Why did my patient develop acute kidney injury?

Using machine learning to understand peri-operative health outcomes

Karandeep Singh, MD, MMSc
Assistant Professor of Learning Health Sciences, Internal Medicine, Urology, and Information
University of Michigan

Disclosures

- No relevant disclosures or conflicts of interest

To answer this question, we will:

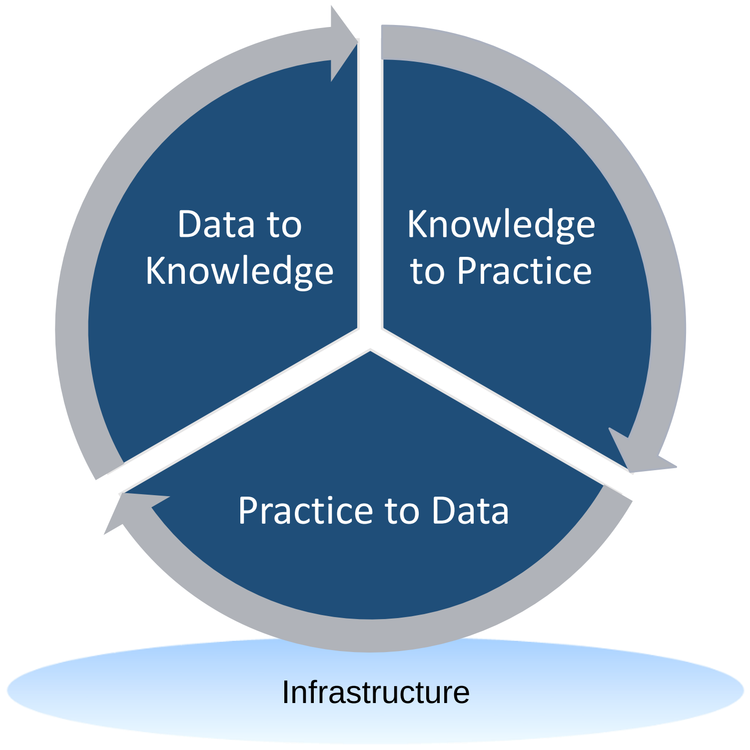
- Describe the role of risk models within learning health systems
- Discuss ways that risk can be quantified
- Understand how risk models are operationalized to improve quality of care
- Discuss issues with attribution for peri-operative outcomes
- Demonstrate how machine learning “model explanations” can be used to explain contributions of different factors (e.g., vital signs, surgeon) to risk in individual cases

My background

- Nephrologist
- Masters in biomedical informatics
- Research lab focuses on Machine Learning for Learning Health Systems
- Member of the MUSIC Collaborative Quality Initiative (urology)
- Co-chair of Clinical Intelligence Committee at U-M health system
- Member of Michigan Artificial Intelligence Advisory Board convened by MEDC and the Center for Automotive Research

What is a learning health system?

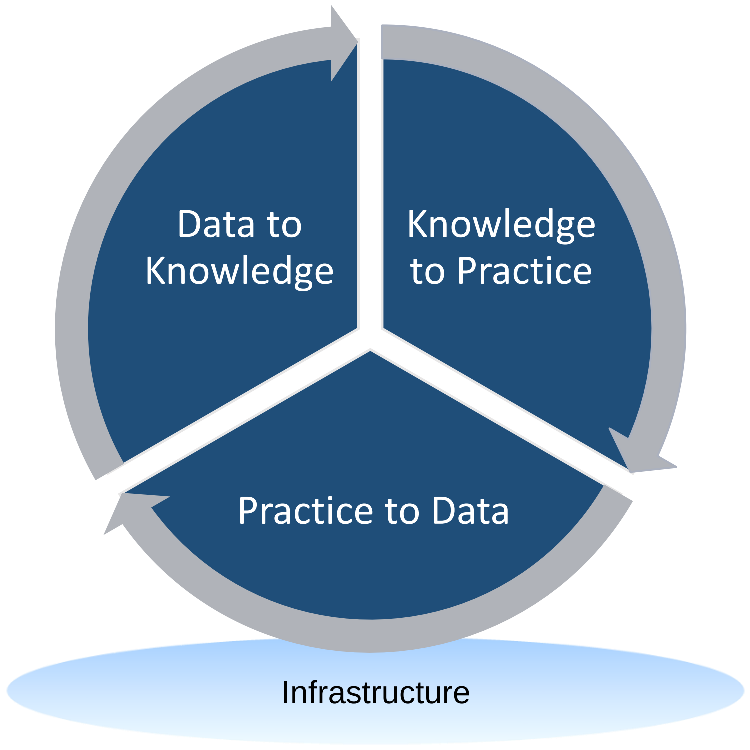
- Any system focused on:
  - Improving people's health
  - Through continuous cycles of knowledge discovery **and** implementation of best practices
  - And doing this at scale



Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, Gunter C, Musen M, Platt R, Stead W, Sullivan K, Van Houweling D: Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J. Am. Med. Informatics Assoc.* 43–50, 2014

What does learning health system require?

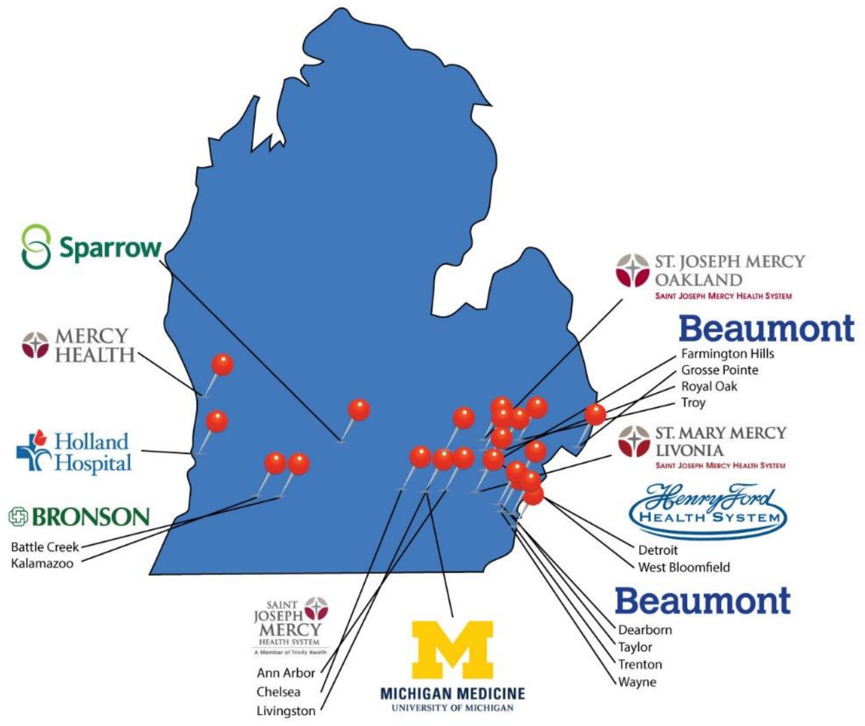
- A community engaged in continuous learning through repeated learning cycles
- Learning cycles
  - Practice to Data
  - Data to Knowledge
  - Knowledge to Practice
- Infrastructure: IT, governance, ethics, policy



What is a learning health system?

- The Collaborative Quality Initiatives are exemplars of learning health systems
  - A community engaged in continuous learning that crosses organizational boundaries
  - Rich data collection through manual and automated mechanisms
  - Ability to learn from multi-institutional data
  - Ability to implement interventions at scale

ASPIRE



What is the role of risk models in learning health systems?

At a patient level

- Select the best diagnostic or treatment approach
  - Example: Mallampati score

- Counsel patients on prognosis

At a system level

- Allocate resources more efficiently
  - Example: Michigan Medicine birthing center delivery volume model

- Early warning systems
  - Example: Michigan Medicine sepsis pilot

- Identify areas of improvement
  - Example: comparing observed versus expected risk

How can we quantify risk?

- Decision tree models
- Regression models
- Machine learning models

How can we quantify risk?

- Decision tree models

RESPIRATION AND THE AIRWAY

Prospective validation of a new airway management algorithm and predictive features of intubation difficulty

F. Cook[1], D. Lobo[1], M. Martin[1], N. Imbert[1,3], H. Grati[1], N. Daami[1], C. Cherait[1], N.-E. Saïdi[1], K. Abbay[1], J. Jaubert[1], K. Younsi[1], S. Bensaid[1], B. Ait-Mamar[1], V. Slavov[1], R. Mounier[1], P. Goater[2], S. Bloc[1,3], J. Catineau[1], K. Abdelhafidh[2], H. Haouache[2] and G. Dhonneur[2,3,*]

[1]Department of Intensive Care Medicine, Henri Mondor University Hospital, Créteil, France, [2]Curie Cancer Institutes of Paris, Paris, France and [3]Paris 12 School of Medicine, Créteil, France

Table 1 Strategy for rating seven assessed airway features

Ranks	Class 0	Class 1	Class 2	Class 3
Features	No problem	Relevant difficult direct laryngoscopy to severe problems		
Mouth opening (cm) or inter-incisor gap	>5 >3 Fingers	5–4 3 Fingers	4–3 2 Fingers	3–2 Thumb
BMI (kg m $^{-2}$)	<25	25–30	31–40	>40
Modified Mallampati (grade)	1	2	3	4
Previous difficult intubation	None	Failed direct laryngoscopy	Macintosh laryngoscope+ Stylet failure	Previous awake intubation
Mandible profile	Normal	Slightly erased	Clearly erased upper lip bite test positive	Retrognatia upper lip bite test negative
Thyromental distance (cm)	>7 >4 Fingers	7–5 4 Fingers	5–3 3 Fingers	< 3 2 Fingers
Cervical spine mobility (°)	>90° Flexion/ extension	90–45° Flexion/extension	44–15° Flexion/extension	14–0°, or flexion fixed deformity

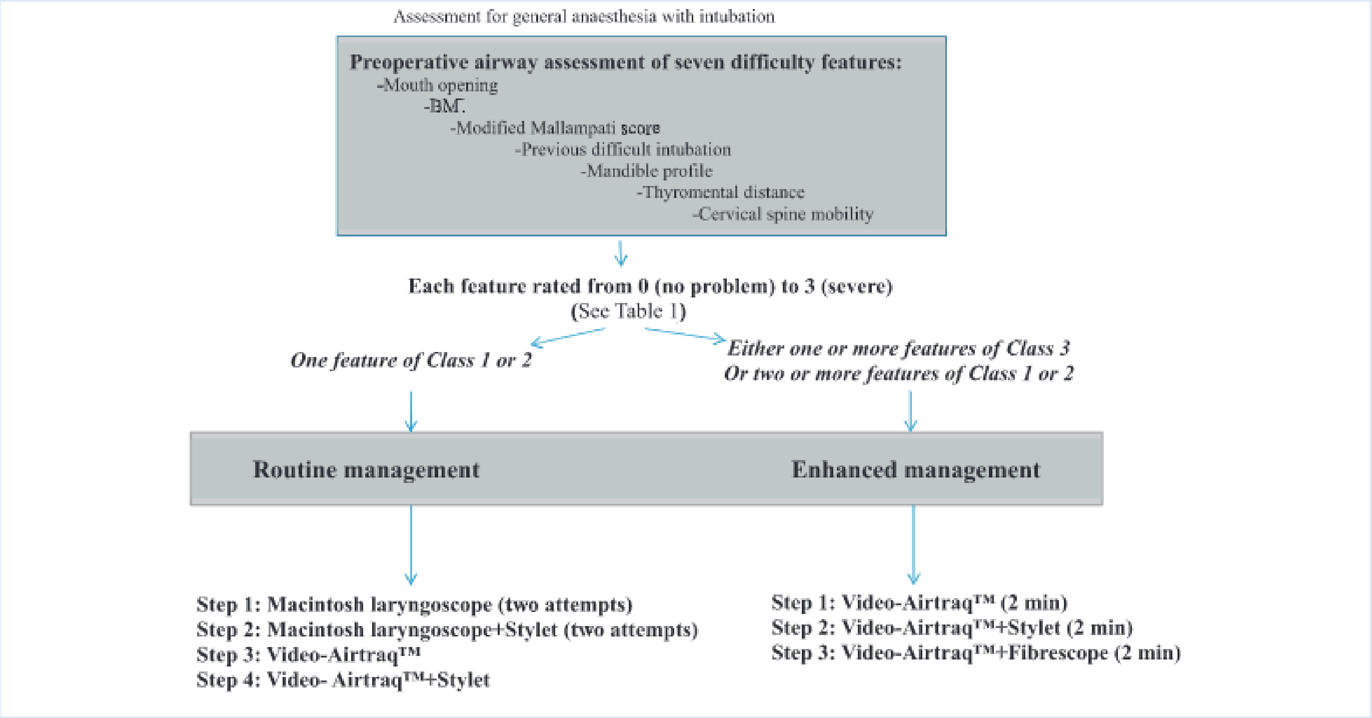


Fig 1. **Decision tree**. Selection of patients and management strategies.

