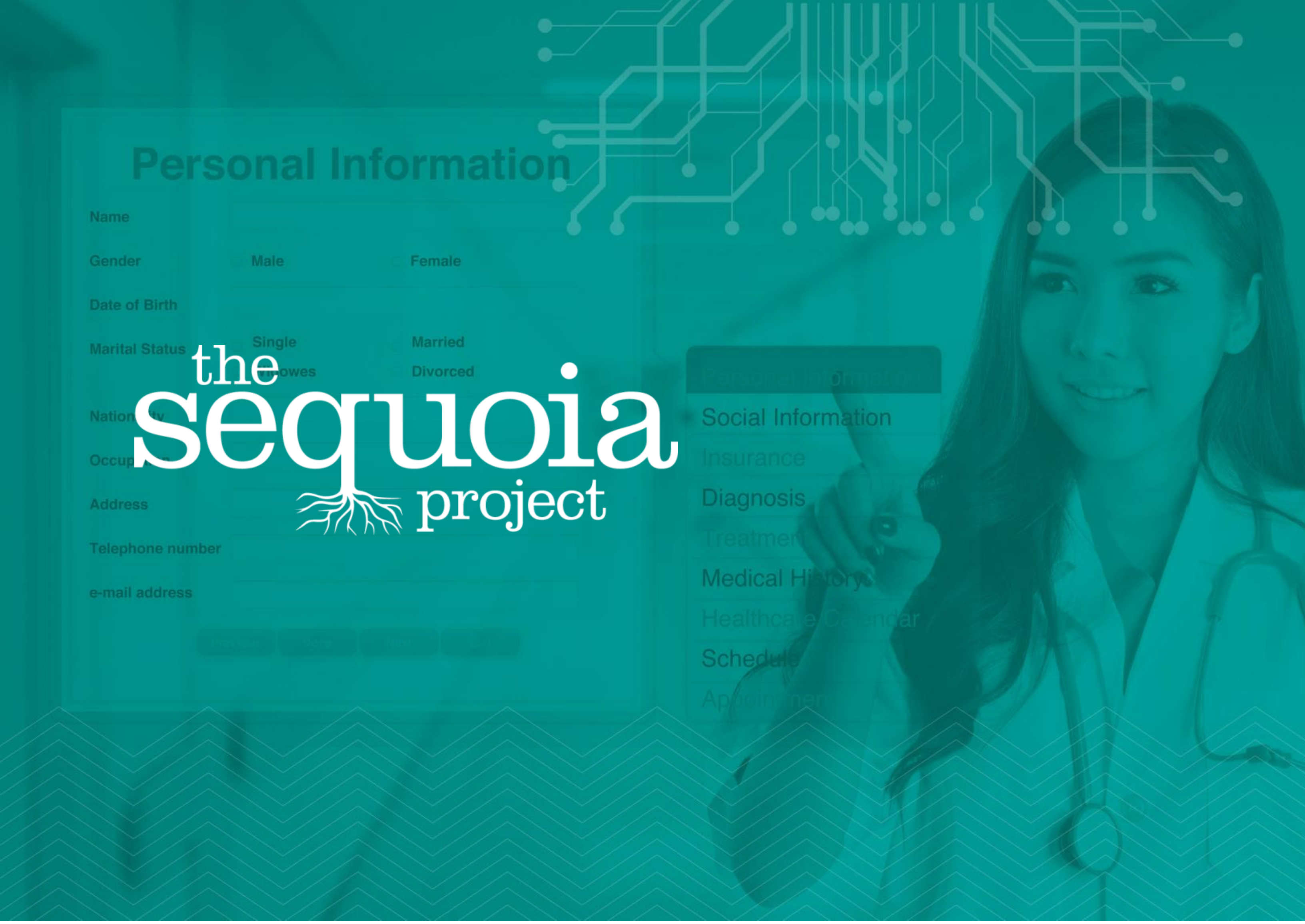


A Framework for Cross-Organizational Patient Identity Management

TABLE OF CONTENTS

An overview of The Sequoia Project, the Care Connectivity Consortium, and a list of the significant contributing authors.

This chapter provides an introduction to the topic of patient identity management, and the development of a framework. This chapter also addresses patient privacy concerns and the patient identity blind spot phenomenon.

This chapter discusses the patient matching strategies taken by Intermountain Health Care and its exchange partners to drive matching success rates from 10% to more than 95%. In addition to providing steps to replicate their results, it also provides lessons learned, future state projections, and key questions you should be asking your organization.

This chapter shares an evolving cross-organizational patient matching maturity model for public comment and future adoption. The maturity model details the characteristics of the initial maturity levels, and will expand to include additional levels as each is accepted.

This chapter details specific cross-organizational patient matching minimal acceptable principles for public comment and future adoption. This technical section includes concrete suggestions for consistent and accurate patient matching nationwide.

Table of Figures

Table of Tables

May 30, 2018

Dear Reader,

Since health data exchange began more than a decade ago, we've seen incremental improvement in patient identity management between data exchange partners. Without a national patient ID system, health IT leaders developed best-in-class strategies to ensure disparate patient data is matched to the right patient every time. Most of these solutions were focused on matching within a health care provider organization or system. Little focus was given, however, to matching between different organizations that have different patient identity management practices and varying levels of maturity. A few years ago, The Sequoia Project and the Care Connectivity Consortium turned their attention to best practice patient matching principles across organizational boundaries.

In a joint case study with Intermountain Health Care, we found cross-organizational matching deficiencies and opportunities for improvement among even our most advanced institutions. In 2016, we published a proposed framework for patient identity management that included actionable best practices and a maturity model roadmap for future growth and improvement in nationwide patient matching strategies. We asked the industry for input on this proposal and we received overwhelmingly positive feedback and many helpful comments which were reviewed by a balanced stakeholder group of the public and private sectors.

The following framework incorporates this feedback, along with new insights and recommendations to improve patient matching nationally across different organizations, disparate technologies, and networks. We believe that adoption of these minimum acceptable matching principles and the maturity model will create a floor upon which the entire industry can build and improve, with continued refinements to incrementally increase match rates toward 100%. For that reason, we do not consider these recommendations "final." As the nation adopts new best practices and improves match rates, we will continue to raise the floor. Likewise, as interoperability evolves, we will evolve the recommendations to adapt to new technologies and best practices.

Thank you to our volunteers from across the industry and government who submitted feedback or participated in the workgroup. Your thoughtful analysis and feedback were key to bettering this proposed framework, and ultimately, bettering our national patient matching capabilities.

Yours in good health,

Mariann Yeager　　Michael Matthews

Mariann Yeager & Michael Matthews
CEO, The Sequoia Project　　Chief Strategy Officer, Cedarbridge Group

Contributors to The Framework

The Sequoia Project

The Sequoia Project is a non-profit 501c3 chartered to advance implementation of secure, interoperable nationwide health data sharing. The Sequoia Project supports health IT interoperability initiatives, most notably: eHealth Exchange, a rapidly growing community of exchange partners who share information under a common trust framework and a common set of rules; and Carequality, a public-private collaborative effort to build consensus among existing data sharing networks regarding technical specifications and best practices.

The Care Connectivity Consortium

The Sequoia Project has teamed with the Care Connectivity Consortium (CCC) to address critical challenges to effective health information exchange. As pioneers, they share the goal of making data sharing a reality on a national scale to support health care providers and patients with easy, fast access to information at the point of care. The CCC, representing Geisinger Health System, Group Health Cooperative, Intermountain Health Care, Kaiser Permanente, Mayo Clinic, and OCHIN, works to develop and incubate new capabilities for adoption by its initiatives and the broader health IT community. The CCC are piloting a set of operational shared services focused on patient identity management, record location, and consent.

Contributing Authors

- Lead Author: Eric Heflin, Chief Technology Officer/Chief Information Officer, The Sequoia Project
- Shan He, Medical Informatics, Intermountain Health Care/Care Connectivity Consortium
- Kevin Isbell, Executive Director of Health Information Exchange, Kaiser Permanente/Care Connectivity Consortium
- Andy Kling, Director of IT, Geisinger Health System/Care Connectivity Consortium
- Katherine Lusk, Chief Health Information Management and Exchange Officer, Children’s Health (Dallas, Texas)

- Odysseas Pentakalos, Ph.D., CTO, SYSNET International, Inc.
- Chris Ross, Chief Information Officer, Mayo Clinic/Care Connectivity Consortium
- Seth Selkow, Director of CCC Program and HIE Engagement, Kaiser Permanente/Care Connectivity Consortium
- Sid Thornton, Medical Informatics, Intermountain Health Care/Care Connectivity Consortium
- Jim Younkin, Senior Director, Audacious Inquiry
- Kelly Carulli, Manager, Audacious Inquiry
- Dawn Van Dyke, The Sequoia Project (Editor)

Patient Identity Management Workgroup

We thank the following individuals for their participation on the Sequoia Project's Patient Identity Management Workgroup. Comprised of industry experts, the workgroup members were charged with dispositioning comments on the Patient Identity Management White Paper. The Workgroup virtually met two times per week between September and December 2017. Special recognition goes to the co-chairs that facilitated the working sessions: Ryan Bramble, Zach Gillen, and Rebecca Madison. The time and energy spent by workgroup members to further advance patient matching as it applies to health care is greatly appreciated.

*Denotes Co-chairs

In alphabetical order:

Jamie Bennett
Healthcare Systems Specialist
JP Systems Inc.
Veterans' Health Administration

*Ryan Bramble**
Senior Director of Technology
Chesapeake Regional Information System for our Patients (CRISP)

Karon Casey
IT Manager
Coastal Connect Health Information Exchange

Adam Culbertson
Innovator in Residence
HIMSS

John T. Donnelly
President
IntePro Solutions Inc.

*Zachary Gillen**
Senior Director, Care Delivery Technology Services
Kaiser Permanente IT

Eric Heflin
Chief Information Officer (CIO)/Chief Technology Officer (CTO)
The Sequoia Project

Al Jackson
Vice President of Information Management & System Performance
Surescripts LLC

Lesley Kadlec, MA, RHIA, CHDA
Director, Practice Excellence
American Health Information Management Association (AHIMA)

Katherine Lusk, MHSM, RHIA, FAHIMA
Chief Health Information Management and Exchange Officer
Children's Health (Dallas, Texas)

*Rebecca Madison**
Executive Director
Alaska eHealth Network

Shelley Mannino-Marosi
Senior Director, Program Management
Michigan Health Information Network Shared Services

Greg Mears, MD
Medical Director
ZOLL

Wendi Melgoza, RHIA, CPHI
HIM Data Quality Manager
Sutter Health Shared Services

Ben Moscovitch
Manager, Health Information Technology
The Pew Charitable Trusts

Marty Prahl
Health IT Consultant
Social Security Administration

Catherine Procknow
Software Developer
Epic

Carmen Smiley
IT Specialist (Health System Analysis)
Office of Standards & Technology
HHS Office of the National Coordinator for Health Information Technology

Chapter 1: A Framework for Patient Identity Management

Introduction

The health care industry is making significant progress towards technical interoperability but continues to fall short of the promise of ubiquitous interoperable health data sharing. Until we can consistently send and receive accurate and useful patient data nationwide, we miss the opportunity to fully realize the documented benefits of seamless, interoperable health data sharing, including improvements in clinical decision making and patient safety, business process improvement, and support for value-based payment. Among the remaining challenges to successful nationwide exchange are patient matching and identity management.

Inadequate patient matching is impeding our ability to provide physicians with accurate, timely, and useful information

The inability to consistently and accurately match patient data creates a number of problems for physicians and other health care providers. A mismatched record is one where a patient's record is linked with a record belonging to a different patient, or a separate record is added for an already established patient, thus creating two separate medical records for the same patient. With respect to patient safety and satisfaction, providers may have an incomplete view of a patient's medical history, care may not be well coordinated with other providers treating the patient, patient records may be overlaid, unnecessary testing or improper treatment may be ordered, and patient confidence may be eroded. In addition, patient privacy preferences may not be honored across organizations without accurate matching.

With insufficient matching practices, providers may experience a number of clinical workflow inefficiencies that are costly. Those include prolonged troubleshooting to find the correct patient record, a reversion to manual telephone and fax information exchange workflows, ordering duplicate tests, and failure to detect and honor patient privacy preferences. Intermountain estimates that the operational costs of fixing a duplicate record is $60 per manual review and intervention.[1] Further, adverse impact to patient safety remains the primary concern in accurately identifying patients. A College of Healthcare Information Management Executives (CHIME) survey indicated that one in five

[1] https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf

CIOs had at least one patient suffer an adverse event in the past year due to mismatched records.[2]

The following chapters will provide an educational overview of patient matching concepts, a case study examining one organization's efforts to improve patient matching internally and across other data sharing organizations, and finally minimal acceptable principles for cross-organizational patient matching. This paper is intended to provide industry guidance on how to improve patient matching across organizations. Organizations should endeavor to establish a plan of achieving higher and higher levels of patient matching scoring to decrease adverse patient outcomes and operational costs, and ultimately streamline health care delivery.

Patient Matching Overview and Concepts

The purpose of this section is to provide background information on patient identification matching and increase the awareness of best practices implementing patient matching methodologies. This applies to participants, vendors supporting the exchange of data, and those that are responsible for setting technical objectives and strategy for their gateways.

Patient Matching includes a broad array of concepts having to do with correctly associating health records with the appropriate patient within a specific health care organization or across multiple systems exchanging health data. The concepts within this paper apply to multiple types of health data exchange within and across organizational boundaries including:

- Query Based Exchange: Requesting a patient's record or a batch of records, also known as "pull" communication.
- Push: Updating a patient's record and sending the update to other systems so that they have the most up-to-date information for patients.
- Publish/Subscribe: Pushing a patient's records to a receiver based upon previously established criteria. Organizations that subscribe will receive updated or new information.

We present a concrete method of evaluating various matching approaches, called sensitivity/selectivity analysis.

[2] Summary of CHIME survey on patient data matching. May 16, 2012. www.ciochime.org/advocacy/resources/download/Summary_of_CHIME_Survey_on_Patient_Data.pdf

Specifically, it has been discovered that some organizations are using patient matching deployments that:

1. Do not detect shared patients (high false negative match rates);
2. Use insufficient rigor to avoid false positive matches;
3. Use sub-optimal deployment architecture;
4. Are too slow; and
5. Require demographic traits that are not uniformly available.

Implications:

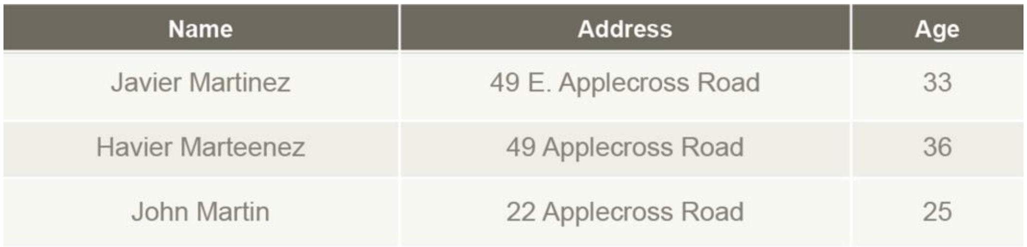
1. Requires human intervention (slow, expensive, not automatable);
2. Could prevent health data exchange from scaling;
3. Cause health data exchange to achieve a lower level of success and perceived value;
4. Reduced quality of patient care; and
5. Decreased patient safety.

Background

Health care data are captured in many different settings such as hospitals, clinics, labs, and physician offices. According to a report by the CDC, patients in the United States made an estimated 1.1 billion visits to physician offices, hospital outpatient departments, and hospital emergency departments just in 2006. This corresponds to a rate of about four visits per patient. In addition to the large volume of visit data generated on an annual basis, the data are also distributed across different health care settings as patients visit hospitals, primary and specialty physicians and move across the country. In order to assumble a longitudinal health record of a patient, all relevant data need to be integrated accurately and efficiently despite the fact that the data are captured using disparate and heterogeneous systems.

The heterogeneity of the systems used to capture patient data at different health care systems cause patient records to have multiple unrelated patient identifiers assigned to them, with the possibility of multiple identifiers assigned to a given patient within a single institution. The lack of precise standards on the format of patient demographic data elements results in incomplete data sharing among health care professionals, patients, and data repositories. In addition to the syntactic heterogeneity of data, unavoidable and all too common data entry errors further exacerbate the inconsistency of those data.

