When does a negative assay exclude a finite target count?

Sharp false-exclusion bounds for native-reference reporting rules

17 September 2026

Abstract

A negative readout need not exclude targets in the original specimen when extraction failures are shared across targets and splitting consumes material. We study a finite, well-mixed specimen with conditionally independent target paths and arbitrary dependence between two preparation recoveries. A sharp moment bound reduces the all-negative probability to four recovery states. An explicit reporting rule accepts each of m independent native-reference pairs whenever at least one reference is recovered, and issues a count-exclusion certificate only if the future assay is also negative. For balanced effective fractions a, the exact worst-case probability of a false certificate is $G_m((1-a)^K)$, where $G_m(H) = \max_{0 \le z \le 1} z^m[1-(1-H)z]$ has a closed two-branch formula; no equality of marginal recovery means is needed, and an explicit source attains the value. We then quantify four things that a reporting rule must survive. Reusing one calibration across T independent specimens has an exact familywise value with an irreducible floor $1-(1-H)^T$, so a five percent familywise level is *unreachable* for two specimens in the worked design however many references are bought. Replacing the exactly-one reference input by Poisson loading of mean λ leaves the guarantee intact for $\lambda \le 1$, has a closed sharp value $d_\lambda^m G_m((1-2a)^K)$ under an explicit rational criterion, and in general reduces to a one-dimensional concave-envelope problem; heavy loading destroys the guarantee outright. An event-specific allowance for observation-law mismatch replaces a product-coupling bound that was an order of magnitude looser. Spurious control positives admit an exact reduction to the same scalar problem, and the resulting specificity budget — a pair false-positive rate below about $1/300$ — is the binding practical constraint. Finally, imposing equal means and nonnegative covariance moves the feasibility threshold from $(1-a)^K$ to $(1-2a)^K$, which makes an otherwise impossible exclusion attainable. The central chain and the new batch, false positive and Poisson results are mechanized in Lean 4; biological transport remains an experimental premise.

1 Question and contribution

Given limited specimen volume, uncertain extraction and recovery, and shared interference, when does a negative result justify exclusion—and should the next experiment use more specimen, another extraction, another well, or a different measurement?

The relevant denominator is the original sampled specimen. Increasing the number of wells does not itself increase the amount of original material observed, and recoveries in separate preparations need not be independent. We ask a bounded version of this decision problem: how should a fixed specimen be allocated to one or two preparations, and what explicit native-reference reporting rule permits a uniform exclusion guarantee? The target is the count N in that specimen. Excluding $N \ge K$ means reporting $N < K$; it does not establish sterility, patient absence, source concentration, or a population prevalence.

Multistage detection errors and their identifiability are established subjects. Guillera-Arroita et al. [6] analyze detection errors at multiple levels and the additional observations needed to distinguish model parameters. Smart et al. [13] and Lugg et al. [10] develop cost-aware environmental-DNA sampling designs. Their survey objectives differ from the finite-specimen, worst-case reporting probability studied here. Bounds on the expectation of a convex function through its values at the vertices of a box are classical [7, 11], as is the fact that an extremal law for finitely many moment constraints needs few support points [8, 15]; our corner and envelope arguments are applications, not new convex-order principles. Binary performance-demonstration tests also explicitly trade sample size against acceptance risk [9], and multiplicity control for families of decisions has a standard vocabulary [1].

Our contribution is the complete connection from conserved target paths to a sharp allocation bound and an observable native-reference rule, followed by an exact accounting of what happens when the four premises a laboratory is most likely to break are relaxed one at a time. Concretely:

1. A sharp four-state envelope for the all-negative probability (Theorem 3.1) and the exact worst-case error of the at-least-one native-reference gate, $G_m((1-a)^K)$, with an attaining source (Theorems 4.1 and 4.2).
2. The exact familywise value when one calibration is reused for T independent specimens, its irreducible floor, and the resulting impossibility and reachability statements (Section 7).
3. Poisson-loaded native references: an exact reduction to a one-dimensional concave-envelope problem for every loading mean, a closed form under a sharp and checkable criterion, a proof that loading means at most one cannot break the original guarantee, and a counterexample showing that heavy loading does (Section 8).
4. An event-specific allowance for observation-law mismatch that is evaluated exactly on a five-point critical set, and an exact reduction of the spurious-positive problem to the same scalar maximum, which converts a vague specificity concern into a number (Section 9).
5. A sharp value under equal means and nonnegative covariance, whose floor is $(1-2a)^K$ rather than $(1-a)^K$; the restriction converts an impossible exclusion into a merely expensive one (Section 10).

The mathematical scope throughout is a prescribed family of reporting policies, not unrestricted optimal experimental design. We provide ordinary proofs, with a machine-checked core; Appendix A states exactly which claims are mechanized. The analysis is conditional on a stated experimental model. No clinical validation is claimed and no priority claim about a named open conjecture is made.

2 Material, recovery and the reporting event

2.1 A finite specimen and shared preparation states

Let N targets be uniformly located in a homogeneous specimen of volume V_0. Two disjoint effective fractions $a, b \geq 0$ satisfy $a + b \leq 1$. They combine usable original volume and downstream coverage. With available volume $V \leq V_0$, representative setup loss ℓ per preparation and full downstream coverage, balanced splitting gives

$$a = b = \frac{V/2 - \ell}{V_0}, \qquad s = \frac{V - \ell}{V_0} \tag{1}$$

for the split and single-preparation designs, respectively. Each formula is used only when its usable volume is nonnegative. Representative loss means that the discarded volume carries targets with the same allocation law as the retained volume. Selective retention or changed inhibitor concentration is not automatically represented by subtracting ℓ.

A recovery state $(X, Y) \in [0, 1]^2$ is drawn once for the specimen. Conditional on that complete state, the targets have independent allocation and retention marks. Recovery is shared by all targets allocated to a preparation; X and Y may be dependent. The same recovery law must apply at the compared volumes and at every admitted target count. Abundance-dependent recovery, aggregation and competition are outside this contract unless incorporated in a new source model.

For one target, the five terminal path masses are

$$1 - a - b, \quad a(1 - X), \quad aX, \quad b(1 - Y), \quad bY. \tag{2}$$

They represent unprocessed material, loss and detection in the first preparation, and loss and detection in the second. They are nonnegative and sum to one. Summing histories with no detection gives

$$R_N(a, b) = \mathbb{E}[(1 - aX - bY)^N]. \tag{3}$$

The expectation is taken *after* the N-target product. Taking a product of marginal expectations would incorrectly renew the preparation state for each target. For $N = 0$ the probability is one. Because the base lies in $[0, 1]$, $R_N \leq R_K$ whenever $N \geq K$.

For specificity, the formal model uses any finite mixture of real-valued recovery states. The analytic arguments also hold for any probability law on the unit square by integrating the same pointwise inequalities. The latter measure-theoretic extension is not part of the mechanized claim.

2.2 Native references and independence

Each reference pair contains one known native target in each usable preparation input, before the native recovery gate. Conditional on its state (X, Y), the two reference counts $A, B \in \{0, 1\}$ are independent Bernoulli outcomes. Thus

$$\mathbb{P}((A, B) = 00, 10, 01, 11 \mid X, Y) = ((1 - X)(1 - Y), X(1 - Y), (1 - X)Y, XY). \tag{4}$$

Pairs are independent draws from the same state population, and the future target specimen is another independent draw. This assumption concerns independent experiments, not unconditional independence within a pair. A common unmodeled batch effect across reference pairs and the specimen can invalidate it. Section 8 replaces the exactly-one input count by an explicit stochastic loading model, and Section 9 prices the failure of the observation law itself.

The references must transport to native targets at the selected volumes, matrices and target forms. Known input counts, accurate terminal counting and absence of competition are required. A soluble spike added downstream does not supply this contract. Setup-volume loss is counted in a, b, s, rather than counted a second time in reference recovery. External reference material is separate from the limited target specimen.

2.3 A prespecified complete-event guarantee

Fix $K \geq 1$, $m \geq 1$, an action and error budget α before the experiment. Let $\mathcal{A}_m$ be the event that every reference pair has at least one recovery, and let $\mathcal{N}$ be the event that the future target assay is all negative. The rule certifies $N < K$ only if $\mathcal{A}_m$ occurs, $\mathcal{N}$ occurs, and the action's proved bound is at most α. Otherwise it issues no such certificate. Failed or missing references are unresolved; invalid records are not converted to passes. The guarantee is

$$\sup_{\mathcal{L}(X,Y),\, N \geq K} \mathbb{P}(\mathcal{A}_m \cap \mathcal{N}) \leq \alpha. \tag{5}$$

This is one complete experiment, including its reference campaign and one future specimen. Blank-certificate probability is reported separately as usefulness.

3 A sharp four-state reduction

Write $p_1 = \mathbb{E}X$, $p_2 = \mathbb{E}Y$, $q = \mathbb{E}[XY]$. Feasibility implies

$$\max(0, p_1 + p_2 - 1) \leq q \leq \min(p_1, p_2), \tag{6}$$

since $XY \geq 0$, $XY \leq X, Y$, and $(1-X)(1-Y) \geq 0$.

Theorem 3.1 (Sharp moment envelope)**.** *For $N \geq K \geq 1$ and $a, b \geq 0$ with $a + b \leq 1$,*

$$\begin{aligned} R_N(a,b) \leq B_K(a,b) := 1 - p_1 - p_2 + q& \\ + (p_1 - q)(1-a)^K + (p_2 - q)(1-b)^K + q(1-a-b)^K&. \end{aligned} \tag{7}$$

At $N = K$ equality is attained by the masses

$$(c_{00}, c_{10}, c_{01}, c_{11}) = (1 - p_1 - p_2 + q,\ p_1 - q,\ p_2 - q,\ q) \tag{8}$$

on $(0,0), (1,0), (0,1), (1,1)$*. The same law attains the envelope for every allocation.*

Proof. On the unit square, $(1 - ax - by)^K$ is convex in each coordinate: for $K = 1$ it is affine, and for $K \geq 2$ its coordinate second derivatives are nonnegative. Successive endpoint interpolation in x and then y bounds it above by

$$(1-x)(1-y) + x(1-y)(1-a)^K + (1-x)y(1-b)^K + xy(1-a-b)^K.$$

Taking expectations gives (7), and count monotonicity covers $N \geq K$. The masses in (8) are nonnegative by (6), sum to one, and have moments p_1, p_2, q. At every corner the interpolation is exact. □

The integrated reference law (4) has precisely the four masses (8), even when native recovery is continuously valued. Thus paired reference outcomes measure the moments relevant to the bound without asserting that actual recoveries are binary. This is the reason the calibration experiment and the bound speak the same language.

