From Implementation to Analytics: The Future Work of Informatics

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: informaticsprofessor.blogspot.com

References

Adams, J. and J. Klein (2011). Business Intelligence and Analytics in Health Care - A Primer. Washington, DC, The Advisory Board Company http://www.advisory.com/Research/IT-Strategy-Council/Research-Notes/2011/Business-Intelligence-and-Analytics-in-Health-Care.

Alschuler, L., B. Beighe, R. Dolin and J. Kuhl (2012). Health Story Project http://www.healthstory.com/pdf/HIMSS_EHRA_Final.pdf.

Bedrick, S., T. Edinger, A. Cohen and W. Hersh (2012). Identifying patients for clinical studies from electronic health records: TREC 2012 Medical Records Track at OHSU. The Twenty-First Text REtrieval Conference Proceedings (TREC 2012), Gaithersburg, MD, National Institute for Standards and Technology.

Berlin, J. and P. Stang (2011). Clinical Data Sets That Need to Be Mined. Learning What Works: Infrastructure Required for Comparative Effectiveness Research. L. Olsen, C. Grossman and J. McGinnis. Washington, DC, National Academies Press**:** 104-114.

Blumenthal, D. (2011). "Wiring the health system--origins and provisions of a new federal program." New England Journal of Medicine **365**: 2323-2329.

Bourgeois, F., K. Olson and K. Mandl (2010). "Patients treated at multiple acute health care facilities: quantifying information fragmentation." Archives of Internal Medicine **170**: 1989-1995.

Boyd, D. and K. Crawford (2011). Six Provocations for Big Data. Cambridge, MA, Microsoft Research http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431.

Buntin, M., M. Burke, M. Hoaglin and D. Blumenthal (2011). "The benefits of health information technology: a review of the recent literature shows predominantly positive results." Health Affairs **30**: 464-471.

Chapman, W., P. Nadkarni, L. Hirschman, L. D'Avolio, G. Savova and O. Uzuner (2011). "Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions." Journal of the American Medical Informatics Association **18**: 540-543.

Chaudhry, B., J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S. Morton and P. Shekelle (2006). "Systematic review: impact of health information technology on

quality, efficiency, and costs of medical care." Annals of Internal Medicine **144**: 742-752.
Cimino, J. (2012). "The false security of blind dates: chrononymization's lack of impact on data privacy of laboratory data." Applied Clinical Informatics **3**(4): 392-403.
Davenport, T. (2006). "Competing on Analytics." Harvard Business Review http://hbr.org/product/competing-on-analytics/an/R0601H-PDF-ENG.
Davenport, T. (2012). Analytical integration in healthcare. Analytics**:** 19-23 http://www.analytics-magazine.org/januaryfebruary-2012/504-analytical-integration-in-healthcare.
Davenport, T., J. Harris and R. Morison (2010). Analytics at Work: Smarter Decisions, Better Results. Cambridge, MA, Harvard Business Review Press.
Demner-Fushman, D., S. Abhyankar, A. Jimeno-Yepes, R. Loane, F. Lang, J. Mork, N. Ide and A. Aronson (2012). NLM at TREC 2012 Medical Records Track. The Twenty-First Text REtrieval Conference Proceedings (TREC 2012), Gaithersburg, MD, National Institute for Standards and Technology.
Denny, J., M. Ritchie, M. Basford, J. Pulley, L. Bastarache, K. Brown-Gentry, D. Wang, D. Masys, D. Roden and D. Crawford (2010). "PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations." Bioinformatics **26**: 1205-1210.
Diamond, C., F. Mostashari and C. Shirky (2009). "Collecting and sharing data for population health: a new paradigm." Health Affairs **28**: 454-466.
Finnell, J., J. Overhage and S. Grannis (2011). All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. AMIA Annual Symposium Proceedings, Washington, DC.
Goldzweig, C., A. Towfigh, M. Maglione and P. Shekelle (2009). "Costs and benefits of health information technology: new trends from the literature." Health Affairs **28**: w282-w293.
Hersh, W. (2010). "The health information technology workforce: estimations of demands and a framework for requirements." Applied Clinical Informatics **1**: 197-212.
Horner, P. and A. Basu (2012). Analytics & the future of healthcare. Analytics**:** 11-18 http://www.analytics-magazine.org/januaryfebruary-2012/503-analytics-a-the-future-of-healthcare.
Hripcsak, G. and D. Albers (2012). "Next-generation phenotyping of electronic health records." Journal of the American Medical Informatics Association: Epub ahead of print.
Hsiao, C., E. Hing, T. Socey and B. Cai (2011). Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-based Physician Practices: United States, 2001–2011. Atlanta, GA, National Center for Health Statistics, Centers for Disease Control and Prevention http://www.cdc.gov/nchs/data/databriefs/db79.htm.
Jha, A., K. Joynt, E. Orav and A. Epstein (2012). "The long-term effect of Premier pay for performance on patient outcomes." New England Journal of Medicine: Epub ahead of print.