In order to integrate all of a patient's health care data, the various patient identifiers assigned to the patient either at different institutions, or erroneously by a single institution, must be linked together despite the presence of syntactic and semantic differences in the associated demographic data captured for the patient. This problem has been known for more than five decades as the record linkage or the record matching problem. The goal in the record matching problem is to identify records that refer to the same real-world entity, even if key identifying attributes in the records do not match completely. If each record carried a unique, universal, authentic, and error-free identification code, then the record matching problem would be easily solved. But in the absence of such an identifier, the records must be matched using the available demographic attributes in the records. The syntactic and semantic differences between the data as well as the data entry errors introduced during capture result in the need to use identification codes that are neither unique nor error-free. For example, only a sophisticated record matching algorithm could automatically make the decision on whether to link the identifiers assigned to the following two patient records that belong to the same person.



Name	Address	Age
Javier Martinez	49 E. Applecross Road	33
Havier Marteenez	49 Applecross Road	36
John Martin	22 Applecross Road	25

Table 1: Record Matching

Sophisticated record matching algorithms approach this complex problem by decomposing the overall process into three tasks: the data preparation phase, the search phase and the matching phase.

1. The data preparation phase is a pre-processing phase which parses and transforms patient damgraohic data in an effort to remove the syntactic and semantic heterogeneity. In the absence of unique patient identifiers, the matching algorithms must use patient demographic data as the matching variables; in the data preparation phase the individual fields are transformed to conform to the data types of their corresponding domains.

2. In the search phase, the algorithms search through the data to identify candidates for potential matches. The brute-force approach of an exhaustive search across all pairs is very expensive (a search over the Cartesian product of

the two data sets is of quadratic order with respect to the number of records in the two data sets) so, this is not a feasible approach for linking large data sets. For example, attempting to link two data sets with only 10,000 records each using an exhaustive search would require 100 million comparisons. To reduce the total number of comparisons, blocking algorithms are used which partition the full Cartesian product of possible record pairs into smaller subsets.

3. In the matching phase, the record pairs identified during the searching phase are compared to identify matches. Typically, a subset of the patient's demographic attributes, which are referred to as the matching variables, is used to identify matches. The corresponding matching variables for each pair of records are compared with one another forming the comparison vector. The matching decision must determine whether a pair of records should be regarded as linked, not linked or possibly linked, depending upon the various agreements or disagreements of items of identifying information. The specification of a record linking procedure requires both a method for measuring closeness of agreement between records and a method using this measure for deciding when to classify records as matches. For example, if we are linking person records, a possible measurement would be to compare the family names of the two records and assign the value of "1" to those pairs where there is absolute agreement and "0" to those pairs where there is absolute disagreement. Values in between 0 and 1 are used to indicate how close the values in the two different domains are.

Key Patient Matching Concepts

The following table helps in visualizing the measures that are described below. The rows describe the classification of a record pair in reality as being either a match or a non-match, whereas the columns indicate the classification decision of the matching algorithm.

	Algorithm Match	**Algorithm Non-match**
Actual Match	True Positive (TP)	False Negative (FN)
Actual Non-match	False Positive (FP)	True Negative (TN)

Table 2: Algorithm Match vs. Non-Match

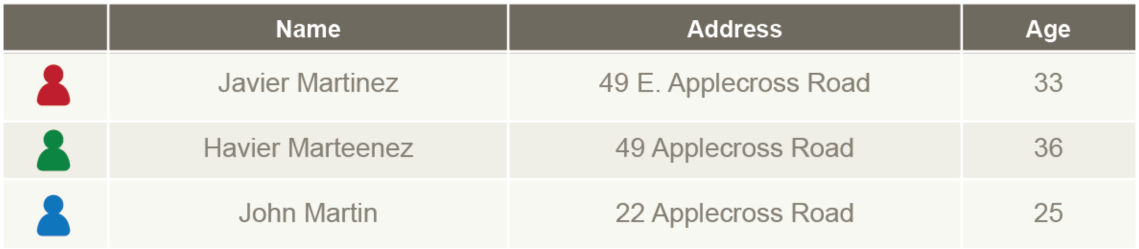
False Positive: Also referred to as a **Type I error**, refers to a classification error by the matching algorithm where a record pair is marked as a match but in reality the two records refer to two distinct patients.

True Positive: Refers to the correct classification by the matching algorithm of two patient records as a match when both records refer to the same person.

False Negative: Also referred to as a **Type II error**, refers to a classification error by the matching algorithm where the two patient records are marked as referring to two distinct patients but in reality the two records refer to the same person.

True Negative: Refers to the correct classification by the matching algorithm of two patient records as a non-match when the two records refer to two different patients.

The following example may make these metrics easier to understand. There are three records that need to be evaluated by the matching algorithm. In reality, the first two records represent two different variations on the demographics of the same person whereas the third record refers to a different individual. The optimal matching algorithm should link the first two records together but should not link the third record with any other records.



	Name	Address	Age
🔴	Javier Martinez	49 E. Applecross Road	33
🟢	Havier Marteenez	49 Applecross Road	36
🔵	John Martin	22 Applecross Road	25

Table 3: Optimal Matching Algorithm

The following figure provides a graphical representation of the various outcomes that a matching algorithm can generate after comparing the three records shown in the previous table and how those decisions would be labeled using the metrics we introduced. The first example is the case where the matching algorithm decided to link the first and second records together and since that is the correct classification in reality, this is considered a true positive. In the second case the first and third records were linked together by the matching algorithm, but in reality, that is an incorrect classification so it is labeled a false positive. In the third example the matching algorithm decided that the first and third records should not be linked together and since in reality these two records belong to two different individuals, this classification is marked as a true negative. The last example shows the case where the matching algorithm decided not to link the first two records together but since in reality both records belong to the same person, this classification is marked a false negative.

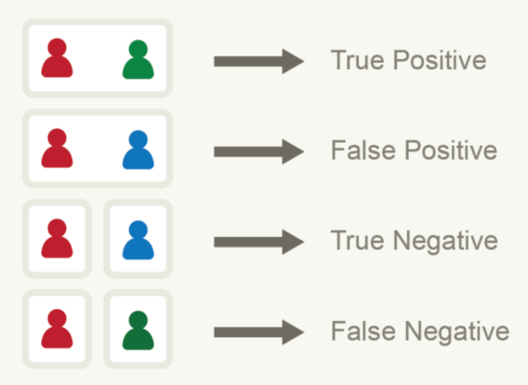


Table 4: True and False Positives and Negatives

The following metrics build on top of the basic metrics described previously. They are commonly used to evaluate the performance of matching algorithms and configuration changes to those algorithms. In the analysis of the data that is used to generate the guidelines in this paper we will use the following metrics to drive the analysis.

Sensitivity: Also referred to as **recall rate**, is a measure of the capability of the matching algorithm to correctly classify two records that refer to the same patient as true matches. It is calculated as the ratio of the number of true positives divided by the sum of true positives and false negatives.

$$Sensitivity \ = \frac{TP}{TP + FN}$$

Specificity: Is a measure of the capability of the matching algorithm to correctly classify two records that refer to two different patients as non-matches. It is calculated as the ratio of the number of true negatives divided by the sum of true negatives and false positives.

$$Specificity \ = \frac{TN}{TN + FP}$$

Precision: is a measure of the fraction of pairs that the matching algorithm classified accurately as matches. It is also called the positive predictor value and can be used along with the sensitivity to jointly evaluate the performance of a matching algorithm.

$$Precision = \frac{TP}{TP + FP}$$

It is important that two metrics are used in the analysis because the metrics evaluate the performance of the matching algorithms from different and conflicting viewpoints resulting in a more balanced assessment as opposed to using just a single metric. For example, if we take the trivial matching algorithm that always links two records together, the sensitivity of this matching algorithm would be 100% since the number of false negatives generated by this algorithm is zero. If we were to look at the specificity of this algorithm though, we would find that it is 0% since it does not identify any true negatives.

It is also important to keep in mind that precision should be preferred as a metric over specificity. Specificity is calculated using the counts of True Negatives and False Positives. In most cases the majority of the record pairs under consideration are true negatives whereas only a relatively small number of record pairs are False Positives. This difference in the scale between the two numbers involved in the calculation causes false negatives to overwhelm the calculation and reduce the value of considering specificity as a metric.

Reducing False Positives

The impact of the patient matching system in making a false positive classification error is that two identity records that belong to two distinct patients are linked together and the clinical data associated with those two records are considered by the health care provider as belonging to the same patient. This can have negative consequences for the patient and these errors should be eliminated or reduced as much as possible. This is a significant error impacting patient safety as it can result in an overlay of two different patients into the same record resulting in an unreliable view of the patient's medical history. This is also why medical identity theft can be life threatening.

Reducing False Negatives

The impact of the patient matching system making a false negative classification error is that two identity records that belong to the same patient are not linked together and as a result the patient's clinical record is not complete. That is because the clinical data associated with each of the two identity records are kept separate and the health care provider is not able to view the patient's complete health record. Although these errors are undesirable they can have less damaging consequences to the quality of care a patient receives compared to false positive errors. This can be a potential patient safety issue since it is possible that existing patient records that may contain valuable information are not found.

Probabilistic/Fuzzy Matching

Probabilistic matching, also referred to as fuzzy matching, uses statistical analysis to determine the overall likelihood or probability that two records match. It is generally accepted that probabilistic matching is a more sophisticated approach to record linking than deterministic matching. Probabilistic matching can take phonetic sounds and nicknames into account as well as "edit distances" when there is variation in fields. For example, probabilistic matching will recognize that records with a first name of Tim and Timothy may belong to the same individual and rank higher probability for similar sounding names like White and Wight. Finally, the probabilistic algorithm will often give a higher probability to 123 Main Street vs. 163 Main Street, recognizing the edit distance between the two as one-character difference.

Rules-based (Deterministic)

Rules-based or deterministic matching uses a one-to-one comparison approach for record linking. Deterministic matching compares exact character-by-character values in fields to determine whether two records should be linked. This strict algorithm does not account for spelling variations or differences in using abbreviations, such as St. versus Street. It should be noted that many consumer-facing applications use deterministic matching over probabilistic, as it is necessary to only return one exact consumer record to that specific record. On the other hand, provider-facing applications can return numerous records, for the provider to then choose the appropriate patient record. It should be noted that even the first academic paper on record linkage recognized that strict deterministic matching is inadequate.

Development of a Framework for Patient Identity Management

To address critical patient matching and identity management issues, The Sequoia Project, in collaboration with the Care Connectivity Consortium (CCC), has developed (and continues to refine) minimal acceptable cross-organizational patient matching rules, suggested matching traits, a framework for methodical improvement, and a maturity model serving as a roadmap for future growth and improvement. The intent of this framework is to facilitate broad collaboration to evaluate, measure, and improve patient matching across organizational boundaries. These work products have been created in response to real-world, known issues from years of experience supporting large-scale, nationwide data sharing initiatives.

As an introduction to this work, we present our observations and a representative case study conducted by the Care Connectivity Consortium and Intermountain Health Care.

Patient Privacy

The CCC and The Sequoia Project believe patient privacy should be at the center of patient identity management strategies. Specifically, we want to help advance the ability of patients to protect the confidentiality and integrity of their data, and to help patients stay aware of and in control of their data. We believe this includes: (1) allowing for anonymous or pseudonymous patient identities; (2) correct identification of patients so that their privacy preferences can be determined and honored; and (3) enabling correct matching of patients to their records (whether anonymous or identifiable). Note that successful patient matching is a precursor for supporting and enforcing a patient's privacy wishes. It may seem counterintuitive, but in order to support a given patient's "opt out" request, that patient first has to be known so his/her preferences can be determined and enforced. We have taken these objectives into account in this framework.

The Patient Identity Blind Spot

Correct, optimal, and safe patient matching has been a high priority for health care organizations since computers were introduced into those enterprises. And even after this multi-decade focus, patient matching is far from perfect. Many believe it will never be successful in the absence of a national identifier. This is described in the RAND Corporation's study, *IDENTITY CRISIS: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System*.

The following pros and cons have been identified for a unique patient identifier approach as a solution to the challenges of accurate patient identity matching:

Pros	Cons	
• Could represent a consistent national approach • Likely improvement in correct patient matching • Could reduce patient safety related errors and improve data sharing • Improves the quality of health care and lowers costs	• Fraud and risk of security breach/privacy compromise potentially increase • Patients who do not have an ID may not be afforded the same access to care • Will require a process for replacing the unique identifier, similar to reporting a lost credit card • Makes it easier for companies to use data for questionable purposes • Expensive and administratively burdensome implementation process	• Risk that the identifier becomes a de facto ID for other purposes (e.g. use of SSN for other purposes) • A unique identifier does not equate to a high assurance identity proofing process or a high assurance authentication process • Political opposition to the unique patient identifier • Some care is anonymous or pseudo-anonymous

Table 5: Pros and Cons of a Unique Patient Identifier Approach

If the United States adopted a unique national patient identifier it could likely improve upon the existing patient matching framework, but it would not lessen the need for the patient matching principles described within this document to fully achieve the ability to accurately identify all records associated with a given patient. Without a unique national identifier, as is the case today, we are left with optimizing a less-than-perfect, complex, mission-critical system – a system upon which patients' lives depend. And because patients receive care across international boundaries, the correct identification of complete and accurate patient data should be both a national and international priority.

As discussed in more detail elsewhere in this paper, patient identification issues are manifested in two broad forms, false negatives and false positives:

- Organizations are unable to detect shared patients within or among their organizations, known as a **false negative**, or;
- Organizations erroneously match the records of different patients, known as a **false positive**.

One example of the impact of a false negative scenario is when a patient presents to a hospital with an infection. The hospital registrar does not recognize the individual as a prior patient and creates a new patient identification number. In this situation, the doctor looking at a new patient record may be unaware of the patient's penicillin allergy, and potentially prescribe this drug, causing an adverse reaction.

On the other hand, a false positive occurs when records belonging to two separate individuals are merged as one. This file will include the medical histories for both patients and therefore include erroneous information for both individuals. It is generally accepted that false positives are more worrisome than false negatives in maintaining the integrity of a patient's medical record.

This Framework addresses these and other related difficult issues.

Your organization, like many others, may have a "blind spot" in terms of patient matching. You likely have acceptable patient matching rates within your enterprise (i.e., your hospital or integrated delivery network). Here, your staff can identify problems, measure, apply fixes, re-measure, and continue to improve until problems are at an acceptable level for patient matching, consent issues, linking, merging, unlinking, and complex unmerging activities. Maintaining organizational awareness of data quality issues is an essential first step in improving data quality and patient matching overall. Organizations are encouraged to analyze false positives, false negatives, duplicates, and overlays in their data systems and to make staff members aware of these metrics for improvement.

But patient matching across organizations is a very different problem. The vast majority of key factors influencing the correctness of patient matching are now out of your direct control, including:

- Default or temporary values;
- Data quality;
- Data completeness;
- Data field consistency;
- Software (vendors, update lifecycle, configuration);
- Vocabulary adoption and versioning;
- Consent, security, sensitive data sharing, and other policies;
- Research Institutional Review Board stipulations;
- Vastly different data characteristics;
- Human and system workflows (latency, variations, definitions, etc.);

Patient matching across organizations is vastly different than matching inside an organization

- Corporate cultures (accepting "friendly" patient nick names vs. meticulous accuracy);
- Data exchange latency;
- Vastly different scope of data (specialty practice vs. large integrated delivery network);
- Organizational size, resource allocation, project timelines, commitment, skill levels;
- Diagnostic capabilities;
- Change management;
- Vendor engagement, version updating strategy, staffing;
- Internal enterprise software architecture (presence/absence of an enterprise-wide active master patient index (MPI), use of multiple MPIs, different tolerances in terms of matching accuracy, different patient matching rules and algorithms, services levels/response times, etc.);
- Legal jurisdictions and requirements (minors, reproductive health, etc.);
- Dependence on the accuracy of patient-supplied data (which studies have shown can be incorrect at an alarming rate) and the associated organizational practices for the rigor of collection of these data;
- Dependence on data provided by family members;
- Differing data items and formats; and/or
- Differing policies for exceptions (e.g. hyphenated names).

Resolution of patient identity issues are more daunting when they cross organizational lines. Such patient identity issues often involve six or more organizations including: the two health information organizations, their two vendors, and often an intermediary such as a health information organization, and their vendor. In such an environment, even mundane items such as scheduling cross-organizational working sessions often introduce days and weeks of delay in resolving each issue due to lack of availability of key personnel. In essence, health data sharing introduces dependencies upon these independent organizations, and intertwines the workflows of the organizations, where no single organization has direct control over the other. This plays heavily into cross-organizational diagnostics, manual fallback procedures when automated patient matching does not work, manual intervention to correct patient records, and manual intervention to gather consent.

> Resolution of patient identity issues are more daunting when they cross organizational lines where they often involve six or more organizations.

