AMIA
2016 Summit in CRI
CRI Year-In-Review

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

Associate Prof & Vice Chair, Biomedical Informatics
Associate Professor of Medicine & Public Health
Chief Research Information Officer/Assoc Dean for Research Informatics
Director, Biomedical Informatics, CCTS
The Ohio State University

San Francisco, California
March 24, 2016

Disclosures

- Associate Editor, IJMI
- Editorial board, JAMIA
- 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 five years
  - Tracked manuscripts throughout the year
  - Intended to spread work out…

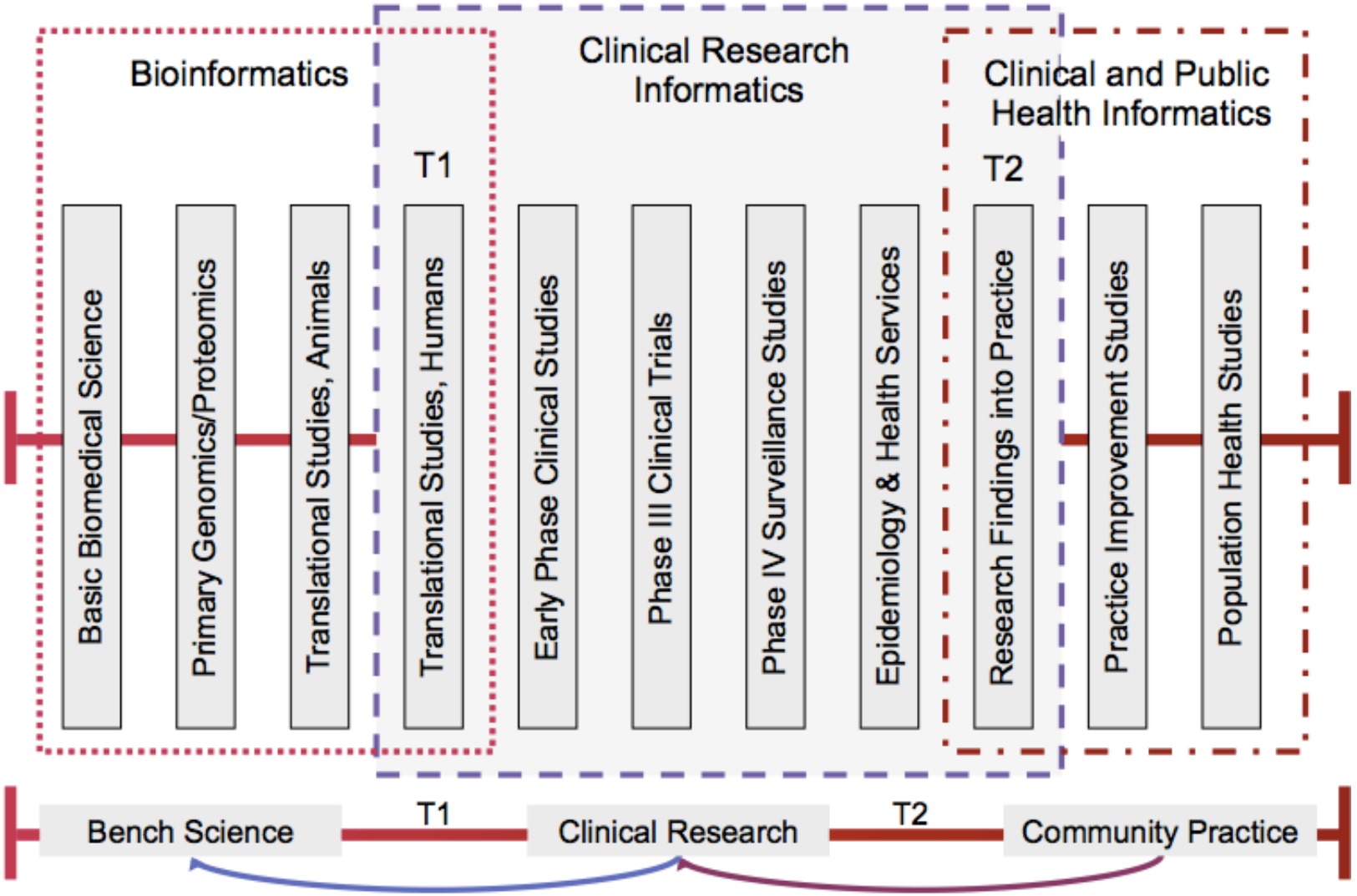
    …still worked down to the wire
- So, what was my approach…

Source of Content for Session

- Literature review:
  - Initial search by MESH terms:
    - "Biomedical Research"[Mesh] AND "Informatics"[Mesh] AND "2015/01/01"[PDat] : "2016/02/01"[Pdat]
    - Resulted in **481** articles
    - Limiting to English and Abstracts: **357**
  - 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 **474** total, from which **185** were CRI relevant
  - From those, I've selected **40** 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



Embi & Payne, JAMIA 2009

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 **40** articles into several CRI categories (admittedly, not *all* CRI areas)
  - Data Sharing, and Re-Use
  - CRI Methods and Systems
  - Recruitment and Eligibility
  - EHRs and Learning Health Systems
  - Emerging Trends in CRI Literature
  - Policy & Perspectives
- In each category, I'll highlight a few key articles and then given a quick(er) "shout out" to some 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 Sharing and Re-Use for Research

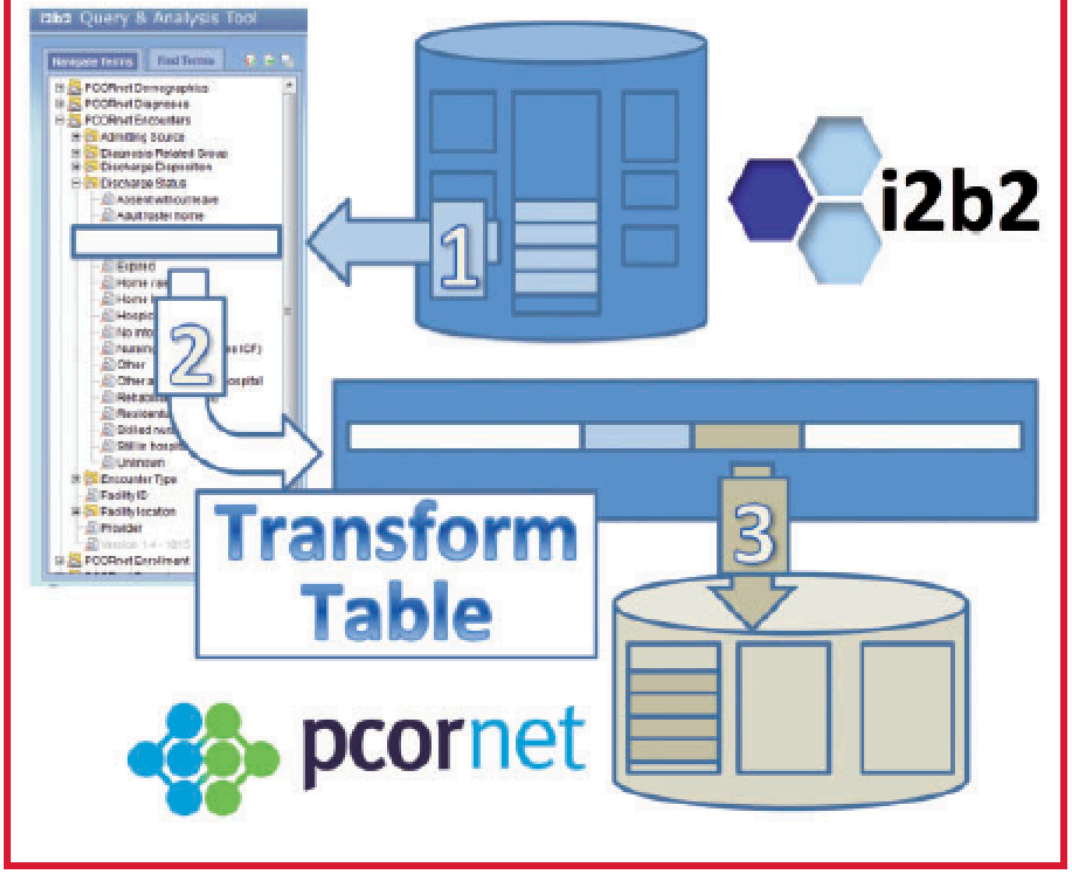
Data Interchange using i2b2
(Klann, et al. JAMIA. Feb. 2016)

- **Objective:** Given multiple different data models across data sharing initiatives, want to exchange data between them without new data extraction, transform, and load (ETL) processes – to reduce the time and expense needed to participate. Goal: Use i2b2 as a hub, to rapidly reconfigure data to meet new analytical requirements without new ETL programming.
- **Methods:** A 12-site PCORnet CDRN that used i2b2 to query data. Developed a process to generate a PCORnet Common Data Model (CDM) physical database directly from existing i2b2 systems, thereby supporting PCORnet analytic queries without new ETL programming. This involved:
  - A formalized process for representing i2b2 information models (the specification of data types and formats);
  - An information model that represents CDM Version 1.0;
  - A program that generates CDM tables, driven by this information model.
- Approach should be generalizable to any logical information model.

Data Interchange using i2b2
(Klann, et al. JAMIA. Feb. 2016)

- **RESULTS:**
- 8 PCORnet CDRN sites implemented this approach and generated a CDM database without a new ETL process from the EHR.
- Enabled federated querying across the CDRN and compatibility with the national PCORnet Distributed Research Network.

Figure 1: The process of populating a new information model schema through look-ups mediated by the i2b2 ontology.



Data Interchange using i2b2
(Klann, et al. JAMIA. Feb. 2016)

- **CONCLUSION:** Useful way to adapt i2b2 to new information models without requiring changes to the underlying data.
- Eight different sites vetted this methodology, resulting in a network that, at present, supports research on 10 million patients' data.
- New analytical requirements can be quickly and cost-effectively supported by i2b2 without creating new data extraction processes from the EHR.

- Useful approach/model for satisfying competing needs to exchange data across initiatives that require different information models with less effort and maintenance.

Transparent Reporting of Data Quality in Distributed Data Networks

(Kahn MG, et al. eGEMs. 2015)

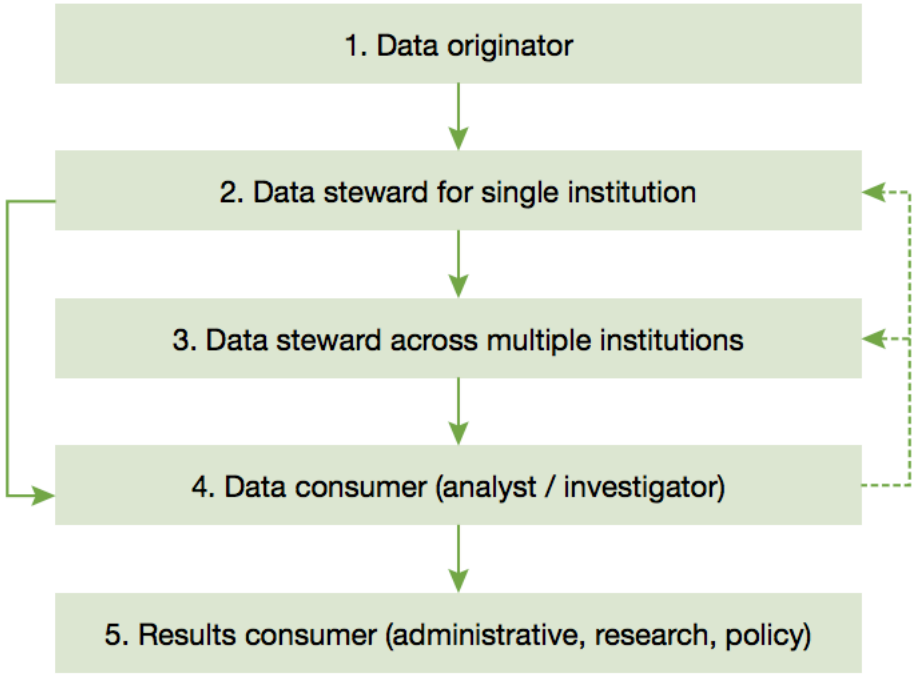
- **Introduction:** As data sharing efforts and availability of electronic administrative and clinical data grow, also a growing concern about the quality of these data for observational research and other analytic purposes.
- Currently, no widely accepted guidelines for reporting quality results that would enable investigators and consumers to independently determine if a data source is fit for use to support analytic inferences and reliable evidence generation.

Transparent Reporting of Data Quality in Distributed Data Networks

(Kahn MG, et al. eGEMs. 2015)

- **Methods:** Developed a conceptual model that captures the flow of data from data originator across successive data stewards and finally to the data consumer.
- The “data lifecycle” model illustrates how data quality issues can result in data being returned back to previous data custodians
- Highlights the potential risks of poor data quality on clinical practice and research results.

Figure 1. Chain of Data Stewardship with Key Data Stewards



Notes: Dashed lines represent data quality issues referred back to previous data stewards.

Transparent Reporting of Data Quality in Distributed Data Networks

(Kahn MG, et al. eGEMs. 2015)

- Given need to ensure transparent reporting of a data quality issues, they created a unifying data-quality reporting framework, including…
- A set of **20 data-quality reporting recommendations** for studies that use observational clinical and administrative data for secondary data analysis.
- Stakeholder input on perceived value of each recommendation via face-to-face meetings, through multiple public webinars targeted to the health services research community, and with an open access online wiki.