4 The at-least-one gate and its exact error

Define

$$w = p_1 + p_2 - q = \mathbb{P}(A \vee B), \qquad q = \mathbb{P}(A = B = 1).$$

Then $0 \le w \le 1$, and independent reference pairs give $\mathbb{P}(\mathcal{A}_m) = w^m$. Independence from the future specimen yields the complete-event identity

$$\mathbb{P}(\mathcal{A}_m \cap \mathcal{N}) = w^m R_N(a, b). \tag{9}$$

For balanced $a = b \le 1/2$, set

$$H = (1-a)^K, \qquad J = (1-2a)^K. \tag{10}$$

The envelope simplifies to

$$B_K(a, a) = 1 - w(1-H) - q(H-J) \le 1 - w(1-H), \tag{11}$$

because $0 \le J \le H \le 1$. No equality of p_1, p_2 is used. Both H and J recur throughout: H is the miss probability of a source in which exactly one preparation recovers, and J that of a source in which both do.

Theorem 4.1 (Scalar maximum)**.** *For an integer $m \ge 1$ and $0 \le H \le 1$,*

$$G_m(H) := \max_{0 \le z \le 1} z^m[1 - (1-H)z] = \begin{cases} H, & H \ge 1/(m+1), \\ \dfrac{m^m}{(m+1)^{m+1}(1-H)^m}, & H < 1/(m+1). \end{cases} \tag{12}$$

The maximizer is $z_* = 1$ in the first branch and $z_* = m/[(m+1)(1-H)]$ in the second. G_m is nondecreasing in H and nonincreasing in m.

Proof. At $H = 1$ the function is z^m, with maximum one. For $H < 1$ its derivative is $z^{m-1}\{m - (m+1)(1-H)z\}$. If $H \ge 1/(m+1)$ this is nonnegative throughout $[0, 1]$. Otherwise its sign changes from nonnegative to nonpositive at z_*, which lies strictly between zero and one. Evaluating at the respective maximizers gives (12), and the branch values agree at equality. Monotonicity in H holds because G_m is a maximum of functions affine and nondecreasing in H; monotonicity in m holds because z^m is nonincreasing in m on $[0, 1]$. $\square$

Theorem 4.2 (Sharp calibrated count exclusion)**.** *Under the model of Section 2, balanced fractions $0 \le a \le 1/2$ and integers $K, m \ge 1$ satisfy*

$$\sup_{\mathcal{L}(X,Y),\, N \ge K} \mathbb{P}(\mathcal{A}_m \cap \mathcal{N}) = G_m((1-a)^K). \tag{13}$$

The supremum is attained in the finite-mixture class, without any constraint on the marginal means. At a fixed source, blank-certificate probability is w^m when the design is eligible to certify.

Proof. Equations (9)–(11) and Theorem 4.1 give the upper bound. For attainment choose $N = K$ and corner masses

$$(1 - z_*,\ z_*/2,\ z_*/2,\ 0), \tag{14}$$

an admissible law with $w = z_*$, $q = 0$ and equal means $z_*/2$. Its source miss probability is $1 - z_* + z_* H$ and its gate probability is z_*^m; their product is $G_m(H)$. At $N = 0$ all target records are negative, so (9) reduces to w^m. $\square$

Corollary 4.3 (Three pairs and two targets)**.** *For $K = 2$, $m = 3$ and balanced $a \le 1/2$, the sharp complete-event error is $(1-a)^2$. An attaining law assigns probability $1/2$ to each single-good corner.*

Proof. $H = (1-a)^2 \geq 1/4$, so the endpoint branch applies. Equivalently the bound follows from the nonnegative identity

$$H - z^3[1 - z(1-H)] = \frac{(1-z)^2(3z^2+2z+1)}{4} + (H - \tfrac{1}{4})(1-z^4). \qquad \square$$

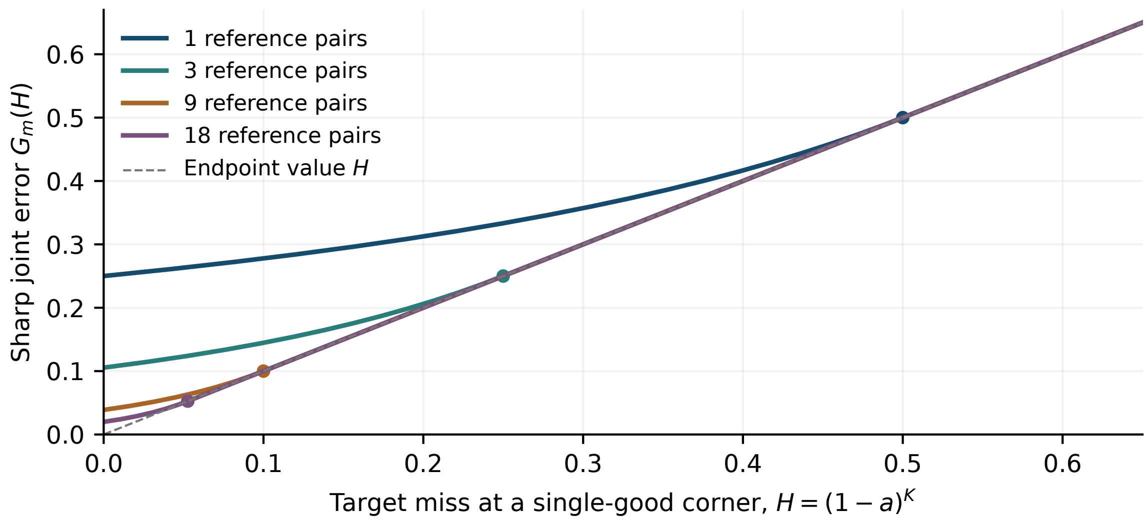

Figure 1: Exact worst-case error for the prespecified gate family. Dots mark $H = 1/(m+1)$, separating an interior adverse source from a source that passes calibration surely. The dashed line is the irreducible endpoint value.

4.1 Why the larger gate is useful

Requiring *exactly* one recovered reference per pair would give acceptance r^m, where $r = p_1 + p_2 - 2q = w - q$. That event is a subset of $\mathcal{A}_m$. For balanced allocation its envelope is $1 - r(1-H) - q(1-J) \leq 1 - r(1-H)$, and the same $q = 0$ source (14) attains $G_m(H)$. Consequently the larger gate increases blank-report availability while preserving the entire sharp worst-case curve. It does not decrease actual error pointwise: accepting more records may increase the actual probability of issuing a false certificate at a given source.

At independent Bernoulli recovery $p_1 = p_2 = 0.8$, $q = 0.64$, three pairs give $r^3 = 0.032768$ against $w^3 = 0.884736$. Double-positive reference pairs are favorable observations; rejecting them imposes an unnecessary availability cost.

5 Allocation and calibration-policy choices

5.1 Balancing is minimax at fixed total coverage

For unequal fractions define $H_a = (1-a)^K$, $H_b = (1-b)^K$ and $H_{\max} = \max(H_a, H_b)$. Because $c_{00} = 1 - w$ and $c_{10} + c_{01} + c_{11} = w$, the corner envelope is

$$c_{00} + c_{10}H_a + c_{01}H_b + c_{11}(1-a-b)^K \leq 1 - w + wH_{\max},$$

using $(1-a-b)^K \leq \min(H_a, H_b)$. The sharp complete-event error is therefore $G_m(H_{\max})$: to attain it, put mass $1 - z_*$ at $(0,0)$ and mass z_* at the single-good corner belonging to the *smaller* allocated

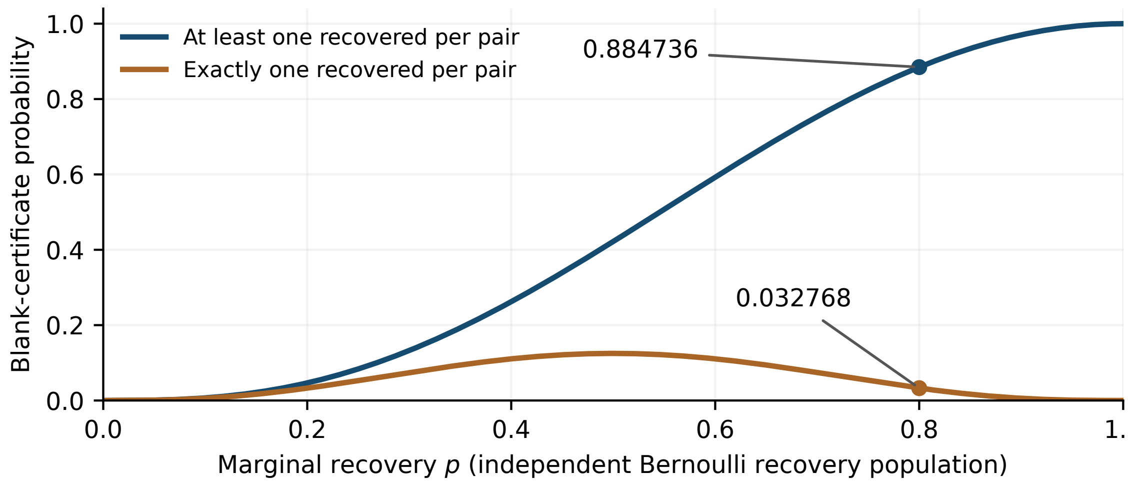

Figure 2: Blank-certificate availability with three pairs in a synthetic independent Bernoulli recovery population. The gate change uses the same six external native inputs. Equality of worst-case error does not imply equality of source-specific error.

Table 1: Finite policy menu at $K = 2$, $a = .45$, $s = .95$. Errors are uniform sharp bounds; utilities use the synthetic independent .8 recovery population only. Reference costs count native inputs and preparations.

