A Practical Guide To Improving Patient Matching Accuracy

IATRIC systems

Table of Contents

Click on topic or page number to navigate.

Introduction

Hospitals and health systems need to make sure they deliver the right care to the right patient — otherwise the results can be disastrous. However, they often have hundreds of thousands of duplicate medical records in various clinical and accounting systems, making it hard to match the right patient with the right record.

Many healthcare organizations suspect they have a problem and want to take action before something really bad happens. Maybe they had a wake up call, such as a duplicate billing or a physician finding wildly inaccurate clinical information about a patient.

This eBook examines the challenges organizations face trying to identify patients, and steps they can take to prevent errors and provide a common patient record across the healthcare enterprise.

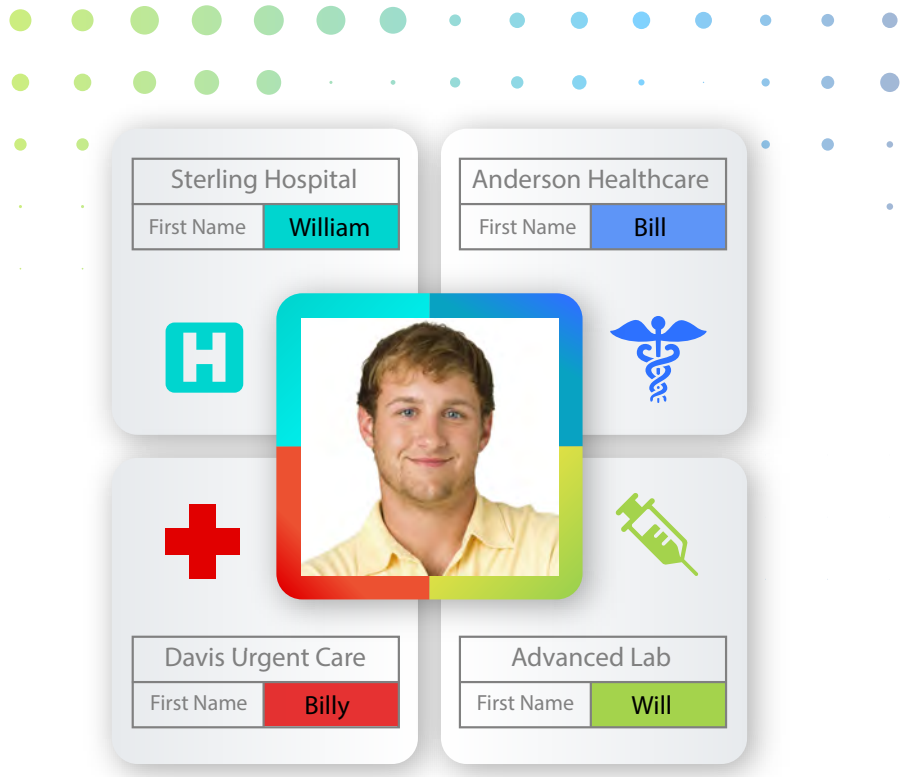
Why It's So Hard to Keep Track of Everybody

In an ideal world, each patient has a unique name that isn't shared with anyone else. For their care, they go to one hospital that has one EHR. They have one address where they receive their mail and there's never any confusion about who's who. Dream on.

In the real world, patients share their names with other people, or have names that are very similar. A single EHR will contain multiple instances of last names and first names as well as variations — John/Jonathan, Linda/Lynda, not to mention William/Will/Bill/Billy. Sometimes there's a middle name, sometimes a middle initial, sometimes it's left out. Multiple instances of the same or similar names are also found in the hospital's registration, billing, and clinical systems.

