After the Fact Co-author

As a professional statistician, you are called by a colleague to examine and "bless" a biomedical experimental report. You are urged to do it quickly, because the report has already been submitted and accepted for publication in a prestigious journal in the author's field. One of the reviewers, however, had suggested that a quick review by a statistician might be in order. To your horror, the report appears to be utter statistical nonsense. The data were not sampled according to any plan, but rather were drawn from various similar experiments done for different purposes. There is no reason to assume the observations were random or independent within or among the data sets. There was no definition of how many data points had been originally available or how those used had be selected. The scatter plots within the paper were plainly skewed, but the computer statistical tests which had been run would have presumed a normal distribution. You explain gently that the statistical work is not an asset to the paper and could prove to be embarrassing to the author and the institution if published. You suggest that he eliminate the statistical portions and describe his work based on the qualitative reasoning which he obviously used. Initially very angry, he clams down and says, "I'll leave the contents alone, but I will add you as a coauthor. How's that?"

How do you reply? How is your reply conditioned by the relative power positions you may hold? If you are unable to reach an accommodation with the author, under what circumstances, if any, would you write to the journal editor to preclude publication? Under what circumstances, if any, would you decline to comment on the paper yourself, but refer the author to another colleague whose statistical expertise you consider to be so minimal that he or she might approve the paper as written?

Source: ASA Ethics Case Study #1

Scintillation Counter

Armstrong is a first year graduate student, working in a molecular biology laboratory. She has a great admiration for Hayes, who is just finishing his thesis work. He seems to have a golden touch in the laboratory. His experiments produce clean data, with scatter consistently less that or equal to theoretical predictions. Because his experiments seldom need to be repeated, Hayes has produced a thesis full of fascinating and demonstrably correct results. The laboratory has already followed up on several of these with success. One day Armstrong notices Hayes leaving the scintillation counter and can't help noticing he has 80 vials. This barely registers in her subconscious until later in the day he shows her his experimental results with 40 data points. When she asks about the missing points, he explains that it is standard practice to eliminate outliers from the analysis. He goes on to mention that the scintillation counter is a scientific instrument that frequently produces murky readings distorted by many different kinds of factors. The more Armstrong thinks about this, the more distraught she becomes. A week later she summons up her courage and tells her story to the professor in whose lab she and Hayes work. He seems uninterested an irritated. He hoped she had come to present him her experimental results, which she hasn't done for months.

What should Armstrong do next, and why?

Would it make a difference if she noticed the discrepancy while reviewing a draft paper for publication? Would it make a difference whether she was to be a co-author on the paper? Would if make a difference if she were not a student, but a professor of statistics in a different department of the same university? Would it make a difference if she and Hayes were professional colleagues in a cancer research laboratory? Would if make a difference if she were Hayes' advisor or mentor?

Source: ASA Ethics Case Study #2

Data Quality

A large company serves both government and private clients. During normal operations, it collects a huge amount of data which is unavailable anywhere else. These data are used internally and also used to meet information requests from clients, the media, researchers, and the general public. The data are shared with government agencies. In some cases, the data have significant social impact.

A statistician who meets requests for information based on these data is concerned because there are not control processes in place to assure uniform quality of the data. There are no audit procedures by which any particular counts or compilations could be verified independently. She feels that, on the whole, the data are probably "pretty good" but are likely to vary widely in quality from one data set to another. She has proposed creation of a statistical services group, which would institute data quality standards and procedures, as well as improve the availability of analytic products using these data. Her proposal has been applauded by management but perpetually left unfunded.

Colleagues with whom the statistician has discussed this matter point out that thousands of data sets lacking data quality standards exist and are widely used. They also point out that even where data quality control standards are in place, it can take years or decades to identify and resolve specific data quality problems. Still, the individual involved is highly uncomfortable ethically with her role in preparing compilations and reports on these data. She does not want her professional reputation on the line with such products given that the recipients do not know what they are getting. She is considering adding a disclaimer to each data product to inform customers about the lack of data quality control. She is also very tempted to resolve the issue by taking other available employment.

You are a close friend of this person, and she has asked for your advice. You are not employed by the same organization and do not know its internal politics or culture. Still, she values your judgment highly, especially in matters of professional ethics. Your advice is quite likely to be the deciding factor in her decision about what course to take.

What issues of statistical ethics are involved here?

What would you advise this statistician to do?

Does you advice change depending on the data subject matter, say demography versus transportation versus product safety data, utility services versus public health-related data?

Does your advice depend on the individual's level of responsibility in the organization, say technician versus middle management versus executive?

Source: ASA Ethics Case Study #3

Ethical Case No. 3

A non-statistician principal researcher requests a statistician providing analysis services to remove outliers from a data set. A report on the analysis, possibly to be submitted for publication, will be prepared by the principal researcher, who has indicated that under no circumstance will he include analytic results obtained with the outliers included.

Which of the following do you consider to be the most ethical response?

a Refuse to comply with the researcher's request.

b Perform the analysis with and without the outliers and provide comments on the implications.

c Perform the analysis with and without the outliers and, if it seems to matter, attempt to negotiate at least a mention by the researcher that the outliers were removed.

d Comply with the researcher's wishes.

Which, if any, of the following could have a substantial impact on your reasoning?

1. The nature of the relationship between the principal researcher and the statistician (e.g., coauthors, supervisor-subordinate, employer-paid consultant).
2. Vehicle for dissemination of the report (e.g., in house report, subject matter journal, statistical journal).
3. Acknowledgment of the removal of outliers in the written report.
4. Acknowledgment of the statistician by name in the report or publication.
5. Subject matter of the study or nature of the data and the use to be made of the results.

