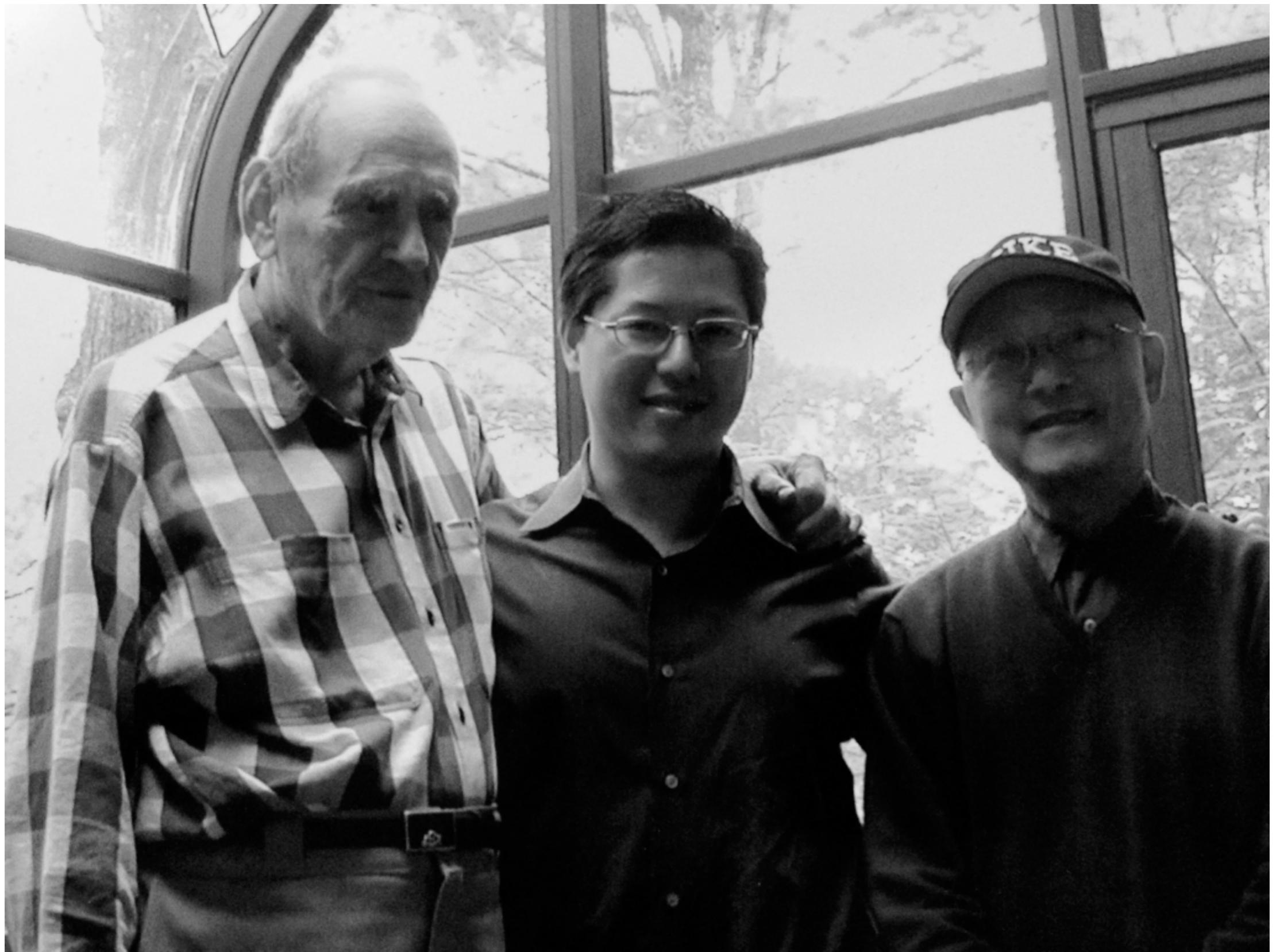


MMCI Practical Data Science

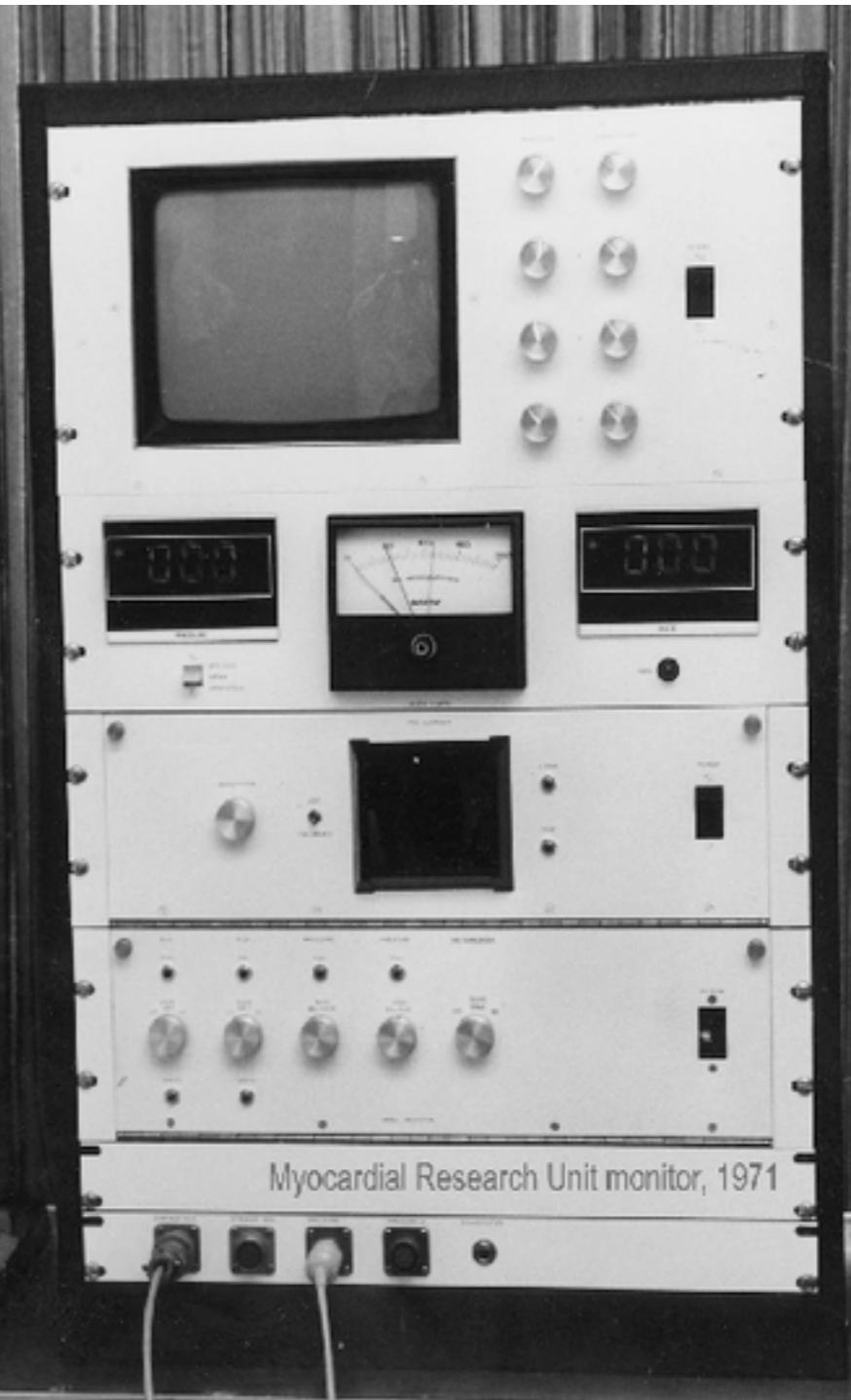
lecture 01

erich s. huang
erich.huang@duke.edu

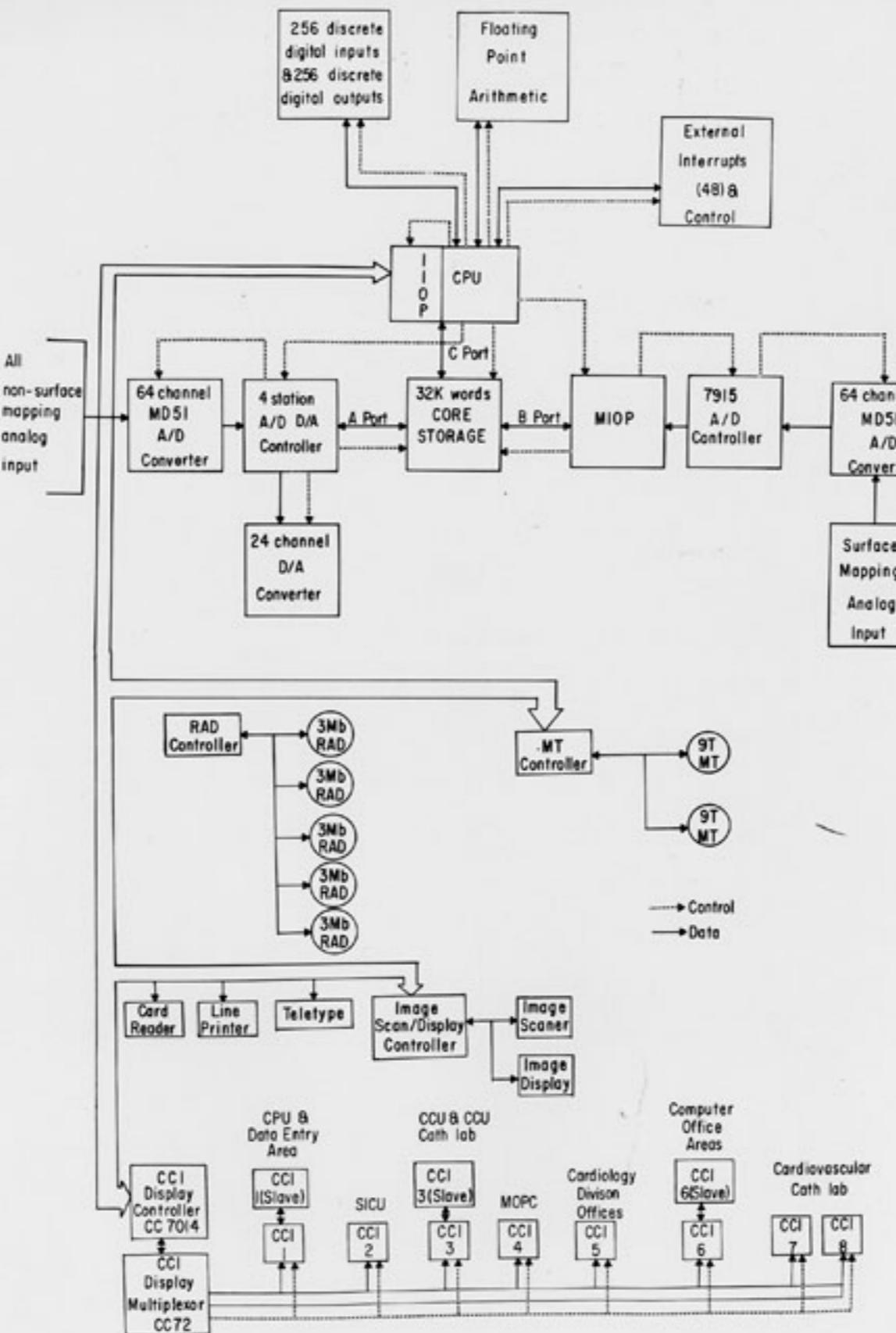


Improving Patient Care by Capturing Computerized data:

A glimpse into the creation of The Duke Databank for Cardiovascular Disease



HARDWARE CONFIGURATION



COST CONSTRAINTS
IMPLEMENTATION
DATA-DRIVEN



MONEYBALL
IN THEATERS SEPTEMBER 23



DukeHealth

Inpatient Admissions 63,312

Outpatient Visits 1,280,514

Surgeries/Endoscopies 85,248

ER visits 66,860

Hospital Labs 5,428,178





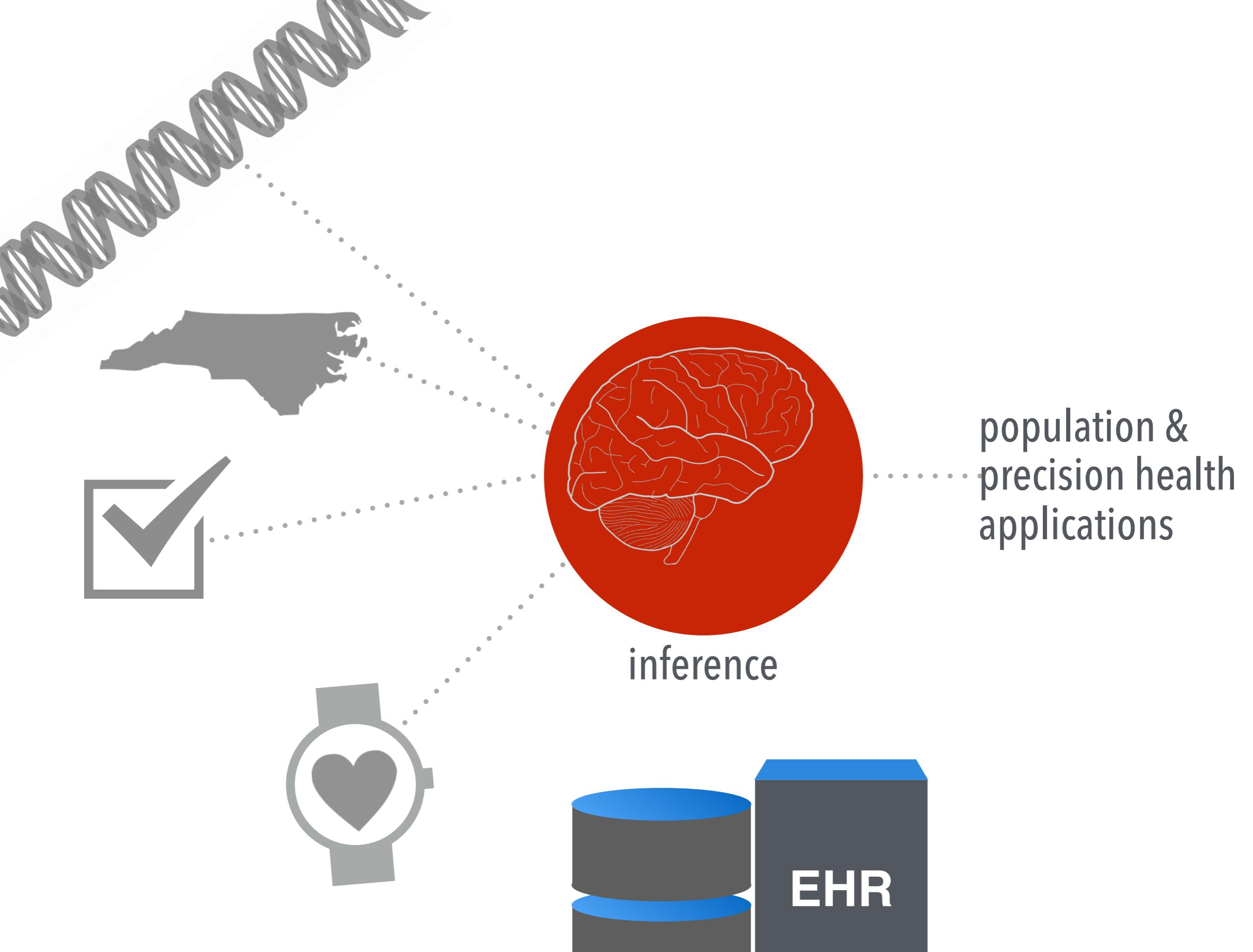
**Necessary but
not sufficient**

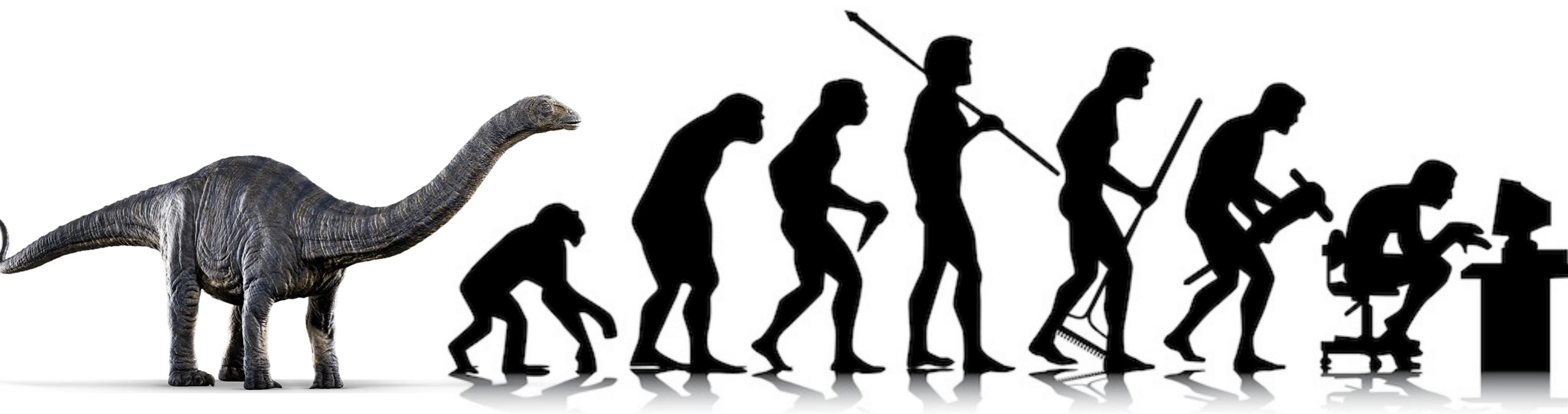
... IT'S STILL EARLY IN THE EVOLUTION...





< 1 pound







"5 Tool Player"



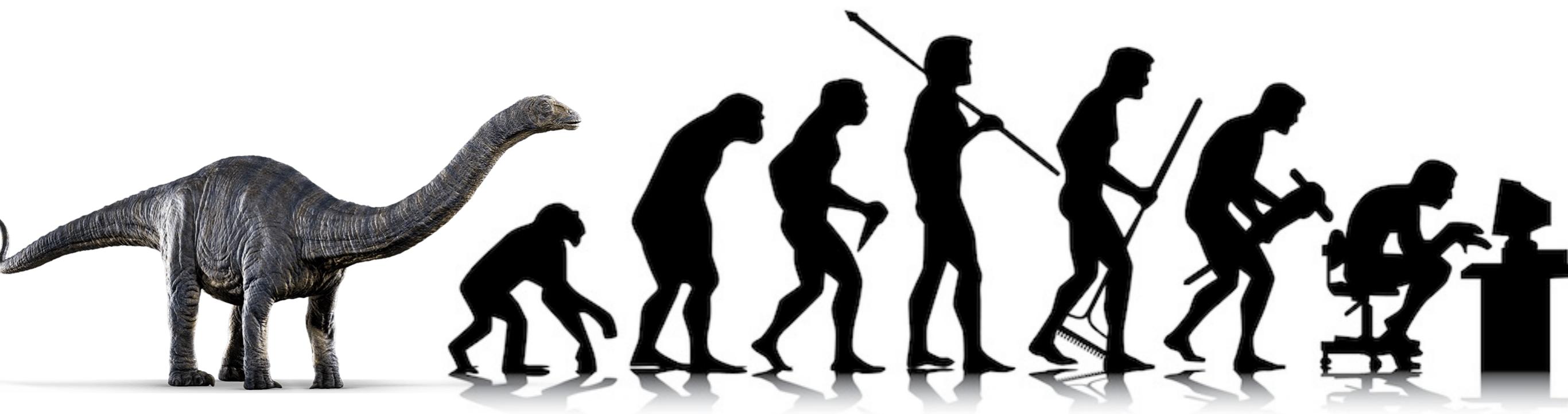
AP

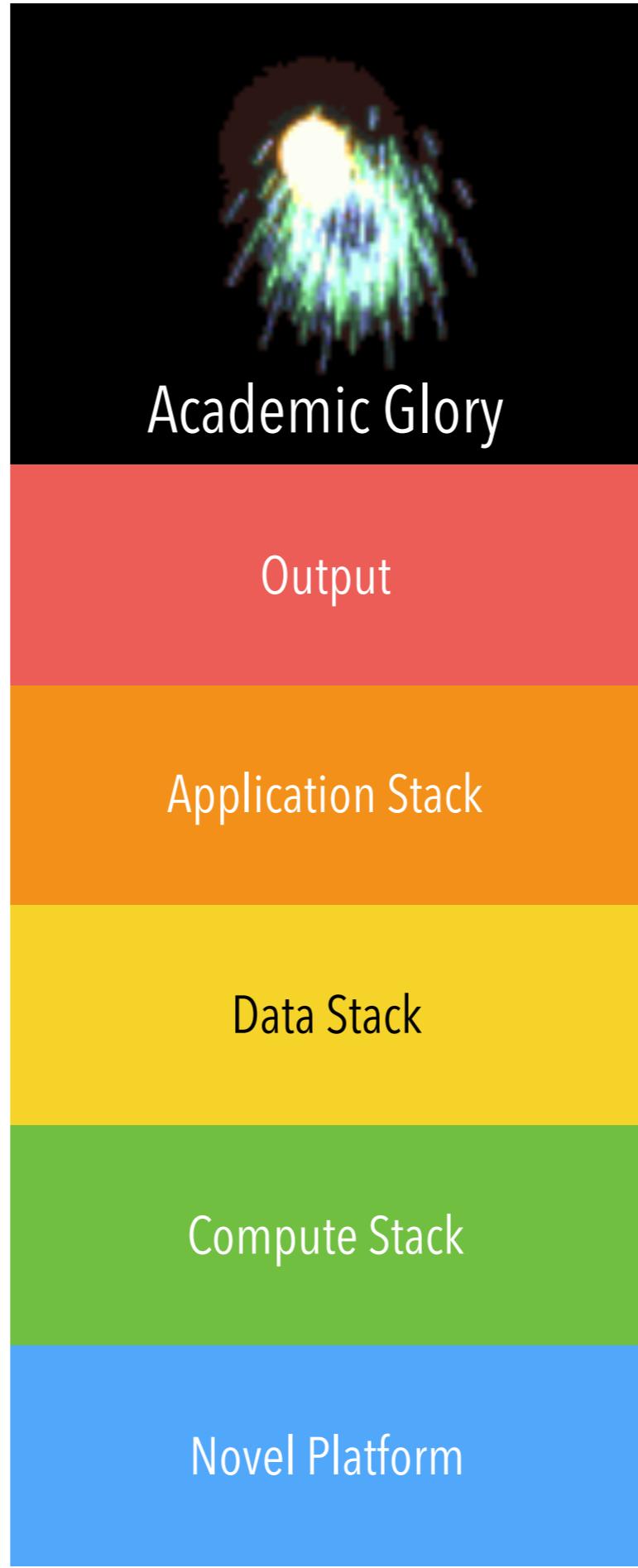
"Catcher with a bad elbow converted to 1st baseman"

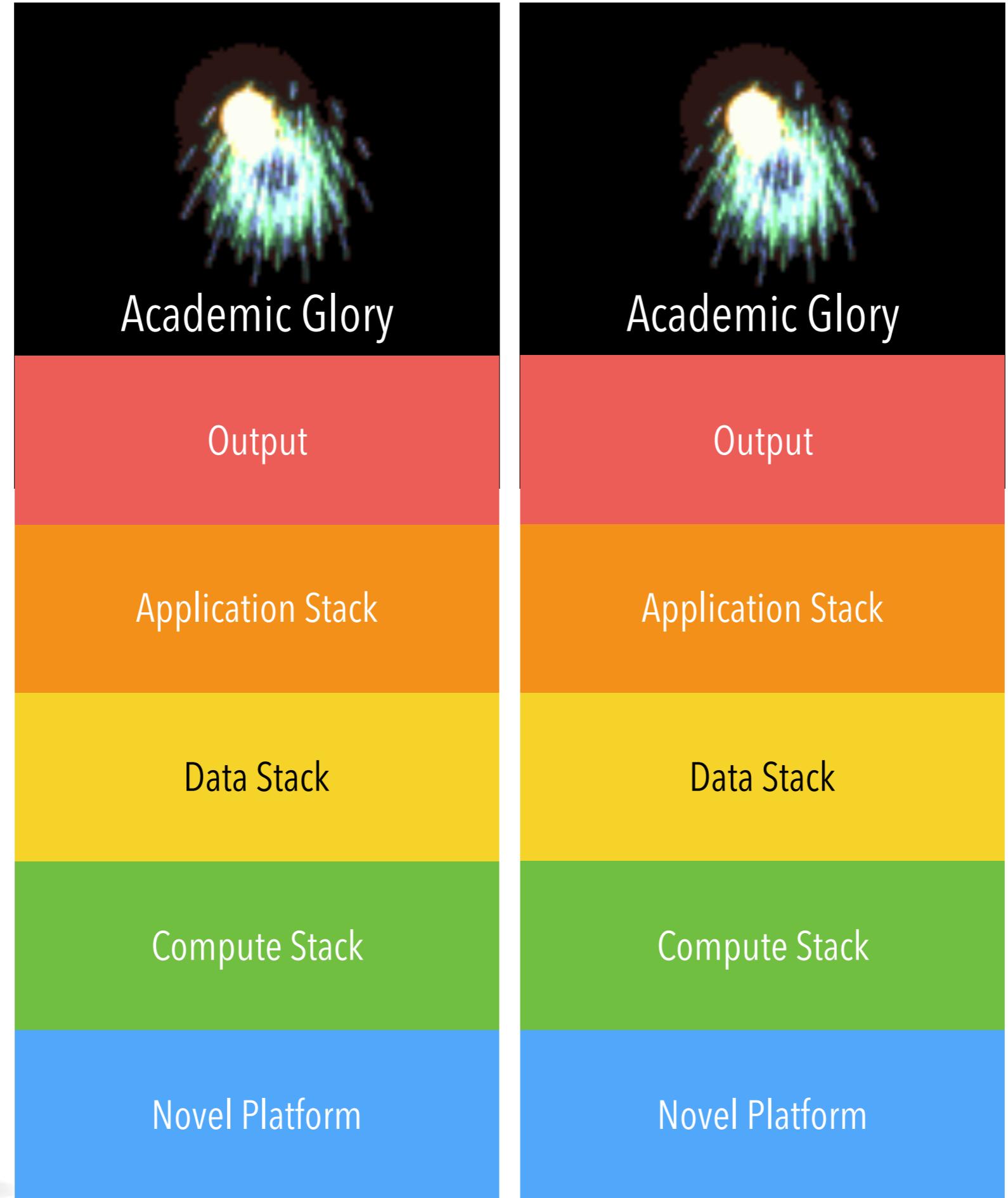
IDENTIFYING PREVIOUSLY HIDDEN VALUE



AN ACADEMIC HEALTH SYSTEM THAT LEARNS







Epic
 **Cerner**TM

SCALABLE? IMPLEMENTABLE?

