When Fit and Test Summaries Do Not Support Later Estimation

Albaraa Mesawi

Preprint

Abstract

When a retained archive fails a reliability specification, an inquiry can change its procedures, target, tolerance, state class, weights, or resources. This paper assesses which repairs answer a declared question at a required uniform reliability. A contributory point specification has disjoint success intervals whenever live states require different answers. More generally, sufficiently frequent decisive successes, which lie in no success interval for a different answer, certify reliable decoding. This certificate is sufficient: its failure leaves open decoding that uses the known procedure's report distribution. An elementary obstruction gives negative judgments directly. When two live states with different answers give the archive the same probability law, no decoder answers correctly with probability above one-half at both. Total-variation bounds quantify the repair and exclusion limits for unequal archive laws. In a hypothetical Gaussian drift archive that omits the fitted drift's sign, two class restrictions are each attainable yet issue opposite answers on almost every report. The archive cannot discriminate between the paired states to justify either restriction. Unequal weights change the guarantee even when justified; they cannot recover the original uniform requirement. A coarser tolerance and a different target are attainable but do not recover that requirement. Retaining the sign supports a relaxation of a tolerance that no procedure can meet on the protocol's class. A hypothetical pooled-count archive gives a second application. These results concern reliable answering, leave broader inquiry continuation and intermediate evidence open, and supply no empirical justification for the hypothetical aims or parameter bounds.

Keywords: data reuse; adequacy for purpose; measurement; reliability specification; identification; decision theory.

1 What an archive permits

A retained archive is assessed against an aim. Beauchemin and Staley tie the later usefulness of a measurement result to its sensitivity to the distinctions a subsequent inquiry needs, and their critical assessment can revise tasks, resources, and aims (Beauchemin and Staley, 2025a). Bokulich and Parker define adequacy given resources through access to some coherent way of using data to achieve a purpose, and they distinguish reuse, which addresses again a question the data already addressed, from repurposing, which serves a purpose other than the one initially envisioned (Bokulich and Parker, 2021). Parker separates success in a particular use from reliability across a type of use and recommends, when adequacy for one purpose is unclear, asking whether there is evidence of adequacy for a related, easier purpose that is also of interest (Parker, 2020). This paper examines six formal responses to a failed reliability specification: reprocessing, changing the target, loosening the tolerance, weighting states, narrowing the admissible class, and adding resources.

Beauchemin and Staley relate the tradeoff between adaptable evidence and discriminatory power to Staley's account of the security and strength of an evidence claim (Beauchemin and Staley,

2025b). They list responses to a result that is insensitive to a targeted difference: re-estimating the uncertainties, for example with a method that uses the data more efficiently; redesigning the measurement; changing the objective so that the evidence serves a new function; or obtaining access to a more sensitive result. They expect inquirers in the end to adopt some objective that the uncertainty does not defeat (Beauchemin and Staley, 2025b). The question here is narrower: which responses answer a declared question at its required reliability, and when is re-estimation closed for fixed retained inputs?

The accounts require that a changed purpose be of interest or serve core objectives. Those requirements allow responses that contribute to an inquiry without completing it. This paper assesses the narrower claim that a repaired report answers a specified question at a declared uniform reliability. For comparison, define a deliberately permissive *success-rate standard*: it accepts a repair whenever some procedure meets the repaired specification at its stated level. No cited account endorses this standard, and its failure identifies no defect in those accounts. Call a report a *decisive success* when it meets the specification at a live state and would not also meet it at a live state with a different core answer. Call a specification *contributory* when every success is decisive. Meeting such a specification at the required reliability certifies reliable answering. A graded version permits rare indecisive successes. These are sufficient certificates; a known procedure can encode the answer even in reports that fall in overlapping success intervals.

An archive obstruction supplies negative judgments independently of those certificates. When two live states with different core answers give the retained information the same law, every decoder has correct-answer probability at most one-half at one of them. Every contributory change of target, tolerance, or uniform level remains unattainable. A changed point specification that the archive meets at level $1 - \gamma$ at both obstructed states issues, with probability at least $1 - 2\gamma$, a report that counts as a success at both. Keeping the original uniform answer task requires additional information and a performance check. A justified change of states or guarantees changes the task being assessed. Merely distinguishing one pair does not establish adequate performance over the live class.

The repair comparisons make these distinctions exact. Consider a hypothetical laboratory that measures a constant quantity while the instrument's offset drifts after calibration. Its archive retains the stable-instrument likelihood and an exact two-sided drift test but omits the fitted drift's sign. Two restrictions of the drift class each permit reliable discrimination between two mechanistic models, yet their rules issue opposite answers at the obstructed states with probability above 0.999999998. The archive cannot discriminate between those states to select a restriction. A prior-weighted rule achieves average success 0.99 while failing at one live model; justified weights would change the required guarantee. A coarser tolerance succeeds with probability at least 0.998650, and a different target has failure probability below 0.000000574, but the equal-law bound still excludes reliable discrimination from this archive. Retaining the sign permits relaxation from 1/5 to 1/2, while a two-point bound excludes the finer tolerance for every procedure using the complete observations on the protocol's class. A second hypothetical archive pools event counts across two windows. It illustrates the same distinctions with a different information loss, nuisance background, and repair.

The claims are conditional on a declared question, live class, and reliability requirement. Both cases are hypothetical, with all premises stated. The statistics are established: likelihood factorization, Gaussian regression, two-point lower bounds, and exponential-moment bounds for Poisson counts. The proposed contribution is an exact conditional analysis of repairs using these results and sufficient reading certificates. The paper offers no general theory of inquiry identity and supplies no empirical assessment of an actual archive or measurement.

Sufficiency concerns a specified statistical model. In the Gaussian models used here, a sufficient statistic permits a possibly randomized counterpart to every full-data rule with the same risk. Simulating a full record from its parameter-independent conditional law given the statistic supplies that counterpart (Casella and Berger, 2002). Sufficiency under the original model therefore justifies a reduction within that model. Preserving a diagnostic after model expansion raises a separate issue: whether the remaining distinctions support the later aim. The example assumes familiarity with likelihood factorization and elementary Gaussian regression.

An audit before choosing a correction

Assessing a proposed reduction starts with the target, admissible parameter states, acceptable outputs, and accessible information. Next, look for states that give the retained information the same probability law. If those states require disjoint sets of acceptable outputs, every procedure using that information has the same output law at both states and cannot succeed with probability above one-half at both. A random seed independent of the state and observations does not change this conclusion. A proposed repair needs a separate positive check: exhibit a procedure using the additional information and calculate its accuracy.

The following lemma states the elementary two-state argument used in both designs.

Lemma 1 (Retained-law obstruction)**.** *Let T be a measurable retained statistic, q a scalar target, and $a > 0$. Suppose two admissible states η, η' satisfy*

$$\mathcal{L}_\eta(T(Y)) = \mathcal{L}_{\eta'}(T(Y)), \qquad |q(\eta) - q(\eta')| > 2a.$$

Every measurable estimator $\delta(T(Y), U)$, with one fixed seed law independent of Y at each state, has success probability at most $1/2$ at one of these states. Equality of the retained laws follows, in particular, if a measurable transformation g satisfies $T \circ g = T$ and the law of $g(Y)$ at η equals the law of Y at η'.

Proof. The joint law of $(T(Y), U)$ is the same at both states, so the estimator has one output law Q. The closed success intervals of radius a are disjoint. Their Q-probabilities sum to at most one, giving the bound. Under the transformation condition, the law of $T(Y)$ at η' equals that of $T(g(Y)) = T(Y)$ at η. □

The strict target inequality matters because success includes the interval endpoints. The lemma gives a sufficient failure certificate. Failure to find such a pair does not establish adequacy: sampling uncertainty or restrictions on available procedures can still defeat the task. Equality of summary values for two individual records is also insufficient; the argument requires equality of the entire retained-data laws. The transformations below establish that equality directly.

Lemma 1 is the equal-law case of a quantitative bound. Write $\mathrm{TV}(P, P') = \sup_B |P(B) - P'(B)|$ for the total-variation distance between two laws on the same space, and let Φ denote the standard normal distribution function.

Lemma 2 (Two-point bound)**.** *Let Z be an information source with law P_η at state η, let q be a scalar target, and let $a > 0$. If two admissible states satisfy $|q(\eta) - q(\eta')| > 2a$, then every measurable estimator $\delta(Z, U)$, with one fixed seed law independent of Z, has failure probability at least $\frac{1}{2}\{1 - \mathrm{TV}(P_\eta, P_{\eta'})\}$ at η or at η'. For Gaussian observations $Y \sim N(\mu_\eta, \sigma^2 I)$ with a common $\sigma > 0$,*

$$\mathrm{TV}(P_\eta, P_{\eta'}) = 2\Phi(\|\mu_\eta - \mu_{\eta'}\|/(2\sigma)) - 1.$$

Proof. Let ν be the seed law and put $K(B \mid z) = \nu\{u : \delta(z,u) \in B\}$ for Borel sets B; joint measurability of δ makes $K(B \mid \cdot)$ measurable. The output laws at the two states are $Q(B) = \int K(B \mid z)\, P_\eta(dz)$ and $Q'(B) = \int K(B \mid z)\, P_{\eta'}(dz)$. Since $0 \le K \le 1$, the difference $Q(B) - Q'(B) = \int K(B \mid z)\,(P_\eta - P_{\eta'})(dz)$ is at most the mass of the positive part of the signed measure $P_\eta - P_{\eta'}$, which equals $\mathrm{TV}(P_\eta, P_{\eta'})$. The same bound holds for $Q'(B) - Q(B)$. The closed success intervals I and I' of radius a about $q(\eta)$ and $q(\eta')$ are disjoint, so

$$Q(I) + Q'(I') \le Q(I) + Q(I') + \mathrm{TV}(P_\eta, P_{\eta'}) \le 1 + \mathrm{TV}(P_\eta, P_{\eta'}).$$

The smaller of the two success probabilities is therefore at most $\frac{1}{2}\{1 + \mathrm{TV}(P_\eta, P_{\eta'})\}$. For the Gaussian formula, write $\Delta = \mu_{\eta'} - \mu_\eta$ and rotate coordinates so that $\Delta = \|\Delta\| e_1$. Total variation is unchanged by this bijection. In the new coordinates both laws are products of a law on the first coordinate and a common law on the remaining coordinates, and for such products the distance equals the distance between the first factors. That distance is between $N(0, \sigma^2)$ and $N(\|\Delta\|, \sigma^2)$. The first density exceeds the second exactly on $\{x < \|\Delta\|/2\}$, where the difference of the two probabilities is $\Phi(\|\Delta\|/(2\sigma)) - \Phi(-\|\Delta\|/(2\sigma)) = 2\Phi(\|\Delta\|/(2\sigma)) - 1$. □

The proof uses only that the two success sets are disjoint Borel sets of outputs. The bound therefore holds for any rule whose success sets at η and η' are disjoint, including a rule that issues one of two distinct labels.

