

# Test-Retest Reliability of the Default Mode Network in a Multi-Centric fMRI Study of Healthy Elderly: Effects of Data-Driven Physiological Noise Correction Techniques

**Rocco Marchitelli,<sup>1,\*</sup> Ludovico Minati,<sup>1,2</sup> Moira Marizzoni,<sup>3</sup> Beatriz Bosch,<sup>4</sup> David Bartrés-Faz,<sup>5</sup> Bernhard W. Müller,<sup>6</sup> Jens Wiltfang,<sup>6,7</sup> Ute Fiedler,<sup>6</sup> Luca Roccagliati,<sup>8,9</sup> Agnese Picco,<sup>10</sup> Flavio Nobili,<sup>10</sup> Oliver Blin,<sup>11</sup> Stephanie Bombois,<sup>12</sup> Renaud Lopes,<sup>12</sup> Régis Bordet,<sup>12</sup> Julien Sein,<sup>13</sup> Jean-Philippe Ranjeva,<sup>13</sup> Mira Didic,<sup>14,15</sup> Hélène Gros-Dagnac,<sup>16,17</sup> Pierre Payoux,<sup>16,17</sup> Giada Zoccatelli,<sup>18</sup> Franco Alessandrini,<sup>18</sup> Alberto Beltramello,<sup>18</sup> Núria Bargalló,<sup>19</sup> Antonio Ferretti,<sup>20,21</sup> Massimo Caulo,<sup>20,21</sup> Marco Aiello,<sup>22</sup> Carlo Cavalieri,<sup>22</sup> Andrea Soricelli,<sup>22,23</sup> Lucilla Parnetti,<sup>24</sup> Roberto Tarducci,<sup>25</sup> Piero Floridi,<sup>26</sup> Magda Tsolaki,<sup>27</sup> Manos Constantinidis,<sup>28</sup> Antonios Drevelegas,<sup>28,29</sup> Paolo Maria Rossini,<sup>30,31</sup> Camillo Marra,<sup>32</sup> Peter Schönknecht,<sup>33</sup> Tilman Hensch,<sup>33</sup> Karl-Titus Hoffmann,<sup>34</sup> Joost P. Kuijper,<sup>35</sup> Pieter Jelle Visser,<sup>36,37</sup> Frederik Barkhof,<sup>36</sup> Giovanni B. Frisoni,<sup>3,38</sup> and Jorge Jovicich,<sup>1</sup> The PharmaCog Consortium**

<sup>1</sup>Center for Mind/Brain Sciences (CIMEC), University of Trento, Rovereto, Italy

<sup>2</sup>Scientific Department, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

<sup>3</sup>LENITEM Laboratory of Epidemiology, Neuroimaging, & Telemedicine—IRCCS San Giovanni Di Dio-FBF, Brescia, Italy

<sup>4</sup>Alzheimer's Disease and Other Cognitive Disorders Unit, Department of Neurology, Hospital Clinic, and IDIBAPS, Barcelona, Spain

<sup>5</sup>Department of Psychiatry and Clinical Psychobiology, Universitat De Barcelona and IDIBAPS, Barcelona, Spain

<sup>6</sup>LVR-Clinic for Psychiatry and Psychotherapy, Institutes and Clinics of the University Duisburg-Essen, Essen, Germany

<sup>7</sup>Department of Psychiatry and Psychotherapy, University Medical Center (UMG), Georg August University, Göttingen, Germany

<sup>8</sup>Department of Neuroradiology, IRCCS San Martino University Hospital and IST, Genoa, Italy

Additional Supporting Information may be found in the online version of this article.

Giovanni B. Frisoni and Jorge Jovicich contributed equally to this work.

Contract grant sponsor: Pharmacog is funded by the EU-FP7 for the Innovative Medicine Initiative; Contract grant number: 115009; Contract grant sponsors: Fondazione IRCCS Istituto Neurologico Carlo Besta (Milano, Italy) and Scienze Mente-Cervello (Rovereto, Italy) (to L.M.)

\*Correspondence to: Rocco Marchitelli, Ph.D. for Mind/Brain Sciences, University of Trento, Italy. E-mail: rocco.marchitelli@unitn.it

Received for publication 17 August 2015; Revised 16 February 2016; Accepted 17 February 2016.

DOI: 10.1002/hbm.23157

Published online 17 March 2016 in Wiley Online Library (wileyonlinelibrary.com).

<sup>9</sup>Department of Health Sciences, University of Genoa, Genoa, Italy

<sup>10</sup>Department of Neuroscience, Ophthalmology, Genetics and Mother-Child Health (DINOGLMI), University of Genoa, Genoa, Italy

<sup>11</sup>Pharmacology, Assistance Publique — Hôpitaux De Marseille, Aix-Marseille University—CNRS, UMR, Marseille 7289, France

<sup>12</sup>University of Lille, INSERM, CHU Lille, U1171 - Degenerative and Vascular Cognitive Disorders, Lille, France

<sup>13</sup>CRMBM-CEMEREM, UMR 7339, Aix Marseille Université—CNRS, Marseille, France

<sup>14</sup>APHM, CHU Timone, Service De Neurologie Et Neuropsychologie, Marseille, France

<sup>15</sup>Aix-Marseille Université, INSERM INS UMR\_S 1106, Marseille, 13005, France

<sup>16</sup>INSERM, Imagerie Cérébrale Et Handicaps Neurologiques, UMR 825, Toulouse, France

<sup>17</sup>Université De Toulouse, UPS, Imagerie Cérébrale Et Handicaps Neurologiques, UMR 825, CHU Purpan, Place Du Dr Baylac, Toulouse Cedex 9, France

<sup>18</sup>Department of Neuroradiology, General Hospital, Verona, Italy

<sup>19</sup>Department of Neuroradiology and Magnetic Resonance Image Core Facility, Hospital Clínic De Barcelona, IDIBAPS, Barcelona, Spain

<sup>20</sup>Department of Neuroscience Imaging and Clinical Sciences, University "G. d'Annunzio" of Chieti, Italy

<sup>21</sup>Institute for Advanced Biomedical Technologies (ITAB), University "G. d'Annunzio" of Chieti, Italy

<sup>22</sup>IRCCS SDN, Naples, Italy

<sup>23</sup>University of Naples Parthenope, Naples, Italy

<sup>24</sup>Section of Neurology, Centre for Memory Disturbances, University of Perugia, Perugia, Italy

<sup>25</sup>Perugia General Hospital, Medical Physics Unit, Perugia, Italy

<sup>26</sup>Perugia General Hospital, Neuroradiology Unit, Perugia, Italy

<sup>27</sup>3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>28</sup>Interbalkan Medical Center of Thessaloniki, Thessaloniki, Greece

<sup>29</sup>Department of Radiology, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>30</sup>Department of Geriatrics, Neuroscience & Orthopaedics, Catholic University, Policlinic Gemelli, Rome, Italy

<sup>31</sup>IRCSS S.Raffaele Pisana, Rome, Italy

<sup>32</sup>Center for Neuropsychological Research, Catholic University, Rome, Italy

<sup>33</sup>Department of Psychiatry, University Hospital Leipzig, Leipzig, Germany

<sup>34</sup>Department of Neuroradiology, University Hospital Leipzig, Leipzig, Germany

<sup>35</sup>Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, the Netherlands

<sup>36</sup>Alzheimer Centre and Department of Neurology, Vrije Universiteit University Medical Center, Amsterdam, the Netherlands

<sup>37</sup>Department of Psychiatry and Neuropsychology, Alzheimer Center Limburg, University of Maastricht, Maastricht, the Netherlands

<sup>38</sup>Memory Clinic and LANVIE, Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland

---

**Abstract:** Understanding how to reduce the influence of physiological noise in resting state fMRI data is important for the interpretation of functional brain connectivity. Limited data is currently available to assess the performance of physiological noise correction techniques, in particular when evaluating longitudinal changes in the default mode network (DMN) of healthy elderly participants. In this 3T harmonized multisite fMRI study, we investigated how different retrospective physiological noise cor-

rection (rPNC) methods influence the within-site test-retest reliability and the across-site reproducibility consistency of DMN-derived measurements across 13 MRI sites. Elderly participants were scanned twice at least a week apart (five participants per site). The rPNC methods were: none (NPC), Tissue-based regression, PESTICA and FSL-FIX. The DMN at the single subject level was robustly identified using ICA methods in all rPNC conditions. The methods significantly affected the mean *z*-scores and, albeit less markedly, the cluster-size in the DMN; in particular, FSL-FIX tended to increase the DMN *z*-scores compared to others. Within-site test-retest reliability was consistent across sites, with no differences across rPNC methods. The absolute percent errors were in the range of 5–11% for DMN *z*-scores and cluster-size reliability. DMN pattern overlap was in the range 60–65%. In particular, no rPNC method showed a significant reliability improvement relative to NPC. However, FSL-FIX and Tissue-based physiological correction methods showed both similar and significant improvements of reproducibility consistency across the consortium (ICC = 0.67) for the DMN *z*-scores relative to NPC. Overall these findings support the use of rPNC methods like tissue-based or FSL-FIX to characterize multisite longitudinal changes of intrinsic functional connectivity. *Hum Brain Mapp* 37:2114–2132, 2016. © 2016 Wiley Periodicals, Inc.

**Key words:** task-free fMRI; resting-state networks; default mode network; physiological noise correction; multisite; test-retest reliability; independent component analysis

## INTRODUCTION

The default mode network (DMN) refers to a set of anatomical regions that are functionally connected, especially during periods of cognitive rest, and that deactivate during externally directed cognitive tasks [Raichle and Snyder, 2007]. The activity in the DMN has been associated with inner cognitive processes [Buckner et al., 2008] that can be affected by a variety of brain diseases [Castellanos et al., 2013]. Further, DMN connectivity has been correlated to validated biomarkers, such as  $\beta$ -amyloid (A $\beta$ ) and total tau protein (T-tau) in cerebral spinal fluid (CSF) [Li et al., 2013]. These biomarkers have been used in preclinical contexts to predict the onset or track the staging

of dementia-related cognitive deterioration in elderly [Hedden et al., 2009; Vannini et al., 2012]. Overall, such results make resting-state functional Magnetic Resonance Imaging (fMRI) DMN connectivity a promising biomarker candidate that is non-invasive and easy to acquire.

The functional connectivity (FC) of the DMN is measured via correlations of blood oxygen level dependent (BOLD) time series among its distinct regions [Buckner, 2012]. Despite its potential, the analysis of resting-state fMRI data remains challenging due to several confounds present in the BOLD signal [Murphy et al., 2013], in particular the physiological signal fluctuations related to time-varying signals of non-brain origin including cardiac and respiratory activities [Birn, 2012].

A variety of techniques and algorithms have been proposed to reduce the “biological noise” component of the BOLD signal. These techniques include the use of cardiac and respiration measures during the experiment, short TR acquisitions to avoid aliasing of cardiac and respiratory fluctuations [Birn et al., 2008; Rasmus et al., 2006; Lund, 2001], tissue-based regressions like averaged white matter (WM) and CSF fMRI signals as nuisance regressors [Jo et al., 2010; Behzadi et al., 2007] or use Bayesian methods to track the frequency trajectories of cardiac and respiration for removing physiological noise [Särkkä et al., 2012]. Other methods exploit canonical correlations analysis to identify autocorrelated physiological noise [Churchill and Strother, 2013] or independent component analysis (ICA) to identify and remove noise signals defined by a spatial reference atlas like in PESTICA [Beall and Lowe, 2007; Beall, 2010] manually selected like in FSL-FIX [Griffanti et al., 2014; Salimi-Khorshidi et al., 2014] or spectral properties of the ICA components [Soldati et al., 2009].

