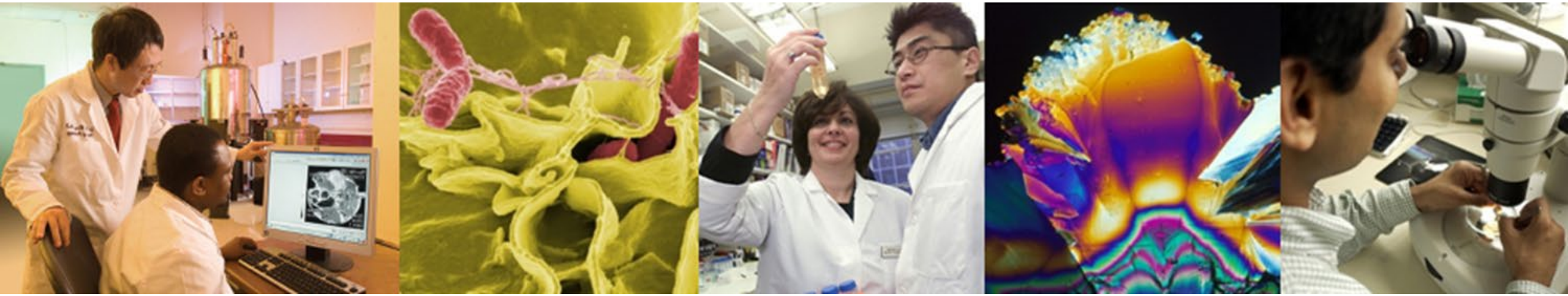
ACD Working Group on Enhancing Reproducibility and Rigor in Animal Research

Interim Report

120th Meeting of the Advisory Committee to the Director (ACD)

June 12, 2020



Barbara Wold, PhD
Bren Professor and Davis Leadership Chair
Director, Merkin Institute for Translational Research
California Institute of Technology

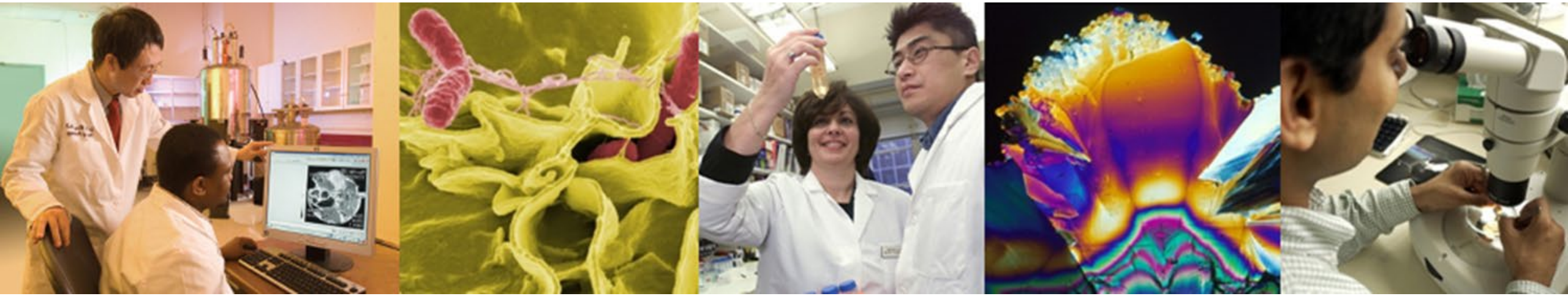
Lawrence A. Tabak, DDS, PhD
Principal Deputy Director, NIH
Department of Health and Human Services

ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research

Interim Report

120th Meeting of the Advisory Committee to the Director (ACD)

June 12, 2020



Barbara Wold, PhD
Bren Professor and Davis Leadership Chair
Director, Merkin Institute for Translational Research
California Institute of Technology

Lawrence A. Tabak, DDS, PhD
Principal Deputy Director, NIH
Department of Health and Human Services

Agenda

- **The charge**
- Interim Progress
- Next Steps

Charge to the Working Group (October, 2019)