Equal laws give distance zero and recover Lemma 1. A positive distance gives a smaller floor on the failure probability, and the floor still holds for every procedure. It separates a specification that no procedure meets from one that more information could meet. It also gives a necessary condition for any new information source: a source that meets such a specification at level $1 - \gamma$ at both states must have total-variation distance at least $1 - 2\gamma$ between its two laws.

2 Certificates for repairs

This section distinguishes sufficient certificates for reliable answering from an obstruction that excludes every decoder.

Setting. Fix a class $\mathcal{K}$ of live states, an information source Z with law P_η at each $\eta \in \mathcal{K}$, and the measurable rules $\delta(Z, U)$ with one fixed seed law independent of Z. A *core aim* is a map A from $\mathcal{K}$ to a finite set of answers with at least two elements. A rule for A issues an answer, and it achieves A at η when it issues $A(\eta)$. The states in $\mathcal{K}$ are live in the sense that the inquiry requires its answer to be reliable at each of them. A *point specification* $S = (q, a, \gamma)$, with $a > 0$ and $0 < \gamma < 1/2$, requires a real report within a of $q(\eta)$ with probability at least $1 - \gamma$ at every state of a declared class that contains $\mathcal{K}$. A fair choice between two answers meets any level $\gamma \ge 1/2$ at a pair of states, so such levels are excluded. Write $I_S(\eta) = [q(\eta) - a,\, q(\eta) + a]$ for the success interval at η.

> **Definition.** S is *contributory* to A on $\mathcal{K}$ when $I_S(\eta) \cap I_S(\eta') = \emptyset$ for every pair $\eta, \eta' \in \mathcal{K}$ with $A(\eta) \ne A(\eta')$.

The term follows Beauchemin and Staley's classification of objectives as overarching or contributory (Beauchemin and Staley, 2025a); the definition specifies one sufficient relation between a point specification and an answer. A *label specification* requires the issued answer to equal $A(\eta)$ with probability at least $1 - \gamma$ at each state of a declared class. Its success sets at states with different answers are disjoint, so it is contributory by construction. For labels, a decisive success means issuing the correct answer. Uniform and weighted assessments of that event impose different performance requirements.

Lemma 3 (Reading)**.** *S is contributory to A on $\mathcal{K}$ if and only if some map f from reports to answers satisfies $f(x) = A(\eta)$ whenever $\eta \in \mathcal{K}$ and $x \in I_S(\eta)$. If S is contributory and a rule δ meets S, then $f \circ \delta$ achieves A with probability at least $1 - \gamma$ at every state of $\mathcal{K}$.*

Proof. If S is contributory, define $f(x) = A(\eta)$ for any $\eta \in \mathcal{K}$ with $x \in I_S(\eta)$, and assign one fixed answer to reports in no success interval. Two states whose intervals contain x have the same answer, so f is well defined. For each answer, the reports lying in a success interval of some state with that answer form a union of nondegenerate intervals; such a union has countably many connected components, each an interval, so it is a Borel set, and f is Borel measurable. If S is not contributory, some report x lies in $I_S(\eta) \cap I_S(\eta')$ with $A(\eta) \neq A(\eta')$, and no value $f(x)$ equals both answers. The last claim holds because success at η gives $f(\delta) = A(\eta)$. □

Contributoriness is a property of the specification alone: it guarantees a correct reading of every successful report without using its distribution under the adopted procedure. A specification that is not contributory can still be met by a procedure whose successes are almost always decisive. For a point specification S and $\eta \in \mathcal{K}$, let $O_S(\eta)$ be the union of the success intervals $I_S(\eta')$ over the states $\eta' \in \mathcal{K}$ with $A(\eta') \neq A(\eta)$. A report x is a *decisive success* at η when $x \in I_S(\eta)$ and $x \notin O_S(\eta)$. A success that is not decisive also counts as a success at a live state with another answer. This interval compatibility alone says nothing about how likely the procedure is to issue that report at either state.

Lemma 4 (Graded reading)**.** *For every point specification S there is a Borel map f from reports to answers with $f(x) = A(\eta)$ whenever x is a decisive success at $\eta \in \mathcal{K}$. If a rule δ issues a decisive success with probability at least $1 - \gamma''$ at every state of $\mathcal{K}$, then $f \circ \delta$ achieves A with probability at least $1 - \gamma''$ at every state of $\mathcal{K}$. The specification S is contributory to A on $\mathcal{K}$ if and only if, at every $\eta \in \mathcal{K}$, every report in $I_S(\eta)$ is a decisive success.*

Proof. For each answer c, let U_c be the union of the intervals $I_S(\eta)$ over the states $\eta \in \mathcal{K}$ with $A(\eta) = c$. As in the proof of Lemma 3, each U_c is a Borel set, and $O_S(\eta)$ is the union of the sets U_c with $c \neq A(\eta)$. Put $f(x) = c$ when $x \in U_c$ and $x \notin U_{c'}$ for every answer $c' \neq c$, and assign one fixed answer otherwise. There are finitely many answers, so f is Borel. If x is a decisive success at η, then $x \in U_{A(\eta)}$ and x lies in no U_c with $c \neq A(\eta)$, so $f(x) = A(\eta)$. The second claim follows because the event of a decisive success at η is contained in the event $f(\delta) = A(\eta)$. For the last claim, S is contributory exactly when $I_S(\eta) \cap O_S(\eta) = \emptyset$ for every $\eta \in \mathcal{K}$. □

Scope of the certificates. Beauchemin and Staley require successful inquiry to satisfy its core objectives, even when other objectives remain unmet (Beauchemin and Staley, 2025a). A response can continue that inquiry or supply intermediate evidence before completing it. The following certificate concerns completion of the declared answer task at its required reliability. It does not classify every research response as continuing or replacing an inquiry.

> **Sufficient answer certificate.** Fix A, the live class $\mathcal{K}$, and an accepted uniform level $1 - \gamma'' > 1/2$. If the procedure adopted for a repaired point specification issues a decisive success with probability at least $1 - \gamma''$ at every $\eta \in \mathcal{K}$, Lemma 4 supplies one measurable reading that answers A at that level. A contributory specification met at that reliability supplies the same certificate for every procedure meeting it. Failure of this sufficient certificate leaves reliable answering open: a decoder can use information in the known procedure's report distribution. Any added resource must be accessible, and its performance requires a separate check.

A change of class also needs error accounting. If a procedure selects a restricted class from the archive and then answers, its claimed uniform reliability must cover the complete selection-and-answer procedure at every original live state. An archive that reveals state identity can safely exclude the false state. Lemma 7 gives a narrower prohibition: an archive cannot discriminate between an equal-law pair to support excluding one of its members. Separate information can remove that obstruction. Statistical independence from the archive is neither required for such information nor sufficient to make it useful. Noncircular justification for a changed task means reasons other than its attainability; it is distinct from either statistical independence or a separate information source.

The analysis retains the uniform requirement. Unequal weights, even when justified, replace it with an average guarantee and do not certify its recovery while both states remain live. A lower uniform level also changes the guarantee; whether it is acceptable depends on the consequences of error. The certificates do not choose the level. The results below establish limits for any decoder, independently of whether a sufficient certificate succeeds.

Proposition 5 (Obstruction inheritance)**.** *Suppose $\eta, \eta' \in \mathcal{K}$ satisfy $A(\eta) \neq A(\eta')$ and $\mathcal{L}_\eta(Z) = \mathcal{L}_{\eta'}(Z)$. Then:*

1. *every rule $\delta(Z, U)$ for A achieves A with probability at most $1/2$ at η or at η';*
2. *every specification contributory to A on $\mathcal{K}$ fails under Z: for each rule, one of η, η' has failure probability at least $1/2 > \gamma$;*
3. *no change of q, of a, or of $\gamma \in (0, 1/2)$ that keeps the specification contributory to A on $\mathcal{K}$ is attainable under Z.*

Proof. The joint law of (Z, U) is the same at both states, so a rule has one output law Q at both. For item 1, $Q\{A(\eta)\} + Q\{A(\eta')\} \leq 1$ because the two answers differ. For item 2, a contributory specification has disjoint success intervals at η and η', so their Q-probabilities sum to at most one, and one of them is at most $1/2$. Item 3 applies item 2 to the repaired specification. □

Corollary 6 (Repairs met at a near-law pair)**.** *Let $\eta, \eta' \in \mathcal{K}$ satisfy $A(\eta) \neq A(\eta')$ and $\mathrm{TV}(P_\eta, P_{\eta'}) \leq \epsilon \in [0, 1]$. Every procedure uses the same measurable kernel at both states, with one fixed seed law independent of the state and Z. Its output laws Q, Q' satisfy $\mathrm{TV}(Q, Q') \leq \epsilon$. For correct labels, or for decisive successes under any point specification S', the smaller success probability is at most $(1+\epsilon)/2$. If a rule meets S' with failure probability at most $\gamma' < 1/2$ at both states, then, under each state, its report lies in $I_{S'}(\eta) \cap I_{S'}(\eta')$ with probability at least $\max\{0, 1 - 2\gamma' - \epsilon\}$. If $\epsilon < 1 - 2\gamma'$, no contributory repair can be met at this level. At $\epsilon = 0$, the bounds are $1/2$ and $1 - 2\gamma'$, and every such met S' is noncontributory.*

Proof. The kernel contraction proved in Lemma 2 gives $\mathrm{TV}(Q, Q') \leq \epsilon$. Correct-label success sets D, D' are disjoint. Decisive-success sets are also disjoint: writing $I = I_{S'}(\eta)$ and $I' = I_{S'}(\eta')$, they lie in $I \setminus I'$ and $I' \setminus I$. In either case,

$$Q(D) + Q'(D') \leq Q(D) + Q(D') + \epsilon \leq 1 + \epsilon.$$

If the rule meets S' at both states, the union bound gives

$$Q(I \cap I') \geq Q(I) + Q(I') - 1 \geq Q(I) + Q'(I') - 1 - \epsilon \geq 1 - 2\gamma' - \epsilon.$$

Exchanging Q, Q' gives the same bound under the other state. Probabilities are nonnegative, and a positive lower bound excludes disjoint success intervals. □

The overlap bound is sharp on a two-state class. For $0 \le \epsilon \le 1 - 2\gamma'$, use reports $-3/2, 0, 3/2$ with laws $(\gamma' + \epsilon, 1 - 2\gamma' - \epsilon, \gamma')$ and $(\gamma', 1 - 2\gamma' - \epsilon, \gamma' + \epsilon)$, and success intervals $[-2, 0]$ and $[0, 2]$. Each succeeds with probability $1 - \gamma'$, the total variation is ϵ, and the overlap has probability $1 - 2\gamma' - \epsilon$ under each law. For $\epsilon \ge 1 - 2\gamma'$, use disjoint intervals of radius 1 centered at $-2, 2$, with masses $((1 + \epsilon)/2, (1 - \epsilon)/2)$ and their reversal at those reports. Each success probability is $(1 + \epsilon)/2 \ge 1 - \gamma'$, and the overlap is empty. This two-report construction attains the label and decisive-success bound for every $\epsilon \in [0, 1]$.

