Alternates, Assemble! Selecting Optimal Alternates for Citizens' Assemblies

ANGELOS ASSOS, MIT
CARMEL BAHARAV, ETH Zurich
BAILEY FLANIGAN, Harvard
ARIEL D. PROCACCIA, Harvard

An increasingly influential form of deliberative democracy centers on *citizens' assemblies*, where randomly selected people discuss policy questions. The legitimacy of these panels hinges on their representation of the broader population, but panelists often drop out, leading to an unbalanced composition. Although participant attrition is mitigated in practice by *alternates*, their selection is not taken into account by existing methods. To address this gap, we introduce an optimization framework for alternate selection. Our algorithmic approach, which leverages learning-theoretic machinery, estimates dropout probabilities using historical data and selects alternates to minimize expected misrepresentation. We establish theoretical guarantees for our approach, including worst-case bounds on sample complexity (with implications for computational efficiency) and on loss when panelists' probabilities of dropping out are mis-estimated. Empirical evaluation using real-world data demonstrates that, compared to the status quo, our method significantly improves representation while requiring fewer alternates.

1 Introduction

Motivated by mounting threats to democratic governance, new methods for facilitating public participation in democracy are taking root. One of the most prominent such methods is the *citizens' assembly*, in which members of the public are chosen randomly to serve on a panel. These panelists convene for several days to learn about an issue, deliberate, and then weigh in on what policy measures should be taken. Across continents, citizens' assemblies and other similar processes, falling under the broader umbrella of *deliberative minipublics*, are increasingly becoming a part of politics at all scales (e.g., see the 2021 Global Climate Assembly [Participedia, 2021], France's recent national assemblies [Bürgerrat, 2023, Giraudet et al., 2022], Ireland's constitutional convention [Irish Citizens' Assembly Project, 2019], Germany's national assembly on nutrition [Deutscher Bundestag, 2023], and the plethora of local and regional assemblies occurring worldwide [OECD, 2020]).

Because citizens' assemblies are costly per participant[1] and deliberation is difficult to facilitate at large scale, only a small fraction of the population can partake — typical panel sizes range from tens to hundreds. Consequently, *representation* is paramount: it is important that the panelists who partake can credibly represent the views of the broader populace. In practice, assembly organizers typically approach this problem by ensuring that the panelists satisfy proportionally representative *quotas*: upper and lower limits on how many panelists must be from various groups. Usually these quotas are set to achieve proportional representation of the population on many dimensions at once, including gender, age, race/ethnicity, education level, and some measure of political opinion. For example, on a panel of 100 people serving the state of Wisconsin, the quotas for *political opinion* might require that the panel contain 40-44 democrats, 40-44 republicans, and 12-20 undecided voters, mirroring the state's party affiliation rates of 42%, 42%, 16%. The quotas on *education level* might additionally require that 30-36 panelists have a college degree and 64-70 do not.

Over the past several years, computer science research has produced algorithms for randomly sampling a panel that is guaranteed to satisfy practitioner-chosen quotas of this form [Baharav and Flanigan, 2024, Flanigan et al., 2021a,b, 2024]. The panel selection process (both as it is studied in past work *and* as it typically works in practice) consists of two steps: First, thousands of invitations are sent out to the population, inviting them to partake in the process. Those who respond affirmatively form the *pool* of willing participants. Then, the final *panel* is randomly selected from the pool, and it must satisfy the quotas. Existing algorithms address the second step of this process, aiming to randomize within the quotas in a way that is maximally fair, transparent, and strategyproof.

This past work goes to great lengths — sometimes at the expense of other ideals [Flanigan et al., 2023] — to ensure that the panel satisfies representative quotas. However, past work does not address that this representation can be undone by what happens downstream: often, months pass between the selection of the panel and its convention, and in this time, some of those who were initially selected for the panel *drop out*. Many times, people drop out just before the panel convenes, or simply do not show up on the first day. These dropouts cause quota violations, which are especially problematic because they do not affect all groups equally: Figure 1 shows dropout rates among different groups across 25 citizens' assemblies. These trends show that certain groups — trending more financially and socially marginalized — tend to drop out at substantially higher rates. We see that young people, indigenous people, racial minorities, people with less permanent housing, and people with the least education drop out at disproportionately high rates — trends that largely persist across datasets collected from assemblies run in two separate countries by two separate organizations. This is not at all surprising: dropout is likely due to random shocks (e.g., illness, childcare falling through, last-minute work commitments, transportation issues), and

[1] Participants are usually compensated and, if the event is in person, their travel expenses are covered.

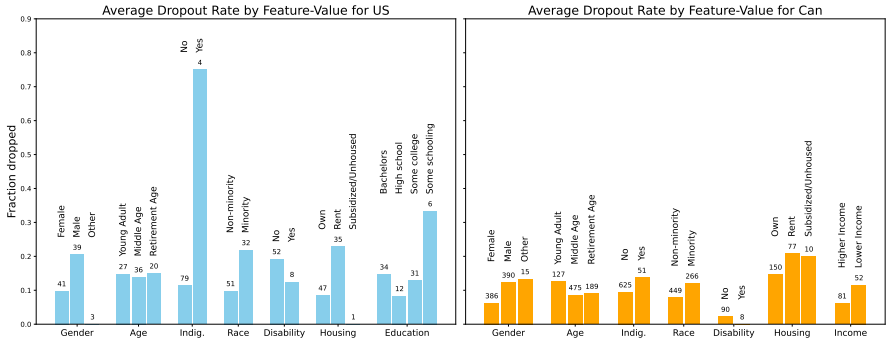

Fig. 1. Average dropout rates among different groups over 3 US assemblies (blue) and 22 Canadian assemblies (orange). Not all datasets contained all attributes; numbers above bars show how many people across datasets had that attribute. Data cleaning methods are in Appendix E.1.

those with less resources may be less able to overcome such shocks in order to attend the panel.

State-of-the-art: heuristics for selecting *alternates*. To hedge against major compromises in representation due to dropout, practitioners often select *alternates* — extra participants to have on standby to replace panelists who drop out. The key challenge is that the alternates must be selected *before* seeing who drops out; this is because by the time dropouts occur (just before or on the first day of the panel's convention), it is too late to recruit replacements for a multi-day process. Subject to this constraint, alternates are used in different ways across organizations: some organizations select an entire duplicate panel using the same method as was used to select the original, and then compensate these alternates for attending the panel's first day. Others directly inflate the size of the panel, imposing quotas that require extra panelists in groups they expect to drop out based on past experience.

Such methods are heuristic and have two main potential inefficiencies, both of which we aim to address in this paper. First, these methods either ignore differential dropout rates altogether, or they rely on practitioners to guess at dropout rates. Given the wealth of historical data on which kinds of people have dropped out from *past* panels, there is a natural opportunity to instead learn statistical predictions of different groups' likelihoods of dropping out. Second, these methods do not consider the problem's combinatorial structure (i.e., which *combinations* of attributes alternates should have), operating instead on the level of quotas, which are imposed on single attributes in isolation. Beyond their potential to lead to suboptimal representation, these inefficiencies can be costly: each additional alternate comes at a cost as they need to be compensated. These opportunities to improve the efficiency of alternate selection motivate our research question.

> ***Research question:*** *Given panelists' dropout probabilities — which are estimated from data and are thus subject to prediction errors — can we design a practical and performant algorithm that selects an optimal set of alternates of a given size?*

Problem setup. More formally, the inputs to our problem are the quotas we want to satisfy; the originally-selected panel K of size k, which respects these quotas; a pool N of size n from which can choose alternates; a budget $a \in \mathbb{N}^+$ limiting the number of alternates we can choose; and of panelists' *dropout probabilities* $\rho_i | i \in K$, describing each panelist's probability of dropping out. As will be technically consequential, we assume that panelists drop out *independently*, reflecting the reality that panelists are randomly chosen from a large population and unlikely to know each other, and thus the event that one drops out should not affect the dropout event of another. These dropout probabilities induce a *dropout set distribution* $\mathcal{D}$, describing the probability that each *panel*

subset drops out. Given all these inputs, the pipeline of selecting and deploying alternates works as depicted in Figure 2. Though our solution must engage with all steps 1-3 in the figure, our main goal is to design a procedure for step 1.

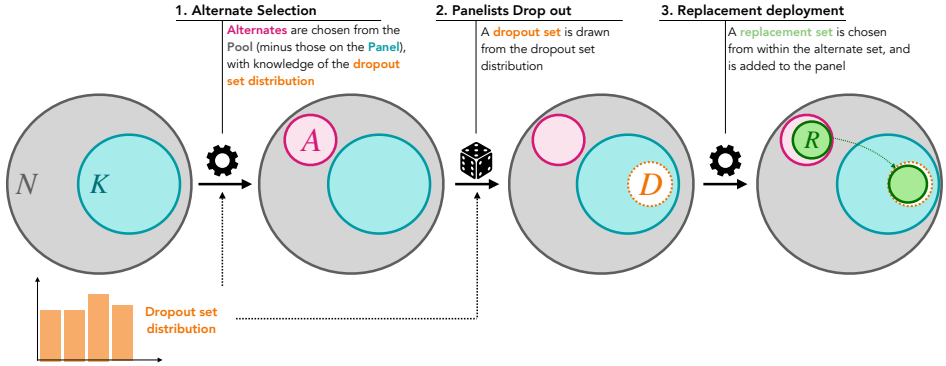

Fig. 2. The alternate selection pipeline. **1.** With knowledge of all inputs, an *alternate selection algorithm* chooses an alternate set $A \subseteq N : |A| = a$. Importantly, A must be chosen *before seeing the random draw of which panelists drop out* (the *dropout set* D), reflecting that alternates need to be placed on retainer ahead of time so that, e.g., they can be invited to the first day of the panel's convention in anticipation of some panelists not showing up. **2.** A dropout set $D \subseteq K$ is drawn from the dropout set distribution. **3.** The best possible *replacement set* $R \subseteq A$ is chosen replace D. The final panel is then $(K \setminus D) \cup R$. To ensure we respect the initial panel size constraint (for budgetary and other logistical reasons), we require that $|R| \leq |D|$, so that adding the replacement set does not exceed the original panel size.

Our goal is to compute the alternate set that can best restore the quotas after dropouts — in expectation over the randomness of which panelists drop out. The optimal alternate set A^* is then expressed (semi-formally) as follows, where the *loss* measures how well the quotas are satisfied:

$$A^* = \text{argmin}_{A \subseteq N} \sum_{D \subseteq K} \Pr_{D \sim \mathcal{D}}[D \text{ drops out}] \cdot \left(\min_{R \subseteq A : |R| \leq |D|} loss(K \setminus D \cup R) \right). \quad (1)$$

We next discuss the difficulty of computing A^*. For this, it is useful to have in mind that n is typically on the order of hundreds or thousands, k is typically on the order of tens or hundreds, and a may range widely, from < 10 up to k or more.

Key challenge. Solving Equation (1) presents a combinatorial minefield, consisting of three separate, nested tasks. First and most generally, we must compute the best alternate set (size a) within the pool (size n), which by brute force would require examining $\binom{n}{a}$ alternate sets. Secondly, we must evaluate at least one alternate set A, which requires computing its expected performance over all possible dropout sets, i.e., all subsets of the panel (size k). By brute force, this amounts to computing a sum with 2^k terms. Finally, to compute *each* term of this sum, corresponding to a specific dropout set D, we must determine how well A can replace D. This means finding the best subset of A, which requires looking at all subsets of A of size at most D, which is at least $\binom{a}{|D|}$. In fact this third problem — the sub-sub-routine among these three combinatorial problems — is *itself* NP-hard in asymptotic parameter a, as follows from an existing theorem (Appendix A.1).

Approach: *Integer Linear Programming* (ILP) and *Empirical Risk Minimization* (ERM). In light of our problem's NP-hardness, a natural choice would be to pursue a polynomial-time approximation algorithm. For practical reasons, we do not go quite so far: of the three tasks above, the two subset selection tasks (the first and third tasks) can be solved quickly and exactly via ILP

solvers. We therefore opt to present an algorithm that is not polynomial time but in exchange does not compromise on practical performance in these aspects. It is the *second* task — computing the expected performance of a given alternate set A over a distribution with support size 2^k — which does not lend itself to ILP solving. This is where we will pursue polynomial dependency on the problem's parameters, and in exchange forgo exact optimality.

For this second task, our approach makes use of the observation that, because agents drop out independently, we have *sample access* to the dropout set distribution even though its support is too large to write down: we can sample a dropout set $D \sim \mathcal{D}$ by flipping a ρ_i-biased coin for every panelist; those which come up 1s compose D. Taking inspiration from ERM, we use this sample access to construct an *empirical estimate* of $\mathcal{D}$, and then we solve our problem over this empirical distribution $\hat{\mathcal{D}}$ instead. We then use an ILP to perform both set selection tasks at once, choosing the alternate set $\hat{A}^*$ that optimizes the expected performance *over the empirical version of the distribution* — just Equation (1) where $\mathcal{D}$ is replaced with $\hat{\mathcal{D}}$. Ideally, at sufficient sample size, $\hat{\mathcal{D}}$ should be "similar enough" $\mathcal{D}$ that $\hat{A}^*$ should exhibit loss close to that of the optimal alternate set A^* with high probability. For this ERM-based approach to be a sufficient improvement on the brute-force method, it must be that we can ensure that the loss of $\hat{A}^*$ closely approximates that of A^* with only polynomially-many samples.

Contributions. We answer our research question affirmatively via the following series of results.

In **Section 3**, we define our ERM algorithms and analyze their sample complexity. We first formally connect our problem to the PAC learning model, which will allow us to leverage known learning theory bounds. We draw this connection for two standard loss functions: the *binary loss*, capturing whether A can perfectly restore the quotas; and the *linear loss*, capturing *how far* A is from restoring the quotas. Then we show that for the ERM algorithm using either loss definition, only $O(a \log n)$ samples are sufficient to ensure that $loss(\hat{A}^*) - loss(A^*) \leq \varepsilon$ with high probability. This bound is the result of tight bounds on the dimensionality of our hypothesis classes. This very low sample complexity is important, because we must solve an ILP whose size depends on it.

In **Section 4**, we analyze the extent to which our two ERM algorithms (using binary and linear loss, respectively) are robust to inaccuracies in the dropout probabilities. While our two loss functions were indistinguishable on the basis of sample complexity, here a distinction emerges: the binary loss can be arbitrarily non-robust to small errors, while the linear loss is extremely robust, its loss growing linearly with the prediction error.

In **Section 5**, we evaluate our entire algorithmic pipeline on data from real citizens' assemblies. We first estimate dropout probabilities from historical data, and then we compare the performance of our binary- and linear-loss ERM algorithms against heuristic benchmark algorithms emulating those used in practice. We find that while our estimates are likely imperfect in some instances (due largely to our data being highly suboptimal for this task), our ERM algorithms have lower loss, more consistent performance, and greater robustness to mis-estimates than our heuristic benchmarks.

In **Section 6**, we first discuss how our algorithms extend to the more general case where *alternates* drop out, and to various other ways of hedging against dropouts beyond selecting an alternate set. Finally, we close by situating our algorithmic solutions — and the problem of alternate selection in general — within the broader goal of *randomly* selecting citizens' assembly participants.

Related work. The application domain of this paper, *citizens' assemblies*, is the subject of substantial recent computer science research [Baharav and Flanigan, 2024, Benadè et al., 2019, Ebadian et al., 2022, Ebadian and Micha, 2023, Flanigan et al., 2021a, 2020, 2021b, 2024]. However, existing work on this application focuses on the the task of *sortition*, the random selection of the original panel, leaving untouched the issue of subsequent panel dropouts. Methodologically, our use of the PAC learning framework to (probably, approximately) solve an NP-hard problem is analogous to the

approach used by Peters et al. [2022] to compute a (probably, approximately) optimal allocation and pricing of rented rooms for prospective tenants. Finally, the key challenge of our problem, estimating a sum over 2^k possible dropout sets, is reminiscent of the core challenge in computing the partition function for graphs or other combinatorial objects. While our analysis does not utilize this technical connection, typical methods for estimating partition functions also often involve sampling techniques, though these techniques differ from ours [Ma et al., 2013].

2 Model

An *instance* of the alternate selection problem $\mathcal{I}$ consists of a set of *quotas*, a *panel* that satisfies these quotas, a *pool*, a vector of *dropout probabilities*, and an *alternates budget*.

The ***panel*** K is a set of the k agents who were originally chosen as assembly members (we often call these agents *panelists*). The ***pool*** N is a set of n agents that is disjoint from K. The pool corresponds to the agents that were willing to participate in the assembly but were *not* chosen, and thus are now available to be chosen as alternates. We will refer to individual agents as i.

To define the quotas, we first define ***features*** and ***feature-values***. A feature is an attribute category, such as "gender" or "age". Each feature f takes on a predefined set of feature-values V_f, an exhaustive and mutually exclusive set of categorical values of that feature. For example, the feature f = "age" might take on the values $V_{age} = \{< 50, \geq 50 \text{ years old}\}$. Formally, a feature is a function $f : (N \cup K) \to V_f$ that maps each agent to their value for feature f. In practice, the feature-values of all agents in K and N are known, and we assume this is the case. We will often use f, v to refer a feature, value pair (e.g., *age*, < 50). We express the set of all feature, value pairs as $FV := \{f, v | f \in F, v \in V_f\}$.

On each $f, v \in FV$, there are upper and lower ***quotas*** describing the maximum and minimum number of agents with value v for feature f the panel should contain. These upper and lower quotas are denoted as $u_{f,v} \in \mathbb{N}$ and $l_{f,v} \in \mathbb{N}$, respectively, where $u_{f,v} \geq l_{f,v}$ and $u_{f,v} > 0$ (otherwise, the group defined by f, v is effectively dropped from the instance). We summarize these quotas as $\boldsymbol{l} := (l_{f,v} | f, v \in FV)$ and $\boldsymbol{u} := (u_{f,v} | f, v \in FV)$. We say that a set of agents S *satisfies the quotas* iff

$$\sum_{i \in S} \mathbb{I}(f(i) = v) \in [l_{f,v}, u_{f,v}] \qquad \text{for all } f, v \in FV.$$

We assume the original panel K satisfies the quotas.

Finally, our ***alternates budget*** $a \in \mathbb{N}$ is the maximum number of alternates we can choose. An instance is formally expressed by the tuple $(N, K, \boldsymbol{l}, \boldsymbol{u}, a)$. In instance $\mathcal{I}$, we will let $\mathcal{A}(\mathcal{I}) := \{A \subseteq N : |A| = a\}$ be the space of all possible alternate sets (we will drop the $\mathcal{I}$ when clear from context).

Dropouts. Each panelist $i \in K$ has some ***dropout probability*** $\rho_i \in [0, 1]$, describing the probability that they drop out after the original panel K is selected. As discussed in Section 1, panelists do not know each others' identities prior to the panel (and thus their dropout events should not be related), so we assume that each agent i drops out independently; that is, if $X_i \in \{0, 1\}$ is the indicator of the event that i drops out, then the X_i's are independent and $Pr[X_i = 1] = \rho_i$. We summarize these dropout probabilities in the vector $\boldsymbol{\rho} := (\rho_i | i \in K)$. Let $D := \sum_{i \in K} X_i$ be the ***dropout set***, the random variable describing the subset of the panel that drops out. Note that $D \in 2^K$ (where 2^K is the power set of K), i.e., any subset of the panel can drop out. Let the ***dropout set distribution*** $\mathcal{D}_{\boldsymbol{\rho}} : 2^K \to [0, 1]$ be the distribution over dropout sets induced by the dropout probabilities $\boldsymbol{\rho}$. Then, by our assumption of independent dropouts,

$$\mathcal{D}_{\boldsymbol{\rho}}(D) = \prod_{i \in D} \rho_i \cdot \prod_{j \in K \setminus D} (1 - \rho_j).$$

While $\mathcal{D}_{\boldsymbol{\rho}}$ is the *true* dropout distribution, under various circumstances we end up working with only *estimates* of this distribution. We will thus use $\mathcal{D}$ to refer to a generic dropout set distribution.

Selecting and evaluating an alternate set. An alternate set is chosen by an ***alternate selection algorithm***: a mapping that takes as input an instance $\mathcal{I}$ and a dropout set distribution $\mathcal{D}$, and outputs an alternate set $A \in \mathcal{A}(\mathcal{I})$. Designing this alternate selection algorithm is the main challenge, and the task we will focus on in this paper.

Intuitively, we evaluate our chosen alternate set according to its ability to "replace" a randomly drawn dropout set D. Formalizing this intuition requires defining precisely how A is used to "replace" D, which works as follows: With our alternate set A selected, we then observe the realization of $D \sim \mathcal{D}_{\boldsymbol{\rho}}$, at which point we are left with only the agents in $K \setminus D$. We must then restore the quotas to the greatest degree possible using a ***replacement set*** R: any subset of A is of size at most $|D|$ (that is, $R \subseteq A : |R| \leq |D|$). The best replacement set in A can simply be computed by a simple integer linear program (see Appendix B.1 for formulation), which solves following problem:

$$\min_{R \subseteq A : |R| \leq |D|} \quad dev(K \setminus D \cup R, \boldsymbol{l}, \boldsymbol{u}). \tag{2}$$

Here, we are minimizing a ***deviation function*** dev, which conceptually measures how well R restores the quotas $\boldsymbol{l}, \boldsymbol{u}$ after D drops out from K. Formally, a deviation function is a mapping that takes in a set of agents and quotas and outputs a real number describing how far that set of agents is from satisfying those quotas.

