PATIENT IDENTIFICATION AND MATCHING INITIAL FINDINGS

December 16, 2013
Prepared for the Office of the National Coordinator
for Health Information Technology by:
Genevieve Morris, Senior Associate, 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

Table of Contents

Initial Findings

From June through November 2013 an environmental scan was performed to assess the current industry capabilities and best practices for patient identification and matching, with a focus on matching across organizations. The scan included a formal interview process with a diverse set of large health systems; health information organizations (HIOs); electronic health record (EHR) vendors; and master data management (MDM)/master person index (MPI) and health information exchange (HIE) vendors. Informal informational discussions were also held with a wide variety of associations representing patients, providers, hospitals, public health organizations, and a variety of health industry workforce sectors, as well as several federal agencies. Over 50 organizations participated in the formal and informal interviews. The following findings represent the best prospects that were developed based on feedback received during the environmental scan and through an intensive review of historical literature on patient matching. The findings are limited by the methodology which did not allow quantitative analysis of these findings.

Standardization of Data Attributes

1. *Require standardized patient identifying attributes in the relevant exchange transactions.*

Patient identifying attributes are used in HL7 messages, (consolidated clinical document architecture (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. Consistently and completely populating a defined and standardized set of data attributes may 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. The table below details the recommended set of required data attributes for relevant exchange transactions and the strategy for improving each attribute. Where possible, the strategies for improvement utilize existing standards from the HIPAA X12 transaction sets, CAQH CORE initiative, HL7, ISO, and various recognized Internet standards.

Data Attribute	Strategy for Improvement
First/Given Name	1) Improve data consistency and normalize data
Last/Family Name	1) Improve data consistency and normalize data 2) 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.)

Data Attribute	Strategy for Improvement
Middle/Second Given Name (includes middle initial)	1) Improve data consistency and normalize data
Suffix	1) Improve data consistency and normalize data 2) 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) 3) If no suffix exists, should be null.
Date of Birth	1) YYYYMMDDHHMMSS 2) If hhmmss is not available, the value should be null 3) 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
Historical Address (street address, city, state, zip code)	1) Evaluate the use of an international or USPS format 2) If unavailable, the value should be null
Phone Number (if more than one is present in the patient record, all should be sent)	1) Utilize an ISO format that allows for the capture of country code. 2) Allow for capture of cell phone, home and work.
Gender	1) ValueSet Administrative Gender (HL7 V3): M, F, UN

Capturing Data Attributes

2. 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 recommendation 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. In addition, while all vendors capture first and last name, some do not have the capability to capture hyphens or apostrophes in the name fields. This leads to inconsistencies when sharing first and last name that may cause false negatives for systems utilizing deterministic matching. 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:

- Middle name or initial
- Suffix
- Current address
- Historical address(es)
- Phone (including home, business, and cell)

In addition, certification criteria should be introduced that require CEHRT to demonstrate the ability to capture apostrophes and hyphens in the first/given name and last/family name fields. Formats for

capturing these data attributes should not be further specified, as they will be standardized on outgoing and incoming messages.

Data Attributes Requiring Additional Study

3. *Study the ability of additional, non-traditional data attributes to improve patient matching.*

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. These data attributes could include email 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. Should such a requirement be considered for the future, it could perform an analysis of the ability of the data attributes to improve patient matching. 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. Additionally, the analysis should include a review of potential patient privacy and security concerns or issues that may be impacted by these additional data elements.

Patient Matching Algorithms

4. *Develop or support 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.*

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 great, investments in implementing and refining their patient matching solutions. During the environmental scan, many indicated that replacing their current systems would be cost prohibitive. As such, it is not suggested that a standardized patient matching algorithm be developed or required. Such a requirement would significantly impact technical and financial operations of health systems, HIOs, and vendors. In addition, imposing a federal standard could hinder market innovation and ultimately be detrimental to improving patient matching.

In a more limited way, however, there is 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.

Identifying Duplicates

5. *Certification criteria should be introduced that requires certified EHR technology (CEHRT) that performs patient matching to demonstrate the ability to generate and provide to end users reports that detail potential duplicate patient records.*

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.

Policies and Best Practices

6. *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.

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. The final public report of the Patient Identification and Matching Initiative, expected to be issued in the first quarter of 2014, will outline the most promising of these practices and serve as a starting place for a workgroup that could further evaluate their merit.

7. *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.

8. *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. As such, these organizations are well positioned to develop a marketing campaign that would include best practices and educational materials for collecting and verifying patient demographics. To have maximum impact, associations have suggested that specific best practices to address the issue of data accuracy be weaved into a broader campaign emphasizing the positive impact of accurate patient data on clinical quality, care coordination, and the efficiency of payment processes.

In addition to working with the appropriate associations, the SAFER Guides initiative should be utilized. The initiative has developed a phased implementation approach of best practices for improving patient identification at the point of care. 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.