Reproducibility and Measurement Assurance
In Biomedical Research:

A NIST Perspective

Anne Plant
Program Director for Biosciences
National Institute of Standards and Technology
Anne.plant@nist.gov
C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Nature 483, 2012

“Unquestionably, a significant contributor to failure in oncology trials is the quality of published preclinical data...The results of preclinical studies must ... be very robust to withstand the rigours and challenges of clinical trials, stemming from the heterogeneity of both tumours and patients.”
Statistical Modeling, Causal Inference, and Social Science

Replication backlash
Posted by Andrew on 17 December 2013, 9:25 am

Raghuveer Parthasarathy pointed me to an article in Nature by Mina Bissell, who writes, “The push to replicate findings could shelve promising research and unfairly damage the reputations of careful, meticulous scientists.”

I can see where she’s coming from: if you work hard day after day in the lab, it’s gotta be a bit frustrating to find all your work questioned, for the frauds of the Dr. Anil Pottis and Diederik Stapels to be treated as a reason for everyone else’s work to be considered guilty until proven innocent.

That said, I pretty much disagree with Bissell’s article, and really the best thing I can say about it is that I think it’s a good sign that the push for replication is so strong that now there’s a backlash against it. Traditionally, leading scientists have been able to simply ignore the push for replication. If they are feeling that the replication movement is strong enough that they need to fight it, that to me is good news.

I’ll explain a bit in the context of Bissell’s article. She writes:

Articles in both the scientific and popular press have addressed how frequently biologists are unable to repeat each other’s experiments, even when using the same materials and methods. But I am concerned about the latest drive by some in biology to have results replicated by an independent, self-appointed entity that will charge for the service. The US National Institutes of Health is considering making validation routine for certain types of experiments, including the basic science that leads to clinical trials.

But, as she points out, such replications will be costly. As she puts it:

Isn’t reproducibility the bedrock of the scientific process? Yes, up to a point. But it is sometimes much easier not to replicate than to replicate studies, because the techniques and reagents are sophisticated, time-consuming and difficult to master. In the past ten years, every paper published on which I have been senior author has taken between four and six years to complete, and at times much longer. People in my lab often need months — if not a year — to replicate some of the experiments we have done. If we require everything to be replicate, it will reduce the resources that are available to do new research.

[Bissell’s students and postdocs], and others, have always managed to replicate our previous data.

Inability to Reproduce Published Studies Gets Attention
The governments of the OECD, a club of mostly rich countries, spent $59 billion on biomedical research in 2012.

One of the justifications for this is that basic-science results provided by governments form the basis for private drug-development work.

If companies cannot rely on academic research, that reasoning breaks down.

...researchers would find it hard to reproduce at least three-quarters of all published biomedical findings.
Challenge in measurements of cell therapy products:

Characterization of product critical quality attributes

What to measure?
How to measure?
Is the measurement correct?
Can it be compared to and combined with other measurements?
Is it meaningful to clinical outcome?
Non-quantitative Methods Can Lead to Ambiguity

Confusion matrix of experts’ scores illustrating correspondences and differences.

<table>
<thead>
<tr>
<th>Expert 1 Scores</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>31</td>
<td>19</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>21</td>
<td>52</td>
<td>49</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>35</td>
<td>105</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

479 colonies scored by two experts
• Data that are difficult to quantify: such as imaging data/ activity

• Response function is often nonlinear, response is often not discrete, sampling issues

  • Have to contend with both measurement dispersion and biological variability

• Common statistical methods are not always applicable

• Identity (such as nucleotide sequence) is difficult to qualify statistically

We need more research in measurement science.
Measurement assurance: Method validation

- Benchmarking microscope performance.
- Measurement assurance of NA sequence.
- Confirmation of cell line identity.
- Accuracy of image analysis algorithms.

Halter et al. Cytometry A 2014

Dima et al. Cytometry A 2011

Munro et al. 2014 Nat Met

Almeida et al. Cytotechnology 2013
The Need for Measurement Assurance Scales with the Size of the Dataset

Today:

Hypothesis → Data collection → Analysis

With big datasets:

Data collection → Analysis → Hypothesis

Big Data could profoundly change biology

Reduce bias
Generate new hypotheses
Expose rare events
Expose patterns
Indicate what is important to measure
### Sharing/Combining Data

<table>
<thead>
<tr>
<th>Would allow:</th>
<th>To be combined requires:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re-analysis of rich datasets</td>
<td>Evaluation of uncertainties</td>
</tr>
<tr>
<td>Development and testing of analytical methods and theoretical models</td>
<td>Statistically weighted combination of data based on confidence</td>
</tr>
<tr>
<td>Comparing results from different experimental systems to test generalizability</td>
<td>Benchmark data from reference materials that allows normalization of scales</td>
</tr>
<tr>
<td>Discovering what experimental parameters are responsible for differences in biological response</td>
<td>Control data that establish confidence in the quality of the data</td>
</tr>
</tbody>
</table>
How NIST sees the problem of irreproducibility

Confidence in Measurements: reproducibility is one element of a good measurement process

**Accuracy:** Orthogonal method

**Precision:** Repeatability (replicates);

**Reproducibility:** (different days, locations, technicians)

**Ruggedness:** sensitivity to assay parameters

**Dynamic range and Response function:**
Instrument benchmarking. +/- controls. Calibration curve. Limit of detection

**Specificity:** sensitivity to matrix effects

Can be reproducible without being accurate.
Role of Agencies and Institutions

- Awareness of bias and sources of error
- Good measurement practice
  - Tools for recording and reporting
  - Analytical methods testing and validation
  - Benchmarking materials and methods
- Community effort (funders, reviewers, mentors)
Beyond Reproducibility

It is important that the public have confidence in the scientific method, and that all researchers, research reviewers and funders have a good understanding of the hallmarks of scientific investigations that produce results with a high level of confidence.