Policy	Joint error	Blank utility	Refs.	Specimen preps.
At-least-one, 3 pairs	0.30250000	0.88473600	6	2
Discordance, 3 pairs	0.30250000	0.03276800	6	2
Single, 3 successes	0.10626374	0.51200000	3	1
Single, 6 successes	0.05751006	0.26214400	6	1

fraction, which is permitted when marginal means are unrestricted. Since G_m is nondecreasing in H and balancing maximizes the smaller fraction at fixed $a + b$, balancing minimizes the sharp error.

If equal marginal means are imposed then $c_{10} = c_{01}$ and the sharp expression instead uses $H_{\text{avg}} = (H_a + H_b)/2$, attained by the symmetric law (14). Convexity of $t \mapsto (1 - t)^K$ again favors balance. These are different source classes; the unrestricted result must not be justified by silently imposing equal means.

5.2 A fair single-preparation comparator

For a single preparation with effective fraction s, put $u = (1 - s)^K$ and $p = \mathbb{E}X$. The one-coordinate chord inequality gives $R_N(s) \le 1 - p(1 - u)$. Require all m independent native reference preparations for that protocol to recover. Its complete error and blank utility satisfy

$$\sup_{\mathcal{L}(X),\, N \ge K} \mathbb{P}(\text{references pass} \cap \mathcal{N}) = G_m(u), \qquad \mathbb{P}_0(\text{certificate}) = p^m. \tag{15}$$

The upper bound is Theorem 4.1; a Bernoulli recovery population with success probability z_* attains it. No second protocol or equal-mean premise is needed. Here m references mean m independent latent draws: six readouts nested within three shared-state pairs are not six independent references.

The discordance choice in Table 1 is removed from the efficient menu: the at-least-one rule has the same uniform error and cost and greater utility at the displayed operating point. The remaining three choices trade error, availability and cost, and none dominates the others in those objectives. With a six-reference ceiling and a blank-availability requirement $\beta = .8$ on the subclass of independent protocol recoveries each of mean at least .8, only the pair rule in this menu qualifies. A stricter error target can instead favor a single-preparation success gate, at reduced availability. Positive blank utility is impossible uniformly over the unrestricted recovery class, which includes total failure.

5.3 Known moments answer a separate question

When equal means p and cross-moment q are known, the single and balanced two-preparation bounds are

$$B_1 = 1 - p + p(1-s)^K, \qquad B_2 = 1 - 2p + q + 2(p-q)(1-a)^K + q(1-2a)^K,$$

whose difference is $pA - qC$ with $A = 1 + (1-s)^K - 2(1-a)^K$ and $C = 1 - 2(1-a)^K + (1-2a)^K$. At $K = 2$, $p = .8$, $s = .95$, $a = .45$, splitting wins exactly when $q < 106/135 \approx .785185$. For $q = .64$ its risk is .1432 against .202; for $q = .79$ it is .20395 and the single preparation wins. These known-moment risks exclude calibration uncertainty and must not be interchanged with Table 1.

5.4 What additional resources change

Under (3), another well with unchanged unique effective coverage and shared recovery leaves the risk unchanged. Assaying unused extract or changing inhibition changes a premise and can help. Another extraction changes coverage, setup loss and joint recovery. More original specimen changes feasible fractions only under an applicable sampling and volume-transport contract. A different measurement requires a new validated observation or recovery law. These results do not choose actions after an arbitrary history of previous target negatives; such conditioning requires a separate sequential analysis.

6 Calibration cost and a worked exclusion

For $H > 0$, the endpoint $G_m(H) = H$ is reached exactly when $m \geq 1/H - 1$, so the smallest permitted pair count is $\max\{1, \lceil 1/H \rceil - 1\}$. At full balanced coverage $a = 1/2$ this requires $m \geq 2^K - 1$. That exponential cost is the cost of reaching the *exact endpoint*, not of meeting every fixed error tolerance.

For fixed m and fixed $a > 0$, increasing K sends H to zero and leaves the floor

$$c_m = \frac{m^m}{(m+1)^{m+1}}, \qquad (m+1)c_m \longrightarrow e^{-1}.$$

In the interior branch $G_m(H) = c_m(1-H)^{-m}$, which for fixed m equals $c_m[1 + mH + O(H^2)]$; no uniform expansion in growing m is asserted. At $K = 2$ we have $H \geq 1/4$, so this balanced gate family cannot provide a universal five-percent certificate regardless of the number of reference pairs. Section 10 shows that this particular impossibility is a property of the unrestricted source class and not of the design.

Table 2: Balanced $a = .45$ at error budget .05. Utility assumes independent Bernoulli .8 recoveries, hence $w = .96$. Each accepted reference pair needs at least one recovery. "Infeasible" refers to this policy family and this source class.

Null $N \geq K$	$H = .55^K$	Min. pairs	Joint bound	Blank utility	Refs.
2	0.30250000	infeasible	—	—	—
5	0.05032844	infeasible	—	—	—
6	0.02768064	9	0.04987721	0.69253400	18
8	0.00837339	8	0.04631823	0.72138958	16
10	0.00253295	7	0.04996622	0.75144748	14

6.1 A five-percent example with a higher target threshold

Consider a synthetic $V_0 = V = 1\,\mathrm{mL}$ specimen, representative loss $\ell = .05\,\mathrm{mL}$ per preparation and full downstream coverage. Two inputs of $.5\,\mathrm{mL}$ leave $.45\,\mathrm{mL}$ usable each, so $a = .45$. The target unit is one native analyte unit obeying the independent-marking model, not an organism or clinical case by implication. Table 2 uses exact rational values behind the rounded displays.

For $K = 6$, nine pairs suffice and eight do not: the reporting bound is $G_9((11/20)^6) \approx .04987721425$, and minimality follows because $z^m[1-(1-H)z]$, hence its maximum, is nonincreasing in m on $0 \leq z \leq 1$. Passing the reference gate and obtaining an all-negative specimen result permits the report "fewer than six target units in the original specimen." It does not establish zero targets. The cost is 18 external native inputs, 18 reference preparations and readouts, nine independent paired latent draws, and two specimen preparations and readouts. Reference-input volumes must match the preparation protocol; in the illustrated split architecture this is $18 \times .5\,\mathrm{mL}$ of external matched input matrix, distinct from the $1\,\mathrm{mL}$ target specimen.

The blank-certificate probability at the synthetic operating point is $.96^9 \approx .692534$. For nonblank $N < 6$, the specified Bernoulli population gives

$$\mathbb{P}_N(\text{certificate}) = .96^9\left[.04 + .32(.55)^N + .64(.1)^N\right], \tag{16}$$

which falls from .692534 at $N = 0$ to .038859 at $N = 5$. This quantity, rather than blank utility, describes low-count reporting. A failed gate, missing or invalid record, insufficient resource budget or positive target result gives no exclusion certificate. The design and threshold are fixed before calibration; the guarantee does not cover searching many gates and reporting whichever happens to pass.

The margin of this example below .05 is only .000122786. The three sections that follow ask what that margin survives.

7 Reusing one calibration across a batch

7.1 Three questions that must be separated

A laboratory that has paid for m reference pairs will want to use them for more than one specimen. Three distinct questions arise, and conflating them is the main way a correct per-experiment guarantee becomes an incorrect batch claim.

1. *Per-specimen unconditional error.* Reusing a gate does not by itself invalidate the marginal bound (13) for each eligible independent future specimen from the stated population.

2. *Expected count of false certificates.* If T_0 specimens satisfy $N_i \geq K$, then the expected number of false certificates is at most $T_0\, G_m(H)$ by linearity, even though the reports share a gate.

3. *Familywise error.* The probability of *at least one* false certificate in the batch is a different quantity and is not $G_m(H)$.

So the slogan that calibration cannot be amortized is too absolute. It can be amortized for a specified marginal claim, with shared-report dependence acknowledged; it cannot be amortized while silently retaining the original familywise guarantee.

7.2 The exact familywise value

Assume one common gate, the same balanced design for every specimen, and independent future latent draws and target paths, independent of the references. Fix the null counts rather than letting batch composition depend on calibration.

Theorem 7.1 (Sharp familywise error on reuse)**.** *For $T \geq 1$ specimens all satisfying $N_i \geq K$, with $H = (1-a)^K$,*

$$\sup_{\mathcal{L}(X,Y),\, N_i \geq K} \mathbb{P}(\mathcal{A}_m \cap \{\textit{some specimen reads all negative}\}) = F_{m,T}(H) := \max_{0 \leq w \leq 1} w^m \big[1-(1-H)^T w^T\big]. \tag{17}$$

For $0 \leq H < 1$, writing $c = (1-H)^T$,

$$F_{m,T}(H) = \begin{cases} 1-c, & c \leq m/(m+T), \\ \dfrac{T}{m+T}\left[\dfrac{m}{(m+T)c}\right]^{m/T}, & c > m/(m+T), \end{cases} \tag{18}$$

and $F_{m,T}(1) = 1$. At $T = 1$ this is $G_m(H)$.

Proof. Condition on the population law. The T specimens are independent, so their all-negative probabilities R_i satisfy $R_i \leq 1-(1-H)w$ by (11) and Theorem 3.1; hence $1 - R_i \geq (1-H)w$ and the probability that at least one reads all negative is at most $1 - [(1-H)w]^T$. Multiplying by the independent gate probability w^m gives the bound (17). The corner source (14) with z_* replaced by w and every $N_i = K$ makes both steps equalities and is admissible for every $w \in [0,1]$, so the supremum is attained. For the closed form, the derivative of $w^m(1-cw^T)$ is $w^{m-1}[m-(m+T)cw^T]$, whose unique positive zero is $w_* = [m/((m+T)c)]^{1/T}$; the two branches correspond to $w_* \geq 1$ and $w_* < 1$, and in the interior case the value is $w_*^m(1 - m/(m+T))$. □

7.3 An impossibility floor, and what remains reachable

Taking $w = 1$ in (17) gives a bound that no reference budget can beat.

