NeuroImage 134 (2016) 281–294

Contents lists available at ScienceDirect

NeuroImage

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

Power estimation for non-standardized multisite studies

Anisha Keshavan [a,b,*], Friedemann Paul [c,af], Mona K. Beyer [d], Alyssa H. Zhu [a], Nico Papinutto [a], Russell T. Shinohara [e], William Stern [a], Michael Amann [f,r], Rohit Bakshi [g], Antje Bischof [a,f,h], Alessandro Carriero [i], Manuel Comabella [j], Jason C. Crane [k], Sandra D'Alfonso [l], Philippe Demaerel [m], Benedicte Dubois [n], Massimo Filippi [o], Vinzenz Fleischer [p], Bertrand Fontaine [q], Laura Gaetano [f,r], An Goris [n], Christiane Graetz [p], Adriane Gröger [p], Sergiu Groppa [p], David A. Hafler [s], Hanne F. Harbo [t], Bernhard Hemmer [u,v], Kesshi Jordan [a,b], Ludwig Kappos [f], Gina Kirkish [k], Sara Llufriu [w], Stefano Magon [f,r], Filippo Martinelli-Boneschi [o], Jacob L. McCauley [x], Xavier Montalban [j], Mark Mühlau [u,y], Daniel Pelletier [s], Pradip M. Pattany [ag], Margaret Pericak-Vance [x], Isabelle Cournu-Rebeix [q], Maria A. Rocca [o], Alex Rovira [j], Regina Schlaeger [a,f,h], Albert Saiz [w], Till Sprenger [f,z], Alessandro Stecco [aa], Bernard M.J. Uitdehaag [ab], Pablo Villoslada [a,w], Mike P. Wattjes [ab], Howard Weiner [g], Jens Wuerfel [c,ac], Claus Zimmer [ad], Frauke Zipp [p], International Multiple Sclerosis Genetics Consortium [ae], Stephen L. Hauser [a], Jorge R. Oksenberg [a], Roland G. Henry [a,b,k]

[a] Department of Neurology, University of California, San Francisco, CA, USA
[b] UC Berkeley—UCSF Graduate Program in Bioengineering, San Francisco, CA, USA
[c] NeuroCure Clinical Research Center and Clinical and Experimental Multiple Sclerosis Research Center, Department of Neurology, Charité University Medicine Berlin, Berlin, Germany
[d] Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
[e] Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
[f] Department of Neurology, Basel University Hospital, University of Basel, Basel, Switzerland
[g] Brigham and Women's Hospital, MA, United States
[h] Clinical Immunology, University Hospital Basel,University of Basel, Basel, Switzerland
[i] Department of Translational Medicine, Department of Radiology, UPO University, Via Solaroli 17, 28100 Novara, Italy
[j] Hospital Universitari Vall d'Hebron, Barcelona, Spain
[k] Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA
[l] Department of Health Sciences, UPO University, Novara, Italy
[m] Department of Radiology, University Hospitals Leuven, Leuven, Belgium
[n] KU Leuven—University of Leuven, Department of Neurosciences, Leuven, Belgium
[o] Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy
[p] Department of Neurology, Focus Program Translational Neuroscience (FTN) and Immunotherapy (FZI), Rhine-Main Neuroscience Network (rmn2), University Medical Centre of the Johannes Gutenberg University Mainz, Germany
[q] Hôpital Pitié-Salpêtrière, ICM, UPMC 06 UM 75, INSERM U 1127, CNRS UMR 7225, IHU-A-ICM, AP-HP: Pôle des maladies du système nerveux, 47 boulevard de l'Hôpital, 75013 Paris, France
[r] Medical Image Analysis Center (MIAC AG), Basel, Switzerland
[s] Departments of Neurology and Immunobiology, Yale School of Medicine, CT, USA
[t] Department of Neurology, Oslo University Hospital and University of Oslo, Oslo, Norway
[u] Dept. Neurology of the Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[v] Munich Cluster of Systems Neurology (SyNery), Germany
[w] Center for Neuroimmunology, Hospital Clinic Barcelona, IDIBAPS, Barcelona, Spain
[x] John P. Hussman Institute for Human Genomics and the Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, USA
[y] TUM–Neuroimaging Center, Technische Universität München, Munich, Germany

