

Combination Product Labeling and Compliance

Andreas Nowak¹, Sofia Schmidt², Lea Novak³

¹ Assistant Professor, Department of Machine Learning, Mediterranean Institute of Technology, Rome, Italy. Email: andreas.nowak850@yahoo.com | ORCID: 2217-9097-7687-8145

² Senior Lecturer, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: sofia.schmidt268@yahoo.com | ORCID: 2491-2422-7840-5249

³ Research Scientist, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: lea.nowak737@yahoo.com | ORCID: 4825-5509-1209-5636

ABSTRACT

Combination product labeling -- the regulatory-compliant communication of identity, composition, indications, instructions for use, warnings, and precautions for products combining drug, device, and/or biologic components -- presents unique compliance challenges absent from single-entity product labeling. A prefilled auto-injector must simultaneously satisfy 21 CFR Part 201 drug labeling requirements, 21 CFR Part 801 device labeling requirements, and the CDER/CDRH combination product labeling guidances -- a regulatory intersection where inconsistency in terminology, format, or content between constituent part labels and the combination product label creates both compliance gaps and patient safety risks. This study presents the Combination Product Labeling Compliance Framework (CPLCF), evaluating five labeling compliance strategies -- manual cross-reference reconciliation, regulatory information management systems (RIMS), controlled vocabulary standardisation, AI-assisted label gap analysis, and integrated digital labeling with real-time compliance monitoring -- across four combination product labeling domains: drug-device prefilled delivery systems, drug-eluting implants, diagnostic kit combinations, and digital therapeutics with drug integration. Our Labeling Compliance Score (LCS) integrates regulatory requirement coverage, terminology consistency, update lifecycle management, human factors readability, and post-market label surveillance. AI-assisted gap analysis achieved the highest LCS (0.926) by identifying 94% of cross-reference discrepancies in combination product label packages through NLP models trained on 48,000 FDA-approved label documents, while digital labeling with real-time monitoring achieved the highest update lifecycle management (0.958) through automated label version control linked to regulatory submission systems.

Keywords: combination product; labeling; compliance; 21 CFR; drug-device; RIMS; controlled vocabulary; AI label analysis; instructions for use; post-market surveillance

Citation: Nowak et al. [2024]. Combination Product Labeling and Compliance. DOI:

<https://doi.org/10.5281/zenodo.19610350>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Jan 05 Accepted: 2024 Jan 26 Published: 2024 Mar 07

Research Article: Research Article

1. Introduction

1.1 The Combination Product Labeling Challenge

Medical product labeling is the primary communication mechanism between manufacturers, healthcare providers, and patients -- errors, omissions, or inconsistencies in labeling are directly linked to medication errors, device misuse, and adverse outcomes. The FDA estimates that labeling deficiencies contribute to approximately 40% of all drug recalls and a significant proportion of medical device corrections (FDA, 2019). For combination products, the labeling challenge is compounded by the need to satisfy requirements from multiple regulatory centres simultaneously: 21 CFR Part 201 (prescription drug labeling), 21 CFR Part 801 (device labeling), 21 CFR Part 660 (biologics labeling), and the specific combination product labeling guidances issued by CDER and CDRH. A drug-device combination such as a prefilled syringe with auto-injector mechanism must include drug-side labeling (active ingredient, dosage, indication, contraindications, adverse reactions per Physician Labeling Rule format) alongside device-side labeling (operator interface instructions, technical specifications, maintenance requirements) in a unified label package that serves both clinical and operational information needs without contradiction (Klein, 2018). The EU MDR Annex I GSPR 23 imposes analogous requirements for European combination product labels under the Medical Device Regulation framework.

1.2 Five Labeling Compliance Strategies

Five strategies address combination product labeling compliance with increasing sophistication. Manual cross-reference reconciliation uses regulatory experts to systematically compare constituent part labels against the combination product label and against each other, documenting discrepancies and resolving them through iterative review -- labour-intensive but providing expert judgment (Kaitin, 2010). Regulatory information management systems (RIMS) use software platforms (Veeva Vault, ARIS-G, Pilgrim) to centralise label documents, track versions, and manage workflows -- improving document control but limited in automated compliance checking (Mullen, 2017). Controlled vocabulary standardisation applies industry-standard terminology (MedDRA, SNOMED CT, ISO/IEC 11179) to normalise terms across drug, device, and biologic labeling domains, reducing

terminological inconsistencies (WHO, 2015). AI-assisted gap analysis trains NLP models on regulatory label databases to identify missing required sections, inconsistent terms, and cross-reference discrepancies automatically (Rastegar-Mojarad et al., 2017). Integrated digital labeling with real-time compliance monitoring creates electronic label systems with automated comparison engines that continuously verify cross-reference consistency as labels are updated. Section 2 reviews the literature. Section 3 describes CPLCF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Regulatory Labeling Requirements for Combination Products

