PATIENT IDENTIFICATION AND MATCHING FINAL REPORT

February 7, 2014 Prepared for the Office of the National Coordinator for Health Information Technology under Contract HHSP233201300029C by: Genevieve Morris, Director, Audacious Inquiry Greg Farnum, Director, Audacious Inquiry Scott Afzal, Principal, Audacious Inquiry Carol Robinson, Principal, Robinson & Associates Consulting LLC Jan Greene, Associate, Robinson & Associates Consulting LLC Chris Coughlin, Associate, Robinson & Associates Consulting LLC



Audacious Inquiry, LLC 5523 Research Park Drive, Suite 370 Baltimore, MD 21228 301-560-6999



Table of Contents

EXECUTIVE SUMMARY	2
INTRODUCTION	5
OVERVIEW OF ENVIRONMENTAL SCAN	5
FINDINGS	15
APPENDIX A: LITERATURE REVIEW	24
APPENDIX B: BEST PRACTICES FOR DATA QUALITY	34
APPENDIX C: ENVIRONMENTAL SCAN: HEALTH SYSTEMS	44
APPENDIX D: HEALTH INFORMATION ORGANIZATIONS (HIOS)	50
APPENDIX E: EHR VENDORS	59
APPENDIX F: MDM/MPI/HIE VENDORS	63
APPENDIX G: ASSOCIATIONS AND OTHER ORGANIZATIONS	71
APPENDIX H: GLOSSARY	82
APPENDIX I: STRUCTURED INTERVIEW QUESTIONS	87

This report was created by Audacious Inquiry, LLC under a contract with the Office of the National Coordinator for Health Information Technology (ONC). The content, views, and opinions do not necessarily reflect those of the Department of Health and Human Services or ONC.



Executive Summary

The United States healthcare system is marching diligently toward a more connected system of care through the use of electronic health record systems (EHRs) and electronic exchange of patient information between organizations and with patients and caregivers. The Patient Identification and Matching Initiative, sponsored by the Office of the National Coordinator for Health Information Technology (ONC), focused on identifying incremental steps to help ensure the accuracy of every patient's identity, and the availability of their information wherever and whenever care is needed.

Matching records to the correct person becomes increasingly complicated as organizations share records electronically using different systems, and in a mobile society where patients seek care in many healthcare settings. Many healthcare organizations use multiple systems for clinical, administrative, and specialty services, which leads to an increased chance of identity errors when matching patient records. Additionally, many regions experience a high number of individuals who share the exact name and birthdate, leading to the need for additional identifying attributes to be used when matching patient records.

While many hospitals, health systems, medical groups, and other organizations serving the health care industry have developed ways to improve the accuracy of matching patient records, these methods have not been adopted uniformly across the industry. For instance, differences in how names and addresses are formatted in various systems has led to high rates of unmatched records, when unaffiliated organizations are participating in health information exchange (HIE). Other issues and circumstances that lead to unmatched or mismatched records include the quality of data as it is entered into systems at the source of patient registration, and the creation of duplicate records for the same patient within a system.

Driven by concerns for patient safety in the event of mismatched or unmatched records and the national imperative to improve population health and lower costs through care coordination, this initiative studied both technical and human processes, seeking improvements to patient identification and matching that could be quickly implemented and lead to near-term improvements in matching rates. The findings presented in this report have been developed with the participation of healthcare stakeholders across many sectors and a wide geography. Lessons were drawn from individuals and organizations working to improve patient identification and matching edge of technology, as well as from those taking a systematic approach to improving data quality within the technical systems storing patient data, through review and rework of business processes. The combination of those approaches has contributed to findings that suggest the standardization of specific demographic fields within health information systems, broad collaboration on industry best practices that could both inform policy and be shared nationally, and areas for further study where additional advances could be made in the future.

Building on existing patient matching strategy work of ONC, from June through November 2013 an <u>environmental scan</u> was performed to assess the current industry capabilities and best practices for patient identification and matching, with a focus on matching records between disparate



organizations providing care for an individual.¹ The scan included a formal interview process with a diverse set of organizations representing four sectors of the healthcare industry, with a specific set of questions for large <u>health systems and integrated delivery networks (IDNs); health</u> information organizations (HIOs); EHR vendors; and <u>master data management (MDM)/master</u> person index (MPI) and HIE vendors. Informal informational discussions were also held with a wide variety of associations representing patients, providers, hospitals, vendors, health information management workforce, and organizations maintaining patient registries for public health and research purposes, as well as several federal agencies. In total, more than 50 organizations participated in the formal and informal interviews, which are summarized in Appendices C through F.

An intensive review of historical literature on patient matching was conducted, a summary of which can be found in <u>Appendix A</u>. In addition, two in-person meetings were held during which comments and advice were sought. The first of those meetings was sponsored by the Bipartisan Policy Center and held in early November, 2013, while the second was a public meeting, held in mid-December 2013, in Washington, D.C. and accessible via the web. At the second meeting, feedback was sought from over 150 stakeholders on a preliminary version of proposed findings. Written feedback from the industry regarding the initial findings was solicited after the December 2013 meeting, and additional outreach was conducted to help inform <u>Appendix B: Best Practices for Data Quality</u>.

The patient identification and matching initiative was led by four guiding principles: 1) patient safety is the driving force for improvement in patient matching, 2) the real-world impacts on the workflow of administrative and clinical personnel must be carefully considered, 3) patient matching is a complex problem, and therefore, improvements will be multifaceted and incremental with no single solution or step that is final, and 4) potential improvements should apply all sizes and types of provider settings, a range of health IT adoption levels, and a broad set of "use cases." The guiding principles, literature review, environmental scan, feedback received at each stakeholder meeting, and the written comments received after each meeting, informed the findings.

Findings

- 1. Standardized patient identifying attributes should be required in the relevant exchange transactions.
- 2. Any changes to patient data attributes in exchange transactions should be coordinated with organizations working on parallel efforts to standardize healthcare transactions.
- 3. Certification criteria should be introduced that require certified EHR technology (CEHRT) to capture the data attributes that would be required in the standardized patient identifying attributes.
- 4. The ability of additional, non-traditional data attributes to improve patient matching should be studied.
- 5. Certification criteria should not be created for patient matching algorithms or require organizations to utilize a specific type of algorithm.

¹ ONC formally launched the Patient Matching Initiative in September 2013.



- 6. Certification criteria that requires CEHRT that performs patient matching to demonstrate the ability to generate and provide to end users reports that detail potential duplicate patient records should be considered.
- 7. Build on the initial best practices that emerged during the environmental scan by convening industry stakeholders to consider a more formal structure for establishing best practices for the matching process and data governance.
- 8. Work with the industry to develop best practices and policies to encourage consumers to keep their information current and accurate.
- 9. Work with healthcare professional associations and the Safety Assurance Factors for EHR Resilience (SAFER) Guide initiative to develop and disseminate educational and training materials detailing best practices for accurately capturing and consistently verifying patient data attributes.
- 10. Continue collaborating with federal agencies and the industry on improving patient identification and matching processes.



Introduction

Over the last few years, the healthcare industry has been working towards meeting the Triple Aim of improving patient care, improving population health, and reducing the cost of healthcare.² Accurately identifying patients and correctly matching their data, particularly as electronic data becomes more prevalent, is a key component to ensuring the Triple Aim can be met. If patient data is not accurately matched within and across organizations, treatment and diagnosis decisions will be made in the absence of valuable information and patients could be subject to adverse events and significant harm. To facilitate the best possible treatment and care coordination among a care team, exchanging and linking patient data from each point of care is critical. In order to improve the efficiency of the healthcare system, duplicative testing, treatment decisions made in the absence of data, and inaccurate billing must be reduced; none of which can be accomplished without correctly identifying patients at the point of care and accurately matching their records as they are shared across organizations. Current methods to match patient records cannot achieve a zero percent error rate in which every possible match is correctly made and erroneous matches are avoided. Indeed, no single solution can accomplish this feat given the underlying contributors to the challenge of accurate record linking. However, the environmental scan and the December 2013 stakeholder meeting indicate that a multi-faceted approach to improve patient matching can make incremental progress to improve the current average match rates in the industry.

Overview of Environmental Scan

Objective

The environmental scan of patient matching practices describes current practices and trends in the health care industry, particularly pertaining to sharing clinical information. Health care organizations and their technology vendors continue to implement varying patient matching processes and technology. This lack of commonality poses a threat to patient safety and organizational efficiency, and both problems are likely to be magnified as HIE activities continue to grow in various forms across the country. The goal of the environmental scan is to document what is happening among larger health systems and HIOs throughout the United States, the approaches software vendors are taking to address customer needs, and whether there are best practices that could be gleaned from thoughtful organizations. The scan is only a snapshot of organizations rather than an academic survey. However, special emphasis was placed on gathering information from as wide a variety of stakeholders as possible within the time constraints of the initiative.

Choice of Participants

Interviewees selected for the structured portion of the scan were organized into four industry categories to provide varied perspectives on patient matching. The categories included health systems, HIOs, MDM/MPI solution vendors, HIE vendors, and inpatient or outpatient EHR vendors. Up to nine participants (hereafter referred to as organizations) were chosen in each category for

² Institute for Healthcare Improvement: Triple Aim Initiative. <u>http://www.ihi.org/offerings/Initiatives/TripleAim/Pages/default.aspx</u>



interviews with a predetermined list of questions unique to their industry category.³ Interview participants were chosen to reflect a balance of geography, size, vendor partners, and the type of organization. In addition, the project team held unstructured conversations with other interested industry parties including federal agencies, associations, solution companies, consumer organizations, and individuals who represent much of the continuum of stakeholders, or who have an innovative approach to matching.⁴

Data Collection Process

In-person and telephone interviews were held with more than 50 organizations and individuals.⁵ Interviews with organizations included members of the organization's C-Suite, health information management (HIM) department, information technology (IT), and government affairs. The interviews took place between September and November 2013.

Limitations

As noted above, the scan was not an academic survey. Therefore, this report will not provide specific tallies of patient matching practices, but offer generalized observations based on data collected during the interviews. These observations are limited by the relatively small subset of organizations, which may also have more focus on patient matching issues than the typical health care organization.

Hospitals and Health Systems

The environmental scan interviews included seven health systems that manage multiple hospitals, physicians, and ancillary services.⁶ In aggregate, the number of patients to whom they provide care is into the millions of individuals. In general, these organizations were aware of the patient matching problem from an internal medical records management and data exchange perspective, even if concern about the issue did not typically rise above managers in the business and health information technology or management departments. While several of the health systems have begun sharing patient records externally, in general it appeared that they were just beginning to think about the implications of incorrectly matched patients at the community level.

Data Attributes

There are several common data attributes that are used by many, if not most, organizations for matching, though the format of the data attributes varies, sometimes within the organization itself (see standardization section below). The three most common attributes used for matching are the patient's name, date of birth, and gender. Often the patient's phone number, address (full or ZIP code only) is also included in the core set of attributes. The variation in data attributes used by health systems tends to be caused by geographical and cultural variation around the stability of

³ The Paperwork Reduction Act of 1980 (Pub. L. No. 96-511, 94 Stat. 2812, codified at 44 U.S.C. § 3501-3521) places limitations on federally funded surveys.

⁴ These organizations and individuals did not receive a formal list of questions prior to or during the conversation.

⁵ This number includes the 32 organizations that received structured questions, and other interested industry parties.

⁶ Two health systems chose not to be included in the final report.



each attribute. In more rural areas attributes such as phone number and address were less transient when compared to urban areas.

While some health systems indicated that they only require three or four very stable data attributes to achieve high quality matching, several with deep experience in patient matching methods noted that they are better served by a larger number of attributes, no matter what those attributes are. For example, Geisinger Health System collects a long list of demographics that are separated into fields: first name, middle name, last name, date of birth, gender, whole social security number, street address, city, state, ZIP, and phone.

Many organizations continue to use the Social Security number (either the full number or the last four digits) at least internally. Fewer share Social Security numbers with providers who are neither employed nor affiliated with their health system. Most acknowledged that patients are becoming increasingly unwilling to share Social Security numbers in a health care context, and that some medical practices, hospitals, and insurance carriers have stopped collecting it. There was a strong feeling among the organizations that a unique number can be very useful in working through records of patients with the same name and birth date, but a few said Social Security number is not that helpful. Proponents of a unique number, whatever it may be, said it helps the system be more automated and reduces the need for costly manual review.

Organizations had an interest in historical demographics such as previous names, addresses, or phone numbers. Mayo, for instance, has been trying to collect maiden name at its Rochester site and in other areas is trying to standardize previous name. Some systems already collect previous data attributes, but others do not or cannot. Geisinger believes that using mother's maiden name for matching might be possible, but patients may not know it, know how to spell it, or be willing to share it. Additionally, some hospital information systems may not have a field available for an additional attribute such as mother's maiden name. Health systems urged caution about any national effort to require additional, atypical data attributes, arguing that the change in workflow for a large organization could be quite costly. However, at least one health system encouraged greater national commonality through regulation.

A number of organizations, including those speaking at the December 16 stakeholder meeting, repeatedly mentioned the value of matching using demographic attributes that are unlikely to change. However, they disagree on which data attributes are stable, with each noting situations where nearly every data attribute would be unstable. For instance, while date of birth might seem stable, some said they had seen people report varying birth dates to appear younger. Social Security numbers can be shared or otherwise misused or stolen. Cell phones may be increasingly stable, but that depends on local trends in phone use. Organizations had difficulty identifying the ideal, stable data attributes (other than a unique numerical identifier), but with further study it could be possible to identify those least likely to have variability.

Standardization

Even if similar data attributes are used across the continuum of care, often they are not stored in a standard format, introducing variation across organizations. Organizations noted that even a patient's name can become quite complicated, with variations in whether middle name or initial is



used, how hyphens are handled, whether previously used names or nicknames are recognized, and how people of different ethnic backgrounds use their family names.

Health systems frequently indicated that there needs to be a standardized way demographics such as name and date of birth are entered into systems and into electronic messaging formats. Systems that have strong patient identity integrity governance tended to have thought through this issue for internal sharing; Mayo Clinic, for instance, said it has established 17 standardized formats for demographic fields to be used by anyone in the organization creating or modifying a record. The differences in how various systems and organizations format their demographic attributes becomes exponentially more of a problem, when they share patient information externally.

Match Methods and Algorithms

Some organizations have developed their own matching methods and algorithms, while others purchased products such as IBM's Initiate, QuadraMed, or use their EHR vendor's product for internal and external matching. The success of the matching methods varied based on whether they were matching patients internally or externally across organizations. Health systems with a tightly managed data quality program and feedback loop appeared to be most successful in attaining match rates of 90 percent or higher. The match rates tended to decrease as organizations began sharing records with systems that are managed differently or have different EHR systems. Furthermore, the match rates dropped even if the same MDM/MPI vendor was used by both organizations. For example, Kaiser Permanente (which has 17 instances of Epic across its regions) reported a match rate of greater than 90 percent within each instance; that rate fell to around 50 percent to 60 percent when sharing between regions using a separate instance of Epic or with outside Epic partners. Other organizations cited similar declines in match rates as data is shared across unaffiliated organizations.

Organizations varied in the type of matching processes used, including: pure deterministic (one-forone character match on specified data attributes), weighted deterministic (one-for-one character match with some data attributes receiving higher "weighting" if they match or did not match; e.g. an Social Security number match is given high weighting), and probabilistic algorithms (complex set of weights, variable matching thresholds, accounting for transposition of characters, misspellings, and other data quality issues). Some of these approaches offer the ability to modify or tune the matching algorithm to manage the false positives and false negative rates. Organizations often described adjusting matching methods to fit the kinds of data entry errors that typically occur in their patient records, such as adjusting the weights of different data attributes to account for the most common errors or to account for policies such as not asking patients for social security numbers.

Manual Review

Health systems typically have several people in medical records or HIM departments devoted to periodically reviewing a potential duplicates report generated by their EHR or MPI and manually identifying the records that should be merged. Most said this activity would likely always be necessary because matching methods must be set to produce duplicates rather than overlays (false



positives) as a patient safety policy.⁷ However, there was definite interest in keeping manual review to a minimum because of the cost and the potential disruption in patient care when a record, such as a test result, is missing from the patient's record. Intermountain reported that it had calculated the cost of fixing a duplicate record at \$60 in operational costs. There is also widespread concern about the implications of poorly matched records being shared outside an organization through HIE, amplifying the errors.

Registration

The majority of health systems recognized that accurate patient matching does not rely solely on patient matching processes and algorithms, but that data quality, particularly at the point of entry, is equally important. Typos, misspellings, transpositions, fields left empty, or fields filled with false data can all cause problems downstream from the point of entry. Health system representatives acknowledged that the personnel at hospital front desks experience frequent turnover, making training a challenge. Regenstrief noted that the average time of employment for their registrars is roughly four months. At the same time, they often put effort into training programs (Intermountain shared its enterprise Master Patient Index (EMPI) training module with the Audacious Inquiry team). There is also a certification program for health care registration personnel through the National Association for Health Access Management (NAHAM) (see NAHAM's interview summary). There were mixed reactions to the idea of a national requirement for certification of health care registrars, but interviewees were generally positive about the idea of sharing best practices through a patient-safety campaign that could bring these training resources to smaller hospitals and to medical practices.

Patient registration processes also offer the opportunity to confirm a patient's identity, helping to reduce the chance of a mismatched or duplicate record. Some organizations reported that their registrars check a photo ID, which is also helpful for reducing the risk of fraudulent use of a patient's insurance coverage. These methods sometimes tie into a health system's patient safety efforts on the clinical side, where the standard of care is to confirm a patient's identity using two attributes (name and date of birth, typically) each time a medical test or intervention is carried out. These methods are also used to decrease the creation of a duplicate patient record. Most health systems, as a best practice, require registrars to search for an existing patient record before creating a new record. For example, Geisinger registration employees are trained to search for a patient record using a few letters of the first and/or last name to increase the chances of finding duplicate records for that patient under a nickname or misspelling. Other health systems, such as Group Health, only allow highly-trained, upper-level staff to create a new patient record, rather than the first line registration staff.

Group Health and Mayo Clinic both described higher-level, advanced governance initiatives around patient identity. Group Health's includes an ongoing process improvement project within the HIM department that is seeking to reduce the number of duplicate records. A cross-functional team standardized the process for assigning new member/patient numbers and reduced the number of departments that could assign numbers. Mayo Clinic described a homegrown application that is

⁷ It is generally recognized in the industry that false positives (incorrectly matching two records together) present a larger patient safety issue than false negatives (no match at all), since the wrong care could be provided based on an incorrect match.



linked to the MPI and automatically notifies all systems when a false positive or false negative is found. The system operators must fix the error and send a message back to the application indicating the error has been remediated.

Finally, some organizations routinely ask patients to confirm their data at registration. Care must be taken during this process not to mistakenly share the screen of a different patient's medical information. Health systems are increasingly focusing more on the potential for fraudulent sharing of insurance identities, so many have added verification of picture IDs or an insurance card with a photo or embedded electronic information, to their registration process. Others are currently utilizing or are implementing biometric methods (palm scans, fingerprints, etc.), with many indicating that they believed the use of biometrics would be common in the near future. All of the organizations currently using biometrics indicated that the biometric identification programs were voluntary for patients. Kaiser representatives also mentioned the organization's experiments with patient kiosks, where members can easily access medical records and adjust incorrect demographic information; this raises the issue of whether patients should have the ability to change demographic or clinical information about themselves, something that was mentioned frequently. When mentioned, organizations indicated the importance of provenance metadata, so that staff would know that the source of the update was the patient herself.

HIOs

Overall, the eight HIOs interviewed viewed their role as a facilitator of patient data sharing.⁸ They see the quality and formatting of data as the responsibility of participating organizations. The HIOs have generally approached each participating organization separately to set up customized interfaces and to adjust any matching methods for the data sets involved. Some also work with new participants to "clean up" data before sharing begins. Colorado Regional Health Information Organization (CORHIO), for instance, spends six to nine weeks integrating data from a new member's EHR system into the HIO.

Data Attributes

Like the health systems, the HIOs also select data attributes for matching and use various matching methods. The HIOs in the scan utilized a number of different MPI vendors, including: IBM Initiate, Medicity, Orion Health/NextGate, QuadraMed, and home-grown products, with two HIOs saying they are adding IBM Initiate to improve matching. HIOs use certain core data attributes such as name, gender, date of birth, and sometimes Social Security number. One HIO suggested that it is helpful to include other fields to get better matches, including; mother's first name and maiden name and father's full name. The HIOs noted that even with a sophisticated algorithm in place, data attribute choice and management is important, including data quality efforts, since at registration there is an average error rate of seven percent in the entry of first name and five percent for last name.

The use of Social Security numbers by HIOs is variable. At least one HIO, CORHIO, requires its participants to send them, and they are then used for matching. Other HIOs encourage participants to send them, but do not require it; this is typically spelled out in the participant's data sharing agreement. How the Social Security number is actually used by HIOs is also variable; while at least

⁸ All HIOs are included in the final report



one HIO sends out the Social Security number on patient records when it is provided, CurrentCare does not. CurrentCare argues that revealing it raises too great a risk under financial service breach regulations. Nevertheless, most HIOs reported that they are seeing fewer participants include Social Security numbers in the records they share.

Standardization

Many of the HIOs expressed concern about varying use of Health Level Seven (HL7) standards, and how medical practices will be able to manage arriving Consolidated Clinical Document Architectures (CCDAs). One HIO has been told by its vendor that an Admit, Discharge, and Transfer (ADT) feed will need to be sent along with each CCDA, in order to successfully match patient records. Conversely, another HIO has been able to improve match rates by working with each participant to standardize the data attributes on the HL7 messages they share.

Helping Participants Correct Matching Errors⁹

Some of the HIOs interact with participants to alert them to duplicate records. These interactions tend to take place through secure email and work best when individuals at the HIO can work with people they know at the medical records department of the hospital. Representatives of HealthInfoNet mentioned having this kind of relationship with participants for feedback. Significantly though, several HIOs said they do not provide feedback on potential duplicate records or mismatches, either stating that this has not been part of their business model or that their member hospitals would not want to look to the HIO for feedback on the quality of their patient data.

HIOs, like health systems, often have staff assigned to managing potential duplicate records. HealthInfoNet representatives said they expect this activity to require more staff hours as more organizations join the HIO, and that it could find more false negatives if it had sufficient personnel to look for them. CORHIO has roughly one full time employee who works with participants to identify duplicates, which can then be merged by its vendor Medicity. New York eHealth Collaborative (NYeC) has started to provide IBM Initiate's Inspector tool as a service that participants in its member RHIOs can use to find duplicates. CurrentCare is unique in that patients go through an opt-in registration process for their records to be shared, so data entry quality can be more closely controlled.

HIO Data Sharing Policies

The local culture of data sharing may impact how well HIOs and their participants interact to maintain a high-match environment. One HIO indicated they have developed a cooperative atmosphere around exchange over the many years that organizations have worked together, but others have struggled to build such an environment. HIOs also have governance policies and procedures of varying types. CORHIO, for example, has a policy committee and a user group where participants can work on the specifics of standardizing the data formats that they share. Utah Health Information Network (UHIN) maintains an MPI committee that has been discussing ways to update patient records as they change, as well as ways to involve patients in reviewing data for accuracy.

⁹ Participants refers to organizations or individuals that contribute data to or request data from HIOs.



False Positive and False Negative Rates

The HIOs were variable in their policies on maintaining false positive and false negative rates. Some HIOs were not aware of tracking these rates, while other HIOs do not need to track these rates because they use an opt-in exchange that matches everyone who is enrolled. Interviewees also noted that match rates are dependent on the threshold set for matching and the organizations' tolerance for error. One HIO described a process improvement project that rejected laboratory messages that do not contain required data, which helped to reduce false positives and false negatives, but raises the risk of delaying lab results delivery.

EHR Vendors

Nine EHR vendors participated in the scan, including inpatient and ambulatory vendors.¹⁰ Some EHR products have matching capabilities built in, while others integrate with a separate matching tool. Overall the capabilities of the vendors varied widely. A range of matching methods were described, from simple deterministic character-by-character matching to more complicated combinations of rules and probabilistic algorithms. When it comes to merging and unmerging records, vendors differ in their ability to easily undo an incorrect merge. One vendor indicated that merged records cannot be unmerged; rather, a user must quarantine the incorrectly merged record and employ audit logs to recreate two separate records. A number of other vendors allow a record to be unmerged, but it must be manually done and audit logs must be used to determine which data should be entered into each individual record. Vendors also varied on standardization or lack thereof of data attributes. While there was variation, most vendors do allow at least a minimum level of customization of the matching algorithm by customers.

Data Attributes and Standardization

In general, EHR vendors are aware of patient matching as an issue for their customers, and some are working on improvements to respond to that need. For instance, one vendor is building more reporting around matching and analytics, presumably in response to the provider reporting requirements of Meaningful Use. Several vendors encouraged better standardization of data attribute entry. Vendors described HIOs requesting different data formats, depending on the community or region. While vendors are able to customize their products to support these requests, it is time-consuming and adds to the expense of the products. A few vendors do not feel that matching is a factor in their relationships with EHR customers. Many vendors have found that customers who implement strong governance policies for their staff members around who can create or change demographic information within a record have fewer duplications and better matching. While the majority of EHR vendors were supportive of standardized data attributes, particularly within the CCDA, a few were opposed to requiring certification of additional data attributes, particularly those not currently captured within their EHR. Other ideas suggested by one or more EHR vendors included adding some testing measures during the certification process for product algorithms and for duplication reports.

