TARGET-SPECIFIC CAUSAL RECOVERY UNDER FEDERATED SUPPORT FRAGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A federation can have enough information overall to resolve a causal target, such as an edge direction, while no client can form some required conditional-independence (CI) answers. We study a fixed set of answers sufficient to determine that target. Some losses leave it recoverable; others admit a graph with a different target answer. When unavailable queries abstain and available answers are correct, we identify the inclusion-minimal groups whose joint loss permits such an alternative. Recovery fails exactly when one such group is entirely missing. Queries with shared conditioning requirements rely on the same client counts. These counts determine when destructive losses occur and distinguish insufficient pooled support from failure caused solely by data placement. For K clients with independently sampled data, we derive an exact failure probability accounting for dependent query losses. Individual missing rates generally do not determine this probability. Holding the same global counts fixed isolates risk caused only by repartitioning observations across clients; averaging these conditional risks recovers the overall failure probability. Controlled enumeration shows different risks with each query's missing probability held fixed. Paired ALARM and CHILD experiments isolate pure fragmentation. Added evidence improves ALARM recovery through alternatives that clients can support, and the exact law closely predicts the transition. In CHILD, these alternatives share a support bottleneck and yield no observed recovery gain.

1 INTRODUCTION

A federation can have enough information overall to resolve a causal target while some required evidence remains unavailable locally. Consider two hospitals evaluating whether a treatment and an outcome are conditionally independent given a binary patient characteristic. Suppose answering this query locally requires at least m observations at each level. The hospitals may hold counts $(m, m-1)$ and $(m-1, m)$, so neither can answer locally even though the pooled counts $(2m-1, 2m-1)$ meet the same requirement. Whether that loss matters depends on what the remaining answers establish. For a simple example, consider two candidate directed acyclic graphs (DAGs) that disagree on whether a target edge is $T \to Y$ or $Y \to T$. If each of two CI answers rules out the reversed-edge candidate, either answer alone determines the orientation, but losing both makes it ambiguous. The risk depends on how often both answers are missing, which their individual missing rates do not determine. More generally, the same pooled support can yield different usable local evidence, and the same amount of missing evidence can yield different target consequences. *Which joint losses of otherwise sufficient causal evidence destroy a target guarantee, and when can federation create those losses purely through support placement?*

We study this question for a fixed target and a supplied candidate DAG class. A fixed finite set of CI queries whose complete, correct answers determine the target is a *target-sufficient CI certificate*. Federated causal methods study how CI evidence, local structures, scores, or sufficient statistics are formed and combined across clients (Wang et al., 2023; Guo et al., 2024a; 2025). This leaves open which losses from an already sufficient evidence system destroy the target guarantee, how federated support produces those losses, and what target risk follows. Under the *local whole-query interface*, at least one client must support every conditioning stratum required by a query. The *matched aggregate interface* applies the same readiness requirement after client counts are combined. Comparing them separates losses caused by dividing the data from those that remain even after pooling.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

When available answers are correct, a graph with a different target answer remains possible exactly when every certificate query that excludes it becomes unavailable. The inclusion-minimal such loss sets are the *minimal destructive sets*; recovery fails exactly when one is entirely missing. Shared conditioning requirements translate these target-level destructive losses into exact conditions on client support: queries that share a conditioning set depend on the same weakest-stratum counts, so several query-level losses can collapse to the same support bottleneck rather than behave as separate missing events. This distinguishes *pure fragmentation*, in which local target recovery fails despite aggregate recovery, from insufficient pooled support. Even when missing answers after pooling do not prevent recovery, they can reduce how much further evidence loss the target can tolerate. With data sampled independently across clients, we derive the exact failure probability for K clients while allowing query losses to remain dependent. Holding the same realised global table fixed isolates risk caused only by its allocation across clients. Averaging these conditional risks over global tables recovers the overall failure probability for those K clients.

Experiments test these mechanisms in fixed causal tasks. Controlled enumeration holds each query's unavailability probability fixed while changing how joint losses align with the destructive sets, producing sharply different target risks. On ALARM, paired trials keep each realised global count table unchanged and vary only its allocation across clients; local recovery can change while aggregate recovery remains unchanged. The exact risk calculation closely tracks this transition. Across fixed evidence families, added evidence improves ALARM robustness by creating effective alternative support routes. In CHILD/GRUNTINGREPORT, those routes share a rare support bottleneck and yield no observed recovery gain. Together, these experiments test the joint-loss, support-fragmentation, and probabilistic parts of the theory. We make three contributions:

- **Minimal joint losses of causal evidence determine target failure.** For a fixed target-sufficient certificate, target-discrepant candidates induce inclusion-minimal destructive sets. When unsupported queries abstain and available answers retain their reference labels, failure occurs exactly when every query in one such set is unavailable. This characterisation yields a target-specific resilience to evidence loss and explains why matched query-wise unavailability can produce different risks.
- **Federated support placement can cause target failure while aggregate recovery remains possible.** Shared conditioning requirements connect destructive losses to client support, yielding an exact local failure representation. Comparing local and aggregate interfaces separates aggregate scarcity from pure fragmentation. The surviving certificate quantifies how nonfatal aggregate scarcity reduces resilience to further fragmentation.
- **Target risk generally depends on how support is jointly placed across clients.** Under independent client support-state vectors, we derive the exact fixed-K failure probability while retaining arbitrary dependence across support blocks within each client. Query-wise unavailability probabilities do not generally determine this risk. Conditioning on realised global support yields fixed-global placement risks whose expectation recovers the unconditional law.

2 RELATED WORK

Restricted and target-local CI evidence. Constraint-based causal discovery recovers graphical information from CI relations and their sample analogues (Spirtes et al., 2000; Kalisch & Bühlmann, 2007). Recent theory asks what remains identifiable when the available CI interface is deliberately restricted: Shiragur et al. (2024) characterise causal information recoverable with fewer CI tests, while Lee et al. (2025) study discovery under a prescribed collection of conditioning sets. Target-local methods likewise seek target-relevant graphical information without learning the full graph (Loranchet & Assaad, 2026). These works vary or restrict the evidence family and ask what that family can identify. We instead take a supplied target-sufficient certificate as fixed and characterise its target-specific fragility under partial evidence availability. The resulting minimal destructive sets identify exactly which losses from that supplied interface re-admit a target-discrepant candidate.

Federated causal learning and evidence formation. Federated causal-discovery methods study how causal evidence is estimated, communicated, or aggregated across clients. FedC2SL forms constraint-based decisions through a federated CI test without collecting raw client data, whereas FedCSL combines locally learned CI-based structural information at the server (Wang et al., 2023; Guo

et al., 2024a). More recent work addresses variable-space sample-quality heterogeneity, limited local sample sizes, and target-local federated structure learning (Guo et al., 2024b; 2025; Yu et al., 2025). We take the target, candidate class, and sufficient certificate as fixed and study what happens when each whole CI answer must first become support-ready under the local whole-query interface. This fragmentation mechanism arises at interfaces that form CI evidence locally; methods that combine suitable sufficient statistics before the CI decision, such as FedC2SL, can aggregate complementary support before the certificate evidence is formed (Wang et al., 2023).

CI redundancy, dependence, and reliability representations. Redundant CI information can detect or correct errors beyond the statements used by a discovery procedure; Faller & Janzing (2026) develop several notions of such redundancy in CI-based graphical-model discovery. Our fixed-certificate analysis instead characterises how redundancy reshapes the joint evidence losses that re-admit target-discrepant candidates. Dependence among CI properties and test outcomes is also an established phenomenon (Mazaheri et al., 2026); our pathwise failure characterisation accommodates arbitrary erasure dependence, while the fixed-K risk law preserves within-client dependence through the joint support-state distribution. Multi-state coherent-system theory provides classical representations of monotone failure through ordered component states and cut/path vectors, together with standard reliability calculations (Barlow & Wu, 1978; Ohi, 2019). Here these tools operate on a failure system induced jointly by target-discrepant causal candidates and federated support constraints, yielding the target-specific recovery-risk law studied in this paper.

3 Problem Formulation

3.1 Target Task and Fixed Certificate

Fix a finite discrete variable set $\mathcal{V}$, a full-system reference DAG $G^\star_{\text{full}}$ on $\mathcal{V}$, a target $T \in \mathcal{V}$, and a target-relevant scope $\mathcal{V}_T \subseteq \mathcal{V}$ containing T. Let $\mathcal{G}$ be a nonempty supplied class of candidate DAGs on $\mathcal{V}_T$. Each candidate induces a target summary through $\mathcal{L}_T : \mathcal{G} \longrightarrow \prod_{h\in H} \mathcal{R}_h$, where H indexes the target components and each finite set $\mathcal{R}_h$ contains their admissible graph-level values. Let $\mathcal{L}^\star_T \in \prod_{h\in H} \mathcal{R}_h$ denote the intended full-system target summary. We also write $G^\star_T := G^\star_{\text{full}}[\mathcal{V}_T]$ for the graph induced by the chosen target scope; Assumption 1 states when this reduced representation is adequate for the target task.

A conditional-independence (CI) query is a triple $I = (X_a, X_b, S)$ asking whether $X_a \perp\!\!\!\perp X_b \mid S$, with $X_a, X_b \in \mathcal{V}_T$ and $S \subseteq \mathcal{V}_T \setminus \{X_a, X_b\}$. Queries need not contain T directly: relations among other variables in $\mathcal{V}_T$ may be needed to exclude target-discrepant candidates. For $G \in \mathcal{G}$, let $d_G(I) := \mathbf{1}[X_a \perp\!\!\!\perp_G X_b \mid S]$, and define the target-reference label from the full-system DAG by $d^\star_T(I) := \mathbf{1}[X_a \perp\!\!\!\perp_{G^\star_{\text{full}}} X_b \mid S]$. When S is nonempty, Ω_S denotes its finite state space and $J_I := |\Omega_S|$ its number of conditioning strata. For $S = \varnothing$, set $\Omega_\varnothing := \{\varnothing\}$ and $J_I = 1$.

Definition 1 (Target-Sufficient CI Certificate)**.** A finite query set $\mathcal{C}$ is a *target-sufficient CI certificate* relative to $(\mathcal{G}, \mathcal{L}_T, d^\star_T, \mathcal{L}^\star_T)$ if the all-reference compatible set

$$\mathsf{VS}^\star_T(\mathcal{C}) := \{G \in \mathcal{G} : d_G|_{\mathcal{C}} = d^\star_T|_{\mathcal{C}}\} \tag{1}$$

is nonempty and every graph in that set has the intended target summary: $G \in \mathsf{VS}^\star_T(\mathcal{C}) \implies \mathcal{L}_T(G) = \mathcal{L}^\star_T$.

The certificate $\mathcal{C}$ is supplied and fixed before evidence formation, may be nonminimal, and may contain redundant queries. Our object is its robustness under partial availability. To interpret that robustness relative to the intended full-system target, we require the chosen target scope to preserve both the certificate labels and the target summary.

Assumption 1 (Certificate-Local Target-Scope Adequacy)**.** The induced target-scope graph $G^\star_T$ belongs to $\mathcal{G}$, matches the full-system reference on the fixed certificate, $d_{G^\star_T}(I) = d^\star_T(I)$ for every $I \in \mathcal{C}$, and represents the intended target summary, $\mathcal{L}_T(G^\star_T) = \mathcal{L}^\star_T$.

This certificate-local assumption ensures that the chosen target scope correctly represents both the fixed certificate and the intended target summary, so the subsequent analysis studies evidence availability within an adequate target representation. We place this fixed target task in a horizontal federation of clients $[K] = \{1, \ldots, K\}$ over the same variable set $\mathcal{V}$, where client k contributes a finite local data set $\mathcal{D}_k$ of size n_k and is associated with a population P_k. The first two main results

are pathwise in realised evidence and realised support states, while a stochastic model is introduced only for the recovery-risk calculation in Section 4.3.

3.2 Evidence Interface and Analysis Regimes

Definition 2 (Three-State Evidence Interface)**.** A *three-state evidence interface* on the fixed certificate is a map $\widetilde{d} : \mathcal{C} \longrightarrow \{0, 1, \perp\}$, where 0 and 1 are definite CI labels and $\perp$ denotes erasure. Its visible and erased query sets are $\mathcal{C}^{\mathrm{vis}} := \{I \in \mathcal{C} : \widetilde{d}(I) \in \{0, 1\}\}$ and $\mathcal{U}^{\perp} := \{I \in \mathcal{C} : \widetilde{d}(I) = \perp\}$. A definite output with $\widetilde{d}(I) \neq d_T^{\star}(I)$ is *target-discordant*; an erasure imposes no compatibility constraint.

Assumption 2 (Clean-Erasure Regime)**.** For every certificate query, $\widetilde{d}(I) \in \{d_T^{\star}(I), \perp\}$. Thus every surfaced definite label equals the target-reference label, while unsupported evidence may abstain.

This is the regime used by Theorems 1–3. It isolates support loss from incorrect definite evidence: unsupported queries remove compatibility constraints, while every surfaced definite label remains aligned with the target reference.

