www.ispor.org

ISPOR
Improving healthcare decisions

ISPOR Special Interest Group

W3: DEMYSTIFYING MACHINE LEARNING: WHY SHOULD WE BE OPEN TO IT?

ISPOR Europe 2019 | Copenhagen, Denmark

Monday, 4 November 2019 | 14:15 – 15:15

ISPOR Statistical Methods in HEOR Special Interest Group (SIG)

Mission: To provide statistical leadership for strengthening the use of appropriate statistical methodology in health economics and outcomes research and improve the analytic techniques used in real world data analysis.

Co-Chairs of SIG

- **Rita M. Kristy, MS**, Senior Director, Medical Affairs Statistics, Astellas Pharma Global Development, Northbrook, IL, USA
- **David J. Vanness, PhD,** Professor, Health Policy and Administration, The Pennsylvania State University, State College, PA, USA

Co-Chairs of ISPOR Missing Data in HEOR Working Group

- **Necdet Gunsoy, PhD, MPH,** Director of Analytics and Innovation for Value Evidence and Outcomes at GlaxoSmithKline (GSK), England, United Kingdom
- **Gianluca Baio, PhD, MSc,** Professor of Statistics and Health Economics, University College London (UCL), England, United Kingdom

ISPOR Special Interest Group

Discussion Leaders:

- **Hélène Karcher, PhD**
  - Parexel Regulatory and Access Consulting, Basel, Switzerland
- **Gorana Capkun, PhD**
  - Novartis Oncology, Basel, Switzerland
- **Christoph Gerlinger, PD Dr.**
  - Bayer AG, Berlin, Germany and
  - University Medical School, Homburg/Saar, Germany
- **Jacqueline Vanderpuye-Orgle, PhD**
  - Paraxel Regulatory and Access Consulting, Los Angeles, CA, USA

3

Q&A

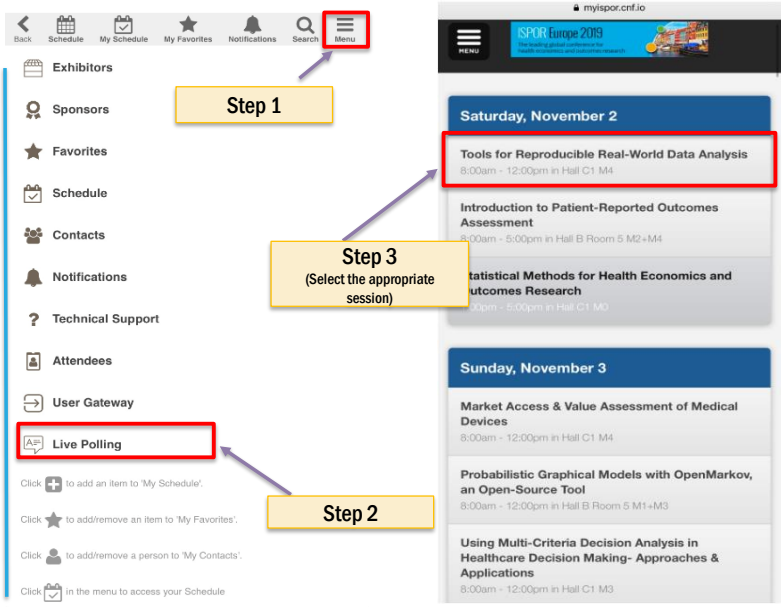
ISPOR **Conference Platform**

Web Platform

https://myispor.cnf.io/

Mobile App

Search "**ISPOR Europe 2019**" in the App Store or on Google Play!



4

SECTION

1

A new era for healthcare data

Why use Machine Learning now?

Helene Karcher, PhD

Machine Learning: the use of algorithms to learn from large volumes of data and predict future outcomes

- Machine learning is a two-step process
  - **Learning**: Learn from relationship between many variables and a few outcomes
  - **Predicting**: Use this learned relationship to predict future outcomes in new situations
- Term originated from Computer Science and Statistics
  - Examples include: predictive text, image recognition, etc.
  - Traditional industries using ML: data-rich industries such as web (google,FB, etc), advertisement, sales analyses....
- Applications to healthcare are more recent

Why is machine learning „suddenly“ big in healthcare?

1. **Progress in data availability**
   - New type and/or large volume of data, e.g., from wearables, genotypes, medical images etc.
   - Data integration: linkages between RWD sources (EMR, lab data, demographics, medications), enabled by progress in nomenclature: Common Data Model (OMOP), tools for conversion from OHDSI (www.ohdsi.org), ICD-10-CM codes, CPT codes, etc.
   - Data became (more) accessible to payers, researchers and manufacturers
     - Opening of registries to academic networks, commercialization of claims data, ...
     - Development of rules around data privacy and sharing

7

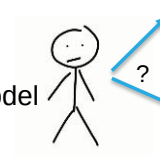
ISPOR www.ispor.org

Why is machine learning „suddenly“ big in healthcare?

2. **Better computing power, data exchange and storage**
   - Larger, shared storage platforms. e.g., EMIF for Alzheimer´s disease, “Health Data Hug” in France, Flatiron
   - Algorithm running speed: models that use to run in weeks now run in minutes
   - Possibility to run algorithms using data collected in real-time
     - Predictions of individual risks can use up-to-date population and individual data

8

Which questions does machine learning algorithms answer in healthcare?



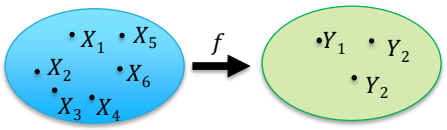
- Individual risk prediction
  - Gorana: predicting NASH risk with a gradient boosting model
- Choice of population for a new treatment
  - Jackie: predicting the risk of pulmonary hypertension with Bayesian networks
- Predicting outcomes or complications
  - Christoph: predicting bleeding patterns using Random forest & Lasso regression
  - Jackie: prediction of OS and AEs in oncology with Bayesian networks
- Treatment effect predictions (Sherri Rose´s seminar mid Nov)
- Testing of scenarios, e.g., for outcomes-based agreements
- ...