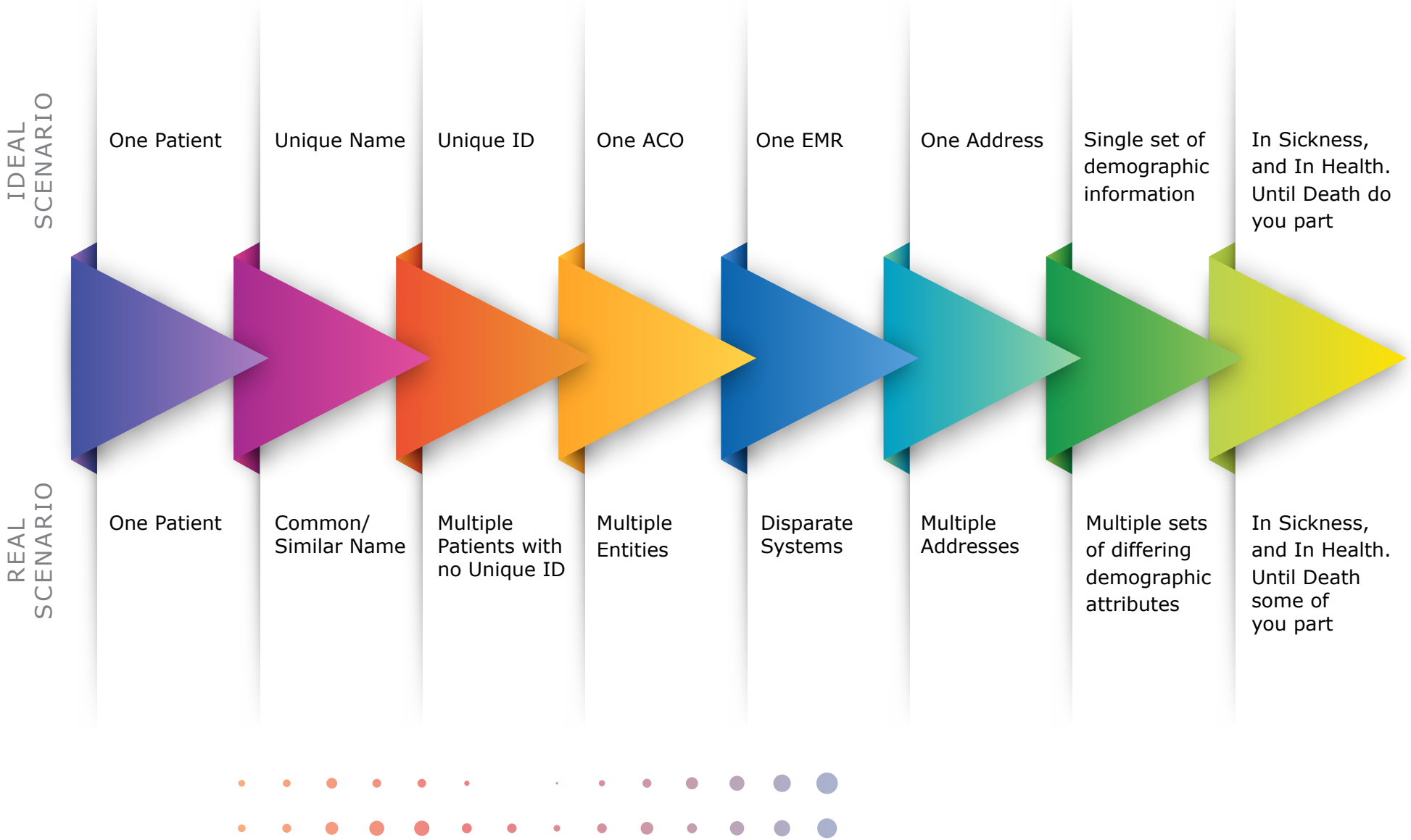
That's just one hospital. Patients go to multiple hospitals, labs, and other providers where they register and where new records are generated. They go on vacation, they move, they marry, they divorce. At various times their records are sent to HIEs, Integrated Delivery Networks (IDN), and other entities, where they may not be accurate or in sync. We also have to remember that patients omit details and provide inconsistent information, which can cause their identity to be linked to someone else and have serious consequences if not corrected in time.

Healthcare organizations have their hands full trying to keep up with all this activity. In order to properly care for patients and properly document their care, they need to share and combine demographic data correctly so there is one record for each patient. However this gets progressively harder as thousands or millions of patients come and go.



Fortunately, there are strategies and tools that healthcare organizations can use to greatly reduce patient identification errors. In addition to improving patient safety and quality of care, these strategies make life easier for the people who have to investigate errors and ensure records are accurate.

Ideal Scenario vs Real Scenario



Back to Table of Contents

EMPI 101 – Determining if Records Match

An Enterprise Master Patient Index, or EMPI, is the primary tool that healthcare organizations use to keep track of patient identity. EMPIs use specialized software to examine patient records across one or more data sources and apply algorithms and business rules to determine if the records match. Getting those algorithms right — tuned to the demographics of your patient population — is the name of the game.

First, some basic patient matching definitions:

Deterministic Matching: An exact match on a data attribute. For example, if two records share the same social security number they refer to the same patient, otherwise they identify different people or John to John would be a match, and John to Jon would not, even though those could be the same person.

Handling exceptions involves creating rules, such as comparing name or DoB in the event of a missing SSN.

Advantages

- Requires an exact match — there's no ambiguity
- Quick to implement and test
- Less risk of errors (initially)

Disadvantages

- Overwhelming cleanup efforts that are difficult or impossible to keep up with.
- Records with spelling variations, typos, abbreviations, etc. create mismatches that result in separate records.
- As multiple information sources start feeding the EMPI system, deterministic matching is outgrown very quickly.
- In a deterrministic world, there is usually no "list" to identify a "close" match. It just creates another record (or duplicate). Finding it is like a needle in the haystack, within a 1M record set.

Probabilistic Matching: A statistical approach that evaluates the probability that two records represent the same person. By assigning a score to each data element and adding scores to produce a final score, matches can be made with a degree of confidence if a predetermined threshold is met.

Business logic can also be applied after probabilistic matching in a deterministic manner to help with reducing remediation efforts. This can also help prevent false positives.

Advantages

- Far fewer duplicate records (when done correctly)
- More efficient matching

Disadvantages

- Algorithms require care and feeding, especially as you add new data sources with different demographics (more on this later)
- Poorly tuned algorithms may cause mismatches
- More expensive to implement — but still far cheaper than a mismatch that has a negative (or disastrous) impact on patient care

Deterministic vs. Probabilistic Matching

In the example below, first and last names are spelled slightly differently, streets and cities are spelled out in EMR1 and abbreviated in EMR2, and the state, zip code, and date of birth are identical. In a deterministic matching environment the match would fail, resulting in two separate medical record numbers.

	EMR 1	EMR 2	Deterministic Match? (In its purest form)	Probabilistic Match? (With algorithms in place)
First	JOHN	JON	No	Yes
Last	SMITH	SMYTHE	No	Yes
Middle	R	R	Yes	Yes
Address	110 SOUTH MAIN ST.	110 S. MAIN STREET	No	Yes
City	SOUTH WINDSOR	S. WINDSOR	No	Yes
State	CT	CT	Yes	Yes
Zip	06074	06074	Yes	Yes
DOB	19900808	19900808	Yes	Yes
WILL THE SYSTEM MAKE THE MATCH?			**NO**	**YES**

In a probabilistic matching environment, whether you get a match depends on the weight the system assigns to each data attribute. For example, let's review the following scores assigned to these data elements in this scenario above.

