PROFESSIONAL INTEGRITY: BEST PRACTICES FOR PUBLISHING YOUR RESEARCH

Data Management and Integrity

Christina N. Bennett, PhD and Marsha Lakes Matyas, PhD

Collaborating Partners

American Physiological Society

Biomedical Engineering Society

Society for Biological Engineering

Data Management and Integrity

Christina N. Bennett, Ph.D.
Associate Publisher, Ethics and Policy
and
Marsha Lakes Matyas Ph.D.
Director of Education Programs
American Physiological Society

Published by
The American Physiological Society
Bethesda, MD

In collaboration with
The Biomedical Engineering Society
Landover, MD
and
The Society for Biological Engineering
New York, NY

This module is part of a series of seven teaching modules designed to promote best practices in publication ethics for life scientists and biomedical engineers who publish research papers. All materials in this module have been peer-reviewed by both the Advisory Board (below) and a panel of Responsible Conduct of Research (RCR) course instructors. They also were field-tested with graduate students in physiology. The modules were developed with support from the National Science Foundation (NSF) (#SES -1238368) and in collaboration with staff and members of the American Physiological Society (APS), Biomedical Engineering Society (BMES), and Society for Biological Engineering (SBE). The modules represent the views of the authors and do not necessarily represent the views of NSF, APS, BMES, or SBE. The information in these modules is designed to represent a summary of best practices and advice at the time of publication. They are not meant to serve as legal advice or publisher policy and do not in any way guarantee protection from professional ethics charges. For more information on how the materials were developed and tested, please contact the authors: Marsha Lakes Matyas: education@the-aps.org; Christina N. Bennett: apsethics@the-aps.org, American Physiological Society, 9650 Rockville Pike, Bethesda, MD 20814, www.the-aps.org.

Project Staff

Principle Investigators

Marsha Lakes Matyas, Ph.D., APS
Christina N. Bennett, Ph.D., APS
Jerry Collins, Ph.D., BMES
June C. Wispelwey, SBE/AIChe

Society Directors

Martin Frank, Ph.D., Executive Director, APS
Edward L. Schilling, III, Executive Director, BMES
June Wispelwey, Executive Director, AIChE

Project Advisory Board

Kim E. Barrett, Ph.D., U. of California, San Diego
Jason Borenstein, Ph.D., Georgia Inst of Tech.
Hannah V. Carey, Ph.D., U. of Wisconsin
Lydia Contreras, Ph.D., U. of Texas at Austin
Jane Zimmer Daniels, Ph.D., JZD Consulting
Michael M. Domach, Ph.D., Carnegie Mellon U.
Gregory L. Florant, Ph.D., Colorado State U.
Elizabeth Heitman, Ph.D., Vanderbilt U.
Charles H. Lang, Ph.D., Penn State Coll. of Med.
Rudy M. Ortiz, Ph.D. U. of California, Merced
Thomas A. Pressley, Ph.D., Texas Tech U. HSC-Lubbock
Gary C. Schoenwolf, Ph.D., U. of Utah
Annie Whitaker, Ph.D., Louisiana State U. HSU-New Orleans

Suggested citation: Bennett, C. N. and Matyas, M. L. Data management and integrity. Bethesda, MD: American Physiological Society, 2017, http://www.lifescitrc.org/resource.cfm?submissionID=11068.

Copyright @ 2017 American Physiological Society
This material is copyrighted under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International copyright (https://creativecommons.org/licenses/by-nc-sa/4.0/): 1) Attribution: Give appropriate credit, provide a link to the license, and indicate if changes were made, but not in any way that suggests the licensor endorses you or your use; 2) NonCommercial: You may not use the material for commercial purposes; 3) ShareAlike: If you change or build upon the material, you must distribute under the same license as the original; 4) You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.
Primary graphics source: clipart.com

Data Management and Integrity Module

Table of Contents

	Page
I. Pedagogy	1
II. Module Objectives	4
III. Instructor Guide	5
Activity A: "May I Have the Data?"	8
Activity B: Critique Your Lab Notebook	10
Activity C: Critique a Poor Example	11
Activity D: Evaluating Journal Guidelines on Data Presentation	12
Activity E: Two-Minute Challenges	13
Activity F: Best Practices for Legends, Tables, Graphs, and Images	15
Activity G: My Data Management and Integrity Checklist	19
Presentation Slides and Text	20
References and Resources	34
IV. Student Handouts	35
Activity A: "May I Have the Data?"	37
Activity B: Critique Your Lab Notebook	38
Activity C: Critique a Poor Example	40
Activity D: Evaluating Journal Guidelines on Data Presentation	42
Activity E: Two-Minute Challenges	44
Activity F: Best Practices for Legends, Tables, Graphs, and Images	46
Activity G: My Data Management and Integrity Checklist	58
Student Slide Handout	60

I. Pedagogy

This module is designed to promote best practices in publication ethics for life scientists and biomedical engineers who publish research papers. The goal is for students to not only understand professional standards of practice in research manuscript development but also to be able to apply these standards to their own work AND to be prepared to teach them to their own students in the future. Toward that end, this module employs student-centered learning strategies that engage students across the spectrum of Bloom's taxonomy (see below). For best impact, students should not simply sit and listen or read and answer questions. Instead, we encourage you to use multiple teaching methods and activities that engage students in actively exploring the topic. Some suggestions you will find in this module include:

- **Interactive Lecture:** The lecture slides and notes include a number of places to stop and engage students in working out a problem, discussing a policy, or reviewing a case study.
- **Think/Pair/Share:** Often part of an Interactive Lecture, students are given a problem to address first on their own, and then they are asked to share their responses with a partner, followed by sharing with the whole class.
- **Voting Cards:** Particularly when discussing ethics issues, students prefer not to raise their hands to indicate their answer to a group question. Consider using voting cards with a simple large-print "Yes" on one side and "No" on the other. Everyone raises their hands and votes and you can quickly visualize the class response. An alternative is "thumbs up/thumbs down" but this is harder to see.
- **My Best Practice Checklists:** These are working documents each student develops to use now and in the future as their personal checklists of best practice in publication ethics.
- **PASS IT ON:** As part of their My Best Practice Checklists, students should make a plan for teaching publication ethics to their future trainees.

Instructors can pick and choose which activities and resources they want to use from the module. However, we encourage you to consider using the Learning Cycle approach because of its rich opportunities for student-centered learning. Alternatively, the Homework/Interactive Lecture/Activities (HILA) approach can be used when class time is limited. Both approaches are outlined below.

Learning Cycle

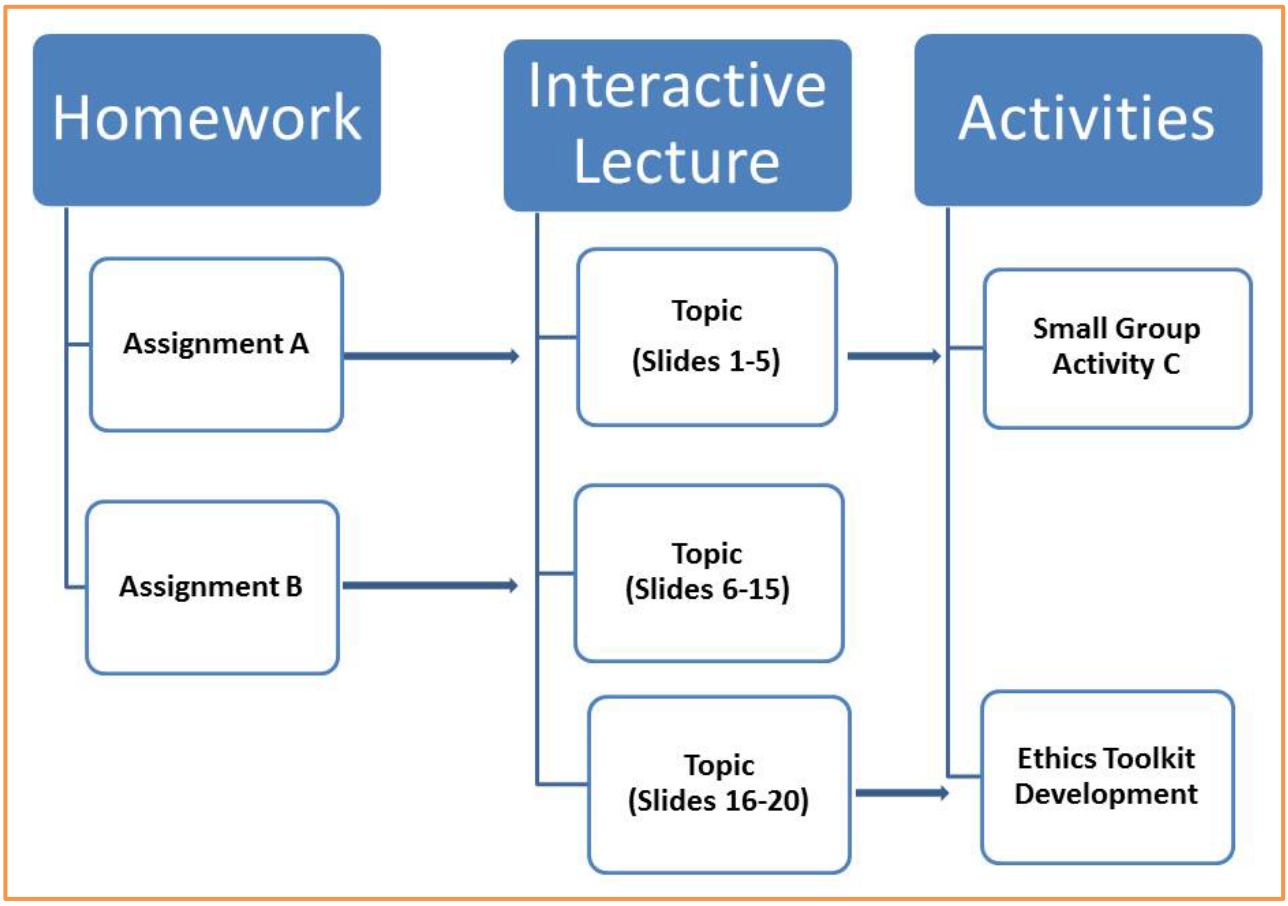
- **Engage**: Piques students' interest in the topic and poses questions or issues that capture their thinking. *Examples:* News articles on ethics violations and examples of manipulated figures.
- **Explore**: Students explore and ask questions, investigate via inquiry, make observations, and test hypotheses. Students should generate additional questions by the end of the exploration phase. *Examples:* Case study that students must try to resolve individually or in groups without additional information on professional standards of practice (these would be readdressed in the elaborate phase below), compare CV's of researchers, interpret letters from editors including comments/questions from reviewers, or write a letter to the editor describing figure manipulation in a manuscript to be submitted.

©American Physiological Society 2017

- **Explain**: Students and instructors use questioning/discussion, reference materials (print and online), expert presentations, and other resources to gain a better understanding of the key principles of the lesson and how they apply to the questions raised by students in the explore phase.
- **Elaborate**: Students apply what they have learned to real scenarios. *Examples:* Students revise their response to the explore phase case study using the principles and knowledge gained in the explain phase, and then do the same for a new case study or, ideally, their own work. Create a personal action plan or checklist for professional standards to use in the future.
- **Evaluate**: Evaluation occurs through each phase, with evidence collected of both student understanding of key principles and information and their ability to apply it to new situations and problems. *Examples:* Changes in approach to case study before and after the explain phase. Personal action plan/checklist addresses the key principles of professional practice. Key principles are applied appropriately to new case studies. Can also include quizzes or tests of content knowledge of professional standards of practice.

Homework/Interactive Lecture/Activities (HILA)

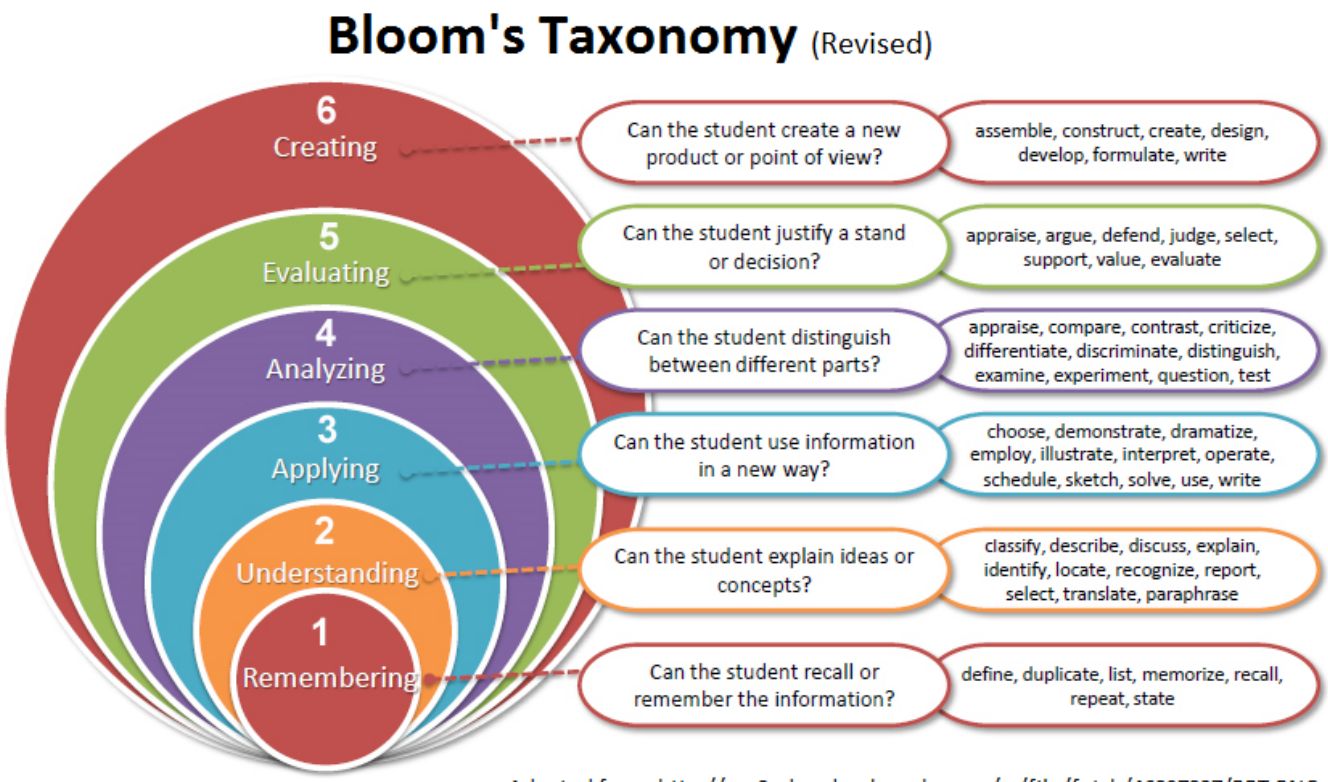
Homework activities are discussed either during the Interactive Lecture or during follow up activities.



©American Physiological Society 2017

Bloom's Taxonomy

Bloom's Taxonomy (established 1956, revised 2001) helps educators more effectively structure their teaching, student learning, and assessment of skills and knowledge. Organizing learning objectives by Blooms Taxonomy helps educators assure that lessons do not focus solely on memorizing basic knowledge but also challenge students to apply what they learn, evaluate new situations, and create solutions to challenging problems. Higher level objectives engage students in learning situations that are more complex and abstract. Overall, the professional ethics lessons in this series of seven modules focus strongly on the higher Bloom's levels (5 – Evaluating (20%) and 6 – Creating (21%)) in addition to including objectives for basic knowledge (Level 2 – Understanding (30%)) and application (Level 3 – Applying (14%)).



Adapted from: http://pcs2ndgrade.pbworks.com/w/file/fetch/46897827/RBT.PNG

Student Handouts

The student section of this guide is formatted for easy duplication. This guide is also available as an MS Word (.doc) file (See References). We encourage you to provide both printed and .doc formats to students. The lessons are designed to help students create a personalized guide for their future work; developing their notes and best practices plans in a .doc format will help students use as well as modify their plans in the future.

