www.ispor.org

ISPOR

Improving healthcare decisions

How to Apply Machine Learning to Health Economics and Outcomes Research:
Findings from the ISPOR Machine Learning Task Force

Bill Padula, PhD
Bill Crown, PhD
Blythe Adamson PhD
Dave Vanness, PhD

ISPOR Annual 2022

Disclosures

- Bill Padula declares consulting and equity for Monument Analytics
- Blythe Adamson declares consulting for Infectious Economics
- Dave Vanness declares consulting for Apriori Bayesian Consulting, LLC

Task Force Co-Chairs: Bill^2

Bill Padula, PhD
University of Southern California

Bill Crown, PhD
Brandeis University

Task Force Members




Blythe Adamson, PhD
Flatiron Health

Dave Vanness, PhD
Penn State University

Noemi Kreif, PhD
University of York

Maarten IJzerman
University of Melbourne

Task Force Members

Atul Butte, MD, PhD
University of California
San Francisco

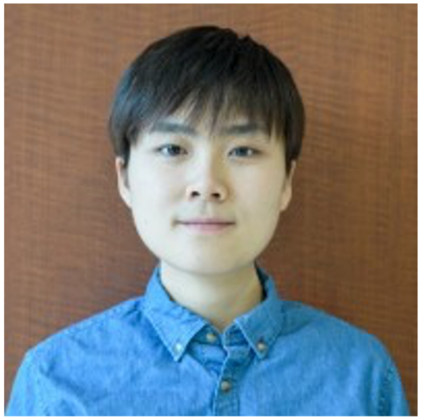
Pall Johnson, PhD
NICE, UK

Federico Felizzi
Novartis

Juan David Rueda, MD PhD
Astra Zeneca

Task Force Collaborators

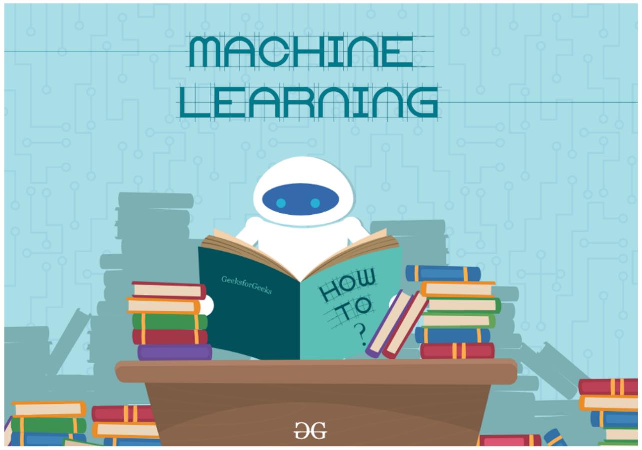
Suzanne Belinson, PhD
Tempus



Hanke Zheng, MS
University of Southern California

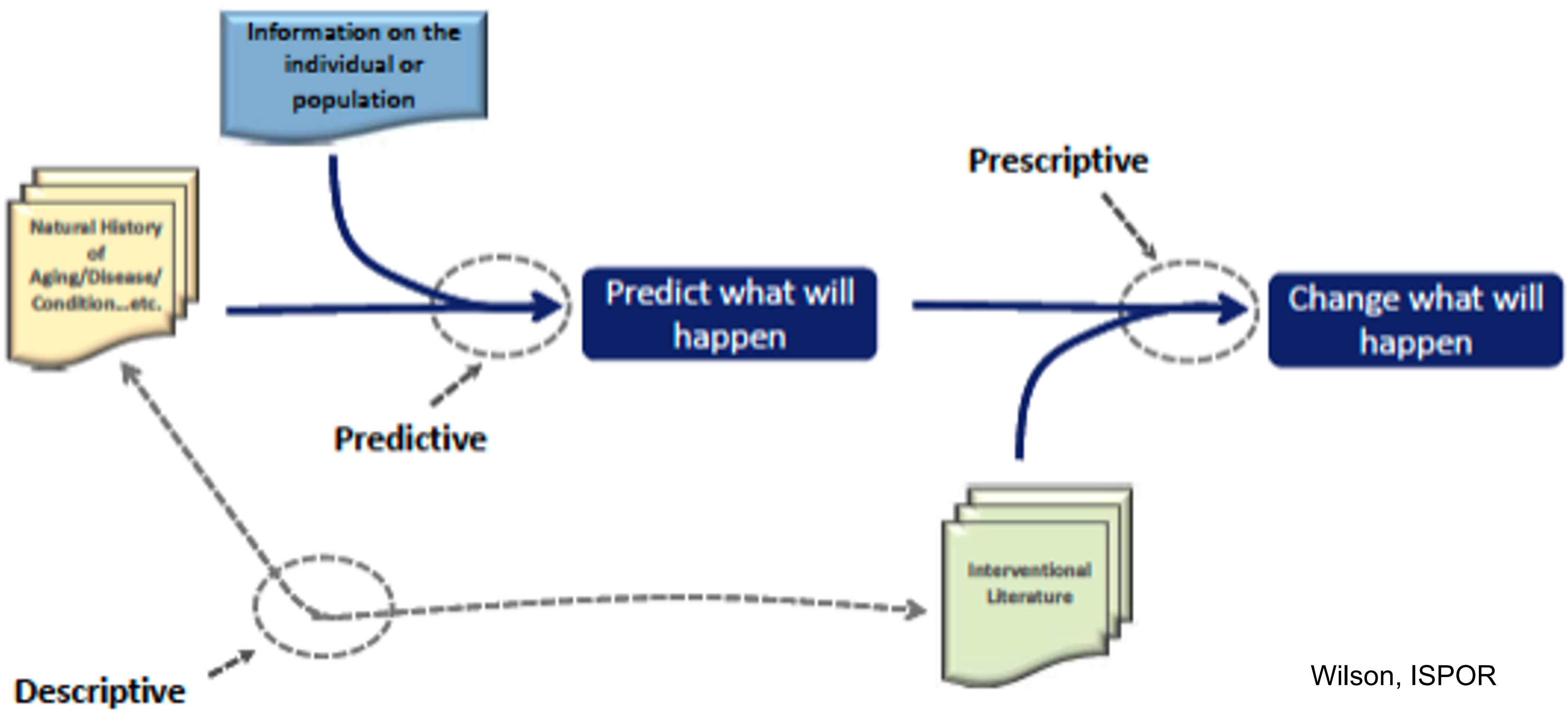
Elizabeth Molsen, RN
ISPOR

What is Machine Learning?



- A large family of mathematical and statistical methods for classification and prediction
- Used to automate analytical process with high volumes of information
- Two general domains
  - **Unsupervised** methods are focused mainly on dimension reduction and learning the underlying structure of the data
  - **Supervised** methods require the specification of an outcome variable and are focused on prediction or classification
- Enormous potential to couple ML methods with big data to undertake classification and clustering tasks to a higher degree of accuracy

Potential for Machine Learning (ML) in HEOR



Wilson, ISPOR

Task Force Objectives

To establish guidance for emerging good practices in the application of machine learning (ML) methodology to traditional ISPOR methods, including economic evaluation, decision sciences and outcomes research in order to improve the value of healthcare delivery.

