



## **Patient Identity Integrity**

**A White Paper by the  
HIMSS Patient Identity Integrity Work Group**

**December 2009**

## **Table of Contents**

<b>Executive Summary</b> .....	<b>3</b>
<b>Standards</b> .....	<b>9</b>
<b>Interfaces</b> .....	<b>14</b>
<b>Algorithms for Linking Records</b> .....	<b>17</b>
<b>Unique Identifiers</b> .....	<b>20</b>
<b>Business Processes</b> .....	<b>24</b>
<b>Data Accuracy</b> .....	<b>26</b>
<b>Data Quality</b> .....	<b>30</b>
<b>Training</b> .....	<b>35</b>
<b>Medical Devices</b> .....	<b>38</b>
<b>Barriers to Patient Identification Integrity</b> .....	<b>42</b>
<b>Appendix A: VUHID</b> .....	<b>46</b>
<b>Appendix B: Algorithmic-based Matching—The Basics</b> .....	<b>48</b>
<b>Appendix C: Patient Identification Training Checklist</b> .....	<b>55</b>
<b>Appendix D: Additional Sources</b> .....	<b>56</b>
<b>Appendix E: Glossary of Key Terms</b> .....	<b>58</b>
<b>Appendix F: HIMSS Patient Identity Integrity Work Group Members</b> .....	<b>60</b>

## **Executive Summary**

### **Introduction**

In June 2008, the HIMSS Privacy and Security Steering Committee created a Patient Identity Integrity (PI Integrity) Work Group composed of volunteer industry experts to address concerns raised from a variety of industry sources about the need for guidance in understanding the complex issues surrounding PI Integrity.<sup>1</sup>

PI Integrity is the accuracy and completeness of data attached to or associated with an individual patient. Data must be reliable, reproducible, and sufficiently extensive for matching purposes. Completeness refers not only to having adequate data elements present but also the correct pairing or linking of all existing records for that individual within and across information systems. PI Integrity is of central importance to achieving quality of care, patient safety, and cost control.

While it is relatively easy to see the implications of PI Integrity on quality, safety, and cost, it is far more difficult to grasp the complexity of maintaining identity integrity in the real operational environment. To solve the problem of assuring a state of high quality PI Integrity, one must look at the entire process of patient identity management (PIM). The PI Integrity Work Group identified nine variables that influence, in varying degrees, our ability to build and sustain a database in a high state of identity integrity. These key influencers are: industry standards, interfaces, algorithms, unique identifiers, business processes, data accuracy, data quality, training, and medical devices.

This paper discusses each variable at a high level and its impact on PI Integrity. Due to the complexity of detail, this paper does not provide an in-depth discussion of each variable; rather it seeks to offer sufficient understanding to provide professionals with a sound basis for planning and decision-making. References are cited in the document for those seeking more in-depth information, along with a select bibliography at the end. Barriers and specific recommendations are discussed in each section.

Specifically excluded from the scope of this paper are the topics of provider identity and user identification, authentication, and authorization for system access. Initial discussions within the HIMSS Work Group uncovered confusion between these topics and PI Integrity. When people speak of privacy and security, they first go to the technical concepts of identification, authentication, and authorization for system access and how they are supported by the system. This paper does not deal with these issues. This paper deals with the holistic process for matching records for one individual person within and across multiple systems. It begins to address the challenges of overlapping concepts, shared terminology and limited understanding and misunderstanding of the scope of PI Integrity.

### **Description of Problem**

The ultimate goal is the accurate identification of the patient and linking of all related information to that individual within and across systems. Linking the wrong clinical

---

<sup>1</sup> The list of work group members can be found in Appendix F.

information to a person can not only cause great personal harm to the patient, but can also incur huge costs to the healthcare provider in correcting and mitigating the error. Incorrect information impacts patient safety and compromises quality of care. Good clinical decisions based on bad data become bad clinical outcomes. For example, the wrong patient who received the wrong lens implant must undergo a second procedure to correct it. A third procedure is performed on the correct patient who never got the procedure due to the identity error. In addition to the obvious negative impact on the patient, providers are also negatively impacted in that they must absorb the costs of correcting their mistakes. Also, there are potentially substantial legal costs that could damage the reputation of the institution based on bad clinical outcomes.

The healthcare industry is moving aggressively to expand the use of electronic health records (EHRs), electronic medical records (EMRs), personal health records (PHRs), and health information exchanges (HIEs). Recent legislation creates new financial and regulatory incentives for increased use by physician practices and HIEs. As health IT makes deeper inroads into the healthcare community (President Obama's call for an EHR for everyone by the year 2014), and HIEs and the Nationwide Health Information Network (NHIN) ramp up to connect information locally, regionally, and nationally, PI Integrity becomes a critical issue that must be understood and addressed. A local system with a poorly maintained or "dirty" master person index (MPI) will only proliferate and contaminate all of the other systems to which it links. These events have brought PI Integrity to the forefront because of its critical importance to the successful implementation of these information systems. HIEs magnify the problem for several reasons: (a) they do not have control of the patient identity data capture process, (b) they receive data from many different provider MPIs and the data fields being sent to the HIE are frequently not consistent and (c) they do not control the interfaces coming into the HIE's database. Since a major portion of the activities of an HIE involve the exchange of clinical information between independent (and therefore heterogeneous) entities, these issues create significant additional stress on the HIE's ability to effectively maintain their system with high data integrity.

Without identity integrity, information pertaining to one individual may exist in one or multiple databases where it resides as a "duplicate," inaccessible or unknown to those needing to see the complete or most current picture. Conversely, information on two individuals may be combined erroneously into one record; this is called an "overlay." These conditions are common symptoms of poor (or lack of) data management and can result in:

- uninformed or marginalized clinical decision-making that impacts quality outcomes and patient safety;
- poor utilization of healthcare resources leading to repeated tests or procedures due to lack of access to existing reports or results;
- inability to drop a bill to collect payment or missed billing opportunities when lab results are posted to an old account or wrong account; and,
- manipulation of the system for illegal purposes such as drug seekers, drug diversion or medical identity theft, to name a few.

Because of the enormous impact that PI Integrity has on the clinical, financial, and administrative business of healthcare, it is imperative that the quality of an organization's identity integrity be addressed prior to sharing data externally with other stakeholders.

### **Standards and Interfaces**

There is a common belief that technology and standards can solve the problem of matching records accurately. As the industry has moved forward in standards development, it has not solved the problem with consistency of implementation of those standards. Two major gaps in standards have been identified. The first gap relates to data standards and is discussed below under Data Quality. The second gap in standards relates to the effectiveness of the variety of matching approaches currently used. These algorithmic approaches are discussed in detail in the section on Algorithms.

In the area of interfaces, major strides have been achieved to improve the manner in which computer systems share information. Most notably, Integrating the Healthcare Enterprise (IHE) promotes the use of established standards and has developed profiles for patient identification, although some issues still exist in the area of patient privacy. A problem also exists in the implementation of interfaces at the local level. The number (30 to 60 in an average hospital) and complexity of applications and systems being interfaced creates a daunting task for IT staff to implement and maintain. Mapping data among the systems requires both technology and domain expertise. The lack of domain expertise impacts the effectiveness of many IT installations. Collaboration between information management and information technology staff in the development of thorough testing plans would improve the efficiency as well as the effectiveness of the systems.

### **Algorithms for Record Matching**

Various technological methods based on algorithmic formulas are used to match and link data. There are three critical areas that impact this process: 1) the quality of the data used for matching, 2) the quality of implementation of the matching criteria, and 3) the effectiveness of the actual algorithm itself. These algorithms are frequently proprietary "black box" solutions.

The matching criteria are built within the MPI solutions or can be positioned above several MPIs in an Enterprise. These solutions include criteria based settings capabilities along with weighting mechanisms to assign a level of match probability. Unfortunately, the settings are configured differently and inconsistently within and across organizations; not implemented correctly to the point of total ineffectiveness; or implemented too stringently or too loosely for adequate matching purposes. There is a lack of industry knowledge and scientific study on the reliability of these proprietary applications. Consequently, no standards have been set for performance expectations or successful outcome ratio. In an ideal scenario, the matching outcome ratio would be 100% successful matches. Without any data on the effectiveness of the matching solutions, there is no way of knowing if they are functioning at a 99 percent or 75 percent level of successful match. Canned data reports provide incomplete information based on their formulaic view of the data.

### **Unique Individual Identifier**

The virtues of a unique identifier for each individual have long been recognized. Testament to that is the gradual recognition and evolution of the Social Security Number (SSN) as a desirable data element to manage person identity. No available identifier comes close to its pervasiveness and consistency in the population. In 1996, the Health Insurance Portability and Accountability Act (HIPAA) mandated a Unique Individual Identifier for healthcare purposes. There is a current Congressional proscription<sup>2</sup> against considering unique identifiers. Due to public concerns over privacy, Congress prohibited DHHS from using the authority under HIPAA to promulgate a final rule or standard. This is an issue that affects both the public and private sector. Progress on a unique identifier solution has been slow due to concerns about the cost to implement, privacy risks in amassing large centralized databases, and technical issues on compatibility with existing systems, as well as lack of national consensus on what identifier to use.

A unique identifier is linked to one individual, provides unambiguous identification, is immutable over time with consistent syntax, is simple of concept to implement, and is cost effective when compared with other solutions. More importantly, it is tremendously effective in reducing false negatives in the identity matching process.

Due to the concerns discussed in this document, we raise readers' attention to the HIMSS member-created and Board-approved Principles on Government Initiatives, which states that "HIMSS calls for the Secretary of HHS, under the direction of the U.S. Congress, to establish an informed patient identity solution. As part of this solution, steps need to include:

- a) Congressional lifting of the prohibition against HHS studying UI solutions;
- b) HHS conducting a study of the cost/benefit and practicality of implementing a UI solution; and,
- c) HHS establishing pilot implementations of unique identifiers to document the challenges and benefits."<sup>3</sup>

An identifier to be implemented at a national level should come with specific limitations on its use and strict penalties for inappropriate use to forestall problems related to privacy, identity theft, and potential for abuse. One can only look at the creative uses of the SSN to understand the value of such a unique identifier to general business processes.

---

<sup>2</sup> The text from the 1998 Omnibus Appropriations Act (not the official title) signed into law (PL 105-277): "SEC. 516. None of the funds made available in this Act may be used to promulgate or adopt any final standard under section 1173(b) of the Social Security Act (42 U.S.C. 1320d-2(b)) providing for, or providing for the assignment of, a unique health identifier for an individual (except in an individual's capacity as an employer or a health care provider), until legislation is enacted specifically approving the standard."

<sup>3</sup> [2009 HIMSS Principles on Government Initiatives](#).

## **Business Processes**

Business processes and the decisions that surround and support the technology play an important role in PIM. There is a general lack of understanding, recognition, and ultimate funding of the business processes required to support and maintain PI Integrity. Business processes should identify the workflow, policies and procedures to ensure PI Integrity. These would include who has responsibility and authority to correct patient identity information. This becomes more challenging when considered in the context of an HIE record or even a provider record. The governance and stewardship discussions, as well as policy development and execution, must involve a multi-stakeholder group that has executive-level involvement and support. Authority must be assigned in every organization for official active and ongoing oversight.

To be fully utilized, technology requires human input through design, intervention, and maintenance. This includes not only the intelligent design and consistent implementation of the matching rules, but also the quality of the data being collected and entered in the systems. It also includes the secondary business and regulatory functional capabilities related to such things as medical identity theft and the Red Flag Rules.<sup>4</sup>

Technology is only as good as the data that go into the system. The processes surrounding the technology are critical to the success of a solid PI Integrity program. While there is a body of knowledge that surrounds proper patient identification business processes, in general, there is a lack of adequately defined business process standards or guidance. Data accuracy and data quality are critical business processes. This includes guidance on measuring and reporting error rates in healthcare databases, and identifying an acceptable error rate. Manufacturing has sophisticated methods for measuring its tolerance for error on the production line. The public would not accept a 97 percent accuracy rate for successful landings in the aviation industry. In healthcare, there are no standards for data accuracy or data quality.

Data accuracy measurement standards include defining identity data errors and differentiating between them and data discrepancies. Guidance should define how to calculate a duplicate rate for a static snapshot, as well as a duplicate creation rate. PI Integrity standards should define what information or data should be considered “minimum” in identifying a patient accurately.

Further adding to the challenge of PI Integrity is the issue of dealing with historical records in legacy systems. Consideration must be given to connecting decades of records that are fraught with inaccurate, inconsistent, and incomplete data.

The finance industry has a high degree of interaction between the consumer and its data to provide continued review and assessment of the currency and accuracy of the information. In healthcare, the consumer (patient) does not control the accuracy of his/her own data.

---

<sup>4</sup> [“Protect your Patients, Protect Your Practice: What You Need to Know about the Red Flags Rule.”](#) American Medical Association.

Accuracy is driven by the organization's internal abilities to collect, enter, and process data. Healthcare does not have the same level of interaction, although that could change with the advent of the PHR and other available interactive formats.

The complexity of the business processes underpinning PI Integrity and its high error opportunity index is not recognized by the industry. Consequently, the business processes are not properly funded or staffed. Some of the lowest-paid and least-trained staff are creating the initial data entries in the information systems. Correcting errors is a low priority and it is also costly. There is limited understanding of the impact this has on an organization's general operations, quality of care, patient safety and financial operations. Rather than being viewed as a form of risk prevention, reduction in funding is justified with the belief that savings will be achieved through technology.

### **Training**

A key business process is the routine monitoring of and feedback on the quality and accuracy of employee performance. Performance cannot be measured without training in the expected knowledge and outcomes. PI Integrity training must include all individuals who have an impact on the PI Integrity processes and training must be ongoing. Frequently only patient access/registration staff receive PI Integrity training without thought to the variety of other individuals who enter data into the system. In addition to the registration staff, these may include staff in health information management, scheduling (OR, clinics, clinician offices, etc.), laboratory, IT, patient finance/billing office, and medical device clinicians.

### **Medical Devices**

Device incompatibility is a major problem in the PI Integrity sector. Neither hardware connections nor data are compatible. Data incompatibility due to no common format means lack of data interoperability for exchanging data.

Current medical device standards have not adequately addressed the issues surrounding connectivity between devices and healthcare systems. Because there is no connectivity, the devices do not receive patient identifiable information or ADT information. Devices are tracked by location and device number for correlation with a particular patient. There is a need to develop a process to ensure that the patient and the device in legacy systems are being integrated with clinical systems. Standards must be developed for patient demographic data sets, data capture, matching criteria across systems, with coordination of these standards. National policy, backed up by legislation, may be a consideration to address compliance for medical devices with standards.

The following sections of this report provide an in-depth discussion of the key influencers in PI Integrity. Each section ends with specific recommendations for moving forward on efforts to ensure true PI Integrity in our healthcare systems. To have true PI Integrity, there must be a holistic approach that provides ongoing support for the people, processes, and training as well as the technology.

## **Standards**

In 1996, Congress passed HIPAA. In Title II of HIPAA, development of a national patient identifier was required and has since been considered by many as fundamental to improving healthcare administration and patient safety. The lack of strong security and privacy controls at that time has delayed research and development of unique patient identifier (UPI) standards, leaving the industry to wholly depend upon the application of deterministic and mathematical matching to link patient records. This process, while widely used and considered effective, is applied without any formal standards.

