Many thanks for the helpful comments, including the “significant impact this method can have on the current Covid-19 pandemic”, “of great interest to the NeurIPs community” and “The authors do an excellent job at summarizing ... their extensions on a very intuitive level...follow without getting lost”. Below we address all reviewers’ points and provide **additional simulation** and **comparison experiments**.

R#1: The method used here provides qualitative instead of quantitative results, in the sense that it can detect the causes of a target even in presence of latent confounders, but not the causal strength. Epidemiological studies in Germany focused on the patient 0 and the first 15 infections in Bavaria, but not on the wider spread. While this leaves us without epidemiological ground truth, we did consult reports from the RKI institute on events that could have contributed to the spread. Since the paper submission, we have presented/discussed the work with virologists/epidemiologists. They were intrigued by the fact that causal inference provides tools that work already on the relatively weak data (case numbers). We will be happy to update the paper to reflect these discussions.

R#2: 1. Had we not extended the theory to account for non-sink targets, when applying SyPI on them, it would result in less detected causes: Th 1 would still prevent false acceptances, but Th 2 would no longer provide necessary conditions for all unconfounded targets. We relaxed this strict assumption allowing Y to have descendants which do not belong in its candidate causes $\mathbf{X}$. In the Covid-19 dataset, we approach this assumption in practice by assuming that only the infections of the regions that occurred before the target belong to its candidate causes. This way, Y can have descendants in the observed (and unobserved) time series, but not in the subset that contains its candidate causes.

2. We agree that the more factors and samples become available, the more useful the method will be.

3. This is an incremental extension of a theoretical method with a twofold goal: 1. it aims at making the aforementioned method easily applicable to real time series data of the Covid-19 pandemic, so that it can later on be used when more data are available, 2. it provides a causal perspective of the current spread of the pandemic in Germany.

Clarifications: In Section 6.2.3 we provide Theorem B, which presents the conditions of Th 1 and 2 combined. The two theorems of SyPI, and the proposed extension theorem do not depend on the type of the variable (e.g., binary or not).

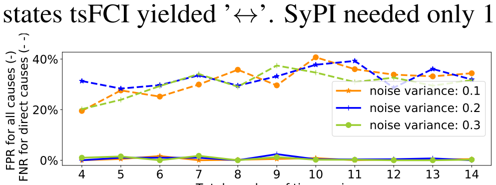
R#3: 1-a: In l 67 and 90 we deliberately omit the “$_t$” index to denote the whole time series, which is in line with our notation in l 59-60. 1-b: As we explain in Section 2.4 we ran two different experiments: first at the federal state level, and then at the district level. In both cases the time series are the daily reported Covid-19 infections in a (federal state or district) region. We assign every time one regional infection series to be the target Y, and all the remaining regional time-series that have reported infections before the target to be the pool of its candidate causes, from which SyPI will then identify the true causes. As we explain, we do this to comply with our proposed modified assumption $\mathbf{DE}^{\mathcal{G}}_{\mathbf{Y}} \notin \mathbf{X}$. For the federal state-level analysis, in addition to the regional infection series, we use as candidate causes the binary time series of the policies that were applied in the target state. The fact that a time series is assigned in the pool of candidate causes does not mean that it will indeed be a true cause. This is what the proposed method identifies. Therefore, the correct phrasing would be “if state A reported on 1/3 and state B on 2/3, then the daily infections of A will be used as a candidate cause of target B, and then SyPI will identify if indeed A causes B”. 1-c: We discuss this case in the last par. of Section 2.2. As we mention, SyPI relies on the stationarity of the causal relationships in the graph. If this is violated (i.e. it could be that the policies not only cause the reported infection time series but also be caused by it in different time windows), then the method will no longer correctly detect the causes.

2-a: Please see added simulations in point 3 below. 2-b: As there is no ground truth, we can only provide evidence that our results seem meaningful. This is why in Section 4.1 we provide (admittedly limited) information about the location of the detected causes with respect to airports and major events that took place, as well as comparison of our findings about the role of the political interventions with other methods which used similar dataset [9].

3: For the potential violation of the proposed assumption please see answer to point 1. of R#2.

R#4: 1. The main difference between the proposed approach and tsFCI is that SyPI pre-calculates a very concise conditioning set for each target and only requires two conditional independence (CI) tests per candidate cause, to decide if it is a true cause of the target. In contrast, tsFCI performs exhaustively CI tests for all possible combinations of conditioning sets and lags, which results in very ambiguous statistical results and very large computational times in large graphs. Of course, tsFCI aims at the full graph discovery and not only at causal feature selection (SyPI). This also justifies tsFCI’s more computationally intensive conditions.

2. As was requested, we performed **additional comparisons with tsFCI** for the infections in the federal states. For fair comparison we used the same threshold for all the statistical tests of both methods (0.05). Due to lack of space here we describe the results and we will add the figure in the manuscript. tsFCI detected 8, while SyPI 44 directed edges (causes). 4 of the tsFCI were a subset of the ones detected by SyPI. For the majority of the remaining states tsFCI yielded '$\leftrightarrow$'. SyPI needed only 19 seconds to run, while tsFCI needed 15 minutes for the same dataset.



3. As requested, here we provide 33 **experiments on simulated graphs** (100 graphs/experiment) with 1000 samples, varying noise and number of time-series, with two hidden, allowing the target to have descendants that do not belong to its candidate causes. The FNR for direct causes (dashed) remains below 40% as in [6], and the FPR (continuous) is close to 0. As expected from the proof, SyPI’s performance was not affected.