* Corresponding author at: Department of Neurology, University of California, San Francisco, USA.
E-mail addresses: anisha.keshavan@ucsf.edu (A. Keshavan), Friedemann.Paul@charite.de (F. Paul), mona.beyer@lyse.net (M.K. Beyer), Alyssa.Zhu@ucsf.edu (A.H. Zhu), Nico.Papinutto@ucsf.edu (N. Papinutto), rshi@mail.med.upenn.edu (R.T. Shinohara), William.Stern@ucsf.edu (W. Stern), michael.amann@usb.ch (M. Amann), rbakshi@post.harvard.edu (R. Bakshi), antje.Bischof@ucsf.edu (A. Bischof), alessandro.carriero@med.unipmn.it (A. Carriero), manuel.comabella@vhir.org (M. Comabella), Jason.Crane@ucsf.edu (J.C. Crane), sandra.dalfonso@med.uniupo.it (S. D'Alfonso), Philippe.Demaerel@uzleuven.be (P. Demaerel), Benedicte.dubois@uzleuven.be (B. Dubois), zonca.lucia@hsr.it (M. Filippi), Vinzenz.Fleischer@unimedizin-mainz.de (V. Fleischer), bertrand.fontaine@upmc.fr (B. Fontaine), laura.gaetanosf@gmail.com (L. Gaetano), An.Goris@med.kuleuven.be (A. Goris), Christiane.Graetz@unimedizin-mainz.de (C. Graetz), adriane.groeger@unimedizin-mainz.de (A. Gröger), Sergiu.Groppa@unimedizin-mainz.de (S. Groppa), david.hafler@yale.edu (D.A. Hafler), h.f.harbo@medisin.uio.no (H.F. Harbo), hemmer@tum.de (B. Hemmer), Kesshi.Jordan@ucsf.edu (K. Jordan), Ludwig.Kappos@usb.ch (L. Kappos), Gina.Kirkish@ucsf.edu (G. Kirkish), SLLUFRIU@clinic.ub.es (S. Llufriu), Stefano.Magon@usb.ch (S. Magon), martinelli.filippo@hsr.it (F. Martinelli-Boneschi), jmccauley@med.miami.edu (J.L. McCauley), xavier.montalban@cem-cat.org (X. Montalban), muehlau@lrz.tu-muenchen.de (M. Mühlau), daniel.pelletier@yale.edu (D. Pelletier), PPattany@med.miami.edu (P.M. Pattany), MPericak@med.miami.edu (M. Pericak-Vance), isabelle.rebeix@upmc.fr (I. Cournu-Rebeix), rocca.mara@hsr.it (M.A. Rocca), alex.rovira@idi.gencat.cat (A. Rovira), regina.schlaeger@ucsf.edu (R. Schlaeger), ASAIZ@clinic.ub.es (A. Saiz), Till.sprenger@usb.ch (T. Sprenger), a.stecco@libero.it (A. Stecco), bmj.uitdehaag@vumc.nl (B.M.J. Uitdehaag), Pablo.VilloslaDiaz@ucsf.edu (P. Villoslada), m.wattjes@vumc.nl (M.P. Wattjes), hweiner@rics.bwh.harvard.edu (H. Weiner), jw@miac.ch (J. Wuerfel), claus.zimmer@tum.de (C. Zimmer), Frauke.zipp@unimedizin-mainz.de (F. Zipp), Stephen.Hauser@ucsf.edu (S.L. Hauser), Jorge.Oksenberg@ucsf.edu (J.R. Oksenberg), Roland.Henry@ucsf.edu (R.G. Henry).

http://dx.doi.org/10.1016/j.neuroimage.2016.03.051

[z] DKD Helios Klinik Wiesbaden, Wiesbaden, Germany
[aa] Section of Neuroradiology, Department of Radiology, Maggiore Hospital, Corso Mazzini 18, 28100, Novara, Italy
[ab] MS Center Amsterdam, VU University Medical Center Amsterdam, The Netherlands
[ac] Medical Image Analysis Center, Basel, Switzerland
[ad] Dept. Neuroradiology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[ae] International Multiple Sclerosis Genetics Consortium, USA, EU, AU
[af] Experimental and Clinical Research Center, Max Delbrueck Center for Molecular Medicine and Charité University Medicine Berlin, Berlin, Germany
[ag] Department of Radiology, University of Miami Miller School of Medicine, Miami, FL, USA

ARTICLE INFO

Article history:
Received 9 December 2015
Accepted 21 March 2016
Available online 1 April 2016

