We thank the reviewers for their comments and discuss the points that have been found problematic.

In terms of communication, the difficulty of the paper is to disclose a null result, ie the absence of gain of causal methods in a the context of multivariate models of lesion-behavior mappings. Null results are harder to establish !

the experimental results did not fully support the conclusion

- The results clearly show that in average (Fig. 3), multivariate models such as random Forests outperform alternatives, in particular linear or causal models.
- What is clearly visible in all figures (paper and suppmat), and is the message of the paper, is that model assumptions (amount and nature of non-linearity in the outcome generation) matter more than the causal nature of the model. We agree thus with R3 that the opposition between causal and non-causal methods is not essential.
- It is also important to note that random variations obscure the results in some cases (supp mat Figs), so that averaging across target regions is necessary to conclude. We thus agree that **causal model performs better in some cases and worse in others**, but there is no strong support for their use in the context of lesion-behavior mapping.

We will clarify the point in the paper: namely that i) important fluctuations exist, and that our conclusions are drawn on average across challenging cases ii) that the nature and amount of non-linearity are actually more important than causal vs non-causal methodology.

lack of a clearly defined modeling goal

We wrote this paper because we noticed that causality has been wrongly invoked as the method to use for brain-lesion mapping. This would be the case if i) some confounding were present in lesion/behavior relationship, and ii) this confounding were captured by other brain regions. We show (lemma 1) that this is not consistent with the classical formulation of brain-lesion mapping. As predicted by theory, we then observe that a standard multivariate model is able to cope with the system identification, and does so more accurately in average than a causal model.

This leads to the conclusion that "causal models" should not be invoked as the solution to lesion/behavior mapping, but only in cases where observed confounders are present ; such cases do not correspond to lesion-behavior mapping models used so far (see discussion).

potentially biased data generation process

Reviewer R3 does not explain what he/she means with bias? To avoid self-fulfilling prophecies, we have decided to systematically rely on the existing simulations schemes proposed in the domain from ref[12] in the paper, as we think that this is a reasonable standpoint.

unclear causal modeling approach

The use of observational causality is meant to address the question of confounding. Hence the whole paper is aimed at analyzing the potential impact on confounders in the problem of lesion-behavior mapping. We have not restated the usual assumptions (such as causal faithfulness) for the sake of brevity, but these are implicit, as in most studies. We will add that.

We use Pearl's formalism to describe the causal structure but turn to causal treatment effect theory to get tractable and statistically sound estimates. The two models are logically equivalent, as stated by Pearl himself in section 4.5 of "The Causal Foundations of Structural Equation Modeling", 2012 (https://ftp.cs.ucla.edu/pub/stat_ser/r370.pdf)

Regarding causal methods We used BART and doubly robust estimators. For mildly high dimensional problems BART is the reference method for this type of problem, as it has outperformed alternatives in the ACIC 2016 and 2019 competitions.

Interpretation of Table 1 (R3):

Table 1 provides a result on Recovery (accuracy in identification), which is the objective of the present work. It is not related to explained variance.