NIH Initiative to Enhance Reproducibility and Transparency of Research Findings

Update for National Advisory Council on Drug Abuse

September 4, 2013

Elena Koustova, PhD, MBA
Director, Office of Translational Initiatives and Program Innovations (OTIPI)
Presentation Goals

- Discuss reproducibility and transparency of research findings with the Council, alert to the issues, inform about NIH and NIDA efforts, and solicit and gather feedback.

- Understanding of the inherently difficult nature of human disease/subject of biological research, the complex data sets generated, the shifting basic meaning of scientific knowledge, the limitations of preclinical tools: advances through trial, error and revision is simply a cost of science.

- Ability to translate research to clinical success has been remarkably low - Focus on foundational work that leads to costly clinical trials.
An Epidemic of False Claims

Competition and conflicts of interest distort too many medical findings

False positives and exaggerated results in peer-reviewed scientific studies have reached epidemic proportions in recent years.

- 45 positive randomized controlled trials from 3 major medical journals (led to the spread of treatments such as hormone replacement therapy for menopausal women and daily low-dose aspirin to prevent heart attacks and strokes) analyzed in 2005, JAMA.
- Of the 34 claims that had been subject to replication, 41% had either been directly contradicted or had their effect sizes significantly downgraded.
- In genetic studies of sex differences, out of 432 claims, only a single one was consistently replicable.

The unspoken rule is that at least 50% of the studies published even in top tier academic journals – Science, Nature, Cell, PNAS, etc... – can’t be repeated with the same conclusions by an industrial lab. In particular, key animal models often don’t reproduce. This 50% failure rate isn’t

Epidemics of replication can increase the value of the company.
A Matter of Increasing Concern

- Costs of drug development have increased along with the number of late-stage clinical-trial failures
- Demand for more effective therapies
- Low success rate is neither sustainable nor acceptable
- “Failure to repeat” puts biotech out of business
- VC reluctance to invest into early stages and to fund academic spin-outs
- Waste of limited research funding
- May erode public support for research
- ...Problem is threatening the reputation of NIH which funds many of the studies in question...
- ...These finding come at a time of growing lack of respect for research institutions and bode ill for the public’s support of tax dollars for the scientific establishment.
September 2011  Bayer HealthCare finds inconsistencies between in-house and published data in almost two-thirds of 67 projects. Replication rate - \(\sim 25\%\)

March 2012  Amgen publication shows that the findings from only 6 of 53 landmark papers can be replicated by company scientists. Replication rate - \(\sim 11\%\)

- contact the original authors, repeat experiments repeat experiments under the authors' direction, in the laboratory of the original investigator, CDAs
- close attention to controls, reagents, investigator bias and describing the complete data set
- no blinding, published Fig not reflective of the entire data set

May 2013  A survey at the MD Anderson Cancer Center finds that more than half of its respondents have tried and failed to reproduce published data.

July 2013  Science Exchange launches a verification program for commercially sold antibodies.

July 2013  Evaluation of excess significance bias in animal studies of CNS disorders reveals that twice as many studies as expected report statistically significant conclusions.
Why don't the data hold up?

- Incorrect statistical methods, regression to the mean = as the experiment is repeated an early statistical fluke gets cancelled out;
- “Significance chasing,” or finding ways to interpret the data so that it passes the statistical test of significance;
- Failure to work to industry standards; lack of industry perspective;
- Potential lack of true independence between academic groups;
- “Shoehorning” process= subtle omissions and unconscious misperceptions;
- Publishing bias, whereby journals are more likely to publish positive and novel findings than negative or incremental results;
- Competition in “hot” fields and rushing findings into print;
- Overemphasis on publishing conceptual breakthroughs in high-impact journals;
- The number of investigators, experiments, analyses increased exponentially (50% never cited) - adequate safeguards against bias, etc. are lacking;
- Research conducted for reasons other than pursuit of truth (academic promotions, “publish or perish”);
- Lack of incentives to retract irreproducible findings;
- Investigator prejudice;
- Outright fraud
Problem

