

# Managing the Integrity of Patient Identity in Health Information Exchange (2014 update)

Save to myBoK

Accurate patient identification is foundational to the successful linking of patient records within care delivery sites and across the healthcare ecosystem to underpin care delivery, data exchange, analytics, and critical business and clinical processes. These goals have increased in importance as health information exchange has evolved over the last decade with the healthcare industry striving to reduce costs, increase interoperability, and transform to a patient-centric care delivery model.

Strong information governance that addresses patient identity integrity and accurate patient matching is key to a patient-centric health system and patient-centric processes. This Practice Brief explores the complexity of patient identification integrity, including how organizations can manage patient identification systems from front end data capture to back end quality control as an ongoing process and carry local quality operations into health information exchange efforts. It urges industry stakeholders to recognize that now is a critical time to address accuracy in patient identification systems.

## Patient Identity Integrity Vital to Healthcare

Various components of the healthcare ecosystem will address these goals and execute patient identification integrity activities to:

- Support care delivery and care coordination within an enterprise, as well as data exchange across healthcare systems.
- Underpin analytics within and across organizations, including pattern recognition, Big Data, and predictive analytics.
- Support information governance strategy and practices. This governance must underscore the complexity of patient identity integrity, including people, processes, and technologies. There is no single solution.
- Ensure HIPAA compliance when managing and releasing individually identifiable health information (IIHI).
- Support compliance with various patient matching guidelines and/or standards from accreditation and standard development organizations (SDOs), as well as governmental agencies or departments.
- Calculate “unique patient” records for meaningful use reporting.<sup>1</sup>
- Support patient-centered medical home initiatives and accountable care organizations.

Patient identification integrity is a complex concept, and one that is not well understood throughout the healthcare industry. The complexity stems from many factors including variability in practices of authentication, data collection, technology, and the historical silo approach to patient identification. Previously, patient identity integrity was seen as a health information management (HIM) or registration/patient access function, with limited staffing devoted to the issue and a site-specific approach. Today’s health transformation initiatives push this function to the front of the health information exchange effort.

Incorrect or incomplete data capture within the healthcare setting can create critical patient care issues and risk privacy breaches, thus degrading consumer and user trust. Health information organizations (HIOs) support, oversee, or govern the exchange of health-related information among organizations according to nationally recognized standards. HIOs are the recipients of the stewardship and governance applied to patient identification processes, thus HIOs are today highlighting many of the weaknesses in the historical systems and practices. As data exchange methods through Direct messaging, private exchange, or state HIOs continue to evolve, patient identification errors will increase significantly.

Policymakers and industry leaders are beginning to recognize the importance of patient identification, as exhibited by the early 2011 patient matching recommendations from the Office of the National Coordinator for Health IT (ONC) Health IT Policy Committee following a public hearing on patient matching hosted by the Privacy and Security Tiger Team in late 2010.<sup>2</sup>

More recent ONC activities such as an environmental scan of vendors, providers, and data exchange organizations are exploring current approaches, practices, and processes related to patient identification. The output from this activity will be presented to the Health IT Policy and Health IT Standards Committees, and could potentially influence requirements for stage 3 of the “meaningful use” EHR Incentive Program.<sup>3</sup>

## Ensuring Patient Identity Integrity

Establishing and maintaining patient identity integrity is a complex issue with challenges that encompass standards, technologies, and business practices. The scope of the challenges includes factors such as:

- Proof of identification is not routinely required at the time of data capture and lack of accountability in validating patient identity compounds duplicate and overlay creation.
- Accurate registration is secondary to patient treatment in the emergency department, which in turn causes the creation of overlays and duplicate patient records after acquiring the patient's identification information.
- Registration is a high staff turnover area where entry-level employees typically do not have adequate education and training.
- High-volume registration areas such as scheduling have a much higher risk of duplicate creation and overlays due to the lack of direct patient contact.
- Specimen registration information typically contains minimal identifiable patient information to locate existing patient records.
- The abundance of poor quality patient identification data stored and managed in siloed legacy systems causes the potential for data integrity issues.
- Data quality issues are magnified when source systems are not kept in sync throughout the exchange network.
- Data error corrections and duplicate remediation practices are not always performed in a timely and comprehensive fashion. Corrections that do occur are often performed by understaffed teams or inadequately trained staff.
- Variation in the tools and solutions that measure or address patient identification integrity also may compromise data integrity.

As mentioned above, identification requirements vary greatly across provider organizations. For example, the request for proof of identity is not always required at registration or check-in. Differences may include one or two forms of identification with or without a photo. Government-issued identification should be the standard.

There is a lack of data standards addressing accurate and complete data capture and data matching for patient identification. And the standards that do exist are limited in scope and adherence is suboptimal.

