This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material: Same Anatomy, Same Score, Different Verdict

Behnood Mashayekhi[1][0009−0002−3062−4001], Arian Amiramjadi[1], and Parsa Hosseini[1]

Amesha, London, United Kingdom
behnood@amesha.co, arian@amesha.co, parsa@amesha.co
https://amesha.co

This supplement reports proof details, complete frozen-grid denominators, and validation boundaries. It is not needed to understand the main-paper claims. No raw medical volume is included, and no result below expands the permitted interpretation in the main paper.

Operational Definitions

A *no-training sufficiency control* is a deterministic rule with no fitted parameters; here it answers the two-candidate laterality query by physical-x or pair-centered rank. A *GT-intimate unit* is a patient–tooth unit with $d(G,T) \le \tau$. The *patient-equal away-minus-toward difference* averages $a_\tau^{\text{away}} - a_\tau^{\text{toward}}$ across all retained exact rows within each patient, then averages patients equally. The *33-cell simultaneous band* is one 95% maximum-deviation patient-bootstrap envelope over the 33 threshold-by-magnitude cells with at least five patients; sparse cells receive no band.

1 Full Proof of the Critical-Region Proposition

Let Ω be a finite physical voxel lattice with Euclidean metric ρ, let $T \subset \Omega$ be a nonempty tooth mask with 6-neighbour surface $S(T)$, and let $G \subset \Omega$ be a nonempty ground-truth canal mask. Define

$$\delta_T(x) = \min_{t \in S(T)} \rho(x,t), \qquad C_\tau = \{x \in G : \delta_T(x) \le \tau\},$$

and put $N = |G|$, $m = |C_\tau|$, with $1 \le m < N$. Let $P_{\text{crit}} = G \setminus C_\tau$. Totally order G by decreasing δ_T, breaking ties by a fixed lexicographic index order; let F_τ be the first m voxels and $P_{\text{far}} = G \setminus F_\tau$. For

$$a_\tau(P;T) = \mathbb{I}\left\{ \min_{p \in P} \delta_T(p) > \tau \right\},$$

we have $a_\tau(P_{\text{crit}};T) = 1$ and $a_\tau(P_{\text{far}};T) = 0$.

Every voxel retained in P_{crit} lies outside C_τ and therefore has distance greater than τ. Because $m < N$, deleting the m farthest voxels leaves a nonempty

prediction. Every noncritical voxel precedes every critical voxel in the order. An m-element prefix cannot contain all m critical voxels, since doing so would also require the preceding $N - m$ noncritical voxels and hence N deletions. At least one critical voxel remains in P_{far}, proving its action is zero.

Both predictions are subsets of G obtained by deleting exactly m voxels, so

$$(\text{TP}, \text{FP}, \text{FN}) = (N - m, 0, m),$$

and their predicted volumes and symmetric differences are equal. Substitution with $\kappa = m/N$ yields

$$\text{Dice} = \frac{2(1-\kappa)}{2-\kappa}, \quad \text{IoU} = \text{Recall} = 1 - \kappa, \quad \text{Precision} = 1.$$

Consequently no deterministic function of these counts and volumes can recover the action on both pair members. This statement does not cover surface, topology, calibration, or tooth-conditioned metrics.

For any $q < 1$, choose a finite instance with $\kappa < 2(1-q)/(2-q)$. Then $1 - \text{Dice} = \kappa/(2-\kappa) < 1 - q$, so P_{crit} has Dice above q and the opposite action from G. This is an existence statement, not a frequency statement.

2 Audit Chronology and No-Rescue Boundary

The original transfer analysis was audited before any exact-family result. The corrected role/convergence and same-world protocols were then frozen and executed. Their positive promotion gates failed. A separate directional replay route anchored to actual predictions was also frozen and executed; its 2-mm primary family had only 24/49 feasible intimate units and 10 discordant patients against a required 20, while its 90% certificate covered a median 3.94% of decisions against a required 20%. Those failed gates were not changed.

Only after that failure was the exact-equivalence benchmark designed. Its $\tau \in \{1, 2, 3\}$ mm and 0.3–6.0 mm displacement grid, builder, summarizer, validator, and minimum 20-patient rule were frozen before enumeration. Family A and Family B are therefore pre-enumeration frozen for their own questions, but they are not an external preregistration and do not retroactively rescue the earlier replay experiment.

