BLIND BY CONSTRUCTION: RECORD-LEVEL UNLEARNING AUDITS MISS RETAINED-CARRIER RECONSTRUCTION ON FROZEN TIME-SERIES FEATURES

Anonymous authors
Paper under double-blind review

ABSTRACT

A record-level unlearning audit compares a model that unlearned a deleted record with one that never saw it, so it is blind by construction to what the retained records reveal about the deleted subject. When a deletion covers a subject with many records, a correlated record from another subject stays in the pool; we call such a record a retained carrier. For frozen time-series features with a ridge readout, we derive an exact identity for how much a carrier adds to reconstructing the deleted subject. Measuring this contribution requires finding carriers, and cosine similarity on frozen features fails here: the features share one dominant direction, so it marks almost every retained record as a carrier and inflates the contribution. After de-centering, the artifact disappears: the all-subject median excess is 0.000 on CWRU, PTB-XL and MIMII and +0.046 on Paderborn. Planted carriers with known effects often go undetected, so these zeros state the measurement's detection limit rather than the absence of leakage. Membership inference on the same data remains accurate. The two channels measure different quantities: whether a deleted record is still distinguishable, and how much a retained carrier lets an attacker reconstruct the deleted subject. Unlearning evaluations should report both, each with its measurement range.

1 INTRODUCTION

A trained model may be asked to forget. When a patient or a machine operator asks that their records be deleted, the model should behave as if those records had never been used. An audit checks that it does (Cao & Yang, 2015; Bourtoule et al., 2021). The standard audit is record-level. It compares two versions of the model: one trained on the deleted record and then unlearned it, and one never saw it. It passes if a membership test cannot tell them apart (Shokri et al., 2017; Carlini et al., 2022). In industrial and clinical time series, a deletion request covers a subject with many records, such as a bearing, a patient, or a machine. Records correlated with that subject stay in the training pool after its own records are gone. Both compared versions keep those records, so the audit is blind by construction to what they still reveal about the deleted subject. This paper builds the instrument that measures that channel, shows that the standard cosine detector fabricates it, and reports what the corrected instrument sees together with its detection limit.

We call a retained record that is correlated with the deleted subject a retained carrier. Another bearing from the same test rig is the simplest example. A reconstructed signal could re-identify a patient or reveal a machine's operating state. Records of relatives or of similar units are known to reveal a person's information even when that person's own records are absent (Humbert et al., 2013; Jayaraman & Evans, 2022). Duplicates are known to confound unlearning (Ye et al., 2025). Two questions follow: how much a retained carrier adds to the reconstruction of a deleted subject, and how reliably that contribution can be measured (Figure 1).

We answer both questions for a common deployment in which a time-series foundation model serves as a frozen encoder with a small learned readout, so that a deletion refits only the readout (Ansari et al., 2024a; Das et al., 2024). The carrier's contribution is the change in the deleted subject's

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.

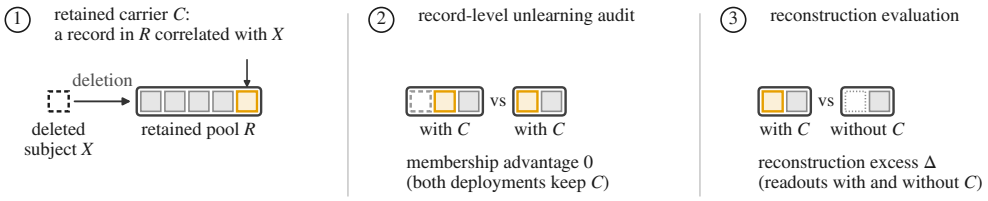


Figure 1: A record-level audit compares two versions that both keep the correlated record C; a reconstruction evaluation compares readouts with and without C. In this work, we measure how much C adds to reconstructing the deleted subject X; Figure 3 reports the headline readings.

reconstruction error when the carrier is removed from the pool. For ridge readouts on fixed features we derive this change in closed form. The closed form attributes the change exactly to a stated set of carriers. On real data an analyst must then find the carriers, refit the readout with and without them, and summarize the result. Our primary summary is the median contribution across all deleted subjects. To learn how small a contribution this procedure can detect, we also plant carriers with known effects.

The first step, finding the carriers, already fails. Carriers are usually found by cosine similarity between frozen features. Frozen time-series features share one dominant direction, so cosine similarity flags at least 88% of retained records as carriers on three encoder families. Refitting the readout without so many records leaves it under-determined, and the resulting change appears as a large carrier contribution. The detector therefore inflates the contribution it is meant to measure. Subtracting the shared direction before computing similarity, which we call de-centering, removes the artifact. With de-centering, the pre-registered median contribution is 0.000 on CWRU, PTB-XL, and MIMII and small, with an interval that includes zero, on Paderborn. A subject with no detected carrier contributes exactly zero, and on these three corpora most subjects have none. Five positive-control readings on planted carriers fell short of the criteria fixed in advance. Membership tests on the same data remain accurate; they measure whether the deleted record is still distinguishable, not what the retained carriers reveal, so the two audits answer different questions.

We make three contributions.

- **The record-level audit does not measure the carrier channel.** Both compared versions keep every retained record (Proposition 1), so its outcome is unchanged by whatever those records reveal, and membership-test success and carrier contribution are different quantities that an evaluation should report side by side.
- **An exact formula for the carrier contribution.** For ridge and kernel readouts on fixed features, we give a closed-form expression for the change in reconstruction error caused by a stated set of carriers. It requires no asymptotic assumption and states when an exact duplicate leaks.
- **A repaired detector and a measurement with a stated detection limit.** We show that cosine similarity on frozen features is an artifact-prone carrier detector and repair it by de-centering. We report these measurements on four corpora: zero on three and small on the fourth, at the resolution that the injected-effect ladder and the positive-control readings establish.

2 Setting, Retained Carriers, and the Two Channels

A deletion leaves two channels open to measurement: whether the deleted record is still distinguishable, which the record-level audit tests, and how much a retained carrier lets the deployed readout reconstruct the deleted subject, which that audit holds fixed (Contribution 1).

2.1 Roles and Access

Three roles share one interface. The maintainer owns the pre-deletion population P and the deletion request G, runs training and unlearning, and deploys the feature map ψ and the readout w_R. The post-only analyst receives only what remains deployed after the request is honored: ψ, w_R, and the retained pool $R = P \setminus G$, grouped into one retained unit per subject; it sees no earlier checkpoint,

parameter update, or logit trace, and no raw record of the deleted subject, the unit that owns every record in G. The evaluator holds the deleted subject X, sends the analyst one signed descriptor z_X (the mean feature of X minus the retained-pool mean μ_R that the analyst supplies), and alone scores every readout the analyst returns (Appendix F.1). A retained carrier is a record in R that is correlated with the deleted subject and that, by staying in the pool, helps or hurts the readout's reconstruction of that subject.

2.2 Two channels

Fix a record r, a carrier set C_r that excludes r, the map ψ, and the deterministic ridge rule; write R for the pool without r. In one version the maintainer trains the readout with r and then unlearns r exactly; in the other, r never joined. Both versions keep every retained record.

Proposition 1 (audit invariance to retained records) *For every retained unit $u \in R$ and every membership test T of the deployed (ψ, w_R, R), T accepts u with the same probability in both versions; the record-level membership advantage at u is zero.*

Proof: Appendix A; the equality is definitional. A membership test therefore measures one channel; what a retained carrier reveals about the deleted subject is a second, with its own quantity to estimate (Section 6.2 reports both).

2.3 Estimand and request unit

The analyst detects the carrier set $C \subseteq R$ only after the maintainer honors G. We measure the second channel by refitting the readout without the carriers. For an evaluation window j and the readout fit on pool Q, write $\mathrm{err}_j(Q) = a_j^\top w_Q - y_j$. The per-window change $\delta_j = \mathrm{err}_j(R \setminus C)^2 - \mathrm{err}_j(R)^2$ is the carrier's contribution of Section 1; we call it the reconstruction excess and write Δ for its expectation over label noise. Removing the carriers leaves the rest of the pool unchanged, so the excess comes from the carriers alone. A subject that the rest of the pool already predicts well has an excess of zero, even if the subject itself is easy to reconstruct. The request unit is the subject: a bearing for CWRU and Paderborn, a patient for PTB-XL, a machine for MIMII. The pre-registered primary quantity, the all-subject estimand, is the median excess over all deleted subjects, with a subject that has no detected carrier counted as zero. The secondary quantity, the carrier-conditioned estimand, is the same median taken over the subjects that do have a detected carrier. The audit is therefore invariant to every retained record (Proposition 1); the next section gives the quantity it leaves unmeasured a closed form.

3 Exact accounting for a fixed-feature readout

3.1 The rank-k identity

Conditioning on a fixed feature map ψ and a ridge or kernel-ridge readout with penalty $\lambda > 0$, the change in the deleted subject's reconstruction error caused by a stated set of carriers has an exact closed form; this section derives it and states what it does not decide (Contribution 2). Let Φ_R be the carrier-free retained design, $A = \Phi_R^\top \Phi_R + \lambda I$, $b = \Phi_R^\top y_R$, and $w_b = A^{-1} b$. Let $a_X = \psi(X)$ be the deleted subject's feature direction with clean value μ_X, and write $p_b = a_X^\top w_b$ and $e_b = p_b - \mu_X$. Suppose k retained carriers have feature columns U and labels $y_c = c_0 + \varepsilon$ with mean-zero noise of covariance $\sigma^2 I$, and define $M = U^\top A^{-1} U$, $g = U^\top A^{-1} a_X$, and $q = U^\top A^{-1} b$.

Theorem 1 (exact fixed-feature identity) *The prediction after retaining the k carriers is $a_X^\top w_k = p_b + g^\top (I + M)^{-1}(y_c - q)$. With $m_k = e_b + g^\top (I + M)^{-1}(c_0 - q)$, the carrier-induced reduction in squared reconstruction error is Equation (1):*

$$\Delta_k = e_b^2 - m_k^2 - \sigma^2 \, g^\top (I + M)^{-2} g. \tag{1}$$

The prediction identity says that a retained record helps only through its residual $y_c - q$, the part of its label that the carrier-free readout does not already predict. In Equation (1) the last term, $-\sigma^2 \, g^\top (I + M)^{-2} g$, is never positive: carrier labels that are noisy relative to the baseline error hurt

reconstruction, so on a natural pool (real retained records rather than injected carriers) the excess can be zero or negative. The prediction identity holds exactly, with no asymptotic or distributional assumption; only the noise-averaged form of Δ_k uses the stated second moment. Appendix A gives the proof, the kernel-ridge form, and each assumption's status. In the experiments the identity also checks the implementation and supplies the predicted floor that admits a fine-tuning pair (Appendix C.2; Section 6.4).

3.2 THE DUPLICATE ENDPOINT

For $k = 1$ and an exact duplicate, $a' = a_X$ and $c_0 = \mu_X$, write $h = a_X^\top A^{-1} a_X$ for the leverage and $B = \mathbb{E}[e_b^2]$ for the baseline second moment. Then $\Delta_1 = h[(2+h)B - h\sigma^2]/(1+h)^2$, which gives Equation (2):

$$\Delta_1 > 0 \iff (2+h)\, B > h\, \sigma^2. \tag{2}$$

Equation (2) states when an exact duplicate leaks and is sharp at the exact-duplicate endpoint only, where excess and residual influence trade one for one and carrier noise caps the recovery ratio (Appendices A.7, A.8).

3.3 WHAT THE IDENTITY DOES NOT DECIDE

Away from the endpoint, with feature-space correlation $\rho < 1$ between carrier and deleted subject, Equation (2) is a screen. A screen predicts only the sign of the excess; the identity gives its exact value. Given the carrier set, a verdict uses Δ_k from Equation (1) with measured carrier labels and an abstention band, a margin around zero inside which no verdict is returned. Scoring windows must not share noise with retained-pool windows, or the realized excess drifts from the mean-scored one (Appendix A.4). For carriers other than exact duplicates, the identity gives no sharp rule for when they leak; that is what a screen is for.

The random-design analysis of Appendix C supplies a second screen, the deterministic-equivalent sign screen of Section 6.5, whose agreement with measured signs is consistency evidence, not a population guarantee. The identity attributes the change exactly to a stated carrier set, so the remaining task is to find that set on real data, which is where the standard detector fails (Section 4).

4 FINDING CARRIERS: THE RAW-COSINE ARTIFACT AND DE-CENTERING

The standard detector for the carriers whose contribution Section 3 attributes, cosine similarity on frozen features, fabricates the channel it is meant to find: frozen time-series features share one dominant direction, so it flags almost every retained record, and de-centering removes the artifact (Contribution 3).

4.1 THE ARTIFACT

The standard detector, the raw cosine detector, counts a retained unit as a carrier when the cosine between its mean feature and the deleted subject's mean feature passes a threshold, 0.9 here. Frozen time-series features, however, share one dominant direction. The raw cosine then measures this shared feature direction, common to all units, instead of the carrier. When the shared feature direction dominates, the raw cosine of any two units rises toward one, whether or not they share a source (Appendix I). The detector therefore saturates: on Chronos-Bolt and TimesFM-2.5 features over CWRU, Paderborn, PTB-XL, and the auxiliary ELD electricity-load set, it flags 92 to 100 percent of retained series in every cross-family cell, one encoder family on one corpus. With Chronos-t5-tiny features, another family, the fraction is 88 to 100 percent, so the saturation covers all three encoder families. The saturated detector inflates the measurement: flagging 48 to 52 of 59 series removes most of the pool, and the under-determined refit produces an apparent excess.

In 4 of 8 cross-family cells that excess passes the material-effect threshold of 0.1, on Paderborn by a wide margin (Figure 2).

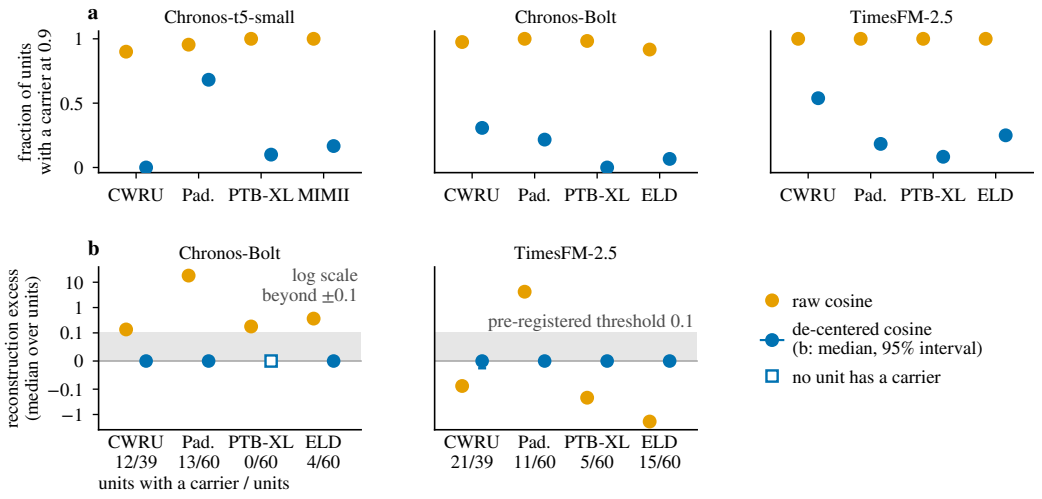


Figure 2: Raw cosine flags most retained units as carriers and reads apparent excess up to +18.27; de-centering removes both. (a) Fraction of deletion units with a retained unit above cosine 0.9, one panel per encoder family, raw against de-centered. (b) Median reconstruction excess of the cross-family runs in predictability units, raw against de-centered with 95% intervals; the value axis is linear between -0.1 and 0.1 and logarithmic beyond, so every raw value is drawn at its value; grey band: below the pre-registered threshold 0.1.

4.2 De-centering and the corrected protocol

De-centering subtracts the retained-pool mean feature from every unit before computing the cosine; the de-centered detector applies the same threshold. With the shared direction removed, the flagged fraction falls to 0 to 55 percent, and in all eight cross-family cells we obtain a median of 0.000 with an upper bound at the same value.

The corrected protocol also fixes what the analyst sees and how the excess is scored and reported. The evaluator builds the signed descriptor from the deleted subject's windows and sends only this descriptor to the analyst. The analyst flags every retained unit whose de-centered cosine with the descriptor reaches the threshold. These units form the detected carrier set C, and the analyst refits the readout without C.

Scoring compares both readouts on the deleted subject. The evaluator scores each readout by its predictability: how much better it predicts the deleted subject than it predicts held-out windows of retained units outside C. The reconstruction excess is the change in predictability when C leaves the pool. This unit is the squared-error excess of Section 2.3 expressed relative to the generic loss (Appendix F.7). We report the median over subjects of the three-seed mean, a bootstrap interval (95 percent, 10,000 resamples), and a matched random control that removes as many retained units as C contains. Under the pre-registered rule, a cell shows a material effect when the median exceeds the material-effect threshold and the interval excludes zero. It supports the null when the upper bound stays below the threshold, and is undetermined otherwise.

Two conventions apply, one to zeros and one to windows. A subject with no detected carrier has an excess of exactly zero, because no refit takes place. We count such zeros apart from a measured zero, where a detected carrier's excess is measured at zero. The descriptor and the scoring share the deleted subject's windows. Rebuilding the descriptor from held-out windows alone moved no headline median beyond the 0.012 bootstrap resolution (Appendix F.5). With de-centering the detector no longer flags carriers through the shared direction, so the next question is how small a true contribution the corrected measurement can register (Section 5).

5 CALIBRATION: WHAT THE CORRECTED MEASUREMENT CAN RESOLVE

Planted carriers with known effects fix how small a contribution the corrected measurement could have seen, and that detection limit is what the zeros of Section 6 inherit (Contribution 3).

5.1 INJECTED-EFFECT LADDER

Controlled-carrier calibration plants a retained unit whose effect we know or construct and measures, for that unit, the detector's response and the excess. The injected-effect ladder mixes a fraction φ of the deleted subject's windows into one retained unit and forces that unit to count as a carrier, because the detector flags it only at large φ. Along the ladder the estimator resolves an excess of 0.02 to 0.04 at φ between 0.35 and 0.5; we call this the estimator's resolution once a carrier is flagged. Appendix B.3 reports the detection limit along the same ladder.

5.2 POSITIVE CONTROLS

Each positive control is a pre-registered planted carrier that the corrected measurement should register (Table 2, Appendix G). Its criterion requires a flag from the detector in at least half of the subject-by-seed runs and a material excess (Section 4.2). None meets the criterion; the controls fall short in different ways. The later-window control keeps part of the deleted subject's own data in the pool: its later windows, bundled into one retained unit. The clean-signal control plants a retained unit that re-observes the deleted subject's clean signal under fresh noise at 6 dB, which fixes its excess at +0.63. Figures A5 to A7 (Appendix H) show how often the detector flagged each planted carrier and the excess that carrier then produced: the later-window unit counts as detected only on CWRU and Paderborn, the clean-signal unit almost never does, and no reading is material (Appendix B.4).

The forced reading on the clean-signal control recovers about one ninth of its known effect. Two further units, a new realization of the deleted subject's autoregressive dynamics and a unit carrying only its fitted spectrum in the real CWRU pool, separate the detector's response from the excess. Five positive-control readings (the later-window control on CWRU, MIMII, Paderborn, and PTB-XL, and the clean-signal control) fell short of the criteria fixed in advance, and the next section measures the natural pools at this resolution.

6 RESULTS ON NATURAL CORPORA

On the four natural corpora the corrected measurement reads zero on three and small on the fourth, membership inference stays accurate on the same pools, and fine-tuning and record-level deletion mark where the measurement stops (Contributions 1 and 3). The primary quantity, pre-registered and unchanged after the runs, is the median excess over all subjects; the secondary quantity is the median over subjects with a detected carrier. The record-level arm of Section 6.5 was added after pre-registration and changes no subject-level number.

6.1 THE PRE-REGISTERED MEDIAN

The pre-registered all-subject median is an exact zero on three of the four corpora and small on Paderborn, and the zeros record how rare detected carriers are rather than how small their effect is. This is the reconstruction channel; the membership channel follows on the same pools. Each request deletes one subject of CWRU (Smith & Randall, 2015), Paderborn (Lessmeier et al., 2016), PTB-XL (Wagner et al., 2020), or MIMII (Purohit et al., 2019); the frozen encoder is Chronos-t5-small (Ansari et al., 2024a) unless stated (Figure 3). The all-subject median is 0.000 on CWRU, PTB-XL, and MIMII, because the de-centered detector flags a carrier for 0 of 10 CWRU bearings, 4 of 40 PTB-XL patients, and 2 of 12 MIMII machines. A subject with no detected carrier has an excess of exactly zero, because no refit takes place. On these three corpora most subjects have no detected carrier. The four PTB-XL carrier patients range from -0.02 to +0.06; the two MIMII carrier machines lie above the threshold at negative predictability, where a loss ratio above one amplifies the excess (Appendix F.7). On Paderborn, where most bearings have a carrier, we measure +0.056 [+0.018, +0.087] on the original 22 bearings and +0.046 [0.000, +0.066] on all 26; the 15 and 16

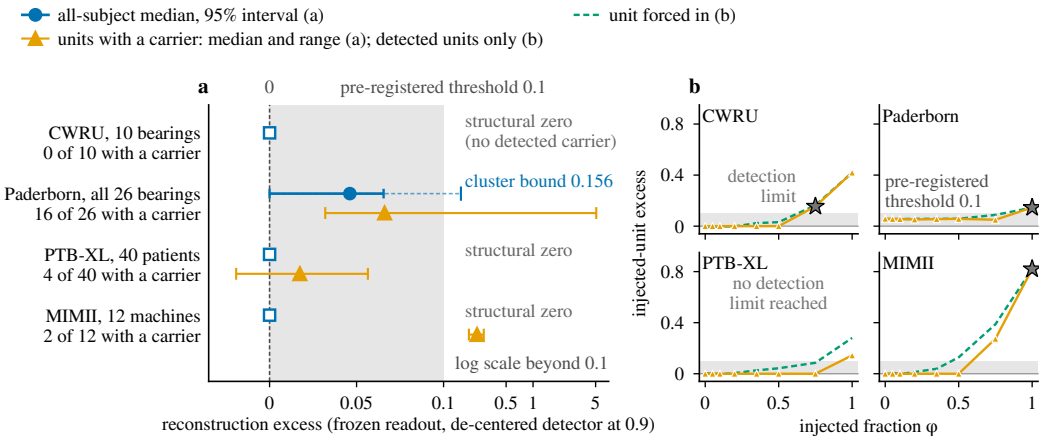


Figure 3: After de-centering, the pre-registered all-subject median excess reads 0.000 on CWRU, PTB-XL and MIMII and +0.046 on Paderborn, while a planted unit forced into the fit resolves 0.02 to 0.04 at φ between 0.35 and 0.5. (a) Chronos-t5-small, frozen readout; value axis linear to 0.1, logarithmic beyond. Blue: all-subject median, 95% interval, Paderborn with the cluster-bootstrap bound; hollow squares: structural zeros (fewer than half of the subjects have a carrier, so median and interval are zero by construction; no interval drawn). Orange: median and range over the units with a carrier. (b) Injected-effect ladder; star: detection limit, φ 0.75 (CWRU), 1 (Paderborn, MIMII), none (PTB-XL).

bearings with a detected carrier give +0.084 and +0.066. Under the rule of Section 4.2 both cells fall below the material-effect threshold; the 22-bearing subject-level interval excludes zero, and the cluster-bootstrap intervals reach it (Appendix F.3). The matched random control carries a basis bias that lifts the net reading, so we use the control-free value (Appendix F.5). These readings cover the carriers the detector flags; a membership test on the same pools comes next.

6.2 The membership channel stays open

Membership inference stays accurate on the very pools where the reconstruction excess reads zero, consistent with the record-level audit and the carrier measurement answering different questions (Contribution 1). This is the membership channel. On the same pools, encoder, and deletion units we run subject-level LiRA (Carlini et al., 2022) and RMIA (Zarifzadeh et al., 2024) with sixteen shadow readouts (Figure A1a). The stronger attack on each corpus, scoring subjects by their mean window score, reaches an area under the ROC curve (AUC) of 0.98 to 1.00 on CWRU, 0.93 to 0.95 on Paderborn, 0.98 on PTB-XL, and 0.99 to 1.00 on MIMII. LiRA alone reaches 0.71 to 0.77 on Paderborn and 0.84 to 0.89 on PTB-XL; every lower bound stays above 0.5 in all 16 corpus-attack-unit combinations. This is consistent with the two channels being different estimands (Proposition 1). These attacks bound no stronger one; next we vary the protocol.

6.3 Robustness of the zero

We test whether the exact zero survives a change of readout, encoder size, pool size, repeated deletion, or feature family. A kernel-ridge readout, four Chronos-t5 sizes, a fivefold larger PTB-XL pool, and five sequential deletions leave the three zero corpora unchanged (Table 1, Appendix G). Under the same variants Paderborn stays positive, and its upper bound reaches past the threshold (Appendix F.3). The larger PTB-XL pool lifts the fraction of patients with a carrier only to 0.135 to 0.215, so most patients still have none and the all-patient median stays zero; carrier patients range from +0.001 to +0.007. At 1000 patients the median is still 0.000 [0.000, 0.000], and six patients sit above the material-effect threshold (Appendix B.2). The headline readings come from Chronos-t5-small features, so we repeat the measurement on 504 non-learned MiniRocket features (Dempster et al., 2021). There raw cosine flags every retained unit as a carrier of every other on all four corpora, and de-centering removes this saturation. After de-centering the subject-level zero replicates

on CWRU, PTB-XL, and MIMII, and the Paderborn median is +0.102 [0.000, +0.139], larger than on Chronos and undetermined. On Chronos-Bolt features the three zeros repeat and the Paderborn median is +0.03 [+0.00, +0.09]. The three zeros therefore do not depend on the readout, the encoder family, or the pool size tested here. All of these variants keep the encoder frozen; the next subsection fine-tunes it.

6.4 Full fine-tuning is a boundary

We now measure what fine-tuning the encoder after the deletion does to the excess. For 72 record-level requests we delete one record, inject one duplicate of it with label noise, and fine-tune the encoder on the retained pool (Figure A2). We then compare each request with the same request under the frozen readout, where the duplicate already produces a positive excess, so the paired change measures what fine-tuning adds to or takes from that excess. A pre-registered rule admits a pair only when the identity of Section 3, evaluated on the fine-tuned encoder, still predicts a positive excess for the duplicate; this prediction is the predicted floor. Table 3 lists the change over the admissible pairs, over all pairs, and for CWRU and PTB-XL; the full-set, record-resampled, cluster, and retraining intervals are in Appendix F.4. Pooled over the admissible pairs, the change stays below the threshold, and the excluded pairs change by at most zero. The per-corpus changes differ in sign, and the CWRU gain excludes zero at the pair level only, since the record-clustered interval touches zero. The direction of the change thus depends on the corpus, and the identity assumes fixed features, so it no longer applies once the encoder moves. Fine-tuning is therefore a boundary of the measurement, not a reading inside it, and the next subsection turns to the other boundary, the deletion unit.

6.5 Record-level deletion and the sign screen

The subject-level measurement deletes whole subjects; we now delete one record and keep the other records of the same subject in the pool (Figure A3). Those retained records are carriers within one subject, and each request is labeled LEAK or SAFE from its own retained windows. The sign screen of Section 3.3 then predicts each request's label without seeing those windows, and we count how often the prediction agrees with the label. The labels follow two rules: the original rule reads each request's label from its own retained windows, and window-clustered relabeling first merges the overlapping windows of one recording and leaves far fewer leaks (Table 3). On most of these requests the subject-level control is an exact zero, because deleting the whole subject removes the sibling records as well. Under the relabeling the screen recovers every positive, three recordings of one healthy bearing; under the original rule it recovers only part of the positives. Table 3 gives its sensitivity, specificity, and balanced accuracy with the leak counts. Within-subject carriers therefore lie outside the subject-level estimand, and this record-level test is boundary evidence, not a reading of the subject-level quantity.

