Almost a Policy

Transmission through Administrative Handoffs

Chae-Yeon Xon

Yonsei University

September 2026

Abstract

Which records should an agency preserve for a given evaluation question? This paper follows target information through administrative handoffs and studies retention under a record budget. A local ledger distinguishes identification from precision and gives the score dimensions required for repair. An Oregon Medicaid lottery exercise evaluates a constructed release. In its Gaussian summary model, approval following invitation loses one direction in the aggregate and four across groups. Under independent-person information, a 25-percent link budget gives an equal-weight group covariance criterion 1.04 times full linkage with targeted retention, compared with 1.83 under uniform retention.

JEL Codes: C14, C31, C44, H51, I13, I18

Keywords: administrative handoffs, partial identification, information retention, record design, Oregon Health Insurance Experiment, policy transmission

Yonsei University. Email: eunixon@yonsei.ac.kr. This draft is circulated for discussion; comments and suggestions are welcome.

Which records should an agency preserve for a given evaluation question? A receiving office may observe the approval rate among applicants while the records connecting invitations to applications remain with the producing agency. The approval rate describes adjudication. Approval following an invitation also depends on the probability of applying. With limited record capacity, the agency chooses which of these distinctions to carry across the handoff. The value of that choice depends on what the receiving evaluator will estimate.

The analysis holds the evaluation target fixed and takes the available retention rules as given. The program produces linked records upstream. The design problem begins after those records have been generated. A release rule determines the experiment available downstream, and the receiving unit may collect further evidence. The resulting ledger locates the target directions lost at a handoff and recovered by new measurement. It also distinguishes the information needed for local identification from the information needed to attain upstream precision.

A two-step application process gives the basic idea. Let a be the probability of applying after an invitation and b the probability of approval among applicants. Approval following an invitation is $d = ab$. A report of b retains the adjudication margin. In a local experiment that leaves a unrestricted, the missing target derivative is $b\,da$. One application-score coordinate therefore suffices to restore the aggregate target. Four separately valued group probabilities generally require four such coordinates. The target determines which part of the missing record matters.

The Oregon Health Insurance Experiment supplies a linked setting for this exercise (Finkelstein et al. 2012). Its public-use records connect lottery selection, notification, reported receipt, application, adjudication, and later Medicaid enrollment.[1] I evaluate a constructed handoff[2] that delivers conditional approval rates and preserves the lottery–enrollment experiment while holding applicant totals and invitation–application links upstream. The lottery supplies experimental variation in access. The reporting rule supplies the information comparison. The exercise keeps the program and its realized records fixed as it changes the evaluator's record.

The retained enrollment experiment supports its assignment-effect target. In the application–adjudication model, the same release has loss rank one for aggregate approval following invitation and four for the four reception-group probabilities. These groups use language requested and phone availability at signup.[3] Their estimated application and approval margins show why a conditional approval report can give an incomplete account of access: the phone comparison is concentrated at application, whereas the language comparison is concentrated at adjudication among applicants. These are descriptive comparisons within the selected cohort. The conditional approval rates also reflect who reaches adjudication. Estimating the effect of application assistance or a change in review procedures would require variation in those interventions and follow-up on their outcomes.

[1] The stage labels follow the recorded sequence in the public-use files. Random assignment occurs at the lottery stage.

[2] The handoff is an analytical release built from linked public-use variables; its stages correspond to observed program records.

[3] Both group indicators are pre-adjudication fields in the linked records and index reception probabilities. The lottery remains the assigned intervention.

A record budget gives the distinction a practical scale. Under the independent-person information benchmark,[4] a 25-percent person-link budget yields an equal-weight group covariance criterion 1.04 times its full-linkage value with targeted retention, compared with 1.83 under uniform retention. This criterion sums the variance bounds for the four group probabilities. For the population approval probability, whose weights are population shares, the two allocations both give about 1.80 at the same budget. The retained records are valued for a specified question. Changing that question changes the value of reallocating links.

The theory makes this comparison explicit. A parameter-independent transformation describes the record released at each handoff. Fisher-score null spaces track first-order ambiguity, and the target derivative maps that ambiguity into the quantities of interest. A loss–acquisition identity assigns changes to named stages. Within regular Gaussian retention channels, one theorem gives a lower bound and an attaining channel for identification-preserving retention; a second gives the corresponding minimum dimension for full-information precision. A convex frontier then allocates feasible retained information under a declared cost constraint and supplies its shadow value. In Oregon, that constraint is an expected share of auditable person links.

The empirical assessment keeps the observation design close to each claim. Household-adjusted lottery regressions describe assignment responses; stage and contact-group comparisons describe observed continuation; and the retention frontier is a model-based calculation on linked records. Simultaneous intervals, a bounded stage-category contamination exercise, and missing-record bounds assess the four-group rank conclusion. Twenty household partitions with rotated training and evaluation folds assess allocation stability.[5] They provide a held-out design check alongside sampling uncertainty for the estimated frontier.

Work on take-up and implementation shows how administrative procedures enter the delivery of benefits (Bhargava and Manoli 2015; Muralidharan et al. 2016; Deshpande and Li 2019; Giannella et al. 2024). Studies combining experiments, administrative records, and structural models make the link from a measured response to a feasible institutional choice explicit (Duflo et al. 2012; Basri et al. 2021; Handel and Kolstad 2015). The choice here concerns the record passed to an evaluator. The agency controls a retention rule, the target specifies the value of the retained distinctions, and the budget limits the available links. Research on approval organizations likewise assigns information and control to named participants (Henry and Ottaviani 2019); the present analysis takes the evaluation target and reporting authority as given.

Comparison of experiments supplies the information ordering (Blackwell 1953; Goel and DeGroot 1979), and missing-information identities separate the retained score from the discarded increment (Louis 1982). Target-oriented compression and optimal design provide related tools (Alsing and Wandelt 2018, 2019; Attia et al. 2018; Elfving 1952; Sagnol 2011). The analysis brings these requirements to a named reporting stage: it asks which target-relevant score directions are

[4]The benchmark treats retained person links as conditionally independent within the stated stage and reception group. It is the information model used for the allocation calculation.

[5]The forty train–evaluation comparisons reuse the same source sample and should be read as design repetitions on that sample.

held there, which channels can preserve them, and what precision those channels provide at a given cost. This also places the exercise alongside statistical-agency design and record-linkage research (Abowd and Schmutte 2019; Lin and Kolaczyk 2024; Bailey et al. 2020; Tahamont et al. 2021).

I. Records, delivery, and the reporting experiment

A. Linked records and the assignment design

Oregon opened a reservation list for the Oregon Health Plan Standard program in 2008 and selected names by lottery in eight draws. Selection offered household members an opportunity to apply, followed by eligibility review. The design and subsequent outcomes are described by Finkelstein et al. (2012). The public-use files link reservation-list characteristics, lottery draw, notification date, initial mail survey, application, application-decision date, approval, and Medicaid enrollment through person and household identifiers (National Bureau of Economic Research 2026).

The initial mail-survey frame contains 58,405 reservation-list members in 50,094 households, including selected members and controls. Its receipt item asks, "As of today, have you received an application?" Documented receipt equals a returned item reporting receipt and is defined for every member of this frame. It consequently combines response and reported receipt. Administrative enrollment is measured in its corresponding linked frame.

