Tools of the Trade

Group comparisons: imaging the aging brain

Gregory R. Samanez-Larkin,[1] and Mark D'Esposito[2]
[1]Department of Psychology, Stanford University, Stanford, and [2]Henry H. Wheeler, Jr Brain Imaging Center, Helen Wills Neuroscience Institute and Department of Psychology, University of California Berkeley, Berkeley, CA, USA

With the recent growth of functional magnetic resonance imaging (fMRI), scientists across a range of disciplines are comparing neural activity between groups of interest, such as healthy controls and clinical patients, children and young adults and younger and older adults. In this edition of Tools of the Trade, we will discuss why great caution must be taken when making group comparisons in studies using fMRI. Although many methodological contributions have been made in recent years, the suggestions for overcoming common issues are too often overlooked. This review focuses primarily on neuroimaging studies of healthy aging, but many of the issues raised apply to other group designs as well.

Keywords: functional magnetic resonance imaging; group studies; aging; hemodynamics; spatial normalization; ROI

INTRODUCTION

Group comparisons are not new to psychology or the neurosciences. Since the inception of functional magnetic resonance imaging (fMRI), scientists have compared the neural activity of groups of interest. In recent years, there has been particular growth in the number of studies of the aging brain using neuroimaging (Cabeza *et al.*, 2005; Grady, 2008). Scientists may compare the brain activity of younger and older adults to explore a wide range of research interests. Group comparisons across the life span can potentially reveal the effects of age-related neural decline, effects of accumulated experience, motivational differences across adult development, or the influence of global differences in time perspective (Reuter-Lorenz, 2002; Reuter-Lorenz and Lustig, 2005; Carstensen, 2006). However, the inferences that can be drawn from such comparisons are highly limited unless care is taken at all stages of the research process from study design to data analysis to interpretation of the results.

Consider the hypothetical results in Figure 1. This figure will be referenced throughout the review to help illustrate a number of the concerns discussed. The figure displays the results from a hypothetical cross-sectional fMRI study. The results of the study reveal a significant main effect of age group such that younger adults show greater activation than older adults across several brain regions (Figure 1A). Plotting

the group-averaged 'neural signal' (either percent signal change from the event of interest within a trial or parameter estimates of a regressor fit) for the two groups confirms the significant difference (Figure 1B). Further, combining the younger and older adult samples, a simple correlation reveals that the neural signal correlates with behavior in the task (Figure 1C; i.e. the number of pictures encoded in a memory task). The region of interest (ROI) data used for the analyses were extracted within individuals using a single functional mask (Figure 1D–E) created from the results of the group comparison analysis in Figure 1A. The results of the study may seem clear: the older adults show less brain activation and perform worse than the younger adults. The authors of this particular study might speculate further that the group differences in task performance are due to the group differences in neural signal in these regions. However, as we will discuss throughout this review, there are several reasons to be skeptical of these conclusions.