ABSTRACT

A concern for researchers planning multisite studies is that scanner and T1-weighted sequence-related biases on regional volumes could overshadow true effects, especially for studies with a heterogeneous set of scanners and sequences. Current approaches attempt to harmonize data by standardizing hardware, pulse sequences, and protocols, or by calibrating across sites using phantom-based corrections to ensure the same raw image intensities. We propose to avoid harmonization and phantom-based correction entirely. We hypothesized that the bias of estimated regional volumes is scaled between sites due to the contrast and gradient distortion differences between scanners and sequences. Given this assumption, we provide a new statistical framework and derive a power equation to define inclusion criteria for a set of sites based on the variability of their scaling factors. We estimated the scaling factors of 20 scanners with heterogeneous hardware and sequence parameters by scanning a single set of 12 subjects at sites across the United States and Europe. Regional volumes and their scaling factors were estimated for each site using Freesurfer's segmentation algorithm and ordinary least squares, respectively. The scaling factors were validated by comparing the theoretical and simulated power curves, performing a leave-one-out calibration of regional volumes, and evaluating the absolute agreement of all regional volumes between sites before and after calibration. Using our derived power equation, we were able to define the conditions under which harmonization is not necessary to achieve 80% power. This approach can inform choice of processing pipelines and outcome metrics for multisite studies based on scaling factor variability across sites, enabling collaboration between clinical and research institutions.

Introduction

The pooled or meta-analysis of regional brain volumes derived from T1-weighted MRI data across multiple sites is reliable when data is acquired with similar acquisition parameters (Cannon et al., 2014; Ewers et al., 2006; Jovicich et al., 2006). The inherent scanner- and sequence-related noise of MRI volumetrics under heterogeneous acquisition parameters has prompted many groups to standardize protocols across imaging sites (Boccardi et al., 2013; Cannon et al., 2014; Weiner et al., 2012). However, standardization across multiple sites can be prohibitively expensive and requires a significant effort to implement and maintain. At the other end of the spectrum, multisite studies without standardization can also be successful, albeit with extremely large sample sizes. The ENIGMA consortium, for example, combined scans of over 10,000 subjects from 25 sites with varying field strengths, scanner makes, acquisition protocols, and processing pipelines. The unusually large sample size enabled this consortium to provide robust phenotypic traits despite the variability of non-standardized MRI volumetrics and the power required to run a genome wide association study (GWAS) to identify modest effect sizes (Thompson et al., 2014). These studies raise the following question: Is there a middle ground between fully standardizing a set of MRI scanners and recruiting thousands of subjects across a large number of sites? Eliminating the harmonization requirement for a multisite study would facilitate inclusion of retrospectively acquired data and data from sites with ongoing longitudinal studies that would not want to adjust their acquisition parameters.

Towards this goal, there is a large body of literature addressing the correction of geometric distortions that result from gradient nonlinearities. These corrections fall into two main categories: phantom-based deformation field estimation and direct magnetic field gradient measurement-based deformation estimation, the latter of which requires extra hardware and spherical harmonic information from the manufacturer (Fonov et al., 2010). Calibration phantoms, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Gunter et al., 2009) and LEGO® (Caramanos et al., 2010), have been used by large multisite studies to correct for these geometric distortions, which also affect regional volume measurements. These studies have outlined various correction methods that significantly improve deformation field similarity between scanners. However, the relationship between the severity of gradient distortion and its effect on regional volumes, in particular, remains unclear. In the case of heterogeneous acquisitions, correction becomes especially difficult due to additional noise sources. Gradient hardware differences across sites are compounded with contrast variation due to sequence parameter changes. In order to properly evaluate the reproducibility of brain segmentation algorithms, these phantoms should resemble the human brain in size, shape, and tissue distribution. Droby and colleagues evaluated the stability of a post-mortem brain phantom and found similar reproducibility of volumetric measurements to that of a healthy control (Droby et al., 2015). In this study, we propose to measure between-site bias through direct calibration of regional volumes by imaging 12 common healthy controls at each site. Quantifying regional bias allows us to overcome between-site variability by increasing sample size to an optimal amount, rather than employing a phantom-based voxel-wise calibration scheme that corrects both contrast differences and geometric distortions.

We hypothesized that all differences in regional contrast and geometric distortion result in regional volumes that are consistently and linearly scaled from their true value. For a given region of interest (ROI), two mechanisms simultaneously impact the final boundary definition: (1) gradient nonlinearities cause distortion and (2) hardware (including scanner, field strength, and coils) and acquisition parameters modulate tissue contrast. Based on the results of Tardiff and colleagues, who found that contrast-to-noise ratio and contrast inhomogeneity from various pulse sequences and scanner strengths cause regional biases in VBM (Tardif et al., 2009, 2010), we hypothesized that each ROI will scale differently a teach site. Evidence for this scaling property can also be seen in the overall increase of gray matter volume and decrease of white matter volume of the ADNI-2 compared to the ADNI-1