Randomized stopped lineage experiments separate phenotypic switching, selection and correlated inheritance

20 September 2026

Abstract

The same resistant outgrowth can reflect pre-existing state, selective survival, within-cell switching or altered daughter production. We ask which observations separate these mechanisms in a finite-state continuous-time branching model with correlated sister states. An independently randomized exponential deadline, competing with the founder's first division or death, turns the observation law into a resolvent: calibrated initial and endpoint markers together with joint daughter records recover the switching generator, the demographic rates and the daughter kernel at every finite rate and for any finite number of states, without matrix logarithms and without a diagonalizability assumption. A monitored delayed daughter readout preserves identification through an explicit channel with interruption flags. For two states we give a rational formula for the switching rate that cancels every marker determinant before it is evaluated, and we replace Hoeffding radii by Chernoff–Kullback–Leibler feature sets, of which the Hoeffding radii are exactly the Pinsker relaxation. On a synthetic two-arm dataset with uncertain calibration these two changes narrow the conservative 95% induction interval from $[0.362, 1.498]$ to $[0.578, 1.092]$ per hour, around a true contrast of 0.8 per hour, using the same observations and the same coverage event; and they lower the certified budget of a separated-class prospective decision from 2×10^6 to 3×10^5 founders. Known calibration mixtures replace pure-state calibration exactly, and a six-category coarsening suffices when only within-cell switching is targeted. Finally, increasing sister concordance at fixed daughter marginals increases finite-time extinction weakly, and strictly in a stated case, while leaving every mean trajectory unchanged. These are model-conditional observation and prediction theorems, not a calibrated assay. The probabilistic arguments are conventional proofs; a declared finite-algebra subset is verified in Lean.

1 The question and its observation model

Phenotypic inheritance, selection and induced plasticity are compatible mechanisms. A lineage experiment should estimate their quantitative contributions, rather than choose one of three mutually exclusive explanations. Induction identification is already established in controlled ODE settings [5]; branching inference distinguishes the information in counts and fractions [6], and division-associated plasticity is explicitly modelled in drug-screen inference [11]. Lineage-tree inference motivates tracking before treatment and after withdrawal, but a lineage relationship is not itself a calibrated phenotype measurement [9].

We give an observation-design result for a declared source. The distinguishing step is to stop following a founder at its first demographic event, with an independent random deadline. This exposes both a killed within-cell chain and its joint daughter exit kernel. The resolvent method is classical in generator estimation [8]; stopped-process potential kernels are likewise standard [3]. Our claim is the explicit connection of those tools to this lineage protocol, its delayed observation channel, its finite-data target and its offspring-dependence consequence. We do not assert priority for resolvent inversion or generating-function comparison in isolation.

Three questions are kept apart throughout. *Structural identifiability* asks whether a record law determines the source at all; it is answered by algebra. *Finite-sample inference* asks what a fixed

number of independent founders certifies; it is answered by concentration inequalities and exact interval arithmetic. *Biological calibration* asks whether a real reporter measures the stipulated latent states; nothing in this paper answers it. Every numerical value below is synthetic.

1.1 Source, emissions and mechanisms

The state space is $\{1, \dots, m\}$. At a fixed exposure u, a cell of type i switches to $j \neq i$ at rate $q_{ij}(u)$, dies at rate $d_i(u)$, and divides at rate $b_i(u)$. Division replaces the mother by an ordered pair (j, k) with joint law $K_i^u(j, k)$. The kernel is stochastic; exchange symmetry is imposed when the daughter order is arbitrary. Conditional on both realized daughter states, their future subtrees have independent clocks. Daughter states need not be independent conditional on the mother.

All per-cell rates are finite. The founder has probability column vector π, with $\pi_i > 0$. Experimental units are independent founder preparations in controlled environments; treatment is randomized independently of founder state. Known marker matrix $E \in \mathbb{R}^{p \times m}$ is column stochastic and has full column rank. Its errors are independent conditional on the complete latent history, and it does not perturb the source. Columns refer to the same externally anchored states across arms. Distinct calibrated matrices at the initial and subsequent reads are allowed; a common E simplifies notation. Rates and channels stay constant during founder and daughter follow-up.

Write $Q_{ii} = -\sum_{j \neq i} q_{ij}$, and define

$$H = Q - \operatorname{diag}(b + d), \qquad B = [\, d \mid (b_i K_i(j, k))_{i,(j,k)} \,]. \tag{1}$$

Thus $H\mathbf{1} + B\mathbf{1} = 0$. Matrices act on row-state distributions. For two labelled states S, T, the principal contrasts include

$$\begin{aligned} &\pi_T, \quad q_{ST}(1) - q_{ST}(0), \quad q_{TS}(0), \quad (b_T - d_T) - (b_S - d_S), \\ &L_T(T) - L_S(T), \quad c_i = K_i(T, T) - L_i(T)^2, \quad \zeta_i(u) = 2 b_i(u) L_i^u(T), \end{aligned} \tag{2}$$

where $L_i(j) = \sum_k K_i(j, k)$ by exchange symmetry. The last contrast measures the rate of tolerant daughter production at division; its treatment change is distinct from within-cell induction. It is an expected output intensity, not a probability of switching per division. The demographic contrast alone does not describe composition dynamics when division changes phenotype.

Here "predisposition" means a pretreatment state predictive of a specified later response. It does not mean an additional unobserved primed state. If $b_i = 0$, only $b_i K_i = 0$ is identifiable: the unused kernel is immaterial. An unanchored global label permutation is harmless; separate permutations in different treatment arms are not.

1.2 Existence and the branching law

Independent competing exponential clocks, followed by kernel sampling at division, construct the chronological source. If $b_i \leq b_{\max}$, its population is dominated by a Yule process of rate $b_{\max}$ per cell. The Yule holding-time sum diverges almost surely: its centered independent terms have summable variances, while their expectations form a divergent harmonic sum. On every finite horizon the dominating population has finite integrated cell time, so bounded switching and death intensities cannot accumulate infinitely many additional events. The source is nonexplosive.

For $z \in [0, 1]$ let $F_i(t, z) = \mathbb{E}_i z^{N(t)}$. First-event conditioning gives

$$\partial_t F_i = d_i(1 - F_i) + \sum_{j \neq i} q_{ij}(F_j - F_i) + b_i \left(\sum_{jk} K_i(j, k) F_j F_k - F_i \right), \quad F_i(0, z) = z. \tag{3}$$

The product follows after conditioning on both daughter states. The bounded first-event integral equation and uniqueness of the polynomial ODE in the unit cube identify this with the actual

source PGF. Extinction by time t has probability $F_i(t,0)$; extinct founders remain in the observed denominator.

1.3 A precise limited-design obstruction

Proposition 1.1. *Fix one treatment schedule and a time $t > 0$, and retain only $Y = \mathbf{1}\{N(t) > 0\}$. Write $r_i = 1 - F_i(t,0)$ and $p = \pi_T$. At a two-state source with $r_T \neq r_S$ and $0 < p < 1$, nearby changes in the rate/kernel parameter θ can be compensated by*

$$p' = \frac{(1-p)r_S + pr_T - r'_S}{r'_T - r'_S}. \tag{4}$$

This is a local family on an observational equivalence level set. The full binary observation law is unchanged, although within-cell induction, selection or daughter inheritance may change.

Proof. The law is Bernoulli with parameter $(1-p)r_S + pr_T$. Continuous dependence in (3) keeps the denominator nonzero and p' interior locally. Substitution proves equality of the entire law. A nondegenerate class exists: choose all rates positive and $d_S > d_T$. At zero, $(r_T - r_S)' = d_S - d_T > 0$, so the response contrast is positive for sufficiently small positive t. One may then vary $q_{ST}(1)$ while holding the untreated rates fixed. Any test of two such aliased sources has type-I plus type-II error one, since both product observation laws are identical. □

The result concerns this binary threshold and an unknown preparation. It is not an impossibility theorem for full fluctuation counts, all trees or multiple randomized schedules.

2 Randomized stopping and an algebraic source inverse

2.1 The retained record

Read one founder marker Y_0 and draw an independent deadline $T \sim \mathrm{Exp}(\lambda)$ with known $\lambda > 0$. Follow the founder until $\sigma = \min(T, \tau_{\mathrm{dem}})$, where τ_{dem} is its first death or division; unobserved switches do not stop follow-up. At a deadline, read an endpoint marker. At death, retain a death flag. At division, read the joint daughter marks. Neither the realized deadline nor the event timestamp enters the retained record. All founders are retained. For binary markers this immediate-read design has $2(2+1+4) = 14$ outcomes. Figure 1 summarises the protocol.

Theorem 2.1 (Finite-state randomized-deadline identification)**.** *Under the source and measurement assumptions above, put $R_\lambda = (\lambda I - H)^{-1}$. The latent deadline and demographic-exit kernels are*

$$A_\lambda = \lambda R_\lambda, \qquad D_\lambda = R_\lambda B. \tag{5}$$

The observed law determines $\pi, A_\lambda, D_\lambda$ and hence

$$H = \lambda(I - A_\lambda^{-1}), \qquad B = \lambda A_\lambda^{-1} D_\lambda. \tag{6}$$

It identifies Q, b, d and each K_i for which $b_i > 0$, at arbitrary finite rates and for any finite number of states. Reducibility, zero switching and singular H are allowed. No diagonalizability assumption is required.