For outcome $H_i^{(k)}$, the reported assignment regression has the form

$$H_i^{(k)} = \alpha_k + \beta_k Z_i + \gamma_{k,d(i)} + \eta_{k,s(i)} + u_i^{(k)}, \tag{1}$$

where Z_i is lottery selection, $d(i)$ is lottery draw, and $s(i)$ is listed household size. Scores are summed within households. Table 1 gives the four assignment coefficients with simultaneous intervals from their joint household-score covariance.

TABLE 1. Lottery-assignment responses in the linked outcome frames

Outcome	Control	Selected	Effect [simultaneous 95% CI]	F
Survey return	45.6	44.9	-0.8 [-1.8, 0.3]	2.8
Receipt-item response	44.6	44.2	-0.4 [-1.5, 0.7]	0.9
Documented receipt	6.8	31.3	24.7 [24.0, 25.5]	5745.1
Medicaid enrollment	14.1	39.1	25.6 [24.7, 26.4]	5316.4

Notes. Each row is the coefficient on lottery selection in a linear probability model with lottery-draw and household-size fixed effects. Scores are aggregated by household. The four intervals use one Gaussian household-score max-t critical value (2.389) from 99,999 draws; adjusted p-values use the same reference distribution. Survey outcomes use the initial mail-survey frame, and enrollment uses the administrative enrollment frame. F is the squared household-clustered t statistic for the assignment coefficient.

Selection raises documented receipt by 24.7 percentage points and enrollment by 25.6 points. The survey-return and receipt-item response coefficients are −0.8 and −0.4 point, respectively. Their simultaneous intervals include zero. The squared clustered t statistics measure assignment relevance for the recorded outcomes. Causal responses to receipt, application, or adjudication would use additional variation in those stages.

Timing records describe the intervening process. Among 13,069 selected respondents with both dates observed, reported receipt rises from 15.0 percent at days 0–28 after notification to 89.9 percent at days 56–84. The median notification-to-decision interval is 76 days among 18,100 selected applicants with a decision date on or after notification. Survey receipt is current status at a selected response date. These timing summaries describe the observed chain; Supplemental Appendix A reports the full figures, continuation comparisons, and an arrival-response extension.

B. Application, adjudication, and the target

The access sample contains 29,799 of 29,834 selected people with complete application and approval fields, in 24,901 households. The four reception groups are defined by requested language and phone availability at signup, both recorded before selection. Let A_i indicate application and Y_i approval for an applicant. For group g, write

$$(2) \qquad a_g = \mathbb{P}(A_i = 1 \mid g, \text{selected}), \qquad b_g = \mathbb{P}(Y_i = 1 \mid A_i = 1, g, \text{selected}), \qquad d_g = a_g b_g.$$

The group target is $d = (d_1, \ldots, d_4)'$. The population target is $\bar{d} = \sum_g p_g d_g$, where p_g is the selected-population group share. These are approval probabilities following an invitation. The lottery effect on subsequent Medicaid enrollment is a separate target based on the retained assignment–enrollment experiment.

TABLE 2. Reception of the Oregon Medicaid invitation

Lottery-list record	People	Applied	Approved | applied	Approved
English requested, phone	23,216	62.2	48.8	30.4
English requested, phone absent	3,663	51.0	51.2	26.1
Other language requested, phone	2,652	62.5	37.3	23.3
Other language requested, phone absent	268	54.1	46.2	25.0

Notes. Rates are percentages among lottery-selected people with complete administrative application and approval fields. The conditional-approval column uses submitted applications as its denominator. Approval is the application decision recorded by Oregon. Confidence intervals and tests resample households within lottery draw.

In the complete-record sample, 60.8 percent apply and 48.0 percent of applicants receive approval, giving approval following invitation of 29.2 percent. Within English-requested records, application is 11.2 percentage points lower when the phone field is empty, whereas conditional approval is 2.4 points higher. Among records with a phone number, application differs by 0.3

point across language requests and conditional approval is 11.6 points lower for other-language requests. The two contact fields locate different margins in the observed process. Their effects as interventions remain a separate empirical question.

C. The constructed reporting handoff

The producing agency holds the invitation–application–adjudication links. The proposed downstream record contains group invitation counts n_g and conditional approval rates, together with the retained lottery–enrollment experiment. Applicant totals and invitation–application links remain upstream. This is a reporting counterfactual evaluated on the linked source records. The public-use archive provides the source experiment used to construct it.

To specify the conditional-rate statistic, let $K_g = \sum_{i:g_i=g} A_i$ and $R_g = \sum_{i:g_i=g} A_i Y_i$. The release rounds $\widehat{b}_g = R_g/K_g$ to a public grid:

$$\widetilde{b}_g = \delta_g \operatorname{round}(\widehat{b}_g/\delta_g), \qquad \delta_g = n_g^{-3/4}. \tag{3}$$

The grid depends on the invitation count. Appendix A.9 derives the statistic's Gaussian local limit, with a mean shift in b_g and zero first-order mean shift in a_g. The identification and precision calculations use this local summary experiment. The finite-sample law retains random-denominator dependence, and the scope of the Gaussian approximation is kept explicit.

Table 3 anticipates the target comparison. In the two-step score model, aggregate approval loses one application-score direction and the four-group vector loses four. The enrollment contrast continues to use its retained experiment. Each row holds its target and statistical model fixed. Additional restrictions linking the enrollment and application–adjudication laws would change the experiment available for the approval calculation.

TABLE 3. Target-specific ranks at the reporting handoff

Target	Estimate	Loss rank	Repair ranks
Lottery effect on Medicaid enrollment	25.6 [24.9, 26.3]	0	0 / 0
Approval following invitation, population	29.2 [28.6, 29.8]	1	1 / 1
Approval following invitation, four groups	23.3–30.4	4 {4}	4 / 4

Notes. The comparison uses the Gaussian conditional-rate summary model for approval and the retained lottery–enrollment experiment for the assignment contrast. Applicant totals and invitation–application links remain upstream. The first estimate is the coefficient on lottery selection in a regression with lottery-draw and household-size fixed effects; its interval clusters by household. The other intervals resample households within draw. Repair ranks report minimum identification rank followed by minimum full-precision rank. The braces give the 95% confidence set for the four-group loss rank. Estimates are percentage points in the first row and percentages in the remaining rows.

The receiver may be authorized to obtain target-specific summary counts or to retain auditable person links. Summary counts are priced here by local score dimension; person links are priced by expected retention capacity. Sending group application totals and retaining person links therefore belong to different menus. Each comparison holds its menu fixed. A link permits audit and later record-level analysis, while a target summary preserves the named statistical distinction.