- Introduce ML methods and their value in conducting research on health economics, as well as patient- and system-level outcomes research
- Describe problems for which machine learning methods are appropriate

Task Force Process

The task force began meeting in November, 2019

(in-person, ISPOR EU)

→ Virtual Meetings in 2020-2021

→ Broke into 6 SWGs

- One lead for each SWG
- 2 editors for each SWG

→ Bill Crown and Bill Padula edited each working groups section

→ Compile full report

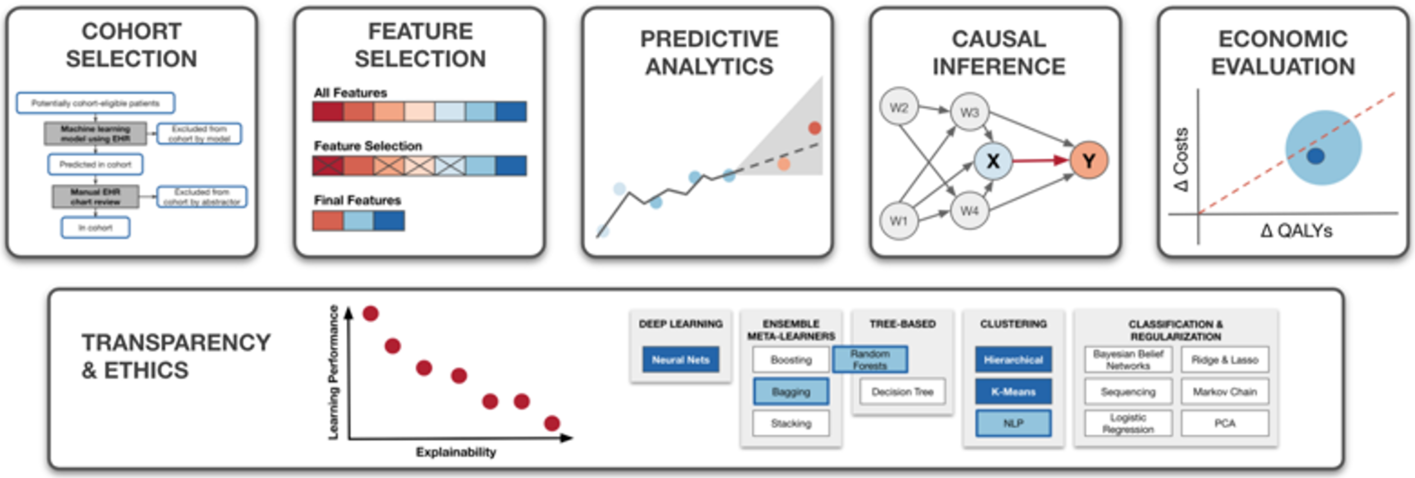
→ 2 rounds of peer-review with leading experts in HEOR and computer science

→ June/July 2022 Publication in

Value in Health

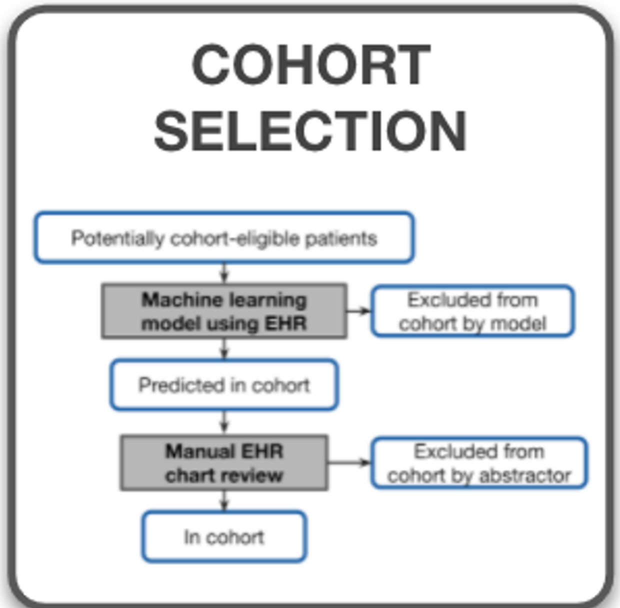
Task Force SWGs: Overview

- Cohort Selection
- Feature Selection
- Predictive Analytics
- Causal Inference
- Economic Evaluation
- Transparency and Explainability