Proof. Before demographic exit the founder is a killed chain: its row distribution satisfies $v' = vH$, so its substochastic semigroup is $S_t = e^{tH}$. Given state i at time t, exit mark a has intensity B_{ia}. Competition with the independent deadline gives

$$A_\lambda = \int_0^\infty \lambda e^{-\lambda t} e^{tH}\, dt, \qquad D_\lambda = \int_0^\infty e^{-\lambda t} e^{tH} B\, dt.$$

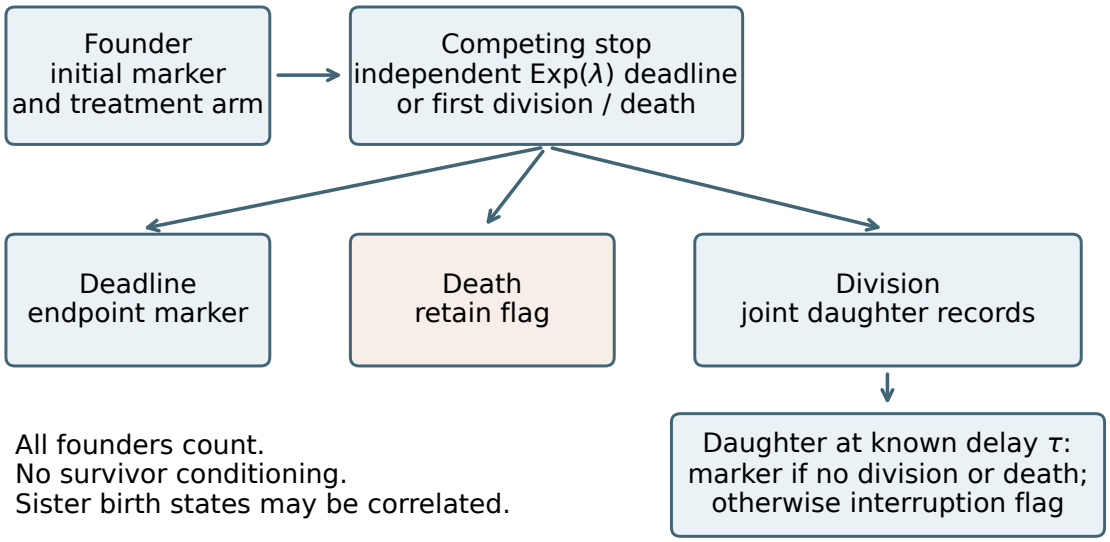

Figure 1: The observation protocol. The deadline clock is independent of the source. Switching is hidden throughout. A delayed daughter read retains an interruption flag for any intervening division or death; records are never conditioned on both daughters surviving. The delayed binary-marker experiment has 24 joint outcomes, versus 14 with immediate daughter readout.

Since $\left\|e^{tH}\right\|_\infty \le 1$, these integrals converge. If R denotes the integral without λ or B, integrate the derivative of $e^{t(H-\lambda I)}$ from zero to infinity. Its limit is zero, and its derivative commutes with $H-\lambda I$, giving both $(\lambda I - H)R = I$ and $R(\lambda I - H) = I$. This proves the inverse and (5), even if H is singular. Moreover

$$A_\lambda \mathbf{1} + D_\lambda \mathbf{1} = R_\lambda(\lambda I - H)\mathbf{1} = \mathbf{1},$$

so death and division account for all probability missing from deadline records.

Let $C = \operatorname{diag}(1, E \otimes E)$ be the exit emission channel, with rectangular blocks when necessary. For an arbitrary left inverse $LE = I$, a left inverse of C is $L_C = \operatorname{diag}(1, L \otimes L)$. The observed deadline and exit blocks are

$$J = E \operatorname{diag}(\pi) A_\lambda E^\top, \qquad V = E \operatorname{diag}(\pi) D_\lambda C^\top. \tag{7}$$

These follow by summing over the initial, current and daughter birth states; only measurement noise and daughter futures are factored, not the birth-state kernel K. The marginal m_0 of the initial marker sums over every outcome, so $\pi = Lm_0$. Thus

$$A_\lambda = \operatorname{diag}(\pi)^{-1} L J L^\top, \qquad D_\lambda = \operatorname{diag}(\pi)^{-1} L V L_C^\top.$$

Positive founder masses permit row division. Equation (6) now follows algebraically. Offdiagonal entries of H give q_{ij}; the death column of B gives d, its daughter row sums give b, and division by positive b_i gives K_i. The diagonal of H checks the recovered total demographic killing. Two randomized, label-aligned exposure arms identify both within-cell induction and treatment changes in daughter production. □

2.2 Finite horizons, stability and a model check

The expected founder observation time is

$$\mathbb{E}\sigma = \pi^\top R_\lambda \mathbf{1} \le \lambda^{-1}. \tag{8}$$

For a physical cap L_0, use $\min(T, L_0)$ and label a cap endpoint as a deadline endpoint. Coupling to the uncapped source shows

$$\mathrm{TV}(P_{\text{capped}}, P_{\text{uncapped}}) \le e^{-\lambda L_0} \pi^\top e^{L_0 H} \mathbf{1} \le e^{-\lambda L_0}. \tag{9}$$

Only paths with $T > L_0$ and no demographic event by L_0 can differ. A cap flag could instead be retained for an exact capped likelihood. The capped kernel itself is not a resolvent. Choosing $L_0 = \log(1/\varepsilon)/\lambda$ yields a feature-bias reserve ε.

If $\|H\|_\infty, \|H'\|_\infty \le R_0$, then $\left\|A_\lambda^{-1}\right\|_\infty \le 1 + R_0/\lambda$. The inverse difference identity yields

$$\left\|H - H'\right\|_\infty \le \lambda(1 + R_0/\lambda)^2 \left\|A_\lambda - A'_\lambda\right\|_\infty. \tag{10}$$

For binary markers, if every observed category differs by at most e and $\pi_i \ge \rho$, marker inversion and row normalization give

$$\left\|A_\lambda - A'_\lambda\right\|_\infty \le \frac{2\|L\|_\infty \left\|L^\top\right\|_\infty + 7\|L\|_\infty}{\rho} e.$$

Indeed the four-entry deadline block has maximum row-sum error $2e$; the initial marginal has entrywise error at most $7e$; and A' has row sums at most one. Larger record spaces use their actual category count. The prefactor in (10) is $\lambda + 2R_0 + R_0^2/\lambda$, minimized at $\lambda = R_0$ when $R_0 > 0$. This is a conditioning guide, not an optimal design theorem: event yield, feature variance and costs also depend on λ.

Two independently assigned deadline rates for the same source must obey

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu. \tag{11}$$

This is the inverse difference identity, since both inverse matrices differ by $(\mu - \lambda)I$. Recovered H, B, π must agree across groups. Construct simultaneous source confidence sets for each group and reject compatibility when their intersection is empty. A union bound controls rejection of a true common source. Passing finitely many checks does not establish a Markov model or uniquely diagnose an incompatible biological assumption.

2.3 Fixed deadlines as an alternative

Proposition 2.2. *For a deterministic deadline $h > 0$, replace the kernels in (5) by $A_h = e^{hH}$ and $D_h = W_h B$, $W_h = \int_0^h e^{tH} dt$. For two states the same observations identify the source at every $h > 0$. For general m, it suffices that $h\|H\|_\infty < \log 2$ throughout the admitted class.*

Proof. The killed-chain occupation derivation is unchanged. In two states write $H = \left[\begin{smallmatrix} -\alpha & a \\ r & -\beta \end{smallmatrix}\right]$. Its eigenvalues are $[-(\alpha+\beta) \pm \sqrt{(\alpha-\beta)^2 + 4ar}]/2$, real and nonpositive. If distinct, A_h has distinct positive eigenvalues with the same projectors; their unique real logarithms recover H. Any competing two-state killed generator has real spectrum and commutes with A_h, so uses the same projectors. In the repeated case $H = \xi I + N$, $N^2 = 0$, and $A_h = e^{h\xi}(I + hN)$ determines both ξ and N, including $N = 0$. The eigenvalues of W_h are $\int_0^h e^{t\xi} dt > 0$; hence $B = W_h^{-1} D_h$. For general m under the bound, the convergent logarithm series on $\|A_h - I\|_\infty < 1$ gives $\log(e^{tH}) = tH$ by differentiating along $0 \le t \le h$. Also $\|W_h/h - I\|_\infty < 1$, so a Neumann series inverts W_h. Marker cancellation completes both cases. Randomized stopping avoids these spectral and small-step restrictions. □

2.4 Implementing a random deadline

Randomized stopping does not require physically interrupting a culture. For a nonperturbing live reporter with sufficiently resolved complete traces, draw $T_i \sim \mathrm{Exp}(\lambda)$ independently for each founder and retain the record only up to $\min(T_i, \tau_{\mathrm{dem},i})$. The deadline may be drawn before imaging, or afterwards provided it is drawn independently of the recorded paths. It must not be chosen using the observed fate, and founders must not be included only because their follow-up happened to be long enough.

