Our thanks to each of the reviewers for their work and their suggestions. We discuss their questions below, and we will include such discussions in a final version of the paper.

Generalisations beyond the Ising Model (Reviewer 1) While the generic techniques of Le Cam's method, and in our case the lifting trick of §3.1 should naturally extend to any graphical model, most of the calculations that offer specific control on χ^2-divergences are strongly affected by the law of the graphical model under consideration. It is, of course, for this reason that most papers in this space commit to studying either Ising models or Gaussian MRFs (GMRFs), in that these are both natural models with relatively tractable calculations. For the particular choice of GMRFs, we feel that the same constructions should extend naturally to show similar results w.r.t. the separation of the sample complexity of recovery and testing. This is due to the (non-rigorous) intuition that the pairwise properties of GMRFs behave similarly to high-temp Ising models (with obvious caveats), where we have established these effects.

Experimental Validation in the Tree Setting (Reviewers 2 and 3) This is a very valid point, and should be tractable. We will include such an experiment in a final version of the paper.

Elaboration of Proposition 1 (Reviewer 2) The main thing that allows comparison is that risks for $\mathbb{SL}$ and $\mathbb{EOF}$ are defined in terms of the probability of an error event. The intended reductions are as follows.

1. Suppose we have a $\lfloor s/2 \rfloor$-approximate structure learner with risk δ. Then we can construct the following $\mathbb{EOF}$ estimator with the same sample costs: Take a sample from Q, and pass it to the structure learner. With probability at least $1-\delta$, this gives a graph $\hat{G}$ that is at most $\lfloor s/2 \rfloor$-separated from $G(Q)$. Now compute $G(P) \triangle \hat{G}$ ($G(P)$ is determined because P is given to the $\mathbb{EOF}$ tester). By the triangle inequality applied to the adjacency matrices of the graphs under the Hamming metric, this identifies $G(P) \triangle G(Q)$ up to an error of $\lfloor s/2 \rfloor$.

2. For $\mathbb{GOF}$, we can use a scheme for $\mathbb{EOF}$ as follows: take enough samples so that $\mathbb{EOF}$ can be solved with risk at most $\delta/2$. Take the samples from Q and pass them to the $\mathbb{EOF}$ solver. With probability at least $1-\delta/2$, if $G(P) = G(Q)$, then this solver must output a graph with at most $s/2$ edges, which if they are separated, this must output a graph with at least $s/2$ edges. So, thresholding the number of edges in this output yields a $\mathbb{GOF}$ tester. Net risk is $\delta/2$ for size and for power, giving sum risk δ.

We must thank you for bringing this up, because writing out this proof shows that there are small bookkeeping errors in the definitions (the complexities of $\mathbb{SL}$ and $\mathbb{EOF}$ should be defined as the level needed to get error $1/8$ and not $1/4$, since going from $\mathbb{EOF}$ to $\mathbb{GOF}$ is doubling the risk in the above, and the $\mathbb{EOF}$ risk should penalise errors of at least $s/2-1$ instead of $s/2$ to make the thresholding work out correctly). These do not affect the results besides small constant corrections, but are important to get right.