II. Target information at a handoff

The analysis uses a finite-dimensional regular local experiment, with quadratic-mean differentiable scores and a common normalization of information. Identification refers to first-order target distinguishability in its Gaussian limit. Let the local parameter be $\theta \in \Theta \subseteq \mathbb{R}^p$, and let the policy counterfactual be $\tau : \Theta \to \mathbb{R}^q$ with derivative $C = D\tau(\theta_0)$. A stage contains a handoff and a measurement. The upstream experiment is $Y_{j-1} \sim P_{\theta,j-1}$. A θ-independent Markov kernel K_j produces the retained record Z_j, and the receiving unit then observes X_j with conditional law $Q_{\theta,j}(\cdot \mid Z_j)$. The experiment available after the stage is $Y_j = (Z_j, X_j)$.[6]

Write I_{j-1} for upstream Fisher information and I_j^- for the information in Z_j. Data processing gives

$$L_j = I_{j-1} - I_j^- \geq 0.$$

The score of the conditional measurement X_j has conditional mean zero given Z_j, so it is orthogonal to the score of Z_j. If its expected conditional information is $A_j \geq 0$, the stage ends with

$$I_j = I_j^- + A_j.$$

The pair (L_j, A_j) is the information ledger of stage j. The first entry records information removed at the handoff, and the second records information produced by measurement at the receiving unit. A sequence containing only garblings has $A_j = 0$ at every stage. A receiving office can have both entries: it receives a summary and then collects documents or decisions of its own.

Define

$$N_{j-1} = \ker I_{j-1}, \qquad N_j^- = \ker I_j^-, \qquad N_j = \ker I_j = N_j^- \cap \ker A_j$$

and their target images

$$M_{j-1} = C(N_{j-1}), \qquad M_j^- = C(N_j^-), \qquad M_j = C(N_j).$$

The inclusions $N_{j-1} \subseteq N_j^-$ and $N_j \subseteq N_j^-$ yield two nonnegative ranks:

$$\ell_j = \dim M_j^- - \dim M_{j-1}, \qquad \rho_j = \dim M_j^- - \dim M_j.$$

[6]The Markov-kernel ordering follows the comparison-of-experiments formulation in Blackwell (1953). Goel and DeGroot (1979) discuss related information measures, while Basu (1978); Andhivarothai and Ramamoorthi (1994) survey partial-sufficiency formulations.

The loss rank ℓ_j counts target directions removed by the handoff. The acquisition rank ρ_j counts target directions recovered by the receiving unit's measurement. Both are relative to the derivative C and the maintained observation law. A change in either object calls for a new ledger.

PROPOSITION 1 (Loss–acquisition ledger). *For every finite sequence of regular stages,*

$$\dim M_J - \dim M_0 = \sum_{j=1}^{J} (\ell_j - \rho_j).$$

The ranks are invariant to smooth invertible reparameterizations of θ*, invertible recodings of each experiment, and invertible linear changes of target coordinates.*

The identity attributes final target ambiguity to named handoffs and measurements. A stage may remove several target directions and recover some of them from evidence produced there, giving a net change of $\ell_j - \rho_j$. If every A_j vanishes, the null spaces form a filtration and the identity reduces to the sum of handoff losses.

A. Identification and precision

The loss rank records a change in identification. Fisher information also measures weakening inside the identified region. For a positive-semidefinite information matrix I, the target is identified in the Gaussian local experiment when

$$\ker I \subseteq \ker C.$$

Under this range condition its local asymptotic covariance bound is

$$\mathbb{V}(I, C) = C I^{\dagger} C',$$

where $I^{\dagger}$ is the Moore–Penrose inverse. Let $\Omega > 0$ specify the units in which components of the target are valued and define

$$\mathcal{R}_{\Omega}(I, C) = \operatorname{tr}\{\Omega \mathbb{V}(I, C)\}.$$

The value is $+\infty$ when the range condition fails. For a scalar contrast $w'\tau$, the corresponding target strength is $\kappa(I; w)^2 = \{w' \mathbb{V}(I, C) w\}^{-1}$.

The weight matrix has a decision interpretation. Let an institution choose $a \in \mathbb{R}^q$, and suppose local loss around the full-information action $\tau(\theta)$ is $\frac{1}{2}\{a - \tau(\theta)\}' H \{a - \tau(\theta)\}$ with $H > 0$. With per-unit information, $\mathbb{V}$ bounds the covariance of $\sqrt{n}(\widehat{\tau} - \tau)$. The corresponding n-scaled quadratic regret bound is $\frac{1}{2}\operatorname{tr}(H\mathbb{V})$. Setting $\Omega = H/2$ makes $\mathcal{R}_{\Omega}$ the local asymptotic regret bound, while W in Section B records the resource cost of retaining score information.

At stage j, the risk created by the handoff and the risk removed by measurement are

$$\Delta^{\text{loss}}_{j,\Omega} = \mathcal{R}_{\Omega}(I_j^{-}, C) - \mathcal{R}_{\Omega}(I_{j-1}, C), \qquad \Delta^{\text{acq}}_{j,\Omega} = \mathcal{R}_{\Omega}(I_j^{-}, C) - \mathcal{R}_{\Omega}(I_j, C).$$

Both quantities are nonnegative whenever their endpoints are finite. The exact ranks (ℓ_j, ρ_j) locate changes in target identification, while $(\Delta^{\text{loss}}_{j,\Omega}, \Delta^{\text{acq}}_{j,\Omega})$ record precision changes that preserve rank. This pairing separates a handoff that deletes a target direction from a handoff that leaves the direction weakly measured.

COROLLARY 1 (Reception regret). *Under the local quadratic decision problem with $\Omega = H/2$, $\Delta^{\text{loss}}_{j,\Omega}$ is the increase in the receiver's local asymptotic regret bound created by the handoff, and $\Delta^{\text{acq}}_{j,\Omega}$ is the decrease produced by measurement at the receiving stage.*

The corollary gives a decision interpretation once the action and its local loss have been specified. For the reporting exercise, the receiver estimates a declared evaluation target, and H values errors in that estimate. A household application decision or an agency eligibility decision would supply its own action, loss, and observation experiment. The Oregon covariance comparisons use the stated target weights; a welfare interpretation would additionally specify the losses those weights represent.

Sampling uncertainty in rank. A thresholded rank estimator uses spectral separation. A confidence set instead inverts uncertainty about the primitive matrices. Supplemental Appendix C states the conditions for each.[7] The Oregon quotient map is diagonal; simultaneous intervals for its four coefficients give the rank conclusion directly.

III. Retention under a record budget

A. A channel drawn from the upstream experiment

Fix a handoff and write $I = I_j^-$ and $L = L_j$, so the upstream information is $I + L = I_{j-1}$. Its local Gaussian limit admits independent central sequences with information I and L. Write the lost increment as

$$U \sim \mathcal{N}(L^{1/2}h, \mathbf{I}_p)$$

under local parameter h. A regular k-coordinate retention channel has Gaussian limit $T_R = R'U$, where R has k columns. Its information is

$$Q_R = L^{1/2} P_R L^{1/2},$$

with P_R the orthogonal projection onto the column space of R. This definition ties the channel to data held upstream: T_R is a statistic of the lost score increment generated by a parameter-independent Markov kernel.

[7] For inference on directionally differentiable maps and numerical implementations at spectral boundaries, see Fang and Santos (2019); Hong and Li (2018); Chen and Fang (2019) develops rank inference that includes lower-rank strata.

THEOREM 1 (Minimum identification-preserving retention). *Among regular Gaussian retention channels, the minimum score rank required to preserve the upstream local identification of the target is*

$$k^* = \dim C(\ker I) - \dim C\{\ker(I+L)\} = \ell_j.$$

There is a k^*-coordinate channel satisfying

$$C\{\ker(I+Q_R)\} = C\{\ker(I+L)\},$$

and every channel satisfying this equality has score rank at least k^*.