Comparing age groups using fMRI is not as simple as collecting a sample of older participants and assuming that any differences between age groups are due to differences in underlying neural computations. Many neurovascular and morphological changes accompanying healthy aging can confound results. This review highlights potential problems with comparing younger and older adults using fMRI, with an emphasis on solutions. We review three areas (group differences in hemodynamics, brain morphology and variance/noise) and summarize solutions that have been proposed in the literature. Although, a number of excellent methodological papers and chapters have appeared in recent years, they are rarely cited and the suggestions for overcoming common issues are often not implemented. Our goals are to summarize this research and provide a succinct and easily accessible
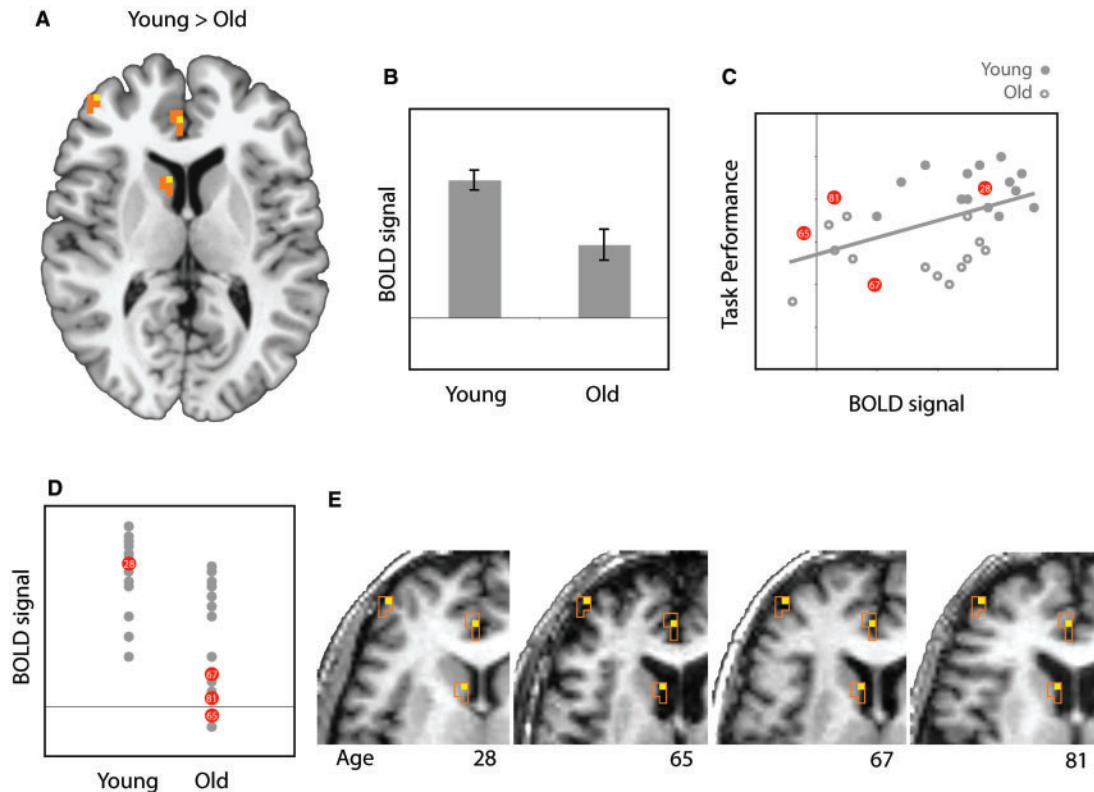
Fig. 1 This figure displays hypothetical results from a cross-sectional study examining differences in brain activity between younger and older adults. Although, the results of this hypothetical study may seem to suggest that the poorer performance of older adults is due to their showing less brain activation than the younger adults in these regions, there are several reasons detailed in the text to be skeptical of these conclusions. (**A**) A whole brain analysis reveals a main effect of age (Young > Old) in three regions: lateral prefrontal cortex, anterior cingulate and medial caudate. (**B**) An example of BOLD signal change plotted from one of these regions reveals significantly greater activation in younger adults. (**C**) A scatter plot displays a significant simple correlation between neural signal (*x*-axis) and behavior (*y*-axis) across all participants (younger and older adults combined). (**D**) The individual data points (used in the previous bar graph and the *x*-axis in the scatter plot) plotted here by age group reveal a nonnormal (bimodal) distribution of the data for the older adults. About half of the sample has BOLD signal change values near zero. (**E**) The data plotted in panels (**B**)–(**D**) were extracted using a single ROI mask shown here overlaid on a random sample of individual participants' T1 anatomy (axial; TT S = 10; MNI S = 12). The participant ages are listed below each image. It is clear in these images that this single mask does not align well with the individual participants' anatomy. The data points for each of these four sample participants have been highlighted with red dots and labeled with the participant ages in panels (**C**) and (**D**).

set of rules of thumb for conducting studies of the aging brain using functional neuroimaging.

PROBLEMS

Group differences in hemodynamics

The most investigated methodological question in the neuroimaging and aging literature is whether the coupling between neural activity and the blood-oxygen-level-dependent (BOLD) signal changes with age (D'Esposito *et al.*, 2003). Common age-related changes in vasculature can lead to age-related differences in the BOLD signal that might not be due to true differences in neural computation (for a review, see Gazzaley and D'Esposito, 2005). One primary goal of functional brain imaging is to identify regions of the brain whose activation correlates with psychological events of interest. Standard brain imaging analysis programs begin by constructing a regression model with the predicted timecourses for these psychological events. These predictors are then convolved with a standard

hemodynamic response function (HRF) to account for the lag and shape of the BOLD response measured by fMRI. If the shape of this HRF significantly deviates from the HRF of an individual—for example, if vascular rather and not neural changes result in age-related differences in the peak amplitude or shape of the HRF—the model may fit less well and become biased.

Group differences in brain morphology

A second important issue in studies comparing groups differing in age is that there is great anatomical variability between the brains of younger and older adults and within a random sample of older adults (Raz *et al.*, 2005; Raz *et al.*, 2007). Gray matter atrophy and sulcal expansion in older adults contributes to a more uneven cortical surface which can lead to distortions in automatic spatial normalization (Crinion *et al.*, 2007). Spatial normalization is an image processing step commonly utilized in brain imaging analysis packages to enable group-averaged statistical comparisons

across the whole brain. These normalization algorithms attempt to account for differences between individual participants brain anatomy by aligning each participant's anatomy to a template image. Unlike age differences in neurovascular coupling, differences in spatial normalization between younger and older adult groups has received little attention. Because all spatial normalization algorithms warp an individual brain according to a template brain, the more participants deviate from the template, the more likely the warp will lead to errors. In fact, inaccurate spatial normalization in the form of cortical overinflation in a sample of older adults has been previously documented (Buckner *et al.*, 2004).