At equal laws, Proposition 5, item 1, excludes every decoder with success probability above $1/2$ at both states. Decoding a repaired report cannot escape that bound, regardless of its decisive-success probability. For unequal laws, Corollary 6 bounds the smaller correct-answer probability by $(1 + \epsilon)/2$; it excludes a required level $1 - \gamma''$ when $\epsilon < 1 - 2\gamma''$. A positive ϵ does not exclude every level above $1/2$. The overlap bound describes how often an attained point specification accepts the issued report under both answers. Equal report laws establish absence of discrimination at this pair; unequal laws give a quantitative limit. Other uses or intermediate contributions remain open.

Two lemmas cover the remaining kinds of repair.

Lemma 7 (Exclusions at a near-law pair)**.** *Under the pair and kernel assumptions of Corollary 6, let $\rho(Z, U)$ output a subset of $\mathcal{K}$ to exclude, with measurable exclusion events. Then*

$$\left| \Pr_{\eta}\{\eta' \in \rho(Z, U)\} - \Pr_{\eta'}\{\eta' \in \rho(Z, U)\} \right| \le \epsilon,$$

and the same holds with η and η' exchanged. A rule that excludes η' with probability r when η holds excludes the true state with probability at least $\max\{0, r - \epsilon\}$ when η' holds. At $\epsilon = 0$, the probabilities are equal, so the true state is excluded with exactly probability r.

Proof. Apply the output-law total-variation bound to the fixed measurable event $E = \{\eta' \in \rho(Z, U)\}$. Exchange the states for the other claim. □

Lemma 8 (Candidate resources)**.** *Let M be a candidate additional resource. If the joint law of (Z, M) is the same at η and at η', then Proposition 5 holds for (Z, M) in place of Z. A resource $M = f(Z, V)$, with one state-independent law of V independent of Z, preserves total variation exactly when retained jointly with Z:*

$$\mathrm{TV}\left(\mathcal{L}_{\eta}(Z, M), \mathcal{L}_{\eta'}(Z, M)\right) = \mathrm{TV}(P_{\eta}, P_{\eta'}).$$

Such a resource never removes an equal-law obstruction. Processing Z alone can only contract the distance. Differing joint laws remove this pair's equal-law certificate only; the near-law bounds can still exclude the required reliability. Adequacy for the core aim requires a separate performance check.

Proof. The first claim applies Proposition 5 to (Z, M). For $M = f(Z, V)$, kernel contraction bounds the joint total variation by $\mathrm{TV}(P_{\eta}, P_{\eta'})$, and projection onto Z gives the reverse bound. Processing Z alone gives the contraction bound. Differing joint laws defeat the equality hypothesis, leaving the quantitative bounds and performance check applicable. □

Corollary 9 (Weighted criteria withdraw protection)**.** *Under the hypothesis of Proposition 5, give η weight $p > 1/2$ and η' weight $1 - p$, and let s and s' be a rule's success probabilities at the two states, either for A or for a specification contributory to A. If $ps + (1 - p)s' \ge 1 - \gamma$, then*

$$s' \le \frac{p + \gamma - 1}{2p - 1}.$$

If $p + \gamma < 1$, no rule meets the weighted requirement.

Proof. The two success events are disjoint under one output law, so $s + s' \le 1$ and $s \le 1 - s'$. Substituting gives $p(1 - s') + (1 - p)s' \ge 1 - \gamma$, which rearranges to $s'(1 - 2p) \ge 1 - \gamma - p$. Since $1 - 2p < 0$, this is the displayed bound. A negative bound is infeasible. □

With $p = 0.99$ and $\gamma = 0.05$ the bound is $2/49 \approx 0.0408$, and a rule that issues the two answers with probabilities $47/49$ and $2/49$ attains it. With $\gamma = 0.01$ the bound is 0. More generally, the bound minus γ equals $(1 - p)(2\gamma - 1)/(2p - 1) \le 0$ for $1/2 < p \le 1$ and $\gamma < 1/2$. A weighted criterion changes the performance required at the less weighted state. Reasons for those weights can justify an average task, but cannot make its guarantee uniform. At the equal-law pair the archive also supplies no evidence for updating the relative state weights.

The positive assessment requires a procedure and a performance check. When no pair of live states with different core answers shares a Z-law, Proposition 5 does not apply. A contributory relaxation can then be attainable where the original specification was not; meeting it at the required reliability certifies the answer task. Lemma 2 can also establish that the original tolerance was unattainable for every procedure. Section 6 exhibits both a valid certificate and its limits.

3 The experiment and the reduction

A constant physical quantity θ is measured at fixed known times

$$t = (0, 1, 2, 3).$$

The working model assumes a stable instrument:

$$M_0 : \quad Y_i = \theta + \epsilon_i, \qquad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad \sigma > 0.$$

The expanded model allows linear drift in the instrument's offset:

$$M_1 : \quad Y_i = \theta + bt_i + \epsilon_i.$$

Both models assume calibration fixing zero systematic offset at the initial time. This exact calibration is an idealization that gives θ its physical interpretation. An arbitrary choice of time origin would not establish it. An unknown initial offset would instead confound the physical quantity with that offset.

Define

$$s_1 = \sum_i Y_i, \qquad s_2 = \sum_i Y_i^2, \qquad \bar{t} = \frac{3}{2}, \qquad k = \sum_i (t_i - \bar{t})Y_i, \qquad D = \sum_i (t_i - \bar{t})^2 = 5.$$

The fitted drift slope is k/D, so the sign of k gives the estimated drift direction. The reduction retains the design, units, calibration interpretation, model assumptions, and

$$T(Y) = (s_1, s_2, k^2). \tag{1}$$

An estimator based on the reduced arrangement may use any measurable function of T and an auxiliary random seed independent of the state and observations. No additional observation or retained resource distinguishes the discarded sign.

The summary retains the working-model fit and unsigned drift information; the later two-sided slope-test calculation in equation (6) verifies its relation to the exact test output.

Fix the parameter class

$$\mathcal{G} = \{(\theta, b, \sigma) : \theta \in \mathbb{R},\ |b| \le 4/5,\ 0 < \sigma \le 1/5\}. \tag{2}$$

The task requires a real-valued estimate within $a = 1/2$ of θ. Section 6 varies a and declares a core aim; this section fixes $a = 1/2$. Let $\mathcal{D}_T$ contain the measurable procedures using T and the permitted independent random seed. Adequacy for this task requires

$$\exists \delta \in \mathcal{D}_T \quad \forall \eta \in \mathcal{G} : \qquad \Pr_{\eta}\{|\delta(T, U) - \theta| \le 1/2\} \ge 0.95. \tag{3}$$

One rule precedes the universal state quantifier because the analyst must choose a procedure without knowing the state. The negation says that every permitted rule fails the requirement at some admissible state; that state can depend on the rule. It does not say that one state defeats every rule. Refusal or unresolved alternatives do not complete this point-estimation task.

The requirement could be warranted if a measurement protocol is required to deliver this accuracy throughout an independently justified operating class, and failures at any included state would defeat its scientific use. Calibration evidence, permissible drift, noise bounds, and the consequences of half-unit errors would need to justify that class and tolerance. Merely lacking prior weights does not entail a uniform criterion. The present hypothetical construction supplies none of this empirical or evaluative justification; it examines the conditional consequence of adopting these requirements.

A successful token use is one accurate issued answer. Reliability in a type of use concerns how often success occurs under its stated circumstances. Equation (3) specifies one demanding form of that reliability by requiring the same lower bound at every allowed state. For a fixed rule, evaluating success at one state is a pointwise claim and leaves its performance elsewhere open. Reversing the displayed quantifiers would allow a different rule for each known state, including a constant equal to its target. A prior-weighted criterion instead averages the statewise success probabilities. These objectives require separate assessments; the two-state contrast below makes the difference exact.

Figure 1 connects the lost drift direction to the acceptable estimates. The two hypothetical cases use the same four measurement times and the allowed drift magnitudes.

Recovering the calibrated quantity from the shared expected mean requires subtracting 1.2 in the increasing-drift case and adding 1.2 in the decreasing-drift case. Panel (b) shows why one common output law cannot assign success probability above one-half to both calibrated targets. The proof below supplies the needed equality of laws by reversing the full noisy readings; panel (a) displays the corresponding reversed expectations.

For an information source Z, let

$$R_Z^* = \inf_{\delta} \sup_{\eta \in \mathcal{G}} \Pr_{\eta}\{|\delta(Z, U) - \theta| > 1/2\}, \tag{4}$$

where $\eta = (\theta, b, \sigma)$ and U is the permitted independent random seed. This minimax risk is the infimum of the attainable worst-case failure probabilities; no minimizing estimator is assumed to exist. The comparison will use $Z = T(Y)$ and $Z = Y$.

4 What survives and what fails

The likelihood and test calculations identify what the archive preserves. The lower bound excludes every permitted estimator for the fine task, and the upper bound shows that the unreduced observations can meet it. The signed summary supplies the repair.

Proposition 10. *For the experiment and task just specified:*

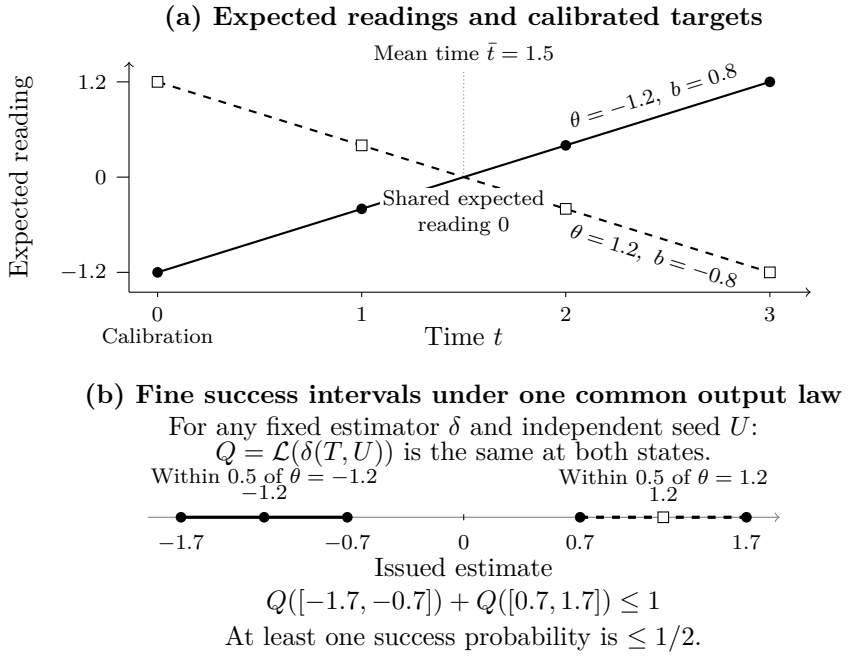


Figure 1: Opposite drift directions and incompatible fine estimates. (a) Lines show model expectations; circle and square markers identify expectations at the four observation times. Both expected mean-time readings equal zero, while the calibrated targets at $t = 0$ are -1.2 and 1.2. Measurements have independent Gaussian errors with a common $0 < \sigma \leq 1/5$. The reversal proof in the text establishes equality of the unsigned archive laws; equal expected means alone do not establish it. (b) Every permitted archive-based estimator, including one using an independent random seed, has the common output law Q. Its closed half-unit success intervals are disjoint, so their probabilities sum to at most one. The common-law argument gives a lower bound on worst-case failure probability.

