MCBK
Mobilizing Computable Biomedical Knowledge

#MobilizeCBK

Mobilizing Computable Biomedical Knowledge

Welcome and Opening

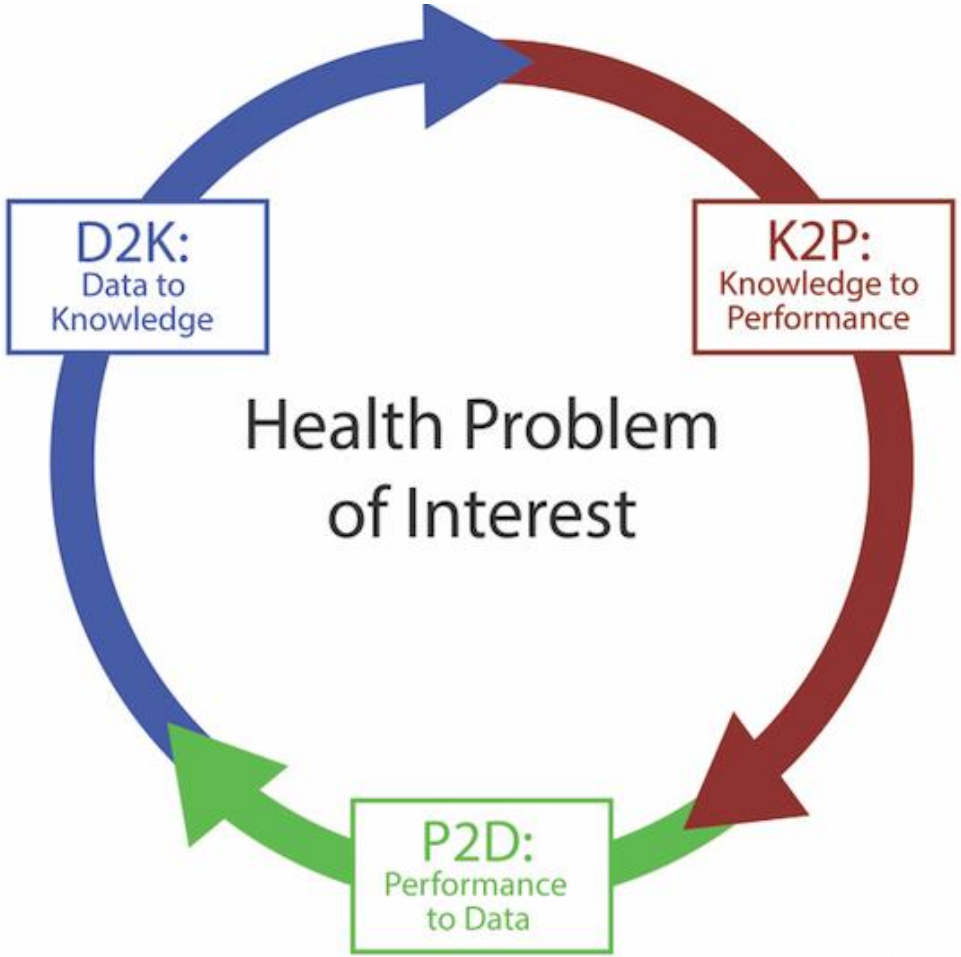
Charles P. Friedman, PhD
Rachel Richesson, PhD
MCBK Co-Chairs

2nd Annual Public Meeting
July 18, 2019
NIH Natcher Conference Center

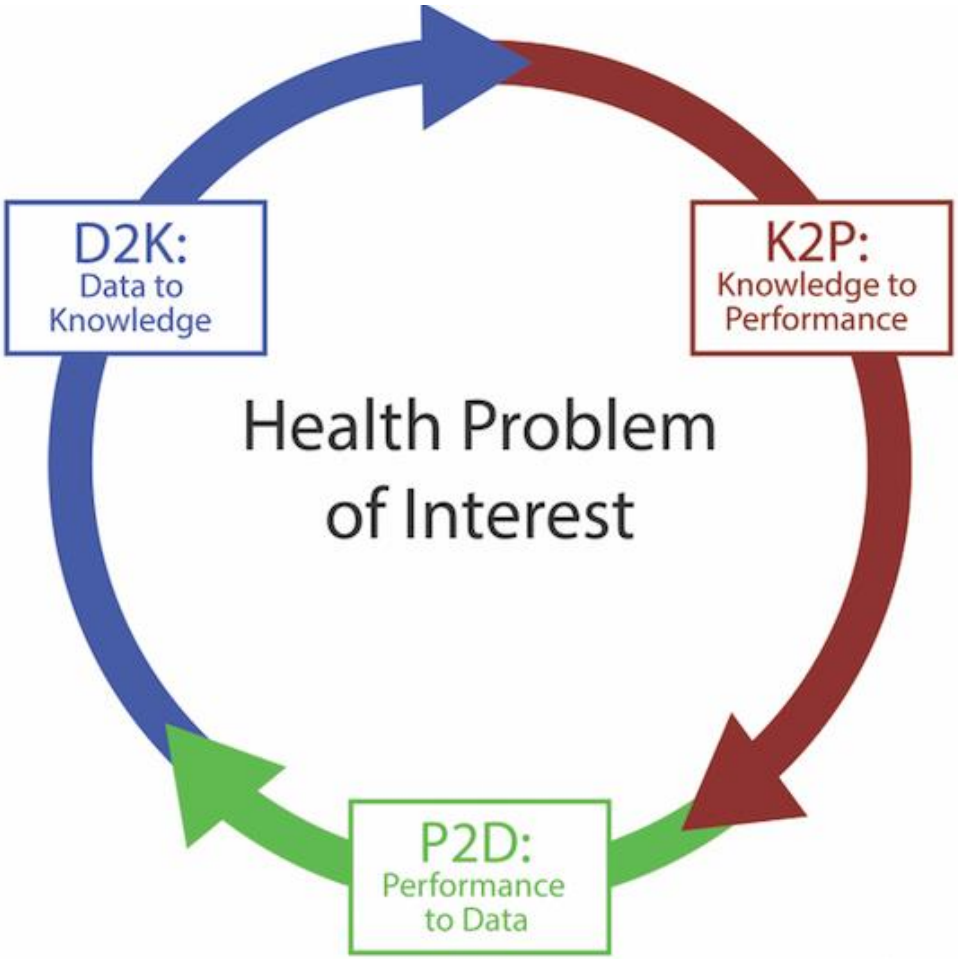
Outline

- Foundations for MCBK
- Why We're All Here
- MCBK Community & Workgroups
- Goals and Plans for This Meeting