The theorem concerns a statistic constructed from the upstream experiment and preserves the target's local identification set. Matching precision within that set imposes an additional requirement. The proof works on $\ker I/\ker(I+L)$. The quadratic form induced by L is an inner product on this quotient, so every target functional there can be represented as a linear functional of $L^{1/2}h$. A basis of the target image gives ℓ_j retained score coordinates. Score rank measures local statistical dimension; a database may encode the resulting statistic in several fields or in one composite record.

Full-information precision generally requires a larger channel. Let $F = I + L$ and suppose $F > 0$. Define

$$k^{\text{prec}} = \operatorname{rank}\{L^{1/2}F^{-1}C'\}.$$

THEOREM 2 (Minimum precision-preserving retention). *Among Gaussian retention channels satisfying* $\ker(I+Q_R) \subseteq \ker C$*, the minimum score rank that attains the upstream covariance bound*

$$C(I+Q_R)^{\dagger}C' = CF^{-1}C'$$

is k^{prec}. An attaining channel has information

$$Q^{\text{prec}} = L^{1/2}P_G L^{1/2}, \qquad G = L^{1/2}F^{-1}C',$$

where P_G projects onto the column space of G. Moreover,

$$k^{\text{prec}} \geq \ell_j.$$

The matrix G collects the lost-score components used by the full-information efficient influence function for the target. Retaining their span reproduces the upstream covariance bound. This gives two design thresholds. The first restores local identification; the second restores full-information precision. The repair frontier below traces the interval between them.

COROLLARY 2 (Placement). *Every statistic formed solely from the retained experiment has a null space containing* $\ker I$ *and target ambiguity containing* $C(\ker I)$*. Preservation of the directions in*

$C(\ker I)/C\{\ker(I+L)\}$ therefore occurs at the handoff or upstream. Evidence collected after the handoff enters the ledger as acquisition information A_j.

The placement statement distinguishes two operations with the same eventual identification outcome. Retention carries an available distinction across a handoff. Acquisition creates a measurement at the receiving stage. Their costs, legal authorities, and data requirements can differ even when $I+Q$ and $I+A$ have the same numerical value.

B. The information menu and vector repair frontier

Gaussian garblings of the lost increment generate the information menu

$$\mathcal{Q}(L) = \{Q : 0 \preceq Q \preceq L\}.$$

Every element can be written $Q = L^{1/2}PL^{1/2}$ for a positive-semidefinite contraction P. Its score rank is $\operatorname{rank} Q$, and a Gaussian channel with that score rank implements it. This is the full Gaussian-channel menu. A record-retention rule induces an element of $\mathcal{Q}(L)$; its operational constraints determine which elements are implementable by that rule. In an applied design, the map from records to Q supplies the units of the budget. The Oregon exercise uses independent stratified retention of person-level links, so q_g is an expected record share and the budget is the expected fraction of links retained.

PROPOSITION 2 (Stratified link retention). *Suppose the lost score is a sum of independent person-level contributions from groups $g = 1, \ldots, G$, with group share p_g and conditional information L_g. Retain each upstream–downstream link in group g independently with probability q_g, and record the sampling indicator. The retained links contribute information*

$$Q(q) = \sum_{g=1}^{G} p_g q_g L_g$$

per person and use expected record capacity $\sum_g p_g q_g$. Thus a capacity bound B gives the feasible set $\sum_g p_g q_g \le B$, with $0 \le q_g \le 1$.

The sampling rule uses group membership and an external random draw. Its information formula maintains independent person-level lost scores. The Oregon calculation uses this working score model; household resampling then quantifies sampling variation in the estimated model criterion. Group weights in the target determine the allocation across q_g, while p_g converts each allocation into an expected number of retained links.

Let $W \succ 0$ price retained information, let b be a budget, and let $\Omega \succ 0$ weight a q-dimensional counterfactual. Define the repair frontier

$$\Phi(b; I, L, W, C, \Omega) = \min_{Q,V} \operatorname{tr}(\Omega V)$$

subject to

$$\begin{pmatrix} I+Q & C' \\ C & V \end{pmatrix} \geq 0, \qquad 0 \leq Q \leq L, \qquad \operatorname{tr}(WQ) \leq b.$$

The value is $+\infty$ when finite V is infeasible. The block constraint then records failure of target identification directly.

THEOREM 3 (Vector repair frontier). *The displayed program equals*

$$\min_{0 \leq Q \leq L,\, \operatorname{tr}(WQ) \leq b} \mathcal{R}_\Omega(I+Q, C).$$

Its value is weakly decreasing and convex in b. Full retention gives $\Phi\{\operatorname{tr}(WL)\} = \mathcal{R}_\Omega(I+L, C)$. Suppose an optimizer Q_b satisfies $I + Q_b > 0$ and Slater's condition. With

$$H_b = (I+Q_b)^{-1} C' \Omega C (I+Q_b)^{-1},$$

there are $A_b, S_b \geq 0$ and $\lambda_b \geq 0$ such that

$$-H_b + \lambda_b W - A_b + S_b = 0,$$

$$\operatorname{tr}(A_b Q_b) = 0, \qquad \operatorname{tr}\{S_b(L - Q_b)\} = 0, \qquad \lambda_b\{\operatorname{tr}(WQ_b) - b\} = 0.$$

At differentiability points,

$$\partial_b \Phi = -\lambda_b, \qquad D_L \Phi[D] = -\operatorname{tr}(S_b D)$$

for a feasible positive-semidefinite menu expansion D.

The scalar case $q = 1$ gives the c-optimal design criterion (Elfving 1952; Sagnol 2011). A vector target can allocate score rank across several counterfactual coordinates. The matrix S_b records which unavailable score directions constrain the target, while λ_b gives the marginal reduction in the covariance criterion from relaxing the stated budget. The shadow value inherits the units of W and Ω. In the person-link specialization, the relevant resource is expected link capacity.

With orthogonal lost increments across handoffs, the institution can solve one program with $Q_j \leq L_j$ and $\sum_j \operatorname{tr}(W_j Q_j) \leq B$. Each stage-specific cap names the lost increment from which a retention channel can be drawn. The orthogonality assumption makes these increments additive in the stated local experiment. Sequential information acquisition and rational-inattention models price experiments or posterior beliefs (Miao et al. 2022; Pomatto et al. 2023; Hébert and Woodford 2023); the caps here attach those choices to the score increments removed at specific handoffs.

Estimated allocation frontiers. Frontier inference uses the stated uncertainty conditions. The empirical exercise distinguishes population score dimension, a feasible link-retention rule, and evaluation after that rule has been selected.

IV. The Oregon reporting and retention exercise

A. The two-step information benchmark

The target and release are given in equations (2) and (3). Conditional on group counts, the working independent-person likelihood is

$$\prod_{g=1}^{4}\prod_{i:g_i=g} a_g^{A_i}(1-a_g)^{1-A_i}\{b_g^{Y_i}(1-b_g)^{1-Y_i}\}^{A_i}. \tag{4}$$

Its application and adjudication scores are orthogonal by conditional centering. Total sample information in the two blocks has entries

$$I_{a,g}=\frac{n_g}{a_g(1-a_g)},\qquad I_{b,g}=\frac{n_g a_g}{b_g(1-b_g)}. \tag{5}$$

The retained Gaussian summary experiment uses the adjudication block. Application scores remain upstream. The derivative of $d_g = a_g b_g$ on this missing block is b_g, so the quotient target map is $\operatorname{diag}(b_1, b_2, b_3, b_4)$.