1. *T preserves the likelihood under M_0.*

2. *T determines the ordinary two-sided slope-test statistic under M_1 pointwise, so every threshold rule based on it is unchanged by the reduction.*

3. *$R_T^* \geq 1/2$, whereas*

$$R_Y^* \leq 2\left[1 - \Phi\left(\frac{1/2}{(1/5)\sqrt{7/10}}\right)\right] \approx 0.00280744.$$

In particular,

$$R_T^* - R_Y^* \geq 2\Phi\left(\frac{1/2}{(1/5)\sqrt{7/10}}\right) - \frac{3}{2} \approx 0.497193.$$

4. *Retaining the sign of k in addition to T restores a sufficient statistic for the full Gaussian drift model and an estimator meeting the task.*

Proof. Under M_0, the likelihood is

$$(2\pi\sigma^2)^{-2} \exp\left\{-\frac{s_2 - 2\theta s_1 + 4\theta^2}{2\sigma^2}\right\}.$$

It depends on the observations through (s_1, s_2), which T retains. This is the factorization theorem applied to a normal sample (Casella and Berger, 2002).

Ordinary least squares under M_1 gives

$$\hat{b} = \frac{k}{5}, \qquad \hat{\theta} = \frac{s_1}{4} - \frac{3k}{10}, \tag{5}$$

and the residual sums of squares are

$$\mathrm{RSS}_0 = s_2 - \frac{s_1^2}{4}, \qquad \mathrm{RSS}_1 = \mathrm{RSS}_0 - \frac{k^2}{5}.$$

The ordinary two-sided slope test uses

$$F(Y) = \frac{k^2/5}{\mathrm{RSS}_1/2}, \tag{6}$$

with null distribution $F_{1,2}$ under the Gaussian assumptions (Casella and Berger, 2002). The summary determines this statistic. Define $F = +\infty$ when the denominator is zero and the numerator is positive, and $F = 0$ when both are zero. These exceptional events have probability zero at every state with $\sigma > 0$; using the same conventions on both sides gives pointwise preservation everywhere.

This summary corresponds to retaining the working-model fit and the two-sided test output. Write $r = s_2 - s_1^2/4$. For finite F in equation (6), $k^2 = 5rF/(F+2)$; if $F = +\infty$, $k^2 = 5r$. Conversely, T determines F. Keeping (s_1, s_2, F) therefore retains exactly the same information as T under these conventions. These are exact outputs; a rounded statistic or a rejection decision alone would define a different archive. The example concerns a choice of retained outputs. Both (s_1, s_2, k^2) and the signed summary (s_1, s_2, k) have three real coordinates, so this comparison establishes no storage advantage for the unsigned archive. The unsigned archive omits the sign of nonzero k.

For every threshold c and state η of M_1,

$$\Pr_\eta\{F(Y) > c\} = \Pr_\eta\{F(T(Y)) > c\},$$

where $F(T(Y))$ denotes evaluation by equation (6). This equality includes both null rejection probabilities and power at alternatives. It does not assert that the test has adequate power for every purpose.

To prove the estimation bound, reverse the observation order:

$$JY = (Y_4, Y_3, Y_2, Y_1).$$

Reversal preserves s_1 and s_2 and changes k to $-k$, so $T(JY) = T(Y)$. Since the error vector is exchangeable, the law of JY at (θ, b, σ) equals the law of Y at $(\theta + 3b, -b, \sigma)$. In particular, the two states

$$\eta_- = \left(-\frac{6}{5}, \frac{4}{5}, \sigma\right), \qquad \eta_+ = \left(\frac{6}{5}, -\frac{4}{5}, \sigma\right) \tag{7}$$

have identical summary laws for every common $\sigma > 0$.

Lemma 1 applies with $g = J$. Successful outputs at η_- lie in $[-17/10, -7/10]$; successful outputs at η_+ lie in $[7/10, 17/10]$. The disjoint intervals give the bound for every reduced-data estimator, including a randomized estimator. Choosing any $0 < \sigma \leq 1/5$ puts both states in $\mathcal{G}$, proving $R_T^* \geq 1/2$. Since the signed estimator in item 4 has failure probability below $1/2$ on $\mathcal{G}$, T cannot be sufficient under M_1: a sufficient statistic permits a randomized rule matching any full-data estimator's risk (Casella and Berger, 2002). The lower bound is therefore also a proof of insufficiency under the expanded model.

For the unreduced observations, the full-rank Gaussian regression formula gives

$$\hat{\theta} - \theta \sim N\left(0, \sigma^2\left(\frac{1}{4} + \frac{(3/2)^2}{5}\right)\right) = N(0, 7\sigma^2/10)$$

(Casella and Berger, 2002). The failure probability increases with σ and is independent of θ and b. Its supremum on $\mathcal{G}$ occurs at $\sigma = 1/5$, giving the displayed upper bound. Since an infimum over estimators cannot exceed the risk of this estimator, the minimax-risk difference has the claimed lower bound. Neither minimax risk is claimed to equal its bound.

Finally, T together with the sign of k determines signed k. If $c = \sum_i t_i Y_i$, then $c = k + (3/2)s_1$. The expanded-model likelihood depends on the data through

$$\mathrm{RSS}(\theta, b) = s_2 - 2\theta s_1 - 2bc + 4\theta^2 + 12\theta b + 14b^2.$$

The factorization theorem makes (s_1, s_2, k) sufficient for (θ, b, σ^2) under M_1 (Casella and Berger, 2002). It also determines the estimator in equation (5), whose success probability exceeds 0.95 throughout $\mathcal{G}$. □

The lower bound is nearly attained. Put $R = 2[1 - \Phi((1/2)/((1/5)\sqrt{7/10}))]$. The T-based estimator

$$\widetilde{\theta} = \frac{s_1}{4} - \frac{3}{2}\frac{|k|}{5}\xi,$$

with ξ an independent fair sign, gives

$$\frac{1}{2} \leq R_T^* \leq \frac{1}{2} + \frac{R}{2} \approx 0.501404.$$

Indeed, replacing $|k|\xi$ by a fair mixture of k and $-k$ leaves its conditional distribution unchanged. With $v^2 = 7\sigma^2/10$, the error law is the equal mixture of $N(0, v^2)$ and $N(3b, v^2)$. The centered component has failure probability at most R, and the other component has failure probability at most one. This proves the upper bound without identifying the exact minimax risk.

Changing the reliability criterion

The same archive can meet a different reliability requirement without recovering the lost direction. Fix any $0 < \sigma_0 \le 1/5$ and restrict attention to the two states in equation (7) at that scale. Give η_- prior probability p and η_+ probability $1 - p$.

Corollary 11. *On this two-state subexperiment, the greatest prior-weighted success probability at tolerance* $1/2$ *is* $\max(p, 1-p)$. *The greatest minimum success probability over the two states is* $1/2$.

Proof. Every permitted estimator has a common output law Q. Its two success probabilities s_-, s_+ are probabilities of the disjoint intervals in the preceding proof, so $s_- + s_+ \le 1$. Therefore

$$ps_- + (1-p)s_+ \le \max(p, 1-p)(s_- + s_+) \le \max(p, 1-p).$$

Constantly issuing the more probable target attains this bound. The inequality $\min(s_-, s_+) \le 1/2$ is attained by independently choosing the two targets with equal probability. □

With $p = 0.99$, always reporting $-6/5$ achieves prior-average success 0.99 and fails with probability one at η_+. It meets a 0.95 average-success requirement while the uniform requirement remains impossible. Because the two archive laws coincide, observing T leaves these prior state weights unchanged. The improvement in the adequacy verdict comes from weighting the task differently. No direction is learned. It applies to this specified two-state prior and gives no average guarantee for an unspecified prior over $\mathcal{G}$. Corollary 9 gives the general form: at these two states, any rule with prior-weighted success at least 0.95 under weights 0.99 and 0.01 has success at most $2/49$ at η_+.

Changing the accuracy requirement

A coarser task succeeds without the sign. Under M_1,

$$\bar{Y} - \theta \sim N(3b/2, \sigma^2/4).$$

Throughout $\mathcal{G}$, its mean has absolute value at most $6/5$ and its standard deviation is at most $1/10$. Its probability of error at most $3/2$ is minimized at those bounds and equals

$$\Phi(3) - \Phi(-27) \approx 0.998650.$$

For a symmetric interval with radius larger than the absolute mean, its normal probability decreases with the absolute mean and the standard deviation. This verifies the stated minimum. The same reduced arrangement meets the coarse task and fails the finer one.

5 When the lost direction matters

The general design identifies which targets the unsigned archive leaves ambiguous. Its proof uses elementary linear algebra and invariance of centered spherical Gaussian noise under orthogonal transformations.

Drift changes the expected reading between calibration and observation. The time separating calibration from the center of the observation design determines the effect on the target. To make that dependence explicit, suppose $n > 2$ measurements obey

$$Y_i = \theta + b(t_i - \tau) + \epsilon_i, \qquad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

where calibration fixes zero offset at the known time τ. Put

$$d = \bar{t} - \tau, \quad v_i = t_i - \bar{t}, \quad D = \sum_i v_i^2 > 0, \quad k = \sum_i v_i Y_i.$$

Let $T = (\sum_i Y_i, \sum_i Y_i^2, k^2)$ and let the admissible class be $\theta \in \mathbb{R}$, $|b| \le B$, $0 < \sigma \le s$, with $B \ge 0$ and $s > 0$. For this general design, let $a > 0$ be an arbitrary tolerance.