Decision tree models

Advantages

- Easy to interpret and implement
- Can model simple interactions between predictor variables

Disadvantages

- Rarely the most accurate

How can we quantify risk?

- Regression models

TECHNOLOGY, COMPUTING, AND SIMULATION
SECTION EDITOR
STEVEN J. BARKER

SOCIETY FOR TECHNOLOGY IN ANESTHESIA

Predictive Performance of Three Multivariate Difficult Tracheal Intubation Models: A Double-Blind, Case-Controlled Study

Mohamed Naguib, MB, BCh, MSc, FFARCSI, MD*, Franklin L. Scamman, MD‡,
Cormac O'Sullivan, CRNA‡, John Aker, CRNA§, Alan F. Ross, MD‡,
Steven Kosmach, MSN, RN*, and Joe E. Ensor, PhD†

Departments of *Anesthesiology and Pain Medicine and †Biostatistics and Applied Mathematics, The University of Texas M. D. Anderson Cancer Center, Houston; ‡Department of Anesthesia, The University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City; and §Department of Anesthesia, Children's Mercy Hospitals & Clinics, Kansas City, Missouri

Table 1. Wilson Risk Sum Score (1)

Risk factor	Level	Variable
Weight	0	<90 kg
	1	90–110 kg
	2	>110 kg
Head and neck movement	0	>90°
	1	About 90° (i.e., ±10°)
	2	<90°
Jaw movement	0	IG ≥5 cm or SLux > 0
	1	IG < 5 cm and SLux = 0
	2	IG < 5 cm and SLux < 0
Receding mandible	0	Normal
	1	Moderate
	2	Severe
Buck teeth	0	Normal
	1	Moderate
	2	Severe

IG = Interincisor gap; SLux = Subluxation (maximal forward protrusion of the lower incisors beyond the upper incisors).

Table 2. Simplified Score Model Described by Arné et al. (20) for Prediction of Difficult Intubation

Risk factor	Score
Previous knowledge of difficult intubation	
No	0
Yes	10
Diseases associated with difficult intubation	
No	0
Yes	5
Clinical symptoms of airway pathology	
No	0
Yes	3
IG and mandible subluxation	
IG ≥ 5 cm or SLux > 0	0
IG < 5.0–3.5 cm and SLux = 0	3
IG < 3.5 cm and SLux < 0	13
Thyromental distance	
≥ 6.5 cm	0
< 6.5 cm	4
Maximum range of head and neck movement	
More than 100°	0
About 90° (±10°)	2
Less than 80°	5
Mallampati score	
Class 1	0
Class 2	2
Class 3	6
Class 4	8
Total possible	48

IG = interincisor gap; SLux = subluxation (maximal forward protrusion of the lower incisors beyond the upper incisors).

Logistic regression analysis identified four risk factors correlated with the prediction of difficult laryngoscopy and intubation: thyromental distance, interincisor gap, height, and Mallampati score. The prediction (l) was determined by the equation

$$l = 0.2262 - 0.4621 \times \text{thyromental distance} + 2.5516 \times \text{Mallampati score} - 1.1461 \times \text{interincisor gap} + 0.0433 \times \text{height},$$

Regression models

Advantages

- Easy to interpret
- Not too difficult to implement
- Sometimes the most accurate

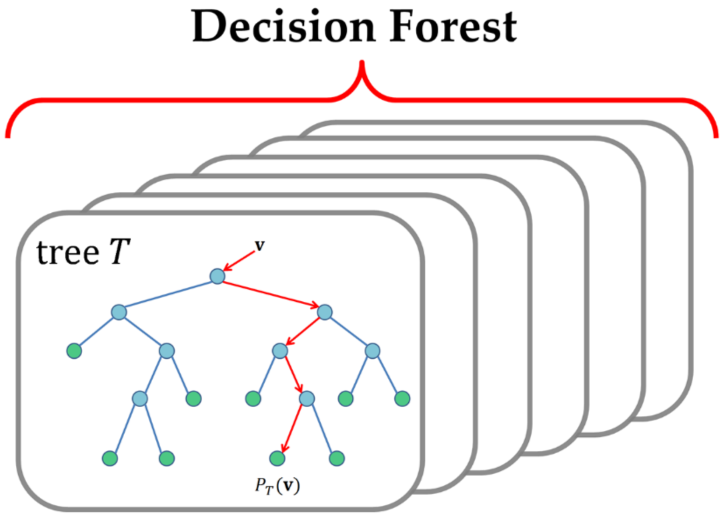
Disadvantages

- Will not capture non-linear relationships unless you use “polynomials” or “splines”

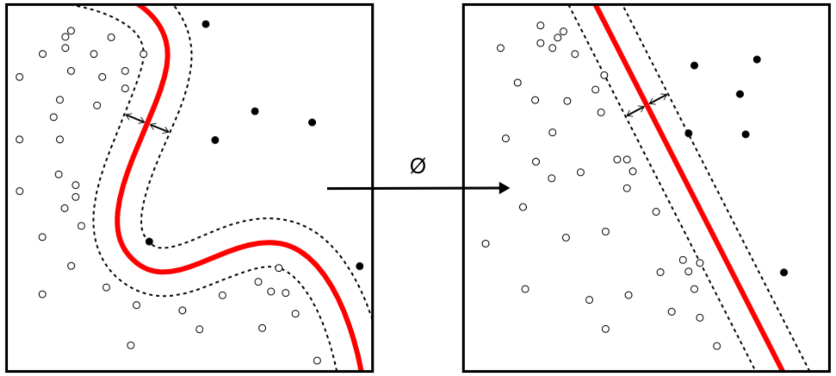
- Will not capture interactions unless explicitly included

How can we quantify risk?

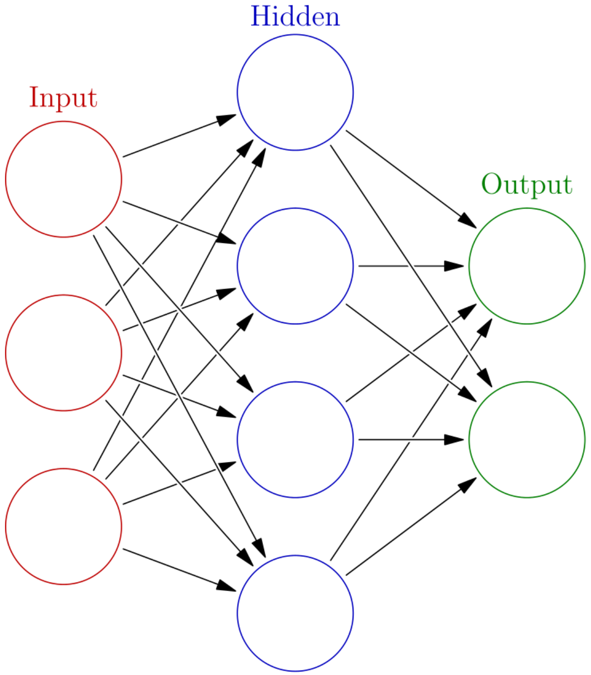
- Machine learning models
  - Refers to many types of algorithms



Random forest or gradient boosted decision trees



Support vector machines



Neural networks (or deep learning if multiple hidden layers)

Machine learning models

Advantages

- Often the most accurate
- Can capture non-linear relationships
- Can capture interactions between variables
- Neural networks can handle imaging and signal waveform data

Disadvantages

- Difficult to implement (and share)
- May “overfit” the training data if not carefully trained

How can risk models be operationalized to improve quality of care?

- At the point of care
    - Shared decision-making tools or decision aides
    - Clinical decision support alerts
    - Changing default selections for order sets

- Population health management
    - Run the model at a fixed interval
    - Use it to:
        - Find patients who need immediate attention (early warning systems)
        - Allocate resources efficiently by prioritizing the “sickest” or “modifiable risk” patients
        - Identify areas where observed outcomes worse than expected risk

Example of shared decision-making: askMUSIC

Michigan **U**rological **S**urgery **I**mprovement **C**ollaborative

A consortium of urologists and urology practices throughout the State that aims to improve the quality and cost-efficiency of urologic care provided to patients in Michigan

Our Goal: Make Michigan #1 in Urologic Care

Example of shared decision-making: askMUSIC

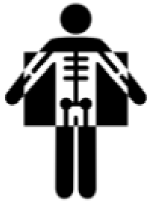
ask.musicurology.com

A digital platform designed to help patients and healthcare professionals make the best possible decisions about urological care

