Costs of Reproducibility Crisis

Brent Goldfarb
University of Maryland
Public Comments; Committee on Reproducibility and Replicability in Science
National Academic of Sciences
April 18, 2018
Scientific Apophenia


• Scientific Apophenia: *assigning of inferential meaning when limited statistical power should prevent such a conclusion or when the data are actually random.*

- Goldfarb & King, 2016
Breast Cancer & Bone Marrow Transplants (HDC/ABMT)

• Devastating and salient disease:
  • Young Women
  • Low survival (particularly Stages III and IV).

• Idea (1982):
  • Super-lethal doses of chemotherapy to aggressively target fast growing cancer cells.
  • Bone marrow transplants to replace destroyed marrow
Timeline of Narrative (HDC/ABMT)

- Increases treatment response at 3 years
- Early evidence based on archival interpretation of previous studies
- Not wrong, but speculative
Timeline: HDC/ABMT

- **Hryniuk & Bush (1984)**
- **Hyrniuk & Levine (1986)**
- **Endorsed by NCI Director**
- **Ignored critique Henderson et al. (1988)**

**Clinical Trials** (~1,000 women) 1988-2001

**Clinical Treatment** (35,000 Women) 1988-2001

- $50,000
- $50,000
- $1,750,000,000
- $60,000,000
Application Costs: HDC/ABMT

Hyrniuk & Levine (1986)
(Ignored critique)
Henderson et al. (1988)

Clinical Trials (~1,000 women) 1988-2001
Clinical Treatment (35,000 Women) 1988-2001

$50,000 $50,000 $1,750,000,000

$60,000,000

29X
Timeline of Narrative

• ~ 40,000 patient underwent HDCT + BMT
  • Harder to recruit patients
  • ~600 died from treatment
• No survival increase at 5 years
• gastrointestinal toxicity, nausea, vomiting and diarrhea, infection and organ toxicity” - Howard et al. 2011
## Error Cost Escalation (Software)

<table>
<thead>
<tr>
<th>Requirements</th>
<th>Method 1 Cost Factors</th>
<th>Software Cost Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>1X</td>
<td>1X</td>
</tr>
<tr>
<td>Design</td>
<td>8X</td>
<td>5X – 7X</td>
</tr>
<tr>
<td>Build</td>
<td>16X</td>
<td>10X – 26X</td>
</tr>
<tr>
<td>Test</td>
<td>21X</td>
<td>50X – 177X</td>
</tr>
<tr>
<td>Operations</td>
<td>29X</td>
<td>100X – 1000X</td>
</tr>
</tbody>
</table>

*Table 4: Comparison of Method 1 and Software Cost Factors*

Stecklein et. al. (2004)
Error Cost Escalation (Software)

General Science

- Original Study
- Citing Studies
- Applications

<table>
<thead>
<tr>
<th>Requirements</th>
<th>Method 1 Cost Factors</th>
<th>Software Cost Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>1X</td>
<td>1X</td>
</tr>
<tr>
<td>Design</td>
<td>8X</td>
<td>5X – 7X</td>
</tr>
<tr>
<td>Build</td>
<td>16X</td>
<td>10X – 26X</td>
</tr>
<tr>
<td>Test</td>
<td>21X</td>
<td>50X – 177X</td>
</tr>
<tr>
<td>Operations</td>
<td>29X</td>
<td>100X – 1000X</td>
</tr>
</tbody>
</table>

Table 4: Comparison of Method 1 and Software Cost Factors
Costs of Poor Reproducibility

• Direct Costs
  • Costs of producing bad research

• We can be reasonably certain about these.

• House of Cards – Indirect Costs
  • Science that builds on incorrect results
    • Building on
    • Striking down

• Applications

• Much harder to Guesstimate
Direct Costs
Pre-clinical bio research

https://doi.org/10.1371/journal.pbio.1002165
http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165
Examples of Harmful Applications

• HDC/ABMT

• Hormone Replacement Therapy for Menopausal Women

• Stents for Low Severity SCAD (Stable Coronary Arterial Disease)
  • ~$ 3.5B annually (2002-2005; extrapolated from Greenwood et al., 2017)

• But most studies implications do not make it into practice.

$90B @ 50%  •  Original Study

$18B @ 20%  •  Building on / Striking Down

$13.5B – $270B  •  Applications to practice

National Science Board, Science & Engineering Indicators, 2018 – Table 4-4
We really don’t know

• Rare events will have outsized effects on these estimates

• Major assumptions drive numbers in the cost of misguided applications

• So... more research needed.
Most policy is decided under uncertainty

• Decisions *must* be made with incomplete information

• What if HDC/ABMT had been effective?
  • Preliminary evidence was *inconclusive*.
  • *Breast cancer mortality has fallen substantially!* (Mortality rate down 50% 1996-2006) (https://www.prb.org/breastcancer/).

• No excuse for poor documentation

• Uncertainty gets lost in compelling “too good to be true” narratives.
Costs of remediation are small — from Center for Open Science – Brian Nosek

• Better Archives / Infrastructure – trivial, and mostly one time.
  • ~3,000,000 datasets posted!

• Training of Research Community - $750k Annually

• Replications
  • Center for Open Science Estimate Approximately ~$100K/study.
Recommendations

• Show your work / Share your outputs

• Open as default for all research outputs

• Much work to be done in establishing standards for disclosure

• Registration as the default as having been conducted.
  • Discovery oriented – register that it exists.
Sources / Further Reading


Thank you!

Brent Goldfarb
Associate Professor
Robert H. Smith School of Business
University of Maryland, College Park 20742
202-730-9484
bgoldfarb@rhsmith.umd.edu
@brentdg2