Table 1. Data Quality Assessment Documentation and Reporting Recommendations

	Item #	Recommendation
Data Capture		
1. Original data source		
Data origin	1	A description of the source of the original or raw data prior to any subsequent processing or transformation for secondary use. Examples would be "clinical practices via AllScripts EHR 2009," "interviewer-administered survey," or "claim for reimbursement."
Data capture method	2	A description of the technology used to record the data values in electronic format. Examples would be "EHR screen entry via custom form," "automated instrument upload," and "interactive voice response (IVR)."
Original collection purpose	3	A description of the original context in which data were collected. Examples would be "clinical care and operations," "reimbursement," or "research"—and in which kinds of facilities data were collected—such as "ambulatory clinic," "same-day surgery clinic," and "clinical research center."
2. Data steward information		
Data steward	4	A description of the type of organization responsible for obtaining and managing the target data set. Examples could be "PBRN," "Registry," Medical group practice," and "State agency."
Database model/data set structure	5	A description of how the data tables and variables are structured and linked in the target database or data set. Includes information on variable types (integer, date, string), min/max ranges if defined, and allowed values for enumerated categorical variable. Includes rules for mandatory/optional fields (variables), especially for fields used to link rows across tables.
Data dictionary/data set definitions	6	A description of data definitions used for data elements, including the URL to documentation if available on the Internet, that provides table- and field-level descriptions of data types and content for each element, and any required context for interpreting data within a patient or across the population. Whereas Recommendation #5 focuses on how the data are *structured* (data syntax), this requirement focuses on descriptions on what the data *mean* (data semantics) as described in the data definitions.
Data Processing/Data Provenance		
Data extraction specifications, including use of natural language processing to extract variables from text documents	7	Documentation on how the target data was obtained from the source data. Examples would be "direct data entry by medical personnel," "indirect data entry by medical record chart abstraction guidelines," and "natural language processing algorithms." Should include the URL to the documentation of the data creation specifications if available on the Internet.
Mappings from original values to standardized values	8	Documentation on how original data values were transformed to conform to the target data model format. Documentation should list source values and describe the logic or mappings used to transform from the original source to the required target values.
Data management organization's data transformation routines, including constructed variables	9	Documentation of any additional data alterations that were performed by the data management team in creating the final data set, such as replacing missing values by imputed values, removal of extreme values, and creation of additional computed values, such as BMI from raw height and weight observations. Should include the URL to documentation if available on the Internet. The documentation should allow an independent reader to trace a value in the target data set to the original source value(s) and should explain all operations performed on the data.
Data processing validation routines	10	Documentation of all data validation rules to which the data were subjected. Rules should identify both data elements and validation algorithms. Examples include comparisons of row counts between source and target data sets and an explanation for any differences in row count or documentation, and a listing of differences in the distribution of categorical data values across source-to-target mappings. Should include the URL to documentation if available on the Internet.
Audit trail	11	Documentation of all changes made to data values, user/system making the change and date/time of the change in the process of "cleaning" a data set prior to use. Reason for the change should be evident from data transformation routines or documented issues (e.g., correction of isolated error, replacement of missing values with standardized "missing value" flag).

Transparent Reporting of Data Quality in Distributed Data Networks

(Kahn MG, et al. eGEMs. 2015)

- **Recommendations**:
- Report on both general and analysis-specific data quality features.

- **Data Capture**
  - Original Source info
  - Data Steward info
- **Data Processing/ Data Provenance**
- **Data Element Characterization**
- **Analysis (specific data quality steps taken)**

Table 1. Data Quality Assessment Documentation and Reporting Recommendations (Cont'd)

	Item #	Recommendation
Data Elements Characterization		
Data format	12	For required data variables verify the format, proper storage, and that required elements are not missing. Examples include verifying that floating point values are not rounded to integer values, conversions across units of measures are correct, and that precision and rounding rules are as expected based on transformations.
Single element data descriptive statistics	13	For each variable, calculate the following descriptive statistics: • Available or not (#/% missing) • For continuous variables: min, max, mean, median, range, percentiles, etc. • For categorical variables—frequencies & proportions by category • If a specific distribution is anticipated, report on goodness-of-fit tests
Temporal constraints	14	Evaluate whether expected temporal constrains are violated or not. Examples include: • Start date and times occur before stop dates and times, • Distribution of intervals between successive measurements, • For time-series—changes in adjacent values and expected directionality in changes meet expectations, and • Conformance to state transition/sequencing rules.
Multiple variables cross validations/ consistency	15	Across two or more data variables that are known to be linked:[80] • Report violations of data model cardinality rules. A cardinality rule determines when zero, one, or more than one data rows in one table can be linked to one or more data rows in another table. • Report violations of data model primary/foreign key rules. A primary/foreign key requires that a row in one table (the foreign key) must point to a row in another table (the primary key). The primary key row must be present. • Report violations of cross-variables dependency rules. A cross-variables dependency states that one row can only exist if another row or value exists. For example, the state of pregnancy should exist only if the patient sex is female. • Report violations of co-occurrence rules. Systolic and diastolic blood pressures should always occur as a pair. • Report violations of co-measurement rules (two distinct measurements of the same observation). Age and date of birth should agree. • Report violations of mutual exclusivity rules. A patient should not be recorded as being dead and alive at the same time.
Analysis—Specific Data Quality Documentation (As Applied by Investigators or Analytic Team)		
Data cleansing/customization	16	Analytic- or study-specific additions to Item# 9
Data quality checks of key variables used for cohort identification	17	Analytic or study specific additions to Items #13–15 that focus on variables that identify cohorts, detect outcomes, define exposures, and participate as covariates. Where these variables may be affected by other related (perhaps causal) variables, these influential variables should also be included. The list of variables contained in these assessments will vary by intended analysis/clinical study. However variables assessed should be organized according to the following categories: cohort, outcome, exposure, confounding.
Data quality checks of key variables used for outcome categorization	18	
Data quality checks of key variables used to classify exposure	19	
Data quality checks of key confounding variables	20	

Notes: "Source data" refers to the original originating data. "Target data" refers to the data as received by the data user.

Transparent Reporting of Data Quality in Distributed Data Networks

(Kahn MG, et al. eGEMs. 2015)

- **SUMMARY:** If followed, these recommendation could help improve:
- A. the reporting of data quality measures for studies that use observational clinical and administrative data,
- B. to ensure transparency and consistency in computing data quality measures, and
- C. to facilitate best practices and trust in the new clinical discoveries based on secondary use of observational data.

Benefits and Risks in Secondary Use of Digitized Clinical Data: Views of Community Members Living in a Predominantly Ethnic Minority Urban Neighborhood

(Lucero RJ, et al. AJOB Empirical Bioethethics. 2015)

- **BACKGROUND:** Use of clinical data in research raises concerns about the privacy and confidentiality of personal health information. While few studies, even less known about perceptions of underrepresented populations
- Study explored urban community members' views on the secondary use of digitized clinical data to:
  - (1) recruit participants for clinical studies;
  - (2) recruit family members of persons with an index condition for primary studies; and
  - (3) conduct studies of information related to stored biospecimens.

- **METHODS:** A qualitative descriptive design was used to examine the bioethical issues outlined from the perspective of urban-dwelling community members.
- Focus groups were used for data collection, and emergent content analysis was employed to organize and interpret the data.
- Group based at Columbia U – subjects from surrounding area – diverse racially, ethnically, and experience with research (some had refused)

Benefits and Risks in Secondary Use of Digitized Clinical Data: Views of Community Members Living in a Predominantly Ethnic Minority Urban Neighborhood

(Lucero RJ, et al. AJOB Empirical Bioethethics. 2015)

- **RESULTS:** 30 community members attended one of four focus groups ranging in size from 4 to 11 participants.
- Five critical themes emerged from the focus-group material:
  - (1) perceived motivators for research participation;
    - Wanted to give back – “pass the ball”
  - (2) objective or "real-life" barriers to research participation;
    - Not knowing about opportunities; communicate barrier
  - (3) a psychological component of uncertainty and mistrust;
    - Trust issues; preferred computer to paper; skeptical about being “called”
  - (4) preferred mechanisms for recruitment and participation;
    - More comfortable if coming from “my doctor” not a “stranger” with no relationship
  - (5) cultural characteristics can impact understanding and willingness
    - Ownership of specimens data – gone once provided; for hispanics – direct translations led to confusion: eg “investigación” both research and procedure

Benefits and Risks in Secondary Use of Digitized Clinical Data: Views of Community Members Living in a Predominantly Ethnic Minority Urban Neighborhood

(Lucero RJ, et al. AJOB Empirical Bioethethics. 2015)

- **CONCLUSIONS:** The overriding concern of community members regarding research participation and/or secondary clinical and nonclinical use of digitized information was that their involvement would be safe and the outcome would be meaningful to them and to others.
- Participants felt that biospecimens acquired during routine clinical visits or for research are no longer their possession. Although the loss of privacy was a concern for some
- **They preferred that researchers access their personal health information using a digitized clinical file rather than through a paper-based medical record**

- *More work is needed to understand ways to approach sub-populations to improve recruitment – but our current perceptions may be off-base*
- *Yet another reason for patient/public participation in our research endeavors*

Other notable papers in this (Sharing/Reuse) category:

- **Using Electronic Health Records to Support Clinical Trials: A Report on Stakeholder Engagement for EHR4CR** (De Moor G, et al. JBI. 2015)
  - Description of progress toward a shared, multi-national infrastructure aiming to demonstrate, a scalable, widely acceptable and efficient approach to interoperability between EHR systems and clinical research systems.
- **A system to build distributed multivariate models and manage disparate data sharing policies: implementation in the scalable national network for effectiveness research** (Meeker G, et al. JAMIA. 2015)
  - Description of system to enable distributed multivariate models for federated networks in which patient-level data is kept at each site and data exchange policies are managed in a study-centric manner.
  - Study illustrates the use of these systems among institutions with highly different policies and operating under different state laws.
  - An increasingly important use-case, and potential model for emerging networks to consider.

Other notable papers in this (Sharing/Reuse) category:

- **De-identification of Medical Images with Retention of Scientific Research Value** (Moore, SM, et al. Radiographics. 2015)
  - Sharing data is essential, but de-identifying imaging data often considered infeasible. Key is to understand DICOM details well enough to remove PHI without compromising the scientific integrity of the data.
  - Authors describe set tools and process to help researchers and vendors de-identify DICOM images according to current best practices, despite wide variability and lack of standardization regarding PHI in images making this a challenge.
  - An informative article for those working on this problem and a step in the right direction.

Other notable papers in this (Sharing/Reuse) category:

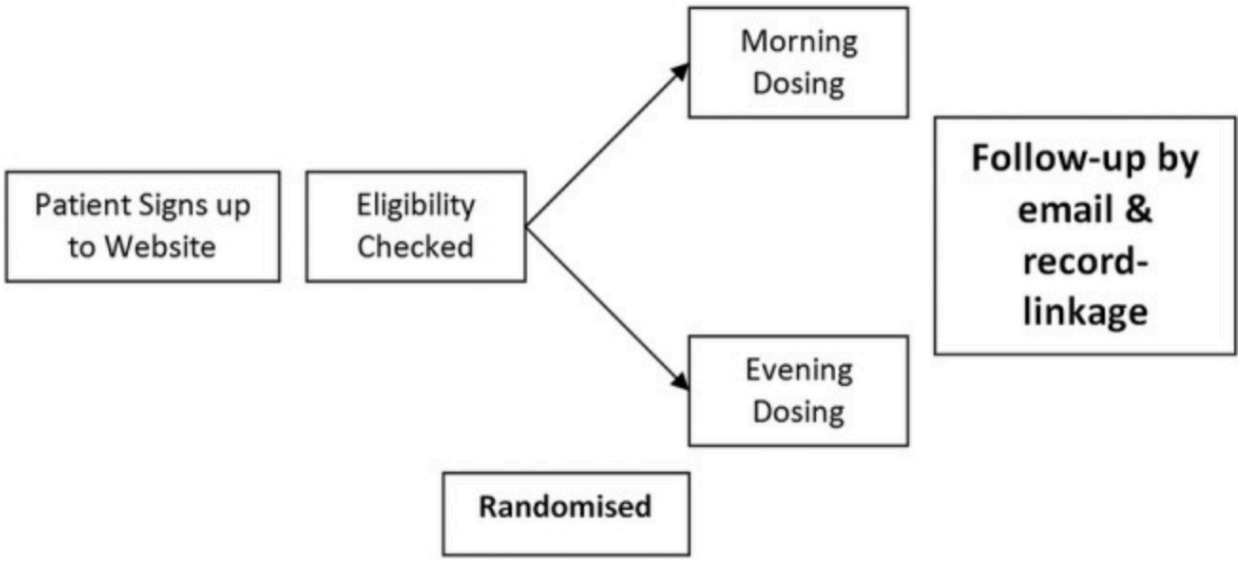
- **Facilitating biomedical researchers' interrogation of electronic health record data: Ideas from outside of biomedical informatics** (Hruby GW, et al. JBI. 2016)
  - Current progress in **BMI literature** is largely in query execution and information modeling, primarily due to emphases on infrastructure development for data integration and data access via self-service query tools
  - In contrast, **the information science literature** has offered elaborate theories and methods for user modeling and query formulation support.
  - Article outlines future informatics research to improve understanding of user needs and requirements for facilitating autonomous interrogation of EHR data by biomedical researchers.
  - **Suggestion: more** cross-disciplinary translational research between biomedical informatics and information science can benefit our research in facilitating efficient data access in life sciences.

CRI Methods and Systems

Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **INTRODUCTION:** Nocturnal blood pressure (BP) appears to be a better predictor of cardiovascular outcome than daytime BP. The BP lowering effects of most antihypertensive therapies are often greater in the first 12 h compared to the next 12 h. The Treatment In Morning versus Evening (TIME) study aims to establish whether evening dosing is more cardioprotective than morning dosing.