Kho, A., J. Pacheco, P. Peissig, L. Rasmussen, K. Newton, N. Weston, P. Crane, J. Pathak, C. Chute, S. Bielinski, I. Kullo, R. Li, T. Manolio, R. Chisholm and J. Denny (2011). "Electronic medical records for genetic research: results of the eMERGE Consortium." Science Translational Medicine **3**: 79re71 http://stm.sciencemag.org/content/3/79/79re1.short.
Lee, S. and L. Walter (2011). "Quality indicators for older adults: preventing unintended harms." Journal of the American Medical Association **306**: 1481-1482.
Lewis, M. (2004). Moneyball: The Art of Winning an Unfair Game. New York, NY, W. W. Norton & Company.
Longworth, D. (2011). "Accountable care organizations, the patient-centered medical home, and health care reform: what does it all mean?" Cleveland Clinic Journal of Medicine **78**: 571-589.
Manor-Shulman, O., J. Beyene, H. Frndova and C. Parshuram (2008). "Quantifying the volume of documented clinical information in critical illness." Journal of Critical Care **23**: 245-250.
Meystre, S., F. Friedlin, B. South, S. Shen and M. Samore (2010). "Automatic de-identification of textual documents in the electronic health record: a review of recent research." BMC Medical Research Methodology **10**: 70 http://www.biomedcentral.com/1471-2288/10/70.
Miller, K. (2011). Big Data Analytics in Biomedical Research. Biomedical Computation Review**:** 14-21 http://biomedicalcomputationreview.org/sites/default/files/w12_f1-big_data.pdf.
Mirnezami, R., J. Nicholson and A. Darzi (2012). "Preparing for precision medicine." New England Journal of Medicine **366**: 489-491.
O'Malley, K., K. Cook, M. Price, K. Wildes, J. Hurdle and C. Ashton (2005). "Measuring diagnoses: ICD code accuracy." Health Services Research **40**: 1620-1639.
Rebholz-Schuhmann, D., A. Oellrich and R. Hoehndorf (2012). "Text-mining solutions for biomedical research: enabling integrative biology." Nature Reviews Genetics **13**: 829-839.
Rhoads, J. and L. Ferrara (2012). Transforming Health Care Through Better Use of Data. Falls Church, VA, Computer Sciences Corp. http://www.csc.com/health_services/insights/80497-transforming_healthcare_through_better_use_of_data.
Rosenbloom, S., J. Denny, H. Xu, N. Lorenzi and W. S. K. Johnson (2011). "Data from clinical notes: a perspective on the tension between structure and flexible documentation." Journal of the American Medical Informatics Association **18**: 181-186.
Ruiz, E., V. Hristidis, C. Castillo, A. Gionis and A. Jaimes (2012). Correlating financial tme series with micro-blogging activity. Fifth ACM International Conference on Web Search & Data Mining, Seattle, WA http://www.cs.ucr.edu/~vagelis/publications/wsdm2012-microblog-financial.pdf.
Safran, C., M. Bloomrosen, W. Hammond, S. Labkoff, S. Markel-Fox, P. Tang and D. Detmer (2007). "Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper." Journal of the American Medical Informatics Association **14**: 1-9.