	Match Score	Partial Match/ Nickname Score	Completely different	Missing
First	8	5	0	0
Last	15	10	-4	-3
Middle	4	1	-2	0
Gender	4	0	-4	0
Address	12	8	0	0
City	6	4	-2	0
State	4	2	-5	0
Zip	6	0	0	0
DOB	8	4	-4	0
SSN	20	6	-10	0
Threshold for Matching			**45**	

Now let's add the scores using the elements from the example.

First name (partial match)	5
Last name (partial match)	10
Middle (match)	4
Address (partial match)	8
City (partial match)	4
State (match)	4
Zip (match)	6
DOB (match)	8
Total score:	**49**

Since the threshold for matching is 45, we have a match! But let's not get too excited. While we have a match, that doesn't mean it's accurate.

The system has linked the records based on the threshold set by the person tuning the data. A common misconception is that these records are now merged. That is not what a good EMPI should do. It should rely on the source system to provide a "merge" if they are from the same source (ie. duplicates).

Note: There are anonymous or bogus values that will receive a zero weight score. These are values that don't represent real information and are place-holders until accurate patient data can be used. ex. John Doe, trauma 1. These should eventually be overwritten with the correct value so they can be scored.

Types of Mismatches

- **False Positive** — Two records are thought to represent the same person, but they apply to different individuals.

The matching score is met — but the match is invalid.

- **False Negative** — Two records are thought to represent different people, but they both apply to the same person. The records are thought to relate to separate individuals.

The two records fail to match — but they should.

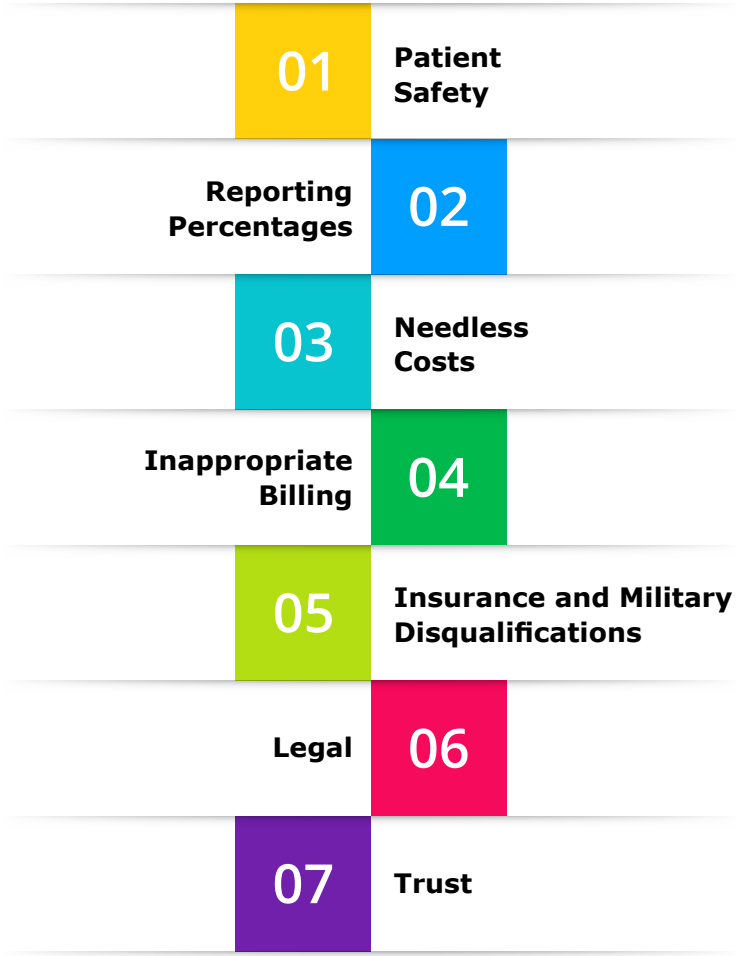
Obviously, healthcare organizations want to avoid both. That's why it's important to tune the algorithms — the data elements we choose to evaluate, the weights we apply to each, and the total score that determines if we have a match.

To be successful, you need to understand the community you're working with. For example, in some ethnic groups, many people share the same last name while in others many have the same first name. If the address is a large assisted living facility, hundreds of people will share that address.

THE POINT: if you apply too much weight to a data element without accounting for your demographics, the result could be a false positive; too little weight could produce a false negative. Obviously you want to minimize either possibility.

Mismatches Have Consequences: Why Patient Identity Is So Important

It's important to reduce patient identity errors for a number of reasons. There's a lot at stake, for patients, caregivers, and the organization as a whole — which is why education and spreading awareness needs to be a large part of your improvement efforts.



- **Patient safety**: Incomplete or erroneous clinical information about patients can have dire consequences.

  A classic example is when a patient is allergic to a certain medication but that information is associated with the wrong patient, who then is given a less safe or less effective medication. Meanwhile, the patient who really has the allergy receives the medication they are allergic to. Lawyers, start your engines.

  Another scenario is that the information on the allergy is spread across multiple records in the IDN that are not linked and aggregated in the clinical record. The doctor won't see all the available information.

