**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

# WHEN AUDIT FINDINGS CONCEAL SKILL GAPS: JOINT VERIFICATION AND TRAINING UNDER A SHARED TIME BUDGET

**Jumarto, Jumarto[1]; Andriansyah, Andriansyah[2]; Evi, Satispi[3]**
[1,2,3]*Doctoral Program in Public Administration, Faculty of Social and Political Sciences, Universitas Muhammadiyah Jakarta, Jakarta 15419, Indonesia*
Email: jumartorildo@gmail.com[1]

**Abstract:** Audit findings depend on both the occurrence of irregularities and the ability to detect them. This creates an identification problem when the same findings guide investment in auditor skills. We formulate a two-stage decision model in which blinded verification, routine auditing and training compete for fixed novice and expert time. A constructed pair of environments has identical single-review finding rates but different returns to training. We derive the resulting information-limited regret, specify the assumptions under which paired reviews distinguish the environments, and solve the finite verification-budget problem by exact multinomial enumeration. In the nominal benchmark, with 100 future cases per domain, two paired reviews increase expected unique detections from 52.680 to 52.903 while reducing distinct-case coverage from 226 to 224. Verification is not selected at shorter workloads, and the single-pair benefit reverses under modest conditional dependence between reviewers. Twenty thousand Monte Carlo episodes check the exact calculations without serving as additional empirical evidence. The contribution is a decision-focused benchmark linking imperfect detection to costly skill investment, not a new capture–recapture estimator or a field-validated personnel policy. Repeated adaptive allocation and institutional calibration remain outside the implemented model.

***Keywords (English):*** *Internal auditing; imperfect detection; verification budget; skill investment; Bayesian decision analysis; opportunity cost*

## INTRODUCTION

An audit unit observes findings, not every irregularity that exists. A low finding rate can therefore indicate a well-controlled process, an unsuitable audit procedure, or insufficient task-relevant expertise. These explanations imply different staffing decisions. The first may justify reducing coverage; the others may justify additional verification or professional development. Treating the observed finding rate as the underlying risk can select the wrong intervention even when the scheduling algorithm is implemented correctly.

This distinction changes the question addressed by an auditor-allocation model. Competency–project matching is already represented in the multi-objective framework of Wang, Ferreira and Yan (Wang, Ferreira, Yan, 2024). Their integrated audit-planning study also connects risk assessment, activity selection and scarce resources.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

Skill development is not a sufficient novelty claim either: recent project-scheduling research explicitly incorporates apprenticeship into workforce improvement (Mousazadegan et al., 2026). A further model combining allocation, risk and mentoring must explain a decision or identification problem that these familiar ingredients alone do not resolve (Kumar & Bawankar, 2025).

The present study examines one such problem. Audit outcomes may not reveal whether the organization needs more expertise or is simply facing fewer irregularities. Moreover, obtaining that information uses the same expert time that could deliver assurance or train a colleague. Verification is consequently both an information-acquisition decision and a capacity-allocation decision. Its value depends on what the organization will do with the information, how many future engagements can benefit, and which engagements are displaced while the information is collected.

The problem of incomplete detection is established, not introduced here. Feinstein separates the occurrence of noncompliance from its detection in detection-controlled estimation and tax-evasion research (Feinstein,. J, 1990). Contemporary work on tax auditing addresses a related but different issue: outcomes are observed for cases selected by human decision makers (Battaglini, Guiso, Lacava, Miller, & Patacchini, 2025). Selection into audit and imperfect detection within an audit should not be conflated. This study holds the initial case environment deliberately simple so that incomplete detection, rather than selection bias, generates the information limit (Lin & Zhan, 2025).

We develop a finite, two-stage benchmark. The first stage allocates a fixed number of novice and expert review slots between separate cases and blinded paired reviews. The second stage chooses whether to train in domain A, train in domain B, or preserve all remaining audit capacity. Both the learner's time and the instructor's time are charged. The latent environment determines the return to training but is not supplied to a feasible policy. Exact enumeration provides a transparent reference against which simulation and later approximate algorithms can be checked.

The contribution is deliberately bounded. We derive a training-regret result for observationally indistinguishable environments, connect the established paired-detection identity to a paid verification decision, and provide an exact restricted-design solver with falsification tests. We do not claim to invent partial observability, capture–recapture estimation, or the value of information. Nor do the synthetic results establish the performance of any defense inspectorate. The institutional application motivates the decision problem (Pfister & Hamann, 2023); it is not an empirical sample (Yang, 2023).

The relevant comparison is not whether earlier papers mention knowledge sharing. It is whether risk occurrence, reviewer sensitivity, verification effort and the future return to training are distinguished in the decision formulation. Table 1 identifies what the reviewed sources establish and the narrower question pursued here. Where only an abstract or publisher preview was available, the comparison is limited to that material; absence of a mechanism from an abstract is not proof that it is absent from the full paper.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

**Table 1** Closest research streams and the bounded contribution.

| Research stream | What prior work establishes | Difference examined here |
|---|---|---|
| Auditor assignment | Suitability and preferences in a multi-objective allocation model. | Our question is whether recorded findings identify the domain worth training. |
| Integrated audit planning | Risk assessment, activity selection and resource allocation. | Verification consumes the same capacity as assurance and skill investment. |
| Apprenticeship scheduling | Workforce skill development through apprenticeship. | Learning alone is not claimed as novelty; incomplete detection changes targeting. |
| Detection-controlled models | Occurrence and detection are different latent processes. | We derive a decision-regret consequence for paid training, not a new detection estimator. |
| Tax-audit prediction | Learning and policy evaluation from human-selected audits. | Selection into audit differs from misses within an audited case. |
| Capture–recapture and test accuracy | Identification depends on observation and error assumptions. | The established identity informs a costly verification decision; dependence is stress-tested. |
| Information and control | Value of information and belief-state decisions. | Exact results are restricted to a precommitted two-stage design. |

**Source:** Compiled by the authors based on the findings of the study, 2026.

Paired detection is mathematically related to two-list capture–recapture. That literature emphasizes matching, closure, independence and homogeneity (International Working Group for Disease Monitoring and Forecasting, 1995). In the present application, case identifiers provide matching, the underlying case is held unchanged between reviews, and each reviewer records a result before seeing the other's. These procedural safeguards support interpretation but do not establish statistical independence. Shared evidence, common training and unobserved case difficulty can still produce correlated errors (Wu, 2025).

Unknown false-positive rates create a further identification problem. Hui and Walter estimate test accuracy under specific multi-population assumptions (Hui & Walter, 1980), and subsequent work reviews the difficulties of estimation without a gold standard (Sun & Zhou, 2025). The implemented benchmark assumes perfect specificity. It does not import the Hui–Walter result into a setting with different reviewer sensitivities across domains, nor does it suggest that two opinions generally establish the truth (Tai & Zhou, 2025).

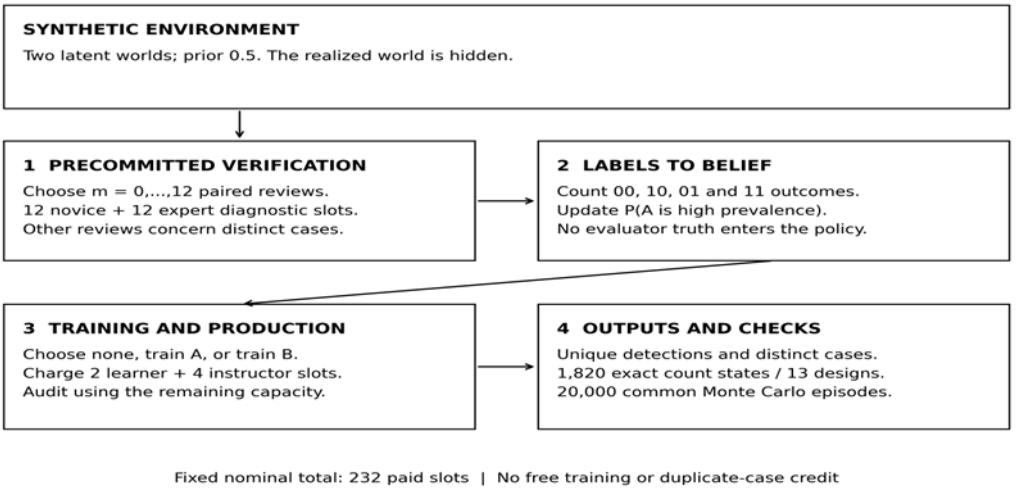
**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

Decision-theoretic information value is also well established (Blackwell, 1953). More information can improve a decision when obtaining it is free and the original action remains available. Here it is not free: duplicate reviews displace distinct cases. The computational question is whether better targeting of a paid skill investment compensates for that displacement. The value of information must be measured in the same units as the service capacity sacrificed to obtain it.