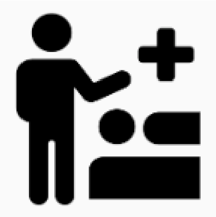
Prostate Cancer Resources for Doctors

Prostate Cancer Apps

Cancer Risk



Imaging Appropriateness



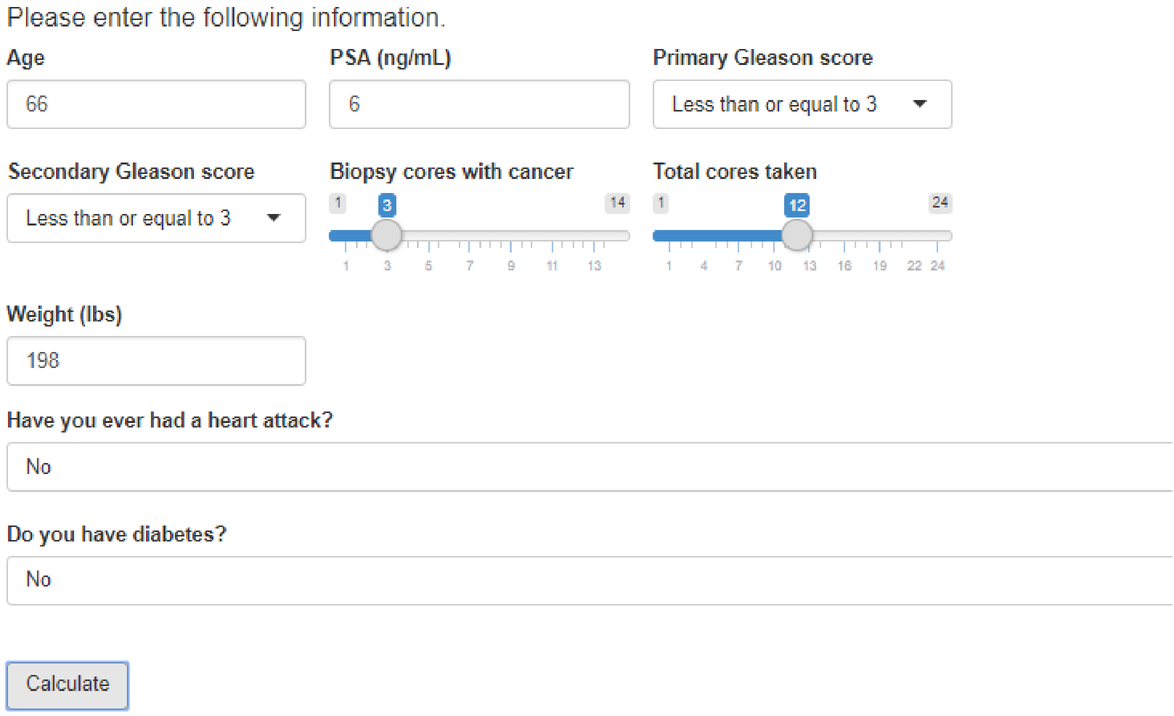
Treatment Options

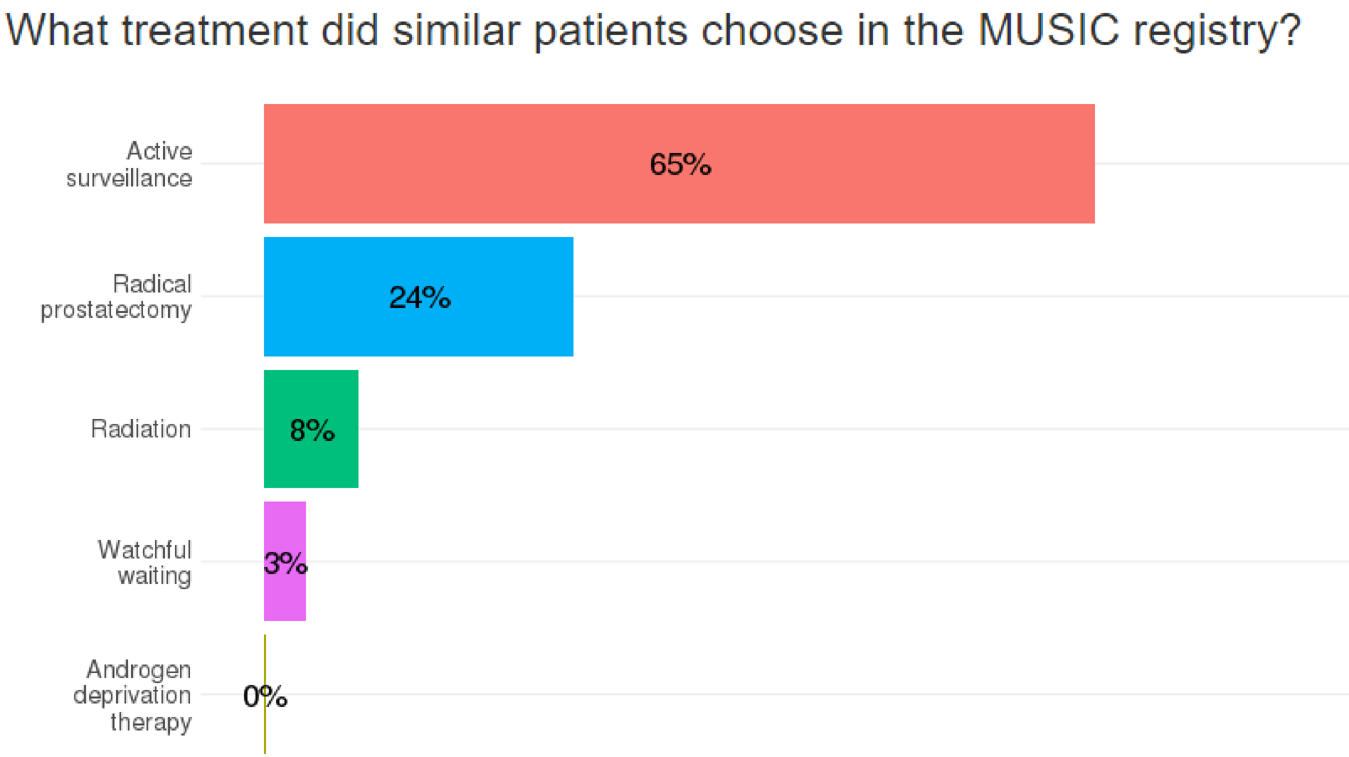
Radical Prostatectomy Pathologic Outcomes

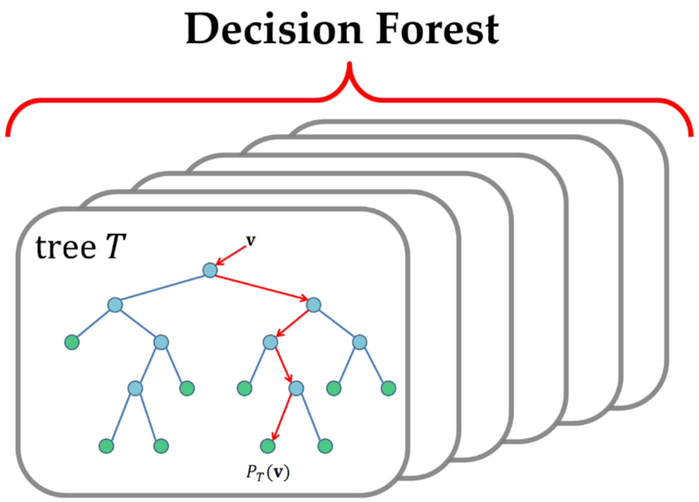
Click on a resource to learn more.

- What is my patient's risk of finding cancer if he undergoes a biopsy?
- Does my patient qualify for active surveillance? (Active Surveillance Roadmap)
- Should I consider imaging to evaluate for metastatic disease?
- What pathologic outcomes can I expect if my patient undergoes a radical prostatectomy?

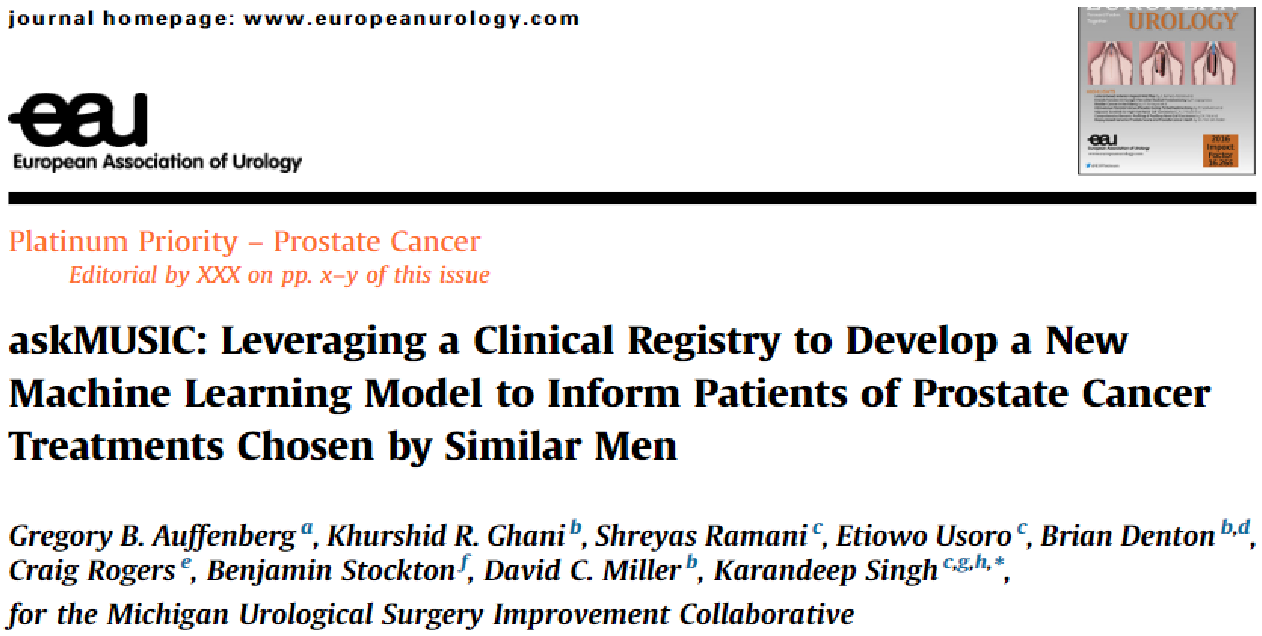
- Home
- For Patients
  - Prostate Cancer Resources for Patients
  - Kidney Stone Resources for Patients
- For Doctors
  - Prostate Cancer Resources for Doctors
- Resources
- Contact Us







Random forest model



journal homepage: www.europeanurology.com

European Association of Urology

Platinum Priority – Prostate Cancer

Editorial by XXX on pp. x–y of this issue

askMUSIC: Leveraging a Clinical Registry to Develop a New Machine Learning Model to Inform Patients of Prostate Cancer Treatments Chosen by Similar Men

Gregory B. Auffenberg [a], Khurshid R. Ghani [b], Shreyas Ramani [c], Etiowo Usoro [c], Brian Denton [b,d], Craig Rogers [e], Benjamin Stockton [f], David C. Miller [b], Karandeep Singh [c,g,h,], for the Michigan Urological Surgery Improvement Collaborative*

Example of clinical decision support

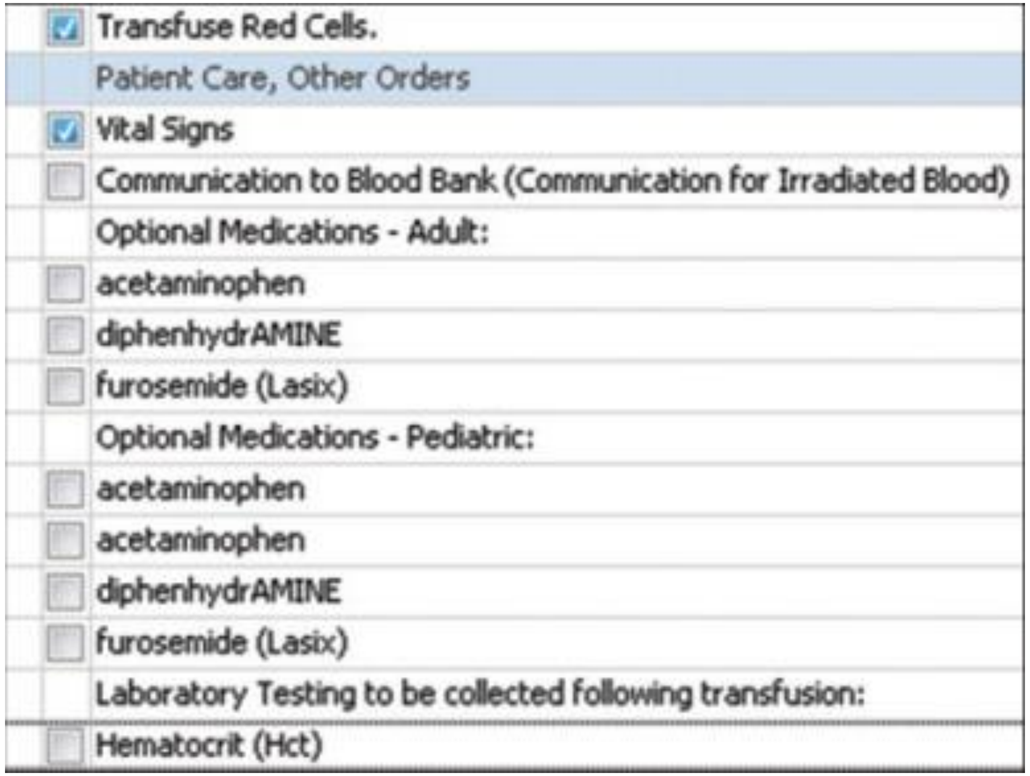
- Will not show an example here to minimize further alert fatigue

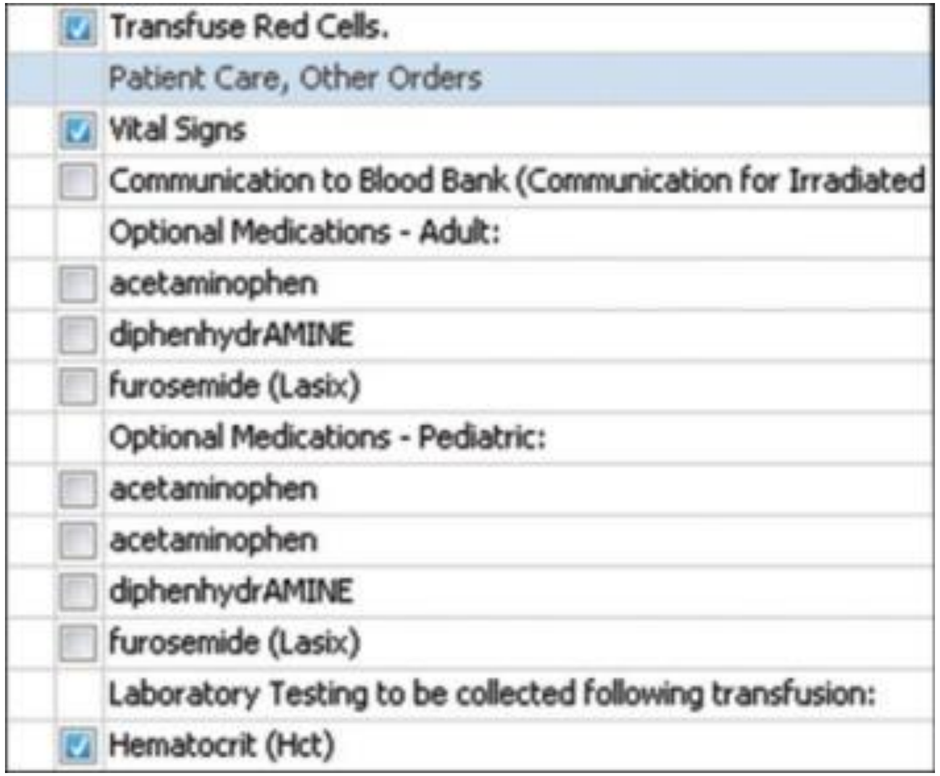
ALERT

Sorry, I know you are giving a talk right now but your version of Windows is slightly out of date. I went ahead and installed the latest version but now your computer needs to be restart. Press OK to restart.