Differences in brain morphology between groups can also impact ROI analyses. Even a highly accurate and reliable normalization algorithm is not perfect. As such, regardless of the algorithm used, ROIs cannot be reliably specified in single subjects from a mask created by a group (or between-group) activation map (Swallow *et al.*, 2003; Poldrack, 2007). This is a serious concern even for relatively homogeneous samples of younger adults (Devlin and Poldrack, 2007), and the concern is only magnified when comparing participants across the life span. Normal, age-related atrophy of the brain results in an increased likelihood of partial volume effects (sampling both gray and nongray matter or multiple neighboring regions in one voxel or region of interest) with increasing age across a range of structures. Once again, consider the hypothetical results of Figure 1. The age group differences in the activation map (Figure 1A) could simply be due to the fact that some older adults have less gray matter in those specific regional sites. In fact, whether ROI masks are created from thresholded clusters in a group comparison map (orange outlines in Figure 1E) or the peak voxels of difference (solid, yellow voxels in Figure 1E) are used to extract the signal in each participant, there is a reasonable likelihood that these masks will not sample gray matter in all participants. Due to age-related structural atrophy, that likelihood is correlated with age. As illustrated in Figure 1E, the peak voxels and the majority of the cluster masks are sampling CSF in the older adults, which could be contributing to the group differences in Figure 1A–D.

Group differences in variance/noise

In addition to structural atrophy and sulcal and ventrical expansion with age, although previous studies have found an increasing spread of activation across regions of the brain (Cabeza, 2002), clusters of functional activation are often smaller in extent within a region in older samples (D'Esposito *et al.*, 1999; Buckner *et al.*, 2000; Hesselmann *et al.*, 2001; Huettel *et al.*, 2001; Handwerker *et al.*, 2007). However, previous studies have revealed that there is similar BOLD signal amplitude between groups at peak voxels within these clusters (D'Esposito *et al.*, 1999; Aizenstein *et al.*, 2004). One of the reasons for these smaller activation extents around a peak voxel might be that older adults are more likely to

have noisy or deactivated voxels within an ROI (D'Esposito *et al.*, 1999; Huettel *et al.*, 2001; Aizenstein *et al.*, 2004). These group differences in noise can influence or invalidate the results of most common statistical tests.

SOLUTIONS

Fortunately, despite the observed differences in hemodynamics, morphology and noise, comparisons between younger and older adults are possible. Several solutions have been proposed in the literature to address these potential problems.

Controlling for differences in hemodynamics

To address the potential problems that can arise from group differences in hemodynamics, at least three solutions have been proposed: (i) local measurement of individual participant HRFs; (ii) global control of HRFs using hypercapnia; and (iii) improvements in experimental design.

One solution to control for potential group differences in the HRF is for participants, while undergoing fMRI, to perform both the tasks of interest and a task to derive an HRF. The data from this latter task (commonly a short, simple visual or motor task) can then be used to estimate the individual HRFs for each participant. This individualized HRF estimate can then be used in the regression model to convolve regressors. However, it is important to note that there is considerable variability not only between individuals but also across regions of the brain within individuals (Aguirre *et al.*, 1998; Handwerker *et al.*, 2004). Further, typically a visual or motor task is used to estimate the HRF but the brain regions of interest are not primary cortex but association cortex. Nevertheless, convolving model regressors with an individualized HRF derived from motor cortex will produce a better estimate of other areas of cortex within an individual than a canonical HRF (Handwerker *et al.*, 2004).

A second alternative is to use breathholding (which induces hypercapnia) to produce global changes in the BOLD signal (Cohen *et al.*, 2004; Handwerker *et al.*, 2007). Both hypercapnia and task-related BOLD signal should be influenced by the same vascular differences. This strategy would allow researchers to normalize the task-related signal change with the hypercapnia-induced signal change from the same voxels to get closer to examining true group differences in neural activity (Handwerker *et al.*, 2007). With a short additional scanning run, researchers can obtain this voxelwise global normalization factor. Whether using a visuomotor task to derive an HRF or a breathholding task to induce hypercapnia to normalize the BOLD signal within individual participants, it would be a good practice to always display the signal (timecourses) for each group for the task of interest, even if this is relegated to a supplementary figure.

