



HAL
open science

MRI Post-processing Platforms and How they Compare

Camille Maumet

► **To cite this version:**

Camille Maumet. MRI Post-processing Platforms and How they Compare. MRI Together 2022, Dec 2022, Online, United Kingdom. 2022. hal-04032032

HAL Id: hal-04032032

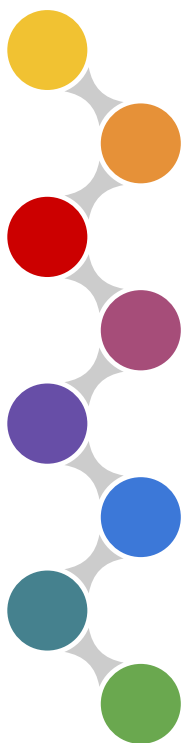
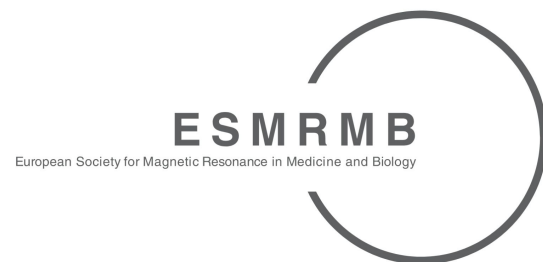
<https://inria.hal.science/hal-04032032>

Submitted on 16 Mar 2023

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.

MRI Together Global workshop on open,
reproducible, and inclusive MR research



Hands-on Juggling with data
**MRI Post-processing Platforms
and How they Compare**

December 8, 2022

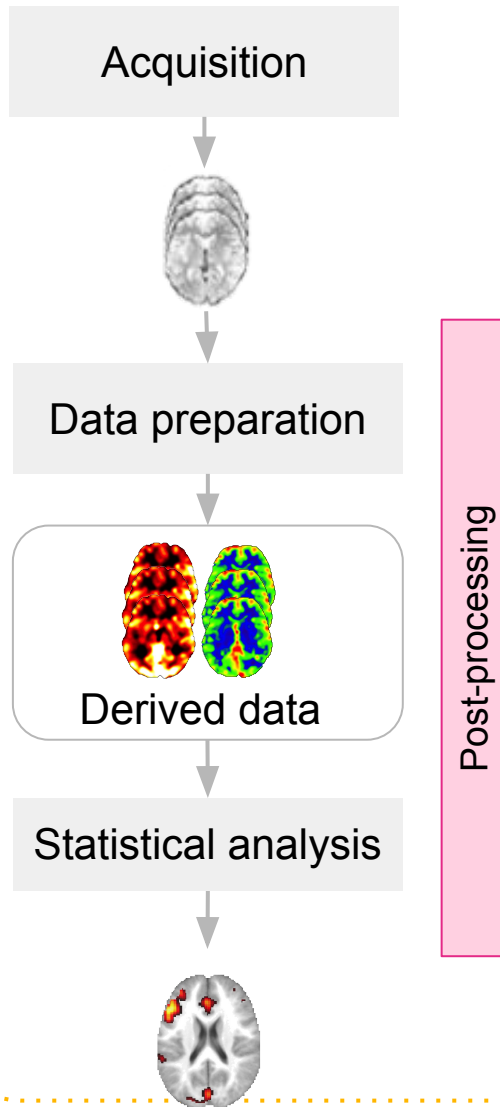
Camille Maumet

Inria, Univ Rennes, Inserm, CNRS, IRISA

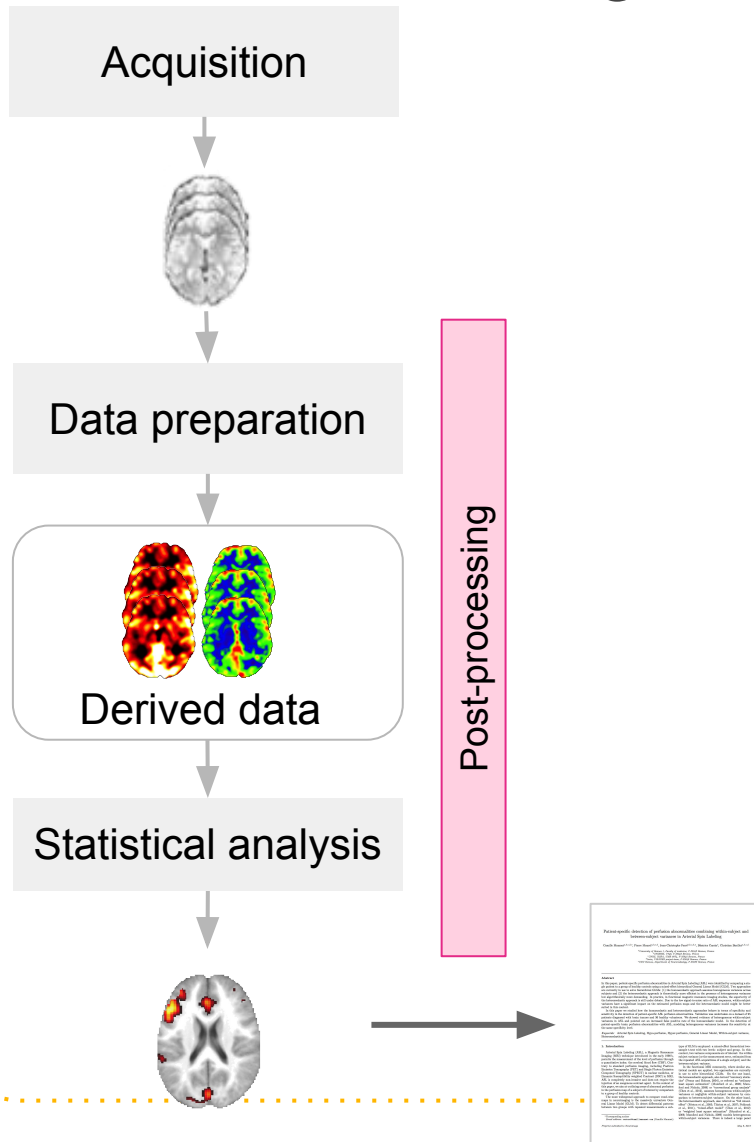


UMR **IRISA**

An MRI study



An MRI study

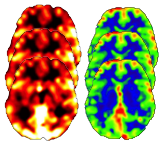


An MRI study

Acquisition

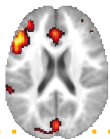


Data preparation



Derived data

Statistical analysis

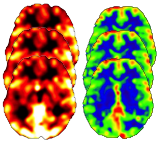


An MRI study

Acquisition

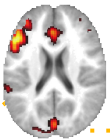


Data preparation



Derived data

Statistical analysis



An MRI study

tion

aration

data

analysis



Reproducing an MRI study



Reproducing an MRI study



Reproducing an MRI study



Which release of SPM was used?

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... I can't reproduce the original results...

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

Reproducing an MRI study



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

... or is this due to my use of a different dataset?

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Reproducing an MRI study



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Did I understand the method correctly?

... or is this due to my use of a different dataset?

... a different computing platform?

Reproducing an MRI study



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Did I understand the method correctly?

... or is this due to my use of a different dataset?

... a different computing platform?

... something else!?

Reproducing an MRI study



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

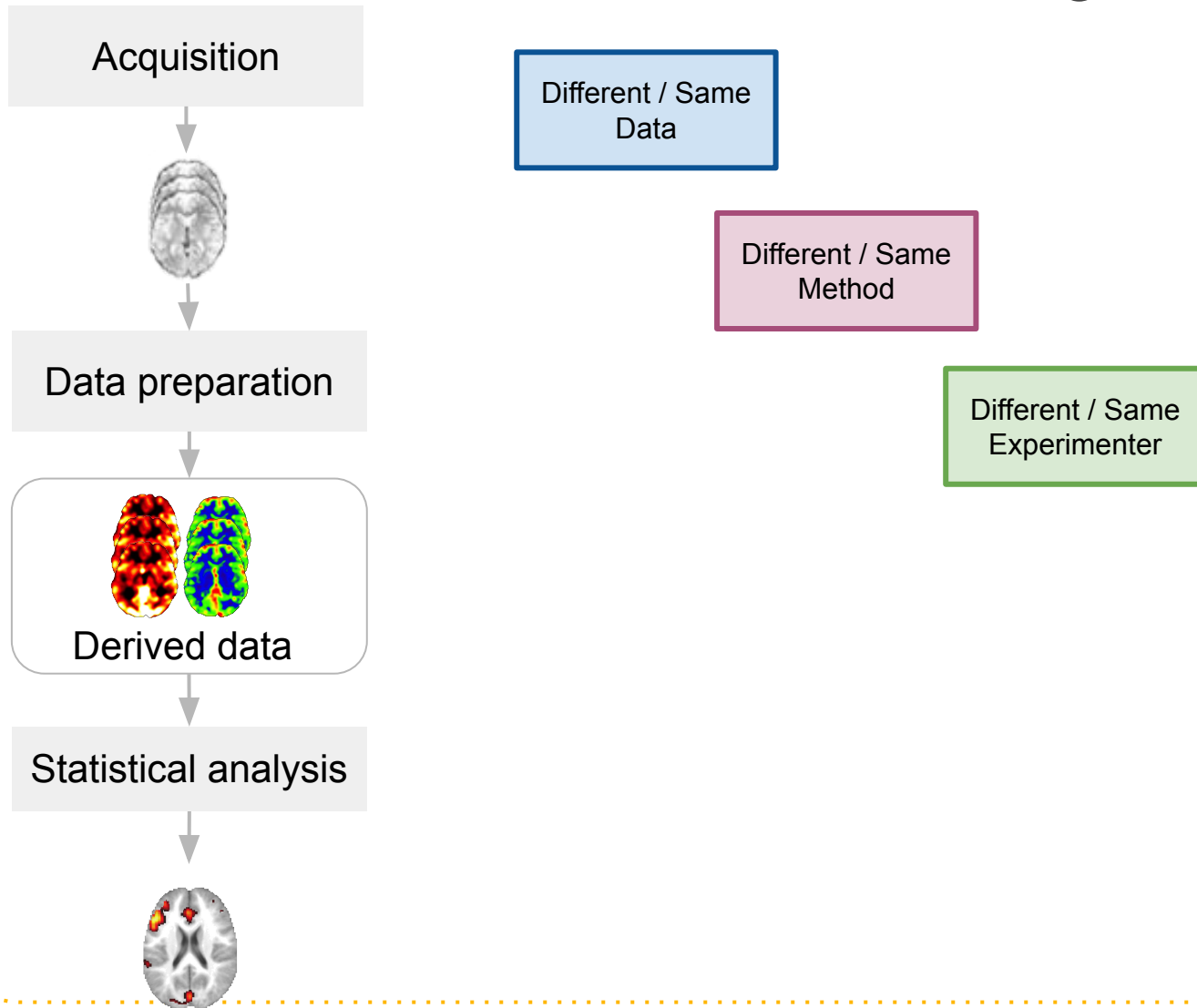
... or is this due to my use of a different dataset?

... a different computing platform?

... something else!?