- **Reporting Percentages (population health, Meaningful Use, Joint Commission, etc.)**: When patients have multiple Medical Record Numbers (MRNs) it increases the measurement denominator and decreases your payments or other benefits. (It could also mean penalties if requirements are not met.)

- **Needless costs**: Duplicate testing when orders are placed on a patient with multiple MRNs across systems.

- **Inappropriate billing**: A patient could be billed twice for the same encounter, or bills could be sent to the wrong patient. In either case, these snafus take time to sort out and can even drive away patients.

- **Insurance and military disqualifications**: If a patient is erroneously linked to a diagnosis or condition, it can be very difficult to undo — especially for a young patient unaccustomed to navigating these bureaucracies.

- **Legal**: Hospitals can be subject to steep litigation and compensation costs for inadvertently sharing patient PHI - not to mention a negative outcome due to a record mismatch.

- **Trust**: Patients, clinicians, and other stakeholders need to trust the data. If a patient accesses his or her portal and it shows they have a condition that they know they don't or any number of other mismatch scenarios, it will not end well for the organization responsible.

Causes of Incorrect Patient Matching

The Human Factor
— Patients and Healthcare Workers

- **Patients — Is Identity on Their Radar Screen?**
  Patients often don't understand the importance of providing accurate and consistent information. They're unpredictable — they may use nicknames, misspell street names, transpose letters and numbers, and omit small but critical details like Jr. and Sr. The potential for confusion is compounded when there's a language barrier, when people don't fully understand what you're asking for and may be too uncomfortable to ask.

  A different problem is card-sharing — using another person's health insurance card to avoid paying or because it's more convenient (at least for them). However this wreaks havoc for staff who have to untangle the mess and for clinicians trying to deliver care.

- **Registrars — Do They Know Their Value?**
  Registrars are on the front lines of ensuring accurate patient identity but they may not understand the importance of their role. They often see themselves as far down the hospital food chain. This perception is reinforced when there's lack of communication, lack of training, and lack of any real recognition for what they do.

Policies and Processes

To get accurate information, registration staff need clear policies about the demographic data that needs to be collected for each patient, the required format, and what clarifying questions to ask patients when they register. However, this information is not always clearly communicated to registration staff and others who need it. In addition, processes for remediating mismatched records may not be applied consistently across the organization, so there's no certainty that problems get fixed.

Technology

Lack of EMPI technology will increase your risk of errors, but so will technology that is poorly implemented, not used consistently, or not receiving the routine maintenance to keep it running at its best. If you don't have an EMPI, it should be something for you to take up with your leadership before something really bad happens. If you have an EMPI but you're getting matching errors (whether false positives or false negatives), it's time to consider analysis and remediation.

How to Improve Your EMPI Scoring

Fortunately there's a lot you can do to ensure the quality, accuracy, and completeness of your patient identity information. Improvement efforts need to address the key areas of people, processes, and technology, because they all work together.

The Human Factor

- **Patients Need to Know What's at Stake**

  Healthcare organizations need to educate patients so they understand why being identified accurately matters to them — how it helps you take better care of them, improve their outcomes, and their safety. Once they understand, most people make a sincere effort to comply.

- **Reach out to your patients with a campaign to spread awareness**

  It should be a family-wide effort since some people are poor historians or can't speak for themselves very well. It should also set goals for patients:

  – To be accurate, consistent, and complete in the information they provide

  – To check their patient portal or paper records for accuracy and update the portal when things change

  – To always carry ID

  – To share their information with a designated family member

  Follow up later to make sure patients get the message. If you have a large population of foreign language speakers, consider doing your outreach in those languages as well.

- **Registrars Need to Understand How Important They Are**

  The ability to identify the patient correctly is the foundation for quality care. Registrars have the most important role in patient identity and we need to educate them on their role in ensuring patient safety, on the rules and standards they need to follow, and potential red flags.

Examples for Educating Registrars

- Recognize when a patient is having trouble communicating and know what to do, including requesting assistance if necessary
- Call for an interpreter if there is a language barrier
- Recognize potential trouble spots, such as twins, Jr./Sr. and places where many people share the same address — assisted living centers, convents, extended care facilities, etc.
- If there's confusion or ambiguity, not to just "let it go," but address the issue proactively, or flag it for investigation and follow-up

Share use cases with them — for example, the number of hours a nurse or secretary spent to resolve a mismatch, lab work on the wrong account, all the gory details involved in trying to unmerge records. Remember that a real story is more meaningful to people than a dry lecture.

Also remember that people want to do a good job, and just setting expectations is often all that's needed. When problems occur, hold people accountable — not in a punitive way, but to educate them and show them where they went wrong so the same mistakes don't happen again.

Finally, we have to make sure that we reward and celebrate registrars when they do a good job. Make sure to celebrate their success as you see numbers increasing over time.

