



Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2023 February ; 12464: . doi:10.1117/12.2653903.

Harmonization of repetition time and scanner effects on estimates of brain hemodynamic response function

Lucie Dole^a, Kurt G. Schilling^{b,c}, Hakmook Kang^d, John C. Gore^{b,c,e}, Bennett A. Landman^{a,b}

^aDepartment of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA;

^bVanderbilt University Institute of Imaging Science, Nashville, TN, USA;

^cDepartment of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN, USA;

^dDepartment of Biostatistics, Vanderbilt University, Nashville, TN, USA;

^eDepartment of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

Abstract

Multisite contributions are essential to improve the reliability and statistical power of imaging studies but introduce a complexity because of different acquisition protocols and scanners. The hemodynamic response function (HRF) is the transform that relates neural activity to the measured blood oxygenation level-dependent (BOLD) signal in MRI and contains information about the latency, amplitude, and duration of neuronal activations. Acquisition variabilities, without adding harmonization techniques, can severely limit our ability to characterize spatial effects. To address this problem, we propose to study and remove variabilities of the sampling rate and scanners on estimates of the HRF. We computed the HRF using a blind deconvolution method in 547 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) across 62 sites and 18 scanners. The approach consists of studying the changes of the response according to repetition times (TR) and scanner models. We applied ComBAT, a statistical multi-site harmonization technique, to evaluate and reduce the scanner and repetition time effects and used the Wilcoxon rank sum test to assess the performance of the harmonization. Results show high scanner and repetition time variabilities ($|d| \geq 0.38$, $p = 4.5 \times 10^{-5}$) across features, indicating that using harmonization is crucial in multi-site studies. ComBAT successfully removes the sampling effects and reduces the variance between scanners for 7 out of 10 of the HRF features ($|d| \leq 0.05$, $p = 0.0052$). Scanners effects have been characterized on multi-site datasets, but the repetition time impact has been less studied. We showed that the use of different values of repetition time leads to changes in HRF behavior. Regression modeling changes in the HRF on the harmonized data are not significant ($p = 0.0401$) which does not allow to conclude how HRF changes with aging.

Keywords

Hemodynamic Response Function; Functional Magnetic Resonance Imaging; Harmonization; ComBAT; scanner effects; multi-site studies

1. INTRODUCTION

Blood oxygenation level-dependent (BOLD) effects produce contrast in functional magnetic resonance imaging (fMRI) of the brain that indicate changes in blood flow, volume and/or oxygenation[1]. BOLD signals are indirect measures of neuronal activity and are characterized by a hemodynamic response function (HRF) which represents the transient response in MRI signal after a brief neural stimulus [2]. Studies of the HRF have focused on evoked BOLD signal changes during stimulus-response tasks. More recently, studies [3], [4] have shown that by identifying relatively large amplitude peaks in a resting state acquisition, it is possible to reliably estimate the HRF without the use of stimuli. Aggregation of data across multiple sites is essential for improving statistical power and reliability of results but introduces variability between acquisition protocols and scanners used. Studies of functional MRI data [5]–[7] have demonstrated large variability between acquisition sites, but few [8] have focused on reducing scanner variability. Recently the ComBAT [9] harmonization technique has been shown to remove site effects in multisite diffusion tensor imaging (DTI) studies[10] while keeping the biological variabilities of subjects. In this work, we examined the effects of the variabilities from scanners and repetition time TR on the HRF. To extract the HRF and its features we used a blind-deconvolution method with resting state fMRI data [4] and coregistered the data in a standard space. We quantified the differences between three manufacturers using statistical tests and applied ComBAT to harmonize the scanners and repetition times. To then investigate possible changes in HRF with aging, we used a cubic spline regression between each feature and age. We also studied the differences in the behavior of each HRF feature between white matter and gray matter to compare our results to previous findings.

