NeuroImage 172 (2018) 478–491

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Improved estimation of subject-level functional connectivity using full and partial correlation with empirical Bayes shrinkage

Amanda F. Mejia[a,*], Mary Beth Nebel[b,c], Anita D. Barber[d], Ann S. Choe[e,f], James J. Pekar[e,f], Brian S. Caffo[g], Martin A. Lindquist[g]

[a] Department of Statistics, Indiana University, USA
[b] Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, USA
[c] Department of Neurology, Johns Hopkins University, USA
[d] Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, USA
[e] Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, USA
[f] F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, USA
[g] Department of Biostatistics, Johns Hopkins University, USA

ARTICLE INFO

Keywords:
Functional connectivity
Connectome
Partial correlation
Reliability
Bayesian statistics
Shrinkage
Measurement error
Resting-state fMRI

ABSTRACT

Reliability of subject-level resting-state functional connectivity (FC) is determined in part by the statistical techniques employed in its estimation. Methods that pool information across subjects to inform estimation of subject-level effects (e.g., Bayesian approaches) have been shown to enhance reliability of subject-level FC. However, fully Bayesian approaches are computationally demanding, while empirical Bayesian approaches typically rely on using repeated measures to estimate the variance components in the model. Here, we avoid the need for repeated measures by proposing a novel measurement error model for FC describing the different sources of variance and error, which we use to perform empirical Bayes shrinkage of subject-level FC towards the group average. In addition, since the traditional intra-class correlation coefficient (ICC) is inappropriate for biased estimates, we propose a new reliability measure denoted the mean squared error intra-class correlation coefficient (ICC_{MSE}) to properly assess the reliability of the resulting (biased) estimates. We apply the proposed techniques to test-retest resting-state fMRI data on 461 subjects from the Human Connectome Project to estimate connectivity between 100 regions identified through independent components analysis (ICA). We consider both correlation and partial correlation as the measure of FC and assess the benefit of shrinkage for each measure, as well as the effects of scan duration. We find that shrinkage estimates of subject-level FC exhibit substantially greater reliability than traditional estimates across various scan durations, even for the most reliable connections and regardless of connectivity measure. Additionally, we find partial correlation reliability to be highly sensitive to the choice of penalty term, and to be generally worse than that of full correlations except for certain connections and a narrow range of penalty values. This suggests that the penalty needs to be chosen carefully when using partial correlations.

Introduction

Measurement reliability is a persistent concern in psychological science (Button et al., 2013; Munafò et al., 2014; Collaboration, 2015). Functional connectivity (FC) of the brain, as measured using resting-state functional magnetic resonance imaging (rs-fMRI), is no exception (Shehzad et al., 2009). Driven by the growing role of subject-level FC estimates in fingerprinting (Finn et al., 2015; Airan et al., 2016), precision functional connectomics (Gordon et al., 2017), brain-behavior studies (Smith et al., 2015), and surgical planning (Tie et al., 2014), determining the best practices for reliable estimation of FC is an important and ongoing topic of research (e.g., Anderson et al., 2011; Birn et al., 2013; Laumann et al., 2015; Noble et al., 2017b). An analysis technique that has been shown to improve reliability of subject-level FC and related measures is *shrinkage*, a statistical estimation method in which individual observations "borrow strength" from a larger group of observations (Su et al., 2008; Varoquaux et al., 2010; Shou et al., 2014; Mejia et al., 2015; Dai et al., 2016; Chong et al., 2017; Rahim et al., 2017).

* Corresponding author.
E-mail address: afmejia@iu.edu (A.F. Mejia).

https://doi.org/10.1016/j.neuroimage.2018.01.029
Received 9 October 2017; Received in revised form 7 January 2018; Accepted 12 January 2018

Shrinkage belongs to the more general family of Bayesian approaches. Fully Bayesian approaches, such as that proposed by Warnick et al. (2017), use a latent variable model in which the unknown connectivity for each subject gives rise to the unobserved "true" time series plus random noise, and subjects are drawn from some population distribution. In this framework, prior distributions are assumed on the parameters controlling the population distribution and the random noise, including the variance within and across subjects. Bayesian computation techniques like Markov chain Monte Carlo (MCMC) or variational Bayes (VB) are used to estimate or sample from the posterior distribution of each parameter and latent variable in the model. The posterior distribution of the connectivity for each subject can then be used to obtain estimates through the posterior mode as well as inference through the posterior quantiles.

Since fully Bayesian approaches tend to be computationally intensive, empirical Bayesian approaches are often employed as an efficient and convenient alternative. In the empirical Bayesian framework, certain parameters are estimated a-priori using the data or prior knowledge obtained from existing studies. Given these parameter estimates, the desired posterior quantities often have a closed-form solution, greatly facilitating computation. Empirical Bayes shrinkage estimators are an example of this approach and result from assuming a measurement error model on a set of estimates. For example, in our case of a Gaussian population prior with independent Gaussian errors, the empirical Bayes shrinkage estimates are weighted combinations of the subject-level observation and the group average, where the degree of shrinkage towards the group average that gives rise to the posterior mean and minimizes mean squared error (MSE) relative to the truth is equal to the ratio of within-subject variance to total (within-subject plus between-subject) variance (James and Stein, 1961; Efron and Morris, 1975). Therefore, lower within-subject variance combined with higher between-subject variance leads to less shrinkage of subject-level estimates toward the group, while higher within-subject variance and lower between-subject variance leads to greater shrinkage.

