We thank all the reviewers for their insightful comments, suggestions, and references.

Reviewer 1:

Novelty of tandem loss: it is not new, but we were not aware of the prior work, we thank Reviewer 2 for bringing it up.

While most of the computed bounds are non-vacuous, they look to be not that tight. Some discussion of this would be valuable. Also a discussion of potential ways to obtain tighter bond values, or whether there is a fundamental limitation.

We provide some discussion in Sections 3.2 and 4.4. The major challenge is the estimation of the tandem loss, which is based on overlaps of OOB samples, which are small [see Section 4.4]. It could be that a better estimation technique could be designed in the future. Another limitation is the oracle bound. For example, in the independent case with the growth of the number of classifiers it converges to $4L(h)^2$, whereas $L(\mathrm{MV}_\rho)$ converges to zero [see Section 3.2].

(*) *Reproduction of C-bounds:* Reproduction of C-bounds requires definition of a margin. We believe that adding it to the body might divert the attention from the main thread of the paper, but we will be very glad to add a section in the appendix, where we introduce the margin, discuss the relation with the tandem loss, and provide the C-bounds.

Line 471: Yes, you are right, thank you!

Reviewer 2:

Thank you for providing references to "joint error" and Equations 7 and 8 in Lacasse et al., we will add them.

Reproduction of C-bounds: see our reply **(*)** to Reviewer 1.

(**) *Adaptation of the strategy from the "second C-bound form":* Note that in our application the first order loss and the disagreements are estimated on different subsamples. The first order loss is estimated on OOB samples, whereas the disagreements are estimated on overlaps of OOB samples and unlabeled data when available. It is not immediately clear whether joint estimation would give an advantage, we will look at it in future work.

Reviewer 3:

Posterior optimization

We agree that posterior optimization did not improve the test error. However, we note that in prior work posterior optimization was either impossible (in C-bounds, except for highly limiting cases of aligned posteriors in binary classification) or led to considerable deterioration of the test error (as we demonstrate for the first order bound). Therefore, we see absence of deterioration of the test error as a step forward relative to prior work.

1) compute the multi-class C-bound

The multi-class C-bound based on the w-margin with $w = 1/2$ (Corollary 1 of Laviolette et al.) is equivalent to our oracle C-tandem bound in Theorem 6. The values of empirical C-tandem bound are reported in the paper. We note that there may be multiple ways of going from the oracle to an empirical bound, but not all of them are directly applicable in the OOB setting, see our reply **(**)** to Reviewer 2. We also note that the general multi-class C-bound (Theorem 2 of Laviolette et al.) cannot be evaluated of the OOB setting, because the max operator in their definition of the margin in equation (3) cannot be exchanged with expectation and the estimation cannot be done using pairs of hypotheses.

2) Please, re-prove Lemma C.14.

Oh, sorry, we missed that the square was outside the expectation on the left and inside on the right when we canceled the terms. Thanks for catching it! The fix is easy. Instead of canceling, take $-\mathbb{E}\left[X\right]^2 \varepsilon^2$ to the right hand side. Then $\mathbb{E}\left[X\right]^2 \varepsilon^2 - 2\varepsilon\mathbb{E}\left[X\right]\mathbb{E}\left[X^2\right] + \mathbb{E}\left[X^2\right]^2 = (\mathbb{E}\left[X\right]\varepsilon - \mathbb{E}\left[X^2\right])^2 \geq 0.$

3) move the posterior optimization experiment to the supplementary and insert some other results instead (Fig H.15?)

We will have an extra page if accepted, so we can have both in the body.

4) Line 281

The kl-inequality has an upper and a lower inverse, which give an upper and a lower bound, respectively. We have used the lower bound.

Thanks a lot for the references to additional work on multiclass classification!

Reviewer 4:

lines 95, 97, and 98: You are right, thanks!

line 109: We will add extra brackets, thanks!