The inability to uniquely identify patients through a single identifier has contributed to inadequate PI Integrity. Improving this process will require the industry to take a standards-based approach that includes assessing the various record matching techniques (deterministic, probabilistic, etc.) to formalize a consistent, reliable process across applications. PI Integrity management must promote usability at the most basic levels (e.g., linking with other identifiers, consistency of patient attributes, supporting electronic transactions, providing integrity of the data elements) and interoperate across systems and organizations. Security and privacy of the identifier must be at the forefront of standards development. And, the identifier must reflect rigidity and be defensible when linking records. Thus, standards will play a significant role not only in streamlining business processes, but in the reliability of patient information, as well.

Over the past decade, data reliability and information flow have evolved through the application of technical and clinical standards that reduce the complexities of business processes, ensure consistency in the design of technologies, and facilitate interoperability across health IT. While there are two ASTM International standards for identifiers (E 1714 "Standard Guide for Properties of a Universal Healthcare Identifier and E 2553 "Guide for Implementation of a Voluntary Universal Healthcare Identification System"), no standard exists for the data elements used in algorithmic record-matching. The existing standards help facilitate PIM. These standards are technology-agnostic and flexible in the context of health information technologies. Standards can be defined as 'specifications' that are based on a distinct set of requirements as determined by the industry they support—healthcare—or the function they serve—claims processing, coordination of benefits, referral certification, auditing, identity management, role based access, etc.

When applied to information systems, standards are a valuable tool for ensuring interoperability, extensibility, improved workflow, increased functionality, security, privacy, and/or compliance with governance constructs. As such, identifying a patient must give consideration not only to the components that form the basis for uniquely identifying individuals, but to how the data are formatted, the necessary specifications for leveraging a unique identifier within clinical workflows, and the specifications necessary to protect the unique identifier from unlawful use or disclosure. This section of the document provides a brief overview on the current state of the industry, barriers to a UPI, and recommendations.

## **Current State of the Industry**

Health information technologies, while fundamental to patient care, are challenged in their capability to ensure security, privacy, and to interoperate across systems and organizations. The lack of interoperability limits the usability of patient information and creates increased complexity between local, regional, and national health IT initiatives. When systems communicate, data can seamlessly be moved from system-to-system and between disparate organizations, facilitating high-quality care. The need to develop interoperable health IT systems has led to standardization and harmonization of clinical terminologies, security, privacy, and technical constructs. Health IT standards deliver structured content and formatting to facilitate the accurate interpretation of messages between clinical and business systems.

IT standards are developed by both national and international standards bodies, as well as associations. Some of the more generic standards bodies include the International Organization for Standardization (ISO), the American National Standards Institute (ANSI), and the National Institute of Standards and Technology (NIST). In addition to these organizations, there also exist health specific standards bodies which include, but are not limited to, Health Level Seven (develops standards for clinical and administrative data), Accredited Standards Committee (develops electronic data interchange standards), and the International Health Terminology Standards Development Organization (develops and promotes the use of SNOMED clinical terminology). In addition, there are several organizations whose responsibility it is to profile, harmonize, and promote health IT standards, including the IHE and the Healthcare Information Technology Standards Panel (HITSP). Both of these organizations play a significant role in the development, adoption, and integration of health IT standards.

## **Patient Identity Standards**

Patient identity standards have been conceptualized in multiple forums over the years. Those that have been tested and approved for use nationally are based on health information priorities as specified by the American Health Information Community (AHIC).<sup>5</sup> AHIC was chartered in 2005 to make recommendations on how to accelerate the development and adoption of health IT. In its role, AHIC established priorities for HIE in the areas of Consumer, Provider, and Population Health. As these priorities were set, they were turned over to HITSP in the form of ‘use cases.’ Use cases describe specific healthcare scenarios that involve information exchanges.

HITSP was established to harmonize and integrate standards to facilitate clinical and business needs for sharing information across organizations and systems.<sup>6</sup> HITSP evaluates the use cases and provides recommendations for health IT standards (e.g., Specifications, Implementation Guides, Code Sets, Terminologies, and Integration Profiles) that promote efficient, streamlined workflows in support of interoperable and secure HIEs. HITSP recommendations have been formalized as “Interoperability Specifications” and

---

<sup>5</sup> [American Health Information Community. Health IT](#), U.S. Department of Health and Human Services.

<sup>6</sup> <http://www.hitsp.org/>

constructs including Transaction Packages.<sup>7</sup> Interoperability Specifications focus on the standards necessary to facilitate one or more use cases in the exchange of information.

The result of HITSP's work is a series of capabilities and specifications for operationalizing, among other concepts, PIM across health IT. Integration profiles that have been adopted for use in PIM are the work of IHE. This organization and the integration profiles are discussed below.

### **IHE and Patient Identity Standards**

IHE ([www.ihe.net](http://www.ihe.net)) is a global initiative that creates technical frameworks—freely available in the public domain—for passing vital health information seamlessly—from application to application, system to system, and setting to setting—across multiple healthcare settings. IHE leverages widely-accepted and approved clinical and functional standards for developing profiles and accompanying frameworks that offer a consistent technical approach to interoperability. The aligning technical frameworks offer organizations and vendors a roadmap to integrate profiles, facilitating reduced complexity and open functionality across health IT. HITSP has incorporated numerous IHE Technical Framework and Profile components into their Interoperability Specifications (HITSP IS).

IHE is organized by clinical and operational domains.<sup>8</sup> The IT Infrastructure (ITI) domain, which focuses on standards-based interoperability solutions,<sup>9</sup> consists of 15 Integration Profiles. ITI profiles are grouped into three distinct information exchange components—Patient Identity and Administration, Patient EMR Exchange, and Information Security. This paper focuses on Patient Identity and Administration only.

### **IHE Patient Identity and Administration**

Patient Identity and Administration profiles facilitate transactions that deal with patient registration, coordination, and discharge. The NHIN has adopted some of these profiles to facilitate information exchanges between health information organizations. These profiles are described in detail in the section on Interfaces.

IHE integration profiles serve as a foundation for leveraging and aggregating patient information. Profiles are technology agnostic and can work in conjunction with whatever architecture is put in place. Standardizing patient identification components and specification will increase the usability of profiles, as well as to simplify queries between applications, and enable more complex processes across enterprises. While the current profiles reflect priority information sharing needs across healthcare settings, additional standards may be required to support priorities within the broader healthcare and public health community.

---

<sup>7</sup> [HITSP Interoperability Specification Overview, V 1.0, July 20, 2007](#)

<sup>8</sup> [IHE Domains](#)

<sup>9</sup> [IHE IT Infrastructure](#)

The Association for Information and Image Management and ASTM International ASTM E 1714 – A Standard Guide for Properties of a Universal Healthcare Identifier (UHID) offers a distinct set of characteristics necessary to create a UPI. UHID characteristics provide a conceptual model from which standards could be derived for application in current PIM mechanisms. While not all characteristics would necessarily apply, those that assert security of the data elements, usability across systems and enterprises, and reliability of the linked data should be adopted as a specification for PIM. By applying these characteristics, granularity of specifications can be achieved over time. Some of the key characteristics that should be leveraged include standardized policies and procedures, patient confidentiality and security of data elements, and the application of industry based standards.

### **Barriers to Patient Identity Standards**

Irrespective of the adoption of a standardized patient identifier, many issues exist that pose risks to PI Integrity. Some of the issues identified below can be easily addressed with algorithms as discussed in that section. Others, however, will require focused attention and examination in order to reduce the implication of their impact on patient identifiers.

1. **Lack of a standard for patient identifier:** The SSN is still valued by many in that it enhances matching capability. This includes back-end machine matching, as well as human record validation. Use of the SSN increases risks to patient privacy. However, removal of the SSN can negatively impact accuracy rates in linking patient records.
2. **Interface mapping:** Record mapping is highly dependent on demographics data and personal identifiers. People move, which means that demographics data can be unreliable. Surnames may be changed or are highly duplicative, often requiring human intervention to reconcile.

A standardized patient identifier will, over the long term, reduce impediments to PIM, leading to improved quality of care and safety. However, at the outset, all risks must be identified and considered as a standardized approach to patient identifiers.

### **Recommendations**

**Privacy and Confidentiality.** When applying probabilistic matching, demographic and personally identifying data from multiple databases and systems must be collected and aggregated in order to reduce the error rate of record linking. As the amount of data increases, this process becomes progressively vulnerable to loss of privacy, thus requiring strong security measures to be in place along with good solid business processes. The healthcare sector should always consider the minimum amount of data that is required in order to maintain the integrity of this process while minimizing privacy impacts. Adequate security safeguards must be in place for assuring minimal risks to privacy.

**Data Security.** The matching process requires access to data from multiple information systems and resources. As the data move from system to system and potentially across settings in an HIE environment, the process becomes increasingly complex. In order to assure that patient confidentiality cannot be compromised, effective security controls and processes must be embedded in the workflow.

**Incident Management and Auditing.** There will need to be traceability (auditing) across systems, domains, and enterprises of patient records to ensure reliability of data and to provide a mechanism for responding to unauthorized uses and disclosures. The IHE Audit Trail and Node Authentication (ATNA) Integration Profile<sup>10</sup> should be adopted to achieve this level of auditing. The IHE ATNA standard has been selected by HITSP and implemented in the NHIN specifications.

### **Summary**

PIM standards serve to enforce predefined methods for connecting patient records by applying specific attributes, demographic data, terminology, and definitions to link records. PI Integrity is fundamental to improved patient safety, quality of care, and reduced healthcare costs. The characteristics of patient identity standards ensure integrity as well as availability of the data across disparate enterprises. The application of current standards in developing a patient identifier will significantly benefit business and process requirements such as automated e-health transactions, patient record mapping, use of patient personal and demographic data, governance standards over primary and secondary data uses, and security, privacy and confidentiality controls. These standards can be integrated into health information technologies such as EHRs to create trusted sustainable mechanisms for moving and sharing data at all levels of patient care.

Today, standards provide interoperability across communications mechanisms and enable broad use of health data to support such secondary uses as bio-surveillance, research, marketing, and public health. Furthermore, standards serve as the underlying foundation for efficient, cost effective healthcare by ensuring that, in accordance with the National Health IT Agenda:

- Medical information follows consumers so that they are at the center of their own care;
- Consumers are able to choose physicians and hospitals based on clinical performance results made available to them;
- Clinicians have a patient's complete medical history;
- Public health and bioterrorism surveillance can be seamlessly integrated into care; and,
- Clinical research can be accelerated and post-marketing surveillance will be expanded.<sup>11</sup>

Standardized PIM will also serve to increase the adoption of a single patient identifier solution by vendors and technology development organizations. Developers will be able to

---

<sup>10</sup> The Audit Trail and Node Authentication (ATNA) Integration Profile establishes security measures which, together with the Security Policy and Procedures, provide patient information confidentiality, data integrity, and user accountability. This environment is considered the Security Domain and can scale from a department, to enterprise or affinity domain.

<sup>11</sup> [Health IT](#), U.S. Department of Health and Human Services.

take advantage of reduced design and integration costs to incorporate patient identity functions into their products. Resources will be security-ready to meet the challenges of regional and national health information networks, providing a reliable model for healthcare enterprises to embrace and integrate into their IT infrastructure. Also, time constraints in bringing such technologies to bear would be significantly decreased. Technologies that leverage HITSP-approved standards and IHE profiles will become widely accepted as trusted for their improved workflows, the efficacy of security and privacy, and their sustainability as standards specifications evolve. The consistent application of standards across health information technologies will enable information sharing throughout the healthcare and public health environment—federal, state, local, public, private, and tribal.

## **Interfaces**

In the area of interfaces, major strides have been achieved to improve the manner in which computer systems share information. IHE promotes the coordinated use of established standards such as DICOM and HL7 to address specific clinical needs in support of optimal patient care. Systems developed in accordance with IHE communicate with one another better, are easier to implement, and enable care providers to use information more effectively.

IHE has three manager profiles that address patient identification: 1) Patient Identification Cross Reference, 2) Patient Demographic Query, and 3) Patient Administration Management. These profiles use relevant HL7 messages including HL7 V2 ADT messages Q22 and Q23 and for HL7 V3 the primary ones are PRPA 01, 03, 04, and 05.<sup>12</sup> The profiles give organizations options for advancing patient identification in a simple or complex healthcare environment.

- The *Patient Identifier Cross-referencing Integration Profile* (PIX) links patient information by cross-referencing identifiers linked to the same patient. Each time a patient registers, the PIX cross references the patient identifiers to determine whether an ID matches this patient. If a match occurs, PIX creates a link. From a technical perspective, the PIX profile does not define how records will be linked within a PIX Manager. It does define the rules for data ingest and ID query response. The PIX profile stipulates that the PIX manager will accept standard HL7 messages with patient identifiers and demographics, link the records as appropriate based on those demographics, and then fulfill any properly defined HL7 ID Query request. Organizations that employ Enterprise Master Patient Indexing (EMPI) can leverage PIX to link to the EMPI.
- The *Patient Demographics Query* (PDQ) lets applications query the central patient information server and retrieve the patient demographics and patient visit information. The PDQ profile provides for the dynamic query responses via HL7 messages. When patient identities are not known, demographic queries can be used

---

<sup>12</sup> HL7 assigns each unique attribute with an identifier. In this case, PR refers to the Practice subsection and PA to the Patient Administration domain.

to form a list of appropriate candidate patients. All queries and responses are defined within the profile and utilize the HL7 Q22 and Q23 defined requests.

- The Patient Administration Management (PAM) Profile is related to both the PIX and PDQ profiles. The PAM profile documents patient registration and encounter status. PAM can support multiple patient registration systems that collaborate as peers. The intent of this profile is to know “where” a patient is, was, or is going to be. Information about patient location is handled just like a PIX manager and queries about the patient are like the PDQ profile. Encounter Management information is additional demographic information carried with the patient. In most cases the PIX and PDQ profiles can be thought of as quasi-interactive interactions (i.e. query and response), while portions of the PAM profile are true publish/subscribe notifications. The exception is in support of the Patient Update Notification within PIX and PDQ, which mirrors the requirements of PAM. The Patient Synchronized Application (PSA) profile pulls patient information from multiple applications and populates user workstations. PSA can work in tandem with PIX to display information pertaining to a patient on a single screen.
- The *Basic Patient Privacy Consents* (BPPC) profile is a record of the privacy consents allowed by the patient. This information is crucial for those patients who transfer between environments and who have specific requests for information to remain private during their stay. This profile is undergoing additional review to assess its adequacy in meaningful HIE given the increased emphasis on data exchange.

This list should not be considered exhaustive. Information sharing requirements will need to be further examined as HIEs evolve.

On an annual basis, IHE tests the current IHE profiles in a controlled environment called Connectathons. Connectathons are held throughout the world including the United States, Europe, and Asia Pacific. Additionally, there are Interoperability Showcase events where vendors who have passed the Connectathons can be included in real-world demonstrations utilizing the use cases at events sponsored by HIMSS (in the United States, Asia Pacific, and Europe) and at the eHealth conference in Canada. Some clinical domains also hold special IHE demonstrations such as the Radiological Society of North America (RSNA), a worldwide radiology and imaging event held annually in Chicago.

IHE’s membership includes vendors, physicians, nurses, country delegations, and other industry standards development organizations (SDO) including HL7, ISO and Object Management Group (OMG). One of the “proof points” where HITSP Transaction Packages (TP) are first tested and certified is at IHE Connectathons. For example, many vendors who sign up to support PIX also sign up to test HITSP TP related to PIX. Therefore, a harmonization of the HITSP requirements and IHE profiles exist today with real proof that they work.

