Research Highlights

RAND HEALTH

Identity Crisis?

Approaches to Patient Identification in a National Health Information Network

Key findings:

- Compared with a statistical matching approach to electronic patient identification, a universal patient identifier (UPI) approach should reduce medical errors and improve systems' interoperability.
- A UPI approach would not significantly increase the risk of security or privacy breaches to patient information.
- A UPI for everyone in the United States would be more expensive to implement, but the additional costs should be viewed in the context of potential improvements in patient safety, system efficiency, and improved privacy protection.

Improvements in healthcare information technology (HIT), properly implemented and widely adopted, should save money and significantly improve the quality of health care in the United States. A 2005 RAND study estimated that annual savings from efficiency alone could be upwards of $77 billion (http://www.rand.org/pubs/research_briefs/RB9136/). A key component of these improvements is a National Health Information Network (NHIN) that would link disparate health care information systems across the United States to allow sharing of critical health information swiftly and easily.

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 mandated the development of a unique patient identifier (UPI) to enable physicians, hospitals, and other authorized users to share clinical and administrative records more efficiently. A UPI could serve as a building block for the new NHIN. Since 2004, the Department of Health and Human Services (DHHS) has moved forward with steps to develop the NHIN. However, development of a UPI, a key to linking records across the emerging network, has been completely sidetracked by privacy concerns. These concerns eventually led Congress to ban DHHS from expending funds to develop the UPI.

The congressional ban has led to reliance on the alternative approach to creating a patient identifier: the use of statistical matching techniques to identify and access patient information. This method involves the identification of patients by matching patient data, such as name, address, zip code, or other information, with medical records. Debate in policy circles continues about the relative merits of the UPI versus statistical matching approaches.

To provide a more factual basis for this debate, a RAND study team led by Richard Hillestad analyzed and compared these two approaches across a number of dimensions, including error rates, operational efficiency, costs, and privacy and security issues. The team conducted reviews of the research literature and relevant

This Highlight summarizes RAND Health research reported in the following publication:

Hillestad R, Bigelow JH, Chaudhry B, Dreyer P, Greenberg MD, Meili RC, Ridgely MS, Rothenberg J, Taylor R, *Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System*, Santa Monica, Calif.: RAND Corporation, MG-753-HLTH, 2008.

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of published, peer-reviewed documents.

Headquarters Campus
1776 Main Street
P.O. Box 2138
Santa Monica, California
90407-2138
TEL 310.393.0411
FAX 310.393.4818

www.rand.org

statutes, interviews with health and IT practitioners involved in patient identification and health information exchange, and discussions with key national providers, consumers, insurers, and privacy organizations. The study found that, compared with a statistical matching approach, a UPI should reduce errors and improve interoperability without significantly increasing the risk of security or privacy breaches. A UPI would be more expensive to implement, but the additional costs should be viewed in the context of potential improvements in patient safety, system efficiency, and improved privacy protection.

UPI and Statistical Matching: Pros and Cons

Potential for Errors. One advantage of a properly implemented UPI system is its freedom from errors. If patients have a single, unique identifier that follows them throughout their lives and is used only for health records, there is relatively little chance of a mismatch between individuals and their records. Because statistical matching attempts to substitute for a UPI by using other kinds of information, such as names, birth dates, addresses, zip codes, or employer information, this technique has a higher potential for error than the UPI option because the other kinds of information may not be unique to the individual, may change over time, and may also be entered in varying formats in different databases.

Because statistical matching involves the probabilistic pairing of patient data with medical records, two types of errors may occur: *false positives*, linking to the wrong patient's records, and *false negatives*, missing the link between a patient and some part of his or her record. Published analyses have found false-negative error rates of about 8 percent in medical databases, trending higher in large databases with millions of records. These errors can pose significant risks for patient safety if providers act on incorrect or incomplete patient information.

