

Replication and Reproducibility in Spinal Cord Injury Research



Take home

- There has been a preponderance of failures to replicate.
- Lack of replication is not a bad thing.
 - It can lead to critical adjustments in approach and save the field huge amounts of money pursuing false leads.
 - Non-replication of early findings is part of the natural history of discovery (Kevin Staley).



The Problem:

- Many reports of treatments that improve outcome after SCI; No translation. Why?
- Rumors: “We repeated that experiment and it didn’t work”.
- Failure of clinical trials for a variety of disorders including stroke and TBI.



In recognition of the problem, NINDS launched the FACILITIES OF RESEARCH EXCELLENCE IN SCI (FORE-SCI) contracts,

- Program Officer, Naomi Kleitman, NINDS
- Contract Officer, Laurie Leonard, NINDS

- PI's and Sites
 - ◆ Oswald Steward (UC Irvine, 2003, 2008)
 - ◆ Dalton Dietrich (U. Miami, 2003)
 - ◆ Philip Popovich (Ohio State U, 2008)



Contracts are different from grants

- NIH buys a service / deliverable
- NIH stipulates the scope and desired product
 - ◆ Faithful replication of published studies
 - ◆ Facilities provide, in one location, resources, capabilities and expertise in SCI research
 - ◆ Activities are defined; conduct of additional studies is limited
- Advisory Committees advise PI and NINDS about studies chosen for replication

◆ Slide from Naomi Kleitman



FORE-SCI – Replication studies

- Specific performance goals of the Contracts:
 - ◆ Try to replicate promising, preclinical studies relating to therapies that could lead to effective treatments for human SCI,
 - ◆ Compare the efficacy of treatments in a standardized environment with a minimum of variability in surgery, animal care, outcome evaluation and cellular analyses,
 - ◆ Promptly report the methodology and results.
- The desired result is that, if proven to provide reliable and robust benefit, these promising strategies would be appropriate to move to the next level of translation or, if appropriate, clinical testing.
- If studies are NOT reproducible, this could save \$millions that would otherwise be spent on dead ends and failed clinical trials.
 - ◆ Slide from Naomi Kleitman



Criteria for study selection

- Clinically relevant endpoints (usually means sparing or recovery of function).
 - ◇ Is treatment potentially translatable to the clinic?
 - ◇ Some were already in or on the way to clinical trials
- Degree of improvement (effect size)
- Scientific merit of the publication
- General strengths and weaknesses
 - Slide from Naomi Kleitman



Findings to date:

- **Surprising preponderance of failures to replicate (1/12)**
 - ◇ *What does it mean to the field?*
- **Methods sections are often incomplete or misleading**
 - ◇ Randomization is rarely explained and often is NOT DONE.
 - ◇ Communication with original authors is essential, but often reveals that the experiment was NOT done as the Methods imply.
- **Significant technological hurdles**
 - ◇ Reproducibility of SCI models; control deficit levels.
- **Publishing negative results is doable and generally well-received by the field.**

Slide from Naomi Kleitman



Important methodological issues:

- Many papers describe work carried out over a period of several years. Groups were not run simultaneously. There is no description of this in Methods.
- This is true of most SCI experiments, and is always true when there are multiple groups involving many animals.
- Batching of animals/non-simultaneity of group assessment is almost never explained.



It is sometimes impossible to remain blind.
Here, treatment turns rats blue!



Peng et al., PNAS, 2009



Why is there a failure to replicate?

- ◆ The file drawer problem. Studies that “work” are published; studies that don’t aren’t.
- Type I statistical error. Multiple comparisons, only one of which is significant.
- ◆ Methodological details that are not reported (non-simultaneous group assessment).
- ◆ Effects are not robust.
- ◆ Inadvertent bias:
 - ☞ Unrecognized tendency to be more careful with the “experimental” group during the surgery for example.
 - ☞ Non-random order of surgery/treatment/testing.
 - ☞ Important or difficult procedures may be done first.
 - ☞ Post-operative care is a treatment variable.



Some points

- ◆ Preclinical is any study that tests a biological concept in an animal model of disease.
- ◆ A large percentage of preclinical studies by the above definition are R01-funded.
- ◆ R01 review does not currently emphasize the importance of replication, optimization, etc.
- ◆ Blinding and randomization requires a larger staff than most R01 grants can support.
- ◆ Replication and optimization studies are not career-builders.



How do we think about failures to replicate?

Does a failure to replicate mean that the basic biology is invalid?

Or does it simply mean that the effect depends on experimental details that are not easily recognized?

Either way, the important conclusion is that the effects are not robust.

Treatments that do not produce robust effects are unlikely to be translatable.

If it's too good to be true, it's probably not true.

- Extraordinary claims require extraordinary documentation.
- The level of documentation for regeneration after spinal cord injury is difficult to compress into the space allowed by high profile journals.
- So, maybe studies reporting regeneration should not be published in high profile journals.
 - (Except for my studies of course).



Roadblocks to solutions

- NIH review criteria? Optimization and replication are not “innovative”.
- Academia: Adjust reward structures?



IACUC issues

- Minimizing animal use (thus reducing “n”) vs. ensuring sufficient power.
- IACUC requirements to avoid duplication. Replication is by definition duplication.



Some fallacies:

- It's hard to publish negative results.
- FALSE:
 - ◆ Reviewers have been very positive.
 - ◆ Every replication paper has been accepted.



Some fallacies:

- Repeating an experiment is not interesting, especially if the results are negative.
- FALSE:
 - ◆ There is increasing recognition that reporting negative results is important and interesting.
 - ◆ And there have been unexpected findings that add value.



Some fallacies:

- You'll make enemies.
 - ◇ Hmm; well maybe this isn't a fallacy.