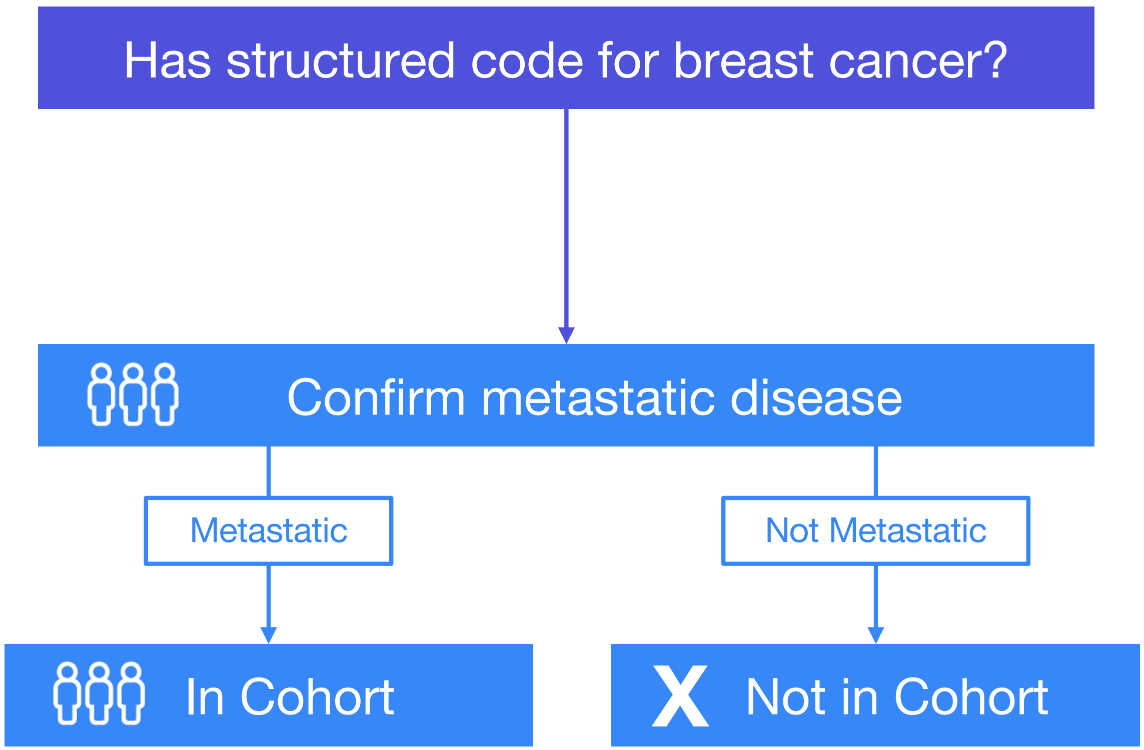


Cohort Selection

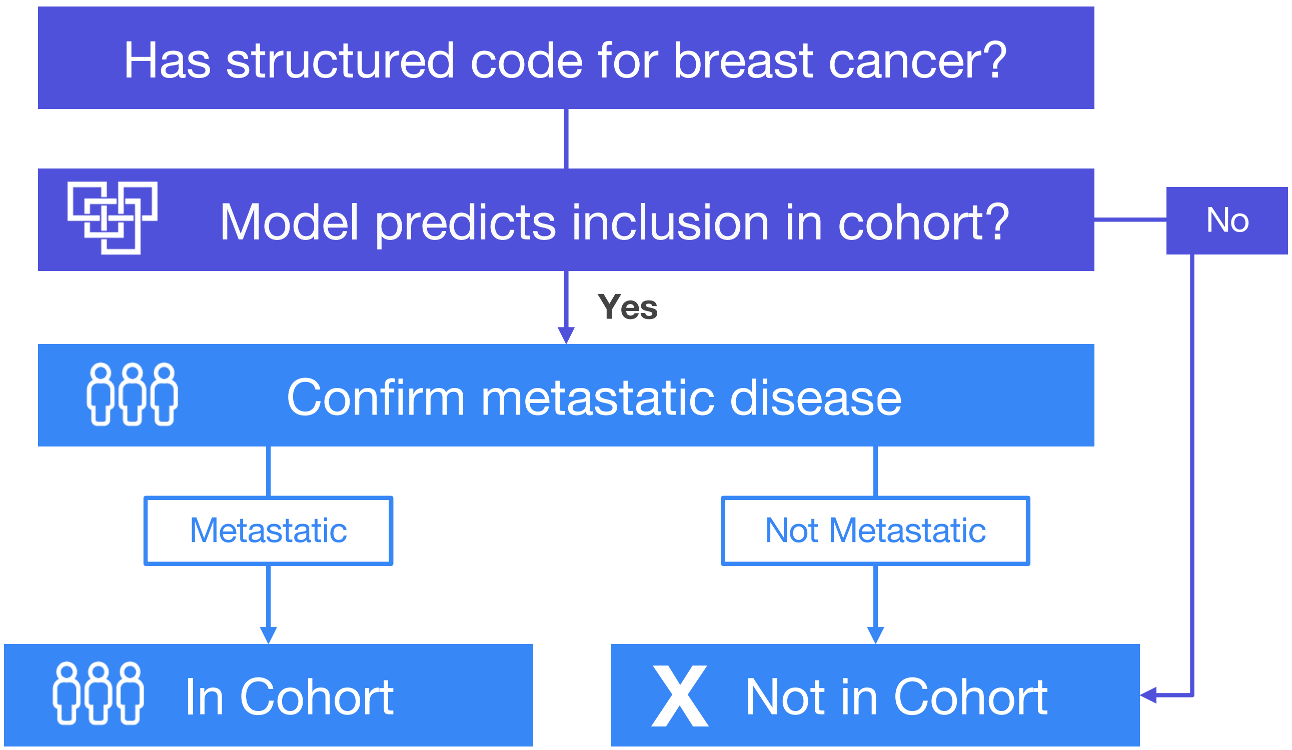
Blythe Adamson, PhD, MPH
Flatiron Health



Cohort selection for metastatic breast cancer

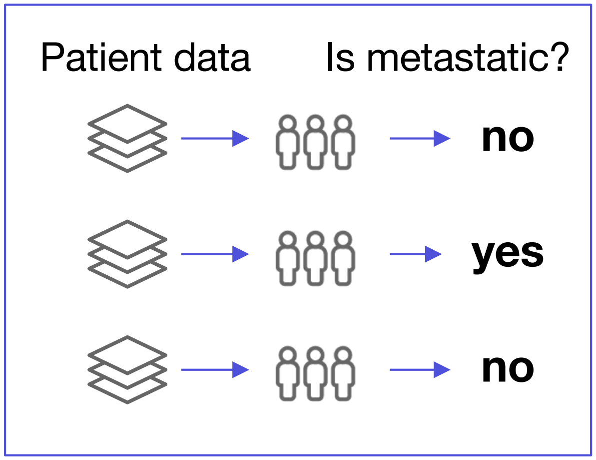


Cohort selection for metastatic breast cancer



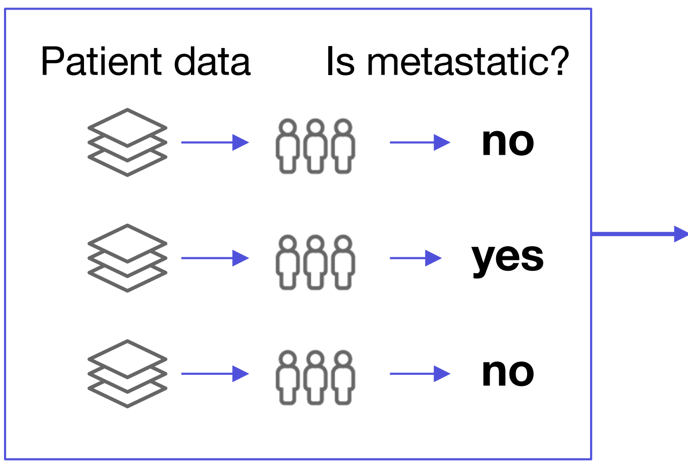
Approach for model-assisted cohort selection

1. Abstractors label some of the patients

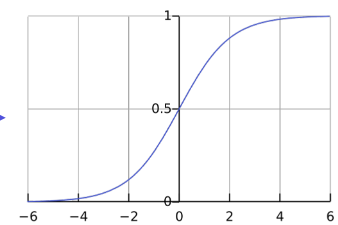


Approach for model-assisted cohort selection

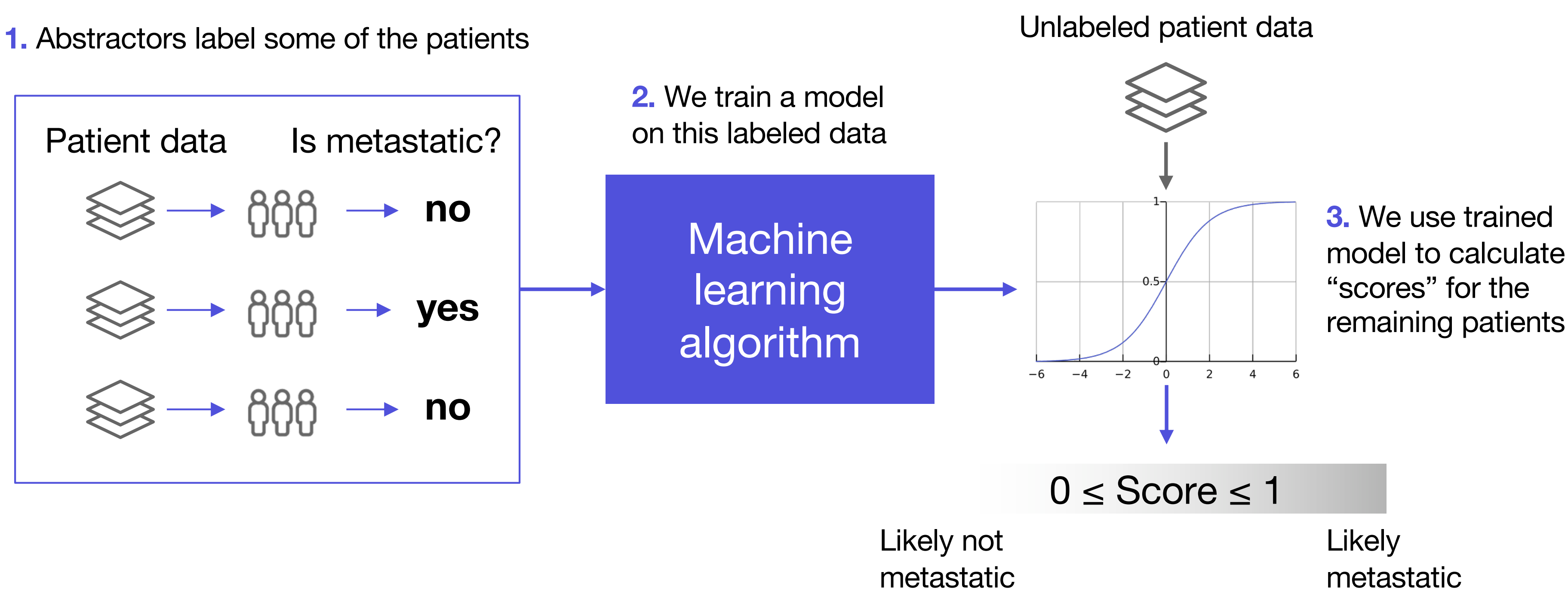
1. Abstractors label some of the patients



