

Registries and Patient Identity Management

Draft White Paper for Third Edition of “Registries for Evaluating Patient Outcomes: A User’s Guide”

Introduction

Electronic health care data are increasingly being generated and linked across multiple systems, including electronic health records (EHRs), patient registries, and claims databases. In general, every system assigns its own identifier to each patient whose data they maintain. This makes it difficult to track patients across multiple systems and identify duplicate patients when different systems are linked. Efforts to address this challenge are complicated by the need to protect patient privacy and security.

Patient identity management (PIM) has been defined as the “ability to ascertain a distinct, unique identity for an individual (a patient), as expressed by an identifier that is unique within the scope of the exchange network, given characteristics about that individual such as his or her name, date of birth, gender [etc.].”¹ For the purposes of this chapter, the scope of this definition will be expanded to refer to PIM as the process of accurately and appropriately identifying, tracking, managing, and linking individual patients and their digitized health care information, often within and across multiple electronic systems.^{2,i} There is an increased need for PIM strategies in the realm of health care data, and the primary reason for this is the continued rise in the quantity and linkage of electronic health care data.

The quantity of electronic health care data continues to grow. EHRs are increasingly being used to generate electronic health care data – about 50% of office-based physicians in the U.S. now use some form of EHR.³ This number is likely to increase significantly in response to the EHR incentive programs

ⁱ A related idea is the concept of patient identity integrity, which is defined as “the accuracy and completeness of data attached to or associated with an individual patient.” Efficient patient identity management leads to high patient identity integrity. See HIMSS Patient Identity Integrity Work Group. Patient Identity Integrity. 2009. Available at: <http://www.himss.org/content/files/PrivacySecurity/PIIWhitePaper.pdf>. Last accessed 24 June 2011.

enacted by the Centers for Medicare and Medicaid Services (CMS), which “provide a financial incentive for the “meaningful use” of certified EHR technology to achieve health and efficiency goals.”⁴ In addition to office-based EHRs, electronic health care data may be created by hospital EHRs, billing systems, insurance claims systems, pharmacy record systems, medical devices, and even by patients themselves via electronic patient health record systems. Large amounts of electronic health care data are also being generated from clinical research. Patient registries, for example, may be designed to study the natural history of a disease, determine the clinical- or cost-effectiveness of a drug or device in real-world clinical practice, monitor drug safety, and/or measure quality of care. Registries can collect cross-sectional (i.e., short-term) or longitudinal (i.e., long-term) data, and often use electronic data capture tools to collect and manage their data.

This increase in the quantity of electronic health care and research data creates new opportunities and need for data linkage. Pharmaceutical companies conducting clinical trials on specific genetic markers are seeking ways to more easily identify and recruit potential patients. EHRs and patient registries are interfacing with each other to minimize the burden of data entry on participating centers and practices (see the “Interfacing Registries with Electronic Health Records” chapterⁱⁱ). Data from patient registries and other electronic sources are being pooled together to form larger, more statistically powerful datasets for research and analysis (see the “Linking Registry Data: Technical and Legal Considerations” chapter and the “Analytic Challenges in Studies That Use Administrative Databases or Medical Registries” chapter).

As more electronic health care data are generated and linked with each other, PIM has become crucial in order to (1) enable health record document consumers to obtain trusted views of their patient subjects, (2) facilitate data linkage projects, (3) abide by the current regulations concerning patient information-related transparency, privacy, disclosure, handling, and documentation,⁵ and (4) make the most efficient use of limited health care resources by reducing redundant data collection.

ⁱⁱ Chapters referenced in this document can be found in the second edition of “Registries for Evaluating Patient Outcomes: A User’s Guide,” available at: <http://www.effectivehealthcare.ahrq.gov/ehc/products/74/531/Registries%20nd%20ed%20final%20to%20Eisenberg%209-15-10.pdf>.

Technical Stakeholders

To address this growing need, many stakeholders are involved in the development of PIM strategies and standards. Several major stakeholders currently include: Integrating the Healthcare Enterprise (IHE); Health Level Seven International (HL7); The Regenstrief Institute, Inc.; and The Healthcare Information Technology Standards Panel (HITSP).

IHE is an initiative focused on using standards for interoperability between health care data systems to produce profiles, or implementation guides, which help organizations implement interoperability in a practical way. HIMSS, the Radiological Society of North America (RSNA), and the American College of Cardiology (ACC) sponsor IHE.⁶ In recent years, IHE has developed the Patient Identifier Cross Referencing (PIX) Integration Profile, which supports the cross-referencing of patient identifiers from multiple domains,⁷ and the Patient Demographics Query (PDQ) Integration Profile, which facilitates the querying of a patient database to retrieve demographics data.⁸

Health Level Seven International (HL7) is a non-profit organization dedicated to developing interoperability standards for electronic health information.⁹ HL7 has developed many of the industry standards currently being used by initiatives such as IHE, including the HL7 Version 2.x and Version 3 messaging standards.

The Regenstrief Institute, Inc. is an informatics and healthcare research organization and a joint enterprise of the Regenstrief Foundation, Inc., the Indiana University School of Medicine, and the Health and Hospital Corporation of Marion County. Regenstrief is active in developing healthcare informatics standards, including the widely-used Logical Observation Identifiers Names and Codes (LOINC®) terminology. It is currently conducting research for the Agency for Healthcare Research and Quality (AHRQ) entitled “Advancing Patient Identity Management in the Context of Real-World Health Information” which focuses on “creat[ing] a more robust and efficient global patient matching algorithm.”¹⁰

Patient Identity Management Strategies

The challenge of patient identity management is not a new one, and has existed since health care information was first digitized. In general, PIM is conducted in one of two environments: either shared identifiers are present or they are absent. When shared identifiers exist, the main PIM strategy that has emerged is to assign a unique patient identifier (UPI) to each patient. In situations where shared

identifiers do not exist, the most common PIM strategy is to use patient matching algorithms to determine whether two sets of information belong to separate patients or the same patient.