IHE does not have a formal relationship with the Office of the National Coordinator for Health Information Technology (ONC) or the NHIN activities. However, there is strong communication between these groups and activities through cross-member representation.

IHE will look to address provider profiles for identification, perhaps in the same manner they have addressed patient identification, with HITSP being in the driver's seat. Additionally, IHE will likely address more aspects of security and data exchange (and the associated patient identification) given the global emphasis on data sharing. Likely candidates for this focused work include Basic Patient Privacy Consent (BPPC) profile and the Audit Trail and Node Authentication (ATNA) profile.

### **Barriers**

The variation in standards is a significant issue that is being resolved through HITSP and other organizations. The pressure to move to a service-oriented architecture (SOA) or Web Services environment is ever increasing. IHE has quickly moved to support these environments with HL7 V3 and XDS.b profiles, but vendor "uptake" in released products still lags. New Web Service definitions from the OMG present challenges for true interoperability.

From an operational perspective the barriers include:

- Inadequate testing of the standards and profiles, both by the vendors and the healthcare organizations, which compromises all organizational goals.
- Organizations not following data standards or profiles as prescribed, thus having non-standardized data that inhibit data quality, data exchange, and interoperability.
- Defective data content (poor quality data or lack of data) to support robust patient identification, which is a foundation for the entire electronic record.
- Erroneous implementation of interface standards or profiles further confounding the business process of patient identity.
- Inadequate testing of implementation, thus allowing error to exist in the system. Assuming that because an ADT message is sent means that it was received is faulty thinking. There are too many variables that can impact the process, one being incompatibility of versions. For instance, ADT message sending and receiving capability may not be identical between two communicating systems. System Beta is sending an A34 merge message but System Alpha was purchased without that functionality. So the merge is never processed. Another scenario is a "kill" instruction that is legitimate 99 percent of the time also impacts the one percent when it should not apply but no one ever thought about it.

### **Recommendations**

- Organizations should use IHE and HL7 standards for patient identification, and require that their vendors demonstrate compliance with the patient identification standards through the most recent IHE Connectathons.
- Test, test, and retest. Do not assume that because a message was sent that it was received. Test thoroughly, repeatedly, and with numerous messages. Give careful consideration to the downstream impact of patient identification standards. Testing

the patient identification standards and profiles in the organization's environment should be conducted with a multi-disciplinary approach since so many different areas of an organization are impacted by patient identification.

- Ensure that information system employees are properly trained in understanding patient identity and their role in ensuring its integrity through creation of interfaces that transmit patient data between applications.
- Develop protocols and tools to ensure that information flowing across the organization arrives at the intended destination without error. Involve the business owners of the information in testing the veracity of the interfaced data to help ensure accurate information in all applications.

## Algorithms for Linking Records

The goal of the NHIN is to allow authorized users to quickly and accurately share health information in an effort to enhance patient safety and improve efficiency of the healthcare system. Achieving this goal is dependent on the ability to link individual health records.

There are basically two different approaches to linking records. The first involves the use of a UPI, and the second involves matching data based on a given set of demographic information. This statistical or mathematical matching technique requires the use of an algorithm. Simply defined, an algorithm is a procedure for solving a mathematical problem. In January 2006, the American Health Information Management Association (AHIMA) published a practice brief, *Surveying the RHIO Landscape: A Description of Current RHIO Models, with a Focus on Patient Identification*.<sup>13</sup> This paper describes the three most commonly used statistical matching tools in today's healthcare environment: basic, intermediate, and advanced. The paper also concludes that advanced record linking methods are far more accurate than other matching methods. Appendix B provides a more extensive analysis of record matching algorithms and their potential effectiveness.

In 2008, the RAND Corporation published a monograph, *IDENTITY CRISIS: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System*.<sup>14</sup> This report advocates for the adoption of a UPI as a means to improve the accuracy of record linking, yet also recognizes the barriers to adoption and acknowledges that statistical matching will be required for the foreseeable future. That is, the use of statistical matching would persist during a transition period to a mandated UPI solution, and statistical matching would also be necessary in the event of a voluntary UPI. Lack of a unique, invariable, and ubiquitous healthcare identifier requires an algorithm to perform matching based on patient demographic data. Algorithms would also be required for historical databases created prior to the implementation of the unique identifier.

---

<sup>13</sup> E-HIM Work Group on Patient Identification in RHIOs. "Surveying the RHIO Landscape: A Description of Current RHIO Models, with a Focus on Patient Identification." *Journal of AHIMA* 77, no.1 (January 2006): 64A-D.

<sup>14</sup> 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 Monograph. October 2008, No. 753. Available at [http://www.rand.org/pubs/monographs/2008/RAND\\_MG753.pdf](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf). Accessed on Aug. 26, 2009.

The lack of standardization of data sets, a standard field or attribute definition and limitations of statistical matching methods contribute to the challenges associated with record linking. Without national standards, the marketplace has generally adopted the expedient approach of using the same statistical matching methods that most large healthcare provider systems have used in their MPI systems for years.<sup>15</sup> This means that most health information systems utilize only deterministic or exact match solutions on partial sets of data fields. These matching capabilities only identify the record for which the user is searching 85 - 92 percent of the time at best. Reports have documented that 30 percent of the time, physicians could not locate previous results on their patients.<sup>16</sup> This is a pandemic problem that only gets worse as the databases get larger, as demonstrated in the RAND report.

The vast majority of information systems utilize basic or deterministic record matching methods, which rely on exact or partial matches of key demographic data elements. Some systems have been enhanced with rules engines to improve the performance of the basic record matching methodology. However, the limited research that has been performed clearly demonstrates that advanced mathematical tools produce the best matching results. In addition to the most commonly known probabilistic or frequency-adjusted models, advanced mathematical algorithms include Bayesian pattern recognition, bi-partite graph matching, machine learning, and neural networks. These advanced systems appear to significantly improve record matching capabilities. Scientific studies have not been performed to determine which of these methods most effectively reduces false positives and false negatives, thereby producing the most accurate and reliable patient matching.

### **Barriers**

Accurate record linking is an important key to the success of interoperability, yet several barriers exist to meeting this challenge:

- **Limited understanding of challenges in linking data.** Both people and processes impact the performance of technology. Generally speaking, record linkage has been viewed as a simple technical challenge solved by computer algorithms without a good understanding of how the algorithm's performance is impacted by the people and business processes associated with creating, identifying, and resolving discrepancies and errors in demographic data and the non-standard processes for collecting those data. The algorithms performance is also dependent on the particular characteristics of the population being matched (e.g., a first name of Jose is more discriminating in a city in Wisconsin than in one on the Mexican border.)
- **Limited understanding of impact on operations, patient safety, and financials.** Until increased automation highlighted the record linkage problems by removing the clerical "middle-man," most users were not aware of the numbers of false positives and false negatives present in the patient indices of the hospital information systems.

---

<sup>15</sup> Ibid, p. 40.

<sup>16</sup> Electronic Medical Records – Getting it Right and Going to Scale. W. Edward Hammond, II, PhD. [Commonwealth Fund](#) background paper, Jan. 2004, #695.

These errors create patient safety risks because of incomplete or inaccurate data, and also contribute to operational inefficiency, billing errors/payment delays, and poor adoption of new technology.

- **Dependence of algorithms on data elements being present and accurate.** Statistical matching is highly dependent on key demographic data elements being present, accurate, and consistent on all records. Variations in data and missing data reduce the accuracy of patient matching.
- **Lack of patient involvement in validating their demographic data.** Use of automated algorithmic record matching without the patient's ability to verify the accuracy of the record matching creates significant opportunity for false positive matches to occur. This leads to substantial patient safety risks.
- **Lack of scientific studies related to effectiveness of algorithms.** Evidence regarding the performance of patient matching algorithms is anecdotal, with few scientific studies to evaluate algorithm effectiveness and accuracy. This is particularly true if one is trying to understand the tradeoffs between false positive and false negative error rates as the algorithm is being 'tuned.'
- **Lack of definitions standards.** Standards need to be defined for data elements, algorithms, and record matching requirements within health data exchanges.
- **Lack of standard methods for computing a static duplicate rate in an MPI database and for calculating duplicate creation rates.** The first such industry standard for this computation was published in a practice brief titled *Managing the Integrity of Patient Identity in Health Information Exchange* by AHIMA in July 2009.<sup>17</sup>
- **Collecting and entering demographic data.** Demographic data utilized by statistical matching methods is traditionally collected and entered into the information systems by the lowest paid and least trained employees. By the nature of the data, errors in names, addresses, phone numbers, etc. cannot always be readily detected at the time of data entry.
- **Correction of errors.** The healthcare system assigns a low priority to correcting data errors and preventing risk associated with fragmented or incorrectly linked patient records. The prevalent view seems to be that some level of error is acceptable and occasional adverse outcomes are less costly than correction or prevention. Low accuracy is unacceptable in aviation, but healthcare tolerates errors and creates work-around solutions for data quality and process issues.
- **Funding.** The cost may be high to study algorithm effectiveness, create standards, and implement change. Funding must be allocated to these initiatives.

### Recommendations for Algorithms

Several actions are recommended to address the accuracy and effectiveness of record linking methodologies:

---

<sup>17</sup> AHIMA. "Managing the Integrity of Patient Identity in Health Information Exchange." *Journal of AHIMA* 80-7 (July 2009): 62-69.

**1) Short Term (within the next five years):**

- Create data definitions (data dictionary) for all key demographic data fields utilized in record matching algorithms to facilitate consistency across providers in the collection of these data fields. Recommend minimum data elements to be utilized in record matching algorithms.
- Perform a research study to validate algorithm effectiveness for electronically linking patient records.
- Adopt an industry standard method of computing duplicate record rates in MPI databases and a standard formula for computing the “creation” rate of newly created duplicate records.
- Provide industry guidance on the process that providers and health information exchange organizations should follow to resolve potential duplicate records within their database. Guidance on staff education and experience requirements of individuals who are capable of monitoring potential duplicate records and resolving them is needed.

**2) Intermediate (five to ten years):**

- Using study results, recommend algorithm standards including search threshold minimums and record auto-linking minimums.
- Using study results, provide industry standards for maximum duplicate record rates.
- Using study results, improve tools and industry standards for IT systems.

**3) Long Term (ten years and beyond):**

- Adoption of a patient identifier solution.
- Reduce dependence on algorithms.

## **Unique Identifiers**

There is a very simple and straightforward way to avoid the false positive and false negative errors associated with probabilistic demographic matching techniques: assign a unique identifier to each person and enable them to use it across all of their healthcare encounters. As long as the identifier is reliably unique and it is not corrupted by mistakes, this approach will eliminate both false positive and false negative errors in patient identification. In fact, this approach is frequently taken today by various electronic healthcare systems. Once the system has identified the individual, it assigns some internal identifier and then uses that identifier for its operations. However, each clinical system does this on its own and the patient is never provided with an externalized identifier they can use across systems.

Unfortunately, it has proven to be very difficult to implement the simple unique identifier solution as an approach to patient identification in the United States.<sup>18</sup> Concerns about the cost to implement such a system, privacy concerns about the huge demographic database

---

<sup>18</sup> Toward Health Information Liquidity: Realization of Better, More Efficient Care from the Free Flow of Health Information, S. Penfield, K Anderson, M Edmund, M Belanger, Booz Allen Hamilton, January, 2009.

needed to support its operation, technical issues about how to retrofit existing automation systems to use the new identifier, and lack of a national consensus about what the identifier should be have all combined to make it extremely difficult to progress on this issue for well over two decades.

Of all these barriers, the issue of privacy has been perhaps most problematic. Privacy advocates have long and correctly argued that the creation of a massive centralized database containing patient identity, patient demographic, and/or patient clinical information represents an unacceptable privacy risk. Unfortunately, this has led to an “either/or” mentality. *Either* we protect patient privacy by drastically limiting healthcare information sharing *or* we promote healthcare information sharing by sacrificing patient control over the privacy of their information. This is believed to be a false choice. It is possible to put in place mechanisms, procedures, software, and policies that permit balancing these two needs without sacrificing either one. Fortunately, options are now emerging which enable the accomplishment of this goal.

Recently, a new project has been launched to create a Voluntary Universal Healthcare Identifier (VUHID) system (see Appendix A). The VUHID project takes a radically different approach to the creation of a universal healthcare identifier and, as a result, has the potential to avoid all of the barriers noted above. It is too early to know whether the VUHID project, or any similar project, will succeed in the healthcare market but unique identifiers have a number of potential advantages that make them highly desirable.

### **Unique Identifier Strengths**

1. **Linked to a person:** Unique identifiers can be linked to a person independent of the healthcare organization that issued the identifier. Any identifier has validity only within the domain of the organization that issues it. But people are mobile and may often require healthcare services across multiple organizations and even multiple countries. Since a personal identifier travels with the person, it can be valid no matter where that person requires medical care.
2. **Unambiguous identification:** Individual identifiers can be made globally unique so that a particular identifier is only used once to identify one person. An individual can thus travel the world using their identifier without worrying that someone else might have the identical number. In addition, clinical automation systems can use the identifier as the ‘key’ to link clinical information on that individual without having to worry whether the key will conflict with any other person’s key in their system. In the VUHID system, each individual’s identifier is globally unique.
3. **Error detection:** This can be an integral part of capturing a person’s identifier. The check digits included in a modern identifier make it possible to detect common errors such as transposition of digits or misreading a digit. A person who improperly enters an identifier can be notified immediately so that the error can be caught and corrected rather than being entered into the system, thus requiring (expensive) later remediation. Also, identifiers can be represented in a machine-readable form that virtually eliminates errors in data entry.

4. **Incorporation capability:** Because the identifier represents a single data element, it can be incorporated into other processes and data structures to improve integrity. It can be encrypted in circumstances where that is necessary and it can be included in authentication processes where that is appropriate.
5. **Immutable over time:** Because the identifier is an abstract entity that does not depend on a person's personal characteristics, it does not vary over time. If the person moves to a new address, gets married and changes their name, gets a job with a new employer, has their name changed through a legal process, or changes their telephone number, there is no impact on their unique identifier. In contrast, each of these situations leads to a need to re-establish patient identity when using demographic matching procedures.
6. **Consistent syntax:** The syntax of a standardized identifier remains consistent across time and across various automation systems. This means that healthcare automation systems only need to add the definition of a unique healthcare identifier once and they are then able to process any individual identifier without further syntax modifications to their software.
7. **Cost effective:** Because of the simplicity and consistency of a standardized healthcare identifier, it is very cost effective to modify healthcare automation systems to incorporate them into use. In the case of the VUHID proposal, it is not even necessary for most automation systems to be modified because the system's existing person identifiers can be cross-mapped to their personal healthcare identifier by the EMPI.
8. **Counterfeit resistant:** Another identification issue is the ability to avoid counterfeiting. It is possible to equip a unique identifier with mechanisms to prevent a person from being able to successfully 'make up' a new identifier and have it accepted as being valid. This ability to avoid counterfeit identifiers is important to avoid the possibility of some outside agency maliciously 'making up' a series of healthcare identifiers for financial gain, for example through healthcare fraud.
9. **Cultural independence:** Names may be difficult to represent in a different language, which may lead to problems trying to establish a person's identity using demographic matching. Numbers, however, are virtually universal, thus making a numeric-based unique healthcare identifier much more usable and robust for people who may travel to foreign countries or who have unusual names.
10. **Simplicity:** The unique identifier approach makes identifying a person simple. This simplicity inherently makes the identification process less prone to errors and it also dramatically reduces the associated costs.