Salant, J. and L. Curtis (2012). Nate Silver-Led Statistics Men Crush Pundits in Election. Bloomberg http://www.bloomberg.com/news/2012-11-07/nate-silver-led-statistics-men-crush-pundits-in-election.html.
Scherer, M. (2012). Inside the Secret World of the Data Crunchers Who Helped Obama Win. Time http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/.
Schoen, C., R. Osborn, D. Squires, M. Doty, P. Rasmussen, R. Pierson and S. Applebaum (2012). "A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas." Health Affairs: Epub ahead of print.
Serumaga, B., D. Ross-Degnan, A. Avery, R. Elliott, S. Majumdar, F. Zhang and S. Soumerai (2011). "Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study." British Medical Journal **342**: d108 http://www.bmj.com/content/342/bmj.d108.
Shortliffe, E. (2011). "President's column: subspecialty certification in clinical informatics." Journal of the American Medical Informatics Association **18**: 890-891.
Smith, M., R. Saunders, L. Stuckhardt and J. McGinnis (2012). Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC, National Academies Press.
Steffes, T. (2012). Predictive analytics: saving lives and lowering medical bills. Analytics**:** 24-27 http://www.analytics-magazine.org/januaryfebruary-2012/505-predictive-analytics-saving-lives-and-lowering-medical-bills.
Ugander, J., B. Karrer, L. Backstrom and C. Marlow (2011). The Anatomy of the Facebook Social Graph. Palo Alto, CA, Facebook http://arxiv.org/abs/1111.4503.
Uzuner, O. (2009). "Recognizing obesity and comorbidities in sparse data." Journal of the American Medical Informatics Association **16**: 561-570.
Uzuner, O., A. Bodnari, S. Shen, T. Forbush, J. Pestian and B. South (2012). "Evaluating the state of the art in coreference resolution for electronic medical records." Journal of the American Medical Informatics Association: Epub ahead of print.
Uzuner, O., I. Goldstein, Y. Luo and I. Kohane (2008). "Identifying patient smoking status from medical discharge records." Journal of the American Medical Informatics Association **15**: 14-24.
Uzuner, O., Y. Luo and P. Szolovits (2007). "Evaluating the state-of-the-art in automatic de-identification." Journal of the American Medical Informatics Association **14**: 550-563.
Uzuner, O., I. Solti and E. Cadag (2010). "Extracting medication information from clinical text." Journal of the American Medical Informatics Association **17**: 514-518.
Uzuner, Ö., B. South, S. Shen and S. DuVall (2011). "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text." Journal of the American Medical Informatics Association **18**: 552-556.
Versel, N. (2012). AMIA: Why interoperability is 'taking so darn long'. Healthcare IT News http://www.healthcareitnews.com/news/amia-why-interoperability-taking-so-darn-long.

Voorhees, E. and W. Hersh (2012). Overview of the TREC 2012 Medical Records Track. The Twenty-First Text REtrieval Conference Proceedings (TREC 2012), Gaithersburg, MD, National Institute for Standards and Technology.
Voorhees, E. and R. Tong (2011). Overview of the TREC 2011 Medical Records Track. The Twentieth Text REtrieval Conference Proceedings (TREC 2011), Gaithersburg, MD, National Institute for Standards and Technology.
Wang, T., D. Dai, A. Hernandez, D. Bhatt, P. Heidenreich, G. Fonarow and E. Peterson (2011). "The importance of consistent, high-quality acute myocardial infarction and heart failure care results from the American Heart Association's Get with the Guidelines Program." Journal of the American College of Cardiology **58**: 637-644.
Weiner, M. (2011). Evidence Generation Using Data-Centric, Prospective, Outcomes Research Methodologies. San Francisco, CA, Presentation at AMIA Clinical Research Informatics Summit.
Wharam, J. and J. Weiner (2012). "The promise and peril of healthcare forecasting." American Journal of Managed Care **18**: e82-e85.