9

Unsupervised vs. supervised machine learning

- Unsupervised: to find patterns in data. No pre-specified output.

  - Often a preparatory step to supervised learning
- Supervised: links input to output variables $Y_i(t) = f(Xi(t))$



  - Often the ultimate goal (prediction)

Algorithms to specify f() numerically from datasets, while mimimizing bias, variance

X_i : e.g., (time-dependent) demographics, lab data, real-time exercise level, genotype, etc.

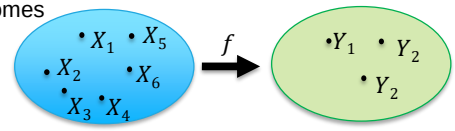
Y_i : e.g., (time-dependent) patient risk to develop an illness, have a complication, be hospitalized, etc.

10

Types of machine learning methods, differences from other „traditional“ models

- Machine learning algorithms are different from disease models and encompass usual statistical functions
  - Hypothesis-free
  - The results from machine learning applications **may not be readily interpretable**.
    - However, they may be more precise because they integrate more data and allow for non-linear behaviors

- Single model methods
  - Regression models (possibly with regularization) – when Y_i are continuous outcomes
  - Classification methods / trees – when Y_i are binary outcomes



- Ensemble methods (computationnally intensive)
  - Random forest (ensemble of trees)
  - Bagging: random resampling, model averaging
  - Boosting (gradient boosting): sequences of models
  - Random feature selection: choosing the best predictor at every modeling step

SECTION

2

Case study: How machine learning techniques was used to develop a patient-centered prediction model for the menstrual bleeding for women who start a new hormonal intrauterine device for contraception

Christoph Gerlinger, PD Dr.

Acknowledgement

- Joint work with
  - Ann-Kathrin Frenz and Christiane Ahlers, Statistics and Data Insights
  - Dr. Vita Beckert, Medical Affairs

13

What is a hormonal intrauterine device?

- Long-acting reversible contraceptive method
- Placed into uterus
- Contraception for up to five years

14

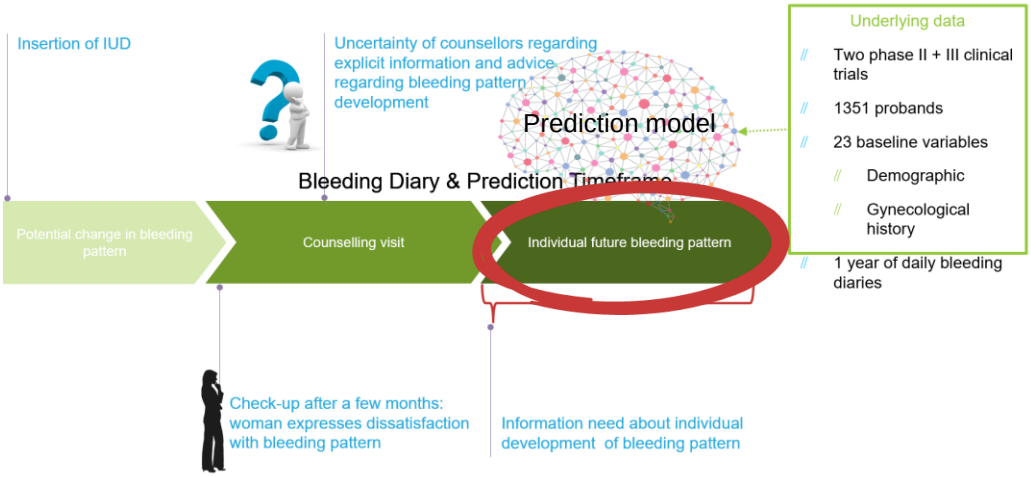
Motivation

- Hormonal influence on menstrual bleeding cycle
- Effect: potential alternation of previous familiar menstrual bleeding pattern
  - → Unfamiliarity with “new” bleeding pattern might lead to insecurities and questions regarding its future development
- Long-term development of bleeding patterns highly individual and heterogenous
  - → Need of personalized information in IUD counselling

„I got my IUD inserted 2 months ago, and I still have very intense irregular bleeding episodes. Will this ever get better?“