## RESEARCH METHODS

**Figure 1** sets out the implemented chronology and the information available at each step.

Picture 1 Implemented two-stage protocol.




**Source:** Compiled by the authors based on the findings of the study, 2026.

### Latent Irregularities and Observed Findings

Let Z denote the presence of an irregularity in a case, and let $D_r$ denote reviewer r's detection of that irregularity. Under the benchmark's no-false-positive assumption, the recorded finding is $Y_r = Z\,D_r$. Within a homogeneous domain, write q for prevalence and $s_r$ for sensitivity conditional on an irregularity (V, 2026). Then:

$$Y_r = ZD_r, \qquad \Pr(Y_r = 1) = qs_r. \tag{1}$$

Observed positives identify a product, not its factors. We use two domain environments with $q_H = 0.8$ and $q_L = 0.4$. A novice has a positive-finding probability a = 0.2 and an expert has probability b = 0.3 in either environment. Sensitivities are a/q and b/q. An expert remains more sensitive than a novice in each environment, but both perform less sensitively in the higher-prevalence environment. Table 2 specifies the complete construction.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

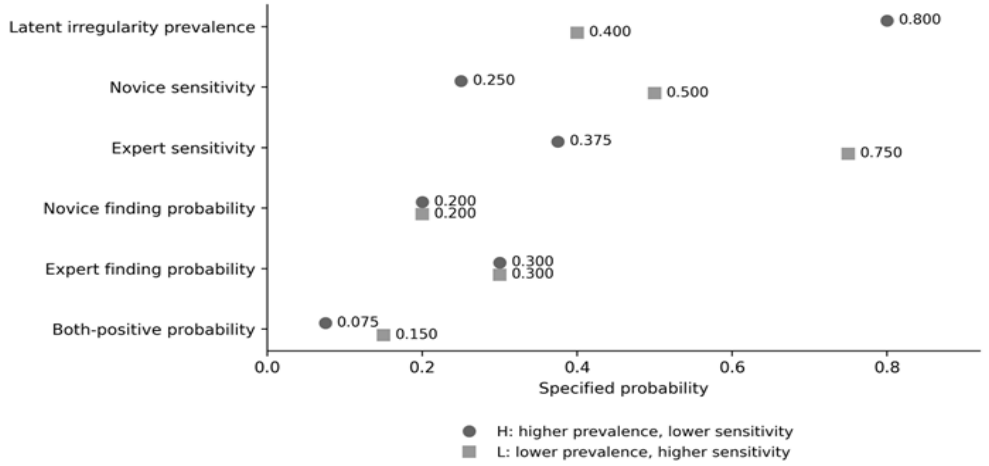
**Table 2** Constructed environments with identical single-review findings.

| State | q | Novice sensitivity | Expert sensitivity | Finding rates: novice / expert | Joint positive |
|---|---|---|---|---|---|
| H | 0.800 | 0.250 | 0.375 | 0.200 / 0.300 | 0.075 |
| L | 0.400 | 0.500 | 0.750 | 0.200 / 0.300 | 0.150 |

**All values are synthetic probabilities. In the two worlds, H and L exchange domains A and B. Both-positive rates assume conditional independence.**

The two possible worlds swap these environments between domains A and B. The prior probability that A is the high-prevalence domain is one half. This two-world catalog and the marginal finding probabilities are supplied as benchmark assumptions, not estimated from institutional records. They permit an exact information experiment. A real application would need a richer uncertainty model and evidence for its parameter ranges (Liu, 2023).

**Figure 2** contrasts the separate-review marginals with paired overlap




**Picture 2** The two synthetic environments share novice

### Two Stages and a Shared Time Budget

The diagnostic stage provides 12 novice slots and 12 expert slots. One slot completes one standardized review. Selecting m paired reviews uses m novice slots and m expert slots on the same m cases in domain A. Remaining slots audit distinct cases separately. A paired case contributes one detected irregularity when either reviewer finds it, never two when both do. Thus m is bounded by both role-specific capacities.

The production stage provides N novice slots in each domain and eight expert slots. At most one training intervention is permitted. Training consumes k = 2 novice slots in the selected domain and h = 4 expert slots. It raises the novice's subsequent sensitivity in that domain by $\eta = 0.10$. The trained domain consequently processes $N - k$ cases, the other domain N cases, and the expert

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

processes 8 − h cases. Without training, all these slots remain available for auditing. Training does not generate a finding by itself.

These are accounting constraints, not fatigue penalties. No policy receives extra hours, free mentoring, a recovery bonus, or an assumed reduction in paired-review duration. The unique-case count falls by one for every duplicate review and by k + h for every training intervention. The model abstracts from scheduling overhead beyond these charged slots. Adding such overhead would change the verification threshold; zero additional overhead is an assumption, not an observed efficiency.

For posterior probability π that A is the high-prevalence domain, the expected domain prevalences are $q_A(\pi) = q_L + (q_H - q_L)\pi$ and $q_B(\pi) = q_H - (q_H - q_L)\pi$. The incremental production yield from training in domain d, relative to no training, is:

$$g_d(\pi) = (N-k)\eta q_d(\pi) - ka - hb. \qquad \textbf{(2)}$$

The first term counts the higher sensitivity over the cases still audited. The second term counts detections forgone in the learner's displaced cases and the instructor's displaced expert cases. The terminal decision is the action with the largest value among zero, $g_A$ and $g_B$. Zero represents preserving capacity rather than training. Ties are resolved in favor of no training, then domain A.

**Blinded Paired Observations**

Under conditional independence of reviewer detections given Z = 1, a pair produces four outcomes. Write c = P($Y_1$ = 1, $Y_2$ = 1). In cell order (00,10,01,11), the observation law is:

$$P_q = (1-a-b+c,\ a-c,\ b-c,\ c), \qquad c = ab/q. \qquad \textbf{(3)}$$

Here c = ab/q. The two environments have the same marginals but different overlap probabilities. In the high-prevalence environment, c is 0.075; in the low-prevalence environment, it is 0.150. Blinded paired reviews therefore contain information that separate, pre-training reviews do not contain in this construction.

