Towards reproducible neuroimaging - across different analysis pipelines
Camille Maumet

To cite this version:
Camille Maumet. Towards reproducible neuroimaging - across different analysis pipelines. QLS 2024 - McGill Seminar Series in Quantitative Life Sciences and Medicine, Feb 2024, Montreal (Online), Canada. 2024. hal-04479231

HAL Id: hal-04479231
https://inria.hal.science/hal-04479231
Submitted on 27 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Towards reproducible neuroimaging across different analysis pipelines

McGill Seminar Series in Quantitative Life Sciences and Medicine

Camille Maumet
Inria, Univ Rennes, Inserm, CNRS, IRISA
A crisis in experimental research

The reproducibility crisis has led to reduced confidence in research findings

Low reproduction rates in many fields
Cancer research: <11%  Psychology: 36%
Medicine: 44%

(Begley & Ellis 2012) (Open Science Collab 2016) (Ioannidis 2005)
A crisis in experimental research

The **reproducibility crisis** has led to **reduced confidence in research findings**

Low reproduction rates in many fields
- Cancer research: <11%
- Psychology: 36%
- Medicine: 44%

(Begley & Ellis 2012) (Open Science Collab 2016) (Ioannidis 2005)

- **Wasted money & effort for research**
- **Delayed translation into clinical practice**
- **Reduced trust in science**

Credits: Money waste CC-BY ProSymbols thenounproject, Hospital bed CC-BY LAFS thenounproject, Yeah right! CC-BY Alfredo thenounproject
Reproducible research
Reproducible research

Original study

Data — Analysis — Results
Reproducible research

Original study

Data → Analysis → Results

≠ Data → ≠ Analysis → ≠ Results
Reproducible research

Original study

Data → Analysis → Results
Reproducible research

Original study

Data → Analysis → Results

≠ Data ≠ Analysis ≠ Results
Reproducible research

Original study

Data → Analysis → Results

Which release of SPM was used?
Which registration algorithm was selected?
How was the brain mask computed?

Did we understand the method correctly?
... or is this due to my use of a different dataset?
... a different computing platform?
... something else!?

Hum... we can’t reproduce the original results...
Reproducible research
Reproducible research

Original study

Data → Analysis → Results

Computational Reproducibility

≠ Data ≠ Analysis → Results

Replicability
Reproducible research

Original study

Data → Analysis → Results

same data

Different data

≠ Data

≠ Analysis

Results

Computational Reproducibility

Replicability

Data → Analysis → Results

same analysis

Different analysis
Reproducible research

Original study

Data → Analysis → Results

same data

Computational Reproducibility

Data → Analysis → Results

different data

≠ Data → Analysis → Results

≠ Analysis

Replicability

Data → ≠ Analysis → Results

≠ Data → ≠ Analysis → Results

same analysis → different analysis
Reproducible research

Original study

Data → Analysis → Results

Reproductions

Data → Analysis → Results
≠ Analysis → ≠ Results

Computational Reproducibility

Different data

≠ Data → Analysis → Results
≠ Data → ≠ Analysis → Results

Replicability

Same data

Data → Analysis → Results
≠ Analysis → ≠ Results

Data → Analysis → Results
≠ Analysis → ≠ Results

Same analysis

Different analysis
On our way to fix irreproducibility...
Reproducible research

Original study

Data → Analysis → Results

Reproducibility

Same data

Data → Analysis → Results
≠ Analysis → Results

Compared Reproducibility

Different data

Data ≠ Analysis → Results
≠ Data → Results

Replicability

Same analysis

Data ≠ Analysis → Results
≠ Data → Results

Different analysis
Fixing irreproducibility with different data

Small samples
Low statistical power
Power of neuroscience studies

High rate of false positives in the literature

(Button et. al, 2013)
Fixing irreproducibility with different data

Small samples
Low statistical power
Power of neuroscience studies

High rate of false positives in the literature

Lack of representativity and diversity of the samples

X-ray: Lung opacity detection
Model trained on male images, tested on female images

Lack of generalisability and fairness

(Button et. al, 2013)

(Larrazabal et. al, 2020)
Fixing irreproducibility with different data

Small samples
Low statistical power
Power of neuroscience studies

High rate of **false positives** in the literature

Lack of representativity and diversity of the samples

**X-ray: Lung opacity detection**
Model trained on male images, tested on female images

Solution: Data Sharing

(Button et. al, 2013)

(Larrazabal et. al, 2020)

Lack of **generalisability** and **fairness**
Neuroimaging data sharing

Single study
30 sub.
Neuroimaging data sharing

Single study
30 sub.

Consortium
1000 sub.

OpenNEURO
NEUROVAULT
NITRC
OSF
ABIDE
ADHD
CRR
1000 Functional Connectomes Project
nki
Neuroimaging data sharing

- **Single study**
  - 30 sub.

