Bias in group-level EEG microstate analysis

Murphy M[1], Wang J[2], Jiang C[2], Wang L[3], Kozhemiako N[4], Wang Y[3], the GRINS consortium[^], Pan JQ[3] & Purcell SM[3,4,*]

[1] Department of Psychiatry, McLean Hospital, Harvard Medical School, United States

[2] The Affiliated Wuxi Mental Health Center of Nanjing Medical University, Wuxi, China

[3] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, United States

[4] Department of Psychiatry, Brigham & Women's Hospital, Harvard Medical School, United States

^ = see members listed below

* = corresponding author

Abstract

Microstate analysis is a promising technique for analyzing high-density electroencephalographic data, but there are multiple questions about methodological best practices. Here we focus on a commonly used approach to the analysis of group differences in microstate maps and the derived metrics based on those maps. Using simulations seeded on real data from healthy control subjects, we compared error rates for analyses performed using microstates derived from the entire dataset to analyses performed using microstates derived separately within each subgroup. The latter approach resulted in substantially higher type I error rates. These results suggest that even subtle differences in microstate topography can have profound effects on derived microstate metrics and that future studies using microstate analysis should take steps to mitigate this large source of error.

Background

Multichannel electroencephalographic (EEG) data can be modeled as a set of semi-stable recurring voltage topographies called microstates. Each microstate persists for 50-150 milliseconds before transitioning to a different microstate. Multiple studies have demonstrated that only a small number of unique microstates are needed to explain most of the variance in the EEG signal[1]. These microstates are extracted via data-driven clustering, most commonly modified *k*-means[2] or atomize and agglomerate hierarchical clustering[3]. While different studies have reported different sets of microstates, four canonical microstates (A, B, C and D) have been widely replicated in studies of resting state EEG[4]. The neurobiological generators of these microstates remain unclear, but evidence suggests that each microstate corresponds to a distinct large-scale neural assembly that overlaps with, but is distinct from, a resting state network as measured by functional magnetic resonance imaging[5].

Microstate analysis has been used to study normal human physiology as well as psychiatric and neurological disease[6–9]. However, comparing results across studies is often difficult because of multiple analytic decision points during both the extraction of microstate topographies and when "backfitting" these topographies to the original signal data to derive various microstate parameters, as described below. Despite these challenges, there are results which appear to be consistent across studies – for example, that schizophrenia may be associated with increased canonical microstate C[7,10,11].

In many studies using microstate analysis to identify differences between groups, microstate topographies are first extracted from each group independently (see for example [6–8]). Then, authors compare the sets of topographies and align them based on spatial correlation and similarity to canonical microstate topography. Each group is then backfit with its own topographies. Parameters such as microstate duration, occurrence, explained variance, and transition properties are calculated from these fits and then compared across groups.

Microstate topographies calculated from different data sets may be broadly similar but are typically not identical. That is, canonical microstate A derived from one data set will not be identical to canonical microstate A derived from a different data set. Multiple studies have shown the canonical microstates C and D have more varied topography both within and between studies than do microstates A and B[5,10,12,13]. It is unclear to what extent subtle differences in microstate topography may impact subsequently estimated parameters. When considering group differences, an alternative approach is to derive a common set of microstates from all individuals in a study and use that common set for backfitting and calculation of microstate parameters. We speculated that the former approach may inappropriately capitalize on chance group differences

in microstate topologies, especially when applied to relatively small samples, and thus lead to spurious differences in derived parameters which no longer reflect "apples-to-apples" comparisons. In contrast, using a common set of microstates should reduce the risk of reporting spurious group differences (referred to below as "combined map" versus "subgroup maps" conditions).

