AMIA 2014 Summit in CRI CRI Year-In-Review

Peter J. Embi, MD, MS, FACP, FACMI

Assoc Prof & Vice Chair, Dept of Biomedical Informatics
Associate Professor of Medicine
Chief Research Information Officer
Co-Director, Biomedical Informatics, CCTS
The Ohio State University

San Francisco, California
April 11, 2014

Disclosures

- Co-founder and consultant: Signet Accel LLC
- Consultant to various universities, research organizations

Approach to this presentation

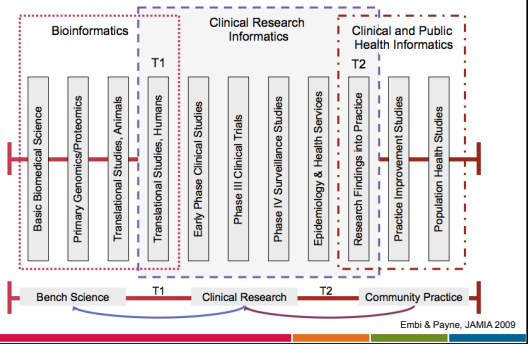
- Mixed approach to article identification:
  - Started with structured approach
    - (akin to ACP "update" sessions)
  - Augment with "what seemed interesting" approach
- Learned a lot from doing this last three years
  - Tracked manuscripts throughout the year
  - Intended to spread work out…
  - …still worked down to the wire
- So, what was my approach…

Source of Content for Session

- Literature review:
  - Initial search by MESH terms:
    - "Biomedical Research"[Mesh] AND "Informatics"[Mesh] AND "2013/01/01"[PDat] : "2014/02/01"[Pdat]
    - Resulted in **75** articles
  - Additional articles found via:
    - Recommendations from colleagues
    - Other keyword searches using terms like:
      - Clinical Trials, Clinical Research, Informatics, Translational, Data Warehouse, Research Registries, Recruitment
    - Yielding **349** more, from which…
    - **115** were CRI relevant
  - From those, I've selected **34** representative papers that I'll present here (*briefly*)

Clinical and Translational Research & Informatics: T1, T2, and Areas of Overlap for Informatics

Shaded CRI Region is Main Area of Focus



Session caveats

- What this is not…
  - A systematic review of the literature
  - An exhaustive review
- What this is…
  - My best attempt at *briefly* covering *some* of the representative CRI literature from the past year
  - A snap-shot of excellent CRI activity over past year
  - What I thought was particularly notable

Topics

- Grouped 34 articles into several CRI categories (admittedly, not *all* CRI areas)
  - Clinical Data Re-Use for Research
  - Data Sharing & Discovery
  - Methods in CRI *(new topic)*
  - Policy & Perspectives
  - Trends in CRI
- In each category, I'll highlight a few key articles and then given a quick "shout out" to a few others
- Conclude with notable events from the past year

Apologies up front

- I'm CERTAIN I've missed a lot of great work
- I'm REALLY SORRY about that

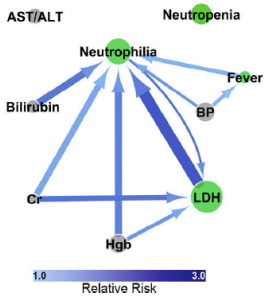
Clinical Data Re-Use for Research

Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. (Patel VN, Kaelber DC. J Biomed Inform. 2013)

- **Goal:** To test the use of aggregated and de-identified electronic health record (EHR) data for multivariate post-marketing pharmacosurveillance in a case study of azathioprine (AZA).
- **Methods:** Aggregated, standardized, normalized, and de-identified, population-level data from the Explore platform (Explorys, Inc.) for over 10 million individuals, of which 14,580 were prescribed AZA based on RxNorm. Based on logical observation identifiers names and codes (LOINC) and vital sign data, examined the following side effects: anemia, cell lysis, fever, hepatotoxicity, hypertension, nephrotoxicity, neutropenia, and neutrophilia. Patients prescribed AZA compared to patients on one of 11 other anti-rheumatologic drugs to determine the relative risk of side effect pairs.
- **Results:** Compared to AZA case report trends, hepatotoxicity did not occur as an isolated event more frequently in patients prescribed AZA than other anti-rheumatic agents. Neutropenia occurred in 24% of patients (RR 1.15, 95% CI 1.07-1.23), but neutrophilia was also frequent (45%) and increased in patients prescribed AZA (RR 1.28, 95% CI 1.22-1.34). Overall findings supported classic clinical knowledge that agranulocytosis is a largely unpredictable phenomenon. Rounding errors propagated in the statistically de-identified datasets for cohorts as small as 40 patients only contributed marginally to the calculated risk.

Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. (Patel VN, Kaelber DC. J Biomed Inform. 2013)

- **Figure:** Compared to other anti-rheum drugs, AZA side effects shown. Those in green have an increased risk for occurrence in patients prescribed AZA; gray nodes have a decreased or non-significant risk. Sizes correspond to the proportion of patients experiencing that side effect. Edge widths scaled to the proportion of patients experiencing the side effect pair; edge color to relative risk.



Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. (Patel VN, Kaelber DC. J Biomed Inform. 2013)

- **CONCLUSION:** Demonstration that aggregated, standardized, normalized and de-identified population level EHR data can provide both sufficient insight and statistical power to detect potential patterns of medication side effect associations, serving as a multivariate and generalizable approach to post-marketing drug surveillance

Secondary use of clinical data: The Vanderbilt approach. (Danciu I, et al. Journal of biomedical informatics (JBI). 2014)

- **Goal:** Significant growth in re-use of data nationally. Goal here was to share best practices and lessons learned related to data re-use framework and infrastructure at one institution with extensive experience.
- **Methods:** Description of the Vanderbilt research data warehouse framework. It consists of identified and de-identified clinical data repositories, fee-for-service custom services, and tools built atop the data layer to assist researchers across the enterprise. Provides resources dedicated to research initiatives benefits not only the research community, but also clinicians, patients and institutional leadership.

Secondary use of clinical data: The Vanderbilt approach. (Danciu I, et al. Journal of biomedical informatics. 2014)

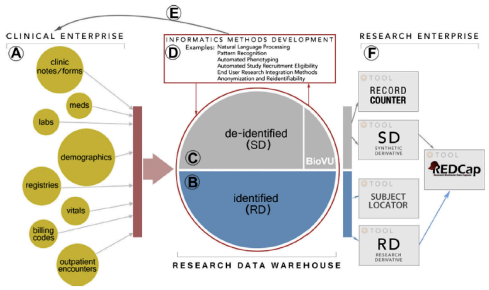


Fig. 1. Vanderbilt clinical and research enterprise.

Secondary use of clinical data: The Vanderbilt approach. (Danciu I, et al. Journal of biomedical informatics. 2014)

- **Conclusion:** Nice summary of one institution's approach to the secondary use of clinical data for research.
- Describes and emphasizes not just key technical components and a list of lessons learned, but also governance considerations.
- Common lessons that should assist others assembling similar services and infrastructure.

Caveats for the use of operational electronic health record data in comparative effectiveness research. (Hersh WR, et al. Medical care. 2013.)

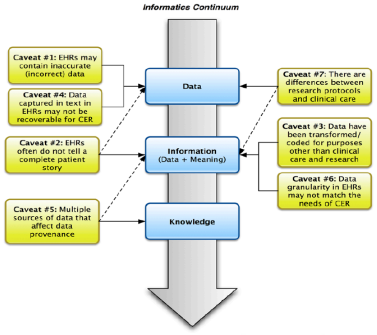
- **Goal:** Great opportunities for clinical data reuse for many tasks, including research. However, there are many caveats to the use of such data
- **Methods:** EDM Forum sponsored effort to convene experts in field to develop list of caveats to inform would-be users of clinical data as well as provide an informatics related recommendations to help improve re-use for comparative effectiveness research
- **Results:** Electronic health record data must be used with knowledge that it could be limited in various ways. Such data may be:
  - Inaccurate
  - Incomplete
  - Transformed in ways that undermine their meaning
  - Unrecoverable for research
  - Of unknown provenance
  - Of insufficient granularity
  - Incompatible with research protocols
- Examples and framework…

Caveats for the use of operational electronic health record data in comparative effectiveness research. (Hersh WR, et al. Medical care. 2013.)

TABLE 1. Data Idiosyncrasies for Use of Operational Electronic Health Record Data in Comparative Effectiveness Research