On an auxiliary request set the screen's flags coincide with a ranking by baseline error, consistency evidence, not a population guarantee (Table 3). Fine-tuning and record-level deletion mark the boundaries of the corrected measurement; Section 7 places it among neighboring work and Section 8 states where each claim stops.

7 Related work

Record-level unlearning audits test membership (Bourtoule et al., 2021; Carlini et al., 2022; Hayes et al., 2025), and recent work also examines retained records (Ye et al., 2025; Fu et al., 2026; Cheng et al., 2026). Reconstruction attacks, interdependent-privacy studies, and attribute inference expose a person whose own records are gone (Bertran et al., 2024; Wu et al., 2025; Humbert et al., 2013; 2017; Chourasia & Shah, 2023; Jayaraman & Evans, 2022). Time-series foundation models serve as frozen encoders whose ridge readouts admit closed-form deletion (Ansari et al., 2024a; Das et al., 2024; Guo et al., 2020; Izzo et al., 2021; Quan et al., 2026). Cosine similarity detects near duplicates in embeddings, and their anisotropy has known corrections (Abbas et al., 2023; Mu & Viswanath, 2018; Ethayarajh, 2019; Su et al., 2021; Timkey & van Schijndel, 2021). Appendix D grades the overlap with each neighboring paper and gives the full discussion. None of them measures what

a retained correlated record adds to the reconstruction of a deleted subject after a lawful deletion, which is the quantity Sections 3 to 6 define, correct, calibrate, and report.

8 SCOPE AND LIMITATIONS

Each claim of this paper stops at a stated boundary: the identity at a fixed feature map and a ridge or kernel-ridge readout, the measurement at its detection limit and one deployment domain, and the normative reading at a single estimated quantity.

Formal scope. The formal results are exact only where their premises hold. Theorem 1 and its duplicate endpoint hold for a fixed feature map and a ridge or kernel-ridge readout; a learned reconstructor lies outside that class, so they do not bound it. Population inference, missed carriers, and plug-in uncertainty lie outside them; the random-design analysis decides no verdict (Appendix C). Proposition 1 holds for exact unlearning; under approximate unlearning its zero advantage becomes a bound (Appendix A, Corollary 1′b).

Measurement scope. The measurement stops at its detection limit and covers few settings. The estimator's resolution and its forced reading on the clean-signal control are stated in Section 5. The detection limit along the injected-effect ladder is in Appendix B.3. The five positive-control readings test two carrier families at one descriptor and readout interface. The headline readings come from one encoder; the robustness checks cover MiniRocket and Chronos-Bolt features and a larger PTB-XL pool. The natural-corpus evidence comes from one deployment domain; every check keeps the same threshold and windows. The record-level arm takes its carriers from declared sibling records, not from the detector (Section 6). Paderborn's cluster-level intervals rest on three damage-family clusters, and the extension to 32 bearings reports a 256-point cluster resampling alongside the subject interval (Appendix F.3). The fine-tuning boundary rests on injected duplicates and Paderborn's natural carriers.

Normative scope and open channels. A positive excess estimates one quantity: it establishes no legal noncompliance, shows no record-deletion failure, and authorizes no removal of another subject's lawfully retained data. For a frozen readout and a detected carrier set, the analyst can remove the carriers, refit, and certify a one-coordinate budget ε (Appendix E). Outside that one-coordinate certificate, however, no recorded quantity says how much of the deleted subject a fine-tuned encoder absorbed, and a learned adversary's channel we measure but do not price: we put no bound on it (Appendices E.4, F.6). Carriers below our detection threshold stay an open channel, priced only at lower thresholds. Table 2 lists every channel with its status (Appendix G). Within these limits the identity is exact, the primary quantity pre-registered, and the resolution stated, so Section 9 can ask evaluations to report the measurement beside the record-level audit.

9 IMPLICATIONS AND CONCLUSION

This paper makes three points. First, a record-level unlearning audit is blind by construction to reconstruction from retained correlated records. It compares two models that both keep those records, so whatever they reveal about the deleted subject, the audit's verdict does not change (Proposition 1). Second, the standard way of finding such records, cosine similarity on frozen time-series features, flags almost every retained record and thereby inflates the measured contribution; subtracting the shared feature direction removes this artifact. Third, after that correction the all-subject median contribution is 0.000 on three of four corpora and small on Paderborn, while membership inference on the same pools stays accurate; the planted carriers show that these zeros sit at the measurement's detection limit. Two consequences follow. Audits should de-center features before searching for carriers, and they should report the membership channel and the reconstruction channel side by side, each with its measurement range. Box 1 (Appendix H) lists what a post-only analyst should report.

AI USE STATEMENT

LLM-based coding and writing agents, operated under the author's direction, were used extensively: to draft and refine hypotheses and pre-registration documents, design experiments, implement all analysis code and the synthetic-data generators, clean and analyze data, search and screen related work, interpret results, translate working notes, draft and revise the text, and run internal adversarial reviews. They also helped develop the fixed-feature carrier framework, formulate the mathematical claims, and write the proofs of Theorem 1; every algebraic claim carries an executable numerical check (Appendix A). The author set the research questions, approved every pre-registered criterion before any run, provisioned the GPU compute and authorized every run, and verified every reported number against a per-number ledger backed by time-stamped pre-registration files and independent re-computations. No AI system is an author. We have reviewed all AI-assisted work and take full responsibility for the final content, including text, claims, code, and artifacts.

ETHICS STATEMENT

All data come from public benchmark corpora (CWRU, Paderborn, PTB-XL, MIMII, ELD, traffic) under their license terms; no data were collected from people and no human-subjects study was conducted. PTB-XL is de-identified; the patient is the subject unit, we report medians, intervals, and counts over patients, and identify no recording or patient; the industrial corpora hold no personal data. A positive carrier reading is a measurement about an estimand: it establishes no legal noncompliance, shows no failure of record deletion, and authorizes no removal of another subject's lawfully retained data (Section 8). The released code implements the carrier detector, the excess measurement, and the membership-inference baselines; no per-record leak list is released. The author declares no conflict of interest.

REPRODUCIBILITY STATEMENT

Appendix A gives proofs of Theorem 1, its endpoint, the shared-noise convention, and the kernel-ridge form (Appendix A.5), with the executable check in the supplementary archive and the assumption list of Appendix A.6. Section 4.2 states the corrected protocol and its pre-registered decision rule; Appendix F lists the protocol constants (feature model, ridge penalty, context length, threshold, seeds, bootstrap size) and every boundary-ladder cell; Appendix B.3 gives the injected-effect ladder and Appendix B.2 the threshold scan. All corpora are public (Section 6.1). The supplementary archive contains the measurement, detection, scoring, and audit code, the pre-registration documents with their time stamps, the number ledger, the archived result files behind every reported number, and the scripts that regenerate Tables 1 and 2, Figures 1 to 3, and Figures A1 to A7.

REFERENCES

Amro Abbas, Kushal Tirumala, Dániel Simig, Surya Ganguli, and Ari S. Morcos. SemDeDup: Data-efficient learning at web-scale through semantic deduplication. arXiv preprint arXiv:2303.09540, 2023. URL https://arxiv.org/abs/2303.09540.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, et al. Chronos: Learning the language of time series. *Transactions on Machine Learning Research (TMLR)*, 2024a. arXiv:2403.07815.

Abdul Fatir Ansari, Caner Turkmen, Oleksandr Shchur, Lorenzo Stella, et al. Chronos-bolt: Fast and accurate pretrained time-series forecasting models. Model release, https://huggingface.co/amazon/chronos-bolt-small (AutoGluon-TimeSeries 1.2), 2024b.

Martin Bertran, Shuai Tang, Michael Kearns, Jamie Morgenstern, Aaron Roth, and Zhiwei Steven Wu. Reconstruction attacks on machine unlearning: Simple models are vulnerable. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2405.20272.

Lucas Bourtoule, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine unlearning. In *IEEE Symposium on Security and Privacy (S&P)*, pp. 141–159, 2021.

Yinzhi Cao and Junfeng Yang. Towards making systems forget with machine unlearning. In *IEEE Symposium on Security and Privacy (S&P)*, pp. 463–480, 2015.

Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramèr. Membership inference attacks from first principles. In *IEEE Symposium on Security and Privacy (S&P)*, pp. 1897–1914, 2022.

Min Chen, Zhikun Zhang, Tianhao Wang, Michael Backes, Mathias Humbert, and Yang Zhang. When machine unlearning jeopardizes privacy. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 896–911, 2021.

Jingpu Cheng, Ping Liu, Qianxiao Li, and Chi Zhang. Machine unlearning under retain–forget entanglement. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=4WMBSHHJEr. arXiv:2603.26569.

Rishav Chourasia and Neil Shah. Forget unlearning: Towards true data-deletion in machine learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6028–6073. PMLR, 2023. URL https://proceedings.mlr.press/v202/chourasia23a.html.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *International Conference on Machine Learning (ICML)*, pp. 10148–10167, 2024.

Angus Dempster, François Petitjean, and Geoffrey I. Webb. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, 2020.

Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. MiniRocket: A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 248–257, 2021.

Polina Dolgova and Sebastian U. Stich. Forgetting has neighbors: Localized collateral forgetting in machine unlearning. arXiv preprint arXiv:2605.31317, 2026.

Kawin Ethayarajh. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 55–65. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1006. URL https://aclanthology.org/D19-1006/.

Jie Fu, Nima Naderloui, Da Zhong, Yuan Hong, and Wendy Hui Wang. Revisiting privacy leakage in machine unlearning: Membership inference beyond the forgotten set. In *2026 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2026. URL https://arxiv.org/abs/2605.01129. arXiv:2605.01129.

Ji Gao, Sanjam Garg, Mohammad Mahmoody, and Prashant Nalini Vasudevan. Deletion inference, reconstruction, and compliance in machine (un)learning. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2022(3):415–436, 2022.

Antonio Ginart, Melody Guan, Gregory Valiant, and James Zou. Making AI forget you: Data deletion in machine learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *International Conference on Machine Learning (ICML)*, pp. 16115–16152, 2024.

Chuan Guo, Tom Goldstein, Awni Hannun, and Laurens van der Maaten. Certified data removal from machine learning models. In *International Conference on Machine Learning (ICML)*, pp. 3832–3842, 2020.

Jamie Hayes, Ilia Shumailov, Eleni Triantafillou, Amr Khalifa, and Nicolas Papernot. Inexact unlearning needs more careful evaluations to avoid a false sense of privacy. In *IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 497–519, 2025.

Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. Addressing the concerns of the Lacks family: Quantification of kin genomic privacy. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2013.

Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. Quantifying interdependent risks in genomic privacy. *ACM Transactions on Privacy and Security*, 20(1):1–31, 2017. doi: 10.1145/3035538. URL https://doi.org/10.1145/3035538.

Mathias Humbert, Benjamin Trubert, and Kévin Huguenin. A survey on interdependent privacy. *ACM Computing Surveys*, 52(6):122:1–122:40, 2020.

Zachary Izzo, Mary Anne Smart, Kamalika Chaudhuri, and James Zou. Approximate data deletion from machine learning models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

Bargav Jayaraman and David Evans. Are attribute inference attacks just imputation? In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1569–1582. ACM, 2022. doi: 10.1145/3548606.3560663. URL https://doi.org/10.1145/3548606.3560663.

Nicolas Johansson, Tobias Olsson, Daniel Nilsson, Johan Östman, and Fazeleh Hoseini. Privacy risks in time series forecasting: User- and record-level membership inference. arXiv preprint arXiv:2509.04169, 2025. URL https://arxiv.org/abs/2509.04169.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017. doi: 10.1109/TIT.2017.2685505.

Nikhil Kandpal, Eric Wallace, and Colin Raffel. Deduplicating training data mitigates privacy risks in language models. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022.

Noam Koren, Abigail Goldsteen, Guy Amit, and Ariel Farkash. Membership inference attacks against time-series models. In *Asian Conference on Machine Learning (ACML)*, volume 260 of *Proceedings of Machine Learning Research*, pp. 319–334, 2025.

Hithem Lamri, Manaar Alam, Haiyan Jiang, and Michail Maniatakos. DRAUN: An optimization-agnostic data reconstruction attack on federated unlearning. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 4906–4914. PMLR, 2026. URL https://proceedings.mlr.press/v300/lamri26a.html.

Katherine Lee, Daphne Ippolito, Andrew Nystrom, Chiyuan Zhang, Douglas Eck, Chris Callison-Burch, and Nicholas Carlini. Deduplicating training data makes language models better. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022.

Christian Lessmeier, James Kuria Kimotho, Detmar Zimmer, and Walter Sextro. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *European Conference of the PHM Society*, volume 3, 2016.

Hongkai Li, Shifeng Xie, Lefei Shen, Zhuo Li, Mouxiang Chen, Xiaobin Zhang, Han Fu, Jianling Sun, Xiaoxue Ren, and Chenghao Liu. TSFMAudit: Data contamination auditing in forecasting time series foundation models. arXiv preprint arXiv:2605.26161, 2026a. URL https://arxiv.org/abs/2605.26161.

Qingwen Li, Xiao Han, Ruiyan Wang, Junjie Wu, and Lanjuan Liu. Membership inference attack against time-series prediction models. *IEEE Transactions on Dependable and Secure Computing*, 23(2):3464–3476, 2026b. doi: 10.1109/TDSC.2025.3637371. URL https://doi.org/10.1109/TDSC.2025.3637371.

Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. Dependence makes you vulnerable: Differential privacy under dependent tuples. In *Network and Distributed System Security Symposium (NDSS)*, 2016.

Marcel Meyer, Sascha Kaltenpoth, Kevin Zalipski, and Oliver Müller. Rethinking evaluation in the era of time series foundation models: (un)known information leakage challenges. arXiv preprint arXiv:2510.13654, 2025.

Jiaqi Mu and Pramod Viswanath. All-but-the-top: Simple and effective postprocessing for word representations. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=HkuGJ3kCb. arXiv:1702.01417.

Seth Neel, Aaron Roth, and Saeed Sharifi-Malvajerdi. Descent-to-delete: Gradient-based methods for machine unlearning. In *Algorithmic Learning Theory (ALT)*, pp. 931–962, 2021.

V. S. Raghu Parupudi. Anisotropy decides cosine vs. rank metrics for text embeddings. arXiv preprint arXiv:2606.29571, 2026. URL https://arxiv.org/abs/2606.29571.

Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. In *Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2019. arXiv:1909.09347.

Yijun Quan, Wentai Wu, and Giovanni Montana. Exact federated continual unlearning for ridge heads on frozen foundation models. arXiv preprint arXiv:2603.12977, 2026. URL https://arxiv.org/abs/2603.12977.

Ahmed Salem, Apratim Bhattacharyya, Michael Backes, Mario Fritz, and Yang Zhang. Updates-Leak: Data set inference and reconstruction attacks in online learning. In *29th USENIX Security Symposium (USENIX Security 20)*, pp. 1291–1308, 2020.

Ayush Sekhari, Jayadev Acharya, Gautam Kamath, and Ananda Theertha Suresh. Remember what you want to forget: Algorithms for machine unlearning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 18075–18086, 2021.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy (S&P)*, pp. 3–18, 2017.

Wade A. Smith and Robert B. Randall. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*, 64–65: 100–131, 2015.

Jianlin Su, Jiarun Cao, Weijie Liu, and Yangyiwen Ou. Whitening sentence representations for better semantics and faster retrieval. arXiv preprint arXiv:2103.15316, 2021. URL https://arxiv.org/abs/2103.15316.

Faiz Taleb, Ivan Gazeau, and Maryline Laurent. Uncovering memorization in timeseries imputation models: LBRM membership inference and its link to attribute leakage. arXiv preprint arXiv:2603.24213, 2026.

Jiashu Tao and Reza Shokri. Range membership inference attacks. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 346–361. IEEE, 2025. doi: 10.1109/SaTML64287.2025.00026. URL https://doi.org/10.1109/SaTML64287.2025.00026.

Anvith Thudi, Hengrui Jia, Ilia Shumailov, and Nicolas Papernot. On the necessity of auditable algorithmic definitions for machine unlearning. In *31st USENIX Security Symposium*, 2022.

William Timkey and Marten van Schijndel. All bark and no bite: Rogue dimensions in transformer language models obscure representational quality. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 4527–4546. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.emnlp-main.372. URL https://aclanthology.org/2021.emnlp-main.372/.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7:154, 2020.

Ziyi Wang, Lixing Chen, Zhongqi Miao, Junhua Tang, Yang Bai, and Jianhua Li. TS-Unlearn: A dual-objective unlearning framework for time series forecasting. In *Advances in Knowledge Discovery and Data Mining (PAKDD), Part II*, volume 16598 of *Lecture Notes in Computer Science*, pp. 341–353. Springer, 2026.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *International Conference on Machine Learning (ICML)*, 2024.

Xiaoyu Wu, Yifei Pang, Terrance Liu, and Zhiwei Steven Wu. Unlearned but not forgotten: Data extraction after exact unlearning in LLM. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Shizhou Xu and Thomas Strohmer. Machine unlearning via information theoretic regularization. arXiv preprint arXiv:2502.05684, 2025. URL https://arxiv.org/abs/2502.05684.

Dayong Ye, Tianqing Zhu, Jiayang Li, Kun Gao, Bo Liu, Leo Yu Zhang, Wanlei Zhou, and Yang Zhang. Data duplication: A novel multi-purpose attack paradigm in machine unlearning. In *34th USENIX Security Symposium (USENIX Security 25)*, pp. 6399–6418, 2025.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Santiago Zanella-Béguelin, Lukas Wutschitz, Shruti Tople, Victor Rühle, Andrew Paverd, Olga Ohrimenko, Boris Köpf, and Marc Brockschmidt. Analyzing information leakage of updates to natural language models. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 363–375, 2020.

Sajjad Zarifzadeh, Philippe Liu, and Reza Shokri. Low-cost high-power membership inference attacks. In *International Conference on Machine Learning (ICML)*, pp. 58244–58282, 2024.

Binchi Zhang, Zihan Chen, Cong Shen, and Jundong Li. Verification of machine unlearning is fragile. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 58717–58738. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhang24h.html.

Benjamin Zi Hao Zhao, Aviral Agrawal, Catisha Coburn, Hassan Jameel Asghar, Raghav Bhaskar, Mohamed Ali Kaafar, Darren Webb, and Peter Dickinson. On the (in)feasibility of attribute inference attacks on machine learning models. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 232–251. IEEE, 2021. doi: 10.1109/EuroSP51992.2021.00025. URL https://doi.org/10.1109/EuroSP51992.2021.00025.

A Proofs and conditioning conventions

This appendix supplies the proofs referred to in Sections 2.2 and 3. All quantities below are conditional on the carrier-free design (Φ_R, y_R), the deleted subject's feature direction a_X, and selected carrier features U. The expectation is only over carrier-label noise unless an additional average over the deleted subject is stated. A.1–A.4 prove Theorem 1, its endpoint, and the shared-noise convention; A.5 derives the kernel-ridge form; A.6 tabulates every assumption with its status in the experiments; A.7 relates reconstruction to influence at the duplicate endpoint; A.8 bounds the noisy recovery ratio; A.9 explains the shared-process control post hoc.

Proof of Proposition 1. Both versions run the same deterministic ridge rule on the same frozen map ψ. In the first, exact unlearning of r returns the readout that the rule produces from R; in the second the rule runs on R directly, so both versions deploy the readout w_R. The retained pool R is the same object in both versions, hence every retained unit $u \in R$, and every carrier set $C \subseteq R$, is a member in both. A membership test T of the deployed (ψ, w_R, R) therefore receives identical inputs in the two versions, and its acceptance probability at u, over its own randomness, is the same in both; the difference is zero. □

Remark. The equality is definitional and says nothing about the strength of membership estimators; the reconstruction excess of Section 2.3 is a different object on the same pool: the two-record construction of Appendix B.6 has record-level excess $480/121$ against a subject-level excess of exactly 0, and Section 6.5 measures the record-level reading on natural pools.

Proposition $1'$ (audit invariance to retained records, as a bound). Let $V_i = (\psi, w^{(i)}, R)$ be the view deployed in version $i \in \{0, 1\}$: $w^{(1)}$ is the readout obtained by training with r and then unlearning r exactly, and $w^{(0)}$ is the readout trained on R alone. Let $s_i = f(V_i, X)$ be any side information computed by one measurable map f from the deployed view and the deleted subject, the evaluator's descriptor z_X included, and write P_i for the law of (V_i, s_i). For a membership test T that reads (V_i, s_i) and its own randomness, write $\alpha_i(T, u)$ for its probability of accepting the retained unit $u \in R$ in version i and $\mathrm{Adv}(T, u) = |\alpha_1(T, u) - \alpha_0(T, u)|$ for the record-level membership advantage at u. Then, for every $u \in R$, every such s and every such T,

$$\mathrm{Adv}(T, u) \;\le\; \mathrm{TV}(P_1, P_0) \;=\; \mathrm{TV}\big(\mathrm{law}\, V_1, \mathrm{law}\, V_0\big) \;=\; 0,$$

where $\mathrm{TV}(P, Q) = \sup_S |P(S) - Q(S)|$ is the total variation distance.

Proof. The test's own randomness ω is independent of the version, so T accepts with probability $\alpha_i(T, u) = \mathbb{E}_{P_i}[\tau(V, s)]$ for the function $\tau(v, s) = \mathbb{P}_\omega(T \text{ accepts } u \mid v, s)$ with values in $[0, 1]$. Write $\Delta = P_1 - P_0$. Its total mass is zero, so its positive part and its negative part have the same mass, and that mass is $\mathrm{TV}(P_1, P_0)$; hence $|\alpha_1 - \alpha_0| = |\int \tau \, d\Delta| \le \mathrm{TV}(P_1, P_0)$, with equality for the likelihood-ratio test $\tau = \mathbf{1}\{\Delta > 0\}$. This is the first inequality. For the first equality, $s_i = f(V_i, X)$ with the same map f and the same deleted subject X in both versions, so both laws P_i are supported on the graph $\{(v, f(v, X))\}$, which is in bijection with the alphabet of V; total variation is invariant under a bijection of the support, so $\mathrm{TV}(P_1, P_0) = \mathrm{TV}(\mathrm{law}\, V_1, \mathrm{law}\, V_0)$. For the second equality, both versions run the same deterministic ridge rule on the same frozen map ψ. Exact unlearning of r is defined as returning the readout that the ridge rule produces from R (a Woodbury downdate of the trained state reproduces it in exact arithmetic), and in the second version the rule runs on R directly; ψ and R are the same objects in both versions, so $V_1 = V_0$, the two laws coincide, and the distance is zero. □

Corollary $1'$a (side information). In particular, a test that also sees z_X, μ_R, their hashes, or any quantity computed by one map from (ψ, w_R, R, X) has record-level membership advantage zero at every retained unit. No similarity between a carrier and the deleted subject enters, because there is no quantity for it to bound.

Corollary $1'$b (approximate unlearning). Suppose the unlearning step is only (ε, δ)-indistinguishable from retraining on R, that is, $P_1(S) \le e^\varepsilon P_0(S) + \delta$ and $P_0(S) \le e^\varepsilon P_1(S) + \delta$ for every event S. Then, for every u, s and T as above,

$$\mathrm{Adv}(T, u) \;\le\; \mathrm{TV}(P_1, P_0) \;\le\; \delta + (1 - \delta)\tanh(\varepsilon/2) \;=\; \frac{e^\varepsilon - 1 + 2\delta}{e^\varepsilon + 1},$$

and the right side, the largest total-variation distance compatible with (ε, δ)-indistinguishability (Kairouz et al., 2017), is $(e^\varepsilon - 1)/(e^\varepsilon + 1)$ at $\delta = 0$. The unlearning of this paper is exact, so the bound is stated for comparison and for implementations whose downdate differs from retraining by rounding.

Proof. The first inequality is the data-processing step of the previous proof. For the second, fix an event S and write $d = P_1(S) - P_0(S)$ and $Q = P_0(S)$. The first constraint gives $d \le (e^\varepsilon - 1)\, Q + \delta$.

The second constraint applied to the complement S^c gives $1 - Q \le e^{\varepsilon}(1 - Q - d) + \delta$, that is, $d \le [(e^{\varepsilon} - 1)(1 - Q) + \delta]/e^{\varepsilon}$. The first bound increases and the second decreases in Q; they meet at $Q^* = (1 - \delta)/(e^{\varepsilon} + 1) \in [0, 1]$, where both equal $d^* = (e^{\varepsilon} - 1 + 2\delta)/(e^{\varepsilon} + 1) = \delta + (1 - \delta)\tanh(\varepsilon/2)$. Hence $\sup_S d \le d^*$, and by symmetry of the two constraints the same bound holds for $\sup_S(P_0(S) - P_1(S))$. □

Remark. Proposition $1'$ restates Proposition 1 as a bound and adds four things: the measure of the advantage (total variation), the test that attains it (likelihood ratio), the side information z_X written into the statement with the conclusion unchanged, and the way the bound leaves zero when unlearning is only approximate. Like Proposition 1 it says nothing about the strength of membership estimators, and the reconstruction excess of Section 2.3 remains a different object on the same pool.

A.1 Rank-k prediction identity

Let $A = \Phi_R^\top \Phi_R + \lambda I$, $b = \Phi_R^\top y_R$, and $w_b = A^{-1}b$. After adding the carrier block, the ridge normal matrix and right-hand side are $A + UU^\top$ and $b + Uy_c$. Woodbury gives

$$(A + UU^\top)^{-1} = A^{-1} - A^{-1}U(I + U^\top A^{-1}U)^{-1}U^\top A^{-1}.$$

Multiplying by $b + Uy_c$ and collecting terms yields

$$w_k = w_b + A^{-1}U(I + M)^{-1}(y_c - q),$$

where $M = U^\top A^{-1}U$ and $q = U^\top A^{-1}b$. Left-multiplication by $a_X^\top$ gives

$$a_X^\top w_k = p_b + g^\top (I + M)^{-1}(y_c - q),$$

with $p_b = a_X^\top w_b$ and $g = U^\top A^{-1}a_X$.

A.2 Noise-averaged squared-error reduction

Let $e_b = p_b - \mu_X$, $y_c = c_0 + \varepsilon$, and $v = (I + M)^{-1}g$. The deleted subject's error after adding carriers is $m_k + v^\top \varepsilon$, where $m_k = e_b + g^\top (I + M)^{-1}(c_0 - q)$. If $\mathbb{E}[\varepsilon \mid \Phi_R, y_R, a_X, U] = 0$ and $\mathrm{Cov}(\varepsilon \mid \cdot) = \sigma^2 I$, then

$$\mathbb{E}[(m_k + v^\top \varepsilon)^2 \mid \cdot] = m_k^2 + \sigma^2 v^\top v.$$

Subtracting this from the carrier-free squared error e_b^2 proves Eq. (1):

$$\Delta_k = e_b^2 - m_k^2 - \sigma^2\, g^\top (I + M)^{-2} g.$$

The theorem is finite-sample and conditional. If $B = \mathbb{E}[e_b^2]$ is used later, that is an additional expectation over the stated deleted-subject/pool population; it is not part of the conditional identity.

A.3 Exact duplicate endpoint

For $k = 1$, $U = a_X$, $c_0 = \mu_X$, and $h = a_X^\top A^{-1} a_X$, one has $M = h$, $g = h$, and $q = a_X^\top w_b = p_b$. Hence $m_1 = e_b/(1 + h)$ and

$$\Delta_1 = e_b^2 - \big(e_b/(1 + h)\big)^2 - \sigma^2 h^2/(1 + h)^2.$$

Taking the additional expectation $B = \mathbb{E}[e_b^2]$ over the deleted subject gives

$$\mathbb{E}[\Delta_1] = h\,[(2 + h)B - h\sigma^2]/(1 + h)^2,$$

which is positive exactly when $(2 + h)B > h\sigma^2$, with limit $B - \sigma^2$ as the leverage grows. The condition is only for this exact-duplicate endpoint.