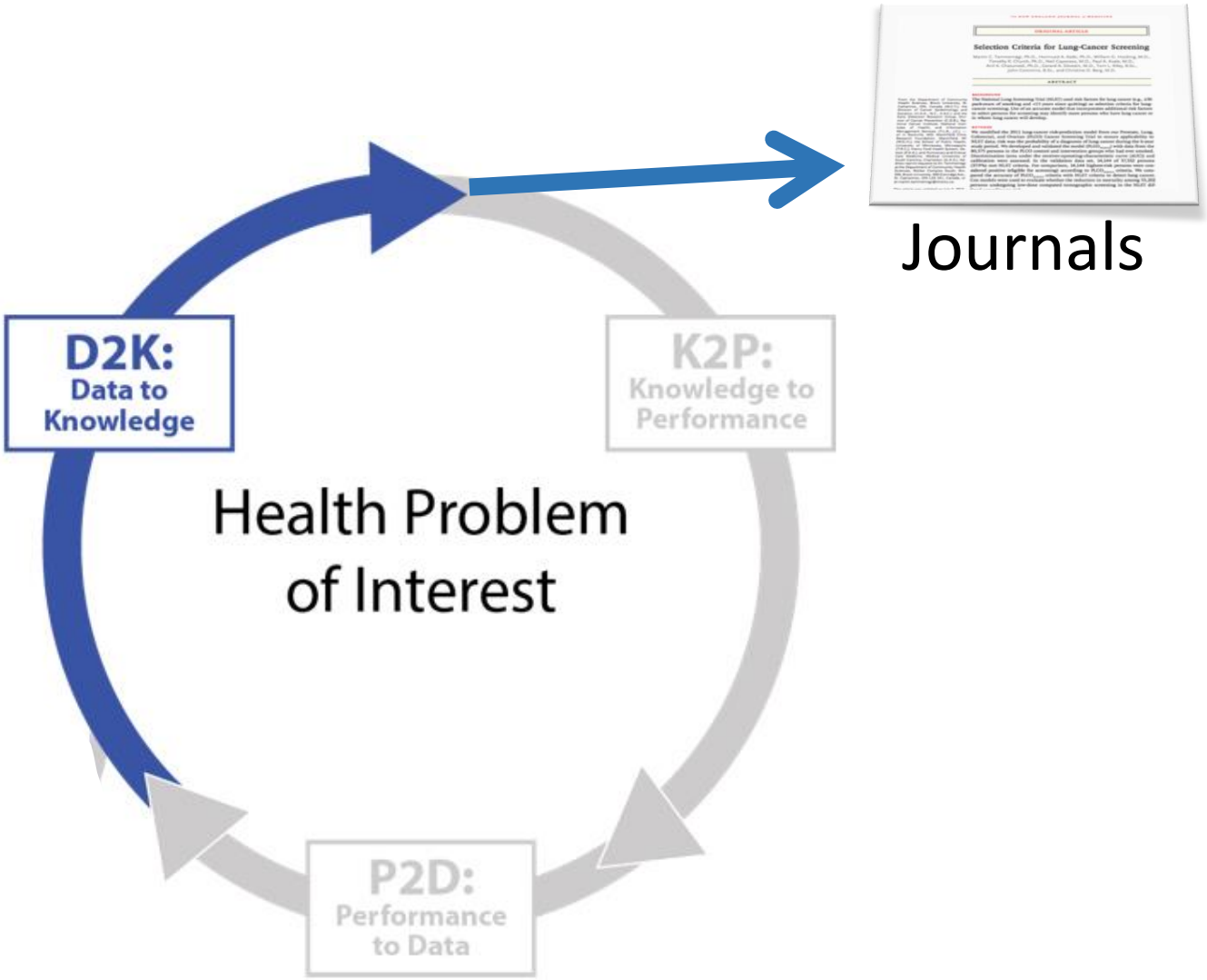
MCBK is Anchored in the "Learning Health System" Concept of Continuous Cycles of Improvement



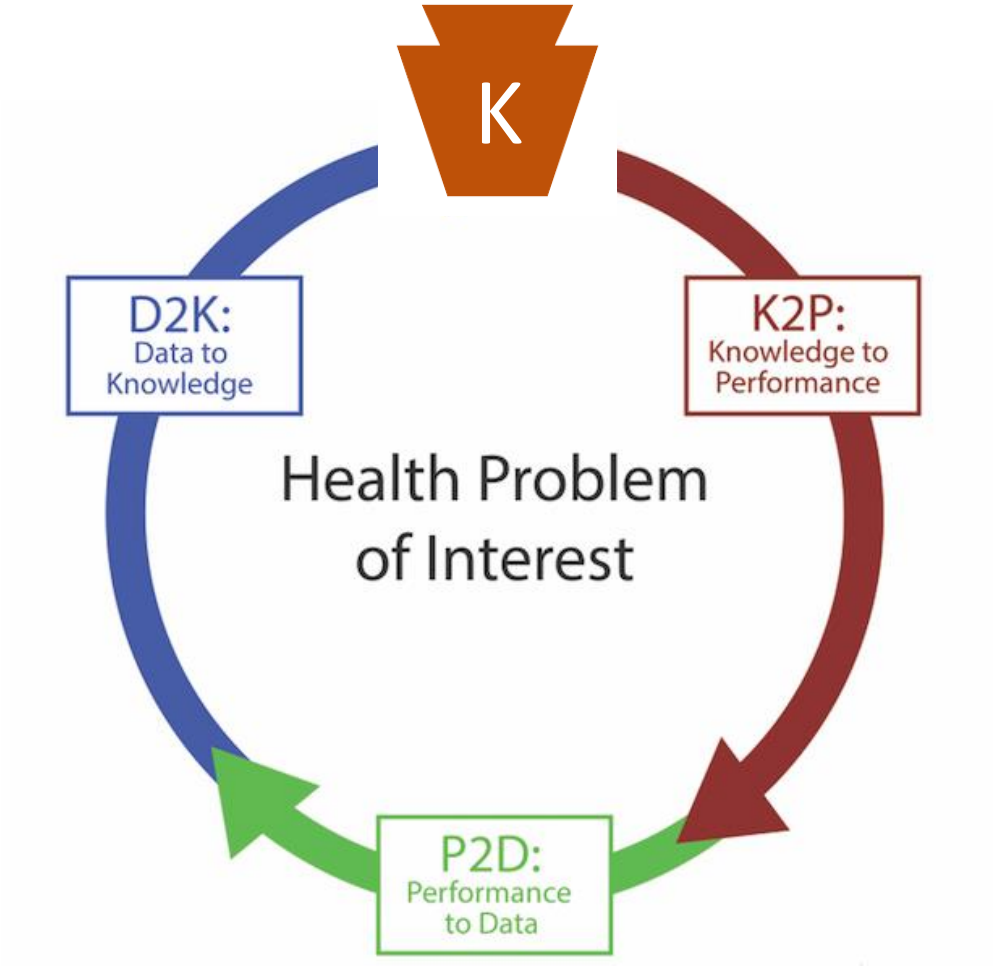
Better Health Requires This



Not Just This



Knowledge is the "Keystone" that Holds the Cycle Together



A Way to Think About Knowledge

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

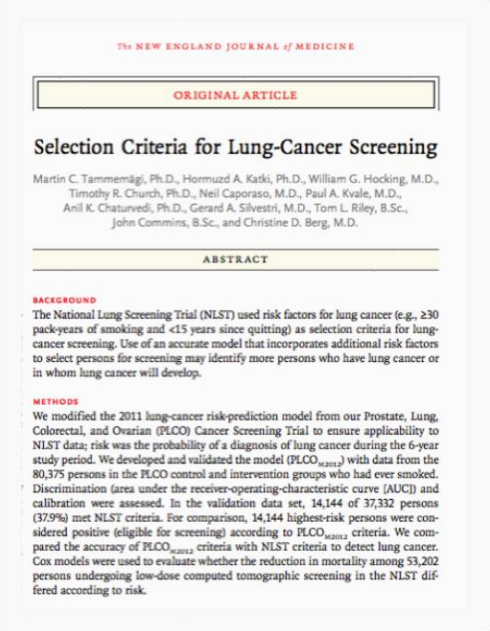
Examples of Biomedical Knowledge

From primarily analytical to primarily deliberative:

- Predictive/explanatory models
- Computable phenotypes
- Causal/propositional networks
- Best practices (guidelines)
- Decision Trees
- Policies

Two Complementary Ways to Represent Knowledge

Human readable in words, pictures, equations



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

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

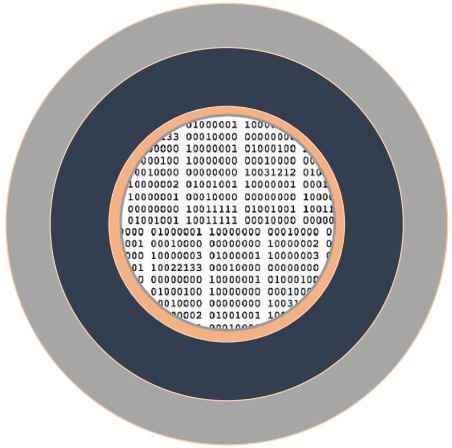
The National Lung Screening Trial (NLST) used risk factors for lung cancer (e.g., ≥30 pack-years of smoking and <15 years since quitting) as selection criteria for lung-cancer screening. Use of an accurate model that incorporates additional risk factors to select persons for screening may identify more 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 ($PLCO_{M2012}$) 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.