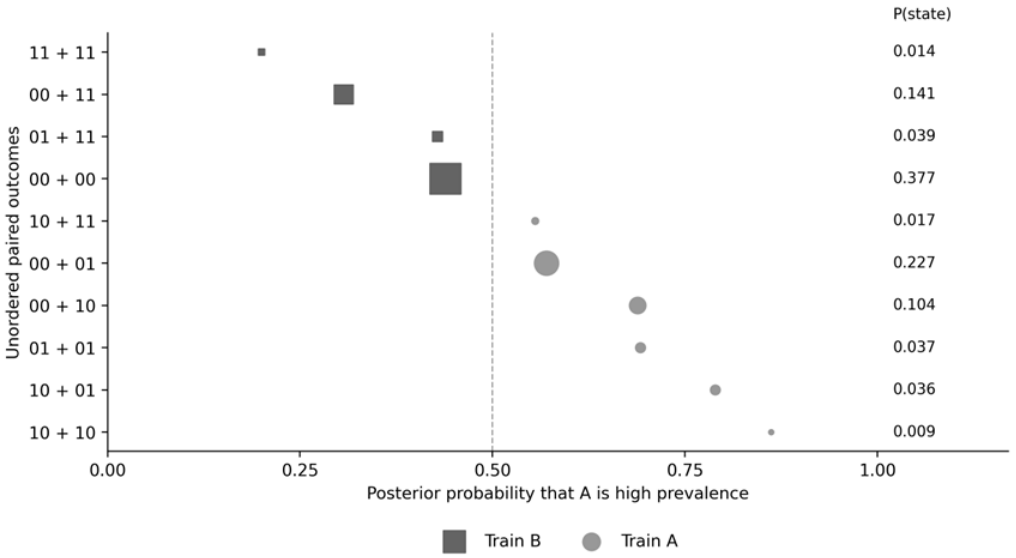
A count vector n = ($n_{00}$,$n_{10}$,$n_{01}$,$n_{11}$), with components summing to m, has a multinomial likelihood $L_w$(n) in world w. The posterior follows Bayes' rule:

$$\pi_{\mathbf{n}} = \frac{\pi_0 L_H(\mathbf{n})}{\pi_0 L_H(\mathbf{n}) + (1-\pi_0) L_L(\mathbf{n})}. \qquad \textbf{(4)}$$

The learner does not receive Z, q's realized world assignment, or an adjudicated truth label. The policy sees the paired results, budgets and declared model catalog. In the synthetic data exports, case-level truth is stored separately and is not accepted by the posterior function. The inaccessible known-world reference is reported only as a bound on the restricted benchmark.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

Figure 4 enumerates the ten count states for the nominal two-pair design.




**Picture 4** All ten unordered count states at m = 2, N = 100.

**Objective and Exact Solution**

Let B denote expected unique detections if every slot were used for separate audits without training. With $R_N$ and $R_E$ diagnostic slots and E production expert slots:

$$B = (R_N + 2N)a + (R_E + E)b. \quad \textbf{(5)}$$

Two separate reviews detect an expected a + b irregularities because they concern different cases. A paired review detects a + b − c unique irregularities. Its immediate opportunity cost is therefore c, even though it improves information. Write $G(\pi) = \max\{0, g_A(\pi), g_B(\pi)\}$ (Delgado & George, 2026). Integrating over the prior and the multinomial count vectors gives the finite objective (Kraszkiewicz, 2024):

$$V(m) = B - m\mathbb{E}[c] + \mathbb{E}_{\mathbf{n}}[G(\pi_{\mathbf{n}})]. \quad \textbf{(6)}$$

$$m^* \in \operatorname{argmax}_{m=0,\ldots,\min(R_N,R_E)} V(m). \quad (7)$$

The solver enumerates every four-cell count vector for each feasible m, updates the posterior and evaluates the best terminal training action. There are binomial(m + 3,3) count vectors at a fixed m and binomial(R + 4,4) across m = 0,…,R when both diagnostic budgets equal R. With R = 12, the complete design search contains 1820 sufficient-statistic vectors. Log-likelihood calculations avoid underflow; probability mass is checked after enumeration.

Optimality is exact only within this policy class: m is selected before collecting paired outcomes, verification occurs in domain A, and only one terminal training intervention is allowed. The algorithm does not solve adaptive stopping, switching verification domains after each result, concurrent scheduling, or a general partially observable Markov decision process.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

Those extensions require additional state and action representations (Kaelbling, Littman, & Cassandra, 1998).

**Experimental Protocol and Reproducibility**

The nominal parameters are listed in Table 3. We evaluate eight future workloads, from 10 to 400 cases per domain. A separate 100-configuration grid varies workload, training sensitivity increment and instructor time. Zero learning is included as a mechanism test, not omitted as an unfavorable case. Structural stress tests generate positively correlated reviewer detections while an independence-based policy remains unchanged (Kumar, Revathi, Alsharif, Gaur, & Nanthaamornphong, 2026).

**Table 3** Nominal assumptions and experiment dimensions.

| Quantity | Setting | Interpretation |
|---|---|---|
| Diagnostic capacity | 12 novice + 12 expert slots | Identical for all feasible policies. |
| Production capacity | N novice slots in each domain + 8 expert slots | N = 10, 20, 30, 40, 50, 100, 200, 400. |
| Training cost | 2 learner + 4 instructor slots | Both are removed from case-review capacity. |
| Training response | Sensitivity increment $\eta = 0.10$ | Declared response; not fitted to personnel data. |
| Prior | Pr(A is H) = 0.50 | Two known candidate worlds; realized world hidden. |
| Observation law | No false positives; nominal $\rho = 0$ | Dependence stress: $\rho = 0, 0.1, 0.2, 0.3$. |
| Parameter grid | 100 configurations | Five workloads × five gains × four instructor costs. |
| Monte Carlo | 20,000 common episodes; four policies | 80,000 paired policy evaluations, not independent observations. |

**Source:** Compiled by the authors based on the findings of the study, 2026.

The exact calculations are checked against 20,000 independent synthetic episodes at N = 100. Four policies are evaluated on common exogenous random arrays within each episode: no training, training without verification, a fixed four-pair rule and the exact precommitted verification design. This yields 80,000 policy evaluations, not 80,000 independent organizations. The same arrays generate case irregularities and review uniforms; policy-specific branching does not redraw the case environment (Nhaca, 2023).

Reported Monte Carlo standard errors concern finite simulation error. They do not quantify uncertainty in the assumed prevalence, learning response or observation law. Twenty-six automated tests check probability conservation, identification identities, feasible time budgets, exact special cases, reproducibility and the separation of policy inputs from evaluator truth. The source, notebook and run-level exports permit independent inspection. No results from earlier manuscript versions are included in this new benchmark.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

## RESULTS AND DISCUSSION

### An Information Limit Before Training

**Proposition 1 (pre-training observational equivalence).** In the two-world construction, any adaptive sequence of separate pre-training novice and expert reviews has the same observation-history distribution in both worlds. Consequently, under the symmetric prior, such histories cannot identify the high-prevalence training domain better than chance.

**Proof.** Conditional on every past history and the next selected domain, a novice finding is Bernoulli(a) and an expert finding is Bernoulli(b), in either world. The case draws are independent across separate reviews. Induction over the history length shows that every history has the same probability in the two worlds. Bayes' rule therefore leaves the prior unchanged. Randomization based on that history is independent of the realized world. Its probability of selecting the high-prevalence domain is at most one half when a training domain must be chosen. □

When training is worthwhile in both the uninformed and known-world decisions, the unavoidable difference in expected production yield is (Elazab, 2026):

$$\mathcal{R}_0 = (N-k)\eta\left(q_H - q_L\right)/2. \qquad (8)$$

This is a benchmark-specific decision consequence of non-identification, not a new general theorem about imperfect detection. The qualification "pre-training" is essential. If training changes sensitivity by a known amount, post-training findings can themselves reveal prevalence. A multi-stage policy may deliberately train to learn. Proposition 1 does not exclude that alternative (Sun X. , 2023); the implemented two-stage chronology excludes later retraining decisions (Dias, Cook, Paglia, & Massara, 2024).

### What Two Reviews Can Identify

**Proposition 2 (paired-review identity under restricted assumptions).** If specificity is one, reviewers are conditionally independent, sensitivities are homogeneous within the analyzed stratum and the joint positive probability is nonzero, the population probabilities identify prevalence and both sensitivities:

$$q = ab/c, \qquad s_1 = c/b, \qquad s_2 = c/a. \qquad (9)$$

**Proof.** The marginal probabilities are a = q $s_1$ and b = q $s_2$; their overlap is c = q $s_1$ $s_2$. Dividing ab by c gives q, after which division of each marginal by q gives the corresponding sensitivity. □

This is the familiar two-list identification logic (International Working Group for Disease Monitoring and Forecasting, 1995), used here to define an observation design rather than presented as an original estimator. Population identification also does not guarantee accurate finite-sample estimation. When the observed overlap count is zero, the plug-in ratio is undefined. The finite benchmark uses complete multinomial likelihoods over its stated candidate worlds instead of dividing sparse empirical frequencies (Bonrath & Eulerich, 2024).