Definition 3 (Compatible Candidate Set)**.** Given realised three-state evidence, the *compatible candidate set* is

$$\widehat{\mathsf{VS}}_T = \left\{ G \in \mathcal{G} : d_G(I) = \widetilde{d}(I) \text{ for every } I \in \mathcal{C}^{\mathrm{vis}} \right\}. \tag{2}$$

For a nonempty compatible set, the associated *safe target report* $\widehat{\mathcal{L}}_T$ reports a target component only when all compatible candidates assign it the same value; otherwise it returns ABSTAIN on that component. If the compatible set is empty, the report is INC. Under Assumptions 1 and 2, $G_T^{\star}$ remains compatible, so erasure enlarges the compatible set by removing constraints and manifests as target ambiguity when compatible candidates disagree.

Proposition 1 (Maximal Componentwise Soundness)**.** *Suppose $\widehat{\mathsf{VS}}_T \neq \varnothing$. The safe target report is definite exactly on components whose values are constant over $\widehat{\mathsf{VS}}_T$. If two compatible candidates disagree on component h, no componentwise rule that is sound for every graph in $\widehat{\mathsf{VS}}_T$ can report a definite value on h.*

4 Target-Specific Failure under Federated Support Fragmentation

With the reporting rule fixed by Proposition 1, the theory characterises which losses of the certificate enlarge the compatible set enough to destroy the target guarantee. Target failure separates into three layers: the candidate DAG class determines which certificate losses can admit a target-discrepant explanation, the local federated interface determines which of those losses occur for a realised support allocation, and a client sampling law determines how often the resulting support states arise. The first two layers are pathwise; probability enters only in the third.

4.1 Candidate-Induced Failure Geometry

A safe reporting rule tells us how to act once the compatible set is known, but not which missing answers allow a target-discrepant candidate to enter it. Call $G \in \mathcal{G}$ *target-discrepant* when $\mathcal{L}_T(G) \neq \mathcal{L}_T^{\star}$, and write $\mathcal{A}_T := \{G \in \mathcal{G} : \mathcal{L}_T(G) \neq \mathcal{L}_T^{\star}\}$. For such a candidate, its *certificate discrepancy set* $\mathrm{Disc}_T(G) := \{I \in \mathcal{C} : d_G(I) \neq d_T^{\star}(I)\}$ records exactly the reference answers that exclude G when they are visible.

Definition 4 (Minimal Destructive Set)**.** A set $D \subseteq \mathcal{C}$ is a *minimal destructive set for target T* if $D = \mathrm{Disc}_T(G)$ for some $G \in \mathcal{A}_T$ and no target-discrepant candidate has a discrepancy set that is a proper subset of D. The family of all such sets is

$$\mathsf{MinDes}_T := \mathrm{Min}_{\subseteq}\{\mathrm{Disc}_T(G) : G \in \mathcal{A}_T\}. \tag{3}$$

When $\mathcal{A}_T \neq \varnothing$, define the *erasure resilience* $\rho_T := \min_{D \in \mathsf{MinDes}_T} |D|$; when $\mathcal{A}_T = \varnothing$, set $\mathsf{MinDes}_T = \varnothing$ and $\rho_T = +\infty$. Certificate sufficiency implies that every destructive set is nonempty whenever target-discrepant candidates exist. Let $\mathsf{Fail}_T := \{\widehat{\mathcal{L}}_T \neq \mathcal{L}_T^{\star}\}$ denote failure to recover the complete target summary.

Theorem 1 (Exact Target-Recovery Failure under Erasure)**.** *Under Assumption 2, if* $\mathsf{MinDes}_T = \varnothing$*, then* $\mathsf{Fail}_T = \varnothing$*. Otherwise, for every realised erased set* $\mathcal{U}^\perp$*,*

$$\mathsf{Fail}_T = \bigcup_{D \in \mathsf{MinDes}_T} \{D \subseteq \mathcal{U}^\perp\}. \tag{4}$$

Remark 1 (Equal Erasure Counts Can Have Opposite Outcomes)**.** Suppose, for illustration, that $\mathcal{C} = \{I_1, I_2, I_3\}$ and the only minimal destructive set is $\{I_1, I_2\}$. Erasing $\{I_1, I_2\}$ causes target failure, whereas erasing $\{I_1, I_3\}$ does not, although both masks erase two queries. The theorem therefore depends on *which* answers disappear jointly, not on the erasure count alone.

Proof sketch. Under clean erasure, every all-reference candidate remains compatible. A target-discrepant candidate G is compatible exactly when $\mathrm{Disc}_T(G) \subseteq \mathcal{U}^\perp$. Such a candidate exists exactly when $\mathcal{U}^\perp$ contains a member of MinDes_T. The complete proof is in Appendix A.2.

The result is pathwise and imposes no stochastic model on query erasures. The scalar ρ_T records only the smallest destructive-set size; exact risk also depends on the full antichain MinDes_T and its overlap structure. Inclusion-minimal sufficient certificates are necessarily singleton-fragile under erasure, a consequence formalised in Appendix A.3.

4.2 Federated Support Generates Destructive Erasures

Theorem 1 identifies the destructive erasure masks without referring to federation. We now ask when a local support interface produces exactly those masks. Each CI remains an atomic evidence item: conditioning strata determine whether the whole query can surface, rather than becoming separately composable evidence pieces. For client population P_k, write $d_{P_k}(I) := \mathbf{1}[X_a \perp\!\!\!\perp_{P_k} X_b \mid S]$. Linking local support to the clean-erasure evidence model requires the certificate labels carried by client populations to share the target-reference semantics.

Assumption 3 (Shared Certificate Semantics)**.** Every client population agrees with the target reference on the fixed certificate: $d_{P_k}(I) = d^\star_T(I)$ for every $k \in [K]$ and $I \in \mathcal{C}$. No agreement is required for client marginals, supports, parameters, or off-certificate CI relations.

This assumption aligns the meaning of locally surfaced certificate evidence with the target reference while allowing client populations to differ in marginals, supports, parameters, and off-certificate CI relations. It enters at the support-to-evidence step, where local readiness is converted into surfaced reference labels or erasures.

Definition 5 (Local Whole-Query Readiness)**.** For query $I = (X_a, X_b, S)$, let $m_I \in \mathbb{N}_{\geq 1}$ be its prespecified per-stratum readiness threshold, and let $N_k(S = s)$ be the number of records in $\mathcal{D}_k$ with conditioning state $S = s$. Client k is *locally ready for the whole query* I when every conditioning stratum reaches that threshold. Equivalently,

$$W_{k,S} := \min_{s \in \Omega_S} N_k(S = s), \qquad M_{k,I} := \mathbf{1}\{W_{k,S} \geq m_I\}, \tag{5}$$

with $W_{k,\varnothing} := n_k$. The local whole-query interface can form I when at least one client is ready: $M^{\mathrm{loc}}_I := \max_k M_{k,I}$.

To distinguish local fragmentation from genuine aggregate scarcity, we use the matched aggregate interface, which applies the same threshold after combining stratum counts across clients:

$$M^{\mathrm{agg}}_I := \mathbf{1}\left\{ \min_{s \in \Omega_S} \sum_{k=1}^{K} N_k(S = s) \geq m_I \right\}. \tag{6}$$

Queries sharing the same conditioning set rise and fall with the same weakest-stratum count $W_{k,S}$. The exact count contains more information than the certificate needs: for availability, it suffices to record which of the certificate's distinct thresholds have been crossed. We call the distinct conditioning sets $\mathfrak{S}_\mathcal{C} := \{S_I : I \in \mathcal{C}\}$ the certificate's *support blocks*.

Definition 6 (Client Support State)**.** For block $S \in \mathfrak{S}_\mathcal{C}$, let $0 < t_{S,1} < \cdots < t_{S,d_S}$ be the distinct readiness thresholds among certificate queries using S. The *support state* of client k on block S is the highest threshold level reached by its weakest conditioning stratum,

$$Z_{k,S} := \max\big(\{j : W_{k,S} \geq t_{S,j}\} \cup \{0\}\big). \tag{7}$$

The *client support-state vector* is $\mathbf{Z}_k := (Z_{k,S})_{S \in \mathfrak{S}_\mathcal{C}}$.

For the matched aggregate interface, define

$$W_S^{\mathrm{agg}} := \min_{s \in \Omega_S} \sum_{k=1}^{K} N_k(S = s). \tag{8}$$

$$Z_S^{\mathrm{agg}} := \max\big(\{j : W_S^{\mathrm{agg}} \geq t_{S,j}\} \cup \{0\}\big). \tag{9}$$

and let $\mathbf{Z}^{\mathrm{agg}} := (Z_S^{\mathrm{agg}})_{S \in \mathfrak{S}_c}$. Let $j(I)$ satisfy $m_I = t_{S_I, j(I)}$. The local federation exposes the coordinatewise maximum of the client states, $\mathbf{Z}^{\mathrm{loc}} := \bigvee_{k=1}^{K} \mathbf{Z}_k$, so, by construction,

$$\begin{aligned} M_{k,I} = 1 &\iff Z_{k,S_I} \geq j(I), \\ M_I^{\mathrm{loc}} = 1 &\iff Z_{S_I}^{\mathrm{loc}} \geq j(I), \\ M_I^{\mathrm{agg}} = 1 &\iff Z_{S_I}^{\mathrm{agg}} \geq j(I). \end{aligned} \tag{10}$$

Moreover, $W_S^{\mathrm{agg}} \geq W_{k,S}$ for every client k and support block S, so $\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{Z}^{\mathrm{agg}}$ coordinatewise and pathwise. Under Assumption 3, the readiness-gated population-label interface returns $d_T^\star(I)$ when I is locally ready and $\perp$ otherwise, hence it satisfies the clean-erasure regime. The readiness threshold specifies evidence availability. Finite-sample validity of a particular CI test is a separate statistical property.

A destructive set D is erased only when every query in D stays below its readiness threshold. In support-state coordinates, there is therefore a coordinatewise largest state that still keeps all queries in D unavailable. We call this state the *support ceiling* of D, denoted $\mathbf{b}(D) = (b_S(D))_{S \in \mathfrak{S}_c}$, with

$$b_S(D) = \begin{cases} \min\limits_{I \in D:\, S_I = S} j(I) - 1, & \text{if } D \text{ uses block } S, \\ d_S, & \text{otherwise.} \end{cases} \tag{11}$$

Different destructive sets can induce the same lower support region, and one such region can be contained in another. The *coordinatewise-undominated support-blocker family* retains only the coordinatewise-maximal ceilings after duplicate removal:

$$\mathcal{B}_T := \mathrm{Max}_{\leq}\{\mathbf{b}(D) : D \in \mathsf{MinDes}_T\}, \tag{12}$$

where $\leq$ is coordinatewise order. Removing smaller ceilings removes only principal downsets already contained in another failure region. Thus $\mathcal{B}_T$ is irredundant in the formal product-state poset; a particular attainable-state support or sampling law can make further blockers operationally redundant or impossible.

Theorem 2 (Exact Support-Block Representation of Target Failure)**.** *Under Assumption 3 and the readiness-gated population-label interface,*

$$\mathsf{Fail}_T = \bigcup_{\mathbf{b} \in \mathcal{B}_T} \{\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{b}\}. \tag{13}$$

The representation is pathwise in the realised support allocation.

Proof sketch. Every query in a destructive set D is locally erased exactly when, for each support block used by D, the federation level lies below the smallest threshold required by a query from that block. This is precisely $\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{b}(D)$. Theorem 1 then gives the union over destructive sets, and duplicate or contained principal downsets can be removed without changing the event. The complete proof is in Appendix B.2. Thus Theorem 2 provides the deterministic bridge from candidate-induced destructive sets to realised federated support regions.

Proposition 2 (Aggregate Scarcity and Pure Fragmentation)**.** *Let* $\mathsf{Fail}_T^{\mathrm{loc}}$ *and* $\mathsf{Fail}_T^{\mathrm{agg}}$ *denote complete target-recovery failure under the local whole-query and matched aggregate interfaces, respectively. Then, pathwise,* $\mathsf{Fail}_T^{\mathrm{agg}} \subseteq \mathsf{Fail}_T^{\mathrm{loc}}$. *Define the* pure-fragmentation event $\mathsf{Frag}_T := \mathsf{Fail}_T^{\mathrm{loc}} \cap (\mathsf{Fail}_T^{\mathrm{agg}})^c$. *Then*

$$\mathsf{Fail}_T^{\mathrm{loc}} = \mathsf{Fail}_T^{\mathrm{agg}} \,\dot\cup\, \mathsf{Frag}_T, \tag{14}$$

$$R_T^{\mathrm{loc}} = R_T^{\mathrm{agg}} + R_T^{\mathrm{frag}}, \tag{15}$$

where the three R-terms are the probabilities of the corresponding events.

