in: The Scientist 10 (March 4): 10 (1996)

Wishful Thinking and the Fallacy of Single-Subject Experimentation

by Michal Jasienski

Few elements in the public view of science and medicine rival the romantic image of a scientist drinking a self-concocted potion and risking his or her life to benefit humanity. Recently, K.S. Brown (The Scientist, Dec. 11, 1995, page 1) outlined the history of self-experimentation in biomedical science. In but two paragraphs, however, was there any suggestion that, apart from being unconventional or potentially dangerous to researchers, self-experimentation as a methodological option might have weaknesses. Non-objective data analysis is the least of problems, however.

Experimenting on oneself suffers from an extreme form of the observer's effect, when the observer becomes the observed. Self-observation is not only, as philosophers of science put it, "theory-laden," but also heavily introduced by the subject's mood and yesterday's diet. One might assert that self-experimentation represents only a minor current in research and as such poses little threat to the overall quality of biomedical science. However, self-experimentation is a special case of much more widespread research involving single subjects.

Single-subject studies, such as a widely publicized attempt to use baboon bone marrow for AIDS treatment involving a single patient, named Jeff Getty (see story on page 3), represent a seriously flawed scientific methodology. Consequently, the results of all single-subject studies are bound to be erroneous. Do the errors matter? After all, every result obtained during the process of a scientific inquiry has an error attached to it, because the scientific procedure that generated this result may be biased or inaccurate. The imprecision and bias of our results may be evaluated only be statistical analysis of repeated observations. Without statistics, there is no way of knowing if the errors matter, and there can be no statistical analysis on samples of one. Can we then base our biomedical policies on estimates with unascertainable errors?

The essence of empirical research is a constant tension between the Scylla of rejecting a correct explanation of a phenomenon and the Charybdis of accepting a false explanation. Science is about navigating between them and finding an explanation through separating unique from repeatable properties of subjects under study. Both types of properties are hopelessly confounded within any single individual, and no amount of technological progress will undo such confounding. The only way is through "averaging out" the superficial differences among individuals, to see what meaningful similarities remain (representing, for example, the effects of a medical treatment).

Averaging out with single-subject experiments does not work; at best, with a proper experimental design one can demonstrate on one subject the presence or absence of some treatment effects. The critical proviso is "with a proper experimental design": lack of it is transparent in the Getty experiment. This ambitious trial was crippled by the absence of control subjects and by the placebo effect of unknown magnitude.

Contrary to the claims of success, this particular experiment must be considered a failure, becoming victim of its own novelty and the politics surrounding the issue. The shortsightness of the Food and Drug Administration which permitted this single-subject trial, rather than promoting larger, controlled experiments, is indeed baffling. The shortsightedness of some AIDS activists opposing fully controlled studies--those in which some subjects do not receive treatment--(L.K. Altman, New York Times, Feb. 6, 1996, page A1), on the grounds that they are unethical, is indeed tragic.

Jeff Getty's uniqueness, resulting in part from the pre- and post-transplant treatment he received from his doctors (including radiation and a battery of antiviral, immunosuppressive, and other drugs) and the media spotlight, makes the failure of the transplant (L.K. Altman, New York Times, Feb. 9, 1996, page A14) uninterpretable. Even if the treatment did turn out to be harmless and efficacious, how far from optimal would it be for different people? What if Getty also had a rare metabolic error that, unbeknownst to him, influenced his response to the new treatment? Would the whole line of research be modified, leading us astray? Can therefore a single person represent all potential target patients? Of course not!

Experimentation on single subjects (and self-experimentation in particular) violates basic assumptions of scientific inquiry and must be abandoned. It precludes statistical analysis and lets the dark forces of wishful thinking and preconceived notions creep into objective experimental methodology. With clinical medicine becoming more scientific (through awareness of inter-individual variability and need for statistical methodology, and so forth), it should have no place for single-subject experimentation. The progress in medicine requires well-designed trials that yield recommendations of optimal procedures, optimal drug dosages and optimal ways of administering them, leading to the maximization of the treatment efficacy. Although less romantic, Dr. Jekyll's self-imposed transformation or Jeff Getty's brave determination could have been much more informative had they involved appropriate experimental design.

Michal Jasienski is a postdoctoral fellow in the Department of Organismic and Evolutionary Biology at Harvard University. His E-mail address is mjasienski@oeb.harvard.edu.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

in: The Scientist 10 (October 14): 12 (1996)

N = 1, Fortuitous Effects, and AIDS Research

by Michal Jasienski

A recent article on xenotransplantation (R. Finn, The Scientist, Aug. 19, 1996, page 1) and two published responses to my commentary (M. Jasienski, The Scientist, March 4, 1996, page 10), which was critical of a single-subject baboon bone-marrow transplant trial, prompt me to explicate some methodological issues further.

What insights can a single-subject study provide? While, as pointed out by J.E. Janosky (The Scientist, May 27, 1996, page 12), there could be some room for N = 1 studies (see e. g. H. Motulsky, Intuitive Biostatistics, New York, Oxford

University Press, 1995), a letter by J.G. Llaurado (The Scientist, July 8, 1996, page 12) unintentionally strengthens my skepticism in this matter. Llaurado writes that the currents that Alessandro Volta used on himself to study electricity are "today believed sufficient to cause lethal arrhythmias on a less corpulent man than Volta." Is it then not quite fortunate for the development of science that Volta was overweight (that is, not representative of the whole population)? How many Voltas were there who were scrawny and, consequently, electrocuted? I suspect that many similar single-subject experiments gain from selective memory.

There is a deeper point, though. Any empirical science moves from general demonstrations to refinements. Single-subject demonstrations of the applicability of laws of nature rely on the subject's being roughly representative of its kind. You've seen one, you've seen them all apples falling to the ground. This is because the property we focus on (mass, rather than, say, color) has indeed the greatest bearing on the phenomenon we are trying to explain (apple's falling down). How often can we be sure, however, which properties as expressed in a John Doe are robustly representative of the responses of all mankind? Granted, it has certainly worked in the past, as Llaurado correctly points out, but the lucky single subjects were representative of the human species only because the questions were about some of the most fundamental phenomena. Our knowledge becomes more and more subtle, however. Rough ideas, approximate explanations, and nonmechanistic predictions are no longer sufficient, especially in medicine.

While I am full of admiration for Jeff Getty's perseverance and fortitude in his involvement in the baboon bone-marrow transplant trial, I would wish the results of his efforts were not rendered uninterpretable by the weakness of the experimental design. I hope that the Food and Drug Administration and the medical establishment are equally determined to follow Getty's leadership with the state-of-the-art statistical methodology. A high-tech approach is a necessary, but not a sufficient, component of a successful study. Statistically sound design will always be a condition sine qua non. Alas, N = 1 is not it.

Michal Jasienski
Department of Organismic and Evolutionary Biology
Harvard University
Cambridge, Mass. 02138
E-mail: mjasiens@oeb.harvard.edu