Mobilizing Computable Biomedical Knowledge for Precision Medicine

Rachel Richesson, PhD, MPH, FACMI
September 27, 2018
Outline

• Learning Health Systems

• The “Keystone” Role of Computable Biomedical Knowledge (CBK)

• Vision of a Computable Knowledge Ecosystem and a Community to Advance It

• Mobilizing CBK
  • July 2018 Meeting
  • Next Steps
Better Health Through Learning Health Systems

Health systems--at any level of scale--become learning systems when they can, continuously and routinely, study and improve themselves.

The NEW ENGLAND JOURNAL of MEDICINE
Perspective: Jan 3, 2013
“Code Red and Blue — Safely Limiting Health Care’s GDP Footprint”
Arnold Milstein, M.D., M.P.H.

...U.S. health care needs to adopt new work methods, outlined in the Institute of Medicine’s vision for a learning health system...
LHS “Anthems”

• Learn from every health event!
• A system problem needs a system solution!
• 17 years to 17 months
  - to 17 weeks to 17 days (to 17 hours)!
Properties of a Health System That Can Learn & Improve

✓ A record of every health event is available to learn from

✓ Best practice knowledge is immediately available to support choices

✓ Improvement is continuous through ongoing study

✓ An infrastructure enables this to happen routinely and with economy of scale

✓ All of this is part of the culture
Learning Cycles
Better Health Requires a Flow

Interpret Results

Analyze

D2K: Data to Knowledge

Health Problem of Interest

Design Intervention

K2P: Knowledge to Performance

Assemble Data

Take Action

P2D: Performance to Data

Capture Practice as Data
Better Health Requires This

Health Problem of Interest

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data
Not This

Health Problem of Interest

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data

Journals
Persistent Computable Knowledge: The “Keystone” that Holds the Cycle Together

Health Problem of Interest

D2K: Data to Knowledge

K: Persistent Computable Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data
The Keystone Enables Discovery Systems to Become Learning Systems

Discovery System

Learning System
Persistent Knowledge

- **Knowledge**: The result of an analytical and/or deliberative process that holds significance for an identified community.

- **Persistence**: An explicit representation exists at any point in time.

- Persistent ≠ Static

- Persistent knowledge can be represented in two ways:
  - human readable
  - machine-executable
Two Complementary Ways to Represent Knowledge

Present: Human readable in words & pictures

Future: Computable (machine-executable) in code

Library Holdings: Books & Journals

Library Holdings: Will add Digital Knowledge Objects
Selection Criteria for Lung-Cancer Screening


ABSTRACT

BACKGROUND
The National Lung Screening Trial (NLST) used risk factors for lung cancer (e.g., ≥30 pack-years of smoking and <15 years since quitting) as selection criteria for lung-cancer screening. Use of an accurate model that incorporates additional risk factors to select persons for screening may identify more persons who have lung cancer or in whom lung cancer will develop.

METHODS
We modified the 2011 lung-cancer risk-prediction model from our Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial to ensure applicability to NLST data; risk was the probability of a diagnosis of lung cancer during the 6-year study period. We developed and validated the model (PLCO_{M2012}) with data from the 80,375 persons in the PLCO control and intervention groups who had ever smoked. Discrimination (area under the receiver-operating-characteristic curve [AUC]) and calibration were assessed. In the validation data set, 14,144 of 37,332 persons (37.9%) met NLST criteria. For comparison, 14,144 highest-risk persons were considered positive (eligible for screening) according to PLCO_{M2012} criteria. We compared the accuracy of PLCO_{M2012} criteria with NLST criteria to detect lung cancer. Cox models were used to evaluate whether the reduction in mortality among 53,202 persons undergoing low-dose computed tomographic screening in the NLST differed according to risk.
The New Knowledge is Expressed in a Model