A.4 Shared-noise convention

Theorem 1 averages over carrier noise that is independent of the deleted subject. For the realized-versus-mean comparison we use the explicit convention $\mathrm{err} = d + \varepsilon_0 - h_v^\top \varepsilon_c$, with $h_v = (I + M)^{-1}g$ and $\Sigma_{c0} = \mathrm{Cov}(\varepsilon_c, \varepsilon_0)$. Expanding the square gives the cross term $-2h_v^\top \Sigma_{c0}$ in realized error. Under the paper's excess convention (baseline squared error minus the squared error after the carrier is added), $\mathrm{excess}_{\mathrm{realized}} - \mathrm{excess}_{\mathrm{mean}} = +2h_v^\top \Sigma_{c0}$. An injected copy shares the deleted subject's realized noise, so its realized excess follows a different estimand. Reversing either covariance orientation or excess direction reverses the displayed sign; the convention above is used in Section 3.3 and Appendix C.

A.5 KERNEL-RIDGE FORM OF THEOREM 1

Premises. (K1) κ is a fixed positive-semidefinite kernel on the input space. By Moore–Aronszajn there is a Hilbert space H and a lift ψ with $\kappa(x, x') = \langle \psi(x), \psi(x') \rangle$, and the same ψ is used in the carrier-free and in the carrier-retained fit. (K2) $\lambda > 0$. (K3) The readout is kernel ridge regression: trained on inputs Z with labels y_Z it predicts $f_Z(x) = k_{Zx}^\top (K_{ZZ} + \lambda I)^{-1} y_Z$, where $(K_{ZZ})_{ij} = \kappa(z_i, z_j)$ and $(k_{Zx})_i = \kappa(z_i, x)$. Nothing about the raw input is used: ψ need not be linear in the series and $\dim H$ may be infinite. The explicit-feature case of Section 3.1 is $\kappa(x, x') = \psi(x)^\top \psi(x')$ with the frozen feature map as ψ.

Notation. $\Phi_R : H \to \mathbb{R}^n$, $(\Phi_R v)_i = \langle \psi(r_i), v \rangle$, with adjoint $\Phi_R^\top \alpha = \sum_i \alpha_i \psi(r_i)$; $a_X = \psi(X)$; $U = [\psi(u_1), \dots, \psi(u_k)] : \mathbb{R}^k \to H$. Gram blocks: $K_{RR} = \Phi_R \Phi_R^\top$ $(n \times n)$, $k_{RX} = \Phi_R a_X$ (n), $K_{RU} = \Phi_R U$ $(n \times k)$, $K_{UU} = U^\top U$ $(k \times k)$, $k_{UX} = U^\top a_X$ (k), $\kappa_{XX} = \kappa(X, X)$; $G = K_{RR} + \lambda I_n$. Under (K1)–(K2), $G \succ 0$, and $A = \Phi_R^\top \Phi_R + \lambda I_H \succeq \lambda I_H$, so A^{-1} exists with $\|A^{-1}\| \le 1/\lambda$; $M = U^\top A^{-1} U \succeq 0$, so $I_k + M \succ 0$. The Woodbury step of A.1 uses only these two facts and the finite rank of U, and A.2–A.3 use only the $k \times k$ and scalar quantities (p_b, g, M, q). Hence A.1–A.3 hold on H verbatim; what the kernel case adds is the computation of (p_b, g, M, q, h) from Gram matrices (Steps 1–3) and an independent finite-matrix derivation (Step 4).

Step 1: push-through and resolvent identities. From $A\Phi_R^\top = (\Phi_R^\top \Phi_R + \lambda I_H)\Phi_R^\top = \Phi_R^\top (\Phi_R \Phi_R^\top + \lambda I_n) = \Phi_R^\top G$,

$$A^{-1} \Phi_R^\top = \Phi_R^\top G^{-1}. \tag{A.5.1}$$

From $A = \Phi_R^\top \Phi_R + \lambda I_H$ we get $I_H = A^{-1} \Phi_R^\top \Phi_R + \lambda A^{-1}$, and inserting (A.5.1),

$$\lambda A^{-1} = I_H - \Phi_R^\top G^{-1} \Phi_R. \tag{A.5.2}$$

(A.5.1) evaluates A^{-1} only on the range of $\Phi_R^\top$, i.e. on vectors $\Phi_R^\top \alpha$; $b = \Phi_R^\top y_R$ is such a vector. a_X and the columns of U in general are not (they are new inputs, and for $\dim H > n$ almost never lie in the span), so for them (A.5.2) is the identity to use. For any $u, v \in H$, (A.5.2) gives

$$u^\top A^{-1} v = \lambda^{-1} \big[\langle u, v \rangle - k_{Ru}^\top G^{-1} k_{Rv} \big], \qquad k_{Ru} = \Phi_R u, \tag{A.5.3}$$

that is, the A^{-1} inner product is λ^{-1} times the residual inner product after ridge-regressing both vectors on the retained features. For $u = v$ the bracket equals $\lambda \cdot v^\top A^{-1} v \ge 0$.

Step 2: Gram forms. Apply (A.5.1) to b and (A.5.3) to a_X and U:

$$\begin{aligned}
p_b &= a_X^\top A^{-1} b = k_{RX}^\top G^{-1} y_R && \text{(the carrier-free kernel-ridge prediction at } X\text{)}, \\
q &= U^\top A^{-1} b = K_{RU}^\top G^{-1} y_R && \text{(its predictions at the carrier inputs)}, \\
g &= U^\top A^{-1} a_X = \lambda^{-1} \big[k_{UX} - K_{RU}^\top G^{-1} k_{RX} \big], \\
M &= U^\top A^{-1} U = \lambda^{-1} \big[K_{UU} - K_{RU}^\top G^{-1} K_{RU} \big], \\
h &= a_X^\top A^{-1} a_X = \lambda^{-1} \big[\kappa_{XX} - k_{RX}^\top G^{-1} k_{RX} \big].
\end{aligned}$$

The λ^{-1} in g, M, h is where the kernel case differs from a naive substitution. Decompose $a_X = \Phi_R^\top \alpha + a_\perp$ with $a_\perp$ orthogonal to $\mathrm{span}\{\psi(r_i)\}$. Then $k_{RX} = K_{RR}\alpha$ and the bracket in h equals $\lambda \alpha^\top K_{RR} G^{-1} \alpha + \|a_\perp\|^2$, so

$$h = \alpha^\top K_{RR} G^{-1} \alpha + \lambda^{-1} \|a_\perp\|^2 :$$

the in-span part stays bounded as $\lambda \to 0$, while the out-of-span part is the squared distance from a_X to the retained span divided by λ. The same decomposition applies column-wise to U in g and M. Writing g as the raw similarity k_{UX}, or as the fitted similarity $K_{RU}^\top G^{-1} k_{RX}$, or applying (A.5.1) to a_X, all drop the λ^{-1} term and mis-state the leverage exactly in the case that matters for a new deleted subject.

Step 3: the λ^{-1} cancels in the theorem. Set $\hat{g} := \lambda g = k_{UX} - K_{RU}^\top G^{-1} k_{RX}$ and $S := \lambda(I_k + M) = K_{UU} + \lambda I_k - K_{RU}^\top G^{-1} K_{RU}$. Then $g^\top (I + M)^{-1} = \hat{g}^\top S^{-1}$ and $g^\top (I + M)^{-2} g = \hat{g}^\top S^{-2} \hat{g}$, so Theorem 1 and Eq. (1) read

$$f_{R \cup U}(X) = p_b + \hat{g}^\top S^{-1} (y_c - q), \qquad m_k = e_b + \hat{g}^\top S^{-1} (c_0 - q), \qquad \Delta_k = e_b^2 - m_k^2 - \sigma^2 \hat{g}^\top S^{-2} \hat{g}, \tag{A.5.4}$$

in which every symbol is a finite Gram-matrix quantity and λ enters only additively. S is the Schur complement of G in the bordered matrix $K_{ZZ} + \lambda I_{n+k}$ of Step 4, which is positive definite for $\lambda > 0$, so $S \succ 0$. This form is also the numerically stable one when λ is small and carriers lie outside the retained span.

Step 4: independent dual derivation. By (K3), kernel ridge on $R \cup U$ predicts

$$f_{R\cup U}(X) = \begin{bmatrix} k_{RX} \\ k_{UX} \end{bmatrix}^\top \begin{bmatrix} G & K_{RU} \\ K_{RU}^\top & K_{UU} + \lambda I_k \end{bmatrix}^{-1} \begin{bmatrix} y_R \\ y_c \end{bmatrix}.$$

Block inversion with the Schur complement S gives the blocks $G^{-1} + G^{-1}K_{RU}S^{-1}K_{RU}^\top G^{-1}$, $-G^{-1}K_{RU}S^{-1}$, $-S^{-1}K_{RU}^\top G^{-1}$, S^{-1}. Multiplying out and collecting terms with $p_b = k_{RX}^\top G^{-1} y_R$ and $q = K_{RU}^\top G^{-1} y_R$,

$$f_{R\cup U}(X) = p_b + \left[k_{UX} - K_{RU}^\top G^{-1} k_{RX}\right]^\top S^{-1}(y_c - q),$$

which is (A.5.4). This derivation uses finite matrices only, so the kernel statement does not rest on the operator-level Woodbury step; the two routes agree. The affine dependence of $f_{R\cup U}(X)$ on y_c is explicit, so A.2 goes through with $v = S^{-1}\hat{g}$.

Endpoint. For $k = 1$ and $u_1 = X$ (so $\psi(u_1) = a_X, c_0 = \mu_X$): $\hat{g} = \lambda h$, $S = \lambda(1 + h)$, $q = p_b$, and (A.5.4) reduces to A.3 with $h = \lambda^{-1}[\kappa(X, X) - k_{RX}^\top G^{-1} k_{RX}]$.

Conclusion and scope. Under (K1)–(K3), Theorem 1 holds for kernel ridge regression as the same identity with (p_b, g, M, q) given by Step 2, equivalently in the λ-free form (A.5.4). The prediction identity needs no distributional assumption; the noise-averaged form needs the conditional second moment of ε exactly as in A.2. The statement does not cover a kernel that is learned or that changes between the two fits (an encoder that moves during unlearning, or a neural tangent kernel after feature learning: then A itself changes and no rank-k update exists), $\lambda = 0$ (A is singular when $\dim H > n$, and the interpolant is a different object), indefinite kernels (no ψ; G need not be positive definite), or a learned adversary. Numerical check: `work/verify_theorem.py` verifies (A.5.1)–(A.5.2) and the Step-2 forms against the primal A^{-1} quantities on an explicit random-Fourier lift with $D < n$ and $D > n$, and for an RBF kernel (infinite-dimensional lift, no primal available) verifies (A.5.4) against the bordered kernel-ridge refit, the exact ε-expectation of the squared error, and the kernel-form endpoint, all to below 10^{-10} (see `work/verify_output.txt`).

A.6 Assumptions behind Theorem 1

Columns: *Stated* records whether the paper states the assumption (explicit) or only uses it (implicit) and where; *Used in* names the statement that needs it: P = prediction identity (A.1), N = noise-averaged Eq. (1) (A.2), E = duplicate endpoint (A.3), R = realized-versus-mean comparison (A.4, Section 3.3), K = kernel form (A.5); *Necessary* says whether the statement fails without it; *Experiments* says whether the reported protocol satisfies it.

#	Assumption	Stated	Used in	Necessary	Experiments
1	Fixed feature map ψ: the same map in the carrier-free and the carrier-retained fit, not a function of y_c, of the deleted subject's labels, or of the deletion	Explicit: Section 3 preamble, Section 8, A.5 (K1)	P, N, E, K	Yes. The Woodbury step needs one common A; if ψ moves, A changes and no rank-k update exists	Satisfied on every frozen-feature cell (Chronos-t5 tiny to large, Chronos-Bolt, TimesFM-2.5). Violated by construction on the fine-tuned-encoder arms, which Section 6.4 and Section 8 report as a boundary outside Theorem 1
2	$\lambda > 0$	Explicit: Section 3 preamble, A.5 (K2)	P, N, E, K	Yes when $\dim H \geq n$ (A singular at $\lambda = 0$); the min-norm interpolant is a different object. Also needed for G^{-1} and the λ^{-1} Gram forms	Satisfied: the headline readout is closed-form ridge with penalty 10^{-3} (Section 6); the kernel-ridge ladder uses a fixed positive penalty
3	Positive-semidefinite kernel (kernel case)	Explicit: A.5 (K1); vacuous for explicit features	K	Yes for the kernel form: it gives ψ (Moore–Aronszajn) and $G \succ 0$. Automatic for explicit features	Satisfied: linear ridge on explicit features; the RBF kernel of the boundary ladder is positive definite

continued on next page

#	Assumption	Stated	Used in	Necessary	Experiments
4	Carrier noise conditionally mean zero, $\mathbb{E}[\varepsilon \mid \Phi_R, y_R, a_X, U] = 0$	Explicit: Section 3.1, A.2	N, E	Only as a normalization: any conditional mean can be moved into $c_0 := \mathbb{E}[y_c \mid \cdot]$. P needs nothing about ε	Not directly verifiable: c_0 is not observed separately from y_c on natural pools. The deployed score plugs in the realized y_c and widens the abstention band by the plug-in half-width (Section 3.3, Appendix E.2)
5	$\mathrm{Cov}(\varepsilon \mid \cdot) = \sigma^2 I$ (homoscedastic, uncorrelated across the k carriers)	Explicit: Section 3.1, A.2	N, E	No. With $\mathrm{Cov}(\varepsilon \mid \cdot) = \Sigma_c$ the noise term is $v^\top \Sigma_c v$, $v = (I + M)^{-1} g$, still ≤ 0 for $\Sigma_c \succeq 0$, so the sign reading of Section 3.1 survives; only the scalar σ^2 form needs $\sigma^2 I$, and for $k = 1$ (Eq. (2)) σ^2 is simply $\mathrm{Var}(\varepsilon)$. A.4 already uses the covariance form	Not verified: several carriers of one deleted subject are often windows of one retained subject, whose noise may be correlated; the protocol does not estimate Σ_c. If it is estimated, use $v^\top \Sigma_c v$
6	Carrier noise independent of the baseline fit (Φ_R, y_R) and of the deleted subject's own noise ε_0	Implicit in the conditioning of A.2 (only conditional first and second moments given (Φ_R, y_R, a_X, U) enter); explicit for the deleted subject's side in Section 3.3	N (conditional moments), R	For the mean-scored Eq. (1) only the conditional moments are needed. For the realized-error estimand: yes, shared noise adds $+2h_v^\top \Sigma_{c0}$ (A.4)	Natural pools: assumed. The disjoint descriptor/scoring split that would support it is enforced only in the certificate procedure (Appendix E.2); in the archived Section 6 measurements the scored windows are the deleted subject's training windows, which also formed the descriptor (Section 4.2), and a disjoint-window rerun moved no headline median by more than the 0.012 bootstrap resolution (Appendix F.5). Injected ladder: violated by construction (the copy shares the deleted subject's realized noise) and therefore scored as a different estimand (Section 5.1, Section 3.3). A positive control built from the subject's own later windows, at least 970 samples from descriptor and scoring windows (Appendix B.4), satisfies it by time separation, with a noise-independence caveat on CWRU where the innovation autocorrelation at the gap lag exceeds 0.1 on 22 of 40 records; an injected copy does not
7	Carrier set U fixed given the conditioning: selected from the post-deletion retained pool and from features only, not from y_c or ε	Implicit: A.1–A.2 condition on U; the selection rule is in Section 4.2	N, E	Yes for N: selecting carriers on their labels makes $\mathbb{E}[\varepsilon \mid U] \neq 0$ and breaks Eq. (1). P holds for any U	Satisfied: the detector selects C by de-centered feature cosine against z_X on the deployed retained pool after deletion; y_c enters only the score
8	Independent noise across the windows that form one subject-level unit (several carrier windows of one retained subject; the deleted subject's scoring windows E_X)	Implicit: $\sigma^2 I$ presumes it when a subject contributes several carrier windows; the average $B = \mathbb{E}[e_b^2]$ over the deleted subject (A.2) presumes exchangeable scoring windows	N (form of the noise term), E (through B), and the width of the reported intervals; not P	No for P and for the identity in Eq. (1) with general Σ_c; dependence changes the noise term (row 5) and the effective sample size behind B and the bootstrap interval	Not guaranteed: adjacent windows of one recording are dependent. Mitigations: the interval is a median bootstrap over subjects, not windows (Section 4.2); Paderborn adds a cluster bootstrap (Section 6.1, Section 8). The descriptor/scoring window split mitigates only inside the certificate procedure (Appendix E.2); in the archived Section 6 measurements descriptor and scored windows coincide (Section 4.2), and the disjoint-window rerun moved no headline median by more than the 0.012 bootstrap resolution (Appendix F.5)
9	A clean value μ_X of the deleted subject exists and e_b is measured against it	Implicit: Section 3.1 defines $e_b = p_b - \mu_X$	N, E	Definitional. Against realized labels the estimand shifts by the deleted subject's noise (A.4)	μ_X is not observed; the deployed estimator scores against held-out realized windows, so it measures the realized-error version, which differs from Eq. (1) by the $+2h_v^\top \Sigma_{c0}$ term of A.4, zero under row 6

Rows 1–3 and 7 carry the theorem: without them there is no common A, no inverse, or no valid conditioning on U. Row 4 is a normalization and row 5 a convenience, both replaceable by their covariance forms without changing the sign of the noise term. Rows 6, 8 and 9 concern the estimand

the experiments measure and leave the identity itself untouched; they are the reason Section 3.3 separates realized from mean-scored excess and the reason the injected ladder is scored under a different estimand.

A.7 Two coordinates at the duplicate endpoint

Section 3.3 refers to the following statement: at the duplicate endpoint the reconstruction excess and a residual-influence coordinate trade one for one along a fixed-χ fiber. At the duplicate endpoint, with $\eta = h/(1+h)$, $s = \mu_X^2 + \sigma^2$, $\iota = B/s$ and $\chi = \iota\eta$, the reconstruction excess Δ_{rec} and the residual-influence coordinate J_{res} satisfy $\Delta_{\mathrm{rec}} + J_{\mathrm{res}} = 2s\chi$, and along a fixed-$\chi$ fiber the two trade one for one. On a two-dimensional cohort support no continuous scalar statistic represents both, so a one-dimensional summary is exact only when cohort leverage is a function of χ: the map from the pair (χ, η) to the two observables is injective while a continuous scalar restricted to a circle cannot be. Here $J_{\mathrm{res}} := \mathbb{E}[I_1^2]$ with $I_1 := a_X^\top w_1 - p_b = \eta\,(y_c - p_b)$, the shift of the prediction at the deleted subject caused by retaining the duplicate (leverage times the duplicate's residual against the carrier-free fit), and $\Delta_{\mathrm{rec}} := \mathbb{E}[\Delta_1]$ of A.3; both expectations are over ε and then over deleted subjects at fixed leverage, as B in A.3. *Proof.* By A.1 with $g = M = h$ and $q = p_b$, $I_1 = \eta(y_c - p_b) = \eta(\varepsilon - e_b)$ and the post-carrier error is $e_b + I_1$, so $\Delta_1 = e_b^2 - (e_b + I_1)^2 = -2e_bI_1 - I_1^2$ for every realization. Averaging with $\mathbb{E}[\varepsilon \mid \cdot] = 0$, $\mathbb{E}[\varepsilon^2 \mid \cdot] = \sigma^2$ and $\mathbb{E}[e_b^2] = B$ gives $\Delta_{\mathrm{rec}} + J_{\mathrm{res}} = 2\eta B = 2s\chi$ and $J_{\mathrm{res}} = \eta^2(B + \sigma^2)$. On a fixed-$\chi$ fiber $B = s\chi/\eta$, so $J_{\mathrm{res}} = s\chi\eta + \sigma^2\eta^2$ and $\Delta_{\mathrm{rec}} = 2s\chi - J_{\mathrm{res}}$; (χ, η) is recovered from the pair as $\chi = (\Delta_{\mathrm{rec}} + J_{\mathrm{res}})/(2s)$ and the positive root of $\sigma^2\eta^2 + s\chi\eta - J_{\mathrm{res}} = 0$, while a continuous scalar on a circle cannot be injective (its image is an interval, and removing an interior point disconnects it). □

A.8 A noise ceiling for the recovery ratio

Section 3.3 refers to the following statement: with carrier noise the ratio Δ_k/e_b^2 has an attainable supremum independent of M. Fix the carrier side: M, the clean carrier residual $r := c_0 - q \neq 0$, and $\sigma > 0$, with U of rank k. Let the deleted subject's side vary over every direction a_X and clean value μ_X, hence over every $g = U^\top A^{-1} a_X$ and every $e_b \neq 0$.

Proposition 2 (noise ceiling) *Over all such deleted subjects,*

$$\sup \frac{\Delta_k}{e_b^2} \;=\; \frac{\|r\|^2}{\|r\|^2 + \sigma^2}.$$

The supremum is attained exactly when $(I+M)^{-1}g$ is parallel to r and $e_b = -(1+\sigma^2/\|r\|^2)\, g^\top (I+M)^{-1} r$, and it does not involve M. At $\sigma = 0$ the ratio takes every value in $(-\infty, 1]$, so no ceiling below 1 exists: the ceiling is a property of the noise term.

Proof. Write $v = (I+M)^{-1}g$ and $t = v^\top r$. By Theorem 1, $\Delta_k = e_b^2 - (e_b + t)^2 - \sigma^2\|v\|^2$. If $t = 0$ then $\Delta_k = -\sigma^2\|v\|^2 \leq 0$. Otherwise set $e_b = -\tau t$ with $\tau \neq 0$; then $\Delta_k/e_b^2 = (2\tau - 1 - K)/\tau^2$ with $K = \sigma^2\|v\|^2/t^2$. For fixed K the right side is largest at $\tau = 1 + K$, where it equals $1/(1+K)$. By the Cauchy-Schwarz inequality $t^2 \leq \|v\|^2\|r\|^2$, so $K \geq \sigma^2/\|r\|^2$, with equality exactly when v is parallel to r. Hence $\Delta_k/e_b^2 \leq \|r\|^2/(\|r\|^2 + \sigma^2)$, with equality exactly at the stated (v, e_b). The map $a_X \mapsto g$ is onto $\mathbb{R}^k$ because $A^{-1}U$ has rank k, and $e_b = p_b - \mu_X$ is free through μ_X, so every (v, e_b) occurs and the supremum is attained. At $\sigma = 0$, $K = 0$ and $(2\tau - 1)/\tau^2$ ranges over $(-\infty, 1]$ as τ ranges over the nonzero reals. □

A.9 The process-only carrier (post hoc explanation of the shared-process control)

Setting. Fix the carrier-free fit (Φ_R, y_R), $A = \Phi_R^\top\Phi_R + \lambda I$, $w_b = A^{-1}\Phi_R^\top y_R$, and write $\langle x, y\rangle_{A^{-1}} = x^\top A^{-1} y$. A window of the deleted subject has feature a and clean label μ, with baseline error $e_b = a^\top w_b - \mu$. A carrier block $U \in \mathbb{R}^{d\times k}$ with clean labels c and noise ε ($\mathbb{E}\varepsilon = 0$, $\mathrm{Cov}\,\varepsilon = \sigma^2 I$, independent of everything else) shifts the readout value at the deleted subject by $I = \langle z, a\rangle_{A^{-1}}$ with the carrier-side transfer vector $z = U(I+M)^{-1}(c + \varepsilon - U^\top w_b)$, $M = U^\top A^{-1} U$ (Appendix A.1), and the realized excess is $\Delta = e_b^2 - (e_b + I)^2 = -2e_bI - I^2$. Let P_θ be the law of (a, μ) for one window of a process θ under the fixed feature map, and let $s(w; \theta) = \mathbb{E}_\theta[a(a^\top w - \mu)]$

be the population score of a readout w on that process; $s(w;\theta) = 0$ exactly when w minimizes the population risk of θ. Write $s_b = s(w_b;\theta_X)$, $\zeta = \mathbb{E}z$, $S = e_b a$, and let $\varrho \in [-1,1]$ be the A^{-1}-correlation of z with $-S$ (the realization correlation). A carrier block is *process-only* for θ_X if (U,c) is independent of (a,μ) and every column-label pair has law P_{θ_X}: the block shares the deleted subject's process and nothing of its realization. Within this subsection I, S, v and z denote the shift, the score vector, the process signal and the transfer vector.

Lemma A.9. *(i)* For any carrier block,

$$\mathbb{E}[\Delta] \;=\; -2\langle s_b, \zeta\rangle_{A^{-1}} \;+\; 2\varrho\,\sigma_z\sigma_s \;-\; \mathbb{E}[I^2],$$

with $\sigma_z^2 = \mathbb{E}\|z-\zeta\|_{A^{-1}}^2$ and $\sigma_s^2 = \mathbb{E}\|S-s_b\|_{A^{-1}}^2$. For a process-only block $\varrho = 0$, so $\mathbb{E}[\Delta] = -2\langle s_b,\zeta\rangle_{A^{-1}} - \mathbb{E}[I^2]$: a process-transfer term, which is zero whenever the carrier-free readout already fits the deleted subject's process ($s_b = 0$), minus a fluctuation penalty. In that case $\mathbb{E}[\Delta] = -\mathbb{E}[I^2] \le 0$, and in general $\mathbb{E}[\Delta] \le 2\|s_b\|_{A^{-1}}\|\zeta\|_{A^{-1}}$; with the leverage replacement of Appendix C.1 and $k=1$ the transfer term is $2\|s_b\|_{A^{-1}}^2/(1+h) \le 2hB/(1+h)$, the cross term of the duplicate endpoint of Appendix A.3. For a pool-fitted readout the transfer term and the penalty are of comparable order, so the lemma fixes the dependence of $\mathbb{E}[\Delta]$ and leaves its sign open. *(ii)* Let $D_X = v + \xi_X$ and $D_u = v + \xi_u$ be the de-centered mean-pooled descriptors of the deleted subject and of the unit, v the de-centered process signal, ξ_X, ξ_u independent mean-zero fluctuations. If each fluctuation law is centrally symmetric and invariant under the orthogonal maps fixing v, then $\mathbb{E}\cos(D_X, D_u) = \varphi_X(\|v\|)\,\varphi_u(\|v\|)$ with $\varphi(t) = \mathbb{E}[(t+\xi_1)/\|te_1+\xi\|]$ nondecreasing in t, so the expected cosine is nondecreasing in $\|v\|$; without the symmetry, $\mathbb{E}\cos(D_X,D_u) \ge 1 - (\pi^2/4)\big(\mathbb{E}\|\xi_X\|^2 + \mathbb{E}\|\xi_u\|^2\big)/\|v\|^2$. The bound is informative once $\|v\|^2$ exceeds $(\pi^2/4)\big(\mathbb{E}\|\xi_X\|^2 + \mathbb{E}\|\xi_u\|^2\big)$. Below that threshold the bound is vacuous and the lemma makes no statement about the expected cosine. Both factors depend on the fluctuation laws as well as on $\|v\|$; the monotone reading holds with those laws fixed.