Existing standards are largely targeted at vendor and source system data format and position, not content accuracy, completeness, or relevance to industry changes. Examples include:

- The Patient Identification (PID) segments in the Health Level Seven (HL7) message not being consistently populated
- The use of components in the PID segment is not consistently implemented
- Vendors and providers typically do not upgrade to newer versions of the HL7 standards that incorporate better support of patient identification integrity

The healthcare industry initiative Integrating the Healthcare Enterprise (IHE) has long embraced the need for protocols that address patient identification, including Patient Identifier Cross Referencing Integration (PIX), Patient Data Query (PDQ), and Patient Administration Management (PAM) integration profiles. However these protocols and standards are not routinely adopted or consistently implemented by vendors, enterprises, or HIOs. Organizations rely instead on data being captured in compliance with older HL7 standards.

## Understanding HIE Participation

Today's health data exchange is working toward the goal of interoperability by increasing information sharing across disparate provider organizations using networks and internet-based technologies. To support high quality data exchange, AHIMA has published data quality standards that promote accurate, comprehensive, current, consistent, relevant, timely, granular, precise, accessible, and well-defined data. For more information on data quality attributes, refer to the 2012 AHIMA Data Quality Management Model practice brief [here](#).

While data exchange is on the uptake, electronically exchanged data rarely meet the standard for each data quality attribute listed above. This incompatibility compromises healthcare's ability to achieve truly successful interoperability. Exchange relies entirely upon demographic information contained in the individual person's record to initially match their clinical information amongst multiple provider organizations. Errors in the collection and transcription of demographic data, aging data, and lack of complete data contained within each provider's patient record severely limits interoperability.

HIOs currently use a variety of data delivery methods, which determine how patient records are sought and matched. Data from one provider may be “pushed” to another. For example, all electronic prescriptions generated within a hospital are automatically sent to a specific pharmacy, or all transcribed documents are forwarded to the provider(s) listed in the HL7 message. In these scenarios, the receiving provider generally already knows which patient the messages concern and thus uses relevant internal procedures to process the incoming transaction.

Data trading partnerships between providers may dictate the content and format of the HL7 message. “Pull” technology is used in scenarios where an HIO provides a record locator service (RLS) that requires a centralized index of all patients (Enterprise Master Person Index or EMPI) collected from HIO participants and a cross-reference to each participating organization’s medical record number (MRN) is assembled using advanced demographic matching techniques and is stored in the EMPI. A provider searches the HIO EMPI using either the patient’s demographic information or the provider’s previously assigned patient identifier. Once the corresponding HIO patient record identifier is located in the EMPI, the provider can use the RLS to request specific clinical information from other participating providers.

Electronic messages are then sent to each of the participating organizations that have stored records pertaining to the patient. Specific types of electronic clinical results required by the federal meaningful use requirements are extracted from each participating organization to be shared or exchanged with the requesting provider or in transition of care cases. The continuity of care document is a required transport mechanism in stage 2 of the “meaningful use” EHR Incentive Program. The continuity of care document is a core data set of the most relevant administrative, demographic, and clinical information facts about a patient’s healthcare, covering one or more healthcare encounters. It provides a means for one healthcare practitioner, system, or setting to aggregate all of the pertinent data about a patient and forward it to another practitioner, system, or setting to support the continuity of care. Its primary use is to represent a snapshot in time in the patient’s health history.

The Consolidated Clinical Document Architecture (C-CDA) is an HL7 document markup standard that specifies the structure and semantics of clinical documents. It is flexible XML-based clinical document architecture. There are many types of C-CDA documents including the consultation note, discharge summary, imaging integration and DICOM Diagnostic Imaging Reports, History and Physical, Operative Note, Progress Note, Procedure Note, and Unstructured documents.

A centralized data model can use both push and pull technologies. Each organization participating in the HIO sends (pushes) patient demographic along with clinical results to a central database managed by the HIO. Providers search the EMPI of this centralized database and pull the corresponding information for the applicable patient across to their system.

### **Data Trading Partner Definition**

Data Trading Partner: Person or entity who has executed an HIO’s data sharing agreement abiding by the HIO’s data governance policies and receiving health data from or providing health data across the HIE. Such a person or entity may be an individual provider, a clinic composed of multiple providers, an integrated delivery network, a clearinghouse, or a billing agent.

### **Record Matching Unifies Patient History**

A fundamental and critical success factor for the federated and centralized models is creating a comprehensive view of a patient’s entire clinical record from amongst the multiple participating organizations. To achieve this objective it is imperative that each contributing organization eliminate the duplicate and overlay records within their enterprise master patient index (EMPI) files. By correcting the integrity of the patient records at each contributing organization the advanced matching algorithm within the HIO EMPI will be able to accurately link records for the same patient from the disparate participating organizations.

The data trading partner’s source system master patient index (MPI) identifies an individual with records at their respective locations. Each data trading partner will have different unique identifiers for their patients and these will not correlate across exchanges.