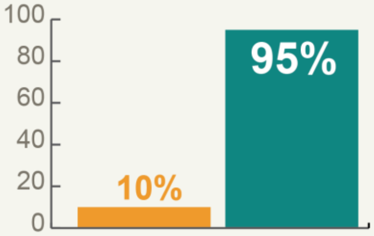
More subtly, it is also a significant issue when determining the "truth." How can we measure the actual, predicted, and targeted patient matching behavior across organizations? Often that entails creating a manually validated subset of mutual patients to become a benchmark to measure patient matching performance. This is a significant effort.

The result: patient matching practices across organizations are inconsistent and often subpar, with match rates as low as 10-30%. In the next chapter, we present a case study of how one organization increased the cross-organizational patient match from only 10% to over 95%, including specific steps, avoidable missteps, and recommendations intended for application to your organization. It should be noted that matching rates and what each organization considers an acceptable match rate will vary significantly by organization. The following chapters present what we assert are the minimal acceptable criteria and practices for improving matching rates.

Chapter 2: Case Study

In this chapter, we present a collaborative study by the Care Connective Consortium (CCC) and The Sequoia Project evaluating traits and processes for successful patient matching across organizations. This study is based on a live production pilot using CCC Services between Intermountain Health Care and local exchange partners. The goal of the shared community case study including a State HIE and an eHealth Exchange Participant is to produce a community wide resource to improve patient matching.

This case study illustrated the ability to increase patient match rates from as low as 10% to more than 95%.



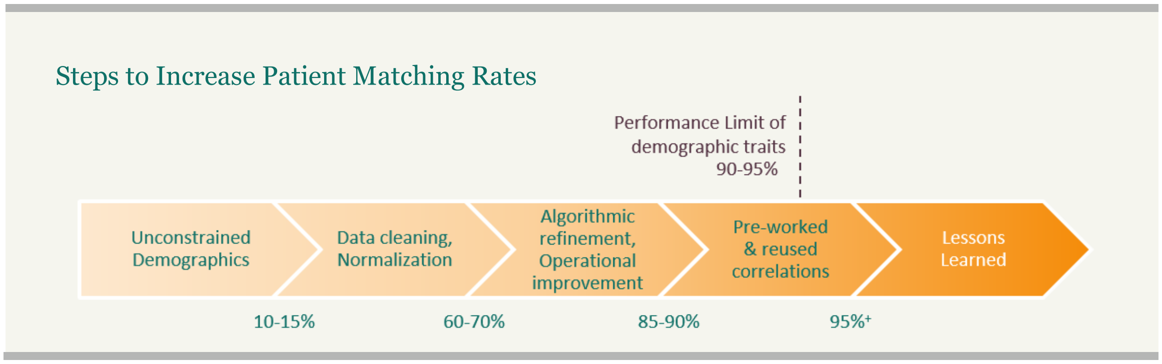
Intermountain Health Care is a not-for-profit health system based in Salt Lake City, Utah, with 22 hospitals, a broad range of clinics and services, about 1,400 employed primary care and secondary care physicians at more than 185 clinics in the Intermountain Medical Group, and health insurance plans from SelectHealth.

Intermountain's willingness to share their incredibly valuable knowledge gain so that the rest of the industry can build upon their work is laudable and provides an example for the industry in terms of "open sourcing" knowledge so we can all benefit from each other's experiences.

The Goal

Intermountain Health Care was seeking to establish exchange of clinical information with two regional organizations as a preliminary step towards broader exchange. Intermountain had invested heavily in health care IT for many years, and frequently shares innovative ways to use IT with the global community. They are very highly regarded in the industry in terms of IT sophistication in the clinical domain. As such, the project was expected to achieve a reasonably high degree of success from the outset. Unfortunately, this proved not to be the case.

Case Study Executive Overview


Steps to Increase Patient Matching Rates
Performance Limit of demographic traits 90-95%
Unconstrained Demographics
Data cleaning, Normalization
Algorithmic refinement, Operational improvement
Pre-worked & reused correlations
Lessons Learned
10-15%
60-70%
85-90%
95%+

Demographics-based patient matching has inherent limitations in performance no matter how sophisticated the matching algorithm because demographic attributes by nature are often not unique across individuals and because many demographics evolve over time as an individual traverses the health care community. Nonetheless, with proper data quality control and algorithmic adjustment, demographic-based patient matching can achieve mathematically promising matching rates around 90-95%. Unfortunately, diverse operational issues within health care data sharing networks often compromise inter-organizational patient record matching performance.

When the process of patient demographics collection is not governed among exchange partners, significant data quality issues can be introduced and the match rate can be as poor as 10-15%. Common data quality issues include missing information, typographical errors, misspellings, and transpositions. Simple process improvements such as data validity checking, normalization, and downstream data cleansing can increase the patient matching rate to 60-70%. While no technology can completely overcome human ability to force errors, systems should be designed so that it is relatively difficult to make an error and easy to correct one. Where feasible, organizations should implement error checking processes. This should include examination of records that may not make sense with certain cohorts, examining overlays, and ultimately identifying patterns in patient matching errors.

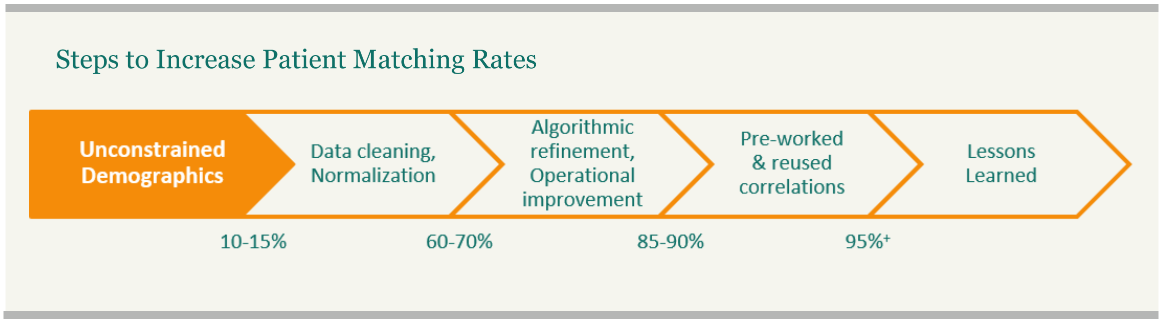
Further improvements in matching rates among organizations accrue as health care data sharing network operational environments are refined to address challenges such as network timeouts, message encoding/decoding inconsistencies, synchronicity of patient consent, etc. Successful patient matching and how it is defined will vary significantly across organizations. Matching rates take into account the number of records that are correctly linked with an existing patient identity, or a new identity created for a truly new patient. Appropriate technical and workflow solutions can increase patient matching rates that approach the mathematical limit of 100%, with an empirically determined limit in the 90% range.

To break through the inherent limitations of demographics-based patient matching, identity correlations of a so called "fragile" population should be proactively curated. (The term "fragile" in this context is a group of patients that frequently match incorrectly.) Such correlations can be established by pre-working and subsequent reuse of identity correlations determined from human review and investigation of the problematic (fragile) record pairs. Essentially, this cooperative approach allows patient matching based on reliable knowledge among disparate organizations and bypasses identity resolution based solely on demographic matching. As an additional benefit, the approach may compensate for minor lapses in operational rigor among the cooperating organizations. The matching rate can thus be increased to reach beyond 95% in Intermountain's case. The matching rate was improved through data normalization efforts and fixing operational barriers, such as consent and timeouts. The 95% match rate was validated through a human review which revealed that data quality and associated rules are a major factor in improving match rates.

Exploration of definitive technologies based on immutable personal attributes, devices, or traits demonstrate the potential for perfect identity resolution but have not been broadly adopted across communities.

Next, we will review the step-by-step approach taken by Intermountain Health Care as they share their process of optimizing patient matching with exchange partners.

Step 1: Small Sample Trial to Establish Baseline



Steps to Increase Patient Matching Rates

Unconstrained Demographics

Data cleaning, Normalization

Algorithmic refinement, Operational improvement

Pre-worked & reused correlations

Lessons Learned

10-15%

60-70%

85-90%

95%$^{+}$

The first step in the project was to establish a baseline for patient matching success across organizational boundaries. The following attributes were used for linking: First Name, Last

Name, Sex, Date of Birth, Phone Number, and Zip code. Empirical testing was established with a small sample selection of 10,000 patients that were known to both Intermountain Health Care and one of its exchange partners. Given that all patients in this small sample were known to have been treated by both organizations, they expected that a large majority of the patients would be successfully matched. A “gold standard” (known correct) dataset composed of accurate matched patient pairs was established. This gold standard dataset was built by leveraging human-reviewed linked patient record pairs from previous operational transactions. It included 340,000 pairs of linked patient demographic records. This important dataset established the benchmark upon which performance of various matching approaches could be accurately assessed.

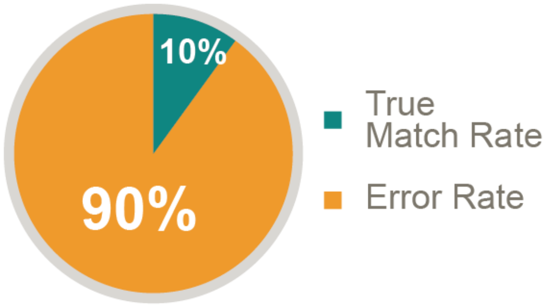
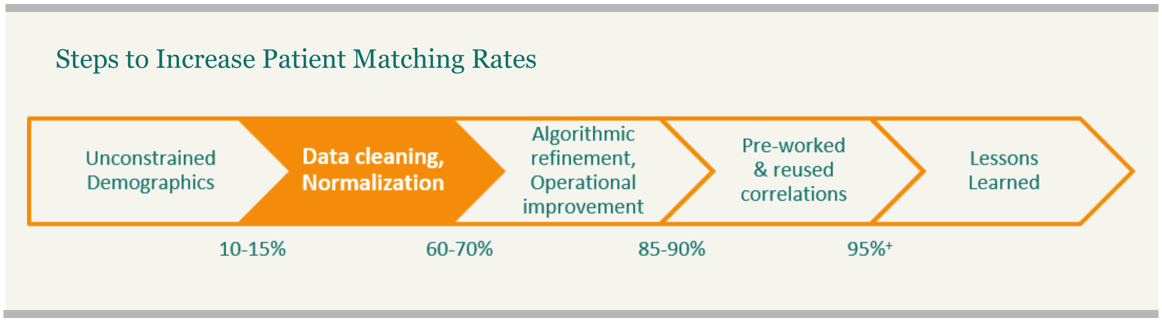


Figure 1: Initial Cross-Organizational Patient Match Error Rate

Surprisingly, the initial attempt in matching patients across organizational boundaries resulted in only a 10% match rate. Even though this was a test that was not intended for production, the outcome of only achieving a 10% match rate was unexpected. **The sample data were fraught with data quality issues.**

Step 2: Trait Analysis


Steps to Increase Patient Matching Rates
Unconstrained Demographics
Data cleaning, Normalization
Algorithmic refinement, Operational improvement
Pre-worked & reused correlations
Lessons Learned
10-15%
60-70%
85-90%
95%+

The next step for the organizations was to enhance the matching rate. They started by characterizing trait data to identify the identity attributes that contributed most to patient matching across these two organizations. This analysis was conducted using Intermountain’s internal Master Patient Index (MPI) database that includes 6.6 million patient records. Several characteristics where analyzed to determine those traits most likely to be useful.

Completeness: At what rate is this trait captured and available?

Validity: Is this trait known to be correct? Patient demographics consisting of default or temporary values (e.g. “Baby Smith” for newborn’s name) are complete but not valid.

Distinctiveness: Is the trait able to uniquely identify a person? For example, a trait such as sex (i.e. administrative gender) is not associated to a single individual, whereas a trait such as an MPI value is distinctive.

Comparability: Is the trait structured, coded (or numerical), or is it free text in string format? An address is an example of a relatively difficult to compare trait, whereas a social security number (SSN) can be easier to compare.

Stability: How much does the trait remain constant over a patient’s lifetime? Although examples exist to the contrary, traits such as gender, birth date, and Social Security Number tend to be relatively consistent over time. Other traits, such as current address, tend to change relatively frequently.

The table on the next page shows the results of an analysis of potential traits and their suitability for use in patient matching.

Patient Attributes Analysis

Attribute Name	Completeness	Validity	Distinctiveness	Comparability	Stability
MPI Identifier	100%	--	100%	Very High	Very High
Last Name	99.85%	99.84%	5.1%	Medium	High
First Name	99.85%	99.33%	3.1%	Medium	High
Middle Name	60.54%	60.54%	2.6%	Medium	High
Suffix Name	0.08%	0.08%	0.08%	Medium	Medium
SSN	61.40%	60.92%	98.0%	High	High
Sex (Admin. Gender)	99.98%	99.98	0.00008%	High	High
Date of Birth	98.18%	97.38%	0.8%	High	Very High
Date of Death	3.36%	3.36%	3.4%	High	Very High
Street Address (1 or 2)	95.00%	94.61%	44.4%	Low	Low

City	94.84%	94.83%	0.8%	High	Low
State	94.81%	94.39%	0.8%	High	Low
Facility MRN	99.90%	99.90%	99.90%	High	Low
Postal Code	92.31%	92.0%	0.6%	High	Low
Primary Phone Number	90.68%	87.26%	51.6%	High	Medium
Work Phone Number	20.28%	19.79%	51.6%	High	Low
Ethnicity	25.25%	25.25%	0.0003%	High	Very High
Race	76.25%	76.25%	0.0001%	High	Very High

Table 6: Patient Attributes Analysis

Items in Table 6 highlighted in green indicate desirable matching characteristics and are explored below. Items in yellow were identified as promising identifiers for future exploration.

The **MPI Identifier**, which is the internal enterprise-wide unique patient identifier, value has many desirable characteristics. It is always internally available since the systems in this study required an MPI value to be assigned as a prerequisite for all other clinical or data entry activities. It also should be valid, distinctive, very comparable, and very stable. However, the ability of an MPI number to be used across organizations is problematic. Since the EMPI identifier is normally specific to one organization, it may not be accepted by the exchange partner. Alternatively, if a Master Patient Index is **shared across** organizations, it can be a very valuable trait and is perhaps sufficient to establish high-confidence matching provided that demographic confirmation is also used to check the correctness of the link. Once this correctness has been confirmed, the linkage between patient identities is assumed.

However, the deployment of a cross-organizational MPI can be expensive and difficult from policy, legal, and technical perspectives. It becomes a new operational system that must be managed via a feed of demographic information from individual systems to the shared MPI, or the shared MPI must be used to actively assign patient IDs in near-real-time as patients are initially entered into their respective systems. Additionally, to remain accurate, the MPI must receive a constant feed of updates to the patient traits as they are corrected and/or change over time. The

> ... deployment of a cross-organizational MPI can be expensive and difficult from policy, legal, and technical perspectives.

processes of achieving acceptable MPI accuracy levels and diagnosing identified inaccuracies can be a very time consuming and expensive cross-organizational process. A cross-organizational shared MPI often receives a delayed feed, and in many cases, only receives incomplete information. The result is that a cross-organizational MPI can be of value, but it has significant limitations that do not exist when an MPI is used within a single organization. These limitations must be addressed in order to for a cross-organizational MPI to be successful.

One strategy for maintaining patient data integrity over time is involving the patient in data entry, correction, and maintenance. This includes making it a practice to ask the patient whether their address or phone number has changed at every visit and having the patient review the demographic information themselves to ensure its correctness. Some organizations require patient validation of existing treating physician relationships (which can also be used to identify impermissible disclosures).

Patients' first and last names also stand out in several regards. These traits are generally complete, valid, and stable. While last names can change during life events such as marriage and divorce, the trait is still considered relatively highly stable, as these changes typically only occur a limited number of times in a person's life. They are not, however, very distinct (5.1% and 3.1% respectively). In addition, they are only moderately comparable largely due to spelling variations, inconsistent use of special characters, inconsistent use of middle name, the use of nicknames vs. formal given names, and software support of patient names with more than three components. It should be noted that some standards (such as the IHE Patient Demographics Query profile) support additional demographics for newborns and other fragile populations, such as mother's maiden name, birth order number, and multiple birth flags. To account for naming convention variations, some matching algorithms process names by removing spaces and special characters, account for aliases, and/or tokenize names for patient matching purposes. For example, in a tokenization system, the name "Javier" and "Havier" would normally be considered the same token.

Patients' middle name is a very different trait in almost all regards (except for comparability and stability) than the patients' first and last names. The middle name is only present in the data set about 60% of the time and it is only valid about 60% of the time. It is also less distinctive, 2.6%, compared to first and last names.

