Demographic Data Quality in a Multi-Stakeholder Health Information Exchange

November 29, 2010

Bill Beighe
Chief Information Officer
Physicians Medical Group of Santa Cruz
Santa Cruz Health Information Exchange

Demographic Data Quality in a Multi-Stakeholder Health Information Exchange

Item	Summer 2009	End of year 2010	Comment
Population in Medical Trading Area	300,000	300,000	Collecting data electronically since 1992 via electronic feed (ADT and batch file load) as well as manual entry
Demographic Data Sources. (There are more clinical data sources that are not allowed to update patient demographics.)			Brought on in stages starting in 1996 through 2010. 2 Hospitals, 2 Safety Net Clinics, 2 large medical groups, 2 large practices 400 Doctors total.
Total Number of Patient ID's in Community Master Patient Index	527,000	570,000	Includes duplicates, out of area patients visiting ED, and deceased patients. We brought on 1 large clinic serving the underserved in 2010 that added 70,000 patients, many were already in the patient index.
Total number of suspected duplicate patient records identified based on a number of algorithms	135,000	74,500	The number of duplicates is likely larger due to complications identifying married persons and other factors
Number of patients that share a SSN	41,000	17,000	Usually due to a data source feeding in the same SSN for all members of a family.
Number of patients with the first name BABY	17,000	16,000	Since new babies are added every day, more progress is being made on this than is reflected in the number.

A fundamental principle is that clinical data must be matched to the correct patient. Should a clinical message arrive that cannot exactly match, a new patient record is created in the community master patient index or depending on the source, the clinical message is delivered to the named providers but no new patient is created. This leads to duplicate patient records and suspected duplicate patient records that must be resolved manually.

The HIO operated with little manual reconciliation of patients or effort to eliminate duplicates for 14 years, simply relying on the built in tools provided by the vendor as well as various trained people at the end points to match and "clean up" where possible. Various sources fed clinical data in and delivered to providers as either matched or unmatched to a unique patient in the master community patient index.

In the summer of 2009, the HIO undertook a project prompted by meaningful use and the realization that all data needed to be accurately and uniquely matched to support Personal Health Records, create accurate longitudinal health records across a community, support medication reconciliation, support XDS.B and effective CCD/CCR exchange. Demographic quality is also at the heart of clinical quality initiatives and a root cause of payment delays/denials/rebilling.

The HIO invested 3 months in assessing the issue and created a plan to address patient demographic data quality. We determined that every entity contributed to the problem starting with enrollment, registration; check in, providers and the various vendors. Contributing to the issue in some locations were a lack of training and guidelines, outdated systems, lack of standards and best practices, lack of quality measurement of data quality and inadequate supervision.

The core HIO Demographic Quality Team

- Project Lead (8 hours / week)
- Data Analyst (5 hours / week)
- Subject Matter Expert (4 hours / week)
- Data Clean Up (1 FTE)
- Plus site administrators, interface engineers, vendor representatives

In late 2009 the team sought to address those issues and began identifying and ranking data sources according to quality of data. For example, some sources were routinely feeding in bogus data such as sending patient SSN as 123-45-6789 or a number pattern such as 999-99-9999 or 121-21-2121 for thousands of patients. Either those sources changed or we turned off SSN for that source.

We discovered that some data sources were using demographic fields to provide local work arounds, such as adding (DM) to the first name of the patient to support some sort of internal reporting project to identify patients with Diabetes Mellitus. This not only caused a matching challenge but was likely a HIPAA violation. Another source entered the address of the clinic itself in order to prevent mail from going to the patients' home. Yet another source was entering and sending an accession number in the patient MRN field as a work around to get the accession number to the recipient. All of these items caused duplicate records or confusing demographic entries requiring manual clean-up. Changes to the data feeds and/or education were provided to change the behavior.

Babies present a challenge. Babies are initially given the name "BABY" JONES, where Jones is the mother's maiden name. After a period of time when the baby is given a legal name the name might auto update to TOM SMITH. Smith being the family name. Sometimes the auto update does not happen. Although we have a data feed from the vital records department, they are by law prevented from sharing the mother's maiden name, making accurate linking challenging to impossible in some cases. The HIO has 16,000 individuals with the first name BABY. A separate project is underway to address this.

Hyphenated names, names with special characters, marriages, suffixes such a III or MD and people using something other than their legal name all present challenges and we are developing best practices suggesting how these are addressed. Some systems have difficulty accommodating these real life situations. Standards will help here once implemented in production versions of software. Historically, many healthcare systems are years behind the current release from their vendor. So standards alone will not address the gap.

Ironically, e-prescribing actually contributed to the duplicates issue. A patient must exist in the system in order for an eRx to be written. In many cases the patient DID exist, but the provider could not locate the patient for some reason. Providers and staff had the ability to create a patient on the fly with minimal demographic data. In the example below, are these individuals the same person? No computer and few humans could possible know. Thus many duplicates were created.

First	Middle	Last	Sex	DOB	SSN	Phone	Address	City
THOMAS	ROBERT	JONES	M	01/10/1950	585-34-9812	831-465-7800	123 MAIN ST	VISTA
TOM		JONES	M	01/10/1950	585-44-9812			
TR		JONES	M	01/10/1950				

In early 2010, the HIO team included the MPI vendor in the weekly calls regarding issues and by mid-year had identified a number of enhancements in the software. These enhancements are due to be installed in early 2011.

The data cleanup role is split across 3 people. The work requires intense concentration and attention to detail however it is boring and can only be performed for several hours at a time. The total effort on an on-going basis toward cleanup is 1 FTE.

The community wide training and best practices guide will be offered to all stakeholders in early 2011 and each stakeholder will be encouraged to incorporate this material into their organization's policies and procedures. Additionally, HIO interface staff will work with feeds to standardize the data.

The team has observed that quality feeds on itself. As duplicate patients are consolidated, fewer new duplicates are being created since the incoming data conclusively is linked to an identified unique patient. Conversations with site administrators led to changes that immediately improved the data matching.

In the year since this project was established we have seen dramatic improvements; nearly a 50% drop is suspected duplicates. A 70% drop in people who share a SSN. A reduction in unidentified patients named BABY.

The HIO believes that over time, as providers adopt new systems that are compliant with meaningful use along with the training and best practices that the effort to manage this population of patients in the community master patient index will decrease.