

Standards of Scientific Conduct: Disciplinary Differences

Michael Kalichman · Monica Sweet · Dena Plemmons

Received: 21 March 2014 / Accepted: 10 September 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Teaching of responsible conduct of research is largely predicated on the assumption that there are accepted standards of conduct that can be taught. However there is little evidence of consensus in the scientific community about such standards, at least for the practices of authorship, collaboration, and data management. To assess whether such differences in standards are based on disciplinary differences, a survey, described previously, addressing standards, practices, and perceptions about teaching and learning was distributed in November 2010 to US faculty from 50 graduate programs for the biomedical disciplines of microbiology, neuroscience, nursing, and psychology. Despite evidence of statistically significant differences across the four disciplines, actual differences were quite small. Stricter measures of effect size indicated practically significant disciplinary differences for fewer than 10 % of the questions. This suggests that the variation in individual

Electronic supplementary material The online version of this article (doi:[10.1007/s11948-014-9594-0](https://doi.org/10.1007/s11948-014-9594-0)) contains supplementary material, which is available to authorized users.

M. Kalichman (✉)

Research Ethics Program, University of California, San Diego, 9500 Gilman Drive #0612, La Jolla, CA 92093-0612, USA
e-mail: mkalichman@ucsd.edu

M. Sweet

Department of Psychology and CREATE, University of California, San Diego, 9500 Gilman Drive #0036, La Jolla, CA 92093-0036, USA
e-mail: msweet@ucsd.edu

D. Plemmons

Department of Anthropology and Research Ethics Program, University of California, San Diego, 9500 Gilman Drive #0612, La Jolla, CA 92093-0612, USA
e-mail: dplemmons@ucsd.edu

D. Plemmons

Division of Research Affairs, San Diego State University, San Diego, CA, USA

standards of practice within each discipline is at least as great as variation due to differences among disciplines. Therefore, the need for discipline-specific training may not be as important as sometimes thought.

Keywords Responsible conduct of research · Research ethics · Standards · Authorship · Collaboration · Data management

Introduction

Despite over 20 years of National Institutes of Health (NIH) requirements for responsible conduct of research (RCR) education (NIH 1989), and prescriptions for teaching particular topics (NIH 2009), it can be argued that the requirements are lacking in at least two ways. First, none of the existing NIH or National Science Foundation (NSF) requirements is specific about what should be taught about those topics. For example, three topics commonly considered to be part of RCR education (e.g., NIH 2009) are “collaborative research including collaborations with industry”, “data acquisition and laboratory tools”, and “responsible authorship and publication”. To what extent are there standards of practice that can be said to be commonly held in any of these three domains? Based on a recent study of faculty from four disciplines (microbiology, neuroscience, nursing, and psychology), the answer is that there is little common ground to serve as a basis for teaching (Kalichman et al. 2013).

Second, existing requirements for RCR education do not resolve whether teaching needs to be handled primarily at the departmental or disciplinary level. The NIH (2009) encourages “participation of research training faculty members in instruction in RCR,” but that does not necessarily mean that the courses are conducted only for people from the same department or discipline. Many institutions (e.g., UC San Diego) have largely approached RCR education through courses that are open to trainees and faculty from diverse disciplines. The question remains open as to whether this approach is “right,” or if training should be left solely to the departments or specific disciplines.

Several arguments can be made in favor of discipline-specific training. It appears that what is taught in the research setting (i.e., through mentoring) has more of a positive impact on researchers than what is taught in formal courses (Anderson et al. 2007). This makes intuitive sense since research trainees spend much more time with their mentors than in any one research ethics course. Further, it is likely that there will be issues and standards to be learned that are specific to a research discipline or “community of practice” (Becher and Trowler 2001; Wenger 1998). Consistent with this emphasis on discipline-specific training, some have argued that courses—and, by inference, educational interventions to promote good research practices—are best “taught by those at the cutting edge of research” (Lee 2004). However, despite these arguments in favor of discipline-specific training, it is important to balance that goal with whether research faculty have the time, will, motivation, and even knowledge to carry out such training (Kalichman 2013). And

whether or not these hurdles can be overcome, standards vary widely across the scientific community (Kalichman et al. 2013), suggesting that much is to be gained from cross-disciplinary education and much lost if RCR education is relegated to the department level.

