Managing Data for Integrity: Policies and Procedures for Ensuring the Accuracy and Quality of the Data in the Laboratory

Chris B. Pascal*
Director, Office of Research Integrity, Department of Health and Human Services, USA

Keywords: research integrity; data management; training; mentor; principal investigator

ABSTRACT: *Management of the research data is an extremely important responsibility of the Principal Investigator (PI) and other members of the research team. Without accurate data, no worthwhile conclusions can be drawn from the research study. Integrity in data management is critical to the success of the research group and to public trust in the research outcomes. One of the primary responsibilities of the PI is to provide proper training to the junior members of the lab. This effort can be buttressed by institutional data policies that are implemented at the group level. Extensive and frequent guidance in good research practices by the PI and other senior research staff is critical to the proper training of new scientists.*

Many scientists think of data management as how you collect the data, record and maintain it. These are important but not sufficient to ensure the quality of the data for purposes of interpretation, reporting, and publication. This article is intended to explore data management in a more comprehensive way that takes into consideration the positive contributions that can be made by institutions and laboratories in adopting data policies, holding regular lab meetings to review and discuss the data, and providing formal training of young scientists in the principles and skills in interpreting and reporting the data.

***** The opinions expressed herein are those of the author and do not necessarily represent the views of the Office of Research Integrity, the U.S. Department of Health and Human Services, or any other federal agency.

This paper was presented at the 6th International Bioethics Conference on the subject of 'The Responsible Conduct of Basic and Clinical Research', held in Warsaw, Poland, 3-4 June 2005.

Address for correspondence: Chris B. Pascal, J.D., Director, Office of Research Integrity, Office of Public Health and Science, Department of Health and Human Services, 1100 Wootton Parkway, Suite 750, Tower Building, Rockville, MD. 20852, USA; email: cpascal@osophs.dhhs.gov.

1353-3452: 2006. Published by Opragen Publications, http://www.opragen.co.uk.

As many readers already know, the U.S. Office of Research Intergrity (ORI) has an abiding interest in providing education in the responsible conduct of research (RCR). ORI has provided financial resources for research institutions to develop educational products in various RCR areas such as managing conflicts of interest, peer review, authorship, and data management. Many of these products are web based and are posted on the ORI website.[1]

An important part of ORI's mission is focused on the integrity of the research process, such as the accuracy of the research data and research publications, and the prevention of, or reduction of, research misconduct. Because accurate scientific results are critical to the advancement of science, an argument can be made that the proper management of the data, and in particular, the quality and accuracy of the data, is the most important element in ensuring scientific integrity and public confidence in research results and findings.

Responsible Research at the Individual Level

In 2002, the Institute of Medicine,[2] issued a report on research integrity stating that "the responsible conduct of research is not distinct from research; on the contrary, competency in research *encompasses* the responsible conduct of that research and the capacity of ethical decision making." Another way to address this issue, is to say that one cannot separate responsible research from the competency or quality of that research. By the same token, lack of knowledge, skills, and experience in research is in most instances invariably fatal to research quality.

In discussing RCR at the individual level, the IOM report[2 (p.5)] lists the following qualities needed by investigators to conduct responsible research–

- intellectual honesty in proposing, performing, and reporting research
- accuracy in representing contributions
- fairness in peer review
- transparency in conflicts of interest
- protection of human subjects; humane care of animals and
- adherence to mutual responsibilities between investigators and research teams

Almost all of these principles and concepts are related to the quality of the research data in some way. Intellectual honesty in proposing, performing, and reporting research certainly encompasses accuracy in describing data collection efforts, reporting data from prior studies, and the inclusion of preliminary data in the application in order to demonstrate the feasibility of the research proposal.

Accuracy in representing contributions includes a fair and accurate description of the contributions of the lead author and making sure that other authors receive appropriate credit for their data or data interpretations included in the manuscript or proposal.

Fairness in peer review requires that the reviewer give appropriate acknowledgment to the author if the data appear sound, data interpretations are

appropriate, the methodology is appropriate to the type of research involved and the proposed article makes a contribution to the literature. If data criticisms are made, the reviewer must have adequate justification for those criticisms.

Conflicts of interest may affect the interpretation and reporting (or withholding of data). When conflicts are disclosed, the reader of a journal article or user of the data may take into account the potential bias of the investigator. However, this may be insufficient to ensure the integrity of the data. If the conflict is very impactful, such as a large stake in the research project, the conflict may lead to apparent or real bias. Although an outsider may be aware that bias could exist, he or she may have a difficult or impossible task in coming up with an accurate account of the level of the bias and how harmful it may be to the safety of the study and the quality of the data. In such circumstances, the institution or investigator should assume direct responsibility for ensuring the quality and reliability of the data. If that cannot be accomplished, the institution may need to completely bar the investigator from participation in the study.