The benchmark conditions on group counts and uses independent person-level scores. It evaluates population-target weights at the stated group shares. Joint estimation of population composition would add the score and target derivative for those shares. Household resampling propagates the estimated rates and shares through the reported criterion.[8] A household-robust efficiency analysis of an implemented retention rule would further specify its cluster-level score experiment or sandwich covariance.

B. Target loss and the scope of repair

Every estimated b_g lies between 0.373 and 0.512. Simultaneous household-bootstrap intervals exclude zero for all four coefficients. In the maintained diagonal quotient map, inversion gives the reported 95-percent loss-rank set $\{4\}$. The diagonal form and the missing application block come from the specified release model. The intervals assess how far its four target coefficients lie from the lower-rank boundary. The application-score direction for aggregate approval is instead $(p_1 b_1, \ldots, p_4 b_4)$ and has rank one. Theorems 1 and 2 give the corresponding repair ranks in Table 3. The orthogonal model makes identification and full-precision repair ranks coincide here; the general theory allows the two thresholds to differ.

A producer authorized to send target summaries can add the four group application totals. Together with conditional approval, those totals carry the four application-score directions. The person-link frontier values auditable microdata under a limit on retained links. Its cost unit is an

[8]The information model defines the bound; household resampling describes uncertainty in the estimated model criterion.

expected number of links. The coordinate threshold and the link budget describe the same target under their respective retention technologies.

The loss-rank conclusion is stable within two declared record-error exercises. Thirty-five of the 29,834 selected records have incomplete application or approval fields; completing those records within the logical constraint that approval requires application leaves the aggregate stage rates in narrow ranges. A second exercise combines simultaneous sampling lower bounds with bounded replacement of stage categories within fixed reception groups. All four conditional-approval coordinates remain positive at replacement radii of 0.1, 0.5, and 1 percent.

C. Allocating an auditable-link budget

Retain each invitation–application link in group g with probability q_g, independently of outcomes conditional on the group. Under Proposition 2, the application block becomes $q_g I_{a,g}$ and expected capacity is $\sum_g p_g q_g$. With equal weight on the four approval probabilities, the model covariance criterion multiplied by N is

$$\mathcal{V}(q) = \sum_{g=1}^{4} \left\{ \frac{b_g^2 a_g(1-a_g)}{p_g q_g} + \frac{a_g b_g(1-b_g)}{p_g} \right\}. \tag{6}$$

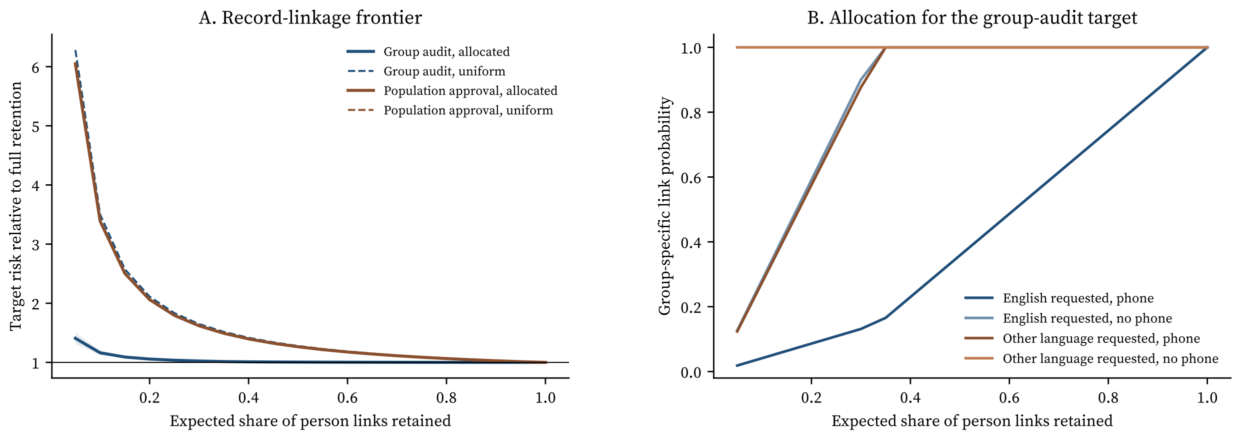
Only the first term varies with application-link retention. For budget $\sum_g p_g q_g \le B$, the optimum is

$$q_g(B) = \min\left\{1, \frac{b_g\{a_g(1-a_g)\}^{1/2}}{p_g\sqrt{\lambda(B)}}\right\}, \tag{7}$$

with $\lambda(B)$ chosen to exhaust the budget. The factor p_g enters twice in the first-order condition: it converts group sample size into target variance and group retention probability into total record capacity.

At 25 percent capacity, the allocation probabilities are 0.109, 0.746, 0.726, and 1.000 in the group order of Table 2. Equal weights value precision for each group probability separately. The group with 268 people is consequently fully retained, using less than one percent of all person links. This is where the group objective and the record budget interact. Figure 1 displays the allocation and its covariance criterion.

FIGURE 1. Target precision and the allocation of person links



Notes. Panel A gives model covariance criteria relative to full linkage. Solid lines allocate link probabilities by target risk; dashed lines use a common fraction across groups. The shaded band reports the simultaneous 95% household-bootstrap band for the estimated group-audit criterion. Panel B gives group-specific retention probabilities. The information benchmark maintains independent person-level scores; bootstrap uncertainty concerns the estimated criterion.

TABLE 4. Precision under a record-linkage capacity constraint

Linked record share	Audit, allocated	Audit, uniform	Population, allocated	Population, uniform
10%	1.16	3.50	3.39	3.40
25%	1.04	1.83	1.80	1.80
50%	1.01	1.28	1.26	1.27

Notes. Entries are independent-person model covariance bounds divided by the corresponding full-linkage bound. Audit gives equal weight to the four group approval probabilities. Population uses the approval probability for the selected population. Allocated retention solves the record-share frontier; uniform retention links the displayed share in every group.

For the group target, the reported ratio to full linkage is 1.04 under allocation by target weight and 1.83 under uniform retention. Fourteen percent capacity reaches 1.10 times the full-linkage group bound. Population approval requires 72 percent capacity to reach that ratio. At 25 percent, its allocated and uniform ratios both round to 1.80. The population comparison shows how the selected objective shapes the gains from reallocating links. Equal group weights protect the four subgroup probabilities. Population weights protect the aggregate approval probability. The 1.04 ratio belongs to the former objective and its working covariance bound. Relative to uniform use of the same link budget, this is about a 43-percent reduction in the group covariance criterion. The largest retention probability is assigned to a small group whose precision receives the same objective weight as that of a large group. The comparison values subgroup estimation accuracy under the stated score model.

An agency implementing this comparison would first choose aggregate approval or the four-group vector as its evaluation target, declare the eligible record cohort, and estimate the allocation inputs from pilot records. It would then sample links with the stated group probabilities and retain those probabilities with the released records. This sequence holds the question and sampling rule fixed when evaluating the resulting release. The budget measures links under equal unit record cost; monetary costs, privacy constraints, and institutional authority would further restrict an operational menu. The Gaussian frontier and the link-sampling frontier each apply to their declared menu.

