*Bio*medical Informatics: One Discipline*

Proceedings

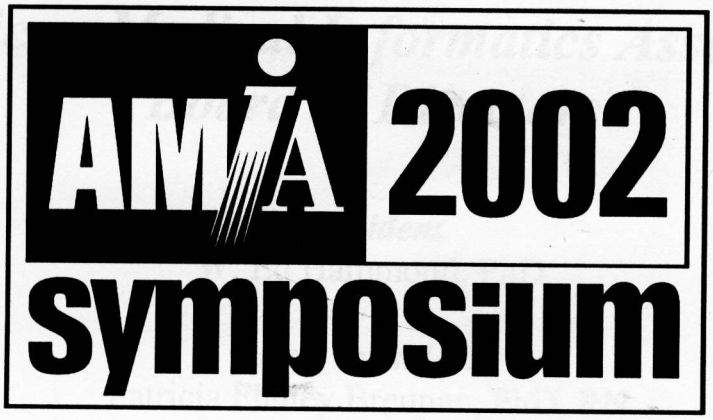
AMIA 2002 symposium

November 9–13, 2002
Henry B. Gonzalez Convention Center
San Antonio, TX

Edited by
Isaac S. Kohane, MD, PhD

AMIA®
Published by Hanley & Belfus, Inc.
www.hanleyandbelfus.com
ISBN 1-56053-600-4

Datavant Exhibit 1009



Bio*medical Informatics:
One Discipline

The Annual Symposium of the
American Medical Informatics Association

PROCEEDINGS

November 9-13, 2002
Henry B. Gonzalez Convention Center
San Antonio, TX

Edited by
Isaac S. Kohane, MD, PhD

HANLEY & BELFUS, INC.
Medical Publishers
Philadelphia

Datavant Exhibit 1009

AMIA 2002 Annual Symposium

Published by Hanley & Belfus, Inc.

Copyright © 2002 by American Medical Informatics Association

Printed in United States of America

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source.
Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 21 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For other copying, reprint or republication permission, write to Executive Director, American Medical Informatics Association, 4915 St. Elmo Avenue, Suite 401, Bethesda, MD 20814

1 2 3 4 5 6 7 8 9 0 MALMAL 9 8 7 6 5 4 3

Library of Congress Number 79-641114

ISBN: 1-56053-600-4

This book was prepared camera-ready.

Port City Press was printer and binder.

Datavant Exhibit 1009

Contents

Papers

A Framework for Evidence-adaptive Quality Assessment that Unifies Guideline-based and Performance-indicator Approaches 2
Primary Author: A. Advani, MD, MPH

Machine Learning Models for Lung Cancer Classification Using Array Comparative Genomic Hybridization 7
Primary Author: C.F. Aliferis, MD, PhD

Open Source Handheld-based EMR for Paramedics Working in Rural Areas 12
Primary Author: V. Anantraman, MBBS

Information Model and Terminology Model Issues Related to Goals 17
Primary Author: S. Bakken, RN, DNSc

Health Care Provider Quality Improvement Organization Medicare Data-sharing: A Diabetes Quality Improvement Initiative 22
Primary Author: D.J. Ballard, MD, PhD

Controlling the Vocabulary for Anatomy 26
Primary Author: R.H. Baud, PhD

Internal Medicine Resident Satisfaction with a Diagnostic Decision Support System (DXplain®) Introduced on a Teaching Hospital Service 31
Primary Author: B.A. Bauer, MD

The Helpful Patient Record System: Problem Oriented and Knowledge Based 36
Primary Author: E. Bayegan, MS

Evidence for the Rapid Construction of Preference During Utility Assessments 41
Primary Author: J. Bennett, MD

A Rapid Usability Assessment Methodology to Support the Choice of Clinical Information Systems: A Case Study 46
Primary Author: M.C. Beuscart-Zéphir

Getting Patients to the Right Healthcare Sources: From Real-world Questions to *Strategy Hubs* 51
Primary Author: S.K. Bhavnani

A Modern Optical Character Recognition System in a Real World Clinical Setting: Some Accuracy and Feasibility Observations 56
Primary Author: P.G. Biondich, MD

Evaluation of the UMLS as a Terminology and Knowledge Resource for Biomedical Informatics 61
Primary Author: O. Bodenreider, MD, PhD

Datavant Exhibit 1009

Automatic Generation of a Metamodel from an Existing Knowledge Base to Assist the Development of a New Analogous Knowledge Base 66
Primary Author: J. Bouaud, PhD

Evaluating a Medical Error Taxonomy 71
Primary Author: J. Brixey, MSN, MPH, RN

Knowledge-based Bioterrorism Surveillance 76
Primary Author: D.L. Buckeridge, MD, MSc

An Evaluation of Unmediated versus Mediated Retrieval Services 81
Primary Author: J.D. Buntrock, MS

Representation of Roles in Biomedical Ontologies: A Case Study in Functional Genomics 86
Primary Author: A. Burgun, MD, PhD

Computerized Reminders to Physicians in the Emergency Department: A Web-based System to Report Late-arriving Abnormal Laboratory Results 91
Primary Author: Z. Cai, MD, PhD

The Internet and Locus of Control in Older Adults 96
Primary Author: R.J. Campbell, EdD

A Knowledge Framework for Computational Molecular-disease Relationships in Cancer 101
Primary Author: M.N. Cantor, MD

Using Medication List - Problem List Mismatches as Markers of Potential Error 106
Primary Author: J.D. Carpenter, RPh, MS

The Implementation of a Personal Digital Assistant (PDA)-based Patient Record and Charting System: Lessons Learned 111
Primary Author: A.E. Carroll, MD

Initializing the VA Medication Reference Terminology Using UMLS Metathesaurus Co-occurrences 116
Primary Author: J.S. Carter, MBA

Analysis and Visualization of Functional Relationships Between RNA Expression and Clinical Annotation Using PathlinX 121
Primary Author: S.L. Carter

Machine Learning for Sub-population Assessment: Evaluating the C-section Rate of Different Physician Practices 126
Primary Author: R. Caruana, PhD

Design and Implementation of a Real-time Clinical Alerting System for Intensive Care Unit 131
Primary Author: H.T. Chen

Unsupervised Imaging, Registration, and Archiving of Tissue Microarrays 136
Primary Author: W. Chen

Datavant Exhibit 1009

YMD: A Microarray Database for Large-scale Gene Expression Analysis 140
Primary Author: K.H. Cheung, PhD

Software Engineering Risk Factors in the Implementation of a Small Electronic Medical Record System: The Problem of Scalability 145
Primary Author: M.F. Chiang, MD

Looking for French-English Translations in Comparable Medical Corpora 150
Primary Author: Y.C. Chiao, MSc

Test of the Technology Acceptance Model for the Internet in Pediatrics 155
Primary Author: W.G. Chismar, PhD

A Comparison of the Charlson Comorbidities Derived from Medical Language Processing and Administrative Data 160
Primary Author: J.-H. Chuang, MD, MS

The Horizontal and Vertical Nature of Patient Phenotype Retrieval: New Directions for Clinical Text Processing 165
Primary Author: C.G. Chute, MD, DrPH

Theoretical, Empirical, and Practical Approaches to Resolving the Unmet Information Needs of Clinical Information System Users 170
Primary Author: J.J. Cimino, MD

A Generic Computerized Method for the Estimate of Familial Risks 175
Primary Author: I. Colombet, MD, MPH

Usability of Expressive Description Logics – A Case Study in UMLS 180
Primary Author: R. Cornet, MSc

Towards a Sharable Numeric and Symbolic Knowledge Base on Cerebral Cortex Anatomy: Lessons Learned from a Prototype 185
Primary Author: O. Dameron

Quality Management Practices in Telehealth Programs: A Stakeholder Accountability Framework 190
Primary Author: K.H. Dansky, PhD

SYNCHRONUS: A Reusable Software Module for Temporal Integration 195
Primary Author: A.K. Das, MD, PhD

WEB-WAP Based Telecare 200
Primary Author: G. DeLeo, MSc

End-to-end Performance Measurement of Internet Based Medical Applications 205
Primary Author: P. Dev, PhD

Selective Retrieval of Pre- and Post-coordinated SNOMED Concepts 210
Primary Author: R.H. Dolin, MD

Expected Value Prioritization of Prompts and Reminders 215
Primary Author: S.M. Downs, MD, MS

Datavant Exhibit 1009

Delivering Bioinformatics Training: Bridging the Gaps Between Computer Science and Biomedicine 220
Primary Author: C.J. Dubay, PhD

The Stanford MediaServer Project: Strategies for Building a Flexible Digital Media Platform to Support Biomedical Education and Research 225
Primary Author: J.C. Durack, MD

Computerized Physician Order Entry and Communication: Reciprocal Impacts 230
Primary Author: R. Dykstra, MD

Implementation of a Classification Hierarchy for the GeneTests/GeneClinics Genetic Testing Databases 235
Primary Author: J. Edwards, PhD

WITH: A System to Write Clinical Trials Using XML and RDBMS 240
Primary Author: P. Fazi

A Proposed Key Escrow System for Secure Patient Information Disclosure in Biomedical Research Databases 245
Primary Author: T.A. Ferris, MD

Implementing Outpatient Order Entry to Support Medical Necessity Using the Patient's Electronic Past Medical History 250
Primary Author: F. FitzHenry, RN, PhD

Email Recruitment to Use Web Decision Support Tools for Pneumonia 255
Primary Author: J.R. Flanagan, MD, PhD

Relationships among Internet Health Information Use, Patient Behavior and Self-efficacy in Newly Diagnosed Cancer Patients Who Contact the National Cancer Institute's Atlantic Region Cancer Information Service (CIS) 260
Primary Author: L. Fleisher, MPH

Clinical Decision Support Systems: A Discussion of Quality, Safety, and Legal Liability Issues 265
Primary Author: J. Fox, PhD

Informatics Tools to Monitor Progress and Outcomes of Patients with Drug Resistant Tuberculosis in Peru 270
Primary Author: H.S.F. Fraser, MBChB, MSc

Exploring the Boundaries of Plausibility: Empirical Study of a Key Problem in the Design of Computer-based Clinical Simulations 275
Primary Author: C.P. Friedman, PhD

