



Corruption of the Scientific Method

Nancy Flournoy

Professor, University of Missouri, Columbia;
Fellow, World Academy of Art and Science

Abstract

The scientific method is accepted worldwide as a major group decision-making process. Countries and international organizations rely upon the scientific method in constructing laws, regulations and treaty agreements. We describe how the scientific method has been corrupted by thoughtlessly following a historical prescription that is reinforced by stakeholders with large financial interests. Finally, we conclude with some options for stemming the corruption.

1. Introduction

We begin by reviewing the scientific method, as it is widely espoused but narrowly defined in the Western world. This method is accepted worldwide as a major group decision-making process. Countries and international organizations rely upon the scientific method in constructing laws, regulations and treaty agreements. We describe how the scientific method has been corrupted by thoughtlessly following a historical prescription that is reinforced by stakeholders with large financial interests. We conclude with some options for stemming the corruption.

2. The P-Value

From a scientific experiment comes the p-value, which is a number summarizing the research effort. Indeed the research effort may produce many p-values, but considering only one simplifies our expose and only one p-value is needed to make our point. In the simplest experiment, the p-value is used to choose between two options:

- The Null Hypothesis and
- The Alternative

The p -value is a number between zero and one; it is an estimate of the chances of observing the data that were in fact observed if the null hypothesis were true. The validity of this estimate depends on the quality of the experiment's design, procedures, modeling assumptions and sample size. A small p -value indicates that it was very unlikely that the observed data were produced under operation of the null hypothesis. Thus a small p -value leads to selecting the alternative hypothesis.

Commonly, the alternative is selected if $p \leq 0.05$. Why 0.05? Before computers, Karl Pearson published tables for just a few decision-making criteria, including $p \leq 0.05$ which was

widely adopted.¹ In spite of now having computers that permit scientists to easily make decision-making criteria reflect context-specific risks, the “ $p \leq 0.05$ ” criterion has virtually become a universal law and it is hard to get experimental results published unless this outcome is obtained.

3. The Method as Intended

When there are only two possible choices after an experiment (i.e., accept the null or the alternative hypothesis), there are two possible ways in which to be wrong. See Table 1.

Table 1: Possible Choices after an Experiment

The Truth is	Decide to go with The Null Hypothesis	Decide to go with the Alternative
The Null Hypothesis	Correct	Error Type 1
The Alternative	Error Type 2	Correct

It is intended that null hypothesis represents the current situation, the status quo, and that it should be hard to reject the null hypothesis in favor of something different or new.

An analogy common in elementary statistics texts is judicial systems in which the defendant is assumed to be innocent until proven guilty. The burden of proof is placed on the prosecution under the belief that the two possible errors are not equal, specifically, that it is worse to convict someone who is innocent than to acquit someone who is guilty. Such decision-making systems reflect the dominant underlying value system.

The two possible errors are assumed to differ in significance in applications of the scientific method. The widely used decision-making strategy considers an error of Type 1 to be much worse than an error of Type 2. It is expected that the terms “null” and “alternative” will be matched with experimental outcomes in such a way that the worse mistake is to wrongly reject the null hypothesis and thus wrongly accept the alternative as truth.

Before the experiment, the scientist establishes a limit on the chances of making a Type 1 error (now almost always 5/100); then after the experiment, the scientist estimates the chances of a Type 1 error assuming the null hypothesis to be true and accepts the alternative only if this estimate is very low (almost always $p \leq 0.05$).

The Type 2 error is given little attention. When a Type 2 error occurs, the alternative is true but is not chosen. For example, a successful alternative therapy is disreputed. Given the rule is to select the alternative if $p \leq 0.05$, a larger sample size reduces the chances of making a Type 2 error. Recognizing the harm that results from making Type 2 errors, many statisticians invest considerable effort trying to convince their scientific colleagues that their planned sample sizes are too small to provide reasonable protection from such erroneous conclusions. But only the p -value matters to most scientific journal reviewers and editors, and so it is hard for scientists to value the impact of making a Type 2 error.

While the systematic acceptance of high Type 2 error rates should be a matter of great concern among those responsible for allocating public resources to the scientific community,

this is not the topic of this paper. It is related; but the focus of this treatise is on errors of intent and not errors of omission.

4. How to Cope with the Scientific Method

The scientific method is easily corrupted by making the hypothesis you want to win acceptance of the null hypothesis!

This assertion is easily demonstrated by two examples in which the null and alternative hypotheses are matched with two possible realities with dramatically different implications for societal health and well-being. These two examples reflect the dominant decision-making strategies of different regulating agencies in the United States.

First consider new drug development, which is regulated by the U.S. National Institutes of Health. The operating null hypothesis is that drugs do no good, or are harmful. The burden of proof is on the drug developer to provide data that strongly support the contention that drugs are beneficial or that they are at least as good as available alternatives.

The second example concerns chemical and physical alterations of natural foods and the environment. These include pesticides, additives to food and food containers, and additives to cosmetics. Also included are genetic, chemical and physical manipulations to plants, animals and the environment. In the development of regulations under the jurisdictions of the U.S. Food and Drug Administration, the Consumer Protection Agency and the Department of Agriculture and in the name of the free market enterprise, the operating null hypothesis is that the substances in question are safe.