In the absence of a nationally recognized patient identifier, the HIO must rely on sophisticated matching technology and the quality and completeness of the demographic data collected and maintained by each participating healthcare provider to create its own unique patient identifier. Therefore, once the HIO EMPI obtains demographic data from its participating

organizations, it must link all of these individual demographic records for one patient into a single record. In this model, the HIO links the different provider records for one patient into one record and assigns that patient a unique numeric identifier.

For example, patient John Smith has been seen at a physician office, hospital, and lab. Each organization submits their respective information to the HIO. The HIO ascertains, based on the demographic information from the respective organizations, that their records all belong to the same John Smith. Advanced matching algorithms are applied to key demographic attributes such as first, middle, and last name, gender, date of birth, Social Security number (if present), phone number, and address in order to reach this conclusion. Once the HIO associates each of their local identifiers to the HIO unique identifier, any provider can query the HIO EMPI using their local identifier and retrieve pertinent information submitted from any other provider. This process provides a transparent bridge from any provider's patient record to clinical information from other contributing providers via the HIO's EMPI utilizing an advanced demographic matching process.

Records from various providers are frequently linked by matching algorithms embedded in the HIO's software. The algorithms available to perform this linking function fall into three main groups: basic, intermediate, and advanced.

Whatever algorithm an organization uses to link records, the results should be verified by staff using record matching validity procedures during the initial system deployment and periodically thereafter. When applying designated HIO system requirements, a percentage of records from different participating organizations will be automatically linked if a sufficiently sophisticated algorithm is used. However, a statistically significant sample should always be reviewed to ensure only true overlap records are auto-linked.

Even with a sophisticated algorithm, the HIO will achieve significantly higher rates of record linkage if potentially overlapped records that have a record match weight lower than the auto-link threshold are reviewed and manually linked.

False positives (incorrectly linking similar records belonging to two different people), false alarms (detecting records that do not belong to the same person), and false negatives (not detecting multiple records belonging to the same person) will always occur with any algorithmic or manual system for identifying potential duplicates, linkages, or overlays. Common pitfalls include linking:

- Two closely related people with very similar names and dates of birth who live near each other, such as cousins who are named after the same individual
- Two individuals living in a dense urban area with the same common name, date of birth, and address
- Twins with the same first name

Failure to catch such errors can result in fragmented data due to missing clinical information or overlaid medical records and, subsequently, negative health outcomes, serious privacy breaches, and legal ramifications. HIOs and data trading partners should adopt a process of periodically measuring their false positive and false negative rates and include formal communication mechanisms to alert appropriate departments and staff when it is determined that duplicates or overlays may impact clinical or business operations.

It is important to realize that most healthcare information systems employ basic algorithm matching techniques and these techniques usually only identify 10 to 40 percent of the existing duplicate records. A few healthcare information systems employ intermediate algorithm matching techniques, which will identify 50 to 70 percent of the existing duplicate records. Sophisticated EMPI solutions incorporate advanced algorithm matching techniques that can identify up to 98 percent of the existing duplicate records. However, the advanced algorithms are not commonly used in healthcare. When choosing algorithm software or vendor consultant services, organizations are advised to investigate and understand proposed measurement techniques to ensure the highest degree of accuracy. The record matching algorithm and procedures employed by the provider should be examined via an independent audit using advanced matching algorithms to validate that they work correctly. This process will help to minimize or eliminate those records that match inappropriately (false positives) and any excessive reporting of records that do not belong to the same person (false alarms) as well as records that should match but fail to do so (false negatives).

### **Basic Algorithms for Linking Records**

Basic algorithms are the simplest technique for matching records and this approach is used by most healthcare information systems today. Comparisons are made on selected data elements—usually the name, date of birth, SSN, and sometimes the gender.

Exact match and deterministic algorithms are both basic matching tools. With exact matching, the data elements used to search must match exactly with those in the database in order to return a particular record. Deterministic matching is slightly more sophisticated; in addition to exact matches, partial matches may be used to return a record.

With an exact match algorithm, if the name “Smith” is entered in a search for a patient with the last name of Smith, the Smithe record would not be located, as it does not match exactly. However, a deterministic match using a substring of the first three letters (partial name) of the last name would return both Smith and Smithe. This is often accomplished via a “wild-card” search, which falls under the basic algorithm definition.

Wild-card searching involves the user entering a few letters of the value being searched and adds a character (frequently a common keyboard symbol) that instructs the database program to return every record that matches the limited letters entered. A wild-card search on “Smi” will return both Smithe and Smith, in addition to any other name in the database beginning with the first three letters and that match any other data element entered to refine the search, such as date of birth or gender.

It is important to note that most healthcare information systems employ basic search and matching techniques to locate patient records and identify potential duplicates. Therefore, many organizations that rely on their healthcare information system to maintain the accuracy of their patient index are often dependent on information technology staff to write customized reports, usually based on SQL queries. These queries and reports, although better than the basic duplicate reports contained in most healthcare information systems, still utilize basic matching techniques to identify potential duplicates. Basic algorithms typically only identify between 10 and 40 percent of the existing duplicate records within the MPI.

