International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

The Probabilistic Sentinel System: A Formally Specified Scoring Engine for Privacy-Safe Member Identity Resolution in Healthcare MDM

Vikram Nafria

Abstract: Healthcare payers routinely fail to unify a single member's history across enrollment, claims, pharmacy, and care management systems, and the operational cost of that failure falls into two opposite failure modes: fragmentation that hides a member's full history, and over-consolidation that can expose Protected Health Information (PHI) to the wrong record. This article presents the Probabilistic Sentinel System (PSS), a practitioner-informed scoring and governance framework for managing that tradeoff in healthcare payer master data management (MDM). PSS extends the Fellegi-Sunter probabilistic linkage model with differential attribute weighting across eight member identifiers, an explicit negative penalty for Social Security Number (SSN) conflicts distinct from a missing SSN, calibrated decision thresholds separating auto-link, manual-review, and auto-reject zones, and an override layer that routes behavioral health, HIV, and substance-use records to human review regardless of numeric score. The framework is motivated by MDM defect patterns drawn from payer operations, but all quantitative results in this paper come from a fully synthetic payer-style benchmark containing 101,000 labeled candidate pairs. The synthetic evaluation supports reviewer inspection through confidence intervals, threshold tables, subgroup fairness results, candidate-pair examples, and audit-log examples without exposing PHI, member-level records, steward notes, proprietary matching rules, or internal governance artifacts.

Keywords: *Fellegi-Sunter Model; Healthcare Identity Resolution; HIPAA Privacy Risk; Master Data Management; Probabilistic Record Linkage*

1. Introduction

In payer MDM operations, a small enrollment update-a changed middle initial, a corrected mailing address-can be enough for identity resolution logic to treat two records of the same member as belonging to different people. When that happens, the member's clinical and claims history splits across two enterprise IDs, and neither record shows the full picture. This defect pattern recursed in production incident queues, steward review backlogs, and audit findings, and it has also been documented in master data management case studies outside healthcare [1].

The opposite failure carries a sharper cost. Two members who share a household address, a common last name, or a reused identifier get consolidated under a single enterprise ID (EID), and one member's behavioral health, HIV status, or substance-use claims history becomes visible inside the record view of an unrelated person. This is a Protected Health Information (PHI) exposure event under the HIPAA Privacy Rule [2]. It triggers breach assessment, notification obligations, and reputational exposure that a payer's compliance office tracks closely [3], [4].

Most production MDM platforms resolve member identity with deterministic matching rules or off-the-shelf probabilistic scoring that treats every attribute conflict the same way. Many scoring configurations treat a missing SSN and a populated, conflicting SSN as equivalent evidence, or ignore the distinction entirely. PSS treats them as fundamentally different signals. The deterministic-versus-probabilistic tradeoff discussed further below does not resolve this problem so much as relocate it: deterministic rules

Independent Researcher, USA

** Corresponding Author Email: vikramnafria26@gmail.com*

ORCID: https://orcid.org/0009-0004-4828-2083

under-link, and probabilistic scoring over-links unless conflict evidence carries its own weight.

PSS is a scoring engine designed from healthcare payer MDM practitioner experience. PSS keeps the Fellegi-Sunter probabilistic framework [5] as its mathematical foundation but adds four operational elements the underlying theory does not by itself provide: differential positive weighting across eight member attributes, an explicit negative penalty for SSN conflict distinct from missing-value handling, calibrated thresholds that separate auto-link, manual-review, and auto-reject decisions, and a sensitive-condition override that routes high-risk PHI categories to human review independent of the numeric score.

1.1. Research Questions

This article investigates five questions that follow directly from the operational problem above. RQ1 through RQ3 are evaluated through a fully synthetic payer-style benchmark calibrated to reproduce a field-informed edge-case structure relevant to payer operations. RQ4 and RQ5 are evaluated as governance and validation-design contributions.

RQ1: How can probabilistic scoring improve identity resolution accuracy while preserving the auditability regulators expect from a healthcare payer's MDM function?