To provide a baseline for discussing responsible practices, faculty from four different disciplines (microbiology, neuroscience, nursing, and psychology) were queried about standards and practices in each of three domains: authorship, collaboration, and data management (Kalichman et al. 2013). In addition, the faculty were asked for their perceptions of how researchers learn those standards. As noted above, the overall finding of that study was that standards in these three domains are sufficiently variable that there is little agreement about what standards might form an RCR curriculum. However, that wide variation in standards may have a hidden implication for discipline-specific training. Specifically, it is possible that there is consensus within specific research disciplines and the perceived variation in standards is due to differences among those disciplines. The goal of this study is to test that possibility by comparing the target disciplines and summarizing the extent to which those disciplines differ from one another.

Methods

This survey study was reviewed and approved by the UC San Diego Institutional Review Board (Protocol #101447SX). The final survey consisted of 132 distinct questions divided among 5 sections: authorship, collaboration, data management, teaching and learning, and demographics (Kalichman et al. 2013; Appendix of supplementary material). The survey was developed with a multi-step process involving an expert panel, focus groups, interviews, selection of graduate programs, selection of faculty sample, and a pilot test. The survey was conducted between November 2010 and March 2011, and resulted in a total of 6,616 valid e-mail invitations. Further details about the preparation of the survey and the survey itself can be found supplementary material.

Analysis of cross-discipline differences consisted of a series of ANOVAs for the Likert questions (5-point scale with 1 = strongly disagree and 5 = strongly agree) and X^2 analyses for the forced-choice yes/no questions, with discipline as the independent variable. Significant results were considered preliminary evidence of disciplinary differences. Finding statistical significance, however, did not always equate to finding substantial cross-discipline differences. Because the number of respondents was sufficiently large, statistical significance might be found even for very small actual differences in averages across disciplines (e.g., for Q96, “In my experience, standards of conduct can be taught explicitly for authorship.” mean Likert scores ranged only from 4.18 to 4.47 across the four disciplines, but the difference was still highly statistically significant, $p < .0001$). For this reason, stricter standards were also defined for practically (i.e., more likely to be of practical importance) significant cross-disciplinary differences: large average differences (>2 points, or 40 % on the Likert scale) and medium average differences (>1 point, or 20 % on the Likert scale) across disciplines.

Results

Faculty from each of four graduate disciplines were invited to participate in the survey, but participants were also asked to self-identify the discipline with which they most closely identified. The breakdown of response rates and self-reported disciplinary affiliation by the original discipline designations is shown in Table 1. While self-identifications typically matched the pre-selected disciplinary affiliations, this was not always the case. Within the four disciplines, the highest response rate based on departmental affiliations was for nursing (27 %), while response rates for the other 3 disciplines ranged from 18 to 20 %. These numbers would change nominally based on differences in self-identification; however, analyses were based on departmental affiliation rather than self-identified discipline.

Results of cross-discipline analyses are summarized in Table 2. Over 80 % of statements resulted in statistical significance of discipline ($p < 0.05$), and 67 % were highly statistically significant ($p < 0.001$). These statistically significant differences among the four disciplines initially suggested a *lack of cross-disciplinary consensus* (i.e., disciplines were different from one another). However, on closer examination of the magnitude of the differences, it became clear that average differences were often quite small. Further analysis of the magnitude of observed differences was conducted using defined a priori standards for practical significance. Using a standard of large average differences (>2 points or 40 % on the Likert scale), practical significance was achieved for only 1 % of the questions. Even using a more liberal standard for differences (>1 point or 20 % on the Likert scale), practically significant cross-discipline differences were found for only 9 % of questions.

Further analysis of those cases with a difference of at least one point between two or more of the four disciplines revealed a few primary findings (Table 3). First, only 8 of the 94 questions revealed even this small one point average difference. This was the case for two questions about authorship, one on collaboration, five on the topic of data management, and none in the category of teaching and learning. Second, only one question resulted in a difference between disciplines of 2 points or more (Q78): “In my experience with research records and collection of raw/primary

Table 1 Graduate program affiliation for survey respondents

Discipline	Response rate (%)	Valid invitations	Respondents	Self-identification
Microbiology	20	1,221	241	254
Neuroscience	18	2,876	522	472
Nursing	27	1,531	414	408
Psychology	18	1,249	219	247
Total	20	6,877	1,396	1,381

Respondents (N = 1,396) were recruited from the named graduate program disciplines in US, Universities in 2010. Response Rate calculated as percent of Respondents who had received Valid Invitations. Invitations were sent to individuals in each discipline based on named affiliation on websites for one of the 50 selected graduate programs in that discipline. Some Respondents selected for a given discipline self-identified for different disciplines

