P1 P2 P3 P4 P5 P6 P7 P8

T1 T2 T3 T4 T5 T6 M1 M2

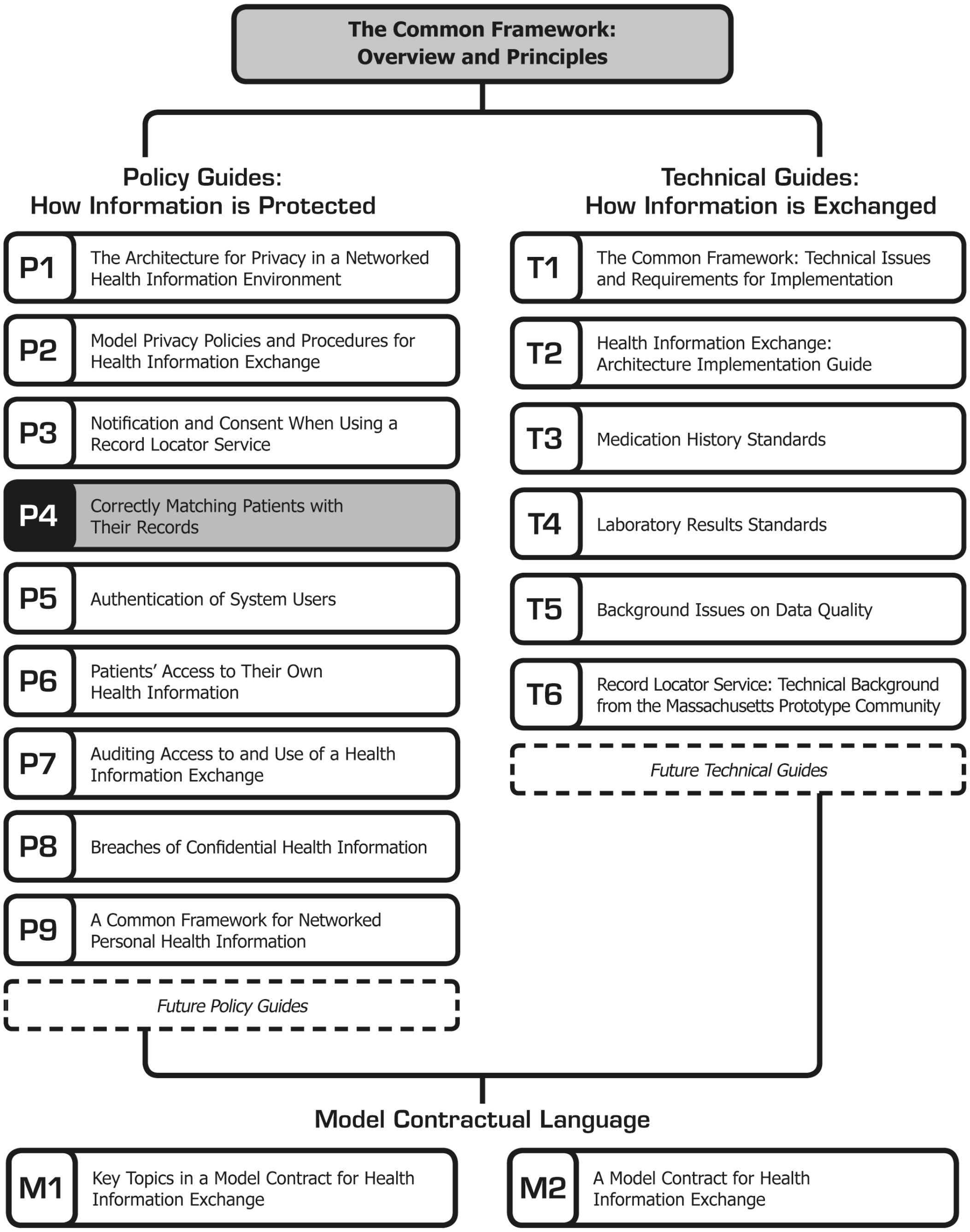
Correctly Matching Patients with Their Records

