We thank the reviewers for their insightful comments. In addition to the changes described below, we will make clearer in the paper our contributions, algorithmic guarantees, and the conditions required for applying specific results.

Reviewer 1 (R1)
1. Non-asymptotic guarantees. R1 is correct that our proofs immediately translate to non-asymptotic bounds: In Appendix B, we bound all error probabilities by simple terms that depend only on α; simply adding up these terms (for their union bound) produces results for finite (but large enough for some inequalities to hold) α. For practical instances with low α, however, bounding dependencies in $n, r, k, q*$ and the $n_{f,v}$ only in terms of α can be lossy, and (as R2 rightfully points out) one would want to loosen quotas on some groups in exchange for lower failure probability. The only obstacle in getting bounds for general $n, r, k, q^*, n_{f,v}$ are the messy algebraic dependencies that don't make for a nice general formula. If the values of these parameters are specific numbers, it is easy to extract much sharper bounds from our proofs. ($*$) We will discuss these points in the paper.

Reviewer 2 (R2) (for discussion of point #2, see response to R1 #1)
1. End-to-end guarantee is unfair when q_i is a valuation. Foremost, we do not consider q_i as purely an agent's valuation for a panel seat, but rather as also capturing their *ability* to join the panel. Constraints on participation ability are documented in a survey by Jacquet[1], and include scheduling conflicts, social anxiety, and family/work. Secondly, the rule R2 proposes (one weakly increasing in q_i) is fundamentally incompatible with creating proportional panels. To see why, suppose men and women are split 50/50 in the population, but women have low $q_i = 1/4$, and men have high $q_i = 3/4$. Then, the pool is likely split about 25/75. If our algorithm sampled women with lower probability than men, women would comprise less than $1/4$ of the panel in expectation, while their proportional share is $1/2$.

3. Explored solutions limited to existing selection procedure. Our modeling choices are dictated by how panel selection is (and in many ways, must be) done in practice. In particular, forming a pool (step 2) is required due to limited participation, and the process of sampling letter recipients (step 1) is constrained by the fact that practitioners usually do not have detailed individual-level demographic population data, so systematically oversampling subpopulations that participate at lower rates is not practicable. This said, more complicated sampling methods that *do not* require such data (e.g., sending multiple rounds of letters) could worth exploring as an improvement on the status quo.

4. No baseline comparison in experiments. There are two reasonable baselines: uniform sampling, and the currently-used greedy algorithm. Uniformly sampling the pool will simply give each agent i an end-to-end probability proportional to q_i (up to negligible differences), since that the probability of them entering the pool is $r/n\, q_i$, so this baseline doesn't warrant experiments. We have done experiments showing that the Sortition Foundation's greedy algorithm gives substantially worse individual fairness guarantees than our algorithm, and ($*$) we will add these results to the paper.

5. Definition of fair representation (gerrymandering, prediction on sensitive covariates). It is true that group fairness guarantees (quotas) *alone* do not address gerrymandering concerns; this is a main reason that existing algorithms, which only guarantee quotas, can be very unfair. At least in terms of expected representation, our algorithm cannot have gerrymandering issues: since all agents' selection probabilities are $\approx k/n$, *any* subset of the population will be represented near-proportionally by linearity of expectation. To prevent gerrymandering in ex-post representation, we have seen practitioners use the cross product of features (in R2's example, gender and race) as a single feature. To R2's point about q_i prediction methods being perceived as discrimination: in the fairness in classification literature, using protected attributes to counteract inequality is an accepted practice. That said, the concern about errors in the q_is being seen as discrimination is relevant; ($*$) we will discuss this and all points made here in the paper/broader impact section.

6. Additional Beck-Fiala Results. Thanks for suggesting Banaszczyk's result: an algorithmic version is in [2,Thm 5.3]. However, we want the per-person marginal probability to deviate by only $\pm\delta k/n$, and this would incur a discrepancy of $O(\sqrt{|F|\log(n/\delta k)})$, which now depends on n. (We confirmed this with the author of [2].) ($*$) That said, we are happy to mention this related result.

Reviewer 3 (R3) (for discussion of point #2 (clarification on individual guarantees) see response to R2 #5)
1. Unclear why greedy is less fair. At a high level, the greedy algorithm is unfair because it permits gerrymandering (see R2 #5 and our response). In more detail, sampling each subgroup in greedy fashion (e.g. according to which quotas are furthest from filled) can result in people of certain groups having near-zero probability of being on the panel. ($*$) We can provide a simple example in the manuscript illustrating this problem, as well as experimental results (see R2 # 4).

Reviewer 5 (R5)
1. Not sold on the claim that q_i values are known or can be estimated. The reviewer is correct not to take on faith that the q_is are known by the algorithm; this is why, in our experiments, we show they can be estimated. We explicitly address the estimation of q_i values in lines 296–304. Additionally, in Appendix D.4, we provide several pieces of experimental evidence showing that, using data available to practitioners, we can estimate q_i values that fit the data well.

[1]Vincent Jacquet. Explaining Non-Participation in Deliberative Mini-Publics (2017). Eur J Polit Res 56.3 640–659.