When Shared Identifiers are Present - Unique Patient Identifier

Definition and Context

One of the most straightforward PIM strategies is the creation of a unique health identifier for individuals, or a unique patient identifier (UPI). Generally, a UPI is defined as a “unique, non-changing alphanumeric key for each patient”¹¹ in a health care system, and which is associated with each medical record or instance of health care data for that patient. Some proposed desirable characteristics of a UPI include that it be unique, non-disclosingⁱⁱⁱ, invariable, canonical, verifiable, and ubiquitous.¹²

The concept of a *universal* UPI (i.e., a UPI that is assigned to a patient for life, and is consistent across all electronic healthcare systems in the U.S.) has been discussed and debated for a number of years. The Health Insurance Portability and Accountability Act (HIPAA) of 1996 called for the adoption of “standards providing for a standard unique health identifier for each individual, employer, health plan, and health care provider for use in the health care system.”¹³ Since the passage of HIPAA, the concept of a UPI has generally been welcomed by the health care industry, which views it as a tool to reduce administrative workload and increase efficiency in exchanging electronic health data.¹⁴ Other groups, including private citizens and experts attending a National Committee on Vital and Health Statistics hearing in July 1998, have expressed serious concerns about the effects that a universal UPI might have on patient privacy and data security.¹⁵ These concerns have halted further efforts at creating a UPI in the United States until appropriate privacy legislation is in place^{16,iv} even though recent research has shown that adoption of a universal UPI would actually strengthen patient privacy and security (by limiting the number of access points to patient health care data) and, while requiring a significant upfront cost, could pay for itself in cost savings from error reduction and administrative efficiency.¹⁷ The adoption of a

ⁱⁱⁱ In this context, “non-disclosing” means that the UPI does not contain any personal information about the patient, such as date of birth or social security number.

^{iv} Privacy and security concerns did not prevent CMS from developing the National Plan & Provider Enumeration System (NPPES) to assign unique identifiers to health plans and health care providers. The National Provider Identifier (NPI) has been implemented since 2006, and a standard identifier has not yet been implemented for health plans. (Available at: <https://nppes.cms.hhs.gov/NPPES/Welcome.do>. Last accessed on 21 October 2011.)

universal UPI is also viewed as the logical next step in strengthening and developing the national health information network.¹⁸

Current Uses of UPIs

UPIs have long been used within individual patient registries and datasets, especially those with prospective data collection, to track and link a particular patient's data over time. One of the most familiar types of UPI is a medical record number – a unique number assigned by a hospital or physician practice that links a patient with their medical record at that institution. Some hospitals have multiple electronic health information systems (e.g., EHRs, administrative/billing systems, lab systems, pharmacy dispensing systems) that assign UPIs to the patients within their domains, and a patient may not necessarily have the same UPI from system to system. Many patient registries also assign a UPI to patients upon screening or enrollment, and UPIs remain the simplest and most straightforward way to uniquely identify patients in a controlled dataset.

UPIs have also been used on a slightly larger scale in aggregated datasets and to link existing databases with administrative datasets. For example, the National Database for Autism Research (NDAR) aggregates data from many different collections of autism data and biospecimens and generates a global unique identifier (GUID) for each patient represented in the aggregated dataset.¹⁹ Similarly, in 2008 the Society of Thoracic Surgeons (STS) Database began collecting HIPAA-compliant unique patient, surgeon, and hospital identifier fields to facilitate long-term patient follow-up via linking to the Social Security Death Master File and the National Cardiovascular Data Registry.²⁰

Outside the United States, UPIs have been used on a wider scale. In Sweden, for example, the personal identity number (PIN) is a unique administrative identifier assigned to all permanent residents in Sweden since 1947. The PIN is used to track vital statistics and also link patients between several national-scale patient registries, including the Patient Register (containing inpatient and outpatient data), Cancer Register, Cause of Death Register, Medical Birth Register,²¹ and Knee Arthroplasty Register.²² In England, a new health identifier was introduced in 1996 – the NHS number is a 10-digit unique identifier used solely for the purpose of patient identification.²³

Future Directions for UPIs

Recently, interest in expanding the use of existing administrative identifiers (such as the Social Security Number in the United States) to serve as UPIs in the healthcare arena has increased. In 2009, the U.S.-based non-profit Global Patient Identifiers proposed the Voluntary Universal Healthcare Identifier

project, which aims to make unique healthcare identifiers available to any patient who uses the services of a regional health information organization (RHIO) or health information exchange (HIE).²⁴ In May 2011, production deployment on the system began. The voluntary nature of this project and its capacity for patients to have both an “open” voluntary identifier and a “private” voluntary identifier (which can be used to control which caregivers have access to clinically sensitive information) make it an interesting alternative to a mandated universal UPI that would likely be assigned and administered by a Federal government agency. In March 2011, the eCitizen Foundation began requirements-gathering work on the Patient Identity Service Project, an open-source, open standards-based patient identity service that will be able to identify and authenticate a patient across multiple systems to gain access to their health records and services.²⁵ The project is funded by the OpenID Foundation of Japan, and future goals include research and development, design, implementation, and testing of the service.

Registries and UPIs

UPIs offer a straightforward way to identify specific patients within a particular registry. A universal UPI would exponentially increase the value of patient registry data by allowing the linking of data across registries and other health care data sources. However, the implementation of a universal UPI in the United States has been halted by concerns over patient privacy, security, and confidentiality which are unlikely to be resolved soon.

