



Reproducibility in rehabilitation research and how data science (& open science) can help

Sook-Lei Liew, PhD, OTR/L

Assistant Professor & Director, Neural Plasticity and Neurorehabilitation Lab

University of Southern California

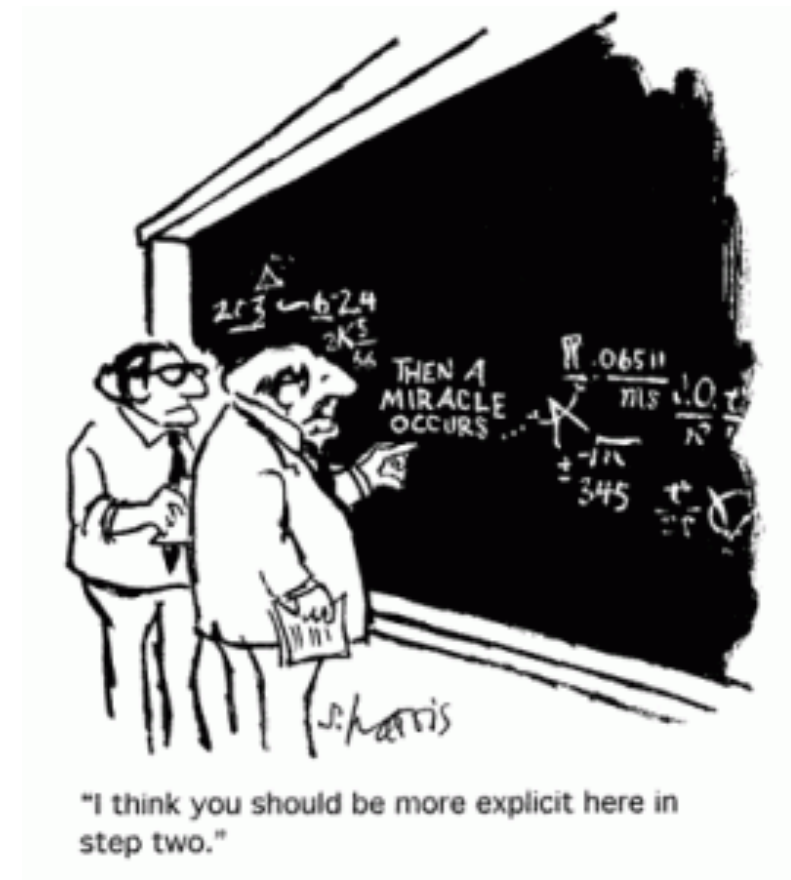
sliew@usc.edu | <https://chan.usc.edu/nprl/>

Scientific reproducibility and replicability

Reproducibility: The ability for someone else (or yourself) to reproduce an experimental paradigm

Replicability: The ability for someone else (or yourself) to obtain consistent results, given the same experiment

1. If I read a paper, is there sufficient detail for me to implement the same experiment?
2. If I implement someone else's experiment, will I get the same results?



What is the reproducibility crisis?

- More than 70% of scientists have tried and failed to reproduce another scientist's experiments:
 - <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>
- Psychology – only 39 of 100 replication attempts were successful
 - <https://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248>



Factors contributing to the problem

Methods (Reproducibility)

- Underutilized reproducible methods:
 - Human error in manual processes (data entry, analysis)
 - Inconsistent keeping record across different team members

Results (Replicability)

- Positive publication bias
- Logistical limitations:
 - Limited money, time, and participant availability can lead to biased and underpowered samples



Potential solutions

Methods (Reproducibility) → Data Science

- Underutilized reproducible methods:
 - Human error in manual processes (data entry, analysis)
 - Inconsistent keeping record across different team members

Results (Replicability) → Open Science

- Positive publication bias
- Logistical limitations:
 - Limited money, time, and participant availability can lead to biased and underpowered samples



What can be done?