The mean reading is centered at $\theta + bd$. When $d \neq 0$ and both nonzero drift orientations remain admissible, that displacement can have either sign. The largest magnitude permitted by the drift bound is $B|d|$. The next proposition tests whether that uncertainty in direction can be removed using the unsigned archive and whether retaining the sign yields sufficient accuracy.

Proposition 12. *The summary retains the stable-model likelihood and every threshold decision of the ordinary two-sided slope test. Two admissible states have the same T-law exactly when they have the same*

$$(\alpha, |b|, \sigma), \qquad \alpha = \theta + bd.$$

There exist admissible states with the same T-law and calibrated targets separated by more than $2a$ if and only if $a < B|d|$. When this condition holds, every T-based estimator has worst-case probability of error greater than a at least one-half. Retaining signed k permits an estimator with worst-case failure probability

$$2\left[1 - \Phi\left(\frac{a}{s\sqrt{1/n + d^2/D}}\right)\right]. \tag{8}$$

When $d = 0$, the sample mean already attains this bound using T alone.

Proof. The stable likelihood depends on $s_1 = \sum_i Y_i$ and $s_2 = \sum_i Y_i^2$. The full-model residual sum of squares is $s_2 - s_1^2/n - k^2/D$, so T determines the slope statistic

$$\frac{k^2/D}{(s_2 - s_1^2/n - k^2/D)/(n-2)}.$$

Use the same zero-denominator conventions as before. To obtain equal retained-data laws, define the orthogonal reflection

$$H = I - 2vv^{\mathsf{T}}/D.$$

It fixes the constant vector, sends v to $-v$, preserves the squared norm, and gives $T(HY) = T(Y)$. Gaussian noise with covariance $\sigma^2 I$ is invariant under H. Since the mean is $(\theta + bd)\mathbf{1} + bv$, reflection sends the state (θ, b, σ) to $(\theta + 2bd, -b, \sigma)$. The states $(-Bd, B, \sigma)$ and $(Bd, -B, \sigma)$ have the same summary law and targets separated by $2B|d|$. Lemma 1 with $g = H$ proves the lower bound when $a < B|d|$.

For the exact description of the law classes, the archive is equivalent to $(\bar{Y}, k^2, W)$, where

$$W = s_2 - s_1^2/n - k^2/D.$$

The variables $\bar{Y}$, k, and W are independent, with laws $N(\alpha, \sigma^2/n)$, $N(Db, \sigma^2 D)$, and $\sigma^2 \chi^2_{n-2}$. These follow by projecting the spherical Gaussian vector onto the constant vector, the centered design vector, and their orthogonal complement. The law of T therefore depends only on $(\alpha, |b|, \sigma)$. Conversely, the law of $\bar{Y}$ identifies α and σ, and $\mathbb{E}[k^2] = D^2 b^2 + \sigma^2 D$ identifies $|b|$. At fixed $(\alpha, |b|, \sigma)$, the possible calibrated targets are $\alpha - |b|d$ and $\alpha + |b|d$. Their separation is at most $2B|d|$, and the boundary states attain that value. This proves the asserted equivalence for a separated equal-law pair.

Table 1: What the summaries support for the original four-observation design and class $\mathcal{G}$. The target $\alpha = \theta + 3b/2$ is the expected mean-time reading. Rows 2 to 4 report worst cases over $\mathcal{G}$, attained at $\sigma = 1/5$ and, in row 3, at $|b| = 4/5$.

Information	Target	Tolerance	Verified probability statement
$T = (s_1, s_2, k^2)$	θ	$1/2$	Every estimator has worst-case failure probability at least $1/2$.
(s_1, s_2, k)	θ	$1/2$	The regression estimator has worst-case failure probability approximately 0.00280744.
T	θ	$3/2$	The sample mean has success probability at least $\Phi(3) - \Phi(-27) \approx 0.998650$.
T	α	$1/2$	The sample mean has worst-case failure probability $2[1 - \Phi(5)] < 0.000000574$.

The signed estimator is $\hat{\theta} = \bar{Y} - dk/D$. Its error is centered normal with variance $\sigma^2(1/n + d^2/D)$, since the mean-noise component and the centered-design component have zero covariance. Its failure probability is largest at $\sigma = s$, giving equation (8). At $d = 0$, the estimator reduces to $\bar{Y}$, which T retains. □

The condition $a < B|d|$ exactly characterizes this equal-law obstruction. It does not characterize all failures of accurate estimation: when $a \geq B|d|$, sampling error can still matter. If $a < B|d|$ and equation (8) is below the permitted failure probability $0 < \gamma < 1/2$, the unsigned archive fails a task that the signed archive achieves. The four-observation design has $d = 3/2$, $D = 5$, and $B|d| = 6/5 > a = 1/2$.

A physically centered design $t = (-3/2, -1/2, 1/2, 3/2)$ with calibration at $\tau = 0$ has $d = 0$ and the same D. Its retained mean has worst-case failure probability $2[1 - \Phi(5)]$, below 0.000000574 at the same a and $s = 1/5$. This is a different measurement arrangement: observations occur on both sides of the calibration time. Its feasibility for the opening laboratory, which measures after calibration, has not been established.

For the original observations, the retained mean achieves the same bound for the different target $\alpha = \theta + bd$, the expected reading at the mean observation time. The sign is unnecessary for that target. Merely changing coordinate labels leaves the calibrated target $\theta = \alpha - bd$ unchanged. Estimating α instead of θ changes the question answered; that change is justified only if α serves the scientific inquiry.

More generally, let $q_c = \theta + cb$ denote the expected instrument reading at time $\tau + c$. On an equal-law class it takes the two values $\alpha \pm |b|(c - d)$. A target is identifiable from the retained experiment if it is constant whenever the retained-data laws are equal. With $B > 0$, this linear target is identifiable exactly when $c = d$. Its largest ambiguity is $2B|c - d|$, so the same lemma supplies a failure certificate when $a < B|c - d|$. Identification concerns the map from states to probability laws. It does not say that one realized summary determines the target, or that a given error probability is attainable. The explicit normal-risk calculations address that further question.

Each response needs its own performance argument. The signed summary is sufficient for the expanded Gaussian model and preserves more than this estimation task requires; the centered case meets the task with the unsigned summary. A zero target contrast only removes the displayed obstruction. It does not by itself control sampling error or establish accuracy. Table 1 collects the checked conclusions for the original observations.

6 A discrimination aim

This section applies the certificates and obstruction to a hypothetical inquiry built on the experiment of Section 4. Every premise is stipulated; no actual thermal study, calibration record, budget, or decision is asserted.

Premises. A laboratory measures a stable equilibrium temperature deviation θ, in kelvin, from a controlled reference at the times $t = (0, 1, 2, 3)$ after an exact zero-offset calibration. Drift satisfies $|b| \le 4/5$ kelvin per time unit, and the noise scale satisfies $0 < \sigma \le 1/5$ kelvin, so the parameter class is $\mathcal{G}$ of equation (2). The laboratory's measurement protocol, written for several uses of the instrument, requires estimation of θ within $1/5$ with probability at least 0.95 throughout $\mathcal{G}$. Two predeclared mechanistic models remain live: H_- predicts $\theta = -6/5$, and H_+ predicts $\theta = 6/5$. The core aim A is to identify which model holds, and the inquiry requires the issued label to be correct with probability at least 0.95 at every state of

$$\mathcal{G}_A = \{(\theta, b, \sigma) : \theta \in \{-6/5, 6/5\},\ |b| \le 4/5,\ 0 < \sigma \le 1/5\} \subset \mathcal{G}.$$

The inquiry requires the same protection at both live models. The team can hand over either the unsigned archive $T = (s_1, s_2, k^2)$ or T together with the observed sign of k, and the signed handoff is stipulated to be feasible. If a raw log were accessible to the later analyst, its sign information would belong to the full resource assessment; the negative verdicts below concern T alone.

Contributory tolerances. Let S_a denote estimation of θ within a at level 0.95. On $\mathcal{G}_A$ the success intervals at the two predictions are $[-6/5 - a, -6/5 + a]$ and $[6/5 - a, 6/5 + a]$, which are disjoint exactly when $a < 6/5$. So S_a is contributory to A exactly when $a < 6/5$, and the reading map of Lemma 3 is the sign of the report. At $a = 6/5$ the report 0 meets S_a under both models.

A specification that is not contributory can still supply the graded certificate. Take $a = 13/10$ and the signed archive on $\mathcal{G}_A$. The success intervals $[-5/2, 1/10]$ and $[-1/10, 5/2]$ overlap on $[-1/10, 1/10]$. The regression estimator of equation (5) has error law $N(0, 7\sigma^2/10)$, with standard deviation at most $(1/5)\sqrt{7/10}$. At $\theta = -6/5$ it fails to be a decisive success only when the error is at least $11/10$ or below $-13/10$; this event has probability below 3×10^{-11}, and the positive-target case is symmetric. It certifies the required 0.95 answer reliability.

The certificate is unnecessary even for this estimator. With $a = 3$, the success intervals are $[-21/5, 9/5]$ and $[-9/5, 21/5]$. At $\theta = -6/5$, decisive success requires a report in $[-21/5, -9/5)$, or error in $[-3, -3/5)$. Its probability is below $1/2$ at every permitted noise level and approximately 0.00016810 at $\sigma = 1/5$. Yet the report's sign identifies the true model with worst-case error approximately 3.71×10^{-13} on $\mathcal{G}_A$. The estimator meets S_3 throughout $\mathcal{G}$ with worst-case failure approximately 7.08×10^{-72}. The report retains discriminating information under the known procedure despite compatibility with both acceptance intervals. With T alone, Proposition 5, item 1, rules out every decoder at the witness pair, for both tolerances and every other repaired report.

The protocol tolerance fails for every procedure. Consider the two states

$$\eta_1 = (0, 0, 1/5), \qquad \eta_2 = (21/50, -9/50, 1/5)$$

of $\mathcal{G}$, whose calibrated targets are separated by $21/50 > 2/5$. The mean vectors of the complete observations differ by $\Delta = (21/50)\mathbf{1} - (9/50)t = (21/50, 6/25, 3/50, -3/25)$, with $\|\Delta\|^2 = 63/250$. Lemma 2 gives, for every estimator using Y and therefore for every estimator using T,

$$\max\{\text{failure at } \eta_1,\ \text{failure at } \eta_2\} \ \ge\ \tfrac{1}{2}\Big\{1 - \big[2\Phi\big(\tfrac{5}{2}\sqrt{63/250}\big) - 1\big]\Big\} \ =\ 1 - \Phi(1.25499\ldots) \ \approx\ 0.104741.$$

The slope shift $-9/50 = -\frac{3}{7} \cdot \frac{21}{50}$ minimizes $\|\Delta\|$ for the given intercept separation: $\|\Delta\theta\,\mathbf{1} + \Delta b\, t\|^2 = 4\Delta\theta^2 + 12\Delta\theta\Delta b + 14\Delta b^2$ is minimized at $\Delta b = -3\Delta\theta/7$ with value $10\Delta\theta^2/7$. The protocol

specification $S_{1/5}$ therefore fails on $\mathcal{G}$ for every procedure, including those with the complete record. The regression estimator of equation (5) has worst-case failure $2[1-\Phi(\sqrt{10/7})] \approx 0.231998$ at that tolerance. The two witness states lie in $\mathcal{G}$ and outside $\mathcal{G}_A$; on $\mathcal{G}_A$ alone, the signed archive meets $S_{1/5}$ by reporting the prediction of the identified model. The failure concerns the protocol's class. This paper does not determine the smallest tolerance attainable at level 0.95 on $\mathcal{G}$; Proposition 10 shows that $1/2$ is attainable with signed k.