Proof. Since $\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{Z}^{\mathrm{agg}}$ pathwise and the failure region in Theorem 2 is a down-set, aggregate failure implies local failure. The disjoint event and risk decompositions then follow directly from the definition of Frag_T. □

Remark 2 (Fragmentation versus Aggregate Scarcity)**.** For a binary conditioning variable and threshold m_I, clients with stratum counts $(m_I, m_I - 1)$ and $(m_I - 1, m_I)$ are both locally unready even though the aggregate counts $(2m_I - 1, 2m_I - 1)$ are ready. Proposition 2 lifts this query-level observation to the complete target event. Proposition 3 formalizes a non-singleton causal witness in which block marginals and each query's fragmentation probability are held fixed while target risk changes; Appendix B.5 gives the construction and proof.

Corollary 1 (Remaining Fragmentation Resilience)**.** *Let $E \subseteq \mathcal{C}$ be the set of queries already unavailable under the aggregate interface, and suppose the surviving certificate $\mathcal{C}_E := \mathcal{C} \setminus E$ still recovers the complete target summary. Then its remaining erasure resilience is*

$$\rho_T(\mathcal{C}_E) = \min_{D \in \mathsf{MinDes}_T(\mathcal{C})} |D \setminus E|. \tag{16}$$

Thus nonfatal aggregate scarcity can consume part of the remaining fragmentation margin. Appendix B.5 gives the full destructive-family restriction and proof.

4.3 Exact Client-Generated Recovery Risk

The support representation is conditional on a realised allocation. To obtain a recovery probability, we now place a stochastic law on the client support states. This changes the quantifier, not the structural failure geometry.

Theorem 3 (Exact Client-Generated Target-Recovery Risk)**.** *Assume the conditions of Theorem 2 and suppose that the client support-state vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$ are mutually independent. The client vectors may have different distributions, and support blocks within each client may be arbitrarily dependent. For $F_k(\mathbf{z}) := \Pr(\mathbf{Z}_k \leq \mathbf{z})$, the federation-level lower CDF satisfies*

$$\Pr(\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{z}) = \prod_{k=1}^{K} F_k(\mathbf{z}). \tag{17}$$

Consequently,

$$\Pr(\mathsf{Fail}_T) = \sum_{\varnothing \neq \mathcal{J} \subseteq \mathcal{B}_T} (-1)^{|\mathcal{J}|+1} \prod_{k=1}^{K} F_k\left(\bigwedge_{\mathbf{b} \in \mathcal{J}} \mathbf{b}\right), \tag{18}$$

where $\bigwedge$ is coordinatewise minimum. If $\mathcal{B}_T = \varnothing$, the risk is zero by convention.

Proof sketch. The event $\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{z}$ occurs exactly when $\mathbf{Z}_k \leq \mathbf{z}$ for every client, giving Equation (17) under mutual client independence. Intersections of support-failure regions are principal downsets with ceiling equal to their coordinatewise meet, so inclusion–exclusion gives Equation (18). The complete proof is in Appendix B.3.

The independence assumption is across client support-state vectors; each joint CDF F_k retains arbitrary dependence across support blocks within a client, while threshold nesting within each block is encoded by the corresponding support-state coordinate. That joint information cannot in general be reduced to query-wise or blockwise marginals.

Proposition 3 (Joint Support Is Necessary for Target Risk)**.** *There exists a five-variable causal target and a two-query target-sufficient certificate with $\mathsf{MinDes}_T = \{\{I_1, I_2\}\}$ and $\rho_T = 2$, together with two fixed global support profiles having the same client sizes, the same aggregate readiness, the same global histogram for each conditioning block, and the same conditional local-unavailability probability for each query, but pure-fragmentation risks $1/3$ and 0.*

This finite witness already shows that query-wise support summaries are not sufficient to determine target risk. The separation is not confined to the particular $1/3$-versus-0 construction: the marginal support information can remain matched while the target-level effect becomes arbitrarily severe.

Corollary 2 (Sharp Joint-Support Separation)**.** *There exists a sequence of CI problems with paired fixed-global support profiles having identical query-wise global support margins and conditional local-unavailability probabilities, but target-risk gaps tending to one. Within each pair, the target task, candidate class, and certificate are fixed.*

Proposition 3 and Corollary 2 therefore separate marginal support agreement from joint target robustness: the former can remain unchanged while the latter ranges from near agreement to an almost maximal risk separation. Appendix B.5 gives the five-variable witness, the asymptotic construction, and the proofs. The next result conditions on a realised global table and relates these fixed-global placement risks to the unconditional client-generated risk.

Proposition 4 (Conditional Fixed-Global Disintegration)**.** *Fix K and the client-size vector $(n_1, \dots, n_K)$. Let C denote the realised support-relevant global joint histogram under a fixed-size random repartition, and define the* fixed-global placement risk $\psi_T(c) := \Pr(\mathsf{Fail}_T^{\mathrm{loc}} \mid C = c)$*. If the aggregate interface succeeds at c, then $\psi_T(c) = \Pr(\mathsf{Frag}_T \mid C = c)$. Unconditionally,*

$$\Pr(\mathsf{Fail}_T^{\mathrm{loc}}) = \mathbb{E}_C[\psi_T(C)]. \tag{19}$$

Thus the unconditional fixed-K target risk is the average, over realised global support, of the corresponding fixed-global placement risk. The conditional inclusion–exclusion representation and proof are in Appendix B.5.

4.4 Finite-Sample and Capacity Consequences

The exact risk law uses the joint client support-state distribution. When that full joint law is unavailable, finite stratum counts still provide coarser readiness and risk consequences; Appendix B.4 develops those reduced-information results. For the deterministic capacity consequence below, write $b_I := J_I m_I$.

Corollary 3 (Fixed-Budget Local Capacity Boundary)**.** *If some $D \in \mathsf{MinDes}_T$ satisfies $\max_{k\in[K]} n_k < b_I$ for every $I \in D$, then target recovery fails deterministically. Under remainder-balanced allocation of N observations, define*

$$K_{I,0}(N) := \begin{cases} \lceil N/(b_I - 1) \rceil, & b_I \geq 2, \\ \infty, & b_I = 1. \end{cases} \tag{20}$$

$$K_{T,0}(N) := \min_{D \in \mathsf{MinDes}_T} \max_{I \in D} K_{I,0}(N). \tag{21}$$

with $K_{T,0}(N) := \infty$ when $\mathsf{MinDes}_T = \varnothing$. Every feasible $K \geq K_{T,0}(N)$ then has $\Pr(\mathsf{Fail}_T) = 1$.

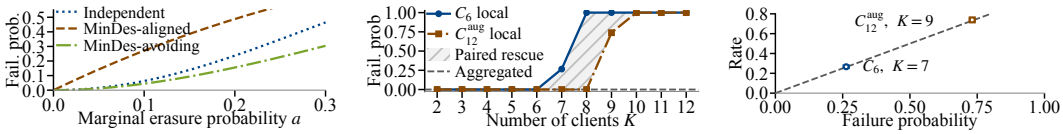
Corollary 3 shows that, even at fixed total sample size, sufficiently fine client fragmentation can force failure by capacity alone. Appendix B.4 gives the finite-sample readiness bounds, the complete proof of this capacity consequence, and computable upper and lower risk envelopes when only marginal readiness summaries are available; Appendix C.1 characterises the additional failure channel created by incorrect definite evidence. Together, Theorems 1–3 and these consequences close the chain from finite local support to query readiness, from readiness to destructive federated support regions, and from those regions to exact or bounded target-recovery risk.

5 Evaluation

We evaluate destructive-loss geometry, support fragmentation, and fixed-K risk prediction. Controlled experiments specify erasures directly; paired experiments derive availability from realised client counts. Under the evaluation interface, available queries carry fixed reference CI labels. Figure 1 reports the results; details appear in Appendices D.2, D.4, D.5, and D.9.

Joint destructive geometry controls target risk. The controlled five-variable instance uses target summary $A \to T$, $B \to T$, $T \to C$, together with nonadjacency of D and T. Exact enumeration yields 18 non-singleton minimal destructive sets with $\rho_T = 2$. Holding all ten query-erasure marginals at 0.1 while changing only their dependence alignment produces more than a six-fold risk spread (Figure 1(a)), as predicted by the joint-loss criterion in Theorem 1. All 2,800 query-group partitions are enumerated in Appendix D.2. For Alarm, the inclusion-minimal six-query certificate C_6 and its inactive one-query extension C_7 share destructive geometry and independent-erasure risk. The prespecified twelve-query extension C_{12}^{aug} and full $|S| \leq 1$ family Q_{40} improve robustness (Appendix D.3).

Repartitioning identical global information creates local target failure. On the five-node Alarm benchmark target scope with $T =$ LVEDVolume and target relations Hypovolemia $\to T$,



(a) **(b)** **(c)**

Figure 1: Three evaluations aligned with the three main theoretical results. (a) Exact controlled enumeration gives 18 non-singleton minimal destructive sets with $\rho_T = 2$. With identical query-wise erasure marginals $a = 0.1$, destructive-set-avoiding, independent, and destructive-set-aligned mechanisms have exact target-failure risks 0.0442, 0.0628, and 0.2710, showing that matched marginal missingness can yield sharply different target risks. (b) Paired ALARM trials repartition the same realised global table with $N = 160{,}000$. At $K = 8$, C_6 fails in all 5,000 trials, whereas C_{12}^{aug} and Q_{40} recover in all of them; the aggregated comparator also recovers in every trial. For C_{12}^{aug}, 17 query-level minimal destructive sets contract to 12 support rectangles and six coordinatewise-undominated blockers, isolating pure fragmentation from aggregate scarcity. (c) Exact fixed-K risks match observation: 0.2631 versus 0.2662 for C_6, $K = 7$, and 0.7315 versus 0.7388 for C_{12}^{aug}, $K = 9$; the latter theory value lies in the Wilson 95% interval $[0.7264, 0.7508]$. Full controlled, paired, and transition details are in Appendices D.2, D.4, and D.5.

LVFAILURE $\to T$, $T \to$ CVP, and $T \to$ PCWP, each paired trial draws one global joint-count table at $N = 160{,}000$ and reuses that same realised table across client counts K, changing only the record partition. Figure 1(b) shows the resulting partition-only failure transition while the aggregated comparator continues to recover. For C_{12}^{aug}, the causal geometry contracts from 17 query-level minimal destructive sets to 12 support rectangles and six undominated blockers, with the $K = 9$ transition dominated by the LVEDVOLUME block at threshold $m = 1645$. This realises the support-block mechanism of Theorem 2 in data: shared conditioning support compresses causal loss geometry into lower-dimensional failure regions, and every observed local failure in the paired experiment is a pure-fragmentation event by Proposition 2.

The exact fixed-K law predicts the observed transition. At fixed K, the paired construction has the unconditional law of independent multinomial client tables, satisfying the independence assumption of Theorem 3; Appendix B.1 proves this equivalence. Conditioning on the realised global table yields the fixed-global placement risk in Proposition 4, while averaging over realised global tables recovers the unconditional fixed-K risk. The observed agreement in Figure 1(c) confirms the quantitative prediction of Theorem 3.

Cross-target validation reveals a shared support bottleneck. On CHILD/GRUNTINGREPORT, the compact C_5, fixed augmentation C_8^{aug}, and Q_{40} have different destructive geometries but share an operationally dominant DISEASE support bottleneck. With $N = 160{,}000$, all three families recover in all 5,000 paired trials at $K = 4$ and fail in all 5,000 at $K = 5$, while the matched aggregate interface recovers throughout. Thus geometry-changing additional evidence need not improve realised robustness when its alternative routes remain tied to the same support bottleneck; Appendix D.7 gives the full analysis, while Appendix D.8 reports the separate CHILD/DISEASE complete-evidence insufficiency case and the broader 84-target applicability census.

6 CONCLUSION

Federated causal recovery can fail even when the federation as a whole has enough support, because the evidence needed to exclude a target-discrepant explanation may be split across clients. We show that target failure is determined exactly by the minimal joint losses of certificate evidence that re-admit such alternatives, and that shared conditioning requirements translate these losses into exact support conditions for local failure. A matched aggregate interface separates aggregate scarcity from pure fragmentation, while independent client sampling yields an exact fixed-K target-failure risk and conditioning on global support isolates the effect of support placement. Controlled and benchmark experiments confirm the predicted fragmentation and risk transitions, while also showing that additional evidence need not improve realised robustness when alternative routes share the same support bottleneck. Federated target robustness therefore depends jointly on the causal role of the missing evidence and on where the support required to realise that evidence resides across clients.

AI USE STATEMENT

AI was used to assist with research refinement, theoretical formulation and proof checking, experimental design, analysis and interpretation of experimental results, literature organisation, and manuscript polishing. All AI-assisted mathematical claims, proofs, experimental protocols, reported results, and manuscript content were reviewed and validated by the authors. The authors take full responsibility for the correctness, originality, and integrity of the work.

REPRODUCIBILITY STATEMENT

Complete proofs are provided in Appendices A–B.5. The supplementary experimental material documents the controlled destructive-geometry construction, matched-partition dependence analysis, fixed ALARM certificate geometry, paired realised-count fragmentation experiment, exact fixed-K risk calculation, conditional placement diagnostic, cross-target CHILD/GRUNTINGREPORT fragmentation validation, applicability checks, and the pre-specified readiness specification; implementation and integrity details are summarised in Appendix D.9. The accompanying anonymous artifact contains the code, configurations, and final outputs used to reproduce the reported numerical results.