### **Intermediate Algorithms for Linking Records**

Intermediate algorithms use more advanced techniques to compare records. Fuzzy logic, nickname tables, phonetic encoding and arbitrary or subjective scoring systems are added to exact match and deterministic tools. A field match weight is subjectively assigned to key patient identifying attributes such as last name, first name, date of birth, and SSN.

For example, a match on the SSN may be assigned a score of 40 points, while a match of the last name scores 25. Conversely a mismatch would also be assigned a subjective negative value on SSN and last name. Records presented to the searcher must reach a minimum cumulative scoring threshold to qualify for inclusion.

Fuzzy logic and rules-based algorithms also may be a component of intermediate algorithms. These tools may utilize nickname tables, rules to address transposition of characters or names, digit rotations, and typographical errors within the MPI database.

Phonetic encoding is typically utilized in intermediate algorithms. These encoding systems, such as Soundex, the New York State Identification and Intelligence System (NYSIIS), or single, double, or triple metaphone, attempt to identify records with similar sounding names. For instance, a person with the last name of “Johansen” would be returned for a search on “Johnson” because the phonetic codes for these names are an exact match.

Intermediate algorithms may include a limited automated frequency adjustment. This adjustment will decrease the score assigned to a field match across two records if the actual field value (such as a common last name or a common date of birth like 01/01/2001) is computed to be present in a high volume of records in that data set.

With these types of algorithms, a search for Elizabeth Jones would return records for Betty Jones as well as Elizabeth Jones. A search for Richard David would return records for both Richard David and David Richard. A search for James Smith with a date of birth of 6/17/1978 would return Jim Smith with a date of birth of 6/17/1987 as a possible match. However, many of these intermediate algorithms are not tolerant of more than one data discrepancy in these key demographic data fields. Due to the subjective scoring and the varying number and type of matching techniques incorporated into intermediate matching algorithms, healthcare organizations can expect to identify anywhere from 50 to 70 percent of the actual duplicated patient records with this technology.

### **Advanced Algorithms for Linking Records**

Advanced algorithms contain the most sophisticated set of tools for matching records and rely on mathematical theory. Ideally, an advanced algorithm is used with the source systems to accurately identify the existing duplicate patient records and is incorporated in the patient search and selection process to accurately locate records irrespective of the variances between

what is entered during the patient search process and how many discrepancies reside in the existing patient record. However, as mentioned above, most source systems use a deterministic or intermediate approach.

The core intelligence within advanced algorithms can include bipartite graph theory, probability theory, mathematical and statistical models, and machine learning, which are applied to determine the likelihood or probability of a match on specified data elements.

Probabilistic matching uses the frequency of specific demographic data elements with an objective probability score assigned to each to adjust the relative value of the match or mismatch for the specified elements. The weight assigned to each field is relative to the weights assigned to other fields, but only after thorough analysis across millions of records (as opposed to a simple frequency analysis with a subjective field weight adjustment as found in intermediate matching approaches).

Advanced algorithms can also include machine learning such as natural language processing and neural networks, which use forms of artificial intelligence that simulate human problem solving. These systems actually “learn” as the program processes more data and will automatically tune the field weights as required based upon the learning achieved during the processing of the entire data set of records. In addition, some advanced matching approaches can also apply machine learning techniques to incorporate decisions made during human adjudication to improve their matching accuracy.

For example, a search for “Maria A Rodriguez, date of birth 8/18/1965, female, SSN 555-45-8888, 110 E 3rd Street” would return a record with the values “Mary **Ann** Rodrigues, date of birth 8/13/1956, female, SSN 555-44-8888, 110 E 3rd Street” and

“M. **Anna** Rodriguez **Jones**, date of birth 8/18/1965, female, SSN 555-45-8889, 110 **Third** Street.”

Despite a variance in multiple data elements, the advanced algorithm using the technologies listed above could identify the appropriate records, whereas many basic and even intermediate algorithms are not tolerant of such multiple data discrepancies.

## Measuring Duplicate Rates

Different methods can be found within the healthcare industry to measure the duplicate rate at a given point in time (the “existence” rate) or measure the ongoing duplicate creation rate. Additionally, different numerators and denominators are used in calculating these rates depending on the goal of the measurement. For instance, if the goal is to measure the overall volume of additional records created during a given time period or by a specific user, the numerator in the formula would be the number of “extra” or additional records created when an existing record already existed for the patient. However, the numerator should include both records when the goal is to measure the overall volume of records that are now in “error” (due to the original and the duplicate record being incomplete for the patient).

It is critical that HIOs measure the data quality of transactions from its participant’s source systems, including the volume of duplicate records sent from that source system. Additionally, HIOs should measure their overlap (record linking) rate as the higher this rate the more value the HIO is able to demonstrate to its data trading partners.