The NEW ENGLAND
JOURNAL of MEDICINE

VOL. 366 NO. 2

ESTABLISHED IN 1812

JANUARY 12, 2012

NEJM.ORG

108 THIS WEEK AT NEJM.ORG

PERSPECTIVE
101 The Politics of Emergency Contraception
A.J.J. Wood, J.M. Drazen, and M.F. Greene
103 The Constitutionality of the ACA's Medicaid-Expansion Mandate I.G. Cohen and J.F. Blumstein
105 Expanding Eligibility, Cutting Costs — A Medicaid Update J.K. Iglesias
104 Achieving Accountable Care — "It's Not About the Bike" J. Walker and A. McKethan

ORIGINAL ARTICLES
109 Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer J. Baselga and Others
120 Subclinical Atrial Fibrillation and the Risk of Stroke J.S. Healey and Others
130 Cardiac Arrest during Long-Distance Running Races J.H. Kim and Others
141 Germline Mutations in HOXB13 and Prostate-Cancer Risk C.M. Ewing and Others
SPECIAL ARTICLE
150 Fitness Memberships and Favorable Selection in Medicare Advantage Plans A.L. Cooper and A.N. Trivedi

CLINICAL PRACTICE
158 Hidradenitis Suppurativa G.B.E. Jemec
IMAGES IN CLINICAL MEDICINE
165 Paragonimiasis M.A. Barrientos and A.U. Carrasco
e5 Jugular Venous C-V Wave in Severe Tricuspid Regurgitation N.B. Senguttuvan and G. Karthikeyan

CASE RECORDS OF THE MASSACHUSETTS GENERAL HOSPITAL
166 A Man with Persistent Ulcers on the Hands D. Kroshinsky, M.P. Hoang, and R.P. Hasserjian
EDITORIALS
176 HER2 Therapy — An Abundance of Riches W.J. Gralishar
178 How Much Atrial Fibrillation Is Too Much Atrial Fibrillation? G. Lamas
CLINICAL IMPLICATIONS OF BASIC RESEARCH
181 Diet and Intestinal Immunity H. Tilg

CORRESPONDENCE
184 Acute Coronary Thrombosis in Boston Marathon Runners An Impedance Threshold Device in Out-of-Hospital Cardiac Arrest Breast-Cancer Adjuvant Therapy with Zoledronic Acid Inflammatory Bowel Disease and ADAM17 Deletion Breast-Cancer Screening

Owmed & published by the MASSACHUSETTS MEDICAL SOCIETY © 2012.
All rights reserved. ISSN 0028-4793.

Academic Glory

The NEW ENGLAND
JOURNAL of MEDICINE

VOL. 366 NO. 2

ESTABLISHED IN 1812

JANUARY 12, 2012

NEJM.ORG

108 THIS WEEK AT NEJM.ORG

PERSPECTIVE
101 The Politics of Emergency Contraception A.J.J. Wood, J.M. Drazen, and M.F. Greene
103 The Constitutionality of the ACA's Medicaid-Expansion Mandate I.G. Cohen and J.F. Blumstein
105 Expanding Eligibility, Cutting Costs — A Medicaid Update J.K. Iglesias
104 Achieving Accountable Care — "It's Not About the Bike" J. Walker and A. McKethan

ORIGINAL ARTICLES
109 Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer J. Baselga and Others
120 Subclinical Atrial Fibrillation and the Risk of Stroke J.S. Healey and Others
130 Cardiac Arrest during Long-Distance Running Races J.H. Kim and Others
141 Germline Mutations in HOXB13 and Prostate-Cancer Risk C.M. Ewing and Others
SPECIAL ARTICLE
150 Fitness Memberships and Favorable Selection in Medicare Advantage Plans A.L. Cooper and A.N. Trivedi

CLINICAL PRACTICE
158 Hidradenitis Suppurativa G.B.E. Jemec
IMAGES IN CLINICAL MEDICINE
165 Paragonimiasis M.A. Barrientos and A.U. Carrasco
e5 Jugular Venous C-V Wave in Severe Tricuspid Regurgitation N.B. Senguttuvan and G. Karthikeyan

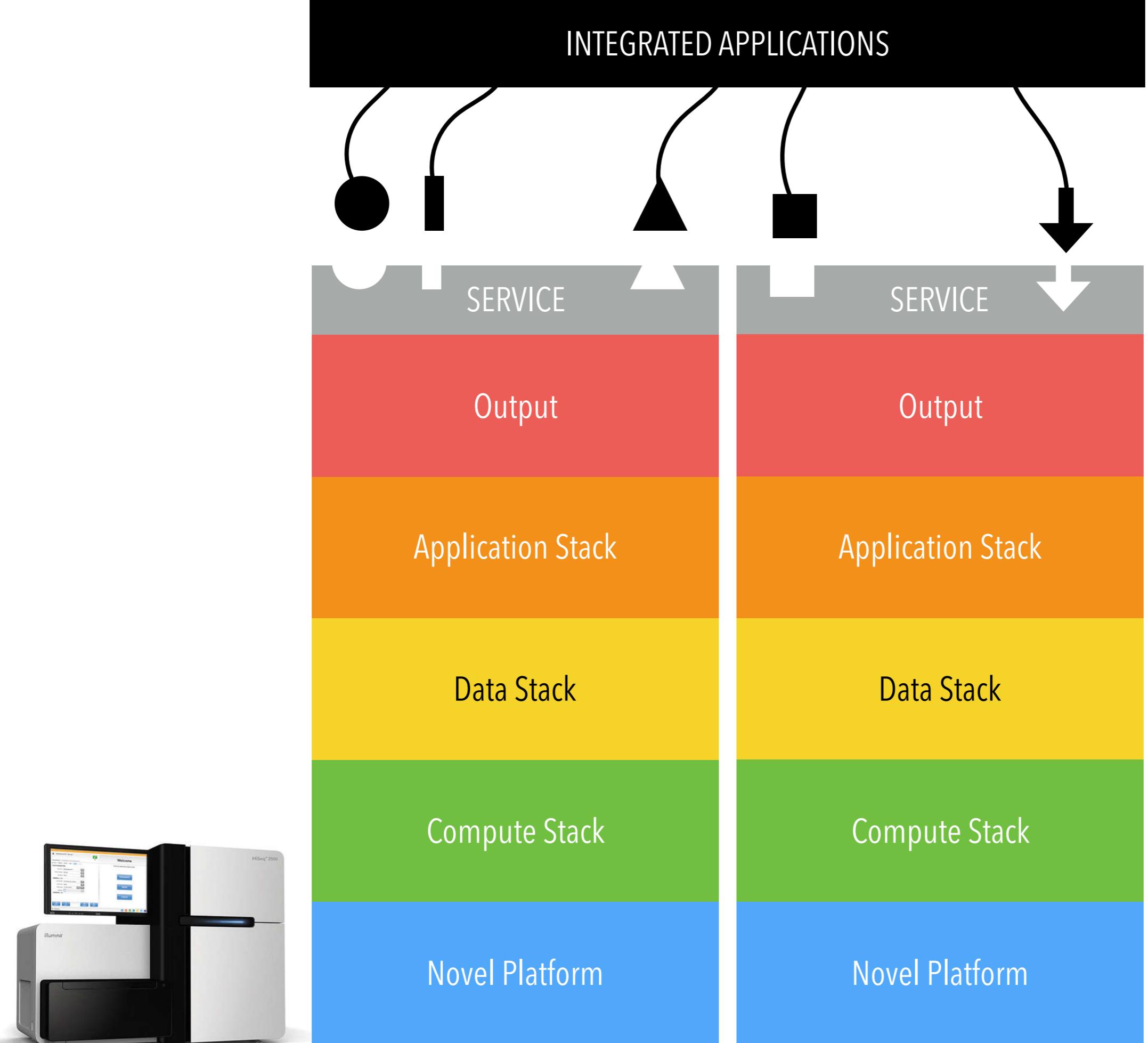
CASE RECORDS OF THE MASSACHUSETTS GENERAL HOSPITAL
166 A Man with Persistent Ulcers on the Hands D. Kroshinsky, M.P. Hoang, and R.P. Hasserjian
EDITORIALS
176 HER2 Therapy — An Abundance of Riches W.J. Gralishar
178 How Much Atrial Fibrillation Is Too Much Atrial Fibrillation? G. Lamas
CLINICAL IMPLICATIONS OF BASIC RESEARCH
181 Diet and Intestinal Immunity H. Tilg

CORRESPONDENCE
184 Acute Coronary Thrombosis in Boston Marathon Runners An Impedance Threshold Device in Out-of-Hospital Cardiac Arrest Breast-Cancer Adjuvant Therapy with Zoledronic Acid Inflammatory Bowel Disease and ADAM17 Deletion Breast-Cancer Screening

Owmed & published by the MASSACHUSETTS MEDICAL SOCIETY © 2012.
All rights reserved. ISSN 0028-4793.

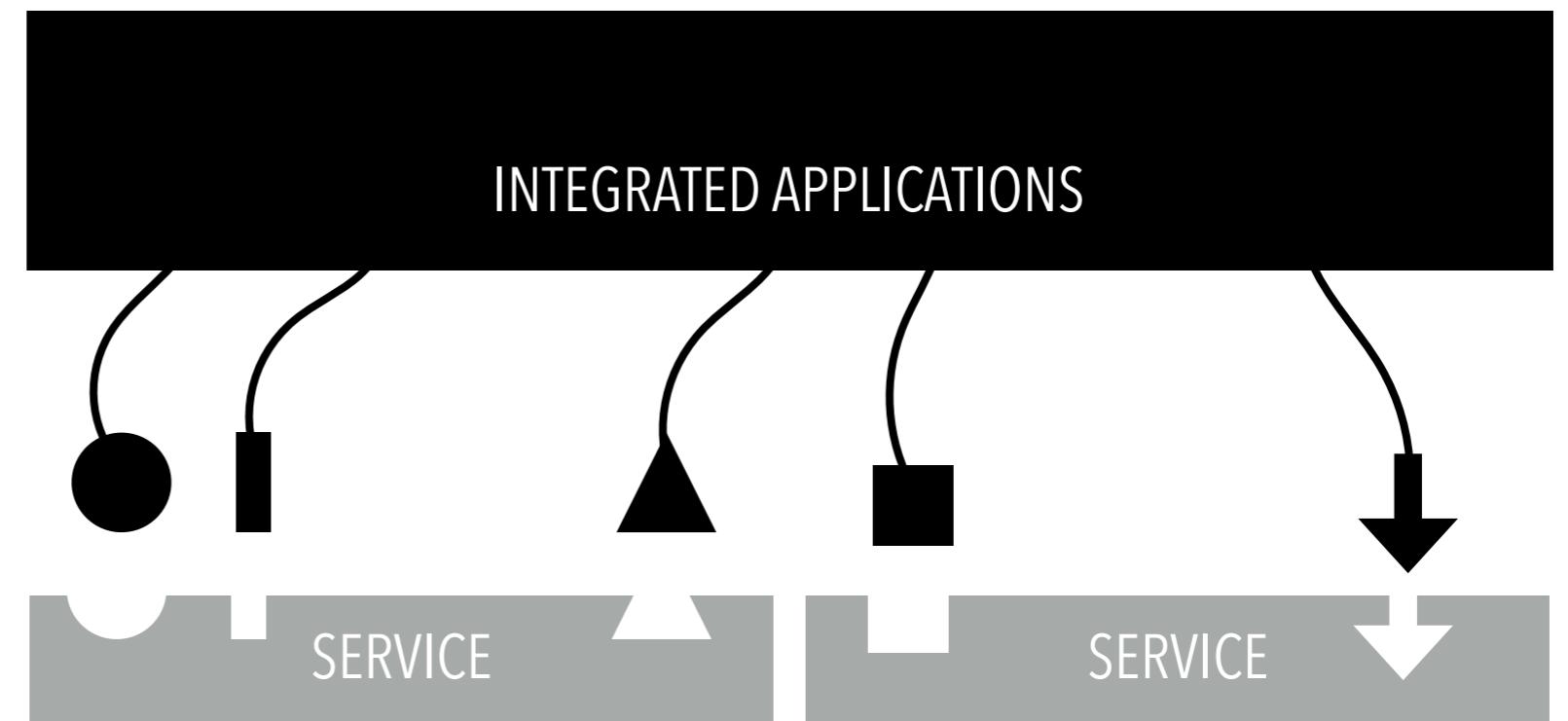
Academic Glory

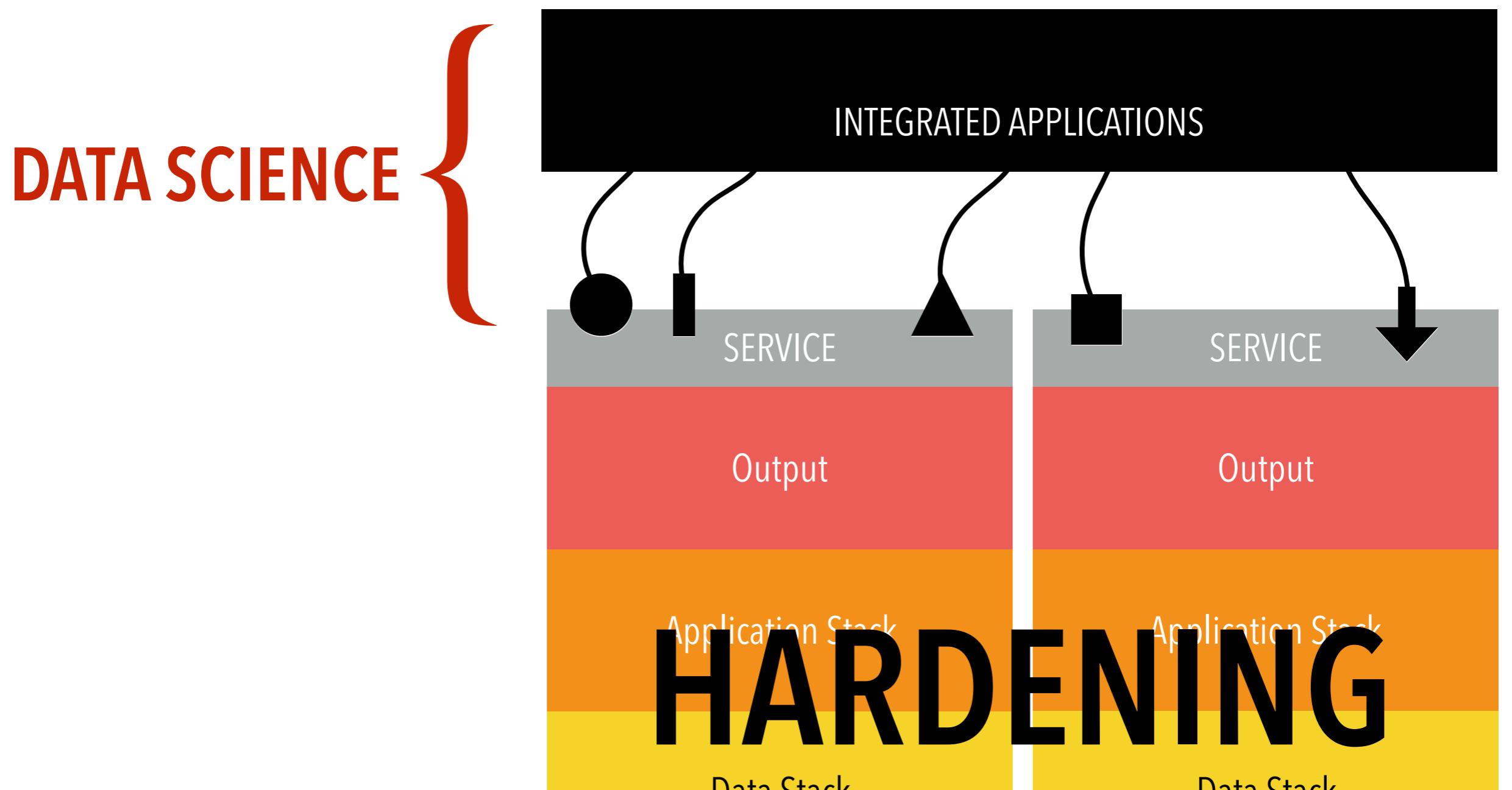
INTEGRATED APPLICATIONS



Epic
Cerner

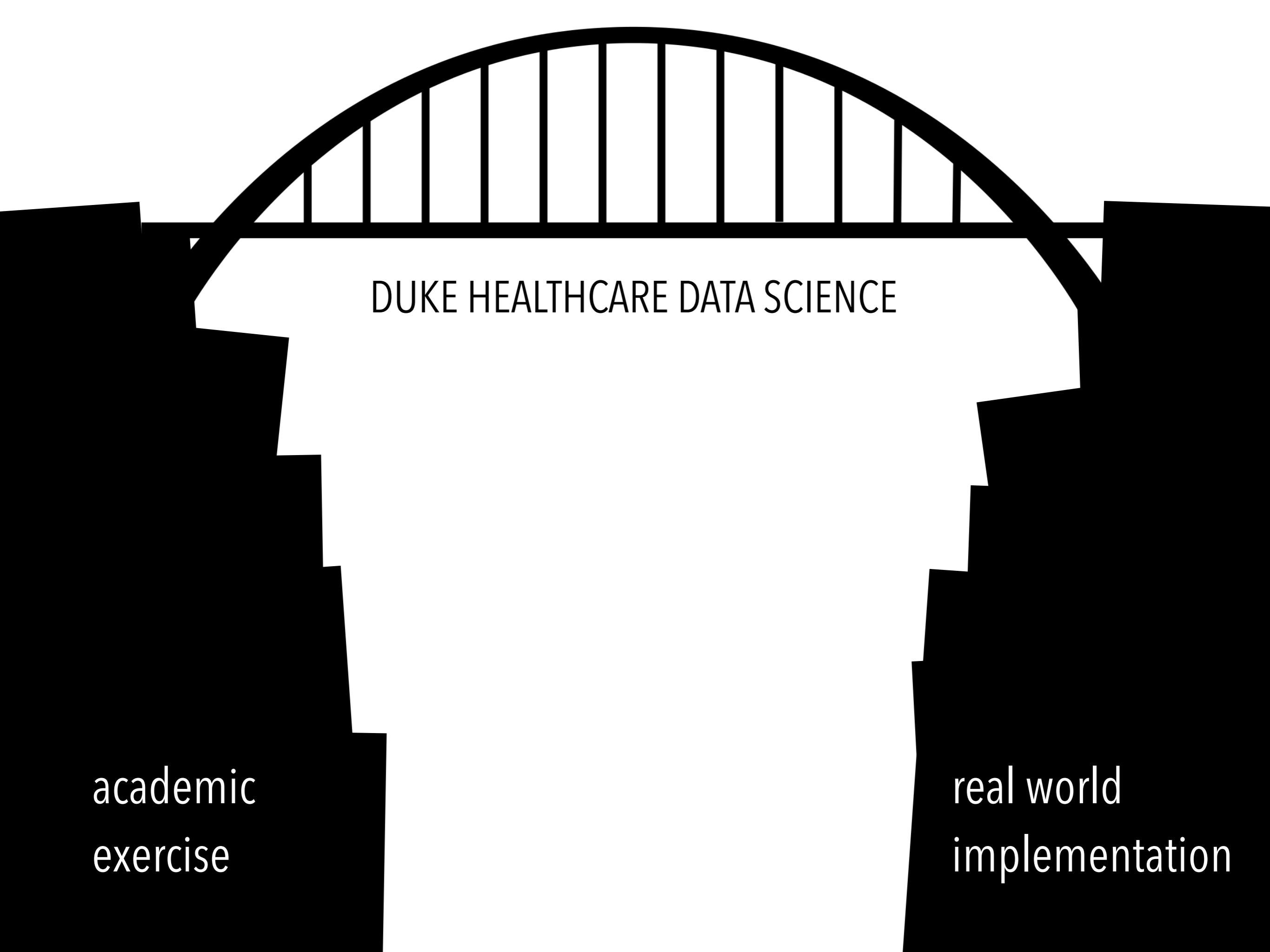
DATA SCIENCE





Knowledge among computer scientists about how to think of and approach the analysis of data is limited, just as the knowledge of computing environments by statisticians is limited. A merger of the knowledge bases would produce a powerful force for innovation.

–Bill Cleveland on *Data Science* (2001)



DUKE HEALTHCARE DATA SCIENCE

academic
exercise

real world
implementation

[BRIEFING ROOM](#)[ISSUES](#)[THE ADMINISTRATION](#)[PARTICIPATE](#)[1600 PENN](#)

Search

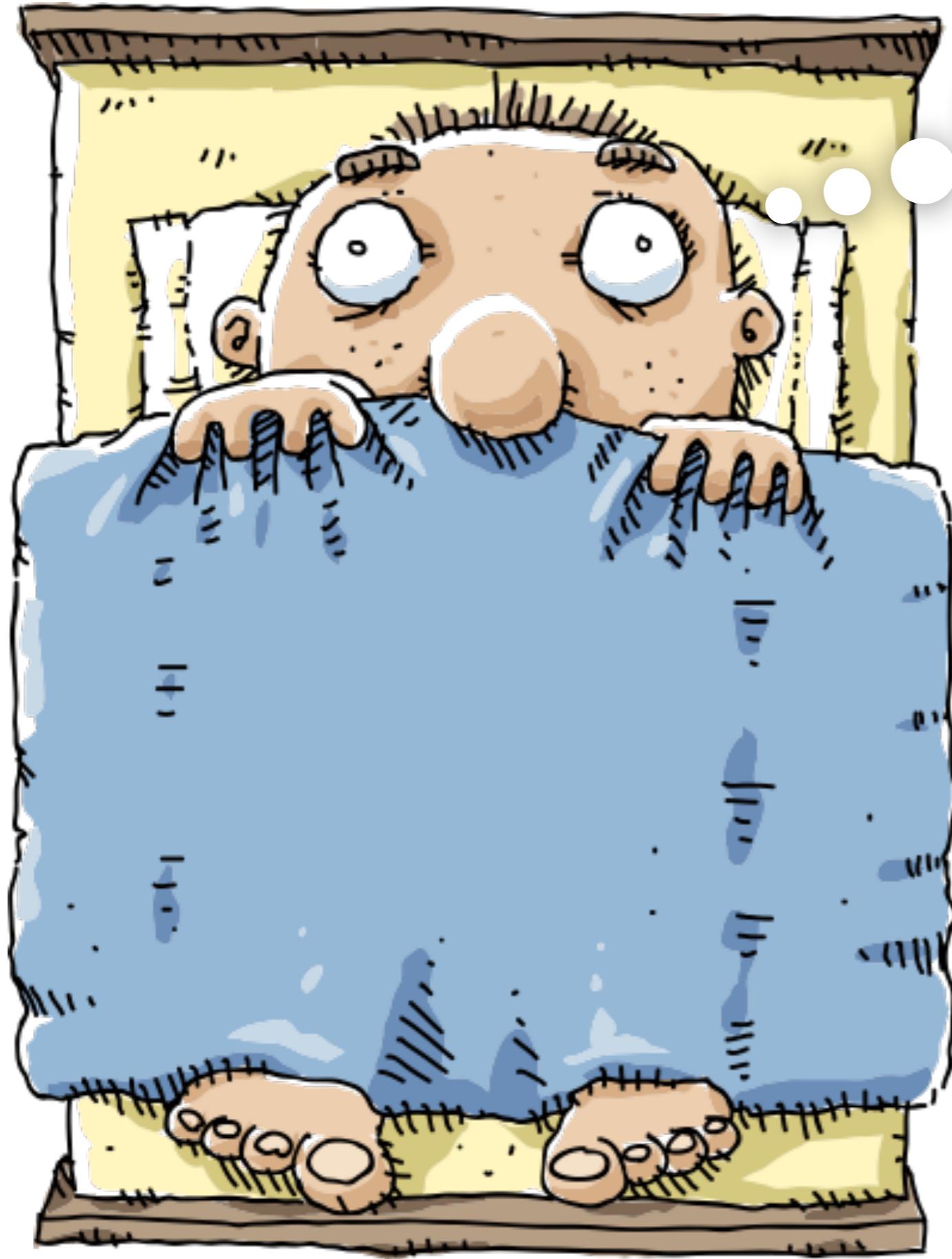


THE PRECISION MEDICINE INITIATIVE

[PRECISION MEDICINE](#)[THE INITIATIVE](#)[PRINCIPLES](#)[STORIES](#)

GO TO TOP

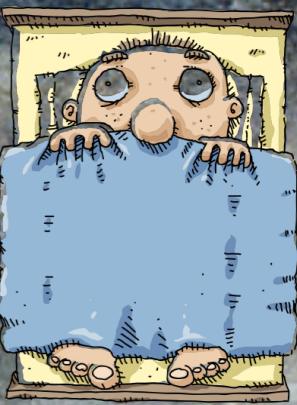
"Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as



PRECISION MEDICINE INITIATIVE

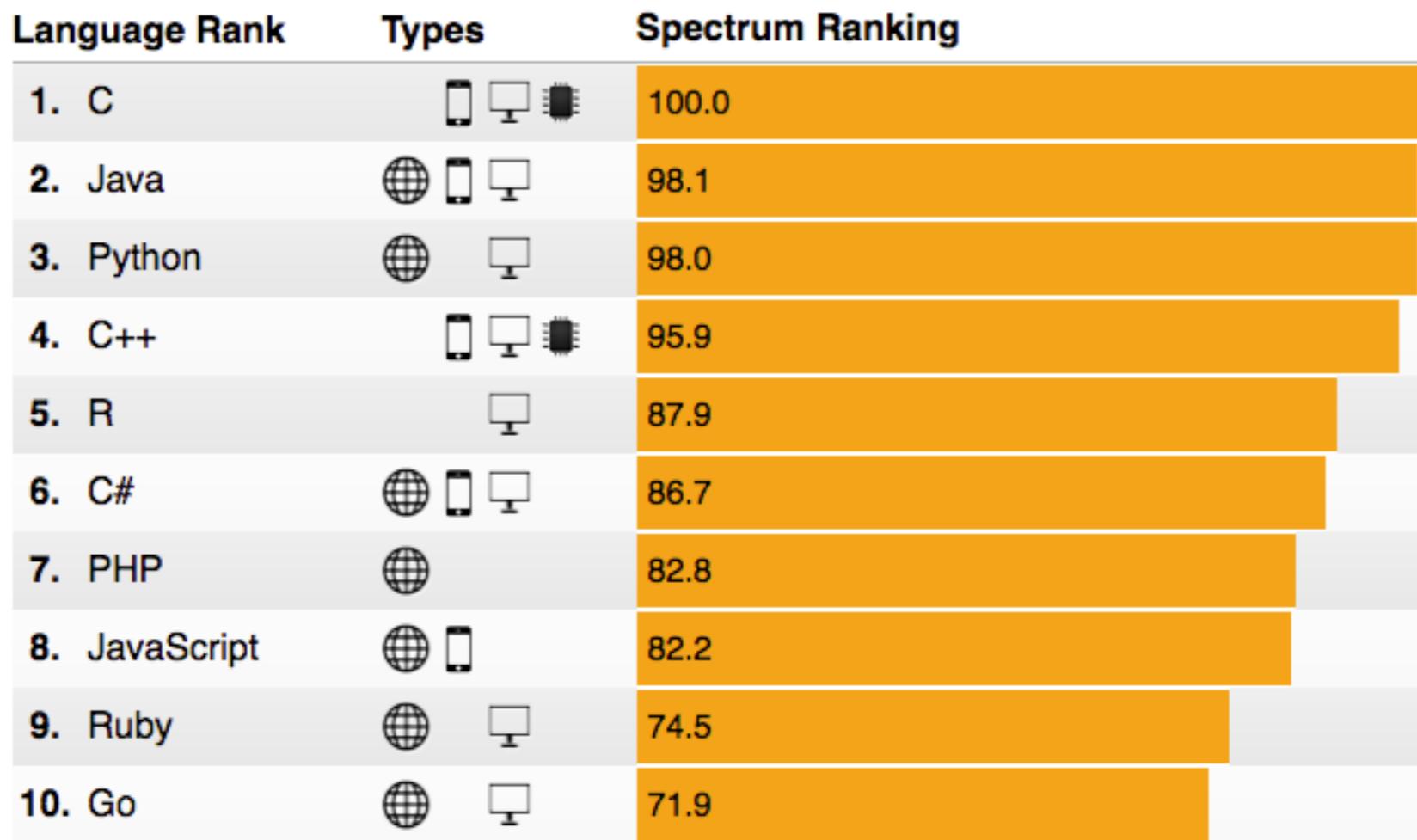


NIGHTMARE #1



Book of Genesis

*8 — “Therefore its name is called Babel,
because there the Lord confused the
language of all the earth; and from there the
Lord scattered them abroad over the face of
all the earth.”*