D. Sampling, allocation selection, and interpretation

The rank intervals and frontier bands resample households within each of the eight lottery draws. Thus 29,799 people describe the access outcome frame, while 24,901 households supply the independent resampling units.[9] Positive coefficients and the scalar capacity constraint give a unique allocation in this specialization. The reported household bootstrap propagates uncertainty through estimated application rates, conditional approval rates, and group shares. General directional-delta inference additionally allows changing active constraints.

The fitted allocation is also evaluated on held-out households. Twenty seeds partition households within draw; each training fold estimates the 25-percent allocation and the other fold evaluates its covariance criterion, after which the roles reverse. Across the forty rotated comparisons, median allocated/full risk is 1.037 and the maximum is 1.044, compared with median uniform/full risk of 1.830 at the same held-out capacity. The allocated criterion is lower in all forty comparisons. These overlapping partitions check sensitivity to fitting the allocation on a particular part of the source sample. Each evaluation targets the same working covariance criterion. The forty comparisons share observations across partitions; the frontier band provides the reported sampling uncertainty. Supplemental Table A8 gives the split protocol and results.

Inverse lottery-selection weights and draw deletions provide additional cohort checks. Weighted group approval rates differ from the selected-cohort rates by at most 1.5 percentage points. Deleting one draw at a time gives group ranges of 29.3–31.6, 25.2–27.3, 22.1–24.0, and 22.0–27.1 percent. These checks retain the defined selected-cohort target and the same four reception groups.

The exercise therefore has three evidentiary layers. Lottery assignment supports the reported access responses under its design. Linked stage records describe application and adjudication within the selected cohort. The reporting calculation evaluates the information and precision associated with a constructed release under an explicit score model. An implemented change in contact services, eligibility review, or reporting practice would supply a further intervention and its own outcomes.

[9] People index stage outcomes; households are the independent units used by the reported bootstrap.

V. Conclusion

The record delivered to an evaluator helps determine which questions about a program remain answerable. In the Oregon reporting exercise, the enrollment experiment is retained, while approval following an invitation requires information on both application and adjudication. The same handoff therefore has different consequences for the enrollment contrast, aggregate approval, and the four-group approval vector.

A target-specific information ledger organizes these consequences. Its loss and acquisition ranks count changes in first-order target ambiguity. Retention channels then have two thresholds: preservation of upstream identification and restoration of upstream precision. A feasible menu and a declared cost turn those thresholds into a design problem.

Under the independent-person information benchmark, targeted use of a 25-percent person-link budget gives an equal-weight four-group covariance criterion 1.04 times its full-linkage value. Uniform retention gives 1.83. The population target gives much closer ratios across the two rules. The size of the gain follows from the target weights, group sizes, and the precision supplied by a retained link.

An operational application starts with that question, the information held at each office, and the retention or acquisition rules available there. Sampling uncertainty, linkage quality, and household dependence then enter the evaluation of a proposed design. Naming these objects makes the reporting handoff a part of the empirical design itself.

A. Core derivations and proofs

A.1. Proof of the loss–acquisition ledger

Proof of Proposition 1. The Markov handoff gives $I_{j-1} \succeq I_j^-$, hence $N_{j-1} \subseteq N_j^-$ and $M_{j-1} \subseteq M_j^-$. Positive semidefiniteness gives $\ker(I_j^- + A_j) = \ker I_j^- \cap \ker A_j$, hence $N_j \subseteq N_j^-$ and $M_j \subseteq M_j^-$. Thus ℓ_j and ρ_j are nonnegative and

$$\ell_j - \rho_j = \dim M_j - \dim M_{j-1}.$$

Summation telescopes. Under a local reparameterization $\phi = R\theta$ with invertible R, null spaces become RN and the target derivative becomes CR^{-1}, giving $(CR^{-1})(RN) = C(N)$. An invertible recoding preserves the experiment and its Fisher null space. An invertible target transformation maps every M bijectively to its transformed image. Each dimension difference is therefore invariant. □

A.2. Proof of reception regret

Proof of Corollary 1. For a local decision loss with Hessian H, the Gaussian local asymptotic risk contribution of the target error is $\mathcal{R}_{H/2}(I, C)$. Substitution of $\Omega = H/2$ into the definitions of $\Delta^{\text{loss}}_{j,\Omega}$ and $\Delta^{\text{acq}}_{j,\Omega}$ gives the two changes in the attainable regret bound. □

A.3. Proof of minimum upstream repair

Proof of Theorem 1. Set $K = \ker I$ and $K_0 = \ker(I + L) = \ker I \cap \ker L$. On the quotient $\mathcal{K} = K/K_0$, define $\langle [h], [v] \rangle_L = h' L v$. This expression is well defined because $Lk_0 = 0$ for $k_0 \in K_0$, and it is positive definite on $\mathcal{K}$. The target induces

$$\bar{C} : \mathcal{K} \longrightarrow C(K)/C(K_0), \qquad \bar{C}[h] = Ch + C(K_0),$$

with rank $\dim C(K) - \dim C(K_0) = \ell_j$.

A k-coordinate channel $T_R = R'U$ induces $\bar{B}_R[h] = R' L^{1/2} h$ on $\mathcal{K}$. Its augmented null space is

$$\ker(I + Q_R) = K \cap \ker(R' L^{1/2}).$$

Equality of upstream and repaired target ambiguity implies $\ker \bar{B}_R \subseteq \ker \bar{C}$. Rank–nullity then gives $k \geq \operatorname{rank} \bar{B}_R \geq \operatorname{rank} \bar{C} = \ell_j$.

For attainment, choose a linear map $B : \mathcal{K} \to \mathbb{R}^{\ell_j}$ with $\ker B = \ker \bar{C}$. By the Riesz representation theorem on $(\mathcal{K}, \langle \cdot, \cdot \rangle_L)$, each coordinate has a vector $[v_s]$ satisfying $B_s[h] = v_s' L h = (L^{1/2} v_s)' L^{1/2} h$. Let the columns of R be $L^{1/2} v_s$. Then $\ker \bar{B}_R = \ker \bar{C}$. Since K_0 lies in the channel null space, this equality yields $C\{\ker(I + Q_R)\} = C(K_0)$. The constructed channel has ℓ_j coordinates. □

A.4. Proof of minimum precision repair

Let $B = I + Q$ and $F = I + L$, with $F \succ 0$. Normalize the full experiment by setting $\widetilde{B} = F^{-1/2}BF^{-1/2} \preceq \mathbf{I}_p$ and $\widetilde{C} = CF^{-1/2}$. The target is identified under B when $\operatorname{range}(\widetilde{C}') \subseteq \operatorname{range}(\widetilde{B})$. On this range, every eigenvalue of $\widetilde{B}^\dagger$ is at least one. Hence $CB^\dagger C' = CF^{-1}C'$ holds exactly when every column of $\widetilde{C}'$ belongs to the unit-eigenvalue space of $\widetilde{B}$. Transforming back gives

$$(L - Q)F^{-1}C' = 0. \tag{A.1}$$

Represent any feasible Q as $L^{1/2}PL^{1/2}$ with $0 \preceq P \preceq \mathbf{I}_p$. Positive semidefiniteness and (A.1) imply

$$(\mathbf{I}_p - P)L^{1/2}F^{-1}C' = 0.$$

Thus P acts as the identity on the column space of $G = L^{1/2}F^{-1}C'$, which gives $\operatorname{rank} Q \geq \operatorname{rank} G$. Setting $P = P_G$ attains this lower bound and satisfies (A.1), proving the first claim.