## Managing Patient Identity Integrity

Participating organizations such as hospitals and other healthcare delivery systems within the HIO are responsible for maintaining the integrity of the patient-identifying data within their own systems. Organizations that fail to carry out this responsibility not only compromise care within their own four walls, but also contaminate the HIO database and cause administrative complications or compromise care at other participating organizations. Additionally, HIOs are responsible for the overall management of the data they store and need to develop data governance policies and practices to optimize the quality of their data. The accuracy and completeness of key patient demographic data is essential to the effective management of patient identity integrity and is required to foster trust amongst HIO participants and data trading partners.

### MPI Definitions

HIOs and organizations going through the process of linking their records and cleaning up their MPIs may need to brush up on common MPI terms.

**Duplicate:** More than one entry or file for the same person in a single facility-level MPI. This causes one patient to have two different medical records within the same facility.

**Overlap:** More than one MPI entry or file for the same person in two or more facilities within an enterprise or HIO. For example, patient John Smith has medical record number 12345 at facility A and medical record number 447788 at facility B within the same enterprise-wide system. When both MPI databases are loaded into an enterprise MPI, the database does not link the two records. Thus, Smith ends up with two different enterprise identifiers and providers cannot view all clinical information across the enterprise for that patient.

**Overlay:** One MPI entry or file for more than one person (i.e., two people erroneously sharing the same identifier). Overlaid records are frequently caused when patient access staff select another patient's record during a scheduling or registration event. Sometimes interfaces cause the error if the receiving system lacks a robust patient record-matching program and "overlays" another patient's record from that inbound interface transaction. On occasion, overlays are caused by an incorrect merge of two records that belong to two different people.

Source: AHIMA MPI Task Force. "Building an Enterprise Master Person Index." *Journal of AHIMA* 75, no. 1 (Jan. 2004): 56A–D.

## **Strong Information Governance Provides Foundation for Data Standards, Record Matching**

A strong information governance program refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. Information governance is the policies, processes, and practices that address the accuracy, validity, completeness, timeliness, and integrity of data (data quality). It is the foundation for establishing data quality standards and record matching thresholds for auto-linking of records across the HIO.

An HIO must establish a strong governance program to:

- Ensure that records are linked accurately
- Ensure the notification to the data trading partner of identified uncertain or likely matches that may fall out of an automated patient matching algorithm

HIOs should establish maximum duplicate record tolerance thresholds for the source systems contributing data as well as minimum data integrity and quality standards—and hold data source partners accountable to these standards. Adherence to these standards is critical to the HIO's long term viability and trust relationship with the participants. Routine monitoring and reporting to the HIO board regarding compliance of these measures is a must. Although HIOs are not able to validate the data presented by a data trading partner, the HIO has a role of oversight and notification of poor data integrity.

Data sharing agreements between the HIO and data trading partners should address common patient identity integrity issues, such as:

- Null fields in the minimum required patient identity data fields
- Use of pseudo or default values (i.e., DOB 01011900, SSN 999999999, baby- or trauma-naming conventions )
- Un-reconciled duplicate records
- Excessive potential and unreconciled overlaps
- Excessive overlaid records from data trading partners

As part of the HIO's data governance practices, there are several important tasks that should be deployed if staffing, technology, and governance allow. These tasks will improve data quality for both the HIO and participating organizations:

- Measuring and reporting to a governance board the rate of duplicate records and the integrity of the data by de-identified data source.
- Providing to the appropriate provider organization the list of potential duplicates for internal reconciliation.
- Notification to the appropriate provider organization for potential overlaid records.
- Validating and linking potential overlap pairs verified to be the same patient. These potential overlaps have a record match score that is below the auto-link score threshold, but are weighted high enough to warrant manual review. Often an additional 50 percent of the auto-linked pair volume can be manually linked after human review.

- Periodically measure the HIO's record matching algorithm's performance to ascertain the false positive and false negative rate.

Without such data maintenance processes, large databases will become riddled with data integrity problems. Providers using the system will start seeing duplicate and overlaid records, lose confidence in the system, and potentially stop using it.

HIO information governance policies should ensure that after source system staff correct internal duplicates, the source system sends electronic merge messages to the HIO. The HIO must process these. This is a key step to keeping the HIO's data in sync with the data source. Additionally, the HIO data governance policies should ensure that data trading partners are not directly making corrections of source-specific duplicates or overlaid records in the HIO's system. Rather all source level data issues must be corrected on a timely basis in the source systems and then electronically communicated to the HIO via the appropriate HL7 message. Data correction rests with the source system processes and policies, thus data integrity is established at the source.

