Mobilizing Computable Biomedical Knowledge

Charles P. Friedman, PhD
Josiah Macy, Jr. Professor
Chair, Department of Learning Health Sciences
Professor of Information and Public Health
University of Michigan
October 18, 2017
Main Menu

• Preliminaries

• The LHS and Infrastructural Services to Support It

• The “Keystone” Role of Persistent Computable Knowledge
  • And the Concept of a “K2P” Service

• Doing this at Scale: Vision of a Computable Knowledge Ecosystem...

• Goals and Plans for This Meeting
A Definition for Purposes of This Meeting

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

• Preliminaries

• The LHS and Infrastructural Services to Support It

• The “Keystone” Role of Persistent Computable Knowledge
  • And the Concept of a “K2P” Service

• Doing this at Scale: Vision of a Computable Knowledge Ecosystem...

• Goals and Plans for This Meeting
Properties of a Health System That Can Learn & Improve

✓ Every participating patient’s characteristics and experience are available to learn from

✓ Best practice knowledge is immediately available to support decisions

✓ 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:
D2K -> K2P -> P2D

Interpret Results

Health Problem of Interest

D2K: Data to Knowledge
K2P: Knowledge to Performance
P2D: Performance to Data

Assemble Data
Analyze Data
Communicate Tailored Change Strategies
Take Action
Record Actions

Formation of Learning Community
Better Health Requires This

A Health Problem of Interest

D2K

K2P

P2D
Not This

D2K

A Health Problem of Interest

Result

Journals
LHS Infrastructure: A *Platform* Supporting Multiple Simultaneous Learning Cycles
LHS Platform as a Set of Integrated Services

Technology for Sharing and Analyzing Data

Technology for Generating & Delivering Tailored Messages to Decision Makers

Policy & Technology for Making Knowledge Actionable & Sharable

Methods and Processes for Promoting Behavior Change

Policies and Mechanisms Governing Access to and Use of Data

Methods and Processes for Supporting Learning Communities

Formation of Learning Community

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data

Health Problem of Interest
In Relative Terms, What Exists Today

Policy & Technology for Making Knowledge Actionable & Sharable

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

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data

In Relative Terms, What Exists Today

Policies and Mechanisms Governing Access to and Use of Data

Technology for Sharing and Analyzing Data
And Our Focus is Here, at K2P…

Policy & Technology for Making Knowledge Actionable & Sharable

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

D2K: Data to Knowledge

P2D: Performance to Data

K2P: Knowledge to Performance

Policies and Mechanisms Governing Access to and Use of Data

Technology for Sharing and Analyzing Data

Health Problem of Interest
Main Menu

• Preliminaries

• The LHS and Infrastructural Services to Support It

• The “Keystone” Role of Persistent Computable Knowledge
  • And the Concept of a “K2P” Service

• Doing this at Scale: Vision of a Computable Knowledge Ecosystem...

• Goals and Plans for This Meeting
Persistent Knowledge

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

Martin C. Tammemägi, Ph.D., Hormuzd A. Katki, Ph.D., William G. Hocking, M.D.,
Timothy R. Church, Ph.D., Neil Caporaso, M.D., Paul A. Kvale, M.D.,
Anil K. Chaturvedi, Ph.D., Gerard A. Silvestri, M.D., Tom L. Riley, B.Sc.,
John Commins, B.Sc., and Christine D. Berg, M.D.

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\textsubscript{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\textsubscript{M2012} criteria. We compared the accuracy of PLCO\textsubscript{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<sub>mc2012</sub>) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.<sup>*</sup>

<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>−1.822606</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

Encodable: Model

Computable: Code

Library

Expanded Library
Computable, Persistent Knowledge is the LHS “Keystone”
The Keystone Enables Discovery Systems to Become Learning Systems.
Folding in the Concept of Infrastructure...
The LHS Requires a “K2P” Service
Minimum Requirements for a K2P Service

- Representation of knowledge in machine-executable forms
- Capability for rapid knowledge revision as the system learns
- Modular linking of related knowledge
- Sharing of knowledge across an ecosystem
- Scalable computation and delivery of tailored messages to inform practice
K2P Use Cases: CDS & Beyond

• **Clinical**: Bringing advice, generated from computable knowledge, to inform decisions of providers, consumers, and managers
• **Research**: Enhancing the scientific record, computable phenotypes, analytic “packages”
• **Public Health**: Event detection objects; rapid response deployment
• **Education**: Learning analytic objects, preparation for practice in an environment of ubiquitous knowledge
Main Menu

• Preliminaries

• The LHS and Infrastructural Services to Support It

• The “Keystone” Role of Persistent Computable Knowledge
  • And the Concept of a “K2P” Service

• Doing this at Scale: Vision of a Computable Knowledge Ecosystem...

• Goals and Plans for This Meeting
Much Has Been Said About Data FAIRness
This Meeting is about Knowledge FAIRness

Making Knowledge:

• 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’
And Digital Libraries to Manage and Share Computable Knowledge

Capability to curate and manage online collections of knowledge objects
And Networks of Digital Libraries to Enable a Computable Knowledge Ecosystem
Main Menu

• Preliminaries

• The LHS and Infrastructural Services to Support It

• The “Keystone” Role of Persistent Computable Knowledge
  • And the Concept of a “K2P” Service

• Doing this at Scale: Vision of a Computable Knowledge Ecosystem...

• Goals and Plans for This Meeting
Thanking the Planning Committee

- Julia Adler-Milstein
- Jane Blumenthal
- Milton Corn
- Chris Dymek
- Peter Embi
- Bob Greenes
- Ken Mandl
- Dan Masys
- Blackford Middleton
- Mark Musen
Meeting Goals

• To begin exploring a set of issues – not primarily about technology – that need to be understood to advance a community interested in computable biomedical knowledge.

• To begin building the core group that anticipates and works in support of establishing a larger computable biomedical knowledge community.
Four Organizing Themes

A briefing paper on each has been distributed:
1. Knowledge infrastructure requirements
2. Establishing a trusted system
3. Metadata
4. IP and Copyright
The Plan: Following This Talk

Today

- Panel: Three ongoing efforts
- Briefing sessions keyed to the four themes (and briefing papers)
- Small group discussions keyed to the four themes and focused on specific questions
- Report out
- Dinner at the Gandy Dancer

Tomorrow

- Small groups resume around four themes
- Cross-fertilization groups
- Synthesizing discussion
- Moving forward from here
What Success Might Look Like

- One or more articles
- An open meeting in mid-2018
- A nascent organization or association with an existing organization
- Computable meta-knowledge (putting what we learned in computable form)
- Funding
- Opening a public library of computable knowledge
Thanks
cpfried@umich.edu
lhs.medicine.umich.edu