Figure 1

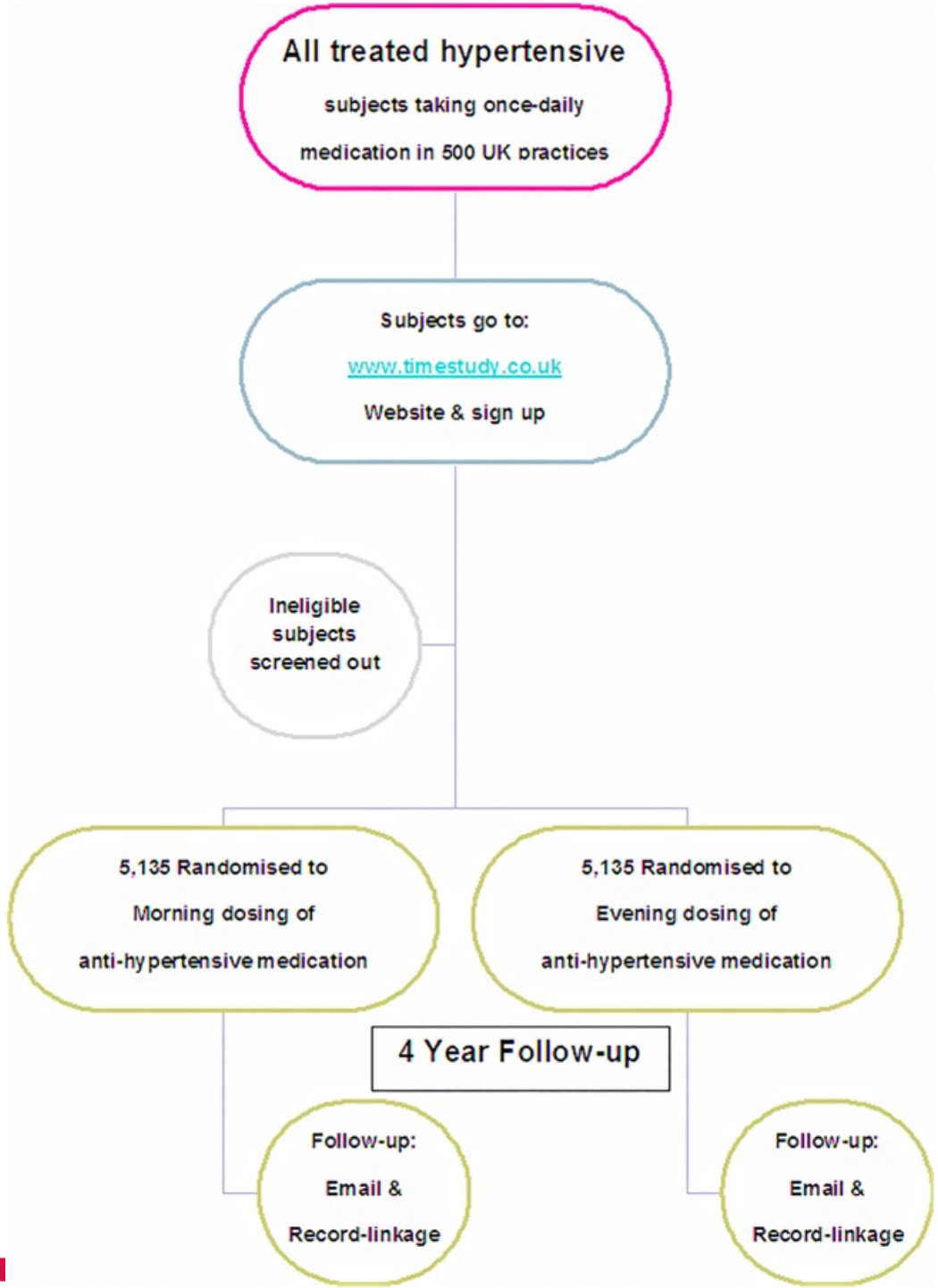


Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **METHODS:** The TIME study uses the prospective, randomised, open-label, blinded end-point (PROBE) design.
- TIME recruits participants by advertising in the community, from primary and secondary care, and from databases of consented patients in the UK.
- Participants must be aged over 18 years, prescribed at least one antihypertensive drug taken once a day, and have a valid email address.
- After the participants have self-enrolled and consented on the secure TIME website (http://www.timestudy.co.uk) they are randomised to take their antihypertensive medication in the morning or the evening.
- Participant follow-ups are conducted after 1 month and then every 3 months by automated email.

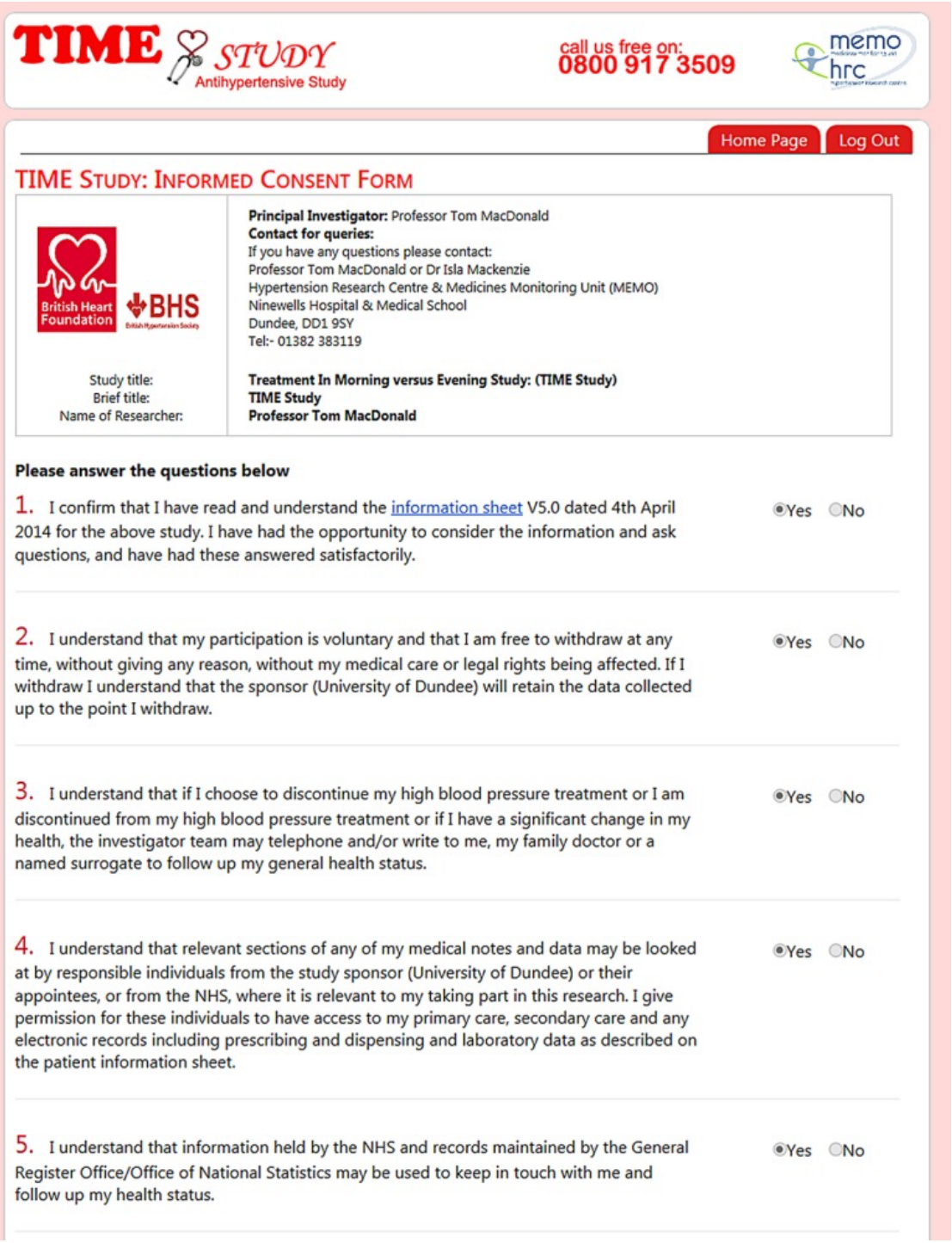
Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **Figure 2: TIME study flow diagram**



Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **Figure 3: Informed consent document**



TIME STUDY
Antihypertensive Study

call us free on:
0800 917 3509

memo hrc

Home Page | Log Out

TIME STUDY: INFORMED CONSENT FORM

British Heart Foundation | BHS British Hypertension Society

Principal Investigator: Professor Tom MacDonald
Contact for queries:
If you have any questions please contact:
Professor Tom MacDonald or Dr Isla Mackenzie
Hypertension Research Centre & Medicines Monitoring Unit (MEMO)
Ninewells Hospital & Medical School
Dundee, DD1 9SY
Tel:- 01382 383119

Study title: **Treatment In Morning versus Evening Study: (TIME Study)**
Brief title: **TIME Study**
Name of Researcher: **Professor Tom MacDonald**

Please answer the questions below

1. I confirm that I have read and understand the information sheet V5.0 dated 4th April 2014 for the above study. I have had the opportunity to consider the information and ask questions, and have had these answered satisfactorily. ◉Yes ○No

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. If I withdraw I understand that the sponsor (University of Dundee) will retain the data collected up to the point I withdraw. ◉Yes ○No

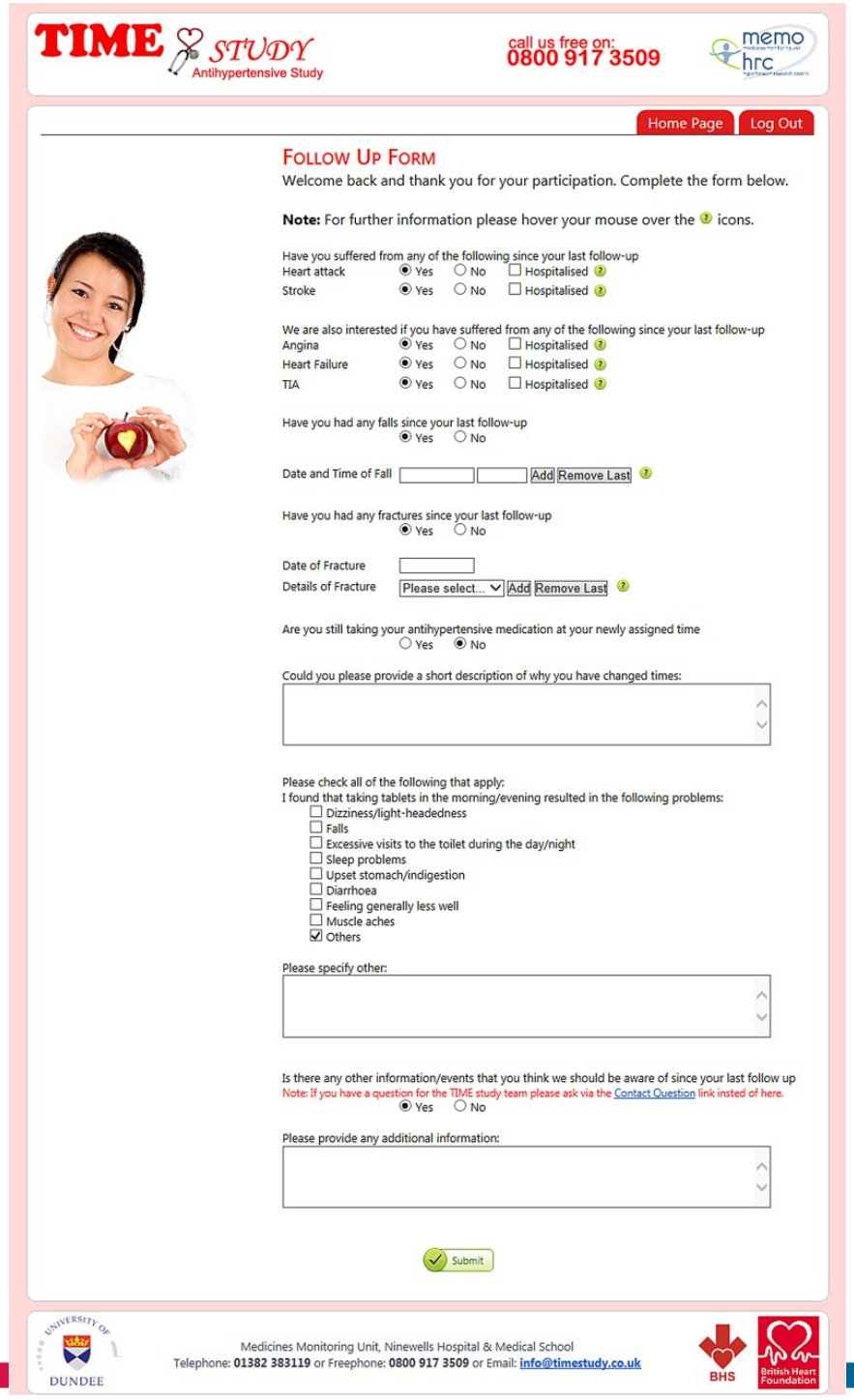
3. I understand that if I choose to discontinue my high blood pressure treatment or I am discontinued from my high blood pressure treatment or if I have a significant change in my health, the investigator team may telephone and/or write to me, my family doctor or a named surrogate to follow up my general health status. ◉Yes ○No

4. I understand that relevant sections of any of my medical notes and data may be looked at by responsible individuals from the study sponsor (University of Dundee) or their appointees, or from the NHS, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my primary care, secondary care and any electronic records including prescribing and dispensing and laboratory data as described on the patient information sheet. ◉Yes ○No

5. I understand that information held by the NHS and records maintained by the General Register Office/Office of National Statistics may be used to keep in touch with me and follow up my health status. ◉Yes ○No