Restrictions of the drift class. Under M_1, $\bar{Y} \sim N(\theta + 3b/2, \sigma^2/4)$, with standard deviation at most $1/10$. On $\mathcal{G}_A \cap \{b \geq 0\}$, the mean of $\bar{Y}$ lies in $[-6/5, 0]$ under H_- and in $[6/5, 12/5]$ under H_+, so the rule that issues H_+ exactly when $\bar{Y} > 3/5$ errs with probability at most $1-\Phi(6) < 10^{-9}$; the threshold lies at least six standard deviations from each range. On $\mathcal{G}_A \cap \{b \leq 0\}$, the means lie in $[-12/5, -6/5]$ and $[0, 6/5]$, and the rule that issues H_+ exactly when $\bar{Y} > -3/5$ has the same error bound. Both rules use T alone. They issue opposite labels whenever $-3/5 < \bar{Y} \leq 3/5$, an event with probability $2\Phi(6) - 1 > 0.999999998$ at either witness state $\eta_\pm$ of equation (7) with $\sigma = 1/5$. Each restriction makes the core aim attainable from the archive, and the archive's report cannot say which restriction to adopt.

Verdicts. Table 2 records the answer assessment for the repairs available to the team. Rows 2 and 3 use Proposition 10 and Lemma 1, whose target condition holds because $12/5 > 2$. Row 4 uses the witness states $\eta_\pm$, where $\bar{Y}$ has law $N(0, \sigma^2/4)$. Its sign is correct with probability exactly $1/2$, and with $\sigma = 1/5$ it lies in $[-3/10, 3/10]$, the overlap of the two success intervals of radius $3/2$, with probability $2\Phi(3) - 1 \approx 0.997300$. With γ' taken as the rule's worst-case failure probability $1 - \Phi(3) + \Phi(-27) \approx 0.0013499$ on $\mathcal{G}$, the bound $1 - 2\gamma'$ of Corollary 6 equals $2\Phi(3) - 1 - 2\Phi(-27)$ and is attained here up to $2\Phi(-27) < 10^{-100}$; at the specification level $\gamma' = 0.05$ the corollary gives 0.90. Row 5 uses $\alpha = \theta + 3b/2 = 0$ at both witness states. Row 6 uses Corollary 9, row 7 uses the restriction rules above and Lemma 7, and row 8 uses item 4 of Proposition 10 and Lemma 8: the law of $(T, \operatorname{sign} k)$ differs at η_- and η_+, since $k \sim N(5b, 5\sigma^2)$ has mean 4 at η_- and -4 at η_+.

What changes in the warranted judgment and action. Failure to reconstruct one correction leaves a search for another T-based procedure open. Proposition 5, item 1, closes the search for any report and decoder meeting the uniform answer task on $\mathcal{G}_A$. Rows 4 to 7 remain attainable for their changed specifications. Row 7 shows why attainability cannot select a restriction: the two rules issue opposite labels on almost every report at the obstructed pair. By Lemma 7, a choice from T that drops the false state η_+ with probability r when η_- holds drops the true state with that probability when η_+ holds. The stated premises supply no additional information to resolve that pair. Row 6 changes the guarantee: even justified weights leave success at η_+ at most $2/49$. Evidence for a weighted objective could justify that change, but could not recover the original uniform task. Rows 4 and 5 also fail that task directly by the equal-law bound, independently of the sufficient certificate.

Row 1 shows that the protocol, written for several uses, was stricter than any procedure could meet on $\mathcal{G}$. Its other uses need separate assessment. Rows 2 and 3 certify the answer once the sign is retained. A further repair declares $S_{1/5}$ on $\mathcal{G}_A$: it removes no live state, is contributory there, and is met with the signed archive by reporting the identified model's prediction. Where $\mathcal{G}$ must be kept, any $a \in [1/2, 6/5)$ is a checked relaxation. The warranted handoff is the observed sign, or another accessible resource shown to meet the required uniform performance throughout the live class. If no such resource is accessible and both witnesses stay live, the warranted report states that T cannot adjudicate the models at the declared standard. That conclusion remains conditional on the model, class, and information boundary.

The assessment imposes no general duty to retain every observation. A changed budget, evidence retiring a model, a justified weighted objective, or a new purpose can warrant a changed

Table 2: Repairs of the failed protocol specification in the hypothetical discrimination inquiry. "Attainable" means a checked procedure meets the stated specification or core aim at level 0.95 on the stated class, uniformly except for row 6's weighted average. Row 1 records failure for every procedure. Probabilities are worst cases unless a state or weighted average is named. Rows 6 and 7 concern label aims, whose success is issuing the correct label, so they are contributory by construction (Section 2).

Repair	Attainable	Contributory to A	Answer assessment
1. Keep $S_{1/5}$ on $\mathcal{G}$	No, for every procedure using Y (failure ≥ 0.1047)	Yes	Over-demanding for any archive; the protocol must change.
2. Relax to $S_{1/2}$	Not with T (failure $\geq 1/2$); yes with signed k (failure 0.002807)	Yes	Certifies the answer once the sign is retained.
3. Relax to S_1	Not with T (failure $\geq 1/2$); yes with signed k (failure 2.28×10^{-9})	Yes	Certifies the answer once the sign is retained.
4. Relax to $S_{3/2}$, report $\bar{Y}$	Yes with T (success $\geq$ 0.998650)	No	Fails the uniform answer task by the equal-law bound; overlap probability 0.9973 at $\eta_{\pm}$.
5. Estimate α within $1/2$	Yes with T (failure $<$ 0.000000574)	No	Fails the uniform answer task by the equal-law bound; $\alpha = 0$ at both witnesses.
6. Weights 0.99 on η_- and 0.01 on η_+, average level 0.95	Yes with T (constant label)	Yes, for labels	Changes the guarantee: success at η_+ is at most 2/49. Justified weights still fail the original uniform task.
7. Restrict $\mathcal{G}_A$ to $b \geq 0$, or to $b \leq 0$	Yes with T in either case (error $< 10^{-9}$)	Yes	Meets either restricted task. T cannot select between the equal-law witnesses; added information or changed protection needs justification.
8. Retain the sign of k with T	Yes ($S_{1/2}$ with failure 0.002807)	Yes	Meets the uniform answer task; the warranted handoff, given the stipulated feasibility.

task. Attainability alone establishes none of those reasons. Research may also continue through intermediate evidence without completing the uniform answer task.

7 A second case: pooled counts

The Gaussian archive loses a sign. The following hypothetical archive loses a split of counts, contains a nuisance parameter shared by the states it confounds, and admits a different repair. Every premise is stipulated.

Premises. Events are counted in two consecutive observation windows of equal known length $\ell = 20$ time units. Window j has event rate $\lambda_j + \beta$, where $\lambda_j \geq 0$ is the signal rate and $\beta \in [0, 1]$ is an unknown background rate common to both windows. The counts N_1, N_2 are independent, with $N_j \sim \text{Poisson}(\ell(\lambda_j + \beta))$. Two predeclared models are live: $H_\uparrow$ predicts $(\lambda_1, \lambda_2) = (1, 3)$, and $H_\downarrow$ predicts $(3, 1)$. The live class is $\mathcal{K}_P = \{(1, 3, \beta), (3, 1, \beta) : \beta \in [0, 1]\}$. The core aim A is to identify which model holds, and the inquiry requires the issued label to be correct with probability at least 0.95 at every state of $\mathcal{K}_P$. The archive retains the pooled count $N = N_1 + N_2$, the window length, and the model assumptions.

Proposition 13. *Write $\Lambda = \lambda_1 + \lambda_2 + 2\beta$ and $\psi(r) = (1 + r)\log(1 + r) - r$ for $r > -1$.*

1. *$N \sim \text{Poisson}(\ell\Lambda)$, so the pooled law depends on the state only through Λ. At every β, the states $(1, 3, \beta)$, $(3, 1, \beta)$, and $(2, 2, \beta)$ have the same pooled law. Every N-based rule for A has success probability at most $1/2$ under $H_\uparrow$ or under $H_\downarrow$, and every specification contributory to A fails under N.*

2. *The rule using (N_1, N_2) that issues $H_\uparrow$ when $N_2 > N_1$ and $H_\downarrow$ otherwise has error probability at most $\exp\{-\ell(\sqrt{3+\beta} - \sqrt{1+\beta})^2\}$ under either model, and so at most $\exp\{-20(2-\sqrt{2})^2\} < 0.00105$ on $\mathcal{K}_P$.*

3. *The report $N/(2\ell)$ lies within 1 of $\Lambda/2$ except with probability at most $\exp\{-m\,\psi(2\ell/m)\} + \exp\{-m\,\psi(-2\ell/m)\}$, where $m = \ell\Lambda$. On $\Lambda \in [4, 6]$ this bound is largest at $\Lambda = 6$, where it is below 0.00293. The target $\Lambda/2$ equals $2 + \beta$ under $H_\uparrow$, under $H_\downarrow$, and at $(2, 2, \beta)$.*

Proof. Item 1: the sum of independent Poisson variables is Poisson with the summed mean, $\ell(\lambda_1 + \beta) + \ell(\lambda_2 + \beta) = \ell\Lambda$. The three listed states share $\Lambda = 4 + 2\beta$. Proposition 5 applies with $Z = N$ and the pair $(1, 3, \beta)$, $(3, 1, \beta)$.

