effective computing for research reproducibility

Laura Fortunato
University of Oxford
Santa Fe Institute

@anthrolog
http://www.santafe.edu/~fortunato/
preliminaries

disclaimer

- I am not a statistician
- I am not a data scientist
- I am not a computer scientist
preliminaries

disclaimer

▶ I am not a statistician
▶ I am not a data scientist
▶ I am not a computer scientist
▶ I am a computational researcher
preliminaries

disclaimer

▶ I am not a statistician
▶ I am not a data scientist
▶ I am not a computer scientist
▶ I am a computational researcher

broader context

“open science” < “open research” < “open scholarship”
Reproducible Research Oxford

@RR_Oxford
http://rroxford.github.io/

The Carpentries

@thecarpentries
http://carpentries.org/
Reproducible Research Oxford

@RR_Oxford
http://rroxford.github.io/

The Carpentries

@thecarpentries
http://carpentries.org/

UK Reproducibility Network

@ukrepro
http://ukrn.org/

Project TIER

@Project_TIER
https://www.projecttier.org/
research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently
research reproducibility

the problem, loosely defined
results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues
research reproducibility

the problem, loosely defined

results can only be trusted if they can be \textit{re-derived}
by the original researchers and/or by others working independently

trust issues

\begin{itemize}
\item can I trust the result?
\end{itemize}
research reproducibility

the problem, loosely defined
results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues

- can I trust the result?
- can I trust the researcher?
research reproducibility

the problem, loosely defined
results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues

▷ can I trust the result?
▷ can I trust the researcher?
▷ can I trust the process?
reproducibility crisis
What does research reproducibility mean?

Title or abstract including one or more of:

“research reproducibility”
“reproducibility of research”
“reproducibility of results”
“results reproducibility”
“reproducibility of study”
“study reproducibility”
“reproducible research”
“reproducible finding”
“reproducible result”
# “research on research”

Ioannidis et al. (2015) *PLOS Biology* 13(10): e1002264

Meta-research: evaluation and improvement of research methods and practices

<table>
<thead>
<tr>
<th>Meta-research area</th>
<th>Specific interests (nonexhaustive list)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methods</strong>: &quot;performing research&quot;—study design, methods, statistics, research synthesis, collaboration, and ethics</td>
<td>Biases and questionable practices in conducting research, methods to reduce such biases, meta-analysis, research synthesis, integration of evidence, cross-design synthesis, collaborative team science and consortia, research integrity and ethics</td>
</tr>
<tr>
<td><strong>Reporting</strong>: &quot;communicating research&quot;—reporting standards, study registration, disclosing conflicts of interest, information to patients, public, and policymakers</td>
<td>Biases and questionable practices in reporting, explaining, disseminating and popularizing research, conflicts of interest disclosure and management, study registration and other bias-prevention measures, and methods to monitor and reduce such issues</td>
</tr>
<tr>
<td><strong>Reproducibility</strong>: &quot;verifying research&quot;—sharing data and methods, repeatability, replicability, reproducibility, and self-correction</td>
<td>Obstacles to sharing data and methods, replication studies, replicability and reproducibility of published research, methods to improve them, effectiveness of correction and self-correction of the literature, and methods to improve them</td>
</tr>
<tr>
<td><strong>Evaluation</strong>: &quot;evaluating research&quot;—prepublication peer review, postpublication peer review, research funding criteria, and other means of evaluating scientific quality</td>
<td>Effectiveness, costs, and benefits of old and new approaches to peer review and other science assessment methods, and methods to improve them</td>
</tr>
<tr>
<td><strong>Incentives</strong>: &quot;rewarding research&quot;: promotion criteria, rewards, and penalties in research evaluation for individuals, teams, and institutions</td>
<td>Accuracy, effectiveness, costs, and benefits of old and new approaches to ranking and evaluating the performance, quality, value of research, individuals, teams, and institutions</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pbio.1002264.t001
xkcd: trouble for science
https://xkcd.com/1574/

MANY COMMERCIAL ANTIBODY-BASED IMMUNOASSAYS ARE UNRELIABLE.

PROBLEMS WITH THE P-VALUE AS AN INDICATOR OF SIGNIFICANCE.

OVERFEEDING OF LABORATORY RODENTS COMPROMISES ANIMAL MODELS.

REPLICATION STUDY FAILS TO REPRODUCE MANY PUBLISHED RESULTS.

CONTROLLED TRIALS SHOW BUNSEN BURNERS MAKE THINGS COLDER.
perceptions among scientists
1,500 scientists lift the lid on reproducibility
perceptions among the public

*The Economist*, issue October 19, 2013
crisis vs. “crisis narrative”

PNAS Sackler Colloquium on Improving the Reproducibility of Scientific Research
published March 13, 2018 by the NAS — National Academy of Sciences USA

Crisis or self-correction: Rethinking media narratives about the well-being of science

Kathleen Hall Jamieson

Opinion: Is science really facing a reproducibility crisis, and do we need it to?