3 Complete Corrected Transfer Results

At 40 epochs, refit-specific target-minus-source ranges were $[-.0316, .0947]$ and $[-.0497, .0995]$ for Dental CBCT and PM lower. At 320 epochs they were $[0, 0]$ and $[0, .0105]$. Deterministic physical-x and pair-centered rank accuracy was 1.000 in both corpora; constant accuracy was .500 and case-indexed random accuracy was .505 and .482.

Table 1. Patient-disjoint accuracy averaged over 20 actual refits. Intervals resample patients after averaging refits.

Target	Epoch	Target fit	Source fit	Δ (95% CI)
Dental CBCT	40	.9432	.9292	.0139 [.0053, .0234]
	80	.9887	.9813	.0074 [.0000, .0155]
	160	1.0000	1.0000	.0000 [.0000, .0000]
	320	1.0000	1.0000	.0000 [.0000, .0000]
PM lower	40	.8801	.8603	.0198 [.0089, .0304]
	80	.9546	.9395	.0151 [.0013, .0285]
	160	.9997	.9976	.0021 [.0007, .0039]
	320	1.0000	.9978	.0022 [.0007, .0043]

4 Same-World Verification and Frozen Gate

The complete run contained 5 folds, 3 fitted representations, 2 training corpora, 20 seeds, and 4 epoch budgets: 2,400 jobs and 1,204,800 indexed patient/model/transform rows. Independently sorted world-point sets matched exactly for all 286 masks; maximum pointwise drift and count drift were exactly zero, while maximum extent drift was numerical zero (1.11×10^{-16}).

Table 2. Primary $\pm$20-mm same-world result.

Target	Complete patients	Feasibility	Absolute flips (95% CI)	Seeds $\geq .05$
Dental CBCT	38/95	.400	.0112 [.0063, .0168]	0/20
PM lower	191/191	1.000	.0093 [.0075, .0113]	0/20

World-coordinate, pair-centered, and deterministic-rank controls had zero flips. Neither patient-only nor crossed patient–refit simultaneous bands supported a transfer-sign reversal. Thus materiality, seed stability, Dental CBCT feasibility, and overall promotion failed.

5 Exact Family A: Analytic Construction

All 2,427 attempts remained in the denominator. Every one of the 169 retained rows had exact integer count equality and opposite clearance actions. The independent validator reconstructed all 92 cases and every retained row from raw labels. Connectivity changes were reported, never used as post-result exclusions.

Table 3. Patient-first Family-A summaries. Connectivity is the fraction unchanged for critical/far deletion.

τ (mm)	Attempts	Pairs	Patients	κ (95% CI)	Dice (95% CI) / connectivity
1	809	27	19	.0090 [.0064,.0117]	.9955 [.9941,.9968] / 1.000, 1.000
2	809	49	27	.0222 [.0158,.0289]	.9887 [.9853,.9920] / .980, 1.000
3	809	93	42	.0263 [.0182,.0350]	.9864 [.9816,.9905] / .720, .989

6 Exact Family B: Actual-Prediction Anchor

Family B starts from a thresholded predicted canal. A direction is derived from the nearest predicted-canal/tooth surface pair without canal truth. The tooth-side predicted half is translated by exact opposite integer vectors; the other half is fixed. Canal truth does not choose this direction; it is used for exact-match retention and truth-relative metric evaluation. Every attempt is assigned one of the statuses in Table 5.

6.1 Preselected worked witness

The illustrative row was selected before visualization as the first under the same prespecified deterministic ordering of all 2-mm-discordant exact rows; selection did not filter on appearance, Dice, magnitude, or clearance. It is not GT-intimate at 2 mm and is not a false-clear example. Table 4 shows how the row contributes +1 before all exact rows are averaged within its patient.

Table 4. Complete worked witness. Counts and named scores are identical; only the tooth-conditioned clearance proxy differs.