From Implementation to Analytics: The Future Work of Informatics

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: informaticsprofessor.blogspot.com

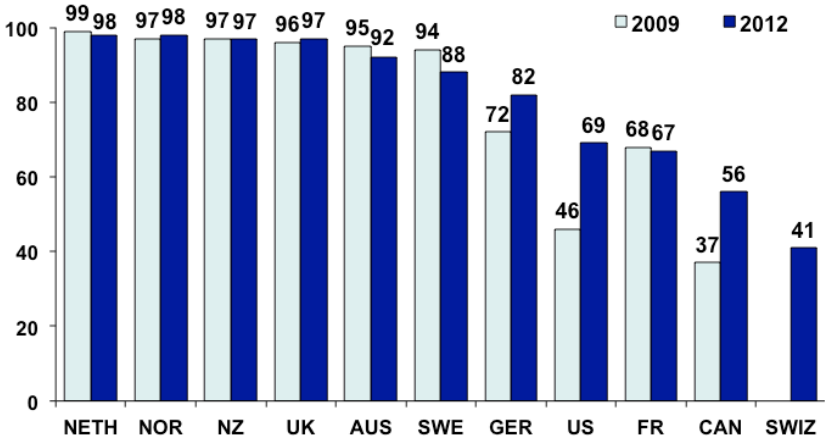
OREGON HEALTH & SCIENCE UNIVERSITY

Outline

- The world is making great strides in electronic health record (EHR) implementation
- "Will there be work for informatics after we are done implementing?"
- Toward analytics, business intelligence, and the learning health system
- Challenges in getting there

OREGON HEALTH & SCIENCE UNIVERSITY

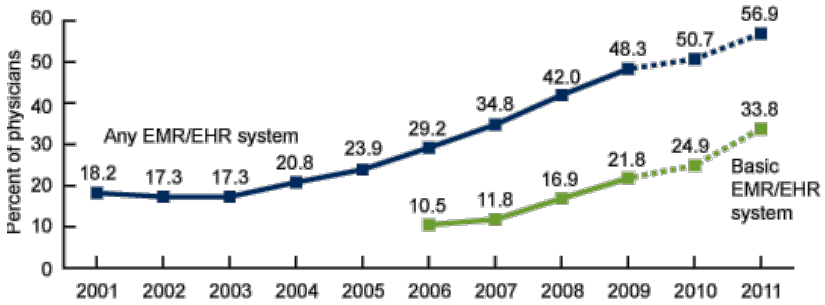
The world is adopting EHRs (Schoen, 2012)



3

OREGON HEALTH & SCIENCE UNIVERSITY

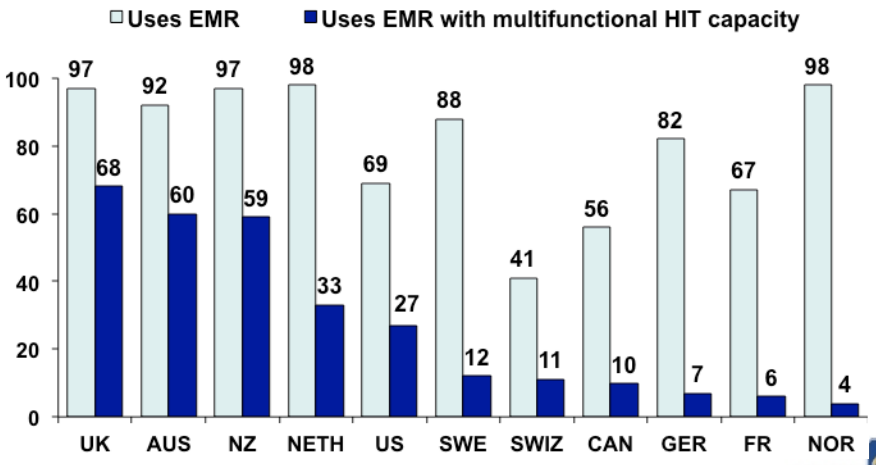
Even in the US (Hsaio, 2011)