RQ2: How does an explicit SSN-conflict penalty affect false-positive consolidation in identifier-reuse and household-sharing scenarios compared to scoring that treats a conflict as equivalent to a missing value?

RQ3: Can ROC-calibrated decision thresholds sustain high auto-link precision while meaningfully reducing manual steward review workload?

RQ4: How does sensitive-condition override logic reduce privacy exposure risk for behavioral health, HIV, and substance-use record categories that a purely score-based routing scheme would otherwise treat like any other match?

RQ5: What statistical validation approach can evaluate classification performance, score drift, and subgroup fairness together, rather than treating each as a separate concern?

1.2. Novel Contribution of This Article

This work contributes nine elements to healthcare payer identity resolution:

Scoring: (1) a formally specified scoring function combining positive and negative components; (2) differential attribute weighting calibrated across eight member identifiers; (3) explicit negative scoring for SSN conflicts, distinct from missing-value treatment.

Privacy and governance: (4) calibrated decision thresholds defining auto-link, manual-review, and auto-reject zones; (5) sensitive-condition override logic bypassing score-based routing for high-risk PHI categories; (6) an immutable audit trail reconstructing model version, score, attributes, and steward rationale for any historical decision.

Lifecycle and validation: (7) a monthly governance cadence tracking score drift and steward reversal rates; (8) a validation framework spanning classification performance, threshold sensitivity, and subgroup fairness; (9) a practitioner-informed synthetic validation backbone enabling inspection of payer-realistic edge cases without exposing production PHI.

2. Literature Review

The theoretical basis for probabilistic record linkage was established by Fellegi and Sunter [5], who framed matching as a decision problem under uncertainty: each candidate record pair receives a composite score built from per-attribute match probabilities, and the pair is classified as a match, non-match, or possible match requiring clerical review depending on where that score falls relative to two calibrated thresholds. The three-zone decision structure, link, review, and non-link, remains the conceptual skeleton underneath most production record linkage systems more than five decades later, including PSS [6].

Fellegi and Sunter's original framework assumed attribute comparison functions that were largely binary, agree or disagree, an assumption that performs poorly against free-text name fields where transcription errors, nicknames, and spelling variation are common. Jaro [7] introduced a string comparison metric tolerant of transpositions and partial character agreement, and Winkler [8] extended it with a prefix-weighting adjustment and decision-rule refinements suited to census-style record linkage. The resulting Jaro-Winkler similarity measure is still the default fuzzy-matching function inside most commercial and open-source entity resolution tooling, including the name-matching component of PSS.

Entity resolution research since Fellegi and Sunter has broadly split into deterministic approaches, exact or rule-based attribute agreement, and probabilistic approaches that assign continuous match likelihoods. A comprehensive survey of end-to-end entity resolution pipelines documents this split and notes that

neither family dominates across all data conditions; deterministic pipelines are easier to audit but brittle against data entry variation, while probabilistic pipelines generalize better but require careful threshold calibration to avoid over-linking [9]. Work extending privacy-preserving record linkage frames the central design tension similarly, as a tradeoff between linkage recall and disclosure risk rather than a solvable optimization [10].

Healthcare-specific applications of the Fellegi-Sunter framework have grown substantially. A data-adaptive extension incorporating missing-data handling and automated field selection improved linkage performance on clinical record pairs where attribute completeness varies by source system [11]. Separate work applying machine learning classifiers to master patient index linkage found that supervised models trained on labeled historical linking decisions outperformed static weight configurations on held-out validation pairs [12], which supports the labeled-pairs evaluation design PSS adopts. Probabilistic linkage without direct patient identifiers has also been validated at national registry scale in the United Kingdom [13].

A second literature thread relevant to PSS concerns fairness in automated linkage decisions. Work on fairness- and cost-constrained privacy-aware record linkage shows that linkage error rates are not uniformly distributed across subpopulations, and that subgroups with sparser or less standardized identifying attributes experience systematically lower match confidence [14]. A framework for fairness-constrained entity resolution proposes explicit fairness objectives alongside accuracy objectives during model training [15], and a related evaluation of support-vector approaches to record linkage reports measurable accuracy-fairness tradeoffs depending on how conflict evidence is weighted [16]. These findings motivate the subgroup fairness testing PSS's validation plan includes.