Two finite-resource effects must be represented in the observation model rather than assumed away. A finite movie horizon is the cap of (9), or an exact censored likelihood with a retained cap flag. A fixed camera frame interval gives a discretized reading rule, not an exact continuously timed marker read; the rounded-time channel must be derived or its discrepancy bounded. Repeated randomized truncations of one founder trace remain dependent observations and cannot be counted as additional independent roots.

Raw times and images should be retained when feasible, even though Theorem 2.1 shows they need not enter the minimal statistical record. Compression proves sufficiency for structural identification on this model; it does not imply that timestamps carry no additional finite-sample information. Randomization creates no information that was absent from the original trace.

3 Delayed daughter observations with interruption flags

After a founder division, follow each daughter for a known duration τ. If she divides or dies during that interval, retain an interruption flag and stop following her. Otherwise read her marker at τ. Switching during the interval remains hidden. The two daughter records are retained jointly, including all interrupted cases. No inference conditions on joint survival.

Theorem 3.1. *For every finite $\tau \geq 0$, the daughter observation channel is*

$$F_\tau = \begin{bmatrix} E e^{\tau H^\top} \\ (\mathbf{1} - e^{\tau H}\mathbf{1})^\top \end{bmatrix}. \tag{12}$$

Its columns sum to one and it has full column rank. The monitored delayed protocol identifies the same source as Theorem 2.1, or Proposition 2.2 within its scope.

Proof. The unnormalized live-state distribution of a daughter born in state j is row j of $e^{\tau H}$. Applying E gives the top block of (12); its missing mass is the interruption probability. Since $e^{\tau H}$ is invertible and E has full column rank, the upper block has full column rank. An explicit left inverse is $e^{-\tau H^\top} L[\, I_p \mid 0\,]$.

The founder deadline block, unaffected by daughter readout, first recovers π, H. Summing delayed pair categories recovers founder division totals; the two demographic columns are $R_\lambda[d \mid b]$, so recover d, b. Now H determines F_τ. Conditional on the two daughter birth states, the two delayed records are independent; hence their joint channel is $F_\tau \otimes F_\tau$. If $L_\tau F_\tau = I$, its left inverse is $L_\tau \otimes L_\tau$, as follows by multiplying and factoring the two finite sums. Substitute this channel in (7) to recover $b_i K_i(j,k)$, and then K_i wherever $b_i > 0$. This ordering avoids circular inference of H from an unknown daughter channel. □

The binary-marker record has $2(2+1+9) = 24$ categories. A full-law Hoeffding region over two arms therefore has radius $\sqrt{\log(96/\delta)/(2n)}$, rather than the immediate-record constant. The direct switching interval of Section 4 uses only its ten stated features and does not require all 48 categories to be estimated separately.

Invertibility is not good conditioning: in spectral norm, $\sigma_{\min}(F_\tau \otimes F_\tau) = \sigma_{\min}(F_\tau)^2$. If the founder division probability is $y_\lambda = \pi^\top R_\lambda b$, additional daughter observation is at most $2\tau y_\lambda$ cell-hours per founder. Equal root counts and equal observation-time budgets are different

comparisons (Figure 4). Monitoring still must detect the first founder division and daughter interruptions. Unknown reporter maturation, inherited fluorescence, unknown delay or missed events are not covered by (12).

Discarding the second daughter loses genuine information: the perturbation $K_i \mapsto K_i + \varepsilon(1,-1,-1,1)$ preserves each daughter marginal but changes c_i by ε. For interior kernels small nonzero ε is feasible. The whole one-daughter stopped law remains equal, whereas the invertible pair channel separates the joint kernels. Thus the requirement of joint daughter records is substantive.

4 Induction intervals with unknown nuisance rates

The exact inverse need not be applied to every parameter in order to estimate a particular mechanism. For two states let $M = \operatorname{diag}(\pi)A_\lambda = LJL^\top$. Direct inversion of a two-by-two matrix gives

$$q_{ST} = \lambda\frac{(A_\lambda)_{ST}}{\det A_\lambda} = \lambda\frac{\pi_T M_{ST}}{\det M}, \qquad q_{TS} = \lambda\frac{\pi_S M_{TS}}{\det M}. \tag{13}$$

The factors π_T and π_S follow from the row-state convention. The determinant is positive: $\lambda I - H$ has positive diagonal entries and determinant $(\lambda + a + b_S + d_S)(\lambda + r + b_T + d_T) - ar > 0$. Consequently $\det A_\lambda > 0$ and $\det M > 0$ for positive founder masses.

Only five features per arm are needed: the four joint initial/deadline marker probabilities and $\mathbb{P}(Y_0 = 1)$ over *all* founders. Demographic rates and K influence these probabilities but need not be estimated or assumed known. The same formula works with delayed daughter records, since none of these five features uses daughter outcomes. We therefore say the construction proceeds *without estimating demographic nuisance rates*; calibration, founder prevalence and the source model still matter, and q_{ST} is a switching rate whereas induction is the between-arm contrast.

4.1 Cancelling the marker determinant

Evaluating (13) by first widening each entry of M and then taking a difference of products repeats the same uncertain calibration factors several times. They can be cancelled exactly beforehand.

Proposition 4.1 (Cancelled-determinant target)**.** *Write* $e_S = \mathbb{P}(Y = 1 \mid S)$, $e_T = \mathbb{P}(Y = 1 \mid T)$, $\delta_E = e_T - e_S > 0$, $\mu = \mathbb{P}(Y_0 = 1)$ *and* $J = (j_{ab})_{a,b\in\{0,1\}}$. *Put*

$$N = -e_T e_S j_{00} + e_T(1-e_S)j_{01} + (1-e_T)e_S j_{10} - (1-e_T)(1-e_S)j_{11}. \tag{14}$$

Then $M_{ST} = N/\delta_E^2$, $\det M = \det J/\delta_E^2$ *and* $\pi_T = (\mu - e_S)/\delta_E$, *so that*

$$\boxed{q_{ST} = \lambda\,\frac{(\mu - e_S)\,N}{\delta_E\,(j_{00}j_{11} - j_{01}j_{10})}.} \tag{15}$$

Proof. With

$$L = \frac{1}{\delta_E}\begin{bmatrix} e_T & -(1-e_T) \\ -e_S & 1-e_S \end{bmatrix}, \qquad \pi_T = \frac{\mu - e_S}{\delta_E}, \tag{16}$$

one has $\det L = (e_T(1-e_S) - (1-e_T)e_S)/\delta_E^2 = 1/\delta_E$. Hence $\det M = \det(LJL^\top) = (\det L)^2 \det J = \det J/\delta_E^2$. Expanding $M_{ST} = (LJL^\top)_{01}$ gives exactly N/δ_E^2. Substituting both into (13) cancels δ_E^{-2} and leaves (15). □

Identity (15) is exact, so it changes no modelling assumption. Its value is that in interval arithmetic every factor appears as few times as possible: the calibration contrast δ_E occurs once instead of four times, and the four observed probabilities enter one bilinear numerator and one

determinant instead of four separately widened entries of M. We do not claim a universal width inequality between the two evaluations; we report the measured reduction on the dataset of Section 4.4 and across the matched diagnostics of Figure 2.

4.2 Chernoff feature sets

Let $\mathrm{kl}(a\|p) = a\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}$ be the binary relative entropy, with the usual conventions at $a \in \{0,1\}$.

Lemma 4.2 (Chernoff set and its Pinsker relaxation)**.** *Let $\hat{p}$ be the mean of n independent Bernoulli(p) variables and $L > 0$. Put $C_L(\hat{p}) = \{q \in (0,1) : n\,\mathrm{kl}(\hat{p}\|q) \le L\}$. Then $\mathbb{P}(p \notin C_L(\hat{p})) \le 2e^{-L}$, and*

$$C_L(\hat{p}) \subseteq \Big[\hat{p} - \sqrt{L/(2n)},\ \hat{p} + \sqrt{L/(2n)}\Big]. \tag{17}$$

Proof. The map $a \mapsto \mathrm{kl}(a\|p)$ is convex with minimum 0 at $a = p$, so $\{p \notin C_L(\hat{p})\} = \{\hat{p} \ge a_+\} \cup \{\hat{p} \le a_-\}$, where $\mathrm{kl}(a_\pm\|p) = L/n$ and $a_- \le p \le a_+$ (an empty event if the corresponding root does not exist in $[0,1]$). The Chernoff bound for binomial tails [2, 7] gives $\mathbb{P}(\hat{p} \ge a) \le e^{-n\,\mathrm{kl}(a\|p)}$ for $a \ge p$ and $\mathbb{P}(\hat{p} \le a) \le e^{-n\,\mathrm{kl}(a\|p)}$ for $a \le p$; each term is at most e^{-L}. Pinsker's inequality in the binary case is $\mathrm{kl}(a\|q) \ge 2(a-q)^2$, so $q \in C_L(\hat{p})$ forces $2(\hat{p}-q)^2 \le L/n$, which is (17). □

Taking $L = \frac{25}{4}$ recovers exactly the radius $r_n = \sqrt{25/(8n)}$, and $L = 6$ recovers exactly $s_c = \sqrt{3/n_c}$. In other words, the explicit Hoeffding radii are the Pinsker relaxation of the Chernoff sets: the sets have the same error allocation and are never wider, strictly narrower whenever the proportion is away from one half. Such kl-based intervals are standard in the concentration literature [4]. Because rational interval arithmetic is inclusion preserving, we obtain the following free improvement.