We tested eyes-closed, resting state EEG data from 30 healthy control participants. For each of 5,000 replicates, we randomly divided this group into two equal-sized subgroups. Using the Luna analysis package (developed by S.M.P.), we then calculated microstate parameters by either 1) fitting all the data to a single set of microstates derived from the pooled data ("common map"), or 2) fitting the data from each subgroup to a set of microstates derived only from that subgroup ("subgroup maps"). For the subgroup-derived microstates, in keeping with previous work, we paired each microstate with the microstate in the other subgroup it most closely resembled (see Methods). We then calculated whether there were statistically significant ($p < 0.05$) differences in any of five microstate parameters (coverage, duration, occurrence, variance explained, mean spatial correlation) between the two subgroups.

Below we arbitrarily refer to the randomized groups as "cases" and "controls", for convenience and to mimic a typical analysis context, although as noted above, all original data are from healthy controls. As group labels are assigned under the null hypothesis of no group differences, only 5% of replicates are expected to be significant at $p < 0.05$, if the procedure appropriately controls false positive rates.

Distinct from testing for group differences in microstate parameters, we also tested for case/control differences in the microstate topographies themselves, based on summaries of pairwise distances between 30 individual-level maps, e.g. asking whether case/control pairs had more dissimilar maps than case/case or control/control pairs on average (see Methods). Figure 1 shows an overview of the simulation and testing procedure, repeated 5,000 times.

Results

By design, the simulation procedure did not induce systematic differences between the topographies of randomly assigned cases and controls. Consistent with this, for all tests comparing the topographical similarity of case and control maps (based on individual-level, $30 \times N{=}1$ segmentation) we observed type I error rates close to their expected value, i.e. 5% (data not shown). Further, based on subgroup ($2 \times N{=}15$) maps, cases and controls typically showed very high spatial correlations with respect to i) to the assigned canonical template map (median $r =$

0.98), ii) to the matched combined ($1 \times N$=30) map (median $r = 0.99$), as well as iii) to each other (median $r = 0.95$). That is, as expected, randomly assigned cases and controls typically exhibited similar $K = 4$ microstate topographies. Further, the median correlation of microstates between subgroups were 0.90, 0.93, 0.95, and 0.97 for canonical microstates A-D respectively, which is comparable to values reported in [6–8].

In contrast, despite similar topographies, error rates for group differences in derived microstate metrics were markedly inflated when using subgroup maps, but not when using a common map (Figure 2a,b). For the shared microstate analyses, 5.01% of all comparisons were statistically significant at a $p < 0.05$ threshold and the median number of false positives per permutation was 0 (of 20 tests performed). In contrast, when using subgroup maps, 31.11% of all comparisons were statistically significant at a $p < 0.05$ threshold and the median number of false positives per permutation was 6 of 20. Based on a conservative Bonferroni-corrected error rate of $0.05 / 20$, the experiment-wide error rate was 68.78% when using subgroup maps, versus 3.72% when using a combined map.

Investigators have long been aware that differences between microstate topographies in different subgroups complicate the interpretation of subgroup differences in microstate parameters[10,12,14,15]. One widespread approach to this complication is to measure the spatial correlation of the microstate maps between subgroups. These correlations are often quite high (for example, see da Cruz et al.[7], where all correlations are > 0.91). These high correlations are then used to explicitly justify using subgroup-specific templates. We note that we observed elevated type I error rates despite spatial correlations that were comparable with values reported by other groups. Furthermore, we restricted our analyses to the 1,026 replicates with the closest case-control topographies, defined as i) case-control spatial correlations $r > 0.95$ for all four states, ii) no 'suboptimal' template map assignments (see Methods) and iii) no significant ($p < 0.01$) case-control differences for tests of topographical similarity. Even in this set, the average error rate was still inflated for analyses based on subgroup maps (16.97%) and but not shared maps (4.99%). Of note, error rates for microstate C were particularly inflated for all five metrics (44.6%) in this reduced set. Taken together, these results suggest that the widespread practice of using spatial correlations to support the use of subgroup derived maps can still be prone to inflated rates of type I error even when subgroup maps are highly correlated.

Overall, inflated error rates were most apparent for comparisons involving microstate B and C, and to a lesser extent D. That microstate A had lower rates of type I error (at least for some metrics) likely reflects different levels of sample-specific variability in topographies, peculiar to this particular study, rather than general properties of microstate analysis or neurophysiology.

