

A Survey on Data Reproducibility and the Effect of Publication Process on the Ethical Reporting of Laboratory Research



Delphine R. Boulbes¹, Tracy Costello², Keith Baggerly³, Fan Fan¹, Rui Wang¹, Rajat Bhattacharya¹, Xiangcang Ye¹, and Lee M. Ellis^{1,4}

Abstract

Purpose: The successful translation of laboratory research into effective therapies is dependent upon the validity of peer-reviewed publications. However, several publications in recent years suggested that published scientific findings could be reproduced only 11% to 45% of the time. Multiple surveys attempted to elucidate the fundamental causes of data irreproducibility and underscored potential solutions, more robust experimental designs, better statistics, and better mentorship. However, no prior survey has addressed the role of the review and publication process on honest reporting.

Experimental Design: We developed an anonymous online survey intended for trainees involved in bench research. The survey included questions related to mentoring/career development, research practice, integrity, and transparency, and how the pressure to publish and the publication process itself influence their reporting practices.

Results: Responses to questions related to mentoring and training practices were largely positive, although an average of approximately 25% did not seem to receive optimal mentoring. A total of 39.2% revealed having been pressured by a principle investigator or collaborator to produce "positive" data. About 62.8% admitted that the pressure to publish influences the way they report data. The majority of respondents did not believe that extensive revisions significantly improved the manuscript while adding to the cost and time invested.

Conclusions: This survey indicates that trainees believe that the pressure to publish affects honest reporting, mostly emanating from our system of rewards and advancement. The publication process itself affects faculty and trainees and appears to influence a shift in their ethics from honest reporting ("negative data") to selective reporting, data falsification, or even fabrication. *Clin Cancer Res*; 24(14); 3447–55. ©2018 AACR.

Introduction

The successful translation of laboratory research into effective new treatments is dependent upon the validity of peer-reviewed publications. Scientists performing research in either academia or pharmaceutical companies, and developing new cancer therapeutics and biomarkers, use these initial published observations as the foundation for their projects and programs. In 2012, Begley and Ellis reported on Amgen's attempts to reproduce the seminal findings of 53 published studies that were considered to support new paradigms in cancer research; only 11% of the key findings in these studies could be reproduced (1). In 2011, another scientific team from Bayer pharmaceuticals reported being unable to repro-

duce 65% of the findings from a different selected set of biomedical publications (2). In 2013, "The Reproducibility Project: Cancer Biology" was launched. The project aims to reproduce key findings and determine the reliability of 50 cancer articles published in *Nature*, *Science*, *Cell*, and other high-impact journals (3). Final results should be published within the next year, but in the initial five replication studies already completed, only two of the manuscripts had their seminal findings confirmed (<https://elifesciences.org/articles/23693>).

Ever since the Begley and Ellis report in 2012, several surveys (refs. 4, 5; American Society for Cell Biology go.nature.com/kbzs2b) have attempted to address the issue of data reproducibility by elucidating the fundamental causes of this critical problem, ranging from honest mistakes to outright fabrication. According to a survey published in May 2016, compiling responses from 1,500 scientists and *Nature* readers (5), 90% of respondents acknowledged the existence of a data reproducibility crisis. More than 70% of researchers reported that they had failed to reproduce the results of published experiments, and more surprisingly, more than 50% of them reported that they had failed to reproduce the same results from their own experiments. The survey revealed that the two primary factors causing this lack of reproducibility were pressure to publish and selective reporting. Very similar results were found through another online survey published by the American Society for Cell Biology, representing the views of nearly 900 of its members (see go.nature.com/kbzs2b). According to respondents who took the survey for *Nature* (5), possible solutions to the data reproducibility problem

¹Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas. ²Office of Postdoctoral Affairs, Moffitt Cancer Center, Tampa, Florida. ³Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas. ⁴Department of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Lee M. Ellis, University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Unit 1484, Houston, TX 77030-1402. Phone: 713-792-6926; Fax: 713-745-1462; E-mail: lellis@mdanderson.org

doi: 10.1158/1078-0432.CCR-18-0227

©2018 American Association for Cancer Research.

Translational Relevance

The successful translation of laboratory research into effective new treatments is dependent upon the validity of peer-reviewed published findings. Scientists developing new cancer therapeutics and biomarkers use these initial published observations as the foundation for their projects. However, several recent publications suggested that published scientific findings could be reproduced only 11% to 45% of the time. We developed an anonymous survey specifically for graduate students and postdoctoral fellows involved in bench research. The results indicate that the pressure to publish and the publication process greatly affect the scientific community and appear to influence a shift in their ethics from honest reporting to selective reporting or data falsification. We believe these findings may have an impact on the scientific community regarding our methods of mentoring and training, and the importance of honest reporting, that should preempt the temptation to present data simply to publish in a journal with a higher impact factor.

include more robust experimental designs, better statistics, and, most importantly, better mentorship. One third of respondents said that their laboratories had taken concrete steps to improve data reproducibility within the past 5 years.

Because only a minority of graduate students seek and obtain a career in academic research (6, 7), we sought to determine whether these young scientists were concerned about issues of research integrity and the publication process. This is particularly important in an era of "publish or perish" and the pressure to publish in journals with a high impact factor. In the current study, we asked postdoctoral fellows and graduate students whether the pressure to publish is a factor in selective reporting and transparency and how the publication process itself influences selective reporting and transparency.

Materials and Methods

Survey content

We developed a 28-item anonymous online survey that included questions related to mentoring and career development, research practice, integrity and transparency, and the publication process, as well as general questions aiming to characterize the population of respondents (field of expertise and career goals). Only responses from graduate students and postdoctoral fellows involved in bench research were considered in our analysis. After general questions aiming to characterize the population of respondents (field of expertise and career goals), participants were asked additional questions concerning their perceived academic pressure, mentoring, best research practices, career advancement requirements, and publication standards. The survey was first validated by a randomly selected sample of 20 investigators in the Texas Medical Center. This study and questionnaire were approved by the MD Anderson Institutional Review Board.