2. We train a model on this labeled data

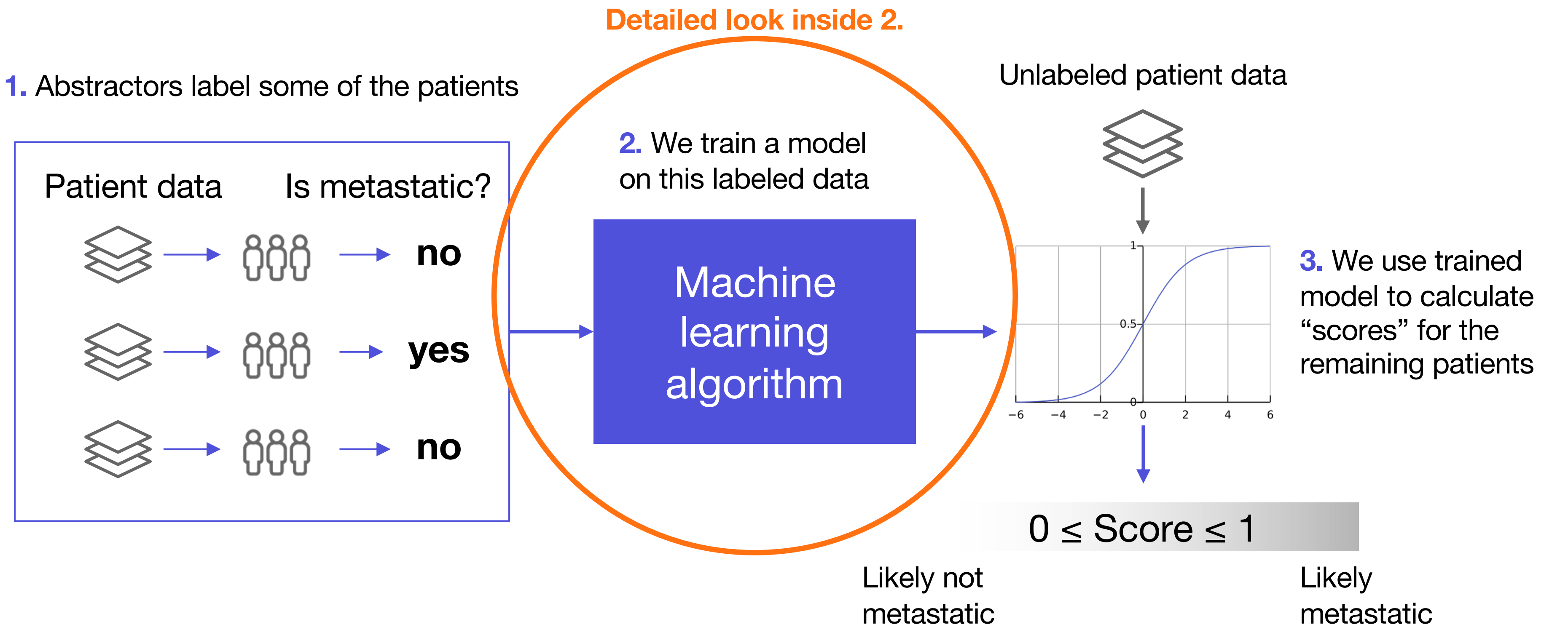


Approach for model-assisted cohort selection



4. Use the trained model as an extra pre-filter to reduce cohort selection abstraction

Approach for model-assisted cohort selection



4. Use the trained model as an extra pre-filter to reduce cohort selection abstraction

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

3. Create features for phrases in those snippets

Feature	Value
with stage	1
stage iv	1
...	...

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

3. Create features for phrases in those snippets

Feature	Value
with stage	1
stage iv	1
bone mets	1
mets were	1
...	...

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

3. Create features for phrases in those snippets

Feature	Value
with stage	1
stage iv	1
bone mets	1
mets were	1
not metastatic	0
...	...

Overview: Building the model

1. Construct list of relevant search terms

'metastatic'
'mets'
'recurrent'
'stage'

2. Extract snippets of text around search hits

History of Present Illness

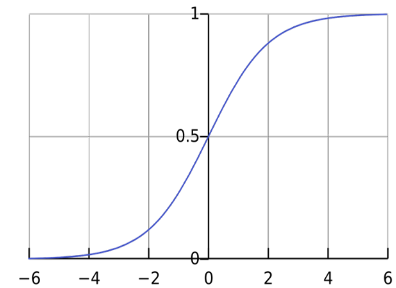
65-year-old female with stage iv breast ca diagnosed in Nov 2015

We are still awaiting records from Boca Raton, but she reports that she was diagnosed when bone mets were discovered during an ER visit for a broken pelvis…

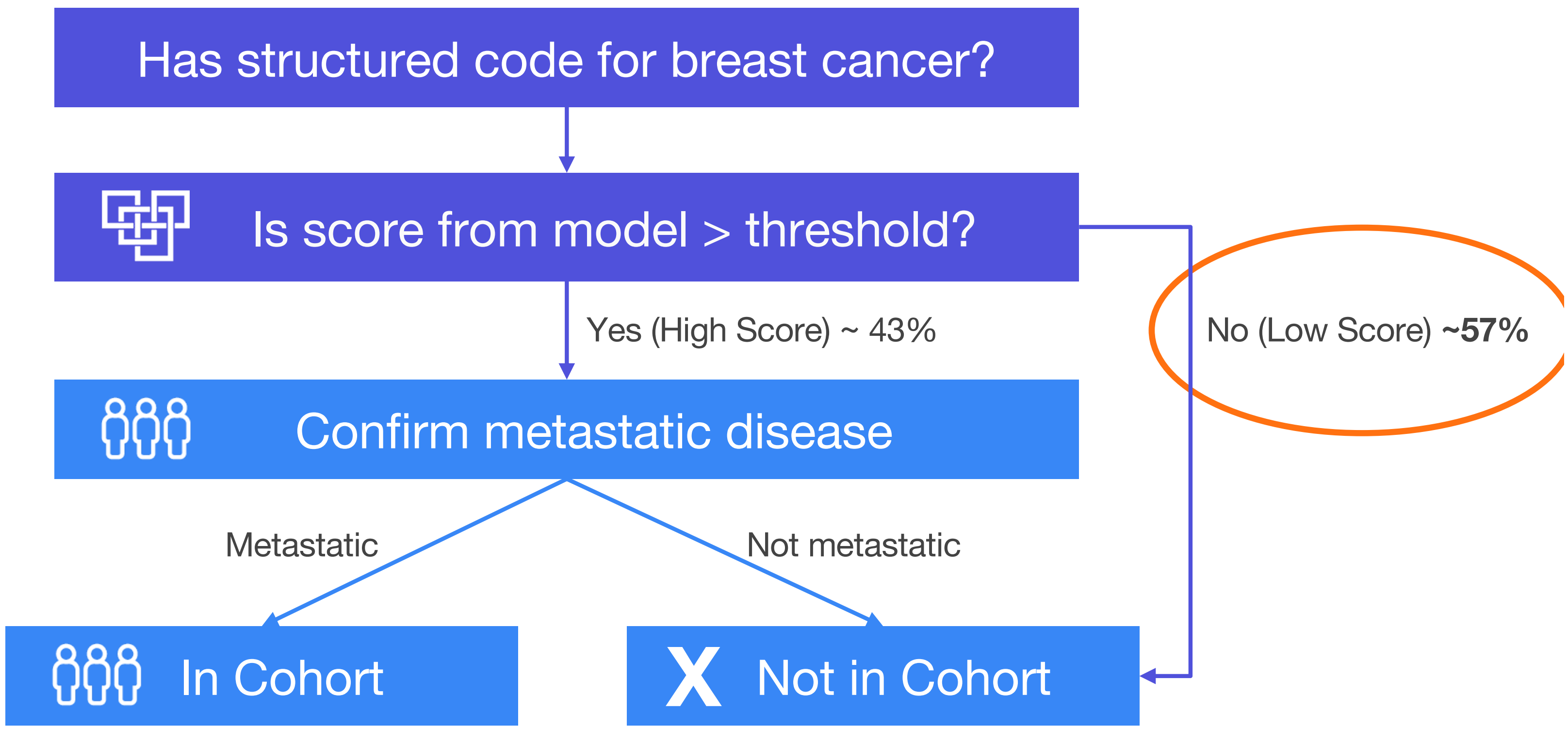
3. Create features for phrases in those snippets

Feature	Value
with stage	1
stage iv	1
bone mets	1
mets were	1
not metastatic	0
...	...