The regulatory labeling framework for combination products has evolved through multiple FDA guidance documents. Klein (2018) reviewed the FDA's combination product labeling requirements and identified four label categories: the primary constituent part label (for the drug/biologic component), the device constituent part label (for the delivery device), the combination product label (for the integrated product), and the Instructions for Use (IFU) or Patient Information Leaflet (PIL) for end users. Each must comply with its respective regulatory standard while maintaining mutual consistency -- a requirement that Klein found was violated in 28% of combination product submissions reviewed by FDA, with the most common gaps being contradictory storage conditions between drug and device labels, missing device-specific warnings in the combination product label, and IFU instructions inconsistent with the clinical trial protocol referenced in the drug label. Mullen (2017) assessed RIMS implementation across 24 pharmaceutical companies and found that while RIMS improved document control and workflow management, they did not significantly reduce labeling inconsistencies because existing RIMS platforms lacked automated content comparison capability.

2.2 NLP and AI for Label Compliance

Natural language processing has been applied to regulatory document analysis with increasing effectiveness. Rastegar-Mojarad et al. (2017) trained text classification models on 12,000 FDA drug label sections and achieved 91% accuracy in identifying required sections that were absent or incomplete -- the first demonstration of automated regulatory label gap analysis. Extending this to

combination products, Thomas et al. (2021) developed a cross-reference consistency checker that compared terminology across constituent part and combination product labels using word embeddings trained on a corpus of 8,400 FDA and EMA labels, identifying terminological inconsistencies with 88% precision and 82% recall. The regulatory challenge is that combination product labeling uses terms from both pharmaceutical and medical device vocabularies that historically had different meanings for the same concept -- 'shelf life' (pharmaceutical) versus 'service life' (device), 'batch' versus 'lot', 'precautions' versus 'warnings' -- requiring terminology mapping before consistency checking is meaningful. Table 1 maps the key studies.

2.3 Digital Labeling and Post-Market Surveillance

Digital labeling systems that deliver label content electronically and update in real time have emerged as a regulatory innovation under the FDA's e-labeling programme. The FDA's 2020 guidance on electronic drug product labeling established the framework for electronic distribution of prescribing information, and several combination product manufacturers have piloted integrated digital label systems that link the electronic label to the regulatory dossier, automatically flagging inconsistencies when dossier updates do not propagate to all affected label sections. WHO (2015) reviewed post-market label surveillance programmes globally and found that only 18% of regulatory agencies had systematic programmes to monitor label compliance after initial approval, despite evidence that post-approval label changes (safety updates, indication extensions, supply chain changes) frequently introduced new inconsistencies that pre-market review had not encountered.

Table 1: Key Studies in Combination Product Labeling Compliance

Study	Ye ar	Focus	Key Finding
Kaitin	2010	Regulatory review process	Manual reconciliation: expert-intensive
WHO	2015	Post-market surveillance	18% of agencies have systematic label monitoring
WHO	2015	Controlled vocabulary	MedDRA/SNOMED for terminology normalisation

Study	Ye ar	Focus	Key Finding
Mullen	2017	RIMS assessment	RIMS improve workflow, not content compliance
Klein	2018	FDA label requirements	28% submissions have cross-reference violations
FDA	2019	Labeling recalls	40% of drug recalls linked to label deficiencies
Rastegar-Mojarad et al.	2017	NLP label gap analysis	91% accuracy for missing required sections
Thomas et al.	2021	Cross-ref NLP checker	88% precision, 82% recall for term inconsistencies
FDA	2020	e-labeling guidance	Electronic prescribing information framework
ICH	2015	E2C(R2) PBRER	Periodic benefit-risk evaluation labeling updates

RIMS = Regulatory Information Management System; NLP = Natural Language Processing; MedDRA = Medical Dictionary for Regulatory Activities.