Corollary 7.2 (Irreducible familywise floor)**.** *For every $m \geq 1$, $F_{m,T}(H) \geq 1-(1-H)^T$, and $F_{m,T}(H) \downarrow 1-(1-H)^T$ as $m \to \infty$. Consequently a familywise level α is reachable within this architecture if and only if $\alpha \geq 1-(1-H)^T$.*

Proof. The lower bound is the value of the objective at $w = 1$, attained by the corner source with $z_* = 1$: that source passes every reference pair surely and each specimen then misses with probability exactly H. For the limit, note that $F_{m,T}$ is nonincreasing in m (as w^m is), and for $m \geq T(1-c)/c$ the first branch of (18) applies and gives exactly $1-c$. □

Table 3: Reusing one nine-pair calibration at $K = 6$, $a = .45$. The familywise column evaluates the closed form (18); the floor column is exact rational arithmetic and is what decides reachability.

Specimens T	Sharp familywise error, $m = 9$	Floor $1 - (1 - H)^T$	Level .05 reachable?
1	0.04987721	0.02768064	yes
2	0.09488141	0.05459506	no
5	0.20757145	0.13095020	no
10	0.34586453	0.24475244	no
24	0.57519502	0.49018291	no

At $a = .45$, $K = 6$ the floor for $T = 2$ is $1 - (1 - (11/20)^6)^2 = .054595\ldots > .05$. *No number of reference pairs gives a five percent familywise guarantee for two such specimens.* This is stronger and cleaner than observing that nine pairs happen to fail, and it is the exact statement mechanized as `batch_floor_exceeds` in Appendix A.

Corollary 7.2 is also constructive, because the floor is decreasing in H and $H = (1 - a)^K$ falls quickly in K. At $a = .45$ the largest batch with a reachable five percent familywise level is $T = 1$ for $K = 6$, but $T = 6$ for $K = 8$ (needing $m = 96$ pairs), $T = 20$ for $K = 10$ ($m = 330$) and $T = 66$ for $K = 12$ ($m = 1070$). Batch reporting is therefore not forbidden; it is priced, and the price is set by the exclusion threshold rather than by the calibration budget. Figure 3 shows both regimes.

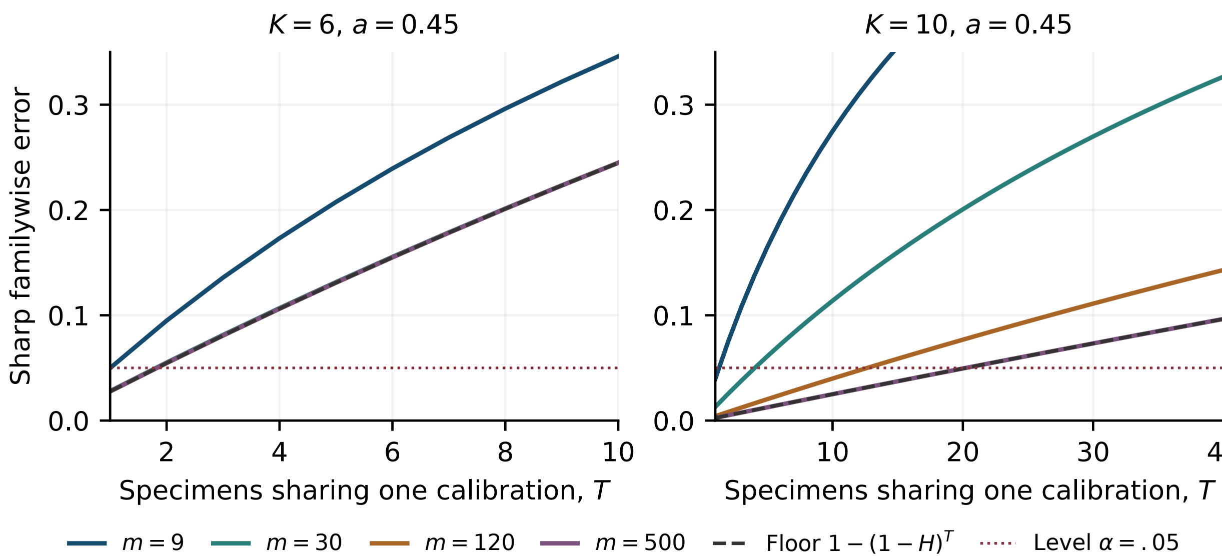

Figure 3: Sharp familywise error when one calibration is reused for T independent specimens. Curves are (18); the dashed line is the floor of Corollary 7.2, which no reference budget can cross, and the dotted line is the level .05. At $K = 6$ (left) the floor already exceeds the level at $T = 2$; at $K = 10$ (right) batches of twenty remain reachable.

7.4 False discovery rate, and repeated calibration attempts

Treating a false exclusion as a false discovery does not by itself escape Corollary 7.2. Let D be the number of exclusion reports and V the number of them whose specimens actually satisfy $N_i \geq K$, and put $\mathrm{FDR} = \mathbb{E}[V/\max(D, 1)]$. Then $\mathrm{FDR} \leq \mathbb{P}(V > 0) \leq F_{m,T_0}(H) \leq F_{m,T}(H)$, and in an all-null batch $V = D$, so the worst-case FDR over batch compositions equals the all-null familywise value. A genuinely less conservative procedure would need per-specimen valid tests or explicitly conditioned calibration [1]; it cannot be inferred from the present all-pairs gate.

A rare gate must also not be rerun until it passes while the one-campaign error claim is retained. With at most L independent reference campaigns and one independent target specimen tested after the first passing campaign, the fixed-source error is $[1-(1-w^m)^L]R_N$ rather than $w^m R_N$. Choosing $w = L^{-1/(2m)}$ makes the first factor tend to one and the second to one, so

$$\sup_{\mathcal{L}(X,Y),\, N \geq K} \mathbb{P}(\text{false certificate}) \longrightarrow 1 \qquad (L \to \infty):$$

unlimited retries destroy the guarantee entirely rather than degrading it gracefully. Prespecify the retry budget and account for it. This matters most when low reference loading is used to trade availability against joint error, as in the next section.

8 Stochastically loaded native references

Section 2.2 assumed exactly one known native target per usable preparation input. That is an experimental burden, not a law of nature. We replace it by an explicit stochastic loading model and ask what survives.

8.1 The observation link

Suppose each reference preparation receives an independent Poisson(λ) native count, independent of its recovery state, with conditionally independent retention and no loading-induced competition. Thinning gives

$$\mathbb{P}(\text{nothing recovered in preparation } 1 \mid X) = e^{-\lambda X}, \tag{19}$$

so the pair-acceptance probability conditional on the state is $1 - e^{-\lambda(X+Y)}$ in place of $1-(1-X)(1-Y) = X+Y-XY$. Equivalently, the *gate* now sees transformed recoveries

$$U = 1 - e^{-\lambda X}, \qquad V = 1 - e^{-\lambda Y}, \tag{20}$$

since $1 - e^{-\lambda(X+Y)} = 1-(1-U)(1-V)$, while the *target* risk (3) still involves X and Y. The entire effect of loading is this mismatch between what the gate measures and what the specimen suffers. Substituting $w_\lambda = \mathbb{E}[1 - e^{-\lambda(X+Y)}]$ for w in Theorem 4.2 without proof is therefore invalid.

Note that the future specimen count is still the fixed unknown N: loading describes the reference inputs only.

8.2 Two conservative transports

Proposition 8.1 (Modest loading is safe)**.** *For $0 < \lambda \leq 1$ and balanced $a \leq 1/2$, the complete-event error under Poisson reference loading is at most $G_m((1-a)^K)$, for every source and every $N \geq K$.*

Proof. For $t \in [0,1]$ and $\lambda \leq 1$ we have $1-t \leq e^{-t} \leq e^{-\lambda t}$, so $(1-X)(1-Y) \leq e^{-\lambda X}e^{-\lambda Y} = e^{-\lambda(X+Y)}$ pointwise, i.e. $1 - e^{-\lambda(X+Y)} \leq X+Y-XY$. Taking expectations gives $w_\lambda \leq w$, and raising to the mth power and multiplying by the unchanged nonnegative target risk gives $w_\lambda^m R_N \leq w^m R_N \leq G_m((1-a)^K)$ by Theorem 4.2. □

The chain $1-e^{-\lambda t} \leq \lambda t \leq t$ does *not* prove this: it compares the Poisson acceptance with $X+Y$, not with $X+Y-XY$. The correct comparison is the exponential one above. Proposition 8.1 says the old certificate survives mean-one loading, generally conservatively; it does not say the old value is still sharp, and availability falls because references may be empty.

For arbitrary λ a rescaling is still available. With $c = \max(1,\lambda)$ we have $U \leq \min(1, \lambda X) \leq cX$ and likewise for V, so $1-(a/c)U-(a/c)V \geq 1-aX-aY \geq 0$. The pair acceptance is exactly the at-least-one acceptance of the recovery pair (U,V), so Theorem 4.2 applied to (U,V) at fractions a/c gives

$$w_\lambda^m R_N(a,a) \leq G_m\big((1-a/c)^K\big). \tag{21}$$

This is safe but weak: at $K=6$, $a=.45$, $m=9$ it degrades from .049877 at $\lambda \leq 1$ to .216676 at $\lambda = 2$ and .973302 at $\lambda = 100$.

8.3 The exact value: a concave-envelope reduction

Both the gate probability and the target risk depend on the state only through $S = X+Y \in [0,2]$, and every law of S on $[0,2]$ is realizable, for instance by $X=Y=S/2$. This makes the problem exactly one-dimensional. Write

$$g(s) = 1-e^{-\lambda s}, \qquad h(s) = (1-as)^K, \qquad d_\lambda = g(2) = 1-e^{-2\lambda}, \tag{22}$$

and let $\Gamma = \{(g(s),h(s)) : s \in [0,2]\} \subset [0,d_\lambda] \times [0,1]$, a compact planar curve. Define the *upper concave envelope*

$$\hat{h}(z) = \max\{y : (z,y) \in \operatorname{conv}\Gamma\}, \qquad z \in [0,d_\lambda]. \tag{23}$$

Theorem 8.2 (Exact Poisson-loaded value)**.** *For every $\lambda > 0$, $K \geq 1$, $m \geq 1$ and balanced $0 < a \leq 1/2$,*

$$\sup_{\mathcal{L}(X,Y),\, N \geq K} \mathbb{P}(\mathcal{A}_m^{(\lambda)} \cap \mathcal{N}) = \max_{0 \leq z \leq d_\lambda} z^m \, \hat{h}(z), \tag{24}$$

and the supremum is attained by a source supported on at most two points of the diagonal $\{X=Y\}$ with $N=K$.

