Mobilizing Computable Biomedical Knowledge

Charles P. Friedman, PhD
University of Michigan

Rachel Richesson, PhD
Duke University

July 10, 2018

#MobilizeCBK
Main Menu

• Improving Health and Learning Health Systems

• The “Keystone” Role of Persistent Computable Knowledge

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

• Goals and Plans for This Meeting
Better Health...

This is why we’re all here today
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

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:
P2D -> D2K -> K2P -> P2D

Interpret Results

Analyze
Assemble Data

Design Intervention
Take Action

Health Problem of Interest

Capture Practice as Data

P2D: Performance to Data
D2K: Data to Knowledge
K2P: Knowledge to Performance
Better Health Requires This

Health Problem of Interest

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data
Not This

D2K: Data to Knowledge

Health Problem of Interest

K2P: Knowledge to Performance

P2D: Performance to Data

Journals
LHS “Platform” as a Set of Integrated Services

Policies and Mechanisms Governing Access to and Use of Data

Technology for Sharing and Analyzing Data

Policy & Technology for Making Knowledge Actionable & Sharable

Technology for Generating & Delivering Tailored Messages to Decision Makers

Methods and Processes for Promoting Behavior Change

Methods and Processes for Supporting Learning Communities

Technology for Capturing Practice Change

Health Problem of Interest

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data
Main Menu

• Improving Health and Learning Health Systems

• The “Keystone” Role of Persistent Computable Knowledge

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

• Goals and Plans for This Meeting
Persistent Computable Knowledge: The “Keystone” that Holds the Cycle Together

Health Problem of Interest

D2K: Data to Knowledge

K2P: Knowledge to Performance

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

![Diagram showing the transition from a Discovery System to a Learning System. The diagram illustrates the flow from Data to Knowledge (D2K), Knowledge to Performance (K2P), and Performance to Data (P2D).]
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 (PLCOw2012) 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>Reference group</td>
<td></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>0</td>
<td></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
This Idea is “On the Street”

The Atlantic, 2018


National Library of Medicine, 2017

LHS “Platform” as a Set of Integrated Services

Policies and Mechanisms Governing Access to and Use of Data

Technology for Sharing and Analyzing Data

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

D2K: Data to Knowledge

K2P: Knowledge to Performance

P2D: Performance to Data

Health Problem of Interest
Platform Services Specific to Mobilizing Computable Knowledge

Health Problem of Interest

- Technology for Sharing and Analyzing Data
- 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
- Policies and Mechanisms Governing Access to and Use of Data
- Methods and Processes for Supporting Learning Communities

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