Table 2. Modified Logistic-Regression Prediction Model (PLCO_{2012}) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
<th>Beta Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, per 1-yr increase†</td>
<td>1.081 (1.057–1.105)</td>
<td>&lt;0.001</td>
<td>0.0778868</td>
</tr>
<tr>
<td>Race or ethnic group‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.000</td>
<td></td>
<td>Reference group</td>
</tr>
<tr>
<td>Black</td>
<td>1.484 (1.083–2.033)</td>
<td>0.01</td>
<td>0.3944778</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.475 (0.195–1.160)</td>
<td>0.10</td>
<td>−0.7434744</td>
</tr>
<tr>
<td>Asian</td>
<td>0.627 (0.332–1.185)</td>
<td>0.15</td>
<td>−0.466585</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>2.793 (0.992–7.862)</td>
<td>0.05</td>
<td>1.027152</td>
</tr>
<tr>
<td>Education, per increase of 1 level†‡</td>
<td>0.922 (0.874–0.972)</td>
<td>0.003</td>
<td>−0.0812744</td>
</tr>
<tr>
<td>Body-mass index, per 1-unit increase†</td>
<td>0.973 (0.955–0.991)</td>
<td>0.003</td>
<td>−0.0274194</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease (yes vs. no)</td>
<td>1.427 (1.162–1.751)</td>
<td>0.001</td>
<td>0.3553063</td>
</tr>
<tr>
<td>Personal history of cancer (yes vs. no)</td>
<td>1.582 (1.172–2.128)</td>
<td>0.003</td>
<td>0.4589971</td>
</tr>
<tr>
<td>Family history of lung cancer (yes vs. no)</td>
<td>1.799 (1.471–2.200)</td>
<td>&lt;0.001</td>
<td>0.587185</td>
</tr>
<tr>
<td>Smoking status (current vs. former)</td>
<td>1.297 (1.047–1.605)</td>
<td>0.02</td>
<td>0.2597431</td>
</tr>
<tr>
<td>Smoking intensity¶</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of smoking, per 1-yr increase†</td>
<td>1.032 (1.014–1.051)</td>
<td>0.001</td>
<td>0.0317321</td>
</tr>
<tr>
<td>Smoking quit time, per 1-yr increase†</td>
<td>0.970 (0.950–0.990)</td>
<td>0.003</td>
<td>−0.0308572</td>
</tr>
<tr>
<td>Model constant</td>
<td></td>
<td></td>
<td>−4.532506</td>
</tr>
</tbody>
</table>

* To calculate the 6-year probability of lung cancer in an individual person with the use of categorical variables, multiply the variable or the level beta coefficient of the variable by 1 if the factor is present and by 0 if it is absent. For continuous variables other than smoking intensity, subtract the centering value from the person’s value and multiply the difference by the beta coefficient of the variable. For smoking intensity, calculate the contribution of the variable to the model by dividing by 10, exponentiating by the power −1, centering by subtracting 0.4021541613, and multiplying this number by the beta coefficient of the variable. Add together all the previously calculated beta-coefficient products and the model constant. This sum is called the model logit. To obtain the person’s 6-year lung-cancer probability, calculate $e^\text{logit}/(1+e^\text{logit})$. CI denotes confidence interval.
Envisioning An Extended Publication Pipeline

Human Readable: Article

Extraction

Library

Computable: Code

Encodable: Model

Programming

Expanded Library
This Idea is “On the Street”

The Atlantic, 2018

The Scientific Paper Is Obsolete


National Library of Medicine, 2017

Use Cases: CDS & Beyond

• **Clinical:** Present advice to inform decisions of providers & patients; evaluate treatments and measure performance for health organizations

• **Research:** Enhance scientific data & patient records, computable phenotypes, analytic “packages”

• **Public Health:** Event detection; deploy rapid response actions to threats; support behavior change and chronic disease management

• **Education:** Learning analytics, training for big data or ubiquitous knowledge environments
LHS “Platform” as a Set of Integrated Services

- **Policy & Technology for Making Knowledge Actionable & Sharable**
- **Technology for Sharing and Analyzing Data**
- **Policies and Mechanisms Governing Access to and Use of Data**
- **Methods and Processes for Supporting Learning Communities**
- **Methods and Processes for Promoting Behavior Change**
- **Technology for Generating & Delivering Tailored Messages to Decision Makers**
- **Technology for Capturing Practice Change**

**Health Problem of Interest**

**D2K: Data to Knowledge**

**K2P: Knowledge to Performance**

**P2D: Performance to Data**
Making Knowledge FAIR

• Findable
• Accessible
• Interoperable
• Reusable

FROM: https://www.force11.org/group/fairgroup/fairprinciples
Approach to Knowledge FAIRness: Machine-executable Knowledge Objects

Guidelines

Articles

Local Analytical Results

Knowledge Objects

Description

Interface

Computer-processable Knowledge ‘Payload’
Digital Libraries to Manage and Share Computable Knowledge

Curate and manage online collections of knowledge objects
Networks of Digital Libraries Enable Computable Knowledge Ecosystem
How can CBK transform health?