15

Motivation – 2



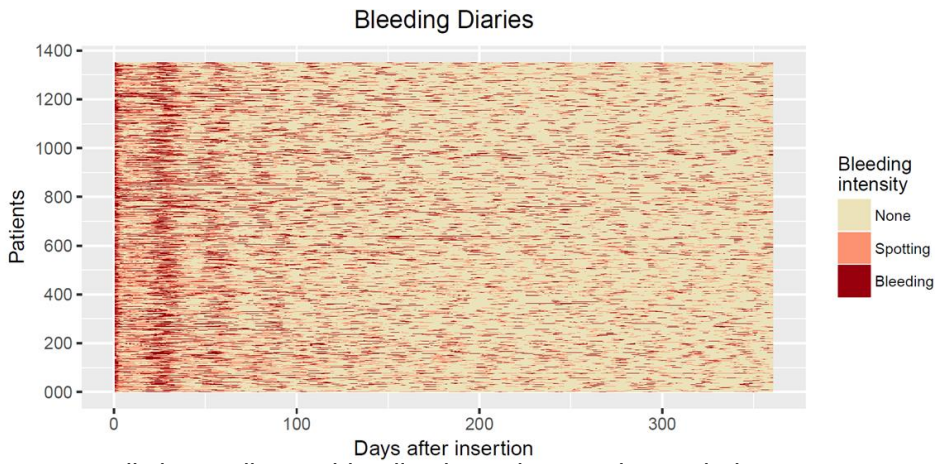
16

Bleeding Diary & Prediction Timeframe

17 relevant prediction attributes: bleeding intensity & cycle regularity

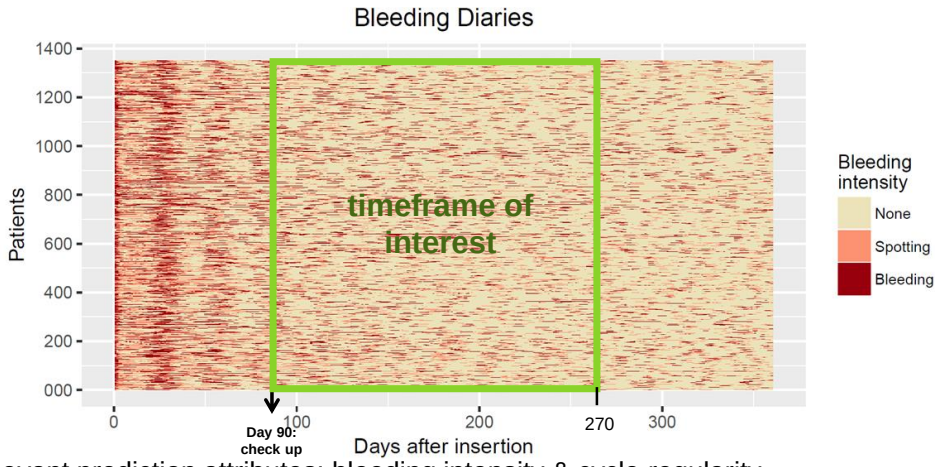
ISPOR www.ispor.org

Bleeding Diary & Prediction Timeframe



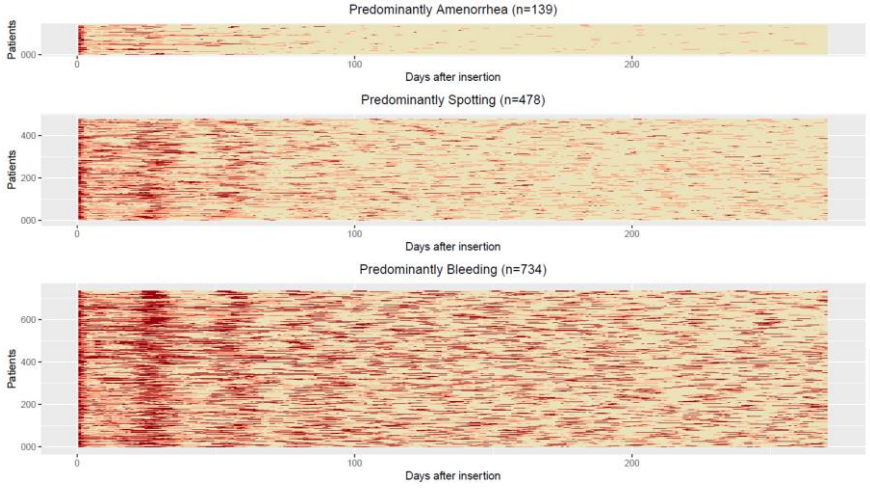
18 relevant prediction attributes: bleeding intensity & cycle regularity

Bleeding Diary & Prediction Timeframe



19 relevant prediction attributes: bleeding intensity & cycle regularity

Bleeding Intensity Clusters



- < 5% spotting days
- < 1% bleeding days

- ≥ 5% spotting days
- < 5% bleeding days

- ≥ 5% bleeding days

20

Assessment of Menstrual Cycle Regularity

- Assessment of regularity difficult due to unknown cycle lengths of patients
- Nearly impossible to identify intermenstrual bleeding

Step 1: Assessment of cycle length

Calculation of autocorrelation function for values between 21 and 35 days

- Similarity measure between original and shifted time series
- Highest for time after which time series repeats itself → cycle length
- Maximum value of function → identified cycle length

Step 2: Assessment of regularity

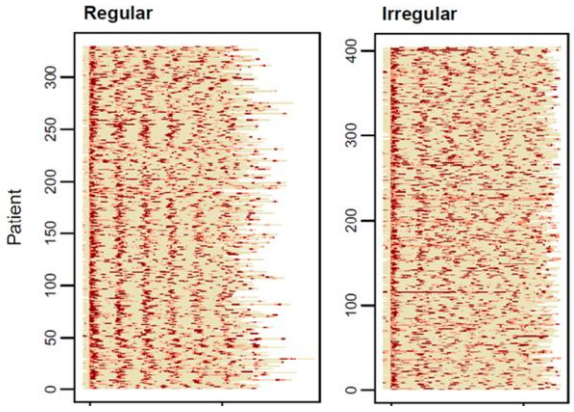
Comparison of two time series models: with and without seasonal component

If seasonal model provides better AIC → regularity identified

21

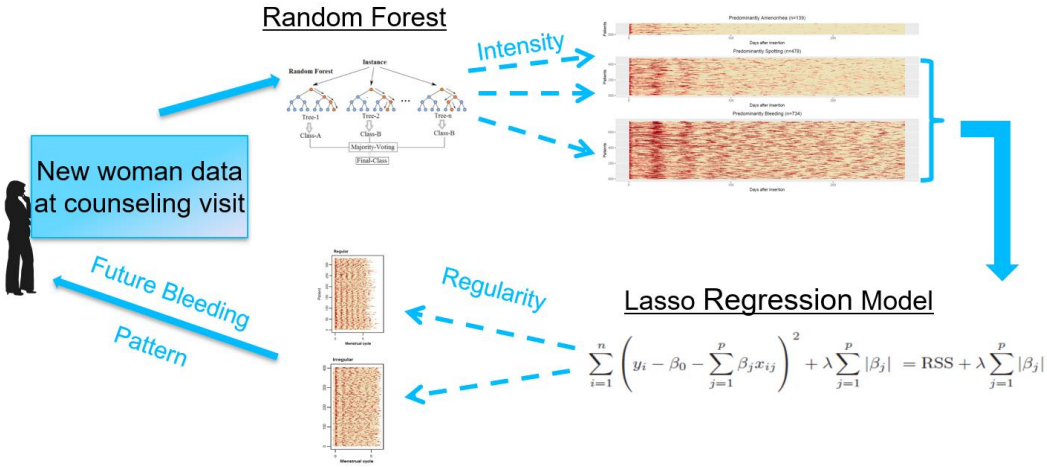
Assessment of Menstrual Cycle Regularity

- Very complex due to individually varying and unknown cycle lengths
- State-of-the-art time series methodology to identify regularity



22

Prediction Model



23

Prediction Performance

- Model developed on data from one dose (LNG-IUD 19.5 mg)
- Random forest predicts cluster for woman
- In our data true clusters are known → prediction can be compared to truth and tested for performance

Performance rates

- Correct classifications: **70.4%**
- Classification in correct or "neighboring" clusters: **98.7%**
- Severe misclassifications: 1.3%

→ Model proved to be very well working on LNG-IUD 19.5 mg data
→ Does it also work on new, independent data?