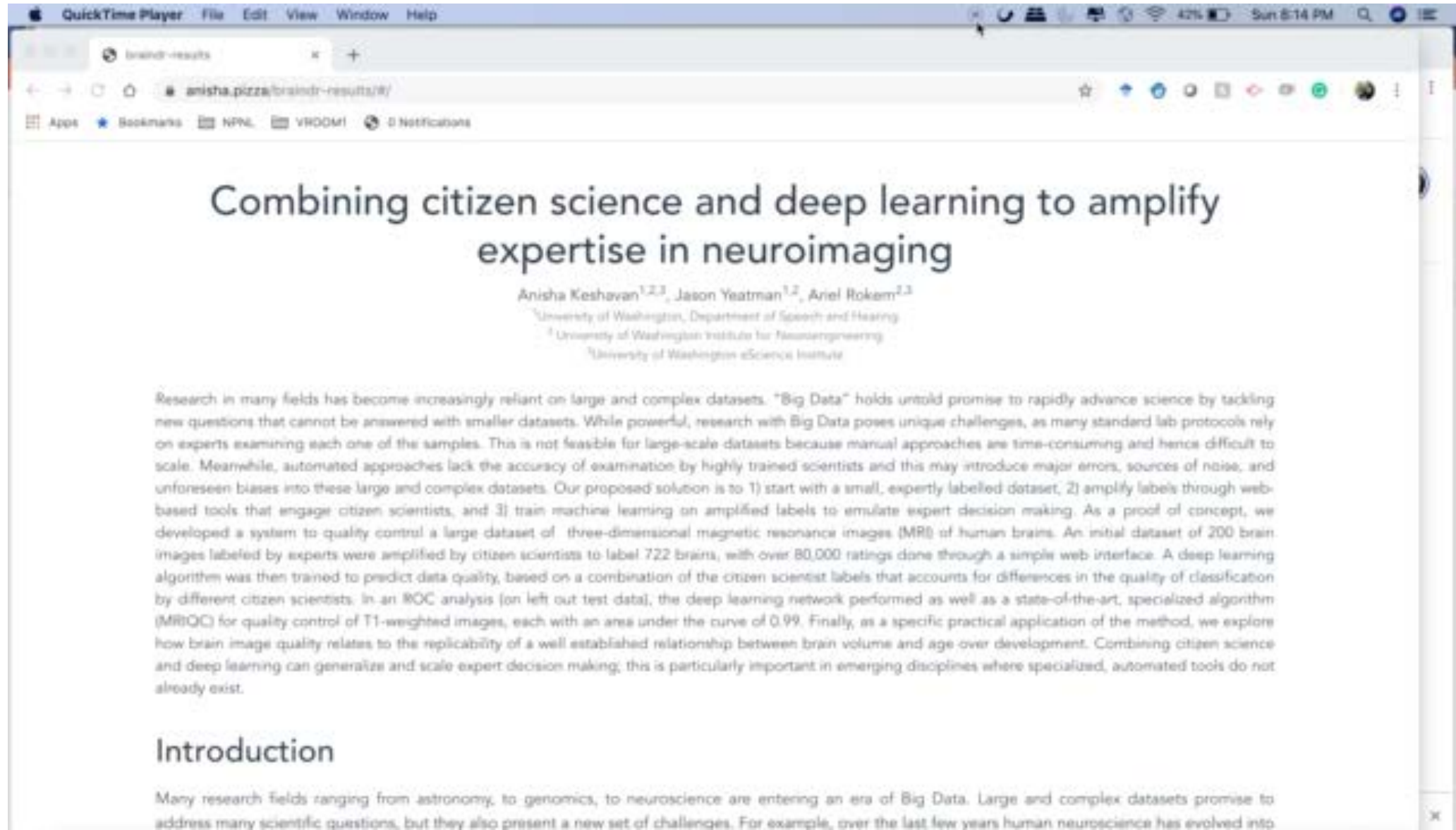
Methods (Reproducibility) → **Data Science**

- Reproducible methods from data science:
 - Data management with consistent formatting
 - Data analysis using executable scripts (Matlab, R, Python)
 - Version control across across different team members, analyses
 - End goal: **Reproducible papers**
- See Center for Reproducible Neuroimaging (ReproNim) as an example: <https://www.repronim.org/>



Reproducible paper example (Keshavan et al., 2019)

<https://anisha.pizza/braindr-results/#/>



Resources for data science in rehab research

- Mobilize Center: <http://mobilize.stanford.edu>
- Center for Large Data Research and Data Sharing in Rehabilitation: <https://www.utmb.edu/cldr>
- 2019 ASNR Symposium: Reliability and Reproducibility in Neurorehabilitation Research
 - Hands-on tutorials and slides on Github: https://github.com/nprl/ASNR_2019
- ReproNim (<https://www.repronim.org/>), NeuroHackademy (https://neurohackademy.org/neurohack_year/2020/)
- Coursera, Udemy, etc.



What can be done?