2. When results cannot be reproduced, **it is often very difficult to understand why** this is happening

Reproducible MRI study



Definitions of reproducible neuroimaging (COBIDAS)

| Levels of generalization | Participants | | MRI acquisition | | | Experiment | | Analysis | | Personnel | |
|--|--------------|--------|-----------------|-------|------|---------------------|-----------------|----------|------|----------------------|------|
| | Population | Sample | Scanner | Visit | Data | Stimulus population | Stimulus sample | Method | Code | Experimenter analyst | Data |
| Generalization over measurements | | | | | | | | | | | |
| ISO repeatability (e.g., 30-min intrascanner reliability) | • | • | • | • | D | • | • | • | • | • | • |
| ISO intermediate reproducibility (e.g., 7-d intrascanner reliability) | • | • | • | D | D | • | • | • | • | • | • |
| ISO reproducibility (e.g., 7-d interscanner reliability) | • | • | D | D | D | • | • | • | • | • | • |
| Generalization over analyses | | | | | | | | | | | |
| Analysis replicability | • | • | • | • | • | • | • | • | • | • | • |
| Collegial analysis replicability | • | • | • | • | • | • | • | • | • | • | D |
| Peng5 reproducibility | • | • | • | • | • | • | • | • | D | D | D |
| Generalization over materials and methods | | | | | | | | | | | |
| Near replicability (different subjects) | • | D | • | – | – | • | • | • | • | • | • |
| Intermediate replicability (different labs) | • | D | D | – | – | • | • | • | • | D | D |
| Far replicability (different experimental & analytical methods) | • | D | D | – | – | • | D | D | D | D | D |
| Hypothesis generalizability (different subject populations & types of stimuli) | D | D | D | – | – | D | D | D | D | D | D |

Table 1. A partial taxonomy of reproducibility in neuroimaging

Definitions of reproducible neuroimaging (COBIDAS)

| Levels of generalization | Participants | | MRI acquisition | | | Experiment | | Analysis | | Personnel | | |
|--|--------------|--------|-----------------|-------|------|---------------------|-----------------|----------|------|----------------------|------|---|
| | Population | Sample | Scanner | Visit | Data | Stimulus population | Stimulus sample | Method | Code | Experimenter analyst | Data | |
| Generalization over measurements | | | | | | | | | | | | |
| ISO repeatability (e.g., 30-min intrascanner reliability) | • | • | • | • | D | • | • | • | • | • | • | |
| ISO intermediate reproducibility (e.g., 7-d intrascanner reliability) | • | • | • | D | D | • | • | • | • | • | • | |
| ISO reproducibility (e.g., 7-d interscanner reliability) | • | • | D | D | D | • | • | • | • | • | • | |
| Generalization over analyses | | | | | | | | | | | | |
| Analysis replicability | • | • | • | • | • | • | • | • | • | • | • | |
| Collegial analysis replicability | • | • | • | • | • | • | • | • | • | • | D | |
| Peng5 reproducibility | • | • | • | • | • | • | • | • | D | D | D | |
| Generalization over materials and methods | | | | | | | | | | | | |
| Near replicability (different subjects) | Dataset | | | - | - | • | • | • | • | • | • | |
| Intermediate replicability (different labs) | | | | - | - | • | • | • | • | D | D | |
| Far replicability (different experimental & analytical methods) | | | | - | - | • | D | D | D | D | D | D |
| Hypothesis generalizability (different subject populations & types of stimuli) | | | | - | - | D | D | D | D | D | D | D |

Table 1. A partial taxonomy of reproducibility in neuroimaging

Definitions of reproducible neuroimaging (COBIDAS)

| Levels of generalization | Participants | | MRI acquisition | | | Experiment | | Analysis | | Personnel | | | |
|--|--------------|--------|--------------------|-------|-------------------|---------------------|------------------------|----------|------|----------------------|------|---|---|
| | Population | Sample | Scanner | Visit | Data | Stimulus population | Stimulus sample | Method | Code | Experimenter analyst | Data | | |
| Generalization over measurements | | | | | | | | | | | | | |
| ISO repeatability (e.g., 30-min intrascanner reliability) | • | • | Method Acquisition | | | • | • | • | • | • | • | | |
| ISO intermediate reproducibility (e.g., 7-d intrascanner reliability) | • | • | | | | • | • | • | • | • | • | • | |
| ISO reproducibility (e.g., 7-d interscanner reliability) | • | • | | | | • | • | • | • | • | • | • | |
| Generalization over analyses | | | | | | | | | | | | | |
| Analysis replicability | • | • | • | • | • | • | • | • | • | • | • | | |
| Collegial analysis replicability | • | • | • | • | • | • | • | • | • | • | D | | |
| Peng5 reproducibility | • | • | • | • | • | • | • | • | D | D | D | | |
| Generalization over materials and methods | | | | | | | | | | | | | |
| Near replicability (different subjects) | Dataset | | - | - | • | • | • | • | • | • | • | | |
| Intermediate replicability (different labs) | | | - | - | • | • | • | • | • | • | D | D | |
| Far replicability (different experimental & analytical methods) | | | - | - | Method Experiment | | Method Post-processing | | • | • | • | • | D |
| Hypothesis generalizability (different subject populations & types of stimuli) | | | - | - | | | | | • | • | • | • | • |

Table 1. A partial taxonomy of reproducibility in neuroimaging

Definitions of reproducible neuroimaging (COBIDAS)

| Levels of generalization | Participants | | MRI acquisition | | | Experiment | | Analysis | | Personnel | | | |
|--|--------------|--------|--------------------|-------|-------------------|---------------------|------------------------|----------|------|------------------------------|------|---|---|
| | Population | Sample | Scanner | Visit | Data | Stimulus population | Stimulus sample | Method | Code | Experimenter analyst | Data | | |
| Generalization over measurements | | | | | | | | | | | | | |
| ISO repeatability (e.g., 30-min intrascanner reliability) | • | • | Method Acquisition | | | • | • | • | • | • | • | | |
| ISO intermediate reproducibility (e.g., 7-d intrascanner reliability) | • | • | | | | • | • | • | • | • | • | • | |
| ISO reproducibility (e.g., 7-d interscanner reliability) | • | • | | | | • | • | • | • | • | • | • | |
| Generalization over analyses | | | | | | | | | | | | | |
| Analysis replicability | • | • | • | • | • | • | • | • | • | Who is doing the replication | | | |
| Collegial analysis replicability | • | • | • | • | • | • | • | • | • | | | | |
| Peng5 reproducibility | • | • | • | • | • | • | • | • | D | | | | |
| Generalization over materials and methods | | | | | | | | | | | | | |
| Near replicability (different subjects) | Dataset | | - | - | • | • | • | • | • | • | • | | |
| Intermediate replicability (different labs) | | | - | - | • | • | • | • | • | • | D | D | |
| Far replicability (different experimental & analytical methods) | | | - | - | Method Experiment | | Method Post-processing | | • | • | • | • | D |
| Hypothesis generalizability (different subject populations & types of stimuli) | | | - | - | | | | | • | • | • | • | • |

Table 1. A partial taxonomy of reproducibility in neuroimaging

On our way **to fix irreproducibility**

Irreproducible with...

Same Data

On our way **to fix irreproducibility**

Irreproducible with...

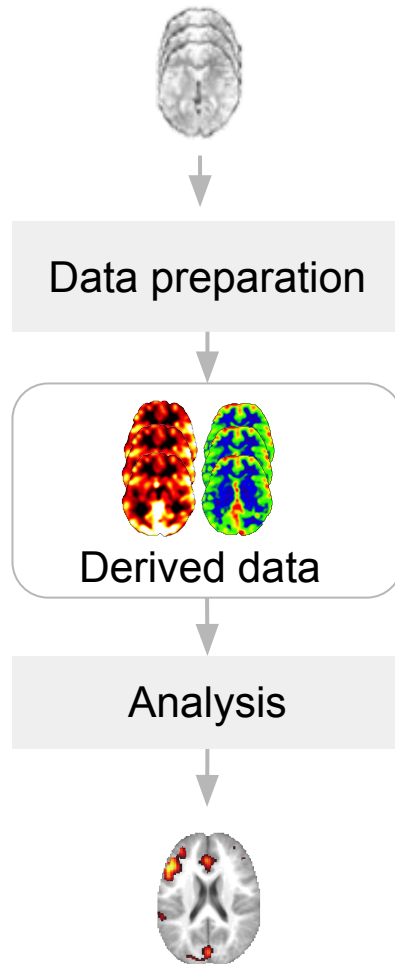
Same Data

Solutions: Sharing code,
containerization, etc.

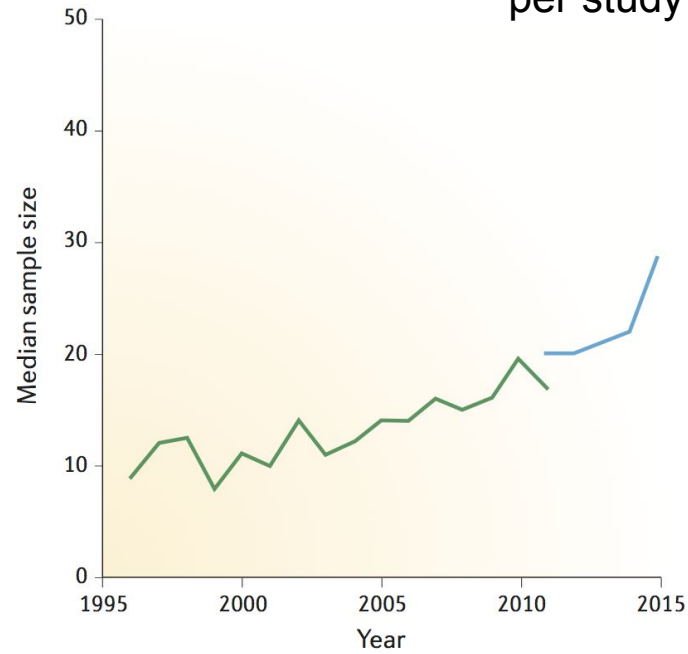
On our way **to fix irreproducibility**

Irreproducible with...

Different Data



about 30 participants
per study



[Poldrack et. al, Nature Neuroscience 2017]

1. False positive finding

Low statistical power

SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



SEAN NEL / SHUTTERSTOCK

In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

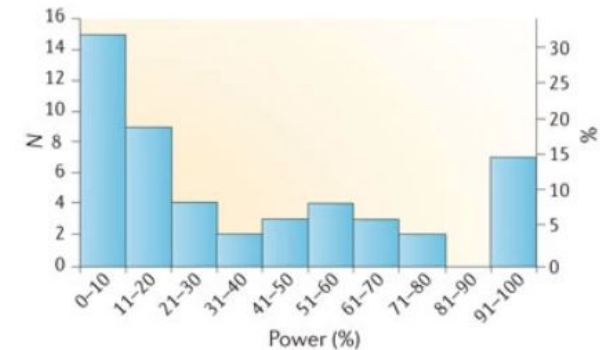
It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

1. False positive finding

Low statistical power

Power of neuroscience studies

Power = Prob. to correctly find a significant effect when a the alternative hypothesis is true.



[Button et. al, Nat Rev Neurosci 2013]

SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



SEAN NEL / SHUTTERSTOCK

In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

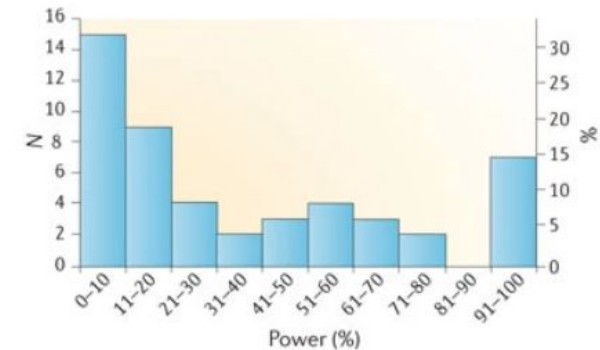
It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

1. False positive finding

Low statistical power

Power of neuroscience studies

Power = Prob. to correctly find a significant effect when a the alternative hypothesis is true.



[Button et. al, Nat Rev Neurosci 2013]

Towards bigger datasets

SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



SEAN NEL / SHUTTERSTOCK

In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

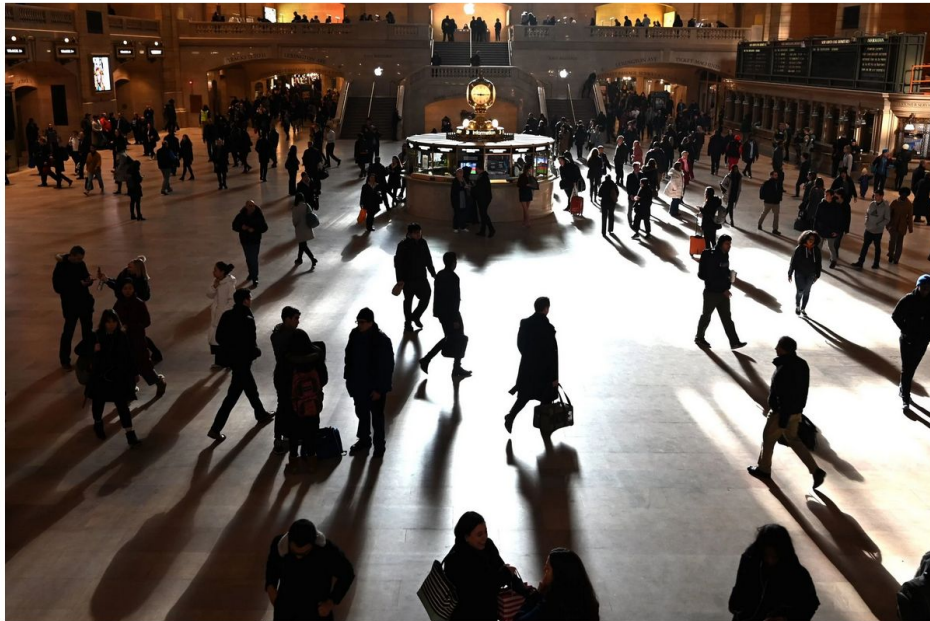
It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