©American Physiological Society 2017

II. Module Objectives

Students will be able to:	Bloom's Levels
1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.	2, 3
2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.	2, 3
3. Develop criteria for assessing quality of data and use those criteria to evaluate the quality of both data prepared by colleagues/collaborators and findings reported in the literature.	5, 6
4. Formulate best practices for presenting digital and non-digital data to enhance clarity and avoid inappropriate manipulation via figures and legends.	6

©American Physiological Society 2017

III. Instructor Guide

Target Audience
This module can be used with both graduate students and undergraduate students. It was initially designed for early career graduate students in biological science, medical science, or biological engineering graduate programs. Graduate students are likely to be somewhat aware of the academic publishing process but may not have had first-hand experience. Undergraduate students engaged in research and scientific writing may also find the materials useful.

Instructor Tips

1) Select the objectives and related activities that you want to address. Edit the PowerPoint Presentation to include the activities and objectives selected.
2) The script/key points for the presentation are in the notes section of the PowerPoint slides.
3) We encourage you to share 1-2 minute personal stories, when appropriate. Keep the stories positive (i.e., "I had a dilemma and I utilized a best practice…dilemma resolved").
4) Allow students to reach conclusions on their own. You are their guide through this class. Facilitate discussion to keep them on task and within time limits.
5) **Be sure to include the "My Checklist" activity in each unit. This is the major "take away" lesson through which students integrate what they have learned in order to develop: 1) their personal checklists for ethical writing; and 2) their plans for teaching publication ethics best practices to their future trainees.**

Teaching Approaches
Learning Cycle and Homework/Interactive Lecture/Activities (HILA) approaches are outlined below.

Evaluation Rubrics and Test Questions
Evaluation rubrics for assignments and test questions are available on request from the authors (email: education@the-aps.org).

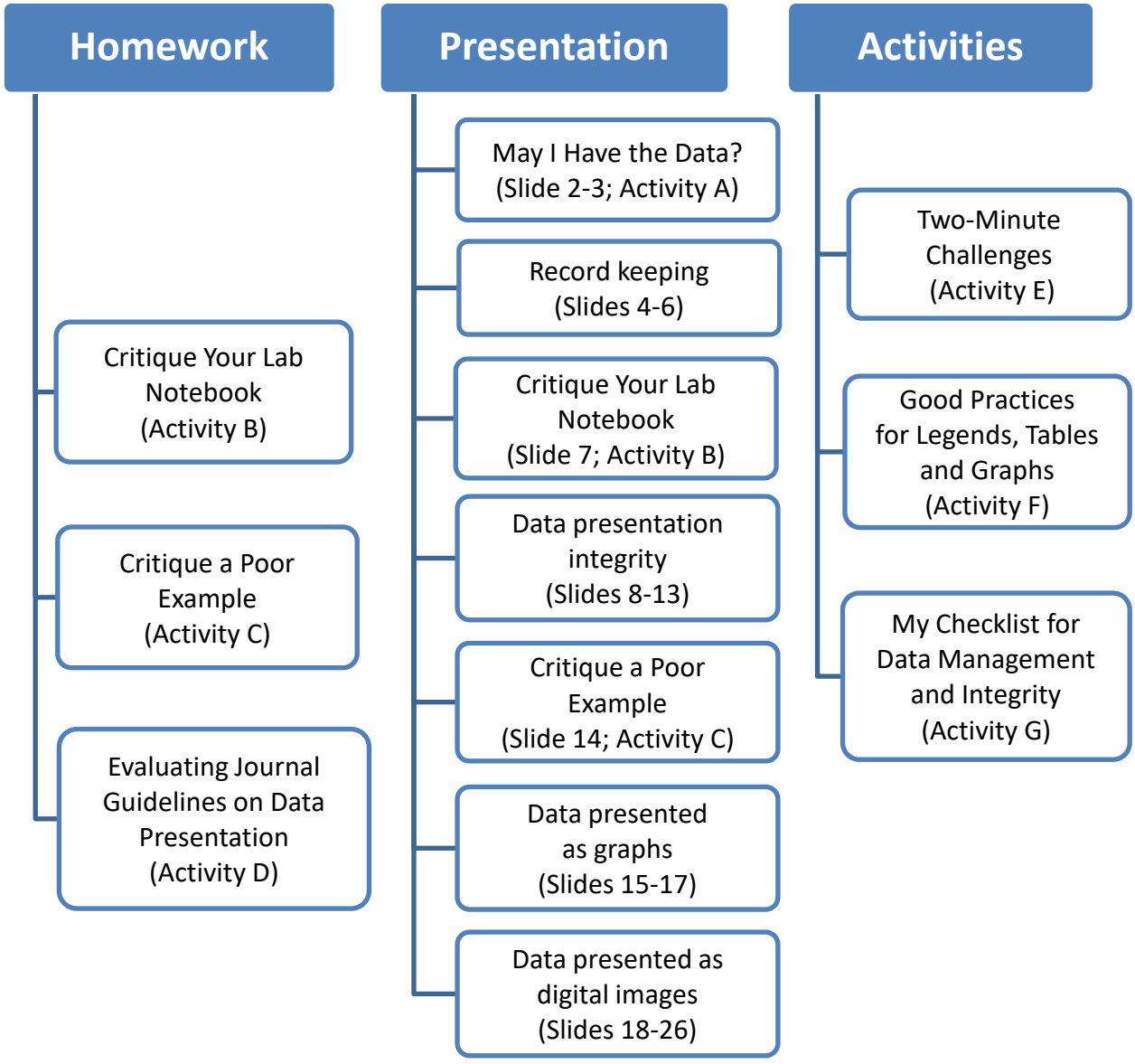
©American Physiological Society 2017

Data Management and Integrity
Learning Cycle

Engage	• Complete **Activity A:** "May I Have the Data?" (all parts) and Part 1 of **Activity D:** Evaluating Journal Guidelines on Data Presentation. If time allows, students should discuss their work in small groups. • Complete **Activity C:** Critique a Poor Example and bring a copy of their example(s) to class along with their completed worksheet. Read "Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm."
Explore	• Complete **Activity B:** Critique Your Lab Notebook and **Activity E:** Two-Minute Challenges. • Read the article, "Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm."
Explain	• Present **Interactive Lecture**.
Elaborate	• Finish **Activity D:** Evaluating Journal Guidelines on Data Presentation in small groups or as a whole class. • Complete **Activity F:** Best Practices for Legends, Tables, Graphs, and Images and review answers in class.
Evaluate	• Complete **Activity G:** My Data Management and Integrity Checklist individually; should be reviewed by instructor. • Quiz/test questions and answer keys are available from the authors.

©American Physiological Society 2017

Data Management and Integrity
Homework/Interactive Lecture/Activities



©American Physiological Society 2017

Activity A
"May I Have the Data?"

Purpose This activity will help students understand best practice in storing original data so it is both accessible and understandable to the broader scientific community.

Objective

1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.

Procedure **Learning Cycle:** Students complete Activity A as an Explore activity. They should view the video (as homework or in class), answer the questions, and then discuss them with a partner (as homework or in class).
HILA: Students complete the Activity A during the Interactive Lecture.

The video is a simple animation of two bears (scientists) talking about a request to share published data. While the animation is childlike, the conversation is important.

Answers for the Instructor are provided in *italics*.

Discussion Questions

1. **Is Judy (the black and white panda bear) justified in asking Brown Bear for a copy of his original data? Why or why not?**
   - *Yes, she was. Being prepared to share original data was a condition of both the research funding and the article publication.*

2. **Brown Bear has trouble remembering where he stored the original data. Is it his job to know where the data is stored?**
   - *Yes, that would be part of being prepared to share the data.*

3. **Brown Bear sent Judy his thumb drive containing the ONLY copy of his original data to Judy. Would you have done that? Why or why not?**
   - *No! The chance that it could get lost in the mail, the drive could become corrupted, or Judy would not return it make it risky to send the only file. Best practice is to have multiple copies of all data stored in secure locations. He should have sent her a copy.*

©American Physiological Society 2017

4. **Accessing the data requires a specific program. Do you think Brown Bear is required to have a copy of that program available?**
   - *This is tricky. Brown Bear should be prepared to share the data and make it accessible. However, it may not be appropriate to distribute a commercial product without the manufacturer's permission.*

5. **Judy asks for a guide to the data on the thumb drive (that is, what each variable name means). Is Brown Bear required to have a guide to the data available to other scientists?**
   - *Data is useless if you don't know what it represents. Brown Bear should have the guide to the data stored along with the data. Yes, it is his obligation to both store and share that information.*

©American Physiological Society 2017

Activity B
Critique Your Lab Notebook

Purpose

This activity will help students develop their understanding of what should be included in a lab notebook and how it should be presented to facilitate reader understanding and correct interpretation. After completing this activity, students will be able to identify key information on experiments, protocols, reagents and other experimental details that promote good data management and storage and reproducibility of data.

Objective

1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.

Procedure

Learning Cycle: Activity B should be done in the Explore phase. Instructor should review results during the Interactive Lecture.
HILA: Activity B should be done before the Interactive Lecture; Instructor will review during the lecture.

Ask students to bring their lab notebook or a portion of their notebook to class for this activity. This is a "Think/Pair/Share" activity. Students should work in pairs, exchange lab notebooks, and use the worksheet to determine whether each type of detail is included in his/her partner's notebook. Review what they learned about their notebooks and summarize class findings. Allow 10-15 minutes for the activity and discussion.

REMINDER: ***Encourage students to note ideas they want to add to their My Data Management and Integrity Checklist.***

©American Physiological Society 2017

Activity C
Critique a Poor Example

Purpose

This activity will allow students to identify poor examples of data presentation in journal article and critique the aspects of the presentation that make them less effective. After completing this activity, students will have improved skills in assessing the quality of data reported in the literature.

Objective

3. Develop criteria for assessing quality of data and use those criteria to evaluate the quality of both data prepared by colleagues/collaborators and findings reported in the literature.

Procedure

Learning Cycle: Students complete Activity C as an Engage activity. Instructor should discuss their responses during the Interactive Lecture.
HILA: Students complete Activity C worksheet before coming to class and review their answers during the Interactive Lecture.

Encourage students to look for examples of poor data presentation in the journals they normally read in their research area. It is less useful if they seek out journals of lesser quality to find these examples.

Also ask students to read:
Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm, Tracey L. Weissgerber, Natasa M. Milic, Stacey J. Winham, Vesna D. Garovic, *PLOS Biology*, April 22, 2015. DOI:10.1371/journal.pbio.1002128 (http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128).

©American Physiological Society 2017

Activity D
Evaluating Journal Guidelines on Data Presentation

Purpose

This activity will allow students to identify and compare data presentation policies in journals. After completing this activity, students will be able to evaluate the journal guidelines on data presentation and recognize the major expectations for best practice.

Objective

2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.

Procedure

Learning Cycle: Students complete the Activity D worksheet as an Explore activity. Instructor discusses their responses during the Interactive Lecture.
HILA: Students complete the Activity D worksheet before coming to class and review their answers during the Interactive Lecture.

Prior to attending the class, each student should identify three journals to which they are likely to submit in the future. They should visit each journal website and read the data presentation guidelines. They also should PRINT OUT the data presentation guidelines and bring them to class. Each student should summarize the journal data presentation guidelines on the worksheet.

REMINDER: ***Encourage students to note ideas they want to add to their My Data Management and Integrity Checklist.***

©American Physiological Society 2017

Activity E
Two-Minute Challenges

Purpose

These case studies will help students to think about data management in the laboratory and to apply data management and presentation best practices to common scenarios. After completing this activity, students will be able to able to evaluate common scenarios and determine how best to address concerns regarding data management and presentation.

Objectives

1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.
2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.

Procedure

Learning Cycle: Activity E should be done in the Explore phase either as homework or in class.

HILA: Activity E should be done during the Interactive Lecture. Can be done as homework before or after the lecture, if preferred.

If done as homework, answers should be discussed in class. If done as a whole class activity, the instructor should introduce the topic by reviewing the questions to be considered and then reading the scenario. Participants should answer the questions by voting yes or no (use voting cards, if preferred). Areas in which responses differ should be explored further. Participants should explain their reasoning.

Answers for the Instructor are provided in *italics*.

Scenario

Norah (Lab PI): "We just received an email from the journal that is reviewing our manuscript on the peanut study. Apparently, one of the reviewers is concerned that the antibody we used to detect the novel peanut variant is non-specific and he or she does not believe our results. The reviewer wants to see the validation data for the antibody study before making a decision. This request comes at a bad time, as Suzanna is not in email contact while she is on her safari vacation (her PhD graduation gift). Maria, can you find Suzanna's data? It should be in one of those five boxes on the shelf above her old lab bench.

©American Physiological Society 2017

Challenge 1: Stored Data

If you were Maria, what questions would you ask Norah before looking through all of the boxes?

- *Do you know if Suzanna did an antibody validation study?*
- *Do you remember when? Or any specific details about the studies that could help me narrow down where to look?*
- *Is there a list for each box noting what data are stored in them?*
- *Did Suzanna save the files someplace else? On a computer server?*

Challenge 2: Storage System

If you were Norah, what systems would you have in place so that such a request from a reviewer would be easy to answer?

- *I would set up a recordkeeping system for the lab so that every experiment run in the lab is well-labeled, saved appropriately based upon the type of data, and easily discoverable by anyone in the lab.*
- *I would keep a record of the reagents used in the lab along with information about how/whether they have been validated/verified/guaranteed.*
- *I would flag all data related to a manuscript in our lab recordkeeping system and have the first author/lead authors go through the data with me to be sure that I am ready to answer any questions that may arise.*

Challenge 3: Reviewer Questions

If you were the reviewer, what experimental details regarding the use of an antibody to detect a protein variant might you want to see reported in the manuscript?

- *Source of the antibody with catalog number*
- *Details about the experimental protocol*
- *Results/explanation about positive and negative controls used to validate specificity of the antibody*
- *Uncropped versions of the Western blots for reviewers to visually see specificity of the antibody*

REMINDER: ***Encourage students to note ideas they want to add to their My Data Management and Integrity Checklist.***

©American Physiological Society 2017

Activity F
Best Practices for Legends, Tables, Graphs and Images

Purpose

This activity will help students develop their understanding of professional standards of practice for data presentation (graphs, digital figures) and how to decide which data presentation type (text, table, graphic, etc.) is best for each situation. It also will help students understand the important role that legends play in data presentation and the information that should be included in legends. They also will be able to describe how digital figures should and should not be edited for use in publications. After completing this activity, students will be able to list key elements of the legend that are needed to effectively describe the experimental design and subsequent modifications. They will be able to identify and justify why data should be presented in a specific format (table, text, graph, etc.). And they will be able to detail how data presentation can affect interpretation of results.

Objective

2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.

Procedure

Learning Cycle: Activity F should be done in the Elaborate phase; Instructor should review during the class.
HILA: Activity F should be done after the Interactive Lecture; Instructor will review during class.

This is a small group activity. Students should work in groups of 3-5, and use the worksheet to guide their discussions and answers. Instructor should discuss answers after groups have time to work. Allow 25-30 minutes for the activity and discussion.

Answers for the Instructor are provided in *italics*.

©American Physiological Society 2017

Part 1: Legends

In small groups, students should review the figures WITHOUT reading the legends. They should discuss whether the results can be interpreted without the information in the legend. They should list questions they have about the figure. THEN they will read the legend and see if enough information is included to address their questions.

Example 1-1: Sarcomere structure

The images in panel A are labeled and large enough to compare the structures based upon the treatment. The images do not have scale bars so I assume that all images were captured at the same magnification but it could make it difficult to compare the images to other figures in other papers. The graph in panel B is well labeled (x vs. y axis). The box-whisker plots appear to show some variability in sample #'s per treatment. It would be nice to have the number of samples/animals treated noted on the graph.

Example 1-2: Daily body weight change

The y-axis range is quite specific (225-300) and the various treatments may not be as different as they appear, if the scale range was a bit broader. It is difficult to see the six treatment groups on the line graph, particularly the Day 1 treatment group. The number of animals in each group is not listed. Some days appear to be statistically significant (*) but there is no statement of level of significance.