That is, if based on a different set of individuals, or using a different number of microstates, we might expect qualitatively different (albeit still generally biased) profile of results.

To give a concrete example, Figure 3 shows the maps and coverage metrics obtained from one simulation replicate. Although maps are visually similar and have high spatial correlations between groups, and with the template maps, they are not identical, as the plot of map differences shows (Figure 3a). As a consequence, although most coverage parameters derived from subgroup and combined maps are correlated with each other, they can also diverge (Figure 3b). Further, this divergence occurs differentially between case and control groups, leading to a spurious difference in coverage for microstate B, C, and D (Figure 3c).

Finally, in order to test whether this finding was likely to be robust across microstate segmentation methods, we tested a single permutation using the analytic approach implemented in CARTOOL[16] (for details, see[6]). We selected a permutation with a high degree of case-control spatial correlation between subgroup maps (all $r > 0.95$). In this specific analysis, we found that 66% of the comparisons were statistically significant using subgroup maps versus none when using shared maps. This suggests that our results are not specific to any one microstate segmentation procedure but rather a common feature of this kind of analysis.

Conclusions

We show that, when microstate maps are estimated separately within subgroups, small differences in microstate topography — *arising purely by chance* — can nonetheless induce detectable differences in microstate parameters and lead to increased rates of spurious differences between groups. We suggest that between-group differences in microstate parameters should be calculated using a shared set of microstates derived from all the data to eliminate this bias.

We expect this will be a general property of any microstate analysis pipeline that directly compares metrics based on maps that were estimated separately: as above, we found that CARTOOL yielded similar results when applied to an exemplar dataset as Luna. We used Luna for the primary analysis as it is computationally efficient and easier to parallelize in a compute cluster.

Further work is needed to determine whether larger studies are similarly susceptible to inflated type I error rates when microstate topographies are calculated independently for each group to be analyzed. Another approach is to use a pre-defined set of topographies for all analyses. In a study of microstate parameters in patients with schizophrenia, Giordano et al.[10]

used a single set of microstate templates derived from a large dataset of healthy controls. This is an appealing approach, however only 19 channels were used to collect the data for the template and the authors down-sampled their high-density EEG data to 19 channels. Recent work suggests that microstate analyses using 19 or fewer channels may be inherently unreliable[17].

Naturally, in applications to real data, two groups (or physiological states) may in fact differ in their characteristic microstate topographies. In this scenario, analyses using a combined map may not be appropriate either, as underlying topographical differences may be reflected as differences in microstate parameters such as coverage. In practice, if two groups show i) significant differences in microstate topographies, and/or ii) significant differences in the mean spatial correlation of sample points to their assigned map for a given microstate, one should avoid making strong inferences about differences (or lack thereof) in other derived parameters for that state or states. Although beyond the scope of this note, an alternative may be to derive subgroup maps but then backfit all individuals to the combined set (e.g $K = 8$), appropriately merging labels when calculating metrics such as coverage (i.e. which would represent *either* the case-derived state A, or the control-derived state A, for example). This would still require that both case and control A maps are sufficiently similar to each other, or to a canonical template, for meaningful interpretation. From a statistical perspective, however, applying the same set of maps to all individuals should at least alleviate the inflation in type I described above, arising from comparing metrics based on different maps, which should be avoided as it is fundamentally anti-conservative under the null.