Library Holdings:
Books & Journals

Computable (machine-executable) in code



Library Holdings: Will add
Digital Knowledge Objects

Example: Human Readable Knowledge

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

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.

From the Department of Community Health Sciences, Brock University, St. Catharines, ON, Canada (M.C.T.); the Division of Cancer Epidemiology and Genetics (H.A.K., N.C., A.K.C.) and the Early Detection Research Group, Division of Cancer Prevention (C.D.B.), National Cancer Institute, National Institutes of Health, and Information Management Services (T.L.R., J.C.) — all in Rockville, MD; Marshfield Clinic Research Foundation, Marshfield, WI (W.G.H.); the School of Public Health, University of Minnesota, Minneapolis (T.R.C.); Henry Ford Health System, Detroit (P.A.K.); and Pulmonary and Critical Care Medicine, Medical University of South Carolina, Charleston (G.A.S.). Address reprint requests to Dr. Tammemägi at the Department of Community Health Sciences, Walker Complex South, Rm. 306, Brock University, 500 Glenridge Ave., St. Catharines, ON L2S 3A1, Canada, or at martin.tammemagi@brocku.ca.

This article was updated on July 3, 2013, at NEJM.org.

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.

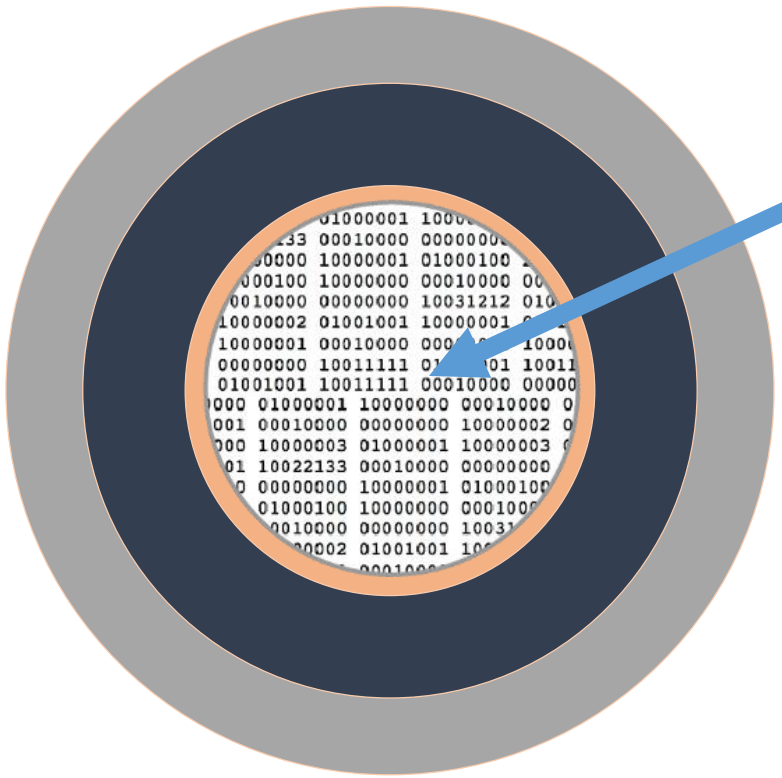
The New Knowledge is Expressed in a Model

Table 2. Modified Logistic-Regression Prediction Model ($PLCO_{M2012}$) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.*

Variable	Odds Ratio (95% CI)	P Value	Beta Coefficient
Age, per 1–yr increase†	1.081 (1.057–1.105)	<0.001	0.0778868
Race or ethnic group‡			
White	1.000		Reference group
Black	1.484 (1.083–2.033)	0.01	0.3944778
Hispanic	0.475 (0.195–1.160)	0.10	–0.7434744
Asian	0.627 (0.332–1.185)	0.15	–0.466585
American Indian or Alaskan Native	1		0
Native Hawaiian or Pacific Islander	2.793 (0.992–7.862)	0.05	1.027152
Education, per increase of 1 level†§	0.922 (0.874–0.972)	0.003	–0.0812744
Body-mass index, per 1-unit increase†	0.973 (0.955–0.991)	0.003	–0.0274194
Chronic obstructive pulmonary disease (yes vs. no)	1.427 (1.162–1.751)	0.001	0.3553063
Personal history of cancer (yes vs. no)	1.582 (1.172–2.128)	0.003	0.4589971
Family history of lung cancer (yes vs. no)	1.799 (1.471–2.200)	<0.001	0.587185
Smoking status (current vs. former)	1.297 (1.047–1.605)	0.02	0.2597431
Smoking intensity¶			–1.822606
Duration of smoking, per 1-yr increase†	1.032 (1.014–1.051)	0.001	0.0317321
Smoking quit time, per 1-yr increase†	0.970 (0.950–0.990)	0.003	–0.0308572
Model constant			–4.532506

* 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 $e^{logit}/(1+e^{logit})$. CI denotes confidence interval.

And the Knowledge Can be Made Computable by Representing It as Coded "Knowledge Objects"

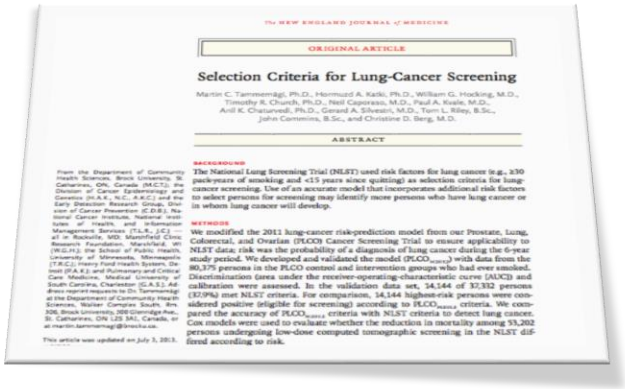


Example: A computer program that takes in characteristics of a person and computes a risk score for that person

Envisioning An Extended Publication Pipeline

Extraction

Programming



THE NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

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 ($PLCO_{m2012}$) 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, 34,144 of 37,332 persons (32.9%) 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.

This article was updated on July 3, 2013.