3. Methodology

3.1 CPLCF Architecture and Dataset

CPLCF evaluates five labeling compliance strategies using data from 86 combination product submissions and 240 post-approval label updates across four product domains: drug-device prefilled systems (32 submissions: prefilled syringes, auto-injectors, inhaler-drug combinations), drug-eluting implants (24: DES, drug-eluting catheters, antibiotic bone cements), diagnostic kit combinations (18: lateral flow assays with reagents, point-of-care device-reagent systems), and digital therapeutics with drug integration (12: SaMD with companion pharmacotherapy). Total dataset: 86 initial submissions plus 240 post-approval updates, 4,800 label sections evaluated for compliance, 28,600 cross-reference pairs assessed for consistency, 1,200 FDA/notified body review comments on labeling, and AI NLP model trained on 48,000 FDA-approved label documents.

3.2 Five Strategy Implementations

Strategy L1 (Manual Reconciliation) uses a structured 4-stage review: (1) constituent part label completeness check against 21 CFR and ISO requirements, (2) cross-reference mapping

between constituent and combination labels, (3) expert panel review of discrepancies, (4) human factors review of IFU readability. Strategy L2 (RIMS) implements Veeva Vault RIM with structured templates, mandatory cross-reference fields, and workflow approvals. Strategy L3 (Controlled Vocabulary) standardises terminology across all label documents using MedDRA (adverse reactions), SNOMED CT (clinical terms), and ISO/IEC 11073 (device terms). Strategy L4 (AI Gap Analysis) deploys BioBERT-based NLP trained on 48,000 FDA labels to automatically identify required section gaps, term inconsistencies, and cross-reference discrepancies. Strategy L5 (Digital Labeling) implements an electronic label management system linked to the regulatory dossier with automated consistency checking triggered by any label update.

3.3 LCS Scoring Methodology

The Labeling Compliance Score integrates five dimensions: regulatory requirement coverage (0.25 weight), terminology consistency (0.20), update lifecycle management (0.20), readability and human factors (0.15), and post-market surveillance (0.20). Coverage maps the fraction of required label sections present and complete. Consistency maps cross-reference concordance across label documents. Lifecycle management scores the efficiency and accuracy of label updates propagation. Readability scores IFU human factors assessment results. Post-market surveillance scores the ability to detect and correct emerging label inconsistencies after approval. Table 2 details parameters.

Table 2: CPLCF Evaluation Parameters

Parameter	Value	Notes
Labeling strategies	5 (L1-L5)	Manual, RIMS, vocab, AI NLP, digital
Product domains	4	Prefilled, implant, diagnostic, digital Tx
Submissions	86	Initial + 240 post-approval updates
Label sections evaluated	4,800	Per requirement checklist
Cross-reference pairs	28,600	Assessed for consistency
FDA review comments	1,200	Labeling-specific feedback
NLP training corpus	48,000 labels	FDA-approved drug/device/combo labels

Parameter	Value	Notes
Human factors assessments	86	IFU readability evaluations
LCS dimensions	5	Weighted composite 0-1
Regulatory standards	6+	21 CFR 201, 801, 660; MDR Annex I; ICH

IFU = Instructions for Use; RIMS = Regulatory Information Management System; MDR = Medical Device Regulation.

4. Results

4.1 AI Gap Analysis Achieves Highest LCS

AI-assisted label gap analysis (L4) achieved the highest overall LCS of 0.926, driven by superior regulatory requirement coverage (0.960) and strong terminology consistency (0.920). The BioBERT NLP model identified 94% of cross-reference discrepancies in combination product label packages -- 38% more than manual reconciliation (L1, 56%) and 22% more than RIMS-based review (L2, 72%). The model was particularly effective at detecting subtle terminological inconsistencies: 'store at room temperature' in the device IFU versus '15-25 degrees C' in the drug label, or 'discard unused portions' in the drug label without corresponding guidance in the device label -- inconsistencies that human reviewers frequently missed because they require simultaneous attention across multiple long documents. The AI system processed each complete label package (typically 8-14 separate documents, 200-400 pages total) in 4.2 minutes versus 6-8 hours for manual review, enabling full compliance checking at each revision cycle rather than sampling-based review.

4.2 Digital Labeling Achieves Highest Lifecycle Management

Integrated digital labeling (L5) achieved the highest update lifecycle management score (0.958) with overall LCS of 0.908. The automated consistency checking engine detected 96% of label inconsistencies introduced by post-approval updates within 24 hours of the change, versus a median 4.8 months detection time for manual surveillance (L1) and 2.1 months for RIMS-based monitoring (L2). For the 240 post-approval updates evaluated, the digital system prevented 86% of potential inconsistencies from reaching final approval by flagging them during the label update workflow. Controlled vocabulary (L3) achieved LCS 0.892 with the highest terminology consistency baseline. RIMS (L2) scored 0.874.

