Certified Decisions

Isaiah Andrews and Jiafeng Chen

Abstract. Hypothesis tests and confidence intervals are ubiquitous in empirical research, yet their connection to subsequent decision-making is often unclear. We develop a theory of *certified decisions* that pairs recommended decisions with inferential guarantees. Specifically, we attach P-certificates—upper bounds on loss that hold with probability at least $1 - \alpha$—to recommended actions. We show that such certificates allow "safe," risk-controlling adoption decisions for ambiguity-averse downstream decision-makers. We further prove that it is without loss to limit attention to P-certificates arising as minimax decisions over confidence sets, or what Manski (2021) terms "as-if decisions with a set estimate." A parallel argument applies to E-certified decisions obtained from e-values in settings with unbounded loss.

Date: February 25, 2025. Andrews: Department of Economics, Massachusetts Institute of Technology, iandrews@mit.edu; Chen: Stanford Institute for Economic Policy Research, Stanford University, jiafeng@stanford.edu. We thank Guido Imbens, Lihua Lei, José Montiel Olea, Jonathan Roth, Suproteem Sarkar, Chris Walker, and Davide Viviano for useful discussions. We thank ChatGPT for intermittently expert research assistance.

1. **Introduction**

Inferential tools, such as hypothesis tests and confidence intervals, are ubiquitous in empirical practice. At the same time many empirical papers aim, directly or indirectly, to improve decision-making by potential consumers of research, such as policymakers, firms, or households. As highlighted by Manski (2021), the connection between inference and decision-making is unclear: outside of special cases it is rare that optimal decisions, in the tradition of Wald (1949), depend on the data through the results of conventional inference procedures.[1] There is thus an apparent tension between the goal of informing decision-making and conventional, inference-focused, methods for empirical research.

Despite this tension—or perhaps because of it—some procedures in the literature aim to bridge the gap between inference and decision-making. In particular, when researchers report confidence sets, one natural recipe for decisions is to proceed *as if* the confidence set enumerates all plausible parameter values, and choose actions that minimize the worst-case loss over the confidence set. For $\mathcal{A}$ an action space, $L(a, \theta)$ a loss function, and $\hat{\Theta}(Y) \subset \Theta$ a data-dependent confidence set with $\mathrm{P}\{\theta \in \hat{\Theta}(Y)\} \geq 1 - \alpha$, we may choose actions

$$\delta(Y) \in \arg\min_{a \in \mathcal{A}} \sup_{\theta \in \hat{\Theta}(Y)} L(a, \theta). \tag{1}$$

Manski (2021) terms decision rules of this form "as-if decisions with set estimates." Decision rules of this form are implicitly envisioned by Neyman, who writes (Neyman, 1977, emphasis ours)

> The problem of confidence intervals consists in determining two functions of the observables [, $Y_1(\cdot) \leq Y_2(\cdot)$,] to be used in the following manner: Whenever the observable variables X assume some values [x], we shall calculate the corresponding values of Y_1 and Y_2, say $Y_1(x) < Y_2(x)$, and *then assert (or act on the assumption) that*
>
> $$Y_1(x) \leq \theta \leq Y_2(x).$$

More recently, Ben-Michael *et al.* (2021) and Chernozhukov *et al.* (2025) select policies by minimizing the worst-case loss over a confidence set for a social welfare function.

A number of recent papers have also considered the other direction, first making a statistical decision and then constructing confidence sets for the resulting loss or

[1] Though see Tetenov (2016); Bates *et al.* (2024); Viviano *et al.* (2025) for some interesting exceptions.

welfare. Kitagawa and Tetenov (2018) select a policy by maximizing an empirical welfare function, and in their supplementary materials discuss how one may construct a confidence set for the true welfare via simultaneous inference and projection (see, also, Ponomarev and Semenova, 2024). The recent literature studying inference on winners (Benjamini *et al.*, 2019; Andrews *et al.*, 2024; Zrnic and Fithian, 2024a,b) generalizes this approach, considering selection of the "best" of several options based on noisy estimates of loss or welfare, and then constructing bounds for the true loss associated with estimated "best" choice. For $\hat{L}(a;Y)$ an estimate of the loss associated with option a based on data Y, this corresponds to choosing

$$\delta(Y) = \arg\min_{a \in \mathcal{A}} \hat{L}(a;Y)$$

and then (in the one-sided case) constructing a loss bound $R(Y)$ with

$$P\{L(\delta(Y),\theta) \leq R(Y)\} \geq 1-\alpha. \tag{2}$$

What, if anything, do these exercises—forming decisions based on confidence sets or confidence sets for the loss of recommended decisions—accomplish? This paper provides an answer, considering the value of providing *certificates* for the loss associated with a recommended decision. In particular, we consider a data analyst who provides a decision-maker with a recommended decision and estimated risk bound $(\delta(Y), R(Y))$. If this pair satisfies (2), we say that $R(Y)$ is a P-certificate for $L(\delta(Y),\theta)$, and say that $(\delta(Y), R(Y))$ is a *P-certified decision*. The loss bounds $R(Y)$ derived in the literature studying inference on winners are thus P-certificates by construction. For as-if decision-making, note that if we define

$$R(Y) = \sup_{\theta \in \hat{\Theta}(Y)} L(\delta(Y),\theta) \tag{3}$$

then $\theta \in \hat{\Theta}(Y)$ implies $R(Y) \geq L(\delta(Y),\theta)$, so

$$P\left\{L(\delta(Y),\theta) \leq R(Y)\right\} \geq P(\theta \in \hat{\Theta}) \geq 1-\alpha.$$

Hence, both procedures discussed above imply corresponding P-certificates.

We show that P-certified decisions provide useful guarantees for an ambiguity-averse decision-maker who has limited capacity for data analysis and delegates it to an analyst. Specifically, we consider a decision-maker who needs to decide whether to adopt an action recommended by the data analyst, or to instead implement a safe default action with known (but potentially large) loss. This decision-maker believes that the analyst's estimate $R(Y)$ contains information about the loss of the recommended action, but has limited knowledge of the data-generating process and wishes to protect themselves against a large increase in risk uniformly over a potentially large class of data generating processes.

P-certificates are useful to such a decision-maker. Specifically, for a decision-maker with nonnegative loss bounded by 1 and a default action costing $C \in [0, 1)$, any rule that adopts the recommended action $\delta(Y)$ only when $R(Y) \leq C$ yields expected loss no greater than $C + \alpha(1 - C)$, meaning that the decision-maker faces a risk increase of an most $\alpha(1 - C)$. Moreover, over the class of non-randomized adoption rules, the optimal adoption decision is simply the threshold rule $\mathbb{1}(R(Y) \leq C)$.

This result speaks only to the optimal action for a decision-maker presented with a given P-certified decision. We next study optimal provision of P-certificates, and show that it is closely connected to as-if optimization. We find that the class of as-if decisions à la (1) and (3) is essentially complete under a minimally weak preference ordering for P-certified decisions.[2] That is, any P-certified decision $(\tilde{\delta}, \tilde{R})$ can be weakly improved by *some as-if decision* (δ, R) corresponding to some confidence set, in the sense that $R(\cdot) \leq \tilde{R}(\cdot)$ almost surely and the loss estimate $R(\cdot)$ is thus weakly tighter than the loss estimate $\tilde{R}(\cdot)$. This establishes that all reasonable P-certified decisions can be cast as as-if optimization (1), but it does not distinguish certain confidence sets from others. This result has some bite, and implies for instance that the studentized projection approach discussed in e.g. Andrews *et al.* (2024) is dominated by the approaches of Chernozhukov *et al.* (2025).

With more structure on the decision problem, we further show that some optimality results for inference translate to optimality for P-certificates. In particular, if parameter is scalar and the loss function prefers high parameter values (e.g. implementing treatments with higher effects yields lower loss), then uniformly most accurate confidence lower bounds for θ are optimal so long as we prefer smaller values of $R(Y)$.

Finally, as an extension, we consider certified decisions that correspond to alternative approaches to frequentist inference. A rapidly growing literature proposes *e-values* as alternatives to p-values and confidence sets (see Ramdas and Wang, 2024, for a review). We show that e-variables generate a type of as-if decisions—called E-posterior decisions by Grünwald (2023)—that provide an *E-certificate* of the form

[2]This result echoes—though it is developed independently from—a recent result by Kiyani *et al.* (2025), in the context of conformal prediction.

$\mathbb{E}[L(\delta,\theta)/R(Y)] \leq 1$. Similarly, we show that the class of such as-if decisions is essentially complete among E-certified decisions: for any E-certified $(\tilde{\delta}, \tilde{R})$, there exists e-variables and corresponding E-posterior decisions $(\delta(\cdot), R(\cdot))$ where $R(\cdot) \leq \tilde{R}(\cdot)$ almost surely. We show that E-certified decisions imply risk bounds for downstream decision-makers even in contexts with unbounded losses.

This paper is most similar to recent work by Kiyani *et al.* (2025), who study decision-making in the context of conformal inference. Similar to our results in Section 3 but preceding this paper, Kiyani *et al.* (2025) show that the problem of minimizing the expected value of an $1-\alpha$ upper bound for loss is equivalent to the problem of searching for a conformal set with coverage $1-\alpha$ and minimizing the expected minimax loss against the conformal prediction set (see (7)). The key to both arguments is the observation that a probabilistic upper bound on loss can be inverted to form a confidence set.