Human Readable: Article

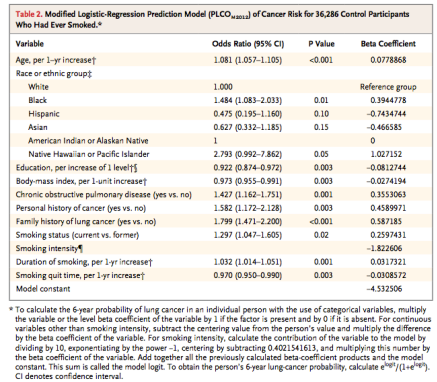
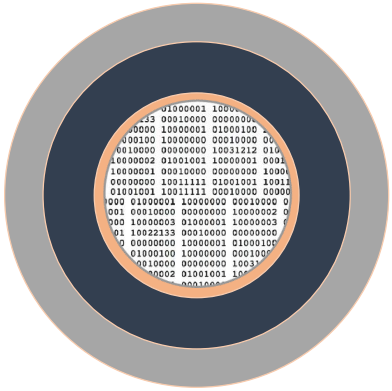


Table 2. Modified Logistic-Regression Prediction Model ($PLCO_{m2012}$) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.*

Variable	Odds Ratio (95% CI)	P Value	Beta Coefficient
Age, per 1-yr increase†	1.081 (1.057–1.105)	<0.001	0.0778868
Race or ethnic group‡			
White	1.000		Reference group
Black	1.484 (1.083–2.033)	0.01	0.3944778
Hispanic	0.475 (0.195–1.160)	0.10	−0.743474
Asian	0.627 (0.332–1.185)	0.15	−0.466585
American Indian or Alaskan Native	1		0
Native Hawaiian or Pacific Islander	2.793 (0.992–7.862)	0.05	1.027152
Education, per increase of 1 level†§	0.922 (0.874–0.972)	0.003	−0.0812744
Body-mass index, per 1-unit increase†	0.973 (0.955–0.991)	0.003	−0.0274194
Chronic obstructive pulmonary disease (yes vs. no)	1.427 (1.162–1.751)	0.001	0.3553063
Personal history of cancer (yes vs. no)	1.582 (1.172–2.128)	0.003	0.4589971
Family history of lung cancer (yes vs. no)	1.799 (1.471–2.200)	<0.001	0.587185
Smoking status (current vs. former)	1.297 (1.047–1.605)	0.02	0.2597431
Smoking intensity¶			−1.822606
Duration of smoking, per 1-yr increase†	1.032 (1.014–1.051)	0.001	0.0317321
Smoking quit time, per 1-yr increase†	0.970 (0.950–0.990)	0.003	−0.0308572
Model constant			−4.532506

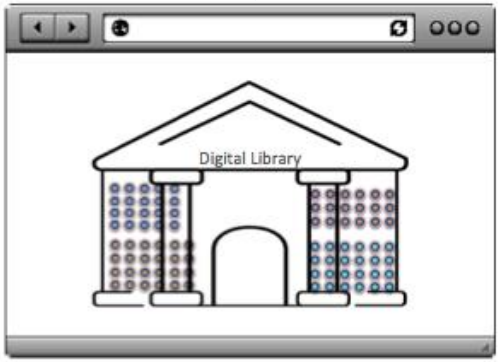
* To calculate the 6-year probability of lung cancer in an individual person with the use of categorical variables, multiply the variable or the level beta coefficient of the variable by 1 if the factor is present and by 0 if it is absent. For continuous variables other than smoking intensity, subtract 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 $e^{logit}/(1+e^{logit})$. CI denotes confidence interval.

Encodable: Model



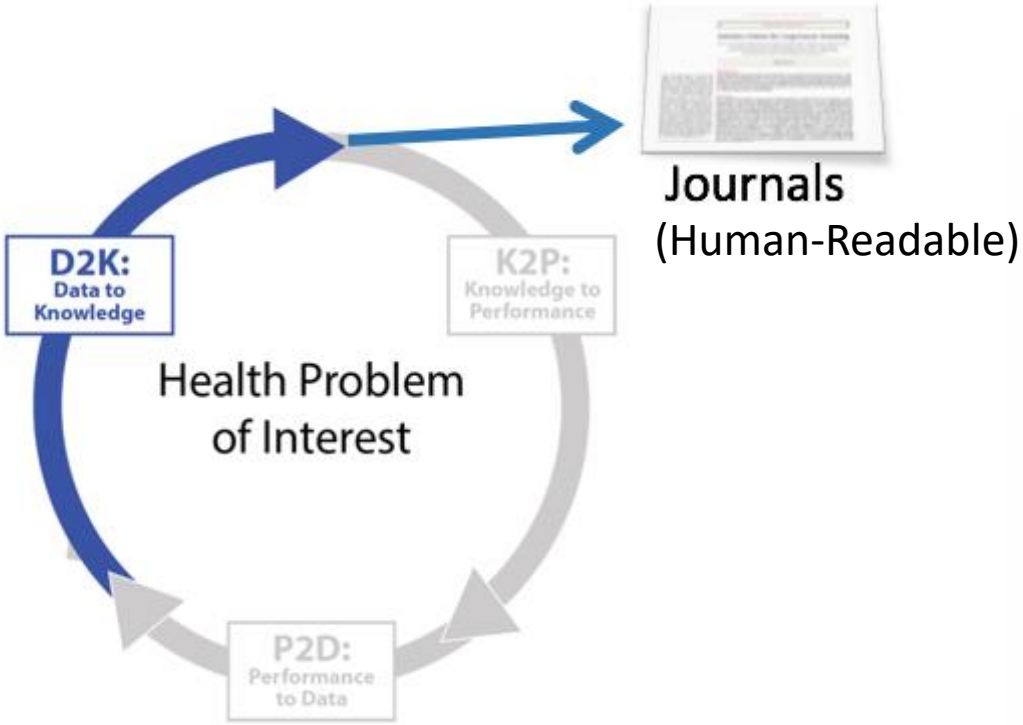
Computable: Code

Library

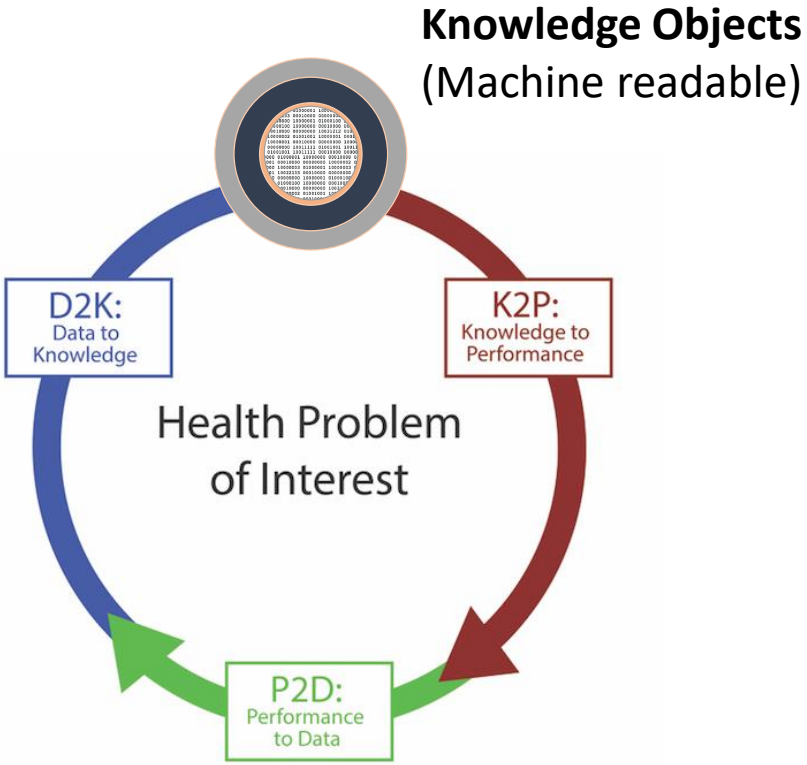


Expanded Library

Computable Knowledge Enables a Continuous Connection Between Discovery and Improvement (and Further Discovery)

