Multiple Imputation Toolkit for Group Level fMRI Statistics

KI Vaden[1], M Gebregziabher[2], SE Kuchinsky[1], MA Eckert[1]

[1] Department of Otolaryngology - Head and Neck Surgery, Medical University of South Carolina

[2] Division of Biostatistics and Epidemiology, Medical University of South Carolina

MUSC COLLEGE *of* MEDICINE

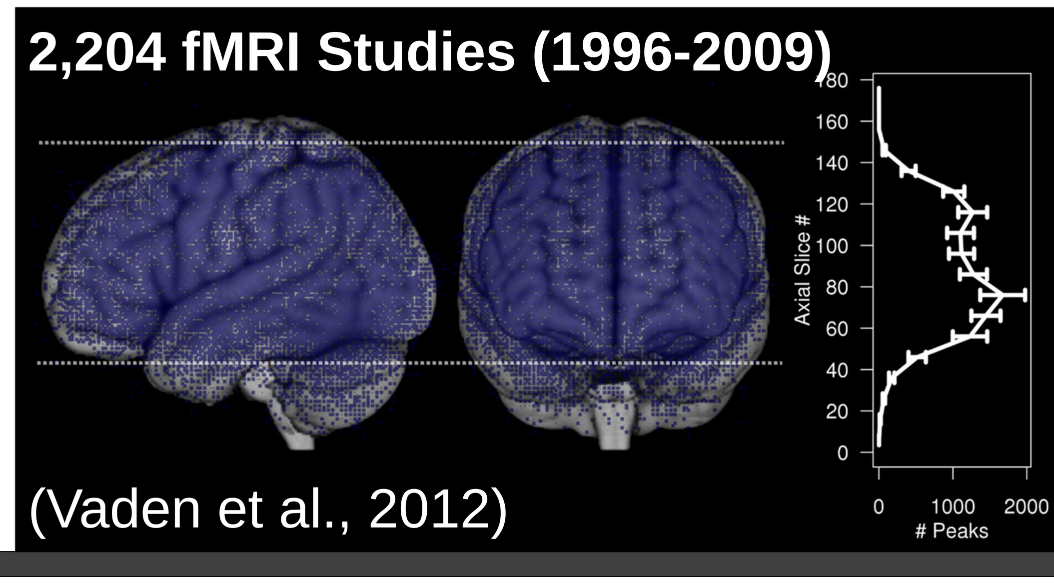
HEARING RESEARCH PROGRAM

Missingness in fMRI Datasets

Widespread missingness occurs in group fMRI analyses because tests are not performed for voxels when any subject has missing data.

Missingness typically is due to susceptibility artifact and limited brain coverage, and is a more significant problem in studies of aging and degenerative disorders because of increased anatomical variability and motion.

Multiple imputation can address missing data in fMRI datasets (Vaden et al., 2012).



(Vaden et al., 2012)

Multiple Imputation

Multiple imputation is a principled "filling in" statistical technique for inferences based on data with missing values assumed to be at random (Rubin, 1987).

- A regression model with predictors is fitted to the observed contrast values (*e.g., age, gender, motion, neighbors: mean observed value within 18 mm of each voxel with missing data*).
- *m* datasets are generated by replacing missing values according to the regression equation; pooled tests approximate a result with no missing data.

Multiple imputation reduces bias, Type I errors, and Type II errors compared to omission, available case or mean replacement analyses in voxels with predictable missingness and fewer than 33% missing cases (Vaden et al., 2012).

Toolkit: Group Level Imputation of Statistic Maps

Three command line scripts that are used to measure missingness, evaluate predictors, and perform single sample t-tests on group level fMRI datasets.

- Identifies voxels with missing data.
- Performs multiple imputation in voxels with missing data.
- Creates statistic maps with and without multiple imputation.
- Uses the multiple imputation method from Vaden et al. (2012).

Freely available from http://www.nitrc.org/projects/multimpute

Requires Matlab and SPM5/8 toolbox (www.fil.ion.ucl.ac.uk/spm), as well as [R] and MICE package (van Buuren & Groothis-Oudshoorn, 2011).

Demonstration

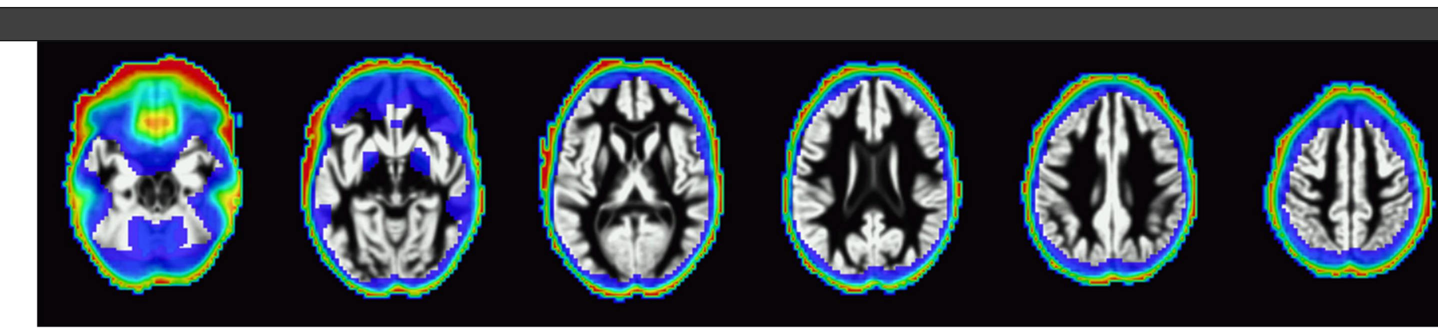
Outcome measure: contrast maps for word listening from an fMRI experiment (N = 36).