Type	Description	Examples
Diagnostic uncertainty	Diagnosis may be recorded when there is only a suspicion of disease	Patient with suspected diabetes mellitus before diagnosis confirmed by laboratory testing
	Some overlapping clinical conditions are difficult to distinguish reliably	Various forms of upper respiratory and related infections, eg, sinusitis, pharyngitis, bronchitis, rhinitis, etc.
	Patients may only partially fit diagnostic criteria	Patients with nondiagnostic gastrointestinal symptoms may partially fit diagnostic criteria for 1 or multiple diseases
	Patients in whom diagnostic testing is done but is negative are still more likely to have disease	Patients undergoing echocardiography for shortness of breath and edema who are found to have normal left ventricular function are different from asymptomatic patients with normal left ventricular function
Diagnostic timing	Repeated diagnosis codes over time may represent a new event or a follow-up to an earlier event	Two hospitalizations with a primary diagnosis of MI are likely 2 events, but a code for MI in outpatient setting is more likely a follow-up to an inpatient MI
	First diagnosis in a database is not necessarily an incident case of disease	A new patient in the system with diabetes may have had diabetes for many years before presentation
	Chronic diseases may vary in severity over time	Patient with congestive heart failure with waxing and waning of symptoms
Treatment choice and timing	Many conditions do not require immediate drug or other treatment	Hyperlipidemia or hypertension may have a trial of lifestyle changes before initiation of drug therapy
	Patient comorbidities may affect timing and choice of treatment	Patient with hypertension may have related diagnoses that were not recorded before initiation of treatment, but may be recorded later to indicate the compelling reason for a treatment choice, such as the use of ACE inhibitors in hypertensive patients with heart failure
Treatment decisions	Treatment decisions not randomized	Physician choosing treatment based on personal views or biases regarding efficacy
	Some treatment decisions are remote to the patient-provider interaction	Restrictions by patient insurance or institutional drug formulary
	Some treatments not reliably recorded	Medications available over the counter and not requiring a prescription may not be recorded, eg, aspirin, proton pump inhibitors
Treatment follow-up	Some treatments confounded by clinical factors unrelated to condition being treated	Patient with multiple comorbidities may be seen more frequently and have conditions treated faster, eg, hyperlipidemia in otherwise healthy person versus patient with diabetes and its complications
	Nonclinical factors impact availability of data	Patient access to resources to follow treatment recommendations may be limited due to travel, payor systems, or other nonclinical factors

Caveats for the use of operational electronic health record data in comparative effectiveness research. (Hersh WR, et al. Medical care. 2013.)



Validity of physician billing claims to identify deceased organ donors in large healthcare databases. (Li AH, et al. PloS one. 2013)

- Often Important to know if subjects in database are living for research.
- **Goal:** Goal here to evaluate validity of physician billing claims to identify deceased organ donors in large provincial healthcare databases.
- **Methods:** Population-based retrospective validation study of all deceased donors in Ontario, Canada from 2006 to 2011 (n = 988). Included all registered deaths during the same period (n = 458,074). Main outcome measures included sensitivity, specificity, positive predictive value, and negative predictive value of various algorithms consisting of physician billing claims to identify deceased organ donors and organ-specific donors compared to a reference standard of medical chart abstraction.
- **Results:** The best performing algorithm consisted of any one of 10 different physician billing claims. It had a sensitivity of 75.4% (95% CI: 72.6% to 78.0%) and a positive predictive value of 77.4% (95% CI: 74.7% to 80.0%) for the identification of deceased organ donors. As expected, specificity and negative predictive value were near 100%. The overall number of organ donors identified by the algorithm each year was similar to the expected value, but those to identify organ–specific donors performed poorly (e.g. sensitivity ranged from 0% for small intestine to 67% for heart; positive predictive values ranged from 0% for small intestine to 37% for heart).

Validity of physician billing claims to identify deceased organ donors in large healthcare databases. (Li AH, et al. PloS one. 2013)

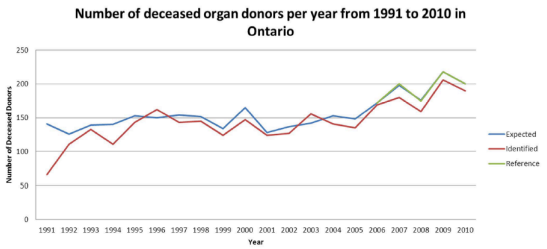


Figure 1. Number of deceased organ donors per year from 1991 to 2010 in Ontario. Expected (blue line) represents the number of deceased organ donors as reported by the Canadian Organ Replacement Registry. Identified (red line) represents the number of deceased organ donors as reported by our physician billing algorithm. Reference (green line) represents the number of deceased organ donors as reported by manual chart reviews from 2006 to 2010.

Validity of physician billing claims to identify deceased organ donors in large healthcare databases. (Li AH, et al. PloS one. 2013)

- **Conclusion:** Billing data not nearly as accurate as manual chart review in this population/setting. Similar findings to those previously reported in Japanese cohort. Varies by condition.
- Primary data abstraction to identify deceased organ donors should be used whenever possible, particularly for the detection of organ-specific donations.
- The limitations of physician billing claims should be considered whenever they are used.
- More caveats to keep in mind…

Other notable papers in this (Re-use) category:

- **Computing health quality measures using Informatics for Integrating Biology and the Bedside.** (Klann JG, Murphy SN. Journal of medical Internet research. 2013)
- **Using electronic dental record data for research: a data-mapping study.** (Liu K, Acharya A, Alai S, Schleyer TK. Journal of dental research. 2013)
- Series of articles from "i2b2 challenge for temporal characteristics" in JAMIA last year:
  - **Evaluating temporal relations in clinical text: 2012 i2b2 Challenge.** Sun W, et al. JAMIA. 2013)
  - **A hybrid system for temporal information extraction from clinical text.** (Tang B, et al. JAMIA. 2013)
  - *More on CRI-related special issues later!*

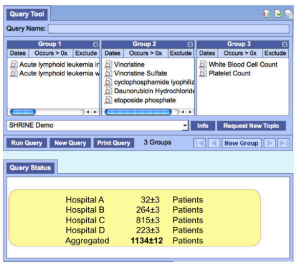
Data Sharing & Discovery

SHRINE: enabling nationally scalable multi-site disease studies. (McMurry AJ et al. PloS one. 2013.)