24

Prediction Performance – 2

- Same model, different independent data:

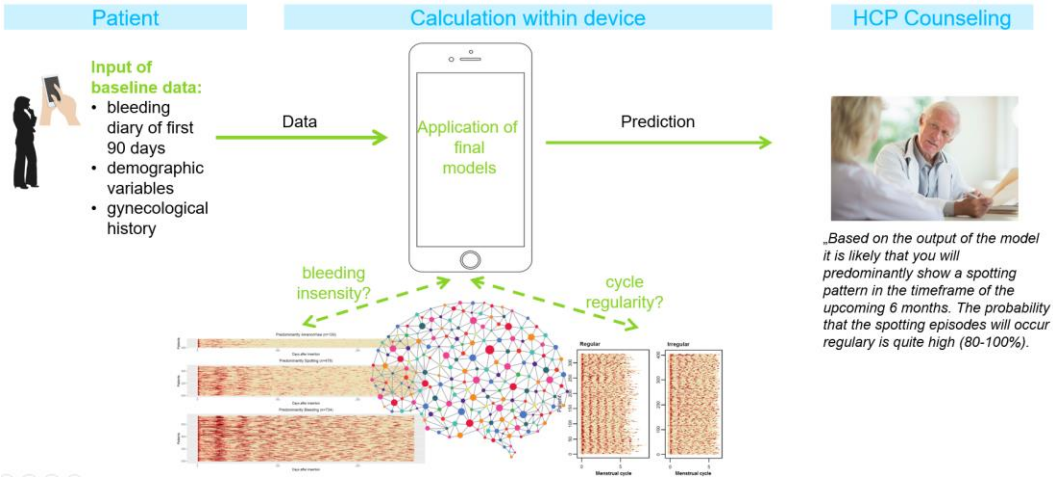
- LNG-IUD 52 mg (n = 216)
  - 72.2% correct classifications
  - 100.0% correct or neighboring classifications
- LNG-IUD 13.5 mg (n= 1300)
  - 69.0% correct classifications
  - 98.8% correct or neighboring classifications

➢ Model works equally well on independent data which was not used for model setup
➢ Works despite different dose in products → application to further portfolio possible

25

ISPOR www.ispor.org

Digital Solution Overview



26

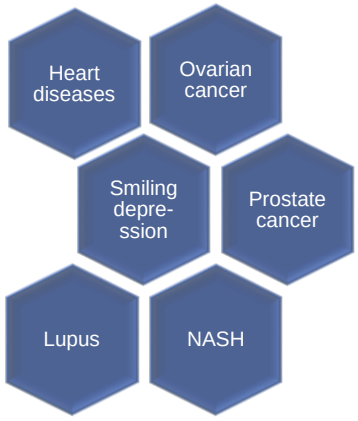
SECTION

3

Use of machine learning for patient identification, improving referrals for silent diseases with difficult diagnostic patterns and no existing therapy

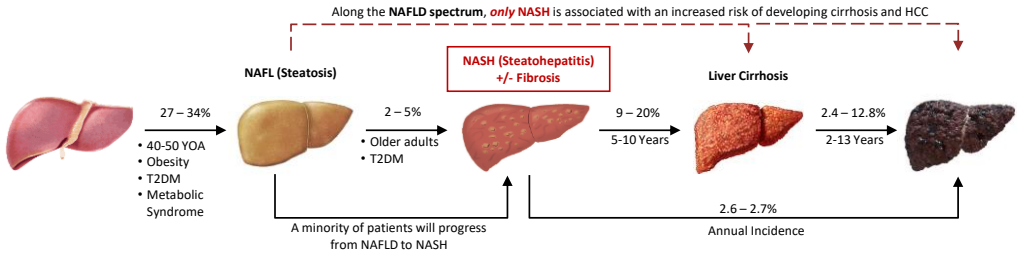
Gorana Capkun, PhD

Creating an alert for silent killers



28 NASH, Non alcoholic steatohepatitis

Aim of treatments: resolution of hepatic inflammation, reduction of fibrosis, avoid cirrhosis, HCC, liver transplantation



1 million or 1.8% of all deaths world-wide annually are caused by liver cirrhosis, the common result for all chronic liver diseases[1]

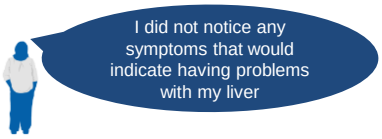
NASH is expected to be the leading cause of liver transplant in the US by 2020[2]

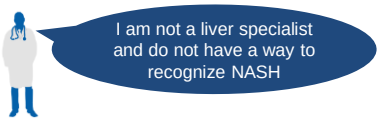
* HCC, Hepatocellular carcinoma; NASH, Non alcoholic steatohepatitis; NAFLD, Non alcoholic fatty liver disease; T2DM, Type 2 diabetes mellitus

29

1.Not prevalence but annual variceal bleeding cases in cirrhotic population with GE varices. Source: Mokdad et al. (2014)
2. Incident Population. Source: Zaman et al. (2008)

Lack of differential symptoms in addition to biopsy based diagnosis lead to late NASH referal to hepatologists