Example 1-3: Change in the accuracy…

The y-axis represents % change. There is no way to know whether the % change is really significant without knowing the absolute numbers. The error bars are quite large for all treatment groups suggesting that there may be little difference between treatments. Numbers of samples per group are not listed on the graph but would be helpful for interpreting whether the changes presented are appropriate.

Part 2: Table or Graph?

In small groups, students should review the raw data provided and discuss how it should be presented in a manuscript (table, line graph, bar graph). Then groups should share with the instructor their decision and reasoning.

Example 2-1: Growth factor

These data most likely are best presented as a bar graph, with paired bars with and without the risk factor for each subject group. The important comparisons are between the various subjects, and for the effect of the risk factors. The absolute data values are less important, suggesting a table is not needed. However, the substantial differences in the scale for the two growth factors suggest that either two bar graphs should be used, or the data should be transformed (or plotted using two y-axes) so as not to obscure the smaller values.

Example 2-2: Fibroblast growth

At first glance, these time-course data suggest that a line graph would be the most appropriate

©American Physiological Society 2017

choice. However, with only three time points, and unequal spacing between the times at which measurements were made, a line graph is unlikely to be aesthetically pleasing. A bar graph can adequately depict the data and clearly show the effect of substance A over time (as well as its statistical significance, if any).

Example 2-3: Normalized Na current

This is a classic current/voltage dataset showing sodium currents in control cells and those treated with a stimulus, and is traditionally depicted as a line graph. However, without even knowing about this area of research, features of the data point to the best way to depict it. First, the values of the independent variable, voltage, are regularly spaced. Second, there is a clear progression of the current values with each change in voltage under both conditions. In fact, it is the slope and shape of the two plots that tells you something about the effect of A23187. A bar graph would not illuminate this as clearly, and a table would not show the relationship at all.

Example 2-4: Adipose tissue weight

This is a complex dataset that details various independent variables. Because these parameters are unrelated, a table would save space. However, if the point of the study was to illustrate the effect of the treatment to influence the various parameters in the obese subjects, this could be convincingly displayed with a series of bar graphs showing the impact of obesity compared to normal controls, and the tendency of the treatment to reverse that impact.

Part 3: Digital Details

In small groups, students should assess the digital images presented in these figures and discuss whether the data are presented clearly. They should develop suggestions on what could be improved.

Example 3-1: Movat Pentachrome-stained cross-sections

Figure elements presented well: You are visually able to see changes between controls over time in two types of mice. The author notes how many animals were in each group and includes the scale bar (100µM) for the images. It is important that all photos are comparable. The author also uses arrows to direct the reader's attention to the results.
Suggestions for improvement: Perhaps the figure could have included an inset image that showed a magnified view of a particular area of the artery.

Example 3-2: CIS treatment after Mg deficiency

Figure elements presented well: The authors provide two magnifications for the images. The legend description is detailed. Arrows are used to direct reader's attention to the findings.
Suggestions for improvement: There is no information on how many samples were evaluated. Scale bars are not included on the images.

Example 3-3: Chronic hypoxia upregulates VEGF receptors

Figure elements presented well: The authors show the entire Western blots making it easy to see

the specificity of the antibodies. The legend is very detailed and describes the analysis. It includes standards (set of protein controls) to compare expression levels.
Suggestions for improvement: The legend contains quite a bit of detail. Some could have been reported in the results section. Relative expression is reported in a bar graph but, in this case, this is not a bad choice because one can see the amount of expression detected in the images.

Example 3-4: Cyclooxygenase
Figure elements presented well: The authors provided a visual image to go along with the relative expression bar graph.
Suggestions for improvement: Controls are set to 1; note how "1" is not the same for all proteins. Some are poorly expressed while others are highly expressed. Also, the images are in pieces. Thus, one cannot know if the results are from one experiment and can be directly compared. There is also no information on sample numbers. In addition, loading controls for the protein expression is not presented.

REMINDER: *Encourage students to note ideas they want to add to their My Data Management and Integrity Checklist.*

©American Physiological Society 2017

Activity G
My Data Management and Integrity Checklist

Purpose

Students will develop a checklist based on course material that they can use now and in the future to guide ethical text preparation in terms of data management and integrity. They should use materials from the activities, readings, and Interactive Lecture. After completing the activity, students should have a checklist for data management best practices AND a plan for teaching these best practices to their students.

Objectives

1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.
2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.
3. Develop criteria for assessing quality of data and use those criteria to evaluate the quality of both data prepared by colleagues/collaborators and findings reported in the literature.
4. Formulate best practices for presenting digital and non-digital data to enhance clarity and avoid inappropriate manipulation via figures and legends.

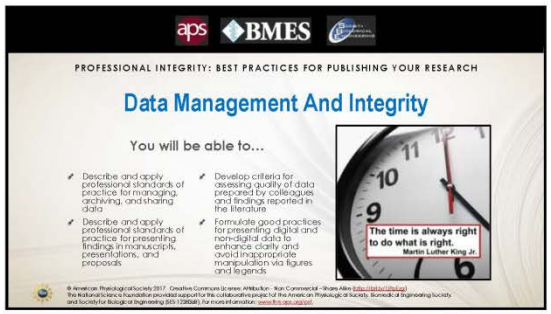
Procedure

Learning Cycle: Complete in the Evaluate phase. Students should do this individually but will want to share their lists in class or with the instructor.
HILA: Should be done after the Interactive Lecture. Students should do this individually but will want to share their lists in class or with the instructor

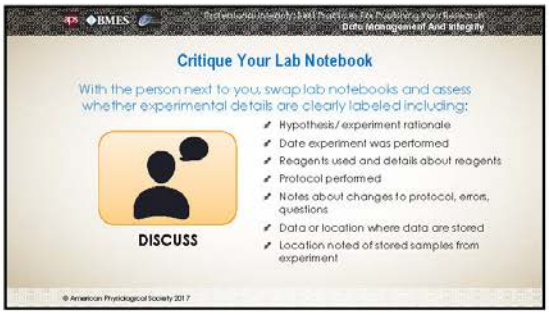
Students should look back at their previous work, readings, and the Interactive Lecture. They should work individually but it may be worthwhile for students to trade plans and review them in teams. Be sure to review each student's plan and provide feedback.

©American Physiological Society 2017

Presentation Slides

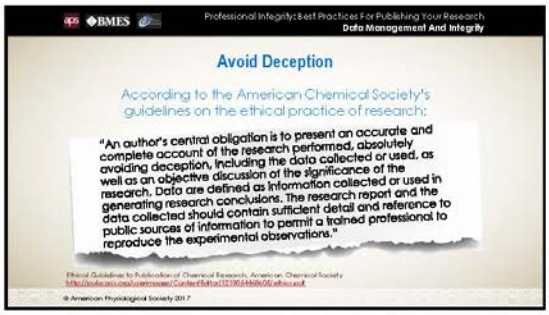


©American Physiological Society 2017

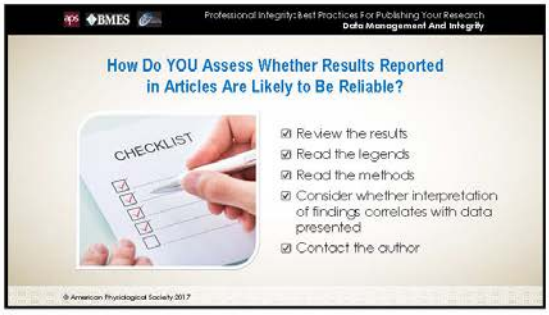

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Critique Your Lab Notebook
With the person next to you, swap lab notebooks and assess whether experimental details are clearly labeled including:
DISCUSS
Hypothesis/ experiment rationale
Date experiment was performed
Reagents used and details about reagents
Protocol performed
Notes about changes to protocol, errors, questions
Data or location where data are stored
Location noted of stored samples from experiment
© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
There Is an Art and Science in Turning Raw Data into Polished Figures
© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Data Presentation Integrity
Readers trust that experiments were performed as described, and interpreted in an unbiased manner
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Avoid Deception
According to the American Chemical Society's guidelines on the ethical practice of research:
"An author's central obligation is to present an accurate and complete account of the research performed, absolutely avoiding deception, including the data collected or used, as well as an objective discussion of the significance of the research. Data are defined as information collected or used in generating research conclusions. The research report and the data collected should contain sufficient detail and reference to public sources of information to permit a trained professional to reproduce the experimental observations."
© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
What Steps Do You Take to Prepare Good Data in Your Lab?
Review the literature
Make a hypothesis
Establish a protocol
Determine sample size
Define method to analyze results including statistics
Repeat experiments
Analyze results
Troubleshoot problems
Discuss findings with colleagues
Validate reagents
Log process, changes, results, and analyses
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
How Do YOU Assess Whether Results Reported in Articles Are Likely to Be Reliable?
CHECKLIST
Review the results
Read the legends
Read the methods
Consider whether interpretation of findings correlates with data presented
Contact the author
© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

What Else Should Be Considered When Assessing the Quality and Validity of Data Presented in a Scholarly Article?

- Were experiments performed blinded?
- Were basic experiments repeated?
- Were all results presented?
- Were there positive and negative controls?
- Were reagents validated?
- Were statistical tests appropriate?

C.G. Begley, Six red flags for suspect work, Nature, 23 May 2013, Vol 497, pg 433

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

"The results show what?"

Share examples of poorly presented figures with the class.

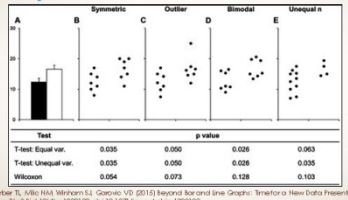
DISCUSS

- Explain why the presentation is not ideal.
- What could be corrected?

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Fig 1. Many different datasets can lead to the same bar graph.



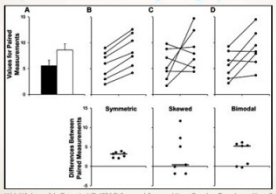
Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128 http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Fig 2. Additional problems with using bar graphs to show paired data.

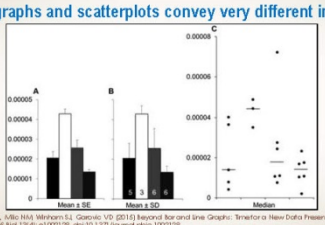


Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128 http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Fig 3. Bar graphs and scatterplots convey very different information.



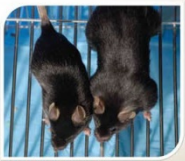
Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128 http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Compare 3 Experiments

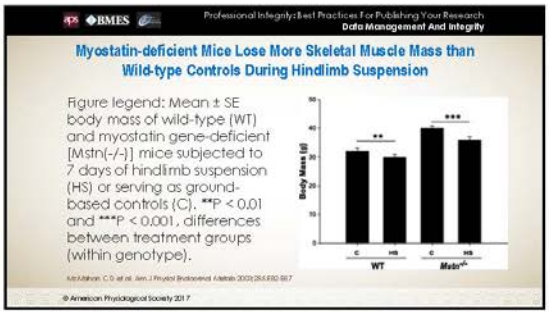
- These experiments all use hindlimb suspension for various periods of time to assess changes in body weight and muscles mass of the animals.
- Two experiments analyze Myostatin knockout mice (Mstn-/-). Myostatin is a protein that regulates muscle growth. Without myostatin, muscles grow much larger than normal.

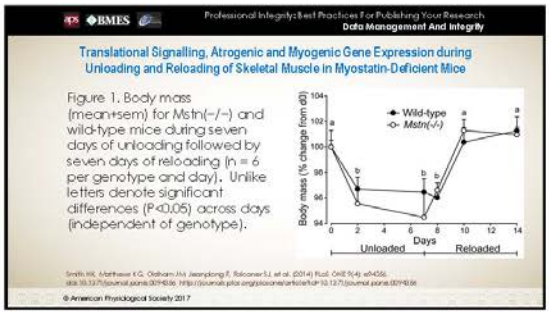


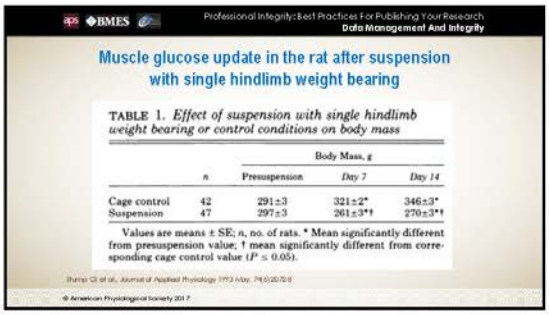
Lee S-J (2007) Quadrupling Muscle Mass in Mice by Targeting TGF-β Signaling Pathways. PLoS ONE 2(9): e789. doi:10.1371/journal.pone.0000789. Used with permission per Creative Commons Attribution 2.5 Generic license.

© American Physiological Society 2017

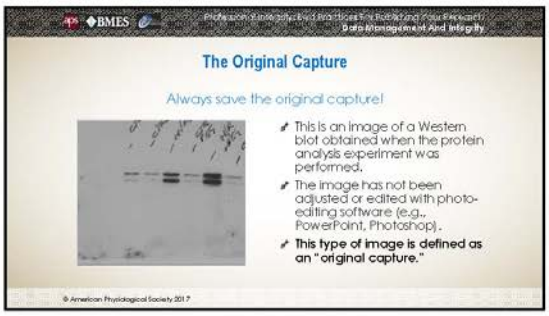
©American Physiological Society 2017

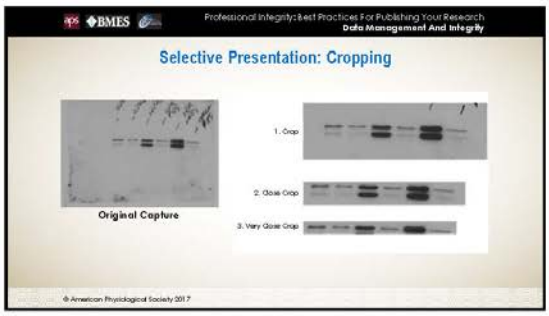

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Myostatin-deficient Mice Lose More Skeletal Muscle Mass than Wild-type Controls During Hindlimb Suspension
Figure legend: Mean ± SE body mass of wild-type (WT) and myostatin gene-deficient [Mstn(-/-)] mice subjected to 7 days of hindlimb suspension (HS) or serving as ground-based controls (C). **P < 0.01 and ***P < 0.001, differences between treatment groups (within genotype).
Body Mass (g)
C
HS
WT
Mstn-/-
© American Physiological Society 2017

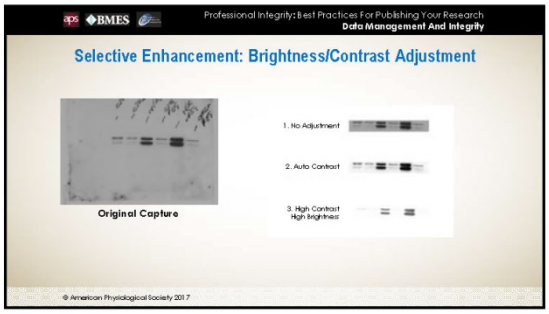

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Translational Signalling, Atrogenic and Myogenic Gene Expression during Unloading and Reloading of Skeletal Muscle in Myostatin-Deficient Mice
Figure 1. Body mass (mean±sem) for Mstn(−/−) and wild-type mice during seven days of unloading followed by seven days of reloading (n = 6 per genotype and day). Unlike letters denote significant differences (P<0.05) across days (independent of genotype).
Wild-type
Mstn(-/-)
Body mass (% change from d0)
Days
Unloaded
Reloaded
© American Physiological Society 2017

