JCI The Journal of Clinical Investigation

Figure errors, sloppy science, and fraud: keeping eyes on your data

Corinne L. Williams, Arturo Casadevall, Sarah Jackson

J Clin Invest. 2019;129(5):1805-1807. https://doi.org/10.1172/JCI128380.

Editorial

Recent reports suggest that there has been an increase in the number of retractions and corrections of published articles due to post-publication detection of problematic data. Moreover, fraudulent data and sloppy science have long-term effects on the scientific literature and subsequent projects based on false and unreproducible claims. At the *JCI*, we have introduced several data screening checks for manuscripts prior to acceptance in an attempt to reduce the number of post-publication corrections and retractions, with the ultimate goal of increasing confidence in the papers we publish.

Figure errors, sloppy science, and fraud: keeping eyes on your data

Recent reports suggest that there has been an increase in the number of retractions and corrections of published articles due to post-publication detection of problematic data. Moreover, fraudulent data and sloppy science have long-term effects on the scientific literature and subsequent projects based on false and unreproducible claims. At the *JCI*, we have introduced several data screening checks for manuscripts prior to acceptance in an attempt to reduce the number of post-publication corrections and retractions, with the ultimate goal of increasing confidence in the papers we publish.

Scrutiny of published scientific literature is increasing, with valid concerns regarding fraud, data integrity, authorship, and plagiarism. These issues are not new to the *JCI* (1–3) or the scientific publishing world at large, but as editors, we feel a responsibility to diligently monitor any data that may be questionable. Readers too are savvier to problematic data, and the introduction of PubPeer (pubpeer.com) in 2012 provided an open forum to raise concerns about data presentation in published papers. At the *JCI*, we have taken all concerns brought to our attention seriously, even those made anonymously, and have also implemented changes to increase the scrutiny of data in manuscripts prior to publication.

In 2012, the journal began requiring all authors to submit their uncropped, unedited blots for review, and we have been manually screening blots and their corresponding unedited versions for all our published papers. At that time, we did not have a formal screening process in place for other types of data; however, image duplications and/or anomalies were sometimes detected after acceptance and prior to publication by production staff or by science editors preparing write-ups for promotional purposes. Other instances were not noticed until after publication, thereby requiring some form of post-publication action, such as correction, expression of concern, or even retraction. As recently noted by Bik et al., formal screening of images of papers prior to publication dramatically reduces the requirement for post-publication correction, which requires additional staff time to handle (4). Moreover, it has been our experience that authors are much more responsive to requests for corrective action prior to formal acceptance. We began formally screening the images in the main text and supplement of all papers on a path toward acceptance in the fall of 2016 and began to include a basic evaluation of statistics around this time. Although initially we were not tracking the number of papers with detected anomalies, based on our anecdotal data, we felt that we were seeing issues and rates similar to those observed by Bik and colleagues (4, 5). Following suggestions that journals need to take an active role in improving the scientific literature by publishing the results of their efforts at the editorial level (6), we began formally tracking the results of our prepublication screens.

Images and blots and stats, oh my!

Between July 1, 2018 and February 5, 2019, we screened 200 papers that were on a clear path toward acceptance for publication. These papers are typically revised manuscripts that have only minor points left to address prior to acceptance (i.e., no additional experiments required). We screen three categories, each of which is handled by a different science editor. All Western blots in the main and supplemental figures are compared to the uncropped, unedited blots provided by the authors. For blots, we are looking to make sure that there are not excessive and/or nonlinear adjustments to contrast or unindicated splicing of images, and to determine if images within a figure panel are derived from the same blot or run on separate gels. A basic statistical review is performed to make sure the tests used are largely appropriate and account for multiple comparisons and repeated measures. Finally, images, including histology panels, flow cytometry plots, and graphs, are visually compared to look for duplications, manipulations, and/or reuse.