Mobilize

Computable
Biomedical
Knowledge

https://youtu.be/DZ2WTs-J-zE
We welcome to the meeting anyone interested in building a community to accelerate the application of biomedical knowledge into practice, and more specifically to achieve this by representing knowledge in computable forms so it can be better curated, managed, and disseminated. The meeting will explore issues of policy and governance as well as issues of technology and standards. The topics to be discussed are applicable to the basic, clinical and translational sciences.

We have designed the meeting to be highly inclusive. We welcome participation from the public and private sectors as well as individuals with backgrounds in informatics, information and knowledge management, population health, and the full range of biomedical sciences. We hope the meeting will include individuals who bring a patient perspective on the issues to be examined as well as those whose interests emphasize health care quality and value.
MCBK - Inaugural Public Meeting

July 10-11, 2018
Lister Hill National Center for Biomedical Communications
National Library of Medicine

Meeting co-chairs:
Charles Friedman
University of Michigan
Rachel Richesson
Duke University

We Must Develop a Communications System so that the Miraculous Triumphs of Modern Science Can Be Taken from the Laboratory and Transmitted to All in Need.

SENATOR LISTER HILL, 1965
Planning Committee

Julia Adler-Milstein, PhD, Associate Professor UCSF School of Medicine

Jane Blumenthal, MSLS, AHIP, FMLA, Associate University Librarian and Director, Taubman Health Sciences Library, University of Michigan

Greg Cooper, MD, PhD, Professor, Department of Biomedical Informatics and of Intelligent Systems, Vice Chairman, Department of Biomedical Informatics, University of Pittsburgh School of Medicine

Milt Corn, MD, FACP, FACMI, Deputy Director for Research and Education, National Library of Medicine

Chris Dymek, Ed.D., Director, Health IT Division, Agency for Healthcare Research and Quality

Peter Embi, MD, MS, Chief Executive Officer, Regenstrief Institute

Bob Greenes, MD, PhD, Professor, Department of Biomedical Informatics, Arizona State University

Nancy Lorenzi, PhD, MA, MS, Vice President for Strategic Change Management, Professor of Biomedical Informatics, Clinical Professor in Nursing, Vanderbilt University Medical Center

Dan Masys, MD, FACMI, Affiliate Professor, Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine

Blackford Middleton, MD, MPH, MSc, FACMI, Chief Informatics and Innovation Officer, Apervita, Inc.

Mark Musen, MD, PhD, Professor of Medicine (Biomedical Informatics Research) and Biomedical Data Science, Stanford Health Policy

Jerry Perry, MLS, Associate Dean and Director AHSL, The University of Arizona Health Sciences Library

Chris Shaffer, University Librarian and Assistant Vice Chancellor, Parnassus Library, University of California San Francisco

Umberto Tachinardi, MD, Director, Biomedical Informatics, University of Wisconsin- Madison

John Wilbanks, Chief Commons Officer, Sage Bionetworks
Meeting Goals

• **Strengthen** our foundation of shared recognition of values and principles for mobilizing CBK

• **Frame** and address important dimensions for mobilizing CBK

• **Grow** the CBK community

• **Develop** Action Plans

• **Identify** priorities

• **Plan** next steps!!
A Manifesto

• Offers clarity of purpose
• Sets direction
• Engages participants
• Helps guide decisions
• Specifies the technical drivers

A famous manifesto

Thanks to Nancy Lorenzi for providing summary and leading reactor panel.
Why—CBK is Needed

• Knowledge has the potential to improve health of individuals and populations

• Every decision affecting health should be informed by the best available knowledge

• Computable forms of knowledge can generate and deliver needed health advice where and when it is needed
Why—The CBK Manifesto

• VISION is to create an ECOSYSTEM that will work toward ensuring that:

• biomedical knowledge that reflects the best and most current evidence and science can be mobilized to support health.

• knowledge can be shared by multiple people and organizations and integrated into health information systems and applications for all.
How - CBK ecosystem

• Sustained through public-private partnerships

• CBK ecosystem will:
  • have a participatory governance and be collaborative
  • be identified as a source of trust
  • be diverse, inclusive, and provide equity and accessibility
  • have high standards of privacy and security
  • support open standards
  • have value for knowledge creators, users and the public
  • be agile for changes in knowledge
  • support transparency, validity and provenance of knowledge
  • have a pipeline that transitions knowledge from human-readable to fully computable
## Levels of Knowledge Format - CDS