For the rank comparison, take w in the null space of G. For every $h \in \ker I$,

$$w'Ch = h'FF^{-1}C'w = h'IF^{-1}C'w + h'LF^{-1}C'w = 0.$$

The first term vanishes because $Ih = 0$, and the second vanishes because $Gw = 0$. Hence $\ker G \subseteq \{w : w'C(\ker I) = 0\}$, and rank–nullity gives $\operatorname{rank} G \geq \dim C(\ker I) = \ell_j$ because $F \succ 0$ implies $\ker F = \{0\}$.

A.5. Proof of placement

Proof of Corollary 2. Let T be any statistic formed from the retained experiment. The data-processing inequality gives $I_T \preceq I$, hence $\ker I \subseteq \ker I_T$. Applying C gives $C(\ker I) \subseteq C(\ker I_T)$, so a statistic of the retained data preserves the ambiguity created at the handoff. Preserving the quotient directions in $C(\ker I)/C\{\ker(I + L)\}$ therefore requires a channel at or before the handoff. Information generated later is represented by the receiving-stage acquisition term. □

A.6. Gaussian information menu

Information contraction gives $Q \preceq L$ for every Gaussian garbling of the lost increment, and Fisher information is positive semidefinite. Conversely, fix $0 \preceq Q \preceq L$ and set $P = L^{\dagger/2}QL^{\dagger/2}$ on the range of L. Then $0 \preceq P \preceq \mathbf{I}_p$ on that range and $Q = L^{1/2}PL^{1/2}$. For an independent $\varepsilon \sim \mathcal{N}(0, \mathbf{I}_p)$, observe

$$T = P^{1/2}U + (\mathbf{I}_p - P)^{1/2}\varepsilon.$$

The covariance of T is $\mathbf{I}_p$, its local mean derivative is $P^{1/2}L^{1/2}$, and its Fisher information is $L^{1/2}PL^{1/2} = Q$. Diagonalizing P and discarding zero-score coordinates gives score dimension $\operatorname{rank} Q$.

A.7. Proof of the vector frontier

For $A = I + Q \succeq 0$, the generalized Schur complement gives

$$\begin{pmatrix} A & C' \\ C & V \end{pmatrix} \succeq 0 \iff \operatorname{range}(C') \subseteq \operatorname{range}(A), \qquad V - CA^{\dagger}C' \succeq 0.$$

Since $\Omega \succ 0$, minimization over V selects $V = CA^{\dagger}C'$ whenever the range condition holds. This proves equivalence with the weighted target-risk program and assigns value $+\infty$ when the target lacks identification.

The feasible correspondence expands with b, so the value decreases. If (Q_1, V_1) and (Q_2, V_2) are feasible at b_1 and b_2, their convex combination is feasible at the corresponding convex combination of budgets. The objective is linear in V, which proves convexity. At $b = \operatorname{tr}(WL)$, $Q = L$ is feasible and dominates every feasible information matrix, giving the full-retention endpoint.

Under $I + Q_b \succ 0$, differentiation of $\operatorname{tr}\{\Omega C(I+Q)^{-1}C'\}$ with respect to Q gives $-H_b$. Attach multipliers $A_b \succeq 0$ to $-Q \preceq 0$, $S_b \succeq 0$ to $Q - L \preceq 0$, and $\lambda_b \geq 0$ to $\operatorname{tr}(WQ) - b \leq 0$. Stationarity and complementary slackness give the system in Theorem 3. The envelope formula gives the two directional prices.

A.8. Proof of stratified link retention

Let S_i be the sampling indicator for a person in group g, with $\mathbb{P}(S_i = 1 \mid g) = q_g$, independent of the record and the local parameter. If s_{ig} is the lost score contribution, the observed contribution is $S_i s_{ig}$. Its mean is zero and its information is $\mathbb{E}[S_i s_{ig} s_{ig}'] = q_g L_g$. Averaging over groups gives $\sum_g p_g q_g L_g$. The expected number of retained links divided by the person count is $\sum_g p_g q_g$.

A.9. The conditional-rate summary experiment

The following result concerns the first-order limit of the rounded statistic. It supplies the Gaussian summary benchmark used in the application.

PROPOSITION 3 (Gaussian limit of a conditional-rate release). *Let $(A_i, A_iY_i)_{i=1}^n$ be independent with $\mathbb{P}(A_i = 1) = a$ and $\mathbb{P}(Y_i = 1 \mid A_i = 1) = b$, where (a, b) lies in a compact subset of $(0,1)^2$. Set $K_n = \sum_i A_i$, $R_n = \sum_i A_iY_i$, and $\widehat{b}_n = R_n/K_n$, using any fixed value on $K_n = 0$. For a public grid width δ_n satisfying $\sqrt{n}\delta_n \to 0$ and $n\delta_n \to \infty$, release $\widetilde{b}_n = \delta_n \operatorname{round}(\widehat{b}_n/\delta_n)$. Under $(a_n, b_n) = (a_0 + h_a/\sqrt{n}, b_0 + h_b/\sqrt{n})$,*

$$\sqrt{n}(\widetilde{b}_n - b_0) \rightsquigarrow \mathcal{N}\left(h_b, \frac{b_0(1-b_0)}{a_0}\right).$$

The limiting Gaussian shift has information $\operatorname{diag}\{0, a_0/[b_0(1-b_0)]\}$ for (a, b). In that limit experiment, the adjudication direction is measured and the application direction has zero information.

A.10. Proof of the conditional-rate release

Write $\overline{A}_n = K_n/n$ and $\overline{R}_n = R_n/n$. The map $\psi(x, y) = y/x$ has gradient $(-b/a, 1/a)$ at (a, ab). The joint influence function of $\psi(\overline{A}_n, \overline{R}_n)$ is

$$\frac{A_i(Y_i - b)}{a},$$

which has mean zero, variance $b(1 - b)/a$, derivative zero in the a direction, and derivative one in the b direction. The multivariate central limit theorem and the delta method under the stated local alternatives give the displayed limit for $\widehat{b}_n$ under (a_n, b_n). Rounding satisfies $\sqrt{n}|\widetilde{b}_n - \widehat{b}_n| \leq \sqrt{n}\delta_n/2 \to 0$, so Slutsky's theorem gives the same limit for the released statistic. The additional condition $n\delta_n \to \infty$ makes the grid coarser than the $1/n$ count scale. The resulting Gaussian shift has information $a_0/[b_0(1 - b_0)]$ for b and zero information for a. On every compact interior parameter set, $\mathbb{P}(K_n = 0) = (1 - a)^n$ converges to zero uniformly, so the convention on that event vanishes from the limit.

