

TRAINING STATISTICS GRADUATE STUDENTS IN ETHICS, REPRODUCIBLE RESEARCH, AND OTHER BEST PRACTICES

Jeffrey D. Dawson
University of Iowa College of Public Health, Iowa City, USA
jeffrey-dawson@uiowa.edu

Graduate students in many fields are required to receive formal training in scholarly integrity. Some institutions offer general courses to address this need, but statistics students may benefit from training more specific to the discipline. We have developed a “Scholarly Integrity in Biostatistics” course, which addresses the competencies mandated by several U.S. funding agencies, while emphasizing the key roles that statisticians have in collaborative research. The course meets one hour per week, and its topics include student/mentor relationships, communication skills, gender bias, conflict resolution, authorship, human and animal subjects’ issues, statistical review of manuscripts, plagiarism, copyright laws, conflicts of interest and commitment, reproducible research, and publication bias. Course credit and grades are based on attendance, class participation, short writing assignments, and a term project. The course has been offered four times, and has been received favorably by the students and by the administration.

INTRODUCTION

For decades, U.S. graduate students supported by pre-doctoral training grants have been required to receive instruction in ethical aspects of research, i.e., scholarly integrity. In recent years, this requirement has been expanded to include all graduate students supported by federal grants, even if such grants are not identified as training grants. More specifically, the U.S. National Institutes of Health (2009) has required the following topics to be covered:

1. Conflict of interest – personal, professional, and financial
2. Policies regarding human subjects, live vertebrate animal subjects in research, and safe laboratory practices
3. Mentor/mentee responsibilities and relationships
4. Collaborative research including collaborations with industry
5. Peer review
6. Data acquisition and laboratory tools; management, sharing and ownership
7. Research misconduct and policies for handling misconduct
8. Responsible authorship and publication
9. The scientist as a responsible member of society, contemporary ethical issues in biomedical research, and the environmental and societal impacts of scientific research

To meet these requirements, the Graduate College of the University of Iowa (UI) organized a multi-semester sequence in Scholarly Integrity in 2010. The format of this course included a 4-hour introductory event during orientation week, followed by two 2-hour workshops per semester. Students enrolling in MS programs who were supported by Federal grants were required to attend such workshops for two consecutive semesters, while their PhD counterparts were required to attend such workshops for four consecutive semesters. During the introductory event and the subsequent workshops, students heard general presentations and then were divided into groups of size 8-10, each led by a faculty facilitator, to discuss several case studies that illustrated the topics that were presented.

Although important content was delivered in the UI Graduate College offering, several important issues were raised. First, the course seemed to be primarily geared towards those in the laboratory sciences, resulting in less than stellar enthusiasm by those in other fields (e.g., statistics, biostatistics, mathematics). Second, several of the case studies seemed too vague, with no “right answer”, leading some students to conclude that it really does not matter what they do when they are faced with a difficult situation. Also, some faculty who served as small-group facilitators did

not feel that their department heads gave them adequate credit for the time that they put into the course, perhaps because it was not a course within their specific department.

In 2012, the UI Department of Biostatistics, housed within the UI College of Public Health, offered a pilot seminar course in scholarly integrity in biostatistics, with the audience being a few students who were participating in a T32 Pre-doctoral Training program. This course was taught jointly by Drs. Kathryn Chaloner and myself. Topics included reproducible research, data falsification, and some recent case studies that had received widespread media attention. Based on this pilot course, and to address the issues that had been raised with respect to the UI Graduate College course, Dr. Chaloner asked me to develop and teach a formal course in Scholarly Integrity in Biostatistics, with the intended audience being biostatistics and statistics graduate students. In this report, I give details of the development and content of this course.

COURSE DEVELOPMENT AND TOPICS

The topics from the pilot offering of the course were expanded to include the following topics: student/mentor relationships, communication skills, gender bias, conflict resolution, authorship, human and animal subjects' issues, statistical review of manuscripts, plagiarism, copyright laws, conflicts of interest and commitment, reproducible research, and publication bias. These topics were mapped to the NIH competencies listed above to ensure that the course met federal requirements; the mapping appears as Table 1. Additionally, it was noted that the course would help meet the following competencies of the UI Biostatistics MS and PhD programs:

- Effectively collaborate on a research team.
- Prepare reports and publications resulting from health science studies.
- Effectively communicate key statistical principles to non-statistical audiences.