BTW: MUMPS/Cache isn't even a blip

The
cheesecake
Factory® & APIs
Application Programming Interfaces

The Cheesecake Factory®



kitchen

<p>sandwiches</p> <p><i>All Served with French Fries or Green Salad Sweet Potato Fries 1.00 extra</i></p> <p>RENEE'S SPECIAL 10.50 <i>One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich, a Cup of Our Soup and a Small Green Salad with a Small Caesar Salad 1.00 extra</i></p> <p>CHICKEN SALAD SANDWICH 11.50 <i>Homemade with Roasted Almonds, Lettuce, Tomato and Mayonnaise. Served on Grilled Brioche Bread</i></p> <p>GRILLED CHICKEN AND AVOCADO CLUB 12.95 <i>Grilled Chicken Breast with Avocado, Bacon, Tomato, Melted Swiss and Herb Mayonnaise</i></p> <p>THE CLUB 12.95 <i>Freshly Roasted Turkey Breast, Bacon, Lettuce, Tomato and Mayonnaise on White Toast</i></p> <p>THE NAVAJO 11.95 <i>Warm Fry-Bread Stuffed with Grilled Chicken, Avocado, Lettuce, Tomato, Red Onion and Mayonnaise</i></p> <p>CUBAN SANDWICH 11.95 <i>Slow Roasted Pork, Ham, Swiss Cheese, Pickles, Mustard and Mayonnaise on a Grilled and Pressed Cuban-Style Roll</i></p> <p>CHICKEN PARMESAN SANDWICH 11.95 <i>Tender Chicken Lightly Coated in Parmesan Breadcrumbs, Roasted Peppers, Tomato Sauce and Melted Cheese on a Freshly Grilled French Roll</i></p> <p>BLACKENED CHICKEN SANDWICH 11.95 <i>Grilled with Melted Fontina Cheese, Lettuce, Tomato, Red Onion and Spicy Mayonnaise</i></p> <p>CALIFORNIA CHEESESTEAK 11.95 <i>Thinly Sliced Grilled Steak Covered with Sautéd Mushrooms, Onions, Peppers and Cheese on a Toasted Roll</i></p> <p>GRILLED SHRIMP & BACON CLUB 13.95 <i>Charbroiled Shrimp, Bacon, Lettuce and Tomato with Our Special Dressing</i></p> <p>CRABCAKE SANDWICH 14.50 <i>With Lettuce, Tomato and Tartar Sauce on Our Brioche Bun</i></p> <p>GRILLED CHEESE 7.95 <i>Served on Egg Bread with Your Choice of Cheese</i></p>	<p>cheesecakes</p> <p>ORIGINAL <i>The One that Started it All! Our Famous Creamy Cheesecake with a Graham Cracker Crust and Sour Cream Topping</i></p> <p>FRESH STRAWBERRY <i>The Original Topped with Glazed Fresh Strawberries. Our Most Popular Flavor for 30 Years! REESE'S® PEANUT BUTTER CHOCOLATE CAKE CHEESECAKE*</i> <i>Reese's Peanut Butter Cups in Our Original Cheesecake with Layers of Delicious Fudge Cake and Caramel</i></p> <p>30th ANNIVERSARY CHOCOLATE CAKE CHEESECAKE <i>Layers of Our Original Cheesecake, Fudge Cake and Chocolate Truffle Cream</i></p> <p>WHITE CHOCOLATE RASPBERRY TRUFFLE* <i>Creamy Cheesecake Swirled with White Chocolate and Raspberry</i></p> <p>ULTIMATE RED VELVET CAKE CHEESECAKE" <i>Layers of Red Velvet Cake and Cheesecake Covered in Cream Cheese Frosting. Finished with White Chocolate</i></p> <p>GODIVA® CHOCOLATE CHEESECAKE <i>Flourless Godiva Chocolate Cake, Topped with Godiva Chocolate Cheesecake and Chocolate Mousse</i></p> <p>FRESH BANANA CREAM CHEESECAKE <i>Banana Cream Cheesecake Topped with Bavarian Cream and Fresh Sliced Bananas</i></p> <p>ADAM'S PEANUT BUTTER CUP FUDGE RIPPLE <i>Creamy Cheesecake Swirled with Caramel, Peanut Butter, Butterfingers® and Reese's Peanut Butter Cups</i></p> <p>WHITE CHOCOLATE CARAMEL MACADAMIA NUT CHEESECAKE <i>White Chocolate Cheesecake Swirled with Macadamia Nuts and Caramel on a Blond Brownie Crust</i></p> <p>LEMON RASPBERRY CREAM CHEESECAKE <i>Raspberry-Vanilla Cake, Creamy Lemon Cheesecake, Raspberry Lady Fingers and Lemon Mousse</i></p> <p>DULCE DE LECHE CARAMEL CHEESECAKE <i>Caramel Cheesecake Topped with Caramel Mousse on a Vanilla Crust</i></p> <p>CHOCOLATE COCONUT CREAM CHEESECAKE <i>Coconut Cheesecake Topped with Coconut Cream Custard, all on a Chocolate Macaroon Crust</i></p> <p>TIRAMISU CHEESECAKE <i>Our Wonderful Cheesecake and Tiramisu Combined into one Amazing Dessert!</i></p> <p>CHOCOLATE MOUSSE CHEESECAKE <i>Silky Chocolate Cheesecake Topped with a Layer of Belgian Chocolate Mousse</i></p> <p>VANILLA BEAN CHEESECAKE <i>Creamy Vanilla Bean Cheesecake, Topped with Vanilla Mousse and Whipped Cream</i></p> <p>CHOCOLATE TUXTEDO CREAM™ CHEESECAKE <i>Layers of Our Fudge Cake, Chocolate Cheesecake, Vanilla Mascarpone Mousse and Chocolate Ganache</i></p> <p>KAHLUA® COCOA COFFEE CHEESECAKE <i>A Rich Brownie, Topped with Kahlua Cheesecake, Creamy Chocolate Mousse and Chocolate Ganache</i></p> <p>PINEAPPLE UPSIDE-DOWN CHEESECAKE <i>Pineapple Cheesecake Between Two Layers of Moist Buttery Pineapple Upside-Down Cake</i></p> <p>CHOCOLATE RASPBERRY TRUFFLE* <i>Layers of Chocolate Cake, Chocolate-Raspberry Swirl Cheesecake, Chocolate Mousse and Chocolate Ganache</i></p> <p>DUTCH APPLE CARAMEL STREUSEL <i>Our Original Cheesecake, Baked Apples, Caramel and Brown Sugar Cinnamon Walnut Streusel</i></p> <p>CHOCOLATE CHIP COOKIE-DOUGH CHEESECAKE <i>Creamy Cheesecake Loaded with Our Chocolate Chip Cookie-Dough with Walnuts</i></p>
--	--

menu = "abstraction"

The Cheesecake Factory®

sandwiches

All Served with French Fries or Green Salad

Sweet Potato Fries 1.00 extra

RENEE'S SPECIAL 10.50

One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich,
a Cup of Our Soup and a Small Green Salad
with a Small Caesar Salad 1.00 extra

CHICKEN SALAD SANDWICH 11.50

Housemade with Roasted Almonds, Lettuce,
Tomato and Mayonnaise.
Served on Grilled Brioche Bread

THE CLUB 12.95

Freshly Roasted Turkey Breast, Bacon,
Lettuce, Tomato and
Mayonnaise on White Toast

GRILLED CHICKEN AND AVOCADO CLUB 12.95

Grilled Chicken Breast with Avocado, Bacon, Tomato, Melted Swiss and Herb Mayonnaise

THE NAVAJO 11.95

Warm Fry-Bread Stuffed with Grilled Chicken, Avocado, Lettuce,
Tomato, Red Onion and Mayonnaise

CUBAN SANDWICH 11.95

Slow Roasted Pork, Ham, Swiss Cheese, Pickles, Mustard and Mayonnaise
on a Grilled and Pressed Cuban-Style Roll

CHICKEN PARMESAN SANDWICH 11.95

Tender Chicken Lightly Coated in Parmesan Breadcrumbs, Roasted Peppers, Tomato Sauce
and Melted Cheese on a Freshly Grilled French Roll

BLACKENED CHICKEN SANDWICH 11.95

Grilled with Melted Fontina Cheese, Lettuce, Tomato, Red Onion and Spicy Mayonnaise

CALIFORNIA CHEESESTEAK 11.95

Thinly Sliced Grilled Steak Covered with Sauteed Mushrooms, Onions,
Peppers and Cheese on a Toasted Roll

GRILLED SHRIMP & BACON CLUB 13.95

Charbroiled Shrimp, Bacon, Lettuce and Tomato with Our Special Dressing

cheesecakes

ORIGINAL

The One that Started it All!

Our Famous Creamy Cheesecake with a Graham Cracker Crust and Sour Cream Topping

FRESH STRAWBERRY

The Original Topped with Glazed Fresh Strawberries. Our Most Popular Flavor for 30 Years!

REESE'S® PEANUT BUTTER CHOCOLATE CAKE CHEESECAKE™

Reese's Peanut Butter Cups in Our Original Cheesecake with Layers of Delicious Fudge Cake and Caramel

30th ANNIVERSARY CHOCOLATE CAKE CHEESECAKE

Layers of Our Original Cheesecake, Fudge Cake and Chocolate Truffle Cream

WHITE CHOCOLATE RASPBERRY TRUFFLE®

Creamy Cheesecake Swirled with White Chocolate and Raspberry

ULTIMATE RED VELVET CAKE CHEESECAKE™

Layers of Red Velvet Cake and Cheesecake Covered in Cream Cheese Frosting. Finished with White Chocolate

GODIVA® CHOCOLATE CHEESECAKE

Flourless Godiva Chocolate Cake, Topped with Godiva Chocolate Cheesecake and Chocolate Mousse

FRESH BANANA CREAM CHEESECAKE

Banana Cream Cheesecake Topped with Bavarian Cream and Fresh Sliced Bananas

ADAM'S PEANUT BUTTER CUP FUDGE RIPPLE

Creamy Cheesecake Swirled with Caramel, Peanut Butter, Butterfingers® and Reese's Peanut Butter Cups™

WHITE CHOCOLATE CARAMEL MACADAMIA NUT CHEESECAKE

White Chocolate Chunk Cheesecake Swirled with Macadamia Nuts and Caramel on a Blond Brownie Crust

LEMON RASPBERRY CREAM CHEESECAKE

Raspberry-Vanilla Cake, Creamy Lemon Cheesecake, Raspberry Lady Fingers and Lemon Mousse

DULCE DE LECHE CARAMEL CHEESECAKE

Caramel Cheesecake Topped with Caramel Mousse on a Vanilla Crust

CHOCOLATE COCONUT CREAM CHEESECAKE

Coconut Cheesecake Topped with Coconut Cream Custard, all on a Chocolate Macaroon Crust

TIRAMISU CHEESECAKE

Our Wonderful Cheesecake and Tiramisu Combined into one Amazing Dessert!

CHOCOLATE MOUSSE CHEESECAKE

Silky Chocolate Cheesecake Topped with a Layer of Belgian Chocolate Mousse

VANILLA BEAN CHEESECAKE

Creamy Vanilla Bean Cheesecake, Topped with Vanilla Mousse and Whipped Cream

CHOCOLATE TUXEDO CREAM™ CHEESECAKE

Layers of Our Fudge Cake, Chocolate Cheesecake, Vanilla Mascarpone Mousse and Chocolate

KAHLUÀ® COCOA COFFEE CHEESECAKE

A Rich Brownie, Topped with Kahlua Cheesecake, Creamy Chocolate Mousse and Chocolate Ganache



RENEE'S SPECIAL 10.50

*One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich,
a Cup of Our Soup and a Small Green Salad
with a Small Caesar Salad 1.00 extra*

item = "service"

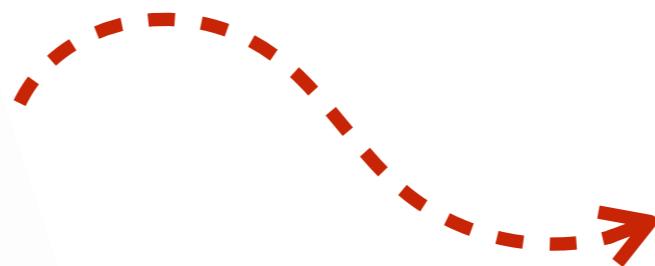
Customer requests "Renée's Special"



The Cheesecake Factory®



service request

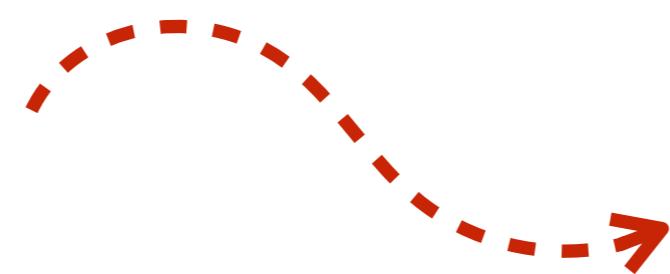


kitchen

The Cheesecake Factory®



kitchen



service fulfilled



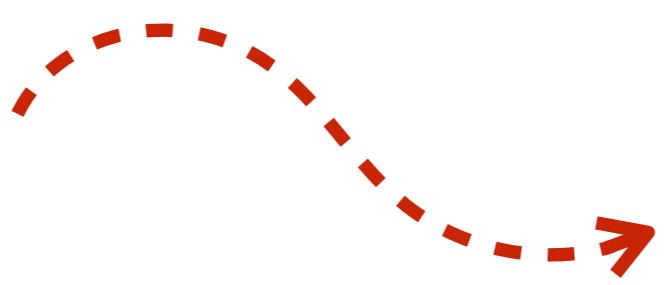
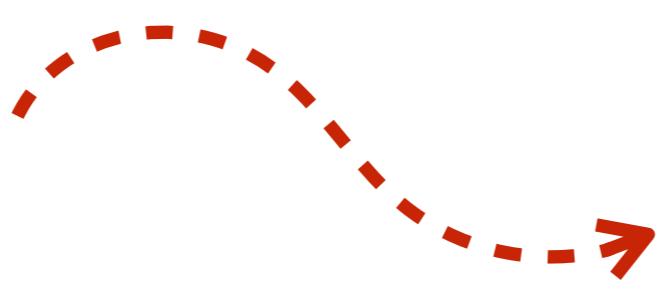
The Cheesecake Factory®



kitchen

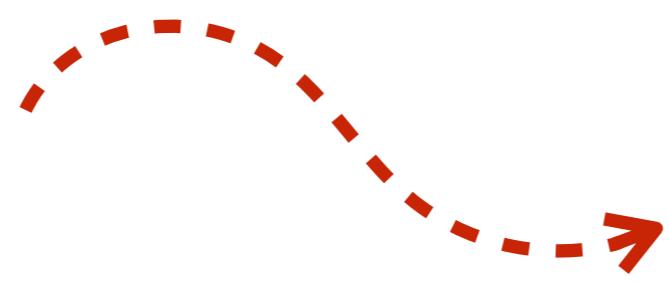


kitchen



The PMI Cheesecake Factory®

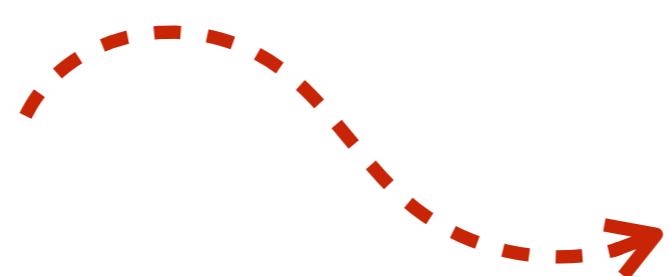
Tensor Flow (Deep Learning) model for acute myocardial infarction



58% Probability of AMI in 5 years

service fulfilled

"kitchen"



kitchen

A meal ~ An Application



Duke Translational Bioinformatics | Duke Surgery | Duke Statistical Sciences
Duke Institute for Health Innovations



Calypso API

Samples

Get Patient

<http://54.186.43.170/api/patients/1>

Predict Patient

<http://54.186.43.170/api/analysis/predict/1>

sorted targets

<http://54.186.43.170/api/targets/patient/1>

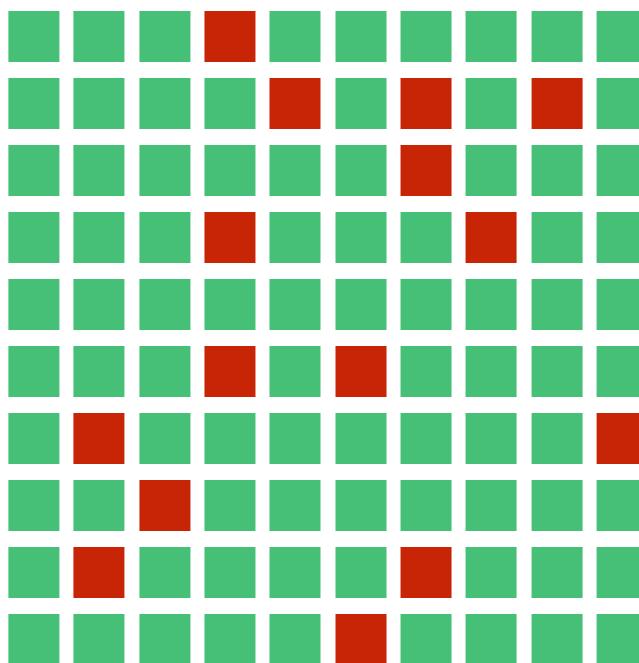
orders of a target

<http://54.186.43.170/api/orders/target/27>

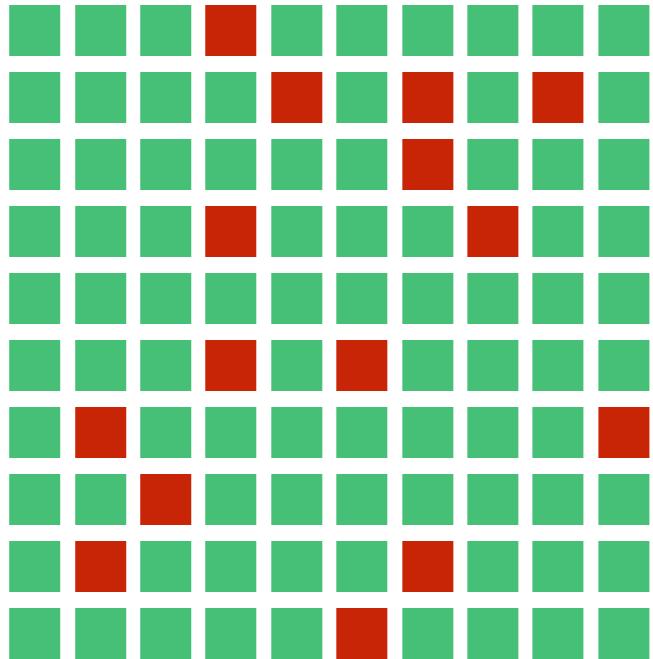
Table of contents

- [TODO](#)
- [Setup](#)
- [Grunt](#)
- [API](#)
 - [Patient](#)
 - [/patients](#)
 - [/patients/:patient_id](#)
 - [Analysis](#)
 - [/analysis/:patient_id](#)
 - [/analysis/histogram/:patient_id](#)
 - [/analysis/percentile/:patient_id](#)
 - [/analysis/predict/:patient_id](#)

surgical complications

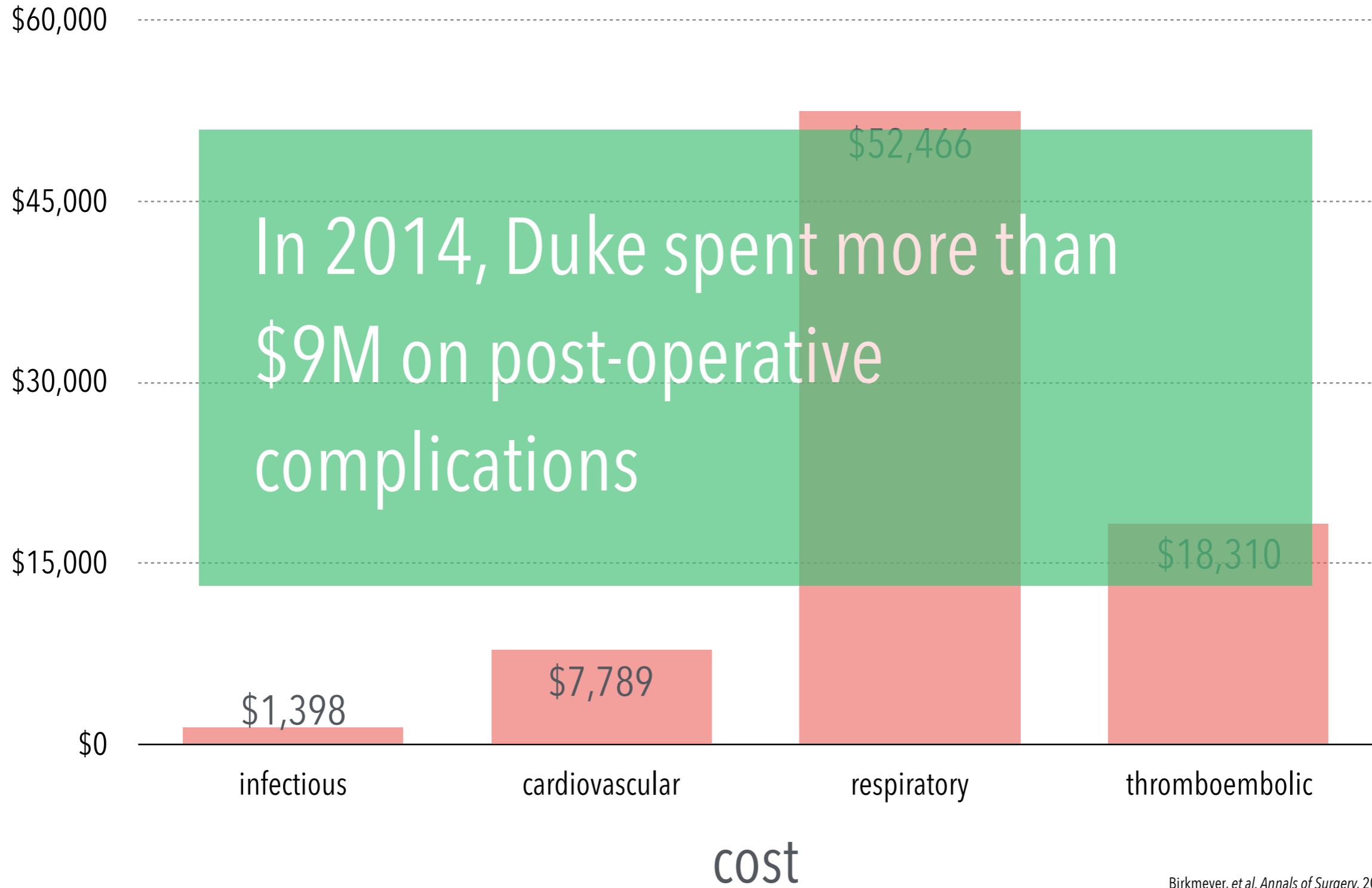


surgical complications



approximately 15 out of
every 100 surgical
procedures performed
results in a complication

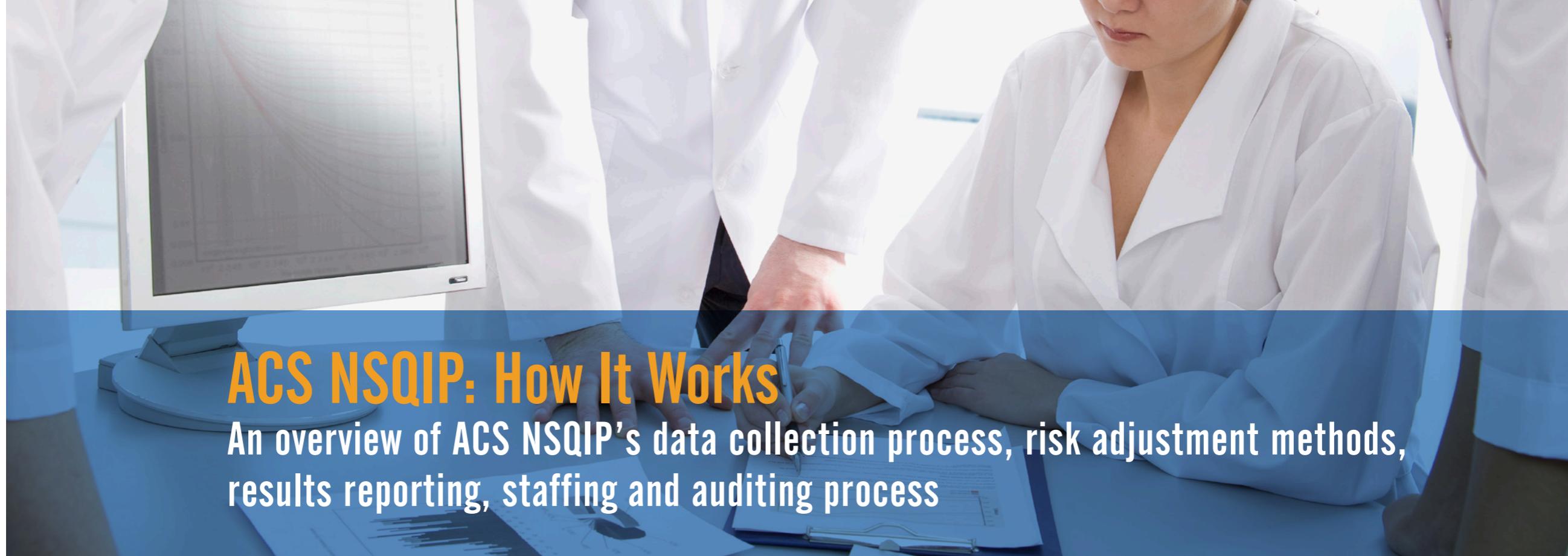
surgical complications





American College of Surgeons
National Surgical
Quality Improvement Program
(ACS NSQIP®)





ACS NSQIP: How It Works

An overview of ACS NSQIP's data collection process, risk adjustment methods, results reporting, staffing and auditing process

The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP®) was developed by surgeons a decade ago to help hospitals measurably improve patient outcomes and save lives. Today, ACS NSQIP remains **the first and only nationally validated, risk-adjusted, outcomes-based program to measure and improve the quality of surgical care across surgical specialties in the private sector.**

The program dates back to the mid-1980s, when the Department of Veterans Affairs (VA) developed NSQIP to help its 133 hospitals measure quality of care based on preoperative risk factors and postoperative outcomes. VA hospitals found great success with the program. Hospitals were able to decrease postoperative mortality rates by 47 percent and morbidity rates by 43 percent between 1991 and 2006.¹ Additionally, VA hospitals saw median length of stay fall from nine to four days, and patient satisfaction improved.