OK **Cancel**

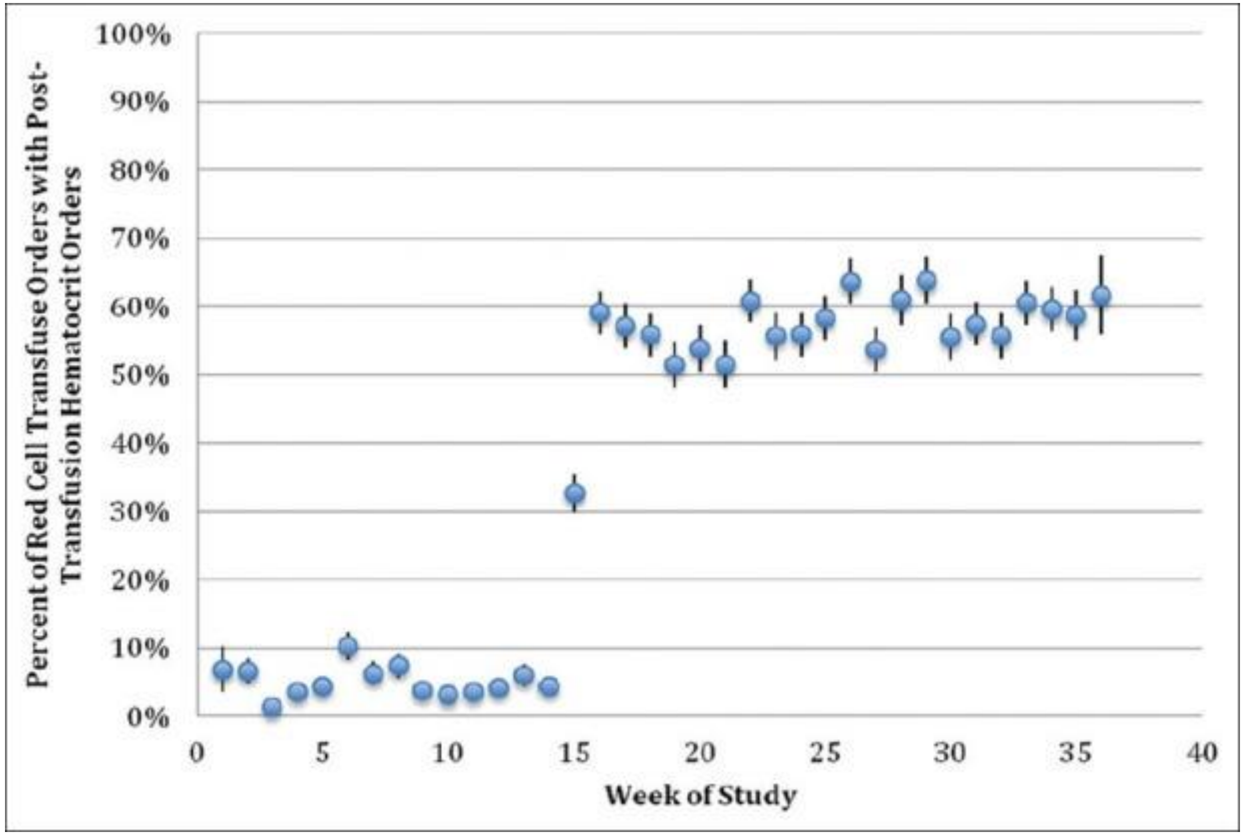
Changing default selections for order sets





Olson J, Hollenbeak C, Donaldson K, Abendroth T, Castellani W: Default settings of computerized physician order entry system order sets drive ordering habits. *J. Pathol. Inform.* 6: 16, 2015.

Changing default selections for order sets



Olson J, Hollenbeak C, Donaldson K, Abendroth T, Castellani W: Default settings of computerized physician order entry system order sets drive ordering habits. *J. Pathol. Inform.* 6: 16, 2015.

Example of population health mgmt: State Innovation Model

- Use a risk model to identify highest emergency department utilizers over the next 6 months in Livingston and Washtenaw Counties

- Run this model every 2 months

- Update the model with new data every 2 months

- Enroll highest risk patients in care coordination intervention

Example of population health mgmt: State Innovation Model

Model version 1

- UM data received: 9/8/2017
- No Trinity Health data
- Created 11 predictors
- Trained on 185,597 patients
    - RMSE: 0.364 visits
- **Run on 188,089 patients**
- Results shared on: 11/1/2017

Model version 2

- UM data received: 11/10/2017
- TH data received: 11/29/2017
- Created 56 predictors
- Trained on 303,514 patients
    - RMSE: 0.511 visits
- **Run on 311,962 patients**
- Results shared on: 1/25/2017

Model version 3

- Awaiting next round of data
- Geocoded addresses to latitude + longitude

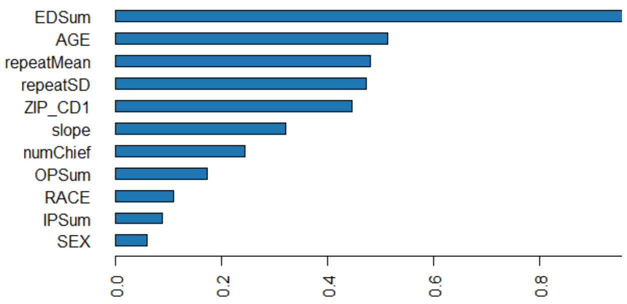
Model version 4

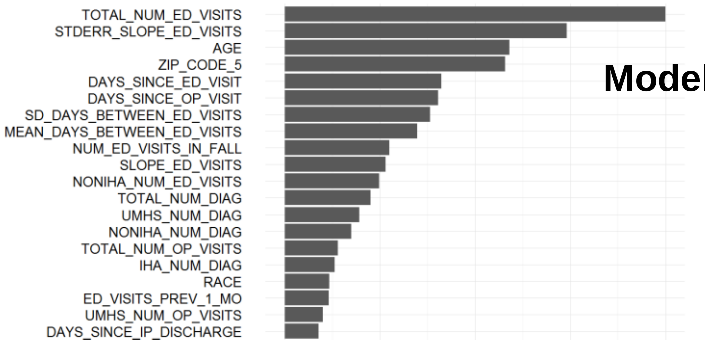
- Add billing codes

....

Model version 11

For context:
Total population of
2 counties =
538,683 people





Towards a Learning Health System to Reduce Emergency Department Visits at a Population Level

Elliott Brannon, MPH[1], Tianshi Wang[2], Jeremy Lapedis, DrPH , MSPH[3], Paul Valenstein, MD[4], Michael Klinkman, MD, MS[5], Ellen Bunting, MA[6], Alice Stanulis[6], Karandeep Singh, MD, MMSc[1,2,7]

[1]Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI; [2]School of Information, University of Michigan, Ann Arbor, MI; [3]Center for Healthcare Research & Transformation, Ann Arbor, MI; [4]Integrated Health Associates, Ann Arbor, MI; [5]Department of Family Medicine, University of Michigan, Ann Arbor, MI; [6]Michigan Data Collaborative, University of Michigan, Ann Arbor, MI; [7]Department of Internal Medicine, University of Michigan, Ann Arbor, MI

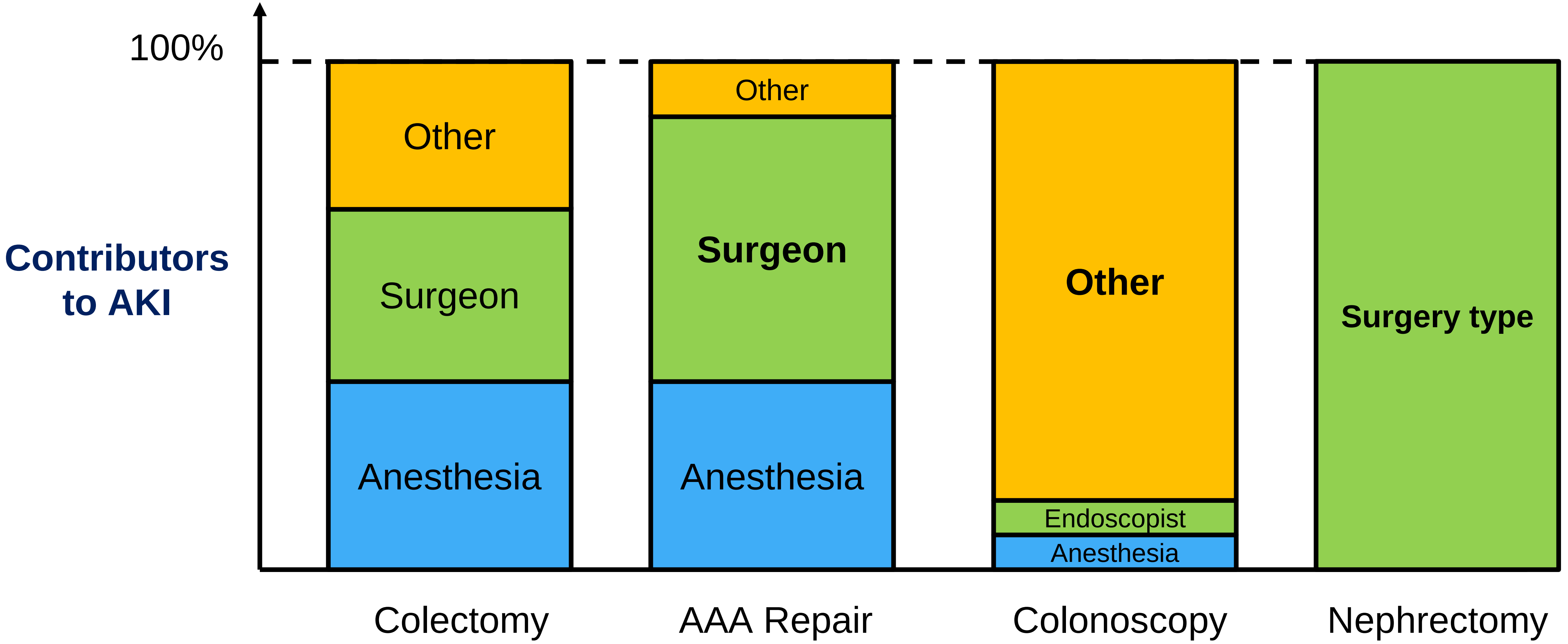
Example of population health mgmt: Identify areas of improvement

- Find where observed outcomes worse than expected risk
- Why did my patient develop AKI?

Provider	AKI Mechanism
Anesthesiologist	Hemodynamics, fluid management, diuretics, glycemic control
Surgeon / Proceduralist	Direct injury, physiologic insult, nephrotoxins
Other	Pre-existing CKD, comorbid conditions, lifestyle factors

Source: Michael Mathis, MD ASPIRE talk 7/20/2018

Attribution of AKI varies by case type



Source: Michael Mathis, MD ASPIRE talk 7/20/2018

Risk adjustment is the "usual" way to attribute risk

- What is it?
  - Method to more accurately assess performance, accounting for baseline risk

- Why do we need it?
  - Establishes basis for comparison across providers/institutions with varying baseline risk
  - Isolates component of outcome *attributable to the anesthesiologist*

- How does it work?
  - Compares a provider's <u>*observed*</u> performance to what was <u>*expected*</u>

Source: Michael Mathis, MD ASPIRE talk 7/20/2018

How does risk adjustment work?