## Unique Identifier Weaknesses

1. **Privacy risks:** Contemporary attempts to establish a unique identifier require a centralized database of patient-specific identification information represents a privacy risk.
2. **Prohibitive costs:** If implementing a unique identifier requires retrofitting all of the existing clinical automation systems in the United States, it represents an investment in terms of cost, time, and complexity.
3. **Data security risks:** Gaining access to just one data element, the unique identifier, can in theory provide access to a patient's entire medical record.
4. **Patient education:** Patients will be responsible for managing their healthcare identifiers and for actively managing the privacy of their clinical information. Not all patients will have the knowledge, capability, motivation, or desire to assume this responsibility. Even those highly motivated will lack the knowledge of where these identifiers could be used.

In summary, unique identifiers offer a number of substantial benefits to help address PI Integrity problems but still have significant challenges. They are only part of a total solution. New approaches to identifiers can eliminate the vast majority of the prior barriers to creating a UPI system. Removing such barriers makes it possible that unique identifiers can be successfully deployed as a part of an overall PI Integrity solution in the coming years, with substantial consequent benefits for all of healthcare.

## Recommendations for Unique Identifiers

### 1) Short Term (within the next five years):

- Secretary of HHS, under the direction of the U.S. Congress, should establish an informed patient identity solution. As part of this solution, steps need to include:
  - Congressional lifting of the prohibition against HHS studying UI solutions;
  - HHS conducting a study of the cost/benefit and practicality of implementing a UI solution; and,
  - HHS establishing pilot implementations of unique identifiers to document the challenges and benefits.”<sup>19</sup>

### 2) Intermediate (five to ten years):

- Use the results of work above to determine implementation process and timeline for unique identifiers as component of patient identity solution.

### 3) Long Term (ten years and beyond):

- Implement unique identifiers as component of patient identity solution.

---

<sup>19</sup> [2009 HIMSS Principles on Government Initiatives.](#)

## **Business Processes**

The business processes that most impact accurate patient identification start with patient registration or patient access, which for some organizations includes patient scheduling. Whether an organization has centralized or decentralized patient intake, this is the area that interfaces first with a patient or customer and gathers the information that builds the patient's identity and the associated medical record number. This customer facing area is currently undergoing considerable process reengineering in many organizations due not only to the installation of patient kiosks and internet portals, but also a higher sensitivity to the "business" of healthcare. There is a new recognition of the patient as a customer, and organizations are seeking to provide a positive customer interaction starting with the appropriate staff, tools, and processes at the first point of communication.

In fact, poor processes cannot be offset by the best of technology. Failure to establish accurate patient identity at the front end means all subsequent business processes are potentially hindered due to compromised patient identification. Establishing an incorrect medical record number at the onset means all systems and associated data will be compromised and ultimately require correction. This is an expensive, time consuming proposition in today's electronic world. The group most likely charged with corrective action when it is required is the Health Information Management Department, with assistance from Patient Registration/Access and Information Services.

## **Barriers**

The most common causes of process failure that lead to inaccurate patient identification are:

1. Lack of adequate "tools" or software to make patient identification easier
2. Inattention to detail by the registration/intake/scheduling staff
3. Poor or absent training (initial and ongoing) of the involved staff
4. Lack of understanding how an inaccurately assigned medical record number impacts all associated electronic data
5. Lack of corporate or organization commitment to accurate patient identification

The most common barriers to effective business processes include lack of standardized ongoing training, inadequate compensation of registration staff, and a lack of organizational awareness of the importance of patient identification to the capture, sharing, and dissemination of patient data. Organizational awareness can be exacerbated by departmental barriers that are raised, as opposed to a collaborative, multi-stakeholder approach to patient identification.

## **Recommendations for Business Processes**

Best practices start with an organization's commitment to accurate patient identification that includes a quantifiable expectation or performance standard for accuracy of medical record number assignment. Executive level support for a multi-stakeholder administrative group that identifies problem areas, monitors relevant data, prioritizes corrective actions, and practices structured, ongoing communication is the foundation. This group should include

at minimum representation from patient registration, patient finance, information systems, and health information management departments. Emphasis for this group must include ensuring adequate tools and resources, and process support, including initial and ongoing competency-based training.

Additionally, explicit organizational guidelines for data stewardship and data governance that reach beyond patient identification are critical to the success of patient identification processes. Data governance is the organization's management of its business data, which includes data quality, data management, business process management, and risk management.

Data stewardship defines the persons or "owners" of data and the responsibility for the quality of the business data. This requires a level of domain expertise and a clear understanding of the impact of inaccurate patient identification. The governance and stewardship discussions, as well as policy development and execution, must involve a multi-stakeholder group that has executive level involvement and support. Ongoing training and administration of all intake and scheduling areas should include appropriate compensation, recognition, and professional certification.

Corrective action required by ineffective business processes or human error should have a multi-disciplinary approach. While the approach for the corrective action may be centralized or decentralized, clear ownership must be established. And, standardized business processes for corrective action, irrespective of the approach, are crucial—lest even greater errors be created.<sup>20</sup>

Reports are key business practices that should be designed to support the monitoring of data accuracy and staff performance. Measurement of patient identification accuracy should be through standard reports structured to monitor:

1. Accuracy of patient identification by the responsible registrar
2. Identification of contributing factors to the error
3. Trending of the data by data entry person, (e.g. registrar), their location (e.g. registration area), and organization

These data should be reported to appropriate management personnel, involved staff, and the multi-stakeholder group. Registrar staff not responding to remedial training should be subject to progressive disciplinary action.

With a solid corporate and multi-stakeholder departmental commitment to accurate patient identification, most risk areas can be identified and collaboratively resolved.

---

<sup>20</sup> Ibid

## **Data Accuracy**

The efficacy of a probabilistic matching system is dependent on the accuracy of the data utilized for both locating and linking patient information. Inefficiencies or errors in the data collection, data entry, or data query can lead to faulty results. In the presence of a clean database, searching an organization's information system with a unique identifier such as a medical or health record number can bypass the query process and be a more effective way to locate an individual within the information system, yet a query of the patient's demographic data is still the most common method used at healthcare entities when the patient is speaking directly with the intake staff.

As a result, data accuracy is critical in three different instances in order to maintain its integrity: the data must be collected correctly, the data must be entered correctly, and the data must be queried correctly. If errors occur during any of these three instances, then the integrity of the organization's data has been compromised. The larger the organization, the greater the number of instances, and there is a proportionately increased opportunity for errors. This only multiplies when the organization becomes part of a larger network or incorporates other entities' information into its own. The need for establishing an ongoing program to measure and monitor data integrity quickly becomes a prudent risk management decision.

There are primarily two types of errors that can occur when using any matching system. The first type is a data identity error. These errors in the actual identification process arise from the act of matching the information within a record to the query information. Examples of data identity errors would be intake staff selecting the wrong patient from a list of potential matches or typing the wrong spelling of a name into the system when performing the initial query.

The second type of error is data discrepancies or errors in the actual data itself. Examples of data discrepancies might be a name that is spelled incorrectly in a system or a date of birth that is transposed. Either of these error types will lead to compromise of data integrity. Tracking and measuring the rate at which these reductions in accuracy manifest themselves will allow organizations to better assess how well they are adding, identifying, and matching patients within their systems.

Accuracy in an EMPI system should be measured with several different sets of metrics. Since each calculation looks at a different aspect of data integrity, a single measurement will not provide an adequate idea of how well an organization is managing their processes and data integrity. Two key metrics for an EMPI's accuracy are static (snapshot) error rate and creation error rate.

The static error rate is a raw calculation of how many times errors occur in an EMPI system. Calculating a static error rate usually requires dedicated staffing resources and/or an automated analysis system to review the database for potential duplicates. A duplicate is defined as two (or more) individual records existing separately in the database for the same person. The static error rate is effectively an organization's raw error rate for their EMPI; it

includes any errors created since the inception of the EMPI. This number has great value as a baseline measurement of a system's level of accuracy.

An arguably more relevant calculation for data accuracy is a creation error rate. A creation error rate takes a static error rate and analyzes it over time. The creation error rate would be an organization's calculation of the number of new errors created last week, month, or year. A calculation of a creation error rate as part of the administrative dashboard allows management to monitor the number of duplicates created in real time. This provides the organization with a much needed indicator to assess changes in their data accuracy.

Increases in an error creation rate might alert management to an issue that needs to be corrected, new staff that needs additional training, or a system (i.e., algorithm, interface) change that is causing more false negatives to enter the system. Decreases in an error creation rate might indicate that a policy or system change has resulted in more accurate data queries or searches. Robust reporting mechanisms can provide error creation rates for an entire network, an entity within the network, a department, or even an individual. The individual error rate can be used as a metric for employee evaluations or incentives to help drive performance.

One of the most common measures of how well a data matching process works is the calculation of a duplicate rate. Duplicates occur when data discrepancies upon intake allow for the creation of several different data records for the same individual. These are often referred to as "false negatives" when discussing algorithm accuracy. Duplicates lead to fragmented data and an incomplete clinical picture for the patient. A simple example would be a patient named Jon Smith who registered four years ago and has not returned to the organization since that initial visit. Mr. Smith arrives today and is registered under the name John Smith, thus creating a duplicate record.

Due to the limitations of automated systems, it is common practice to audit potential duplicate results and validate them before merging records. As a result, a calculation of potential duplicates will yield information on the sensitivity of a system's matching algorithm. The confirmed duplicates will yield information on an organization's ability to properly query the system and make an appropriate selection.

To demonstrate how these calculations intersect, a fictional hospital, Baystate General, has 4.3 million records in their EMPI system. A basic duplicate checker would compare each record against all the other records in the EMPI based on name, date of birth (DOB), and SSN. Any records that matched on two of those three criteria would be returned as a potential duplicate. The duplicate checker returned 250,000 potential duplicates for Baystate General, thus the static error rate for potential duplicates at Baystate General is 5.8 percent ( $250,000/4,300,000$ ).

If the potential duplicates were aggregated over time, based on the date the second (duplicate) record was created, this would show the organization their creation rate. A historical creation rate is useful for pinpointing a timeframe for the root cause of issues. For example, Baystate General Hospital had merged with Young Hospital three years ago and at

that time, their respective MPIs were combined. The historical creation rate report for Baystate General shows that there was a large spike in duplicate records created at the time of the merger. This would indicate that there were several patients in both MPIs that were not combined when the organizations came together. Process changes could then take place so the next time Baystate General combined their MPI with another organization, this issue would be addressed at the time of the organization merger.

The correlating measure of data matching accuracy is the rarer and arguably more serious issue of an overlaid record. Overlays occur when an error is made in the identification process that leads to a patient being incorrectly identified as another person. A common example of an overlay would be twins identified as the same person. A particular challenge with this type of issue is identifying its existence. Another example would be patients fraudulently using another person's identification data to secure services. It is often not noticed until a discrepancy in the clinical or billing record is detected.

Another accuracy measure that can offer value for an organization is an analysis of the organization's core data changes. Core data changes are an organization's "near misses." Many matching algorithms rely on static data elements to provide the stability needed to perform a match. These include first and last name (generally, last name is considered static for men), date of birth (DOB), full or part of a SSN, and gender. If these data elements are being changed in a system, this is usually an indicator that there is an issue with data collection at the initial intake or a record is being incorrectly updated at the time of the edit, which can lead to an overlay. Either of these situations is a danger to the data stability and accuracy of an EMPI. An analysis of core data changes offers the opportunity to review data collection accuracy and can reveal valuable information regarding current core data collection processes and help identify data collection problems with particular intake sites or individuals.

Core data change tracking requires a cataloguing of changes to the core matching elements in an organization. For example, if a facility matches on first name, last name, gender, DOB, and SSN, these would be the core elements that would be monitored for changes. Core data changes offer the opportunity to examine both the raw data numbers and the arguably more relevant trend data. If an enterprise finds that data items such as first name or DOB are consistently being changed in their system, this would indicate an opportunity for further training or standardization of data entry processes.

The organizational structure, workflow, and information technology available in healthcare lead to some inherent vulnerabilities in maintaining data accuracy. Inequalities in training of new staff and pressure on staff to process a greater number of registrations in less than optimal environments combine to create inconsistencies in the intake processes. Variability in the method of entering common data elements such as numeric and directional street names, suffixes, and joined names, will inevitably begin to occur. In addition to variability, evolving societal changes in traditional definitions of familial relationships such as same sex marriage, single parents, etc. add a further level of complexity. Additionally, many enterprises do not employ, require, and enforce a consistent methodology for both the querying of information from patients and the manner of data entry. Some areas

dangerously rely on their familiarity with patients, which can lead to assumptions regarding a patient's identity and can lead to a lost opportunity to verify and potentially update key demographic information.

Furthermore, human error is inevitable and even with completely stable data, keystroke errors will occur in the data entry process. However, the data collection requirements for many healthcare organizations are numerous and many of these elements change frequently. In our modern society, many data elements that were relatively stable a few decades ago, such as marital status, employer, or even address, now change on a much more frequent basis. According to the U.S. Census Bureau, 14.9 percent of Americans move to a new address each year.<sup>21</sup> Moreover, there are more than 2,230,000 marriages and 957,200 divorces that more often than not result in demographic changes.<sup>22</sup> These frequent social changes multiply the need for data edits and updates, thus increasing the risk for an error.

Many organizations have attempted to address the accuracy of the data entered into their system by creating "specialist" intake of data. As health insurance information became more complicated and difficult to collect, hospitals and networks developed call centers filled with data entry specialists whose primary focus is making certain all essential data are collected and the appropriate pre-appointment insurance approvals are in place. However, decentralized data entry sites with "specialists" are not immune from accuracy concerns. Maintaining operational consistency among staff that may be located in different cities, states, or even countries is a particular challenge. Moreover, decentralizing data collection deemphasizes its importance to those who are interacting with the patient face-to-face, thus often missing a golden opportunity to properly correct a data inaccuracy or provide clarification to partially completed data collection. The key to any decentralized registration site is making certain an easy communication conduit is set up between the points of service and the data entry specialist and that the point of service staff remains motivated to refer registrants to the call center whenever appropriate.

The inability of many information systems to offer a robust reporting or auditing functionality can be a significant barrier to data integrity as well. These inadequacies can only be addressed by allocating additional personnel resources, seeking pricey outside assistance, or purchasing expensive system add-ons to handle the work that a robust auditing functionality can provide.

### **Recommendations for Data Accuracy**

One of the most effective methods to ensure data accuracy is to create standardized procedures for the collection, input, and query of information. As previously stated, data accuracy must occur in each of these three business processes in order to maintain data integrity. Data collection should be done with the patient present whenever possible to allow for any follow up questions or data validation. Staff should be trained to verify data

---

<sup>21</sup> "Why People Move: Exploring the March 2000 Current Population Survey". Jason Schacter. US Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau Special Report. May 2001.

<sup>22</sup> [www.cdc.gov/nchs/fastats/divorce.htm](http://www.cdc.gov/nchs/fastats/divorce.htm)

elements before entering them to prevent assumptions. If the patient is available, staff should provide the opportunity for the patient to review the data entered before it is accepted into the EMPI. This allows the opportunity for correction: "Anne Marie is one word, not two" or "It's Macdonald, not Mc Donald." Whenever verifying data elements either at the initial collection or when querying the EMPI for a match, patients should be asked to answer all questions directly regarding the data elements. In other words, never state, "Is your date of birth 01/01/01" or "Do you still live at 123 Main Street in Anytown, USA." Rather, ask patients, "Can you please tell me your date of birth" or "Can you please provide me with your full address."

