We thank all the reviewers for their feedback. We will incorporate them in the future version.

Overall, reviewers agree that our paper is technically sound. Both Reviewers 1 & 3 acknowledge that our setup is novel and well-motivated, and our techniques and results are non-trivial and significant. R2 has some questions about our setting and comparison against previous work [22], which we can clarify.

Problem Setting (R2).

Our problem setting lies at the intersection of active learning and counterfactual inference, and has connections to learning under sample-selection bias and covariate shift. All of these very well-studied problems in machine learning that are far from solved. The novelty in our setting lies in using active learning to reduce the need for labels in counterfactual inference.

"the learner knows the logging policy": This is a fairly standard assumption in prior work [1,5,19,22] and holds for example in online advertisement. In many applications where it does not hold, one can estimate the logging policy by fitting a model (for example, Athey et al, “Approximate residual balancing: debiased inference of average treatment effects in high dimensions.” and references therein). Note that this fitting can be done from unlabeled data only.

"the number of unlabeled examples is at most the size of logged data set": We look at this setting mostly for simplicity, and our algorithm will work (with minor modifications) with more unlabeled data so long as the labeling budget in the online phase is limited. Our algorithm is most useful in a setting where the learner has already collected some logged data and would like to query a few more labels to build a classifier for the population. Allowing unlimited data and labeling budget in the online phase will lead to trivial solutions that do not use the logged data.

Comparison with Prior Work (R2)

Novelty and significance: We apply and analyze two variance control techniques (regularization and clipping), and demonstrate how to combine them with the disagreement-based active learning (DBAL) framework in order to derive a better sample selection bias correction method. Note that combining DBAL with regularization is technically non-trivial, and the outcome (regularized DBAL) may be of independent interest.

The role of $\tilde{\theta}$: Our results are in line with a long line of prior work on active learning theory. Prior work shows that changing the interaction mode or setup in active learning typically leads to a constant factor improvement in label complexity – which is measured by a modified form of the disagreement coefficient – see for example Zhang and Chaudhuri. "Active learning from weak and strong labelers." NeurIPS 2015; Huang, et al. "Active learning with oracle epiphany." NeurIPS 2016.

Value of q_0 ("the results are only meaningful if q_0 is not too small"): Quite contrary, because we do active learning, our method does work well even if q_0 is small (unlike passive learning solutions). In particular, our label complexity depends on an average term $\mathbb{E}[\frac{1}{1+\alpha Q_0(X)}]$, while the label complexity of [22] and many other baselines is either proportional to $1/q_0$ or $\frac{1}{1+\alpha q_0}$ which can be much worse.

Experiments (R1 and R2)

The main contribution of this paper is a new algorithm with theoretical analysis. We agree that it would be interesting to see how the proposed algorithm works practically, and we will add some experiments to the final version.

Discussion of Results (R1, R2, R3)

Gain from the clipping technique: The exact gain from the clipping technique depends on the data distribution. We provide a concrete example (Example 30) in our paper. We will make it clearer and provide more quantitative analysis in the final version.

Difference between error bounds for the variance regularizer and the second moment regularizer: The error bound is about $\tilde{O}(\sqrt{\frac{1}{m}\mathbb{E}\frac{1[h^*(X)\neq Y]}{Q_0(X)}})$ with the second moment regularizer while $\tilde{O}(\sqrt{\frac{1}{m}\mathrm{Var}(\frac{1[h^*(X)\neq Y]Z}{Q_0(X)})})$ with the variance regularizer. The latter is smaller, but the difference is almost negligible since $\frac{1}{m}\mathbb{E}\frac{1[h^*(X)\neq Y]}{Q_0(X)} - \frac{1}{m}\mathrm{Var}(\frac{1[h^*(X)\neq Y]Z}{Q_0(X)}) = \frac{1}{m}l(h^*)^2$ diminishes as $m \to \infty$.

Parameters in Section 5.3: Thanks for pointing this out. We have provided some examples in Appendix (proof of Theorem 2, Examples 30, 31) , and we will elaborate them and make it clearer in the final version. In terms of R3’s question about the inequality at line 298, the difference between its LHS and RHS depends can be quite significant for some data distribution and logging policy. For example, if $Q_0(X) = q_0$ with a very small probability and is close to 1 elsewhere, then LHS is still about $\frac{\nu}{1+\alpha q_0}$ while RHS is about $\frac{\nu}{1+\alpha}$ which is much smaller if α is large and q_0 is small.