THE CONNECTING FOR HEALTH COMMON FRAMEWORK

Correctly Matching Patients with Their Records

CONNECTING FOR HEALTH COMMON FRAMEWORK

The document you are reading is part of *The **Connecting for Health** Common Framework,* which is available in full and in its most current version at: http://www.connectingforhealth.org/. The Common Framework will be revised and expanded over time. As of October 2006, the Common Framework included the following published components:


The Common Framework: Overview and Principles
Policy Guides: How Information is Protected
Technical Guides: How Information is Exchanged
P1 The Architecture for Privacy in a Networked Health Information Environment
P2 Model Privacy Policies and Procedures for Health Information Exchange
P3 Notification and Consent When Using a Record Locator Service
P4 Correctly Matching Patients with Their Records
P5 Authentication of System Users
P6 Patients' Access to Their Own Health Information
P7 Auditing Access to and Use of a Health Information Exchange
P8 Breaches of Confidential Health Information
P9 A Common Framework for Networked Personal Health Information
Future Policy Guides
T1 The Common Framework: Technical Issues and Requirements for Implementation
T2 Health Information Exchange: Architecture Implementation Guide
T3 Medication History Standards
T4 Laboratory Results Standards
T5 Background Issues on Data Quality
T6 Record Locator Service: Technical Background from the Massachusetts Prototype Community
Future Technical Guides
Model Contractual Language
M1 Key Topics in a Model Contract for Health Information Exchange
M2 A Model Contract for Health Information Exchange

Correctly Matching Patients with Their Records*

Introduction

Health institutions with large numbers of records must rely on probability to declare that a given record or set of records matches a set of identifiers (name, gender, date of birth, etc.). The risk of this strategy, of course, is that the matches so recorded may not be accurate. There is some risk of "false negatives"—records that pertain to a patient but are not found. There is a much greater risk, however, from "false positives"—matches with records that do not pertain to the subject patient, but are wrongly returned in a search.

False positive matches carry two forms of risk—privacy risk and clinical risk. The privacy risk is that records pertaining to patients not under the care of a particular clinician will be delivered, exposing personal details to those who have no need for them. The clinical risk is that a clinician will make a decision based on information that is erroneous because it is actually information about a different person, not the subject patient. Although clinicians are trained to make allowances for the fact that there is a significant error rate in clinical information when they make important decisions, the technology for handling such matches still needs to be optimized for a high degree of certainty, and where incorrect matching does occur, the system should err on the side of returning false negatives rather than false positives.

In addition to the technology, however, there also need to be policies spelling out how systems containing patient information should operate. This document outlines a set of policies for matching patient records with patient demographic details, so as to minimize incidental disclosures of personal health information within the nationwide electronic health information exchange. This document was developed by the **Connecting for Health** Policy Subcommittee.

Part of the **Connecting for Health** effort is to define and develop the "Common Framework"—the set of technical and policy specifications designed to help sub-network organizations (including regionally based networks and national networks, such as the VA) create data exchanges among their participating members, while creating interoperability between sub-network organizations (SNOs)[1] (see http://www.connectingforhealth.org).

The goal of the Common Framework is to define a minimal set of commonly adhered to standards and policies that allow for the SNO-based implementation of health information networks that are nationally interoperable. One component of this system is the Record Locator Service (RLS), which is a file of the location of patient records, queryable only by authorized participants. The RLS is the "White Pages" of any given sub-network, the coordinating entity that allows institutions within that sub-network to know whether other institutions hold records relevant to a particular patient. The RLS is designed to take a query from authorized users in the form of demographic details and return only the location of one or more matching records.

The RLS must implement a matching algorithm for queries using a sometimes incomplete subset of the possible constellation of demographic details. Authorized queriers

* **Connecting for Health** thanks William Braithwaite, Senior Vice President and Chief Medical Officer, eHealth Initiative, and Clay Shirky, Adjunct Professor, New York University Graduate Interactive Telecommunications Program, for drafting this paper.