If data collection is handled via a form (either distributed in advance to patients or a form completed at the site of care) forms should be created with a "box per letter" format to ensure that each letter is separately written. All elements on intake forms should be finite (i.e., month of birth, day of birth, year of birth) and provide examples of proper completions, such as "MM/DD/YYYY." It is best to avoid asking patients to complete certain non-demographic questions that could cause confusion, such as insurance plan number, insurance group number, or guarantor. These types of questions often lead to inaccurate data collection and are best collected by reviewing documentation or obtaining information through an insurance verification system.

The utilization of look-up entries as opposed to free text entries and rules based data validation (e.g., zip code to city validation) should be incorporated in data input methodologies whenever possible. This reduces the opportunities for data entry errors as well as creates data entry checks within the intake process.

Ideally, the best way to ensure data accuracy is to minimize the number of times the information should be collected, input, and queried. The combination of utilizing a unique identifier for queries and using best practice data verification measures for the collection and input of updates will minimize the chance of data errors.

## **Data Quality**

The accuracy of a probabilistic matching algorithm is increased by the number of data elements that can be used in the search. The amount of information that can potentially be collected on an initial admission or encounter is rich with data. These data can be grouped into a hierarchical pyramid<sup>23</sup> that provides a categorical breakdown of personal information. The higher levels of the pyramid are associated with more effective identification. The variety of information used by healthcare systems can be grouped into four sections.

---

<sup>23</sup> Clarke, R. Human Identification in Information Systems: Management Challenges and Public Policy Issues. *Information Technology & People* 7, 4 (1994) 6-37.

- **Names** - what the person is called by others
- **Knowledge** - what the person knows
- **Codes or Tokens** - what the person is called or given by an organization,
- **Biometrics** - appearance (e.g., gender, eye color), bio dynamics (e.g., signature), and natural physiography (e.g., fingerprints, iris scans, DNA).

For example, data collection for a new patient to an enterprise might involve the capture of demographic information (names and knowledge), linking that to a health record number (codes) and giving the patient an identification card (token) that might contain a photo and/or a signature (biometrics). The patient would also need to sign some consent forms (biometrics) that would be filed for potential comparison at a future date and provide support documents (tokens) to verify their information.

Little of that data is usually readily available when doing a query for the patient upon their return to an enterprise. The patient may forget their ID card or the enterprise may not provide one, appearances change over time, the signature on file is not captured electronically or is not available for comparison. Most enterprises are resigned to searching by the health record number or a patient's demographic information.

As indicated above, many organizations capture a large amount of data but very little of the captured data is actually used in the matching search. The industry standard for required data elements in a probabilistic search includes full name, gender, and DOB.<sup>24</sup> Elements such as SSN, address, phone number and mother's maiden name are often viewed as very helpful to a search, but optional.

Additionally, there are very few truly static identifiers that can be used to match two individuals. Two factors have created a pendulum shift in data collection: the discouraged and largely discontinued use of SSN as a required field and the reluctance of some organizations to require other identifiers (e.g., mother's maiden name) The limited medical record data collected for identification in the early 1900s is very similar to the information that is being captured today, yet the pool of potential matches has increased significantly.

The importance of having a breadth of data elements to examine has been stressed in the algorithm section (see Table 1). These results were also demonstrated in a study performed by the RAND Corporation of 80 million records in a SSN death registry database.<sup>25</sup> It is important to note that using the Death Registry creates an almost ideal dataset for this experiment. The parameters of this database assume that there are no duplicates in the data set and that the values will remain constant. An individual's last name and/or zip code will

---

<sup>24</sup> Fernandes, Lorraine and O'Connor, Michele. "Future of Patient Identification." *Journal of AHIMA* 77-1 (January 2006): 36-40.

<sup>25</sup> 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 Monograph. October 2008, No. 753. Available at [http://www.rand.org/pubs/monographs/2008/RAND\\_MG753.pdf](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf). Accessed on Aug. 26, 2009.

not change after being entered in the registry and the same person will not expire more than once.

The study utilized different combinations of data elements to create a composite key for matching. These included first name, last name, DOB, last 4 digits of SSN, zip code, and birth year. The data showed that by using a combination of any of the elements except the SSN, the best false positive rate that could be achieved was 1 in 80,000 (using first and last name, DOB, and zip). By adding the SSN element, the false positive rate drops to 1 in 39,000,000.

Conversely, by removing the zip code, this number drops to a false positive rate of 1 in 640,000. While this is eight times better than any accuracy rate achieved without an SSN, it becomes obvious that in order to achieve a satisfactory identification, several different pieces of data, including a relatively unique data element, needed to be added to the query.

Data captured for a probabilistic match must be distinct enough to provide value to the match, yet not revealing enough to pose a threat to the data security of the individual. There are data elements that allow quality data to be captured while minimizing risk to data accuracy. One element previously discussed was the last 4 digits of an individual's SSN or another unique PIN. Another is creating an amalgam identifier from different elements of the various categories of identity information.

As previously mentioned, identity information has several levels. One idea for creating a match is combining a variety of those elements to create an amalgam identifier. This identifier could be searched on its own or be recreated by entering the specific data elements. For example, some state driver's licenses are coded based on the resident's name, date of birth, gender, and eye color. These elements are combined to create an amalgam ID that can also be recreated using the separate elements. In addition, these elements can be verified using an outside document for data security.

For example, using a very simple fictitious amalgam code, patient Sylvia Black comes to the hospital. She is a female with a date of birth of 5/12/1961; the last four digits of her SSN are 7591. Using these data elements, Baystate General Hospital could create a health record number for her (S25-6591-6112). This alpha numeric code utilizes the first initial of Sylvia's name (S), a numeric code for the second letter (B=25), a code for her gender (Female=6), the last three digits of her SSN (591), and the year and day of her birth (61 and 12, respectively). More sophisticated systems would be able to run an algorithm on each element to create a more intricate key.

This type of identifier provides a high level of inherent value. The identity information levels run the spectrum from names to biometrics (name, known, biometric) and the information is largely static. In addition, based on the data elements collected, a registration clerk would be able to query the database and retrieve the patient's health record number with a remote probability of false positives.

Smart Cards, which bypass the demographic search and provide a one-to-one retrieval of demographics, are another tool that, when implemented properly, can help improve data quality. A photo on the card or another verification check needs to occur separately as a back-up to ensure the identity of the presenting individual. The smart card itself or the additional verification step needs to have a safeguard in place that prevents someone from simply finding an individual's card on the street and using it to seek treatment or perpetrate medical identity theft. The information pulled from the organization's information system when a smart card is used for identification must still be confirmed with the presenter. Reliance on smart cards for identification without frequently confirming the data behind it can lead to missed opportunities to correct or verify the accuracy of data.

A particular challenge with data quality comes as a result of the import or export of one organization's MPI to another. The recipient is often at the mercy of the quality of data from the sending organization. This is a particular challenge for HIEs as they struggle with definitively matching patients seen at multiple entities within the HIE. Many organizations do not have the resources to perform the necessary pre-conversion cleanup on the incoming data prior to the initial conversion of a new MPI to the EMPI as a whole, but this cleanup is essential. An abundance of vendors have powerful cleanup tools to assist with these imports and exports. Utilizing a vendor for this work helps avoid the strain on existing staff that results from these onetime events and provides industry expertise on a specialized topic that is not easily understood by temporary employees. Unfortunately, at many organizations, the decisions around acquiring or merging with another entity are made at a management level and too often, these decisions do not consider the MPI cleanup costs (either handled by a vendor or handled in-house) that should be factored into the acquisition decisions.

Despite strict adherence to all best practices regarding data collection, entry, and query, a number of significant barriers still exist. Identity thieves or individuals intent on committing fraud will relentlessly search for ways to continue their criminal behavior. Patients intentionally seeking to hide their medical history will continue to seek ways to avoid identification in a hospital's MPI. A narcotic addicted patient can arrive at a hospital and claim they have never been to the entity previously and provide bogus registration information, then try to deceive clinicians into providing prescriptions for narcotics. The same individual could arrive at that same hospital and claim again they have not been a patient previously and create a new identity, seek treatment again and attempt to obtain additional narcotics. To stop these potential fraud incidents some entities are now turning to biometrics as a more certain method for identifying their patients.<sup>26</sup>

Biometrics are technologies that capture, store, and automatically confirm an individual's identity by comparing patterns of physical or behavioral characteristics in real time against enrolled computer records. The use of automated biometric technologies has increased rapidly over the last few decades, in the wake of international terrorism and financial fraud, and corresponding with advances in computer processing technology. It is not uncommon now in Japan for bank customers to approach an ATM and gain access to their accounts only after being subjected to a scan of their palm to determine their hand's unique deep vein

---

<sup>26</sup> Lawrence, Daphne. "Who Are You"? *Healthcare Informatics* 2009 March; 26(3):32-35.

pattern. The same technology has been piloted in U.S. healthcare organizations, although there is currently no widespread deployment. The key to making this technology work properly is making certain the initial scan of the biometric is appropriately assigned to the correct individual. Biometrics can also be cost prohibitive and make implementation across a large network with many intake sites too expensive. In addition, improvements in the level of security protections, as well as the accuracy and sensitivity of the technology, need to occur before acceptance in the United States becomes more widespread.

Another possibility to address similar fraud situations is to employ identity verification services through an outside vendor. These verification systems tend to take advantage of publicly available information and charge users for querying the database. The systems can often be cumbersome but more and more vendors are forming partnerships with these services and offering them as an add-on or benefit to the MPI system they are selling. With great power comes great responsibility, so it is important that access to these services be properly monitored or it could lead to staff using information from these systems to steal identities or commit fraud.

Additional barriers revolve around patients who do not possess some of the key data elements the organization attempts to collect. Immigrant patients often do not possess a SSN and in some instances have no government-issued document to assist with verifying their identity. Certain biometrics cannot be collected on some patients (e.g., amputees cannot provide a fingerprint). It is best to have policies and procedures in place that address these situations specifically and what actions should be taken to verify identities in lieu of the required elements.

Name changes occur due to various factors, including marriage, divorce, adoption, and citizenship. But the degree to which an organization confirms these types of changes prior to making a change in their MPI is important. Patients have nicknames, aliases, or “AKAs” they often use interchangeably and provide to registration staff during intake. Organizations need to educate their intake staff to ask specifically for "complete legal names" and to only update an account's name when legal documentation can be provided to validate the name change.

### **Recommendations on Data Quality**

While there are several steps required to get to an ideal environment for data quality and accuracy, the industry is heading in the right direction. The accuracy of biometrics as a tool for identification has been improving, and as the technology advances and makes the acquisition more cost efficient, this could be considered the gold standard for patient identification.

In the short term, there are other technologies that can be used as an interim step, such as smart cards with an amalgam ID. As hospitals, insurers, and other healthcare entities issue ID cards to patients or customers, they should utilize cards that take advantage of technology, which at a minimum, captures a photograph and signature of the patient accompanied by an amalgam ID.

Currently the best practice is to standardize the collection and entry of three or four static identifiers that are unique enough to provide value to the match, yet not so revealing they pose a threat to the individual's data security. Since data quality is dependent upon the information within the data set being queried, minimizing data components that could have null values or default/generic entries (e.g., using all "9's" in lieu of a SSN not provided) will increase the number of unique entries in the data set.

## **Training**

Training individuals who add or modify patient identity information within the EMPI is the cornerstone for ensuring accurate, complete patient data. This includes individuals on the front-end data collection, entry, and query, as well as the back-end information system staff who ensure proper functionalities are in place to support the technology involved. The ability to link an individual patient's data either within an organization or with other healthcare providers and stakeholders relies on the presence of accurate and consistent patient information. Providing basic training to staff members responsible for this important function improves data quality, which in turn, enhances an organization's ability to link patient records accurately.

Any individual who originates or modifies records within the EMPI requires training. The target audience includes individuals who schedule appointments in ancillary departments, nursing staff who register newborns and/or Emergency Department patients, physician office personnel—if practice data are stored in the EMPI, laboratory personnel registering specimens in a reference lab, blood bank employees creating donor records, health information management and business office staff who modify information in the EMPI as well as all registration staff members. Training is also required of the information system staff who manage the systems on the backend to ensure sound interfaces, proper and consistent settings across multiple systems, etc. A basic understanding of the organization's standards surrounding creating and modifying records within the EMPI coupled with information regarding the characteristics of common data errors serves as the educational foundation for good patient identification practices.

The current environment requires healthcare organizations to formulate and deploy procedures aimed at ensuring an individual is properly identified when being registered and treated. This necessitates use of positive patient identification tools, which run the gamut from requiring a picture ID at registration to biometric identification of incoming patients. Consistent care in all phases of patient identification—from sign-in, interview and placement of an identification band through ensuring the right patient receives the right treatment—underlies good identification practices. Employees must also be trained to recognize potential identity fraud instances that must be monitored to comply with Red Flag Rules.<sup>27</sup>

---

<sup>27</sup> Identity Theft Red Flags & Address Discrepancies under FACTA – Federal Register November 11, 2007, Available at <http://www.ftc.gov/redflagrule>. Accessed on Aug. 26, 2009.

Once the basic concepts surrounding proper patient identification are mastered, performance must be monitored with periodic feedback provided to each employee. Routine quality measures and consequences for failure to meet standards should be established and maintained.

Careful identification practices must be maintained throughout the life of the data, whether the individual is a clerk who enters or updates information about an individual patient in the hospital information system or an information systems employee creating interfaces transmitting patient data between applications. Protocols and tools must be developed to ensure that information flowing across the organization arrives at the intended destination without error. Involving the business owners of the information in testing the veracity of the interfaced data helps ensure accurate information in all applications.

Training protocols must reinforce that updates and/or correction of key patient identifiers (last name, first name, middle name, gender, date of birth, and/or SSN) must be controlled and performed in the appropriate location within the organization's information system. Historically, hospital employees lacked understanding of the data structure of the hospital information system (HIS) database. Consequently, updates frequently occurred in applications that lacked the ability to change all modules within the database. In older HIS systems, HL7 update messages are not routinely sent when the patient does not have a visit episode within a specified date range; this is another challenge in managing accurate patient information. The organization's training policy should address all aspects of data correction, front-end and back-end, to ensure synchronization across all of their various applications.

Individuals involved with billing processes must be keenly aware of clues and patterns of identity fraud, both at an individual level and more organized fraudulent practices. Again, the organization's procedures to comply with Red Flag Rules must be understood and consistently followed. This places greater demand on staff in terms of business knowledge and performance expectations.

Proper patient identification requires individuals to provide personal information in order to ensure they are appropriately identified and treated. While it is necessary for healthcare personnel to acquire personal information about patients, it is important to guard against inadvertent or intentional dispersal of this information. It is imperative that staff have adequate privacy and security training. Not only is this training required by law, it is prudent risk management for the healthcare organization.

Privacy must be addressed throughout patient identification procedures. These include, but are not limited to: techniques used for patient sign-in and notification to approach the registration areas, protection of computer screens when performing patient searches and/or registrations, care in confirmation of existing stored personal information when checking for the proper person, processes for providing a private environment for all registration or scheduling activities and for triage purposes in the Emergency Department.

Training in all of these areas should be role-based, at time of hire and ongoing to ensure staff have the appropriate knowledge to ensure data integrity at its highest level. A

comprehensive training checklist for patient access staff is available in the Appendix at the end of this white paper. Portions of the checklist are applicable to other staff as discussed above.