Quantity	Toward arm	Away arm
Realized displacement (mm)	5.739	5.739
Truth intersection (voxels)	5,010	5,010
Predicted volume (voxels)	22,439	22,439
Dice	0.230525	0.230525
IoU	0.130279	0.130279
Precision	0.223272	0.223272
Recall	0.238265	0.238265
Volume similarity	0.967515	0.967515
Minimum tooth distance (mm)	0.000	10.607
Action at $\tau = 2$ mm	0 (CLOSE)	1 (CLEAR)

Exact yield was 10.86% of all attempts and 29.19% of geometry-feasible attempts. The 1,757 exact rows covered 64 patients. At least one action- discordant row occurred in 52, 57, and 61 patients at $\tau = 1, 2, 3$ mm, respectively. Table 6 distinguishes all exact rows from the action-discordant subset.

Table 5. Complete Family-B attempt accounting.

Status	Count	Status	Count
Toward clips volume	6,260	Away clips volume	458
Toward overlaps fixed half	1,806	Away overlaps fixed half	245
Duplicate smaller nominal	1,272	Undefined zero-distance direction	120
Geometry feasible, not exact	4,262	Retained exact row	1,757
Total	16,180	Geometry feasible	6,019

Table 6. Family-B patient-equal action difference, conditional on an exact row. All exact rows stay in the estimator; "discordant" counts the subset whose two actions differ.

τ	Exact rows / patients	Discordant rows / patients	Away > toward	Difference (95% CI)
1	1,757 / 64	632 / 52	632 / 0 reverse	.378 [.301,.458]
2	1,757 / 64	827 / 57	827 / 0 reverse	.481 [.402,.559]
3	1,757 / 64	1,038 / 61	1,038 / 0 reverse	.605 [.530,.679]

6.2 Magnitude dependence

Only five exact rows occurred below 3 mm, and none was action-discordant. The table reports the full magnitude grid; effects with fewer than five patients are displayed but were excluded from the simultaneous band.

Table 7. Exact-row yield and 2-mm patient-equal action difference by nominal displacement. n_p is the number of patients with an exact row.

mm	Exact	n_p	Difference	mm	Exact	n_p	Difference
.3	0	0	–	3.3	58	33	.303
.6	1	1	.000	3.6	74	33	.298
.9	0	0	–	3.9	132	56	.376
1.2	0	0	–	4.2	152	57	.421
1.5	1	1	.000	4.5	180	60	.433
1.8	0	0	–	4.8	200	62	.464
2.1	1	1	.000	5.1	215	62	.513
2.4	1	1	.000	5.4	239	63	.520
2.7	1	1	.000	5.7	231	62	.569
3.0	21	13	.141	6.0	250	62	.614

The 33 cells with at least five patients entered a simultaneous patient bootstrap. Its maximum-deviation critical value was .193816; an independent implementation reproduced all points, intervals, bands, and 10,000 complete draws exactly. Exact-pair Dice was low: row mean .273, median .286, range [.017,.493]; patient-equal mean .276, 95% CI [.254,.298]. Family B is therefore an actual-prediction-anchored equivalence construction, not an equal-high-Dice or subtle-error result.

7 Independent Validation and Claim Firewall

Table 8. Independent validation checks. All were evaluated against the frozen source rows.

Check	Status
Frozen protocol matches the independently verified implementation	PASS
All 92 cases, 809 teeth, and 16,180 attempts reconstructed	PASS
All 1,757 retained rows reconstructed without omissions or duplicates	PASS
Truth intersection and predicted volume equal pairwise	PASS
All named-score, norm, and vector-sum gaps are zero	PASS
Independent action labels agree for every row	PASS
Attempt, failure, magnitude, tooth, patient, and cell totals match	PASS
All patient-first intervals and the simultaneous band reproduced	PASS
Frozen grid contains at least 20 discordant patients	PASS (61)

Permitted claim: within this frozen constructed grid, exact equality of the named confusion-count/volume metrics does not identify the stated clearance proxy. Forbidden claims include natural-error prevalence, clinical harm, safety, deployment, validated clinical cutoffs, equality of surface/topology metrics, or superiority of a new metric.

8 Data Use and Availability

The study uses de-identified research datasets released by their maintainers under their source access terms. No new participant recruitment, intervention, or clinical decision occurred. The submission does not redistribute raw CBCT volumes, labels, or checkpoint weights; source permissions and institutional statements remain those of the cited dataset/model releases. DentVoxel v1 is attributed to You Zhou, Yan Xu, and Mihai-Adrian Tarce, is licensed CC BY 4.0, and is identified by DOI `https://doi.org/10.6084/m9.figshare.31239889.v1`.