Best practice in data-sharing in neuroimaging... and more

Thomas Nichols
University of Warwick
http://warwick.ac.uk/tenichols
OHBM Committee on Best practice in Data Analysis & Sharing (COBIDAS) History

• Created at OHBM 2014
  – Prompted by "OHBM Council Statement on Neuroimaging Research and Data Integrity"

• Charged with
  1. Identifying best practices of data analysis and data sharing in the brain mapping community
  2. Preparing a white paper organising and describing these practices
  3. Seeking input from the OHBM community
  4. Publishing these recommendations
COBIDAS Status

• Membership

Simon Eickhoff       Alan Evans       Michael Hanke       Nikos Kriegeskorte
Michale Milham      Russel Poldrack   Jean-Baptiste Poline Erika Proal
Bertrand Thirion    David van Essen    Tonya White        BT Thomas Yeo
Thomas Nichols

• Will not prescribe practice (mostly)

• Rather focus on what to report
  – To support open and reproducible research
COBIDAS Status

• fMRI
  – Task & rest

• Divisions of fMRI Practice
  – Experimental design reporting
  – Image acquisition reporting
  – Preprocessing reporting
  – Statistical modeling
  – Results reporting
  – Data sharing
  – Replication and reproducibility

• Produce white paper
  – To be commented on, and ultimately approved by OHBM members
COBIDAS Document Form

- For each division
  - Principals of open and reproducible research
  - When feasible, recommendations for practice
  - Detailed tabular listing of what to report
# Example: Experimental Design Reporting

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Notes</th>
<th>Mandatory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of subjects</strong></td>
<td><em>Elaborate each by group if have more than one group.</em></td>
<td></td>
</tr>
<tr>
<td>Subjects approached</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Subjects consented</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Subjects refused to participate</td>
<td>Provide reasons.</td>
<td>N</td>
</tr>
<tr>
<td>Subjects excluded</td>
<td>Subjects excluded after consenting but before data acquisition; provide reasons.</td>
<td>N</td>
</tr>
<tr>
<td>Subjects participated and analyzed</td>
<td>Provide the number of subjects scanned, number excluded after acquisition, and the number included in the data analysis. If they differ, note the number of subjects in each particular analysis.</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Inclusion Criteria and Descriptive Statistics</strong></td>
<td><em>Elaborate each by group if have more than one group.</em></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Mean, standard deviation and range.</td>
<td>Y</td>
</tr>
<tr>
<td>Gender</td>
<td>Absolute counts or relative frequencies</td>
<td>Y</td>
</tr>
<tr>
<td>Race &amp; Ethnicity</td>
<td>Per guidelines of NIH or other relevant agency</td>
<td>N</td>
</tr>
<tr>
<td>SES, Education</td>
<td>Specify measurement instrument used; may be parental SES and education if study has minors.</td>
<td>N</td>
</tr>
<tr>
<td>IQ</td>
<td>Specify measurement instrument used.</td>
<td>N</td>
</tr>
<tr>
<td>Handedness</td>
<td>Absolute or relative frequencies; basis of handedness-attribution (self-report, EHI, other tests)</td>
<td>Y</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>Describe any screening criteria, including those applied to “normal” sample such as MRI exclusion criteria.</td>
<td>Y</td>
</tr>
<tr>
<td>Clinical criteria</td>
<td>Detail the area of recruitment (in- vs. outpatient setting, community hospital vs. tertiary referral center etc.) as well as whether patients were currently in treatment.</td>
<td>Y</td>
</tr>
</tbody>
</table>
COBIDAS: 100+ reporting items
**Scientific Data’s Meta Data Standard**

<table>
<thead>
<tr>
<th>Subject</th>
<th>organism</th>
<th>organism part</th>
<th>developmental stage</th>
<th>age</th>
<th>sex</th>
<th>laterality</th>
<th>Additional columns…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient1</td>
<td>Homo sapiens</td>
<td>brain</td>
<td>adult</td>
<td>43</td>
<td>male</td>
<td>Right</td>
<td>…</td>
</tr>
<tr>
<td>Patient2</td>
<td>Homo sapiens</td>
<td>brain</td>
<td>child</td>
<td>37</td>
<td>male</td>
<td>Left</td>
<td>…</td>
</tr>
<tr>
<td>Patient3</td>
<td>Homo sapiens</td>
<td>brain</td>
<td>adult</td>
<td>23</td>
<td>female</td>
<td>Right</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

**Voxel measurements**

<table>
<thead>
<tr>
<th>field of view</th>
<th>field of view unit</th>
<th>matrix size</th>
<th>matrix size unit</th>
<th>resolution value</th>
<th>resolution unit</th>
<th>repetition time</th>
<th>repetition time unit</th>
<th>echo time</th>
<th>echo time unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>256x256x192</td>
<td>millimeter</td>
<td>1x1x1</td>
<td>millimeter</td>
<td>2170 millisecond</td>
<td>4.33 millisecond</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64x64x24</td>
<td>millimeter</td>
<td>3.125x3.125x6</td>
<td>millimeter</td>
<td>4000 millisecond</td>
<td>12 millisecond</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 square centimeter</td>
<td>millimeter</td>
<td>256 x 256</td>
<td>millimeter</td>
<td>7.3 millisecond</td>
<td>3 millisecond</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ISA-Tab:** ‘Investigation’ (the project context), ‘Study’ (a unit of research) and ‘Assay’ (analytical measurement) in Tabular format
Replicability, Reproducibility, Repeatability

• COBIDAS report
  – Peng’s “Reproducibly”: Same data, same code, different researchers
  – Replication: Different data, different researchers, different methods… same conclusion

• ISO *repeatability* (ISO 3534-2:2006 3.3.5) *same-site test-retest*
  – Same “method”, “test or measuring facility”, same “operator” & “equipment” on “identical test/measurement items” “within short intervals of time”.