During the course of our study, 28.5% (57 of 200) of the papers screened were flagged for issues with statistical tests, 21% (42 of 200) of papers had some issue with the blots, and 27.5% (55 of 200) of papers had issues with images. For screening of statistical analysis issues, the most common concern was a lack of accounting for multiple comparisons in the chosen analysis: for example, the use of multiple paired *t* tests for experiments that were run contemporaneously with a shared control. Western blots were typically flagged for lack of the corresponding raw images and use of a loading control that was not derived from the same gel as other samples in the figure panel. It may seem like common sense, but we ask that authors provide a loading control for all gels run, and in cases where it makes sense to run parallel gels contemporaneously, we ask that the experimental design is transparently acknowledged. Papers flagged for statistics and blot issues during our study period did not contain major errors, and authors were allowed to revisit their statistical analysis or address any blot discrepancies in their next submission. In all cases, the authors successfully responded to concerns in the revised version. It is worth mentioning that outside of the study period, we have had papers that were rejected for inappro-

Reference information: *J Clin Invest*. 2019;129(5):1805–1807. https://doi.org/10.1172/JCI128380.

priate manipulation of blot images and instances in which authors were asked to remove blot images that were "enhanced" in a manner that distorted the results or to replace figure panels with blots derived from a single experiment.

We detected issues with images in 27.5% (55 of 200) of papers screened. The journal is extremely fortunate that one of the editors (CLW) has an excellent eye for image duplications. Of the papers with image issues, 89.1% (49 of 55 papers) had what we consider to be minor transgressions, such as undisclosed reuse of representative images in more than one figure or unaltered duplications of a panel that appeared to be inadvertent (i.e., a copy-and-paste error). In these cases, authors were alerted to our concerns and asked to confirm figure assembly and/or asked to disclose reuse of representative images in the legend of the second occurrence. While these mistakes are relatively minor, had they not been detected prior to publication, several would likely have required a formal correction.

We found what we consider moderate issues in 7.3% (4 of 55) of papers with image concerns. This group includes papers in which multiple panels were reshown that may or may not represent the same treatment/condition or in which a different crop of the same image was used to represent distinct treatment groups. In cases such as these, the authors were asked to clarify how figures were prepared and to explain how many times experiments were performed. A subset of the Editorial Board, including the Editor in Chief, Deputy Editors, and the handling Editor reviews the authors' response and determines the next course of action. In these cases, authors were able to quickly respond to the concerns raised, provide feasible explanations as to how the errors occurred, and easily provide original data that were correct. In one instance, we requested institutional oversight to verify the integrity of the data, and the authors were allowed to submit corrected figures. These issues almost certainly would have required post-publication correction. Moreover, any extended length of time between publication and detection of errors can make attempts to find and interpret the original data files and experimental records challenging for the corresponding author, especially in cases where the person that generated the data in question is no longer in the lab.

Lastly, manuscripts with major issues accounted for 3.6% (2 of 55) of the papers flagged and 1% (2 of 200) of the total papers screened during our tracking period. These issues included multiple instances of the same image being differentially cropped and used to represent different conditions and treatments as well as images that appeared to have been altered and possibly fabricated. The authors of the papers in question were asked to provide explanations for the noted discrepancies, and in both cases, the Editorial Board members determined that the explanations provided were not sufficient, resulting in loss of confidence in the data and ultimately rejection of the manuscript. In cases where data appear to be fabricated, we attempt to inform the Office of Research Integrity or equivalent at the corresponding author's institute. Our experience (both within and outside of the tracking period) is that institutions vary widely in their response to data integrity concerns, and there is unfortunately reluctance to deal with identified problems that have not been published.

Of note, there did not appear to be a correlation between issues in one area and issues in another area. Of the 200 papers screened, only 4 (2%) were flagged for concerns with statistics, blots, and images, 12 papers (6%) had problems with both blots and images, and 14 papers (7%) were marked with concerns about statistics and blots.

Take-home lessons

It should be noted that the vast majority of issues detected by our team during this study were not caught by our excellent reviewers. We have found that detection of data anomalies, especially image duplication and manipulation, requires a dedicated person with a keen eye and excellent pattern-recognition skills. We suspect that most scientists would be surprised to learn that there is not automated software to screen all images in a manuscript for reuse, partial overlap, and rotation. While we have tools to directly compare two images, it is computationally much more complicated to compare every possible rotation of images without a fixed point of orientation. There is interest in developing such methods, and Acuna et al. have recently reported the development of a pipeline with potential for large-scale detection of inappropriate image reuse (7). Currently, our screening is limited to images and blots within a single manuscript, and we do not have the capacity to screen for use of images in other published papers. The tool described by Acuna and colleagues could be an invaluable aid for our screening processes, and we look forward to further development and reporting on their pipeline.