Weak convergence of the statistic gives the displayed Gaussian limit and its information matrix. Equivalence of the full sequence of released experiments to that limit would require a corresponding likelihood or experiment-convergence argument. The application uses the Gaussian summary model explicitly. In a finite release, the random denominator and the available companion records remain part of the observation law.

A.11. The Oregon likelihood and record-share allocation

The score components from equation (4) are

$$s_{a,i} = \frac{A_i - a_g}{a_g(1 - a_g)}, \qquad s_{b,i} = \frac{A_i(Y_i - b_g)}{b_g(1 - b_g)}.$$

Conditional centering of $Y_i - b_g$ among applicants gives $\mathbb{E}[s_{a,i}s_{b,i}] = 0$. Summing the marginal score variances gives equation (5). Under independent link sampling and the retained adjudication benchmark, the delta-method covariance for d_g is

$$\frac{b_g^2 a_g(1 - a_g)}{n_g q_g} + \frac{a_g b_g(1 - b_g)}{n_g}.$$

Multiplication by N and summation yields equation (6). Define $c_g = b_g^2 a_g(1-a_g)/p_g$. At an uncapped coordinate, the Lagrangian derivative is

$$-\frac{c_g}{q_g^2} + \lambda p_g = 0, \qquad q_g = \sqrt{\frac{c_g}{\lambda p_g}} = \frac{b_g\sqrt{a_g(1 - a_g)}}{p_g\sqrt{\lambda}}.$$

Capping at one and choosing λ to clear the record-share budget gives equation (7). The calculation prices a group retention probability by the group's share of all records.

A.12. Reading the diagonal loss-rank intervals

In the Oregon Gaussian summary model the target map on the discarded application block is

$$B(b) = \operatorname{diag}(b_1, b_2, b_3, b_4).$$

Let $[\underline{b}_g, \overline{b}_g]$, $g = 1, \ldots, 4$, be intervals with joint coverage at least $1 - \alpha$ for the four coefficients, under the sampling approximation used to construct them. Form the coefficient box $\mathcal{B} = \prod_g [\underline{b}_g, \overline{b}_g]$ and its rank image

$$\mathcal{R} = \{\operatorname{rank} B(\widetilde{b}) : \widetilde{b} \in \mathcal{B}\}.$$

Whenever the coefficient vector belongs to $\mathcal{B}$, its rank belongs to $\mathcal{R}$. Thus joint coefficient coverage passes directly to rank coverage. If all four lower endpoints are positive, every matrix in this box has rank four and $\mathcal{R} = \{4\}$. The reported Oregon intervals satisfy this condition. Their interpretation maintains the diagonal target map and the discarded application block specified in the main text. With coefficients strictly inside the parameter space, rank four follows from the model; the intervals quantify separation from its lower-rank boundary.

The household bootstrap describes sampling uncertainty in the coefficients entering this calculation. Its validity is a requirement on the simultaneous coefficient intervals. The rank-image argument transfers that coverage and leaves the observation model fixed. Supplemental Appendix C gives the corresponding confidence-region formulation for general primitive matrices.

References

Abowd, John M. and Ian M. Schmutte, "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review*, 2019, *109* (1), 171–202.

Alsing, Justin and Benjamin Wandelt, "Generalized Massive Optimal Data Compression," *Monthly Notices of the Royal Astronomical Society: Letters*, 2018, *476* (1), L60–L64.

— **and** — , "Nuisance Hardened Data Compression for Fast Likelihood-Free Inference," *Monthly Notices of the Royal Astronomical Society*, 2019, *488* (4), 5093–5103.

Andhivarothai, N. and R. V. Ramamoorthi, "Partial Sufficiency and Least Favorable Experiments," *Scandinavian Journal of Statistics*, 1994, *21* (4), 493–500.

Attia, Ahmed, Alen Alexanderian, and Arvind K. Saibaba, "Goal-Oriented Optimal Design of Experiments for Large-Scale Bayesian Linear Inverse Problems," *Inverse Problems*, 2018, *34* (9), 095009.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform? Lessons from U.S. Historical Data," *Journal of Economic Literature*, 2020, *58* (4), 997–1044.

Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken, "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia," *American Economic Review*, 2021, *111* (12), 3827–3871.

Basu, D., "On Partial Sufficiency: A Review," *Journal of Statistical Planning and Inference*, 1978, *2* (1), 1–13.

Bhargava, Saurabh and Dayanand Manoli, "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

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

Chen, Qihui and Zheng Fang, "Improved Inference on the Rank of a Matrix," *Quantitative Economics*, 2019, *10* (4), 1787–1824.

Deshpande, Manasi and Yue Li, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 2019, *11* (4), 213–248.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 2012, *102* (4), 1241–1278.

Elfving, Gustav, "Optimum Allocation in Linear Regression Theory," *The Annals of Mathematical Statistics*, 1952, *23* (2), 255–262.

Fang, Zheng and Andres Santos, "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 2019, *86* (1), 377–412.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group, "The Oregon Health Insurance Experiment: Evidence from the First Year," *The Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

Giannella, Eric, Tatiana Homonoff, Gwen Rino, and Jason Somerville, "Administrative Burden and Procedural Denials: Experimental Evidence from SNAP," *American Economic Journal: Economic Policy*, 2024, *16* (4), 316–340.

Goel, Prem K. and Morris H. DeGroot, "Comparison of Experiments and Information Measures," *The Annals of Statistics*, 1979, *7* (5), 1066–1077.

Handel, Benjamin R. and Jonathan T. Kolstad, "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare," *American Economic Review*, 2015, *105* (8), 2449–2500.

Hébert, Benjamin and Michael Woodford, "Rational Inattention When Decisions Take Time," *Journal of Economic Theory*, 2023, *208*, 105612.

Henry, Emeric and Marco Ottaviani, "Research and the Approval Process: The Organization of Persuasion," *American Economic Review*, 2019, *109* (3), 911–955.

Hong, Han and Jessie Li, "The Numerical Delta Method," *Journal of Econometrics*, 2018, *206* (2), 379–394.

Lin, Shurong and Eric D. Kolaczyk, "Data Privacy for Record Linkage and Beyond," Working Paper 32940, National Bureau of Economic Research 2024.

Louis, Thomas A., "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society: Series B*, 1982, *44* (2), 226–233.

Miao, Jianjun, Jieran Wu, and Eric R. Young, "Multivariate Rational Inattention," *Econometrica*, 2022, *90* (2), 907–945.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, "Building State Capacity: Evidence from Biometric Smartcards in India," *American Economic Review*, 2016, *106* (10), 2895–2929.

National Bureau of Economic Research, "Oregon Health Insurance Experiment Public Use Data," NBER Research Data 2026. Accessed August 29, 2026.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz, "The Cost of Information: The Case of Constant Marginal Costs," *American Economic Review*, 2023, *113* (5), 1360–1393.

Sagnol, Guillaume, "Computing Optimal Designs of Multiresponse Experiments Reduces to Second-Order Cone Programming," *Journal of Statistical Planning and Inference*, 2011, *141* (5), 1684–1708.

Tahamont, Sarah, Zubin Jelveh, Aaron Chalfin, Shi Yan, and Benjamin Hansen, "Dude, Where's My Treatment Effect? Errors in Administrative Data Linking and the Destruction of Statistical Power in Randomized Experiments," *Journal of Quantitative Criminology*, 2021, *37* (3), 715–749.