Supplementary Information

Public regulatory registries remain registration infrastructures rather than governance infrastructures for AI-enabled medical devices

Yu Han, Aaron Ceross

This Supplementary Information accompanies the manuscript. All machine-readable artefacts referenced below are in the OSF/Zenodo deposit, organised under `data/`, `code/`, `results/`, and `methodology/`. Dates follow the four terms defined in Methods: audit cut-off (2026-05-01), source-archive lock, stress-test reference-data download (2026-05-14), and post-review correction.

Appendix A Source inventory and operational capture notes

Twenty-four public regulatory sources across five jurisdictions were locally archived on 2026-05-01 (HTML capture + screenshot + `meta.json` per source). The complete machine-readable manifest is deposited as `data/sources/archive_manifest.csv` and `sources_inventory.json`; capture status, HTTP status, retrieval timestamp, and stored paths are recorded per source.

Access mode by jurisdiction. Bulk-downloadable structured data: FDA 510(k), PMA, De Novo, GUDID, MAUDE, Recalls (openFDA REST API + download partitions), the FDA AI-Enabled Device List (CSV), and the NMPA UDI full release (XML). Search-interface-only structured data: PMDA Shonin, the ARTG, and the four first-cohort EUDAMED modules. Documentation-only at cut-off: EUDAMED Vigilance and Market Surveillance modules.

Operational capture problems (moved here from the main text). (i) FDA `accessdata.fda.gov` search front-ends (510(k), PMA, De Novo, MAUDE, Recalls) timed out on headless capture because the pages navigate/rewrite content client-side; the underlying data was obtained via openFDA bulk endpoints, so these are recorded as search front-ends, not as the data source. (ii) NMPA endpoints returned HTTP 412 (`nmpa.gov.cn`, `datasearch`) or partial captures (`udi.nmpa.gov.cn`) from the audit network; landing-page HTML was rescued via the Wayback Machine CDX API and flagged `cutoff_status = wayback`. The NMPA UDI *bulk release* was nonetheless retrieved directly (see Appendix **??**). (iii) TGA URLs (ARTG, AI list, DAEN) returned `ERR_HTTP2_PROTOCOL_ERROR`/404 on capture because the TGA Akamai CDN refuses non-browser user agents; canonical landing pages were recorded and remain machine-readable at the live database front-ends. (iv) The EUDAMED interface is a single-page web application whose module pages are JavaScript routes rather than separately archivable URLs; module timelines are sourced to Commission Decision (EU) 2025/2371 and the Commission's EUDAMED overview.

Appendix B Inter-annotator disagreement patterns

Headline reliability (pre-adjudication, 100 cells both-coded) is in the main text (Table 3). This appendix reports the full breakdown.

Table Appendix B.1: Reliability by dimension (pre-adjudication).

Dimension	n	Raw agr.	κ	Interpretation
Availability score (0–3, linear-weighted)	100	0.780	0.724	Acceptable
Gap type (5 categories)	93	0.796	0.703	Acceptable
Confidence rating (3 categories)	100	0.740	0.555	Cautious

Table Appendix B.2: Per-function agreement on availability score (pre-adjudication).

Function	n	Raw agr.	Weighted κ	Interpretation
F1 Identity	20	0.650	0.605	Cautious
F2 Version & Change	20	0.950	0.706	Acceptable
F3 Evidence	20	0.800	0.000	Codebook failure (low variance)
F4 Linkage	20	0.750	0.697	Acceptable
F5 AI Specificity	20	0.750	0.565	Cautious

Category prevalence by gap type: structuring 0.409, schema 0.339, publication/access 0.151, linkage 0.086, population 0.016. Confidence: Medium 0.540, High 0.335, Low 0.125. The F3 weighted-κ of 0.000 reflects near-zero score variance (almost all F3 cells coded 0–1), not coder disagreement; raw agreement is 0.800. As an adjudication diagnostic only, post-revision Cohen's κ on the availability score is 0.854 (raw 0.890); the pre-adjudication figure remains the headline measure.

Appendix C Adjudication log

Of 22 pre-adjudication disagreements, 18 resolved by one annotator accepting the other's reading after re-application of the codebook, and 4 by the higher-coding annotator revising downward. Two two-step availability disagreements (EU F2a, EU F4d) were resolved under the codebook Section 6 mismatch rule and the Section 5 governance-probe note respectively; no codebook change was required. The complete per-cell log (both annotators' codes, the final code, the codebook section invoked, and the decisive evidence source) is deposited as `data/coding/adjudication_log.csv` (human-readable version `adjudication_log.md`).

