group, the conclusion would be that the two treatments are equal.

Ideally the pills would look identical so neither the patients nor their doctors would know who was getting which medication. That would prevent bias from creeping into the results. For example, let's say your doctor considered



warfarin superior to aspirin. If he knew he was giving you warfarin, he might appear more enthusiastic than he

would if he was giving you aspirin. His enthusiasm could trigger a placebo effect in you, which is an actual improvement produced not by the drug but by your hope—bolstered by your doctor's attitude—that the pill will make you better. The placebo effect would make the drug appear more effective than it really is.

Research on both warfarin and aspirin for stroke prevention has produced Class I evidence, with aspirin and warfarin producing approximately equal results in patients with strokes caused by blood-vessel narrowing or stenosis. Since warfarin requires expensive monitoring and carries slightly greater risks, most people take aspirin. However, in patients with atrial fibrillation—a heart flutter that causes blood clots that can travel to the brain—warfarin has been found to be superior to aspirin. This highlights another important point about interpreting study results: You have to know if the people enrolled in the study are like you to be able to know if the results apply to you.

ANECDOTAL EVIDENCE

What about less persuasive evidence? Should it be ignored? Drugs that have never been tested can still be very effective.

“Before aspirin came into widespread use no one conducted a rigorous placebo-controlled study to show that it reduced fever,” says Robert Gross, M.D., Ph.D., of the University of Rochester and author of a recent editorial in the journal *Neurology* about levels of evidence. “People just chewed on bark and their fever went down, so Bayer got the idea to isolate the effective compound in the bark and sell it as aspirin.”

Also, drugs that have been shown effective for one condition may show benefits for another, even though they've never been tested for it.

Gabapentin, for example, a drug for controlling seizures, was tested on thousands of subjects and found to be effective. But some menopausal women who took the drug claimed it also reduced the number of hot flashes they experienced. That's called “anecdotal evidence,” based on casual observations by patients or physicians. Anecdotal evidence qualifies as Class IV—the lowest form. Still, such anecdotes, even in the absence of rigorous testing, may point to a useful secondary effect of the drug.

“Absence of evidence of effectiveness is not evidence of the absence of effectiveness,” says Gary Gronseth, M.D., a neurologist at the University of Kansas who wrote a manual on evidence assessment for the AAN. In other words, the fact that no study exists proving that the treatment works is not proof that the treatment is useless. “This is a logical fallacy that insurance companies sometimes use to deny payment.”

When tested scientifically, gabapentin turned out to be effective in reducing hot flashes in women. Those anecdotal reports prompted a rigorous scientific study that produced the highest level of evidence.

SHADES OF GREY

The four categories help physicians judge the quality of the evidence they are examining. While Class I evidence is the best, Class II is almost as good. If too many subjects in the study drop out before the end of the study, for example, the evidence may be downgraded to Class II.

Class III evidence contains even more potential for bias. Perhaps the group receiving the placebo contained patients who were younger or older than the subjects receiving the therapy. Or maybe the therapy produced a visible side effect that alerted examining physicians to which subjects were receiving the actual drug. Such bias doesn't necessarily discredit the results of the experiment, but it renders the results less reliable than Class I and II evidence.

Based on the quality of available evidence, the AAN issues recommendations for various treatments. But these “practice parameters” are not rigid guidelines that physicians must follow, according to Jacqueline French, M.D., a neurologist at New York University's Comprehensive Epilepsy Center and the Chair of the AAN's Quality Standards Committee,

which creates the practice parameters. Rather, they serve primarily as aids to the judgment a physician must exercise.

“If a practice parameter assesses the effectiveness of a drug with patients up to the age of 65, and you're standing there with a patient who is 67, it's biologically plausible that the parameter applies to this patient,” Dr. French explains. “But if the parameter states that with age the effects of the drug diminish, and you're with an 85-year-old patient, then biological plausibility might say I shouldn't use this evidence.”

Patients and caregivers also benefit from

knowing about classes of evidence, according to Dr. French, because such information helps them understand why their doctors favor one treatment over another.

“If patients understand the difference between strong and weak evidence,” Dr. French says, “they will be in a better position to assess the risks versus the likelihood of benefit for proposed treatments.” NN

For a version of the AAN practice guidelines that are written specifically for patients and caregivers, please go to aan.com and click on “Practice Guidelines.”

Classes of Evidence

Class I evidence is produced by a type of study known as the randomized controlled trial (RCT). People are selected for this kind of study based on rigorous criteria and then assigned randomly (in a sense, by coin toss) to either a specific medical treatment or placebo (sometimes called a “sugar pill”). A Class I study requires the following:

- ▶ The team who rates how well patients do with the treatment must be “blinded,” which means they do not know whether a patient received the treatment or a placebo. As a result, the team's assessment of the treatment response is unbiased.
- ▶ The researchers must decide their main question in advance of the study.
- ▶ There must be clear rules for who is included in the study and for who cannot participate.
- ▶ At least 80 percent of people who enter the study must complete it.
- ▶ If a treatment is being compared to an existing “standard” treatment, certain strict rules must be followed. For example, the dose of the drug used in the current study must match the dose that was used in the original studies that defined the standard treatment.

Class II evidence is produced by a study that—like a randomized controlled trial—compares two groups of people that are similar: One group receives a treatment and the other receives a placebo. However, a Class II study falls short of Class I in the following ways:

- ▶ It lacks one of the five Class I criteria listed above. For example, if 30 percent of patients dropped out of the study, this would raise concern that there was something different about these patients, which might introduce bias into the results. Therefore the study would be downgraded to Class II.
- ▶ The treatment decision was made by the person's doctor and not randomly. However, as in Class I, the team who rates how well patients did with the treatment must not know which treatment the patient received.

Class III evidence is produced by studies of all other types in which the team rating the outcome is unaware of what treatment the person received, or in which the outcome of the treatment can be easily measured in a standard way (like a number) that does not require much judgment and so is not likely to be subject to bias.

Class IV evidence is produced by studies that do not meet the above rules. These could be reports of a series of patients that a doctor has observed, or a study asking a panel of experts what is the best treatment for seizures.