Materials and Methods

Participants and Procedure

The analyzed data was collected as part of the Global Research Initiative on Neurophysiology of Schizophrenia (GRINS), a research collaboration between the Broad Institute of Harvard and the Massachusetts Institute of Technology (USA), Harvard Medical School (USA), and Wuxi Mental Health Center (China). All study procedures were approved by the Institutional Review Board of Harvard TH Chan School of Public Health and the Institutional Review Board of the Wuxi Mental Health Center. Data collection procedures for this study were previously reported in [18]. All participants were between 18 and 45 years old, had no history of serious neurologic or psychiatric illness, and were not taking medications. All participants provided informed consent before completing the study. For each participant, eight minutes of eyes-closed resting states EEG data was recorded using a 64-channel Ag/AgCl EasyCap system (Brain Products GmbH, Germany). After semi-automatic artifact rejection (including segments with excessive eye movement), the final recording durations were 4.5 minutes on average (ranging from 2 to 7 minutes). Data were recorded at 500 Hz with a left clavicle ground and FPz reference electrode. Electrode impedances were kept below 10 kΩ. For further details, please see [18].

Microstate segmentation

Microstate segmentation and fitting was performed using Luna (http://zzz.bwh.harvard.edu/luna/). For microstate segmentation, based on 57 average-referenced and bandpass filtered (0.3 to 35 Hz) EEG channels, we calculated the global mean field power (GFP) across all electrodes. For each recording, we i) identified all GFP local maxima ("GFP peaks"), ii) rejected peaks that were of unusually high (>3 standard deviation (SD) units above the GFP peak mean) or low (>1 SD below) power, iii) as well as peaks with high spatial kurtosis (>1 SD above mean GFP kurtosis), as can result if one or two electrodes have high amplitude artifacts. From the remaining set, in keeping with past work[16], we extracted 1000 time points for the subgroup map condition; we extracted 500 time points per recording for the combined map condition, thereby ensuring that all segmentation analyses were based on the same total number of data points (i.e. 30×500 = 15×1000 = 15,000). For segmentation only, we concatenated GFP peaks across individuals and applied a polarity-insensitive *k*-means clustering with number of distinct states K=4, following refs[1,19].

Automated labelling and alignment of microstates

This resulted in four microstate maps for each of three conditions: a single combined set ($N = 30$), and for each null replicate, a "case" ($N = 15$) and "control" ($N = 15$) set, each arbitrarily ordered

and labelled 1, 2, 3 and 4. To align similar maps across conditions, we identified pairings between each set of four maps (1-4) and an external, labelled set of maps (A-D) previously derived from a larger set of equivalent EEG recordings. Considering all possible pairings (e.g. 1→B, 2→A, 3→D & 4→C), the optimal set was selected to minimize the total distance $\Sigma_i(1-r_{ij})^x$ where i=1,2,3,4 is the observed map, j=A,B,C,D is the putative matched template for that map, and r_{ij} is the corresponding spatial correlation. By default, we set x=2 (i.e. to place more weight on a single dissimilar pair of maps, versus multiple only slightly divergent pairs, when calculating distance) although similar results were obtained with x=1. Each observed map was matched with exactly one template map and vice versa. We tracked if a replicate had a "suboptimal" match with the template, such that a map was assigned to one canonical microstate (e.g. B) but in fact had a higher spatial correlation with a different canonical microstate in the template (e.g. C, implying that one of the three other maps had an even higher spatial correlation with C). On average, 4.6% of assignments exhibited this type of pattern (more so for B, less so for A and C). Finally, as well as matching all three sets (i.e. case, control and combined maps) against a labelled template, we also matched each case and control map against the combined (N = 30) map, with equivalent results in terms of type I errors (data not shown).

Testing for group differences in microstate parameters

For the backfitting to derive per individual microstate parameters, we labeled each time point in the original recordings with the closest microstate map (again, polarity-insensitive). Putative microstates of short (<20 milliseconds) duration were replaced by the next most likely state until all states were at least 20 milliseconds. Microstate duration was defined as the average amount of time a given map persisted before transitioning to another microstate. Microstate coverage was the portion of the data that was mapped to each microstate. Microstate global explained variance (GEV) was the amount of variance in the data that was explained by that microstate. Microstate occurrence was how often a given microstate appeared. Spatial correlation was the average correlation of all assigned maps to the microstate. Group differences were assessed using an independent-sample t-test, after removing any outliers (+/- 2 SD units from the mean).

