Previously funded projects

Projects previously funded through this mechanism

These are the projects funded through a previous round of this call.

UTILIZING ESTABLISHED YOUTUBE INFRASTRUCTURE FOR TRAINING MODULE DEVELOPMENT

- **DESCRIPTION (provided by applicant):** In order to ensure that studies are well designed, well reported, and therefore better able to generate reproducible results, we must train researchers adequately in these areas. We propose to develop a series of webisodes targeted at graduate students, postdoctoral fellows, and beginning investigators that will address critical features of experimental design and analysis/reporting. We will capitalize on our experience explaining (1) how research works, (2) what makes good research, (3) how to understand results, and (4) what they mean to the healthcare system and the practice of medicine. Currently, our project team has a very successful blog (The Incidental Economist) in which we explain how research and evidence inform both the practice of health care and health policy. It has developed enough of a following that we now contribute to blogs at Academy Health and JAMA and we write weekly for the New York Times. We also have a popular YouTube Channel (i.e., Healthcare Triage) where we explain how research and policy works, and discuss what studies "mean" for a lay audience. For this proposal, we will create two innovative and unique Learning Modules using our existing writing and production infrastructure. The Experimental Design Learning Module will focus on the intricacies of designing research that is robust, with a specific eye towards making it reproducible. This module will be comprised of four distinct learning units: 1) Replication, 2) Pitfalls with Experimental Design, 3) Measurement, and 4) Randomization. The Analysis/Reporting Learning Module will cover the various factors that are critical to writing about research with enough clarity that those reading about it can reproduce the experiment with fidelity. This Learning Module will be comprised of four distinct learning units: 1) Pitfalls ith Analysis, 2) Scientific Writing, 3) How to Write a Scientific Paper, and 4) Paper Submission. Each learning unit will have one or more sub-topics that will be the subject of an individual webisodes. These webisodes will be short units of training, of sufficient depth and coverage to empower learners with real-world, practical knowledge and skills to modify their conduct and reporting of research. And perhaps more importantly, these webisodes will be entertaining, engaging, and readily accessible. The webisodes will be marketed and distributed using "traditional" methods (i.e., brochures, flyers, CME credit) as well as 'new media' (YouTube, Twitter, blogs). Because the content will be free of charge, and accessible over any Internet connection, the potential audience is at least national, if not international. Our goal is to provie learners with a new educational option for gaining the knowledge they will need regarding factors affecting reproducibility in experimental design, data analysis, and reporting, in a user-friendly, accessible format. Our approach also allows learners then to return to specific webisodes when they need reminders on certain topics, or answers to specific questions. It is our hope that our two modules will supplement their learning and provide them new opportunities to make their research writing clearer, and the research it describes more reproducible.

- **PUBLIC HEALTH RELEVANCE:** Currently, a lack of awareness or adherence to sufficiently high standards in the planning, execution, and reporting of scientific experiments and their results exists. In order to ensure that studies are well designed, well reported, and therefore better able to generate reproducible results, we must ensure that researchers are receiving adequate training in these areas. In order to promote better training and ensure the reliability and reproducibility of our research, our project team proposes to develop a series of short webisodes targeted at graduate students, postdoctoral fellows, and beginning investigators that will address critical features of experimental

design, and analysis/reporting that will be marketed and distributed utilizing both traditional methods (i.e., brochures, flyers, CME) and 'new media' (i.e., YouTube, blogs, Twitter).

- **NIH Spending Categories**: Behavioral and Social Science; Clinical Research; Complementary and Alternative Medicine; Neurosciences

- **Project terms**: Academy; Address; Adherence; Affect; Area; Awareness; Basic Science; Charge; Data Analyses; design; Development; Educational process of instructing; Electronic Mail; empowered; Ensure; experience; experimental analysis; Experimental Designs; Eye; Goals; graduate student; Health; Health Policy; Healthcare; Healthcare Systems; high standard; Housing; Individual; innovation; Institution; International; Internet; Knowledge; Laboratories; Learning; Learning Module; lectures; Length; Link; Marketing; Measurement; medical schools; Medicine; Methods; New York; Pamphlets; Paper; Policy Research; Postdoctoral Fellow; pre-clinical; Production; programs; Publications; Publishing; Randomized; Reading; Reporting; Reproducibility; Research; Research Design; Research Infrastructure; Research Methodology; Research Personnel; research study; Research Training; Schools; Science; Series; skills; statistics; Time; trafficking; Training; Training Activity; Training and Education; Triage; user-friendly; web site; Work; Writing