In 2001, ACS launched a pilot program funded by the Agency for Healthcare Research and Quality (AHRQ) to show that NSQIP was also effective in private-sector hospitals. Based on the successful pilot, in 2004 ACS began enrolling new private sector hospitals into NSQIP.

Research

Original Investigation

Association of Hospital Participation in a Surgical Outcomes Monitoring Program With Inpatient Complications and Mortality

David A. Etzioni, MD, MSHS; Nabil Wasif, MD, MPH; Amylou C. Dueck, PhD; Robert R. Cima, MD;
Samuel F. Hohmann, PhD; James M. Naessens, ScD; Amit K. Mathur, MD, MS; Elizabeth B. Habermann, PhD, MPH

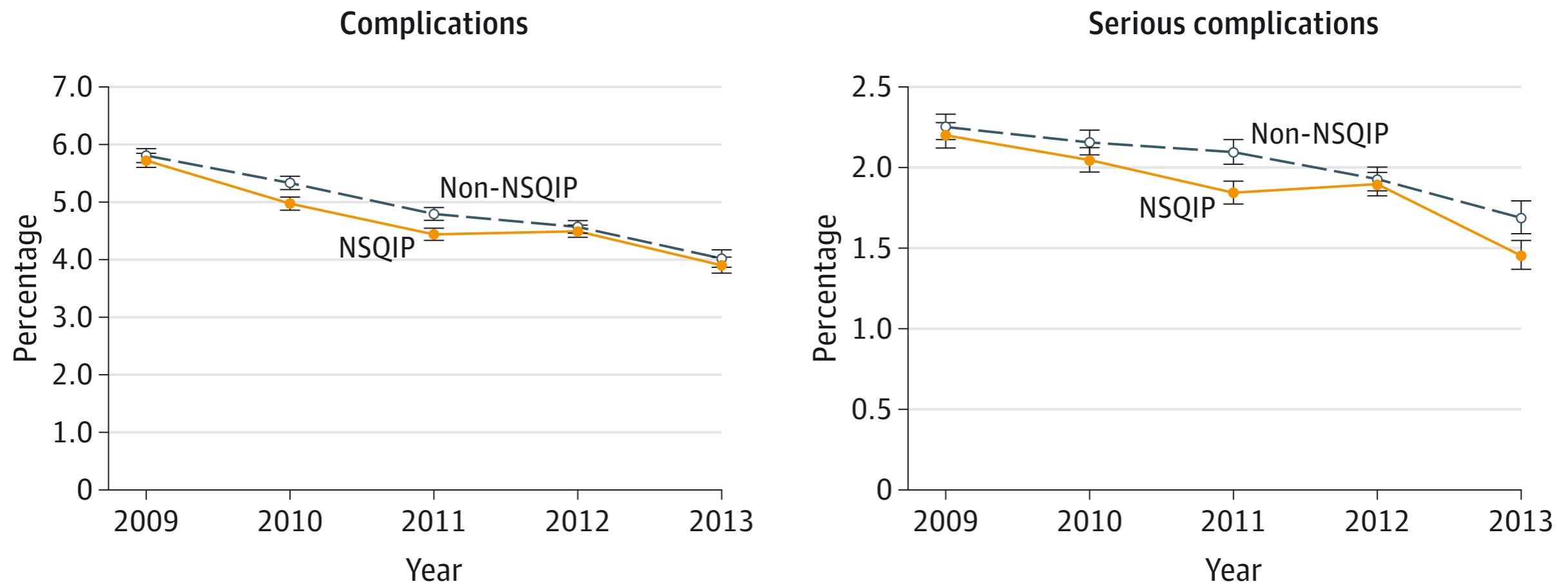
Research

Original Investigation

Association of Hospital Participation in a Quality Reporting Program With Surgical Outcomes and Expenditures for Medicare Beneficiaries

Nicholas H. Osborne, MD, MS; Lauren H. Nicholas, PhD; Andrew M. Ryan, PhD;
Jyothi R. Thumma, MPH; Justin B. Dimick, MD, MPH

Figure 2. Adjusted Rates of Complications, Serious Complications, and Mortality by Hospital NSQIP Participation and Year



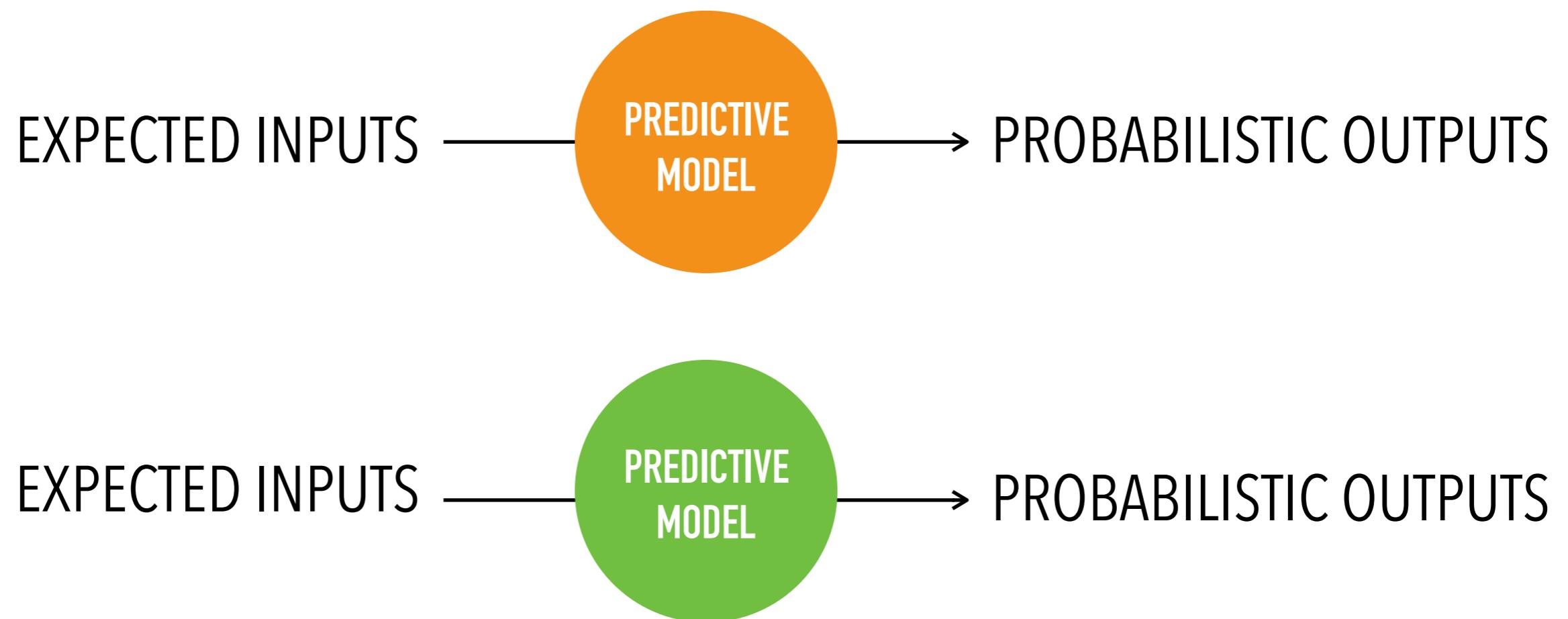
NSQIP, National Surgical Quality Improvement Program. Error bars indicate 95% CIs. Adjusted for patient comorbidity, operation type, age, and sex.



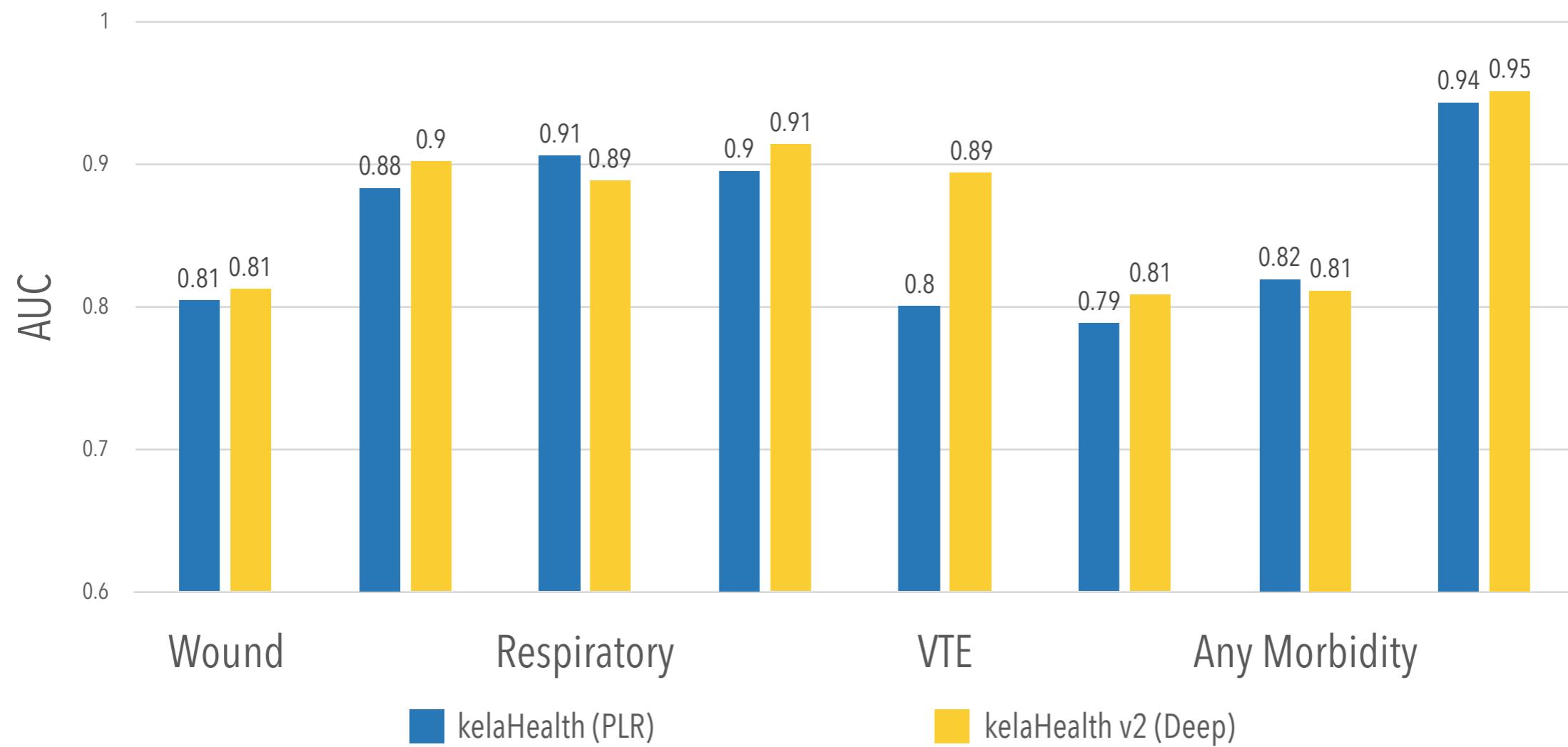
**“Weighing a
pig does not
make the pig fatter”**



cognition



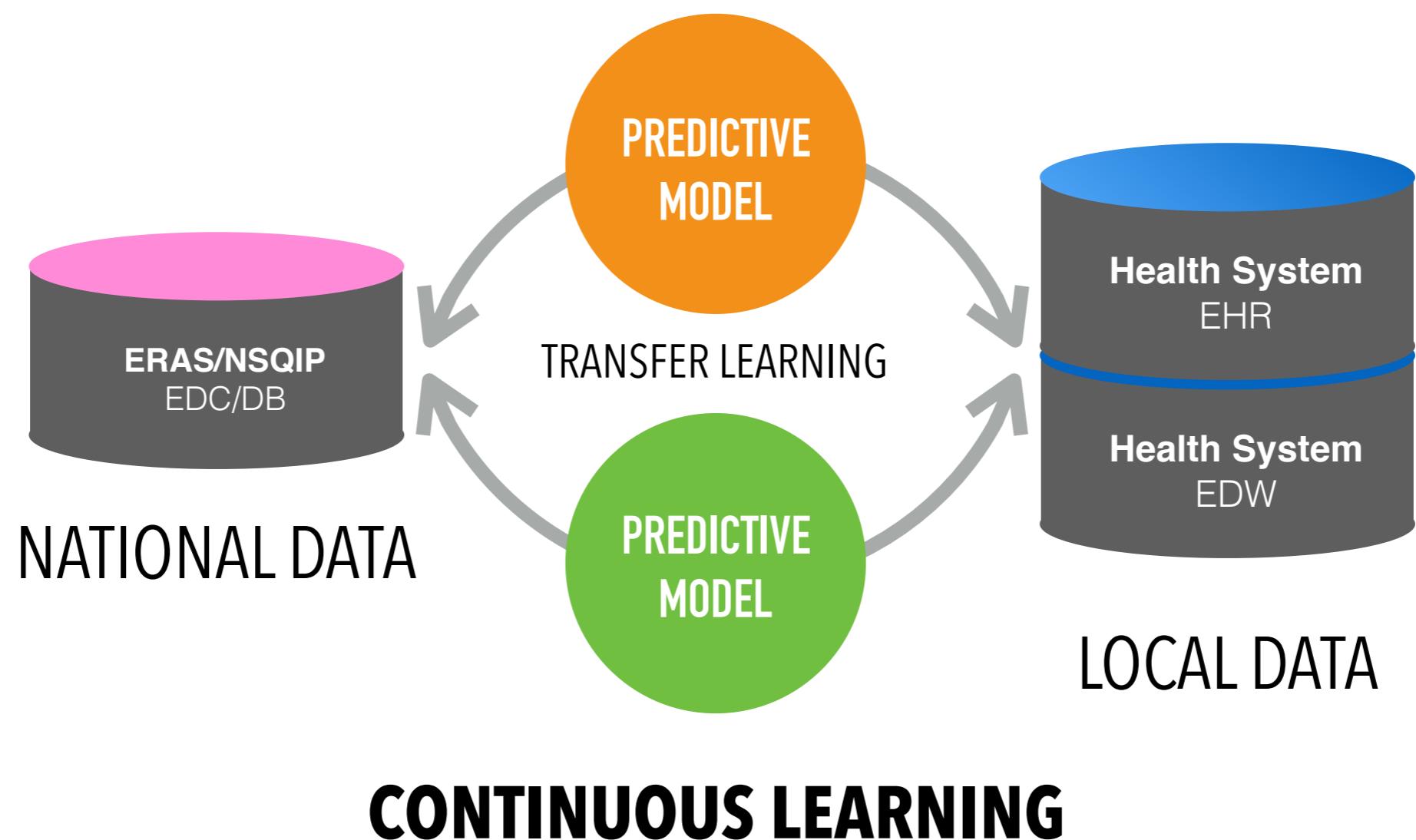
Model Performance





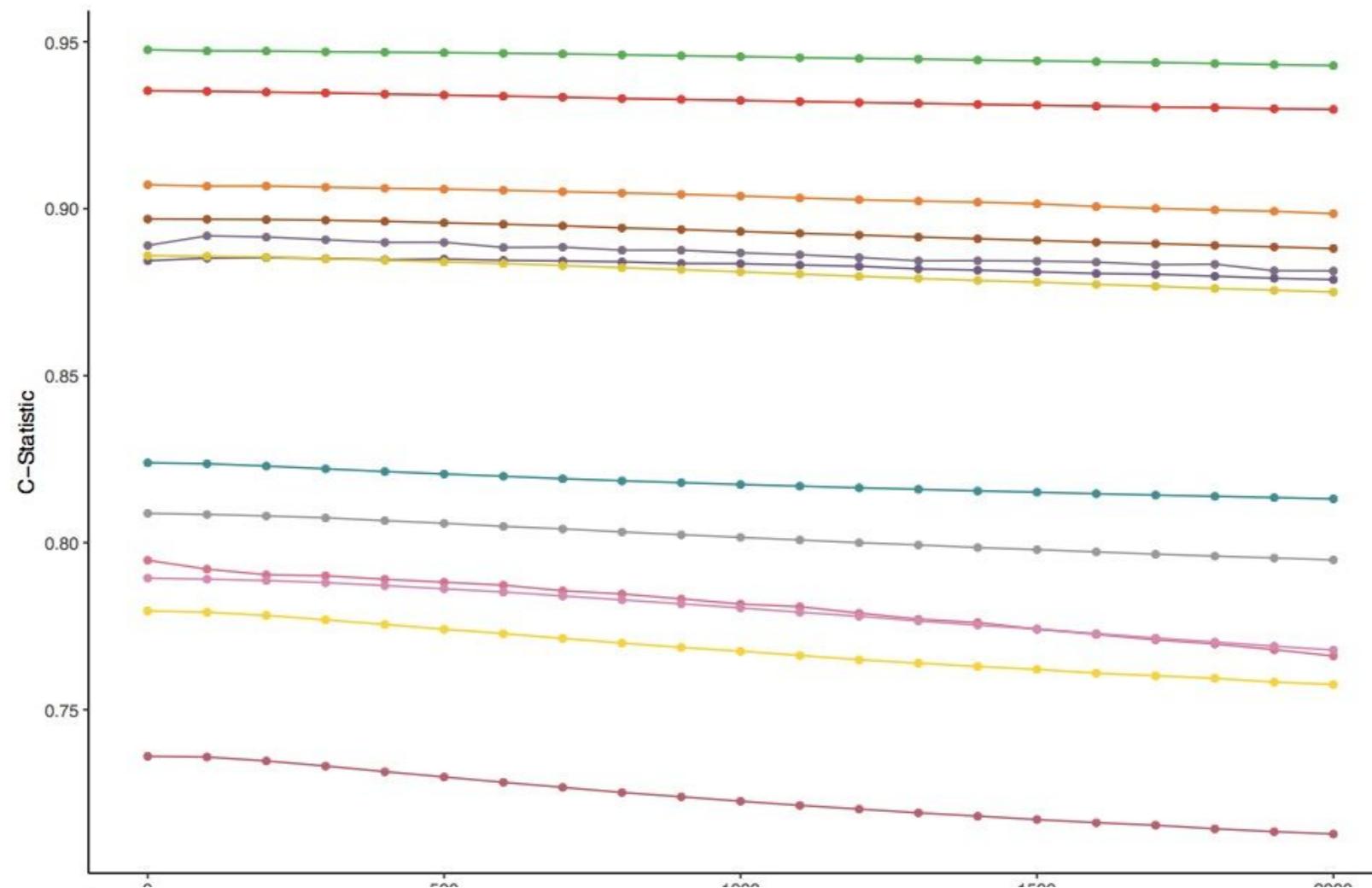


cognition



Transfer Learning

Random Subcohorts



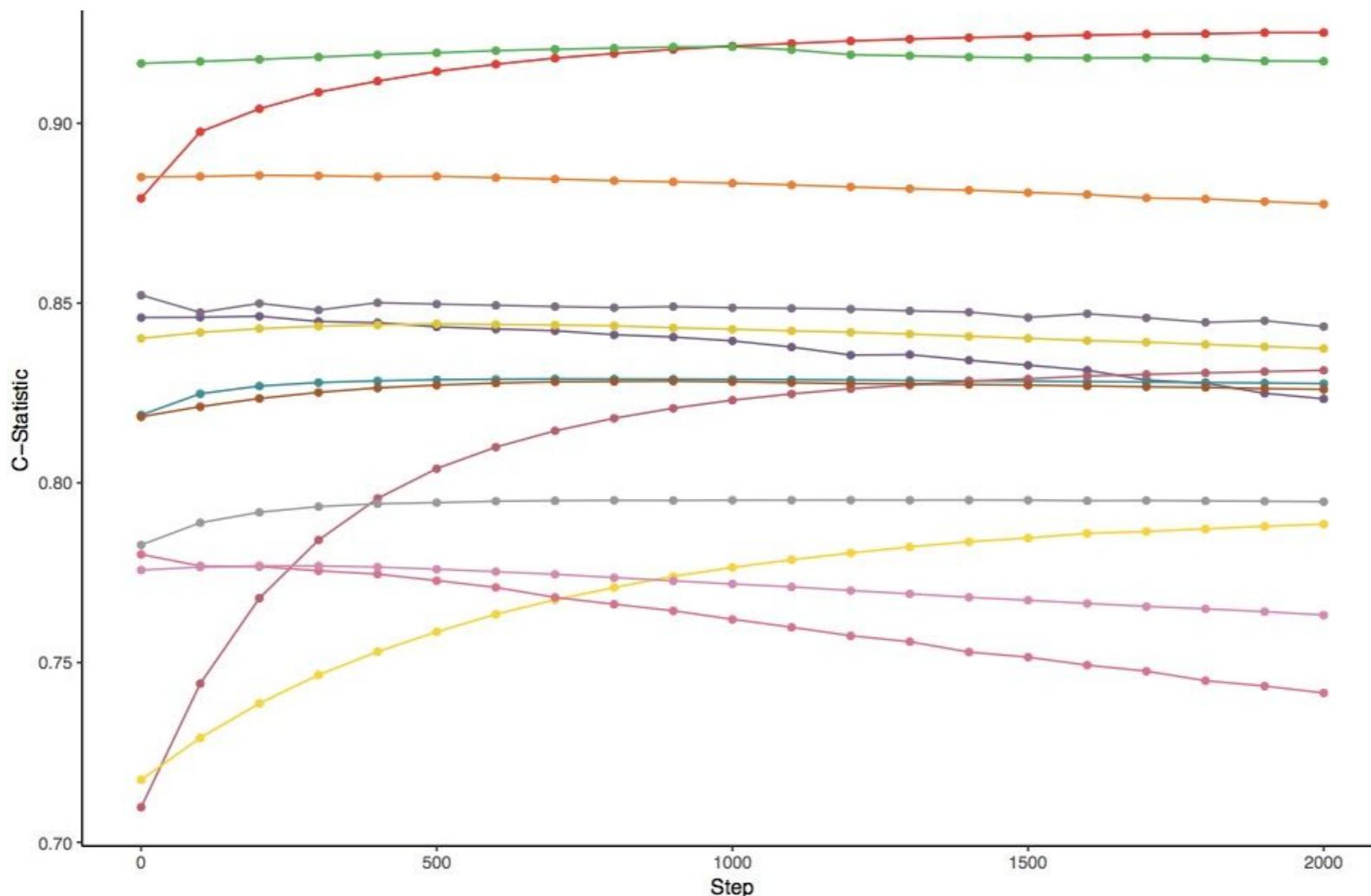
tuning interation

Models

- Wound complication
- Cardiac complication
- Respiratory complication
- Thrombotic complication
- Renal complication
- Neurologic complication
- Urinary tract infection
- Bleeding complication
- Septic complication
- Reoperation
- Readmission
- Mortality
- Any morbidity

Transfer Learning

Local Subcohort

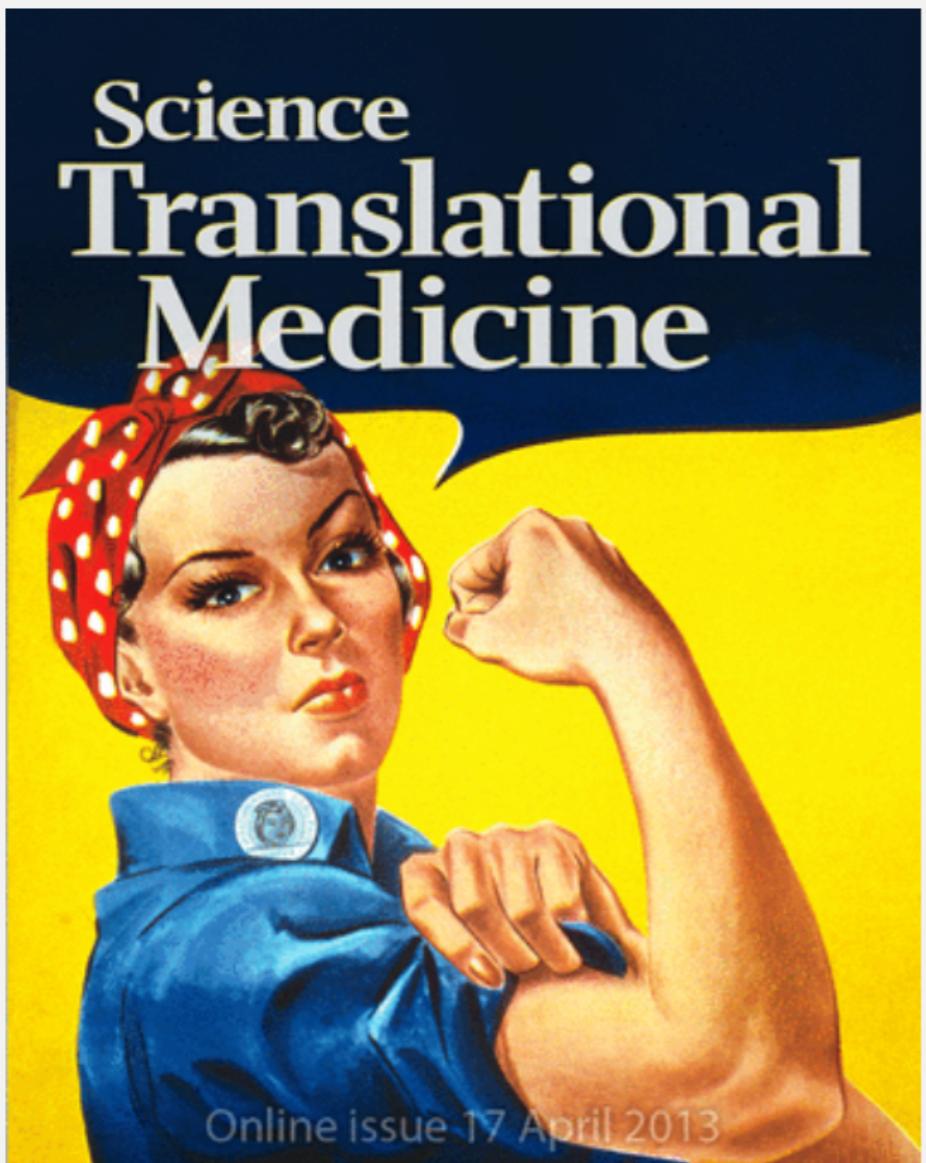


tuning interation

Models	Wound complication	Thrombotic complication	Urinary tract infection	Reoperation	Mortality
	Cardiac complication	Renal complication	Bleeding complication	Readmission	
	Respiratory complication	Neurologic complication	Septic complication	Any morbidity	



About The Cover



ONLINE COVER "Doin' It for Themselves"—and Future Generations. A U.S. cultural icon, Rosie the Riveter represents millions of American women who entered the workforce during World War II. Rosies worked in factories of all kinds, including some that manufactured munitions and other supplies needed for the war effort. In the realm of biomedical research, women from around the world are now supplying

REPORT

CROWDSOURCING

Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer

Adam A. Margolin,^{1,*†} Erhan Bilal,^{2†} Erich Huang,^{1,3,4†} Thea C. Norman,¹ Lars Ottestad,⁵ Brigham H. Mecham,^{1,6} Ben Sauerwine,⁷ Michael R. Kellen,¹ Lara M. Mangravite,¹ Matthew D. Furia,^{1,8} Hans Kristian Moen Volland,^{5,9,10,11} Oscar M. Rueda,¹¹ Justin Guinney,¹ Nicole A. Deflaux,¹ Bruce Hoff,¹ Xavier Schildwachter,¹ Hege G. Russnes,^{9,10,12} Daehoon Park,¹³ Veronica O. Vang,^{9,10} Tyler Pirtle,⁷ Lamia Youseff,⁷ Craig Citro,⁷ Christina Curtis,¹⁴ Vessela N. Kristensen,^{9,10,15} Joseph Hellerstein,⁷ Stephen H. Friend,^{1*} Gustavo Stolovitzky,² Samuel Aparicio,^{16,17,18†} Carlos Caldas,^{11,19,20†} Anne-Lise Børresen-Dale^{9,10†}

Although molecular prognostics in breast cancer are among the most successful examples of translating genomic analysis to clinical applications, optimal approaches to breast cancer clinical risk prediction remain controversial. The Sage Bionetworks–DREAM Breast Cancer Prognosis Challenge (BCC) is a crowdsourced research study for breast cancer prognostic modeling using genome-scale data. The BCC provided a community of data analysts with a common platform for data access and blinded evaluation of model accuracy in predicting breast cancer survival on the basis of gene expression data, copy number data, and clinical covariates. This approach offered the opportunity to assess whether a crowdsourced community Challenge would generate models of breast cancer prognosis commensurate with or exceeding current best-in-class approaches. The BCC comprised multiple rounds of blinded evaluations on held-out portions of data on 1981 patients, resulting in more than 1400 models submitted as open source code. Participants then retrained their models on the full data set of 1981 samples and submitted up to five models for validation in a newly generated data set of 184 breast cancer patients. Analysis of the BCC results suggests that the best-performing modeling strategy outperformed previously reported methods in blinded evaluations; model performance was consistent across several independent evaluations; and aggregating community-developed models achieved performance on par with the best-performing individual models.