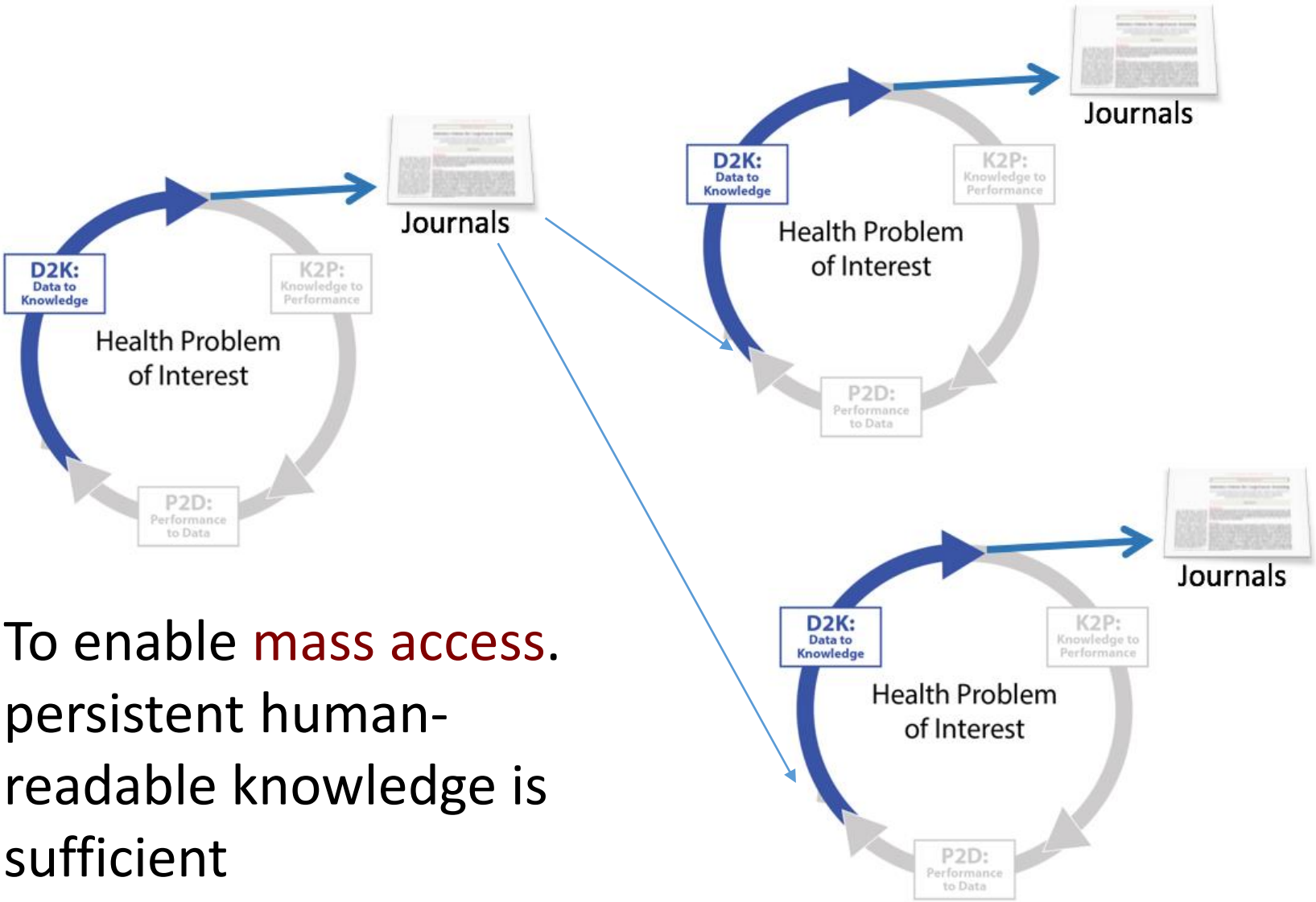


Discovery Systems



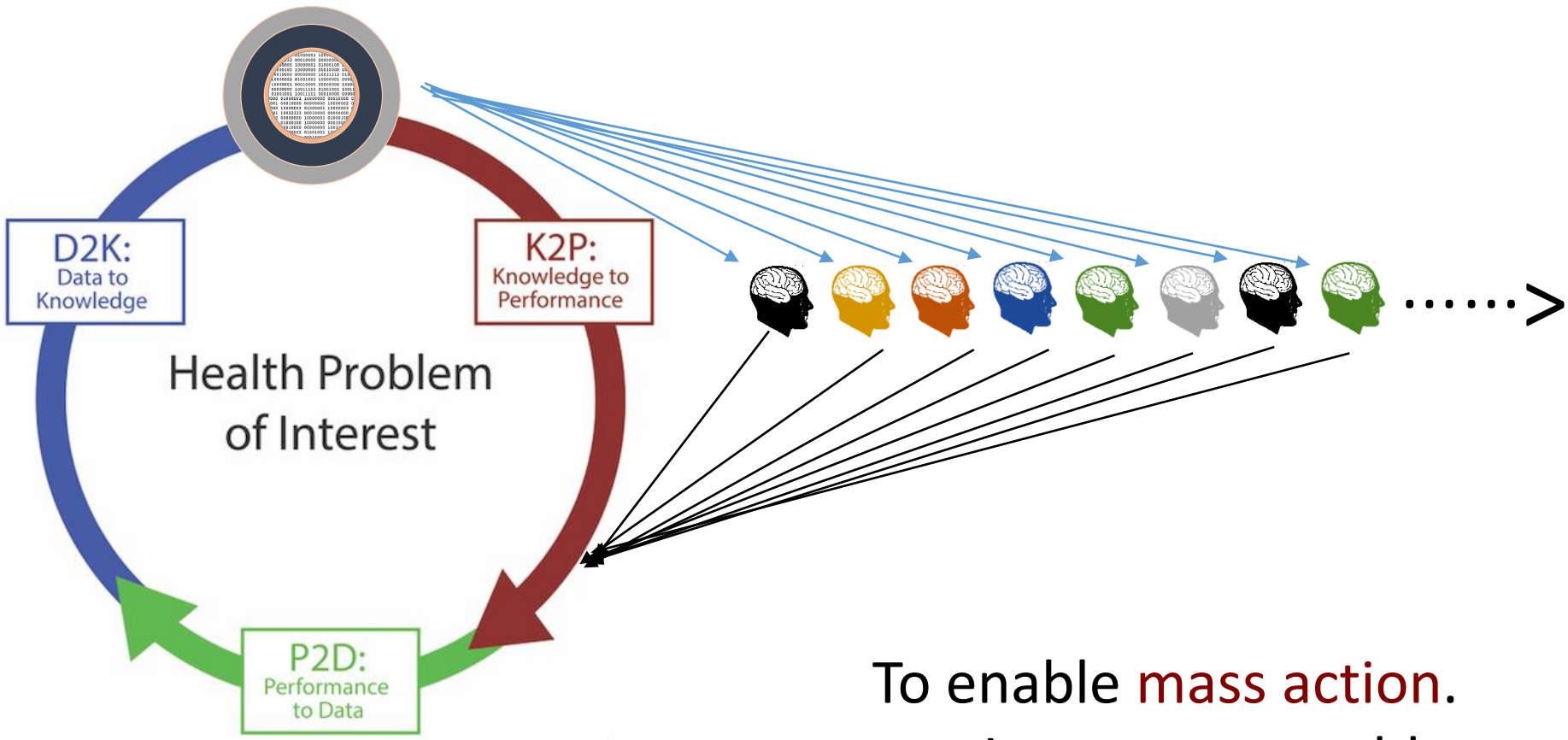
Learning Systems

Serial Discovery Systems Require Only Mass Access to Knowledge



To enable mass access. persistent human-readable knowledge is sufficient

Learning Systems Operate in Parallel, Requiring Mass Action



To enable mass action. persistent computable knowledge is essential.