This course (BIOS 7270, for one semester hour of credit) was submitted for consideration to the Department of Biostatistics, the College of Public Health curriculum committee, the UI Graduate College, and the UI Office of the Vice-President for Research in 2013, and was approved at all four levels. It has been offered during every spring semester since 2014. During the initial offering, it was noted that some course topics seemed to invoke more discussion than others, so discussion-invoking topics were moved to earlier in the semester as a way to improve camaraderie and discussion among the students for the entire course. It was also noted that the topics seemed to fall into three general areas: social and cultural issues, regulations, and best practices. Table 1 displays the topics covered in the course in order of coverage, the areas in which they fall, and the NIH competency numbers (as outlined in the Introduction section) that the topics address.

Table 1: Areas, topics, and NIH competency numbers (see NIH competency list in Introduction).

Area	Topic	NIH competency number(s)
Social and cultural issues	Responsibility: Taking ownership of a project	4, 6
	Proactively avoiding misconduct	7, 9
	Student/mentor relationships and communication	3
	Gender Bias	3, 7, 9
	Conflict resolution	3, 9
Regulations	IRBs and human/animal subjects	2, 6
	Authorship and peer review	5, 8
	Intellectual property	6
	Plagiarism and copyright laws	7
	Conflicts of interest and commitment	1, 9
Best practices	Being an effective collaborator	4, 6
	Organizing and documenting your work	4, 8
	Reproducible research	4, 8, 9
	Multiple comparisons and Type I/II errors	8, 9
	Publication bias	8, 9

The course also had the learning objectives that, at the end of the course, the student would be able to do the following:

- Understand the ethical importance of obtaining, managing, and storing high quality data.
- Document statistical analysis plans and organize computer programming code to assure integrity in the analytical aspects of the research process.
- Perform independent and collaborative research that maintains high standards of quality, integrity, and ethics.
- Communicate with their peers, supervisors, and future trainees more effectively, so that ethical issues can be proactively addressed in a timely manner.
- Reduce the occurrence of misunderstandings that may lead to unethical behavior.

SPECIFIC EXAMPLES OF TOPICS

Below are a few examples of how we introduced and discussed some specific topics, with emphasis on the topics that seemed to generate the most discussion.

Proactively avoiding misconduct

The U.S. Department of Health and Human Services Office of Research Integrity (2017) has developed an interactive video called “The Lab”, wherein the user sees the news reports of a hypothetical case of research misconduct, and then gets to choose which character to play as part of the interactive experience: a post-doctoral fellow, a graduate student, a principal investigator, or a research integrity officer. Once the user chooses the character to play, a video is shown that illustrates what happened before, during, and after the misconduct incident from the perspective of that character. Periodically, the video stops and the narrator asks the user to make a decision on behalf of the character on how to proceed, which then dictates the ensuing sequence of events.

In the Scholarly Integrity in Biostatistics class, the students are randomly assigned to play the role of a specific character, and then given the homework assignment to go through the video program and try various answers to the queries to see what happens. Then the students’ experiences and reactions are discussed in class, first in small groups and then with the entire class. Despite the fact that this video program is set in the context of a basic science laboratory, the biostatistics students seem to understand and appreciate the lessons learned from this video. It should be noted that none of the four characters who are the subject of role-playing are the perpetrator of the academic misconduct; however, each character can make decisions that can help preserve the integrity of the lab. Hence, it is in everyone’s best interest to be proactive in developing a working environment where research misconduct is less likely to occur.

Conflict resolution

There are several different taxonomies for describing styles of communication in the face of conflict. The Scholarly Integrity in Biostatistics course emphasizes the terminology used by UI Counseling services, as some students may end up using such services at some point. Quoting from the website of UI University Counseling Services (2017), these are:

- Withdrawal—Person tries to solve problem by denying its existence. Results in win/lose.
- Smoothing—Differences are played down; surface harmony exists. Results in win/lose in forms of resentment, defensiveness, and possible sabotage if issue remains suppressed.
- Forcing—One’s authority, position, majority rule, or a persuasive minority settles the conflict. Results in win/lose if the dominated party sees no hope for self.
- Compromise—Each party gives up something in order to meet midway. Results in win/lose if ‘middle of the road’ position ignores the real diversity of the issue.
- Collaboration--Abilities, values, and expertise of all are recognized; each person’s position, is clear but emphasis is on group solution. Results in win/win for all.

