Responses to Gelman:

I need to respond to the column by Andrew Gelman about ethics (Vol. 24, No. 4). Most of the column is about a paper published by the principal investigator, Carl Blackman, and me, as the statistician on the project.

There are basically two parts to his column. The first is a claim of us being unethical and the second is his assertion of a flawed statistical analysis. With regard to ethics, he said "The ethics violation, as I see it, by Blackman and his statistician colleague came not in their design, data collection, or even their flawed analysis, but when they had the opportunity to subject their data to an outside analysis."

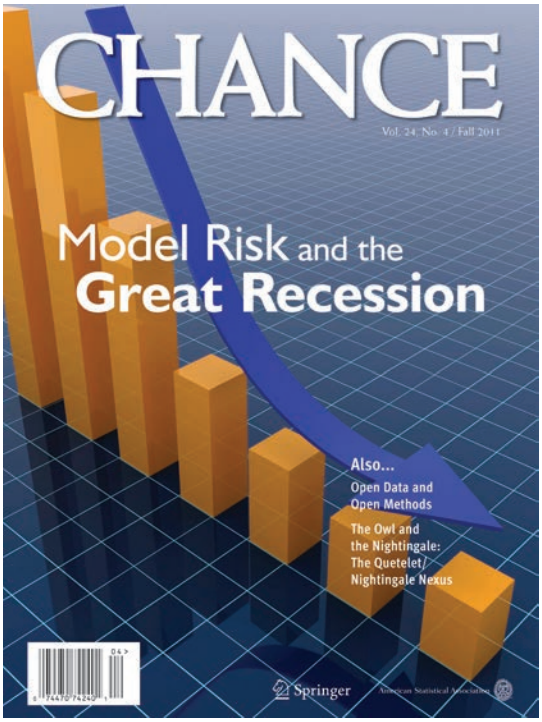
I worked at EPA [Environmental Protection Agency] for about 29 years and prided myself on making sure that all research I was involved in was of the highest quality that I could make it with my limited powers. Never did I do anything that could be construed as unethical—being ethical has always been important to me. I thought Carl Blackman was the most ethical researcher at EPA. He used to say "that's the data" a lot and would go where that led and not be led by personal biases. Carl had no problem with outside scrutiny of his work.

Gelman is saying we were unethical because the data in that paper was not given to him. I had no final say in any request about sharing data. I worked on hundreds of projects and was never once asked to make a decision about sharing data. A few times, including on this paper, I was asked for input on sharing the data, but the final decision was made by the lead researcher and/or management. Since I did not make the decision about sharing data, it is wrong to accuse me of being unethical. Carl will give details of why the request was denied in his response to Gelman's column.

On Page 53 of his column, Gelman admits to less than full disclosure about who he was when making the request for the data. His less than full disclosure about himself and his accusation without all the facts about my part in the decision is far worse ethically than anything he is accusing me of doing.

In addition, Gelman says the analysis was flawed and, as he pointed out several times, his "proof" seems to be that he had a PhD (although not at the time) and I only had a master's degree. That is correct, but I successfully finished all the coursework for a PhD in biostatistics at The University of North Carolina and, at the time the paper was published, I had almost 20 years of experience in medical and biological research. He continually asserts that he doesn't want to be a credentials snob, but that is exactly what he is.

On Page 52 of the column, Gelman says (to paraphrase) his Figure 1a summarized our results based on the statistical significance level. True, our Figure 1 had significance levels, but it also showed the estimated treatment effect. And a detailed data table was also provided with the means of the sham and exposed groups, their corresponding SE, and the *p*-values for each exposure condition. His implication is that we only



used *p*-values to make conclusions, which is clearly not true. Then he proceeds on a long rebuke of our using *p*-values to put the data into three groups, but ignores our statement (Page 221, third paragraph) that we did that for hypothesis-generating purposes and not as proof as any underlying patterns.

Last, Gelman discusses why we were wrong in our analysis of the design points A1, A2, B1, and B2. Carl will address this in his reply.