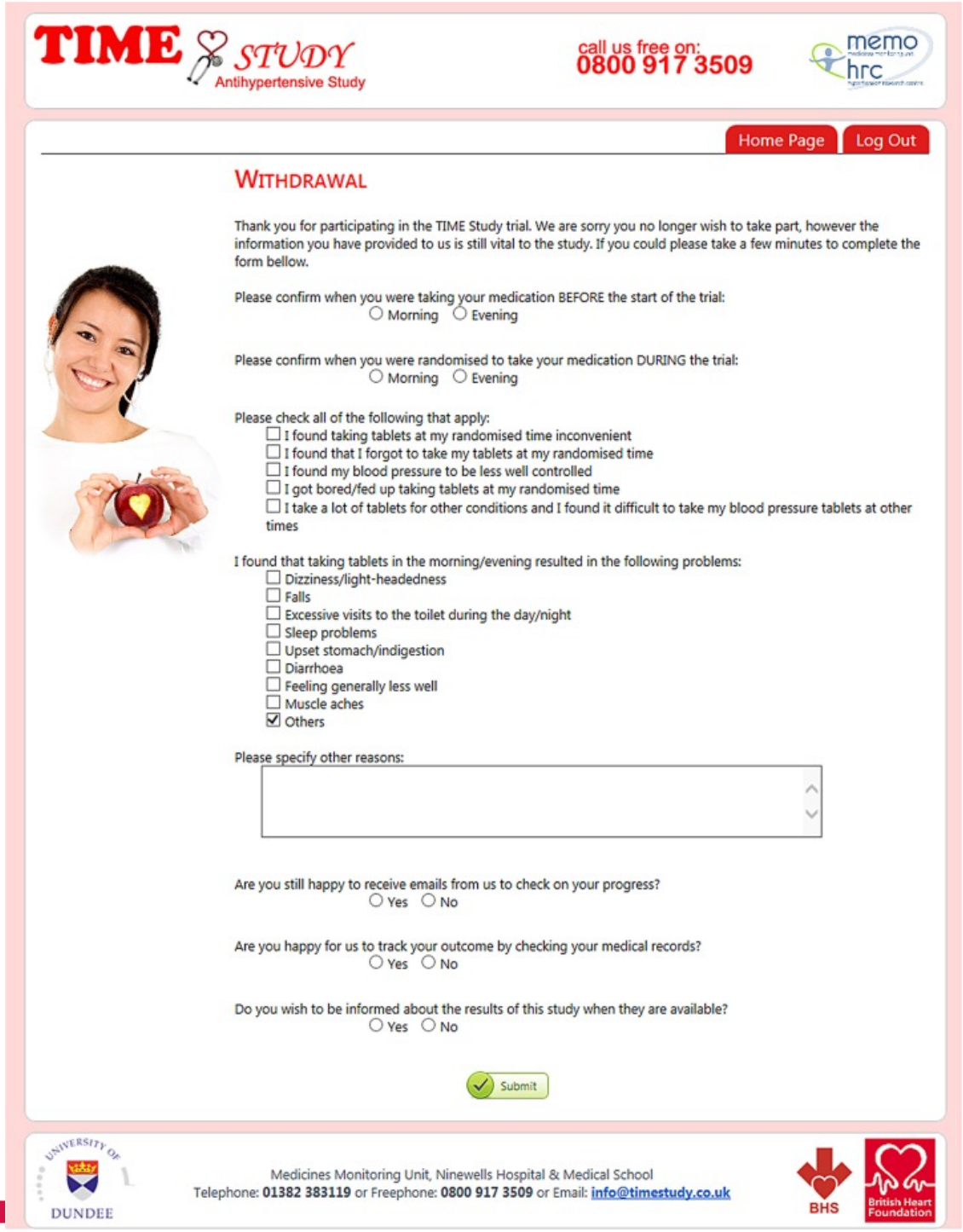
Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **Figure 4: Follow up form**



Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **Figure 5: Withdrawal form**



Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **ANALYSIS:**
- The trial is expected to run for 5 years, randomising 10,269 participants, with average participant follow-up being 4 years.
- The primary end point is hospitalisation for the composite end point of non-fatal myocardial infarction (MI), non-fatal stroke (cerebrovascular accident; CVA) or any vascular death determined by record-linkage.
- Secondary end points are: each component of the primary end point, hospitalisation for non-fatal stroke, hospitalisation for non-fatal MI, cardiovascular death, all-cause mortality, hospitalisation or death from congestive heart failure.
- The primary outcome will be a comparison of time to first event comparing morning versus evening dosing using an intention-to-treat analysis. The sample size is calculated for a two-sided test to detect 20% superiority at 80% power.

Methods of a large prospective, randomised, open-label, blinded end-point study comparing morning versus evening dosing in hypertensive patients: the Treatment In Morning versus Evening (TIME) study. (Rorie DA, et al. BMJ Open. 2016)

- **My Conclusion:**
- Ability to run a pragmatic clinical trial, entirely online, direct to patient, thanks to modern web-based methods, approaches to consent and treatment, and UK's linkage of primary care records.
- **A very interesting model for all of us to watch…**

Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project (Beresniak, A, et al. Contemp. Clin. Trials. 2015)

- INTRODUCTION: The widespread adoption of electronic health records (EHR) provides a new opportunity to improve the efficiency of clinical research. The European EHR4CR (Electronic Health Records for Clinical Research) 4-year project has developed an innovative technological platform to enable the re-use of EHR data for clinical research. The objective of this cost-benefit assessment (CBA) is to assess the value of EHR4CR solutions compared to current practices, from the perspective of sponsors of clinical trials.
- MATERIALS AND METHODS: A CBA model was developed using an advanced modeling approach. The costs of performing three clinical research scenarios (S) applied to a hypothetical Phase II or III oncology clinical trial workflow (reference case) were estimated under current and EHR4CR conditions, namely protocol feasibility assessment (S1), patient identification for recruitment (S2), and clinical study execution (S3). The potential benefits were calculated considering that the estimated reduction in actual person-time and costs for performing EHR4CR S1, S2, and S3 would accelerate time to market (TTM). Probabilistic sensitivity analyses using Monte Carlo simulations were conducted to manage uncertainty.

Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project (Beresniak, A, et al. Contemp. Clin. Trials. 2015)

- RESULTS: Should the estimated efficiency gains achieved with the EHR4CR platform translate into faster TTM, the expected benefits for the global pharmaceutical oncology sector were estimated at euro161.5m (S1), euro45.7m (S2), euro204.5m (S1+S2), euro1906m (S3), and up to euro2121.8m (S1+S2+S3) when the scenarios were used sequentially.
- CONCLUSIONS: The results suggest that optimizing clinical trial design and execution with the EHR4CR platform would generate substantial added value for pharmaceutical industry, as main sponsors of clinical trials in Europe, and beyond.

Table 6
Cost–benefit assessment results.*

Scenarios (S)	Mean costs current practices (€) [SD]	Mean costs EHR4CR** (€) [SD]	Benefit (€)	Absolute cost–benefit§ (€)	Relative cost–benefit§§
S1: Protocol feasibility assessment	183,959 [24,035]	216,491 [44,999]	161,522,390	−161,305,616	0.01
S2: Patient identification for recruitment	48,142 [7,162]	185,393 [51,078]	45,712,633	−45,527,322	0.02
S3: Clinical Study execution and SAE reporting	2,448,030 [690,712]	1,597,410 [361,624]	1,906,506,416	−1,904,911,265	0.01
S1 + S2	–	402,111 [67,555]	204,573,175	−204,171,064	0.01
S1 + S2 + S3	–	1,999,261 [368,549]	2,121,810,208	−2,119,810,946	0.04

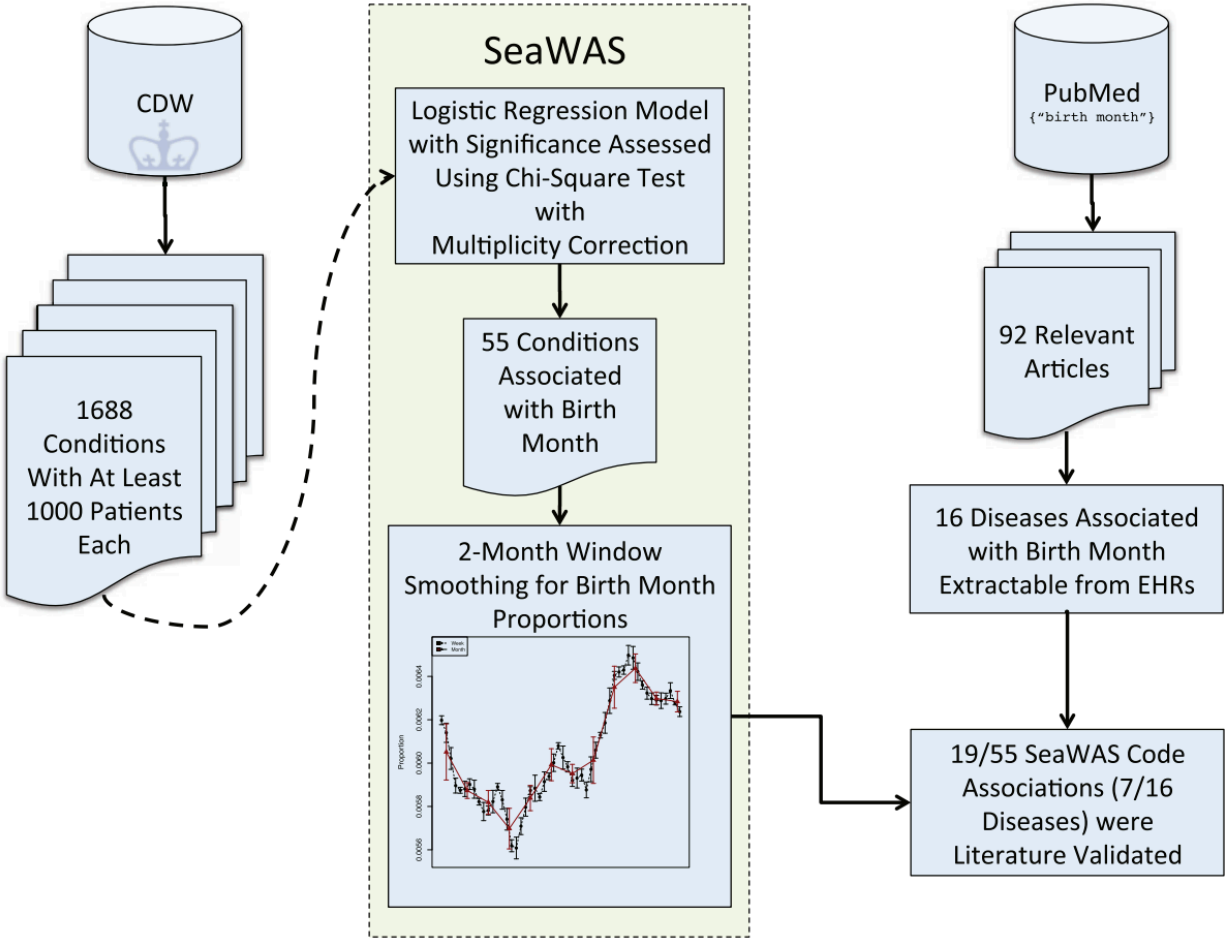
Birth Month Affects Lifetime Disease Risk: A Phenome Wide Study (Bowland, MR, et al. JAMIA. 2015)

- **Objective:** An individual's birth month has a significant impact on the diseases they develop during their lifetime. Previous studies reveal relationships between birth month and several diseases including atherothrombosis, asthma, attention deficit hyperactivity disorder, and myopia, leaving most diseases completely unexplored. This retrospective population study systematically explores the relationship between seasonal affects at birth and lifetime disease risk for 1688 conditions.

- **Methods:** We developed a hypothesis-free method that minimizes publication and disease selection biases by systematically investigating disease-birth month patterns across all conditions. Our dataset includes 1 749 400 individuals with records at New York-Presbyterian/Columbia University Medical Center born between 1900 and 2000 inclusive. We modeled associations between birth month and 1688 diseases using logistic regression. Significance was tested using a chi-squared test with multiplicity correction.

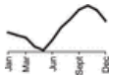
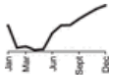
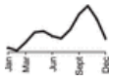
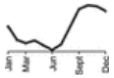
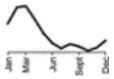
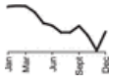
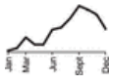
Birth Month Affects Lifetime Disease Risk: A Phenome Wide Study (Bowland, MR, et al. JAMIA. 2015)

Figure 1: Overview of the SeaWAS algorithm. The algorithm takes all 1688 conditions as initial input, finds significant associations over all months, then it models each birth month's association with the condition by smoothing the birth month proportions using a 2-month window. We then extracted all relevant birth month articles from PubMed (n = 92) and mapped the results to extractable codes from electronic health records. SeaWAS found 7 of the 16 diseases reported as associated with birth month in the literature corresponding to 19/55 associated codes.



Birth Month Affects Lifetime Disease Risk: A Phenome Wide Study (Bowland, MR, et al. JAMIA. 2015)

- **Results:** We found 55 diseases that were significantly dependent on birth month.
- Of these 19 were previously reported in the literature (P < . 001), 20 were for conditions with close relationships to those reported, and
- 16 were previously unreported. We found distinct incidence patterns across disease categories.

EHR Condition in SeaWAS	N	Passed Internal Validation?	Adjusted *P*[1]	Seasonal Pattern	Birth Month Risk	
					High	Low
Other (n = 7)						
Acute upper respiratory infection	112 487	Yes	<0.001		October	May
Bruising	8904	Yes	0.015		December	April
Nonvenomous insect bite	7435	Yes	0.001		October	February
Venereal disease screening	69 764	Yes	0.003		October	June
Primary malignant neoplasm of prostate	20 353	Yes	0.002		March	October
Malignant neoplasm of overlapping lesion of bronchus and lung	2714	Yes	0.014		February	November
Vomiting	30 495	No	0.029		September	January

Birth Month Affects Lifetime Disease Risk: A Phenome Wide Study (Bowland, MR, et al. JAMIA. 2015)

- **Conclusions:** Lifetime disease risk is affected by birth month.
- Seasonally dependent early developmental mechanisms may play a role in increasing lifetime risk of disease.

