Commentary

Head motion: the dirty little secret of neuroimaging in psychiatry

Carolina Makowski, BSc; Martin Lepage, PhD; Alan C. Evans, PhD

Psychiatry is at a crossroads when choosing final samples for analysis of neuroimaging data. Many patient populations exhibit significantly increased motion in the scanner compared with healthy controls, suggesting that more patients would need to be excluded to obtain a clean sample. However, this need is often overshadowed by the extensive amount of time and effort required to recruit these valuable and uncommon samples. This commentary sheds light on the impact of motion on imaging studies, drawing examples from psychiatric patient samples to better understand how head motion can confound interpretation of clinically oriented questions. We discuss the impact of even subtle motion artifacts on the interpretation of results as well as how different levels of stringency in quality control can affect findings within nearly identical samples. We also summarize recent initiatives toward harmonization of quality-control procedures as well as tools to prospectively and retrospectively correct for motion artifacts.

Introduction

Neuroimaging studies of psychiatric disorders often face the dilemma of how to handle patient head motion — a dilemma we are reluctant to confront. On one hand, assembling a sufficiently large cohort for meaningful study is a painstaking process, and there is a natural desire to use all the data collected. On the other hand, many patient populations exhibit significantly increased motion in the scanner compared with healthy controls, suggesting that more scans must be excluded to obtain a clean enough sample for a significant result. Do common artifacts such as motion really make a difference in large samples? If so, what should be done about it?

Defining motion is deceivingly simple at the core, but it has been surprisingly difficult for researchers to reach consensus over the threshold of acceptable motion that can be tolerated within an MRI study. Moreover, it has become apparent that motion artifacts in neuroimaging research cannot be ignored. Both structural and functional imaging domains are affected by motion artifacts; studies have highlighted that results obtained with the original sample compared with a quality-controlled subset of the data yield significantly different effect sizes and even different neuroanatomical substrates for interpretation. Recently, an editorial by Weinberger and Radulescu brought this issue to the attention of researchers in psychiatry, challenging the common interpretation of findings derived from case-control MRI studies as altered “neurobiology” in patients compared with controls. Instead, the authors urged the field to more critically and carefully acknowledge MRI-derived confounds, such as head motion, that may be clouding key findings in the literature.

In 2002, Blumenthal and colleagues were among the first to report a significant negative association between grey matter brain volumes and severity of motion artifact. This finding is of particular importance in the study of pediatric populations and neurodevelopmental disorders, given children’s tendency to be more restless and, in turn, exhibit more movement in the scanner. This realization has rebounded in the literature as researchers attempt to disentangle and reconcile the disparate structural trajectories associated with normal brain development — a critical concept that requires consolidation if we are to broach the topic of abnormal brain trajectories in psychiatric disorders. Nearly a decade ago, Shaw and colleagues conducted a seminal study on the cortical thickness trajectories underlying normal development in a sample ranging from 3.5 to 33 years of age and found predominantly nonlinear associations between cortical thickness and age, with notable peaks within the first decade of life. However, the age window leading up to these peaks of cortical development encompasses an age group (i.e., 5–10 yr) where children have been shown to exhibit the most movement in the MRI scanner. These confounds of movement provide a feasible explanation for the challenge in replicating these findings subsequently in independent samples. Ducharme and colleagues addressed such inconsistencies...
using a sample with a similar age range as that in the study by Shaw and colleagues, but with the additional component of three levels of quality control: none, standard and stringent. With increasing stringency of quality control (i.e., removal of more scans with motion), the reported nonlinear cortical thickness associations with age disappeared; instead, predominantly linear associations across most of the brain remained. Characterization of neurodevelopmental trajectories of white matter have also been of great interest and are similarly likely to be significantly influenced by motion. A recent investigation by Reafa and colleagues confirmed this sentiment, showing that significant correlations between commonly reported diffusion metrics (i.e., fractional anisotropy and mean diffusivity) and age are weakened with increased motion estimates.

These examples are convincing evidence that motion artifacts have a significant impact on structural imaging results, but why? As a simple point for comparison, motion in an MRI is akin to motion in any image taken with a camera: the higher degree of motion present, the more blurred and fuzzy the image will be. The basic physics underlying MRI data acquisition add another layer of complexity to the effects of motion or head displacement in a resulting image. When a patient moves in the scanner, it is the spatial frequencies of the MRI, or k-space, that are perturbed. The errors introduced give rise to motion artifacts that are not localized, but propagate throughout the image (e.g., ghosting, ringing, blurring). The signal derived from functional MRI (fMRI) can also be greatly perturbed by motion; two detailed reports by blurring). The signal derived from functional MRI (fMRI) can propagate throughout the image (e.g., ghosting, ringing, blurring). The signal derived from functional MRI (fMRI) can also be greatly perturbed by motion; two detailed reports by

Cortical thinning and grey matter volume loss have been reported in many psychiatric disorders (for a relevant meta-analysis, see Goodkind and colleagues), although it is difficult to disentangle the degree to which grey matter reductions are due to the underlying neurobiology of the disorder itself or due to motion. Yao and colleagues sought this answer in bipolar disorder and schizophrenia samples, finding that reduced surface area and cortical thickness do in fact characterize these psychotic disorders, albeit effect sizes were attenuated when taking head motion into account.