A third solution to overcome the issues related to hemodynamics is improvement in experimental design that allows for investigation of interaction effects (Buckner *et al.*, 2000;

Gazzaley and D'Esposito, 2005; Rugg and Morcom, 2005). Unlike main effects of age, interaction effects with age within a region are highly unlikely to be due to group differences in neurovascular coupling, because age-related differences in neurovasculature should influence all conditions equally. With an interaction design, a study may reveal similar signal across age in one condition, but different (diminished or increased) signal in older adults in another condition (Mather *et al.*, 2004; Gazzaley *et al.*, 2005). For research questions in which either it is difficult to design an ideal control condition or the primary interest is in a main effect, a parametric design should be used (Gazzaley *et al.*, 2005). An ideal design is to have parametric manipulations nested within conditions so that effects of age within one condition can be safely explored (Samanez-Larkin *et al.*, 2007). The BOLD signal in one specific condition or trial type should never be directly compared between groups. Instead, the size of a within-group condition or parametric effect should be compared between groups.

This issue of assessing between-group differences in within-group effects is also relevant to individual difference analyses. As discussed above, older adults may have reduced BOLD signal due to vascular differences or an increased likelihood of sampling nongray matter. They may also perform worse in a task for unrelated reasons. In the example in Figure 1C, the simple correlation between BOLD signal and task behavior is significant, but not the partial correlation controlling for age. In fact, the correlation is close to zero within each age group. For example, there may be a main effect of age on the number of pictures encoded in a memory task, the response latency in an attentional interference task, or overall reaction time. There may also be a main effect of age on global BOLD amplitude. However, if the effects are not assessed within the two groups or the individual difference analysis across all participants does not partial out the group difference (i.e. age), illusory relationships between brain activity and behavior may emerge. Correlations between behavioral measures and brain activation must control for age to be meaningfully interpreted (Samanez-Larkin *et al.*, 2008). Likewise, correlations between age and brain activation should control for behavioral performance (Samanez-Larkin *et al.*, 2007).

An important caveat related to the effectiveness of interaction designs in assessing true group differences in neural activity is that this solution relies on the assumption that the vascular changes are regionally limited and the linearity of neurovascular coupling is preserved into old age. Interaction effects alone are not sufficient for ruling out artifactual differences between groups. Consider a case in which there is double the BOLD signal amplitude in younger than older adults. If there is an additional main effect of task condition, an artifactual age-related interaction might emerge as well. In summary, we recommend the use of both interaction designs and HRF normalization (using a hypercapnic control or a visuomotor localizer as described above). Improvements

in experimental design and controls should lead to stronger inferences that can be drawn not only from functional neuroimaging studies of aging in particular[1] but also functional neuroimaging studies in general.

Controlling for differences in brain morphology

To address the potential problems that can arise from group differences in brain structure, at least two solutions have been proposed: (i) improvements in spatial normalization and (ii) within-individual adjustments in ROI analyses.

Software packages differ in the accuracy of their spatial correction (Crivello *et al.*, 2002; Robbins *et al.*, 2004; Ardekani *et al.*, 2005). Many of the brain imaging analysis software programs will continue to evolve and it is up to responsible researchers to compare the normalization results using a few different methods in order to find the most optimal warp given their specific population. Researchers should routinely examine the normalization results and qualitatively report these results. At minimum, we recommend a visual inspection of the normalization results of each individual participant to ensure accurate alignment, especially in structures of interest, between individuals and the template brain. If one normalization method is unsatisfactory, it may be necessary to try another. With the introduction of a universal file format by the Neuroimaging Informatics Technology Initiative (NIFTI; http://nifti.nimh. nih.gov/), the same data files can easily be used in different brain imaging analysis software programs. Researchers can normalize in one program and run all of the other analysis and visualization routines in another program. One important caveat is that software programs differ in whether they use the Montreal Neurological Institute (MNI) or Talairach (TT) coordinate system as the default normalized space. In addition to the standard algorithms used in common brain imaging analysis packages, a number of new techniques are continually being developed (Suckling *et al.*, 2006; Joshi *et al.*, 2007; Postelnicu *et al.*, 2007) but are not yet integrated into the processing pipelines of standard packages. In the future, these different approaches will be more easily accessible so that all researchers can more simply take advantage of the features of many programs without having to apply additional warping when translating from one program to another. Such an effort is underway in the Neuroimaging in Python (NiPy; http://neuroimaging.scipy.org/) project (Millman and Brett, 2007).

In addition to identifying optimal normalization algorithms, one study has suggested that researchers could improve the accuracy of results by using an appropriate, population-specific template brain as the target for normalization (Buckner *et al.*, 2004). This study contains instructions in the appendix for creating your own population-specific template (Buckner *et al.*, 2004). In addition, the Open Access Series of Imaging Studies (OASIS) contains a

---

[1] For a more thorough discussion of experimental design in studies of human aging, see Rugg and Morcom (2005).

growing number of sample anatomical data files spanning adult age as well as anatomy from both healthy older adults and older adults clinically diagnosed with dementia (Marcus et al., 2007). As new normalization algorithms are developed, software developers could take advantage of this database to optimize the accuracy of their programs by testing the performance of the spatial normalization algorithms on a population varying in age and clinical status (Mega et al., 2005; Suckling et al., 2006).