We will consider two specific deviation functions. First, the *binary deviation* ($dev^{0/1}$) simply checks whether the quotas are satisfied, outputting 0 if yes and 1 if no. The *linear deviation* (dev^{ℓ_1}) is more continuous, measuring how far away each quota is from being satisfied by normalized ℓ_1 distance. For a set of agents S and quotas $\boldsymbol{l}, \boldsymbol{u}$, we define these deviation functions as follows.

$$dev^{0/1}(S, \boldsymbol{l}, \boldsymbol{u}) := \begin{cases} 0 & \text{if } \sum_{i \in S} \mathbb{I}(f(i) = v) \in [l_{f,v}, u_{f,v}] \qquad \forall f \in F, v \in V_f \\ 1 & \text{else} \end{cases} \tag{binary dev.}$$

$$dev^{\ell_1}(S, \boldsymbol{l}, \boldsymbol{u}) := \sum_{f \in F} \sum_{v \in V_f} \frac{\max\left\{0,\ l_{f,v} - \sum_{i \in S} \mathbb{I}(f(i) = v),\ -u_{f,v} + \sum_{i \in S} \mathbb{I}(f(i) = v)\right\}}{u_{f,v}} \tag{linear dev.}$$

Note that the linear deviation is strictly more expressive than the binary deviation: the functions correspond exactly when $dev^{0/1} = dev^{\ell_1} = 0$, but otherwise the range of $dev^{0/1}$ is simply 1 while the range of dev^{ℓ_1} encompasses a spread of rational numbers. Throughout the paper, we will often drop the $\boldsymbol{l}, \boldsymbol{u}$ from the dev inputs when they are clear from context.

Finally, we evaluate an alternate selection algorithm that outputs A according to the *expected* deviation dev, over the randomness of drawing D from the true dropout set distribution $\mathcal{D}_{\boldsymbol{\rho}}$:

$$L^{dev}(A; \mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I}) = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}\left[\min_{R \subseteq A : |R| \leq |D|} dev(K \setminus D \cup R, \boldsymbol{l}, \boldsymbol{u})\right] \tag{3}$$

Abusing notation slightly, we will write the loss with respect to the binary deviation $dev^{0/1}$ as $L^{0/1}$ and linear deviation dev^{ℓ_1} as L^{ℓ_1}. We will sometimes compute the loss over a generic dropout set distribution $\mathcal{D}$, which is defined by Equation (3) where all instances of $\mathcal{D}_{\boldsymbol{\rho}}$ are replaced with $\mathcal{D}$. We will often drop the $\mathcal{I}$ out of our loss function when it is clear from context.

An ILP for computing the optimal alternate set. Finally, we express the optimal alternate set as the solution of an ILP. This ILP — expressed here informally and written fully in Appendix B.2 — is the program we solve when computing an optimal alternate set over a given dropout distribution $\mathcal{D}$. Note that by the program definition below, $\textsc{Opt}^{dev}(\mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I})$ is the true optimal alternate set in instance $\mathcal{I}$ with dropout probabilities $\boldsymbol{\rho}$. In $\textsc{Opt}^{dev}$, the objective function is exactly the expected

loss (Equation (3)), but for illustrative purposes, we have expanded the expectation into its sum form, where $support(\mathcal{D}) := \{D \in 2^K : \mathcal{D}(D) > 0\}$ be the collection of unique dropout sets with nonzero probability in $\mathcal{D}$ (noting that any D outside this support cannot contribute to the loss).

$\textsc{Opt}^{dev}(\mathcal{D}, \mathcal{I})$:

$$\begin{aligned}
&\underset{A, R_{A,D} | D \in support(\mathcal{D})}{\arg\min} \quad \sum_{D \in support(\mathcal{D})} Pr[D \sim \mathcal{D}] \cdot dev(K \setminus D \cup R_{A,D}, \boldsymbol{l}, \boldsymbol{u}) \\
&\text{s.t.} \quad |A| \le a \\
&\qquad A \subseteq N \\
&\qquad dev(K \setminus D \cup R_{A,D}, \boldsymbol{l}, \boldsymbol{u}) \le dev(K \setminus D \cup R, \boldsymbol{l}, \boldsymbol{u}) \qquad \forall\, R \subseteq A : |R| \le |D|,\ D \in support(\mathcal{D})
\end{aligned}$$

Observe that the number of terms in the objective and the number of constraints grows in the size of the $support(\mathcal{D})$. In our problem, this will be the dominant source of combinatorial blowup, as such, when $support(\mathcal{D})$ is small, the ILP is also "small" and can be solved quickly in practice with ILP solvers. However, when this support is large (it would be 2^k when $\mathcal{D} = \mathcal{D}_{\boldsymbol{\rho}}$), this ILP becomes intractable. This is precisely the advantage of our ERM-based approach, to be presented next: it allows us to solve Opt over an *empirical version* of $\mathcal{D}_{\boldsymbol{\rho}}$, whose support will be polynomial in size. In this way, we handle the most problematic source of combinatorial blowup – the 2^k possible dropout sets – while outsourcing the "practically easy" combinatorial aspects of the problem to Opt.

3 ERM-Alts, an ERM-based alternate selection algorithm

While we cannot explicitly write down or optimize over $\mathcal{D}_{\boldsymbol{\rho}}$, we can *sample* it: to draw a $D \sim \mathcal{D}_{\boldsymbol{\rho}}$, we must simply flip a ρ_i-weighted coin for each agent, and those whose coins turn up 1 are those in D. Using this observation, we now present our algorithm, ERM-Alts, which finds the alternate set that minimizes the *empirical risk*: the expected loss on an empirical approximation of $\mathcal{D}_{\boldsymbol{\rho}}$.

ALGORITHM 1: $\text{ERM-Alts}^{dev}(s, \boldsymbol{\rho}, \mathcal{I})$

$\hat{\mathcal{D}}_{\boldsymbol{\rho}} \leftarrow$ Draw s samples (dropout sets) from $\mathcal{D}_{\boldsymbol{\rho}}$ to produce *empirical* approximation of $\mathcal{D}_{\boldsymbol{\rho}}$, called $\hat{\mathcal{D}}_{\boldsymbol{\rho}}$.
$\hat{A}^* \leftarrow$ Solve $\textsc{Opt}^{dev}(\hat{\mathcal{D}}_{\boldsymbol{\rho}}, \mathcal{I})$.
return $\hat{A}^*$

Our theoretical analysis of ERM-Alts^{dev} will seek to bound its sample complexity: *how large must s be to ensure that $\hat{\mathcal{D}}_{\boldsymbol{\rho}} \approx \mathcal{D}_{\boldsymbol{\rho}}$ such that the loss of $\text{ERM-Alts}^{dev}(s, \boldsymbol{\rho}, \mathcal{I})$ closely approximates the loss of $\textsc{Opt}^{dev}(\mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I})$?* To ensure that the ILP $\textsc{Opt}^{dev}(\hat{\mathcal{D}}_{\boldsymbol{\rho}}, \mathcal{I})$ (solved in step 2 of the algorithm) is small enough to solve quickly, we ideally want s — and therefore $support(\hat{\mathcal{D}}_{\boldsymbol{\rho}})$ — to be polynomial in the parameters of the instance. We prove such sample complexity bounds by applying known results from PAC learning, so we now cast our problem within the PAC learning framework.

3.1 Casting our problem within the PAC learning framework

From PAC learning, we will primarily use the concept of *agnostic PAC learnability*, which is defined as follows (see Definition 3.3, [Shalev-Shwartz and Ben-David, 2014]).

Definition 3.1 (Agnostic PAC Learnability). A hypothesis class of functions $\mathcal{H}$ with domain $\mathcal{X}$ and range $\mathcal{Y}$ is agnostic PAC learnable if there exists a function $s_{\mathcal{H}} : (0,1)^2 \to \mathbb{N}$ and a learning algorithm with the following property: for every $\delta, \varepsilon \in (0,1)$ and every distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$, when running the learning algorithm on $s \ge s_{\mathcal{H}}(\varepsilon, \delta)$ i.i.d. examples generated by a fixed distribution

$\mathcal{D}$, the algorithm returns a hypothesis $\hat{h} \in \mathcal{H}$ such that

$$\Pr[L(\hat{h}) - \min_{h \in \mathcal{H}} L(h) \leq \varepsilon] \geq 1 - \delta,$$

where the probability is computed over the random draw of the choice of training set given to the learning algorithm (assumed to be drawn from $\mathcal{D}$) and the loss L measures the extent to which h fails to output the *true label* $y \in \mathcal{Y}$ on input $x \in \mathcal{X}$, in expectation over $x, y \sim \mathcal{D}$.

Mapping this onto our problem, we want to find an alternate set $\hat{A}$ (hypothesis $\hat{h}$) within $\mathcal{A}$ (hypothesis class $\mathcal{H}$) such that with high probability (probability at least $1 - \delta$), $\hat{A}$ has near-optimal expected loss (has loss within ε of that of the loss-minimizing hypothesis) over dropout sets (samples) drawn from $\mathcal{D}_{\boldsymbol{\rho}}$ (a fixed distribution $\mathcal{D}$). Our learning algorithm is empirical risk minimization (Algorithm 1), corresponding to a classical algorithm for PAC learning.

Fix an instance $\mathcal{I}$. As illustrated above, our hypothesis class in $\mathcal{I}$ is just the collection of all alternate sets $\mathcal{A}(\mathcal{I})$. To distinguish actual alternate sets (sets of agents) and hypotheses (functions), we let h_A be the hypothesis corresponding to A. For all $a \in \mathcal{A}(\mathcal{I})$, each h_A takes as input a given dropout set (and technically the instance, but we leave this implicit) and outputs a real number describing how well A can replace that dropout set, as measured by dev. Because $dev^{0/1}$ and dev^{ℓ_1} have different ranges, we must define two hypothesis classes per instance: a *binary* hypothesis class $\mathcal{H}^{0/1}(\mathcal{I}) = \{h_A^{0/1} | A \in \mathcal{A}(\mathcal{I})\}$, and a *linear* hypothesis class $\mathcal{H}^{\ell_1}(\mathcal{I}) = \{h_A^{\ell_1} | A \in \mathcal{A}(\mathcal{I})\}$. The output of h_A^{dev} is exactly the minimum deviation dev over all replacement sets, as defined in Equation (2):

$$h_A^{0/1}(D) = \min_{R \subseteq A : |R| \leq |D|} dev^{0/1}(K \setminus D \cup R) \qquad \text{and} \qquad h_A^{\ell_1}(D) = \min_{R \subseteq A : |R| \leq |D|} dev^{\ell_1}(K \setminus D \cup R). \quad (4)$$

Across these two hypothesis classes, the *labeling* function is the same: the "true" label for any D is 0, reflecting that A can ideally maintain quota satisfaction after D drops out. Formally, we define a labeling function as $\tau(D; \mathcal{I}) = 0$, which we shorten to $\tau(D)$. Of course, there may exist no $A \in \mathcal{A}(\mathcal{I})$ such that $h_A(D) = 0$ for all D; this consideration is encompassed by the definition of agnostic PAC learning, where we aim to compete with the best available hypothesis.

Finally, we show that by minimizing the $L^{0/1}$ and L^{ℓ_1} loss as in Section 2, we are respectively minimizing the standard 0-1 loss and ℓ_1 loss from PAC learning. Fixing $A \in \mathcal{A}$, we apply the definition of the loss (Equation (3)) and our definition of hypotheses (Equation (4)):

$$L^{0/1}(A; \mathcal{D}_{\boldsymbol{\rho}}) = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}} \left[\mathbb{I}(h_A^{0/1}(D) \neq 0 \right] = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}} \left[\mathbb{I}(h_A^{0/1}(D) \neq \tau(D) \right], \text{ the 0/1 PAC learning loss;}$$

$$L^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}}) = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}} \left[\left| h_A^{\ell_1}(D) - 0 \right| \right] = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}} \left[\left| h_A^{\ell_1}(D) - \tau(D) \right| \right], \text{ the } \ell_1 \text{ PAC learning loss.}$$

This completes our reduction to agnostic PAC learning, which we will use to apply known sample complexity bounds (i.e., bounds on $s_{\mathcal{H}(\epsilon,\delta)}$ in Definition 3.1) from the PAC learning literature. These known bounds, stated in Appendix C.3, depend on the "dimension", or richness, of $\mathcal{H}$. To apply them, we must therefore characterize the respective dimensions of $\mathcal{H}^{0/1}(\mathcal{I})$ and $\mathcal{H}^{\ell_1}(\mathcal{I})$. Because $\mathcal{H}^{0/1}(\mathcal{I})$ consists of hypotheses with range $\{0, 1\}$, its relevant dimensionality measure is the *VC-dimension* (VC); hypotheses in $\mathcal{H}^{\ell_1}(\mathcal{I})$ have real-valued range, so this class's dimension is measured by a generalization of the VC-dimension called the *Pseudodimension* (Pdim). We define these notions of dimension below in the same notation as in Definition 3.1; then, in the next section, we tightly bound our hypothesis classes' respective dimensions.

Definition 3.2 (VC-Dimension). The *VC-Dimension* of $\mathcal{H}$ (whose hypotheses have range $\{0, 1\}$) is the size of the largest set of points in $\mathcal{X}$ that can be *shattered* by $\mathcal{H}$. A set of points X can be shattered by $\mathcal{H}$ iff for all labeling functions $b : X \to \{0, 1\}^{|X|}$, there exists an $h_b \in \mathcal{H}$ such that $h_b(x) = b(x)$ for all $x \in X$.

Definition 3.3 (Pseudodimension). The *Pseudodimension* of $\mathcal{H}$ (whose hypotheses have range $[0,1]$) is the size of the largest set of points in $\mathcal{X}$ that can be *pseudo-shattered* by $\mathcal{H}$. A set of points X can be pseudo-shattered by $\mathcal{H}$ iff there exists a *witness vector* $r \in [0,1]^{|X|}$ (indexed as r_x) such that for all labeling functions $b : X \to \{0,1\}^{|X|}$, there exists an $h_b \in \mathcal{H}$ such that $\mathbb{I}(h_b(x) - r_x > 0) = b(x)$ for all $x \in X$.

3.2 Bounds on VC dimension of $\mathcal{H}^{0/1}$ and Pseudimension of $\mathcal{H}^{\ell_1}$

Our hypothesis classes are instance-dependent (since the set of available alternate sets depends on who is in the instance's pool), but we want to give dimension bounds that apply across instances. Thus, for each type of hypothesis class we will examine its *worst-case* dimensionality over all instances within a given collection. We define a collection of instances as follows: let $\mathfrak{I}(n,a)$ denote all instances $\mathcal{I}$ with $|N| = n$ and an alternate budget of a. Similarly let $\mathfrak{I}(n,a,|F|)$ denote all instances $\mathcal{I}$ with $|N| = n$, an alternate budget of a, and $|F|$ features.

Theorem 3.4 (Main VC, Pdim Bounds) *Fix any $n, a \in \mathbb{N}_{\geq 1}$ such that $n \geq a$. Then,*

$$\max_{\mathcal{I} \in \mathfrak{I}(n,a)} \mathrm{VC}(\mathcal{H}^{0/1}(\mathcal{I})), \max_{\mathcal{I} \in \mathfrak{I}(n,a)} \mathrm{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \in \Theta(a \log n).$$

Proof: Upper bounds. Fix an arbitrary instance $\mathcal{I}$. Both upper bounds are proven by the same argument, which relies on the simple fact that both our hypothesis classes $\mathcal{H}^{0/1}(\mathcal{I})$ and $\mathcal{H}^{\ell_1}(\mathcal{I})$ are finite. Their sizes are upper-bounded by $|\mathcal{A}|$, which is of size at most $\binom{n}{a}$, as there are at most this many a-subsets of N. We now apply the well-known fact that $\mathrm{VC}(\mathcal{H}) \leq \log(|\mathcal{H}|)$, and likewise, $\mathrm{Pdim}(\mathcal{H}) \leq \log(|\mathcal{H}|)$. It follows that

$$\mathrm{VC}(\mathcal{H}^{0/1}(\mathcal{I})), \mathrm{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \leq \log \binom{n}{a} \leq \log \left(\left(\frac{en}{a} \right)^a \right) = a(\log n - \log a + 1) \in O(a \log n). \quad \square$$

The lower bounds both arise from the same instance, but are more involved. We thus give a sketch here, and defer the details to Appendix C.1. The most efficient way to reason about the problem is in terms of agents' feature *vectors*, describing their values across the features; formally, agent i's feature vector is defined as $w(i) := (f(i) | i \in F)$.

Proof Sketch: Lower Bounds. We first construct the instance, then construct a shattered set to get the VC lower bound, and then extend the instance to get the Pdim lower bound.

Instance construction. Fix $n, a \in \mathbb{N}_{\geq 1}$ with $n \geq a\, 2^{|F|/a}$. Let our instance be such that $k = |F| = \lfloor \log \binom{n}{a} \rfloor - 2a$, and let the features be numbered $1 \ldots |F|$. Each feature $f \in F$ is binary-valued, so $V_f = \{0,1\}$ for all $f \in F$. The quotas are $l_{f,0} = 0$, $l_{f,1} = 1$ and $u_{f,0} = u_{f,0} = \infty$ for all $f \in F$, so the only requirement is that the panel contains at least one person with a 1 for every feature. We have established that the panel contains $k = |F|$ panelists; let K be constructed such that the i-th panelist's feature vector be the i-th standard $|F|$-length basis vector. That is, $w(i) = 0 \ldots 010 \ldots 0$, where the 1 occurs at the i-th feature. Note that K satisfies the quotas.

We specify the pool N only partially, leaving most of its contents to be arbitrary. To construct the part of the pool we care about, we partition the $|F|$ features into a sets of equal size $m = |F|/a$ (assume m is integer — we handle indivisibility in the full proof). As such, these sets of features are

$$g_1 = \{1, 2, \ldots, m\}, \quad g_2 = \{m+1, m+2, \ldots, 2m\}, \quad \ldots, \quad g_a = \{(a-1)m+1, (a-1)m+2, \ldots, am\}$$

We then define a corresponding subgroups of the pool $G_1 \ldots G_a$. Each G_j contains 2^m agents, composed of agents who have 0s for all features other than those in g_j, and for the m features in g_j have values described by a unique vector $u \in \{0,1\}^m$. That is, each $i \in G_j$ is defined by some unique $u \in \{0,1\}^m$, and their feature vector is $0 \ldots 0u0 \ldots 0$, where u occurs over the indices in g_j.

Lower bound on the VC-dimension. We construct a collection of $k = \left\lfloor \log \binom{n}{a} \right\rfloor - 2a$ dropout sets to be shattered as follows, where each dropout set consists of a single panelist:

$$D_1 = \{1\},\ D_2 = \{2\}, \ldots,\ D_k = \{k\}.$$

Fix a labeling $b \in \{0,1\}^k$, where assuming 0 means we can satisfy the quotas after dropouts, and 1 means we cannot. We will construct an alternate set A_b to realize labeling b by taking one agent from each set G_j, thereby taking a agents overall. In particular, from group G_j we take the agent whose values are the opposite of b on all indices in g_j (i.e., they match the vector $c \in \{0,1\}^k$ such that $c_i = 1 \iff b_i = 0$). To see why this achieves the labeling, recall that achieving a 1 label for dropout set D_i just means making sure at least one agent in A_b has a 1 for the i-th feature, as this is the 1-value we lose when $D_i = \{i\}$ drops out. On any 0-labeled dropout set $i \in [k]$, A_b must contain an agent with a 1 in the i-th position: to see this, note that i must lie in exactly one g_j, and the agent we took from set G_j must have a 1 for the i-th feature because we choose this value to match $c_i = 1$. Negating this argument proves correctness for 0-labeled dropout sets, with the additional argument that all agents in G_j have value 0 for all features outside g_j. We have shattered a collection of dropout sets of the following size, completing the VC dimension lower bound:

$$\left\lfloor \log \binom{n}{a} \right\rfloor - 2a \geq \log(n/a)^a - 1 - 2a = a(\log n - \log a - 2) - 1 \in \Omega(a \log n).$$

Extension to the pseudodimension. To lower bound the pseudodimension, we actually show something stronger: for all $\mathcal{I}$,

$$\mathrm{VC}(\mathcal{H}^{0/1}(\mathcal{I})) \geq t \implies \mathrm{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \geq t \qquad \text{for all } t \in \mathbb{R}^+.$$

The fundamental insight used to show this claim is that there exists a threshold $1/(k+1)$ (corresponding to all entries of our witness vector), which transforms our linear hypothesis class into the corresponding binary hypothesis class. To see this, note that the range of any linear hypothesis $h_A^{\ell_1}$ can take on rational values at minimal intervals of $1/k$, so the smallest nonzero output of any linear hypothesis is $1/k$. Thus, if for some D a linear hypothesis $h_A^{\ell_1}(D) > 1/(k+1)$, there must be a quota violation and the corresponding binary hypothesis $h_A^{0/1}(D) = 1$; else, it must be that $h_A^{\ell_1}(D) = 0$, there is no quota violation, and the corresponding binary hypothesis $h_A^{0/1}(D)$ must output 0. □

The construction giving this lower bound uses many features: $|F| = k = \lfloor \log \binom{n}{a} \rfloor - 2a$. This prompts the question: can we prove sample complexity bounds that improve as $|F|$ gets smaller? We present such bounds below. We give a proof sketch here, and defer the details to Appendix C.2. We remark that the condition $n \geq a2^{|F|/a}$ is required only for the lower bound.

Theorem 3.5 ($|F|$-dependent VC, Pdim Bounds) *Fix any $n, a, |F| \in \mathbb{N}_{\geq 1}$ such that $n \geq a2^{|F|/a}$. Taking $|F|$ to be a varying parameter and $\max_{f \in F} |V_f|$ to be a constant, we have that*

$$\max_{\mathcal{I} \in \mathfrak{I}(n,a,|F|)} \mathrm{VC}(\mathcal{H}^{0/1}(\mathcal{I})),\ \max_{\mathcal{I} \in \mathfrak{I}(n,a,|F|)} \mathrm{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \in \left[|F|, O\left(a|F| \log \max_{f \in F} |V_f| \right) \right] \in \left[|F|, O(a\,|F|) \right].$$

Proof sketch. The upper bound is proven by bounding the size of the hypothesis class in terms of $|F|$ rather than n and a. The insight here is that the maximum number of *unique* possible alternate sets in an instance is upper-bounded by the number of ways to choose combinations of unique feature vectors (with repeats). The number of possible unique feature vectors is bounded by $\prod_{f \in F} |V_f| \leq \left(\max_{f \in F} |V_f|\right)^{|F|}$. From there, the bound follows from straightforward identities from combinatorics to bound the number of ways to choose a feature vectors, with repeats, from a universe of feature vectors of this size.