<table>
<thead>
<tr>
<th></th>
<th>Narrative</th>
<th>Semi-structured</th>
<th>Structured</th>
<th>Executable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Format</strong></td>
<td>Narrative text</td>
<td>Organized text</td>
<td>Coded; interpretable by computer</td>
<td>Coded; interpretable by CDS systems; variety of formats</td>
</tr>
<tr>
<td><strong>Sharability</strong></td>
<td>Broad</td>
<td>Broad</td>
<td>Broad</td>
<td>Very limited</td>
</tr>
<tr>
<td><strong>Tool independent</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Site independent</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Author</strong></td>
<td>Guideline developer</td>
<td>Clinical domain expert</td>
<td>Knowledge engineer</td>
<td>CDS Implementer</td>
</tr>
</tbody>
</table>

*Boxwala et al., JAMIA 2011, [https://doi.org/10.1136/amiajnl-2011-000334]*
Breakout Groups

• Standards
  Mark Musen
  Bob Greenes
  Rachel Richesson

• Technical Infrastructure
  Christopher Shaffer
  Gregory Cooper

• Sustainability for Mobilization & Inclusion
  Chris Dymek
  Nancy Lorenzi
  Jerry Perry

• Policy & Coordination to Ensure Quality & Trust
  Blackford Middleton
  Josh Richardson
  Jodyn Platt
Standards

- Metadata standards to include important dimensions of CBK
  - Use – input/output, interoperability
  - Relationships – dependencies, correspondences
  - Discoverability/Findability – identifier, title, domain
  - Trust – provenance, evidence levels, source(s)
  - Object Mgt – versions, cryptographic keys
  - Access – ownership, policies, licenses, terms of use

- Exemplars for artifacts and metadata for range of use cases & knowledge-sharing scenarios

- Other ideas: values, principles, process...
Technical Infrastructure

• Learn from other industries
• Collaborate w/ and harmonize knowledge users
• Describe use cases and activity models in order to:
  • Identify infrastructure requirements
  • Identify “reference implementations”
  • Minimize # standards required
  • Promote standards as services
• Define meta-information about knowledge and its use
• Establish governance infrastructure to help manage CBK ownership, intellectual property and access
Sustainability for Mobilization & Inclusion

• Energize and promote MCBK
• Articulate value proposition for CBK
• Define and execute plans to engage stakeholders:
  • Professional societies (AMA, ANA)
  • Patient advocacy groups
  • Clinical guideline organizations
  • Libraries
  • Industry vendors
Policy & Coordination to Ensure Quality & Trust

• Key question: What is the process for ensuring high quality and trustworthy knowledge in a CBK ecosystem?

• Define and convey transparency in the CBK ecosystem
  • Talk to other groups and obtain information about current practices
  • Assemble, compare, and contrast metadata schema to identify gaps
  • Evaluate FAIR systems for transparency
  • Extend to FAIR-TLC (Traceability, Licensure, Connectedness)
  • Map the regulatory space
Transparency – Starting Ideals

• Make public all info & metadata about object to inform users to apply to specific use case
• Include limitations; include derivatives
• Values are explicitly stated (e.g., tradeoffs, COI)

• Provide metadata for artifacts, development process, system, people, experience (implementation), sources, COI, legal parameters
• Capacity to “look under the hood”
• User design
How we see this progressing...

• **Grow** the CBK community
  • Spread the Word – publish our Manifesto
  • Identify and Include - those we have missed
  • Sustain and Coordinate - our efforts

• **Develop** the community and our work
  • Harmonize and Disseminate Action Plans
  • Implement Challenge Problems

• **Plan** a range of activities to advance our work and realize our vision of mobilized CBK
Actions Going Forward

• Website – MobilizeCBK.org

• Webinar – October 25, 2018

• Parallel Action of the 4 theme-based groups

• Interim “governance”
  • Univ. of Michigan as interim home & resource hub
  • Meeting Planning Committee to act as Steering Committee (interim governance)

• Road Map – under development

• Future Meeting – July 2019
Onward!

Policy & Technology for Making Knowledge Actionable & Shareable

Technology for Generating & Delivering Tailored Messages to Decision Makers

Methods and Processes for Promoting Behavior Change

Technology for Capturing Practice Change

Methods and Processes for Supporting Learning Communities

Formation of Learning Community

P2D: Performance to Data

K2P: Knowledge to Performance

D2K: Data to Knowledge

Policies and Mechanisms Governing Access to and Use of Data

Technology for Sharing and Analyzing Data

Health Problem of Interest

Mobilize CBK
Questions?

Rachel.Richesson@dm.duke.edu

If you're interested in learning more or joining the discussion, please email us at MCBK-Info@umich.edu

MobilizeCBK.org