- Lack of reproducibility and transparency of published research findings.
- Underlying issues:
  - Poor education
  - Poor evaluation
  - Perverse reward incentives for academic scientists
Addressing the Issues

0 Two NIH Workshops -> Nature, 2012, 490, p.189-191
0 Ad-hoc group met to develop approaches.
   0 Discussions were also informed by existing IC efforts (e.g., NCI, NINDS).
0 Group came to a consensus on guiding principles to address the underlying issues.

A call for transparent reporting to optimize the predictive value of preclinical research


The US National Institute of Neurological Disorders and Stroke convened major stakeholders in June 2012 to discuss how to improve the methodological reporting of animal studies in grant applications and publications. The main workshop recommendation is that at a minimum studies should report on sample-size estimation, whether and how animals were randomized, whether investigators were blind to the treatment, and the handling of data. We recognize that achieving a meaningful improvement in the quality of reporting will require a concerted effort by investigators, reviewers, funding agencies and journal editors. Requiring better reporting of animal studies will raise awareness of the importance of rigorous study design to accelerate scientific progress.
Principles for Addressing Issues

1. Raise community awareness.
2. Enhance formal training.
3. Improve the evaluation of applications.
4. Protect the integrity of science by adoption of more systematic review processes.
5. Increase stability for investigators.

NIH and NIDA efforts to implement those principles
Principle 1: Raise community awareness

- NIDA will issue a Notice (NOT) in the NIH Guide “Improving Reporting of Research Methods and Results in Translational Addiction Research Involving Animals is a NIDA Commitment”.
- NIDA will include relevant language from the NOT when it issues new FOAs.
- NIDA OTIPI and OEA will develop workshops for upcoming scientific meetings (e.g., College on Problems of Drug Dependence; Society for Neuroscience) to alert the drug abuse and addiction research community on the issues and NIDA’s efforts to address them.
- NIDA Director’s blog on this topic will be disseminated by various social media channels (e.g., Facebook, Twitter).
- Will welcome NIDA Council’s feedback.
Principle 2: Enhance formal training

- Integrate modules and/or courses on experimental design into existing required training courses and award terms and conditions.

  **Action:**

  - NIH: OIR will create and pilot a new module on research integrity as it relates to experimental biases and study design to ethics training course required for NIH intramural fellows.
  
  - NIDA: solicited Small Business Innovation Research (SBIR) proposals to create a fee-for-service “bundle” to guide, assist or educate the addiction and other life science investigators on the design, execution and interpretation of animal experiments.
2014 SBIR Contact RFP posted on Aug 29, 2013

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (HHS), THE NATIONAL INSTITUTES OF HEALTH (NIH) AND THE CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC) SMALL BUSINESS INNOVATIVE RESEARCH (SBIR) PROGRAM

PROGRAM SOLICITATION PHS 2014-1

Closing Date: November 13, 2013, 4:30PM Eastern Time

Participating HHS Components:

- The National Institutes of Health (NIH)
- The Centers for Disease Control and Prevention (CDC)

National Institute on Drug Abuse (NIDA)

151 Web Resource System for Prescription Drug Providers, Researchers and Users: The Prescription Drug Abuse Policy System (PDAPS)

152 Technological Tools to Facilitate Implementation of Evidence-Based Substance Abuse Prevention Interventions among the Military

153 Products to Prevent (Lethal) Drug-induced Respiratory Depression

154 Bundled Service for Designing Methodologically Rigorous Animal Studies
Principle 3: Improve the evaluation of applications

- Consider the use of guidelines and/or checklists to systematically evaluate grant applications.