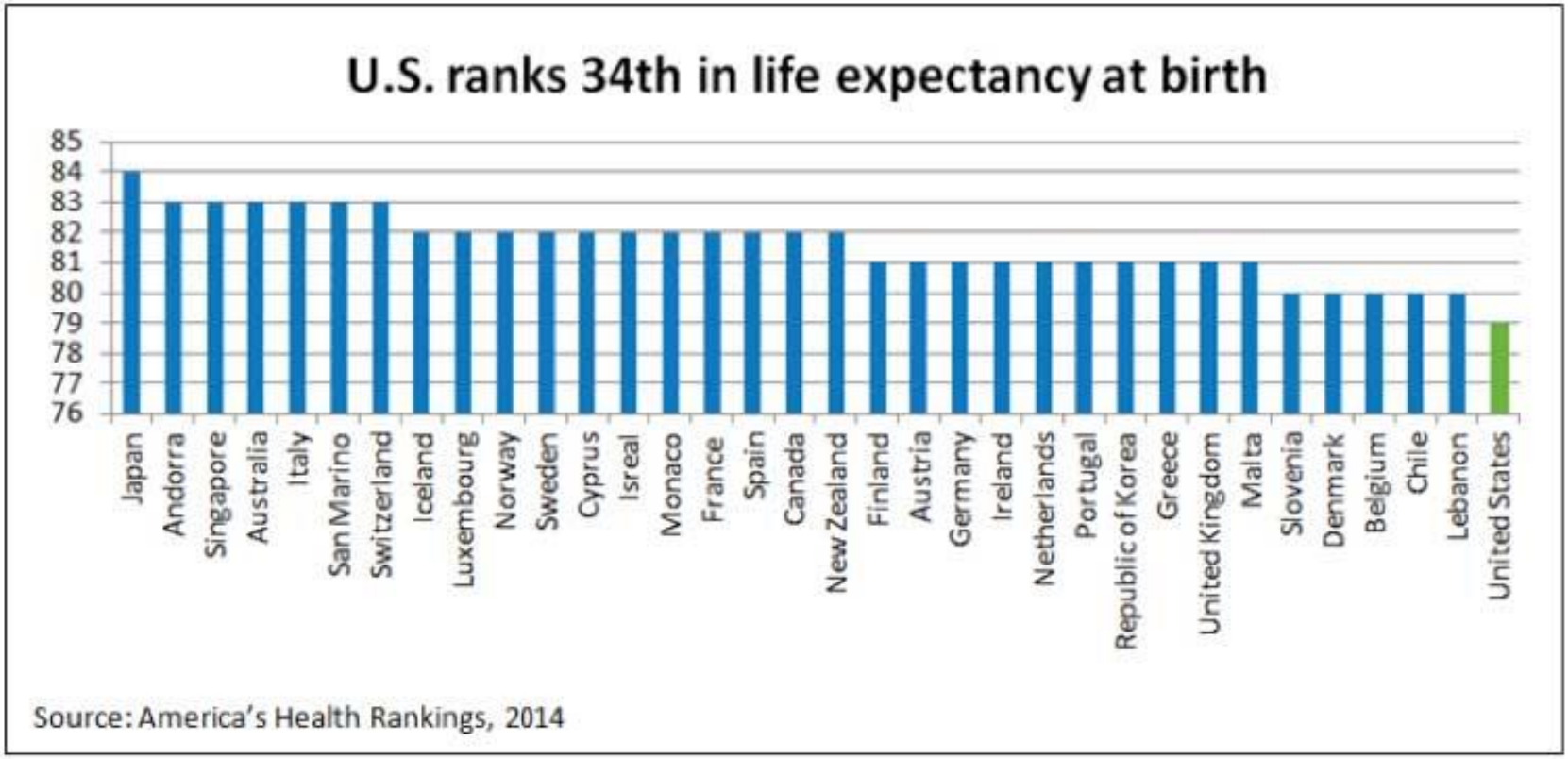
Outline

- Foundations for MCBK
- Why We're All Here
- MCBK Community & Workgroups
- Goals and Plans for This Meeting

Better Health…

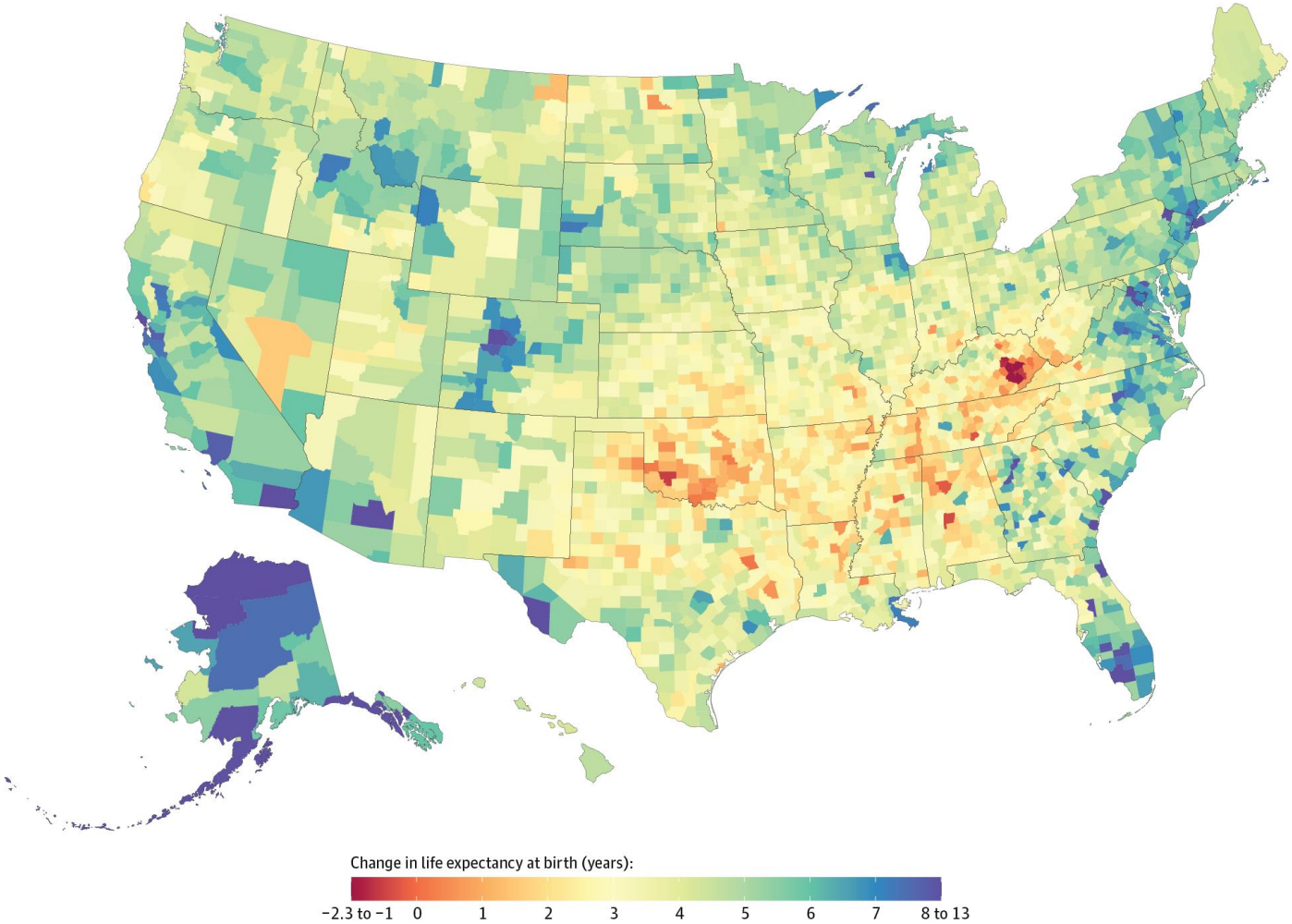
This is why we're all here today

Life Expectancies in U.S. and Other Countries



Source: https://www.apha.org/topics-and-issues/health-rankings

Inequalities in Life Expectancy Among US Counties, 1980 to 2014



Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA Intern Med.* 2017;177(7):1003–1011. doi:10.1001/jamainternmed.2017.0918