Corollary 4.3. *Fix any evaluation of (13) or (15) by inclusion-preserving interval arithmetic. Replacing each Hoeffding feature interval by the corresponding Chernoff set C_L yields an output interval contained in the original output interval, with an unchanged coverage guarantee.*

4.3 The finite-sample interval

Theorem 4.4 (Direct finite-sample interval)**.** *Take n independent founders per arm and n_c independent labelled calibration readouts per state, independent of the experimental units. Replace each of the ten empirical feature proportions by the Chernoff set $C_{25/4}$ of Lemma 4.2, and each of the two calibration proportions by C_6. Evaluate (16) and (15) using inclusion-preserving interval arithmetic. If the marker-contrast lower bound or the lower bound on $\det J$ is nonpositive, return $[0,\infty)$ for that arm; otherwise intersect the result with $[0,\infty)$, an empty result being a model-conflict flag. Subtract the arm intervals to obtain an induction interval. Its coverage is at least 0.95, uniformly over the admitted source, without knowing division, death or reverse-switching rates, K or π.*

Proof. By Lemma 4.2 each founder feature is missed with probability at most $2e^{-25/4}$ and each calibration proportion with probability at most $2e^{-6}$. The ten features may overlap within a unit; the union bound does not require their independence. Hence all twelve true values lie in their sets except with probability at most $20e^{-25/4} + 4e^{-6} < 0.04 + 0.01 = 0.05$. On that joint coverage event, rational interval addition, multiplication and division by a sign-separated interval contain the true intermediate values, by their endpoint definitions. The true δ_E and $\det J$ are positive, the latter because $\det J = \delta_E^2 \det M$ and $\det M > 0$; when the enclosures certify those two signs the evaluation proceeds, and otherwise the unbounded fallback is returned. Thus (15), which by Proposition 4.1 equals q_{ST} exactly, is contained whenever it is evaluated, and the fallback contains it otherwise. The nonnegative intersection cannot remove the true switching rate. Subtracting the intervals retains the true contrast. No asymptotic likelihood argument and no uncertified global optimization is used. □

Table 1: The retained two-arm dataset. J columns refer to the endpoint marker only when the deadline precedes demographic exit; the fifth column also includes death and division records. The last two columns are the demographic-exit totals of the six-category coarsening of Section 4.5.

	deadline first					demographic exit	
Arm	J_{00}	J_{01}	J_{10}	J_{11}	all initial $Y_0 = 1$	$Y_0 = 0$	$Y_0 = 1$
Untreated	19497	4654	4270	10288	18023	7826	3465
Treated	13937	10895	3754	11178	18242	6926	3310

The implementation encloses each Chernoff set by certified rational endpoints: candidate endpoints are located in floating point and then *verified* to lie outside the set by an exact rational lower bound on kl, built from an argument-reduced arctangent-hyperbolic series with an explicit geometric tail (Appendix A). The inequalities $e^6 > 400$ and $e^{25/4} > 500$ follow from finite positive Taylor sums. Arithmetic guarantees containment; concentration together with the independence and model assumptions supplies the probability guarantee. Both are needed.

The exact binomial (Clopper–Pearson) set is narrower still, and it is the natural next candidate. We do not use it, for an implementation reason that is specific to an exact-arithmetic pipeline: certifying a Clopper–Pearson endpoint p requires bounding a partial binomial tail $\sum_{k \le x} \binom{n}{k} p^k (1-p)^{n-k}$, and at $n = 5 \times 10^4$ a rational p with a 10^{-12} grid makes p^k carry a denominator of order 10^{12k}, so the exact sum is not computable. The Chernoff set is the tightest object we know how to certify in exact rational arithmetic, because its defining inequality involves only two logarithms, each of which has a short rational enclosure. A floating-point Clopper–Pearson interval would reintroduce exactly the uncertified numerics that the rest of this construction avoids.

For a cap or other justified per-feature bias η, widen each feature interval by η. Channel calibration uncertainty is already carried as a region for (e_S, e_T) and is not silently replaced by a point estimate. With known E, omit the calibration error allocation. If reporter errors depend on unobserved histories rather than the stipulated current state, a separate observation model or quantified error bound is necessary.

4.4 A worked unknown-nuisance dataset

All quantities in this example are synthetic. The generating source uses

$$\begin{gathered} \pi_T = .35, \quad b = (.22, .16), \quad d = (.12, .04), \quad q_{TS} = .07, \\ K_S = (.62, .12, .12, .14), \quad K_T = (.10, .10, .10, .70), \\ E = \begin{bmatrix} .9 & .15 \\ .1 & .85 \end{bmatrix}, \quad q_{ST}(0) = .05, \quad q_{ST}(1) = .85, \quad \lambda = 1. \end{gathered} \tag{18}$$

Rates are per hour. The inference routine receives only category counts, calibration counts and λ, not these source parameters. For $n = 50{,}000$ per arm and $n_c = 100{,}000$ per state, seed 20260922 gives calibration successes $(10{,}211, 85{,}049)$ and the counts of Table 1.

Table 2 evaluates the same data and the same coverage event by four routes. The original construction — Hoeffding radii and entrywise inversion of M — returns the induction interval $[0.362, 1.498]$. Cancelling the marker determinant removes 39.47% of the width; replacing the radii by Chernoff sets removes 30.58%; together they remove 54.79%. The outward-rounded publication interval is

$$[0.578,\ 1.092]\ \text{hour}^{-1}, \tag{19}$$

around a true contrast of 0.8 per hour. It excludes zero. The improvement uses the same observations, the same error allocation and the same confidence event as the original construction;

Table 2: Four evaluations of the same data under the same 95% coverage event. Endpoints are outward-rounded to four decimals; the exact rational endpoints are in the accompanying numerical record. The Chernoff rows are contained in the Hoeffding rows of the same evaluator (Corollary 4.3); the cancelled columns are narrower here but are not claimed to be narrower at every dataset.

Feature sets	Evaluator	Untreated q_{ST}	Treated q_{ST}	Induction	Width reduction
Hoeffding	(13)	$[0, .1608]$	$[.5232, 1.4978]$	$[.3624, 1.4978]$	—
Hoeffding	(15)	$[0, .1282]$	$[.6316, 1.1908]$	$[.5034, 1.1908]$	39.47%
Chernoff	(13)	$[.0054, .1190]$	$[.5999, 1.2746]$	$[.4809, 1.2692]$	30.58%
Chernoff	(15)	$[.0130, .1010]$	$[.6793, 1.1045]$	$[.5783, 1.0916]$	54.79%

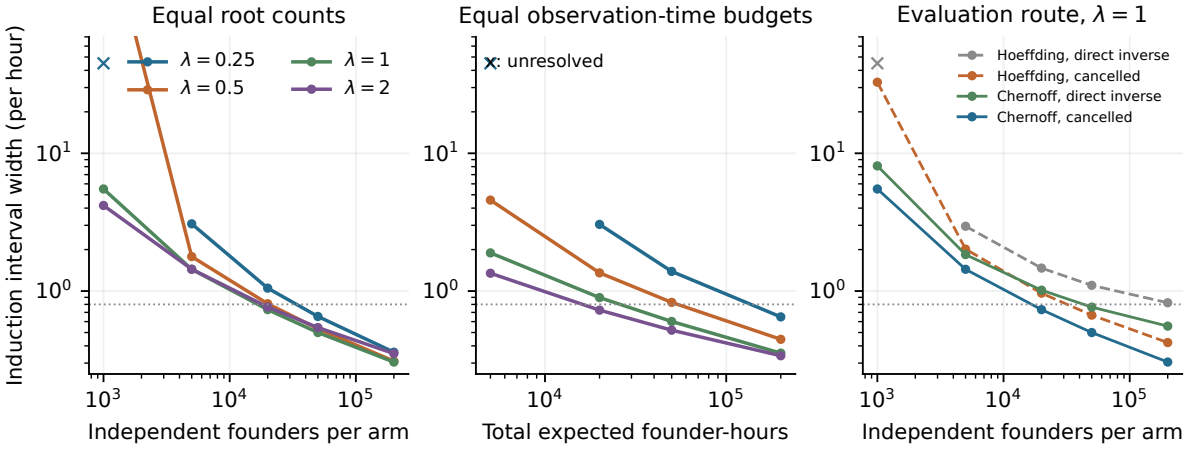

Figure 2: Matched design diagnostics for the sources in (18). All panels use the same source pair and 200,000 total calibration readouts, rounded expected category counts and the conservative interval algorithm; they are not Monte Carlo coverage or prospective guarantees, and unbounded intervals are shown as unresolved rather than as finite widths. Left: varying the number of founders in each arm at four deadline rates, using the improved route. Middle: varying total expected founder observation time with equal allocation to the two arms; calibration time is not included in founder-hours and its sample count is held fixed. Right: the four evaluation routes of Table 2 at $\lambda = 1$.

it introduces no new statistical assumption, extra data, asymptotic approximation or additional error allocation, and it does not by itself establish a universal sample-size requirement.