Data validity and reliability is certainly important in both human research and animal research. If the plan for data collection, interpretation, and reporting is vague or piecemeal, it may not be possible for researchers to make sound conclusions about the data. Hence, poorly designed studies may pose unnecessary risks to human subjects or animals, suggesting that it may not be inappropriate to conduct the study at all. Certainly, if there is more than minimal risk involved, the investigator, institution, and IRB should seriously consider not going forward with the proposed research. If the study is worth doing, it is important to collect and use the data in a way that advances science and benefits the public. This means the data must be valid, reliable, and interpreted and reported in a reasonable fashion.

Adherence to mutual responsibilities of research teams certainly includes data responsibilities. The lead investigator has responsibility to establish the data collection procedures, make sure the research team understands its responsibilities, and provide training and supervision as needed. Likewise, the members of the research team need to be diligent in following the procedures, asking questions as needed, and letting the PI know if there are problems with the data.

Data issues are clearly important throughout the research process as indicated by how often they are linked to the RCR issues identified by the IOM report concerning responsibilities at the individual level of research.

Responsible Research at the Institutional Level

The IOM[2 (p.5)] recommends that the institution address RCR by the following–

- provide leadership in RCR
- encourage respect for everyone involved
- promote productive interactions between trainees and mentors
- advocate adherence to rules regarding the conduct of research
- conduct thorough inquiries and investigations into alleged misconduct
- offer educational opportunities in RCR

- monitor and evaluate the institutional environment supporting integrity and RCR and use this knowledge for continuous quality improvement
- ORI also recommends adoption of specific institutional policies and guidelines that support research integrity

These institutional RCR elements can impact data integrity in a positive way, especially through RCR education programs, particularly programs that focus on research data, regular mentoring of young scientists in the laboratory, and in particular, institutional efforts to adopt formal data policies or guidelines for the institution, that are then implemented at the laboratory or group level. Before we discuss these positive inputs to data integrity, let's discuss the risks involved in handling, interpreting, and reporting data.

Risks to Data Integrity

There are a number of risks to data integrity, including research misconduct, questionable research practices, poor mentoring or training in handling data, lack of proper guidance from the institution and lab, and poor data practices in the lab.

Research misconduct[3] is a clear danger to data integrity because experience has shown that most ORI misconduct cases involve data fabrication or falsification. By definition, such fabrication and falsification is the absolute antithesis of proper data practices. Also, many cases of misconduct lead to misrepresentations in the scientific literature. ORI and the institutions that respond to research misconduct allegations under the ORI regulation[4] have had over 120 scientific articles corrected or retracted due to research misconduct or, in a few cases, a determination by the institution that the article was not supportable.

Also, two studies by Kalichman and Eastwood[5] on research students and post-docs suggest that they will modify data to get published or funded over 10% of the time. Although the studies do not indicate that these incidents are clear misconduct violations under the Federal definition of research misconduct, i.e., fabrication, falsification, or plagiarism, it is likely that some unknown percentage would constitute formal research misconduct. Also, a recent article in *Nature* by Martinson,[6] et al, indicates that misconduct occurs in young and mid-career investigators funded by NIH over 1% of the time.

Questionable Research Practices

Questionable research practices (QRPs) are also a threat to data integrity and probably more common than incidents of formal research misconduct. These are practices identified by the National Academy of Sciences (NAS)[7] as practices that are not acceptable to the community and not in conformity with normative behavior, but not serious enough to be considered research misconduct. The NAS has identified the following behaviors as being QRPs:

- failure to retain research data
- maintaining inadequate research records
- authorship without a significant research contribution
- refusing reasonable requests for access to unique research materials or data
- misrepresenting speculations as fact or releasing preliminary research results, without sufficient data to allow critical review
- inadequate supervision or exploitation of subordinates
- using inappropriate statistics to enhance the significance of research findings

The NAS[7 (p6)] goes on to say that QRPs erode "confidence in the integrity of the research process, violate traditions associated with science, **and affect scientific conclusions**."

All of the QRPs identified by the NAS affect the data in one way or the other. Certainly, failure to retain the data or maintaining inadequate research records directly affects the ability of the research study to reach valid or reliable conclusions. Authorship without a significant contribution may give unfair credit (including credit for the data) to an investigator who does not deserve it. Refusing reasonable requests for data may violate scientific norms and violate the NIH[8] data sharing policy.

Misrepresenting speculations as fact or releasing preliminary results before adequate review may undermine the credibility of the scientific enterprise and unfairly raise expectations of scientists and the public in new discoveries. Inadequate supervision of subordinates may lead to poor data practices, and harm the training of young scientists. The use of inappropriate statistics may inappropriately inflate the apparent significance of research and may harm human subjects or the public if the findings lead to clinical treatments based on faulty data.