Proof. Fix a law of S and put $z = \mathbb{E}[g(S)]$. The pair $(\mathbb{E}[g(S)], \mathbb{E}[h(S)])$ is the barycenter of a probability measure carried by the compact set Γ, hence lies in $\operatorname{conv}\Gamma$; therefore $\mathbb{E}[h(S)] \leq \hat{h}(z)$ by (23). Since $R_N \leq R_K = \mathbb{E}[h(S)]$ for $N \geq K$ and the gate contributes z^m, the left side of (24) is at most the right side.

Conversely let z_* maximize $z^m \hat{h}(z)$. The point $(z_*, \hat{h}(z_*))$ lies on the boundary of the compact convex set $\operatorname{conv}\Gamma \subset \mathbb{R}^2$, hence in a proper face of it, and so is a convex combination of at most two extreme points; extreme points of $\operatorname{conv}\Gamma$ belong to Γ. Writing that combination as $t(g(s_1),h(s_1)) + (1-t)(g(s_2),h(s_2))$, the law placing mass t at $(s_1/2, s_1/2)$ and $1-t$ at $(s_2/2, s_2/2)$ is an admissible finite source with $\mathbb{E}[g(S)] = z_*$ and $\mathbb{E}[h(S)] = \hat{h}(z_*)$; at $N=K$ its complete-event probability is $z_*^m \hat{h}(z_*)$. □

The envelope is completely described by the curvature of h as a function of z. Substituting $t = 1-as = 1+(a/\lambda)\log(1-z)$ into $h = t^K$ and differentiating twice gives

$$\frac{d^2h}{dz^2} = \frac{Ka}{\lambda(1-z)^2}\, t^{K-2}\, [\theta - t]\,, \qquad \theta := \frac{(K-1)a}{\lambda}, \tag{25}$$

so h is convex in z where $t \le \theta$ and concave where $t \ge \theta$. As z increases from 0 to d_λ, t decreases from 1 to $1-2a$. Hence h is concave on $[0, z_c]$ and convex on $[z_c, d_\lambda]$, with $z_c = 1 - e^{-(\lambda/a)(1-\theta)}$; the concave part is empty when $\theta \ge 1$ and the convex part is empty when $\theta \le 1-2a$. For such a concave-then-convex h, the envelope $\hat{h}$ equals h on an initial interval $[0, \tau]$ and is affine on $[\tau, d_\lambda]$, the affine piece being tangent to h at $\tau \in [0, z_c]$ and passing through $(d_\lambda, h(2))$. Thus (24) is a genuine one-dimensional computation, not a search over recovery laws.

8.4 A closed form when the envelope is a chord

The degenerate case $\tau = 0$, in which $\hat{h}$ is the single chord from $(0,1)$ to (d_λ, J), is both the most useful and the easiest to detect.

Corollary 8.3 (Sharp Poisson value)**.** *Let $K \ge 2$, $0 < a \le 1/2$, $m \ge 1$ and $\lambda > 0$, and put $d_\lambda = 1 - e^{-2\lambda}$ and $J = (1-2a)^K$. If*

$$\frac{Ka}{\lambda} \ \ge \ \frac{1-J}{d_\lambda}, \tag{26}$$

then

$$\sup_{\mathcal{L}(X,Y),\, N \ge K} \mathbb{P}(\mathcal{A}_m^{(\lambda)} \cap \mathcal{N}) = d_\lambda^m \, G_m(J), \tag{27}$$

attained at $N = K$ by the two-point law placing mass $1-u_*$ at $(0,0)$ and mass u_* at $(1,1)$, where u_* is the maximizer of Theorem 4.1 at $H = J$. Condition (26) *is exactly the condition that $\hat{h}$ be the chord, and it holds in particular whenever $\lambda \le a(K-1)$.*

Proof. Write $\psi(z) = 1 - (z/d_\lambda)(1-J) - h(z)$ for the excess of the chord over the curve, so $\psi(0) = \psi(d_\lambda) = 0$. Since h is concave then convex, h' decreases then increases, so ψ' increases then decreases and changes sign at most once from positive to negative. If $\psi'(0) \ge 0$ then ψ rises and then falls between two zeros, hence $\psi \ge 0$ throughout and the chord is a concave majorant of Γ; being affine with both endpoints on Γ it is the least one, so $\hat{h}(z) = 1 - (z/d_\lambda)(1-J)$. If $\psi'(0) < 0$ then $\psi < 0$ immediately to the right of 0 and the chord is not a majorant. One differentiation gives $h'(z) = -Kat^{K-1}/[\lambda(1-z)]$, so $h'(0) = -Ka/\lambda$ and $\psi'(0) \ge 0$ is precisely (26). When $\lambda \le a(K-1)$ we have $\theta \ge 1 \ge t$, so (25) is nonnegative, h is convex throughout and the chord is its envelope a fortiori.

Given $\hat{h}$, write $u = z/d_\lambda \in [0,1]$; the right side of (24) becomes $\max_u d_\lambda^m u^m[1 - u(1-J)] = d_\lambda^m G_m(J)$. The stated two-point law has $S \in \{0, 2\}$, so it sits at the two ends of the chord, where the envelope bound is an equality, and $\mathbb{E}[g(S)] = u_* d_\lambda$. ☐

Criterion (26) is strictly weaker than convexity. At $K = 6$, $a = .45$ convexity requires $\lambda \le 2.25$, while (26) holds up to $\lambda \approx 2.68$; at $K = 4$, $a = .3$ the two thresholds are 0.9 and ≈ 1.09. The closed form therefore covers a strictly larger range of loading means than a curvature argument alone would suggest, and it is decided by a single rational inequality.

It is also the exact range. Just past the threshold the formula ceases to be even an upper bound: at $K = 6$, $a = .45$, $m = 9$ and $\lambda = 3$, the two-point source placing mass 1/5 at $S = 23/100$ and 4/5 at $S = 2$ has complete-event probability at least .03931189, against the chord value .03788662. That witness is verified in exact rational arithmetic with certified Taylor enclosures of the exponentials, so (26) cannot be relaxed. Figure 4 shows the two quantities agreeing exactly up to $\lambda^* \approx 2.68$ and separating immediately afterwards.

Two features deserve comment. First, the relevant miss parameter is J, not H: under Poisson loading the adverse source is *perfectly positively correlated*, because concentrating targets in states

Table 4: Poisson-loaded references at $K = 6$, $a = .45$, budget .05, each row inside the sharp regime (26); the $\lambda = 5/2$ row lies outside the convexity range $\lambda \le a(K-1) = 2.25$ and is covered only by the sharper criterion. The specimen still uses two preparations. Availability is evaluated at the synthetic independent .8 recovery population only. The last row is the exactly-one-target policy of Section 6.

Mean load λ	Min. pairs	Sharp joint error	Blank availability	Ref. preps.	Expected native targets
1/2	3	0.02663943	0.14927118	6	3
1	4	0.04579126	0.32607306	8	8
2	7	0.04312931	0.49709756	14	28
9/4	7	0.04539504	0.55430150	14	31.5
5/2	7	0.04681837	0.59908132	14	35
exactly one	9	0.04987721	0.69253400	18	18

where both preparations work is now the efficient way to buy gate passage. Changing the observation law changes the adverse mechanism, and worst-case behavior must not be identified with antagonistic recoveries in general. Second, since $J \le H$ and $d_\lambda \le 1$, (27) can be far below $G_m(H)$; that is a lower joint error bought with rarer gate passage, not higher assay sensitivity.

8.5 Costs, availability and a warning about heavy loading

Table 4 reports the minimum pair counts meeting $\alpha = .05$ under (27) at $K = 6$, $a = .45$, where $J = 10^{-6}$ and criterion (26) permits $\lambda \le 2.68$. Availability uses the same synthetic independent Bernoulli .8 recovery population as before, for which $w_\lambda = .32(1 - e^{-\lambda}) + .64(1 - e^{-2\lambda})$. Rational Taylor enclosures of the exponentials certify every integer threshold in the table.

Loading buys a striking reduction in reference preparations — three pairs instead of nine at $\lambda = 1/2$ — at the cost of availability, which falls from .6925 to .1493 because most references are now empty. At $\lambda = 2$ the policy uses fourteen reference preparations and 28 expected native targets to reach availability .497. None of these dominates the others; Figure 4 displays the whole tradeoff.

Heavy loading is not harmless. At $\lambda = 100$ with deterministic $X = Y = .05$, $a = .45$, $K = 6$ and $m = 9$, the complete-event probability is $(1 - e^{-10})^9(.955)^6 \ge .758$, fifteen times the nominal .049877: many loaded targets make poor recovery look like successful calibration. This example lies outside the regime of Corollary 8.3 and shows why the loading mean must appear in the theorem rather than in a footnote.

Finally, λ is itself an assumption. A dilution from an uncertain stock, shared dispensing fluctuation, overdispersion, adsorption or competition can violate it. If λ is a fixed unknown in a justified interval inside the regime of Corollary 8.3, the worst-case bound uses its upper endpoint, since d_λ increases in λ; availability may be reported at a lower endpoint for a specified source. A shared random loading effect is a dependence model, not a substitution.