- Growing need to pool data from multiple sites/sources, and do so in a federated manner. This is first of two articles that address this…
- **Goal:** Demonstrate ability to query data across sites, different source platforms, to create a research system that could aggregate as many patient observations as possible from a large number of hospitals in a uniform way.
- **Methods:** Built upon i2b2 framework, the 'Shared Health Research Information Network', has the following properties: (1) reuse electronic health data from everyday clinical care for research purposes, (2) respect patient privacy and hospital autonomy, (3) aggregate patient populations across many hospitals to achieve statistically significant sample sizes that can be validated independently of a single research setting, (4) harmonize the observation facts recorded at each institution such that queries can be made across many hospitals in parallel, (5) scale to regional and national collaborations.

SHRINE: enabling nationally scalable multi-site disease studies. (McMurry AJ et al. PloS one. 2013.)

Figure 1. Investigator's perspective of the SHRINE Webclient. **Group 1** defines searches for patients with Acute Lymphoid Leukemia (ALL). **Group 2** refines the search result to only those patients having one of the medications listed. The medications shown are all chemotherapeutic agents administered during intensive phase. **Group 3** further refines the result to require a lab test administered during diagnosis. Lab test values can be set directly or flagged as 'abnormally high/low'. In the **Query Status window**, patient counts are displayed with a Gaussian blur to provide additional privacy



eguards of small patient populations. Results are shown for each hospital and the aggregated patient set size.

SHRINE: enabling nationally scalable multi-site disease studies. (McMurry AJ et al. PloS one. 2013.)

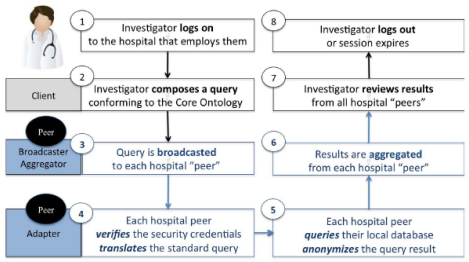


Figure 2. Investigator logs in and composes a query in steps 1–2. SHRINE securely queries multiple hospital peers and returns aggregated results in steps 3–6. The process of securing and translating queries across multiple hospitals is invisible to the investigator user. Lastly, the investigator reviews the results and logs out in steps 7–8.

SHRINE: enabling nationally scalable multi-site disease studies. (McMurry AJ et al. PloS one. 2013.)

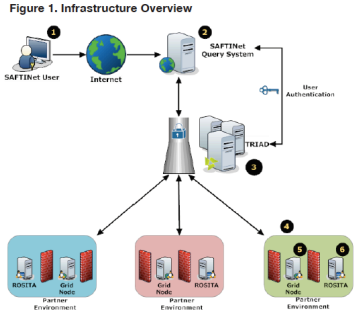
- **Conclusion:** Open source software for multi-site clinical studies can enable federated queries.
- SHRINE implementation in use for multi-site studies of such conditions as autism co-morbidity, juvenile idiopathic arthritis, peripartum cardiomyopathy, colorectal cancer, diabetes, and others.
- An emerging model for such work that is important to our field.

Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) Technology Infrastructure for a Distributed Data Network. (Schilling, LM.et al. eGEMs 2013)

- **Goal:** Distributed Data Networks (DDNs) often require connecting disparate data sources and platforms to support comparative effectiveness research. This approach addresses technical and governance concerns without regard to source, including network security and data disclosure laws and regs.
- **Methods:** The Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) deploys TRIAD grid technology, a common data model, detailed technical documentation, and custom software for data harmonization to facilitate data sharing in collaboration with stakeholders in the care of safety net populations. Data sharing partners host TRIAD grid nodes containing harmonized clinical data within their internal or hosted network environments. Authorized users can use a central web-based query system to request analytic data sets.