2. METHOD

2.1 Dataset

ADNI is a longitudinal multisite study that aims to improve clinical trials for the prevention and treatment of Alzheimer's disease¹ [11]. Our subset is composed of the rollover of patients from ADNI 1 to ADNI 2 [12]; new patients of ADNI 2 and their visit after 3 and 6 months; the rollover of patients from ADNI 2 to ADNI 3 [13]; and new patients of ADNI 3. Subjects have been followed up to 5 years after the first visit for ADNI 2 and up to 4 years for ADNI 3. In total the dataset includes 160 subjects from ADNI 2 and 387 subjects from ADNI 3 and is composed of 271 men and 276 women aged between 56 and 96. Subjects are categorized into six groups: 230 cognitively normal (CN), 65 Significant Memory Concern (SMC), 98 Early Mild Cognitive Impairment (EMCI), 68 Middle Cognitive Impairment (MCI), 52 Late Mild Cognitive Impairment (LMCI), and 34 Alzheimer's Disease (AD). The fMRI data were collected on 3T scanners from 62 acquisition sites, 18 different models of scanners (from GE, Philips, and Siemens) in a resting state, with TR varying between 0.607 to 3s. Processing of the fMRI data included slice timing correction, alignment, detrending, and z-normalization.

¹ <https://adni.loni.usc.edu/>

2.2 HRF and features estimations

We extracted the HRF using a blind deconvolution approach from the open-source toolbox rsHRF² [4], [14]. We used the toolbox with a set of three gamma functions estimating the features of height, time to peak, and full width at half maximum (FWHM), and we computed the dip height, trough height, time to dip, time to trough, peak integral, dip integral, and trough integral (shown in Figure 1). To align each feature to the Montreal Neurological Institute (MNI) standard space for all subjects, we performed a two-step registration. First, we registered the fMRI data to the structural space using the anatomical image of the subject with FSL³ [15], [16]. Then, we estimated the transformation from the structural space to the MNI space using ANTs⁴ [17], which is applied to the functional images registered in the structural space. To assess variabilities between scans and repetition times and to examine whether there is a change in HRF with aging, we considered the average of each feature in gray matter (GM) and white matter (WM). The average of each subject is computed using the binary brain mask of the two tissues expressed in the MNI space.

2.3 Harmonization

We first showed the differences between scanners and repetition time on the white matter to highlight the need for harmonization techniques. To remove these effects, we used ComBAT⁵ [10], [18] and harmonized the WM and GM mean values. ComBAT is a method for correcting batch effects based on an empirical Bayes model initially developed for genomic data. It has been shown to remove site effects while preserving biological variability in multi-site studies. We selected 11 scanners with more than 26 samples per scanner and two repetitions times (TR=0.607s and TR=3.0s) with more than 90 samples. To quantify the differences between manufacturers and evaluate the performance of the harmonization, we compared the effect sizes before and after harmonization using Cohen's d value. To assess if the differences between manufacturers were significant across features, we used the Wilcoxon rank sum [19], because the data distribution across manufacturers is not normal (p-value of the Shapiro-Wilk test [20] $\leq 10^{-5}$).

2.4 Modeling of HRF changes in aging

To study whether there is a variation in the HRF features according to age, we studied only the CN group obtained with a TR=3s and used the following spline-based model:

$$Y \sim b_0 + \sum_{k=1}^{K+d+1} b_k B_k(X)$$

where X is the age, Y is the mean value of the studied feature in the white or gray matter, b_0 is the intercept, and the b_k are the coefficients associated with the B-spline B_k . The spline order has been chosen as cubic ($d=3$) with one knot ($k=1$), located at 77. We used

² <https://github.com/compneuro-da/rsHRF>

³ <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>

⁴ <http://stnava.github.io/ANTs/>

⁵ <https://github.com/Jfortin1/ComBatHarmonization>

bootstrapping to compute the confidence intervals of the regression curve at 5% and 95%. Bootstrapping [21] was applied to 50% of the age distribution restricted between [65; 90], used to predict the regression curve. During the bootstrapping procedure, we randomized the knot value between the years 74 and 80. To evaluate if there are changes in aging, we computed a likelihood ratio test[22] between our model and the same model without the B-spline terms. If the p-value is significant, it means that the feature studied is associated with age.

