Moving Toward Zero: How to Solve the United States Healthcare Person Matching Challenge

How to make healthcare person identification work

October 17, 2013

Abstract: This document describes how to augment existing EMPI demographic matching processes in order to achieve person identification error rates that approach zero.

Contents

Executive Summary

This white paper is a response to the recent decision by the ONC to launch an investigation into the performance of demographic matching for patient identification. We contend that error levels in patient matching must be very close to zero in order to enable trusted exchange of health information. We propose that existing EMPI capabilities must be augmented with a single patient-controlled data element that guarantees uniqueness. Only then can the potential value of healthcare information exchange be realized with confidence.

The only "solution" to the patient matching challenge is to implement techniques that can move us toward an identification error rate of zero

Intended audience

This white paper is intended for anyone interested in understanding the strengths and weaknesses of the current EMPI patient matching approach and how it can be improved. It also should be of interest to anyone who needs medical care (and that is all of us) because inaccurate person matching places all of us at risk when errors occur.

Defining the need

On September 11, 2013 the Office of the National Coordinator for healthcare technology announced that it is sponsoring an effort to evaluate patient matching[1]. This announcement was made in recognition of the critical nature of accurate person identification as a prerequisite to healthcare information interoperability. If two healthcare entities cannot accurately identify the same person then they cannot safely exchange information about that individual and they cannot safely provide service to that person. Indeed, if the two agencies are erroneously dealing with two *different* individuals, believing them to be the same person, then any services they offer are actually likely to cause harm for both individuals.

This white paper is entitled "Moving Toward Zero" because it contends that the only "solution" to the patient matching challenge is to implement techniques that can move us toward an identification error rate of zero. Current identification techniques based on enterprise master patient index (EMPI) systems, also known as master data management (MDM) systems, do a good job of patient matching. They use very sophisticated demographic matching algorithms to determine the probability that a given set of demographic data represents one of the persons in their existing database. However, despite years of tuning these algorithms and refining the procedures that support them, it has proven to be impossible to eliminate matching errors. There is ongoing discussion about the actual error level but persistent industry reports describe error rates ranging from 5% to 10% and often considerably more.

The Problem

EMPI matching errors

EMPI demographic matching leads to two distinct categories of error.

[1] ONC Launches Patient Matching Initiative, Sept. 11, 2013, http://www.healthit.gov/buzz-blog/health-innovation/onc-launches-patient-matching-initiative/.

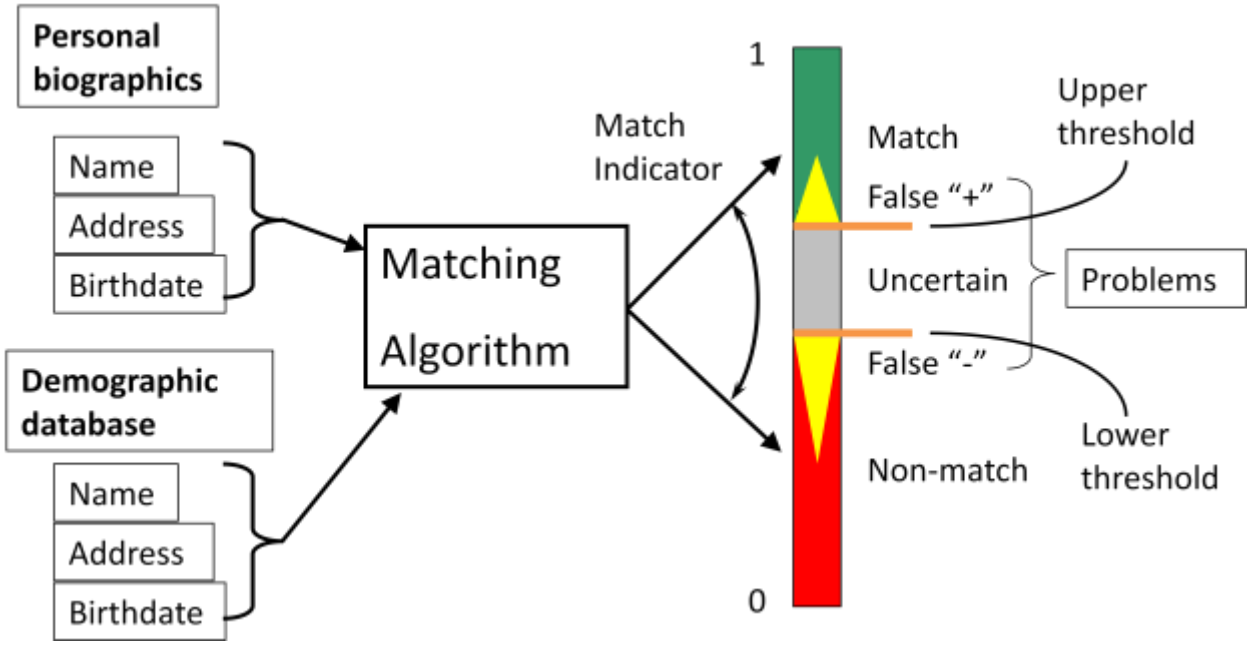
A false positive match occurs when the matching algorithm determines that two data sets represent the same person but in reality they actually are different people who happen to have a similar set of demographic information.