The United States is a melting pot of cultures and peoples with multiple naming conventions and practices. Whenever possible, the matching process should include considerations for known practices such as cultures where persons born on the same day as a deceased revered ancestor or tribal leader are given the same name as the deceased. This situation creates the potential for a mismatch unless the "year of birth" is more heavily weighted in relation to other identifiers. Many persons from countries around the world try to accommodate an English spelling of a name when translating a name from the language of origin. Some of these translations can create long and unfamiliar names. This can be confusing to provider office staff and may create a reversal of first and last name from one health system to another. Even common English-origin names can also suffer from this problem, such as in the case of a person named "John George" where both the surname and the given name are both common surnames and common given names.

The **patients' suffix name** can be quickly disregarded given its exceedingly low completeness, validity, and distinctiveness (0.08%).

Outstanding Challenge

The **SSN** looks like a much more promising trait, as it scores very highly for all criteria used for this assessment (including a 98% distinctiveness score). But the SSN is fraught with other challenges including fraud, medical and financial identity theft, sharing by multiple individuals, and more. These issues are compounded by the fact that some organizations require the use of SSNs for matching purposes, and even have made internal assumptions that the SSN will be provided. In contrast, other organizations have banned the use of SSNs, or only allow SSNs to be shared under limited conditions, such as only sharing the last 4 digits. The outcome of these SSN-related issues is that it remains a contentious trait. It holds promise but is not in itself a solution. In addition, local policy makes it impossible, at the current time, to have cross-organizational patient matching that depends on this value being consistently available. The SSN trait remains a technical hurdle as well as an opportunity. One significant opportunity for improved use of SSNs will be explored later in this chapter.

Sex (more accurately referred to as administrative gender) is, as expected, largely complete, valid, comparable, and stable. However, also as expected, it is not distinctive (0.00008%). As part of Stage 3 Meaningful Use, vendors will leverage vocabulary standards for birth sex (male, female, and unknown), sexual orientation, and gender identity. It will be up to providers to establish workflows for accurately and consistently collecting this patient information in a private and efficient manner.

Date of birth stands out in multiple ways, including completeness, validity, comparability, and very high stability. Its distinctiveness (0.8%) in this specific analysis, when combined with other traits, made it a useful trait. Overall, it was one of the most promising traits.

Street address looked promising from the perspective of completeness and validity. It also provided good (44.4%) distinctiveness. However, it was ranked low from the perspectives of comparability and stability. It may be of use for patient matching approaches that look at a patient's address history.

Postal code and **primary telephone number** also look promising when combined with other demographics. They are relatively complete and valid (above 80%) and they are easy to compare with minor normalization effort. Even though postal code by itself if not highly distinctive, the fact of it being numerical and collected by most organizations can make it an attractive trait, when combined with other traits, for patient matching.

Additional traits to improve patient matching efforts may include: Maiden Name, Multiple Birth Indicator, Birth Order, Telephone Number(s) and types, and Email Address(es) and types. While many organizations store these data, many are not using these attributes specifically during patient matching or exchange of data.

In the future, the authors feel it is likely that biometrics will play a significant role in patient matching and identity proofing efforts. Internationally, biometrics are already used for use cases such as secure entry and unique identification. Examples of biometrics include fingerprint, palm veins, face recognition, DNA, palm print, hand geometry, iris recognition, and retinal scanning. Biometric devices are used to capture these metrics in a systematic and reliable way. Biometrics are considered immutable attributes, in that they are innate, entrenched, and would take significant effort to change. As such, biometric attributes are ideal for patient matching use cases and identity proofing. However, it should be noted that biometrics are often expensive to implement, maintain, operate and in some cases are rejected by end-users.

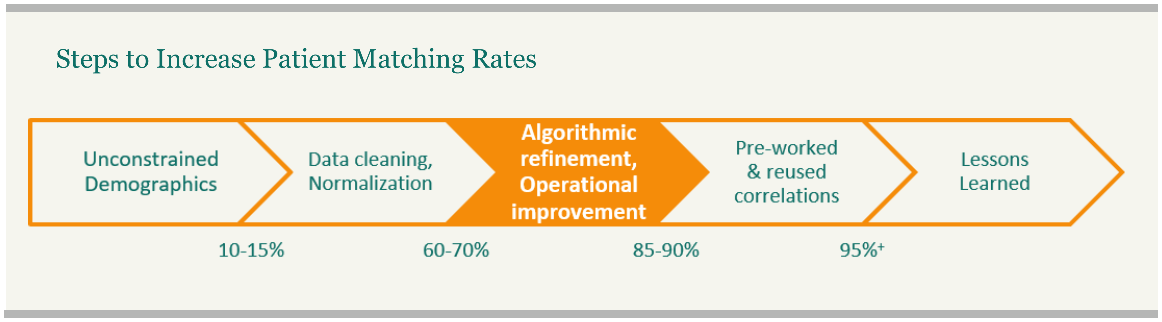
Some attributes may be affected by time range considerations, such as a timeframe when a patient lived at a given address or time period when they were known by a different name. Accuracy of patient addresses may be improved if matched with data from payer systems. To become a viable attribute, systems would need the ability to capture this information and organizations would have to commit to using this information.

Lessons learned from this analysis include:

1) More data do not necessarily mean better patient matching results. Depending on how sophisticated the matching algorithm is, traits with poor validity and comparability may cause a decrease in matching performance. Each organization should conduct a trait analysis on their internal patient population. The best combination of traits should be determined for each pair of exchange partners (or data sharing network), where feasible.

2) Most patients (>90%) can be uniquely identified by a combination of common demographic data elements (e.g. name, date of birth, address, etc.) when available.

Step 3: Offline Algorithmic Performance Measurement and Refinement


Steps to Increase Patient Matching Rates
Unconstrained Demographics
Data cleaning, Normalization
Algorithmic refinement, Operational improvement
Pre-worked & reused correlations
Lessons Learned
10-15%
60-70%
85-90%
95%+

After identifying and analyzing potential traits to leverage in the development of more effective patient matching algorithms, the organizations next identified and implemented relatively easy, high-value improvements. **Missing data were gathered; inaccurate data were corrected.** This was completed by referencing a gold master data set that was used across organizations. The master data set leveraged for this exercise was previously reviewed manually to confirm each pair represents the same person. Creation of this gold master data set was essential as it allowed for an authoritative result to be obtained.

Default values (also known as temporary values) merit special consideration. In this context, we define a default value as a data item that is known to be fictitious due to lack of information. Default values are commonly employed when organizational policy or software limitations require certain fields to be supplied even if the correct value for that field cannot be ascertained at the time. A common example is a hospital admission system which requires a patient name to be entered upon admission. If the patient's name is not known, staff are typically instructed to enter a value such as "Jane Doe." In this case, "Jane Doe" is the default value. Inside a given organization, default values are often not harmful. But across organizational boundaries, default values can be harmful if not properly managed. Default values, when sent across organizational boundaries, can sometimes have the effect of “contaminating” the receiving organizations’ MPI traits for that patient. Consequently, default values should never be exchanged across organizational boundaries.

For this case study, all known default values were inventoried and excluded from matching algorithms. Alternative name representations, such as nicknames or common misspellings, were accommodated during matching.

During this stage of the case study, it was observed that the patient identity data appears to become incorrect at a rate of 1% per month. Examples for which patient identity data (for valid reasons) “becomes incorrect” include patients legally changing their names, address, or telephone number(s). It has been noted that similar data, such as mailing lists and provider directories, also have been found to age at this approximate rate. Without intervention, these incorrect data can propagate across organizations. By implementing patient identity improvement processes, data will become incorrect at a slower rate, and ultimately be more manageable for an organization to remediate. Networks must continuously adapt as their patient databases grow and change over time. As networks grow, more sophisticated scalable solutions will be necessary for handling data degradation.

> ... patient identity data appears to become incorrect at a rate of 1% per month.

During this stage, seven combinations of traits were assessed to determine their predicted success in terms of completeness and uniqueness. Table 2 has been simplified, but remains “directionally correct” for these seven patient trait combinations. The completeness value indicates the percentage of the combined traits that had all the data present in the patients’ records, while the uniqueness value represents the percentage of matches that resulted in a unique match (identifying a single person). The table does not include estimates for true and false negatives, true and false positives, selectivity, or sensitivity.

Analysis of Patient Trait Combinations

Sequence	Combination of Traits	Completeness	Uniqueness
1	FN+LN+DoB	98.2%	95.7%
2	FN+LN+DoB+Sex	98.2%	95.9%
3	FN+LN+DoB+Sex+ZIP(first 5)	91.1%	99.2%
4	FN+LN+DoB+Sex+Phone	76.2%	99.5%
5	FN+LN+DoB+Sex+MN	59.9%	98.9%
6	FN+LN+DoB+Sex+MN(initial)	60.0%	97.7%
7	FN+LN+DoB+Sex+SSN(last 4)	61.9%	99.7%

Table 7: Analysis of Patient Trait Combinations

The first combination explored (sequence 1) was that of **first name**, **last name**, and **date of birth**. This combination resulted in 98.2% completeness while relatively good in terms of completeness, was relatively poor in terms of uniquely identifying a given individual (95.7%).

The next combination reviewed (sequence 2) was **first name**, **last name**, **date of birth**, and **administrative gender**. This had no discernable impact to completeness (compared to sequence 1) with a near-trivial improvement on uniqueness from 95.7% to 95.9%. This was regarded as an insignificant improvement over sequence 1.

In sequence 3, the prior sequence (sequence 2) combination was expanded to include **first name**, **last name**, **date of birth**, **administrative gender,** and the addition of the **five-digit zip/postal code**. This combination of traits reduced completeness by a significant amount, down to 91.1%, but dramatically increased uniqueness to 99.2%. This was a significant finding and allowed the creation of an algorithmic rule essentially stating that if these traits are all available for a given patient, then use this set of traits to match across organizations.

For sequence 4, the same set of traits was used as in sequence 3 with the substitution of a **telephone number** for the five-digit zip/postal code. This further reduced completeness to 76.2% reflecting a lower availability of telephone numbers, but it also provided one of the best uniqueness scores at 99.5%. This resulted in another rule indicating that if these five traits are available for a given patient, they should be used. This became the highest precedence rule—if these traits are available, then this is the first rule to be applied. However, if these five traits are not all available, lower precedence rules are utilized.

Sequence 5 used the same set of traits as was used in sequence 4, but with the substitution of a **middle name** for the telephone number. This reduced completeness to 59.9% and resulted in a relatively high uniqueness of 98.9%. A rule was created to use these traits, when available.

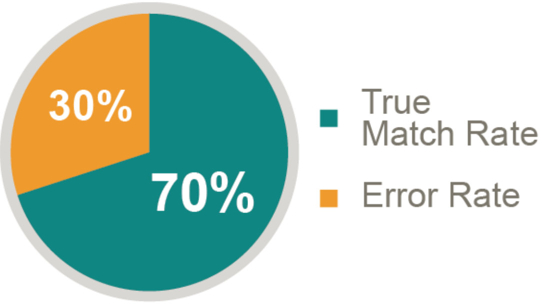


Figure 2: Initial Performance Analysis

In sequence 6, a rule similar to sequence 5 was created but it only used the **first character of the middle name** instead of the full middle name. This resulted in 60% completion with a reduced uniqueness, as expected, to 97.7%. So, this resulted in a rule that if no higher precedence rules were first applied, and if that patient's full middle name is available, then it is used to match. If the full middle name is not available, but the patient's middle name first character is available, then it will be used to match.

Finally, sequence 7 used the **first name**, **last name**, **date of birth**, **administrative gender**, and **last 4 numbers of the SSN**. This resulted in low completeness of 61.9% but the highest uniqueness at 99.7%. This too became a rule, with careful placement between the other patient matching rules.

It should be noted this was an inward-facing analysis. That is, it was a review of these traits, for those patients, only at Intermountain Health Care. It remains to be determined how these introspectively-derived rules would work across organizational boundaries.

Expectations Based on Low Effort, High Yield Data Rework

After performing the above steps, a new analysis was conducted using the same small sample data set to determine the matching results. **The analysis showed a marked improvement with a true match rate of approximately 70%.** This represents a significant improvement over the prior results. At this point, the involved organizations felt that match rates within their organization were adequate, but at the same time recognized the need to improve internal controls to improve external patient matching across organizations.

Next, testing of these new rules was expanded to include a larger sample data set of 340,000 patients.

When analyzed, Intermountain found that the patient matching success rate for this larger sample set was still unacceptably low. Note this match rate is where Intermountain expected to start from in terms of cross-organizational patient matching.

Additional data quality interventional work was performed, focusing on:

- Data entry control;
- Enforced data integrity checks; and
- Data transcription problems (from paper to EHRs)

A new analysis was conducted after making these internal data quality improvements. This analysis established the new **expected** values.

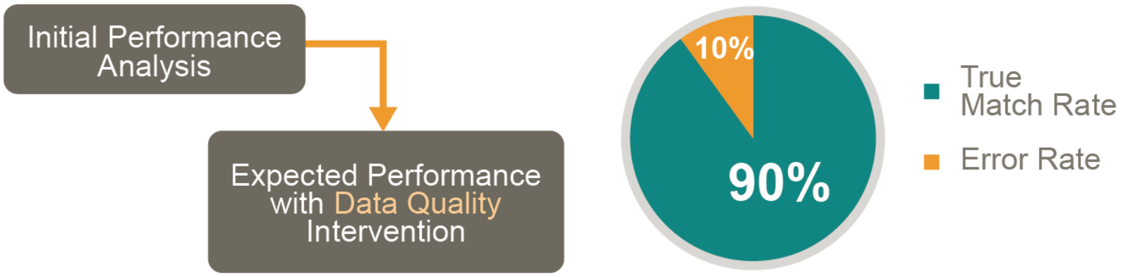


Figure 3: Expected Results After Data Rework and New Rules.

Next, the actual performance with data intervention was determined.

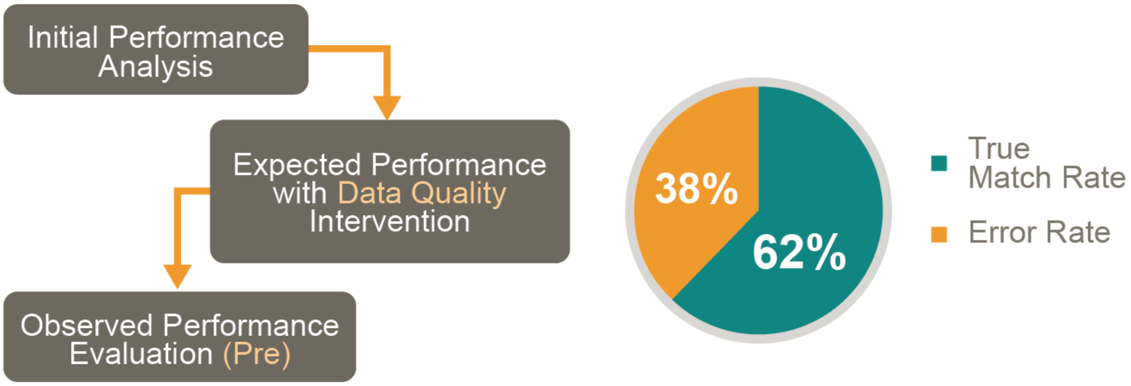
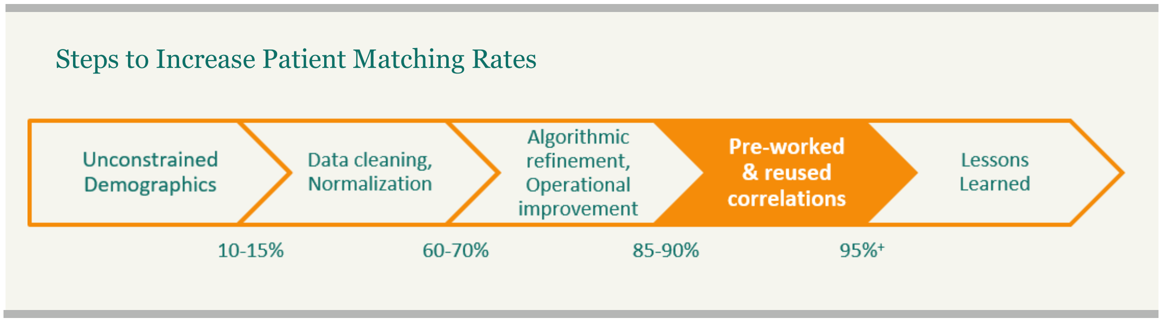


Figure 4: Actual Observed Performance After Data Interventional Improvements

This resulted in a 28% worse performance than expected (a 10% error rate was expected, not a 38% error rate). Further investigation was conducted.

Step 4: Operational Performance Measurement and Improvement


Steps to Increase Patient Matching Rates

Unconstrained Demographics
Data cleaning, Normalization
Algorithmic refinement, Operational improvement
Pre-worked & reused correlations
Lessons Learned
10-15%
60-70%
85-90%
95%+

Step 3 resulted in an observed correct match rate of 62%, which was far below the predicted 90% match rate. The disparity between the predicted match rate of 90%, compared to the actual results of 62% in Step 3, warranted further analysis to ascertain the causes of the failure rates.