An XML-based System for the Flexible Classification and Retrieval of Clinical Practice Guidelines 280
Primary Author: T. Ganslandt, MD

Rapid Deployment of an Electronic Disease Surveillance System in the State of Utah for the 2002 Olympic Winter Games 285
Primary Author: P.H. Gesteland, MD

Datavant Exhibit 1009

Pediatric Cancer CareLink™ –Supporting Home Management of Childhood Leukemia 290
Primary Author: D.M. Goldsmith, RN, MPH

Quality of Life Assessment Software for Computer-inexperienced Older Adults: Multimedia Utility Elicitation for Activities of Daily Living 295
Primary Author: M.K. Goldstein, MD

Web-Based Informatics Education: Lessons Learned from Five Years in the Trenches 300
Primary Author: L.K. Goodwin, RN, PhD

Analysis of Identifier Performance Using a Deterministic Linkage Algorithm 305
Primary Author: S.J. Grannis, MD

Using the Metaschema to Audit UMLS Classification Errors 310
Primary Author: H. Gu, PhD

Fuzzy Logic Controller for Weaning Neonates from Mechanical Ventilation 315
Primary Author: G.E. Hatzakis, PhD

Neural Networks Morbidity and Mortality Modeling During Loss of HIV T-Cell Homeostasis 320
Primary Author: G.E. Hatzakis, PhD

Professional's Information Link (PiL): A Web-based Asynchronous Consultation Service 325
Primary Author: W. Hersh, MD

XML Representation of Hierarchical Classification Systems: From Conceptual Models to Real Applications 330
Primary Author: S. Hoelzer, MD

Combat Medical Informatics: Present and Future 335
Primary Author: R.W. Hoyt, PhD

Detecting Relationships Between Physiological Variables Using Graphical Modeling 340
Primary Author: M. Imhoff, MD

Accuracy of Three Classifiers of Acute Gastrointestinal Syndrome for Syndromic Surveillance 345
Primary Author: O. Ivanov, MD, MPH

A Customizable Similarity Measure Between Histological Cases 350
Primary Author: M.-C. Jaulent, PhD

Challenges in Implementing a Knowledge Editor for the Arden Syntax: Knowledge Base Maintenance and Standardization of Database Linkages 355
Primary Author: R.A. Jenders, MD, MS

A Rule-based Model for Local and Regional Tumor Spread 360
Primary Author: I.J. Kalet, PhD

Datavant Exhibit 1009

National Pharmaceutical Stockpile Drill Analysis Using XML Data Collection on Wireless Java Phones 365
Primary Author: B.T. Karras, MD

An Online Heart Simulator for Augmenting First-year Medical and Dental Education 370
Primary Author: R. Kelsey

Generating a Mortality Model from a Pediatric ICU (PICU) Database Utilizing Knowledge Discovery 375
Primary Author: C.E. Kennedy, MD

Using a Neural Network with Flow Cytometry Histograms to Recognize Cell Surface Protein Binding Patterns 380
Primary Author: E.-Y. Kim, MD

A Situational Approach to the Design of a Patient-oriented Disease-specific Knowledge Base 385
Primary Author: M.I. Kim, MD

DNA Splice Site Detection: A Comparison of Specific and General Methods 390
Primary Author: W. Kim, PhD

Agent-based Monitoring and Alert Generation for a Home Care Telemedicine System 395
Primary Author: V.G. Koutkias

Clinician's Assessments of Outpatient Electronic Medical Record Alert and Reminder Usability and Usefulness Requirements 400
Primary Author: M.A. Krall, MD, MS

Representing Nested Semantic Information in a Linear String of Text Using XML 405
Primary Author: M. Krauthammer, MD

Web-based Tailoring and Its Effect on Self-efficacy: Results from the MI-HEART Randomized Controlled Trial 410
Primary Author: R. Kukafka, DrPH, MA

Gene Expression Levels in Different Stages of Progression in Oral Squamous Cell Carcinoma 415
Primary Author: W.P. Kuo, DDS, MS

Usability Evaluation of an Experimental Text Summarization System and Three Search Engines: Implications for the Reengineering of Health Care Interfaces 420
Primary Author: A.W. Kushniruk, PhD

Validation of Electronic Student Encounter Logs in an Emergency Medicine Clerkship 425
Primary Author: J.S. Lee, MSIV

Terminology Extraction from Text to Build an Ontology in Surgical Intensive Care 430
Primary Author: S. Le Moigno, MD

Datavant Exhibit 1009

Provider Error Prevention: Online Total Parenteral Nutrition Calculator 435
Primary Author: C.U. Lehmann, MD

Use of the Internet to Study the Utility Values of the Public 440
Primary Author: L.A. Lenert, MD, MS

NLP-based Information Extraction for Managing the Molecular Biology Literature 445
Primary Author: B. Libbus, PhD

Neural Network Modeling to Predict the Hypnotic Effect of Propofol Bolus Induction 450
Primary Author: C.-S. Lin, MD

Using Binning to Maintain Confidentiality of Medical Data 454
Primary Author: Z. Lin, RN, MS

A Spatial Analysis of Obesogenic Environments for Children 459
Primary Author: G.C. Liu, MD, MS

A Study of Abbreviations in MEDLINE Abstracts 464
Primary Author: H. Liu, MS

An Integrative Model for In-silico Clinical-genomics Discovery Science 469
Primary Author: Y.A. Lussier, MD

Introducing an Academic Data Warehouse into the Undergraduate Medical Curriculum 474
Primary Author: J.A. Lyman, MD, MS

Communication Infrastructure in a Contact Center for Home Care Monitoring of Chronic Disease Patients 479
Primary Author: N. Maglaveras

Correlating Web Usage of Health Information with Patient Medical Data 484
Primary Author: B.A. Malin

Free-text Medical Document Retrieval Via Phrase-based Vector Space Model 489
Primary Author: W. Mao, MS

Connecticut RxData Project 494
Primary Author: M.K. Martin, DVM, MPH

Improved Coding of the Primary Reason for Visit to the Emergency Department Using SNOMED 499
Primary Author: J. McClay, MD

The Lexical Properties of the Gene Ontology 504
Primary Author: A.T. McCray, PhD

Usability Testing and Outcomes of an Interactive Computer Program to Promote Smoking Cessation in Low Income Women 509
Primary Author: A.M. McDaniel, DNS, RN

Datavant Exhibit 1009

Use of a MeSH-based Index of Faculty Research Interests to Identify Faculty Publications: An IAIMSian Study of Precision, Recall, and Data Reusability 514
Primary Author: K.A. McKibbon, MLS

The Effect of Sample Size and Outcome Prevalence on Supervised Machine Learning of Narrative Data 519
Primary Author: L.K. McKnight, MD

Comparing Communication Technology on Chinese, English, and Spanish Diabetes Web Sites 523
Primary Author: Y. Michea, MD

The USU Medical PDA Initiative: The PDA as an Educational Tool 528
Primary Author: L. Moore, PhD

PQL: A Declarative Query Language over Dynamic Biological Schemata 533
Primary Author: P. Mork, MS

Medical Informatics as a Market for IS/IT 538
Primary Author: T.A. Morris, PhD

A Comparison of Communication Needs of Charge Nurses in Two Operating Room Suites 543
Primary Author: J. Moss, RN, MS

A Neural Network Approach to Treatment Optimization 548
Primary Author: P. Munro, PhD

A Security Architecture for Query Tools Used to Access Large Biomedical Databases 552
Primary Author: S.N. Murphy, MD, PhD

A Semantic Normal Form for Clinical Drugs in the UMLS: Early Experiences with the VANDF 557
Primary Author: S.J. Nelson, MD

Implementation of an Automated Guideline Monitor for Secondary Prevention of Acute Myocardial Infarction 562
Primary Author: L.A. Noirot

The Chronus II Temporal Database Mediator 567
Primary Author: M.J. O'Connor, MSc

Comparing Imperfect Measurements with the Bland-Altman Technique: Application in Gene Expression Analysis 572
Primary Author: L. Ohno-Machado, MD, MHA, PhD

Impact of a Computerized Alert During Physician Order Entry on Medication Dosing in Patients with Renal Impairment 577
Primary Author: M.I. Oppenheim, MD

A UML-based Meta-framework for System Design in Public Health Informatics 582
Primary Author: A.O. Orlova, PhD

Datavant Exhibit 1009

Maximum Entropy Modeling for Mining Patient Medication Status from Free Text 587
Primary Author: S.V. Pakhomov, PhD

Computable Decision Modules for Patient Safety in Child Health Care 592
Primary Author: R.M. Pakpahan, MD, MHA

Knowledge Management: Evaluating the Organizational Requirements and Culture for an Emerging Technology 597
Primary Author: C. Parton

Characteristics and Override Rates of Order Checks in a Practitioner Order Entry System 602
Primary Author: T.H. Payne, MD

Support for Guideline Development through Error Classification and Constraint Checking 607
Primary Author: M. Peleg, PhD

Auditing the UMLS for Redundant Classifications 612
Primary Author: Y. Peng

Supporting Collaboration in Multidisciplinary Home Care Teams 617
Primary Author: D. Pinelle

Tracking Meaning Over Time in the UMLS® Metathesaurus® 622
Primary Author: T. Powell, MS

SmartQuery: Context-sensitive Links to Medical Knowledge Sources from the Electronic Patient Record 627
Primary Author: S.L. Price, MD, MS

Mining Time-dependent Patient Outcomes from Hospital Patient Records 632
Primary Author: R.B. Rao, PhD

Providing Context-sensitive Decision-support Based on WHO Guidelines 637
Primary Author: H.N. Ray, DDS, MS

Scale and Context: Issues in Ontologies to Link Health- and Bio-informatics 642
Primary Author: A. Rector, MD, PhD

Asking Questions: Information Needs in a Surgical Intensive Care Unit 647
Primary Author: M.C. Reddy, MS

Using LOINC to Link an EMR to the Pertinent Paragraph in a Structured Reference Knowledge Base 652
Primary Author: J.C. Reichert, MD

Cognitive Analyses of a Paper Medical Record and Electronic Medical Record on the Documentation of Two Nursing Tasks: Patient Education and Adherence Assessment of Insulin Administration 657
Primary Author: S.M. Rinkus, RN