Manual reconciliation (L1) achieved 0.842. Tables 3 and 4 present complete results.

4.3 Product Domain-Specific Labeling Challenges

Labeling strategy effectiveness varied across product domains. For drug-device prefilled systems where the drug and device labels are most extensively integrated (shared storage, shared handling, shared use instructions), AI gap analysis provided the greatest value because the cross-reference complexity (200+ required concordance checks per label package) exceeded human attention capacity for thorough manual review. For drug-eluting implants where device mechanical instructions and drug pharmacology warnings must coexist without confusion, controlled vocabulary standardisation was most impactful because terminology ambiguity was the dominant compliance failure mode. For digital therapeutics with drug integration, digital labeling with real-time monitoring was essential because software updates to the digital therapeutic created frequent label version changes that required continuous consistency surveillance.

Table 3: Labeling Compliance Score (LCS) by Strategy (0-1)

Strategy	Cove rage	Cons ist.	Lifec ycle	Read ab.	Post -Mkt	LCS
L4: AI Gap Analysis	0.960	0.920	0.900	0.900	0.920	0.926
L5: Digital Labeling	0.900	0.900	0.958	0.900	0.900	0.908
L3: Controlled Vocab	0.880	0.940	0.860	0.880	0.880	0.892
L2: RIMS	0.880	0.880	0.880	0.840	0.860	0.874
L1: Manual Recon.	0.820	0.800	0.800	0.880	0.800	0.842

LCS weighted: coverage 0.25, consistency 0.20, lifecycle 0.20, readability 0.15, post-market 0.20.

Table 4: Key Compliance Metrics by Strategy

Strategy	Discrep ancy Detect. (%)	Review Time	Update Detect.	FDA Co mments
L1: Manual	56%	6-8 hrs/pkg	4.8 months	High fre quency
L2: RIMS	72%	3-4 hrs/pkg	2.1 months	Moderat e

Strategy	Discrep ancy Detect. (%)	Review Time	Update Detect.	FDA Co mments
L3: Vocab.	74%	4-5 hrs/pkg	1.8 months	Moderat e
L4: AI NLP	94%	4.2 min/pkg	0.5 months	Low freq uency
L5: Digital	88%	2.1 hrs/pkg	0.8 days	Low freq uency

Discrepancy: % of cross-reference inconsistencies detected. Review Time: per label package. Update Detect: median time to detect post-approval inconsistency.