Identity matching standards work has documented substantial variability in which attributes health systems and payers use for matching and how those attributes are standardized [17]. The HIPAA Privacy Rule [2] and Security Rule [18] establish the regulatory floor for handling PHI exposure risk, and HHS's own compliance reporting confirms that breach investigations frequently trace back to identity and access-control failures [3]. A federal oversight review of HHS breach-reporting practices found gaps in how identity-related incidents are tracked and communicated [4]. Work cataloguing attack vectors against privacy-preserving record linkage systems demonstrates that scoring logic itself is a legitimate target for privacy risk analysis [19].

None of this literature, however, specifies a payer-operational scoring function that treats a populated identifier conflict as categorically different evidence from a missing identifier, nor does it integrate sensitive-condition override logic directly into the routing decision. Healthcare record-linkage studies improve linkage performance under missingness and supervised training [11], [12], but they do not foreground PHI exposure risk from false-positive member consolidation as the primary threshold objective. Privacy-preserving linkage research protects disclosure during linkage computation [10], [19], but does not specify how a payer should route behavioral health, HIV, or substance-use categories when a high score conflicts with privacy governance. PSS is built to close that specific gap.

Existing entity-resolution platforms and probabilistic linkage frameworks, including tools such as Splink, IBM InfoSphere MDM, Oracle Enterprise Data Quality, and traditional configurable probabilistic linkage systems, provide configurable matching, thresholding, and data-quality controls. PSS is not positioned as a replacement for those platforms. Its contribution is the formal specification of a healthcare-payer-specific decision layer that integrates explicit SSN conflict penalties, sensitive-condition override routing, healthcare payer privacy governance, steward-adjudicated auditability, ROC-calibrated thresholding, and operational deployment constraints for regulated healthcare identity resolution into a single auditable scoring framework.

3. Method

3.1. PSS Architecture and Scoring Model

PSS computes a composite match score S(a,b) for a candidate pair of records a and b as the sum of positive per-attribute match evidence minus negative conflict penalties, as in (1).

$$S(a,b) = \sum_{i=1}^{n} w_i \, s_i(a,b) + \sum_{j=1}^{m} p_j \, c_j(a,b)$$

Here w_i is the positive weight assigned to attribute i, s_i(a,b) is the per-attribute match score for that attribute across records a and b, p_j is a negative penalty weight, and c_j(a,b) is a conflict indicator that activates when two records actively disagree on a

governed attribute rather than simply lacking a value for it. This additive structure keeps the model interpretable enough for a steward or auditor to reconstruct why a given pair scored the way it did, attribute by attribute.

PSS scores eight member attributes drawn from enrollment, claims, pharmacy, and government-program source systems, each weighted according to its discriminative strength and expected data quality behavior in payer operations [20], [21]. Table I summarizes the weighting logic and privacy/quality rationale behind each attribute.

TABLE I. PSS Attribute Weighting Logic and Privacy/Quality Rationale

Attribute	Weighting Logic	Rationale
SSN	High positive; strong conflict penalty	Populated conflict is stronger non-match evidence than missing value
MBI	High positive weight	Govt-issued identifier, strong discriminative value
DOB	High positive; partial match	Stable, though transcription errors occur
Last Name	Secondary; fuzzy match	Name changes, spelling variation common
First Name	Secondary; fuzzy match	Nicknames, spelling variation common
Address	Moderate; conflict penalty	Useful but changes, shared by households
Phone	Lower positive	Shared by families, changes over time
Email	Lower-moderate positive	Self-reported, not always stable

The design decision that most distinguishes PSS is the treatment of SSN. A missing SSN is common among pediatric members and newly enrolled dependents, and PSS treats it as a low-information neutral case rather than as conflict evidence. A populated SSN that disagrees between two records signals a known failure mode-reuse, transcription error, or a household member mistakenly entering a guardian's SSN. PSS assigns this conflict a dedicated negative penalty p_j, larger in magnitude than any other single-attribute disagreement.