Section 2 sets up a stylized model of an ambiguity-averse decision-maker and discusses guarantees provided by P-certificates. Section 3 discusses optimal P-certificates. Section 4 extends our results to E-certificates.

2. Ambiguity-averse decision-makers and P-certified decisions

Consider a decision-maker facing a loss function $L(a,\theta) \in [0,1]$ for an action $a \in \mathcal{A}$ and an unknown parameter $\theta \in \Theta$. The decision-maker tasks an analyst to analyze the data $Y \in \mathcal{Y}$ and to provide a recommended action $\delta(Y) \in \mathcal{A}$. The data Y is sampled from a distribution P. To model uncertainty, we assume that Nature may choose any (θ, P) pairs from a set $\mathcal{M}$. For instance, if (θ, P) specifies a statistical model indexed by θ, then $\mathcal{M} = \{(\theta, P_\theta) : \theta \in \Theta\}$. More generally, we allow θ may fail to fully describe P, for instance due to nuisance parameters or model incompleteness.

The decision-maker has access to a default action $a_0 \notin \mathcal{A}$, which yields a loss C which is known to the decision-maker and does not depend on θ. To assess whether to adopt the analyst's recommendation instead, the decision-maker also asks for an assessment of the loss $L(\delta(Y),\theta)$. In this section and the next, we assume that this assessment takes the form of a high-probability upper bound on the loss, which we call a $(1-\alpha)$ *P-certificate* $R(Y)$ for $\delta(Y)$.[3]

[3] $\delta(\cdot), R(\cdot)$ are allowed to depend on external randomization devices. For compactness, we subsume such external random variables into Y.

Definition 2.1. For some $\alpha \in (0,1)$, the pair $(\delta(\cdot), R(\cdot)) : \mathcal{Y} \to \mathcal{A} \times [0,1]$ is called a $1-\alpha$ *P-certified decision* if, for every $(\theta, P) \in \mathcal{M}$,

$$P\left(L(\delta(Y), \theta) \leq R(Y)\right) \geq 1-\alpha. \tag{4}$$

When this holds, $R(Y)$ is called a $(1-\alpha)$ *P-certificate* for $\delta(Y)$.

The guarantee (4) is both familiar and convenient, since many familiar statistical procedures yield guarantees of this form. In particular, if the analyst computes a $1-\alpha$ confidence set $\hat{\Theta}(Y)$, as-if minimax optimization over the confidence set provides such a guarantee: For $\delta(Y) \in \arg\min_{a \in \mathcal{A}} \sup_{\theta \in \hat{\Theta}(Y)} L(a, \theta)$ and $R(Y) = \min_{a \in \mathcal{A}} \sup_{\theta \in \hat{\Theta}(Y)} L(a, \theta)$,[4] the guarantee (4) is a consequence of coverage:

$$P\left\{L(\delta(Y), \theta) \leq R(Y)\right\} \geq P\{\theta \in \hat{\Theta}(Y)\} \geq 1-\alpha. \tag{5}$$

We shall see in Section 3 that the converse is also essentially true: Any reasonable (δ, R) pair may be derived from as-if optimization from some confidence set.

For now, we consider a decision-maker's adoption decision given a P-certified decision. We suppose that the decision-maker encounters considerable ambiguity, considering a model $\mathcal{M}_{\text{DM}}$ that is potentially much larger than $\mathcal{M}$. Given a P-certified decision, however, the decision-maker confines attention to those (θ, P) pairs under which (4) holds

$$\mathcal{M}_{\text{DM}}^{(P,\alpha)} = \left\{(\theta, P) \in \mathcal{M}_{\text{DM}} : P\left[L(\delta(Y), \theta) \leq R(Y)\right] \geq 1-\alpha\right\}.$$

Assume that the decision-maker has some prior π over $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$, which could for instance be a prior on $\mathcal{M}_{\text{DM}}$ truncated to $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$.

Suppose that after observing $(\delta(Y), R(Y))$ the decision-maker adopts the recommended action, $Q = 1$, with probability probability $q(a, r) \equiv \mathrm{P}(Q = 1 \mid \delta(Y) = a, R(Y) = r)$ while otherwise, $Q = 0$, they select the safe action a_0. The decision-maker wishes to choose the option rule q from some class $\mathcal{Q}$ to minimize minimize expected loss under her prior, subject to the constraint that the worst-case expected loss is not much worse than C. That is, let $\delta_Q(Y) = Q\delta(Y) + (1-Q)a_0$, where for $\tau \in (0, 1-C]$ the decision-maker chooses q to solve:

$$\min_{q(\cdot,\cdot) \in \mathcal{Q}} \mathbb{E}_{Q \sim q, \pi}\left[L(\delta_Q(Y), \theta)\right] \text{ subject to } \sup_{(\theta, P) \in \mathcal{M}_{\text{DM}}^{(P,\alpha)}} \mathbb{E}_P[L(\delta_Q(Y), \theta)] \leq C + \tau. \tag{6}$$

[4] If the arg min does not exist, then for any $\epsilon > 0$, we may choose $(\delta(Y), R(Y))$ such that

$$\sup_{\theta \in \hat{\Theta}(Y)} L(\delta(Y), \theta) \leq \inf_{a \in \mathcal{A}} \sup_{\theta \in \hat{\Theta}(Y)} L(a, \theta) + \epsilon$$

and $R(Y) = \sup_{\theta \in \hat{\Theta}(Y)} L(\delta(Y), \theta)$.

The preference (6) minimizes Bayes risk under π, subject to the constraint that the worst-case risk is bounded. The preference (6) encodes that the decision-maker is conservative, ambiguity-averse, but persuadable. If the decision-maker were consumed by ambiguity aversion, then she would implement the default action and would have no use for the analyst. In contrast, under (6), she is willing to hazard most an excess risk of τ for plausible rewards when the worst case over $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$ is overly paranoid.

The preference (6) is also reminiscent of the ambiguity-averse preference (DP) considered in Banerjee *et al.* (2020). Banerjee *et al.* (2020) consider a weighted average between a Bayes expected utility and the worst-case expected utility. One may interpret this weighted average as a Lagrange-multiplier form of (6).

P-certified decisions let the decision-maker ensure that the worst-case risk constraint is respected, while still sometimes adopting the recommendation. Suppose the decision-maker never accepts a recommendation when $R(Y) > C$. If they accepts the recommendation with probability at most $u \in [0,1]$, $\sup_{a,r} q(a,r) \le u$, then they incur no more than $u\alpha(1-C)$ in excess risk, regardless of how large $\mathcal{M}_{\text{DM}}$ is.

Proposition 2.2. *For $u \in [0,1]$, any adoption decision $q(a,r) \le u\mathbb{1}(r \le C)$ has maximum risk*

$$\sup_{(\theta,P)\in\mathcal{M}_{\text{DM}}^{(P,\alpha)}} \mathbb{E}_P[L(\delta_Q(Y),\theta)] \le C + u\alpha(1-C).$$

Proposition 2.2 assures the decision-maker that she does not have to tolerate much excess risk if she adopts whenever $R(Y)$ indicates that $\delta(Y)$ is plausibly preferable to the default action.[5] To further control the excess loss, she may either adopt less, in the sense of choosing a smaller u, or ask that the analyst provide recommendations with higher confidence, requesting a larger $1-\alpha$.

Motivated by Proposition 2.2, we may ask for what beliefs π and sets of rules $\mathcal{Q}$ the problem (6) yields the simple the simple threshold decision rule $q(a,r) = u\mathbb{1}(r \le C)$. To this end, we impose a few restrictions.

First, suppose the decision-maker's beliefs π, the expected loss is upper bounded by $R(Y)$: $\mathbb{E}_\pi[L(\delta(Y),\theta) \mid R(Y) = r] \le r$. Moreover, suppose that the decision-maker's model $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$ is sufficiently rich so that any joint distribution over $(L,R) = (L(\delta(Y),\theta), R(Y)) \in [0,1]^2$ consistent with the P-certificate constraint (4) is rationalized by some member of $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$.

[5] Since $\text{P}\{L(\delta_Q(Y),\theta) > R(Y)\} \le \alpha$ and $L(\delta_Q(Y),\theta) \le 1$, we also obtain the following P-specific bound, which is markedly better if $R(Y)$ can be much smaller than C:

$$\mathbb{E}_P\left[L(\delta_Q(Y),\theta)\right] \le \alpha + \mathbb{E}_P[QR(Y) + (1-Q)C].$$

If $Q = \mathbb{1}(R(Y) \le C)$, this bound simplifies to $\alpha + \mathbb{E}_P[\min(R(Y),C)]$.

Second, suppose that the decision-maker only considers simple adoption decisions based on $R(Y)$ and not on $\delta(Y)$: i.e., $q(a, r) = q(r)$. This restricts to only "simple" adoption decisions, for instance because the decision-maker either does not observe $\delta(Y)$ when deciding to adopt or has a difficult time assessing different recommendations $\delta(Y)$ on their own merits.

Third, suppose the decision-maker commits to adopting with at least probability u for some assessment $R(Y) = r$: $\sup_{r\in[0,1]} q(r) \geq u$. For instance, it may be reasonable to impose $q(0) = 1$ so that decisions with the maximally favorable assessment is always adopted. If we further restrict decision rules to be non-randomized, then this condition (with $u = 1$) is equivalent to the decision-maker committing to sometimes accept the analyst's recommendation.

Finally, under the preceding assumptions, one can show (see Lemma A.1 below) that the worst-case excess risk is bounded above by $u\alpha(1 - C)$. Assume the decision-maker is sufficiently conservative that this constraint binds, $\tau = u\alpha(1 - C)$.