Table 2 Differences across four research disciplines (microbiology, neuroscience, nursing, and psychology) in responses to 94 Likert scale (1 = strongly disagree; 5 = strongly agree) statements and questions on the topics of authorship, collaboration, data management, and teaching and learning

	Authorship	Collaboration	Data management	Teaching and learning	Total
# of questions	31	24	21	18	94
% Statistically significant ($p < 0.05$)	94	79	86	67	83
% Highly statistically significant ($p < 0.001$)	87	42	71	61	67
% Not statistically significant	6	21	14	33	17
% Significant based on >2-point difference	0	0	5	0	1
% Significant based on >1-point difference	6	4	23	0	9

Respondents (N = 1,396) were recruited from graduate programs in each of the four research disciplines in US. Universities in 2010. Statistical significance was tested by one-way ANOVA, and practical significance was assessed by finding more than a 2- or 1-point difference in average responses to 5-point scale Likert questions

data, research trainees are permitted to take copies of research records or data when they leave for a new research position.” The biggest disciplinary difference in this case was between microbiology (average Likert response = 3.9) and nursing (average response = 1.6). Although microbiology rankings were more than neuroscience, and psychology rankings were more than nursing, *both* nursing and psychology scored this question at least one point lower than either microbiology or neuroscience. Third, there was a very clear pattern of responses to these questions. Microbiology and neuroscience were typically very similar to one another, as were nursing and psychology. Further, with one exception, scores were statistically significantly lower for both nursing and psychology as compared to microbiology and neuroscience. The exception (Q75), regarding a preference for electronic over printed research records, resulted in much higher agreement from nursing and psychology as compared to microbiology and neuroscience.

Discussion

In an initial summary of this study (Kalichman et al. 2013), consensus about standards of conduct was minimal, if not absent, across all survey respondents. However, it remained possible that the apparent lack of consensus was due to differences among research disciplines and that standards within any one discipline might come closer to consensus. To test this possibility, responses from each of the four research disciplines (microbiology, neuroscience, nursing, and psychology) were compared statistically to assess whether, in fact, two or more of these disciplines differed significantly from one another. Assessed in this way, differences among the disciplines were found to nearly always reach statistical significance.

Table 3 Authorship, collaboration, and data management survey items for which average (\pm standard deviation) responses differed by 1 point or greater between 2 or more of the disciplines surveyed (microbiology, neuroscience, nursing, psychology)

	Microbiology	Neuroscience	Nursing	Psychology
Authorship				
Q23: In my opinion, it is <i>not possible</i> to establish criteria for authorship before the work begins	3.0 \pm 1.3	2.8 \pm 1.3	1.9 \pm 1.1	2.2 \pm 1.1
Q32: In my opinion, an individual <i>should</i> be listed as an author even if her or his sole contribution was to provide the funding for the research project	3.4 \pm 1.4	3.2 \pm 1.4	2.2 \pm 1.2	2.6 \pm 1.4
Collaboration				
Q47: In my experience, the term "collaboration" is reserved for relationships between peers, not between senior faculty and post-docs or graduate students	2.9 \pm 1.5	2.6 \pm 1.4	2.0 \pm 1.2	1.6 \pm 1.0
Data management				
Q72: In my experience, standards for data management are typically more stringent in industry than in academia	4.1 \pm 1.0	3.7 \pm 1.1	2.7 \pm 1.2	2.6 \pm 1.4
Q73: In my experience with research records and collection of raw/primary data, research records are summarized in a bound lab notebook, with numbered pages, dated and written in ink	3.6 \pm 1.3	3.6 \pm 1.3	2.4 \pm 1.3	2.0 \pm 1.2
Q75: In my experience with research records and collection of raw/primary data, electronic records are preferable to printed research records	2.8 \pm 1.1	3.1 \pm 1.1	3.9 \pm 1.0	3.7 \pm 1.1
Q78: In my experience with research records and collection of raw/primary data, research trainees are permitted to take copies of research records or data when they leave for a new research position	3.9 \pm 1.2	3.6 \pm 1.3	1.6 \pm 0.9	2.6 \pm 1.3
Q90: Research records should be retained indefinitely	3.7 \pm 1.3	3.4 \pm 1.4	2.5 \pm 1.4	2.8 \pm 1.4

In all cases, except Q75, Tukey post hoc pairwise comparisons were statistically significant for both nursing and psychology being less than microbiology and neuroscience. For Q75, both nursing and psychology were statistically significantly greater than microbiology and neuroscience

However, the magnitude of the differences was generally quite small. To identify substantive difference among the disciplines, the standard for differences was defined as a >2 -point difference on the 5-point Likert scale across the four disciplines. With this relatively modest standard, only 1 of 94 statements (1 %) met the criterion for differences among the four disciplines. Even for a more modest definition of differences defined as a greater than one point difference, disciplinary differences were found for only 9 % of the questions. In short, the patterns of

responses for the four research disciplines were largely similar to one another. To be clear, this does not mean that the individual disciplines were in agreement about standards of conduct. Instead, this meant that the various disciplines were similarly diverse in the range of answers provided by the survey respondents.