Survey method

In July 2016, the survey was sent to graduate students and postdoctoral fellows in the Texas Medical Center [The University of Texas MD Anderson Cancer Center, University of Houston, Rice

University, Baylor College of Medicine, University of Texas Health Science Center at Houston (UTHealth), Houston Methodist Hospital, Texas A&M University, and Texas Children's Hospital] via multiple listservs. Reminder e-mails were sent out approximately once per month for the following year. In April 2017, in order to increase the power of our study, the population was extended to graduate students and postdoctoral fellows affiliated with the National Postdoctoral Association and Moffitt Cancer Center via additional listservs, and the survey was also distributed via Twitter and LinkedIn (via the corresponding author, L.M. Ellis). The e-mail or social media invitation provided a link to the study that included a consent statement, description of the study, and the survey items. The survey was closed in July 2017.

Survey analyses

Many of the survey questions allowed the respondent to select "other, please explain." When respondents selected this option and provided comments, these comments were either extrapolated to fit into one of the original response choices or used to create new response categories. Most responses were compiled as percentages of the total number of responses received to a specific question. When several answers could be selected, the responses were compiled in absolute values. In addition, throughout the survey and in particular at the end, respondents had the opportunity to share thoughts and comments that they believed could be relevant to the issue of data reproducibility. Of note, after reviewing every comment, we chose to redact the parts of the comments that we deemed could have led to the identification of the respondent because of their specificity (uncommon field, institution name, . . .). The parts redacted were replaced by [. . .].

Results

Respondent characteristics

With our eligibility criteria of (1) being a graduate student or postdoctoral fellow and (2) performing bench science, 467 of our total 576 respondents were deemed eligible. The raw data are provided in Supplementary Data S1. Throughout the survey and, in particular, the final question, we gave the respondent the opportunity to share thoughts and experiences that they deemed relevant to this issue. Slightly more than 100 respondents (~20%) decided to share their opinion with us in the open-ended comment section at the end of the survey. Comments were compiled in a separate file (Supplementary Data S2).

Respondent characteristics are summarized in Table 1. Roughly 10.7% of respondents were graduate students, and 89.3% were postdoctoral fellows. To ensure inclusivity, we considered respondents in all scientific fields eligible, but our easy access to MD Anderson Cancer Center and Moffitt Cancer Center listservs resulted in a large proportion of respondents conducting laboratory research in cancer biology (60.6%), which is the authors' primary interest. The remaining respondents were involved in other fields of biology (e.g., metabolism, microbiology, infectious diseases, neurology, immunology, botany; 29.8%) or physics, chemistry, or biotechnology (9.6%). The most commonly selected career goals were principal investigator in academia (39.4%), industry or private sector (11.8%), and undecided (40.9%).

Mentoring supervision

Previous surveys revealed that one potential solution to the data reproducibility problem was to provide better mentorship.

Table 1. Population characteristics

Characteristic	Respondents, %
Current position	
Graduate student	10.7
Postdoctoral fellow	89.3
Field of study	
Cancer biology	60.6
Biology (other)	10.5
Neuroscience	6.9
Microbiology/virology	6.2
Biotechnology	4.5
Immunology	2.6
Chemistry	2.5
Physics	2.6
Molecular biology/biochemistry	1.9
Plant biology	1.7
Career goal	
Principal investigator in academia	39.4
Undecided	40.9
Industry/private sector	11.8
Academia/government (other)	2.6
Writing/editing/publishing	1.4
Science policy/regulatory affairs	1.3
Other	2.6

Therefore, our survey included a series of questions on how respondents felt about mentor supervision and leadership provided in their laboratory. More than 70% of respondents said that they had meetings with their mentors at least weekly to discuss data, and 74% said that they had the opportunity to present their data outside of their work environment at least once per year (Fig. 1A and B). However, although those numbers are definitely encouraging, 28% of the trainees said that they met only once per month or less with their mentor, and 24% said that they presented their work outside of their laboratory only when they had enough data for a manuscript. Responses to the final question on mentorship methods showed that 220 (47.2%) of respondents had been pressured to produce "positive" data (Supplementary Data S1, question 8). Among the 220 respondents that answered "yes," 83.1% (39.2% of all respondents) said that this pressure came from either their principal investigator (171) and/or colleague/collaborator (54) and 123 from themselves or a mix of all three (Fig. 1C; detailed results in Supplementary Data S1, question 8a and additional comments; Supplementary Data S2), and some indicated via additional comments that threats (such as losing their position or their visa status) were sometimes used as a means of pressure (Supplementary Data S2).

Best research practices

Most respondents in previous surveys (ref. 5; go.nature.com/kbzs2b) indicated that poor research methodology is also a major reason for lack of reproducibility. Good mentoring also means teaching trainees good research practices. For example, it is generally accepted that good cell culture practices require regular testing to ensure that the cells are both uncontaminated and correctly identified (8–11). However, many respondents in our survey said that cell line authentication (38.5%) and mycoplasma contamination (29.4%) were never tested in their laboratory (Fig. 2A and B).

Similarly, strong and independent statistical analysis and/or blinded studies are known approaches to avoid detrimental bias in data analysis. However, almost 60% of respondents said that

blinded studies were rarely or never performed in their laboratory, and most respondents said that the issue was not even discussed in their research team (Fig. 2C; Supplementary Data S1, question 19).

Respondents of previous surveys also suggested that poor statistical knowledge plays a large part in poor reproducibility (ref. 5; go.nature.com/kbzs2b). However, 66.5% of our survey respondents said that their laboratories did not consult with a statistician regarding their experiments (Fig. 2D). Respondents reported that this was either because they believe they have a strong enough understanding of statistics to do their own analyses (16.7%) or because they believe their statistical software is adequate for their studies (49.8%; Fig. 2D; Supplementary Data S1, question 12).