• ISO *reproducibility* (ISO 3534-2:2006 3.3.10) *between-site test-retest*
  – Same “method” on “identical test/measurement items” in different “test or measurement facilities”, different “operators using different equipment”

• ISO terms adopted by...
COBIAS Status

• 20\textsuperscript{th} October
  – Manuscript posted on OHBM website
  – Open for comment for 4 weeks (until 17\textsuperscript{th} Nov.)

• End of November
  – Member comments integrated
  – Final draft, approved by Council

• December
  – Finalized manuscript posted for up/down vote
  – Upon approval, submission for publication
Some Useful Tools & Projects

- NIDM
- Neurosynth & Reverse Inference
- NeuroVault
INCF’s Neuroimaging Data Model (NIDM)

- Collaborative effort to represent all aspects of neuroimaging experiments, data collection & analysis in semantic web, machine-readable form
NIDM Overview

NIDM-API

NIDM-Workflows

NIDM-Results

NIDM-CORE

Subject and experiment information

Acquisition details

Raw data

Analyses+Workflows

Derived data

Publications

Distributed, shared and queried via a common API

Databases+Webservices

Provenance - PROV-DM
Semantic Web

- **PROV-DM**: Semantic web model for provenance
- **NIDM**: PROV-DM for neuroimaging
NIDM Results

- **Semantic web**
  - “serialized” into a text file

**Diagram Notes**:
- Design of the experiment, pre-processed data...
- Statistical estimation
- Inference
- Effects, standard error and statistical maps...
- Inference
- Thresholded maps, list of peaks and clusters...

**Legend**:
- Activity
- Entity
- Agent
Common Format for all software

FEAT Report
Analysis methods

Cluster List

Cluster & Peak Definition

Threshold

Statistic map

Excursion set

SPM(T_{84})

Contrast weights

Examined contrast(s):

Cluster & peak definition

Statistic map

Excursion set

Threshold

Cluster & Peak Definitions

Cluster

Peak

Contrast weights

Excursion set

Cluster

Peak
NIDM Exporters

• SPM12 done!
  SPM 12 batch system: SPM ➔ Stats ➔ Results report
  – SPM8 extension underway

• FSL – in beta testing

• AFNI
  – In planning stage

• For more see
  http://nidm.nidash.org
Neurosynth & Reverse Inference
Reverse Inference & Brain Imaging

• Politics study from 2007
  – Voters viewed images of Democratic candidates (N=20)
  – Subset that disliked Clinton:
    • “...exhibited significant activity in the anterior cingulate cortex, an emotional center”..., activated when one “feels compelled to act in two different ways but must choose one.”

Reverse Inference & Brain Imaging

• Logic
  – Emotion conflict resolution task
    ➔ Anterior Cingulate activation
      *known from the literature*
  – Hillary Clinton
    ➔ Anterior Cingulate activation
      *observed in this experiment*
  – Ergo
    ➔ Hillary Clinton induces emotional conflict

▶ Faulty Reverse Inference
  – High $P(\text{A.C. Act.} \mid \text{Emot. Conf.})$ *doesn’t imply*
    high $P(\text{Emot. Conf.} \mid \text{A.C. Act.})$ ！！！
Reverse Inference: Correctly!

- Bayes Rule
  - Cognitive Domain $C$, Activation $A$
  - $P(C=c | A) = \frac{P(A|C=c)P(C=c)}{\Sigma_{c^*} P(A|C=c^*)P(C=c^*)}$
    summation over all cognitive domains!

- Can we find “$P(\text{Emot. Conflict} | \text{ACC Act.})$”? 
  - Need to run 100’s of experiments!
  - Or, use meta analysis!
  - But best Neuroimaging Meta Analysis databases are still limited
    - BrainMap.org has 2,757 studies (started in 1988)
    - Pubmed finds 28,694 refs “fMRI” in title/abstract
Neurosynth

A Term-based search
“Pain”

Related studies
Mechanisms of Directed
An fMRI Investigation of
Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain

 Automated coordinate extraction

P(X, Y, Z, Study)

-23 18 45 1
19 3 12 1
-40 0 -16 1
35 -41 29 2
2 18 33 2

Meta-analysis
P(Pain|Activation)

4,393 studies (in < 12 months!)

B Forward inference

Pain

Reverse inference
Working Memory?
Emotion?
Pain?
...

C Classification

Working mem. Emotion Pain

P = 78% P = 64% P = 87%

Select highest probability

“Pain”
Neurosynth Methods

• 17 Neuroscience-focused journals used

• Tagging
  – Each article ‘tagged’ with psychological terms
  – Scored as high frequency (>1/1000 words) or not

• Coordinate harvesting
  – Tables parsed for x,y,z coordinates

• Not exhaustive, but already massive
  – 4,400+ studies, 145,000+ foci
What about Anterior Cingulate?

• It’s always there!

• Finally, can do real reverse inference...
NeuroVault
NeuroVault

- **BrainMap | Neurosynth**
  - Only coordinates
  - Huge loss of information
- **Coordinate based vs. Intensity based Meta Analysis**
  - Substantial information loss

Thank you!