Appendix D Aggregate jurisdiction totals (diagnostic)

Per the manuscript's rationale, requirement- and function-level deficit patterns are the substantive findings; jurisdictional totals are diagnostic, not a league table. The full adjudicated 5×20 matrix with per-cell scores, confidence ratings, and gap types is deposited as `results/full_5x20_matrix.csv`. For reference, the China column carries a mean availability of 0.55/3, dominated by schema and publication/access gaps. Low-confidence cells (12/60 in the core; concentrated in Japan, China F2–F5, and Australia F4c) are reported here as indicative and are excluded from the High/Medium-confidence sensitivity analysis.

Appendix E FDA PMA per-record classifications and De Novo list

PMA pathway (n=17). Each P-numbered entry on the FDA AI/ML-Enabled Device List was matched into openFDA `/device/pma.json` (100% match) and classified by the strongest AI signal in primary searchable fields. Six carried a narrow AI term; **five of the six are mammographic computer-aided detection devices carrying FDA product code `MYN`** (the radiological computer-assisted-detection code), the sixth being an opto-acoustic breast-imaging device (`QNK`). The "CAD" family was credited as a narrow term only where product code / approval statement indicated computer-aided detection or diagnosis; the manufacturer name "iCAD" was not treated as narrow evidence.

Table Appendix E.3: FDA PMA AI-identifiability per-record classification (n=17).

PMA	Trade name	Product code	Strongest tier
P230022	Jewel Patch Wearable Cardioverter Defi.	MVK	broad
P140011/S008	MAMMOMAT B.brilliant w/ Tomosynthesis	OTE	broad
P210011	xT CDx	PQP	none
P210015	Avive Automated External Defibrillator	MKJ	broad
P200003	Imagio Breast Imaging System	QNK	narrow
P150046	Nevisense	ONV	none

PMA	Trade name	Product code	Strongest tier
P160009	PowerLook Tomo Detection Software	MYN	narrow (CAD)
P150043	QVCAD System (Version 3.3)	MYN	narrow (CAD)
P110014	MarginProbe	OEE	none
P090012	MELAFIND	OYD	none
P040028	Cervical Imaging System	MWM	none
P010034	Second Look Digital CAD System	MYN	narrow (CAD)
P000041	RapidScreen RS-2000D	MYN	narrow (CAD)
P980025	Logicon Caries Detector	MYN	none
P970058	ImageChecker D	MYN	narrow (CAD)
P940029	PapNet Testing System	MNM	broad
P950009	BD FocalPoint GS Imaging System	MNM	broad

Pooled across 510(k) and PMA (n=1,393): 41 (2.9%) carry a narrow term, 699 (50.2%) carry no AI signal. Per-record raw data: `data/fda/pma_identifiability_full_raw.csv`.

De Novo pathway. The 37 De Novo (`DEN`-prefixed) entries on the FDA AI list were not included in the scripted openFDA primary-field test (no `/device/de_novo.json` endpoint at cut-off). FDA De Novo records are publicly searchable and downloadable through a separate FDA database not incorporated into the scripted protocol. The identifier list is deposited in `data/fda/fda_ai_list_published_20260501.csv` (rows with submission numbers beginning `DEN`).

Appendix F Product-family traceability: sampling and per-record notes

Six known multi-submission AI-device families were purposively sampled as a stress test of structured product-line traceability in the richest public API: Aidoc Medical (n=34), iCAD (n=15), HeartFlow (n=8), Caption Health (n=6), Cleerly (n=4), Paige.ai (n=3); total n=70 510(k) records. Selection was purposive (manufacturers known to have multi-submission histories), not a representative sample. **Zero of 70** expose a structured predecessor/predicate/related-510(k) field in `/device/510k.json`; 7/70 carry a version string in `device_name`; all 70 can be grouped by `applicant` but without temporal lineage direction. Per-record openFDA dump: `data/fda/product_family_n70_raw.json`; rebuilder script in `code/`.

Appendix G Exploratory gap-profile similarity

As an exploratory analysis only, pairwise Jaccard similarity on binarised (met/unmet) gap profiles across the five jurisdictions is computed from the adjudicated matrix (`results/full_5x20_matrix.csv`). We draw no causal claim about regulatory tradition from a five-jurisdiction, hand-coded sample; the analysis is reported for transparency and is not used in any substantive conclusion.