2. Lack of generalizability and equitability

The New York Times

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images

Lack of representativity and diversity

2. Lack of generalizability and equitability

The New York Times

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

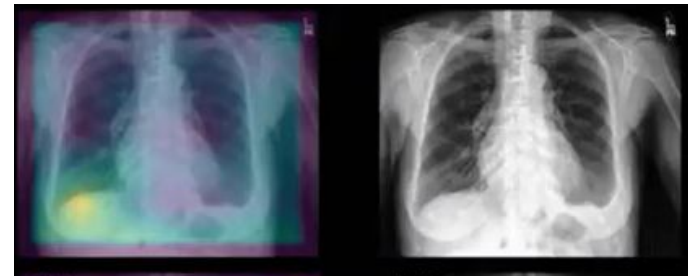


Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images

Lack of representativity and diversity

X-ray: Lung opacity detection

Model trained on male images, tested on female images



[Larrazabal et. al, PNAS 2020]

2. Lack of generalizability and equitability

The New York Times

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

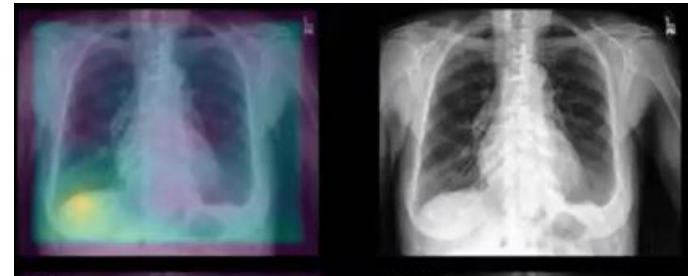


Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images

Lack of representativity and diversity

X-ray: Lung opacity detection

Model trained on male images, tested on female images



[Larrazabal et. al, PNAS 2020]

We need representative
and diverse data
and teams

Open data

Unique study
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T Q R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R

NITRC

OSF

- + Images
- + Homogenous
- Datasets



Open data

Unique study
30 participants



OpenNEURO

studyforrest.org

Consortium
1000 participants



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project

NEUROVAULT

L I E L N L F T Q K T Q R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R



NITRC

OSF

CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



- + Images
- + Homogenous
- Datasets



Open data

Unique study
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T Q R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R

Consortium
1000 participants



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project



NITRC

OSF

CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



Cohort
100 000 participants



HUMAN
Connectome
PROJECT

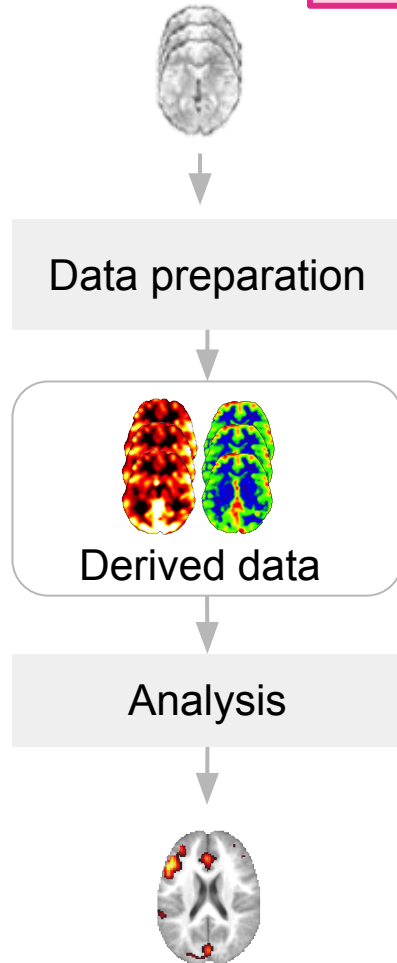
biobank^{uk}
Improving the health of future generations

- + Images
- + Homogenous
- Datasets

On our way **to fix irreproducibility**

Irreproducible with...

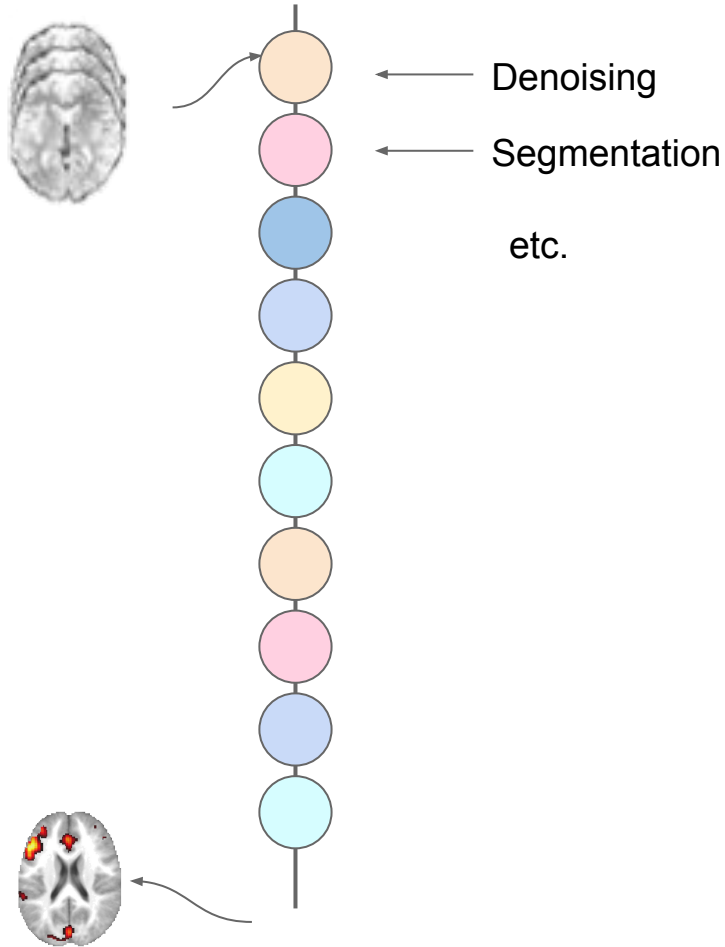
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

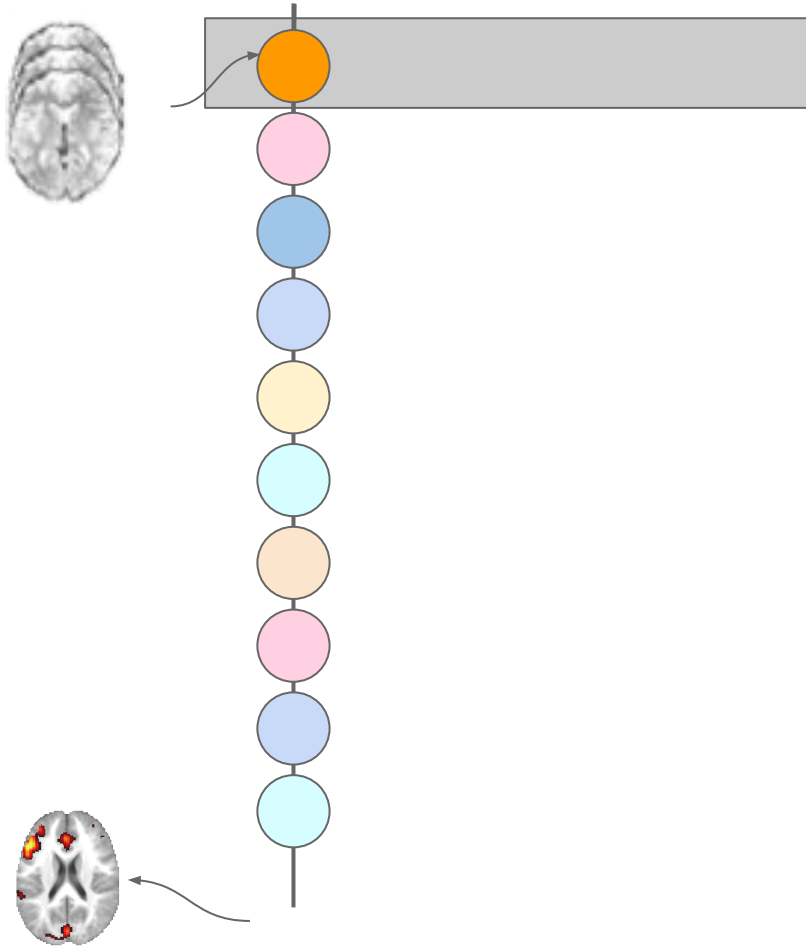
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

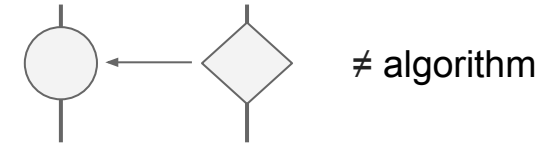
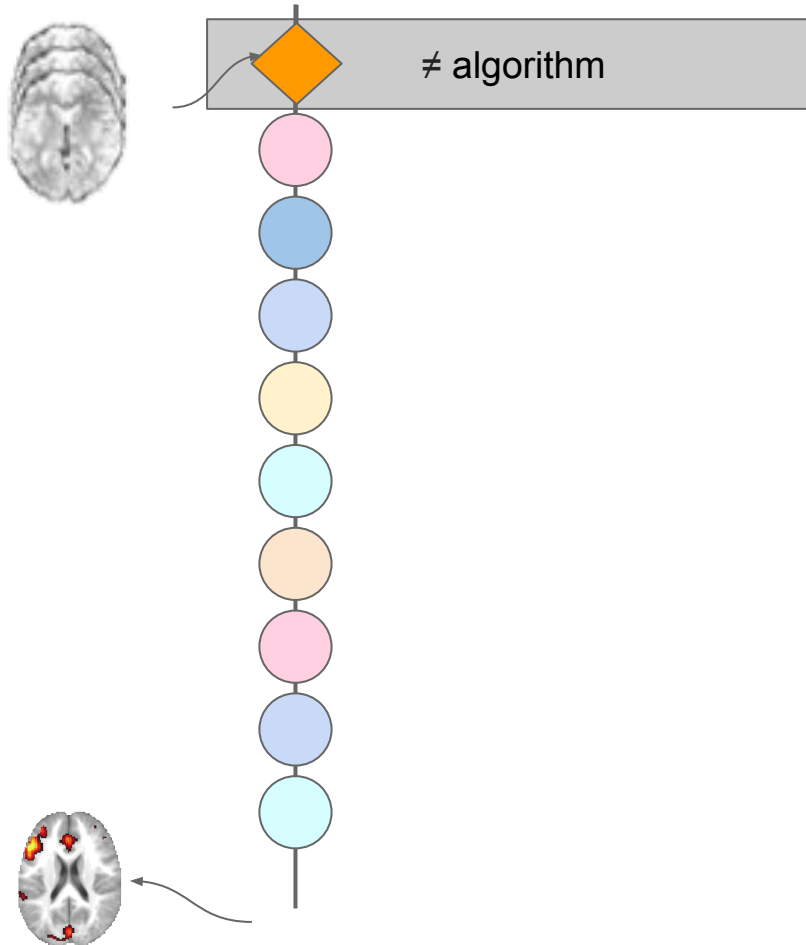
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

