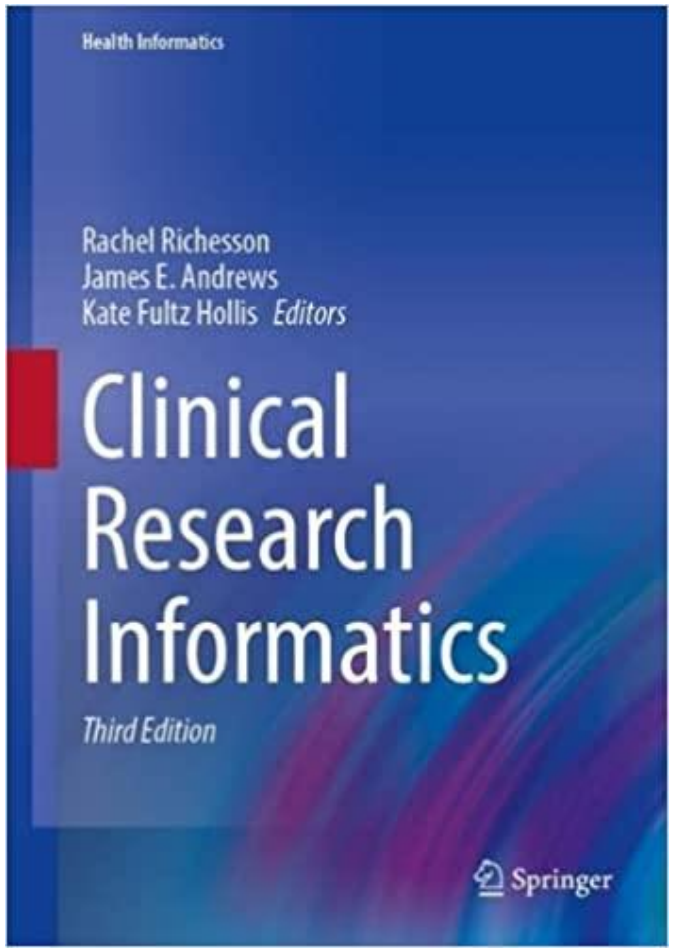
5.2 Clinical Research Informatics (1/2)

What is Biomedical and Health Informatics? - http://informatics.health/
William Hersh, MD, FACMI, FAMIA, FIAHSI

Clinical research informatics

- Definitions
- Opportunities
- Challenges
- Solutions

- Overview text: Richesson, 2023



Definitions

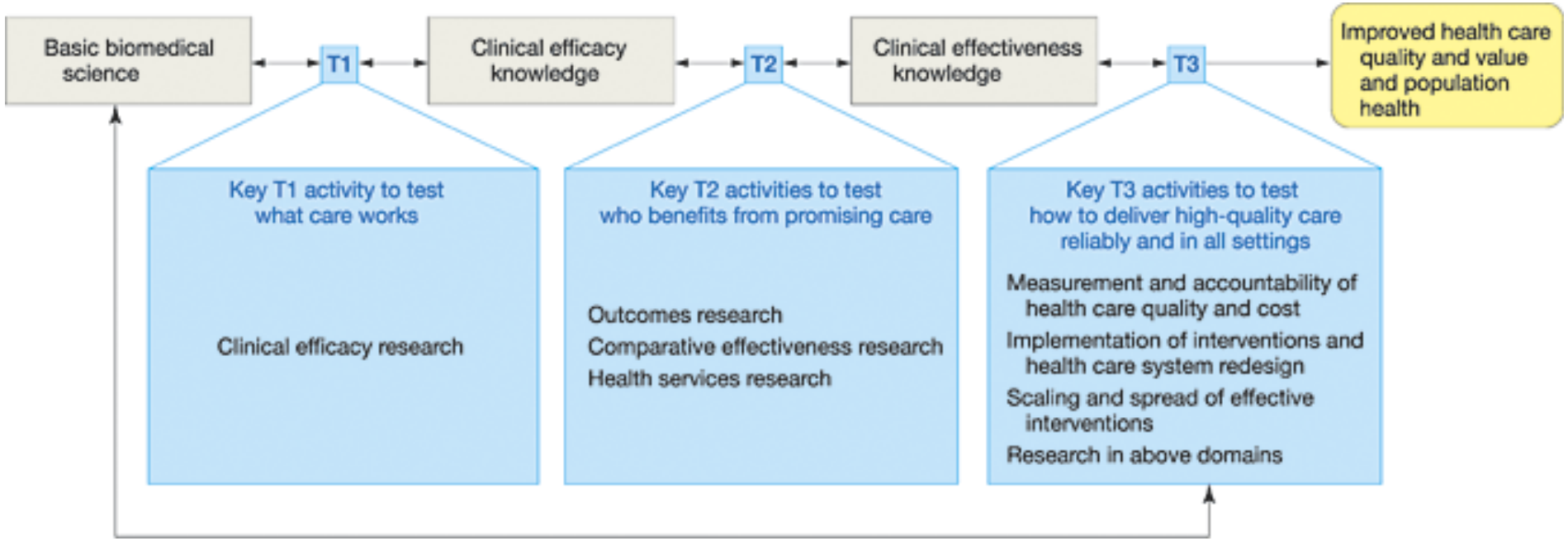
- Clinical research
- Translational research
- Clinical research informatics

