

# Functional MRI brain state occupancy in the presence of cerebral small vessel disease -- pre-registration for a replication analysis of the Hamburg City Health Study

PCI RR, Revision 01

## Comments by the Editor

1. Testing against a null hypothesis of no association is arguably relatively uninformative, given a sample size of ~1500, because even a very weak relationship could be statistically significant at this sample size. It would be good to include a more nuanced a priori discussion of what size of effect would be of theoretical and/or practical importance in the present context.

**Many thanks for this comment. We agree that it is important to consider not only statistical but also theoretical/practical significance and to discuss what a minimal interesting effect size from that perspective would be. Since the field of dynamic functional connectivity and imaging-derived brain states is relatively young, such a discussion will necessarily remain somewhat speculative. (2020 Cornblath, Communications Biology) reported differences in fractional occupancy of DMN-related states between 6 and 7 percentage points between different task demands (rest vs n-back) in healthy subjects. This would constitute an upper bound for a minimal effect size of interest, but along the continuum of ischemic white matter disease severity, even smaller effects would be of importance. Given the high prevalence of cerebral small vessel disease and its known association with cognitive impairment, we argue that variations in state occupancy equaling half of the rest/n-back difference would be of interest, and include the following explanation in the revised manuscript:**

Sample size calculation is based on an effect size on the odds ratio scale of 0.95, corresponding to an absolute difference in the probability of occupying a DMN-related brain state between the first and third WMH-load quartile of 1.3 percentage points, and between the 5% and 95% percentile of 3.1 percentage points. Approximating half the difference in fractional occupancy of DMN-related states between different task demands (rest vs n-back) in healthy subjects, which was estimated to lie between 6 and 7 percentage points (2020, Cornblath), this value represent a plausible choice for the smallest effect size of theoretical and practical interest. It also equals the effect size estimated based on the data presented in (2022, Schlemm).

2. The power analysis is predicated directly upon the central estimate of effect size from Schlemm et al (2022). It seems likely that this effect size will overestimate the true population effect, because it has been selected non-randomly as the target of this

replication (i.e. you chose to propose a replication for this effect, and not for effects that did not come out as significant in the previous analysis). It would be more realistic to use a lower-bound estimate of the effect size or, to motivate the smallest effect size of interest in the present context a priori (as noted above).

**We appreciate this comment and agree that non-random selection of effects for replication carries the risk of overestimating their true size, thus biasing the power analysis. The 95%-confidence interval of the estimate of the central effect from (2022 Schlemm) chosen for replication was 0.915 - 0.983. Rather than using this lower bound we argue for a smallest effect size of interest of 3.1 percentage points absolute difference in FO across the cSVD spectrum (5% and 9% quartiles of WMH load distribution), corresponding to half the difference seen between rest and n-back task conditions in healthy subjects.**

3. It is stated that the multiverse analysis will be run if the hypothesized association can be replicated using the primary analytical choices. If the point is to test the robustness of the outcome to different analytical choices, it is not clear to me why this is only relevant if the main hypothesis test is significant.

**The aim of the proposed procedure was indeed to inform about, not to test, the robustness of the outcome of the test of the primary hypothesis to different analytical choices. The rationale behind planning the multiverse analysis only if the primary null hypothesis is significant was the desire to create a more severe test procedure. We agree that this rationale was misguided and did not take into account the conceptual issues laid out in the next point. We therefore revised the manuscript to explain in more detail our inferential strategy and to plan the multiverse analysis as exploratory, independent of the outcome of the main analysis.**

No hypothesis testing and will be carried out in these [exploratory] multiverse analyses. They rather serve to inform about the robustness of the outcome of the test of the primary hypothesis. Any substantial conclusions about the association between severity of cerebral small pathology and time spent in high-occupancy brain states, as stated in the Scientific Question in Table 1, will be drawn from the primary analysis using pre-specified methodological choices.

4. This leads into a deeper conceptual issue for your Stage 1 plan. You wish to preregister a multiverse analysis, but you state that “For each combination... we will quantify the association between WMH load and average time spent in high-occupancy brain states... no hypothesis testing and therefore no adjustment for multiple comparisons will be carried out.” (The statement that no adjustment for multiple comparisons will be made may not be necessary here, because – as you note – it is entailed by the fact that no hypothesis testing is being done.)  
The problem here is that you do not specify how the results of the multiverse analysis will determine your conclusions. The point of preregistration (at least in a formal registered Report) is not just to say what analyses you will do, but to specify unambiguously all of the logical linkages between possible outcomes and theoretical conclusions. The title of your manuscript suggests that the multiverse analysis is the focus of your study; but, within the methods, it is an additional exploratory step which is contingent on the outcome of the main hypothesis test and will serve a descriptive

and not an inferential purpose.