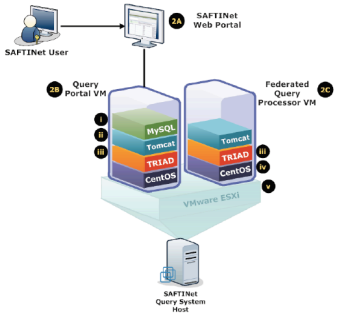
Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) Technology Infrastructure for a Distributed Data Network. (Schilling, LM.et al. eGEMs 2013)

Figure 1. Infrastructure Overview



Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) Technology Infrastructure for a Distributed Data Network. (Schilling, LM.et al. eGEMs 2013)

Figure 2. Query System Components



Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) Technology Infrastructure for a Distributed Data Network. (Schilling, LM.et al. eGEMs 2013)

- **Conclusions:** SAFTINet DDN infrastructure achieved a number of data sharing objectives, including scalable and sustainable systems for ensuring harmonized data structures and terminologies and secure distributed queries.
- In addition to technical description, issues of development and implementation of technical documentation, governance, and related solutions also discussed.
- Platform now in use by other DDNs as well

Data Sharing & Discovery (others)

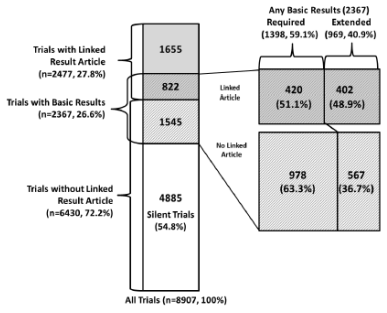
- **Privacy technology to support data sharing for comparative effectiveness research: a systematic review**. (Jiang X, Sarwate AD, Ohno-Machado L. Medical care. 2013)
  - Nice survey showing that state-of-the-art privacy-preserving technologies can guide the development of practical tools that will scale up for CER. But, many challenges to address
- **Pathways to Success for Multi-Site Clinical Data Research**. (McGraw, D & Leiter, AB. eGEMs 2013)
  - Details experiences of 9 multi-site research initiatives across the country, as well as a number of resources used by project participants to work through various challenges – informs governance in data sharing and re-use.
- **Identifying clinical/translational research cohorts: ascertainment via querying an integrated multi-source database.** (Hurdle JF et al. JAMIA. 2013)
  - Cohort identification tool combining multiple, publically available resources for pre-research finding.
- **The secure medical research workspace: an IT infrastructure to enable secure research on clinical data.** (Shoffner M, et al. Clinical and translational science. 2013)
  - *Collaborative* led out of UNC developed Secure Medical Research Workspace (SMRW) that enables researchers to use clinical data securely for research.

Methods in CRI

Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. (Huser & Cimino. PloS one. 2013)

- **Goal:** To understand how results of human clinical trials (the "*Trialome*") are made public.
- **Methods:** Analyzed large set of trials at ClinicalTrials.gov, considering two result artifacts: (1) existence of a trial result journal article that is formally linked to a registered trial or (2) the deposition of a trial's basic summary results within the registry.
- **Results:** Sample = 8907 completed, interventional, phase 2-or-higher clinical trials, 2006-2009. Majority (72.2%) had no structured trial-article link. 2367 trials (26.6%) deposited basic summary results within the registry. Of those, 969 trials (10.9%) had extended results and 1398 trials (15.7%) had only required basic results. Most (54.8%) had no results, and few (9.2%) report results through both registry deposition and publication.

Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. (Huser & Cimino. PloS one. 2013)



Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. (Huser & Cimino. PloS one. 2013)

- ***Conclusion:*** Even when combining publications and registry results, and despite availability of several information channels, trial sponsors often do not meet the mandate to inform the public either via a linked result publication or basic results submission.
- Authors identify several mechanisms by which the linkages between trials and their published results can be increased.

The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. (Ida Sim et al. JBI 2013)

- **Goals:** Need for more attention on CRI science than operational processes for conducting clinical studies. The activities of these scientific processes – the science of clinical research – are centered on the study protocol, which is the abstract representation of the scientific design of a clinical study.
- **Methods:** The Ontology of Clinical Research (OCRe) is an OWL 2 model of the entities and relationships of study design protocols for the purpose of computationally supporting the design and analysis of human studies. OCRe's modeling is independent of any specific study design or clinical domain. It includes a study design typology and a specialized module called *ERGO Annotation* for capturing the meaning of eligibility criteria.

The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. (Ida Sim et al. JBI 2013)

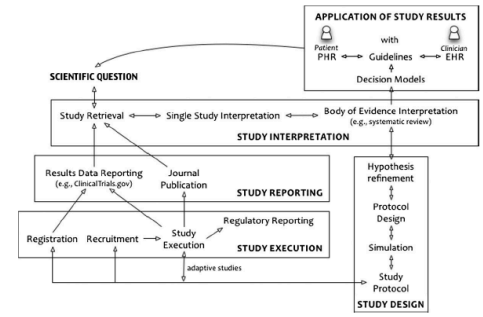


Fig. 1. Idealized scientific lifecycle of a human study within a learning health system.

The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. (Ida Sim et al. JBI 2013)

- **Conclusions:** OCRe captures the central semantics that underlies the scientific processes of clinical research.
- It serves as an informatics foundation for supporting the entire range of knowledge activities that constitute the science of clinical research.
- Great overview of key issues in CRI and represents methodological advance in fundamental CRI science with widespread implications.