Additional HIO data governance policies should address data fields required to be entered when users are searching the system. Some HIOs have opted to keep their search criteria very conservative to minimize false positives in the list of returned records. A conservative approach that minimizes false positives lessens the potential for inadvertent disclosure of the wrong patient records during a search. However it also increases the false negatives and records are then not linked and presented to the end user.

All HIO system users must understand how and when to validate their selection of the correct patient record and how and when to determine that their search does not include the patient record sought. HIOs need to ensure their data trading partner users are trained on these record search and validation techniques, and that ongoing training is also required. Preferably, the training should include competency-based testing. HIOs should develop the training materials and distribute to each data trading partner for use in training their own staff.

Technology solutions that enable HIOs to support a comprehensive information governance program include:

- Defining custom processing logic to be used in assessing data quality of an incoming transaction before it is stored in the database.
- Defining validations to perform against specific fields, such as properly formatted Social Security numbers and dates of birth or checking for pseudo values.
- Identifying records that do not contain the minimum necessary valid fields to perform accurate matching.
- Defining custom logic based on predefined decision points for how records are matched during a transaction.
- Defining exceptions to prevent auto-linking of challenging record linking situations such as multiple births or common names in densely populated areas.
- Incorporating technology solutions that facilitate address verification.

## **Methods to Improve Patient Record Validation**

While it is the duty of every healthcare organization to protect the privacy of patient information within its control, this identifying information is vital in establishing appropriate links within an HIO. Organizations must balance the amount of information sent to an HIO to ensure that privacy requirements and patient care needs are addressed.

One method organizations may employ to improve the validity of links is to ensure that adequate identifying information accompanies the records sent to the HIO. Key demographic identifiers are essential, including at minimum last name, first name, middle initial or name, date of birth, gender, address, telephone number (land and/or cell), and ideally the patient's SSN. For more information on how to manage the SSN, see AHIMA's practice brief "[Limiting the Use of the Social Security Number in Healthcare](#)."

Historically, healthcare providers have collected SSNs and stored them in MPI databases to improve patient identification. With the increasing problem of identity theft, however, many individuals and organizations are reluctant to use this identifier for fear of compromising an individual's identity.

Some HIOs have addressed this concern by including only the last four digits of the SSN in the identifiers accompanying a patient's record. Absent the full SSN value, the last 4 digits of the patient's SSN significantly increases the accuracy of record matching. Other identifiers also may be provided that are used only in the background and are not visible to the user searching for a record. For example, mother's maiden name, birthplace, guarantor, next-of-kin, previous names, e-mail address, and

insurance policy information is helpful in determining what two records should be linked between disparate sources within the HIO. These identifiers can be used in the background by the algorithm and not displayed to the individual searcher.

Guidelines and procedures for HIO staff to utilize in determining overlap record validity should be developed and periodically updated. Ideally, HIM professionals who are skilled in record match validation methods or data analysis should perform a manual review of the proposed linked records during an initial implementation, with periodic review of the matching thresholds thereafter. This review will help determine at what point records represent the same individual consistently.

Validity procedures for human review of overlap matches that fall below the auto-linking threshold provide HIO staff with guidance on how to consistently decide whether two different records belong to the same patient. Overlap guidelines should be developed to ensure all HIO staff perform this validity process and are consistent in their overlap validity decision-making.

While false positives are usually relatively easy to identify (although common names and multiple birth records frequently pose challenges), identifying a false negative (two records that fail to match) is more challenging. In the absence of research-based studies of record-matching algorithms, it is recommended that HIOs and participating organizations undertake a periodic audit of their respective patient indexes using an advanced matching algorithm to validate the accuracy of their current matching process.

## **Identification Models Continue to Emerge**

Today, many groups continue to work on various technology solutions to ensure accurate patient identification. The following are emerging identity detection models that are available both within and beyond the United States.

Smart card technology includes the use of plastic cards, fobs, subscriber identity modules (SIMs) that are used in mobile phones, electronic passports, and USB tokens.

In comparison, biometric technologies identify individuals using unique biological characteristics such as anatomical, physiological, or behavioral. These technologies include approaches such as facial recognition, fingerprinting, hand printing, palm-vein or iris scanning, and voice recognition.

It should be noted that usage of a biometric approach satisfies one part of the two-factor authentication required by stage 2 of the meaningful use EHR Incentive Program and CMS' Conditions of Participation. Although emerging identification models offer significantly greater identification accuracy once a patient is enrolled, attention must still be given to linking the biometric data to the historical data. Therefore, biometrics should be coupled with advanced algorithms to facilitate linking of current and historical data and provide a total solution.

## **Organizational Training Enables HIM Professionals to Step Up**

HIM professionals have the expertise and ability to manage MPI database integrity and play a critical role in directing enterprise-wide and HIO activities that affect the MPI and the subsequent exchange of health information. HIM professionals in MPI leadership roles provide data and feedback necessary to improve registration accuracy, contribute to software selection, and maintain accurate data transference among downstream databases.