In Sweden, the ability to link data from separate national patient registries using the PIN has allowed researchers to pull from a pool of millions of Swedish residents to address difficult epidemiological questions. Concerns about patient privacy and confidentiality have been addressed by requiring that an ethical review board review and approve the planned study before any data are released to researchers. Past precedent has been that the review boards allow most PIN-based registry linkages, on the condition that the PINs are removed from the combined dataset and replaced with different, unique serial numbers. Researchers also sign a legal agreement ensuring secure storage of the data and agreeing not to attempt to re-identify the patients in the de-identified dataset they are given.²⁶

When Shared Identifiers are Not Present - Patient Matching Algorithms

Definition and Context

In the absence of a national UPI in the United States, most researchers and hospital administrators have turned to patient matching algorithms and other statistical matching techniques as a way to manage patient identities within the confines of a specific patient registry, research project, institution, or other

grouping of healthcare data. This method of PIM involves comparing identifiable patient attributes (often demographics such as date of birth, gender, name, and address, but sometimes other individually-identifiable information) using a logic model which then classifies each pair as a match, a non-match, or a possible match that may require manual review.

In the realm of patient and record matching, algorithms can be either deterministic or probabilistic. Deterministic algorithms are more straightforward and classify a pair of records as a match if they meet a specified threshold of agreement. The definition of agreement can vary depending on which data elements are available, the quality and missingness of the data, and the desired sensitivity and specificity of the algorithm. Probabilistic algorithms treat the match status of individual data elements as observable variables and the match status of the record pair as a latent variable, and model the observable variables as a pattern mixture. This method characterizes the uncertainty in the matching process, making it a more sophisticated (and less straightforward) method than deterministic matching.²⁷

One major consideration in choosing an appropriate matching algorithm is the accuracy with which it matches patients. Matching accuracy is affected by the number of patients being compared, the number and type of common data elements being compared, and the mathematical validity of the algorithm itself. An algorithm that returns close to 100% matching in a pool of few patients with many data elements may perform less accurately in a pool of many patients with fewer data elements. Importantly, an algorithm that does not perform accurately may limit the conclusions and results able to be drawn from a particular dataset.

Current Uses of Patient Matching Algorithms

Patient matching algorithms are widely used when disparate healthcare data sources are combined and no unique, common patient identifier is available. The two main options are to use an existing record linkage software program or to develop a new matching algorithm independently. Commercial software options, such as Link Plus and The Link King, apply probabilistic algorithms that have been found to provide a higher sensitivity than matching using a basic deterministic algorithm.²⁸ As described in the case example “Integrating Data From Multiple Sources With Patient ID Matching,”^v an open-source

^v Case examples referenced in this document can be found in the second edition of “Registries for Evaluating Patient Outcomes: A User’s Guide.”

product (Febrl) was used to combine data from eleven different data sources into KIDSNET, a computerized registry that gives providers an overall view of childrens' use of preventative health services.²⁹

Many patient matching algorithms have been developed to meet the needs of specific projects. For example, a group at Partners HealthCare developed an algorithm to compare data in the Social Security Death Master File with demographic data in the Partners EHR system to identify patient deaths that may have occurred outside of Partners institutions (and therefore not recorded in the patients' medical record). They then developed another algorithm using clinical data to identify false-positives resulting from the first algorithm (e.g., if clinical data for a 'deceased' patient is recorded more than 30 days after the date of death in the SSDI, that patient must have been falsely matched to an SSDI entry).³⁰ In another example, researchers at the University of Alabama Birmingham used matching algorithms to link emergency medical services (EMS) data with hospital EHRs and a statewide death index to characterize the medical conditions and comorbidities of patients who receive out-of-hospital endotracheal intubation.³¹

New and innovative algorithms that are unrelated to specific projects also continue to be developed, with the goal of advancing patient matching algorithm science. Recent examples include algorithms proposed by groups at Vanderbilt University,³² John Radcliffe Hospital in the United Kingdom,³³ and the University of Duisburg-Essen in Germany.³⁴

Future Directions of Patient Matching Algorithms

Any statistical matching approach is dependent on three factors, which are listed below.

1. **The quality of the data it is comparing.** Are the data entered correctly, without mistakes? Are the data complete, or is there a high level of missingness? The quality of data within a particular registry will always be a factor of the practices employed by that registry. See the "Data Collection and Quality Assurance" chapter for recommended best practices.
2. **The comparability of the data it is comparing.** Are the data from the different sources collected in the same format and in the same way? There are a number of current initiatives to improve the standardization of data elements being used in patient registries,³⁵ but the area with the most need for future work is the testing and standardization of the algorithms themselves.
3. **The accuracy of the matching algorithm.** What is the likelihood of the algorithm returning a false positive match or missing true matches? While there has been some scientific research validating specific matching algorithms,^{36,37,38} the Health Information Technology Policy Committee recently called for increased standards around patient matching, including standardized formats for demographic data fields; internal evaluation of matching accuracy within institutions and projects; accountability to acceptable levels of matching accuracy; the

development, promotion, and dissemination of best practices in patient matching; and supporting the role of the patient.³⁹

Another emerging trend in patient matching algorithms is privacy-preserving record linkage, or “finding records that represent the same individual in separate databases without revealing the identity of the individuals”.³⁴ This concept was expanded upon by researchers at University of Duisburg-Essen in Germany, mentioned in the previous section, who propose a method which encrypts patient identifiers while allowing for errors in identifiers. Given the concerns about patient privacy and confidentiality surrounding patient identity management, this method may be increasingly used in the future.