- Using Risk Adjustment:

	Case Type Performed	# Cases *observed* to have AKI
Provider A	100 AAA repairs	20
Provider B	100 colonoscopies	3

  - Incidence of AKI:
    - Provider A → 20%
    - **Provider B → 3%**

Source: Michael Mathis, MD ASPIRE talk 7/20/2018

How does risk adjustment work?

- Using Risk Adjustment:

	Case Type Performed	# Cases *observed* to have AKI	# Cases *expected* to have AKI
Provider A	100 AAA repairs	20	50
Provider B	100 colonoscopies	3	2

  - Incidence of AKI:
    - Provider A → 20%
    - **Provider B → 3%**
  - Comparing provider's observed performance to what was expected:

Source: Michael Mathis, MD ASPIRE talk 7/20/2018

How does risk adjustment work?

- Using Risk Adjustment:

	Case Type Performed	# Cases *observed* to have AKI	# Cases *expected* to have AKI	Observed / Expected (O/E) Ratio
Provider A	100 AAA repairs	20	50	20/50 = 0.40
Provider B	100 colonoscopies	3	2	3/2 = 1.50

  - Incidence of AKI:
    - Provider A → 20%
    - **Provider B → 3%**
  - Comparing provider's observed performance to what was expected:
    - $\boxed{\textbf{Provider A} \rightarrow \textbf{0.40}}$
    - Provider B → 1.50

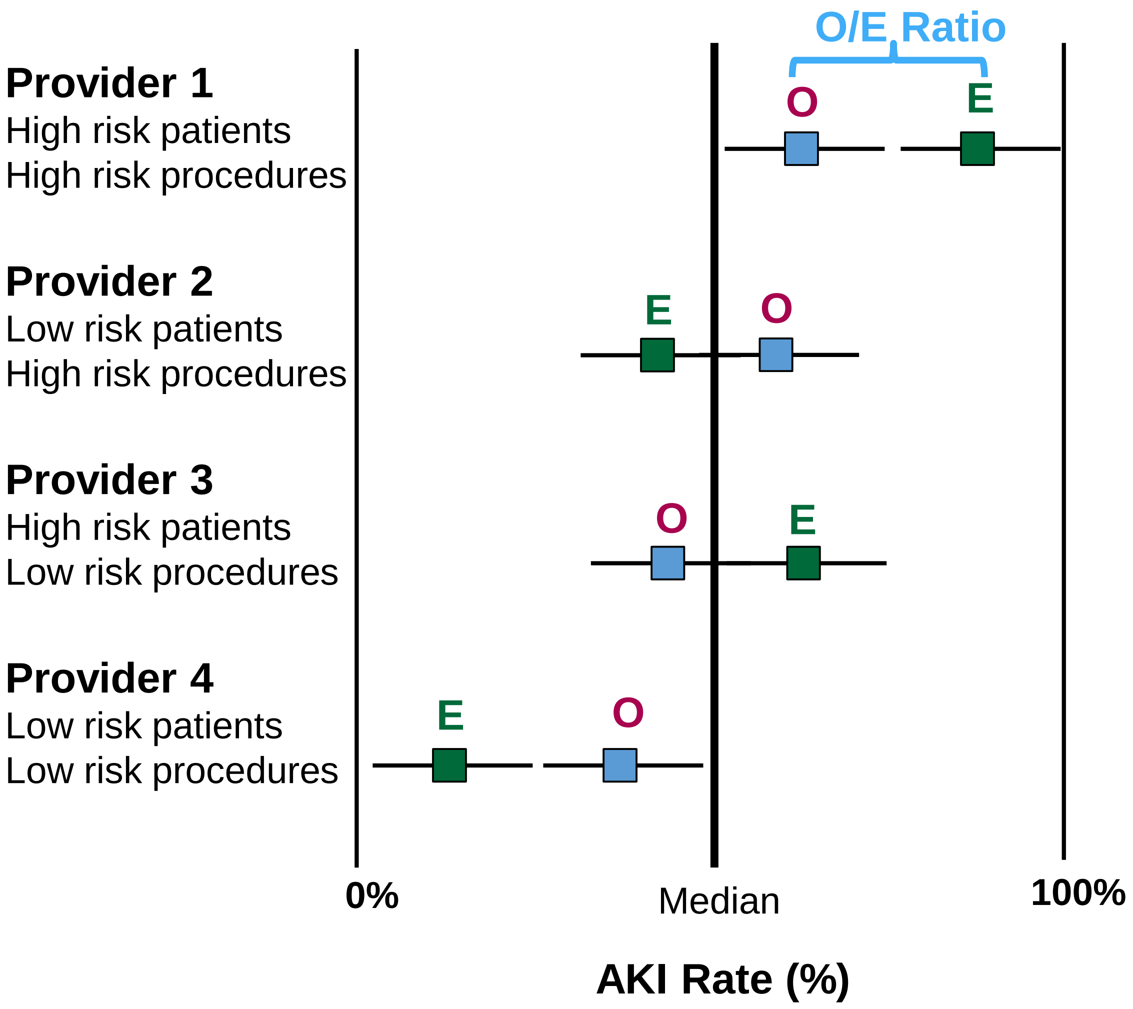
Source: Michael Mathis, MD ASPIRE talk 7/20/2018

How is “expected” risk calculated?

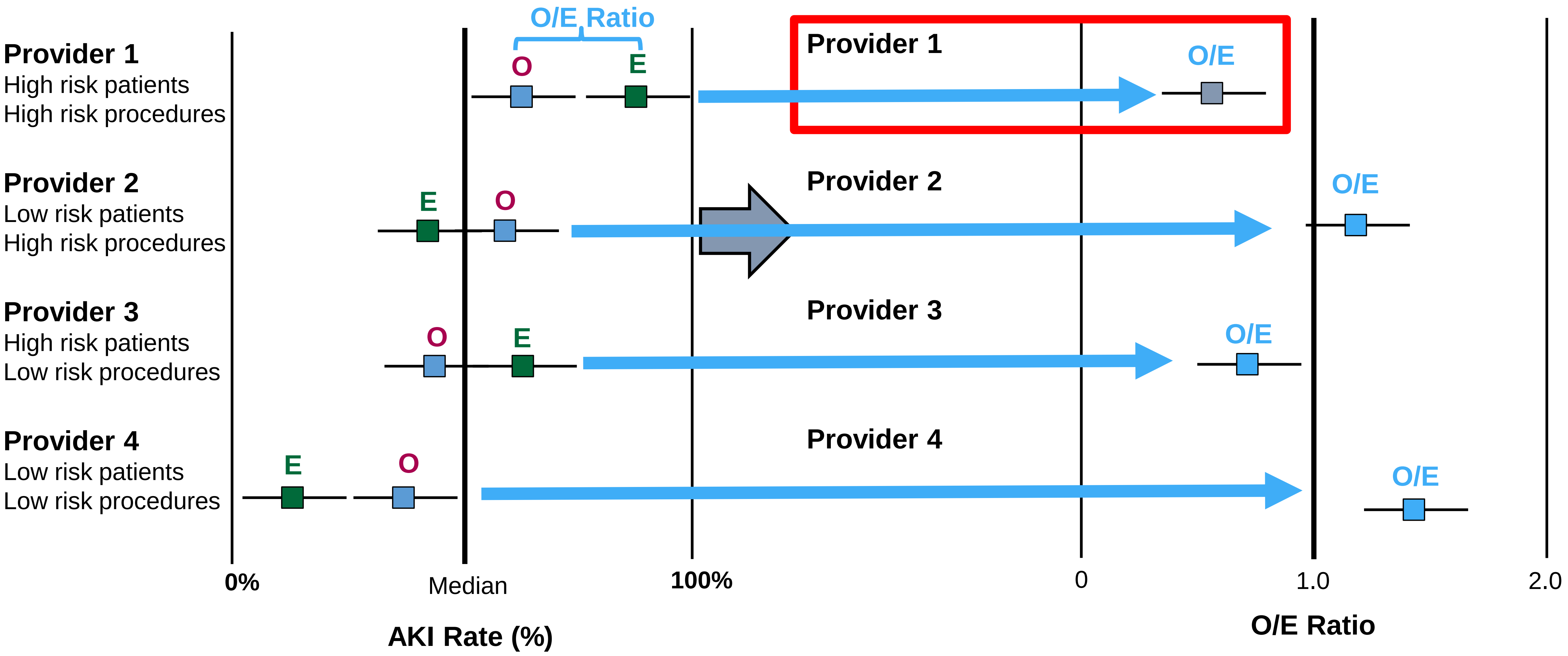
- **Use variables to develop a risk model considering:**
  - **Patient characteristics**
    - Demographics: age, gender, BMI
    - ASA status
    - Comorbidities: renal insufficiency, HTN, HF, diabetes, CAD, liver disease, etc.
    - Labs: hemoglobin, creatinine
  - **Surgical characteristics**
    - Procedure type (anesthesia CPT code)
    - Emergent / elective
  - **Center characteristics**
    - Type of hospital

Source: Michael Mathis, MD ASPIRE talk 7/20/2018

Observed/expected ratio is what matters



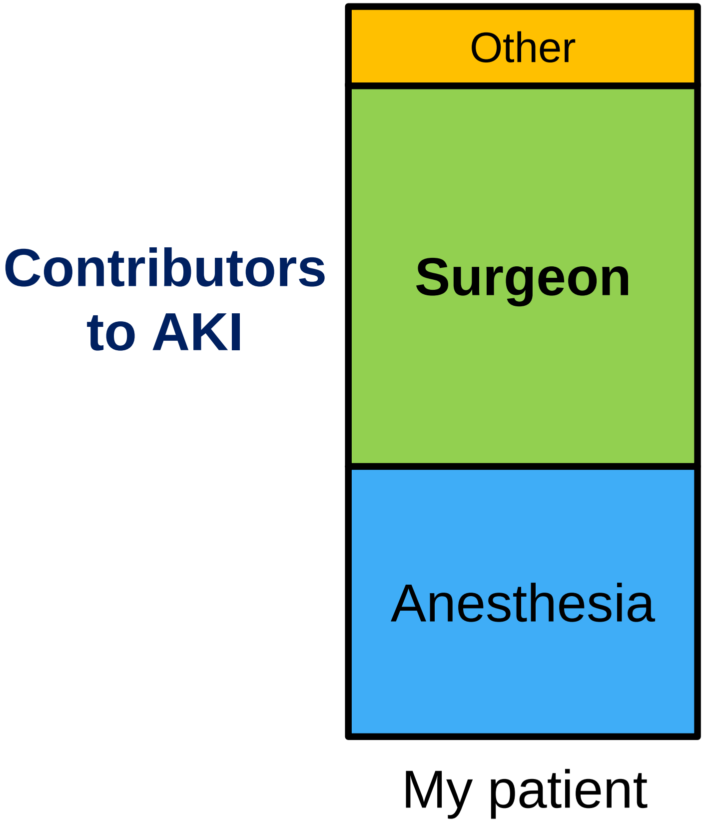
Observed/expected ratio is what matters



Why not settle for risk adjustment?

- Risk adjustment is useful at an anesthesiologist level
- Helpful for identifying “underperforming” anesthesiologists
- **Not helpful for attributing risk at a case level**
- If you want to identify which AKI outcomes were preventable from the perspective of the anesthesiologist
  - **Cannot do it with risk adjustment alone**

Why did my patient develop AKI?



- To answer this question, we need to know the contributions of each factor to the patient's risk

← **Can we generate this for each surgery?**

Yes, we can.

If we know which cases have a large anesthesiologist contribution to AKI risk,

We can focus our QI efforts and learning on those patients.

How do we attribute risk at a case level?