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

### A Paid-Verification Threshold

Let G(π) be the maximum terminal training gain. Relative to no verification, the design value separates into an information benefit and a service loss:

$$V(m) - V(0) = \{\mathbb{E}[G(\pi_n)] - G(\pi_0)\} - m\mathbb{E}[c]. \quad \textbf{(10)}$$

The first term is nonnegative under the correctly specified Bayesian model because the original action remains available (Blackwell, 1953). The second term is strictly positive when duplicate positive findings are possible. Verification should therefore be purchased for its effect on the downstream decision, not simply because it increases statistical information (Cochrane, Wills, & Johnson, 2025).

**Proposition 3 (nominal single-pair threshold).** For the nominal values in Table 3 and N ≥ 29, training is chosen under every posterior reached by one pair. The incremental value of one blinded pair is:

$$V(1) - V(0) = 0.003(N - 2) - 0.1125. \quad \textbf{(11)}$$

**Proof.** The two cell distributions are (0.575,0.125,0.225,0.075) and (0.650,0.050,0.150,0.150). Their total-variation distance is 0.150. Under equal prior probabilities, one pair raises the probability of selecting the correct training domain from 0.500 to 0.575. The training benefit is therefore (N − 2) × 0.10 × (0.8 − 0.4) × 0.075 = 0.003(N − 2). The expected duplicate loss is the mean of 0.075 and 0.150, or 0.1125. Subtracting gives the expression. Since $q_d$ at the selected action is at least 0.6, training has positive gain for N ≥ 29. □

Thus one pair becomes beneficial only when N > 39.5, or at least 40 future cases per domain for integer N. This is not an institutional staffing recommendation: the threshold depends directly on synthetic detection rates, training effects and time costs. Its purpose is to expose the mathematical condition that a practical design would have to estimate.

## Computational Results

### The Verification Budget Is Not Monotone in Its Own Value

Table 4 reports the exact best precommitted design at each workload. No verification is selected at N = 10, 20 or 30. At N = 40, one pair is selected but its expected gain is only 0.0015 detections. The threshold is consequently a break-even point, not evidence of a material operational improvement. The optimal number rises to two pairs at N = 100, six at N = 200 and the tested maximum of 12 at N = 400.

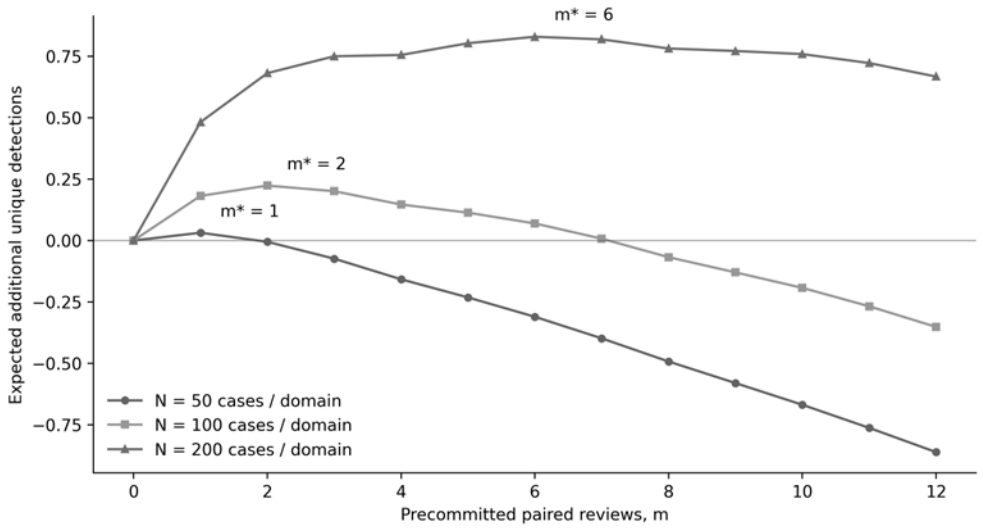
**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

Table 4 Exact verification designs and expected unique detections.

| Future cases per domain | Best pairs m* | No verification | Best design | Gain | Known-world regret |
|---|---|---|---|---|---|
| 10 | 0 | 12.40000 | 12.40000 | 0.00000 | 0.00000 |
| 20 | 0 | 16.40000 | 16.40000 | 0.00000 | 0.00000 |
| 40 | 1 | 25.08000 | 25.08150 | 0.00150 | 0.75850 |
| 50 | 1 | 29.68000 | 29.71150 | 0.03150 | 0.92850 |
| 100 | 2 | 52.68000 | 52.90335 | 0.22335 | 1.73665 |
| 200 | 6 | 98.68000 | 99.50896 | 0.82896 | 3.13104 |
| 400 | 12 | 190.68000 | 193.38423 | 2.70423 | 5.25577 |

**Source:** Compiled by the authors based on the findings of the study, 2026.

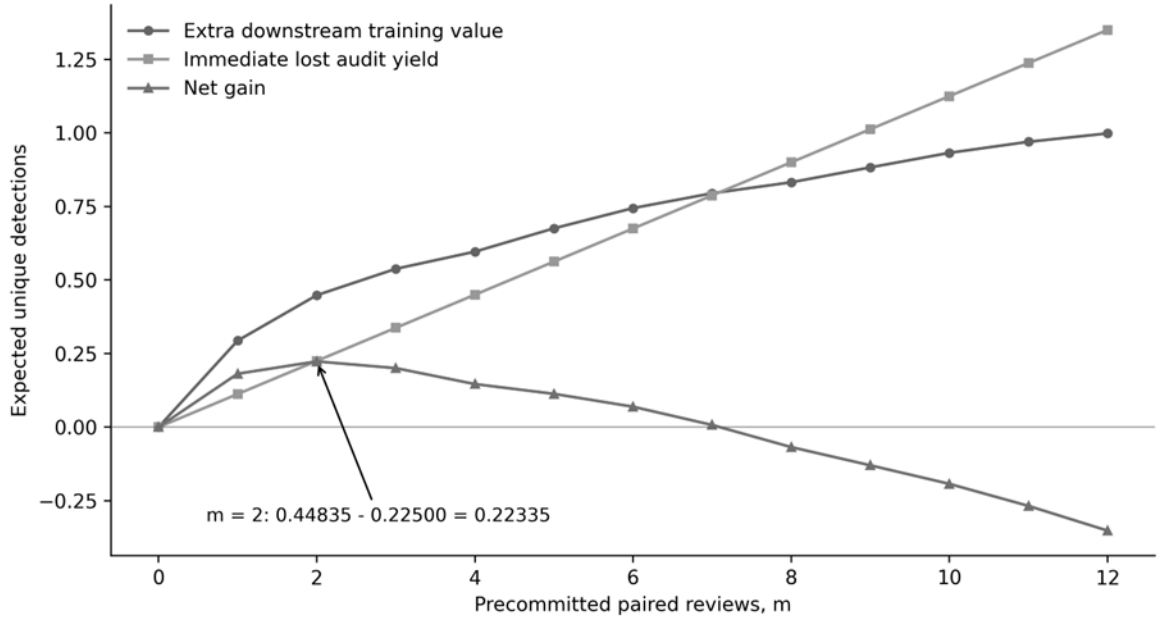
**Exact optima within the stated precommitted policy class.**




**Picture 5** Exact expected gain relative to training without verification at N = 50

The decomposition in Figure 6 separates better targeting of training from the audit yield forgone during verification. At two pairs, the increase in expected downstream training value is 0.44835 detections and the duplicate-review cost is 0.22500. Their difference, 0.22335, is the net gain already reported in Table 4. The larger information benefit at higher pair counts does not by itself establish a better decision: its service cost increases at the same time (Bagodi & Sinha, 2023).