A qualitative gain accompanies the quantitative one. Under the original route the untreated arm returned $[0, .161]$, consistent with no spontaneous switching at all; the improved route returns $[.0130, .1010]$, which certifies a strictly positive baseline switching rate at the same confidence level. The true value is .05 per hour.

A retained 20-founder example has a determinant interval reaching zero and returns unresolved under all four routes. The procedure never replaces that failure with a clipped finite answer.

In 300 independent synthetic repetitions of the improved route, including fresh calibration each time, the interval covered the true contrast and excluded zero in all 300. For either empirical proportion the two-sided 95% exact binomial Monte Carlo interval is $[.9878, 1]$. This is a diagnostic of conservatism and power at this source, not the coverage proof and not a universal power claim.

4.5 A switching-only record with six categories

If the scientific target is only within-cell switching, the daughter phenotype readout can be dropped entirely.

Proposition 4.5. *Retain, per arm, the four deadline categories (Y_0, Y_T) and two demographic-exit categories indexed by Y_0 alone, with counts $(n_{00}, n_{01}, n_{10}, n_{11}, n_{0D}, n_{1D})$. These six categories*

determine all five features of (15)*, hence Theorem 4.4 applies verbatim to this record.*

Proof. The four deadline proportions are immediate. The fifth feature is $\mu = (n_{10}+n_{11}+n_{1D})/n$, and n_{0D} is determined by subtraction from n. No other functional of the record enters (15). □

Here D means *either* first death *or* first division, for this deliberately coarsened target; it does not mean death alone. The corresponding counts for the worked dataset are the last two columns of Table 1. This six-category multinomial retains all five features including their dependencies, so no information about q_{ST} is lost relative to the full record. Recovering b, d, K, or predicting the sister-dependence consequence of Section 6, still requires the richer record. This is a useful experimental cost distinction, not a claim that all mechanisms can be recovered from six categories.

4.6 Prospective guarantees and the source of cost

A prospective statement must control every source in a specified class, not just the realized dataset. Here is a deliberately local, nonempty example. Allow $\pi_T \in [.345, .355]$, $q_{TS} \in [.068, .072]$, $b_S \in [.218, .222]$, $b_T \in [.158, .162]$, $d_S \in [.118, .122]$, $d_T \in [.038, .042]$, and arbitrary stochastic K in each arm. Set the true marker probabilities to $.1, .85$, but estimate them by the calibration procedure. Control q_{ST} lies in $[.049, .051]$; the null treated range is the same, whereas the alternative treated range is $[.849, .851]$. Thus neither hypothesis hides a known demographic nuisance value, although these performance classes are intentionally narrow. Classify using the threshold .4 per hour.

For this calculation, enclose the exact two-state inverse $(\lambda I - H)^{-1}$ over the input boxes and apply the emissions, obtaining a box of true feature values. Then enlarge that box to cover every interval the procedure could report. Under the original construction the enlargement is by $2r_n$ and $2s_c$, the factor two accounting for sampling displacement and the reported confidence radius. Under Chernoff sets the corresponding enlargement of a true value p is

$$\mathcal{E}(p) = \bigcup \{C_L(a) : \; n \; \mathrm{kl}(a\|p) \le L\} = \left[\, \ell(a_-(p)), \; u(a_+(p)) \,\right], \tag{20}$$

where ℓ, u are the endpoints of C_L and $a_\pm$ those of the ball in the first argument; both maps are nondecreasing, so the union is the stated interval and a box of true values is enlarged by applying ℓ and u at its two ends. By (17), $\mathcal{E}(p) \subseteq [p - 2r_n, p + 2r_n]$, so the prospective enclosure is never wider than before. Inclusion arithmetic encloses every possible output on the coverage event; no source grid or optimizer establishes the extrema.

At $\lambda = 1$ the original route certifies the decision at $n = n_c = 10^6$ per arm and per calibration state: all returned null contrast intervals lie inside $[-.129, .129]$ and all returned alternative intervals inside $[.441, 1.299]$. The improved route certifies the same decision at $n = n_c = 1.5 \times 10^5$, with null intervals inside $[-.144, .144]$ and alternative intervals inside $[.408, 1.314]$; it does not certify at $n = n_c = 1.4 \times 10^5$, where the alternative lower bound is .3998. The sufficient budget therefore falls from two million founders plus two million labelled calibration readouts to three hundred thousand of each, a factor of 6.7. These remain sufficient budgets for this narrow class, not recommended experimental sizes and not lower bounds. At smaller budgets the rectangular method still loses information through interval dependencies, marker subtraction and determinant conditioning.

There is nevertheless an intrinsic small-effect limit for the immediate-read record, or its founder-only coarsening. Compare sources differing only in $q_{ST} = a$ versus $a + \epsilon$, $a \ge a_0 > 0$. The latent stopped-path relative entropy is expected occupation in S times $a \log(a/(a+\epsilon)) + \epsilon \le \epsilon^2/a_0$. Expected occupation is at most $1/\lambda$; discarding latent information cannot increase relative entropy. Consequently any test on n treated founders has error sum at least

$$\max\left\{0, 1 - \sqrt{\frac{n\epsilon^2}{2a_0\lambda}}\right\}. \tag{21}$$

This follows from the variational testing bound, Pinsker's inequality and product additivity; the unchanged control adds no information about this pair. It explains why arbitrarily small induction cannot be resolved uniformly at a fixed sample size. It does not establish that the sufficient budgets above are necessary.

4.7 What a joint region could still buy

Rectangular propagation discards the dependence between the features, and a better region is plausible. It is worth stating precisely what would be needed, because a Dirichlet posterior credible set is not automatically a uniformly valid frequentist confidence set, and an ordinary likelihood ellipsoid is generally only asymptotic.

Proposition 4.6 (Normalized-mixture region)**.** *For one ordered categorical sample with counts $n_1, \dots, n_k$ let $\mathcal{L}(p; x) = \prod_r p_r^{n_r}$ and $m(x) = \int \mathcal{L}(v; x)\, \Pi(dv)$, where the mixing distribution Π on the simplex is fixed before the data are seen. Then $\mathcal{C}_\alpha(x) = \{p : \mathcal{L}(p; x) \geq \alpha\, m(x)\}$ has coverage at least $1 - \alpha$ for every p. For the uniform Dirichlet mixture, $m(x) = (k-1)! \prod_r n_r!/(n+k-1)!$.*

Proof. For fixed p, $\mathcal{L}(p; \cdot)$ is a probability distribution on ordered samples of length n, so $\sum_x m(x) = \int \sum_x \mathcal{L}(v; x)\Pi(dv) = 1$ by Fubini. Hence $\mathbb{E}_p[m(X)/\mathcal{L}(p; X)] = \sum_{x:\mathcal{L}(p;x)>0} m(x) \leq 1$, where samples of zero probability contribute nothing. Markov's inequality gives $\mathbb{P}_p(m(X)/\mathcal{L}(p; X) > 1/\alpha) \leq \alpha$, which is the complement of $\{p \in \mathcal{C}_\alpha(X)\}$. The Dirichlet integral is the Beta function identity. □

Multiplying the mixture distributions for the two independent founder samples and the two independent calibration samples gives one normalized joint numerator, so a single threshold $\alpha = .05$ delivers simultaneous coverage without splitting the error budget across twelve marginal sets. Intersecting the region with the calibrated common-label and valid-source constraints and then projecting the induction target preserves coverage, provided the projection is outer-validated; an inner feasible region alone does not, because it can exclude the truth. The exact arithmetic numerator is straightforward, and certified projection of the nonlinear target is the real computational task; it is not implemented here. A closely related established alternative is the split likelihood-ratio construction of universal inference [10], which supplies finite-sample likelihood-ratio confidence sets and nuisance handling; its precise construction should be used rather than attaching a finite-sample label to an ordinary likelihood ellipsoid.

Mixture penalties can themselves be conservative, so superiority over the simplified rectangle is not promised. The improvements actually carried out in this paper — Proposition 4.1 and Lemma 4.2 — are both exact and both reuse the original coverage event, which is why they compose without any further union bound. We also note that one must not select the most favourable among several separately valid procedures without accounting for the selection; intersecting enclosures derived from the *same* simultaneous event, as here, is a different and legitimate operation.

5 Calibration without pure populations

Theorem 4.4 requires independently justified state labels in the calibration sample. Pure sorted populations are a strong demand, and more calibration replicates reduce sampling uncertainty but never unknown systematic label contamination. Known mixtures suffice instead, exactly.

Proposition 5.1 (Known-mixture calibration)**.** *Let two calibration preparations have independently known tolerant fractions $p_0 \neq p_1$* at readout*, and let $r_g = \mathbb{P}(Y = 1 \mid \text{preparation } g)$. Under the same state-conditional channel, $r_g = e_S + (e_T - e_S)p_g$, hence*

$$e_T - e_S = \frac{r_1 - r_0}{p_1 - p_0}, \quad e_S = \frac{p_1 r_0 - p_0 r_1}{p_1 - p_0}, \quad e_T = \frac{(1 - p_0)r_1 - (1 - p_1)r_0}{p_1 - p_0}. \tag{22}$$

Pure states are the special case $p_0 = 0$, $p_1 = 1$. More generally, if P is a known matrix of preparation compositions with full row rank and $Y = EP$ are the observed marker distributions, any right inverse gives $E = YP^+$.