Clinical research (from NIH)

- Clinical research comprises studies and trials in human subjects that fall into the three categories
  - Patient-oriented research – research conducted with human subjects (or on material of human origin such as tissues, specimens and cognitive phenomena) for which an investigator (or colleague) directly interacts with human subjects, including
    - Mechanisms of human disease
    - Therapeutic interventions
    - Clinical trials
    - Development of new technologies
  - Epidemiologic and behavioral studies
  - Outcomes research and health services research

Translational research (Zerhouni, 2007; Fort, 2018)

- Accelerating research results from laboratory to clinical environment to community along T1/T2/T3 axis (Dougherty, 2008; National Center for Advancing Translational Science [NCATS])
- Many aspects facilitated by informatics (Richesson, 2023)



OHSU

Clinical research informatics (CRI)

- CRI – “application of informatics principles and techniques to support the spectrum of activities and business processes that instantiate clinical research” (Richesson, 2023)
- Historical focus on management of research protocols and data capture and analysis, but increasingly focuses on integration of clinical systems and re-use of clinical data (Embi, 2013; Embi, 2019)
- Another area where there is distinction between informatics and related disciplines (e.g., computer science) (Bernstam, 2009)
- Growing numbers of research organizations employ a Chief Research Informatics Officer (CRIO) (Sanchez-Pinot, 2017)

Opportunities for CRI

- Growing quantity and breadth of data from electronic health record (EHR)
  - Early motivation for EHRs was “secondary use” or re-use of data (Safran, 2007; Meystre, 2017)
  - Opportunities for “evidence-generating medicine,” i.e., connecting clinical practice to research (Embi, 2013; Embi, 2019) or “real-world evidence” (Shaywitz, 2018)
  - Expand data beyond narrow scope and time frame of research data (Franklin, 2024)
  - Facilitate hypothesis generation as well as testing (Tang, 2024)

Opportunities (cont.)

- Beyond EHRs to health systems
  - Practice-based research networks (PBRNs) the address pertinent research questions and enabled by informatics (Westfall, 2007; Peterson, 2012; Peterson, 2012)
  - Learning health system – learning from data collected in care (IOM, 2012)
  - Serving as “honest brokers” for use of data for research (Keller, 2025)
- And more
  - Emulating experimental studies from observational data (Hernan, 2025)
  - Enhancing research in cancer and precision medicine (Gomez, 2024)
  - Leveraging generative AI (Foote, 2025)

Challenges for CRI

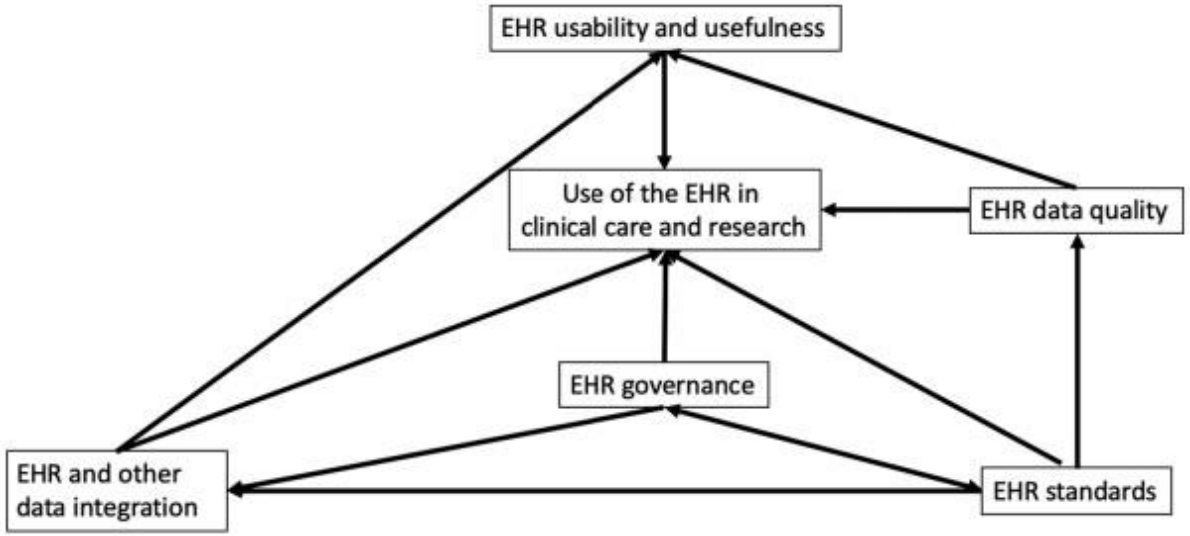
- Some issues already discussed
  - Data quality – Unit 5
  - Trust and privacy – Unit 7
  - Patient matching – Unit 8
- To discuss here
  - EHR tools for CRI
  - Data sharing
  - Trust