- First, we need to measure AKI risk factors that are cleanly attributable to a provider

- Ideally, some should be modifiable

Variables

- Anesthesiologist
  - fluids administered, intraoperative BP, electrolytes (modifiable)
- Surgeon
  - skill, prior rate of AKI, whether an artery was lacerated (operative report)
- Patient factors
  - Comorbidities, presence and severity of CKD
- Other
  - surgery type

How do we attribute risk at a case level?

- Next, we need to measure contribution of each variable to the case risk

- In a linear model, this is straightforward
  - Let’s say AKI risk is a number
  - If AKI risk = 0.5 x anesthesiologist’s years of experience + 0.1 x surgeon’s years of experience, then:
    - For anesthesiologist with 10 years experience and surgeon with 10 years of experience
    - Anesthesiologist’s contribution is 5/(5+1) or 83%, surgeon’s contribution 17%

How do we attribute risk at a case level?

- Case-level attribution is more difficult for machine learning models

- The contribution of each variable is not fixed – it depends on the other variables for that case

- Why does this matter?
  - Machine learning models often outperform regression models for modeling risk
  - What good are they if we can’t attribute risk for individual cases?

Machine learning can attribute risk using 2 recently described methods

- These methods generate “model explanations”
- Shapley values
- Locally interpretable model explanations (LIME)

Machine learning can attribute risk using 2 recently described methods

Shapley values

- A way to calculate how much to pay each person when people are working together to earn money
- Originally described in 1953 by Lloyd Shapley
- Shapley won the Nobel prize for economics in 2012
- Shapley values rediscovered by machine learning community in 2013
  - Variables are “working together” to generate prediction

Knowl Inf Syst (2014) 41:647–665
DOI 10.1007/s10115-013-0679-x

REGULAR PAPER

Explaining prediction models and individual predictions with feature contributions

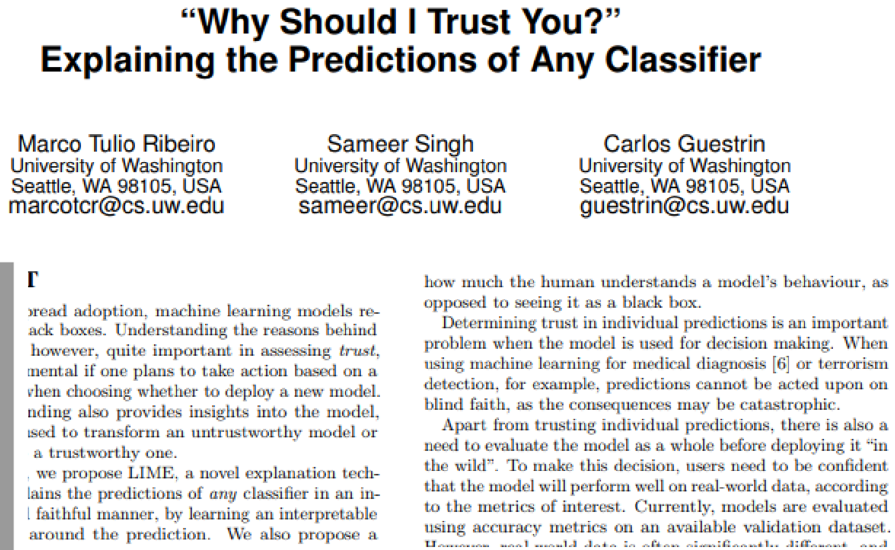
Erik Štrumbelj · Igor Kononenko

Received: 12 November 2012 / Revised: 2 August 2013 / Accepted: 17 August 2013 /
Published online: 30 August 2013
© Springer-Verlag London 2013

Machine learning can attribute risk using 2 recently described methods

LIME

- A less computationally intensive way of generating concise “model explanations” for an individual data point in a dataset



“Why Should I Trust You?”
Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

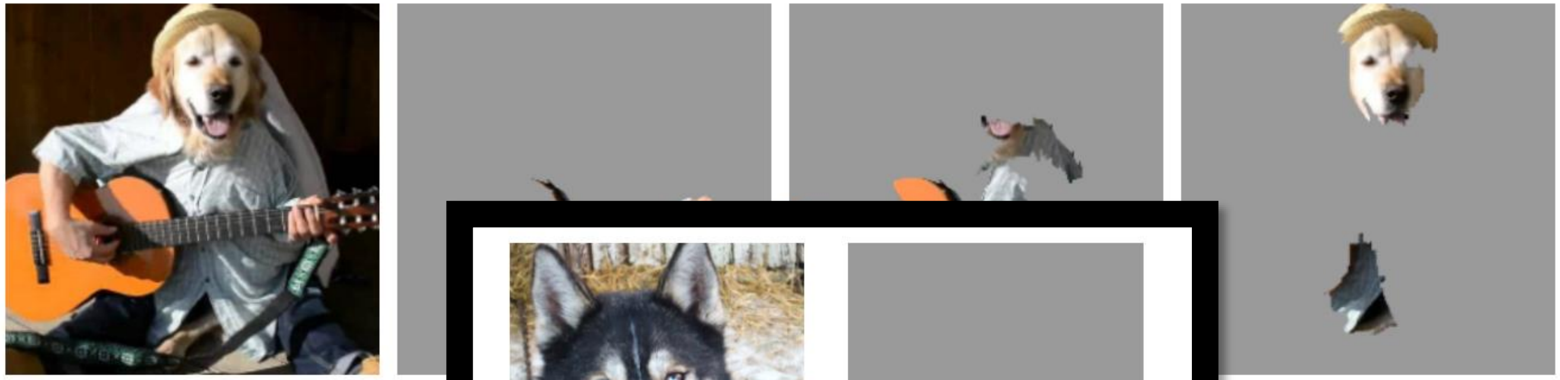
pread adoption, machine learning models re- ack boxes. Understanding the reasons behind however, quite important in assessing *trust*, mental if one plans to take action based on a when choosing whether to deploy a new model. nding also provides insights into the model, used to transform an untrustworthy model or a trustworthy one.

, we propose LIME, a novel explanation tech- lains the predictions of *any* classifier in an in- faithful manner, by learning an interpretable around the prediction. We also propose a

how much the human understands a model’s behaviour, as opposed to seeing it as a black box.

Determining trust in individual predictions is an important problem when the model is used for decision making. When using machine learning for medical diagnosis [6] or terrorism detection, for example, predictions cannot be acted upon on blind faith, as the consequences may be catastrophic.

Apart from trusting individual predictions, there is also a need to evaluate the model as a whole before deploying it “in the wild”. To make this decision, users need to be confident that the model will perform well on real-world data, according to the metrics of interest. Currently, models are evaluated using accuracy metrics on an available validation dataset.



(a) Original Image (b) … d) Explaining *Labrador*

Figure 4: Explaining an image … eural network. The top 3 classes predicted are “Electric … abrador” ($p = 0.21$)

(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model’s prediction in the “Husky vs Wolf” task.

Shapley and LIME in action

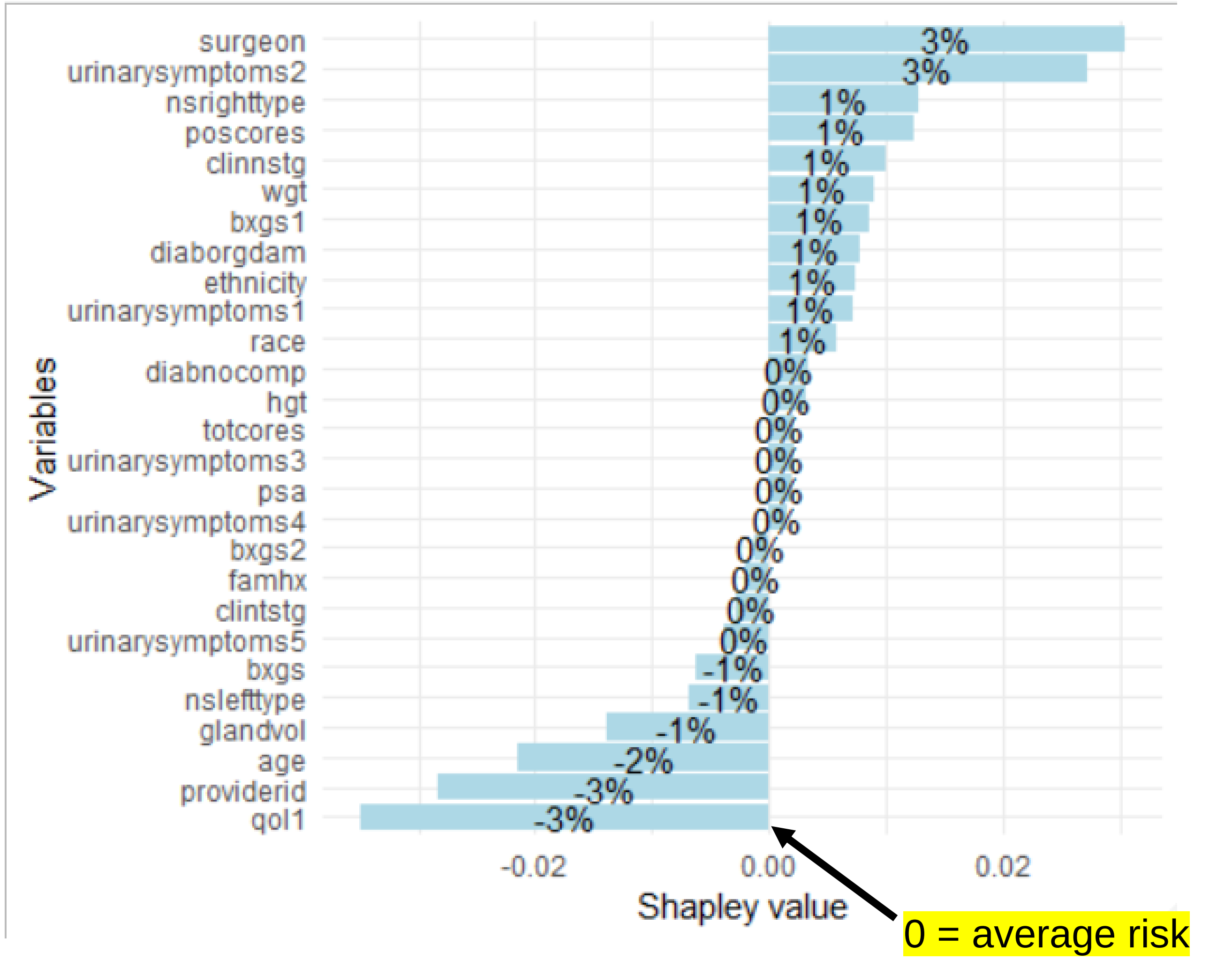
- Let’s use these methods on MUSIC data
- We will fit a random forest model to predict 3-month continence following prostatectomy
- For a handful of cases, we will take a look at the contributions of individual predictors
- Think about whether the cases are worth investigating (from a surgeon’s perspective)

Example: Why did my patient develop incontinence following prostatectomy?