INTRODUCTION

Breast cancer is the leading female malignancy in the world (1) and is one of the first malignancies for which molecular biomarkers have exhibited promise for clinical decision making (2–5). Biomarkers can be used to divide the disease into predictive (the likelihood that a patient

responds to a particular therapy) or prognostic (a patient's risk for a defined clinical endpoint independent of treatment) subcategories. Such molecular subcategorization highlights the possibilities for precision medicine (6), in which biomarkers are leveraged to identify disease taxonomies that distinguish biologically relevant groupings beyond standard clinical measures and can potentially inform treatment strategies.

A decade after early achievements in the development of prognostic molecular classifiers (called signatures) of breast cancer based solely on gene expression analysis (4, 5, 7), a large number of signatures proposed as markers of clinical risk prediction either fail to surpass the performance of conventional clinical covariates or await meaningful prospective validation [for example, the MINDACT (8, 9) and TAILORx/RxPONDER (10) trials]. Although gene expression-based breast cancer prognostic tests have been successfully implemented in routine clinical use, the application of molecular data to guide clinical decision making remains controversial (11).

Slow progress to evolve useful molecular classifiers may relate to poor study design, inconsistent findings, or improper validation studies (12). Data and code that underlie a potential new disease classifier are often unavailable for diligence. In addition, the rigor and objectivity of assessing molecular models are confounded by the tendency of data generation, data analysis, and model validation to be combined within the same study. This leads to the “self-assessment trap,” in which the desire to demonstrate improved performance of a researcher's own methodology may cause inadvertent bias in elements of study design, such as data set selection, parameter tuning, or evaluation criteria (13).

¹Sage Bionetworks, 1100 Fairview Avenue North, MS: M1-C108, Seattle, WA 98109, USA.

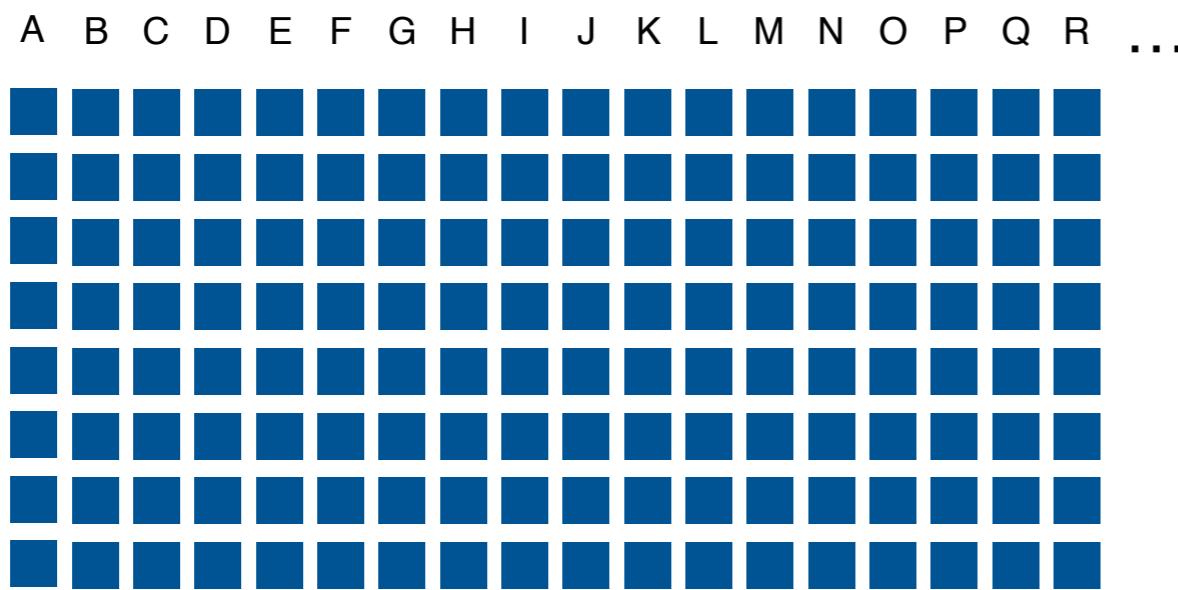
²Functional Genomics and Systems Biology, IBM Computational Biology Center, P. O. Box 218, Yorktown Heights, NY 10598, USA. ³Institute for Genome Sciences & Policy, Duke University, Durham, NC 27708, USA. ⁴Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA. ⁵Department of Oncology, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, 0450 Oslo, Norway. ⁶Trialmatics, LLC, Seattle, WA 98103, USA. ⁷Google Inc., 651 North 34th Street, Seattle, WA 98103, USA. ⁸Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA. ⁹Department of Genetics, Institute for Cancer Research, Oslo University Hospital, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway. ¹⁰KG. Jebsen Centre for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, 0313 Oslo, Norway.

¹¹Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ¹²Department of Pathology, Oslo University Hospital, 0450 Oslo, Norway. ¹³Department of Pathology, Drammen Hospital, Vestre Viken HF, 3004 Drammen, Norway. ¹⁴Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. ¹⁵Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, 1478 Åhus, Norway. ¹⁶Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. ¹⁷Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ¹⁸Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. ¹⁹Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. ²⁰Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.

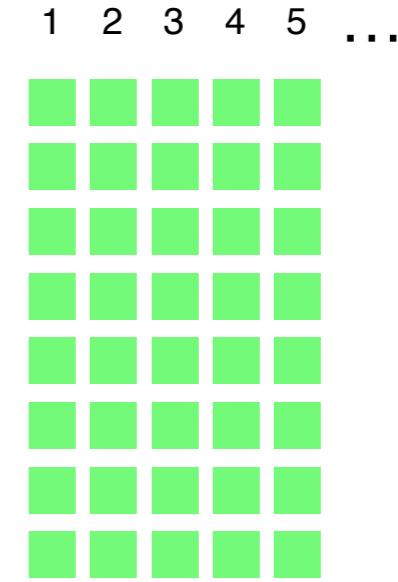
*Corresponding author. E-mail: margolin@sagebase.org (A.A.M.); friend@sagebase.org (S.H.F.)

†These authors contributed equally to this work.

PROGNOSIS CHALLENGE BREAST CANCER COHORTS

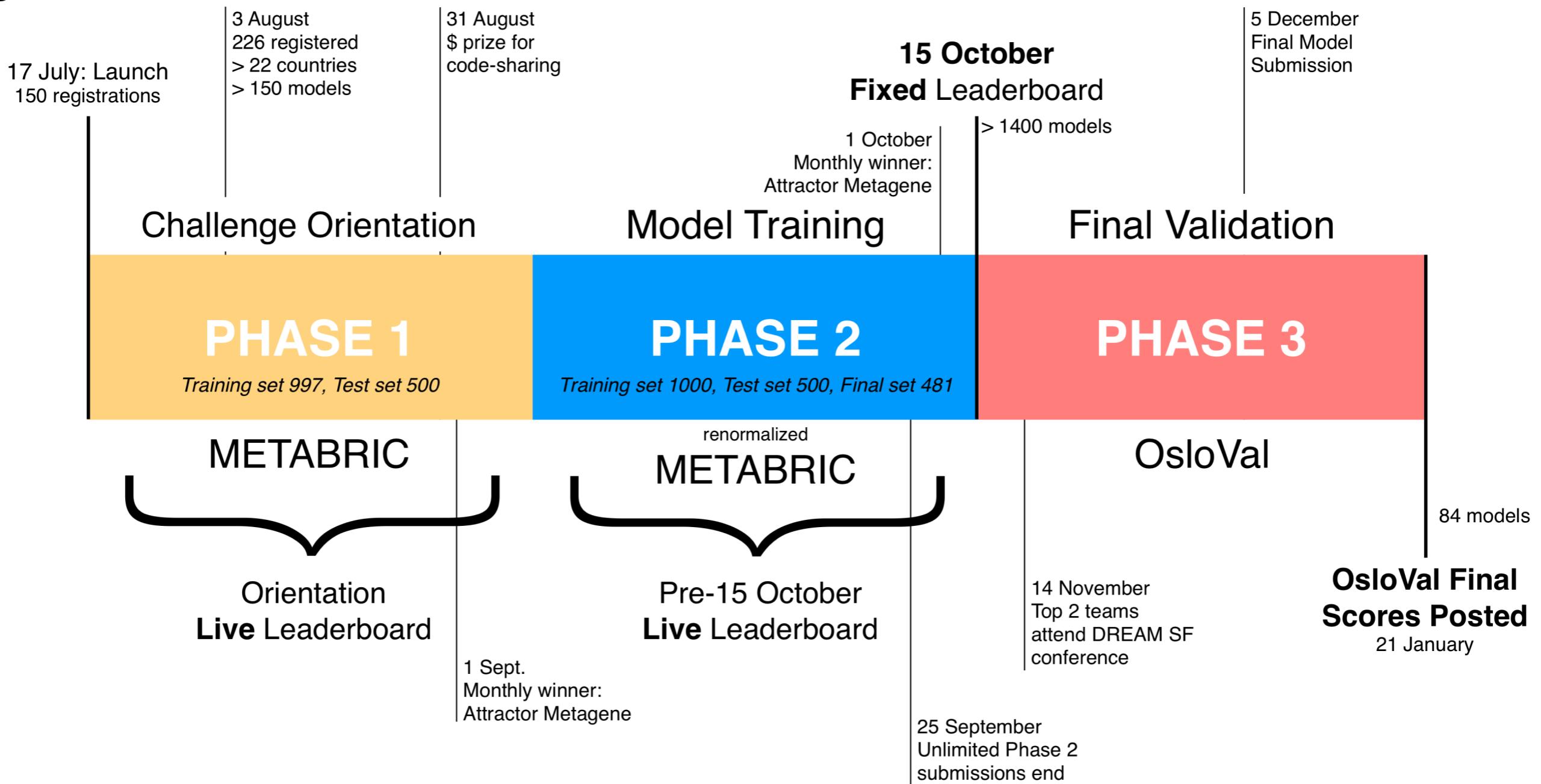


"METABRIC DATASET"
2000 PATIENTS:
expression, copy #, and clinical data

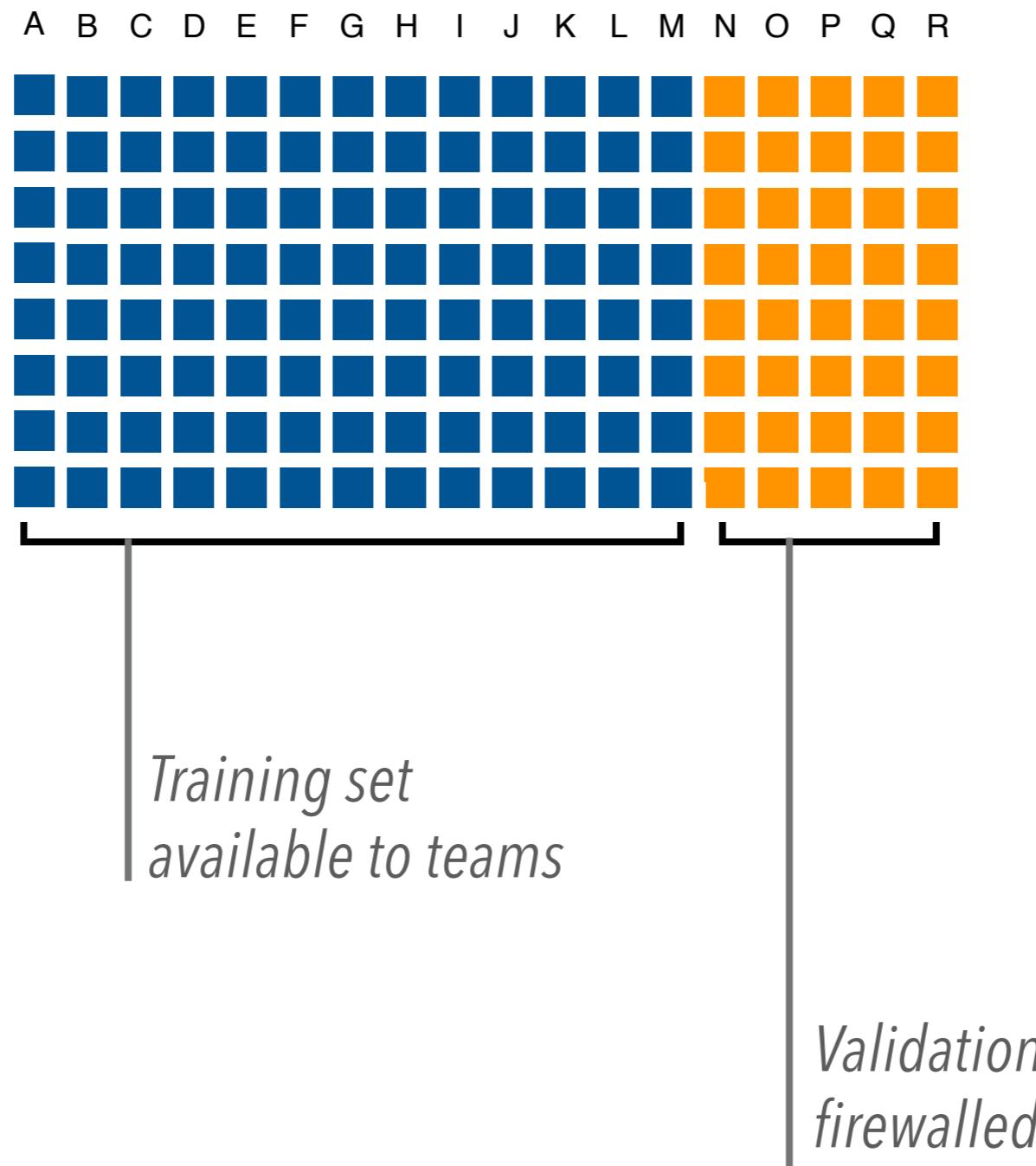


"OSLOVAL DATASET"
200 PATIENTS:
expression, copy #, and clinical data

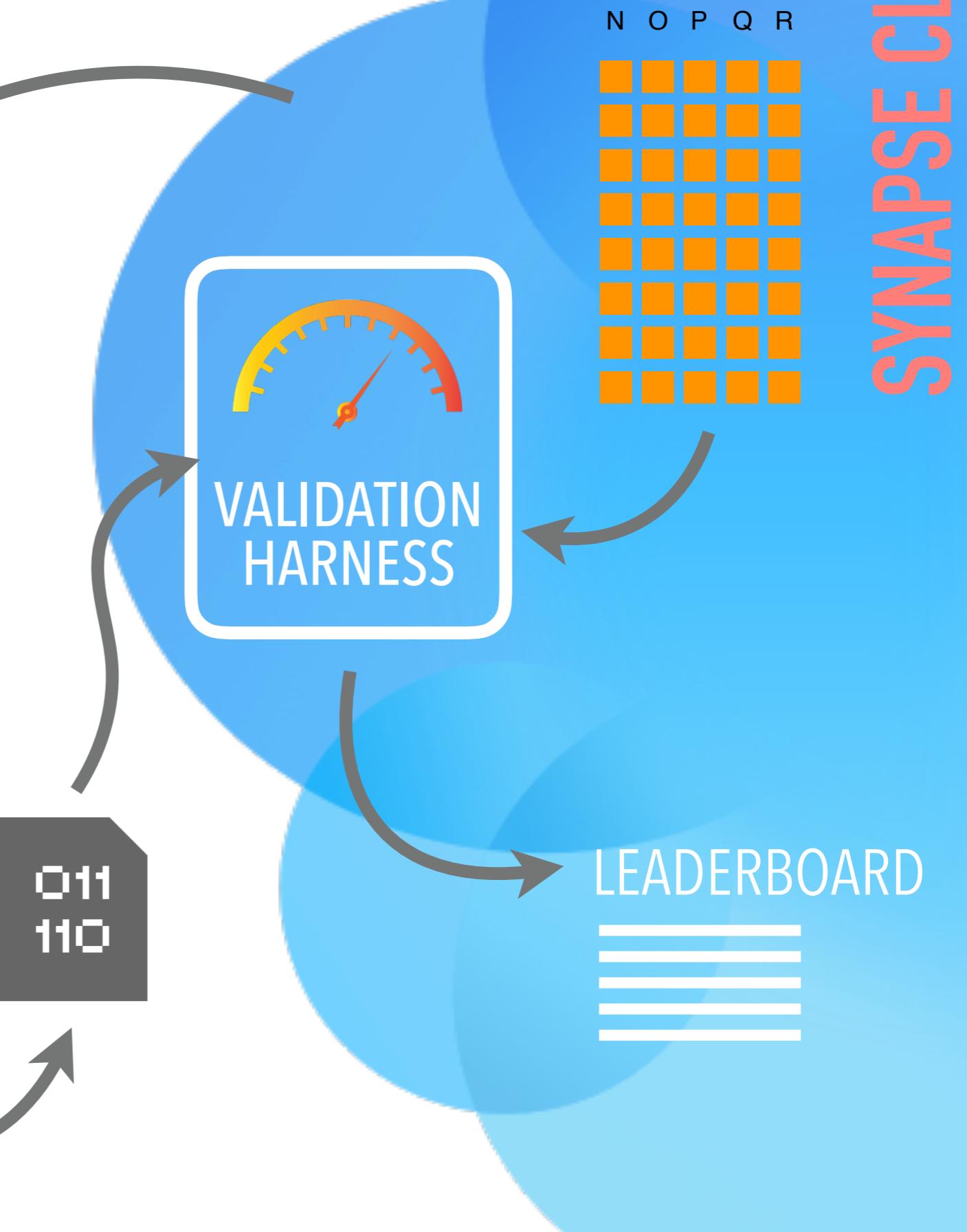
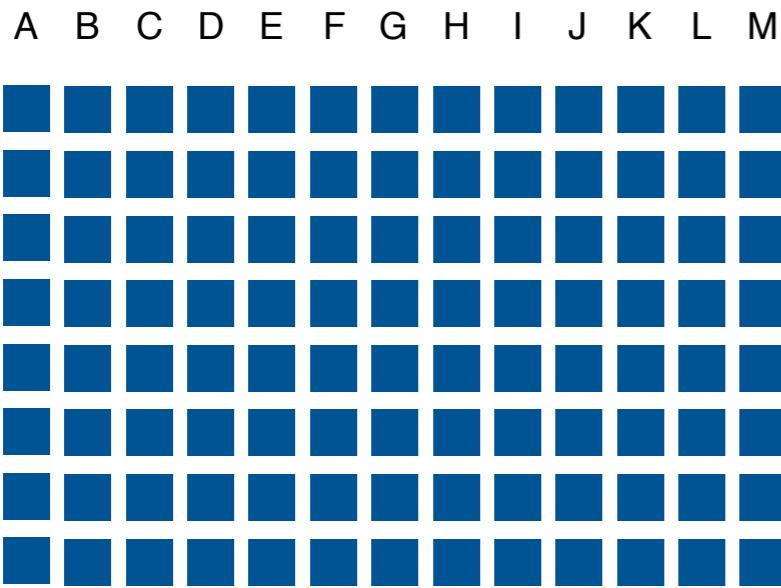
Figure 1



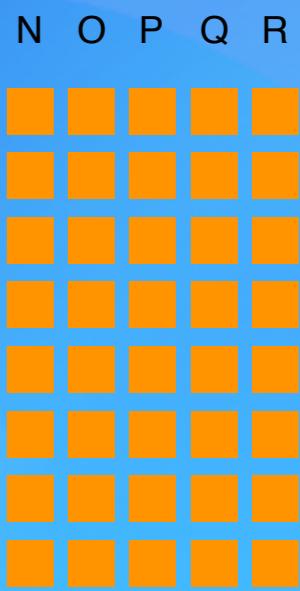
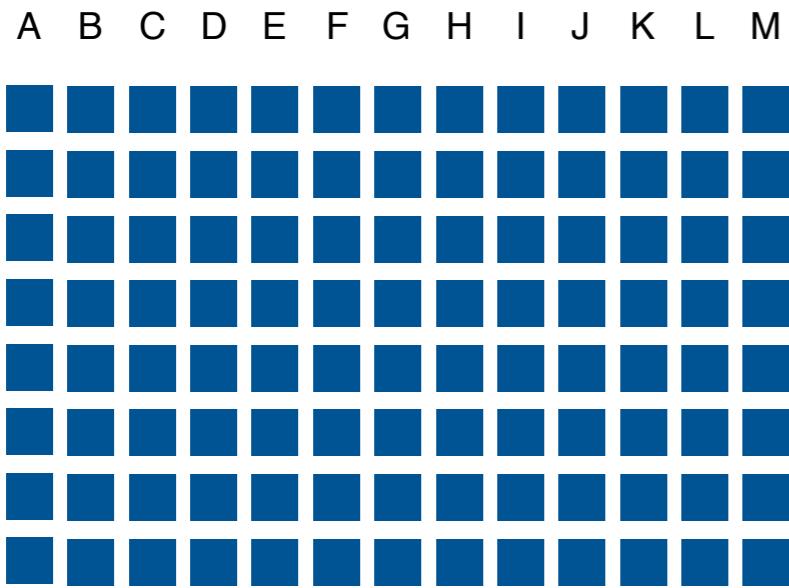
“METABRIC DATASET”



"METABRIC DATASET"



"METABRIC DATASET"



*Multiple Tranches/
Rounds of
Test & Validation*

VALIDATION
HARNESS

PREDICTIVE
MODEL

011
110

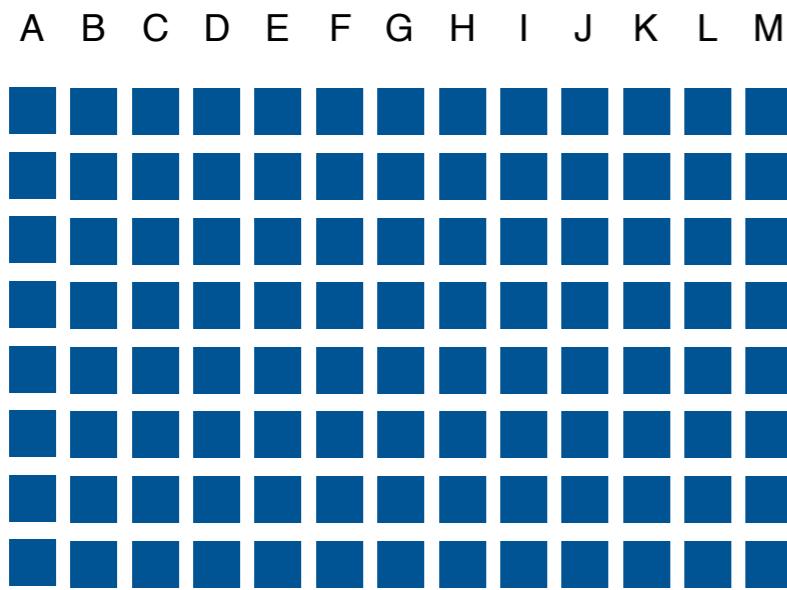
Google

LEADERBOARD

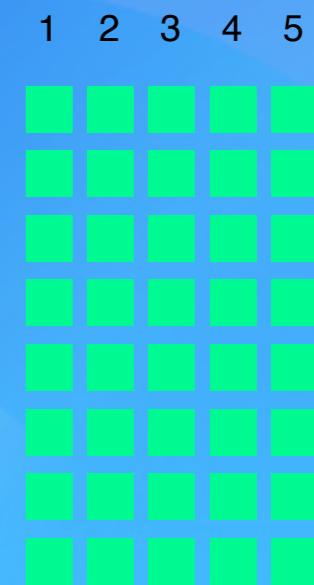


SYNAPSE CLOUD

"METABRIC DATASET"



"OSLOVAL"