Accessing Genomic Data through XML-based Remote Procedure Calls 662
Primary Author: A. Riva, PhD

Datavant Exhibit 1009

A Usability Study of Physicians Interaction with a Paper-based Patient Record System and a Graphical-based Electronic Patient Record System 667
Primary Author: N.J. Rodriguez, PhD

Evaluating the Prevalence, Content, and Readability of Complementary and Alternative Medicine (CAM) Web Pages on the Internet 672
Primary Author: S. Sagaram, MBBS

Discovering Protein Similarity Using Natural Language Processing 677
Primary Author: I.N. Sarkar, MPhil

An Unsupervised Self-optimizing Gene Clustering Algorithm 682
Primary Author: A. Schachter, MD, MMSc

A Knowledge Representation View on Biomedical Structure and Function 687
Primary Author: S. Schulz, MD

A Rule Driven Bi-directional Translation System for Remapping Queries and Result Sets Between a Mediated Schema and Heterogeneous Data Sources 692
Primary Author: R. Shaker

In Their Own Words? A Terminological Analysis of E-mails to a Cancer Information Service 697
Primary Author: C.A. Smith, MA, MILS, MSIS

Development and Pilot Testing of an Organizational Information Technology/Systems Innovation Readiness Scale (OITIRS) 702
Primary Author: R. Snyder-Halpern, PhD, RN

A Context-sensitive Methodology for Automatic Episode Creation 707
Primary Author: R.Y. Son, MS

Role Grouping as an Extension to the Description Logic of Ontylog Motivated by Concept Modeling in SNOMED 712
Primary Author: K.A. Spackman, MD, PhD

Pilot Study Optimizing MEDLINE Queries in an Automated Disease Management Telemedicine System 717
Primary Author: J.A. Spaeder, MD

Exploring Text Mining from MEDLINE 722
Primary Author: P. Srinivasan, PhD

Finding UMLS Metathesaurus Concepts in MEDLINE 727
Primary Author: S. Srinivasan, MS

Towards Elicitation of User Requirements for Hospital Information System: From a Care-process Modelling Technique to a Web-based Collaborative Tool 732
Primary Author: P. Staccini, MD, MS

Introduction of a Hierarchy to LOINC to Facilitate Public Health Reporting 737
Primary Author: S.J. Steindel, PhD

The Sublanguage of Cross-coverage 742
Primary Author: P.D. Stetson, MD

Datavant Exhibit 1009

Modeling Fatigue 747
Primary Author: W. Sumner, II, MD

Automated Concept Matching Between Laboratory Databases 752
Primary Author: Y. Sun, MD, PhD

Identification of Patient Name References within Medical Documents Using Semantic Selectional Restrictions 757
Primary Author: R.K. Taira, PhD

Extended Attributes of Event Monitor Systems for Criteria-based Notification Modalities 762
Primary Author: Y. Tao, MD

Head and Neck Lymph Node Region Delineation with 3-D CT Image Registration 767
Primary Author: C.-C. Teng

Supporting Physicians in Taking Decisions in Clinical Guidelines: The GLARE "What If" Facility 772
Primary Author: P. Terenziani, PhD

A Successful Technique for Removing Names in Pathology Reports Using an Augmented Search and Replace Method 777
Primary Author: S.M. Thomas, MD

Using Point of Service Clinical Documentation to Reduce Variability in Charge Capture 782
Primary Author: S.N. Thornton, PhD

Distributed Medical Informatics Education Using Internet2 787
Primary Author: P. Tidmarsh, JD

Crossing the "Digital Divide:" Implementing an Electronic Medical Record System in a Rural Kenyan Health Center to Support Clinical Care and Research 792
Primary Author: W.M. Tierney, MD

Defining Medical Concepts by Linguistic Variables with Fuzzy Arden Syntax 796
Primary Author: S. Tiffe, MS

Integration of a Standard Gastrointestinal Endoscopy Terminology in the UMLS Metathesaurus 801
Primary Author: M. Tringali, MD

Understanding Usage Patterns of Handheld Computers in Clinical Practice 806
Primary Author: M. Tschopp, MD

CHIP TUNER: A Web Tool for Evidence-based Noise Reduction in Gene Discovery 810
Primary Author: C.L. Tsien, MD, PhD

Data, Network, and Application: Technical Description of the Utah RODS Winter Olympic Biosurveillance System 815
Primary Author: F.-C. Tsui, PhD

Datavant Exhibit 1009

An Approach to Object-relational Mapping in Bioscience Domains 820
Primary Author: D. Tuck, MD

Implementing Security and Access Control Mechanisms for an Electronic Healthcare Record 825
Primary Author: F.-K. Ückert

Timely Redistribution of Information for Epidemiological Surveillance and Alert: The Experience from the French Communicable Diseases Network 830
Primary Author: A.-J. Valleron, PhD

Research for Research: Tools for Knowledge Discovery and Visualization 835
Primary Author: E.M. van Mulligen, PhD

Implementation of an Electronic Logbook for Intensive Care Units 840
Primary Author: C.J. Wallace, RN, PhD

The SNOMED Clinical Terms Development Process: Refinement and Analysis of Content 845
Primary Author: A.Y. Wang, MD

The Cognitive Demands of an Innovative Query User Interface 850
Primary Author: D. Wang, MD

GLEE - A Model-driven Execution System for Computer-based Implementation of Clinical Practice Guidelines 855
Primary Author: D. Wang, MPhil

A Clinical Rules Taxonomy for the Implementation of a Computerized Physician Order Entry (CPOE) System 860
Primary Author: J.K. Wang, MD

End of Visit: Design Considerations for an Ambulatory Order Entry Module 864
Primary Author: S.J. Wang, MD, PhD

Building an Asynchronous Web-based Tool for Machine Learning Classification 869
Primary Author: G. Weber

Conducting a Study of Internet-based Video Conferencing for Assessing Acute Medical Problems in a Nursing Facility 874
Primary Author: M. Weiner, MD, MPH

Temporal Knowledge Representation for Scheduling Tasks in Clinical Trial Protocols 879
Primary Author: C. Weng, MS

The German Specialist Lexicon 884
Primary Author: G. Weske-Heck, MD

Optimization of Clinical Teaching Unit Call Schedules at the Ottawa Hospital Through Tabu Search Heuristics 889
Primary Author: C.A. White, MD

Datavant Exhibit 1009

Using Client-side Event Logging and Path Tracing to Assess and Improve the Quality of Web-based Surveys 894
Primary Author: T.M. White, MD, MS, MA

Using Natural Language Processing to Analyze Physician Modifications to Data Entry Templates 899
Primary Author: A. Wilcox, PhD

Robustness of Empirical Search Strategies for Clinical Content in MEDLINE 904
Primary Author: N.L. Wilczynski, MSc

NLM Tele-educational Application for Radiologists to Interpret Mammography 909
Primary Author: M. Wu, MSc

Discovery of Gene-regulation Pathways Using Local Causal Search 914
Primary Author: C. Yoo, MS

Automatic Extraction of Gene and Protein Synonyms from MEDLINE and Journal Articles 919
Primary Author: H. Yu, MS, MPhil

Introducing Information Technology into the Home: Conducting a Home Assessment 924
Primary Author: T. Zayas-Caban

Matching of Flow-cytometry Histograms Using Information Theory in Feature Space 929
Primary Author: Q. Zeng, PhD

Toward a Cognitive Taxonomy of Medical Errors 934
Primary Author: J. Zhang, PhD

Enriching the Structure of the UMLS Semantic Network 939
Primary Author: L. Zhang

Progressive Display of Very High Resolution Images Using Wavelets 944
Primary Author: Y. Zhang

Filtering for Medical News Items Using a Machine Learning Approach 949
Primary Author: W. Zheng, MBBS, MCSc

An Assessment of the Visibility of MeSH-indexed Medical Web Catalogs through Search Engines 954
Primary Author: P. Zweigenbaum, PhD

Posters

Tracking Residents' Conference Attendance: A Reliable Solution at Last? 960
Primary Author: L.B. Afrin, MD

Gathering Customer Input Prior to Home Page Redesign: An Ontological Study 961
Primary Author: K. Alexander, MLS

Datavant Exhibit 1009

Artificial Neural Networks in Stroke Surveillance: The Brain Attack Surveillance In Corpus Christi (BASIC) Study 962
Primary Author: A. Al-Wabil, MS

Developing and Testing a Theory-guided Website to Promote Early Melanoma Detection Among Young Adults 963
Primary Author: M. Arnold, MPH, ScD

Facilitating Rapid Login to Clinical Information Systems: A Scalable Strategy 964
Primary Author: D. Aronsky, MD, PhD

Mining Access Logs Using Visual Linking of Entity Characterizations 965
Primary Author: P.V. Asaro, MD

Evaluation of Data Correction Methods for Positive Blood Culture Results 966
Primary Author: O. Assadian, MD

Home Monitoring of Hypertensive Patients through Intelligent Dialog System 967
Primary Author: I. Azzini, MSc

Case-base Reasoning in a Web-based Clinical Decision Support System for Thoracic Surgery 968
Primary Author: A. Babic, PhD

A Prospective, Controlled Trial of Technology-assisted Pharmacist Intervention to Improve Adherence to Coronary Artery Disease Secondary Prevention Medication Guidelines in Hospitalized Patients 969
Primary Author: T.C. Bailey, MD

SuML: A Survey Markup Language for Generalized Survey Encoding 970
Primary Author: M.W. Barclay

Assessing the Personal Digital Assistant Uses and Needs of Medical Residents 971
Primary Author: J.R. Barrett, PhD

Bridging the Digital Divide: Report on Focus Groups with Low-income Urban and Rural Residents in Pennsylvania 972
Primary Author: S.B. Bass, PhD, MPH

Standards-based Sharable Active Guideline Environment (SAGE): A Project to Develop a Universal Framework for Encoding and Disseminating Electronic Clinical Practice Guidelines 973
Primary Author: N. Beard, MD

Multi-access Services for the Management of Diabetes Mellitus: The M2DM Project 974
Primary Author: R. Bellazzi, PhD