A false negative match occurs when the matching algorithm determines that two data sets represent different individuals but they actually both contain information concerning the same person.

Both false positive and false negative matches cause problems. When a false positive match occurs a person may be treated (incorrectly) using information that applies to someone else. This is clearly a very dangerous situation. When a false negative match occurs, information needed to guide patient therapy may not be available because that information is "lost" in the form of a fragmented medical record.

In addition to these two sources of error, most matching algorithms have the potential to report an "indeterminate" match where the algorithm declares that it cannot make a definitive decision. This requires human intervention and, due to the ambiguous nature of the matching data, represents a substantial additional potential source of both false positive and false negative matches.

Figure 1: EMPI Matching Errors



Each EMPI client needs to establish upper and lower match thresholds as represented by the two orange lines. The sources of error in this diagram include the two yellow triangles and the gray rectangle which represents uncertain results that need to be processed by human operators.

What is the incidence of false positive and false negative matches? This is a critical question but unfortunately it is very difficult to answer. The incidence of matching errors depends on a number of factors but there are four main influences:

1. How sophisticated is the matching algorithm being used by the EMPI?
2. What are the matching threshold settings being used?
3. What is the nature of the person population represented in that EMPI's database?

4. What is the overall quality of the set of person demographic information presented to the EMPI for matching?

It is mathematically not possible to reduce the overall matching error rate beyond a certain level by 'tuning'.

Figure 1 shows a diagram of how EMPI systems make demographic match decisions. The result of a matching episode is reported as a value which is shown in this example as ranging between 0 and 1. The EMPI system allows its owner to establish two thresholds for matching. If the matching indicator value is less than the lower orange threshold then the EMPI reports "no match". If the matching indicator value is greater than the upper orange threshold then the EMPI reports a "match". If the matching value lies between the two thresholds then the EMPI reports an "indeterminate" match. Indeterminate matches require further analysis, typically involving a human operator trying to further clarify the situation.

Because the potential negative consequences of a false positive match are so serious, most organizations set their upper match threshold to a high value in order to minimize the frequency of false positive matches. In order to minimize the frequency of false negative matches, it is also desirable to set the lower match threshold to a low value. This however has unacceptable consequences because it means that the EMPI reports a large proportion of indeterminate matches. These require time and human processing; resources that are often not readily available in the client organization. In addition, because of the ambiguous nature of indeterminate matches, this human processing can lead to a significant additional set of person identification errors.

It is of course possible to minimize or entirely eliminate the incidence of indeterminate matches by moving the lower match threshold to a value near the higher match threshold. In fact, some organizations set both thresholds to the same value in order to avoid indeterminate matches completely. However, doing this causes the incidence of false negative matches to skyrocket and this in turn leads to an unacceptable rate of fragmented medical information.