**Action:**
- NIDA will be one IC to pilot the use of a checklist to enhance systematic review of applications.
- NIDA’s OEA will dedicate additional time to train/prepare reviewers to apply the criteria outlined in NOT to grant applications and to encourage the reviewers to consider aspects important for research reproducibility.
- NIDA will pilot the use of a checkbox for program staff to indicate whether applicants have included information related to the reporting standards. If not, they will ask for a description of the challenges in addressing the reporting requirement.
NIDA Notice: Improving Reporting of Research Methods and Results in Translational Addiction Research Involving Animals is a NIDA Commitment

- NIDA's mission is to lead the Nation in bringing the power of science to bear on drug abuse and addiction. To make meaningful and powerful research progress, foundational data upon which new advances will hinge must be reliable and reproducible. This is especially important in order to enable translation of preclinical findings into human applications intended to facilitate the development of new therapies. Toward this end, NIDA is committed to the support of translational studies involving animals marked by transparency in reporting on the design, conduct and analysis of experiments.

- With this Notice, NIDA aims to heighten awareness to this commitment in the drug abuse and addiction research communities. Investigators proposing translational studies involving animals are encouraged to address a core set of research parameters/reporting standards including animal selection, sample size estimation and statistical methods, as well as randomization, blinding and data handling in their applications. Transparency in the reporting of these parameters will facilitate reproducibility among laboratories and ultimately the translation of findings to the clinic.

- To help insure that investigators are informed of NIDA’s commitment to transparency in reporting research methods and results in the translational studies involving animals, the information contained herein will be included in all appropriate FOAs on which NIDA is the lead Institute. Investigators seeking NIDA grant support are urged to discuss these issues with program staff prior to submission of applications.
Possible Review Checklist

Council participation welcome

BOX 1  A core set of reporting standards for rigorous study design

Randomization

- Animals should be assigned randomly to the various experimental groups, and the method of randomization reported.
- Data should be collected and processed randomly or appropriately blocked.

Blinding

- Allocation concealment: the investigator should be unaware of the group to which the next animal taken from a cage will be allocated.
- Blinded conduct of the experiment: animal caretakers and investigators conducting the experiments should be blinded to the allocation sequence.
- Blinded assessment of outcome: investigators assessing, measuring or quantifying experimental outcomes should be blinded to the intervention.

Sample-size estimation

- An appropriate sample size should be computed when the study is being designed and the statistical method of computation reported.
- Statistical methods that take into account multiple evaluations of the data should be used when an interim evaluation is carried out.

Data handling

- Rules for stopping data collection should be defined in advance.
- Criteria for inclusion and exclusion of data should be established prospectively.
- How outliers will be defined and handled should be decided when the experiment is being designed, and any data removed before analysis should be reported.
- The primary endpoint should be prospectively selected. If multiple endpoints are to be assessed, then appropriate statistical corrections should be applied.
- Investigators should report on data missing because of attrition or exclusion.
- Pseudo replicate issues need to be considered during study design and analysis.
- Investigators should report how often a particular experiment was performed and whether results were substantiated by repetition under a range of conditions.
Principle 4: More systematic review processes

- Collaborate further with scientific journals and the scientific community on efforts to improve rigor.

**Action:**

- NIH will continue outreach to Journals to partner with them to determine value of recently adopted reporting guidelines.
- NIH will evaluate the PubMed Commons Community Response Effort, which is a pilot program testing options for scientists to post online comments on original research articles.
Nature and family announces new editorial measures

- http://www.nature.com/nature/focus/reproducibility/index.html

- Checklist to disclose technical and statistical information and experimental and analytical design elements

- Reporting of key methodological details: precise characterization of key reagents, such as antibodies and cell lines

- Will abolish space restrictions on the methods section

- Authors to provide, in tabular form, the data underlying the graphical representations used in figures

- More precise descriptions of statistics; Nature journals will now employ statisticians as consultants
Principle 4: More systematic review processes

- NIDA will conduct outreach to Journals in addiction area to partner with them to improve research reproducibility

- Publishers invited to workshops (College on Problems of Drug Dependence; Society for Neuroscience) be developed by NIDA OTIPI and OEA
Principle 4: More systematic review processes

- Consider the advisability and approach to supporting replication/reproducibility studies or centers.