To date, very few studies have directly compared how different physiological noise correction (PNC) methods

## ABBREVIATIONS

(f)MRI	(functional) Magnetic resonance imaging
(group) ICA	(group) Independent component analysis
(r)PNC	(retrospective) Physiological noise correction
BOLD	Blood oxygen level dependent
DMN	Default mode network
EPI	Echo planar imaging
FC	Functional connectivity
FD	Framewise displacement
FSL-FIX	FMRIB's ICA-based Xnoiseifier
GE	General Electric
ICC(s)	Intraclass correlation coefficient(s)
MNI	Montreal neurological institute
NPC	No physiological correction
PESTICA	Physiologic Estimation by Temporal ICA.
QA	Quality assurance
ROI	Region of interest;
TRT reliability	Test-retest reliability
tSNR	Temporal signal-to-noise ratio
WM	White matter

affect the test-retest (TRT) reliability of FC within the DMN, which is of high relevance for the minimization of unwanted variability in longitudinal studies. In a recent study, Birn et al. [2014] investigated a group of young healthy subjects at a 3T site with a focus on PNC methods, using direct measures of cardiac and respiratory activity while also including tissue-based regressions (no ICA methods were used). They found that PNC, in general, did not improve TRT reliability.

TRT datasets are unique and often challenging to obtain, so their public availability may help generalizing the results from different analyses methods. The Consortium for Reliability and Reproducibility [Zuo et al., 2014] is a recent effort that created a public repository of various TRT resting state fMRI datasets independently collected worldwide on 3T MRI machines (Siemens, Philips, GE) using a variety of MR acquisition protocols. To the best of our knowledge it remains unclear how PNC may affect TRT reliability in the context of a multicenter MRI study. In particular, reliability (or lack thereof) may be exacerbated in heterogeneous populations such as healthy elderly subjects. In the context of the current study, this age group is relevant because the patterns of physiological noise variability might not be the same in young and elderly people [Nicolini et al., 2012; Schulz et al. 2013; Taylor and Johnson, 2010]. The contributions of cardiac and respiratory activity on BOLD signal fluctuations as a function of age are unknown, so a conservative approach is to consider that physiological effects in young and elderly people may be different. In the context of multisite experiments, retrospective fMRI based PNC (rPNC) methods are of special interest because short TR acquisitions or direct physiological measurements are more challenging to implement uniformly across clinical sites that may not have special equipment or access to special MRI sequences.

The purpose of this study was to evaluate whether different rPNC approaches, intended as methods that minimize the cardiac and respiratory components in the resting state BOLD fMRI signal, would differently affect the within-site TRT reliability of DMN connectivity (z-scores, cluster-size, cluster-overlap) and the consistency of the reliability across a consortium. The rPNC methods were four, all based on ICA or region of interest (ROI) estimates from the fMRI signal (no physiological correction (NPC), Tissue-based regression, PESTICA and FSL-FIX). The raw data of this study is public so it can be used to evaluate additional physiological or head motion correction methods.

## MATERIALS AND METHODS

Several aspects of the subjects, study design, and data preparation steps used in this FC study were already described in recent morphometric [Jovicich et al., 2013; Marizzoni et al., 2015], diffusion [Jovicich et al., 2014] and

resting state [Jovicich et al., 2016] MRI studies but are here repeated for completeness and with the appropriate modifications.

## Subjects

Fifteen clinical sites (13 MRI sites) participated in this study across Italy (Brescia, Verona, Genoa, Rome, Chieti, Perugia and Naples), Spain (Barcelona), France (Marseille, Lille, and Toulouse), Germany (Essen, Leipzig), Greece (Thessaloniki) and The Netherlands (Amsterdam). The Brescia site was responsible for the coordination and analysis of the whole study and did not acquire MRI data. Each MRI site recruited five local volunteers within an age range of 50–80 years. The subject's age range corresponds to that of the clinical population that will be studied with the protocols tested in this reproducibility study. Each subject underwent two MRI sessions completed at least 7 days (but no more than 60 days) apart at the site. This short period between the test and retest sessions was chosen to minimize biological changes that could affect the reliability of the measures and to mimic the variability expected from separate sessions, as measured in longitudinal studies. Table I summarizes information about age, gender and TRT interval times of the subjects scanned at each MRI acquisition site. All participants were volunteers with no history of major psychiatric, neurological or cognitive impairment (referred to as healthy in this study), were Caucasian and provided written informed consent in accordance with the local ethical committee for each institution. Exclusion criteria were described in previous work [Jovicich et al., 2014].

## MRI Acquisitions

The thirteen 3T MRI sites that participated in this study used a variety of MRI system vendors and models (Siemens, GE, and Philips). Table I summarizes the main MRI system and EPI acquisition differences across sites. Only vendor-provided sequences were used.

Each subject had a total of two resting state EPI acquisitions (rsfMRI), one from the test session and one from the retest session. In each session the following single shot EPI acquisition parameters were common across sites: nominal voxel size  $3 \times 3 \times 3 \text{ mm}^3$ , TE = 30 ms, TR = 2.7 s, Ernst flip angle of 85°, one-pass interleaved axial slices acquired (equidistant choice in Philips, default options in GE and Siemens) oriented parallel to the AC-PC line covering the full brain, 0.45 mm slice gap (15% slice thickness), 40 slices, 200 volumes, fat suppression, no parallel imaging. TR was chosen to be the shortest common possible value in the consortium allowing full brain coverage while keeping consistent temporal resolution across sites. The total resting state acquisition duration was 9 min, a duration that has been shown appropriate for reliable intrinsic connectivity results [Liao et al., 2013; Van Dijk et al., 2010; Zuo

**TABLE I. Outline of MRI site, resting state fMRI acquisition and demographic differences**

MRI site location	Subjects' age: mean $\pm$ std dev, (range) years	Test-retest time interval (days)	Gender (females/N)	3T MRI scanner	MR system software version	TX/RX coil	Fat suppression	In-plane matrix	tSNR median (IQR)
Site 1 Verona	67.8 $\pm$ 9.9 (26)	28 $\pm$ 23	2/5	Siemens Allegro	V25A	Birdcage	Fat Sat.	64x64	69 $\pm$ 8.8
Site 2 Barcelona	74.6 $\pm$ 2.7 (6)	10 $\pm$ 3	5/5	Siemens Trio Tim	B17	Body/ 8-chan.	Fat Sat.	72x72	118.5 $\pm$ 8.1
Site 3 Leipzig	62.8 $\pm$ 2.6 (6)	13 $\pm$ 3	3/5	Siemens Trio Tim	B17	Body/ 8-chan.	Fat Sat.	72x72	110 $\pm$ 18.3
Site 4 Marseille	66.0 $\pm$ 8.3 (20)	23 $\pm$ 22	4/5	Siemens Verio	B17	Body/ 12-chan.	Fat Sat.	72x72	134.5 $\pm$ 6
Site 5 Essen	52.4 $\pm$ 1.5 (3)	11 $\pm$ 5	2/5	Siemens Skyra	D11	Body/ 20-chan.	Fat Sat.	72x72	125 $\pm$ 62
Site 6 Naples	59.0 $\pm$ 3.5 (9)	19 $\pm$ 15	2/5	Siemens Biograph mMR	B18P	Body/ 12-chan.	Fat Sat.	72x72	94 $\pm$ 19.4
Site 7 Lille	64.2 $\pm$ 5.3 (13)	15 $\pm$ 11	3/5	Philips Achieva	3.2.2	Body/ 8-chan.	SPIR	80x80	120.5 $\pm$ 20
Site 8 Toulouse	59.2 $\pm$ 4.5 (12)	14 $\pm$ 10	3/5	Philips Achieva	3.2.2	Body/ 8-chan.	SPIR	80x80	100.4 $\pm$ 12.7
Site 9 Chieti	68.8 $\pm$ 4.3 (11)	11 $\pm$ 5	5/5	Philips Achieva	3.2.2	Body/ 8-chan.	SPIR	80x80	114.8 $\pm$ 51
Site 10 Perugia	60.8 $\pm$ 10.3 (24)	10 $\pm$ 4	3/5	Philips Achieva	3.2.2	Body/ 8-chan.	SPIR	80x80	118.6 $\pm$ 19
Site 11 Genoa	58.2 $\pm$ 2.2 (5)	24 $\pm$ 17	2/5	GE HDxt	15 M4A	Body/ 8-chan.	Water only	64x64	136.3 $\pm$ 19
Site 12 Thessaloniki	56.6 $\pm$ 5.5 (5)	32 $\pm$ 8	3/5	GE HDxt	15 M4A	Body/ 8-chan.	Water only	64x64	118.4 $\pm$ 20.3
Site 13 Amsterdam	62.8 $\pm$ 8.2 (21)	11 $\pm$ 6	3/5	GE HDxt	DV22.0	Body/ 8-chan	Water only	64x64	158.6 $\pm$ 83.3
					DV23.1				

<sup>1</sup>A software upgrade took place during the study at site 13; DV22 was used for the first participant and DV23.1 for the rest.

et al., 2013]. Each MRI session included the acquisition of two T1 weighted anatomical scans [Jovicich et al., 2013]. Subjects were instructed to keep their eyes closed, stay relaxed and try to avoid falling asleep.

Other acquisition parameters including head RF coil characteristics, pulse sequence and fat suppression methods were difficult to standardize due to system differences. The choices for these parameters were made based on the optimal or possible options available at the different platforms (see Table I). All images from multi-channel coils were reconstructed online as the sum of the squares across channels. When allowed by the MRI system, images were reconstructed and saved disabling additional filtering options that could differ across scanners introducing different degrees of smoothing.

### Pre-Processing of Functional MRI Data with NPC

Pre-processing was performed in the individual space of each subject using a combination of FSL [Jenkinson et al., 2012] and AFNI [Cox, 1996] programs in the same order as they are listed. The following pre-processing steps were performed with NPC (Fig. 1). The first four volumes were discarded (fslroi, FSL) to allow for steady state stabilization of the BOLD fMRI signal. Then, EPI volumes were realigned (mcflirt, FSL) and six head movement parameters were calculated; slice timing correction based on the slice acquisitions parameters at each site (slicetimer, FSL); non-brain voxels removal (bet2, FSL); temporal filtering with a bandpass filter (0.01–0.1 Hz) (fslmaths, FSL) to remove low and high-frequency signal fluctuations. Additional confounds, such as the six head movement parameters and their derivatives (1d\_tool.py, AFNI), temporally filtered (1dBandpass, AFNI) as the main signal [Hallquist et al., 2013], were removed from the data using multiple linear regressions (3dDeconvolve, AFNI) for an overall of 12 regressors of no interest plus linear and quadratic trends. Neither regressors nor volumes were censored. 4D EPI volumes were spatially smoothed using 6 mm full-width at half-maximum Gaussian filter (susan, FSL) and normalized to mean signal intensity by a single factor (fslmaths, FSL).

### Retrospective Physiological Noise Correction Methods

In addition to the case of NPC, the following three methods for rPNC were evaluated: (i) PESTICA correction [Beall and Lowe, 2007]; (ii) Tissue-based correction [Weissenbacher et al., 2009]; and (iii) FSL-FIX correction [Griffanti et al., 2014; Salimi-Khorshidi et al., 2014]. Each of these methods followed a slightly different preprocessing workflow, as schematically shown in Figure 1 and described below. All preprocessing pipelines, with or without rPNC, were performed in single subject space.

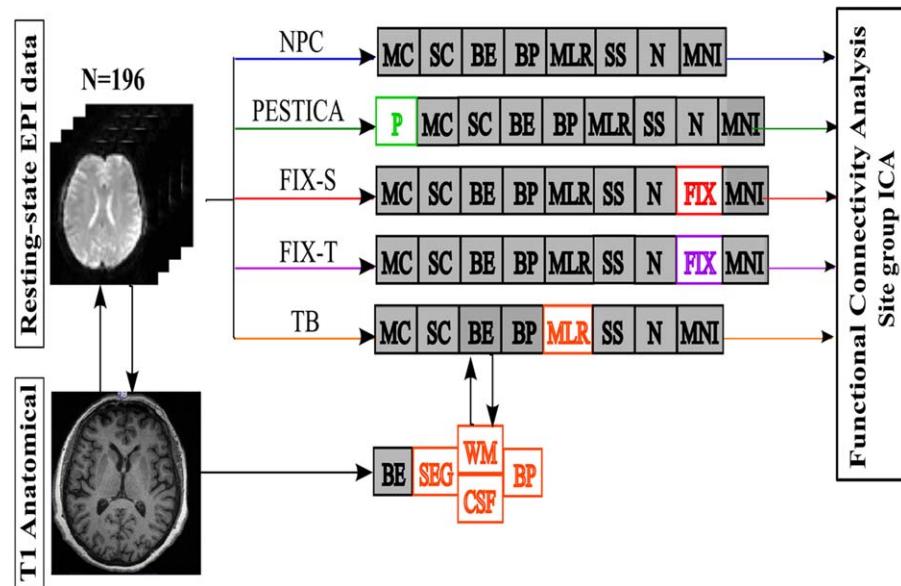


Figure 1.