[1] A sub-network organization (SNO) operates as a health information data exchange organization (whether regionally or affinity-based) that operates as a part of the National Health Information Network (NHIN), a nationwide environment for the electronic exchange of health information made up of a "network of networks."

present a set of demographic details and receive in return zero or more matching record locations. Probability weighted matching can improve the quality of record matching by taking the specific characteristics of records in particular databases into account.

Issue: the "false positive match" and the RLS

- What should our recommendations or requirements be for optimizing matching probabilities so as to minimize incidental disclosures and clinical risk caused by false positive matches within the RLS?

Example:
Attempt to match: John Q Public, 1043 W. Easy St., Phoenix, AZ 85535, 5556060, 10-24-1950, 482891822.

Which of the potential matches should be returned in response to this query?

Sample Data Listed in Order of Probability of Match

Comparison Scoring
Part of Initiate's Identity Hub Software, http://www.initiatesystems.com
This example from their literature

Rec#	Name	Address	Phone	DOB	SSN	Example Score
101	John Q Public	"1043 W. Easy St, Phoenix, AZ.85535"	5556060	10-24-1950	482891822	20.0
102	Jon Public	"1043 W. Easy St, Phoenix, AZ.85535"	5556060	10-24-1950	482891822	18.0
103	J Public		5553232	10-25-1950	482891822	11.0
104	John Q Long	"552 Green Dr, Phoenix, AZ.85535"		11-15-1962	57265225	5.0
105	Danny Smith		5552745	10-24-1950	48289244	5.0
106	Kevin Dohert	"1028 W. Easy Ave, Phoenix, AZ .85535"	5554289		48224857	4.0

**Note:* The example score on a scale of 1 to 20 is an arbitrary placeholder for different levels of matching for purposes of discussion, but does not represent an absolute scale of probability.

Questions

1. *Does this false positive match scenario qualify as an "incidental disclosure" pursuant to HIPAA?*
2. *What should our recommendations be regarding prevention of such disclosures?*
3. *What should our recommendations be regarding what actions to take when such disclosures occur?*
4. *Is this a Common Framework issue?*

HIPAA

Pursuant to HIPAA privacy regulations, a covered entity is permitted to use or disclose protected health information for treatment, payment, or health care operations.[2] An entity is also permitted to use or disclose protected health information *incident to* an otherwise permitted use or disclosure, provided that it has complied with applicable requirements of the minimum necessary standard and required security safeguards.[3]

In proposing the addition of the incidental disclosure provision to the Privacy Rule in its 2002 guidance, the United States Department of Health & Human Services (HHS) described an incidental use or disclosure as a secondary use or disclosure that *cannot reasonably be prevented, is limited in nature, and that occurs as a by-product of an otherwise permitted use or disclosure*. As described in the preamble to the Privacy Rule, an incidental use or disclosure is permissible only to the extent that the covered entity has applied reasonable safeguards and implemented the minimum necessary standard. In addition, covered entities are not required to document permitted incidental disclosures in an accounting of disclosures.[4]

HIPAA's minimum necessary standard requires covered entities to limit how much protected health information is used, disclosed, and requested for certain purposes. These minimum necessary policies and procedures also reasonably must limit who within the entity has access to protected health information, and under what conditions, based on job responsibilities and the nature of the business. The minimum necessary standard does not apply to disclosures, including oral disclosures, among health care providers for treatment purposes.[5]

HIPAA security standards require that a covered entity must have in place appropriate administrative, technical, and physical safeguards that protect against uses and disclosures not permitted by the Privacy Rule and that limit incidental uses and disclosures.[6] It is not expected that a covered entity's safeguards guarantee the privacy of protected health information from any and all potential risks.