In terms of repeating experiments, 85% of our respondents said that they always repeat either all experiments or at least the main ones in triplicate (Supplementary Data S1, question 17). Also, 47.2% said that they use at least one method, or even two or more methods (35.6%), to validate the antibodies they use in their experiments, and only 11.6% referred only to the manufacturer catalog (Supplementary Data S1, question 16).

Research integrity and transparency

Most respondents (56.8%) stated that they feel it is necessary to have a first-authored publication in the journals *Cell*, *Nature*, or *Science* when seeking an academic position (Fig. 3A); one quote paraphrases the comments of others: "There is no future in science without publishing (in) high impact factor journals" (Supplementary Data S1, question 27). We asked our respondents if the pressure of the "publish or perish" system influenced the way they report their data, and most (62.8%) responded that it did (Fig. 3B; Supplementary Data S1, question 27 and additional comments; Supplementary Data S2). Responses to subsequent questions about data falsification and fabrication suggest that this pressure may contribute to some investigators' rationale to do whatever it takes to publish their work in high-impact journals. Almost all respondents (99.6%) said that they had never falsified data (although 24.2% said they had omitted results that did not support their working hypothesis). However, 27.1% said that they had witnessed someone fabricating or falsifying data to complete a project or manuscript (Fig. 3C and D).

The publication process

As noted above, the "publish or perish" paradigm was pervasive among our respondents and influenced the way they reported their data. Taking this parameter into account, we wondered how the publication process influences research integrity and transparency. As stated above, 24.2% of respondents said that they had never falsified data but had omitted results that did not support their working hypothesis. Likewise, a similar proportion of respondents (23%) felt that manipulating their data or withholding the disclosure of negative data was demanded, expected, or necessary to prove a hypothesis (Supplementary Data S1, question 9 and additional comments; Supplementary Data S2). Anticipating this response, we asked a series of questions about the manuscript revision process.

Most respondents estimated that the revision process in lower-impact factor journals (impact factor <10) could take up to 6 months, and the revision process in higher-impact factor journals (impact factor >20) could take up to a year or

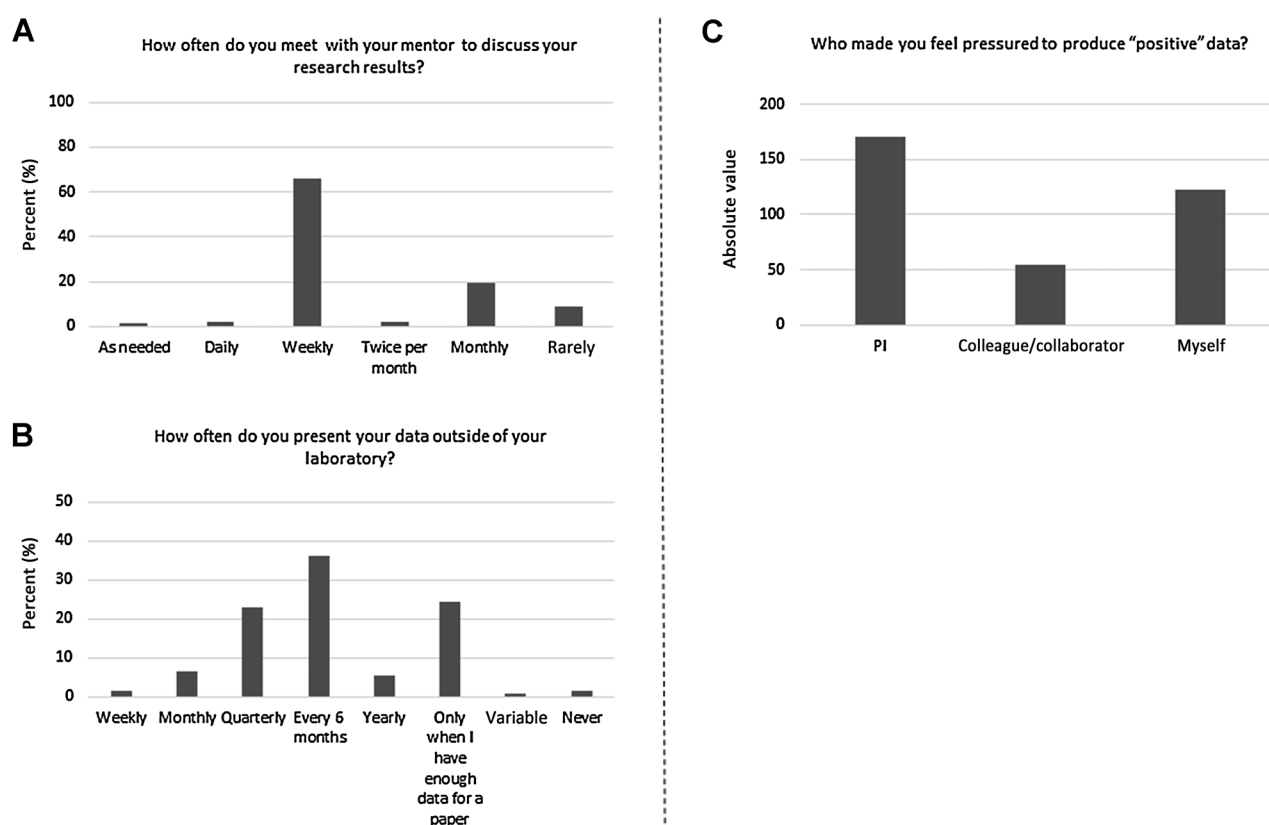


Figure 1.