An informatics framework for the standardized collection and analysis of medication data in networked research. (Richesson RL. JBI. 2014)

- **Goal:** Medication exposure is important to nearly all clinical research, but great variation in how the data are collected, coded, and analyzed. Coding and classification systems for medication data are heterogeneous in structure, and there is little guidance for implementing them, especially in large research networks and multi-site trials. Largely related to evolution from paper-based environments with limitations, not thoughtful to our purposes and environment today (with electronic systems and solutions).
- This paper provides a framework for standardizing such data.
- **Methods:** Review approaches to coding medication data in multi-site research contexts, and propose a framework for the classification, reporting, and analysis of medication data.

An informatics framework for the standardized collection and analysis of medication data in networked research. (Richesson RL. JBI. 2014)

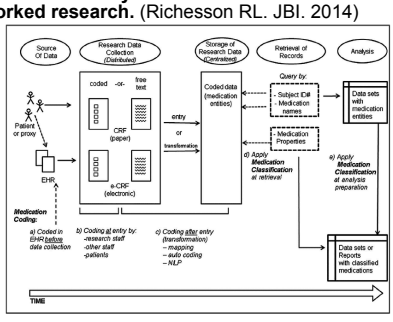


Fig. 1. Collection, coding, storage, retrieval, and classification of medication data in research studies. Medication coding systems can be applied (a) before research data collection, (b) at data entry (electronic or paper), or (c) after collection. Medication classification systems can be applied (d) at the time of data retrieval or (e) as part of analysis preparation and reporting.

An informatics framework for the standardized collection and analysis of medication data in networked research. (Richesson RL. JBI. 2014)

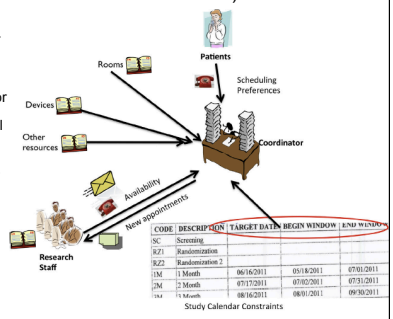
- **Conclusion:** Great review and framework paper for anyone working on challenges of how to manage medication data for research.
- The framework proposed can be used to develop tools for classifying medications in coded data sets to support context appropriate, explicit, and reproducible data analyses by researchers and secondary users across clinical research domains.

An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. (Weng C, et al. Journal of biomedical informatics. 2013)

- **Goal:** Develop a clinical research visit scheduling system to coordinate clinical research visits with patient care visits and increase efficiency at clinical sites where clinical and research activities occur simultaneously (dual purpose visit scheduling).
- **Methods:** Participatory Design to create this software called Integrated Model for Patient Care and Clinical Trials (IMPACT). Using a multi-user constraint satisfaction and resource optimization algorithm, IMPACT automatically synthesizes temporal availability of various research resources and recommends the optimal dates and times for pending research visits.
- Conducted scenario-based evaluations with 10 clinical research coordinators (CRCs) from diverse clinical research settings to assess the usefulness, feasibility, and user acceptance of IMPACT. We obtained qualitative feedback using semi-structured interviews with the CRCs.

An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. (Weng C, et al. Journal of biomedical informatics. 2013)

- **Figure 1:** The current workflow for scheduling a research visit. The scheduler or research coordinator needs to manually synthesize temporal constraints of multiple research resources and staff, as well as patients.



An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. (Weng C, et al. Journal of biomedical informatics. 2013)

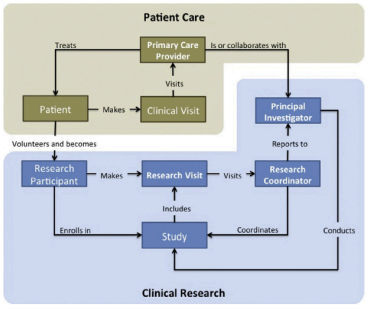


Fig. 2. The semantic relationships of entities in the domain model.

An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. (Weng C, et al. Journal of biomedical informatics. 2013)

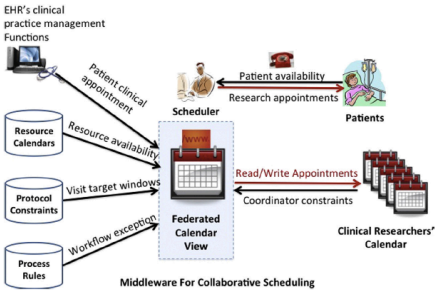


Fig. 3. The IMPACT architecture.

An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. (Weng C, et al. Journal of biomedical informatics. 2013)

- **Results:** Most CRCs acknowledged the usefulness of IMPACT features. Support for collaboration within research teams and interoperability with electronic health records and clinical trial management systems were highly requested features. Good user acceptance.
- **Conclusion:** Nice example of using classic informatics approaches to develop needed solution where no good one currently exists. Concepts are generalizable.
- Next step to compare the effectiveness of IMPACT with that of existing scheduling solutions on the market and conducting field tests.