Source: *Amstat News* (early 1990s)

Ethical Case No. 4

Consider a research statistician, working for a non-profit institution, who has taken on a joint project with a non-statistician colleague. Assume that the colleague has collected data from a large group of human subjects and that, in doing so, s/he guaranteed anonymity to those subjects. A record was created for each subject containing, among other demographic data, name, age, sex, social security number, and a unique identification number. The study was concerned with both sensitive data (e.g. drug usage and sexual orientation) and non-sensitive data (e.g. food preference and product preference). This file was provided to the statistician.

After analysis and publication in a professional journal with the statistician as first author, a request was made by a non-statistician for a copy of the data base—for the stated purpose of replication. The statistician, citing the confidentiality assurances provided to the subjects, refused to provide either the entire data base or an extract omitting names and social security numbers. The individual then demanded of the journal that either the data base be provided or the article be retracted.

To what extent to you agree of disagree with each of the statements below? Do you consider any of them self evident?

1. The statistician should not have agreed to coauthor an article if s/he was unwilling to comply with a legitimate request for the underlying data.

2. The statistician should have supplied a file extract with name and social security number removed.

3. The statistician may ignore the confidentiality assurance since it was made by the second author and not the statistician.

4. Providing the names and social security numbers to the statistician was a breach of confidentiality. The non-statistician should not have provided this information to the statistician and the statistician should not have accepted it from the non-statistician.

5. My answers to the above would be the same regardless of

   (a) whether this was a market research study or a product liability study; or

   (b) whether the requestor was from a federal regulatory agency or a company competing with a study sponsor.

Source: *Amstat News* (early 1990s)

Job survival

After you received your doctorate in statistics last year you took the only statistical job you could find near your aged parents. It is with a large private for-profit hospital chain, where you are now the only statistician with any research responsibilities. You spend most of your time supporting some randomized clinical trials initiated during the tenure of your predecessor, who left on very short notice and for reasons referred to rather vaguely in terms of "mutual agreement." The hospital chain has been losing money, and the Board now counts on a "successful" research program to attract both referrals of patients and new investors.

The protocol for the largest of the ongoing trials, now coming to a close, clearly states the null hypothesis to be that cancer patients treated by a new Regimen Y will have the same 2-year survival as those treated by the standard Regimen X, but no statistical method for the comparison is specified. Your proportional hazards analysis, with adjustment for half a dozen covariates that are commonly used in the field, produces a two-tail p-value of 0.18. Disappointing, expecially since the new treatment seems to have caused a few deaths from drug toxicity. But the Principal Investigator, a world-renowned physician who once took a short course in the use of the statistical package SSIMPLE, says (correctly) that an exponential survival model with adjustment for only three covariates, one of them distinctly non-standard, yields p=0.04. He says that this is an important new "positive" finding; he has written it up as such; he will be mailing it to a leading journal tomorrow morning; you can be the sixth author, take it or leave it; and he considers publication to be essential to getting his next budget (with your salary in it) from the Board of Directors.

What do you do?

OK, what do you really do now?

Source: John Bailar and Mike Ginevan via ASA

Monitoring study

A statistician was called upon to provide design and analysis expertise for an exposure monitoring network at a large industrial facility. The study design was conducted without incident and resulted in a sampling network of multiple stations which provided information on particle and gas phase toxicants. Sampling continued for a number of months, so the network provided information on both spatial and temporal variability in toxicant concentrations. During the data analysis, which was designed to estimate worker exposures, the statistician discovered that there were substantial discrepancies between the results of gas and particle phase monitoring. Specifically, the data suggested that the particle phase monitors had failed periodically during the measurement period. Moreover, the pattern of failure was nonrandom, such that it raised an issue of deliberate sabotage. The statistician communicated these findings to the project manager. After some inquiries, the project manager informed the statistician that careful auditing had been done and that the results were accurate. However, he also asked the statistician to confine his reported analyses to the particle phase monitoring only, because the gas phase toxicant concentrations were too low to be of interest from a worker protection standpoint.

What should the statistician do? He was not a monitoring expert, but had learned a lot about the processes while developing the study design. On this basis, the monitoring results suggested problems. On the other hand, he had assurances that the monitoring results were accurate, but these came with a request which could be considered suspicious (without considering the gas phase data, the particle phase data seemed unremarkable). Should he submit a particle-phase-only report, as requested, or should he submit a complete report, which would likely jeopardize further consulting? Would it make a difference if he thought the particulate toxicant concentration would or would not threaten worker health? Should he try to "go over the project manager's head?" Should he consider "whistle blowing?" Why or why not?

Source: John Bailar and Mike Ginevan via ASA

Example 6.2

Scott Lasser is a statistician at a biostatistical consulting unit. Fro three years, part of his salary support came from a research grant awarded to Dr. Satya Sethuramen, a scientist in the field of nutrition. She was preparing to conduct a clinical trial of the effects of a specific vitamin on the health of nursing mothers and their infants. This study required that the study subjects be randomized into three groups that differed in the dosage of the vitamin to be given to the mother. The effect of the vitamin would be determined from period blood and milk samples from the mother, and blood samples from the infant, and other health measures. The study was expensive to conduct and the research budget was very limited. Based on some previous data in the literature, Scott estimated that Dr. Sethuramen would need about 30 mother-infant pairs in each of the three groups in order to detect a clinically important effect. Because of the limited budget and the cost of recruiting and enrolling mother-infant pairs, Dr. Sethuramen wanted Scott to analyze the data repeated as it accrued. She would then stop the study at the point at which statistical significance was obtained.

Scott was concerned about this; should he be? What should he do.

Source: Janice Derr 2000