Proof. Condition on the latent state within a preparation and use that the channel does not depend on the preparation. The displayed formulas solve the resulting linear system, whose determinant is $p_1 - p_0$. The matrix statement is the same computation with P in place of (p_0, p_1). □

With uncertain mixture fractions, propagate their independently justified confidence or bias regions jointly with r_0, r_1 and enforce a nonzero mixture contrast; the additional factor $|p_1 - p_0|^{-1}$ quantifies the price of similar preparations. This does not solve unknown ground truth by itself: the compositions must be established by information independent of the uncalibrated marker being assessed. Sorting on that same marker and declaring the gates pure is circular.

Remark 5.2 (Bounded contamination)**.** If nominal S and T preparations have contamination fractions at most η_S, η_T at readout, their marker means differ from e_S, e_T by at most $\eta_S|e_T - e_S|$ and $\eta_T|e_T - e_S|$, hence by at most η_S and η_T. Widen the calibration sets by these justified bounds and propagate the resulting channel region. Unknown contamination must not be converted into an arbitrarily small reserve. If cells switch between sorting and measurement, the relevant composition is the composition at readout; a known transition model and delay can propagate it, and without that knowledge sorting purity at time zero does not establish readout purity. If the resulting channel region admits rank loss, report partial or unresolved identification.

6 Sister concordance changes extinction without changing means

The joint kernel is identifiable for a reason beyond fitting lineage correlations: it determines a population outcome that mean growth cannot. Offspring generating-function orders have an established role in extinction theory [1]. The following finite-time consequence is specialized to the identified two-state source and a fixed-marginal kernel change; it is not a new general principle that offspring dependence affects extinction.

Theorem 6.1. *For an exchange-symmetric daughter kernel define*

$$G = Q - \operatorname{diag}(b + d) + 2\operatorname{diag}(b)L. \tag{23}$$

An expected row count vector obeys $m(t) = m(0)e^{tG}$. For two states let

$$K_i^+ = K_i + \varepsilon_i(1, -1, -1, 1), \qquad \varepsilon_i \geq 0, \tag{24}$$

with all entries nonnegative. All mean trajectories are unchanged, but for every initial type and $t \geq 0$,

$$F_i^+(t, 0) \geq F_i(t, 0). \tag{25}$$

The same weak ordering holds for eventual extinction. If $d_S \neq d_T$ and $b_i\varepsilon_i > 0$, then

$$F_i^+(t, 0) - F_i(t, 0) = \frac{b_i\varepsilon_i(d_S - d_T)^2}{3}t^3 + O(t^4), \tag{26}$$

so the finite-time inequality is strict for sufficiently small positive t.

Proof. A type-i event contributes $e_j - e_i$ for a switch, $-e_i$ for death, and $e_j + e_k - e_i$ for division. Taking conditional expectations gives row i of (23). Yule domination gives finite first moments and justifies the stopped generator calculation and its limit, so $m' = mG$. The perturbation (24) preserves row mass and each daughter marginal, hence G.

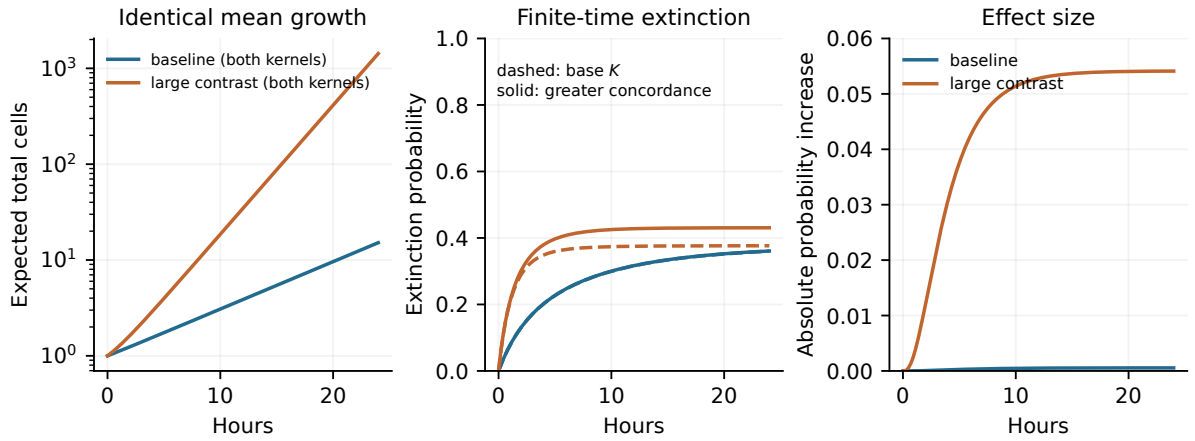

Figure 3: Model predictions, not biological measurements. Each kernel pair has exactly the same mean generator, so its two mean curves coincide (left). Numerical solutions of the branching ODE give extinction (middle) and the absolute extinction difference (right). The baseline perturbation increases 24-hour extinction by approximately 0.0005464; the explicitly synthetic large-contrast source gives approximately 0.05412. The theorem proves the ordering, not these floating-point effect sizes. Parameters are specified in the text and retained with the figure data.

The offspring PGF difference is the exact polynomial identity

$$\sum_{jk}(K_i^+ - K_i)(j,k)x_j x_k = \varepsilon_i(x_S - x_T)^2 \geq 0. \tag{27}$$

Let f, f^+ be the right sides of (3). The unit cube is invariant: on a face $x_i = 0$ the field points inward, and on $x_i = 1$ the switch terms and $b_i(\sum K_i x_j x_k - 1)$ are nonpositive. Both fields are cooperative, because for $j \neq i$ their partial derivatives are q_{ij} plus a nonnegative derivative of the offspring polynomial on the cube. Equation (27) implies $f^+(x) \geq f(x)$ there.

Comparison follows by a strict-supersolution argument. On any fixed finite horizon, add $\eta e^{Ct}\mathbf{1}$ to the new solution, with C larger than a Lipschitz constant of f on a neighbouring positive cube. For sufficiently small $\eta > 0$ this is a strict supersolution of the old ODE. At a first contact with the old solution, the touching coordinate is equal and the other coordinates are no smaller, contradicting cooperativity and the strict differential inequality. Letting η decrease to zero proves (25). Extinction-by-time events increase with t; their limits yield the eventual ordering.

At zero, $F_i'(0,0) = d_i$, so $(F_S^+ - F_T^+)^2 = (d_S - d_T)^2 t^2 + O(t^3)$. The difference equation is the old field difference plus this forcing multiplied by $b_i\varepsilon_i$. The old field is Lipschitz, so Gronwall's inequality first gives $F^+ - F = O(t^3)$. Its old-field contribution to the derivative is then $O(t^3)$. Integrating the leading t^2 forcing proves (26). □

If $b_i = b$ and $d_i = d$ are state independent, the total count process is a birth–death process independent of Q, K, π. Common demographic clocks therefore give equality of all unmarked count laws. This is a zero-effect boundary: increased sister concordance need not strictly increase eventual or finite-time extinction in every source. Eventual strictness is not proved by the cubic expansion, since both models may eventually go extinct with probability one.

Figure 3 illustrates Theorem 6.1 on two sources. Using (18) with $q_{ST} = .18$ and $\varepsilon_S = \varepsilon_T = .05$, the 24-hour extinction increase is approximately 0.0005464. It is small and is not presented as practically important. Guided by the forcing in (27), a second synthetic source uses $q_{ST} = q_{TS} = .01$, $b = (.6, .4)$, $d = (.5, .005)$, $\pi_T = .35$, independent-daughter base marginals $L_S(T) = .25$, $L_T(T) = .75$, and $\varepsilon_i = .18$. All kernel entries remain positive. Its 24-hour increase is approximately .05412; mean growth is still identical within the pair. No compatible

Table 3: Parameter provenance for an AU565-inspired assay. Numerical model inputs are synthetic unless explicitly described as reported.

Quantity	Value or requirement	Status and source
Context	AU565; lapatinib or gemcitabine lineage imaging	Reported architecture, Mohammadi et al. [9], methods
Cadence/horizon	Images every 30 minutes; AU565 follow-up 96 hours	Reported; not the instantaneous event detection assumed here
Recorded reporter	Nuclear/cell-cycle reporter and terminal fates	Reported; not a calibrated tolerance-state channel
States S, T	Predictive phenotypic response classes	Assumed model; no validated AU565 mapping supplied
Marker E	$e_S = .1$, $e_T = .85$	Synthetic; labelled or known-mixture calibration required
Rates and K	Equation (18); hours and inverse hours	Synthetic, not estimated from AU565 records
Clock/delay	$\lambda \in \{.25, .5, 1, 2\}$ per hour; known τ	Design choices; detect interruptions explicitly
Independent units	Independent founders/blocks	Required; related cells in one lineage are not iid roots

biological calibration is known here to justify that larger death contrast. A long-horizon prediction additionally assumes that the short-assay source continues to apply.