Resting-state fMRI preprocessing outline. The diagram shows main pre-processing steps for each rPNC method under investigation: NPC (blue); PESTICA (P, green); Tissue-based (TB, orange); FSL-FIX using standard (FIX-S, red) or site-specific classifier (FIX-T, violet). MC, motion correction; SC, slice-timing correction; BE, brain extraction; BP, band-pass filter (0.01–0.1 Hz); MLR, multiple linear regression; SS, spatial smoothing (FWHM = 6 mm); N, mean intensity normalization; MNI, MNI

coregistration. TB correction required anatomical image pre-processing: BE, brain extraction; SEG, segmentation; WM/CSF, white matter/CSF regressors; BP, band-pass filtering of regressors (WM/CSF regressors were bandpass filtered). Anatomical BE images were also used for FSL-FIX (see text). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

### PESTICA Correction

Physiologic Estimation by Temporal ICA (<http://www.nitrc.org/projects/pestica/>) (PESTICA) is a freely available retrospective physiologic denoising algorithm (AFNI, MATLAB) that estimates the cardiac and breathing cycle directly from the BOLD fMRI data using temporal ICA [Beall and Lowe, 2007]. The software defines two regressors-of-no-interest from the two most correlated independent components to spatial template references of heart beating and breathing, converts and regresses them out from the main signal using IRF-RETROICOR [Beall, 2010].

Following the developers' recommendations, PESTICA v2.0 was run before slice timing correction. Default parameters suggested by the developers were used, such as temporal ICA dimensionality (15 components) and the full range for cardiac and respiratory estimators (48–85 bpm for cardiac and 10–24 bpm for breathing). All recommended PESTICA QA tools were used to verify that the algorithm performed properly: (1) visual inspection to assess the correct coregistration of the current EPI volume and the mean EPI template provided within PESTICA; (2) Plots of cardiac and breathing estimators; (3) coupling maps of cardiac and respiratory profiles.

### Tissue-Based Correction

Brain Tissue-based correction is a common method to remove physiological confounds without a predefined model of noise [Birn et al., 2009; Bright and Murphy, 2013; Weissenbacher et al., 2009]. This method relies on the assumption that BOLD fluctuations in WM and CSF are dominated by physiological noise, whereas BOLD fluctuations in gray matter (GM) combine signals from neural origin and physiological noise. One approach, the global regression method, removes the mean signal from the full brain. Therefore it might cause the loss of neural signal from GM. Instead, regression of the average WM and CSF time series from GM is expected to reduce the impact of physiological fluctuations in GM. Therefore we chose the latter approach, here referred to as Tissue-based correction. Tissue-based correction is hence applied during multiple linear regressions together with movement confounds (Fig. 1).

Tissue-based correction requires processing of T1 anatomical images in order to spatially define WM and CSF masks. In this study, for each subject, we only used the first individual anatomical volume acquired at test to define WM and CSF masks. Non-brain tissue was initially removed from center-oriented T1 anatomical volumes (bet,

FSL); the anatomical volume was then registered to the original EPI (`align_epi_anat.py`, AFNI); segmented in WM and CSF using partial volume segmentation (fast, FSL). Masks were thresholded at 0.99 (3dcalc, AFNI) to reduce the inclusion of gray matter voxels into the masks. In order to avoid gross co-registration and segmentation errors visual inspection was used. Two regressors of no interest were calculated from brain-extracted EPI as the mean time series within the non-gray matter masks (3dcalc, AFNI), temporally filtered as the main signal (1dBandpass, AFNI), and removed together with movement confounds by multiple linear regressions (3dDeconvolve, AFNI). The total number of nuisance regressors in this analysis pipeline was therefore 14 (six head movement, six derivative of head movement, mean WM and CSF time series).

### FSL-FIX Corrections

FMRIB's ICA-based Xnoiseifier (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIX>) (FSL-FIX) exploits single-subject spatial ICA to auto-classify noisy components and remove them from the main signal [Griffanti et al., 2014; Salimi-Khorshidi et al., 2014]. Components classified as noise include physiological, head movement and other artifacts relative to image acquisition. The classification of noise components can be performed either by using one of several training files provided by the authors or by generating a data-specific training set on one's own data. In this study both approaches were investigated.

Following the developers' suggestions, FSL-FIX was run after the NPC preprocessing pipeline. No changes were made to the preprocessing pipeline to be consistent across all rPNC methods examined. Temporal filtering was used consistently with common practice and works implementing FSL-FIX [Griffanti et al., 2014]. Spatial smoothing was also performed, since it returned high reproducibility of the tSNR across multiple sites in our previous works [Jovicich et al. 2016]. Single-subject ICA was run with automatic estimation of dimensionality using MELODIC on preprocessed data including the brain-extracted anatomical images. From the various pre-defined FIX training sets offered, we selected the one (herein referred to as FIX-standard) generated from a set of acquisition and preprocessing conditions as close as possible to those employed in this study (Standard.RData: TR = 3s, Resolution =  $3.5 \times 3.5 \times 3.5 \text{ mm}^3$ , session = 6 mins, default FEAT preprocessing including default spatial smoothing).

The data-specific FIX training set (herein referred to as FIX-training) was generated separately for each MRI site using the two acquisition sessions of each subject (i.e., a total of 10 resting state runs per site). FIX-standard and FIX-training corrections were run using a signal/noise threshold of 75 and 20, respectively. Thresholds were determined as the minimum value to consistently remove at least those components that were visually related to

physiological artifacts across subjects, sessions and sites. For FIX-standard this could not be achieved with a threshold lower than 75, whereas the implementation of site-specific classifiers in FIX-training allowed the usage of considerably smaller thresholds.

### Characterization of Group DMN For Each MRI Site

After pre-processing with each rPNC method in the individual space of each subject across all preprocessing workflows, we performed group independent component analysis (group ICA), using MELODIC (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC>) to extract the site-specific group DMN, for each site and each rPNC method separately. For each site and rPNC method, all subjects and sessions (i.e., 10 resting state runs) were spatially normalized to the MNI template via linear (affine) registration [Jenkinson et al., 2002] and subsampled at a resolution of 4 mm isotropic voxels, then decomposed into 10 independent components [Jovicich et al., 2016], using the multi-session temporal concatenation procedure in MELODIC. A higher number of components were not extracted to avoid splitting of the DMN [Abou-Elseoud et al., 2010; Jovicich et al., 2016].

An automatic selection procedure was used for each MRI site and rPNC method to select the group DMN from the 10 ICs in each dataset. In particular, an overlap measure was used to select the DMN using only the posterior regions (posterior cingulate and precuneus, left/right parietal cortex) from an independent functional DMN template [Rosazza et al., 2012] to avoid circularity. The component with the highest number of voxels in common with the template was chosen as representative of the DMN. Both our components and the DMN template were thresholded at  $z$ -scores  $> 2.3$ ,  $P < 0.01$  [Beckmann and Smith, 2004]. The automatically selected group DMN for each site was visually inspected in order to assess overall resemblance with standard DMN patterns including its main co-activation areas.

### Characterization of Single-Subject DMN

For each site and rPNC method, dual-regression was then used to derive the single subject and session DMN from the site-specific group DMN [Beckmann and Filippini, 2009; Zuo et al., 2010]. Single-subject DMN volume maps were thresholded at  $z > 2.3$ ,  $P < 0.01$  (Beckmann and Smith, 2004). To account for spatial DMN variability across sessions and subjects, we used a functional cluster-criterion that defined the DMN by four main clusters (3dclust, AFNI). These clusters were (i) posterior cingulate and precuneus (BA31, BA30, BA29, BA23), (ii) left/right parietal cortex (BA39, BA40, BA22, BA7) and (iii) medial prefrontal cortex (BA9, BA10, BA32, BA24) (Franco et al., 2009).

To avoid inaccuracies for the definition of DMN activation maps, clusters were defined as made of voxels no more than 4mm apart and a cluster volume of at least 1,800 micro-liters, per each ROI. These clusters were anatomically constrained by a reference DMN template derived from conducting a group ICA overall the consortium with NPC implemented. For each rPNC method, spatial convergence defined as the percent overlap between each individual DMN component and this reference template was calculated to quantify single-subject DMN variability across MRI sites after image processing.

Mean z-scores in the DMN (i.e., the mean z-score across all voxels in the four main clusters) and relative activation cluster-size (i.e., the total number of voxels in the four main nodes) were characterized for each subject, session, and rPNC method. Furthermore, the DMN activation cluster-overlap between sessions (i.e., the total number of common voxel coordinates between sessions) was also calculated for each subject and rPNC method.

### Within-Site TRT Reliability of DMN

The main goal of this study was to evaluate the effects of different rPNC techniques on the precision of the DMN-derived measurements. To this end, we considered a measure of TRT reliability within each site in the consortium. Within-site TRT reliability was studied for each rPNC method using the following three metrics: (1) absolute percent change of mean DMN connectivity; (2) absolute percent change of DMN cluster-size, and (3) Jaccard index for DMN cluster-overlap across sessions [Bennett and Miller, 2010; Maitra, 2010; Meindl et al., 2010]. These measurements were then used for the statistical analysis of within-site TRT reliability and its reproducibility consistency across the consortium.

### Across-Site Reproducibility Consistency of the DMN

For each DMN reliability measure and each rPNC condition, we used intra-class correlations, ICC(2,1) to quantify the degree of consistency of TRT reliability scores (here defined as reproducibility consistency) across all 13 sites in the consortium [Bennet and Miller, 2010]. Considering subjects as “raters” and sites as “targets,” ICC measures the proportion of variance between-sites out of the total variance. This formulation returns positive coefficients in the range 0–1 with values close to 1 indicating no strong site differences or biases in TRT reliability scores. Reproducibility consistency was measured using an in-house Matlab script; we used a leave-one-out approach to define the standard deviation of the ICCs.

### Head Movement Measures

Head movement is known to negatively influence FC estimates [Power et al., 2012; Van Dijk et al., 2012]. In order to estimate the amount of head movement, we calculated the framewise displacement (FD) to summarize instantaneous head movement in the data [Power et al., 2012]. Head movement estimations were derived from the six realignment parameters for each subject and session. Head rotation parameters were converted from radians to millimeters in terms of the corresponding displacement on a  $r = 50$  mm sphere, representing the average distance between cortex and head center [Power et al., 2012]. To summarize head movement, we calculated the mean and maxima FD in each subject and session. These movement metrics were used to test for significant differences between sessions (paired *t*-test) as well as MRI site effects (Kruskal-Wallis test).

### Statistical Measures

All statistical measures were performed using IBM SPSS Statistics for Macintosh, Version 22.0. We conducted non-parametric Kruskal-Wallis tests to evaluate MRI site effects for each DMN-derived measurements and related TRT reliability scores under each rPNC approach. The variance estimated was used to calculate reproducibility consistency of reliability across sites, for each DMN-derived measurement, respectively. The Kruskal-Wallis test was used on the various TRT reliability scores to obtain a statistical measure of site independence. The ICC analysis was used as a descriptive measure of reproducibility consistency across sites, for each DMN-derived measurement and rPNC method. We used non-parametric Friedman test to evaluate rPNC effects in each site of the consortium and, if MRI site effects were not found (Kruskal-Wallis,  $P > 0.05$ ), even on the pooled data across the consortium. rPNC method effects on the reproducibility consistency values were also evaluated using Friedman test.