4. Use features for input to logistic regression



This example reduced hours of human abstraction for cohort selection by 57%



ML models need to be continually refreshed and tested to account for

Changes in treatment patterns

Changes in documentation patterns

Changes in network of health clinics

Feature Selection

Blythe Adamson, PhD, MPH
Flatiron Health

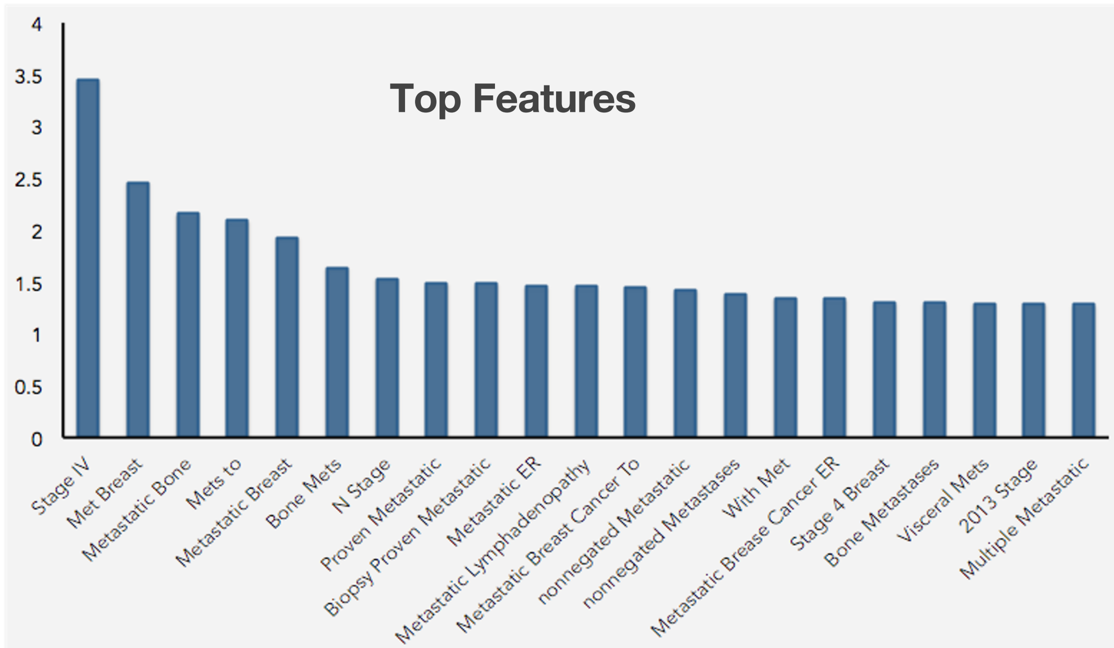
Approaches for ML feature selection from high-dimensional structured data

Filter	Generate and evaluate the subset of variables without the involvement of a model. Normally used as a preprocessing step.
Wrapper	“Brute force” feature selection techniques. The subset of variables is measured by the performance of the model.
Embedded	Propose and evaluate a subset of variables during the construction of the model.
Hybrid	Combine filter and wrapper approaches.

Traditional methods allow explainability of important phrases discovered in feature selection

Search Terms

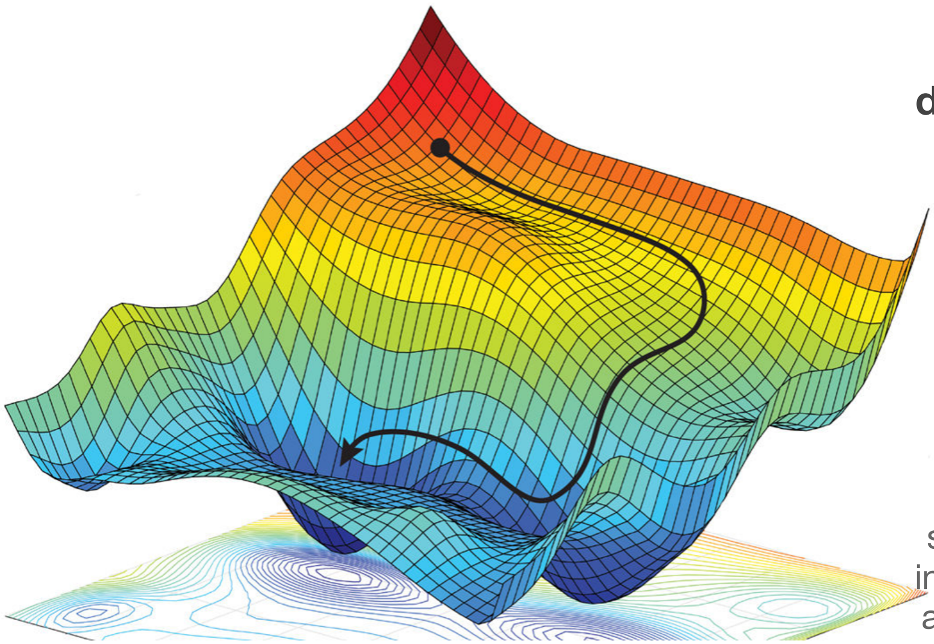
'distant'
‘met’
'metastases'
'metastasis'
'metastatic'
'mets'
'recurred'
'recurrence'
'recurrent'
'stage'
'systemic'



Deep learning methods allow the model to determine which features are important to select

Search Terms

'distant'
'met'
'metastases'
'metastasis'
'metastatic'
'mets'
'recurred'
'recurrence'
'recurrent'
'stage'
'systemic'



There is **unlimited dimensionality** in the unstructured data domain of free text and semantics in medical charts

i.e., memory cells in LSTM architecture are necessary for words at the beginning of a sentence to change the interpretation of a phrase at the end of a sentence

www.ispor.org

Predictive Analytics

Dave Vanness, PhD
Pennsylvania State University

Predictive Analytics: Definition

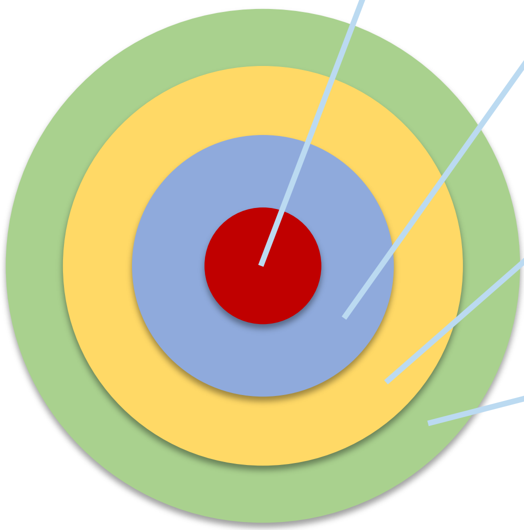
Predictive analytics is the systematic examination of the associative structure of observed data to generate estimates of data not yet observed. It comprises a set of statistical models and/or algorithms used to approximate or query the associative structure of observed data, a set of rules to generate estimated values from those models and/or algorithms, and an implied consequence incurred when the estimated value of the variable differs from its true value.