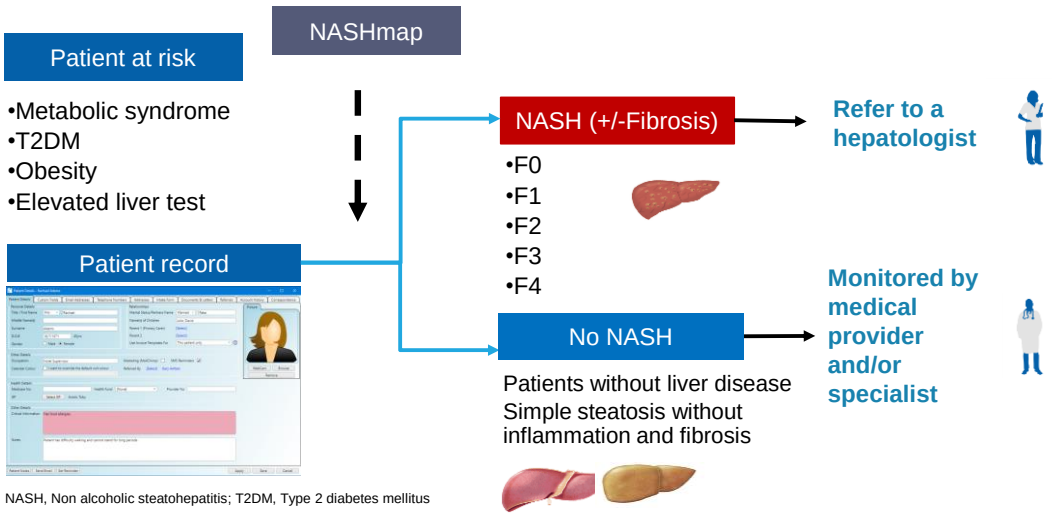


- NASH is mostly asymptomatic or patient report unespecific symptoms and common to many diseases
- Biopsy is the “gold standard” to diagnose NASH performed **only when high suspicion of NASH by a hepatologist**
- The **aim** was to create an algorithm: NASHmap
  - that accurately predicts NASH and
  - can be applied to EHR systems or in clinical practice to identify likely NASH patients

30

NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record

Application of NASHmap in clinical practice

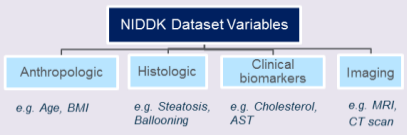


31 NASH, Non alcoholic steatohepatitis; T2DM, Type 2 diabetes mellitus

Two data sets: NIDDK registry and Optum EHR

NIDDK Registry

- US observational research data for NAFLD
- 450+ NASH patients (biopsy confirmed)
- Follow up over ~4 years (2002 - 2006)
- Variables collected

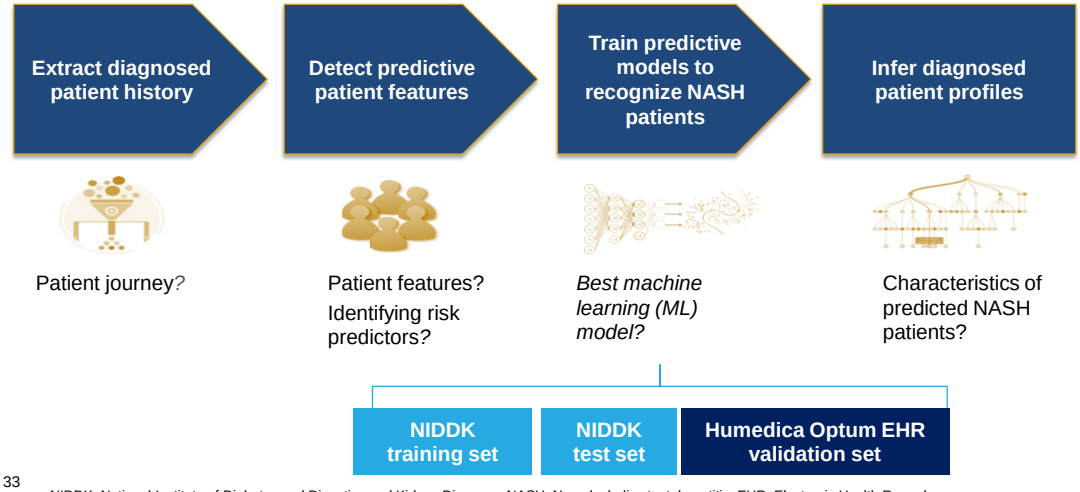


Optum Humedica EHR Data

- US RWD with information from 2007 – 2017
- 23K NASH/NAFLD diagnosed patients
- Variables collected:
  - patient demographics,
  - diagnoses,
  - procedures,
  - prescribed medications,
  - vital signs,
  - lab values.
  - physician notes (NLP processed)

32 NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases; NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record; NAFLD, Non alcoholic fatty liver disease; BMI, Body Mass Index; AST, aspartate aminotransferase; MRI, magnetic resonance imaging; CT, computed tomography; NLP, natural language processing

Modeling process



33

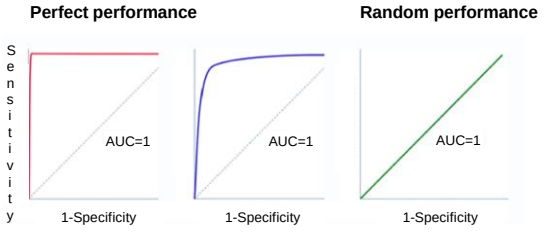
NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases; NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record

Model is evaluated using performance metrics

	Predicted Class	
Actual Class	***NASH***	***Non-NASH***
NASH	True Positive (TP)	False Negative
Non-NASH	False Positive (FP)	True Negative

Performance*	Area Under the ROC Curve (AUC)*
Fail	50-60%
Poor	60-70%
Fair	70-80%
Good	80-90%
Excellent	90-100%

- **Sensitivity**: % of NASH patients correctly identified out of the NASH diagnosed
  - If low, model fails to identify NASH patients
- **1 - Specificity**: % of non-NASH patients predicted to have NASH
  - If high, model suggests to biopsy patients without NASH



34

* Assuming that undiagnosed patients don't have the disease

NASH, Non alcoholic steatohepatitis; ROC, receiver operating characteristic

The most important features for predicting NASH were identified and included in the model

These are the **best 14 features** from the starting set of 35 features ranked on their contribution to predictions.