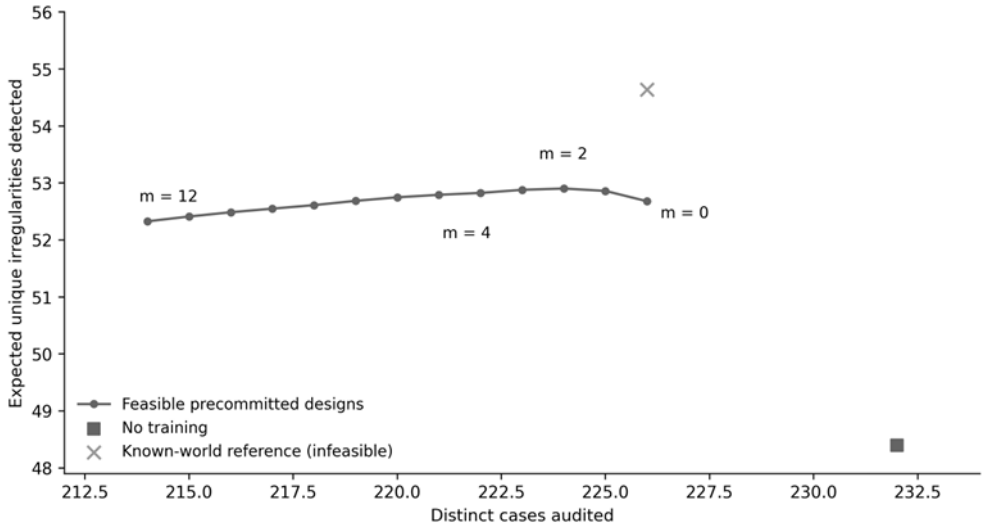
**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal




**Picture 6** At N = 100, the incremental downstream training value equals net gain.

At N = 100, two pairs reduce expected distinct-case coverage from 226 to 224. The yield gain is 0.22335, approximately 0.424% of the no-verification total. This small conditional gain should not be reframed as a large general improvement in audit effectiveness. The known-world reference yields 54.64, leaving a substantial information gap even after the paid verification design is optimized (Li J. , 2026).

**Figure 7** locates every nominal verification budget in detection–coverage space.




Picture 7 Exact design points at N = 100.

**Learning and Verification Have Different Roles**

Across the 100-configuration parameter grid, zero sensitivity gain always selects zero verification. There is no downstream skill decision that can benefit from improved targeting, so duplicate reviews only reduce unique-case yield. This result follows from the specified

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

opportunity-cost objective and is independently checked in the tests. It should not be described as evidence that real verification has no quality-assurance value; assurance (Chang, 2023), deterrence and dispute resolution are outside this objective (Li, Zhu, Zhao, & Zhao, 2024).

Training also remains a separate decision from verification. At sufficiently short workloads it is better to preserve the instructor's and learner's capacity. At longer workloads, training without verification can already improve yield. Verification buys a better choice of domain, not the entire gain from professional development. Comparisons against a no-training policy alone would therefore overstate its contribution (Hlazunov, 2024).

**Correlated Misses Can Reverse the Decision**

The independence assumption is stress-tested rather than treated as a consequence of blinding. With conditional detection correlation ρ, the joint positive probability becomes:

$$c_\rho = q\left[s_1 s_2 + \rho\sqrt{s_1(1-s_1)s_2(1-s_2)}\right]. \quad \textbf{(12)}$$

Positive dependence raises overlap and reduces the unique yield of a paired review at unchanged marginal finding rates. It also changes the likelihood that should be used for inference. Table 5 holds the nominal independence-based design fixed and evaluates its actual performance under the altered data-generating law. The comparison labeled “dependence-aware” is supplied the true correlation; it is a diagnostic reference, not an estimated deployment solution.

**Table 5** Dependence stress: net gain over no verification.

| Future cases | True ρ | Nominal pairs | Misspecified gain | Aware pairs | Aware gain |
|---|---|---|---|---|---|
| 40 | 0.0 | 1 | +0.00150 | 1 | +0.00150 |
| 40 | 0.1 | 1 | -0.02354 | 0 | +0.00000 |
| 40 | 0.2 | 1 | -0.04859 | 0 | +0.00000 |
| 40 | 0.3 | 1 | -0.07363 | 0 | +0.00000 |
| 100 | 0.0 | 2 | +0.22335 | 2 | +0.22335 |
| 100 | 0.1 | 2 | +0.15610 | 2 | +0.15610 |
| 100 | 0.2 | 2 | +0.08724 | 1 | +0.09248 |
| 100 | 0.3 | 2 | +0.01676 | 1 | +0.04798 |

**Source:** Compiled by the authors based on the findings of the study, 2026.

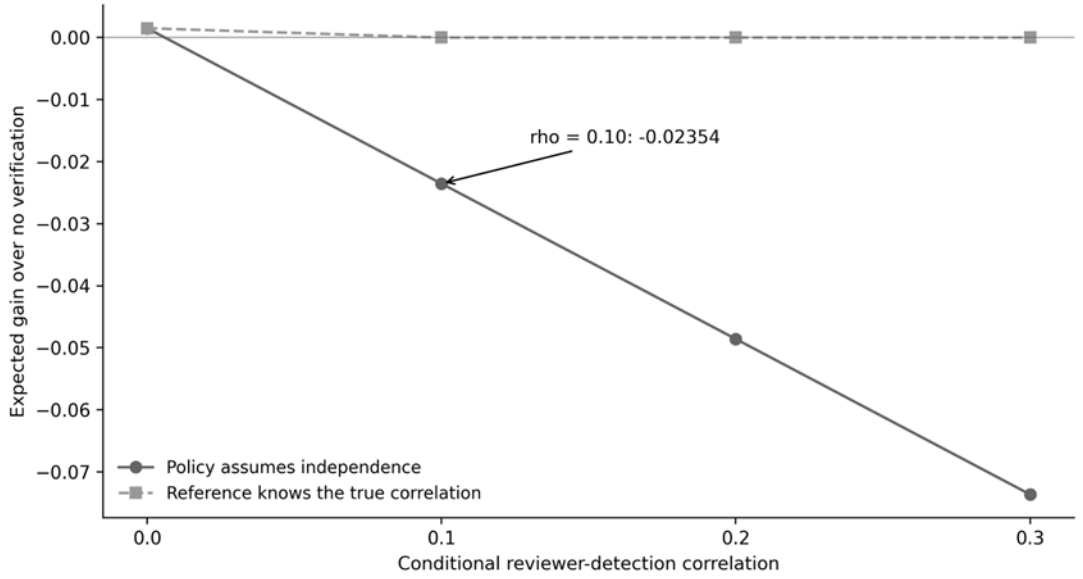
***The dependence-aware reference is supplied the true conditional correlation. This is a diagnostic comparison, not a claim that correlation is known in practice.***

At N = 40, correlation 0.10 turns the nominal gain of 0.00150 into a loss of 0.02354 detections. A design using the correct dependence selects no verification. At N = 100, the nominal

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

two-pair design remains beneficial over the tested correlations, but its gain declines. The appropriate conclusion is conditional: a longer use horizon can absorb some information-cost misspecification, while a near-break-even design is fragile (N & Ahmed, 2025).

Figure 9 retains the near-break-even N = 40 case, where misspecification reverses the decision.