Registries and Patient Matching Algorithms

As mentioned above, patient matching algorithms have become the default PIM strategy for registries that link with outside data sources, due to the lack of a universal UPI in the United States. As a result, many different algorithms have been developed – some are commercially available, some are open-source; some were developed for specific projects and some were developed with broader applications in mind. The performance and effectiveness of matching algorithms can impact the results produced by the registries that are using them. The type of registry also impacts the type of patient matching algorithm needed. Registries used for direct patient care may require an algorithm with different sensitivity, specificity, and timeliness than registries used for population-based research efforts. Registry owners and operators would benefit from standards surrounding patient matching algorithms, which would allow them to more confidently and effectively use appropriate algorithms for linking projects.

Emerging Strategies and Related Ideas

In addition to a universal patient identifier and patient matching algorithms, other strategies are emerging to manage patient identities in disparate electronic health care data sources. These include biometrics, master patient indexes, and health information exchanges.

Biometrics

One new option in the PIM field is the use of biometrics – that is, “automated methods of recognizing an individual based on measurable biological (anatomical and physiological) and behavioral characteristics.”⁴⁰ Some examples of biometric measurements are: fingerprint, palm print, hand geometry, DNA, handwriting, finger or hand vascular pattern, iris/retina, facial shape, voice pattern and gait.

Biometrics are attractive because of their difficulty to fabricate, their resistance to change over time (unlike demographic information such as name and address), and their high degree of uniqueness – making them effectively biological UPIs. For biometrics to be used as UPIs, though, there would need to be agreement on which biometric to use and the format in which it should be collected. Also, some biometric measurements are more unique than others. For example, a fingerprint is highly unique to an individual while a person's hand geometry is not as unique. Hand geometry therefore is often used to confirm a person's identity (i.e., in combination with another identifier) rather than as a sole identifier.

One drawback to using biometrics is the investment in specialized technology and equipment that is required to capture many of these measurements. There is also concern about the privacy and security implications surrounding using biometrics, connected with their history of use in law enforcement and their potential misuse to derive information other than identity (e.g., analyzing DNA for genetic diseases).⁴¹

Some hospitals have begun using biometrics to verify provider identity and restrict access to EHRs. Biometrics are also being used in some hospitals to verify patient identity upon hospital admission⁴² and to identify critically injured, unconscious patients presenting to an emergency room.⁴³

Registries and Biometrics

Many registries, particularly those with biobanks associated with them, already collect biometric data (e.g., DNA). However, the data are often used for purposes other than PIM, including investigating genetic components of disease⁴⁴ and risk factors for disease.⁴⁵

Biometrics remains an attractive option for PIM; the largest obstacle to their use in patient registries is likely the investment in technology and equipment that they require, although this would vary depending on where registry data are collected. A multi-site, practice-based registry would probably be less able to accommodate the collection of biometrics, while a registry based out of a single hospital that already collects biometric data for other purposes would be able to begin collecting biometrics for a registry more easily, since the initial investment in technology has already been made. Registries using biometrics would also be subject to the same concerns about privacy and security as biometric use in other disciplines.

Master Patient Index

A master patient index (MPI) is an index that facilitates the identification and linkage of patients' clinical information within a particular institution. The term "enterprise master patient index" (EMPI) is sometimes used to distinguish between an index that serves a single institution (i.e., MPI) and one that contains data from multiple institutions (i.e., EMPI). MPIs are not themselves patient identity management strategies, but rather informational infrastructures within which those strategies are applied. Most MPIs use a patient matching algorithm to identify matches and then assign a UPI that is associated with that patient record going forward. MPIs and EMPIs are created for the purpose of assigning a UPI to each patient treated within a certain healthcare system – providers can then use that identifier to have a global view of the patient's care across multiple institutions within that system.

Several leading software companies have released commercially-available MPI and EMPI products. Oracle has published a thorough description of the design and functionality of their EMPI product.⁴⁶ Open-source options are also available, including one developed by Project Kenai called OpenEMPI.⁴⁷

EMPIs are used as supplemental tools to apply PIM strategies for data sharing efforts such as Health Information Exchanges (HIEs, described in the next section). For example, the Michigan Clinical Research Collaboratory at the University of Michigan created the "Honest Broker" system which serves three functions: facilitating the actual exchange of data between members of the collaboratory for research, maintaining an MPI to manage patient identities within that data, and de-identifying datasets in conformance with HIPAA limited dataset standards.⁴⁸

Figure 1 is adapted from the IHE integration profile⁴⁹ and illustrates the actors that participate in the Patient Identifier Cross-referencing profile. The entity often called an MPI is represented by the combination of the Patient Identity Source ("Source") and the Patient Identity Cross-reference Manager ("Manager"). The Source provides patient identity information (Patient Identity Feed) to the Manager. It is common to have multiple patient identity sources which provide patient ID feeds to the Manager. The Manager is responsible for managing patient identities by detecting matches and creating and maintaining cross-references of patient identifiers across these various sources. The Patient Identifier Cross-reference Consumer ("Consumer") retrieves Patient Identity Cross References or aliases. This allows patients to be linked across multiple systems or domains that use different patient identifiers to represent the same patient.