- **Results**: This results in two modules: (1) Experimental Design and (2) Analysis and Reporting. Each is a YouTube playlist, with each video 4–5 minutes. Similar videos are grouped in the same section. https://iu.cloud-cme.com/aph.aspx?EID=36408&P=3000&CaseID=764#

Experimental Design:

1. Replication / Pseudoreplication (10 minutes)
2. Randomization (5 minutes)
3. Hypothesis / Block Design / Factorial Design / Recruitment / Adherence (20 minutes)
4. Negative controls / Positive controls / Calibration / Precision and Accuracy / Intra-observer Variability / Baseline Assessment / Blinding / Observer Bias / Recording Data (30 minutes)

Analysis and Reporting

1. Introduction (4 minutes)
2. Sample Size and Power / P Values / P Value Meaning (15 minutes)
3. Titles / Introductions / Methods / Results / Discussion / References / Tables and Graphs / Jargon (30 minutes)
4. The Review Process (4 minutes)

PROMOTING AWARENESS AND KNOWLEDGE TO ENHANCE SCIENTIFIC RIGOR IN NEUROSCIENCE

- **DESCRIPTION (provided by applicant):** While the foundation for rigorous scientific conduct is created primarily at the university level through strong mentor relationships, scientific training, and the peer review system, there is little systematic, formal training for graduate students and postdoctoral fellows in experimental design and data analysis. To meet this need, the Society for Neuroscience (SfN) will implement a multimodal training and awareness building program targeted towards early career scientists on experimental design, data analysis, and reporting communicated through the specific lens of conducting neuroscience research. The aim of the program emphasizes development of an interactive modular webinar series featuring practical strategies and tools taught by renowned field experts. The program will be combined with communications and related activities that will be held through many of SfN's leading venues for research, including the SfN annual meeting, and includes partnerships with academic institutions. The long term goal of this program is to increase awareness about this essential topic, impart knowledge of best practices, and change practices related to scientific rigor and reproducibility for early career scientists.

- **PUBLIC HEALTH RELEVANCE:** It is estimated that neurological illnesses affect one billion individuals worldwide (World Health Organization, 2007), including 50 million Americans, requiring

over $500 billion annually in U.S. healthcare costs, not including costs borne by caretakers. It is critical to develop a strong and diverse neuroscience research workforce to find new and innovative ways to prevent and treat neurological illnesses. Further, the research in which this workforce engages needs to be well designed, well vetted, accessible, and replicable; as other areas of science and scientific discovery will benefit from new discoveries.

- **NIH Spending Category**: Clinical Research; Drug Abuse (NIDA only); Neurosciences; Substance Abuse

- **Project Terms:** Address; Advertisements; Affect; American; Area; Awareness; base; behavior change; Biological Sciences; Biomedical Research; Brain; career; Case Study; Communication; cost; Data Analyses; Data Collection; data management; Data Reporting; design; Development; early-career faculty; Educational process of instructing; Effectiveness; Ensure; Event; Experimental Designs; Faculty; Feedback; Focus Groups; Foundations; Future; Goals; graduate student; Health; Health Care Costs; improved; Individual; innovation; insight; Institution; Journals; Knowledge; lectures; lens; Location; meetings; member; Mentors; Modeling; multidisciplinary; Neurologic; Neurosciences; Neurosciences Research; Peer Review; physical science; Positioning Attribute; Postdoctoral Fellow; prevent; programs; Published Comment; Reporting; Reproducibility; Research; research study; Resources; Science; Scientist; Series; Societies; Surveys; System; Testing; tool; Training; Training Activity; Training Programs; United States National Institutes of Health; Universities; webinar; World Health Organization

- **Results**: Seven webinars, plus PDFs of Webinar Discussion Questions to follow each webinar (about 5 discussion questions for each webinar). They also put together around five–ten papers for people to read pre-webinar for each webinar. This is an example web page for one webinar: http://neuronline.sfn.org/Articles/Professional-Development/2016/Improving-Experimental-Rigor-and-Enhancing-Data-Reproducibility-in-Neuroscience. It looks like each webinar was lead by a different person, and they picked more senior scientists in many cases. Each webinar is an hour.