Importance Rank	Feature
1	HbA1c
2	AST
3	ALT
4	**Total Protein**
5	AST/ALT
6	BMI
7	Triglycerides

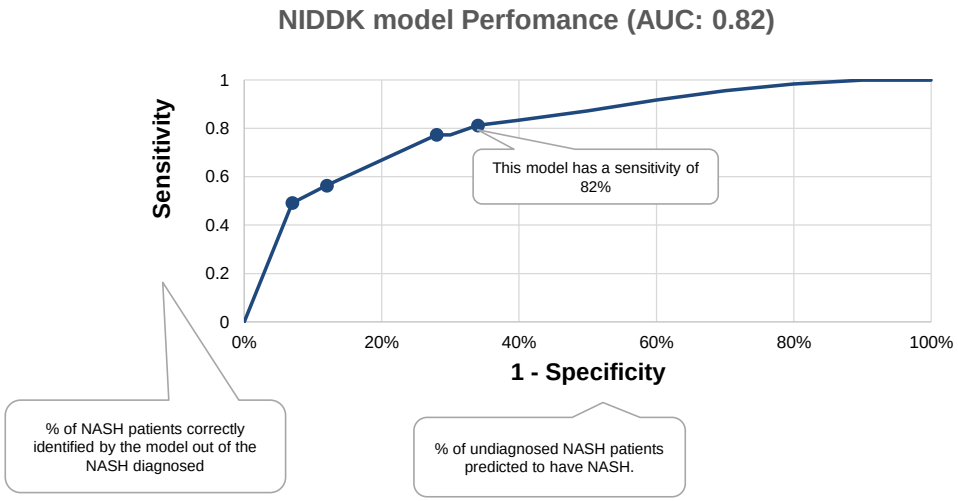
Importance Rank	Feature
8	Height
9	Platelets
10	**WBC**
11	**Hematocrit**
12	Albumin
13	**Hypertension**
14	Gender

Five features used in the reduced 5 feature model.

Bolded variables are not used in the existing scores for fibrosis detection

NASH, Non alcoholic steatohepatitis; BMI, Body Mass Index; AST, aspartate aminotransferase; ALT, alanine transaminase; HbA1c, average blood glucose levels for the last two to three months; WBC, white blood cells

The new model demonstrates high performance in the NIDDK registry data with a sensitivity of 81%



NASH, Non alcoholic steatohepatitis; AUC, area under the curve; NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases

Performance in Optum data can only be evaluated on sensitivity

Challenge	Metrics	Interpretation
Difficult to identify NASH patients in real world data: ▪ Undiagnosed patients may have NASH ▪ NASH diagnosis may not be accurate **In EHR, NASH positive cases area defined as patients with NASH diagnosis and liver biopsy**	We focused on two metrics where NASH positive cases are clearer: ▪ **Sensitivity:** % of NASH patients correctly identified among NASH diagnosed ▪ **1 – Specificity:** % of undiagnosed NASH patients* predicted to have NASH **We can balance these metrics to capture undiagnosed patients**	NASHmap performance assessed with sensitivity **at a controlled specificity**: ▪ Allows comparison against other models ▪ Allows customization to different use cases **Based on this approach, we've evaluated the model**

*Ideally, we want to optimize sensitivity and precision**. We can optimize these metrics in NIDDK, but have to strike balance with our metrics above in Optum EHR*

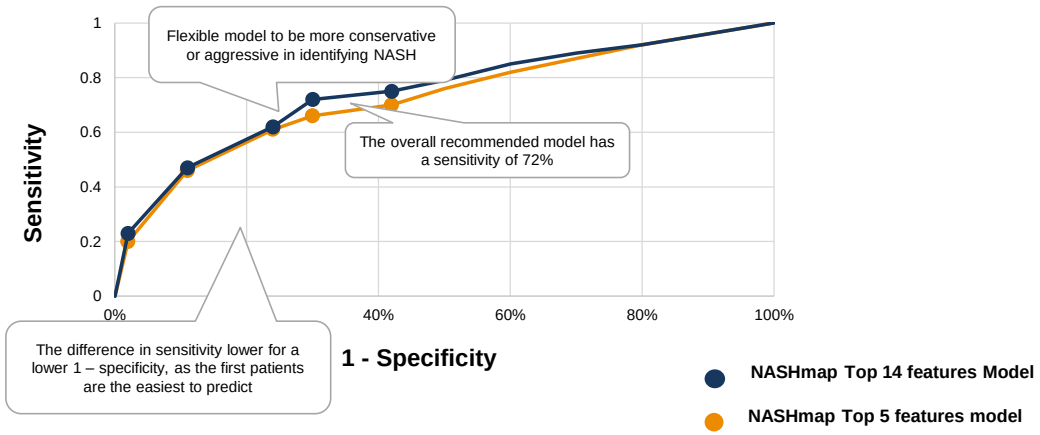
37

Undiagnosed NASH patients: Patients without NASH diagnosis or NAFLD diagnosis ;
**Precision: % of NASH diagnosed patients out of the patients identified as NASH by the model

NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases; NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record

ISPOR www.ispor.org

The NASHmap model also demonstrates high performance in Optum EHR data



38 NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record

NASHmap is a good predictor of NASH patients

Best Models

- *The highest performing model: is **14 variable gradient boosting model***
- *Balanced model: 5 input variables slightly lower performance, can be implemented on 30% more patients (~1M)*