Responses to questions about mentoring supervision from 467 respondents. **A**, Responses to question 6. Comments provided in response to "other, please explain" were either extrapolated to fit into one of the original choices or used to create the new categories of "twice per month," "daily," "as needed," and "rarely." **B**, Responses to question 7. Comments provided in response to "other, please explain" were either extrapolated to fit into one of the original choices or used to create the new categories of "monthly," "yearly," "variable," "never," and "not applicable." **C**, Responses to question 8a. The 220 respondents who said they felt pressured to provide "positive" data (question 8) indicated that the pressure came from a principal investigator or colleague/collaborator or was self-induced. Because several answers could be selected, the data are shown as absolute values of all responses selected (359 total).

even more (Fig. 4A and B). Out of the 248 respondents able to give an answer to this question, almost 90% estimated that the total cost of the revision, including salaries and supplies and services, could reach up to \$100,000 (Fig. 4C). These extra financial and time costs were estimated by a large majority to lead to only 0% to 25% improvement in support of the major findings reported in the original version of the manuscript (Fig. 4D).

Finally, to the question "When asked to revise a manuscript in the review process, have you ever been unable to provide the 'positive' data requested by the reviewer?" for which multiple answers were possible, 334 respondents gave 356 responses (we are assuming that the multiple responses represented different publications or revisions). Among these 356 responses, most indicated that the respondent had either never encountered this situation (154) or reported the negative results as well (122). However, a worrisome 80 responses (22.5%) indicated that when this issue occurred, the respondent either repeated the experiment until obtaining the outcome that would appease the reviewer or failed to report the negative data and made an "excuse" as to why the study was not conducted (Supplementary Data S1, question 24). Some responses even indicated that either the principal investigator or the journal itself sometimes requested elimination

of unsupportive data or negative results to meet word limits on the manuscript (Supplementary Data S1, question 24 and additional comments; Supplementary Data S2).

Discussion

In 2016, in a report in *Nature*, Baker suggested that data reproducibility in science could be improved by more robust experimental designs, better statistics, and, most importantly, better mentorship, with a positive note that one third of the respondents had taken concrete steps to improve reproducibility through one of these factors within the previous 5 years (5). We asked graduate students and postdoctoral fellows their opinions about data reproducibility. Our data indicate that graduate students and postdoctoral fellows believe that the pressure to publish and the publication process greatly influence selective reporting and transparency, potentially leading to data reproducibility problems.

Responses to questions related to mentoring, methodology, and good research practice teaching results were largely positive: guidance, meetings with mentors, presentations, replication of experimental results, and blinded studies were all noted to be important. However, some data collected during this set

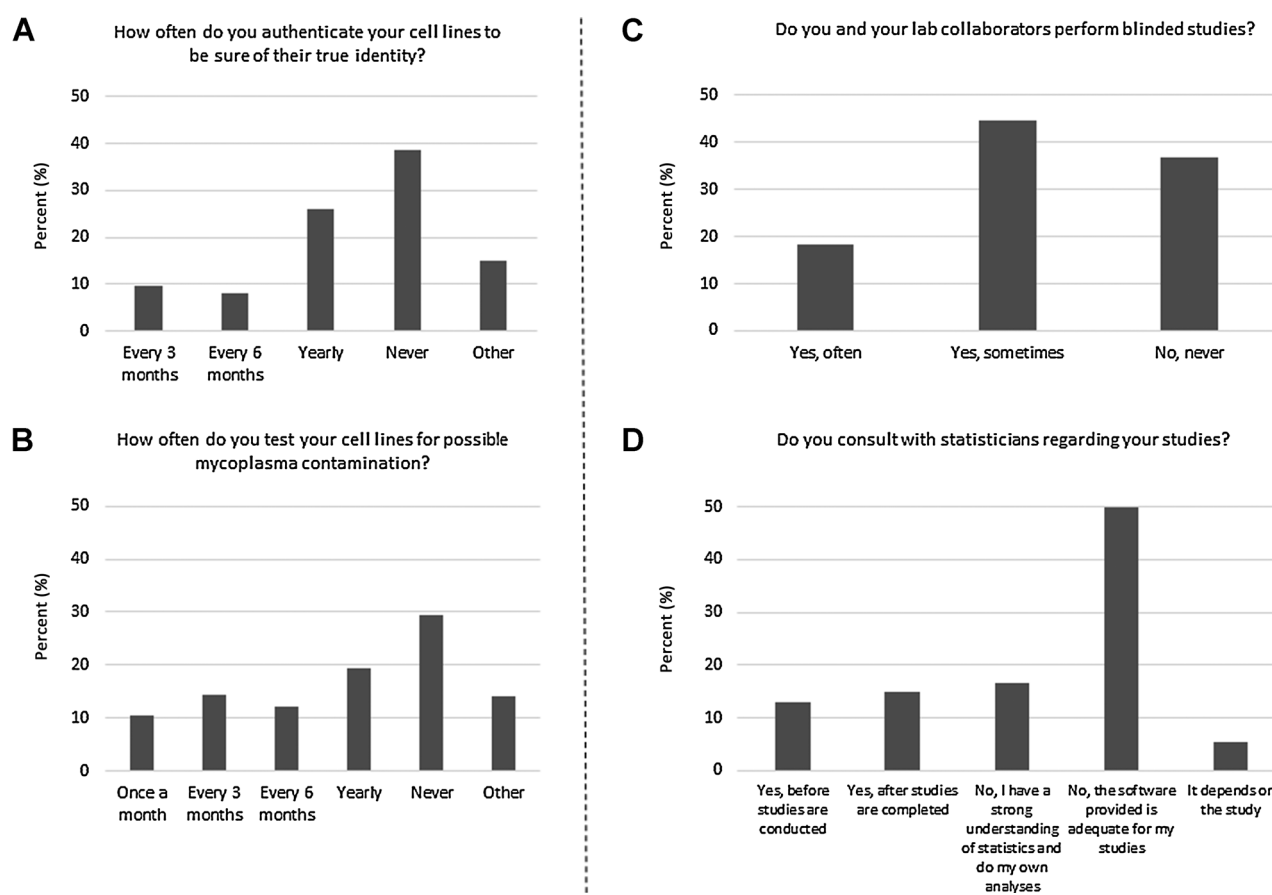


Figure 2.

