We thank the reviewers for their positive feedback and valuable suggestions. Overall, as the reviewers point out, our work is the first analysis of a nonzero-sum adversarial hypothesis testing model bridging multiple areas and is meant to be followed up upon. Our aim was to obtain a tractable game-theoretic model that is general enough to capture some of the problems that arise in adversarial classification yet simple enough to state concrete results on the behavior of any NE and we hope that our work is a first step in that direction. Yet, we greatly appreciate the reviewers suggestions of improvement and further work—below we discuss some of the main questions/comments in the reviews.

Computation of equilibria: We have indeed not looked into the problem of computing a NE in our game. We believe that obtaining the complete structure of a NE and computing it is a difficult problem in general because the strategy spaces of both players are uncountable (and there is no pure-strategy NE in general), and we cannot use the standard techniques for finite games; but it is definitely an interesting direction for future developments. We note, however, that in this paper we are able to show a partial structure of NE (the defender performs a certain likelihood ratio test at NE (Proposition 4.1) and the attacker's strategy concentrates (Lemma 4.5)) and that we are able to derive error exponents associated with classification error (at any NE) using a small set of assumptions and without explicitly computing a NE.
Existence of pure-strategy equilibria: In the numerical example of Section 5 in one dimension (see Appendix C.1 for the corresponding discussion), it appears that there is no pure-strategy NE for small n but there may be a pure-strategy NE for large n (though we did not prove it). Given the simplicity of this example, we are not sure what kind of natural assumption would lead to obtain existence of a pure strategy NE for all n but investigating such conditions is certainly an interesting direction; and we might be able to show that there exists a pure NE for large-enough n.
Defender's mixed strategy over Φ_n: In order to show the existence of an equilibrium, it is necessary to consider randomization over Φ_n. However, once the existence of a NE is established, then the test proposed by the reviewer will indeed achieve the same objective- this is shown in Proposition 4.1. If we do not randomize over Φ_n, it is not clear how one can establish existence of an equilibrium, since the objective functions are not quasiconcave—see Remark 4.2.

Assumption (A4): As the reviewers pointed out, we agree that this is in fact a strong assumption. This is the condition that naturally appears in the study of error exponents (see line 563 in the proof of Lemma 4.4); and we provide a numerical counter-example where our results do not hold when (A4) is not satisfied. Still, there could be a weaker assumption under which our results hold, but this needs to be further investigated and looking more precisely at what happens when (A4) does not hold is in any case indeed an interesting direction for future work. However, we believe that our work is a good starting point to understand the equilibrium behavior and obtain error exponents for our model.
Assumption (A4) for the Neyman-Pearson formulation: We agree with the reviewer that, in one dimension, the acceptance region of an optimal Neyman-Pearson test for a fixed alternative q will be a "vanishingly small neighborhood of the null distribution p" and that while it can still intersect Q for finite n, it may not for large-enough n; so that Lemma A.6 may always hold. However, it is unclear to us how this might generalize to higher dimension. Our intuition is that in higher dimension, the acceptance region may become close to p only in certain directions. We also note that our proof of Lemma A.6 actually uses Assumption (A4) and not a weaker version of it—see the expression of Γ_n in line 690. Overall, we believe that (A4) is needed in higher dimensions even for the Neyman-Pearson case; although it is possible that a weaker assumption will suffice in one dimension—we still need to check that carefully. We will include a discussion about this in the appendix.

Applications and attacker's model: Our model is relevant for problems that arise in the context of adversarial classification, as mentioned in Section 3.2 on model discussion; but our focus in this study was indeed more on developing a model that allows analytical investigation while containing the key elements of a nonzero-sum adversarial setting and that could be the basis of further works extending our results and our model. A model where the attacker changes the distribution gradually has been considered in a previous work by Brandão et al. [5]. However, they study a non-game-theoretic setting in the sense that they look for an optimal decision rule in which the adversary can generate each sample from a potentially different distribution (among a given set of distributions). Our nonzero-sum game-theoretic model on the other hand better captures the interaction between rational agents that may arise in some adversarial classification problems. We will look into a possible game-theoretic formulation for the model suggested by the reviewer with a suitable application in the future. We also plan to study a sequential version of our problem where data samples arrive over time and the defender can make a decision in an online fashion.

Gain from minimax strategy and Stackelberg equilibrium: We agree with the reviewer that it would be interesting to understand the gain we get from knowing that the game is nonzero-sum and in particular knowing $c(\cdot)$ (note that the gain in utility will be in the exponents). We note, however, that to obtain a completely meaningful comparison, we would need to model the information available to the attacker and the strategy that he adopts, which would lead to a significantly more complex incomplete information game. We also agree that looking at the Stackelberg equilibrium could be interesting and help solve computational issues, although we note that most of the security games literature using Stackelberg games assumes finite action spaces.

We thank the reviewers again for their encouraging feedback and numerous thoughtful suggestions for future extensions.