MBI receives a high positive weight comparable to SSN. DOB carries a high positive weight and supports partial-value matching. Last and first name receive secondary weights with fuzzy matching via the Jaro-Winkler function [7], [8]. Address carries a moderate weight with its own conflict penalty option. Phone and email receive the lowest weights, since both are frequently shared, mutable, or incomplete.

3.2. Evaluation Design

The quantitative evaluation uses a fully synthetic payer-style benchmark containing 101,000 labeled

candidate pairs, 228,000 membership records, 278,017 claims, 10 companies, 15 payer groups, and 18 sub-groups. The benchmark is calibrated to reproduce field-identified identity-resolution patterns, but it does not contain production member records, incident narratives, steward notes, proprietary matching rules, PHI, or internal governance artifacts. Edge cases including twins, household sharing, newborn SSN reuse, name changes, sex/name transitions, COB dependencies, and sensitive claims were sampled or stress-tested at plausible stress-test rates.

Labeled pairs were partitioned into an 80/20 train-validation split, with five-fold cross-validation on the training partition. In this evaluation, the formally specified additive PSS score is the operational decision score used for routing. Logistic regression was used as a calibration and validation benchmark on the synthetic labeled pairs, consistent with prior findings that supervised approaches trained on labeled historical decisions outperform static weight configurations [12]. The normalized weights disclosed in Appendix A represent the auditable scoring configuration evaluated against that benchmark, not an opaque replacement for the additive PSS rule. Class weighting was applied asymmetrically during calibration checks: false-positive consolidations were penalized more heavily than false-negative fragmentations, reflecting that a false positive carries direct PHI exposure risk. The synthetic legacy baseline represents a conventional payer MDM configuration without an explicit SSN-conflict penalty, without sensitive-condition override routing, and with simpler score-threshold routing.

Decision thresholds were calibrated using ROC analysis, targeting a minimum of 99.9% auto-link precision. Three routing zones were defined relative to a calibrated link threshold, T_link, mirroring the Fellegi-Sunter clerical-review zone [5]. The numeric threshold is necessary but not sufficient for auto-linking: hard-stop routing flags, including sensitive-condition indicators, COB or multiple-coverage dependencies, household-sharing patterns, and governed identifier-conflict cases, can still route a high-scoring pair to manual review. Appendix B provides worked synthetic examples showing how these flags affect routing despite scores above T_link. Table II summarizes the evaluation method applied to each article component.

TABLE II. Evaluation Method by Article Component

Component	Method	Evidence
Scoring model	Train-validation split, 5-fold CV	Precision, recall, F1, ROC-AUC, PR-AUC
SSN conflict penalty	With/without penalty comparison	FP reduction, review-routing impact
Threshold routing	ROC-based calibration	Auto-link precision, recall, FP risk
Sensitive override	Override volume/outcomes	PHI categories receive human review
Audit trail	Sample decision reconstruction	Regulatory defensibility
Lifecycle governance	Monthly drift monitoring	Drift detection, retraining triggers

Independent of score-based routing, any candidate pair where either record carries a behavioral health, HIV, or substance-use claim indicator is routed to manual steward review regardless of computed score, through minimum-necessary access controls. This override was evaluated separately from the scoring model, by measuring override volume against total review volume and confirming steward outcomes on overridden pairs.

3.3. Statistical Validation Plan

Classification performance was assessed using precision, recall, specificity, F1, ROC-AUC, and PR-AUC, with confusion matrices at each candidate threshold. Auto-link precision was the primary metric throughout. Threshold sensitivity was tested at three candidate T_link values (35, 40, 45).

The paired synthetic comparison between the baseline configuration and PSS was evaluated using McNemar's test for paired categorical outcomes, chi-square tests for independence, confidence intervals around each headline metric, and paired t-tests or Wilcoxon signed-rank tests for continuous outcomes.

Ongoing drift was monitored using control charts tracking monthly score distributions and steward reversal rates. Subgroup fairness analysis was conducted across members without SSNs, non-Anglophone names [7], [8], pediatric records, Medicaid enrollees, members with frequent address changes, and members sharing household attributes [14]-[16].