Policies and Processes

Hospitals should to take a close look at the policies and processes they use for ensuring patient identity and update them on an annual basis as needed. Here are key points to watch for:

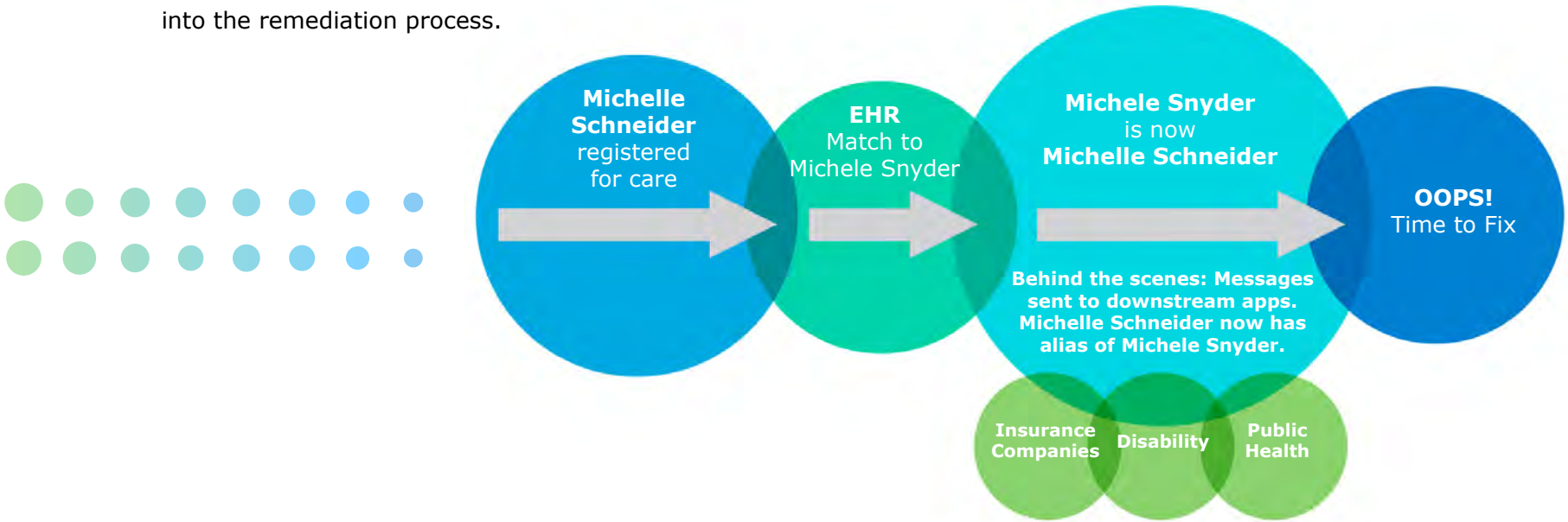
- **Make sure expectations are clear, communicated to all who need them, and readily accessible.** For example, you need clear rules about what information patients need to provide, the format it must have and what supporting documentation is required. The goal should be complete, accurate, and consistent data for identifying patients, anywhere it is collected. Consider a spot check — such as 20 records each week — to provide quantifiable stats to share with the team.

- **Have a documented process for remediating mismatched records.** Sorting out mismatches is usually ugly, and a very tedious task. It's important to have a process in place for people to follow. People also tend to just fix the record they're working on (especially when they're in a hurry), while historical records are overwritten with incorrect information. All this needs to be factored into the remediation process.

- **Consider the ripple effect.** Keep in mind that mistakes almost always flow downstream to other EMR systems, healthcare providers, or organizations. Remediation needs to include processes for fixing those downstream records — and the people who maintain the systems need to work to be kept in the loop.

  The ripple effect (see below): During registration, Michelle Schneider is mistakenly matched to Michele Snyder. Downstream, the two individuals share the same information — that of Michelle Schneider.

- **Who has the final say?** Remember that different people might be managing different systems. Policies and processes for data governance need to extend across the enterprise. Someone needs to "own" patient identity and be able to have the final say, ensure consistency, and make sure errors aren't ignored and perpetuated.



Back to Table of Contents

Technology

When implemented correctly, EMPI technology can make a dramatic difference in the quality, accuracy, and completeness of patient identity information. The technology uses analytics to detect patterns that would be impossible to detect manually, sets thresholds for linking records automatically, flags potential matching errors for remediation, and helps you fix them before they escalate.

EMPI analytics, if included with your system, have another important function: to measure and show improvements over time, so you can provide feedback, adjust processes as necessary, and celebrate success.

EMPI Technology Best Practices

As you investigate the right EMPI technology, the following best practices will help you use the technology to its full advantage:

- **Establish consistency across all data sources**
  Hospitals and health systems usually have a broad landscape of data sources feeding demographic data:
  - Acute care/Ambulatory EHRs
  - Interfaces (ADT, EMR, physician practices, pharmacies, payer organizations, etc.)
  - Flat files
  - Order messages
  - CCDs

  The EMPI solution needs to work across this complex environment, spanning the different technologies and data structures to provide a common patient record. It's also important to ensure data is normalized when it is collected, with complete, accurate, and consistent data across all data sources rather than having to sort it out later.

- **Fast reporting of issues and anomalies**
  The technology needs to provide fast and easy reporting capabilities for any patient identity issue that needs further investigation — especially when urgent attention is needed. Examples include:
  - When two people share the same SSN
  - Hospital re-admits or drug seekers at multiple locations
  - Significant changes that almost always occur because someone has overwritten a field by accident (first name, gender, date of birth, SSN, etc.)
  - Data poor records which don't have enough information to identify a person (ie..first and last name only)

- **Visibility and control across the enterprise**
  Your EMPI technology needs to provide HIM staff with the visibility to detect records that need to be reviewed, wherever those records are maintained in the healthcare enterprise. It also needs to provide incident management capabilities so they can take whatever remediation action is necessary — to evaluate, merge/unmerge, and report.

- **Personal device integration and biometrics**
  Patient identification is particularly challenging in critical/trauma situations, when the patient is incoherent/non-responsive and has no ID. Some EMPI providers can now integrate with biometrics technology such as retina recognition, palm recognition, and fingerprint recognition which can be critically important in a trauma situation.