[Vanness DJ. Predictive analytics monograph. ISPOR. Forthcoming at https://www.ispor.org/member-groups/special-interest-groups/statistical-methods]

Predictive Analytics: Context

1. A decision must be made.
2. The decision depends on a prediction.
3. The prediction is informed by existing data.
4. When the prediction differs from the outcome, we incur a loss.
5. The loss depends on the magnitude and sign of the difference.
6. The data generating process may be complicated and high-dimensional.
7. We only need to know about the data generating process to the extent it helps make a better prediction.

Predictive Analytics: Key Concepts

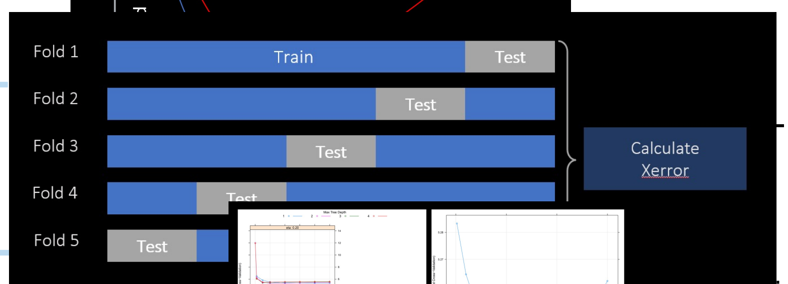


Bias-Variance Tradeoff

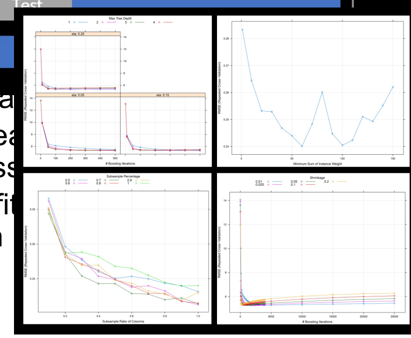
Highly complex models may give better fit to the data at hand but worse predictions for outcomes not yet observed

Cross

Creati the bias-variance tradec



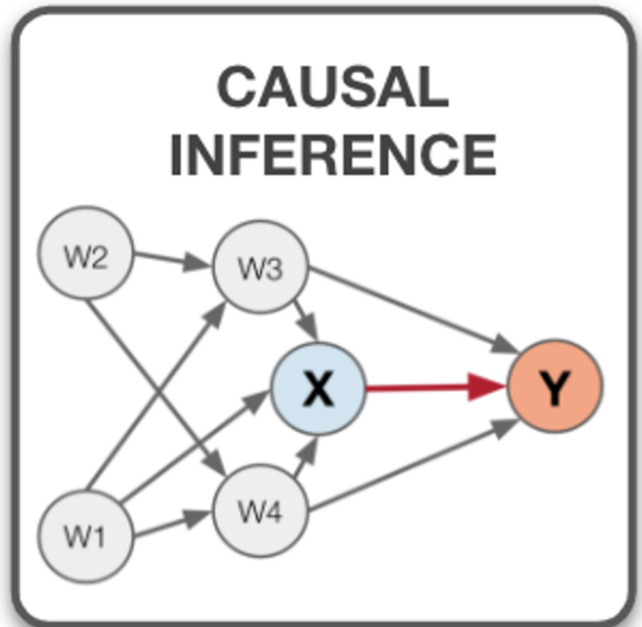
- Bagging: ra d aggregating multiple models
- Random fea est predictor from a random subset (less
- Boosting: fi he errors of the previous generation



www.ispor.org

Causal Inference

Bill Crown, PhD
Brandeis University



Causal Inference: Definition

- Causal inference is the unbiased estimation of the effect of a particular intervention (e.g., clinical or policy) on an outcome of interest. It requires consideration of the assumptions, study designs, and estimation strategies necessary to draw causal conclusions from data.

The Y-hat vs. B-hat Problem

Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106

Machine Learning: An Applied Econometric Approach

Sendhil Mullainathan and Jann Spiess

Machines are increasingly doing “intelligent” things: Facebook recognizes faces in photos, Siri understands voices, and Google translates websites. The fundamental insight behind these breakthroughs is as much statistical as computational. Machine intelligence became possible once researchers stopped approaching intelligence tasks procedurally and began tackling them empirically. Face recognition algorithms, for example, do not consist of hard-wired rules to scan for certain pixel combinations, based on human understanding of what constitutes a face. Instead, these algorithms use a large dataset of photos labeled as having a face or not to estimate a function $f(x)$ that predicts the presence y of a face from pixels x. This similarity to econometrics raises questions: Are these algorithms merely applying standard techniques to novel and large datasets?

Two Causal Frameworks

- Rubin. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology 1974. 66(5):688-701.

- Pearl. Causality. Cambridge University Press, 2000.

Journal of Economic Literature 2020, 58(4), 1129–1179
https://doi.org/10.1257/jel.20191597

Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics†

Guido W. Imbens*

In this essay I discuss potential outcome and graphical approaches to causality, and their relevance for empirical work in economics. I review some of the work on directed acyclic graphs, including the recent The Book of Why *(Pearl and Mackenzie 2018). I also discuss the potential outcome framework developed by Rubin and coauthors (e.g., Rubin 2006), building on work by Neyman (1990 [1923]). I then discuss the relative merits of these approaches for empirical work in economics, focusing on the questions each framework answers well, and why much of the the work in economics is closer in spirit to the potential outcome perspective.* (*JEL* C31, C36, I26)

1. *Introduction*

Causal inference (CI) in observational studies has been an integral part of econometrics since its start as a separate field in the 1920s and 1930s. The simultaneous equations methods developed by Tinbergen (1930), Wright (1928), Haavelmo (1943), and their successors in the context of supply and demand settings were from the beginning, and continue to be, explicitly focused on causal questions. Subsequently, the work by the Cowles Commission, and both the structural and reduced form approaches since then, have thrived by focusing on identifying and estimating causal and policy-relevant parameters. Over the last thirty years close

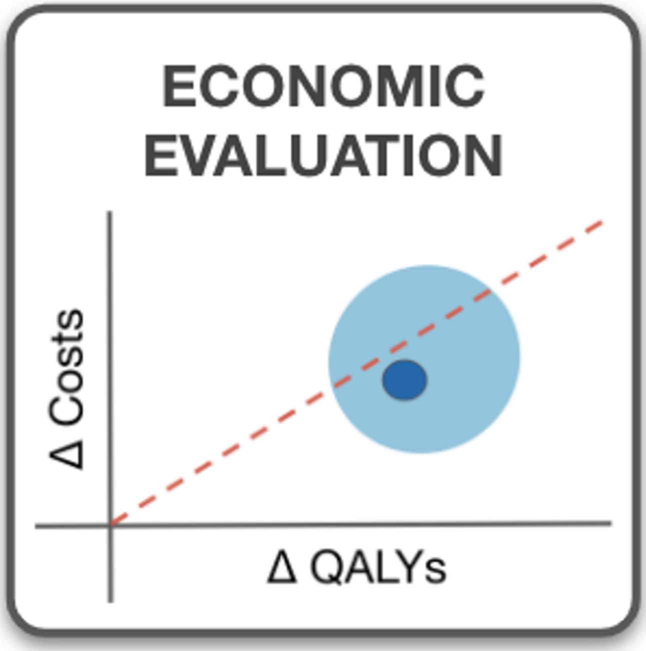
*Stanford University, SIEPR, and NBER. I am grateful for help with the graphs by Michael Pollmann and for comments by Alberto Abadie, Jason Abaluck, Alexei Alexandrov, Josh Angrist, Susan Athey, Gary Chamberlain, Stephen Chaudoin, Rebecca Diamond, Dean Eckles, Ernst Fehr, Avi Feller, Paul Goldsmith-Pinkham, Chuck Manski, Paul Milgrom, Evan Munro, Franco Perrachi, Michael