An incidental use or disclosure that occurs as a result of a failure to apply reasonable safeguards or the minimum necessary standard, where required, is a violation of the Privacy Rule. Failure to comply with HIPAA regulations can result in general fines of up to $25,000 per incident.[7]

1. Does the false positive match scenario qualify as an "incidental disclosure" pursuant to HIPAA?

Recommendation

The **Connecting for Health** Policy Subcommittee assumes that the RLS false positive match is an incidental disclosure pursuant to HIPAA, with the understanding that such a disclosure is permissible under the law only to the extent that the covered entity or entities involved have applied reasonable safeguards and implemented the minimum necessary standard. The parameters recommended in this document for such matches are believed to require such safeguards.

NB: It has been noted that, as a legal matter, it is unclear whether "false positive match" disclosures are "incidental" according to HIPAA. There may be a legal argument that they can be considered permissible treatment disclosures. In either case, the disclosure would

[2] 42 CFR 164.502(a)(1)(ii).

[3] 42 CFR 164.502(a)(1)(iii).

[4] 45 CFR 164.528(a)(1).

[5] 45 CFR 164.502(b), 164.514(d).

[6] 45 CFR 164.530(c).

[7] HIPAA is a federal law. Individual states may place additional requirements/restrictions on the communication/transmission of protected health information.

be permissible under HIPAA, so the result would be the same. At the time of this writing, there has been no authoritative guidance on the issue from HHS, although it is possible that an FAQ on the topic could be sought in the future.

2. What should our recommendations be regarding prevention of such disclosures?

Recommendations

The **Connecting for Health** Policy Subcommittee *assumes* that the covered entities who could be involved in a request for information from the RLS, including the requester of information, the RLS (which could be defined as a "business associate" pursuant to HIPAA[8]), and the entity holding information pointed to by the index are in compliance with HIPAA's minimum necessary standard and required security safeguards. Indeed, if the entities are not in compliance with HIPAA's requirements and disclosures of protected health information occur, they are in violation of federal law and subject to the penalties described above.[9]

Beyond the strictures of HIPAA, the false positive match scenario may still produce incidental disclosures, albeit "permissible ones" according to the law. For example, in the sample data presented above, it is possible that J Public is not the person for whom information was requested. If J Public had, perhaps, visited a psychiatric hospital and that information was both recorded in the index, even without any additional clinical information, and returned to the requester, J Public might feel that his privacy had been violated, despite the entities' compliance with the letter of the law.

Therefore, the **Connecting for Health** Policy Subcommittee recommends the setting of a minimum level of certainty before the RLS returns information to the requester; and that whenever that level of certainty is not reached, the RLS could request additional demographic fields until either the level of certainty is reached or no record can be returned. This recommendation is based on ethical and public policy reasons, as opposed to the merely legal requirements of the HIPAA "reasonable safeguards" standard. It is also based on the need to reduce errors in record linkage.

For example, if a requester submits information in five demographic fields for a patient to the RLS, but the RLS does not find a match with a certain level of certainty on any one record, the RLS will report back that there is no match.

In the case that the RLS can return no matches with the specified certainty level, the RLS could require additional demographic data in order to determine a match. For example, at this point, the requester could be asked to supply data for additional demographic fields.

These levels could be set in order to minimize to the extent possible incidental disclosures of protected health information in an effort to respect the privacy of patients for ethical and public policy reasons.

Issues considered in formulating these recommendations include:

1. Should the Policy Subcommittee specify a level of accuracy for matching? Yes.
2. Should the level of accuracy be different for different use cases? No. The Policy Subcommittee made it clear that the RLS will not accept "wild-card" queries and can only respond to attempts to locate records on an individually identifiable patient. Other than that, the RLS has no mechanism to distinguish one use case from another, so the level of accuracy should not change.
3. Assuming the Policy Subcommittee specifies a level of accuracy for matching, how should it be determined? At least for external requests for matches, the level of certainty should be high enough that the probability of data being returned on the wrong patient would be very unlikely. One in 100,000 and one in a million were mentioned as potential levels, but the level could be different for different databases. The Policy Subcommittee recommends that the figure of one in 100,000 be set as the initial maximum probability of a false positive error when querying an RLS. It is expected that this set point may be adjusted as experience with operational RLSs gives us more real

[8] 45 CFR 160.103.