Operating Issues. There are significant operational differences between the two methods:

- A statistical matching approach presents a higher chance of ambiguity—that a record might belong to more than one individual. The process for parsing ambiguous records to ensure a correct match between a patient and records—known as *disambiguation*—is both more essential and more complex in statistical matching, because the number of potential matches and the types of information available are greater. As the need for this kind of parsing becomes greater, it may require human involvement, at which point the advantages of automation may be lost, particularly efficiency and interoperability.
- A statistical matching system may be less complicated to implement, because it does not require new registration or issuing new identification data.
- A UPI approach would allow more architectural flexibility. The complexity of a matching system that uses multiple personal attributes to identify patients increases exponentially with the number of attributes and the size of the network. Such a network would require a complex, hierarchical architecture to decrease the complexity for end users. By contrast, a UPI approach would permit end users to share information directly with other end users, bypassing the need for a hierarchical design.

Costs. The cost of a patient identifier depends on several variables, including the architecture chosen to achieve connectivity. To estimate the costs of a statistical matching approach, the authors examined one proposal that would consist of a "network of networks," in which individual providers would subscribe to a hierarchical structure that allows linking of patients to patient data in a particular region. This approach would require a onetime investment of $90 million and an annual maintenance cost of about $18 million to fund the Record Locator Services required for the matching. If patients must be enrolled in this system, the enrollment cost is estimated to be $1.5 billion. In comparison, a mandatory UPI system could be substantially more expensive. One estimate, based on adapting and enhancing Social Security numbers to be more secure, for use as UPIs, put total costs at between $3.9 and $9.2 billion, depending on the security features.

Security and Privacy. Security and privacy could actually be strengthened with a UPI. A unique patient identifier, once developed, would immediately become protected health information under federal and (applicable) state law. UPIs would be sensitive information and could be a target for illicit access. Unlike the demographic components of a statistical matching algorithm (such as the Social Security number), however, the UPI would not link to financial records that are the specific target of identity thieves. If the UPI were to facilitate the development of a more efficient national network, any potential negative effects of such a network could be ameliorated directly through other aspects of systems architecture, such as encryption, access controls, and audit trails. And use of a UPI would actually improve privacy by limiting the transmission of more sensitive identifiers, such as the combination of names, address, date of birth, and Social Security numbers.

Related RAND analysis examined privacy and security issues from a legal perspective. RAND researchers Michael Greenberg and Susan Ridgely studied the implications of UPIs in the context of an NHIN (http://www.rand.org/pubs/research_briefs/RB9376/). They found that the emerging NHIN faces legal hurdles regardless of which approach to patient identification is adopted. They contend that the controversy over UPIs has distracted from the key privacy issue connected with an NHIN: the need for stronger protection for medical information under HIPAA in the context of an NHIN. They make the case that HIPAA's privacy rules are not adequate for an NHIN, regardless of whether the network involves a uniform national system or a patchwork arising from regional health information organizations. Therefore, strengthening HIPAA rules, not patient identification schemes, should be at the center of the national debate.

Conclusions and Implications for Policy

- *Adoption of a UPI would bolster the U.S. health care system*. Although a UPI approach would be more expensive to implement, it should reduce errors, improve patient safety, and enhance the interoperability and efficiency of health information networks.
- *A UPI approach could strengthen patient privacy and security*. In the context of a networked health information system, security and privacy have much more to do with how access is managed and records maintained than with a specific identifier approach. Password protection and encryption of a UPI are relatively easy, whereas encryption of personal keys used in matching algorithms decreases the power of the algorithms. Repeated disclosure of personal information and its link to health information, required in statistical matching in a network, carries a greater security risk of disclosure of sensitive information than a UPI. It may also be more difficult to recover from errors using demographic matching. Once a person's health information is compromised, it is not possible to give the person a new identity with a new set of personal attributes.
- *Prohibiting development of a UPI actually sidesteps the larger problem*. Development of an NHIN requires first establishing a legal environment that best protects privacy while also encouraging the advances that interoperability would bring to health care quality and efficiency. Therefore, the authors recommend that Congress remove constraints on pursuing a UPI. Continuing a *de facto* endorsement of statistical matching as the only approach to linking patients with their health care records is likely to inhibit the effective development of the NHIN. ■

Abstracts of all RAND Health publications and full text of many research documents can be found on the RAND Health Web site at www.rand.org/health. This research brief was written by David M. Adamson. The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

RAND Offices
Santa Monica, CA • Washington, DC • Pittsburgh, PA • New Orleans, LA/Jackson, MS • Doha, QA • Cambridge, UK • Brussels, BE