Not all EMPI systems have these capabilities, but these are some of the important things you should consider in your search.

Start by Understanding Your System

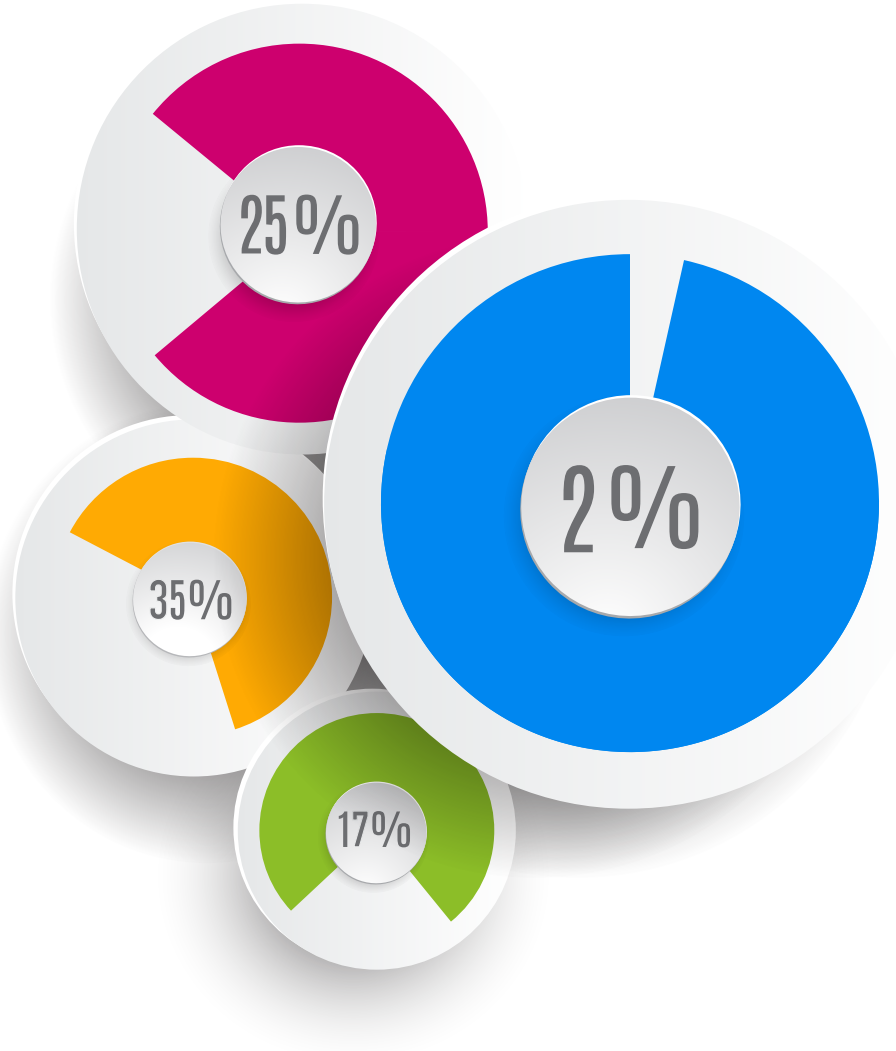
The first step in reducing patient identity errors is to know your current duplicate rate and where problems are occurring. The ONC suggests a duplicate rate of 2% or less, but many healthcare organizations have a much higher rate before they recognize they have a problem. (Rates in the teens are typical, but some hospitals have duplicate rates higher than 30%.)

You also need to know all the data sources that provide your patient identity information, and specific issues that are skewing the overall results.

For example:

- Are there thousands of generic IDs that have never been updated?
- Are they accounting for patient demographics?
- How much data is in a non-standard format, missing, or clearly wrong?

Knowing your current patient identification system will provide a baseline so you'll know where to focus your improvement efforts.



Get the Facts with an Assessment

To bring clarity to these issues, Iatric Systems offers a Patient Matching Quality Assessment. As part of the assessment, we analyze your patient identification data from each data source, and provide a report that you can share with your leadership that details duplicate rates, problem areas, and recommendations for improvement.

The assessment covers a wide range of patient identification issues in your environment including:

- **Your duplicate rate:** This is a compelling number, providing a snapshot of how dirty your data actually is.
- **Matched records and record sets:** For each data source, the number of total records, the number that need to be reviewed, the number of unique patients, and the number of groups of multiple medical record numbers (MRNs) for the same patient (representing records that can be retired to produce a clean data set).
- **Human errors:** Whether it's a particular data source, certain registrars, or specific issues

Example: Analysis of a sample data source, with 148 occurrences of four records representing the same patient. (Think of what that does to your public health and Meaningful Use reporting!)

Site Statistics	Record Count	Percent
Total Records	500,438	
Total Records in Groups	41,212	6.95%
Total Duplicate Groups = 23,188*		
Groups with 2 records	29,075	90.46%
Groups with 3 records	1,060	8.88%
Groups with 4 records	148	0.64%
Groups with >4 records	4	0.02%
** Largest group contains 7 records	4	

- **Field level analysis (e.g. mother's maiden name, previous address):** How good are you at getting those details right? What's the percentage that are missing? (This is an interesting factoid that often surprises people.)
- **Generic IDs (i.e. baby girl/boy, SSN, DOB):** Formats for generic IDs are often inconsistent. For example, how does your organization indicate an unknown social security number? All nines? All zeros? 123456789? Or whatever comes to mind? The analysis uncovers the formats you use for generic IDs and the number of occurrences of each.