No matter how much work is done to tune the upper and lower match threshold values it is mathematically not possible to reduce the overall matching error rate beyond a certain level by 'tuning' the match thresholds. Careful work to standardize demographic inputs and apply more sophisticated matching algorithms can offer some incremental help but neither of these approaches fully addresses the problem of matching errors.

Trends that exacerbate the problem

In today's evolving healthcare environment there are two fundamental trends that lead people to conclude that unassisted EMPI demographic matching cannot meet the industry's needs for accurate person identification.

Problems with the data

The first trend is that EMPIs are being asked to perform demographic matching across an increasing number of independent healthcare "silos". Many of the demographic elements generated by these silos represent text fields (such as names and addresses) that cannot be automatically corrected. A last name spelled Smith, Smithe, Smyth or Smit might be the person's actual name or it might be a typographical error. There is no automated way to know and there is no automated way to correct such a mistake.

Names with double letters, hyphenated names, foreign names and names with non-standard formats all represent other problematic situations which increase the rate of matching errors.

Population size

The second trend involves increasing population size. EMPIs are being asked to perform demographic matching across larger and larger populations. As more people reside in an EMPI database there are more instances of multiple people with similar demographic profiles. This increase of similar demographics increases the probability of incorrect EMPI demographic matching.

When an EMPI system is searching for a person against a small patient population such as the several thousand persons who might be patients of a specific physician there is a low likelihood of two persons having significantly similar demographic information. Consequently there should be a very low rate of errors, especially of false positive errors.

It is a different situation, however, when the population grows to significant size. As EMPI matching is extended from single provider organizations to related ancillary providers such as lab, pharmacy and radiology; to clusters of provider organizations; or to regional or statewide health information exchanges, the patient population to be matched can increase significantly. These increasing populations represent a significant challenge for demographic matching.

Consider Table 1 which illustrates the experience of a health information exchange located in Houston.

Table 1: Population Demographics[2]

Harris County, Texas

- 12 years of data
- 3.4 million patients in hospital district's database
- 249,213 patients have the same first and last name
- 76,354 patients share both names with 4 others
- 69,807 pairs share both names and a birth date with each other
- 2,488 patients named Maria Garcia
- 231 "Maria Garcia"s have the same birth date

This illustrates how increasing the size of patient populations is a challenge for identifying patients using demographic matching regardless of the quality of the input data or the sophistication of the matching algorithm. Clearly, any attempt at identifying Maria Garcia using demographic matching against the Harris County population will produce unsatisfactory error rates and require exorbitant amounts of manual analysis.

It is worth noting that Harris County represents roughly 1% of the U.S. population. As projects such as Healtheway unfold with a goal of serving the entire United States population it is clear that unassisted EMPI demographic matching will not be able to provide satisfactory performance.

Table 2 demonstrates potential daily frequencies of EMPI matching and resulting error rates when applying 'optimum' EMPI matching error rates of 0.001 for false positives (99.9% accuracy) and 0.05 for

[2] Source: Houston Chronicle, 4/5/11

false negatives (95% accuracy)[3] as defined by the Patient Matching Power Team of the HIT Standards Committee.

Table 2: Error Rates

Daily number of EMPI matches	Number of false positives @ 0.001 error rate	Number of false negatives @ 0.05 error rate
10,000	10	500
100,000	100	5,000
1,000,000	1,000	50,000
10,000,000	10,000	500,000

It is worth noting that almost no real-world EMPI systems have been able to operate at the 0.001 and 0.05 accuracy levels established by the Patient Matching Power Team so that actual performance will likely create significantly more errors than indicated in Table 2. However, even the indicated error rates will not be sustainable.

Finally, we should also expect a more frequent need for EMPI matching operations as the population grows larger because there will be a larger set of healthcare organizations trying to exchange information. The result is that unaided EMPI matching cannot deliver adequate accuracy to support automated interoperability of clinical information across large populations.