Model Performance

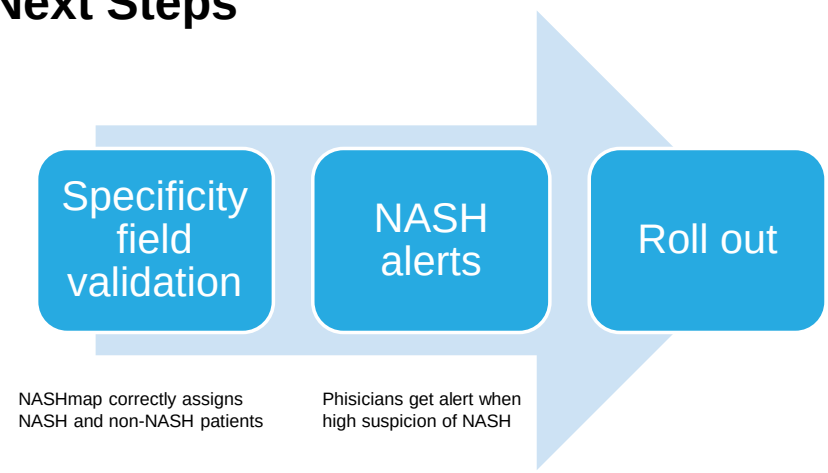
Data/Model Type	Sensitivity (Top 14 features)	Sensitivity (Top 5 features)
NIDDK Registry Model *(test data performance)*	81%	80%
Optum EHR Data	72%	66%

For Optum EHR, 30% of undiagnosed patients are predicted as NASH positive

39 NASH, Non alcoholic steatohepatitis; EHR, Electronic Health Record

ISPOR www.ispor.org

Next Steps



40 NASH, Non alcoholic steatohepatitis

Your question?

Your thoughts/ideas?

Striking aspects?

41

SECTION

4

Applications and case studies of machine learning in oncology and pulmonary hypertension for patient identification, clinical decision support, prognostic tools.

Jackie Vanderpuye-Orgle, PhD

Application of Machine Learning in Oncology

Situational assessment

- Current trends in the healthcare ecosystem highlight the need to identify strategies to better predict individual patient response to innovative IO therapies
  - Increased development and use of precision medicines/targeted therapies
  - American Joint Committee on Cancer's recent call for cancer risk prediction models to move towards individualized risk prediction
  - Payers/managed care driving towards the use of therapies for the "right patient at the right time" to help curb costs

Response

- Goal is to develop a transparent individualized risk prediction tool
  - To identify patients who are most likely to benefit from IO therapies and those at high risk for severe AEs
  - To develop algorithms to inform drug development, access/regulatory decision-making and to guide prescribing decisions

43

Bayesian Network Machine-Learned Predictive Modelling

A Bayesian network is a **graphical representation** of a set of variables, represented by "nodes", and the relationships between them, represented by "edges".

Machine-learning algorithms are applied to **"learn" the structure** of the network illustrating these relationships in such a way as to maximize fit to the data and predictive power.

The quantitative relationships between variables underlying a structure is encoded within the nodes and edges through **probability tables**.

44

Case Study: Bayesian Networks Machine Learning Approach for Individualized Risk Prediction

Rationale

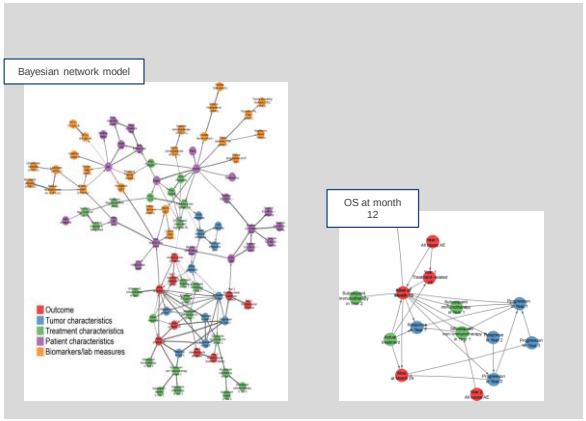
- The client's product was an immunotherapy drug in mRCC
- Trials were characterized by heterogenous treatment response
- The client was interested in developing a "personalized" risk prediction model to identify subsets of patients who respond better to therapy

Analytic approach

- Apply a novel risk prediction tool to understand variation in patient response to immunotherapy in mRCC
- Predicted OS, all-cause AEs and treatment-related AEs based on patient characteristics. Assessed model performance using logistic regression

Results

- Risk scores (MSKCC and Heng), biomarkers (hemoglobin, albumin) and performance status were most prognostic for predicting OS, with ideal patient responders having high EQ-5D-3L scores and health state utility values



45

Application of Machine Learning in Pulmonary Hypertension

Clinical Decision Support Tool for Optimal Therapy using Bayesian Networks

PHORA: Pulmonary Hypertension Outcomes Risk Assessment

46

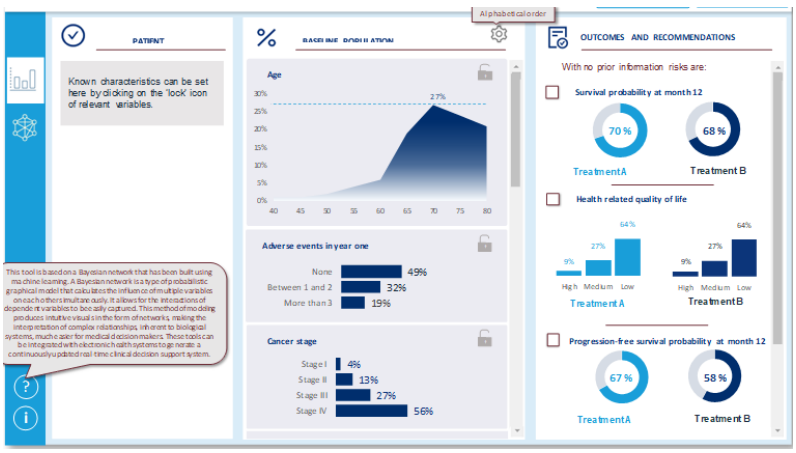
Key Features of PHORA for clinical decision support

- Provides a prognostic tool for physicians to stratify patients
- Assists the clinical team with managing workflow
- Capitalizes on advanced machine learning and data mining technology
- Can be integrated with Electronic Health Records

https://myphora.org/for-physicians

47

Applications for Adaptive Decision Making: Starting with Baseline Population Data