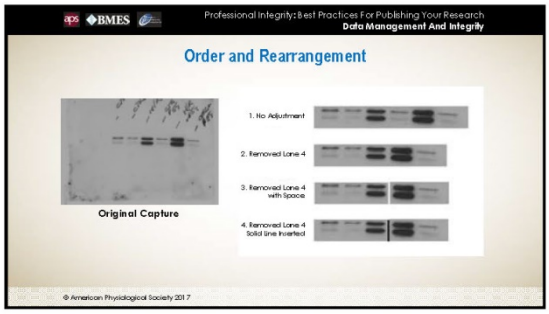

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Muscle glucose update in the rat after suspension with single hindlimb weight bearing
TABLE 1. Effect of suspension with single hindlimb weight bearing or control conditions on body mass
Body Mass, g
n
Presuspension
Day 7
Day 14
Cage control 42 291±3 321±2* 346±3*
Suspension 47 297±3 261±3*† 270±3*†
Values are means ± SE; n, no. of rats. * Mean significantly different from presuspension value; † mean significantly different from corresponding cage control value (P ≤ 0.05).
© American Physiological Society 2017

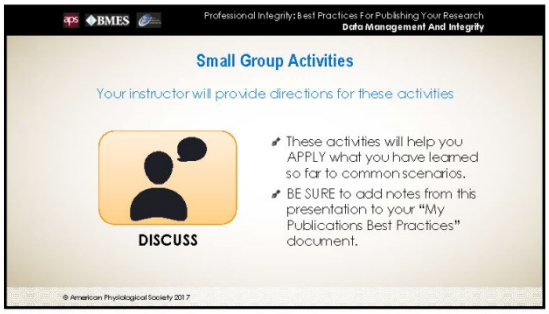
Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Digital Data Presentation Guidelines
Does your favorite journal provide data presentation guidelines? Share one of the guidelines with the class.
DISCUSS
Selective Presentation
Selective Enhancement
Order and Rearrangement
© American Physiological Society 2017

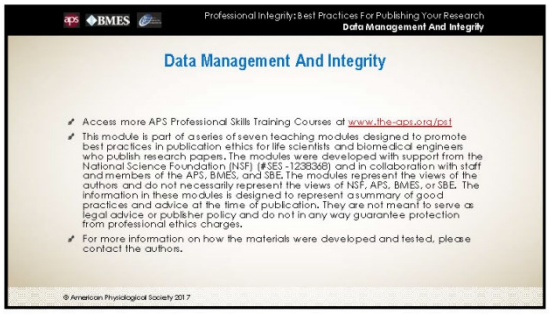

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
The Original Capture
Always save the original capture!
This is an image of a Western blot obtained when the protein analysis experiment was performed.
The image has not been adjusted or edited with photo-editing software (e.g., PowerPoint, Photoshop).
This type of image is defined as an "original capture."
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Selective Presentation: Cropping
Original Capture
1. Crop
2. Close Crop
3. Very Close Crop
© American Physiological Society 2017









 ©American Physiological Society 2017

Presentation Slide Text

Slide #	Text
1	This presentation is part of the professional skills training series on professional integrity best practices for publishing your research. Today we will review best practices for Data Management and Integrity. This presentation will help you to: • **Describe and apply** professional standards of practice for managing, archiving, and sharing data • **Describe and apply** professional standards of practice for presenting findings in manuscripts, presentations, and proposals • **Develop** criteria for assessing quality of data prepared by colleagues and findings reported in the literature; and • **Formulate** best practices for presenting digital and non-digital data to enhance clarity and avoid inappropriate manipulation via figures and legends
2	Have you ever tried to repeat an experiment that was reported in the literature? Have you ever been unable to get the same results as reported in the literature? It is quite frustrating to discover that the information provided in a manuscript is not sufficient to repeat the work. So, what do you do? If you completed the assignment in Activity A for homework, the video depicting a dialogue between an author and a reader showed just how frustrating it can be to receive clarification on the methods and data reported in a research article. Ideally, complete experimental methodology and findings should be included in your research articles. If the information does not fit in the manuscript itself, it should be readily available to readers who inquire. At this time, if you have not completed Activity A, please pause the presentation to do so.
3	[No text for this slide]
4	While publishing interesting results is important, it is absolutely critical that you have the original data that led to the results available for you, your co-authors, colleagues, or journal editors to assess. Yes, even journals may ask you to upload raw data as part of the submission process or they may ask to see the raw data during review. Thus, you have to have a well-organized method for saving your data. One that can be easily interpreted by your advisor, a lab mate or someone unfamiliar with your work.

©American Physiological Society 2017

	Think about it: Would your advisor be able to readily locate a piece of your data if you had to instruct him or her over the phone or by email? If not, now is the time to set up a record keeping process. At a minimum, • **Label your data clearly** including the date of the experiment, the lanes or samples tested, and the treatments or variables given. • **Save all versions of your data**. If you modify the data, save the original and the modified version with understandable file names and descriptions. • **Note the data that are used in a manuscript or public presentation.** These pieces of data should be easily located and saved in multiple places. Again these data should be clearly labeled and have descriptions. Does your lab have a data storage policy? Did you receive clear instructions or training on that data policy? If your lab does not have a data storage policy, take the initiative to meet with your advisor and establish a lab policy.
5	Grant funders like NSF and NIH require grantees to save all components related to a grant for at least three years. They define the three years as starting after the final report is submitted. Since you may not know when this occurs, it is reasonable to save raw data and data used in publications for even longer, likely 5 to 6 years, which may be longer than your time in the lab group. Papers, notebooks, and computer files should be saved. In particular, computer files should be saved on BOTH your institution's servers and on external hard-drives. Lastly, data are owned by the institution. Data do NOT go with you when you leave the institution. You may make copies, but the originals should stay, clearly labeled, and accessible.
6	Let's stop for a moment to assess your lab notebook and answer the question, "Could your advisor or colleague find data in your lab notebook?" If you have not already done so, pause the presentation to complete the worksheet for Activity B. This activity will help you assess whether critical details, such as those listed on the slide, are included in your lab notebook. Ideally, you should do this activity with a colleague. Trade notebooks and do the assessment on each other's notebooks. Lab notebooks should be a detailed record of the experiments performed. The more details they include, the more likely you will be able to understand what you did months or even years from now. Moreover, it should be legible and clear, so that any

©American Physiological Society 2017

	colleague can understand what is reported. If your colleague cannot find this information in your notebook in this exercise, then others will not be able to find it either. To complete Activity B, pause the presentation now.
7	While data storage practices are absolutely critical, the raw data that are analyzed and presented as results in the figures of a published manuscript are often the most exposed and evaluated component of your work. Each figure may represent many experiments. Thus, how you analyze and present your raw data as a final, polished result is an aspect of research that takes a lot of science and a little bit of art. We will focus the rest of this session on data presentation integrity.
8	When readers assess the results reported in your paper they have to assume that the experiments were performed, described, and interpreted in an unbiased manner. They trust that you are reporting the data honestly and completely. While authors want to present their best work in manuscripts, they must be careful not to make their results more impressive than they actually are.
9	We, as researchers should not create the illusion of good data by manipulating poor results. As noted in the American Chemical Society's guidelines on the ethical practice of research, "An author's central obligation is to present an accurate and complete account of the research performed, **absolutely avoiding deception**, including the data collected or used, as well as an objective discussion of the significance of the research. Data are defined as information collected or used in generating research conclusions. The research report and the data collected should contain sufficient detail and reference to public sources of information to permit a trained professional to reproduce the experimental observations." These are very serious words. While the vast majority of researchers have no intention to deceive the public, some data preparation practices can promote inappropriate data enhancement. We all have to be careful about how we prepare our results for presentation in publications as well as in grant applications, at meetings, and any other public venues.
10	Regardless of your area of expertise, the general process for generating data are likely similar. What are some of the steps that you take before starting an experiment?

©American Physiological Society 2017

	Do you: review the literature, establish a protocol, determine sample size, or consult with a statistician? What about during an experiment? Do you: validate reagents, keep detailed notes and record all results, follow the protocol, and avoid distractions? What about after an experiment? Do you: analyze results, troubleshoot problems, repeat experiments, and discuss findings with colleagues? If your lab is an expert on a certain technique, readers will ask you for your protocol. By providing more details, you increase the likelihood that your work will be repeated and promoted.
11	You know your criteria for good results in your research area but how you do assess whether results reported in a manuscript are valid? One can review the results, read the legends, read the methods, and consider whether the interpretation of the data agrees with the data presented. Are there other ways to assess the validity of published data? You could email the authors to ask for more information, share the work with colleagues, or discuss the findings in journal clubs. Remember that one of the strengths of the scientific enterprise is the ongoing review and discussion of research findings among scientists.
12	When you are reading an article that is in your field, you are likely to be much more critical of the findings reported because you know what type of information the author should share with the reader. What happens when you read a paper that is not in your area of expertise? What do you assess to help you determine whether the study was performed well? The following are a few criteria that could help to determine whether the information presented is adequate for interpreting the results. • Were experiments performed blinded? • Were the basic experiments repeated? • Were all results presented? • Were there positive and negative controls? • Were reagents validated? and • Were statistical tests appropriate?
13	One of your homework assignments for Activity C was to find examples of poorly

©American Physiological Society 2017

	presented experiments. Sometimes seeing poorly designed or poorly presented figures helps you determine what aspects make you trust one study more than another. At this time, work individually or in groups, to review your selected figures and note what could be corrected. Pause the presentation to complete the activity if you have not already done so.
14	The article "Beyond Bar and Line Graphs" discusses how data presented in graphs can be misleading. Graphs often present a summary of multiple data points or observations; therefore, the data behind the graphs usually are not visible. Pause a moment to look at Figure 1 from the paper. . This figure shows how a bar graph can hide the range of results that were collected. Panel C shows possible outliers. Panel D shows that the results may be bimodal. Panel E shows that numbers of samples are not equal.
15	Now pause to look at Figure 2. Additional problems can arise when using bar graphs to show paired data. While the bar graph suggests a difference in the two groups of measurements, the panels show how the same bar graph could be generated from very different sets of data. Clearly, in Panels C and D, not all pairs responded similarly. In fact, the data in panel D suggests that there may be two unique populations, responders and non-responders.
16	Now pause to look at Figure 3. Here the bar graphs and scatterplots convey very different information. These examples should help you consider whether your data are being presented as clearly as possible. Try different presentation approaches and get feedback from other colleagues as well as your co-authors. How do THEY interpret the different graphical presentations?
17	Clear labeling of data is also very important. Next, we will go through a few examples of figures that report similar findings. Each study assesses whether animals change body weight due to muscle loss after having their hind-limbs suspended for a period of time. Two of the studies use myo-statin knockout mice. Animals, and humans, without myo-statin protein expression grow very large muscles.
18	The first example compares the body weight, on the y axis, of four groups of mice: wild type mice and myo-statin knockout mice both with and without their hind-limbs being suspended, as noted on the x-axis. Pause the video to look at this figure and its legend

©American Physiological Society 2017

	and consider what information the figure shared that is useful to the reader and whether there is information that is missing. What information does this figure share that is useful to the reader? First, the body mass is clearly displayed. Also, the description of the experiment, as well as the details of the statistics used, is noted in the legend. However, neither the figure nor the legend indicates how many animals were evaluated. Without that information it is hard to know whether the results are derived from a meaningful sample size. Would you believe the findings if the results were from 1 or 2 animals per group? What if the results were derived from 6 or more animals?
19	The next example uses control mice and myo-statin knockout mice to assess changes in body weight, both during tail suspension and after tail suspension. Pause the video to look at this figure and its legend and consider what information the figure shared that is useful to the reader and whether there is information that is missing. In this figure, the x-axis is clearly labeled with a description of the experiment variables and the changes in weight over time are easy to see. The legend describes the experimental design including the number of animals analyzed. However, the graph does not report the absolute body weights for the animals. Rather the graph shows percent change in body mass. With this presentation, the reader cannot identify absolute changes. We can only infer relative changes. In addition, the scale bars on the y-axis range from 94% to 104%, which may alter your impression of the change in body mass, particularly if you expected the scale to start at zero.
20	This last example analyzes rats. One set was free to walk in the cage. The other set had their hind-limbs suspended for 7 and 14 days. Pause the video to look at this table and its legend and consider what information the table shared that is useful to the reader and whether there is information that is missing. In this presentation, the use of a table, rather than a graph, allows the information to be easily read. However, using a table makes it more difficult to visualize differences between the two experimental variables. In summary, while the three examples are similar, they all shared their data with the reader differently. It is worth taking the time to consider how you want to share your results in order to convey the most important information to the reader.
21	Data can also be captured digitally via photographs of results in real time. Often these

©American Physiological Society 2017

	provide stunning images that help to visualize the results reported. Journals have developed guidelines for presenting digital data. The guidelines may include information on selective presentation, selective enhancement, and rearrangement. Activity D asks you to find the data presentation guidelines for your favorite journal. If you have not already done so, review the guidelines that you found. Work individually, or in pairs, to compare and contrast the guidelines. At this time, pause the presentation to complete the activity.
22	The next few slides contain examples of image adjustments that help explain why the data presentation guidelines are a part of the journal's information for authors. First, let's start with an image of a Western blot. The image has not been adjusted or edited with photo-editing software, including PowerPoint or Photoshop. This type of image is defined as an "original capture." As a best practice, always save, clearly label, and securely store the original captures of your data.
23	In this example, the original capture has been edited by cropping the image in several ways. Pause the video to look at this image. Which images do you think were appropriately cropped? [INSERT 3 SECOND PAUSE]. **What are the benefits and drawbacks of cropping a large image?** Cropping an image makes better use of the figure space in the journal. Readers only need to see the relevant portions of the image that contain the data, not empty space. However, a very close crop may remove important information related to the specificity of the antibody. **What features in the image should remain after cropping?** The relevant data should remain along with space above and below the signal after cropping the image. In this case, it is best to keep both rows visible. Even if the data are supposed to be the upper band, it may be that the lower band has value to readers. Some journals have guidelines regarding how much space above and below a signal must remain visible. **Which images were appropriately cropped**? Numbers 1 and 2 were appropriately cropped. In contrast, Number 3 removes the lower band, and that may not be appropriate. This of course assumes that one actually "knows" that the protein of interest is represented by the top band and not the

	bottom band. There are instances where antibodies cross-react more strongly with irrelevant proteins and the band of interest is actually lighter in density.
24	In the next example, the contrast of the original capture has been edited. Pause the video to look at this image. In which images do you think the brightness and contrast were adjusted appropriately? **What are the benefits and drawbacks of adjusting brightness and contrast?** Contrast adjustment can highlight the important features of the image while reducing the background noise. It can help emphasize the data of interest. **What features in the images should remain after brightness and contrast adjustment?** All of the features in the image should remain. The adjustment should not completely remove any elements of the images, including the background. The adjustment should be applied to the entire image, not a specific feature or area. **Which images have an appropriate use of brightness and contrast adjustment?** Numbers 1 and 2 have appropriate contrast adjustment. Number 3 is not appropriate because some of the lanes have been removed by the extreme contrast adjustment.
25	In this last example, the arrangement of the lanes in the original capture has been edited. Pause the video to look at this image. In which images do you think the rearrangements were appropriately done? **What are the benefits and drawbacks of rearranging images?** Image rearrangement can make interpretation of relevant data easier if irrelevant lanes are removed or the lanes of interest are presented in an order that makes the most sense. **What features in the images should remain after rearrangement?** All of the features should remain in an image that has been rearranged. If a portion of the image has been rearranged, removed, or added, it should be noted in the legend. **Which images show appropriate image rearrangement?** Numbers 3 and 4 show appropriate image rearrangement because the images make it clear that a lane has been removed. Image number 2 is problematic because it gives the impression that the experiment was run in the order presented, which is not true. **How could figure legends help explain the rearrangements in images Number 3 and Number 4?** The legend should include a statement such as “All samples were run on the same gel,

©American Physiological Society 2017

	but an irrelevant lane was removed." This would make it clear to the reader that the original image has been modified. In summary, these three examples show that while some modification of images is reasonable, it is important to keep adjustments to a minimum so that the image that appears in the publication still conveys the same information as the original capture.
26	The remaining activities in this module will help you to APPLY what you have learned so far to common scenarios and to your own work. BE SURE to add notes from this presentation to your "My Checklist" document.
27	Thank you for listening to this presentation. To access more information about APS Professional Skills Training Courses, visit www.the-aps.org/pst.