¹⁰ Five vendors chose not to be identified in the final report, and one vendor requested its data be anonymized.



MDM/MPI/HIE Vendors

Seven vendors that sell software specifically for MDM, MPI and HIE were included in the scan. These vendors proved to be quite knowledgeable about both the technical aspects of matching patient records and the ways in which their customers manage data quality. Most vendors offer variations on a hybrid algorithm system that uses both deterministic and probabilistic methods. Vendors offer a combination of methods that assigns a score to how closely two records match on an attribute by attribute basis, ending up with a probability score. Vendors also utilize many tools that adjust for data quality issues such as transposition. Vendors' out-of-the-box matching methods are typically adjusted for a given data set and are optimized based on the kinds of mistakes that tend to be made during creation of a record by a particular health system. However, at least one vendor indicated that its solution is self-learning and should not be tweaked by users.

Data Attributes and Standardization

Vendors have differing views about the potential utility of adding certain demographic attributes on a national basis. Mirth believes that standardization of attribute use can help (such as regular use of mother's maiden name). IBM Initiate's matching typically relies on name, date of birth, address, phone number, and Social Security number (when available), and having these attributes available consistently can help improve matching. Many of the vendors indicated that historical data attributes would be helpful for matching (such as historical address). However, Mirth noted that getting universal use of an attribute such as previous address could be difficult because laboratories and some EHRs do not, or cannot, collect that information. NextGate feels that the common choice of data attributes is not as important as customers actually filling in the fields, because algorithms depend on having enough information to work with.

While none of the vendors believe there is a silver bullet to prevent all patient matching inaccuracies, a few technical suggestions were offered. Several vendors expressed that standardization of some core data attributes and more constrained HL7 implementation guides are potentially helpful steps, although there were differing opinions on the value of adding additional non-traditional demographic attributes as a standard requirement of patient records. There was also some support for the development of an open-source algorithm that could be used by smaller vendors, as well as repeated calls for better testing tools in the MDM space to verify the capabilities of an algorithm. Finally, several vendors believed that biometrics are a logical next step.

Data Quality

Vendors' models leave different parties responsible for merging and unmerging records. Vendors indicated that this was typically part of the customer agreement, with some customers desiring to shift liability away from themselves if records are incorrectly merged. The ability of the customer to adjust the algorithm, make rule changes, and adjust weighting independent of the vendor is variable across vendors. Vendors place a strong emphasis on data quality as a key issue for accurate patient matching. Vendors repeatedly indicated that education on the importance of accurate data collection was important to improving data quality. Training programs and the provision of timely feedback on data entry errors and creation of duplicative records were cited as highly variable across the health care industry. Suggestions for improvement included requiring the completion of certain fields such as first name, middle name, last name, and Social Security number (or the last four numbers of it); providing a feedback loop back to registration staff so they know what errors



they are making; spelling out month in the date of birth field to avoid transposition with day; and including expectations of data governance policies and procedures in HIE participation agreements to improve data quality of all HIO members.

Associations and Other Unstructured Interviews

Medical and Physician Groups

The physician groups interviewed did not believe that patient matching is on the radar of most ambulatory physicians or their staff. They are very aware of Meaningful Use requirements and wary of additional ones coming out in future stages. In fact, two groups noted that physicians may choose to ignore some requirements in Meaningful Use Stages 2 and 3, if they are more expensive to implement (updated software, workflow impediments) than the Medicare penalty. There is great concern in many quarters—not just among provider groups—about how practices will manage the flow of CCDAs in 2014. Groups focused on advancing better practice management techniques, suggested education of office staff about the importance of clean patient demographic data, but noted that it can be quite difficult to reach medical practices with information that doesn't appear to have an immediate impact on patient care or lightening their administrative burdens. It was suggested that physicians would best respond to educational messages when they include a focus on the financial bottom line.

One association felt that while hospitals are spending more on managing duplicate records and other bad outcomes of poor patient matching, they are also wary of liability issues on incorrect matches and an increasing occurrence of identity theft and fraud. Many hospitals are having to hire additional staff to focus solely on patient matching, and there is a growing unhappiness with the perception that MPIs are a black box that they do not understand and cannot control. Training registrars is a possibility, but associations noted that registration personnel are often low-paid and do not remain long in their positions.

Health Information Management Professionals

Several of the organizations representing professionals who manage health IT or patient information for hospitals and providers were enthusiastic about the Patient Matching Initiative and desire to work actively with ONC on disseminating best practices around patient identity integrity. This is an issue several of the groups have already developed materials around, and they look forward to new avenues for raising the issue's profile.

There is widespread support for a patient identifier among HIM professionals, but acknowledgment that it is not likely to happen soon and that matching processes and techniques would still need to be improved, even if there was a common number for use across the continuum of care. HIM professionals recognize that matching algorithms of varying types can be effective, and indicated that there would not be value in mandating a common algorithm. However, there were suggestions about establishing some basic standards for home-grown algorithms, such as Healtheway's interest in banning the simplest character-by-character deterministic methods.

Common themes from HIM professionals include the need to standardize data fields (some suggested this should be addressed in the implementation guides rather than the standards



themselves) and require filling out certain data fields within an EHR or registration system. Many mentioned the need to improve the accuracy of data entry at the source. It was suggested that vendors offer more training so that users are likely to take advantage of the report functions of matching software (such as regular lists of potential duplicate records created that day). There was widespread interest in a set of common terms or dictionary for patient demographics as well as matching methods (for instance, a common understanding of what a hybrid algorithm is).

Consumer Groups

Three consumer groups were interviewed, and two of them encouraged ONC to keep the patient's point of view in mind when coming up with solutions to patient matching. Associations indicated that work to improve patient matching should be aware that some patients may be uncomfortable sharing certain personal information with a registration clerk, to be sure to explain to patients why demographic information will make their health care easier to manage, and to encourage providers to share the computer screen with patients to confirm demographic and clinical information.

Associations Serving Government

Associations for public health agencies that collect and manage data for registries and other purposes suggested that data attribute standardization would be helpful for their work. They were also supportive of improving the quality of data entry through certification of registration personnel. One of the organizations would like to have a national-level forum for state public health officials to swap ideas on health data management.

Findings

The following findings represent near-term prospects that were developed based on feedback received during the <u>environmental scan</u> and through an intensive <u>review of historical literature</u> on patient matching. The findings are limited by the methodology, which did not allow quantitative analysis of these findings.

Standardization of Data Attributes

1. Standardized patient identifying attributes should be required in the relevant exchange transactions.

Patient identifying attributes are used in HL7 messages, CCDA, Integrating the Healthcare Enterprise (IHE), and eHealth Exchange standards to identify the patient to whom the message or clinical document relates. The attributes are generally highly variable from an implementation standpoint, with few fields being required, and little to no standardization of the data attributes themselves. The lack of data attributes that are populated consistently and in a standardized format within messages has been identified by the industry as a major impediment to more accurate patient matching. During the December 16, 2013 meeting, numerous stakeholders, including MPI vendors, EHR vendors, HIOs, associations, and hospitals indicated a desire for standardized patient data attributes, with many indicating that consistently and completely populating a defined and standardized set of data attributes would have a positive impact on match rates across a broad range of matching scenarios. The scenarios include, but are not limited to: querying for patient data, linking lab results or documents that are pushed to a provider (such as a CCDA), and linking within



a MPI serving a specific multi-entity domain. Additionally, many ambulatory providers, particularly those in small practices, are unlikely to have sophisticated algorithms supporting matching processes. Standardization of the data attributes on all transactions is an approach that supports multiple matching scenarios across the healthcare community to ensure that all providers have a base level of standardized demographic data to facilitate patient matching processes.

Through the environmental scan, the attributes most utilized for current matching processes were identified. Subsequently, a review of existing standards and a gap analysis was performed to ascertain which of the identified data attributes are required on the various standards and any existing formatting requirements. Based on this analysis and feedback received from the stakeholder meeting, the table below details a set of data attributes that should be required for relevant exchange transactions and the strategy for improving each attribute. Where possible, the strategies for improvement utilize existing standards from the Health Insurance Portability and Accountability Act (HIPAA) X12 transaction sets, Council for Affordable Quality Healthcare (CAQH) CORE initiative, HL7, International Organization for Standardization (ISO), and various recognized Internet standards.

Data Attribute	Strategy for Improvement
First/Given Name	1) Improve data consistency and normalize data
Current Last/Family Name	 Improve data consistency and normalize data Follow the CAQH Core 258: Eligibility and Benefits 270/271 Normalizing Patient Last Name Rule version 2.1.0 (Addresses whether suffix is included in the last name field.)
Previous Last/Family Name	 Improve data consistency and normalize data Follow the CAQH Core 258: Eligibility and Benefits 270/271 Normalizing Patient Last Name Rule version 2.1.0 (Addresses whether suffix is included in the last name field.)
Middle/Second Given Name (includes middle initial)	1) Improve data consistency and normalize data
Suffix	 Improve data consistency and normalize data Suffix should follow the CAQH Core 258: Eligibility and Benefits 270/271 Normalizing Patient Last Name Rule version 2.1.0 (JR, SR, I, II, III, IV, V, RN, MD, PHD, ESQ) If no suffix exists, should be null.
Date of Birth	 YYYYMMDDHHMMSS If hhmmss is not available, the value should be null Precise year, month, and day are required
Current Address (street address, city, state, ZIP code)	1) Evaluate the use of an international or USPS format



Data Attribute	Strategy for Improvement
Historical Address (street address, city, state, ZIP code)	 Evaluate the use of an international or USPS format If unavailable, the value should be null
Current Phone Number (if more than one is present in the patient record, all should be sent)	 Utilize an ISO format that allows for the capture of country code. Allow for capture of cell phone, home and work.
Historical Phone Number	 Utilize an ISO format that allows for the capture of country code. Allow for capture of cell phone, home and work.
Gender	1) ValueSet Administrative Gender (HL7 V3): M, F, UN

Upon an initial review of transactions used for Meaningful Use, the following transaction standards have been identified that require modification to conform to the above:

- HL7 v2.x messages
- HL7 Implementation Guide for CDA® Release 2: IHE Health Story Consolidation
- HL7 Version 2.5.1 Implementation Guide: S&I Framework Lab Results Interface
- IHE Volume 3 (ITI TF-3) Cross-Transaction Specifications and Content Specifications
- Nationwide Health Information Network (NHIN) Patient Discovery Web Service Interface Specification V 2.0
- 2. Any changes to patient data attributes in exchange transactions should be coordinated with organizations working on parallel efforts to standardize healthcare transactions.

There are a number of organizations working to standardize healthcare transactions. The Accredited Standards Committee (ASC) X12 and the CAQH CORE to standardize administrative transactions. CAQH and X12 have standardized a number of transactions, but there is variation across the transactions on the data that is required for patient demographics. In particular the 837 transaction standard which is used for submitting claims to payers lacks standardization of patient data attributes but has similar underlying data needs to clinical transactions. X12 has discussed internally updating the 837 transaction standard, but has not yet begun work.

A second parallel effort is by the National EMS Information Systems (NEMSIS) is a non-profit organization that creates and maintains a standard data set for emergency services. States require emergency medical service organizations to send structured data on a regular basis. EMS personnel record data in electronic systems that are certified by NEMSIS to meet the data set standards. In an effort to more closely align with the electronic transactions used by EHRs, NEMSIS is working with HL7 to ballot version 3 of the data set. There is an opportunity for alignment of the NEMSIS and X12 data standards with work to standardize patient data attributes in clinical transactions, since the majority of the required data attributes are the same.



Capturing Data Attributes Electronically

3. Certification criteria should be introduced that require certified EHR technology (CEHRT) to capture the data attributes that would be required in the standardized patient identifying attributes.

In order to include the data attributes listed in the previous finding in all relevant exchange messages, CEHRT must first have the ability to capture the data attributes. The majority of the data attributes listed above are currently captured by EHR systems. However, at least one attribute, historical address, is not consistently captured across all vendors. Certification criteria should be introduced that require CEHRT to demonstrate the ability to capture the following list of data attributes, not currently required in the 2014 certification criteria:

- Previous last/family name
- Middle name or initial
- Suffix
- Current address
- Historical address(es)
- Phone (including home, business, and cell)
- Historical phone(s)

An area requiring further study is the capturing and exchange of hyphens and apostrophes. The ability of vendors to capture and exchange these characters is inconsistent, which could lead to difficulties when sharing first and last name, particularly when using deterministic matching. Vendors have varying technical approaches and perspectives on the use of hyphens and apostrophes when matching. Prior to introducing certification criteria, the benefits and shortcomings of including hyphens and apostrophes when exchanging patient data attributes should be explored.

It is important to recognize that much of this data is initially captured in a registration system, which may be a separate product from an organization's CEHRT. Regulating CEHRT is the first step in the process to begin capturing or storing these attributes.

Data Attributes Requiring Additional Study

4. The ability of additional, non-traditional data attributes to improve patient matching should be studied.

One way of potentially improving the accuracy of matching is to introduce into the matching process additional data attributes that are not typically captured in today's workflows or used for matching. These data attributes could include but are not limited to email address, Direct address, mother's first and maiden name, father's first and last name, place of birth, driver's license number, passport number, or eye color. Currently, EHR systems cannot capture the majority of these data attributes in a structured field. Mother's maiden name was the most common data attribute able to be captured; however, it is rarely captured today. Introducing a requirement to capture and exchange these data attributes would require significant changes to current registration processes and vendor system capabilities.



In addition, these data attributes are often used to establish a patient's identity through knowledgebased authentication or are used by patients as security questions for secure online credentials. As such, the capture and exchange of these data attributes may contribute to decreased security of consumer information. Consequently, should such a requirement to capture and exchange these data attributes be considered for the future, an analysis of the ability of the data attributes to improve patient matching and the potential security issues that may arise from their use should be studied. Working with a number of different organizations that have the capability today, a statistical analysis on a set of representative test patient data could help to determine if the presence of these data attributes improves the likelihood of a positive match.

Patient Matching Algorithms

5. Certification criteria should not be created for patient matching algorithms or require organizations to utilize a specific type of algorithm.

Included in the environmental scan were small and large health systems utilizing a range of patient matching products, including internally and commercially developed products. Vendors that offer a range of patient matching products, including those that utilize deterministic matching and a range of probabilistic matching tools, were also included. The majority of solutions depend on a version of the same base algorithm (Fellegi-Sunter), with each company building a complementary set of proprietary tools (that account for data quality, geographic differences, and data attribute availability) that make their product unique.

Healthcare organizations have made at least modest, and in some cases, significant investments in implementing and refining their patient matching solutions. During the environmental scan, many indicated that replacing their current systems or modules would be cost prohibitive. As such, it is not suggested that a standardized patient matching algorithm be developed or required of vendors. Such a requirement could have significant technical and financial implications for health systems, HIOs, and vendors. In addition, imposing a federal standard could hinder market innovation and ultimately be detrimental to improving patient matching. However, it should be noted a small number of organizations have voiced their preference for ONC to regulate algorithms and essentially ban the use of pure deterministic algorithms. However, such a ban has the potential to cause significant hardship for health systems and ambulatory providers in particular. At this time, it is not recommended that there be any prescriptive action on the use of particular algorithms.

Identifying Duplicates

6. Certification criteria that requires CEHRT that performs patient matching to demonstrate the ability to generate and provide to end users reports that detail potential duplicate patient records should be considered.

Identifying duplicate patient records within an EHR system is important to ensuring accurate matching of patient records. The environmental scan revealed that many EHR systems with built-in matching processes offer reports that identify potential duplicate records, though not all systems offer such a capability. Additionally, some systems have the capability, but do not make the reports accessible to end users. Certification criteria should be introduced that requires CEHRT that



performs patient record matching to demonstrate the ability to generate and provide to end users reports that detail potential duplicate patient records. Further, CEHRT should clearly define for users the process for correcting duplicate records, which typically requires the merging of records.

Best Practices, Education, and Outreach

7. Build on the initial best practices that emerged during the environmental scan by convening industry stakeholders to consider a more formal structure for establishing best practices for the matching process and data governance.

The environmental scan revealed that many organizations are making strides in establishing and refining their practices for improving the accuracy of patient data and matching for clinical and administrative purposes. It is important to note that this environmental scan was developed to look for best practices in identity verification and patient matching processes, and that while focusing on those areas, many other areas where best and promising practices could be established to improve the accuracy of patient data were highlighted. Practices include regular reviews of potential duplicates, data governance programs that work to establish current rates and then improve false positive and false negative rates, training programs that can be replicated, policies that apply across a health system with multiple sites, and processes for a central entity, such as an HIO or Accountable Care Organization (ACO), to notify participants of matching errors and corrections. (See <u>Appendix B: Best Practices for Data Quality</u> for additional detail.)

While the environmental scan identified some methods with potential for use throughout the healthcare industry, it is unclear whether these best practices could be universally utilized, particularly in small ambulatory practices. Industry stakeholders, including health systems, HIOs, vendors, and associations should be convened to develop a set of best practices for matching processes and to research methods for measuring current and future practices for their effectiveness in improving matching rates. This evaluation of best practices could inform the educational campaign envisioned in finding nine.

8. Work with the industry to develop best practices and policies to encourage consumers to keep their information current and accurate.

Patients are the primary source of demographic data used in matching and are consequently pivotal to ensuring data quality during the registration and admission process, and throughout the healthcare continuum. Patients are typically not aware of the matching processes used when their data is shared and may not understand that ensuring their providers have accurate and up-to-date information in their systems can actually have a positive impact on the quality of their care. Processes vary significantly across organizations for having patients update their demographic information. Some organizations ask patients to complete paper forms that are later used to update the practice management/EHR system, or they use telephone registration in advance of scheduled appointments. Other organizations utilize electronic methods, such as a patient portal, waiting room kiosks, iPads, etc. to prompt patients to update their demographic information. This data can then be fed directly to the EHR to populate or update the patient record. Understanding these processes, and how they vary, is important in meeting the goal of better engaging and activating patients. Regardless of the process an organization uses, raising awareness among patients of the importance of correct, current demographics is a worthy goal in itself.



Meaningful Use Stage 2 places an increased emphasis on patient engagement with their health information. This emphasis should be extended to ensuring patients are engaged in maintaining accurate demographic data. Policies, best practices, and outreach activities should be developed for educating and activating patients to take responsibility for the accuracy of their demographic data. Examples of best practices could include allowing patients to manage their own demographics via a patient portal, training registrars and clinicians to verify patient demographic information, and verification of a patient's identity via a photo ID and/or insurance card.

9. Work with healthcare professional associations and the Safety Assurance Factors for EHR Resilience (SAFER) Guide initiative to develop and disseminate educational and training materials detailing best practices for accurately capturing and consistently verifying patient data attributes.

Accurate patient identification and matching across organizations cannot be adequately addressed through standardization of data attributes alone. The accuracy of the data attributes themselves is important for minimizing false positives and false negatives. While some systems are equipped with algorithms that can compensate for data accuracy issues using probabilistic matching techniques, these systems have limitations. Additionally, EHR systems are not universally equipped with such algorithms to compensate for some data inaccuracies. Consequently, improving the accuracy with which data attributes are captured and the consistency with which they are verified with patients is a more efficient and effective method for improving patient match rates across organizations.

Data integrity programs should acknowledge the key role of the front office staff and registrars who are typically responsible for verifying the patient demographic information that is used in matching. They are critical to any effort to improve patient matching industry-wide, and should be involved as partners in data integrity initiatives. Ensuring that staff members have adequate and appropriate training is a necessary component to improving data integrity. This could include training that emphasizes the importance of filling in demographic fields accurately and completely, and an explanation of the implications of incorrect information and duplicated records on patient care downstream. Other potential best practices related to registration that were noted in the environmental scan included restricting the number and type of hospital personnel who can create a patient record and encouraging registration staff to obtain appropriate certification.

The American Hospital Association, American Health Information Management Association, American Medical Association, American Academy of Family Physicians, American College of Physicians, Medical Group Management Association, National Association of Healthcare Access Management, and other associations have a long history of developing best practices and training materials for providers, nurses, medical assistants, registrars, and front office staff. Additionally, organizations such as the National Patient Safety Foundation and its national Partnership for Patients Initiative, the Leapfrog Group, the Robert Woods Johnson Aligning Forces for Quality Initiative, and Campaign Zero have a long history of working to improve patient safety. These organizations are well positioned to develop a marketing campaign that would include best practices and educational materials for collecting and verifying patient demographics.

This representative set of a broader diverse group of patient safety organizations, with a common mission to improve the quality and safety of health care, could be a powerful force in leading an educational campaign about the importance of patient identification data quality, and could assist



in the distribution of best practice materials suggested in this report's findings. Such a broadly supported campaign would be particularly useful for smaller hospitals, physician groups, and ancillary parts of the health care system, such as laboratories and long-term care facilities, which often lack adequate funding to support sophisticated health information management infrastructures. As smaller and ancillary organizations are more integrated into coordinated care teams, they will need support in improving their policies and processes to better ensure the integrity of their own patients' information, which will improve the accuracy of record matching across their health information exchange ecosystem.

In addition to working with experienced patient safety groups and associations, the SAFER Guides initiative should be utilized to promote practices that improve data quality for better patient identification and matching. The initiative has developed a phased implementation approach of best practices for improving patient identification at the point of care. There are nine SAFER guides, each focusing on a different aspect of health IT use, such as organizational responsibilities or test results review and follow-up. One of the guides is focused on patient identification and includes a series of checklists for self-assessment on specific best practices. While a few of the group's recommendations would require modifications to EHR systems, a number address workflow processes:

- Patients are registered using a centralized, common database using standardized procedures.
- Patient identity is verified at key points or transitions in the care process (e.g., rooming patient, vital sign recording, order entry, medication administration, and check-out).
- The use of test patients in the production (i.e., "live") environment is carefully monitored. When they do exist, they have unambiguously assigned "test" names (e.g., including numbers or multiple ZZ's) and are clearly identifiable as test patients (e.g., different background color for patient header).
- The organization regularly monitors their patient database for erroneous patient identification information.
- 10. Continue collaborating with federal agencies and industry on improving patient identification and matching processes.

A critical aspect of making progress against the overall patient matching challenge is to ensure the continuation of focus and the coordination across federal agencies and industry stakeholders to develop the best practices for data governance, consumer engagement, and data quality. The environmental scan and literature review have identified a number of best practices in each category that provide a solid starting point; however, additional work is needed to identify which best practices can be implemented by a wide range of organizations, particularly ambulatory practices which typically have fewer resources. Collaboration with federal agencies and the industry will be necessary to ensure wide adoption of best practices.

The December 2013 presentation included an initial finding that an open source algorithm that could be utilized by vendors to test the accuracy of their patient matching algorithms or be utilized by vendors that do not currently have patient matching capabilities built into their systems, should be developed or supported. There may be value in developing an open source algorithm or



updating and supporting an existing open source algorithm that EHR vendors may choose to utilize in their products. This approach would likely be most beneficial to smaller organizations that have not invested heavily in patient matching to date. The environmental scan highlighted that some EHR vendors, particularly the larger inpatient vendors, have developed matching algorithms; however, many vendors do not have such matching capabilities. Ambulatory providers in particular are likely to rely on their EHR vendor to effectively match patient records (or at least present an initial list of potential matches). In addition, an open source algorithm could be utilized as a testing tool for vendors to benchmark the accuracy of their proprietary techniques. Open source algorithms currently exist, but would require updates to ensure they utilize the proposed required data attributes for matching, and can accept the data attributes in the proposed format. Existing algorithms should be evaluated to see if they could be updated and supported as needed or whether a new algorithm should be developed.

During the December 2013 meeting, a number of stakeholders indicated concern about ONC developing or supporting such an open source algorithm. Concerns included liability issues should the algorithm be utilized and fail to match patient records, building and maintaining a large enough set of test patient data to test algorithms, and the methods for ensuring the algorithm is of high quality are meets a specific threshold of matching. Additional comments received after the meeting, indicate a support for an open source algorithm, but only with a separation of the algorithm for use in EHRs as opposed to being used for testing against commercial MPI products. Consequently, additional work needs to be performed to identify legal issues, identify if existing open source algorithms can be utilized, and to define the use of an algorithm for testing commercial products against.



Appendix A: Literature Review

Introduction

Accurate identification and matching of patient records is pivotal for ensuring patient safety as records are stored and exchanged electronically. For instance, one fifth of CIOs surveyed by College of Healthcare Information Management Executives (CHIME) indicated that at least one patient in the last year suffered an adverse event, due to mismatched records.¹¹ Other studies have found an association between duplicate patient records and lab results not being reported to patients to receive adequate follow-up care.¹² As the use of EHRs and electronic exchange of patient data has increased and accurately identifying and matching patient records has been recognized as a major challenge to the industry, many organizations (both public and private) and industry leaders have researched and evaluated patient record matching challenges. The result is a growing body of research on the subject.