EHR tools for CRI

- Subjects in clinical trials do not always match population from which they are sampled
  - Demographics of those recruited through patient portals different from general population (Obeid, 2019)
  - Analysis of 1645 clinical trial participants in 202 trials found participants had fewer comorbidities and less use of medication than nonparticipants across variety of diseases (Rogers, 2021)
- Eligibility not always precise enough for computer representation (Ross, 2010)
- Data may not always be suitable for use in observational studies (Shang, 2018)
- Existing clinical workflows may not support learning health system (or research generally) (Goldhaber, 2024)

EHR and related tools can help (Obeid, 2017) or hinder (Holmes, 2021)

- Patient portals to notify patients of research opportunities
- Electronic alerts to care providers about patients who meet eligibility requirements
- Electronic alerts to research team about patients who meet eligibility requirements
- Access to data warehouse via staff member/analyst
- Self-service tools to run de-identified queries
- EHR-derived registries to aid in recruitment



Data sharing

- There is growing advocacy that scientific publications also publish their underlying data
  - Especially since most research funded by government sources (i.e., tax dollars)
- Relatively easy to do in era of ubiquitous Internet
  - Calls by journal editors, e.g., International Committee of Medical Journal Editors (Taichman, 2017), and others (Ross, 2013; Mello, 2018)
  - Adherence to policies incomplete (Naudet, 2018)
  - Give academic credit to those who produce data (Bierer, 2017)?
- Calls for
  - Data available based on FAIR principles of Findability, Accessibility, Interoperability and Reusability (Wilkinson, 2016)
    - And also traceable, licensed, and connected (TLC) (Haendel, 2016)
  - Transition toward larger open science ecosystem (NAP, 2018)
- NIH now requires Data Management and Sharing Plans (Kozlov, 2022)

Why make data available?

- Demonstration of value of repeating analyses
  - Re-analysis of 37 RCTs found 13 (35%) would have different interpretation of results, with majority identifying more who would benefit (Ebrahim, 2014)
  - Reproduction of analysis in studies of treatment of major depression in adolescence overturned original analyses (LeNoury, 2015)
- Calls for more transparency in research (Baggerly, 2018)
- Willingness of other research stakeholders, including researchers themselves, to allow sharing of data, though concern expressed about potential increased costs and loss of research validity (Mazor, 2017)

Concerns for data release

- Time-limited protection of those who generate data for rewards of their work and from those who aim to discredit or undermine original research (International Consortium of Investigators for Fairness in Trial Sharing Data, 2016)
- Cost to prepare and maintain, estimate to be about $30-50K per trial (Rockhold, 2016)
- Actual applications of data re-use so far have been small (Strom, 2016)
- Potential re-identification of de-identified data (Rocher, 2019)?
- Need to address informatics issues, such as adherence to standards (Kush, 2014) and proper attention to workflows, integration with other data, and research subject engagement (Tenenbaum, 2016)

Could synthetic data be the answer to privacy concerns?

- Generated from models of prevalence of patient findings, test results, diagnoses, etc. in population
  - Generation may employ machine learning methods (Choi, 2017; Tucker, 2020), including large language models (Yoon, 2023)
- Well-known example is Synthea
  - Many downloadable sets in different formats (Walonoski, 2018)
  - Including data set specific to COVID-19 (Walonoski, 2020)
- Shown in number of disease states to associate well with general population from which derived (Benaim, 2020; Foraker, 2020), including for COVID-19 (El Emam, 2021)
  - Although one analysis found effectively reproduced population in modeling demographics and services used, but less so for heterogeneous outcomes often used in quality measures (Chen, 2019)

OHSU

Trust and protection of participants

- Rights of patients/participants/subjects historically protected by institutional review boards (IRBs) (Grady, 2015)
- US Common Rule guides protection of research subjects and has been updated (Bierer, 2017; Hodge, 2017)
  - Making consent forms more readable
  - Allowing one IRB for multisite studies
  - Wider exemption for low-risk research
  - Allowing broad nonspecific future consent for biospecimens and re-use of data
- Still gaps between ethical principles and healthcare organizations' data-sharing policies and practices (Jackson, 2025)

Patients willing for data to be shared with appropriate safeguards

- Survey of US adults found patients amenable to secondary use of their EHR data; more toward research than quality or marketing uses (Grande, 2013)
- Breast cancer patients expressed low level of concerns about privacy of data but do want it protected (Rogith, 2014)
- 93% of patients are willing to have data shared with academics and 82% with industry researchers (Mello, 2018)
- From two academic medical centers, 87% willing to have their de-identified data shared and 56% willing to allow identified data to be shared for research, with somewhat lower proportions for those with mental health or abuse issues (Weng, 2019)
- Patients more willing to share data with local researchers and when well-designed consent forms used (Kim, 2019)
  - Higher age and health literacy also associated with willingness to share