A Markov Chain Model for Evidence-adaptive Clinical Decision Support Systems 975
Primary Author: G. Benoit, PhD

Datavant Exhibit 1009

Automated Information Extraction and Analysis for Information Synthesis 976
Primary Author: C. Blake, MS

The Potential Economic Impact of Ambulatory Physician Order Entry at a Large Integrated Delivery Network 977
Primary Author: B.H. Blumenfeld, MD, MS

GenNav: Visualizing Gene Ontology as a Graph 978
Primary Author: O. Bodenreider, MD, PhD

The Effect of Text Templates on Physician Data Entry 979
Primary Author: W.A. Bowes, III, MD, MS

Applying Axiomatic Design Methodology to Create Guidelines That Are Locally Adaptable 980
Primary Author: A.A. Boxwala, MBBS, PhD

A Usability Engineering Approach to Development of a Computer-based System for Continence Health Promotion 981
Primary Author: A.R. Boyington, RN, PhD

Endomebase: A Model Cross-species Tissue-specific Gene and Protein Expression Database 982
Primary Author: P.C. Bradshaw, PhD

An Online Image Management System for Anatomy Teaching 983
Primary Author: J.F. Brinkley, MD, PhD

Redesign and Prototype for the Interface of a Volumetric Infusion Pump Following the Usability Engineering Lifecycle 984
Primary Author: J. Brixey, MSN, MPH, RN

Semi-automatic Semantic Mapping between Nomenclatures: Knowledge Interlingua Transfer-knowledge Acquisition Technique (KITKAT) 985
Primary Author: P.J.B. Brown, MD, MRCGP

Case-based Reasoning for Medical Risk Stratification: Contrast Associated Nephropathy 986
Primary Author: D. Bu, MD

Returning to the Single Vendor HIS? An Example for the Integration of Department Specific Applications in a HIS 987
Primary Author: T. Burkle, PhD, MD

LISA: A Clinical Information and Decision Support System for Childhood Acute Lymphoblastic Leukaemia 988
Primary Author: J.P. Bury

Cancer Electronic Quality of Life Project: Delivering Wireless Web-based Technology at the Point of Care 989
Primary Author: K.M. Cam

The Delphi Method Online: Medical Expert Consensus Via the Internet 990
Primary Author: K.M. Cam

Datavant Exhibit 1009

The Use of an Emergency Department Screen Saver Linked to Web Resources to Raise Awareness of Bioterrorist Threats 991
Primary Author: L.L. Casebeer, PhD

Using Text Analysis to Examine ICD-9 Codes to Determine Uniformity in the Reporting of Medpar® Data 992
Primary Author: P.B. Cerrito

Image Misregistration and Lesion Signal-difference-to-noise Ratio Impact on Spectral Distance Images Formed from Multispectral Images 993
Primary Author: B.E. Chapman, PhD

Exchangeable Multimedia Framework for Digital Library in Medicine 994
Primary Author: B.-H. Chen

Issues in Developing Clinical Applications for the Wireless Environment 995
Primary Author: E.S. Chen, MA

Communicating the Molecular Epidemiology of Poliovirus by Using Geographical Information Systems for the Global Polio Eradication Initiative 996
Primary Author: P. Chenoweth, RN, MPH, ND

A Medical Decision Support System for Polyp Screening by Using Fuzzy Classification Trees 997
Primary Author: I.-J. Chiang

Text Boundary Detection of Medical Reports 998
Primary Author: P.S. Cho, PhD

Metadata Management for Clinical Research Databases 999
Primary Author: Q. Chong

The Use of a Pharmacy Data Mart as a Patient Safety Tool 1000
Primary Author: N.A. Coblio

Evaluating the Use of Computer-based Interactive Technology for Improving Outpatient Procedure Education 1001
Primary Author: C.T. Cowl, MD, MS

Design Approach for Integration of Demographic, Biologic, and Clinical Data 1002
Primary Author: J. Dalton, MS

Delivery of Evidence Based Medicine Resources on Wired Mobile Devices 1003
Primary Author: D. Davis, MA

Reframing Telemedicine in Kentucky: Developing an Evaluation Plan to Move from Electronic Clinical Encounters to an e-Health Resource for Population Health 1004
Primary Author: E. Drake, MSIS, MSPH

Is the Lack of Emphasis on Learning a Barrier to the Effective Deployment of Informatics? Content Analysis of NHS Strategy and Information Strategy Since 1998 1005
Primary Author: S. de Lusignan, MD

Datavant Exhibit 1009

Using New Object Oriented Expression Language (GELLO) to Encode Arden Syntax's Medical Logic Modules 1006
Primary Author: Y. Denekamp, MD

A New Tool to Identify Key Biomedical Concepts in Text Documents, with Special Application to Curriculum Content 1007
Primary Author: J.C. Denny

An Information System for 911 Dispatch Monitoring and Analysis 1008
Primary Author: M.R. Dockrey

CentraView: A Real-time ICU Patient Data Integration and Presentation System 1009
Primary Author: W.H. Drummond, MD, MS

Shape Characterization of the Corpus Callosum in Schizophrenia Using Template Deformation 1010
Primary Author: A. Dubb, MD

Implementation, Usage, and Impact of Automated Guidelines in the DMA: A Web-based Diabetes Disease Management System 1011
Primary Author: A.K. Dubey, MD

Integration of Patient Acuity Scoring and Reporting into a Clinical Data Repository and Web-based Clinical Information Viewing System 1012
Primary Author: R.G. Duncan, MD

Delivering Data Mining Without the Warehouse: MUSC's Experience 1013
Primary Author: M.D. Ebeling

So They Tell You Your Display Is "Wonderful," So What? 1014
Primary Author: J.A. Effken, PhD, RN

Existing Data Resources as Reference Standards for a Clinical Data Warehouse 1015
Primary Author: J.S. Einbinder, MD, MPH

Rule-enforcing, Frame-based Terminology Environment for Controlled Medical Dictionaries 1016
Primary Author: G. Elhanan, MD

MedCIRCLE – Collaboration for Internet Rating, Certification, Labeling and Evaluation of Health Information on the Semantic World Wide Web 1017
Primary Author: G. Eysenbach, MD

The Effectiveness of the Electronic Patient Record to Track and Control Multi-drug Resistant Organisms (MDRO) 1018
Primary Author: K. Feng

Introducing Handheld Computing for Interactive Patient Education 1019
Primary Author: J. Finkelstein, MD, PhD

A Service-based Approach to Integrating Epidemiological Information in Public Health Departments 1020
Primary Author: D.W. Forslund, PhD

Datavant Exhibit 1009

Evaluating the Process of Generating a Clinical Trial Protocol 1021
Primary Author: L.G. Franciosi, MSc

Principal Component Analysis and Variable Selection in Intensive Care Monitoring 1022
Primary Author: R. Fried, ProfDr

Modeling a Concept-based Information System to Promote Scientific Discovery: The Telemakus System 1023
Primary Author: S. Fuller, PhD

Supporting Loose Confederacies of Collaboration: Experiences from Computational Biology and 3D Imaging Collaboratories 1024
Primary Author: C.S. Gadd, PhD, MBA, MS

MOPPDB: Database for Translational Investigations at H. Lee Moffitt Cancer Center and Research Institute 1025
Primary Author: D. Garcia, MS

Diabetes and Your Eyes: A Pilot Study on Multimedia Education for Underserved Populations 1026
Primary Author: B. Gerber, MD

Validation of Survival Data in a Commercial Cancer Registry Database Using the U.S. Social Security Death Master File 1027
Primary Author: J.W. Goldwein, MD

A Database-driven Web-based System for Managing Research Protocol Change in a Multi-protocol, Multi-institutional Setting 1028
Primary Author: V. Gowda

A Study of the Adequacy of User and Indexing Vocabularies in Natural Language Queries to a MeSH-indexed Health Gateway 1029
Primary Author: N. Grabar, MSc

Conceptual Heterogeneity Complicates Automated Syndromic Surveillance for Bioterrorism 1030
Primary Author: J. Graham, MD, MS

Integrating Evidence and Data into Decision Models for Ongoing Technology Evaluation 1031
Primary Author: A. Grant, MD, PhD

Patient Enrollment for a Web-based Patient Portal System 1032
Primary Author: J.A. Greim, RN

Age Did Not Influence Use of the HeartCare System 1033
Primary Author: A. Ground, RN, MBA

The Integration of an e-Library and Computer-based Course Support Resources into a Dental School Curriculum 1034
Primary Author: G.F. Guest, DDS

Development of the CISQ-MA to Assess Nursing Attitudes Towards Electronic Medication Administration Modules of Healthcare Information Systems 1035
Primary Author: B. Gugerty, DNS, RN

Datavant Exhibit 1009

A Rule-based Computer System to Facilitate Public Health Surveillance during the 2002 Olympic Winter Games 1036
Primary Author: A.V. Gundlapalli, MD, PhD

An Evaluation of a De-identification Software Engine: Progress Towards Sharing Clinical Documents and Pathology Reports 1037
Primary Author: D. Gupta, MD

Re-using a Knowledge Base in Different Guideline Models: Experience in PRODIGY 1038
Primary Author: R.W. Hall, MSc

Analysis of Discordance on the 'Sign or Symptom' Concepts between the UMLS and Clinical Terms in Korean Medical Records 1039
Primary Author: S.-B. Han

MEDLINE Metric: A Method to Assess Medical Students' Search Effectiveness 1040
Primary Author: G.G. Hannigan, MLS, MPH, PhD

Conceptual Inferencing for Real-time Clinical Decision Support Using Hierarchical Queries in a Relational Database Management System 1041
Primary Author: R.R. Hausam, MD

Prognostic Model for Terminally Ill Cancer Patients on Laboratory Data 1042
Primary Author: K. Hira, MD

Mining Interesting Patterns in Time-series Medical Databases: A Hybrid Approach of Multiscale Matching and Rough Clustering 1043
Primary Author: S. Hirano, PhD

Experience with Message Format and Code Set Standards for Early Warning Public Health Surveillance Systems 1044
Primary Author: W.R. Hogan, MD, MS

A Framework for Characterizing the Development of Expertise in the Domain of Orthodontics 1045
Primary Author: J.L. Hong, DDS, MS