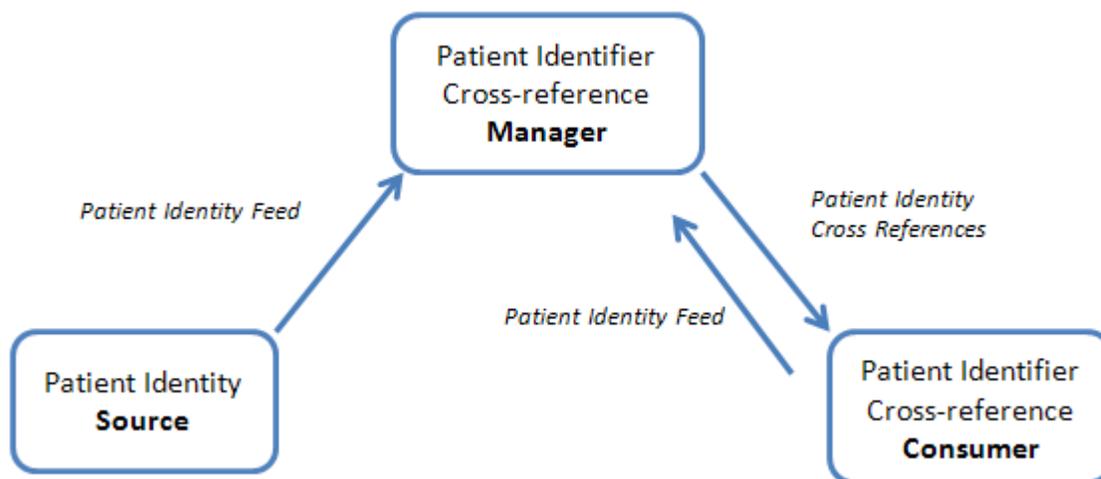


Figure 1: Basic Process Flow with Patient Identifier Cross-referencing

Illustrating how users may interact with an MPI in daily practice may be helpful. In one possible scenario, an emergency room physician sees a patient presenting at the emergency room with vague and poorly-defined pain who specifically asks to be prescribed narcotics. A new quality improvement program being implemented in this emergency room requires the physician to check the patient's history of filling prescriptions before issuing a prescription for a narcotic drug. The emergency room's EHR system and the hospital pharmacy's electronic dispensing record system each assign their own patient IDs to patients within their systems, and send patient feeds to the hospital's MPI (the Manager in this scenario) each time a new patient ID is assigned. The MPI creates and maintains cross-references of all identifiers for patients and provides the cross-references to consumers who seek that information. The consumer in this scenario would be the emergency room system which sends the MPI a patient identity cross-reference or demographic query with information about the patient in question. The MPI notifies the emergency room system that the patient identified in the emergency room as "ER703" matches the patient whose pharmacy records are under the pharmacy system identifier "012". The emergency room system then queries the pharmacy system for the identifier "012", and presents the dispensing record data to the emergency room physician.

Registries and MPIs

Health care institutions that utilize MPIs to manage patient identities across their multiple data sources (e.g., EHRs, pharmacy records, administrative and billing records) are desirable partners for data linkage projects and for inclusion in patient registries, since they are able to draw from a broader pool of data than

any one of the data sources alone. By addressing PIM needs upfront, they minimize the work needed for outside sources to link to their data for research uses.

In the relational infrastructure shown in Figure 1, registries can act as Patient Identity Sources, Patient Identifier Cross-reference Consumers, or both. Registries that contain patient identifiers and other demographic information can act as Patient Identity Sources and send patient identity feeds to a Manager. Registries can also act as Patient Identifier Cross-reference Consumers, if they request and receive patient identity cross references from an MPI or other Patient Identity Cross-reference Manager. This may be done to add new patients to a registry or to augment existing data in a registry with additional information on the same patients.

Health Information Exchange

A Health Information Exchange (HIE) is an integrated platform solution to enable information sharing across disparate healthcare applications. (See the “Technical and Security Issues in Creating a Health Information Exchange” case example, which describes the Oakland Southfield Physicians HIE.) HIEs are sometimes referred to as “databanks”, since (unlike MPIs) they contain actual patient data. HIEs are not themselves patient identity management strategies, but they implement those strategies to manage their data. Most HIEs achieve this by partnering with an MPI to manage the identity of patients within the HIE.

HIEs can be powerful research tools. A group at the Swansea University School of Medicine has developed the Secure Anonymized Information Linkage (SAIL) databank, containing over 500 million records from multiple health and social care service providers in the UK.⁵⁰ The SAIL databank has already been used to demonstrate the feasibility of identifying potential clinical trial participants at the primary care level, which may be especially useful for disease areas in which recruitment of clinical trial participants is historically difficult (e.g., chronic conditions such as diabetes).⁵¹

Because they contain patient data, HIEs are subject to the same privacy and security concerns and regulations as patient registries. A white paper published in April 2011 by the AHIMA/HIMSS HIE Privacy & Security Joint Work Group provides a summary of these considerations.⁵²

Registries and HIEs

A patient registry may contribute data to an HIE, but registries and HIEs are distinct and separate endeavors. Data contained in HIEs are not necessarily collected using observational study methods, as

patient registry data are – rather, they are often collected and aggregated by linking to existing databases (which may be, for example, registries, administrative databases, or public health surveillance systems). The purpose of an HIE is not just to evaluate specified outcomes in a defined patient population or even to serve any one predetermined scientific, clinical, or policy purpose, but to provide an aggregated database that can be used for a variety of purposes (which may include identifying patients to recruit for clinical trials, conducting ecological studies, etc).

Major Challenges and Barriers

The process of patient identity management introduces several technical, ethical, and operational challenges, including selecting the appropriate PIM strategy, discussed earlier in this paper. Additional challenges include the obligation to protect the privacy and security of patient data and the technical interoperability (or lack thereof) of disparate health care data sources.

Protecting Patient Privacy and Security

One of the most pressing challenges in PIM is addressing the tension between linking patient data in order to manage their identities and protecting the privacy and security of those data. This challenge has inherent ethical, regulatory, and technical considerations.

Ethical and Regulatory Considerations

The concepts of protecting patient privacy and security and PIM have always been intertwined. Managing patient identities is essential for protecting the privacy and security of those patients (i.e., in order to protect someone's information, one needs to first know who they are and which information is theirs). Conversely, regulations and ethical considerations compel the protection of patients' privacy and security when managing their identities (i.e., it is not enough to know who they are and which information is theirs, one must also protect this information).