Attention-deficit/hyperactivity disorder (ADHD) poses a particularly interesting challenge in psychiatry, as some of the prominent clinical features characterizing this neurodevelopmental disorder include motion-related symptoms. Two recent studies assessing functional connectivity in ADHD have used rigorous quality-control procedures to more confidently draw upon biologically relevant networks underlying the disorder. Mirroring findings in structural imaging, one of these studies also reported reduced effect sizes compared with those reported in previous literature on functional connectivity patterns in ADHD, which the authors largely attribute to their cleaner dataset. Of note, the study by Fair and colleagues demonstrated that correcting for motion can have clinical utility in predicting subtypes of ADHD (i.e., combined v. inattentive subtypes). Specifically, the authors used three different models of motion correction in a machine learning algorithm to classify subgroups of patients with ADHD, finding a relatively high degree of accuracy (i.e., 71%–77%) with two of these models.

One of the largest and most comprehensive investigations examining motion bias across clinical cohorts (i.e., ADHD, ASD and schizophrenia) and different post-processing software was recently conducted by Pardoe and colleagues. This
study used resting-state fMRI acquisitions to inform the degree of movement present in $T_1$-weighted images and examined this quantitative metric of motion in the context of brain morphometry, namely cortical thickness, grey/white matter contrast and grey matter/subcortical volumes. As expected, motion estimates were higher in all clinical populations than in controls and for the extreme ends of the age distribution (i.e., < 20 yr and > 40 yr). Intriguingly, cortical thickness and measures of cortical contrast were more affected by motion than volumetry, although the latter was affected to a greater degree by the choice of segmentation method. This paints a rather intricate and complex picture of the various levels by which motion may impact neuroimaging analyses of clinical cohorts and underscores the importance of, at minimum, having a clear and consistent quality-assurance protocol to exclude scans visibly affected by motion artifacts. Furthermore, given the manner by which different pipelines may handle motion, it is equally important to study thoroughly the processed outputs — successful runs may not always be linked to trustworthy results.

Many of the examples discussed thus far have come from cross-sectional study designs. However, longitudinal neuroimaging studies are integral in capturing dynamic brain changes that underlie the clinical course of psychiatric illness. Notably, it has been shown previously that head motion shows some degree of test–retest reliability.26–30 Thus, for longitudinal studies, motion correction should be considered for all time points. It may also be the case that only a subset of longitudinal scans from a single participant are retained for analysis after completing quality control. Fortunately, statistical methods, such as multilevel modelling (i.e., mixed-effects models, hierarchical linear models) are designed to handle missing data points. For a more complete description of these methods, see Singer and Willet.31

Guidelines to minimize and correct for head motion artifacts

In the sections that follow, we propose a quality-control workflow that can serve as a framework for both prospective and retrospective datasets. This workflow is depicted in Figure 1. In addition to psychiatric patient samples, it should be noted that the confounds of head motion in the scanner also exist in other fields of biomedical imaging, such as neurology. As such, we propose that this quality-control workflow can be readily generalized to other patient populations, such as patients with neurologic disorders characterized by motor symptoms.

Considerations for prospective datasets

Behavioural training

One of the simplest and arguably most effective techniques in minimizing head motion is to ensure the participant remains as still as possible within the MRI scanner. Sedation and general anesthesia have been successful in minimizing motion, particularly within pediatric populations, but such methods introduce additional confounds and are not always feasible or desirable in most research studies.32–34 Further, several studies have shown that it is possible to obtain successful scans within pediatric and psychiatric patient samples without the use of more invasive procedures.7,32,33,35–37 For instance, it is good practice to acclimatize the participant to the MRI setting using a mock scanner and to provide adequate training tailored to the population of interest. If a mock scanner is not available, videos simulating the MRI environment may be used (e.g., http://vimeo.com/32255381).38 Acclimatizing children to the scanner environment requires more personnel and preparation time than required for adolescents or adults. Detailed pediatric protocols, such as the inclusion of visual aids or allowing the child to view a movie during the scan, have been reported with high success rates.7,39,40 Neurofeedback paradigms have also been explored, where real-time feedback is provided to participants on their patterns of movement in the scanner,41 although this technique has had limited returns thus far. In summary, incorporating pre-scan protocols with attention to participant compliance and behaviour during scan acquisition will enhance the likelihood of a successful scan. These points are highlighted in Figure 1, under the “behavioural training” portion of the quality-control workflow.