9 Observation-law mismatch and control false positives

9.1 An event-specific perturbation allowance

Let each actual reference-pair observation law be within total variation $\varepsilon_C \in [0, 1]$ of its ideal law, and the future target observation law within $\varepsilon_T \in [0, 1]$, with independence across actual experiments as well as ideal ones. Maximal coupling of each whole pair and of the future observation gives the

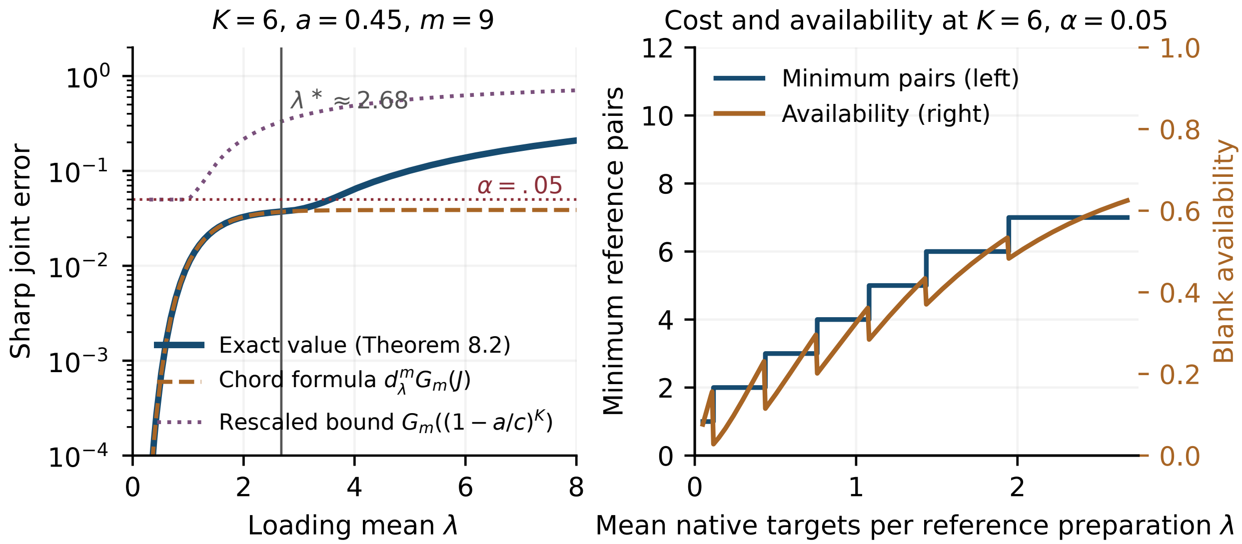

Figure 4: Left: the exact value of Theorem 8.2, evaluated from the concave envelope, against the chord formula (27) and the rescaled bound (21). The two agree to plotting accuracy exactly while criterion (26) holds, and separate immediately past λ^*; the rescaled bound is safe but far above both. Right: minimum reference pairs and the resulting blank availability against the loading mean, at $K = 6$, $a = .45$, $\alpha = .05$.

product allowance

$$\mathbb{P}_{\text{actual}}(\text{false certificate}) \le \min\{1, B + 1 - (1-\varepsilon_C)^m(1-\varepsilon_T)\} \le \min\{1, B + m\varepsilon_C + \varepsilon_T\}, \tag{28}$$

where B is any valid ideal bound. This is correct but wasteful, because it adds the whole coupling defect to a bound that was itself a maximum over sources. A bound exceeding α means only that *that bound* no longer certifies α; it does not show the actual error exceeds α.

Proposition 9.1 (Event-specific allowance)**.** *Under the same perturbation class,*

$$\mathbb{P}_{\text{actual}}(\textit{false certificate}) \le \mathcal{R}_m(H; \varepsilon_C, \varepsilon_T) := \max_{0 \le w \le 1} \left[\min(1, w + \varepsilon_C)\right]^m \min\{1, 1 - (1-H)w + \varepsilon_T\}, \tag{29}$$

and for $H < 1$ the maximum is attained in the five-point set consisting of 0, 1, $\varepsilon_T/(1-H)$, $1 - \varepsilon_C$ *and* $\frac{m(1+\varepsilon_T)}{(m+1)(1-H)} - \frac{\varepsilon_C}{m+1}$, *so* (29) *is an exact finite evaluation for rational inputs.*

Proof. Actual pair acceptance is at most $\min(1, w + \varepsilon_C)$ and the actual all-negative probability at most $\min(1, R_N + \varepsilon_T) \le \min(1, 1 - (1-H)w + \varepsilon_T)$ by (11); the two events are independent, and maximizing over the ideal source parameter w gives (29). For the critical set write $c = 1 - H$ and $A = 1 + c\varepsilon_C + \varepsilon_T$. On $\{w \le \varepsilon_T/c\}$ the second factor is capped and the objective increases in w; on $\{w \ge 1 - \varepsilon_C\}$ the first factor is capped and the objective decreases; in between the objective is $u^m(A - cu)$ with $u = w + \varepsilon_C$, which is unimodal with stationary point $u = mA/((m+1)c)$. The listed points include every endpoint, cap transition and interior stationary point. □

When both optima lie in the uncapped interior branch, (29) simplifies to

$$\mathcal{R}_m = G_m(H)\left[1 + (1-H)\varepsilon_C + \varepsilon_T\right]^{m+1}, \tag{30}$$

an inflation by a power $m+1$ of a quantity within $O(\varepsilon)$ of one, rather than the additive $m\varepsilon_C + \varepsilon_T$ of (28). The improvement is an order of magnitude at the scales that matter.

Table 5: Sensitivity of the $K = 6$ example to observation-law mismatch, with $\varepsilon_C = \varepsilon_T = \varepsilon$. The perturbations are illustrative budgets, not estimated assay errors. "Meets" refers to the event-specific column.

Pairs m	ε	Ideal $G_m(H)$	Product bound (28)	Event-specific (29)	Meets .05?
9	0	0.04987721	0.04987721	0.04987721	yes
9	1e-05	0.04987721	0.04997721	0.04988705	yes
9	0.0001	0.04987721	0.05087676	0.04997568	yes
9	0.001	0.04987721	0.05983233	0.05086973	no
10	0.001	0.04640782	0.05735298	0.04742465	yes
10	0.003	0.04640782	0.07891724	0.04951931	yes
12	0.003	0.04122913	0.07953480	0.04451556	yes

Table 5 shows the practical consequence. At nine pairs and $\varepsilon = 10^{-4}$ the product allowance gives .050877 and stops certifying, while the event-specific bound gives .049976 and still certifies; the apparent fragility at that budget was an artifact of the cruder bound. The genuine fragility appears at $\varepsilon = 10^{-3}$. Ten pairs give a less brittle design: 20 reference preparations, ideal blank availability $.96^{10} \approx .664833$, and a bound below .05 even at a 3×10^{-3} budget. Note that increasing m reduces ideal error but multiplies the exponent in (30), so m should be selected against a stated error-and-availability objective using (29), not by minimizing ideal error alone.

This remains a conditional sensitivity calculation. No experiment here establishes that a real assay meets such budgets, and pipetting coefficients of variation, recovery percentage errors and fluorescence errors are not numerically interchangeable with total variation distance. If dependence rather than marginal accuracy is the uncertainty, per-experiment budgets are insufficient and the joint law must be constrained.

9.2 Spurious positives: an exact model and a closed form

Suppose each reference channel additionally reports a spurious positive with probability η_1, η_2, independent of true recoveries and across experiments, so that the observed outcomes are $A' = A \vee C_1$ and $B' = B \vee C_2$. Writing

$$\kappa = 1 - (1 - \eta_1)(1 - \eta_2) \tag{31}$$

for the probability that a pair shows at least one spurious positive, the observed pair acceptance is

$$\widetilde{w} = 1 - (1 - \eta_1)(1 - \eta_2)(1 - w) = \kappa + (1 - \kappa)w. \tag{32}$$

If the target experiment has independent probability η_T of a spurious positive, its observed all-negative probability is $(1 - \eta_T)R_N$. For conservative control of the exclusion error one should not rely on a beneficial target artifact whose probability is unknown from below, so we take $\eta_T = 0$.

At first sight (32) creates a new optimization problem. It does not.

Proposition 9.2 (Exact reduction and closed form)**.** *Let $0 \le \kappa < 1$ and $0 \le H \le 1$, and put*

$$A_\kappa = \frac{1 - \kappa H}{1 - \kappa}, \qquad H_\kappa = \frac{H(1 - \kappa)}{1 - \kappa H}. \tag{33}$$

Then for every w and every m,

$$[\kappa + (1 - \kappa)w]^m [1 - (1 - H)w] = A_\kappa \cdot v^m [1 - (1 - H_\kappa)v], \qquad v = \kappa + (1 - \kappa)w, \tag{34}$$

Table 6: Effect of spurious control positives on the $K = 6$, $m = 9$, $a = .45$ example. κ is the probability that a reference pair shows at least one spurious positive; $\eta_T = 0$ is assumed for the target.

Pair false-positive rate κ	decimal	Sharp joint error	Meets .05?
0	0	0.04987721	yes
1/1000	0.001	0.04991332	yes
1/300	0.00333333	0.04999787	yes
1/250	0.004	0.05002210	no
1/100	0.01	0.05024174	no
1/30	0.0333333	0.05112301	no

so the sharp complete-event error under this model is at most $A_\kappa G_m(H_\kappa)$, with equality whenever the maximizer $v_* = \min\{1, m(1-\kappa H)/[(m+1)(1-H)]\}$ satisfies $v_* \geq \kappa$; otherwise it equals κ^m. In particular $\kappa = 0$ recovers Theorem 4.2, and in the interior branch

$$A_\kappa G_m(H_\kappa) = \frac{m^m(1-\kappa H)^{m+1}}{(m+1)^{m+1}(1-H)^m(1-\kappa)}. \tag{35}$$

Proof. Substituting $w = (v-\kappa)/(1-\kappa)$ gives $1-(1-H)w = [1-\kappa H-(1-H)v]/(1-\kappa) = A_\kappa[1-(1-H_\kappa)v]$, since $1-H_\kappa = (1-H)/(1-\kappa H)$; this is (34). As w ranges over $[0,1]$, v ranges over $[\kappa, 1] \subseteq [0,1]$, so the maximum is at most $A_\kappa G_m(H_\kappa)$ by Theorem 4.1, with equality when the unconstrained maximizer is admissible; the stated v_* is Theorem 4.1's maximizer after simplifying $1-H_\kappa$. When $v_* < \kappa$ the objective is decreasing on $[\kappa, 1]$ and the maximum is at $v = \kappa$, i.e. $w = 0$, where the value is κ^m. For (35), substitute $(1-H_\kappa)^m = (1-H)^m/(1-\kappa H)^m$ into the interior branch of (12) and multiply by $A_\kappa = (1-\kappa H)/(1-\kappa)$. $\square$

At total recovery failure $w = 0$ the gate still passes with probability κ^m, which is why positive reference signals without specificity controls are insufficient evidence. Quantitatively, Table 6 evaluates (35) at the worked design.