Under these conditions, we can show that $u\mathbb{1}(r \leq C)$ is optimal for (6).

Proposition 2.3. *Let $u \in [0, 1]$. Consider*

$$\mathcal{Q}(u) = \left\{ q(a, r) = q(r) : \sup_{r\in[0,1]} q(r) \geq u \right\}.$$

Suppose the decision-maker's beliefs satisfy:

(1) π is optimistic: $\mathbb{E}_\pi[L(\delta(Y), \theta) \mid R(Y) = r] \leq r$

(2) $\mathcal{M}_{\text{DM}}^{(P,\alpha)}, R(\cdot), \delta(\cdot)$ is sufficiently rich: For any joint distribution G over $(L, R) \in [0, 1]^2$ with $G(\{L \leq R\}) \geq 1 - \alpha$, there exists $(\theta, P) \in \mathcal{M}_{\text{DM}}^{(P,\alpha)}$ under which $(L(\delta(Y), \theta), R(Y)) \sim G$.

Then the optimal adoption decision over $\mathcal{Q} = \mathcal{Q}(u)$ for (6) with $\tau = u\alpha(1 - C)$ is $q(r) = u\mathbb{1}(r \leq C)$.

The intuition for Proposition 2.3 is simple. Since the decision-maker commits to accepting with probability at least u, this exhausts the excess risk budget τ. Accepting any recommendation with $R(Y) > C$ strictly increases the worst-case risk over a rich $\mathcal{M}_{\text{DM}}^{(P,\alpha)}$. On the other hand, accepting any $R(Y) = r \leq C$ with probability less than u worsens the objective in (6) without relaxing the constraint. Thus, the optimal acceptance decision is $u\mathbb{1}(r \leq C)$. In particular, if the decision-maker commits to non-randomized decision rules $q(r) \in \{0, 1\}$ and to accept some recommendations ($\sup_r q(r) = 1$), then the optimal acceptance decision is immediately $q(r) = \mathbb{1}(r \leq C)$.

Appendix B furthermore characterizes the form of optimal adoption rules when we do not restrict to $q \in \mathcal{Q}(u)$, though their form depends on the details of the prior.

Together, Proposition 2.2 and Proposition 2.3 show that P-certified decision rules are compatible with simple adoption decisions for an ambiguity-averse decision-maker willing to tolerate some increase in her worst-case risk relative to a safe option. Thus, P-certified decisions offer downstream decision-makers—who may be less familiar with the statistical environment—simple insights from data that guide decisions, without sacrificing (much) safety.

Provision of P-certified decisions requires that the analyst knows the decision-maker's loss function $L(a, \theta)$, but does not require knowledge of the prior π nor the cost of default option C. To further reduce information requirements, via (5) it is even possible for the analyst to simply communicate a confidence set $\hat{\Theta}(Y)$ without knowledge of the loss function, and ask the decision-maker to find $(\delta(Y), R(Y))$ via as-if optimization.

We make two remarks before illustrating with an example. First, we note that the results Proposition 2.2 and Proposition 2.3 apply to *any* P-certified decisions, including the trivial certificate where $R(Y) = 1$ with probability $1 - \alpha$, independent of $\delta(Y)$. Thus, these results do not require that the analyst be particularly skilled in data analysis, nor that they necessarily shares the same objective as the decision-maker, just that the analyst be able to credibly provide P-certificates.

Second, the validity of P-certificates is to some extent verifiable. To prove that $(\delta(\cdot), R(\cdot))$ is a P-certified decision, when the loss function L is money-metric the analyst can in principle offer insurance policies that pay out 1 unit of utility if it turns out that $L(\delta(Y), \theta) > R(Y)$, in exchange for α units of insurance premium. The decision-maker can purchase any amount of such policies before seeing the data. If $(\delta(\cdot), R(\cdot))$ is indeed a P-certificate, offering such insurance has nonnegative expected value for the analyst.[6]

2.1. **Example: inference on winners.** To illustrate our results, we consider an example based on the recent literature on inference on winners (Benjamini *et al.*, 2019; Andrews *et al.*, 2024; Zrnic and Fithian, 2024a,b). Consider a decision-maker interested in potentially implementing an action a from a finite set $\mathcal{A}$. For instance, the

[6]As Neyman (1937) writes, "The theoretical statistician constructing [a confidence interval] may be compared with the organizer of a game of chance in which the gambler has a certain range of possibilities to choose from while, whatever he actually chooses, the probability of his winning and thus the probability of the bank losing has permanently the same value, [α]."

decision-maker could be a municipal policymaker deciding considering whether to implement a program, as studied in Bergman *et al.* (2024), that encourages low-income households with children to move to high-opportunity neighborhoods. In this context, each $a \in \mathcal{A}$ could index a potential set of "recommended" neighborhoods, while the default action a_0 corresponds to the status quo of not making any recommendation.

The decision-maker's goal is to maximize some bounded outcome, for instance the average household income rank in adulthood (relative to the adult income distribution) for children in the targeted households. Specifically, for $\theta(a) \in [0, 1]$ the (unknown) true outcome associated with selection a, the policymaker would like to choose actions a that solve $\max_{a\in\mathcal{A}} \theta(a)$. The policymaker also faces a fixed cost κ of making a recommendation at all, for instance because such recommendations have not been made in the past and this new program needs to be explained to the public. To represent this decision-maker's preferences we can consider the loss function $L(a, \theta) = 1 - \theta(a)$, and take $C = 1 - \theta(a_0) - \kappa$ where for simplicity we treat $\theta(a_0)$ as known.

Suppose the decision-maker is assisted by an analyst, who uses data (e.g. Census-tract level measures for economic opportunity together with a model for long-term outcomes as in Bergman *et al.* (2024)) to form an estimate $X(a)$ for the outcome associated with each possible action a, along with an associated standard error $\sigma(a)$, which we treat as fixed for simplicity. The analyst, but not the decision-maker, knows the joint distribution of estimation error, where for $a \in \mathcal{A}$, $\frac{X(a)-\theta(a)}{\sigma(a)} \sim Z(a)$. In this setting, we can take $Y = (X(\cdot), \sigma(\cdot))$ to collect the estimates and standard errors.

The analyst may select a recommended action via empirical welfare maximization, taking $\delta(Y) = \arg\max_{a\in\mathcal{A}} X(a)$. To form a P-certificate, they could take $R(Y) = 1 - (X(\delta(Y)) - c_{1-\alpha})$, for $c_{1-\alpha}$ the $1-\alpha$ quantile of the maximum error $\sup_{a\in\mathcal{A}} Z(a)\sigma(a)$. This bound, corresponding to an unstudentized version of what Andrews *et al.* (2024) call the projection approach, was suggested in the supplementary materials of Kitagawa and Tetenov (2018). Alternatively, the analyst could consider studentized projection and report $R^*(Y) = 1-(X(\delta(Y))-c^*_{1-\alpha}\sigma(\delta(Y)))$, for $c^*_{1-\alpha}$ the $1-\alpha$ quantile of the maximum studentized error $\sup_{a\in\mathcal{A}} Z(a)$. Studentization ensures that the width of the interval reflects the standard error for the recommended policy $\delta(Y)$, and further ensures that e.g. the inclusion of a single, very nosily estimated option a in the choice set does not greatly influence the critical value. For this reason, the theoretical discussion of projection in Andrews *et al.* (2024) focuses primarily on the studentized approach.

Given either $(\delta(Y), R(Y))$ or $(\delta(Y), R^*(Y))$, the decision-maker must decide whether to implement to action $\delta(Y)$. The preference (6) captures that they decisonmaker seek to maximize the expected outcome under some joint prior π on θ and the distribution of Y. However, since the decision-maker does not know the distribution of Y (beyond their knowledge that the analyst provides a P-certificate), they limit attention to adoption rules q such that risk is not greatly increased relative to the status quo, even in the worst case. Under the conditions of Proposition 2.3, we have shown that it is then optimal for the decision-maker to adopt (perhaps with some probability) when $R(Y) \leq C$. By the duality between tests and confidence sets, however, $\mathbb{1}\{R(Y) \leq C\}$ and $\mathbb{1}\{R^*(Y) \leq C\}$ are size-α tests for the null $L(\delta(Y), \theta) \geq C$, so Proposition 2.3 implies that the decision-maker adopts the recommended action only when the analyst rejects the null that it increases loss relative to the status quo.

3. **Optimal P-certificates**

In settings where P-certificates are useful, it seems reasonable that both analysts and decision-makers will prefer tighter bounds on the loss. That is, given the choice between level-$1-\alpha$ P-certificates $R(Y)$ and $\tilde{R}(Y)$ such that $R(Y) \leq \tilde{R}(Y)$ for all Y, it is natural to prefer R to $\tilde{R}$. For instance, if the decision-maker uses the non-randomized adoption rule $Q = \mathbb{1}(R(Y) \leq C)$ derived in Proposition 2.3, then

$$\mathbb{E}_P[L(\delta_Q, \theta)] \leq \alpha + \min(R(Y), C),$$

and using $R(Y)$ rather than $\tilde{R}(Y)$ yields a tighter risk bound on the loss. More generally, if we consider any loss function $B(R, \theta)$ that is increasing in R for all θ, then $\mathbb{E}_P[B(R(Y), \theta)] \leq \mathbb{E}_P[B(\tilde{R}(Y), \theta)]$ for all (P, θ), so R yields a lower expected loss than $\tilde{R}$.