**Action:**

- Select ICs will pilot additional use of supporting replication studies.
- Evaluate ongoing efforts by NINDS.
  - Pilot work has been done by NINDS in supporting replication studies.
- Evaluate ongoing efforts by NIA.
  - NIA is currently supporting the Interventions Testing Program, where preclinical studies are conducted with multi-site duplication, rigorous methodology and statistical analysis.
- NIDA: in DPMC’s MDP, internal go/no go only after blind claim testing by independent contractor.
We plan to discuss how NIDA can support the validation of NIDA's researcher's studies.
Principle 5: Increase stability for investigators

Adapt NIH bio-sketch to allow investigators to place their work into a functional context.

**Action:**

- Select ICs will perform pilot evaluations of changes to bio-sketch to include elements that aid in framing the PIs work and describing the applicant’s contribution to the publications cited.
- Select ICs will also pilot additional experiments to reduce “perverse incentives”.
- Efforts by NCI to reduce “perverse incentives” will be evaluated.
  - NCI recently developed an Outstanding Investigator Award to address perverse incentives by providing substantial, longer-term support to experienced investigators.
Conclusions

- Addressing the systemic issues of irreproducibility and reporting transparency will require tremendous commitment and a desire to change the prevalent culture.
- Perhaps the most crucial element for change is to acknowledge that the bar for reproducibility in performing and presenting discovery and preclinical studies must be raised.
- Ensuring systematic attention to reporting and transparency is only a small step toward solving the issues of reproducibility that have been highlighted across the life sciences.
- Tackling these issues is a long-term endeavor that will require the commitment of NIH, academic institutions, researchers and publishers.
- NIH and NIDA are committed to leading the community discussions on this topic. We urge the community to improve research reproducibility.
Deficient Experimental procedures

Education

Good Experimental design

Lack of Transparency in Reporting

Review/Evaluation

Transparency in Reporting

Publication of negative outcomes

Publication Bias

Culture

Poor Reproducibility

Better Reproducibility

NIH Meeting, December 13, 2012
NIH Initiative to Enhance Reproducibility and Transparency of Research Findings

Update for National Advisory Council on Drug Abuse

September 4, 2013

Elena Koustova, PhD, MBA
Director, Office of Translational Initiatives and Program Innovations (OTIPI)
# Trends in attrition rates

<table>
<thead>
<tr>
<th>Reasons for Attrition (~ %), Ph II and III</th>
</tr>
</thead>
<tbody>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Safety</td>
</tr>
<tr>
<td>Efficacy</td>
</tr>
</tbody>
</table>

Nature RDD 3, 711-716 (August 2004)  
Nature RDD 12, 569, (August 2013)
Trends in attrition rates

Of the 148 failures between Phase II and submission in 2011 and 2012, reasons were reported for 105; the majority of failures were due to lack of efficacy, as shown on the left. On the right, the 105 reported failures are broken down according to therapeutic area. Comparison of the reasons for failures in Phase II and Phase III trials in 2011 and 2012 with those in earlier periods that we reported previously (see main text for details). Data are from Thomson Reuters, Drugs of Today © Prous Science S.A.
a) Pie chart showing distribution:
- Anticancer (n=23): 28%
- Nervous system (n=15): 19%
- Alimentary and/or metabolism (n=11): 18%
- Anti-infectives (n=11): 13%
- Cardiovascular (n=7): 13%
- Other (n=16): 8%

b) Pie chart showing financial and/or commercial:
- Safety (including risk-benefit) 21%
- Efficacy:
  - Versus placebo: 32%
  - As add-on therapy: 29%
  - Versus active control: 5%
- Not disclosed 66%
Correlation between impact factor and retraction index.

