Data Provenance, Data Integrity, Scientific Rigor, and Science Culture Action Plan
Department of Neurology, Duke University School of Medicine

The outside world and regulatory bodies are looking at how we acquire, maintain, validate, and report our research findings. There is an unfortunate brewing culture of public mistrust of biomedical science brought about by the combination of a few instances of serious, highly publicized scientific misconduct, and a widespread perception that many laboratory findings are difficult to replicate due to bias, poor design, lack of appropriate blinding or worse. These issues have been highly publicized, and we are committed to being a part of the solution to these issues.

Data provenance and integrity as well as explicit rules for data acquisition ensure that the knowledge we report is supported by the primary data and technical approaches, and the primary data are retained in a form that allows us to be certain of the veracity of our knowledge. Scientific rigor ensures the proper application of the scientific method using the highest standards and appropriate statistical approaches in the field. Scientific rigor is essential to conduct of the scientific enterprise.

There are multiple reasons to take codify our approach to these issues. First, science is publicly funded and its credibility in the public's eyes is vital to continuation and expansion of funding. We must do all that we can to eliminate misconduct and minimize bias. Second, research builds on previous reports in the pursuit of accurate knowledge. We must be confident of the truth in prior publications to further the scientific enterprise. Third, proper research standards avoid the waste time or money following up inaccurate or erroneous reports. We must keep the scientific enterprise moving forward. Finally, science drives translational and clinical research which must be built on a solid foundation.

As a Department, we commit to following six general principles:

- Ensure that every member of the lab maintain and update (daily) a lab notebook
- Know where your data are.
- Know what has been done to acquire and modify your data.
- Make all efforts to ensure that data collection and analysis are, at a minimum, unbiased and blinded when possible.
- Follow proper statistical procedures.
- Empower Departmental staff to understand these principles and monitor their implementation.

Overview of Best Practices

1. Insofar as possible given the nature of the research, best practices in scientific rigor, including statistics, should be followed.

2. In recognition that no one size fits all, each laboratory should establish its own specific plan for scientific accountability and scientific rigor, per established standards of its field, integrating industry or other perspectives when appropriate.

3. Record keeping should track all primary data and should provide a way to "audit" the data for each figure of each paper expeditiously. In general, this means the data should be organized and indexed in adherence to the FAIR principle: data should be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable.

4. Daily work logs should be expected of all lab members.

5. All modifications of raw data should be performed on copies of the original data, if possible, and should be tracked, dated, and documented fully. A copy of the original unmodified data should be expeditiously (prior to any analysis) stored in a central repository under the control of the PI.

6. The laboratory head should avoid allowing his/her expectations about the nature of the results affect the attitudes, or behavior of the laboratory staff.

7. Scientific accountability and scientific rigor should be a frequent discussion between the laboratory head and the laboratory staff, to establish a sense of common purpose and a shared goal to discover the truth.

8. There should be no impediment to reporting scientific behavior outside the norms, which may be done via the Anonymous Accolade/Unwanted Behavior link, or any of the physical drop boxes that are located outside office 227D in the Bryan Research Building, outside office 238 in the Neurology Clinic on Morreene Road, and outside lab 5200 in MRSBIII.

Best practices in experimental design and data collection

- Employ systematic random sampling for data collection, including selection of subjects, brain area, cells (collected via, e.g., FAX sorting, or sampled in microscopic fields) or cell parts.
- Strive to eliminate bias in experimental procedures and analysis. If practical, experimenters should be blinded to treatment. The timing of experiments should be balanced to account for sources of bias over time (e.g., evolution of surgical skills, fatigue, change in personnel; test-order effects, circadian rhythms in experimental animals).
- Use positive and negative controls.
- Use replicate samples, including both technical and biologic replicates, for experimental groups, when appropriate.

- Use validated and/or well-characterized reagents (such as antibodies and pharmacological agents). If validation is not available, conduct full validation.
- Consider limitations of behavioral, animal, or cellular models including possible contributions of genetic background and gender.
- Find a proper balance between increasing numbers of animals for replication and the goals of "replacement, reduction, and refinement" in animal research.
- Obtain and study the raw data for any results provided by shared research cores.
- Use best practices and reporting standards for collecting, analyzing and stating '–omics' and epigenetics data.

Best practices in data analysis and statistics

- Consult with a bio-statistician both before and after data collection, if statistical analysis is needed. "Stats shopping" (finding the one test that shows significance) is unacceptable.
- Primary expected outcomes should be noted prior to experimentation and analysis.
- Determine sample size by pre-experiment power analyses, when possible. Identify stopping points *a priori* to avoid testing to a foregone conclusion.
- Conduct a thorough characterization of experimental effects.
- Repeat key experiments within the laboratory to reduce likelihood of statistical flukes or biased results.
- Use care in pooling of data across experiments done at different times, multiple time points, or different experimental groups. Validate and fully describe normalization practices when pooling is necessary.
- Avoid data exclusion except for predetermined criteria. If it is necessary, define and report objective procedures for dealing with attrition or other missing data and data exclusion. Unless there is a compelling, transparent reason to exclude data, include all runs of each experimental procedure. This applies to exclusion of individual points or complete data sets. While technical issues often arise, days of "no results" in a laboratory notebook should be explained.
- Perform theoretically correct analysis of data using appropriate statistics and sample sizes.
- Do not use statistics to draw misleading or erroneous conclusions. Take advantage of resources that provide professional statistical expertise (e.g., the Biostatistics consultation service).
    - Perform statistical tests to validate what is seen in the data, rather than to reveal effects that may be statistically significant but functionally non-significant.
    - Select appropriate statistical tests, including testing of statistical assumptions, such as adherence of data to a normal distribution.
    - Control for multiple comparisons when appropriate.