- **Consortium**
  - 1000 sub.

- **Cohort**
  - 100,000 sub.
Different pipelines... different results?
Reproducible research

Original study

Data → Analysis → Results

Reproducibility

Data → Analysis → Results
≠ Data → ≠ Analysis → ≠ Results
≠ Data → ≠ Analysis → ≠ Results
≠ Data → ≠ Analysis → ≠ Results

Computational Reproducibility

Replicability

→ same data

→ different data

→ same analysis

→ different analysis
Reproducible research

Original study

Data → Analysis → Results

Reproducibility

- Data
- ≠ Analysis
- Results

Computational Reproducibility

same data

- Data
- Analysis
- ≠ Results

different data

- ≠ Data
- Analysis
- ≠ Results

same analysis

- ≠ Data
- Analysis
- ≠ Results

different analysis

Replicability
A multiplicity of analysis pipelines

Data ≠ Analysis

Data → ≠ Analysis → Results
A multiplicity of analysis pipelines

1. Input data
2. Data preparation
3. Derived data
4. Statistical analysis
5. Results
A multiplicity of analysis pipelines
A multiplicity of analysis pipelines
A multiplicity of analysis pipelines

≠ algo.

≠ algorithm

Input data

Results
A multiplicity of analysis pipelines
A multiplicity of analysis pipelines

≠ algorithm

≠ software

≠ software version
A multiplicity of analysis pipelines

- Input data
- ≠ param.
- ≠ algorithm
- ≠ software
- ≠ software version
- ≠ parameters
- Results
A multiplicity of analysis pipelines

- ≠ algorithm
- ≠ software
- ≠ software version
- ≠ environment
- ≠ parameters
- ≠ environment
A multiplicity of analysis pipelines

... and many more...

≠ algorithm ≠ software ≠ software version ≠ parameters ≠ environment
Many-analyst project: NARPS

1 dataset

9 ‘yes/no’ questions

[Botvinik-Nezer et. al, Nature 2020]
Many-analyst project: NARPS

1 dataset

9 ‘yes/no’ questions

70 analysis pipelines by different teams

[Botvinik-Nezer et. al, Nature 2020]

Credits: Teams undraw.co
Many-analyst project: NARPS

Results over the 70 post-processing pipelines:

9 ‘yes/no’ questions

[Botvinik-Nezer et. al, Nature 2020]
Many-analyst project: NARPS

Results over the 70 post-processing pipelines:

- **Significant finding** (>80% of teams agree)
  - All 'yes'

- **Non-significant finding** (>80% of teams agree)
  - All 'no'

9 ‘yes/no’ questions

[Botvinik-Nezer et. al, Nature 2020]
Many-analyst project: NARPS

Results over the 70 post-processing pipelines:

- **Significant finding** (>80% of teams agree)
  - All 'yes'

- **Contradictory results** (Teams disagree)

- **Non-significant finding** (>80% of teams agree)
  - All 'no'

9 ‘yes/no’ questions

[Botvinik-Nezer et. al, Nature 2020]
Variability across neuroimaging software

Principles of functional MRI

Input data → Data preparation → Derived data → Statistical analysis → Results
Variability across neuroimaging software

Principles of functional MRI

Input data → Data preparation → Derived data → Statistical analysis → Results
Variability across neuroimaging software

Principles of functional MRI

Input data → Data preparation → Derived data → Statistical analysis → Results

fMRI results: Activations maps

Visual

Hand motion

Music

Credits: Examples of activations maps from Neurosynth [https://neurosynth.org/](https://neurosynth.org/) (Yarkoni et al, 2011)
Goal 1: Investigate the variability observed in fMRI results when using different software packages
Variability across neuroimaging software

Principles of functional MRI

Goal 1: Investigate the variability observed in fMRI results when using different software packages

- Reproduced 3 studies from the literature
- Using 3 fMRI software: AFNI, FSL, SPM (used in 80% literature (Carp, 2012))
- Pipelines aligned using available configurations.
Variability across neuroimaging software

- Variability in the activation maps

<table>
<thead>
<tr>
<th>study ds001</th>
<th>study ds109</th>
<th>study ds120</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Brain Activation Map" /></td>
<td><img src="image2" alt="Brain Activation Map" /></td>
<td><img src="image3" alt="Brain Activation Map" /></td>
</tr>
</tbody>
</table>