Item 2: under $H_\uparrow$ the rule errs when $N_1 - N_2 \geq 0$. For $u > 0$ and independent $N_1 \sim \text{Poisson}(m_1)$, $N_2 \sim \text{Poisson}(m_2)$, Markov's inequality gives $\Pr\{N_1 - N_2 \geq 0\} \leq \mathbb{E}\, e^{u(N_1 - N_2)} = \exp\{m_1(e^u - 1) + m_2(e^{-u} - 1)\}$. With $m_2 > m_1$, the choice $e^u = \sqrt{m_2/m_1}$ gives $u > 0$ and makes the exponent $2\sqrt{m_1 m_2} - m_1 - m_2 = -(\sqrt{m_2} - \sqrt{m_1})^2$. With $m_1 = \ell(1 + \beta)$ and $m_2 = \ell(3 + \beta)$, the exponent is $-\ell(\sqrt{3+\beta} - \sqrt{1+\beta})^2$. Under $H_\downarrow$ the rule errs when $N_2 - N_1 > 0$, an event contained in $\{N_2 - N_1 \geq 0\}$, and the same bound holds with the means exchanged. Since $\sqrt{3+\beta} - \sqrt{1+\beta} = 2/(\sqrt{3+\beta} + \sqrt{1+\beta})$ decreases in β, the bound is largest at $\beta = 1$, where $(2 - \sqrt{2})^2 = 6 - 4\sqrt{2}$ and $\exp\{-20(6 - 4\sqrt{2})\} \approx 0.0010459$.

Item 3: the event $|N/(2\ell) - \Lambda/2| > 1$ is $|N - m| > 2\ell$, with $m = \ell\Lambda \in [80, 120]$ and $2\ell = 40 < m$. For $N \sim \text{Poisson}(m)$ and $0 < x < m$, Markov's inequality applied to e^{uN}, whose mean is $\exp\{m(e^u - 1)\}$, gives $\Pr\{N \geq m + x\} \leq \exp\{-m\,\psi(x/m)\}$ at $e^u = 1 + x/m$, and applied to e^{-uN} gives $\Pr\{N \leq m - x\} \leq \exp\{-m\,\psi(-x/m)\}$ at $e^{-u} = 1 - x/m$. Both exponents decrease in m for fixed x, because $\frac{d}{dm}[m\,\psi(x/m)] = \log(1 + x/m) - x/m < 0$ and $\frac{d}{dm}[m\,\psi(-x/m)] = \log(1 - x/m) + x/m < 0$. The bound is therefore largest at $m = 120$, where it equals $\exp\{-120\,\psi(1/3)\} + \exp\{-120\,\psi(-1/3)\} \approx 0.002408 + 0.000519$. The last claim is arithmetic. □

Item 3 meets the deliberately permissive success-rate standard: the window-averaged rate is estimated within 1 with failure probability below 0.003. Yet Proposition 5, item 1, excludes every decoder of the pooled report for the original uniform label task. The shared target $\Lambda/2$ also makes its success intervals identical at each paired background. A prior-weighted label criterion is attainable by a constant label, with the changed statewise performance described in Corollary 9. Item 2 supplies a positive repair: the sign rule using the two window counts meets the uniform core aim with error below 0.00105.

The information loss is aggregation, so the repair retains the split of the counts. Equality of the pooled laws follows from additivity of Poisson means, with no symmetry transformation. The obstruction holds at every background level in $[0, 1]$, while the Gaussian pair with targets $\pm 6/5$ requires the extreme drift $|b| = 4/5$. In both cases the fixed archive fails the declared uniform answer task, and a checked repair changes the accessible information.

One further verdict shows the resource test applied twice. The law of (N_1, N_2) depends on the state only through $(\lambda_1+\beta, \lambda_2+\beta)$, so the split archive identifies the difference $\lambda_2-\lambda_1$ and the labels, but not the individual signal rates. On a class that admits both states, a specification of λ_1 within $a < 1/2$ is obstructed by the pair $(\lambda_1, \lambda_2, 1)$ and $(\lambda_1+1, \lambda_2+1, 0)$. Removing that obstruction needs a resource that separates the pair, such as a count from a background-only window. Separation is necessary; adequate estimation also needs a performance check throughout the declared class.

8 What the analysis adds to critical assessment

A failed correction and an archive obstruction support different judgments. Failure to reconstruct a correction excludes one method. Lemma 1 excludes every procedure for the specified uniform estimation task, and Proposition 5 excludes every decoder for the core answer. Corollary 6 describes the overlapping successes of point specifications that remain attainable. Section 5 makes the law classes explicit: $(\alpha, |b|, \sigma)$ fixes the retained law, while calibrated targets within it differ by $2|b||d|$. Limited computation, unsuccessful search, and uncertain premises warrant weaker assessments than these proofs. After any negative judgment, sufficient reading certificates can support a proposed repair. Failure of a sufficient certificate alone leaves reliable decoding open. Equal laws exclude discrimination at the pair. When laws differ, Corollary 6 and Lemma 7 quantify the remaining limits.

Sensitivity, contributory specifications, and changes of objective

Beauchemin and Staley make sensitivity target-relative and quantitative. A measurement result is useful for learning about possibilities to the extent that it is sensitive to their differences, and it is sensitive to the difference between possibilities only if its uncertainty is smaller than the difference between the predictions based on them (Beauchemin and Staley, 2025a). In their 2025 preprint, a sensitivity limitation arises when two theories' predictions differ by less than twice the reported uncertainty (Beauchemin and Staley, 2025b). The two thresholds differ by a factor of two when the uncertainty is read as the half-width of an interval, and the correspondence below uses the preprint's form. With a tolerance a in place of that uncertainty, a point specification is contributory to discriminating two predictions exactly when this limitation is absent and the predictions do not differ by exactly $2a$, the case in which the closed success intervals touch. This correspondence concerns interval membership. It does not establish a necessary condition for decoding under a known procedure, as the S_3 example shows.

Their account lets critical assessment revise tasks, resources, and aims, and permits a change in objectives with a shift in sensitivity expectations when improving the measurement process or

uncertainty estimate cannot resolve a lack of sensitivity (Beauchemin and Staley, 2025a). A contributory specification met at the required reliability certifies answering the declared core question. Under an equal-law obstruction no report from the archive can answer it uniformly, by Proposition 5. This limits one claim about what the changed objective achieves. It does not imply that the objective has no value for the original inquiry or that a response failing to complete the task begins a new inquiry.

Their examples also distinguish the justification of a change from its numerical success. They reject an ad hoc reduction of an uncertainty interval made to beat competitors, because it would not survive critical scrutiny and would not contribute to an inquiry into the uncertainty at all (Beauchemin and Staley, 2025a). A new rule applied to the same information cannot remove an existing equal-law obstruction, by Lemma 8. This observation does not establish that their example satisfies the lemma's hypothesis. They describe the ATLAS reanalysis of its 2018 W-mass data as using the same data with improved theoretical inputs, a more advanced statistical method, and an improved method for estimating an important background, in pursuit of an objective the 2018 result had not met (Beauchemin and Staley, 2025a). As they describe it, that response changes resources and tasks and keeps the core objective; it is the branch Lemma 8 leaves open. This paper does not model the ATLAS measurement and asserts nothing about it beyond their description. The certificate concerns the complementary case, in which the inputs are fixed and no separating resource is available.

The same consequence holds for uncertainty reports. Let $C(T, U)$ be any set-valued report with coverage at least 0.95 at each paired state. Under the common archive law, the union bound gives

$$\Pr\{\theta_- \in C(T,U),\ \theta_+ \in C(T,U)\} \geq 0.90,$$

and the diameter of the report is at least $12/5$ on that event. This is the set-valued counterpart of Corollary 6: with probability at least 0.90 the report contains both predictions, so it is not contributory to discriminating them. Beauchemin and Staley distinguish warranted interval attribution from a claim that the true quantity lies in the interval (Beauchemin and Staley, 2025a). The coverage condition here is an additional task specification, and the paper does not impose it on every uncertainty interval in their account.

Easier purposes, reuse, and repurposing

Parker recommends that, when adequacy for one purpose of interest is unclear, the evaluator consider a related, easier purpose that is also of interest (Parker, 2020). Reliable decoding of the original answer is one sufficient relation between purposes. It does not exhaust relatedness: identifying possibilities or supplying intermediate evidence can serve the original question without settling it. Proposition 5 excludes recovery of the uniform answer task from an obstructed archive. It leaves these broader purposes open. Rows 4 and 5 of Table 2 are attainable uses of the same measurements; they fail the declared discrimination requirement, and their further interest needs separate assessment.

Bokulich and Parker distinguish reuse, which reanalyses data to better answer a question the data already addressed, from repurposing, which serves a purpose other than the one initially envisioned. Data inadequate in a particular instance of use can still be adequate given accessible resources, such as additional data and a correction, that permit further processing (Bokulich and Parker, 2021). The certificates here assess a narrower accomplishment: answering the declared question at a specified reliability. Failing that certificate establishes neither repurposing nor the absence of useful reuse. Under an equal-law obstruction, further processing alone cannot complete the uniform answer task; broader reanalysis can remain useful. An additional resource must remove

the obstruction and meet a separate performance check. Their adequacy-given-resources definition already requires access to a coherent method; this analysis adds no quantifier to it.

Corollaries 11 and 9 distinguish retaining the same answer labels from retaining the same guarantee. Parker leaves the interpretation of reliability probabilities open (Parker, 2020). Here the uniform requirement is fixed. At an obstructed pair, an average requirement can be met while success at the less weighted live state is at most $2/49$ for the stated weights and level. Reasons about the states or error consequences can justify adopting that different requirement. They do not make the resulting performance uniform, and its attained success rate supplies no justification for the change.

The warrant remains conditional. A theorem cannot establish the calibration, operating bounds, target, or error consequences that make a task scientifically appropriate, and a proof under a poorly justified class gives an irrelevant negative result. Provenance metadata can help later users assess processing choices and suitability (Bokulich and Parker, 2021). By Lemma 8, metadata common to the paired states cannot remove their ambiguity, while metadata carrying direction information changes the resource assessment.

Uncertain information boundaries

The certificate presupposes a definite information boundary. In practice the boundary can be uncertain: a raw log may or may not be accessible to the later analyst, and an auxiliary record may or may not carry the needed distinction. The assessment handles such uncertainty in three steps.

First, nested descriptions of the boundary are ordered. Suppose one description gives the analyst the source Z and a rival description gives (Z, M). Every rule using Z is a rule using (Z, M) that ignores M, so every specification attainable under the first description is attainable under the second. An equal joint law of (Z, M) at two live states with different answers implies an equal law of Z there, so an obstruction under the second description is an obstruction under the first. The assessment therefore needs to settle the boundary only when the verdicts differ. Among nested descriptions that the evidence leaves open, if the core aim is attainable under the narrowest, the positive verdict holds under all of them, and if an obstructed pair survives under the widest, the negative verdict holds under all of them. Descriptions that are not nested need separate assessments.