Methods in CRI (others)

- **Ethical and practical challenges to studying patients who opt out of large-scale biorepository research.** (Rosenbloom ST, et al. JAMIA. 2013)
  - Good considerations for those doing bio-repositories and universal consenting
- **Clustering clinical trials with similar eligibility criteria features.** (Hao T, et al. JBI 2014)
  - Important innovation that could assist with screening, recruitment, etc.
- **Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health.** (Crutzen R, et al. Health promotion international. 2013)
  - Not CRI per se, but novel use of existing tool for those doing Internet-based health research (e.g. with consumers)
- **Confounding adjustment in comparative effectiveness research conducted within distributed research networks.** (Toh S, et al. Medical care. 2013)
  - Good approaches to address common sources of bias in multi-site CER

Policy and Perspectives

CRI Policy & Perspectives:

- **Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting.** (Hripcsak G, Bloomrosen M, Flately Brennan P, Chute CG, Cimino J, Detmer DE, et al. JAMIA. 2014)
- **Trustworthy reuse of health data: a transnational perspective.** (Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, et al. International journal of medical informatics. 2013)
- **People, organizational, and leadership factors impacting informatics support for clinical and translational research.** (Payne PR, Pressler TR, Sarkar IN, Lussier Y. BMC medical informatics and decision making. 2013)
- **Don't take your EHR to heaven, donate it to science: legal and research policies for EHR post mortem** (Huser & Cimino. JAMIA 2013)
- **Evidence generating medicine: redefining the research-practice relationship to complete the evidence cycle.** (Embi & Payne. Medical care. 2013)

Other CRI Trends

Other CRI Trends:

- **Using Evernote as an electronic lab notebook in a translational science laboratory.** (Walsh E, Cho I. Journal of laboratory automation. 2013)
- **Measure once, cut twice--adding patient-reported outcome measures to the electronic health record for comparative effectiveness research.** (Wu AW, Kharrazi H, Boulware LE, Snyder CF. Journal of clinical epidemiology. 2013)
- **Participatory surveillance of hypoglycemia and harms in an online social network.** (Weitzman ER, Kelemen S, Quinn M, Eggleston EM, Mandl KD. JAMA internal medicine. 2013)
- **Describing the relationship between cat bites and human depression using data from an electronic health record.** (Hanauer DA, Ramakrishnan N, Seyfried LS. PloS one. 2013)
- Many more publications focused in research informatics…

Other CRI Trends:

- **Recent trends in biomedical informatics: a study based on JAMIA articles.** (Jiang X, Tse K, Wang S, Doan S, Kim H, Ohno-Machado L. JAMIA. 2013)

Most cited 2011 articles in 2011–12			
1	2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text[34]	23	Perspective
2	Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings[35]	20	Review (FREE)
3	Factors motivating and affecting health information exchange usage[36]	15	Research and applications
4	Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting[37]	15	Perspective
5	Translational bioinformatics: linking knowledge across biological and clinical realms[38]	13	Perspective (OPEN)
6	Data from clinical notes: a perspective on the tension between structure and flexible documentation[39]	13	Perspective
7	Anaphoric relations in the clinical narrative: corpus creation[40]	11	Research and applications
8	Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access[41]	11	Case report (FREE)
9	Ability of pharmacy clinical decision-support software to alert users about clinically important drug–drug interactions[42]	10	Research and applications
10	Factors influencing alert acceptance: a novel approach for predicting the success of clinical decision support[43]	10	Research and applications
Most cited 2012 articles in 2012 to June 2013			
1	Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study[44]	15	Research and applications
2	iDASH: integrating data for analysis, anonymization, and sharing[45]	10	Brief communication
3	A novel signal detection algorithm for identifying hidden drug–drug interactions in adverse event reports[46]	10	Research and applications (FREE)
4	The National Center for Biomedical Ontology[47]	9	Brief communication
5	The dangerous decade[48]	7	Perspective
6	The financial impact of health information exchange on emergency department care[49]	7	Research and applications (OPEN)
7	High-priority drug–drug interactions for use in electronic health records[50]	7	Research and applications
8	A translational engine at the national scale: informatics for integrating biology and the bedside[51]	6	Brief communication
9	A systematic review of the psychological literature on interruption and its patient safety implications[52]	6	Review
10	Portability of an algorithm to identify rheumatoid arthritis in electronic health records[53]	6	Research and applications (FREE)