Proof. (i) $e_b I = \langle z, S\rangle_{A^{-1}}$; splitting both vectors into mean and centered parts gives $\mathbb{E}[e_b I] = \langle \zeta, s_b\rangle_{A^{-1}} - \varrho\sigma_z\sigma_s$, and $\Delta = -2e_bI - I^2$. Under independence the centered inner product has mean zero. The bounds are Cauchy's inequality; for the leverage form, $\mathbb{E}[u(c - u^\top w_b)] = -s_b$ because (u,c) has law P_{θ_X}, and $\|s_b\|_{A^{-1}}^2 = \|\mathbb{E}[A^{-1/2}a\,e_b]\|^2 \le \mathbb{E}[a^\top A^{-1}a]\,\mathbb{E}[e_b^2] = hB$. (ii) Independence gives $\mathbb{E}\langle \hat D_X, \hat D_u\rangle = \langle \mathbb{E}\hat D_X, \mathbb{E}\hat D_u\rangle$ and the symmetry gives $\mathbb{E}\hat D = \varphi(\|v\|)\,v/\|v\|$; pointwise, $t \mapsto (t+\xi_1)/\sqrt{(t+\xi_1)^2 + \|\xi_\perp\|^2}$ has derivative $\|\xi_\perp\|^2/\|te_1+\xi\|^3 \ge 0$, and both factors are nonnegative. For the bound, when $\|\xi\| < \|v\|$ the angle between v and $v+\xi$ is at most $\arcsin(\|\xi\|/\|v\|) \le (\pi/2)\|\xi\|/\|v\|$, and $1-\cos(\alpha+\beta) \le (\alpha+\beta)^2/2$; when $\|\xi\| \ge \|v\|$ the trivial bound $1-\cos \le 2$ is already below $(\pi^2/4)\|\xi\|^2/\|v\|^2$. □

Reading. The estimator of Section 4.2 responds to a process-only unit only through the misfit of the carrier-free readout on the deleted subject's process, and this is an expectation over realizations; for the detection caliber the statement holds up to a selection term that vanishes as detection becomes certain; the descriptor of Section 4.2 responds to the de-centered process signal $\|v\|$. Neither statement was available before the shared-process control was run; both are post hoc explanations of its end state, and neither assigns a value to any reading. The selection term is at most $\sigma_\Delta\sqrt{(1-p)/p}$, σ_Δ the standard deviation of Δ and p the detection rate; the forced caliber carries no such term. The bound is one Cauchy-Schwarz step: the selection term equals $\mathrm{Cov}(\Delta, \mathbf{1}\{\text{detected}\})/p$, and $|\mathrm{Cov}(\Delta, \mathbf{1}\{\text{detected}\})| \le \sigma_\Delta\sqrt{p(1-p)}$.

B Calibration and detection limit

This appendix is the complete record behind Sections 4 and 5. It follows the artifact chain: the raw-cosine artifact (B.1), what de-centering changes (B.2), the resolution of the corrected estimator (B.3), the controlled carriers (B.4), the three zeros a natural cell can print (B.5), and the record-level arm (B.6). Throughout, a deletion unit with no detected carrier has a reconstruction excess of exactly zero by construction; the definitions of B.5 call this a structural zero and keep that label for the figure legends.

B.1 THE ARTIFACT

This section shows why raw cosine similarity flags most of a frozen time-series pool as carriers and why the refit without them reads a spurious excess. On Chronos-Bolt and TimesFM-2.5 features (Ansari et al., 2024b; Das et al., 2024) over CWRU, Paderborn, PTB-XL, and the auxiliary ELD set, raw cosine exceeds 0.9 for 92 to 100 percent of retained units in every model-by-corpus cell. With Chronos-t5-tiny the fraction is 88 to 100 percent. Frozen time-series features share one large common direction, the shared feature direction of Section 4.1, and raw cosine measures that direction instead of the carrier (Appendix I). As a result, 4 of 8 cross-family cells pass the 0.1 threshold, with an apparent excess up to +18.27 on Paderborn (Figure 2). Flagging 48 to 52 of 59 series removes most of the pool, and the under-determined refit produces that excess. On Chronos-Bolt, the continuous-valued family of Section 4.1, raw cosine flags 0.90 to 1.00 of retained subjects across the four headline pools, so the artifact belongs to mean-pooled frozen features, whichever encoder produced them. Appendix I gives the geometry; the next section reports what removing the shared direction changes.

B.2 THE CORRECTION

This section reports what de-centering changes: the flagged fraction, the eight natural cells, a second encoder, non-learned features, the threshold scan, and six other anisotropy calibers. Subtracting the retained-pool mean before taking the cosine removes the shared direction, and Section 4.2 prints what the eight cells of Section 4.1 then read. The carrier-conditioned subsets of those cells read medians from -0.60 to +2.62 (Table 1), outside the pre-registered estimand.

With the encoder replaced by Chronos-Bolt, the pre-registered all-subject median reads +0.00 on CWRU, +0.03 [+0.00, +0.09] on Paderborn, +0.00 on PTB-XL, and +0.00 on MIMII, the same three states as on Chronos-t5-small. The two kinds of zero (Appendix B.5) do not depend on the encoder family, and de-centering brings the Chronos-Bolt flagged fraction of Appendix B.1 down to 0.00 to 0.59. Raw-cosine saturation also appears without pretraining: on 504 non-learned MiniRocket features (Dempster et al., 2021) every retained unit passes raw cosine 0.9 against every other on all four corpora, and de-centering removes it (Table 1). The de-centered subject-level zero replicates on CWRU, PTB-XL, and MIMII; Paderborn reads +0.102 [0.000, +0.139], larger than on Chronos and undetermined. A PTB-XL pool of 1000 patients, 180 with a carrier, keeps the median at 0.000 [0.000, 0.000], and six patients read above 0.1 (Table 1).

The threshold scan reports the same 20 natural cells at de-centered cosine 0.70, 0.80, 0.85, and 0.90. The fraction of subjects with at least one carrier rises from 0.000 to 0.700 on CWRU, from 0.100 to 0.650 on PTB-XL, and from 0.167 to 0.833 on MIMII between 0.90 and 0.70. The largest median excess on the grid is +0.072, on Paderborn at 0.70, and no cell passes the materiality gate. The headline stays at 0.90 as pre-registered; the scan is a sensitivity analysis and not a choice of threshold. Six other anisotropy corrections flag similar carrier sets and decide no cell (Table J.1; Appendix I.4). The de-centered detector therefore flags few natural carriers, and the next section asks how small an effect the estimator resolves once a carrier is flagged.

B.3 RESOLUTION OF THE CORRECTED ESTIMATOR

This section measures how small an injected carrier effect the estimator resolves and at which mixing fraction the detector first sees the injected unit. The injected-effect ladder injects one synthetic record that mixes a fraction φ of the deleted subject's training windows with windows of another randomly chosen subject. It z-normalizes the mixture, featurizes it with the same frozen encoder, and adds it to the pool as one more retained unit. We report two calibers. Under the detection caliber the injected unit counts as a carrier only if its de-centered cosine with the deleted subject reaches 0.9, which is the deployed detector. Under the forced caliber the injected unit always counts, which isolates the estimator from the detector.

Under the forced caliber the estimator resolves an excess of +0.024 [0.000, +0.059] at $\varphi = 0.35$ and +0.031 [+0.004, +0.083] at $\varphi = 0.5$ on CWRU. On PTB-XL it resolves +0.027 [+0.015, +0.047] at $\varphi = 0.35$. On MIMII it resolves +0.039 [+0.014, +0.096] at $\varphi = 0.35$ and +0.128 [+0.087, +0.193] at $\varphi = 0.5$. On Paderborn it resolves +0.088 [+0.065, +0.265] at $\varphi = 0.75$ above that run's natural baseline of +0.058, that ladder's own $\varphi = 0$ cell under the detection caliber. Under the detection

caliber the injected unit passes 0.9 for 83 percent of CWRU deleted subjects at $\varphi = 0.75$ and for 100 percent of Paderborn and MIMII deleted subjects at $\varphi = 1$. On PTB-XL it passes for 62 percent of deleted subjects at $\varphi = 1$, where the median de-centered similarity of an exact copy is 0.931 and the distribution straddles the threshold. The limit of detection is the smallest φ at which all three seeds have a lower interval bound above 0. It is $\varphi = 0.75$ on CWRU, $\varphi = 1$ on Paderborn and MIMII in increment terms, and unreached on PTB-XL. On the same ladder the exact identity of Theorem 1 reproduces the measured change in the deleted subject's error to within 1.9×10^{-12} in all 32 corpus-by-dose cells; this checks the implementation and is not a prediction. The ladder fixes the resolution for a carrier the detector sees; the next section plants carriers with a known or constructed effect and asks whether the detector sees them at all.

B.4 Controlled carriers

This section reports the five pre-registered controls in full: what the detector flagged, what the estimator read, and the known effect where one exists. The five controls test whether the de-centered detector of Section 4.2 flags a retained unit that shares the deleted subject's later windows, clean signal, process, or fitted spectrum but none of its realized noise. The first two control units are within-subject siblings of the deleted subject, the second of them the later-window control of Section 5.2. The third, the clean-signal control of Section 5.2, is a synthetic re-observation of its clean signal. The fourth is a new realization of its AR dynamics under fresh noise, and the fifth a new realization of only its fitted spectrum in the real pool. All five lie outside the subject-level estimand of Section 2.3; they test the sensitivity of the detector and the estimator, not the natural-pool estimand.

The first control synthesized the unit as the subject's low-pass component at the spectral median frequency plus fresh Gaussian noise with the residual's spectrum, at four noise ratios and three seeds. Under the detection caliber the unit stayed below the 0.9 threshold at every noise level on both main corpora, so the detected-caliber excess was 0.000 by definition, the pre-registered PC-BLIND state. Forced to count as a carrier, the same unit produced a material excess on CWRU, so the estimator, and not the detector, admits such a unit. The synthesis had treated the high-frequency texture that carries a subject's identity as noise, a design flaw recorded in its report. Its numbers came from a local execution with a reduced MIMII pool and enter no claim of this paper.

The later-window control retains one unit built from the deleted subject's own later windows, at least 970 samples after both the descriptor and the scoring windows, so no synthesis intervenes and the noise is separated in time. A pre-registration frozen before the run fixed its grid, constants, and end-state mapping, and the run on the full CWRU, MIMII, and Paderborn pools returned with no deviation. Because the AR(32) innovation autocorrelation at the gap lag exceeds 0.1 on 22 of 40 CWRU records, median 0.11, the CWRU end state carries a noise-independence caveat. On MIMII and Paderborn the gap-band autocorrelation, 0.05 and 0.04, sits at the white-noise level. At de-centered cosine 0.9 the detector flags the unit for 0.800 of CWRU deleted subjects, 8 of 10 bearings on all three seeds, with a Wilson interval [0.49, 0.94]. Yet the detected-caliber excess reads -0.003 [-0.017, +0.053] and the forced-caliber excess -0.01 [-0.03, +0.08]: the pre-registered state PC-DETECT-ONLY. Only the healthy Normal bearing reads a material +0.65; 6 of 10 deleted subjects read negative. On the full MIMII pool the detector flags the unit for 0.111 of deleted subjects, 4 of 36 rows, so the detected-caliber excess is a structural zero on 32 of 36 rows. The forced-caliber excess there reads +0.032 [+0.009, +0.064]: the state PC-BLIND. The descriptive Paderborn cells read detection 0.700 and 0.650 on the 22 and 26 bearings with excess +0.057 and +0.048, the natural-carrier readings of Section 6.1 unchanged by the injected unit. The reading for the paper is therefore split by corpus. On CWRU the detector sees a same-subject fresh unit and the estimator reads no excess above the material-effect threshold from it, so the null of Section 6 states the estimator's resolution for such units. On MIMII the null does not cover such units, because the detector misses them. On Paderborn the detector flags the deleted subject's own later windows in 0.68 of rows and another subject's in 0.09. Yet the unit's own contribution stays at +0.01 [+0.00, +0.02] (forced: +0.01), below the material-effect threshold, the same shape as on CWRU. On PTB-XL the detector flags the deleted subject's own later windows in only 0.03 of rows (median de-centered cosine 0.65) against 0.00 for another subject's; forced into the fit the unit reads +0.01 [-0.00, +0.06].

The clean-signal control constructs the truth. Twenty-four synthetic subjects, each an AR(4) process with its own dynamics, supply two records apiece. The injected unit re-observes the deleted subject's frozen clean signal under fresh white noise at a noise-to-signal ratio of 0.25, that is +6 dB,

so its reconstruction excess is known analytically, +0.63 to +0.68 between the two reference readouts. A negative arm injects an independent subject whose excess is 0 by construction. The same Chronos-t5-small features and the same estimator read both units at three seeds. At de-centered cosine 0.9 the detector flags the constructed carrier in 1 of 72 subject-seed rows, median similarity 0.696, so the detected-caliber excess is a structural zero on 71 of 72 rows. Forced to count, the estimator reads +0.072 [+0.044, +0.097], about one ninth of the truth, and at +3 and 0 dB it reads +0.032 and +0.018 against truths of +0.43 and +0.25. That fraction belongs to one carrier family at one noise level. Across the three noise levels it reads 0.115, 0.075 and 0.074 of the truth, the detected caliber reads 0 at all three, and no natural-pool reading converts through it. The independent unit is never flagged and reads -0.002 [-0.011, +0.010] when forced. The pre-registered end state is PC-BLIND-SYNTH with a SPECIFIC negative arm. The miss is specific to this family of carriers, which share a realization but carry weak dynamics. The mean-pooled descriptor averages the realization away and keeps the dynamics, which is why the real fresh windows of the later-window control pass the threshold on CWRU and this constructed unit does not. The null of Section 6 therefore states the detector's range for this family. With the same encoder, the window-max, covariance and whitened-mean descriptors flag this carrier in 0.00 of rows, against 0.01 for the mean-pooled one, and flag the independent unit in 0.00. A descriptor that does not average over windows is no more sensitive to the shared realization than the mean-pooled one on these features.

Remark. Theorem 1 makes the reconstruction excess an exact rank-k functional of the feature geometry: the residual $y_c - q$ reaches the scoring windows (design Φ_E) through $\Phi_E A^{-1} U (I+M)^{-1}$ and pays for carrier noise through $(I + M)^{-2}$. The one-ninth reading above is a measurement on this carrier family and at this descriptor and readout interface, and no bound.

The fourth control shares the process and changes the realization. With the twenty-four AR(4) subjects of the clean-signal control, the planted unit is a new realization of the deleted subject's dynamics under fresh white noise at noise ratio 0.25. It runs at pole radii 0.40, 0.70 and 0.90 with the same encoder, threshold, estimator, and three seeds. At pole radius 0.90 the detector flags it in 0.51 of the 72 subject-seed rows, median similarity 0.903, and the detected-caliber excess reads 0.00 [0.00, 0.01] and the forced estimator 0.01 [-0.02, +0.02]. The pre-registered end state is B-DETECTED-NONMATERIAL with a SPECIFIC negative arm. Its pre-registered end states are detected-material, detected-nonmaterial, and blind; it is a response control and has no pass criterion of its own. At noise ratio 0.25 the detection rate reads 0.00, 0.22 and 0.51 across the three pole radii; the independent unit is flagged in 0.03 of rows. The window-max, covariance and whitened-mean descriptors flag neither unit at the same threshold. Figure A7 and Figures A5 and A6 draw the flagged fractions and the readings of the later-window control and the three controls after it side by side. Lemma A.9 (Appendix A.9) states the process-only case: transfer vanishes at a process-optimal readout, and the descriptor cosine rises.

The fifth control plants, in the real CWRU pool, one retained unit that shares only the deleted bearing's fitted AR(16) spectrum: a Gaussian realization under fresh noise. The detector flags it in 0.07 of rows (median de-centered cosine 0.32), against 0.80 for the bearing's own later windows; forced into the fit, the estimator reads -0.01 [-0.03, +0.03]. No window-level descriptor sees this unit either. The window-max, covariance and whitened-mean descriptors flag it in 0.00 of rows and the other bearing's unit in 0.00, against 0.07 and 0.10 for the mean.

On the clean-signal control's design, with Chronos-Bolt features, the constructed carrier is flagged in 0.50 of rows, median similarity 0.90, and the detected-caliber excess reads +0.08 [0.00, +0.17], undetermined by the rule of Section 4.2. Forced to count, the estimator reads +0.22 [+0.18, +0.26], 0.35 of the truth. The independent unit is flagged in 0.01 of rows yet reads -0.03 [-0.04, -0.01] at the forced caliber. At that caliber we report the second-encoder arm as inconclusive under its pre-registered criterion, which requires the independent unit's interval to contain zero and does not depend on whether the two readings coincide.

Post hoc explanation (Lemma A.9). For a retained unit that shares only the deleted subject's process, with an independent realization, Theorem 1 gives an expected excess equal to a process-transfer term minus a fluctuation penalty. The transfer term vanishes once the carrier-free readout already fits that process. The mean-pooled de-centered cosine instead rises with the de-centered process signal for fixed fluctuation laws. Five positive-control readings therefore fell short of their pass states (Section 5), and the next section defines what a zero on a natural cell can then mean.

B.5 THREE ZEROS AND LEMMA B.0

This section defines the three zeros a natural cell can print and proves when the carrier-free majority fixes the all-subject median and its bootstrap interval at zero.

Setting. A cell is a corpus, an encoder, a deletion granularity and a caliber, with n deletion units. For unit i the detector returns the carrier set $C_i \subseteq R$ at de-centered cosine τ, and the estimator returns a reading e_i, the mean over seeds of the reconstruction excess after removing C_i. When C_i is empty in every seed no refit takes place and $e_i = 0$ by definition. Write K for the number of units with a nonempty carrier set, and n_+ and n_- for the numbers of units with $e_i > 0$ and $e_i < 0$. Write m for the median of (e_i) and $[\mathrm{lo}, \mathrm{hi}]$ for its percentile bootstrap interval at level $1 - \alpha$ from B resamples of the units. The pre-registered rule of Section 4.2 with gate g reads: material when $m > g$ and $\mathrm{lo} > 0$; null when $\mathrm{hi} < g$; undetermined otherwise.

Definition (three zeros). (Z-S) A unit with an empty carrier set is a structural zero; a cell with $K < n/2$ is structurally dominated. (Z-M) A unit with a nonempty carrier set whose reading lies inside the gate is a measured zero. A cell with $K \geq n/2$ that satisfies the null rule with a median of 0 is a measured zero at the cell level. A cell with $K \geq n/2$ that satisfies the null rule with a nonzero median is a below-gate cell with a carrier-determined median, whether or not its interval contains 0. The term measured zero is reserved for the two cases above. (Z-R) Let $\mathbb{F}$ be the set of pairs (carrier family, strength) for which a pre-registered positive control ended in its pass state: detection in at least half of its rows and a material reading at the detection caliber. The pass states are A-DETECTED-MATERIAL for the fourth control and A-CALIBRATED, B-MISCALIBRATED-ATTENUATED or B-MISCALIBRATED-INFLATED for the third control, whose own end state was C-BLIND, printed PC-BLIND-SYNTH in Appendix B.4. A cell that satisfies the null rule is a zero within the measurement range: it bears on carriers of a pair in $\mathbb{F}$ and is silent about every other pair. In this archive no control reached its pass state, so $\mathbb{F}$ is empty: each natural null states the measurement range (Section 5) and bears on no carrier family.

Lemma B.0. (a) $m = 0$ exactly when $n_+ < n/2$ and $n_- < n/2$, barring an exact cancellation of the two middle readings (even n, no zero unit, $n_+ = n_- = n/2$). In every case $K < n/2$ forces $m = 0$ whatever the K carrier readings are. (b) Up to the Monte Carlo error of the resampling and to resamples that contain no zero unit, an event of probability at most $(K/n)^n$, the following holds. The upper bound hi is 0 exactly when the resample median is positive with probability below $\alpha/2$ and nonnegative with probability at least $\alpha/2$, and the lower bound lo is 0 likewise with the signs exchanged. For a cell with $n_+ < n/2$ and $n_- < n/2$ the second condition holds automatically. A count $\mathrm{Binomial}(n, k/n)$ with an integer $k < n/2$ has median k and therefore reaches $\lceil n/2 \rceil$ with probability at most $1/2$. Hence $K < n/2$ with $\mathbb{P}(\mathrm{Binomial}(n, K/n) \geq \lceil n/2 \rceil) < \alpha/2$ prints $[0, 0]$ and satisfies the null rule for every value of the carrier readings. (c) A positive control is such a cell with the planted unit as carrier. The end state C-BLIND (C-BLIND-PROCESS of the fourth control) is (a) and (b) on that cell. The end state B-DETECTED-NONMATERIAL (C-DETECT-ONLY of the third control) is a measured zero on it when it satisfies the null rule with a median of 0, and a below-gate or undetermined cell otherwise. Only the pass state (A-DETECTED-MATERIAL of the fourth control; A-CALIBRATED, B-MISCALIBRATED-ATTENUATED or B-MISCALIBRATED-INFLATED of the third) registers the carrier family at that strength, so a natural-cell null bears on registered families only.

Proof. Order the readings as negatives, zeros, positives; for odd n the middle position is a zero exactly when both n_- and n_+ are at most $(n-1)/2$. For even n the median is the mean of the two middle positions, and both are zeros exactly when both counts are at most $n/2 - 1$. If one is a zero and the other is not, the mean is nonzero. If neither is a zero, the mean is zero only when the two readings cancel exactly, which requires $n_+ = n_- = n/2$ and no zero unit, the excluded case. This is (a). A resample of the units has median above zero exactly when its positive count reaches $\lceil n/2 \rceil$, provided that for even n the resample contains a zero unit. Without one, a positive count of exactly $n/2$ pairs a negative with a positive in the middle; that count is $\mathrm{Binomial}(n, n_+/n)$. The upper percentile of the resample medians is zero exactly when the median is positive in fewer than $\alpha/2$ of the resamples and nonnegative in at least $\alpha/2$ of them; the lower percentile is zero under the mirrored conditions. For $n_- < n/2$ the negative count is $\mathrm{Binomial}(n, n_-/n)$ with the integer mean

n_-, which is also its median. The resample median is therefore negative with probability at most $\mathbb{P}(\text{Binomial}(n, n_-/n) \geq n_- + 1) \leq 1/2$, and the upper percentile cannot fall below 0. This is (b). Part (c) is the end-state rule of the pre-registration read on the control cell □.

These three zeros carry no statement about the effect in each subject: with fewer than half of the subjects flagged, the all-subject median stays at 0 whatever the flagged subjects read, even at readings of +10 or −10. On CWRU the detector flags no carrier, yet 7 of 10 bearings have a nearest retained unit at de-centered cosine between 0.7 and 0.9, above the 0.696 median of the known constructed carrier, so missed carriers stay possible. The record-level arm, where carriers are declared and not detected, closes the chain.

B.6 Within-subject carriers under record-level deletion

This section reports the record-level arm, where the deleted record's own siblings stay in the pool: which requests read LEAK under each label rule and what the sign screen recovers. Record-level deletion supplies natural positives. When one record leaves and its within-subject siblings stay, 30 of 249 record-level requests on CWRU, Paderborn, and PTB-XL read LEAK under the original label rule of Section 6.5, with 90 SAFE and 129 undetermined. The carriers are the declared sibling records, and the estimand is the record-level within-subject increment. It is the measured form of a two-record construction, where the sibling gives a positive record-level increment and a subject-level increment of exactly 0 (Appendix A, Remark). On these 120 two-class requests the deterministic-equivalent sign screen of Section 3.3 reads 12 true positives, 9 false positives, 81 true negatives, and 18 false negatives. That is sensitivity 0.400 [0.227, 0.594], a value inside the 69-request interval [0.158, 1.000], and specificity 0.90 [0.82, 0.95]. Balanced accuracy is 0.65 with class-stratified bootstrap interval [0.56, 0.74], and AUROC is 0.83 [0.76, 0.90], against 0.80 for a ranking on the deleted subject's baseline error alone. The 18 misses have median baseline error 0.18 against 1.11 for the 12 hits. The exact identity flags all 30 positives at its zero cut, sensitivity 1.000, with specificity 0.59, and ranks them with AUROC 0.998. Deleting the whole subject on the same windows leaves 221 of 249 requests at a structural zero and one LEAK, so no record-level positive is a subject-level positive. Figure A3 draws the 249 requests with the subject-level control on the same windows.

Two readings must be reported together. The original label rule treats the ten seeds' rows as independent. Window-clustered relabeling (Section 6.5), a bootstrap that resamples the distinct label windows instead, leaves 3 positives: three Normal rows of the healthy CWRU bearing, three distinct recordings, and the screen recovers all three. That is a record-level balanced accuracy of 0.985, the same value as Section 6.5 by coincidence, with interval [0.955, 1.000] on the 36 two-class requests that remain. Under the abstention band of Section 3.3 the screen abstains on 26 of the 30 positives and issues SAFE on 4, a mis-issued SAFE rate of 0.13 [0.04, 0.31]. On PTB-XL only 1 of the 13 record-level positives has a sibling that the 0.9 detector would flag, so a declared subject map finds these carriers and the descriptor does not. These are within-subject carriers; they are outside the subject-level estimand of Section 6.1 and raise nothing in it. On the 69 auxiliary requests the plug-in rule abstains on 68 and mis-issues no SAFE; the screen abstains on 28 and decides 41, one of them a leak, without error. Both auxiliary-set leaks are of this kind: a retained third of the deleted subject's own windows produces each label, and the nearest other unit stays below the de-centered threshold.

On a second encoder the same picture appears at series level. In the Chronos-Bolt CWRU cell the request unit is one series and the same bearing's other three records stay in the pool. There 12 of 39 series have a carrier and read a median excess of +0.90, range −0.28 to +5.88, 11 of 12 positive, against a pre-registered all-series median of 0.000. The pool's 2 HP Normal row is a copy of its 1 HP row, the pair at cosine 1.000 (Appendix F.1; Table J.2). The readings above drop the copy; the 1 HP row keeps its archived reading from a pool that held the copy, and without both rows the subset median is +0.76 over 11 series. On the record-level arm the same double removal leaves 29 of 248 requests LEAK, sensitivity 0.379 [0.207, 0.577], and 2 window-clustered positives on 35 two-class requests. Under the archive's record order the 12 series fall on four bearings and their largest similarities pair up within a bearing, so the flagged carriers are consistent with the same bearing's other records. The subset is outside the pre-registered estimand and decides no cell; it is the series-level counterpart of the record-level instance above, not the subject-level estimand of Section 6.1. Across this appendix the detector flags few natural carriers, the estimator resolves the injected effects of B.3 once a carrier is flagged, and declared within-subject carriers lie outside the subject-level estimand; Section 8 states where each reading stops.

C THE DETERMINISTIC-EQUIVALENT LAYER

This appendix states the closed form behind the deterministic-equivalent sign screen and explains why the injected-effect ladder cannot grade it.

C.1 THE CLOSED FORM AND THE TERMS OF SECTION 3.3

This section states the random-design deterministic equivalent that Section 6.5 evaluates with plug-in inputs and defines the isotropic factorization, the deviation from the full criterion, and the margin regime. Fix k carriers of feature-space correlation ρ to the deleted subject, subject leverage $\theta_x = h$, and typical pool-row leverage θ_u. Write $S = \mathbb{E}[\mu_X^2]$, $W = \mathbb{E}[p_b^2]$, and $C = \mathbb{E}[\mu_X p_b]$ for the second moments, $V_\zeta = \mathrm{Var}[u^\top w_b]$ for the background fluctuation, and σ^2 for the carrier label noise. The auxiliary quantities are

$$(\alpha, \beta) = \big(1 + (1-\rho^2)\,\theta_u,\ \rho^2\theta_x\big), \quad a = \frac{\rho\,\theta_x k}{\alpha + k\beta}, \quad G = \frac{(\rho\,\theta_x)^2 k}{(\alpha + k\beta)^2}, \quad \sigma_{\text{eff}}^2 = \sigma^2 + (1-\rho^2)\,V_\zeta.$$

The deterministic-equivalent excess is

$$\mathbb{E}[\Delta_k] \approx (W + S - 2C) - \big[(1-a\rho)^2 W + (a-1)^2 S + 2(1-a\rho)(a-1)\,C\big] - \sigma_{\text{eff}}^2 G.$$

It equals the pool-typical baseline error $W + S - 2C$ minus the post-carrier bias term and the noise term, and at $k = 1$ it reduces to the factored form behind Equation (2). The subject-anchored variant replaces the pool-typical baseline error by the deleted subject's own μ-scored baseline error B at fixed S. It shifts C to $C - (B - (W + S - 2C))/2$ and keeps the pool geometry θ_u and V_ζ. Section 6.5 reports the sign of this variant, with θ_x, B, σ^2, S, W, C, and ρ estimated on the 12 estimation windows and the pool. Its magnitude is rejected on real features in 9 of 9 pre-registered fits, so the screen is descriptive. The exact identity of Theorem 1 evaluated on the same windows with realized labels, the plug-in rule of Section 3.3 without an abstention band, flags 35 of 67 safe requests on the same set.