Other notable papers in this (Methods/Systems) category:

- **Data management in clinical research: Synthesizing stakeholder perspectives** (Johnson SB, et al. JBI. 2016)
- **OBJECTIVE:** This mixed-methods employing sub-language analysis in an innovative manner to study data management needs in clinical research from the perspectives of researchers, software analysts and developers.
- **FINDINGS:** Researchers perceive tasks related to study execution, analysis and quality control as highly strategic, in contrast with tactical tasks related to data manipulation. Researchers have only partial technologic support for analysis and quality control, and poor support for study execution.
- **CONCLUSION:** Software for data integration and validation appears critical to support clinical research, but may be expensive to implement.
- Features to support study workflow, collaboration and engagement have been underappreciated, but may prove to be easy successes
  - Potential low hanging fruit for software developers!

Other notable papers in this (Methods/Systems) category:

- **How accurately does the VIVO Harvester reflect actual Clinical and Translational Sciences Award-affiliated faculty member publications?** (Eldredge JD, et al. J Med Libr Assoc. 2015)
  - Spoiler alert: Not very accurately!
  - More work needed here

- **Understanding data requirements of retrospective studies** (Shenvi, EC. Int J. Med. Inform. 2015)
  - Nice overview of the issues we should consider when facilitating access to data for retrospective studies

Other notable papers in this (Methods/Systems) category:

- **Feasibility and utility of applications of the common data model to multiple, disparate observational health databases** (Voss, EA. J. Am Med Inform Assoc. 2015.)
  - Standardizing to OMOP CDM improved data quality, increased efficiency, and facilitated cross-database comparisons to support a more systematic approach to observational research. Comparisons across data sources showed consistency in the impact of inclusion criteria, using the protocol and identified differences in patient characteristics and coding practices across databases.
  - CONCLUSION: Standardizing data structure (through a CDM), content (through a standard vocabulary with source code mappings), and analytics can enable an institution to apply a network-based approach to observational research across multiple, disparate observational health databases.
- **Online accesses to medical research articles on publication predicted citations up to 15 years later** (Perneger, TV. J. Clin Epidemiol. 2015)
  - CONCLUSION: Early interest in a medical research article, reflected by online accesses within a week of publication, predicts citations up to 15 years later. This strengthens the validity of online usage as a measure of the scientific merit of publications.

Participant Recruitment and Eligibility

The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials.

(Ateya MB, et al. BMC Med Inform Decis Mak. 2016)

- **BACKGROUND:** An increasing number of clinical trials are conducted in primary care settings. Making better use of existing data in the electronic health records to identify eligible subjects can improve efficiency of such studies
- Aim: to quantify the proportion of eligibility criteria that can be addressed with data in electronic health records and to compare the content of eligibility criteria in primary care with previous work.
- **METHODS:** Eligibility criteria were extracted from primary care studies downloaded from the UK Clinical Research Network Study Portfolio.
- Criteria were broken into elemental statements. Two expert independent raters classified each statement based on whether or not structured data items in the electronic health record can be used to determine if the statement was true for a specific patient.
- Statements were also classified based on content and the percentages of each category were compared to two similar studies reported in the literature.

The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials.

(Ateya MB, et al. BMC Med Inform Decis Mak. 2016)

- **RESULTS:** Eligibility criteria were retrieved from 228 studies and decomposed into 2619 criteria elemental statements. **74 % of the criteria elemental statements were considered likely associated with structured data in an electronic health record.**
- **79 % of the studies had at least 60 % of their criteria statements addressable with structured data likely to be present in an electronic health record**

Table 2

Percentages of likely present CES per study

Percentage of Likely Present CES per Study	Number of Studies (%)
=100 %	31 (14)
≥80–<100 %	76 (33)
≥60–<80 %	74 (32)
≥40–<60 %	31 (14)
≥20–<40 %	10 (4)
≥0–<20 %	6 (3)
Total	228 (100)

The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials.

(Ateya MB, et al. BMC Med Inform Decis Mak. 2016)

- **RESULTS:** Based on clinical content, most frequent categories were: "disease, symptom, and sign", "therapy or surgery", and "medication" (36 %, 13 %, and 10 % of total criteria statements respectively). Also identified new criteria categories related to provider and caregiver

Table 3

Eligibility criteria classification based on semantic categories

CES Category	This Study[a] % (n)	Weng et al. %	Köpcke et al. %
Medical Condition (Health Status)			
Disease, symptom and sign	36 % (944)	28 %	22.52 %
Pregnancy conditions	4 % (110)	3 %	5.24 %
Allergy	2 % (42)	1 %	5.95 %
Disease stage	1 % (20)	6 %	2.27 %
Cancer	0.3 % (9)	12 %	3.4 %
Organ or tissue status	0.1 % (2)	1 %	5.38 %
Life expectancy	0.1 % (2)	0 %	0.38 %
Treatment or Healthcare			
Therapy or surgery	13 % (352)	15 %	10.2 %
Medication (Pharmaceutical substance or drug)	10 % (255)	17 %	7.37 %
Device	0.5 % (14)	0 %	0 %
Diagnostic or Lab Tests			
Diagnostic or lab results	5 % (129)	14 %	19.41 %
Receptor status	0 % (0)	0 %	0 %
Demographics			

The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials.

(Ateya MB, et al. BMC Med Inform Decis Mak. 2016)

- **CONCLUSIONS:** Electronic health records readily contain much of the data needed to assess patients' eligibility for clinical trials enrollment.
- Eligibility criteria content categories identified by this study can be incorporated as data elements in electronic health records to facilitate their integration with clinical trial management systems.

- More information to inform improvements for research use of EHRs

Assessing the readability of ClinicalTrials.gov (Wu DTY, et al. JAMIA. 2015)

- **Objective**: ClinicalTrials.gov serves critical functions of disseminating trial information to the public and helping the trials recruit participants. This study assessed the readability of trial descriptions at ClinicalTrials.gov using multiple quantitative measures.
- **Materials and Methods** The analysis included all 165 988 trials registered at ClinicalTrials.gov as of April 30, 2014. To obtain benchmarks, the authors also analyzed 2 other medical corpora:
  - (1) all 955 Health Topics articles from MedlinePlus and
  - (2) a random sample of 100 000 clinician notes retrieved from an electronic health records system intended for conveying internal communication among medical professionals.
  - The authors characterized each of the corpora using 4 surface metrics, and then applied 5 different scoring algorithms to assess their readability. The authors hypothesized that clinician notes would be most difficult to read, followed by trial descriptions and MedlinePlus Health Topics articles.

Assessing the readability of ClinicalTrials.gov (Wu DTY, et al. JAMIA. 2015)

- **Results:** Trial descriptions have the longest average sentence length (26.1 words) across all corpora;
- 65% of their words used are not covered by a basic medical English dictionary.
- In comparison, average sentence length of MedlinePlus Health Topics articles is 61% shorter, vocabulary size is 95% smaller, and dictionary coverage is 46% higher.

Assessing the readability of ClinicalTrials.gov (Wu DTY, et al. JAMIA. 2015)

- **Results**
- All 5 scoring algorithms consistently rated ClinicITrials.gov trial descriptions the most difficult corpus to read, even harder than clinician notes.
- On average, it requires 18 years of education to properly understand these trial descriptions according to the results generated by the readability assessment algorithms.

Table 2: Readability scores

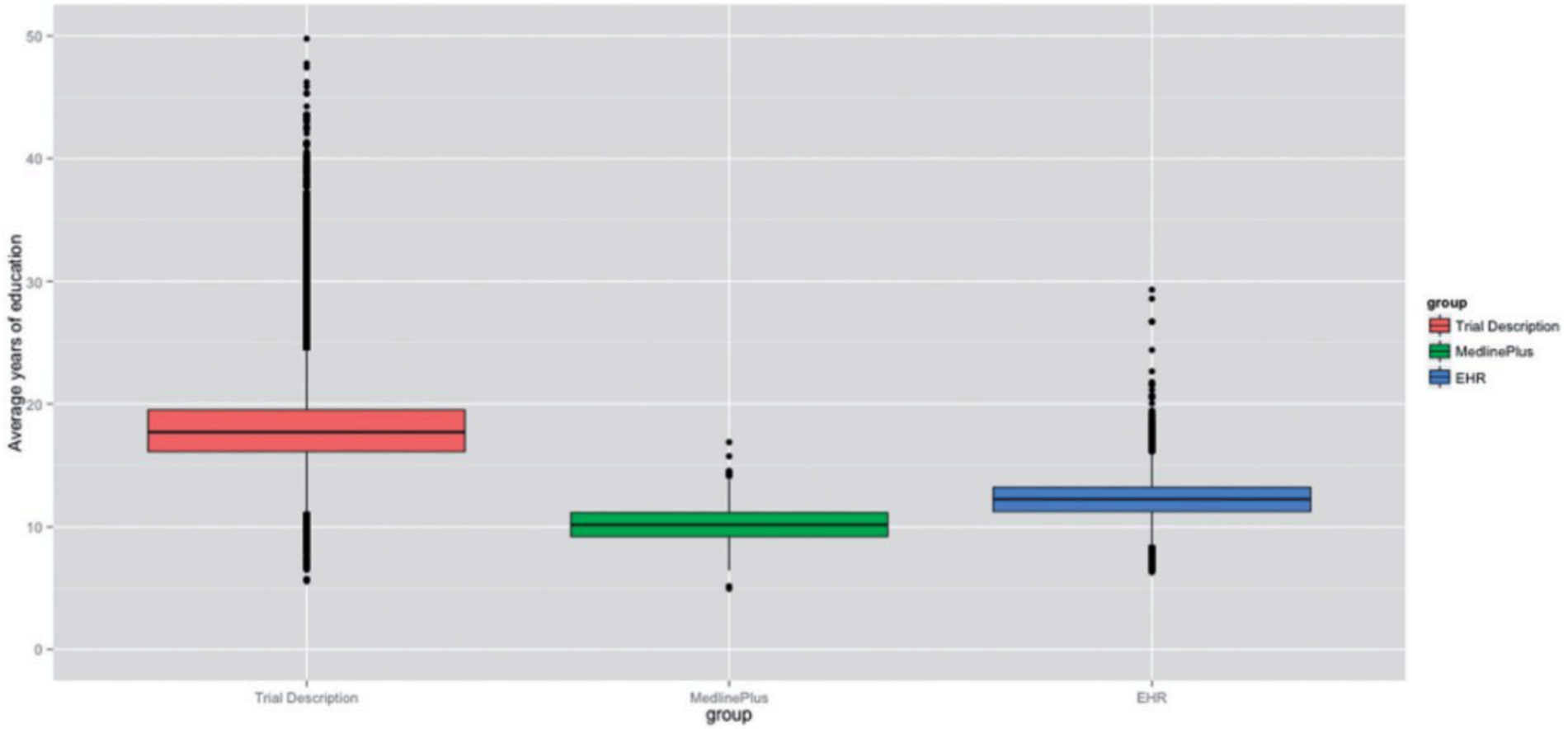
Scoring algorithm	Trial Description	MedlinePlus	EHR
NDC	15.8 ± 0.8 👎	11.3 ± 2.3 👍	15.1 ± 1.7
FKGL	17.2 ± 4.2 👎	8.0 ± 1.4 👍	9.1 ± 1.9
SMOG	17.9 ± 3.1 👎	10.7 ± 1.3 👍	11.9 ± 1.5
GFI	21.1 ± 4.6 👎	10.9 ± 1.9 👍	12.9 ± 2.2
Average of NDC, FKGL, SMOG, and GFI	18.0 ± 3.0 👎	10.2 ± 1.5 👍	12.2 ± 1.5
MSRM	−0.44 ± 0.28 👎	− 0.10 ± 0.23 👍	− 0.36 ± 0.18

👍 Best readability; 👎 Worst readability.

Abbreviations: NDC, New Dale-Chall; FKGL, Flesch-Kincaid Grade Level; SMOG, Simple Measure of Gobbedygook; GFI, Gunning-Fog Index; MSRM, Medical-Specific Readability Measure.

Assessing the readability of ClinicalTrials.gov (Wu DTY, et al. JAMIA. 2015)

- **Conclusion:** Trial descriptions at ClinicITrials.gov are extremely difficult to read. Significant work is warranted to improve their readability in order to achieve ClinicalTrials.gov's goal of facilitating information dissemination and subject recruitment.



Other notable papers in this (Recruitment) category:

- **Initial Readability Assessment of Clinical Trial Eligibility Criteria** (Kang, T, et al. AMIA Ann Symp. 2015)
  - Similar finding to previous paper. Overall use of technical jargon leads to a college reading necessary to understand eligibility criteria text.
- **Design, development and deployment of a Diabetes Research Registry to facilitate recruitment in clinical research** (Tan MH, et al. Cont Clin Trials. 2016)
  - After 5 years, the DRR has over 5000 registrants. DRR matches potential subjects interested in research with approved clinical studies using study entry criteria abstracted from their EHR. By providing large lists of potentially eligible study subjects quickly, the DRR facilitated recruitment in 31 clinical studies.