The lower bound is proven by generalizing the proof of the lower bound in Theorem 3.4 such that $|F|$ is no longer set to k, but rather allowed vary so long as $|F| \leq k$. □

3.3 Sample complexity bounds for ERM-Alts

We now plug our upper bounds from Theorem 3.4 into known dimension-dependent sample complexity bounds (Lemmas C.1 and C.2 respectively, stated in Appendix C.3), to give sample size-dependent formal guarantees on the loss of ERM-Alts. The key takeaway of these bounds is that we need few samples (only logarithmic in n), which is important because our ILP Opt scales commensurately. One could achieve analogous $|F|$-dependent results by plugging in Theorem 3.5.

Corollary 3.6 (**Bound for** ERM-Alts$^{0/1}$) *Fix $\mathcal{I}, \boldsymbol{\rho}$ and constants $\varepsilon, \delta > 0$. There exists*

$$s \in O\left((a \log n + \log(1/\delta)) \cdot 1/\varepsilon^2\right) \quad \textit{such that}$$

$$\Pr\left[L^{0/1}(\textit{ERM-Alts}^{0/1}(s, \boldsymbol{\rho}, \mathcal{I}); \mathcal{D}_{\boldsymbol{\rho}}) - L^{0/1}(\textit{Opt}^{0/1}(\mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I}); \mathcal{D}_{\boldsymbol{\rho}}) \le \varepsilon\right] \ge 1 - \delta,$$

where the probability above is over the random sampling of $\hat{\mathcal{D}}_{\boldsymbol{\rho}}$ in ERM-Alts.

Corollary 3.7 (**Bound for** ERM-Alts$^{\ell_1}$) *Fix $\mathcal{I}, \boldsymbol{\rho}$ and constants $\varepsilon, \delta > 0$. There exists*

$$s \in O\left((a \log n + \log(1/\varepsilon) + \log(1/\delta)) \cdot 1/\varepsilon^2\right) \quad \textit{such that}$$

$$\Pr\left[L^{\ell_1}(\textit{ERM-Alts}^{\ell_1}(s, \boldsymbol{\rho}, \mathcal{I}); \mathcal{D}_{\boldsymbol{\rho}}) - L^{\ell_1}(\textit{Opt}^{\ell_1}(\mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I}); \mathcal{D}_{\boldsymbol{\rho}}) \le \varepsilon\right] \ge 1 - \delta.$$

4 Robustness to mis-estimated dropout probabilities

So far, we have assumed that panelists' dropout probabilities $\boldsymbol{\rho}$ are known. However, in practice they must be estimated from historical data based on known features. This prompts the question: How robust is our ERM algorithm when dropout probability estimates are subject to prediction errors?

To formalize this, let $\tilde{\rho}_i$ be our estimate of the agent i's dropout probability, and correspondingly, let $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_i | i \in K)$. Then, let our *prediction error* γ be defined as the largest difference between any dropout probability estimate and its true value, i.e., $\gamma := \|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_\infty$. Ideally, the loss of an algorithm given $\tilde{\boldsymbol{\rho}}$ should approach its loss given $\boldsymbol{\rho}$ as $\gamma \to 0$. In fact, we find that whether this is true depends on our choice of deviation function: ERM-Alts$^{0/1}$ suffers arbitrary loss under vanishingly small prediction errors (Theorem 4.1), while ERM-Alts$^{\ell_1}$ is highly robust (Theorems 4.2 and 4.3).

For all proofs in this section, we weaken one assumption: we permit the replacement set R to be larger than D.[2] Formally, we use the following modified version of the loss (Equation (3)), where the optimization occurs over all $R \subseteq A$ instead of just $R \subseteq A$ such that $|R| \le |D|$:

$$\mathcal{L}^{dev}(A; \mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I}) := \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}\left[\min_{R \subseteq A} dev(K \setminus D \cup R, \boldsymbol{l}, \boldsymbol{u})\right].$$

Even with this relaxation, we will still be able to prove separation between the two deviation functions' robustness. Note that this relaxation does not affect our results from Section 3.

In our proofs we will use the following shorthand, so ERM-A^{dev} and $\widetilde{\text{ERM-}A^{dev}}$ are the alternate sets given by our algorithm given the *true* and *estimated* dropout probabilities, respectively, and OPT-A^{dev} and $\widetilde{\text{OPT-}A^{dev}}$ are the analogous optimal alternate sets.

$$\text{ERM-}A^{dev} := \text{ERM-Alts}^{dev}(s, \boldsymbol{\rho}, \mathcal{I}) \qquad \widetilde{\text{ERM-}A^{dev}} := \text{ERM-Alts}^{dev}(s, \tilde{\boldsymbol{\rho}}, \mathcal{I})$$

$$\text{OPT-}A^{dev} := \text{OPT-}A^{dev}(s, \mathcal{D}_{\boldsymbol{\rho}}, \mathcal{I}) \qquad \widetilde{\text{OPT-}A^{dev}} := \text{OPT}^{dev}(s, \mathcal{D}_{\tilde{\boldsymbol{\rho}}}, \mathcal{I}),$$

Henceforth dropping $\mathcal{I}, \boldsymbol{l}, \boldsymbol{u}$ from the notation, our formal goal is to bound the added loss $\mathcal{L}$ of our algorithms due to mis-estimation,

$$\mathcal{L}^{dev}(\widetilde{\text{ERM-}A^{dev}}, \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{dev}(\text{ERM-}A^{dev}, \mathcal{D}_{\boldsymbol{\rho}}).$$

[2] Theorem 4.1 holds without this relaxation, but we have not been able to determine whether Theorem 4.2 does.

We will generally bound this quantity by bounding the analogous difference for $\widetilde{\text{OPT-}A^{dev}}$ and OPT-A^{dev} (which will be *deterministic*), and then use that, for sufficient s, ERM-A^{dev} = OPT-A^{dev} and $\widetilde{\text{ERM-}A^{dev}}$ = $\widetilde{\text{OPT-}A^{dev}}$ (which will be *with high probability*). This last translation step will require our lower bound instances to be constructed with some care.

We first show that ERM-Alts$^{0/1}$ is *arbitrarily* non-robust. Recalling that the maximum possible binary loss is 1, this means is that for arbitrarily small γ, there exists an instance where estimation error γ increases the binary loss by a quantity arbitrarily close to 1 with high probability.

Theorem 4.1 (**Binary loss lower bound**) *Fix any constants $\alpha \in (0,1), \delta \in (0,1/2], \gamma \in (0,1)$. There exists $\mathcal{I}$, $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ with $\|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_\infty \le \gamma$, and $s(\alpha,\delta,\gamma)$ such that for all $s \ge s(\alpha,\delta,\gamma)$,*

$$\Pr\left[\mathcal{L}^{0/1}(\widetilde{\textit{ERM-}A^{0/1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{0/1}(\textit{ERM-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}}) \ge 1-\alpha\right] \ge 1 - 2\delta.$$

Proof. We construct a simple instance $\mathcal{I} = (N, K, \boldsymbol{l}, \boldsymbol{u}, a)$ as follows: there is only one binary feature, so $F = f_1$, and $V_{f_1} = \{0,1\}$. The panel K is of size $k = 2\lceil \log_{(1-\gamma)} \alpha \rceil$ (note: must be ≥ 2, even, and positive) and it is comprised of $k/2$ agents with feature-value 0 and $k/2$ agents with feature-value 1. The quotas are tight, so $l_{f_1,1} = u_{f_1,1} = l_{f_1,0} = u_{f_1,0} = k/2$. The pool N contains at least k agents with feature-value 0 and at least k agents with feature-value 1, so our alternate set construction is unencumbered by any limitations of the pool. Finally, $a = k/2$. We set $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ as follows, noting that indeed $\|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_\infty = \gamma$.

$$\rho_i = \begin{cases} \gamma & \text{if } f_1(i) = 1 \\ 0 & \text{if } f_1(i) = 0 \end{cases} \qquad \tilde{\rho}_i = \begin{cases} 0 & \text{if } f_1(i) = 1 \\ \gamma & \text{if } f_1(i) = 0 \end{cases} \qquad \text{for all } i \in K.$$

First, observe that OPT-$A^{0/1}$ just consists of $a = k/2$ agents i with $f_1(i) = 1$, and $\widetilde{\text{OPT-}A^{0/1}}$ consists of $a = k/2$ agents i with $f_1(i) = 0$. This is because under both $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, exactly $a = k/2$ members of the panel have a non-zero chance of dropping out, and both of these optimal sets simply have backups for all of these agents, and no other agents. Note that $\mathcal{L}^{0/1}(\text{OPT-}A^{0/1}; \mathcal{I}) = 0$. Further,

$$\mathcal{L}^{0/1}(\widetilde{\text{OPT-}A^{0/1}}; \mathcal{D}_{\boldsymbol{\rho}}) = 1 \cdot \Pr_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}[\exists\, i \in D\colon f_1(i) = 1] = 1 - (1-\gamma)^{k/2} \ge 1 - \alpha,$$

because $\widetilde{\text{OPT-}A^{0/1}}$ contains only agents with $f(i) = 0$ and therefore must incur $dev^{0/1}$ if *any* panelist with $f(i) = 1$ drops out. It follows that

$$\mathcal{L}^{0/1}(\widetilde{\text{OPT-}A^{0/1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{0/1}(\text{OPT-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}}) \ge 1 - \alpha.$$

Finally, it remains to show that for sufficient s,

$$\Pr[\text{ERM-}A^{0/1} = \text{OPT-}A^{0/1} \wedge \widetilde{\text{ERM-}A^{0/1}} = \widetilde{\text{OPT-}A^{0/1}}] \ge 1 - 2\delta,$$

which implies the claim. We defer the full argument to Appendix D.1, where the argument proceeds in two steps: first, we show a constant separation between the loss of OPT-$A^{0/1}$ and that of any other alternate set A, and a symmetric separation for $\widetilde{\text{OPT-}A^{0/1}}$. Then, we apply Corollary 3.6 to derive $s(\alpha,\delta,\gamma)$ such that for all $s \ge s(\alpha,\delta,\gamma)$, the desired event occurs with at least $1-2\delta$ probability. □

Fortunately, we next find that ERM-Alts^1 *is* highly robust, incurring additional loss proportional to $\gamma|FV|$ (Theorem 4.2). We then show that this bound is essentially tight (Theorem 4.3).

Theorem 4.2 (**Linear loss upper bound**) *Fix any constants $\varepsilon, \delta > 0, \gamma \in (0,1]$. Fix any $\mathcal{I}$, $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ with $\|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_\infty \le \gamma$. Then, there exists $s(\varepsilon,\delta)$ such that for all $s \ge s(\varepsilon,\delta)$,*

$$\Pr\left[\mathcal{L}^{\ell_1}(\widetilde{\textit{ERM-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\textit{ERM-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \le 2\gamma|FV| + \varepsilon\right] \ge 1 - \delta$$

Proof sketch. Fix all required entities; as usual, $\mathcal{I}$ will be dropped from our notation. The core of the proof is showing the following bound on the change in linear loss for *any* fixed alternate set A, when evaluated with respect to $\mathcal{D}_{\boldsymbol{\rho}}$ versus $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$:

$$|\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\tilde{\boldsymbol{\rho}}})| \leq \gamma|FV|. \tag{5}$$

Once we have this bound, we can apply it to OPT-A^{ℓ_1} to show the following chain of inequalities, where the first inequality is by the optimality of $\widetilde{\text{OPT-}A^{\ell_1}}$ for $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$:

$$\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}};\mathcal{D}_{\tilde{\boldsymbol{\rho}}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1};\mathcal{D}_{\boldsymbol{\rho}}) \leq \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1};\mathcal{D}_{\tilde{\boldsymbol{\rho}}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1};\mathcal{D}_{\boldsymbol{\rho}}) \leq \gamma|FV|.$$

This gives us an almost analogous version of our desired bound for OPT-A^{ℓ_1} and $\widetilde{\text{OPT-}A^{\ell_1}}$. To relate $\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}};\mathcal{D}_{\boldsymbol{\rho}})$ to $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}};\mathcal{D}_{\tilde{\boldsymbol{\rho}}})$, we apply Equation (5) once more and derive $s(\varepsilon,\delta)$ based on Corollary 3.7 to ensure that with probability $\geq 1-\delta$, the loss of $\widetilde{\text{ERM-}A^{\ell_1}}$ on $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$ with $s \geq s(\varepsilon,\delta)$ is within ε of its respective corresponding optimal alternate set (for the upper bound, we do not need to ensure that ERM-A^{ℓ_1} is close to OPT-A^{ℓ_1}). This additional application of Equation (5) is the source of the 2-factor on $\gamma|FV|$ in the statement, and completes the proof.

It remains to show Equation (5). The argument is quite intricate, but we will present the broad strokes of the argument here and give the fully formal version in Appendix D.2. The high level approach is to construct $k+1$ intermediate probability vectors that transform $\boldsymbol{\rho}$ to $\tilde{\boldsymbol{\rho}}$ by altering one agent's dropout probability at a time. As such, let $\boldsymbol{\rho}^i = (\tilde{\rho}_1,\ldots,\tilde{\rho}_i,\rho_{i+1},\ldots,\rho_k)$ for all $i \in [k]$. Then, using a telescoping sum and the triangle inequality, we show that

$$|\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\tilde{\boldsymbol{\rho}}})| = |\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^0}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^k})| \leq \sum_{i=1}^{k} \left|\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^{i-1}}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^i})\right|.$$

We first bound each individual term of this resulting sum separately. We do so via a coupling argument, where we couple the dropouts drawn per $\boldsymbol{\rho}^i$ and $\boldsymbol{\rho}^{i+1}$. We represent these dropouts as

$$\boldsymbol{Y} = (Y_j \sim \text{ Bernoulli}(\rho_j^{i-1}) | j \in [k]) \quad \text{ and } \quad \boldsymbol{Y'} = (Y'_j \sim \text{ Bernoulli}(\rho_j^{i}) | j \in [k]),$$

where Y_j (likewise Y'_j) indicates whether j dropped out. We couple $\boldsymbol{Y}$ and $\boldsymbol{Y'}$ by drawing each Y_j and Y'_j via the same independent draw of $X_j \sim Unif[0,1]$. Letting $\boldsymbol{X} = (X_j | j \in [k])$, we denote the coupled dropout vectors as $\boldsymbol{Y}(\boldsymbol{X}) = (Y_j(\boldsymbol{X}) | j \in [k])$ and $\boldsymbol{Y'}(\boldsymbol{X})$ analogously. Let $D(\boldsymbol{Y}(\boldsymbol{X})) = \{j \in [k] | Y_j = 1\}$ be the dropout set specified by $\boldsymbol{Y}$ and define $D(\boldsymbol{Y'}(\boldsymbol{X}))$ analogously. Define $R(\boldsymbol{Y}(\boldsymbol{X})) := \arg\min_{R \subseteq A} dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R)$ to be the optimal replacement set for $D(\boldsymbol{Y}(\boldsymbol{X}))$, and define $R(\boldsymbol{Y'}(\boldsymbol{X}))$ analogously.

This coupling will serve to ensure that with substantial probability, $D(\boldsymbol{Y}(\boldsymbol{X}))$ and $D(\boldsymbol{Y'}(\boldsymbol{X}))$ are the same, and otherwise they are similar. In particular, there is some $cond(X_i)$ such that

$$\begin{aligned} Pr[cond(X_i)] &= 1-\gamma \quad \text{and} \quad cond(X_i) \implies D(\boldsymbol{Y}(\boldsymbol{X})) = D(\boldsymbol{Y'}(\boldsymbol{X})) \\ Pr[\neg cond(X_i)] &= \gamma \quad \text{and} \quad \neg cond(X_i) \implies D(\boldsymbol{Y}(\boldsymbol{X})) \triangle D(\boldsymbol{Y'}(\boldsymbol{X})) = \{i\}. \end{aligned}$$

Fix an $i \in [k]$. Applying the definition of the loss along with the law of total expectation, linearity of expectation, and Jensen's inequality, we deduce that

$$\begin{aligned} &|\mathcal{L}^{\ell_1}(A;\, \mathcal{D}_{\boldsymbol{\rho}^{i-1}}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^i})| \\ &\qquad \leq \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R(\boldsymbol{Y}(\boldsymbol{X}))) - dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R(\boldsymbol{Y'}(\boldsymbol{X})))|\right]. \end{aligned}$$

From here, there are three key steps. First, we carefully choose a replacement set R_X that is common across terms such that

$$\leq \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}) - dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R_{\boldsymbol{X}})|\right]$$

Next, we apply the first consequence of our coupling argument: that $cond(X_i)$ (which occurs with probability $1-\gamma$) implies $D(\boldsymbol{Y}(\boldsymbol{X})) = D(\boldsymbol{Y'}(\boldsymbol{X}))$, conditional on which the expectation is 0. Then,

$$\leq \gamma \, \mathbb{E}_{\boldsymbol{X}} \left[\left| dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}) - dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}) \right| \neg cond(X_i) \right]$$

Finally, we apply the second consequence of our coupling argument: that $\neg cond(X_i)$ implies that $D(\boldsymbol{Y}(\boldsymbol{X})) \triangle D(\boldsymbol{Y'}(\boldsymbol{X})) = \{i\}$, which means the symmetric difference of the two sets on which dev^{ℓ_1} is evaluated above is also $\{i\}$. Over several steps of reasoning, we use this observation to bound on difference of deviations above in terms of only i's feature-values. The deduction concludes with

$$\leq \gamma \sum_{f \in F} 1/u_{f,f(i)}.$$

Putting it all together, $|\mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^0}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^k})|$ is upper bounded by the above expression summed up over all i, concluding the proof:

$$\sum_{i=1}^{k} \gamma \sum_{f \in F} \frac{1}{u_{f,f(i)}} = \gamma \sum_{f,v \in FV} \frac{\sum_{i=1}^{k} \mathbb{I}(f(i)=v)}{u_{f,v}} \leq \gamma \sum_{f,v \in FV} \frac{u_{f,v}}{u_{f,v}} = \gamma |FV|. \qquad \square$$

Finally, we show that the above bound is essentially tight. The proof proceeds similarly to that of Theorem 4.1 but the construction is more complex, requiring generic numbers of feature-values. We defer the proof to Appendix D.3.

Theorem 4.3 (Linear loss lower bound) *Fix any constants $\gamma \in (0,1), \delta \in (0, 1/2]$ and $|FV|, |F|$. There exists a $\mathcal{I}$ with $|F|$ features and $|FV|$ feature-values, $\boldsymbol{\rho}$, and $\tilde{\boldsymbol{\rho}}$ with $\|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\| \leq \gamma$, and $s(\delta, \gamma)$ such that for all $s \geq s(\delta, \gamma)$,*

$$\Pr\left[\mathcal{L}^{\ell_1}(\widetilde{ERM\text{-}A}^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(ERM\text{-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \geq \gamma |FV| - \gamma |F| \right] \geq 1 - 2\delta.$$

5 Empirical Evaluation

Estimating dropout probabilities from historical data. We estimate the dropout probabilities by fitting the same parametric model as in prior work on sortition, where it was used to estimate agents' probabilities of opting into the pool [Flanigan et al., 2020]. This model, sometimes referred to as *simple independent action*, assumes that each feature-value independently contributes to an agent's probability of dropping out. In particular, the parameters of the model are $\{\beta_0\} \cup \{\beta_{f,v} | f \in F, v \in V_f\}$, and we assume these parameters relate to the dropout probabilities as follows:

$$\rho_i = \beta_0 \prod_{f \in F} \beta_{f,f(i)}. \tag{6}$$

Colloquially, $\beta_{f,v}$ is the probability of dropping out due to having value v for feature f, and β_0 is the baseline probability of dropping out irrespective of identity. We fit these β parameters using maximum likelihood estimation.

Datasets. We fit the βs in Equation (6) based on datasets whose rows are agents i, and columns are all i's feature-values plus the binary outcome variable describing whether i dropped out. Each *dataset* corresponds to a single instance (real-world assembly); a data *cluster* is a group of datasets with common feature-values (this commonality is necessary for coherently estimating the βs). Datasets within each cluster correspond to assemblies organized by the same practitioner group.

We train βs from seven total datasets divided into two separate data clusters: the first cluster contains 3 datasets from US assemblies, named *US-1*, *US-2*, *US-3*, and the second contains 4 datasets from Canadian assemblies, named *Can-1* - *Can-4*. We train on as many features present in Figure 1 as possible. Data structure and cleaning details are in Appendix E.1.

We focus on the US data cluster in the body because these datasets contain extra information relevant for evaluation: the precise quotas used in panel selection, and the alternates used in practice. Replication of results in the Canadian datasets are described in Appendix E.7; we see basically the same trends, but they are less pronounced because the quotas used in these Canadian datasets tend to be far less restrictive (this is a known trend, see, e.g., [Flanigan et al., 2020]).

Algorithms. We run three ERM-based algorithms, ERM-ALTS^1, ERM-ALTS$^{0/1}$, and ERM-ALTS^1-EQ (see below), all with $s = 300$ samples (see Appendix E.2 for demonstration of convergence). We compare these algorithms to the benchmarks below, formally specified in Appendix E.3.

QUOTA-BASED is a proxy for what at least one practitioner group reportedly does in practice: chooses alternates by selecting an entire duplicate panel. Interpreted literally, this method requires $a = k$; to capture this reasoning for generic a we extend their method by scaling down the quotas proportionally to the size of the alternate set. This algorithm is the natural benchmark for what one could do using only existing tools and no knowledge of the dropout probabilities.

GREEDY captures the marginal benefit of knowing the dropout probability estimates but not reasoning about the problem's combinatorial structure. GREEDY orders the panelists in decreasing order of dropout probability. Then, taking panelists $i \in \{1 \dots a\}$ in order, it finds the remaining pool member whose attributes match i on the most features and adds them to the alternate set.

ERM-ALTS^1-EQ roughly captures the opposite of greedy: it captures the marginal benefit of reasoning about the combinatorial structure of the problem *without* access to dropout probability estimates, so dropout probabilities are assumed to be equal. Formally, this is ERM-ALTS^1 run with dropout probabilities all equal to the average of the dropout probabilities in that instance, so that the expected number of dropouts is the same as in ERM-ALTS^1.

PRACTITIONER ALTERNATES (represented by a dot) reflects the actual alternate set that was used in that instance. These alternates were essentially chosen by QUOTA-BASED, with a few context-driven adjustments. Note that there are many alternate sets satisfying a given set of quotas, so we do not expect the practitioner alternates to perfectly match the performance of QUOTA-BASED.