So, if your main purpose is your inferential replication, then the multiverse analysis should probably be dropped from the Stage 1 plan, unless it can be sufficiently thoroughly specified to elevate it to a proper part of the registered plan (which could be challenging, given the vast range of possible outcomes). It could still be added at Stage 2 as an additional exploratory analysis, but the format would require that your conclusions could not be substantially driven by non-registered parts of your analysis. On the other hand, if your main purpose is the multiverse analysis (as your title suggests), then there may be relatively little benefit in the Registered Reports format for your study, because your approach is to cover the range of possible analysis pipelines rather than nail a specific one down. In a sense, Registered Reports and multiverse analyses can be considered to represent alternative strategies to achieve analytic transparency (notwithstanding that, in principle, it is possible to preregister a multiverse analysis).

**We are very grateful for this insightful comment. We agree that our submission was not clear enough on the descriptive vs inferential purpose of the multiverse analysis and lacked a specification of how conclusions would be drawn from statistical test results. We have removed reference to the multiverse analysis from the title of the manuscript and clarify in the text that the multiverse analysis is planned as an exploratory exercise not affecting our substantial conclusions.**

5. As a minor point, the legend for Figure 2 is insufficient for me to understand the Figure (e.g. how should we know which effects in this plot are significant? I do not see 36p in the symbol key...), and should be improved. In passing, I note that (if I interpret correctly) the vast majority of analyses return results that are smaller than the effect size reported by Schlemm et al (2022).

**Thanks for suggesting to improve the legend for Figure 2, which we have done. Effects are significant if their 95%-confidence interval does not contain zero, which is indicated more clearly now. The legend key refers to different brain parcellations; regression strategies, including 36p, are listed on the vertical axis. Most analyses using variants of the 36p regression model and not relying on the very coarse Desikan-Killiany or Harvard-Oxford atlas return results that are compatible with the effect size of (2022 Schlemm). That said, we agree that the majority of results are smaller.**

## Reviews

### Reviewed by anonymous reviewer, 07 Mar 2023 00:45

The author presents a pre-registration of a replication study to examine associations between dynamic resting-state fMRI, small vessel disease (WMH), and executive function (TMT-B). The research question is scientifically valid, but the theoretical rationale for the hypothesis (specifically regarding the focus on dynamic FC and high-occupancy states) requires further clarification and justification. The sample and proposed methods are mostly appropriate and feasible, but I offer some suggestions for improved clarity and rigor.

**We thank the reviewer for taking the time to carefully read and comment on our manuscript.**

1. Line 47: The author suggests that limited reproducibility of functional connectivity “in the presence of cSVD” might contribute to “heterogeneous” findings in the prior

literature. I believe it is important to acknowledge here that functional connectivity measurements are generally quite noisy with low reliability, even in healthy control samples, especially when measured with short (~5 minutes) resting-state scans (e.g., Laumann et al., 2015, Figure 4), which are quite common in clinical studies, including the present study. This is not a problem specific to cSVD or any clinical population.

**We thank the reviewer for this comment and agree that shorter resting-state scans are more prone to noisy connectivity measurements. We also agree that low reliability is a general problem not restricted to cSVD, but maintain that it might be exacerbated in clinical populations due to small samples, increased motion or indeed different brain properties, such as vasoreactivity. We have clarified the introduction as follows:**

[ll. 47-53] “Findings with respect to functional connectivity (FC), on the other hand, are more heterogeneous than their SC counterparts, perhaps because FC measurements are prone to be affected by hemodynamic factors and noise, resulting in relatively low reliability, especially with resting-state scans of short duration. This problem is exacerbated in the presence of cSVD and made worse by the arbitrary processing choices.”

2. The justification for choosing to focus on dynamic FC as opposed to traditional FC is not clear. Given the context of the paragraph beginning in line 40, it seems to imply that dynamic FC might be more “reproducible” or less “arbitrary” than traditional FC. However, this is counter-intuitive, as dynamic FC has not been demonstrated to lead to either of these improvements. If anything, dynamic FC is likely less reproducible than traditional FC in the sense that resting-state scans are broken into even shorter “state” bins. If the argument is that dynamic FC is more likely to be relevant for cognition, this claim has also been disputed in the literature, as dynamic shifts in FC have been suggested to arise from non-cognitive processes, which should be acknowledged as a potential interpretation of the data (Laumann et al., 2016; Laumann & Snyder, 2021).

