Reviewer 1 *[The] methodology combines multiple different ideas in causal inference (multi-headed deep learning models and targeted learning) in a novel way, and their empirical evaluation seems very strong.* Thank you for your support and your insightful remarks and suggestions.

claim that their methodology is stable because it does not involve any propensity terms in denominators. However [...] their learnt $\tilde{Q}$ function, which involves propensities in denominators in its second term (which is weighted by $\hat{\epsilon}$. This is a good point. The idea is that the model can use the ϵ term and the loss term to downweight the influence of extreme propensities. In the case $\hat{\epsilon} = 0$ the propensity score is not used for the estimate. We have clarified the language, and emphasized that insensitivity to extreme propensity-score values is an intuition.

The baselines in their evaluations are not completely clear. We have clarified this. We chose hyperparameters (mainly, the optimizer) to produce a stronger baseline version of TARnet (lns 231-235).

The authors claim that part of strength of model is insensitivity to very low/high propensity scores due to lack of propensity scores in denominators. However in their evaluations they exclude data with extreme propensity scores which makes this claim difficult to verify. In addition [...] We have clarified this. We used a standard trimming procedure (e.g., as in the expts of van der Laan and Rose). We have also added a table to the appendix comparing trimmed and untrimmed estimators—treg is substantially less affected than AIPTW/TMLE.

It seems weird that Equation 2.2 has no hyperparameter... We have clarified this. Indeed, there is a hyperparameter. We used an arbitrary fixed value (1.0) to avoid unfairly advantaging our method via hyperparam search.

claim that the third head in their model regularizes the model such that finite-sample performance should be improved..." What we meant is that there is no infinite-data advantage to this procedure. It's not clear what happens in small vs. moderately-sized data. We have clarified the language. We have added the suggested experiment to the appendix. We subsampled the ACIC data. It shows that the Dragonnet's improvement is more significant with smaller-sized data.

Reviewer 2 *My main concerns are the novelty. Both of the methods seem like we have A and B, so we can try to combine to see how it works.* Both proposed methods are new and non-trivial. R1 and R3 both agree the paper is novel.

Reviewer 3 Thank you for your comments. We've clarified where you requested. We address your main concerns:
Experiments We are deeply surprised by the dismissal of the experiments as 'limited'. The range and detail of experiments go well beyond the level typical for this area (e.g., the related work). The paper makes a good-faith attempt to test using (necessarily) semi-synthetic data that is (i) realistic, (ii) defined by an existing benchmark, and (iii) covers a wide range of simulation approaches (we use 101 ACIC benchmark simulated datasets). Further, we produce a strong baseline (improving on the published version of TARnet) and do not engage in any unfair hyperparameter tuning.

As suggested, we have made various clarifications to the captions of the tables and figures in experiments. We've also clarified the following: TMLE is combined with treg by plugging the $\hat{Q}$ and $\hat{g}$ values from targeted regularization into the TMLE. Figure 2 and 3 are out-of-sample with ACIC data. In 5.4, we divide according to absolute ATE estimation error for the baseline (<1 or >1, chosen arbitrarily as a small value). We've also changed notation away from "ϵ", which is overloaded. Our comment "targeted regularization essentially never hurts" refers to the dragonnet/treg combo, where degradation is very small (0.01) on average. Note table 3 gives improvement/degradation amount conditioned on improvement/degradation. Note, however, **there is a typo in table 3**. We miscopied some numbers; the +dragon values should be 54%, 1.42, 0.32. Note this is better than the incorrectly reported values. We apologize for the error.

Theory Our purpose here is to produce practical methods *inspired* by existing theory results. As we note in the paper, consistency of the neural networks is a key ingredient for asymptotic results. You correctly point out that we do not attempt to prove consistency. However, establishing such consistency for neural networks is an active area of research. Extending such results is a valuable direction for future work, but is outside the scope of the paper. We note that the theory inspiring Dragonnet is non-asymptotic and we do not make any claims about Dragonnet's asymptotic properties.

With respect to targeted regularization, you write, "*It seems that coupling the estimators for Q and g will in general lead to loss of consistency and of the double-robustness property. The authors claim that "consistency is plausible – even with the addition of the targeted regularization term" because "the model can choose to set epsilon to zero"*." Our point here is that if Q and g models are consistent without the targeted regularization, then they will also be consistent with it (thus, the conditions for good asymptotics are satisfied). The reason is that treg embeds the original (Q, g) model in a larger model class by introducing ϵ (with $\epsilon = 0$ the original model class). In detail, consistency in the original model means $\hat{Q} = \mathbb{E}[Y|x, t]$ and $\hat{g} = P(T = 1|x)$ at $n = \infty$. The treg model preserves finite VC dimension (we add only 1 param), so the limiting model is an argmin of the true (population) risk. The true risk for the treg loss has a minimum at $\hat{Q} = \mathbb{E}[Y|x, t]$, $\hat{g} = P(T = 1|x)$, and $\hat{\epsilon} = 0$. This is because the original risk is minimized at these values (by consistency), and the treg term (a squared error) is minimized at $\hat{Q} + \hat{\epsilon}H(\hat{g}) = \mathbb{E}[Y|x, t]$, which is achieved at $\hat{\epsilon} = 0$. This gives efficiency under the same consistency condition as TMLE/AIPTW. We have clarified this point in the exposition and added the proof. Note that there *is* a trade-off in the (finite-data) empirical risk objective or in the presence of model-misspecification (indeed, that's the point).