We further contextualize our results by showing the loss under two extremes: $A = \emptyset$, the *worst case* where no alternates are used, and $A = N$, the (unattainable) *best case*, where one can choose alternates from the entire pool (akin to being able to choose alternates *after* seeing the dropout set).

Loss Estimation. We cannot precisely evaluate an alternate set A's true expected loss $L^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}})$ due to the exponential support of $\mathcal{D}_{\boldsymbol{\rho}}$, which motivated our ERM approach. We thus take a sampling approach to evaluation as well: we compute the approximate loss $L^{\ell_1}(A; \hat{\mathcal{D}}_{\boldsymbol{\rho}})$, where $\hat{\mathcal{D}}_{\boldsymbol{\rho}}$ is the result of drawing 300 samples from $\mathcal{D}_{\boldsymbol{\rho}}$ — the same sample size used in *running* our ERM algorithms.

5.1 Evaluation on *realized* dropout set with *estimated* $\tilde{\rho}$

First, in each instance *US 1-3*, we evaluate the loss of each algorithm on the *actual* dropout set that occurred in that instance, corresponding to a single draw from the distribution $\mathcal{D}_{\boldsymbol{\rho}}$. As such, this experiment aims to test *what would have happened had each algorithm been run in this assembly.*

To emulate having trained on historical data, in *US-1* we estimate $\tilde{\boldsymbol{\rho}}_1$ on only datasets *US-2* and *US-3*, and analogously we get $\tilde{\boldsymbol{\rho}}_2$ and $\tilde{\boldsymbol{\rho}}_3$ for the other datasets. As will be the case for multiple analyses in this section, we study how the loss changes as the alternate budget a is increased. Figure 3 shows that while no algorithm is clearly dominant, the two non-ERM algortihms, QUOTA-BASED and GREEDY, have the highest losses for most values of a. In *US-2*, ERM-ALTS^1 seems to perform the best; in the *US-1* and *US-3*, ERM-ALTS^1-EQ performs the best (although in *US-3*, only one person dropped out, so the maximum possible loss is very low).

It may seem surprising that ERM-ALTS$^{\ell_1}$ does not always dominate, given that it is precisely optimizing the L^{ℓ_1} loss. There are two possible explanations: ERM-ALTS$^{\ell_1}$ performs well on the

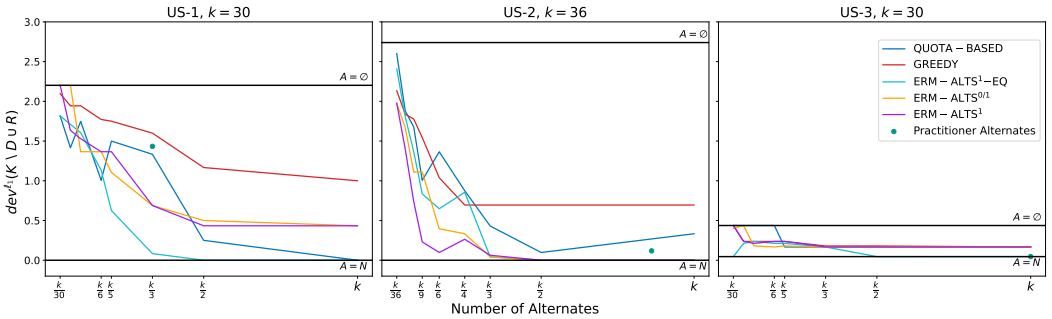

Fig. 3. In each instance *US-j* we are given all elements of $\mathcal{I}$, plus the realization of D and dropout probability estimates $\tilde{\boldsymbol{\rho}}_j$. We run each algorithm on all inputs except D, and each produces an A. What is plotted is then $dev^{\ell_1}(K \setminus D \cup R)$, where R is chosen from each A to minimize $dev^{\ell_1}(K \setminus D \cup R)$.

dropout set *distribution*, but we got unlucky with this realized draw of D; or, our estimated dropout probabilities are inaccurate. While one cannot tease apart these explanations conclusively, the dominance of ERM-ALTS$^{\ell_1}$-EQ suggests that inaccurate dropout probability estimates is a contributing factor (since it is just ERM-ALTS$^{\ell_1}$ without the dropout probabilities). This is supported empirically: we find that our probability estimates are *well-calibrated* to the realized dropouts in *US-2*, but far from it in *US-1* and *US-3*, precisely tracking where ERM-ALTS$^{\ell_1}$ dominates (Appendix E.4).

We draw two conclusions from these findings. First, the lower loss of ERM algorithms suggests the importance of considering the problem's combinatorial structure. Second, ERM-ALTS$^{\ell_1}$ exhibits strong performance when it has reasonably well-calibrated predictions — and such predictions are possible to achieve. Importantly, the quality of our dropout probability estimates here should be taken as a weak lower bound on the achievable prediction quality in practice. The data we had access to was on-hand and not collected or curated for this purpose, so it was highly imperfect for our prediction task compared to what practitioners could have access to at steady state.[3]

5.2 Evaluation on the *dropout set distribution* when $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}$

We now evaluate these algorithms' L^{ℓ_1} loss *in expectation over dropout sets, assuming the ground-truth dropout probabilities are known*. For the sake of realism, we let these ground-truth probabilities $\boldsymbol{\rho}$ be those learned by training on *US-1*, *US-2*, and *US-3*. Here, we expect ERM-ALTS^1 to outperform all other algorithms, as it is precisely optimizing our evaluation objective; what we are testing here is *the extent* to which various algorithms dominate others, and *how efficiently* each algorithm uses alternates, i.e., how quickly the loss approaches the unattainable optimum of $A = N$.

In Figure 4, we see that ERM-ALTS$^{\ell_1}$ indeed outperforms all other algorithms. It also uses alternates highly efficiently, achieving loss *matching* that of the offline solution (choosing $A = N$) with an alternate budget of only $k/2$, which amounts to 7%, 8%, and 16% of n across *US-1*, *US-2*, and *US-3*, respectively. Comparing the algorithms more broadly, we see that QUOTA-BASED and GREEDY are again reliably dominated by all three ERM algorithms, reinforcing the weakness of heuristics for this problem. Furthermore, the fact that ERM-ALTS$^{\ell_1}$ dominates both GREEDY (ignores combinatorics)

[3]Our data originally contained 25 panels, but most were dropped due to inconsistent features. The data also did not include features that practitioners anecdotally believe to be predictive of dropping out (e.g., living far from the panel's physical location). Finally, we did not have records of recruitment process issues that might have affected participation in specific datasets (e.g., emails going to spam, or other external shocks). If panel organizers collected consistent features, recorded more dropout-related attributes, and systematically noted external shocks, predictions could likely be significantly improved.

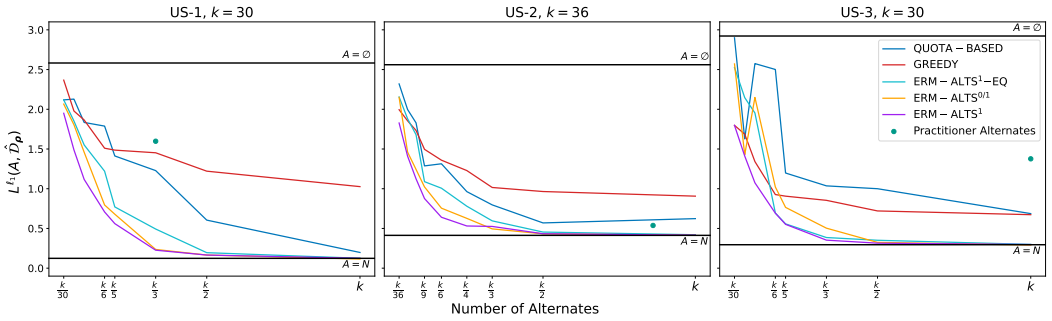

Fig. 4. The dropout probabilities known to the relevant algorithms is $\boldsymbol{\rho}$. We evaluate loss according to an empirical estimate of $\mathcal{D}_{\boldsymbol{\rho}}$, as described under **Loss Estimation**. Versions of these plots that include standard deviations over $D \sim \mathcal{D}_{\boldsymbol{\rho}}$ are in Appendix E.5.

and ERM-Alts^1-Eq (ignores dropout probabilities) shows us that both our innovations — learning $\boldsymbol{\rho}$ *and* considering the problem's combinatorics — are individually important to achieving low loss.

Because ERM-Alts^1's dominance here is by construction, we repeat this analysis for four other performance metrics: *the L^{ℓ_1} loss counting only violations of lower quotas*, since too few people is generally worse than too many; *the maximum deviation from any quota*, either raw or normalized by the quota size; and *the number of groups excluded from the panel*. The results, located in Appendix E.6, show that ERM-Alts$^{\ell_1}$ is reliably the lowest-loss algorithm on these other metrics as well.

5.3 Robustness: Evaluation on the *dropout set distribution* when $\tilde{\boldsymbol{\rho}} \neq \boldsymbol{\rho}$

Finally, we isolate the effects of prediction errors on the performance of these algorithms. We begin with the ground-truth probabilities $\boldsymbol{\rho}$ from the previous experiment, and then perturb each entry of $\boldsymbol{\rho}$ such that $\tilde{\rho}_i \sim Unif([\max\{0, \rho_i - \gamma\}, \min\{1, \rho_i + \gamma\}])$. The true dropout distribution remains $\mathcal{D}_{\boldsymbol{\rho}}$, but the algorithms now receive $\tilde{\boldsymbol{\rho}}$. We repeat the experiment $r = 25$ times per error level γ.

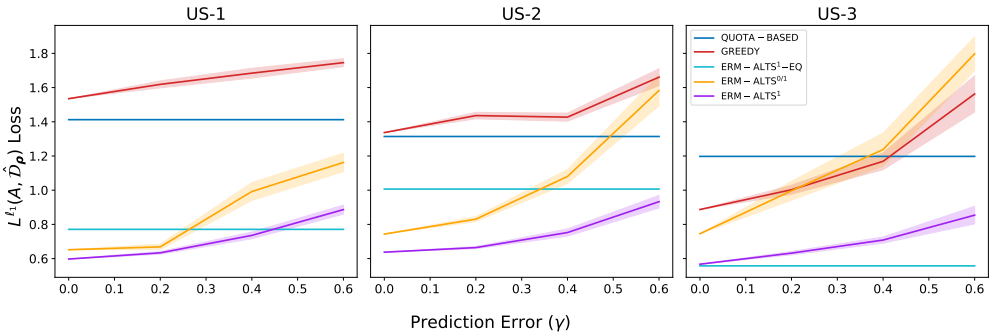

Fig. 5. $a = 6$ (between $k/6$ and $k/5$ for all instances). Per each of the $r = 25$ reps: we run each algorithm (other than the two which don't use dropout probabilities) on $\mathcal{I}$ with perturbed dropout probabilities $\tilde{\boldsymbol{\rho}}$. Then, we evaluate A's loss according to an empirical estimate of $\mathcal{D}_{\boldsymbol{\rho}}$, as described under **Loss Estimation**. The plot shows the mean and standard error of the loss over reps, with losses of Quota-Based and ERM-Alts$^{\ell_1}$-Eq shown for comparison.

As expected, all algorithms' loss increases as the prediction error γ increases. First comparing algorithms that use predictions, the separation is clear: ERM-Alts$^{\ell_1}$ begins with lower loss *and* its

loss increases more slowly as γ grows. Comparing ERM-ALTS$^{\ell_1}$ and ERM-ALTS$^{\ell_1}$-EQ, we see that in *US-1* and *US-2*, large prediction errors are required ($0.4, > 0.6$) before it is worth ignoring dropout probability estimates and opting for ERM-ALTS$^{\ell_1}$-EQ. In *US-3*, ERM-ALTS$^{\ell_1}$ and ERM-ALTS$^{\ell_1}$-EQ have identical performance under perfect predictions, but as γ grows to 0.6, the gap in loss between the algorithms grows only by 0.25 (where the max possible L^{ℓ_1} loss in *US-3* is 2.9, by Figure 4).

6 Discussion

Based on these results, our ERM algorithms are immediately practical. They all — but ERM-ALTS$^{\ell_1}$ especially — offer the prospect of drastically decreasing the number of alternates practitioners must select and compensate, while still maintaining or even decreasing the extent to which dropouts compromise representation. Even if a practitioner group does not have the requisite data to predict dropout rates (or simply does not want to), ERM-ALTS$^{\ell_1}$-EQ still outperforms both heuristic benchmarks significantly, thereby offering substantial practical improvement even for practitioners without access to predictions.

Our ERM algorithms can also be generalized to handle additional practical concerns and other approaches to handling dropout. Chief among such practical concerns is that *alternates* may drop out in addition to panelists. In fact, our ERM algorithms can be directly applied to hedge optimally against such additional dropouts, even when alternates drop out according to a different distribution. Furthermore, our ERM algorithms can be used to hedge against dropouts in many different ways beyond selecting alternates: they can be used to select loss-minimizing *extra panelists* while obeying additional quotas; to select loss-minimizing *panels* directly; or to select loss-minimizing panels and alternates in conjunction. We describe the details of all these extensions in Appendix F.

Future work: retaining the randomness of sortition. We now examine how alternate selection fits within the broader pipeline of *sortition*, the ubiquitously-used process of *randomly* selecting citizens' assembly members. Past work on sortition algorithms has overlooked dropout, focusing on the selection of the original panel. It has gone to great lengths to design the randomness of sortition such that pool members' chances of selection for the panel are *as equal as possible* within the quotas — a property which confers normative ideals like fairness, resistance to manipulation, and more [Flanigan et al., 2023]. Now, as we expand the sortition model to encompass dropouts and the subsequent deployment of alternates, we encounter a barrier to equalizing pool members' chances of selection: existing alternate selection methods and ours alike choose alternates *deterministically*, thereby creating a deterministic route by which agents can reach the panel within a selection process that is supposed to be random. In fact, in trying to mitigate dropout, alternate selection algorithms may be distinctly *unfair*, privileging groups with lower likelihood of dropping out.

This motivates the natural next research question: *How would one randomize alternate selection?* We give some intuition here. First, suppose we wanted to "randomize" our ERM algorithms *without* compromising on robustness to dropout. To do this, we would find as many alternate sets as possible with optimal loss, and then randomize over them with the goal of maximally equalizing pool members' chances of selection, exactly as state-of-the-art sortition algorithms randomize over quota-satisfying panels [Flanigan et al., 2021a]. This approach is not guaranteed to work, however; in the worst case, there may be only one such optimal alternate set, and our algorithm would remain deterministic. Intuitively, as we permit ourselves to randomize over alternate sets with increasingly suboptimal loss, we gain the flexibility to randomize over more alternate sets and can thus further equalize pool members' chances of being selected. This exposes a fundamental trade-off between *the randomness that defines sortition* and *robustness of representation to dropout*; we leave the development of algorithms that strike this trade-off optimally to future work.

References

Carmel Baharav and Bailey Flanigan. 2024. Fair, Manipulation-Robust, and Transparent Sortition. In *Proceedings of the 25th ACM Conference on Economics and Computation*. 756–775.

Gerdus Benadè, Paul Gölz, and Ariel D Procaccia. 2019. No stratification without representation. In *Proceedings of the 2019 ACM Conference on Economics and Computation*. 281–314.

Bürgerrat. 2023. French citizens' assembly supports assisted dying. Available at https://www.buergerrat.de/en/news/french-citizens-assembly-supports-assisted-dying/ (2024/02/11).

Deutscher Bundestag. 2023. The Legislative Process: Focus on Nutrition. https://www.bundestag.de/en/parliament/process-nutrition-995912. Accessed: 2025-02-03.

Soroush Ebadian, Gregory Kehne, Evi Micha, Ariel D Procaccia, and Nisarg Shah. 2022. Is Sortition Both Representative and Fair? *Advances in Neural Information Processing Systems* 35 (2022).

Soroush Ebadian and Evi Micha. 2023. Boosting Sortition via Proportional Representation. Manuscript.

Ethan Fetaya. 2016. Introduction to Statistical Learning Theory - Lecture 5. Online. Available: https://www.wisdom.weizmann.ac.il//%7Eethanf/teaching/ItSLT_16/lectures/lec_fat_no_anim.pdf.

Bailey Flanigan, Paul Gölz, Anupam Gupta, Brett Hennig, and Ariel D Procaccia. 2021a. Fair algorithms for selecting citizens' assemblies. *Nature* 596, 7873 (2021), 548–552.

Bailey Flanigan, Paul Gölz, Anupam Gupta, and Ariel D Procaccia. 2020. Neutralizing self-selection bias in sampling for sortition. *Advances in Neural Information Processing Systems* 33 (2020), 6528–6539.

Bailey Flanigan, Paul Gölz, and Ariel Procaccia. 2023. *Mini-Public Selection: Ask What Randomness Can Do for You*. Ash Institute for Democratic Governance and Innovation.

Bailey Flanigan, Gregory Kehne, and Ariel D Procaccia. 2021b. Fair sortition made transparent. *Advances in Neural Information Processing Systems* 34 (2021), 25720–25731.

Bailey Flanigan, Jennifer Liang, Ariel D Procaccia, and Sven Wang. 2024. Manipulation-Robust Selection of Citizens' Assemblies. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Louis-Gaëtan Giraudet, Bénédicte Apouey, Hazem Arab, Simon Baeckelandt, Philippe Begout, Nicolas Berghmans, Nathalie Blanc, Jean-Yves Boulin, Eric Buge, Dimitri Courant, et al. 2022. "Co-construction" in deliberative democracy: lessons from the French Citizens' Convention for Climate. *Humanities and Social Sciences Communications* 9, 1 (2022), 1–16.

The Irish Citizens' Assembly Project. 2019. http://www.citizenassembly.ie/work/.

Jianzhu Ma, Jian Peng, Sheng Wang, and Jinbo Xu. 2013. Estimating the partition function of graphical models using Langevin importance sampling. In *Artificial Intelligence and Statistics*. PMLR, 433–441.

OECD. 2020. *Innovative Citizen Participation and New Democratic Institutions*. Policy Report. OECD Publishing. Accessed: 2025-02-03.

Participedia. 2021. Global Assembly on the Climate & Ecological Crisis. https://participedia.net/case/global-assembly-on-the-climate-ecological-crisis. Accessed: 2025-02-03.

Dominik Peters, Ariel D Procaccia, and David Zhu. 2022. Robust rent division. *Advances in Neural Information Processing Systems* 35 (2022), 13864–13876.

Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

A Supplemental Materials from Section 1

A.1 Argument showing NP-hardness

Lemma A.1 *Given a panel K, an alternate set A, a dropout set D, and a set of quotas $\boldsymbol{l}, \boldsymbol{u}$, it is NP-hard (in asymptotic parameter a) to compute the optimal replacement set $R \subseteq A : |R| \leq |D|$.*

Proof. We show that simply *deciding* whether there exists an R such that $K \setminus D \cup A$ satisfies the quotas is NP-hard. Let it be the case that $a = k$, and let the dropout probabilities be constant so that with high probability, $|D| \in \Theta(a)$ (note that we know that $|D| \leq a$). Let the quotas $\boldsymbol{l}, \boldsymbol{u}$ be tight. Let $D_{f,v} = \sum_{i \in D} : f(i) = v$ be shorthand for the number of agents in D with value v for feature f. Now, let our *modified* quotas $\boldsymbol{l'}, \boldsymbol{u'}$ also be tight, such that $l_{f,v} = u_{f,v} = D_{f,v}$ for all $f, v \in FV$.

Note that the problem of deciding whether there exists an R such that $K \setminus D \cup A$ satisfies the quotas $\boldsymbol{l}, \boldsymbol{u}$ is equivalent to checking whether there is a $|D|$-sized subset of A satisfying the modified quotas $\boldsymbol{l'}, \boldsymbol{u'}$. By Theorem 1 of [Flanigan et al., 2021a] (via a reduction from Set Cover), this problem is known to be NP-hard in a. □

B Supplemental Materials from Section 2

B.1 ILP formulation for computing best replacement set

Given a deviation function dev, an alternate set A, and a dropout set D, one can compute the deviation-minimizing replacement set via the following ILP. Because a generic deviation function dev directly intakes a set (a strange variable to represent in an ILP), we will write two ILPs, one for $dev^{0/1}$ and one for dev^{ℓ_1}. we will write ILPs for variables are $y_1, \ldots, y_a$, where $y_i \in \{0,1\}$ is an indicator of whether alternate i is included in the replacement set R.

$dev^{0/1}$:

$$\begin{aligned}
\min \quad & \sum_{f,v} z_{f,v} \\
& \sum_{i \in A} y_i \leq |D| \\
& l_{f,v} - \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in A} y_i \, \mathbb{I}(f(i) = v) \right) \leq z_{f,v} \, (|K| + |A|) && \text{for all } f \in F, v \in V_f \\
& - u_{f,v} + \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in A} y_{i,D} \, \mathbb{I}(f(i) = v) \right) \leq z_{f,v} \, (|K| + |A|) && \text{for all } f \in F, v \in V_f \\
& y_i \in \{0,1\} && \text{for all } i \in A \\
& z_{f,v} \in \{0,1\} && \text{for all } f \in F, v \in V_f
\end{aligned}$$

dev^{ℓ_1}:

$$\begin{aligned}
\min \quad & \sum_{f,v} z_{f,v} \\
& \sum_{i \in A} y_i \leq |D| \\
& l_{f,v} - \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in A} y_i \, \mathbb{I}(f(i) = v) \right) \leq z_{f,v} \, u_{f,v} && \text{for all } f \in F, v \in V_f \\
& - u_{f,v} + \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in A} y_{i,D} \, \mathbb{I}(f(i) = v) \right) \leq z_{f,v} \, u_{f,v} && \text{for all } f \in F, v \in V_f \\
& y_i \in \{0,1\} && \text{for all } i \in A \\
& z_{f,v} \in \mathbb{R}_{\geq 0} && \text{for all } f \in F, v \in V_f
\end{aligned}$$

B.2 Formulation of $\textsc{Opt}(\mathcal{D}, \mathcal{I})$

For concreteness, we define this ILP not for generic dev, but for our two specific deviation functions $dev^{0/1}$ and dev^{ℓ_1}. We express these programs simultaneously, where the $(* \, dev^{0/1} \, *)$ and $(* \, dev^{\ell_1} \, *)$ tags denote constraints appearing only in the $dev^{0/1}$ and dev^{ℓ_1} versions of Opt, respectively.