Although “collaboration” is generally preferred, we discuss situations where each of the above approaches are appropriate and inappropriate. We also consider the classic psychological labels of “passive”, “aggressive”, “passive-aggressive”, and “assertive” communication, and relate

these four terms to the five bulleted terms listed above. For example, “smoothing” may result in passive-aggressive behavior if feelings of resentment are unresolved, while “compromise” and/or “collaboration” are most likely to occur when someone is appropriately “assertive”.

The motivation for covering conflict resolution is that research misconduct often occurs in the presence of unresolved conflict and stress, as seen in video of “The Lab”. The intent is that students who use appropriate communication to resolve conflict will be more effective and happier students, graduate research assistants, and full-time statisticians, and will have more sound scholarly integrity.

Reproducible research

Peng (2009) has used the term “reproducible research” to mean, in part, that the analysis data set and the analysis code would be included in a manuscript submission. However, not all agree that such a submission process would really be helpful to improve research integrity. Keiding (2010) has stated that, “[The] number crunching of the [final dataset used for analysis] is a minor part of the totality of the statistical analysis, and it ridicules our profession to believe that there is a serious check on reproducibility in seeing if somebody else’s computer reaches the same conclusion using the same code on the same data set as the original statistician’s computer did.” Supporting Keiding’s statement is the fact that it is often necessary to merge data from several different sources (some confidential and/or proprietary in nature) into a final dataset that will be actually analyzed, and substantial errors that might happen during the merging process cannot be checked if one only shares the final dataset and code. Reproducible research continues to be an important topic that is discussed in meetings, presentations, and statistical blogs. In some cases, the emphasis is on a particular software, such as using the knitr package in R (Gandrud, 2015). Regardless of the tools used, the concept is very important.

In our class, we take a holistic approach to reproducible research, emphasizing the importance of organizing one’s work to make a clear path between the raw data and the analytical results. To illustrate this concept, we do a mock exercise wherein I pretend to receive an email that has a dataset attached and a request to calculate an odds ratio relating a numeric predictor to a binary response. In front of the class, I read in the data, run a logistic regression analysis to come up with an odds ratio and a confidence interval, and then email the result. The students time this process, and find that I can usually get these tasks done in less than two minutes. We then discuss the pitfalls of doing a rush job in this manner. Next, I forward the mock email and dataset to the class (complete with an intentional data entry error that I want them to catch), and ask them to spend 1-2 hours doing a more careful analysis, along with careful documentation of their work. After they turn in their results, I share with them a template that I have developed that can be used to organize such collaborative research, containing such information as the following:

- Specific goals of the study, including hypotheses
- Definitions of key variables
- Date of request, promised completion date, and actual completion date
- Source of the data
- Any eligibility criteria that were imposed on the data
- Simple descriptive statistics
- Primary results
- Next steps to take
- Path names and filenames to run the analyses
- Output from statistical software in an appendix

I have found that the items corresponding to the first eight bullets can often fit on one or two pages, which helps clarify the scientific questions and answers. The path names and filenames in the next-to-last bullet are extremely helpful when one needs to repeat or update the analyses, or when it is necessary to hand the project over to a new analyst (which happens often when a project lasts for several years). The final bullet, the output, is very helpful when putting together any written deliverable based on the analyses, such as a report or a grant application.

EVALUATION AND STUDENT PROJECTS

Students are assigned points based on attendance, class participation, short assignments, and a course project. The choice of project is primarily left up to the students, who may work alone or in groups of size two or three. The general instruction is that students may prepare relevant case studies and/or do a follow-up presentation on a topic that was covered in class. Students are graded on their oral presentations (10-20 minutes, depending on class size and group size) as well as polished Powerpoint or PDF slides that the students turn in during finals week. Some examples of student projects include the following.