Society for Neuroscience Rigor and Reproducibility Training Webinars:

1. Improving Experimental Rigor and Enhancing Data Reproducibility in Neuroscience
2. Minimizing Bias in Experimental Design and Execution
3. Best Practices in Post-Experimental Data Analysis
4. Best Practices in Data Management and Reporting
5. Statistical Applications in Neuroscience
6. Experimental Design to Minimize Systemic Biases: Lessons from Rodent Behavioral Assays and Electrophysiology Studies
7. Tackling Challenges in Scientific Rigor: The (Sometimes) Messy Reality of Science

BEYOND TEXTBOOK, YET SIMPLE, STATISTICAL TOOLS FOR REPRODUCIBLE ANIMAL RESEARCH

- **DESCRIPTION (provided by applicant):** The capacity of laboratory animal research to be translated to interventions for human health is dependent on research reproducibility, replicability, and generalizability (RRG). It is often the case that challenges with RRG arise from differences in scientific protocols, unknown or uncontrolled confounding factors, incomplete communication of scientific protocols, and unintentional misapplication of statistical principles, rather than the oft-discussed research misconduct. Many of these challenges with RRG involve practical and theoretical aspects of statistics and study design that go beyond what is taught in basic statistics courses or textbooks. We therefore propose creating a set of animated vignettes to educate early career, laboratory animal investigators (namely, graduate students, postdoctoral fellows, and beginning investigators) about these issues. We specifically propose modules on variability, power, randomization, hypothesis testing, and inferences that address common, and yet not routine-textbook needs of laboratory researchers. Each module will be composed of several short, animated vignettes that will 1) discuss a statistical or design principle

that is important for RRG; 2) foster dialogue and collaboration between statistical and laboratory animal scientists; and 3) model diversity of statistical and laboratory animal scientist with respect to expertise and demographic characteristics. Each vignette will also be evaluated for content validity, face validity, and educational value by consulting with statistical experts, experienced lab animal researchers, and early career investigators, respectively. Finally, vignettes will be reinforced with additional online instructional content, including tutorial literture and self-assessment quizzes. Our team will widely disseminate the instructional materials leveraging our experience and resources creating and sharing online educational content, and we commit to maintain the materials in an openly available web portal without cost to end users. By supporting principles of RRG in this way, we contribute broadly to the mission of the NIH in fostering fundamental creative discoveries, innovative research strategies, and their applications; developing, maintaining, and renewing scientific human resources; and exemplifying and promoting the highest level of scientific integrity in the conduct of science. We specifically contribute to the mission of the NIGMS by "training the next generation of scientists, in enhancing the diversity of the scientific workforce, and in developing research capacities throughout the country."

- **PUBLIC HEALTH RELEVANCE:** Improving the reliability of published laboratory animal research is imperative to being able to translate basic science to interventions for human health. The proposed training modules will support this mission by creating educational materials that address commonly encountered but unexpectedly complex statistical and research design issues.

- **NIH Spending Category**: Aging; Clinical Research; Neurosciences

- **Project terms**: Address; Advocate; American; Animal Experimentation; Animals; Basic Science; Biometry; career; Characteristics; Collaborations; Communication; Communities; Complex; Consult; cost; Country; Crossover Design; Data; design; Education; Educational Materials; Educational process of instructing; Equation; Evaluation; experience; experimental analysis; Experimental Designs; Face; Failure; Fostering; graduate student; Health; help-seeking behavior; Human; Human Resources; Immersion Investigative Technique; improved; innovation; Internet; Intervention; Investigation; Knowledge; Laboratories; Laboratory Animals; Laboratory Research; Laboratory Scientists; Learning; Measures; men; Methods; Mission; Modeling; National Institute of General Medical Sciences; Natural Sciences; next generation; Paper; Postdoctoral Fellow; Procedures; Protocols documentation; public health relevance; Publishing; Randomized; Reading; Recommendation; Reproducibility; Research; Research Design; Research Misconduct; Research Personnel; research study; Resources; Science; Scientist; Self Assessment; statistics; Testing; Textbooks; tool; Training; Translating; Uncertainty; United States National Institutes of Health; web site; Woman; Writing