Multiple imputation was used to generate $m = 5$ versions of each contrast image voxels with fewer than 30% missing subjects. Resultant t-scores were pooled within each voxel.

1 Import contrast data

Input: list of contrast image files

Output: contrast data with neighborhood averages of each missing voxel, map: proportion of missing data in each voxel.



2 Evaluate predictors (*e.g. age, gender, motion*) on missing data voxels. Perform t-tests / multiple imputation and pooled t-tests.

Input: predictor list, contrast data / neighbor averages (step 1).

Output: sensitivity analysis results and t-scores for each voxel.

Result: Predictors were significantly related to missingness in 48% of voxels ($p < 0.0001$) and perfect predictors in 52%.

3 Export results into t-score maps.

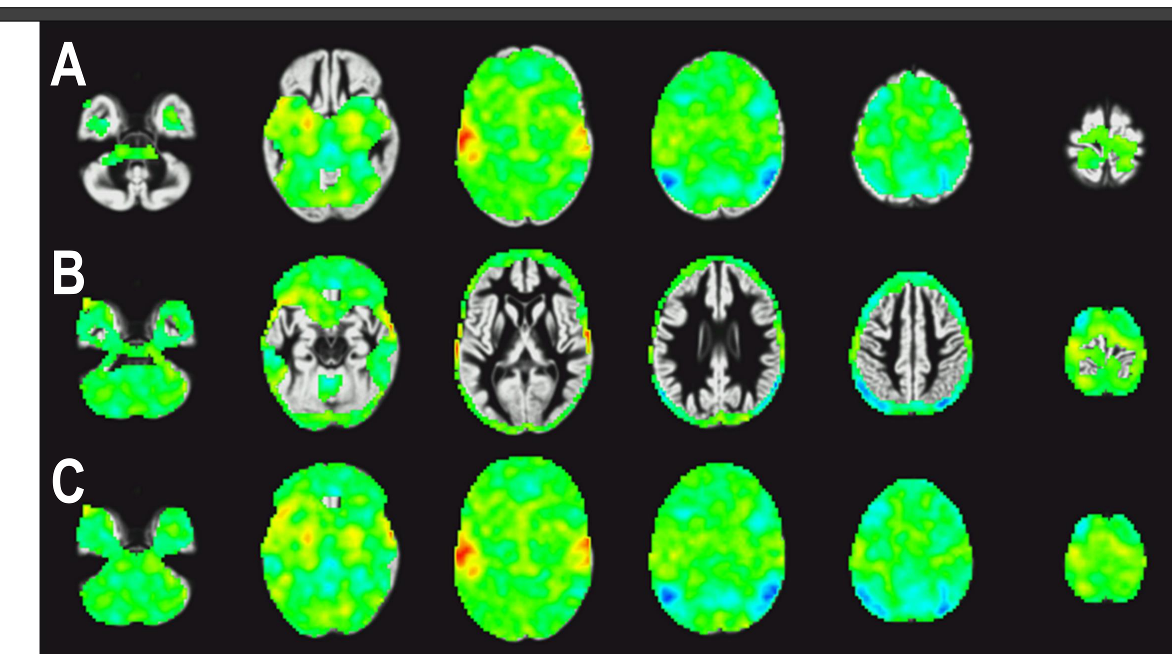
Input: t-scores in text files (step 2).

Output files: t-score maps.

A. spmT complete: 33,456 voxels

B. spmT missing: 14,927 voxels (+45%)

C. spmT combined: 48,383 voxels



4 Generate result tables for each t-score map in SPM.

T-score maps and SPM.mat are used to summarize significant voxels and clusters.

Conclusion

Group level fMRI results are negatively affected by missing data. Multiple imputation offers an alternative to omitting voxel datasets from tests, thereby reducing Type II errors (Vaden et al., 2012). The toolkit analyzes missingness and performs multiple imputation for group level statistics (version 1: single sample t-tests).

Acknowledgments: This work was supported (in part) by the NIH and National Institute on Deafness and Other Communication Disorders (P50 DC 00422) and the MUSC Center for Biomedical Imaging, in a facility constructed with support from (NIH-NCRR C06 RR14516) and the SC Clinical and Translational Research Institute (NIH-NCRR UL1 RR029882). We also thank our study participants.

Kenny Vaden
vaden@musc.edu
http://eckertlab.org

Title

Multiple Imputation Toolkit for Group Level fMRI Statistics

Authors

Kenneth I. Vaden; Mulugeta Gebregziabher; Stefanie E. Kuchinsky; Mark A. Eckert

Abstract

Group level fMRI statistics typically exclude a significant number of voxels from statistical tests because those voxels contain partial to complete missing data across subjects. Excluding data from statistic tests can result in increased Type II errors for voxels that contain significant effects and possibly increased Type I errors for some whole brain comparison methods. This missingness problem can be exaggerated in studies of normal aging and degenerative disorders, as a result of increased anatomical variability and motion artifact that can be compounded by missingness due to susceptibility artifact that is observed in most subjects. Multiple imputation is a principled "filling in" statistical technique for making inferences using datasets with values that are assumed to be missing at random. We adapted the use of multiple imputation for functional imaging data (Vaden et al., 2012) and present a new open source toolkit to allow other researchers to examine missingness in their data and perform multiple imputation of group level fMRI statistics.

Works Cited

Vaden, K.I., Gebregziabher, M., Kuchinsky, S.E., Eckert, M.A. (2012). Multiple imputation of missing fMRI data in whole brain analysis. *NeuroImage, 60(3)*, 1843-1855.

van Buuren, S., Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations. R package, version 2.13. Retrieved from http://www.stefvanbuuren.nl.

URL for the toolkit: **http://www.nitrc.org/projects/multimpute**