Given the length of the survey—132 total questions including 94 Likert scale questions—it is arguably surprising that faculty response rates were as high as they were (20 %) without incentives to participate. However, within this set of respondents, there is one noteworthy difference that deserves comment. While response rates ranged between 18 and 20 % for the disciplines of microbiology, neuroscience, and psychology, the rate among nursing faculty was considerably higher: 27 %. This substantial difference might represent something distinctive about nursing researchers *per se*, but it is worth noting that the nursing participants were characterized by a much higher percentage of female respondents than for the other three disciplines.

It might be hypothesized that differences associated with the nursing group might reflect either disciplinary differences or differences in gender response rates. Across all disciplines, female respondents were more highly represented than male (53 vs. 47 %), but this was almost entirely due to the high percentage of female respondents from nursing (90 %) and, to a lesser extent, psychology (52 % female respondents). Both microbiology and neuroscience respondents were more often than not male (just over 30 % female respondents in both cases). Despite these possible factors in response rates, and in responses, it is important to emphasize that answers about acceptable standards varied widely across all survey respondents, within the individual disciplines, and for those identifying themselves as male or female. According to the NSF Survey of Science and Engineering Indicators, women now comprise just over half of the faculty across all biomedical sciences (NSF 2012). However, by discipline, the percentage of female faculty in neuroscience is on the order of 21–30 % (Lorden et al. 2011), 32 % in microbiology (AAMC 2012), and 45–48 % in psychology (American Psychological Association 2012), but over 90 % in nursing (American Association of Colleges of Nursing 2001). These percentages are sufficiently close to the response rates by discipline for females in this study to suggest that differences among the disciplines were not due to under or over sampling by gender.

One consistent pattern to emerge was that in those cases where differences were most dramatic, nursing and psychology were typically more similar to one another than to microbiology and neuroscience (Table 3). While this study was designed to see if there were differences, it cannot answer the question as to *why* those differences occur. Some possible reasons for differences include: (1) structural characteristics of research done in each of these disciplines (e.g., perhaps much of microbiology and neuroscience dictates on balance a different approach to research than in nursing or psychology); (2) the type of people who choose to work in each of these disciplines; and/or (3) the degree to which research involves human subjects. The latter may be particularly important for Q78 (“In my experience with research records and collection of raw/primary data, research trainees are permitted to take copies of research records or data when they leave for a new research position.”). Nurses scored this question lower than any other discipline, at an average of 1.6 on

the Likert scale. Psychologists scored this item considerably higher, at 2.6, but still much less than found for microbiology and neuroscience, 3.9 and 3.6, respectively. While not all researchers in nursing and psychology will necessarily conduct studies involving human subjects, it is more likely than not to be the case as compared to researchers in microbiology and neuroscience. If so, it makes sense that there would be more reluctance among psychology and nursing researchers to have trainees taking sensitive human research records with them when leaving for a new research position.

Given the pattern of disciplinary differences (i.e., microbiology and neuroscience vs. nursing and psychology), it is of interest to see which questions resulted in differences of at least one point (or 20 % of the Likert scale) (Table 3). On the subject of authorship, microbiologists and neuroscientists were more skeptical of the possibility of establishing authorship criteria before the work begins on a project (Q23) and more open to assigning authorship credit even if the sole contribution was to provide funding for the project (Q32). On the subject of collaboration, microbiologists and neuroscientists were less likely than nurses and psychologists to think of trainees (e.g., postdocs and graduate students) as collaborators (Q47). Finally, on the subject of data management, microbiologists and neuroscientists perceived stronger standards in industry than in academia (Q72), but favored a “bound lab notebook, with numbered pages, dated and written in ink” (Q73), and were more inclined to allow trainees “to take copies of research records or data when they leave for a new research position” (Q78) and to keep research records indefinitely (Q90). Given these differences “on average,” it is worth cautioning researchers against assuming wide agreement on standards for recordkeeping, having access to copies of research records, or how long research records should be retained.