When R and $\tilde{R}$ are ordered almost surely, or more generally in the sense of first-order stochastic dominance, we say R weakly dominates $\tilde{R}$.

Definition 3.1. We say that a $1-\alpha$ P-certified decision (δ, R) *weakly dominates* another $1-\alpha$ P-certified decision $(\tilde{\delta}, \tilde{R})$ if

$$P(R(Y) \geq r) \leq P(\tilde{R}(Y) \geq r)$$

for all $r \in [0, 1]$ and all $(\theta, P) \in \mathcal{M}$.

Our main result in this section shows that the class of P-certified decisions from as-if optimization against a confidence set is an *essentially complete* class with respect to dominance. That is, any P-certified decision rule $(\tilde{\delta}, \tilde{R})$ is weakly dominated by some decision rule (δ, R) formed by as-if optimization via (5). As an implication, these decision rules are also essentially complete with respect to any risk function of the form $\mathbb{E}_P[B(R, \theta)]$, for any weakly increasing $B(\cdot, \theta)$.

To state this result, denote by $\mathcal{C}_{P,\alpha,\epsilon}$ the class of decision rules formed by (ϵ-approximate) as-if optimization against a confidence set $\hat{\Theta}(Y)$ in the following fashion: For $\epsilon > 0$, consider $(\delta(Y), R(Y))$ such that

$$\sup_{\theta \in \hat{\Theta}(Y)} L(\delta(Y), \theta) \leq \inf_{a \in \mathcal{A}} \sup_{\theta \in \hat{\Theta}(Y)} L(a, \theta) + \epsilon$$
$$R(Y) = \sup_{\theta \in \hat{\Theta}(Y)} L(\delta(Y), \theta)$$
$$P(\theta \in \hat{\Theta}(Y)) \geq 1 - \alpha \text{ for all } (\theta, P) \in \mathcal{M}.$$

where sup over an empty set is defined to be 0. The constant ϵ accounts for cases where the minimax loss over $\hat{\Theta}(Y)$ is not achieved by any element of $\mathcal{A}$; our result holds for any $\epsilon > 0$.

Theorem 3.2. *For any P-certified decision $(\tilde{\delta}, \tilde{R})$ and any $\epsilon > 0$, there exists some $(\delta, R) \in \mathcal{C}_{P,\alpha,\epsilon}$ that weakly dominates $(\tilde{\delta}, \tilde{R})$.*

Theorem 3.2 shows a sense in which it is without loss to limit attention to P-certificates derived from as-if optimization with confidence sets. Thus, not only is as-if optimization a convenient way to generate P-certified decisions, but there is, in a sense, no need to look beyond this recipe.

In the context of conformal prediction, recent work by Kiyani *et al.* (2025) (their Theorem 2.3) shows a result similar to Theorem 3.2.[7] To connect these results, let ϑ be a random prediction target and consider H the joint distribution over (ϑ, X). In a prediction context, ϑ is the unknown true label of a unit and X is the observed features. A conformal prediction set $\hat{\Theta}(X)$ satisfies $H(\vartheta \in \hat{\Theta}(X)) \geq 1 - \alpha$, where the probability is taken over $(\vartheta, X) \sim H$. Kiyani *et al.* (2025) show that the following optimization problems are equivalent, with essentially the same argument as that underlying Theorem 3.2

$$\min_{R(\cdot), \delta(\cdot)} \mathbb{E}_H[R(X)] \text{ subject to } H\left\{L(\delta(X), \vartheta) \leq R(X)\right\} \geq 1 - \alpha$$
$$\min_{\hat{\Theta}(\cdot)} \mathbb{E}_H\left[\min_{a \in \mathcal{A}} \max_{\vartheta \in \hat{\Theta}(X)} L(a, \vartheta)\right] \text{ subject to } H\{\vartheta \in \hat{\Theta}(X)\} \geq 1 - \alpha. \tag{7}$$

[7] We developed Theorem 3.2 independently of their result.

In contrast, we consider a parameter inference setting, where probabilities are solely taken over Y fixing θ, rather than jointly over (X, ϑ). We also consider evaluations of the certificate $R(\cdot)$ beyond its expectation, using a weaker preference ordering (Definition 3.1).

3.1. **Example: inference on winners, continued.** Theorem 3.2 establishes a strong sense in which as-if optimization is reasonable. Dominance ordering is demanding, however, and there are many (R, δ), $(\tilde{R}, \tilde{\delta})$ pairs that cannot be ordered, in the sense that neither $R(Y)$ nor $\tilde{R}(Y)$ (weakly) stochastically dominates the other. Consequently, weak dominance in the sense of Definition 3.1 provides only a partial order over the class of certified decisions, and does not generally provide clear recommendations for "optimal" procedures. This partial order is nevertheless sufficient to rule out some methods discussed in the literature.

To illustrate, let us return to the example, introduced in Section 2.1, of an analyst who selects a recommended action via empirical welfare maximization, $\delta(Y) = \arg\max_{a\in\mathcal{A}} X(a)$, and forms P-certificate $R(Y)$ or $R^*(Y)$ using unstudentized or studentized projection, respectively. The certified decision (δ, R) can be obtained via as-if optimization using the confidence set

$$\hat{\Theta}(Y) = \{\theta : \theta(a) \geq X(a) - c_{1-\alpha} \text{ for all } a \in \mathcal{A}\},$$

and so is easily cast into the essentially complete class constructed in Theorem 3.2.

By contrast, (δ, R^*) is less naturally understood via as-if optimization, since R^* corresponds to the studentized confidence set

$$\hat{\Theta}^*(Y) = \left\{\theta : \theta(a) \geq X(a) - \sigma(a)c^*_{1-\alpha} \text{ for all } a \in \mathcal{A}\right\}.$$

If the analyst conducts as-if optimization over $\hat{\Theta}^*(Y)$, they recommend decision $\tilde{\delta}(Y) = \arg\max_{a\in\mathcal{A}} \left\{X(a) - \sigma(a)c^*_{1-\alpha}\right\}$, which is a special case of the risk-aware empirical maximization approach proposed by Chernozhukov *et al.* (2025) and can differ from $\delta(Y)$ when $\sigma(\cdot)$ is non-constant. The corresponding risk bound $\tilde{R}(Y) = 1 - \max_{a\in\mathcal{A}} \left\{X(a) - \sigma(a)c^*_{1-\alpha}\right\}$ is no larger than $R^*(Y)$ in every realization of Y, so $(\tilde{\delta}, \tilde{R})$ weakly dominates (δ, R^*).[8]

[8]Formally, however, if $\Theta = [0,1]^{\mathcal{A}}$ then (δ, R^*) still belongs to the essentially complete class we study, since it results from as-if optimization against the confidence set $\hat{\Theta}^{**}(Y) = \left\{\theta \in \Theta : L(\delta(Y), \theta) \leq 1 - (X(\delta(Y)) - \sigma(\delta(Y))c^*_{1-\alpha})\right\}$. Note, however, that $\hat{\Theta}^{**}(Y)$ is a super-set of $\hat{\Theta}^*(Y)$, which helps explain the superior performance of $(\tilde{\delta}, \tilde{R})$ relative to (δ, R^*).

3.2. **Optimality under monotonicity.** The non-uniqueness of optimal certified decisions, in the sense of weak dominance, is intuitive: Theorem 3.2 shows that P-certificates are closely linked to confidence sets, and in most problems the class of admissible (i.e. undominated) confidence sets is large. If we narrow attention to settings where optimal (uniformly most accurate) confidence sets exist, one might hope that optimal P-certified decisions will exist as well and are found by as-if optimization. The next theorem shows that hope is borne out, provided the loss is decreasing in the parameter and we limit our comparison of stochastic dominance to risk values exceeding $\inf_{a\in\mathcal{A}} L(a,\theta)$.[9]

To state the result, we recall the definition uniformly most accurate confidence bounds:

Definition 3.3 (Lehmann and Romano (2024), p.79–80)**.** Suppose $\Theta \subset \mathbb{R}$. A random variable $\hat{\theta}(Y)$ is a $1-\alpha$ *confidence lower bound* if $P(\hat{\theta}(Y) \leq \theta) \geq 1-\alpha$ for all $(\theta, P) \in \mathcal{M}$. Furthermore, it is *uniformly most accurate* if for any $1-\alpha$ confidence lower bound $\tilde{\theta}(Y)$, any $t > 0$, and any $(\theta, P) \in \mathcal{M}$,

$$P\{\hat{\theta}(Y) \leq \theta - t\} \leq P\{\tilde{\theta}(Y) \leq \theta - t\}.$$

Theorem 3.4. *Suppose $\Theta \subset \mathbb{R}$ is compact, $L(a,\theta)$ is weakly decreasing in θ, and $R(\theta) = \inf_{a\in\mathcal{A}} L(a,\theta)$ is achieved by some $a \in \mathcal{A}$ for every $\theta \in \Theta$. Let $\hat{\theta}(Y)$ be a $1-\alpha$* uniformly most accurate *confidence lower bound, assumed to exist. Let (δ, R) be a certified decision that as-if optimizes against $[\hat{\theta}(Y), \infty) \cap \Theta$:*

$$R(Y) \equiv L(\delta(Y), \hat{\theta}(Y)) = \inf_{a\in\mathcal{A}} L(a, \hat{\theta}(Y)).$$