Poor Data Practices in the Lab Can Harm the Training of Young Scientists

A key responsibility of the Principal Investigator (and other senior members of the lab) is the training of students and young investigators. Training in good data practices is critical to the training of young scientists and certainly impacts on the quality of data generated in the lab. But what happens if little or no training is provided in the lab or worse yet, training is provided but it consists of poor data practices. This can positively harm both the trainees and the lab itself by generating poor quality data that may be incorrect and lead to faulty conclusions.

Based on ORI experiences with individual cases of research misconduct and allegations of misconduct that do not result in misconduct findings, ORI has identified a number of lab practices that do not result in good data outcomes.

Poor Data Practices in the Lab: Things You Want to Avoid

1. When the PI does not review the raw data, but instead relies on summary tables or figures that are prepared for publications or grant applications, he or she may end

up being surprised that the data is faulty or worse constitutes research misconduct due to fabrication or falsification. This can be very damaging to the reputation and morale in the lab and very embarrassing to the PI. This has occurred in several ORI cases.

2. If the PI or other senior staff in the lab puts unusual pressure on a lab technician or junior researcher to produce data on short notice for publication or to include preliminary data in a grant application, he or she may get what was asked but not what was desired. The data may be manipulated, falsified, or fabricated to meet the deadline. Of course, the lab really wants high quality data that will meet the objectives of the research study and advance scientific progress.

3. When the lab hires a new, highly recommended graduate student, new Ph.D., or post-doc to help with the research, the expectations are high to match the recommendations. What happens when something goes wrong and the data generated by the post-doc is suspect? Did the PI or other senior, trusted lab member provide adequate supervision and review and discuss the data with the new person during the early months in the lab. If not, the lab may be surprised by the suspicious data. Make sure someone experienced and responsible in the lab trains the new member in quality data practices that meets the lab's high standards.

4. Deeds speak louder than words. When the Lab Chief or PI models good data practices in recording, maintaining, interpreting, and reporting data, the lab staff pay great attention. If the Chief or PI meet regularly with lab staff to review data, discuss interpretations and potential findings, and talk about how the data might be presented for grant applications or publications, he or she is providing quality mentoring to the staff on data practices. Unusual questions about the data deserve additional discussion, including the all important judgment calls on omitting data due to experiment failures, poor technique, incorrect dosages, and extreme outliers. The PI has a most important role in explaining the rationale for these decisions and making sure that the staff do not use expediency to justify data omissions. The reasonable scientist approach should apply, i.e., would a knowledgeable scientist from another laboratory in the same discipline and familiar with the type of research conducted find the explanation for dropping data points appropriate. If on the other hand, the PI talks about good data practices but in fact makes convenient data decisions not justified by the scientific results, the lab staff are likely to become cynical and follow the PI's behavior and not what the PI says. The really good and confident scientific staff may decide to move to another lab under those circumstances.

5. Many scientists complain that while in training they had bad experiences in getting credit for their work. A typical complaint goes something like this: "I developed the hypothesis for the study, in consultation with my mentor, conducted the experiments, and wrote up the results but still did not receive first authorship on the paper. I tried to talk to my PI about it but she was not willing to

discuss it. Even though this occurred years ago, I still get angry about it when I recall the experience." This experience is not unusual and creates cynicism and morale problems in the lab. One of the reasons this occurs is because the lab often has a very informal policy on authorship and basically the PI calls the shots. If the lab adopted a written authorship policy and discussed it periodically or whenever a new scientist joined the lab, it would likely improve communication and reduce the number of problems and misunderstandings that occur for the junior members of the lab.

6. Is there a data policy in the lab? Are data practices in the lab very informal, with the staff keeping data any way they choose? Lack of a data policy and lack of data "norms" in the lab can result in inconsistent practices and confusion about how to handle and report data. In ORI's experience, lack of mentoring, lack of clear expectations by the PI and senior staff, and failure to discuss expectations for data with lab staff on a regular basis can result in irregular data practices and in some cases provide breeding grounds for research misconduct. Active involvement of the PI and senior staff is crucial to good practices in the lab. Hands on involvement by senior staff with graduate students, scientists in training, and new members to the lab is crucial to maintaining high quality data standards.

7. Is your lab too large to provide sound guidance to junior staff? If you work in a large lab and only a few senior staff are available to mentor young trainees, the risk of non-existent training or poor training in proper data practices increases. This problem is sometimes referred to as "absentee mentors." Either the PI or other senior staff are out of the lab frequently going to meetings, working on collaborations out of the lab, or involved in too many projects. The junior staff are left to their own devices or the one or two senior staff are overwhelmed and not able to adequately train the other staff. This too can lead to poor data practices or wide variability in how data are handled, interpreted and reported, thus reducing the quality of the data. Corrective actions may be needed such as having the PI spend more time in the lab or hiring more senior staff to share in the mentoring and training of junior scientists.