Technical and methodological considerations

A clear study protocol is required before any scans are collected. The mantra of “quality over quantity” is an important consideration when deciding upon a scanning protocol for a psychiatric neuroimaging study. Focusing on one or two modalities and optimizing their quality (e.g., by acquiring multiple structural imaging scans or functional imaging runs) may be a more efficient use of scan time than acquiring images using many modalities at suboptimal quality. Minimizing motion during scan acquisition may also require technical intervention. In recent years, various research groups have worked diligently toward developing and improving imaging protocols to correct for motion as scans are being acquired in real time. These new techniques hold promise in significantly reducing the proportion of scans that might otherwise need to be excluded owing to motion confounds. For instance, prospective motion correction (PROMO) has been proposed to counter the effects of inevitable participant movement in a proactive manner.5,42,43 The PROMO framework acts in real time during the acquisition of a scan, where motion can be detected through maintenance of a fixed coordinate system in relation to the participant’s position within the scanner, and images are automatically rescanned if significant motion artifact is sensed. This procedure has been shown to drastically reduce the caveats associated with motion in children, and intuitively is expected to hold similar benefits for clinical populations. Framewise integrated real-time MRI monitoring (FIRMM) has also been developed recently to quantify degree of motion from frame to frame as the data are being acquired, allowing MRI technicians and researchers on site to actively monitor head movement accurately and proactively.44 Similarly, measures of head position and orientation while scanning have been used to improve the quality of images acquired using positron emission tomography (PET).45 These points can be found in the “technical/methodological” portion of the quality-control workflow in Figure 1.
Fig. 1: Proposed quality control workflow for MRI data sets, both prospective and retrospective. Note: this figure is designed as a simplified guide and is not a comprehensive workflow for all imaging modalities and clinical populations. DTI = diffusion tensor imaging; fMRI = functional MRI; QC = quality control.
Considerations for retrospective datasets

The previous section introduced practical suggestions for prospective data collection. A natural question that follows is what solutions exist to handle motion artifacts within the abundance of clinical neuroimaging data that have already been amassed? Having a quantitative measure of motion — a variable that could be used as a means to better match patient and control samples — could be beneficial. We provide some suggestions in Figure 1 that can be applied to raw imaging data to help mitigate the effects of motion artifacts on retrospective datasets.

Correcting for motion after image acquisition

Motion estimates from functional imaging can be used to inform analysis of T1-weighted structural scans, although caution should be exercised when using a proxy measure of motion from another imaging modality. Motion during one acquisition does not necessitate motion during another, even if both acquisitions are collected within the same session. In the realms of diffusion imaging, a total motion index (TMI) has been defined based on several proxies for motion, including calculations of translation, rotation and signal dropout, to obtain a collective “score” of motion. The TMI can then be used as a regressor in the comparison of group differences. Scores of modality-specific toolboxes and workflows have also been developed to help manage the effects of motion on image quality, for application to structural T1-weighted images, diffusion-weighted imaging, and task-based fMRI positron emission tomography and arterial spin labelling.

Choosing MRI scans for analysis

We have argued the idea that a “clean” dataset, with minimal motion impact, will yield a more biologically valid finding. However, a clear consensus on data cleaning standards has not yet emerged, despite worthy efforts in that direction. This is largely attributable to the fact that defining inclusion/exclusion criteria of an MRI scan rests largely upon the research question at hand and on the imaging modality. Standards in fMRI may be more explicit on this point; for instance, Power and colleagues proposed a widely cited method of “scrubbing” fMRI data to remove frames with a high degree of motion and significant amplitude changes in the blood oxygen level–dependent (BOLD) signal. Alternatives to handle different types of motion-related variance in fMRI acquisitions have been described recently by Caballero-Gaudes and Reynolds. Image-quality metrics can also be derived automatically (e.g., signal-to-noise and contrast-to-noise ratios), which provides a quantitative metric that can allow researchers to quickly pinpoint “outliers” that should be flagged (see the “raw image quality control” section of Fig. 1). 

Conclusion

The quantification of motion to be accounted for when analyzing data is certainly attractive, compared with mere exclusion of scans, in the analysis of retrospective data. However, sometimes a bad scan is just a bad scan, and it may be worthwhile to exercise the art of “letting go” in severe cases. Neuroimaging technology is developing quickly, and it is reasonable to expect that better algorithms and solutions for handling the blurred edges in our scans will be coming our way. Until then, do not shy away from data cleaning; the rewards gained in validity are worth the loss of a few scans.

Acknowledgements: C. Makowski is supported by a doctoral graduate award from the Canadian Institutes of Health Research (CIHR; CGS-D) and funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives Initiative. M. Lepage and A.C. Evans are supported by a James McGill Professorship at McGill University and hold operating grants from CIHR. The authors thank members of the McGill Centre for Integrative Neuroscience, past and present, for valuable discussions that contributed to some of the ideas presented in this manuscript.

Affiliations: From the McGill Centre for Integrative Neuroscience, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Que., Canada (Makowski, Evans); and the Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Verdun, Que., Canada (Makowski, Lepage).
Competing interests: A.C. Evans has received consulting fees from Biospective. No other competing interests declared.

Funding: The funding sources had no role in the writing of this article.

Contributors: This commentary was conceptualized by all authors. C. Makowski wrote the initial versions of the manuscript, and M. Lepage and A.C. Evans revised the content of the article. All authors approved the final version to be published and can certify that no other individuals not listed as authors have made substantial contributions to the paper.

References