- **Identify gaps and opportunities** to improve the rigor, reproducibility, translational validity, and transparency of animal models studies

- Evaluate **how animal models of human disease are currently developed, validated, and accepted into routine use**, and how this process could be improved

- Assess the current state of science for **validating alternative models** to animal research

Charge to the Working Group (October, 2019)

- Consider benefits and burdens of **registering animal studies** that aim to lead to first human trials

- Model **financial implications of potential changes in the average costs** of grants using animal models, the number of studies funded, or the need to develop consortia to achieve appropriate statistical power

- Consider how rigor in animal research is incorporated into **training**

ACD Enhancing Reproducibility and Rigor in Animal Research Working Group Members

EXTERNAL MEMBERS

Barbara Wold, PhD (Co-Chair)
Bren Professor, Davis Leadership Chair
Director, Merkin Institute
California Institute of Technology

Nancy Ator, PhD
Professor of Behavioral Biology
Johns Hopkins School of Medicine

Lais Berro, PhD
Postdoctoral Fellow
University of Mississippi Medical Center

Eliza Bliss-Moreau, PhD
Associate Professor; Core Scientist
University of California, Davis

Romer A. Gonzalez Villalobos, MD, PhD, FAHA
Senior Principal Scientist
Janssen Research and Development, LLC

F. Claire Hankenson, DVM, MS, DACLAM
Attending Veterinarian; Director, Campus Animal Resources; Professor
Michigan State University

Veronique Kiermer, PhD
Publisher & Executive Editor
PLOS

Keisa W. Mathis, PhD
Assistant Professor
University of North Texas Health Science Center

Sarah Nusser, PhD
Vice President for Research
Professor of Statistics
Iowa State University

Regina Nuzzo, PhD
Senior Advisor for Statistics
American Statistical Association

Eric Prager, PhD
Associate Director, External Affairs
Cohen Veterans Bioscience

F. Daniel Ramirez, MD, MSc
Cardiac Electrophysiology Fellow
CHU Bordeaux, IHU Liryc

Karen Svenson, PhD
Senior Scientific Program Manager and Research Scientist
Jackson Laboratory

ACD Enhancing Reproducibility and Rigor in Animal Research Working Group Members

USG MEMBERS

Lawrence A. Tabak, DDS, PhD (Co-Chair)
Principal Deputy Director, NIH

Brian Berridge, DVM, PhD, DACVP
Associate Director, National Toxicology Program;
Scientific Director Division National Toxicology Program
National Institute of Environmental Health Science, NIH

Paul Brown, PhD
Associate Director for Pharmacology and Toxicology
Office of New Drugs, Center for Drug Evaluation and Research, FDA

Janine Clayton, MD
Director
Office of Research on Women's Health, NIH

Joshua A. Gordon, MD, PhD
Director
National Institute of Mental Health, NIH

Michael Lauer, MD
Deputy Director for Extramural Research
Office of Extramural Research, NIH

Robyn Lee-Stubbs, MS, CPIA, PStat®
IACUC Chair/Statistician
United States Army Medical Research Institute of Chemical Defense

Glenn Merlino, PhD
Scientific Director for Basic Research
National Cancer Institute, NIH

Shai Silberberg, PhD
Director for Research Quality
National Institute of Neurological Disorders and Stroke, NIH

Carrie Wolinetz, PhD
Acting Chief of Staff; Associate Director
Office of Science Policy, NIH