Gelman is correct that ethics is important. We should all be ethical in our research, and so too should we be ethical in our complaints about ethics.

Dennis House

I appreciate being given the opportunity to comment on an editorial column by Andrew Gelman, titled "Ethics and Statistics," that appeared in 2011, Vol. 24, No. 4, pages 51–53, in which my colleagues and I and our 1988 publication were prominently featured.

Let me state from the outset that I fully support the principles of data sharing and open and transparent exchange of scientific information, as well as the scientific publication and peer review process which supports open criticism, independent reproducibility, and scrutiny of scientific results. Furthermore, I believe that I have been a faithful adherent to those principles throughout my 40+-year career as a researcher at the [U.S.] Environmental Protection Agency, with many of my most useful research products and contributions resulting from open collaborations with a great many scientists. Dr. Gelman's column purported to present some important issues in scientific ethics from a statistician's viewpoint, elaborating on an experience from his graduate school years to illustrate his point. He begins his article with the statement: "An ethics problem arises when you are considering an action that (a) benefits you or some cause you support, (b) hurts or reduces benefits to others, and (c) violates some rule." Dr. Gelman then proceeds anecdotally to relate a distantly recalled experience, during the course of which he dismisses published, peer-reviewed scientific results of over 20 years ago, and impugns the scientific competency, integrity, and motives of me and my experienced EPA statistician coworker. I would counter that the unprofessional and self-serving manner in which this article was written, misrepresenting the original personal interaction and ignorant of (and uninterested in) the biophysical complexities and underpinnings of this and careful follow-up studies (that have been independently reproduced and considered influential), should be considered beneath the standards of this magazine and, ironically, would seem to violate the very ethical principles that Dr. Gelman espouses.

Dr. Gelman levels the charge, 20 years after the fact, that I violated the principle of openness in scientific research by denying his request to send him copies of my logbooks and that I designed experiments and data analyses that led to a "waste of effort," presumably because I and my coworker misapplied statistical principles in the analysis of the experimental findings. Both assertions are based on misleading and incomplete information, and in my view, are groundless.

I indeed recall receiving Gelman's letter in 1988 or 1989 requesting data from our publication (Blackman et al. 1988). The letter addressed a statistical issue, so I discussed the request with my co-author, an experienced and practiced biostatistician. The data in question had been collected in 1981–1983, nearly six years previously, and by 1984, we had tested and confirmed the fundamental hypothesis suggested from the original p-value plot, viz., that the local static magnetic field was involved in determining which frequencies could be biologically effective (Blackman et al., 1985). Thus, in 1988, I concluded that it was unlikely that further statistical examination of the data that was used to generate the original p-value plots would be profitable in advancing the understanding of that science. When Gelman called me to repeat his request, I inquired about his reason for wanting to use these data and he indicated that he wished to test some statistical hypotheses. When I offered to describe some of the sources of variance we had discovered in the 10 years during which we had refined the test, he appeared uninterested. As we talked, I inferred that his research status at Harvard was likely to be a graduate student and I recall having the strong impression that he was searching for data that he could work with in his spare time. It was disconcerting to me that he was not forthcoming about his status or motives (Was it in fact ethical to "be careful not to identify myself as a PhD student"?), and that he expressed no interest in learning of the larger context or implicit constraints in the experimental design and data analysis that were not obvious in the paper, but elaborated in earlier studies. I made the decision at that point that the time and effort I would have had to invest to locate and photocopy data scattered across many lab notebooks from six years earlier was excessive, since I believed that results of a statistical reanalysis of that data, whatever the outcome, would not advance the understanding of the science underpinning our results. I remember thinking that if he really were a graduate student, I should speak to his research adviser to determine the scientific and educational value of his working with these data. Since Gelman chose to not acknowledge that he was a graduate student at the time, I made a judgment that his efforts would not be mutually beneficial, and likely did not have the support or endorsement of his mentor. In fact, I was no longer performing studies in this research area because the EPA research program was discontinued in early 1986, and I had archived research materials from the 1976–1985 years. I was aggressively pursuing other experimental techniques and establishing collaborations that were completely outside the electromagnetic research area. Further, the size of the data set for this particular paper was very large and not easily assembled, as it was scattered across two projects. There was another hurdle, viz., I was required to get permission from EPA to share the raw data, something I believe I could have done had I considered this to be a worthwhile pursuit or had I been able to indicate that I was collaborating with a Harvard professor and his student.