Terms of Section 3.3. The **isotropic factorization** is the $k = 1$ closed form above with an isotropic carrier component orthogonal to the deleted subject's direction (V_ζ evaluated as W). The **deviation from the full criterion** is $\mathcal{E} := \mathbb{E}[\Delta_1] - H$, the $k = 1$ deterministic-equivalent excess above minus the scalar rule value $H = \theta_x[(2 + \theta_x)B - \theta_x\sigma_{\text{eff}}^2]/(1 + \theta_x)^2$. The **margin regime** is the set of plug-in states with $|H| > |\mathcal{E}|$; there H and $\mathbb{E}[\Delta_1]$ share their sign, and every sign disagreement lies outside it.

A random-design deterministic equivalent describes how the collective effect varies with ρ and k under a proportional limit and a Gaussian design with bounded spectrum, up to an $O(1/d)$ remainder. On real frozen features the effective dimension is small and the magnitude prediction fails in the pre-registered fits above, so the layer explains sign structure, supplies the descriptive screen of Section 6.5, and never signs a deployment. The next section explains why the injected-effect ladder of Appendix B.3 cannot grade this layer.

C.2 WHY THE INJECTED-EFFECT LADDER CANNOT GRADE IT

This section explains why the injected-effect ladder and the deterministic-equivalent rules address different estimands, so the ladder neither validates nor refutes the rules. On the ladder, for each deleted subject and dose, we evaluate the three adjudication quantities of the pre-registered predicate suite. The inputs are the signing state, the retained pool without the deleted subject, and the injected unit's own features and labels. The quantities are the scalar margin $\log[(2 + h)B] - \log[h\sigma^2]$, the subject-anchored deterministic-equivalent excess, and the exact rank-k identity scored on the deleted subject's windows. The deterministic-equivalent excess is negative in every cell, with medians from -0.05 on CWRU to -1.09 on Paderborn. The measured realized change instead grows with φ to +0.52 on CWRU, +0.03 on Paderborn, +0.14 on PTB-XL, and +0.81 on MIMII at φ = 1. These ladder readings score the mixed unit at the subject level and are a different construction from the injected duplicate of Section 6.4 (Appendix F.4), so the two are not comparable. Sign agreement at φ of 0.5 and above is 0.20 to 0.38 for the deterministic equivalent and 0.23 to 0.39 for the scalar rule; Spearman rank correlation with the measurement is +0.35 to +0.54.

Section 3.3 and Appendix A.4 explain the divergence. The injected unit is a copy of the deleted subject's own windows, so its label noise is the deleted subject's realized noise; the realized excess therefore includes the term $2h_v^\top \Sigma_{c0}$ that the mean-scored excess excludes by construction. The rules predict the mean-scored quantity, which for these typical deleted subjects, whose baseline error B is small relative to the pool's held-out noise, is near zero or negative. The ladder and the rules therefore address different estimands, and the ladder cannot validate or refute the rules. It does show that ladders built from copies overstate mean-scored leakage, which is why the issuance premise of Appendix E requires scoring windows on the deleted subject's side to be disjoint from the retained pool's noise clusters. The forced-caliber excess of Appendix B.3 inherits the same term at high φ; the resolution statement of Section 5 uses the low-φ range, where the injected mixture shares little of the deleted subject's realized noise. The ladder therefore calibrates the estimator (Appendix B.3) and leaves the deterministic-equivalent layer to the natural-pool screens of Section 6.5.

D Related work and positioning

This appendix discusses four neighboring lines of work and then grades, paper by paper, what each establishes and what this paper adds (Appendix D.5).

D.1 Unlearning evaluation

Machine unlearning removes the influence of requested records from a trained model, exactly by retraining or approximately through certified updates (Cao & Yang, 2015; Bourtoule et al., 2021; Ginart et al., 2019; Guo et al., 2020; Sekhari et al., 2021; Neel et al., 2021). Its evaluation is record-level: an auditor compares the unlearned model with one that never trained on the record, usually through membership inference, and later work documents when such tests overstate removal (Shokri et al., 2017; Carlini et al., 2022; Zarifzadeh et al., 2024; Hayes et al., 2025; Thudi et al., 2022; Zhang et al., 2024). Retained records enter this literature as injected duplicates that confound unlearning and its verification (Ye et al., 2025), and through their own membership risk after unlearning, including range-level membership (Fu et al., 2026; Tao & Shokri, 2025). They also enter through retain-forget entanglement, where retained samples that share features with the forget set lose utility when it is unlearned (Cheng et al., 2026; Dolgova & Stich, 2026). We keep the record-level unlearning audit as our reference and run two membership-inference attacks on the same pools, but both compared models keep every retained record, which holds a retained carrier's reconstruction excess fixed. Those studies ask what happens to retained records, or what injected ones do to unlearning, whereas we quantify the reconstruction excess that a naturally retained record causes for the deleted subject. This measurement replaces neither retained-set membership inference nor duplicate-injection attacks, and exact-copy deduplication at intake stays an existing baseline.

D.2 Reconstruction and privacy auditing

Reconstruction attacks on unlearning recover a deleted point from a parameter update (Bertran et al., 2024; Lamri et al., 2026), extract deleted text with pre- and post-unlearning checkpoints (Wu et al., 2025), or exploit the difference between two releases (Chen et al., 2021; Salem et al., 2020; Zanella-Béguelin et al., 2020; Gao et al., 2022). Interdependent-privacy work shows that records of relatives or correlated individuals expose a person who shared nothing (Humbert et al., 2013; 2017; 2020; Liu et al., 2016; Xu & Strohmer, 2025), and deletions that depend on published models make records interdependent and constrain sound deletion (Chourasia & Shah, 2023). We share the reconstruction question and the premise that correlated data leak, but our analyst is post-only: it receives the deployed encoder, readout, and retained pool, and no earlier checkpoint or parameter difference. Any leakage through the readout must therefore come from a record that remains in the final deployment, and we quantify that record's reconstruction excess. Attribute-inference work shows that much apparent leakage equals population-level imputation (Jayaraman & Evans, 2022; Zhao et al., 2021), so we measure each carrier against a carrier-free baseline and a subject predictable from population structure alone has zero excess. These studies establish the risk; this paper adds the post-only reconstruction excess as an operational estimand, its exact identity, and the audit on time-series foundation models.

D.3 TIME-SERIES FOUNDATION MODELS AND FIXED-FEATURE READOUTS

Time-series foundation models are often deployed as frozen encoders with a small learned readout (Ansari et al., 2024a;b; Das et al., 2024; Woo et al., 2024; Goswami et al., 2024), a pattern that random convolutional kernels and reservoirs established earlier (Dempster et al., 2020; 2021; Jaeger & Haas, 2004). For a ridge readout over fixed features, deletion has a closed-form update, and exact unlearning of ridge heads on frozen foundation models has been derived (Guo et al., 2020; Izzo et al., 2021; Quan et al., 2026). Privacy studies of time-series models measure record- or user-level membership, memorization in imputation models, and pretraining contamination (Koren et al., 2025; Li et al., 2026b; Johansson et al., 2025; Taleb et al., 2026; Meyer et al., 2025; Li et al., 2026a), and time-series unlearning work designs the unlearning objective itself (Wang et al., 2026). We adopt the frozen-encoder deployment, keep the deletion exact, and reuse the rank-k algebra behind these updates. The algebra serves a different purpose here: an exact fixed-feature identity that attributes a deleted subject's reconstruction excess to a stated set of retained carriers. Those privacy studies ask whether a series or user was a training member, whereas we measure the reconstruction excess that a retained series causes for a deleted subject.

D.4 SIMILARITY-BASED DETECTION AND REPRESENTATION GEOMETRY

Cosine similarity between embeddings is the standard detector of semantic near duplicates (Abbas et al., 2023), and deduplication is known to reduce memorization and privacy risk (Lee et al., 2022; Kandpal et al., 2022). Learned embeddings are anisotropic: a shared mean vector, a few dominant directions, or rogue dimensions inflate cosine similarity, and mean removal, whitening, or standardization corrects it (Mu & Viswanath, 2018; Ethayarajh, 2019; Su et al., 2021; Timkey & van Schijndel, 2021; Parupudi, 2026). We use a cosine detector to find retained carriers, inherit mean removal from this literature, and apply no whitening. Our addition is the consequence for unlearning audits: on frozen time-series features the raw cosine detector marks almost every retained record as a carrier and inflates the apparent excess. De-centering removes the artifact, controlled carriers with known effects calibrate what the de-centered detector can resolve, and cosine detectors in other embedding spaces keep their validity.

D.5 POSITIONING MATRIX

Each row states what the cited work establishes and what this paper adds; the rows record proximity of threat model and estimand, and this paper claims no discovery that correlated records leak. What it claims is a post-only threat model, the reconstruction excess and its exact fixed-feature identity, a de-centered detector that avoids the spurious carriers of the shared feature direction, and membership and reconstruction as different estimands on the same pools.

Work	**What they establish**	**What we add**
Bertran et al. (2024)	Near-complete reconstruction of a deleted point in linear regression from the parameter or model update.	Post-only access with no earlier checkpoint or parameter difference; the reconstruction excess of retained carriers on frozen time-series foundation-model features; a subject-level estimand.
Wu et al. (2025)	Extraction of deleted data from a language model, guided by pre- and post-unlearning checkpoints or logits.	The post-only analyst of Section 2.1 has no history access; the object is the reconstruction excess of records retained in the final deployment; no language model and no checkpoint attack.

continued on next page

Work	What they establish	What we add
Ye et al. (2025)	Attacker-injected exact and near duplicates that confound unlearning and its verification and evade deduplication.	Natural retained carriers after a lawful deletion; the raw versus de-centered detector comparison; injected-duplicate results kept apart from natural-pool conclusions.
Chourasia & Shah (2023)	Records become interdependent when deletions depend on published models; retraining from scratch after a deletion can still expose the deleted record; a sound deletion guarantee.	The risk from retained correlated data becomes a measurable fixed-feature carrier coordinate, kept separate from the membership channel.
Tao & Shokri (2025)	A range-level membership game that captures risk from similar or partially overlapping data.	A reconstruction excess instead of range membership; carrier-free baseline, rank-k accounting, subject-level deletion.
Jayaraman & Evans (2022)	Attribute inference often matches population-level imputation; white-box access can add identification power.	We separate population predictability from the retained carrier's reconstruction excess.
Humbert et al. (2017)	Data of relatives or correlated individuals exposes people who shared nothing; the interdependent-privacy view.	The same intuition becomes a retained-pool reconstruction estimand on time-series foundation models, with a readout-level identity.
Cheng et al. (2026)	Retain-forget entanglement: retained samples correlated with the forget set suffer collateral damage from unlearning; a two-phase optimizer evaluated across datasets and architectures.	No change to the unlearning step; measurement of what retained carriers add to reconstruction of the deleted subject, with an exact rank-k identity and time-series evidence.
Fu et al. (2026)	Retained-set membership risk can rise after unlearning; TC-UMIA, a three-class attack that compares black-box predictions from before and after unlearning.	Membership and reconstruction as different estimands on the same pools; a post-only retained pool and a fixed readout carrier functional.
Abbas et al. (2023)	Embedding cosine similarity detects semantic duplicates at web scale.	Raw cosine on frozen time-series features saturates under the shared feature direction; a de-centered detector with a materiality gate.
Xu & Strohmer (2025)	Deleting an explicit feature leaves correlated information behind; an information-theoretic account of residual dependence.	Retained carriers on time-series foundation models, the rank-k readout identity, and the natural-pool and adaptation boundaries.
Lamri et al. (2026)	DRAUN: gradient and parameter reconstruction attacks on federated unlearning.	No federation, no gradients, a post-only retained pool.

continued on next page

Work	What they establish	What we add
Mu & Viswanath (2018); Ethayarajh (2019); Su et al. (2021); Timkey & van Schijndel (2021)	Learned embeddings are anisotropic: a shared mean vector, a few dominant directions, or a few rogue dimensions inflate cosine similarity; mean removal, whitening, and standardization correct it.	The consequence of this known artifact for unlearning audits: raw cosine on frozen time-series features produces spurious carriers and a spurious reconstruction excess; a de-centered detector with a materiality gate and a stated detection limit.

E CONDITIONAL REMOVAL AND THE ONE-COORDINATE CERTIFICATE

For a frozen readout and a detected carrier set, the analyst can remove the carriers, refit, and certify a one-coordinate budget ε; this appendix states the certificate's premises, its abstention rule, and its coverage in this archive. In this archive it covers no measured leak in the natural pools, abstains on every fine-tuned deployment, and leaves sub-threshold carriers and learned adversaries open.

E.1 WHAT THE CERTIFICATE SIGNS

For a fixed feature map, a ridge or kernel readout, a signed descriptor z_X from the evaluator, and a detected carrier set, the certificate states that the carrier-attributable excess left after carrier-aware removal is at most a budget ε. The post-only analyst uses the deployed readout, the retained pool, and z_X; the evaluator supplies only held-out aggregate scores. The certificate does not cover reconstruction from population structure, a learned adversary, carriers below the threshold, or an encoder that absorbed information about the deleted subject during adaptation.

E.2 THE PROCEDURE FAILS CLOSED

Carrier-aware removal has four steps. The evaluator freezes disjoint descriptor windows D_X and scoring windows E_X and issues z_X from D_X. The analyst detects a candidate set under the declared threshold, refits the readout without it, and computes the pool-side operators g, M, and q of Equation (1). The evaluator supplies e_b and the held-out score on E_X, so Δ_k is assembled without the raw deleted subject leaving the evaluator. The analyst stops when the collective score is within ε, or abstains. Issuance also requires four premises; each unverified premise becomes a named reason for abstention. Detection requires that no above-threshold unit remain and reports detector recall against the measured leak band even when it fails. Estimation requires held-out noise estimation on windows disjoint from the deleted subject's noise clusters and a minimum evaluator-held sample. The class premise requires a fixed-feature ridge or kernel readout and an attribution route for any adapted deployment. The archive premise checks the feature, window, label, protocol, subject-map, descriptor, and role and time-order hashes; the gate lists every failed conjunct.

E.3 THE BELOW-THRESHOLD CHANNEL

Carriers below the threshold lie outside the detected coordinate and need a separate price. We bound their excess with a leverage bound at most one, a displacement bound taken as the minimum of a weight-norm term, a two-constraint spectral family, and a rowwise Hölder term, and the archive-realized displacement. On the flagship CWRU cell with Chronos-t5-small and four archived deleted subjects, the leverage bound reads 1.0000 for detection thresholds between 0.9 and 0.99 while the realized leverage is 0.998 to 0.999. The bound is tight because the channel is open there, and the old scaled-identity bound overstated it by four to five orders of magnitude. At a detection threshold of 0.5 the displacement bound is 1.002 to 6.0 times the realized displacement, a non-vacuous price; at thresholds of 0.9 and above it is 200 to 700 times, the true worst case over the threshold set.

E.4 Adaptation makes attribution undecidable

Adaptation can move information about the deleted subject from the readout coordinate into the encoder. Write the total as a population part, an absorbed part, and a carrier part; the certificate covers the last. Two versions can agree on every archived quantity, including weights, retained data, absorption fraction, and collective score, and still differ between zero and maximal absorbed mass, so no predicate of the signing archive separates them. An adapted deployment must buy attribution with lineage of the encoder-training pool, a class-matched counterfactual re-adaptation arm, or a pre-registered fresh cohort, and abstain otherwise. On our fine-tuned arms the absorption gate fires on 69 of 72 deploy requests, and the class and member-state premises fail on all 30 fine-tuned requests of the intervention run.

E.5 Coverage in this archive

Section 6 supplies no cell with a material natural leak band, and the intervention run confirms it: across 84 requests on four corpora and one fine-tuned arm, no request reaches CERTIFY while covering measured leakage. On the flagship CWRU cell all 30 requests abstain, 17 on the archive premise and 13 on the archive premise plus a formal residual; on the fine-tuned arm all 30 abstain on class match and the member-state gate. On PTB-XL, MIMII, and Paderborn the procedure issues 13 certificates on requests whose measured band is empty, which the pre-registered rule classes as vacuous. It abstains on the other 11 for a formal residual, and the below-threshold ledger reads OPEN on every request, because a CERTIFY never closes an OPEN channel. Removal costs little utility, a median change of 0.000 and at most 0.047 in mean retained-subject error, and measured learned residuals of 0.17 to 0.995 stay in the ledger. We measure that learned-adversary channel but do not price it; the reasons and the calibration arithmetic are in Appendix F.6. The observed coverage is therefore an archive-level diagnostic, not a coverage probability or a whole-model guarantee.

F Protocol constants and supplementary results

This appendix records the constants that every run shares and the supplementary results that the main text cites by pointer.

F.1 Protocol constants and data flow

This section lists the fixed constants of every run, the data flow of the three-role interface, and the one data-provenance correction, the CWRU loader copy. Features come from Chronos-t5-small unless stated. The readout is a closed-form ridge with penalty 10^{-3}, context length 96, and one-step-ahead targets, with three seeds fixed before the runs. In the A-half protocol of the natural-pool cells each record contributes 32 superset windows, 24 for training and 8 held out, so the training mask does not depend on the seed. The seeds vary only the matched random control and the injected units, so the three seeds of an A-half cell are not repeated measurements of its excess.

The interface of Section 2.1 passes six quantities between the three roles; the table lists who produces each quantity, who holds it, and who refits or scores with it.

Quantity	**Produced by**	**Held by**	**Refit / scored by**
ψ, the deployed feature map	maintainer	maintainer, analyst	analyst refits nothing in ψ; frozen throughout
R, the retained pool with subject IDs and hashes	maintainer	maintainer, analyst	analyst refits the readout on R and on $R \setminus C$

continued on next page

Quantity	Produced by	Held by	Refit / scored by
w_R, the deployed readout	maintainer	maintainer, analyst	analyst hands w_R and the refit $w_{R\setminus C}$ to the evaluator
X, the raw deleted subject and its labels	requesting subject	evaluator only	evaluator scores X on w_R and $w_{R\setminus C}$ and returns the aggregate
μ_R, the retained-pool mean feature (de-centering anchor)	analyst, from ψ and R	analyst, evaluator	enters z_X; hash attested
z_X, the signed descriptor of the deleted subject	evaluator, from X and μ_R	evaluator, analyst	analyst detects C from z_X; evaluator supplies e_b and the held-out score for Δ_k

z_X is released to an auditing analyst under the assumption that the analyst never tries to invert it, and to no adversary. The CWRU loader copy, one recording that entered the archived CWRU pool twice, and the three pools it affects are in Table J.2.

F.2 The auxiliary sign check and the audit of its code

This section reports the sign check on the auxiliary request set and the code audit that rules out window-level circularity. The auxiliary set holds 69 auxiliary requests across five corpora, CWRU, Paderborn, PTB-XL, ELD, and traffic, with 2 leak and 67 safe labels measured independently. The undetermined requests stay out, and both leaks are auxiliary-corpus requests, one ELD meter and one traffic sensor. On this set the sign of the subject-anchored deterministic-equivalent excess, the Section 3.3 layer fed from the estimation windows, reaches specificity 0.970 with exact interval 0.896 to 0.996, sensitivity 1.000 with exact interval 0.158 to 1.000, and AUROC 1.000. The stratified bootstrap interval for the balanced accuracy of Section 6.5 is degenerate on the positive side, the exact intervals imply 0.53 to 1.00, and the AUROC interval is degenerate with two positives. Both false positives are CWRU bearings; the balanced accuracy rests on a 10-seed majority vote, and single seeds give 0.69 to 0.97.

Each deleted subject's 36 windows split into disjoint thirds of 12 for carriers, estimation, and labels, seeded per subject and per run. The label windows feed only the ground-truth labeler, the estimation windows feed only the screen's plug-in inputs, and the deterministic-equivalent screen reads no window-level error at all, so the auxiliary-set result has no window-level circularity. The confusion matrix of the screen reproduces from the per-subject scores as 2 true positives, 2 false positives, 65 true negatives, and 0 false negatives. The exact identity evaluated on the estimation windows with realized labels gives 2, 35, 32, and 0, specificity 0.478 with exact interval 0.354 to 0.603 and balanced accuracy 0.739. On the 120-request set the same rule reads specificity 0.589; Section 3.3 states the abstention band it needs. The scalar rule (2) gives 2, 4, 63, and 0, specificity 0.940 and balanced accuracy 0.970. Ranked by baseline error and cut at four flags, as the screen is, the auxiliary requests give the screen's own matrix, so the screen adds no separation over baseline error. The ranking of the identity and the deleted subject's baseline error alone both reach AUROC 1.000. The label intervals treat 120 rows drawn from 36 windows as independent, so some labels may move to undetermined under a window-clustered bootstrap; 51 of 120 requests were undetermined and excluded from the two-class set.

F.3 BOUNDARY LADDERS IN FULL

This section lists the four boundary ladders and the Paderborn pools in full. Table 1 lists the ladders; no ladder moves the reading off the floor, and Paderborn remains the uncertain positive case. An RBF kernel-ridge readout on the same pools gives median excess 0.000 on CWRU, PTB-XL, and MIMII and +0.041 on Paderborn, against the ridge reading of Section 6.1 in the same run. On PTB-XL it reconstructs the deleted patients' windows better than ridge, with predictability 0.29 against 0.07, and still finds no carrier effect. Across Chronos-t5 tiny, small, base, and large the three null corpora stay at 0.000; Paderborn medians range from +0.036 to +0.080 with upper bounds up to +0.101, so the null reading does not apply uniformly there. Enlarging the PTB-XL pool to 200 patients raises the fraction of patients with a carrier to 0.135 for the small model and 0.215 for the base model, so 27 to 43 patients carry one. Every PTB-XL median stays at 0.000 with interval 0.000 to 0.000. Five sequential subject deletions show no clear accumulation: the step-2-to-5 increment is 0.000 on three corpora and +0.008 with interval +0.001 to +0.022 on Paderborn.

For 137 of the 186 subject-seed points of the natural-pool runs the pool already predicts the subject to within its label noise, so by Equation (2) not even an exact duplicate can raise the mean-scored excess. The null of Section 6.1 is therefore first a statement about these pools' geometry.

The Paderborn medians and subject-bootstrap intervals of Section 6.1 carry cluster-bootstrap intervals of 0.000 to 0.156 on all 26 bearings and 0.000 to 0.127 on the original 22, each over the three damage-family clusters. On 32 bearings the subject interval is +0.014 to +0.071 and the cluster interval +0.000 to +0.072, both below 0.1. The cluster interval has 256 support points and its top atom is still one bearing's reading, so the cluster-level reading remains a sensitivity band. The original 22-, 26-, and 29-bearing readings are unchanged (Table 1); the 32-bearing pool adds the KB family and reads +0.061. The cluster resampling moves from 27 to 256 equiprobable atoms; the 97.5th percentile is carried by 19 atoms instead of one.

F.4 INJECTED DUPLICATE ON BOTH ARMS, AND NATURAL CARRIERS UNDER FINE-TUNING

This section gives the absolute levels of the injected duplicate under the frozen and fine-tuned encoders and the natural-carrier arm on Paderborn. The absolute levels are medians over the 18 pairs per corpus of the archived pair table `admissible_pairs_72.csv`, columns `excess_l0` and `excess_fullft`. The carrier is one $\varphi = 1$ duplicate of the deleted record with labels y_X plus noise of scale $\sigma_{\text{inj}} = 0.1$ (the archived environment file). The unit is the record, 40 records of 10 bearings on CWRU, so the 18 CWRU pairs come from 15 records and the 17 admissible pairs from 14 (the rule is dated by file timestamp). From the frozen readout (l0) to the fine-tuned encoder (fullft) the duplicate moves from +1.01 to +1.19 on CWRU, from +0.25 to +0.10 on PTB-XL, from +0.43 to +0.63 on MIMII, and from +0.02 to 0.00 on Paderborn. On CWRU 18 of 18 pairs sit above the material-effect threshold on the frozen arm and the range is +0.23 to +1.83; on Paderborn 17 of 18 adapted floors are non-positive. The paired change of Section 6.4 is the difference of these two arms. Over all 18 pairs per corpus CWRU gains +0.17 [+0.01, +0.39] and PTB-XL reads -0.15 [-0.27, -0.06]; the full set of 72 changes by -0.01 [-0.03, -0.00], and the 20 excluded pairs change by at most zero. Resampling by record widens the CWRU admissible interval to [-0.00, +0.57] and lifts the pooled admissible upper bound to +0.12. On CWRU, PTB-XL, and MIMII the duplicate already reads above the gate under the frozen readout, so the fine-tuning result is an increment on an open channel and says nothing about natural carriers. The natural-carrier arm below measures them under fine-tuning on Paderborn.

The natural-carrier arm of Section 6.4 was pre-registered before the run and measures natural carriers under full fine-tuning on Paderborn. Over the 15 deleted bearings with a carrier, the same subject-level de-centered readout on the fully fine-tuned encoder, 3 seeds, moves the reconstruction excess from the frozen readout's +0.08 to 0.00 with interval +0.00 to +0.01. The paired change over the 45 pairs and its pair-level interval are the headline of Section 6.4; the bearing-clustered interval is -0.26 to -0.05, so both intervals agree and the arm ends in the pre-registered SHRINK branch. Over all 22 deleted bearings, 7 without a carrier at a change of exactly 0 by construction, the frozen +0.06 falls to 0.00. The paired change is -0.05 with pair-level interval -0.06 to -0.03, while the bearing-clustered interval -0.08 to 0.00 touches zero. The deleted bearing's retained predictability rises from +0.82 to +0.95 over the same bearings, 11 of 15 upward. Because the adapted readout predicts every positive-scale deleted bearing at 0.95 to 0.97 with or without its carriers, part of

the shrinkage may be range compression of the estimator in the adapted space. The pre-registered encoder-level control, the same full retrain with each deleted bearing's carriers removed from the pool, 15 bearings and 3 seeds, lowers the deleted bearing's predictability by +0.01 with interval +0.01 to +0.12. That reading is positive, with the median under the gate and the upper end above it, so it is undetermined by the rule of Section 4.2. It exceeds the readout-level excess within the bearing by +0.01, pair-level interval +0.00 to +0.01 and bearing-clustered +0.00 to +0.22, on 12 of 15 bearings, so neither reading of the shrinkage is excluded. The upper end moves between +0.12 and +0.60 across resampling seeds; with the upper end above the gate, the pre-registered branch depends only on the median and the lower end. Re-identifying carriers in the adapted feature space leaves the carrier sets unchanged, Jaccard median 1.000, so the shrinkage is not a change of carrier set. A member-arm check reads the deleted bearing at +0.98 against +0.96 when its records enter the encoder's training set, a small separation.

F.5 Random control and descriptor windows

This section explains the count-matched random control's basis bias and the disjoint-window rerun of the descriptor. In the archived natural-pool runs the random arm removes as many retained units as the carrier set contains but keeps their held-out windows in the generic basis of the predictability score. The control arm therefore scores higher than the retain arm on 13 of 15 Paderborn deleted bearings with carriers, median +0.027 with interquartile range +0.006 to +0.064 on the 22-bearing pool. On the 26-bearing pool it scores higher on 15 of 16 deleted bearings, median +0.023. The net reading of +0.071 equals the excess of Section 6.1 plus this bias; the excess itself compares two fits on the same fixed basis and never passes through the control. In the same runs the descriptor is the mean feature of the deleted subject's training windows and the scored windows are those same windows; on PTB-XL and MIMII the descriptor averages all of the subject's windows, held-out included. Detection reads features only, never labels or errors, so this is a protocol mismatch with the interface of Section 4.2 and no label leakage. The rerun, pre-registered before the run on the same cached features, rebuilt the descriptor from the subject's held-out windows only, assert-checked disjoint from the scored windows. It changed the carrier set of 5 of 22 Paderborn bearings, 4 of 40 PTB-XL patients, 1 of 12 MIMII machines, and no CWRU bearing. The medians moved to +0.053 on the 22 bearings (+0.018 to +0.077) and to +0.050 on all 26 (0.000 to +0.064), and stayed at 0.000 with upper bound 0.000 elsewhere. Each move lies within the bootstrap resolution of Section 4.2, and every cell kept its category. A descriptor of the same size drawn from the training windows moved the 22-bearing median to +0.039, so descriptor size matters more than descriptor origin. Restricting the random control's basis to units it did not remove lowers the control arm by a median +0.014 per carrier bearing and leaves the 22-bearing net at +0.070.