Many "Use Cases"

- Clinical: Applying pharmacogenomics to health care (making precision health actually work)
- Public Health: Detecting and automatically reporting notifiable diseases
- Translational ("T1+") Research: Using computable phenotypes to identify study-eligible persons
- Policy: Identifying non-compliant documents
- Education: Early detection of and remediation for students who will encounter difficulties

MCBK Manifesto

Mobilizing Computable Biomedical Knowledge (CBK):
A Manifesto

Preamble

Knowledge has the potential to improve health care, the health of individuals, and the health of populations. Every decision affecting health should be informed by the best available knowledge. For moral and ethical reasons, it is imperative that each and every member of society has access to what is known at the time they are making health-related choices and decisions.

It is no longer sufficient to represent knowledge in the form of printed words and static pictures. The increasingly rapid rate of scientific discovery needs knowledge representations that are more agile and amenable to scalability and mass action. This in turn can enable the continuous cycles of discovery and improvement envisioned as Learning Health Systems.

Contemporary digital technology enables knowledge to be represented in *computable* forms expressed in machine-executable code. Computable knowledge unleashes the potential of information technology to generate and deliver useful information—and particularly, decision-specific advice--to individuals and organizations with great speed on a world-wide scale. It is essential to take full advantage of these capabilities, while continuing established practices that validate knowledge, preserve it, and ensure that it can be trusted.

There is work to do to mobilize best available health knowledge for the greater good. To begin, biomedical knowledge in computable form must be made interoperable using open standards, and widely available so that it can be used to immediately impact health.

It is time for action on a global scale.

Computable Biomedical Knowledge

Computable Biomedical Knowledge is the result of an analytic and/or deliberative process about human health, or affecting human health, that is explicit, and therefore can be represented and reasoned upon using logic, formal standards, and mathematical approaches.

Vision

We are dedicated to:

Mobilizing biomedical knowledge that can support action toward improving human health. This should be done using computable formats that can be shared and integrated into health information systems and applications.

Efficiently and equitably serving the learning and knowledge needs of all participants, as well as the public good. This will work to significantly reduce health disparities.

Ensuring that the knowledge properly reflects the best and most current evidence and science. This will ensure that knowledge can be trusted for use to improve health and health care.

*Achieving this through evolution of an open Computable Biomedical Knowledge ecosystem dedicated to achieving the FAIR principles: making Computable Biomedical Knowledge easily findable, universally accessible, highly interoperable, and readily reusable.** The current interest in making data "FAIR" should be matched by equally intense interest in making knowledge "FAIR".

Mechanisms of Activity

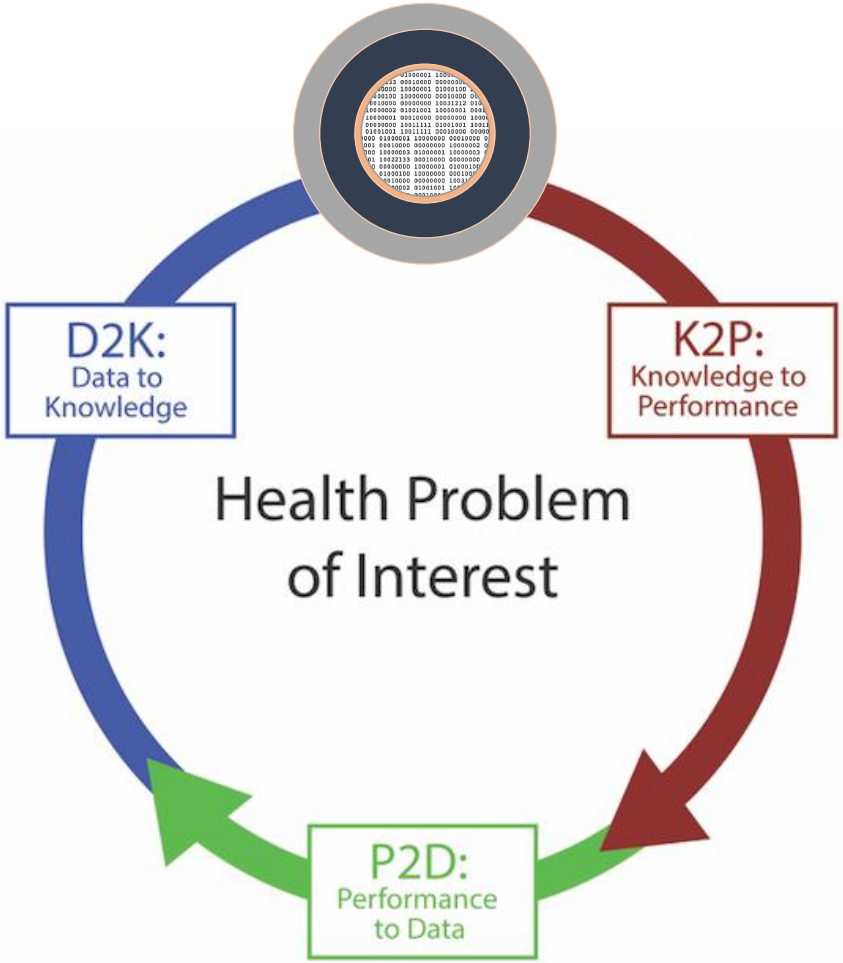
We believe that all of the following are important:

- The CBK Concept
  - Sustain the Computable Biomedical Knowledge ecosystem through public-private partnerships.
  - Establish broadly-based participatory governance of the ecosystem.
  - Make the ecosystem diverse and inclusive.
  - Explore the sciences of Computable Biomedical Knowledge collaboratively.
  - Be agile to reflect the increasingly rapid changes in knowledge.
- The CBK Technical System
  - Enable the ecosystem with open standards.
  - Build and uphold trust in Computable Biomedical Knowledge through the ecosystem.
  - Ensure robust and unbiased methods to support transparency and expose the currency, validity and provenance of Computable Biomedical Knowledge.
  - Implement the highest standards of privacy and security for all stakeholders.
  - Enable a pipeline that transitions knowledge from human-readable to fully computable through successive stages.
- The CBK Use/User System
  - Ensure the safe and effective use of Computable Biomedical Knowledge through the ecosystem.
  - Generate value for the creators of the knowledge, the users of the knowledge, and the general public.
  - Engender equity in health and in knowledge accessibility

*- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.