TMLE: An Example of how ML Prediction Can Be Used for Causal Inference

- Estimate (Y|T,X)
- Predict Y for (T=1,X) and (T=0,X)
- Estimate the exposure mechanism P(T=1|X) and use predicted probabilities to update E(Y|T,X) using doubly robust methods
- Calculate updated estimates of Y for (T=1|X) and (T=0|X) and calculate ATE as the mean of individual differences
- Finally, using ensemble methods (weighted average of TMLE estimate produced by a family of models of exposure and outcomes using cross-validation criteria) produces estimates that are asymptotically as good as the best performing model.

Economic Evaluation

Bill Padula, PhD
University of Southern California



Structural Uncertainty

ORIGINAL RESEARCH

OPEN ACCESS

Value of hospital resources for effective pressure injury prevention: a cost-effectiveness analysis

William V Padula,[1] Peter J Pronovost,[2,3] Mary Beth F Makic,[4] Heidi L Wald,[5] Dane Moran,[6] Manish K Mishra,[7] David O Meltzer[8]

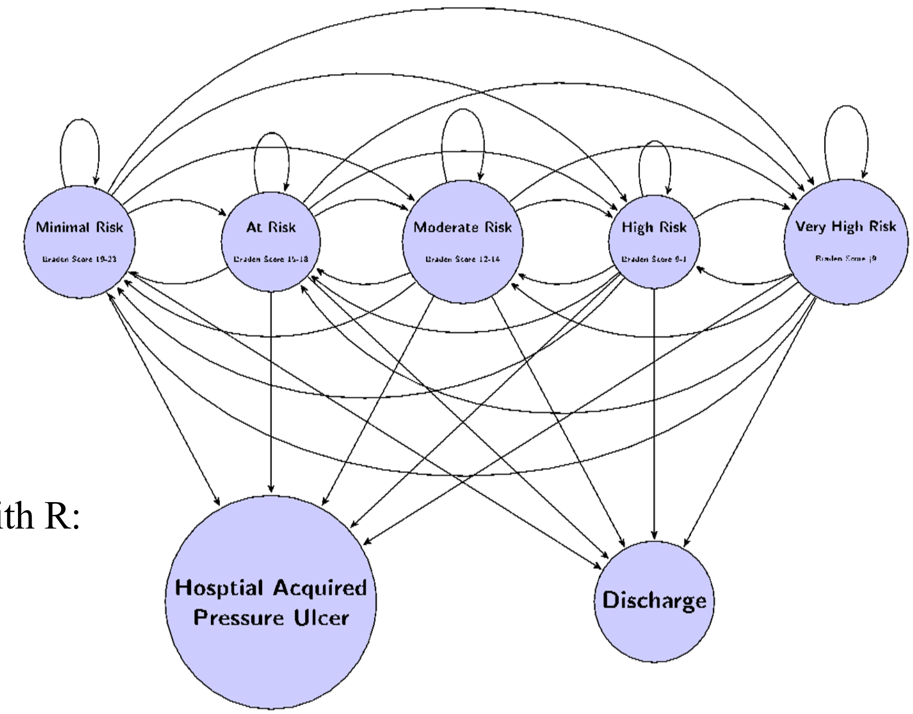
Padula WV, et al. BMJ Qual Saf 2019;28:132–141. doi:10.1136/bmjqs-2017-007505

BMJ

Multi-state Modelling with R:
the `msm` package

Version 1.6.4

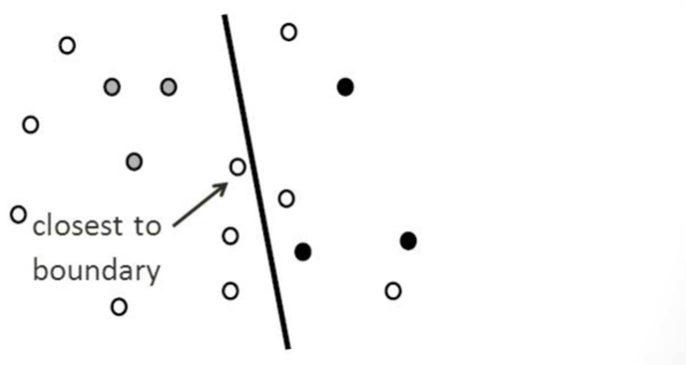
Christopher Jackson
MRC Biostatistics Unit
Cambridge, U.K.



Parameter and Sampling Uncertainty

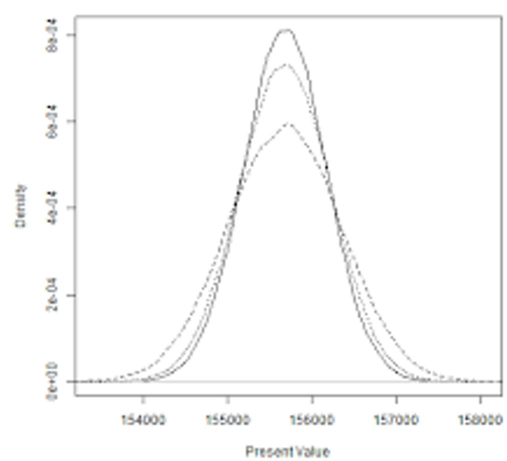
Improving certainty that sample observations used in cohort models are appropriate

Increasing the number of samples to draw on model findings



Finding the appropriate parameters in big data that are fit for purpose for model parameters