Match Rates

The patient matching challenge can be quantified many different ways: the percent of records that are matched, duplicate record rates, false positive rates, false negative rates, sensitivity, and specificity. While an error rate of less than eight percent is the industry-recognized standard, many organizations exceed this rate. Studies have found that duplicate record rates in health care databases can be high. One study evaluated 112 MPIs from 2000 to 2003 and found a mean duplication rate of eight percent, with a quarter of the indexes having duplicate record rates of more than ten percent. An examination of a smaller set of 11 EMPIs, which are combined files of MPIs from multiple organizations, found a higher rate of duplication, ranging from seven percent to 39 percent.¹³

In 2008, RAND Health worked with a 42,000-record sample of the Social Security Death Master File database to pinpoint the theoretical risk of a false positive match using specific identity attributes. It found that when using first name, last name, birth year, and ZIP code, the number of possible false matches is one in 3,500. Adding a unique part of the Social Security number reduced the probability to near zero. When RAND carried out a similar analysis of the larger database of 80 million records of the Social Security Death Master File, the false positive rate began at 98 percent with just last name (it was about 66 percent in the smaller sample size), going down to 33 percent when date of birth was added, and to one in 39 million when first name, ZIP code, and last four digits of the Social Security number were also used.¹⁴ A CHIME survey found that most respondents

<u>ftp://ftp.software.ibm.com/common/ssi/ecm/en/imw14343usen/IMW14343USEN.PDF</u> ¹⁴ Identity Crisis? Approaches to patient identification in a national health information network. RAND Health. 2008. www.rand.org/pubs/research_briefs/RB9393/index1/html

¹¹ Summary of CHIME survey on patient data matching. May 16, 2012. <u>www.cio-</u> <u>chime.org/advocacy/resources/download/Summary_of_CHIME_Survey_on_Patient_Data.pdf</u>

¹² Duplicate Patient Records – Implication for Missed Laboratory Results. November 3, 2012. AMIA Annual Symposium. <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540536/</u>

¹³ Integrating Patient Medical Records in Pursuit of the EMR. IBM Software. 2010. (Originally published in 2008 before IBM purchased Initiate)



have less than an eight percent error rate; however, about a third of those surveyed reported an eight to 14 percent error rate.¹⁵

Data Quality's Impact on Match Rates

Data quality has a large impact on a system's ability to accurately match patients and provide the appropriate care. A 2008 study found that the majority of patient identification errors in the emergency department could be traced to registration, where the patient's information was entered incorrectly. ¹⁶ Researchers at Johns Hopkins undertook a Plan-Do-Study-Act (PDSA) process to identify the root of patient identification errors and found that there were approximately seven to 15 registration errors per month, prior to implementing new processes. The new processes, implemented through three PDSA cycles, reduced registration errors by more than 80 percent for inpatients. ¹⁷

A number of organizations have recognized the importance of improving the accuracy of initial data entry. In 2009, the AHIMA encouraged better training of hospital registry personnel: "Comprehensive training in record search routines coupled with data quality mechanisms that provide feedback to registration and scheduling staff are crucial," the AHIMA report concluded.¹⁸ More recently AHIMA worked with hospitals to identify the registration workflows and identify registration improvement programs to ameliorate data quality.¹⁹ The HIMSS Work Group also noted that business process quality improvement efforts should include anyone in the organization who touches patient identity in an information system, not just those in registration.²⁰ Some organizations have put forth registrar certification as a way to improve data entry; the NAHAM maintains a certification program for health care registrars and their managers.²¹

Patient Confirmation of Identity Information

Patients have a growing role in confirming the accuracy of their own identity information in health care systems, research shows. The ONC Tiger Team has suggested that Meaningful Use measures addressing patient engagement consider the patient's role in managing demographic data. Additionally, the Tiger Team concluded, individuals should have a simple method for reporting corrections to their health and demographic information, and HIOs should support patient access to

http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_044000.hcsp?dDocName=bok1_044000

¹⁵ Summary of CHIME survey on patient data matching. May 16, 2012. <u>www.cio-</u> <u>chime.org/advocacy/resources/download/Summary_of_CHIME_Survey_on_Patient_Data.pdf</u>

¹⁶ The nature and occurrence of registration errors in the emergency department. AF Hakimzada, RA Green, OR Sayan et al. Int J Med Inform. 2008; 77:169-175. <u>http://www.ncbi.nlm.nih.gov/pubmed/17560165</u>

¹⁷ Registration-associated patient misidentification in an academic medical center: causes and corrections. MJ Bittle, P Charache, DM Wassilchalk. Jt Comm J Qual Patient Saf. 2007 Jan; 33(1):25–33. http://www.ncbi.nlm.nih.gov/pubmed/17283939

¹⁸ Managing the Integrity of Patient Identity in Health Information Exchange. Journal of AHIMA 80, no.7 (July 2009): 62-69.

¹⁹ Exposing Double Identity at Patient Registration. Chris Dimick. Journal of AHIMA 80, no. 11 (November 2009): web extra.

http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_045561.hcsp?dDocName=bok1_045561 ²⁰ Patient Identity Integrity. HIMSS Patient Identity Integrity Work Group. December 2009.

http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf ²¹ Candidate Guide to Certification 2014. National Association of Healthcare Access Management.

http://c.ymcdn.com/sites/www.naham.org/resource/resmgr/Certification/NAHAM_Candidate_Guide_to_Cer.pdf



their records in the language of business associate and participation agreements.²² The ONC Power Team supported the Tiger Team and recommending that "Providers should allow patients to verify the patient attributes the provider has recorded for them through a method such as sharing the data entry screen with the patient for review, providing the patient with a printed summary or online access to the data to help identify quality issues and utilize the methods provided by HIT developers to identify missing/unavailable data and approximate or questionable values at the time of data entry."²³ Suggested methods for providing patients access to update their demographic information electronically is through the use of kiosks and online patient portals.^{24, 25}

Patient Matching Processes

Existing Standards

While there are few formalized standards for patient matching, IHE has developed two profiles that are used by eHealth Exchange for querying patient data, Patient Identifier Cross Referencing (PIX)/Patient Demographics Query (PDQ) and Cross-Community Patient Discovery (XCPD). These profiles govern the process for an organization to transmit a patient's demographic information to a receiving organization to determine if it has records available for the patient; the process for the receiving organization to respond to the query with the patient's demographic information so that the originating organization can match the patient to its records; and cross reference the patient identifiers from each organization for future queries.^{26, 27} ONC's Power Team has recommended that patient query patterns follow these standards and that Clinical Document Architecture (CDA) R2 header formats be used to represent patient attributes.²⁸

Matching Methods and Algorithms

In the absence of a unique patient identifier, the industry has implemented two primary matching methods, with variations of each method: deterministic matching and probabilistic algorithms. Deterministic matching performs a character-by-character match on a specified set of data attributes, one attribute at a time. Probabilistic algorithms are predominantly based on the Fellegi-Sunter theory, which is a complex statistical analysis of a set or string of patient data attributes that when considered in concert determine whether there is an automatic match, no match, or manual

²⁶ IHE USA Certification Program (slides). IHE. November 12, 2013. <u>http://www.iheusa.org/certification.aspx</u>
 ²⁷ Patient Identity Management Webinar (slides). Eric Heflin. IHE. August 18, 2005.

http://wiki.ihe.net/index.php?title=Current_Published_ITI_Educational_Materials ²⁸ HIT Standards Committee Patient Matching Power Team Recommendations to ONC. August 2011.

http://www.healthit.gov/sites/default/files/standards-

²² ONC Privacy and Security Tiger Team Recommendations on Patient Matching. HIT Policy Committee Hearing. February 2011. <u>http://www.healthit.gov/sites/default/files/archive/HIT%20Policy%20Committee/2011/2011-02-02/5b%20-%20Deven_Tiger%20Team%20Recommendations%20on%20Patient%20Matching.doc</u>

²³ HIT Standards Committee Patient Matching Power Team Recommendations to ONC. August 2011. http://www.healthit.gov/sites/default/files/standards-

certification/8_17_2011Transmittal_HITSC_Patient_Matching.pdf

²⁴ Touchscreen Check-In: Kiosks Speed Hospital Registration. California HealthCare Foundation. March 2009. http://www.chcf.org/publications/2009/03/touchscreen-checkin-kiosks-speed-hospital-registration

²⁵ Identification, Authentication and Matching to Support Consumer Access to HIE Data. Margaret Mary Community Hospital (Batesville, Indiana), Nor More Clipboard, Indiana Health Information Technology Inc. (report on ONC Challenge Grant). <u>https://www.nomoreclipboard.com/wiki/images/a/ac/NMC_Case_Study_MMCH.pdf</u>

certification/8 17 2011Transmittal HITSC Patient Matching.pdf



review. Probabilistic algorithms typically use additional tools to account for data entry errors, such as Soundex, edit distance calculations, frequency indexing, etc. These two methods can also be combined and used as a hybrid matching method.²⁹ The industry has stopped short of recommending a standard matching method or algorithm because these are often customized and fine-tuned for specific data sets, taking into account the ways in which demographic attributes differ across community and ethnicity. Additionally, there is a lack of substantiated research on the variability of success with different matching methods, particularly with real-world data sets, although one small study of sample data found that simple deterministic methods did not perform as well as either probabilistic or hybrid methods.^{30, 31, 32, 33, 34} While the industry has not called for a standard algorithm, it has indicated a desire for standards of algorithm performance and the thresholds used for matching.35

The matching methods used by vendors, health systems, and HIOs are typically proprietary, with each organization having its own variation of the method. There are however a number of open source MPI/MDM matching methods, most of which utilize probabilistic algorithms. Open source options include: OpenEMPI,³⁶ OpenMRS,³⁷ Febrl,³⁸ and ChoiceMaker (partially open source).³⁹

Data Attributes

Matching methods rely on patient demographic data to identify and match records, with most methods using some combination of local patient identifier (medical record number), first and last name, date of birth, phone number, address, and Social Security number. While standardized sets of data attributes have been proposed as best practices, the industry has not accepted one standard set of attributes to be used for matching, leaving each organization to decide for itself. Suggestions for standardized attributes have come from a number of industry stakeholders, including the ONC Power Team, which made the following recommendation in 2011:

http://jamia.bmjjournals.com/content/16/5/738.full?related-urls=yes&legid=amiajnl;16/5/738

²⁹ Record linkage software in the public domain: A comparison of Link Plus, the Link King, and a "basic" deterministic algorithm. Campbell, K. M., Deck, D., & Krupski, A. Health Informatics Journal, 14(1), 5–15: 2008. http://jhi.sagepub.com/content/14/1/5.long

³⁰ Patient Linkage. Presentation to NCVHS (slides). Shaun Grannis. 2005. http://www.ncvhs.hhs.gov/051207p1.pdf. ³¹ Real World Performance of Approximate String Comparators for use in Patient Matching. Shaun Grannis, J. Marc Overhage, Clement McDonald, Medinfo 2004: Amsterdam: IOS Press, 2004.

http://www.researchgate.net/publication/8353095 Real world performance of approximate string comparators f or use in patient matching/file/d912f5143154550077.pdf

³² Identity Resolution and Data Quality Algorithms for Master Person Index. Oracle Health Sciences. August 2010. http://www.himss.eu/sites/default/files/identity-resolution-algorithm-wp-171743.pdf

³³ An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. VJ Zhu, MJ Overhage, J Egg et al. J Am Med Infom Assoc 2009 16: 738-745.

³⁴ Record linkage software in the public domain: A comparison of Link Plus, the Link King, and a "basic" deterministic algorithm. Campbell, K. M., Deck, D., & Krupski, A. Health Informatics Journal, 14(1), 5-15: 2008. http://jhi.sagepub.com/content/14/1/5.long

³⁵ Patient Identity Integrity, HIMSS Patient Identity Integrity Work Group, December 2009.

http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf ³⁶ https://openempi.kenai.com/#Home-Latestnews

³⁷ https://wiki.openmrs.org/display/docs/Patient+Matching+Module

³⁸ http://sourceforge.net/projects/febrl/

³⁹ http://oscmt.sourceforge.net/



Patient attributes for matching should ideally be discriminating (some authors discuss uniqueness of attributes such as biometrics, but in most cases we are simply hoping for attributes that discriminate one patient from another). Patient attributes: ubiquitous (e.g., last name, date of birth, eye color), unchanging or invariable (e.g., date of birth, gender, given name, DNA), uncomplicated (e.g., last name, date of birth, gender), easily and inexpensively accessible and uncontroversial. Having a common "base" set of patient attributes across entities that are matching patients is important if the entities are going to achieve an acceptable level of sensitivity and specificity.⁴⁰

The ONC Power Team stopped short of recommending which data attributes should comprise the standard set. Other organizations, including HIMSS⁴¹ and the Bipartisan Policy Center, ⁴² have echoed the ONC Power Team's call for a standard set of data attributes, particularly in lieu of a unique patient identifier. While these organizations have made note of the most commonly used data attributes, they have not recommended which should be included in the standard set of attributes, although HIMSS has emphasized the use of data attributes that are unique enough to provide a match but are not likely to pose a threat to the patient's privacy or security.⁴³ Additionally, some organizations have suggested that a standard set of data attributes should include attributes that support research on health disparities by providing more granular information about patients.⁴⁴

In addition to a common set of data attributes used for matching, a number of organizations have recommended that the data attributes be represented in a standard format. The ONC Privacy and Security Tiger Team strongly recommended the standardization of the data attribute formats used for matching. It cited the required use of a middle name as one example of standardizing the name field. This would require EHRs to be tested to ensure that standardization is feasible. The team also noted that there should be consistency in how organizations handle a data field when the information is not available, and how missing data should be represented. The ONC Power Team subsequently suggested the use of existing formats, recommending that patient query patterns should follow the Nationwide Health Information Network (NwHIN) (now known as Healtheway) patient query implementation guide and that the CDA R2 header formats should be used to represent patient attributes.⁴⁵

certification/8 17 2011Transmittal HITSC Patient Matching.pdf

⁴⁰ HIT Standards Committee Patient Matching Power Team Recommendations to ONC. August 2011. <u>http://www.healthit.gov/sites/default/files/standards-</u>

⁴¹ Patient Identity Integrity, a White Paper by the HIMSS Patient Identity Integrity Work Group. December 2009. http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf

⁴² Challenges and strategies for accurately matching patients to their health data. Bipartisan Policy Center. June 2012.

http://bipartisanpolicy.org/sites/default/files/BPC%20HIT%20Issue%20Brief%20on%20Patient%20Matching.pdf ⁴³ Patient Identity Integrity, a White Paper by the HIMSS Patient Identity Integrity Work Group. December 2009. http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf

⁴⁴ Leveraging Meaningful Use to Reduce Health Disparities: An Action Plan. Consumer Partnership for eHealth. August 2013. <u>http://www.nationalpartnership.org/research-library/health-care/HIT/leveraging-meaningful-use-to.pdf</u>

⁴⁵ HIT Standards Committee Patient Matching Power Team Recommendations to ONC. August 2011. http://www.healthit.gov/sites/default/files/standards-

certification/8 17 2011Transmittal HITSC Patient Matching.pdf



Best Practices and Processes

Patient matching is not solely a statistical problem. Issues of data integrity and workflow are significant factors, and some research has assessed the topic from the perspectives of those working in the fields of health information, financial, and information management. Sharp HealthCare has worked for decades to improve patient matching across its organization. In 2011, the organization began a Six Sigma process to implement best practices to improve patient matching, particularly at the point of registration, to eliminate the creation of duplicate records. The best practices include the use of biometrics (a palm scan) and a multi-factor lookup. Registrars must first search for a patient using common data attributes, before a new record can be created. The use of these best practices has allowed Sharp HealthCare to reduce its error rate.⁴⁶ Additional industry best practices include the implementation of a strong master data governance program and ongoing manual review of patient records to eliminate duplicate records and ensure records have not been improperly merged.^{47,48}One such model for an ongoing quality improvement program is the Improvement Focused Model. This model makes its goal to maintain a clean MPI, meaning no duplicate records and no overlaid records are present. The model takes organizations through a six step process from assessment of the MPI to the creation of processes and procedures to ensure a clean MPI is maintained over time.⁴⁹

HIMSS developed the Patient Identity Integrity Toolkit in 2011 to disseminate information to organizations about best practices and processes for matching patient records. Within the toolkit, HIMSS developed a set of key performance indicators (KPIs) that allow an organization to evaluate its patient matching processes and technology and make continuous improvements. The KPIs are geared towards maintaining a clean MPI. The toolkit provides both the list of data attributes to be collected and the formulas used to calculate each performance indicator. Additionally, the toolkit indicates the use and appropriate audience for each indicator. Below is the list of data attributes and indicators developed by HIMSS.⁵⁰

http://api.ning.com/files/TPjnN*zfKPQnIW2DugUalXh9kkTjem2PBusrW2XB7Mzs-

⁴⁶ Patient Matching: Sharp HealthCare's Journey. May 2012.

⁴TDu9*MiIqxn0hdkdW6gOQxBDNOEm9w4QSGm5tvC4hDUGfQGOMO/SharpMPICaseStudy.pdf

 ⁴⁷ The Five Most Dangerous Health Information Practices. Presentation to AHIMA 2013 Convention (slides). Beth Just and Grant Landsbach. October 2013. Also in Journal of AHIMA 84, no.11 (November–December 2013): 40-42.
 ⁴⁸ Minimizing Duplicate Patient Records to Maximize Cash Flow; Case Study of Texas Health Resources. Patricia Consolver. Healthcare Financial Management Association. May 2013.

http://www.quadramed.com/quadramed/QMMedia/documents/Articles/Minimizing-Duplicate-Patient_0513_RCSreprints_consolver.pdf

⁴⁹ Establishment of a Quality Program for the Master Patient Index. Robin Altendorf. AHIMA's 79th National Convention and Exhibit Proceedings, October 2007.

http://library.ahima.org/xpedio/idcplg?IdcService=GET_HIGHLIGHT_INFO&QueryText=%28baby+girl%29%3C and%3E%28xPublishSite%3Csubstring%3E%60BoK%60%29&SortField=xPubDate&SortOrder=Desc&dDocNam e=bok1_039331&HighlightType=HtmlHighlight&dWebExtension=hcsphttp://li

⁵⁰ Patient Identity Integrity Toolkit: Patient identity integrity key performance indicators. HIMSS. December 2012. http://www.himss.org/files/HIMSSorg/content/files/PII03 Key Performance Indicators Final.pdf



Data to be collected:

- EMPI Database Size
- EMPI Person Population
- EMPI Database Growth
- EMPI Patient/Person Additions
- Total EMPI Matches
- Total EMPI Non-Matches
- EMPI Database Duplicates
- False Positive Matched Pairs
- True Matched Pairs
- False Negative 'Non-Match' Pairs
- Indeterminate Matched Pairs
- EMPI Matching Validation
- Database Error Growth
- Total Registrations Performed

Performance Indicators

- EMPI Database Activity Rate
- EMPI Database Duplicate Rate
- Duplicate Creation Rate
- True Match Rate
- False Positive Match Rate
- False Negative (Non-Match) Rate
- Indeterminate Match Rate
- Matching Accuracy Rate
- Matching Error Rate

Unique Patient Identifier

In the absence of work toward an official national patient identification number, some research has continued to explore the possibilities of a voluntary patient identifier, largely at the local or regional level. Global Patient Identifier, Inc. (GPII) has been working on the voluntary universal healthcare identification (VUHID) project, which issues a unique identifier to patients in the form of a token (such as a smart card). Patients can then choose to use the identifier at each point of care. Additionally small research projects have studied identifier cards that patients can use to positively identify themselves at multiple provider sites, and found that the cards can accommodate providers' workflow and are technologically feasible.⁵¹ In 2011, the White House launched the NSTIC, the goal of which is to decrease identity theft caused by the multiple passwords and identities individuals must maintain across websites. NSTIC intends to allow individuals to create a voluntary trusted identities or credentials that can be used across the internet for sensitive information. Individuals maintain control of the credential(s) and choose the amount of personal information to share with each site. In 2013, NSTIC launched a number of pilots, which include the use of the credentials for healthcare; however, results of the pilots are not yet available.⁵²

Security and Privacy Considerations

Much of the patient matching literature also recognizes the risk of exposing patient information inadvertently in an effort to make a correct identification. The ONC Power Team stated, "Responses to patient queries should not return any patient attributes that were not included in the original query, though it may be appropriate for the response to indicate other data that could be useful in matching this patient." The group also suggested additional research on specific metrics that should

⁵¹ A Health Care Identifier for Each Patient: Testing a system of establishing voluntary patient identification across multiple health care records to improve outcomes and reduce costs. Robert Wood Johnson Foundation, Grant ID 68195. October 29, 2012. <u>http://www.rwjf.org/en/research-publications/find-rwjf-research/2012/10/a-health-care-identifier-for-each-patient.html</u>

⁵² National Strategy for Trusted Identities in Cyberspace: Enhancing Online Choice, Efficiency, Security, and Privacy. The White House. April 2011. <u>http://www.nist.gov/nstic/</u>



be returned in response to a query, and recommended that a query response provide a uniform resource locator (URL) link, explaining the matching approach used and providing a point of contact for more information. One study also reached the conclusion that patients who are concerned about the exposure of their personal health information may be more likely to withhold information from their physician.⁵³ Health care organizations are increasingly concerned about patient identification from the perspective of fraudulent use of individuals' medical identity.⁵⁴ Meanwhile, privacy advocates are concerned about the use of patient data without the express permission of the individual,⁵⁵ patient access to their electronic health information,^{56,57} and about potential security risks around the increased use of databases for research and public health.⁵⁸

Private Industry Work on Patient Identification and Matching

Various industry collaborations are currently underway to approach health IT issues that have included new approaches to patient identification and matching.

Healtheway

In 2011, Healtheway created a task group to study the issue of patient matching. The group developed a number of recommendations, including:

- Share lessons learned from the eHealth Exchange, (e.g. educational sessions, testimony to the HIT Standards Committee and HIT Policy Committee, etc.).
- Standardize the list of identity attributes used for correlation purposes.
- Develop best practices for incorporating patient discovery into clinical workflow (e.g. checking as patients make appointments, etc.).
- Expand the use and testing of IHE Patient Discovery Profile to leverage additional attributes (e.g. use of past identifiers).
- Further explore whether/how broadcast queries may be used among Exchange participants, and if so, the parameters under which global queries are necessary and appropriate.

Healtheway is also working with the Care Connectivity Consortium (CCC) to use eHealth Exchange standards to create interoperability amongst the CCC members. The project includes an agreed-upon method of matching the patient records of participants.

⁵³ Concern About Security and Privacy, and Perceived Control over Collection and Use of Health Information Are Related to Withholding of Health Information from Healthcare Providers. Israel Agaku et al. J Am Med Inform Assoc. amiajnl-2013-002079 Published Online First: 23 August 2013.

http://www.ncbi.nlm.nih.gov/pubmed/23975624

 ⁵⁴ Medical Identity Theft: Recommendations for the Age of Electronic Medical Records. California Attorney General. October 2013. <u>https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/medical_id_theft_recommend.pdf</u>
 ⁵⁵ Patient Identification and Matching. Patient Privacy Rights letter to Dr. Jacob Reider. Dec. 16, 2013. http://patientprivacyrights.org/wp-content/uploads/2013/12/PPR-Patient-Matching-Testimony-for-12.16.13.pdf

⁵⁶ Consumer Principles for Health and Care Planning in an Electronic Environment. Consumer Partnership for eHealth. November 2013. <u>http://www.nationalpartnership.org/research-library/health-care/HIT/consumer-principles-for-1.pdf</u>

⁵⁷ Patient Access to Information and Interoperability. Center for Democracy & Technology. Letter to Kathleen Sebelius. April 22, 2013. <u>https://www.cdt.org/files/pdfs/Interoperability-RFI.pdf</u>

⁵⁸ Matching known patients to health records in Washington State data; v0.6. Latanya Sweeney. Data Privacy Lab, Harvard University. <u>http://www.Dataprivacylab.org/projects/was</u>



CommonWell Health Alliance (CommonWell)

CommonWell Health Alliance, a vendor consortium launched in March 2013, has as a main goal patient linking and matching, and that it "supports patient matching by explicit patient identifier, key patient demographic data, or clinical data. Over time, the patient linking and matching solution will track and normalize a patient's identity across care settings regardless of the native patient indexing system or EMPI." ⁵⁹ CommonWell currently has seven vendor members representing 42 percent of the acute and 23 percent of the ambulatory EHR market in the United States. The organization announced in December 2013 that it plans to launch interoperability services in early 2014 at select provider locations in Chicago, Illinois, Elkin and Henderson, North Carolina, and Columbia, South Carolina.⁶⁰

Workgroup for Electronic Data Interchange (WEDI)

The business-oriented WEDI and its charitable WEDI Foundation arm issued a roadmap for health information technology in December 2013 with 10 recommendations focused on patient engagement. Among these were recommendations on patient matching that called for standardizing the patient identification process across the healthcare system, a consumer awareness campaign, and convening the industry to identify best practices for patient matching.⁶¹

American Health Information Management Association (AHIMA)

AHIMA, whose members manage health records in hospitals and other organizations, has produced a number of reports and presentations focused on improving the quality of data maintained in health records. AHIMA has focused on best practices on such topics such as data governance, reconciling and managing enterprise master patient indexes, assessing and improving EHR data quality, and limiting the use of the Social Security Number in healthcare.^{62, 63, 64, 65}

Healthcare Information and Management Systems Society (HIMSS)

HIMSS created a Patient Identity Integrity Work Group in 2008, composed of volunteer industry experts, to address concerns raised from a variety of sources about the need for guidance in understanding the complex issues surrounding the integrity of patient demographic data. The work group issued a report in December 2009 that suggested setting data definitions for key demographic data, defining minimum data attributes to be used in matching algorithms, researching algorithm effectiveness, and establishing an industry standard method of computer duplicate record rates in MPI databases. Longer term recommendations included following up on study results to determine recommended algorithm standards, and the eventual adoption of a

⁶¹ 2013 WEDI Report. WEDI Foundation. Dec. 5, 2013. <u>http://www.wedi.org/topics/2013-wedi-report</u>

⁵⁹ Commonwell Health Alliance website. <u>http://www.commonwellalliance.org/services</u>

⁶⁰ CommonWell Announces Launch Geographies and Participants. Press release. December 11, 2013. <u>http://www.commonwellcommunity.com</u>

⁶² Butler, Mary. "Keeping Information Clean: New Information Governance Efforts Challenge HIM to Sort Out Dirty Data." *Journal of AHIMA* 84, no.11 (November–December 2013): 28-31.

http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_050467.hcsp?dDocName=bok1_050467 ⁶³ Reconciling and Managing EMPIs (Updated). *Journal of AHIMA* 81, no.4 (April 2010): 52-57.