References and Resources

1. Begley, C.G., Six red flags for suspect work, *Nature*, *Vol 497*, pg. 433, 23 May 2013.
2. Lorsh, J.R., Collins, F.S. and Lippincott-Schwartz, J., Cell Biology: Fixing problems with cell lines, *Science*, 346(6216):1452-3, 19 Dec 2014. doi: 10.1126/science.1259110
3. National Institutes of Health. Video vignettes and resources about data reproducibility. Retrieved from: http://www.nih.gov/science/reproducibility/training.htm.
4. Rennie D. and Gunsalus C.K. (2015, June18) "If you think it's rude to ask to look at your co-authors' data, you're not doing science": Guest post, Retraction Watch. Retrieved from: http://retractionwatch.com/2015/06/18/if-you-think-its-rude-to-ask-to-look-at-your-co-authors-data-youre-not-doing-science-guest-post/
5. Weissgerber, T.L., Milic, N.M., Winham, S.J. and Garovic, V.D. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm, *PLOS Biology*, 2015 April 22. doi:10.1371/journal.pbio.1002128

Course Resources

Each of the **Professional Skills Training Courses on Best Practices for Publishing Your Research** has multiple resources to accompany the Instructor Guide. All of the following resources are available at www.the-aps.org/pst/ethics.

1. PowerPoint (.ppt) files for the Interactive Lecture. These slides are editable.
2. Instructor and Student Guides are available as editable .doc files.
3. Request form for assessment tools (quizzes and key).
4. Links to video versions of the Interactive Lecture on YouTube.
5. Links to online, on demand version of the module.

Publication Ethics Community

In addition, APS hosts a Publication Ethics Community on the Life Science Teaching Resource Community. The community posts ethics cases for comment by participants and experts. See www.lifescitrc.org and click on My Community.

Ethics CORE (Collaborative Online Resource Environment)

This website is coordinated by the National Center for Professional and Research Ethics. The site provides resources for Responsible Conduct of Research courses and seeks to create communities of responsible research and professional practice. It is an excellent source of case studies, simulations, role-play scenarios, videos, and lectures. See https://nationalethicscenter.org.

We welcome your questions and feedback on these materials.
Email us at education@the-aps.org.

©American Physiological Society 2017

Data Management and Integrity Module

Student Handouts

These activities will help you:

1. Describe and apply the professional standards of practice (that is, the expectations of the research community) for managing, archiving, and sharing data.
2. Describe and apply the professional standards of practice (that is, the expectations of the research community) for presenting findings and conclusions within manuscripts, presentations, and proposals.
3. Develop criteria for assessing quality of data and use those criteria to evaluate the quality of both data prepared by colleagues/collaborators and findings reported in the literature.
4. Formulate best practices for presenting digital and non-digital data to enhance clarity and avoid inappropriate manipulation via figures and legends.

This module is part of the series, *"Professional Integrity: Best Practices for Publishing Your Research"* developed by:
American Physiological Society www.the-aps.org
Biomedical Engineering Society www.bmes.org
Society for Biological Engineering www.aiche.org/sbe

For information on the other modules or to take an online, interactive version of one or more modules, go to www.the-aps.org/pst.

©American Physiological Society 2017

About Your Publication Ethics Checklists

In these modules, you will be encouraged to create your OWN checklists for preparing manuscripts using ethical and professional standards of practice for researchers.

WRITE

Why do I need a checklist?

As your training progresses, your research and writing skills develop along with your knowledge of the field, your professional network, and your independence as a professional. This also means that understanding and following best practices for professional behavior, including research and publication ethics, increasingly rests on your shoulders. YOU become the person who is setting the standards for your laboratory group. YOU are the person who must establish protocols for assuring ethical behavior. And YOU are the person who has to teach standards and protocols to every trainee in your lab and, sometimes, to those with whom you collaborate. You cannot assume that they come with an understanding of best practice…you must inform, guide, and monitor their adherence to best practices.

What should I include in the checklist?

You are investing time and effort to learn best practice for publication ethics through this module (and possibly the other modules in this series). **This activity is the big "take away" from this module**. It is YOUR checklist of things to remember about publication ethics. In each module in this series, you will add a checklist of the things you want to remember from that module. You also will add notes on how you would teach this to your students in the future. For most modules, we encourage you to add three sections to your checklist:

1. **Definitions to Remember Table:** Consider adding the terms and definitions from the lecture. Also add the links for professional standards you want to access later (e.g., ICMJE criteria for authorship). Remember to add the source of your definition or text if you are copying it.
2. **My Best Practices Checklist:** What are the things you want to check as you develop or revise your manuscripts?
3. **PASS IT ON:** How will you teach this to YOUR trainees in the future? How will you share this with those with whom you collaborate?

When you are done with these modules, we encourage you to make a copy of your checklists and keep them handy for use as you develop manuscripts in the future.

©American Physiological Society 2017

Activity A
"May I Have the Data?"

Purpose This activity will help you understand best practice in storing original data so it is both accessible and understandable to the broader scientific community.

Procedure Watch this video: "Data Sharing and Management Snafu in 3 Short Acts" by Karen Hanson, Alisa Surkis, and Karen Yacobucci, NYU Health Sciences Libraries, August 3, 2012. https://www.youtube.com/watch?v=N2zK3sAtr-4

WATCH

The video is a simple animation of two bears (scientists) talking about a request to share published data. While the animation is childlike, the conversation is important. Answer the questions below and bring your answers to class.

"May I Have the Data? "

1. **Is Judy (the black and white panda bear) justified in asking Brown Bear for a copy of his original data? Why or why not?**

2. **Brown Bear has trouble remembering where he stored the original data. Is it his job to know where the data is stored?**

3. **Brown Bear sent Judy his thumb drive containing the ONLY copy of his original data to Judy. Would you have done that? Why or why not?**

4. **Accessing the data requires a specific program. Do you think Brown Bear is required to have a copy of that program available?**

5. **Judy asks for a guide to the data on the thumb drive (that is, what each variable name means). Is Brown Bear required to have a guide to the data available to other scientists?**

©American Physiological Society 2017

Activity B
Critique Your Lab Notebook

Purpose This activity will help you understand what should be included in a lab notebook and how it should be presented in order to facilitate reader understanding and correct interpretation. After completing this activity, you will be able to identify key information on experiments, protocols, reagents and other experimental details that should be documented in order to maintain best data management, storage, and sharing practices.

Procedure Bring your lab notebook or a portion of your notebook to class for this "Think/Pair/Share" activity.

You should work with a partner, exchange lab notebooks, and use the worksheet to determine whether each type of detail is included in his/her notebook. Be prepared to discuss with the class what you learned about your own notebook from your partner's review.

Does the lab note book contain…?	Yes/No	Comments
A hypothesis/experiment rationale		
A date the experiment was performed		
Details about reagents used		
Details about the protocols used		
Notes about changes to protocol, errors, and/or questions		
Notes about where data are stored		
Notes about where samples are stored		
References to prior published work		
Neat and legible writing		
Enough information to repeat the experiment		

©American Physiological Society 2017

Other:		
Other:		
Other:		
If some of the above information is not included in the lab notebook, ask the owner to explain where the information is saved. Please list the details and storage locations here.		
Are there details that have not been recorded in any location? If so, please list here.		

WRITE

REMEMBER: **Note ideas that you want to add to your My Data Management and Integrity Checklist.**

©American Physiological Society 2017

Activity C
Critique a Poor Example

Purpose In this activity, you will identify poor examples of data presentation in journal articles and critique the aspects of the data presentation that make them less effective. This activity can help you hone your skills in assessing the quality of data reported in the literature and in your own lab and collaborations.

Procedure Look for two examples of poor data presentation in the journals you normally read in your research area. For each, PRINT OUT the data presentation (figure, table, or text) with the legend (if available).

WRITE

Then write down why the presentation is not ideal and how it could be corrected. Bring your print outs of the data presentation and the worksheet below to class to share.

READ

Be sure to read:
Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm, Tracey L. Weissgerber, Natasa M. Milic, Stacey J. Winham, Vesna D. Garovic, *PLOS Biology*, April 22, 2015. DOI:10.1371/journal.pbio.1002128 (http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128).

Example 1 (write citation here)
1. What type of presentation is it (figure, table, text, etc.)?
2. Why is the presentation not ideal?
3. What would you do to correct it?

©American Physiological Society 2017

Example 2 (write citation here)
1. What type of presentation is it (figure, table, text, etc.)? **2.** Why is the presentation not ideal? **3.** What would you do to correct it?

WRITE

REMEMBER: **Note ideas that you want to add to your My Data Management and Integrity Checklist.**

Activity D
Evaluating Journal Guidelines on Data Presentation

Purpose This activity will allow you to identify and compare data presentation policies in journals. After completing this activity, you will be able to evaluate the journal guidelines on data presentation and recognize the major expectations for best practice.

Procedure Prior to attending class, identify three journals to which you are likely to submit manuscripts in the future. Visit each journal website and read the data presentation guidelines.

PRINT OUT the data presentation guidelines and bring them to class. Summarize the journal data presentation guidelines on the worksheet below.

In class, you will work with a partner to do the following:

1. Compare the journal guidelines provided by you and your partner.
2. Work together to summarize what you learned from the various journals to answer the question: "What are the major expectations from the journals regarding data presentation?" List the major expectations on your worksheet.

Journal 1 Name:

Please list the data presentation guidelines:

Journal 2 Name:

Please list the data presentation guidelines:

©American Physiological Society 2017

Journal 3 Name:

Please list the data presentation guidelines:

SUMMARY

What are the major expectations from the journals regarding data presentation?

WRITE

REMEMBER: Note ideas that you want to add to your My Data Management and Integrity Checklist.

Activity E
Two-Minute Challenges

Purpose These case studies will help you think about data management in the laboratory and apply data management and presentation best practices to common scenarios. After completing this activity, you will be able to evaluate common scenarios and determine how best to address concerns regarding data management and presentation.

Procedure Read the scenario and write your answers for each Challenge.

DISCUSS

You will either share these in a small group or with the whole class. BE SURE to explain your reasoning for each answer (that is, WHY do you think that?)

Scenario

Norah (Lab PI): "We just received an email from the journal that is reviewing our manuscript on the peanut study. Apparently, one of the reviewers is concerned that the antibody we used to detect the novel peanut variant is non-specific and he or she does not believe our results. The reviewer wants to see the validation data for the antibody study before making a decision. This request comes at a bad time, as Suzanna is not in email contact while she is on her safari vacation (her PhD graduation gift). Maria, can you find Suzanna's data? It should be in one of those five boxes on the shelf above her old lab bench.

Challenge 1: Stored Data

If you were Maria, what questions would you ask Norah before looking through all of the boxes?

©American Physiological Society 2017

Challenge 2: Storage System

If you were Norah, what systems would you have in place so that such a request from a reviewer would be easy to answer?

Challenge 3: Reviewer Questions

If you were the reviewer, what experimental details regarding the use of an antibody to detect a protein variant might you want to see reported in the manuscript?

WRITE

REMEMBER: Note ideas that you want to add to your My Data Management and Integrity Checklist.

©American Physiological Society 2017

Activity F
Best Practices for Legends, Tables, Graphs and Images

Purpose

This activity will help you understand best practice for data presentation (graphs, digital figures) and how to decide which data presentation type (text, table, graphic, etc.) is best for each situation. It also will help you understand the important role that legends play in data presentation and the information that should be included in legends. After completing this activity, you will be able to describe how digital figures should and should not be edited for use in publications. You also will be able to list key elements of the legend that are needed to effectively describe the experimental design and subsequent modifications. You will be able to identify and justify why data should be presented in a specific format (table, text, graph, etc.). And, you will be able to detail how data presentation can affect interpretation of results.

Procedure

Work in groups of 3-5 and use the worksheet to guide your discussions and answers. There are three parts to this activity.

DISCUSS

Part 1: Legends

Review each figure WITHOUT reading the legend. In small groups, discuss whether the results can be interpreted without the information in the legend. List questions you have about the figure. THEN read the legend and see if enough information is included to address your questions.

Part 2: Table or Graph?

In small groups, review the raw data provided and discuss how they should be presented in a manuscript (table, line graph, bar graph). Then share with the instructor your decision and reasoning.

Part 3: Digital Image Details

In small groups, assess the digital images presented in these figures and discuss whether the data are presented clearly. Develop suggestions on what could be improved.

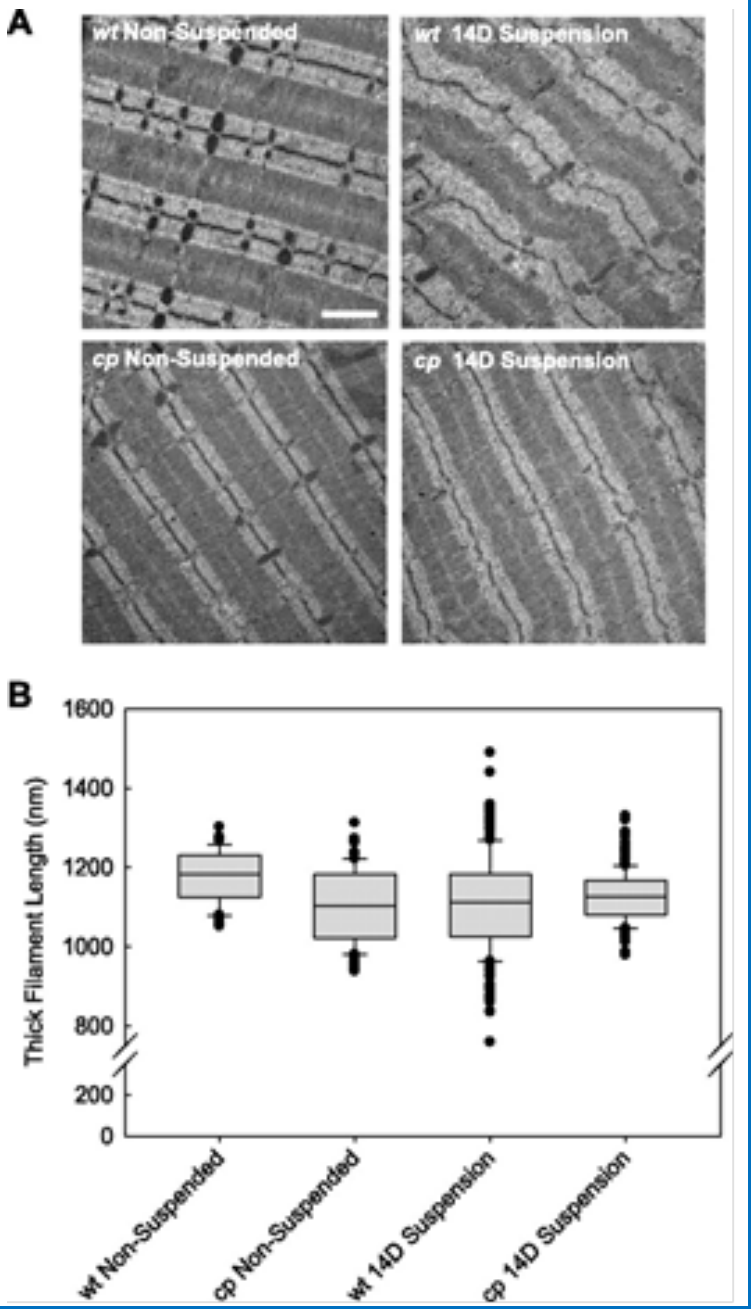
©American Physiological Society 2017

Part 1: Legends

Review each figure **WITHOUT** reading the legend. In your group, discuss whether the results can be interpreted without the information in the legend. List questions you have about the figure. **THEN** read the legend and see if enough information is included to address your questions.