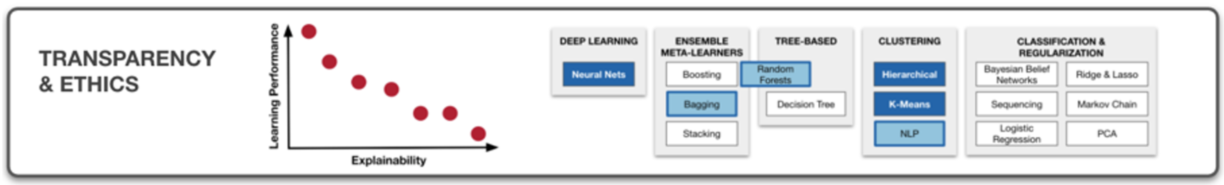
Fewer modeling assumptions



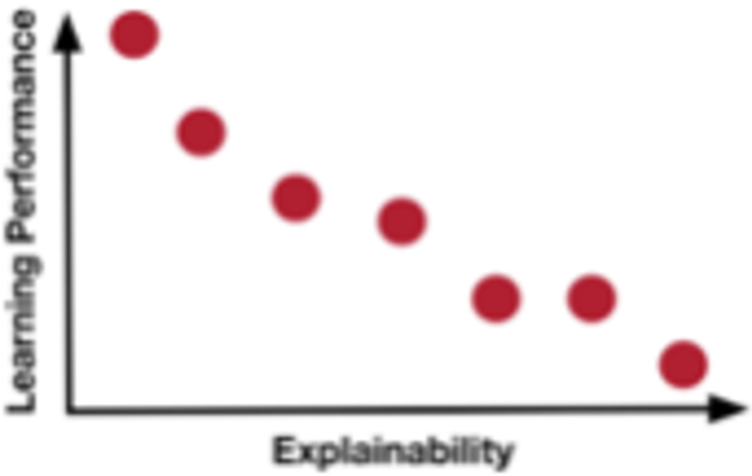
www.ispor.org

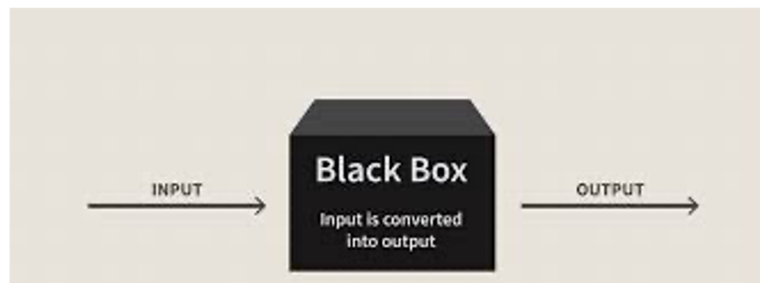
Transparency & Explainability

Bill Padula, PhD
University of Southern California



Transparency: finding the equilibrium between performance and explainability





The PALISADE Checklist

Key considerations for evaluating the transparency of machine learning (ML) to stakeholders and decision makers

Element	Definition
Purpose	Is the purpose of the algorithm clearly stated at the outset? Is the implementation of the algorithm in a healthcare setting fair and ethical?
Appropriateness	Is there a clear justification that the algorithm is acceptable in the context within which it is being applied?
Limitations	Have the strengths and limitations, in the context of the purpose, been identified? This should cover both the algorithm and any data used.
Implementation	Consideration of access, implementation and resource issues when implemented in healthcare settings
Sensitivity & Specificity	For classification algorithms, has the model performance and accuracy (specificity and sensitivity) been appropriately evaluated?
Algorithm characteristics	Has the ML mechanism been clearly characterized and described? Is there sufficient transparency for the results to be reproducible?
Data characteristics	Is the selection of datasets justified and are the key characteristics known? This should extend to training sets, test sets and validation sets.
Explainability	Are the outputs of the algorithm clearly understandable by both the healthcare professional and the patient?

ISPOR www.ispor.org

The PALISADE Checklist

Key considerations for evaluating the transparency of machine learning (ML) to stakeholders and decision makers

P	**Purpose**	Is the purpose of the algorithm clearly stated at the outset? Is the implementation of the algorithm in a healthcare setting fair and ethical?
A	**Appropriateness**	Is there a clear justification that the algorithm is acceptable in the context within which it is being applied?
L	**Limitations**	Have the strengths and limitations, in the context of the purpose, been identified? This should cover both the algorithm and any data used.
I	**Implementation**	Consideration of access, implementation and resource issues when implemented in healthcare settings
S	**Sensitivity & Specificity**	For classification algorithms, has the model performance and accuracy (specificity and sensitivity) been appropriately evaluated?
A	**Algorithm Characteristics**	Has the ML mechanism been clearly characterized and described? Is there sufficient transparency for the results to be reproducible?
D	**Data Characteristics**	Is the selection of datasets justified and are the key characteristics known? This should extend to training sets, test sets and validation sets.
E	**Explainability**	Are the outputs of the algorithm clearly understandable by both the healthcare professional and the patient?

Padula et al, Value in Health (2022). *In Press.*

www.ispor.org

Case #1

Bill Padula, PhD
University of Southern California

Case #1

- A graduate student has obtained a dataset from a physician
  - N = 50 patient observations; all patients diagnosed with hypercholesterolemia
  - 17 variables/columns: HDL/LDL level; Statin Rx; Time; Age; Sex; Race; BMI…
- The physician wants to know if there is a particular statin associated with low cholesterol levels over time
- The physician has heard the amazing breakthroughs of machine learning for predicting health outcomes
- The physician asks the graduate student to use a Neural Network to predict the best Statin Rx for lowering HDL/LDL
- Is this a good case for Machine Learning? Use the PALISADE Checklist to determine your answer

It’s Time for a Poll!

- **CASE #1: Is this a good case for Machine Learning? Use the PALISADE Checklist to determine your response.**
  a. **Yes**
  b. **No**

Case #1 Analysis

- ISPOR Task Force Recommendation: **No – don't use ML**
- The dataset is small and easily manageable 'by hand'
- These data could easily be fit to an OLS or Logistic Regression model
  - Then use stepwise regression to identify the fit of covariates
- No causality
- Logistic Regression achieves predictive quality that is Explainable
- A Neural Network could accomplish the task, but with limited transparency that the physician could explain in practice settings

Case #2

Bill Padula, PhD
University of Southern California

Case #2

- A PharmD at the FDA has been asked whether an NSAID anti-inflammatory Rx is causing stroke in the general population
- The NSAID is a blockbuster drug, and has been on the market for 7 years, prescribed to millions of patients
- Most patients who take this Rx sometimes take a combination of multiple other drugs – some other drugs have black box warnings for stroke
- You have access to a database with the following numbers:
  - N = 17.2 million observations
  - M = 9,000 variables, including over 3,500 columns of data on Rx's
- Initial sampling indicates that ~11% of patients have taken this NSAID
- The pharmacist calls the FDA biostatistical department to ask if they can apply machine learning to this sample to help come to a better answer
- Is this a good case for Machine Learning? Use the PALISADE Checklist to determine your answer

It’s Time for a Poll!

- **CASE #2: Is this a good case for Machine Learning? Use the PALISADE Checklist to determine your response.**
  a. **Yes**
  b. **No**

Case #2 Analysis

- ISPOR Task Force Recommendation: **Yes – use ML**
- The dataset is large and complex
  - **Cohort Selection** methods, including clustering, could help isolate the 11% of over 1 million patients that have used the drug, and identify comparable samples of control patients with stroke
  - **Feature Selection** methods such as random forests could help boil down thousands of variables to several dozen predictors, including Rx interactions that could be predictive of the main outcome measure, stroke
- Predictive analytic methods are suited to analyze which combination of covariates accurately assess the probability of stroke in an at-risk population taking this drug
  - A combination of supervised and unsupervised methods may need to be tested to see which approach is best predicting the outcome before drawing conclusions
- TMLE could be applied to these statistical methods in order to determine whether the drug causes stroke in this cohort study

Conclusions

Reminder:
Task Force Discussion Hour
11:45am-12:45pm
Discussion Lounge
Prince George's Exhibition Hall A-C

Thank you for participating!

- Methods for machine learning are advancing more quickly than HEOR can keep pace with
- Seek consultation with computer scientists and applied mathematicians/statisticians to determine whether ML fits your research question
- Consider a range of ML and traditional HEOR methods to serve your analytical needs
- *Is the juice worth the squeeze*? ML requires a great deal of effort and computational resources to initiate for clinical research purposes

ISPOR, the professional society for health economics and outcomes research (HEOR), is an international, multistakeholder, nonprofit dedicated to advancing HEOR excellence to improve decision making for health globally. The Society is the leading source for scientific conferences,
peer-reviewed and MEDLINE-indexed publications, good practices guidance, education, collaboration, and tools/resources in the field.

ISPOR's community of more than 20,000 individual and chapter members from 120+ countries includes a wide variety of healthcare stakeholders, including researchers, academicians, regulators and assessors, public and private payers, healthcare providers, industry, and patient representatives. The Society's leadership has served as an unbiased resource and catalyst for innovation in the field for more than 20 years.

www.ispor.org

ISPOR

Improving healthcare decisions