Second, differing verdicts require evidence about access and performance. Lemma 8 asks first whether the resource's joint law with the archive differs at an obstructed pair. Lemma 7 prevents archive-based discrimination between those equal-law states. Separate records, such as provenance or access arrangements, can support treating a resource as accessible. A difference between joint laws is insufficient for the answer task. For example, a bit independent of the archive at each witness, with Bernoulli probabilities 0.51 and 0.49, gives joint total variation 0.02; Lemma 2 bounds the smaller correct-label probability by 0.51. Meeting level $1 - \gamma$ at a pair requires joint total variation at least $1 - 2\gamma$. Even that necessary condition leaves performance throughout the live class to be checked.

Third, until that evidence is in, the warranted report is conditional and names its condition: the archive is inadequate for the core aim unless an accessible resource meets the required performance. Neither an unconditional negative verdict nor an unconditional positive verdict is warranted. In the case of Section 6, the narrow description (T alone) and the wide one (T with the sign of k) give opposite verdicts, so the assessment turns on whether the sign is accessible; an accessible raw log that records the readings with their times would supply the sign, since the readings and times

determine k. The analysis supplies no evidence about the resource and assigns no probabilities to the competing descriptions.

The repair menu

Row 4 of Table 2 falls under Beauchemin and Staley's sensitivity condition, since $12/5 < 2(3/2)$ (Beauchemin and Staley, 2025b). Proposition 5 applies the equal-law argument to this repair menu: every change of target, tolerance, or uniform level that stays contributory remains unattainable under the fixed obstruction. Reliable answering at the original level requires resolving that obstruction; a justified change of class or guarantee instead changes what is required. These results do not limit every way of continuing research.

Beauchemin and Staley list responses to an insensitive result: re-estimating uncertainties, redesigning the measurement, changing the objective, and obtaining access to a more sensitive result (Beauchemin and Staley, 2025b). The obstruction closes re-estimation from fixed inputs for the declared uniform answer task. Redesign or additional resources can support that task after a performance check. Changing the objective can remain useful without completing it. Their expectation that inquirers adopt objectives the uncertainty does not defeat is compatible with this assessment. Lemma 2 gives a necessary quantitative condition for recovering the answer task: joint total variation at the obstructed pair must reach $1 - 2\gamma$ to meet level $1 - \gamma$.

Boyd describes the limit of reuse from the side of provenance. A result mal-adapted to a new context can sometimes yield a constraint if enough provenance information allows backtracking through the stages of processing to an earlier, adaptable product and reprocessing it; when provenance shows that a useful constraint cannot be generated in practice, little remains except generating constraints in another way (Boyd, 2018). The certificate is a formal version of the second judgment for a declared task. When no earlier product is retained and the retained product gives two live states one law, no processing of that product meets a contributory specification, and a checked resource can establish recovery of that task. The counterfactual muon trigger that Boyd discusses, from Franklin's history of the weak neutral current experiments, has the same structure at the level of design: its evidence would have been consistent both with theories positing weak neutral currents and with theories omitting them.

A documented decision at an uncertain boundary

A documented harmonization decision illustrates why access to coding information matters. Two-sample Mendelian randomization needs, for each genetic variant, its effects on the exposure and on the outcome for the same effect allele (the genetic form to which a reported effect refers), and the MR-Base platform harmonizes the two sources using knowledge of the effect alleles and, where necessary, the effect allele frequencies (Hemani et al., 2018). For variants with A/T or G/C alleles, the palindromic variants, the identity of the effect allele can be ambiguous. The effect allele frequency can indicate the reference strand, but the authors state that it may not be a reliable indicator when the frequency is close to 0.5 (Hemani et al., 2018). In their applied example on LDL cholesterol and coronary heart disease, they harmonized the data with the algorithm set to assume that all variants were coded on the forward strand, and they report that disabling that option would have excluded seven palindromic variants with allele frequencies close to 0.5 (Hemani et al., 2018).

These responses raise distinct assessment questions. Assuming forward coding requires justification about how the sources were coded. Using allele frequencies adds information whose reliability the source explicitly qualifies near 0.5. Excluding ambiguous variants changes the inputs, so the

resulting estimator needs a performance assessment on those retained. Identical palindromic allele letters alone do not establish equal joint laws of the effect estimates. This paper supplies no reporting model for those estimates and applies neither an exact-law obstruction nor a total-variation bound to the genetic case. It reports no estimate under any option and makes no recommendation about which option the authors should have taken. The cited article establishes the documented ambiguity and response; the constructed examples supply the proved probability-law comparisons.

Established statistical comparisons

The construction illustrates insufficiency under model expansion. The sufficiency principle applies within the model in which sufficiency is established; Casella and Berger explicitly warn that the normal sample's mean and variance can cease to be sufficient in another family (Casella and Berger, 2002). Here a signed summary is sufficient under the expanded model and retains more than the coarse and mean-time tasks require. Loss of sufficiency alone supplies neither the fine-task impossibility nor a failure of every later purpose.

Decision-relative information comparison is also established. Blackwell relates simulation that does not depend on the state to attainable loss vectors for finite collections of state laws and bounded convex decision problems (Blackwell, 1953). Brooks, Frankel, and Kamenica compare signals for restricted classes, including a single decision problem, and define discriminating classes through disjoint optimal-action sets. Their setup uses finite states and a fixed interior prior (Brooks et al., 2024). The contributory condition of Section 2 resembles that requirement at the level of a single aim and a uniform criterion, with closed success intervals in place of optimal-action sets; no result from their framework is invoked. The continuous-state calculation here invokes neither characterization theorem. Lemma 2 gives the standard two-point argument and Gaussian distance formula; Corollary 6 and Lemma 7 apply the same total-variation reasoning to repairs and exclusions. The Poisson bounds in Proposition 13 are standard exponential-moment bounds. Each result is proved in the text. The proposed contribution lies in the conditional repair analysis and the distinctions it makes, and the paper offers no new theory of information or adequacy.

9 Limits

The reduction deliberately discards a sign. A reader familiar with regression can identify the missing component immediately; the calculations determine which uses that omission defeats. A practitioner who already retains signed k has supplied the repair. The design calculation describes the consequences of a target and tolerance, leaving their scientific justification to the inquiry.

The certificates presuppose a declared core aim, live class, and reliability level. They neither choose these commitments nor classify every contribution to a continuing inquiry. Several core aims need separate assessments. For a specification that is not contributory, the graded certificate needs a procedure-specific calculation; failure of that certificate leaves other decoding open. The weighted calculations concern a different guarantee from the fixed uniform task. Both constructed cases stipulate their budgets, accessibility premises, and predictions. The ATLAS and harmonization descriptions remain limited to the cited accounts, with no model of either and no claim about results under alternative choices. The pooled-count case assumes independent Poisson counts and a background common to both windows; changing either premise changes the identification analysis.

The lower bound depends on keeping both paired states live. Evidence about drift orientation, a new measurement, or access to signed k can change the assessment. Any archive-based selection and subsequent answering need a joint error assessment over the original class. A prior-weighted

objective can be justified while failing the original uniform task. None of these possibilities shows that T alone meets that task as specified.

Refusal cannot improve the unconditional probability of issuing an accurate estimate at both paired states: accurate issued answers still occupy disjoint success intervals.

The test result is conditional on the selected diagnostic. A one-sided slope test uses the sign that T removes; nonlinear departures call for other checks. Gaussianity supplies the likelihood and exact performance calculations. The reversal argument itself needs only the invariance of the noise law under reversal, which independent identically distributed errors provide. These qualifications prevent an inference from this example to all diagnostics, models, or targets.

10 Conclusion

An attainable repair and recovery of a declared answer task are different achievements. Contributory specifications and sufficiently frequent decisive successes provide sufficient certificates for reliable answering. Their failure alone leaves decoding open, as the signed S_3 example demonstrates. Equal retained laws at states with different answers give a direct negative result: no report and decoder based on that archive answers correctly with probability above one-half at both states. Attainable point specifications then have overlapping successes, while weighted criteria and class restrictions change the required protection. Justified unequal weights still do not restore the original uniform guarantee. Archive-based selection cannot discriminate between an equal-law pair, though it can support narrowing elsewhere when selection and answering meet the stated error requirement. In the Gaussian example, retaining the sign supports a checked relaxation or a specification on the live model class. The pooled-count example repairs aggregation loss by retaining the window counts. Added resources require both access and performance evidence throughout the live class. These conditional results assess reliable answering within the cited accounts' broader treatments of useful purposes and continuing inquiry. They do not identify every valuable response with successful completion, and their scientific application needs justification for the aim, class, and information boundary.

Acknowledgments

AI-assisted writing. I designed and directed this paper, made the editorial decisions, and edited and verified the manuscript. OpenAI GPT-6 Astra and GPT-6.1 Sol assisted with drafting and manuscript review. I take full responsibility for the paper.

References

Beauchemin, Pierre-Hugues, and Kent W. Staley. 2025a. Making Measurement Useful: Integrating Measurement, Uncertainty, and Sensitivity. *Synthese* 206:110. https://doi.org/10.1007/s11229-025-05187-0.

Beauchemin, Pierre-Hugues, and Kent W. Staley. 2025b. How Uncertainty and Underdetermination Allow Measurement to Produce Useful Results. Author preprint consulted for the cited formulations: submitted to Elsevier, June 27, 2025. https://philsci-archive.pitt.edu/id/eprint/25918/. Published successor: *Measurement* 267 (March 2026):120330. https://doi.org/10.1016/j.measurement.2026.120330.

Blackwell, David. 1953. Equivalent Comparisons of Experiments. *Annals of Mathematical Statistics* 24(2):265–272.

Bokulich, Alisa, and Wendy Parker. 2021. Data Models, Representation and Adequacy-for-Purpose. *European Journal for Philosophy of Science* 11:31. https://doi.org/10.1007/s13194-020-00345-2.

Boyd, Nora Mills. 2018. Evidence Enriched. Author manuscript consulted: https://philsci-archive.pitt.edu/14286/. Publication: *Philosophy of Science* 85(3):403–421. https://doi.org/10.1086/697747.

Brooks, Benjamin, Alexander Frankel, and Emir Kamenica. 2024. Comparisons of Signals. *American Economic Review* 114(9):2981–3006. https://doi.org/10.1257/aer.20230430.

Casella, George, and Roger L. Berger. 2002. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury.

Hemani, Gibran, Jie Zheng, Benjamin Elsworth, Kaitlin H. Wade, Valeriia Haberland, Denis Baird, Charles Laurin, Stephen Burgess, Jack Bowden, Ryan Langdon, Vanessa Y. Tan, James Yarmolinsky, Hashem A. Shihab, Nicholas J. Timpson, David M. Evans, Caroline Relton, Richard M. Martin, George Davey Smith, Tom R. Gaunt, and Philip C. Haycock. 2018. The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome. *eLife* 7:e34408. https://doi.org/10.7554/eLife.34408.

Parker, Wendy S. 2020. Model Evaluation: An Adequacy-for-Purpose View. *Philosophy of Science* 87(3):457–477. https://doi.org/10.1086/708691.