The Chrysalis Effect

Ernest H. O’Boyle
University of Iowa

Presentation for CASBS Best Practices in Science Conference
QRPs occur when hypotheses are altered to support data or when data are altered to support hypotheses.

The question lies in motivation for the changes

- Bad psychometrics
- Discovery of an outlier after initial analysis
- Reviewer recommendation to test an interaction
- Hypotheses/theoretical model not supported
Tracked 142 dissertations in MGMT and I-O psychology

Across dissertations and journal articles, we recorded 2,311 hypotheses
• 645 common or retained hypotheses
• 333 added hypotheses
• 1333 dropped hypotheses
Questionable Research Practices

- QRP #1: Altering the data after hypothesis testing
- QRP #2: Deletion or addition of data after hypothesis tests
- QRP #3: Selective deletion or addition of variables
- QRP #4: Reversing the direction or reframing hypotheses to support data
- QRP #5: Post hoc dropping or adding of hypotheses
QRPs within the common hypotheses

<table>
<thead>
<tr>
<th>QRP</th>
<th>Unsupported Dissertation Hyp.</th>
<th>Supported Dissertation Hyp.</th>
<th>Risk ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Δsupport</td>
<td>%</td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>56</td>
<td>21%</td>
</tr>
</tbody>
</table>
QRPs within the common hypotheses

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Signif.</th>
<th>%</th>
<th>% Diff.</th>
<th>Risk ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Added Hyp.</td>
<td>333</td>
<td>233</td>
<td>70%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropped Hyp.</td>
<td>1333</td>
<td>516</td>
<td>39%</td>
<td>31%</td>
<td>1.81</td>
</tr>
</tbody>
</table>

(1.64; 1.99)
Collective Effect of QRPs

- Of the 1,978 hypotheses contained in the dissertations (i.e., dropped and common hypotheses), 889 (45%) were statistically significant.
- Of the 978 hypotheses contained in the journal articles (i.e., added and common hypotheses), 645 (66%) were statistically significant.
- 21.0% inflation & a more than doubling of the ratio of supported to unsupported hypotheses from 0.82:1 in the dissertations to 1.94:1 in the journal articles.
Interaction Terms in OLS

- In our six highest impact I-O journals
  - 1995-1999: 44% significant
  - 2000-2004: 54% significant
  - 2005-2009: 61% significant
  - 2010-2014: 74% significant

- Mean sample size has not changed over the 20 year period
Meta-analyzed the relation between sample size and effect size for more than 1200 moderators \((r = -.33)\)

Recalculated \(p\)-values less than .05 were 100% consistent with author's conclusion

Recalculated \(p\)-values between .05 and .10, were 17% consistent with author's conclusion
40% of degrees of freedom in structural equation models don’t match the proposed model (Green, Keeler, Vandenberg, and Cortina, 2015)
Response rates in studies of workplace deviance are upwardly biased by 16%. (Greco, O’Boyle, & Walter, 2015)
Why do researchers do it?

- “To be honest, I am not sure if these are considered be "questionable" research practices (except falsifying data/results).”
- “Publish or perish! When you consider the style of writing that is expected by JAP, AMJ, JPSP, and the like then you basically MUST do some of these things.”
- “If the first story does not appeal, you rewrite it to appease the editors/reviewers. This is a key to getting one's work accepted. Anyone who pretends otherwise is a fool and probably unpublished. Sorry to burst your bubble. This is not science, it is the art of persuasion.”