3. RESULTS

3.1 Harmonization

Figure 2.A shows the Cohen's d and p values obtained for each feature for the three manufacturers before harmonization. If the p-value is less than 0.005, this indicates that the differences between the two distributions are statistically significant. We show important differences between scanners for height, dip height, trough height, peak integral, dip integral, trough integral, time to dip (GE/Philips), time to trough (GE/Siemens and Philips/Siemens), and FWHM (GE/Siemens and Philips/Siemens). However, the differences between manufacturers for the time to peak are not significant. In addition, there are large effect sizes between manufacturers, especially between Siemens and Philips, that are effectively reduced by ComBAT for all the features of the HRF, highlighting the need for harmonization in multi-site studies.

For most features, the p-value does not reject the null hypothesis after harmonization, showing that the differences between manufacturers are no longer significant, except for the trough integral (GE/Siemens), the time to peak (Philips/Siemens), and FWHM (Philips/Siemens) where differences remain. For all features, the Cohen's d values are significantly reduced since they are lower than 0.054. In Figure 3, we show visually the effect of ComBAT on the data. Each point represents a scan, colored by site-specific attributes TR for the first two columns and scanner model for the latter two. We see a successful harmonization of repetition times since each scan obtained with a TR=0.607s is moved into the scans obtained with a TR= 3s. The harmonization of the scanners is to a lesser extent successful since the variance associated with the different scanners decreases as well as the difference between the means is decreased.

3.2 Modeling of HRF changes with aging

In figure 4 we show the regression curves for all HRF parameters. Each point represents the parameters average in WM (orange) or GM (blue) of a subject. For all features, changes with aging are not significant ($p = 0.0401$). We show a time to peak, time to dip, and FWHM higher in the WM than in the GM, which also are consistent with previous studies[23]. For the parameters of the height, trough height, peak integral, and trough integral, the values of WM and GM are consistent with the literature because it has been shown that the amplitude of the BOLD signal is less in the WM than in the GM[23], [24]. However, this study has also shown a dip height in the WM lower than the GM, which is not shown by our results. In Figure 5 we show the HRF average across a population where the images were acquired with a repetition time of 3s and 0.607s. The curves obtained with a low TR are consistent

with the results of previous studies[23], which makes it possible to validate the method and results. This shows the impact of the repetition time during the acquisition of images as it seems to indicate that low temporal resolution acquisitions do not precisely capture the variations of the HRF.

4. CONCLUSION

In this study, we quantified the scanner and TR effects on the hemodynamic response function. The HRF and its features were computed using a blind deconvolution approach and we applied ComBAT on each feature. We also studied how the HRF changes in healthy aging using a spline-based model. We showed significant scanners and TR effects on the HRF, that ComBAT was able to eliminate, but there remains variability in the data, which did not identify a change in HRF with aging. There is an important sensitivity of the HRF to the repetition time where its shape changes significantly between low and high values.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Career Award #1452485, the National Institutes of Health under award numbers R01EB017230, R01MH123201, K01EB032898, and in part by ViSE/VICTR VR3029 and the National Center for Research Resources, Grant UL1 RR024975-01.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI)(National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

- [1]. Ogawa S, Menon RS, Kim SG, and Ugurbil K, "On the characteristics of functional magnetic resonance imaging of the brain," *Annu Rev Biophys Biomol Struct*, vol. 27, pp. 447–474, 1998, doi: 10.1146/annurev.biophys.27.1.447. [PubMed: 9646874]
- [2]. Buxton RB, Uludağ K, Dubowitz DJ, and Liu TT, "Modeling the hemodynamic response to brain activation," *NeuroImage*, vol. 23, pp. S220–S233, Jan. 2004, doi: 10.1016/j.neuroimage.2004.07.013. [PubMed: 15501093]
- [3]. Tagliazucchi E, Balenzuela P, Fraiman D, and Chialvo DR, "Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis," *Frontiers in Physiology*, vol. 3 FEB, 2012, doi: 10.3389/fphys.2012.00015.
- [4]. Wu GR, Liao W, Stramaglia S, Ding JR, Chen H, and Marinazzo D, "A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data," *Medical Image Analysis*, vol. 17, no. 3, pp. 365–374, Apr. 2013, doi: 10.1016/j.media.2013.01.003. [PubMed: 23422254]