Testing for group differences in microstate topographies

Prior to generating random case/control groupings, we also estimated K=4 microstate maps separately for each individual (Figure 1). To assess topographical within- and between-group similarities based on these individual-level maps, we computed i) a 30-by-30 distance matrix, comparing each individual's map to all others (using distance as defined above), and ii) each individual's distance to the external template set of K=4 maps. Then, based on these distances,

we estimated a number of metrics to quantify potential global differences in microstate topographies between groups, separately for each null replicate: i) mean case-to-case distance, ii) mean control-to-control distance, iii) the absolute difference between these two values, iv) the ratio between the mean distance between phenotypically concordant pairs versus the mean distance between all phenotypically discordant pairs, v) mean case-to-template distance, vi) mean control-to-template distance, and vii) the absolute difference between these last two values. For each replicate, we randomly permuted case/control labels R=1,000 times to estimate the distributions of these seven statistics under the null hypothesis that differences in microstate topography are independent of case/control status. If Q is the number of times a permuted statistic was equal to or greater than the observed statistic, the associated empirical p-value was estimated as $(Q+1)/(R+1)$. As group labels were generated under the null (i.e. independent of microstate topography), we expected these seven tests to follow the distribution expected under the null, i.e. to be unbiased and conform to nominal type I error rates.

Figure Legends

Figure 1. Schematic of the study. Red violin plots are "cases" and black plots are "controls".