⁶⁴ Assessing and Improving EHR Data Quality. AHIMA. *Journal of AHIMA* 84, no. 2 (March 2013): 48-53. <u>http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_050085.hcsp?dDocName=bok1_050085</u> ⁶⁵ imiting the Use of the Social Soci

⁶⁵Limiting the Use of the Social Security Number in Healthcare. AHIMA. *Journal of AHIMA* 82, no.6 (June 2011): 52-56.

http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_049016.hcsp?dDocName=bok1_049016



unique patient identifier to reduce dependence on algorithms. ⁶⁶ HIMSS is currently working on an update to the 2009 Patient Identity Integrity report.

Conclusion

Work on patient identity and matching has accelerated in the past few years, with all sectors of the industry working to develop a better understanding of the root problems and identifying potential solutions. Much of the work has demonstrated that there is no single solution that will ensure patients' health information is accurately matched 100 percent of the time, with zero false positives and false negatives. While a universal patient identifier has been discussed by many in the industry as a solution, they acknowledge that many of the matching mechanisms and data quality methods that are in practice today would still be required. Emerging technologies such as biometrics and smart cards have some stakeholders optimistic that they could reduce reliance on complicated statistical analyses of demographics in data sets. Yet, veteran health information management professionals point out that improving the accuracy of data entry and maintenance will always be needed, and they offer a number of fundamental best practices in patient identity integrity and registration training.

⁶⁶ Patient Identity Integrity, a White Paper by the HIMSS Patient Identity Integrity Work Group. December 2009. <u>http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf</u>



Appendix B: Best Practices for Data Quality

Patient demographic data must be gathered more accurately and completely wherever patients are interacting with a healthcare facility and processes must be implemented to ensure that errors and duplicate records are corrected. Quality improvement, in any field, is recognized as a set of systematic and continuous actions that lead to ongoing improvement. Professionals who manage patient data have been aware of data quality issues for many years, and have developed some well-accepted best practices for quality improvement around patient identity integrity (PII). This section provides an overview of some of the industry's current thinking on how healthcare organizations can improve data quality and maintain PII and highlights a few best practices in the field today.

Data Entry Policies and Processes

Nearly every conversation about patient identification and matching invariably comes around to the role of data quality. Typos, misspellings, transpositions, and fields left empty or filled with false data can all cause problems downstream from the registration desk. Registrars and front desk staff are primarily responsible for the correct entry of patient data into the system. Unfortunately, the average annual wage for registration personnel at hospitals is less than \$29,000, and people in these positions tend to move to other jobs frequently, making training and staff retention a challenge.⁶⁷

Patient registration processes offer the opportunity to confirm a patient's identity to reduce the chance of a mismatched or duplicate record. Some organizations require that their registrars request to check a photo ID when a patient arrives at a health care facility. While this practice is also helpful for reducing the risk of fraudulent use of a patient's insurance coverage, it has significant limitations in populations of children and in those facing social disparities. Another option is provided by some EHR systems that feature the ability to store a photo within a patient record. In addition to helping prevent duplicate records in the system, this feature can also add value to other patient safety efforts to confirm a patient's identity with two data attributes (name and date of birth, typically) each time a medical test or intervention is carried out.

The HIMSS PII Work Group pointed out in its comprehensive 2009 white paper on patient matching that standardized procedures for the collection, input, and query of information are among the most effective methods for maintaining high quality. Among its specific recommendations:

- Data collection should be done with the patient present for any follow-up questions or data validation, and the patient should be given an opportunity to review the record.
- Staff should be trained to verify elements before entering them to prevent assumptions.
- If patients are asked to fill out a form, they should be created with a "box per letter" format so each letter is separately written. The form should make clear the format of dates (i.e. MM/DD/YYYY).
- The use of look-up entries rather than free-text entries should be incorporated whenever possible.

⁶⁷ http://www.indeed.com/salary/q-Medical-Registrar-l-United-States.html



• In data processes, the number of times information is collected, input, and queried should be minimized. 68

Cultural competence

Registrants and other healthcare personnel must maintain cultural and socioeconomic sensitivity when they are asking patients for personal information for their patient record. Healthcare organizations are increasingly attuned to this issue, with Stage 2 Meaningful Use measures requiring the collection of race and ethnicity.⁶⁹ Registrars may feel uncomfortable at times asking for these and other data attributes where a patient presents as a different race or gender than the record indicates or than might be physically observed. The American Hospital Association offers advice on cultural competence in these circumstances.⁷⁰ In addition, the Fenway Institute has produced a document on how to gather data on sexual orientation and gender identity in a healthcare setting.⁷¹

Registrar/Front Office Staff Training

To maintain a high-quality registration process, where accurate data is correctly entered into the demographic fields of a record, and the creation of duplicate records is minimized, strong training programs are essential. Many training programs begin with an emphasis on the importance of accurate patient data to patient safety, as well as organizational efficiency, with examples of how incomplete, inaccurate, duplicate, or overlaid patient records can result in harm to patients. The individuals who may need training should include anyone who originates or modifies a record within the MPI or EMPI in addition to personnel who create records that are maintained in other information systems, such as lab or billing systems. The HIMSS PII Work Group notes this could include registrars, schedulers, physicians, nursing staff, and laboratory personnel, along with health information management and business office staff.⁷² Information system staff also need to be aware of how the system works on the front end to better manage the system's back end. The HIMSS panel made the following specific recommendations on training:

- Training should take place upon initial employment as well as through continuing education.
- Performance should be assessed on an ongoing basis on topics such as sound patient interview techniques and data entry conventions and requirements.

⁶⁸ Patient Identity Integrity, a White Paper by the HIMSS Patient Identity Integrity Work Group. December 2009. http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf

⁶⁹ Stage 2 Meaningful Use. Centers for Medicare and Medicaid Services. <u>http://www.cms.gov/Regulations-and-</u> Guidance/Legislation/EHRIncentivePrograms/Stage 2.html

⁷⁰ Reducing Health Care Disparities: Collection and Use of Race, Ethnicity and Language Data. Health Research & Educational Trust. August 2013. www.hpoe.org/EOC-real-data

⁷¹ How to Gather Data on Sexual Orientation and Gender Identity in Clinical Settings. The Fenway Institute. http://www.fenwayhealth.org/site/DocServer/Policy_Brief_HowtoGather..._v3_01.09.12.pdf?docID=9142 ⁷² Patient Identity Integrity, a White Paper by the HIMSS Patient Identity Integrity Work Group. December 2009.

http://www.himss.org/files/HIMSSorg/content/files/PrivacySecurity/PIIWhitePaper.pdf
counties in central and

northeastern Pennsylvania.



Registrar/Front Office Staff Certification

In addition to rigorous training programs and policies for requiring training within an organization, there is also value in encouraging more

Best Practice Highlight: Geisinger Health System

Geisinger Health System is an integrated health services organization widely recognized for its innovative use of the electronic health record, and the development and implementation of innovative care models. The system serves more than 2.6 million residents throughout 44

widespread use of certification programs for registration and front-desk personnel. The NAHAM maintains a certification program for healthcare registration personnel, offering two levels of certification training and testing for healthcare access employees (access managers and access associates).⁷³ There are currently over 5,700 certified associates and approximately 600 certified managers in the United States, though a large percentage do not maintain their certifications from year to year. It is not clear what percentage of hospitals in the United States currently require certification of registration personnel, but according to NAHAM, some hospitals do require certification to be achieved during a probationary period or to be considered for promotions.

Managing Duplicate Records

Minimizing the initial creation of duplicate records and having rigorous systems in place to find and resolve duplicates are two concrete ways that healthcare organizations can improve patient matching and reduce the risks of false positive and false negative matches. There are existing tools available to help organizations analyze patient records and indices to identify potential duplicates, but human review of the potential duplicate list is also critical. Organizations should evaluate where duplicates are being created, and work with staff members who may need more training to lower the rate of duplicate creation. Steps should be taken to resolve the duplicates including notification of other business units in the organization where the duplicate record might also be stored. Even in smaller healthcare organizations where they may not have the technical

Geisinger has invested in a comprehensive training program for personnel who access and create patient records. Because Geisinger maintains a strict security process for its registration and scheduling system, staff members must attend training before they are given access to patient records.

Upon hire, Geisinger Heath System employees attend a nine to 14 day training track, before they are able to access and create patient records. Geisinger's Education Department can accommodate about 40 employees within each training track. Registration staff are trained on the organization's process for identifying patients (confirming three of five core demographic elements), the proper way to register patient demographics, insurance processes, and scheduling for inpatient and outpatient appointments and procedures.

Registration staff members learn about the importance of patient identity integrity to the whole organization, and its impact on quality of care, patient safety, and controlling costs. Trainees learn about the potential impacts of the wrong information being used in patient care. They are instructed to report any potential duplicate record to the Medical Records Department.

Emergency Department (ED) registrants are required to pass a validation test to confirm they know how to identify patients properly in a time-sensitive situation. An ED staff member who does not pass the test must attend an additional training course related to patient search methods, and be retested.

The Education Department also provides additional training as necessary to registration staff to ensure they are following the most current processes. This training can be instructor-led in a classroom, or computer-based. The training program was developed internally by the Education Department with support from external subject matter experts, as well as subject matter experts from within the organization.

As part of Geisinger's data quality program, the Quality Assurance Department monitors registration errors and provides feedback to management to be shared with registration staff. The Quality Assurance Department works with the Education Department to provide additional training to staff who make continued registration errors.

⁷³ Candidate Guide to Certification 2014. National Association of Healthcare Access Management. http://c.ymcdn.com/sites/www.naham.org/resource/resmgr/Certification/NAHAM Candidate Guide to Cer.pdf



tools for such data analysis, a staff member with the specific skill set known as data profiling should be responsible for reviewing records for potential duplicates on a regular basis. Loosely defined as "the practice of learning things about a data set by someone who may not be an expert in the subject matter of those data," a data profiler uses analytic techniques to determine the attributes of the information in a data set, its quality and whether the data set contains the information that is needed.

Texas Health Resources, a 13-hospital health system in North Texas, reports that it reduced its duplicate records rate to 0.36 percent over several years by focusing on four tasks: scrubbing the MPI, identifying and selecting the right patient records, educating key stakeholders about avoiding duplicate creation, and monitoring performance. The project started by studying when and why duplicate records were created (for instance, an inconsistent approach to use of legal names versus nicknames and other formatting issues) and developing education and training programs for registration and medical records staff. The organization also incorporated tools into the search function of the record system, so that results are sorted based on the likelihood of a records match. Among the benefits of reducing the duplication rate has been a decrease in the number of accounts receivable days, from 48.4 in 2006 to 38 in 2012. Texas Health Resources' new process includes registration data scrubbing, electronic identification card readers, stakeholder education with follow-up training, and real-time alerts to prevent duplicate records from being created.⁷⁴

In another published case study, Children's Medical Center in Dallas, Texas, found it needed to clean up its MPI when its high rate of duplicate records made it difficult to find patients after a new EHR was installed. The EHR was configured to search for exact matches, so physicians looking up patient information were often missing key health data because records containing slightly different versions of the patient's name were not found (false negatives). After determining that the average cost of each duplicate record was 96 dollars and affected clinical care in four percent of the cases (delayed surgeries, duplicated lab tests, and imaging) the organization hired a contractor to assist in the data clean up. The process took 10 months and included an analysis of the hospital's entire identity integrity process, detailing the cause of every duplicate created, and an action plan to address those duplicates. Ultimately the duplicate rate was cut from 22 percent to .14 percent.⁷⁵

Consumer Involvement

Permitting consumer review of demographic information through electronic means, such as allowing patients to look at the computer screen being used by the registration clerk, or requesting validation of a patient's data through the use of a patient portal, is somewhat controversial among healthcare organizations and associations representing them. While most concerns center on the potential for a HIPAA breach in the event that a patient is mistakenly shown another person's

⁷⁴ Minimizing Duplicate Patient Records to Maximize Cash Flow; Case Study of Texas Health Resources. Patricia Consolver. Healthcare Financial Management Association. May 2013. http://www.guadramed.com/guadramed/OMMedia/documents/Articles/Minimizing-Duplicate-Patient 0513 RCS-

reprints_consolver.pdf ⁷⁵ Duplicate Records Compromise EHR Investment. Healthcare Financial Management. August 2009.

http://www.healthcaretechnologyonline.com/doc/duplicate-records-compromise-ehr-investment-0002



record, there is also growing support for finding solutions that can include the patient in data quality efforts and also ensure the privacy of their information.

ONC's Health IT Policy and Standards Committees each addressed the need for consumers to have access to their electronic record and an easy method for requesting corrections to it. The Standards Committee Power Team, focusing on specific patient matching recommendations, suggested the following:

Providers should allow patients to verify the patient attributes the provider has recorded for them through a method such as sharing the data entry screen with the patient for review, providing the patient with a printed summary or on-line access to the data to help identify quality issues and utilize the methods provided by HIT developers to identify missing/unavailable data and approximate or questionable values at the time of data entry.⁷⁶

It remains unclear how prevalent this kind of practice is among healthcare organizations. Physicians who follow the principles of patient-centered care are accustomed to sharing the screen with the patient. Some organizations, such as Kaiser Permanente, have installed kiosks in patient

"With the increased focus on patient engagement through initiatives such as Blue Button and Stage 2 of Meaningful Use, there will be more and more opportunities for patients to digitally access information in their medical records, including personal demographic information. The next step should be providing patients with a simple means to alert providers to errors and be able to submit corrected information. Involving patients in the process can help improve data accuracy." Deven **McGraw, Center for Democracy and Technology**

care areas to allow patients to check in and see some of their record and have an opportunity to ask for corrections. However, these practices do not seem to be the norm. The 2008 Most Wired survey found just 13 percent of responding hospitals were using kiosks, but that number dropped to 11 percent in the 2013 survey. In 2008, five percent of respondents were using kiosks to allow patients to pre-register, and that rose to seven percent in 2013.77,78 A 2009 review of the use of patient kiosks by the California HealthCare Foundation concluded that a self-service touch-screen kiosk in a common area tends to be used for way finding and check-in, but that registration is time-consuming and would not work well when there is a line of patients waiting. To solve that problem, some vendors are offering a mobile option, such as a tablet, that a patient could take back to their seat in

⁷⁶ Letter to ONC. Health IT Standards Committee. August 17, 2011.

http://www.healthit.gov/archive?dir=HIT%20Standards%20Committee/2011/2011-08-17

⁷⁷ Survey: Hospital kiosk use expands beyond wayfinding information systems. Health Facilities Management. November 2008.

http://www.hfmmagazine.com/hfmmagazine/jsp/articledisplay.jsp?dcrpath=HFMMAGAZINE/Article/data/11NOV 2008/0811HFM_Upfront_Survey&domain=HFMMAGAZINE

⁷⁸ Phone call with Suzanna Hoppszallern, Most Wired/Health Forum. Jan. 16, 2014.



an ambulatory setting or emergency department.79

The Medical Center of Central Georgia has used patient registration kiosks and laptop computers since 2007 and reports that patients fill out demographic fields more accurately than when a registration staff person asks questions verbally and keys in the responses. The self-service registration process has had the side benefit of reducing registration times and improving both patient and staff satisfaction. ⁸⁰ However, NAHAM cautions that care must be taken with any process established for patients to electronically enter their own information or change their records, to ensure that standardized formats for demographic information are maintained and that all the needed fields are filled out completely. They believe that review by a registration staff member will still be required for patient facing electronic registration tools. However, new mobile apps for iOS and Android devices, such as the award-winning iBlueButton® app, are moving forward with capabilities to give patients and caregivers a secure electronic way to notify their providers of both clinical and demographic changes needed in their records with the Blue Button+ Direct Protocols.^{81 82}

There is also discussion within the context of the NSTIC initiative to work toward solutions where consumers could choose and manage their own unique digital identity for use in a variety of situations, including interactions with the healthcare system. This initiative is building on the idea of an identity ecosystem: "individuals can choose among multiple identity providers and digital credentials for convenient, secure, and privacy-enhancing transactions anywhere, anytime." It is being led by the private sector and supported by government. The advantage for consumers is that they could maintain a voluntary, secure identity that could be used in banking, shopping, accessing health records, and other sensitive transactions. This system is in the early stages of development, and its impact on patients interacting with the healthcare system remains uncertain.

⁷⁹ Touchscreen Check-In: Kiosks Speed Hospital Registration. California HealthCare Foundation. March 2009. http://www.chcf.org/publications/2009/03/touchscreen-checkin-kiosks-speed-hospital-registration

 ⁸⁰ Kiosks Let Patients Register Themselves. HFMA Patient Friendly Billing newsletter. <u>https://www.hfma.org/Content.aspx?id=3403</u>
⁸¹ http://www.ibluebutton.com/

⁸² <u>nup://www.ibiuebutton.com</u>

⁸² <u>http://bluebuttonplus.org/</u>



HIOs as Data Intermediaries

Once patient data leaves a healthcare facility for sharing in a health information exchange scenario, it becomes increasingly important that patients are Best Practice Highlight: Maine HealthInfoNet

HealthInfoNet is the not-for-profit statewide organization operating the statewide HIE for the State of Maine. With 34 of Maine's 38 hospitals connected to HealthInfoNet in December 2013, 100% of the hospitals are expected to be HIE participants by early 2014. HealthInfoNet is connected to 395 ambulatory provider sites including primary and specialty care practices, federally qualified health centers, mental health agencies, home health and long-term care providers. It is managing 1.3 million records in the exchange.

properly identified and duplicate records are kept to a minimum. There is an imperative on HIOs, whether private, community or region-based, or statewide, to ensure that patient record matching is as accurate as possible, and there are some up-and-coming best practices for organizations to consider. For example, while many HIOs do not currently provide feedback on potential duplicate records or mismatches to their members, some HIOs are beginning to see value-add to the overall quality of patient identity data across their member base, by reviewing records for potential duplicates or errors and notifying the organizations where those records exist, to resolve the issues. These interactions tend to take place through secure email or phone calls and work best when individuals at the HIO can work with people they know at the medical records departments of hospitals or clinics, and when the service is expressly mentioned as part of a member participation agreement.

Governance Initiatives

Competing priorities for C-suite attention and financial resources, can make it difficult to focus attention on the steps necessary to develop and

HealthInfoNet maintains two FTE analyst positions, with the duty of managing work queues of records flagged as potentially "same patient" records. When the IBM Initiate application encounters records from the same healthcare organization that appear to be the same patient, they are flagged as potential duplicate records. The HealthInfoNet analysts review the file of potential duplicates daily. Any records that appear to be duplicates are communicated back to the member organization (either by phone or by secure DIRECT email) to a specific contact person at the member organization. The organization is asked to look at the records and determine if the records indeed belong to the same person. If the organization determines that the two medical records should be combined into one surviving record, a merge is performed by the organization which is transmitted to HealthInfoNet via an HL7 merge message. For any organization that is unable to provide HealthInfoNet with an HL7 merge message, a manual merge is performed by the analyst at HealthInfoNet.

The HealthInfoNet analysts are able to make this process work, says Deborah Wilson (HIN Business Analyst), because they have developed good working relationships with individuals at the participating organizations. She says: "They look at this as a value-add service that can help reduce their liability and improve the safety of their patients. It may also help them with timely payments by identifying these problems sooner in their systems."

HealthInfoNet maintains a policy that it is a steward of its member organizations' data, not the owner. The participant agreements with its member organizations outline precisely how patient data can be used by the HIE. Through their efforts to identify potential data quality issues, and notify their members when those are found, HealthInfoNet is fulfilling its obligation in keeping patient records as accurate as possible at all points in the continuum of care.

maintain ongoing data integrity programs. These programs include tools to measure data quality and use of root-cause analysis and process improvement projects to identify and fix specific workflow problems. Historically, discrepancies within individual patient records were primarily seen as the domain of business departments within a hospital, such as registration and HIM departments, and had lower visibility to other parts of the organization. Managers responsible for maintaining patient records acknowledge challenges in their ability to elevate data integrity to the highest levels of healthcare as a strategic priority for improving patient safety. This appears to be changing as payment models are increasingly tied to the management of specific health conditions,



and Meaningful Use requires increased electronic exchange of patient records amongst providers. An emerging best practice in healthcare organizations is the establishment of more robust data integrity programs with

Best Practice Highlight: Group Health Cooperative

Group Health provides medical coverage and care to more than 600,000 residents in the State of Washington and North Idaho who are covered by health plans offered by Group Health Cooperative or their subsidiaries, Group Health Options, Inc., and KPS Health Plans. Nearly two-thirds of

members receive care at Group Health Medical Centers.

data governance structures that reach up into the Csuite and across all of the business units in the organization.

Once a data integrity program is established including senior executives, the assignment of interdisciplinary teams across many business units to work on specific data quality problems is one example of an action step that can be taken. For example, Peninsula Regional Medical Center in Salisbury, Maryland, created a committee with "representatives from HIM, patient registration, finance, IT, and labs that meets monthly to review duplication creation rates, discusses trends in registration data errors, and create new processes to correct the mistakes." A data governance program can also use data maturity models to evaluate the level of an organization's oversight of its data overall. This assessment assigns a maturity rating to each data system and the processes of managing the data in those systems. With maturity ratings between 1 (initial) to 5 (optimized), the organization can prioritize investments in moving different data systems to a higher level of maturity over time.

Group Health has established a Data Governance Board that includes senior business leaders from across the organization. They maintain four practice councils focused on various sections of the organization: group practice (which includes health information management), the health plan, Group Health's research institute, and enterprise services (departments supporting legal, financial, security, etc.).

Data quality projects are initiated by individuals within departments who see a problem and seek help from their practice council. Data Governance Board members themselves may also present issues for consideration. Those potential projects are then considered by the full board. An example of a recent project included two different teams who were measuring inpatient utilization but coming up with conflicting trends. Their processes were analyzed to figure out what they were doing differently.

The data governance staff helps manage these projects, along with data quality programs such as tracking data metrics and clarification of roles around data stewardship – what it means to be responsible for an application and its data. The staff would like to address other issues as well, including how to determine where in the data chain a correction to a patient record gets made. They also carry out regular education throughout the organization, raising the awareness of the importance of data quality to the organization.

Patient Identity Integrity Metrics

Patient identity integrity (PII) – the accuracy, quality, and completeness of data attached to or associated with an individual patient – impacts quality of care, patient safety, patient privacy, medical identity theft, health information exchange, and coordination of care, all top strategic priorities for healthcare organizations.⁸³ The accuracy of PII also impacts administrative and financial efficiencies, from prior authorizations to claims processing and patient billing. Improving PII requires measurement. To that end, in December 2012 HIMSS released a set of key performance indicators that can be used by organizations to assess their data governance performance. The

⁸³ Measuring Patient Identity Integrity: Foundation to Healthcare Analytic Success. Presentation to AHIMA 2013 Convention (slides). Barbara Demster, Stacie Durkin, Lorraine Fernandes. October 2013.



indicators "quantify the state of existing performance of business process and the technology that supports the Master Person (Patient) Index

Best Practice Highlight: Mayo Clinic

Mayo Clinic serves more than a million patients each year and has 131,000 annual hospital admissions.

(MPI) including algorithms, matching logic, and threshold selections, among others...(and) provide management tools to identify business process variances across time, location, source, and employee."⁸⁴

The HIMSS document includes a list of the data that should be collected about an EMPI to measure performance, such as database size, person population, total matches, etc. An organization identifying these data points for measuring performance could use them in a variety of ways: to support its continuous improvement of PII, as a basis for performance reviews, and to assess vendor performance.⁸⁵

Sisters of Charity of Leavenworth Health System (SCLHS) reports that it uses many of the HIMSS metrics to manage the work done by its data integrity team. The Denver-based system has eight acute care hospitals in three states using an Epic patient record, along with about 200 physician clinics that are being incorporated into a single master patient index. The organization established a data integrity team in 2008 to centralize the patient record "cleanup" process across the entire system. The team reports to the data integrity manager, who reports to the vice president for health information management, and in turn reports to the systems' chief information officer. Mayo maintains a strong data governance and standardization process that runs through its clinic sites and hospital sites, and which is now being extended to the Mayo Clinic Health System sites. These efforts are overseen by Mayo's Enterprise Data Governance, which sets and enforces policies, principles and standards through a Data Governance Committee, composed of 15 members from all three campuses. The committee includes Mayo's chief information officer, chief medical information officer, chief planning officer and others. The data governance work is an essential part of Mayo's Culture of Safety.