A Query Analysis of Consumer Health Information Retrieval 1046
Primary Author: Y. Hong, MLIS

Variability in User Interaction with Physician Order Entry System 1047
Primary Author: J. Horsky, MA

Modification of a PDA-based System for Data Collection 1048
Primary Author: Y. Hsieh, RN, MS

Outcome Prediction after Moderate and Severe Head Injury Using an Artificial Neural Network 1049
Primary Author: M.H. Hsu

A Decision Support System to Promote Oral Antibiotic Dosing 1050
Primary Author: T.M. Hulgan, MD

Datavant Exhibit 1009

Proposed Design of a Clinical Information System for the Management of Bronchial Asthma 1051
Primary Author: S. Huq, MBBS

Applying Telemedicine to Outpatient Physical Therapy 1052
Primary Author: J.L. Jacobs, MD

The Feasibility of Using Propositional Analysis as a Method for Understanding Differences Between Physicians and Nurses Reading EMRs 1053
Primary Author: C.M. Johnson, MS, RN

Implementing Medical Algorithms to Reduce Medical Errors 1054
Primary Author: K.A. Johnson, PhD

Searching for Patient Educational Material on the World Wide Web: An Exploration of Nurses' Search Behavior 1055
Primary Author: J. Jones, RN, PhD

EMR Tutor: Specifications for the Use of the Elecronic Medical Record Environment for Teaching and Learning 1056
Primary Author: V.J. Joshi, MBBS

A New Method for "User-centered" Indexing of Documents in Information Retrieval 1057
Primary Author: Y. Kagolovsky, MD

Cleansing and Geocoding Spatial Data for an Academic Medical Center 1058
Primary Author: J. Kamal, PhD

Use and Evaluation of MedWeaver by Medical Students in a Clinical Setting 1059
Primary Author: K. Kawamoto

A Stake in HL7 1060
Primary Author: S. Kay, PhD

Barriers to Adolescent Ability to Evaluate the Accuracy of Information about HIV 1061
Primary Author: A. Keselman, Ph.D.

Constructing a Non-human Subset in SNOMED® CT 1062
Primary Author: J.T. Kilbourne, MD

Applying Social Network Analysis Techniques to Measure the Efficacy of Computerized Medication Records 1063
Primary Author: J.Y. Kim, PhD

A Web Delivered Research Laboratory Information System for Pharmacokinetic Trials 1064
Primary Author: T. Grose

Building a Human Brain Atlas Optimized for Normal Adult Anatomy 1065
Primary Author: P. Kochunov, PhD

Datavant Exhibit 1009

DAERI – A Case Database for Adverse Events Related to the Internet 1066
Primary Author: C. Kohler

Searching for New Patterns in Medical Data with Evolutionary Programming 1067
Primary Author: P. Kokol

Evidence-based Automated Preventive Care Scheduling 1068
Primary Author: S. Krishna, PhD, MBA

Cognitive Models of HIV Decision-making in Young Adults 1069
Primary Author: T.T. Kubose, PhD

Integration of Clinical and Administrative Information in Medical Treatment 1070
Primary Author: E. Kuhn, ProfDr

A Model for the Use of Information Technology in Oral Anticoagulation Management 1071
Primary Author: G.J. Kuperman, MD, PhD

Evidence Based Order Sets as a Nursing Care Planning System 1072
Primary Author: L.M. LaCrosse, MSN, RN

DXplain Evoking Strength – Clinician Interpretation and Consistency 1073
Primary Author: T.A. Lasko, MD

Virtual Patient/Eye Simulator – Evolution of a Futuristic Teaching Tool 1074
Primary Author: R. Lasslo, MS, MD

A Video-streaming Platform for e-Learning 1075
Primary Author: P. Le Beux

Assessment of a French Code of Ethics for Health Teaching Resources on the Internet 1076
Primary Author: F.F. Le Duff, MD, MPH

Online Discharge Notes Increase Utilization of Computerized Medication Lists 1077
Primary Author: J. Lee, MD

A CORBA-based Medical Decision Support System for Medication Error 1078
Primary Author: Y. Lee, PhD

Patient-focused Wireless Messages for Diabetes 1079
Primary Author: M.G. Leu, MD, MS

Efficient Therapeutic Guideline Production Across Multiple Output Formats 1080
Primary Author: B. Lewis, PhD

Iterative Development of a Web Application to Support Teleconferencing of a Distributed Tumor Board 1081
Primary Author: H. Li

Datavant Exhibit 1009

Defending Clinician Values: Quality-in-use of Decision Support Systems for Thoracic Surgery 1082
Primary Author: L. Lidman, MSc

Test-retest Reliability of Internet-based Self-assessment Program for Depression 1083
Primary Author: C.-C. Lin, MD

Mining of Correlated Rules in Genome Sequences 1084
Primary Author: L. Lin

Medical Data Mining - Experience of Knowledge Discovery in Two Clinical Databases 1085
Primary Author: C.C.H. Liu

Cache Table Design for Disease Surveillance System 1086
Primary Author: Z. Liu, MS

Collection of Information Directly from Patients through an Adaptive Human-computer Interface 1087
Primary Author: D.F. Lobach, MD, PhD, MS

Syndromic Surveillance for Bioterrorism Using Automated Collection of Computerized Discharge Diagnoses 1088
Primary Author: W.B. Lober, MD

Optical Report Archiving of the Medicolegal Electronic Medical Record for an Enterprise Clinical Information System 1089
Primary Author: M. LoBue

Feasibility Studies in IP Videoconferencing, Streaming, and Wireless Communication 1090
Primary Author: C. Locatis, PhD

BIKMAS-II: A Knowledge Management System for Biomedical Informatics 1091
Primary Author: V. López-Alonso

A Simple Bayesian Network for Tuberculosis Detection 1092
Primary Author: H. Ma, MD

Defining an Open Metadata Framework for Proteomics: The PROMIS Project 1093
Primary Author: W.J. MacMullen, MSIS

EPHIP: An Innovative Program for Health Informatics Education 1094
Primary Author: J.E. MacNeill

Senex II: Ordering of Molecular Mechanisms in Cancer 1095
Primary Author: V. Mah, MD

Who Needs Evidence? Pediatric Protocol Development Using a Multi-disciplinary Consensus Model 1096
Primary Author: C.G. Maloney, MD

Datavant Exhibit 1009

Mining an Administrative Pharmacy Database for Practice Patterns 1097
Primary Author: M.K. Martin, DVM, MPH

Application of an Administrative Inpatient Pharmacy Database for Analysis of Community Acquired Pneumonia Treatment 1098
Primary Author: M.K. Martin, DVM, MPH

A Database-based Application for Management and Statistical Analysis of High-throughput Gene Expression Data 1099
Primary Author: M. Masseroli, PhD

Robotic Assistance During Ambulation by Older Adults 1100
Primary Author: J. Matthews, PhD, RN

Interactive Voice Response Technology for Smoking Cessation Relapse Prevention 1101
Primary Author: A.M. McDaniel, DNS, RN

Improving Clinical Communication with a Virtual Whiteboard 1102
Primary Author: L.K. McKnight, MD

Using Patient Data to Rank Records of Literature Retrieval 1103
Primary Author: E.A. Mendonça, MD

Oral Pathology Report System: Improving an Existing Database System 1104
Primary Author: Y. Michea, MD

Development and Evaluation of Computer Assisted Learning Modules for Medical Students 1105
Primary Author: Y. Michea, MD

Modular Design of Health Education Interactive Multimedia 1106
Primary Author: Y. Michea, MD

Towards a Hospital Geographical Information System Dedicated to Nosocomial Infection Control 1107
Primary Author: R. Micol, MS

Speech Language Pathology Assessments On-line Project: Small Steps Toward Electronic Clinical Documentation in a Long Term Care Facility 1108
Primary Author: C. Miranda

From Phenotype to Genotype: Experiences in Navigating the Available Information Resources 1109
Primary Author: J.A. Mitchell, PhD

A Declarative Query Interface for Large Semantic Networks 1110
Primary Author: P. Mork, MS

Primary Care Physicians' Satisfaction with Their Methods for Tracking Abnormal Results and Their Attitudes Concerning Clinical Decision Support Systems 1111
Primary Author: H.J. Murff, MD

Datavant Exhibit 1009

Disease-specific Data Sheets in the Management of Chronic Conditions - Case Example: Diabetes 1112
Primary Author: S.P. Narus, PhD

The Procedure Hierarchy in SNOMED® CT: What's New and Why 1113
Primary Author: S. Nash, DO

Protocol Design Patterns: Domain-oriented Abstractions to Support the Authoring of Computer-executable Clinical Trials 1114
Primary Author: J.H. Nguyen

A Dynamic Imaging Database for 3-D Morphologic Analysis and Clinical Assessment in Diagnostic Radiology 1115
Primary Author: G. Niculescu

Health Information Retrieval Tool (HIRT) 1116
Primary Author: M.T. Nyun, MD, MPH, MS

Accomplishments and Obstacles of PKI Implementation Within a Small Family Practice 1117
Primary Author: T. O'Brien

Training in Medical Informatics in Northeastern Brazil 1118
Primary Author: L. Ohno-Machado, MD, PhD

Geocoding Patient Addresses for Biosurveillance 1119
Primary Author: K.L. Olson, PhD

Systematic Monitoring for a Result Reporting System at Muenster's University Hospitals 1120
Primary Author: N. Osada, Dr.

