Addressing reproducibility in the neurosciences and neuroimaging: Statistical, computational, and sociological aspects

Fall 2018

Instructor:	Jean-Baptiste Poline	**Time:**	TR 14:30 – 17:30
Email:	jean-baptiste.poline@mcgill.ca	**Place:**	Rasmussen Reading Room

Course Pages:

1. figshare collection: https://figshare.com/s/a80c1bec9a5996d36b0c
2. GitHub classroom: https://classroom.github.com/classrooms/42415240-reprocourse

Objectives: This series of journal club readings will give students a solid background in the reproducibility and replication issues in the neurosciences and brain imaging, and introduce solutions that can be implemented for reproducible and replicable science. The statistical, methodological, computational and sociological aspects of the topic will be reviewed through these articles, as well as the implementation of solutions. After this course, students should be able to identify reproducibility issues in neuroscience literature and apply the principles of reproducible, reusable, and efficient research in their work.

Prerequisites: An undergraduate-level understanding of statistics is assumed.

Office Hours: Fridays 11:00–12:30 in WB210 or by appointment. Students are also encouraged to post questions in the reprocourse channel on the Brainhack Slack.

Class Policy:

- Regular attendance is essential and expected.
- In accordance with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Grading Policy: There is no exam for this course. Participation and ability to critically analyse information for each class will account for 50% of the grade. The remaining 50% of the grade is derived from a mock grant proposal (25%), and an oral presentation of the grant (25%). In an effort to train students for oral presentation, the students will present their mock grant orally on the last day during a 15 minute talk during which they will have the opportunity to address comments from the instructors on the mock grant and expose the background, hypothesis and experimental strategy.

Academic Integrity: McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see www.mcgill.ca/students/srr for more information).

Course Outline: Each week, students will be provided with course readings as well as practical exercises related to reproducibility in the neurosciences and neuroimaging. These are intended to complement one another and provide students with hands-on experience implementing methodological solutions to problems detailed in the week's readings. Discussions for each week will be broadly grouped as considering either the sociological, statistical, or computational aspects of reproducibility.

Reproducibility in life sciences: Some background ———————————————————

Sept. 6: Landmark papers in the reproducibility crisis in life sciences

- Begley & Ellis (2012): Raise standards
- Baggerly & Coombes (2009): Forensic analysis
- Working in a UNIX environment

Sept. 13: General description of the issue and a first open science response

- Academy of Medical Sciences (2015): Reproducibility and Reliability of Biomedical Research
- Open Science Collaboration (2015): The reproducibility project
- Introduction to Git for version control

Sept. 20: Some sociological and general aspects

- Smaldino & McElreath (2016): The natural selection of bad science
- Allison et al. (2016): A tragedy of errors
- Introduction to GitHub for collaboration

Sept. 27: Some general solutions and principles

- Wilkinson et al. (2016): The FAIR principles
- Bosman et al. (2017): The scholarly common principles
- Introduction to NumPy and Pandas

Some statistical aspects ———————————————————

Oct. 4: Power issues: the initial reports

- Ioannidis (2005): Why most research results are false
- Button et al. (2013): Power failure
- Using Python for statistics

Oct. 11: Power issues: the more recent reports

- Poldrack et al. (2017): Scanning the horizon
- Dumas-Mallet et al. (2017): Three biomedical examples

Oct. 18: Guest lecture by Dr. Paramita S. Chaudhuri

- Rosenthal (1979): The file drawer effect
- Simmons et al. (2011): p-hacking
- Simonsohn et al. (2014): p-curve

Oct. 25: Proposal for redefining p-values and response

- Benjamin et al. (2018): Redefining p-value
- Lakens et al. (2018): Justify your alpha
- An experiment in p-hacking

Some computational aspects, with examples from neuroimaging

Nov 1st: Software or software use issues

- Eklund et al. (2016): (fMRI) Cluster failure
- Varoquaux (2017): Cross-validation failure
- Introduction to containers

Nov 15: Some Computational aspects

- Glatard et al. (2015): OS dependencies
- Bowring et al. (2018): Same data, different results
- Carp (2012): Pipeline flexibility
- Containers, continued

Nov 22: Examples of replications and of reproducible articles

- Boekel et al. (2015): A pure replication study
- Waskom et al. (2014): An entirely reproducible article

Nov 29: Community based standards

- Nichols et al. (2017): The COBIDAS report
- Gorgolewski et al. (2016): The BIDS standard

References

Academy of Medical Sciences. (2015). Reproducibility and reliability of biomedical research: improving research practice. In *Symposium report.* Retrieved from https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature News*, *530*, 27. doi: https://doi.org/10.1038/530027a

Baggerly, K. A., & Coombes, K. R. (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 1309–1334. doi: https://doi.org/10.1214/09-AOAS291

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*, 531. doi: https://doi.org/10.1038/483531a

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... others (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6. doi: https://doi.org/10.1038/s41562-017-0189-z

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115–133. doi: https://doi.org/10.1016/j.cortex.2014.11.019

Bosman, J., Bruno, I., Chapman, C., Tzovaras, B. G., Jacobs, N., Kramer, B., ... others (2017). The scholarly commons-principles and practices to guide research communication. Retrieved from https://osf.io/kbuq6/

Bowring, A., Maumet, C., & Nichols, T. E. (2018). Exploring the impact of analysis software on task fmri results. *BioRxiv*. doi: https://doi.org/10.1101/285585

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365. doi: https://doi.org/10.1038/nrn3475

Carp, J. (2012). On the plurality of (methodological) worlds: estimating the analytic flexibility of fmri experiments. *Frontiers in neuroscience*, *6*, 149. doi: https://doi.org/10.3389/fnins.2012.00149

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in biomedical science: a review of three human research domains. *Royal Society open science*, *4*, 160254. doi: https://doi.org/10.1098/rsos.160254

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 201602413. doi: https://doi.org/10.1073/pnas.1602413113

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., ... others (2015). Reproducibility of neuroimaging analyses across operating systems. *Frontiers in neuroinformatics*, *9*, 12. doi: https://doi.org/10.3389/fninf.2015.00012

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... others (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044. doi: https://doi.org/10.1038/sdata.2016.44

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*, e124. doi: https://doi.org/10.1371/journal.pmed.0020124

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... others (2018). Justify your alpha. *Nature Human Behaviour*, *2*, 168. doi: https://doi.org/10.17605/OSF.IO/9S3Y6

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., ... others (2017). Best practices in data analysis and sharing in neuroimaging using mri. *Nature Neuroscience*, *20*, 299. doi: https://doi.org/10.1038/nn.4500

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. doi: https://doi.org/10.1126/science.aac4716

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115. doi: https://doi.org/10.1038/nrn.2016.167

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*, 638. doi: https://doi.org/10.1037/0033-2909.86.3.638

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*, 1359–1366. doi: https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534. doi: https://doi.org/10.1037/a0033242

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*, 160384. doi: https://doi.org/10.1098/rsos.160384

Varoquaux, G. (2017). Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage*. doi: https://doi.org/10.1016/j.neuroimage.2017.06.061

Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal representations of task context support the flexible control of goal-directed cognition. *Journal of Neuroscience*, *34*, 10743–10755. doi: https://doi.org/10.1523/JNEUROSCI.5282-13.2014

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... others (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, *3*. doi: https://doi.org/10.1038/sdata.2016.18