The implication is specific: expected burden alone can miss an extinction-relevant component of inheritance. A clinical recommendation would require further assumptions about intervention effects, toxicity, competition and the persistence of the fitted source, all outside this paper.

7 Experimental meaning, model challenges and practical scope

7.1 A single biological context, without invented calibration

Consider the AU565 lineage-imaging architecture studied by Mohammadi et al. [9]: drug-treated and control cells are tracked for cell-cycle phase, division and fate. It is a concrete architectural starting point, not a calibrated instance of our phenotype model. Table 3 keeps source provenance separate from synthetic choices. We do not combine rates from unrelated cell lines to construct a supposedly calibrated organism.

With the marker and rate assumptions met, the experiment estimates both within-cell switching and division-associated daughter production. For the synthetic source (18), $L_S(T) = .26$ and $L_T(T) = .8$, so the inheritance contrast is .54 and the sister covariances are .0724, .06. Tolerant daughter-production intensities are $\zeta_S = .1144$ and $\zeta_T = .256$ per hour. They are equal across the two example arms by construction; induction differs by .8 per hour. In an experiment neither equality need be assumed. Selection, inheritance and induction may all coexist, and none of them alone identifies genetic versus epigenetic causation.

Multiplying every biological rate and λ by $c > 0$, and dividing τ, L_0 by c, leaves the ideal record law unchanged. Thus ratios such as switching/division, death/division and delay times event rate govern this design. Scaling time cannot validate those ratios or supply a missing marker. A population doubling time, a mean cell-cycle duration and an exponential single-cell division hazard are not interchangeable parameters.

7.2 Age dependence and what a phase channel must supply

Constant hazards are a substantive assumption. If the division time D has Laplace transform $\varphi(\lambda) = \mathbb{E}e^{-\lambda D}$ and there is no death or switching, then $\mathbb{P}(T < D) = 1 - \varphi(\lambda)$ and the one-state

exponential source inverse produces an apparent birth rate

$$b_{\mathrm{app}}(\lambda) = \lambda(A_\lambda^{-1} - 1) = \lambda \frac{\varphi(\lambda)}{1 - \varphi(\lambda)}. \tag{28}$$

For $D \sim \mathrm{Erlang}(2, \nu)$, $\varphi(\lambda) = (\nu/(\nu + \lambda))^2$ and $b_{\mathrm{app}}(\lambda) = \nu^2/(2\nu + \lambda)$. At $\nu = 1$, the deadline rates $\lambda = 1, 2$ give $1/3$ and $1/4$, so no common homogeneous one-state generator explains both deadline groups. This is a model obstruction, not sampling noise, and it is experimentally testable: it converts the cross-deadline check (11) into a falsification instrument.

Multiple deadline rates therefore measure aspects of the waiting-time distribution even when the Markov inverse fails. A continuum of Laplace transform values determines the waiting-time law under standard uniqueness conditions; finitely many noisy values do not identify an unrestricted distribution. Estimating a constrained waiting-time law is a well-posed future target, and it is not solved here.

A finite-phase extension is available but is not automatic. Expand the state to (phenotype, cell-cycle phase), specify the phase transitions and the division reset, and apply Theorem 2.1 *only if its full observation and preparation assumptions hold for the expanded chain.* Three conditions are easy to lose. First, the calibrated channel must distinguish all expanded states: if a two-state tolerance channel and a two-phase channel have independent errors conditional on the expanded state and each has rank two, their tensor channel has rank four; otherwise the joint channel must be calibrated directly. A phenotype-only marker has identical columns for the hidden phases and is not full rank, and FUCCI colour classes need not separate every hidden Erlang stage. Second, founder support must be positive on every expanded state: a cohort of newborn cells has zero mass in later phases, violating the positive founder masses used for row normalization, so an asynchronous calibrated preparation or a new reachability argument is required. Third, nondividing phases have $b_i = 0$, where only the zero exit mass is recovered. The initial age distribution matters in its own right: an asynchronously sampled cell has a residual lifetime distribution that need not equal its newborn lifetime distribution, and age stratification alone does not make an age-dependent hazard constant. Internal-state and age-dependent branching models are developed in [6]; a phenotype-plus-phase extension should be positioned against that literature.

Likewise, F_τ describes delayed *biological state evolution*, not reporter maturation. If two hidden states have identical emissions and identical aggregated transition, demographic and daughter laws under every admitted intervention, they decorate the same observed process: all record laws agree, while the founder split and hidden inheritance can vary.

7.3 Destructive readouts and well-level dependence

Destructive initial staining cannot be applied to the same founder that is then followed. The initial marker requirement therefore needs a compatible live readout, a justified alternative preparation design, or a different theorem. Randomizing fixation times at the well level does not by itself supply the missing initial measurement or an immediate daughter-state measurement.

If k cells share one random well deadline T_w, their outcomes may be independent conditional on T_w under an idealized source, but they are generally dependent after marginalizing it. For a category probability $p(T_w)$, independent wells and conditionally iid cells give

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(p(T_w))}{W} + \frac{\mathbb{E}[p(T_w)(1 - p(T_w))]}{Wk}. \tag{29}$$

Increasing cells per well does not remove the first term. One safe concentration construction treats each well's bounded average as a single independent observation and applies the bounded-difference argument across the W wells; a richer conditional-on-time likelihood can use within-well information more efficiently, but it is not the iid-founder multinomial analysed above. Arbitrary

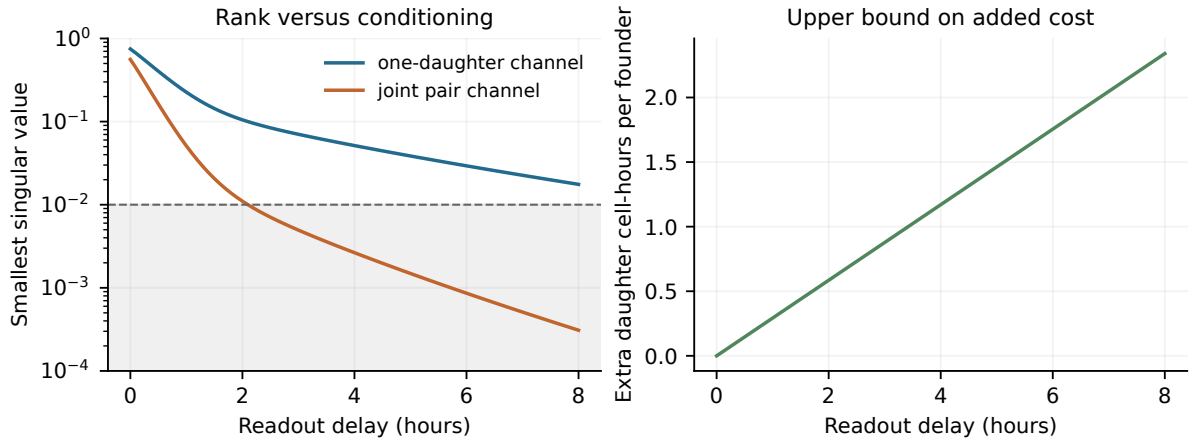

Figure 4: Delay conditioning for the treated source in (18). Singular values are numerical evaluations of the exact channel (12); the pair-channel value is their square. The dashed 0.01 spectral-error level illustrates where a norm perturbation bound alone cannot certify pair-channel rank; it is not an empirical calibration error and not a theorem that information vanishes at that delay. Right: additional daughter cell-hours per founder are bounded by $2\tau y_\lambda$, with $y_1 \simeq .14624$. Founder observation averages .79690 hours in this arm. Long delay retains structural rank but can destroy practical precision.

shared well effects can create clone-like correlations, and treatment randomization alone does not turn a mixture over environments into a homogeneous Markov source. Use measured or controlled blocks and replicate independent units at the level on which errors are computed.

Throughput is a real constraint, but no universal hardware ceiling follows from a platform name. What must be quantified for a specific instrument is the number of independent preparations, imaging fields, tracking failure rate, calibration acquisition, total cell-hours and the actual observed state channels.

7.4 A usable design decision

Table 4 distinguishes identification from a ready assay. A cap of $14/\lambda$ hours has reserve below 10^{-6}; known finite caps can instead be modelled exactly. Daughter monitoring may continue for τ after founder division, so elapsed follow-up is at most $L_0 + \tau$ under a cap. The independently labelled calibration samples in the worked example total 200,000; their acquisition cost is additional to founder observation time.

For an explicit reporter failure, suppose daughter interruption hazards are state independent and the delayed reporter has identical columns for the two birth states. The whole daughter channel then has identical columns. Any two kernels with the same row mass, including different sister covariances, give the same delayed pair law. Founder data can still recover H, b, d, but cannot repair that rank loss in the daughter readout. Thus an uncalibrated maturation channel cannot be substituted for the full-rank channel in Theorem 3.1.

For every limitation there is a specific observation that would address it: an initial or live state channel for preparation; labelled or independently characterized mixtures for calibration; demographic flags for denominator integrity; joint daughters for K; independent blocks for well effects; phase information or waiting-time profiles for age dependence. More uninformative replicates do not resolve an exact alias.