Other notable papers in this (Recruitment) category:

- **Videoconferencing for site initiations in clinical studies: Mixed methods evaluation of usability, acceptability, and impact on recruitment** (Randell R, et al. Informatics for Health and Social Care. 2015)
  - Site-visits for trail initiations can be time consuming and costly. The group performed formal usability study and interviews and found very good usability and acceptance of using an off-the-shelf solution to initiate studies – across multiple end-users. Some variation based on hardware.
- **Visual aggregate analysis of eligibility features of clinical trials** (He, et al, JBI, 2015)
  - Goal: to develop a method for profiling the collective populations targeted for recruitment by multiple clinical studies addressing the same medical condition using one eligibility feature each time. Tool enables visual aggregate analysis of common clinical trial eligibility features and includes four modules for eligibility feature frequency analysis, query builder, distribution analysis, and visualization, respectively.
  - Users found it usable and potential useful. Another tool in our arsenal!

Other notable papers in this (Recruitment) category:

- **Simulation-based Evaluation of the Generalizability Index for Study Traits (GIST).** (He et al. AMIA Annu Fall Symp 2015) *(Distinguished Paper Award)*
- The paper addresses an important and yet understudied representativeness issue that occurs when certain population subgroups are systematically underrepresented in clinical studies across major medical conditions.
- It demonstrates the validity and effectiveness of a quantitative metric called Generalizability Index for Study Traits (GIST), which is able to assess the population representativeness of a set of related clinical trials.
- It quantified the population representativeness of a set of trials that differed in disease domains, study phases, sponsor types and study designs. Simulation-based evaluation experiments show that this GIST metric is reliable and truthful.
- **Contribution:** Provides a reliable quantitative metric for improving the transparency in potential biases in clinical research study population definitions. Could health consumers assess the applicability of clinical research evidence.

EHRs and Learning Health Systems

Operationalizing the Learning Health Care System in an Integrated Delivery System (Psek WA, et al. eGEMs. 2015)

- **INTRODUCTION:** Enabling Learning Health Care Systems presents many challenges but can yield opportunities for continuous improvement.
- At Geisinger Health System (GHS) a multi-stakeholder group is undertaking to enhance organizational learning and develop a plan for operationalizing the LHCS system-wide.
- They present here a framework for operationalizing continuous learning across an integrated delivery system and lessons learned through the ongoing planning process.

Operationalizing the Learning Health Care System in an Integrated Delivery System (Psek WA, et al. eGEMs. 2015)

- **FRAMEWORK:** The framework focuses attention on nine key LHCS operational components:
  - Data and Analytics;
  - People and Partnerships;
  - Patient and Family Engagement;
  - Ethics and Oversight;
  - Evaluation and Methodology;
  - Funding;
  - Organization;
  - Prioritization; and
  - Deliverables.
  - Definitions, key elements and examples for each are presented. The framework is purposefully broad for application across different organizational contexts.

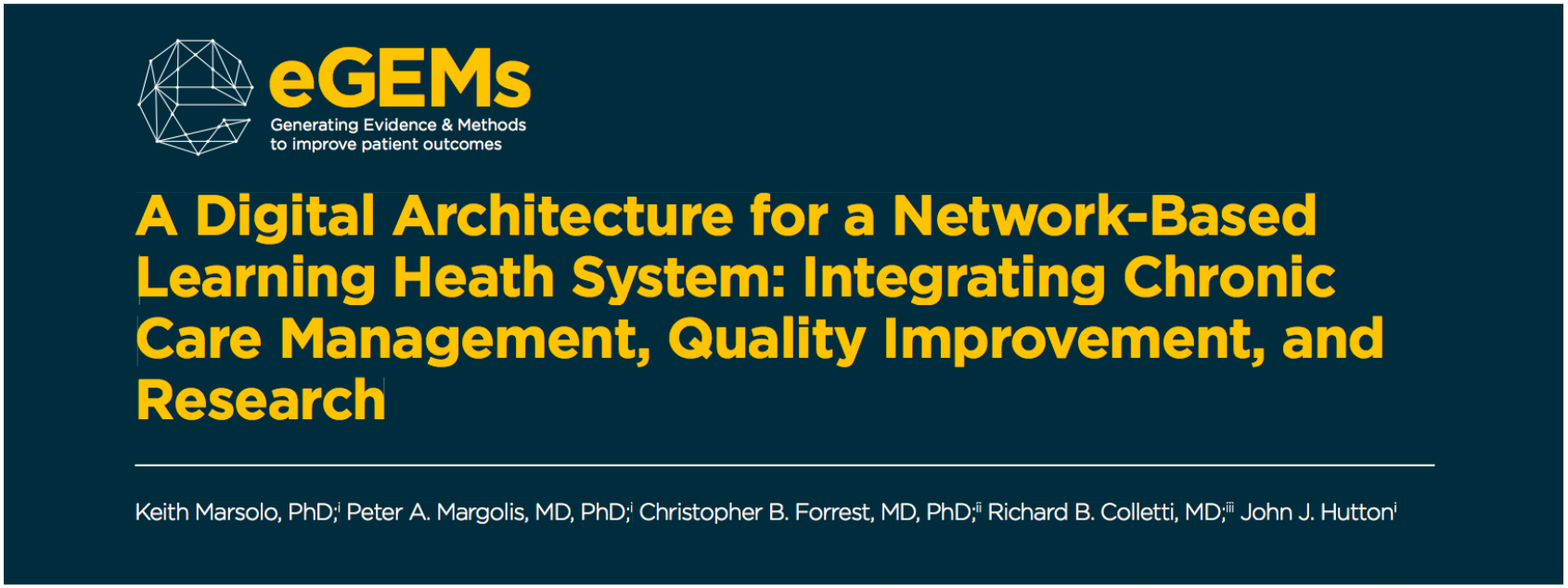
Operationalizing the Learning Health Care System in an Integrated Delivery System (Psek WA, et al. eGEMs. 2015)

- **CONCLUSION:** A realistic assessment of the culture, resources and capabilities of the organization related to learning is critical to defining the scope of operationalization.
- Engaging patients in clinical care and discovery, including quality improvement and comparative effectiveness research, requires a defensible ethical framework that undergirds a system of strong but flexible oversight.
- Leadership support is imperative for advancement of the LHCS model.
- Should inform other organizations considering a transition to an LHCS.

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

- **Introduction:** ImproveCareNow Network to create a proof-of-concept architecture for a network-based Learning Health System.
- Collaboration involved transitioning an existing registry to one that is linked to the electronic health record (EHR), enabling a “data in once” strategy.



A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

- Sought to automate a series of reports that support care improvement while also demonstrating the use of observational registry data for comparative effectiveness research.
- **Goals**:
  - Collect registry data through EHR/routine practice
  - Deploy automated reports to support care, QI, increasing value for all
  - Demonstrate ability to observational data for CER
  - Test feasibility of federated registry architecture – with all benefits
  - Pilot informatics tools to increase patient/family engagement

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

- **Table 1**
- Selection of the most important functional requirements of the network-based Learning Health System
- Along with the components of the architecture that satisfy each functional requirement.

Table 1. Functional Requirements and Corresponding Architectural Components of the Network-Based Learning Health System

FUNCTIONAL REQUIREMENT	ARCHITECTURE COMPONENT
DATA CAPTURE	
Clinical documentation functionality (discretely capture data at the point of care)	EHR-based data collection forms—see Implementation section, subsection (*C*) below
Electronic transfer of clinical documentation responses to the registry	Electronic data transfer (see *D*)
Form responses added to a progress note or referral letter	EHR-based data collection forms with links to note templates (see *C*)
Ability for all centers to participate, regardless of EHR maturity; ability to capture non-EHR data	Web-based data collection forms (see *B*)
Capture of all common data elements for the condition of interest	Process for defining outcome measures and data elements necessary for computation; process for calculating derived variables (see *E*)
MANAGING CLINICAL CARE AND QUALITY IMPROVEMENT (QI)	
QI & chronic care management reports with daily refresh	Automated reporting—performance measurement for quality improvement, pre-visit planning, and population management (see *F*)
Reports to monitor data entry compliance	Automated reporting—data quality and exception reports (see *F*)
USE OF DATA FOR RESEARCH	
Ability to use the data to support clinical care, QI, and research	Standardized IRB protocols, Data Use and Business Associate Agreements (see *A*)

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

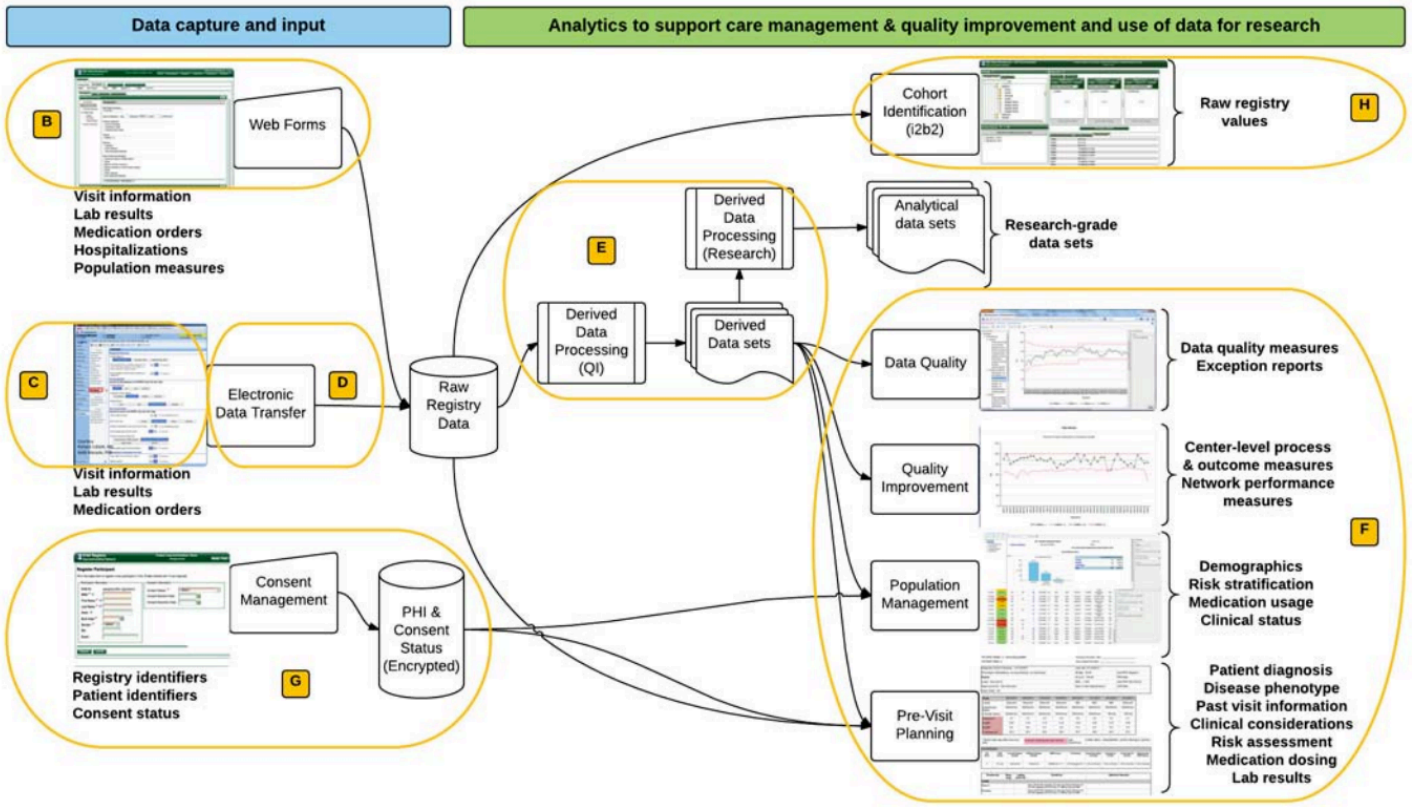
(Marsolo K, et al. eGEMs. 2015)

- **Architecture:**
  - Worked with 3 leading EHR vendors to create standardized EHR-based data collection forms
  - Automated many of ImproveCareNow's analytic reports and developed an application for storing protected health information and tracking patient consent.
  - Deployed a cohort identification tool (based upon i2b2) to support feasibility studies and hypothesis generation.

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

Figure 1. Functional Architecture of the Learning Health System that Supports the ImproveCareNow Network



Note: The components that support data input are listed on the left and the components that support research and analytics on the right. All tools are accessible from the same user interface. The labeled ovals correspond to the subheading of the Implementation section where a more detailed description can be found. The informatics components described in Table 1 that support patient engagement are not shown.

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

- **Results:**
  - As of pub, 31 centers had adopted the EHR-based forms and 21 centers are uploading data to the registry.
  - High usage of reports
  - Investigators using cohort identification tools for several clinical trials
- **Lessons/future directions:**
  - The current process for creating EHR-based data collection forms requires groups to work individually with each vendor.
  - A vendor-agnostic model would allow for more rapid uptake.
  - Interfacing network-based registries with the EHR would allow them to serve as a source of decision support - but additional standards are needed to achieve this

A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research

(Marsolo K, et al. eGEMs. 2015)

- **Conclusions:**
- Creating the learning health system is highly complex, information intensive, and lacking in best practices
- Many component parts
  - Technical, Organizational, Workflows, Legal/regulatory, etc.
  - A great model for Informatics – as achieving this requires all our skills/ knowledge
- Progress is being made!

- An excellent article outlining, using real-world experience, some of the key elements to architecting a network-based LHS

Systematic reviews examining implementation of research into practice and impact on population health are needed

(Yoong SL, et al. J Clin Epidemiol. 2015)

- **Objective:** Examine the research translation phase focus (T1-T4) of systematic reviews published in the Cochrane Database of Systematic Reviews (CDSR) and Database of Abstracts of Reviews of Effects (DARE).
  - T1 includes reviews of basic science experiments;
  - T2 includes reviews of human trials leading to guideline development;
  - T3 includes reviews examining how to move guidelines into policy and practice;
  - T4 includes reviews describing the impact of changing health practices on population outcomes.