Recently, for example, without general discussion, education or debate, the public in the United States has witnessed the widespread introduction of nanoparticles of titanium dioxide into toothpaste, creams and lotions and of high fructose corn syrup into drinks and canned goods. Titanium dioxide is regarded as an inert, non-toxic substance by many regulatory bodies such as the U.S. MSDS (Material Safety Data Sheets). Yet a quick Google search reveals that safety questions are not resolved. On the University of Rochester, New York, website alone one can find scientists advocating for and against the use of nanoparticle formulations of titanium dioxide in sunscreen.^{2,3,4}

What is wrong with changing the glucose from corn into fructose, and loading it into processed foods? Companies must have a reason for investing energy, and hence money, into changing a simple sugar into a more complex one. Our digestion of fructose bypasses the switch in the liver that signals us that we are full and should stop eating.^{5,6} So, while high fructose syrup may not be harmful per se, it encourages over-eating and hence obesity.

Are titanium dioxide and high fructose corn syrup public health problems? As the scientific method is currently applied, acceptance of this assertion requires that experimental data be generated that strongly supports it. A major reason that it is hard to establish that chemical and physical alternations to food and the environment are harmful with the “scientific rigor” that would command stronger regulation is because it is labeled the alternative hypothesis.

Who determines what is and what is not the null hypothesis has tremendous advantage in winning acceptance of their position.

This is not lost on stakeholders with large financial interests, as is apparent from their frequent cries that their products and additives have not been proven harmful. Focus here on who they are saying should be doing the proving.

When companies do not have to prove their products are safe, who will pay to prove otherwise? The American Cancer Society has a list of priorities for further research, but the burden is on them, their donors and the public.⁷ A commercial company's structural advantage in the decision-making process forces the initiative for and cost of proving the alternative away from them, often it seems like into the air.

Because there is no well-funded public, systematic, institutionalized process for studying the effects of chemical and physical alterations to our food and environment, individuals who believe the alternative hypothesis (e.g., endocrine-disrupting chemicals are harmful at low doses) are largely disempowered and silenced in their efforts to substantiate their claims.^{8, 9, 10}

“The public should be educated concerning the scientific method as it is commonly employed, and the implications this construct has for their well-being.”

5. What can be Done?

- The public should be educated concerning the scientific method as it is commonly employed, and the implications this construct has for their well-being.
- In designing scientific studies and in reporting scientific results, the question ‘what is the null hypothesis?’ should be critically debated. Which type of error is worse? Before debating important study results, and well before accepting them, the public should debate whether the labeling of possible realities as null and alternative hypotheses actually match society's values, recognizing the significance of these labels.
- Attention should be paid to the chances of Type 2 error in designing studies and in reporting results. If journal editors, government regulators and the media demanded an estimate of the chance of a Type 2 error, assuming the alternative hypothesis is true, along with every p-value reported, it would shed considerable light on scientific evidence for and against the two hypotheses.

6. Conclusion

We have shown how outcomes from the scientific method depend strongly on how the questions are formulated, in particular, what is and is not the null hypothesis. Because the errors associated with different decisions are not treated equally, it is relatively hard to accept the alternative and relatively easy to accept the null hypothesis. Thus, which is which matters and should be of great interest when discussing any research of public interest.

Author Contact Information

Email: flournoyn@missouri.edu

Notes

1. Karl Pearson, *Tables for Statisticians and Biometricians* (Oxford: Biometric Laboratory, University College, 1931)
2. Zahra Naghdi Gheshlaghi, Ghoiam Hossein Riazi, Shahin Ahmadian, Mahmoud Ghafari, Roya, Mahinpour, "Toxicity and interaction of titanium dioxide nanoparticles with microtubule protein," *Acta Biochimica et Biophysica Sinica* 40, no. 9 (2008): 777-82
3. XiaoBo Li, ShunQing Xu, ZhiRen Zhang, Hermann J Schluessner, "Apoptosis induced by titanium dioxide nanoparticles in cultured murine microglia N9 cells," *Chinese Science Bulletin* 54, no. 20 (2009): 3830-3836
4. Ivo Iavicoli et al., "Toxicological effects of titanium dioxide nanoparticles: a review of in vitro mammalian studies," *European review for medical and pharmacological sciences* 15, no. 5 (2011): 481-508
5. Steven R. Gundry, *Dr. Gundry's Diet Evolution: Turn Off the Genes That Are Killing You and Your Waistline* (New York: Three Rivers Press, 2008)
6. Luc Tappy and Kim-Anne Lê, "Metabolic effects of fructose and the worldwide increase in obesity," *Physiological reviews* 90, no. 1 (2010): 23-46
7. Elizabeth M. Ward, Paul A Schulte, Kurt Straif, Nancy B Hopf, Jane C Caldwell, "Research Recommendations for Selected IARC-Classified Agents" *Environmental Health Perspective* 118, no. 10 (2010): 1355-62 <http://dx.doi.org/10.1289/ehp.0901828>
8. Wade V. Welshons and Frederick Susan vom Saal, "Large effects from small exposures: II. The importance of positive controls in low-dose research on bisphenol A," *Environmental Research* 100, no. 1(2006): 50-76
9. Frederick S. vom Saal, "Bisphenol A eliminates brain and behavior sex dimorphisms in mice: how low can you go?" *Endocrinology* 147, no. 8 (2006): 3679-3680
10. Wade V. Welshons, Susan C. Nagel and Frederick S. vom Saal, "Large effects from small exposures: III. Mechanisms mediating responses to the low doses of the plastic monomer bisphenol A," *Endocrinology* 147, no. 6 (2006): S56-S69