Comprehensive training requires that written policies and procedures for patient search routines are readily available to authorized users of the HIO search function. Essential to maintaining MPI integrity is ongoing training, not just initial training. Ideally, competency standards are incorporated as part of the training program.

Targeted user training should include:

- Review of the search methodology specific to enterprise and HIO MPI software. Inclusion of standardized patient-naming conventions in training programs is an essential key to prevent duplicate creation at all levels.
- Appropriate use of spaces, hyphens, and apostrophes must be established.
- Provision for use or exclusion of titles such as Rev., Dr., and the proper designation of prefixes or suffixes such as Sr., Jr., II, or III should be included in documentation guidance.
- Strong emphasis must be placed on collecting and using the patient's legal name, excluding nicknames, and proper provision for the middle name (or, at minimum, middle initial).

Capturing minimum demographic fields to properly identify the patient should be strictly enforced as it greatly impacts accurate patient identification.

## To Manage One Must Measure

Patient identity integrity aligns neatly with the messages of multiple management gurus—you can't manage what you can't measure. This wisdom is central to healthcare reform, quality of care improvement, administrative efficiency, and reduced healthcare costs. A single error rate snapshot is only the start in the plan-do-study-act cycle that, when done well, can make a critical difference to the quality of care a patient receives. Striving for accuracy is as important in healthcare as it is in landing an airplane. Setting this goal for patient identity accuracy will be a significant step in achieving high healthcare quality standards throughout the entire healthcare delivery system.

## Notes

1. Department of Health and Human Services. "Medicare and Medicaid Programs; Electronic Health Record Incentive Program—Stage 2; Proposed Rule." *Federal Register* 77, no. 45 (March 7, 2012). <http://www.gpo.gov/fdsys/pkg/FR-2012-03-07/html/2012-4443.htm>.
2. National Committee on Vital and Health Statistics. Meeting Proceedings from December 1, 2010. <http://www.ncvhs.hhs.gov/101201tr.htm>.
3. Office of the National Coordinator for Health IT. "Patient Identification and Matching Final Report." February 7, 2014. [http://www.healthit.gov/sites/default/files/patient\\_identification\\_matching\\_final\\_report.pdf](http://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf).

## References

AHIMA Data Quality Management Task Force. "Data Quality Management Model." June 1998.

AHIMA e-HIM Workgroup on EHR Data Content. "Data Standard Time: Data Content Standardization and the HIM Role." *Journal of AHIMA* 77, no. 2 (February 2006): 26–32.

AHIMA e-HIM Workgroup on HIM in Health Information Exchange. "HIM Principles in Health Information Exchange." *Journal of AHIMA* 78, no. 8 (September 2007).

AHIMA e-HIM Work Group on Patient Identification in RHIOs. "Surveying the RHIO Landscape." *Journal of AHIMA* 77, no. 1 (Jan. 2006): 64A–D.

AHIMA e-HIM Work Group on Regional Health Information Organizations (RHIOs). "Using the SSN as a Patient Identifier." *Journal of AHIMA* 77, no. 3 (March 2006): 56A–D.

AHIMA MPI Task Force. "Building an Enterprise Master Person Index." *Journal of AHIMA* 75, no. 1 (January 2004): 56A–D.

AHIMA MPI Task Force. "Maintenance of a Master Patient (Person) Index—Single Site or Enterprise." October 1997.

AHIMA MPI Task Force. "Master Patient (Person) Index: Recommended Core Data Elements." July/August 1997.

AHIMA MPI Task Force. "Merging Master Patient Indexes." September 1997.

Altendorf, Robin L. "Establishment of a Quality Program for the Master Patient Index." AHIMA's 79th National Convention and Exhibit Proceedings, October 2007.

Angeloni, Magaly and Mike Berry. "Probabilistic Record Matching and De-duplication Using Open Source Software." Presentation at CDC National Immunization Conference. October 20, 2004. [https://cdc.confex.com/cdc/irc2004/techprogram/paper\\_5997.htm](https://cdc.confex.com/cdc/irc2004/techprogram/paper_5997.htm).

Arellano, Max G. and Gerald I. Weber. "Issues in Identification and Linkage of Patient Records Across an Integrated Delivery System." *Journal of Healthcare Information Management* 12, no. 3 (Fall 1998).

Department of Health and Human Services. “Voluntary 2015 Edition Electronic Health Record (EHR) Certification Criteria; Interoperability Updates and Regulatory Improvements; Proposed Rule.” *Federal Register* 79, no. 38 (February 26, 2014). <http://www.gpo.gov/fdsys/pkg/FR-2014-02-26/pdf/FR-2014-02-26.pdf>.