Many stakeholders in the health information technology field recognize this relationship. The Health Information Security and Privacy Collaboration (HISPC) count patient and provider identification as one of their nine domains of privacy and security.⁵³ The Commission on Systemic Interoperability released a report in 2005 in which they recommended that Congress authorize the Department of Health and Human Services to “develop a national standard for determining patient authentication and identity,” and “develop a uniform federal health information privacy standard for the nation, based on HIPAA and pre-

empting state privacy laws [...]”. These recommendations were made simultaneously, “to advance progress of the connectivity of health information technology.”⁵⁴ Thus, it is widely recognized that PIM and patient privacy and security are closely related, but there continues to be disagreement about *how they should* relate.

The regulatory framework that guides this discussion is the Health Insurance Portability and Accountability Act (HIPAA), enacted in 1996. As mentioned previously in this paper, HIPAA mandated the implementation of a nationwide unique patient identifier, but concerns about patient privacy and security prompted the barring of any funding for this endeavor, in 1999. While HIPAA has not led to the implementation of a standard PIM method, it does set forth a framework for the protection of patient privacy and health information security. This framework is summarized in Table XX, Chapter XX, “Principles of Registry Ethics, Data Ownership, and Privacy.”

Technical Considerations

Data holders employ three main technical methods of ensuring the privacy and security of patient data: anonymization, encryption, and pseudonymization.

Anonymization

Anonymization is the practice of removing information that is identifiable to an individual, or that may enable an individual’s identity to be deduced. This is a viable option in some data use situations (e.g., conducting a research study that does not require patient follow-up), but not an option in others (e.g., maintaining comprehensive health records for patients in an EHR). It is also not a reversible process – once identifiers are removed from data, they cannot be reinserted.

Encryption

Encryption involves applying a mathematical calculation or algorithm to transform a patient’s original data (plain text) into coded data (cypher text). In order to read the cypher text, a user or system must have access to a key that de-encrypts the data back into plain text. This is an attractive option because it does not involve deleting or removing patient data, and because the coded data is not in a readable format if it falls into the wrong hands. However, encryption requires robust data management policies and resources to implement successfully.⁵⁵

Pseudonymization

Pseudonymization is a more sophisticated approach to patient privacy protection. It involves two steps: depersonalization, where identifiable data is separated from other clinical data and stored in a separate location, and pseudonymization, where a unique identifier is generated and applied to the depersonalized dataset. The unique identifier, or pseudonym, does not change for a given patient over time, and is not derived from any identifiable attributes of the patient. Pseudonymization can be reversible, if the relationship between the pseudonym and the identifiable data is maintained in a secure way and can facilitate re-identification of the patient under specific circumstances (e.g., a trusted third party maintains the relationship, and only discloses that relationship if the requestor has knowledge of a particular key or password). Pseudonymization can also be irreversible, where the relationship between the pseudonym and the identifiable data is not maintained, and re-identification is not possible.^{56,57}

Interoperability

In the same way that healthcare enterprises such as hospitals, clinics, and physician offices require patient identifier cross-referencing, that is the linking of patients across different domains, it is necessary to consider how registries may fit within this model and the challenges that level of interoperability may impose. Separate patient registries may use the same PIM infrastructure to register their patient identifiers within a shared patient identifier cross-reference manager, allowing the identifiers to be linked back to relevant healthcare and related systems. This approach may represent a possible solution whereby registries can more easily and securely be linked to other systems across known domains such as an HIE, but challenges still remain in terms of how this approach could successfully be used more broadly across non-participating healthcare enterprises.

Conclusion

Patient identity management is a fast-growing and evolving field, influenced by emerging technologies, regulations, and opportunities to use electronic health care data. The current status of PIM in the United States is primarily a factor of the provision in HIPAA for “standards providing for a standard unique health identifier for each individual [...] for use in the health care system,”⁵⁸ the debate this provision has generated over implications for patient privacy and security, and the subsequent blocking of any funding being allocated to the pursuit of a national UPI. As a result, most PIM endeavors in the U.S. (including attempts to link patient registries with other health care data sources) utilize patient matching algorithms

to identify duplicates and manage patient identities. The lack of standards in this area means that the accuracy and effectiveness of these algorithms can vary widely.

Debate continues around how to best address the challenge of PIM, and stakeholders generally hold one of two views. Some view a national UPI as the best solution, provided the long-standing concerns about protecting patient privacy and security can be adequately addressed in the future. Others believe that resources would be better spent developing and standardizing the PIM methods that have grown organically in the absence of a national UPI – namely, EMPIs and patient matching algorithms. These two endeavors are not necessarily mutually exclusive, and patient registries and data linkage projects would benefit from the advancement of either or both.