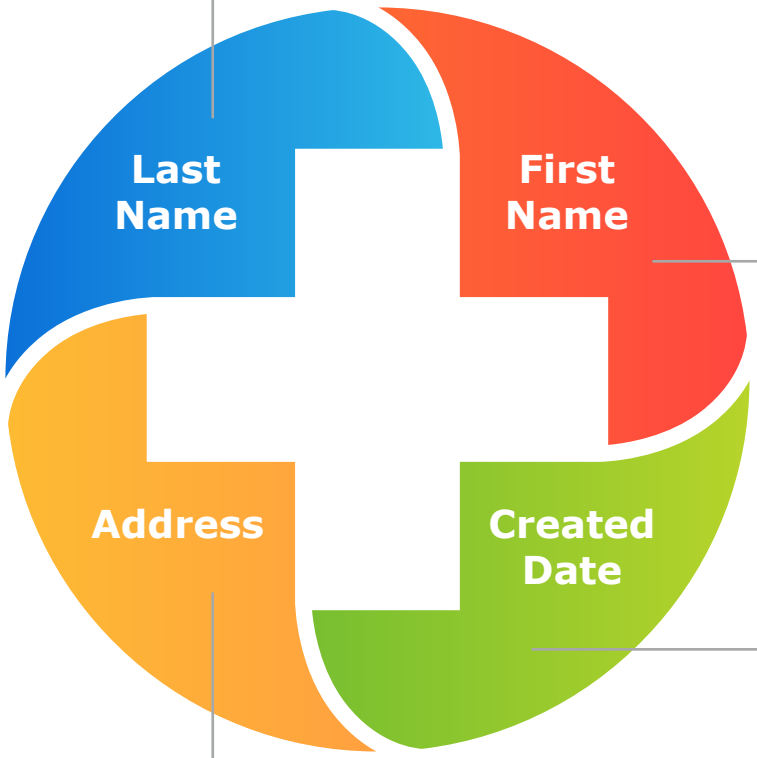
Generic Names	Record Count	Generic SSN	Record Count	Generic DOB	Record Count
BABY GIRL	1126	999-99-9999	465636	01/01/1901	202
BABY BOY	944	000-00-0000	1022	2001/01/01	33
MALE	312	0111111111	80	01/01/1950	48
BABY	222	0999999999	69	01/01/2001	60
GIRL	99	00	56	1950/01/01	54
BOY	150	111-11-1111	72	01/01/1900	222
BABY GIRL	1003	099-99-9999	4	09/09/1999	13
FEMALE	71	000-99-9999	1	01/01/1880	23

- **Data Poor Records:** Iatric Systems uses an algorithm to see if records contain enough data to be considered for analysis, or if too many fields are missing. (These are incomplete and should not be considered for matching.) This is also useful for staff education — you can show how scores improve when fields are completed properly.
- **Geographic/demographic issues:** For example, how well does your EMPI account for naming conventions of ethnic groups in your population? Does it account for people with the same address — assisted living facilities, convents, rehab centers, nursing homes, etc.? Evaluating these details will help eliminate false positives and result in more successful algorithms.

Sample "Top" name and address lists from a full extract

(Some of the details have been changed to protect privacy)

The assessment results will help your hospital make informed decisions about what changes are needed to reduce your duplicate rate, whether reeducating staff, normalizing your data, or investing in new EMPI technology. Iatric Systems offers various services for remediation; we also partner with leading EMPI vendors to help healthcare organizations meet their patient identity challenges.



Last Name	Count
MARTINEZ	4110
JOHNSON	990
YANG	1802
SMITH	703
PATEL	737

First Name	Count
MICHAEL	3800
JOSHUA	3541
MICHELLE	3058
JOHN	2998
MARIA	2954

Created Date	Count
19890901	854
19900124	197
19900531	190
19900228	155
19900309	136

Address	Count
2005 MAIN ST MANCHESTER MANOR MANCHESTER RI 06220	1077
1235 SOUTH ST SHADY LANE HOME VERNON RI 06220	582
2409 S ALVERNO RD MOTHER MARY CONVENT COLCHESTER MA 02415	495
1235 S 24 ST WILSON REHAB CENTER EAST HAMPTON NY 02587	460
960 S RAPIDS RD RAPID RECOVERY HEALTH AND REHAB HEBRON MA 98903	312

Keeping Your EMPI Running at Its Best

Once you have your EMPI system in place, it's important not to become complacent. The data feeding your EMPI is constantly evolving as new systems are integrated, existing systems are upgraded, patients move, get married, etc. Thus, ensuring accurate patient identity is an ongoing process:

- **Continuing education** to remind patients and registration staff about why accurate patient identity is important, practices to follow, pitfalls to avoid, how important patient information is to the facility and how deeply it is protected (PHI)
- **Reaching out** to patients from time to time to confirm they are who you think they are
- **Periodically revisiting your patient identity policies** and processes to make sure they are working as intended, and revising as necessary
- **Researching, testing, and tweaking your algorithms** to adjust the weight given to each search term as patient demographics evolve
- **Managing remediation** — reviewing records to determine if they represent the same person or not, merging/unmerging patient records
- **Building, modifying, and testing interfaces** between systems containing patient information to ensure the information is consistent, complete, and in the right format
- **Staying on top of new regulations**, organizational requirements, and technology associated with patient identity, and adapting as needed