Responses to questions about best research practices. **A**, Responses to question 13. Among the 465 respondents, 112 responded "other" and explained it as "not applicable." This graph represents the responses of the 353 other respondents. **B**, Responses to question 14. Among the 465 respondents, 121 responded "other" and explained it as "not applicable." This graph represents the responses of the 344 other respondents. **C**, Responses to question 18. Among the 467 respondents, 14 responded "other" and explained it as "not applicable." This graph represents the responses of the 453 other respondents. **D**, Responses to question 12. When the answer given was "other, please explain," the comments provided by these respondents were either extrapolated to fit into one of the original categories or used to create the new category "It depends on the study." All 467 respondents answered this question, but 15 of the "other" responses did not fit into any of the existing categories or the new category. This graph represents the responses of 452 other respondents.

of questions could explain at least, in part, a lack of reproducibility. For example, about a quarter of respondents reported that they met only once per month or less with their mentor and that they presented their work outside of their laboratory only when they had enough data for a manuscript. These findings indicate that roughly a quarter of respondents might not be optimally supervised for the duration of their project, underscoring a need for improvement in mentoring methods. It is of paramount importance for trainees to be exposed to an interactive scientific environment, such as laboratory meetings and seminars, and to be able to discuss findings and challenges in research outside of their daily work environment.

Also, about one third of our respondents also admitted to never authenticating their cell lines or testing for potential mycoplasma contamination. As suggested by several studies, these types of tests ensure valid and reproducible experimental results by reducing the use of misidentified or cross-contaminated cell lines that could lead to problems with reproducibility. Indeed, in 2007, Hughes and colleagues estimated that the

use of contaminated or misidentified cancer cell lines ranged from 18% to 36% (9). A more recent publication estimated a 20% cross-contamination rate (10, 11). Cell line identification ensures valid and reproducible experimental results by reducing the use of misidentified or cross-contaminated cell lines that could result in questionable or even invalid or uninterpretable data (12, 13). Therefore, our data suggest that about one third of graduate students and postdoctoral fellows could receive better training in research practices, and correcting these practices could help reduce questionable data and misinterpretation of results.

One finding from our survey stands out as deeply disturbing and worrisome: almost half of our respondents (47.1%) said they had been pressured to produce "positive" data, and 83.1% of these respondents said that this pressure came from a mentor or a colleague. In addition, threats such as losing their position or their visa status were reported. Beyond the obvious inappropriateness of such pressure, it seems evident that these types of pressure and threats could lead trainees to select only the data supporting their or their mentor's hypothesis; in the worst case scenario, this could

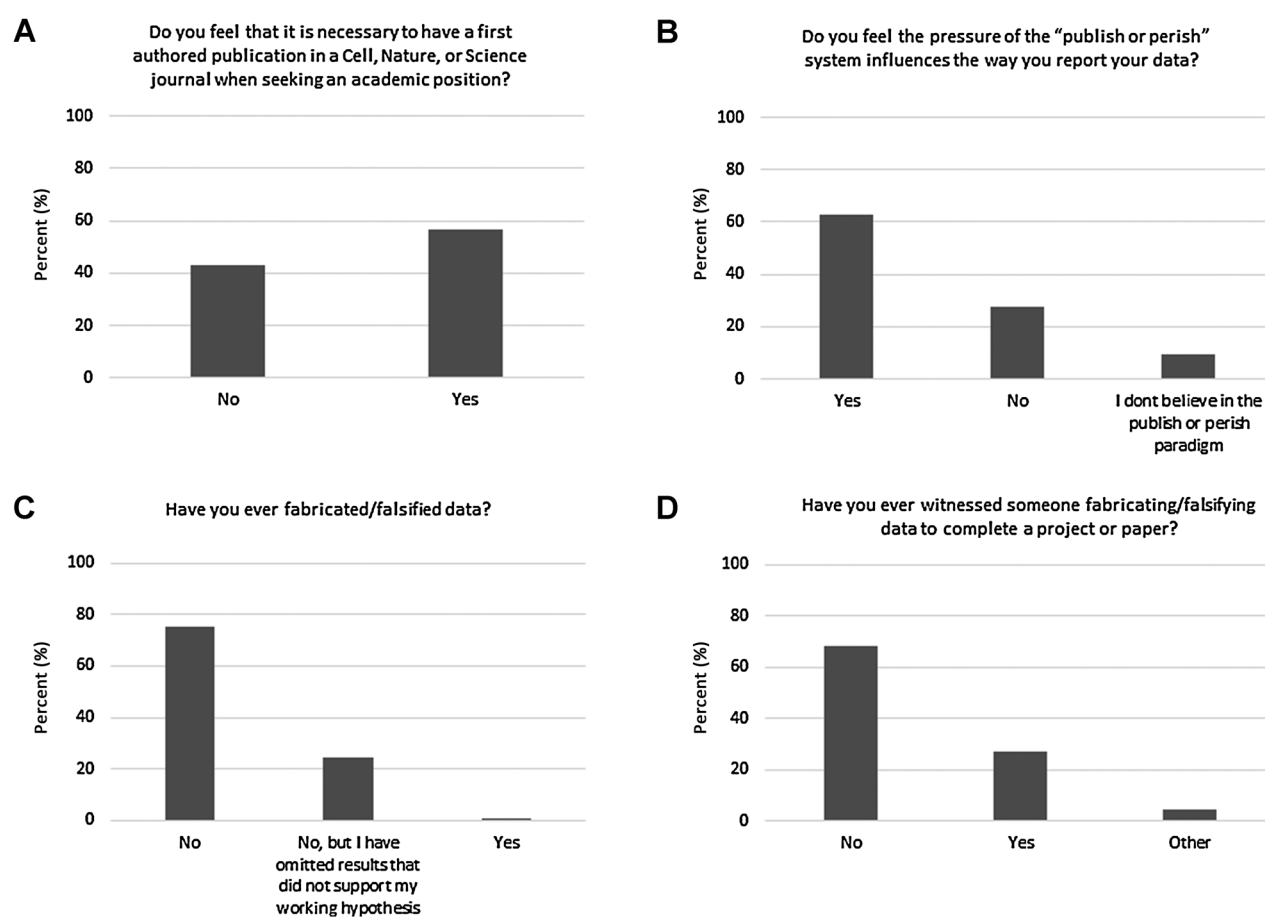


Figure 3. Responses to questions about research integrity and transparency. Responses were provided by all 467 respondents to questions 5 (A), 27 (B), 10 (C), and 11 (D).