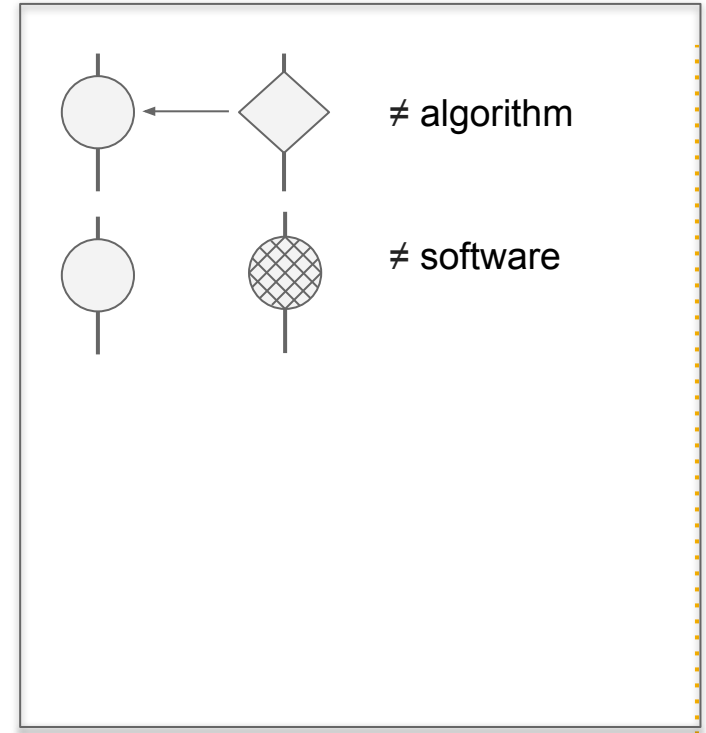
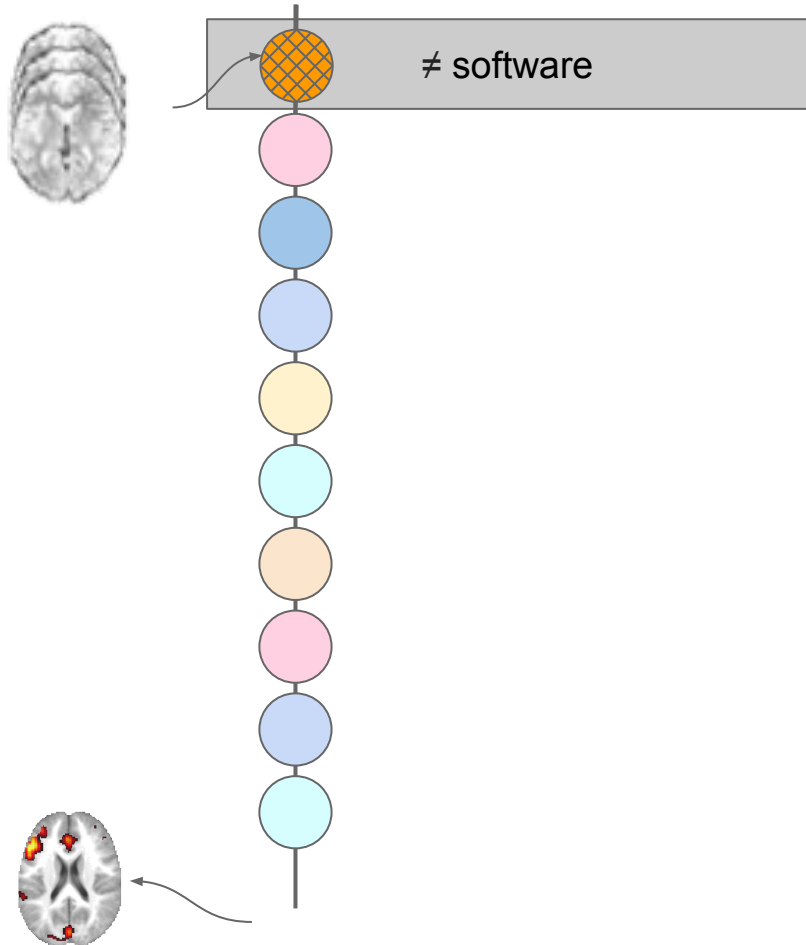
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

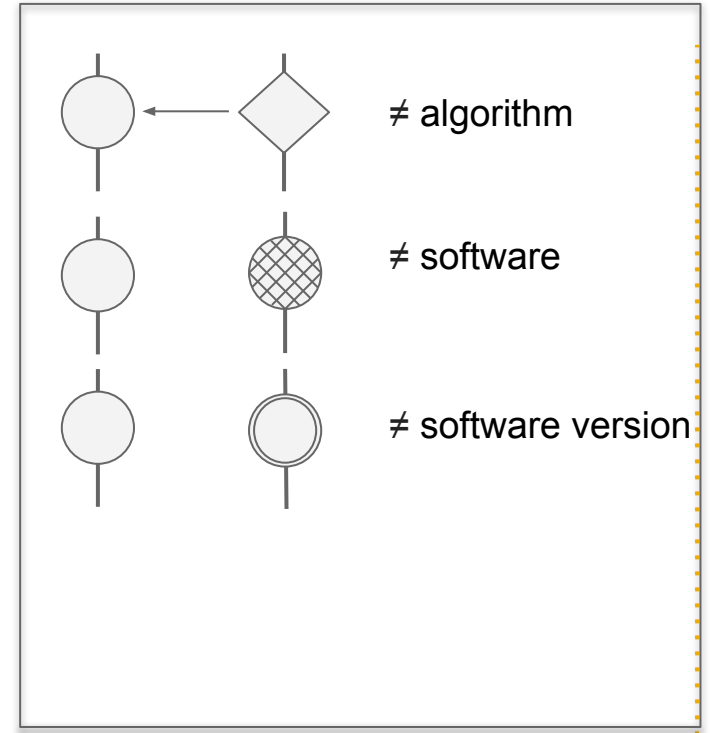
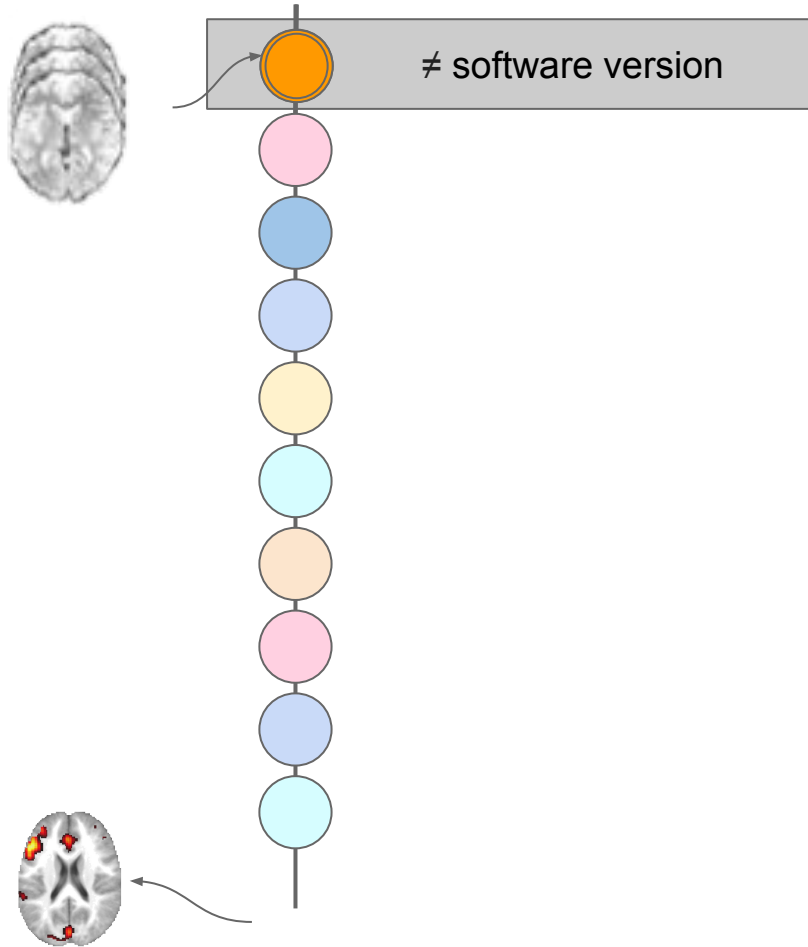
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

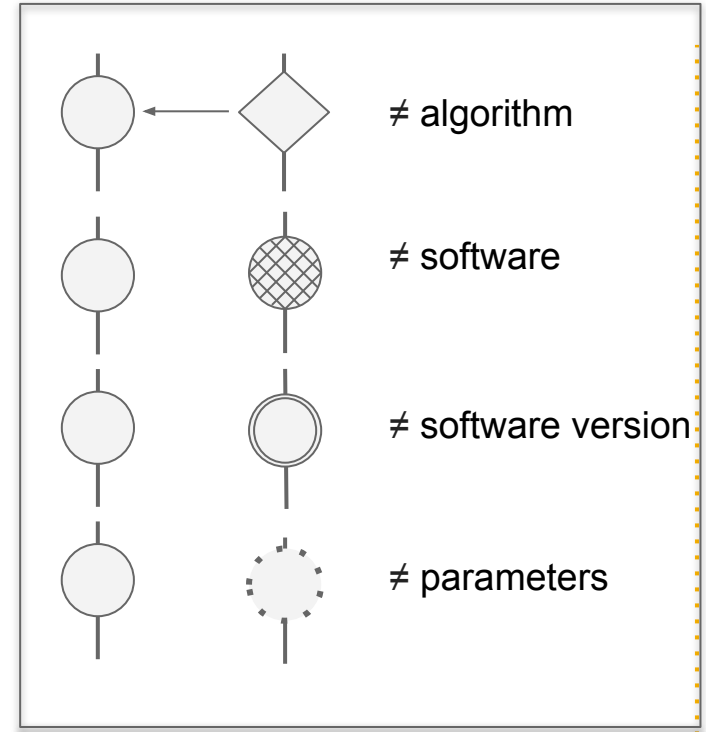
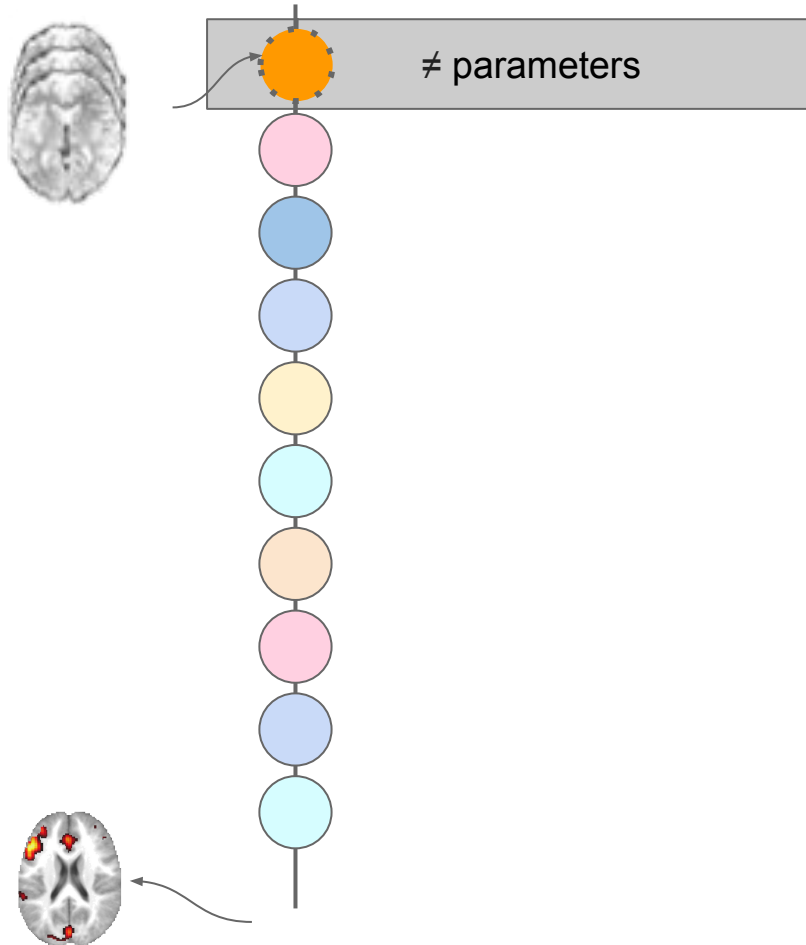
Different Post-processing Methods



On our way **to fix irreproducibility**

Irreproducible with...