REFERENCES

Richard E. Barlow and Alexander S. Wu. Coherent systems with multi-state components. *Mathematics of Operations Research*, 3(4):275–281, 1978.

Philipp Michael Faller and Dominik Janzing. On different notions of redundancy in conditional-independence-based discovery of graphical models. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 1027–1035. PMLR, 2026.

Xianjie Guo, Kui Yu, Lin Liu, and Jiuyong Li. FedCSL: A scalable and accurate approach to federated causal structure learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 12235–12243, 2024a.

Xianjie Guo, Kui Yu, Hao Wang, Lizhen Cui, Han Yu, and Xiaoxiao Li. Sample quality heterogeneity-aware federated causal discovery through adaptive variable space selection. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 4071–4079, 2024b.

Xianjie Guo, Liping Yi, Xiaohu Wu, Kui Yu, and Gang Wang. Enhancing causal discovery in federated settings with limited local samples. In *Federated Learning in the Age of Foundation Models*, pp. 164–179. Springer Nature Switzerland, Cham, 2025.

Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295, 1983.

Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8:613–636, 2007.

Kenneth Lee, Bruno Ribeiro, and Murat Kocaoglu. Constraint-based causal discovery from a collection of conditioning sets. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 2486–2516. PMLR, 2025.

Timothée Loranchet and Charles K. Assaad. Local Markov equivalence for PC-style local causal discovery and identification of controlled direct effects. In *Proceedings of the Fifth Conference on Causal Learning and Reasoning*, volume 323 of *Proceedings of Machine Learning Research*, pp. 84–118. PMLR, 2026.

Bijan Mazaheri, Jiaqi Zhang, and Caroline Uhler. Meta-dependence in conditional independence testing. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 4427–4440. PMLR, 2026.

Fumio Ohi. Mathematics of systems' reliability: Systems' structure and stochastic performance. *IEICE ESS Fundamentals Review*, 12(4):269–279, 2019.

Kirankumar Shiragur, Jiaqi Zhang, and Caroline Uhler. Causal discovery with fewer conditional independence tests. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 45060–45078. PMLR, 2024.

Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

Zhaoyu Wang, Pingchuan Ma, and Shuai Wang. Towards practical federated causal structure learning. In *Machine Learning and Knowledge Discovery in Databases: Research Track*, volume 14170 of *Lecture Notes in Computer Science*, pp. 351–367. Springer, 2023.

Kui Yu, Chen Rong, Hao Wang, Fuyuan Cao, and Jiye Liang. Federated local causal structure learning. *Science China Information Sciences*, 68(3):132105, 2025.

Appendix

A Structural Failure Geometry

A.1 Safe Compatible Recovery

Proof of Proposition 1. Fix a target component h. By the safe-report rule following Definition 3, the report on h is definite exactly when every graph in $\widehat{\mathsf{VS}}_T$ has the same value on h; in that case the reported value is that common value. Hence every definite report is sound for every compatible graph.

Conversely, suppose $G, G' \in \widehat{\mathsf{VS}}_T$ disagree on h. Any rule that outputs a definite value v on h must disagree with at least one of the two values $\mathcal{L}_{T,h}(G)$ and $\mathcal{L}_{T,h}(G')$. Such a rule is therefore not sound for every graph in the compatible class. Thus abstention is forced whenever the compatible class does not agree componentwise, proving maximal componentwise soundness. □

A.2 Exact Destructive-Erasure Geometry

Lemma 1 (Compatibility Criterion under Clean Erasure)**.** *Under Assumption 2, a candidate $G \in \mathcal{G}$ is compatible with the surfaced certificate answers if and only if*

$$\mathrm{Disc}_T(G) \subseteq \mathcal{U}^{\perp}. \tag{22}$$

Consequently, complete target recovery fails if and only if the erased set contains the discrepancy set of at least one target-discrepant candidate.

Proof. Under Assumption 2, a visible query carries its reference label. Thus, for every $I \in \mathcal{U}^{\mathrm{vis}}$, the surfaced value is $d_T^\star(I)$. A candidate G is compatible exactly when

$$d_G(I) = d_T^\star(I) \qquad \forall I \in \mathcal{U}^{\mathrm{vis}}. \tag{23}$$

Equivalently, every certificate query on which G disagrees with the reference is absent from the visible set, which is Eq. (22).

Assumption 1 guarantees that $G_T^\star$ matches the reference certificate labels, so $G_T^\star$ remains compatible. If a target-discrepant G also remains compatible, the compatible class contains two graphs with different target summaries, and the safe rule cannot recover the complete reference summary. Conversely, if complete target recovery fails while $G_T^\star$ remains compatible, some target component is not constant over the compatible class; hence at least one compatible graph is target-discrepant. Applying Eq. (22) gives the result. □

Proof of Theorem 1. By Lemma 1,

$$\mathsf{Fail}_T = \bigcup_{G \in \mathcal{A}_T} \{\mathrm{Disc}_T(G) \subseteq \mathcal{U}^{\perp}\}. \tag{24}$$

If $\mathcal{A}_T = \varnothing$, the union is empty and the theorem follows. Otherwise, certificate sufficiency implies that every $\mathrm{Disc}_T(G)$ for $G \in \mathcal{A}_T$ is nonempty.

Because $\mathcal{C}$ is finite, every discrepancy set in Eq. (24) contains at least one inclusion-minimal discrepancy set $D \in \mathsf{MinDes}_T$. Hence

$$\mathrm{Disc}_T(G) \subseteq \mathcal{U}^{\perp} \quad \Longrightarrow \quad D \subseteq \mathcal{U}^{\perp} \tag{25}$$

for some $D \in \mathsf{MinDes}_T$. The reverse implication is immediate because every $D \in \mathsf{MinDes}_T$ is itself the discrepancy set of at least one target-discrepant candidate. Therefore Eq. (24) reduces exactly to Eq. (4). □

Corollary 4 (Structure-Aware Erasure Bounds)**.** *Under Assumption 2 and $\mathsf{MinDes}_T \neq \varnothing$,*

$$\max_{D \in \mathsf{MinDes}_T} \Pr(D \subseteq \mathcal{U}^{\perp}) \leq \Pr(\mathsf{Fail}_T). \tag{26}$$

$$\Pr(\mathsf{Fail}_T) \le 1. \tag{27}$$

$$\Pr(\mathsf{Fail}_T) \le \sum_{D \in \mathsf{MinDes}_T} \Pr(D \subseteq \mathcal{U}^{\perp}). \tag{28}$$

No independence among query-erasure events is required.

Proof. Each event $\{D \subseteq \mathcal{U}^{\perp}\}$ is contained in Fail_T by Theorem 1, which gives the lower bound after maximising over D. The upper bound is the union bound applied to the exact failure union; truncation at one is immediate. □

A.3 Minimal Certificates and Extension-Consistent Nested Evidence

Proposition 5 (Singleton Fragility of Inclusion-Minimal Certificates)**.** *Let $\mathcal{C}$ be a nonempty target-sufficient certificate such that no proper subset of $\mathcal{C}$ is target-sufficient. Then*

$$\mathsf{MinDes}_T(\mathcal{C}) = \{\{I\} : I \in \mathcal{C}\}, \qquad \rho_T(\mathcal{C}) = 1. \tag{29}$$

Proof of Proposition 5. Fix $I \in \mathcal{C}$. Since $\mathcal{C}$ is inclusion-minimal, $\mathcal{C} \setminus \{I\}$ is not target-sufficient. The reference graph remains compatible with its labels on this reduced set, so failure of sufficiency means that there exists $G_I \in \mathcal{G}$ satisfying

$$d_{G_I}(J) = d^{\star}_T(J) \quad \forall J \in \mathcal{C} \setminus \{I\}, \tag{30}$$

while $\mathcal{L}_T(G_I) \neq \mathcal{L}^{\star}_T$. Since the full $\mathcal{C}$ is target-sufficient, G_I cannot also agree on I. Hence

$$\mathrm{Disc}_T(G_I) = \{I\}. \tag{31}$$

Thus every singleton $\{I\}$ belongs to the discrepancy family and is inclusion-minimal. Conversely, any discrepancy set containing two or more queries contains one of these singleton discrepancy sets as a strict subset and therefore cannot be inclusion-minimal. This proves

$$\mathsf{MinDes}_T(\mathcal{C}) = \{\{I\} : I \in \mathcal{C}\}, \tag{32}$$

and the minimum cardinality is one. □

Thus complete-evidence minimality can coincide with maximal singleton fragility under erasure.

The following appendix-only consequence is used to interpret nested certificate families in the empirical analysis. It is deliberately weaker than a certificate-selection or repair-ordering result.

Corollary 5 (Extension-Consistent Nested-Evidence Monotonicity)**.** *Let $\mathcal{C} \subseteq \mathcal{C}'$ be target-sufficient certificates for the same target, candidate class, and reference labels. Suppose the evidence interface is* extension-consistent*: adding queries in $\mathcal{C}' \setminus \mathcal{C}$ does not change which queries in $\mathcal{C}$ surface or the labels they carry. Under Assumption 2, for every realised evidence state,*

$$\mathsf{Fail}_T(\mathcal{C}') \subseteq \mathsf{Fail}_T(\mathcal{C}). \tag{33}$$

Under any common law for the extension-consistent evidence states,

$$R_T(\mathcal{C}') \le R_T(\mathcal{C}), \tag{34}$$

with strict inequality if and only if

$$\Pr[\mathsf{Fail}_T(\mathcal{C}) \setminus \mathsf{Fail}_T(\mathcal{C}')] > 0. \tag{35}$$

Proof. At a common realised state, extension consistency ensures that every surfaced constraint from $\mathcal{C}$ is also present under $\mathcal{C}'$, possibly together with additional target-consistent constraints. Hence

$$\widehat{\mathsf{VS}}_T(\mathcal{C}') \subseteq \widehat{\mathsf{VS}}_T(\mathcal{C}). \tag{36}$$

If the larger compatible class on the right already has the unique target summary $\mathcal{L}^{\star}_T$, then every graph in the smaller class on the left has that same summary. Therefore recovery under $\mathcal{C}$ implies recovery under $\mathcal{C}'$, which is Eq. (33). Taking probabilities gives the risk inequality. The difference between the two risks is exactly the probability of the set difference in Eq. (35). □

B Federated Support Fragmentation and Exact Risk

B.1 Support-State Representation

For $\mathbf{z} = (z_S)_{S \in \mathfrak{S}_\mathcal{C}}$, define the readiness map

$$\Phi(\mathbf{z})_I := \mathbf{1}\{z_{S_I} \geq j(I)\}, \qquad I \in \mathcal{C}. \tag{37}$$

The support-state space $\prod_{S \in \mathfrak{S}_\mathcal{C}} \{0, \ldots, d_S\}$ is ordered coordinatewise.

Proposition 6 (Support-Threshold Chain Representation)**.** *For every client k, its query-readiness mask is $\Phi(\mathbf{Z}_k)$. Moreover,*

$$\Phi(\mathbf{z} \vee \mathbf{z}') = \Phi(\mathbf{z}) \vee \Phi(\mathbf{z}'), \tag{38}$$

and therefore

$$(M_I^{\mathrm{loc}})_{I \in \mathcal{C}} = \Phi\left(\bigvee_{k=1}^{K} \mathbf{Z}_k \right). \tag{39}$$

The ambient number of threshold signatures is at most $\prod_{S \in \mathfrak{S}_\mathcal{C}} (d_S + 1)$; dependencies among support blocks may make some states unreachable.

Proof. By definition of $Z_{k,S}$,

$$Z_{k,S} \geq j \iff W_{k,S} \geq t_{S,j}. \tag{40}$$

Substituting $S = S_I$ and $j = j(I)$ gives $M_{k,I} = \Phi(\mathbf{Z}_k)_I$. For each query coordinate,

$$\Phi(\mathbf{z} \vee \mathbf{z}')_I = \mathbf{1}\{\max(z_{S_I}, z'_{S_I}) \geq j(I)\} = \Phi(\mathbf{z})_I \vee \Phi(\mathbf{z}')_I, \tag{41}$$

which proves Eq. (38); iterating over clients gives Eq. (39). The state-count bound is the cardinality of the ambient product of chains. □

Lemma 2 (Multinomial Splitting Equivalence)**.** *Let N i.i.d. observations have a finite cell distribution $\mathbf{p}$, and partition their fixed indices into disjoint client groups of sizes $n_1, \ldots, n_K$ with $\sum_k n_k = N$. Then the client cell-count vectors are mutually independent and*

$$\mathbf{C}_k \sim \mathrm{Mult}(n_k, \mathbf{p}). \tag{42}$$

Equivalently, the same unconditional joint law is obtained by first drawing $\mathbf{C} \sim \mathrm{Mult}(N, \mathbf{p})$ and then, conditional on $\mathbf{C}$, partitioning the realised cells among fixed-size client groups according to the corresponding multivariate-hypergeometric law.