The organization's Master Patient Identification Index (MPII) has two components, operational and system support. Operationally, MPII User Support maintains the quality of patient demographic information for the enterprise. MPII IT and users work daily to provide system controls, promote user awareness, and focus on high-risk areas such as emergency departments.

Mayo also built a system called Patient Misidentification Notification System (PMNS) which provides automatic notification to more than 400 other systems within Mayo whenever two patient records need to be merged or unmerged. That notification includes a level of criticality that can be set by the user, so the recipient knows how urgent it is for the change to be implemented.

Mayo assigned a "One is Enough" project team to provide registrant supervisors with a weekly list of misidentifications from their staff to facilitate staff awareness and root causes of misidentifications. In managing the accuracy of patient identification at its hospital and clinic sites, Mayo has also found that explaining to patient record users how much each error costs the system in efficiency encourages them to strive for greater accuracy and data quality.

The data integrity manager at SCLHS works closely with the registration department to understand its processes and challenges, provide feedback on the creation of duplicate records, and work with registration managers to identify root causes of duplicates and overlays that may require additional training of individuals or workflow changes. Maintaining internal metrics helps SCLHS monitor the rate of problems in patient records and the success of interventions.

Even if an organization does not move directly to adapting the HIMSS indicators for process improvement or internal quality measurement purposes, the simple act of pulling together a

⁸⁴ HIMSS Patient Identity Integrity Key Performance Indicators, December 2012 <u>http://www.himss.org/files/HIMSSorg/content/files/PII03_Key_Performance_Indicators_Final.pdf</u>

⁸⁵ Measuring Patient Identity Integrity: Foundation to Healthcare Analytics Success. Presentation to AHIMA 2013 convention. October 2013.



multidisciplinary team to determine how to define and track metrics, could be of value in elevating patient identity integrity as an issue and could be a new way to bridge communications between health information management departments and the other business units where patient data is being collected and stored.

Any effort to compare organizations using data integrity metrics should take into account those measures that may be affected by the use of varying data quality methods and algorithms. For instance, if an algorithm is tuned to be quite strict, it will produce a larger number of potential duplicate records; if that measurement is then used for benchmarking against other databases or systems where the algorithm may not be so finely tuned, it may not give an accurate picture of the organization's data integrity program.

Conclusion

There are many best practices currently being implemented by healthcare organizations across the country, and it is critical that these best practices be shared and adapted, particularly for smaller organizations. It is also true that there is a rapidly increasing need to improve the accuracy of patient identifying information in electronic record systems through careful human processes, as well as through more standard electronic data fields in those systems. In a more connected and coordinated healthcare environment, where the electronic sharing of patient records will be essential to improving care and lowering costs, the imperatives for building on current and emerging best practices through a broad and collaborative industry effort cannot be overstated.



Appendix C: Environmental Scan: Health Systems

Beth Israel Deaconess Medical Center, Boston, MA

Vendor

Beth Israel has used the eIndex product from Oracle (originally a Sun Microsystems product) since the 1990s.

Match Method

The internal matching algorithm is self-developed, proprietary, and probabilistic. Beth Israel achieves a match rate up to the high 90's percentile. The internal match process will offer a list of potential matches, when a match cannot be determined. The external algorithm, used between hospitals, looks for an exact match and is probabilistic and somewhat deterministic. The external matching process does not offer a list of potential matches, when a match cannot be determined with 100 percent certainty. They do not utilize XCPD or PIX/PDQ standards, as they feel the standards are lag the technology and are too long and complex to use.

Data Attributes

The main data attributes used by Beth Israel are name, gender, date of birth, and ZIP code. As for assigned numbers, they do not believe Social Security numbers would be helpful; however, they do use an internal ID but that is not left on the record when it is shared outside the organization. In a manual process review, address and phone number may be added to determine a match, but they have found that gender is easy to get wrong and is becoming more complicated over time. Parents' names could be added and help with specificity but may wreck sensitivity.

Manual Processes

Beth Israel maintains three full-time staff in medical records that are responsible for identifying duplicate records and merging/unmerging records. Beth Israel was not able to provide a dollar amount on what it costs to maintain that function.

Comments

Concerns were shared about certification criteria being too prescriptive, and that the market may be better suited to manage some of these issues. Beth Israel suggested that criteria should be more outcomes-based rather than prescriptive. They also noted that ONC could suggest matching processes organizations need to use to get above an 80% match rate.

Geisinger Health System, Danville, PA

Vendor

NextGate v7.0 is their MPI vendor. Geisinger utilizes Epic throughout its health system and has started a small pediatrics practice on Care Everywhere.

Data Attributes

Geisinger uses first, middle, and last names (including suffix), DOB, gender, whole Social Security number, address line 1 (break into street name and house number), city, state, ZIP code, and phone



number. Geisinger separates out each element of DOB and weights them separately. They have found that mother's maiden name reporting is inconsistent and additional data attributes would likely be a lot of work for changing their algorithm, however the Social Security number is critical. The system uses whole Social Security number when sending data to the HIE, and uses final four when pulling data from the HIE. Geisinger does not disclose the Social Security number in Epic Exchange. Geisinger has found that in central Pennsylvania, where most of their patients live, address is very stable because patients do not move a great deal. However, Geisinger has had to adjust their matching to address their Amish population that does not have a Social Security number or phone number. Additionally, many older female patients share a Social Security number with their husbands and many are widows, making matching of their records challenging.⁸⁶

Match Method

Geisinger utilizes NextGate, which is a probabilistic algorithm. Geisinger also uses subscription services to verify demographic data, such as Passport. The system does not use IHE PIX/PDQ protocols.

Manual Processes

Geisinger's HIM department reviews a daily report for obvious registration errors. They proactively seek out death data from sources such as Lexus Nexus, to update records on a regular basis. For staff members who need assistance with searching for a patient, the HIM department provides a help desk.

Best Practices

Geisinger maintains a registration training team and requires registration staff to pass a nine-day class. Approximately 40 people each month go through this classroom-based training. It covers the principles of patient identity integrity, the organization's process for identifying patients (confirming three of five core demographic attributes), insurance processes, and details of inpatient and outpatient registration. Only people who have been through this course and received clearance can access the Geisinger registration system to create records. Emergency department staff who work with patient records are also given a written test to confirm they know how to identify patients properly in a time-sensitive situation. Recently, Geisinger restricted the DOB field for all registrars except the ED, Admitting, and HIM. If the registrar needs to change a DOB, they must call HIM. HIM will decide whether or not the DOB can or should be changed based on the information provided by the registrar. With these processes in place, Geisinger has about 1,200 people who can create a new patient record. In addition, Geisinger provides employees with online modules and occasional webinars on registration topics. The organization provides feedback to registrars on errors, and they are also expected to self-identify any time they suspect they have created a duplicate record. Geisinger also has a patient portal that offers patients an opportunity to check their demographic data, but they cannot update their data in the Epic system directly.

⁸⁶ Until the mid-1970s, SSNs were not regularly assigned to non-workers. The Railroad Retirement program, which a large number of central Pennsylvania residents are beneficiaries of, typically assigned an SSN to the worker only, and their wives utilized the same SSN.



Comments

For Geisinger patients, Social Security number is the best identifier, with DOB also being fairly important. Geisinger does not feel that adding mother's maiden name as a data field would be useful, as many people do not know it or knowhow to spell it and adding it as a requirement would cause large changes to the customized algorithm. Geisinger will potentially move toward biometrics in the future.

Group Health Cooperative, Seattle, WA

Vendor

Group Health uses one instance of Epic, and is in the process of implementing IBM Initiate, primarily to aid matching between its health plan (administrative data) and delivery system (clinical data). In addition, Group Health has three internally developed legacy products, which meet individual business unit needs, and for matching medical record numbers (MRNs) from outside to internal data when Group Health is building interfaces to external organizations.

Data Attributes

Epic Care Everywhere uses first and last name, DOB, gender, address, and phone number. Initiate uses a weighted probabilistic algorithm for name (first, middle, last, suffix) alternate name (first, middle, last, suffix), DOB, Social Security number, residential and mailing addresses (all lines), Medicare ID, Medicaid ID, and Railroad Retirement ID. Historic address would be helpful, but is not collectible in Epic. Group Health does not use Social Security number except for queries with the Social Security Administration (SSA), using Care Everywhere. Epic does not standardize formats, while Baseline standardizes a few. Initiate uses a number of standardization functions; however, Group Health is not currently using the standard USPS format but will move to that in the future to reduce mail costs. Group Health has a plan to launch foundational master data management with Initiate in December 2013.

Match Method

Group Health currently uses Epic's weighted deterministic matching process for health information exchange. There are five weighted attributes, if four match entirely and a fifth does not it highlights the discrepancy. Initiate creates an enterprise identifier (EID) that is used internally and can change as records are linked and unlinked, but it is not used externally.

Best Practices

Group Health manages potential duplicates daily. Recently they began a process improvement project with a cross-functional team (HIM, business operations, health plan membership) to reduce duplicates. The project is working to standardize the process for assigning new member/patient numbers. Part of the process is reducing the number of departments that can create new patient records. Additionally, registrars are trained to confirm that the information displayed is correct and to ask searching questions such as "Have you ever gone by a different name?" The team is now monitoring monthly duplicate number assignment volumes at the user and department level and reporting defects to department managers. Group Health's goal is to reduce duplicates by 50%, but they have not yet achieved this benchmark. Approximately 100 duplicate records are created every



month, but interviewees said that number could be higher, and since they are likely not catching all duplicates. This process will become more automated once IBM Initiate is in place.

Comments

Group Health feels that improving patient matching is a combination of people, process, and technology. Standard ways of entering and editing data in source systems must be required. Increasing the number of required data points to verify, before assigning a new patient number, would also help with accuracy. Currently Group Health requires matching on three data points. Requiring validation by asking for patient driver's license or other form of picture ID should also be considered. Group Health does favor a unique identifier for solving some of these patient matching issues, and are hoping Initiate makes a significant difference in their match rates.

Group Health would like to see standards for data quality among trading partners. Vendors need to help smaller organizations because it is not feasible for a small organization like a medical practice to be thinking about matching algorithms. Group Health would not support adding any new data attributes until there is statistical analysis to show it improves things, as it would be expensive and the organization responds to many audit requests already. Finally, though the HIM department places a high value on accurate matching of patient records, it is not currently a visible priority for the overall organization as it has not been linked to other patient safety initiatives.

Kaiser Permanente, Oakland, CA

Vendor

Kaisers uses the Epic EHR, which they call KP HealthConnect, as well as Care Everywhere for sharing across organizations, and KPOrchestra for their internal HIE. For matching, Kaiser utilizes a customized version of Epic's matching capabilities. Kaiser has control over the algorithms and has shared them with Epic to improve the product. Kaiser has 17 instances of Epic across its seven regions. This has created an environment where a patient has multiple MRNs if they have visited an office in more than one region. Kaiser is working to fix this issue, potentially through the implementation of IBM Initiate as an EMPI. Health plan membership information is maintained separately from clinical data, but users can check patients against the health plan data to locate their clinical record.

Data Attributes

Kaiser uses last name, first name, DOB, and gender for matching. They are looking at alternate address and cell phone to deal with separate regions of California, and are also considering attributes that will not change over time like eye color. Helpful data attributes suggested by Kaiser include maiden name, cell phone, and previous address because it links patients to a household. Kaiser has had a strict organizational policy against using Social Security number (even for transactions with the SSA), but are evaluating its use with its policy committee.

Manual Process

Kaiser has a full time employee (FTE) assigned on a national level to work on potential duplicate reports. In addition, Kaiser offers system users the ability to review a problematic match by phone with medical records staff to resolve the issue.



Rates

Match rates across regions or with outside Epic partners are around 50 to 60 percent. The internal rate within one instance of Epic is over 90 percent.

Comments

Kaiser is starting to put the infrastructure in place to assign an EID across regions. Kaiser indicated that it is difficult to measure false negatives since no one knows there is a false negative unless they discover there is actually a match. Kaiser has a fairly successful patient portal and has utilized kiosks in the past. Kaiser's view is that they need to trust their patients on the information they put in their record but feels that self-reporting is good for some things and not others.

Kaiser's suggestions for improving interoperability include: changing the suffix on the CCD header from null to none value; and generally on the CCD header a value of null should signify the data was not collected, and none should mean that the patient does not have that data attribute (i.e. suffix). Kaiser representatives were favorable to the idea of Meaningful Use being used to make these changes.

Mayo Clinic, Rochester, MN

Vendor

Mayo has developed an in-house MPI, which is common across all facilities. They use the GE Centricity EHR in the Rochester facility and the Cerner EHR in Florida, Arizona, and the facilities and providers in the Mayo Clinic Health System.

Data Attributes

Mayo uses first, middle and last names, DOB, gender, a local patient identifier from other Mayo facilities or Mayo Clinic number, and Social Security number when available (Arizona does not allow for the release of Social Security number). Mayo's Rochester clinic has been trying to collect maiden name, while other clinics are working to standardize previous name. In addition, while Mayo accepts the USPS format from other systems, at the Rochester clinic, they standardize address to the Gregg Reference Manual style.

Match Method

Mayo uses deterministic matching.

Manual Processes

Mayo has built a Patient Misidentification Notification system. When false positive or false negative is found, the system notifies the 130 applications that are downstream of the issue. Each system must then correct the problem and respond that they have fixed the record. If it is a critical merge/unmerge they will receive a robo page, telling them to fix the error immediately. When searching for a patient record, staff searches for critical data attributes three times, three different ways. In addition, Cerner users are launched into a MPI query using a web tool called Patient Demographic Inquiry, if the patient is not found initially in the Cerner system.



Governance

Mayo maintains a strong data governance and standardization process. Patient misidentification numbers are high across the enterprise for many reasons. Fortunately, most misidentifications are found and corrected before the patient arrives for services. The Patient Misidentification "One Is Enough" team had a plan to visit Mayo Clinic Health System sites and offer training, but the resources for the program were reduced. There was also a Patient Misidentification charter that they intended to begin in the emergency departments of the Mayo Group Practices and use LEAN methodology to improve misidentifications in those three sites, but the charter did not make it through the approval process.

Registration and Training

There is a standardized registration process at Mayo hospitals and clinic sites, and they are working to improve consistency at smaller sites. Mayo uses photos in the patient record and require ID to be presented if a patient is changing their name or critical demographic information. The patient receives a prefilled form to confirm her demographic data at the appointment; forms are scanned into the system, and staff enters data from work queues, though this will be automated in the future. Prospective patients can also complete all information through Mayo's patient portal.

Costs

Mayo has calculated that each misidentification costs at least \$1,200, and some have cost hundreds of thousands of dollars and taken months to resolve. A misidentified patient in the hospital is the most costly because of the potential for treatment errors and/or an extended stay.

Rates

Mayo has good data on misidentifications, with a false positive rate of 0 to .1%. However, they would like to do more research on their false negative rates.

Comments

Mayo maintains one portal for sharing between sites, but has not joined regional HIOs, as it is very complicated in their multi-state environment. They are however a founding member of the CCC, and are working to exchange data via Direct and Healtheway.



Appendix D: Health Information Organizations (HIOs)

Colorado Regional Health Informational Organization (CORHIO), Denver, CO

Vendor

Medicity ProAccess, rebranded by CORHIO as PatientCare 360.

Data Attributes

The primary attributes used for matching include: first and last names, gender, DOB, and SSN. The secondary attributes include: middle initial, complete address, and nicknames. SSN is collected and retained whenever available, but is not displayed in query results. All participants are sending CORHIO a partial SSN at minimum, some send all nine digits. Currently Medicity does not establish an enterprise ID (EID), but they are considering maintaining a CORHIO EID in the future.

Match Method

Medicity utilizes a deterministic match that is not weighted.

Manual Processes

CORHIO mostly opts-out patients manually. This process has allowed them to find many duplicate patient records. In addition, CORHIO runs a monthly presumed matching report; however, they do not perform manual merges themselves. Rather, Medicity performs all merges when they are notified by CORHIO. CORHIO attempts to fix problems at the source, with approximately one FTE working on data integrity with source system participants.

Feedback to Participants

CORHIO notifies hospitals when a possible mismatch or duplicate is found and works to help them clean up data. They also work with new members on data cleanup on the front end, spending six to nine weeks to integrate a new member's EHR data into the HIE.

Governance

CORHIO has a policy committee and user group, which has representatives from all users statewide. The user group sets the standards for data feeds from users, including which patient demographic attributes will be sent to CORHIO.

Rates

CORHIO has a mismatch rate of less than one percent, as patient records are not matched unless all five primary matching criteria are matched. The false negative rate is higher than 1% because CORHIO does not manually merge confirmed matched patients, but works with data providers to improve attributes at the source. As a subset of the Users' Group, CORHIO will be establishing a HIM committee to work across the community to address patient identity.

Comments

CORHIO has an opt-out system for patients, where information is still collected by the HIO from hospital ADT feeds. Currently 31 ADT feeds are populating CORHIO's MPI, with 2.6 million unique



patients (57 percent of Colorado's population). CORHIO will cover 93 percent of hospital beds in Colorado by mid-2014, and is connected to Quality Health Information Network, the other large HIO in the state. CORHIO recommended a review of the CCDA header row to see which demographics are required and optional and whether there are standards for the demographics. Medicity has stated they will need an ADT feed along with the Direct message in order to match the patient record to make the CCDA available for query.

CurrentCare Rhode Island Quality Institute (RIQI), Providence, RI

Vendor

InterSystems provides the HIE technology, and QuadraMed Smart IX and Smart Merge v 9 provide the MPI.

Data Attributes

CurrentCare uses first, middle, and last names, DOB and gender. Common aliases, address and phone number are also employed for first name matching. They do not collect Social Security number since collection of Social Security number opens the door to the financial services breach reporting requirements, which is more onerous than HIPAA. CurrentCare adds an EID to the record, but does not share it outside of the HIO. External MRNs are used for future matching when they have already been attached to an MPIID (EID).

Match Method

CurrentCare has a hybrid match method for initial MRN linking (as it must match some fields but still meet a probabilistic point threshold), deterministic for matching where an MRN already exists for the matching assigning authority.

Feedback to Participants

CurrentCare's opt-in process provides an opportunity for improving data quality, as a patient must enroll for their record to be included in the HIO. They work with 400 enrollment partners around the state of Rhode Island and have enrolled about 350,000 patients so far. Record duplication does still occur when patients sign up more than once through different providers, but a non-enrolled patient will never be matched. QuadraMed has a logic process that manages duplicates and does automatic merging, and CurrentCare has to manually handle the duplicates that are not automatically matched. CurrentCare has found that record reconciliation is a challenge when working with member organizations, as the issue of overlay records may take time to resolve when hospital members have competing priorities and/or capacity concerns. When a potential overlay is discovered, CurrentCare's compliance officer reaches out to the member organization with the information, if it is confirmed that there is an incorrectly merged record. CurrentCare revokes the records (they do not try to separate the data) and creates a new record so only the correct data is collected in the future. CurrentCare is currently receiving CCDs from practice-based EHRs, and works directly with the EHR vendors to standardize the demographic data that should be in the CCD. They have collected nearly 300,000 CCDs, with active clinical data to be used during care processes.



Rates

CurrentCare was not able to report estimated false positive or false negative rates as they do not know how they would find the errors. In the CurrentCare environment, if the data does not meet the matching threshold it will fail. However, if the data does meet the threshold and there are multiple potential record matches, it will be considered ambiguous and will ultimately fail to match. The records fall into an ambiguous match queue to be worked, and additional information or corrections are made so that future records will match.

Costs

CurrentCare pays \$92,000 annually for QuadraMed and approximately \$70,000 for operational staff to work duplications and ambiguous matches.

Comments

A Universal ID is the alternative to the current matching process that CurrentCare believes will have the largest impact.

HealthInfoNet, Portland, ME

Vendor

HealthInfoNet utilized 3M for two years but moved to IBM Initiate in January 2011. Thus far, they feel that IBM Initiate is proving to be much more robust, including a longitudinal history of a patient record.

Data Attributes

Attributes utilized by the IBM Initiate software to evaluate records for matching include: first and last names, middle initial, suffix, title, Social Security number, DOB, full address, home phone, alternate phone, and gender. In addition to the current attribute values, the matching logic uses the attribute history when it evaluates the records for matching. HealthInfoNet has discovered that the DOB field can be highly variable. Maine has a relatively high Somali population, which culturally does not recognize birthdays. When they immigrate to the U.S., they are assigned January 1 in the decade they were born. Consequently, many have the same name and same birthday. HealthInfoNet is also finding that increasingly their members do not provide Social Security number, with less than 50 percent of records having a full Social Security number.

Automated and Manual Processes

HealthInfoNet's automated match process uses an IBM Initiate application to link records for patients being seen at unaligned health delivery organizations in the State. When probabilistic attributes do not score at 99.6 percent, patient records are not linked until a person determines that they should be linked. HealthInfoNet staff review potential linkages and duplicates, using all available demographic information (current demographic data plus any historic demographic data residing on the record) as the basis for their determination. Overlay records (records that have had updates applied to the demographic information which result in significant changes to the record) are flagged for evaluation. Overlay records are resolved daily. HealthInfoNet staff work directly with organizations to resolve any overlays and when deemed necessary, remove incorrectly transmitted demographic data from the record. They have two FTEs who work to identify overlays,



linkages and duplicates. With additional staff capacity, they believe more false negatives would be discovered.

Feedback to Participants

HealthInfoNet communicates potential duplicates to data sources via email or phone. They feel that this works better than sending regular reports, but will take more manpower as more organizations join.

Rates:

HealthInfoNet was not able to report on estimated false positive or false negative rates.

Comments:

HealthInfoNet provides a mechanism for patients to request an audit of who has seen their record and when. They are working under a grant to develop a Blue Button portal for patients to access their HIE data and will point the patient to the holder of records if an error in their record is discovered. They have found that there is variability with the demographics in CCDs, and have struggled to parse the data out before sending it to IBM Initiate for matching and have instead focused on HL7-based exchange instead . HealthInfoNet's policy is to act as steward of data, without taking ownership, and they have created a new set of data release policies. They also suggested exploring some other validation methods for IDs, such as using the insurance exchange to validate Medicare/Medicaid ID for a state or regional HIO. They proposed that using Joint Commission accreditation criteria or payment criteria could help drive improvements in data quality.

Indiana Health Information Exchange (IHIE), Indianapolis, IN (and Regenstrief Institute)

Vendor

IHIE utilizes a matching product built and maintained by Regenstrief Institute.

Data Attributes

IHIE uses first and last names, gender, DOB, zip code (5 digit), and SSN when available, but a third of the time SSN is not provided. The SSN is provided when sending information out, and since there is a high level of trust among the HIO participants, this has been an acceptable practice in their region. IHIE was skeptical of hashing the SSN, because hashing hinders the ability to correct for transposition errors, and thus limits the discriminating power of SSN for automated matching.

Rates

Multiple analyses by the Regenstrief team reveals an average seven percent error rate for first name and five percent for last name. Of the millions of daily transactions processed by IHIE, 20,000 to 50,000 require adjudication using complex automated matching algorithms; the remaining transactions contain sufficiently discriminating identifiers enabling rapid matching using deterministic logic. IHIE's rates are on average 92 percent sensitivity and greater than 99 percent specificity.

Manual Processes

IHIE has a process to routinely review mismatches (typically monthly), but they do not have a process to discover the false negatives in the system.



Feedback to Participants

IHIE does not currently provide their members with information about potential duplicates, stating that historically feedback on data quality has not been a top priority for hospitals who are more focused on maintaining integrity internally for billing and administrative purposes.

Comments

IHIE and Regenstrief believe that additional processes allowing consumers to review their own data for accuracy should be explored. They also feel that increasing the consistency and standardization of data reported for quality and payment purposes would improve data quality across the system. Regenstrief is looking toward the future and anticipates that as matching systems operate against increasingly larger patient populations, added discriminating power will be needed to avoid false positives and confirm true matches. In the absence of a national unique identifier (which they believe would improve matching accuracy), the needed discriminating power may be obtained from additional demographics such as parent's names, city of birth, and also biometric identifiers. However, they noted that costs for implementing new technology and workflow changes may pose significant challenges and are concerned that the return on investment (improved matching accuracy) will be perceived as insufficient. They support creating a testing model for evaluating which data attributes have the most impact when added to a list of data fields in probabilistic matching. They also feel improvements can be made in the consistency of how data fields are filled out.

Michigan Health Connect (MHC)

Vendor

MHC uses Medicity and maintains roughly 2.5 million records in its MPI.

Data Attributes

MHC uses a number of demographic values contained in the messages, including first and last names, address, DOB, gender, and MRN or whatever unique organizational number is included by new organizations as they onboard to MHC. They then assign an overall MHC number that fits on top of all other numbers from other sources, but they do not send that out with query responses.

Match Method

MHC's system uses deterministic matching, with an HL7 merge to keep the database clean. There are 2 levels of MPI management within the system. Each organization has a data stage where their data is stored. This data stage is where the first layer of MPI management is done and will represent as clean an MPI as the organization maintains itself, through its registration and MPI management (merge/unmerge) activities. During the onboarding of new members, MHC (Medicity) looks at the ADT feed and builds the interface based on how the member organization is handling any merging of potential duplicates. Medicity does any unmerging that is necessary.