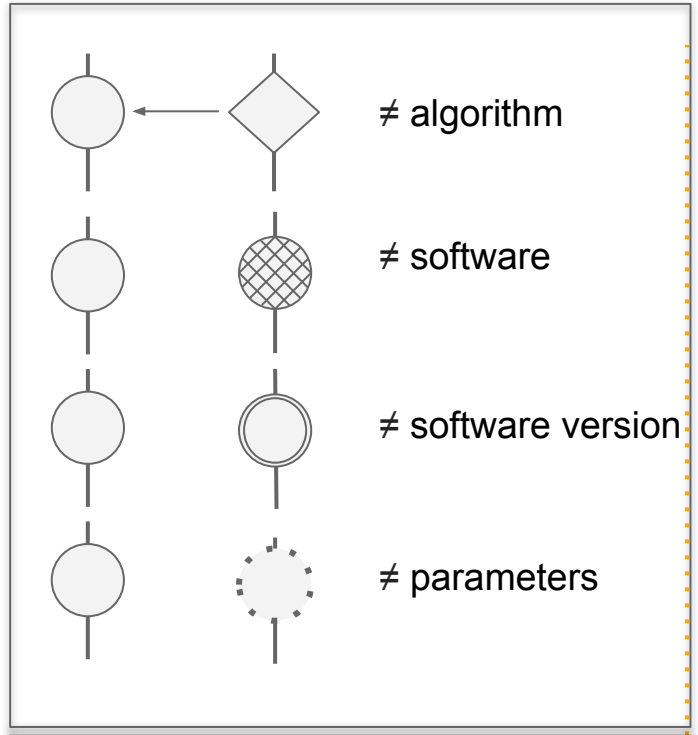
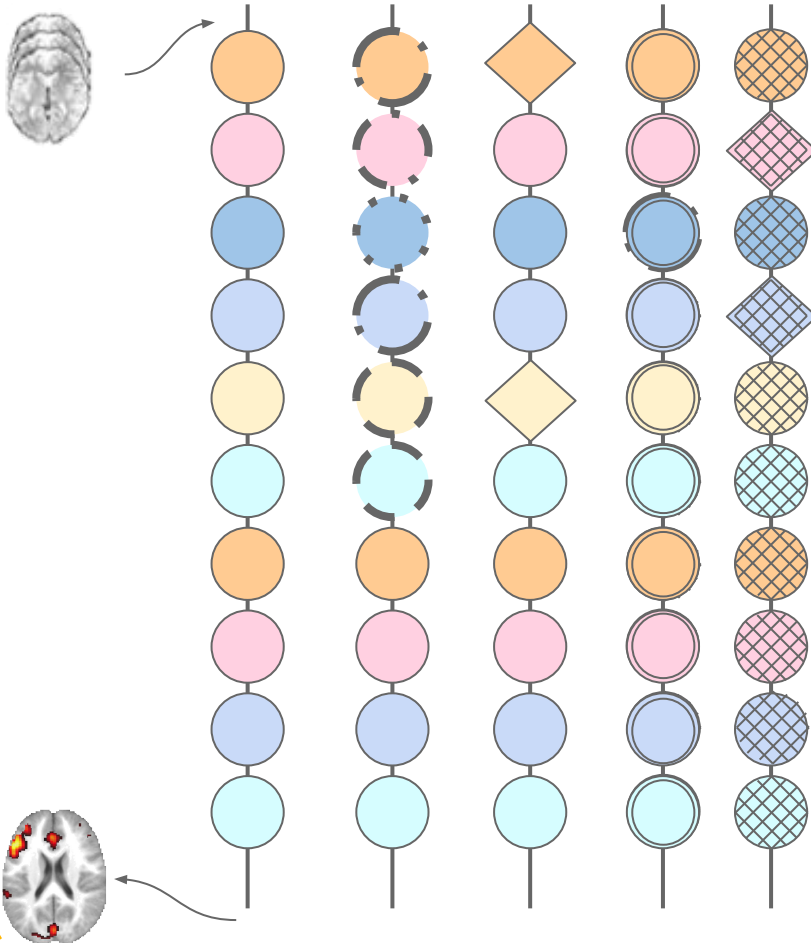
Different Post-processing Methods



On our way to fix irreproducibility

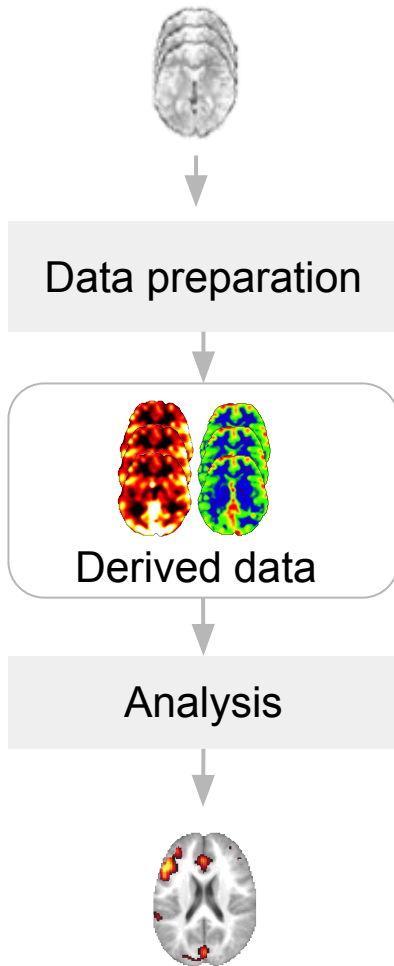
Irreproducible with...

Different Post-processing Methods

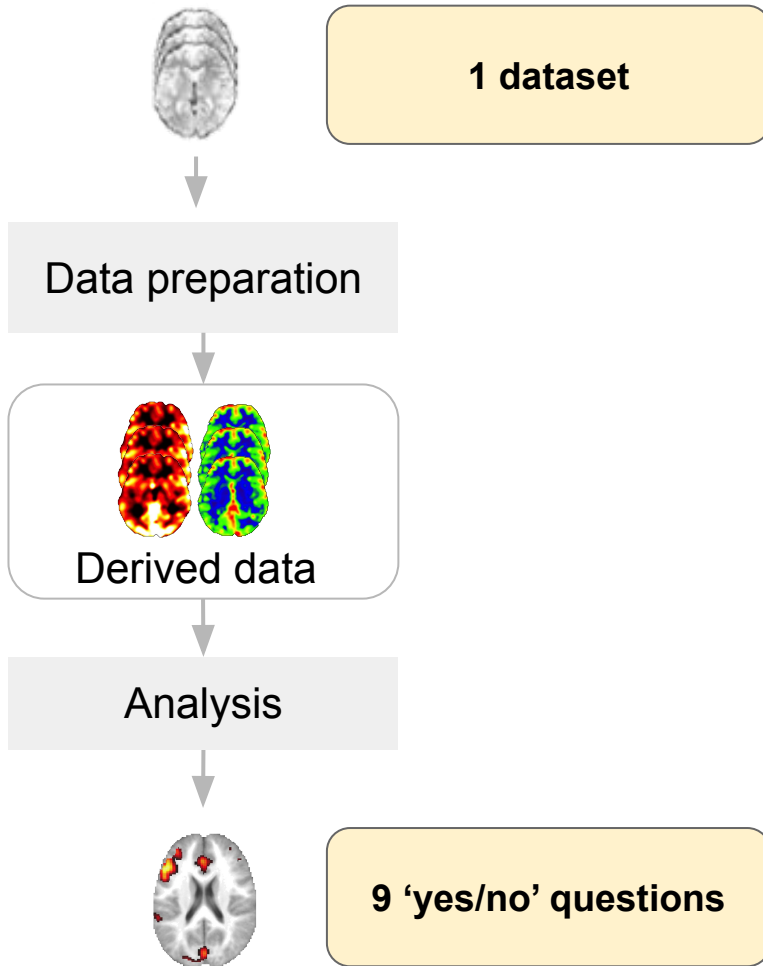


A family of acceptable pipelines
over 10^{30} combinations...

Many analysts project : NARPS

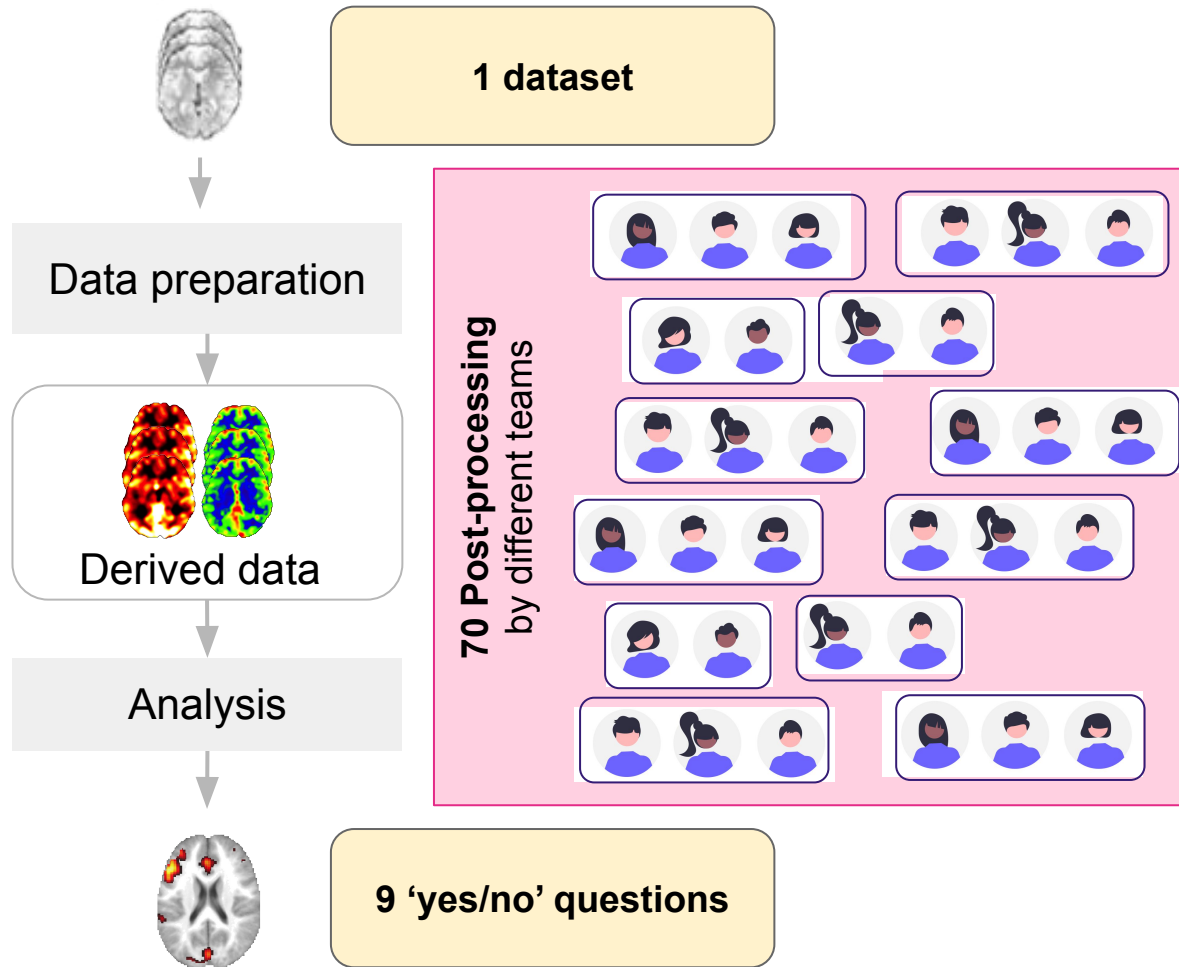


Many analysts project : NARPS



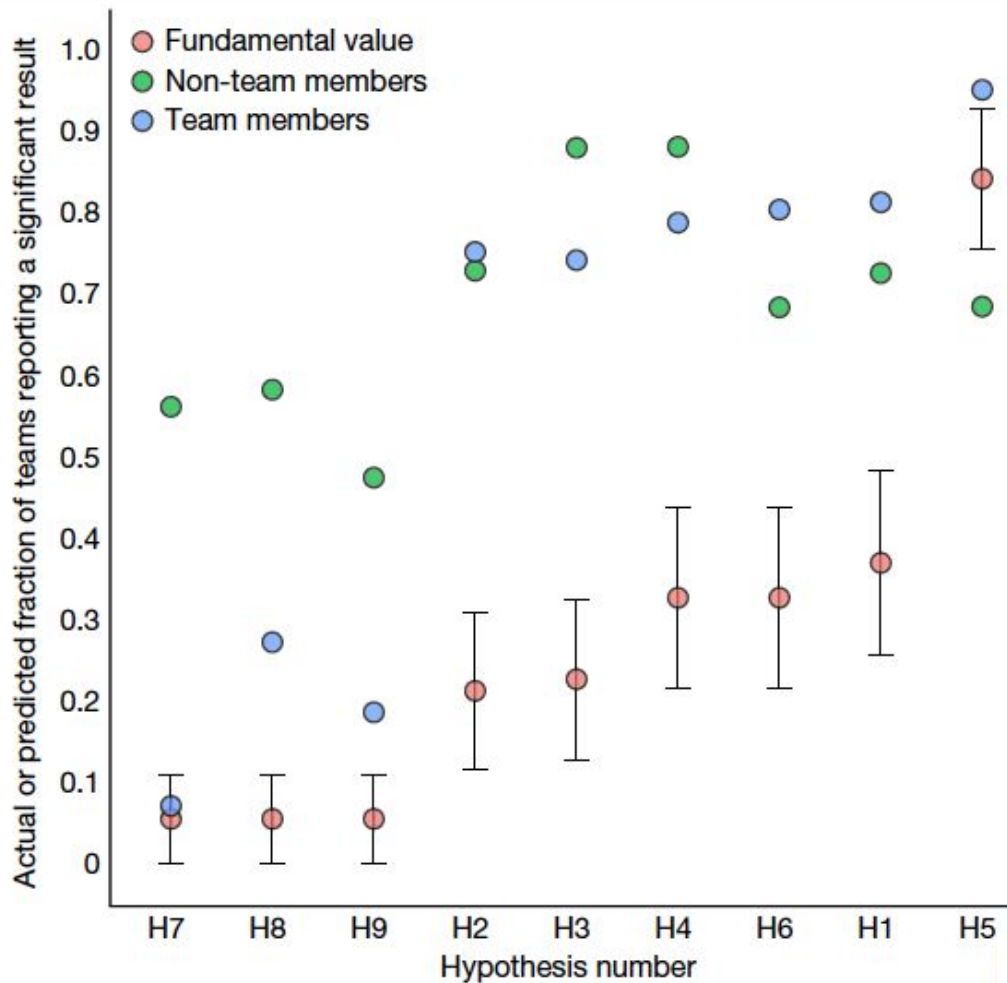
*e.g. Is there a parametric effect of gain:
Positive effect in ventromedial PFC -
for the equal indifference group?*