Proof. Each client count vector is a function of a disjoint subset of the i.i.d. observations, so the vectors are mutually independent and have the multinomial laws in Eq. (42). Conditional on the global cell counts, exchangeability of the i.i.d. sample makes every assignment of the realised cells to the fixed client index groups equiprobable subject to the group sizes. This is exactly the multivariate-hypergeometric partition law. Marginalising the conditional representation therefore recovers the same unconditional joint experiment. □

Remark 3 (Conditional versus Unconditional Partition Laws)**.** Lemma 2 justifies the client-independence assumption used by Theorem 3 for the unconditional fixed-K model. Conditional on one realised global count vector, the client tables are generally dependent. The pathwise Theorem 2 remains exact under that conditioning, whereas a conditional risk calculation must use the induced hypergeometric law rather than the product-CDF factorisation.

B.2 Proof of Support-Induced Factorisation

Proof of Theorem 2. Fix $D \in \mathsf{MinDes}_T$. By Proposition 6, and under the readiness-gated population-label interface (hence clean erasure by Assumption 3), we have

$$\begin{aligned} \{D \subseteq \mathcal{U}^\perp\} &= \bigcap_{I \in D} \{M_I^{\mathrm{loc}} = 0\} \\ &= \bigcap_{I \in D} \{Z_{S_I}^{\mathrm{loc}} < j(I)\}. \end{aligned} \tag{43}$$

For a support block S unused by D, the ceiling coordinate $b_S(D) = d_S$ imposes no restriction. If D uses block S, all inequalities in Eq. (43) hold exactly when

$$Z_S^{\mathrm{loc}} \le \min_{I \in D : S_I = S} j(I) - 1 = b_S(D). \tag{44}$$

Hence

$$\{D \subseteq \mathcal{U}^{\perp}\} = \{\mathbf{Z}^{\mathrm{loc}} \le \mathbf{b}(D)\}. \tag{45}$$

Substituting Eq. (45) into Theorem 1 gives the union over all $\mathbf{b}(D)$. If $\mathbf{b} \le \mathbf{b}'$, then $\{\mathbf{Z}^{\mathrm{loc}} \le \mathbf{b}\} \subseteq \{\mathbf{Z}^{\mathrm{loc}} \le \mathbf{b}'\}$. Therefore duplicate ceilings and coordinatewise dominated ceilings can be removed by keeping exactly the coordinatewise-maximal members, yielding $\mathcal{B}_T$ and Eq. (13). No probability law is used. □

B.3 Proof of Exact Fixed-K Recovery Risk

Proof of Theorem 3. For any support-state ceiling $\mathbf{z}$,

$$\{\mathbf{Z}^{\mathrm{loc}} \le \mathbf{z}\} = \bigcap_{k=1}^{K} \{\mathbf{Z}_k \le \mathbf{z}\}, \tag{46}$$

because $\mathbf{Z}^{\mathrm{loc}}$ is the coordinatewise maximum of the client states. Mutual independence of $\mathbf{Z}_1, \dots, \mathbf{Z}_K$ therefore gives

$$\Pr(\mathbf{Z}^{\mathrm{loc}} \le \mathbf{z}) = \prod_{k=1}^{K} F_k(\mathbf{z}), \tag{47}$$

which is Eq. (17).

For every nonempty $\mathcal{J} \subseteq \mathcal{B}_T$,

$$\bigcap_{\mathbf{b} \in \mathcal{J}} \{\mathbf{Z}^{\mathrm{loc}} \le \mathbf{b}\} = \left\{ \mathbf{Z}^{\mathrm{loc}} \le \bigwedge_{\mathbf{b} \in \mathcal{J}} \mathbf{b} \right\}. \tag{48}$$

Apply finite inclusion–exclusion to the exact union in Theorem 2, substitute Eq. (48), and then use Eq. (17). This gives Eq. (18). If $\mathcal{B}_T = \varnothing$, Theorem 2 gives an empty failure event, so the risk is zero. □

B.4 Finite-Sample Consequences

Corollary 3 is stated in Section 4.4. We state and prove the finite-sample readiness result here, give the complete proof of the capacity consequence, and record a reduced-information consequence for settings in which the full joint client-state law is unavailable.

For $[x]_+ := \max\{x, 0\}$, fix query $I = (X_a, X_b, S)$ and client k, let $p_{k,s} := P_k(S = s)$ and $F_{k,I,s} := \Pr[\mathrm{Bin}(n_k, p_{k,s}) \le m_I - 1]$, and write $r_{k,I} := \Pr(M_{k,I} = 1)$ and $b_I := J_I m_I$.

Lemma 3 (Finite-Sample Readiness Bounds)**.** *If the n_k local observations are i.i.d. from P_k, then*

$$\left[1 - \sum_{s \in \Omega_S} F_{k,I,s}\right]_+ \le r_{k,I} \le \mathbf{1}\{n_k \ge b_I\} \prod_{s \in \Omega_S} (1 - F_{k,I,s}). \tag{49}$$

Both inequalities are equalities when $J_I = 1$.

Define the marginal readiness intensity $\Lambda_I := \sum_{k=1}^{K} r_{k,I}$ and $\mathbf{\Lambda} := (\Lambda_I)_{I \in \mathcal{C}}$. If $\mathsf{MinDes}_T = \varnothing$, set $L_T(\mathbf{\Lambda}) = U_T(\mathbf{\Lambda}) = 0$; otherwise let

$$L_T(\mathbf{\Lambda}) := \left[1 - \min_{D \in \mathsf{MinDes}_T} \sum_{I \in D} \Lambda_I\right]_+, \tag{50}$$

$$U_T(\mathbf{\Lambda}) := \min\left\{1, \sum_{D \in \mathsf{MinDes}_T} \min_{I \in D} e^{-\Lambda_I}\right\}. \tag{51}$$

Corollary 6 (Marginal Readiness-Intensity Risk Bounds)**.** *Under Assumption 3, mutually independent client support states, and the readiness-gated population-label interface,*

$$L_T(\mathbf{\Lambda}) \le \Pr(\mathsf{Fail}_T) \le U_T(\mathbf{\Lambda}). \tag{52}$$

Proof of Lemma 3. For each $s \in \Omega_S$, let

$$B_s := \{N_k(S = s) \le m_I - 1\}. \tag{53}$$

The marginal distribution of $N_k(S = s)$ is $\mathrm{Bin}(n_k, p_{k,s})$, so $\Pr(B_s) = F_{k,I,s}$. Readiness fails exactly when at least one required stratum is short:

$$\{M_{k,I} = 0\} = \bigcup_{s \in \Omega_S} B_s. \tag{54}$$

The union bound therefore gives

$$r_{k,I} \ge \left[1 - \sum_s F_{k,I,s}\right]_+. \tag{55}$$

If $n_k < J_I m_I = b_I$, no allocation of n_k observations can place at least m_I observations in all J_I disjoint conditioning strata, hence $r_{k,I} = 0$. Otherwise, set $A_s := B_s^c$. Multinomial cell counts are negatively associated (Joag-Dev & Proschan, 1983), and the events A_s are increasing events on disjoint coordinates; therefore

$$r_{k,I} = \Pr\left(\bigcap_s A_s\right) \le \prod_s \Pr(A_s) = \prod_s (1 - F_{k,I,s}). \tag{56}$$

Combining this with the deterministic capacity screen gives Eq. (49). If a required stratum has $p_{k,s} = 0$, then $F_{k,I,s} = 1$ and both sides correctly force $r_{k,I} = 0$. If $J_I = 1$, readiness is exactly one binomial tail event, so both inequalities are equalities. □

Proof of Corollary 6. For query I, let $E_I := \{M_I^{\mathrm{loc}} = 0\}$. Mutual client independence gives

$$\Pr(E_I) = \prod_{k=1}^{K} (1 - r_{k,I}) \le e^{-\Lambda_I}, \tag{57}$$

and the union bound gives

$$\Pr(E_I^c) \le \sum_{k=1}^{K} r_{k,I} = \Lambda_I. \tag{58}$$

For any $D \in \mathsf{MinDes}_T$,

$$\begin{aligned} \Pr(\mathsf{Fail}_T) &\ge \Pr\left(\bigcap_{I \in D} E_I\right) \\ &\ge 1 - \sum_{I \in D} \Pr(E_I^c) \\ &\ge 1 - \sum_{I \in D} \Lambda_I. \end{aligned} \tag{59}$$

Taking the positive part and choosing the D with smallest total readiness intensity yields $L_T(\mathbf{\Lambda})$.

For the upper bound, Theorem 1 and a union bound give

$$\Pr(\mathsf{Fail}_T) \le \sum_{D \in \mathsf{MinDes}_T} \Pr\left(\bigcap_{I \in D} E_I\right). \tag{60}$$

Each intersection is contained in every one of its constituent erasure events, so by Eq. (57),

$$\Pr\left(\bigcap_{I \in D} E_I\right) \le \min_{I \in D} e^{-\Lambda_I}. \tag{61}$$

Summing over D and truncating at one yields $U_T(\mathbf{\Lambda})$. □

Proof of Corollary 3. A query I with J_I conditioning strata can be locally ready only if a client contains at least $J_I m_I = b_I$ observations. Therefore, if $\max_k n_k < b_I$, then $M_{k,I} = 0$ for every client and $M_I^{\mathrm{loc}} = 0$ deterministically. If $\max_k n_k < b_I$ holds for every I in some $D \in \mathsf{MinDes}_T$, then $D \subseteq \mathcal{U}^\perp$ deterministically, and Theorem 1 gives target failure.

For remainder-balanced allocation, the largest client has size $\lceil N/K \rceil$. If $b_I \geq 2$, then

$$\left\lceil \frac{N}{K} \right\rceil < b_I \iff \left\lceil \frac{N}{K} \right\rceil \leq b_I - 1 \iff K \geq \frac{N}{b_I - 1}. \tag{62}$$

The smallest integer satisfying this condition is therefore

$$K_{I,0}(N) = \left\lceil \frac{N}{b_I - 1} \right\rceil. \tag{63}$$

When $b_I = 1$, every nonempty client has capacity at least one. Total sample size alone therefore cannot force zero readiness, and the boundary is set to infinity.

If $K \geq K_{T,0}(N)$, choose a destructive set D attaining the minimum in Eq. (21). Then $K \geq K_{I,0}(N)$ for every $I \in D$, so every query in D is locally unavailable. The exact erasure theorem again gives $\Pr(\mathsf{Fail}_T) = 1$. If $K_{T,0}(N) = \infty$, the statement is vacuous, as intended. □

B.5 Fragmentation versus Aggregate Scarcity

The main text separates aggregate scarcity from pure fragmentation at the target-event level. This subsection records two further consequences: how the destructive geometry restricts after some aggregate evidence is already unavailable, and why query-wise support summaries do not in general determine fragmentation risk.

Proposition 7 (Surviving-Certificate Destructive Geometry)**.** *Let $E \subseteq \mathcal{C}$ be an aggregate-erased query set, and suppose the surviving evidence $\mathcal{C}_E := \mathcal{C} \setminus E$ still recovers the complete target summary under the reference labels. Then $\mathcal{C}_E$ is target-sufficient and*

$$\mathsf{MinDes}_T(\mathcal{C}_E) = \mathrm{Min}_{\subseteq}\{D \setminus E : D \in \mathsf{MinDes}_T(\mathcal{C})\}. \tag{64}$$

Consequently, its remaining erasure resilience is

$$\rho_T(\mathcal{C}_E) = \min_{D \in \mathsf{MinDes}_T(\mathcal{C})} |D \setminus E|. \tag{65}$$

Proof. For every target-discrepant candidate G, restricting the certificate removes exactly the erased coordinates,

$$\mathrm{Disc}_{T,\mathcal{C}_E}(G) = \mathrm{Disc}_{T,\mathcal{C}}(G) \setminus E. \tag{66}$$

Taking the inclusion-minimal nonempty discrepancy sets over target-discrepant candidates gives Eq. (64). The assumed aggregate recovery excludes the empty discrepancy set, so $\mathcal{C}_E$ remains target-sufficient. Equation (65) follows by taking minimum cardinality. □

Thus aggregate scarcity can be nonfatal while still reducing the number of additional local-only losses needed for target failure. The construction is a restriction of the same destructive geometry characterised by Theorem 1 rather than a new certificate-selection problem.