The specificity budget is severe: the nine-pair design tolerates $\kappa \leq 1/300$ and fails at $\kappa = 1/250$. Compared with the transport budget of Table 5, which tolerates $\varepsilon = 10^{-4}$ comfortably, control false positives are the binding practical constraint on this design, and they should be budgeted first.

9.3 What negative controls establish

Extraction blanks, no-template controls and matrix-matched negatives interrogate different routes to false positivity. They do not measure native recovery by themselves, they must be counted in the resource budget, and they cannot be added silently as free validation.

If n independent negative-control *pairs* show no spurious pair pass, the exact one-sided binomial upper confidence bound at failure probability δ is the Clopper–Pearson value [4]

$$\kappa_U = 1 - \delta^{1/n}. \tag{36}$$

At $n = 96$ and $\delta = .05$ this is $\kappa_U \approx .030724$, which by Table 6 is roughly a factor of nine too large to support the nine-pair design. Certifying $\kappa \leq 1/300$ at $\delta = .05$ requires $(1-1/300)^n \leq 1/20$, that is $n = 898$ clean independent control pairs. Zero observed positives do not establish a zero false-positive rate, and pairing, matrix matching and independent sampling remain assumptions. If a data-dependent bound for κ is used to choose a reporting policy, its coverage failure belongs in the procedure's error budget; plugging an estimated rate into (35) is not justified.

This is a prospective validation requirement, not a validated dataset. It is also the clearest instance of a general point: the cheap part of this design is the mathematics, and the expensive part is the evidence that its premises hold.

10 What stronger source information buys

The corner laws of Theorem 4.2 show what an unrestricted model permits. They do not assert that extreme failure is common. Excluding them requires evidence expressed as a mathematical restriction, and different restrictions buy very different amounts.

10.1 Nonnegative covariance alone buys nothing

Two distinctions matter. First, $q > 0$ is not positive correlation: nonnegative covariance means $q \geq p_1 p_2$. Second, nonnegative covariance *alone*, with unrestricted marginal means, does not improve Theorem 4.2 at all. Take $Y \equiv 0$ and X Bernoulli with mean z_*: then $q = 0 = p_1 p_2$, the covariance is zero, and the source attains $G_m(H)$ exactly. An assumption that protocols are comparable, or that recoveries are similar, therefore has no content unless it constrains the marginals.

10.2 Equal means and nonnegative covariance

Theorem 10.1 (Sharp value under a constrained source class)**.** *Restrict the source class to equal marginal means $p_1 = p_2 = p$ and nonnegative covariance $q \geq p^2$. Then for balanced $a \leq 1/2$, with H, J as in* (10)*,*

$$\sup \mathbb{P}(\mathcal{A}_m \cap \mathcal{N}) = \max_{0 \leq p \leq 1} (2p - p^2)^m \left[(1-p)^2 + 2p(1-p)H + p^2 J\right], \tag{37}$$

attained by independent Bernoulli recovery coordinates with success probability p, at $N = K$.

Proof. By (11) the envelope is $1 - w(1-H) - q(H-J)$ with $w = 2p - q$, which is nonincreasing in q at fixed w because $H \geq J$. At fixed w the constraint $q = 2p - w \geq p^2$ forces $p \geq 1 - \sqrt{1-w}$, and $q = 2p - w$ increases with p, so q is minimized at $p = 1 - \sqrt{1-w}$, where $q = p^2$ exactly. The remaining feasibility constraints (6) hold there, since $p^2 \leq p$ and $p^2 \geq 2p - 1$. Hence the extremal source in this class is the independent Bernoulli law with corner masses $((1-p)^2, p(1-p), p(1-p), p^2)$, for which $w = 2p - p^2$ and the envelope equals $(1-p)^2 + 2p(1-p)H + p^2 J$ by direct expansion. As p ranges over $[0,1]$, $w = 2p - p^2$ ranges over all of $[0,1]$, so no value of w is lost. □

10.3 The floor moves from H to J

The practical content of Theorem 10.1 is not the formula but where its floor sits.

Corollary 10.2 (Feasibility threshold)**.** *Denote by G_m^{cov} the right side of* (37)*. Then $G_m^{\mathrm{cov}} \geq J$ for every m, and $G_m^{\mathrm{cov}} \downarrow J$ as $m \to \infty$. Consequently, within this source class a level α is reachable by some finite m if and only if $\alpha > J = (1-2a)^K$, whereas in the unrestricted class of Theorem 4.2 it is reachable if and only if $\alpha \geq H = (1-a)^K$.*

Proof. The value at $p = 1$ is exactly J, giving the lower bound. For the limit, fix $\eta \in (0,1)$ and split at $p = 1 - \eta$: for $p \leq 1 - \eta$ the gate factor $(2p - p^2)^m = (1 - (1-p)^2)^m \leq (1 - \eta^2)^m \to 0$, while

Table 7: Reference pairs needed for $\alpha = .05$ at $a = .45$, comparing the unrestricted source class with equal means and nonnegative covariance. The constrained minima are certified in exact rational arithmetic from both sides.

Null $N \geq K$	$H = (1-a)^K$	$J = (1-2a)^K$	Pairs, unrestricted	Pairs, constrained	Sharp value
2	0.30250000	0.01000000	infeasible	38	0.04989975
3	0.16637500	0.00100000	infeasible	16	0.04888960
4	0.09150625	0.00010000	infeasible	11	0.04765854
5	0.05032844	0.00001000	infeasible	9	0.04757889
6	0.02768064	0.00000100	9	8	0.04822688

for $p \geq 1-\eta$ the bracket is at most $\eta^2 + 2\eta H + J$. Hence $\limsup_m G_m^{\text{cov}} \leq \eta^2 + 2\eta H + J$ for every $\eta > 0$, so the limit is J. The unrestricted statement is Theorem 4.1: $G_m(H)$ is nonincreasing in m with minimum value H, attained once $m \geq 1/H - 1$. $\square$

This converts an impossibility into a cost. At $a = .45$ and $K = 2$ we have $H = .3025$ and $J = .01$: no reference budget can certify $\alpha = .05$ over unrestricted recovery laws, but 38 pairs suffice once equal means and nonnegative covariance are assumed, with sharp value .04989975. Table 7 gives the corresponding counts for $K = 2, \ldots, 6$, and Figure 5 shows the two floors.

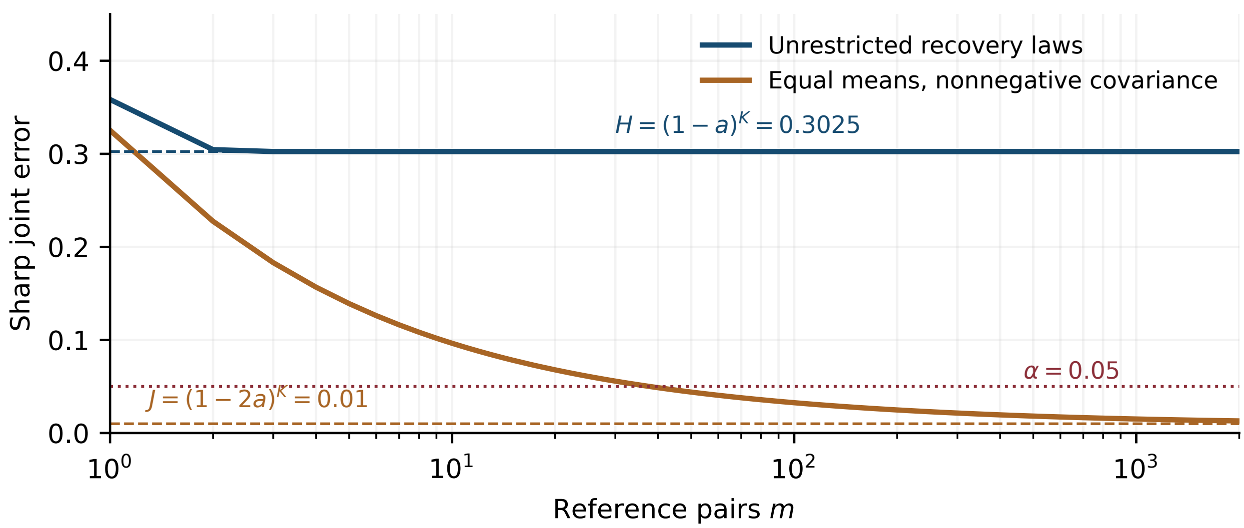

Figure 5: Sharp joint error against reference budget at $K = 2$, $a = .45$. Over unrestricted recovery laws the value is pinned at $H = .3025$ for every $m \geq 3$; under equal means and nonnegative covariance it decreases to the floor $J = .01$ and crosses $\alpha = .05$ at $m = 38$.

The interpretation is concrete. The floor H is the miss probability of a source in which exactly one preparation works; the floor J is that of a source in which both do. Assuming equal means and nonnegative covariance is precisely the assumption that a state in which one preparation systematically works while the other systematically fails is not the whole story, and the exclusion threshold that becomes reachable is governed by the total coverage $2a$ rather than the per-preparation coverage a. Whether this transports from a validation population to actual specimens is an empirical question, and Corollary 10.2 makes explicit exactly how much it is worth paying to answer it.

Corollary 8.3 cautions against the opposite error. Its sharp witness has $X = Y$, so shared all-or-nothing failure can be the limiting mechanism even under perfect positive dependence. Worst-case behavior is a property of the observation law, not a fixed picture of antagonistic recoveries.

11 Relation to detection-capability validation

Detection-capability validation addresses neighboring but distinct questions. A blank decision threshold concerns how often blank material produces an apparent positive, and a limit of detection concerns a low level detected with specified probability under a defined measurement procedure; CLSI EP17-A2 is the standing guidance [3]. The present claim instead bounds false exclusion of a *count in an original specimen*, over a source model and an explicit reference-and-target reporting event. These are complementary. A properly designed whole-process detection study can and should include extraction and matrix failures, and nothing here shows that such a study is inadequate.