The goal of spatial normalization is to equate whole brain size and align key structures so that when averaged maps are created across a group of subjects, there can be reasonable confidence in the localization of effects (Ashburner and Friston, 1999). Activation maps should be the starting, not ending point for group analyses. These maps should be used only to identify key regions of interest to follow up with more careful visualization of activation timecourses and ROI analyses.

For ROI analyses, manual adjustments need to be made within individuals. There are at least two options for correcting ROIs within individuals. If a single mask is created from a functional group map or group difference map, this map should be overlaid on each individual participant and if partial volume effects occur, the mask should be nudged so that only gray matter in the brain structure of interest is sampled (for an example, see the supplementary methods of Samanez-Larkin et al., 2007). If the ROIs are anatomically defined, they should be defined on individual participants. In summary, one solution for appropriate ROI analyses is to first spatially normalize (to approximately equate the size of ROIs relative to whole brain across participants), and then manually define or adjust ROI masks for each individual's anatomy (Poldrack, 2007).

Controlling for differences in noise/variance

To address the potential problems that can arise from group differences in noise/variance, at least two solutions have been proposed: (i) censoring outliers and (ii) the use of appropriate statistical tests.

It has been previously suggested that increasing spatial smoothing will help reduce both spatial normalization errors and noise in cross-sectional functional neuroimaging studies. We do not recommend this strategy for two reasons. First, increasing the smoothing kernel will only exacerbate the problem of increased noise by including more outlier voxels (i.e. voxels whose values lie standard deviations away from the surrounding voxels). Second, a general issue with large smoothing kernels is partial volume effects. Increasing the smoothing kernel may help reduce noise but will lead to partial voluming of neighboring cortex and CSF in cortical regions of interest due to the tortuosity of the gyri and sulci and partial voluming of neighboring ventricles or structures in subcortical regions of interest. Importantly, many subcortical regions of interest are small relative to standard voxel sizes, so a large kernel will blur over these

regions entirely. For these reasons, we recommend small spatial smoothing kernels. We cannot recommend a specific smoothing kernel because optimal smoothing is dependent on voxel size and contrast to noise ratio (Weibull et al., 2008).

Instead of increasing the spatial smoothing, the problem of increased noise in an older adult sample can be addressed by excluding voxels that can be considered outliers within each ROI within participants. A recent study used a probabilistic atlas to define anatomical ROIs in individual participants, masked these ROIs for gray matter only, and then excluded outlier voxels within each region (Aizenstein et al., 2006). These trimmed regions will likely include fewer voxels in the older adults, which can lead to noisier or less reliable estimates in this group. If outlier voxels are censored within participants, the statistical threshold for significance should be set within-individual by taking into account the number of voxels in the ROI (Handwerker et al., 2007). Due to group differences not only in outliers but also structural atrophy, it is best to use small or censored volumes to maximize comparisons of the peak signal and minimize noise across age groups.

For reasons described above (differences in neurovascular coupling, structural and physiological variability, outlier voxels) and others, neuroimaging data from an older group of participants may be noisier than from a younger group. This difference can lead to unequal variance between age groups, violating the core assumptions of many common statistical analyses. It is standard practice in behavioral research to assess the equivalence of variance between groups. Unfortunately, variance equivalence tests are rarely reported in the functional neuroimaging literature. Because variance may increase with age, group comparisons should include at least as many, if not more, older than younger adults. Importantly, nearly all of the problems raised in this review are exaggerated when studies are underpowered. Even with sufficient and equal sample sizes, statistical tests of variance between groups should be reported. If variance is unequal, there are at least two options for addressing this problem. Investigators can either use non-parametric tests or take advantage of multilevel modeling techniques currently implemented in FSL (Woolrich et al., 2004; http://www.fmri.ox.ac.uk/fsl/) and under development in the Multi-level Mediation/Moderation (M3) toolbox (Davidson et al., 2008; http://www.columbia.edu/cu/psychology/tor/).