3.4. Operational Motivation vs Experimental Evaluation

The paper intentionally separates field motivation from experimental validation. Payer MDM practice informs the failure modes, attribute choices, privacy-governance constraints, and review-routing design. The operational rationale is based on professional MDM experience and is used to define design requirements, not to assert measured production prevalence. Quantitative performance claims are reported only from the fully synthetic benchmark.

This framing demonstrates field relevance: the model is built around identity-resolution defects, steward workflows, and privacy controls that healthcare payer MDM teams encounter in routine operations. The experimental layer evaluates the proposed scoring framework using a fully synthetic benchmark where experiments can be reproduced without exposing regulated healthcare data. Because the benchmark intentionally includes payer-relevant edge cases, it should be interpreted as a construct-validity and stress-test evaluation rather than proof of production prevalence or external generalizability.

The objective of this operational framing is not to establish statistical significance or report production performance. Its purpose is to explain why the proposed decision framework is operationally relevant under enterprise conditions. Statistical significance, comparative evaluation, and reproducibility are addressed separately using the fully synthetic benchmark described in this paper.

4. Results and Discussion

The fully synthetic benchmark provides the quantitative evidence layer for this article. Detailed statistical outputs come from 101,000 scored candidate pairs (90,000 true matches, 11,000 true non-matches), December 1-20, 2025. PSS accuracy is 0.9970 (Wilson 95% CI [0.9967, 0.9973]) versus 0.9406 for the legacy baseline. The paired comparison is significant: 5,912 pairs correct only under PSS, 212 only under legacy, continuity-corrected McNemar chi-square = 5303.4946, $p < 0.001$, Cramer's V = 0.2292. These results should be read as reproducible synthetic validation, not as audited production KPI reporting.

The synthetic benchmark is a confidentiality-safe reproducibility layer, not a substitute for prospective production validation. Auto-link precision in the benchmark is calibrated to approximately 99.9% at the selected T_link, so most surviving risk sits in the manual-review and override zones by design.

4.1. Synthetic Validation and Field Applicability

The strongest publishable evidence package in this version is a repeatable fully synthetic validation package tied to auditable decision records. Table 3 reports the synthetic benchmark threshold-sensitivity output. The selected operating threshold is T_link = 40, the lowest evaluated threshold that preserves approximately 99.9% auto-link precision while keeping manual-review escalation inside the 1%-3% operating band. Recall in Table III is auto-link recall at the selected threshold; Table IV reports overall classification and routing recall across the full scored benchmark. Appendix B provides worked synthetic examples showing how hard-stop review flags can route high-scoring pairs to manual review despite exceeding T_link.

TABLE III. Routing Threshold Sensitivity Comparison

T_link	Precision	Recall	Review Rate	FP Links	Note
35	.9990	0.9792	0.0100	95	Lower review burden, higher FP exposure
40	0.9990	0.9768	0.0200	88	Selected threshold
45	0.9991	0.9705	0.0300	79	Higher review burden, lower recall

Table IV reports the full validation performance metrics with Wilson 95% confidence intervals.

The relatively lower ROC-AUC should be interpreted alongside the high PR-AUC because the benchmark is intentionally imbalanced and optimized around very high auto-link precision. In this setting, PR-AUC and threshold-specific precision are more informative than global rank ordering across all possible thresholds.

TABLE IV. Validation Performance Metrics

Metric	Value	95% CI
Labeled pairs	101,000	-
True matches	90,000	-
True non-matches	11,000	-
PSS accuracy	0.9970	[0.9967, 0.9973]
Legacy accuracy	0.9406	[0.9391, 0.9420]
PSS precision	0.9990	[0.9988, 0.9992]
PSS recall	0.9976	[0.9973, 0.9979]
PSS specificity	0.9920	[0.9902, 0.9935]
PSS FP rate	0.0080	[0.0065, 0.0098]
PSS F1	0.9983	[0.9981, 0.9985]

PSS auto-link precision	0.9990	[0.9988, 0.9992]
PSS ROC-AUC	0.8621	-
PSS PR-AUC	0.9983	-

TABLE V. Confusion Matrix at Selected Operating Point (T_link = 40)

Actual class	Linked by PSS	Not linked / review-routed by PSS
True match	89,784 TP	216 FN
True non-match	88 FP	10,912 TN

These counts reconcile the headline validation metrics in Table IV: recall = 89,784 / 90,000 = 0.9976, specificity = 10,912 / 11,000 = 0.9920, and precision = 89,784 / (89,784 + 88) = 0.9990.