- Gender bias
- Publication/media bias
- Vioxx case study
- Importance of accommodating random effects in data analyses
- False positive findings (Ioannidis, 2005)
- Important sample size calculations
- Data encryption
- Ethical issues in the peer review process
- Use/misuse of journal impact factors
- Reducing import errors when accessing multiple datasets
- Case studies unethical research on vulnerable populations
- Anti-vaccination controversy
- Knitr package in R

STUDENT EVALUATIONS OF THE COURSE

All registered students are encouraged to complete anonymous course evaluations during the last week of class, the results of which are not shared with the instructor until after the instructor has submitted final grades to the registrar's office. The evaluation forms contain a total of 24 items, dealing with matters concerning the course itself, as well as issues related to the skills, effort, and effectiveness of the instructor. Each item is in the form of a positive statement, e.g., "Class time is used efficiently", and the students are asked to indicate their level of agreement with the statement on a scale of 1 to 6, with 6 being the highest level of agreement.

Overall, the student evaluations of the course have been quite positive. Across all four years of the course's offering, the mean of the average rating has been 5.54 (across-question SD= 0.25). Table 2 lists the six questions that relate, at least in part, to the content and usefulness of the course, rather than solely to the instructors efforts. For these six questions, the average rating across students varies from 5.01 to 5.62.

Table 2: Mean Responses for Course-Related Items from Student Evaluations, 2014-2017
(Scale of 1 to 6)

Item	Mean Response
I have developed a greater appreciation for this subject.	5.29
I have learned a great deal in class.	5.01
Important points are clarified with good examples.	5.62
This course requires an appropriate amount of work.	5.54
The instructor increase my interest in the course material.	5.23
Overall, this is an excellent course.	5.23

In addition to numerical scores, the students are also encouraged to make specific comments and suggestions. After the first offering in 2014, a common theme of the comments was that topics should be arranged to encourage more discussion during the first few weeks of the course, as it turned out that the liveliest discussions seemed to occur during the last couple of weeks of the course. Additionally, there was some indication that although the topics were interesting and important, they were somewhat segmented in nature. In response to these sorts of comments, and as described earlier in this paper, the organization that is listed in Table 1 was

developed and has been used since 2015. This restructuring seemed to improve the course, as all six items listed in Table 2 showed improvement over time from 2014 to 2017.

CONCLUSION

The UI Graduate College has phased out its general course in Scholarly Integrity, and now requires individual departments or colleges to develop their own versions of the course. The syllabus of Scholarly Integrity in Biostatistics is used as a template by some other units when developing their own courses. In all, 67 students have taken this course, including two advanced PhD students who are taking it a second time, as part of retraining required by the graduate college. In addition to graduate students in biostatistics and statistics, some students from mathematics, pharmacy, medicine, epidemiology, and dentistry have completed the course. The course will soon reach a somewhat steady state of approximately 10-14 first-year students and 5-6 advanced students per year, with the advanced students acting as peer mentors for less experienced students.

Scholarly integrity involves more than being honest when doing research. The word, “integrity”, implies completeness, robustness, and strength, such as when we think about the importance of the structural integrity of a bridge or a building. Thus, scholarly integrity necessitates being skilled in many key “best practices” for collaborative statisticians, such as communicating clearly and doing research that is well organized and reproducible. If we provide appropriate education to our students concerning scholarly integrity, this will help them avoid many of the pitfalls of scientific research.

REFERENCES

- Gandrud, C. (2015). *Reproducible Research with R and RStudio* (2nd ed.). Boca Raton, FL: CRC Press.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 696–701.
- Keiding, N. (2010). Reproducible research and the substantive context. *Biostatistics*, 11(3), 376–378.
- National Institutes of Health. (2009). *NOT-OD-10-019: Update on the Requirement for Instruction in the Responsible Conduct of Research*. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-10-019.html>
- Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics*, 10(3), 405–408.
- UI University Counseling Services (2017). *Five Basic Methods for Resolving Conflict*. Retrieved from <https://counseling.uiowa.edu/self-help/five-basic-methods-for-resolving-conflict/>
- U.S. Department of Health & Human Services, Office of Research Integrity (2017). *The Lab*. Retrieved from <https://ori.hhs.gov/thelab>