Agenda

- The Charge
- **Interim Progress**
- Next Steps

Discussions Topics

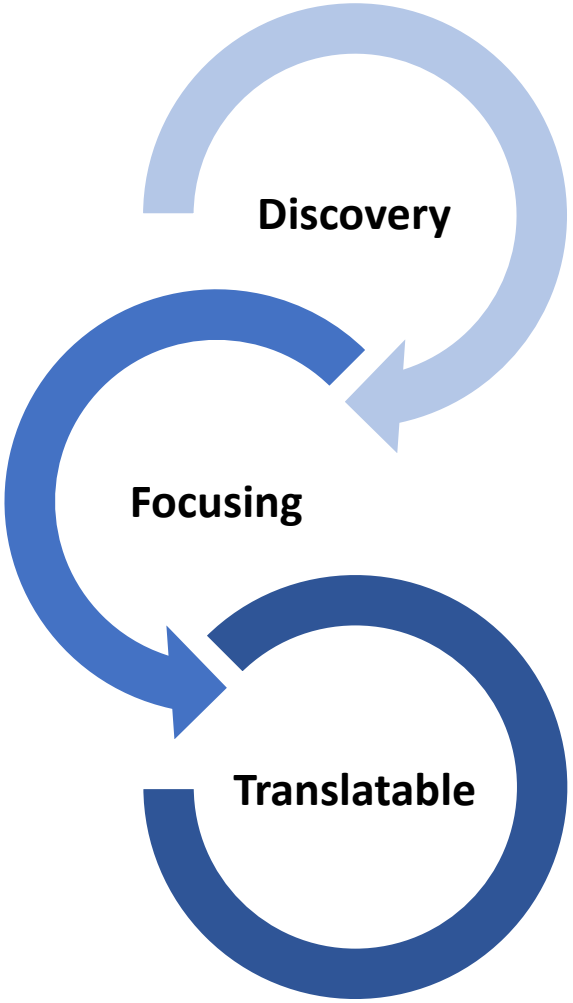
Major discussion foci:

- Statistical design and analysis
- Animal model evaluation
- Preregistration
- Translatability
- Research Culture

Active working group subcommittees:

- Vocabulary
- Registration, Checklists and Guidelines
- Financial Implications of any recommendations

Quality -> Rigor, Transparency, Reproducibility, and Translatability



- Scientific rigor and transparency are important across the research enterprise.
- Animal research has unique needs and reproducibility expectations through the stages of research
- Animal experiments often serve as the foundation for human clinical trials. Thus, there is a cost when translatability fails. When translation is the goal our approaches might differ.
  - Standards for design and analysis, including all statistical aspects
  - Preregistration
  - Follow-up reproducibility/replicability studies

Vocabulary and Statistical Analysis

Vocabulary for Reproducibility

F. Daniel Ramirez, MD, MSc
Cardiac Electrophysiology Fellow
CHU Bordeaux, IHU Liryc

Regina Nuzzo, PhD
Senior Advisor for Statistics
Communication and Media Innovation
American Statistical Association

We must avoid a black-and-white binary approach to reproducibility, especially when judging whether two results are “consistent” or “the same”

Vocabulary

- Reproducibility is important, but does not necessarily equate to rigorous research.
- Ultimately scientific rigor and transparency are the building blocks for getting consistent results across studies.
- It is important to recognize the interdependencies of rigor, transparency, and reproducibility. No one alone ensures translatability to humans.

Vocabulary

Methods Reproducibility	***Results Reproducibility***	***Inferential Reproducibility***
Providing enough detail about study procedures and data so the same procedures could be repeated.	Getting the same results from a new study with procedures as close to the original as possible.	Drawing similar conclusions or making knowledge claims of similar strength from study replications and re-analyses.
Transparency; a **prerequisite** for all else.	Similar to earlier definitions of **replicability.** Understand the expectations. May be hard to achieve for good reasons.	The process by which **a scientific field decides** which research claims or effects are "**true**."

Goodman, Steven N., Daniele Fanelli, and John PA Ioannidis. "What does research reproducibility mean?." *Science translational medicine* 8.341 (2016): 341ps12-341ps12.