www.MobilizeCBK.org

Making Knowledge FAIR



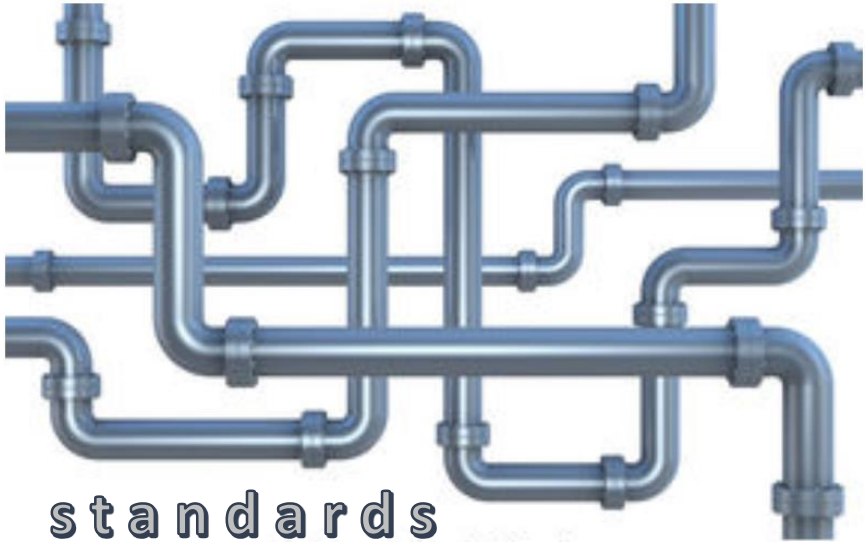
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

FAIR: https://www.force11.org/group/fairgroup/fairprinciples

Access Fuels Innovation

Data

Knowledge



standards
infrastructure
policies

Research, discovery, generation of evidence

Applications/Action

-- targeted, personalized, useful, usable, …

Outline

- Foundations for MCBK
- Opportunities for Action
- MCBK Community & Workgroups
- Goals and Plans for This Meeting

Standards for MCBK

Bob Greenes
Co-Chair
Arizona State University

Bruce Bray
Co-Chair
University of UT

Technical Infrastructure for MCBK

Chris Shaffer
Co-Chair
University of California San Francisco

Leslie McIntosh
Co-Chair
Research Data Alliance

Policy and Coordination to Ensure Quality and Trust

Blackford Middleton
Co-Chair
Apervita, Inc.

Jody Platt
Co-Chair and MCBK Liaison
University of Michigan

Sustainability for Mobilization and Inclusion

Jerry Perry
Co-Chair
University of Arizona

Chris Dymek
Co-Chair
AHRQ

Outline

- Foundations for MCBK
- Why We're All Here
- MCBK Community & Workgroups
- Goals and Plans for This Meeting

Who is here?

Meeting highlights

- 190+ participants (!)
- 24 posters (!)
- 17 technical demos (!)
- 4 working groups – action sessions
- Remarks and perspectives from national and international leaders
- Presentations and discussions
  - Use cases, critical stakeholders

Meeting Goals

- **Strengthen** foundation of shared recognition of values and principles for mobilizing CBK
- **Frame** and address important dimensions for mobilizing CBK
- **Advance** workgroups action plans
- **Identify** priorities
- **Plan** next steps
- **Grow** the MCBK community

Today's Plan - Success Depends on You

- Remarks - Dr. Patricia Brennan (NLM)
- Panel - MCBK Use Cases
- Lunch & Technical Demonstrations & Posters
- Panel - Engaging Critical Stakeholders
- Breakout Sessions – Workgroup Action Sessions
- "Open Mic"
- Reception at Rock Bottom Brewery

Plan for Tomorrow

- Workgroup Updates and Charge for Day 2
- Remarks - Dr. Don Rucker (ONC)
  - Dr. Dipak Kalra (i~HD)
- Lunch & Technical Demonstrations
- Workgroup Action Sessions
- Panel - Meeting Reflections
- Closing

Steering Committee

- Julia Adler-Milstein
- Bruce Bray
- Milton Corn
- Chris Dymek
- Peter Embi
- Charles Friedman
- Bob Greenes
- Stan Huff
- Dipak Kalra
- Nancy Lorenzi
- Leslie McIntosh
- Blackford Middleton
- Mark Musen
- Jodyn Platt
- Jerry Perry
- Rachel Richesson
- Chris Shaffer
- Umberto Tachinardi
- John Wilbanks

MCBK

Mobilizing Computable Biomedical Knowledge

1. Register for workgroups

2. Sign-up to view Technical demos

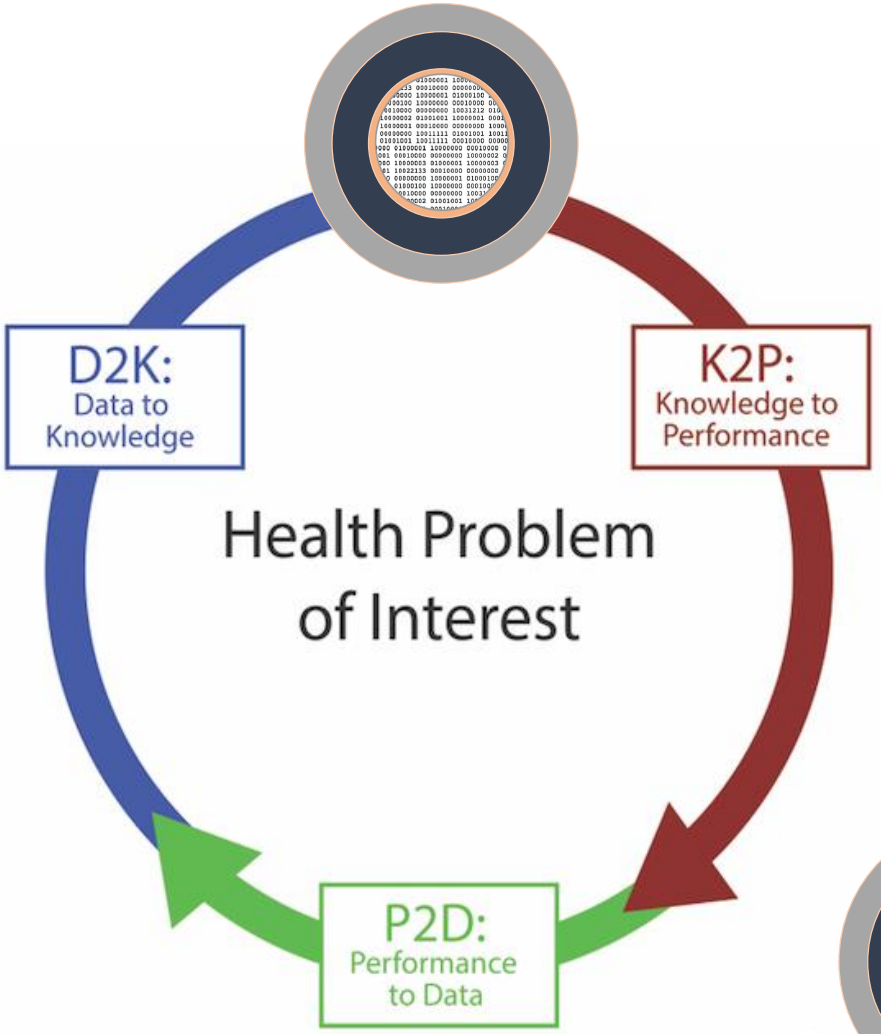
3. #MobilizeMCBK tag @DLHSumich

4. Restrooms location

5. Additional information

MCBK

Mobilizing Computable Biomedical Knowledge

#MobilizeCBK

Onward!