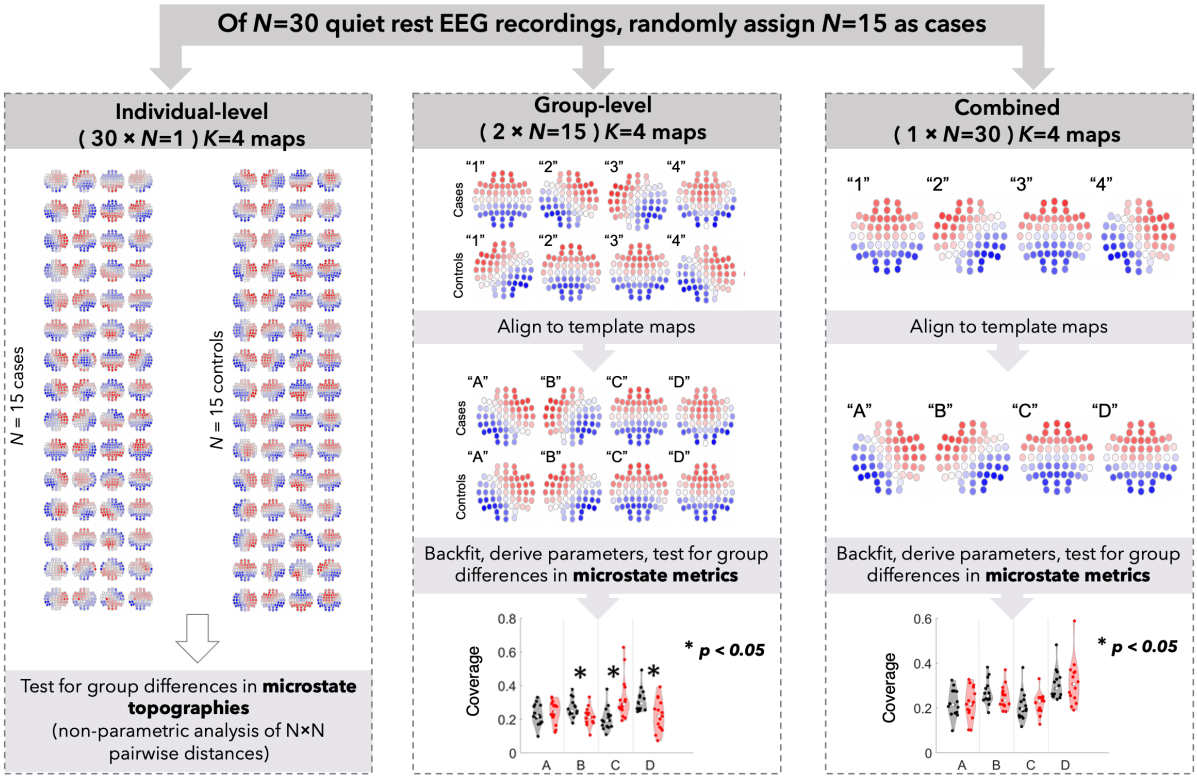


Figure 2. Subgroup maps show consistently higher type I error rates than combined maps. A) Proportion of permutations with false positives for each microstate and parameter. Gray bars are combined maps and black bars are subgroup derived maps. B) Histogram of p-values for all comparisons in all permutations. The dotted line indicates p = 0.05. There is a large peak of p values less than 0.05 In the subgroup derived maps.

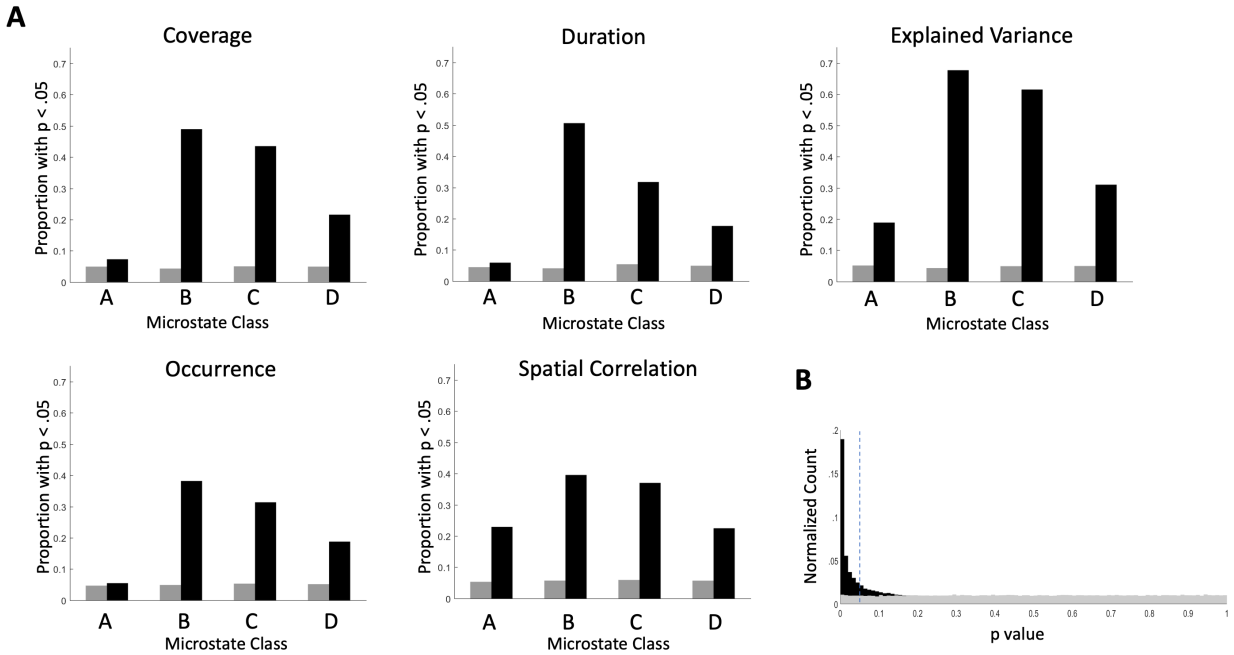
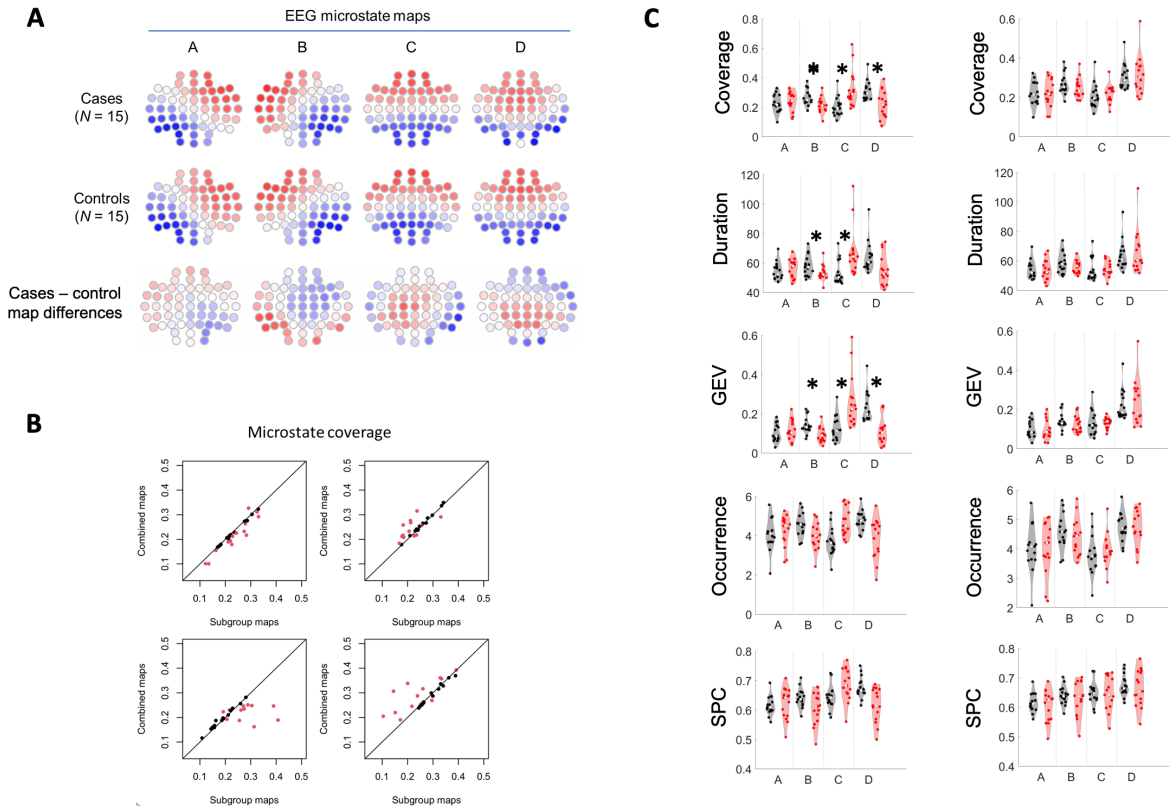