**We appreciate this comment. We did not mean to suggest that dynamic FC is more reproducible or less arbitrary than traditional FC. Our speculation about “limited reproducibility” in line 48 of the original submission refers to FC in general in comparison to structural connectivity (see also R1.1). We also do not claim that dynamic FC is more relevant for cognition than traditional FC, but that it is an exciting development worth exploring in the context of white matter disease and cognitive impairment.**

**We agree that non-cognitive processes may contribute to measurable changes in FC. Indeed the functional network dedifferentiation we aim to replicate is unlikely to arise from moment-to-moment fluctuations in cognitive processes, but rather from the structural changes associated with cSVD.**

**We clarified the Introduction as follows:**

[ll. 56-62] “While the study of dynamic FC measures may not solve the problem of limited reliability, especially in small populations or subjects with extensive structural brain changes, it adds another -- temporal -- dimension to the study of functional brain organisation, which is otherwise overlooked. Importantly, FC dynamics do not only reflect moment-to-moment fluctuations in cognitive processes but are also related to brain plasticity and homeostasis, which may be impaired in cSVD.”

3. What is the expected final sample size after excluding participants who were already included in the prior study?

**(2022 Schlemm) analyzes 988 participants from the HCHS. At the time of planning the current replication, data from 2562 are available. We therefore expect the final sample size to be about 1500 participants, with no overlap to our prior study. [l. 89 of the revised manuscript.]**

4. The Trail Making Test - part B is selected as a measure of executive function. However, unless performance is adjusted for part A, performance on part B alone may be driven in large part by psychomotor processing speed (Arbuthnott & Frank, 2000).

**We agree that the TMT-B/A ratio may be considered a purer measure of executive function than TMT-B. We prefer to use the TMT-B as our primary cognitive endpoint because it reflects executive function and processing speed, both of which are known to be affected in cSVD. We have clarified this choice throughout the manuscript.**

**As per R2.1, we plan to explore multiple cognitive endpoints in the Stage 2 report, and will also include the TMT-B/A ratio there.**

5. The proposed clustering analyses will use a default  $k$  of 5, presumably based on the result from the author’s previous paper. However, I believe it is important to test whether this result replicates in the proposed analyses. A range of  $k$  values should be tested and compared to find the best fitting number of clusters, rather than selecting a value a priori.

**The reviewer is correct in assuming that the choice of  $k=5$  clusters is motivated by our previous work. This number has also been identified and validated in (2020 Cornblath). The proposed strategy of testing a range of  $k$  values again and selecting the number giving the best fit is, in our opinion, at odds with the principle of a registered report. In particular, unless examined in a comprehensive simulation study including the data-dependent selection of  $k$ , it is unclear how such an approach would affect the statistical properties of the proposed test procedures. Also, interpretation of test results becomes more difficult and less generalizable, when aspects of the tested hypothesis are defined using information from the data.**

**We therefore propose to keep the a-priori value of  $k=5$  clusters for the pre-registered main analysis. We are open to exploring a range of  $k$  values as additional analyses in the manuscript.**

6. In addition to the demographic variables, variables relating to resting-state fMRI data quality and motion (e.g., mean FD, % frames retained in scrubbing) should be reported and included as statistical covariates.

**We appreciate this suggestion and agree that it will be helpful to report information on image quality and head motion, which we plan to do. We are reluctant to prespecify motion covariates as nuisance variables for our primary regression analyses. Reasons include the fact that different regression strategies would need to be adjusted for different motion parameters (e.g. scrubbing % not present in all confound regression strategies), making analyses less comparable. More fundamentally, it is not clear in what way the association between WMH load and state occupancy would be confounded by motion, given that state occupancy is not directly linked to connectivity strength. It seems plausible that the identification of brain states could be affected by motion artifacts, but adjusting for motion parameters in the k-means clustering procedure is not well established.**

**We therefore propose to not include motion variables as statistical covariates in the proposed analysis and investigate the effect of head motion on brain state identifiability in future work.**

7. Table 2. It is not clear which pipelines will include GSR, e.g., GSR should be included in the Power and Satterthwaite pipelines if consistent with the references, but it is not listed in the table.