Durkin, Stacie, and Lorraine Fernandes. “Measuring Patient Identity Integrity: Foundation to Analytic Success.” Presentation at 85th AHIMA National Convention and Exhibit. October 2013.

eHealth Initiative. “2012 Annual HIE Survey Results—Report on Health Information Exchange: Supporting Healthcare Reform.” [http://www.ehdc.org/resource-center/publications/view\\_document/43-survey-2012-annual-hie-survey-results-report-on-health-information-exchange-supporting-healthcare-reform-data-exchange](http://www.ehdc.org/resource-center/publications/view_document/43-survey-2012-annual-hie-survey-results-report-on-health-information-exchange-supporting-healthcare-reform-data-exchange).

eHealth Initiative. “Fifth Annual Survey of Health Information Exchange at the State and Local Levels.” September 2008. <http://www.sftvision.com/eHealthInitiativeResults2008SurveyonHIEs.pdf>.

Fernandes, Lorraine et al. “Big Data, Bigger Outcomes.” *Journal of AHIMA* 83, no. 10 (October 2012): 38-43.

Fernandes, Lorraine. “Patient Identification in Three Acts.” *Journal of AHIMA* 79, no. 4 (April 2008): 46-49.

Gartner. “IT Glossary: Big Data.” 2013. <http://www.gartner.com/it-glossary/big-data/>.

Healthcare Information Technology Standards Panel. Comparison of CCR/CCD and CDA Documents, Version 1. December 2009. [http://publicaa.ansi.org/sites/apdl/hitspadmin/Matrices/HITSP\\_09\\_N\\_451.pdf](http://publicaa.ansi.org/sites/apdl/hitspadmin/Matrices/HITSP_09_N_451.pdf).

Hillestad, Richard et al. “Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System.” RAND Corporation. 2008. <http://www.rand.org/pubs/monographs/MG753.html>.

Markle Foundation. “Achieving the Health IT Objectives of the American Recovery and Reinvestment Act: A Framework for ‘Meaningful Use’ and ‘Certified or Qualified EHR.’” Markle Connecting for Health, April 30, 2009. [www.markle.org/downloadable\\_assets/20090430\\_meaningful\\_use.pdf](http://www.markle.org/downloadable_assets/20090430_meaningful_use.pdf).

Office of the National Coordinator for Health IT. “Health Information Exchange (HIE), HIE Benefits.” <http://www.healthit.gov/providers-professionals/health-information-exchange/hie-benefits>.

Office of the National Coordinator for Health IT. “Implementing Consolidated-Clinical Document Architecture (C-CDA) for Meaningful Use Stage 2.” April 5, 2013. [http://www.healthit.gov/sites/default/files/c-cda\\_and\\_meaningfulusecertification.pdf](http://www.healthit.gov/sites/default/files/c-cda_and_meaningfulusecertification.pdf).

Smart Card Alliance. “Smart Cards and Biometrics in Healthcare Identity Applications.” May 2012. <http://www.smartcardalliance.org/pages/publications-smart-cards-and-biometrics-in-healthcare-identity-applications>.

## **Prepared by (Update)**

Julie A. Dooling, RHIA  
Stacie Durkin, MBA, RHIA, RNC  
Lorraine Fernandes, RHIA  
Beth Just, MBA, RHIA, FAHIMA  
Steve Kotyk  
Ellen Shakespeare Karl, MBA, RHIA, CHDA, FAHIMA  
Kathy Westhafer, RHIA, CHPS

## **Acknowledgments (Update)**

Kathleen Addison, CHIM  
Jill S. Clark, MBA, RHIA, CHDA  
Dana DeMasters, MN, RN, CHPS  
Cynthia Hilterbrand, RHIA, MBA  
Kelsey Hirt, MS, RHIA  
Lesley Kadlec, MA, RHIA  
Gavin Krumenacker

Stephanie Luthi-Terry, MA, RHIA, FAHIMA  
Lori McNeil Tolley, RHIA  
Jennifer Miller, MHIS, RHIA  
Erik Pupo, FHIMSS, CPHIMS  
Dan Rode, MBA, CHPS, FHFMA  
Theresa Rihanek, MHA, RHIA, CCS  
Diana Warner, MS, RHIA, CHPS, FAHIMA  
Lou Ann Wiedemann, MS, RHIA, CHDA, CDIP, FAHIMA

### **Prepared by (Original)**

2009 HIE Practice Council  
Beth Haenke Just, MBA, RHIA  
Diane P. Fabian, MBA, RHIA  
Lenore L. Webb, RHIA  
Beth M. Hjort, RHIA, CHPS

### **Acknowledgments (Original)**

2009 Privacy and Security Practice Council  
2009 EHR Practice Council  
Linda Bock, RHIA

---

The information contained in this practice brief reflects the consensus opinion of the professionals who developed it. It has not been validated through scientific research.

---

**Article citation:**

AHIMA Work Group. "Managing the Integrity of Patient Identity in Health Information Exchange (2014 update)" *Journal of AHIMA* 85, no.5 (May 2014): expanded web version.

---

**Driving the Power of Knowledge**

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.