Positive Inputs to Support Data Integrity

To develop and maintain high standards for data quality in the laboratory requires an active support system. At a minimum, it requires a major commitment by senior leadership in the lab, particularly the PI and other senior and experienced staff. It requires a structure to support the importance of data quality, such as an institutional data policy which is then implemented in the individual labs with customized changes to take into account differences between disciplines and the type of research conducted in the specific lab. It also requires regular mentoring and formal training in good data practices. Each of these positive inputs will be discussed below.

Institutional Data Policies

In this section, we will review several data policies adopted by various research institutions to determine what types of policies are available and how they might impact on the quality of data generated at the institution.

The first policy is from Stanford University.[9] The Stanford policy provides that the research data belongs to Stanford, not the investigator. This is the general approach taken by research institutions and is consistent with legal precedent.[10] It also provides that the Principal Investigator (PI) is responsible for the maintenance and retention of the data. Under the policy, research records must be maintained for a minimum of 3 years from the end of the project period which is also consistent with NIH and Public Health Service policies. [11]

The policy also provides "data access" protections for research staff, including students, postdoctoral scholars, and other staff. This ensures that students, research trainees, and other staff are able to access data as appropriate to complete their training, develop publications, or for other purposes consistent with the research project under the overall guidance of the PI. The policy specifically states that data for students must be maintained at least until the student's degree is awarded, which for doctoral candidates could be several years beyond the NIH requirement for 3 years.

Like most institutions, Stanford provides that the data must normally remain with the institution except for special exceptions, such as transfer of a PI to another institution along with the research funding and relevant records. When the institution will not permit transfer of the original records, it normally would permit the investigator to take a copy of the research records.

Two other provisions of the policy are worthy of mention. First, it specifically states that Stanford's data responsibilities include "facilitating the investigation of charges, such as scientific misconduct or conflict of interest." The policy also provides that the PI "should adopt an orderly system of data organization" and "should communicate the chosen system to all members of the research group." This suggests that the PI has some responsibility to communicate expectations to trainees or provide some formal training in the chosen system of data organization.

The next policy is from Duke University.[12] Duke requires maintenance of data records for a minimum of 5 years. It provides that data records must be kept to document the "experimental methods and accuracy of data collection as well as the methods and accuracy of data interpretation." It further provides that co-investigators and trainees "are an integral part of the research project" and have the right to review all records and data. Faculty or other responsible investigators have the obligation to ensure that research records are appropriately documented.

Among the data policies reviewed here, the University of New Hampshire[13] has the most detailed policy. The policy provides a detailed definition of research data that includes:

1. Raw numbers, field notes and observations, detailed experimental protocols, procedures for analysis, data obtained from instruments, interviews, and surveys.

2. Computer files, databases, research notebooks, lab journals, tables, charts, slides, videotapes, sound recordings, and photographs.

3. Physical collections, biological specimens, cell lines, derived reagents, marine life, drilling core samples, genetically altered microorganisms, or other tangible artifacts.

Each PI or other investigator is responsible for "adopting sound policies for procuring and maintaining research data, and for educating those supervised about these practices and the associated rationale." The PI is also responsible for the physical storage of the research data; providing reasonable access to each member of the research group; and is instructed to come to a written understanding with each student investigator regarding data access.

The data policy further provides that the "research data must be retained in sufficient detail and duration to allow appropriate response to questions about research accuracy, authenticity, primacy, and compliance with laws and regulations governing the conduct of research." Consistent with standard practices for the relevant discipline, each PI or other investigator is also "responsible for adopting sound practices for procuring and maintaining research data" and for "educating those supervised about these practices and the associated rationale."

The data policy provides for a general period of data retention of three years from the date of data collection or three years from the date of termination of the sponsored agreement or date of publication of a paper based on the data, whichever is longer. It also acknowledges the need to follow the NIH and PHS policy on data sharing once the results have been published or provided to the sponsor.

The last data policy reviewed here is really two policies from the Harvard Medical School[14] and focuses on both basic and clinical research. Although relatively short, in other ways the Harvard policies are the most ambitious. The guidelines recommend that a faculty member should be assigned to mentor a junior investigator and the ratio of mentors to trainees should be small enough to permit close interaction by the mentor. The mentor is expected to supervise the experimental design, the process of acquiring, recording, examining, interpreting, and storing data. Regular lab meetings are recommended to review and discuss the research and data. Original experiments should be kept in bound notebooks if possible.