- [5]. Friedman L and Glover GH, "Report on a multicenter fMRI quality assurance protocol," *Journal of Magnetic Resonance Imaging*, vol. 23, no. 6, pp. 827–839, 2006, doi: 10.1002/jmri.20583. [PubMed: 16649196]
- [6]. Dansereau C et al. , "Statistical power and prediction accuracy in multisite resting-state fMRI connectivity," *NeuroImage*, vol. 149, pp. 220–232, Apr. 2017, doi: 10.1016/j.neuroimage.2017.01.072. [PubMed: 28161310]
- [7]. Abraham A et al. , "Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example," *NeuroImage*, vol. 147, pp. 736–745, Feb. 2017, doi: 10.1016/j.neuroimage.2016.10.045. [PubMed: 27865923]
- [8]. Yu M et al. , "Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data," *Human Brain Mapping*, vol. 39, no. 11, pp. 4213–4227, Nov. 2018, doi: 10.1002/hbm.24241. [PubMed: 29962049]
- [9]. Johnson WE, Li C, and Rabinovic A, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, Jan. 2007, doi: 10.1093/biostatistics/kxj037. [PubMed: 16632515]
- [10]. Fortin JP et al. , "Harmonization of multi-site diffusion tensor imaging data," *NeuroImage*, vol. 161, pp. 149–170, Nov. 2017, doi: 10.1016/j.neuroimage.2017.08.047. [PubMed: 28826946]
- [11]. Mueller SG et al. , "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, Nov. 2005, doi: 10.1016/j.nic.2005.09.008. [PubMed: 16443497]
- [12]. Aisen PS, Petersen RC, Donohue M, Weiner MW, and Initiative ADN, "Alzheimer's Disease Neuroimaging Initiative 2 Clinical Core: Progress and plans," *Alzheimer's & Dementia*, vol. 11, no. 7, pp. 734–739, 2015, doi: 10.1016/j.jalz.2015.05.005.
- [13]. Weiner MW et al. , "The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement," *Alzheimers Dement*, vol. 13, no. 5, pp. 561–571, May 2017, doi: 10.1016/j.jalz.2016.10.006. [PubMed: 27931796]
- [14]. Wu G-R et al. , "rsHRF: A toolbox for resting-state HRF estimation and deconvolution," *NeuroImage*, vol. 244, p. 118591, Dec. 2021, doi: 10.1016/j.neuroimage.2021.118591. [PubMed: 34560269]
- [15]. Jenkinson M and Smith S, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, Jun. 2001, doi: 10.1016/S1361-8415(01)00036-6. [PubMed: 11516708]
- [16]. Jenkinson M, Bannister P, Brady M, and Smith S, "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002, doi: 10.1006/nimg.2002.1132. [PubMed: 12377157]
- [17]. Avants BB, Tustison N, and Johnson H, "Advanced Normalization Tools (ANTs)," p. 41.
- [18]. Fortin JP et al. , "Harmonization of cortical thickness measurements across scanners and sites," *NeuroImage*, vol. 167, pp. 104–120, Feb. 2018, doi: 10.1016/J.NEUROIMAGE.2017.11.024. [PubMed: 29155184]
- [19]. Hollander M, Pledger G, and Lin P-E, "Robustness of the Wilcoxon Test to a Certain Dependency Between Samples," *The Annals of Statistics*, vol. 2, no. 1, pp. 177–181, 1974.
- [20]. Shapiro SS and Wilk MB, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, p. 591, Dec. 1965, doi: 10.2307/2333709.
- [21]. Efron B, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, Jan. 1979, doi: 10.1214/aos/1176344552.
- [22]. Vuong QH, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989, doi: 10.2307/1912557.
- [23]. Li M, Newton AT, Anderson AW, Ding Z, and Gore JC, "Characterization of the hemodynamic response function in white matter tracts for event-related fMRI," *Nature Communications*, vol. 10, no. 1, Dec. 2019, doi: 10.1038/s41467-019-09076-2.
- [24]. Fraser LM, Stevens MT, Beyea SD, and D'Arcy RCN, "White versus gray matter: fMRI hemodynamic responses show similar characteristics, but differ in peak amplitude," *BMC Neuroscience*, vol. 13, no. 1, p. 91, Aug. 2012, doi: 10.1186/1471-2202-13-91. [PubMed: 22852798]