To increase the statistical power of the present analysis, in addition to the non-parametric analysis we also conducted a 2-way ANOVA on the pooled reliability scores across rPNC methods and sites for each DMN-derived measurement, separately. The indexes of kurtosis and skewness were examined to assess the distributional shape of TRT reliability scores and determine whether the assumption of normality was met (kurtosis range: [-2, 2]/skewness range: [-1, 1]) [Bulmer, 1997; Mat Roni, 2014].

Finally, bivariate Pearson’s correlations were conducted to evaluate the existence of a relationship between the TRT reliability of connectivity measurements (z-scores, cluster-size, cluster-overlap) and across-session averaged movement estimates (mean FD). The significance level for all the statistics was set at  $P < 0.05$ . Statistics were corrected for multiple comparisons over all possible pairwise combinations (13 sites for Kruskal-Wallis and five rPNC

methods for the Friedman test) using the method of Dunn-Bonferroni [Dunn, 1964] at  $\alpha = 0.05$ .

## RESULTS

### Head Movement Measures

The consortium distributions of the head movement parameters are shown in Supporting Information Figure 1. Overall we found low head movement: mean (SD) ( $0.11 \pm 0.1$  mm) with only 8% of time points exceeding 0.25 mm and 1% of them exceeding 0.5 mm. Mean FD range was (0.02–0.37 mm) where only the 4% of subjects exceeded higher mean FD than 0.25 mm. Maxima FD range was (0.07–3.32 mm) with roughly the 3% of subjects exceeding 1.5 mm (half-voxel size). Mean FD showed statistical differences across sites (Kruskal-Wallis,  $\chi^2(12, n = 130) = 31, P = 0.002$  uncorrected) but not across sessions (paired *t*-test,  $P = 0.99$ ). Likewise, maxima FD showed site effects (Kruskal-Wallis,  $\chi^2(12, n = 130) = 24, P = 0.02$  uncorrected) and no statistical difference between the two sessions (paired *t*-test,  $P = 0.51$ ). Given that the overall head movement was considered low and that ICA tends to separate movement components, we decided not to censor volumes in this study [Power et al., 2012]. This also lets us use group ICA on time series of the same length (196 vols).

### Effects of MRI Site and rPNC on DMN: Mean z-Scores and cluster-Size

Despite the low number of subjects (five per sites, two repetitions for each), group ICA successfully revealed DMN activation maps in all 11 MRI sites and for all rPNC methods, consistently detected via automatic selection (sample site shown in Fig. 2). rPNC methods, however, showed no significant impact on the consistency of group-level DMN estimates across MR sites (Kruskal-Wallis,  $\chi^2(12, n = 65) = 19.4, P = 0.08$ ). Template matching was consistently higher than 50% across all MRI sites and rPNC methods. Furthermore, the difference between the first and second ranked components was often higher than 30%, suggesting negligible variability between the selected DMN components across MRI sites (Supporting Information Table I).

Single-subject DMN maps were also successfully obtained using dual regression for each rPNC method and site (sample subject shown in Fig. 3). In NPC and FIX-training conditions, DMN patterns were either not found or did not exceed the threshold in one and five subjects, respectively. This allowed us to measure the median and its interquartile range (IQR) across subjects and sessions for the average DMN z-scores and cluster-size for each rPNC approach and MRI site (Fig. 4).

We found that FIX-standard classified 85% of single subject ICA components as noise across the entire consortium,

whereas FIX-training, performed with a more conservative threshold, classified and removed 71% (see Supporting Information Fig. 2A). FIX-standard removed a larger proportion of components than FIX-training at each site with the exception of site 2 (see Supporting Information Fig. 2A). In any case, both techniques were effective at attenuating physiological noise, overall (see Supporting Information Fig. 2B for sample noise ICs).

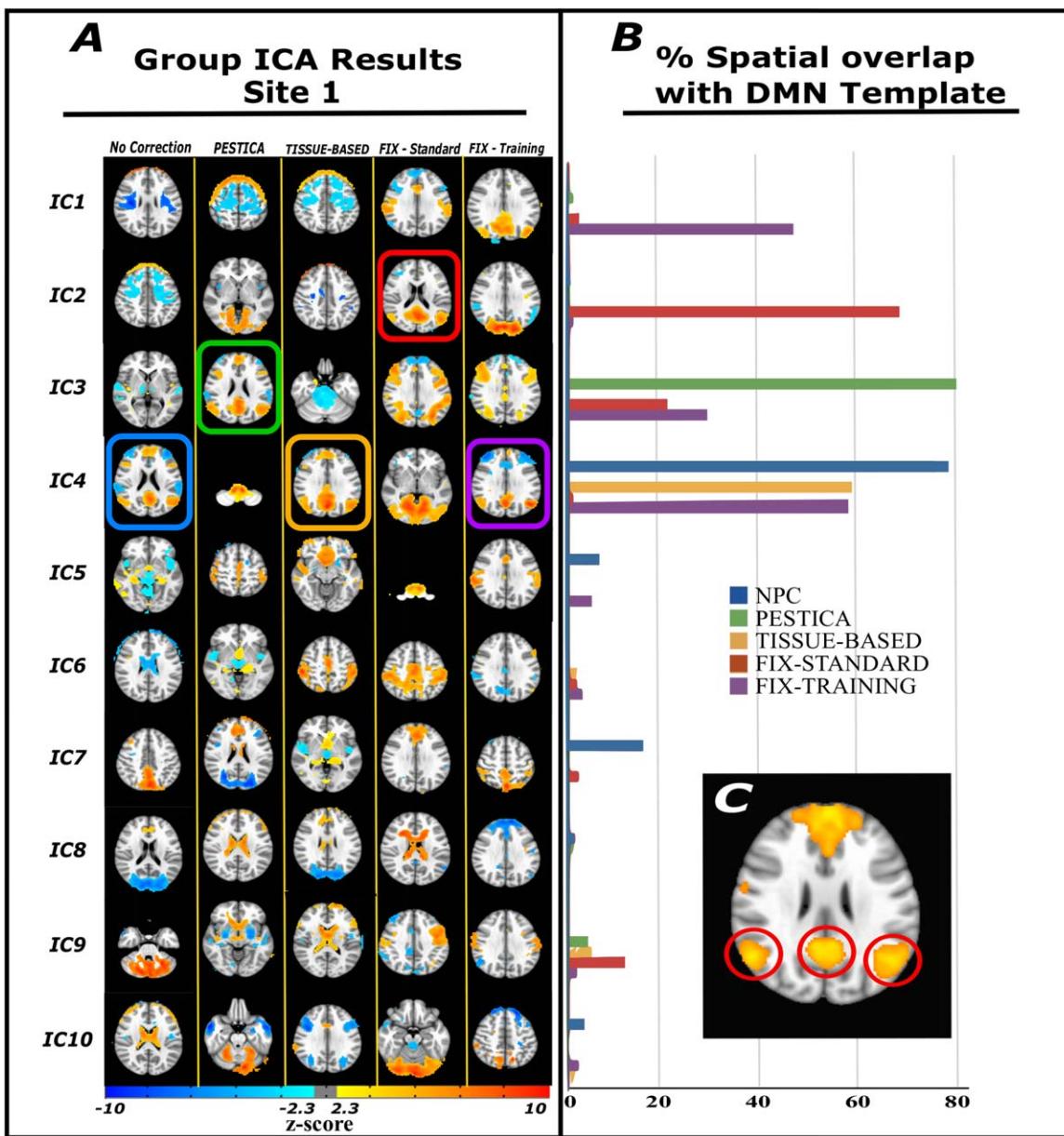
Spatial convergence of single-subject DMN maps was overall medium-to-high within a range of 58–61% (Supporting Information Table II). Despite this convergence, a Kruskal-Wallis test revealed considerable MRI site effects in the spatial convergence of single-subject DMN maps. These were found for NPC, PESTICA, Tissue-based and FIX-standard and abolished only for FIX-training [Kruskal-Wallis,  $\chi^2(12, n = 125) = 17, P = 0.2$ ].

A Kruskal-Wallis test also revealed significant MRI site effects in both DMN mean z-scores [ $\chi^2(12, n = 129) = 29, P = 0.004$  uncorrected] and cluster-size values [ $\chi^2(12, n = 129) = 42, P < 0.001$  uncorrected] when NPC is applied (Table II). Similarly to single-subject DMN spatial convergence, MRI site effects persisted even after applying PESTICA, Tissue-based and FIX-standard methods. Only FIX-training canceled MRI site effects in DMN mean z-scores [Kruskal-Wallis,  $\chi^2(12, n = 125) = 17, P = 0.2$ ], cluster-size [Kruskal-Wallis,  $\chi^2(12, n = 125) = 15, P = 0.3$ ].

In general, FSL-FIX tended to increase median values of mean z-scores within the DMN nodes relative to the other methods (Fig. 4A): FIX-standard significantly increased median values of mean z-scores relative to NPC in four sites and FIX-training in five sites (Table III). We also evaluated how physiological noise correction methods affect the number of active voxels ( $z > 2.3$ ) outside the selected DMN nodes in any of the main tissue compartments: GM, WM and CSF. We found that, outside the DMN, FSL-FIX was the only method to show significant increases of voxels relative to NPC in the GM and surrounding WM and CSF tissue (Supporting Information Fig. 3). Furthermore, these accretions in mean z-scores were not circumscribed to the DMN component but were also found in other components of known resting-state networks (Supporting Information Table III). The other rPNC methods did not exert any significant change on median values of mean z-scores relative to NPC in single sites. We found fewer rPNC effects for DMN cluster-size across sites relative to mean z-scores (Fig. 4B, Table III).

### Effects of rPNC Method on TRT Reliability: DMN Connectivity, Cluster-Size and Cluster-Overlap

Figure 5 shows the TRT reliability of the DMN connectivity metrics within each MRI site and rPNC method. We report the median (IQR) value across subjects at each site and rPNC method, for each DMN-derived measurement under investigation. We found that no rPNC method