Statistical Analysis

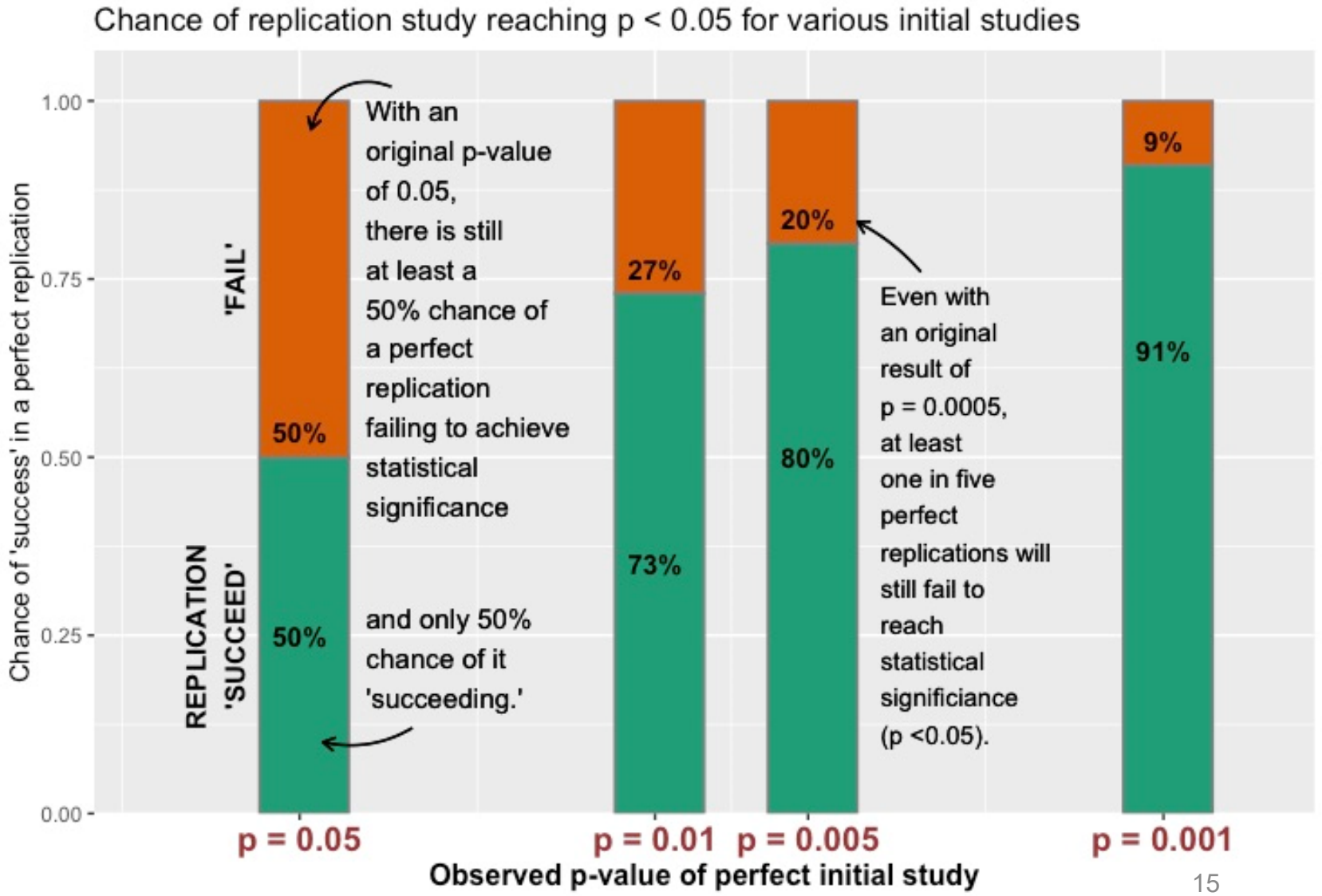
All three types of reproducibility are important.

- To achieve **methods** reproducibility, need:
  - **Widespread transparency** of study protocol, experimental methodology, data, and other details standardized for the field.
- To achieve **results** reproducibility, need:
  - The expectation that **perfect replications may fail** to achieve statistical significance.
  - Attention to **study design, power,** and **statistical analyses.**
  - The **evaluation of cumulative evidence**, not a single study.
- To achieve **inferences/conclusions** reproducibility, need:
  - **Complete reporting** of all analyses and results
  - Minimize the file-drawer problem, **publication bias**, cherry-picking, p-hacking.

Statistical Analysis

Statistical significance is not enough to judge reproducibility.