CONTROLS IN ANIMAL STUDIES PROFESSIONAL SKILLS COURSE

- **DESCRIPTION (provided by applicant):** Animal models are used in biomedical research studies to determine how genetic, therapeutic, and other modifications affect the structure and function of a living system. Some of these studies influence what drug candidates are selected for human clinical trials or what therapeutic intervention should be tested in the clinic. However, there is growing concern that much of the information reported in the scholarly literature, particularly those that use animal models, is unreliable due to poor experimental design and technique, improper analysis of findings or over- interpretation of results. As such, scholarly journals, funding bodies, and scientific associations have developed author, and reviewer, checklists to help ensure that the experimental details reported in manuscripts are accurate, robust and can be repeated by other investigators. Despite these efforts, by the time journals interact with researchers, as authors, the experimental portion of the study is complete including any flawed research practices related to the study. Thus, while journals can enforce transparency of the research reported in manuscripts, they cannot change how the original research was performed. To address this issue, we propose to develop an on-demand education module, for graduate students, postdoctoral fellows and early career investigators who use, or are planning to use animal models in their research. The module will be developed based upon our proven method for designing highly-effective Professional Skills Training (PST) Courses such as the writing and reviewing manuscripts PST and the publication ethics PST. In Aim 1 we will prepare an education module,

focused on developing skills in designing, performing and reporting well- controlled animal studies, using effective student-centered learning pedagogy. In Aim 2, we will pilot, evaluate, and revise the draft module with multiple audiences including graduate students, postdoctoral fellows and other biomedical research investigators. In Aim 3, we will disseminate the revised version of the module and related teaching resources via free digital libraries, professional society workshops, and websites. In Aim 4, we will develop and support a community of practice for animal researchers in order to foster an ongoing dialogue on best practices for performing animal studies. The education materials and community of practice will provide biomedical researchers, who use animal models in their research, the resources they need in the lab to help them develop well designed, properly controlled research studies that are more likely to be accurate and reproducible.

- **PUBLIC HEALTH RELEVANCE:** This project will develop teaching materials to support biomedical researchers who use, or plan to use, animals as research models. Research practices, particularly for animal studies, are learned within the laboratory setting and as such the quality of the training, and the expertise available, varies from lab to lab, leaving room for poor practices to develop in weak laboratory environments. These materials will provide a robust education module that focuses on the professional standards of practice related to designing, performing, analyzing and reporting biomedical studies that use animals as a research model and will be available on-demand for any researcher that wants to develop strong skills in designing animal studies and in reporting findings that are significant, and reproducible.

- **NIH Spending Category**: Alcoholism, Alcohol Use and Health; Clinical Research; Neurosciences; Substance Abuse

- **Project Terms**: Address; Affect; Animal Care and Use Committees; Animal Model; Animals; Applications Grants; base; Biological; Biomedical Research; career; Chemicals; Clinic; Clinical Trials; Community of Practice; Community Practice; Control Animal; design; Digital Libraries; Disease; drug candidate; Education; Educational process of instructing; Educational workshop; Ensure; Environment; experimental analysis; Experimental Designs; Fostering; Funding; Genetic; graduate student; Guidelines; Health; Human; insight; Institution; International; Intervention; Journals; Laboratories; Laboratory Procedures; Learning; Learning Module; Left; Literature; Manuscripts; Methodology; Methods; Modeling; Modification; online community; Organism; pedagogy; Peer Review; Play; Postdoctoral Fellow; Printing; Professional Ethics; Professional Organizations; Publications; Reporting; Reproducibility; Research; Research Design; Research Personnel; research study; Resources; Role; skills; skills training; Staging; Structure; Students; System; Teaching Materials; Techniques; Testing; Therapeutic; Therapeutic Intervention; Time; tool; Training; Translating; Trust; United States National Institutes of Health; web site; Work; Writing

ENHANCING DATA REPRODUCIBILITY THROUGH CELL AUTHENTICATION TRAINING

- **DESCRIPTION (provided by applicant):** "Enhancing Data Reproducibility Through Cell Authentication Training" aims to create an exportable online training module on cell authentication suitable to all postgraduate researchers who use cell lines in their research, including graduate students, postdoctoral scholars, and junior faculty members. NIH currently funds approximately 9,000 projects and sub-projects involving cell lines, at a total spend of \$3.7B. While the National Cancer Institute funds the bulk of these projects, all Institutes at the NIH and the FDA fund studies using cell lines. Cell line contamination and misidentification are major contributors to research reproducibility problems. Best practices exist to reduce these problems significantly and ensure more effective use of research dollars, and yet awareness and implementation of these practices remain low. We believe that an effective "active learning" training module and robust dissemination campaign will improve both awareness and implementation. This module will contain highly interactive training units including "how to" videos that will turn learning into practice by sending the trainees back into the laboratory to practice their skills, and tracking whether their training is incorporated into their future researh. We will