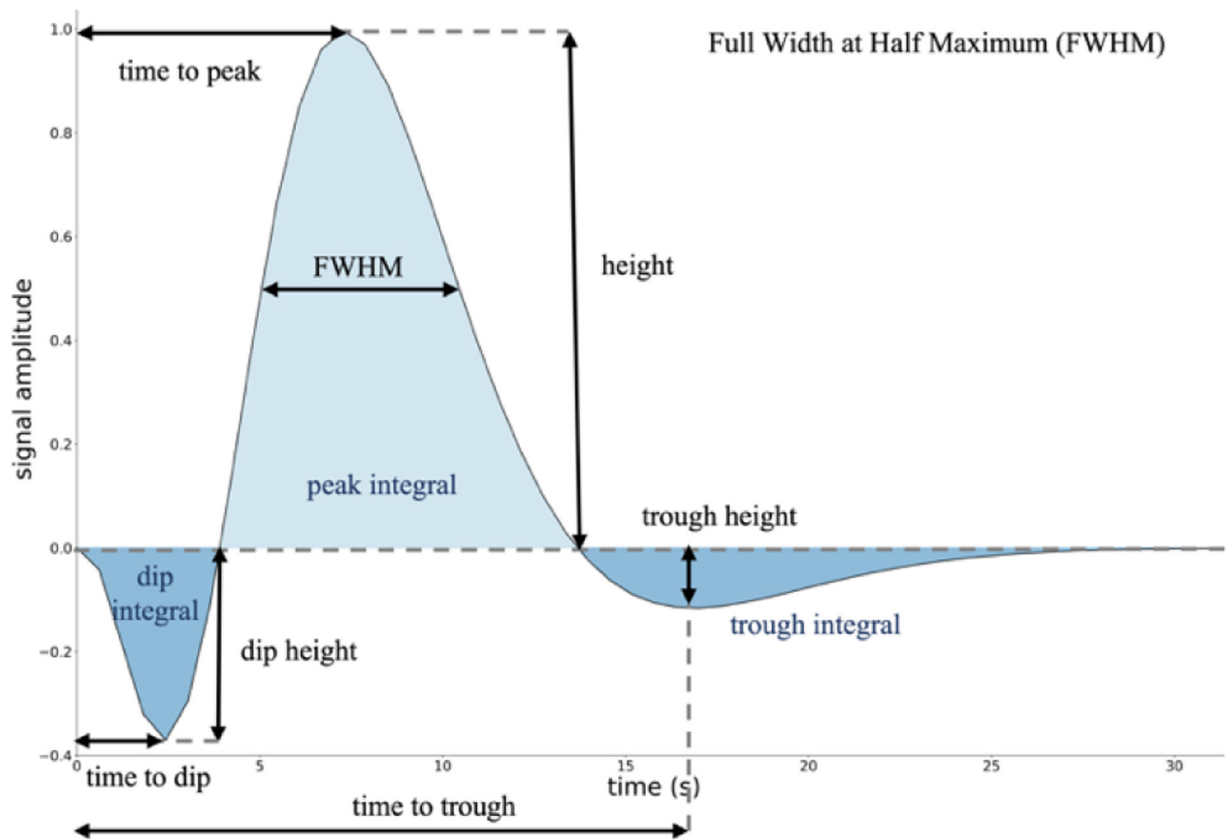