F.6 Population-level qualifiers and the unpriced learned-adversary channel

This section states the qualifiers that population-level statements need and why we measure the learned-adversary channel without pricing it. Population-level statements need their own qualifiers: in the affine second-moment attacker class the population floor has a distribution-free identity, for all measurable attackers a positive floor needs a declared regularity premise, and beyond those classes 'no reconstructor' stays open. The ethics statement records how we handle the per-patient PTB-XL recordings and the potentially harmful reading of a leak result. We measure the learned-adversary channel but do not price it, because the archive lacks a replay-verifiable attacker suite, a same-coordinate sham floor, and a pre-registered exchangeable cohort. With m calibration values the third-largest reading has level $3/(m+1)$, above five percent at the archived cohort size, so we quote no safety factor.

F.7 Scale amplification at negative predictability

This section explains the two MIMII carrier readings above the gate as relative effects amplified by a loss ratio above one. The two MIMII machines with a carrier at 0.9, `slider_id00` and `valve_id04`, read three-seed mean excess +0.192 and +0.281 at predictability -0.25 and -0.65. With predictability defined as $(L_{\mathrm{gen}} - L_{\mathrm{tgt}})/L_{\mathrm{gen}}$, a negative value means the deleted subject's loss L_{tgt} exceeds the generic loss L_{gen}. The excess equals a relative effect times the ratio $L_{\mathrm{tgt}}/L_{\mathrm{gen}}$, which then exceeds 1, so the readings are real relative effects amplified by that scale, not numerical

artifacts. The pre-registered median is stable to such points, the mean is not: the MIMII all-machine mean is +0.039 against an all-subject median of zero. The same reading applies to the 6 of 15 Paderborn carrier bearings with negative predictability (Table 1). Every Paderborn carrier reading above 0.1 comes from a bearing with negative predictability; the carrier bearings with non-negative predictability read a median of +0.058 on the 22-bearing pool and +0.060 on all 26, close to the all-subject medians.

G EVIDENCE AND BOUNDARY TABLES

This appendix collects the evidence table, the boundary table, and the boundary-arm summary that the main text cites. Table 1 is referenced from Section 6.3, Table 2 from Sections 5.2 and 8, and Table 3 from Sections 6.4 and 6.5. Table 1 (continued) continues Table 1 with the membership-channel, sign-check, boundary-ladder, and follow-up rows.

Table 1: Evidence table: request unit and adaptation; the membership channel, the sign check, the boundary ladders, and the pre-registered follow-up runs for the same cells are in Table 1 (continued). Request unit: subject (bearing, patient, machine) in the natural-pool rows and series in the eight cross-family cells (Chronos-Bolt, TimesFM-2.5; ELD is a series set); the adaptation block deletes one record per request and injects one $\varphi = 1$ duplicate of it (record-level units). Protocol tier: frozen readout, de-centered detector at 0.9, unless the row says full fine-tuning. Estimand: reconstruction excess in predictability units; all-subject estimate = median over deletion units of the three-seed mean with its 95% median-bootstrap interval; n = rows / distinct targets; cluster interval = bootstrap over the three damage-family clusters. Carrier-conditioned: readings over the units that have a carrier, the secondary estimand outside the pre-registered all-subject estimand, deciding no cell; readings at negative retained predictability are scale amplifications of a real relative effect. ZERO(structural): no unit has a carrier; ZERO(structural-dominated): fewer than half of the units have a carrier, so the all-unit median is zero by construction; ZERO(mixed): at least half have a carrier and the median still reads zero. [+0.000, +0.000] marks a structural zero (fewer than half the subjects have a detected carrier, so the median and its bootstrap percentiles are zero by construction); it is not a confidence interval.

Cell	n	All-subject estimate [95%]	Carrier-conditioned	State
Request unit: reconstruction excess				
CWRU (10 bearings)	30 / 10	+0.000 [+0.000, +0.000]	—	ZERO(structural), 0/10 bearings with carrier
PTB-XL (40 patients)	120 / 40	+0.000 [+0.000, +0.000]	4 patients: -0.02 to +0.06, median +0.017, none above 0.1 (measured near-zero)	ZERO(structural-dominated), 4/40 patients with carrier
MIMII (12 machines)	36 / 12	+0.000 [+0.000, +0.000]	2 machines: +0.19 and +0.28, above the material-effect threshold, at retained predictability -0.25 / -0.65 (scale amplification); Chronos-t5-base 0/12 with carrier	ZERO(structural-dominated), 2/12 machines with carrier
Paderborn, original 22 bearings	66 / 22	+0.056 [+0.018, +0.087]	15 bearings: median +0.084, range +0.035 to +5.30; 6 at negative retained predictability (scale amplification)	below the material-effect threshold; 15/22 with carrier; cluster interval [0.000, 0.127]
Paderborn, all 26 bearings	78 / 26	+0.046 [+0.000, +0.066]	16 bearings: median +0.066, range +0.032 to +5.08; 6 at negative retained predictability (scale amplification)	interval includes 0; 16/26 with carrier; cluster interval [0.000, 0.156]
Paderborn, 29 bearings (three healthy added)	87 / 29	+0.056 [+0.012, +0.070]; cluster [+0.005, +0.084]	K004 +0.048, K005 +0.056, K006 +1.20 (scale amplification)	below the material-effect threshold; K006 at negative retained predictability
Paderborn 22, count-matched random control	66 / 22	net +0.071 (excess +0.056)	—	control basis keeps removed units' held-out windows: bias median +0.027 on 13/15 carrier bearings; control-free excess is the conservative reading

continued on next page

continued from the previous page

Cell	n	All-subject estimate [95%]	Carrier-conditioned	State
Chronos-Bolt (Bolt) and TimesFM-2.5 (TFM) × CWRU, Paderborn, PTB-XL, ELD	8 cells (series)	8/8 medians 0.000, upper bounds $\leq$ 0.000; lowest lower bound -0.026 (TFM/CWRU)	medians (n): Bolt/CWRU +0.90 (12; 92 percent positive), Bolt/ELD +2.62 (4), Bolt/Paderborn -0.09 (13), TFM/CWRU -0.17 (21), TFM/ELD +0.02 (15), TFM/Paderborn -0.60 (11), TFM/PTB-XL -0.04 (5); Bolt/PTB-XL 0/60	ZERO in 8/8: structural Bolt/PTB-XL (0/60); structural-dominated 6 cells (7 to 31 percent with carrier), among them Bolt/CWRU (one series scored against its own copy); mixed TFM/CWRU (21/39); raw-cosine layer 4/8 material, up to +18.27
Adaptation (full fine-tuning; injected $\varphi = 1$ duplicate of the deleted record; record-level units)				
Absolute level of the duplicate's excess, frozen readout → full fine-tuning (median of 18 pairs per corpus)	18 per corpus	CWRU +1.01 → +1.19; PTB-XL +0.25 → +0.10; MIMII +0.43 → +0.63; Paderborn +0.02 → 0.00	—	the frozen readout already reads the duplicate on CWRU (18/18 pairs above 0.1); the paired changes below are increments on an open channel
Pooled Δ_{adapt}: 52 admissible pairs ‖ all 72 pairs	52 / 38 ‖ 72 / 55	+0.01 [-0.05, +0.09] ‖ -0.01 [-0.03, -0.00]	—	below the material-effect threshold; 52-pair record-clustered (by record) upper bound +0.12
CWRU: 17 admissible pairs (14 records of 10 bearings) ‖ all 18	17 / 14 ‖ 18 / 15	+0.20 [+0.02, +0.39] ‖ +0.17 [+0.01, +0.39]	—	OPEN at the pair level; record-clustered (by record) [-0.00, +0.57] touches zero
PTB-XL 16 ‖ 18; MIMII 18; Paderborn	16 ‖ 18; 18; 1 of 18	PTB-XL -0.09 [-0.29, -0.04] ‖ -0.15 [-0.27, -0.06]; MIMII +0.07 [-0.18, +0.23]	—	PTB-XL NEGATIVE (reduced, still open); MIMII undetermined; Paderborn no valid control for the injected duplicate (17/18 adapted floors $\leq$ 0; natural-carrier arm in the next row); absorption gate fires on 69/72 deploy requests
Paderborn, natural carriers under full fine-tuning (no injection; 15 carrier bearings)	45 / 15	paired Δ -0.07 [-0.11, -0.06]; bearing-clustered [-0.26, -0.05]	frozen readout → full fine-tuning: +0.08 → +0.00 (15 carrier bearings); +0.06 → +0.00 (all 22)	the natural carriers' reconstruction excess shrinks (both intervals below 0); Section 6.4, Appendix F.4

Table 1 (continued): Evidence table, continued: membership channel, sign check, boundary ladders, and the pre-registered follow-up runs for the cells of Table 1; conventions and state words as there; carrier-conditioned readings, where a row has them, are appended to its state.

Cell	n	Estimate [95%]	State
Membership channel (same cells)			
Subject-level LiRA / RMIA AUC, 16 shadow readouts, range over 3 seeds	8 of 16 cells	CWRU LiRA 0.978–1.000 / RMIA 0.970–0.997; Paderborn LiRA 0.714–0.774 / RMIA 0.930–0.952; PTB-XL LiRA 0.844–0.894 / RMIA 0.977–0.984; MIMII LiRA 0.960–0.996 / RMIA 0.991–0.995	OPEN; all 16 lower bounds (subject and record level) > 0.5, min 0.564
Chronos-Bolt / TimesFM-2.5 series-level LiRA, RMIA (3-seed mean per cell)	32 cells	AUC 0.951–1.000	OPEN; 32/32 reportable
Sign check (auxiliary 69-request set, 5 corpora; record-level 120-request two-class set, 3 corpora)			
Deterministic-equivalent sign screen on held-out estimation windows (Section 3.3 layer, descriptive); 2 LEAK / 67 SAFE; 51 undetermined excluded	69 / 69	tp 2, fp 2, tn 65, fn 0; specificity 0.970 [0.896, 0.996]; sensitivity 1.000 [0.158, 1.000]	descriptive; both false positives are CWRU
Balanced accuracy (10-seed majority vote) and AUROC	69 / 69	BA 0.985 (class-stratified interval degenerate on the positive side; the two exact intervals imply 0.53 to 1.00); single seeds 0.69–0.97; AUROC 1.000 (interval degenerate with two positives); always-LEAK BA 0.500	descriptive; no population sign accuracy claimed
Same screen on 120 record-level two-class requests, within-subject sibling carriers, Appendix B.6; 30 LEAK / 90 SAFE; 129 undetermined excluded	120 / 120	tp 12, fp 9, tn 81, fn 18; sensitivity 0.400 [0.227, 0.594]; specificity 0.900 [0.819, 0.953]; BA 0.650 [0.556, 0.744], class-stratified; the two exact intervals imply 0.52 to 0.77; AUROC 0.834 [0.758, 0.902]	descriptive; record-level estimand outside Section 6.1; sensitivity interval excludes 1; 18 misses at median baseline error 0.18 against 1.11 for the 12 hits

continued on next page

continued from the previous page

Cell	n	Estimate [95%]	State
Window-clustered relabeling of the same 249 requests, distinct label windows as the resampling unit	36 / 36	3 LEAK / 33 SAFE; tp 3, fp 1, tn 32, fn 0; sensitivity 1.000 [0.292, 1.000]; specificity 0.970 [0.842, 0.999]; BA 0.985 [0.955, 1.000]; AUROC 1.000 (bootstrap interval degenerate with three positives)	descriptive; 27 of 30 positives become undetermined; both readings belong together
Boundaries (frozen readout)			
RBF kernel-ridge readout	4 corpora	CWRU, PTB-XL, MIMII 0.000; Paderborn +0.041 [+0.005, +0.050]	no ladder above the material-effect threshold; paired KRR−ridge on Paderborn -0.008 [-0.037, -0.001]
Capacity: Chronos-t5 tiny / small / base / large	16 cells	three null corpora 0.000 in 12/12 cells; Paderborn tiny +0.080 [+0.000, +0.101], small +0.056 [+0.018, +0.087], base +0.036 [+0.000, +0.050], large +0.051 [+0.000, +0.084]	no material cell; CWRU structural at small/base/large (prevalence 0)
PTB-XL pool 40 / 100 / 200 patients, small and base	6 cells	all medians 0.000 [0.000, 0.000]	ZERO(structural-dominated): with carrier 0.100–0.215 (small 4/40, 14/100, 27/200; base 5/40, 21/100, 43/200); carrier-conditioned: carrier patients (100 / 200): medians +0.002 (small 100), +0.002 (small 200), +0.007 (base 100), +0.001 (base 200); at most 14 percent above 0.1
Five sequential subject deletions, Δ_{seq} steps 2–5	12 per corpus	CWRU +0.000 [-0.034, +0.000]; Paderborn +0.008 [+0.001, +0.022]; PTB-XL 0.000; MIMII +0.000 [+0.000, +0.077]	no accumulation above the material-effect threshold
Pre-registered follow-up runs (returned; same protocol, de-centered caliber)			
Non-TSFM feature family: MiniRocket, 504 PPV features, no learning; same windows, threshold, ridge, and seeds	4 corpora: 10 / 22 / 40 / 12 units	CWRU 0.000 [0.000, 0.000]; Paderborn +0.102 [0.000, +0.139]; PTB-XL 0.000 [0.000, 0.000]; MIMII 0.000 [0.000, +0.057]	not TSFM-specific: raw-cosine prevalence 1.000, de-centered 0.000 / 0.682 / 0.225 / 0.333; ZERO(structural) on CWRU (0 of 10 with carrier); ZERO(structural-dominated) on PTB-XL (9 of 40; carrier-conditioned median +0.031) and MIMII (4 of 12; carrier-conditioned +0.205); Paderborn undetermined (15 of 22; carrier-conditioned +0.133, 11 above 0.1)
PTB-XL pool 1000 patients (2514 records), Chronos-t5-small	1000 patients	0.000 [0.000, 0.000]	ZERO(structural-dominated): 180 of 1000 with a carrier (0.180); carrier-conditioned median +0.001 (-0.053 to +0.511), 6 patients above 0.1; ladder 40 / 100 / 200 / 1000 patients all 0.000, with carrier 4 / 14 / 27 / 180
Pre-registered positive control: the deleted subject's own later windows, $\geq$ 970 samples from descriptor and scoring windows, as the retained unit	CWRU 10 bearings / MIMII 12 machines / Paderborn 22 bearings / PTB-XL 40 patients; 3 seeds	CWRU detection 0.80 (Wilson [0.49, 0.94]; 24 of 30 rows), detected-caliber excess -0.003 [-0.017, +0.053]; MIMII detection 0.11 (4 of 36 rows), excess +0.000 [+0.000, +0.000]; pre-registered extension: Paderborn detection 0.68 (Wilson [0.47, 0.84]; 66 rows), the unit's own contribution +0.01 [+0.00, +0.02] at the detection caliber and +0.01 [+0.00, +0.02] forced; PTB-XL detection 0.03 (Wilson [0.00, 0.13]; 120 rows), +0.00 [+0.00, +0.00] and +0.01 [-0.00, +0.06]	CWRU PC-DETECT-ONLY (noise-independence caveat: innovation autocorrelation at the gap lag above 0.1 on 22 of 40 records): the detector sees the unit and the estimator reads a measured near-zero. MIMII PC-BLIND: the detector misses the unit, a structural zero on 32 of 36 rows. Paderborn 22 / 26 PC-DETECT-ONLY, descriptive (archived run). Pre-registered extension: Paderborn PC-DETECT-ONLY, the detector sees the unit and its own contribution stays below the gate, the same shape as on CWRU; PTB-XL PC-BLIND, the detector misses the unit
Pre-registered shared-process control: one retained unit sharing only the deleted subject's AR dynamics, a new realization under fresh white noise (24 synthetic AR(4) subjects; pole radii 0.40 / 0.70 / 0.90; noise ratio 0.25)	24 subjects; 3 seeds; 72 rows per tier	at r = 0.90: detection 0.514 (37 of 72 rows), median similarity 0.903; detected-caliber excess +0.0000 [-0.0009, +0.0137]; forced +0.0140 [-0.0175, +0.0242]; independent unit detection 0.028 (2 of 72), forced +0.0049 [-0.0214, +0.0326]	B-DETECTED-NONMATERIAL with a SPECIFIC negative arm: the detector registers the unit and the estimator reads no material excess at either caliber; detection 0.000 / 0.222 / 0.514 across r = 0.40 / 0.70 / 0.90 (descriptive); no analytic truth for this unit
Pre-registered encoder-level control: the same full fine-tuning with the deleted bearing's carriers removed from the pool (Paderborn, no injection; 15 carrier bearings)	45 / 15; 3 seeds	encoder-level excess +0.01 [+0.01, +0.12]; difference to the readout level within target +0.01, pair-level [+0.00, +0.01], bearing-clustered [+0.00, +0.22]; frozen readout +0.08 $\rightarrow$ readout level +0.00 $\rightarrow$ encoder level +0.01	E-POSITIVE-NONMATERIAL: positive; median under the material-effect threshold, upper bound above it, undetermined by the rule of Section 4.2; above the readout level on 12 of 15 bearings; neither reading of the shrinkage is excluded; Section 6.4, Appendix F.4

continued on next page

continued from the previous page

Cell	n	Estimate [95%]	State
Natural pool with descriptor ∩ scored windows = ∅ (descriptor from held-out windows only); control basis without removed units	5 cells: 10 / 22 / 40 / 12 / 26 units	archived → disjoint: Paderborn 22 +0.056 [+0.018, +0.087] → +0.053 [+0.018, +0.077]; Paderborn 26 +0.046 [0.000, +0.066] → +0.050 [0.000, +0.064]; CWRU, PTB-XL, MIMII 0.000 → 0.000	every cell keeps its verdict class; largest shift 0.0035 < 0.012 resolution; carrier sets changed for 5 of 22, 4 of 40, 1 of 12, 0 of 10 units; control basis without removed units: control arm lower by a median +0.014 per carrier bearing, 22-bearing net +0.071 → +0.070
Pre-registered Paderborn extension: 32 bearings (KB family added: three combined-damage bearings), same protocol; cluster resampling over four damage-family clusters	96 / 32; 3 seeds	+0.06 [+0.01, +0.07]; cluster interval [+0.00, +0.07] over 256 support points; KB23 +0.15, KB24 +0.24 (scale amplification), KB27 +0.07	below gate; subject interval excludes 0; the cluster reading is a sensitivity band: its top atom is one bearing's reading and 19 atoms sit at or above the 97.5th percentile; the original 22 / 26 / 29 rows of Table 1 unchanged

Table 2: Boundary table. Columns are the four layers at which the paper makes a statement: the finite-sample identity, the random-design asymptotic, the empirical observation, and the open channel. Each cell carries a state word, CONDITIONAL, OPEN, or ABSTAIN, followed by the premise that fixes it; a dash means the layer makes no statement on that row. Both Paderborn cluster upper bounds are listed in the estimation row; the disjoint-window rerun (Appendix F.5) and the pre-registered positive control (Appendix B.4) are recorded there, and the row stays OPEN. The encoder-migration row reads the injected $\varphi = 1$ duplicate of the deleted record (record-level units, Section 6.4), an adaptation boundary; the natural-carrier arm of Section 6.4 is in Appendix F.4.

Channel	Finite-sample identity	Random-design asymptotic	Empirical observation	Open channel
Encoder migration (adaptation, absorption)	CONDITIONAL: Theorem 1 needs a fixed feature map; no rank-k update exists once the encoder moves (App. A.5 scope)	—	ABSTAIN: absorption gate fires on 69/72 deploy requests; certificate abstains on all 30 fine-tuned requests; for the injected $\varphi = 1$ duplicate (record-level units) the CWRU pair-level interval excludes 0 (+0.20 [+0.02, +0.39]) and the record-clustered interval touches 0 (-0.00 to +0.57)	OPEN: attribution undecidable from the signing archive; needs lineage, a class-matched counterfactual arm, or a fresh cohort (App. E.4)
Learned adversaries	—: the identity bounds no learned reconstructor	—	CONDITIONAL: measured but not priced; learned residuals 0.17 to 0.995 stay in the ledger; with m calibration values the third-largest reading has level $3/(m+1)$, so no safety factor is quoted (App. F.6)	OPEN: no replay-verifiable attacker suite, same-coordinate sham floor, or pre-registered exchangeable cohort in the archive
Population inference	CONDITIONAL: affine second-moment attacker class has a distribution-free floor identity (App. F.6)	CONDITIONAL: a positive floor for all measurable attackers needs a declared regularity premise	CONDITIONAL: on Paderborn the population share of a deleted bearing's predictability is 0.942 against a carrier share of 0.058	OPEN: beyond those classes "no reconstructor" stays open; population statements need their own qualifiers (Section 8; Appendix F.6)
Below-threshold leakage (carriers below the threshold)	CONDITIONAL: leverage bound ≤ 1 and a displacement bound taken as the minimum of three terms (App. E.3)	—	CONDITIONAL at cap 0.5 (bound / realized 1.002 to 6.0); OPEN at caps ≥ 0.9 (200 to 700×; realized leverage 0.998 to 0.999 against a bound of 1.0000), flagship CWRU cell	OPEN in the high-correlation band; below-threshold ledger reads OPEN on every below-threshold request
Estimation (plug-in inputs, resolution, windows)	CONDITIONAL: exact given (e_b, g, M, σ^2); the abstention band widens by the plug-in confidence half-width (Section 3.3)	ABSTAIN: the random-design layer signs no verdict; its magnitude prediction is rejected in 9 of 9 pre-registered fits	CONDITIONAL: forced-caliber reading about one ninth of a constructed truth (synthetic-truth control, Appendix B.4); injected-ladder resolution about 0.03 (φ 0.35 to 0.5); Paderborn +0.056 [+0.018, +0.087] on 22 / +0.046 [0.000, +0.066] on 26; cluster upper bounds 0.127 (22) and 0.156 (26), each the three-member healthy cluster's median; random-control basis bias +0.027 (net +0.071 vs. excess +0.056)	OPEN: descriptor = scored windows in the archived Section 6 runs (Section 4.2); the disjoint-window rerun moved no headline median by more than the 0.012 bootstrap resolution (Appendix F.5); the pre-registered positive control reads PC-DETECT-ONLY on CWRU, PC-BLIND on MIMII, PC-DETECT-ONLY on Paderborn, PC-BLIND on PTB-XL, PC-BLIND-SYNTH on the constructed carrier, and B-DETECTED-NONMATERIAL on the shared-process unit (Appendix B.4)
Legal and normative scope	—	—	CONDITIONAL: a positive increment is a measurement about an estimand; it establishes no legal noncompliance, shows no failure of record deletion, and authorizes no removal of another subject's data (Section 8)	OPEN: legal interpretation is outside the paper

Table 3: Fine-tuning moves the excess in a corpus-dependent direction while the pooled change stays below the threshold, and the sign screen recovers the record-level labels only in part; both arms are boundaries of the measurement. Paired change in reconstruction excess under full fine-tuning of the encoder (pairs admitted by the predicted-floor rule unless stated) and the agreement of the deterministic-equivalent sign screen with the record-level labels; brackets are bootstrap intervals and the last column repeats the verdict words of the text.

Arm	Pairs or requests	Change or reading [interval]	Verdict
Full fine-tuning, admissible pairs (Appendix F.4)	52 of 72 (18 per corpus)	+0.01 [−0.05, +0.09]	pooled median below the threshold
CWRU, admissible pairs	17	+0.20 [+0.02, +0.39]; record-resampled [−0.00, +0.57]	gain excludes zero at the pair level only
PTB-XL, admissible pairs	16	−0.09 [−0.29, −0.04]	reduction that does not reach zero
Paderborn, natural carriers	—	−0.07 [−0.11, −0.06]; deleted bearing's predictability rises from +0.82 to +0.95	shrinkage at the pair level
All pairs	72	−0.01 [−0.03, −0.00]	—
Excluded pairs	20	—	change by at most zero
Record-level sign screen (Appendix B.6)	249; LEAK on 30 (original rule) and on 3 (relabeled); subject-level control exact zero by construction on 221	sensitivity 0.400 [0.227, 0.594], specificity 0.900 [0.819, 0.953], balanced accuracy 0.650	recovers all three relabeled positives; calls 4 original-rule leaks SAFE under its abstention band
Auxiliary sign check (Appendix F.2)	69; 2 leaks	balanced accuracy 0.985, matching the baseline error	consistency evidence, not a population guarantee

H APPENDIX FIGURES

This appendix holds the seven figures that the calibration and results sections cite, each caption opening with the figure's conclusion. Figures A1 and A2 carry one legend per panel, printed above, inside or below it; the light grey band marks readings below the pre-registered threshold 0.1. Figure A1 accompanies Sections 6.2 and 6.3 and the injected-effect ladder of Section 5.1; Figure A2 accompanies Section 6.4 and Appendix E; Figure A3 accompanies Section 6.5 and Appendix B.6. Figure A4 accompanies Section 6.1 and Appendix F.3; Figures A5 and A6 accompany the positive controls of Section 5.2 and Appendix B.4, and Figure A7, the controlled-carrier calibration of Section 5.2, closes the figure sequence. Box 1, cited from Section 9, is printed at the end of this appendix.

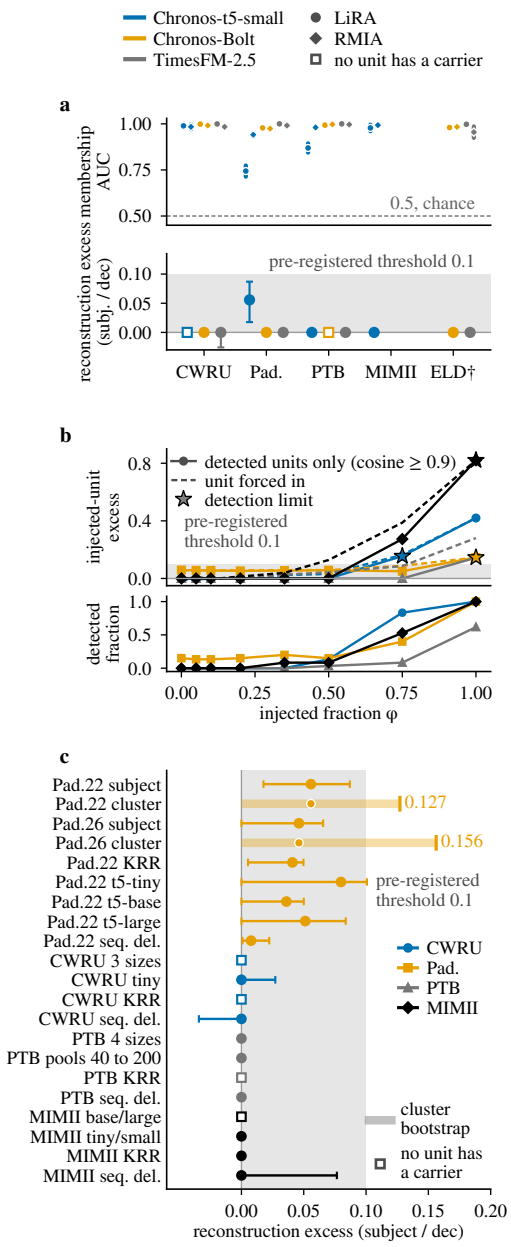


Figure A1: Membership stays open while reconstruction excess sits below the threshold in the same corpora. (a) Subject-level membership AUC (LiRA, RMIA; three-seed range) above the de-centered excess. (b) Injected-effect ladder: median excess of one injected unit, detected only and forced in; stars, detection limit; lower strip, fraction detected. (c) Boundary ladders; each Paderborn cluster bound equals the healthy cluster's median.