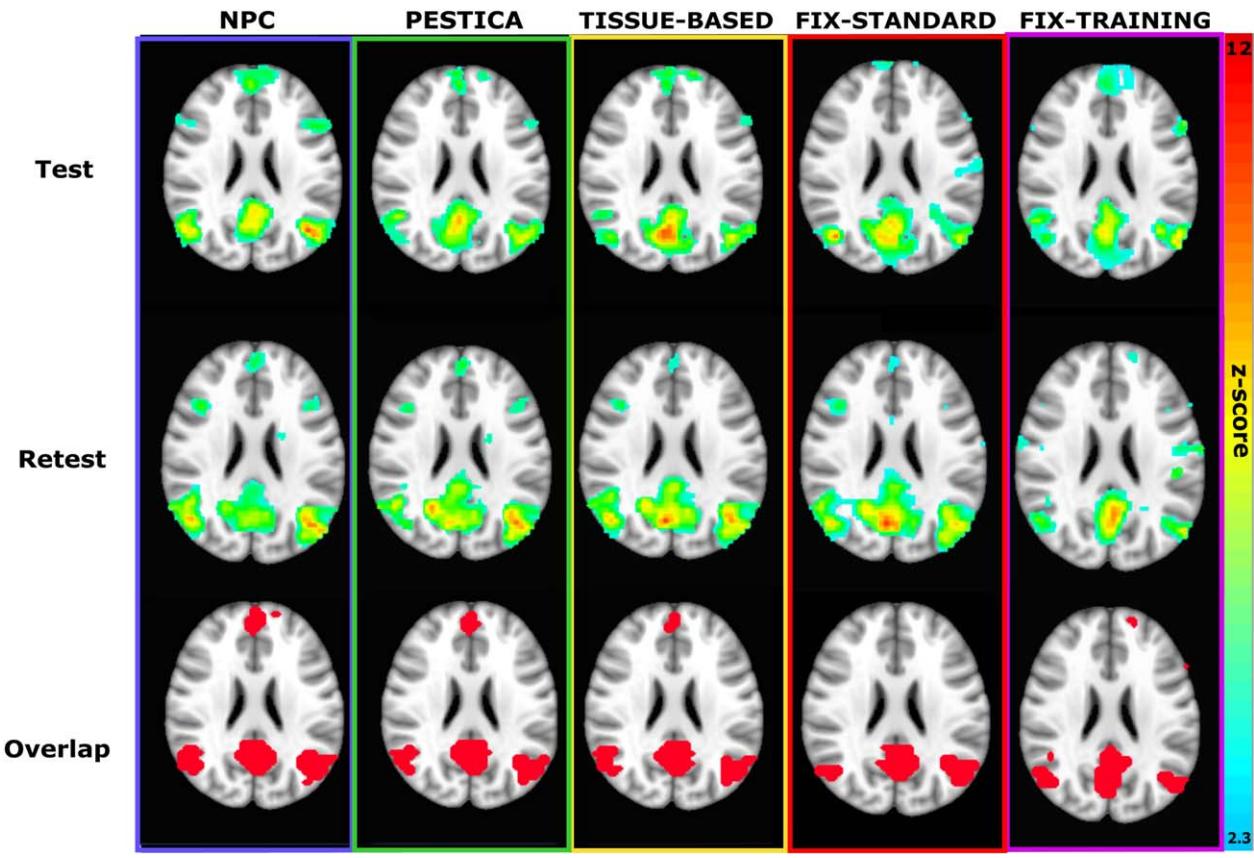
**Figure 2.**

Automatic selection of the group DMN at a sample MRI site. (A) Group ICA decomposition in 10 ICs (rows) for each physiological correction method (columns) is shown for one sample site (site 1). Activation maps include both positive ( $z > 2.3$ ) and negative ( $z < -2.3$ ) co-activations. Colored squares show the components that were automatically selected as the DMN in each rPNC condition. (B) Proportion of overlapping voxels between each inde-

pendent component (IC) and the DMN template is reported for each rPNC method. Only positive z-score ( $z > 2.3$ ) were considered for the selection. (C) The IC with maximum spatial overlap with the posterior regions of the DMN template was selected as the site-group DMN for each rPNC method respectively. See text for more details. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

significantly influenced TRT reliability relative to NPC at each single site for any DMN-derived measurement (Friedman test  $P > 0.05$ , Bonferroni corrected). Furthermore, differences between rPNC methods were not systematic

across MR sites. We found no statistically significant MRI site effects (Kruskal-Wallis,  $P > 0.05$ ) for the TRT reliability of mean z-scores in the DMN or the reliability of its volume cluster-size under all rPNC method conditions. We



**Figure 3.**  
Single-subject DMN example. Reconstructed DMN map after dual regression on a sample single subject (site 1) is shown for each rPNC method (columns) at test (top row), retest (middle row) and its across-session spatial overlap (bottom row). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

found statistically significant MRI site effects (Kruskal-Wallis,  $P < 0.05$ ) for the TRT reliability of cluster-overlap in the DMN. However, these effects did not survive multiple comparisons.

Therefore, we pooled together all the TRT reliability scores of each rPNC method across subjects and sites to examine overall TRT reliability (Fig. 6). Absolute percent errors ranged from 5 to 11% for DMN z-scores and cluster-size reliability. DMN cluster-overlap was in the range 60–65%. There was a statistically significant difference in the TRT reliability of mean z-scores and volume cluster-size in the DMN, depending on the rPNC method implemented, mean z-scores:  $\chi^2(4) = 14$ ;  $P = 0.009$ , uncorrected; Cluster-Size:  $\chi^2(4) = 21$ ;  $P < 0.001$ , uncorrected; (Table III). Multiple comparisons applied via pairwise comparisons revealed that the median (IQR) mean z-scores reliability error of PESTICA, 4.6% (6%), was statistically different from FIX-training, 10.6% (16.9%), ( $Z = -3.6$ ,  $P = 0.003$ ). Pairwise comparisons revealed that the median (IQR) cluster-size reliability error of Tissue-based, 5.3%

(6.3%) was statistically different from FIX-standard, 9.4% (12.1%), ( $Z = -4.08$ ,  $P < 0.001$ ) and FIX-training, 7.3% (12.6%), ( $Z = -3.03$ ,  $P = 0.03$ ). However, no rPNC method showed a statistically significant reduction of TRT reliability of mean z-scores or and cluster-size in relation to NPC within the DMN (Fig. 6).

We found that mean FDs did not correlate with absolute percent errors of DMN mean z-scores and cluster-size for all rPNC methods. However, we found that TRT reliability of cluster-overlap (i.e., the Jaccard index) was significantly anti-correlated with mean FD when using FSL-FIX (FIX-standard:  $r(63) = -0.4$ ,  $P < 0.01$  and FIX-training:  $r(58) = -0.3$ ,  $P < 0.01$ ).

With respect to the ANOVA framework, the distributions of TRT reliability scores were not normally distributed for mean z-scores (kurtosis: 2.5, skewness: 1.6) and cluster-size (kurtosis: 4.5, skewness: 1.9). However, a square root transformation was found to make these distributions more Gaussian-like (mean z-scores: kurtosis: -0.1, skewness: 0.5; cluster-size: kurtosis: 0.2, skewness: 0.6). The TRT reliability

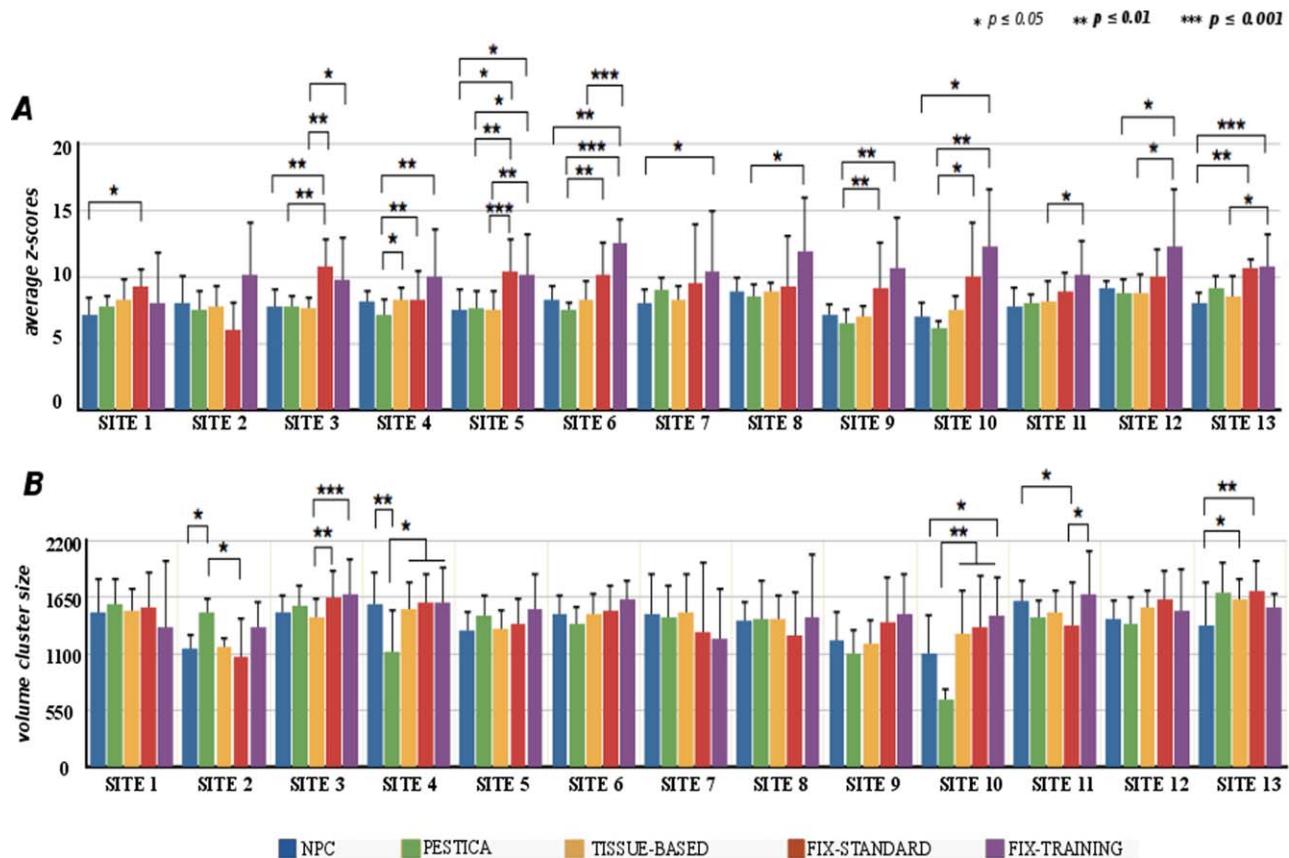


Figure 4.

Within-Site DMN functional connectivity and cluster-size results. Median (IQR) across subjects for supra-thresholded (A) mean z-scores ( $z > 2.3$ ) and (B) volume cluster-size ( $N^{\circ}$  active voxels at  $z > 2.3$ ) within each site and for each PNC method. rPNC effects were evaluated within each site using Friedman test

( $P < 0.05$ , corrected with Dunn-Bonferroni method). Physiological correction using FIX tended to increase FC but not cluster-size in the DMN across sites. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

scores of the cluster-overlap were instead normally distributed (kurtosis: 0.3, skewness: -0.5).

After this transformation, statistically significant rPNC method effects were found in TRT reliability of mean  $z$ -scores [ $F(4,12) = 4.5$ ,  $P = 0.002$  uncorrected]. In addition, statistically significant MRI site effects were found in the TRT reliability of both mean  $z$ -scores [ $F(4,12) = 2$ ,  $P = 0.02$  uncorrected] and cluster-size [ $F(4,12) = 6$ ,  $P < 0.001$  uncorrected]. However, consistently with non-parametric statistical tests, both rPNC method and MRI site effects in the TRT reliability of both DMN-derived measurements did not survive multiple pairwise comparisons. Importantly, none of these metrics showed interaction effects.

Considering the TRT reliability of cluster-overlap (i.e., the Jaccard index), statistically significant MRI site effects [ $F(4,12) = 8.3$ ,  $P < 0.001$  uncorrected] and interaction between rPNC methods and sites [ $F(4,12) = 1.9$ ,  $P < 0.01$  uncorrected] were found. Multiple pairwise comparisons performed between MRI sites for each rPNC method revealed that MRI

site effects were exacerbated under FIX-training correction and minimized under Tissue-based correction, respectively.