Notable CRI-Related Events in Past Year

Major research funding news

- **ARRA funding ended**
  - Major impact to CRI… while it lasted
  - (unintentionally led to research on sustainability)
- **AHRQ re-funded EDM Forum**
  - Speaks to importance of knowledge sharing and extensive work that remains – is worthy of funding
  - Expanded scope to include broader cross-section of stakeholders (not just PROSPECT funded projects)
- **Funding of PCORnet**
  - Ongoing – stay tuned – but will undoubtedly have major impact in CRI
- **NIH funding – could have been worse**
  - BD2K, NCATS, New NCI programs
  - Still recognition of need for CRI (still need more for NLM)

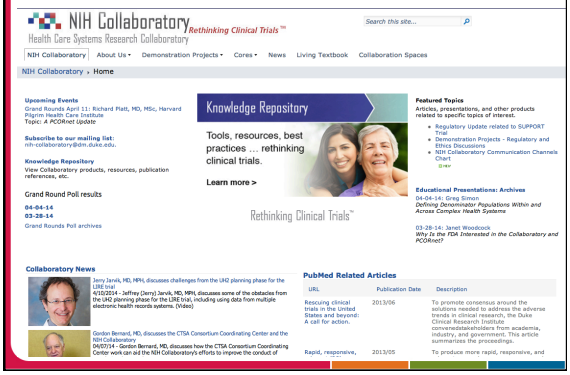
Initiatives to Inform EHR design for Clinical Research

- Broad recognition of need for work, resources in CRI space:
- **Clinical Research Forum – IT Roundtable**
  - October 2013 – 2 day meeting in DC
  - Included researcher and informatics leaders from AHCs, Health and research IT vendors
  - Needs, challenges and opportunities for leveraging EHRs to support research identified
  - White paper and call to action under development
- Another initiative: NIH Collaboratory…

NIH collaboratory

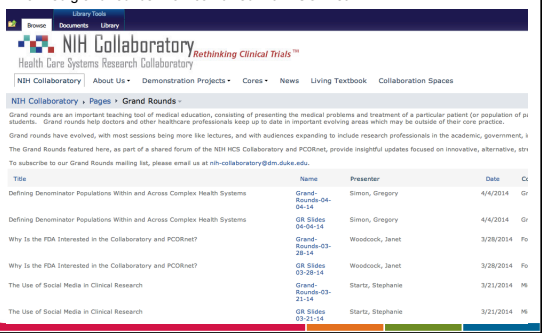
- NIH Common Fund initiaitive – started in 2012
  - Focus on rethinking trials – pragmatic clinical trials – and leveraging EHRs for research.
  - Health Care Systems Collaboratory program, engages healthcare systems as partners in discussing and promoting activities, tools, and strategies for supporting active participation in PCTs.
  - The NIH Collaboratory consists of seven demonstration projects, and seven problem-specific working group 'Cores', aimed at leveraging the data captured in heterogeneous 'real-world' environments for research, thereby improving the efficiency, relevance, and generalizability of trials.

NIH collaboratory: Website



NIH collaboratory: Resources and Education

Archived grand rounds. Now combined with PCORnet



NIH collaboratory

Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory (Richesson RL, et al. JAMIA 2013)

- In addition to introducing NIH Collaboratory, particular focus on its Phenotype, Data Standards, and Data Quality Core.
- Presents early observations from researchers implementing PCTs within large healthcare systems.
- Identifies gaps in knowledge and present an informatics research agenda that includes identifying methods for the definition and appropriate application of phenotypes in diverse healthcare settings, and methods for validating both the definition and execution of electronic health records based phenotypes.

Special Journal Issues dedicated to CRI Topics

Much activity in new eGEMs online journal:

- Generating Evidence & Methods to improve patient outcomes
- Four Thematic Areas:
  - Methods
  - Informatics
  - Governance
  - Learning Health System
- EDM Forum of Academy Health
- Free, Open, Peer-reviewed

- 32 manuscripts last year - >450 downloads/paper
  - Soon will be indexed in Pubmed (already in EBSCO)

New CRI-focused chapters BMI Textbooks

- In "Biomedical Informatics" 4th Edition
  - Editors: Shortliffe & Cimino
  - Chapter authors:
    - Payne, Embi, Cimino
- In "Informatics Education in Healthcare"
  - Editor: Berner
  - Chapter authors: Embi & Payne
- More evidence of CRI as established domain

In Summary…

- Maturing informatics approaches in CRI – accelerating
  - Emergence of methods work important indicator
- CRI infrastructure also maturing and beginning to improve science
- Multiple groups/initiatives converging on common needs to advance the field
- Poised to deploy and test our approaches to realize the "*learning health system*"
- Already collecting articles for next year!
- Exciting time to be in CRI!

Thanks!

Special thanks to those who suggested articles/ events to highlight:

- Philip Payne
- Rachel Richesson
- Adam Wilcox
- Chunhua Weng
- Erin Holve

Thanks!

Peter.Embi@osumc.edu

Slides will be linked to on http://www.embi.net/ (click on "Informatics")