Now consider the hypothetical results displayed in Figure 2. This figure displays results from a more carefully designed and reliable cross-sectional fMRI study. The results of this hypothetical study reveal a significant condition by age group interaction across several brain regions (Figure 2A) such that younger and older adults do not differ in condition R, but younger adults show greater activation than older adults in condition S. Plotting the group-averaged BOLD signal for the two groups clearly confirms the significant and nonsignificant differences (Figure 2B).
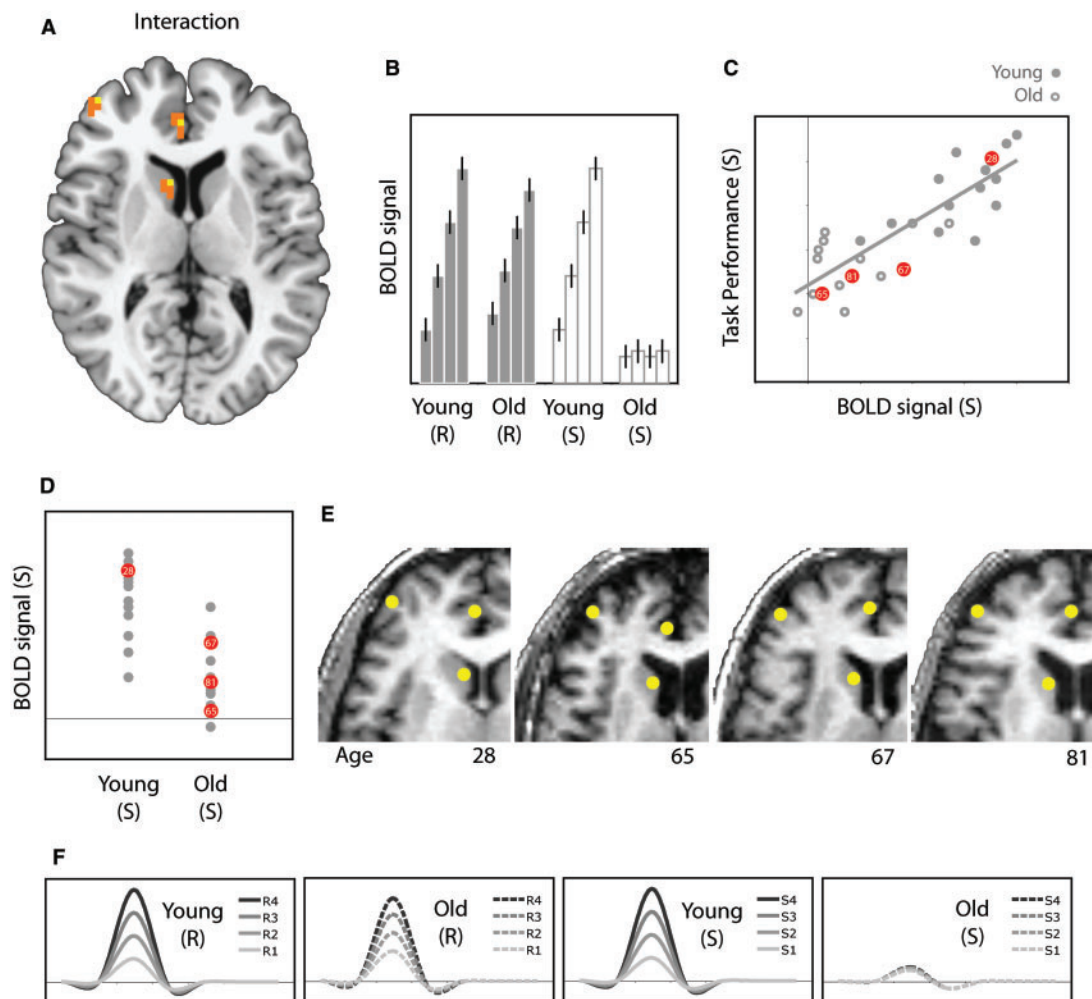
Fig. 2 This figure displays hypothetical results from a more carefully designed and analyzed cross-sectional study. (**A**) A whole brain analysis reveals an interaction effect with age (Age × Condition × Level) in the lateral prefrontal cortex, anterior cingulate and medial caudate. (**B**) An example of BOLD signal change plotted from one of these regions reveals similar activation in younger and older adults in condition R and significantly different activation between age groups in condition S. (**C**) A scatter plot displays a significant partial correlation (controlling for age) between neural signal (a difference score between levels in condition S) and behavior (a difference score in task performance between levels in condition S). (**D**) The individual data points (the difference scores in condition S used in the x-axis in the scatter plot) plotted here by age group reveal approximately normal distributions of the data in both younger and older adults groups. (**E**) In this hypothetical study, ROI masks were individually defined on single participants. These individualized masks are overlaid on a sample of participants' T1 anatomy (axial; TT S = 10; MNI S = 12). The individual participant ages are listed below each image. It is clear from the images that these individual masks provide much better spatial accuracy than the single mask used in Figure 1. The data points for each of these four sample participants have been highlighted with red dots and labeled with the participant ages in panels (**C**) and (**D**). (**F**) A sample set of timecourses of activation from the ROI analysis visually confirm the age group by condition interaction. It is clear from these plots that the age differences are not due to group differences in the general shape, height or latency of the BOLD signal in this region of the brain.