References

- ¹ North Carolina Health & Wellness Trust Fund Commission. North Carolina Health Information Exchange Strategic Plan. Available at: https://www.ncrecover.gov/library/pdf/NC_InformationExchangeStrategicPlanNarrative.pdf. Last accessed on 17 August 2011.
- ² NorthPage Research LLC. 5 Tips for Successful Patient Identity Management in Government Agencies. Available at: <http://govhealthit.com/sites/default/files/resource-media/pdf/northpagereportpatientidentitymanagementtipsforgovtagencies.pdf>. Last accessed on 17 August 2011.
- ³ Hsiao CJ, Hing E, Socey TC, Cai, B. Electronic Medical Record/Electronic Health Record Systems of Office-based Physicians: United States, 2009 and Preliminary 2010 State Estimates. Centers for Disease Control and Prevention, National Center for Health Statistics. December 2010. Available at: http://www.cdc.gov/nchs/data/hestat/emr_ehr_09/emr_ehr_09.htm. Last accessed on 8 August 2011.
- ⁴ Centers for Medicare and Medicaid Services. CMS EHR Meaningful Use Overview. Available at: https://www.cms.gov/ehrincentiveprograms/30_Meaningful_Use.asp. Last accessed on 16 September 2011.
- ⁵ NorthPage Research LLC. 5 Tips for Successful Patient Identity Management in Government Agencies. Available at: <http://govhealthit.com/sites/default/files/resource-media/pdf/northpagereportpatientidentitymanagementtipsforgovtagencies.pdf>. Last accessed on 17 August 2011.
- ⁶ HIMSS. "Integrating the Healthcare Enterprise (IHE)." Available at: http://www.himss.org/ASP/topics_ihe.asp. Last accessed on 8 August 2011.
- ⁷ Integrating the Healthcare Enterprise. Patient Identifier Cross-Referencing. Available at: http://wiki.ihe.net/index.php?title=Patient_Identifier_Cross-Referencing. Last accessed on 17 August 2011.
- ⁸ Integrating the Healthcare Enterprise. Patient Demographics Query. Available at: http://wiki.ihe.net/index.php?title=Patient_Demographics_Query. Last accessed on 17 August 2011.
- ⁹ Health Level Seven International. "About Health Level Seven International." Available at: <http://www.hl7.org/about/index.cfm?ref=nav>. Last accessed on 8 August 2011.
- ¹⁰ Regenstrief Institute, Inc. "Medical Informatics Projects." Available at: <http://www.regenstrief.org/medinformatics/projects>. Last accessed on 8 August 2011.
- ¹¹ Hillestad R, Bigelow JH, Chaudhry B, Dreyer P, Greenberg MD, Meili RC, Ridgely MS, Rothenberg J, Taylor R. 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/content/dam/rand/pubs/monographs/2008/RAND_MG753.pdf. Last accessed on 29 June 2011.
- ¹² American Society for Testing and Materials (ASTM). Standard Guide for Properties of a Universal Healthcare Identifier (UHID). Available at: <http://www.astm.org/Standards/E1714.htm>. Last accessed on 8 August 2011.
- ¹³ Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191 Sec. 1173(b) (August 21, 1996).
- ¹⁴ National Committee on Vital and Health Statistics (NCVHS), Subcommittee on Standards and Security. Hearing Minutes. July 20-21, 1998, Chicago, IL. Available at: <http://ncvhs.hhs.gov/980720mn.htm>. Last accessed on 15 September 2011.
- ¹⁵ National Committee on Vital and Health Statistics (NCVHS), Subcommittee on Standards and Security. Hearing Minutes. July 20-21, 1998, Chicago, IL. Available at: <http://ncvhs.hhs.gov/980720mn.htm>. Last accessed on 15 September 2011.
- ¹⁶ Omnibus Consolidated and Emergency Supplemental Appropriations Act of 1999, Pub. L. No. 105-277 112 Stat. 2681-386.
- ¹⁷ Greenberg M, Ridgely M. Patient Identifiers and the National Health Information Network: Debunking a False Front in the Privacy Wars. *J Health Biomed Law*. 2008;4(1):31-68.
- ¹⁸ Hillestad R, Bigelow JH, Chaudhry B, Dreyer P, Greenberg MD, Meili RC, Ridgely MS, Rothenberg J, Taylor R. 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/content/dam/rand/pubs/monographs/2008/RAND_MG753.pdf. Last accessed on 29 June 2011.
- ¹⁹ Johnson SB, Whitney G, McAuliffe M, Wang H, McCreedy E, Rozenblit L, et al. Using global unique identifiers to link autism collections. *J Am Med Inform Assoc*. 2010 Nov 1; 17(6):689-95.
- ²⁰ Jacobs JP, Haan CK, Edwards FH, Anderson RP, Grover FL, Mayer JE, Jr., et al. The rationale for incorporation of HIPAA compliant unique patient, surgeon, and hospital identifier fields in the STS database. *Ann Thorac Surg*. 2008 Sep; 86(3):695-8.
- ²¹ Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekblom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009; 24(11):659-67.
- ²² Robertsson O, Dunbar M, Knutson K, Lewold S, Lidgren L. Validation of the Swedish Knee Arthroplasty Register: a postal survey regarding 30,376 knees operated on between 1975 and 1995. *Acta Orthop Scand*. 1999 Oct; 70(5):467-72.
- ²³ National Health Service. "Records – The NHS Number." Available at: <http://www.nhs.uk/NHSEngland/thenhs/records/Pages/thenhsnumber.aspx>. Last accessed on 24 October 2011.
- ²⁴ Global Patient Identifiers, Inc. VUHID System. Available at: <http://gpii.info/system.php>. Last accessed on 9 August 2011.
- ²⁵ eCitizen Foundation. Patient ID Services Project Contact Page – Citizen Centered Solutions. Available at: http://civics.typepad.com/files/pids_overview_deliverable_march_3_2011-v1.pdf. Last accessed on 19 August 2011.
- ²⁶ Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekblom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009; 24(11):659-67.
- ²⁷ Li X, Shen C. Linkage of patient records from disparate sources. *Stat Methods Med Res*. 2011 Jun 10.
- ²⁸ Campbell KM, Deck D, Krupski A. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a 'basic' deterministic algorithm. *Health Informatics J*. 2008 Mar; 14(1):5-15.
- ²⁹ Wild EL, Hastings TM, Gubernick R, Ross DA, Fehrenbach SN. Key elements for successful integrated health information systems: lessons from the States. *J Public Health Manag Pract*. 2004 Nov; Suppl:S36-47.
- ³⁰ Turchin A, Shubina M, Murphy SN. I am Not Dead Yet: Identification of False-Positive Matches to Death Master File. *AMIA Annu Symp Proc*. 2010; 2010:807-11. Available at: <http://proceedings.amia.org/127h5i/127h5i/1>. Last accessed on 26 August 2011.
- ³¹ Wang HE, Balasubramani GK, Cook LJ, Yealy DM, Lave JR. Medical conditions associated with out-of-hospital endotracheal intubation. *Prehosp Emerg Care*. 2011 Jul-Sep; 15(3):338-46.
- ³² Durham E, Xue Y, Kantarcioglu M, Malin B. Private medical record linkage with approximate matching. *AMIA Annu Symp Proc*. 2010; 2010:182-6.
- ³³ Finney JM, Walker AS, Peto TE, Wyllie DH. An efficient record linkage scheme using graphical analysis for identifier error detection. *BMC Med Inform Decis Mak*. 2011; 11:7.
- ³⁴ Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. *BMC Med Inform Decis Mak*. 2009; 9:41.
- ³⁵ Agency for Healthcare Research and Quality. Developing a Registry of Patient Registries (RoPR). Available at: <http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=690>. Last accessed on 17 August 2011.
- ³⁶ Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC, et al. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol*. 2008 Dec 1; 168(11):1326-32.
- ³⁷ Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol*. 2007 Sep; 60(9):883-91.
- ³⁸ Alemi F, Loaiza F, Vang J. Probabilistic master lists: integration of patient records from different databases when unique patient identifier is missing. *Health Care Manag Sci*. 2007 Feb; 10(1):95-104.
- ³⁹ U.S. Department of Health & Human Services. The Office of the National Coordinator for Health Information Technology. Health IT Policy Committee: Recommendations to the National Coordinator for Health IT.