Write the formal support-state failure region from Theorem 2 as

$$\mathcal{F}_T := \bigcup_{\mathbf{b} \in \mathcal{B}_T} \downarrow \mathbf{b}. \tag{67}$$

Proposition 8 (Aggregate-Conditioned Support Frontier)**.** *Fix an aggregate support state $\mathbf{a} = \mathbf{Z}^{\mathrm{agg}}$ for which the aggregate interface recovers the target. On the conditioned domain $\mathbf{z} \leq \mathbf{a}$, the local target-failure region satisfies*

$$\mathcal{F}_T \cap \downarrow \mathbf{a} = \bigcup_{\mathbf{b} \in \mathcal{B}_T} \downarrow (\mathbf{a} \wedge \mathbf{b}). \tag{68}$$

After duplicate and coordinatewise-dominated ceilings are removed, define

$$\mathcal{B}_T[\mathbf{a}] := \mathrm{Max}_{\leq}\{\mathbf{a} \wedge \mathbf{b} : \mathbf{b} \in \mathcal{B}_T\}. \tag{69}$$

Then pure fragmentation occurs exactly when

$$\exists\, \mathbf{c} \in \mathcal{B}_T[\mathbf{a}] \quad \textit{such that} \quad \mathbf{Z}^{\mathrm{loc}} \leq \mathbf{c}. \tag{70}$$

Proof. For principal downsets, $\downarrow \mathbf{b} \cap \downarrow \mathbf{a} = \downarrow (\mathbf{a} \wedge \mathbf{b})$. Intersecting Theorem 2 with the pathwise domain $\mathbf{Z}^{\mathrm{loc}} \leq \mathbf{a}$ gives Eq. (68); removing duplicate and contained principal downsets yields Eq. (69). Aggregate success makes every resulting local failure a pure-fragmentation event. □

Support-information hierarchy. Let

$$U_{\mathcal{C}} := \bigcup_{I \in \mathcal{C}} S_I \tag{71}$$

be the union of conditioning variables used by the certificate, and let $C_{U_{\mathcal{C}}}$ denote their realised global joint histogram. Under uniform repartition into fixed client sizes, $C_{U_{\mathcal{C}}}$ determines the joint conditional law of all client support-block counts and hence of the complete local-readiness vector. The separate block histograms $\{C_S : S \in \mathfrak{S}_{\mathcal{C}}\}$ determine each query's conditional readiness probability, while $\mathbf{Z}^{\mathrm{agg}}$ determines only the aggregate-conditioned failure geometry. These summaries are strictly different in strength in general. Proposition 3 states the resulting necessity claim in the main text; the concrete five-variable construction is proved here.

Proof of Proposition 3. Use variables S_1, S_2, X, T, Y and candidate class $\mathcal{G} = \{G^\star, G^-\}$. Let

$$G^\star : S_1 \to X \to T \to Y, \qquad S_2 \to Y, \tag{72}$$

and let G^- reverse only the target edge, giving

$$G^- : S_1 \to X \to T \leftarrow Y \leftarrow S_2. \tag{73}$$

The target summary records the orientation of the T–Y edge. Take

$$\begin{aligned} &I_1 = (X, Y \mid S_1), \quad I_2 = (X, Y \mid S_2), \\ &\mathcal{C} = \{I_1, I_2\}. \end{aligned} \tag{74}$$

In $G^\star$, the path $X \to T \to Y$ is active under either conditioning set, so both reference labels are dependence. In G^-, T is an unconditioned collider on $X \to T \leftarrow Y$, and conditioning on S_1 or S_2 does not open that collider, so both labels are independence. Hence each query alone excludes the sole target-discrepant rival, while losing both re-admits it; therefore $\mathsf{MinDes}_T = \{\{I_1, I_2\}\}$. A generic faithful discrete distribution Markov to $G^\star$ with full support on the two root variables S_1, S_2 makes both support profiles below legitimate positive-probability realised samples while preserving these reference labels.

Now take $N = 4$, $K = 2$, $n_1 = n_2 = 2$, binary S_1, S_2, and thresholds $m_{I_1} = m_{I_2} = 1$. In the aligned profile the four realised support signatures are

$$(0,0), (0,0), (1,1), (1,1). \tag{75}$$

In the crossed profile they are

$$(0,0), (0,1), (1,0), (1,1). \tag{76}$$

Both profiles have block margins $C_{S_1} = C_{S_2} = (2,2)$, so both queries are aggregate-ready. Under a uniform balanced repartition, each query is locally unavailable with probability $2/\binom{4}{2} = 1/3$ in both profiles. In the aligned profile the two unavailability events coincide, so both queries are erased with probability $1/3$ and target failure occurs with probability $1/3$. In the crossed profile the two unavailability events correspond to different balanced cuts and are disjoint, so they never occur together and target failure has probability zero. □

The proposition isolates a support-layer analogue of the controlled dependence analysis for Theorem 1: the causal destructive geometry is fixed, and the query-wise fragmentation probabilities are also fixed, but the shared-record joint support arrangement changes whether the destructive losses occur together. Thus the joint support-state law in Theorem 3 cannot generally be replaced by query-wise readiness probabilities.

Proof of Corollary 2. Take $N = 2n$, two clients of size n, threshold one, and one binary conditioning variable S_j for each unordered balanced bipartition of the $2n$ records. Their number is $M_n = \binom{2n}{n}/2$. Let the finite candidate class be $\mathcal{G} = \{G^\star, G_1, \ldots, G_{M_n}\}$. In the reference DAG $G^\star$, the target T is disconnected from $X_1, \ldots, X_{M_n}$, and each S_j is an isolated full-support root, otherwise disconnected

from $T, X_1, \ldots, X_{M_n}$. Let the target summary record the corresponding T–X_j adjacencies, and let candidate G_j add only $T \to X_j$. Then query $I_j = (T, X_j \mid S_j)$ has singleton discrepancy $\{I_j\}$. Under a reference population in which the S_j are independent full-support roots, every prescribed finite support-signature matrix below has positive probability. In an aligned support profile all S_j induce the same balanced split, so all query-unavailability events coincide and target risk is $1/M_n$. In a separated profile the S_j enumerate all unordered balanced splits, so exactly one query-unavailability event occurs for every balanced repartition and target risk is one. Since $M_n \to \infty$, the gap tends to one. □

Detailed conditional representation and proof of Proposition 4. Let $U_{\mathcal{C}} := \bigcup_{I \in \mathcal{C}} S_I$, let $C := C_{U_{\mathcal{C}}}$ be the realised support-relevant global joint histogram, and define

$$H_c(\mathbf{z}) := \Pr(\mathbf{Z}^{\mathrm{loc}} \le \mathbf{z} \mid C = c) \tag{77}$$

under a fixed-size random repartition. Then the fixed-global placement risk defined in Proposition 4 admits the exact representation

$$\psi_T(c) = \sum_{\varnothing \neq \mathcal{J} \subseteq \mathcal{B}_T} (-1)^{|\mathcal{J}|+1} H_c\left(\bigwedge_{\mathbf{b} \in \mathcal{J}} \mathbf{b}\right). \tag{78}$$

Proof of Proposition 4. Conditioning on $C = c$ changes only the probability law on the support-state failure regions from Theorem 2. Inclusion–exclusion over those regions gives Eq. (78); the pure-fragmentation identity follows from Proposition 2; Eq. (19) is the law of total probability. No client-independence assumption is used conditionally on C. This conditional representation complements rather than replaces Theorem 3, whose product form arises after the client-generation law is specified. □

C Validity and Scope Boundaries

C.1 Incorrect Definite Evidence

The clean-erasure analysis excludes wrong definite evidence. The following proposition records the formal boundary and is followed by a concrete incompatibility witness.

Proposition 9 (Boundary of the Clean-Erasure Geometry)**.** *For general three-state evidence, let F_I^{disc} denote the event $\widetilde{d}(I) \in \{0,1\}$ and $\widetilde{d}(I) \neq d_T^\star(I)$. Then*

$$\mathsf{Fail}_T \subseteq \left(\bigcup_{I \in \mathcal{C}} F_I^{\mathrm{disc}}\right) \cup \left(\bigcup_{D \in \mathsf{MinDes}_T} \{D \subseteq \mathcal{U}^\perp\}\right), \tag{79}$$

and hence

$$\Pr(\mathsf{Fail}_T) \le \sum_{I \in \mathcal{C}} \Pr(F_I^{\mathrm{disc}}) + \sum_{D \in \mathsf{MinDes}_T} \Pr(D \subseteq \mathcal{U}^\perp). \tag{80}$$

Proof of Proposition 9. Let

$$\mathcal{E}_{\mathrm{disc}} := \bigcup_{I \in \mathcal{C}} F_I^{\mathrm{disc}}. \tag{81}$$

On $\mathcal{E}_{\mathrm{disc}}^c$, every definite answer equals its target-reference label, so the realised evidence satisfies Assumption 2. Theorem 1 then gives

$$\mathsf{Fail}_T \cap \mathcal{E}_{\mathrm{disc}}^c \subseteq \bigcup_{D \in \mathsf{MinDes}_T} \{D \subseteq \mathcal{U}^\perp\}. \tag{82}$$

Adding $\mathcal{E}_{\mathrm{disc}}$ to the right-hand side yields the event containment in Eq. (79). Taking probabilities and applying the union bound separately to the discordant-evidence events and destructive-erasure events gives Eq. (80). □

This proposition deliberately does not model why a discordant definite answer occurs. Source semantics, source selection, agreement, and reference-free auditing are outside the present support-fragmentation theory.

Concrete incompatibility witness. In the two-candidate construction of Proposition 3, the reference labels for I_1 and I_2 are dependence (0), whereas the target-discrepant rival has independence labels (1). If, for example, I_1 is fully visible but is assigned the discordant definite label $\widetilde{d}(I_1) = 1$, then $G^\star$ is excluded while G^- remains compatible. The compatible class is nonempty, yet its target report is wrong. Thus complete visibility and nonempty compatibility do not turn incorrect definite evidence into an erasure event.

C.2 Readiness Threshold as an Interface Contract

The main results require only that each query I have a prespecified integer threshold m_I defining when its conditioning support is sufficient for the chosen evidence-formation interface. The value may come from a tester-specific power calculation, an application-level minimum-cell rule, or another externally validated contract. No tester-specific validity theorem is needed for Theorems 1–3. Keeping m_I exogenous prevents a particular CI tester, selector, or multiplicity correction from becoming an additional scientific object of this paper.

C.3 Target-Scope Applicability

Proposition 10 (Target-Scope Adequacy Boundary)**.** *Perfect evidence inside a reduced target scope need not determine the intended full-system target summary. In particular, omitting a mediator can make a direct-versus-mediated causal target observationally indistinguishable within the reduced scope.*

Proof. Fix $\varepsilon \in (0, 1/2)$ and let $X \sim \mathrm{Bernoulli}(1/2)$. In Model A, let

$$Y = X \oplus N, \qquad N \sim \mathrm{Bernoulli}(\varepsilon), \tag{83}$$

with N independent of X, and let the full graph contain the direct edge $X \to Y$.

In Model B, introduce an omitted mediator M and set

$$M = X \oplus N_1, \qquad Y = M \oplus N_2, \tag{84}$$

where N_1, N_2 are mutually independent and independent of X. Choose $q \in (0, 1/2)$ satisfying $2q(1-q) = \varepsilon$, and take $N_1, N_2 \sim \mathrm{Bernoulli}(q)$. Then

$$Y = X \oplus (N_1 \oplus N_2), \qquad \Pr(N_1 \oplus N_2 = 1) = \varepsilon, \tag{85}$$

so Models A and B induce exactly the same reduced joint distribution $P(X, Y)$.

Their full causal interpretations differ: Model A contains a direct edge $X \to Y$, whereas Model B contains only the mediated path $X \to M \to Y$. Hence even perfect observational or CI evidence restricted to $\{X, Y\}$ cannot distinguish the intended direct-versus-mediated full-system target. Target-scope adequacy is therefore logically prior to support fragmentation and finite-sample readiness. □

D Empirical Analyses and Reproducibility

D.1 Experimental Contract and Fixed Settings

The empirical program mirrors the theory chain and reports evidence that instantiates Theorem 1, Proposition 5, Theorems 2–3, Corollaries 3 and 6, and the target-scope boundary in Proposition 10. Experiments outside the fixed-certificate support-fragmentation regime are not used to support these claims.

The main benchmark is ALARM with

$$\begin{aligned} T &= \mathrm{LVEDVOLUME}, \\ \mathcal{V}_T &= \{\mathrm{HYPOVOLEMIA}, \mathrm{LVFAILURE}, \mathrm{LVEDVOLUME}, \mathrm{CVP}, \mathrm{PCWP}\}. \end{aligned} \tag{86}$$

The intended target summary records the four target-local relations

$$\begin{aligned} &\textsc{Hypovolemia} \to \textsc{LVEDVolume}, \\ &\textsc{LVFailure} \to \textsc{LVEDVolume}, \\ &\textsc{LVEDVolume} \to \textsc{CVP}, \\ &\textsc{LVEDVolume} \to \textsc{PCWP}. \end{aligned} \tag{87}$$

The five-node candidate class contains all $29{,}281$ labelled DAGs. The order-one evidence universe contains 40 symmetric CI queries. The compact certificate C_6 is an inclusion-minimal target-sufficient family; C_7 is an inactive one-query extension; C_{12}^{aug} is a twelve-query nested extension fixed before realised-count trials; and Q_{40} is the full order-one family. These families are oracle-certified experimental fixtures used evaluatively rather than outputs of a proposed certificate-learning algorithm; the paper does not optimise over the family of sufficient certificates.