Previous work has clearly illustrated the benefits of shrinkage, with 25–30% gain in reliability of subject-level connectivity (Varoquaux et al., 2010; Shou et al., 2014; Dai et al., 2016; Rahim et al., 2017) and parcellations (Mejia et al., 2015; Chong et al., 2017). However, estimating the relevant variance components to determine the degree of shrinkage has typically relied on having access to repeated measures through test-retest fMRI data. This limits the applicability of shrinkage methods, since in many studies only a single rs-fMRI session is available for most if not all subjects, and even if multiple sessions were available one would want to utilize the full data available for each subject to improve estimation.

In this work, we propose a novel method to compute empirical Bayes shrinkage estimates of FC, where the degree of shrinkage is determined using single-session fMRI data. Previous work has proposed using "pseudo test-retest" data, in which a single scanning session is split into two contiguous sub-sessions, as a proxy for inter-session variance (Mejia et al., 2015; Mueller et al., 2015). However, this will tend to overestimate the sampling variance of FC, since fewer time points are used in its estimation. In Mejia et al. (2015), we proposed using an empirically determined adjustment factor to correct for this, but the generalizability of such an approach is limited. Here, we instead propose a measurement error model for FC and describe how this model can be used to estimate within-subject variance of FC using single-session fMRI data. Leveraging recent developments in the study of moment-to-moment changes in FC, this model assumes that within-subject variance of FC comes not only from sampling error, but from changes in true FC over time, i.e. dynamic connectivity (Allen et al., 2014). The measurement error model, resulting shrinkage estimator, and variance component estimation techniques are described in the Methods section.

Assessing the reliability of shrinkage estimates is also a challenge, since the intra-class correlation coefficient (ICC), a commonly used and interpretable reliability metric, is not appropriate for biased estimators. Therefore, most reliability studies for shrinkage estimates have relied on mean squared error (MSE) using simulations or test-retest fMRI data to illustrate the gains in reliability due to shrinkage. However, MSE is sensitive to measurement scale and lacks the convenient interpretation of ICC, which ranges from 0 to 1 and represents the proportion of variance in the observations due to true between-subject differences rather than within-subject error or deviation. We therefore propose combining ICC and MSE into a novel reliability measure for biased or unbiased estimators, ICC_{MSE}. ICC_{MSE} is equal to ICC for unbiased estimators but is also appropriate for biased estimators and allows for fair and intuitive comparison between shrinkage and traditional estimators. We motivate and describe this measure in the Methods section.

Finally, we explore the role of scan duration in reliability of both shrinkage and traditional estimates of FC, as the effect of scan duration on reliability of FC is a topic of much recent interest (e.g., Shehzad et al., 2009; Van Dijk et al., 2010; Anderson et al., 2011; Birn et al., 2013; Laumann et al., 2015; Noble et al., 2017a). Several recent studies have also observed substantial differences in reliability across connections. For example, connections within the default mode network (DMN) have been found to exhibit particularly high reliability even for short scan duration, while connections involving the motor network tend to exhibit poor reliability (Shehzad et al., 2009; Van Dijk et al., 2010; Anderson et al., 2011; Laumann et al., 2015; Mueller et al., 2015; Finn et al., 2015). Therefore, we also explore the relationship between scan duration and reliability for connections within different resting-state networks. In addition, most of the extant literature on the relationship between scan duration and FC reliability uses Pearson correlation coefficients as the primary measure of the degree of connectivity between brain regions. The issue of sufficient scan duration deserves greater investigation in the context of partial correlations, which are becoming increasingly popular for their ability to distinguish between brain regions that are directly versus indirectly correlated (Smith et al., 2011, 2015; Varoquaux and Craddock, 2013; Wang et al., 2016).

We perform a reliability analysis using data from the Human Connectome Project (HCP) to examine the role of scan duration, shrinkage, and connectivity measure (full and partial correlation) on reliability of functional connectivity. The HCP is ideal for this analysis due to its large sample size and relatively long duration of rs-fMRI scans. We assess reliability at multiple levels: omnibus reliability over all connections, reliability of within-network connections, reliability of all connections with a particular seed, and reliability of individual connections. This multi-resolution approach provides a more complete picture of reliability and illustrates that reliability of FC is more complex than a single measure can detect. For estimating partial correlations through ridge regression, we first perform a reliability study to assess the impact of the regularization parameter, ρ. We find that certain values of ρ lead to partial correlation estimates with improved reliability for *particular* connections but worse reliability overall compared with full correlations. Notably, we find that common choices of ρ, such as 0.01, lead to partial correlations with much worse reliability than full correlations. The reliability study is described in Section 3, and We conclude with a discussion.

Methods

Empirical Bayes shrinkage using single-session data

Let $q = 1, \ldots, Q$ index the nodes (voxels, vertices or regions) between which we wish to estimate pairwise connectivity, and consider a single pair of nodes q and q'. The overall estimate of FC obtained from an fMRI session is often referred to as *static* connectivity, in contrast to *dynamic* connectivity across the session. As we describe in the next section, considering static connectivity as an average over a dynamic connectivity time series enables us to use the central limit theorem to model and estimate the relevant variance components. Hereafter, the measure of FC is assumed to be Fisher-transformed correlation or partial correlation.

A measurement error model for functional connectivity

Assume here that all subjects have the same scan duration, and let