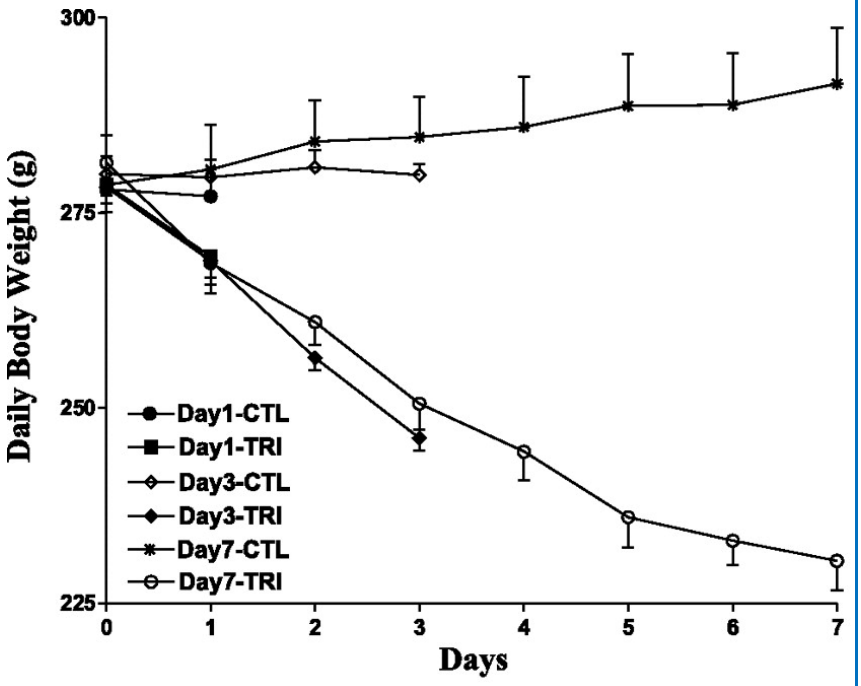
Example 1-1: Sarcomere structure for soleus muscles of nonsuspended wt and cp mice and mice exposed to 14 days of hindlimb suspension

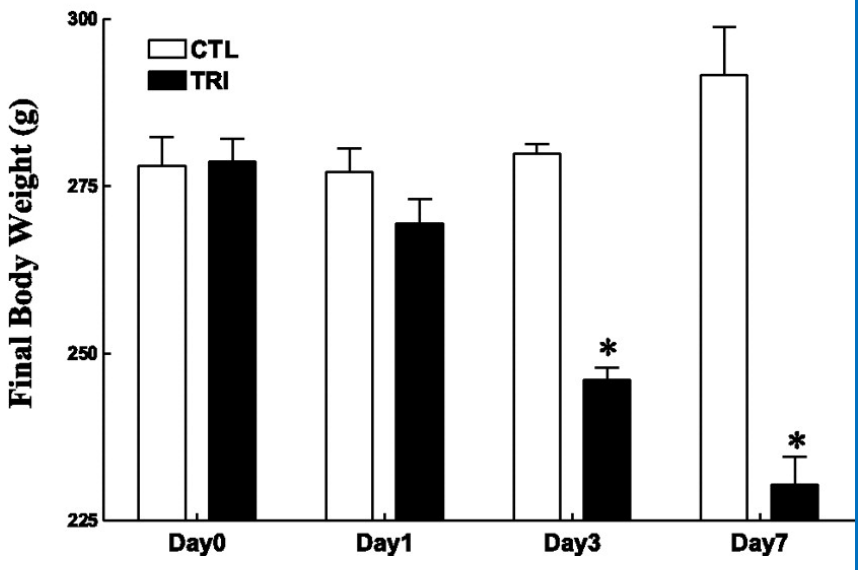
My questions about this figure:



Example 1-2: **Daily body weight change and final body weights in control and triamcinolone-treated groups**

My questions about this figure:

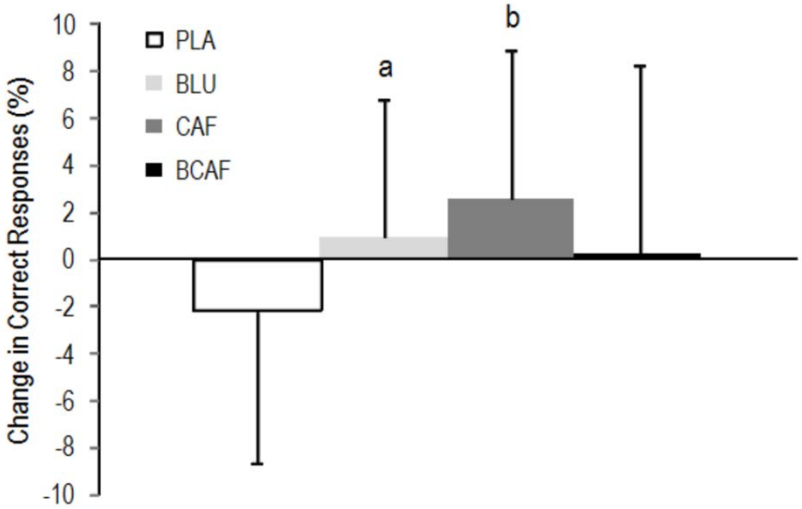




©American Physiological Society 2017

Example 1-3: Change in the accuracy in the visual Go/No-Go psychomotor task

My questions about this figure:



LEGENDS

Figure 1 legend

Sarcomere structure for soleus muscles of nonsuspended wt and cp mice and mice exposed to 14 days of hindlimb suspension. A: representative electron micrographs (EM) from soleus muscles of wt and cp mice that were either not suspended (nonsuspended) or exposed to 14 days of hindlimb suspension (14D suspension). Scale bar = 1 µm. B: thick filament lengths measured from muscle EMs are shown in a box plot representation. The line within the boxes indicates the median, while the top and bottom edges of the box indicate 75th and 25th percentiles, respectively. The brackets above and below the boxes indicate 90th and 10th percentiles, respectively, and the additional symbols are values that fall outside the 10th to 90th percentile range. Neither means nor medians were different among the groups, but note increased variability in thick filament lengths for muscles of wt mice after hindlimb suspension.

Ref: Jay J. Salazar, Daniel E. Michele, Susan V. Brooks, *Journal of Applied Physiology,* 108 (1) 120-127, 2010. DOI: 10.1152/japplphysiol.01080.2009

Figure 2 legend

Daily body weight change (top) over the experimental period and final body weights (bottom) in days 1, 3, and 7 control (CTL) and triamcinolone-treated (TRI) animal groups (n = 7 for each subgroup). Values are means ± SE. *Significantly different from CTL groups.

Ref: John E. Cho, Mario Fournier, Xiaoyu Da, Michael I. Lewis, *Journal of Applied Physiology,* 108 (1) 137-145, 2010. DOI:10.1152/japplphysiol.00704.2009.

Figure 3 legend

Change in the accuracy in the visual Go/No-Go psychomotor task.

a: small difference from PLA, b: moderate difference from PLA. PLA: white light/placebo; CAF: white light/240 mg caffeine; BLU: blue light/placebo; BCAF: blue light/240 mg caffeine**.**

Ref: Beaven, C.M., Ekström J., A Comparison of Blue Light and Caffeine Effects on Cognitive Function and Alertness in Humans. *PLoS ONE* 8(10): e76707, 2013. DOI:10.1371/journal.pone.0076707

 ©American Physiological Society 2017

Part 2: Table or Graph?

Each set of experimental results requires you to make decisions about the best way to present the data to the reader. For each of the examples below, decide how the data should be displayed and **WHY**.

Example 2.1

How should the following data be displayed - table, line graph, or bar graph?

Example 2.1 data	Diabetic	Smoker	Drinker	High Blood Pressure
Growth factor I with cardiovascular risk factor	32 ± 12	39 ± 9	41 ± 14	30 ± 11
Growth factor I without cardiovascular risk factor	30 ± 15	31 ± 10	45 ± 11	39 ± 11
Growth factor II with cardiovascular risk factor	1,470 ± 513	1,334 ± 665	1,528 ± 1,015	1,913 ± 985
Growth factor II without cardiovascular risk factor	1,470 ± 975	1,652 ± 1,014	1,553 ± 774	1,978 ± 836

Explain why:

Example 2.2

How should the following data be displayed - table, line graph, or bar graph?

Example 2.2 data	12h	24h	48h
Relative activity (fold) of fibroblast growth factor	1.0 ± 0.1	1.2 ± 0.5	2.2 ± 0.4
Relative activity (fold) of fibroblast growth factor + Substance A	3.0 ± 0.7	4.5 ± 1.2	4.2 ± 1.4

Explain why:

©American Physiological Society 2017

Example 2.3

How should the following data be displayed - table, line graph, or bar graph?

Example 2.3 data	Normalized Na current in…	
	Control cells	**Cells treated with A-23187**
-120 mV	0.96 ± 0.05	0.96 ± 0.03
-110 mV	0.98 ± 0.03	0.95 ± 0.02
-100 mV	0.99 ± 0.04	0.95 ± 0.02
-90 mV	0.95 ± 0.02	0.92 ± 0.04
-80 mV	0.78 ± 0.02	0.87 ± 0.03
-70 mV	0.72 ± 0.04	0.82 ± 0.03
-60 mV	0.51 ± 0.03	0.75 ± 0.04
-40 mV	0.26 ± 0.01	0.56 ± 0.01
-20 mV	0.20 ± 0.02	0.44 ± 0.01

Explain why:

©American Physiological Society 2017

Example 2.4

How should the following data be displayed - table, line graph, or bar graph?

Example 2.4 data			Adipose tissue weight, g
Brown		Lean mice	0.14±0.02
		Obese mice: Untreated	0.50±0.10
		Obese mice: Treated	0.21±0.05
White	Inguinal	Lean mice	0.40±0.07
		Obese mice: Untreated	5.78±0.50
		Obese mice: Treated	4.36±0.20
	Retrovesical	Lean mice	0.30±0.06
		Obese mice: Untreated	3.37±0.22
		Obese mice: Treated	3.22±0.11
			Plasma lipid levels, mmol/l
Total cholesterol		Lean mice	2.59±0.15
		Obese mice: Untreated	5.42±0.28
		Obese mice: Treated	2.91±0.16
Triglycerides		Lean mice	1.35±0.18
		Obese mice: Untreated	1.85±0.41
		Obese mice: Treated	0.81±0.05

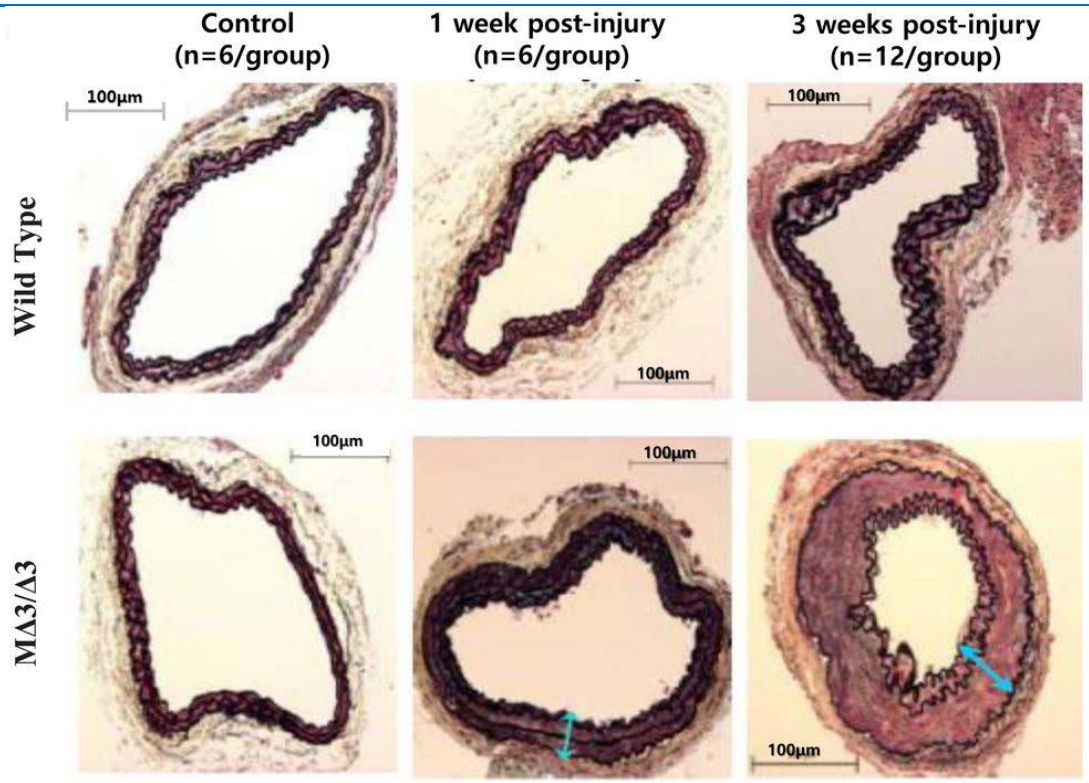
Explain why:

©American Physiological Society 2017

Part 3: Digital Images Details

In small groups, assess the digital images presented in these figures and discuss whether the data are presented clearly. Develop suggestions on what could be improved.

Example 3-1



Legend: Movat pentachrome-stained cross-sections of mouse carotid arteries. Uninjured control arteries in wild-type and MΔ3/Δ3 had similar thickness. At 1 and 3 wk post-injury, MΔ3/Δ3 arteries exhibited significantly increased medial thickening compared with wild-type arteries (denoted by blue arrows); *n* = 6 per group for 1 wk, *n* = 12 per group for 3 wk.

Ref: Lara Gotha, Sang Yup Lim, Azriel B. Osherov, Rafael Wolff, Beiping Qiang, Ilana Erlich, Nafiseh Nili, Sivaram Pillarisetti, Ya-Ting Chang, Phan-Kiet Tran, Karl Tryggvason, Ulf Hedin, Karin Tran-Lundmark, Suzanne L. Advani, Richard E. Gilbert, Bradley H. Strauss. *American Journal of Physiology -Heart and Circulatory Physiology*, Vol. 307 no. 3, H337-H345, 2014. DOI: 10.1152/ajpheart.00654.2013

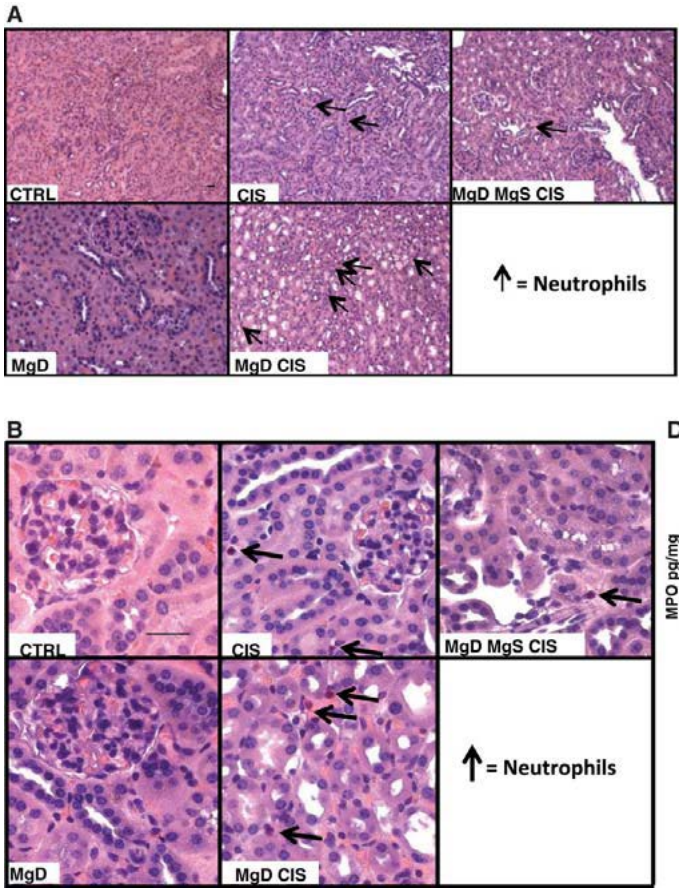
Example 3-1 Analysis

a) List some positive aspects of how these data are presented.

b) Could this figure be improved? If so, how?

