Small p values may not yield robust findings: An example using REST-meta-PD

June 28 2021

Credit: Pixabay/CC0 Public Domain

Thousands of scientific papers describing the inner workings of the brain

and its dysfunction have been published using resting state functional magnetic resonance imaging (RS-fMRI). This powerful tool allows researchers to look at each cubic millimeter of the brain as voxels—the 3D version of a pixel. The average brain is well over 1,000,000 cubic mm, so researchers need to perform multiple comparison correction (MCC) to reduce the possibility of making false claims.

As part of this MCC, a smaller p value threshold is widely recommended for declaring significance. Yet there are many ways to perform MCC, and some methods are considered more liberal, while others are more stringent. A stringent MCC may reduce the number of false positive results and the brain regions surviving MCC are often considered to be true positive results. A true positive would mean that the result could be found again, and again, in different studies. But is that really the case? To answer this question, and to determine how to best reduce false positives and increase reproducibility, researchers around the world are coming together as part of a worldwide large-scale consortium.

The REST-meta-PD study combines RS-fMRI data from 15 independent studies of PD patients and performed the voxel-wise analysis known as 'amplitude of low frequency fluctuations' (ALFF). Two results are worth mentioning. The first is false positivity or low reproducibility: The research team found that around 80% of the voxels declared to be important, or significant after MCC, in each individual cohort did not reflect the results of the entire dataset when the data were all pooled together. The second is false negativity: The most robust result that was identified in the full dataset was the abnormal activity in the left putamen, a brain region with known involvement in Parkinson's disease. Interestingly, most individual studies would not have been able to find this effect on their own when performing the stringent MCC—the p-values were not low enough to meet the threshold.

This international team has found that the use of stringent MCC in

smaller sampled studies may exclude meaningful brain regions because the differences are small and, despite being consistent across populations, may be hard to detect in single cohort studies. Results from studies with small sample sizes are known to be limited in terms of reproducibility and generalizability, and have fueled articles with titles such as "Why Most Published Research Findings Are False", and this is clearly also the case for RS-fMRI studies.

The authors propose a normative modeling method (NMM) which integrates the variability inherent within the healthy control group across all 15 individual studies (cohorts). They found that this model achieved good reproducibility even for results from a single-cohort study. An effective mechanism for sharing imaging data and building a reference model based on NMM is urgently needed in the neuroimaging community. Sharing individual brain scan data remains a challenge because of the potential privacy and ethical concerns, but full, unthresholded, statistical maps can be shared easily and do not contain any individual level information. By sharing these maps rather than simply the results that have passed a stringent MCC, the research community can be more confident that their study results will be replicable, and be relevant to much larger populations. RS-fMRI holds advantages of non-invasiveness, fairly good spatial and temporal resolution, and easy access in most hospitals. The current study provides a way for RS-fMRI clinical application in the near future, e.g., precise localization of the abnormal activity, and hence guide precise brain modulation.

Medical Xpress

Provided by Science China Press