Each unit is expected to develop and adopt "specific guidelines" to identify practices that will enhance the quality of research conducted by the lab members for the relevant discipline or type of research being conducted. In other words, specific labs or departments are expected to provide more detailed guidance consistent with the key principles addressed in the Harvard guidelines. This approach encourages greater specificity and relevance to the type of research conducted in the particular unit, improving the odds that higher quality research will result. This devolution from general principles to specific guidance should improve the quality of research and give more confidence to trainees in how to carry out the details of the research study.

The Harvard[14] guidelines provide that for research involving primary data collection, the PI should retain the original data for as long as practical, but not less than 5 years from the first major publication or from the completion of an unpublished study. Research units are also expected to develop their own authorship guidelines and "honorary authorship" is considered inappropriate.

Policies to Support Data Integrity

We have just discussed several institutional data policies. Adoption of such policies provides structure to institutional efforts to ensure that data is managed properly and that data which is used and reported by the institution is of high quality and supports the scientific findings. Sufficient detail in the data policy is needed to make it useful.

Implementation of the institutional policy at the group level is important. Because it is not possible to develop a policy that fits the needs of all departments or laboratories within the institution, there must be more specific guidance at the group level for handling data. Some data issues can fit all comers, such as stating "it is unacceptable to falsify or fabricate data in the institution." However, some scientific disciplines will have special techniques, methods, or procedures for conducting research that require further, or different, guidance on how the data should be collected, retained, selected, interpreted, reported, or omitted. This could require the research group to develop its own unique guidance on data, or provide additional guidance that supplements the institution-wide policy. Whichever approach is taken, it is important that the data policy is robust enough to make sure that is useful to the research group and particularly to the more junior members who are still being trained in proper scientific practices, including the handling of data.

What Issues Should Be Addressed in a Data Policy

As illustrated by the data policies that were previously reviewed, some of the basics that should be included in data policies are descriptions of–

- the time period for retention of the data
- access of the research staff, including students, to the data
- who has overall responsibility for the data
- the investigator's ability to take the data, or a copy thereof, when leaving the institution
- the institution's ownership of the data and it's right to take control of the data for a variety of reasons such as patents, regulatory issues, audit purposes, or other reasons
- investigator and trainee access and review of the data for purposes of publication, presentations, grant applications, or other reasons
- procedures for security of the data, lab notebooks, physical samples, biological samples, and other research records
- standard procedures for recording and maintaining data in the lab or other research group

- plans for regular lab meetings to review and discuss the data
- plans for training on the lab's data policies
- who has overall responsibility for the data: this is usually the PI or other lead investigator

Training in the Data Policy and Guidelines

Whatever standards for data are adopted by the institution or individual groups, there should be some training component for lab members to inform members of the principles, standards, and expectations for data and provide hands-on guidance by senior staff so that lab members become accustomed to how data is actually managed in practice. This should expressly include discussion of difficult issues, such as the appropriateness of excluding data points that do support the expected findings of the research project. Does the lab have a pre-defined standard for excluding data, such as "all data points that are two standard deviations from the central tendency of the data are excluded." This should also mean that "good" data points are also dropped when they meet that test.

Other difficult issues should be handled similarly. For example, a graduate student conducting an experiment said that she accidentally added too much active chemical solution to the experiment and when she saw the results made no sense, she threw out the data. Was this recorded in the lab notebook? Was there enough detail to permit the PI to determine exactly what happened? Would the rationale for discarding the data convince an experienced scientist outside the lab who was familiar with the type of research and the techniques used?

Regular Lab Meetings to Discuss the Data

Understanding the data and being familiar with the full range of data issues that arise is critical to becoming a qualified scientist. For most types of research, the data is the critical outcome that leads to new findings, generates new publications, and supports new research projects so the beginning scientist can eventually become a PI with his or her own projects. Taking formal classes on statistics, laboratory practice, methods, developing hypotheses, and critical reasoning skills helps develop the scientific skills of young scientists. However, hands-on training by an experienced senior scientist, often the PI, is needed to develop critical data skills over a period of months and years. Without this long term mentoring process by one individual over many years, or multiple individuals over a long period, it is difficult for the young scientist to get enough repetition in handling data issues and developing other scientific skills to become a highly qualified scientist.

Thus, the research group must provide frequent, repetitive opportunities for scientists in training to obtain these skills. One of the best ways to do this is to have weekly or bi-weekly meetings with the research staff, where research trainees can give presentations and discuss their progress for ongoing research, including discussion of

the data, identifying any problems that may have developed, determining whether the data supports the hypothesis, and other data issues that may arise.