**Thanks for catching this error. All 36p pipelines, including the Power and Satterthwaite pipelines, should include GSR. Since the global signals are contained in the 36 motion parameters, we decided to not explicitly mention GSR for these designs.**

8. Hypotheses regarding the dynamic FC are mostly framed in terms of high vs. low occupancy. For instance, the separation of 2 high-occupancy and 3 low-occupancy “states” is shown to be robust in the multiverse analyses, but are the regional network configurations of these states also consistent? Is the optimal solution of  $k = 5$  consistent? These questions seem to have more importance for theoretical interpretation, whereas the observation of greater occupancy in states defined as being “high-occupancy” is nearly circular. Why is it meaningful that WMH would be associated with dwell time in a state defined by its occupancy rate, as opposed to a state defined by its pattern of network configuration?

**We appreciate this comment. We chose to frame our hypotheses algorithmically in terms of occupancy in order to minimize manual interference in identifying the states of interest, and thus maximize reproducibility of our analysis. Based on (2020 Cornblath) and (2022 Schlemm), these states are expected to be characterized by activation or suppression of the default mode network, and it is this pattern of network configuration which makes the hypothesized association meaningful. We agree that testing whether high-occupancy states have higher occupancy states than low-occupancy states would be circular, and no such hypothesis test is proposed. As an outcome-neutral quality check, the separation between high- and low-FO states will be checked, as shown in Figure 2. While it would be interesting to explore how the high-occupancy states identified using different parcellation and regression strategies differ in terms of their network configuration, this is not the focus of the proposed study. As stated under “Further exploratory analyses”, we plan to describe and visualize the alignment of brain states with pre-defined functional networks.**

## References

- Arbuthnott, K., & Frank, J. (2000). Trail Making Test , Part B as a Measure of Executive Control : Validation Using a Set-Switching Paradigm. *Journal of Clinical and Experimental Neuropsychology*, 22(4), 518–528. [https://doi.org/10.1076/1380-3395\(200008\)22](https://doi.org/10.1076/1380-3395(200008)22)
- Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J. un, Chen, M. Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U. F., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., & Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, 87(3), 657–670. <https://doi.org/10.1016/j.neuron.2015.06.037>
- Laumann, T. O., & Snyder, A. Z. (2021). Brain activity is not only for thinking. *Curent Opinion in Behavioral Sciences*, 40, 130–136. <https://doi.org/10.1016/j.cobeha.2021.04.002>
- Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Berg, J. J., Greene, D. J., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2016). On the Stability of BOLD fMRI Correlations. *Cerebral Cortex*, 1–14. <https://doi.org/10.1093/cercor/bhw265>

## **Reviewed by Olivia Hamilton, 03 Apr 2023 14:07**

Reviewer comments on a registered report entitled “Functional MRI brain state occupancy in the presence of cerebral small vessel disease-pre-registration for a multiverse replication analysis of the Hamburg City Health Study”.

N.B. I do not have sufficient practical knowledge of the imaging methodology described in this report to be able to assess its suitability for the intended application. However, the proposed study is a replication of a previously published article (Schlemm, 2022) and will use the same imaging pipeline, so I assume the proposed method has already passed peer review at Biological Psychiatry. Still, the handler at PCIRR might wish to recruit another reviewer for this article who can better cover the imaging aspects.

I enjoyed reading this well-written and well-thought out registered report. The author proposes to replicate their own previous findings (Schlemm, 2022) in a sample from the Hamburg City Health Study:

Greater white matter hyperintensity (WMH) volume is associated with fMRI-derived brain states of high fractional occupancy (i.e. the proportion of BOLD volumes assigned to each brain state) Less time spent in high-occupancy states is associated with poorer scores on the trail making task part B (a test of executive function).

The analysis plan appears sound. The methods seem appropriate to be able to test the stated hypotheses and seem sufficiently detailed that the study might be replicated (with a reminder of the above caveat that I am not able to assess the imaging parameters).

**We thank the reviewer for their time and careful reading of the manuscript. We address their comments below.**

### **Main comments/questions:**

1. The original study (Schlemm, 2022) tested associations between brain dynamics and multiple domains of cognitive ability. Would it not be beneficial to test the same associations in the present replication study, as were carried out in the original article, rather than to attempt replication of a single significant result? The increased statistical power in the present study might reveal associations that were not detected previously (which would be very interesting in itself), or might return null results, which would be in line with the previous article.  
The results of this work might be more compelling if the replication were to be carried

out in a different dataset entirely. Did the author consider carrying out these replication analyses on any alternative samples (i.e. not the HCHS)?