build partnerships and develop social media strategies to disseminate and promote this module to all postgraduate researchers who use cell lines in their work. The module will be optimized for use both by individuals accessing the material directly in an "online-only" format, and by trainees within Responsible Conduct of Research and other classroom settings using a "blended" format in which some materials are viewed independently online and others are used by groups in class. The total instruction time will be 2-4 hours. As a result of this training, knowledge of why and how to perform cell authentication will improve and research reproducibility problems involving cell lines will decrease. This project will save millions of dollars in wasted research expenditures and decrease translation time from bench to clinic to bedside.

- **PUBLIC HEALTH RELEVANCE:** Biological research using cell lines is often one of the first in a long series of steps from basic research to new treatments for disease. Research reproducibility in projects involving cell lines requires that those lines be what researchers think they are, and be free of contamination from other cells or microbes. As a result of the training program proposed in this application, researchers will learn how to test for and decrease the likelihood of cell line contamination and misidentification, and research reproducibility problems involving cell lines will decrease. This project will ensure more effective use of millions of dollars in research expenditures and decrease translation time from bench to clinic to bedside.

- **NIH Spending Category**: Cancer; Clinical Research; Health Disparities; Minority Health

- **Project Terms**: Active Learning; Awareness; Back; Basic Science; bench to bedside; biological research; Cell Line; Cells; Data; Development; Disease; Education Projects; Educational Curriculum; Effectiveness; Ensure; Expenditure; Faculty; Feedback; Foundations; Funding; Future; Generations; graduate student; Hour; improved; Individual; Institutes; Institution; Instruction; Knowledge; Laboratories; Learning; Link; Media Campaign; Medicine; member; Microbe; National Cancer Institute; Participant; programs; public health relevance; Recruitment Activity; Reproducibility; Reproducibility of Results; Research; Research Personnel; Resources; responsible research conduct; Series; skills; social media; Societies; Testing; Time; Training; Training Activity; Training Programs; Translations; United States; United States National Institutes of Health; wasting; Work

DEVELOPMENT OF REPRODUCIBLE INFORMATICS SKILLS AMONG MICROBIOME RESEARCHERS

- **DESCRIPTION (provided by applicant):** Enthusiasm for understanding how changes in the human microbiome affect health has led to an influx of researchers who have little experience in documenting and performing bioinfomatic analyses. Compounding this problem is that over the past 10 years microbiome datasets have grown from hundreds to millions of sequences. With the increased size of the datasets the complexity of the analyses has also grown. Researchers that once used spreadsheets to analyze their data now struggle to use command line tools. Traditional training programs have not been able to meet the needs of these researchers and so it is essential that instructional materials be developed to train these researchers on the best practices for documenting and disseminating their analyses so that they can be reproduced by others. The objective of this proposal is to develop a training module that microbiome researchers can use to improve the reproducibility and overall quality of their research. Aside from the general importance of insuring that all research is reproducible, the significant growth of the community makes it urgent that such instructional materials are developed now. The specific aim of the proposed effort will develop a set of autotutorials to teach microbiome researchers habits for engaging in reproducible research. This aim will be achieved through an iterative process of development, evaluation, and refinement using a wide network of microbiome researchers to assess the materials prior to broader dissemination. The expected outcomes of the proposed modules are the improvement of the reproducibility of research within the microbiome research community, increased accessibility to raw original data, and a greater use of literate programming tools for constructing manuscripts and oral presentations. Furthermore, if it is possible to improve the reproducibility of the original research, then it wil be more likely that other researchers will use those data, results, and

methods to perform additional analyses resulting in a greater understanding of the microbiome. Such a process is rare within the microbiome literature given the vast size of many of these datasets. The long-term goal of this project is to establish a broader community- supported resource devoted to disseminating best practices in performing reproducible microbiome research. Given the significant role of the microbiome in human health and the significant growth in our understanding of how it shapes health and disease, improving the reliability of the results from these studies will have a meaningful positive impact. This project will yield a significant vertica step in the field because it will put tools into the hands of researchers performing microbiome-focused studies empowering them to perform sophisticated and reproducible analyses. The approach taken in the proposed research is innovative because it represents the first concentrated effort to develop formal, public, and open training modules directed at the microbiome research community. Finally, we anticipate that the materials we develop for the microbiome research community will be easily disseminated across other bioinformatics research disciplines.