Focusing Feedback on Interpersonal Skills: Practice Makes Perfect. A Web-based Tutorial for Standardized Patients 1121
Primary Author: T. Owens

Maximizing Data Portability in Patient-controlled Longitudinal Medical Records 1122
Primary Author: M.B. Palchuk, MD

Minimizing Privacy Risks of Biosurveillance 1123
Primary Author: E.C. Pan, MD

Patient-controlled Pediatric Immunization Records 1124
Primary Author: E.C. Pan, MD

Drug Eruption Point of Care Reference System 1125
Primary Author: A. Papier, MD

A Pitfall in Accessing the Genbank Sequence Literature 1126
Primary Author: T.B. Patrick, PhD

Gerios: Recommending Drugs and Dosing for Elderly Patients 1127
Primary Author: J.F. Peterson, MD

Datavant Exhibit 1009

From Documenting for Dollars to Collecting for Quality 1128
Primary Author: J. Phillips, RN

Bayesian Classification of Respiratory Disease and Asthma with Administrative Data Sets 1129
Primary Author: K.T. Phillips, MSW

Rehabilitation Division Clinical Outcomes Web Application and Database 1130
Primary Author: L. Pierce

Ontological Analysis for the Unification of Biology 1131
Primary Author: D.M. Pisanelli

A Comprehensive Outpatient Results Manager with Decision Support: Design Considerations and Architecture 1132
Primary Author: E.G. Poon, MD

Neonatal Nutrition Information Practices 1133
Primary Author: P. Porcelli, MD, MS

An Enhanced Framework for Pattern Detection in Clinical Laboratory Data 1134
Primary Author: A.R. Post, MD

Distance Education in the Management of the Hospital Equipment Maintenance 1135
Primary Author: M.P. Ramos, MSc

User Satisfaction with the Automated Patient Assessment 1136
Primary Author: C.A. Reilly, PhD, RN

Defining Expected Daily Emergency Department Utilization Rates for Detection of Bioterrorist Attacks 1137
Primary Author: B.Y. Reis, PhD

Implementation of State Wide Immunization Registry 1138
Primary Author: D.J. Richards

The Semantic Web and the Integration of Health Data Resources 1139
Primary Author: R.L. Richesson, MS, MPH

Conversion of MeSH® SCRs to Concept Structure 1140
Primary Author: C. Ridley

From Jobs to Work: Scheduling the Right Bioinformatics Tools 1141
Primary Author: J.E. Ries, MS

Visual Representation of the Numeric Bioenvironmental Data Display of the International Space Station's Flight Surgeon Console 1142
Primary Author: S. Rinkus, RN

Integrating Medical Documentation into Provider Order Entry 1143
Primary Author: S.T. Rosenbloom, MD, MPH

Predicting Outcomes of Testing for Decision Support Algorithms 1144
Primary Author: S.T. Rosenbloom, MD, MPH

Datavant Exhibit 1009

Surveying Housestaff Opinions Regarding Clinical Decision Support 1145
Primary Author: S.T. Rosenbloom, MD, MPH

Reducing Errors and Improving Performance by Applying Usability Engineering Principles to a Web-based Prototype of an Integrative Medicine Reference Tool 1146
Primary Author: J.A.Rukab

Clinicians' Opinions about the Usefulness and Effectiveness of Support Systems for Evidence- and Preference-based Patient Care 1147
Primary Author: C.M. Ruland, RN, PhD

Exploring Delivery of SNOMED® CT Solutions for Surgical Pathology Reporting Applications 1148
Primary Author: E. Ryske, MBA

Arden Syntax 2.1: Ten Years of Standardized Knowledge Bases 1149
Primary Author: R.M. Sailors, PhD

Disaster Simulation in the Internet to Train Rescuing Team: Development, Impact Analysis and Performance Evaluation 1150
Primary Author: E. Salvador, RN, MSc

Implementation of Ordering Guidelines for Chest CT Scans in a Provider Order Entry System 1151
Primary Author: D.L. Sanders, MD

Talking about Medical Errors: The Void in Existing Controlled Terminologies 1152
Primary Author: W. Sangster, MD

Exploring SNOMED Using Phylogenetic Analysis Tools 1153
Primary Author: I.N.Sarkar, MPhil

Pharmacokinetic Simulation of Muscle Relaxants: Prediction of Onset Based on Potency 1154
Primary Author: T. Sawa, MD, MS

[Withdrawn by the author] 1155

MORPHOSAURUS: Crosslingual Medical Text Retrieval by Subword Indexing 1156
Primary Author: S. Schulz, MD

The Companion Project – A Wearable Computing Device for the Patient for Reducing Medical Errors and Improving Patient Care 1157
Primary Author: M. Scotch

An Approach to Automate and Individualize Interactive Decision Support for Patients 1158
Primary Author: G.C. Scott

Datavant Exhibit 1009

A Mis-recognized Medical Vocabulary Correction System for Speech-based Electronic Medical Record 1159
Primary Author: H.J. Seo, PhD

Knowledge Acquisition of Generic Queries for Information Retrieval 1160
Primary Author: Y.-H. Seol, MPhil

A Software Tool for Editing and Executing Knowledge Component-based Clinical Guidelines 1161
Primary Author: H. Shah, MBBS, MS

A Hybrid Framework for Representation and Use of Clinical Guidelines 1162
Primary Author: Y. Shahar, MD, PhD

An Application to Integrate the Logistical and Technical Aspects of Data Dictionary Support to Multiple Healthcare Systems 1163
Primary Author: S. Shakib, MPH

Use of Protégé-2000 to Encode Clinical Guidelines 1164
Primary Author: R.D. Shankar, MS

Genetics Education Materials (GEM) Database and Website 1165
Primary Author: K. Silvey, MA

Comprehensive Analysis of Clinical Information Flow in an Integrated Delivery Network 1166
Primary Author: J. Smith

Using Concept Markers to Find Genetics Content in a Medical School Curriculum 1167
Primary Author: J.D. Smithers

Health Consumer Information in a Quality-controlled Gateway 1168
Primary Author: L.F. Soualmia, MSc

Accelerating Improvement of Care for Acute Myocardial Infarction Patients through Improved Data Collection 1169
Primary Author: B. Speadbury, MBA

Design and Implementation Status of a System Supporting Quasi On-line Hospital Operational Cost-monitoring and Performance-oriented Financing 1170
Primary Author: B. Spyropoulos

Software Agents for Extracting, Aggregating, and Updating Data from Web Pages of Genomic Databanks 1171
Primary Author: A. Stella

Design of a Web-based Care Team Scheduler for PalmCIS 1172
Primary Author: P.D. Stetson, MD

Physician Training: Changing Approaches for Maximizing Content Delivery 1173
Primary Author: J.S. Stone, RN

Path to Obtaining Health Insurance Information from the Internet: A Preliminary Observational Study 1174
Primary Author: K.C. Su, MLIS

Datavant Exhibit 1009

Evaluation of the Diagnostic Accuracy of Chest X-rays Acquired Using a Digital Camera for Low-cost Teleradiology 1175
Primary Author: A. Szot, MD, MS

Development and Implementation of Bed-side Network Computer for Providing Various Information to Inpatients 1176
Primary Author: S. Takamatsu

Implementing Computerized Decision Support for Ordering Titratable Drugs in a Pediatric Critical Care Unit 1177
Primary Author: D.A. Talbert, PhD

Mining the Epidemiological Literature for Risk Factors and Incidence Data of Kawasaki Disease 1178
Primary Author: L. Tanaka, MD

Saving Lives in Real-time: A Pre-hospital Telementoring Case 1179
Primary Author: R.D. Tindall

Development of a Wireless Web-based Infrastructure to Support Collection of Patient Self-reported QoL and Symptom Information in a Clinical Setting 1180
Primary Author: L.J. Trigg, MN, RN

A Typology for Modeling Processes in Clinical Guidelines and Protocols 1181
Primary Author: S.W. Tu, MS

Early Clinical Decisions for Rejection Therapy in Liver Transplantation 1182
Primary Author: G. Tusch, PhD

Education and Health Promotion in Remote Area Using a Video Conference System, a Satellite, and a Cable Television Network 1183
Primary Author: T. Utsugi, MD

Choosing Sides. Assigning Laterality as an Attribute in SNOMED® CT 1184
Primary Author: M.M. van Berkum, MD

Alzheimer's Caregiver Internet Support System (ACISS): Evaluating the Feasibility and Effectiveness of Supporting Family Caregivers Virtually 1185
Primary Author: L. Vehvilainen, MPH

Individual Preferences and On-line Public Health Services 1186
Primary Author: V. Vimarlund, PhD

St-Guide: A State/Transition Representation Model for Clinical Guidelines 1187
Primary Author: J. Wainer, PhD

Assessing Physiologic Data Representations for Anesthesia Record-keeping 1188
Primary Author: L.R. Waitman, PhD

Datavant Exhibit 1009

Designing Enterprise-wide Tools to Improve User Accountability for Protected Health Information: a Pilot Study 1189
Primary Author: J.S. Wald, MD, MPH

Controlled Software Evaluation Methodology: Case Simulation and Cognitive User Analysis 1190
Primary Author: K. Walker, MEE, MD

A History of Health Informatics in the Texas Medical Center 1191
Primary Author: K. Walker, MEE, MD

The Grouping of Roles in SNOMED Clinical Terms 1192
Primary Author: A.Y. Wang, MD

The Application of Wireless LAN Technology to Personal Digital Assistants to Access Web-based Clinical Applications 1193
Primary Author: J.K. Wang, MD

Expert System for Determination of Eligibility for Health Insurance Programs 1194
Primary Author: L. Wang, MD

Information Systems Developments to Detect and Analyze Chemotherapy-associated Adverse Drug Events 1195
Primary Author: M.G. Weiner, MD

A Palm OS Peripheral Brain Program for Use by Clinicians 1196
Primary Author: J.M. Weinfeld, MD

Existing Medical Taxonomies Poorly Code Outcome for an Acute Health Care Event 1197
Primary Author: A. Weltge, MD, MPH

Knowledge Discovery Using the Electronic Medical Record 1198
Primary Author: A. Wilcox, PhD

Analyzing Knowledge Base Content Development and Review: Recommendations for a Robust Knowledge Management Infrastructure 1199
Primary Author: S.G. Wilkinson, MS PT

PDA Use among Medical Students and Residents 1200
Primary Author: P. Winkelstein, MD

Use of a Web-based, Patient-oriented, Clinical Practice Guideline for Asthma 1201
Primary Author: C. Winslow, MD

Computerized Clinical Decision Support Systems Have the Potential to Detect Drug-lab Interactions in Outpatient Clinics 1202
Primary Author: J. Witter, MD

A Web-based Dynamic Clinical Informatics Page 1203
Primary Author: C. Wu, MS

Datavant Exhibit 1009

A Bayesian Network Implementation of Differential Diagnosis of Oral Soft Tissue Enlargements 1204
Primary Author: Y. Xing, MD

Developing Quality and Performance Criteria for a Clinical Document Imaging Project 1205
Primary Author: T.R. Yackel, MD, MPH

Redesign of Pedigree Displays through User-centered Visualization 1206
Primary Author: L. Yang

Improving Access to TB Medical Records in Remote Clinics in Peru Using a Personal Digital Assistant Based Application 1207
Primary Author: Z. Yasin

Comparison of Fingerprint and Iris Biometric Authentication for Control of Digital Signatures 1208
Primary Author: A.E. Zuckerman, MD

Theater-Style Demonstrations

Creating a Software Tool for the Clinical Researcher -- the IPS System 1210
Primary Author: B.G. Buchanan, PhD

An Enterprise Web Viewing System for Clinical and Administrative Data 1211
Primary Author: R.G. Duncan, MD

A Distributed Integration Approach to the Electronic Health Record 1212
Primary Author: J.E. Herrewynen

Stopsmoking.ucsf.edu: Web-based Randomized Trials of Smoking Cessation Interventions 1213
Primary Author: L. Lenert, MD, MS

Numerical Scoring of Performance on a Computer-based Patient Simulation 1214
Primary Author: J. Littlefield, Ph.D.