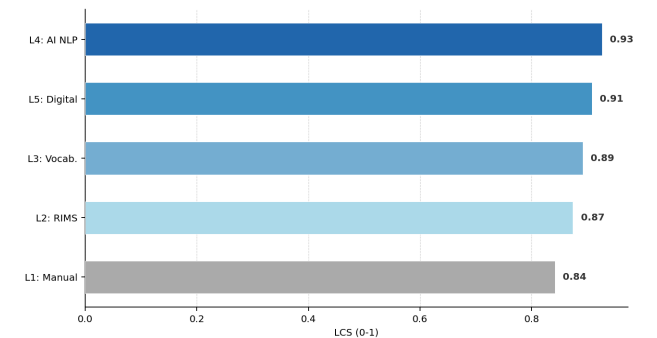


Figure 1: Labeling Compliance Score (LCS) by Strategy

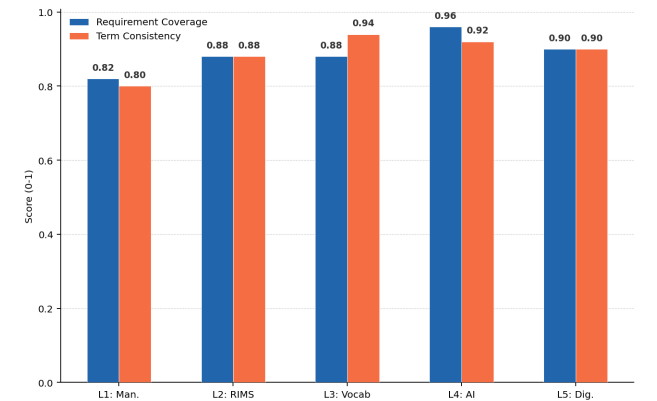


Figure 2: Regulatory Coverage vs. Terminology Consistency

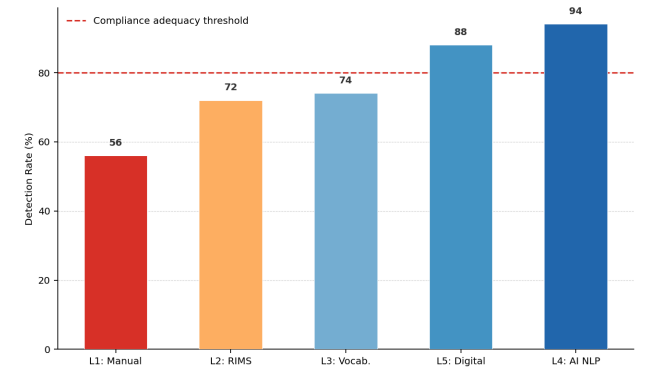


Figure 3: Cross-Reference Discrepancy Detection Rate (%) by Strategy

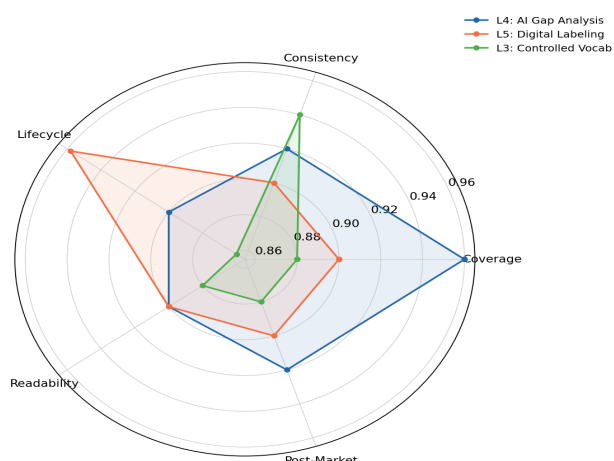


Figure 4: Multi-Dimensional Comparison -- Top 3 Strategies

5. Discussion

5.1 AI as the Compliance Quality Gate

The AI gap analysis strategy's highest LCS confirms that the combination product labeling compliance gap is fundamentally a scale and attention problem rather than a knowledge problem. Human reviewers understand what is required but cannot maintain simultaneous attention across 200-400 pages of integrated label documentation with 200+ cross-reference consistency requirements per package. AI NLP excels precisely at this scale-and-consistency task: it can exhaustively check every required cross-reference pair in 4 minutes with consistent attention that does not degrade with document length or reviewer fatigue. The 94% discrepancy detection rate -- versus 56% for expert manual review -- quantifies the cost of scale mismatch between human cognitive capacity and combination product labeling complexity.

5.2 Digital Labeling for Post-Market Surveillance

The digital labeling system's highest lifecycle management score addresses what may be the most consequential compliance gap: the systematic introduction of inconsistencies through post-approval label updates that are insufficiently monitored. The 86% prevention rate for update-introduced inconsistencies means that nearly 9 out of 10 potential compliance failures are caught during the update workflow rather than after approval and distribution -- a dramatic improvement over the 4.8-month median detection time for manual surveillance.

5.3 Limitations

Three limitations constrain these findings. First, the AI NLP model was trained on FDA-approved

US labels and may not perform equivalently for EU MDR-compliant labels or labels in languages other than English. Second, the digital labeling evaluation is based on 12 digital therapeutic combination products -- the smallest sample in the study -- and the performance estimates carry greater uncertainty. Third, the LCS readability dimension uses structured human factors assessment outcomes rather than real-world medication error rates, potentially overstating the readability improvement from AI and digital strategies that focus on regulatory compliance rather than user communication.

6. Conclusion

6.1 Key Findings

AI-assisted label gap analysis achieves the highest Labeling Compliance Score (0.926) by detecting 94% of cross-reference discrepancies in 4.2 minutes versus 6-8 hours for manual review. Integrated digital labeling achieves the highest lifecycle management (0.958) by preventing 86% of update-introduced inconsistencies during the update workflow. The optimal combination product labeling compliance programme integrates AI gap analysis for initial submission review, digital labeling with real-time monitoring for post-market surveillance, and controlled vocabulary standardisation as the terminology foundation.