The Impact of Matching Errors

When using an EMPI system for person identification to support the exchange of clinical information across independent healthcare silos, one error is one error too many. Depending on the specific situation, the consequence of an incorrect identification match can include:

- time lost looking for missing information
- the need to perform duplicate tests
- poor medical outcomes because treating physicians do not have complete information
- serious inadvertent harm to the patient
- avoidable medical treatment errors
- and even patient fatalities

Not only are these consequences unacceptable because of their impact on patient care, they also constitute a major driver of increased medical costs. Healthcare provider organizations can ill afford the inefficiencies, poor medical outcomes, patient and provider dissatisfaction, malpractice risk, elevated insurance premiums, and reputational damage that can result from identification errors.

The ONC-sponsored effort to continue to limit the rate of false positive and false negative patient matching errors by continuing to improve demographic matching is laudable. However, it is certain that no matter how many resources are put into this effort, it cannot lead to the necessary goal – a patient identification system with virtually zero errors. Instead, what is required is to augment the existing

[3] In August, 2011 the Patient Matching Power Team Federal Advisory Committee formed by the HIT Standards Committee indicated that a matching specificity of 99.9% and sensitivity of 95% "are in the range that will eventually be recommended."

EMPI probabilistic demographic person identification system with a fundamentally different approach that can actually achieve this goal.

The Proposed Solution

Since one error is "one too many" in patient identification systems, any proposed EMPI matching augmentation must make error-free matching possible[4]. The straightforward way to accomplish this is to equip each person with a unique individual identifier[5]. This identifier becomes the single patient-controlled data element that guarantees uniqueness. It is important to note that this data element differs from other identifiers, like the Social Security Number, in that it is used solely for the purpose of patient identification in healthcare, it has no other uses.

A unique identifier is created and is issued to the person in the form of some token which enables automated reading of the embedded ID – a 2-D barcode, a magnetic stripe, a smart card, etc. When the unique identifier is issued to the person a copy of that identifier is included in the demographic information for that person stored in the EMPI. The person can present their ID token at each subsequent medical encounter. The identifier is read by an automatic device and forwarded to the EMPI. The EMPI does a mathematically precise search for that identifier and locates the corresponding patient information. Because each identifier is globally unique and is not shared with any other person there is no potential for error. "Zero errors" thus becomes an achievable goal.

Role of the EMPI

Note that an EMPI will always be an essential part of the person identification solution. The idea is not to replace existing EMPI technologies but rather to enhance them so that they can accomplish their task. The addition of a personal identifier to the EMPI's patient record can eliminate the need to do demographic matching for specific encounters. However, demographic matching will always be needed as a fallback for situations such as linking to historical data and in situations where a person's ID is either not assigned or not available.

Furthermore the EMPI is still required for its other functions.

- It is the repository for the authoritative personal demographics profile (including the unique identifier) for each person.
- It may maintain electronic interfaces to the various ATD systems in use by the HIE.
- It is the repository of cross reference information between the various institution-specific identifiers used by its subscribing care provider organizations.
- It is the interface to the system that distributes and manages the unique identifiers.
- It often plays a critical role in the record locator services provided by an HIE.

In the rare event of an encounter where a person does not have their identification token with them the EMPI may be required to do a demographic match in order to locate that person's record. This of course is not desirable because of the possibility of a false positive or false negative match but if the

[4] It is, of course, always possible for a human to cause errors, for example through intentional malfeasance. However, the solution described in this white paper can indeed lead to virtually error free operation.

[5] We are aware of the congressional prohibition against unique individual identifiers. The language of the prohibition prevents the use of *federal* resources to implement an individual identifier standard but it in no way precludes a private enterprise approach for the creation of such a system.

patient's unique identifier truly is not available there may be no other alternative for that particular encounter.

Requirements for a Solution

Any solution that has the potential to effect a qualitative, not just incremental, improvement in traditional EMPI demographic matching must conform to a number of other requirements. This list summarizes some of those requirements.