The six queries in C_6 are

$$\begin{aligned} &\textsc{Hypovolemia} \perp\!\!\!\perp \textsc{LVFailure}, \\ &\textsc{CVP} \perp\!\!\!\perp \textsc{PCWP} \mid \textsc{LVEDVolume}, \\ &\textsc{Hypovolemia} \not\perp\!\!\!\perp \textsc{LVFailure} \mid \textsc{LVEDVolume}, \\ &\textsc{Hypovolemia} \perp\!\!\!\perp \textsc{PCWP} \mid \textsc{LVEDVolume}, \\ &\textsc{LVFailure} \perp\!\!\!\perp \textsc{CVP} \mid \textsc{LVEDVolume}, \\ &\textsc{CVP} \not\perp\!\!\!\perp \textsc{PCWP} \mid \textsc{Hypovolemia}. \end{aligned} \tag{88}$$

All visible queries receive fixed reference labels; the realised-count experiment changes only support availability. The numerical readiness thresholds m_I were fixed before the realised-count trials. In the benchmark configuration they were generated from the declared signal floor 0.30 and nominal $(\alpha, \beta) = (0.05, 0.10)$ settings; the present theory treats the resulting thresholds as fixed interface inputs and does not use their derivation as a tester-validity claim.

D.2 Controlled Exact Failure Geometry

The controlled five-variable instance uses abstract variables (A, B, T, C, D), reference edges $A \to T$, $B \to T$, $T \to C$, and $B \to D$, and a target summary with four components: $A \to T$, $B \to T$, $T \to C$, and nonadjacency between D and T. It enumerates the same $29{,}281$ DAG catalogue but uses a separate fixed ten-query certificate. Exact enumeration yields Table 1. No target-discrepant DAG has zero certificate discrepancy, so the certificate is target-sufficient for this supplied target summary. The 18 minimal destructive sets form an antichain with minimum size two.

Table 1: Controlled exact discrepancy geometry.

Quantity	Value
Candidate DAGs	29,281
Target-discrepant DAGs	29,033
Unique discrepancy sets	272
Minimal destructive sets	18
Size profile	$6 \times 2,\ 10 \times 3,\ 2 \times 4$
ρ_T	2

For reproducibility, Table 2 fixes the exact ten-query order used by the controlled erasure masks and matched partitions.

For the matched-marginal dependence analysis, query I is erased by $U_I^{\perp} = Z_{g(I)} \vee E_I$, with group-shock probability $\gamma = \kappa a$ and residual-erasure probability $\delta = (a - \gamma)/(1 - \gamma)$. The selected aligned and avoiding partitions have the same group sizes $(3, 3, 3, 1)$ and the same query marginals. At $a = 0.1$ and $\kappa = 1$, the aligned groups each contain a complete minimal destructive set, whereas the avoiding groups contain none. The exact operating point is reproduced in Table 3. Thus the dependence analysis changes only how equal marginal erasure mass aligns with the causal failure family.

Table 2: Fixed controlled-certificate order for the exact dependent-erasure analysis. The symbols T, A, B, C, D are the abstract variables of the controlled construction, not ALARM variable indices.

Pos.	Tested CI statement	Reference label
0	$T \perp\!\!\!\perp D \mid B$	ind.
1	$A \perp\!\!\!\perp B$	ind.
2	$A \perp\!\!\!\perp B \mid T$	dep.
3	$A \perp\!\!\!\perp C \mid T$	ind.
4	$A \perp\!\!\!\perp D \mid C$	dep.
5	$A \perp\!\!\!\perp C \mid B$	dep.
6	$T \perp\!\!\!\perp D$	dep.
7	$C \perp\!\!\!\perp D \mid T$	ind.
8	$A \perp\!\!\!\perp D$	ind.
9	$C \perp\!\!\!\perp D \mid B$	ind.

Table 3: Matched-marginal erasure mechanisms in the controlled instance. All ten queries have identical marginal erasure probability $a_I = 0.1$; only dependence alignment with the fixed destructive family changes.

Mechanism	a_I	$\mathbb{E}\lvert\mathcal{U}^{\perp}\rvert$	Exact failure
Destructive-set avoiding	0.1	1	0.0442
Independent	0.1	1	0.062827
Destructive-set aligned	0.1	1	0.2710

Exhaustive matched-partition analysis. To rule out a cherry-picked dependence construction, we exhaustively enumerate every unlabelled partition of the ten controlled queries into group sizes $(3, 3, 3, 1)$:

$$10\,\frac{9!}{(3!)^3 3!} = 2{,}800. \tag{89}$$

At $a = 0.1$ and $\kappa = 1$, all 2,800 partitions preserve the same query-wise marginals, expected erased-query count, group sizes, and dependence strength. Exact target-failure probability ranges from 0.0280 to 0.2710. The selected destructive-set-aligned construction attains the global maximum, while the selected destructive-set-avoiding construction lies in the low-risk region but is not the minimum-risk partition. Table 4 groups the exhaustive census by the number of shock groups that individually contain at least one complete minimal destructive set.

Table 4: Exact distribution over all 2,800 matched partitions, grouped by the number of shock groups containing at least one complete minimal destructive set. The singleton group cannot itself be destructive because $\rho_T = 2$.

Complete destructive-set groups	Count	Min.	Q1	Median	Mean	Q3	Max.
0	490	0.0280	0.0361	0.0442	0.042778	0.0442	0.0523
1	1070	0.1090	0.1171	0.1171	0.120143	0.1252	0.1252
2	940	0.1900	0.1900	0.1981	0.195377	0.1981	0.1981
3	300	0.2710	0.2710	0.2710	0.271000	0.2710	0.2710

D.3 ALARM FIXED-CERTIFICATE GEOMETRY

Exact candidate enumeration gives the family summary in Table 5. The inactive C_7 extension has zero participation in every minimal destructive set relative to C_6, while the twelve-query extension fixed before the realised-count trials removes five of the six singleton bottlenecks. The full family still contains one singleton minimal destructive set, so all four families have $\rho_T = 1$ even though their total destructive geometry differs substantially.

Under homogeneous independent query erasures with $a = 0.1$, the corresponding exact failure probabilities are 0.468559, 0.468559, 0.152841, and 0.118088. The equality for C_6 and C_7 verifies that adding evidence does not improve fragility unless the added query changes the target-induced

Table 5: Exact ALARM fixed-certificate geometry and target-failure risk under homogeneous independent query erasure $a = 0.1$. Adding an inactive query ($C_6 \to C_7$) changes neither the destructive family nor the risk; added evidence helps when it reshapes target-destructive losses.

Family	$\lvert\mathcal{C}\rvert$	Minimal sets	Singleton	P_{fail}
C_6	6	6	6	0.468559
C_7	7	6	6	0.468559
C_{12}^{aug}	12	17	1	0.152841
Q_{40}	40	19	1	0.118088

destructive sets. Corollary 5 formalises the relevant nested-family monotonicity under extension consistency. The comparison is used only as a fixed-family robustness analysis; no optimality claim over alternative certificates is made.

D.4 PAIRED REALISED-COUNT FRAGMENTATION

Population and paired sampling. The homogeneous ALARM benchmark population is sampled at total budget $N = 160{,}000$. For trial b,

$$\mathbf{C}^{(b)} \sim \text{Multinomial}(N, \mathbf{p}). \tag{90}$$

For each evaluated K, the same global table $\mathbf{C}^{(b)}$ is allocated without replacement to remainder-balanced client sizes, giving client tables $\mathbf{C}_k^{(b,K)}$ with

$$\sum_{k=1}^{K} \mathbf{C}_k^{(b,K)} = \mathbf{C}^{(b)}. \tag{91}$$

All conditioning-stratum counts are marginals of these same client tables, so query erasures can be strongly dependent. Local and aggregated readiness use the threshold rule in Definition 5. The evidence channel satisfies Assumption 2.

Seeds $42, \ldots, 51$ define ten random-number streams for the same fixed population. Each stream contributes 500 independent global draws, for $5{,}000$ trial-level outcomes. The complete client grid is

$$\begin{aligned}\mathcal{K}_{\text{eval}} = \{&2, 4, 6, 7, 8, 9, 10, 11, 12, 14, 16, 20,\\ &24, 25, 26, 27, 28, 31, 32, 33, 34, 35\}.\end{aligned} \tag{92}$$

The full local failure counts are shown in Table 6. Every entry after $K = 10$ is $5{,}000/5{,}000$ for all three families.

Table 6: Local target failures out of $5{,}000$ paired trials. The ≥ 20 column summarises all evaluated $K \geq 20$ points, all of which are complete failures; $K = 11$ is also complete failure for every family.

Family	2	4	6	7	8	9	10	12	16	≥ 20
C_6	0	0	0	1331	5000	5000	5000	5000	5000	5000
C_{12}^{aug}	0	0	0	0	0	3694	5000	5000	5000	5000
Q_{40}	0	0	0	0	0	3694	5000	5000	5000	5000

Define the paired rescue event

$$\mathsf{Rescue}_K = \{C_6 \text{ fails and } C_{12}^{\text{aug}} \text{ recovers}\}. \tag{93}$$

The rescue counts at $K = 7, 8, 9$ are respectively $1{,}331$, $5{,}000$, and $1{,}306$ out of $5{,}000$; at $K = 8$ every compact-family failure is rescued on the same realised global table and client partition. No nested-family recovery-monotonicity violation occurs on the complete evaluated trace.

Support-block contraction. For C_{12}^{aug}, 17 query-level minimal destructive sets map to 12 distinct support rectangles and reduce to the six coordinatewise-undominated rectangles in Table 7. A condition "block S at threshold m unavailable" means no client reaches that weakest-stratum threshold.

Table 7: Coordinatewise-undominated support rectangles for C_{12}^{aug} in the formal support-state poset. Some rectangles can be impossible or operationally redundant under a particular attainable-state law.

Index	Support-level constraints	Role
0	Empty-conditioning threshold $m = 1127$ unavailable	impossible
1	CVP $m = 1347$, HYPOVOLEMIA $m = 1985$, and PCWP $m = 1645$ unavailable	residual
5	CVP $m = 1645$, HYPOVOLEMIA $m = 1985$, and PCWP $m = 1347$ unavailable	residual
6	CVP $m = 1645$ and LVEDVOLUME $m = 2066$ unavailable	residual
9	LVEDVOLUME $m = 1645$ unavailable	dominant
11	LVEDVOLUME $m = 2066$ and PCWP $m = 1645$ unavailable	residual

For the compact family, the first bottleneck is CVP $\perp\!\!\!\perp$ PCWP $\mid$ LVEDVOLUME, with $J_I = 3$, threshold $m_I = 2066$, and conditioning-state probabilities $(0.0886, 0.7019, 0.2095)$. Its local invisibility rate is 0.2662 at $K = 7$ and one at $K = 8$, matching the compact-family target-failure event because it is a singleton minimal destructive set.

D.5 EXACT RISK AND RECOVERY TRANSITION

The paired construction is conditional on the same global realised table across K, but at each fixed K its unconditional distribution coincides with independent multinomial client tables of the prescribed sizes by Lemma 2. This is the distributional regime used for the fixed-K calculation below.

At C_{12}^{aug}, $N = 160{,}000$, and $K = 9$, two client sizes are 17,777 and seven are 17,778. For the dominant rectangle in Table 7, exact three-state multinomial calculation gives

$$\begin{aligned} R_{\mathrm{dom}} &= \Pr\{\text{no client reaches LVEDVOLUME } m = 1645\} \\ &= 0.7314724572117319. \end{aligned} \tag{94}$$

The remaining five rectangles are impossible or have much smaller joint lower-tail probabilities. A log-domain union bound over their residual contributions gives

$$0 \le \Pr(\mathsf{Fail}_T) - R_{\mathrm{dom}} \le 2.1303508961 \times 10^{-186}. \tag{95}$$

Thus the exact risk is numerically determined by Eq. (94) to far beyond the reported precision. The empirical count $3{,}694/5{,}000$ gives rate 0.7388 and Wilson 95% interval $[0.7264, 0.7508]$.

Conditional fixed-global placement diagnostic. To isolate the conditional placement term in Proposition 4, we additionally fix representative realised LVEDVOLUME rare-state totals and repeatedly repartition only the same global support. For $K = 9$, five totals spanning approximately -2 to $+2$ population standard deviations each use 10^6 repartitions over seeds 42–46. Table 8 shows that the conditional dominant-block failure probability varies substantially even though N, K, the certificate, and the readiness threshold remain fixed. At the mean three-state global table $(14176, 112304, 33520)$, a second 10^6-repartition check gives full three-stratum failure rate 0.772405 and rare-state-only failure rate 0.772405, with zero event mismatches. For the same five rare-state totals at $K = 8$, the dominant failure event is deterministically impossible because every total exceeds $8(m - 1) = 13152$.

Table 8: Repeated-repartition diagnostic for the dominant LVEDVOLUME block of C_{12}^{aug} at $K = 9$. Each row conditions on a fixed global rare-state total and uses 10^6 repartitions. This is a placement-only diagnostic, not a replacement for the full $5{,}000$-trial target-risk evaluation.