Understanding Data Irreproducibility: What could be done to improve Research Data Quality?

- **Challenge**: NIDA challenges the American research and non-research communities
  - to explore the possible causes of scientific data irreproducibility and the sources of experimental variation;
  - to identify the negative, potentially addressable, systemic causes of that irreproducibility and
  - to propose the ways to tackle those damaging causes.

- **Solution Type**: NIDA is interested in theoretical solutions, not literature reviews. For each identified cause, obvious and non-obvious, the analysis must be presented that demonstrates, in theory or in practice, a negative impact of that cause on data reliability. The solution for the identified negative cause problem should be provable, that is, the successful problem solver must find a way to justify their train of thought leading to the proposed solution. The solvers can identify and propose a solution to one problem contributing to biological research data irreproducibility, or to identify and propose a solution to multiple problems.

- **Prize**: $10,000
Collins Sets Five Themes for NIH, 2009
- translating research into medicine
- health care reform

NIH’s mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce the burdens of illness and
"Judging by the only criterion that matters to patients and taxpayers—not how many interesting discoveries about cells or genes or synapses have been made, but how many treatments for diseases the money has bought—the return on investment to the American taxpayer has been approximately as satisfying as the AIG bailout" (Newsweek Cover, May 14, 2010).

The goals of the Agency are as follows: expand the knowledge base in biomedical and associated sciences in order to enhance the Nation’s economic well-being and ensure a continued high return on the public investment in research; NIH funding is subject to the need to demonstrate public benefit", and “academic scientist have a social contract … to create a useful product or service”, Science Translat Med, 2012

Small Business Innovation Research (SBIR) Program
Challenges and Prize Competitions
Involvement and Contributions of NIDA IRP
ACCENTUATE THE POSITIVE

A literature analysis across disciplines reveals a tendency to publish only ‘positive’ studies — those that support the tested hypothesis. Psychiatry and psychology are the worst offenders.

Proportion of papers supporting tested hypothesis
21. Chrispy on March 31, 2011 2:29 PM writes...

One of my biggest shocks, upon coming to a large biotech/pharma from an academic background, was how little of what was published is reproducible. Time and time again, results (especially animal results) in top-tier journals cannot be reproduced. At first, I suspected that people at my new company were incompetent. Now I realize that they are simply more realistic.

3. Iridium on March 31, 2011 9:40 AM writes...

At my small start-up we currently use the mouse P6/2 model for some of our neuro-D programs. To date, every assay in our hands as a 'gold standard' model have never given us the expected results. I believe the reason is secondary to the chemical and/or chemical model have been quite real based.

27. Anonymous on March 31, 2011 6:49 PM writes...

The same lax training in academics slowly transfers into industry. CRO's can be dicey, they get paid per compound and they know that it's too much work to double check everything. Process work is different because they are on the hook for everything.

I actually blame the erosion of scientific integrity in academics. Look in the Supp info, 10 mg reactions GC yields (calibrated) do I see the GC traces.. no

The fault lies with the current system which forces people to play this dirty game. Hence you see the most competitive countries suffer from dishonestly the most.

As a scientist I find that we are actually causing the majority of our progress because we publish highly reproducible

2. You're Pfizered on March 31, 2011 9:30 AM writes...

This becomes more of an interesting factoid given how the pharmaceutical industry is doing more and more collaborating with academic groups on early stage stuff.

I'm sure every company has folks out there scouring the academic labs for the next hot target and/or technology. Won't they likely have the most actual reproducible results if they collaborate with academic researchers?

1. Hap on March 31, 2011 9:21 AM writes...

With academics patenting more as well (universities want the licensing money), "Follow the money" might be the appropriate aphorism. People will do whatever gets them paid (or power, or status), in business or academics.

It would be helpful to have data (rather than examples) or irreproducibility - if it is happening, it could be "optimistic results" from the authors (bias), insufficiently specified protocols in the papers, or something else, which would help to figure out what to do about it.
10. BioBrit on March 31, 2011 10:37 AM writes...