- **PUBLIC HEALTH RELEVANCE:** The proposed research is relevant to public health because it supports researchers within the domain of microbiome research, who have shown that it is impossible to separate human health from the structure and function of the human microbiome. Thus the research is relevant to the part of NIH's mission that pertains to the development, maintenance, and renewal of scientific resources that will assure our ability to perform robust and reproducible research in order to prevent disease.

- **NIH Spending Category**: Alcoholism, Alcohol Use and Health; Clinical Research; Complementary and Alternative Medicine; Dental/Oral and Craniofacial Disease; Substance Abuse

- **Project Terms**: Affect; analytical method; Area; Back; base; Bioinformatics; Biomedical Research; Communities; Computer software; Computers; Data; Data Analyses; Data Set; design; Development; Discipline; Disease; Documentation; Educational Curriculum; Educational process of instructing; Electronics; empowered; Ensure; Evaluation; experience; Goals; Growth; Habits; Hand; Health; Human; Human Microbiome; improved; Individual; Informatics; innovation; insight; Laboratories; learning materials; literate; Literature; Maintenance; Manuscripts; meetings; member; Methods; microbiome; Mission; Molecular Biology; Mus; novel; Oral; Outcome; Paper; Positioning Attribute; prevent; Process; programs; Public Health; Publications; Reagent; Reliability of Results; Reproducibility; Research; Research Personnel; research study; Research Support; Resources; Role; Scientist; Sequence Analysis; Shapes; sharing data; skills; Structure; Technology; tool; Training; Training Activity; Training Programs

Other available modules

The full collection is available at https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-dat aspx

Pragmatic and Group-Randomized Trials in Public Health and Medicine

- **Description:** "A free, 7-part, online course presented by Dr. David M. Murray that provides a detailed guide to designing and analyzing group-randomized trials (GRTs). The course includes video presentations, slide sets, suggested reading materials, and guided activities."

https://prevention.nih.gov/resources-for-researchers/nih-methods-training/grt

1. Introduction and Overview: "Part 1 provides an introduction and overview of the three kinds of randomized trials and their distinguishing characteristics. The differences between group-randomized trials (GRTs) and individually randomized group-treatment trials (IRGTs) are discussed." (25 minutes)
2. Designing the Trial: "Part 2 considers the design of group-randomized trials (GRTs), with a focus on internal and statistical validity, the factors that influence precision, and the major features that define

variations in group-randomized trial designs, including cohort vs. cross-sectional designs; matching, stratification, and constrained randomization; and time as a factor." (30 minutes)
3. Analysis Approaches: "Part 3 provides analysis approaches to group-randomized trials (GRTs), including a review of model-based approaches, randomization tests, marginal methods like generalized estimating equation (GEE), methods for complex survey samples, two-stage methods, methods for unbalanced designs, methods used with constrained randomization, and methods for individually randomized group-treatment trials (IRGTs)." (30 minutes)
4. Power and Sample Size: "Part 4 explores power and sample size for group-randomized trials (GRTs). The presentation focuses on Cornfield's two penalties of extra variation and limited degrees of freedom (df); strategies to reduce extra variation and increase df; the seven steps involved in any power analysis; estimating the intraclass correlation coefficient (ICC); and adaptations required for unbalanced designs. The course also provides an example of a sample size calculation for a simple GRT." (27 minutes)
5. Examples: "Part 5 provides examples of group-randomized trials (GRTs) from the Health Care Systems Collaboratory, a project funded by the NIH involving nine independent projects, eight of which are GRTs. All rely on electronic health records as their primary source of data, but they vary in their interventions, outcomes, and design and analytic methods." (20 minutes)
6. Review of Recent Practices: "Part 6 reviews recent practices in group-randomized trials (GRTs) based on a paper published by Murray, et al., in 2008 in the Journal of the National Cancer Institute. The presentation also explores recent practice[s] for individually randomized group-treatment trials (IRGTs) based on a paper published by Pals et al. in 2008 in the American Journal of Public Health. Both papers review published studies and critique the methods used for design, sample size calculation, and analysis." (20 minutes)
7. Alternative Designs: "Many alternative designs have been proposed to evaluate interventions delivered to or through groups, including multiple baseline designs, time series designs, quasi-experimental designs, stepped wedge designs, and regression discontinuity designs. Part 7 examines these designs in comparison to group-randomized trials (GRTs) and individually randomized group-treatment trials (IRGTs) with a focus on internal validity and cost in terms of time and participants." (20 minutes)