Table VI reports subgroup fairness results. No material precision gap was observed for the auto-linked subgroup decisions; the household-sharing subgroup shows recall of 0.0000 because those pairs are intentionally routed to manual review as a privacy-recall tradeoff, reflecting an oversampled stress-test cell rather than production prevalence. Perfect subgroup point estimates in Table 6 reflect controlled synthetic stress-test construction and should not be interpreted as expected production subgroup performance. Wilson lower confidence bounds become materially wider for small cells, especially the sex/name-change subgroup (N = 79) and the multiple-coverages edge case (N = 2), so those rows are reported as stress-test signals rather than inferential evidence of production fairness.

TABLE VI. Subgroup Fairness Results

Subgroup	N	Precision	Recall	Note
Non-Anglophone	3,789	1.0000	1.0000	No material gap
No SSN	2,437	1.0000	1.0000	No material gap
Pediatric	1,913	1.0000	1.0000	No material gap
Household sharing	1,004	1.0000	0.0000	Stress-test cell, conservative routing
Frequent mover	727	1.0000	1.0000	No material gap
Medicaid	659	1.0000	1.0000	No material gap

Sex/name change	79	1.0000	1.0000	No material gap
Multiple coverages	2	1.0000	1.0000	Illustrative edge-case cell; not inferential

4.2. McNemar Paired Comparison

Table VII reports the paired-decision comparison between PSS and the legacy baseline using a continuity-corrected McNemar test.

TABLE VII. McNemar Paired Comparison, PSS vs. Legacy

PSS-Only Correct	Legacy-Only Correct	Chi-Sq / p / V
5,912	212	5303.49 / <0.001 / 0.2292

Override volume, while a small fraction of total synthetic review volume, consistently produced human review for every behavioral health, HIV, or substance-use flagged pair regardless of computed score. Every scoring decision retains its model version, computed score, per-attribute contribution, and steward rationale, reconstructable on demand for regulatory or forensic review [17], [18], [22]-[24]. Monthly governance reviews track score distribution drift, steward reversal rates, and attribute completeness, triggering a retraining or reweighting cycle when indicators move outside established tolerance.

4.3. Limitations

PSS's performance depends directly on the quality of the labeled decisions encoded in the synthetic benchmark, and static attribute weights can degrade over time as upstream source-system data quality shifts. Beyond weight drift, the pairwise scoring approach does not directly model household or family relational context, and Jaro-Winkler similarity may underperform for non-Anglophone names [7], [8]-a limitation that compounds for members without an SSN on file, who receive systematically lower-confidence scores by design. This motivation reflects anecdotal and experiential payer MDM practice rather than measured production outcomes; it should be read as design context, not as empirical evidence of production prevalence. External generalizability should be tested through prospective, governed deployment before production performance claims are made.

4.4. Future Research

Future work could extend PSS toward online learning incorporating steward feedback continuously, graph-based identity resolution modeling household and family relationships, fairness-aware threshold calibration [14], [15], explainable AI dashboards, and integration with data observability platforms. Source-trust scoring by field and feed, temporal identity logic tied to coverage dates, and household-aware graph context would further harden production accuracy.

5. Conclusion

Healthcare payer MDM teams have generally treated identity resolution accuracy and privacy risk as separate problems. The synthetic results reported here suggest that this integrated scoring and governance design can improve both identity-resolution accuracy and privacy-risk control within the evaluated benchmark. The audit trail and monthly governance cadence built into PSS matter as much as the scoring function itself. What remains is extending the same governance discipline into household relationships, non-Anglophone name matching, and subgroups whose identifying attributes are sparser by circumstance.