The 38% error rate was broken down into more detailed contributing factors. As shown in Figure 5: Detailed Analysis of 38% Error Rate, we uncovered five major factors: algorithmic, authorization, network, messaging, and a more intrinsic error category. Systematically, Intermountain Health Care worked with their exchange partners to address each addressable issue.

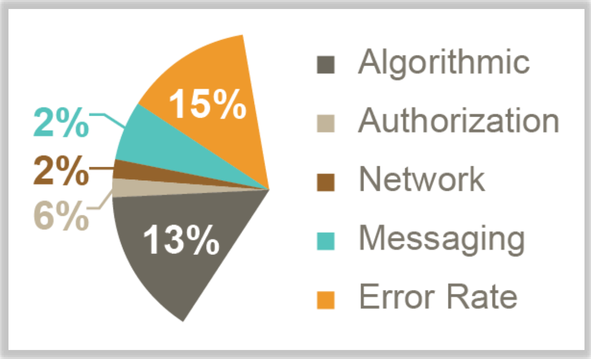


Figure 5: Detailed Analysis

Algorithmic: Various algorithmic patient matching improvements were implemented at this phase including data normalization, selection of traits, blocking strategies, bucketing strategies, as well as additional improvements. Data normalization efforts were completed in coordination with a Statewide MPI committee consisting of subject matter experts, providers, physicians, and payers. Fields and traits were examined and formats were standardized based on HL7 ADT feeds. Train-the-trainer educational documentation covering standard look up procedures and data normalization was shared with each organization to implement efficient (and scalable) staff behavioral changes across the involved organizations.

Authorization: The ability for an organization to honor a patient's authorization choice (such as opt-in or opt-out or fine-grained preferences) requires accurate patient identification. Although outside the scope of this paper, the lack of an appropriate patient authorization resulted in what appeared to be a failure to match when the root cause was a permissions issue. Timing was an unexpected factor with respect to the patient authorization issue. For example, in some cases a patient would "opt-in" but there would be a delay before that status was reflected in all systems. *The CCC is prototyping an innovative approach for addressing this issue. Those interested in learning more on this topic are encouraged to reach out to the CCC.*

It should also be noted that a lack of patient authorization, or an opt-out may skew match rates. Because many federated systems respond with a "not found" message when a patient has opted-out of information-sharing; a "not found" cannot be interpreted as anything but a "non-response".

IT Networking Issues: Some IT technical issues were also identified, namely, network timeouts. This class of issues manifested themselves as apparently failed patient matches, however, the root issue was that the responding system for a given patient matching request was not received before the initiating system gave up. An apt analogy is a telephone call. If the party being called doesn't answer after 15 rings, then the person placing the call may give up and disconnect. If the party being called would have answered reliably at 20 rings, then there was a failure that could have been avoided if the calling party waited longer. In a similar manner, Intermountain worked with a University partner (its exchange partner for this test) to configure both systems' network timeouts to higher values. As an aside: we are intentionally avoiding the issues of service levels and use case-driven response time requirements at this stage, but we will come back to this topic in the future. In summary, some apparent patient matching failures were actually due to network timeout issues.

> ... apparent patient matching failures were actually due to network timeout issues

Security Header Issues: In addition, a number of the failures were attributed to technical errors unrelated to patient traits. As this case study leverages the eHealth Exchange (which is based on an open standard called IHE Cross Enterprise User Assertion), each message has an internal technical component called a security header. This component is part of the wrapper around each patient matching request message between exchange partners. Inside this security header is a very important set of data indicating the purpose of the request (such as for treatment, claims, patient authorized exchange, or emergency), the requesting person, the requesting person's role, the

> ... patient matching was effectively blocked by a technical error unrelated to patient traits

patient's authorization, and much more. These data are contained in a section of the message called the Security Assertion Markup Language (SAML) header. A number of issues were found to exist around the generation and/or consumption of the SAML header where, for example, the SAML header could not be properly understood by the receiver resulting in an error. The implication of this class of issues is that patient matching was effectively blocked by a technical error unrelated to patient traits. An analogy is when a physical package is not delivered because the recipient address on the package was illegible. This class of issues were resolved by the technical implementation teams.

Data Encoding: A number of additional issues were related to data formats. This was a key area of improvement for this case study. It was originally assumed that data would be consistently formatted internally and across organizational boundaries. Unfortunately, this was far from being true. Most fields, in fact, had different representations that required normalization before they could be compared. Telephone numbers had to be tagged so that work telephones were not being compared with mobile telephones. And telephone numbers had very little consistency in the use of special characters. These issues were addressed by removing all special characters and normalizing them to an international standardized format. It was also discovered that in many cases first names were being combined with middle name or initials. This was also corrected. Another large group of problems was associated with the use of names with non-alphabetic characters (e.g. O'Toole) or internal spaces (e.g. Van Der Camp). **Project-wide conventions were developed and applied to normalize these names.**

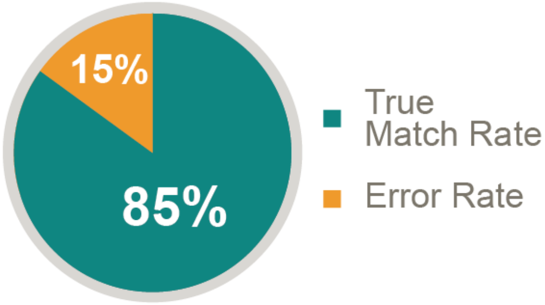
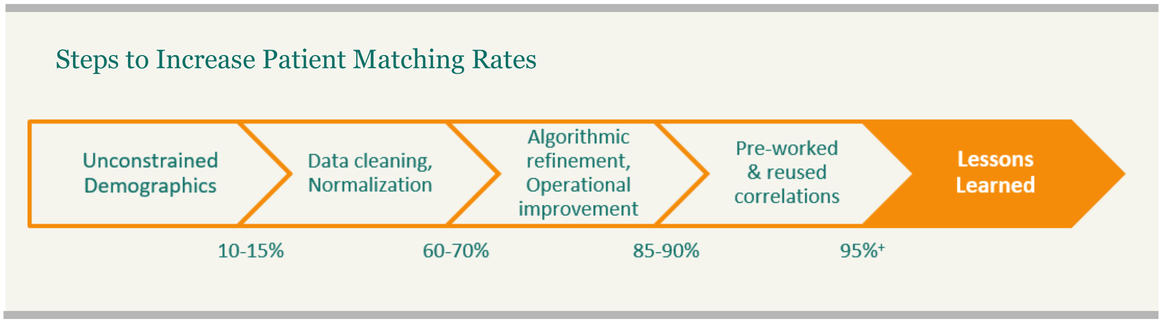


Figure 6: The True Match Rate

As a result of these improvements, the true match rate improved to approximately 85%. That is, about 15% of the matching problems still remained. Obviously, this was a significant improvement with the best results to date for the study. Next, a number of best practices were utilized to make further incremental improvements.

Step 5: Implementation of Best Practices and Lessons Learned


Steps to Increase Patient Matching Rates

Unconstrained Demographics
Data cleaning, Normalization
Algorithmic refinement, Operational improvement
Pre-worked & reused correlations
Lessons Learned
10-15%
60-70%
85-90%
95%+

To address the remaining 15% error rate, a number of best practices were identified and implemented. These practices included:

- Working systematically with partners to reduce or eliminate prior issues;
- Applying results of prior work (see Lessons Learned);
- Agreeing with trading partners on data standardizations;
- Agreeing on consent synchronization;
- Expecting no less than a 90% match rate across organizations;
- Keeping investment reasonable by agreeing when goals are achieved;
- Focusing on scalable solutions;
- Pre-working fragile identities when possible (see Lessons Learned);
- Improving the human workflow; and
- Leveraging CCC Shared Services.

After employing these identified best practices, Intermountain expected a successful match rate of approximately 90% as shown in Figure 7: Expected Error Rate at this Phase. However, as seen in Figure 8: Actual Results, at this stage they achieved a 95% match rate, which was significantly better than expected. The 95% match rate was the most conservative rate of the samples. Data quality and associated rules were a major factor in these rates. The purpose of this white paper is to improve match rates and provide lessons learned that can be replicated throughout the industry through minimal acceptable practices. These lessons learned can later be scaled through conformance rules and testing criteria.

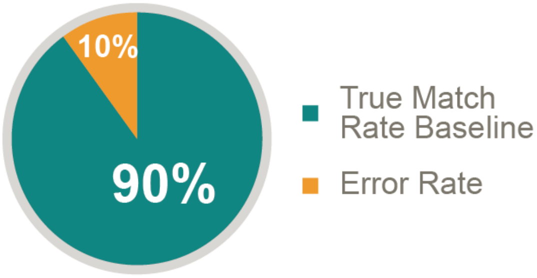


Figure 7: Expected Error Rate at this Phase

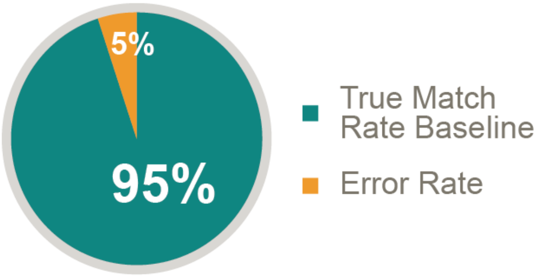


Figure 8: Actual Results

After a review of the reasons for the increased patient matching success rate of approximately 95%, the major factors contributing to this successful result were identified as:

- The use of the CCC Shared Services;
- Collaborating with external exchange partners regarding standardized data formats; and
- Addressing patient consent issues.

In the next section, overall case study lessons learned will be shared.

Lessons Learned

> ... a category of patients emerged that repeatedly failed to match correctly...

Fragile Identities: One interesting outcome of the case study was the emergence of a category of patients that repeatedly failed to match correctly (false positive match or a false negative match) even after multiple interventions. These patients' traits were manually edited to enable successful matching, but later the patients again failed to match. This process repeated several times (failure to match, rework, match, and a failure to match) during this study. As a result, these patients were put into a category called "fragile identities." Examples of fragile populations with thin demographics include newborns and trauma patients. Some of these patient identities were studied in detail and the root causes of their repeated failure to correctly match were identified. In most cases, the repeated matching problem was due to "thin" demographics, such as just a first and middle initial instead of full first and middle names, a missing address, or an address using non-standard abbreviations. There were also a few outlier cases that patients have the same date of birth, address, last name, and very similar first name (twins). A strategy was developed to (1) identify the characteristics of patients in this category, (2) query for and create a work list of patients falling into this category, and (3) implement manual remediation of those patients' identities to **proactively** attempt to resolve future patient matching problems. The case study demonstrated the effectiveness of the approach in a **reactive** scenario, but it is expected that by providing a proactive remediation standard operating procedure, we can eliminate this entire class of patient matching issues. Demographic collisions are more likely with thinner demographic data. These can be reduced by increasing the depth of demographic information by such processes as direct patient outreach.

To help identify fragile populations, staff may use "potential match" reports to identify those records with a high likelihood of matching yet failed to automatically match because one or more records were missing key information. Methods to increase the robustness of demographic information may include the use of patient portals with an identity strength test, or training for registration clerks to work with patients to collect additional information.

Issues Unique to Pediatrics:

- No national naming convention for newborns, specifically, patients who have not yet received their legal name and have a temporary name.
- Multiple birth persons present challenges with same date of birth, address, mother's maiden name and potentially very similar names and identifiers often only differing by a single digit:
    - Children's Health in Dallas, Texas was working with 3 MPI vendors, all well recognized in the market for data integrity: one a regional data warehouse for state reporting, another a regional HIE, and a bolt on HIE product. All 3 reported a 22% duplicate rate. After looking at the list of patients – all were multiple births, there were no duplicates. The 3 vendors could not accept the multiple birth indicator that Children's captured. Children's Health was forced to undertake a very onerous process to prove their data was accurate to each of these three vendors before the vendors would create additional capabilities to prevent an over-lay with potential inaccurate information. All 3 vendors remarked that Children's was the only organization to bring this to their attention.
    - Multiple birth patients are often named similarly:
        - Same first name, different middle name – John David Smith, John Daniel Smith
        - Reversal of first and middle name – John David Smith, David John Smith
        - Similar first and/or middle name – Sarah, Sarai

Outstanding Challenge

        - Different first name, same middle name – John James Smith, Joshua James Smith
- Newborns do not have a Social Security Number (SSN) or government–assigned identification at the time of birth.
    - SSN is often the highest weighted item used in patient matching algorithms.
- Gender identity for patient and families:
    - Ambiguous gender at birth
    - Transgender (gender at birth, legal gender & preferred gender identity)
- Non-traditional families' designation of mother or father
- Care provided in-utero and the need to generate a medical record prior to birth episode
- Guardianship/custody of child
- Proxy designation for patient portal

As the boundaries of health care delivery continue to evolve, nationally there is more emphasis on population health, wellness, and delivery of care at the patient's physical location. An area of increasing focus is delivery of health care in-utero. Traditionally, a medical record is not created until after the birth event. Technology and practice should change to provide complete and accurate medical records both for the mother and fetus. Fetal centers provide services to identify fetal anomalies and provide in-vitro surgeries with a goal of assuring optimum continuum of care. The integrity of the mother's medical record is compromised if a procedure completed on the fetus is recorded on the mother's record, especially in encounters where the treatment provided to the fetus, such as medication, is different to the treatment provided to the mother. The infant's medical record is incomplete without accurate capture of clinical information for in-utero diagnostic and interventional procedures. The authors recommend that the industry follow the Children's Hospital Association's temporary demographic conventions for newborn.

Current standards limit gender identification to male, female and unknown. Standards, technology, and process should be robust enough to accurately capture gender as part of the foundation for a complete medical record.

Considerations for handling Pediatrics:

- Standards adoption
- Information governance, process, and technology
- Vendor capture of multiple birth indicator, birth order, and mother's maiden name
- Creation of medical record prior to birth event

Well Behaved Group: In sharp contrast to the "fragile identities" group, another group of patient identities emerged that exhibited the opposite behavior: their identities seemed to almost always match correctly (true positive or true negative match). This group was regarded as the "well behaved group" with respect to patient matching. Members of this group were analyzed to determine the reason(s) this group exhibited desirable matching behavior. The single most important factor distinguishing this group from the fragile group was the presence of complete demographics. These patients had a full, correctly spelled name including their middle name and any special characters. The patients had a complete current address and telephone number. These patients also had historical name information and historical address information. The traits of this group of patients are being used to inform the best practices for the other patient groups including the human workflow implications such as ensuring hospital admissions staff are trained and motivated to enter and use robust and complete patient traits.

Knowledge Reuse: Manual work on a patient's identity is expensive, slow, and error prone. It has a large negative impact on the speed that patients can begin treatment. However, it represents very valuable information gained. Once a patient's records have been manually analyzed and remediated, that information can and should be leveraged in the future to prevent repeated manual rework on the same patient record. For example, a correction made to a record should not be made in a read-only system rather it should be updated via an approved secure workflow to a master patient record. This allows future patient matching activities to leverage the improvements made from the manual process and ultimately makes the organization more efficient. In a similar manner, the links created between patient records also can and should be leveraged. Those linking decisions should be stored in such a way that they can be reused in the future.

Involve Patients in Identity Management: Patients themselves are valuable allies in helping maintain their identities. Two methods have been identified so far. A critiquing service involves patients, at the point of care, in helping to link, correct, unlink, and otherwise update the patient records. It is also envisioned that in the future a patient portal (or other self-service application) could perhaps help patients understand their identity completeness in a manner similar to a password strength test offered by many sites and applications. Post case study, other organizations have indicated that they have involved patients in identity management practices.

Observations and Recommendations

- The biggest opportunity to immediately impact matching rates is standardized formats for demographic data among data sharing participants. Standardized formats are lacking or have not been widely adopted for many of these important data elements, including telephone (TEL formats) and Social Security Number format.
- Consistent name representation will be a challenge without probabilistic assistance because of data collection workflow issues that favor alternate representations (such as preferred name over legal name).
- Increased patient match rates (above 95%) may require a supplemental identifier in addition to the required fields. A supplemental identifier may include a national or regional shared identifier, such as a driver's license number. High data quality at the point of capture is essential for acceptable patient match rates. This allows for probabilistic linking where alternative representations are allowed among the exchange participants and where established linkages are expected to be reusable for future exchange transactions.
- A data cleanup step is needed for accurate patient matching. However, data cleanup may add cost and can be defined differently by each organization. A cleanup standard is needed, where data cleanup can be standardized to a reference point. AHIMA's Information Governance Structure provides a model for data integrity and long-term data maintenance.