The second layer of MPI management is where identities are matched across data stages at the community level so that the same individual is identified uniquely across different organizations/data stages. MHC has not experienced their members complaining about duplicates; however, they have thought about using an external duplication analysis/reporting service to validate the deterministic rules/logic currently being used within the system. They are able to use



IHE profiles but they not heavily used. They are working with their members to better fit the CCD and CCDAs into their workflows.

Feedback to Participants

When onboarding new members, MHC works with individuals within each organization to standardize the data being sent to the HIO and has found that some organizations are more interested in this assistance than others. They have a provider advisory committee that helps determine how data is displayed, and privacy and security, as well as technology advisory committees that are responsible for the security policies. MHC does have a process for reporting potential duplicates or mismatches back to the data source on an ad hoc basis but does not currently offer a service to routinely send this data as part of the source organization's MPI management activities. This is a challenging service to provide since the HIO relies on the data being sent from the source – if the source does not see the person as being the same (having had that person standing in front of them during a registration process) then on what basis does the HIO know better that indeed it is a duplicate? MHC believes this is inherently a registration issue and if the source registration process/system is loose there is little an HIO can do after the fact to clean up the upstream issues.

Rates

MHC was not able to report on estimated false positive or false negative rates.

Cost

MHC has a fraction of an FTE focused on reviewing data quality and potential duplicates.

Comments

Hospitals are extremely busy with Meaningful Use, ICD-10, and other federal requirements, so any data quality regulations would need to be factored into those other priorities. If federal data sets are available, those should be adopted in registration systems, closer to the front end of a medical record's life cycle.

New York eHealth Collaborative (NYeC), New York, NY

Vendor

NYeC utilizes IBM Initiate version 10.0. NYeC has created four separate instances of the IBM Initiate MPI: two RHIOs each have their own instance; three others share a third instance; and there is a fourth instance that sits over top of the first three to allow sharing of data across NYeC.

Data Attributes

NYeC uses first, middle, and last names, full address, phone(s), hashed and unhashed (depending on the RHIO) Social Security number when available, DOB, gender, and insurance number if it is available. While Social Security number is collected by New York hospitals, it is becoming more common for systems to withhold the Social Security number when sharing data. Additionally, if Social Security number is a match the algorithm will add points, but if it does not match, it will not detract points. NYeC has also begun using source-system EMPIs for deterministic matching.



Feedback to Participants

NYeC does not currently send automated merge/unmerge messages to data sources, since they are not ready to accept the automated messages. However NYeC does send ad hoc emails to the RHIO when an issue is found, and the RHIO works with their data sources to correct the record. NYeC is working to roll out IBM Initiate's Inspector tool to provide reports on data quality and duplicates as a service that can be provided at the RHIO level. The RHIO members of NYeC can also offer the service to their members, allowing hospital staff to look at the Inspector tool to see what is causing failed matches.

Rates

NYeC recently attempted to analyze their false positive/negative rate, but did not yet have enough data to confirm the results.

Cost

In addition to the cost of IBM Initiate, NYeC needs one FTE to provide at a minimum 10,000 clerical reviews per month.

OCHIN, Portland, OR

Vendor

OCHIN provides a fully hosted version of both Epic and eClinical Works (eCW). For data sharing between the various Epic instances, Epic Care Everywhere is utilized. The eCW users utilize the eCW HIE capabilities.

Details

OCHIN, while not a traditional HIO in the same vein as the other interviewees, was included in this category to help provide insight to matching rates and specific challenges within multiple instances of Epic EHR systems across a broad swath of FQHCs in 17 states. OHCIN has 105 organizations currently querying between instances of Epic, and also share with the Veterans Administration (VA) and Social Security Administration (SSA) via Healtheway.

Data Attributes

Epic-to-Epic EHR matching at OCHIN uses the following fields: first and last names, DOB, and gender. Addresses are not used, and OCHIN observed that the address field is less stable in the FQHC patient population due to more frequent moves. OCHIN does not use SSN for matching, except with the VA. OCHIN has not used SSN with the SSA transactions, but has found they have a fairly high match rate (more than 90 percent) with the SSA without it.

Match Method

OCHIN uses the standard Epic algorithm and states that customization would be impractical given the number of organizations they support, although there is customization occurring with the VA connection. The Epic algorithm assigns a weight to each data attribute, but then confirms those attributes with deterministic matching. There is a high mismatch rate on the name fields; OCHIN believes this is due FQHC patient's providing multiple names at different points of care.

Feedback to Participants

OCHIN states that the Epic duplication report works well. The OCHIN clinics tend to use different registration processes depending on the payer source for their patients, as modifications are necessary depending if the patient is on Medicaid (MMIS number) or a commercial plan (driver



license number). Data collected for patient registration does not currently consider the implications of HIE. The priority consideration is given to medical billing and matching the payer's demographic data. OCHIN is having the most success on matching when process improvement programs are followed. One such improvement was conducted between Sutter Health, California and the Santa Cruz Health Services Agency, where staff in both organizations identified registration issues and implemented alternate names in the registration of patients, which improved name matching within the Epic algorithms. Other process improvements include the rejection of HL7 messages that do not meet the specified standard for patient demographics; for example, rejecting lab results that do not contain matching patient demographic data. OCHIN is currently receiving CCDs from 8 different systems, each requiring a customized approach to the parsing of data, and they have found that matching lab results is a particular problem, as many labs have not yet implemented Logical Observation Identifiers Names and Codes (LOINC) vocabulary. OCHIN has not had much success creating improvement programs with the hospital emergency departments to validate data, but they have had some recent success with obstetric centers. Overall, much time is spent with trading partners deciding what the common fields will be and ensuring the HL7 messages follow the specified requirements.

Rates

OCHIN can document a 93 percent match rate with the SSA. In all other environments, the match rates vary greatly, from 14 to 90 percent positive match rates. Currently members are not involved in a process to set and monitor goals for matching rates, but doing so might provide a high benefit for care coordination, as providers often give up in searching for a patient's record when matching rates drop below 50 percent.

Costs

OCHIN was not able to provide an estimate for costs.

Comments

OCHIN feels that the FQHC point of view must be considered, as they tend to have less resources available to address patient matching issues. OCHIN also feels that a common identifier would be very helpful.

Utah Health Information Network (UHIN), Salt Lake City, UT

Vendor

UHIN currently uses Optum for both its HIE solution and its MPI. They will migrate to IBM Initiate in 2014.

Data Attributes

UHIN currently uses first and last names, middle initial, DOB, gender, Social Security number, address and phone numbers for matching. Matching protocols may change as they move to IBM Initiate. UHIN currently collects Social Security number from some members who are willing to send it, but they do not return the Social Security number on query responses.

Match Method

With the Optum product, UHIN's match method is deterministic, but they will move to probabilistic matching with IBM Initiate. UHIN also uses LexisNexis and PeopleSmart for manual data validation. Merging records is mostly a manual process and is not very satisfactory, costing UHIN in both



human resources and social capital with members. UHIN will work with their new vendor to develop more automated processes for sending patient record feedback on to members.

Feedback to Participants

UHIN has sent periodic reports to users, but with the new MPI system wants to implement a regular reporting system showing incomplete and invalid submissions, potential overlays and potential merges in the user MPI system. UHIN is working to create a patient portal that will allow patients to manage consent, and will include a patient authentication component.

Governance

UHIN has an MPI Committee and has been working with their member community to share information between data sources for updating records, perhaps using DIRECT.

Rates

UHIN estimates they currently have a 12 to 15 percent duplication rate for records in their MPI database, but cannot quantify this precisely. They are working actively to improve this rate and to develop quantifiable measures to evaluate the health of the MPI.

Costs

The yearly cost of record reviews is currently one and a half FTEs at approximately \$82,000 per year. Hardware and software MPI costs are bundled in the total cost of the HIE system and are difficult to separate. UHIN anticipates that these annual costs will be substantially decreased with the purchase of the new IBM MPI, while accuracy improves as mentioned above.

Comments

UHIN is approaching 4 million patients in their MPI. UHIN's current system allows all of the ADT feeds from hospitals to create new records as they update the MPI. Improvements could be made by requiring a minimum set of fields to be filled out before records can be updated and by developing a testing environment that MDM/MPI vendors could test their products against. UHIN anticipates that creating such a test environment where testing results could be published, would help vendors offer better products and improve customers' experience with MPI data. Standardization of data fields to handle things like hyphenated names would be a positive improvement. UHIN also suggested increasing community involvement through improved training and education for registrars, and looking at licensing or certification requirements.



Appendix E: EHR Vendors

Allscripts

Solution

Allscripts offers Allscripts Community Identity (ACI), a patient matching product that is an integrated identity management solution and integrates both with its own solutions as well as third parties. Allscripts recently acquired dbMotion, which also offers a solution that integrates with other third party EMPI products, as well as ACI.

Algorithm

Allscripts Community Identity utilizes a combination of deterministic and probabilistic algorithms to identify patient matches, allowing customers to achieve the optimal balance between accuracy and flexibility, according to their preferences. These algorithms identify matches in the cases of common errors like misspellings and transpositions.

Data Attributes

The Allscripts default attributes are: first and last names, DOB, patient identifier (such as Medical Record Number) and gender. Social Security number is also an option, but it has become a choice less likely to be used by clients as part of the primary rule set. There are 43 different fields in the standard product, and customers may add more attributes to the primary set if they choose to. Customers also have five miscellaneous fields that can be used to capture any data attribute, which in some instances is used for matching purposes. Allscripts will be implementing further data attribute normalization in the next release of Allscripts Community Identity, as well.

Additional Details

Allscripts supports open industry standards like PIX (v2/v3)/PDQ standards. The ambulatory product also has checks within the EHR that can be setup to notify the user of potential duplicates. Additionally, because data integrity is critical in this effort, Allscripts recommends that customers clean their data prior to going live using the EMPI. Also, Allscripts will work with their customers to improve the data quality, since it is important for accurate matching.

Comments

Allscripts welcomes further standardization of data attributes such as phone numbers and will continue to be focused on IHE standards moving forward. To further demonstrate this commitment, Allscripts has partnered with five IT vendors to form the CommonWell Health Alliance – a nonprofit focused on developing a national secure network and standards that will:

- 1. Unambiguously identify patients.
- 2. Provide a national, secure record locator service. For treatment purposes, providers can know where a patient's records are located.
- 3. Enable peer-to-peer sharing of patient records requested via a targeted (or directed) query.
- 4. Enable patients and consumers to withhold consent / authorization for participation in the network.



Allscripts recognizes the importance of interoperability between solutions from multiple vendors and is therefore committed to leveraging open industry standards where possible. In a rapidly evolving industry where understanding the full longitudinal record of the patient for optimal patient care competes with protecting the privacy of that same patient, there is always room for transparency and education of the risks and benefits of establishing patient identity across the multiple collaborative care settings.

Inpatient/Ambulatory EHR Vendor

Solution

The EHR vendor uses the Qvera Interface Engine (QIE) for external connections and has an internally developed solution for matching within the EHR. QIE does the initial matching when a record is received from an external source.

Algorithm

The algorithm in the EHR is deterministic with a one-for-one character match, including spaces. Since it is deterministic, the vendor has found that it has trouble matching hyphenated and compound names. The internal algorithm cannot be customized by users; all changes are made by the vendor, typically via a core release. The EHR has a sophisticated method of merging records and the QIE product is more adaptable for providing a total solution.

Data Attributes

The EHR uses first and last names, DOB, and gender as a minimum to match on, but can also use other data attributes on top of that such as location of care and a system's patient identifier (MRN). The attribute that helps customers the most is an external ID of another medical trading partner, whether a practice management system, a lab, or a hospital with a different EHR system. The vendor also notes that they are seeing the use of Social Security number decreasing with their customers, and when it is being used, it is most commonly only four digits.

Comments

The vendor would like to see more data attributes used consistently for matching, such as phone, address, aliases, or historical information on names or addresses. They are looking at two proposed solutions for receipt of CCDA documents within the EHR system via Direct. One workflow is for providers to use the built-in messaging tab, which displays on the user's desktop; this will flag some record discrepancies (such as a nickname), and the system user will then need to manually import the document into the master record. The other option is when the record comes through the QIE tool once that is live, then the matching will be automated and the new document will be consumed into the patient's record with an alert provided to the provider that there is new information about the patient in the record. Matching errors and unmerges are handled with a manual process.

NextGen Healthcare

Solution

NextGen currently uses the NextGen Rosetta product for matching, but with their recent acquisition of Mirth, they plan to use Mirth Match for all of their products as an EMPI. NextGen currently has a



lightweight MPI that can store multiple identifiers for patients; it is not a matching algorithm, but a rather a confirmation based on an exact match of identifiers.

Algorithm

NextGen's current MPI product is deterministic and can only deliver an exact match. Mirth Match is a probabilistic algorithm that allows for weighted confirmation and is adjustable. Currently, NextGen customers are able to modify the strength of the match criteria, though it is tricky to do without vendor support. NextGen does not recommend that customers modify the interface without consulting NextGen Support first, as matching problems are not apparent right away.

Data Attributes

NextGen's standard match criteria will look for a common identifier first, if one is not found, it will match against: first and last names, DOB, Social Security number, and phone number. Users cannot send the DOB, Social Security number or phone with hyphens, as the matching method will not be able to match the hyphens. The product has a mother's maiden name field but they are not sure how much it is used.

Comments

Direct poses a problem for patient matching because Social Security number is not included in the XDR/XDM metadata standard. Users can receive a daily or weekly potential duplicate report and can subsequently merge records and record the merge in an audit log. However, once a record is merged it cannot be unmerged. If patient data has been merged erroneously, the record is quarantined and two new records must be created. The two most recent versions of the software will have audit capabilities to help determine which data goes where, but the current software does not provide this capability. Between 10 and 20 percent of NextGen customers use merge capabilities.

RelayHealth

Solution

RelayHealth developed a deterministic match method, which is built into their EHR products.

Data Attributes

RelayHealth uses the core attributes of first and last names, DOB, ZIP code, and gender. They add any numerical ID they can capture, even when it is a partial number such as the last four digits of a Social Security number or a MRN, as they have found this will significantly improve the match.

Customization

RelayHealth's customers can control the minimum score for a given trading partner, but they cannot customize the attributes.

Additional Details

RelayHealth uses XCPD for document sharing, but does not currently use PIX/PDQ. The product has some tolerance for false data, and RelayHealth is working on fine-tuning their scoring methodology. They are also working to create a standardized format for exchange through their work with CommonWell. RelayHealth also uses the NYSIIS algorithm, which was introduced by the New York



State Identification and Intelligence System (NYSIIS) in 1970 as an improvement to the Soundex algorithm. NYSIIS handles some multi-character n-grams and maintains relative vowel positioning, whereas Soundex does not.



Appendix F: MDM/MPI/HIE Vendors

HealthUnity Corporation

Solution

HealthUnity offers a wide range of solutions including MDM and MPI. In addition, the HealthUnity suite of solutions includes HIE, a DIRECT gateway, messaging, analytics, referral management, care coordination, patient-engagement portal and population health management.

Data Attributes

HealthUnity utilizes a core set of data attributes, including: first, middle, and last names, DOB, Social Security number (or last four digits), gender, home and office phone numbers, Insurance policy number, mother's maiden name, driver's license number, race, religion, language, and full address. Additionally, they offer customers the ability to use a wide variety of attributes, and these are not limited in number.

Algorithm

HealthUnity's matching algorithm is a hybrid based on multiple scoring models that in turn use heuristic-based combinations of probabilistic and deterministic algorithms. Probabilistic models use field-specific algorithms to calculate a match score between a pair of fields. The various fields have unique solutions: person name fields require synonym and diminutives handling, standardization, Soundex, and word distance calculations; long fields such as Social Security number susceptible to transposition use NGrams to correct. In total the product offers out-of-thebox eight different scoring models, and potential matches are shown along with their confidence values. Additional models can be configured using a GUI tool.

Additional Details

HealthUnity supports all IHE protocols, including: PIX/PDQ, XCPD, Audit Trail and Node Authentication (ATNA), and Consistent Time (CT). Customers can configure most aspects of the matching algorithm via a user interface. HealthUnity supports integration of the ADT system through an application programming interface (API) call as well as deferred registration when data is received by MPI software in the HL7 format. HealthUnity cannot share specific metrics due to confidentiality agreements with clients, but they have made a significant impact on improving their customer false positive and false negative rates. When matching, the product can automatically identify potential false positives and potential false negatives, allowing the customer to focus on false positives first. In addition, HealthUnity offers a results routing service that takes records with no MRN, determine which EMR it should be sent to, and deliver the result; through this offering they are able to identify 85 percent of orphan results.

Comments

Standardization of data fields could help match rates. HealthUnity suggested that spelling out the month and using numerals for the day would be a way to avoid mixing up those fields. They also suggest filling in Social Security number twice to reduce errors. HealthUnity would also like a hosted MPI solution that vendors can test against to see how well a given MPI is performing.



IBM Initiate

Solution

IBM Initiate partners with EHR, portal, imaging, and data exchange vendors and can integrate with them via API calls, HL7 messages, web services, and other methodologies. The software solution can also be embedded in a vendor's product. The solution includes 10 standard analytics reports within Workbench to help users tune the system, with the capability to define other reports.

Algorithm

The IBM Initiate algorithm is both deterministic and probabilistic, which is more complex but has a greater potential for improving matching. IBM Initiate believes that using both methods is the only way to get the accuracy organizations require, as well as meet diverse business requirements. The product assigns a score to how closely the records match on an attribute by attribute basis, and ends up with a cumulative probability or likelihood score. They also utilize a number of matching techniques to identify common data errors such as transpositions, misspellings, phonetics, and data entered in wrong fields (such as flipping the first and last names). False positive and false negative rates depend on the threshold set by the customer, and they are usually conservative to avoid false positives. The IBM Initiate product is highly customizable. Customers can make rule changes and adjust the weighting of attributes. The solution optimizes around the particular types of errors that each implementing organization tends to create during registration.

Data Attributes

IBM Initiate has a standard data model that includes about 15 data attributes that they highly recommend customers use. The most commonly used attributes include first, middle, and last names, address, phone (cell is much more valuable now), and Social Security number (value has decreased as sharing of Social Security number has decreased). Customers can add additional attributes to meet their business requirements and goals. There is no limit on the number of attributes IBM Initiate can match against. IBM Initiate has not had any requests to use mother's maiden name for matching, but believes it could have an adverse effect on family matches if not appropriately applied. IBM Initiate indicated that historical data (such as the address or name) has a large impact on accurate matching. The product does standardize address to a format similar to USPS, although not exactly the same. Call outs to third parties can readily be done to validate data, such as address.

Additional Details

PIX/PDQ is the most common mechanism for IBM Initiate to actively integrate with EHRs and HIOs. Web services are also commonly used for EHR integrations. IBM Initiate provides the Inspector tool, which allows users to view potential duplicates, linkages and overlays that fall below the autolink or automerge thresholds. Users can then merge or unmerge records as appropriate. The system can "learn" by flagging records that have been subject to a matching error that is validated by human review, meaning that they are not scored later as a potential match.

Comments

IBM Initiate highly recommends the use of confidence scores and emphasizes that it is equally important to ensure that matching algorithms will indeed assign scores to all records that deserve



evaluation. IBM Initiate does not believe that a national unique identifier in and of itself will resolve the identification challenge. New data attributes may be useful, but if considering new attributes, IBM Initiate recommends testing and validation prior to requiring their use. In addition, biometrics could be a great addition to matching on demographics, particularly for verifying patients' identities at the point of registration.

Informatica Corporation

Solution

Informatica provides the Healthcare Data Management and Identity Resolution products.

Algorithm

The Informatica algorithm is a hybrid model that uses four different algorithm techniques: Hamming Distance, Jaro-Winkler Distance, Edit Distance, and Bigram frequency. The algorithms can be tuned to create "fuzzy matching" keys, based on the sensitivity of the settings. The customer determines the data attributes used for matching and the influence of each attribute by adjusting the weights of the algorithm. Informatica offers training at various levels and maintains an active online user forum to allow their customers to develop the skills to create the customization that best meets their business needs.

Data Attributes

Informatica's MDM solution is customizable, based on the needs of the customer. The most common attributes used for matching are first and last names, address, and Social Security number and/or other numerical ID (medical record number or driver's license number). Customers can choose limitless additional data attributes.

Additional Details

Informatica uses PIX/PDQ, and they described plans to expand on the IHE profiles. They make available a manual unmerge service for their health care customers, but they do not have any method to capture metrics on false positive and false negative rates for their Identity Resolution products. Informatica's solution also has role-based linkages for the various business needs of health care organizations in addition to patient matching for clinical care, eligibility, and payment.

Comments

Informatica is concerned that standardization might be a double-edged sword, potentially improving products at the bottom end, but limiting innovations. They expressed support for the idea of an open source algorithm.

Mirth Corporation

Solution

Mirth Match is the MPI product offered by Mirth.



Algorithm

Mirth Match maintains two standard algorithms, a plugin interface to create custom algorithms, and scripting interface to allow installation-specific deterministic or probabilistic logic to be layered on top of any base algorithm. The probabilistic algorithm is a standard Fellegi-Sunter algorithm. The custom algorithm could be either or hybrid. All matching algorithms in Mirth Match compare pairs of entities, and calculate a score that summarizes how similar the two are to each other (a similarity score). When performing a plugin search operation (as opposed to a register operation), one of the two "entities" will actually be a set of search criteria. The deterministic algorithm uses a hard-coded set of proprietary rules and fixed "magic numbers" to calculate a similarity score based on the values of those traits that are actually present in both entities. The probabilistic algorithm computes a partial similarity score for each trait for which a value is actually present in both entities, using weighting factors (matching agreement rate and mismatching agreement rate) that are calculated separately for each trait from the frequency of the values that appear in the domain, using tools built into Mirth Match (the Estimated U calculator and the Expectation Maximization calculator). The weighting factors can be recalculated at any time using the same tools or be manually tweaked.

Data Attributes

The data attributes used for matching are customizable, but the most commonly used traits are first, middle, and last names, DOB, gender, Social Security number, ZIP code, and phone number. The fields include the entire value of each trait. There is no limit on the number of attributes that can be used for matching, but the total fields are limited to 4,000 characters.

Additional Details

Mirth Match does not rely on IHE profiles but does support them, and they are utilized in other Mirth products, such as Mirth Connect.

Comments

Mirth believes that required field completion of first, middle, and last names, Social Security number or the last four digits would reduce false positives, and they feel that mother's maiden name could help improve matching (while there is an HL7 field, it is rarely used). A solid HL7 validator tool, similar to the CCDA validator tool, could help improve the message structures. A potential policy lever could be working with Medicare and Medicaid to reject reimbursement requests that lack valid information, which would make it in the interest of providers to have good data quality. Mirth would be willing to participate in a validation process that certifies the performance level of their algorithm.

NextGate Solutions, Inc. (with Orion Health)

Solution

NextGate's MatchMetrix is the embedded MPI for the Orion Health HIE solution. NextGate also offers the MatchMetrix solution as a stand-alone EMPI, and uses the same matching technology for its Provider Index.



Algorithm

The NextGate algorithm is a hybrid model, employing both a probabilistic engine for the core matching and a deterministic rules engine to filter out false positives. The customizable engine is tuned to the specific patient population being matched, using statistics gathered during the analysis phase of the implementation. The algorithm also offers advantages beyond its high level of accuracy. It's inclusive of a flexible independent algorithm optimization specific to each data source and has the ability to easily extend and fine tune post go-live.

Data Attributes

Some of the common attributes of the NextGate MPI are first, middle, and last names, historical names, current and historical addresses, current and historical phone numbers (home, work, mobile, etc.), date of birth, Social Security number (or last four digits of Social Security number), gender, and a death flag, but it is important to note that there is no limit to the number or types of fields, and that a comprehensive data analysis is performed for each client to identify the initial attributes set. Examples of additional attributes might be driver's license number, insurance identification numbers, family links such as spouse or emergency contact information, employer, and provider of care attributes, etc.

Additional Details

Interoperability with other systems is provided using the IHE PIX/PDQ v2 and v3 messaging standards. This enables a consistent HL7 based messaging format to allow both insert and updates to the EMPI, as well as queries. To maximize the utility of the data, the engine employs numerous standardization techniques, such as phonetically encoding name attributes using the Metaphone3 algorithm, normalizing identifiers and phone numbers to compensate for character transpositions, and checking equivalence and exclusion tables to find synonyms (e.g. nicknames) and values that should be ignored. The engine can call additional external data validation services, such as address and Social Security number verification, though most of NextGate's customers do not use this capability due to the cost of using these external services.

Comments

NextGate cautioned there are many variations in how people enter data, particularly with dates and missing fields. They recommend that matching algorithms be flexible enough to account for regional demographic differences. NextGate works with their customers to review data before loading into an MPI, to understand the regional qualities within their patient population and adjust the algorithm appropriately. Additionally, NextGate recommends to their customers that they work to improve the quality of data in their MPI through education and training processes, including a systematic feedback loop for registration staff, so they understand how their work impacts other parts of the system. Orion Health also suggests that regional HIOs could offer duplicate identification as a service, and might consider putting data quality review requirements in their participation agreements. They point to a new service offered in the United Kingdom, called the Medical Interoperability Gateway (MIG) as a model for this type of feedback loop from HIOs.