Appendix H Structured-only sensitivity analysis

As a third pre-specified restriction (in addition to the core-only and High/Medium-confidence restrictions reported in the main text), the 0–3 availability scale was collapsed to a binary *structured* indicator (scores 2–3 = structured; 0–1 = not structured), to test whether the headline finding depends on the four-point scale rather than on a structured-vs-narrative distinction. Under this collapse, 13 of 100 cells (13%) are structured. The structured share is generated deterministically from the adjudicated matrix by `code/results_pipeline/regen_structured_only_sensitivity.py`.

Because the structured-only collapse largely duplicates the distinction already encoded in the 0–3 scale (it merges scores 2 and 3, and separately merges 0 and 1), it is reported as a descriptive robustness check rather than as a primary inferential result. It does not alter the conclusion that software-change (F2), evidence (F3), and AI-specificity (F5) information is not exposed as structured public data in any of the five jurisdictions.

Table Appendix H.4: Structured-only sensitivity: share of cells scoring 2–3 (structured) by function and by jurisdiction. The direction of the main finding is unchanged: F2 and F3 expose no structured fields, F5 almost none, and the USA is closer to structured exposure than the other four jurisdictions.

Function	**Structured %**	**Jurisdiction**	**Structured %**
F1 Identity	35%	USA	20%
F2 Version/Change	0%	EU	10%
F3 Evidence	0%	China	15%
F4 Linkage	25%	Japan	10%
F5 AI specificity	5%	Australia	15%
Overall	**13%**		

Appendix I Worked public-record traceability test

To make the audit concrete for a single device, we ran the five governance questions against the richest public record environment in the audit (the FDA), using a real AI-enabled device in our data: Aidoc's *BriefCase* triage family (34 510(k) records in the product-family probe). This is a test of what a public user can recover from public records, not an assessment of the device or its manufacturer, and it makes no claim about safety or about information the FDA holds internally (Table Appendix I.5). Even in the richest public environment, three of the five questions cannot be answered from structured public records, and a fourth only through manual PDF reading.

Table Appendix I.5: Worked public-record traceability test for an FDA-cleared AI device family (Aidoc *BriefCase*, 34 510(k) records). "Recoverable?" refers to recovery from public FDA records by an external user; answers follow the audit's USA-column findings and the product-family probe. The exercise is a traceability test, not a safety or compliance judgment.

Governance question	**Public source checked**	**Recoverable?**	**Form of evidence**	**Req(s)**
Is it AI-enabled?	FDA AI list; primary 510(k) record	Yes via list; no via primary record	Curated supplementary list; primary `device_name` ("BriefCase") carries no AI term	F1a, F5a
Which version is currently marketed?	510(k) record, summary, device name	No	No structured current-version field	F2a
Has it changed since authorisation?	Successive 510(k)s, summaries	Inferential only	0/34 records expose a structured predecessor/predicate field; lineage only by date-ordered same-applicant submissions	F2b–c, F4d
What population/data was it validated on?	Decision summary	Partial	Narrative in 510(k) Summary PDF only	F3b–c, F5c
Can authorisation be linked to recalls / adverse events?	Recalls, MAUDE (openFDA)	Partial / manual	Separate systems, not joined to the record by a stable key	F1b, F4b–c

Appendix J Lifecycle traceability summary

For each jurisdiction, Table Appendix J.6 records the stated lifecycle instrument or rhetoric alongside the audit-observed availability of the four version-and-change requirements (F2a–d), reporting the qualitative

observations directly rather than on a constructed numeric axis. Lifecycle posture is near-universal; public version-and-change traceability is almost entirely absent, present only at the narrative/inferential level in the USA.

Table Appendix J.6: Qualitative lifecycle traceability summary. For each jurisdiction we record the relevant lifecycle instrument or rhetoric, and the audit-observed availability of the four version-and-change requirements (F2a–d). ✓= structured field directly observed; (✓) = narrative or inferential availability; n/o = not observed.

Jurisdiction	**Lifecycle instrument / rhetoric**	**F2a**	**F2b**	**F2c**	**F2d**
USA	PCCP guidance (final 2025); 510(k) supplements	(✓)	(✓)	(✓)	n/o
EU	EUDAMED "living picture"; MDR change framework	n/o	n/o	n/o	n/o
China	UDI updates; product registration changes	n/o	n/o	n/o	n/o
Japan	PMDA SaMD framework; partial change approvals	n/o	n/o	n/o	n/o
Australia	TGA software change guidance; AI list	n/o	n/o	n/o	n/o