Many analysts project : NARPS



Many analysts project : NARPS

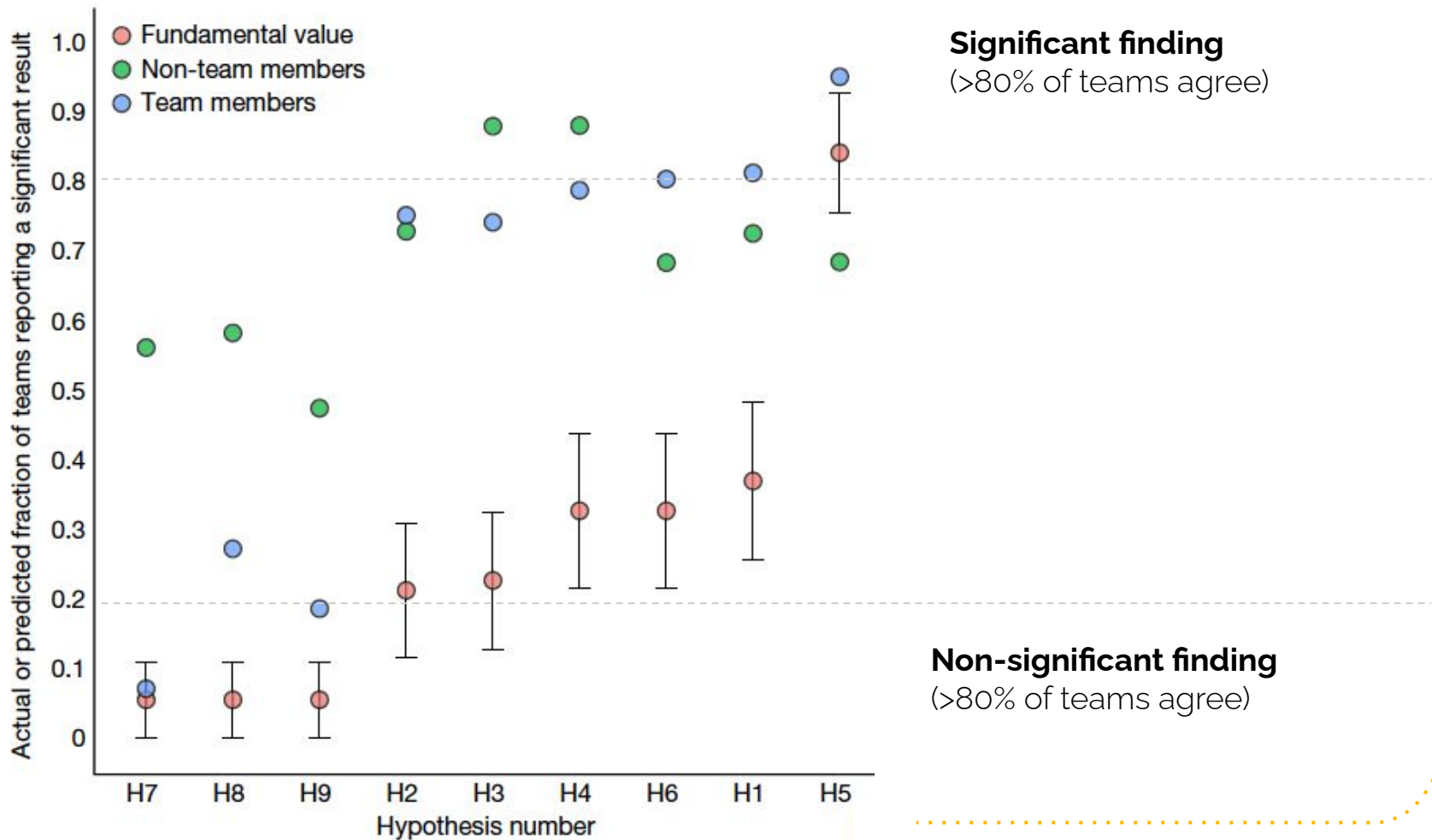
Results over the **70 post-processing pipelines**:



[Botvinik-Nezer et. al, Nature 2020]

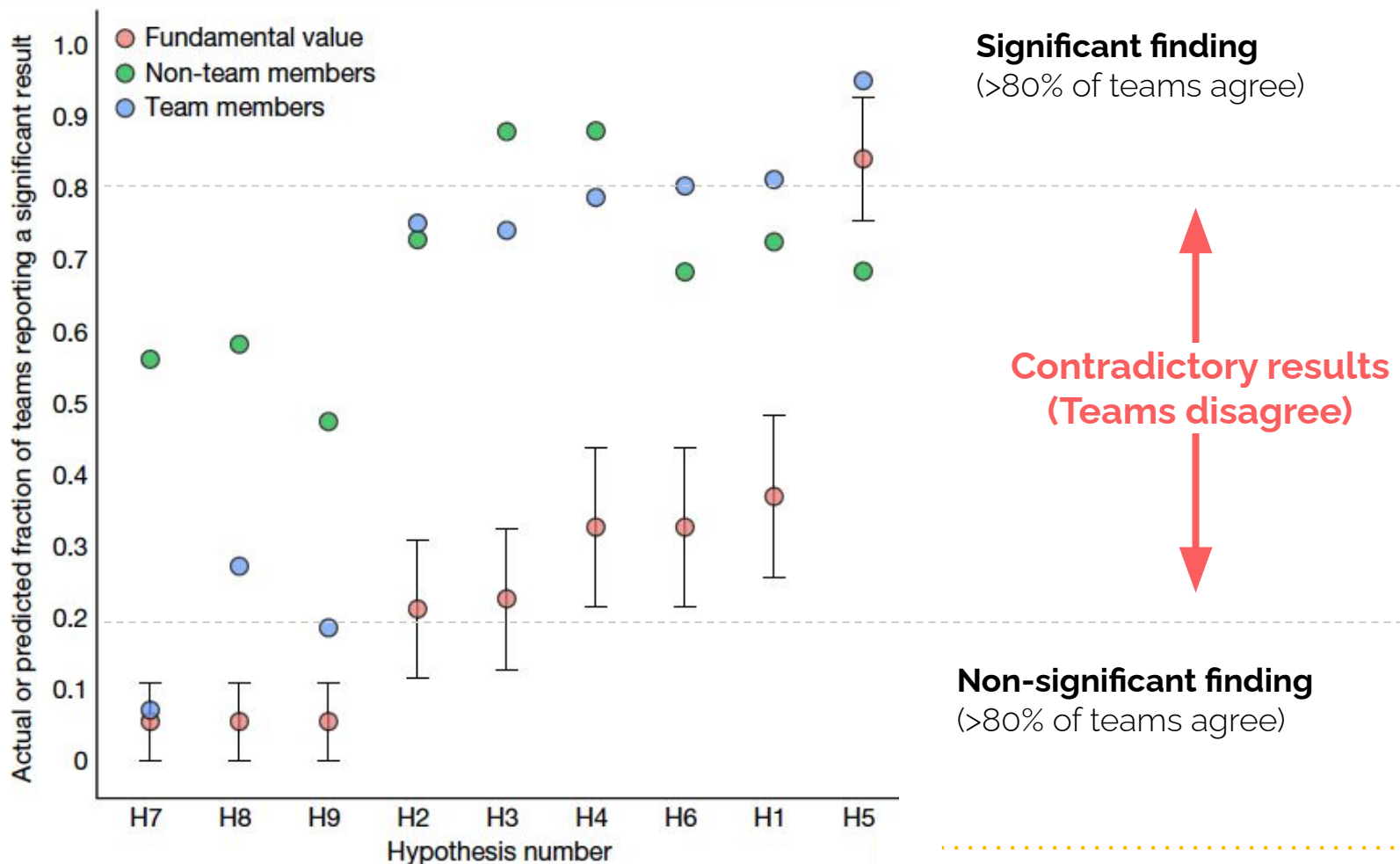
Many analysts project : NARPS

Results over the **70 post-processing pipelines**:



Many analysts project : NARPS

Results over the **70 post-processing pipelines**:



Different levels of variability in post-processing pipelines

- Different end-to-end pipelines [Carp 2013]
- Different Software Version [Gronenschild 2012]
- Different operating system [Glatard 2015]

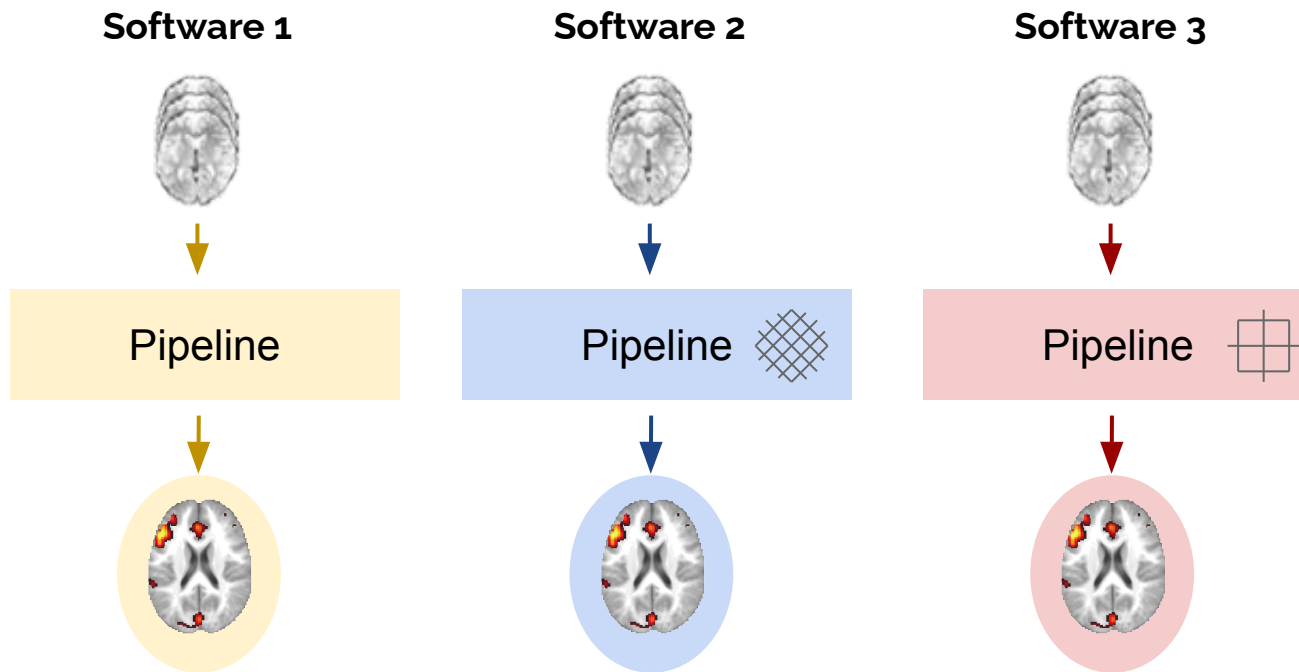
- Different **post-processing software packages** [Bowring 2019, 2021]