Each module includes:

- Video (YouTube)
- Video transcript
- PDF of presentation slides
- Suggested activities for the module (about 10 questions each that go with a journal article)
- Answers to suggested activities
- Participant feedback survey

NIH Rigor and Reproducibility Training Modules

- **Description**: "These modules, developed by NIH, focus on integral aspects of rigor and reproducibility in the research endeavor, such as bias, blinding and exclusion criteria. The modules are not meant to be comprehensive, but rather are intended as a foundation to build on and a way to stimulate conversations, which may be facilitated by the accompanying discussion materials. Currently, the modules are being integrated into NIH training activities."

Modules:

1. Lack of Transparency (3 minutes)
2. Blinding and Randomization (5 minutes)
3. Biological and Technical Replicates (2 minutes)
4. Sample Size, Outliers, and Exclusion Criteria (5 minutes)

Each module includes:

- Professional video (not just video of a lecture) (YouTube)
- Approximately two-page PDF with discussion points

- Starting points
- Lead-in questions
- Follow-up questions

edX Course: Principles, Statistical and Computational Tools for Reproducible Science

This was evidently funded as an extension of a training grant. https://www.edx.org/course/principles-statistical-computational-harvardx-ph527x

- **Description:** "Today the principles and techniques of reproducible research are more important than ever, across diverse disciplines from astrophysics to political science. No one wants to do research that can't be reproduced. Thus, this course is really for anyone who is doing any data intensive research. While many of us come from a biomedical background, this course is for a broad audience of data scientists. To meet the needs of the scientific community, this course will examine the fundamentals of methods and tools for reproducible research. Led by experienced faculty from the Harvard T.H. Chan School of Public Health, you will participate in six modules that will include several case studies that illustrate the significant impact of reproducible research methods on scientific discovery. This course will appeal to students and professionals in biostatistics, computational biology, bioinformatics, and data science. The course content will blend video lectures, case studies, peer-to-peer engagements and use of computational tools and platforms (such as R/RStudio, and Git/Github), culminating in a final presentation of a final reproducible research project. We'll cover Fundamentals of Reproducible Science; Case Studies; Data Provenance; Statistical Methods for Reproducible Science; Computational Tools for Reproducible Science; and Reproducible Reporting Science. These concepts are intended to translate to fields throughout the data sciences: physical and life sciences, applied mathematics and statistics, and computing. Consider this course a survey of best practices: we'd like to make you aware of pitfalls in reproducible data science, some failure - and success - stories in the past, and tools and design patterns that might help make it all easier. But ultimately it'll be up to you to take the skills you learn from this course to create your own environment in which you can easily carry out reproducible research, and to encourage and integrate with similar environments for your collaborators and colleagues. We look forward to seeing you in this course and the research you do in the future!"

What you'll learn:

- "Understand a series of concepts, thought patterns, analysis paradigms, and computational and statistical tools, that together support data science and reproducible research."
- "Fundamentals of reproducible science using case studies that illustrate various practices"
- "Key elements for ensuring data provenance and reproducible experimental design"
- "Statistical methods for reproducible data analysis"
- "Computational tools for reproducible data analysis and version control (Git/GitHub, Emacs/RStudio/Spyder), reproducible data (Data repositories/Dataverse) and reproducible dynamic report generation (Rmarkdown/R Notebook/Jupyter/Pandoc), and workflows."
- "How to develop new methods and tools for reproducible research and reporting"
- "How to write your own reproducible paper."

Course syllabus:

- Module 1: Introduction to Course Overview Introduction to faculty Project assignment
- Module 2: Fundamentals of Reproducible Science Why reproducible research matters Definitions and concepts Factors affecting reproducibility
- Module 3: Case Studies in Reproducible Research Potti 2006 Baggerly and Coombes 2007 Ioannidis 2009 Reproducible Reporting
- Module 4: Data Provenance Project design Journal requirements and mechanisms Repositories Privacy and security

- Module 5: Statistical Methods for Reproducible Science Prediction Models Coefficient of determination Brier score AUC Concordance in survival analysis Cross validation Bootstrap - Module 6: Computational Tools for Reproducible Science R and Rstudio Python Git and GitHub Creating a repository Data sources Dynamic report generation Workflows
- Course Conclusion Final Project: Write a reproducible report that could be submitted at a peer review journal