Further, the neural signal difference in condition S correlates with behavior in the task controlling for age group (Figure 2C). Plotting the individual participants activation confirms that the effects are not driven by outlier subjects (Figure 2D). The ROIs were manually adjusted on the anatomical images of individual participants to include gray matter only (Figure 2E). The authors include timecourses of activation to demonstrate that the age differences in condition S are not due to significant abnormalities in the shape of the HRF between age groups (Figure 2F). The results of this study are not only much more convincing but also lead to stronger inferences that can be drawn from the data. Thus, with the proper care, group comparison studies using

functional imaging have the potential to make lasting theoretical contributions to the psychology and biology of aging.

SUMMARY

We hope this review serves as a helpful reference for aging researchers as well as journal editors and reviewers of aging studies. Peer review of aging studies should carefully consider the issues we have addressed and require that researchers address them in their studies. Thus, we encourage current and future researchers to consistently follow these simple guidelines and report how they addressed between-group differences. Nearly all scientific journals

Table 1 Summary of recommendations

Study design	Data processing	Analyses
Use interaction or parametric interaction designs	Compare and select the best normalization algorithms	Assess effects within groups; control for group differences in performance
Include sufficient sample sizes in each group (equal numbers or more in older adult group)	Use appropriate reference anatomical templates	Test for equal variance between groups
Include appropriate controls for group differences in HRFs (functional or hypercapnia)	Specify or adjust regions of interest on individual participants regardless of age	Use appropriate statistical models (i.e. nonparametric, multilevel, etc.)

now allow appendices to be published alongside an article. If space is limited in the main text of an article, these essential methodological details should be included in a supplement or appendix.

Conducting an appropriate and interpretable group comparison study using fMRI requires a high level of attention to detail. We have addressed three core group differences that can bias results and have suggested many solutions for improving group comparisons. This set of minimum standards summarized in Table 1 should include: (i) appropriate controls for group differences in HRFs; (ii) interaction or parametric interaction designs; (iii) assessment of effects within groups; (iv) controlled behavioral performance or the inclusion of performance measures as covariates; (v) comparison and selection of the best normalization algorithms; (vi) the use of appropriate reference anatomical templates; (vii) the specification or adjustment of regions of interest on individual participants regardless of age; (viii) the assessment of unequal variance between groups; (ix) sufficient sample sizes in each group; and (x) the use of nonparametric or multilevel statistical models where appropriate. Although, this review has focused primarily on studies of human aging, these methodological suggestions are not specific to comparisons of younger and older adults. Although, the group differences (in brain morphology or noise) between younger and older adults may be more pronounced than in other group comparison studies, all of the issues raised and many of the recommended solutions apply to studies comparing children and young adults or clinical patient groups and healthy age-matched controls.

REFERENCES

Aguirre, G.K., Zarahn, E., D'esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8(4), 360–9.

Aizenstein, H.J., Butters, M.A., Clark, K.A., et al. (2006). Prefrontal and striatal activation in elderly subjects during concurrent implicit and explicit sequence learning. *Neurobiology of Aging*, 27(5), 741–51.

Aizenstein, H.J., Clark, K.A., Butters, M.A., et al. (2004). The BOLD hemodynamic response in healthy aging. *Journal of Cognitive Neuroscience*, 16(5), 786–93.

Ardekani, B.A., Guckemus, S., Bachman, A., Hoptman, M.J., Wojtaszek, M., Nierenberg, J. (2005). Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. *Journal of Neuroscience Methods*, 142(1), 67–76.

Ashburner, J., Friston, K.J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4), 254–66.

Buckner, R.L., Head, D., Parker, J., et al. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2), 724–38.

Buckner, R.L., Snyder, A.Z., Sanders, A.L., Raichle, M.E., Morris, J.C. (2000). Functional brain imaging of young, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 12(Suppl. 2), 24–34.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging*, 17(1), 85–100.

Cabeza, R., Nyberg, L., Park, D. (2005). *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*. New York, NY, US: Oxford University Press.

Carstensen, L.L. (2006). The influence of a sense of time on human development. *Science*, 312(5782), 1913–5.

Cohen, E.R., Rostrup, E., Sidaros, K., et al. (2004). Hypercapnic normalization of BOLD fMRI: comparison across field strengths and pulse sequences. *NeuroImage*, 23(2), 613–24.

Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., Friston, K. (2007). Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. *NeuroImage*, 37(3), 866–75.

Crivello, F., Schormann, T., Tzourio-Mazoyer, N., Roland, P.E., Zilles, K., Mazoyer, B.M. (2002). Comparison of spatial normalization procedures and their impact on functional maps. *Human Brain Mapping*, 16(4), 228–50.

Davidson, M., Atlas, L., Lindquist, M., Bolger, N., Wager, T. (2008). *The M3 Toolbox: the Multi-level Mediation/Moderation Framework for Connectivity Analyses in fMRI Data*. Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

D'Esposito, M., Deouell, L.Y., Gazzaley, A. (2003). Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nature Reviews Neuroscience*, 4(11), 863–72.