The experimental conclusion is conditional but specific: with calibrated state channels, intact denominators and the stated stopping rule, switching and daughter production are separately recoverable, and joint daughter observations support an extinction consequence invisible to mean growth. Without those channels, the paper identifies concrete ambiguities rather than promising retrospective recovery from data that never measured them.

Table 4: Design requirements and failure signs. Sample counts are independent roots or labelled calibration readouts, not descendant counts.

Design or resource	What it supplies and costs	Failure sign / limit
One binary endpoint	One count-threshold outcome	Preparation/rate alias, Proposition 1.1
Randomized stopped record	Initial marker, event class, endpoint or joint daughters; mean founder time $\leq 1/\lambda$	Missing denominators or marker rank
Six-category record	Switching target only, no daughter phenotypes, Proposition 4.5	Cannot recover b, d, K or predict extinction
Monitored delay	Each daughter: marker at τ or interruption; at most $2\tau y_\lambda$ extra cell-hours/root	Missed interruptions, unknown reporter memory
Two arms and calibration	Worked example: 50,000 roots/arm, 100,000 labelled reads/state	Report unresolved if the determinant bound reaches zero
Prospective decision	Narrow class of §4.6: 1.5×10^5 roots/arm and reads/state	Sufficient for that class only; not a lower bound
Calibration material	Pure states, or two known mixtures, Proposition 5.1	Circular if gated on the same marker
Division information	At $\lambda = 1$, treated division yield .14624; needed for K	Rare divisions weaken kernel recovery
Physical cap	$L_0 = 14/\lambda$, reserve $< 10^{-6}$ in each feature	Treating a capped kernel as an exact resolvent
Multiple deadline rates	Intersect common-source confidence sets	Incompatible H, B, π; cause not uniquely identified
Extinction prediction	Identified joint K plus continued source validity	Extrapolation beyond validated times/conditions

A Numerical certification and implementation details

The exact rational population law for a randomized deadline is particularly convenient. Rational λ, H, B, E, π give rational matrices $A_\lambda, D_\lambda, J, V$; no exponential approximation is used for those laws. Directed checks include a three-state cycle with

$$H = \begin{bmatrix} -3 & 2 & 0 \\ 0 & -3 & 2 \\ 2 & 0 & -3 \end{bmatrix}, \quad \lambda = \tfrac{3}{2}, \quad A_\lambda = \frac{1}{665} \begin{bmatrix} 243 & 108 & 48 \\ 48 & 243 & 108 \\ 108 & 48 & 243 \end{bmatrix}.$$

The spectrum of H is $-1, -4 \pm i\sqrt{3}$. Rational matrix multiplication verifies (6) exactly. Reducible and stiff-rate checks also include a zero eigenvalue, showing why inversion of H itself is unnecessary. These examples validate implementation conventions; the integral proof of Theorem 2.1 supplies the general result.

Interval operations are endpoint inclusion operations on exact fractions. For a positive rational x, the upward square-root approximation is $(\lfloor 10^{12}\sqrt{x} \rfloor + 1)/10^{12}$, obtained by integer square root of the scaled numerator and denominator. Matrix products are finite sums of interval products. A nonpositive lower determinant or marker contrast leads to an unbounded result, never an arbitrarily chosen denominator floor. For the prospective boxes, interval inversion of the two-by-two $\lambda I - H$ precedes the observed-law and target maps. This enclosure need not be sharp; its inclusion property and the explicit threshold separation are the certificate. No finite grid is substituted for a continuous parameter box.

Chernoff sets require a certified enclosure of $\{q : n\,\mathrm{kl}(\hat{p}\|q) \leq L\}$. We locate each endpoint by floating-point bisection, move it outward to a rational with denominator 10^9, and then *verify* that the candidate lies outside the set by evaluating a rational lower bound on $\mathrm{kl}(\hat{p}\|\cdot)$. That

lower bound uses two-sided rational bounds on the logarithm: $\log x = 2\,\mathrm{artanh}\,\frac{x-1}{x+1}$ with the odd series truncated after a fixed number of terms and the remainder bounded by the geometric tail $|z|^{2N+1}\big((2N+1)(1-z^2)\big)^{-1}$, preceded by exact halving of the argument into $[2/3, 4/3]$ and a rational enclosure of $\log 2$ obtained from the same series at $z = 1/3$. Since kl is a positive combination of two logarithms with known signs, term-wise lower bounds give a sound lower bound on kl. A separate audit re-evaluates every certified endpoint at 200 decimal digits and confirms that it lies outside the exact Chernoff set.

The script `check_paper.py` distributed with the source replays every constant printed above: the retained dataset regenerated from its seed, the six-category totals, the soundness of the logarithm and kl bounds, the containment of each Chernoff set in its Hoeffding radius, the four rows of Table 2 with their width reductions, the unresolved small-sample example, the exact resolvent checks, the prospective budgets, the delay channel values, the Erlang obstruction, the extinction effect sizes and the Monte Carlo diagnostic. All 86 checks pass.

B Reproducibility and proof status

The chronological construction, the observation laws, the resolvent integral, matrix-log uniqueness, confidence coverage and the cooperative comparison are conventional mathematical proofs in this paper. Lean verifies a declared finite algebra subset: resolvent cancellation, the rational two-state target identity, rectangular delayed-channel cancellation, the offspring polynomial identity, and — new in this version — the cancelled-determinant identity (15) in the form of its two marker-block lemmas, the known-mixture calibration formulas (22), the contamination bound of Remark 5.2, and the six-category subtraction identity of Proposition 4.5. Successful compilation of that subset does not constitute a compiled proof of the stochastic theorems, and this paper must not be described as machine verified. In particular, Lemma 4.2, Theorem 4.4 and Proposition 4.6 are conventional proofs.

The pinned environment uses Lean 4.30.0 and its frozen Mathlib dependency. All five modules, including the new `PhenotypeIdentification/CancelledTarget.lean`, compile with warnings promoted to errors, exit code zero and empty output. An axiom probe confirms that each of the six new declarations depends only on `propext`, `Classical.choice` and `Quot.sound`; no custom mathematical axiom and no incomplete proof is admitted. The script `lean_verify.py` reproduces the compilation and the probe and writes `lean_receipt.json` with exit codes and source hashes.

The source package is self-contained: `model.py` holds the synthetic source, the stopped observation law and the branching ODE; `certkit.py` holds exact interval arithmetic, the rational logarithm bounds, the Chernoff sets and the two target evaluators; `prospective_boxes.py` holds the class-level decision; `make_figures.py` regenerates the four vector figures; and `check_paper.py` replays every number. Source-record simulation uses seed 20260922; the 300-repetition coverage and power diagnostic regenerates the calibration sample and both arms each time. Exact binomial intervals describe Monte Carlo uncertainty only. Extinction plots use numerical ODE solutions, not interval-certified effect magnitudes.

All examples are synthetic. No experimental data were fitted, no empirical rate ranges are claimed, and no clinical conclusion is inferred. Author information, affiliations and a data and code availability statement are to be completed before submission.

References

[1] Daniela Bertacchi and Fabio Zucca. Strong survival and extinction for branching random walks via new order for generating functions. *Journal of Applied Probability*, 63(3):950–968, 2026. doi: 10.1017/jpr.2025.10070. arXiv:2403.01565.

[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001.

[3] P. J. Fitzsimmons and Jim Pitman. Kac's moment formula and the Feynman–Kac formula for additive functionals of a Markov process. *Stochastic Processes and their Applications*, 79 (1):117–134, 1999. doi: 10.1016/S0304-4149(98)00081-7.

[4] Aurélien Garivier and Olivier Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th Annual Conference on Learning Theory (COLT 2011)*, volume 19 of *Proceedings of Machine Learning Research*, pages 359–376, 2011. arXiv:1102.2490.

[5] James M. Greene, Jana L. Gevertz, and Eduardo D. Sontag. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. *JCO Clinical Cancer Informatics*, 3:1–20, 2019. doi: 10.1200/CCI.18.00087.

[6] Einar Bjarki Gunnarsson, Jasmine Foo, and Kevin Leder. Statistical inference of the rates of cell proliferation and phenotypic switching in cancer. *Journal of Theoretical Biology*, 568: 111497, 2023. doi: 10.1016/j.jtbi.2023.111497. arXiv:2306.08096.

[7] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

[8] Philipp Metzner, Eike Dittmer, Tobias Jahnke, and Christof Schütte. Generator estimation of Markov jump processes. *Journal of Computational Physics*, 227(1):353–375, 2007. doi: 10.1016/j.jcp.2007.07.032.

[9] Farnaz Mohammadi, Shakthi Visagan, Sean M. Gross, Luka Karginov, J. C. Lagarde, Laura M. Heiser, and Aaron S. Meyer. A lineage tree-based hidden Markov model quantifies cellular heterogeneity and plasticity. *Communications Biology*, 5:1258, 2022. doi: 10.1038/s42003-022-04208-9.

[10] Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890, 2020. doi: 10.1073/pnas.1922664117. arXiv:1912.11436.

[11] Chenyu Wu, Einar Bjarki Gunnarsson, Jasmine Foo, and Kevin Leder. A statistical framework for detecting therapy-induced resistance from drug screens. *npj Systems Biology and Applications*, 11(1):88, 2025. doi: 10.1038/s41540-025-00560-8. arXiv:2409.18879.