©American Physiological Society 2017

Example 3-2



Legend: CIS treatment after Mg deficiency is associated with enhanced neutrophil infiltration and renal myeloperoxidase (MPO). Mice were maintained on either 100% Mg or 10% Mg-deficient diets or were maintained on a 10% Mg diet followed by Mg supplementation, as described in methods, and then treated with saline (CTRL) or CIS (12 mg/kg). All mice were euthanized 48 h post-CIS (or saline); fixed kidney tissues were evaluated for neutrophils by Leder staining. A and B: representative images for each group at ×200 (A) and ×400 (B) magnification. Scale bars = 20 μm.

Ref: Malvika H. Solanki, Prodyot K. Chatterjee, Madhu Gupta, Xiangying Xue, Andrei Plagov, Margot H. Metz, Rachel Mintz, Pravin C. Singhal, Christine N. Metz. *American Journal of Physiology - Renal Physiology*, Vol. 307 no. 4, F369-F384, 2014. DOI: 10.1152/ajprenal.00127.2014

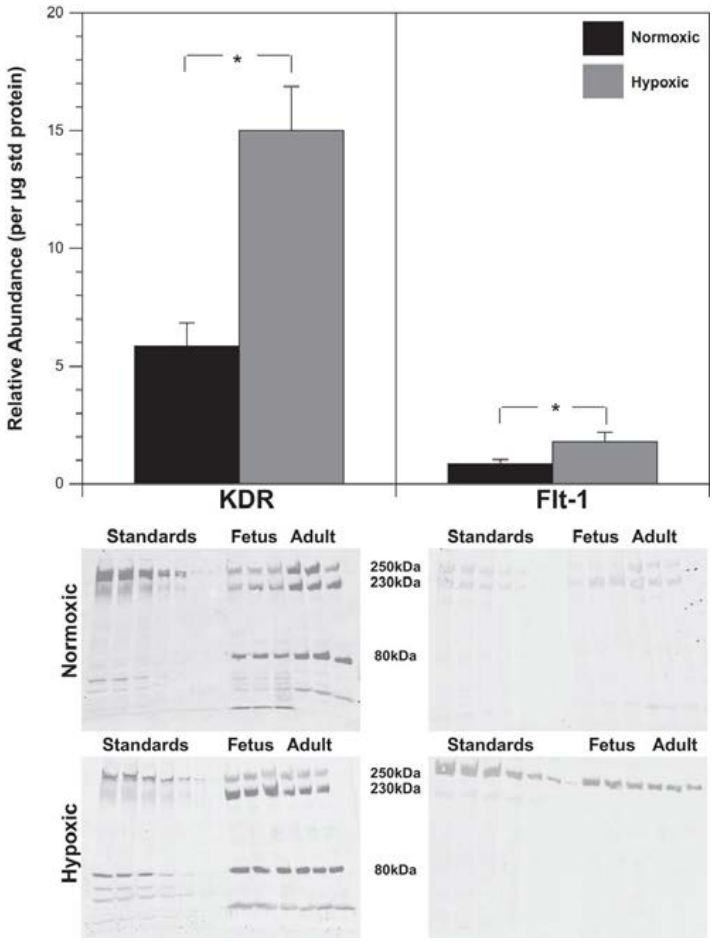
Example 3-2 Analysis

a) List some positive aspects of how these data are presented.

b) Could this figure be improved? If so, how?

©American Physiological Society 2017

Example 3-3



Legend: Chronic hypoxia upregulates VEGF receptors. Whereas Western blot quantification revealed that VEGF levels were not altered by chronic hypoxia, abundances of VEGF tyrosine kinase receptors were significantly increased by chronic hypoxia. VEGF R1 levels were increased by 107% and VEGF R2 levels were increased by 156% in hypoxic arteries. Receptor levels presented here were measured relative to amounts in a standard pool of normoxic carotid arteries. In light of the fact that both fetal and adult samples were run on the same gel with the same standards, the lanes with the adult samples are shown here adjacent to the previously published standard and fetal bands (1) for all 4 groups. Bar graphs for KDR represent averages of all glycosylated bands, but for Flt-1 only the 250-kDa band was used for quantification because the 230-kDa band was absent in the hypoxic arteries. Results are presented as mean ± SD for n = 6 for all experimental groups. Comparative significant differences: *P < 0.05 via ANOVA.

Ref: Olayemi O. Adeoye, Vincent Bouthors, Margaret C. Hubbell, James M. Williams, William J. Pearce. *Journal of Applied Physiology,* 117 (7) 777-787, 2014. DOI: 10.1152/japplphysiol.00012.2014

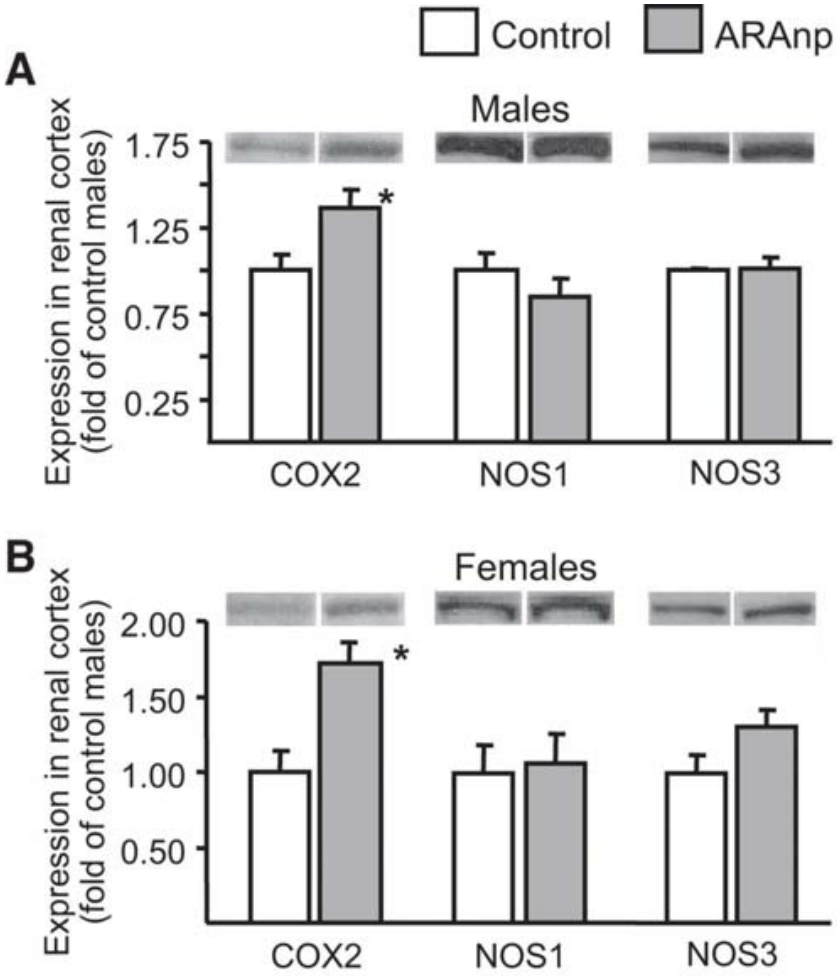
Example 3-3 Analysis

a) List some positive aspects of how these data are presented.

b) Could this figure be improved? If so, how?

©American Physiological Society 2017

Example 3-4



Legend: Cyclooxygenase (COX)-2, neuronal nitric oxide synthase (NOS1), and endothelial nitric oxide synthase (NOS3) expression in the renal cortex of control rats and rats treated with an ANG II type 1 receptor antagonist during the nephrogenic period (ARAnp) at 16–17 mo of age. Protein abundance was normalized for GADPH expression. *P < 0.05 vs. vehicle-treated (control) rats.

Ref: Fara Saez, Virginia Reverte, Alexander Paliege, Juan Manuel Moreno, María T. Llinás, Sebastian Bachmann, F. Javier Salazar, *American Journal of Physiology - Renal Physiology,* Vol. 307 no. 4, F461-F470, 2014.
DOI: 10.1152/ajprenal.00198.2014

Example 3-4 Analysis

a) List some positive aspects of how these data are presented.

b) Could this figure be improved? If so, how?

©American Physiological Society 2017

Activity G
My Data Management and Integrity Checklist

Purpose In this activity, you will use what you have learned to establish a plan for data management and presenting data using professional standards of practice. After completing the activity, you should have a plan and a checklist for data management and integrity AND a plan for teaching best practices to your students. You also should be ready to share your plan with collaborators.

Procedure Look back at your previous work, readings, and the Interactive Lecture. Use the following questions to guide the development of your checklist. Discuss your plan with your instructor.

WRITE

This checklist can be added to other checklists you generate through the modules on publication ethics best practices.

Following are some suggestions on information you may want to include but MAKE THE LIST YOUR OWN. We recommend you recreate these tables in your word processing program so you can modify your checklist later.

My Best Practices Checklist

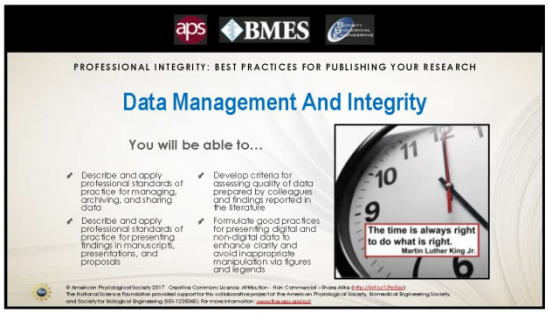
I. DATA MANAGEMENT AND STORAGE	
A. These parts of my research must be available to readers, journals, and institutions, if requested:	
B. This information should be recorded in my lab notebook for every experiment:	
C. This information should be recorded about data and versions of data:	
D. This is my plan for storing data securely:	
E. **PASS IT ON**	This is my strategy for teaching students and new lab personnel about my data management and storage plan:

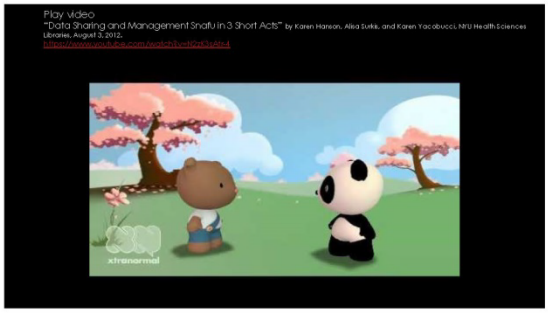
©American Physiological Society 2017

F. This is my strategy for MONITORING that data in my lab/collaborations is being stored securely and managed properly:	
II. DATA INTEGRITY **Following are my rules for…**	
A. Assessing the quality and validity of data in articles, including my OWN:	
B. Writing legends:	
C. Determining whether data are best described by text, table, line graph or bar graph:	
D. Cropping original captures of digital figures:	
E. Adjusting brightness and/or contrast in a digital figure:	
F. Adjusting order and/or arrangement in a digital figure:	
G. PASS IT ON	This is my plan for teaching these critical skills in data management and integrity to my students and sharing this information with my collaborators:

©American Physiological Society 2017

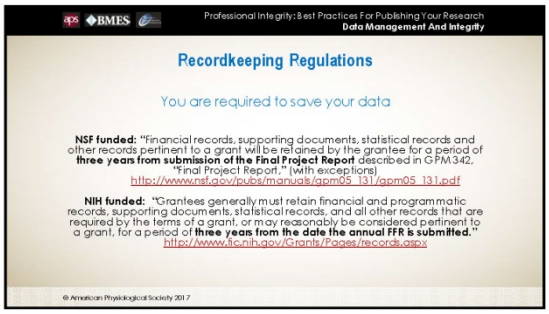
Student Slide Handout

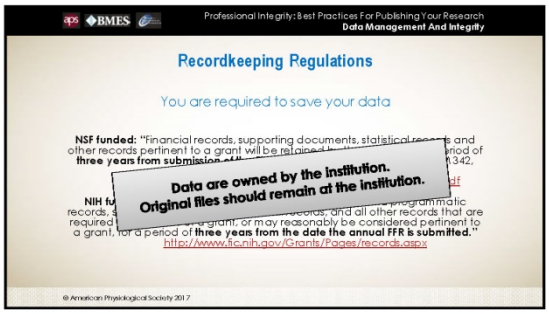


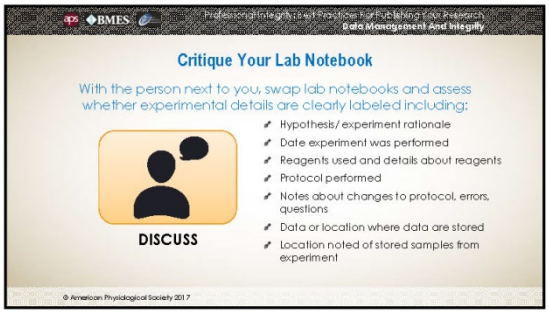


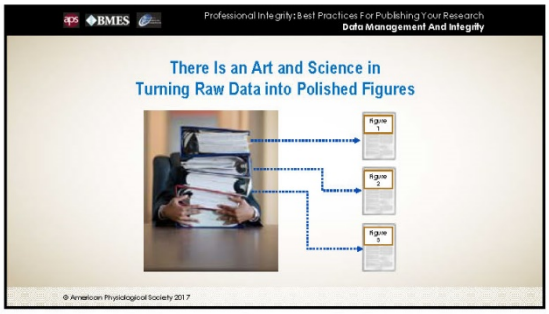
 ©American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
What are Good Practices for Keeping a Record of Your Data?
Label your data clearly (date / experiment / lanes / treatments)
Save versions of data at each step of modification (from raw data to figures prepared for publication)
Specify the pieces of data that are included in your manuscript
Could your advisor readily locate a piece of your data if you had to instruct him/her over the phone or by email?
© American Physiological Society 2017

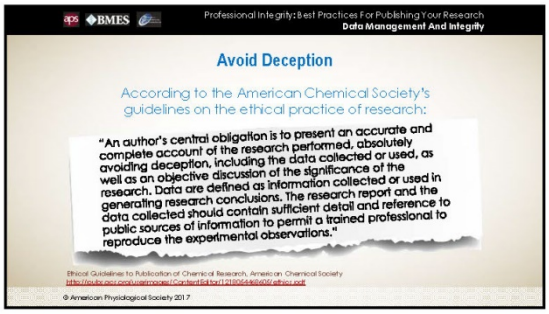

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Recordkeeping Regulations
You are required to save your data
NSF funded: "Financial records, supporting documents, statistical records and other records pertinent to a grant will be retained by the grantee for a period of three years from submission of the Final Project Report described in GPM 342, "Final Project Report," (with exceptions)
http://www.nsf.gov/pubs/manuals/gpm05_131/gpm05_131.pdf
NIH funded: "Grantees generally must retain financial and programmatic records, supporting documents, statistical records, and all other records that are required by the terms of a grant, or may reasonably be considered pertinent to a grant, for a period of three years from the date the annual FFR is submitted."
http://www.fic.nih.gov/Grants/Pages/records.aspx
© American Physiological Society 2017

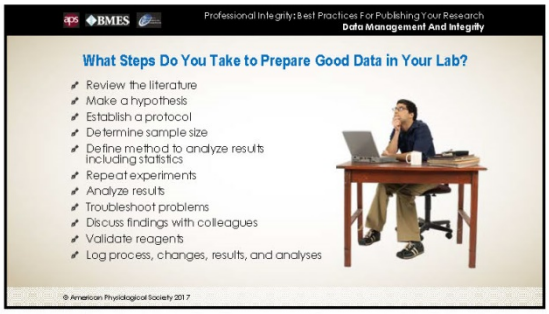

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Recordkeeping Regulations
You are required to save your data
Data are owned by the institution.
Original files should remain at the institution.
© American Physiological Society 2017

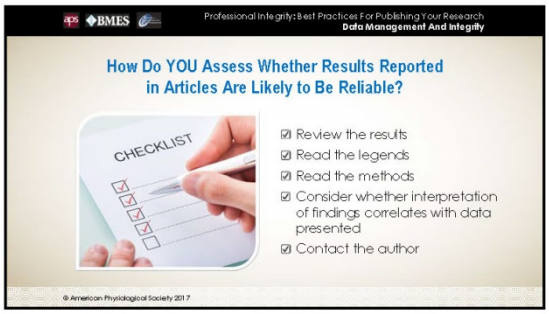




©American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Avoid Deception
According to the American Chemical Society's guidelines on the ethical practice of research:
"An author's central obligation is to present an accurate and complete account of the research performed, absolutely avoiding deception, including the data collected or used, as well as an objective discussion of the significance of the research. Data are defined as information collected or used in generating research conclusions. The research report and the data collected should contain sufficient detail and reference to public sources of information to permit a trained professional to reproduce the experimental observations."
Ethical Guidelines to Publication of Chemical Research, American Chemical Society
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
What Steps Do You Take to Prepare Good Data in Your Lab?
Review the literature
Make a hypothesis
Establish a protocol
Determine sample size
Define method to analyze results including statistics
Repeat experiments
Analyze results
Troubleshoot problems
Discuss findings with colleagues
Validate reagents
Log process, changes, results, and analyses
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
How Do YOU Assess Whether Results Reported in Articles Are Likely to Be Reliable?
CHECKLIST
Review the results
Read the legends
Read the methods
Consider whether interpretation of findings correlates with data presented
Contact the author
© American Physiological Society 2017

What Else Should Be Considered When Assessing the Quality and Validity of Data Presented in a Scholarly Article?

