|  |  |
| --- | --- |
| **Article Title** | Replication of machine learning methods to predict treatment outcome with antidepressant medications in patients with major depressive disorder from STAR\*D and CAN-BIND-1 |
| **Original Article DOI**  Example: pone.1234567 | pone.0253023 |
| **Description of the Error(s)**  Include any relevant information like updated ID numbers (e.g. grant numbers, DOIs, URLs, etc.) | After publication, some bugs were found in the computer code used to generate our results.  Our study looked at whether subjects achieved a certain depression score (e.g., remission defined as QIDS-SR score ≤5) after a certain period of time. Due to a bug, we were looking at whether they achieved a certain score at any point during this period of time, instead of during their last score in this time period. Fixing this bug was what led to most of the different results in our revised results.  As well, the results we reported as being the results of testing our cross-validation models on our STAR\*D holdout set were actually results from the cross-validation itself. This had little effect on the results upon correcting, likely because the holdout set was randomly chosen as described in the initial manuscript.  For our week 9 STAR\*D predictions and external validation, we have also added an explicit requirement that subject data be from Level 1, as a few subjects went to Level 2 early. This also had little effect on the results.  We also added an explicit requirement that subjects have a baseline score of more than 5 for the relevant QIDS score. This was already a requirement for the QIDS-C datasets, so had no change on the QIDS-C models. For the QIDS-SR models, this resulted in dropping around 1% or less of subjects and also had little effect on the results.  All authors have been apprised of the error and accept the corrections and this request. |
| **How do the error(s) affect the results, conclusions, and overall scientific understanding of your study?** | After fixing these bugs, most of the numerical results reported in our paper have changed, generally decreasing by a small amount. Please find the revised manuscript attached, with track changes to show differences.    Few edits were required in the text for how we reported or discussed our results, as the general results are upheld. Some secondary results are now different but largely have the same conclusions; for example, we initially reported that using the fewer, overlapping features between two datasets increased performance, but now we found that using these fewer features results in about the same performance as all features; these results are still interesting compared to prior work which found a decrease in performance.    We also made a few edits to our methodology section with some extra details to improve reproducibility. |