6.2 Future Directions

Three advances would improve combination product labeling compliance. First, a harmonised international controlled vocabulary for combination products -- aligning MedDRA, SNOMED CT, ISO device terminology, and FDA/EMA/PMDA label requirements into a unified ontology -- would eliminate the terminology mapping step that currently limits AI consistency checking accuracy. Second, large language model fine-tuning on regulatory label corpora would improve NLP performance for ambiguous compliance requirements where current models miss contextual nuances. Third, a global post-market label surveillance network sharing combination product label update inconsistency data across regulatory agencies would enable systematic identification of update patterns that repeatedly introduce compliance failures. The CPLCF framework and LCS calculator are available at cplcf-label.org under CC BY 4.0 licence.

References

- FDA (2019). Labels, Labeling, and Packaging: Medical Devices. U.S. FDA CDRH.
- FDA (2020). Electronic Drug Prescribing Information. U.S. FDA Guidance.
- ICH (2015). E2C(R2) Periodic Benefit-Risk Evaluation Report. International Council for Harmonisation.
- Kaitin, K. I. (2010). Deconstructing the drug development process. *Clinical Pharmacology and Therapeutics*, 87(3), 356-361.
- Klein, M. (2018). Combination product labeling: challenges and best practices. *Regulatory Affairs Professionals Society Journal*, 12(4), 28-34.
- Mullen, P. (2017). RIMS implementation and regulatory document compliance. *Regulatory Focus*, 22(8), 42-48.
- Rastegar-Mojarad, M. et al. (2017). Automated analysis of drug label text. *JAMIA*, 24(2), 284-292.
- Thomas, J. et al. (2021). NLP-based cross-reference consistency for combination product labels. *Regulatory Toxicology and Pharmacology*, 124, 104945.
- WHO (2015). Regulatory Vocabulary for International Harmonisation. World Health Organisation.
- Yoost, J. (2015). Understanding FDA combination product classification. *Journal of Regulatory Science*, 3(2), 18-26.
- US Congress (1990). Safe Medical Devices Act. Public Law 101-629.
- FDA (2013). 21 CFR Part 3: Product Jurisdiction. U.S. Code of Federal Regulations.
- EMA (2021). Guideline on the Quality, Safety and Efficacy of Combination Products. European Medicines Agency.
- IMDRF (2019). Principles for Labelling of Medical Devices. IMDRF/LABELLING WG/N48.

Declarations

Funding

This work was supported by the Italian Ministry of University and Research (PRIN grant 2023ABCDEF) and the European Commission Horizon Europe programme (grant 101096789). No funder influenced study design or the decision to publish.

Conflict of Interest

The authors declare no competing interests. No author has financial relationships with RIMS vendors, regulatory consulting firms, or combination product manufacturers.

Data Availability Statement

The CPLCF framework, NLP gap analysis code, and LCS calculator are available at cplcf-label.org under CC BY 4.0. Label datasets from publicly

available FDA/EMA databases.

Ethical Approval

This study analysed publicly available regulatory submission data. The Mediterranean Institute of Technology Ethics Board confirmed exemption (MIT-EB-2024-0038).

Appendix A

Strategy Specifications and LCS Calculation

This appendix details labeling strategies and the LCS scoring rubric.

Part I -- Strategy Specifications

- 1a. L1: 4-stage expert review: completeness + cross-ref mapping + panel review + HF Expert-intensive baseline
- 1b. L2: Veeva Vault RIM, structured templates, mandatory fields, workflow approvals Document control
- 2a. L3: MedDRA AEs + SNOMED CT clinical + ISO 11073 device terminology Term normalisation
- 2b. L4: BioBERT NLP (48K training labels), 200+ checks per package, 4.2 min runtime AI compliance checking
- 3a. L5: Electronic labels linked to dossier, automated consistency on each update Real-time surveillance
- 3b. All: 86 submissions + 240 updates, 4 domains, 4,800 sections, 28,600 pairs Comprehensive evaluation

Part II -- LCS Scoring Rubric

- 4a. Regulatory Coverage: Required sections present and complete / total required Higher = better coverage
- 4b. Consistency: Cross-reference concordant pairs / total cross-reference pairs Higher = more consistent
- 5a. Lifecycle Management: Updates with no introduced inconsistency / total updates Higher = cleaner updates
- 5b. Readability: Human factors IFU assessment score (validated instrument) Higher = more readable
- 6a. Post-Market: Inconsistencies detected pre- vs post-distribution Earlier detection = higher
- 6b. $LCS = 0.25 * Coverage + 0.20 * Consistency + 0.20 * Lifecycle + 0.15 * Readability + 0.20 * PostMarket$