lead trainees to falsify data. Although 99.6% of respondents said that they had never falsified data, 24.2% admitted to practicing selective reporting and 27.1% stated that they had witnessed someone fabricating or falsifying data to complete a project or manuscript. Of note, we are aware that such a discrepancy between admitting falsification and witnessing it raises an important question about the honesty of the respondents in answering queries this section; this issue is one of the limitations of this study. The 27.1% of respondents who witnessed data fabrication or falsification is likely much closer to reality.

Although our survey was completely anonymous, it is reasonable to assume that some respondents were uncomfortable responding honestly to some questions, in particular the questions discussed above on fabrication and falsification. This is a limitation of our study, but one that we cannot control as the survey as anonymous and follow-up is not possible. It is possible that only honest participant and/or participants genuinely concerned with this issue were willing to take this survey, limiting the input from people that may not practice the highest ethical scientific principles. Alternatively, there may be people that may not see this issue as important enough to spend time addressing it in a survey. It is obvious that because the survey was not mandatory, the population that participated was solely selected by their willingness and/or interest to take the survey.

Subsequently, this self-selection process may have introduced another limitation to our study: the fact that most of the queries are asking for opinions or observations, not data or facts, allowing subjective answers to queries. Finally, it is impossible for us to determine how many people actually received/read our e-mails (even if they actually had been sent by each institution) or our links in Twitter and LinkedIn. Therefore, we are not able to determine and report on a reliable denominator, making it impossible to report on the percentage of actual respondents/participants who chose to complete the survey. Despite these limitations, we hope that the data provided encourage leaders in the scientific community to do more to emphasize the importance of ethical rigor among trainees.

According to almost two thirds our respondents, and in accordance with Baker's *Nature* survey results, the "publish or perish" paradigm is alive and well and does influence the way data are reported. This may lead to the temptation to provide the most appealing or sensational data, which in turn may entice some investigators to compromise their research ethics. In response to our questions about the review process itself, a large majority of respondents said that the investment in both time and finances required to revise a manuscript is substantial (up to \$100,000) and often results in only marginal increases in quality, and some respondents' comments suggested that reviewers have unrealistic

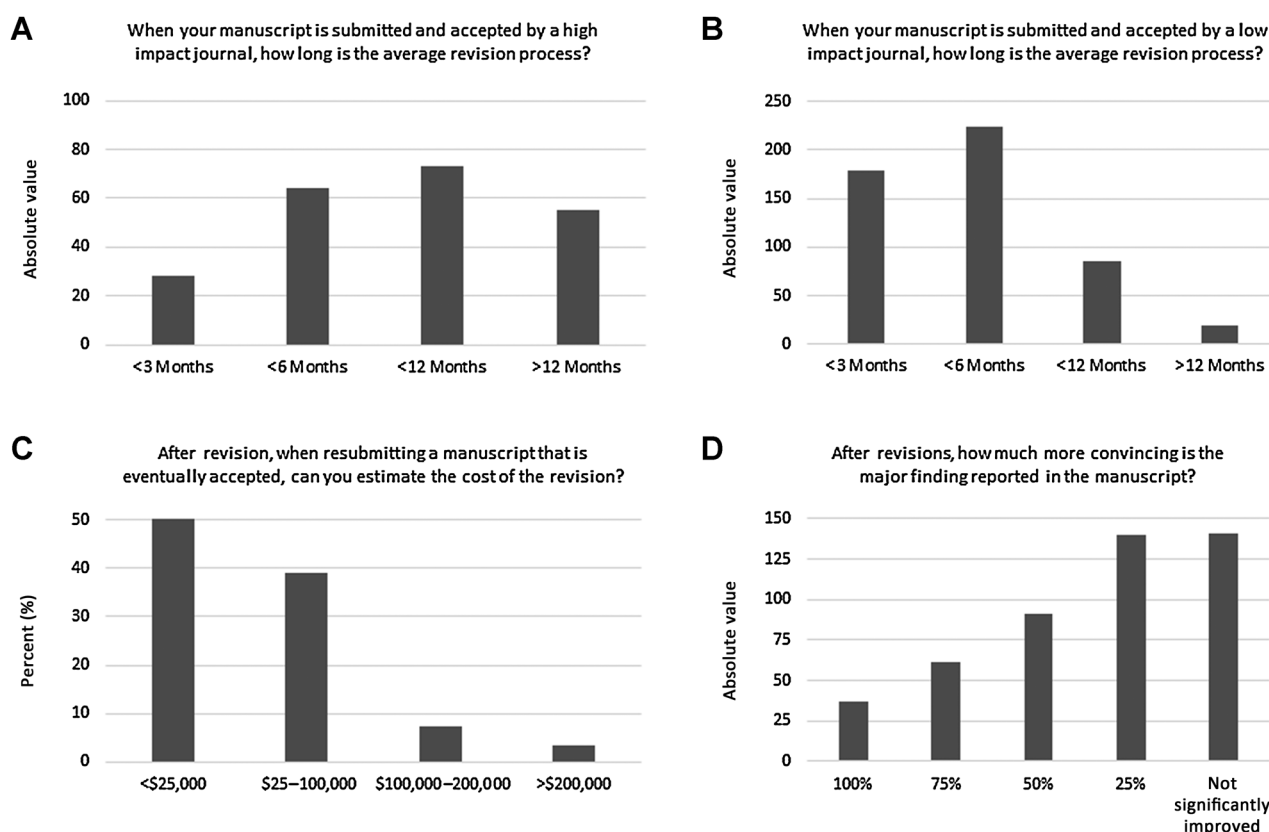


Figure 4.