My Healthe*V*et 1215
Primary Author: D. Maloney

The Regenstrief Medical Record System 2002: Focus on the Medical Gopher Clinical Workstation 1216
Primary Author: C.J. McDonald, MD

Datavant Exhibit 1009

This material may be protected by Copyright law (Title 17 U.S. Code)

A Security Architecture for Query Tools used to Access Large Biomedical Databases

Shawn N. Murphy, MD, Ph.D. and Henry C. Chueh, MD, M.S.
Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA.

Disseminating information from large biomedical databases can be crucial for research. Often this data will be patient-specific, and therefore require that the privacy of the patient be protected. In response to this requirement, HIPAA released regulations for the dissemination of patient data. In many cases, the regulations are so restrictive as to render data useless for many purposes. We propose in this paper a model for obfuscation of data when served to a client application, that will make it extremely unlikely that an individual will be identified. At Partners Healthcare Inc, with over 1.4 million patients and 400 research clinician users, we implemented this model. Based on the results, we believe that a web-client could be made generally available using the proposed data obfuscation scheme that could allow general usage of large biomedical databases of patient information without risk to patient privacy.

INTRODUCTION

At the core of hospital sponsored clinical research programs are services for recruiting patients for clinical research studies. At Partners Healthcare Inc. in Boston, a corporation that includes both the Massachusetts General Hospital (MGH), and the Brigham and Women's Hospital (BWH), there has been created a clinical data warehouse named the Research Patient Data Registry (RPDR). This data warehouse contains all of the inpatient and outpatient data identified as desirable for research as defined by a previous study (1). This data includes diagnoses, medications, procedures, and laboratory tests. It contains about 270 million patient-concept associations from 26 million patient encounters encompassing about 1.4 million patients.

Faculty members at the MGH or BWH are able to find cohorts for their research studies through a visual interface (query tool) supported from any networked Microsoft Windows client workstation in the hospitals. The model for the user interface consists of a hierarchical tree of items for the users to choose from, where items are dragged into a set of containers that roughly model a Venn diagram. Aggregate numbers of patients who match the query criteria are returned in a series of display widgets. The individual patients that represent the aggregates can later be identified, provided the Institutional Review Board has approved the study request.

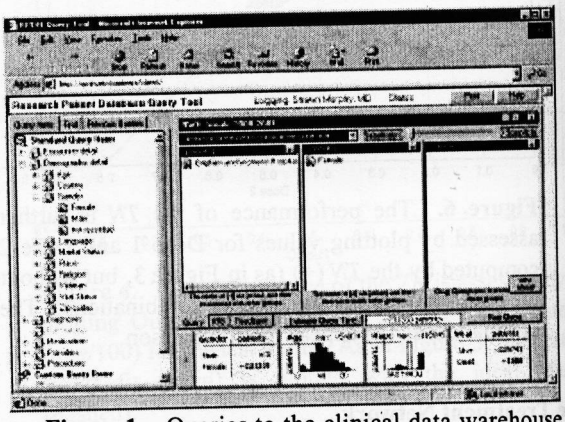


Figure 1 – Queries to the clinical data warehouse are made through this web query tool, where items arranged hierarchically in the left panel are dragged into the group panels on the right to construct a Boolean query. This query is asking for the number of BWH patients who are also female. Results are shown in the lower right displays. Approximate numbers are returned to prevent the identification of specific patients using the query tool.

A fairly general user base has evolved, as many faculty members did not want to use the tool themselves, but rather wished to delegate that task to a fellow or research associate. Another security concern is that traditional audit logs have a difficult time proving that a person is searching for a specific individual with the query tool, making it more necessary to prevent such an attack at the point of the lookup. These two factors heightened concerns about patient confidentiality using a general query tool, and it became imperative that a single individual could not be identified through the user interface.

An attack on the system would occur by someone entering criteria into the query tool that would result in only one matching person. For example, a person wishes to know if someone in their neighborhood has manic-depression, and that person knows the birthday of the son of this woman. They might start by asking the query tool for how many 30-year-old females in the 02054 area code had a normal

Datavant Exhibit 1009

delivery on September 3, 1988. If this resulted in only one matching person, then they would ask how many 30-year-old females in the 02054 area code had a normal delivery on September 3, 1988, plus have manic-depression. If this also resulted in one matching person then they would know that the person in their neighborhood had manic-depression.

An approach to preventing the identification of individuals in the database could be to de-identify patients in the database with a system that complies with HIPAA's 18 points for data de-identification (2). Although this would be a good start, there are two sets of problems with this. The first set of problems results from the destruction of important clinical data. This results mostly from the destruction of location information (zip codes) and the destruction of date information. Just attempting to recruit patients in the general area of the hospital for research studies relies on zip codes, not to mention more complex epidemiological studies. The destruction of date information is equally problematic. The HIPAA regulations state that dates should be rounded to the month if not to the year. However, very often a query will want to identify which event occurred first as a means of finding cause and effect. For example, in a study on drug adverse events, one may want to find all cases where a rise in liver function tests followed the initiation of a drug by 1-3 days. The second set of problems is that even with the destruction of all the above information, there is still no guarantee that someone cannot be identified in the database. As greater quantities of data are added, the chance that a set of clinical values (such as an SGOT of 367 and an SGPT of 1325) will not be able to uniquely define a patient becomes very low.

Another approach might be to use table ambiguation by suppression of selected cells (3). However, this approach would also result in the destruction of large amounts of clinical data. Because cells that uniquely distinguish a patient from other patients are the cells to be suppressed, the most interesting clinical data would be specifically targeted for destruction. This would severely limit the usefulness of the data for many purposes, such as has been extensively analyzed by Ohno-Machado et. al. (4). Furthermore, this approach is also susceptible to re-identification of patients that have supposedly been de-identified (5).

Other approaches include allowing the patient to look at those who have accessed their records, possible spotting their neighbor from the above example. There is a high maintenance cost to such a system (issuing passwords, etc.), but even this heroic solution will not work with query tools which touch thousands if not millions of different patient records with every query.

Our solution to thwarting an attack on this kind of system was to take advantage of the following set of properties of the query tool, and inventing a solution that fit around these properties. Although the fact that these properties must exist in the query tool limits the usefulness of the solution, these are properties many query tools would have (on the web for example) where aggregate numbers of patients need to be returned. The required properties are:

1) It is sufficient that the query tool only returns numbers of distinct patients.

2) The number of distinct patients is useful even when the number is within a probable range of 2 or 3 patients. This range can be varied to obtain optimized signal to noise ratios for detecting an attack.

The requirement that the query tool return distinct patients and not some other item such as hospital encounters or events depends on the nature of the item. One may generalize and conclude that any set of independent items is a candidate to be used as the reported item. However, there are some kinds of relationships between reported items (such as if the same hospitalization is reported as multiple encounters in the database) that could be exploited to attack our data obfuscation system.

METHODS

We developed a method of data obfuscation that was generally applicable to query tools that meet the above requirements. Additionally the server application should satisfy the following technical requirements:

1) The server application that will produce the numbers for the query tool must allow the numbers to be changed in a programmatic fashion before they are sent to the query tool. An obfuscator is going to be placed between the server output and the query tool, so these numbers need to be assessable.

2) An audit trail with a minimum of user ID, date/time of query, and true number of patients returned by the queries needs to be kept. A program is going to continuously analyze this audit trail.

In an abstract sense, the key to the data obfuscation method is to make attempts to discern individuals in the database a recognizable pattern by a computer algorithm. The method works by having the server not pass true numbers of patients to the query tool, but rather numbers that would be (if the

Datavant Exhibit 1009

same query were performed repeatedly) distributed around the true number of patients. For example, if one queried to find out how many people carried the diagnosis of manic-depression at Partners Healthcare Inc., one might get the result of 3,487, 3,488, 3,488, 3,490, or 3,489 patients. The exact nature of this range is the subject of much discussion below, but the range would be quasi-centered on the true number of 3,489. In order to find the true number, one would need to run the same query repeatedly. It is here where a computer algorithm would recognize this attack and mark the user whose queries returned the same true (not reported) number repeatedly as a suspicious pattern.

Forcing a user to do the same query multiple times if they are trying to obtain the true number creates a detectable pattern. This is because the chance that the same number of patients will be returned from many different queries of a large database is low, except perhaps for queries that return zero patients. The mean number of attempts required to hone in on the true number will increase as the range gets larger. This suggests that the range should be large, which, however, needs to be balanced with user dissatisfaction of receiving a wide range of approximate numbers. If one studies the functions available for creating the range, one finds the optimal function is one derived from a Gaussian one as discussed in figure 2. The reason that this function is optimal is essentially that specific features of any other distributions could be exploited to allow fewer attempts for finding the true number.

As the numbers that are being returned to the query tool are obfuscated, a history table of true numbers is maintained. This history table is seen only by the system, and contains for each query the date/time, the user ID, and the number of patients truly returned by the queries. A trigger each time a query is performed queries this history table. The query looks at the numeric results of queries performed by that user to see if the exact same number of patients were returned in previous queries within a specific time period. Depending on the time range employed for obfuscation, when a certain number of queries return the exact same number of patients, the user is prevented from performing further queries. Eventually, the account is reviewed and the lock is released if it is determined that the same number of patients were returned by the queries purely by coincidence.