AFNI software

FSL software

SPM software

Original paper

[Bowring Maumet* Nichols*, HBM 2019]
Variability across neuroimaging software

- Variability in the activation maps

[Bowring, Maumet*, Nichols*, HBM 2019]
Variability across neuroimaging software

- Variability in the activation maps

[Bowring Maumet* Nichols*, HBM 2019]
Variability across neuroimaging software

- Variability in the activation maps

**Dice coefficient**

<table>
<thead>
<tr>
<th>Software</th>
<th>AFNI</th>
<th>FSL</th>
<th>SPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFNI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSL</td>
<td>0.3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SPM</td>
<td>0.5</td>
<td>0.4</td>
<td>1</td>
</tr>
</tbody>
</table>

[Bowring Maumet* Nichols*, HBM 2019]
Variability across neuroimaging software

Variability in the unthresholded statistic maps

<table>
<thead>
<tr>
<th>Study</th>
<th>AFNI software</th>
<th>FSL software</th>
<th>SPM software</th>
</tr>
</thead>
<tbody>
<tr>
<td>ds001</td>
<td><img src="image1" alt="AFNI" /></td>
<td><img src="image2" alt="FSL" /></td>
<td><img src="image3" alt="SPM" /></td>
</tr>
<tr>
<td>ds109</td>
<td><img src="image4" alt="AFNI" /></td>
<td><img src="image5" alt="FSL" /></td>
<td><img src="image6" alt="SPM" /></td>
</tr>
<tr>
<td>ds120</td>
<td><img src="image7" alt="AFNI" /></td>
<td><img src="image8" alt="FSL" /></td>
<td><img src="image9" alt="SPM" /></td>
</tr>
</tbody>
</table>

**Correlations**
- AFNI vs. FSL: 0.62
- AFNI vs. SPM: 0.61
- FSL vs. SPM: 0.68

**Conclusion:** The choice of software impacted both the final activation and statistic maps.

[Bowring Maumet* Nichols*, HBM 2019]
Variability across neuroimaging software

Goal 2: Investigate how remaining differences in software implementations can explain the observed differences

study ds001

1. Preprocessing
2. Subject-level Signal
3. Subject-level Drift
4. Subject-level Noise
5. Group-level Stat

[Bowring Nichols* Maumet*, HBM 2021]
Variability across neuroimaging software

Goal 2: Investigate how remaining differences in software implementations can explain the observed differences

study ds001

1. Preprocessing
   2. Subject-level Signal
   3. Subject-level Drift
   4. Subject-level Noise
   5. Group-level Stat

Conclusion: 1 main contributing factor was found but different for each study

[Bowring Nichols* Maumet*, HBM 2021]
Understanding analytical variability
Understanding analytical variability...
Understanding analytical variability...

- Numerical uncertainty (e.g. ≠ OS)
- Unwanted variability
- Different implem.
Understanding analytical variability...

- Numerical uncertainty (e.g. ≠ OS)
- Assumption violations
- Unwanted variability
- Different implem.
- Unreliable results
- Data induced variability
Understanding analytical variability...

- Numerical uncertainty (e.g. ≠ OS)
- Different implem.
- Unwanted variability
- Assumption violations
- Better fit to the data
- Variability of interest
- Data induced variability
- Unreliable results

?
Describing the neuroimaging pipeline space
NARPS open pipelines

1 dataset

9 ‘yes/no’ questions

1- H1: “yes”
2- Activation + Statistic brain maps
3- Textual description of the methods

70 analysis pipelines by different teams

Credits: Teams undraw.co.
**NARPS open pipelines**

1 dataset

9 ‘yes/no’ questions

70 analysis pipelines by different teams

1- H1: “yes”
2- Activation + Statistic brain maps
3- Textual description of the methods

https://github.com/Inria-Empenn/narps_open_pipelines/

Credits: Teams undraw.co
Credits - This project was presented and received contributions during the following events:

- Empenn hackathon (February 2024)
- [Brainhack Marseille 2023](#) (December 2023)
- [ORIGAMI lab](#) hackathon (September 2023)
- [OHBM Brainhack 2023](#) (July 2023): Arshitha Basavaraj, Boris Clénet, Rémi Gau, Élodie Germani, Yaroslav Halchenko, Camille Maumet, Paul Taylor.
- [e-ReproNim FENS NENS Cluster Brainhack](#) (June 2023): Liz Bushby, Boris Clénet, Michael Dayan, Aimee Westbrook.
- [OHBM Brainhack 2022](#) (June 2022): Elodie Germani, Arshitha Basavaraj, Trang Cao, Rémi Gau, Anna Menacher, Camille Maumet.