Oracle Healthcare

Solution

Oracle Healthcare Master Person Index (OHMPI) is Oracle's product offering an advanced matching solution for person and non-person entities. Oracle works with many EHR vendors and sells both a standalone and integrated solution.

Algorithm

Oracle's match algorithm is probabilistic with some rules-based deterministic matching. Oracle Match Engine is based on Fellegi-Sunter methodology that uses conditional probabilities to estimate the matching weights. The Match Engine is very flexible and allows using custom components to define a deterministic rule-based approach. In this sense, the match engine is flexible enough to use probabilistic and/or deterministic. The product offers a long list of predefined techniques and algorithms and allows users to define additional algorithms using the match framework plug-in. The product offers phonetic encoder matching including Soundex, NYSIIS, and Metaphone, and uses advanced edit-distance algorithms. It also offers dates, numeric and alphanumeric algorithms, and nickname substitution datasets. The product includes advanced power match options such as system-dependent matching, frequency-based matching, conditional matching, and field swapping.

Data Attributes

The core set of data attributes used for matching include: first and last names, DOB, address (could be partial, line one, or city, state, ZIP), Social Security number, and gender. A comprehensive set of templates covering person, patient, provider, and organization are available. In addition, choice of data attributes recorded and attributes used for matching can be customer defined. The product includes a Standardization Engine to parse, normalize, and phoneticize fields such as names and addresses.

Rates

The match rates depend on the data attributes chosen for matching and the product typically achieves match percentage rates in the high nineties range.

Additional Details

OHMPI supports a number of IHE profiles, including: IHE PIX/PDQ HL7 v2 and v3, PAM, XPID, and ATNA. The product carries out extensive transaction logging for greater levels of governance and provides reporting on duplicates, defaults, and high error rates. The product can automatically merge records based on match weight thresholds. Data stewards use a non-programmatic browser-based console to manage all records, including merging and unmerging potential duplicates, viewing reports, and managing/governing the implementation.

A single best record representing the golden record of the person/entity from across multiple source systems is constantly updated using a configurable survivorship engine. The product is highly customizable by users, including introducing custom matching rules and adding data attributes, including biometric data. OHMPI can be implemented in multiple ways including registry-type or transactional and supports real-time matching and batch matching. Systems that



seek updates from OHMPI can subscribe to event propagation directly from OHMPI or use an integration layer. The product is extremely scalable and is used from small departmental projects to very large country-wide implementations.

Comments

Oracle has found that one problem is the lack of a feedback loop in many health care organizations, meaning they do not communicate to registrars the data entry errors being made. Best practices of data quality are not consistent, but they would help improve match rates.

QuadraMed

Solution

QuadraMed provides an EMPI product called SmartIX (which runs on InterSystems Ensemble) with SmartMerge for duplicate resolution and SmartID for patient search & selection.

Algorithm

The product uses a probabilistic algorithm, based on Fellegi-Sunter, which allows for an unlimited history of values and multiple values for each attribute. The probabilistic algorithm is enhanced with auto-link rules, and when a record does not meet the auto-link threshold, secondary rules that are deterministic, are applied. The algorithm acts at the field level as well as the value level and is a statistically self-tuning algorithm, meaning users do not modify the algorithm. QuadraMed is different from other vendors in this regard and feels that once a manual judgment process is introduced, statistical validity is undermined. Ultimately customers cannot opt out of the algorithm but can adjust the match weight thresholds for reporting purposes.

Data Attributes

The most common data attributes used for matching include: first, middle, and last names, gender, DOB, person identifiers (i.e. Social Security number, Driver License #, etc.), full address, telephone number, race, marital status, Mother's maiden name, and birthplace. Customers can add additional data attributes, but there are attributes that QuadraMed deems minimally necessary in order for the algorithm to produce best matches.

Rates

The rate of false positive and false negative detection vary based on the quality of the source data. There is no single rate that can be quoted since each situation and quality of source data is different.

Additional Details

QuadraMed supports the HIE PIX/PDQ profile and expects to support XCPD in 2014. The product provides visual confidence levels for matches to users. Confidence levels of definite and potential matches are indicated using a stack ranked Green – Yellow – Red hierarchy (the top, green highlighted entry has the highest confidence while the bottom, red highlighted entry has the lowest confidence of the potential match candidates). Both the sequence of the list (top to bottom) and the color indicate a stack ranking. Along with this a confidence value is also associated with the match – the higher the number, the stronger the match. In addition, SmartMerge has the ability to both merge and unmerging within the EMPI



is only half the challenge. The ability to support this same process in the data source is also necessary. If the unmerging/unlinking does not occur simultaneously within the source system, then the source system will continue to convey inaccurate details. QuadraMed also provides a merge automation solution, which assists in this simultaneous merge process.

Comments

QuadraMed has found that data quality is often a huge issue for their clients. It is possible for QuadraMed to identify the source of the ADT event that resulted in duplication, and customers that take advantage of that do better with matching than those who do not. QuadraMed does not believe that a universal identifier would make a significant difference in accurate patient matching, particularly since it would be an additional identifier that can be miss-keyed and possibly misapplied, specifically in the case of medical identity theft. QuadraMed has noted that one of the challenges clients face is the multiple HL7 versions and variations that exist between disparate source systems. Additionally, missing data or inaccurate data can be challenging for many healthcare organizations. Sometimes data quality issues are intentional when patients choose not to provide data, but more often they are the result of human error or inadequate systems & tools. This issue will not be resolved until the bigger issues of privacy, data standards, and data governance are addressed. Joint Commission or Meaningful Use Standards though may encourage hospitals and providers to work toward better data quality. The most effective incentives for data quality improvements will come from payers and from patients themselves as a result of engagement in achieving high quality, cost effective healthcare.



Appendix G: Associations and Other Organizations

Provider Organizations

American Academy of Family Physicians (AAFP)

As a membership organization representing providers in mostly smaller practices, the AAFP has not seen the issues of patient identification and electronic matching rise to a level of significant concern for its members. Health IT experts at the AAFP have found that even in newly formed accountable care organizations (ACOs), the participating organizations usually have existing relationships and therefore are not driven to solve patient matching issues. They also noted that the shrinking Meaningful Use incentive payments in coming years may not be effective in driving smaller practices to make major workflow changes, particularly if the requirements are too onerous. The AAFP supports a VUHID, and suggests a second level of VUHID for specially-protected or sensitive health conditions, such as human immunodeficiency virus (HIV) status.

About the American Academy of Family Physicians

Founded in 1947, the AAFP represents 110,600 physicians and medical students nationwide. It is the only medical society devoted solely to primary care. Approximately one in four of all office visits are made to family physicians. That is nearly 214 million office visits each year — nearly 74 million more than the next largest medical specialty. Today, family physicians provide more care for America's underserved and rural populations than any other medical specialty. Family medicine's cornerstone is an ongoing, personal patient-physician relationship focused on integrated care. To learn more about the specialty of family medicine, the AAFP's positions on issues and clinical care, and for downloadable multi-media highlighting family medicine, visit <u>www.aafp.org/media</u>. For information about health care, health conditions, and wellness, please visit the AAFP's award-winning consumer website, <u>www.familydoctor.org</u>.

American College of Physicians (ACP)

ACP counts more than 137,000 members, a large proportion of whom are small practice internists (many in practices with less than 10 providers). Providers in small practices are concerned about the transitions of care measure that takes effect with Stage 2, but most are focused more on sending out the required number of CCDAs and have not started to focus on receiving CCDAs and using them as part of their daily workflow. Providers who are beginning Stage 2 in 2014 are more concerned about having other providers to send electronic CCDAs to than what they will do with CCDAs when they receive them; however, once the first hurdle has been accomplished, providers will have an opportunity build new workflows that incorporate the CCDA into the overall care of patients. Typical practices have an understanding of the importance of up-to-date, accurate patient demographics, but have not yet seen the value that will come when receiving documents from new sources that require matching. ACP is likely to include this scenario in new educational materials and believes it will be easier for practices if they could be provided with a list of the data attributes that are used for matching. Many practices have developed their own idiosyncratic matching strategies. Anecdotes suggest that there is concern that a national strategy may be at odds with their local strategy and cause confusion and possible error during conversion. In addition, ACP


believes that use of a voluntary universal unique healthcare identifier could provide privacy benefits and that its potential use should be studied.

American Hospital Association (AHA)

The AHA has nearly 5,000 health systems, hospitals, and networks, and 43,000 individual members. One of the largest concerns for AHA's members is the reliance on third party software for MPI, software that is typically felt to be a "black box," with users lacking insight concerning the methodology as well as the ability to revise the output as necessary. Additionally, AHA's members recognize the need to hire increasing numbers of employees to handle patient matching, including regularly reviewing MPI data, removing duplicates and overlays, and generally cleaning the data. Data quality is a contributing issue to inaccurate patient matching, but there are multiple challenges to improving data quality. There are numerous departments within a hospital that contribute data, and there is variation in the management of patient matching by hospitals. As a result, this variability challenges the notion of a one size fits all solution. In addition, registrars' primary goal is entering patient data quickly into the system in support of the initiation of patient care. Staff turnover rate in this area is high, so staff training will always be a cost. Patients are also a contributing factor. There are cultural sensitivities to requesting some demographic data, and rather than risk offending patients, hospitals may choose not to ask some questions. If they do request the data, patients may not be willing to provide the data, particularly if the data requested seems intrusive. Finally, AHA members are concerned about liability issues around sharing patient data electronically, and the potential for basing medical decisions on an inaccurately matched patient record. The provider confidence in the validity of the data, necessary to support subsequent reliance on the data, needs to be addressed. The AHA supports a national unique identifier, and believes that if such an identifier is limited to use for treatment purposes only, patients will be more inclined to accept it.

American Medical Association (AMA)

The AMA is the largest professional organization representing physicians and medical students. Patient matching is not an issue that most practicing physicians are currently focused on since it tends to be a fairly technical issue, and the majority of doctors have yet to start exchanging patient data electronically (roughly 10 percent of physicians participate in an HIE). However, it is an issue that is gathering increasing interest and scrutiny. AMA's members are focused on a myriad of competing regulatory requirements, including Meaningful Use, ICD-10, and multiple quality reporting mandates but are struggling to meet them as deadlines collide. In terms of exchanging health information, physicians lament the lack of interoperability among disparate systems and find meeting Meaningful Use to be a significant challenge. Many are trying to determine how to meet the transitions of care measure, if they continue with the Meaningful Use program (many physicians, however, are considering simply dropping out or choosing to take the penalty rather than attempt to comply with such challenging or inapplicable requirements). While accurately collecting demographic data is not currently at the forefront of provider's thinking, it could enable the transitions of care process to work more smoothly once interoperability among EHRs is established. Until that time, physicians find collection of some patient data to be redundant (e.g., some specialists / primary care physicians collect information that other specialists do not readily need). Physicians are also concerned about the liability issues associated with the electronic exchange of information (e.g., if they received information sent to them by another provider but did



not act on it) and the volume of information they could receive. In the future, physicians may require help navigating these issues. While the AMA's policy previously opposed a national unique identifier, newer policy supports the need for more research on this matter and supports the need for a study on the value of such an identifier.

Medical Group Management Association (MGMA)

MGMA's membership consists of professional administrators and medical group practice leaders. There is recognition that many of the data quality issues arise during the registration process, and MGMA believes that automating the registration process and allowing patients to enter their demographic data through websites, kiosks, tablets, etc. could alleviate some of these issues. Machine readable identification cards would eliminate the need for staff to retype information from a health insurance identification card. In addition, machine-readable cards, especially those with "smart chip" technology, could be a tremendous asset for accurate patient identification; more efficient patient matching; capture and storage of critical health, demographic, and insurance information; and automating the patient insurance eligibility verification process. MGMA also feels that at some point the foundational pieces, such as data attributes, have to be standardized in order to move forward effectively and efficiently with electronic exchange patient data. MGMA has a great deal of experience with educational campaigns and recommends that the data accuracy message be weaved into a broader campaign on the importance of exchanging patient data, both for improved care and more efficient billing processes.

Health IT and Health Information Management Organizations

American Health Information Management Association (AHIMA)

AHIMA represents more than 72,000 HIM and health informatics professionals responsible for managing and maintaining the quality of health records. AHIMA and its members have long standing experience linking and matching patients to their records and to their data. Furthermore, AHIMA has recognized that high quality data are need for multiple purposes such as care delivery, performance and outcomes measures, population and public health, research, and policymaking. AHIMA strongly believes that additional attention, such as improved quality controls, standardization, and appropriate validation at data collection points are needed to improve the accuracy of patient matching.

AHIMA highlighted the increasing need for quality data capture at the very beginning of the patient registration process and continuing along the care continuum. AHIMA plans to adapt its existing training and education resources for the patient registration/admission processes. As health care delivery systems and providers consolidate, merge, and are acquired and with the continued implementation of HIOs, accurate and reliable patient matching is even more vital. However, AHIMA emphasized that developing an algorithm or implementing a technology solution is insufficient to solve the problem. AHIMA notes that people, processes and practices, including workflow, are absolutely critical to ensure that data collected and exchanged is accurate. AHIMA is excited to work with ONC and other stakeholders on quality data capture, beginning at patient registration, and repurposing existing AHIMA materials and education and training for diverse audiences. AHIMA is supportive of ONC's efforts to address matching patients and their data and looks forward to collaborating on this industry wide challenge.



College of Healthcare Information Management Executives (CHIME)

More than 1,400 CIOs and senior IT leaders participate in CHIME. CHIME's members identified patient matching as a key issue in 2013 and launched a patient matching workgroup to evaluate potential solutions. CHIME members support standardization of the data attributes used for matching patients, but not necessarily a standard algorithm. Members would however prefer more insight into how current third party MPI products work, with many feeling like they are a "black box" they have little control over. CHIME supports a national unique identifier.

Electronic Health Record Association (EHR Association)

Established in 2004, as a partner of the Healthcare Information and Management Systems Society (HIMSS), the EHR Association is comprised of more than 40 companies that supply the vast majority of operational EHRs to physicians' practices and hospitals across the United States. The EHR Association offered strong interest and support for the ONC Patient Identification and Matching Initiative, welcoming the opportunity to open the door for a more robust conversation about improving matching rates. They do not currently have an official association-approved position on patient matching or a unique patient identifier, but have had many meetings and internal discussions on ways to improve matching rates. They recommend looking at the HL7 implementation guidance and companion guides for opportunities to provide more specific direction that would move vendors towards standardization without increasing the number of requirements or constraining innovation. The association members participating in the conversation also offered suggestions related to increasing interoperability of EHRs, such as using mature, tested standards where possible, rather than prescribing feature functions in certification requirements. They indicated that, as part of patient matching, availability of some form of a unique identifier (e.g., part/whole Social Security number, driver license number, etc.) substantially improves on the matching accuracy. Whether that requires the introduction of a new unique patient identifier and/or improving existing ones that are not as unique as intended or timely in their issuance is not as critical. They also ask for a much stronger emphasis put on developing best practice recommendations for improving data quality, citing concerns for the amount of human intervention that will be required to resolve potential mismatches as HIE becomes much more widespread.

Health IT Now Coalition

Health IT Now is a broad based coalition of patient groups, provider organizations, employers and payers that supports incentives to deploy heath information technology to improve quality, outcomes, and patient safety and to lower costs. As a coalition, Health IT Now suggests that ONC bring together stakeholders to create a model data-sharing agreement that would incorporate the standardization of data attributes in the demographic fields of EHRs. Health IT Now also feels that ONC could support the development of certification standards for a core set of required demographic fields, where the technology will not validate a record until the appropriate information has been entered into required fields. Finally, Health IT Now suggested looking at the Patient Safety and Quality Improvement Act of 2005, where safe harbors protecting privilege and confidentiality are in place to allow for more honest and open study of issues affecting patient safety. To fully integrate EHRs into clinical workflows, providers must fundamentally trust the completeness and integrity of the data stored in patient records. Health IT Now believes that this type of safe harbor protection for Patient Safety Organizations (PSOs) might also be extended to



health IT vendors, who could benefit from "non-punitive, post market" study on the issues of patient identification and matching. Creating a culture of safety that rewards providers and vendors alike for identifying and addressing problems quickly and transparently will be central to establishing a trusted learning health system.

Integrating the Healthcare Enterprise (IHE)

IHE USA is a not for profit organization founded in 2010 that operates as a national deployment committee of IHE International®. IHE USA serves as a voice representing U.S. health IT interests in international health IT efforts for fostering the adoption of a consistent set of information standards to enable interoperability of health IT systems. IHE International hosts technical committees which develop the Domain-specific (e.g. Radiology, IT Infrastructure, etc.) IHE Technical Frameworks. This consensus frameworks provide detailed specifications called integration profiles for implementing information exchange standards such as HL7.

For example, the concept of the PIX/PDQ integration profiles is that the PIX would include a unique ID for the patient, but if data quality is compromised at the onset, this field would be useless. The HL7 2.x messages carry links to the primary account number and the secondary, tertiary account numbers, so an alternative idea would be to use the last field of the IHE PIX profile (260 characters available) to insert a voluntary patient identifier in that field with the other links. However, this method would require broad adoption of the voluntary patient ID. IHE has launched a certification program for vendors for systems using the PIX and PDQ profiles, providing an opportunity to test real-world implementations. IHE also has concerns about data quality and the lack of methods to measure data quality, particularly since the effectiveness of every algorithm is limited by poor data quality. They also believe that some matching issues are created by character limitations within some EHR systems, when a name or address exceeds the number of characters allowed in those fields, leading to inconsistent data collection across systems.

In the short term IHE is concerned, as were a number of other parties at the ONC Patient Matching Meeting, with the lack of inclusion of some key data attributes on slide 18 of the presentation from December 16, 2013. In general this is important because statistically the more, good explanatory attributes to choose from the better the hit rate. The most important attribute group discussed in the meeting was "Prior Name" which often equates to "maiden name." This impacts tens of millions of people who have records in one or more prior names – this has been a serious issue with academic graduate school applicant credentials for many years. In addition to names changing due to marriage there are many changes tied to cultural and religious related changes.

Another area is the lack of a middle name or the facility to deal with multi-part names, often culturally bound. For example, in some cultures a parent's first name can become the child's last or given name. Additional populations need to take account of other potential attributes. For example, prisoners also number in the millions where many prisoners have an alias or several. They have the legal right to use these aliases which can vary by prison facility. Lastly, on attributes, IHE believes that such areas as medical devices and complementary and non-medical care such as physical therapy where different standards may apply should be looked at. In addition, medical devices are governed heavily by Institute of Electrical and Electronics Engineers (IEEE) standards; these need to be harmonized. The whole area of data attributes needs more in-depth analysis and research in



IHE's view to optimize patient matching effectiveness. Of course there is a limit in terms of cost/benefit and drawing the diminishing returns line. In general it looks like there is work to be done to get to that point.

National Association for Healthcare Access Management (NAHAM)

NAHAM is a nonprofit association with the goals of establishing best practices and subject matter expertise; providing networking, education, and certification opportunities; and enabling its members to influence and promote high quality delivery of patient access services (including scheduling, call centers, registration, admissions, patient finance, benefits coordination, and other front-end revenue cycle related services). NAHAM believes that positive patient identification is an essential attribute to patient safety and the delivery of healthcare services to all. Positive patient identification is the first critical step in providing patient care. Incorrect patient identification through the registration process increases the potential for patient harm. Improved patient identification standards, processes, and technology ensure safe and appropriate patient care and can eliminate duplicate medical records and fraudulent billing.

NAHAM supports continuing efforts to create an environment of positive patient identity and believes that the standardization of patient identification protocols and technologies are important means to this goal. NAHAM is investigating appropriate third factors to enhance positive patient identification. NAHAM supports the development of standards for data attributes in electronic systems, whether clinical or administrative, and enhanced common capabilities for all healthcare data systems to input standardized data. NAHAM members have seen errors on insurance cards or people using two different names for their personal and professional lives (for example, some clergy members, or women using maiden and married names interchangeably) produce difficulties in accurate identification and matching of health records and the creation of duplicate records that may not provide the full medical history. Increasing the chance of error are the multiple data systems and their individual users that feed into the EHR (patient registration, patient scheduling, and clinical settings) and the several ways that patient data must be collected (for example, in person, over the phone, and in the emergency room).

NAHAM believes that education and training are important parts of the solution for positive patient identity. NAHAM offers two levels of certification for patient access professionals and believes that certification can become an effective means for providing training and continuing education.

Education and training are also important to ensure personnel at all levels understand the important roles patient data input and patient identification protocols serve in enhancing patient safety. NAHAM supports Stage 3 Meaningful Use requirements to improve patient matching and supports a comprehensive approach that includes the standardization of patient identification attributes, the development of standards for EHR technology solutions, and the development of best practices and protocols for data input. This would include regular feedback from supervisors and audits for quality control.

Workgroup for Electronic Data Interchange (WEDI)

WEDI is a coalition comprised of a broad cross-section of the healthcare industry, including hospitals, providers, health plans, laboratories, pharmacies, clearinghouses, vendors, government



regulators, and other industry stakeholders. WEDI strongly advocates that a national patient identifier is the best solution forward. A major issue in patient matching today is the variations of implementation, even within one institution. While WEDI recommends standardization of the patient identification process, it has not adopted a position on standardizing the attribute fields of EHRs or other health IT systems. WEDI's research indicates that between 10 to 15 percent of all health insurance denials are due to incorrect identification numbers and believes that new technologies, such as Quick Response Codes (QR Codes) and biometric images using mobile technology, offer new avenues in support of better patient matching and identification.

Consumer Groups

Center for Democracy and Technology (CDT)

CDT is a nonprofit, public interest organization developing and promoting civil liberties in the digital age. CDT represents the rights of patients to privacy of their electronic health information. CDT staff have worked on the ONC Federal Advisory Committees (FACAs), including the Health IT Policy Committee. CDT currently chairs the Committee's Privacy and Security Tiger Team. CDT believes that data quality is a major component in the patient matching challenge and that any effort to improve matching must address data quality. Adding additional data attributes to matching processes or even the use of a common unique identifier is unlikely on its own to significantly improve matching (to the desired accuracy levels), if the quality of data is poor. CDT agrees with pursuing the development of standards for seven to 10 data attributes commonly used in matching across organizations. While CDT supports the study of additional non-traditional data attributes, they recommend avoiding SSN or the last four of SSN pending further evaluation of whether the use of SSN or components of SSN leaves patients more vulnerable to identity theft. Additionally, exploration of additional data attributes should include patient selected identifiers. To ensure that organizations have the ability to accept and use patient chosen identifiers, the choice may need to be limited to a particular set of acceptable data attributes. CDT also recommends that ONC monitor the progress of private sector efforts to enable individuals to create trusted cyberspace identities and consider whether this effort has the potential to help improve accuracy in patient matching. CDT also agrees that patients should be empowered to play a role in improving data quality.

National Partnership for Women and Families

The National Partnership has been actively involved in advancing private and secure health IT in ways that measurably improve the lives of women and their families, and is leading the Consumer Partnership for eHealth (CPeH), a coalition of more than 50 consumer, patient, and labor organizations with a combined membership representing more than 127 million Americans. The National Partnership pointed out that patient matching must take into account the diverse characteristics and attitudes among patient populations and be designed accordingly. For example, address might work well for many, but does not work well for homeless individuals. No single attribute will work equally well for all patient populations and regions, and the task at hand should be to identify a combination of attributes that collectively works best across the diversity of patient populations. The National Partnership feels strongly that any development of standards for patient data attributes be done in ways that acknowledge the wide ethnic and cultural differences of patients, to the extent that is possible. For example, for the 60 million people who speak a language



other than English at home, providing electronic access in languages other than English could do much to reduce errors in their data attributes and in matching.

The National Partnership has found that some patients can be reluctant to disclose certain information and points to strong research being conducted by the Fenway Institute as a positive example of how using focus groups and survey information can help with understanding better whether and when patients might disclose certain information to be used for matching, and how developing educational tools for both patients and providers on the value of collecting certain types of information can assist both groups in doing so. Involving patients in the review of their record and explaining why sharing up-to-date, accurate information helps the quality of care are two examples of ways that providers could improve the quality of the data in a patient's record. Additionally, using multiple-factor authentication in patient portals, as they become more commonly used, can be another verification of patients' identity. The National Partnership also recommends concurrent educational efforts that emphasize the importance of accurate patient identification and matching for improving patient safety.

Patient Privacy Rights (PPR)

PPR is a nonprofit organization founded in 2004, whose mission is to ensure that Americans control all access to their health records. PPR believes that further HIE activity needs to wait until true patient control over their data is achieved. They are concerned about health care organizations exchanging patient data without the knowledge and control of the patients and would like to see a system wherein the patient is primarily in control of sharing her medical records. PPR supports the NSTIC process, which would allow patients to create their own electronic identity and choose when and with whom their records and identity are shared. PPR encourages patients not to allow use of their data in HIE until these issues are resolved. PPR supports the use of patient portals, the Blue Button initiative, and other initiatives to give patients direct access and control over their medical information. In a written statement provided by PPR, they offered the opinion that the ONC Patient Identification and Matching Initiative should have been more focused on patient engagement rather than methods for health care organizations to share data without patient involvement. Instead, patients should verify their own identities at each place PHI is disclosed and match the records themselves. The organization's written statement said, "Patient matching' is a method of involuntary, hidden surveillance, much like the NSA's (National Security Agency's) surveillance of phone records and metadata. It enables 1000s of hidden third parties to collect and aggregate our personal health data from many places without our knowledge or consent." Additionally, PPR states that any patient matching system that is unable to notify the patient by simple text message or email each time their ID is matched, is insecure and error-prone by today's standards and should not be deployed.