Then for any other $1-\alpha$ P-certified $(\tilde{\delta}, \tilde{R})$ and any $r > R(\theta)$,

$$P(R(Y) \geq r) \leq P(\tilde{R}(Y) \geq r)$$

3.3. **Example: choosing a treatment proportion.** Theorem 3.4 does not apply to the inference on winners problem since the parameter of interest θ is vector-valued, and there does not in general exist a uniformly most accurate confidence set. Consider instead a setting where the decision-maker is considering a single treatment, for instance enrollment in a job training program, and must decide what fraction $a \in [\varepsilon, 1]$ of a population randomize into treatment. Below $a = \varepsilon > 0$ they treat

[9]Absent the latter restriction, it is impossible to dominate as-if optimization against the trivial confidence set which takes $\hat{\Theta}(Y)$ empty with probability α, and equal to Θ with probability $1-\alpha$.

no one, $a_0 = 0$. The average outcome under treatment is θ, while the (known) average outcome under control is ρ. There is again a fixed cost κ of treating anyone, and there is also a variable cost $\psi(a)$ of treating fraction $a > 0$ of the population where we assume $\psi(a)$ is continuous and increasing, reflecting e.g. the increasing cost of hiring large numbers of qualified instructors. If the decision-maker would like to maximize the average outcome in the population, net of costs, then we can represent their preferences with loss function $L(a, \theta) = a(1 - \theta) + \psi(a)$ for $a \in \mathcal{A}$, along with loss $C = (1 - \rho) - \kappa$ for the default action $a_0 = 0$. Note that the loss is decreasing in θ for all a by construction, and $\inf_{a \in \mathcal{A}} L(a, \theta)$ is attained for all θ.

Suppose the analyst observes a normally distributed estimate $X \sim N(\theta, \sigma^2)$ for the average outcome under the treatment, where σ is fixed and known to the analyst and $Y = (X, \sigma)$. The uniformly most accurate level 1-α lower confidence bound for θ is $\hat{\theta}(Y) = X + \sigma z_\alpha$, for z_α the α-quantile of a standard normal distribution. By Theorem 3.4, the optimal P-certified decision in this example takes $\delta(Y) = \arg\min_{a \in \mathcal{A}} L(a, \hat{\theta}(Y))$ and $R(Y) = (\delta(Y), \hat{\theta}(Y))$. If the decision-maker then implements the recommended treatment when $R(Y) \leq C$, this corresponds exactly to implementing treatment (for some portion of the population) when the analyst is able to reject that $\theta \leq \overline{\theta}$ for a threshold $\overline{\theta}$ such that $\inf_a R(a, \overline{\theta}) = C$.

4. Extension: E-certificates

We have shown that P-certified decisions offer a natural route to combine (frequentist) statistical inference with decision-making. However, high-probability guarantees, as appearing in confidence intervals and hypothesis tests, are not the only type of frequentist guarantee available. A recent literature in statistics proposes *e-values* as an alternative and a complement to traditional statistical inference (see Ramdas and Wang, 2024, for a review). An *e-variable* against the parameter value θ_0 is a nonnegative random variable $E(Y, \theta_0)$ such that

$$\mathbb{E}_P[E(Y, \theta_0)] \leq 1 \text{ for all } P \text{ such that } (\theta_0, P) \in \mathcal{M}.$$

One can interpret $E(\cdot, \theta_0)$ as the realized dollar payoff of a bet, priced at \$1, against the hypothesis $H_0 : \theta = \theta_0$. Since the expected payoff under H_0 is less than \$1, this bet is (at most) fair. Correspondingly, a large realization of $E(Y, \theta_0)$ can be interpreted as evidence against H_0—in the sense that it is an unlikely event under H_0—much like a small p-value is evidence against H_0.

As with confidence sets, we can use e-variables to certify the quality of decisions. Throughout this section, we assume $L(a, \theta) > 0$, but no longer assume $L(a, \theta) \leq 1$.

Definition 4.1. For fixed $\gamma > 0$, a pair $(\delta(\cdot), R(\cdot))$ is an *E-certified decision* at multiple γ if

$$\mathbb{E}_P\left[\frac{L(\delta(Y), \theta)}{R(Y)}\right] \leq \gamma \text{ for all } (\theta, P) \in \mathcal{M}.$$

We abbreviate E-certified decisions at multiple 1 as E-certified decisions.

As Definition 4.1 highlights, $R(Y)$ is a stochastic upper bound for loss in the sense that the ratio $L(\delta(Y), \theta)/R(Y)$ has ex ante expectation bounded below some prespecified γ.

Similar to our previous results, one can derive E-certified decisions using a version of as-if optimization, now defined using e-variables. Grünwald (2023) proposes what he terms the E-posterior: if there exists an e-variable $E(\cdot, \theta)$ for every $\theta \in \Theta$, the E-posterior is defined as $1/E(Y, \theta)$, with the convention $1/0 = \infty$. Grünwald (2023) proposes computing the action that minimizes the worst-case E-posterior-weighted loss, along with the minimized value. Formally, since minimization over $\mathcal{A}$ might not be attained by any element in $\mathcal{A}$, let $\mathcal{C}_{E,\epsilon}$ be the following class of (δ, R) such that, for some collection of e-variables $E(\cdot, \theta)$,

$$R(Y) \equiv \sup_{\theta \in \Theta} \frac{L(\delta(Y), \theta)}{E(Y, \theta)} \leq \inf_{a \in \mathcal{A}} \sup_{\theta \in \Theta} \frac{L(a, \theta)}{E(Y, \theta)} + \epsilon \tag{8}$$

Intuitively, for choosing $\delta(\cdot)$, states θ deemed implausible by the data—those with large $E(Y, \theta)$—have their losses downweighted, while those with small $E(Y, \theta)$ have their loss upweighted. Grünwald (2023) shows that $(\delta(Y), R(Y))$ is an E-certified decision at multiple 1.[10]

We next show that, as with P-certificates in Theorem 3.2, it is without loss to limit attention to as-if optimization, since the set of E-certified decisions of the form (8) is an essentially complete class with respect to the dominance order.

Proposition 4.2. *For any E-certified decision $(\tilde{\delta}, \tilde{R})$ and any $\epsilon > 0$, there exists some collection of e-variables $E(Y, \theta)$ for which the corresponding E-certified decisions $(\delta, R) \in \mathcal{C}_{E,\epsilon}$ chosen via (8) weakly dominates $(\tilde{\delta}, \tilde{R})$.*

[10]Here, our setup is slightly different since the minimization is up to a constant ϵ. Nevertheless, observe that since $R(Y) \geq L(\delta(Y), \theta)/E(Y, \theta)$,

$$\mathbb{E}_P[L(\delta(Y), \theta)/R(Y)] \leq E_P[E(Y, \theta)] \leq 1.$$

Like our results for P-certificates, using E-certificates also provide guarantees when combined with downstream decision-making. For a decision-maker who never adopts when $R(Y) > C$, the decision-maker at most doubles the cost of the default action C. Unlike our results for P-certificates, such a guarantee does not require that the loss function be bounded above.

Proposition 4.3. *Given an E-certified decision rule (δ, R), for any adoption decision $Q \leq \mathbb{1}(R(Y) \leq C)$ a.s.,*

$$\mathbb{E}_P\left[L(\delta_Q(Y), \theta)\right] \leq \mathbb{E}_P\left[\max\left(\frac{L(\delta(Y), \theta)}{R(Y)}, 1\right) C\right] \leq 2C \text{ for all } (\theta, P) \in \mathcal{M}$$

Motivated by Proposition 4.3, if we only sought certificates with $\mathbb{E}_P\left[\max\left(\frac{L(\delta(Y),\theta)}{R(Y)}, 1\right)\right] \leq 1+\gamma$ to start with, $R(Y)$ can often be improved by slightly modifying the E-posterior. We detail this improvement in Appendix C.

5. Conclusion

This paper considers combining statistical inference with statistical decisions. In two leading modes of frequentist inference, statistical inference translates to *certified decisions*—decisions paired with certificates of their performance. We show that these certified decisions provide ambiguity-averse downstream decision-makers with useful risk guarantees and simple adoption decision rules—which may be implemented and understood without statistical sophistication on part of the decision-maker. Moreover, we show that it is without loss to base certified decisions on statistical inference procedures, and thus that inferential and decision goals are, at the very least, partially aligned.