Figure 3. An example of elevated type I errors when using subgroup maps for a single permutation. A) Subgroup derived case and control maps and the difference between them. B) Correlations between the coverage metrics calculated with subgroup maps for cases (red) and controls (black) for each microstate. C) Coverage, duration, global explained variance (GEV), occurrence, and spatial correlation (SPC) for cases and controls with subgroup maps (left column) and combined maps (right column). * $p < 0.05$.



GRINS Consortium Members

Clinical Research Team: Jun Wang, Chenguang Jiang, Guanchen Gai, Kai Zou, Zhe Wang, Xiaoman Yu, Guoqiang Wang, Shuping Tan, Michael Murphy, Mei Hua Hall, Wei Zhu, Zhenhe Zhou

Molecular Genetics: Lu Shen, Shenying Qin. Hailiang Huang

Electrophysiology data analyses: Nataliia Kozhemiako, Lei A Wang, Yining Wang, Lin Zhou, Shen Li, Jun Wang, Robert Law, Minitrios Mylonas, Michael Murphy, Robert Stickgold, Dara Manoach, Mei-Hua Hall, Jen Q. Pan, Shaun M. Purcell

Project management: Zhenglin Guo, Sinead Chapman, Hailiang Huang, Jun Wang, Chenaugnag Jiang, Zhenhe Zhou, Jen Q. Pan

Principle Investigators: Mei Hua Hall, Hailiang Huang, Dara Manoach, Jen Q. Pan, Shaun M. Purcell, Zhenhe Zhou

Bibliography

1. Michel, C. M. & Koenig, T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage* **180**, 577–593 (2018).

2. Pascual-Marqui, R. D., Michel, C. M. & Lehmann, D. Segmentation of Brain Electrical Activity into Microstates; Model Estimation and Validation. *IEEE Trans. Biomed. Eng.* **42**, 658–665 (1995).