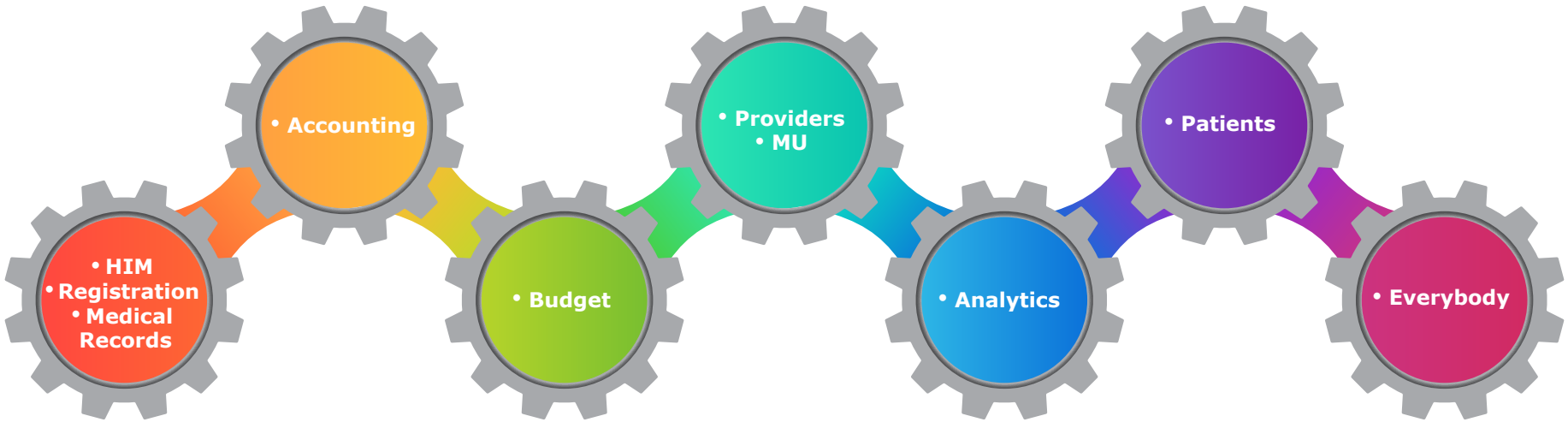
Obviously these tasks require a certain amount of time and expertise. Many hospitals find that complementing their internal staff with the resources of a knowledgeable service provider that can step in when needed is a very attractive option.

Back to Table of Contents

The Benefits of Accurate Patient Identification

Earlier we discussed the negative consequences of patient identity errors, including the risk of improper tests, incorrect diagnoses, and inappropriate treatment. A good way to conclude this eBook is by examining the many benefits when patient identification is done right, and how your improvement efforts pay off. (Stakeholder categories are in bold.)

- Eliminates mismatches and duplicate charts, and the headaches and hours spent investigating and correcting errors. **HIM/Registration/Medical Records**
- Ensures accurate billing using correct insurance information. **Accounting**
- Provides better results for your public health and incentive payment programs by reducing the number of unique patients down to those that are truly valid. **Budget**
- Provides a true picture of patients under a provider's care, or multiple providers at multiple sites. **Providers**
- Reduces costs of unnecessary tests. **Budget**
- Aggregates data from multiple information sources under one entity for an HIE or integrated network. **Providers/MU**
- Reduces overhead due to the lower number of unique patient lives that you need to maintain and monitor. **Budget**
- Reduces the cost of doing business. For example, if EMPI remediation can reduce the number of unique patient lives from one million to 800,000, at 3 cents per record, it adds up. **Budget**
- Improves accuracy of analytics because the patient count is correct. **HIM/Analytics**
- Reduces fines, penalties, and litigation costs. **Budget**
- Helps build an atmosphere of trust. When patients see results in their portal, it's with confidence that the record is complete and all the information pertains to them. **Patients**
- Improves patient safety and quality of care. **EVERYBODY**



Common EMPI Terminology

Algorithm — A complex set of steps to create an outcome. The Fellegi-Sunter model is the base algorithm that probabilistic algorithms employ, with vendors adding additional parsing routines, data attribute availability, etc. to enhance.

Attribute — A variable that identifies a type of information to determine likeness on same values (First Name, Last Name, DOB, zip code, etc.)

Enterprise Master Patient Index (EMPI) — An indexed system of patient records across an enterprise of participating data sources that can represent a single entity of a patient record based on patient demographics.

Entity — Multiple sources of demographic information tied together with a single EMPI number thus creating an entity (single best record, golden record, etc.)

Link or **Match** — Creating an association between two or more records across the same or multiple sources of information.

Merge — Creating a "surviving" record and an "obsolete" record from multiple records containing multiple charts in a single registration system. Not usually performed within an HIE.

Scoring — Adding of multiple attribute scores to give a total value.

Task — An item that may require remediation. Can be a duplicate (two records with different MRN numbers from the same source of information) or overlays (a single MRN that may have been overwritten by someone else's information). Can also include potential links and potential duplicates.

Threshold — A set score that, when achieved, enables a set of records to be linked automatically (auto match) or a task created for manual review (potential match).

Weighting — A numeric value given to specific attributes that, when combined with other attributes, creates an overall score.

Request your Patient Matching Quality Assessment to learn your duplicate rate and get reports detailing your duplicate results.

To stay up to date with patient identity and other interoperability topics, read or subscribe to our blog.

IATRIC systems

Ph: 978-805-4100 Fax: 978-805-4101 info@iatric.com iatric.com