**Source:** Compiled by the authors based on the findings of the study, 2026.

***The nominal design continues to assume conditional independence and retains one pair.***

Using the independence moment formula under positive conditional covariance would also underestimate prevalence. Substituting the altered overlap into ab/c gives q $s_1$ $s_2$ divided by $s_1$ $s_2$ plus the conditional covariance, which is smaller than q for positive covariance. Additional review counts do not remove that asymptotic misspecification (Collins & Huynh, 2014). Independence (Sun & Zhou, 2025), heterogeneity and false positives are therefore substantive threats to identification (International Working Group for Disease Monitoring and Forecasting, 1995).

***Independent Computational Checks***

**The paired simulated difference between the budgeted and no-verification policies is 0.25225 with Monte Carlo standard error 0.02804; its exact expectation is 0.22335. These are compatible, not identical, quantities.**

**Table 6** Exact expectations and independent computational checks.

| **Policy** | **Exact yield** | **Monte Carlo mean** | **Monte Carlo SE** | **Error / SE** |
|---|---|---|---|---|
| No training | 48.40000 | 48.34985 | 0.04359 | -1.151 |
| Training, no verification | 52.68000 | 52.62160 | 0.04709 | -1.240 |
| Four fixed pairs | 52.82634 | 52.76055 | 0.04678 | -1.406 |
| Budgeted verification | 52.90335 | 52.87385 | 0.04689 | -0.629 |

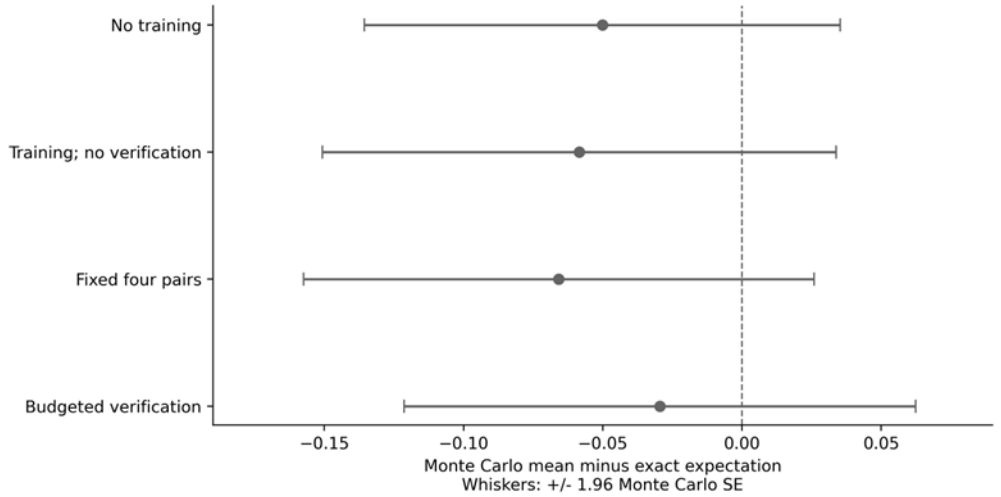
**Source:** Compiled by the authors based on the findings of the study, 2026.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

**20,000 independent episodes, with common case-level arrays across policies. Errors concern simulation precision, not uncertainty about the synthetic model** (Bourg & Gouguet, 2023)**.**

All 26 automated checks pass. The checks include the one-pair threshold identity, conservation of total paid time, the no-learning corner case and rejection of infeasible correlation parameters. They establish consistency of this implementation with the stated finite model. They do not establish that its response coefficients, two-world prior or performance measures describe an actual audit organization.

**Figure 10** plots Monte Carlo errors against their exact targets instead of visually overlapping the absolute yield estimates.




Picture 10 For each policy at N = 100

**Source:** Compiled by the authors based on the findings of the study, 2026.

## Discussion

### The Scientific Contribution and Its Boundary

The central addition to a resource-controlled assignment study is an observation problem that changes the value of skill investment. It is possible to allocate time perfectly against the wrong interpretation of findings. The aliased benchmark makes that failure explicit: arbitrarily many separate pre-training observations cannot reveal which of two domains has the higher return to the same sensitivity improvement. The organization must change its measurement design, its intervention design, or the assumptions it is willing to make (Li & Fang, 2025).

This is distinct from claiming that low detection is always evidence of poor competence. The benchmark deliberately constructs one ambiguity. Real findings may also depend on auditee cooperation, procedure choice, documentation quality and incentives. Without additional data, those mechanisms can be indistinguishable as well. The model should motivate measurement and decision analysis (Katiku, 2023), not personnel judgments based on finding counts (Danielson & Ekenberg, 2023).

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

The proposed novelty lies in the joint decision formulation and its inspectable consequence: the same constrained expertise purchases immediate detections, information about where capability is inadequate, or a future increase in capability. The underlying concepts have established precedents (Wang, Ferreira, & Yan, A, 2024). The present work does not demonstrate literature-wide priority for the combination (International Working Group, 1995). The strongest defensible claim is a specifically formulated verification–training problem, an exact restricted solver, and an adversarial benchmark that can distinguish methods which treat findings as truth from methods which account for incomplete detection (Jena, Chand, & Singh, 2026).

***A General Dynamic Extension: Specified, Not Yet Evaluated***

A repeated workforce model would carry a belief over latent risk and response parameters, current competency, remaining time and queued engagements. Audit, verification and mentoring would affect different parts of that state (Omolere, 2026). For state $s_t = (b_t, C_t, H_t, Q_t)$, a compact formulation is:

$$V_t(s) = \max_{u \in \mathcal{U}(s)} \mathbb{E}_{b_t}[r(u) + \gamma V_{t+1}(s')]. \quad \textbf{(13)}$$

Here $b_t$ is a posterior belief, $C_t$ competency, $H_t$ capacity and $Q_t$ queued work. The reward counts each detected irregularity once and can penalize overdue tasks. Feasible actions charge novice, expert, verification and instruction hours against the same individual budgets. A transition updates both skills and beliefs. This is a partially observed control formulation, not a newly invented solution method (Kaelbling, Littman, & Cassandra, 1998).

The implementation supplied here solves the finite two-stage special case, not this general extension. Before claiming a scalable algorithmic contribution, the extension should be tested against an active-training probe, a policy that selects verification by information gain alone, a certainty-equivalent allocator, a no-verification allocator and a known-parameter diagnostic bound. Training-as-probe is especially important because a known capability intervention can itself break pre-training observational equivalence. Omitting that comparator could exaggerate the necessity of blinded duplicate reviews.

Resource matching must also be preserved. Any approximate dynamic policy should receive the same task arrivals, individual time budgets and admissible measurements as its comparators. Error dependence, false positives, parameter misspecification and changing task mix should be varied independently rather than hidden inside a favorable policy bundle. Out-of-sample evaluation should freeze design choices before scoring a held-out environment.

***Data Requirements and Limitations***

The experiment uses constructed cases and explicitly declared response probabilities. It does not use personnel records, real audit findings or a calibrated learning curve. The numerical example should be read as a discriminating mathematical test, not a secondary-data validation of

defense governance. An external table of aggregate audit findings cannot identify the reviewer sensitivities, paired error dependence or training response required by this model (Labeeb, 2023).

A suitable empirical dataset would link initial risk assessments, case identifiers, independently recorded paired review results, reviewer roles, task characteristics, elapsed review time and subsequent adjudication where available. Training records should distinguish instructor hours, learner hours, targeted capability and post-training task exposure. Assignment and verification probabilities must be recorded before outcomes. Otherwise, disagreement samples selected after a first finding can introduce a new selection problem.

The paired-data assumptions require direct examination. Reviewers should not see each other's labels before submitting their own; the evidence version should be preserved; and heterogeneous case difficulty should be modeled or stratified. Blinding is not proof of conditional independence. Unknown false-positive rates require additional identifying restrictions (Sun & Zhou, 2025), reference information or a suitable additional measurement design (Hui & Walter, 1980). No claim that two reviews alone establish prevalence is warranted in that general setting (Collins & Huynh, 2014).