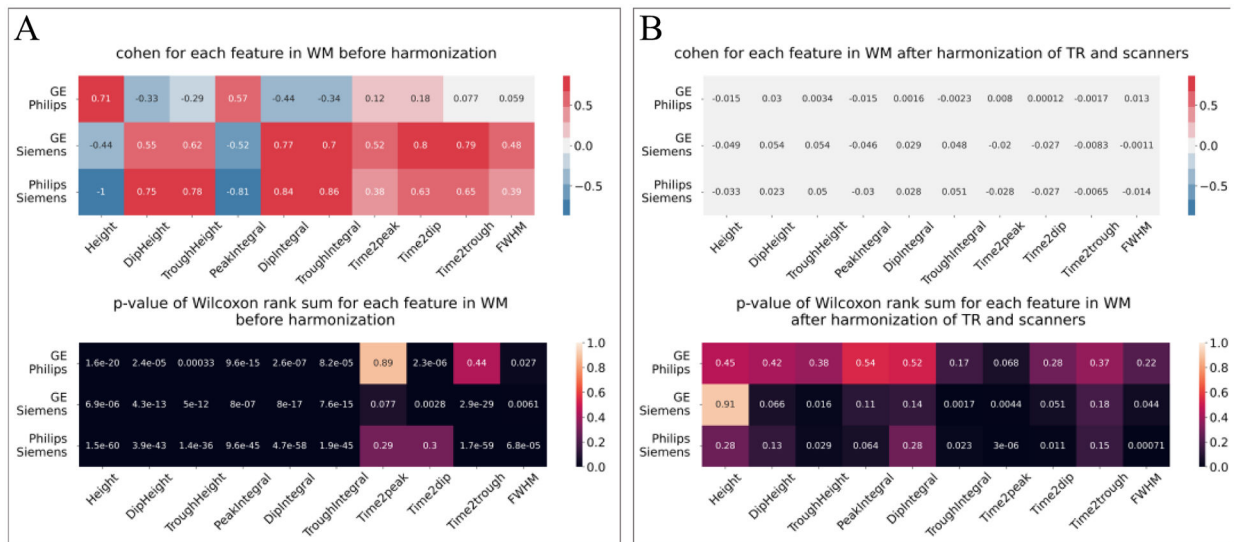