4

OREGON HEALTH & SCIENCE UNIVERSITY

Although advanced functionality is less common (Schoen, 2012)



OREGON HEALTH & SCIENCE UNIVERSITY

5

US investment has been substantial

"To improve the quality of our health care while lowering its cost, we will make the immediate investments necessary to ensure that within five years, all of America's medical records are computerized … It just won't save billions of dollars and thousands of jobs – it will save lives by reducing the deadly but preventable medical errors that pervade our health care system."

January 5, 2009

Health Information Technology for Economic and Clinical Health (HITECH) Act of the American Recovery and Reinvestment Act (ARRA) (Blumenthal, 2011)

- Incentives for electronic health record (EHR) adoption by physicians and hospitals (up to $27B)
- Direct grants administered by federal agencies ($2B, including $118M for workforce development)

OREGON HEALTH & SCIENCE UNIVERSITY

6

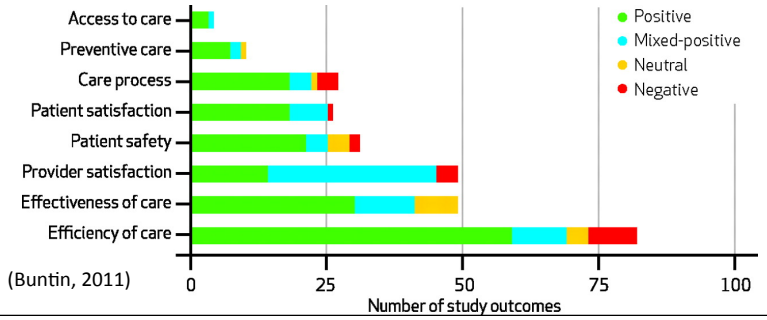
But the US healthcare system is still in need of fixing

- Recent IOM report (Smith, 2012) analyzes data to find annual
  - $750B in waste (out of $2.5T system)
  - 75,000 premature deaths
- Sources of waste
  - Unnecessary services provided
  - Services inefficiently delivered
  - Prices too high relative to costs
  - Excess administrative costs
  - Missed opportunities for prevention
  - Fraud

OREGON HEALTH & SCIENCE UNIVERSITY

Health information technology (HIT) is part of solution

- Systematic reviews (Chaudhry, 2006; Goldzweig, 2009; Buntin, 2011) have identified benefits in a variety of areas
  - Although 18-25% of studies come from a small number of 'health IT leader" institutions



(Buntin, 2011)

Components of the learning healthcare system (Smith, 2012)

- Records immediately updated and available for use by patients
- Care delivered the has been proven "reliable at the core and tailored at the margins"
- Patient and family needs and preferences are a central part of the decision process
- All healthcare team members are fully informed about each other's activities in real time
- Prices and total costs are fully transparent to all participants in the care process
- Incentives for payment are structured to "reward outcomes and value, not volume"
- Errors are promptly identified and corrected
- Outcomes are routinely captured and used for continuous improvement

9

OREGON HEALTH & SCIENCE UNIVERSITY

IOM schematic for the learning healthcare system (Smith, 2012)

10

OREGON HEALTH & SCIENCE UNIVERSITY

A not-uncommon dialogue with a student

- Student: "Will there be any informatics work left to do once everyone has implemented EHRs?"
- Teacher: "The real interesting work is yet to come!"