Variables in Opt. The variables x_i for $i \in N$ are indicator variables of whether pool member $i \in N$ is included in the alternate set. $y_{i,D}$ is then the indicator variable of whether i is chosen to be on the replacement set for dropout set D (note that for $y_{i,D} = 1$, it must be that i was in the alternate set to begin with, implying the constraint that $y_{i,D} \leq x_i$). The variables $z_{D,f,v}$ can be thought of as the linear deviation from the quotas, specifically on feature f and value v, that is incurred when we use the selected replacement set on dropout set D. Finally, the variables d_D capture the deviation (either

linear or binary) on set D. The objective minimizes the expected value of the deviation function over the dropout set distribution $\mathcal{D}$.

$$
\begin{aligned}
&\textsc{Opt}^{dev}(\mathcal{I}, \mathcal{D}) \\
&\quad \min \sum_{D \in 2^K} d_D \cdot \mathcal{D}(D) \\
&\quad \text{s.t.} \sum_{i \in N} x_i = a \\
&\qquad y_{i,D} \le x_i && \forall\, i \in N,\ D \in 2^K \\
&\qquad \sum_{i \in N} y_{i,D} \le |D| && \forall\, D \in 2^K \\
&\qquad l_{f,v} - \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in N} y_{i,D}\, \mathbb{I}(f(i) = v) \right) \le z_{D,f,v}\, u_{f,v} && \forall\, D \in 2^K, f, v \in FV \\
&\qquad -u_{f,v} + \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in N} y_{i,D}\, \mathbb{I}(f(i) = v) \right) \le z_{D,f,v}\, u_{f,v} && \forall\, D \in 2^K, f, v \in FV \\
&\qquad \sum_{f,v} z_{D,f,v} \le d_D \qquad (*\, dev^{\ell_1}\, *) && \forall\, D \in 2^K \\
&\qquad \sum_{f,v} z_{D,f,v} \le d_D \cdot (|K| + a)|FV| \qquad (*\, dev^{0/1}\, *) && \forall\, D \in 2^K \\
&\qquad d_D \in \mathbb{R}_{\ge 0} \qquad (*\, dev^{\ell_1}\, *) && \forall\, D \in 2^K \\
&\qquad d_D \in \{0, 1\} \qquad (*\, dev^{0/1}\, *) && \forall\, D \in 2^K \\
&\qquad x_i \in \{0, 1\},\ y_{i,D} \in \{0, 1\},\ z_{D,f,v} \in \mathbb{R}_{\ge 0} && \forall\, i \in N, D \in 2^K, f, v \in FV
\end{aligned}
$$

C Supplemental Materials from Section 3

C.1 Proof of Theorem 3.4 (Lower Bounds)

Proof. We begin by proving that our relative values of n and a as constructed in our instance are compatible. We will set $k = |F| = \lfloor \log \binom{n}{a} \rfloor - 2a$, and we will require that $n \geq a2^{|F|/a}$. Then, it needs to be the case that for all $a, n \in \mathbb{N}_{\geq 1}$ and $n \geq a$,

$$n \geq a2^{\left(\log \binom{n}{a} - 2a\right)/a}.$$

$$a2^{\left(\log \binom{n}{a} - 2a\right)/a} = a2^{\log \binom{n}{a}/a - 2} = a/4 \cdot \binom{n}{a}^{1/a} \leq a/4 \cdot \left(\left(\frac{ne}{a}\right)^a\right)^{1/a} = a/4 \cdot \frac{ne}{a} = en/4 < n.$$

With our instance parameters established as valid, we proceed by first constructing the instance, then proving the VC lower bound, then extending that result to prove a Pdim lower bound.

Instance construction. Construct $\mathcal{I}$ as follows. Let $k = |F| = \lfloor \log \binom{n}{a} \rfloor - 2a$. Let the features f be numbered $1 \ldots |F|$. Let each feature be binary-valued, so $V_f = \{0, 1\}$ for all $f \in F$. Let the quotas be $l_{f,0} = 0$, $l_{f,1} = 1$ and $u_{f,0} = u_{f,0} = \infty$ for all $f \in F$, so the only requirement is that the panel contains at least one person with a 1 for every feature.

Constructing the panel. Define the panel K as follows: let the panelists are numbered $i \in [k]$ and let the i-th panelist's feature vector be the i-th standard $|F|$-length basis vectors, so that

$$w(1) = \underbrace{100\ldots000}_{\text{length } |F|},\ w(2) = 010\ldots000, \ldots,\ w(k-1) = 000\ldots010,\ w(k) = 000\ldots001.$$

Note that K satisfies these quotas.

Constructing the pool. We will construct only a subset of the pool and leave the rest to be constructed arbitrarily. The subset we will construct consists of a groups of agents $G_1, \ldots, G_a$. To define these groups, first define corresponding integers $m_1, \ldots, m_a$, where $m_j \in \{\lfloor |F|/a \rfloor, \lfloor |F|/a \rfloor\}$. Which m_j are assigned the floor versus ceiling doesn't matter so long as $\sum_{j \in [a]} m_j = |F|$.

Now, we divide up the features $1 \ldots |F|$ into sequential sets of features of size $m_1, \ldots, m_a$; call these sequential sets of features g_j, so

$$g_1 = \{1, 2, \ldots, m_1\}, \quad g_2 = \{m_1 + 1, m + 2, \ldots, m_1 + m_2\}, \quad \ldots, \quad g_a = \{\sum_{j=1}^{a-1} m_i + 1, \ldots, a|F|\}.$$

We now define our a corresponding subgroups of the pool $G_1 \ldots G_a$, where each G_j contains 2^{m_j} agents. G_j is composed of agents who have 0s for all features other than those in g_j, and for the r_j features in g_j have values described by a unique vector $u_i \in \{0, 1\}^{m_i}$. That is, each $i \in G_j$ is defined by some unique $u_i \in \{0, 1\}^g$, and their overall feature vector is then $0 \ldots 0 u_i 0 \ldots 0$, where u_i occurs over the indexes in g_j.

Construction of shattered set for VC **lower bound.** We will now define a collection of k dropout sets to be shattered, where each dropout set consists of a single panelist.

$$D_1 = \{1\},\ D_2 = \{2\}, \ldots,\ D_k = \{k\}.$$

Fix any vector of labels $b \in \{0, 1\}^k$, indexed as b_i. Our goal is to construct an alternate set A_b such that A_b "realizes" this vector of labels; that is, for all $i \in [k]$, we have that

$$b_i = h^{0/1}_{A_b}(D_i).$$

In words, the 1 label at b_i reflects that A_b cannot fully restore the quotas after D_i drops out, and the 0 label reflects that it can. Let $c \in \{0, 1\}^k$ be the "opposite" of b, so $c_i = 1 \iff b_i = 0$. We will construct A_b by taking one agent from each set G_j, thereby taking a agents overall. From the set

G_j, add to A_b the agent whose unique vector of values u_i over features g_j matches c over those same indices. Formally, we add agent $i \in G_j$ to the alternate set for whom

$$f(i) = c_f \qquad \text{for all } f \in g_j.$$

Now, it remains to show that

$$b_i = 0 \iff h^{0/1}_{A_b}(D_i) = 0,$$

which proves the claim. Recall that achieving a 0 label for dropout set D_i just means making sure at least one agent in A_b has a 1 for the i-th feature, this is the 1-value we lose when $D_i = \{i\}$ drops out.

To prove the forward direction, on any $i \in [k]$ such that $b_i = 0$, A_b must contain an agent with a 1 i-th position: this is exactly the agent we took from G_j where $i \in g_j$. This agent must have a 1 for the i-th feature because we choose them to match the labels in c (the opposite of the labels in b) on the features in g_j; their ith value must be 1 then, since $b_i = 0 \implies c_i = 1$. This means that

$$b_i = 0 \implies h^{0/1}_{A_b}(D_i) = 0.$$

On the other hand, fix an $i \in K$ such that $b_i = 1$, and let $g_j \ni i$. The agent added to A_b from group G_j must have a 0 for the ith feature by construction of this agent to match c on the indices of g_j. The agents added from groups $G_{j'}$ for all $j \in [a] \setminus j$ must have 0s for all features outside $g_{j'}$, which includes index i by assumption. This means that

$$b_i = 0 \implies h^{0/1}_{A_b}(D_i) = 1.$$

We have shattered a collection of dropout sets of size k, which implies that $VC(\mathcal{H}^{0/1}(\mathcal{I})) \geq k = \left\lfloor \log \binom{n}{a} \right\rfloor - 2a$, and we obtain our tight lower bound of

$$VC(\mathcal{H}^{0/1}(\mathcal{I})) \geq \left\lfloor \log \binom{n}{a} \right\rfloor - 2a \geq \log \left(\frac{n}{a}\right)^a - 1 - 2a = a(\log n - \log a - 2) - 1 \in \Omega(a \log n). \quad (7)$$

Extension to Pdim **lower bound.** We will prove our Pdim lower bound by proving something stronger: that for all $\mathcal{I}'$,

$$\text{VC}(\mathcal{H}^{0/1}(\mathcal{I}')) \geq t \implies \text{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I}')) \geq t \qquad \text{for all } t \in \mathbb{R}^+. \quad (8)$$

Then, it will follow from Equation (7) that for our instance constructed above,

$$\text{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \in \Omega(a \log n).$$

It remains to prove the statement in Equation (8). Suppose we have $\mathcal{I}$ for which $\text{VC}(\mathcal{H}^{0/1}(\mathcal{I})) \geq t$. That means we can find a collection of t dropout sets $\mathbf{D} = \{D_1, \ldots, D_t\}$ for which for any labeling $b \in \{0,1\}^t$ there exists $h^{0/1} \in \mathcal{H}^{0/1}(\mathcal{I})$ for which $h^{0/1}(D_i) = b_i$ for all $i \in [t]$

Now, we will show that for this same instance $\mathcal{I}$ and same collection of t dropout sets $\mathbf{D}$, there exists a witness vector $r = (r_1, \ldots, r_t)$ such that for all $b \in \{0,1\}^t$, there exists a hypothesis $h^{\ell_1} \in \mathcal{H}^{\ell_1}(\mathcal{I})$ for which

$$\mathbb{I}(h^{\ell_1}(D_i) > r_i) = b_i \qquad \text{for all } i \in [t].$$

This is precisely what is required to show that $\text{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \geq t$.

Define r such that $r_i = 1/k+1$ for all $i \in [t]$, and fix an arbitrary $b \in \{0,1\}^t$. Let the alternate set A_b and its corresponding binary hypothesis $h^{0/1}_{A_b} \in \mathcal{H}^{0/1}$ be the hypothesis that realizes the labeling b in the binary case, i.e., such that

$$\mathbb{I}(h^{0/1}_{A_b}(D_i) = 1) = b_i,$$

Note that such a hypothesis must exist by the fact that $\text{VC}(\mathcal{H}^{0/1}(\mathcal{I})) \geq t$.

Examining the *linear* hypothesis corresponding to the same alternate set, $h^{\ell_1}_{A_b}$, the key observation — which concludes the proof — is that for all dropout sets $D_i, i \in [t]$,

$$\mathbb{I}\left(h^{\ell_1}_{A_b}(D_i) > 1/k+1\right) \overset{(1)}{=} \mathbb{I}\left(h^{0/1}_{A_b}(D_i) = 1\right) \overset{(2)}{=} b_i. \tag{9}$$

Equality (2) holds simply because we chose chose A_b such that $h^{0/1}_{A_b}$ realizes the labeling b. The critical step is equality (1), and it follows from the definition of the range of hypotheses in $\mathcal{H}^{\ell_1}(\mathcal{I})$. By definition, the range of any hypothesis in this class is fundamentally the range of the dev^{ℓ_1}, which consists of the sum of rational numbers whose denominators are $u_{f,v}$ for $f \in F, v \in V_f$. Using that all such $u_{f,v} \leq k$, it follows that the smallest possible nonzero rational number in the range of any hypothesis in $\mathcal{H}^{\ell_1}(\mathcal{I})$ is

$$\frac{1}{\max_{f\in F, v\in V_f} u_{f,v}} \geq \frac{1}{k} > \frac{1}{k+1}.$$

This implies that for any $h^{\ell_1} \in \mathcal{H}^{\ell_1}(\mathcal{I})$ and any dropout set D,

$$h^{\ell_1}(D) > 1/k+1 \iff h^{\ell_1}(D) > 0. \tag{10}$$

The only required remaining observation is that for all alternate sets A and corresponding hypotheses $h^{\ell_1}_A \in \mathcal{H}^{\ell_1}(\mathcal{I})$ and $h^{0/1}_A \in \mathcal{H}^{0/1}(\mathcal{I})$, plus all dropout sets D, we have that

$$h^{\ell_1}_A(D) > 0 \iff h^{0/1}_A(D) > 0 \tag{11}$$

because both inequalities correspond exactly to the case that A cannot restore the quotas when D drops out.

Putting the equivalences together in Equations (10) and (11), we get that for all A and all D,

$$h^{\ell_1}_A(D) > 1/k+1 \iff h^{0/1}_A(D) > 0,$$

exactly as needed to show equality (2) in Equation (9). □

C.2 Proof of Theorem 3.5

Proof: Upper Bounds. The upper bound is proven by bounding the size of the hypothesis class in terms of $|F|$ rather than n and a. Note that the maximum number of possible unique feature vectors that could ever occur in an instance is

$$m = \prod_{f\in F} |V_f| \leq \left(\max_{f\in F} |V_f|\right)^{|F|}$$

Then, the maximum *unique* possible alternate sets in any instance is defined by the number of ways to choose a alternates from a set of m unique elements, potentially with duplicates. It is well-known that the way this is as follows, where we are slightly abusing notation here to now treat $\mathcal{A}$ as the set of all *unique* alternate sets:

$$|\mathcal{A}| \leq \binom{m+a-1}{a} \leq \left(\frac{e(m+a-1)}{a}\right)^a.$$

Taking the logarithm of this quantity and plugging in our upper bound for m, we get

$$\begin{aligned}|\mathcal{A}| \le a\ (\log(m+a-1) - \log a + 1) &\le a\left(\log\left(\left(\max_{f\in F}|V_f|\right)^{|F|} + a - 1\right) - \log a + 1\right)\\ &\le a\left(\log\left(\max_{f\in F}|V_f|\right)^{|F|} + 1\right) \qquad (*)\\ &= a\left(|F|\log\max_{f\in F}|V_f| + 1\right)\end{aligned}$$

The starred step is shown as follows. Consider the inequality written more simply as

$$\log(x+a-1) - \log a + 1 \le \log(x) + 1 \iff \log(x+a-1) - \log a - \log(x) \le 0.$$

Take the derivative of the LHS with respect to x, and get

$$\frac{1}{a+x-1} - \frac{1}{x}.$$

This is weakly negative for all $x \ge 1, a \ge 1$, both of which are true in any non-degenerate instance of our problem. This means the original LHS is maximized at $x = 1$. At $x = 1$,

$$\log(1+a-1) - \log a - \log(1) = 0 \le 0,$$

concluding the proof. □

Proof: Lower Bounds. The proof of the lower bound proceeds very similarly to the lower bound in Theorem 3.4.

Instance construction. Construct $\mathcal{I}$ as follows. Let $|F| \le k$, and let $n \ge a2^{|F|/a}$. Let the features f be numbered $1 \dots |F|$. Again, $V_f = \{0,1\}$ for all $f \in F$ and the quotas are again $l_{f,0} = 0$, $l_{f,1} = 1$ and $u_{f,0} = u_{f,0} = \infty$ for all $f \in F$. The panel is constructed as in the proof of Theorem 3.4, except now there are $|F|$ panelists $1, 2, \dots, |F|$ with unique $|F|$-length basis vectors, with panelist i having value 1 for feature i and 0s for all other features. For the remaining $k - |F|$ panelists, let them have value 0 for all features. We construct the pool exactly as in the proof of Theorem 3.4.

Construction of shattered set for VC **lower bound.** We construct the collection of dropout sets to be shattered as before, except now we construct only $|F|$ of them, from panelists $1 \dots |F|$:

$$D_1 = \{1\},\ D_2 = \{2\}, \dots,\ D_{|F|} = \{|F|\}.$$

Analogously to before, we fix any vector of labels $b \in \{0,1\}^{|F|}$, indexed as b_i. We construct our alternate set A_b exactly as before, taking one person from each set G_j. The rest of the proof of the VC lower bound follows for the same reason, as does the extension to the Pdim lower bound, giving us lower bounds of

$$\mathrm{VC}(\mathcal{H}^{0/1}(\mathcal{I})),\ \mathrm{Pdim}(\mathcal{H}^{\ell_1}(\mathcal{I})) \ge |F|.$$ □

C.3 Restatement of Existing PAC-Learning Bounds

We use the following known PAC learning bounds, along with the fact that uniform convergence is sufficient for agnostic learnability (Theorem 6.7, [Shalev-Shwartz and Ben-David, 2014]).

Lemma C.1 (Theorem 6.8 of [Shalev-Shwartz and Ben-David, 2014]) *Fix constants $\varepsilon, \delta > 0$, and let $\mathcal{H}$ is a hypothesis class of functions with range $\{0,1\}$ such that $\mathrm{Pdim}(\mathcal{H}) < \infty$. Then, there exists a constant $C \in \mathbb{R}^+$ such that $\mathcal{H}$ has the uniform convergence property with sample complexity*

$$s(\delta,\varepsilon) \le C \cdot \frac{\mathrm{VC}(\mathcal{H}) + \log\left(\frac{1}{\delta}\right)}{\varepsilon^2}.$$

Lemma C.2 ([Fetaya, 2016]) *Fix constants $\varepsilon, \delta > 0$, and let $\mathcal{H}$ is a hypothesis class of functions with range in $\mathbb{R}$ such that $\text{Pdim}(\mathcal{H}) < \infty$. Then, there exists a constant $C \in \mathbb{R}^+$ such that $\mathcal{H}$ has the uniform convergence property with sample complexity*

$$s(\delta, \varepsilon) \leq C \cdot \frac{\text{Pdim}(\mathcal{H}) \cdot \log\left(\frac{1}{\varepsilon}\right) + \log\left(\frac{1}{\delta}\right)}{\varepsilon^2}.$$

D Supplemental Materials from Section 4

D.1 Full proof of Theorem 4.1

Proof. We construct a simple instance $\mathcal{I} = (N, K, \boldsymbol{l}, \boldsymbol{u}, a)$ as follows: there is only one binary feature, so $F = f_1$, and $V_{f_1} = \{0, 1\}$. The panel K is of size $k = 2\lceil \log_{(1-\gamma)} \alpha \rceil$ (note: must be ≥ 2, even, and positive) and it is comprised of $k/2$ agents with feature-value 0 and $k/2$ agents with feature-value 1. The quotas are tight, so $l_{f_1,1} = u_{f_1,1} = l_{f_1,0} = u_{f_1,0} = k/2$. The pool N contains at least k agents with feature-value 0 and at least k agents with feature-value 1, so our alternate set construction is unencumbered by any limitations of the pool. Finally, $a = k/2$. We set $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$ as follows, noting that indeed $\|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|_\infty = \gamma$.

$$\rho_i = \begin{cases} \gamma & \text{if } f_1(i) = 1 \\ 0 & \text{if } f_1(i) = 0 \end{cases} \qquad \tilde{\rho}_i = \begin{cases} 0 & \text{if } f_1(i) = 1 \\ \gamma & \text{if } f_1(i) = 0 \end{cases} \qquad \text{for all } i \in K.$$

First, observe that OPT-$A^{0/1}$ just consists of $a = k/2$ agents i with $f_1(i) = 1$, and $\widetilde{\text{OPT-}A^{0/1}}$ consists of $a = k/2$ agents i with $f_1(i) = 0$. This is because under both $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, exactly $a = k/2$ members of the panel have a non-zero chance of dropping out, and both of these optimal sets simply have backups for all of these agents, and no other agents. Note that $\mathcal{L}^{0/1}(\text{OPT-}A^{0/1}, \mathcal{D}_{\boldsymbol{\rho}}) = 0$. Further,

$$\mathcal{L}^{0/1}(\widetilde{\text{OPT-}A^{0/1}}, \mathcal{D}_{\boldsymbol{\rho}}) = 1 \cdot \Pr_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}[\exists\, i \in D \colon f_1(i) = 1] = 1 - (1-\gamma)^{k/2} \geq 1 - \alpha,$$

because $\widetilde{\text{OPT-}A^{0/1}}$ contains only agents with $f(i) = 0$ and therefore must incur $dev^{0/1}$ if *any* panelist with $f(i) = 1$ drops out. It follows that

$$\mathcal{L}^{0/1}(\widetilde{\text{OPT-}A^{0/1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{0/1}(\text{OPT-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}}) \geq 1 - \alpha.$$

Finally, it remains to show that for sufficient s,

$$\Pr[\text{ERM-}A^{0/1} = \text{OPT-}A^{0/1} \wedge \widetilde{\text{ERM-}A^{0/1}} = \widetilde{\text{OPT-}A^{0/1}}] \geq 1 - 2\delta,$$

which implies the claim. This argument proceeds two steps: (1) showing a constant separation between the loss of OPT-$A^{0/1}$ and that of any other alternate set A, and a symmetric separation for $\widetilde{\text{OPT-}A^{0/1}}$; and then (2) applying Corollary 3.6 to derive $s(\alpha, \delta, \gamma)$ such that for all $s \geq s(\alpha, \delta, \gamma)$, the desired event occurs with at least $1 - 2\delta$ probability.