3. Murray, M. M., Brunet, D. & Michel, C. M. Topographic ERP analyses: A step-by-step tutorial review. *Brain Topogr.* **20**, 249–264 (2008).

4. Khanna, A., Pascual-Leone, A., Michel, C. M. & Farzan, F. Microstates in resting-state EEG: current status and future directions. *Neurosci Beiobhav Rev* 105–113 (2015). doi:10.1016/j.surg.2006.10.010.Use

5. Custo, A. *et al.* Electroencephalographic Resting-State Networks: Source Localization of Microstates. *Brain Connect.* **7**, 671–682 (2017).

6. Murphy, M. *et al.* Abnormalities in electroencephalographic microstates are state and trait markers of major depressive disorder. *Neuropsychopharmacology* 1–8 (2020). doi:10.1038/s41386-020-0749-1

7. da Cruz, J. R. *et al.* EEG microstates are a candidate endophenotype for schizophrenia. *Nat. Commun.* **11**, 1–11 (2020).

8. Tait, L. *et al.* EEG microstate complexity for aiding early diagnosis of Alzheimer's disease. *Sci. Rep.* **10**, 1–10 (2020).

9. Al Zoubi, O., Mayeli, A., Tsuchiyagaito, A., Misaki, M. & Zotev, V. EEG Microstates Temporal Dynamics Differentiate Individuals with Mood and Anxiety Disorders From Healthy Subjects. *Front. Hum. Neurosci.* **13**, 1–10 (2019).

10. Giordano, G. M. *et al.* Neurophysiological correlates of Avolition-apathy in schizophrenia: A resting-EEG microstates study. *NeuroImage Clin.* **20**, 627–636 (2018).

11. Nishida, K. *et al.* EEG microstates associated with salience and frontoparietal networks in frontotemporal dementia, schizophrenia and Alzheimer's disease. *Clin. Neurophysiol.* **124**, 1106–1114 (2013).

12. Michel, C. M. & Koenig, T. EEG microstates as a tool for studying the temporal dynamics

of whole-brain neuronal networks: A review. *Neuroimage* (2017). doi:10.1016/j.neuroimage.2017.11.062

13. Brodbeck, V. *et al.* EEG microstates of wakefulness and NREM sleep. *Neuroimage* **62**, 2129–2139 (2012).

14. Koenig, T. *et al.* A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest. *Eur. Arch. Psychiatry Clin. Neurosci.* **249**, 205–211 (1999).

15. Murphy, M., Stickgold, R. & Öngür, D. Electroencephalogram Microstate Abnormalities in Early-Course Psychosis. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 35–44 (2020).

16. Brunet, D., Murray, M. M. & Michel, C. M. Spatiotemporal Analysis of Multichannel EEG: CARTOOL. *Comput. Intell. Neurosci.* 1–15 (2011). doi:10.1155/2011/813870

17. Zhang, K. *et al.* Reliability of EEG microstate analysis at different electrode densities during propofol-induced transitions of brain states. *Neuroimage* **231**, 117861 (2021).

18. Kozhemiako, N. *et al.* Non-rapid eye movement sleep and wake neurophysiology in schizophrenia. *Elife* **11**, 1–33 (2022).

19. Poulsen, A. T., Pedroni, A., Langer, N. & Hansen, L. K. Microstate EEGlab toolbox: An introductory guide. *bioRxiv* 1–30 (2018). doi:10.1101/289850