- Unique – It must be possible to guarantee that the added person-specific demographic element is unique, can never be confused with any other issued by the system, and is never re-used.
- Effective – It must have the ability to eliminate virtually all identification errors.
- Permanent – The added person-specific demographic element must remain with the person to whom it is assigned for life. An exception will be if the identifier is used fraudulently or is compromised in some way.
- Cost effective – Any proposed solution must be inexpensive to implement.
- Use existing infrastructure – It must make as much use as possible of existing systems such as EMPIs and record locator services.
- Patient empowerment – The solution must enable individuals to play an active role in managing their own identity.
- Error correction – With any system, errors will occur. There must be efficient and effective processes to correct errors once they are identified.
- Standards based – A proposed unique identification system must be based on accepted national and international standards.
- Private enterprise – A private enterprise solution is necessary to honor the current congressional prohibition against federal involvement in a unique individual healthcare identifier.
- Simple – It must be simple for patients, providers and provider organizations to understand.
- Trust – It must build trust with all participants.
- Universal – Any proposed solution must be available to any person who needs it
- Healthcare focus – In order to achieve operational simplicity the solution must be dedicated to healthcare purposes only.
- Scalable – The solution must be able to scale to serve a population of any size.
- Non-profit – Using a non-profit corporation to manage the solution ensures that it remains focused on its core purpose and offers services at the minimum cost feasible
- Privacy enabling – The solution must be implemented in such a way that it enhances patient privacy and threatens it in no way. It must not represent any HIPAA privacy liability to those who use it.
- Counterfeit resistance – The added patient specific demographic element used by the system must not be susceptible to data entry typographical errors or intentional attempts to create counterfeit identifiers.
- Phased deployment – It must be possible to implement the solution in a progressive manner in order to minimize disruptions to processes that are already in place.
- Identity theft mitigation – The solution must minimize the opportunity for identity theft and make identity theft remediation simple, rapid and cost effective.

- Anonymization – Since anonymous data sets will play an increasing role in exchange of sensitive clinical information the solution must fully support the creation and use of anonymous data sets.
- Language independent – The design and operation of the solution must be independent of the language spoken by the person.
- Efficient registration – The solution must enable substantial improvements in the efficiency of patient registration for a clinical encounter to the benefit of the patient, the provider and the provider organization.
- Voluntary (optional) – Because an individual identifier is a potentially sensitive item from a number of perspectives each individual should ideally be given a choice as to if and when to participate.
- Other processes – A properly implemented patient identification capability should be able to facilitate other related healthcare processes.
    - Fraud mitigation
    - Recovery from data breaches
    - Public reporting
    - Education
    - Epidemiology
    - Research
    - Billing

Implementing the solution

Global Patient Identifiers Incorporated (GPII) Accurate Identity Matching (AIM) is a solution for all healthcare stakeholders to choose as a way to achieve accurate identification and information sharing across healthcare organizations. AIM, which is based on technology available today, allows organizations to use GPII numbering services behind the scenes to link patient records more accurately with existing EMPI technologies. That is, a unique data element is added into the EMPI to ensure an accurate match of a patient with his or her medical records.

The system of patient identification established by GPII meets all of the solution requirements listed above. The system represents the implementation of two ANSI/ASTM International standards.[6] It is exceptionally cost-effective and has the potential to offer participants the opportunity to implement patient mediated privacy management of their own clinical information. The core system services have been available for more than four years and have been used in test mode to issue more than 1,000,000,000 identifiers.

It is important to note that the proposed solution will never eliminate the need for an EMPI. However, the addition of a unique patient specific demographic element will lead to a decline in the incidence of matching errors. Furthermore, this solution can be initially put in place with minimal modifications to the core EMPI system and the patient registration facilities of the participating care delivery organizations. The solution can be implemented in a phased manner as other components of the healthcare ecosystem decide to realize the benefits offered by the system.