Toward analytics and business intelligence for healthcare

- Analytics is the use of data collection and analysis to optimize decision-making (Davenport, 2010)
- Business intelligence (BI) is the "processes and technologies used to obtain timely, valuable insights into business and clinical data" (Adams, 2011)
- As in many areas of advanced information and technology, healthcare and biomedicine are behind the curve of other industries (Miller, 2011; Rhoads, 2012)

Who employs analytics outside of healthcare?

- Amazon and Netflix recommend books and movies with great precision
- Many sports teams, such as the Oakland Athletics and New England Patriots, have used analytics to select players, plays, strategies, etc. (Lewis, 2004; Davenport, 2006)
- Facebook can target advertising very precisely knowing your friends, your interests, where you go, etc. (Ugander, 2011)
- Twitter volume and other linkages can predict stock market prices (Ruiz, 2012)
- Recent US election: re-election of President Obama (Scherer, 2012) and predictive ability of Nate Silver (Salant, 2012)

13

OREGON HEALTH & SCIENCE UNIVERSITY

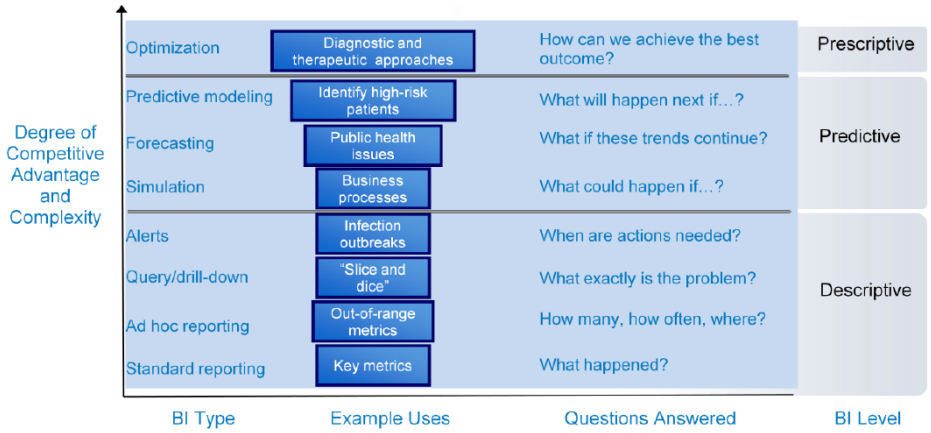
What about analytics in healthcare?

- With shift of payment from "volume to value," healthcare organizations will need to manage information better to provide better care (Diamond, 2009; Horner, 2012)
  - To realize this, they must achieve "analytic integration" starting with completing "implementing basic transaction data systems" (Davenport, 2012)
- Prediction not only of patient response but also behavior, e.g., regimen adherence (Steffes, 2012)
- A requirement of coming "precision medicine" (Mirnezami, 2012)
- HITECH investment and tools such as ONC QueryHealth may help
  - http://wiki.siframework.org/Query+Health

14

OREGON HEALTH & SCIENCE UNIVERSITY

Levels of BI (Adams, 2011)



Analytics and BI part of larger "secondary use" of clinical data

- Many "secondary uses" or re-uses of electronic health record (EHR) data, including (Safran, 2007)
  - Personal health records (PHRs)
  - Clinical and translational research – generating hypotheses and facilitating research
  - Health information exchange (HIE)
  - Public health surveillance for emerging threats
  - Healthcare quality measurement and improvement
- Successful demonstration that the phenotype in the EHR can be used with the genotype to replicate known associations as well as identify new ones, e.g., eMERGE (Kho, 2011; Denny, 2010)

Challenges for secondary use of clinical data

- EHR data does not automatically lead to knowledge
  - Data quality and accuracy is not a top priority for busy clinicians
- Little research, but problems identified
  - EHR data can be incorrect and incomplete, especially for longitudinal assessment (Berlin, 2011)
  - Much data is "locked" in text (Hripcsak, 2012)
  - Many steps in ICD-9 coding can lead to incorrectness or incompleteness (O'Malley, 2005)
- There are also important "provocations" about use of "big data" for research (Boyd, 2011)

17

OREGON HEALTH & SCIENCE UNIVERSITY

Challenges (cont.)

