Patient Identification in Three Acts

by Lorraine Fernandes, RHIA, and Michele O'Connor, MPA, RHIA, FAHIMA

As the curtain slowly rises on interoperable data exchange, the industry is still auditioning for the role of patient identification.

Throughout the past ten years, healthcare has been setting the stage for IT interoperability. With patient identification in a leading role, the script has been outlined, written, reworked, and discussed some more. Now, after years of rehearsals and scripting, the curtain is slowly rising, with patient identification downstage front and center. However, the industry still hasn't decided which identification method will play the role.

Currently there are several possibilities based around two general methods: probabilistic matching, which seeks matches on multiple data elements such as name and date of birth, and a national identification number, in which each person would be issued a unique medical ID number. Industry developments in the coming year are likely to have an influence, including:

- Political changes resulting from the November 2008 elections
- The pending release of the RAND report on patient identification (commissioned by the HIMSS EHR Vendor Alliance)
- Pilot projects related to the nationwide health information network, which will demonstrate use cases and interconnectivity between data exchange organizations
- Additional data exchange activities as more health information exchange organizations become operational
- Ongoing HIPAA compliance efforts
- Exchange and identification activities in Canada, the United Kingdom, and other countries
- And, of course, privacy remains a constant and central character in this play.

In fact, the privacy issues have become more complex than they were five and ten years ago. In that time, the industry has seen the advent of personal health records, the deployment of electronic health records, and a frequent string of headlines reporting security breaches in healthcare, banking, finance, and government.

These developments have heightened consumer unease over data exchange and patient identification. For some, there is a common belief that electronic systems are less secure than traditional paper-based health records. Others fear that employers who gain access to medical information may discriminate against employees.

The Commission on Systemic Interoperability, for one, has stressed the need to resolve the question of patient matching. In 2005 the commission, charged by Congress with developing recommendations and priorities for implementing an electronic health information exchange network, called for a resolution on a national standard for patient authentication and identity.

Has Anyone Else Figured It Out?

Several times a year throughout the world, Integrating the Healthcare Enterprise (IHE) Connectathons demonstrate patient matching using HL7 messaging and the IHE Patient Identifier cross-referencing integration profile and Patient Demographic Query manager. US regional health information organizations are beginning to embrace the IHE approach, namely the profile and manager segments, which use the HL7 standards in the v2.x or v3 worlds. Thus, they are ensuring interoperability of patient matching as standards evolve, and vendors can write standards-based software that can be used globally.

Canada and Europe

In 2001 Canada decided against a new federal healthcare identifier, despite the presence of some quasi-unique provincial identifiers. Instead the country adopted the client registry approach, which requires enterprise master patient index software using probabilistic matching. The approach is already deployed as the foundational technology in the majority of the provinces. A pan-Canadian view of data is the ultimate goal, but privacy differences between provinces must be addressed prior to launching this next level of patient matching on a national scale.

Many European countries have country-specific or regional healthcare identifiers, but concern remains about the integrity of the identifier and privacy of health information. The European Union has begun to explore how to match patient records across the member states, with research and policy discussion just starting. A key focus is on the use of smart cards to achieve interoperable health records. The European Union’s stringent data protection, privacy, and security policies will guide future discussion and research efforts.

Act One: Probabilistic Matching Improved

Probabilistic matching has been used by healthcare and numerous other industries for decades. Though the details get technical, the theory and its evolution are essential for any data quality initiatives that rely on record-linking technologies.

The underlying Fellegi Sunter probabilistic theory, first formalized in 1969, has been advanced significantly in the past decade. This is the theory used by many vendors that offer probabilistic matching products.

New estimation technologies have extended the original work to incorporate closeness measures, which enhance matching through distance editing (identifying closeness of the match, such as one or two characters off). The technology broadens the search to identify possible matches where human keying mistakes may have caused data errors.

Identifying near matches establishes better data quality at the beginning of an episode of care. Probabilistic matching also adapts well to the declining use of the Social Security number as a data element, which is employed less frequently in patient matching today due to privacy and identity theft concerns. The method also adapts well to the increasing use of new data elements such as cell phone numbers and even biometrics.

Probabilistic matching recognizes that the quality and volume of data may vary significantly but takes advantage of the strength of the data in order to present the best results. These systems have recently demonstrated their capabilities, speed, and accuracy on national-sized databases, aiding an array of data quality initiatives.

In a data exchange environment, organizations using this probabilistic approach will decide how to manage data quality at various levels, potentially including patient consent as a part of these discussions. For example, they will have to determine if they will establish a minimum data quality threshold for each stakeholder group, in addition to monitoring and potentially managing data quality at the exchange level.

The probabilistic algorithm must be able to address this complexity, as different approaches will be used throughout the country, and the stakes increase significantly when advancing from a local stakeholder setting to a regional health information exchange organization. Solving data issues at a local level is easy—in a multistakeholder setting with few baseline policies, it is vastly more challenging.

The probabilistic approach to patient matching was supported in 2005 by Connecting for Health in the report “Linking HealthCare Information: Proposed Methods for Improving Care and Protecting Privacy.”

Critics of the method say that 99 percent accuracy cannot be achieved; therefore a national identifier is required.

What Role for Social Security Numbers?

When the US government introduced Social Security numbers (SSNs) in 1936, it assured the public that their use would be limited to the Social Security program. That didn’t last, of course, and in the recent debate over patient matching in healthcare, SSNs have been put forward as a logical candidate for a national healthcare identifier.

The arguments generally voiced for using the SSN include: most people have one, most have it memorized, each person should have only one, no two people should have the same one, and the infrastructure is in place for administering them.