Outstanding Challenge

QUESTIONS TO ASK YOUR ORGANIZATION

1. Have we documented our default or temporary values? How do we prevent these values from being transmitted to our exchange partners?
2. Are our staff trained and actually capturing high-quality patient identity data? Does their workflow encourage them to properly match patients or does it encourage them to create a duplicate patient?
3. Are we normalizing addresses against a standard, such as the USPS? Are we normalizing all other fields so they are comparable?
4. Are all our patient demographics data as complete as possible (full middle name, prior names, prior addresses, etc.)?
5. Are we capturing the telephone type (home, mobile, work) as well as the number itself?
6. Do we capture additional fields that can be of use in matching, such as email addresses?

   Note: Direct Email addresses or standard email addresses may be collected and used to improve matching. It will be necessary to clarify type of email address at the point of capture. In the same vein, indicating type of address and type of phone number can improve match rates.
7. Are our matching rules going to work between organizations, such as with a federal agency or another state, which may not have/use/supply the same patient matching traits and rules? Do we have a clear understanding of which exchange partners will use and supply SSNs to us? Do we understand which organizations require us to provide them with SSNs?
8. How do we handle patient consent with respect to patient matching?
9. Are we using all available information for matching such as prior names, addresses, telephone numbers, etc.?
10. Are we only using strict character-by-character matching? (The answer should be no as this is not a scalable solution. Matching algorithms could use “fuzzy matching” with weighted scoring and phonetic searches to account for similarly spelled names, and should be tested.)

Case Study in Review

The next generation of patient matching is still on the horizon. Health care is still in its infancy with respect to patient identity management between organizations (also more correctly known as record linkage). Many other domains have studied patient matching at an industry and academic level for many years. Several industries, specifically financial services and airline transportation, have legislative support for unambiguous matching of their customer records. However, legislative support for patient identity matching is not assumed or suggested in this draft framework. We can perhaps learn from the consumer credit reporting bureaus whom have been working on this problem for many decades. Today, credit bureaus use approximately 140 separate traits to match people to the correct database record. Not only are consumers matched, but their physical addresses are also matched in a similar way as patients. Additionally, credit bureaus link consumers to a web of other entities that provide credit to that consumer. The net result is not perfect, but it allows for a national or international scale solution that can inform our work within the health IT domain.

The CCC is also innovating and developing multiple next-generation patient matching approaches. Several of their key approaches include:

- A critiquing service to involve patients and providers in the patient matching workflow at the optimal time with a feedback loop to leverage such knowledge gained;
- Authoritative sources on a field-by-field basis;
- Identification and re-work of fragile identities;
- Staff incentives;
- Consent shared services;
- Patient matching/shared record locator services;
- Data quality analysis; and
- Empirical analysis across several organizational boundaries.

As Intermountain continues to strive for perfect patient matching within and across organizations, they anticipate that these improvements will allow for patient matching rates to exceed 99%. Proposed future improvements may include:

- Use of biometrics;
- Proactive correlations; or
- Patient engagement in identity management.

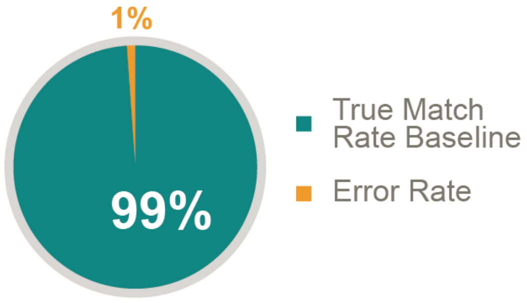


Figure 9: Future Match Rate

Chapter 3: Cross-Organizational Patient Matching Maturity Model

Introduction

As mentioned in Chapter 1, patient identity management has remained in the national spotlight as a key prerequisite to successful health information exchange. The purpose of the maturity model, described in this chapter, is to provide a method to evaluate, measure, and improve patient matching deployments across organizational boundaries. The proposed maturity model is designed to provide a simple framework aiding in the comprehension of this domain and to focus on process change. We believe that more precise definitions of the maturity model will give organizations the ability to adopt more advanced patient identity management in a methodical manner.

This framework is based in part upon the International Organization for Standardization (ISO) framework (which includes people, process, and technology) with the added dimension of governance.

Scope

Patient matching is often thought of in two very different domains: (1) patient identity management within an organization and (2) identity management across organizational boundaries. The scope of this paper is largely focused on patient matching across organizations. While there is overlap, and these areas will also be discussed, this paper does not focus on patient identity management inside organizations otherwise.

Characteristics of Mature and Immature Organizations

While there are many ways for measuring the overall maturity of an organization, here are some general characteristics the authors feel can help delineate immature and mature organizations which, in turn, helps define our patient matching maturity roadmap for improvement.

Immature organizations generally possess the following characteristics:

1. Processes are improvised
2. Known processes are commonly ignored
3. The organization is in reactive mode
4. Schedules, staffing plans, and budgets are not fact-based
5. Quality is sacrificed and results are variable
6. Quality is not objectively measured

Mature organizations generally possess the following characteristics:

1. Coordination, communication, and collaboration across silos
2. Work plans are generally realistic and accomplished for common project types
3. Process and practice are largely in agreement
4. Processes improve over time
5. Staff understand their responsibilities and there are no key gaps in staffing or skills
6. Management and staff are aligned

Summary of Levels

Carnegie Mellon University's Software Engineering Institute (SEI) is generally regarded as having created the first Information Technology maturity model. (http://www.sei.cmu.edu/cmmi/index.cfm). There are five levels defined for their model, which, according to the SEI, "Predictability, effectiveness, and control of an organization's software processes are believed to improve as the organization moves up these five levels. While not rigorous, the empirical evidence to date supports this belief." The SEI Capability Maturity Model's (CMM) Five Maturity Levels of Software Processes are:

1. Initial (chaotic, ad hoc, individual heroics) - the starting point for use of a new or undocumented repeat process.
2. Repeatable - the process is at least documented sufficiently such that repeating the same steps may be attempted.
3. Defined - the process is defined/confirmed as a standard business processes.
4. Managed - the process is quantitatively managed in accordance with agreed-upon metrics.
5. Optimizing - process management includes deliberate process optimization/improvement.

AHIMA's Information Governance Structure, available through the AHIMA website (http://www.ahima.org/topics/infogovernance), provides an overview of AHIMA's identity management practice levels. The AHIMA model is complementary to this Sequoia whitepaper as it provides a maturity model to assess patient matching **inside** an organization, whereas the Sequoia maturity model addresses patient matching **across** organizational boundaries.

The Sequoia Project believes these levels are a valuable construct to help guide our thinking about cross-organizational patient matching. Each level provides minimal acceptable criteria for achieving more advanced patient identity matching processes. Through advanced application of minimal acceptable criteria, cross-organizational patient matching efforts will be improved both internally for organizations, and through increased trust of partner organizations.

Level 0: Indicating ad hoc processes and outcomes, and little to no management oversight or recognition;

Level 1: Indicating adoption of basic defined processes with associated repeatable outcomes, and limited management involvement;

Level 2: Indicating increasing maturation of processes, definitions of most key processes, data governance, algorithm use, active management involvement, and accumulation of quality metrics;

Level 3: Indicating advanced use of existing technologies with associated management controls and senior management awareness and use of quality metrics. Data governance is incorporated into the process with initial implementation of an information governance program; and

Level 4: Indicating innovation, ongoing optimization, and senior management active involvement. An information governance program is firmly in place with executive support and operational integration, and the value of information is recognized. Data integrity with respect to patient matching also includes external reconciliation with quality processes.

At Level 0, people are disorganized, processes are not well understood or defined, and the overall view of cross-organizational patient identity management is that of a chaotic system because of the lack of replicable results. Success often depends on individual heroic efforts. The organization is often in reactive mode instead of one of proactive management.

At Level 1, there is a growing awareness of the critical role of cross-organizational patient matching and a corresponding recognition of the need to apply basic management controls by lower and mid-level management staff. At this level, some processes are repeatable, but not all. Success is more predictable. The quantity of reactive mode events declines. Level 1 organizations, and above, have implemented all applicable *Cross-Organizational Patient Matching Minimal Acceptable Principles,* as described in chapter 4.

At Level 2, all key processes related to cross-organizational patient matching are understood and documented. They may be enforced inconsistently. The organization is normally not in a reactive mode, and unexpected events become relatively rare where they were the norm in Levels 0 and 1.

At Level 3, organizations monitor, analyze, and systematically improve their ability to manage patients across organizational boundaries. Most if not all processes are defined and documented. The processes are somewhat rigid. However, at Level 3, the processes are largely documenting the system behaviors "as is," as opposed to Level 4 where the processes are innovative. At Level 3, the organization achieves consistency, but it is not optimal.

At Level 4, innovation becomes a standard component of patient matching. Management uses data accumulated to model and, as is deemed viable, implement sometimes significant improvements. Key staff members are considered leaders in this domain and contribute to the community.

The time an organization has existed is not necessarily strongly correlated to an organization's maturity. Each level will likely have to introduce innovation to advance to the next level.

Data Governance

Patient Matching is one component of an enterprise data governance strategy. The authors assert that organizations with unsophisticated patient matching strategies are more likely to spend additional funding on third party vendors tasked with data cleanup efforts. These organizational costs are often passed on to patients and families covering the costs of care. As such, to improve overall patient care and quality, organizations should endeavor to establish information governance and data governance practices that best fit organizational needs.

AHIMA defines "information governance" as an organization-wide framework for managing information throughout its lifecycle and supporting the organization's strategy, operations, regulatory, legal, risk, and environmental requirements.

AHIMA defines "data governance" as the responsibility of the business unit. It is the policies, processes, and practices that address the accuracy, validity, completeness, timeliness, and integrity of data. AHIMA is committed to advancing information and data governance in the health care industry to ensure the quality and integrity of all types of information necessary to safe, high quality, cost effective care and the improvement of the health of individuals and populations. For additional resources on information and data governance, please visit AHIMA's website (http://www.ahima.org/).

Because Health information Exchange organizations (HIEs) represent multiple health care organizations, they will likely have a membership comprised of organizations operating at various levels of maturity. While it may not be feasible for the HIE to set the maturity bar for all organizations within its membership, successful patient identity matching between HIE members will accrue from greater identity management maturity among all its members. Therefore, HIEs should be encouraged to provide educational materials about best practices for patient matching and assist immature organizations where possible.

Level Characteristics

Workforce

- At Level 0, an organization has staff focused on **internal** patient identity and matching with no dedicated staff **outwardly** focused (i.e. on patient matching across organizational boundaries). In the unlikely event that some staff are focused on outward patent matching, they do so because of a personal recognition of the need, instead of this recognition occurring at an organizational level. Or they are focused on cross-organizational patient matching due to a limited scope project. Workforce patient identity management formal processes, training, skills development, and career path are not recognized. Staff members are not trained. Job titles do not exist for patient identity management staff; cross-organizational patient identity management is often a responsibility that is added on to other positions.
- At Level 1, management has recognized the need for specific assignments for external patient matching and have started formulating plans.
- At Level 2, staff are devoted, at least part-time, to cross-organizational patient matching.
- At Level 3, staff include formal responsibility for cross-organizational patient matching. Training is accepted as necessary and appropriate. Staff are involved with industry initiatives. The organization has deployed techniques to assess the relative volume of external patient identification is employed by that organization and can thus assign appropriate workforce to meet this organizational need.
- At Level 4, staff involved in patient identity management are involved at more senior levels within the organization and are leading innovation with respect to this topic. See also Standards Development characteristics.

Patient Involvement

Patients should be engaged as an active participant in their identity management as early as possible in order to educate patients about the importance of proper patient identification and further reduce instances of patient matching errors. Accurate patient matching is a pre-requisite for the ability to enable patients to become more engaged in their data exchange.

- At Level 0, Patient involvement in their identity management does not exist. At Level 0, organizational staff do not consider the patient to be a part of the identity management workflow other than to confirm their demographics when checking-in or registering.
- At Levels 1 and 2, policies and practices have been put in place for training, partner outreach, and patient education at the first point of contact. The patient is recognized as a potential active participant in their identity management. This may include manual workflows with visual oversight.
- At Level 3 the patient is involved via manual workflows and processes such as through a patient portal and patient engagement efforts, but no system changes are made to accommodate such involvement.

  At Level 4, the patient is recognized as a key ally in optimal patient identity management. In addition, at Level 4, as patients become involved in their own identity management the knowledge gained is durable, shared across the enterprise, and reused for subsequent cross-organizational patient identity management. Multi-factor identity proofing is added to supplement the use of demographic data.

A challenge for patient involvement in patient identity proofing is the ability to curate data for minors, homeless populations, patients with dementia, etc.

Outstanding Challenge

Use of Technology

Provider use of technology at all levels is assumed, at least for administration purposes, but technology access and literacy for patients remains an outstanding challenge, especially for certain groups such as those at risk for housing stability, those that are not technology literate, etc.

Outstanding Challenge

- However, the deployment of limited technology at Level 0 is largely built around ad hoc processes and standards, such as using custom data interfaces that are not fault tolerant, robust, performant, or well documented.

- At Level 2, the deployment of technology is largely built around ad hoc processes and standards, such as using custom data interfaces that are not fault tolerant, robust, performant, or well documented.
- At Level 3, a technical process for data integrity and reconciliation has been developed by a strong performing third party vendor, or custom developed and is available via the patient identity management solution.
- At Level 4, the organization has developed new technology, is continuously testing their innovative technology, and is submitting refined version to Standards Development Organizations (SDOs) to help advance the industry.

Communication and Community Involvement

- In terms of participation in communications within the health information technology communities, organizations at Level 0 are largely isolated. This results in them being unaware of standard approaches to common problems and in deploying solutions to problems they believe are unique which, in fact, are common.
- At Level 1, organizations become aware of standards and communities and begin formulating plans to begin participating in broader communities.
- At Level 2, organizations are involved in appropriate health IT communities, such as those curating relevant standards, state or regional exchanges, and state-wide approaches.
- At Level 3, organizations are fully integrated into most relevant health IT communities, such as SDO committees.
- At Level 4, organizations exhibit leadership in relevant communities such as by co-chairing workgroups and testifying in front of state and national legislative bodies and agencies. They participate in board, state, and federal advisory committees, etc. At Level 4, reporting capabilities also exist within the product to identify areas of opportunity to enhance patient matching and identity management.
- At Levels 3 and 4, key staff members also frequently share negative and positive knowledge gained to help others understand patient matching problems and solutions better so that they may leverage prior work.

Workflows

- At Level 0, workflows are based on speculative needs and are not driven by confirmed, high-priority use cases.
- Workflows at Levels 1 are largely driven by payment requirements and the desire to meet federal regulatory requirements.
- Workflows at Level 2 fully meet federal regulatory requirements.
- Level 3 workflows are driven by more advanced objectives such as full round-trip immunization query, administration, update, and reporting. Cross-organizational

partners are partially incorporated into workflows. Data governance principles are in place that define data integrity and subsequent requirements.

- Level 4 workflows are driven by advancing the state of the art and tracking adherence to the best demonstrated practices. Level 4 includes optimization of workflow to incorporate partner organizations. Information governance, which includes data reconciliation processes and quality controls, is also in place. Workflows include automated capabilities to maintain a comprehensive data location set for each patient, recover gracefully from episodes of identity theft and data breaches, and permit the patient to participate anonymously in research, education, and public health activities.

External Matching Focus

Organizations that are at Level 0 often do not yet understand that rules of patient identity management that work for them within their enterprise do not necessarily work across organizational boundaries. One example is an organization that makes no distinction between patient matches across organizations from those that occur internally. It assumes that patient demographic feeds, including merges, links, unmerges, unlinks, and demographic updates are occurring both for internal patient matching and for their partner organizations externally. This manifests as policies and procedures for patient identity management that are not viable because they cannot enforce their internal enterprise policies and procedures across organizational boundaries. A more specific example of this is that the organization uses a master patient index (MPI) configuration that is only effective at matching patients if the patient's SSN is provided. Internally, they can enforce that policy; externally, they cannot.

- Level 0 includes organizations that often do not yet understand that identify management rules and processes that work for them within their enterprise do not necessarily work across organizational boundaries. Data attributes that can be standardized at Level 0 include: last name, first name, date of birth, gender, middle initial, race, primary phone number, and address.
- At Level 1, there is no defined reconciliation process. Data attributes that can be standardized include: middle name, mother's maiden name, prefix, and marital status.
- At Level 2, an ad hoc reconciliation process has been implemented. Data attributes that can be standardized include: alias or previous name(s), use of standard data definitions for address (e.g., USPS ZIP code), last four digits of Social Security Number and/or driver's license number and/or passport or alien ID number.
- At Level 3, a periodic (such as daily) reconciliation process has been implemented. Data attributes that can be standardized include: multiple birth

designation, birth order (if a multiple birth), birthplace, e-mail address, previous address(es), and phone numbers.

