

# The Scientific Evaluation of New Claims

George C. Curtis  
University of Michigan Hospitals

*In the scientific world, the burden of proof for new claims is on the claimant. This is consistent with probabilities, because most new suggestions do turn out to be false. There is no single path from claim to proof. This article illustrates four typical stages of the process, using the example of a claim for effectiveness of a new treatment. They are labeled "Tentative or Suggestive Evidence," "Promising or Interesting," "Probably Effective," and "Established Effectiveness." Each is based on characteristic levels of evidence. A few clues for identifying naive enthusiasts or outright frauds who try to shortcut these stages include claims of being ignored or rejected by an "establishment" that is too self-interested or set in its ways to consider new approaches; claims of special skills that others lack; vagueness about how the treatment has been evaluated and tested; and expressions such as "amazing," "unique," "special," and "never dreamed of."*

The world is full of claims of new and exciting things: from a better mousetrap to an instant release for untapped inner powers—or a surefire cure for anxiety. Each one brings with it the problem of deciding whether it is, indeed, a great breakthrough, or just another dose of snake oil. There is no single way of judging, but there are ways of minimizing mistakes at the beginning, and of arriving at the truth in the end.

In the scientific world, a new claim is presumed false until proved otherwise. In other words, the burden of proof is on the claimant. This is consistent with the probabilities, because most new claims or ideas do in fact turn out to be wrong. In science, as in other fields, there are standards of evidence for each of the steps along the way from claim to proof. For the sake of this discussion, let us take the case of a claim of effectiveness of a new treatment.

---

**Author's Note:** Correspondence may be addressed to George C. Curtis, Department of Psychiatry, University of Michigan Medical Center, Ann Arbor, MI 48109.

*Editor's Note:* This article has been adapted with the permission of the author, at the request of the Editor. This work previously appeared under the same title in the *Phobia Society of America Newsletter*, 1987, 6(1), 4-5.

### **Tentative or Suggestive Evidence**

At the beginning, the evidence usually consists of anecdotal observations, single-case reports, or uncontrolled or open trials. An investigator can usually get this sort of evidence published somewhere. Publishing in an independent, scientific journal is the principal way an investigator submits his or her ideas for review and criticism by others. Through these means, other investigators are stimulated to try new ideas and to add their insights into the investigation. To publish, the investigator must describe one or more cases, the new treatment, and the therapeutic benefit, in a clear, objective way that does not convey a sense of exaggeration or uncritical enthusiasm. He or she must also show an awareness that only very tentative conclusions can be drawn from such evidence.

A responsible investigator is expected to report the number of cases studied and the study's success ratio, as well as other factors that may have influenced the outcome of the investigation. The argument for the treatment is strengthened if there are several cases that have responded the same way; if some, or all, were especially difficult cases for which the usually effective treatments had failed; or if the impression of improvement is supported by some type of objective measurement. At best, however, conclusions based on this type of evidence are considered to be tentative or preliminary, warranting further investigation, which may be done by the original investigator or his or her peers. As additional trials are made, fragments of evidence accumulate.

### **Promising or Interesting**

The next study stage is likely to be a prospective open trial, where cases are selected so as to avoid stacking the deck for or against treatment. Cases will be assessed/diagnosed by some method previously proved to be reliable and accurate, and change will be measured by several different methods. Normally some of these measures will be self-ratings or questionnaires filled out by clients, and some will be ratings done by a specially trained clinical investigator. Usually, the number of cases will not be large, and there will be no placebo group (a harmless inactive treatment or drug) or other control group.

If the results of several trials of this sort are fairly positive and consistent, they may be considered promising or interesting. There may now be justification for the expense and effort of a full-scale "controlled," and, if possible, "double-blind" investigation.

### **Probably Effective**

The use of a control condition is crucial in treatment studies of mental disorders because of the difficulty in assessing the cause of improvement. The controlled study will include the features of the open trial plus a control group or control condition. If it is a drug that is being evaluated, the control may be a harmless sugar pill, a placebo, which looks exactly like the real medicine. With psychosocial treatments, the control problem is more difficult. The investigator tries to think of some treatment that approximates the one to be tested but that is known to have little or no effect.

The study should be designed so that about equally difficult cases get the real treatment and the control treatment. The most effective way to do this is to assign cases randomly to treatment or control conditions. Other ways are the crossover design (where people alternate between a control and a trial treatment) or the matched groups design (where each person in the test group is paired with someone in the control group who is similar in age, sex, severity, duration of illness, or other factors thought or known to influence the outcome of treatment or course of illness).

Blindness is a strategy to minimize the very substantial effects that expectations or beliefs can exert on the outcome as perceived by the client, clinician, or observer. The person who assesses the degree of change should be different from the clinician who provides the treatments. In tests of drug treatments, by means of a placebo, both the therapist and the client can be prevented from knowing whether the real or the control treatment is being provided. In a study of psychosocial interventions, it is usually possible to prevent only the person who assesses the change from knowing the client's assigned experimental group (real or placebo therapy).

It is relatively easy to monitor the amount and type of drugs taken through urine and blood analysis. Tests can determine the presence of the prescribed drugs as well as others, such as tranquilizers or street drugs, which may have been taken. It is more difficult to verify that a psychosocial treatment has been delivered as intended, but some helpful methods of verifying this have been recently developed (e.g., audio or videotaping therapy sessions, and having them blindly evaluated by a skilled clinician).

A final requirement is that the number of patients be large enough, and appropriate methods of statistical analysis be chosen, to permit valid probability statements about the meaning of the findings.

If a treatment method is reasonably promising and interesting, a report of a controlled treatment trial with most of the above features will find accep-

tance for publication in one of the leading journals, regardless of what the results show.

If one well-designed trial, as described above, shows a treatment to be substantially superior to another, or if several less rigorous investigations agree in finding it effective, the scientific community will begin to regard the treatment as probably effective.

### **Established Effectiveness**

If several well-designed, controlled studies agree in showing one form of treatment to be substantially superior to another, the scientific community will usually consider the effectiveness of the treatment to be *established*. If there are other treatments that are also effective for the condition, additional research will be necessary to determine the appropriate conditions under which each should be used. If different, well-designed treatment trials give different results, additional research, including a careful examination of possible reasons for the disagreement, will be necessary before a final conclusion about effectiveness can be reached.

### **Application in Practice**

Unfortunately it is not possible to conduct a clinical practice only with methods established as effective. At the moment, the number of such treatments is limited, and not one is effective for every patient or client. Responsible practitioners are expected to know where various treatment plans stand in the research scheme and any limitations on their effectiveness. They also have an obligation to disclose this information fully to their patients or clients.

## **CONCLUSION**

From time to time, someone claims astounding results that no one else is good enough or smart enough to duplicate. Most of these claimants are either naive enthusiasts or outright frauds. Some warning signs:

1. The complaint that they are being ignored or excluded by the "establishment" or its equivalent;
2. Portrayal of the "establishment" (or equivalent) as too set in its ways to accept anything new, or as protecting its self-interest against new ideas;
3. Claims that cures are complete or surefire;

4. Words and phrases such as “amazing,” “never-dreamed-of,” “unique,” or “special”;
5. Belittlement of other approaches;
6. Vagueness about methods.

However, in science nothing is ever taken as an established fact until there exists a published set of instructions (in effect, a “cookbook”) for replicating results and until the scientific community is convinced that any competent investigator can get the same results by following those written instructions. As Winston Churchill said of democracy, “This appears to be the worst possible system, except for all the others.”