Other opportunities to discuss the data come naturally when it is time to develop publications or submit new grant applications. The entire lab staff that worked on the supported research should participate in the discussion by making comments on how the data collection went, discussing whether it supported the study or not, and how the data might fit into a publication or new grant application. Data problems should be acknowledged and explanations for resolving those problems can be explored. If novel issues arise, that should be discussed as well. Even those scientists or students who are not ready for authorship or are not listed on the grant application can learn from hearing the lab discussion on how the project is proceeding or whether additional work is needed and why.

In addition to formal lab meetings where the PI or other senior staff take the lead in lending direction to the discussion and resolution of data issues, the PI can ask the students, post-docs, and other less senior staff to develop case studies on data issues to present to the other lab members. These could be constructed out of actual experiences in the lab or experiences of other colleagues working in a different lab. By asking the scientist in training to develop the case and lead the discussion, the lab member will become more familiar with the different data issues that can arise and will learn from the lab discussion that is generated.

Although most PI's instinctively realize that constant mentoring and hands-on training are required to develop skillful, responsible scientists, ORI data[15] indicate that a significant minority, approximately 25% do not give it a high priority. Sometimes, these PI's are called absentee mentors. In such cases, where junior scientists are not getting regular, quality mentoring, the lab members may develop poor or inconsistent data practices. Of course, it is also possible that frequent mentoring can lead to poor research practices, where the mentor does not take responsible research practices seriously, or cuts corners in his or her research when it is convenient and because it gives the PI a short term benefit, such as getting published or funded based on data that have been "cleaned up."

Over time, labs will develop their own criteria for adjusting data points when they believe extraneous factors have altered the experiment. This may happen when an obvious mistake has occurred, such as use of the wrong chemical, incorrect dosages, failure to follow the proper method, etc. How is this information communicated to new scientists entering the lab? Is it passed on by osmosis? A better approach would be reducing it to writing and adding it to the lab's data policy or having a formal meeting to discuss it with new lab members. To avoid making these judgments on "gut feeling," it would seem to be more beneficial to have an open lab discussion whenever a new example was identified where experimental mistakes seemed to justify correcting the record. By discussing the specific principle or rationale for correcting the scientific record, the lab would reduce the possibility of developing a haphazard approach to the handling of research data and the scientific record. One way to do that is by using a "reasonable scientist" standard for making a decision. In other words, what would a

skilled scientist from outside the lab familiar with the research and the methodologies believe is reasonable.

Procedures for Publishing Papers and Preparing Grant Applications

Getting funding for research projects and publishing the results of the research are critical to the career of PI's and a necessary condition for the training of new scientists who need more experienced scientists to teach them the methods, principles, and specific skills to succeed in science. Of course, the research data generated in the lab must be recorded, retained, and interpreted before it can be properly used in publications and grant applications. This is a wonderful opportunity for the PI or other senior lab staff to get the science trainees involved. All research staff who worked on the various projects should be permitted to review the data, participate in discussions about the results, and make suggestions about how the data should be presented in publications and new grant applications. Some of the trainees may merit authorship on papers based on their substantial contributions to the research project. Likewise, other researchers who performed experiments and generated data may be listed as staff on new applications. The PI usually sets the tone on how these training opportunities are utilized. Active discussion by all lab members who had significant roles in the research may improve the publication or grant application and provides many teaching opportunities for the PI and other senior staff.

Two Studies Related to Data Integrity and Mentoring in the Lab: The Martinson Study and an ORI Study on Research Integrity Measures Utilized in Biomedical Research Laboratories

A recent study published in *Nature* by Martinson,[6] et al, reported data on self reports by early (post-docs) and mid-career scientists (who had already received their first RO1 grants[a]) on a variety of scientific behaviors, either actual research misconduct or questionable research practices that do not meet the normative behavior expected by the scientific community or the public. While some of the questions asked in the study can be criticized for ambiguity, thus raising doubts about the meaning or significance of the results, overall the results indicate that the rate of questionable scientific practice in areas such as dropping data points based on a "gut feeling" (15%), inappropriately assigning authorship credit (10%), inadequate record keeping (27%), failing to present data that contradict one's own previous research (6%) publishing the same data in two or more publications (5%), also referred to as duplicate publication, and overlooking others' use of flawed data or questionable interpretation of data (13%) gives legitimate cause for concern. This suggests that improved graduate education in appropriate scientific practices and better training by Principal Investigators of young scientists in the lab, especially in the areas of data management, interpretation, and reporting, is warranted.

a. An ROI grant is a research project grant that is typically initiated by an individual scientist (a Principal Investigator) who is experienced in the type of research proposed in the application.

In the second study conducted in 2002,[15] ORI surveyed over 6000 NIH funded Principal Investigators (PI's) to determine what measures they utilized in their labs to ensure the integrity of the research. About 2900 individuals responded. A number of important questions concerning how the PI managed the laboratory and provided mentoring and training to the lab staff were addressed. These included issues such as how the laboratory records were kept; how much time the PI spent in the laboratory; how often the PI held laboratory meetings to review research and provide opportunities for the staff to present the status of ongoing work; and how much time the PI spent on mentoring and supervision on average.

