Thank you for your insightful feedback, corrections, and additional references that we will incorporate into our paper. Below, we address several key points.

All Reviewers (particularly R3): “scope is a bit narrow” & “significance seems limited” - Based on consistent feedback from the reviewers, we see that our presentation currently limits the perceived relevance and general importance of our work. In our introduction, we will emphasize that our work is directly applicable to any scenario that requires computing confidence intervals around importance sampling (IS) estimates. More broadly, we will also discuss that the community is interested in our definition of safety [38] and its limitations [39], and IS [4 (in paper), 40, 43, 44]. Lastly, we will mention that the $\alpha$-security formalization also pertains to high-confidence methods that do not use IS [31 (in paper), 41, 42].

R1: “generalize to continuous MDPs” - Our work generalizes to continuous MDPs, but care must be taken to select $\pi_c$ such that IS weights are bounded. For example, the diabetes treatment simulation discussed in the paper has continuous states and actions. We will make sure to discuss this extension in the paper.

“problems in applying Hoeffding’s inequality” - In order to use Hoeffding’s inequality, we assume that the IS weights are bounded. Although WIS is biased, it works very well in practice. Consequently, we introduced the notion of quasi-$\alpha$-security in Definition 2 to specifically allow for the analysis of WIS.

“if the attacker knows we are using Panacea” - The optimal attack does not change (lines 261–262), and therefore, Panacea limits the damage incurred by the attacker. We will move this discussion outside of the proof block.

R2: “include studies on how easy it is to trick those algorithms too” - We are definitely interested in pursuing follow-up directions to ensure security for model-based approaches, which we predict would be quite different and a significant contribution on its own. In a future work section, we will include a discussion of similarities and additional challenges that arise in that setting.

“the way $\pi_c$ is chosen” - We completely agree on the importance of how $\pi_c$ is chosen, even though the violation of safety comes primarily from the safety test. Our current definition of security assumes that $\pi_c$ has lower performance than $\pi_b$, but does not specify how often this occurs. Attacking the data used to select $\pi_c$ can increase this frequency. Notice that attacking the data used to select $\pi_c$ alone would not cause the safety property to be violated. We will add a discussion on this topic.

“whether the trajectory must still have been performed in the real environment” & “single out the few trajectories” - Because the transition and reward functions are not known, one can not distinguish real and fake trajectories. Rare events are critical to account for, and may look like fake trajectories. Perhaps impossible trajectories can be identified using domain-specific knowledge, but that must be analyzed on a per-domain basis. We will mention this in the paper.

“use of any $\alpha \geq 1$ would be pointless” - We see that we did not provide sufficient discussion of Table 1, saying that the behavior you note is what we aim to show! The middle column is usually $\geq 1$, indicating that standard methods can be completely broken (make pessimimal policies appear optimal) easily, as you described. However, the right column shows values of $c$ that make Panacea $\alpha$-secure for any $\alpha \in [0,1]$. E.g., plugging in $\alpha = 0.05$ gives Panacea a meaningful security guarantee. Note that if $c \leq 0$, Panacea is not useful, but that the values of $c$ are positive and grow quickly as $n$ grows relative to $k$.

R3: “a worst-case stand-in” - When the stakes are high – for example, in the application of RL to sepsis treatment in the intensive care unit, wherein training data is generated from hand-written doctors’ notes – we do not want to assume that the data contains only minor errors (such as patient height), but also major ones (such as wrong drug name).

R4: “an upper bound on the number of corrupted samples” - We will add a discussion of the many issues faced by practitioners, including estimating the number of corrupt trajectories (perhaps based on known error rates in the data processing pipeline of NLP and computer vision models) and selecting $\pi_c$. Panacea is only one piece of the puzzle, but provides guarantees that are informative to practitioners.