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 68 (2015) 788–791

Journal of Clinical Epidemiology

REVIEW ARTICLES

Systematic reviews examining implementation of research into practice and impact on population health are needed

Sze Lin Yoong[a,b,c,*], Tara Clinton-McHarg[a,b,c], Luke Wolfenden[a,b,c]

[a]Hunter New England Population Health, Hunter New England Local Health District, Booth Building, Longworth Avenue, Wallsend, New South Wales 2287, Australia

[b]School of Medicine and Public Health, The University of Newcastle, University Drive, Callaghan, New South Wales 2308, Australia

[c]Priority Research Centre for Health Behaviour, The University of Newcastle, Callaghan, New South Wales 2308, Australia

Accepted 23 December 2014; Published online 30 December 2014

Systematic reviews examining implementation of research into practice and impact on population health are needed

(Yoong SL, et al. J Clin Epidemiol. 2015)

- **Methods**: A cross-sectional audit of randomly selected reviews from CDSR (n = 500) and DARE (n = 500). The research translation phase of reviews, overall and by communicable disease, noncommunicable disease, and injury subgroups, were coded by two researchers.

- **Findings:**
  - 898 reviews examined a communicable, non-communicable, or injury-related condition.
  - Of those, **98% of reviews within CDSR focused on T2**, and the remaining 2% focused on T3.
  - In DARE, **88% focused on T2**, 8.7% focused on T1, 2.5% focused on T3, and 1.3% focused on T4.
  - Almost all reviews examining communicable (CDSR 100%, DARE 93%), noncommunicable (CDSR 98%, DARE 87%), and injury (CDSR 95%, DARE 88%) were also T2 focused.

Systematic reviews examining implementation of research into practice and impact on population health are needed

(Yoong SL, et al. J Clin Epidemiol. 2015)

- **Conclusion and Implications:**
- **Few** reviews exist to guide practitioners and policy makers with implementing evidence-based treatments or programs.
- Few T3 (research examining how to move evidence-based guidelines into practice) and T4 (research examining the impact of implementation of evidence-based guidelines on population health) focused systematic reviews are available in CDSR and DARE.
- Practitioners and policy makers may have little evidence to guide practice and policy decisions.
- **Efforts to increase the production of T3 and T4 systematic reviews including issuing targeted calls for such reviews, establishing funding schemes, and creating more specialist journals for dissemination are needed.**

Other notable papers in this (EHRs/LHS) category:

- **Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care**. (Wagner LI, et al. Cancer. 2015)
  - 636 women under gyn-onc care received instructions to complete clinical assessments through Epic MyChart. Patient Reported Outcomes Measurement Information System (PROMIS) computer adaptive tests (CATs) were administered to assess fatigue, pain interference, physical function, depression, and anxiety.
  - 4,042 MyChart messages sent, and 3203 (79%) were reviewed by patients.
  - 1493 patients (37%) started, and 93% (1386 patients) completed
  - 49.8% who reviewed the MyChart message completed the assessment.
- **Conclusion:** Demonstration of increasingly available/embedded, direct-to-patient delivery of validated instruments that can help both with improving care, managing populations, and collecting PRO data for downstream research

Other notable papers in this (EHRs/LHS) category:

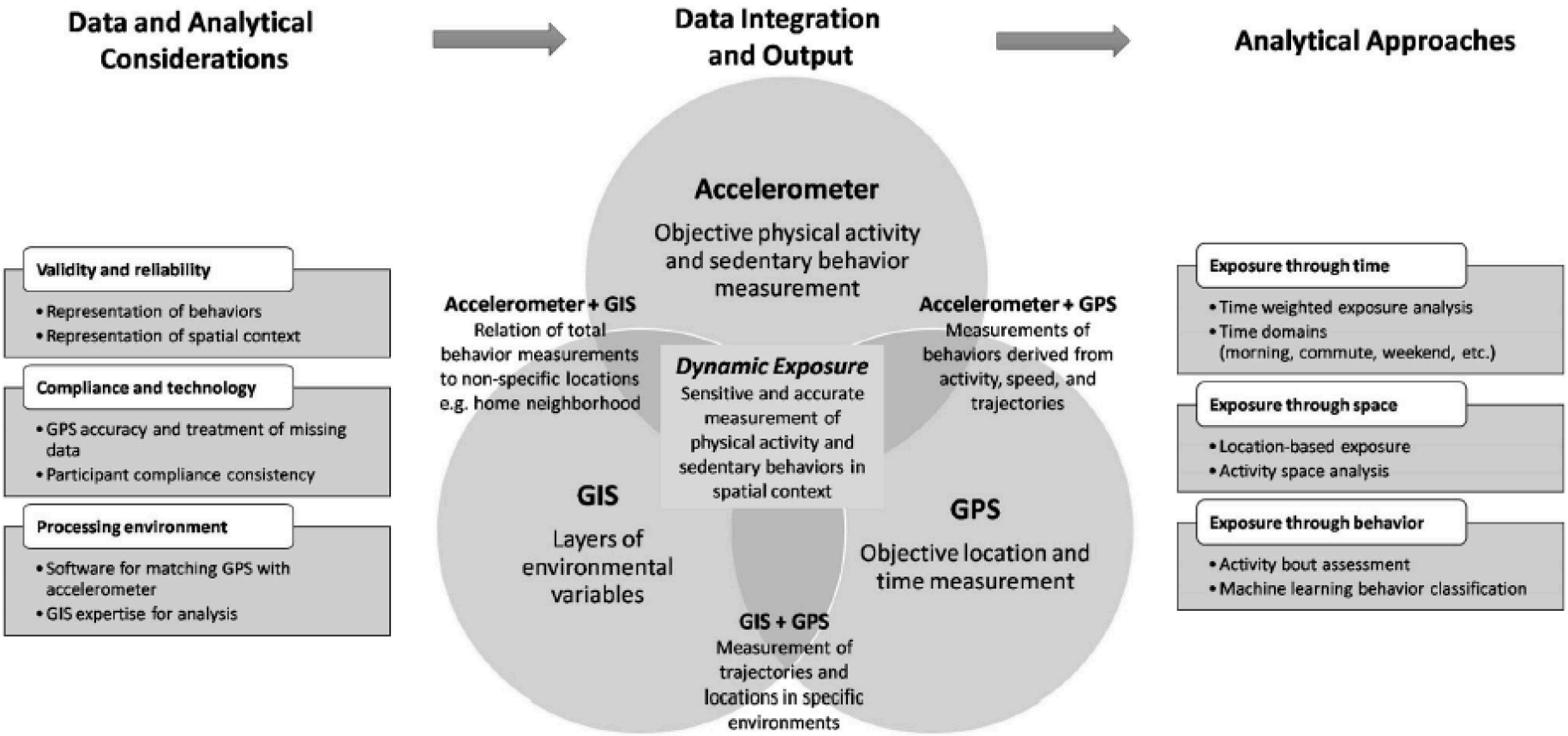
- **The geographic distribution of cardiovascular health in the stroke prevention in healthcare delivery environments (SPHERE) study** (Roth C, et al. JBI. 2016)
  - Embedded, third-party decision support tools, rendered into EHR environment to facilitate driving evidence-to practice, can also yield data that when combined with community derived variables can generate evidence/insights for further care improvements.
- **Using Health Information Technology to Support Quality Improvement in Primary Care** (White Paper. Prepared for AHRQ; Higgins, TC et al. Mathematica Policy Research Princeton. 2015)
  - Guidance toward furthering use of HIT to improve quality of care in primary care. Further informing the connection to the “virtuous cycle” of evidence-generating-medicine (EGM) and evidence-based-medicine (EBM)
- **Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework to Address Challenges** (Amarasingham R, et al, eGEMs 2016)
  - Useful read as CRI increasingly called upon to address such issues

CRI Trends (1): Mobile Data for Research

- Ongoing work on platforms, like Apple's Research Kit, and…
  - Increasing consideration of how to use mobile data directly for research
  - We are looked to increasingly to sort out how to do it… some progress:
- **A framework for using GPS data in physical activity and sedentary behavior studies** (Jankowska, MM, et al. Exerc Sport Sci Rev. 2015)
  - Global Positioning Systems (GPS) are applied increasingly in activity studies, yet significant theoretical and methodological challenges remain. This article presents a framework for integrating GPS data with other technologies to create dynamic representations of behaviors in context. Using more accurate and sensitive measures to link behavior and environmental exposures allows for new research questions and methods to be developed.

CRI Trends (1): Mobile Data for Research

- **A framework for using GPS data in physical activity and sedentary behavior studies** (Jankowska, MM, et al. Exerc Sport Sci Rev. 2015)



CRI Trends (1): Mobile Data for Research

- **The mobilize center: an NIH big data to knowledge center to advance human movement research and improve mobility** (Ku JP, et al. JAMIA. 2015)
  - Vast amounts of data characterizing human movement are available from research labs, clinics, and millions of smartphones and wearable sensors, but integration and analysis of this large quantity of mobility data are extremely challenging.
  - Authors have established the Mobilize Center to harness these data to improve human mobility and help lay the foundation for using data science methods in biomedicine.
  - The Center is organized around 4 data science research cores:
    - biomechanical modeling,
    - statistical learning,
    - behavioral and social modeling, and
    - integrative modeling.
  - Important biomedical applications, such as osteoarthritis and weight management, will focus the development of new data science methods. (http://mobilize.stanford.edu)

CRI Trends (2): Centralized Clinical Research Data/ Resources

- **OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data** (Kass-Hout, TA. JAMIA 2015.)
  - **Objective:** to facilitate access and use of big important Food and Drug Administration public datasets by developers, researchers, and the public through harmonization of data across disparate FDA datasets provided via application programming interfaces (APIs). Since June 2014, openFDA has developed four APIs for drug and device adverse events, recall information for all FDA-regulated products, and drug labeling.
  - There have been **more than 20 million API calls** (more than half from outside the United States), 6000 registered users, 20,000 connected Internet Protocol addresses, and dozens of new software (mobile or web) apps developed.
  - A case study demonstrates a use of openFDA data to understand an apparent association of a drug with an adverse event.

CRI Trends (2): Centralized Clinical Research Data/ Resources

- **Unveiling SEER-CAHPS(R): a new data resource for quality of care research** (Chawla N, et al. JGIM. 2015)
  - Describe a new combined resource for enabling quality of care research based on a linkage between the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) patient surveys and the NCI's Surveillance, Epidemiology and End Results (SEER) data.
  - In total, 150,750 respondents in the cancer cohort and 571,318 were in a non-cancer cohort. The data linkage includes SEER data on cancer site, stage, treatment, death information, and patient demographics, in addition to longitudinal data from Medicare claims and information on patient experiences from CAHPS surveys.
  - Result is **a valuable new resource for information about Medicare beneficiaries' experiences of care across different diagnoses and treatment modalities, and enables comparisons by type of insurance.**

CRI Trends (2): Centralized Clinical Research Data/ Resources

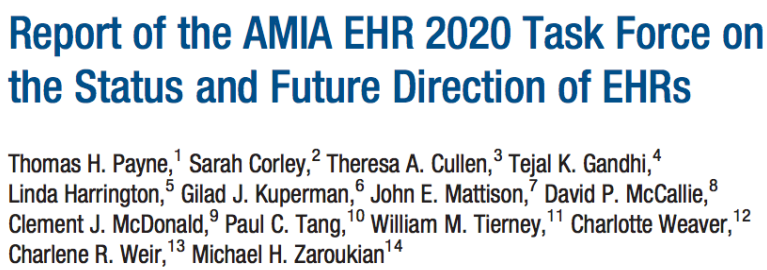
- **NIH/NCATS/GRDR(R) Common Data Elements: A leading force for standardized data collection** (Rubinstein, YR et al. Contemp Clin Trials. 2015)
- The main goal of the NIH/NCATS GRDR(R) program is to serve as a central web-based global data repository to integrate de-identified patient clinical data from rare disease registries, and other data sources, in a standardized manner, to be available to researchers for conducting various biomedical studies, including clinical trials and to support analyses within and across diseases.
- The aim of the program is to advance research for many rare diseases.
- One of the first tasks toward achieving this goal was the development of a set of Common Data Elements (CDEs). A list of 75 CDEs was developed by a national committee and was validated and implemented during a period of 2 year proof of concept.
- **Effort seems to be advancing discussion of data standardization and interoperability for rare disease patient registries.** Notable work.

CRI Policy & Perspectives:

- **Harnessing next-generation informatics for personalizing medicine: a report from AMIA's 2014 Health Policy Invitational Meeting** (Wiley LK, et al. JAMIA. 2016)
- AMIA's 2014 Health Policy Invitational Meeting developed recommendations for updates to current policies and to establish an informatics research related to personalizing care through the integration of genomic or other high-volume biomolecular data with data from clinical systems to make health care more efficient and effective.
- Report summarizes the findings (n = 6) and recommendations (n = 15) from the policy meeting, which were clustered into 3 broad areas:
  - (1) policies governing data access for research and personalization of care;
  - (2) policy and research needs for evolving data interpretation and knowledge representation; and
  - (3) policy and research needs to ensure data integrity and preservation. The meeting outcome underscored the need to address a number of important policy and technical considerations in order to realize the potential of personalized or precision medicine in actual clinical contexts.