Daniele Fanelli

Scientific progress despite irreproducibility: A seeming paradox

Richard M. Shiffrin, Katy Börner, and Stephen M. Stigler
The Irreproducibility Crisis of Modern Science
Causes, Consequences, and the Road to Reform

David Randall & Christopher Welser
National Association of Scholars
April 2018
it’s complicated. . .

Reproducible research in computational science

reproducibility spectrum
it’s complicated. . .

Reproducible research in computational science

reproducibility spectrum

reproducibility *identical results with same data*
it’s complicated. . .


Reproducible research in computational science

reproducibility spectrum

reproducibility *identical* results with same data
replicability *consistent* results with new data
What does research reproducibility mean?

A conceptual framework

it’s complicated... 
What does research reproducibility mean?

a conceptual framework

1. methods reproducibility:
   enough detail about study procedures and data
it’s complicated. . .
What does research reproducibility mean?

**a conceptual framework**

1. methods reproducibility:  
   enough detail about study procedures and data
2. results reproducibility:  
   same results from closely matched independent study
it’s complicated. . .
What does research reproducibility mean?

a conceptual framework

1. methods reproducibility:
   enough detail about study procedures and data
2. results reproducibility:
   same results from closely matched independent study
3. inferential reproducibility:
   qualitatively similar conclusions from reanalysis or replication
it’s complicated. . .
Before reproducibility must come preproducibility

“preproducibility”
it’s complicated...
Before reproducibility must come preproducibility

“preproducibility”
a “scientific recipe” for others to repeat the experiment or analysis
it’s complicated...  
Before reproducibility must come preproducibility

“preproducibility”  
a “scientific recipe” for others to repeat the experiment or analysis

“An experiment or analysis is preproducible  
if it has been described in adequate detail  
for others to undertake it.  
Preproducibility is a prerequisite for reproducibility,  
and the idea makes sense across disciplines.”
e.g. the view from psychology...

hypothetico-deductive model of the scientific method and potential threats
points of view
Reproducibility failures are essential to scientific inquiry

...vs. the view from maths, computer science

- e.g. no statistical problems, no issues with experimental design
- failures to replicate essential to integration of conflicting observations and ideas into coherent theory
points of view
Reproducibility failures are essential to scientific inquiry

vs. the view from maths, computer science

- e.g. no statistical problems, no issues with experimental design
- failures to replicate essential to integration of conflicting observations and ideas into coherent theory

“The discovery that an experiment does not replicate is not a lack of success but an opportunity. ... A failure to reproduce is only the first step in scientific inquiry. In many ways, how science responds to these failures is what determines whether it succeeds.”
research reproducibility

the problem, loosely defined

results can only be trusted if they can be re-derived
by the original researchers and/or by others working independently
research reproducibility

the problem, loosely defined
results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

(part of) the solution
change in approach to computing in research

sharing over email
manual updating
mouse “trails” in GUIs
...
not reproducible

shared repository
version control
dynamic documents
...
fully reproducible
e.g. data management malfunction


Dance reveals symmetry especially in young men
e.g. data management malfunction

Dance reveals symmetry especially in young men

Symmetry study deemed a fraud

“...Brown says that it is unclear which data set is the original because many versions exist.”
e.g. data analysis malfunction

Reinhart and Rogoff (2010)
*American Economic Review* 100(2): 573–578

Growth in a time of debt
Reinhart and Rogoff (2010)
*American Economic Review* 100(2): 573–578
Growth in a time of debt

Herndon et al. (2014)
Does high public debt consistently stifle economic growth?
A critique of Reinhart and Rogoff
e.g. software malfunction

https://support.microsoft.com/en-us/help/214326/excelincorrectly-assumes-that-the-year-1900-is-a-leap-year
If this behavior were to be corrected, many problems would arise, including the following:

- Almost all dates in current Microsoft Excel worksheets and other documents would be decreased by one day. Correcting this shift would take considerable time and effort, especially in formulas that use dates.

- Some functions, such as the WEEKDAY function, would return different values; this might cause formulas in worksheets to work incorrectly.

- Correcting this behavior would break serial date compatibility between Microsoft Excel and other programs that use dates.

If the behavior remains uncorrected, only one problem occurs:

- The WEEKDAY function returns incorrect values for dates before March 1, 1900. Because most users do not use dates before March 1, 1900, this problem is rare.

NOTE: Microsoft Excel correctly handles all other leap years, including century years that are not leap years (for example, 2100). Only the year 1900 is incorrectly handled.

References
An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.
Science should be ‘show me’, not ‘trust me’; it should be ‘help me if you can’, not ‘catch me if you can’.