Dr. Gelman admits that he chose to deliberately conceal his status at Harvard, for whatever reason. I would argue that it was this misrepresentation that prevented a more productive and mutually beneficial exchange, and unfortunately, he is correct that the raw data are no longer available. This electromagnetic research effort was a basic research project for which the data had to be retained in accordance with the EPA records retention schedule; they were destroyed after that date. In view of these facts, I assert that my refusal of the data request was not unethical and should not have been characterized as such.

Dr. Gelman also asserts that I designed experiments that were inadequate and wasteful of resources, and that I misused the concept of statistical significance. I do not intend to defend the classical model of statistical analysis against the newer model that he is championing; we had a more critical audience to convince in the early and mid-1980s: experimental biologists, electrical engineers, and biophysicists. Thus, our experimental design was based on our reported findings that there were temperature changes in the samples during the 20-minute exposure period as the samples rose to the set-point temperature, and that temperature change could influence the endpoint (Blackman et al. 1982). We believed it was essential to ensure that the sham and exposed samples had the same temperature history; as we were not able to get a comparable exposure chamber to serve as a coincident sham control, we always paired the two treatments within a one-hour period. Thus, in Gelman's designation, samples A1 and B1 were subjected to the same temperature change conditions, and samples A2 and B2 were subjected to the same temperature conditions, but the temperature conditions for A1 and B1 were not the same as those for A2 and B2. I do not question Gelman's conclusions that, on the average, the results of samples A2, B1, and B2 were indistinguishable, but, in my considered judgment, the audience I was addressing would not have accepted a revised experimental protocol as he suggests. I believe our procedure was the best we could devise at the time, and it was indeed effective in satisfying many concerns of the target audience.

Last, Dr. Gelman said of our work, "They used this completely inappropriate statistical distinction as the basis of several pages of speculation. This … led to a waste of effort." (Page 52). We clearly stated in the paper that "To analyze the data further, we used a transformation, the *p*-value, which combines the difference between the means of the exposed and unexposed groups, as well as the variance of each group. This transformation may be useful for hypothesis generation, but not as proof of any underlying pattern(s)." We had collected the response data as a function of frequency in the early 1980s, but did not immediately present it because the response profile was a tantalizing enigma.

The speculative use of *p*-values to highlight features of the data was far from "a waste of effort"; rather, it led to the report by Blackman et al. (1985) demonstrating that the earth's magnetic field could influence which frequencies of magnetic fields could be effective. That report quickly led others in the United States and abroad to confirm and extend the conditions for static magnetic-field controlled resonances. The research in this 1985 paper was undertaken because of the use of *p*-values, in the larger context of the biophysical considerations, to propose novel testable hypotheses. Its publication initiated a new direction in the electromagnetic research area by investigators in the United States, former USSR, and Sweden; the publication now has over 200 citations and has spawned a spirited debate regarding the underlying physical mechanism(s) of the action of electromagnetic fields on biological systems. In the 1988 paper under question, I merely provided the answer to the question, what aspect of the data caused us to think that the static magnetic field could be involved?