[9] All covered entities except small health plans were required to have compliant security standards in place before 4/21/05, while small health plans have until 4/21/06 to comply with the HIPAA Security Rule.

data with which to judge whether it continues to be appropriate. The Policy Subcommittee also recommends that a large test data set and standard set of queries be developed so that vendors of matching algorithms can test against this standard.[10]

4. Under what circumstances, if any, would it be acceptable to lower a matching threshold—the "Break the Glass" scenario? For normal external requests, a Break the Glass scenario assumes that the requester can make better judgments about unreliable data than the probabilistic matching algorithm. The Policy Subcommittee concluded that this was a useful "escape" mechanism in the past but that the increasing sophistication of matching algorithms might make such a mechanism anachronistic in the future. Special circumstances such as internal research or audits were considered to be situations when the high probability level for the matching algorithm might be reduced. Breaking the Glass is fraught with technical, practical, and operational problems and may have greater potential for harm than benefit. The Policy Subcommittee concludes that such a mechanism has no place in the RLS. When such special circumstances arise, a requester should go directly to the source of the clinical data and work through local mechanisms for dealing with them.

3. What should our recommendations be regarding what actions to take when incidental disclosures occur?

Recommendations

The **Connecting for Health** Policy Subcommittee *assumes* that the covered entities who could be involved in a request for information from the RLS are in compliance with HIPAA's minimum necessary standard and required security safeguards. The incidental disclosures in this scenario would be permissible according to the law.

However, as discussed above, the false positive match scenario may still produce incidental disclosures, even if the RLS requires a high level of certainty in order to return information from the index. Given the above recommendations of this Subcommittee, if a match meets the criteria for a positive match, they are permissible disclosures and cannot possibly be prevented without making the threshold for a positive match so high that it would create an unacceptable level of false negative matches.

The **Connecting for Health** Policy Subcommittee adopts the following position: In the case in which a requester of information recognizes that information received from the RLS does not apply to the patient about whom information was requested, the requester should take reasonable steps to immediately destroy that information, including, where applicable, deleting the electronic version of that portion of the RLS response and/or any paper copies thereof.

4. Is this a Common Framework issue?

Recommendations

Yes. If the false positive match scenario is *not* approached as a Common Framework issue, it is possible that various SNOs could set different standards for certainty and different numbers of required fields of demographic data in order for the RLS involved to return information to the requester.

This variability in what information will be returned from an RLS raises reliability of information questions: How will I, as a provider, be able to rely upon the information returned from the RLS in Idaho as well as the information from the RLS in Maine? Moreover, it is also unclear how sub-network RLS systems could grow and become increasingly interoperable when different threshold standards for information return are set across the systems.

For these reasons, the **Connecting for Health** Policy Subcommittee adopts the position that the false positive match/incidental disclosure scenario and the recommendations arrived at by this Policy Subcommittee should be

[10] For more information on matching algorithms and probability, *see* "Linking Health Care Information: Proposed Methods for Improving Care and Protecting Privacy," available at: http://www.connectingforhealth.org/assets/reports/linking_report_2_2005.pdf.

considered part of the Common Framework, to allow for increased reliability and scalability of a nationwide electronic health information exchange.

Acknowledgements

The members of the **Connecting for Health** Policy Subcommittee have accomplished an extraordinary task in less than a year's time—the development of an evolving piece of work that can serve as the core of nationwide health information exchange—the policy components of **The Common Framework**. During this time, we have been fortunate to work with respected experts in the fields of health, information technology, and privacy law, all of whom have contributed their time, energy, and expertise to a daunting enterprise. Our consultants and volunteers have worked long hours in meetings and conference calls to negotiate the intricacies of such issues as privacy, security, authentication, notification, and consent in health information exchange.We offer them our heartfelt thanks for taking on this journey with us, and look forward to the remaining work ahead.

