We thank the reviewers' time and energy for reading our paper and providing all these comments and suggestions. We are also appreciative for the positive feedback, including that this work has "solid/important theoretical results" (R2/R5), "meticulous, thorough" development of the graphical characterization (R4), and the complete algorithm that "looks inspiring" (R3). Below, we address some of the concerns raised by the reviewers.

Comparison to (Kocaoglu et al., 2019) [14] (R2). We provide a detailed comparison of our work to that of [14] in Appendix D.3. Unfortunately, we could not include this material in the main paper due to the lack of space. In short, we consider the problem of learning from interventional data with *unknown interventional targets*, a setting which cannot be handled by the work in [14]. When considering the results in [14], we make the following contributions:

1. We formulate the Ψ-Markov property and derive a graphical characterization that subsumes that of [14], which is formally shown in Proposition 7, Appendix D.3.
2. We show that the algorithm introduced in [14] is not applicable under unknown interventional targets and present a complete algorithm (Ψ-FCI) under unknown interventional targets (Example 10).
3. We handle causal sufficiency as a special case of the derived results. Subsection 3.1 establishes a graphical characterization for this case, and Section C, in the Appendix, presents an algorithm for learning an equivalence class from interventional data under causal sufficiency. We prove this algorithm to be complete for both known and unknown interventional targets. The work in [14] does not discuss the causally sufficient case nor completeness.

C-faithfulness versus Faithfulness (R2 & R5). As noted by R2, c-faithfulness is indeed stronger than the faithfulness assumption, which is common for learning Markov equivalence classes from when only data from one distribution, the observational one is available. A similar assumption is also required in other works [10, 18, 32, 14], but it is phrased slightly differently depending on the specific setting. The preliminary experimental results in Appendix D suggest that c-faithfulness largely holds, especially in the discrete case. However, we do understand the concern of R2 that c-faithfulness may not hold in some settings, in a similar fashion as the faithfulness assumption in the observational case. The present work provides necessary conditions to establish the theoretical limitations of inferring causal invariances from the combination of multiple datasets, and it paves the way for heuristic and approximation algorithms that may weaken the c-faithfulness assumption, akin to weakening faithfulness in the observational case (e.g., Zhalama, Zhang, J., Mayer, W. (2017). Weakening faithfulness: some heuristic causal discovery algorithms. JDCA, 3(2), pp. 93-104).

Further Empirical Evaluation (R4 & R5). We agree that additional experiments would be helpful to refine the theoretical understanding under finite samples and other empirical constraints. In this spirit, we have conducted experiments on Sachs data [26] after the submission. We indeed observe discrepancies in the recovered structures compared to JCI [18]. We will provide and discuss these findings in relation with the known ground truth. We also conducted synthetic experiments with two other generating causal graphs (Fig.1(a) and Fig.4(a) with discussions in Ex.6 and Ex.8, respectively) and the results look similar to the ones we included in the paper (Fig.6(a)). For the camera ready, we will conduct more synthetic experiments on random causal graphs and include the results.

Theorem 4 in the Appendix (R4). Thm. 4 establishes that a tuple of distributions that is generated by a causal graph satisfies the Ψ-Markov property relative to the graph and the true set of interventional targets (lines 174-176). This ascertains that the equivalence class learned by Ψ-FCI is non-empty (Theorem 2), i.e., the equivalence class is guaranteed to include at least the generating causal graph and the true unknown set of interventional targets. We will make this point more prominent by referencing the theorem in the main text (instead of only the section, line 176).

Testing the Distributional Invariances (R5). There are different ways of implementing hypothesis testing for the distributional invariances discussed in line 22 of Ψ-FCI, which can be seen as evaluating statements like $|\hat{P}_i(y|w) - \hat{P}_j(y|w)| \leq \epsilon$, where the hat represents the empirical distribution. Ψ-FCI is agnostic to the particular implementation of the test, which is in general chosen based on the specific details of the setting. Still, for concreteness, when the support of two distributions $\hat{P}_i(y, w)$ and $\hat{P}_j(y, w)$ is the same, testing whether $\hat{P}_i(y|w)$ is equal to $\hat{P}_j(y|w)$ can be done as follows. First, define a binary variable F which is set to 0 for a sample from i and set to 1 for a sample from j. Then, test if $I(F; Y|W)$ is zero. In our experiments, we use the previous method to test the distributional invariances.

Hard Interventions (R5). The presented characterization and algorithm are sound under hard interventions; however, the equivalence class can be further refined in this setting due to the change in the adjacencies of the graphical model following from the do-intervention. To understand the subtlety, consider the graphs $\mathcal{G} = \{X \rightarrow Y\}$, $\mathcal{D} = \{X \leftarrow Y\}$ and $\boldsymbol{\mathcal{I}} = \langle \{X\}, \{Y\} \rangle$. The pairs $\langle \mathcal{G}, \boldsymbol{\mathcal{I}} \rangle$ and $\langle \mathcal{D}, \boldsymbol{\mathcal{I}} \rangle$ are Ψ-Markov equivalent according to Thm. 1. However, the graphs are distinguishable under hard interventions since $(X \not\perp\!\!\!\perp Y)_{\mathcal{G}_{\overline{X}}}$ while $(X \perp\!\!\!\perp Y)_{\mathcal{D}_{\overline{X}}}$. Given this realization, we opted to avoid discussing hard interventions since the results are somewhat direct but much weaker.

Comparison to (Rothenhäusler et al., 2015) [25] (R5). We briefly mentioned this work in the introduction, lines 68-69. We ended up not doing a detailed comparison to Ψ-FCI given the very nature of both approaches, they are not really comparable. On the one hand, [25] considers the broader class of cyclic causal models while ours is restricted to acyclic models. On the other hand, our work makes no assumption about the functional form or type of soft intervention, while [25] considers linear causal relations and shift interventions. We will reflect this discussion in the paper.