First, consider any alternate set $A \neq \text{OPT-}A^{0/1}$. A contains $< k/2$ alternates with value 1, and thus it incurs $dev^{0/1}$ of 1 when all $k/2$ panelists with value 1 drop out. This dropout set occurs with probability $\gamma^{k/2}$. Hence,

$$\mathcal{L}^{0/1}(A; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{0/1}(\text{OPT-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}}) \geq \gamma^{k/2} > \gamma^k. \tag{12}$$

Now, recall that $k = 2\lceil \log_{(1-\gamma)} \alpha \rceil$ and let $s(\alpha, \delta, \gamma) \in \Theta\left(\frac{k/2 \cdot \log(k) + \log(1/\delta)}{\gamma^k}\right)$. By Corollary 3.6, we know that for all $s \geq s(\alpha, \delta, \gamma)$,

$$\Pr\left[\left|\mathcal{L}^{0/1}(\text{ERM-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{0/1}(\text{OPT-}A^{0/1}; \mathcal{D}_{\boldsymbol{\rho}})\right| \leq \gamma^k\right] \geq 1 - \delta.$$

Combining this fact with Equation (12), we get that

$$\Pr[\text{ERM-}A^{0/1} = \text{OPT-}A^{0/1}] \geq 1 - \delta.$$

By symmetry, under $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$ the same gap in 0/1 loss is induced when all panelists with *value 0* drop out, and thus for all $s \geq s(\alpha, \delta, \gamma)$, we have that

$$\Pr[\widetilde{\text{ERM-}A^{0/1}} = \widetilde{\text{OPT-}A^{0/1}}] \geq 1 - \delta.$$

By union bounding, we get that

$$\Pr[\text{ERM-}A^{0/1} = \text{OPT-}A^{0/1} \wedge \widetilde{\text{ERM-}A^{0/1}} = \widetilde{\text{OPT-}A^{0/1}}] \geq 1 - 2\delta. \qquad \square$$

D.2 Proof of Theorem 4.2

In these proofs, it will be convenient to use the following shorthand: For any $(f, v) \in FV$, define the feature-value linear deviation of a set S as:

$$dev^{\ell_1}_{f,v}(S, \mathcal{I}) = \frac{\max\{0, l_{f,v} - \sum_{i \in S} \mathbb{I}(f(i) = v), -u_{f,v} + \sum_{i \in S} \mathbb{I}(f(i) = v)\}}{u_{f,v}}$$

Then note that then we can express the overall linear deviation as the sum of the feature-value deviations: $dev^{\ell_1}(S, \mathcal{I}) = \sum_{(f,v) \in FV} dev^{\ell_1}_{f,v}(S, \mathcal{I})$.

Proof. Because the instance will be fixed throughout the proof, we will drop $\mathcal{I}$ from all our $\mathcal{L}^{\ell_1}$ and dev^{ℓ_1} functions, leaving it implicit.

The core of the proof is showing the following bound on the change in linear loss for *any* fixed alternate set A, when evaluated with respect to $\mathcal{D}_{\boldsymbol{\rho}}$ versus $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$:

$$|\mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\tilde{\boldsymbol{\rho}}})| \leq \gamma |FV|. \tag{13}$$

Once we have this bound, we can apply it to OPT-A^{ℓ_1} to show the following chain of inequalities, where the first inequality is by the optimality of $\widetilde{\text{OPT-}A^{\ell_1}}$ for $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$:

$$\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \leq \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \leq \gamma |FV|.$$

This gives us an almost analogous version of our desired bound for OPT-A^{ℓ_1} and $\widetilde{\text{OPT-}A^{\ell_1}}$. To relate $\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}})$ to $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}})$, we apply Equation (5) once more and derive $s(\varepsilon, \delta, \gamma)$ based on Corollary 3.7 to ensure that with probability $\geq 1 - \delta$, the loss of $\widetilde{\text{ERM-}A^{\ell_1}}$ on $\mathcal{D}_{\tilde{\boldsymbol{\rho}}}$ with $s \geq s(\varepsilon, \delta, \gamma)$ is within ε of its respective corresponding optimal alternate set. As we are showing an upper bound, we do not have to ensure that ERM-A^{ℓ_1} is close to OPT-A^{ℓ_1}. This completes the proof.

The high level approach to proving Equation (5) is to construct $k + 1$ intermediate probability vectors that incrementally transform $\boldsymbol{\rho}$ to $\tilde{\boldsymbol{\rho}}$ by altering one agent's dropout probability at a time. As such, let $\boldsymbol{\rho}^i = (\tilde{\rho}_1, \ldots, \tilde{\rho}_i, \rho_{i+1}, \ldots, \rho_k)$ for all $i \in [k]$. Then our quantity of interest can be rewritten as $|\mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^0}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^k})|$. Using a telescoping sum and the triangle inequality, we have that

$$\begin{aligned}
\mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^0}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^k})| &= \left| \sum_{i=1}^{k} \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^{i-1}}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^i}) \right| \\
&\leq \sum_{i=1}^{k} \left| \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^{i-1}}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^i}) \right|.
\end{aligned}$$

We will bound each individual term of this resulting sum separately. This will be done via a coupling argument, where we couple the random dropouts under $\boldsymbol{\rho}^i$ and $\boldsymbol{\rho}^{i+1}$. Formally, we are coupling

$$\boldsymbol{Y} = (Y_j \sim \text{ Bernoulli}(\rho_j^{i-1}) | j \in [k]) \quad \text{ and } \quad \boldsymbol{Y'} = (Y'_j \sim \text{ Bernoulli}(\rho_j^{i}) | j \in [k]),$$

whose entries describe whether each panelist dropped out when dropouts were sampled under the $\boldsymbol{\rho}^i$ and $\boldsymbol{\rho}^{i+1}$ vector, respectively. $\boldsymbol{Y}$ and $\boldsymbol{Y'}$ are sampled the following coupled sampling process, which maintains the marginal distributions of $\boldsymbol{Y}$ and $\boldsymbol{Y'}$ as compared to the Bernoulli processes described above, and therefore does not affect any expected value of interest. Fix a sequence of

random values $\boldsymbol{X} = (X_1, \ldots, X_k)$, each drawn independently from the uniform distribution on $[0, 1]$. Then, Y_j and Y'_j depend on the X_j as follows:

$$Y_i(\boldsymbol{X}) = \begin{cases} 1 \text{ if } X_i \le \rho_i \\ 0 \text{ else} \end{cases} \qquad Y'_i(\boldsymbol{X}) = \begin{cases} 1 \text{ if } X_i \le \tilde{\rho}_i \\ 0 \text{ else} \end{cases} \qquad Y_j(\boldsymbol{X}), Y'_j(\boldsymbol{X}) = \begin{cases} 1 \text{ if } X_j \le \tilde{\rho}_j \text{ and } j < i \\ 1 \text{ if } X_j \le \rho_j \text{ and } j > i \\ 0 \text{ else} \end{cases}.$$

We write $\boldsymbol{Y}(\boldsymbol{X}) = (Y_j(\boldsymbol{X}) | j \in [k])$ and $\boldsymbol{Y'}(\boldsymbol{X}) = (Y'_j(\boldsymbol{X}) | j \in [k])$. Note that the distribution of $\boldsymbol{Y}(\boldsymbol{X})$ (resp. $\boldsymbol{Y'}(\boldsymbol{X})$) per this sampling process is identical to the distribution of $\boldsymbol{Y}$ (resp. $\boldsymbol{Y'}$) induced by sampling each entry of $\boldsymbol{Y}$ (resp. $\boldsymbol{Y'}$) according to independent Bernoulli random draws: the marginals are the same, and each entry of $\boldsymbol{Y}$ (resp. $\boldsymbol{Y'}$) remains independent.

Let $D(\boldsymbol{Y}(\boldsymbol{X})) = \{j \in [k] | Y_j = 1\}$ be the dropout set specified by $\boldsymbol{Y}$ and define $D(\boldsymbol{Y'}(\boldsymbol{X}))$ equivalently. Define $R(\boldsymbol{Y}(\boldsymbol{X})) := \arg\min_{R \subseteq A} dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R)$ to be the optimal replacement set for $D(\boldsymbol{Y}(\boldsymbol{X}))$, and define $R(\boldsymbol{Y'}(\boldsymbol{X}))$ analogously.

The goal of the coupling argument is to ensure that $D(\boldsymbol{Y}(\boldsymbol{X}))$ and $D(\boldsymbol{Y'}(\boldsymbol{X}))$ are similar. In fact:

Fact 1: When $X_i \le \min(\rho_i, \tilde{\rho}_i)$ or $X_i > \max(\rho_i, \tilde{\rho}_i)$, $D(\boldsymbol{Y}(\boldsymbol{X})) = D(\boldsymbol{Y'}(\boldsymbol{X}))$ and consequently $R(\boldsymbol{Y}(\boldsymbol{X})) = R(\boldsymbol{Y'}(\boldsymbol{X}))$.

Fact 2: When $X_i \in (\min(\rho_i, \tilde{\rho}_i), \max(\rho_i, \tilde{\rho}_i)]$, $D(\boldsymbol{Y}(\boldsymbol{X})) \triangle D(\boldsymbol{Y'}(\boldsymbol{X})) = \{i\}$ ($\triangle$ is the symmetric difference).

We can express our quantity of interest in terms of these quantities and then apply the law of total expectation, linearity of expectation, and Jensen's inequality to get that

$$\begin{aligned}
&|\mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^{i-1}}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}^{i}})| \\
&\quad = \left|\mathbb{E}_{\boldsymbol{Y}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y}) \cup R(\boldsymbol{Y}))\right] - \mathbb{E}_{\boldsymbol{Y'}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}) \cup R(\boldsymbol{Y'}))\right]\right| \\
&\quad = \left|\mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_{\boldsymbol{Y}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y}) \cup R(\boldsymbol{Y}))\middle|\boldsymbol{X}\right]\right] - \mathbb{E}_{\boldsymbol{X}}\left[\mathbb{E}_{\boldsymbol{Y'}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}) \cup R(\boldsymbol{Y'}))\middle|\boldsymbol{X}\right]\right]\right| \\
&\quad = \left|\mathbb{E}_{\boldsymbol{X}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R(\boldsymbol{Y}(\boldsymbol{X})))\right] - \mathbb{E}_{\boldsymbol{X}}\left[dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R(\boldsymbol{Y'}(\boldsymbol{X})))\right]\right| \\
&\quad \le \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R(\boldsymbol{Y}(\boldsymbol{X}))) - dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R(\boldsymbol{Y'}(\boldsymbol{X})))|\right].
\end{aligned}$$

To upper bound this expression, we select the same replacement set for both $D(\boldsymbol{Y}(\boldsymbol{X}))$ and $D(\boldsymbol{Y'}(\boldsymbol{X}))$ for a fixed realization of $\boldsymbol{X}$. Let this set $R_{\boldsymbol{X}}$ be defined as follows: if the difference inside the absolute value is nonnegative, let $R_{\boldsymbol{X}} = R(\boldsymbol{Y'}(\boldsymbol{X}))$, otherwise $R_{\boldsymbol{X}} = R(\boldsymbol{Y}(\boldsymbol{X}))$. For each possible realization of $\boldsymbol{X}$, this fixes the optimal replacement set for the smaller term and substitutes it in for the optimal replacement set for the larger term, thereby upper-bounding the absolute value of the difference.

$$\le \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}) - dev^{\ell_1}(K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R_{\boldsymbol{X}})|\right]$$

To simplify notation, we will henceforth let $S(\boldsymbol{X}) = K \setminus D(\boldsymbol{Y}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}$ and $S'(\boldsymbol{X}) = K \setminus D(\boldsymbol{Y'}(\boldsymbol{X})) \cup R_{\boldsymbol{X}}$ be the panels after dropout and replacement:

$$= \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(S'(\boldsymbol{X})) - dev^{\ell_1}(S'(\boldsymbol{X}))\right]$$

By *Fact 1*, when $X_i \le \min(\rho_i, \tilde{\rho}_i)$ or $X_i > \max(\rho_i, \tilde{\rho}_i)$, the expression in the expectation evaluates to 0. Thus, it can be simplified to only consider the case when $X_i \in (\min(\rho_i, \tilde{\rho}_i), \max(\rho_i, \tilde{\rho}_i)]$, which (due to the distribution of X_i) occurs with probability $\max(\rho_i, \tilde{\rho}_i) - \min(\rho_i, \tilde{\rho}_i) \le \gamma$:

$$\le \mathbb{E}_{\boldsymbol{X}}\left[|dev^{\ell_1}(S'(\boldsymbol{X})) - dev^{\ell_1}(S'(\boldsymbol{X}))| \middle| X_i \in (\min(\rho_i, \tilde{\rho}_i), \max(\rho_i, \tilde{\rho}_i))\right] \cdot \gamma$$

Now expanding by f, v and applying the triangle inequality,

$$\leq \mathbb{E}_{\boldsymbol{X}}\left[\sum_{f\in F}\sum_{v\in V_f}\left|dev^{\ell_1}_{f,v}(S(\boldsymbol{X})) - dev^{\ell_1}_{f,v}(S'(\boldsymbol{X}))\right| \,\middle|\, X_i \in (\min(\rho_i,\tilde{\rho}_i), \max(\rho_i,\tilde{\rho}_i)]\right]\cdot\gamma$$

By *Fact 2*, when $X_i \in (\min(\rho_i,\tilde{\rho}_i), \max(\rho_i,\tilde{\rho}_i)]$, $D(\boldsymbol{Y}(\boldsymbol{X}))\triangle D(\boldsymbol{Y'}(\boldsymbol{X})) = \{i\}$. It follows that $S(\boldsymbol{X})\triangle S'(\boldsymbol{X}) = \{i\}$. Then, for all feature-values i does not possess (all $f, v : f(i) \neq v$), we have that $|dev^{\ell_1}_{f,v}(S(\boldsymbol{X})) - dev^{\ell_1}_{f,v}(S'(\boldsymbol{X}))| = 0$. Then,

$$\leq \mathbb{E}_{\boldsymbol{X}}\left[\sum_{f\in F}|dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{X})) - dev^{\ell_1}_{f,f(i)}(S'(\boldsymbol{X}))| \,\middle|\, X_i \in (\min(\rho_i,\tilde{\rho}_i), \max(\rho_i,\tilde{\rho}_i)]\right]\cdot\gamma$$

Now, to prove the final step of the deduction, assume without loss of generality that $S(\boldsymbol{X}) = S'(\boldsymbol{X})\dot{\cup}\{i\}$. Then for any $f \in F$, there is some $z_f \in \mathbb{N}$ such that $\sum_{j\in S'(\boldsymbol{X})}\mathbb{I}(f(j) = f(i)) = z_f$ and $\sum_{j\in S(\boldsymbol{X})}\mathbb{I}(f(j) = f(i)) = z_f + 1$. We can then rewrite the feature-level deviations as

$$dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{X})) = \frac{1}{u_{f,f(i)}}\max\left\{0, l_{f,v} - (z_f+1), -u_{f,v} + z_f + 1\right\}$$

and

$$dev^{\ell_1}_{f,f(i)}(S'(\boldsymbol{X})) = \frac{1}{u_{f,f(i)}}\max\left\{0, l_{f,v} - z_f, -u_{f,v} + z_f\right\}.$$

There are three relevant domains for both $z_f, z_f + 1$ to consider: $< l_{f,v}$, $[l_{f,v}, u_{f,v}]$, and $> u_{f,v}$. Note that because $z_f, z_f + 1, u_{f,v}$, and $l_{f,v}$ are all integers, it cannot be that $z_f < l_{f,v} \leq u_{f,v} < z_f + 1$. Thus, z_f and $z_f + 1$ must either be in the same domain or in neighboring domains. The proof proceeds by showing that, regardless of which of the five possible cases of which domains z_f and $z_f + 1$ fall into, $|dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{X})) - dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{'X}))| \leq 1/u_{f,f(i)}$. Because the arguments all use exactly the same approach, we show only the first such case, where $z_f, z_f + 1 < l_{f,v}$. *Case 1:* If $z_f, z_f + 1 < l_{f,v}$, then the dominating term of both $\max\left\{0, l_{f,v} - (z_f+1), -u_{f,v} + z_f + 1\right\}$ and $\max\left\{0, l_{f,v} - z_f, -u_{f,v} + z_f\right\}$ are the second one; Then we get

$$|dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{X})) - dev^{\ell_1}_{f,f(i)}(S(\boldsymbol{'X}))| = \left|1/u_{f,f(i)}((l_{f,v} - z_f) - l_{f,v} - z_f - 1))\right| = \left|1/u_{f,f(i)}\right|.$$

Carrying out cases 2-5 similarly, we conclude that the final step in the overall deduction:

$$\leq \gamma\sum_{f\in F}\frac{1}{u_{f,f(i)}}$$

Putting it all together gives

$$|\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^0}) - \mathcal{L}^{\ell_1}(A;\mathcal{D}_{\boldsymbol{\rho}^k})|$$
$$\leq \sum_{i=1}^{k}\gamma\sum_{f\in F}\frac{1}{u_{f,f(i)}} = \gamma\sum_{f\in F}\sum_{v\in V_f}\frac{\sum_{i=1}^{k}\mathbb{I}(f(i)=v)}{u_{f,v}} \leq \gamma\sum_{f\in F}\sum_{v\in V_f}\frac{u_{f,v}}{u_{f,v}} = \gamma|FV|.$$

Now that we have proven Equation (5), we apply it to complete the proof. Take $s \geq s(\varepsilon, \delta)$ where $s(\varepsilon, \delta) \in \Theta\left((a \log n + \log(1/\varepsilon) + \log(1/\delta)) \cdot 1/\varepsilon^2\right)$. Then:

$$\begin{aligned}
\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) &\leq \mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) + \gamma|FV| && \text{[eq. (5), } \boldsymbol{\rho} \to \tilde{\boldsymbol{\rho}}\text{]} \\
&\leq \mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) + \varepsilon + \gamma|FV| && \text{[w.p. } 1-\delta \text{ via Corollary 3.7]} \\
&\leq \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) + \varepsilon + \gamma|FV| && \text{[optimality of } \widetilde{\text{OPT-}A^{\ell_1}} \text{ for } \mathcal{D}_{\tilde{\boldsymbol{\rho}}}\text{]} \\
&\leq \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) + \varepsilon + 2\gamma|FV| && \text{[eq. (5), } \tilde{\boldsymbol{\rho}} \to \boldsymbol{\rho}\text{]} \\
&\leq \mathcal{L}^{\ell_1}(\text{ERM-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) + \varepsilon + 2\gamma|FV| && \text{[optimality of OPT-}A^{\ell_1} \text{ for } \mathcal{D}_{\boldsymbol{\rho}}\text{]}
\end{aligned}$$

Rearranging gives us

$$Pr[\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{ERM-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \leq \varepsilon + 2\gamma|FV|] \geq 1 - \delta. \qquad \square$$

D.3 Proof of Theorem 4.3

Proof. We employ a similar proof strategy as in Theorem 4.1: we construct an instance, $\mathcal{I}$, and a pair of dropout probability vectors, $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, such that the *optimal* alternate set for the misestimated dropout probability vector performs substantially worse than the *optimal* alternate set for the true dropout probability vector when evaluated under the true dropout probability vector. We then apply Corollary 3.7 to select an $s(\delta, \gamma)$ such that ERM-Alts$^{\ell_1}$ respectively given $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}$ outputs both the corresponding optimal sets with probability $1 - 2\delta$, thus allowing us to conclude the result.

We construct $\mathcal{I}$ as follows: without loss of generality, assume that all feature-values are numbers, so for all $f \in F$, $V_f = \{1, 2, \ldots, |V_f|\}$. Set our panel size, $k = 2(\max_{f \in F} |V_f| - 1)$. Our panel is comprised of two "subgroups", each of size $\max_{f \in F} |V_f| - 1$. Their features are as follows:

	Group 1	**Group 2**
Index Range	$i \in \{1, \ldots, \max_{f \in F} \lvert V_f \rvert - 1\}$	$i \in \{\max_{f \in F} \lvert V_f \rvert, \ldots, \lvert K \rvert\}$
$\forall f \in F,$	$f(i) = \begin{cases} i, & \text{if } i < \lvert V_f \rvert \\ \lvert V_f \rvert, & \text{otherwise} \end{cases}$	$f(i) = \lvert V_f \rvert$
$\rho_i \ \forall i \in K$	γ	0
$\tilde{\rho}_i \ \forall i \in K$	0	γ

Essentially, all values except for the last value for each feature are only represented by one person on the panel. The last value for each feature, f, is a "filler" value that everyone else has, preserving the rarity of the first $|V_f| - 1$ values. We have tight quotas, so for all $f \in F$ and all $v \in V_f$ such that $v \neq |V_f|$, $l_{f,v} = u_{f,v} = 1$. For all $f \in F$, $l_{f,|V_f|} = u_{f,|V_f|} = k - (|V_f| - 1)$. Our pool is a copy of the panel, so for every $i \in K$ there is an $i' \in N$ with matching feature-values. Our alternate budget, a, is $k/2$.