D'Esposito, M., Zarahn, E., Aguirre, G.K., Rypma, B. (1999). The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *NeuroImage*, 10, 6–14.

Devlin, J.T., Poldrack, R.A. (2007). In praise of tedious anatomy. *NeuroImage*, 37(4), 1033–41.

Gazzaley, A., Cooney, J.W., Rissman, J., D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neurosci*, 8(10), 1298–300.

Gazzaley, A., D'Esposito, M. (2005). BOLD functional MRI and cognitive aging. In: Cabeza, R., Nyberg, L. Park, D., editors. *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*. New York, NY, US: Oxford University Press, pp. 107–31.

Grady, C.L. (2008). Cognitive neuroscience of aging. *Annals of the New York Academy of Sciences*, 1124, 127–44.

Handwerker, D.A., Gazzaley, A., Inglis, B.A., D'Esposito, M. (2007). Reducing vascular variability of fMRI data across aging populations using a breatholding task. *Human Brain Mapping*, 28(9), 846–59.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4), 1639–51.

Hesselmann, V., Zaro Weber, O., Wedekind, C., et al. (2001). Age related signal decrease in functional magnetic resonance imaging during motor stimulation in humans. *Neuroscience Letters*, *308*(3), 141–4.

Huettel, S.A., Singerman, J.D., McCarthy, G. (2001). The effects of aging upon the hemodynamic response measured by functional MRI. *NeuroImage*, *13*(1), 161–75.

Joshi, A.A., Shattuck, D.W., Thompson, P.M., Leahy, R.M. (2007). Surface-constrained volumetric brain registration using harmonic mappings. *IEEE Transactions on Medical Imaging*, *26*(12), 1657–69.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, *19*(9), 1498–507.

Mather, M., Canli, T., English, T., et al. (2004). Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science*, *15*(4), 259–63.

Mega, M.S., Dinov, I.D., Mazziotta, J.C., et al. (2005). Automated brain tissue assessment in the elderly and demented population: construction and validation of a sub-volume probabilistic brain atlas. *NeuroImage*, *26*(4), 1009–18.

Millman, K.J., Brett, M. (2007). Analysis of functional magnetic resonance imaging in Python. *Computing in Science and Engineering*, *9*(3), 52–5.

Poldrack, R.A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*, 67–70.

Postelnicu, G., Zollei, L., Desikan, R., Fischl, B. (2007). Geometry driven volumetric registration. *Information Processing in Medical Imaging*, *20*, 675–86.

Raz, N., Lindenberger, U., Rodrigue, K.M., et al. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral Cortex*, *15*(11), 1676–89.

Raz, N., Rodrigue, K.M., Haacke, E.M. (2007). Brain aging and its modifiers: insights from in vivo neuromorphometry and susceptibility weighted imaging. *Annals of the New York Academy of Sciences*, *1097*, 84–93.

Reuter-Lorenz, P.A. (2002). New visions of the aging mind and brain. *Trends in Cognitive Sciences*, *6*(9), 394–400.

Reuter-Lorenz, P.A., Lustig, C. (2005). Brain aging: reorganizing discoveries about the aging mind. *Current Opinion in Neurobiology*, *15*(2), 245–51.

Robbins, S., Evans, A.C., Collins, D.L., Whitesides, S. (2004). Tuning and comparing spatial normalization methods. *Medical Image Anaslysis*, *8*(3), 311–23.

Rugg, M.D., Morcom, A.M. (2005). The relationship between brain activity, cognitive performance, and aging: the case of memory. In: Cabeza, R., Nyberg, L. Park, D., editors. *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*. New York, NY, US: Oxford University Press, pp. 132–54.

Samanez-Larkin, G.R., Gibbs, S.E.B., Khanna, K., Nielsen, L., Carstensen, L.L., Knutson, B. (2007). Anticipation of monetary gain but not loss in healthy older adults. *Nature Neuroscience*, *10*(6), 787–91.

Samanez-Larkin, G.R., Hollon, N.G., Carstensen, L.L., Knutson, B. (2008). Individual differences in insular sensitivity during loss anticipation predict avoidance learning. *Psychological Science*, *19*(4), 320–3.

Suckling, J., Long, C., Triantafyllou, C., Brammer, M., Bullmore, E. (2006). Variable precision registration via wavelets: optimal spatial scales for inter-subject registration of functional MRI. *NeuroImage*, *31*(1), 197–208.

Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zacks, J.M. (2003). Reliability of functional localization using fMRI. *NeuroImage*, *20*(3), 1561–77.

Weibull, A., Gustavsson, H., Mattsson, S., Svensson, J. (2008). Investigation of spatial resolution, partial volume effects and smoothing in functional MRI using artificial 3D time series. *NeuroImage*, *41*(2), 346–53.

Woolrich, M.W., Behrens, T., Beckmann, C.F., Jenkinson, M., Smith, S.M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–47.