Given a statistically significant initial study, the chance of a replication "succeeding" (another statistical significance; p < 0.05) is surprisingly low.



For details, see: Goodman, Steven N.. *Statistics in medicine* (1992).

Selection of Models

Experimental Systems: Models for, not of

Joshua A. Gordon, M.D., Ph.D.
Director, NIMH

NIH National Institute of Mental Health

Animal Models in NIGMS-Funded Sepsis Research

Jon R. Lorsch, Ph.D.
Director, NIGMS

NIH National Institute of General Medical Sciences

Selection of Models: animal and non-animal

Evaluate model utility:

- Models *for* human Disorders
- How well does the model address the question of interest?
- For translational work, use readouts specific for the translational diagnostic or therapeutic goal

Implementing culture change around Selection of Models:

- Enable critical analysis of models including in peer review. This activity does not fall to one group but requires a domain experts.
- Incentivize making new models when current ones cannot address the question of interest.

Changing models – Sepsis Research

- Sepsis is a rapidly developing, progressive, heterogeneous disorder
- Perceived problems facing sepsis research
  - Ambiguity about how best to define the clinical syndrome(s)
  - Fading consensus about best preclinical models
  - Dozens of failed clinical trials - Translatability
- NIGMS convened a Working Group of their Advisory Council and issued an RIF to develop strategies for advancing sepsis research supported by NIGMS
  - Included examining the utility of current animal models

Changing models – Sepsis Research

- Recommendations to improve model system
  - Support development of models that mimic (1) non-immunological aspects of sepsis and (2) major co-morbidities in human sepsis.
  - Encourage use of human clinical material to confirm observations in nonhuman models.
  - Support discovery science, computational, and cell-culture and organoid-type methods in preclinical sepsis research.
  - Evaluate models by their ability to provide readouts relevant to translation of new diagnostic methods and therapies.
  - Cease support for using a poorly predictive model. Educate study sections
- The process, in this case organized by an NIH institute, brought a range of domain experts together to evaluate the field and then acted to improve model selection

Changing models – Sepsis Research

I. NIGMS assembled a domain expert working group

II. WG Recommendations to improve the model

- Support development of better models that mimic
  1) non-immunological aspects of sepsis
  2) major co-morbidities in human sepsis
- Encourage use of human clinical material to confirm nonhuman model results.
- Support discovery science, computational, and cell-culture and organoid-type methods in preclinical sepsis research.
- Evaluate models by their ability to provide readouts relevant to translation of new diagnostic methods and therapies.

III. Action to move funded research away from unsuccessful model/use pairing

Selection of Models

(Move over, Mice!): PhysioMimetics: Integration of Organs-on-Chips with Systems Biology to Humanize Drug Development

Linda G. Griffith,
Director, Center for Gynepathology
School of Engineering Professor of Teaching Innovation, Biological Engineering, and Mechanical Engineering Research
Massachusetts Institute of Technology

MIT BE
BIOLOGICAL ENGINEERING

Developing, evaluating, adopting better models

- Increase emphasis on deep molecular / cellular / tissue / clinical phenotyping of patient populations, guided by
- Multi-scale “systems biology” conceptualization to better capture the complexity and reality of disease.
  - Patient stratification according to mechanistic hypotheses
  - Integrate biophysical and dynamic phenomena with “big data” (beyond “AI”)
- Non-animal alternatives are critical to consider and continue to evolve. In addition to above:
  - More intense interdisciplinary development and democratization of human “microphysiological systems technologies”
  - Design principles for complex in vitro (non-animal) disease models

Effects and limitations of preregistration

Preregistration as a tool to increase transparency, rigor, and reproducibility of preclinical animal studies

Tim Errington, Center for Open Science

CENTER FOR OPEN SCIENCE

Tim Errington

Director of Research | Metascience

Effects of preregistration

What is preregistration?

- Time-stamped, read-only version of research plan
  - Hypotheses
  - Sampling plan
  - Variables
  - Design plan
  - Analysis

Purposes:

- Discoverability: Study exists
- Interpretability
  - Distinguish exploratory and confirmatory approaches
  - Clear answers require clear questions
  - Exploratory research is allowed and encouraged
- Improves the quality and transparency of your research.