Appendix A. Proofs

Proposition 2.2. *For $u \in [0, 1]$, any adoption decision $q(a, r) \leq u\mathbb{1}(r \leq C)$ has maximum risk*

$$\sup_{(\theta,P)\in\mathcal{M}_{\text{DM}}^{(P,\alpha)}} \mathbb{E}_P[L(\delta_Q(Y), \theta)] \leq C + u\alpha(1 - C).$$

Proof. Fix $(\theta, P) \in \mathcal{M}_{\text{DM}}^{(P,\alpha)}$, and let A be the event that $L(\delta(Y), \theta) \leq R(Y)$. Then

$$\begin{aligned}
\mathbb{E}_P[L(\delta_Q(Y), \theta) - C] &= \mathbb{E}_P[Q(L(\delta(Y), \theta) - C)] \\
&\leq \mathbb{E}_P\left[Q\mathbb{1}(A)(L(\delta(Y), \theta) - C) + Q\mathbb{1}(A^C)(1 - C)\right] \\
&= \mathbb{E}_P\left[q(\delta(Y), R(Y))\mathbb{1}(A)(L(\delta(Y), \theta) - C) + q(\delta(Y), R(Y))\mathbb{1}(A^C)(1 - C)\right]
\end{aligned}$$

Note that

$$q(\delta(Y), R(Y))\mathbb{1}(A)\left(L(\delta(Y), \theta) - C\right) \leq 0$$

since $q(\delta(Y), R(Y)) = 0$ when $R(Y) > C$, and $R(Y) \leq C$, $\mathbb{1}(A) = 1$ jointly imply $L(\delta(Y), \theta) \leq R(Y) \leq C$. On the other hand,

$$\mathbb{E}\left[q(\delta(Y), R(Y))\mathbb{1}(A^C)(1 - C)\right] \leq u(1 - C)\mathbb{E}[\mathbb{1}(A^C)] \leq u\alpha(1 - C).$$

Hence, $\mathbb{E}_P[L(\delta_Q(Y), \theta) - C] \leq u\alpha(1 - C)$, as desired. □

Proposition 2.3. *Let $u \in [0, 1]$. Consider*

$$\mathcal{Q}(u) = \left\{q(a, r) = q(r) : \sup_{r \in [0,1]} q(r) \geq u\right\}.$$

Suppose the decision-maker's beliefs satisfy:

(1) π is optimistic: $\mathbb{E}_\pi[L(\delta(Y), \theta) \mid R(Y) = r] \leq r$

(2) $\mathcal{M}_{\text{DM}}^{(P,\alpha)}, R(\cdot), \delta(\cdot)$ is sufficiently rich: For any joint distribution G over $(L, R) \in [0, 1]^2$ with $G(\{L \leq R\}) \geq 1 - \alpha$, there exists $(\theta, P) \in \mathcal{M}_{\text{DM}}^{(P,\alpha)}$ under which $(L(\delta(Y), \theta), R(Y)) \sim G$.

Then the optimal adoption decision over $\mathcal{Q} = \mathcal{Q}(u)$ for (6) with $\tau = u\alpha(1 - C)$ is $q(r) = u\mathbb{1}(r \leq C)$.

Proof. By Lemma A.1, the constraint is satisfied only if

$$\sup_{r \geq C} q(r)(r - C) = 0$$

by plugging in $a = \alpha$ and $\sup q(r) \geq u$ to (9). This implies that $q(r) = 0$ for all $r > C$. Similarly, one can show that $\sup_{r \in [0,1]} q(r) = u$ for all q that satisfies the constraint of (6). Thus all q satisfying the constraint obey $q(r) \leq u\mathbb{1}(r \leq C)$. It suffices to show that $u\mathbb{1}(r \leq C)$ weakly dominates all other such $q(r)$ in terms of the objective in (6).

Consider $q(r) \leq u\mathbb{1}(r \leq C)$. Note that

$$\mathbb{E}_\pi[(q(R) - u\mathbb{1}(r \leq C))\left(L(\delta(Y), \theta) - C\right)] \geq 0$$

by the optimism of π. Therefore, $q(r)$ is weakly dominated by $u\mathbb{1}(r \leq C)$. Thus $u\mathbb{1}(r \leq C)$ is optimal. □

Lemma A.1. *Suppose $q(a, r) = q(r)$. Suppose $\mathcal{M}_{\text{DM}}^{(P,\alpha)}, R(\cdot), \delta(\cdot)$ is sufficiently rich: For any joint distribution G over $(L, R) \in [0, 1]^2$ with $G(\{L \leq R\}) \geq 1 - \alpha$, there exists $(\theta, P) \in \mathcal{M}_{\text{DM}}^{(P,\alpha)}$ under which $(L(\delta(Y), \theta), R(Y)) \sim G$.*

Then, for any $u \in [0,1]$,

$$\sup_{(\theta,P)\in\mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}} \mathbb{E}_P\left[L(\delta_Q(Y),\theta)\right] \leq C + u\alpha(1-C)$$

if and only if

$$\sup_{a\leq\alpha}\left\{a\sup_{r\in[0,1]} q(r)(1-C) + (1-a)\sup_{r\geq C} q(r)(r-C)\right\} \leq u\alpha(1-C). \tag{9}$$

Proof. First, consider the only if direction. For contrapositive, suppose there exists $(a, r^-, r^+), a \leq \alpha, r^+ \geq C$ such that

$$aq(r^-)(1-C) + (1-a)q(r^+)(r^+ - C) > u\alpha(1-C).$$

Let G be a distribution over $(L,R) \in [0,1]^2$ be such that $G(L=1, R=r^-) = a$ and $G(L = R = r^+) = 1-a$. By assumption, there is some (θ, P) under which $(L(\delta(Y),\theta), R(Y)) \sim G$. The resulting risk under that distribution is

$$\mathbb{E}_P[q(R)L(\delta(Y),\theta) + (1-q(R))C] = aq(r^-)(1-C) + (1-a)q(r^+)(r^+ - C) > ua(1-C).$$

This proves the only if direction.

Now, for the if direction, suppose q satisfies (9). Fix some $(\theta, P) \in \mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}$. Then for $L = L(\delta(Y),\theta)$, $R = R(Y)$,

$$\begin{aligned}
\mathbb{E}_P\left[L(\delta_Q,\theta) - C\right] &= \mathbb{E}_P\left[Q(L-C)\right] \\
&\leq \mathbb{E}_P\left[Q\mathbb{1}(L>C)(L-C)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}(R<L)\mathbb{1}(C<R)Q(L-C)\right] + \mathbb{E}_P\left[\mathbb{1}(C<L\leq R)Q(L-C)\right] \\
&\leq P(L>R)\mathbb{E}_P\left[Q(L-C) \mid L>R\right] \\
&\quad + (1-P(L>R))\mathbb{E}_P\left[Q(R-C) \mid C<L\leq R\right]
\end{aligned}$$

Observe that, by conditioning on R,

$$\mathbb{E}_P\left[Q(L-C) \mid L>R\right] \leq \sup_{r\in[0,1]} q(r)(1-C)$$

$$\mathbb{E}_P\left[Q(R-C) \mid C<L\leq R\right] \leq \sup_{r\geq C} q(r)(r-C)$$

Since $P(L>R) \leq \alpha$, we have that

$$\mathbb{E}_P\left[L(\delta_Q,\theta) - C\right] \leq u\alpha(1-C).$$

by assumption. ∎

Lemma A.2. *Under the assumptions of Lemma A.1, suppose additionally that*

$$\sup_{r\in[0,1]} q(r) = 1.$$

Then

$$\sup_{(\theta,P)\in\mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}} \mathbb{E}_P\left[L(\delta_Q(Y),\theta)\right] \leq C + \alpha(1-C)$$

if and only if

$$q(r) \leq \mathbb{1}(r \leq C).$$

Proof. By Lemma A.1, $\sup_{(\theta,P)\in\mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}} \mathbb{E}_P\left[L(\delta_Q(Y),\theta)\right] \leq C + \alpha(1-C)$ is equivalent to

$$\sup_{a\leq\alpha}\left\{a(1-C) + (1-a)\sup_{r\geq C} q(r)(r-C)\right\} \leq \alpha(1-C).$$

This condition is implied by $q(r) \leq \mathbb{1}(r \leq C)$ by inspection. On the other hand, taking $a = \alpha$ yields

$$\sup_{r\geq C} q(r)(r-C) \leq 0$$

This implies that $q(r)(r-C) = 0$ for all $r > C$, and hence $q(r) = 0$ for all $r \geq C$. Therefore $q(r) \leq \mathbb{1}(r \leq C)$. □

Theorem 3.2. *For any P-certified decision $(\tilde{\delta}, \tilde{R})$ and any $\epsilon > 0$, there exists some $(\delta, R) \in \mathcal{C}_{P,\alpha,\epsilon}$ that weakly dominates $(\tilde{\delta}, \tilde{R})$.*

Proof. Consider

$$\hat{\Theta}(Y) = \left\{\theta \in \Theta : L(\tilde{\delta}(Y),\theta) \leq \tilde{R}(Y)\right\}.$$

Since $(\tilde{\delta}, \tilde{R})$ is P-certified, $\hat{\Theta}$ is a confidence set:

$$P(\theta \in \hat{\Theta}(Y)) = P(L(\tilde{\delta}(Y),\theta) \leq \tilde{R}(Y)) \geq 1-\alpha$$

for all $(\theta, P) \in \mathcal{M}$.