Rare-state total	Approx. location	$K = 9$ conditional failure
13,949	-2σ	0.964142
14,062	-1σ	0.902999
14,176	mean	0.772690
14,290	$+1\sigma$	0.548658
14,403	$+2\sigma$	0.272169

Trace-level minimal-destructive-set event collapse at $K = 9$. The four queries conditioned on LVEDVOLUME that use threshold $m_I = 1645$ share the same realised three-stratum counts, so their erasure indicators are identical on every paired trial. On the complete $5{,}000$-trial $K = 9$ trace, the 17 minimal-destructive-set events of C_{12}^{aug} and the 19 events of Q_{40} each collapse to only two realised event vectors: an identically zero vector and the common erasure vector. The common vector occurs in $3{,}694$ trials and coincides trialwise with target failure. Table 9 records the event-level analysis. Eight minimal destructive sets remain protocol-unresolved outside the realised trace, so this collapse is reported as a complete-trace diagnostic rather than a general identity over all possible count vectors.

Table 9: $K = 9$ minimal-destructive-set event overlap analysis on the complete $5{,}000$-trial trace. "Zero" and "common" classify realised event vectors; "unresolved" counts minimal destructive sets whose zero event is not proved for every possible realised count vector under the stated protocol.

Family	Minimal sets	Vectors	Zero	Common	Unresolved
C_{12}^{aug}	17	2	15	2	6
Q_{40}	19	2	17	2	2

As an additional singleton check, at $C_6, K = 7$ the exact probability that no client satisfies the first bottleneck is 0.263071, versus empirical failure $1{,}331/5{,}000 = 0.2662$ with Wilson interval $[0.2541, 0.2786]$. The marginal readiness-intensity bounds give $[0.462945, 1]$ for $C_{12}^{\mathrm{aug}}, K = 9$; the interval is valid but deliberately loses the joint blocker information used in Eq. (94).

D.6 AGGREGATED COMPARATOR AND CAPACITY BOUNDARY

Across all 22 evaluated client counts, every query in every reported family is aggregated-visible, the aggregated compatible class has size one, and aggregated target failure is zero (Table 10). This is a pathwise control: the aggregated mask and aggregated counts are invariant across K within each trial.

Table 10: Aggregated control, identical at every evaluated K.

Family	Aggregated visible	Mean VS	Failure
C_6	6	1	0
C_{12}^{aug}	12	1	0
Q_{40}	40	1	0

The deterministic local-capacity boundaries are compared with operational transitions in Table 11. The large separation is expected: the capacity boundary uses only total local sample capacity, while the earlier empirical transition is driven by unequal conditioning-state probabilities, especially the rare 0.0886 state.

Table 11: Operational and deterministic landmarks under the fixed-total local non-pooling contract. "Onset" is the first evaluated K with nonzero failure; "complete collapse" is the first with $5{,}000/5{,}000$ failures.

Family	Onset	Observed collapse	Capacity zero
C_6	7	8	26
C_{12}^{aug}	9	10	33
Q_{40}	9	10	33

D.7 CROSS-TARGET CHILD FRAGMENTATION VALIDATION

We use CHILD/GRUNTINGREPORT as a second target-sufficient natural benchmark instance. The five-node scope is {GRUNTING, GRUNTINGREPORT, DISEASE, LUNGPARENCH, SICK}. The target and the three structural query families are fixed before the realised-count Monte Carlo evaluation. In the exact structural analysis, the three eligible CHILD targets GRUNTINGREPORT, RUQO2, and XRAYREPORT have the same compact size, augmentation size, and homogeneous-erasure risks; hence the choice of GRUNTINGREPORT does not depend on its realised fragmentation outcome.

Table 12 reports structural robustness under the homogeneous independent-query-erasure diagnostic with $a = 0.1$. These probabilities are not fixed-K realised-count risks: the former place an abstract erasure law directly on queries, whereas the latter are induced by client support under the fixed-budget sampling model.

Table 12: Exact CHILD/GRUNTINGREPORT fixed-family geometry and homogeneous independent-erasure risk at $a = 0.1$. These risks are structural diagnostics and are not the realised-count fixed-K risks reported below.

Family	$\lvert\mathcal{C}\rvert$	Minimal sets	Singleton	P_{fail}
C_5	5	5	5	0.409510
C_8^{aug}	8	12	1	0.169771
Q_{40}	40	21	1	0.118807

The compact and augmented risks are reproduced by exact structural enumeration, while the Q_{40} value is taken from the exact analysis of the fixed full order-one family. The three-query augmentation in C_8^{aug} is fixed before the realised-count trials. Thus the comparison evaluates how pre-specified evidence families behave after support is allocated across clients rather than selecting queries from the Monte Carlo outcomes.

For the realised-count validation, the total budget is $N = 160{,}000$. Seeds $42, \ldots, 51$ define ten random-number streams, each contributing 500 independent global count draws. Within each trial, one global joint-count table is drawn and the same realised table is reused when records are repartitioned across the evaluated client counts. Local and matched aggregate readiness use the same fixed per-stratum thresholds, so changes across K are induced by support placement rather than by changes in the realised global counts.

A coarse transition scan brackets the change between $K = 4$ and $K = 6$. Before the $5{,}000$-trial evaluation, we therefore add the unique intervening integer $K = 5$, giving the evaluation grid $K \in \{2, 4, 5, 6\}$. This refinement is determined by the bracketed interval rather than by selecting among multiple client counts according to their observed failure rates.

Table 13: Paired CHILD/GRUNTINGREPORT realised-count validation. Entries are local target failures out of $5{,}000$ trials for each query family; the last column uses the matched aggregate interface on the same realised global tables.

K	C_5	C_8^{aug}	Q_{40}	Aggregate
2	0/5000	0/5000	0/5000	0/5000
4	0/5000	0/5000	0/5000	0/5000
5	5000/5000	5000/5000	5000/5000	0/5000
6	5000/5000	5000/5000	5000/5000	0/5000

The empirical failure rate is one at $K = 5$ and $K = 6$, but the corresponding population failure probability is not asserted to equal one exactly. At $K = 4$, the dominant one-block calculation gives $R_T = 2.801468 \times 10^{-11}$ for each of the three families. At $K = 5$, a rarest-state tail calculation gives $R_T \geq 1 - 5.87022 \times 10^{-11}$, and at $K = 6$ the corresponding bound is $R_T \geq 1 - 5.33617 \times 10^{-44}$. Thus the observed $5{,}000/5{,}000$ failures are consistent with a near-deterministic probabilistic transition rather than being used to assert an exact population probability of one.

The three query families have different destructive geometries and do not have identical complete blocker families. Over the $K = 4$–6 transition, however, their risk is controlled by the same operationally dominant DISEASE support block, which has six conditioning states and threshold $m = 1783$. Its rarest state has probability 0.047551016. The expected rarest-state count is approximately $40{,}000(0.047551016) \simeq 1902$ at $K = 4$, above the threshold, but only $32{,}000(0.047551016) \simeq 1522$ at $K = 5$, below it.

This transition is not the deterministic local-capacity boundary of Corollary 3. At $K = 5$, each client still has $n_k = 32{,}000$ observations, whereas the six-state minimum total required by this block is only $6 \times 1783 = 10{,}698$. The sharp change is therefore driven by the lower tail of an unequal conditioning-state distribution rather than by the deterministic condition $n_k < J_I m_I$.

The ALARM and CHILD analyses therefore expose complementary sides of the same mechanism. In ALARM, added evidence improves realised robustness when it creates effective alternative support routes. In CHILD/GRUNTINGREPORT, homogeneous-erasure risk decreases substantially across the enlarged families, yet the realised fixed-budget transition is unchanged because those families retain a common operational bottleneck. Query-level redundancy and support-route redundancy are therefore distinct.

D.8 APPLICABILITY BOUNDARIES

The support-fragmentation theory is interpreted only after checking target-scope adequacy and complete-evidence sufficiency. Table 14 reports the exact census over all targets in ALARM, CHILD, and INSURANCE. Fourteen of 84 targets exceed the five-node degree budget and 26 additional targets have no adequate five-node scope. Of the remaining 44, twenty are still target-insufficient even with the complete within-scope CI family.

Table 14: All-target applicability and minimum sufficient conditioning order. The lower block partitions the 44 adequate target-scope instances.

Diagnostic	Alarm	Child	Insurance	Total
All targets	37	20	27	84
Target degree > 4	4	2	8	14
No adequate five-node scope	7	4	15	26
Adequate five-node scope	26	14	4	44
Minimum sufficient order $r = 0$	7	0	0	7
Minimum sufficient order $r = 1$	5	3	0	8
Minimum sufficient order $r = 2$	5	2	1	8
Minimum sufficient order $r = 3$	1	0	0	1
Complete-CI insufficient	8	9	3	20

A separate CHILD/DISEASE instance makes the complete-evidence boundary explicit. In the five-node scope used for this applicability analysis, all 40 reference labels with conditioning sets of size at most one reduce the $29{,}281$ candidate DAGs to five, but those five disagree on the target-incident orientations. Order-two evidence leaves the same five compatible DAGs. Hence failure in this instance cannot be attributed to federated erasure: the supplied evidence family is not target-sufficient even at complete visibility. This negative-control target is distinct from the target-sufficient CHILD/GRUNTINGREPORT instance in Appendix D.7; the former diagnoses complete-evidence insufficiency, whereas the latter evaluates support fragmentation after target sufficiency has been established.

The detailed INSURANCE signal/readiness sweeps serve only as an applicability boundary because scope, support, and signal limitations coexist there; mechanism validation therefore relies on settings that isolate the support-fragmentation mechanism in Theorems 2–3.

D.9 REPRODUCIBILITY AND INTEGRITY CHECKS

Table 15 consolidates the settings for the experiments reported in this paper.

Seed stability. The ten seed-specific compact-family failure rates at $K = 7$ range from 0.252 to 0.320, and the augmented-family rates at $K = 9$ range from 0.712 to 0.778. At $K = 8$, every seed gives compact-family failure one and augmented-family failure zero. These ranges diagnose Monte Carlo stability for one fixed population; they are not variation across benchmark populations.

Compatible-class recovery and the minimal-destructive-set hit criterion were computed independently on every family-interface row. Across the complete applicable traces and grids, we observed no violations of the recovery criterion, local–aggregate dominance, nested-family monotonicity, fixed-global aggregate-mask invariance, or the deterministic capacity boundary.

The cross-target CHILD/GRUNTINGREPORT validation contains $10 \times 500 \times 4 \times 3 = 60{,}000$ family-level paired trial rows across ten seeds, four client counts, and three query families. Every

Table 15: Parameters and run scale for the reported experiments. In the paired experiment, the pre-specified support thresholds are induced by the readiness specification with minimum-signal input 0.30 and $(\alpha, \beta) = (0.05, 0.10)$; the same values are used throughout the paired trials.

Evaluation component	Fixed setting	Enumeration / replication scale
Controlled exact geometry	Five abstract variables; fixed ten-query certificate	All 29,281 labelled DAGs; exact enumeration
Matched-partition dependence analysis	$a = 0.1$, $\kappa = 1$, group sizes $(3, 3, 3, 1)$	All 2,800 matched partitions; exact probabilities
ALARM fixed-family geometry	Five-node target scope; $C_6, C_7, C_{12}^{\mathrm{aug}}, Q_{40}$	Complete candidate-DAG enumeration under reference labels
Paired realised-count experiment	ALARM; $N = 160{,}000$; fixed thresholds m_I; 22 client counts	Seeds 42–51; 500 paired trials per seed; 5,000 total
Fixed-K risk calculation	$C_{12}^{\mathrm{aug}}, K = 9$; client sizes $17{,}777 \times 2$ and $17{,}778 \times 7$	Exact multinomial tail plus residual union bound
Conditional placement diagnostic	$C_{12}^{\mathrm{aug}}, K = 9$; five fixed rare-state totals; mean three-state table	Seeds 42–46; 10^6 repartitions per representative total; 10^6 three-state checks at the mean table
Cross-target fragmentation validation	CHILD/GRUNTINGREPORT; $N = 160{,}000$; $C_5, C_8^{\mathrm{aug}}, Q_{40}$; $K \in \{2, 4, 5, 6\}$	Seeds 42–51; 500 paired trials per seed; 5,000 per K
Applicability census	ALARM, CHILD, INSURANCE	All 84 targets; exact scope / complete-evidence analysis

$(\text{seed}, \text{trial}, K, \text{family})$ key is unique, each K–family cell contains exactly $5{,}000$ trials, the matched aggregate interface has zero target failures, and no compact-to-augmented or augmented-to-Q_{40} recovery-monotonicity violation is observed. Additional checks confirm the structural, trial-integrity, calibration, and blocker conditions used in the reported analysis.

The random-number streams are the consecutive seeds 42–51, each contributing 500 global draws. Reported figures are generated directly from the tabulated numerical outputs, and no unobserved trace value is imputed into any reported result.