The failure mode we do identify is a modeling one. Suppose a process fails completely with probability ρ and otherwise detects each of N targets independently with probability d. Its miss probability is

$$\rho + (1-\rho)(1-d)^N, \tag{38}$$

whereas an analysis that replaces this by $[1-(1-\rho)d]^N$ resamples the shared failure for each target and can understate risk by orders of magnitude at moderate N. Equation (3) is the general form of (38), and the entire paper is an accounting of the difference. This diagnoses a mismatch between a model and a process; it does not establish a defect in a particular standards-compliant study.

The reporting records needed to evaluate the present contract — sample processing amounts, extraction replicates and blanks, blank and detection limit assessment, inhibitors, positive and negative controls, technical replication, partition volume and copies per partition — are exactly those enumerated by MIQE 2.0 [2] and the digital PCR checklist [12]. Our bound complements detection-capability validation by making the original-specimen denominator, the shared recovery and the reference-based reporting event explicit. It does not replace blank or detection limit estimation, and it does not establish that the native-reference transport assumptions hold in any particular assay.

12 Interpretation and prospective validation

For a fixed recovery law and independent future specimen, whenever the gate has positive probability,

$$\mathbb{P}(\mathcal{N} \mid \mathcal{A}_m) = R_N.$$

The small joint error in (13) is not a claim that the conditional miss probability after acceptance is equally small, nor a posterior probability that the specimen contains fewer than K targets. Reference gating controls how often an experiment issues an erroneous certificate across complete repetitions.

The deployment unit is one reference campaign and one future specimen. Section 7 makes the cost of departing from that unit exact rather than rhetorical, and the same discipline applies to the other three relaxations: each of Sections 8, 9 and 10 replaces an informal worry with a number that a design can be checked against. Collecting them, the worked $K = 6$ design survives mean-one Poisson loading and transport budgets of order 10^{-4}, fails for any batch of two, and requires a control false-positive rate below 1/300 that itself needs some 900 clean control pairs to certify. That profile, not the headline .0499, is the honest summary of what the design can do.

A validation study should use counted native-like inputs and representative matrices, retain the full four-outcome reference table and invalid records, and record original volume, usable volume, target-count uncertainty, extraction and well identities, coverage, and independently sampled batch identifiers. It should first challenge transport across target forms and volumes, representative material loss, competition, reference classification, and dependence across nominally independent pairs. The relevant gate quantity is $w = \mathbb{P}(A \vee B)$, while q remains useful for allocation. A convenient downstream spike is not sufficient evidence, and reporting compliance does not prove native transport. No linked native-reference dataset is supplied here; existing readout-processing records without original-volume and extraction linkage cannot establish the required law.

The practical limitation is structural. At the endpoint, an adverse source passes every reference pair and still misses finite targets because material allocation and preparation recovery remain imperfect; more calibration cannot remove that floor, and Corollary 7.2 shows the same phenomenon governs batch reuse. In the interior regime, partial acceptance balances a poorer future assay. The exact formulas identify whether additional independent references, improved unique coverage, stronger recovery information or a changed exclusion threshold can meet an intended budget. They do not warrant a universal preference for splitting, and they do not license a claim about disease absence.

Several questions remain open. The concave-envelope problem of Theorem 8.2 is solved in closed form only when the envelope degenerates to a chord; a classification of the tangency point τ outside that regime would complete it. Multiple-testing procedures designed for this architecture, rather than inherited from it, could plausibly beat Corollary 7.2 by conditioning on the calibration record. Adaptive policies that choose the next experiment after a history of negatives need a sequential analysis that we have deliberately not attempted.

A Formal scope and reproducibility

The ordinary proofs above are self-contained. The accompanying Lean 4 development [5, 14] uses the repository's pinned Lean 4.30.0 and mathlib environment, with warnings treated as errors. Its finite-mixture model admits any finite number of continuous-valued recovery states. Only the standard axioms `propext`, `Classical.choice` and `Quot.sound` occur; no `sorry`, additional axiom, native decision procedure or numerical oracle is used.

The following are *not* mechanized and are established by the ordinary proofs given above: the unequal-allocation results of Section 5; the extension from finite mixtures to arbitrary probability laws on the unit square by integrating the same pointwise inequalities; the asymptotic expansions of Section 6; the closed form (18) and the limit in Corollary 7.2 (the batch identity, the attaining witness and the numerical impossibility *are* mechanized); the coupling arguments of Section 9 including Proposition 9.1; the concave-envelope reduction Theorem 8.2 and its Corollary 8.3 (the conservative transport Proposition 8.1 *is* mechanized); the sharpness half of Proposition 9.2 (its reduction and upper bound *are* mechanized); and Theorem 10.1 with Corollary 10.2. Numerical plots illustrate established formulas; finite numerical checks do not prove continuum claims, and no assertion of biological validity follows from formal verification.

Every number printed in this manuscript is regenerated and re-checked by `check_paper.py`. The script finally re-reads `main.tex` itself and verifies that each decimal literal in the prose is the correctly rounded form of a quantity it computed, so no figure quoted in the text is unaccounted for. It uses exact rational arithmetic wherever the quantity is rational and certified rational Taylor enclosures for the exponentials in Section 8, so each integer threshold reported is a proved decision. The minimum pair counts of Table 7 are certified from both sides: an adaptive exact-rational bisection proves the constrained maximum lies below α at the stated m, and an exact rational

Table 8: Claim-to-declaration map. All declarations live in the Lean namespace `SpecimenReliability`; the thirteen modules form one dependency chain rooted at `Source.lean`. The last three modules are additions of the present paper.

Paper claim	Lean declarations
Conserved target paths and the source law (3); monotonicity in the count	`path_normalized`, `path_nonneg`, `negative_histories`, `source_law`, `risk_count_mono`
Sharp moment envelope, Theorem 3.1	`pointwise_envelope`, `sharp_upper`, `envelope_attained`, `corners_moments`, `corners_nonneg`
Paired reference law (4) and feasible moments (6)	`reference_law`, `reference_as_corners`, `moments_feasible`
Complete-event identity (9)	`accepted_identity`
Scalar maximum, Theorem 4.1	`gate_opt_properties`, `gate_maximum`, `gate_endpoint`, `gate_interior_formula`
Sharp calibrated count exclusion, Theorem 4.2	`accepted_reduction`, `accepted_general`, `general_attainment`, `witness_family_valid`
Three-pair corollary, Corollary 4.3	`cubic_gate_bound`, `accepted_resolution`, `accepted_sharp`
Reporting availability, Section 4	`accepted_blank_rate`, `utility_comparison`, `utility_examples`, `discordance_subset`
Single-preparation comparator (15)	`single_identity`, `single_general`, `single_attainment`, `single_witness_valid`
Exact worked thresholds, Table 2	`worked_bounds`
Batch composition identity behind Theorem 7.1	`atLeastOne_law`, `batch_identity`
Floor attainment, Corollary 7.2	`batch_witness`, `batch_witness_valid`
Two-specimen impossibility at $K = 6$, $a = .45$, Section 7	`batch_floor_exceeds`
False-positive reduction and bound, Proposition 9.2	`fp_reduction`, `fp_bound`
Specificity thresholds, Table 6	`fp_example_safe`, `fp_example_unsafe`
Poisson comparison and the guarantee for $\lambda \leq 1$, Proposition 8.1	`poisson_pointwise`, `poisson_accept_le`, `poisson_accept_nonneg`, `poisson_general`

witness proves it does not at $m - 1$. The one quantity that is irrational by nature, the interior branch of (18), is evaluated in double precision and labelled as such in Table 3; the reachability decisions in that section depend only on the exact floor. `make_figures.py` rebuilds all five figures. Build instructions, verification commands and the claim-to-declaration map are in the accompanying `BUILD.md`.

References

[1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1): 289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[2] Stephen A. Bustin, Jan M. Ruijter, Maurice J. B. van den Hoff, Mikael Kubista, Michael W. Pfaffl, Gregory L. Shipley, Nham Tran, Stefan Rödiger, Andreas Untergasser, Reinhold Mueller, Tania Nolan, Mojca Milavec, Malcolm J. Burns, Jim F. Huggett, Jo Vandesompele, and Carl T. Wittwer. MIQE 2.0: Revision of the minimum information for publication of quantitative

real-time PCR experiments guidelines. *Clinical Chemistry*, 71(6):634–651, 2025. doi: 10.1093/clinchem/hvaf043.

[3] Clinical and Laboratory Standards Institute. *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline—Second Edition*. CLSI, Wayne, PA, 2012. Document EP17-A2; reaffirmed 2017.

[4] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404.

[5] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction — CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.

[6] Gurutzeta Guillera-Arroita, José Joaquín Lahoz-Monfort, Anthony R. van Rooyen, Andrew R. Weeks, and Reid Tingley. Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, 8(9):1081–1091, 2017. doi: 10.1111/2041-210X.12743.

[7] Wassily Hoeffding. The extrema of the expected value of a function of independent random variables. *The Annals of Mathematical Statistics*, 26(2):268–275, 1955. doi: 10.1214/aoms/1177728543.

[8] Samuel Karlin and William J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers (Wiley), New York, 1966.

[9] Dennis D. Leber, Leticia S. Pibida, and Alexander L. Enders. Confirming a performance threshold with a binary experimental response. Technical Note 2045, National Institute of Standards and Technology, 2019.

[10] William H. Lugg, Josh Griffiths, Anthony R. van Rooyen, Andrew R. Weeks, and Reid Tingley. Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, 9 (4):1049–1059, 2018. doi: 10.1111/2041-210X.12951.

[11] Albert Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746, 1959. doi: 10.1214/aoms/1177706203.

[12] RDML Consortium. dMIQE: Minimum information for publication of quantitative digital PCR experiments. `https://rdml.org/dmiqe.html`. Checklist and link to the 2020 guidelines; accessed 17 September 2026.

[13] A. S. Smart, A. R. Weeks, A. R. van Rooyen, A. Moore, M. A. McCarthy, and R. Tingley. Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, 7:1291–1298, 2016. doi: 10.1111/2041-210X.12598.

[14] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, pages 367–381. ACM, 2020. doi: 10.1145/3372885.3373824.

[15] G. Winkler. Extreme points of moment sets. *Mathematics of Operations Research*, 13(4): 581–587, 1988. doi: 10.1287/moor.13.4.581.