#### Effects of rPNC Method on the Reproducibility Consistency of TRT Reliability

Figure 7 shows reproducibility consistency (ICC) results for the three DMN-derived measurements (mean  $z$ -scores, volume cluster-size and cluster-overlap). We found a statistically significant difference in the reproducibility consistency of all these measurements [mean  $z$ -scores:  $\chi^2(4) = 52$ ,  $P < 0.001$ ; cluster-size:  $\chi^2(4) = 54$ ,  $P < 0.001$ ; cluster-overlap:  $\chi^2(4) = 45$ ,  $P < 0.001$ ], depending on the rPNC method implemented (Fig. 7). For mean  $z$ -scores, pairwise comparisons revealed that the median (IQR) ICC values applying Tissue-based, 0.67 (0.04), and FIX-standard, 0.68 (0.04) was statistically different from NPC, 0.55 (0.02), ( $Z = -3.5$ ,  $P = 0.005$ ;  $Z = -4.2$ ,  $P < 0.001$ ; respectively). For cluster-

**TABLE II. Statistical testing for site-effects**

	Mean Z-score	Test-retest reliability			Cluster-overlap
		Cluster-size	Mean Z-score	Cluster-size	
NPC	$\chi^2(12, n = 129) = 29; P = 0.004$	$\chi^2(12, n = 129) = 42; P < 0.001$	$\chi^2(12, n = 64) = 13; P = 0.3$	$\chi^2(12, n = 64) = 19; P = 0.08$	$\chi^2(12, n = 64) = 23; P = 0.03$
Pestica	$\chi^2(12, n = 130) = 54; P < 0.001$	$\chi^2(12, n = 130) = 58; P < 0.001$	$\chi^2(12, n = 65) = 7; P = 0.9$	$\chi^2(12, n = 65) = 15; P = 0.2$	$\chi^2(12, n = 65) = 29; P = 0.004$
Tissue-based	$\chi^2(12, n = 130) = 28; P = 0.006$	$\chi^2(12, n = 130) = 42; P < 0.001$	$\chi^2(12, n = 65) = 20; P = 0.07$	$\chi^2(12, n = 65) = 13; P = 0.3$	$\chi^2(12, n = 65) = 21; P = 0.05$
Fix-standard	$\chi^2(12, n = 130) = 24; P = 0.02$	$\chi^2(12, n = 130) = 30; P = 0.003$	$\chi^2(12, n = 65) = 21; P = 0.06$	$\chi^2(12, n = 65) = 17; P = 0.1$	$\chi^2(12, n = 65) = 26; P = 0.01$
Fix-training	$\chi^2(12, n = 125) = 15; P = 0.2$	$\chi^2(12, n = 125) = 15; P = 0.3$	$\chi^2(12, n = 60) = 8; P = 0.8$	$\chi^2(12, n = 60) = 8; P = 0.8$	$\chi^2(12, n = 60) = 22; P = 0.04$

Kruskal-Wallis tests performed for each rPNC method (rows) and DMN-derived measurement (columns 1–2) and relative TRT reliability (columns 4–6) scores are here reported (*P*-values uncorrected). All DMN-derived measurements showed site-effects under NPC condition while FSL-FIX tended to reduce site effects; in particular FIX-training canceled site effects. In contrast, TRT reliability of functional connectivity and volume cluster-size did not show site effects. Some subjects did not show DMN functional clusters at the given threshold in the NPC (1 subject) and FIX-training (five subjects) conditions.

size, pairwise comparisons revealed that median (IQR) ICC values are significantly reduced applying all rPNC methods excepted for FIX-standard (PESTICA, 0.47 (0.04), Tissue-based, 0.54 (0.02), and FIX-training, 0.28 (0.03)) compared to NPC, 0.65 (0.04), ( $Z = 4.7, P < 0.001$ ;  $Z = 3.2, P < 0.01$ ;  $Z = 6.5, P < 0.001$ , respectively). For cluster-overlap, pairwise comparisons showed that median ICC values for FIX-training, 0.61 (0.02), were significantly reduced compared to NPC, 0.74 (0.02), ( $Z = 4.1, P < 0.001$ ).

## DISCUSSION

In the present study, we investigated the influence of some rPNC methods on the TRT reliability of intrinsic DMN connectivity measured in elderly participants across a consortium of 13 MRI sites. The main results are as follows: (1) The DMN at the single subject level was robustly identified using ICA methods in all rPNC conditions; (2) Relative to NPC, physiological noise correction methods significantly affected the median DMN z-score connectivity and cluster-size metrics, with FIX-based methods giving significantly higher values; (3) Within-site DMN reliability metrics were consistent across MRI sites and not significantly different across rPNC methods; (4) In particular, none of the physiological correction methods were able to significantly improve the TRT reliability relative to NPC; (5) Relative to NPC, FSL-FIX and tissue-based methods showed similar and significantly higher consistency of median DMN z-score reliability across the consortium.

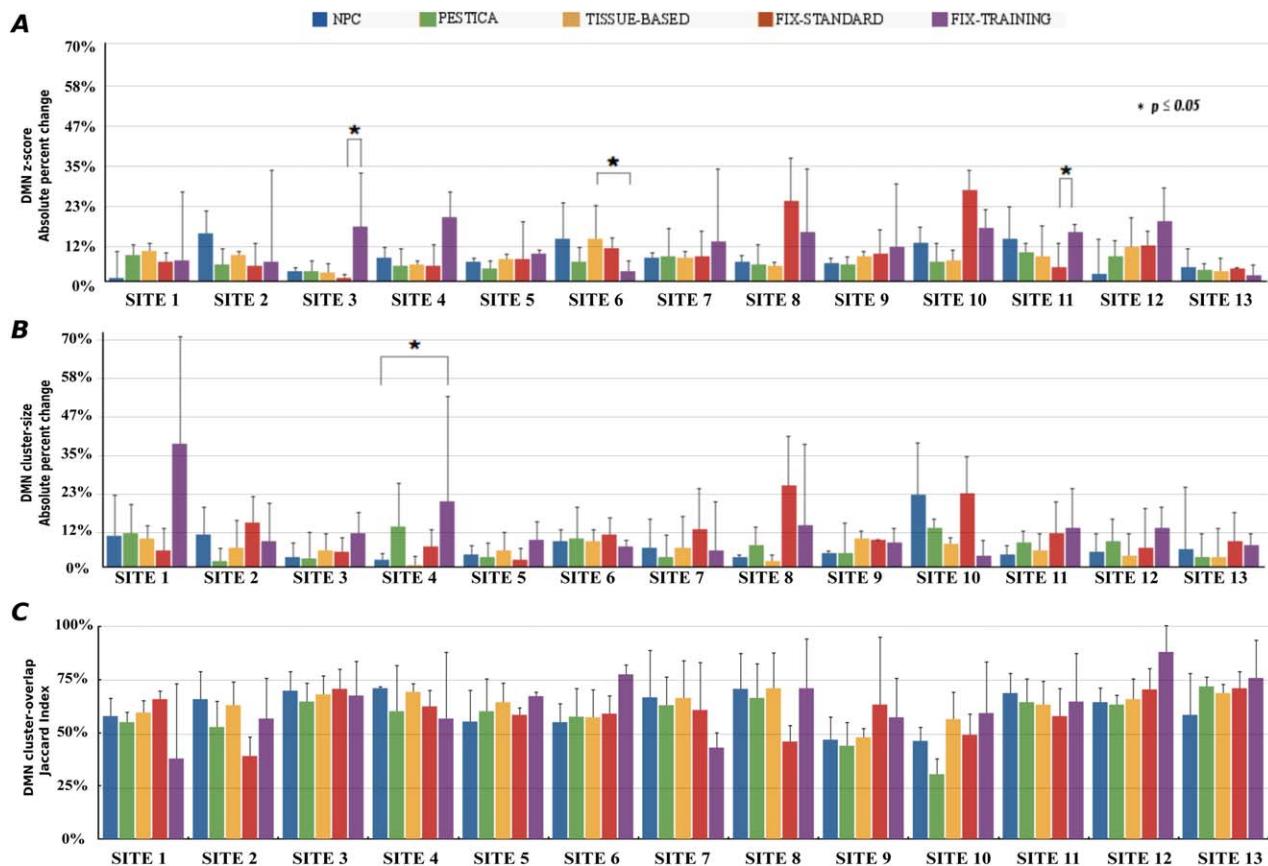
A recent study showed that, despite MRI acquisition harmonization efforts, this consortium showed high variability of temporal signal-to-noise ratio (tSNR) across sites [Jovicich et al., 2016]. In this study we extend those findings to show that, in spite of the high tSNR variability, the site-group DMN was able to be automatically selected for all rPNC methods, allowing robust DMN characterization at the single-subject level. This suggests that the rPNC methods evaluated here are valid candidates for longitudinal multisite resting-state fMRI studies with elderly participants. However, the overlap measure using a full DMN-template was not always successful at automatically detecting the DMN among the components at every site and rPNC condition. The accuracy of the automatic selection procedure improved only once frontal regions were excluded from the template. This was due to high variability of frontal DMN areas in terms of co-activation cluster-size (number of active voxels), spatial coordinate locations of active voxels in relation to the template, and the presence of large structured noise in these regions in other components. The FC variability in frontal regions might be a sign of age-related decline of mean z-scores in the DMN of healthy aging subjects [Damoiseaux et al. 2008, Huang et al., 2015].

An intrinsic challenge of characterizing resting-state networks with ICA methods is the possibility of network splitting [Abou-Elseoud et al., 2010], which can also affect

**TABLE III. Statistical testing for rPNC-effects**

Centers	DMN measurements		Test-retest reliability		
	Mean Z-score	Cluster-size	Mean Z-score	Cluster-size	Cluster-overlap
Site 1	$\chi^2(4) = 13; P = 0.1$	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 2; P = 1$	$\chi^2(4) = 5; P = 1$	$\chi^2(4) = 6; P = 1$
Site 2	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 14; P = 0.1$	$\chi^2(4) = 5; P = 1$	$\chi^2(4) = 6; P = 1$	$\chi^2(4) = 2; P = 1$
Site 3	$\chi^2(4) = 24; P < 0.01$	$\chi^2(4) = 22; P < 0.01$	$\chi^2(4) = 11; P = 0.4$	$\chi^2(4) = 3; P = 1$	$\chi^2(4) = 6; P = 1$
Site 4	$\chi^2(4) = 19; P = 0.01$	$\chi^2(4) = 17; P = 0.03$	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 16; P = 0.04$	$\chi^2(4) = 6; P = 1$
Site 5	$\chi^2(4) = 31; P < 0.01$	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 3; P = 1$	$\chi^2(4) = 1; P = 1$	$\chi^2(4) = 1.6; P = 1$
Site 6	$\chi^2(4) = 34; P < 0.01$	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 13; P = 0.1$	$\chi^2(4) = 31; P = 1$	$\chi^2(4) = 13; P = 0.2$
Site 7	$\chi^2(4) = 10; P = 0.5$	$\chi^2(4) = 3; P = 1$	$\chi^2(4) = 2; P = 1$	$\chi^2(4) = 4; P = 1$	$\chi^2(4) = 5; P = 1$
Site 8	$\chi^2(4) = 11; P = 0.4$	$\chi^2(4) = 1; P = 1$	$\chi^2(4) = 7; P = 1$	$\chi^2(4) = 8; P = 1$	$\chi^2(4) = 7; P = 1$
Site 9	$\chi^2(4) = 21; P < 0.01$	$\chi^2(4) = 9; P = 1$	$\chi^2(4) = 2; P = 1$	$\chi^2(4) = 3; P = 1$	$\chi^2(4) = 4; P = 1$
Site 10	$\chi^2(4) = 19; P = 0.01$	$\chi^2(4) = 29; P < 0.01$	$\chi^2(4) = 7; P = 1$	$\chi^2(4) = 9; P = 0.9$	$\chi^2(4) = 10; P = 0.5$
Site 11	$\chi^2(4) = 14; P = 0.1$	$\chi^2(4) = 13; P = 0.1$	$\chi^2(4) = 10; P = 0.4$	$\chi^2(4) = 7; P = 1$	$\chi^2(4) = 5; P = 1$
Site 12	$\chi^2(4) = 16; P = 0.05$	$\chi^2(4) = 5; P = 1$	$\chi^2(4) = 2; P = 1$	$\chi^2(4) = 5; P = 1$	$\chi^2(4) = 5; P = 1$
Site 13	$\chi^2(4) = 26; P < 0.01$	$\chi^2(4) = 16; P = 0.05$	$\chi^2(4) = 4; P = 1$	$\chi^2(4) = 10; P = 0.5$	$\chi^2(4) = 6; P = 1$
All sites	–	–	$\chi^2(4) = 14; P = 0.1$	$\chi^2(4) = 21; P < 0.01$	$\chi^2(4) = 7; P = 1$

Friedman tests performed for each site (rows) and DMN-derived measurement (columns 1–2) and relative TRT reliability (columns 4–6) scores are here reported (adjusted *P*-values across 13 sites using Bonferroni correction). Mean z-score and volume cluster-size showed significant rPNC effects within many sites. In contrast, the TRT reliability of these DMN-derived measurements showed rPNC effects only rarely. The test was also performed on the scores from the entire consortium (last row) in case no site-effects were found for these estimates (see Table II).