SYNAPSE CLOUD

FINAL ROUND

PREDICTIVE
MODEL

011
110

Google

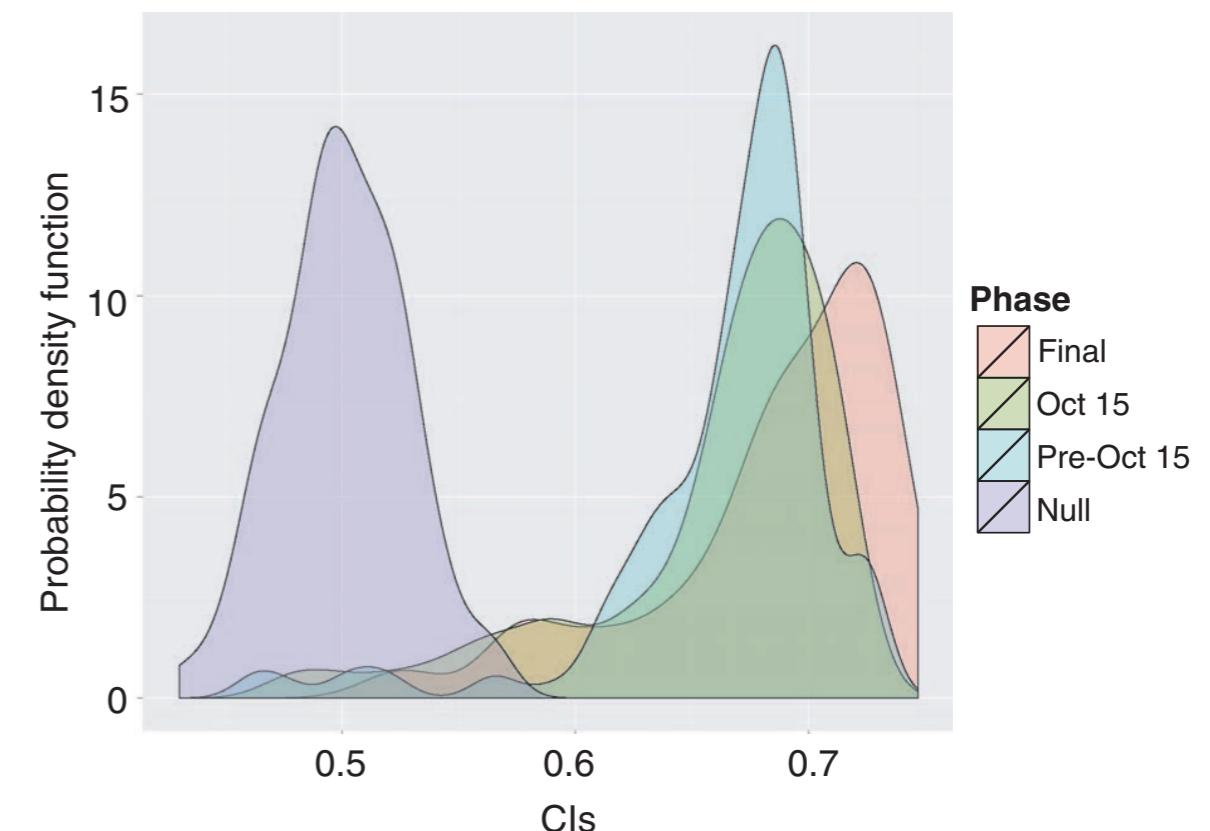
VALIDATION
HARNESS

LEADERBOARD

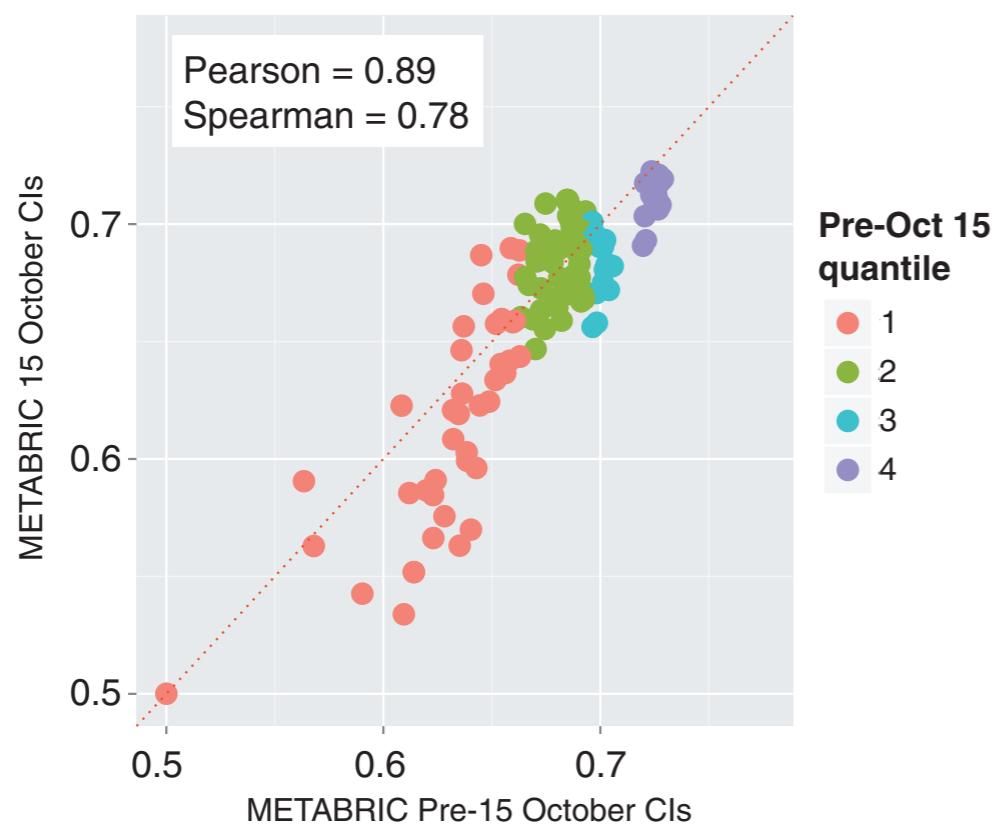
(A) Evolution of model CIs (pre-15 October)



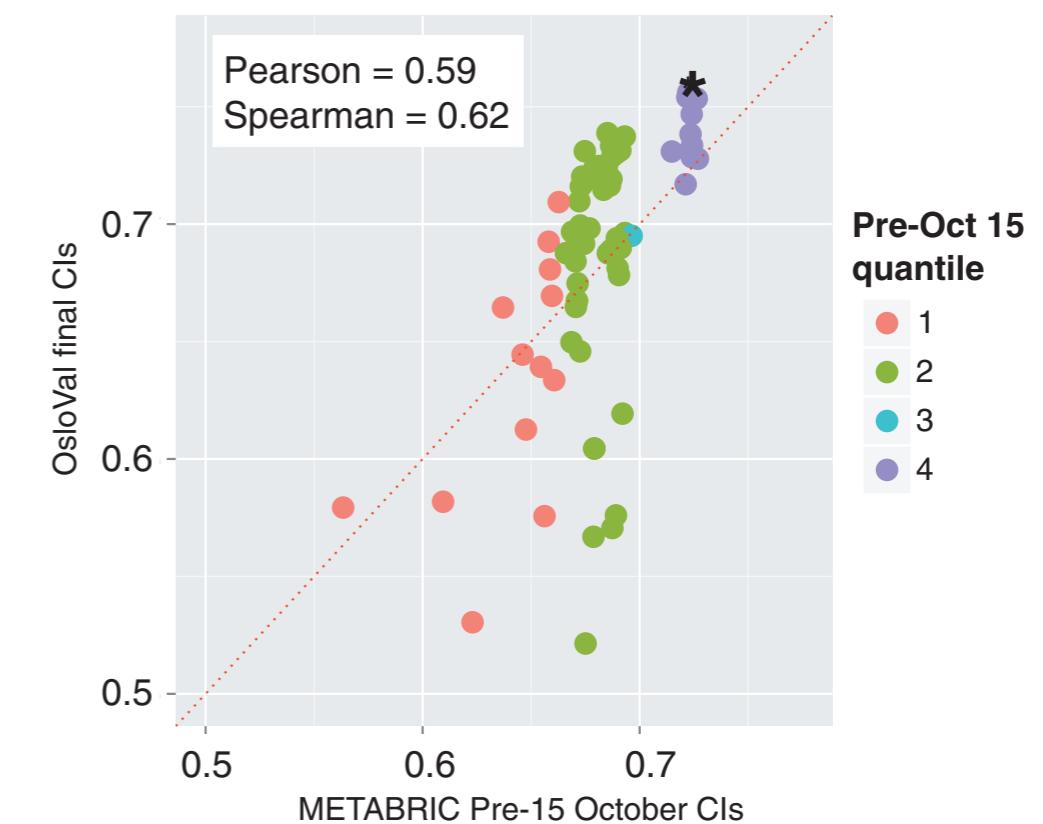
(B) CIs by challenge phase



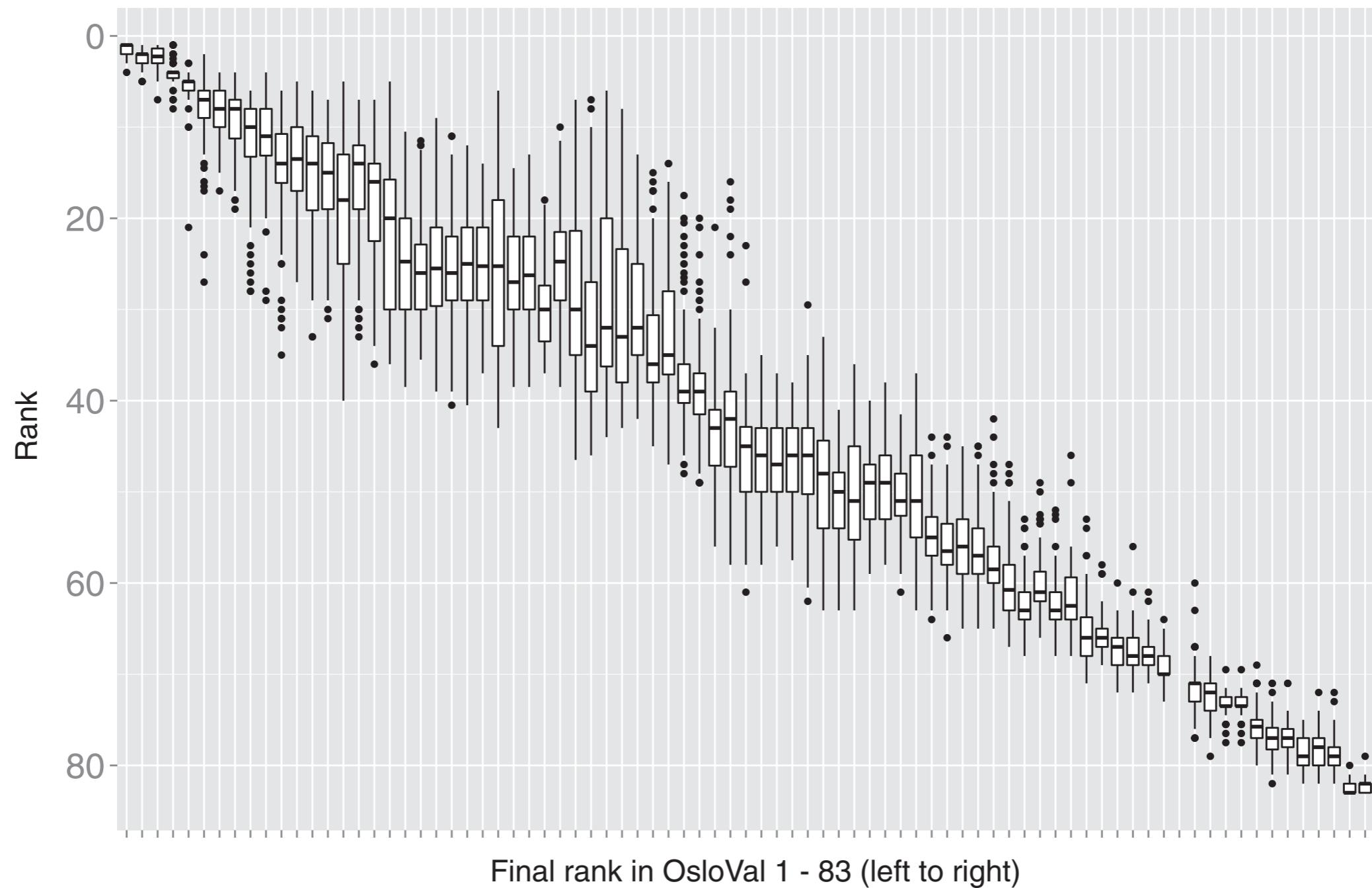
(C) Pre-15 October versus 15 October CIs

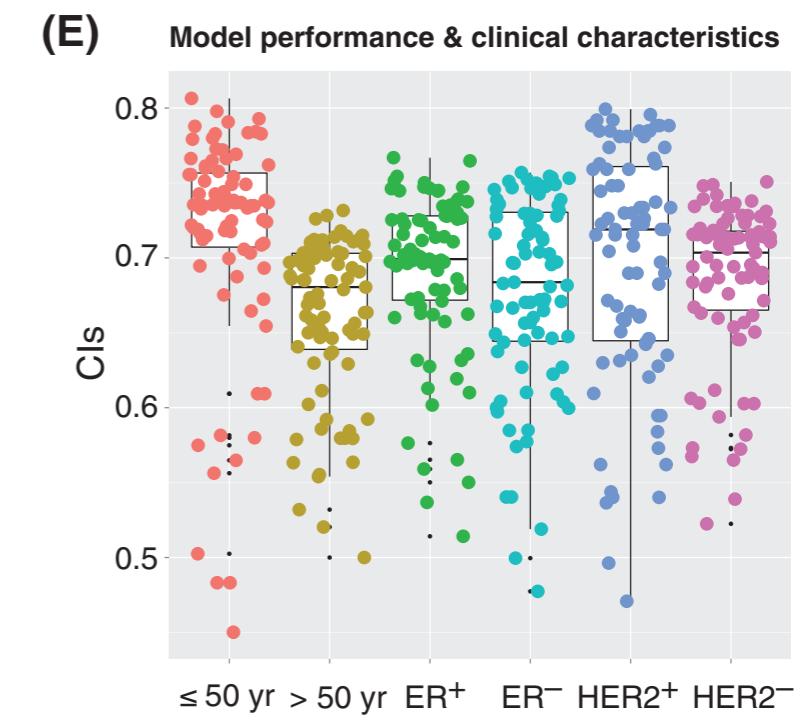
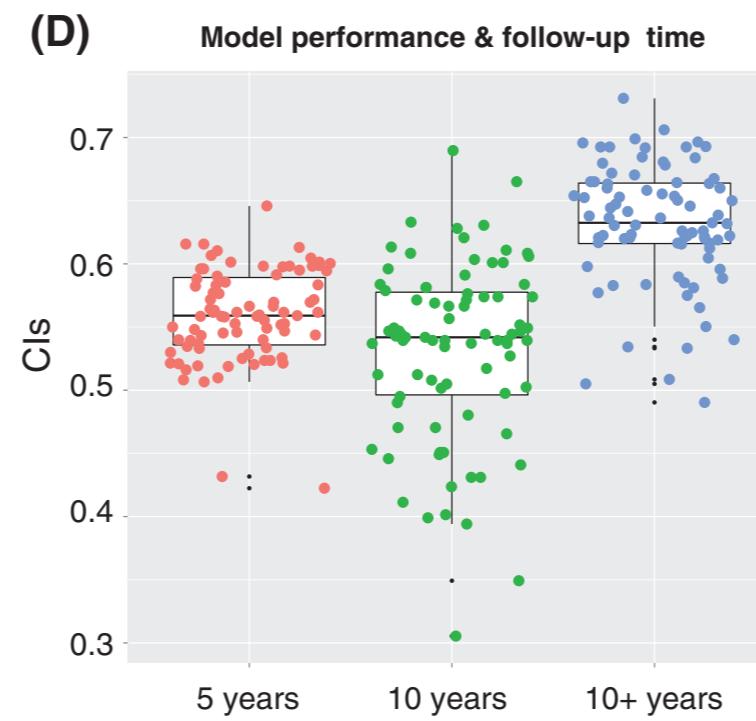
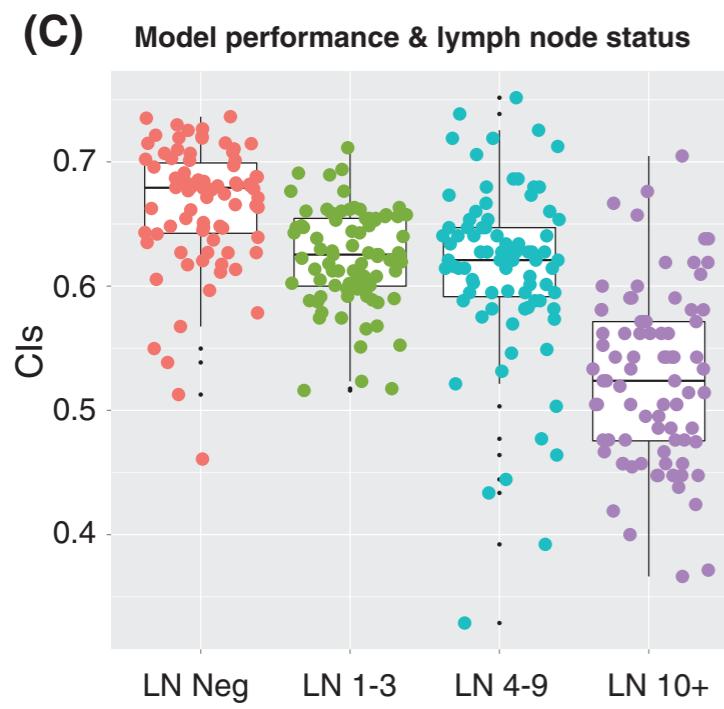
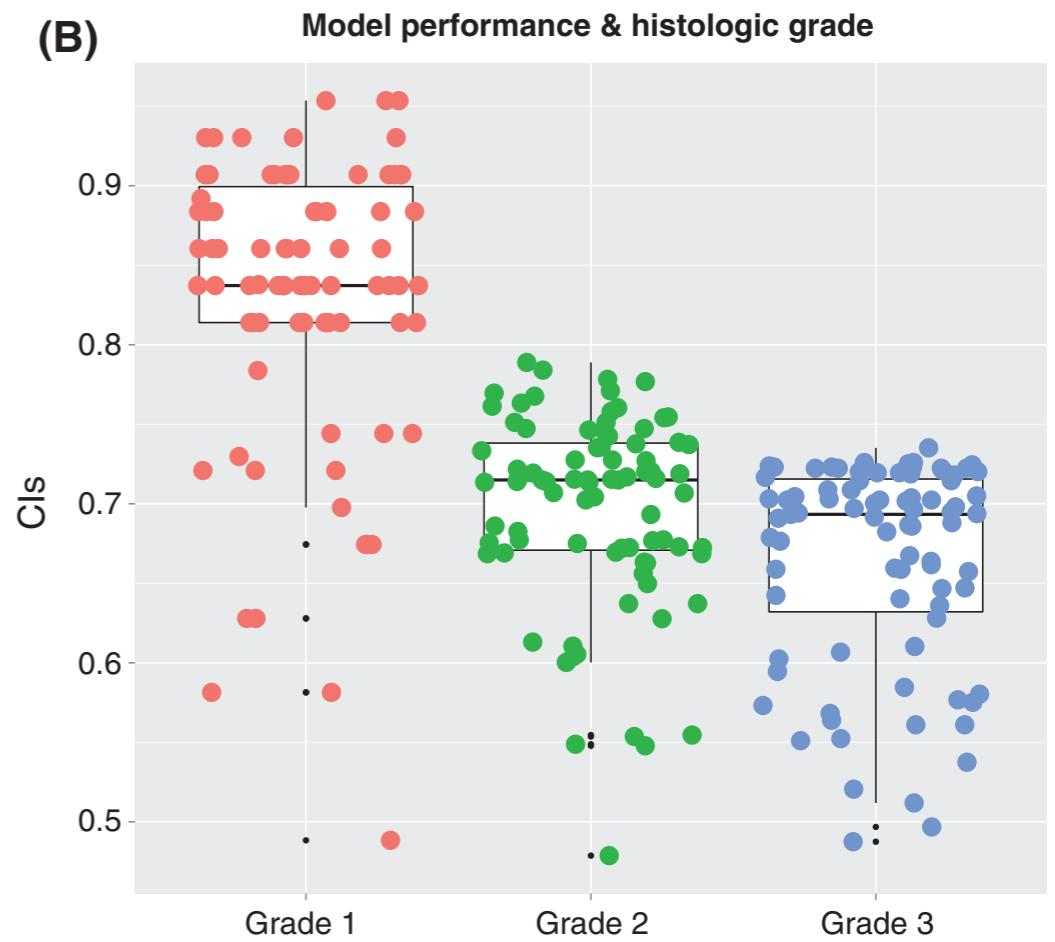
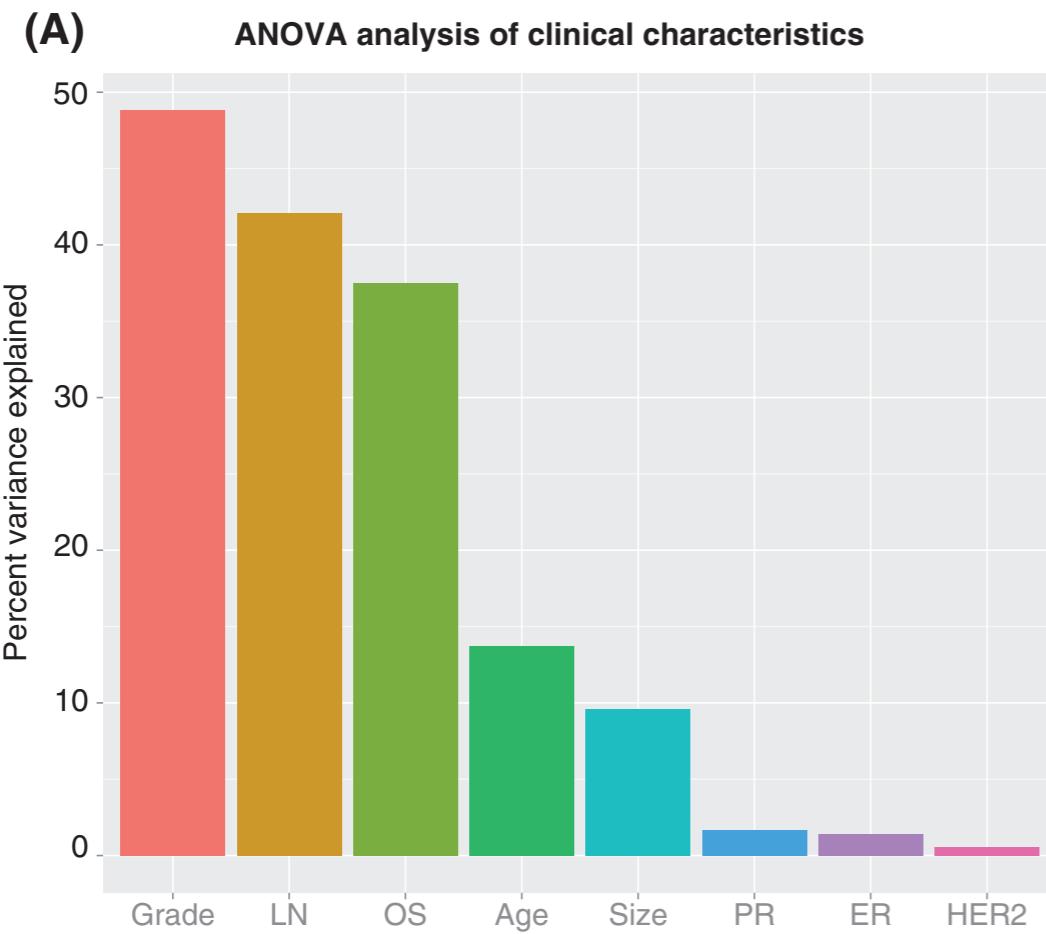


(D) 15 October versus OsloVal CIs



Random subsampling testing of rank stability: Final OsloVal models





Final Oslo Validation Leaderboard (top 20)

Synapse ID	Model Name	Concordance Index	Rank	Created By
syn1417992	Attractor Metagenes Model 093009	0.756186507	1	Wei-Yi Cheng
syn1417956	Attractor Metagenes Model 093005	0.75403751	2	Wei-Yi Cheng
syn1444414	Attractor Metagenes Model 101513	0.75345142	3	Wei-Yi Cheng
syn1444422	Attractor Metagenes Model M	0.748111487	4	Tai-Hsien Ou Yang
syn1444394	Attractor Metagenes Model 101510 OSDS	0.746809065	5	Tai-Hsien Ou Yang
syn1444255	JCModelNewClin2	0.738734045	6	Jie Cheng
syn1444367	Attractor Metagenes Model 101508 OSDS	0.738278197	7	Tai-Hsien Ou Yang
syn1443244	Django11	0.737301381	8	Eric Bonnet
syn1444362	Attractor Metagenes Model 101507 OSDS	0.733589476	9	Tai-Hsien Ou Yang
syn1443545	JCModelNewClin	0.733328992	10	Jie Cheng
syn1444275	JCModelNewClinGenes	0.731375358	11	Jie Cheng
syn1442225	JGAB rsf modded	0.731049753	12	Archit Bhise
syn1417893	WY093001	0.73091951	13	Wei-Yi Cheng
syn1443544	Django15	0.730268299	14	Eric Bonnet
syn1443208	Django9	0.728770513	15	Eric Bonnet
syn1426931	NCIS_S01E06	0.72857515	16	Tai-Hsien Ou Yang
syn1444370	Attractor Metagenes Model 101509	0.727858817	17	Wei-Yi Cheng
syn1443507	ToyModel-05-32	0.724733003	18	Vincenzo Lagani
syn1418229	RATestModel/20	0.724667882	19	Rob Atlas
syn1441207	GSOA 8	0.7240167	20	Scooter Willis



Site

Overall Survival



Site

Overall Survival



Regress Out?

Keep in the model?