Consider $(\delta, R) \in \mathcal{C}_{P,\alpha,\epsilon}$ such that

$$R(Y) \equiv \sup_{\theta\in\hat{\Theta}(Y)} L(\delta(Y),\theta) \leq \left(\inf_{a\in\mathcal{A}} \sup_{\theta\in\hat{\Theta}(Y)} L(a,\theta) + \epsilon\right) \wedge \sup_{\theta\in\hat{\Theta}(Y)} L(\tilde{\delta}(Y),\theta).$$

Such a choice of $\delta(\cdot)$ exists since one can choose $\delta(Y) = \tilde{\delta}(Y)$ if

$$\sup_{\theta\in\hat{\Theta}(Y)} L(\tilde{\delta}(Y),\theta) \leq \inf_{a\in\mathcal{A}} \sup_{\theta\in\hat{\Theta}(Y)} L(a,\theta) + \epsilon.$$

Otherwise, one can choose some action with worst-case risk bounded by $\inf_{a\in\mathcal{A}}\sup_{\theta\in\hat{\Theta}(Y)} L(a,\theta)+\epsilon$ by definition of the infimum. By construction,

$$R(Y) \leq \sup_{\theta\in\hat{\Theta}(Y)} L(\tilde{\delta}(Y),\theta) \leq \tilde{R}(Y).$$

Since $R(Y)$ and $\tilde{R}(Y)$ are ordered almost surely, they are also ordered in the sense of stochastic dominance. This concludes the proof. □

Theorem 3.4. *Suppose $\Theta \subset \mathbb{R}$ is compact, $L(a,\theta)$ is weakly decreasing in θ, and $R(\theta) = \inf_{a\in\mathcal{A}} L(a,\theta)$ is achieved by some $a \in \mathcal{A}$ for every $\theta \in \Theta$. Let $\hat{\theta}(Y)$ be a $1-\alpha$* uniformly most accurate *confidence lower bound, assumed to exist. Let (δ, R) be a certified decision that as-if optimizes against $[\hat{\theta}(Y),\infty) \cap \Theta$:*

$$R(Y) \equiv L(\delta(Y),\hat{\theta}(Y)) = \inf_{a\in\mathcal{A}} L(a,\hat{\theta}(Y)).$$

Then for any other $1-\alpha$ P-certified $(\tilde{\delta},\tilde{R})$ and any $r > R(\theta)$,

$$P(R(Y) \geq r) \leq P(\tilde{R}(Y) \geq r)$$

Proof. Consider

$$\tilde{\theta}(Y) = \inf\left\{\theta \in \Theta : L(\tilde{\delta}(Y),\theta) \leq \tilde{R}(Y)\right\}$$

Since Θ is compact $\tilde{\theta}(Y) \in \Theta$ a.s. Note that $P(\tilde{\theta}(Y) \leq \theta) \geq P\{L(\tilde{\delta}(Y),\theta) \leq \tilde{R}(Y)\} \geq 1-\alpha$, and thus $\tilde{\theta}(Y)$ is a $1-\alpha$ confidence lower bound for θ.

Again let $R(\theta) = \inf_{a\in\mathcal{A}} L(a,\theta)$, and note that $R(Y) = R(\hat{\theta}(Y))$ while $R(\tilde{\theta}(Y)) \leq L(\tilde{\delta}(Y),\tilde{\theta}(Y)) = \tilde{R}(Y)$. Decreasingness of $L(a,\theta)$ in θ implies that $R(\theta)$ is decreasing in θ.

Since $\hat{\theta}$ is a uniformly most accurate confidence lower bound, $P\left\{\hat{\theta}(Y) \leq \theta - t\right\} \leq P\left\{\tilde{\theta}(Y) \leq \theta - t\right\}$ for all $t > 0$. It follows that $P\left\{\hat{\theta}(Y) \wedge \theta \leq t\right\} \leq P\left\{\tilde{\theta}(Y) \wedge \theta \leq t\right\}$ for all $t \in \Theta$, and thus that $\hat{\theta}(Y) \wedge \theta$ first order stochastically dominates $\tilde{\theta}(Y) \wedge \theta$. Since first order stochastic dominance is preserved by monotone transformations, it follows that $R(\tilde{\theta}(Y) \wedge \theta) = R(\tilde{\theta}(Y)) \vee R(\theta)$ first order stochastically dominates $R(\hat{\theta}(Y) \wedge \theta) = R(\hat{\theta}(Y)) \vee R(\theta)$, and thus that for all $r > R(\theta)$,

$$\mathrm{P}\{R(Y) \geq r\} = \mathrm{P}\{R(\hat{\theta}(Y)) \geq r\} \leq \mathrm{P}\{R(\tilde{\theta}(Y)) \geq r\} \leq \mathrm{P}\{\tilde{R}(Y) \geq r\}.$$

This completes the proof. □

Proposition 4.2. *For any E-certified decision* $(\tilde{\delta}, \tilde{R})$ *and any* $\epsilon > 0$*, there exists some collection of e-variables* $E(Y, \theta)$ *for which the corresponding E-certified decisions* $(\delta, R) \in \mathcal{C}_{E,\epsilon}$ *chosen via* (8) *weakly dominates* $(\tilde{\delta}, \tilde{R})$*.*

Proof. Define

$$E(Y, \theta) = \frac{L(\tilde{\delta}(Y), \theta)}{\tilde{R}(Y)}.$$

Because $(\tilde{\delta}, \tilde{R})$ is E-certified, $\mathbb{E}_P[E(Y, \theta)] \leq 1$ for all $(\theta, P) \in \mathcal{M}$. Observe that for all θ,

$$\frac{L(\tilde{\delta}(Y), \theta)}{E(Y, \theta)} = \tilde{R}(Y) = \sup_{\theta \in \Theta} \frac{L(\tilde{\delta}(Y), \theta)}{E(Y, \theta)}$$

Pick $(\delta, R) \in \mathcal{C}_{E,\epsilon}$ such that

$$R(Y) \equiv \sup_{\theta \in \Theta} \frac{L(\delta(Y), \theta)}{E(Y, \theta)} \leq \min \left(\inf_{a \in \mathcal{A}} \sup_{\theta \in \Theta} \frac{L(a, \theta)}{E(Y, \theta)} + \epsilon, \tilde{R}(Y) \right)$$

which exists since $\tilde{R}(Y)$ is the worst-case E-posterior loss of some action, namely $\tilde{\delta}(Y)$, and thus

$$\tilde{R}(Y) \geq \inf_{a \in \mathcal{A}} \sup_{\theta \in \Theta} \frac{L(a, \theta)}{E(Y, \theta)}.$$

By definition, $R(Y) \leq \tilde{R}(Y)$. This completes the proof. □

Proposition 4.3. *Given an E-certified decision rule* (δ, R)*, for any adoption decision* $Q \leq \mathbb{1}(R(Y) \leq C)$ *a.s.,*

$$\mathbb{E}_P \left[L(\delta_Q(Y), \theta) \right] \leq \mathbb{E}_P \left[\max \left(\frac{L(\delta(Y), \theta)}{R(Y)}, 1 \right) C \right] \leq 2C \text{ for all } (\theta, P) \in \mathcal{M}$$

Proof. Note that, P-almost surely,

$$L(\delta_Q(Y), \theta) = \frac{L(\delta_Q(Y), \theta)}{QR(Y) + (1 - Q)C} \left(QR(Y) + (1 - Q)C \right) \leq \max \left(\frac{L(\delta(Y), \theta)}{R(Y)}, 1 \right) C.$$

where the equality holds since $QR(Y) + (1 - Q)C > 0$ a.s. by Lemma A.3. The inequality follows from $Q \leq \mathbb{1}(R(Y) \leq C)$ and thus $QR + (1 - Q)C \leq C$.

Now,

$$\mathbb{E}_P \max \left(\frac{L(\delta(Y), \theta)}{R(Y)}, 1 \right) \leq 1 + \mathbb{E}_P[L/R] \leq 2.$$

This completes the proof. □

Lemma A.3. *Let* $(\delta(\cdot), R(\cdot))$ *be an E-certified decision. For all* $(\theta, P) \in \mathcal{M}$*,*

$$P(R(Y) > 0) = 1.$$

Proof. By Proposition 4.2, it suffices to consider $(\delta, R) \in \mathcal{C}_{E,\epsilon}$ for some $\epsilon > 0$, since otherwise one can find $\tilde{R}$ that lower bounds $R(Y)$. For such a pair,

$$R(Y) = \sup_{\theta' \in \Theta} L(\delta(Y), \theta')/E(Y, \theta') \geq L(\delta(Y), \theta)/E(Y, \theta) \geq 0.$$

Note that $R(Y) = 0$ implies that $E(Y, \theta) = \infty$, since $L(a, \theta) > 0$. However, since $\mathbb{E}_P[E(Y, \theta)] \leq 1$,

$$P(E(Y, \theta) = \infty) = 0 \geq P(R(Y) = 0) = 0.$$

□

Appendix B. Optimal adoption based on P-certificates

Proposition 2.3 derives optimal adoption rules over a constrained class of decision rules $\mathcal{Q}(u)$. This section extends these results, deriving optimal adoption rules over the class of all adoption rules that depend only on $R(Y)$.

Proposition B.1. *Let $u \in [0, 1]$. Consider*

$$\mathcal{Q}^* = \bigcup_{u \in [0,1]} \mathcal{Q}(u) = \{q(a, r) = q(r)\}.$$

Suppose the decision-maker's beliefs satisfy:

(1) π is optimistic: $\mathbb{E}_\pi[L(\delta(Y), \theta) \mid R(Y) = r] \leq r$

(2) $\mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}, R(\cdot), \delta(\cdot)$ is sufficiently rich: for any joint distribution G over $(L, R) \in [0, 1]^2$ with $G(\{L \leq R\}) \geq 1 - \alpha$, there exists $(\theta, P) \in \mathcal{M}_{\mathrm{DM}}^{(P,\alpha)}$ under which $(L(\delta(Y), \theta), R(Y)) \sim G$.

Then an optimal adoption decision over $\mathcal{Q} = \mathcal{Q}^*$ for (6) is

$$q_{q^*}(r) = (u - q^*)\mathbb{1}\{r \leq C\} + \mathbb{1}\{\mathbb{E}_\pi[L|R = r] \leq C < r\} \min\left\{\frac{\alpha(1 - C)}{(1 - \alpha)(r - C)} q^*, u - q^*\right\}$$

where $q^* \in [0, u]$ solves $\min_{q^* \in [0,u]} \mathbb{E}_{Q \sim q_{q^*}, \pi}\left[L(\delta_Q(Y), \theta)\right]$.