Figure 1.

The hemodynamic response function starts with an increase in the concentration of deoxyhemoglobin resulting in a decrease in the signal. It is then followed by an increase in the blood flow with a higher concentration of oxyhemoglobin, increasing the signal amplitude. Finally, the signal returns to baseline through a post-stimulus undershoot. The features of height, peak integral, FWHM, and time to peak are characterizing the peak of the response; dip height, dip integral, and time to dip are characterizing the dip of the response; and trough height, trough integral, and time to trough are characterizing the post-stimulus undershoot.

**Figure 2.**

Effect size and p-value of the Wilcoxon rank sum test between GE, Philips, and Siemens for each HRF feature. (A) before harmonization. (B) after harmonization of TR and scanner effects. There are significant differences and large effect sizes between manufacturers that are reduced with harmonization.

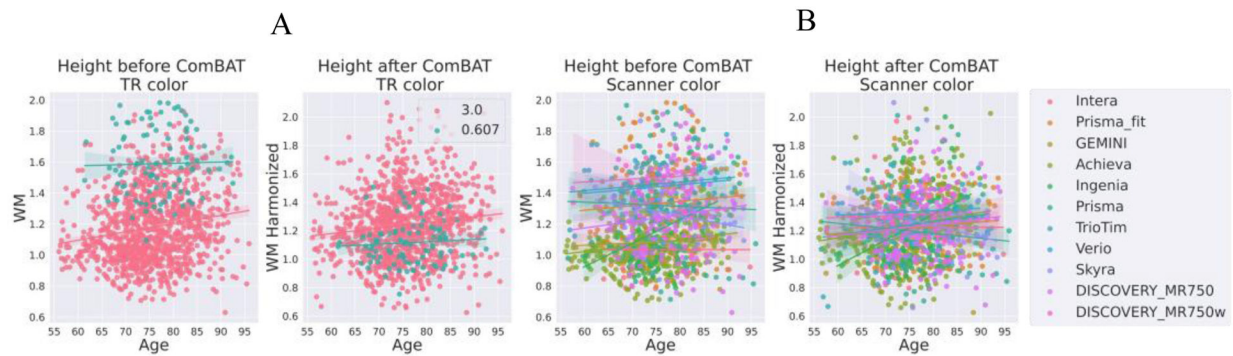


Figure 3.

Effect of the harmonization of scanners and TR on the HRF height. Each point represents the mean value of the height in the white matter and the line models a linear regression between the feature and the age. (A) shows the impact of two TRs on the HRF before and after harmonization where the color represents the TR values. (B) shows the impact of the different scanners on the HRF before and after harmonization where the color represents the scanner models. There are important TR and scanner effects on the HRF that are significantly reduced with the use of harmonization.

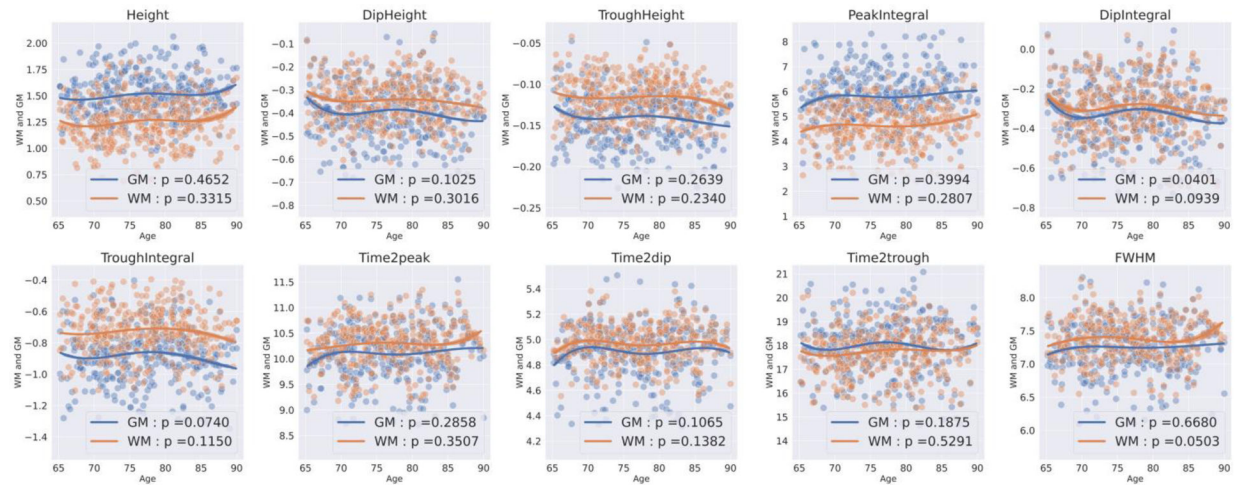


Figure 4. Regression curve based on a spline model modeling changes in HRF with aging in WM (orange) and GM (blue) for each HRF feature. Each point represents the mean value of the parameter for a subject. The p-value for all features is not significant ($p = 0.005$) indicating no changes with aging in WM and GM.