The scientists who responded were assigned to 3 categories of science: basic, clinical, and epidemiology. Overall, scientists reported that they kept their data in permanently bound notebooks 32% of the time. For basic scientists it was 39%. However, the overall data indicated that most researchers used a variety of methods for keeping data, including loose-leaf notebooks, digital files, and audio-visual media, in addition to the bound notebooks.

Table 1: Method for Collecting and Storage of Data
% of Data Stored Using Alternative Media

Principal field	N	Loose-leaf	Permanently-Bound	Digital Files	Audio-Visual Media
Basic	2208	29.5	38.7	42.4	21.4
Clinical	406	23.1	15.4	65.5	14.2
Epidemiological	296	11.5	9.3	76.9	7.9
All	2900	26.7	32.4	49.2	18.9

In response to a question about how long the data is retained, the PI's reported an average time of 12 years for published and unpublished data, far exceeding the 3 year NIH requirement.

Table 2: Data Retention
Minimum length of time data are retained (years)
Based on a sub-sample of 50% of the respondents; approx. 1400 PI's

Principal field of investigator	Number (N)	Data unlikely to be published	Data reported in a publication	After filing a patent application
Basic	1082	12.5	12.9	14.1
Clinical	202	11.3	11.5	13.6
Epidemiological	140	12.2	12.7	12.5
All	1418	12.3	12.7	14.1

One interesting question that directly affects how the data is handled was a question about how the lab documents its rationale for excluding data when outlier data points occur or an atypical experiment is conducted, e.g., there is a failure in the methodology, incorrect dosage, wrong chemical, etc. The PI reported that this information was recorded 53% of the time. On the other hand, a separate question that asked the PI what "percentage of manuscripts clearly described the criteria for

inclusion or exclusion of data" reported 72% of the manuscripts did so, somewhat better than the 53% for documenting outliers. Even so, this means that 47% of the time, exclusion of outliers is not documented, and 28% of the time, no clear inclusion/exclusion criteria for data in manuscripts is provided. Thus, for a substantial percentage of the time, the lab has the opportunity to make "convenient" decisions to change or exclude unhelpful data that does not support the hypothesis. This loose handling of the data is consistent with the Martinson[6] study that suggests that data decisions are often made based on "gut feelings", or other rationales not based on any apparent scientific facts or principles.

A number of questions were asked about supervision and mentoring practices in the lab. Overall, the PI reported meeting with each staff member about 2.5 hours a week. The PI reported reviewing lab notebooks about 20 times a year and holding lab meetings 33 times a year, a little more often than once every 2 weeks. The PI also reported being in the lab (and thus available for mentoring or supervision) about 66% of the time and spending about 10 hours a week mentoring. The majority of PI's report, on average, having 5.5 scientists in the lab to supervise. See table below.

Table 3: Measures of Supervision by Principal Investigator of Researchers in His/Her Laboratory. Based on a sub-sample of the respondents, approx. 1400.

Scientific field of investigator	Number (N)	Hours/week spent with each supervised researcher	Number of examinations of lab notebooks in past 12 months	Number of meetings with each supervised researcher in past 12 months	% of time Laboratory Director is physically present in the laboratory	Hours per week spent mentoring
Survey item no.		(Q19)	(Q21)	(Q22)	(Q24)	(Q27)
Basic	1039	2.66	22.0	37.2	67.2	11.1
Clinical	192	2.01	14.2	17.9	63.8	6.7
Epidemiological	136	2.08	9.7	14.6	62.7	6.4
All	1365	2.52	20.2	33.5	66.7	10.0

While the majority of PI's spent significantly more time with trainees (in the lab 66% of the time and providing mentoring or supervision for 2.5 hours per trainee a week), a subset of the PI's, approximately 25% indicated that they held lab meetings 12 times or less per year, spent less than one hour a week supervising each trainee, and reviewed lab notebooks approximately 3 times a year. This suggests that a substantial minority of PI's allot insufficient time to the difficult, but important, task of training young scientists and might be called "absentee mentors." See Table 4 below.

Table 4: Measures of Supervision by Less Involved PI's

Hours/week spent with each supervised researcher	Examination of Lab Notebook in past 12 months	Lab meetings/year	
Up to 1 hour	Up to 3 times/year	Up to 12 per year	

Overall, these two studies suggest that there is room for improvement by PI's in how they mentor and supervise the scientific staff in the lab. While the majority of PI's spend a substantial amount of time in the lab, thus making themselves available for supervising or mentoring the junior staff, a significant minority are clearly less involved, providing little attention to the mentoring and supervisory responsibilities of the scientific leader of the lab. This latter group seems to be "disengaged" from the leadership responsibilities of the lab, making it more likely that the lab staff will do their "own thing" and raising the risk that the individual scientists will deviate from the accepted scientific norms of handling the research data, increasing the risks that questionable research practices, or even research misconduct, will occur.

