

Standardized Demographic Data Improve Patient Matching in Electronic Health Records

Additional elements such as email addresses could further increase accuracy

Overview

The rapid adoption of electronic health records (EHRs) provides patients and clinicians with greater access to more complete and up-to-date medical information. Yet, EHRs can't fulfill their full potential to help coordinate care until patients and clinicians are assured that all of an individual's records can be accessed no matter where they are stored. To do so, EHRs must better link—or match—records for each patient across the many hospitals, offices, and other sites where that individual seeks care.

Typical matching approaches rely on comparing demographic data—such as names and birthdates—that are stored in different records to determine whether those records refer to the same individual. Automated algorithms typically conduct those comparisons, and either link a record when the data are similar enough or send a notification that no match could be made. Variation in the information that organizations collect and use to compare records can hinder accurate matches. In addition, discrepancies in data formatting among systems—such as abbreviating "Street" or "Avenue," or including different fields for city, state, and ZIP code—can also affect matching.

To address some of the variability challenges, many organizations—including a contractor that examined patient matching for the Office of the National Coordinator for Health Information Technology (ONC)—have recommended standardizing demographic data. Elements used for matching would be captured and formatted in the same way.

Until recently, the effects of standardizing demographic data had not been tested. Research from Indiana University has found significant benefits from standardizing addresses using the U.S. Postal Service's (USPS') standard. To realize even greater increases in matching rates, additional data elements, such as email addresses, should also be used when attempting to link records.

Current approaches to patient matching

Failed or inaccurate matches occur for many reasons, such as incorrect data, typos, and changes in patient information, such as when someone moves. Research has found that match rates between organizations can be as low as 50 percent, even when they share the same EHR system.¹

Currently, ONC requires EHRs to support the use of the following data elements for matching as patients move among health care providers and institutions:

- Address
- Name
- Birthdate
- Phone number
- Sex

ONC also has proposed that application programming interfaces, which are software tools that allow different technologies to communicate, include address and phone number as part of the data elements that EHRs make available for matching and other uses.

However, ONC does not require the use of certain standards for some of these data elements, and it has asked the health information technology industry to report back on new approaches to address the patient matching problem.

A test of recommendations for standardizing demographic data

Research funded by The Pew Charitable Trusts at Indiana University provided first-of-its-kind data on the effect of data standardization on patient matching rates.² To conduct the analysis, the researchers used tens of thousands of records from the Indiana Health Information Exchange (HIE), a county public health registry, Social Security's Death Master File, and a newborn screening laboratory. The dataset representing hospital-to-hospital exchange alone consisted of 100,000 records. Each of these databases had already been reviewed to determine which records belonged to the same person, enabling researchers to understand the number of correct and inaccurate matches both before and after standardization of select demographic data. The fields tested included last name, phone number, Social Security number, birthdate, and address.

Table 1

Data Elements and Standards Examined to Improve Patient Matching

Data element	Applied standard/rules	Effect
Last name	Applied normalization rule from the Council for Affordable Quality Healthcare, a nonprofit alliance of health plans and trade associations	Removal of special characters (such as apostrophes) and suffixes, such as "Jr."
Telephone number	Formatted according to International Telecommunication Union Recommendation E.123	Converting numbers to the format (XXX) XXX-XXXX
Social Security number	Default and invalid values removed	Invalid values made blank
Date of birth	Date rules applied	Invalid dates made blank
Address	Applied U.S. Postal Service certified address standardization rules	Corrected errors that would make an address undeliverable by the U.S. Postal Service and used agreed-upon abbreviations, such as changing "Boulevard" to "BLVD"

© 2019 The Pew Charitable Trusts

Key findings from standardization research

Standardizing addresses according to USPS guidelines yielded improved match rates in both the public health and HIE datasets. Specifically, researchers found that standardizing addresses generally increased match rates by about 3 percentage points. Match rates within the HIE dataset, for example, increased from 84.9 to 87.6 percent. In addition, researchers found that combining addresses in the USPS format with last name standardization produced as much as a 10 percentage point improvement in match rates, from 81.3 to 91.6 percent, which cut the number of unlinked records by half.

On the other hand, standardizing telephone, date of birth, and Social Security numbers had no appreciable effect. One possible reason is that these fields don't have the same degree of variation as addresses and last names.

Government should require USPS format for addresses along with more data

To ensure that health care providers get the information they request from another clinician or organization, both parties' EHR systems should use the same demographic data standards and elements for matching—an approach highlighted by the Government Accountability Office in a report to Congress required by the 21st Century Cures Act, passed in 2016.³ To achieve that target, ONC should start by specifying that addresses be formatted according to the USPS standard. Employing these standards does not necessarily require workflow changes. Address validation software can convert the data into the USPS format, much as this is already done in e-commerce.

In addition, health records contain other demographic data routinely collected that aren't typically used or made available for matching. Separate research published in 2017 found the availability of email increased from 9 to 54 percent between 2005 and 2014, while other identifying data such as Social Security numbers declined from 83 to 50 percent.⁴ The documentation of email is likely higher today given the adoption of patient-facing tools, such as portals, that often require emails to register.

To further improve match rates, ONC should also encourage the use of additional data routinely captured in EHRs—potentially, email addresses, insurance policy identification number, and mother's maiden name—for matching. Making these changes would ensure that more information is available to matching algorithms to better link records. ONC could specify standards and require additional data by updating the U.S. Core Data for Interoperability, which is a set of elements that are considered essential for exchanging information among health care providers, to include this extra information.

Conclusion

Research shows that standardizing data elements in EHRs, such as addresses and last names, would help ensure that patients' records are accurately matched when seeking care across health care facilities. The use of additional data elements already in EHRs, such as email address or mother's maiden name, could further improve matching. These steps, in turn, could provide patients and clinicians with better access to health data and enhance the coordination of care, improving outcomes and reducing costs.

Endnotes

- 1 Genevieve Morris et al., "Patient Identification and Matching Final Report" (2014), https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.
- 2 Shaun J. Grannis et al., "Evaluating the Effect of Data Standardization and Validation on Patient Matching Accuracy," *Journal of the American Medical Informatics Association* 26, no. 5 (2019): 447–56, https://doi.org/10.1093/jamia/ocy191.
- 3 Government Accountability Office, "Health Information Technology: Approaches and Challenges to Electronically Matching Patients' Records Across Providers" (2019), https://www.gao.gov/products/GAO-19-197.
- 4 Adam Culbertson et al., "The Building Blocks of Interoperability: A Multisite Analysis of Patient Demographic Attributes Available for Matching," Applied Clinical Informatics 8, no. 2 (2017): 322-36, https://doi.org/10.4338/ACI-2016-11-RA-0196.

For further information, please visit: pewtrusts.org/healthit

Contact: Ken Willis, communications officer

Email: kwillis@pewtrusts.org

Project website: pewtrusts.org/healthit

The Pew Charitable Trusts is driven by the power of knowledge to solve today's most challenging problems. Pew applies a rigorous, analytical approach to improve public policy, inform the public, and invigorate civic life.