5.1. Appendix

Appendix A. Reproducible Synthetic Scoring Weights. Implementation-specific weights may remain operational controls in real deployments, but the fully synthetic companion package discloses normalized weights (100-point scale) sufficient to reproduce the reported scoring logic:

TABLE VIII. Synthetic Scoring Weights

Attribute	Weight	Penalty	Note
SSN	24	-18	Strongest conflict penalty
MBI	22	-12	High-confidence identifier
DOB	16	-8	Strong conflict evidence
Last name	9	-3	Fuzzy + temporal exception
First name	8	-3	Fuzzy + nickname handling
Address	8	-4	Moves handled conservatively
Email	6	-2	Self-reported, mutable
Phone	5	-2	Shared within households

Appendix B. Candidate-Pair Scoring Examples. Selected fully synthetic rows showing routing behavior for twins, household sharing, name change, sex/name transition, newborn SSN reuse, and multiple-coverage dependency cases. Table 9 routes reflect both the calibrated score threshold and hard-stop review rules; therefore, a high-scoring pair can still be routed to manual review when a governed override or dependency flag is present. PAIR000006 (last-name change after marriage) shows a PSS score of 68.8 and auto-link decision despite name-history variation, given high DOB/SSN/MBI/address agreement and no hard-stop review flag.

TABLE 9. Candidate-Pair Scoring Examples

Scenario	True Match	PSS Score	Route
Newborn using mother's SSN	0	56.8	Manual review
Last name change (marriage)	1	68.8	Auto-link

Multiple people, same account	0	31.2	Manual review
Twins, different gender	0	50.0	Manual review
First name change	1	72.6	Auto-link
Multiple coverages	1	76.0	Manual review
Twins, same gender	0	50.0	Manual review
Different names, same member	1	67.5	Manual review

Appendix C. Audit Reconstructability. The audit log records candidate pair, source systems, model version, total score, threshold route, sensitive-override flag, steward role, rationale, timestamp, and correction/appeal status, protected in production with append-only storage and role-based access.

Appendix D. Real-Time Payer Accuracy Hardening. The next production iteration should add source-trust scoring by field and feed, temporal constraints tied to coverage dates, household graph context, COB-aware dependency rules, and a closed-loop steward feedback process, directly addressing reviewer concerns about Jaro-Winkler limitations, family context, and payer-specific generalizability.

5.2. Acknowledgment

The author declares no acknowledgments beyond those listed as authors.

References and Footnotes

Author contributions

V. Nafria: Conceptualization, Methodology, Scoring architecture design, Formal analysis, Writing - original draft, Writing - review and editing.

Conflicts of interest

The author declares no conflicts of interest.

Ethics and Data Availability

Ethics approval: Not applicable for the submitted quantitative evaluation because it uses only fully synthetic data and does not disclose PHI, member-level records, steward notes, proprietary matching rules, or internal governance artifacts. The study is presented as practitioner-informed synthetic-data research rather than human-subjects production-data analysis. Data availability: The synthetic benchmark, scoring logic, and aggregate validation tables constitute the reproducible data layer for this article; production enterprise data are not available because they may contain regulated healthcare information and confidential operational controls.

References

[1] A. Haug, A. M. Staskiewicz, and L. Hvam, "Strategies for master data management: A case study of an international hearing healthcare company," Information Systems Frontiers, vol. 25, no. 5, pp. 1903-1923, 2023.

[2] U.S. Department of Health and Human Services, "The HIPAA Privacy Rule," HHS, Washington, DC, USA, 2024.

[3] U.S. Department of Health and Human Services, Office for Civil Rights, "Annual report to Congress on HIPAA Privacy, Security, and Breach Notification Rule compliance for calendar year 2022," HHS, Washington, DC, USA, submitted Feb. 14, 2024.

[4] U.S. Government Accountability Office, "Electronic health information: HHS needs to improve communications for breach reporting," GAO, Washington, DC, USA, Rep. GAO-22-105425, 2022.

