Harmonization of resting-state functional MRI data across multiple imaging sites

April 18 2019

Credit: CC0 Public Domain

It has become increasingly apparent that the low reproducibility of results is a ubiquitous problem across many scientific fields such as biomedical science and psychology. This problem is particularly serious

in biomedical studies using functional magnetic resonance imaging (fMRI) data. An increasing number of studies have reported success in constructing machine-learning algorithms (artificial intelligence) that use fMRI data to classify subjects as either healthy or suffering from a psychiatric disorder. However, it has been suggested that if these classifiers were constructed from a small number of samples (e.g., tens of participants) from a single site, it might not be possible to generalize their application to the data acquired from other imaging sites. A solution to this low generalization capability is to collect big data across many sites, but the considerable site-related differences in fMRI data is a formidable obstacle to the feasibility of this solution.

The current study, published in the journal *PLOS Biology*, revealed important mechanisms underlying the site-related differences in fMRI data and developed an effective harmonization method (i.e., reduction of site-related differences) for resting-state fMRI (rs-fMRI) data, thus reducing such differences by 30 percent. Furthermore, this research group publicly released a neuroimaging database, parts of which were used in the current study. This database is internationally unique in being a multi-site neuroimaging database of thousands of people suffering from multiple psychiatric disorders. Their study and dataset provide an innovative path toward accelerating our understanding of the neural mechanisms behind psychiatric disorders as well as the development of neuroimaging biomarkers of various disorders.

The research group collected a traveling-subject rs-fMRI dataset, in which nine participants traveled to 12 sites, and a multi-site, multi-disorder rs-fMRI dataset, in which 805 participants suffered from 4 types of psychiatric disorders (autism spectrum disorder, major depressive disorder, schizophrenia, and obsessive-compulsive disorder), from 9 sites. Researchers estimated the influence of different imaging sites, individual differences, and the effects of disorders on resting-state functional connectivity by applying a mathematical model to the

combined dataset (traveling-subject rs-fMRI dataset + multi-site, multi-disorder rs-fMRI dataset).

Moreover, the research group demonstrated that site-related differences involve biological sampling bias (differences in participant groups) and engineering measurement bias (differences in the properties of MRI scanners). Researchers found that the effects of both bias types on rs-fMRI functional connectivity were greater than or equal to those of psychiatric disorders. They specified the properties of MRI scanners that significantly affected the rs-fMRI connectivity. Furthermore, they revealed that each site can sample only from a sub-population of the participants, which suggests the importance of collecting multi-site data. To overcome the limitations associated with the site differences, they developed a novel harmonization method that removed only the measurement bias by using the traveling-subject dataset, and in this way they reduced measurement bias by 29 percent and improved the signal-to-noise ratios by 40 percent.

Consequently, the research group could construct a valuable multi-site neuroimaging database of 2,409 individuals suffering from multiple psychiatric disorders (ASD: autism spectrum disorder, MDD: major depressive disorder, SCZ: schizophrenia, OCD: obsessive-compulsive disorder, CP: chronic back pain, and several others). This database consists of resting-state fMRI, structural images of the brain, and patient demographics (age, sex, and clinical rating scales) of 125 participants with ASD, 455 participants with MDD, 159 participants with SCZ, 110 participants with OCD, 107 participants with CP, and 42 participants with other disorders as well as 1,421 healthy participants. The researchers released parts of this database to approved users. A person who wants to use this database must download the application form from a website, provide the necessary information, and send it to an e-mail address. After the approval of each application, an ID account is issued. An ID account allows the user to download the data from website. So

far, the research group has released the data of 706 participants with psychiatric disorders and 1,122 healthy participants from 8 sites (12 MRI scanners). In addition, the team is currently constructing a publicly open database that anyone can access.

In the future, analyzing such large data by the proposed harmonization method could lead to developing practical brain-circuit-based biomarkers of psychiatric disorders regardless of the imaging site used. This study is thus expected to make a strong contribution to the diagnosis and treatment of psychiatric disorders.

More information: Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLOS Biology*, DOI: 10.1371/journal.pbio.3000042

Provided by ATR Brain Information Communication Research Laboratory Group