- Avoid "significance chasing" such as interpreting the data in different ways so that it passes the statistical test of significance, or analyzing different measures until finding one on which groups differ.

Best practices in data management

- It should be emphasized to all lab members that data is the property of Duke University.
- The complete primary data should be retained for a minimum of 5 years after publication (https://provost.duke.edu/sites/all/files/FHB_App_P.pdf), backed up, and protected against alterations.
- Experimental records should explain: Why the experimental work was done and what project it was a part of; who designed the experiment and who performed it; what the person making the record did; when (month, date and year) the experiment/work was conducted; how the work and data collection were conducted (methodology); what materials were used, including reagent validation, the findings and interpretation, and next steps.
- The data should be recorded in ways that cannot be altered. Any alterations and modifications of the primary data for analysis, publication, etc., should be performed on copies of the data. Eventually, as data software becomes available, all data should be indexed, dated, and described.
- Lab notebooks should be decided by the PI, and may include bound volumes with page numbers, or LabArchives electronic lab books approved by Duke University. Data that cannot be accommodated in the bound format should be indexed in the lab notebook.
- Data notebooks should be open for viewing, and a brief description of days of "no data" should be included.
- Every figure of every paper should (in draft form) be cross-referenced with the location of the experiment and the original data that contributed to the figure.
- The level of information security should be appropriate for the material, especially for human subject protection and PHI.
- Data should be accessible readily to all data owners, and available to appropriate outside parties if needed, in accordance with the NIH data sharing policy https://grants.nih.gov/policy/sharing.htm (for updates related to genome-wide association studies, see also https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/).

Best practices in publication

- Report full details on methods and experimental design, including technical and biological replicates, methods for randomization and blinding, primary endpoints, and self-replication efforts, and reagent validation.

- Report complete results of all analyses done as part of an experiment, including statistical controls for multiple comparisons and identification of pre- and post-hoc analyses. Methods sections should be too long, rather than too short.
- Avoid "rushing" findings into publication without a full investigation and proper self-replication.
- Target appropriate venues for publication. Avoid pressure to publish in the most glamorous journal at the expense of following the best practices for experimental design, data analysis, statistics, and publication. If a paper requires a long methods section or many figures to document the science thoroughly, do not try to compress it into a short format, no matter how "important" the results seem. Strive to publish well-controlled negative, "uninteresting," or "not novel" results in appropriate venues.
- Resist the emerging trend where the peer review process demands additional experiments on an abbreviated timeline, with the associated pressure for results to be interpreted to conform to previously-reached conclusions.

<u>Creating a functional and proactive scientific culture</u>

- The Department's goal is to instill a culture of 'getting it right', with the expectation of open conversation and a lack of retribution for calling results or procedures into question either within the group, or to the lab head in confidence.
- All Department staff should know that they may bring any and all concerns to the attention of the Chair or the Department's Ombudsperson in confidence, without fear of retaliation or retribution, using the Ethics drop boxes or the Anonymous Accolade/Unwanted behavior link. Staff should also be aware of the Duke Integrity Line to report concerns anonymously.
- Laboratory heads must minimize incentives or pressures (or the appearance thereof) that drives their staff to perform for reasons other than pursuit of truth. It is critical to avoid the real danger that staff will respond to the laboratory head's concerns about academic promotions, choice of publication venue, or competition with other labs.
- Issues of proper scientific conduct and scientific rigor should be discussed regularly with laboratory personnel, in both private and group meetings.
- Laboratory heads should be involved in laboratory procedures, should oversee some of the actual experimental work, and should "know" how things are done in their laboratory.
- Meetings with staff should include inspection of the primary data and discussion of detailed analysis procedures, as well as discussion of final publication-style figures.
- Research team heads should compose Lab Compact and present the document to their teams in 2022. All Research team members i.e., head, staff, and trainees, should comply with the Lab Compact.

Concrete steps to be taken by the Department

1. We will all continue to talk about proper scientific conduct at all levels: faculty meetings, lab meetings, and courses.
2. We will take advantage of any institutional courses, offerings or best practices in these areas.
3. We will strive to create a culture of improvement in ethical research by having individual labs test different approaches and then adopting best practices broadly.
4. Each laboratory should develop a "Data Management Standard Operating Procedure (SOP)" that will provide specific guidelines for data acquisition, storage, and transparency. The SOP should cover 4 basic components of data management:
    a. How the data collected and stored;
    b. How notes taken and stored;
    c. How analysis is conducted, tracked and, if intermediate steps are saved, stored;
    d. How figures are made and linked to both the analysis steps and the original data.
5. The laboratory's Data Management SOP should be discussed with the Chair when it has been completed, and compliance measurement will be a topic in the annual 1-on-1 meeting.
6. All research staff in all laboratories must read the Department's Scientific Culture and Accuracy Plan (SCAP), and sign an affirmation that they have done so.
7. The Chair will serve as a "Data Integrity Liaison" to the School of Medicine. He will advise individuals or laboratories on all of the issues covered in this plan and will work with the school's designated official for scientific integrity.
8. There will be a semi-annual meeting of all lab PI's to discuss alterations in this document and individual lab plans.
9. Faculty who are expert in data analysis will be available to advise students, postdocs, and faculty on how to analyze their data.
10. The Department's System Administrator, will work with each laboratory to implement their chosen procedures for data storage, backup, and tracking.
11. An annual anonymous survey of department research staff and faculty will be conducted to assess the "integrity quotient" of the department.