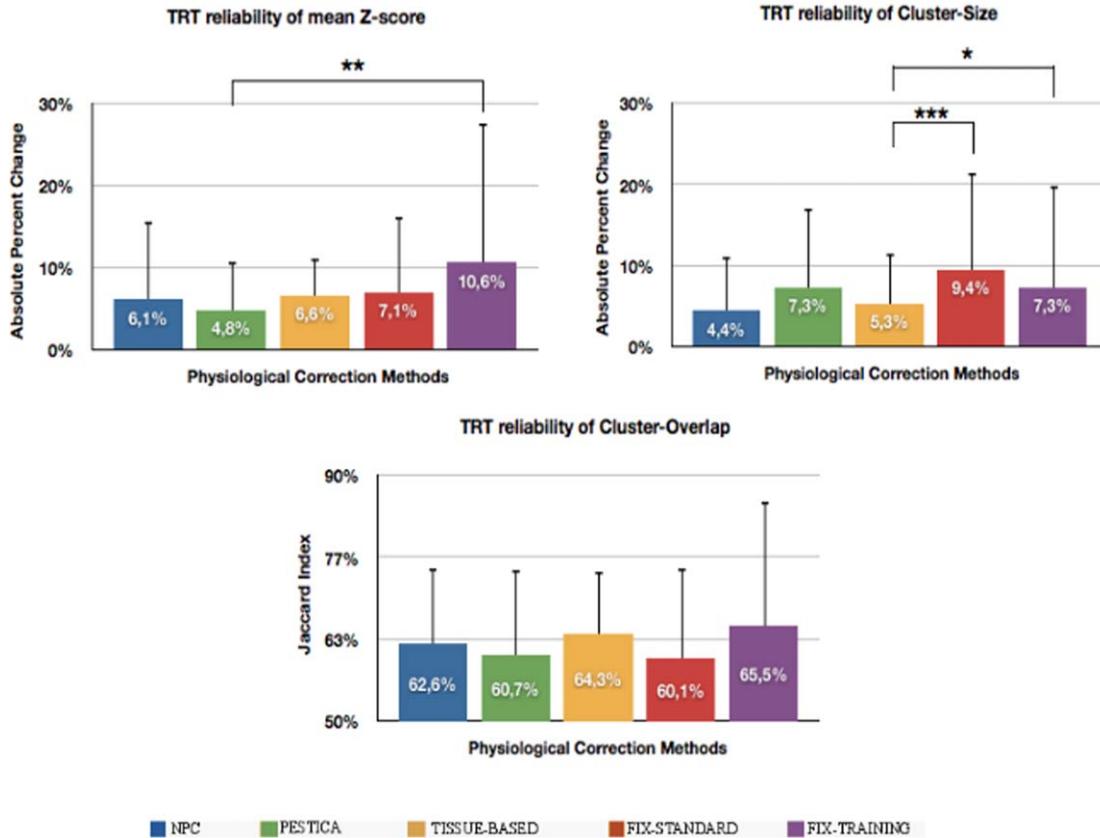


**Figure 5.**

Within-site test-retest reliability of DMN-derived measurements for each rPNC method. Median (IQR) across subjects for test-retest reliability scores in the DMN: (A) absolute percent change for functional connectivity; (B) Cluster-size error (0% highest reliability, 100% lowest reliability); (C) Jaccard index of DMN spatial overlap (0% lowest reliability, 100% highest reliability). High test-retest reliability was found for any rPNC method

and DMN-derived measurement in all sites. rPNC effects were evaluated within each site using Friedman test ( $P < 0.05$ , corrected with Dunn-Bonferroni method). No systematic rPNC effects were found across the different sites. Statistically significant differences between rPNC methods were rarely found. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

\*  $p \leq 0.05$    \*\*  $p \leq 0.01$    \*\*\*  $p \leq 0.001$



**Figure 6.**

Overall test-retest reliability of DMN-derived measurements as function of rPNC method. Consortium median (IQR) of TRT reliability of each DMN-derived measurement: mean z-score (left), cluster-size volume (right), overlap-size (bottom). For each DMN-derived measurement, rPNC effects were evaluated using Fried-

man test ( $P < 0.05$ , corrected with Dunn-Bonferroni method). We found that no rPNC method statistically reduced the test-retest % error of FC and cluster-size nor increased the Jaccard index for the cluster-overlap in the DMN. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

reproducibility [Zuo and Xing 2014]. In this study, the DMN splitting phenomenon was addressed in several ways: first, we kept the number of ICs low and constant for all sites [Abou-Elseoud et al., 2010]. Second, we used a supervised algorithm to select the group component with maximum overlap to a DMN reference template and we visually inspected the group DMN results. As with most studies, it is possible that some components of the DMN were not properly included in our estimations, and instead were distributed through one or more of the other rejected components. However, since the rPNC methods were applied at the single-subject level before group ICA, after which all rPNC methods followed a similar workflow, we expected that comparison of rPNC methods would not be strongly biased by DMN splitting effects.

We found that, among the rPNC methods, FSL-FIX increased mean z-scores in the DMN with no impact on the DMN cluster-size in many sites of the consortium.

This finding is potentially due to the fact that FSL-FIX removes a larger proportion of signal-of-no-interest from the data than the other rPNC methods (such as residual motion artifacts), leading to an increase in mean z-scores within the DMN. This effect was found also in other resting-state networks. This suggests that, the application of FSL-FIX could increase the sensitivity of group ICA to detect the DMN in healthy aging subjects and discern abnormal connectivity patterns in clinical aging, a task known not to be straightforward [Bai et al., 2008; Koch, 2012]. Importantly, training the classifier at each site improved MRI site consistency in DMN-derived measurements and confined individual spatial DMN maps within converging spatial locations across MRI sites. Relative to no physiological correction, FSL-FIX was found to increase the number of active voxels in gray matter areas outside the DMN nodes, together with surrounding white matter and CSF voxels. This finding stresses the importance of

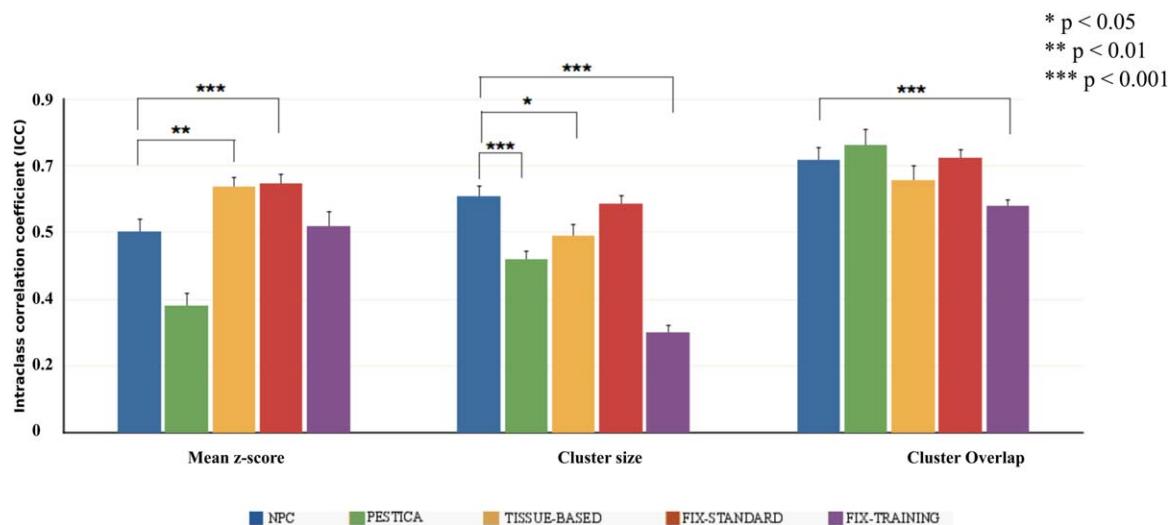


Figure 7.

Reproducibility consistency of DMN-derived measurements across sites. Reproducibility consistency measured via ICC (mean  $\pm$  standard deviation) is here shown for each DMN-derived measurement (bar groups) and rPNC method (color bars). For each DMN-derived measurement, rPNC effects were evaluated using Friedman test ( $P < 0.05$ , corrected with Dunn-

Bonferroni method). Tissue-based and FIX-standard increase the reproducibility consistency of TRT reliability of average z-score reliability (left) with respect to NPC; in contrast, FIX-training increases variability of reliability estimates across different sites. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

the selection of nodes that will be defined as members of the DMN.

These findings are consistent with previous works showing that DMN connectivity estimates strongly depend on how thoroughly physiological and other confounds are removed upstream [Beall and Lowe, 2010; Birn, 2012; Murphy et al., 2013]. Of note, in this work we only considered ICA (with a specific number of components) to characterize the DMN. This choice was mainly motivated by the high DMN reliability obtained via ICA in our previous work [Jovicich et al., 2016].

The TRT reliability of DMN connectivity metrics was consistent across all MRI sites for all rPNC methods, which allowed the averaging across sites to estimate consortium reliability metrics. Absolute percent errors were in the range of 5–11% for DMN z-score and cluster-size reliability. Jaccard indexes were in the range 60–65% for DMN cluster-overlap reliability. Only for FSL-FIX an inverse correlation between motion and Jaccard indexes was found. This dependence might originate from the additional cleaning of residual motion-related spatial maps implemented by FSL-FIX, which leads to higher characterization of subject-specific DMN maps when the overall motion detected is lower.

These results are consistent with similar measures recently reported in a single-site study with young healthy participants [Meindl et al., 2010]. In the present study with elderly participants, we found high reliability of DMN cluster-size and moderate DMN cluster-overlap, consistent

across sites, albeit not influenced by rPNC methods. The statistical evaluations of MRI site effects in TRT reliability scores were performed using both parametric and non-parametric approaches: both tests agreed in the lack of MRI site effects for DMN mean z-scores and cluster-size reliability. DMN cluster-overlap reliability results differed across tests, giving no MRI site effects with the non-parametric test but significant MRI site effects with the 2-way ANOVA. Given the low number of subjects it is difficult to interpret this difference since the non-parametric test is robust to outlier but underpowered with low samples while the 2-way ANOVA has the opposite characteristics.

In particular, no rPNC correction method showed significant DMN reliability improvements relative to NPC. These findings are in agreement with those from a recent single-site study [Birn et al., 2014] that found that neither independent physiological measures (cardiac and respiratory measures) nor Tissue-based correction improved the TRT reliability of mean z-scores among ROIs, including the DMN in young adults. Our study extends those results in several ways: population age, multicentric MRI consortium and rPNC methods.

To the best of our knowledge, this is the first study that investigates the influence of some rPNC methods on the consistency of DMN connectivity reliability in a multisite study. In fact, although DMN reliability is consistent across sites, we found that rPNC methods can still influence the degree of reproducibility consistency across the consortium. When considering mean z-scores in the DMN,

Tissue-based and FIX-standard correction methods significantly improved the reproducibility consistency of the DMN amongst the different 3T MRI sites. This effect may be driven by exploiting individual anatomical information (T1-anatomical) to remove physiological noise from non-gray matter tissue. This suggests that the use of single-site derived spatial priors in PESTICA may not be sufficient to contrast multisite heterogeneity. When considering reproducibility consistency, FIX-training did not perform as well as FIX-standard. This may be related to the deviation of non-brain related percent variance classified between the two methods (FIX-training retained higher signal variance, very likely not to be associated to the DMN). Furthermore, such variability was unequally distributed across sites, leading to overall lower reproducibility consistency of the DMN relative to FIX-standard. Therefore, we recommend adopting a higher signal-noise threshold for FSL-FIX in longitudinal multisite resting-state FC studies.

## CONCLUSIONS

In this study, we found that the application of rPNC allows the consistent characterization of single-subject DMN in elderly participants despite the tSNR variability present in a multicentric consortium. FSL-FIX tended to increase mean z-scores in the DMN with no impact on the cumulative cluster-size. We found that rPNC methods do not significantly influence the TRT reliability of mean z-scores, cluster-size and cluster-overlap in the DMN. In particular, none of the correction methods showed a consistent improvement of reliability relative to performing no correction. However, we found that Tissue-based and FSL-FIX significantly reduce variability of TRT reliability scores of mean z-scores in the DMN in this multicentric study. Overall, these findings support the application of Tissue-based or FSL-FIX physiological noise correction methods in multisite longitudinal resting-state fMRI studies.