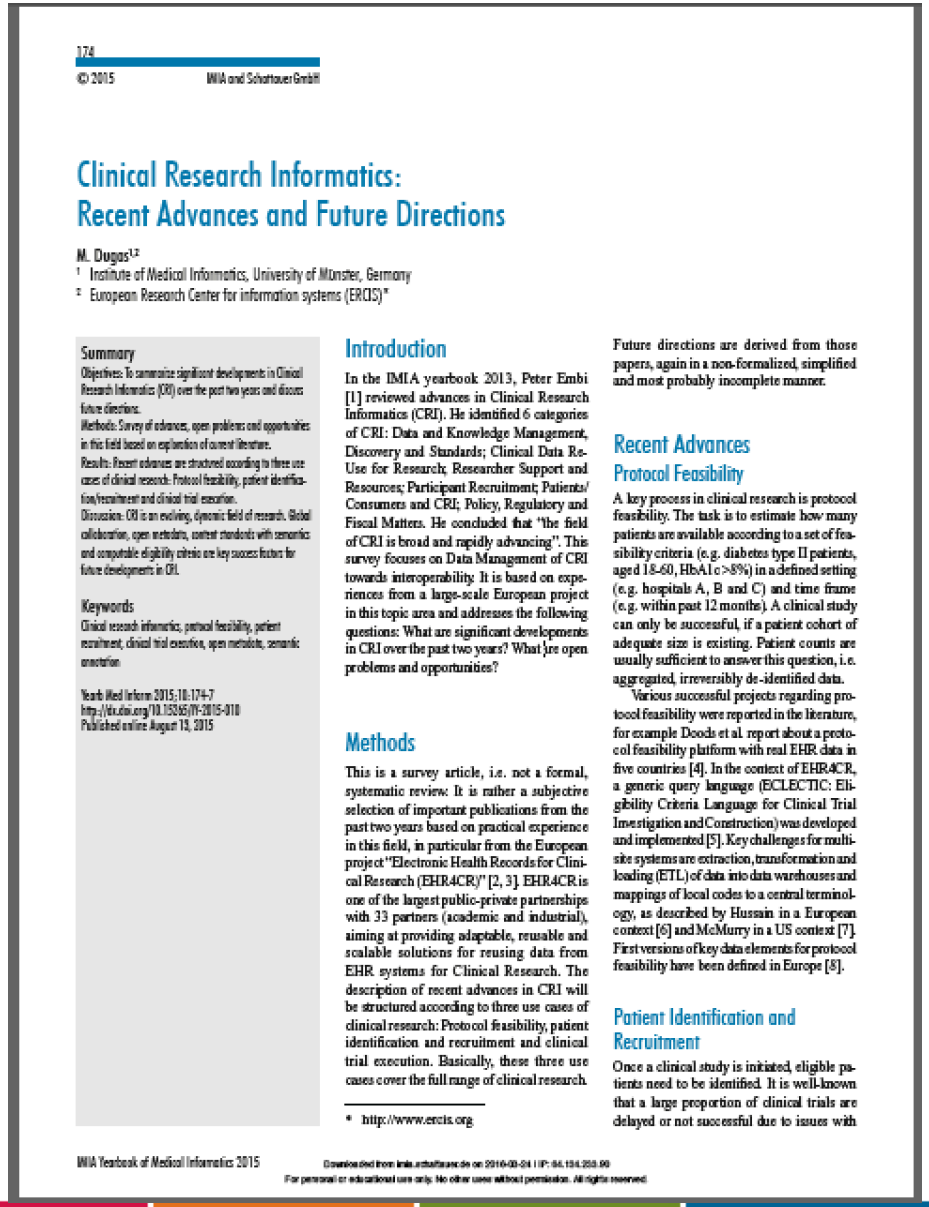
CRI Policy & Perspectives:

- **Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs.** (Payne T, et al. JAMIA 2015)
- 5 areas and 10 recommendation to improve EHRs, including some CRI-focused:
  - Simplify and Speed Documentation
    - **Enable systematic learning and research at point of care**
  - Refocus regulations
  - Increase transparency
  - Foster Innovation
    - **Support standards, APIs, etc. to enable informaticians and others to drive innovation and research**
  - Support Patient-Centered Care Delivery
    - **Enable precision medicine and new models of care**



CRI Policy & Perspectives:

- **Clinical Research Informatics: Recent Advances and Future Directions** (Dugas M. IMIA Yearbook of Medical Informatics. 2015)
- Summarizes developments in Clinical Research Informatics (CRI) over the past two years and discuss future directions.
- Recent advances structured according to three use cases of clinical research: Protocol feasibility, patient identification/ recruitment and clinical trial execution.
- Particular call-outs: global collaboration, open metadata, content standards with semantics and computable eligibility criteria as key success factors for future developments in CRI.



174

Clinical Research Informatics: Recent Advances and Future Directions

M. Dugas[1,2]
[1] Institute of Medical Informatics, University of Münster, Germany
[2] European Research Center for information systems (ERCIS)*

Summary

Objectives: To summarize significant developments in Clinical Research Informatics (CRI) over the past two years and discuss future directions.

Methods: Survey of advances, open problems and opportunities in this field based on exploration of current literature.

Results: Recent advances are structured according to three use cases of clinical research: Protocol feasibility, patient identification/recruitment and clinical trial execution.

Discussion: CRI is an evolving, dynamic field of research. Global collaboration, open metadata, content standards with semantics and computable eligibility criteria are key success factors for future developments in CRI.

Keywords

Clinical research informatics, protocol feasibility, patient recruitment, clinical trial execution, open metadata, semantic annotation

Yearb Med Inform 2015;10:174-7
http://dx.doi.org/10.15265/IY-2015-010
Published online August 13, 2015

Introduction

In the IMIA yearbook 2013, Peter Embi [1] reviewed advances in Clinical Research Informatics (CRI). He identified 6 categories of CRI: Data and Knowledge Management, Discovery and Standards; Clinical Data Re-Use for Research; Researcher Support and Resources; Participant Recruitment; Patients/ Consumers and CRI; Policy, Regulatory and Fiscal Matters. He concluded that "the field of CRI is broad and rapidly advancing". This survey focuses on Data Management of CRI towards interoperability. It is based on experiences from a large-scale European project in this topic area and addresses the following questions: What are significant developments in CRI over the past two years? What are open problems and opportunities?

Methods

This is a survey article, i.e. not a formal, systematic review. It is rather a subjective selection of important publications from the past two years based on practical experience in this field, in particular from the European project "Electronic Health Records for Clinical Research (EHR4CR)" [2, 3]. EHR4CR is one of the largest public-private partnerships with 33 partners (academic and industrial), aiming at providing adaptable, reusable and scalable solutions for reusing data from EHR systems for Clinical Research. The description of recent advances in CRI will be structured according to three use cases of clinical research: Protocol feasibility, patient identification and recruitment and clinical trial execution. Basically, these three use cases cover the full range of clinical research.

Future directions are derived from those papers, again in a non-formalized, simplified and most probably incomplete manner.

Recent Advances

Protocol Feasibility

A key process in clinical research is protocol feasibility. The task is to estimate how many patients are available according to a set of feasibility criteria (e.g. diabetes type II patients, aged 18-60, HbA1c >8%) in a defined setting (e.g. hospitals A, B and C) and time frame (e.g. within past 12 months). A clinical study can only be successful, if a patient cohort of adequate size is existing. Patient counts are usually sufficient to answer this question, i.e. aggregated, irreversibly de-identified data.

Various successful projects regarding protocol feasibility were reported in the literature, for example Doods et al. report about a protocol feasibility platform with real EHR data in five countries [4]. In the context of EHR4CR, a generic query language (ECLECTIC: Eligibility Criteria Language for Clinical Trial Investigation and Construction) was developed and implemented [5]. Key challenges for multi-site systems are extraction, transformation and loading (ETL) of data into data warehouses and mappings of local codes to a central terminology, as described by Hussain in a European context [6] and McMurry in a US context [7]. First versions of key data elements for protocol feasibility have been defined in Europe [8].

Patient Identification and Recruitment

Once a clinical study is initiated, eligible patients need to be identified. It is well-known that a large proportion of clinical trials are delayed or not successful due to issues with

* http://www.ercis.org

CRI Policy & Perspectives:

- **Data Sharing** (Longo & Drazen. NEJM. 2016)

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

CRI Policy & Perspectives:

- **Data Sharing** (Longo M. NEJM. 2016)
- Responses were swift…
  - Forbes to ISCB… perhaps my favorite:
  - Redpen/Blackpen: redpenblackpen.tumblr.com

Notable CRI-Related Events & Trends

Notable CRI-Related Events

- President's Precision Medicine Initiative
  - Several FOAs announced in November 2015
  - Reviews underway, more to come…
- Cancer Moonshot announcement in SOTU '16
  - VP Biden leading effort
- NIH budget increase requested…
  - President has proposed an overall budget increase of $825 million for the NIH in fiscal year 2017 compared with 2016. That money would be reserved for three efforts: the NCI's "moonshot" initiative, the Precision Medicine Initiative cohort program, and the BRAIN program.

Notable CRI-Related Events

- PCORnet phase 2 underway
  - 2 new CDRNs funded (LHSNet, OneFlorida)
  - 13 total now – pragmatic trials already underway

- NLM changes
  - Appointment of new Director coming soon

- Launch of new journals:
  - Learning Health Systems – Friedman Editor-in-Chief
  - Nature: Scientific Data http://www.nature.com/sdata/

Notable CRI-Related Events

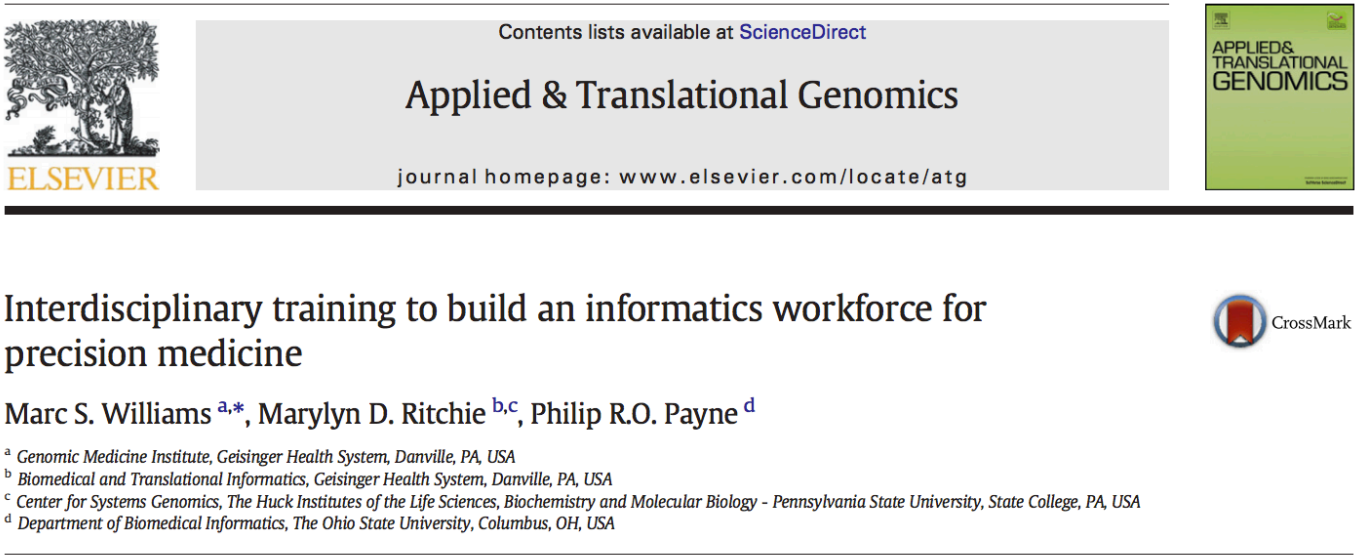
- **NPRM to Common Rule**
  - Many proposed changes that will impact CRI practice
- **Delays in MU-3, talk of MU changing…**
  - Potential for changes given MACRA, MIPS, etc.
  - Impact on EHRs/use, and therefore impacts to CRI
- **Emergence of new leadership roles**: In addition to CIO, CMIO, CRIO – increasing use of such titles as:
  - Chief Learning Officer
  - Chief Data Scientist/Architect/Officer
  - Vice President of Population Health
  - Business Intelligence Officer
  - Chief Analytics Officer
  - Chief Clinical Transformation Officer

Notable CRI-Related Trends…

- Recognition and push for reproducible research
  - NIH building this in future grant application and journals are facilitating open data sharing to promote replication studies
    - (research parasites notwithstanding ;)
- As some articles today reveal:
  - Renewed focus on the use of standards in practice, at least when it comes to data, and recognition of the importance of data quality (though still not quite there…)
- Trends for 2016:
  - Push for more networks & data sharing,
  - More concern about sustainability for research enterprise
  - Need to demonstrate value, even as demand for CRI expertise grows!

Notable CRI-Related Trends…

- Training continues to be key… funding for CRI through NLM and BD2K training grants
  - More and different training may be needed to succeed with ongoing initiatives
  - **CRI/TBI Workforce Development: Interdisciplinary training to build an informatics workforce for precision medicine** (Williams M, et al. Applied and Translational Genomics. 2016)



Contents lists available at ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Interdisciplinary training to build an informatics workforce for precision medicine

Marc S. Williams [a,*], Marylyn D. Ritchie [b,c], Philip R.O. Payne [d]

[a] Genomic Medicine Institute, Geisinger Health System, Danville, PA, USA
[b] Biomedical and Translational Informatics, Geisinger Health System, Danville, PA, USA
[c] Center for Systems Genomics, The Huck Institutes of the Life Sciences, Biochemistry and Molecular Biology - Pennsylvania State University, State College, PA, USA
[d] Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA

In Summary…

- Informatics approaches in CRI continue to accelerate
  - **Much** more activity than in years past
  - I'm sure it will only continue!
- CRI infrastructure also maturing and beginning to drive science
- Multiple groups/initiatives converging on common needs to advance the field
- CRI initiatives and investments beginning to realize the vision of the "*learning health system*"

- A very exciting time to be in CRI!

Thanks!

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

- Stephen Johnson
- Neil Sarkar
- Keith Marsolo
- Erin Holve
- Shawn Murphy
- Philip Payne
- Chunhua Weng

Thanks!

Peter.Embi@osumc.edu

Slides will be linked to on

http://www.embi.net/ (click on “Informatics”)