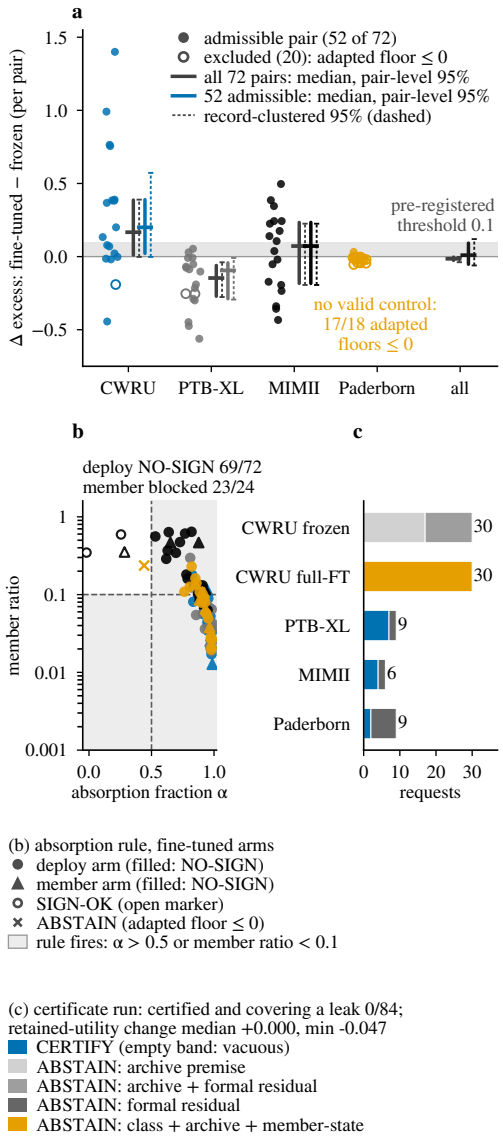


Figure A2: Full fine-tuning moves the injected duplicate's excess in a corpus-dependent direction; pooled, the change stays below the threshold. (a) Paired change for 72 record-level requests, 52 admissible (filled); pair-level and record-clustered 95% intervals. (b) Absorption rule on the fine-tuned arms, deploy and member requests. (c) Certificate run: decisions on 84 requests of five legs.

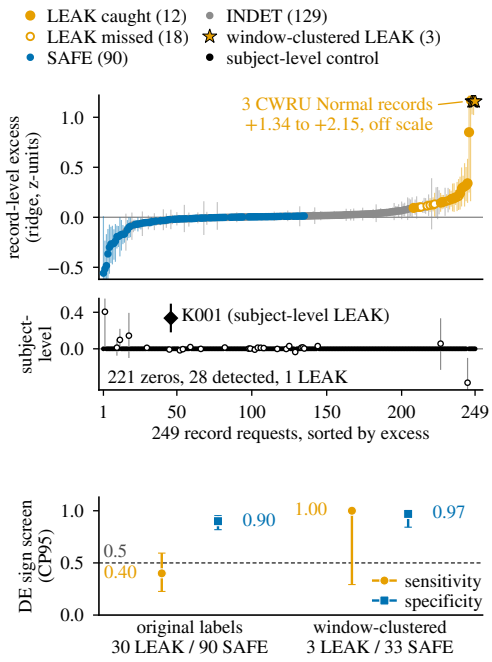


Figure A3: Record-level deletion with within-subject carriers reads LEAK on 30 of 249 requests, boundary evidence outside the subject-level estimand. Top: requests sorted by excess with the original labels; stars, the three window-clustered LEAK, off scale. Middle: subject-level control on the same requests. Bottom: sensitivity and specificity of the sign screen under both labelings, exact 95% intervals.

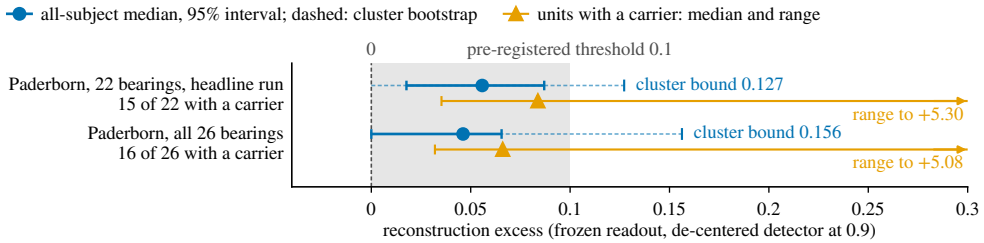


Figure A4: Both Paderborn cohorts read a small all-subject excess and a much wider spread over the bearings with a carrier. Rows: the 22 bearings of the headline run and all 26; predictability units. Blue: all-subject median, 95% interval, cluster-bootstrap bound printed. Orange: median and range over the bearings with a carrier; the range end is arrowed.

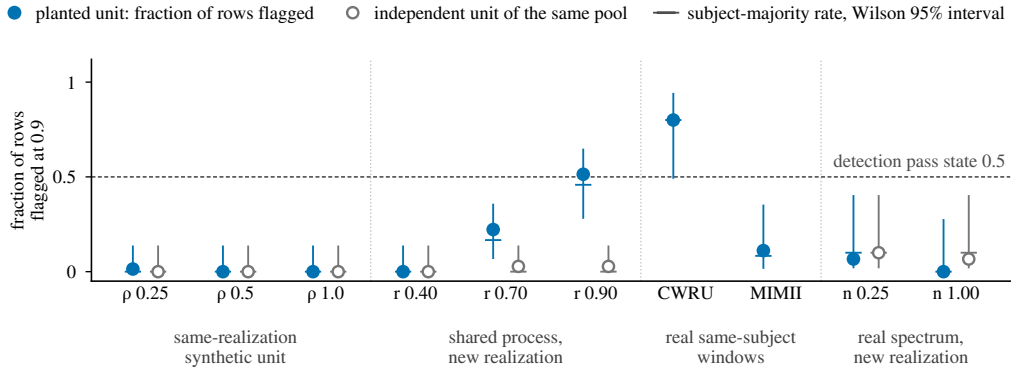


Figure A5: The detector flags the planted unit in most rows only for the real same-subject windows of CWRU and the strongest shared-process unit. Ten planted units and their independent counterparts; markers, fraction of rows flagged at cosine 0.9; tick and line, subject-majority rate with Wilson 95% interval; dashed, the pre-registered detection pass state 0.5.

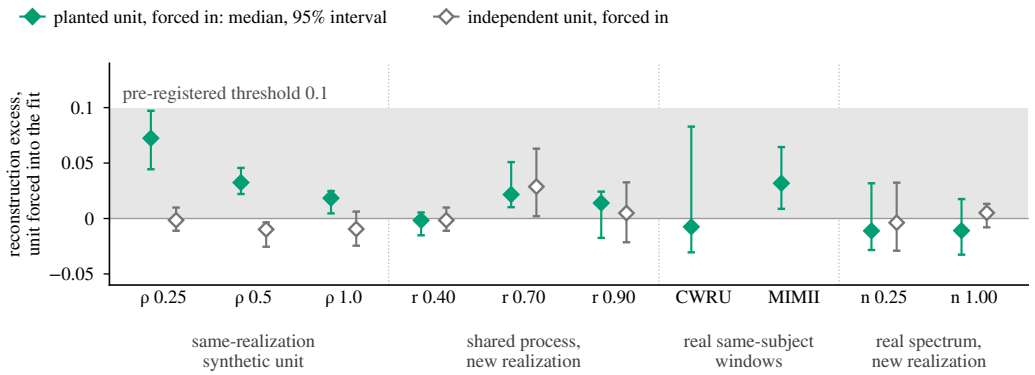


Figure A6: Forced into the fit, the planted units read below 0.1 on every control, and the independent units read near zero. Same ten units as the detector figure, predictability units; green, planted unit; grey hollow, independent unit; median and 95% interval; grey band, below the pre-registered threshold 0.1.

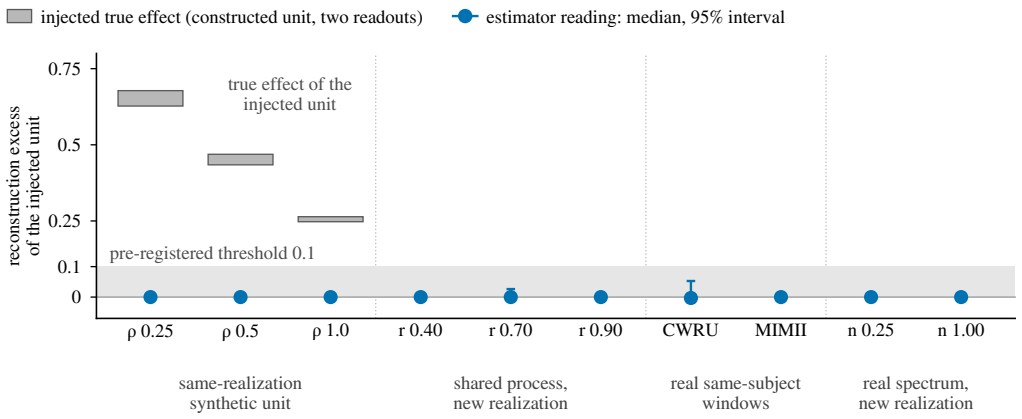


Figure A7: The estimator reads at or near zero on every planted unit, including the constructed units whose true effect lies between 0.25 and 0.68. Ten planted units, predictability units; grey blocks, the true effect where a construction defines one; blue, median reading with 95% interval; grey band, below the pre-registered threshold 0.1 that a passing control must exceed. Detector side: Figures A5 and A6.

Box 1. What a post-only analyst should do.

- Report both channels for the same pool and deletion units: a membership score and a reconstruction excess.
- Delete and audit at the subject level: the bearing, the patient, the machine.
- Detect carriers with the de-centered detector, state its threshold and detection limit, and count carrier-free subjects apart from those whose carrier measured zero.
- Give the deleted subject to an independent evaluator and receive only its signed descriptor.
- Abstain without the signed descriptor, archive hash, or role and time-order attestation.
- State the encoder regime: a frozen readout and a fine-tuned encoder are different measurements.

I THE RAW-COSINE ARTIFACT AND ITS DE-CENTERING CORRECTION

Raw cosine on frozen features is an artifact of a shared component. Write each descriptor as $m_u = \mu + e_u$ with $\gamma = \|\mu\|^2/\mathbb{E}\|e_u\|^2$. For any two units, sharing a source or not, $\cos(m_u, m_v) \geq 1 - 2\max(\|e_u\|, \|e_v\|)^2/\|\mu\|^2$, so the pair passes τ with probability at least $1 - 4/((1-\tau)\gamma)$ (Lemma A); pooling W windows raises γ by up to W. Subtracting the pool mean removes μ exactly; an independent pair then passes τ with probability of order $1/d_{\text{eff}}$ plus an anchor term of order $1/n$, while a re-observed unit still passes whenever its perturbation is below $\sqrt{1-\tau^2}$ of its de-centered residual (Lemma B).

Lemma A speaks to the flagging step only: the apparent excess reported under raw cosine comes from refitting on a pool depleted by the flags (Section 3.1), which the lemma does not model; whether the archived features sit in the large-γ regime is an empirical question that this appendix leaves open, and the synthetic checks behind it are not measurements.

I.1 SETTING AND A DISPLACEMENT LEMMA

Units $u \in \mathcal{U}$, $|\mathcal{U}| = n \geq 2$, are the retained units and the target. The descriptor of unit u is its mean window feature $m_u \in \mathbb{R}^d$ (Section 4.2); the de-centering anchor is the sample mean $\bar{m} = n^{-1}\sum_u m_u$ and $\tilde{e}_u := m_u - \bar{m}$. A pair (u, v) is flagged when its cosine is at least τ. We assume

(M1) $m_u = \mu + e_u$ with a common vector $\mu \neq 0$ and unit residuals e_u; $s^2 := \sup_u \mathbb{E}\|e_u\|^2$, $\gamma := \|\mu\|^2/s^2$, $r_u := \|e_u\|/\|\mu\|$.[1]

At the window level we write (M2) $F_{u,w} = \mu + b_u + \xi_{u,w}$, $w = 1, \ldots, W_u$, with a unit effect b_u and a within-unit term $\xi_{u,w}$, so that $m_u = W_u^{-1}\sum_w F_{u,w}$ satisfies (M1) with $e_u = b_u + W_u^{-1}\sum_w \xi_{u,w}$.

When the e_u are independent with mean zero and covariance Σ we write $s^2 = \operatorname{tr}\Sigma$, $d_{\text{eff}} := (\operatorname{tr}\Sigma)^2/\operatorname{tr}(\Sigma^2)$, $K_4 := \sup_u \operatorname{Var}(\|e_u\|^2)/s^4$ and $\sigma_\parallel^2 := \hat{\mu}^\top \Sigma \hat{\mu}$.

Lemma 0 (angle displacement). *Let $x \neq 0$ and $\|y\| \leq \|x\|$, with $x+y \neq 0$. Then $\langle x, x+y\rangle \geq 0$ and $\angle(x, x+y) \leq \arcsin(\|y\|/\|x\|)$, with equality iff $y \perp (x+y)$. Moreover, writing $y = \alpha x + y_\perp$ with $y_\perp \perp x$: if $1+\alpha > 0$, then $\cos\angle(x, x+y) \geq \tau$ iff $\|y_\perp\| \leq (1+\alpha)(\sqrt{1-\tau^2}/\tau)\|x\|$; if $1+\alpha \leq 0$, then $\cos\angle(x, x+y) \leq 0$. In particular $\|y\| \leq \sqrt{1-\tau^2}\|x\|$ implies $\cos\angle(x, x+y) \geq \tau$.*

Proof. Let $r = \|y\|/\|x\| \leq 1$, $a = \langle y, \hat{x}\rangle$ and $b = \|y_\perp\|$, so $a^2 + b^2 = r^2\|x\|^2$. Then $\langle x, x+y\rangle = \|x\|^2 + a\|x\| \geq (1-r)\|x\|^2 \geq 0$, so $\theta := \angle(x, x+y) \in [0, \pi/2]$ and $\sin\theta = b/\|x+y\|$ with $\|x+y\|^2 = (\|x\| + a)^2 + b^2$. The claim $\sin^2\theta \leq r^2$ is equivalent to $b^2(1-r^2) \leq r^2(\|x\| + a)^2$; substituting $b^2 = r^2\|x\|^2 - a^2$, the right side minus the left side equals $(a + r^2\|x\|)^2 \geq 0$, with equality iff $a = -r^2\|x\|$, i.e. $\langle y, x+y\rangle = a\|x\| + r^2\|x\|^2 = 0$. For the criterion, $x + y = (1+\alpha)x + y_\perp$, so for $1+\alpha > 0$, $\cos\angle(x, x+y) = (1+\alpha)\|x\|/((1+\alpha)^2\|x\|^2 + \|y_\perp\|^2)^{1/2}$, and squaring gives the equivalence; for $1+\alpha \leq 0$ the numerator is ≤ 0. The last sentence is the first part with $\arcsin(\sqrt{1-\tau^2}) = \arccos\tau$. ∎

I.2 RAW COSINE UNDER A COMMON COMPONENT

Lemma A (artifact). *Assume (M1). For every pair $u \neq v$ (with $m_u, m_v \neq 0$; $r_w < 1$ suffices):*

(i) $\cos\angle(m_u, m_v) \geq 1 - 2\max(r_u, r_v)^2$*; in particular the pair is flagged at τ whenever* $\max(r_u, r_v) \leq r_\tau := ((1-\tau)/2)^{1/2}$.

(ii) $\mathbb{E}\cos\angle(m_u, m_v) \geq 1 - 4/\gamma$.

(iii) $\mathbb{P}[\cos\angle(m_u, m_v) \geq \tau] \geq 1 - 4/((1-\tau)\gamma)$.

[1] Notation local to this appendix: μ is the common vector of (M1) and $\bar{m}$ the sample anchor that plays the part of μ_R in Section 2.1; Σ, λ_i, P, and δ denote the residual covariance, the eigenvalues of Σ_w, a projection, and a splitting constant, unrelated to the label-noise covariance Σ_c and the ridge penalty λ of Appendix A, the population P of Section 2.1, and the increment δ_j of Section 2.3.

(iv) If $\mathbb{E}\|e_u\|^2 = s^2$ *for all* u*,* $K_4 < \infty$ *and* $(1-\tau)\gamma > 2$*, then* $\mathbb{P}[\cos\angle(m_u, m_v) \geq \tau] \geq 1 - 2K_4/((1-\tau)\gamma/2 - 1)^2$*. For Gaussian residuals with covariance* Σ*,* $K_4 = 2/d_{\text{eff}}$*.*

No right-hand side involves the joint law of (e_u, e_v)*: the statements hold verbatim for independent units, for units that share a source, and for arbitrarily dependent pairs. The minorants in (ii) to (iv) increase with* γ *and tend to 1.*

Proof. (i) If $\max(r_u, r_v) > 1$ the right side is below -1. Otherwise Lemma 0 with $x = \mu$ and $y = e_w$ gives $\angle(\mu, m_w) \leq \arcsin r_w$ for $w \in \{u, v\}$; the triangle inequality for angles gives $\angle(m_u, m_v) \leq \arcsin r_u + \arcsin r_v \leq 2\arcsin\max(r_u, r_v) \leq \pi$, and cos decreases on $[0, \pi]$, so $\cos\angle(m_u, m_v) \geq \cos(2\arcsin r_{\max}) = 1 - 2r_{\max}^2$. The flag statement is $1 - 2r_\tau^2 = \tau$. (ii) Take expectations in (i) with $\max(r_u^2, r_v^2) \leq r_u^2 + r_v^2$ and $\mathbb{E}\, r_w^2 = \mathbb{E}\|e_w\|^2/\|\mu\|^2 \leq 1/\gamma$. (iii) By (i), $\{\cos\angle(m_u, m_v) < \tau\} \subseteq \{r_u > r_\tau\} \cup \{r_v > r_\tau\}$, and Markov's inequality gives $\mathbb{P}[r_w > r_\tau] = \mathbb{P}[\|e_w\|^2 > r_\tau^2\|\mu\|^2] \leq s^2/(r_\tau^2\|\mu\|^2) = 2/((1-\tau)\gamma)$. (iv) Since $r_\tau^2\gamma = (1-\tau)\gamma/2 > 1$, $\{\|e_w\|^2 > r_\tau^2\gamma s^2\} \subseteq \{|\|e_w\|^2 - s^2| \geq (r_\tau^2\gamma - 1)s^2\}$, whose probability is at most $\mathrm{Var}(\|e_w\|^2)/((r_\tau^2\gamma - 1)^2 s^4) \leq K_4/((1-\tau)\gamma/2 - 1)^2$ by Chebyshev; the union over u, v doubles it. For $e \sim N(0, \Sigma)$, $\|e\|^2 = \sum_i \lambda_i z_i^2$ with independent standard normal z_i, so $\mathrm{Var}\,\|e\|^2 = 2\sum_i \lambda_i^2 = 2\,\mathrm{tr}(\Sigma^2)$ and $K_4 = 2\,\mathrm{tr}(\Sigma^2)/(\mathrm{tr}\,\Sigma)^2 = 2/d_{\text{eff}}$. ■

Remark A.1 (mean pooling raises γ). Under (M2) with within-unit terms $\xi_{u,w}$ uncorrelated across w and uncorrelated with b_u, with covariance Σ_ξ, and unit effects b_u with covariance Σ_b, $e_u = b_u + W_u^{-1}\sum_w \xi_{u,w}$ and $\mathbb{E}\|e_u\|^2 = \mathrm{tr}\,\Sigma_b + \mathrm{tr}\,\Sigma_\xi/W_u$. With no unit effect the pooled ratio is $\gamma_{\text{unit}} = W_u\gamma_{\text{window}}$; in general $\gamma_{\text{window}} \leq \gamma_{\text{unit}} \leq \|\mu\|^2/\,\mathrm{tr}\,\Sigma_b$. Positively correlated windows reduce the gain: for an AR(1) profile with coefficient φ the divisor W_u becomes $W_u/[1 + 2\sum_{k=1}^{W_u-1}(1 - k/W_u)\varphi^k]$. Pooling leaves the common direction untouched and shrinks the residual, so it moves the descriptor layer toward the setting of Lemma A.

Remark A.2 (which anisotropy drives the artifact). Lemma A is driven by γ, the energy of the common component relative to the residual. The uncentred second moment $\mathbb{E}[m_u m_u^\top] = \mu\mu^\top + \Sigma$ has a leading eigenvalue share of at least $(\|\mu\|^2 + \sigma_\parallel^2)/(\|\mu\|^2 + s^2) \geq \gamma/(1+\gamma)$, so a large γ forces a dominant uncentred direction. The centred covariance Σ, and the pooled window covariance Σ_w of Appendix I.4 with its diagnostic $\lambda_1/\,\mathrm{tr}\,\Sigma_w$, do not involve μ: their leading share measures residual anisotropy, which enters Lemma B through d_{eff} but does not create the one-sided cone of Lemma A. Example: $\mu = 0$ and $\Sigma = v_1 v_1^\top$ give a centred leading share equal to 1, while for a sign-symmetric residual law $\cos\angle(e_u, e_v)$ equals $+1$ or -1 with equal probability, so the flag rate at any $\tau < 1$ is one half and does not approach 1. The two diagnostics answer different questions and should be read side by side; a plug-in estimate of γ is $\hat\gamma = \|\bar m\|^2/(n^{-1}\sum_u \|m_u - \bar m\|^2)$, whose numerator has expectation $\|\mu\|^2 + s^2/n$ and whose denominator has expectation $s^2(1 - 1/n)$.

Remark A.3 (hubness is downstream of the cone). Under (M1) with small r, $m_u/\|m_u\| = \hat\mu + e_u^\perp/\|\mu\| + O(r^2)$, where $e_u^\perp$ is the component of e_u orthogonal to $\hat\mu$, so $\angle(m_u, m_v) = \|e_u^\perp - e_v^\perp\|/\|\mu\| + O(r^2)$: the raw nearest neighbour of u is the Euclidean nearest neighbour of $e_u^\perp$ among the residuals. A unit whose residual norm is unusually small is then at distance about $\|e_u^\perp\|$ from every u, closer than a typical pair (distance about $\sqrt{2}\,s$ in high dimension), and becomes the nearest neighbour of many units: a hub. After de-centering the cosine is invariant to the scale of $\tilde e_u$, so residual-norm heterogeneity no longer produces hubs. Raw-cosine hubness is therefore a symptom of the cone combined with norm heterogeneity, not an alternative mechanism; the hub count N_1, the number of units whose nearest neighbour is u, is expected to show high skew under C0 and lower skew under C1 for this reason.

I.3 De-centering

Lemma B (de-centering). *Assume (M1) with* $n \geq 2$ *and the anchor* $\bar m$*.*

(i) [exact removal] $\tilde e_u = e_u - \bar e$ *for every* u *and every* μ*, where* $\bar e = n^{-1}\sum_u e_u$*. Hence* $\cos\angle(\tilde e_u, \tilde e_v)$ *does not depend on* μ*, and the* γ*-driven statements of Lemma A have no counterpart for the de-centered cosine.*

Assume further that the e_u are independent with mean zero, covariance Σ, $s^2 = \operatorname{tr}\Sigma$ and $K_4 < \infty$. Then for $u \neq v$:

(ii) [independent pair, infinite pool] for every $\delta \in (0,1)$, $\mathbb{P}[\cos\angle(e_u, e_v) \geq \tau] \leq 1/(d_{\text{eff}}\tau^2(1-\delta)^2) + 2K_4/\delta^2$; with $\delta = 1/2$ the right side is $4/(d_{\text{eff}}\tau^2) + 8K_4$, and for Gaussian residuals $(4/\tau^2 + 16)/d_{\text{eff}}$. The same majorant holds for $\mathbb{P}[\cos\angle(e_u, e_v) \leq -\tau]$.

(iii) [finite pool] $\mathbb{E}\langle\tilde{e}_u, \tilde{e}_v\rangle = -s^2/n$ and $\mathbb{E}\|\tilde{e}_u\|^2 = s^2(1-1/n)$. For every $b \in (0,1)$ with $\theta_\tau + 2\arcsin b < \pi/2$, $\theta_\tau := \arccos\tau$, $\tau_b := \cos(\theta_\tau + 2\arcsin b) > 0$, and every $\delta \in (0,1)$,

$$\mathbb{P}[\cos\angle(\tilde{e}_u, \tilde{e}_v) \geq \tau] \leq 1/(d_{\text{eff}}\tau_b^2(1-\delta)^2) + 1/(nb^2(1-\delta)) + 2K_4/\delta^2.$$

(iv) [same-realization pair; (iv) needs (M1) only] If unit v re-observes unit u, $m_v = m_u + \eta$, then $\tilde{e}_v = \tilde{e}_u + \eta$ exactly, and $\cos\angle(\tilde{e}_u, \tilde{e}_v) \geq \tau$ whenever $\|\eta\| \leq \sqrt{1-\tau^2}\|\tilde{e}_u\|$; writing $\eta = \alpha\tilde{e}_u + \eta_\perp$ with $1 + \alpha > 0$, the pair is flagged iff $\|\eta_\perp\| \leq (1+\alpha)(\sqrt{1-\tau^2}/\tau)\|\tilde{e}_u\|$. Under the raw cosine the same pair is flagged whenever $\max(r_u, r_v) \leq r_\tau$ whatever η is; de-centering replaces a condition on $\|e\|/\|\mu\|$ by a condition on $\|\eta\|/\|\tilde{e}_u\|$.

Proof. (i) $\bar{m} = \mu + \bar{e}$, so $m_u - \bar{m} = e_u - \bar{e}$. (ii) Let $G = \{\|e_u\|^2 \geq (1-\delta)s^2\} \cap \{\|e_v\|^2 \geq (1-\delta)s^2\}$. On G, $\cos\angle(e_u, e_v) \geq \tau$ implies $\langle e_u, e_v\rangle^2 \geq \tau^2(1-\delta)^2 s^4$. By independence and $\mathbb{E}\, e_v e_v^\top = \Sigma$, $\mathbb{E}\langle e_u, e_v\rangle^2 = \mathbb{E}[e_u^\top \Sigma e_u] = \operatorname{tr}(\Sigma^2) = s^4/d_{\text{eff}}$, so Markov gives $\mathbb{P}[\langle e_u, e_v\rangle^2 \geq \tau^2(1-\delta)^2 s^4] \leq 1/(d_{\text{eff}}\tau^2(1-\delta)^2)$. Chebyshev gives $\mathbb{P}[\|e_w\|^2 < (1-\delta)s^2] \leq K_4/\delta^2$ for each w, so $\mathbb{P}[G^c] \leq 2K_4/\delta^2$. The two-sided statement is identical. (iii) Expanding $\langle e_u - \bar{e}, e_v - \bar{e}\rangle$ and using $\mathbb{E}\langle e_u, e_v\rangle = 0$, $\mathbb{E}\langle e_w, \bar{e}\rangle = s^2/n$ and $\mathbb{E}\|\bar{e}\|^2 = s^2/n$ gives $-s^2/n$; likewise $\mathbb{E}\|e_u - \bar{e}\|^2 = s^2 - 2s^2/n + s^2/n$. For the tail, let $H = \{\|\bar{e}\| \leq b\min(\|e_u\|, \|e_v\|)\}$. On H, Lemma 0 with $x = e_w$ and $y = -\bar{e}$ gives $\angle(\tilde{e}_w, e_w) \leq \arcsin b$ for $w \in \{u, v\}$, so $\angle(e_u, e_v) \leq \angle(\tilde{e}_u, \tilde{e}_v) + 2\arcsin b$, and $\cos\angle(\tilde{e}_u, \tilde{e}_v) \geq \tau$ forces $\cos\angle(e_u, e_v) \geq \tau_b$; on G this gives $\langle e_u, e_v\rangle^2 \geq \tau_b^2(1-\delta)^2 s^4$, an event of probability at most $1/(d_{\text{eff}}\tau_b^2(1-\delta)^2)$ by the Markov step of (ii). On G, $H^c \subseteq \{\|\bar{e}\|^2 > b^2(1-\delta)s^2\}$, whose probability is at most $\mathbb{E}\|\bar{e}\|^2/(b^2(1-\delta)s^2) = 1/(nb^2(1-\delta))$ because $\mathbb{E}\|\bar{e}\|^2 = s^2/n$. Adding $\mathbb{P}[G^c] \leq 2K_4/\delta^2$ gives the claim. (iv) $\tilde{e}_v - \tilde{e}_u = m_v - m_u = \eta$ because the anchor is common to both. Lemma 0 with $x = \tilde{e}_u$ and $y = \eta$ gives both statements. ■

Remark B.1 (what de-centering cannot do). The majorant in (ii) scales as $1/d_{\text{eff}}$. If the residual is confined to a one-dimensional subspace with a sign-symmetric law, $\cos\angle(\tilde{e}_u, \tilde{e}_v) = \pm 1$ and the flag rate is about one half for every $\tau < 1$, and (i) shows that no anchor in $\operatorname{span}(v_1)$, in particular the sample mean, changes this: the false flag is then a property of the residual geometry, not of the common component. Raising d_{eff} is what the spectrum-equalising calibers C3 and C4 do (when Σ is proportional to Σ_w; Proposition C), at the price described there.