There are two options for deploying this solution that can be considered:

[6] ASTM international standards E 1714 and E 2553 are both available at www.astm.org.

- The first is to assign an identifier to a person only when he or she specifically requests one. This "voluntary" option gives each person the ability to choose whether and when to participate.
- The second, "mandatory" option is to assign an identifier to all persons in a specific population.

Both voluntary and mandatory deployments can be chosen by different healthcare organizations and the resulting systems can coexist side by side. The choice of a voluntary or mandatory deployment can be made by the client based on their operational constraints and policy wishes as well as the desires of the patient population that they serve. In either case, individuals retain control over the use of their identifier.

This flexibility, combined with the potential to implement additional capabilities in such areas as privacy and fraud remediation, make possible a major advance in the patient empowerment that can be offered to persons participating in this solution. These persons are able to play an active role in ensuring their accurate identification across all healthcare encounters. They also can play the deciding role in implementing the privacy management of their medical records[7].

As a private enterprise vendor, GPII is able to offer this solution despite the congressional prohibition on involvement in a unique healthcare identifier. As a not-for-profit corporation, GPII maintains a focus to provide unique patient identification at the minimum cost.

Conclusion

Whenever humans are involved in an activity there is always the possibility that errors may occur. We must acknowledge the possibility that even if every patient is equipped with a unique demographic data element, rare incidents of person misidentification may occur. However the rate of these errors will be orders of magnitude less than the current demographic matching error rate. Furthermore, when such errors are detected they will be easily correctable.

For example, if a patient identifier is discovered to have been used in an episode of fraud (such as identity theft); it is possible to deactivate that identifier at "electronic speed" to prevent its use in any further fraudulent activity. This helps ensure that the damage caused by an episode of intentional malfeasance can be kept to a minimum. Furthermore, the patient can then be issued a new replacement identifier that can be used to restore integrity to his or her information so that they can receive appropriate care going forward.

While there are still significant operational issues to be resolved, it is clear that a unique individual demographic data element used to augment the operation of an existing EMPI represents a qualitative improvement in the effort to achieve a zero incidence of person identification errors. The approach scales well to enormous person populations. It is simple to use, extremely cost effective, and makes available other very important healthcare functions[8].

As healthcare attempts to move interoperability to the next level by implementing Accountable Care Organizations, large Health Information Exchanges, and a national clinical information exchange network; accurate person identification is essential. It may be possible to achieve some small incremental improvements in the accuracy of existing demographic matching approaches but it is clear that, without some sort of qualitatively different augmentation, existing EMPI systems cannot meet the emerging identification needs of healthcare. The solution described here offers a straightforward,

[7] A Privacy Strategy for the Unites States Healthcare Industry, http://gpii.info/pub/GPII-Privacy_Whitepaper-May_2013.pdf.
[8] Ibid.

simple, and cost-effective method to enhance these EMPI systems in a way that can get us to "zero" – no errors in patient identification.

Bibliography

- "Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System", Richard Hillestad, et. al., Oct. 2008, RAND Corporation Monograph # 753, http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf
- Patient Identity Integrity, HIMSS Patient Identity Integrity Work Group, December, 2009, http://www.himss.org/ResourceLibrary/ResourceDetail.aspx?ItemNumber=6968
- The growing Threat of Medical Identity Fraud: A Call to Action, Medical Identity Fraud Alliance, July, 2013, http://medidfraud.org/wp-content/uploads/2013/07/MIFA-Growing-Threat-07232013.pdf
- ONC Launches Patient Matching Initiative, Sept. 11, 2013, http://www.healthit.gov/buzz-blog/health-innovation/onc-launches-patient-matching-initiative/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+healthitbuzzblog+%28Health+IT+Buzz+Blog%29

Barry Hieb, M.D.
Chief Scientist
520.320.6220
520-342-8457 (mobile)

Rob Macmillan, CEO
520-449-9840 (mobile)

Global Patient Identifiers, Inc. at www.gpii.info
Sponsor of the Voluntary Universal Healthcare Identifier Project