To our authors, we strongly encourage you and your trainees to keep methodical records of experiments and the data collected for a given paper. It is to your benefit to be able to easily locate the original experiments, thereby allowing easy determination of how errors may have been introduced. When you submit raw blot images, please label each blot provided with the corresponding image and the sample being run. Disclose when samples, especially "loading controls," are run on parallel gels. We will have more confidence in your data if you can present independent representative images in different figures, and you must clearly indicate in the legend of the second instance that an image from a previous figure is being reshown for comparison purposes. Be transparent when the same data set is presented in more than one figure panel, and make sure that your statistical analysis appropriately accounts for all the groups in your experiment, even if you show only a subset in the final figure. If you do not know much about statistical analysis, please don't try to guess — educate yourself and reach out to statisticians for a formal consultation to ensure that you are incorporating an appropriate design and analysis.

Transparency in data reporting is essential to maintain the integrity of the scientific literature. At the *JCI*, we have recently enacted a policy that dictates that all graphs of quantitative data be presented so that the distribution and variation of the data are clear and highly encourage that individual data points be shown (see https://www.jci.org/kiosks/authors#Figures for more information). This change in policy has also been key for us in detecting when a single control has been used for multiple experiments, thereby necessitating a different statistical approach and

further disclosure. We will continue to monitor our own practices to ensure we are holding research published in the *JCI* to the highest standards of integrity.

Is cheating in biomedical science on the rise? There is evidence that the number of articles retracted is increasing (8, 9), and the *JCI* retracted more papers in the period from 2014 to 2018 (15 retractions) compared with 2009 to 2013 (6 retractions). This rise in retractions across journals could be in part due to editors taking accusations of fraud more seriously as well as increased reporting of problems to journals, spurred by more public, online discussion of problematic papers. We would speculate that the widespread use of software such as Photoshop has made it easier for authors who are tempted to cheat to manipulate data. On the journal side, we are limited to catching obvious errors after they are committed. The scientific community as a whole needs to be steadfast in guarding against unreliable data at all stages of planning, acquiring, interpreting, and publishing data. One potential mechanism to reduce figure assembly errors is to have someone other than the primary author who generated the data put together the figures.

We recognize that our screening methods are not perfect and subject to human error. Thus far, none of the papers included in the tracking period have had issues brought to our attention after publication. Of note, we have issued corrections for two papers that were screened prior to publication but the presence of figure errors was missed. Nevertheless, our screening process has been critical for limiting the number of post-publication corrections, which surely benefits our authors and increases the confidence of our readers in the data published by the *JCI*. Additionally, during the tracking period, we prevented the publication of two articles that are likely to have been retracted had they been published and the figure problems detected in post-publication review. Despite the issues that were uncovered during our screening, we believe that the majority of our authors act in good faith and are not intentionally introducing errors to the scientific literature. As manuscripts increasingly become more complicated and involve more co-authors, the chance to introduce errors increases. Authors, reviewers, editors, and readers must all guard against honest mistakes, sloppy science, and fraud. The integrity of the scientific community is on the line.

Corinne L. Williams, Arturo Casadevall, and Sarah Jackson

1. Savla U. When did everyone become so naughty? *J Clin Invest*. 2004;113(8):1072.
2. Neill US. Stop misbehaving! *J Clin Invest*. 2006;116(7):1740–1741.
3. Neill US. All data are not created equal. *J Clin Invest*. 2009;119(3):424.
4. Bik EM, Fang FC, Kullas AL, Davis RJ, Casadevall A. Analysis and correction of inappropriate image duplication: the *Molecular and Cellular Biology* Experience. *Mol Cell Bio*. 2018;38(20):e00309-18.
5. Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. *mBio*. 2016;7(3):e00809-16.
6. Casadevall A, Fang FC. Making the scientific literature fail-safe. *J Clin Invest*. 2018;128(10):4243–4244.
7. Acuna DE, Brookes PS, Kording KP. Bioscience-scale automated detection of figure element reuse. bioRxiv website. https://www.biorxiv.org/content/10.1101/269415v3. Published February 23, 2018. Accessed February 26, 2019.
8. Fang FC, Casadevall A. Retracted science and the retraction index. *Infect Immun*. 2011;79(10):3855–3859.
9. McCook A. Retractions rise to nearly 700 in fiscal year 2015 (and psst, this is our 3,000th post). Retraction Watch website. https://retractionwatch.com/2016/03/24/retractions-rise-to-nearly-700-in-fiscal-year-2015-and-psst-this-is-our-3000th-post/. Published March 24, 2016. Accessed February 26, 2019.