The *p*-value inferences from that early publication directly led to scientific discovery that has had substantial, beneficial consequences for expanding the understanding of how electromagnetic fields can influence biological systems and processes. I understand that Dr. Gelman is waging a battle to help scientists avoid the subtle trap of drawing conclusions from differences between significance levels. A battle is more persuasively argued in black and white terms, so he conveniently ignores our published qualification in the use of *p*-values and follow-up studies building on these inferences. I do not wish to engage in this larger battle, but simply to point out that the use of *p*-values in the context of a complex biophysical problem with real-world considerations and influences led to a very productive outcome. I agree with Dr. Gelman that it is important for researchers in all areas of life sciences to utilize best practices in the application of statistical principles and to be called to account when this is not the case. However, I would add that it behooves statisticians, at all levels of education, to consider the full context of a study and to engage researchers in a constructive, respectful manner before leveling charges of incompetence and unethical behavior in areas of research about which they have limited understanding. Perhaps there are even good reasons why the statistically sophisticated neuroscience research community, in some cases, still draws conclusions from the differences between significance levels.

Carl F. Blackman, PhD
Integrated Systems Toxicology Division (B-105-03)
National Health and Environmental Effects Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711 USA

Editor's Note: *This letter to the editor has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.*

Further Reading

Gelman, A. 2011. Open data and open methods, ethics and statistics. *CHANCE* 24(4):51–53.

Blackman, C.F., S.G. Benane, L.S. Kinney, W.T. Joines, and D.E. House. 1982. Effects of ELF fields on calcium-ion efflux from brain tissue in vitro. *Radiation Research* 92(3):510–520.

Blackman, C.F., S.G. Benane, J.R. Rabinowitz, D.E. House, and W.T. Joines. 1985. A role for the magnetic field in the radiation-induced efflux of calcium ions from brain tissue in vitro. *Bioelectromagnetics* 6(4):327–337.

Blackman, C.F., S.G. Benane, D.J. Elliott, A.R. Wood, D.E. House, and M.M. Pollock. 1988. Influence of electromagnetic fields on the efflux of calcium ions from brain tissue in vitro: Three models consistent with the frequency response up to 510 Hz. *Bioelectromagnetics* 9(3):215–227.

Gelman's Response:

I appreciate Dr. Blackman's explication of the background behind his research on the effects of electromagnetic fields. I am indeed no expert in this research area or in the experimental techniques used to study these effects, and I defer to Dr. Blackman's knowledge here.

I do not, however, defer to his belief that "results of a statistical reanalysis of that data, whatever the outcome, would not advance the understanding of the science underpinning our results." That is not his judgment to make.

I don't think we ever want to put a scientist in the position of deciding whether others' analyses of his data will materially affect the conclusions. Replications and reanalyses are the prerogative of the scientific community.

The reason for sharing data is that other researchers might well come to different conclusions. Central to the scientific method is a recognition that we make mistakes. Science is a collective enterprise. I am not claiming certainty that my analysis of the complete data would be correct. Any analysis I would do would be completely open, and others could feel free to come along later to criticize or improve it.

At a practical level, I accept Mr. House's statement that he did not make the decision to not share the data (beyond the averages and standard errors that were in the published papers), along with Dr. Blackman's argument that he was very busy and did not want to put in the effort required to get the data together. Looking forward, this is one reason that I believe we, the statistics profession (and the scientific community more generally), should make replication and data sharing a standard rather than an option. Looking forward, if future researchers accept the expectation of automatically satisfying data requests (even from students), they will be motivated to get the relevant permissions ahead of time and store their data in a portable form.

One reason we speak of ethical "dilemmas" is because of the difficulty of acting ethically or even determining the most ethical decision at any time. As Dr. Blackman discussed in his letter, he faced the tradeoff between, on one hand, sharing data and, on the other, refusing the request (thus freeing up his time to do additional important research). I stand by my general principle that it is a violation of scientific ethics to not share data when such sharing is possible, but I recognize that life is full of tradeoffs, and my retrospective assessment of the ethics in a single decision made over 20 years ago is not a larger judgment on the people involved (nor would I put myself in a position to make such a judgment).

One reason I wanted to use real examples with real people in my *CHANCE* column was to bring statistical ethics to life. I appreciate the letters of Dr. Blackman and Mr. House and I hope that readers will benefit from seeing both their perspectives and mine—just as researchers in general can benefit from seeing multiple analyses of publicly shared data.

Andrew Gelman
Column Editor of *Ethics and Statistics*