We thank the reviewers for their constructive suggestions. We have (1) added simulations for (i) Skinny Gibbs, (ii) different hyperparameter choices, (iii) coverage of our VB credible sets, and (2) expanded the discussion in the final version to address the reviewer comments. A summary of the added discussion is provided point-by-point below.

Reviewer 1: How much more work is needed versus linear regression. In linear regression, one can do exact computations using the Gaussian likelihood to yield precise oracle results [Ray and Szabo (2019)]. Such exact expressions are not available for logistic regression, so we must instead use a different test-based proof using general ideas from Bayesian nonparametrics (Section 10). The technical details are thus different (and more involved) here.

Novelty of the VB algorithm. A methodological novelty here is using Laplace slabs for the *prior* underlying the VB approximation, rather than Gaussian slabs as all previous works do, and we show this does better empirically (Sections 5 & 8). We agree that deriving the resulting CAVI algorithm is somewhat standard, but the Laplace slabs modify the usual Gaussian update equations and are needed for implementation/simulations. We emphasize our main contribution is to provide theoretical guarantees and show our VB calibration empirically outperforms existing (Gaussian) VB approaches, rather than novelty of the optimization algorithm. We have included these derivations for completeness.

Comment on the theoretical validity of the algorithm. To the best of our knowledge, convergence properties of CAVI is still largely an open problem in even simple models, see e.g. Plummer et al. (July 2020). It is empirically known that the VB optimization problem is typically difficult and non-convex, and that CAVI will not return a global optimizer. However, with proper initialization, one still often recovers a good VB posterior (as we see in our simulations). This is an excellent point, but unfortunately well beyond the scope of our paper.

Comment on the advantage of VB over frequentist approaches. A main advantage is access to variable inclusion probabilities and their credible sets, which are often as equally interesting to practitioners as estimation. Our VB approach performs well empirically for model selection, as demonstrated by its good FDR, TPR and coverage of credible sets, which we have now added to the simulations. We have also added discussion on this point, thank you.

Establishing the frequentist validity of VB credible intervals (i.e. Bernstein-von Mises type results). Bernstein-von Mises results have only been proved for VB in low-dimensional settings [Wang and Blei (2019)], where one can modify classical local asymptotic normality (LAN) arguments for parametric models. Since LAN expansions do not generally hold for high-dimensional models, including logistic regression, new proof techniques are required.

Reviewer 2: Missing references and MCMC method. We have added the missing reference for Skinny Gibbs [Narisetty et al. (2019)] and have added the method in our simulations. Skinny Gibbs is indeed an order of magnitude faster than MCMC using Stan, but generally 50-100 times slower than the VB methods. It provided broadly similar FDR and TPR as VarBVS. Wei and Ghosal (2019) was already cited in the manuscript. Thank you for this suggestion.

Reviewer 3: Practical relevance. While this is a general issue with theory, it can often inform practice. Many methodology/applied researchers are unaware that using light tailed (e.g. Gaussian) prior slabs can yield poor inference for *true* sparse Bayesian inference, while the situation is even more complicated for VB. Indeed, we are unaware of any existing VB papers for logistic regression *not* using prior Gaussian slabs. It is practically important to pick heavy enough slabs and our theory confirms that exponential tails (Laplace) are sufficient for estimation when using VB. These findings are fully reflected in practice, where our use of Laplace slabs consistently and significantly outperforms the usual Gaussian slab approach in almost all simulations (Sections 5 & 8).

Our theory also provides conditions on the design matrix, which include many common examples, under which sparse VB works. The non-asymptotic nature of our full results (Section 10) also confirm these lessons apply for reasonable sample sizes, as demonstrated by our simulations. While our contribution is clearly on the more theoretical side, we think the routine use of VB in machine learning, including for logistic regression, and the practical insights afforded by our results, mean a machine learning conference is the right venue for our work.

Explanation and efficiency of using the surrogate KL for CAVI. The use of a surrogate KL functional (arising from maximizing a lower bound on the marginal likelihood) is a standard technique for VB in Bayesian logistic regression, see e.g. Chapter 10.6 of the textbook Bishop (2006). Its performance and motivation have been studied in several papers, which we now cite more clearly [including Bishop (2006)]. We agree that we were too quick on this point and have expanded the explanation, as well as providing references to more extensive discussions. Thank you.

Reviewer 4: Add application based support with a large model and do a model assessment. It is known that for large variable selection problems, MCMC methods often mix poorly and we should not assume MCMC estimates are close to exact values [Carbonetto and Stephens (2015), Griffin et al. (2017)]. Hence VB methods have been extensively used in the literature. If required, we can add a large real-world dataset with several thousand features in the supplement.

Discuss sensitivity to hyperparameter selection. We have added a simulation study on this: we find that the choice of hyperparameter does indeed affect the small-sample behaviour, which we now report/discuss. Thank you for this point.