[5] I. P. Fellegi and A. B. Sunter, "A theory for record linkage," Journal of the American

Statistical Association, vol. 64, no. 328, pp. 1183-1210, 1969.

[6] B. Williams, "Record linkage in statistical sampling: Past, present, and future," in Recent Advances on Sampling Methods and Educational Statistics, Cham, Switzerland: Springer, 2022, ch. 9.

[7] M. A. Jaro, "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida," Journal of the American Statistical Association, vol. 84, no. 406, pp. 414-420, 1989.

[8] W. E. Winkler, "String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage," in Proc. Section on Survey Research Methods, American Statistical Association, 1990, pp. 354-359.

[9] V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, and K. Stefanidis, "An overview of end-to-end entity resolution for big data," ACM Computing Surveys, vol. 53, no. 6, Art. no. 127, 2020.

[10] A. Gkoulalas-Divanis, D. Vatsalan, D. Karapiperis, and M. Kantarcioglu, "Modern privacy-preserving record linkage techniques: An overview," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 4966-4987, 2021.

[11] X. Li, H. Xu, and S. Grannis, "The data-adaptive Fellegi-Sunter model for probabilistic record linkage: Algorithm development and validation for incorporating missing data and field selection," Journal of Medical Internet Research, vol. 24, no. 9, e33775, 2022.

[12] W. Nelson, N. Khanna, M. Ibrahim, J. Fyfe, M. Geiger, K. Edwards, and J. Petch, "Optimizing patient record linkage in a master patient index using machine learning: Algorithm development and validation," JMIR Formative Research, vol. 7, e44331, 2023.

[13] H. Blake, L. Sharples, K. Harron, J. van der Meulen, and K. Walker, "Linkage of national clinical datasets without patient identifiers using probabilistic methods," International Journal of Population Data Science, vol. 7, no. 3, 2022.

[14] N. Wu, D. Vatsalan, S. Verma, and M. A. Kaafar, "Fairness and cost constrained privacy-aware record linkage," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2644-2656, 2022.

[15] V. Efthymiou, K. Stefanidis, E. Pitoura, and V. Christophides, "FairER: Entity resolution with fairness constraints," in Proc. 30th ACM Int. Conf. Information and Knowledge Management (CIKM 2021), 2021.

[16] C. Makri, A. Karakasidis, and E. Pitoura, "Towards a more accurate and fair SVM-based record linkage," in Proc. 2022 IEEE Int. Conf. Big Data, 2022, pp. 4691-4699.

[17] Y. Deng, L. P. Gleason, A. Culbertson et al., "Evolving availability and standardization of patient attributes for matching," Health Affairs Scholar, vol. 1, no. 4, qxad047, 2023.

[18] U.S. Department of Health and Human Services, "The HIPAA Security Rule," HHS, Washington, DC, USA, 2024.

[19] A. Vidanage, T. Ranbaduge, P. Christen, and R. Schnell, "A taxonomy of attacks on privacy-preserving record linkage," Journal of Privacy and Confidentiality, vol. 12, no. 1, 2022.

[20] IEEE Standards Association, "IEEE recommended practice for the quality management of datasets for medical artificial intelligence," IEEE Std 2801-2022, 2022.

[21] R. Miller, H. Whelan, M. Chrubasik, D. Whittaker, P. Duncan, and J. Gregorio, "A framework for current and new data quality dimensions: An overview," Data, vol. 9, no. 12, p. 151, 2024.

[22] A. A. Mamun, S. Azam, and C. Gritti, "Blockchain-based electronic health records management: A comprehensive review and future research direction," IEEE Access, vol. 10, pp. 5768-5789, 2022.

[23] A. Ahmad, M. Saad, M. Al Ghamdi, D. Nyang, and D. Mohaisen, "BlockTrail: A service for secure and transparent blockchain-driven audit trails," IEEE Systems Journal, vol. 16, no. 1, pp. 1367-1378, 2022.

[24] A. Haddad, M. H. Habaebi, M. R. Islam, N. F. Hasbullah, and S. A. Zabidi, "Systematic review on AI-blockchain based e-healthcare records management systems," IEEE Access, vol. 10, 2022.