In addition, we would like to offer special thanks to the volunteers and consultants who authored the initial drafts of this body of work—their hard work created a strong foundation upon which to focus the Subcommittee's deliberations:Stefaan Verhulst, Clay Shirky, Peter Swire, Gerry Hinkley, Allen Briskin, Marcy Wilder, William Braithwaite, and Janlori Goldman.

Finally, we must note that none of this work would have been possible without the leadership and inspiration of our co-chairs, William Braithwaite and Mark Frisse.They have led us with steady hands and determination of spirit.

Connecting for Health Policy Subcommittee

William Braithwaite, MD, eHealth Initiative, (Co-Chair)

Mark Frisse, MD, MBA, MSc, Vanderbilt Center for Better Health, (Co-Chair)

Laura Adams, Rhode Island Quality Institute

Phyllis Borzi, JD, George Washington University Medical Center

Susan Christensen*, JD, Agency for Healthcare Research and Quality, United States Department of Health and Human Services

Art Davidson, MD, MSHP, Denver Public Health

Mary Jo Deering*, PhD, National Cancer Institute/National Institutes of Health, United States Department of Health and Human Services

Jim Dempsey, JD, Center for Democracy and Technology

Hank Fanberg, Christus Health

Linda Fischetti*, RN, MS, Veterans Health Administration

Seth Foldy, MD, City of Milwaukee Health Department

Janlori Goldman, JD, Columbia College of Physicians and Surgeons

Ken Goodman, PhD, University of Miami

John Halamka, MD, CareGroup Healthcare System

Joseph Heyman, MD, American Medical Association

Gerry Hinkley, JD, Davis, Wright, Tremaine LLP

Charles Jaffe, MD, PhD, Intel Corporation

Jim Keese, Eastman Kodak Company

Linda Kloss, RHIA, CAE, American Health Information Management Association

Gil Kuperman, MD, PhD, New York-Presbyterian Hospital

Ned McCulloch, JD, IBM Corporation

Patrick McMahon, Microsoft Corporation

Omid Moghadam, Intel Corporation

Joyce Niland, PhD, City of Hope National Medical Center

Louise Novotny, Communication Workers of America

Michele O'Connor, MPA, RHIA, MPI Services Initiate

Victoria Prescott, JD, Regenstrief Institute for Healthcare

Marc A. Rodwin, JD, PhD, Suffolk University Law School

Kristen B. Rosati, JD, Coppersmith Gordon Schermer Owens & Nelson PLC

Sara Rosenbaum, JD, George Washington University Medical Center

David A. Ross, ScD, Public Health Informatics Institute

Clay Shirky, New York University (Chair, Technical Subcommittee)

Don Simborg, MD, American Medical Informatics Association

Michael Skinner, Santa Barbara Care Data Exchange

Joel Slackman, BlueCross/BlueShield Association

Peter P. Swire, JD, Moritz College of Law, Ohio State University

Paul Tang, MD, Palo Alto Medical Foundation

Micky Tripathi, Massachusetts eHealth Collaborative

Cynthia Wark*, CAPT, United States Public Health Service Commissioned Corps, Centers for Medicare and Medicaid Services, United States Department of Health and Human Services

John C. Wiesendanger, MHS, West Virginia Medical Institute/Quality Insights of Delaware/Quality Insights of Pennsylvania

Marcy Wilder, JD, Hogan & Hartson LLP

Scott Williams, MD, MPH, HealthInsight

Robert B. Williams, MD, MIS, Deloitte

Joy Wilson, National Conference of State Legislatures

Rochelle Woolley, RxHub

Amy Zimmerman-Levitan, MPH, Rhode Island State Department of Health

**Note: Federal employees participate in the Subcommittee but make no endorsement*