- Many data "idiosyncrasies" (Weiner, 2011)
  - "Left censoring": First instance of disease in record may not be when first manifested
  - "Right censoring": Data source may not cover long enough time interval
  - Data might not be captured from other clinical (other hospitals or health systems) or non-clinical (OTC drugs) settings
  - Bias in testing or treatment
  - Institutional or personal variation in practice or documentation styles
  - Inconsistent use of coding or standards

18

OREGON HEALTH & SCIENCE UNIVERSITY

Patients also get care at multiple sites

- Study of 3.7M patients in Massachusetts found 31% visited 2 or more hospitals over 5 years (57% of all visits) and 1% visited 5 or more hospitals (10% of all visits) (Bourgeois, 2010)
- Study of 2.8M emergency department (ED) patients in Indiana found 40% of patients had data at multiple institutions, with all 81 EDs sharing patients in a completely connected network (Finnell, 2011)

19

OREGON HEALTH & SCIENCE UNIVERSITY

Other challenges for analytics and BI in healthcare

- Volume of information can be challenging
  - Average pediatric ICU patient generates 1348 information items per 24 hours (Manor-Shulman, 2008)
- Forecasting has always been difficult in healthcare (Wharam, 2012)

20

OREGON HEALTH & SCIENCE UNIVERSITY

How to leverage analytics and BI to improve healthcare

- Data standards and interoperability
- Data entry
- Ease of extraction while maintaining privacy and security
- Quality measures that are associated with improved outcomes
- Healthcare system that pays for value over quantity
- Research based on use cases
- Development of highly competent workforce

21

OREGON HEALTH & SCIENCE UNIVERSITY

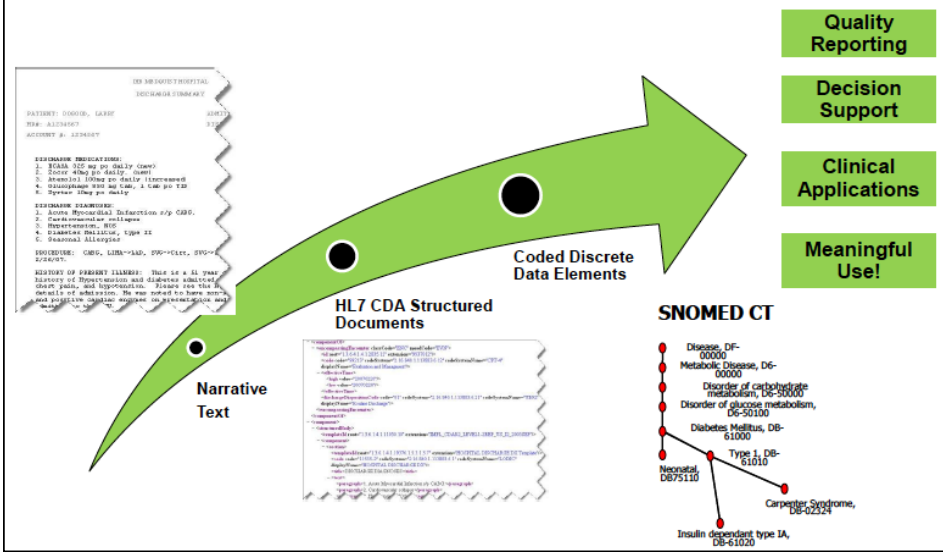
Data standards and interoperability

- Standards and interoperability are "taking too long" (Versel, 2012)
  - Despite slow pace, many standards are mature and ready for use
- HealthStory project (www.healthstory.com) is taking incremental approach in HL7 Consolidated CDA (Alschuler, 2012)
  - Creating standardized templates for various EHR document types now, structured data later