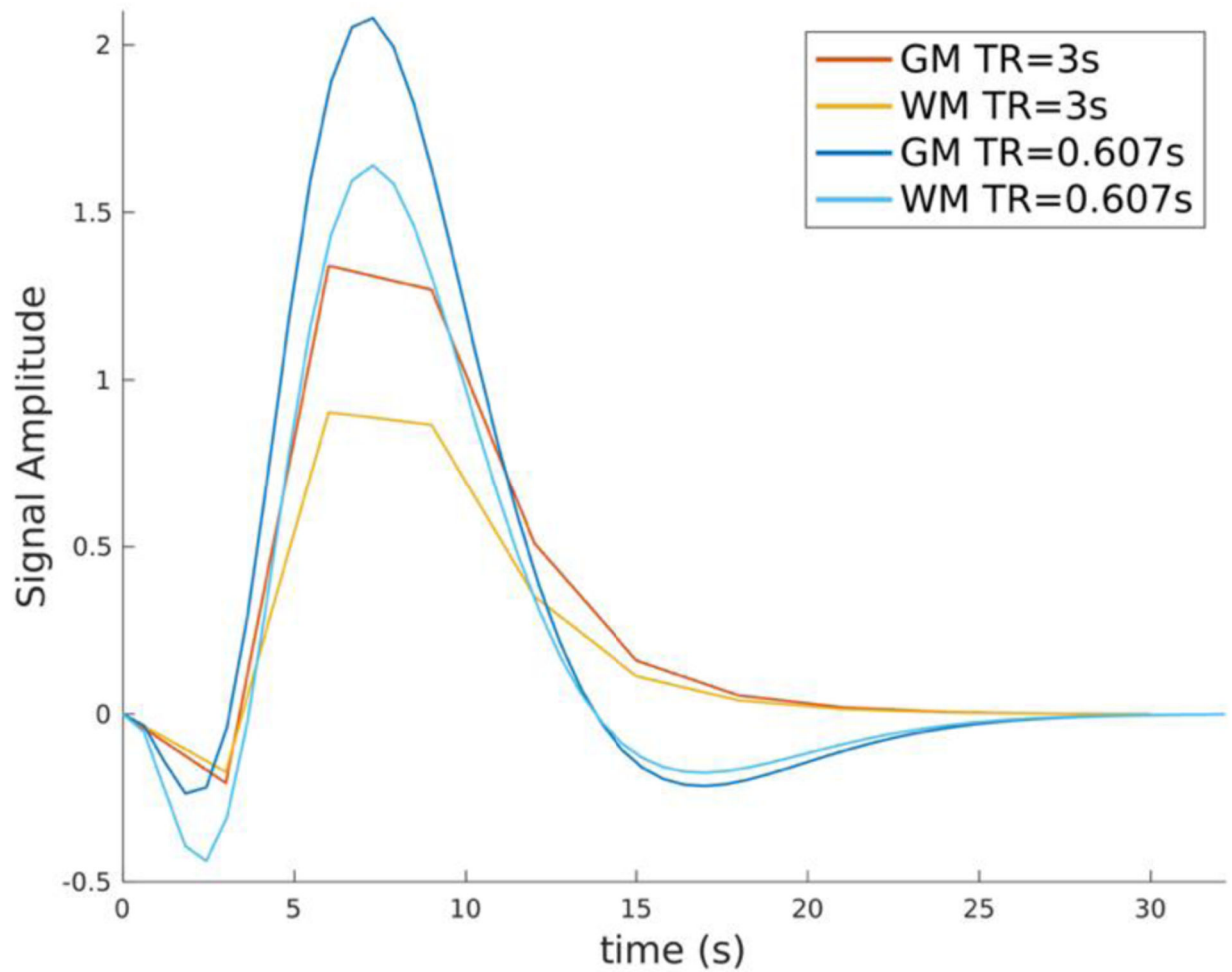


Figure 5.

Comparison of the HRF average in WM and GM across two populations obtained with $TR=0.607s$ (blue curves) and $TR=3s$ (orange curves). It shows significant differences in the HRF behavior according to the TR used during acquisition. The WM is lower than GM in the dip with a $TR=3s$ which shows that an important repetition time does not allow to precisely capture the variations of the HRF.