Cultural variations must be incorporated in all training programs. Proper identification of individuals is complicated by differences in the cultural practices surrounding key identifiers. Each organization should incorporate practices common to their locale in their policies addressing the approved way to enter a patient's name. In training, it is important to note that the organization's approved format may differ from someone's idea of proper or preferred format. The organization should agree for the sake of consistency on the format that will be acceptable for entry into their EMPI, with all staff being trained in that format.

Following are some examples of cultural challenges that may be encountered:

- Use of maternal and paternal surnames in combination with other names in the patient's legal name
- Practice regarding retention of maiden name at marriage
- Order of names may list the surname (last name) before the given (first) name
- Use of compound names, hyphens or absence of such data
- Use of 1/1/YYYY as birth dates for citizens of some regions in the world due to not tracking births via Western calendar
- European practice of listing dates as DD/MM/YYYY instead of MM/DD/YYYY
- Use of particles (*de, da, dos, do, etc.*) in a person's name
- Religious names

### **Recommendations on Training**

Healthcare organizations should strive to educate the individuals who create and modify patient identification information within their health information systems. Education is not one time; it is continuous with appropriate performance markers established and monitored. If the policies described in this section are not currently in place, they should be added to the organization's current practice. Once developed, continuous education must be deployed to ensure compliance.

The first step is to assess the organization's current practice across all areas that impact patient identification. Often an assessment by a third party offers new insights that those individuals involved in day-to-day operations may overlook. Ensure that the assessment includes review of the stored patient data and the processes used for data capture. This includes sound patient interview techniques, as well as data entry conventions and requirements.

Areas that fail to meet expectations or that have not had appropriate historical practices must be corrected. This may involve implementation of new technology paired with effective procedures and on-going training. Any successful long-term solution must incorporate people and processes along with technology.

## **Medical Devices**

PI Integrity is a growing identity management and risk management issue with respect to accurately matching medical data from medical devices and systems to the correct patient record. There is an increasing number of point-of-care devices further driving automation of more monitoring features. There is an increasing amount of medical data being collected, analyzed, and stored in these devices. These stand alone computer systems are disparate medical data systems that need to be interconnected and integrated. To eliminate the “silo effect” there must be consolidation of medical device networks and the enterprise network. The integration of these systems can prove to be challenging.

Typically, hospitals and healthcare organizations have 300 - 400 percent more medical equipment than IT devices.<sup>28</sup> The adoption of the EHR and enterprise-wide deployments of medical device systems are all driving the convergence of private medical device networks with the hospital enterprise network. Medical devices were developed to function as special purpose equipment or stand alone computers. Medical devices record and store clinical values (i.e., numerical values, trending, wave forms), alarm conditions, and error messages continuously or intermittently.

Integration of medical device information, a critical part of EMR planning and implementation, is frequently overlooked. Many medical devices and systems do not have the capacity to send real time transactions without converting the medical device data from one format to another to meet accepted standards. Today, there is an ever-growing requirement by EMR developers that any system feeding patient data into the EMR must have patient demographics. These devices must have an admission, discharge, and transfer (ADT) interface that is compatible with the EMR. Medical devices that only identify a patient by location or a unique device identification number rather than the patient’s identification number are not being supported by enterprise IT systems. Ensuring the integrity of the data and accuracy of patient identification is the most essential infrastructure component of interoperability and communication process, particularly when data from a patient care device are exported to the enterprise HIS system.

Currently, when connecting a patient to a device, the clinician may be entering identifying data through a manual keyboard. Product coding technologies, like barcodes and EPC-numbers, would help improve accuracy at this juncture by eliminating some of the random human error. Clinicians jotting down values generated by the device on scraps of paper and manually entering data into a patient's chart or EMR are faced with delays in information dissemination and even greater risk of errors, typos, or transposed numbers. The impact on quality of care and patient safety can be enormous. The lack of connectivity between devices and EHRs only perpetuates these behaviors, contributing to increasing inefficiency in healthcare. These inefficiencies only drive up cost and increase liability.

---

<sup>28</sup> HIMSS. Medical Device Security. Available at [http://www.himss.org/ASP/topics\\_medicalDevice.asp](http://www.himss.org/ASP/topics_medicalDevice.asp). Accessed on Aug. 26, 2009.

To overcome the challenges of medical device connectivity, there must be an in-depth understanding of point-of-care workflows, medical device connectivity and knowledge of device vendor offerings and product strategies. Device interface development is a specialized task that consumes resources and diverts attention away from core competencies. Obtaining device protocols is generally difficult and sometimes impossible due to competitive issues. The frustration heightens when incomplete connectivity results are obtained, decreasing hospital efficiency.

Each device vendor has its own proprietary protocols. This is not a cost-effective route to connectivity. The device interconnectivity issue is one of safety and efficiency. There is a lack of industry standards for cable connections. There are many devices supporting a variety of independent cable styles making physical connectivity a challenge. Clinicians should not be spending their time sorting cables. Advanced alert systems on items such as bypass pumps and ventilators cannot be implemented without direct communication. The copious volumes of data produced by these types of systems still require manual entry into EHR systems.

Transmitting data is the biggest obstacle due to the lack of a common format. Health IT systems lack the infrastructure to monitor and detect the data. Healthcare decisions are made by comparing results to normal or baseline data. Medical devices lack the capability to do that.<sup>29</sup>

Connecting point-of-care systems, which were designed to run as stand-alone networks, to the clinical information system (CIS) or EHR is not as simple as it may seem. The absence of plug-and-play interoperability can be as primitive as using a non compatible serial port connector or as complex as translating transport protocols and interface development. Cost and complexity are added to the situation when interoperability is lacking.

Efforts to increase interoperability amongst medical devices began in 2005 with the formation of the IHE Patient Care Device (PCD) domain to address integration of medical devices into the healthcare enterprise. As interoperability continues to drive significant improvements in patient safety and quality of care, the medical device transactions must flow from the point-of-care to the EHR. As demonstrated at the HIMSS-sponsored 2007 Connectathon, the IHE PCD domain profiled the successful development and exchange of information from vital signs, physiological monitors, ventilators, infusion pumps, and anesthesia workstations with enterprise applications such as clinical information systems.<sup>30</sup> Work continues to actively extend enterprise level integration of point-of-care devices compatible with new alarm communication management workflow integration needs. Each of these profiles is defined in full detail in the [IHE PCD Technical Framework](#).

---

<sup>29</sup> Technology drives advances in home health monitoring, *WTN News*; April 2, 2008 Available at <http://wistechology.com/articles/4661/>. Accessed on Aug. 26, 2009.

<sup>30</sup> IHE. IHE Patient Care Device Domain. Available at <http://www.ihe.net/pcd/>. Accessed on Aug. 26, 2009.

Every vendor implements data communication differently. Devices transmitting vital signs, particularly in ED, OR and critical care units, are further challenged by the sheer volume of data being generated and the constant shuffling of equipment from one patient to the next. Presently there are standards for device interfaces but no real compliance. Each time new equipment is put into service the interfaces need to be reworked or new interfaces have to be built. “Plug-and-play” should be the ultimate goal.

### **Standards and Regulatory Issues**

Current medical device standards have not adequately addressed the issues surrounding connectivity between devices and healthcare systems. Because there is no connectivity, they are not receiving patient identifiable information or ADT information. They are tracked by location and device number for correlation with the patient on whom it is being used.

The biggest successes have been in HL7 and DICOM transfer protocols. There is still much work that needs to be done integrating error-resistant medical systems in the absence of proven standards for data communication and control, and a lack of reliable and safe system architectures, which support the improvement of patient safety and healthcare efficiency.

Standards-based medical device interoperability can provide real-time comprehensive population of the EMR and lay a foundation for the more comprehensive improvements in patient safety and quality that can arise from the integration of medical devices.

### **Recommendations for Medical Devices**

#### **PI Integrity and Security Issues**

- Point of care devices generally lack the ability to collect and store patient identity data
- Medical devices and systems collecting, storing, and analyzing medical data are designed as special purpose computers and are disparate systems; these systems pose many security risks.
- Remote monitoring devices are specifically assigned to a patient and tracked by device codes but integration to the electronic record is slow

#### **Recommendations:**

- Foster the adoption of Patient Identifier Cross-referencing Integration Profile (PIX) data elements to support accurate record linking
- Develop a process to ensure that the patient and the device in legacy systems are being integrated with clinical systems.
- Conduct security risk assessment of medical devices
- Assign responsibilities and accountability associated with threats and vulnerabilities
- Matching algorithms need to be applied to data transactions to ensure linkage to the correct record

## **Regulatory Issues**

- Lack of patient demographic data capture standards
- Lack of matching criteria across systems standards
- Lack of data set standards
- Lack of coordination of standards

### **Recommendations:**

- Develop patient demographic data capture standards
- Develop matching criteria across systems standards
- Develop standard data sets
- Develop coordination of standards
- Identify the existing legal and regulatory framework surrounding identity management and integrity in medical devices
- National legislation and policy may need to be put in place to address compliance

## **Integration and Interoperability**

- Medical devices are disparate IT Systems
- Point-of-care systems were designed to run as stand-alone systems/networks.
- Absence of plug-and-play interoperability
- Integration of medical device data and enterprise networks continues to create clinical and IT inefficiencies
- Lack of interoperability adds cost and complexity

### **Recommendations:**

- Identify logical connectivity for these disparate systems. Focus on the process of accessing point-of-care stand alone systems.
- Encourage and incentivize further development of technologies and modification to existing systems, making interoperability possible
- Adopt a Web client interface that associates with an enterprise network. A Web client provides a common gateway interface, augmenting the integration of these disparate systems.

## **Device Incompatibility**

- Hardware
  - Hardware connections are not compatible
- Data
  - Data incompatibility slows devices from reaching the mainstream market.
  - Lack of data interoperability

- No common format for exchanging data. A fundamental problem with devices, such as vital sign monitors and infusion pumps, is that there is no common format for exchanging data.
- Data transmission is related to the lack of an infrastructure to monitor and detect the data. Comparison data stored within the EHR is of limited use to physicians and other providers.
- Alarms
  - Develop systems and methods that provide bi-directional interoperability between the medical device and the clinical repository
  - Provide the intelligence to identify variations in the patient's conditions alerting the caregiver or to adjust the machine parameters accordingly

**Recommendations:**

- Standardize the hardware connections for true “plug-and-play” connectivity
- Data standards are needed to improve/reduce the time-to-market.
- Develop bi-directional communication with intelligence to modify the treatment and notify a healthcare professional to assess the situation.

## **Barriers to Patient Identification Integrity**

There are many barriers that need to be overcome in order to “solve” the PI Integrity problem. For some of these barriers, there appear to be ready solutions, while others are not so easily resolved. However, substantial progress can be made on virtually all of these issues if appropriate healthcare agencies develop the consensus and willpower to do so. This section examines the following barriers to achieving accurate patient identification:

- Costs
- Confidentiality and privacy issues
- Impression of this being an ‘unsolvable’ problem
- Lack of enabling legislation
- Inadequate legal infrastructure
- Lack of provider clinical automation
- Poorly defined business processes
- Insufficient analysis of errors
- Lack of data standards

### **Costs**

The cost of implementing a PI Integrity solution must include the cost of creating the solution and the cost of deploying it across all of the relevant healthcare organizations. Traditional ‘solutions’ to the healthcare identification problem involve sums that start in the

billions or tens of billions of dollars. The RAND Corporation argues that even at a cost of \$11 billion dollars the project would be worthwhile because of the benefits that would accrue.<sup>31</sup> However, the Voluntary Universal Healthcare Identifier project described in this document is already operational and could be deployed at a small fraction of this cost estimate. In light of the large sums of money being proposed to ‘fix’ healthcare by the Obama administration and the strong return on this investment, it appears that cost should not be considered a barrier to solving the patient identification conundrum.

### **Confidentiality and Privacy Issues**

Privacy continues to be the most problematic and difficult aspect of creating a patient identity solution. The healthcare privacy community is correct when it argues that we must take proactive and ongoing actions to protect and improve the privacy of patients and their demographic and clinical information. Unfortunately, this has led to a climate where protecting privacy and enabling sharing of clinical information has come to be viewed as an ‘either or’ situation. We must correct this misimpression and offer a variety of options that enable a patient and a healthcare organization to balance these two needs for each specific situation. The VUHID patient identification option described earlier has specific provisions to support a variety of privacy capabilities that enable case-by-case determination of the amount of privacy and information sharing that is appropriate for that situation.

### **An ‘Unsolvable’ Problem**

There is an impression that the UPI is not feasible. Based on the complexity of the issue, the lack of industry consensus and reinforced by the Congressional ban on applying federal resources to an individual healthcare identifier, many consider it impossible to solve this problem. Traditional approaches to patient identification do indeed run into a set of problems that appear to be insolvable. However, new innovative approaches to patient identification that can be easily implemented on top of traditional demographic matching hold the promise of virtually eliminating matching errors while also addressing patient privacy concerns. Intensive efforts to educate the industry concerning the new solutions that are available are needed to ensure that all potential ways to solve this pressing need are evaluated, so that truly optimal solutions can be put in place.

### **Lack of Enabling Legislation**

Currently, there is a lack of enabling legislation. Because accurate patient identification is a national (indeed an international) problem, most people look to the U.S. Congress to craft a legislated solution. The 111<sup>th</sup> Congress has under consideration a healthcare reform bill that includes a machine-readable health plan identification card. This would indicate movement in the direction of a unique ID standard. However, alternative strategies should be explored that are not dependent upon Congressional action in the hope that faster, cheaper, and perhaps even more effective solutions may be found.

---

<sup>31</sup> 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 Monograph. October 2008, No. 753. Available at [http://www.rand.org/pubs/monographs/2008/RAND\\_MG753.pdf](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf). Accessed on Aug. 26, 2009.

### **Inadequate Legal Infrastructure**

There is an inadequate legal infrastructure to support the advances in technology. The whole concept of using technology to communicate patient health information is fairly young in the cycle of life and, as a result, there has yet to be an adequate amount of time spent figuring out all of the legal implications. Over time, the legal expertise in the global environment of electronic information sharing is going to increase. Both providers and the legal establishment will have to become more knowledgeable in the fields of privacy, security, and technology. The legal infrastructure will have to become more robust, enabling providers to make right decisions.

### **Lack of Provider Clinical Automation**

The adoption of clinical automation by providers is very low in the United States today. Hospitals and providers have not made the investment to automate their operations mainly due to cost. As the investment in health IT increases, there will be a significant impact on how healthcare is delivered, with an increase in quality of care and patient safety and reduction in cost.

To receive stimulus financial incentives from the American Recovery and Reinvestment Act of 2009 (ARRA), providers must implement an EMR that meets the “meaningful use” definition. Over the next several years, it is hoped that providers will take advantage of the incentives and become more automated. Solo practice physicians are the most at risk for not adopting EMRs due to the capital outlay required.

### **Poorly Defined Business Processes**

On the whole, healthcare providers have poorly defined business processes. This includes clinical, administrative, and financial processes. Until the last several years, the revenue cycle in hospitals has not received much attention. Some of the lowest paid workers are the first face that a patient sees, yet they are frequently among the lowest paid and least trained. This adds to the problem of significant errors during patient registration. These errors flow through the information cycle and frequently are not corrected until information has been sent out of the source application. This causes significant data quality issues for all systems that are receivers of the patient demographic information. This, in turn, leads to the problem with matching patient demographic information for the purpose of sending and receiving results. In order to address these barriers, hospitals must address the beginning of the revenue cycle process with well-trained associates at registration. This will impact the whole cycle positively, thus feeding those systems both inside and outside of the hospital’s walls with accurate data.