Responses to questions about the publication process. **A**, Responses to question 22. Among the 467 respondents, 285 responded "other" and explained it as "not applicable." This graph represents the responses of the 182 other respondents. Because several answers could be selected, the data are shown as absolute values of all responses selected (220 total). **B**, Responses to question 23. Among the 467 respondents, 43 responded "other" and explained it as "not applicable." This graph represents the responses of the 424 other respondents. Because several answers could be selected, the data are shown as absolute values of all responses selected (506 total). **C**, Responses to question 26. Among the 454 respondents, 206 responded either "not applicable" or "don't have access to information necessary to provide an estimate." This graph represents the responses of the 248 other respondents. **D**, Responses to question 25. Among the 467 respondents, 53 responded "other" and explained it as "not applicable." This graph represents the responses of the 414 other respondents. Because several answers could be selected, the data are shown as absolute values of all responses selected (470 total).

expectations. Although peer review is an undeniably necessary process for publishing reliable science, its limitations have often been discussed in the past (14–17). In our study, when respondents were asked if they were reporting negative results in response to reviewers' requests, 22.5% said that when this issue occurred, they either repeated the experiment until obtaining the outcome that would appease the reviewer or failed to report the negative data and made an "excuse" as to why they did not conduct the study (Supplementary Data S1, question 24). Although these were not the majority of responses, they still represent a substantial proportion of the respondents and likely reflect part of the data reproducibility problem.

At the end of the survey, we gave the respondents the opportunity to share thoughts and experiences that they believed could be relevant to this issue (question 28). About 20% of respondents decided to share their personal view in this "free comments" section. Of note, none of the additional comments provided in the open-ended comment section at the end of our survey reflected *positive opinions* about the current state of scientific research. This is not surprising; we assume that only frustrated investigators felt the need to express their views in a more

comprehensive way. We recognize this point is a limitation to this study.

Respondents mentioned in their comments that competition for funding (highlighting the importance of independent and well-funded public/academic research), career advancement, and "landing the first job," and even pressure from mentors and colleagues, result in a lack of transparency that could explain, at least in part, the lack of data reproducibility observed over the past few decades. As one respondent noted, "This creates an environment where moral principles and scientific rigor are often sacrificed in the pursuit of a 'perfect' story based on spectacular theories. [...] Unfortunately, until the system changes, and no longer equates success with publication in high impact journals, matters will only deteriorate further. I have honestly lost faith in academic science, and believe it is a waste of tax payer and philanthropic money in its current state."

Several respondents also declared that honesty and complete transparency in data reporting are doomed to be affected by the current setup of academic research, because scientific honesty is not rewarded. These respondents voiced concern about the current value system used to judge scientific worthiness (e.g., publication record, top-tier journals, H-index) and blame this value

system for the lack of reward for ethical practice and bias against negative results. The balance between productivity and scientific contribution was presented as a critical element in scientific reporting, and some respondents said that this is why they did not intend to pursue a career in academia. This particular trend was interestingly addressed in a recent study published by Roach and Sauermann showing that the 80% of early in graduate school students seeking an academic career dramatically dropped to 55% by the time they obtain their PhD and started a postdoctoral fellowship (18). Several suggested that the number of published studies and impact factor of the journals in which the studies are published should be de-emphasized and the reporting of negative data encouraged, as a way to help fix the data reproducibility problem. Respondents indicated that a lack of published studies and/or studies published in high-impact journals seems to project a reputation of being unproductive, which directly limits funding and job opportunities. One comment underlined this issue: "We have seen plenty of 'one hit wonder', the quality of one's work, in my opinion, should not be evaluate based on the impact factor of a given publication or manuscript. The merit of science, the depth of knowledge, and the value of the significance of ones hypothesis and data should be the key evaluation for future faculty hiring in a long run." Within the past few years, in an effort to reduce the bias against negative results, several journals have started publishing "negative results" sections that present inconclusive or null findings or demonstrate failed replications of other published work. However, as several respondents in our study pointed out, even with the existence of these journals, negative findings remain a low priority for publication, and changes are needed to make publishing negative findings more attractive (19, 20).

So what now? Several studies have published potential solutions to the issue (and to some, crisis) of data reproducibility. Such recommendations include reducing the importance of the "impact factor mania" or choosing a set of diverse criteria to recognize the value of one's contributions that are independent of the number of publications or where the manuscripts are published (21, 22). This was part of the rationale to change the NIH biosketch format as explained by Francis S. Collins and Lawrence A. Tabak in their *Nature* comment from January 2014: "Perhaps the most vexed issue is the academic incentive system. It currently over-emphasizes publishing in high-profile journals. No doubt worsened by current budgetary woes, this encourages rapid submission of research findings to the detriment of careful replication. To address this, the NIH is contemplating modifying the format of its 'biographical sketch' form, which grant applicants are required to complete, to emphasize the significance of advances resulting from work in which the applicant participated, and to delineate the part played by the applicant. Other organizations such as the Howard Hughes Medical Institute have used this format and found it more revealing of actual contributions to science than the traditional list of unannotated publications" (23). Scientists tend to view or define "important contributions to science" as a new insight or discovery forms the foundation of further advances in a field. However, it is important to point out that negative studies may also have a major impact on a scientific field, allowing investigators to seek why a finding was negative and/or direct investigations in another direction. Also, in addition to the numerous comments about the need of finding a better way to publish negative results, several respondents suggested other potential solutions to the data reproducibility problem. For example, one respondent pointed out that