- At Level 4, a quality assurance process has been implemented. Data attributes that can be standardized include: insurance ID/policy number, insurance plan, previous insurance, Medicaid ID, Medicare ID, and Biometric ID.

Testing

- Testing of identity management solutions at Level 0 is minimal, manual, ad hoc, and does not consistently ensure successful deployment of “passed” systems.
- At Level 1, testing programs should at least validate name, address, and date of birth.
- At Level 2 and above, testing is largely automated, based on significant real-world lessons learned, and is a good predictor of a successful deployment.
- At Level 3, testing programs predict, with a high degree confidence, successful deployments
- At Level 4, an enforced testing program, with a defined level of acceptance, exists with published definitions.

A future goal would be to develop national level guidance for testing, and possibly develop a test harness to ensure standardized patient identity testing is available nationwide.

Use of Patient Matching Quality Metrics

- At Level 0, the value of patient matching quality metrics is not recognized.
- At Level 1, the organization has begun to recognize the value, and has started planning for the future capture of metrics.
- At Level 2, quality metrics are in place with data definitions for each metric. Information governance concepts are also introduced with support by executive leadership.
- At Level 3, the metrics are being used to actively improve. Executive leadership is actively engaged and supports data stewardship. Initial training is implemented and integrated into the fabric of the organization.
- At Level 4, the metrics are being further refined, and include feedback loops to the systems and organizations involved in patient identity management. Their external health IT trading partners join in metrics capture, use, and feedback. As patient matching becomes a recognized success, these successful patient matching data can be used as a strategic asset to help enable aspirational objectives such improving fraud detection, population health, and research. Engagement with external health IT trading partners includes business partners as well as consensus on metrics and definitions.

Diagnostic Approaches

Diagnostic approaches vary with the level of maturity (ad hoc, some automation, full automation, innovative approaches). **At Level 0, all patient matching exceptions and errors require human intervention.** Processes are not well understood or documented. Errors are often not recognized in a timely manner. Manual work queues are not consistently staffed. Management has little to no visibility into exceptions (frequency, types, root causes, impact, remediation plan, etc.).

- At Level 0, no intentional patient matching diagnostic automation exists.
- At Level 1, minimal automation is in place with ad hoc manual intervention required.
- At Level 2, a moderate amount of automation is in place, but human intervention is still needed to use and leverage this automation.
- At Level 3, full automation is in place but there is still a need for human intervention in exceptional cases. Based on metrics (as defined by the organization), the need for human intervention should be less than 0.5 percent.
- At Level 4, an innovative, fully automated approach is implemented that does not require human intervention.

Trust in Matching Processes

- At Level 0, trust in matching processes and performance is brittle. The overall patient matching system is not well regarded by end-users, administrators, or management. It is considered error prone and is not trusted to be reliable or available. It often returns unexpected patient search results or errors. False positive and false negative matches occur frequently. Manual disambiguation of returned patient searches is frequent. User disillusionment and abandonment is common at this level. Clinical users, in particular, see their hopes of improved patient medical records availability dashed and stop using the system.
- At Level 1, trust in matching process is improving but still inconsistent. There is no awareness of the value of data integrity or patient matching.
- At Level 2, trust in matching process is not recognized as having a return on investment "ROI", but the activity is being completed as a task. There is also a general lack of understanding about the impact or importance of patient matching.
- At Level 3, end-users see a personal ROI of patient matching efforts resulting in significant payoff thereby accelerating adoption of cross-organizational patient

matching into their personal processes. Organizations incorporate use of cross-organizational patient records as a best demonstrated practice or standard of care.

- At Level 4, organizations have feedback loops with senior representatives of their staff to identify innovative approaches. This results in the identification of new patient identity management strategies that are novel and valuable. Patient identification is a significantly more stable and reliable capability that requires fewer resources. Resources that were allocated to matching are freed up for new tasks such as monitoring, maintenance, and improvement strategies using feedback loops.

Management Oversight

- At Level 0, management oversight is virtually non-existent. Management may not even have a firm grasp of the definition of cross-organizational patient matching and records exchange. The business and clinical value of cross-organizational patient matching is not recognized.
- At Level 1, management awareness has increased and basic management controls are being defined. Training for staff is ad hoc and formalized training programs are lacking.
- At Level 2, management is actively involved in cross-organizational patient matching. Initial management controls have been implemented, are being used, and being improved. Processes for maintaining staff core competencies are in place. Metrics are being captured but are not yet being fully used.

- At Level 3, management is leveraging metrics. Senior management is aware of the importance of cross-organizational patient matching as being of strategic importance as a prerequisite for other activities such as care summary exchanges. Management ensures that processes for maintaining staff core competencies are in place with a feedback mechanism.   Management has defined what success looks like for the organization. Workflow is reviewed and optimized at a system-wide level to ensure that patient matching dependencies, such as proper staff incentives, are in place.
- At Level 4, management has empowered the organization to assume a leadership role in the industry. Innovation projects are funded and staffed. Innovations, once proven, are incorporated into production operations. Knowledge is shared with SDOs and with the wider community via significant industry involvement. Senior management includes at least one member that is focused on cross-organizational identity management as a formal area of

responsibility. Processes for maintaining staff core competency are in place and have expanded to business partners. Patient matching will require minimal management oversight. The organization has achieved their definition of success for patient matching. In addition, the organization has achieved success as defined by its data exchange trading partners.

Use of Industry Standards

Use of industry standards and calibration/backfill of those standards is very different across levels.

- At Level 0, the organization may use custom solutions that are not intentionally based on standards.
- At Level 1, there is some use of standards that are not well understood. The organization may not be aware of the benefits or drawbacks to using specific standards. Organizations at this level may also have the naïve belief that the standards will solve more problems than they actually do (such as data quality issues).
- At Level 2, organizations have deliberately chosen standards and have passed sample message validation. Organizations understand that standards have limits, but they leverage their capabilities fully and provide backend system support, such as more advanced internal algorithms, to make the best use of these standards.
- At Level 3, use of standards has matured, organizations understand that standards have limits, but they leverage those capabilities fully and provide for backend system support, such as more advanced internal algorithms, to make the best use of standards. Organizations also work with SDOs to fix errors and vagueness in the standards.
- At Level 4, organizations take the initiative to create new standards and to suggest significant improvements to existing standards. A standardized patient identification strategy is fully functioning and evolving to keep up with organizational needs.

Establishment of Feedback Loops

- At Level 0, feedback loops do not exist; nor does the recognition exist that feedback loops are of value.
- At Level 1, feedback loops are established with a few primary data sources.
- At Level 2, feedback loops are established with most data sources and other key workflow participants.

- At Level 3, feedback loops are established with all participants, human and system, in the patient matching process inside an organization.
- At Level 4, feedback loops are established with all participants, human and system, in the patient matching process across internal and external organizational boundaries. Feedback loops are used for improvement of fraud detection, remediation of data breaches, and analysis of public health and research work.

Fragile Identities

Each level includes engagement and management of external business partners when managing fragile identities. Note that the term “fragile identity” is defined elsewhere in this document.

- At Level 0, fragile identities are not recognized.
- At Level 1, data aging due to issues such as legal name changes are recognized and managed.
- At Level 2, cultural variations and conventions are recognized and managed.
- At Level 3, organizations recognize that some patient identities are “fragile” and tend to consistently be false negatively matched or false positively matched. This can be due to demographics that are not sufficiently rich or that are very similar to other people, or errors such as an incorrect identifier.
- At Level 4, organizations recognize fragile identities and implement an associated process to systematically identify identities in this category, assign appropriate staff to remediate these fragile identities, and then measure the results to confirm the resolution. Organizations also analyze well-behaved identities, learn what characteristics make these identities largely immune from mismatching, and then leverage this knowledge to help manage their entire census. Organizations work with their exchange partners to help identify and remediate fragile identities as a formal part of their on-going processes.

Pediatric Matching

As mentioned elsewhere in this paper, there are many significant challenges with respect to neo-natal and pre-natal patient matching, including a potential lack of a name or even a birth date.

- At Level 0, the special challenges associated with pediatric identities are not recognized.
- At Level 1, these challenges are becoming apparent, but little progress is made.

- At Level 2, a program is enacted to assist with the multitude of internal system changes that are required to enable matching of pediatric patients between organizations.
- At Level 3, vendor and custom systems are deployed supporting pediatric matching across organizational boundaries, but significant gaps remain including reliance on algorithms that are not yet fully tuned for pediatric matching.
- At Level 4, the organization is able to successfully match pediatric patients with external exchange partners.

Flow Down

In this context, the term “flow down” refers to legal agreements between organizations such as the eHealth Exchange DURSA.

- At Level 0, organizations do not have any special provisions in their various legal agreements covering the organizations with which they contract to enforce any degree of patient identity management practices.
- At Level 1, organizations have contractual language agreeing to partner with external organizations on patient identity matching.
- At Level 2, organizations have initiated discussions with external organizations on updating contractual language regarding patient identification matching.
- At Level 3, organizations recognize that their patient matching processes will only have limited utility unless they obligate vendors and organizations connected to their exchange partner to patient matching principles. Terms covered in these binding agreements include data quality, data completeness, and key workflow components such as dealing with minimization of duplicate records, minimal patient matching practices, exception handling, and service levels.
- At Level 4, organizations must obligate their internal patient matching data sources, consumers, vendors, and systems to the same patient matching principles. Data sharing networks obligate their network members and HIEs obligate their participants to comply with patient matching principles.

Knowledge Sharing

- At Level 0, knowledge gained is often lost since the organization is largely in reactive mode and is often “fighting fires.”

- At Level 1, the organization has some recognition of the value of capturing knowledge but there is no formal process for capturing it.
- At Level 2, knowledge about patient matching processes is captured and shared internally to a limited extent.
- At Level 3, the knowledge is shared with partners and is starting to be shared with the broader health IT community.
- At Level 4, the knowledge gained about a specific patient match is implemented in automated systems that leverage the information broadly and durably. For example, take an organization that has manually resolved an external patient matching investigation with a partner and identified consent as being the root cause. The organization has a method of incorporating that knowledge into their patient matching data and processes, allowing this entire class of issues to be permanently resolved. Knowledge of patient matching is still important but the organizational focuses on patient matching as a prerequisite for enabling proper medical care while evolving strategies to keep up with system and organizational changes.

Temporary (Default) Values

- At Level 0, default or known temporary values are defined for the organization.
- At Level 1, temporary values are inventoried and defined for the organization and its key exchange partners. The inventory is not enforced through user interface data capture, though, and occasionally temporary values are discovered that were previously missed. Partner temporary values inventories are also incomplete.
- At Levels 2 and above, the inventory of known temporary values is accurate.
- At Level 3, technological enforcement of known temporary values is in place including at the staff data capture levels as well as at the automated data exchange levels. Staff processes are also in place to enforce this list. As new partners begin the onboarding process, temporary value inventories are exchanged as part of the formalized process.
- At Level 4, both new and established partners exchange temporary value inventories.

New Partner Connectivity

- At Level 0, each new data exchange partner is connected using an ad hoc process. Patient matching attributes are not documented or are incompletely documented. Partner data considerations, such as quality and availability, are not accounted for during the planning and implementation process. This leads to many production-level unfulfilled expectations and errors, which is part of the reason why the Level 0 organization is often seemingly fighting fires. They simply have not yet gained enough experience to proactively manage, predict, and resolve common issues with partner patient identity management.

- At Level 1, key patient matching considerations are documented and considered during the implementation process with each new partner or for the network(s) the organization participates within.
- At Level 2, key patient matching processes with new data exchange partners are included into manual testing processes. Patient identity management considerations are well documented into complete, implementable specifications. Standards are adhered to, when possible, and manual testing confirms adherence.
- At Level 3, patient identity management testing processes are automated and patient identity management related on boarding processes are forward-looking and request adherence to best practices.
- At Level 4, heuristics and advanced processes allow for deeper insights into data exchange partners to understand their patient identity systems, processes, and workflows. This results in advanced configuration of partner systems' integrations to optimize their success in patient matching.

Data Quality

- At Level 0, data quality is an unknown. Technical staff and management have very limited awareness of the importance of data quality, or the status of their data.
- At Level 1, data quality has been identified as a key component of patient matching across organizations. The organization may attempt to side-step their own data quality issues by asking their exchange partners to make adjustments to their data. No formal analysis of data quality exists, but there is a growing awareness of the need for formal control of data quality.
- At Level 2, data quality has been assessed and is well understood. The organization understands its data quality situation, and that of its primary exchange partners.
- At Level 3, training is provided to staff on workflow considerations as well as education on the impact of not following training guidelines.
- At Level 4, performance improvement is occurring on a routine basis and is embedded in the project. Performance evaluations include quality measurements of staff for accountability.
- For more information see Chapter 1 and the Feedback Loops topic in this chapter. The majority of topics discussed in the case study presented in Chapter 1 are ultimately focused on the central topic of data quality.

Characteristic	0	1	2	3	4
Workforce	○	◔	◑	◕	●
Patient Involvement	○	◔	◑	◕	●
Use of Technology	○	◔	◑	◕	●
Communication & Community Involvement	○	◔	◑	◕	●
Workflows	○	◔	◑	◕	●
External Matching Focus	○	●	●	●	●
Testing	◔	◑	◑	●	●
Use of Patient Matching Quality Metrics	○	◔	◑	◕	●
Diagnostic Approach	○	◔	◑	◕	●
System Stability	○	◔	◑	◕	●
Management Oversight	○	◔	◑	◕	●
Use of Industry Standards	○	◔	◑	◑	●
Establishment of Feedback Loops	○	◔	◑	●	●
Fragile Identities	○	○	○	○	●
Pediatric Matching	○	◔	◑	●	●
Flow Down	○	○	○	◑	●
Knowledge Sharing	○	◔	◑	◕	●
Temporary (Default) Values	◔	◑	●	●	●
New Partner Onboarding	○	◔	◑	◕	●
Data Quality	○	◔	◑	◕	●

Table 8: Overview of Characteristics by Level

Summary

This chapter is designed to start a broad discussion on the need to define cross-organizational patient identity management characteristics in terms of levels of maturity. The Sequoia Project and the CCC are hopeful that this paper gives management a useful model for methodical assessment and improvement in cross-organizational patient matching. Moreover, we hope this leads to national-scale improvements in our ability to accurately exchange patient information while honoring patient privacy preferences, and ultimately providing better care and outcomes to those patients.

Chapter 4: Cross-Organizational Patient Matching Minimal Acceptable Principles

Introduction

The Sequoia Project, in collaboration with the Care Connectivity Consortium, has identified patient matching and identity management as a key national impediment to successful health data sharing across organizational boundaries. In this chapter we are providing **a list of minimal acceptable cross-organizational patient matching principles**.

This list of principles has been created in response to real-world production experience supporting large-scale health data sharing endeavors. These principles are intended to establish minimally acceptable expectations that an organization must meet in order to exchange with organizations adopting Cross-Organizational Patient Identity Management Maturity Model Level 1, as described in Chapter 3: *Cross-Organizational Patient Identity Management Maturity Model*. Some rules seem obvious; however, experience has shown that there is a lack of consistent application of these rules in production across the nation. This paper aims to begin illuminating and improving the consistency of the application of these rules. Organizations that adopt the model should consider enforcing these principles through their appropriate governance mechanisms. **We believe this list of principles will serve to create a component of a "Level 1" adoption model that organizations can target, test against as appropriate, and declare conformance to.** Additional levels beyond level 1 are described in Chapter 3.

Unlike previous chapters in this paper, Chapter 4 is oriented towards a technical audience.

Context and Next Steps

1. We intend that these 'minimal acceptable principles' will be introduced into the health IT community gradually in order to avoid the occurrence of breaking changes and to provide opportunity for vendors and health IT implementers to adapt and improve their patient matching approaches and success rates. The recommendations will initially be proposed as guidelines and evolve over time to become official policy, and then finally become part of testing programs. The proposed phases are described in more detail, as follows:

   Phase 1: Adopt principles as <u>guidance</u>. During Phase 1, these patient matching rules will be considered guidance, and will not be enforced through testing programs and will not be a condition of joining or participating in health data sharing networks.

   Phase 2: Principles will become official policy but will not be tested. In this phase, the "MAY/SHOULD" constraints will change to "MUST" constraints, other than as noted below. “SHOULD NOT” constraints will change to “MUST NOT” when the underlying issue is requirement. When the principle is a best practice, constraints will remain “SHOULD NOT”. It should be noted that the focus is on minimal acceptable practices, rather than best practices.