Variability across software

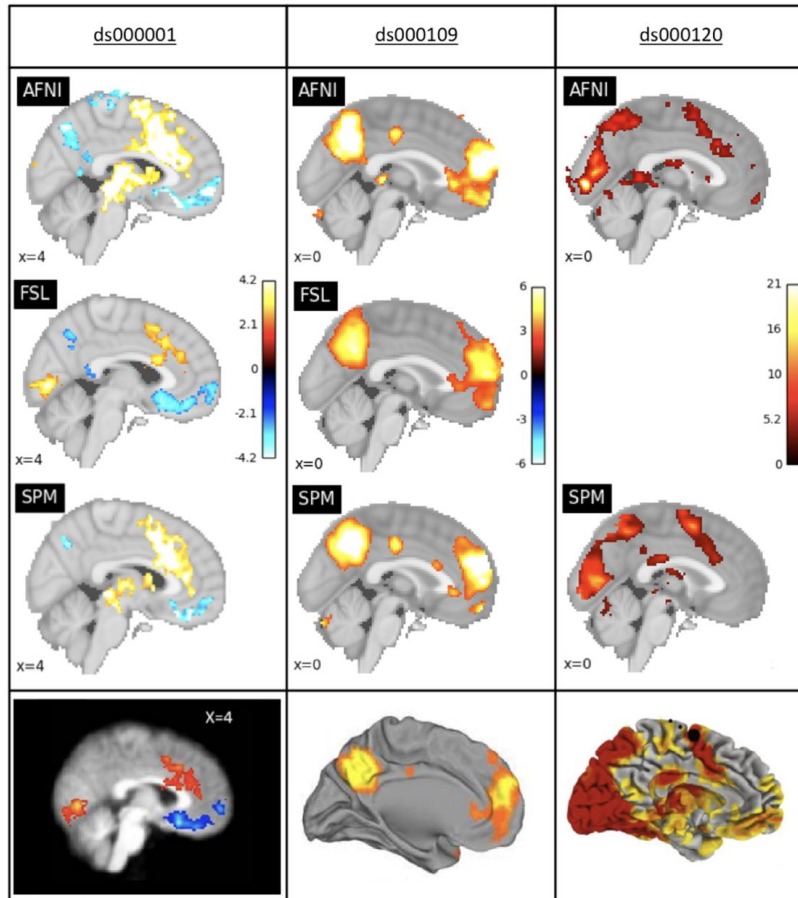
- Reproduced 3 published functional MRI studies
 - ds001, ds109, ds120 (from [Openneuro](#))
- Using 3 different software packages
 - AFNI, FSL, SPM



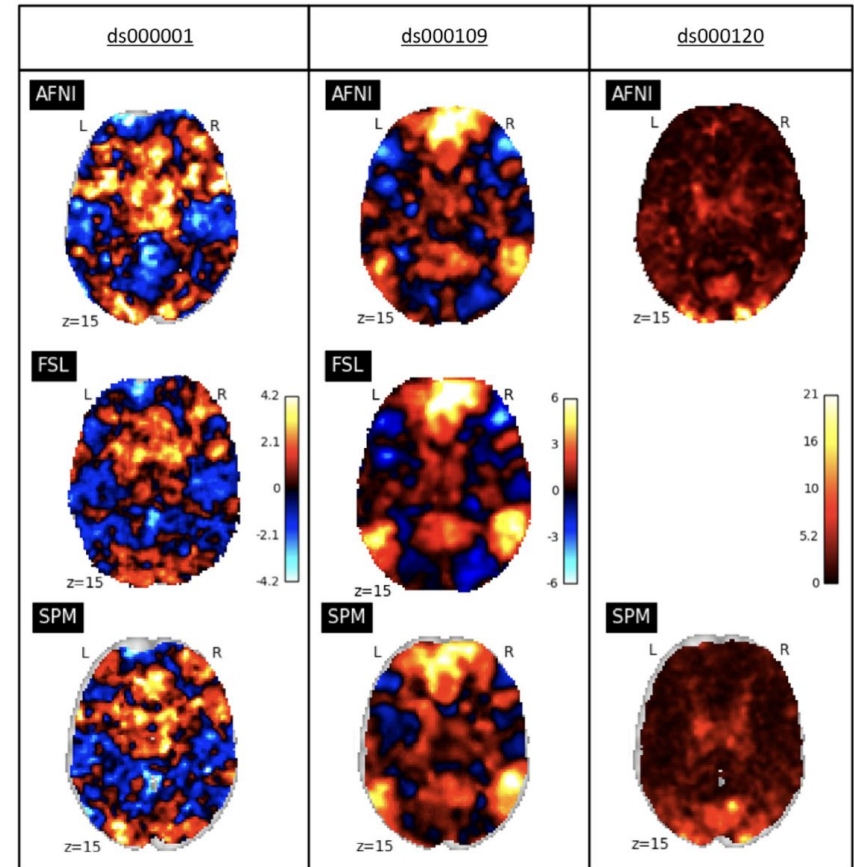
Alex Bowring Tom Nichols



Software Comparison Project



Comparison of the final results



Comparison of the statistic maps

Code overview

https://github.com/nisox-bdi/Software_Comparison/

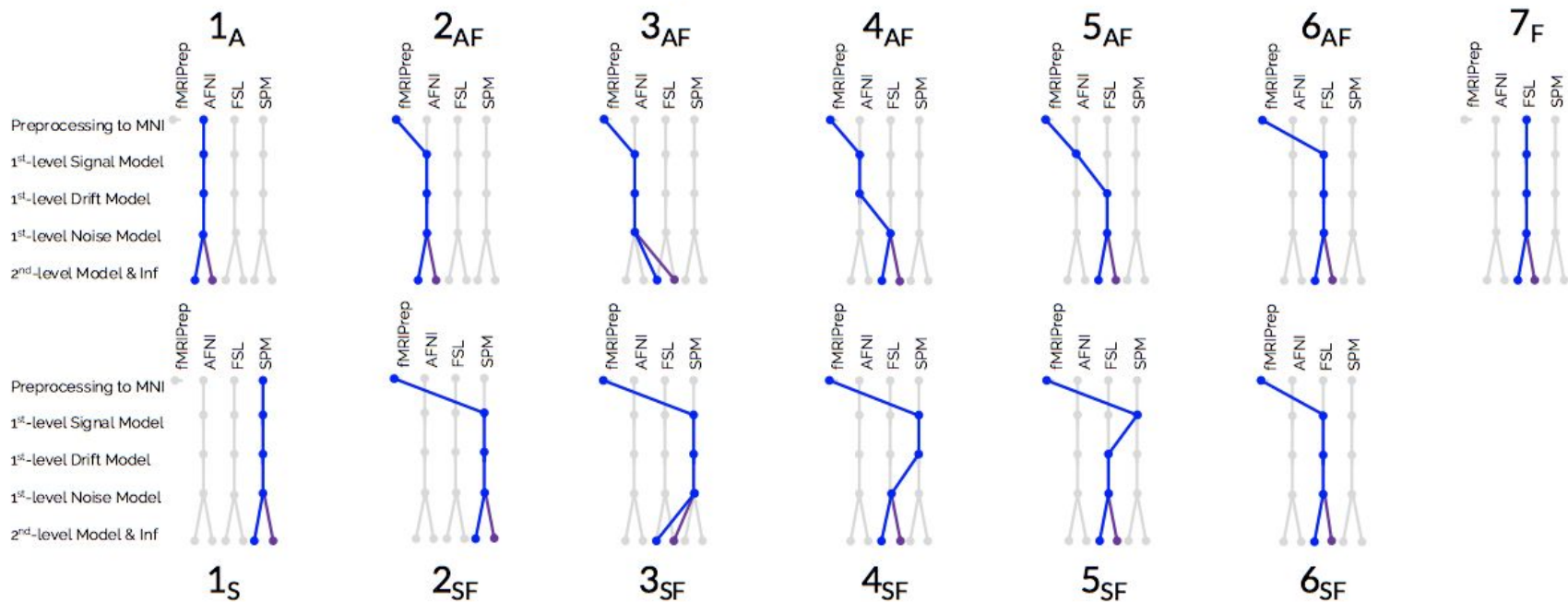
Conclusions

- Observed differences with aligned pipelines across software packages
- More pronounced with thresholded maps
- How to explain those variations

Software Comparison Project 2



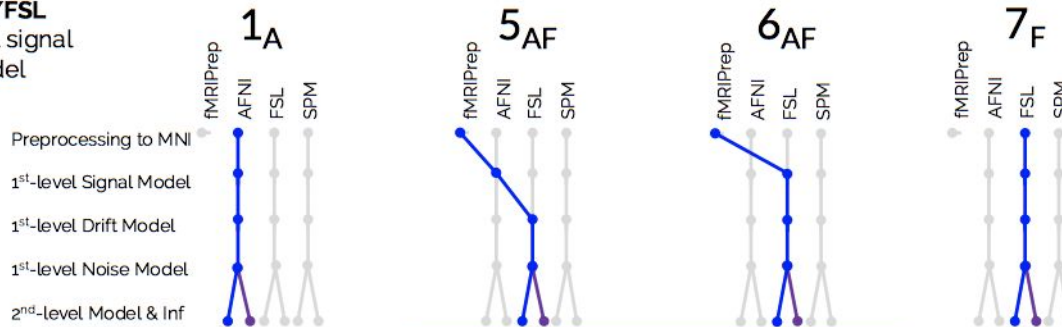
Alex Bowring Tom Nichols



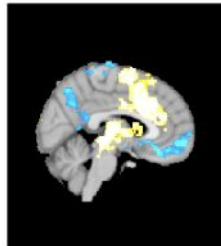
Software Comparison Project 2

ds000001

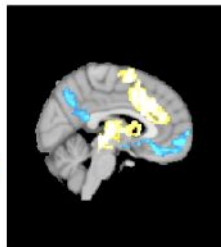
AFNI/FSL
1st-level signal model



AFNI
PARAMETRIC



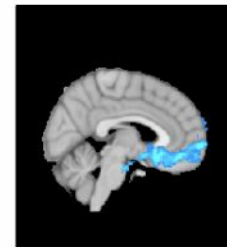
AFNI
PERMUTATION



FSL
PARAMETRIC



FSL
PERMUTATION



Code overview

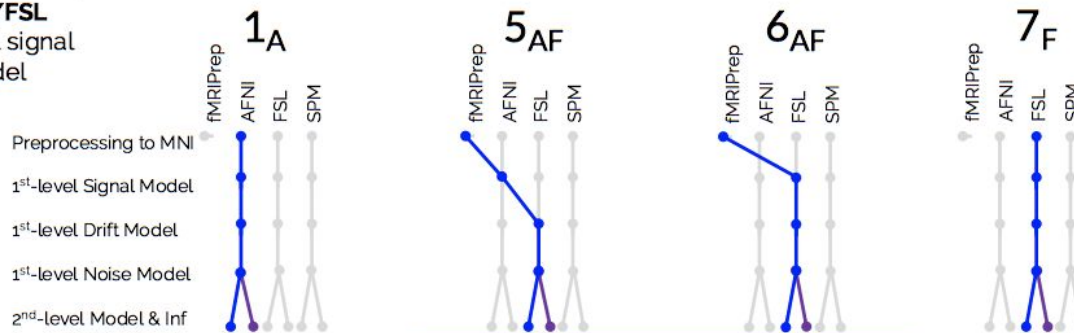
<https://github.com/NISOx-BDI/SC2>

Main sources of variations

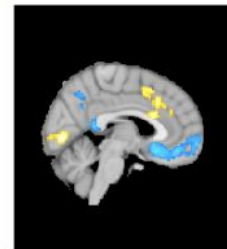
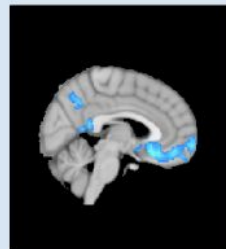
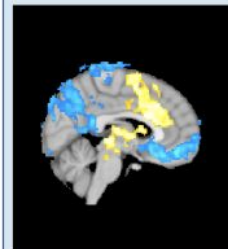
- Different across datasets
 - ds001: software's **first-level signal model**
 - ds109: preprocessing and **first-level noise model**
 - ds120: **group-level model and inference**

ds000001

AFNI/FSL
1st-level signal
model



AFNI
PARAMETRIC



FSL
PARAMETRIC

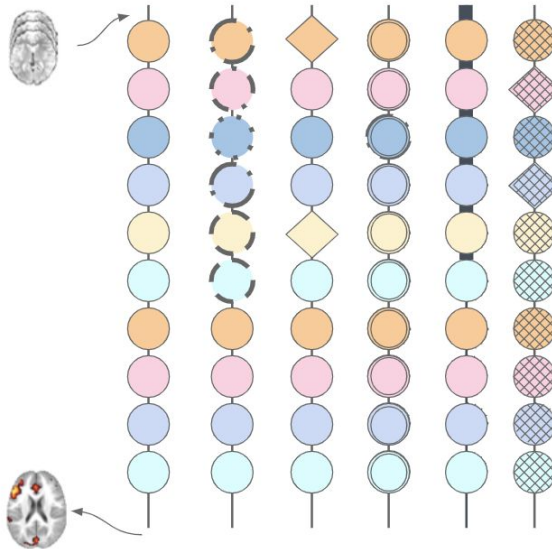
**Going further to
understand** analytical
variability

Why studying analytical variability?

- Impacts the **reliability** of existing findings
- Impacts the **reusability** of derived datasets

Explaining analytical variability...

?



Explaining analytical variability...