investigators can only vouch for the data they generate, and not all of the data in a manuscript. This is especially relevant in an era in which studies may be multidisciplinary, using, for example, a combination of genomics, bioinformatics, complex statistics, high-throughput studies, and other methods (Supplementary Data S2, comment 22). Another respondent who had had the opportunity to work in both academia and industry stated that industry had much more oversight, control, and care for the quality of data. Other interesting comments underlined the fact that postdoctoral fellow salaries and graduate student stipends have only slightly increased in recent years, making rapid career advancement desirable (and therefore increasing the pressure for good publications). Finally, another respondent commented that, because of the extreme competition for positions in both academia and industry, a reduction in the number of PhD trainees (if substantial job growth does not occur in the scientific sector) could have a positive impact on the publication pressure and hence lead to better data reporting. Of course, this could negatively affect the work force in laboratories and have broader negative effects on the entire research field. This question was also addressed in a recently published study by Xu and colleagues who pointed out that although postdoctoral fellows were highly skilled scientists classically trained to pursue academic tenure-track positions, the number of these academic opportunities remains largely unchanged over the years, whereas the number of postdoctoral fellows keeps increasing, which results in the fact that the majority of these young scientists will need to look at other types of careers (24).

In conclusion, our survey results suggest that postdoctoral fellows and graduate students believe that publishing in top-tier journals is a requirement for success or even survival in academia; this pertains to grant applications, job opportunities, and career advancement/promotions. These findings indicate that although current mentoring and research practices may be suboptimal, these factors represent only one component of a bigger issue. Blaming the reproducibility problem solely on poor mentoring and research practices would allow our research and scientific community to avoid having a serious discussion about the pressure felt by many faculty and trainees; our rewarding and advancement system appears to force faculty and trainees to shift their ethics on a continuum from honest reporting to selective reporting, data falsification, or even fabrication. Perhaps our findings are best summed up in the following quote from a respondent (Supplementary Data S2, comment 91): "There is an unavoidable direct link between scientific output and individual career which inherently compromises the conduct of research. It is ironic that we have to report 'conflicts of interest' when the most relevant conflict (that science is a major or the sole part of our livelihood) is considered not to be a conflict."

Disclosure of Potential Conflicts of Interest

L.M. Ellis is a consultant/advisory board member for Roche. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: D.R. Boulbes, T. Costello, L.M. Ellis
Development of methodology: D.R. Boulbes, T. Costello, R. Bhattacharya, L.M. Ellis
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): D.R. Boulbes, T. Costello

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.R. Boulbes, T. Costello, K. Baggerly, R. Wang
Writing, review, and/or revision of the manuscript: D.R. Boulbes, T. Costello, F. Fan, R. Wang, R. Bhattacharya, L.M. Ellis
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D.R. Boulbes, X. Ye
Study supervision: D.R. Boulbes, L.M. Ellis

Acknowledgments

The authors thank Erica Goodoff and Donald Norwood from the Department of Scientific Publications and Rita Hernandez from the Department of Surgical Oncology at The University of Texas MD Anderson Cancer Center for editorial assistance, as well as Amy Wilson at the National Postdoctoral Association for her assistance in distributing the survey to the National Post-

doctoral Association community. The authors also thank all the anonymous participants of this survey.

This work was supported, in part, by National Institutes of Health (NIH) Cancer Center support grant CA016672 and the William C. Liedtke, Jr., Chair in Cancer Research (L.M. Ellis).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 25, 2018; revised March 13, 2018; accepted April 5, 2018; published first April 11, 2018.

References

- Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012;483:531–3.
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011; 10:712.
- Baker M, Dolgin E. Cancer reproducibility project releases first results. *Nature* 2017;541:269–70.
- Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS One* 2013;8: e63221.
- Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016; 533:452–4.
- Woolston C. Graduate survey: uncertain futures. *Nature* 2015;526: 597–600.
- Many junior scientists need to take a hard look at their job prospects. *Nature* 2017;550:429.
- Freedman LP, Gibson MC, Wisman R, Ethier SP, Soule HR, Reid YA, et al. The culture of cell culture practices and authentication—Results from a 2015 Survey. *BioTechniques* 2015;59:189–92.
- Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *Biotechniques* 2007;43:575, 7–8, 81–2 *passim*.
- Neimark J. Line of attack. *Science* 2015;347:938–40.
- Horbach S, Halffman W. The ghosts of HeLa: how cell line misidentification contaminates the scientific literature. *PLoS One* 2017;12:e0186281.
- Chatterjee R. Cell biology. Cases of mistaken identity. *Science* 2007; 315:928–31.
- Drexler HG, Uphoff CC, Dirks WG, MacLeod RA. Mix-ups and mycoplasma: the enemies within. *Leukemia Res* 2002;26:329–33.
- Smith R. Peer review: a flawed process at the heart of science and journals. *J Royal Soc Med* 2006;99:178–82.
- Finn CE. The limits of peer review. *Education Week* 2002;21:30, 34.
- Park IU, Peacey MW, Munafo MR. Modelling the effects of subjective and objective decision making in scientific peer review. *Nature* 2014;506:93–6.
- Bhattacharya R, Ellis LM. It is time to re-evaluate the peer review process for preclinical research. *Bioessays* 2018;40.
- Roach M, Saueremann H. The declining interest in an academic career. *PLoS One* 2017;12:e0184130.
- O'Hara B. Negative results are published. *Nature* 2011;471:448–9.
- Matosin N, Frank E, Engel M, Lum JS, Newell KA. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Dis Mod Mech* 2014;7:171–3.
- Casadevall A, Fang FC. Causes for the persistence of impact factor mania. *mBio* 2014;5:e00064–14.
- Ellis LM. The erosion of research integrity: the need for culture change. *Lancet Oncol* 2015;16:752–4.
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612–3.
- Xu H, Gilliam RST, Peddada SD, Buchold GM, Collins TRL. Visualizing detailed postdoctoral employment trends using a new career outcome taxonomy. *Nat Biotechnol* 2018.