The benchmark omits continuous skill growth, forgetting, specialist exit, variable duration,

For the Indonesian Air Force Inspectorate, the immediate implication is therefore a governed measurement proposal, not automated deployment. Findings should not be used as a direct competence score; developmental records should not become unvalidated disciplinary evidence. Any institutional study would require author and institutional determination of permissions, privacy protections and applicable review obligations before data collection.

## CONCLUSION

This study demonstrates that audit findings cannot always be interpreted as direct representations of underlying organizational risk or auditor capability. When the probability of detecting an irregularity depends simultaneously on its prevalence and the reviewer’s sensitivity, identical observed finding rates may correspond to substantially different latent conditions. Within the constructed two-environment benchmark, separate pre-training reviews remain unable to distinguish these conditions, even when additional observations are collected. The central implication is therefore not that auditors lack information, but that the information generated by conventional observation may be structurally insufficient for identifying where skill investment will produce the greatest return. This finding reinforces the importance of distinguishing between the occurrence of irregularities and the capacity to detect them when audit evidence is subsequently used for workforce development decisions.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

The analysis further shows that verification can reduce this identification problem, but verification itself competes with assurance production and training for scarce expert and novice capacity. Under the benchmark assumptions, paired reviews provide additional information because their joint outcomes contain information that separate reviews do not. However, the informational advantage does not automatically translate into superior audit performance. At 100 future cases per domain, the optimal two-pair design increases expected unique detections from 52.680 to 52.903, while reducing distinct-case coverage from 226 to 224. Thus, the gain is positive but comparatively modest, demonstrating that information acquisition must be evaluated against its opportunity cost rather than treated as costless quality improvement.

The study also identifies an important non-monotonic relationship between verification and decision performance. Verification becomes worthwhile only beyond a workload-dependent threshold, while additional paired observations generate increasing costs and may provide diminishing or fragile benefits. More importantly, the dependence analysis shows that a decision that appears marginally favourable under conditional independence can become unfavorable when reviewers share correlated errors. At a workload of 40 cases per domain, a modest conditional dependence of 0.10 changes the nominal positive gain into a negative outcome. This result emphasizes that model assumptions concerning reviewer independence are not merely technical specifications; they can directly influence managerial decisions about training expenditure.

Overall, the study contributes a bounded computational decision framework linking imperfect detection, verification expenditure, skill development, and opportunity cost. Its contribution should be understood as a proof-of-concept rather than an empirically validated personnel policy. The evidence is generated from synthetic environments, predetermined parameters, exact enumeration, and simulation-based computational checks. Accordingly, the findings establish the internal logic and decision consequences of the proposed benchmark but do not establish that the assumed prevalence, learning response, or reviewer behavior corresponds to a particular institution. Future empirical research is therefore required before the framework can support operational workforce decisions.

## RECOMMENDATIONS

*First*, organizations should avoid treating raw audit-finding frequency as a direct measure of auditor competence. A low number of recorded findings may reflect low underlying irregularity prevalence, limited detection sensitivity, differences in case difficulty, procedural constraints, or other contextual factors. Consequently, training priorities should not be determined solely from aggregate finding counts. A more defensible approach is to combine audit outcomes with information about case characteristics, reviewer role, task exposure, review time, and, where feasible, independently obtained verification evidence. Such an approach would allow managers to distinguish a genuine competency gap from a measurement problem before allocating scarce resources to professional development.

**International Journal of Interdisciplinary Cultural Studies**
ISSN: 2327-008X (Print), ISSN: 2327-2554 (Online)
Volume 21, Issue 2, 2026
https://cgscopus.com/index.php/journal

*Second*, verification should be treated as a targeted information investment rather than an automatic enhancement of audit quality. The model indicates that duplicate reviews are justified only when the expected improvement in the subsequent decision exceeds the detection opportunities sacrificed to obtain additional information. Institutions should therefore establish explicit decision thresholds that incorporate workload, expected training benefit, reviewer time, and the opportunity cost of reduced case coverage. The numerical thresholds reported in this study should not be transferred directly to practice because they arise from synthetic assumptions. Instead, they should serve as a methodological template for estimating institution-specific break-even points using empirical parameters.

*Third*, future empirical implementation should prioritize the construction of high-quality paired-review data. Reviewers should independently record their assessments before exposure to another reviewer's judgment, while case identifiers, task characteristics, reviewer characteristics, and subsequent adjudication should be retained. The study specifically indicates that blinding alone cannot guarantee conditional independence because shared training, common evidence, and differences in case difficulty may generate correlated errors. Therefore, independence assumptions should be tested rather than presumed. Where false-positive rates or reviewer heterogeneity are material, additional identification strategies or reference information will be required.

*Fourth*, subsequent research should extend the present two-stage framework into a repeated and adaptive allocation model. Such a model could jointly represent changing beliefs about latent risk, evolving reviewer competency, remaining expert capacity, queued assignments, and the effects of training. Verification, auditing, and mentoring could then be evaluated as alternative uses of the same constrained resources. Comparative testing should include no-verification allocation, information-gain-based verification, certainty-equivalent allocation, active training as a diagnostic intervention, and a known-parameter benchmark. This would clarify whether the proposed mechanism remains advantageous when decisions can be revised over time.

*Five*, institutional adoption should proceed through a governed empirical validation process rather than immediate automated deployment. For settings such as public-sector or defense auditing, findings should not become unvalidated individual competence scores or disciplinary indicators. A suitable implementation should first establish data governance, privacy safeguards, institutional authorization, transparent parameter estimation, and independent evaluation. The accompanying computational package, exact solver, synthetic outputs, and automated tests provide a reproducible methodological foundation, but institutional deployment requires new evidence and formal authorization. Future studies should therefore move from mathematical identification, to calibrated institutional data, and ultimately to out-of-sample evaluation before the framework is used to guide consequential workforce or training decisions.

**ACKNOWLEDGEMENTS**

The author would like to express sincere gratitude to the Rector of Universitas Muhammadiyah Jakarta, Prof. Dr. Ma'mun Murod; the Dean of the Faculty of Social and Political Sciences, Universitas Muhammadiyah Jakarta, Prof. Dr. Evi Satispi, S.Pd., M.Si.; and the Head of the Doctoral Program in Public Administration, Prof. Dr. Andriansyah, M.Si., for their valuable support, guidance, and encouragement throughout the research process. The author greatly appreciates the opportunities, direction, and continuous guidance provided from the initial stage to the completion of this research.

**REFERENCES**

Ahmad, M., Alam, M., & Ahmed, J. (2026). A Hybrid Multi-Detector Framework with Bayesian Uncertainty Quantification and Machine Learning Verification for Robust QRS Detection and HRV Analysis. *12(3)*, 1013–1021. doi:10.29284/bs8x4n45

Bagodi, V., & Sinha, D. (2023). Analysis of dynamics and opportunity cost of SME operations. *38(2)*, 166–166. doi:10.1504/ijpqm.2023.129605

Battaglini, M., Guiso, L., Lacava, C., Miller, D., & Patacchini, E. (2025). Refining public policies with machine learning: The case of tax auditing. J. *Econom*, *249*, 105847. doi:10.1016/j.jeconom.2024.105847

Blackwell, D. (1953). Equivalent Comparisons of Experiments. *Ann. Math. Stat*, *24*, 265–272. doi:10.1214/aoms/1177729032

Bonrath, A., & Eulerich, M. (2024). Internal auditing's role in preventing and detecting fraud: An empirical analysis. doi:10.1111/ijau.12342

Bourg, J.-F., & Gouguet, J.-J. (2023). Opportunity cost: what policies should be implemented around mega-sporting projects? 152–155. doi:10.25965/ebooks.507

Bychkov, D. (2025). Bridging the Training and Compliance Gap: Evidence from FBI Audit Data. doi:10.2139/ssrn.5417194

Chang, K.-P. (2023). Opportunity Cost and Investment Criteria. 47–57. doi:10.1007/978-981-19-9119-6_5