## ACKNOWLEDGMENTS

Authors are grateful to all members and collaborators of the PharmaCog project and particularly to Luca Venturi, Genoveffa Borsci, Thomas Günther and Aurélien Monnet for their contribution in the start-up phase. Ludovico Minati receives employment funding from Fondazione IRCCS Istituto Neurologico Carlo Besta (Milano, Italy) and Scienze Mente-Cervello (Rovereto, Italy).

## REFERENCES

Abou-Elseoud A, Starck T, Remes J, Nikkinen J, Tervonen O, Kiviniemi V (2010): The effect of model order selection in group PICA. *Hum Brain Mapp* 31:1207–1216.

Bai F, Zhang Z, Yu H, Shi Y, Yuan Y, Zhu W, Zhang X, Qian Y (2008): Default-mode network activity distinguishes amnestic type mild cognitive impairment from healthy aging: A combined structural and resting-state functional MRI study. *Neurosci Lett* 438:111–115.

Beall EB (2010): Adaptive cyclic physiologic noise modeling and correction in functional MRI. *J Neurosci Methods* 187:216–228.

Beall EB, Lowe MJ (2007): Isolating physiologic noise sources with independently determined spatial measures. *NeuroImage* 37: 1286–1300.

Beall EB, Lowe MJ (2010): The non-separability of physiologic noise in functional connectivity MRI with spatial ICA at 3T. *Journal of Neuroscience Methods* 191:263–276.

Beckmann CF, Smith SM (2004): Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Trans Med Imag* 23:137–152.

Beckmann M, Filippini S (2009): Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *NeuroImage* 47:S148

Behzadi et al. (2007): A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37:90–101.

Bennett CM, Miller MB (2010): How reliable are the results from functional magnetic resonance imaging? *Ann N Y Acad Sci* 1191:133–155.

Birn RM (2012): The role of physiological noise in resting-state functional connectivity. *NeuroImage* 62:864–870. Aug 15;

Birn RM, Cornejo MD, Molloy EK, Patriat R, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V (2014): The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity. *Brain Connect* 4:511–522.

Birn RM, Murphy K, Handwerker DA, Bandettini PA (2009): fMRI in the presence of task-correlated breathing variations. *NeuroImage* 47:1092–1104.

Birn RM, Murphy K, Bandettini PA (2008): The effect of respiration variations on independent component analysis results of resting state functional connectivity. *Hum Brain Mapp* 29:740–750. Jul;

Birn RM, Diamond JB, Smith MA, Bandettini PA (2006): Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage* 31:1536–1515..

Bright MG, Murphy K (2013): Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *NeuroImage* 64:526–537.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008): The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 1124:1–38.

Buckner RL (2012): The serendipitous discovery of the brain's default network. *NeuroImage* 62:1137–1145.

Bulmer MG. (1997). Principles of statistics. p. 63.

Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013): Clinical applications of the functional connectome. *NeuroImage* 80:527–540.

Churchill & Strother (2013): "PHYCAA+: an optimized, adaptive procedure for measuring and controlling physiological noise in BOLD fMRI." *NeuroImage* 82: 306–325.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comp Biomed Res Int J* 29:162–173.

Damoiseaux JS, Beckmann CF, Arigita EJ, Barkhof F, Scheltens P, Stam CJ, Smith SM, Rombouts SA (2008): Reduced resting-state brain activity in the "default network" in normal aging. *Cereb Cortex* 18:1856–1864.

Dunn OJ (1964): Multiple comparisons using rank sums. *Technometrics* 6:241–252.

Franco AR, Pritchard A, Calhoun VD, Mayer AR (2009): Interrater and intermethod reliability of default mode network selection. *Hum Brain Mapp* 30:2293–2303.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2014): ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* 15:232–247.

Hallquist MN, Hwang K, Luna B (2013): The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *NeuroImage* 82:208–225.

Hedden T, Van Dijk KRa, Becker JA, Mehta A, Sperling Ra, Johnson Ka, Buckner RL (2009): Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *J Neurosci* 29:12686–12694.

Huang CC, Hsieh WJ, Lee PL, Peng LN, Liu LK, Lee WJ, Huang JK, Chen LK, Lin CP (2015): Age-related changes in resting-state networks of a large sample size of healthy elderly. *CNS Neurosci Ther* 21:817–825.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012): FSL. *NeuroImage* 62:782–790.

Jenkinson M, Bannister P, Brady JM, Smith SM (2002): Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17:825–841.

Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW (2010): Mapping sources of correlation in resting state fMRI, with artifact detection and removal. *Neuroimage* 52(2):571–582.

Jovicich J, Marizzoni M, Bosch B, Bartrés-Faz D, Arnold J, Benninghoff J, Wiltfang J, Roccatagliata L, Picco A, Nobili F, Blin O, Bombois S, Lopes R, Bordet R, Chanoine V, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargalló N, Ferretti A, Caulo M, Aiello M, Cavalieri C, Soricelli A, Salvadori N, Tarducci R, Floridi P, Tsolaki M, Constantinidis M, Drevelegas A, Rossini PM, Marra C, Hoffmann KT, Hensch T, Schönknecht P, Kuij JP, Visser PJ, Barkhof F, Bordet R, Frisoni GB, Jovicich J; PharmaCog Consortium. (2015). Longitudinal reproducibility of automatically segmented hippocampal subfields: A multisite European 3T study on healthy elderly. *Hum Brain Mapp* 36:3516–3527.

Mat Roni S. (2014). Introduction to SPSS, School of Business, Edith Cowan University, Australia. Available at: [www.researchgate.net/publication/262151892\\_Introduction\\_to\\_SPSS](http://www.researchgate.net/publication/262151892_Introduction_to_SPSS).

Meindl T, Teipel S, Elmouden R, Mueller S, Koch W, Dietrich O, Coates U, Reiser M, Glaser C (2010): Test-retest reproducibility of the default-mode network in healthy individuals. *Hum Brain Mapp* 31:237–246.

Murphy K, Birn RM, Bandettini PA (2013): Resting-state fMRI confounds and cleanup. *NeuroImage* 80:349–359.

Nicolini P, Ciulla MM, De Asmundis C, Magrini F, Brugada P (2012): The prognostic value of heart rate variability in the elderly, changing the perspective: from sympathovagal balance to chaos theory. *Pacing Clin Electrophysiol* 35:622–638.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59:2142–2154.

Raichle ME, Snyder AZ (2007): A default mode of brain function: A brief history of an evolving idea. *NeuroImage* 37:1083–1090.

Rosazza C, Minati L, Ghielmetti F, Mandelli ML, Bruzzone MG (2012): Functional connectivity during resting-state functional MR imaging: study of the correspondence between independent component analysis and region-of-interest-based methods. *Am J Neuroradiol* 33:180–187.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014): Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* 90:449–468.

Särkkä S, Solin A, Nummenmaa A, Vehtari A, Auranen T, Vanni S, Lin FH (2012): Dynamic retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. *NeuroImage* 60(2): 1517–1527.

studies: A comparison of cross-sectional and longitudinal segmentations. *Neuroimage* 83:472–484.

Koch W, Teipel S, Mueller S, Benninghoff J, Wagner M, Bokde AL, Hampel H, Coates U, Reiser M, Meindl T (2012): Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. *Neurobiol Aging* 33:466–478.

Li X, Li TQ, Andreasen N, Wiberg MK, Westman E, Wahlund LO (2013): Ratio of A $\beta$ 42/P-tau181p in CSF is associated with aberrant default mode network in AD. *Sci Rep* 3:1339.

Liao XH, Xia MR, Xu T, Dai ZJ, Cao XY, Niu HJ, Zuo XN, Zang YF, He Y (2013): Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. *NeuroImage* 83:969–982.

Lund TE (2001): fcMRI-mapping functional connectivity or correlating cardiac-induced noise? *Magn Reson Med* 46:628–629.

Maitra R (2010): A re-defined and generalized percent-overlap-of-activation measure for studies of fMRI reproducibility and its use in identifying outlier activation maps. *NeuroImage* 50:124–135.

Marizzoni M, Antelmi L, Bosch B, Bartrés-Faz D, Müller BW, Wiltfang J, Fiedler U, Roccatagliata L, Picco A, Nobili F, Blin O, Bombois S, Lopes R, Sein J, Ranjeva JP, Didic M, Gros-Dagnac H, Payoux P, Zoccatelli G, Alessandrini F, Beltramello A, Bargalló N, Ferretti A, Caulo M, Aiello M, Cavalieri C, Soricelli A, Salvadori N, Parnetti L, Tarducci R, Floridi P, Tsolaki M, Constantinidis M, Drevelegas A, Rossini PM, Marra C, Hoffmann KT, Hensch T, Schönknecht P, Kuij JP, Visser PJ, Barkhof F, Bordet R, Frisoni GB, Jovicich J; PharmaCog Consortium. (2015). Longitudinal reproducibility of automatically segmented hippocampal subfields: A multisite European 3T study on healthy elderly. *Hum Brain Mapp* 36:3516–3527.

Schulz S, Adochie FC, Edu IR, Schroeder R, Costin H, Bär KJ, Voss A (2013): Cardiovascular and cardiorespiratory coupling analyses: A review. *Philos Trans A Math Phys Eng Sci* 371:20120191

Soldati N, Robinson S, Persello C, Jovicich J, Bruzzone L (2009): Automatic classification of brain resting states using fMRI temporal signals. *Electron Lett* 45:19

Taylor BJ, Johnson BD (2010): The pulmonary circulation and exercise responses in the elderly. *Semin Respir Crit Care Med* 31:528–538.

Van Dijk KRA, Sabuncu MR, Buckner RL (2012): The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* 59:431–438.

Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010): Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol* 103:297–321.

Vannini P, Hedden T, Becker JA, Sullivan C, Putcha D, Rentz D, Johnson KA, Sperling RA (2012): Age and amyloid-related alterations in default network habituation to stimulus repetition. *Neurobiol Aging* 33:1237–1252.

Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C (2009): Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. *NeuroImage* 47:1408–1416.

Zuo XN, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, Breitner JC, Buckner RL, Calhoun VD, Castellanos FX, Chen A, Chen B, Chen J, Chen X, Colcombe SJ, Courtney W, Craddock RC, Di Martino A, Dong HM, Fu X, Gong Q, Gorgolewski KJ, Han Y, He Y, He Y, Ho E, Holmes A, Hou XH, Huckins J, Jiang T, Jiang Y, Kelley W, Kelly C, King M, LaConte SM, Lainhart JE, Lei X, Li HJ, Li K, Li K, Lin Q, Liu D, Liu J, Liu X, Liu Y, Lu G, Lu J, Luna B, Luo J, Lurie D, Mao Y, Margulies DS, Mayer AR, Meindl T, Meyerand ME, Nan W, Nielsen JA, O'Connor D, Paulsen D, Prabhakaran V, Qi Z, Qiu J, Shao C, Shehzad Z, Tang W, Villringer A, Wang H, Wang K, Wei D, Wei GX, Weng XC, Wu X, Xu T, Yang N, Yang Z, Zang YF, Zhang L, Zhang Q, Zhang Z, Zhang Z, Zhao K, Zhen Z, Zhou Y, Zhu XT, Milham MP (2014): An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data* 1:9140049

Zuo XN, Kelly C, Adelstein JS, Klein DF, Castellanos FX, Milham MP (2010): Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. *NeuroImage* 49:2163–2177.

Zuo XN, Xing XX (2014): Test-retest reliabilities of resting-state fMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci Biobehav Rev*.

Zuo XN, Xu T, Jiang L, Yang Z, Cao XY, He Y, Zang YF, Castellanos FX, Milham MP (2013): Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *NeuroImage* 65:374–386.