Phase 3: Principles will become an enforceable condition of testing and onboarding processes for new or existing data sharing partners. At this point, the list of patient matching rules will become part of the PASS/FAIL testing criteria. Hence, organizations not meeting these criteria will not be allowed to claim adoption of Cross-Organizational Patient Identity Management Maturity Model Level 1, nor enter into production until the identified deficits are remediated. Please keep this in mind when assessing the rules. The testing criteria should be developed by the community via an open, inclusive, consensus-based process and approved via applicable data sharing connections or network formal change management processes, including associated testing of organizations currently in production. This criterion applies to operational networks adopting principles.

2. These practices should be adopted as soon as is practical. It is expected that some of these rules can be implemented at any time, with little to no negative impact to exchange partners. Other rules, such as those based on workflow, will likely require partner coordination in order to more effectively facilitate adoption.

> Workflow rules will require partner coordination before adoption to avoid breaking changes.

3. We anticipate that adoption and additional factors will generate feedback to iteratively improve and refine these patient matching practices. This list of Level 1 rules will likely evolve with implementation and lessons learned.
4. Organizations adopting these principles must obligate their internal patient matching data sources, consumers, and systems to do the same. Similarly, a data sharing network must legally bind its network participants, who, in turn legally bind their participants. An organization must legally bind its vendors and systems.

Adopters of these principles are generally expected to deploy them initially as “SHOULD” type of constraints, with the intention to change them to “MUST” constraints as quickly as possible. In a similar manner, “SHOULD NOT” constraints will be changed to “MUST NOT” constraints, with exceptions as noted below. The below proposed principles are specifically intended for those organizations exchanging using the IHE International Cross-Community Patient Discovery (XCPD) standard. However, it is expected that many of these practices will also assist those exchanging using other standards.

Traits & Identifiers

Below are rules designed to improve the use of traits and identifies for cross-organizational patient identity management and matching. These are intended to be implemented by technical staff who manage and maintain patient identity matching systems. Patient identifiers SHOULD be consistent, not reused, unchanging, and should prevent the creation of duplicate patients at partner sites.

Category	Patient Discovery Initiating Gateways	Patient Discovery Responding Gateways	Details
Required traits	Patient Discovery Initiating Gateways SHOULD query using all traits required by the underlying specifications. In addition, where optional traits are known to be of "high quality", then participants SHOULD query using all possible optional traits.	Patient Discovery Responding Gateways SHOULD use and respond with all traits required by the underlying specifications. In addition, where optional traits are known to be of "high quality", then the query response SHOULD include the best known traits possible optional traits.	The term "high quality" may vary from organization to organization. As a result of variation in the industry, the term "high quality" does not have a defined threshold that can be broadly applied. Organizations and networks are encouraged to identify their threshold for determining "high quality" traits.
Patient Trait truncation	Patient traits transmitted by Patient Discovery Gateways to other Patient Discovery Gateways SHOULD NOT be truncated.	Patient traits transmitted by Patient Discovery Gateways to other Patient Discovery Gateways SHOULD NOT be truncated.	
Temporary or default values	Patient Discovery Initiating Gateways SHOULD NOT transmit any temporary or default value for any patient trait as this can contaminate the partner gateway’s patient traits and/or result in false negative matches. Although XCPD Initiating Gateways SHOULD NOT transmit temporary values, if it is known that this operation will not harm any exchange partner, then XCPD Initiating Gateways MAY transmit temporary values.	Patient Discovery Responding Gateways SHOULD NOT reply with any temporary value for any patient trait as this can contaminate the partner gateway’s patient traits and/or result in false negative matches. Although XCPD Responding Gateways SHOULD NOT transmit temporary values, if it is known that this operation will not harm any exchange partner, then XCPD Responding Gateways MAY transmit temporary values.	See appendix for a definition of temporary values. Temporary values vary by organization. More work is needed to identify consistently used temporary values across organizations. An example of "harm" would be if their partners add patient records to their system based on inbound XCPD queries.

Category	Patient Discovery Initiating Gateways	Patient Discovery Responding Gateways	Details
Specific identifiers	Patient Discovery Initiating Gateways SHOULD NOT require the use of any specific identifier or value such as SSN unless such a trait is required by the applicable specification or standard. Any existing policy or statutory requirements related to the use of SSNs for patient matching still apply.	Patient Discovery Responding Gateways SHOULD NOT require the use of any specific identifier or value such as SSN unless such a trait is required by the applicable specification or standard. Any existing policy or statutory requirements related to the use of SSNs for patient matching still apply.	It is anticipated that networks and members of those networks will either directly reference IHE XCPD or have a more constrained implementation specification. For example, Carequality would reference the Query Implementation Guide, the eHealth Exchange would reference the Patient Discovery Specification.
Patient identifiers	Patient identifiers SHOULD be consistent, not reused, unchanging, and should prevent the creation of duplicate patients at partner sites. A patient identifier SHOULD NOT be constructed in such a way that it dynamically changes based on the known identity of that patient at that time.	Patient identifiers SHOULD be consistent, not reused, unchanging, and should prevent the creation of duplicate patients at partner sites. A patient identifier SHOULD NOT be constructed in such a way that it dynamically changes based on the known identity of that patient at that time.	Systems should not be allowed, for example, to simply concatenate a list of all patient identities together and return that value as the patient ID, since that list of all known patient identities can change at any time.
Multiple patient identifiers	Patient Discovery Initiating Gateways SHOULD NOT supply more than one patient identifier, per assigning authority.	Not applicable	
Identifier life spans	Patient Discovery Initiating Gateways SHOULD NOT make any assumptions about how long a partner's patient identifier will be valid. Organizations that maintain internal correlations between internal patient identifiers and external patient identifiers SHOULD implement the behavior described in Exception Handling #2 and #3 below.	Patient Discovery Responding Gateways SHOULD NOT make any assumptions about how long a partner's patient identifier will be valid. Organizations that maintain internal correlations between internal patient identifiers and external patient identifiers SHOULD implement the behavior described in Exception Handling #2 and #3 below.	For example, one organization may retire a patient identifier after a set period of time with inactivity, whereas others will retain an identifier indefinitely. Varying policies will impact ID lifespans.

Category	Patient Discovery Initiating Gateways	Patient Discovery Responding Gateways	Details
	Alternatively, organizations that maintain internal correlations SHOULD implement their systems so that they always issue a XCPD request before contemporaneous XCA Query for Documents/Retrieve Documents requests.		
Subsequent requests	Not applicable	Patient Discovery Responding Gateways SHOULD NOT require identical demographic traits on subsequent requests, as were used on the initial request, if the same identifier provided on the initial correlation is re-used on subsequent requests.	By identical traits, we are referencing the exact same number of traits and the exact same value in each trait supplied.
Multiple ambiguous matches	Initiating Gateways that are unable to handle multiple ambiguous matches should indicate so prior to querying a Responding Gateway.	If applicable to their internal architecture, Patient Discovery Responding Gateways MAY return multiple ambiguous matches per Assigning Authority. Also note that this “MAY” constraint will remain a “MAY” constraint after the remainder of these rules change to “MUST” constraints. Responding gateways SHOULD handle multiple ambiguous matches per Assigning Authority.	In this use case, only one match will be returned, or nothing will be returned.
Duplicate patient records	Not applicable	Patient Discovery Responding Gateways SHOULD NOT return duplicate patient records or return the same patient record in such a way that a duplicate record will	Internally, the systems behind XCPD/XCA Exchange SHOULD NOT return the same patient with the identical data using different assigning authorities or identifiers.

Category	Patient Discovery Initiating Gateways	Patient Discovery Responding Gateways	Details
		be created by the XCPD Initiating Gateway.	The same patient should use the same identifier for each request or response. This is important to prevent duplicate patients from being created.

Table 9: Rules for Traits & Identifiers

Matching Algorithms

Below are proposed rules to improve matching algorithms for cross-organizational patient identity management and matching. These are intended to be implemented by technical staff who manage and maintain matching algorithms.

1. Patient Discovery Responding Gateways SHOULD track patient identity trait changes and SHOULD respond based on prior or current (historical) demographics. Organizations that use historical records for patient matching have a higher maturity rating.
2. Patient Discovery Responding Gateways SHOULD match based on normalized traits. In addition, the initiator accepting or rejecting the match should also use normalized traits to do the reciprocal match. Normalization of traits is a minimal accepted practice.
3. Both Patient Discovery Responding Gateways and Patient Discovery Initiating Gateways SHOULD use case insensitive matching.
4. Patient Discovery Responding Gateways SHOULD NOT use exact character-by-character matching. In the future, organizations will be able to test minimal acceptable matching criteria.
5. Other than immediately above, these rules will not define the specific algorithms to be used, or avoided, since specific algorithms are system, vendor, data, and organization dependent.

Exception Handling

Below are proposed rules to improve exception handling for cross-organizational patient identity management and matching. These are intended to be implemented by technical staff who configure and manage exception handling for patient identity systems.

1. Patient Discovery Responding Gateways MAY return an error indicating that obtaining patient consent may allow different, presumably more, information to be returned. This change will be implemented as per a timeline determined by

the Patient Identity Management Maturity Model Level 1 Adopter. Note that this behavior only applies to Responding Gateways that would deny access based on lack of consent, and it only applies if returning such an error itself is not an impermissible disclosure.

2. An organization's patient identifiers SHOULD NOT be reused for different patients, but the identifiers are allowed to be permanently decommissioned and a new identifier may be assigned the same patient. If a patient is merged, unmerged, linked, unlinked, or undergoes a similar transaction, the XCPD and XCA Responding Gateway SHOULD permanently decommission the identifier or identifiers formerly used to represent the patients subject to the merge, unmerge, link, or unlink. The XCA Gateway SHALL generate an error for all subsequent Query for Documents or Retrieve Documents requests using that decommissioned patient identifier. Systems are not required to decommission identifiers if their internal logic is such that correct and complete patient data are returned for that identifier.
3. XCA Initiating Gateways SHALL have logic in place to correctly process Query for Documents or Retrieve Documents errors indicating that a patient identifier has been decommissioned such that this triggers a new XCPD Patient Discovery request.
4. Patient Discovery Initiating Gateways SHOULD use the XCPD "revoke" transaction to indicate that a previous correlation made by a partner SHOULD BE revoked.
5. Patient Discovery Responding Gateways MAY accept the XCPD "revoke" transaction and, if they do, they MUST revoke the correlation.

Chapter 5: In Closing

Many organizations are dedicated to implementing the highest possible quality patient matching. At the same time, many of those same organizations have (or will) experienced significant and unacceptable error rates when matching patients across organizational boundaries. The enclosed case study, maturity model, and minimal acceptable principles are meant to share knowledge regarding this important domain for the betterment of all health electronic data exchange.

Term Definitions

Terms	Definition
Algorithmic	A finite set of unambiguous instructions that, given some set of initial conditions, can be performed in a prescribed sequence to achieve a certain goal and that has a recognizable set of end conditions. In the domain of cross organizational patient matching, refers to a set of mutual expectations for a cross organizational model for processing instructions. The algorithmic term is an umbrella term for various processes that can be used for patient matching.
Authorization	The process of giving someone permission to do or have something. For example, authorization can work to prevent what otherwise would have been a successful patient match and should be tracked as a separate workflow that can impact patient matching if not properly mitigated or resolved. In addition, organizations might wish to consider including or deliberately excluding patient match failures due to authorization issues. This is a pre-requisite for matching in some environments such as non-HIPAA covered entities.
Blocking/ Bucketing Strategies	Internal-use attributes with values constructed from attribute values. The intent is for an MPI to define its shared attributes so that identities with shared attributes are much more likely to find matches for each other than identities chosen randomly from the entire data set, reduces burden on the system rather than completing a one to one match against every patient in the system. Bucketing groups specific attributes for use in comparison during the candidate selection process. Examples of possible buckets are: Last Name + Phone, or Last Name + Email Address, or Last Name + First Name. Buckets are defined during the initial configuration of a matching algorithm can have one or more attributes per bucket.
Comparability	Collecting data fields in the same format and way. For example, reviewing an official identification card and recording DOB in the MMDDYYYY format. Comparability may include the standardization of data elements.
Completeness	Availability of all requisite data fields and attributes. Lack of missing data values. Even if a record shows completeness, it may include default/temporary/dummy values.
Data Characteristics	Data characteristics include whether data fields are precise, valid, reliable, timely, complete, available, and unique.

Data Encoding	Encoding is the process of converting data into a format required for a number of information processing needs, including: Program compiling and execution.
Data Normalization/ Standardization	The process of taking a field such as SSN or full name and converting it into a format the follows certain prescribed rules across cross organizational boundaries. A process of achieving standardization.
Default/ Temporary/ Dummy Values	A trait associated with a patient that is known to be incorrect due to lack of information. Temporary values are often created when clinical IT systems require that a value be entered even if that value is not available, such as if a patient has not been identified. Temporary values are often short-term in nature. Examples of temporary values are an SSN of all 1s (111-11-1111) or a newborn name of “Baby Jones.” A pseudonym is not a temporary value as it is intended to be a substitute patient identifier with a specific purpose, such as to protect the privacy of a public figure.
Deterministic	Records are determined to refer to the same patient if they have an exact match based on a subset of data, i.e. name, DOB, SSN. Deterministic matching utilizes a one-to-one comparison of characters and numbers in data fields.
Distinctiveness	Easily distinguishable when comparing a given patient across organizational boundaries and with respect to other similarly characterized patients.
Error Rate	The combined rate of incorrect negative matches and incorrect positive matches.
False Negative	Failure to match two records that represent the same person or patient. Also referred to as a Type II error, refers to a classification error by the matching algorithm where the two patient records are marked as referring to two distinct patients but in reality the two records refer to the same person.
False Positive	A pair of patient records that appear on a duplicate report that are not in fact true duplicates. Also referred to as a Type I error, refers to a classification error by the matching algorithm where a record pair is marked as a match but in reality the two records refer to two distinct patients.
Fragile Identities/ Populations	Populations for which minimal or no data may be available for identification, such as trauma patients and pediatrics.
Identifier	An individual's attribute that may be used to establish identity. This may include demographic attributes such as address or phone number, or biometric attributes such as blood type.
IHE	Formally known as Integrating The Healthcare Enterprise, IHE is a health care information technology standards body.

Immutable Personal Attributes	An immutable object is any sort of physical attribute which is perceived as being unchangeable, entrenched and innate.
Interoperability	Interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged without special effort.
Match Rate	The combined rate of correct negative matches and correct positive matches.
Minimal Data Set	The fewest number of fields needed in a patient record used for comparison against other patient records.
Overlay	An overlay occurs when one identity's information overrides the information of another, different record. For instance, if an existing record is updated from John Smith to Jane Jones, and the demographic information is changed, this is an overlay.
"Patient" in reference to MPI	In a master patient index, the patient refers to the individual for which data attributes are compared against other attributes within the system to determine whether they are the same individual, also referred to as consumers, individuals, covered lives, recipients, etc.
Patient Discovery Initiating Gateways	A system sending outbound IHE XCPD and XCA requests.
Patient Discovery Responding Gateways	A system receiving inbound IHE XCPD and XCA requests.
Pediatric Demographics	A special set of demographics that may be captured based on what is known about a fetus or a newborn.
Proactive correlations	Establishing procedures to identify matches prior to creation of a duplicate record for a single patient identity.
Probabilistic	The process of using statistical analysis to determine the overall likelihood that two records match. Also known as fuzzy matching, it calculates the probability of a match.
Security Header	A security enhancement that includes security context information to the beginning of SOAP message.
Shared Patients	Patients for which data is stored by more than one health organization.
Stability	Traits that are likely to stay the same over time. An example of a stable trait is date of birth. An example of an unstable trait is an address.

Threshold	A threshold is the level at which records are automatically linked, rather than manually linked (manual linkage implying human intervention and disambiguation – sometimes referred to as potential linkage resolution)
True Negative	Refers to the correct classification by the matching algorithm of two patient records as a non-match when the two records refer to two different patients.
True Positive	Refers to the correct classification by the matching algorithm of two patient records as a match when both records refer to the same person.
Unique Patient Identifier	The value (usually a number) assigned to an individual for identification purposes that is unique across the entire national health care system. The United States congress currently does not permit the use of federal funds to research or promote the use of a national UPI.
Usable Period	The duration of time for which patient data is considered accurate and reliable. While data may be complete, it may be outdated. Therefore, patient information is often validated during every patient encounter.
Validity	Accuracy of patient data included in a patient record and matched to the correct individual.
Weight	Attributes or fields that are emphasized more than other data items when comparing patient records. A number (weight) is assigned to each field that reflects its relative importance compared to other data elements.
XCA	IHE Cross-Community Access is an international standard in Final Text status. XCA is focused on standards-based sharing of clinical documents.
XCPD	IHE Cross-Community Patient Discovery is an international standard in Final Text status. XCPD is a standards-based method of discovering mutually known patients between different communities.