22

OREGON HEALTH & SCIENCE UNIVERSITY

HealthStory approach to Consolidated CDA



Data entry

- Clinical documentation is not a top priority for clinicians
- Anecdotally, many clinicians complain that EHR systems prioritize charge capture over clinical value
- Need to reconcile "tension" of structured vs. flexible documentation (Rosenbloom, 2011)

OREGON HEALTH & SCIENCE UNIVERSITY

Ease of extraction while maintaining privacy and security

- Most EHR systems offer little in way of tools for clinicians and others to make use of data
- One concern for doing this is protecting patient privacy and confidentiality
  - One option is to try to de-identify, but can be difficult, especially for narrative records (Uzuner, 2007; Meystre, 2010)
  - However, even lab data can be difficult to truly de-identify (Cimino, 2012)

25

OREGON HEALTH & SCIENCE UNIVERSITY

Quality measures that are associated with improved outcomes

- Quality measures increasingly used in US and elsewhere
- Use has been more for process than outcome measures (Lee, 2011)
- In UK, pay for performance schemes achieved early value but no further gains (Serumaga, 2011)
- In US, some quality measures found to lead to improved patient outcomes (e.g., Wang, 2011), others not (e.g., Jha, 2012)

26

OREGON HEALTH & SCIENCE UNIVERSITY

Healthcare system that pays for value over quantity

- US healthcare system still mostly based on fee for service model
- Affordable Care Act (ACA, aka Obamacare) implements accountable care organizations (ACOs), which provide bundled payments for conditions (Longworth, 2011)

27

OREGON HEALTH & SCIENCE UNIVERSITY

Need research based on use cases

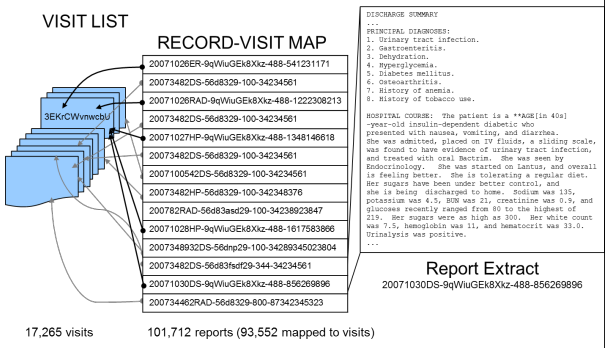
- Challenge evaluations provide common data sets and use cases
  - i2b2 – various natural language processing challenges (Uzuner, 2007-2012)
  - TREC Medical Records Track – information retrieval (IR) to identify potential patients for research (Voorhees, 2011-2012)
  - Text mining of scientific literature (Rebholz-Schuhmann, 2012)
- Research itself has challenges (Chapman, 2011)
  - Small and unrealistic data sets
  - Issues of de-identification and privacy of data sets

28

OREGON HEALTH & SCIENCE UNIVERSITY

TREC Medical Records Track – applying IR (search) to medical records

- Use case: identify patients who are possible candidates for clinical studies
- Uses test collection developed at Univ. of Pittsburgh
- Initial results show mediocre performance
  - Typical IR approaches do not work well
  - Best performance from manual Boolean queries (Demner-Fushman, 2012; Bedrick, 2012)



(Voorhees, 2011; Voorhees, 2012)

29

Need for a competent workforce

- Growing recognition of needs and competencies (Hersh, 2010)
- Coming certification, starting with physicians (Shortliffe, 2011)

- Back to student dialogue: "The work of informatics will transition from implementing systems and workflows to using data, information, and knowledge to improve individual health, healthcare, public health, and biomedical research"

30

Conclusions

- A growing body of evidence supports EHR and other IT to improve health and healthcare
- The world is making great strides in adopting EHRs
- The next step is to make use of the increasing data through analytics and BI
- There are challenges, but also benefits, to this use of data
- The work of informatics will become more interesting and important as we enter this new era

OREGON HEALTH & SCIENCE UNIVERSITY