- Were experiments performed blinded?
- Were basic experiments repeated?
- Were all results presented?
- Were there positive and negative controls?
- Were reagents validated?
- Were statistical tests appropriate?

C.G. Begley, Six red flags for suspect work, Nature, 23 May 2013, Vol 497, pg 433

© American Physiological Society 2017

"The results show what?"

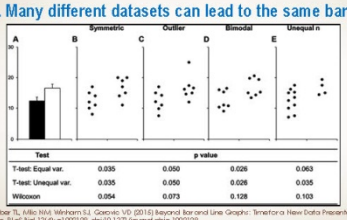
Share examples of poorly presented figures with the class.

- Explain why the presentation is not ideal.
- What could be corrected?

DISCUSS

© American Physiological Society 2017

Fig 1. Many different datasets can lead to the same bar graph.



Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

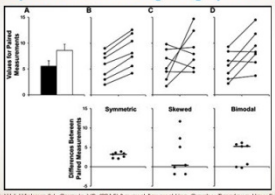
Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128
http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

©American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Fig 2. Additional problems with using bar graphs to show paired data.

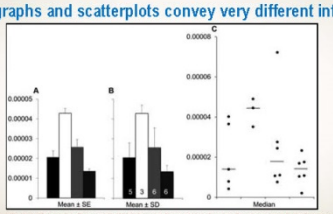


Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128 http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Fig 3. Bar graphs and scatterplots convey very different information.



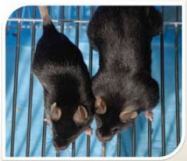
Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128 http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Compare 3 Experiments

- These experiments all use hindlimb suspension for various periods of time to assess changes in body weight and muscles mass of the animals.
- Two experiments analyze Myostatin knockout mice (Mstn-/-). Myostatin is a protein that regulates muscle growth. Without myostatin, muscles grow much larger than normal.



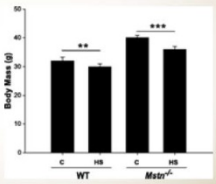
Lee S-J (2007) Quadrupling Muscle Mass in Mice by Targeting TGF-β Signaling Pathways. PLoS ONE 2(9): e789. doi:10.1371/journal.pone.0000789 Used with permission per Creative Commons Attribution 2.5 Generic license.

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Myostatin-deficient Mice Lose More Skeletal Muscle Mass than Wild-type Controls During Hindlimb Suspension

Figure legend: Mean ± SE body mass of wild-type (WT) and myostatin gene-deficient [Mstn(-/-)] mice subjected to 7 days of hindlimb suspension (HS) or serving as ground-based controls (C). **P < 0.01 and ***P < 0.001, differences between treatment groups (within genotype).



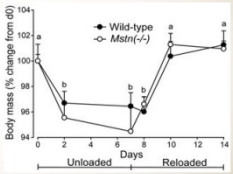
McMahon C D et al. Am J Physiol Endocrinol Metab 2003;285:E82-E87

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Translational Signalling, Atrogenic and Myogenic Gene Expression during Unloading and Reloading of Skeletal Muscle in Myostatin-Deficient Mice

Figure 1. Body mass (mean+sem) for Mstn(−/−) and wild-type mice during seven days of unloading followed by seven days of reloading (n = 6 per genotype and day). Unlike letters denote significant differences (P<0.05) across days (independent of genotype).



Smith HK, Matthews KG, Oldham JM, Jeanplong F, Falconer SJ, et al. (2014) PLoS ONE 9(4): e94366. doi:10.1371/journal.pone.0094366 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094366

© American Physiological Society 2017

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity

Muscle glucose update in the rat after suspension with single hindlimb weight bearing

TABLE 1. *Effect of suspension with single hindlimb weight bearing or control conditions on body mass*

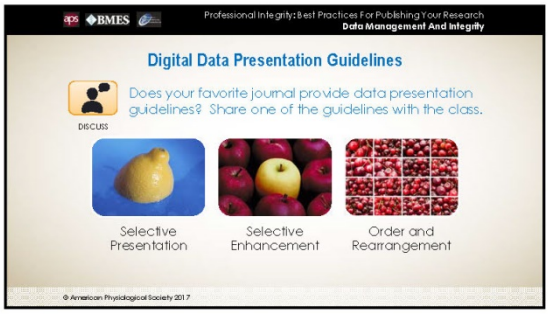
	n	Body Mass, g		
		Presuspension	Day 7	Day 14
Cage control	42	291±3	321±2*	346±3*
Suspension	47	297±3	261±3*†	270±3*†

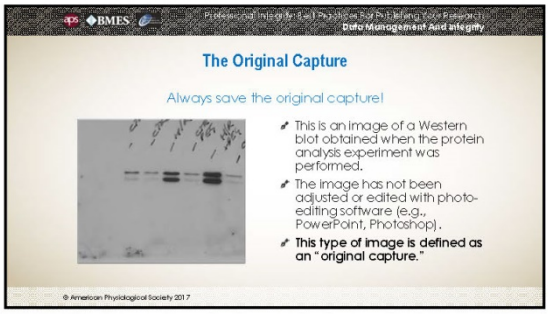
Values are means ± SE; n, no. of rats. * Mean significantly different from presuspension value; † mean significantly different from corresponding cage control value ($P \leq 0.05$).

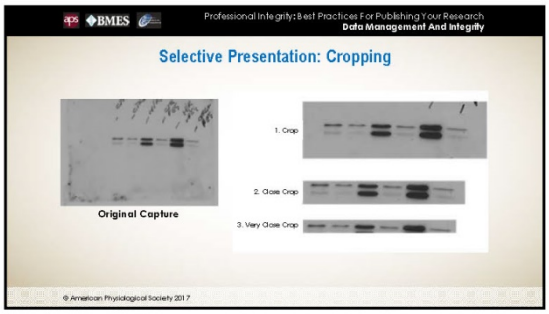
Stump CS et al., Journal of Applied Physiology 1993 May; 74(5):2072-8

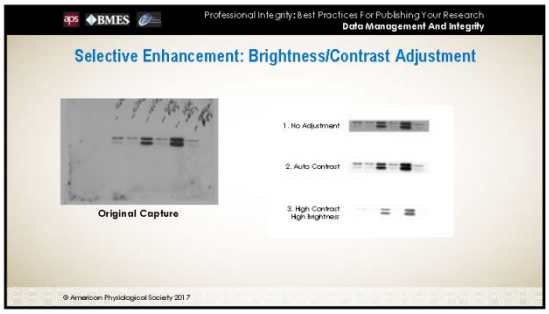
© American Physiological Society 2017

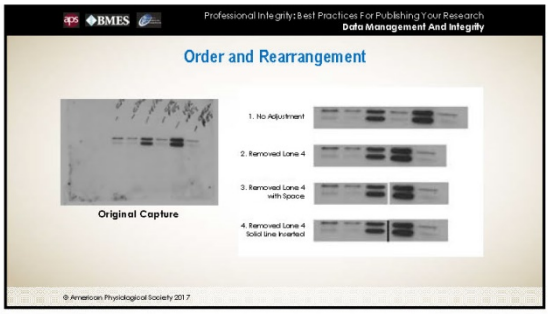
©American Physiological Society 2017

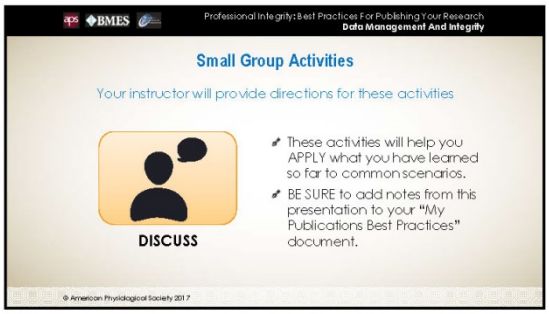

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Digital Data Presentation Guidelines
Does your favorite journal provide data presentation guidelines? Share one of the guidelines with the class.
DISCUSS
Selective Presentation
Selective Enhancement
Order and Rearrangement
© American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
The Original Capture
Always save the original capture!
This is an image of a Western blot obtained when the protein analysis experiment was performed.
The image has not been adjusted or edited with photo-editing software (e.g., PowerPoint, Photoshop).
This type of image is defined as an "original capture."
© American Physiological Society 2017

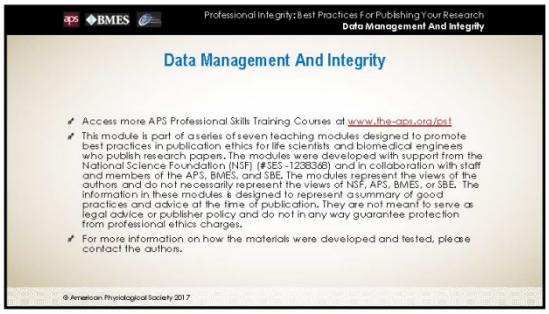

Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Selective Presentation: Cropping
Original Capture
1. Crop
2. Close Crop
3. Very Close Crop
© American Physiological Society 2017

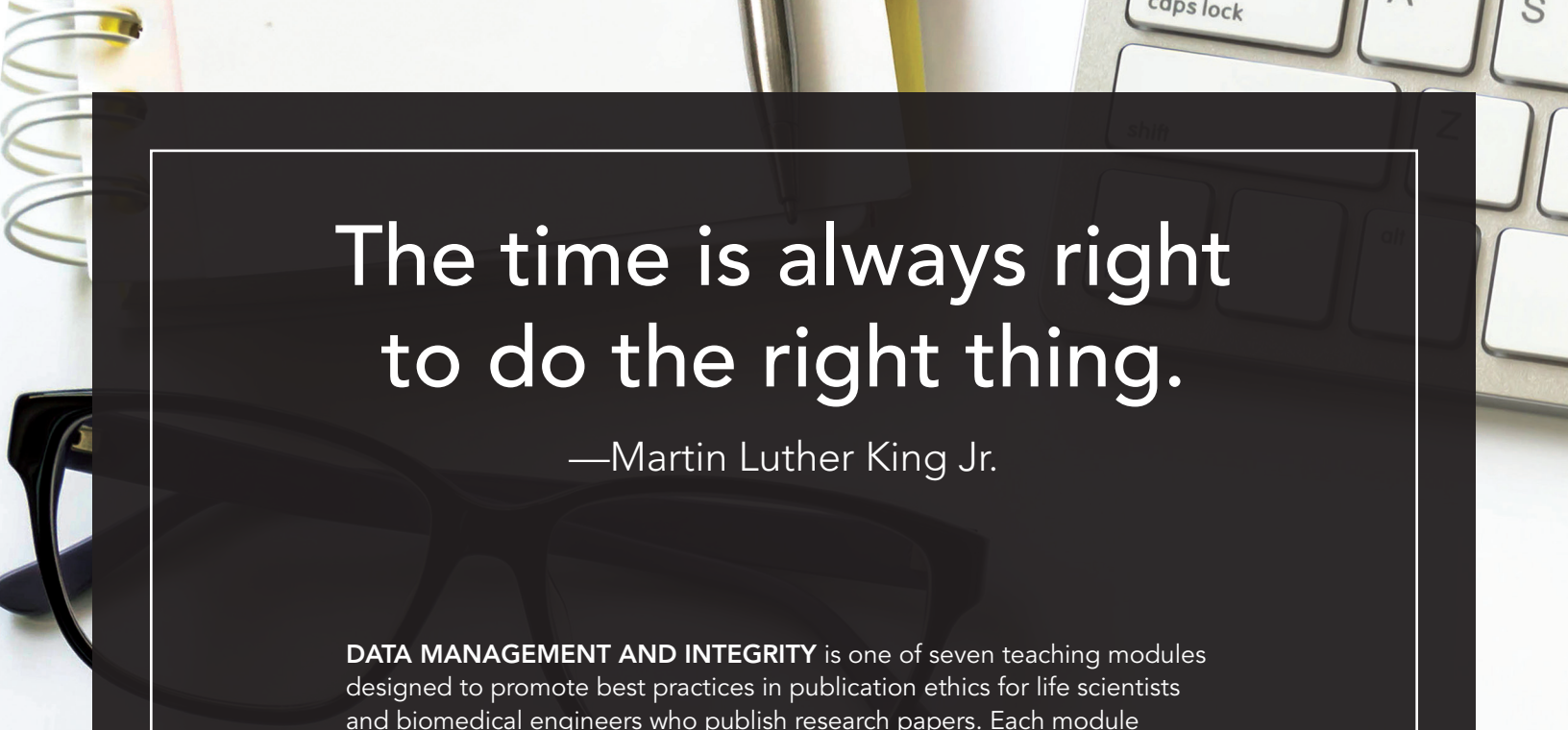






©American Physiological Society 2017


Professional Integrity: Best Practices For Publishing Your Research
Data Management And Integrity
Data Management And Integrity
Access more APS Professional Skills Training Courses at www.the-aps.org/pst
This module is part of a series of seven teaching modules designed to promote best practices in publication ethics for life scientists and biomedical engineers who publish research papers. The modules were developed with support from the National Science Foundation (NSF) (#SES -1238368) and in collaboration with staff and members of the APS, BMES, and SBE. The modules represent the views of the authors and do not necessarily represent the views of NSF, APS, BMES, or SBE. The information in these modules is designed to represent a summary of good practices and advice at the time of publication. They are not meant to serve as legal advice or publisher policy and do not in any way guarantee protection from professional ethics charges.
For more information on how the materials were developed and tested, please contact the authors.
© American Physiological Society 2017



The time is always right to do the right thing.

—Martin Luther King Jr.

DATA MANAGEMENT AND INTEGRITY is one of seven teaching modules designed to promote best practices in publication ethics for life scientists and biomedical engineers who publish research papers. Each module provides information on and principles of the most common publication ethics issues as well as the tools needed to integrate and apply professional standards of practice to real life situations. After finishing each module, students will have a personal checklist to use in the preparation of future manuscripts AND a plan for teaching module principles to their future trainees and collaborators.

Modules are designed to be used by higher education institutions, laboratory groups, individuals, and professional societies. The teaching paradigms used in the modules support various types of learners and were designed to integrate into current Responsible Conduct of Research (RCR) training courses/programs.

Modules were developed with support from the National Science Foundation (NSF) (#SES -1238368) and in collaboration with staff and members of the American Physiological Society, Biomedical Engineering Society, and the Society for Biological Engineers.

Handouts for instructor and students, audio and video resources, and online course links are available at www.the-aps.org/pst for all seven modules:

- Authorship
- Conflicts of Interest
- Considerations for Animal and Human Studies
- Data Fabrication and Falsification
- Data Management and Integrity
- Overlapping Publications
- Text Preparation and Avoiding Plagiarism