### **Insufficient Analysis of Errors**

Data errors have always been a problem for providers, though analyzing the errors is becoming an important process that will need to be addressed. The inability to analyze errors on the source system is a problem for most applications. In order to analyze the data, there is usually an extraction of the data out of the source system into another system, which has the ability to find and analyze the errors. The problem with this is that it is often difficult once the analysis is completed to identify where the error actually originated. Healthcare organizations must develop a robust data governance with strong data

stewardship. Errors must be anticipated, managed, and eliminated prior to flowing to other systems. This is particularly important in this day of HIE.

### **Lack of Data Standards**

The need for data quality standards for electronic transmission of personal health information is evident. Currently, there are very few standards surrounding the data needed in order to have true interoperability. An example of a standard for lab tests that is available which is not incorporated into most institutions operations is the [LOINC codes](#) developed by Regenstrief Institute, Inc. Until data standards are used or developed, there will continue to be issues with the quality of the data collected.

The lack of standards creates issues with quality, data linking, the ability to authenticate the data and the integrity of the data. One study published through AHIMA<sup>32</sup> showed the lack of standards resulted in error rates for false positives ranging from 6 percent to 22 percent. Other issues include lack of information collected due to no accountability, patients unable to provide information, and systems not being able to accept the data from different sources. In order to be able to set data quality standards, the data must be defined, collected according to those standards, and constantly monitored.

### **Summary**

The list of barriers noted above is long and somewhat daunting. The burden placed on healthcare by not solving the patient identification problem is even more significant. Given that a reasonable consensus can be achieved on what solution to attempt, there do not appear to be any barriers in the list given above that would justify *not* trying to achieve a solution. Even if the result were to only reduce the rate of errors by 25 percent from the current unacceptable levels, the results could be considered to be money and resources well spent. If the premise of this white paper is true and it is possible to eliminate the vast majority of patient identification errors, then there is an enormous moral imperative for healthcare to attempt to create such a solution.

---

<sup>32</sup> Fernandes, Lorraine, and O'Connor, Michele. "Future of Patient Identification." *Journal of AHIMA* 77-1 (January 2006): 36-40.

## Appendix A: VUHID

A brief description of the Voluntary Universal Healthcare Identifier (VUHID) project:

The VUHID system is based on two ASTM International standards, E1714 and E2553<sup>33</sup> that were most recently approved in 2007. The VUHID project has three primary departures from previous healthcare identification proposals that combine to make the system exceptionally simple and cost effective: 1) It is a voluntary system, 2) It is layered on top of the NHIN, and 3) It is architected to avoid the need for a central repository of patient demographic data. A very important side effect is that the system can also meet the most demanding privacy and patient confidentiality requirements.

The VUHID system involves the operation of a secure, highly scalable Web site. The EMPI systems at the heart of various HIEs<sup>34</sup> communicate with the VUHID Web site to receive services relating to VUHID identifiers. The fundamental transactions involved are:

1. EMPI requests a new open identifier
2. EMPI requests a new private identifier
3. EMPI requests the status of a specific identifier
4. EMPI requests the retirement or termination of an identifier
5. EMPI requests the location of clinical information associated with a specific identifier
6. VUHID system notifies EMPI of the termination of an identifier

These transactions are designed to be secure, simple, and fast. Note that inherent in the design of these transactions is the fact that none of them require patient identification, patient demographic information or patient clinical information to be exchanged between an EMPI and the VUHID Web site.<sup>35</sup> The VUHID system has been architected to make it operationally impossible for the VUHID servers to ever have identifying information on a person; indeed, VUHID is never aware of the identity of the person associated with any specific VUHID identifier. By taking this approach, users can be assured that participating in the VUHID system does not represent any increased risk to their privacy.

A patient receives a VUHID identifier by requesting one from a caregiver who is a member of a participating HIE. The patient's demographic information is acquired and forwarded to the HIE EMPI system for a match. Once the EMPI has positively identified the person, it sends a request for an identifier to the VUHID server but includes no information about the

---

<sup>33</sup> These standards can be obtained at [www.astm.org](http://www.astm.org).

<sup>34</sup> For purposes of this paper, we will consider the terms health information exchange (HIE) and regional health information organization (RHIO) to be synonymous.

<sup>35</sup> See the VUHID Web site for a description of how VUHID identifiers can enable clinical information exchange without any clinical information ever being sent to the VUHID Web site (<http://gpil.info/>).

person involved. The VUHID system generates a new unique identifier, marks it as 'active', notes the date and time it has been created as well as the EMPI that requested it, and returns the identifier to the EMPI system. The EMPI then links this identifier to any pre-existing identifiers in the HIE for that person and prints out an ID card for the person to carry. The person can then take this card to any other caregiver in the HIE and it can be used to retrieve demographic information from the EMPI by simply having the site enter the identifier and query the EMPI system.

Other aspects of the VUHID system not mentioned here, for reasons of space, include enabling it to serve patient privacy needs and enabling exchange of patient clinical information no matter where the patient goes as long as the caregivers he or she visit are participating in the VUHID network.

The two primary goals of the VUHID network are to enable unambiguous patient identification and to support patient control over the privacy of their information. Further details on the VUHID system are available at <http://gpii.info>.

## Appendix B: Algorithmic-based Matching—The Basics

Statistical- or mathematically-based electronic algorithms are dependent upon key demographic data points for that person or patient. These demographic attributes, such as name, DOB, and address are not unique to a single individual and, with the exception of DOB, may change frequently throughout a person's lifetime. This means that matching errors will continue to occur even if the matching algorithm is standardized. Most matching algorithms perform better when they have more data elements available for comparison (see Table 1). The minimum data set for patient identification used most commonly in the industry includes full legal name (first, middle, last, suffix), DOB, gender, full mailing address, and home telephone number. The last four digits of the SSN should be captured since it provides significant improvement in the algorithm's success in matching records. As fewer patients are willing to provide their full SSN, an effort in the healthcare industry to capture at least the last four digits will improve algorithmic record matching. Additional data, such as race or ethnicity, guarantor, insurance, next of kin, emergency contact, other identifiers (insurance number, driver's license number, UPI), mother's name, father's name, and work address, may also be collected. All of these data elements can be used by sophisticated statistical matching algorithms.

Last name, first name, middle name or middle initial, DOB, gender, SSN or last four digits of SSN, and telephone or address are generally considered the minimum data elements for best algorithm performance and to reduce or eliminate mismatches (false-negatives). Many in the health technology sector tout the absolute minimum for matching as last name, first name, DOB, and gender. Less data results in less confidence in match scores and even exact matches on these four data points across a large data set of records will result in false positive matches and overlaid records. A fifth data point is needed to validate identity, preferably the SSN (or last four digits) or middle name to increase the likelihood that twins or individuals with common names are not linked improperly.

How various record matching algorithms actually compute the duplicate rate also varies significantly. The July 2009 AHIMA Practice Brief "Managing the Integrity of Patient Identity in Health Information Exchange,"<sup>36</sup> cited in the Algorithm Section above, provides guidance on how the duplicate rate should be computed. This Practice Brief is a work product of AHIMA's 2009 HIE Practice Council and reflects the consensus opinion of the professionals who developed it who have worked with this issue extensively in the field. It has not been validated through scientific research. The Patient Identity Integrity Workgroup supports and recommends such validation research.

Using more data elements in algorithmic record matching also increases the probability that there will be an error in at least one of the data elements (see Figure 1). Increasing the number of data elements used in an algorithm may increase the complexity of the programming of that algorithm. Additionally, many of the key data fields needed for algorithmic record matching involve fields where the data value may likely change over

---

<sup>36</sup> AHIMA. "Managing the Integrity of Patient Identity in Health Information Exchange." *Journal of AHIMA* 80, no.7 (July 2009): 62-69.

time (names, addresses, telephone numbers and even the patient's gender). The reality is that no algorithm, no matter how good, can compensate for bad data, and bad data can result from real changes made by the patient (e.g., name change) as well as other errors throughout the system. Biometric devices have been mentioned in the context of matching algorithms and although not yet used frequently in healthcare for patient identity, they are starting to gain notice as a component to solving the patient identity problem.

Table 1 demonstrates that as the percent of data fields present for matching in a given set of three million records decreases, the error rate ("false-negative rate") increases. Even with matches on last name, first name, and 90 percent of field attributes of three additional fields [DOB, zip code, and last four digits of SSN], 3 percent of the time a record for a patient is not found (so the "match" fails).

**Table 1: More Data ... Improved Matching<sup>37</sup>**

Scenario	Percent of Attributes Present				
	Name	Date of Birth	Zip Code	SSN (last 4 digits)	False-negative Rate (%)
A	100	100	100	0	6%
B	100	90	90	0	22%
C	100	90	90	70	7%
D	100	90	90	90	3%

Source: Initiate Systems, Chicago, IL

Differing scenarios illustrate that the greater the number and completeness of data elements present for matching patient records, the lower the error rate (Scenario D). When fewer elements are present, the rate of bad matches increases (Scenario B).

### Errors Created by Algorithmic Methods

All statistical matching methods are subject to false positive and false negative results. False positive results occur when information for two different people appears to be a match representing the same individual. False negative results occur when two different records for the same person are thought to represent different people. The RAND Corporation analyzed an 80 million record demographic database and determined that an error-free combination of name (not specified in the report, but assumed to be last name/first name), DOB, zip code and last four digits of the SSN would be the minimum required data set to accurately identify all patients.<sup>38</sup> Simply removing the partial SSN created nearly 1,000 false positive matches. In a large database, the use of an almost-unique data element such as SSN dramatically reduces the false positive rate. Based on the RAND study, for a random individual in the 80 million record database, there is approximately a 98 percent chance of

<sup>37</sup> Fernandes, Lorraine, and O'Connor, Michele. "Future of Patient Identification." *Journal of AHIMA* 77-1 (January 2006): 36-40.

<sup>38</sup> 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 Monograph. October 2008, No. 753. Available at [http://www.rand.org/pubs/monographs/2008/RAND\\_MG753.pdf](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf). Accessed on Aug. 26, 2009.

finding another person's record with the same last name. If matching on the last name and DOB, there is a 66 percent rate of finding the wrong record. And there is still over a 12 percent chance of finding the wrong record when matching on a combination of last name, first name, DOB, and zip code. This means that even with an algorithmic match of last name, first name, DOB, and zip code, 12 percent of the records identified will not belong to the searched patient. Adding the last four digits of the patient's SSN dropped the false positive (erroneous) matching to 1 out of 39 million—a dramatic improvement in the record matching capability of the algorithm. This study also demonstrated that the false positive rate is sensitive to population size; the larger the population, the higher the false positive rate.<sup>39</sup>

Due to the nature of statistical matching, reducing the number of false positives drives an increase in the number of false negatives, and vice versa. When comparing two records, as more demographic data elements match, the higher the algorithm score, or record match weight. The higher the match weight, the more likely it is that the two records represent the same individual. Conversely, the lower the match weight, the more likely it is that the two records represent different individuals. The process of balancing the two types of errors is called “tuning the algorithm.” Generally, this process results in a match weight threshold that minimizes both false positives and false negatives. If the match weight threshold is set too high, there will be a large number of false negatives or missed real matches. If the match weight is set too low, there will be a higher volume of false positives, or erroneous matches. The goal should be to identify as many real matches as possible while minimizing the erroneous matches. Since the data and the methodology are imperfect by design, there will always be a zone of ambiguity where potential matches require some level of human review and resolution if the business objectives require.

The purpose for which the algorithmic matching is being performed also affects how one would want to “tune” the algorithm. For instance, if you are analyzing an entire database to identify all the potential duplicates within that database, it would be advisable to widen the net and try to identify all possible duplicates. During an EMPI cleanup project, each potential duplicate pair would be evaluated and false positives eliminated. However, on a front-end search across an HIE dataset, the goals may be to tighten the threshold, eliminate chances of false positive matches, and only bring up highly probable results to the clinician. This would ensure that care would not be provided by that clinician based on another patient's medical history but conversely would also result in more frequently not detecting portions of the patient's record that should be presented as part of their clinical data.

Data quality impacts algorithmic performance and it is not possible with current tools and technology to entirely automate patient matching for most healthcare environments. Resolution of data discrepancies frequently requires human involvement. It must be noted that some organizations choose not to resolve ambiguity based upon their business drivers or functions (i.e., e-prescribing, pharma, imaging). If patients cannot be unambiguously

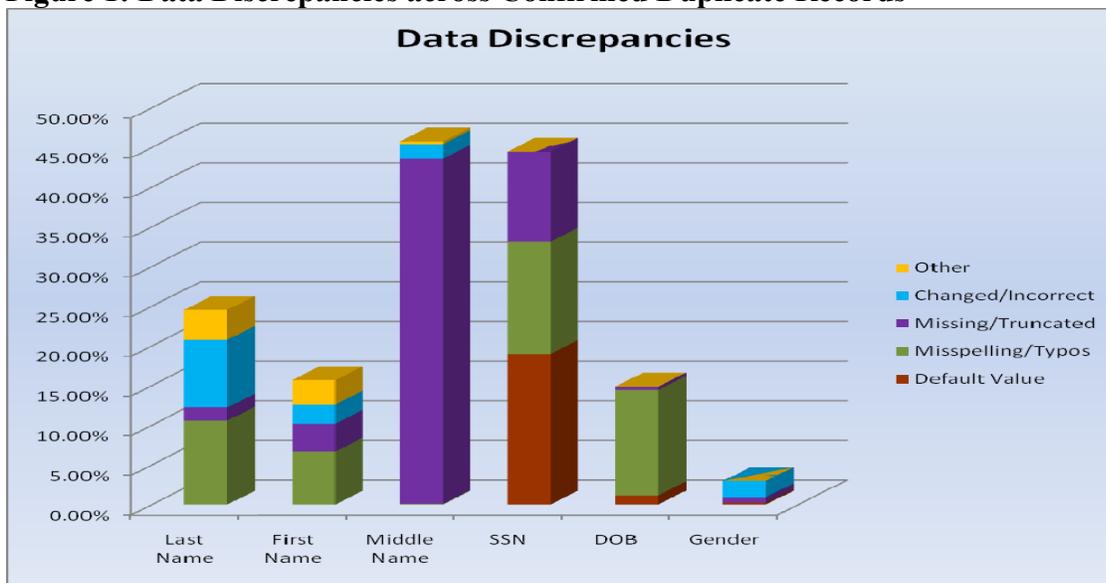
---

<sup>39</sup> Ibid, p.14

identified via a computer-based process, machine level interoperability will be hampered significantly.<sup>40</sup>

False negatives, also represented as duplicate records in an MPI, are usually the result of discrepancies in demographic data elements or missing information. Name or address changes, data entry errors, database formatting changes, and business processes can all contribute to the creation of duplicate records. A study performed by Initiate Systems showed the mean duplicate rate to be 8 percent in 112 MPIs analyzed. Discrepancies in female last name, SSN, and telephone number were the most common. A study by Just Associates determined 88 percent of validated and confirmed duplicates have one or more errors in the six key data elements and over 46 percent have errors in two or more of these six key fields (last name, first name, middle name, DOB, gender and SSN).<sup>41</sup>

**Figure 1: Data Discrepancies across Confirmed Duplicate Records**



Source: Just Associates, Inc., Centennial, Colorado. 2007

The study was completed on 244,356 confirmed duplicate records. The duplicate records were identified via an algorithmic computer analysis and validated and confirmed by trained patient identity experts to be true duplicate records.