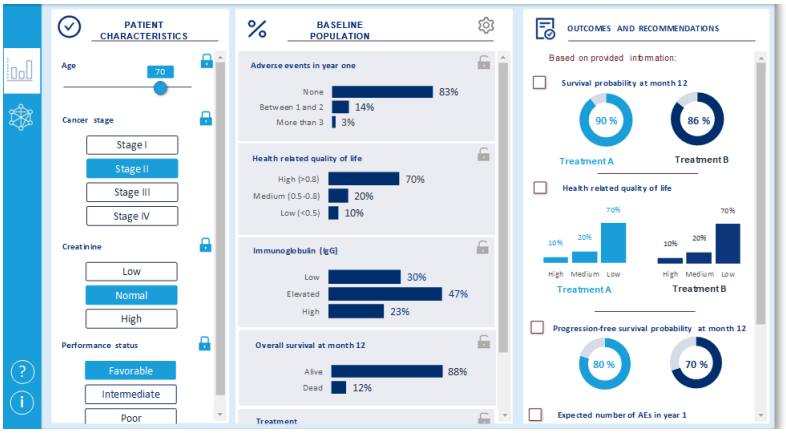


Real-time decision tools

Using individual patient data

48

Applications for Adaptive Decision Making: Leveraging Individual Patient Data



Why Should We Be Open To It? Trends and Some Potential Applications

1. Evolving healthcare ecosystem focused on the use of "Big data"/RWE to augment clinical trial data in regulatory and access decision making is driving the need for advanced statistical methods

Pharmaceutical innovation trending towards on precision medicine/targeted therapies tied to the identification of relevant biomarkers

Reimbursement and access decisions moving from fee-for-service to value-based (US) and general healthcare decisions emphasizing patient centricity

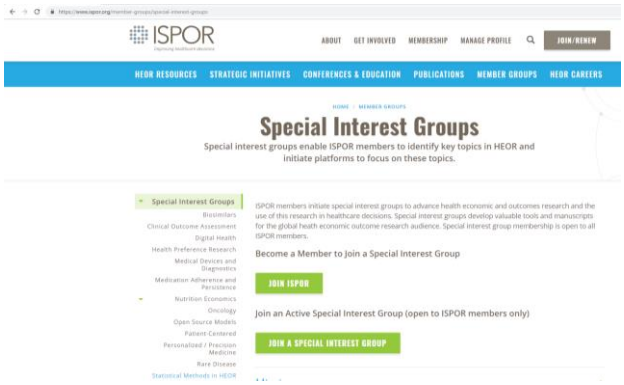
Rising healthcare costs driving the need to better diagnostic tools and optimal prescription

- Support individualized risk prediction and the use of precision medicine
- Characterizing patients who are most likely to benefit from respective therapies
- Providing output for clinical support tools
- Providing output for value-based access discussions
- etc…

Sign up as a Special Interest Group Member

ISPOR Special Interest Group

1. Visit ISPOR webpage www.ispor.org
2. Select “Member Groups”
3. Select “Special Interest Groups”
4. On left side bar, click “Statistical Methods in HEOR”
5. Click button to “Join This Special Interest Group”



For more information, e-mail statisticalmethodssig@ispor.org

51

RECENT WEBINAR – “Principles of Machine Learning for Prediction”

ISPOR Special Interest Group

- **Date:** July 30, 2019
- **Speaker:** David Vanness, PhD, Professor, Penn State University, USA
- **Description:** This webinar introduced learners to the basic principles of machine learning (ML) for prediction in health economics and outcomes research (HEOR). We began by framing the prediction problem and drawing both connections and contrasts with the related problems of explanation and causal inference. We then confronted the “Iron Law of Prediction” – which characterizes the goals of prediction in terms of “minimizing loss” – or the consequences of making poor predictions. We then explored the principles of “cross validation” and the “tuning” of algorithms to minimize loss. Using a simulated dataset, we applied several ML algorithms to illustrate the principles of “complexity penalization,” “bagging,” “random feature selection,” and “boosting.” We concluded the webinar by considering an empirical application of boosting to construct propensity scores – a method of prediction that is used to improve causal inference for other parameters of interest.
- **Learning Objectives:**
  - Describe the Iron Law of Prediction as a tradeoff between bias and variance of predictions to minimize the loss associated with making poor predictions.
  - Describe how cross-validation allows analysts to minimize loss.
  - Describe how the process of tuning ML algorithms is used to optimize their performance in minimizing loss.
  - Describe how the bagging, random feature selection and boosting can be used to minimize loss.
- **Recording link**: https://www.ispor.org/conferences-education/education-training/virtual/webinars

52

UPCOMING WEBINAR – "Machine Learning for HEOR: Prediction and Causal Inference"

ISPOR Special Interest Group

- **Date and Time:** Next Friday, November 15, 2019 at 12:00PM EST
- **Speaker:** Sherri Rose, PhD, Associate Professor, Harvard Medical School, USA
- **Description:** Health care research is moving toward analytic systems that take large health databases and estimate varying quantities of interest both quickly and robustly, incorporating advances from statistics, econometrics, and computer science. This workshop will discuss specific challenges related to developing and deploying statistical machine learning algorithms in prediction and causal inference for health policy research questions, including examples from the areas of health plan payment, mental health care, and medical devices. Considerations go beyond typical measures of statistical assessment, and include concepts such as dataset shift and algorithmic fairness.
- **Learning Objectives:**
  - Understand the shortcomings of standard parametric regression techniques for the estimation of prediction and causal effect quantities
  - Be introduced to the ideas behind machine learning approaches as tools for confronting high-dimensional data
  - Become familiar with the properties and basic conceptual implementation of machine learning for prediction and causal effect estimation
  - Recognize the centrality of fairness and generalizability considerations in quantitative health policy research
- **Registration link**: https://www.ispor.org/conferences-education/education-training/virtual/webinars/machine-learning-for-health-economics-and-outcomes-prediction-and-causal-inference