Society for Participatory Medicine (S4PM)

The Society for Participatory Medicine is a nonprofit organization devoted to promoting the concept of participatory medicine, with a collaborative relationship between the patient and provider, and patient access to their own electronic health records. The S4PM's research shows that patients are eager to interact with their providers through portals, smart phone monitoring of



chronic diseases, and other technology-enabled devices, but that there is a very small percentage of patients who are now empowered with their own health data. The S4PM strongly supports involving patients in validating their demographic data and recommends looking toward technical solutions to assist with that, such as the use of "in case of emergency" apps on smart phones that can provide medical professionals with a patient's identity and emergency contact information. They also recommend that patient disease support organizations be involved in any type of national education campaign on accurate patient identification and matching that may result from this initiative.

Government

Department of Defense (DOD) working with Sysnet International

The DOD has partnered with Sysnet International to evaluate the feasibility of using OpenEMPI within the Nationwide Health Information Exchange and to develop a proposed architecture for a Universal Healthcare Language Service, as outlined in the President's Council of Advisors on Science and Technology (PCAST) 2011 report. Sysnet International led the architecture and development of OpenEMPI, which is an extensible patient identity management platform that has been deployed all around the world, with a nationwide deployment in Rwanda, and with additional countries evaluating a potential implementation. Sysnet International has seen that effective data attributes differ by geographic location. As such, having more data attributes will enable better matching. Sysnet International has found that at least seven to eight attributes are needed to address geographic variability. The project with the DOD did not develop a standard list of data attributes.

Social Security Administration (SSA)

SSA participates in health information exchange using the eHealth Exchange, for the purpose receiving clinical records for determining disability eligibility. The majority of data attributes SSA uses for matching, are based on the data received on the disability claim, and include: first, middle, and last names, suffix, aliases, address (what they report as their contact address at that point in their life when they apply for disability), phone number, Social Security number (full), DOB, and gender. SSA has seen good response rates where a responding organization has used the name, Social Security number, DOB, and gender to identify patients. SSA recommends that ONC explore additional data attributes beyond the basics of the patient identification segment in HL7 to see if there could be other attributes that improve the patient patch. However, any requirement for additional data attributes must first be analyzed, both for their potential impact on matching and for the workflow changes and subsequent burden those changes may cause. SSA suggested that a set of best practices be developed to assist in improving identifying patients.

Other Organizations

Equifax

Equifax is a reliable third party source of consumer demographic information and maintains more than 310 million U.S. consumer records. False negative patient matches are caused by missing or incorrect data attributes, most commonly address and name changes. Matching algorithms rely



upon the accuracy of the patient data attributes that are only supplied by health systems, which typically have conflicting information about the same patient at different facilities. False positive patient matches are caused by patient identity errors. Equifax's current and historical address and name change data can be incorporated as an additional source into a matching algorithm to increase match rates by reducing the number of false negative matches. In addition, Equifax patient identity-proofing and fraud alerts at the point of registration can help to prevent false positive patient matches. Even if a healthcare organization decides to deploy biometrics or smart cards, the patient must first be identity-proofed before being linked to a biometric or smart card. Equifax also recommends better training of registration staff in identifying counterfeit identification documents i.e. driver's license, Social Security number cards, birth certificates. Additionally, setting up a patient identity-proofing service, which uses "out-of-wallet" knowledge-based authentication questions, would help reduce the risk of incorrectly identifying the patient and thus have a positive impact on patient matching. Equifax also believes that optical character recognition (OCR) software is a potentially valuable technology to improve data collection since it can eliminate "fat finger" data entry errors during the registration process.

iTriage, LLC

iTriage is a global health care technology company founded in 2008 by two emergency medicine physicians. The company has developed mobile/Web applications for consumers to track their medical status and information. They suggest patient records could be linked and shared using a consumer-oriented approach, enabled by using OAuth2.0 standards.

Global Patient Identifiers, Inc. (GPII)

GPII is a nonprofit organization that has developed a VUHID system to augment an EMPI, with a portal interface that generates one or more unique identification numbers for patients, allowing them to segment their health care data before sharing with providers. The VUHID system is based on two ASTM International standards, E1714 and E2553 that were most recently approved in 2007. The VUHID system is architected to avoid the need for a central repository of patient demographic data by providing an identification card for patients with a VUHID number. The VUHID system has a "break-the-glass" function for health care emergencies, and can revoke the unique ID number in cases of fraud or misuse. More information on GPII is available at <u>http://gpii.info</u>.

Tascet Inc. (Vendor of Patient ID Services)

Tascet solutions support good risk management practices. Managing inherent risk related to EHRs can be achieved by three different approaches:

- 1. Remedial: a reactive/corrective approach that creates a lot of residual risk.
- 2. Detective: relying upon third party participation (paper documentation or demographics), this is a transfer of risk that may multiply residual risk.
- 3. Preventive: this approach prevents the risk from occurring.

Patient record matching will always have inherent risk that is impossible to mitigate, unless a patient identification protocol is implemented first. Patient identification and patient matching is not the same thing; one cannot exist without the other and they need to occur in the proper sequence. Patient record matching cannot consistently be accurate without a unique patient identification protocol; however, a patient identification process cannot rely upon paper documents



or demographics. Tascet provides a preventive protocol of identifying patients through a process of establishing a Super Identity. This Super ID consists of an identity packet that contains name, address, data of birth, gender, and country of birth along with finger proofs and a face proof. The identity packet is matched against other identities in its Identity Infrastructure to confirm the patient as unique and that a Super Identity has been established. A 16 digit Standard Patient Number (the unique patient identifier) is assigned to the patient. This accurately and consistently links the patient to their medical record and only their medical record within and across health information systems and eliminates record cloning. This process accomplishes the difficult task of ensuring that the same identifier is not attributed to multiple patients and multiple identifiers are not assigned to the same patients. The failure to implement an identification protocol first, undermines electronic record exchange, promotes widespread data contamination, encourages fraud, and is poor risk management.



Appendix H: Glossary

ADT feed: Admissions, discharges and transfers, an ADT feed is a type of HL7 (see Health Level Seven definition) message typically sent from a healthcare registration system to ancillary systems, regarding the status of a patient. ADT messages contain demographic information about a patient in the patient identification (PID) segment of the message, as well as patient status information and insurance information in other segments of the message.

Algorithm: A step-by-step procedure using mathematical calculations and predefined rules to solve a mathematical problem or to complete a computer process.

API call: API stands for application programming interface. An API is a set of commands, functions, and protocols which programmers can use when building software for a specific operating system, and an API call is a specific operation such as querying, adding, or deleting data between systems.

Biometrics: The use of a technological solution to identify a person through their biological or behavioral traits.

Blue Button: The Blue Button is a symbol on a website —for example, an online patient portal provided by a health care provider or insurer — that patients may use to view and download their health information into a personal health record. Developed by the Veterans Health Administration and implemented into various health information systems including those used by the United States Department of Defense and Health and Human Services, Blue Button has been expanded to Blue Button+ with additional data parsing features and transport protocols. It is being expanded for general use through the development of the Blue Button Connector in 2014.

Consolidated-Clinical Document Architecture (CCDA): The CDA is a standard developed by Health Level Seven International that provides a common architecture, coding, semantic framework and markup language for the creation of electronic clinical documents. The CCDA includes a collection of clinical documents including the consultation note, discharge summary, diagnostic imaging reports, history and physical, etc.

Cross-Community Patient Discovery (XCPD): A profile developed by Integrating the Healthcare Enterprise (IHE), an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. It supports locating communities with patient electronic health records and the translation of patient identifiers across communities.

Data attributes: The specific demographic information that an organization maintains to identify a patient's electronic record.

Data governance: Data governance represents the decision-making and accountability structure for managing data. This can include organizational policies and strategies that define the purpose for collecting data, the ownership of data, and the intended use of data. A data governance plan serves as the framework for overall organizational approach to data governance.



Data integrity: The idea that information is correct, complete, whole and has not been altered to conflict with the original intent of its creator.

Data quality management: The mechanisms, such as regular schedule of data review for errors, put into place by an organization to ensure the integrity of its data.

Deterministic matching: Records are determined to refer to the same patient if they have an exact match based on a subset of data, i.e. name, date of birth, address, or social security number. It is the simplest kind of record linkage, sometimes referred to as rules-based.

Direct: Direct, when used singularly in this report, refers to the implementation of the Direct Project's set of standards, protocols, and services that enable simple, secure electronic transport of health information (push messaging) between healthcare participants.

Duplicate: When one patient has two or more different medical records within the same healthcare organization or health information organization.

Enhanced Soundex: Algorithms where additional steps are taken, such as replacing many multiletter sequences that produce unrelated sounds, before performing the steps of the basic Soundex algorithm.

Enterprise identifier (EID): A unique identifier (number) assigned to a patient for use across an organization's facilities.

Enterprise master patient/person index (EMPI): A database that encompasses information for every patient or person in an enterprise (an organization that has multiple systems or a group of organizations with separate systems).

False match rate: The percent of incorrectly matched pairs that are accepted by the system as correct.

False negative: A match result that fails to match two records that represent the same person. The records are thought to relate to separate individuals.

False positive: A match result between two records that do not represent the same person.

Fellegi-Sunter theory: A theory of probabilistic matching pioneered in the 1960s that recognizes that each field-by-field comparison is subject to error.

Frequency indexing: Mathematical calculations used within an algorithm where common names or words receive lower scores and uncommon names or words receive higher scores.

Health Level 7 (HL7): A set of standards defining how information is packaged and communicated from one party to another within the healthcare industry. The standards are developed and maintained by an organization called Health Level Seven International, a nonprofit standards developing organization.



Historical values in matching: The use of previous addresses or names (e.g. maiden) as part of the matching technology that allows for a stronger link between records and consequently a larger number of matches.

Hybrid algorithm: A matching method that combines both probabilistic and deterministic algorithm rules.

IHE Profiles: Profiles developed and approved by IHE that document how standards will be used by each system to cooperate to address the problem. Referencing IHE Integration and Content Profiles helps ensure interoperability between systems by specifying technical details for use by developers and implementers of systems.

Linkage or potential linkage: Two separate records from different sources that may be the same patient and should be linked together in the MPI.

Master data management (MDM): The business processes and technology to support the mastering of core data that supports critical business processes across an enterprise.

Master patient/person index (MPI): An electronic demographic database with information on every patient registered at a health care organization.

Match: Two or more records in a database that have been identified through an electronic or manual process, as potentially containing information about the same individual; an initial match may require further validation.

Matching thresholds: The numeric scores set within an algorithm at which records are automatically linked, rather than manually linked within an electronic record system.

Medical record number (MRN): The unique identifier assigned to a patient's record within a specific electronic health record system or organization.

Merge/Unmerge: Merging two patient records combines them. Unmerging records involves creating two or more records from one record that has been previously merged, usually to separate out the information that is attached to two or more different patients and incorrectly combined into one. There are also merge and unmerge messages that can be sent under HL7 Master Patient Index Standards.

Metaphone3: A phonetic algorithm whose first version was published in 1990, used for indexing words by their English pronunciation. It was developed by Lawrence Philips as an alternative to the Soundex algorithm.

National Strategy for Trusted Identities in Cyberspace (NSTIC): A federally supported private sector initiative designed to provide consumers with a voluntary, secure means of maintaining an identity credential for use in banking, shopping, healthcare, and other aspects of modern life.

New York State Identification and Intelligence System (NYSIIS): A phonetic algorithm devised in the 1970s as an alternative to the Soundex algorithm.



Nickname tables: An electronic file, used within an algorithm, where a formal first name is linked to its associated nicknames or diminutive names.

Overlap: Two or more of the same patients' records from different facilities, using different MRN numbers, aggregated into an enterprise database.

Overlay: Two or more individuals incorrectly assigned the same identifier so that their health information is comingled into one record.

Patient Demographics Query (PDQ): A profile developed by IHE that allows applications to query by patient demographics for patient identity from a central patient information server and retrieve a patient's demographic and visit information.

Patient Identifier Cross-Referencing (PIX): A profile developed by IHE that supports the cross-referencing of patient identifiers from multiple electronic systems either via by query/ response or via an update notification.

Patient identity integrity (PII): The accuracy, quality, and completeness of demographic data attached to or associated with an individual patient.

Patient identification (PID) segment: The portion of the HL7 ADT message that contains 30 different fields of demographic information about the patient with values ranging from name and date of birth to marital status and citizenship. The PID segment is used as the primary means of communicating the identifying information about a patient between electronic systems.

Probabilistic matching: The process of using statistical analysis to determine the overall likelihood that two records match. Also known as fuzzy matching, it calculates the probability of a match.

Soundex: A phonetic algorithm for indexing names by sound as pronounced in English, developed by Robert C. Russell and Margaret K. Odell and patented in 1918.

Unique patient identifier (UPI): The value (usually a number) assigned to an individual for identification purposes that is unique across the entire national health care system. The United States does not permit the use of federal funds to research or promote the use of a national UPI.

Validate: A process used to authenticate, verify, or prove that two or more records identified as possible matches represent the same individual.

Voluntary unique patient identifier (VUPI): A system through which a person chooses to be assigned a unique identifier (usually a number) for use in all encounters the individual has with healthcare providers who are prepared to use the VUPI system to validate identities.

XDM (Cross-Enterprise Document Media Interchange): A profile developed by IHE that provides document interchange using a common file and directory structure over several standard media (USB, CD and email attachments). This permits the patient to use physical media to carry medical documents. This also permits the use of person-to-person email to convey medical documents.



XDR (Cross-Enterprise Document Reliable Interchange): A profile developed by IHE that provides for document exchange between EHRs, PHRs, and other healthcare IT systems using a secure electronic messaging system, such as Direct. This allows for secure exchange of healthcare documents in the absence of a repository or registry infrastructure.



Appendix I: Structured Interview Questions

Health System Questions

- 1. Are you using a commercially available patient matching product?
 - a. If so, which vendor and which version of the product?
 - b. How long have you been using this product?
 - c. If not, what do you use for patient matching or master data management?
- 2. What data elements/attributes are utilized by your patient matching product to facilitate accurate patient matching? Specifically, which fields within the attributes are utilized?
 - a. Is historical data such as previous address/ phone number used in the process?
 - b. Are the data elements different when matching within your organization, versus outside of your organization?
 - c. Is social security number or similar identifiers (e.g. Driver's license) used for matching?
 - d. Does your system put any of the data elements into a standardized format, such as USPS address format?
- 3. How many separate data sources does your system have?
- 4. Does your MPI create an enterprise identifier (EID) and link a patient's various MRNs to it?
- 5. If your solution creates an EID, do you send it outbound back to source systems to use as the primary identifier within each of those solutions, or is it simply an internal identifier?
- 6. Does your MPI solution utilize an algorithm to automatically match patients, if so, is it based on probabilistic or deterministic matching or both?
- 7. What is your current match rate across your enterprise, including sensitivity and positive predictive value?
- 8. What is your current false positive rate across your enterprise?
- 9. What is your current false-negative rate across your enterprise?
- 10. Do you currently have processes in place that would minimize the creation of a duplicate record (look up before create, data collection methodology, etc.)?
- 11. Do you have a process to correct your MPI once a duplicate or false positive has been detected?
- 12. Do have a data governance program and/or remediation program in place?
- 13. Are you working on initiatives to improve your patient matching accuracy? If so ... Elaborate
- 14. Are you focusing on technology or administrative initiatives, or both?
- 15. Do you have an internal goal for false positives/false negatives? How do you establish and measure performance metrics for your matching system?
- 16. Are the metrics reported internally beyond the IS/IT department? If so, to who are they reported?
- 17. Are you participating in data sharing activities (through an HIE or other sharing solution) in which your organization is a source to an external MPI?
 - a. If yes, does that external solution impose minimum data requirements or formats on your organization as a source?



- 18. Does your organization rely on any of the Integrating the Healthcare Enterprise (IHE) patient matching or search profiles (cross-community patient discovery (XCPD), patient identity cross-reference adds or updates (PIX), patient demographic query (PDQ)?
- 19. What additional steps do you take (post initial match) to confirm patient identity at the point of care?
- 20. Do you have a process to utilize patients at the point of care to help with matching?
- 21. Could biometric identifiers improve patient matching? Have you implemented this technology?
- 22. Does your EHR system allow you to collect patient photos? If yes, do you currently utilize the feature and could you use it to match patients at the point of care?
- 23. Do you use a registration system that assigns a unique Medical Record Number for every patient registered?
- 24. Do you use any system to prefill patient information at registration? E.g. card reader, PHR integration, etc.
- 25. Does your EHR or clinical system automate patient linking when an external record is sent inbound to your organization?
 - a. If so, how frequently does the solution require manual record-to-patient linking?
- 26. Do you have any control over the sensitivity of the patient linking process?
- 27. What technology, processes, or combination of both do you believe has the greatest potential to improve your patient matching performance?
- 28. Can you provide annual cost estimates for your organization for identifying and matching patients? Costs include hardware, software, and FTEs.
- 29. What is your best estimate of the annual overall financial cost to your organization for patient matching errors?

HIO Questions

- 1. Are you using a patient matching product that is part of your HIE technology vendor's offerings, or a best of breed approach with an outside product? (i.e. using Medicity or Orion Health's integrated MPI versus IBM Initiate or QuadraMed)
 - a. Which vendors are you using and which product versions?
 - b. How long have you been live with your product?
- 2. What data elements/attributes are utilized by your patient matching product to facilitate accurate patient matching? Specifically, which fields within the attributes are utilized?
 - a. Is social security number (or other ostensibly unique identifier) used for matching?
 - b. Does your system put any of the data elements into a standardized format, such as USPS address format?
- 3. How many data sources feed your MPI?
- 4. Does your MPI create an enterprise identifier (EID) and link a patient's various MRNs to it?
- 5. If your solution creates an EID, do you send it outbound back to source systems to use as the primary identifier within each of those solutions?
- 6. Does your patient matching algorithm leverage probabilistic or deterministic functionality?
- 7. Does your patient matching algorithm leverage a hybrid patient matching function (deterministic and probabilistic)?



- 8. Does your organization rely on any of the Integrating the Healthcare Enterprise (IHE) patient matching or search profiles (cross-community patient discovery (XCPD), patient identity cross-reference adds or updates (PIX), patient demographic query (PDQ)?
- 9. What is your current match rate?
- 10. What is your current false positive rate?
- 11. What is your current false-negative rate?
- 12. Do you currently have processes in place that would minimize the creation of a duplicate record (look up before create, data collection methodology, etc.)? What is that process?
- 13. Do you have a process to correct your MPI once a duplicate or false positive has been detected?
- 14. Do have a data governance program and/or remediation program in place?
- 15. Are you working on initiatives to improve your patient match rate? If so ... Elaborate
- 16. Are you focusing on technology or administrative initiatives to improve patient matching, or both?
- 17. Do you have an internal goal for false positives/false negatives? Or ... how do you establish metrics for measuring and monitoring the performance of your matching algorithm?
- 18. Could federal data sets be used to improve data quality? If yes, which data sets have the most potential?
- 19. What is your policy for data coordination when matching across organizations utilizing your HIO, do both parties need to correlate data quality in their algorithms, i.e. do you require the use of specific algorithms or match rates when exchanging with an external organization, or do they require it of you?
- 20. What technology, processes, or combination of both do you believe has the greatest potential to improve your patient matching performance?
- 21. Can you provide annual cost estimates for your organization for identifying and matching patients? Costs include hardware, software, and FTEs.
- 22. What is your best estimate of the annual financial cost to your organization of patient matching errors?

EHR Vendor Questions

- 1. Does your product utilize an outside patient matching product, or is patient matching functionality built into your product?
 - a. If you utilize an outside patient matching product, which product and what version?
- 2. What data elements/attributes are utilized by your patient matching product to facilitate accurate patient matching? Specifically, which fields within the attributes are utilized?
- 3. Is social security number or a similar (ostensibly unique) identifier such as driver's license used for matching?
- 4. Does your system transform any of the data elements into a standardized format, such as USPS address format?
- 5. If the data elements are received from a registration system, are they changed into a standardized format?
- 6. Does your product utilize an algorithm to automatically match patients? If yes, does leverage probabilistic functionality, deterministic functionality, or both?



- 7. Can your system use a "3rd factor" matching process that would enable patients to select from a limited set of questions and supply the answer (similar to password help in many non-healthcare applications (for example, on-line banking)? For example maternal grandmother's given/1st name?)
- 8. Does your EHR product rely on any of the Integrating the Healthcare Enterprise (IHE) patient matching or search profiles (cross-community patient discovery (XCPD), patient identity cross-reference adds or updates (PIX), patient demographic query (PDQ)?
- 9. When patient data is received from an outside EHR system, is the data automatically linked to a patient record?
 - a. If yes, how is this accomplished?
 - b. If no, how is patient data linked to a record?
- 10. Does your product match patients at the point of care (during registration) through integration with the ADT system or is matching performed after the registration event when the ADT message is routed to the software, or are both methods used?
 - a. Do you use the same algorithms at the point of care matching as in the passive or after registration matching?
- 11. Do you have a technical ability to unmerge erroneously merged patient records?
- 12. Does your system account for default/fake values when matching patients? E.g. Jane Doe, 01-01-1900, invalid SSN such as numbers starting with 000-00, etc.?
- 13. Does your registration system query the EHR for possibly matched patients before creating a new identity in the EHR?
- 14. Do any of your clients have a data governance program and/or remediation program in place?
- 15. Does your system have any mechanism to retrospectively identify and/or correct duplicate patient records or false positive matches within the EHR?
- 16. When a matching error (either false positive or false negative) is detected how does your system adjust to avoid that error in the future?
- 17. Does your EHR product automate patient linking when an external record (CCD or CCDA document) is received via a message such as a Direct message?
 - a. If so, how frequently does the solution require manual record-to-patient linking?
- 18. Do you allow providers and hospitals any control over the performance characteristics (e.g., sensitivity, positive predictive value, etc.) of the record to-patient linking process?
- 19. What have been your customers' experiences with the following metrics (what's the average metric across your clients?)
 - a. Match rate, including sensitivity and positive predictive value
 - b. False positive rate
 - c. False negative rate
- 20. What technology or processes/policies is your company considering to further reduce patient matching errors?

MPI/MDM/HIE Vendor Questions

- 1. HIE Vendors only Does your product utilize an outside patient matching product, or is patient matching functionality built into your product?
 - a. If you utilize an outside patient matching product, which product and what version?



- 2. Does your product's algorithm utilize matching functionality that is probabilistic, deterministic, or a combination?
 - a. Describe how your product uses probabilistic or deterministic functionality.
 - b. If probabilistic or deterministic functionality is used, are rules applied and if so, how and when?
 - c. Are customers allowed to choose which methods they use?
- 3. Does your product perform real-time matching, batch matching, or both? Are customers allowed to choose which method they use?
- 4. Does your product match patients at the point of care (during registration)? If so, how is this performed? Through integration with the ADT system, direct integration with Webservices or an API call, or is matching performed after the registration event when the patient data is received by the software?
 - d. Do you use the same algorithms at the point of care matching as in the passive or after registration matching?
- 5. What specific techniques does your product use, such as edit distance calculations, SoundEx, nickname substitution, etc.?
- 6. Does your product include ETL and data profiling tools? If so, which tools are used, for what purpose, and how?
- 7. Does your product rely on any of the Integrating the Healthcare Enterprise (IHE) patient matching or search profiles (cross-community patient discovery (XCPD), patient identity cross-reference adds or updates (PIX), patient demographic query (PDQ)?
- 8. Does your system transform any of the data elements into a standardized format, such as USPS address format?
- 9. Does your system account for default/fake values when matching patients? E.g. Jane Doe, 01-01-1900, etc.? .? If so, for which attributes and how do you visualize or quantify these default values.
- 10. Do you have a technical ability to unmerge/unlink erroneously merged patient records?
- 11. When a matching error (either false positive or false negative) is detected how does your system adjust to avoid that error in the future?
- 12. Does your product allow customers to customize the algorithm to attain lower false positive and false negative rates?
 - a. How is this achieved? What skills are required to make these changes, i.e. business process, data base management, code developer, etc?
 - b. How does your product report the results of such customization efforts?
- 13. What are the most common data elements/attributes utilized by your patient matching product to facilitate accurate patient matching? Specifically, which fields within the attributes are utilized?
 - a. Is there a limit to the number of attributes that can be leveraged?
 - b. Are customers allowed to choose the data elements/attributes that are used?
- 14. What have been your clients' experiences with the following metrics (what's the average metric across your clients?)
 - a. Match rate, including sensitivity and positive predictive value
 - b. False positive rate
 - c. False negative rate



- 15. Does your product respond to patient inquiries with potential matches in addition to definite matches? If so, how does it indicate the confidence levels of the potential matches?
- 16. What techniques and technologies are you exploring to yield improved patient matching performance?