I heard this suggestion a few days ago (I can't take credit myself). We, as in the journals, insist on standardized chemistry elucidation experiments - NMR, elemental analysis etc, in order to publish. This helps us compare across research labs - like vs like. A (worthy but lofty) goal would be to insist on the same in the biological sciences - standardized assays and animal models. That would help us compare like vs like. Of course there would have to be a place for method development papers, just as there is in chemistry. But, want to publish your brand new ligand for a disease? New pathway hypothesis? You'd better demonstrate it using standard models within defined experimental criteria. I'm sure many would claim this in unattainable and would inhibit their research. May be true, but worth considering. Getting all those elemental analysis was a pain in the butt for me too.

28. Anonymous on March 31, 2011 6:51 PM writes...

Here is the solution:

Publish irreproducible results in Paper, results in banning of all authors from publishing in that journal again.

35. HelicalZz on April 1, 2011 8:59 AM writes...

There have been a few comments here related to publication, and bans or retractions based on work hard to reproduce. I'd say again that I don't think the problem is an insidious one, but rather more often the result of internally optimized models. Results reproduce, but not not 'as well as reported'.

As I noted in the linked blog, publication may be a solution in select cases. A journal that requires (or itself seeks) CRO confirmation of research results prior to publication. That is, replaces part of peer review with contracted confirmation. This would be expensive, and not a model for all, or even most, research. It could however be worthwhile to publish under this type or model when you are providing research that is looking for outside funding. [Whether or when one would want to publish such research is an entirely different topic].
Several approaches were used to reproduce the published data

<table>
<thead>
<tr>
<th>Model reproduced 1:1</th>
<th>Model adapted to internal needs (cell line, assays)</th>
<th>Literature data transferred to another indication</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-house data in line with published results</td>
<td>1 (7%)</td>
<td>12 (86%)</td>
<td>0</td>
</tr>
<tr>
<td>Inconsistencies that led to project termination</td>
<td><strong>11 (26%)</strong></td>
<td><strong>26 (60%)</strong></td>
<td>2 (5%)</td>
</tr>
</tbody>
</table>

- Irreproducibility was high both when Bayer scientists applied the same experimental procedures as the original researchers and when they adapted their approaches to internal needs (for example, by using different cell lines).
- Examples of discrepancies: inabilities to reproduce over-expression of certain genes in specific cell types; decreased cell proliferation via functional inhibition of a target using RNA interference.
- High-impact journals did not publish more robust claims.
- Confirmation of any given finding by another academic group did not improve data reliability.
Avoiding irreproducibility

- Become much more cautious when working with published targets - always check out at least some of the data.
- VC rely on CROs to confirm published data before making investment. Set-aside capital (few hundred thousand dollars) to validate academic claims.
- Cultivate ties with researchers whose work has repeatedly proved to be reproducible.
- But against the backdrop of falling research and development (R&D) budgets, increased financial risk-aversion and high attrition rates, broader strategies to improve data robustness are needed as well.
- Create a precompetitive consortium that could assess the reproducibility of reported findings. The advantages gained by not having to pursue dead-end claims would offset any lost competitive edge. Another, would be to encourage the scientific community — including academic and industry players, as well as journal editors — to publish more negative data.
- Researchers need to be rewarded for actions that improve data robustness.
- Tap into the expertise of university technology transfer offices (TTOs). Whereas many TTOs have set up microseed funds to spin companies out of academic findings, the TTOs would better serve both the universities and the broader community by funding either CROs or other research teams to independently validate claims. Their data packages would then become so much stronger and better able to attract early-stage investors. A lot of biotechs are currently formed sooner than they should be.
- More data may be needed before the field can move forward effectively. Pool data on irreproducibility so that we can unravel the drivers of the translatability of academic findings.