- *Average risk of incontinence at 3 months for all patients is 34%*

- Model predicts 37% chance of incontinence
- Patient is **continent** at 3 months

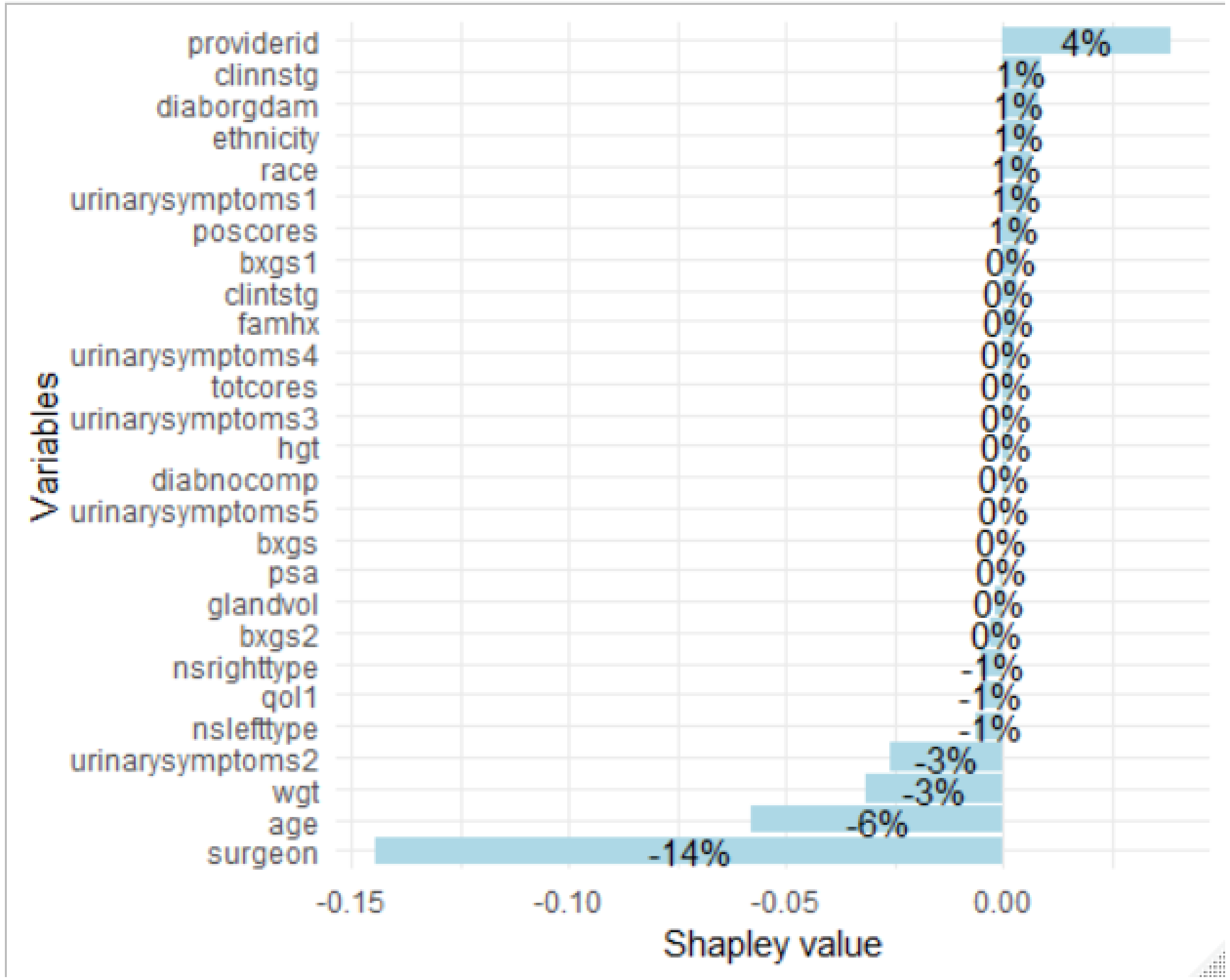
- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

- Model predicts 16% chance of incontinence
- Patient is **continent** at 3 months

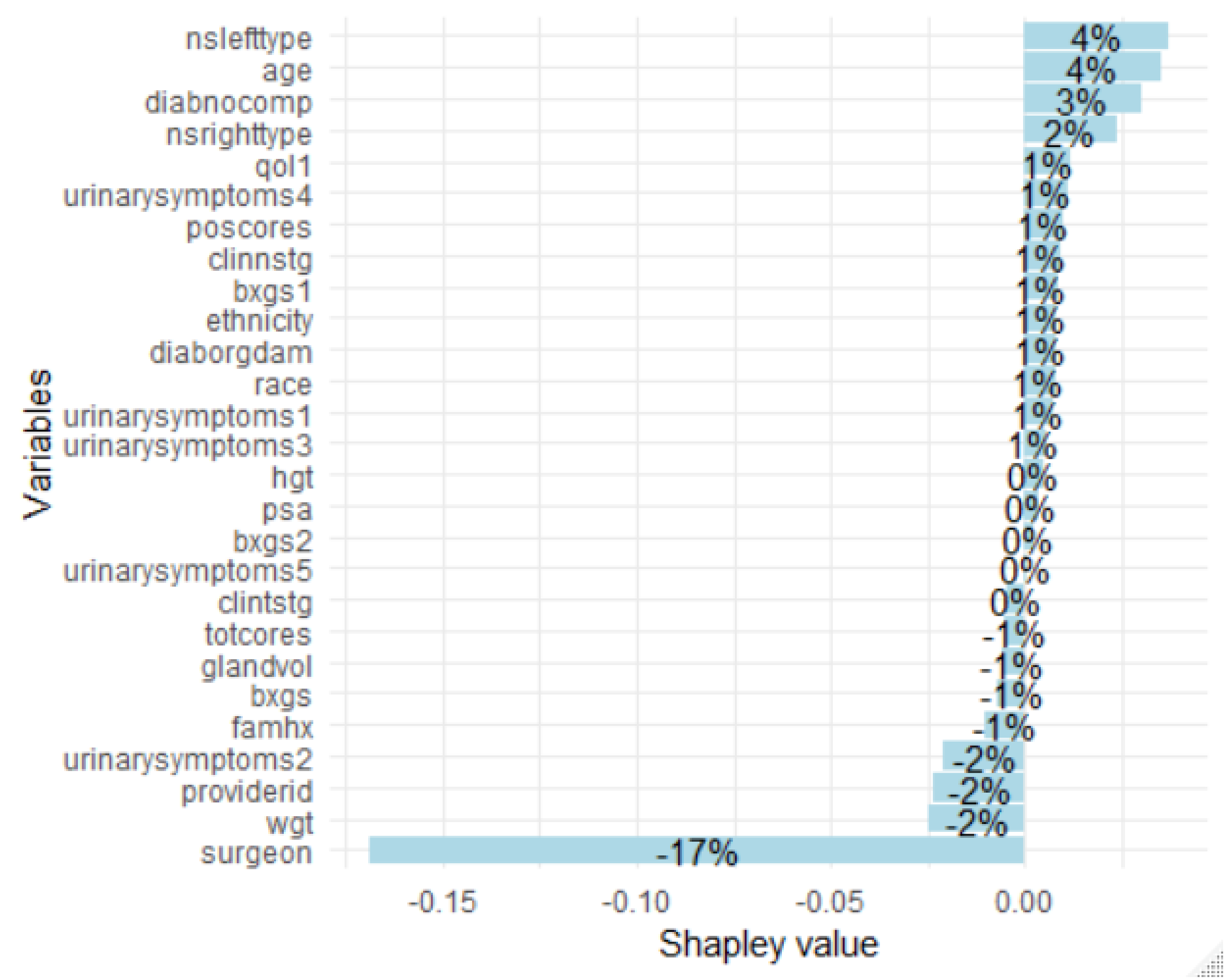
- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

- Model predicts 30% chance of incontinence
- Patient is **incontinent** at 3 months

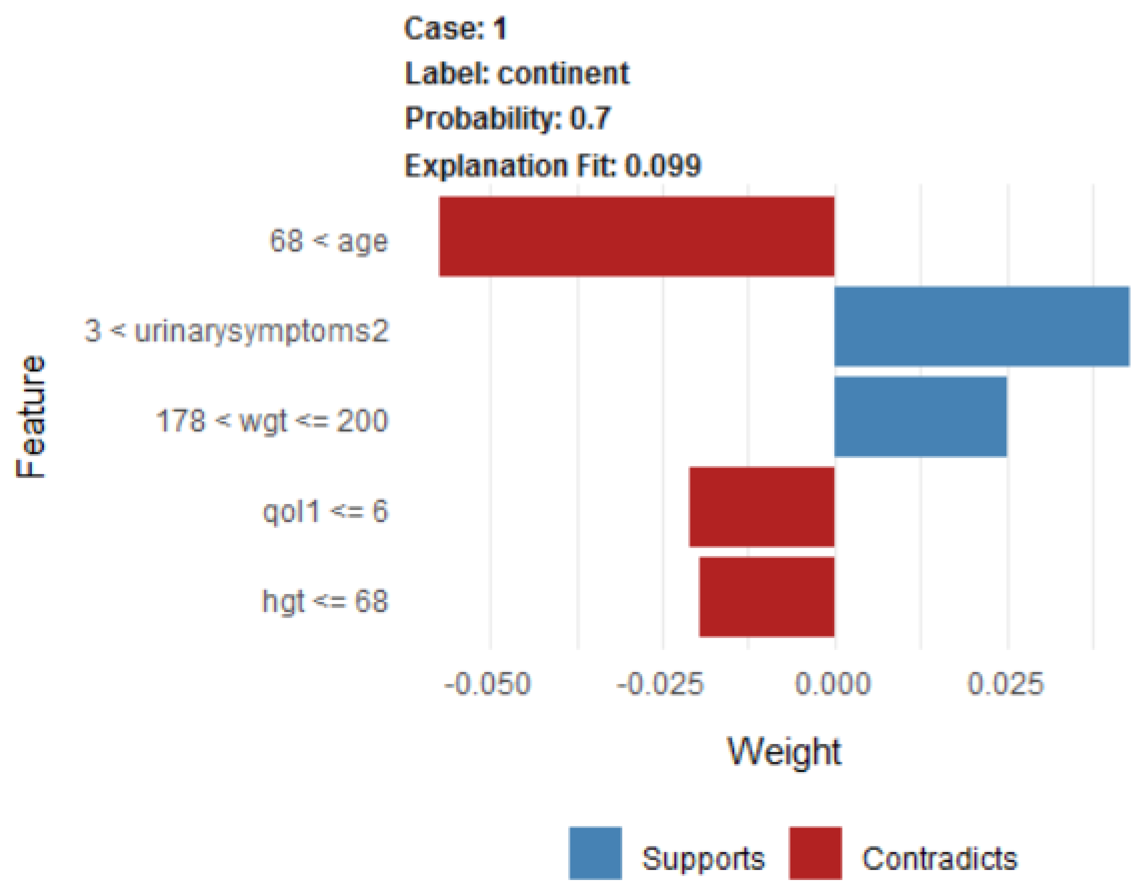
- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

- Model predicts 30% chance of incontinence
- Patient is **incontinent** at 3 months

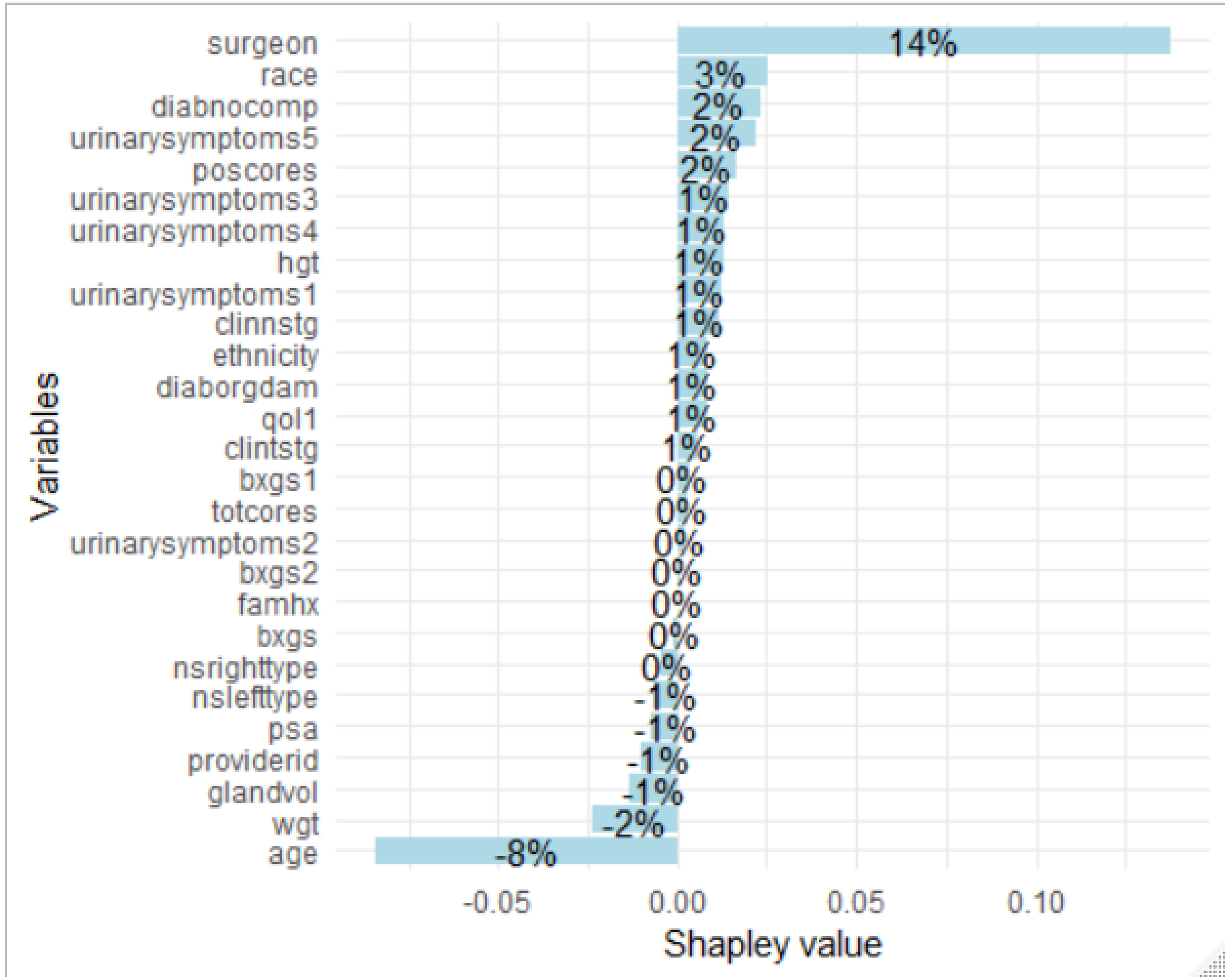
- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