Effects of preregistration

- Requires additional administrative action by the research teams
- If incorporated into the NIH grant application process policy decisions would be needed
  - What type/stage of research needs preregistration?
  - Optional or mandatory?
  - Role in review scoring?
  - Minimum required elements?
  - Tools to preserve intellectual capital e.g. embargo, moderated request

Impact on NIH resources

- Subcommittee led by Mike Lauer to examine financial and resource implications of possible ACD WG changes:
  - Changes in sample size, research organism, or alternative approaches
  - Other changes toward rigor and transparency, even if sample size and organism are unchanged
  - Scientists may abandon research because of new requirements, changing the NIH portfolio
  - Greater use of Contract Research Organizations
  - Costs of increased involvement and incentives for statisticians

Impact on NIH resources

- Anticipated impact on NIH resources and infrastructure:
  - Increased resources to design and document research activities
  - Costs associated with replication studies
  - Costs associated with preregistration, data management, curation, storage, and sharing
  - Costs of meta-analyses and standardization
  - Added activities and staff in review (statistical expertise)

Impact on NIH resources

- First, examine a random set of grants and associated administrative data to determine:
  - Assess proposals and publications for elements of rigor (randomization, blinding, sample size, data management)
  - Were methods described in grants similar to those in resulting papers?
  - Attempt to estimate financial implications of enhanced rigor and other possible measures (e.g. pre-registration)
- Obtain feedback from others who have looked at this, such as the Center for Open Science, other grant funders, and professional societies

Request for Information

- Request for Information (RFI) inviting comments and suggestions on Enhancing Rigor, Transparency, and Translatability in Animal Models Involved in Biomedical Research.
- Three focus areas with specific questions and an online submission portal.
  - *Rigor and Transparency*
  - *Optimizing the Relevance to Human Biology and Disease*
  - *Research Culture*

Refining Themes

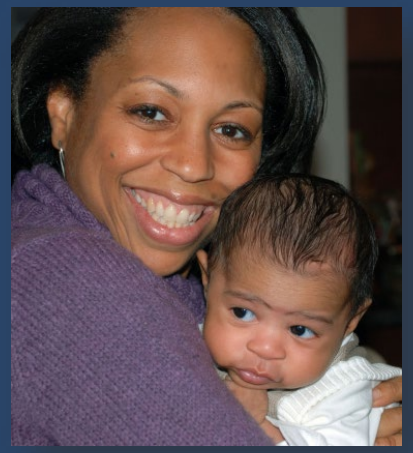
- **Selecting or developing** the most appropriate animal (or other) model for human disease to address the question of interest
- Strengthening experimental **design and analysis**
- The impact of animal **care and husbandry** on experimental outcomes
- Enhancing **transparency**
- Training and continuing **education**, including **vocabulary**
- **Measuring and evaluating** effects of any interventions
- Tackling the **cultural incentives** to keeping the status quo