Proof. By Lemma A.1, the constraint in (6) holds if and only if (9) does. In turn, (9) holds if and only if

$$q(r)(1 - C) \leq \frac{\alpha}{a} u(1 - C) - \frac{1 - a}{a} \sup_r q(r)(r - C)$$

for all (a, r). If $\sup_r q(r)(r - C)/(1 - C) > u\alpha$ this constraint necessarily fails, while if $\sup_r q(r)(r - C)/(1 - C) \leq u\alpha$ the right hand side is minimized at $a = \alpha$, so the constraint is equivalent to

$$\sup_r q(r) \leq u - \frac{1-\alpha}{\alpha} \sup_r \frac{q(r)(r-C)}{1-C}. \tag{10}$$

By the law of iterated expectations,

$$\mathbb{E}_{\pi,q}[QL + (1-Q)C] - C = \int q(r)(\mathbb{E}_\pi[L|R(Y)=r] - C)d\pi_R(r), \tag{11}$$

for $\pi_R(r)$ the marginal distribution of $R(Y)$ under π. The decision-maker wants to choose q to maximize this expression subject to the constraint (10). Since (11) is pointwise increasing in $q(r)$ for all r where $\mathbb{E}_\pi[L \mid R(Y)=r] \leq C$, it is without loss to consider rules which exhaust the constraint on this set,

$$q(r) = (u-q^*)\mathbb{1}\{r \leq C\} + \mathbb{1}\{\mathbb{E}_\pi[L|R=r] \leq C < r\} \min\left\{\frac{\alpha(1-C)}{(1-\alpha)(r-C)}q^*, u-q^*\right\}$$

for a constant $q^* \in [0,u]$. We are minimizing a continuous function over a compact set, so a minimizer exists, and the result is immediate. □

Appendix C. Modifying E-posteriors to optimize post-adoption bound

Let $\mathcal{C}_{E,1+\gamma}$ denote the following class of decision rules: For an e-variable $E(\theta,Y)$, define $E_\gamma(\theta,Y) = \gamma E(\theta,Y) + 1$,[11] consider

$$\delta(Y) \in \arg\min_{a\in\mathcal{A}} \sup_{\theta\in\Theta} \frac{L(a,\theta)}{E_\gamma(\theta,Y)} \quad R(Y) = \min_{a\in\mathcal{A}} \sup_{\theta\in\Theta} \frac{L(a,\theta)}{E_\gamma(\theta,Y)}. \tag{12}$$

In this section, let us assume $\mathcal{A}$ is finite for simplicity, and thus the minimum always exists. Since $E_\gamma/(1+\gamma)$ is an e-variable, such a decision rule is E-certified with multiple $1+\gamma$.

We show that adoptions rules with $Q \leq \mathbb{1}(R(Y) \leq C)$ entail risk at most $(1+\gamma)C$ post adoption, in contrast to Proposition 4.3. This guarantee holds because

$$\mathbb{E}_P\left[\max\left(\frac{L(\delta(Y),\theta)}{R(Y)}, 1\right)\right] \leq 1+\gamma.$$

Moreover, for all decisions (δ, R) satisfying the above, decision rules in $\mathcal{C}_{E,1+\gamma}$ are essentially complete with respect to weak dominance.

[11] We can also view $E_\gamma = (1+\gamma)\left(\frac{\gamma}{1+\gamma}E(\theta,Y) + \frac{1}{1+\gamma}\right) = (1+\gamma)S_\theta^{[\gamma/(1+\gamma)]}(Y)$, where $S_\theta^{[c]}(Y) = cE(\theta,Y) + (1-c)$ is proposed by Grünwald (2023) as a modification to e-variables that produces upper bounded E-posteriors. Our subsequent result gives some guidance on how different choices of γ impact the resulting guarantee.

Proposition C.1. *Given* $(\delta, R) \in \mathcal{C}_{E,1+\gamma}$*, we have that for any adoption decision* $Q \leq \mathbb{1}(R(Y) \leq C)$*,*

$$\mathbb{E}_P\left[L(\delta_Q(Y),\theta)\right] \leq \mathbb{E}_P\left[\max\left(\frac{L(\delta(Y),\theta)}{R(Y)},1\right)C\right] \leq (1+\gamma)C \text{ for all } (\theta,P) \in \mathcal{M}$$

Moreover, for any pair $(\tilde{\delta},\tilde{R})$ *such that* $\mathbb{E}_P\left[\max\left(\frac{L(\tilde{\delta}(Y),\theta)}{\tilde{R}(Y)},1\right)\right] \leq 1+\gamma$*, there exists some* $(\delta,R) \in \mathcal{C}_{E,1+\gamma}$ *that weakly dominates it.*

Proof. Observe that

$$L(\delta_Q(Y),\theta) = \frac{L(\delta_Q(Y),\theta)}{QR(Y)+(1-Q)C}\left(QR(Y)+(1-Q)C\right) \leq \max\left(\frac{L(\delta(Y),\theta)}{R(Y)},1\right)C.$$

Since

$$\mathbb{E}_P\left[\max\left(\frac{L(\delta(Y),\theta)}{R(Y)},1\right)\right] \leq \mathbb{E}_P\left[\max\left(\frac{L(\delta(Y),\theta)}{L(\delta(Y),\theta)/E_\gamma(Y,\theta)},1\right)\right] = \mathbb{E}_P\left[E_\gamma(Y,\theta)\right] \leq 1+\gamma.$$

this proves the first claim.

For the second claim, let

$$E_\gamma(Y,\theta) = \max\left(\frac{L(\tilde{\delta}(Y),\theta)}{\tilde{R}(Y)},1\right).$$

Since its expectation is bounded by $1+\gamma$, $(E_\gamma(Y,\theta)-1)/\gamma$ is an E-variable. Thus there is some $(\delta,R) \in \mathcal{C}_{E,1+\gamma}$ that corresponds to $E_\gamma(Y,\theta)$. Note that

$$R(Y) = \min_{a\in\mathcal{A}}\sup_{\theta\in\Theta}\frac{L(a,\theta)}{E_\gamma(Y,\theta)} \leq \sup_{\theta\in\Theta}\min\left(\tilde{R}(Y)\cdot L(\tilde{\delta}(Y),\theta)/L(\tilde{\delta}(Y),\theta), L(\tilde{\delta}(Y),\theta)\right) \leq \tilde{R}(Y).$$

This completes the proof. □

References

ANDREWS, I., KITAGAWA, T. and MCCLOSKEY, A. (2024). Inference on winners. *The Quarterly Journal of Economics*, **139** (1), 305–358. 3, 4, 9, 10

BANERJEE, A. V., CHASSANG, S., MONTERO, S. and SNOWBERG, E. (2020). A theory of experimenters: Robustness, randomization, and balance. *American Economic Review*, **110** (4), 1206–1230. 7

BATES, S., JORDAN, M. I., SKLAR, M. and SOLOFF, J. A. (2024). Incentive-theoretic bayesian inference for collaborative science. *arXiv preprint arXiv:2307.03748*. 2

Ben-Michael, E., Greiner, D. J., Imai, K. and Jiang, Z. (2021). Safe policy learning through extrapolation: Application to pre-trial risk assessment. *arXiv preprint arXiv:2109.11679*. 2

Benjamini, Y., Hechtlinger, Y. and Stark, P. B. (2019). Confidence intervals for selected parameters. *arXiv preprint arXiv:1906.00505*. 3, 9

Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L. F., and Palmer, C. (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review*, **114** (5), 1281–1337. 10

Chernozhukov, V., Lee, S., Rosen, A. M. and Sun, L. (2025). Policy learning with confidence. 2, 4, 13

Grünwald, P. D. (2023). The e-posterior. *Philosophical Transactions of the Royal Society A*, **381** (2247), 20220146. 4, 16, 24

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, **86** (2), 591–616. 3, 10

Kiyani, S., Pappas, G., Roth, A. and Hassani, H. (2025). Decision theoretic foundations for conformal prediction: Optimal uncertainty quantification for risk-averse agents. *arXiv preprint arXiv:2502.02561*. 4, 5, 12

Lehmann, E. L. and Romano, J. P. (2024). *Testing statistical hypotheses*. Springer. 14

Manski, C. F. (2021). Econometrics for decision making: Building foundations sketched by haavelmo and wald. *Econometrica*, **89** (6), 2827–2853. 1, 2

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **236** (767), 333–380. 9

— (1977). Frequentist probability and frequentist statistics. *Synthese*, **36** (1), 97–131. 2

Ponomarev, K. and Semenova, V. (2024). On the lower confidence band for the optimal welfare. *arXiv preprint arXiv:2410.07443*. 3

Ramdas, A. and Wang, R. (2024). Hypothesis testing with e-values. *arXiv preprint arXiv:2410.23614*. 4, 15

Tetenov, A. (2016). An economic theory of statistical testing. 2

Viviano, D., Wuthrich, K. and Niehaus, P. (2025). A model of multiple hypothesis testing. *arXiv preprint arXiv:2104.13367*. 2

Wald, A. (1949). Statistical decision functions. *The Annals of Mathematical Statistics*, pp. 165–205. 2

Zrnic, T. and Fithian, W. (2024a). A flexible defense against the winner's curse. *arXiv preprint arXiv:2411.18569*. 3, 9

— and — (2024b). Locally simultaneous inference. *The Annals of Statistics*, **52** (3), 1227 – 1253. 3, 9