In contrast the primary PI group shows a strong commitment to mentoring, supervision, holding regular lab meetings, reviewing research notebooks frequently, and providing regular opportunities for lab members to present and discuss their ongoing research. Based on ORI's experience with hundreds of allegations of research misconduct, this latter group is much more likely to teach and encourage good data practices in the lab.

Conclusions

Most PI's clearly recognize the importance of providing regular mentoring and training to junior research staff in appropriate scientific practices, including the management of data. However, changes in the research enterprise the past several years, including a slow down in research funding, the growth of larger labs with more research trainees or senior post-docs to supervise, and the increase in collaborations with interdisciplinary teams spread over many institutions or internationally have made the research process more demanding and time consuming. This has increased the pressure on some PI's to focus more on getting published and bringing in the grant money and to decrease the time spent on mentoring the junior scientists. While this may be understandable, it is short sighted.

In the long run, the success of the lab depends on the quality of its staff, including students, new Ph.D's and post-docs. By renewing the lab's focus on regular and quality training of junior staff on data management issues, the PI can increase the probability of the lab's success and provide quality training for the next generation of scientists. This will not only benefit the lab and its staff, it will benefit the U.S. and international research enterprises through better quality research.

REFERENCES

1. RCR resources posted on the ORI website: http://ori.dhhs.gov/education/rcr_general.shtml; accessed on 11/3/05.
2. Institute of Medicine (IOM) (2002) "Integrity in Scientific Research: Creating an Environment that Promotes Responsible Conduct." National Research Council of the National Academies, Washington, D.C., p. 9.

3. Research misconduct is defined as "fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results."42 CFR 93.103, Federal Register, p.28386, Vol. 70, May 17, 2005.
4. Public Health Service Policies on Research Misconduct, Federal Register, 28370-28400, May 17, 2005, to be codified at 42 CFR Part 93. This regulation is posted at the ORI website at http://ori.dhhs.gov/documents/42_cfr_parts_50_and_93_2005.pdf; accessed on 11/3/05.
5. Kalichman, M.W., and Friedman, P.J. (1992) "A Pilot Study of Biomedical Trainees' Perceptions Concerning Research Ethics." *Academic Medicine* **67**: 769-775. Eastwood, S. et al. (1996) "Ethical Issues in Biomedical Research: Perception and Practices of Post Doctoral Fellows Responding to a Survey." *Science and Engineering Ethics* **2**: 89-106.
6. Martinson, B.C., Anderson, M.S., and de Vries, R. (2005) "Scientists Behaving Badly. *Nature* **435**: 737-738.
7. National Academy of Sciences. (1992) *Responsible Science: Ensuring the Integrity of the Research Process.* National Academy Press, Washington, D.C., pp. 5-6.
8. NIH Guide: Final NIH Statement on Sharing Research Data: February 26, 2003, NOT-OD-03-032. For additional guidance, see http://grants.nih.gov/grants/policy/data_sharing/; accessed 11/3/05.
9. Stanford University Policy: Retention and Access to Research Data. http://www.stanford.edu/dept/DoR/rph/2-10.html accessed 5/10/2005.
10. Forsham v. Harris, 100 S.Ct. 977 (1980). "We hold here that written data generated, owned, and possessed by a privately controlled organization receiving federal study grants are not 'agency records' within the meaning of the Act when copies of those data have not been obtained by a federal agency subject to the FOIA."
11. 45 CFR 74.53, Retention and Access Requirements for Records.
12. Duke University Policy on Data Retention and Access. http://www.ors.duke.edu/ors/policies/data.html; accessed on 5/10/05.
13. University of New Hampshire Policy on Ownership and Management of Research Data. Http://www.unh.edu/orps/downloads/DataOwnershipandManagement.pdf; Accessed on 11/8/05
14. Harvard Medical School "Guidelines for Investigators in Scientific Research" and "Guidelines for Investigators in Clinical Research." http://www.hms.harvard.edu/integrity/scientif.html; http://www.hms.harvard.edu/integrity/clinical.html; both accessed on 5/13/05.
15. Survey of Research Integrity Measures Utilized in Biomedical Research Laboratories: Final Report, pp. 35-36, Exhibit 2D, August 29, 2003. Conducted under contract to the American Institutes for Research. The report is posted on the ORI website at http://ori.dhhs.gov/documents/research/integrity_measures_final_report_11_07_03.pdf; accessed on 11/3/05.