Agenda

- The Charge
- Interim Progress
- **Next Steps**

Timeline

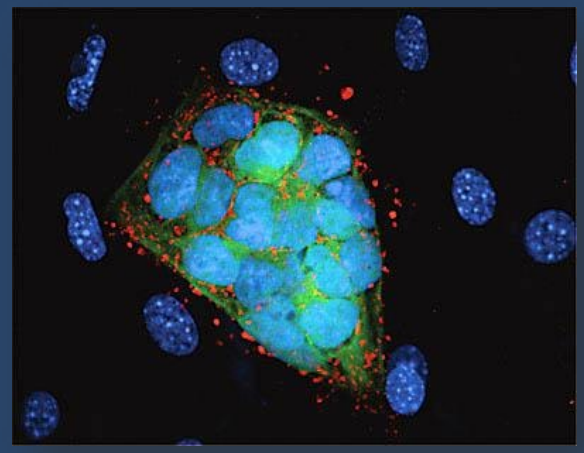
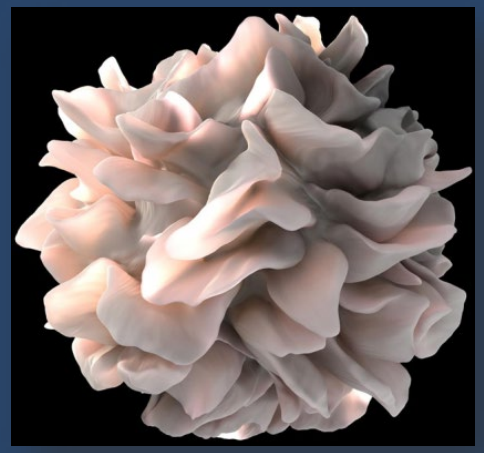
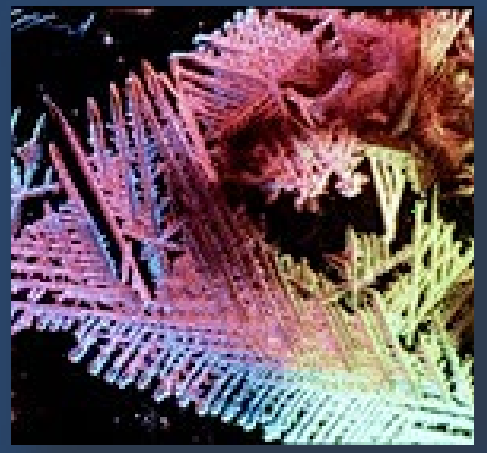
- October/November 2019: Kickoff meetings
- December 2019: Interim report to ACD
- January 2020 - April 2020: In-person and teleconference meetings
- **June 12 2020: Interim report to ACD**
- June 2020 - November 2020: In-person and teleconference meetings
- December: Final recommendations to ACD




NIH…

Turning Discovery Into Health

Lawrence.Tabak@nih.gov

Changing a Research Culture

POLICY
PROVIDING REASONS TO EMBRACE CHANGE

METASCIENCE
ACQUIRING EVIDENCE TO ENCOURAGE CHANGE.

COMMUNITY
FOSTERING CONNECTION AND INCLUSION TO SPREAD CHANGE

TRAINING
DISSEMINATING KNOWLEDGE TO ENACT CHANGE

INFRASTRUCTURE
BUILDING TECHNOLOGY TO ENABLE CHANGE.