We did a Monte Carlo simulation (6) of an attack, where a user is repeatedly performing the same query to find the average number of patients in the repeats. When the range of the Gaussian function that randomizes the results has a standard deviation of 1.33 (giving values about the true number as shown in figure 2), we found that it would take an average of 12.3 repeats before the average of the results converged on the true number of patients. Restated, if one performs the same query repeatedly, then a series of numbers, representing the obfuscated numbers of patients, are going to be returned. If one takes the running average of these numbers, they will eventually converge on the true number. The running average will become less that 0.5 patients from the true number of patients in a consistent fashion after about 12 numbers are returned. This estimate may actually be somewhat low because it would take a few more returned numbers before the attacker was confident that convergence had indeed been obtained. The threshold of tolerance for queries returning the same number of patients depends on the risk one wants to take that a patient will be identified in the database, vs. the inconvenience to the user and hassle to the auditors of access denial based on the same numbers returned coincidentally.

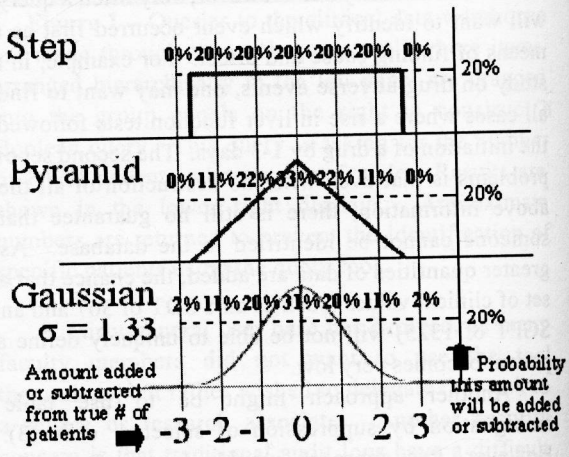


Figure 2 – An obscuration function is used to create small values that will be added or subtracted from the true number of patients to prevent patient identification. We found that functions derived from the Gaussian function were the least vulnerable to attack. Other examples of functions include the Step function and the Pyramid function. The Step function has a 4% chance of being cracked in just 2 tries (when #patients+2 and #patients–2 are obtained). The Pyramid function has a 1.2% chance of being cracked in 2 tries, but almost 6% in 4 tries (calculated via binomial probability function). A Gaussian derived function with roughly the same range of values cannot be definitively cracked in any finite number of tries.

Datavant Exhibit 1009

RESULTS

Using our query tool as an example, we investigated how large, biomedical databases could be set up to be explored for aggregate patient numbers without allowing the identity of an individual to be revealed. We do not mean to indicate that an interesting finding will not boil down to a single patient. Rather, a detailed investigation of a single patient without potentially revealing the identity of that person would be an almost impossible task. Cases that require this kind of detailed investigation of a very small domain of patients are not candidates for study through our query tool, although our query tool may give approximate numbers of patients to determine if enough patients exist to fulfill a study designed for those patients.

Given that we are interested in investigating large patient samples and trends, we studied various methods to obfuscate the results. Our initial approach was to de-identify all of the patients in the database, removing obvious identifiers. This will prevent a direct attack, where someone breaks into the database itself to perform a quick lookup of a patient. However, if someone had more time they could use gender, date of birth, and zip code to identify many patients (7). This kind of attack would even be possible through the query tool client application without direct access to the database. Our next approach was to blur some of our demographic variables, such as convert birth dates to ages, and present any queries that returned 2 or fewer patients as < 3, in theory trying to enforce a bin size of 2. This approach suffered from the weakness that a clever person could create a "step" of patients from a previous query which would return a constant number of patients, 10 for example. This step could then be logically OR'd with criteria of the patient to be identified. Then, to find if our example person had manic-depression, the count would go from 10 to 11. If one implemented some kind of rounded output, an attacker would just need to determine the position and direction of the rounding and put a step exactly where the rounding occurred.

Our final solution was randomizing the output within a certain range using a Gaussian function. This would force any person to statistically search for the mean output if they wanted to find the true answer. By forcing an attacker to perform a query repeatedly in hopes that an average can be ascertained, the system will see the query returning the same number of patients repeatedly. When queries are seen returning the same number of patients repeatedly it removes the access of the user to the system.

Although this method may seem overly complex in some ways, the task of obfuscating data in a foolproof manner is extremely difficult. The method we have described also has some vulnerabilities that need to be explored. One vulnerability is that we allow a larger number of queries to be performed that return zero patients. Unfortunately, naï ve users who are performing incorrect queries will often have zero patients returned. Having the table-watching agent terminate access to these users would be too sensitive a response. However, attacks around zero are more likely to succeed. Therefore we modify our range algorithm to have a wider obfuscation at values close to zero, and indicate the greater range by returning "<3" when the obfuscated number is less that 3.

Another vulnerability is that prior knowledge of data in the database may allow "steps" to be pre-conceived, and thus queries that do not return the same number of patients could be used to get the true mean from several queries. For example, if one wished to know the manic-depressive status of the 30 year old female in the 02054 area code, and one knew the manic-depressive status of 30 year old females in numerous other area codes, one could add each of these area codes to several queries. Thus if one knew that the 02055 area code had 2 manic-depressives, and the 02056 area code had 10 manic-depressives, the result of 3 and 12 when each query was performed with the 02054 area code added, might suggest that 1 manic-depressive lived in the 02054 area code (obtained once the known numbers were subtracted out). Since the queries returned different numbers of patients, the agent watching the numbers of patients returned by the queries would not catch this attack. This type of attack would require that one had extensive specific knowledge of numbers of patients in other areas regarding specifically what one wanted to discern in a given individual. Our environment at Partners Healthcare Inc. makes this an unlikely scenario. On the other hand, this may be an important scenario if the query tool were to be used on a nationwide database. Someone may know all the manic-depressives in Florida, and use these to construct "steps" to illicitly find the manic-depressive status of someone in Massachusetts. To counter such an attack, one would implement a rounding function on the results, such as rounding to the nearest 10. An attack would need to operate at the margin of the rounding, and the agent would be look for a pattern of numbers of patients returned at the margins of the rounding.

Datavant Exhibit 1009

Finally, there is a potential vulnerability concerning breakdowns of patients into different categories by the query tool, such as by males and females. Our strategy for addressing this vulnerability is to add the obfuscating value for the total patient number across random breakdown numbers, and then to obfuscate the breakdown numbers by the same Gaussian function used to obfuscate the total. This approach will prevent the breakdown totals from being used to further converge on the true number of total patients. However, the effect will be to produce less and less reliable numbers as one drills into increasing detail in the breakdown presentation.

DISCUSSION

The objective of our data obfuscation is to allow a general population of users to have access to aggregate data about patients without threatening the confidentiality of the individual patient. It has been repeatedly shown that de-identified data can often be combined with other publicly available data to identify the individual (8). Attempts to obfuscate the raw data in a database can result in a deterioration of the potential knowledge that might be extracted from the database (4). As databases contain greater amounts of specific clinical information about specific patients it would become necessary to increasingly alter the clinical data to allow the patients to fit into larger "bins".

Hospital data consortiums have long been available in the form of de-identified data dumps from hospital admission records and the like. However, it has been shown that most of these data sets are prone to re-identification and therefore violate our concept of privacy (7,8). An alternative to large data dumps is to allow such data to be requested from server-side services that are more easily amenable to data obfuscation and monitoring. Designed correctly, these services could greatly decrease the risk of divulging private information.

Acceptable methods of obfuscation will be dependent on the types of data that need to be served back to the client application. The methods described in this paper apply to clients that need sets of distinct patients where exact numbers of patients are not required.

Naï ve approaches to data obfuscation are often to restrict the types of data elements that may be used in a query. Our example of looking for manic-depressives in certain zip codes may spur a naï ve response to restrict the use of zip codes. But this is a "slippery slope", because any data can potentially be used in combination with other data to identify patients, such as child delivery dates, hospital admission dates (8), lab values, etc. The richer the database is with data, the more problematic this becomes.

The data obfuscation method discussed in this paper does not limit the logic of the query, nor the data elements or values that can be used in the query. It depends exclusively on the numeric quantity of the results. In our environment, this data obfuscation allows general research queries to be performed with relatively little risk to patient confidentiality. This allows a greater domain of users and a greater comfort level from our sponsors. This model allows sophisticated data to be stored intact in the database such as gene sequences and bio-markers of various kinds. Without this obfuscation method, there would be steep reservations about including increasingly patient-specific data in large databases for the MGH and BWH research community.

This work was supported by Partners Healthcare Inc.

References

1. Murphy, S. N., Morgan, M. M., Barnett, G. O., Chueh, H. C. Optimizing Healthcare Research Data Warehouse Design through Past COSTAR Query Analysis. Proc AMIA Fall Symp. 1999; 892-6.
2. Rules and Regulations. Federal Register 65(250), Section 164.514, 82818, Dec 28, 2000.
3. Fischetti M,Salazar J.Model and algorithms for the 2-dimensional cell suppression problem in statistical disclosure control. Mathematical Programming 1999; 84:283-312.
4. Ohno-Machado, L., Dreiseitl,S., Vinterbo, S., Effects of Data Anonymization by Cell Suppression on Descriptive Statistics and Predictive Modeling Performance, Proc AMIA Fall Symp 2001; 503-7.
5. Dreiseitl,S., Vinterbo, S., Ohno-Machado, L., Disambiguation Data: Extracting Information from Anonymized Sources, Proc AMIA Fall Symp 2001; 144-8.
6. J. E. Gentle, Random Number Generation and Monte Carlo Methods (Statistics and Computing), Springer-Verlag, 1998
7. Sweeney L. Guaranteeing anonymity when sharing medical data, the DataFly system. Proc. AMIA Fall Symposium. 1997; 51-5.
8. Malin, B., and Sweeney, L. Re-Identification of DNA through an Automated Linkage Process, Proc AMIA Fall Symp 2001; 423-7.

Datavant Exhibit 1009