The results of this study provide a compelling argument for multi-disciplinary training in RCR, as opposed to dependence solely on discipline-specific training. Even with a liberal definition of disciplinary differences (i.e., differences between disciplines could be as small as one point on the Likert scale), differences were found only 9 % of the time. This means that patterns of commonly accepted standards are largely no different across the four disciplines. More importantly, this does not mean that the disciplines exhibited uniformity in their agreement (or disagreement) with the various questions asked. In fact, as noted previously (Kalichman et al. 2013), there was typically little agreement among scientists surrounding the issues covered in this survey. In short, any semblance of discipline-specific response pattern is largely lost in the noise of individual differences in defining the appropriate standards of scientific practice as a whole. While this project cannot distinguish whether those differences are in the best interests of science, teaching of RCR should be grounded in a recognition of the variations that occur across disciplines as well as among individual researchers. Correspondingly, it will be in the best interests of increasingly multi-disciplinary science to have those scientists learning about RCR in the context of multi-disciplinary discussions.

Acknowledgments Many individuals provided invaluable perspectives at each stage of this study, but the authors particularly want to thank the following for their expertise and guidance: Daniel Cabrera

(Northern Illinois University), Paul Friedman (UC San Diego), Elizabeth Heitman (Vanderbilt University), Francis Macrina (Virginia Commonwealth University), Joan Sieber (California State University East Bay), Connie Ulrich (University of Pennsylvania), David Urban (Virginia Commonwealth University), and Daniel Vasgird (West Virginia University). The authors also thank Tiffany Lagare and Kelli Wing for their assistance in collecting names and e-mail addresses for faculty surveyed in this study. This project was supported by NIH NR009962, UL1RR031980, and UL1TR000100.

Ethical standard Ethical approval for this study involving human subjects was granted.

References

- AAMC. (2012). Women in U.S. academic medicine and science: Statistics and benchmarking report 2011–2012; table 3: Distribution of full-time faculty by department, rank, and gender, 2012. <https://members.aamc.org/eweb/upload/Women%20in%20U%20S%20Academic%20Medicine%20Statistics%20and%20Benchmarking%20Report%202011-20123.pdf>. Accessed 23 Sept 2014.
- American Association of Colleges of Nursing. (2001). Effective strategies for increasing diversity in nursing programs. <http://www.aacn.nche.edu/aacn-publications/issue-bulletin/effective-strategies>. Accessed 23 Sept 2014.
- American Psychological Association. (2012). Graduate study in psychology. Faculty in U.S. and Canadian graduate departments of psychology: 2010–2011. <http://www.apa.org/workforce/publications/12-grad-study>. Accessed 23 Sept 2014.
- Anderson, M. S., Horn, A., Risbey, K. R., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. *Academic Medicine*, 82, 853–860.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines* (2nd ed.). Buckingham: SRHE and Open University Press.
- Kalichman, M. (2013). Rescuing RCR education. *Accountability in Research*, 21(1), 68–83.
- Kalichman, M., Sweet, M., & Plemmons, D. (2013). Standards of scientific conduct: Are there any? *Science and Engineering Ethics*, 1–12. doi:10.1007/s11948-013-9500-1.
- Lee, R. (2004). Research and teaching: making—or breaking—the links. *Planet*, 12, 9–10.
- Lorden, J. F., Kuh, C. V., Voytuk, J. A., (Eds.). (2011). An assessment of research-doctorate programs: Panel on the biomedical sciences; research-doctorate programs in the biomedical sciences: Selected findings from the NRC assessment. Washington, DC: National Academies Press (US); 2011. 6, Neuroscience and Neurobiology: Combining Data from the Program and Student Surveys. <http://www.ncbi.nlm.nih.gov/books/NBK82472>. Accessed 23 Sept 2014.
- NIH. (1989). Requirements for programs on the responsible conduct of research in national research service award institutional training programs. *Guide for Grants and Contracts*. 18(45), 1. http://grants.nih.gov/grants/guide/historical/1989_12_22_Vol_18_No_45.pdf. Accessed 23 Sept 2014.
- NIH. (2009). Update on the requirement for instruction in the responsible conduct of research. Notice number: NOT-OD-10-019. <http://grants.nih.gov/grants/guide/notice-files/not-od-10-019.html>. Accessed 23 Sept 2014.
- NSF. (2012). Chapter 3. Science and engineering labor force. Science and Engineering Indicators 2012. <http://www.nsf.gov/statistics/seind12/c3/c3h.htm>. Accessed 23 Dec 2013.
- Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.