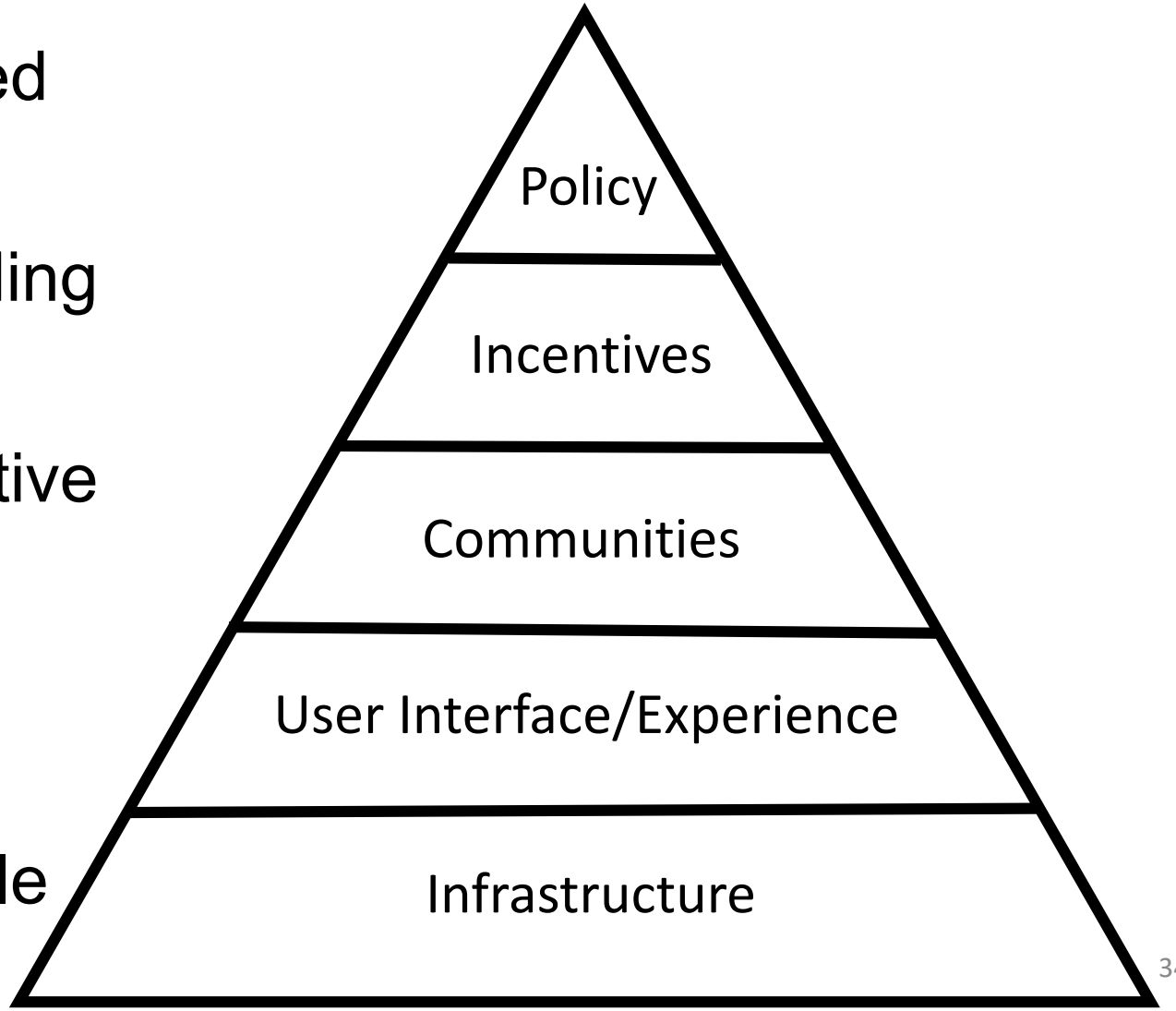
Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible



Vocabulary

- There are differences between reproducibility and replicability, but those words are often used interchangeably. There is no clear standardized definition.
- Achieving reproducibility is thus a difficult concept to discus and has unique changes.

Methods Reproducibility	**Providing enough detail about study procedures and data** so the same procedures could in theory or in actuality be repeated
Results Reproducibility	**Getting the same results** from a new study with procedures as close to the original as possible
Inferential Reproducibility	Drawing "**qualitatively similar conclusions**" or making "**knowledge claims of similar strength**" from a study replication or reanalysis

Adopted from
STEVEN N. GOODMAN, DANIELE FANELLI, JOHN P. A. IOANNIDIS
*SCIENCE TRANSLATIONAL MEDICINE*01 JUN 2016 : 341PS12

Statistical Analysis

- ***Statistical significance is not enough.*** Even perfect replication studies of perfect original studies have a surprisingly low chance of reproducing statistical significance.
  - For a perfect original study with **p = 0.049** (statistically significant), a perfect replication study getting **p < 0.05** has **only a 50% chance**.
  - Original study **p = 0.01**; replication with **p < 0.05** has **73% chance**.
  - Original study **p = 0.001**; perfect replication with **p < 0.05** has **91% chance**.

Statistical Analysis

- ***Even continuous p-values aren't enough.***
  - The chance of *any* replication getting the same p-value as the original is ½.
  - Suppose the replication p-value is the same but the effect is in the opposite direction -- is this a successful replication?
  - Suppose the p-value is the same but the effect size is much smaller -- does this count?
- ***Researchers need to use prior knowledge to decide whether two effect sizes are "close enough."***
  - Two studies, one with increased survival of 200% and another with 0.02% -- is this a replication? What about 10% and 15% -- is that always close enough?