As demonstrated in Figure 1, key demographic data fields stored in a Master Patient Index database are frequently populated with default values, typographical errors, and misspellings. Additionally, key fields such as a patient's middle name are frequently not populated and name field values frequently change. The DOB is in error over 13 percent of the time and even the patient's gender is incorrect 3 percent of the time.

<sup>40</sup> Ibid, p.20.

<sup>41</sup> Other Data (Just Associates Inc., unpublished data, 2007)

These studies demonstrate the need for data definitions and the importance of complete and accurate data for quality and consistency of patient matching. As noted by the AHA, "...there are serious safety risks that could arise from attributing a medical record to the wrong individual . . . a cluster of demographic information may not be sufficient to distinguish between the 37-year-old Mary Jones with diabetes and a penicillin allergy and the 37-year-old Mary Jones in perfect health. Mixing up their records could have serious consequences."<sup>42</sup>

### **Data Definition Standards**

Optimal record matching requires the uniform adoption not only of data transaction standards as are defined in HL7 or X12, but standards must address the definition of field-level data elements and the format for data entry. One example is the definition for the field "mother's name." Does this mean the patient's mother's current name, mother's maiden name, mother's first name, or mother's last name? Is it to be collected for all patients or just newborns? If it is not available, what should be entered in the field? Other examples include defining the data entry protocols for a patient without a middle name, or for a patient with a hyphenated last name, a legal first name vs. a nickname or an unknown DOB. Cultural variations in names may make it difficult to determine which part of the name is the first name and which part is the last name.

The lack of standardization of data sets, a standard field, or attribute definition and limitations of statistical matching methods contributes to the challenges associated with record linking. Without national standards, the marketplace has generally adopted the expedient approach of using the same statistical matching methods that most large healthcare provider systems have used in their MPI systems for years.<sup>43</sup> This means that most health information systems utilize only deterministic or exact match solutions on partial sets of data fields. These matching capabilities only identify the record for which the user is searching 85 – 92 percent of the time at best. Reports have documented that 30 percent of the time, physicians could not locate previous results on their patients.<sup>44</sup> This is a pandemic problem that only gets worse as the databases get larger, as demonstrated in the RAND report.

---

<sup>42</sup> American Hospital Association (AHA), "Protecting and Improving Care for Patients and Communities: Health Information Technology", 2006 Advocacy Position Paper, Washington, D.C., 2006.

<sup>43</sup> 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 Monograph. October 2008, No. 753. Available at [http://www.rand.org/pubs/monographs/2008/RAND\\_MG753.pdf](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf). Accessed on Aug. 26, 2009.

<sup>44</sup> Electronic Medical Records – Getting it Right and Going to Scale. W. Edward Hammond, II, PhD. Commonwealth Fund background paper (Jan. 2004, [www.cmwf.org](http://www.cmwf.org) #695)

## Levels of Algorithms

In its practice brief *Surveying the RHIO Landscape: A Description of Current RHIO Models, with a Focus on Patient Identification*, AHIMA defined three levels of record matching algorithms: Basic, Intermediate, and Advanced.<sup>45</sup> Definitions provided include:

1. Basic record linking compares selected data elements—most frequently name, DOB, SSN, or gender—using exact (identical match of data elements) and deterministic (exact or partial match) linking approaches. This method assumes a high degree of confidence that the match is accurate.
2. Intermediate record linking provides more advanced techniques for comparing records by enhancing exact match and deterministic tools with additional logic and arbitrary or subjective scoring systems. Subjective weighting, ad-hoc weighting, fuzzy logic, and rules-based algorithms are examples of intermediate matching tools.
3. Advanced record linking employs sophisticated mathematical or statistical algorithms such as probabilistic matching, bipartite graph theory, machine learning, and neural networks. They also account for aspects such as missing data, default and invalid data, transposed digits or characters, misspellings, nicknames, and the time variance of a patient’s information.

The vast majority of information systems utilize basic or deterministic record matching methods, which rely on exact or partial matches of key demographic data elements. Some systems have been enhanced with rules engines to improve the performance of the basic record matching methodology. However, the limited research that has been performed clearly demonstrates that advanced mathematical tools produce the best matching results. In addition to the most commonly known probabilistic or frequency-adjusted models, advanced mathematical algorithms include Bayesian pattern recognition, bi-partite graph matching, machine learning, and neural networks. These advanced systems appear to significantly improve record matching capabilities. Scientific studies have not been performed to determine which of these methods most effectively reduces false positives and false negatives, thereby producing the most accurate and reliable patient matching. Table 2 below provides an overview of the three levels of algorithms and the effectiveness of their attributes.

**Table 2: Levels of Computerized Record Matching Algorithms**

Algorithm Attribute	Algorithm Effectiveness		
	Basic	Intermediate	Advanced
Identifies high percent of potential duplicate records (low false negative rate)	Low	Low-Moderate	High
Creates erroneous matches (high false positive rate)	High	Moderate-High	Low
Record match weight indicative of	N/A	Low	Moderate-High

<sup>45</sup> E-HIM Work Group on Patient Identification in RHIOs. "Surveying the RHIO Landscape: A Description of Current RHIO Models, with a Focus on Patient Identification." *Journal of AHIMA* 77, no.1 (January 2006): 64A-D.

true or false match (record match weight can be trusted)			
Adapts to variations in population of records	Low	Low	Moderate-High
Tolerance for data errors or discrepancies in data values	Low	Low-Moderate	High
Complex/expensive to implement	Low	Moderate-High	High

**Note:** There can be significant variations in sophistication in intermediate algorithms and, although probabilistic algorithms are considered advanced, caution should be exercised. Many algorithms marketed to be advanced are not of the same caliber in effectiveness as others.

## Appendix C: Patient Identification Training Checklist

<input type="checkbox"/> <b>Patient access staff should follow an organizational naming conventions policy that documents in detail how to record a patient’s name in the MPI. They should be trained in how to follow the policy and their compliance monitored. The policy should address:</b>
<ul style="list-style-type: none"><li>• Use of legal name</li><li>• Use of punctuation and/or spaces in names – is it allowed or not &amp; when</li><li>• Use of entitles</li><li>• Newborn registrations, designation of multiple births, updating name to legal name</li><li>• Trauma and unknown individuals naming conventions</li><li>• Use of middle names/initials</li><li>• How to handle patient name when the insured name differs from the legal name</li></ul>
<input type="checkbox"/> <b>Name changes policy</b>
<ul style="list-style-type: none"><li>• Who is authorized to make changes/corrections to the patient name and when</li><li>• What sort of documentation must be presented to change a name</li><li>• Do adopted children’s record have to meet alternate standards</li></ul>
<input type="checkbox"/> <b>How to search the MPI to maximize search results</b>
<ul style="list-style-type: none"><li>• Types and number of searches required prior to creating a new record</li><li>• Reason(s) why searches are sometimes unsuccessful</li></ul>
<input type="checkbox"/> <b>How to validate an existing record prior to selecting it for registration</b>
<ul style="list-style-type: none"><li>• How many data elements must match to select a record</li><li>• How to solicit information from the patient to help in validation – a script to ask questions in appropriate format</li></ul>
<input type="checkbox"/> <b>Positive patient ID</b>
<ul style="list-style-type: none"><li>• Requirements for government issued picture IDs – when required or not</li><li>• Secondary identification checks prior to application of an ID band</li></ul>
<input type="checkbox"/> <b>Minimum data requirements to create a new record in the MPI</b>
<input type="checkbox"/> <b>Maintenance of patient privacy throughout the registration process</b>
<ul style="list-style-type: none"><li>• Sign-in sheets, Use of techniques to minimize using full names</li><li>• Protecting face sheet information</li><li>• Private conversations during registration</li></ul>
<input type="checkbox"/> <b>How to recognize a data integrity issue and what steps to take to address it. This includes duplicates, overlaps, overlays and/or suspected identity fraud.</b>
<input type="checkbox"/> <b>The organization’s Red Flag Rules policy and procedure to follow.</b>
<input type="checkbox"/> <b>Patient Access employees need to understand their database: hierarchical or relational; implications on storing and correcting patient identifying information.</b>
<input type="checkbox"/> <b>Background information regarding the “national state” of patient identification</b>
<ul style="list-style-type: none"><li>• Common causes and types of errors encountered in databases</li><li>• Common duplicate &amp; overlap error rates</li><li>• Their organization’s duplicate rates, historical and current</li><li>• Feedback regarding their own performance surrounding duplicate/overlay creation.</li></ul>

Checklist Provided By: Linda Bock, RHIA, Senior Consultant, Just Associates, Inc.

## Appendix D: Additional Sources

1. Health Level 7 (HL7): [www.hl7.org](http://www.hl7.org)
2. Integrating the Healthcare Enterprise (IHE): [www.ihe.net](http://www.ihe.net)
3. IEEE 802.11, Institute of Electrical and Electronics Engineers (IEEE): [grouper.ieee.org/groups/802/11/](http://grouper.ieee.org/groups/802/11/)
4. The Health Insurance Portability and Accountability Act (HIPAA) of 1996: [www.hhs.gov/ocr/privacy](http://www.hhs.gov/ocr/privacy)
5. Healthcare Information Technology Standards Panel (HITSP): [www.hitsp.org](http://www.hitsp.org)
6. "Wireless Networks" [on-line] (June 2, 2003); available from Internet: [www.hipaaadvisory.com](http://www.hipaaadvisory.com)
7. The Wi-Fi Alliance: [www.wi-fi.org](http://www.wi-fi.org)
8. Nikita Borisov, Ian Goldberg, and David Wagner, "Intercepting Mobile Communications: The Insecurity of 802.11": [www.isaac.cs.berkeley.edu/isaac/mobicom.pdf](http://www.isaac.cs.berkeley.edu/isaac/mobicom.pdf)
9. HITSP Patient ID Cross-Referencing Transaction Package Released for Implementation 20081218 V2.3: [http://publicaa.ansi.org/sites/apdl/IOLib/HITSP\\_V1.0\\_2008\\_C62\\_Unstructured\\_Document.pdf](http://publicaa.ansi.org/sites/apdl/IOLib/HITSP_V1.0_2008_C62_Unstructured_Document.pdf)
10. Recommendations for a Unique Health Identifier for Individuals in Ireland: [http://www.higa.ie/media/pdfs/Unique\\_Health\\_Identifier\\_Report.pdf](http://www.higa.ie/media/pdfs/Unique_Health_Identifier_Report.pdf)
11. "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](http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf)
12. Medical Identity Theft Final Report, Booz Allen Hamilton, January 15, 2009: [http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS\\_0\\_10741\\_848096\\_0\\_0\\_18/MedIdTheftReport011509.pdf](http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_10741_848096_0_0_18/MedIdTheftReport011509.pdf)
13. The Connecting for Health Common Framework: Resources for Implementing Private and Secure Health Information Exchange (Particularly sections P4 and T5), Markle Foundation, 2006: <http://www.connectingforhealth.org/>

14. Achieving the Health IT Objectives of the American Recovery and Reinvestment Act: A Framework for "Meaningful Use" and "Certified or Qualified" EHR, Markle Foundation, 2009:

[http://www.markle.org/downloadable\\_assets/20090430\\_meaningful\\_use.pdf](http://www.markle.org/downloadable_assets/20090430_meaningful_use.pdf)

15. References on the state of biometric technology include: -

1. <http://www.businessgreen.com/vnunet/news/2236775/researchers-hack-facial>

2. <http://www.engadget.com/2006/09/22/digital-fingerprint-door-lock-defeated-by-photocopied-print/>

3. <http://www.engadget.com/2008/04/03/researcher-raises-alarm-about-biometric-hacking-with-biologger/>

16. For information on card systems:

[http://www.ftc.gov/opa/2006/02/cardsystems\\_r.shtm](http://www.ftc.gov/opa/2006/02/cardsystems_r.shtm)

## Appendix E: Glossary of Key Terms

<b>Integrity</b>	<p><b>Data Integrity:</b> “The quality of correctness, completeness, wholeness, soundness and compliance with the <i>intention of the creators of the data...</i>”<sup>46</sup></p> <p><b>Message Integrity:</b> “The validity of a transmitted message... (dealing) with methods that ensure that the contents of a message have not been tampered with and altered...” Data Integrity “is achieved by preventing accidental or deliberate but unauthorized insertion, modification or destruction of data in a database...”<sup>47</sup></p> <p><b>Patient Identity Integrity (PI Integrity):</b> The accuracy and completeness of data attached to or associated with an individual patient. This includes the accuracy and quality of the data as it relates to the individual as well as the correctness of the pairing or linking of all existing records of that individual within and across information systems. PI Integrity is of central importance to achieving quality of care, patient safety, and cost control.</p>
<b>Duplicate Record</b>	One person with two or more records in one database or Master Person Index (MPI). Without the two records being linked, information not available for point of care and clinical decisions are made in the absence of data. There is an increase in costs associated with repeat tests, clinical procedures, etc., as well as rework in clinical and business processes.
<b>False Positive</b>	Algorithm match error occurring when information for two different people appears to be a match representing the same individual.
<b>False Negative</b>	Algorithm match error occurring when two different records for the same person are thought to represent different people.
<b>Overlap Record</b>	One person with two or more unique enterprise identifiers. Without the two records being linked, information not available for point of care and clinical decisions are made in the absence of data. There is an increase in costs associated with repeat tests, clinical procedures, etc., as well as rework in clinical and business processes.
<b>Overlay Record</b>	Records of two different people are “combined” into one record in error. Person A is treated with Person B’s clinical information. This has huge implications for quality of care and patient safety.

---

<sup>46</sup> PCmag.com Encyclopedia of IT terminology

<sup>47</sup> Ibid

**Patient Identity  
within Integrity  
Management  
(PIIM)**

The totality of business processes required to assure PI Integrity within and across organizations.

**Unique Patient  
Identifier**

The value permanently assigned to an individual for identification purposes and is unique across the entire (UPI) national healthcare system. A UPI is not shared with any other individual.<sup>48</sup>

---

<sup>48</sup> <http://www.ncvhs.hhs.gov/app3.htm>

## **Appendix F: HIMSS Patient Identity Integrity Work Group Members**

Barbara Demster, MS, RHIA, CHCQM,  
Work Group Chair

John Awad  
Shanda Brown  
Paul Donohoe  
Lydia Duckworth  
Stacie Durkin, RN-C, RHIA, MBA  
Lorraine M. Fernandes, RHIA  
Barry Hieb, MD  
Beth Haenke Just, MBA, RHIA  
James Kragh

Deborah Lafky, MSIS, Ph.D., CISSP  
Martin Larson  
Steve Posnack  
Anil Saldhana  
Kamilah Shepherd, RHIA  
Sara Temnitz  
Sherri Walter  
David Weitzel, MS, JD, CIPP/G

### **Additional White Paper Contributors**

Leah Blackstone  
Linda Bock, RHIA  
Mark Haas, MBA  
Bill Klaver  
Michele O'Connor, RHIA, FAHIMA  
Vicki Wheatley, MS, RHIA

Thanks also to Helen Hill, Richard Hillestad, Michael Nusbaum, Peter Palmer, Eithne Reichert, Kathy Sadler, and Letha Stewart for their contributions to the work group's discussions.

---

HIMSS wishes to thank the volunteers who worked on this paper and put aside their personal business agendas to work in a collaborative manner to address this issue that affects us all. The paper reflects the broad range of industry experience and subject matter expertise they represent. In creating this white paper, participants do not necessarily endorse all of the concepts or recommendations arrived at through this collaboration. These cross-industry viewpoints ensure that HIMSS fulfills its requirement to offer a coordinated voice to the national discussion on these important healthcare issues.