First, we lower bound $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}})$. Observe that OPT-$A^{\ell_1}$ is the set of all Group 1 clones in the pool, and $\widetilde{\text{OPT-}A^{\ell_1}}$ is the set of all Group 2 clones in the pool. This is because, similar to in the proof of Theorem 4.1, under both $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, there are exactly $k/2$ agents with a non-zero probability of dropping out, and these sets contain clones of exactly those agents and no others. As a result, note that $\mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) = 0$ and $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A\ell_1}; \mathcal{D}_{\tilde{\boldsymbol{\rho}}}) = 0$, so:

$$\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) = \mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) = \mathbb{E}_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}\left[\min_{R \subseteq \widetilde{\text{OPT-}A^{\ell_1}}} dev^{\ell_1}(K \setminus D \cup R)\right]$$

For every dropout set D in the support of $\mathcal{D}_{\boldsymbol{\rho}}$, denote the linear deviation minimizing replacement subset of A as $R(D)$. We first lower bound our deviation by only accounting for the error on only some of the feature values.

$$\mathbb{E}_{D\sim\mathcal{D}_{\boldsymbol{\rho}}}\left[dev^{\ell_1}(K\setminus D\cup R(D))\right] \geq \mathbb{E}_{D\sim\mathcal{D}_{\boldsymbol{\rho}}}\left[\sum_{\substack{(f,v)\in FV\\ v<|V_f|}} dev^{\ell_1}_{f,v}(K\setminus D\cup R(D))\right]$$

Noting that all of these feature-values have a quota of 1 and applying linearity, we have:

$$\geq \mathbb{E}_{D\sim\mathcal{D}_{\boldsymbol{\rho}}}\left[\sum_{\substack{(f,v)\in FV\\ v<|V_f|}} \frac{1}{1}\cdot\left(1-\sum_{i\in K\setminus D\cup R(D)}\mathbb{I}(f(i)=v)\right)\right]$$

$$= \sum_{\substack{(f,v)\in FV\\ v<|V_f|}}\left(1-\mathbb{E}_{D\sim\mathcal{D}_{\boldsymbol{\rho}}}\left[\sum_{i\in K\setminus D\cup R(D)}\mathbb{I}(f(i)=v)\right]\right)$$

Finally, recall that for any $(f,v)\in FV$ such that $v<|V_f|$, only one agent on the panel (agent v) has that value. Additionally, as $\widetilde{\text{OPT-}A^{\ell_1}}$ is the set of all Group 2 clones, it has *no* agents with any of these values, and the same must be true for any replacement set taken from it. Therefore, $\sum_{i\in K\setminus D\cup R(D)}\mathbb{I}(f(i)=v)$ is 1 if agent v does not drop, and 0 otherwise.

$$= \sum_{\substack{(f,v)\in FV\\ v<|V_f|}}(1-\Pr[\text{agent } v \text{ does not drop}])$$

$$= \sum_{\substack{(f,v)\in FV\\ v<|V_f|}}(1-(1-\gamma)) = \gamma(|FV|-|F|).$$

Hence, we have our desired bound on the difference between $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}};\mathcal{D}_{\boldsymbol{\rho}})$ and $\mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1};\mathcal{D}_{\boldsymbol{\rho}})$. Now, we show that there exists $s(\delta,\gamma)=\Theta\left(\frac{k/2\cdot\log k+\log(k/\gamma^k)+\log(1/\delta)}{\gamma^{2k}/k^2}\right)$, such that for all $s\geq s(\delta,\gamma)$

$$\Pr[\text{ERM-}A^{\ell_1}=\text{OPT-}A^{\ell_1}\wedge\widetilde{\text{ERM-}A^{\ell_1}}=\widetilde{\text{OPT-}A^{\ell_1}}]\geq 1-2\delta,$$

First we consider the selection of $\widetilde{\text{ERM-}A^{\ell_1}}$. Fix any $A\subseteq N$ that is not $\widetilde{\text{OPT-}A^{\ell_1}}$ —so it is missing at least one Group 2 clone. In the event that our dropout set is all panelists in Group 2, A will incur some non-zero linear loss. In particular, for any feature f with maximal $|V_f|$, the dropout set will have $k/2$ agents with value $|V_f|$. However, *no* Group 1 member, and thus no Group 1 clone in the pool, has value $|V_f|$ for that feature. As A has strictly fewer than $a=k/2$ Group 2 clones, any replacement subset taken from it will fall short of the quota for $f,|V_f|$ by at least one person, incurring a loss of at least $\frac{1}{u_{f,|V_f|}}=\frac{1}{k/2}$. Recall that $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}};\mathcal{D}_{\tilde{\boldsymbol{\rho}}})=0$, so we have that:

$$|\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}};\mathcal{D}_{\tilde{\boldsymbol{\rho}}})-\mathcal{L}^{\ell_1}(A;\mathcal{D}_{\tilde{\boldsymbol{\rho}}})| \geq \min_{R\subseteq A} dev^{\ell_1}(K\setminus\{\text{Group 2}\}\cup R)\cdot\Pr_{D\sim\mathcal{D}_{\tilde{\boldsymbol{\rho}}}}[D=\{\text{Group 2}\}]$$

$$\geq \frac{1}{k/2}\cdot\gamma^{k/2} > \frac{\gamma^k}{k}$$

By our choice of s and Corollary 3.7, we know that

$$\Pr[|\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\tilde{\rho}}) - \mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\tilde{\rho}})| \leq \gamma^k/k] \geq 1 - \delta$$

Hence, this implies that $\widetilde{\text{ERM-}A^{\ell_1}} = \widetilde{\text{OPT-}A^{\ell_1}}$ with probability at least $1 - \delta$.

Next, consider the selection of ERM-A^{ℓ_1}. Fix any $A \subseteq N$ that is not equal to OPT-A^{ℓ_1} . If our dropout set is all panelists in Group 1, then there is at least one person, i, in the dropout set whose clone, i', is not in A. This person has at least one feature, f, for which $f(i) < |V_f|$ (as is the case for all panelists in Group 1 by its definition). Moreover, i and i' are the only agents in the instance with that feature-value. Therefore, A must incur a loss of 1 from that feature-value, as the quota is exactly 1. Recall that $\mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) = 0$, so we have that:

$$\begin{aligned}|\mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(A; \mathcal{D}_{\boldsymbol{\rho}})| &\geq \min_{R \subseteq A} dev^{\ell_1}(K \setminus \{\text{Group 1}\} \cup R) \cdot \Pr_{D \sim \mathcal{D}_{\boldsymbol{\rho}}}[D = \{\text{Group 1}\}] \\ &\geq 1 \cdot \gamma^{k/2} > \frac{\gamma^k}{k}\end{aligned}$$

As before, by our choice of s and Corollary 3.7, we know that

$$\Pr[|\mathcal{L}^{\ell_1}(\text{ERM-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\tilde{\rho}})| \leq \gamma^k/k] \geq 1 - \delta$$

Hence, this implies that ERM-A^{ℓ_1} = OPT-A^{ℓ_1} with probability at least $1 - \delta$. By union bounding, we conclude that indeed $\Pr[\text{ERM-}A^{\ell_1} = \text{OPT-}A^{\ell_1} \wedge \widetilde{\text{ERM-}A^{\ell_1}} - \widetilde{\text{OPT-}A^{\ell_1}}] \geq 1 - 2\delta$. Hence we can combine this with our lower bound on $\mathcal{L}^{\ell_1}(\widetilde{\text{OPT-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{OPT-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}})$ to conclude that

$$\Pr[\mathcal{L}^{\ell_1}(\widetilde{\text{ERM-}A^{\ell_1}}; \mathcal{D}_{\boldsymbol{\rho}}) - \mathcal{L}^{\ell_1}(\text{ERM-}A^{\ell_1}; \mathcal{D}_{\boldsymbol{\rho}}) \geq \gamma(|FV| - |F|)] \geq 1 - 2\delta.$$

□

E Supplemental Materials from Section 5

E.1 Data cleaning details

E.1.1 Data cleaning details for datasets US-1, US-2, and US-3. After trimming features that were not common across these datasets, these datasets contain the following features with the following values (up to renaming for consistency across datasets).

f	V_f
gender	*male*
	female
	non-binary/other
age	*16-24*
	25-34
	35-44
	45-54
	55-64
	65-74
	75+
housing	*own*
	rent
	subsidized housing/unhoused
educational attainment	*bachelor's or higher*
	some college
	high school
	some schooling
race/ethnicity	*native american*
	white
	AAPI
	black
	latinx
	multiracial

Table 1. Feature-values common across *US* instances.

Handling feature-values for Figure 1 / learning the dropout probabilities. The features above were all directly available in the raw data. The only modification to the data we made for these purposes was to merge some of the $f = age$ categories. For Figure 1, this was for consistency across *US* and *Canada* datasets; for training, this was to ensure sufficient sample size and try to capture practically meaningful life stages. These merges combined *16-24* and *25-34* into "Young Adult", *35-44*, *45-54*, and *55-64* into "Middle Age", and *65-74* and *75+* into "Retirement Age".

Handling feature-values for running/evaluating the algorithms. The quotas we imposed exactly reflected the quotas used in practice in each instance. In particular, these quotas were imposed on the raw-feature values above, and sometimes other feature-values that appeared in only a subset of these datasets (which were omitted above because they were not common across datasets).

Handling the "dropped" variable. In this dataset, the main (but still minor) complication was that in *US-2* and *US-3*, there was fairly substantial "dropout" due to people not responding or declining

immediately when notified they were selected for the panel. There were concerns that this may have been due to logistical snags in the recruitment process in these particular cases, meaning that someone "dropping out" at this stage might not be reflective of general "propensity to drop out". Furthermore, there were people recruited to replace these people immediately after the panel was selected (called "panel reselection"), and those who were selected in that process were essentially panelists, meaning they had the opportunity to display dropout/nondropout behavior. We therefore constructed our "dropped" variable in the following way: we ignored people who were selected for the panel but then did not accept the initial notification (just assigning them *n/a* for the binary outcome variable describing dropout), and conversely, we considered those who were selected during panel reselection to be within our dropped/did not drop dataset, even though they were not technically selected for the original panel. Note that this construction was used only for estimating the dropout probabilities, as this is the only purpose of the "dropped" variable.

Handling missing data. There was no missing data.

E.1.2 Data cleaning details for datasets Can-1 - Can-4. These datasets required considerably more unification across the values of each feature. After trimming features that were not common across these datasets and unifying variables, the *Can* datasets contain the following common feature-values.

f	V_f
gender	*male*
	female
	non-binary/other
age	*1* (roughly up to 29)
	2 (roughly 30 - 44)
	3 (roughly 45 - 64)
	4 (roughly 65 - 70)
	5 (roughly 71+)
housing	*own*
	rent
	subsidized housing/unhoused
indigenous	*yes*
	no
racial minority	*yes*
	no
income	*higher income*
	lower income

Table 2. Feature-values common across *Canada* instances.

Of these variables, three were constructed or modified in some way. First, *age* was available in raw form, but the age ranges used across datasets varied substantially. We unified them to the best of our ability, trying to combine similar ranges. The age ranges given in Table 2 are qualified with "roughly" because, while a majority of datasets' ranges respected those guidelines, sometimes there was some leakage between categories. Second, *racial minority* is a constructed variable, which was set to *yes* if people answered "yes" to any number of other raw features: whether they were a *visible minority*, *black*, *indigenous*, *POC*, or *racialized*. Because different datasets contained different subsets of these raw indicators, we combined them into one global indicator to make features as

common as possible across datasets. Finally, *income* is a constructed variable, which was created by (1) unifying several differently-coded variables asking about income, and (2) inferring income level by how many times per year a person took a flight, where greater numbers of flights were assumed to indicate higher income. This variable was ultimately recoded into just two categories, to avoid over-indexing on the assumptions underlying the construction of this variable.

Handling feature-values for Figure 1 / learning the dropout probabilities. The only further modification to the data we made for these purposes was to again merge some of the $f = age$ categories. For Figure 1, this was for consistency across *US* and *Canada* datasets; for training, this was to ensure sufficient sample size and try to capture practically meaningful life stages. These merges defined *1* as "Young Adult", *2* and *3* as "Middle Age", and *4* and *5* as "Retirement Age"

Handling feature-values for running/evaluating the algorithms. In these datasets, we did not have access to the raw quotas used in practice. We therefore simply imposed tight quotas on the feature-values in Table 2, enforcing that if the original panel contained $k_{f,v}$ agents with f, v, then we set the corresponding quotas such that $l_{f,v} = u_{f,v} = k_{f,v}$. We remark that although we made these quotas perfectly tight, the quota *structure* is much less rich in this dataset, owing to the fact that there are far fewer feature-values. This is one likely explanation for why the trends are less pronounced for our results in this dataset.

Handling the "dropped" variable. This variable was taken off-the-shelf, and required no cleaning.

Handling missing data. There were a few limited cases where pool members were missing a feature-value. In these cases, we simply dropped that person from the data. This occurred in two datasets here, *Can-2* and *Can-3*. In *Can-2*, we dropped 16 pool members out of 144 total. Similarly, in *Can-3*, we dropped 2 pool members out of 212 total.

E.2 Convergence plots across sample sizes

Figure 6 shows the loss of all three ERM algorithms at increasing numbers of samples s. Here, s refers to the number of samples used to *train* (i.e., the s provided to ERM-ALTS^{dev}, determininng how many samples to draw to create $\hat{\mathcal{D}}_{\boldsymbol{\rho}}$); to avoid too much noise generated by *evaluation* (rather than non-convergence), we evaluated the loss on 500 samples for all values of s. These tests were performed in instance *US-3*. Based in these plots, we selected a sample size of $s = 300$.

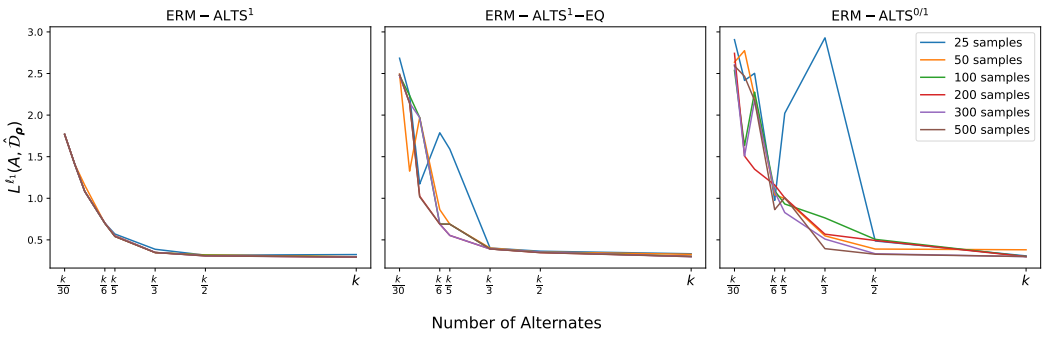

Fig. 6. Plots showing losses of ERM algorithms over varying numbers of training samples (with losses always *evaluated* using 500 samples).

E.3 Benchmark algorithms

The only algorithms that remain to be defined are Greedy and Quota-Based.

First, we define Greedy. Recall that $w(i) = (f(i) | f \in F)$ is agent i's feature vector. We define the hamming distance between two feature vectors w, w' (indexed by f) as the total number of features on which they differ in value:

$$\textit{hamming-dist}(w, w') := \sum_{f \in F} \mathbb{I}(w_f \neq w'_f).$$

ALGORITHM 2: Greedy($\boldsymbol{\rho}, \mathcal{I}$)

$A \leftarrow \emptyset$
$\{1, 2, \ldots, k\} \leftarrow$ Number the panelists in decreasing order of ρ_i, such that $i < j \Rightarrow \rho_i \geq \rho_j$.
for $i \in \{1, 2, \ldots, a\}$ **do**
  $i^* \leftarrow \arg\min_{j \in N \setminus A} \textit{hamming-dist}(w(i), w(j))$
  $A \leftarrow A \cup \{i^*\}$
end
return A

To define Quota-Based, we first define how the quotas are constructed, and then define the ILP used to find a quota-satisfying panel. Fix the original quotas (designed for the size k panel) $\boldsymbol{l}, \boldsymbol{u}$. Now, define our *scaled* quotas as

$$l'_{f,v} := \lfloor l_{f,v} \cdot a/k \rfloor, \qquad u'_{f,v} := \lceil u_{f,v} \cdot a/k \rceil \qquad \text{for all } f, v \in FV.$$

(where if these quotas are infeasible, we loosen all quotas by 1). Then, Quota-Based amounts to solving the following ILP; note that there may be many alternate sets that satisfy these quotas, and we leave this choice to be arbitrarily made by the ILP solver (in practice, a practitioner would use one of the existing panel selection algorithms [Baharav and Flanigan, 2024], whose default functionality makes similarly arbitrary choices when it comes to the precise quota-compliant panel it outputs).

Quota-Based($\mathcal{I}$):

$$\begin{aligned} \max_{x_i | i \in N} \quad & 1 \\ \text{s.t.} \quad & \sum_{i \in N} x_i \cdot \mathbb{I}(f(i) = v) \in [l'_{f,v}, u'_{f,v}] \qquad \text{for all } f, v \in FV \\ & \sum_{i \in N} x_i = a \\ & x_i \in \{0, 1\} \qquad \text{for all } i \in N \end{aligned}$$

E.4 Calibration plots

Here, we attempt to evaluate our empirically-learned dropout probabilities without knowing the ground truth. In each instance j, we compare the *expected* number of dropouts with f, v according to the learned probabilities $\tilde{\boldsymbol{\rho}}_j$, versus the actual number of dropouts from f, v in that instance (examining the *realized* dropout set). Here, $\tilde{\boldsymbol{\rho}}_j$ is exactly as it was defined in the description of Figure 3 (so we train $\tilde{\boldsymbol{\rho}}_j$ on all datasets *other* than *US-j*). We will consider our predictions well-calibrated for the intended dataset if all points on the plot hew to the line $y = x$, describing the case where the expectation of dropouts from f, v perfectly matches the realized number of dropouts.

We see that this is fairly true for *US-2*, but far from true for *US-1* and *US-3*. We remark that *US-3* was a slightly strange case, because there was just one dropout.

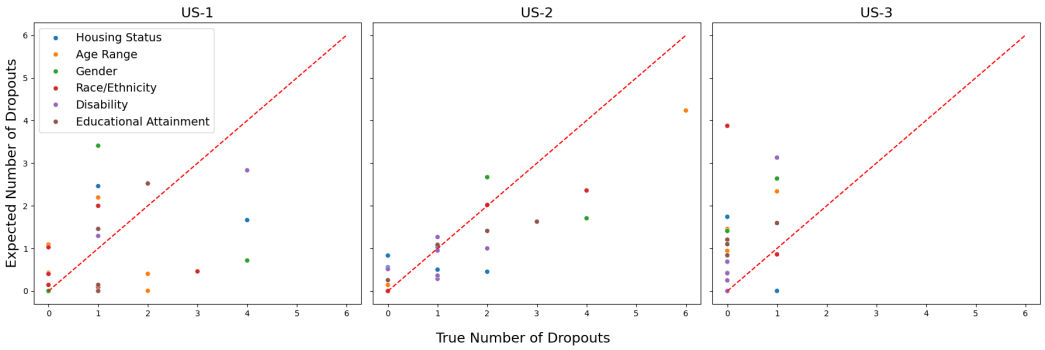

Fig. 7. Points indicate number of dropouts observed in the data in each group f, v (with colors corresponding to values of f). Red line is $y = x$.

E.5 Version of Figure 4 with standard deviations

In Figure 8 we show the version of Figure 4 with standard deviations, as described in the legend. All run parameters are identical to those used to create Figure 4.

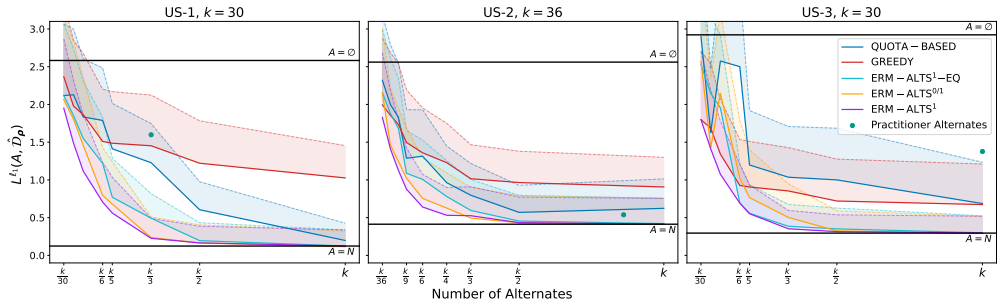

Fig. 8. Analog of Figure 4 showing standard deviation of losses over all samples D drawn during evaluation. Standard deviation shown only above the mean for clarity of plots, but is symmetric above and below the mean.

E.6 Algorithm evaluation on other performance criteria

In Figure 9 we replicate the analysis in Figure 4 on four other metrics of performance, which we define below. As in the body, we compute these expected values over an empirical version of $\mathcal{D}_{\boldsymbol{\rho}}$, called $\hat{\mathcal{D}}_{\boldsymbol{\rho}}$. For each draw of D, S in the definitions below is equal to $K \setminus D \cup R$, where R is chosen according to Equation (2) *with the deviation function specified not as* dev^{ℓ_1}*, but as the current performance benchmark.*

Normalized Deviation Below Quotas:

$$\mathbb{E}_{D\sim\hat{\mathcal{D}}_{\boldsymbol{\rho}}}\left[\sum_{f,v\in FV}\frac{\max\left\{0,\ l_{f,v}-\sum_{i\in S}\mathbb{I}(f(i)=v)\right\}}{u_{f,v}}\right].$$

Max Normalized Quota Deviation:

$$\mathbb{E}_{D\sim\hat{\mathcal{D}}_{\boldsymbol{\rho}}}\left[\max_{f,v\in FV}\frac{\max\left\{0,\ l_{f,v}-\sum_{i\in S}\mathbb{I}(f(i)=v),\ -u_{f,v}+\sum_{i\in S}\mathbb{I}(f(i)=v)\right\}}{u_{f,v}}\right].$$

Max Quota Deviation:

$$\mathbb{E}_{D\sim\hat{\mathcal{D}}_{\boldsymbol{\rho}}}\left[\max_{f,v\in FV}\max\left\{0,\ l_{f,v}-\sum_{i\in S}\mathbb{I}(f(i)=v),\ -u_{f,v}+\sum_{i\in S}\mathbb{I}(f(i)=v)\right\}\right]$$

Number of Unrepresented Feature-Values:

$$\mathbb{E}_{D\sim\hat{\mathcal{D}}_{\boldsymbol{\rho}}}\left[\sum_{f,v\in FV}\mathbb{I}\left(\left(\sum_{i\in S}f(i)=v\right)=0\right)\right]$$

E.7 Results on Canadian data cluster

In Figure 10 and Figure 11 we repeat our two main analyses from (respectively) Figure 3 and Figure 4 in datasets *Can-1 - Can-4*. We see similar relative performance of the algorithms, with the algorithms' performance on the realized dropout sets being somewhat unpredictable, and their relative performance in expectation being fairly consistent. Again, Quota-Based and Greedy tend to have the highest loss, though there are more exceptions in this dataset than the *US* datasets.