Results (Reliability) → Open Science

- Overcoming positive publication bias and logistical limitations by testing samples from:
 - Retrospective datasets that have been archived
 - Pooled samples across retrospective/prospective datasets from diverse research sites
 - Large prospective datasets
- All of these would benefit from data science for accurate data management, analysis across sites



Open Science: What is it?

- Open science movement: Sharing (published & unpublished) data, code, protocols, resources
- Why do it? To improve scientific reproducibility and replicability and build the capacity of the scientific community (especially trainees)
- What's involved? Usually free to download, with some agreement you won't abuse/sell the data. That's it!



Open Data: What types of data are there?

- Data types: Surveys, behavioral measures, demographics, kinematic data, videos, physiological data (e.g., brain imaging)
- Prospective data collections (protocol is set prior to data collection)
- Retrospective data archives (usually study-specific data)
- Health services / medical records

Open Data: What types of data are there?

- **Rehabilitation-Related Data Archives (NCMRR-funded)**

- CLDR: <https://www.utmb.edu/cldr/>
 - Center for Large Data Research and Data Sharing in Rehabilitation
 - Many types including health services research (e.g., medical records) and retrospective study-specific rehabilitation data
- ICPSR/ADDEP:
<https://www.icpsr.umich.edu/web/pages/ADDEP/index.html>
 - Archive of Data on Disability to Enable Policy and research
 - Retrospective study-specific rehabilitation data
- OpenSim: <https://opensim.stanford.edu/>
 - Free motion simulation toolbox and trained models for different populations:
<http://simtk.org/>



Open Data: What types of data are there?

- **Prospective/Coordinated Brain Imaging, Clinical/Behavior**

- Human Connectome Project: <https://www.humanconnectome.org/>
 - Lifespan, young adult, clinical populations, with harmonized behavior
- UK Biobank: <https://www.ukbiobank.ac.uk/>
 - UK health records data including brain imaging, genetics, clinical variables
 - Working up to 100,000 individuals
- All of Us: <https://allofus.nih.gov/>
 - On beta release; will be US health records data including brain imaging, genetics, clinical variables and questionnaires
 - Working up to 1 million individuals



Open Data: What types of data are there?

- **Community (Study-Specific) Brain Imaging**

- Open Neuro: <https://openneuro.org/>
 - 372 MRI, MEG, EEG, ECoG datasets
- INDI: http://fcon_1000.projects.nitrc.org/
 - International Neuroimaging Data-Sharing Initiative: Prospective and retrospective data
 - Resting state fMRI, structural MRI, diffusion MRI with behavioral measures
- NITRC: <https://www.nitrc.org/>
 - Neuroimaging Tools and Resources Collaboratory: Atlases, data, and tons of software/tools



Open Data: But I want something specific?



NPNL at USC
@NPNLatUSC



If you're a student/researcher who can't collect data right now but who needs data to analyze to support your thesis/grant/project, let me know what type of data you're looking for, and I'll try to find an openly shared source! Will be doing an [@ASNRehab](#) webinar on this soon!

- Myelin water fraction MRI with behavior
- EEG data during FES-evoked movements
- Walking data with EMG, kinematics
- Resting state EEG with motor learning



USC University of
Southern California

Open Data: Collaborative data sharing

- If you have a specific need, you may consider reaching out to someone who has published a dataset that you'd like to utilize
- General guidelines:
 - Collaborate on the data (including authorship)
 - Receive useful insight on the data wrt how you use it
 - No one's data is perfect!
 - Maybe help organize their data into a data archive that you both can also publish (see journals like *Scientific Data*, *GigaScience*) or cite

Open Data: I want to share data

- Everyone should think now about data sharing
 - Include consent/IRB language for sharing de-identified data
 - Learn about good data management
- You can apply for a CLDR rehabilitation-specific data sharing grant to cover your time/effort to archive existing data: <https://www.utmb.edu/cldr/pilot-projects/data-sharing-studies>
- Learn more about FAIR principles and reproducible methods for open science: <https://www.repronim.org/index.html>
- Happy to discuss best place to archive or other questions: sliew@usc.edu

Thank You and Fight On!



The Neural Plasticity and Neurorehabilitation Laboratory

<http://npnl.usc.edu>
sliew@usc.edu

 Twitter: @NPNLatUSC

USC Chan Division of Occupational
Science and Occupational Therapy

USC Division of Biokinesiology
and Physical Therapy

Keck Medicine USC Viterbi
of USC *Department of
Biomedical Engineering*

USC Stevens Neuroimaging
and Informatics Institute



USC University of
Southern California