**We appreciate these suggestions. We agree that it would be interesting to re-examine associations between brain dynamics and other measures of cognitive ability than executive function / processing speed as measured by the TMT-B. However, we expect the sample size of this new dataset to be insufficient to test a large number of hypotheses with adequate power. Furthermore, since our previous analysis has not shown significant associations between brain dynamics and, e.g., word recall, we see little justification to include it as a pre-registered analysis in the proposed project. We do, however, agree that it would be interesting to include associations between brain dynamics and other domains of cognitive ability as non-pre-registered exploratory analyses.**

**The suggestion to make our work more compelling by working with a different dataset is intriguing. We did consider this possibility and decided against it in order to maximize the chance of replicating the key finding of (2022 Schlemm) by keeping the same imaging and analysis protocol. We understand that this choice limits the generalizability of a possible positive replication. Further replication studies using different datasets might be carried out in the future.**

2. The authors note that missing data patterns will be reported. How will missing data be handled? Apologies if I've missed this.

**We will perform a complete-case analysis. This is now mentioned explicitly in II. 151-152 of the revised manuscript.**

3. Are there any additional exclusion criteria beyond not having imaging data, failure of pre-processing, or inclusion in Schlemm et al., 2022? I'm wondering whether participants with non-SVD-related brain/cognitive disorders will be excluded?

**Thanks for allowing us to clarify this point. We had originally planned not to use any additional exclusion criteria. Upon further deliberation we have revised the plan to exclude subjects with a radiologically confirmed intra-axial intracranial space occupying lesions. Other possible pathologies, such as inflammatory or ischemic residues, or known dementia, will not be excluded.**

4. The author states that participants will be excluded from analysis if automated processing using Freesurfer or fmriPrep fails. Failure to pre-process through these automated pipelines might be more likely for those with more severe SVD (due to the presence of additional visual SVD markers such as lacunes, microbleeds etc) and may risk biasing the sample (towards milder cases of SVD). Do the imaging/data team perform any manual checks on any of the imaging data at any point, or on those that fail the automated pipelines? If so, it would be useful to detail this in the report - apologies if I've missed this.

**We appreciate this suggestion. In our previous experience with the HCHS sample we have not noted any substantial problems with the imaging pipeline used in (2022 Schlemm) and proposed for the current project. The imaging team does perform manual quality checks on the data, but we chose not to use those in order to ensure replicability of the pipeline. That said, we will look at images that failed processing and, if possible, provide details on why and at what stage the processing failed.**

5. I note that this registered report only has one author. From the previously published



article (Schlemm, 2022), I can see that other researchers have likely been involved in the preparation of this work and should probably also be credited as authors on this report.

**Absolutely! The current replication analysis was planned by the single author of the submitted stage 1 manuscript, partly because some of the other researchers involved with (2022 Schlemm) have, in the meantime, worked with data from HCHS participants we propose to analyze, e.g., (2022 Petersen, NeuroImage). They could not, and were not, therefore, be involved in the preparation of this work, but will contribute to carrying out the proposed research and be credited as authors on the final Stage 2 report.**

**Minor comments:**

6. The author mentions that they will exclude the sample who were included in the previous (Schlemm, 2022) analyses. As the sample size of the previous study (Schlemm, 2022) is known, it would be helpful to include it in the current report.

**The sample size of the previous study is reported as n=988 n 151.**

7. It would be helpful if fractional occupancy was defined in the introduction, to save the reader having to hunt for it further down in the article.

**We agree that such an early definition would be helpful and have included it as follows:**

[Il. 69-71]: “The fractional occupancy of a functional MRI-derived discrete brain state is a subject-specific measure of brain dynamics defined as the proportion of BOLD volumes assigned to that state relative to all BOLD volumes acquired during a resting-state scan.”

8. The author states that ‘gender’ is included as a covariate. Do you perhaps mean sex, rather than gender? See this guidance from the ONS.

**We very much appreciate this comment. Indeed, sex might be the more appropriate covariate to include. In the HCHS, participants were asked both about their sex and their gender, in a way similar to the ONS guidance. In the previous sample, all participants answered either male/man or female/woman. This lack of diversity might indicate limitations with regards to the representativeness of the sample which, if manifested, will be discussed in eh stage two manuscript.**

**Additional question:**

9. Is the author planning to test potential mediation of the association between WMH and cognitive ability, via brain dynamics? I’m not suggesting that they include this as part of this report – I’m just interested.

**Thanks for this question. No, we do not plan to test mediation effects of brain dynamics between WMH and cognitive impairment. We consider mediation a causal concept, about which no reliable information can be inferred from this cross-sectional dataset.**