I.4 The six calibers as linear images of the de-centered residual

Let the pooled window statistics be μ_w, $\Sigma_w = \sum_i \lambda_i v_i v_i^\top$ (population moments over all windows, target included), $\varepsilon := 10^{-2}\operatorname{tr}(\Sigma_w)/d$, $\sigma_i := (\Sigma_w)_{ii}^{1/2}$ with $\sigma_i := 1$ when $\sigma_i < 10^{-12}$, and $T := (\Sigma_w + \varepsilon I)^{-1/2}$, $D := \operatorname{diag}(1/\sigma_i)$, $P := I - \hat{\bar{m}}\hat{\bar{m}}^\top$, $P_1 := I - v_1 v_1^\top$.

Proposition C (representation). *For every pair (u, v) the six detector calibers C0 to C5 read*

$$\text{C0} : \cos\angle(m_u, m_v); \quad \text{C1} : \cos\angle(\tilde{e}_u, \tilde{e}_v); \quad \text{C2} : \cos\angle(P\tilde{e}_u, P\tilde{e}_v);$$
$$\text{C3} : \cos\angle(T\tilde{e}_u, T\tilde{e}_v); \quad \text{C4} : \cos\angle(D\tilde{e}_u, D\tilde{e}_v); \quad \text{C5} : \cos\angle(P_1\tilde{e}_u, P_1\tilde{e}_v).$$

Proof. C2: $Pm_u = P(m_u - \bar{m}) + P\bar{m} = P\tilde{e}_u$ because $P\bar{m} = 0$. C3 to C5: the window transform is $x \mapsto L(x - \mu_w)$ with $L \in \{T, D, P_1\}$ linear; the unit mean of the transformed windows is $L(m_u - \mu_w)$, the mean over units of these is $L(\bar{m} - \mu_w)$, and their difference is $L\tilde{e}_u$, whatever the window counts. ■

C.1 (C1 versus C2). *Let $a_u := \langle\tilde{e}_u, \hat{\bar{m}}\rangle$, so $\tilde{e}_u = P\tilde{e}_u + a_u\hat{\bar{m}}$ and $\angle(\tilde{e}_u, P\tilde{e}_u) = \arctan(|a_u|/\|P\tilde{e}_u\|)$. Then*

$$|\cos_{\text{C1}}(u, v) - \cos_{\text{C2}}(u, v)| \leq \arctan(|a_u|/\|P\tilde{e}_u\|) + \arctan(|a_v|/\|P\tilde{e}_v\|).$$

If $a_u = 0$ for every u, then $\cos_{C1} = \cos_{C2}$ on all pairs; the majorant on the right vanishes iff $a_u = 0$ for every u. Since $\tilde{e}_u$ is free of μ, $\|\mu\|$ enters only through the direction $\hat{\bar{m}} = (\mu + \bar{e})/\|\mu + \bar{e}\|$:

(a) as $\gamma \to \infty$ with a fixed residual law, $\hat{\bar{m}} \to \hat{\mu}$ and $a_u \to \langle e_u - \bar{e}, \hat{\mu}\rangle$, with $\mathbb{E}\langle e_u - \bar{e}, \hat{\mu}\rangle^2 = \sigma_\parallel^2(1 - 1/n)$; C1 and C2 converge iff $\sigma_\parallel^2 = 0$ (no residual energy along the common direction), and otherwise the majorant tends to a nonzero limit set by $\sigma_\parallel^2/s^2$ (equal to $1/d$ for isotropic residuals) (and so, generically, does the gap);

(b) as $\|\mu\| \to 0$, $\hat{\bar{m}} \to \bar{e}/\|\bar{e}\|$ and $a_u \to \langle e_u, \bar{e}/\|\bar{e}\|\rangle - \|\bar{e}\|$, a μ-free limit that is not zero in general; at the same time $\cos_{C0} \to \cos \angle(e_u, e_v)$, which differs from $\cos_{C1}$ by the anchor term of Lemma B(iii).

Hence "C1 and C2 converge as $\|\mu\| \to 0$" is not established; what holds is the exact gap majorant above, which tends to zero exactly when the anchor-direction shares $a_u/\|P\tilde{e}_u\|$ do, and then the two calibers converge.

Proof. $P\tilde{e}_u \perp \hat{\bar{m}}$ gives the arctan form. The triangle inequality for angles gives $|\angle(\tilde{e}_u, \tilde{e}_v) - \angle(P\tilde{e}_u, P\tilde{e}_v)| \le \angle(\tilde{e}_u, P\tilde{e}_u) + \angle(\tilde{e}_v, P\tilde{e}_v)$, and $|\cos x - \cos y| \le |x - y|$. The right side is zero iff $a_u = a_v = 0$. (a) For fixed $\bar{e}$, $(\mu + \bar{e})/\|\mu + \bar{e}\| \to \hat{\mu}$ as $\|\mu\| \to \infty$, and $\mathrm{Var}(\langle e_u, \hat{\mu}\rangle - \langle \bar{e}, \hat{\mu}\rangle) = \sigma_\parallel^2(1 - 2/n + 1/n)$. (b) is the substitution $\mu = 0$. ■

C.2 (ZCA, C3). *(i) If $\Sigma_w = \lambda I$ on the span of the de-centered residuals, T acts there as the scalar $(\lambda + \varepsilon)^{-1/2}$ and $\cos_{C3} \equiv \cos_{C1}$. (ii) If $\Sigma = c\Sigma_w$ for some $c > 0$ (as under (M2) without unit effects, where $\Sigma = \Sigma_\xi/W$), then $\mathrm{Cov}(T\tilde{e}_u)$ has eigenvalues proportional to $\lambda_i/(\lambda_i + \varepsilon) \in [0, 1)$; since $x \mapsto x/(x + \varepsilon)$ is concave with $f(0) = 0$, the normalised spectrum is majorised by that of Σ and the participation ratio d_{eff} does not decrease, tending to $\operatorname{rank}\Sigma_w$ as $\varepsilon \to 0$. With unit effects ($\Sigma = \Sigma_b + \Sigma_\xi/W$) whitening by Σ_w can lower d_{eff}. (iii) Let $x = \tilde{e}_u$ and $\eta \perp x$. C1 flags the pair iff $\|\eta\|^2/\|x\|^2 \le (1 - \tau^2)/\tau^2$. If $x \in \operatorname{span}\{v_i : \lambda_i \ge \lambda_{\text{hi}}\}$ and $\eta \in \operatorname{span}\{v_i : \lambda_i \le \lambda_{\text{lo}}\}$ with $\lambda_{\text{lo}} < \lambda_{\text{hi}}$, then $T\eta \perp Tx$ and*

$$\|T\eta\|^2/\|Tx\|^2 \ge ((\lambda_{\text{hi}} + \varepsilon)/(\lambda_{\text{lo}} + \varepsilon)) \cdot \|\eta\|^2/\|x\|^2,$$

so C3 misses the pair as soon as $((\lambda_{\text{hi}} + \varepsilon)/(\lambda_{\text{lo}} + \varepsilon))\|\eta\|^2/\|x\|^2 > (1 - \tau^2)/\tau^2$, although C1 flags it. With the floor ε above the amplification is at most $(\lambda_1 + \varepsilon)/\varepsilon = 1 + 100\, d\, \lambda_1/\operatorname{tr}\Sigma_w$. Conversely, if x and η lie in one eigenspace of Σ_w, T is a scalar on both and C3 agrees with C1 on the pair.

Proof. (i) and the last sentence: T is a scalar on an eigenspace. (ii) $\mathrm{Cov}(T\tilde{e}_u) = T\Sigma T$; when Σ and Σ_w share eigenvectors it is diagonal in that basis with entries $\lambda_i^{(\Sigma)}/(\lambda_i + \varepsilon)$, which are $\propto \lambda_i/(\lambda_i + \varepsilon)$ only when $\Sigma = c\Sigma_w$ (for $\Sigma = I_4$ and $\Sigma_w = \operatorname{diag}(100, 1, 1, 1)$ the whitened spectrum is $\propto (1/(100 + \varepsilon), 1/(1 + \varepsilon), 1/(1 + \varepsilon), 1/(1 + \varepsilon))$ and the participation ratio falls below 4). With $\Sigma = c\Sigma_w$, $f(x) = x/(x + \varepsilon)$ is concave with $f(0) = 0$, so $f(x)/x$ is non-increasing and the unit-sum normalisation of $(f(\lambda_i))_i$ is majorised by that of $(\lambda_i)_i$; the participation ratio is Schur-concave, hence does not decrease, and as $\varepsilon \to 0$ the entries tend to $c \cdot \mathbf{1}\{\lambda_i > 0\}$, whose participation ratio is $\operatorname{rank}\Sigma_w$. (iii) T is diagonal in the eigenbasis, so the two spans are T-invariant and stay orthogonal; $\|Tx\|^2 = \sum_{i:\lambda_i \ge \lambda_{\text{hi}}} x_i^2/(\lambda_i + \varepsilon) \le \|x\|^2/(\lambda_{\text{hi}} + \varepsilon)$ and $\|T\eta\|^2 \ge \|\eta\|^2/(\lambda_{\text{lo}} + \varepsilon)$. Lemma 0 (criterion with $\alpha = 0$) applied to $(Tx, T\eta)$ finishes. ■

C.3 (per-coordinate standardization, C4). *(i) If all σ_i are equal, D is a scalar and $\cos_{C4} \equiv \cos_{C1}$. In general, with $\sigma_{\min} := \min_i \sigma_i$, $\sigma_{\max} := \max_i \sigma_i$, $\delta_\sigma := (\sigma_{\max} - \sigma_{\min})/(\sigma_{\max} + \sigma_{\min})$ and $\sigma_h := 2\sigma_{\max}\sigma_{\min}/(\sigma_{\max} + \sigma_{\min})$, one has $\|\sigma_h D - I\|_{\text{op}} \le \delta_\sigma$, hence $\angle(Dx, x) \le \arcsin\delta_\sigma$ for every $x \ne 0$ and $|\cos_{C4}(u, v) - \cos_{C1}(u, v)| \le 2\arcsin\delta_\sigma$. (ii) If Σ_w is diagonal (the anisotropy sits in a few coordinates, the rogue-dimension case), then $T = \operatorname{diag}((\sigma_i^2 + \varepsilon)^{-1/2})$ and C4 agrees with C3 up to the ε floor. (iii) If $\Sigma_w = \sigma^2 I + (\lambda_1 - \sigma^2)v_1 v_1^\top$ with $v_1 = d^{-1/2}(1, \ldots, 1)$, every σ_i^2 equals $\sigma^2 + (\lambda_1 - \sigma^2)/d$, so D is a scalar: C4 coincides with C1 and leaves the dominant direction untouched, while C3 and C5 remove or flatten it. Hence per-coordinate standardization changes the de-centered verdict only when the residual variance is heterogeneous across coordinates, and it equals ZCA only when Σ_w is diagonal.*

Proof. $\sigma_h D - I = \operatorname{diag}(\sigma_h/\sigma_i - 1)$, and $\sigma_i \in [\sigma_{\min}, \sigma_{\max}]$ gives $\sigma_h/\sigma_i - 1 \in [-\delta_\sigma, \delta_\sigma]$ (the two endpoints are attained). Lemma 0 with x and $y = (\sigma_h D - I)x$, $\|y\| \le \delta_\sigma\|x\| \le \|x\|$, gives

$\angle(\sigma_h Dx, x) \le \arcsin \delta_\sigma$; the positive scalar σ_h does not change directions. The triangle inequality for angles and $|\cos a - \cos b| \le |a - b|$ give (i). (ii) and (iii) are direct. ■

C.4 (first principal direction, C5, $k = 1$). *(i) C5 and C2 each remove one direction from $\tilde{e}_u$: v_1 (leading eigenvector of the centred Σ_w) and $\hat{\bar{m}}$ respectively. $\|P_1 - P\|_{\mathrm{op}} = \sin \angle(v_1, \hat{\bar{m}})$, so $\|P_1 \tilde{e}_u - P\tilde{e}_u\| \le \sin \angle(v_1, \hat{\bar{m}}) \|\tilde{e}_u\|$ and, whenever $\sin \angle(v_1, \hat{\bar{m}}) \|\tilde{e}_u\| \le \|P\tilde{e}_u\|$, $\angle(P_1\tilde{e}_u, P\tilde{e}_u) \le \arcsin(\sin \angle(v_1, \hat{\bar{m}}) \|\tilde{e}_u\| / \|P\tilde{e}_u\|)$; $P_1 = P$ iff $v_1 = \pm\hat{\bar{m}}$; in that case C5 ≡ C2. (ii) Because Σ_w is centred, v_1 is a residual direction, and $v_1 = \pm\hat{\mu}$ is the coincidence that $\hat{\mu}$ is a leading eigenvector of Σ_w, i.e. $\hat{\mu}^\top \Sigma_w \hat{\mu} = \lambda_1(\Sigma_w)$. In the generic case C5 removes the leading residual direction, which C2 keeps, and keeps the anchor direction, which C2 removes. (iii) [uncentred variant] If v_1° denotes the leading eigenvector of the uncentred second moment $S_w = \mu\mu^\top + \Sigma_w$ (windows not centred; the reading of Parupudi 2026), then $\sin \angle(v_1^\circ, \hat{\mu}) \le 2\lambda_1(\Sigma_w)/\|\mu\|^2 = 2p_1/\gamma_w$ with $p_1 := \lambda_1(\Sigma_w)/\operatorname{tr}\Sigma_w$ and $\gamma_w := \|\mu\|^2/\operatorname{tr}\Sigma_w$. So $v_1^\circ \to \hat{\mu}$ as $\gamma_w \to \infty$, projecting out v_1° tends to C2, and the common component left after that projection has norm $\|\mu\| \sin \angle(v_1^\circ, \hat{\mu}) \le 2\lambda_1(\Sigma_w)/\|\mu\|$. The statement "C5 with $k = 1$ equals C2 when $\mu \approx v_1$" therefore holds for the uncentred leading direction in the mean-dominated regime, and for the centred definition used here only under the coincidence in (ii).*

Proof. (i) For unit vectors a, b the matrix $bb^\top - aa^\top$ has nonzero eigenvalues $\pm \sin \angle(a, b)$. Lemma 0 with $x = P\tilde{e}_u$ and $y = (P_1 - P)\tilde{e}_u$. (ii) is direct. (iii) is the Davis-Kahan $\sin\theta$ theorem in the form of Yu et al. (2015, Theorem 2 with $r = s = 1$): for the symmetric matrix $\mu\mu^\top$ with eigen-gap $\|\mu\|^2$ and the perturbation Σ_w, $\sin \angle(v_1(\mu\mu^\top + \Sigma_w), \hat{\mu}) \le 2\|\Sigma_w\|_{\mathrm{op}}/\|\mu\|^2$. ■

J TWO PRE-REGISTERED FOLLOW-UP TABLES

Table J.1 applies six anisotropy corrections to the archived window-level features at the same 0.9 threshold. The archive has no truth-bearing units, so the table reports flagged carrier sets and their overlap with mean removal; it decides no cell.

Table J.2 places three CWRU pools side by side: the archived pool with the loader copy, the headline pool with the copy dropped, and the pool read by a corrected loader. Headline numbers stay with the middle column.

Table J.1: Six anisotropy corrections on the archived window-level features (chronos-t5-small; pre-registered follow-up run on anisotropy calibers). Each caliber flags a carrier when the similarity of two unit descriptors reaches 0.9: C0 is raw cosine and serves as the artefact reference; C1 is the mean removal of Section 4.2 (the paper's detector); C2 projects out the mean direction; C3 is ZCA whitening of the windows; C4 standardizes each dimension of the windows; C5 projects out the first principal direction of the windows. Block A: share of units with at least one carrier (units flagged). Block B: pool Jaccard of the flagged set against C1 (units flagged only here / only under C1); n/a when both sets are empty. Block C: skewness of the nearest-neighbour occurrence count N_1 (largest N_1). Block D: share of window variance on the leading direction after the caliber's window transform (C0 to C2 leave the windows unchanged). Block E: the pre-registered branch reading. The archive has no held-out windows and no truth-bearing units, so the table reads no excess and no sensitivity and decides no cell; the Paderborn rows are the training-window pool of 8 bearings and differ from the 22-bearing pool of Table 1.

Pool	n	C0 raw	C1 mean removal	C2 mean direction out	C3 ZCA whitening	C4 standardized	C5 first direction out
A. Prevalence: share of units with a carrier at 0.9 (units flagged); record level							
CWRU, 40 records (10 bearings)	40	1.000 (40)	0.675 (27)	0.675 (27)	0.050 (2)	0.675 (27)	0.500 (20)
CWRU without the loader copy, 39 records (headline pool)	39	1.000 (39)	0.667 (26)	0.667 (26)	0.000 (0)	0.667 (26)	0.462 (18)
Paderborn, 60 records (8 bearings; training-window pool)	60	1.000 (60)	0.983 (59)	0.983 (59)	0.000 (0)	0.883 (53)	0.183 (11)

continued on next page

continued from the previous page

Pool	n	C0 raw	C1 mean removal	C2 mean direction out	C3 ZCA whitening	C4 standardized	C5 first direction out
PTB-XL, 160 records (40 patients)	160	0.975 (156)	0.069 (11)	0.044 (7)	0.000 (0)	0.044 (7)	0.000 (0)
MIMII, 12 machines (one record each)	12	1.000 (12)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)
CWRU after full fine-tuning, 40 records	40	1.000 (40)	0.375 (15)	0.450 (18)	0.050 (2)	0.200 (8)	0.325 (13)
CWRU after full fine-tuning, without the loader copy, 39 records	39	1.000 (39)	0.385 (15)	0.410 (16)	0.000 (0)	0.154 (6)	0.282 (11)
A. Prevalence, subject level (deletion units of Section 5.1; training-window pool)							
CWRU, 10 bearings	10	0.900 (9)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)
Paderborn, 8 bearings (training-window pool)	8	1.000 (8)	0.500 (4)	0.500 (4)	0.000 (0)	0.500 (4)	0.250 (2)
PTB-XL, 40 patients	40	0.975 (39)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)
MIMII, 12 machines	12	1.000 (12)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)
CWRU after full fine-tuning, 10 bearings	10	0.900 (9)	0.000 (0)	0.000 (0)	0.000 (0)	0.000 (0)	0.200 (2)
B. Carrier-set overlap with C1: pool Jaccard (only here / only under C1); record level							
CWRU, 40 records (10 bearings)	40	0.68 (13/0)	reference	1.00 (0/0)	0.07 (0/25)	1.00 (0/0)	0.68 (1/8)
CWRU without the loader copy, 39 records (headline pool)	39	0.67 (13/0)	reference	1.00 (0/0)	0.00 (0/26)	1.00 (0/0)	0.63 (1/9)
Paderborn, 60 records (8 bearings; training-window pool)	60	0.98 (1/0)	reference	1.00 (0/0)	0.00 (0/59)	0.90 (0/6)	0.19 (0/48)
PTB-XL, 160 records (40 patients)	160	0.07 (145/0)	reference	0.64 (0/4)	0.00 (0/11)	0.64 (0/4)	0.00 (0/11)
MIMII, 12 machines (one record each)	12	0.00 (12/0)	reference	n/a (0/0)	n/a (0/0)	n/a (0/0)	n/a (0/0)
CWRU after full fine-tuning, 40 records	40	0.38 (25/0)	reference	0.65 (5/2)	0.13 (0/13)	0.35 (2/9)	0.40 (5/7)
CWRU after full fine-tuning, without the loader copy, 39 records	39	0.38 (24/0)	reference	0.72 (3/2)	0.00 (0/15)	0.24 (2/11)	0.30 (5/9)
C. Hubness: skewness of N_1 (largest N_1); record level							
CWRU, 40 records (10 bearings)	40	1.05 (4)	0.65 (3)	0.65 (3)	0.65 (3)	0.65 (3)	0.57 (3)
CWRU without the loader copy, 39 records (headline pool)	39	1.03 (4)	0.57 (3)	0.57 (3)	0.64 (3)	0.57 (3)	0.46 (3)
Paderborn, 60 records (8 bearings; training-window pool)	60	0.55 (3)	0.64 (3)	0.68 (3)	0.73 (3)	0.60 (3)	0.70 (3)
PTB-XL, 160 records (40 patients)	160	1.55 (6)	1.19 (6)	1.19 (6)	1.71 (7)	1.19 (6)	1.77 (7)
MIMII, 12 machines (one record each)	12	1.77 (5)	0.00 (2)	0.66 (3)	0.00 (2)	0.00 (2)	0.50 (3)
CWRU after full fine-tuning, 40 records	40	0.74 (3)	0.65 (3)	0.51 (3)	0.91 (4)	0.91 (4)	0.85 (4)
CWRU after full fine-tuning, without the loader copy, 39 records	39	0.64 (3)	0.90 (4)	0.95 (4)	0.90 (4)	0.84 (4)	0.85 (4)
D. Leading direction: $\lambda_1 / \operatorname{tr} \Sigma_w$ of the window features after the caliber's window transform; n = windows							
CWRU, 40 records (10 bearings)	960	0.253	0.253	0.253	0.003	0.205	0.190
CWRU without the loader copy, 39 records (headline pool)	936	0.254	0.254	0.254	0.003	0.207	0.180
Paderborn, 60 records (8 bearings; training-window pool)	1440	0.310	0.310	0.310	0.003	0.244	0.172
PTB-XL, 160 records (40 patients)	1280	0.207	0.207	0.207	0.003	0.175	0.135
MIMII, 12 machines (one record each)	96	0.288	0.288	0.288	0.011	0.211	0.188
CWRU after full fine-tuning, 40 records	960	0.689	0.689	0.689	0.005	0.436	0.437
CWRU after full fine-tuning, without the loader copy, 39 records	936	0.691	0.691	0.691	0.005	0.437	0.440

E. Pre-registered branch reading (R4, mechanical; descriptive only)

AGREE (C2 to C5 within one flagged unit of C1 and pool Jaccard at least 0.8, in all 14 blocks): does not hold; WHITEN-MORE (C3 or C4 flags more units than C1): no block.

WHITEN-FEWER: C3 flags fewer units than C1 on the CWRU 40, CWRU 39, Paderborn, and PTB-XL records (25, 26, 59, 11 fewer) and on the Paderborn bearings (4); C4 flags 6 and 4 fewer on the Paderborn and PTB-XL records and the same set on CWRU; after fine-tuning C3 and C4 flag 13 and 7 fewer (15 and 9 without the loader copy).

R2-DIFFERS (pool Jaccard of C2 against C1 below 0.8): PTB-XL records 0.64; CWRU after fine-tuning 0.65 (0.72 without the loader copy); the frozen CWRU, Paderborn, and MIMII sets coincide with C1 (1.00).

HUBNESS: the N_1 skewness falls from raw to C1 on the CWRU (1.05 to 0.65), PTB-XL (1.55 to 1.19), and MIMII (1.77 to 0.00) records and rises on the Paderborn records (0.55 to 0.64); the largest N_1 on CWRU falls from 4 to 3.

Table J.2: The CWRU loader copy: archived pool, headline pool, and corrected loader (pre-registered follow-up run on the loader copy). In the archive the file 2HP/Normal_99.mat also stores the 1 HP array and the loader read the first array, so the 2 HP Normal row repeated the 1 HP recording. Column 2: the archived pool (40 series, 39 distinct recordings). Column 3: the pool with the copy dropped (39 series), the read-only deduplication that the headline numbers of Appendix B.3, Figure A3, and Table 1 (continued) use. Column 4: the same three cells rerun with a corrected loader that reads the file's own array, so that the pool holds 40 distinct recordings. Last column: corrected-loader reading minus headline reading, marked when it exceeds the 0.012 bootstrap noise floor of the ledger. Instrument, thresholds, seeds, label rules, and bootstrap constants are unchanged across the columns; only the pool differs. The pair-dropping variant of Appendix B.3 is in the source table.

Quantity	Archived pool (loader copy in the pool)	Copy dropped (headline pool)	Corrected loader (40 distinct recordings)	Corrected − headline
Record-level screen (record units; chronos-t5-small; CWRU, Paderborn, PTB-XL requests together)				
Requests; LEAK; SAFE (two-class set)	250; 31; 90 (121)	249; 30; 90 (120)	250; 29; 88 (117)	-3 (two-class set)
CWRU LEAK; of which with a sibling at 0.9	14; 10	13; 9	12; 8	-1
CWRU records with a sibling at 0.9	27 of 40	26 of 39	27 of 40	+1
Sign screen full_de: tp / fp / tn / fn	13 / 9 / 81 / 18	12 / 9 / 81 / 18	12 / 9 / 79 / 17	tn -2; fn -1
Sensitivity [CP95]	0.419 [0.245, 0.609]	0.400 [0.227, 0.594]	0.414 [0.235, 0.611]	+0.0138 (above 0.012)
Specificity [CP95]	0.900 [0.819, 0.953]	0.900 [0.819, 0.953]	0.898 [0.815, 0.952]	-0.0023
Balanced accuracy [class-stratified bootstrap]	0.660 [0.568, 0.751]	0.650 [0.556, 0.744]	0.656 [0.558, 0.753]	+0.0058
AUROC [bootstrap]	0.838 [0.763, 0.902]	0.834 [0.758, 0.902]	0.847 [0.773, 0.911]	+0.0123 (above 0.012)
Window-clustered relabeling: positives / two-class set	4 / 37	3 / 36	5 / 37	+2
Window-clustered tp / fp / tn / fn; BA [bootstrap]	4 / 1 / 32 / 0; BA 0.985 [0.955, 1.000]	3 / 1 / 32 / 0; BA 0.985 [0.955, 1.000]	5 / 1 / 31 / 0; BA 0.984 [0.953, 1.000]	-0.0005
Archived-rule positives that become undetermined under relabeling	27 of 31	27 of 30	24 of 29	-3
Three-valued reading: abstentions on LEAK; false-SAFE rate [CP95]	27 of 31; 0.129 [0.036, 0.298]	26 of 30; 0.133 [0.038, 0.307]	25 of 29; 0.138 [0.039, 0.317]	+0.0046
Subject-level control on the same requests: structural zeros / requests; LEAK	222 / 250; 1	221 / 249; 1	222 / 250; 1	+1 (denominator 250 again)
Series-level cross-family cells (archived cross-family run; de-centered caliber; CWRU)				
Chronos-Bolt: carrier-subset median (units with a carrier of N)	+1.05 (13 of 40)	+0.90 (12 of 39)	+1.16 (11 of 40)	+0.2577 (above 0.012)
Chronos-Bolt: positive share of the subset; subset range	12 of 13; -0.28 to +5.88	11 of 12; -0.28 to +5.88	10 of 11; -0.14 to +2.42	-3.457 (largest)
Chronos-Bolt: all-series median [CI] (the pre-registered estimand)	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.0000
TimesFM-2.5: carrier-subset median (units with a carrier of N)	-0.16 (22 of 40)	-0.17 (21 of 39)	-0.15 (24 of 40)	+0.0261 (above 0.012)
TimesFM-2.5: all-series median [CI] (the pre-registered estimand)	0.000 [-0.013, 0.000]	0.000 [-0.026, 0.000]	0.000 [-0.048, 0.000]	-0.0224 (lower end, above 0.012)
Bearings carrying the Chronos-Bolt subset	4 (InnerRace_021, Normal, OuterRace_007, OuterRace_014)	4 (InnerRace_021, Normal, OuterRace_007, OuterRace_014)	3 (Normal, OuterRace_007, OuterRace_014)	-1