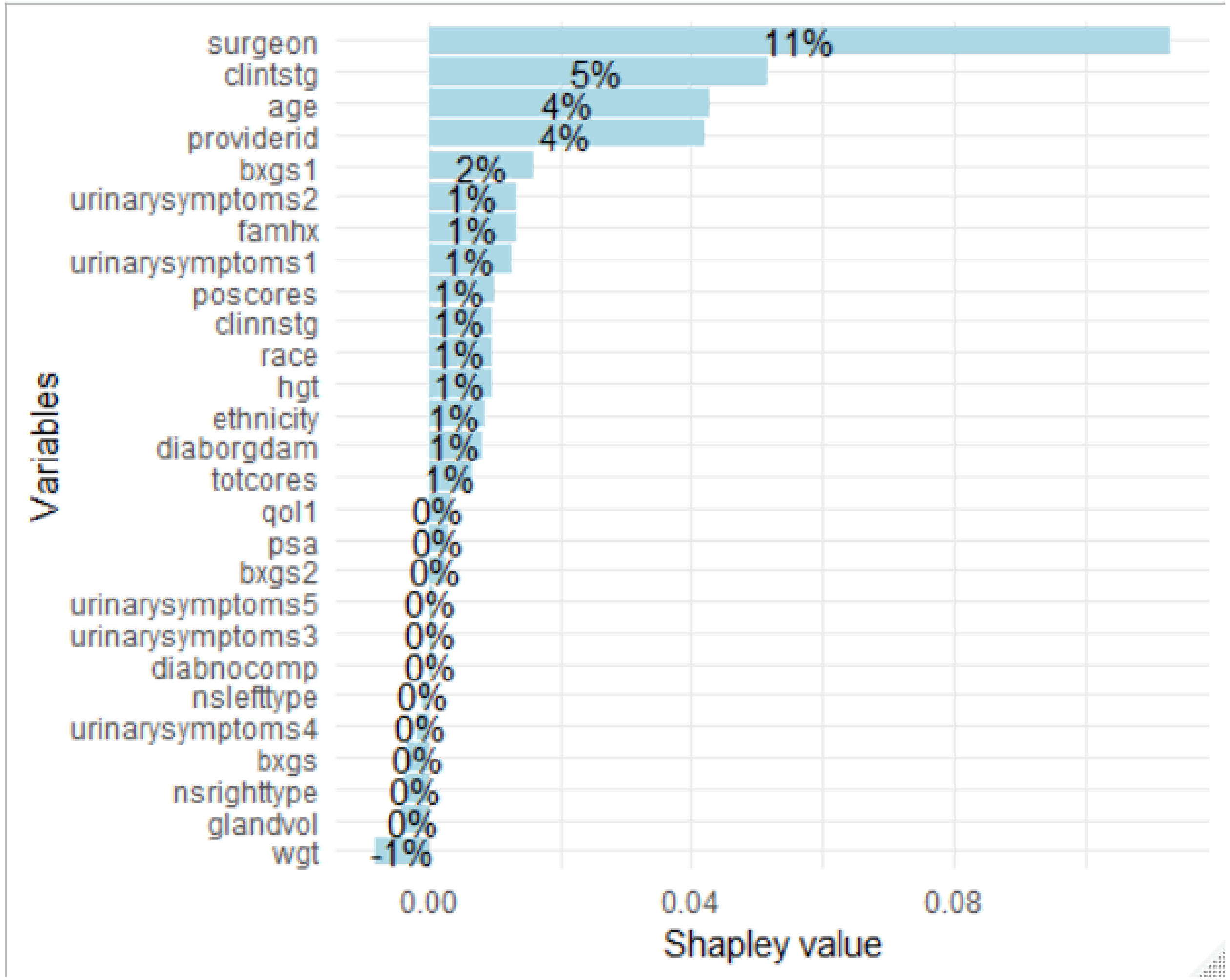
- Model predicts 52% chance of incontinence
- Patient is **incontinent** at 3 months

- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

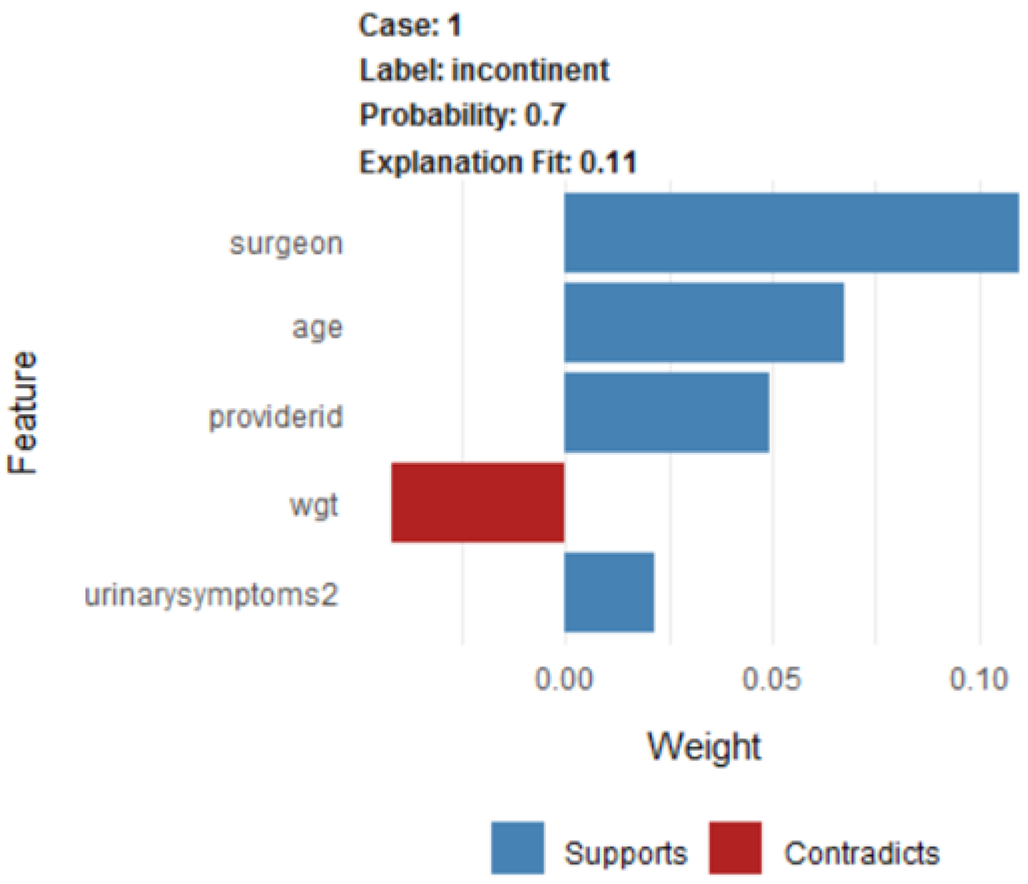
- Model predicts 70% chance of incontinence
- Patient is **incontinent** at 3 months
- Is this worth investigating?



Example: Why did my patient develop incontinence following prostatectomy?

- Model predicts 70% chance of incontinence
- Patient is **incontinent** at 3 months

- Is this worth investigating?



Cases are worth investigating when…

- Model gets it wrong
- Surgeon or practice is primary contributing factor to high risk

One last word about Shapley vs. LIME

Shapley is gaining in popularity

- In the US finance sector and in the UK more broadly, the law requires that any model with financial implications has to be interpretable

Shapley values have an "additive property"

- If you can determine which variables are anesthesiologist-related, you can add up their contributions to get the anesthesiologist contribution
- Cannot do this with LIME

Where do we go from here?

- There are machine learning models peri-operative AKI described in the literature but not openly available

- I am currently participating in a VA contract where we are developing an in-hospital AKI early warning system (PI: Michael Heung, MD)

RESEARCH ARTICLE

Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications

Paul Thottakkara[1,3], Tezcan Ozrazgat-Baslanti[1], Bradley B. Hupf[1], Parisa Rashidi[2], Panos Pardalos[3], Petar Momcilovic[3], Azra Bihorac[1]*

1 Department of Anesthesiology, College of Medicine, University of Florida, Gainesville, Florida, United States of America, **2** Biomedical Engineering Department, University of Florida, Gainesville, Florida, United States of America, **3** Industrial and Systems Engineering, University of Florida, Gainesville, Florida, United States of America

* abihorac@anest.ufl.edu

Model	Acute Kidney Injury		
	Accuracy (95% CI)	AUC (95% CI)	PPV (95% CI)
Logistic Regression Model	0.752 (0.746,0.758)	0.824 (0.818,0.828)[b]	0.725 (0.714,0.737)
GAMs	0.756 (0.751,0.761)	0.827 (0.821,0.832)[a]	0.719 (0.706,0.729)
Naïve Bayes Model	0.744 (0.738,0.749)	0.797 (0.791,0.803)[a,b]	0.545 (0.534,0.558)
SVM	0.767 (0.757,0.774)	0.819 (0.811,0.828)[a,b]	0.662 (0.648,0.676)
After feature selection with LASSO			
Logistic Regression Model	0.753 (0.747,0.757)	0.824 (0.818,0.830)[b]	0.726 (0.714,0.738)
GAMs	0.757 (0.752,0.762)	0.828 (0.822,0.833)[a]	0.72 (0.706,0.732)
Naïve Bayes Model	0.744 (0.737,0.750)	0.797 (0.789,0.804)[a,b]	0.545 (0.533,0.556)
SVM	0.767 (0.759,0.774)	0.82 (0.812,0.829)[a,b]	0.665 (0.646,0.685)
After feature extraction with 5 principal components			
Logistic Regression Model	0.774 (0.769,0.781)	0.853 (0.849,0.859)[a,b]	0.758 (0.746,0.767)
GAMs	0.773 (0.768,0.777)	0.858 (0.853,0.862)[a,b]	0.784 (0.771,0.793)
Naïve Bayes Model	0.741 (0.735,0.747)	0.819 (0.814,0.826)[a,b]	0.666 (0.651,0.677)
SVM	0.777 (0.767,0.782)	0.857 (0.850,0.862)[a,b]	0.735 (0.725,0.750)

Where do we go from here?

- Need to define which variables are attributable to the anesthesiologist versus other factors

- Should operative reports be included in the model?
  - Either extracted variables or the whole report as a sequence of text
  - Hypotension wouldn’t necessarily be attributed to the anesthesiologist in the setting of major surgical bleeding
  - Potential for collaboration with surgical CQIs

Thank you!

Feel free to contact me:

Karandeep Singh, MD, MMSc

kdpsingh@umich.edu

Twitter: @kdpsinghlab

LinkedIn: https://www.linkedin.com/in/kdpsingh