Final Oslo Validation Leaderboard (top 20)

Synapse ID	Model Name	Concordance Index	Rank	Created By
syn1417992	Attractor Metagenes Model 093009	0.756186507	1	Wei-Yi Chen
syn1417956	Attractor Metagenes Model 093005	0.75403751	2	Wei-Yi Chen
syn1444414	Attractor Metagenes Model 101513	0.75345142	3	Wei-Yi Chen
syn1444422	Attractor Metagenes Model M	0.748111487	4	Tai-Hsien Wang
syn1444394	Attractor Metagenes Model 101510 OSDS	0.746809065	5	Tai-Hsien Wang
syn1444255	JCModelNewClin2	0.738734045	6	Jie Cheng

Final Oslo Validation Leaderboard (top 20)

Synapse ID	Model Name	Concordance Index	Rank	Created By
syn1417992	Attractor Metagenes Model 093009	0.756186507	1	Wei-Yi Chen
syn1417956	Attractor Metagenes Model 093005	0.75403751	2	Wei-Yi Chen
syn1444414	Attractor Metagenes Model 101513	0.75345142	3	Wei-Yi Chen
syn1444422	Attractor Metagenes Model M	0.748111487	4	Tai-Hsien Wang
syn1444394	Attractor Metagenes Model 101510 OSDS	0.746809065	5	Tai-Hsien Wang
syn1444255	JCModelNewClin2	0.738734045	6	Jie Cheng

BASELINE NAIVE CLASSIFIER ~ 0.72



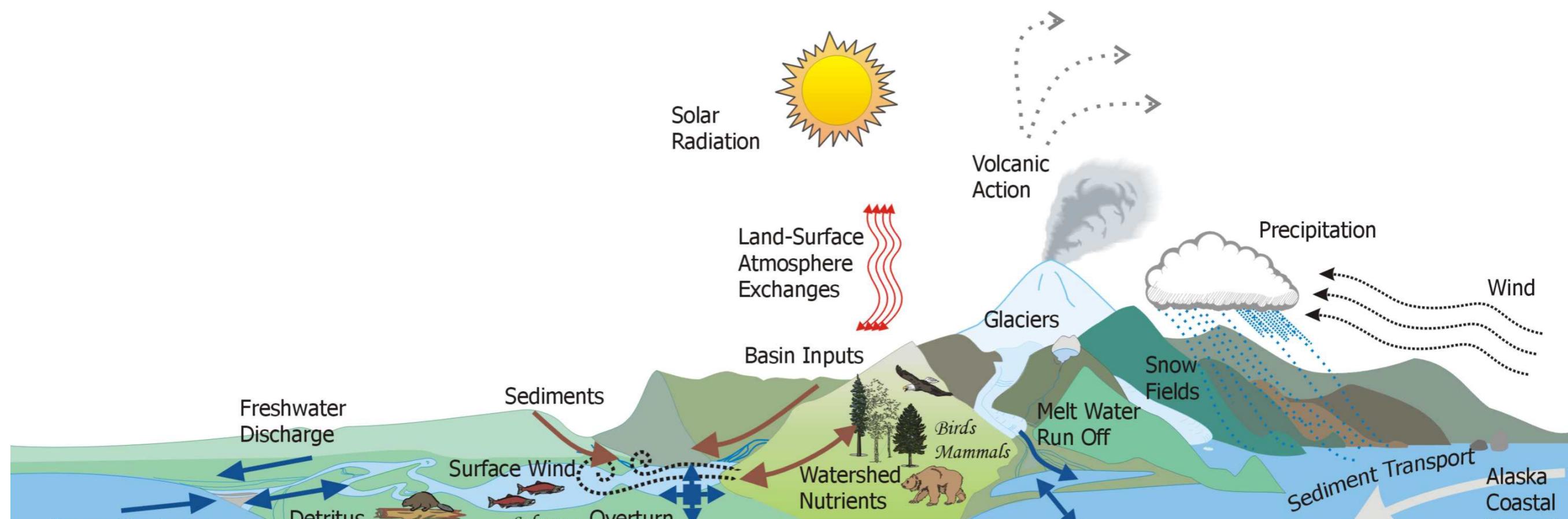
*IS PERFORMANCE THE ONLY
CRITERION FOR A GOOD MODEL?*

Final Oslo Validation Leaderboard (top 20)

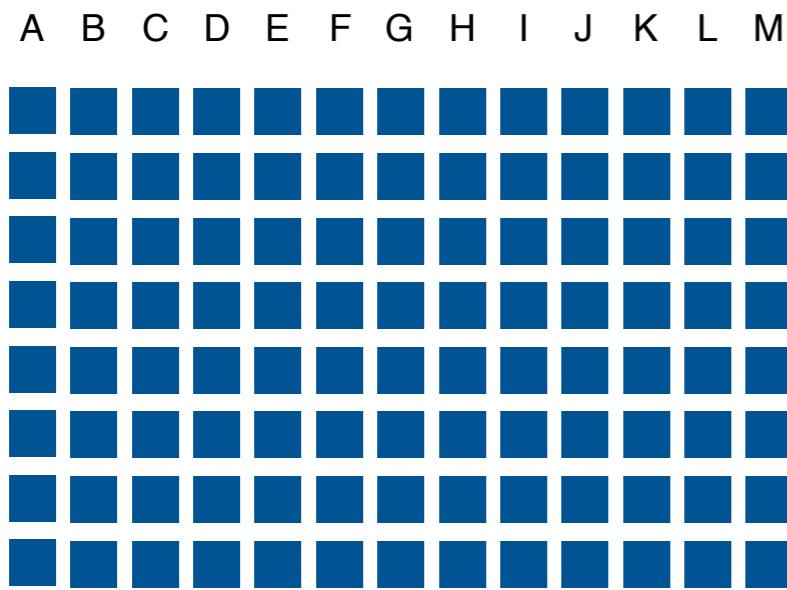
Synapse ID	Model Name	Concordance Index	Rank	Created By
syn1417992	Attractor Metagenes Model 093009	0.756186507	1	Wei-Yi Chen
syn1417956	Attractor Metagenes Model 093005	0.75403751	2	Wei-Yi Chen
syn1444414	Attractor Metagenes Model 101513	0.75345142	3	Wei-Yi Chen
syn1444422	Attractor Metagenes Model M	0.748111487	4	Tai-Hsien Chen
syn1444394	Attractor Metagenes Model 101510 OSDS	0.746809065	5	Tai-Hsien Chen
syn1444255	JCModelNewClin2	0.738734045	6	Jie Cheng

*BASELINE NAIVE CLASSIFIER ~ 0.72
MANY MODELS OUTRIGHT FAILED...*

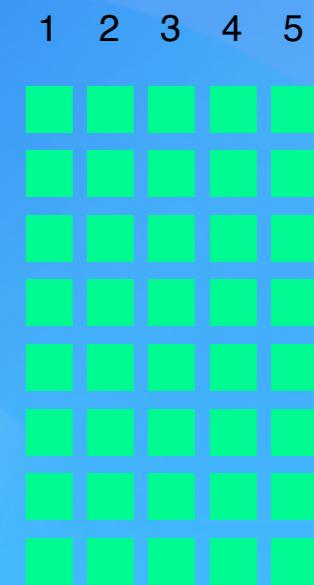
ECOSYSTEM



"METABRIC DATASET"



"OSLOVAL"



SYNAPSE CLOUD

FINAL ROUND

PREDICTIVE
MODEL

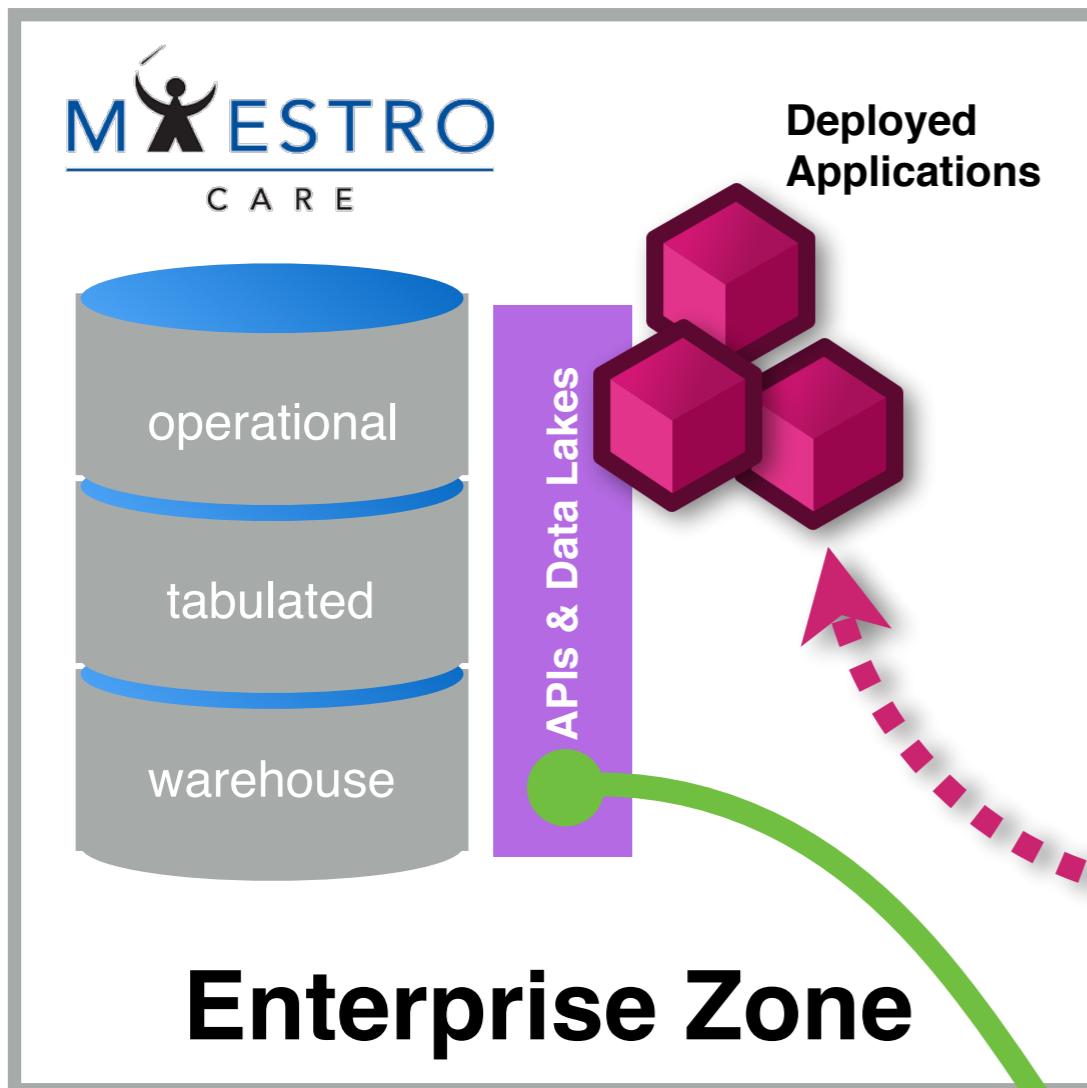
011
110

Google

VALIDATION
HARNESS

LEADERBOARD

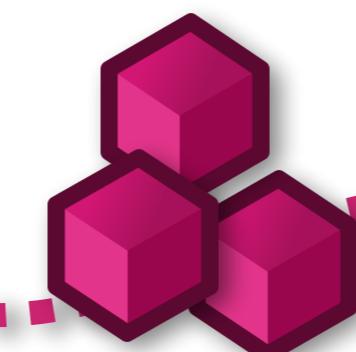
A DUKE UNIVERSITY Precision Health Ecosystem



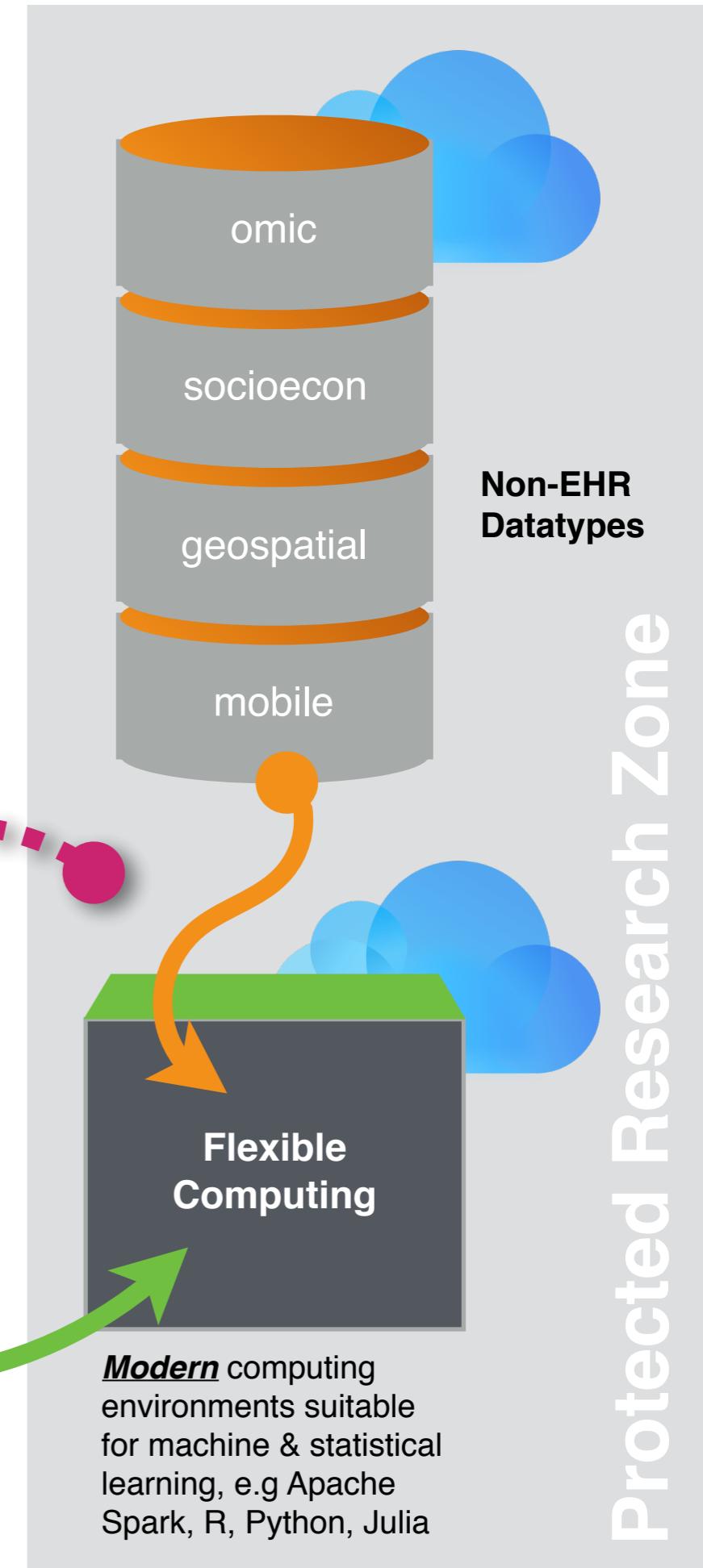
Precision Health is Data Science: Precision Health Applications require the integration of dependent and independent variables from EHRs, -omic data from advanced laboratory platforms and NGS, and other “non-EHR datatypes” such as socioeconomic, geospatial, and mobile health data. Critically, these integrated data need to generate testable **machine learning or statistical inference** for stratifying individuals or populations for health interventions. This **absolutely requires data liquidity combined with more advanced and flexible computing and data platforms**. In their absence, precision health is rhetoric, not action. EHRs are necessary, but far from sufficient for supporting these needs.

Agile Governance Path

automated testing
objective performance metrics



Precision Health Applications



PERFORMANCE TESTING

*objectively
benchmark models*

FUNGIBLE COMPUTE

*capable of handling
arbitrary machine
learning
methodologies*



INTERFACES

*push and pull data in
an enterprise EHR*

≡ MENU

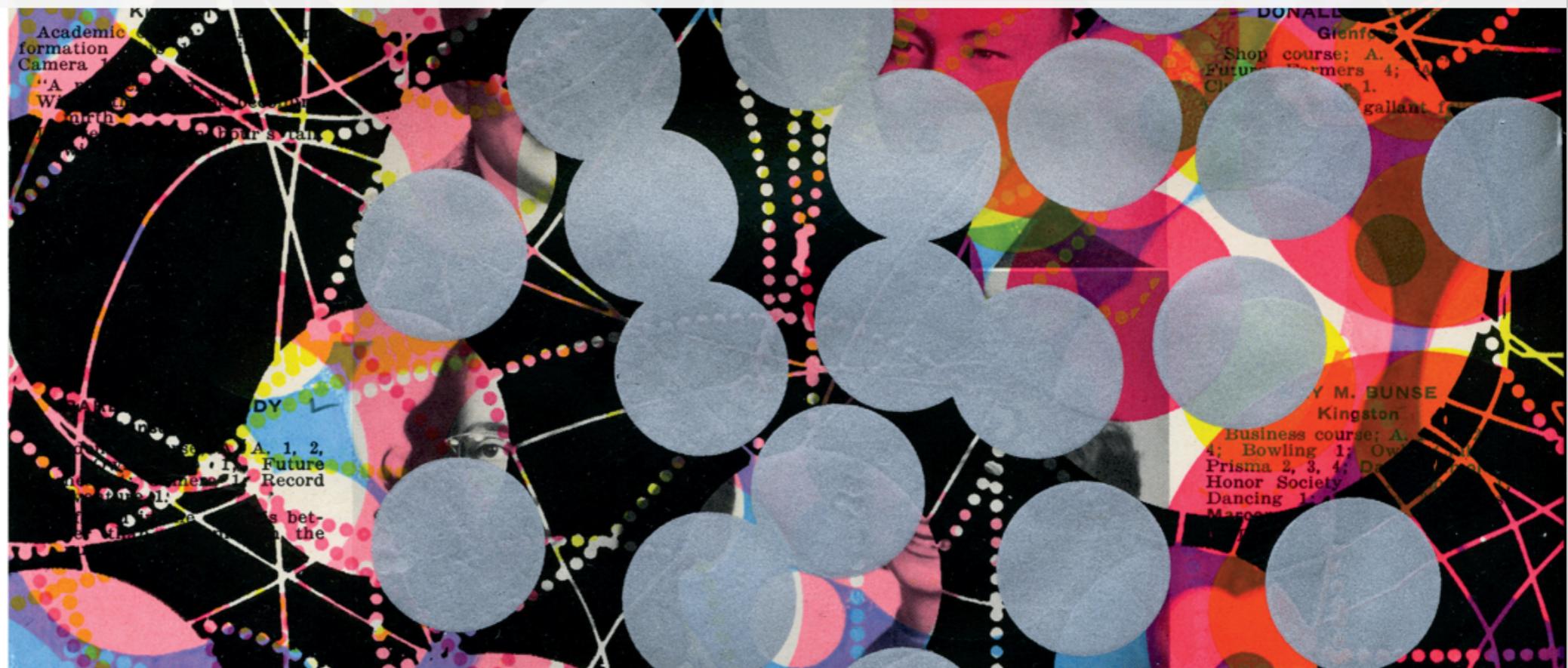
Harvard
Business
Review



SUBSCRIBE >



SIGN IN



ARTWORK: TAMAR COHEN, ANDREW J BUBLITZ, 2011, SILK SCREEN
ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

WHAT TO READ NEXT



Big Data: The Management Revolution



Fast Healthcare Interoperability Resources
Modular, Composable, Queryable, Extensible



SMART®





The Cheesecake Factory®

sandwiches

All Served with French Fries or Green Salad

Sweet Potato Fries 1.00 extra

RENEE'S SPECIAL 10.50

One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich,
a Cup of Our Soup and a Small Green Salad
with a Small Caesar Salad 1.00 extra

CHICKEN SALAD SANDWICH 11.50

Housemade with Roasted Almonds, Lettuce,
Tomato and Mayonnaise.
Served on Grilled Brioche Bread

THE CLUB 12.95

Freshly Roasted Turkey Breast, Bacon,
Lettuce, Tomato and
Mayonnaise on White Toast

GRILLED CHICKEN AND AVOCADO CLUB 12.95

Grilled Chicken Breast with Avocado, Bacon, Tomato, Melted Swiss and Herb Mayonnaise

THE NAVAJO 11.95

Warm Fry-Bread Stuffed with Grilled Chicken, Avocado, Lettuce,
Tomato, Red Onion and Mayonnaise

CUBAN SANDWICH 11.95

Slow Roasted Pork, Ham, Swiss Cheese, Pickles, Mustard and Mayonnaise
on a Grilled and Pressed Cuban-Style Roll

CHICKEN PARMESAN SANDWICH 11.95

Tender Chicken Lightly Coated in Parmesan Breadcrumbs, Roasted Peppers, Tomato Sauce
and Melted Cheese on a Freshly Grilled French Roll

BLACKENED CHICKEN SANDWICH 11.95

Grilled with Melted Fontina Cheese, Lettuce, Tomato, Red Onion and Spicy Mayonnaise

CALIFORNIA CHEESESTEAK 11.95

Thinly Sliced Grilled Steak Covered with Sauteed Mushrooms, Onions,
Peppers and Cheese on a Toasted Roll

GRILLED SHRIMP & BACON CLUB 13.95

Charbroiled Shrimp, Bacon, Lettuce and Tomato with Our Special Dressing

cheesecakes

ORIGINAL

The One that Started it All!

Our Famous Creamy Cheesecake with a Graham Cracker Crust and Sour Cream Topping

FRESH STRAWBERRY

The Original Topped with Glazed Fresh Strawberries. Our Most Popular Flavor for 30 Years!

REESE'S® PEANUT BUTTER CHOCOLATE CAKE CHEESECAKE™

Reese's Peanut Butter Cups in Our Original Cheesecake with Layers of Delicious Fudge Cake and Caramel

30th ANNIVERSARY CHOCOLATE CAKE CHEESECAKE

Layers of Our Original Cheesecake, Fudge Cake and Chocolate Truffle Cream

WHITE CHOCOLATE RASPBERRY TRUFFLE®

Creamy Cheesecake Swirled with White Chocolate and Raspberry

ULTIMATE RED VELVET CAKE CHEESECAKE™

Layers of Red Velvet Cake and Cheesecake Covered in Cream Cheese Frosting. Finished with White Chocolate

GODIVA® CHOCOLATE CHEESECAKE

Flourless Godiva Chocolate Cake, Topped with Godiva Chocolate Cheesecake and Chocolate Mousse

FRESH BANANA CREAM CHEESECAKE

Banana Cream Cheesecake Topped with Bavarian Cream and Fresh Sliced Bananas

ADAM'S PEANUT BUTTER CUP FUDGE RIPPLE

Creamy Cheesecake Swirled with Caramel, Peanut Butter, Butterfingers® and Reese's Peanut Butter Cups™

WHITE CHOCOLATE CARAMEL MACADAMIA NUT CHEESECAKE

White Chocolate Chunk Cheesecake Swirled with Macadamia Nuts and Caramel on a Blond Brownie Crust

LEMON RASPBERRY CREAM CHEESECAKE

Raspberry-Vanilla Cake, Creamy Lemon Cheesecake, Raspberry Lady Fingers and Lemon Mousse

DULCE DE LECHE CARAMEL CHEESECAKE

Caramel Cheesecake Topped with Caramel Mousse on a Vanilla Crust

CHOCOLATE COCONUT CREAM CHEESECAKE

Coconut Cheesecake Topped with Coconut Cream Custard, all on a Chocolate Macaroon Crust

TIRAMISU CHEESECAKE

Our Wonderful Cheesecake and Tiramisu Combined into one Amazing Dessert!

CHOCOLATE MOUSSE CHEESECAKE

Silky Chocolate Cheesecake Topped with a Layer of Belgian Chocolate Mousse

VANILLA BEAN CHEESECAKE

Creamy Vanilla Bean Cheesecake, Topped with Vanilla Mousse and Whipped Cream

CHOCOLATE TUXEDO CREAM™ CHEESECAKE

Layers of Our Fudge Cake, Chocolate Cheesecake, Vanilla Mascarpone Mousse and Chocolate

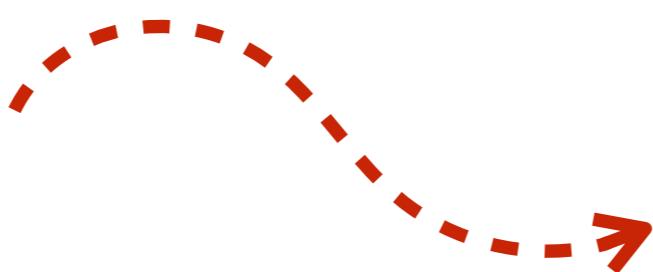
KAHLUÀ® COCOA COFFEE CHEESECAKE

A Rich Brownie, Topped with Kahlua Cheesecake, Creamy Chocolate Mousse and Chocolate Ganache

The Cheesecake Factory®



kitchen



service fulfilled



RENEE'S SPECIAL 10.50

*One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich,
a Cup of Our Soup and a Small Green Salad
with a Small Caesar Salad 1.00 extra*

≡ SECTIONS

HOME

SEARCH

The New York Times

PLAY THE CROSSWORD

esh22 ▾



Congress Moves to Strike
Internet Privacy Rules
From Obama Era



YouTube Advertiser
Exodus Highlights Perils of
Online Ads



Fake Sleuths: Web Gets It
Wrong on London
Attacker



TECH FIX
Crossing the Border?
Here's How to Safeguard
Your Data From Searches

PAID POST: GUINNESS
Why Irish Pubs Are a
Metaphor for American
Immigration

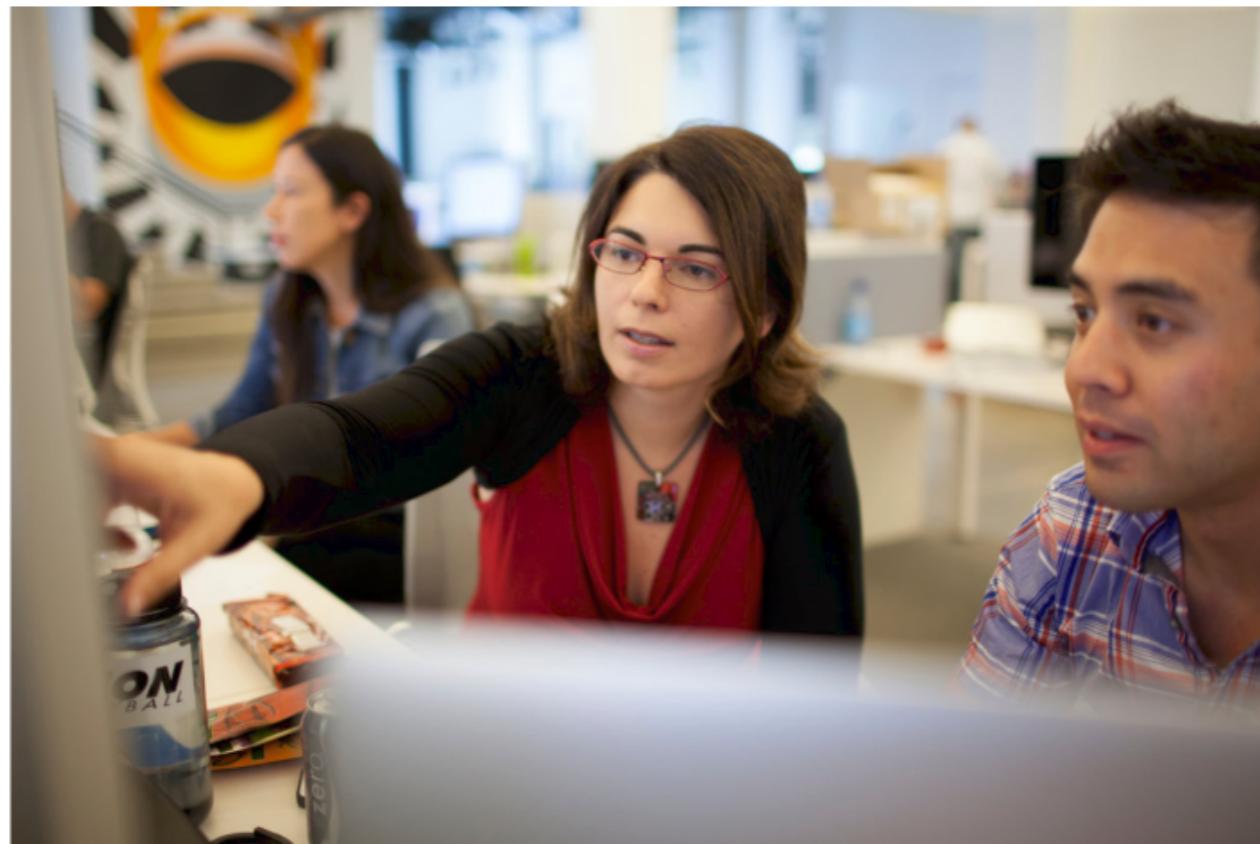


TECH TIP
Smartphones That C

TECHNOLOGY

For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights

By STEVE LOHR AUG. 17, 2014



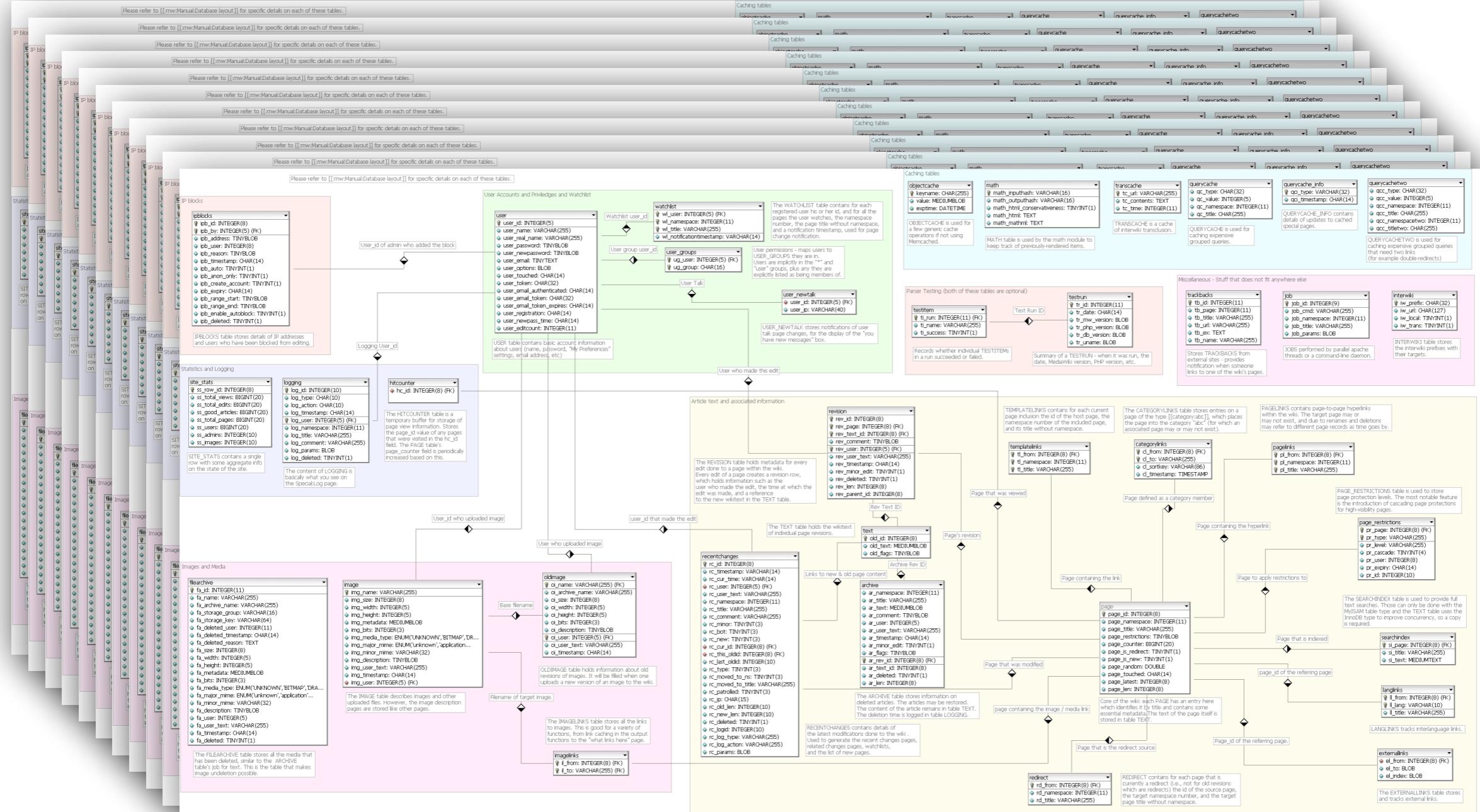
Monica Rogati, Jawbone's vice president for data science, with Brian Wilt, a senior data scientist.
Peter DaSilva for The New York Times