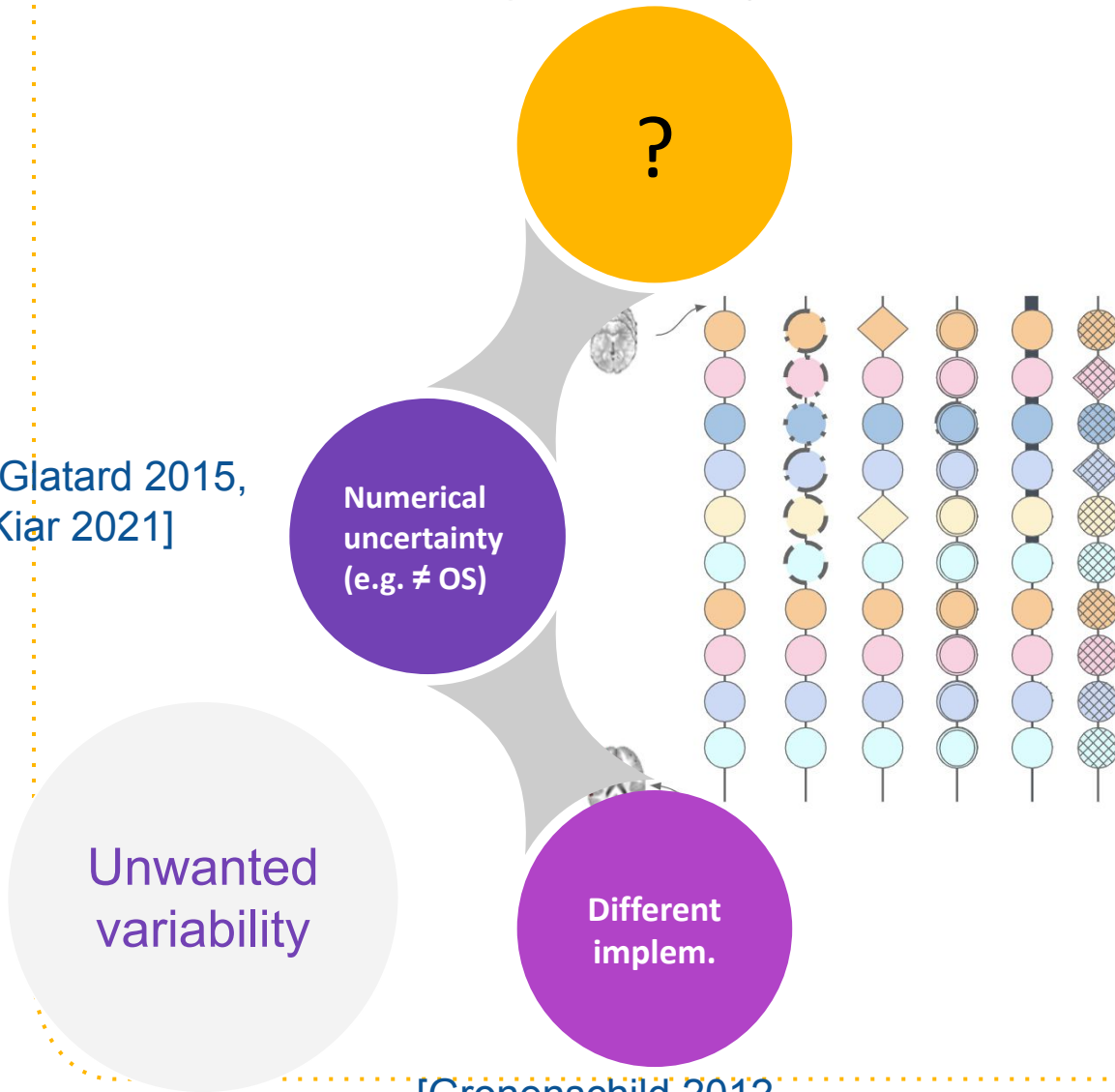
[Glatard 2015,
Kiar 2021]

Numerical
uncertainty
(e.g. \neq OS)

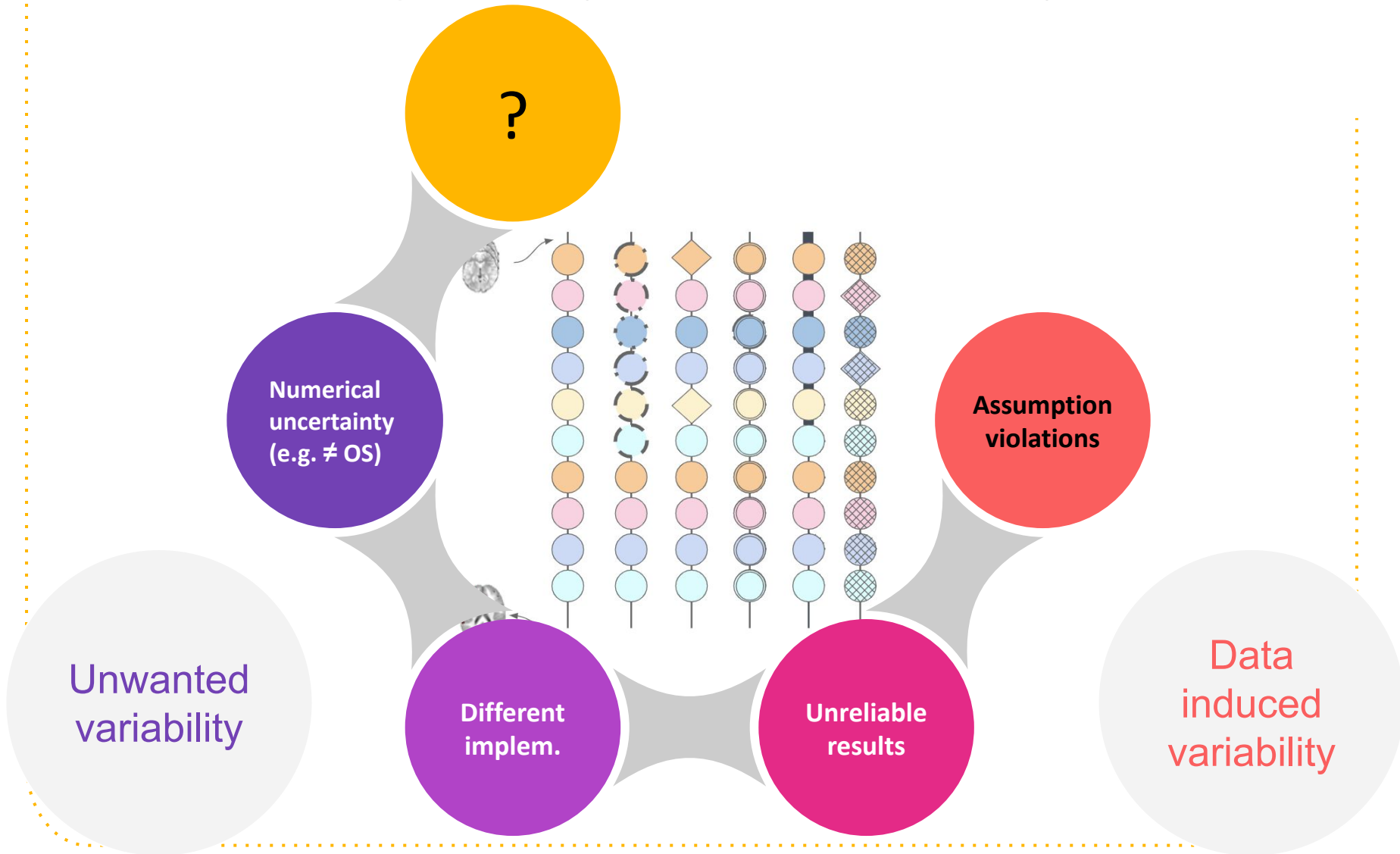
Unwanted
variability

Different
implem.

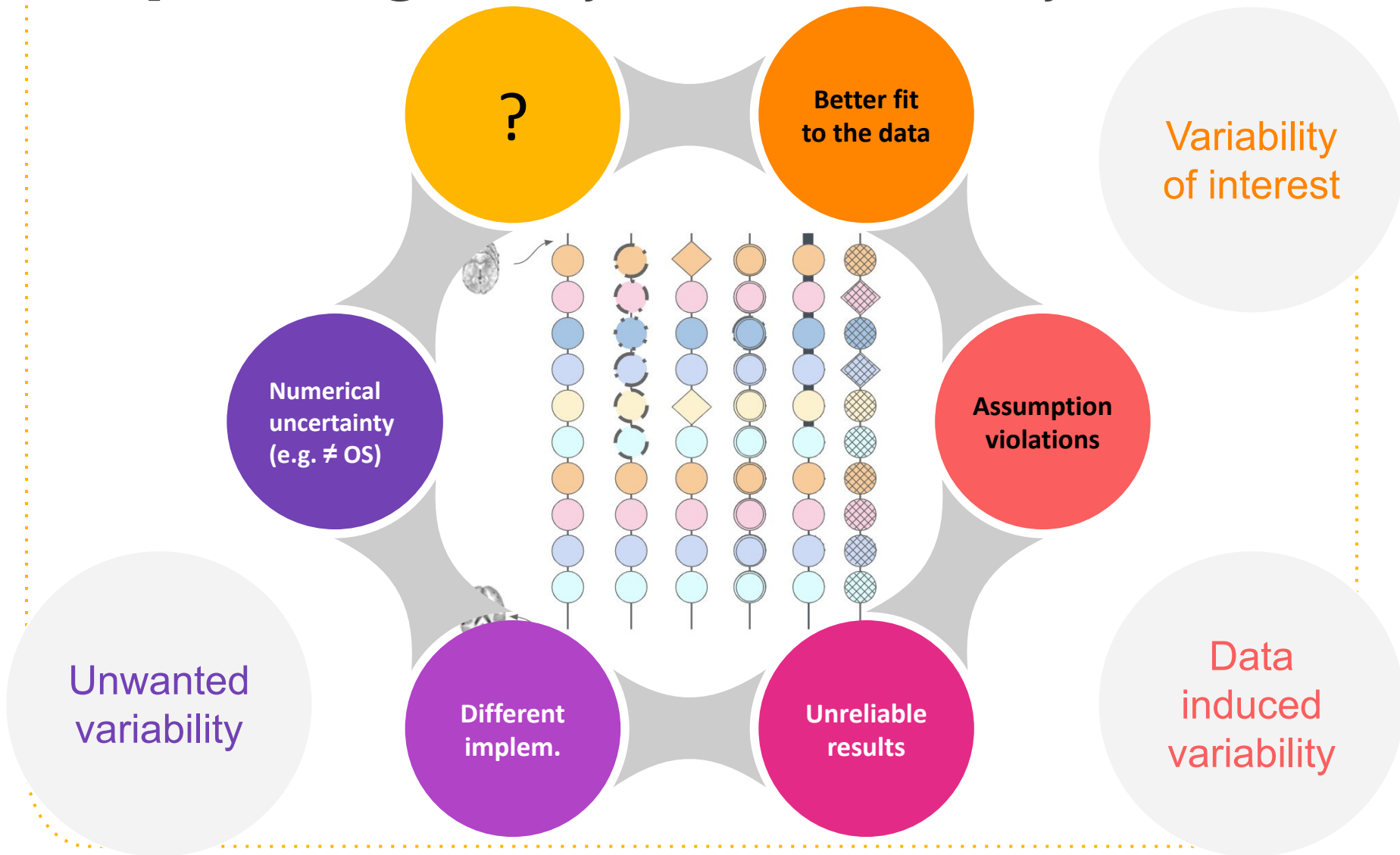
[Gronenschild 2012,
Bowring 2019]



Explaining analytical variability...



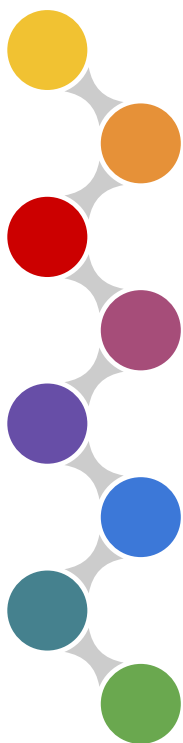
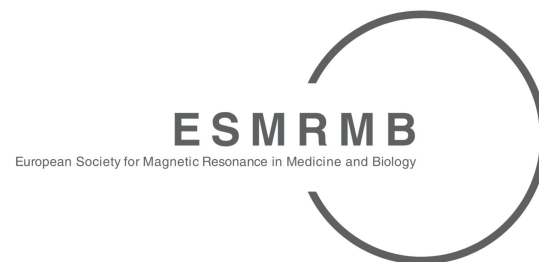
Explaining analytical variability...



On our way...

- Sharing a **variety of research artefacts** and in particular **code** along with paper
- Storing and **sharing provenance** (BIDS-Prov)
- **Benchmarking** different methods

MRI Together Global workshop on open,
reproducible, and inclusive MR research



Hands-on Juggling with data
**MRI Postprocessing
Platforms and How they
Compare**

Camille Maumet

Thank you!

 @cmaumet@fediscience.org
 [@cmaumet](https://twitter.com/cmaumet)