However, there are several problems with using the SSN as a healthcare identifier. The numbering scheme has imbedded intelligence and does not contain check digits; hence, by modern standards it is a poor identifier subject to error in recording. In addition, not all persons in the country have an SSN, in particular foreign students and illegal immigrants. Within this population, there is widespread fraud and misuse of SSNs.

Further, the SSN has already been usurped by the credit industry, making it a compromised, “nonconfidential” identifier. With the rise in identity theft, tied mostly to stolen SSNs, many citizens are reluctant to use their SSNs in nontax situations. In fact, states such as California have laws preventing the use of the SSN as an identifier for healthcare insurance. These states, and many citizens, would look dimly on using the SSN as a healthcare identifier.

However, like it or not, the SSN is widely embedded in healthcare, and it is still a value that legacy systems rely on for patient matching. In these instances, the number may be used for searching but not displayed to any query. Additionally, using only the last four digits safeguards privacy while still lending high rates of matching.

The table here illustrates the SSN’s impact on patient matching in probabilistic matching. The first row shows that when complete name, date of birth, and zip code are present, the false negative rate (or missed potential matches) is 6 percent.

When the capture rates for the date of birth and zip code decrease, the false negative rate increases dramatically. Adding the SSN lowers the rate again, as shown in the third and fourth rows.
Whether the full SSN or just the last four digits are used makes little difference, as the third and fourth rows demonstrate. The false negative rates differ by a point. Thus using the last four digits of the SSN is a prudent approach to addressing privacy while not compromising the matching of patient records.

Name	Date of Birth	Zip Code	SSN	False Negative Rate
100%	100%	100%	0%	6%
100%	90%	90%	0%	22%
100%	90%	90%	70% (all digits)	7%
100%	90%	90%	70% (4 digits)	8%

Source: Initiate Systems, Inc.

Act Two: National Identifier Still on Hold

HIPAA in its initial state called for a national healthcare identifier, but Congress quickly took action to withhold funding for evaluation and policy development. An identifier was never formulated. Each year this funding embargo is renewed with little or no debate. Numerous groups including vendor alliances, government task forces, and professional organizations have advocated for a national identifier as the only way the US will achieve consistent linking of information, both within a single organization and across organizations in a data exchange environment.

The privacy debate surrounding a national identifier generally encompasses the concern that a single identifier creates the key to all healthcare data regarding a person. Although the identifier is intended to be limited to healthcare, the multiple uses of the Social Security number casts doubt as to how "sacred" a healthcare identifier could really be.

Major issues with this approach include the backporting of a national identifier to billions of existing paper and electronic records; the costs of formulating, introducing, and propagating a new healthcare identifier; and the sheer political and organizational challenge in deciding who gets an identifier and how they would be issued. It could take a decade, if not two, for a national identifier to have an impact.

There is another issue. The unique identifier would likely be subject to the same data corruption as other data elements. The same issues that plague organizational identifiers today, such as multiple medical record numbers or accidental misuse, will continue; however, a unique identifier may limit the occurrences.

Act Three: Enter VUHID

Two new leading actors on the interoperability stage are a pair of ASTM standards (E1714 and E2553) passed by committee E31 in 2007. These standards describe the architecture and implementation of the Voluntary Universal Healthcare Identifier (VUHID) system. The standards propose a system where voluntary healthcare identifiers are issued by an organization at no cost to facilitate the linking of clinical information as a patient designates.

The standards define Open Voluntary Identifiers that are used to support unrestricted sharing of information. The standards also allow for Private Voluntary Identifiers, which the patient can use for restricted sharing of information on a wide variety of conditions such as psychiatric, genetic, or behavioral health information. A typical patient would have only one open ID but may have as many private IDs as his or her situation warrants.

One benefit of the system is that healthcare providers would participate in the VUHID network based on their established relationship with their patients. Providers are able to authenticate the patient's identity, and then a private or open ID can be requested on the patient's behalf. Patients are in control of what information will be shared with whom. The system requires an enterprise master patient index with probabilistic matching to manage the identifier assignment and dissemination process.

The underlying assumption is that regional data exchange organizations with an enterprise master patient index will support this standard, and that the index software will be robust enough to scale to the heightened demand. By taking a voluntary approach layered on top of this structure, the VUHID proposal avoids virtually all of the "show stopper" problems that have previously prevented implementation of a national healthcare identifier system. A prototype of the VUHID project, being conducted by ASTM, is expected by mid-2008. Further information can be found at www.vuhid.org.

While not specifically endorsing the VUHID approach, in late 2007 the National Alliance for Health Information Technology issued the point of view paper "Safety in Numbers: Resolving Shortcomings in the Matching of Patients with their Electronic Records," urging the adoption of a voluntary, patient-controlled national healthcare identifier.

Is This the End or Just an Intermission?

With many actors and complex plotlines, enabling true interoperability and data exchange hasn't been easy. But between probabilistic matching, a national identifier, and VUHID, a clearer picture of an ending is beginning to emerge. Throughout all these discussions, the importance of data quality and the need for confidential, flexible, and realistic approaches remains a constant.

Richard Hillestad, principal researcher of the RAND Corporation, previews an upcoming RAND report, noting a potential resolution: "RAND research on patient identification indicates that a unique patient identifier could play an important role in enhancing patient privacy and security while accurately linking patient records. Other methods of linking to patient records such as algorithmic matching will also be necessary but could use the unique patient identifier to advantage in searches across large systems such as a [nationwide health information network]."

Indeed, curtain calls may be in sight.

***Lorraine Fernandes** (lfernandes@initiatesystems.com) is senior vice president of healthcare practice and **Michele O'Connor** (moconnor@initiatesystems.com) is senior director of healthcare practice at Initiate Systems, Inc.*

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