Transmittal Letter. February 8, 2011. Available at:

http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov_policy_recommendations/1815. Last accessed on 18 August 2011.

⁴⁰ National Science and Technology Council. Biometrics Glossary. Available at:

<http://www.biometrics.gov/Documents/Glossary.pdf>. Last accessed on 10 August 2011.

⁴¹ Prabhakar S, Pankanti S, Jain A. Biometric Recognition: Security and Privacy Concerns. *IEEE Security & Privacy*. 2003 March/April; 1(2):33-42.

⁴² Lawrence S. Biometrics bring fingerprint ID to hospitals. *CIO Insight*. 2005 Mar 24. Available at:

<http://www.cioinsight.com/c/a/Health-Care/Biometrics-Bring-Fingerprint-ID-to-Hospitals/>. Last accessed on 10 August 2011.

⁴³ Marohn D. Biometrics in healthcare. *Biometric Technology Today*. 2006 Sep;14(9):9-11.

⁴⁴ Rasmussen A, Sevier S, Kelly JA, Glenn SB, Aberle T, Cooney CM, et al. The lupus family registry and repository. *Rheumatology (Oxford)*. 2011 Jan;50(1):47-59.

⁴⁵ Wolf EJ, Miller MW, Krueger RF, Lyons MJ, Tsuang MT, Koenen KC. Posttraumatic stress disorder and the genetic structure of comorbidity. *J Abnorm Psychol*. 2010 May;119(2):320-30.

⁴⁶ Ouaguenouni S, Sivaraman K, Braun T. Identity Resolution and Data Quality Algorithms for Master Person Index: An Oracle White Paper. Available at: <http://www.oracle.com/us/industries/healthcare/identity-resolution-algorithm-wp-171743.pdf>. Last accessed on 19 August 2011.

⁴⁷ OpenEMPI. An Open Source Enterprise Master Patient Index. Available at: <http://openempi.kenai.com/>. Last accessed on 24 August 2011.

⁴⁸ Boyd AD, Saxman PR, Hunscher DA, Smith KA, Morris TD, Kaston M, et al. The University of Michigan Honest Broker: a Web-based service for clinical and translational research and practice. *J Am Med Inform Assoc*. 2009 Nov-Dec;16(6):784-91.

⁴⁹ Integrating the Healthcare Enterprise (IHE). IHE IT Infrastructure (ITI) Technical Framework. Volume 1 (ITI-TF1) Integration Profiles. Available at: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Rev8-0_Vol1_FT_2011-08-19.pdf. Last accessed on 8 September 2011.

⁵⁰ Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak*. 2009;9:3.

⁵¹ Brooks CJ, Stephens JW, Price DE, Ford DV, Lyons RA, Prior SL, et al. Use of a patient linked data warehouse to facilitate diabetes trial recruitment from primary care. *Prim Care Diabetes*. 2009 Nov;3(4):245-8.

⁵² Durkin S, Sullivan C, et al. The Privacy and Security Gaps in Health Information Exchanges. Available at: http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_049023.pdf. Last accessed on 19 August 2011.

⁵³ Dimitropoulos L, Alakoye A, Anderson H, Apgar C, Banger A, et al. Privacy and Security Solutions for Interoperable Health Information Exchange: Nationwide Summary. 2007. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://healthit.ahrq.gov/portal/server.pt/community/ahrq-funded_projects/654/outcomes_from_the_privacy_and_security_solutions_for_interoperable_health_information_exchange_project/24069. Last accessed on 25 August 2011.

⁵⁴ Commission on Systemic Interoperability. *Ending the Document Game: Connecting and Transforming Your Healthcare Through Information Technology*. 2005. Washington, DC: U.S. Government Printing Office. Available at: <http://endingthedocumentgame.gov/PDFs/entireReport.pdf>. Last accessed on 25 August 2011.

⁵⁵ Miller AR, Tucker CE. Encryption and the loss of patient data. *J Policy Anal Manage*. 2011 Summer;30(3):534-56.

⁵⁶ Noumeir R, Lemay A, Lina JM. Pseudonymization of radiology data for research purposes. *J Digit Imaging*. 2007 Sep;20(3):284-95.

⁵⁷ Neubauer T, Heurix J. A methodology for the pseudonymization of medical data. *Int J Med Inform*. 2011 Mar;80(3):190-204.

⁵⁸ Health Insurance Portability and Accountability act of 1996, Pub. L. No. 104-191 Sec. 1173(b) (August 21, 1996).