

1 We thank all reviewers for the constructive feedback. We address all concerns below.

2 **Response to Reviewer 1** » not rigorous enough... the real-world dataset size ...
3 biases ... physician evaluation size.

4 Due to the costly nature of medical data sets, the original data set size is not big, as
5 you pointed out. However, we note that this is in fact a strong motivation to build
6 such a simulator. Also, as you mentioned, this is the first work to build a simulator
7 that can generate data with causal relations—the data set size will naturally grow
8 in the future and this work will hopefully inspire improved simulation systems.
9 Moreover, a small data set size does not seem necessary to introduce bias. (Here
10 by "bias," we mean bias in the estimated parameter values.) The real-world data
11 set is collected by using two months of clinical records. Our simulation data
12 satisfy distribution properties of the real-world data set. This is further suggested by both of the updated and original
13 physician's evaluations.

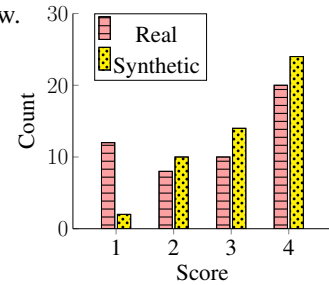


Figure 1: New physician's evaluation results of 100 data examples.

14 Regarding the physician evaluation, the samples given to the physician for evaluation were also sampled uniformly,
15 which is unbiased. Also, we have added another 40 examples (20 simulation records and 20 real-world records), which
16 increased the physician evaluation from 60 examples to 100 examples. The new results with 100 examples are shown in
17 Figure 1. We find that the physician evaluation using 100 samples is consistent with the evaluation using 60 samples.
18 Both of the updated and original evaluation figures in the paper indicate that our simulator simulates realistic patient
19 diagnosis records where the physician cannot distinguish from real-life cases.

20 » The section 2.2... first the format of the real-world dataset. Thanks for the feedback. We will adjust the order and
21 introduce the format of the data set in the beginning of Section 2.2.

22 » ... the heuristic ... It is not heuristic, but rigorous. We clarify it below and also in the updated version of the paper.
23 Please note that components of c in equation (1) can be 1 or 0 (they are not c_i , which is 1); see the line below equation
24 (1). In words, our heuristic estimation considers the case where parents of a variable having value 0, e.g., we estimate
25 $P(X = 1 \mid Pa_1(X) = 0, Pa_2(X) = 1)$ with $P(X = 1 \mid Pa_2(X) = 1)$ which is also supported by the medical
26 insight that $P(X = 1 \mid Pa_1(X) = 0, Pa_2(X) = 1) \leq P(X = 1 \mid Pa_2(X) = 1)$.

27 » The causal relation is relatively simple ... three layers ... In our system, the complexity of the causal relationship
28 is mainly reflected in the coverage of different types of causal structures rather than the depth of the longest chain.
29 Our causal graph covers all the basic d-separation relations, i.e., the chain structure, the collider structure, and the
30 common cause structure. Additionally, domain knowledge and physician's experiences may reduce the complexity of
31 the ground-true causal graph structure, but it will not reduce the complexity of the task, which recovers the true causal
32 graph structure from a complete/empty graph.

33 » ... broader scenarios ... not just neuropathic pain... As you pointed out, this work is the first step towards this goal
34 of systematic causal discovery evaluation. We hope and believe that this work will inspire the development of other
35 simulators with various types of causal relationships.

36 **Response to Reviewer 2** Thanks for your encouraging comments. » ... venue ... Thanks for sharing your concern.
37 Your proposal is great. At the same time, because of the urgent need of simulators for causal discovery method
38 evaluation, we believe that this work will benefit a large range of causality researchers in the NeurIPS community.

39 » ... a layout of the complete graph ... We have presented the complete graph in the form of cause-effect node pairs in
40 the appendix. Due to the size of the graph, we will visualize it using an online graph visualization tool and attach the
41 link to the updated version of the paper.

42 » ... test for the PC and FCI ... For PC and FCI we used Chi-squared test. We will make it explicit.

43 **Response to Reviewer 3** » ... around 800 causal relationships ... In this work, we directly used the diagnostic terms
44 from physicians instead of codes such as ICD codes. Also, we differ the left and right side for both symptom diagnostic
45 terms and pattern diagnostic terms. This leads to around 200 nodes and 800 edges.

46 » ... implementation ... We will open source the entire simulator, including the ground-truth causal graph, the estimated
47 parameters, the simulator for data generation, and examples under different scenarios shown in the paper.

48 » ... generalizability ... time series. Although our simulator has already covered many practical challenges in the causal
49 discovery domain, we agree with the reviewer that it would be of higher impact to further generalize it, for example, to
50 time series. This is feasible but needs some real-world time-series data to support the parameter estimation. We are
51 working closely with medical doctors to enlarge the cohort, and generalizing it to consider time series is one of our
52 future work direction.