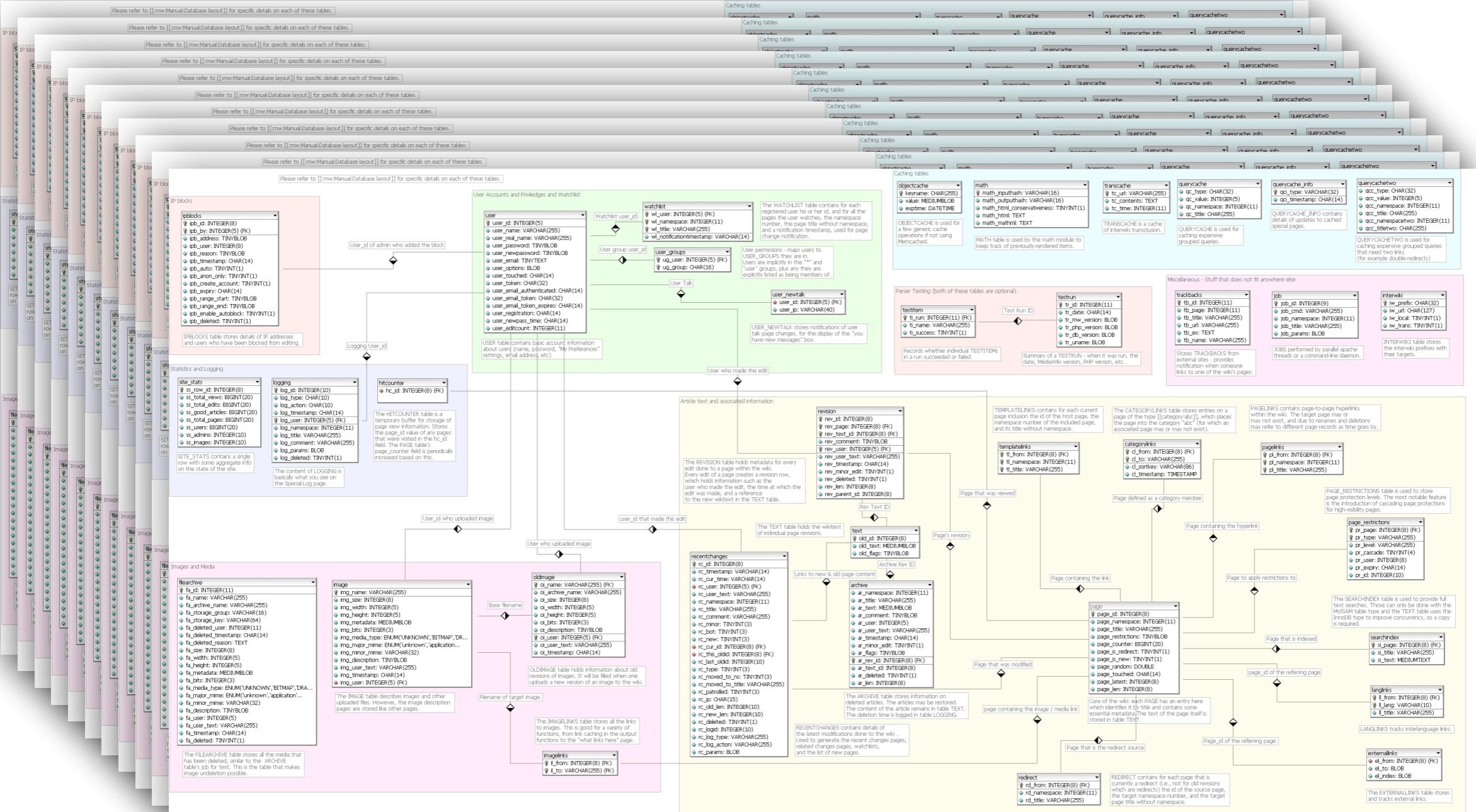
Technology revolutions come in measured, sometimes foot-dragging steps.

The lab science and marketing enthusiasm tend to underestimate the

ETL

RENEE'S SPECIAL 10.50
*One-Half of a Fresh Turkey Sandwich or Chicken-Almond Salad Sandwich,
 a Cup of Our Soup and a Small Green Salad
 with a Small Caesar Salad 1.00 extra*





```
{
  "_id" : ObjectId("5876a35e65f7db4c273bbd9d"),
  "fullUrl" : "urn:uuid:12e74960-3fa7-4b03-a5ff-134a0b45c529",
  "resource" : {
    "id" : "12e74960-3fa7-4b03-a5ff-134a0b45c529",
    "status" : "final",
    "code" : {
      "coding" : [
        {
          "system" : "http://loinc.org",
          "code" : "8302-2",
          "display" : "Body Height"
        }
      ]
    }
  }
}
```

FHIR Encounter Resource

This is the current officially published version of *The Argonaut Data Query Implementation Guide Version 1.0.0*. For list of available versions, see the [Directory of published versions](#).

Argonaut Data Query Implementation Guide

Data Element Query

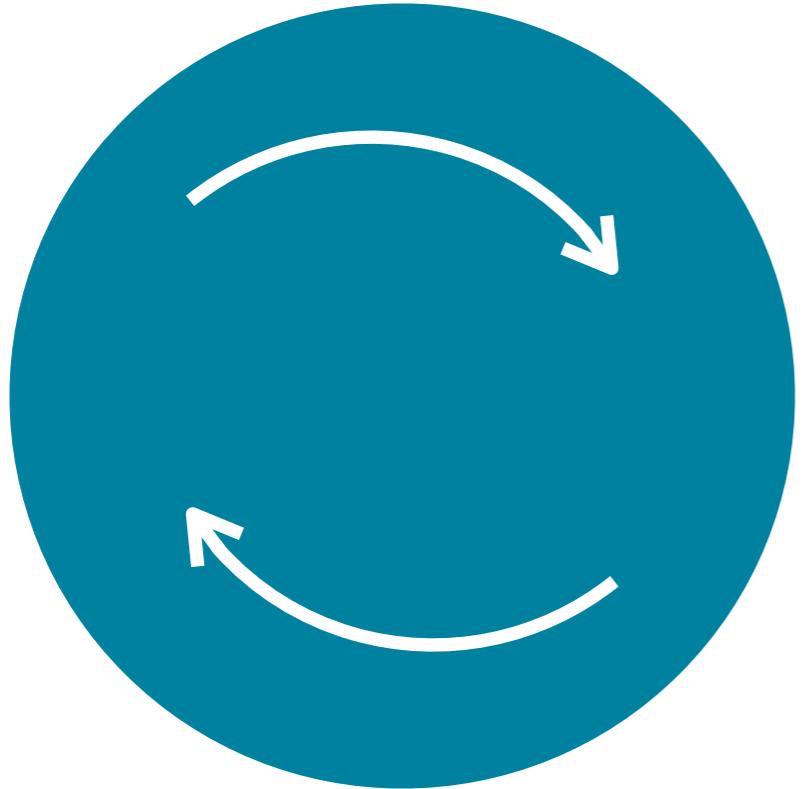
The Argonaut data element query IG is intended to meet the 2015 Edition certification criterion for Patient Selection 170.315(g)(7) and Application Access – Data Category Request 170.315(g)(8). They were created for each of the 2015 Edition Common Clinical Data Set. Where applicable they are based on the HL7 U.S. [Data Access Framework \(DAF\) FHIR DSTU2 Implementation Guide](#). However, the Argonaut use case and requirements per resource are a subset of those of the DAF implementation guide.

The table below lists the FHIR Resources used for the corresponding 2015 Edition Common Clinical Data Set (CCDS) Data elements:

No	CCDS Data Element	FHIR Resource
(1)	Patient Name	Patient
(2)	Sex	Patient
(3)	Date of birth	Patient
(4)	Race	Patient
(5)	Ethnicity	Patient
(6)	Preferred language	Patient
(7)	Smoking status	Observation
(8)	Problems	Condition
(9)	Medications	Medication, MedicationStatement, MedicationOrder
(10)	Medication allergies	AllergyIntolerance
(11)	Laboratory test(s)	Observation, DiagnosticReport
(12)	Laboratory value(s)/result(s)	Observation, DiagnosticReport
(13)	Vital signs	Observation
(14)	(no longer required)	-
(15)	Procedures	Procedure
(16)	Care team member(s)	CarePlan
(17)	Immunizations	Immunization
(18)	Unique device identifier(s) for a patient's implantable device(s)	Device
(19)	Assessment and plan of treatment	CarePlan
(20)	Goals	Goal
(21)	Health concerns	Condition

The Argonaut Profiles for each of the data element queries is listed below. Each profile defines the minimum mandatory elements, extensions and terminology requirements that **MUST** be present. Requirements and guidance are given in the profile narrative summary. A formal hierarchical table that presents a [logical view](#) of the content in both a differential and snapshot view is also provided along with references to appropriate terminologies and examples.

- [Argonaut AllergyIntolerance Profile](#)
- [Argonaut CarePlan Profile](#)
- [Argonaut CareTeam Profile](#)
- [Argonaut Condition Profile](#)
- [Argonaut Device Profile](#)
- [Laboratory Tests and Results](#)
- [Argonaut DiagnosticReport Profile](#)



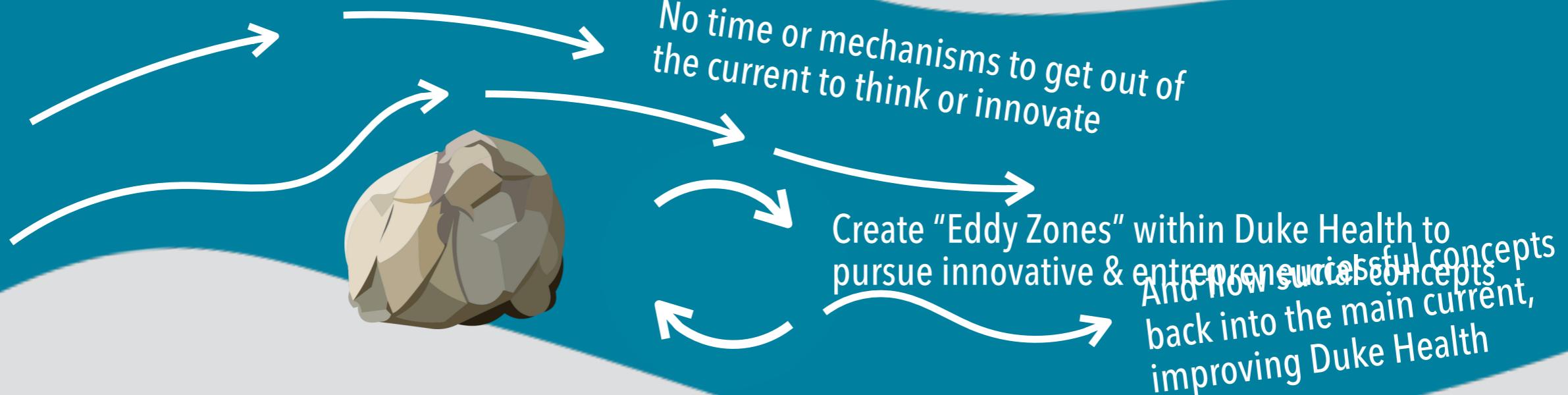
Eddy Zones!

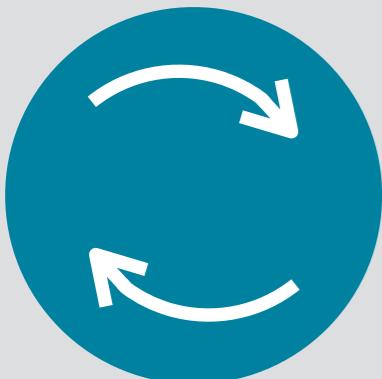
Data Science Initiative at Duke

The need for “Eddy Zones”

Larry Carin, Vice Provost
Michael Pencina, DCRI

Innovative ideas arising
from the “current” of
activities in Duke Health





A Center that creates an “Eddy Zone”

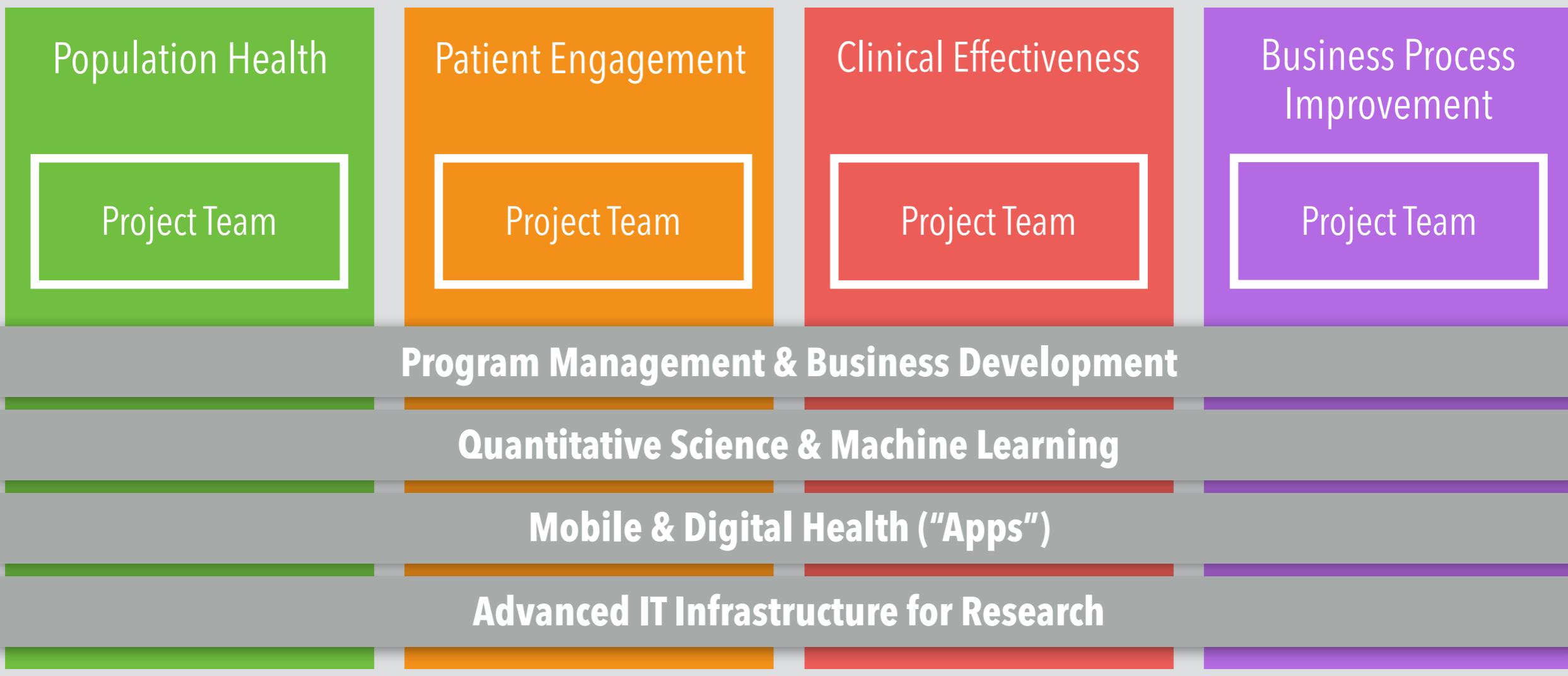
- Data & Analytics align with Population Health and Precision Medicine, as well as clinical and business intelligence
- Center will focus on health data science in all its forms
 - **Clinical:** from Epic, while integrating “non-EHR” data such as genomic and mobile data; and drive clinically-motivated analytic challenges
 - **Business Process:** improving health delivery efficiencies and link to Margolis Center and health policy
 - **Basic Science:** addressing the most challenging and important areas of biomedical health sciences

Data Science Initiative at Duke

Institutionalizing “Eddy Zones”

“CENTER FOR HEALTHCARE DATA SCIENCE”

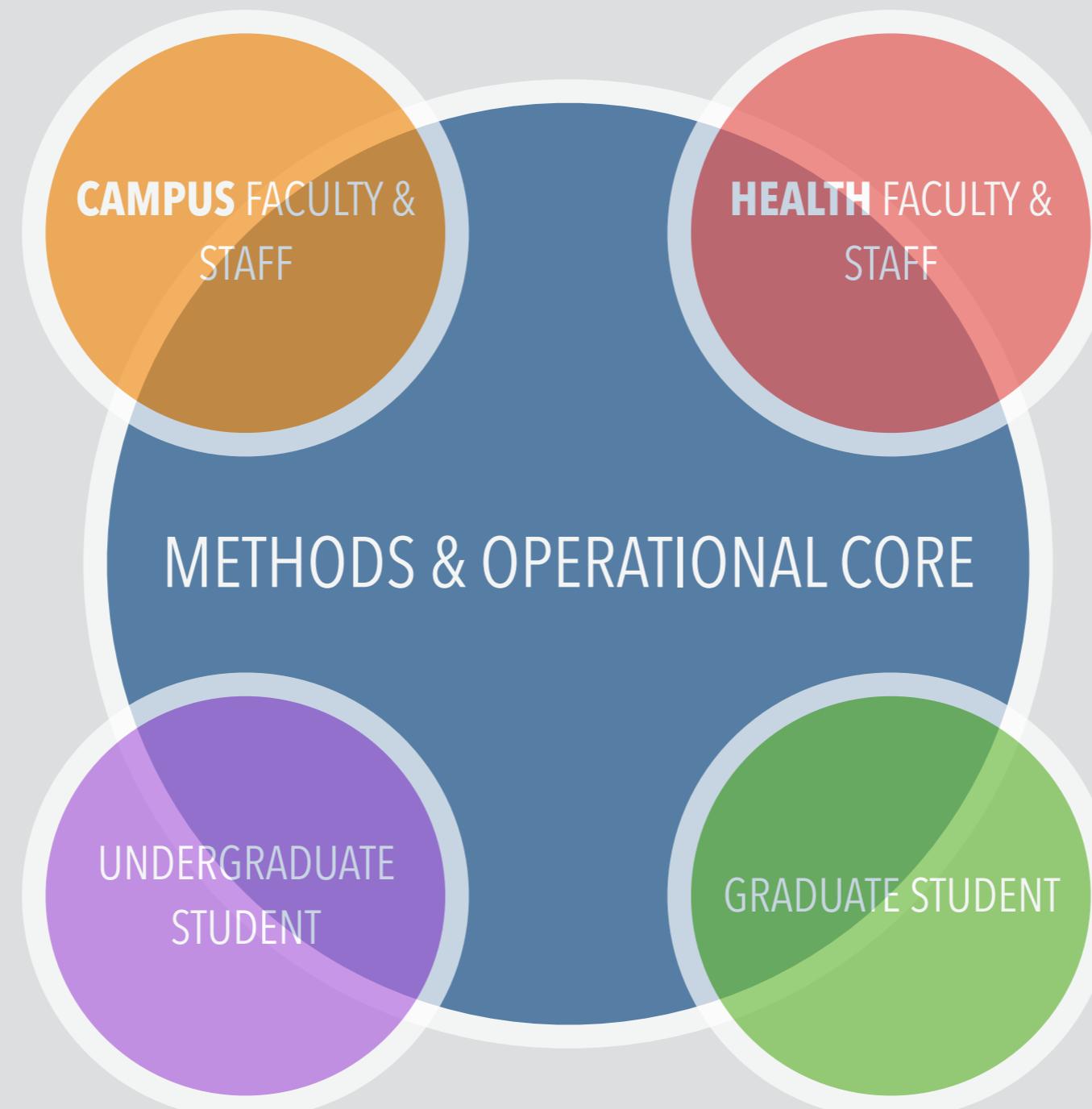
Functional Cores Strategic Domains



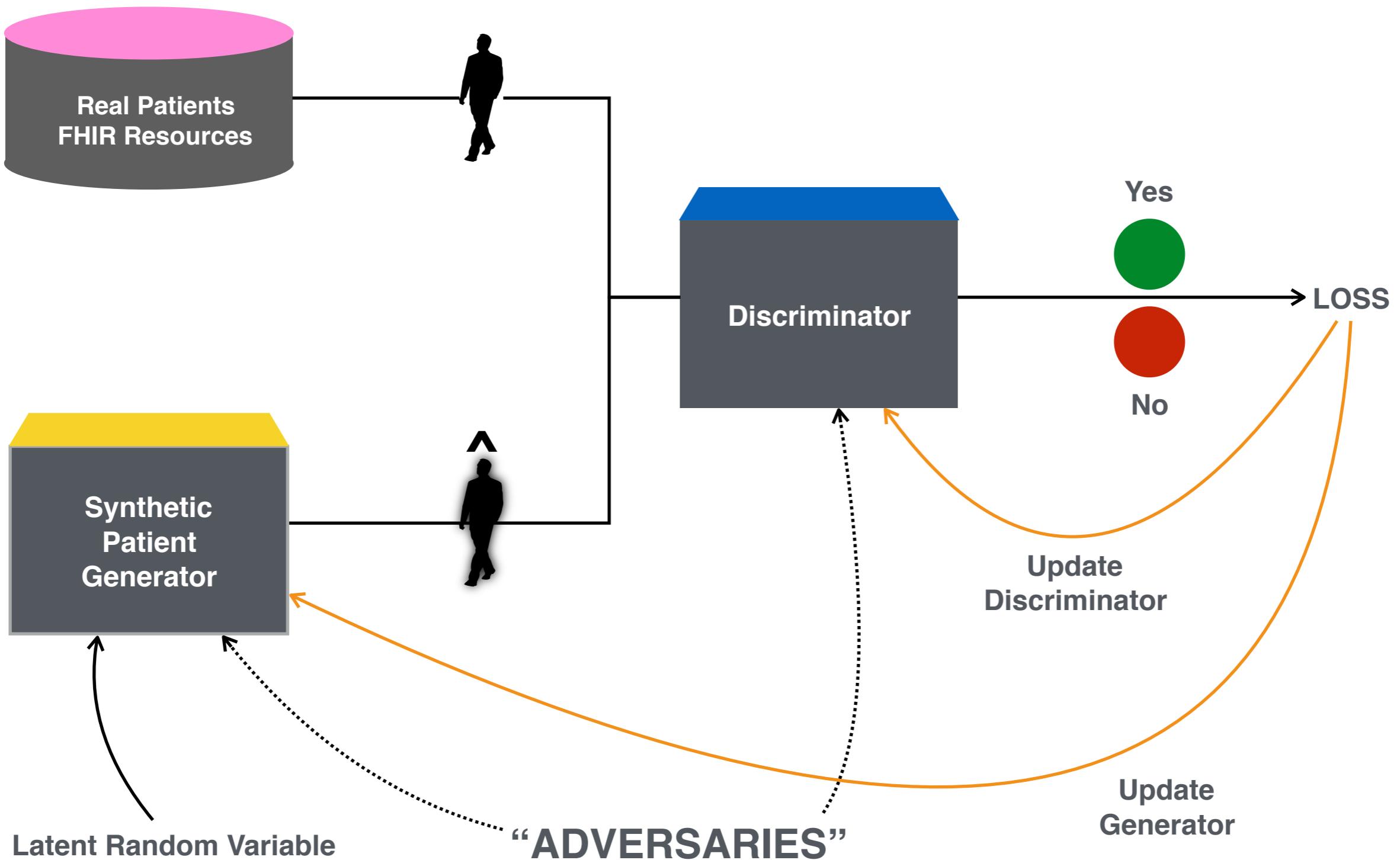
Data Science Initiative at Duke

Institutionalizing “Eddy Zones”

Project Team



Generative Adversarial Networks



Generated Samples of our model



(a) Faces



(b) Street View House Numbers

[Data Science], I think, is like a shark, you know? It has to constantly move forward or it dies.

–Woody Allen in *Annie Hall*