[https://github.com/Inria-Empenn/narps_open_pipelines/](https://github.com/Inria-Empenn/narps_open_pipelines/)
Tracing provenance in neuroimaging

“Provenance describe how an artifact or set of artifacts was produced” Wikipedia

- Based on W3C PROV
- As part of neuroimaging standardization efforts

https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/
Provenance Use Case 1 An unambiguous description of the methods
Provenance Use Case 1: An unambiguous description of the methods

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?
Provenance Use Case 1

An unambiguous description of the methods

Can we reproduce exactly the same results (same data, same methods)?
Provenance Use Case 1: An unambiguous description of the methods

Can we reproduce exactly the same results (same data, same methods)?
Provenance Use Case 2 Dealing with unwanted variability

A previously widely-used parameter is no longer considered as best practice.
A previously widely-used parameter is no longer considered as best practice.

A bug was found in a given tool version, can we retrieve all corresponding results?
Provenance Use Case 2: Dealing with unwanted variability

A previously widely-used parameter is no longer considered as best practice.

A bug was found in a given tool version, can we retrieve all corresponding results?
Provenance Use Case 3 Exploring factors impacting analytical variability

Are my conclusions dependant on the size of the smoothing kernel used at Step 5?

BIDS-Prov will soon be open for community feedback!

http://bids.neuroimaging.io/bep028
Which solutions for practitioners?
Same data meta-analysis

**Goal**: Investigate the validity of same-data meta-analysis methods

[Lefort-Besnard Nichols*, Maumet*, OHBM 2024]
Validity of mega-analyses in neuroimaging

Multi-stage functional MRI analysis

Input data → Data preparation → Derived data → Statistical analysis → Results

Xavier Rolland
Pierre Maurel
Validity of mega-analyses in neuroimaging

Multi-stage functional MRI analysis

Subject 3
Data preparation
Derived data 1
Statistical analysis
Results

Subject 2
Data preparation
Derived data 2

Subject N
Data preparation
Derived data N

Subject-Level

Group-Level
Validity of mega-analyses in neuroimaging

Multi-stage functional MRI analysis
Validity of mega-analyses in neuroimaging

Multi-stage functional MRI analysis

Subject 3

Data preparation

Derived data 1

Statistical analysis

Results

Subject 2

Data preparation

Derived data 2

Subject N

Data preparation

Derived data N

Subject-Level

Group-Level
Validity of mega-analyses in neuroimaging

Multi-stage functional MRI analysis

Subject-Level

Data preparation

Derived data 1

Statistical analysis

Results

Group-Level

Data preparation

Derived data 2

Derived data N

Goal: Investigate the validity of group statistics using subject data prepared with different pipelines

Subject 3

Subject 2

Subject N
Validity of mega-analyses in neuroimaging

**Goal:** Investigate the **validity of group statistics** using **subject data** prepared with **different pipelines**

- HCP Young-Adult (1000 subjects)
- **3 pipeline variations**
  - Spatial smoothing: 5 or 8 mm
  - Modelling of the brain response: with/out derivative
  - Correction for subject motion: 0,6, 24 covariates

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Pipelines 1 &amp; 2</th>
<th>Group-level statistics</th>
<th>Empirical false positive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1000 repetitions</td>
<td></td>
</tr>
</tbody>
</table>

**Subject-level data preparation**

Xavier Rolland
Pierre Maurel
Validity of mega-analyses in neuroimaging

"Reference" pipeline:
- canonical HRF
- smoothing 5mm
- 24 motion regressors

P-p plots (Bland-Altman)

Conclusion: Invalidity of the mega-analyses for multiple variations studied (e.g. different smoothing, 1+ variation)

[Rolland Maurel Maumet, ISBI 2022]
Leveraging large and open datasets

Shared data with limited metadata

500 000+ images
A variety of fMRI tasks, preprocessing and analysis
Minimal metadata required.
Leveraging large and open datasets

Shared data with limited metadata

500 000+ images

A variety of fMRI tasks, preprocessing and analysis

Minimal metadata required.

**Goal:** Can we take advantage of a large pool of unlabelled data to help answer new research questions?
Leveraging large and open datasets

Shared data with limited metadata

500 000+ images
A variety of fMRI tasks, preprocessing and analysis
Minimal metadata required.

**Goal:** Can we **take advantage of a large pool of unlabelled data** to help answer **new research questions**?

- **Step 1:** Build a model-agnostic representation of fMRI statistic maps

![Diagram](image)

- **Step 2:** Fine-tune a classifier (predicting cognitive process: 🤔 🖥️)
Leveraging large and open datasets

- Tested on two datasets:
  - Single large study (HCP Young adult) 23 tasks x 787 subjects
  - Collection of studies (Brainpedia) 29 studies, 36 tasks, 826 subjects

Conclusion: The pre-trained classifier always performed better especially when the problem was complex (small samples, heterogeneity).

[Germani Fromont* Maumet*, Gigascience 2023]
Towards reproducible neuroimaging across different analysis pipelines

Camille Maumet

Thank you!

Credit: Presentation template and icons by SlidesCarnival, adapted

@cmaumet.bsky.social
@cmaumet@fediscience.org