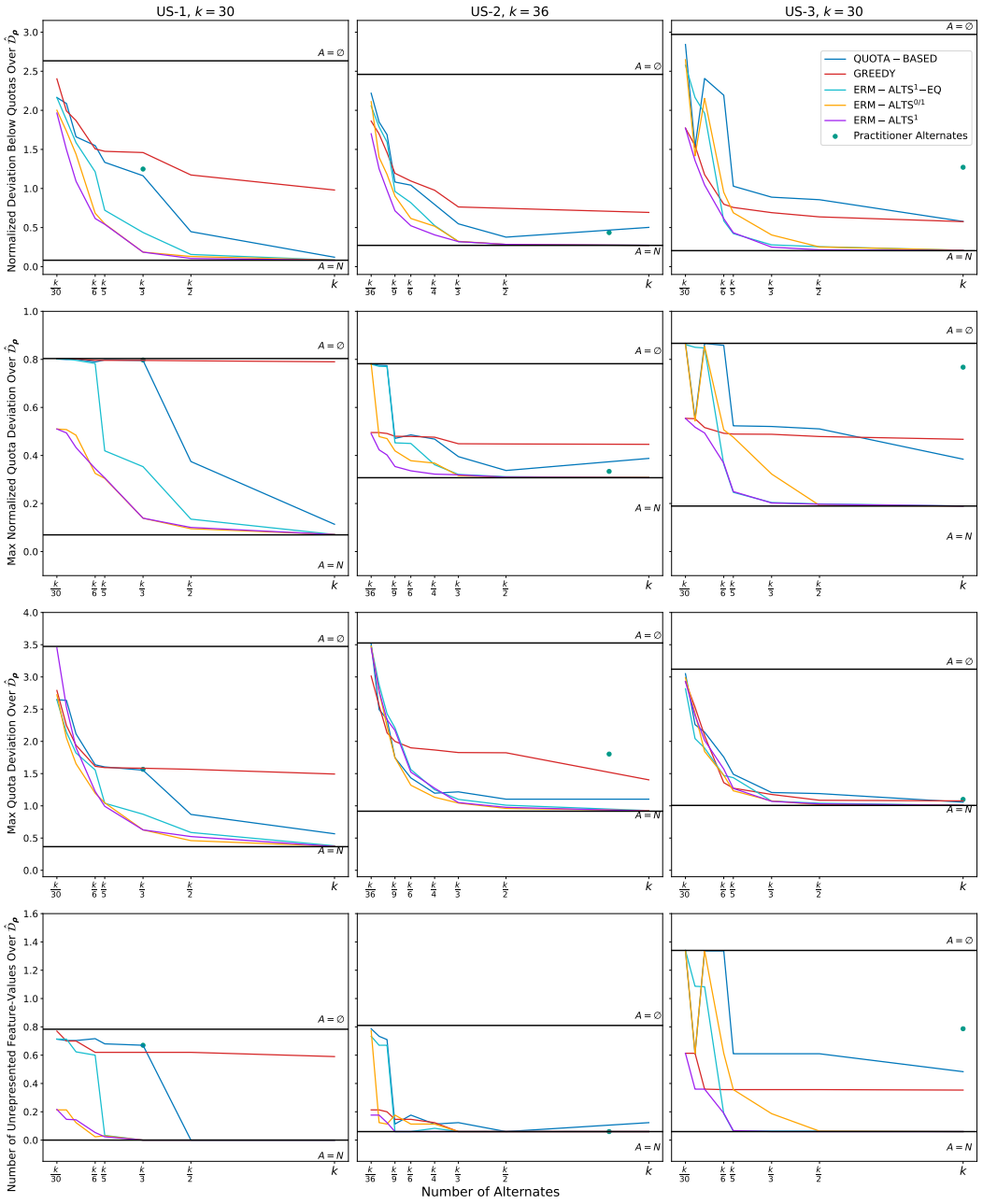

Fig. 9. Analysis of expected algorithmic performance across instances *US-1, US-2, US-3* on other performance metrics, as defined above. All run parameters are the same as in Figure 4.

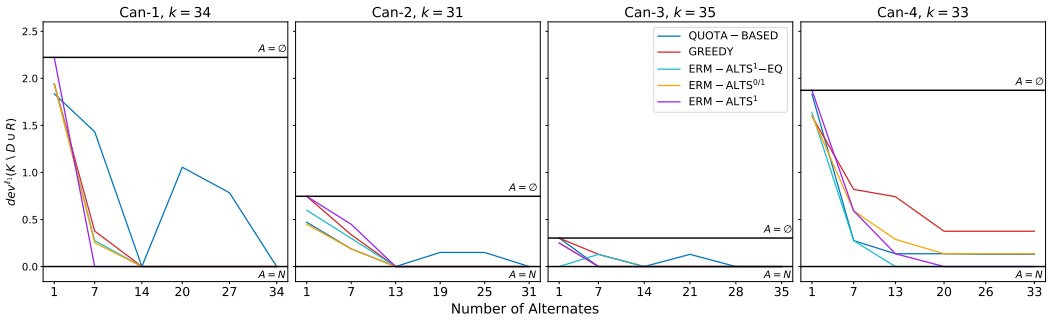

Fig. 10. Analog of Figure 3 in Datasets *Can-1-Can-4*.

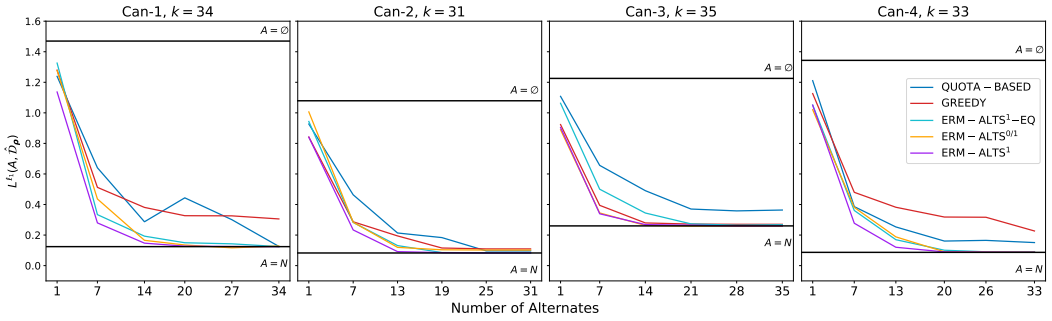

Fig. 11. Analog of Figure 4 in Datasets *Can-1-Can-4*.

F Supplemental Materials from Section 6

F.1 Extension 1: Allowing *alternates* to drop out

In the body we assumed that while panelists may drop out, the *alternates* will always be available to replace them. A reasonable extension would be to assume that alternates will also drop out with some probability — for simplicitly, we will suppose that they do so via the same dropout probabilities as do the panelists, but in principle our algorithm can handle separate distributions. In this case, one would choose the alternates via our algorithm as follows: first, sample dropouts from the panel K *and the remaining pool* N. Then, in every sample we get two dropout sets: the dropouts from the panel D and the dropouts from the pool $\bar{D}$. Now, instead of evaluating A on D, we evaluate $A \setminus \bar{D}$ on D. This weakly worsens the deviation for any given draw of dropouts, but it does not affect the sample complexity required for ERM-Alts, because the size of the hypothesis class has not changed, so running ERM-Alts with order $a \log n$ is still sufficient. The main change to the algorithm occurs in how we implement the ILP Opt, because we are now optimizing the expectation over dropouts from the panel *and* from each candidate alternate set. This new version of Opt is specified below as Opt-Alts-Drop, which takes in a distribution over pool dropout sets ($\bar{\mathcal{D}}$) in addition to the instance and the distribution over panel dropout sets:

$\textsc{Opt-Alts-Drop}^{dev}(\mathcal{I}, \mathcal{D}, \bar{\mathcal{D}})$

$$\min \sum_{(D,\bar{D})\in 2^K\times 2^N} d_{(D,\bar{D})}\cdot \mathcal{D}(D)\cdot \bar{\mathcal{D}}(\bar{D})$$

$$\text{s.t.} \sum_{i\in N} x_i = a$$

$$y_{i,(D,\bar{D})} \le x_i \qquad \forall\, i\in N,\ (D,\bar{D})\in 2^K\times 2^N$$

$$y_{i,(D,\bar{D})} \le \mathbb{I}(i\in\bar{D}) \qquad (*\text{ prevents using alternate dropouts }*) \qquad \forall\, i\in N, (D,\bar{D})\in 2^K\times 2^N$$

$$\sum_{i\in N} y_{i,(D,\bar{D})} \le |D| \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N$$

$$l_{f,v} - \left(\sum_{i\in K\setminus D}\mathbb{I}(f(i)=v) + \sum_{i\in N} y_{i,(D,\bar{D})}\,\mathbb{I}(f(i)=v)\right) \le z_{(D,\bar{D}),f,v}\,u_{f,v} \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N, f,v\in FV$$

$$-u_{f,v} + \left(\sum_{i\in K\setminus D}\mathbb{I}(f(i)=v) + \sum_{i\in N} y_{i,(D,\bar{D})}\,\mathbb{I}(f(i)=v)\right) \le z_{(D,\bar{D}),f,v}\,u_{f,v} \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N, f,v\in FV$$

$$\sum_{f,v} z_{(D,\bar{D}),f,v} \le d_{(D,\bar{D})} \qquad (*\, dev^{\ell_1}\,*) \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N$$

$$\sum_{f,v} z_{(D,\bar{D}),f,v} \le d_{(D,\bar{D})}\cdot(|K|+a)|FV| \qquad (*\, dev^{0/1}\,*) \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N$$

$$d_{(D,\bar{D})} \in \mathbb{R}_{\ge 0} \qquad (*\, dev^{\ell_1}\,*) \qquad \forall\, D\in (D,\bar{D})\in 2^K\times 2^N$$

$$d_{(D,\bar{D})} \in \{0,1\} \qquad (*\, dev^{0/1}\,*) \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N$$

$$x_i\in\{0,1\},\ y_{i,(D,\bar{D})}\in\{0,1\}, \qquad \forall\, i\in N, (D,\bar{D})\in 2^K\times 2^N$$

$$z_{(D,\bar{D}),f,v}\in\mathbb{R}_{\ge 0} \qquad \forall\, (D,\bar{D})\in 2^K\times 2^N, f,v\in FV$$

Given the practical importance of this extension, we replicate our empirical analysis from Figure 4 for this algorithm. The results, shown in Figure 12, show the same trends as previous results, with the three ERM algorithms consistently outperforming the two heuristic algorithms (Quota-Based, Greedy).

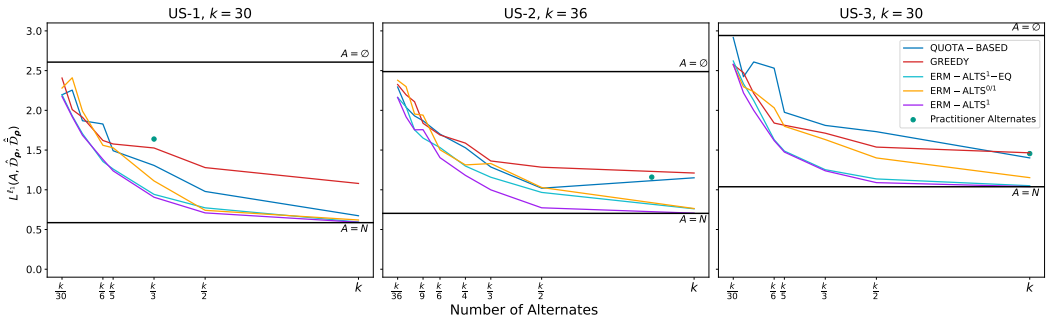

Fig. 12. Replication of Figure 4 when alternates also drop out according to $\boldsymbol{\rho}$. All run parameters are the same as in Figure 4.

F.2 Extension 2: Directly adding extra panelists

Suppose instead of selecting alternates from whom we can choose later, we might just want to select "extra" people outright who will be deterministically added to the panel. The only difference in this setup is that, because these people will be deterministically added to the panel, we may want to make sure that they do not violate any upper quotas too much. Therefore, we also take in a set of additional upper quotas, $\boldsymbol{u}^*$ to impose on the selection of extra panelists. This again does not affect the sample complexity of order $a \log n$, and amounts to a simple change in the ILP Opt to enforce extra quotas on the alternate set it considers. This new version of Opt is specified here as Opt-Preempt:

$$
\begin{aligned}
&\textsc{Opt-Preempt}^{dev}(\mathcal{I}, \mathcal{D}, \boldsymbol{u}^*)\\
&\min \sum_{D \in 2^K} d_D \cdot \mathcal{D}(D)\\
&\text{s.t.} \sum_{i \in N} x_i \le a\\
&\sum_{i \in N} x_i \mathbb{I}(f(i) = v) \le u^*_{f,v} && \forall\, D \in 2^K, f, v \in FV\\
&l_{f,v} - \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in N} x_i\, \mathbb{I}(f(i) = v) \right) \le z_{D,f,v}\, u_{f,v} && \forall\, D \in 2^K, f, v \in FV\\
&-u_{f,v} + \left(\sum_{i \in K \setminus D} \mathbb{I}(f(i) = v) + \sum_{i \in N} x_i\, \mathbb{I}(f(i) = v) \right) \le z_{D,f,v}\, u_{f,v} && \forall\, D \in 2^K, f, v \in FV\\
&\sum_{f,v} z_{D,f,v} \le d_D \qquad (*\, dev^{\ell_1}\, *) && \forall\, D \in 2^K\\
&\sum_{f,v} z_{D,f,v} \le d_D \cdot (|K| + a)|FV| \qquad (*\, dev^{0/1}\, *) && \forall\, D \in 2^K\\
&d_D \in \mathbb{R}_{\ge 0} \qquad (*\, dev^{\ell_1}\, *) && \forall\, D \in 2^K\\
&d_D \in \{0, 1\} \qquad (*\, dev^{0/1}\, *) && \forall\, D \in 2^K\\
&x_i \in \{0, 1\},\ z_{D,f,v} \in \mathbb{R}_{\ge 0} && \forall\, i \in N, D \in 2^K, f, v \in FV
\end{aligned}
$$

F.3 Extension 3: Selecting a maximally robust *panel*

In the body and in Extension 2, we assumed that the panel was given to us by some external selection algorithm and we could only choose alternates or extra panelists after the fact. Now, we ask: what if we could choose the panel *alongside* the alternates? The next two extensions explore successively more general versions of this question.

We now suppose we want to avoid selecting alternates altogether and just try to choose a panel that is within some *initial quotas* $\boldsymbol{l}'$, $\boldsymbol{u}'$, k', so that it has minimal expected loss with respect to some true quotas $\boldsymbol{l}$, $\boldsymbol{u}$, k after dropouts.[4] Just for the remainder of Appendix F, we will refer to the pool N as the *entire pool*, before the panel is selected. What we are now calling the pool would actually be $N \cup K$ in the notation of Section 2. To select a maximally robust panel using our ERM

[4]Note that if $\boldsymbol{l}' = \boldsymbol{l}$, $\boldsymbol{u}' = \boldsymbol{u}$, and $k' = k$, then this is just simply selecting the most robust panel in the actual bounds. If $\hat{k} > k, \hat{u}_{f,v} \ge u_{f,v}$, and $\hat{\ell}_{f,v} = \ell_{f,v}$, then we are in the analog of the "adding extra panelists" scenario where now the extra people are now chosen in conjunction with the panel itself.

approach, one would first sample dropout sets directly from N to get an empirical distribution of dropouts. We will refer to dropout sets from this pool as D', where $D' \in 2^N$. For every dropout set D' and every possible panel K, we are now evaluating the deviation as $dev(K \setminus D', \boldsymbol{l}, \boldsymbol{u})$. As such, we must solve a slightly modified version of the Opt ILP, which we specify below, that takes in a distribution over dropout sets from 2^N. Additionally, given that our hypothesis class is now of size at most $\binom{n}{k}$, applying the same techniques as in Theorem 3.4, we deduce that it is sufficient to run ERM-Alt with a slightly larger sample complexity of order $k \log n$.

Opt-Panel-Select$^{dev}(\mathcal{I}, \mathcal{D}')$

$$
\begin{aligned}
\min \quad & \sum_{D' \in 2^N} d_{D'} \cdot \mathcal{D}'(D') \\
\text{s.t.} \quad & \sum_{i \in N} x_i = |K| \\
& l_{f,v} - \left(\sum_{i \in N \setminus D'} x_i \, \mathbb{I}(f(i) = v) \right) \leq z_{D',f,v} \, u_{f,v} && \forall \, D' \in 2^N, f, v \in FV \\
& - u_{f,v} + \left(\sum_{i \in N \setminus D'} x_i \, \mathbb{I}(f(i) = v) \right) \leq z_{D',f,v} \, u_{f,v} && \forall \, D' \in 2^N, f, v \in FV \\
& \sum_{f,v} z_{D',f,v} \leq d_{D'} \qquad (* \, dev^{\ell_1} \, *) && \forall \, D' \in 2^N \\
& \sum_{f,v} z_{D',f,v} \leq d_{D'} \cdot |K| \cdot |FV| \qquad (* \, dev^{0/1} \, *) && \forall \, D' \in 2^N \\
& d_{D'} \in \mathbb{R}_{\geq 0} \qquad (* \, dev^{\ell_1} \, *) && \forall \, D' \in 2^N \\
& d_{D'} \in \{0, 1\} \qquad (* \, dev^{0/1} \, *) && \forall \, D' \in 2^N \\
& x_i \in \{0, 1\}, \; z_{D',f,v} \in \mathbb{R}_{\geq 0} && \forall \, i \in N, D' \in 2^N, f, v \in FV
\end{aligned}
$$

F.4 Extension 4: Selecting a robust panel and alternates in conjunction

To select alternates and a panel in conjunction, we would again sample the entire pool to get a distribution over dropout sets D'. We would then solve a more complex version of the Opt ILP to find a panel K and corresponding alternate set A. Here, the performance is evaluated as throughout the paper, where some dropouts are drawn from K (which we have chosen) and we replace them as well as possible with the elements of A. We specify the new version of Opt below. Additionally, our hypothesis class is now of size at most $\binom{n}{k+a}$ (the distinction between the chosen K and A applies only in how we compute the deviation), meaning that it is sufficient to run ERM-Alt with sample complexity order $(k + a) \log n$.

OPT-PANEL-AND-ALT-SELECT$^{dev}(\mathcal{I}, \mathcal{D}')$

$$\min \sum_{D' \in 2^N} d_{D'} \cdot \mathcal{D}'(D')$$

$$\text{s.t.} \sum_{i \in N} w_i = |K| \qquad (* \text{ panelists } *)$$

$$\sum_{i \in N} x_i = a$$

$$x_i \leq 1 - w_i \qquad (* \text{ alternates aren't panelists } *) \qquad \forall i \in N$$

$$y_{i,D'} \leq x_i \qquad \forall\, i \in N,\ D' \in 2^N$$

$$y_{i,D'} \leq \mathbb{I}(i \in D') \qquad (* \text{ not using dropped replacers } *) \qquad \forall\, i \in N,\ D' \in 2^N$$

$$\sum_{i \in N} y_{i,D'} \leq \sum_{i \in D} w_i \qquad (* \text{ limit replacement set size } *) \qquad \forall\, D' \in 2^N$$

$$l_{f,v} - \left(\sum_{i \in N \setminus D'} w_i \mathbb{I}(f(i) = v) + y_{i,D'} \mathbb{I}(f(i) = v) \right) \leq z_{D,f,v}\, u_{f,v} \qquad \forall\, D' \in 2^N, f, v \in FV$$

$$- u_{f,v} + \left(\sum_{i \in N \setminus D'} w_i \mathbb{I}(f(i) = v) + y_{i,D'} \mathbb{I}(f(i) = v) \right) \leq z_{D,f,v}\, u_{f,v} \qquad \forall\, D' \in 2^N, f, v \in FV$$

$$\sum_{f,v} z_{D',f,v} \leq d_{D'} \qquad (* \, dev^{\ell_1} \, *) \qquad \forall\, D' \in 2^N$$

$$\sum_{f,v} z_{D',f,v} \leq d_{D'} \cdot |K| \cdot |FV| \qquad (* \, dev^{0/1} \, *) \qquad \forall\, D' \in 2^N$$

$$d_D \in \mathbb{R}_{\geq 0} \qquad (* \, dev^{\ell_1} \, *) \qquad \forall\, D' \in 2^N$$

$$d_D \in \{0, 1\} \qquad (* \, dev^{0/1} \, *) \qquad \forall\, D' \in 2^N$$

$$w_i \in \{0, 1\}, x_i \in \{0, 1\},\ y_{i,D'} \in \{0, 1\} \qquad \forall\, i \in N, D' \in 2^N$$

$$z_{D',f,v} \in \mathbb{R}_{\geq 0} \qquad \forall\, D' \in 2^N, f, v \in FV$$

F.5 Extension 5: Setting robust *quotas*

It may seem unsavory to select the panel based exclusively on the goal of being robust to dropout, as in extensions 3 and 4. This would forgo the guarantees on panel selection algorithms gained in previous work (e.g., [Flanigan et al., 2021a]), and in a potentially problematic way: it might privilege groups whose members are unlikely to drop out (or even worse, who *seem* unlikely to drop out based on past data). A different way of hedging against dropouts in a one-shot method (i.e., without choosing alternates or extra people after the panel is already chosen) is to try to set the quotas in a robust way to begin with. We now explore the extent to which this seemingly intuitive proposal is well-defined and algorithmically feasible.

First, observe that we cannot evaluate the robustness of a set of quotas directly, because many possible panels can be consistent with a set of quotas, and different panels may be robust to different degrees due to their differing combinatorial structure. In order to evaluate the robustness of a set of quotas, we must first fix a panel selection algorithm, which maps these quotas (along with the pool) to a particular *panel distribution* $\mathcal{P}$ — a distribution over all possible panels that satisfy the quotas. Fixing a selection algorithm, we then describe the extent to which our "robustified quotas" $\hat{\boldsymbol{\ell}}, \hat{\boldsymbol{u}}, \hat{k}$

are robust with respect to our true quotas $\boldsymbol{\ell}, \boldsymbol{u}, k$ as the expected deviation from the true quotas of $K \setminus D$, where $K \sim \mathcal{P}$ (satisfying $\hat{\boldsymbol{\ell}}, \hat{\boldsymbol{u}}, \hat{k}$) and $D \sim \mathcal{D}$. The key difference is that now, we are trying to optimize the expected deviation over the randomness of both the dropout set *and the panel*. Our ERM approach still helps us here, because it allows us to evaluate our choice of quotas for a *given* panel distribution; however, even with cutting edge selection algorithms, the process of computing a desirable $\mathcal{P}$ — as is necessary to evaluate a given choice of quotas — is its own complex multilayer optimization problem [Flanigan et al., 2021a]. This makes it is unclear how to directly optimize the quotas without brute-force exploration of all quota settings — we leave this to future work.