Cochrane, J., Wills, A., & Johnson, S. (2025). RJHMC-Tree for Exploration of the Bayesian Decision Tree Posterior. doi:10.1214/25-ba1565

Collins, J., & Huynh, M. (2014). Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Stat. Med, 33*, 4141–4169. doi:10.1002/sim.6218

Danielson, M., & Ekenberg, L. (2023). Bayesian Decision Analysis. 21–34. doi:10.1201/9781003406709-3

Delgado, A., & George, D. (2026). Full Reparation and Opportunity Cost of Renewable Investors. *11*, 83–100. doi:10.54648/eila2026007

Dias, A., Cook, C., Paglia, A., & Massara, R. (2024). Imperfect Detection of Terrestrial Mammals in Environmental Impact Assessment (Eia) Baseline Surveys May Compromise Decision-Making and Mitigation Measures. doi:10.2139/ssrn.5030501

Elazab, R. (2026). Disclosure of material weaknesses in internal control systems: what is the impact of investment decisions. *5(3)*. doi:10.1504/ijaudit.2026.10080410

Feinstein, J. (1990). Detection Controlled Estimation. J. *Law Econ*, *33*, 233–276. doi:10.1086/467205

Feinstein, J. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. *RAND J. Econ*, *22*, 14–35. doi:10.2307/2601005

Grünig, R., & Kühn, R. (2025). Problem Verification and Analysis. 55–64. doi:10.1007/978-3-662-71620-5_7

Hlazunov, A. (2024). DEVELOPMENT OF BAYESIAN NETWORKS FOR A DECISION SUPPORT SYSTEM DURING INTERNAL CYBER THREATS ANALYSIS. *1(25)*, 103–117. doi:10.28925/2663-4023.2024.25.103117

Hui, S., & Walter, S. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, *36*, 167–171. doi:10.2307/2530508

International Working Group for Disease Monitoring and Forecasting. (1995). Capture–Recapture and Multiple-Record Systems Estimation I: History and Theoretical Development. *Am. J. Epidemiol*, *142*, 1047–1058. doi:10.1093/oxfordjournals.aje.a117558

Jena, S., Chand, L., & Singh, A. (2026). AUTOMATED VERIFICATION OF BRSR DISCLOSURES: A NEURO-SYMBOLIC NLP PIPELINE FOR OFFLINE GREENWASHING DETECTION, ETHICAL AUDITING, AND GREEN FINANCE INTEGRATION IN INDIA. *14(4)*, 134–145. doi:10.21474/ijar01/23320

Kaelbling, L., Littman, M., & Cassandra, A. (1998). Planning and acting in partially observable stochastic domains. *Artif. Intell*, *101*, 99–134. doi:10.1016/S0004-3702(98)00023-X

Katiku, S. (2023). Quality Internal Auditing and Management Decision Making in Kenya. *3(2)*, 96–100. doi:10.59413/ajocs/v3.i2.2

Kraszkiewicz, A. (2024). Opportunity Cost (Cost of Lost Profits) as an Element of Decision-Making Analysis in Prosumer Energy. 229–239. doi:10.1007/978-3-031-70955-5_25

Kumar, A., Revathi, V., Alsharif, M., Gaur, N., & Nanthaamornphong, A. (2026). Learning-Based Bayesian Deep Unfolding for Robust MIMO Detection Under Imperfect CSI. *33*, 3651–3655. doi:10.1109/lsp.2026.3719183

Kumar, S., & Bawankar, S. (2025). Dual Pathway CNN-Bayesian Decision Framework for Robust Brain Tumor Analysis. 914–920. doi:10.1109/dasa68193.2025.11499073

Labeeb, G. (2023). Factors Affecting the Relationship between Internal Audit and External Audit in Banks and their Impact on Strengthening the Internal Control System and Reducing the Cost of External Auditing. *12(6)*, 2674–2686. doi:10.21275/sr23623142232

Li, J. (2026). Algorithmic Auditing for Employee Health: Bridging the Verification Gap in GRI 403 Compliance via Topological Data Analysis. doi:10.2139/ssrn.5992894

Li, W., & Fang, C. (2025). Bayesian decision framework for optimizing imperfect preventive maintenance in repairable systems with competing failure modes. *3144(1)*, 12035–12035. doi:10.1088/1742-6596/3144/1/012035

Li, Z., Zhu, Y., Zhao, L., & Zhao, Y. (2024). Safety decision of autonomous driving behavior based on Bayesian network and its verification method. 145–153. doi:10.1109/qrs-c63300.2024.00029

Lin, D., & Zhan, Z. (2025). Automated Financial Statement Auditing via YOLOv5s Object Detection and NLP-Based Semantic Analysis. *49(11)*. doi:10.31449/inf.v49i11.8999

Liu, Y. (2023). Necessity of Multiple Rounds of Nucleic Acid Detection: Application of Bayesian Decision Theory in Real Life. *49*, 476–482. doi:10.54691/bcpbm.v49i.5452

Mousazadegan et al. (2026). A novel approach to improve workforce skills in the multi-skilled project scheduling problem using the apprenticeship method. *Comput. Ind. Eng*, *211*, 111651. doi:10.1016/j.cie.2025.111651

N, P., & Ahmed, L. (2025). Robotic Cloud Automation-Enabled Attack Detection and Command Verification Using Attention-Based RNNs, ConvLSTM, and Bayesian Networks. *10(3)*, 1–19. doi:10.46243/jst.2025.v10.i03.pp01-19

Nhaca, Á. (2023). Analysis of the Contribution of Internal Auditing in Organizations. doi:10.11648/j.frontiers.20230304.11

Omolere, O. (2026). AI and Machine Learning in Auditing: Enhancing Cybersecurity Risk Management through Real-Time Monitoring, Anomaly Detection, and Predictive Analysis. doi:10.2139/ssrn.5968654

Pfister, K., & Hamann, H. (2023). Collective Decision-Making and Change Detection with Bayesian Robots in Dynamic Environments. 8814–8819. doi:10.1109/iros55552.2023.10341649

Sun, A., & Zhou, X.-H. (2025). Estimation of Diagnostic Test Accuracy Without Gold Standards. *Stat. Med*, *44*, e10315. doi:10.1002/sim.10315

Sun, X. (2023). Data decision based promoting effect of non-destructive detection skill competition on detection technology course teaching in colleges and universities. *9(1)*. doi:10.2478/amns.2023.2.00760

Tai, Q., & Zhou, X.-H. (2025). Simultaneous adjustment of nonignorable verification bias and imperfect gold standard bias in diagnostic test studies, with an application to Alzheimer's disease detection. *19(4)*. doi:10.1214/25-aoas2081

V, R. (2026). Digital Transformation in Accounting and Auditing: An Analysis of Its Impact on Financial Reporting, Internal Controls and Fraud Detection. *22(1)*, 341–355. doi:10.69889/3qtqf402

Wang, X., Ferreira, F., & Yan, P. (2024). A multi-objective competency-based decision support system for the assignment of internal auditors to multiple projects. *Ann. Oper. Res, 338*, 303–334. doi:10.1007/s10479-024-05855-3

Wang, X., Ferreira, F., & Yan, P. (2025). A multi-objective optimization approach for integrated risk-based internal audit planning. *Ann. Oper. Res, 346*, 1811–1840. doi:10.1007/s10479-023-05228-2

Wu, X. (2025). A Unified Global and Local Outlier Detection Framework with Application to Chinese Financial Budget Auditing. *13(11)*, 978–978. doi:10.3390/systems13110978

Yang, T. (2023). Analysis of Application Status of Bayesian Decision. *38*, 888–893. doi:10.54097/hset.v38i.5974

Youssef, S. (2025). OCR and Object Detection Synergy: A Novel Framework for Automated Price Tag Verification and Shelf Auditing. doi:10.2139/ssrn.5555720