# Machine Learning, High Energy Physics to Medicine

**Claudio Campagnari UCSB Physics**

**(Main) ML in Medicine Collaborators:**

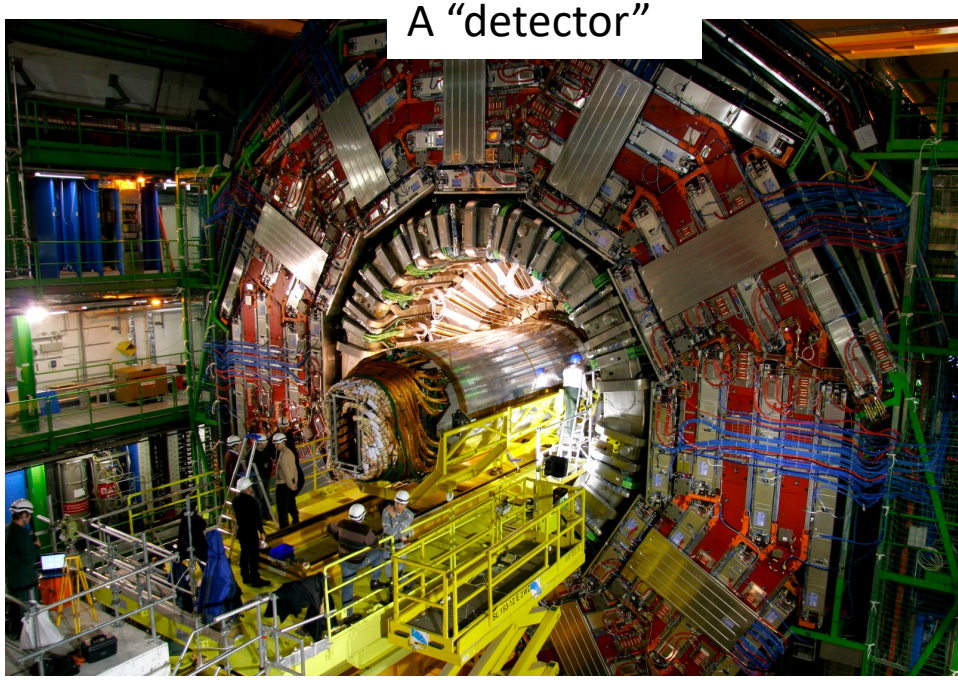
A. Yagil
UCSD Physics
E. Adler, B. Greenberg
UCSD Medical School
K. S. Jering, B. Clagget, S. Solomon
Brigham and Women Hospital and Harvard Medical School
F. Ahmad
Northwestern Medical School

A. A. Voors
University Medical Center, Groeningen (NL)
S. Y. Yang
Kyungpook National University Medicine (S. Korea)
J. J. Park
Seoul National University Medicine (S. Korea)
L. Klein
UCSF Medical School

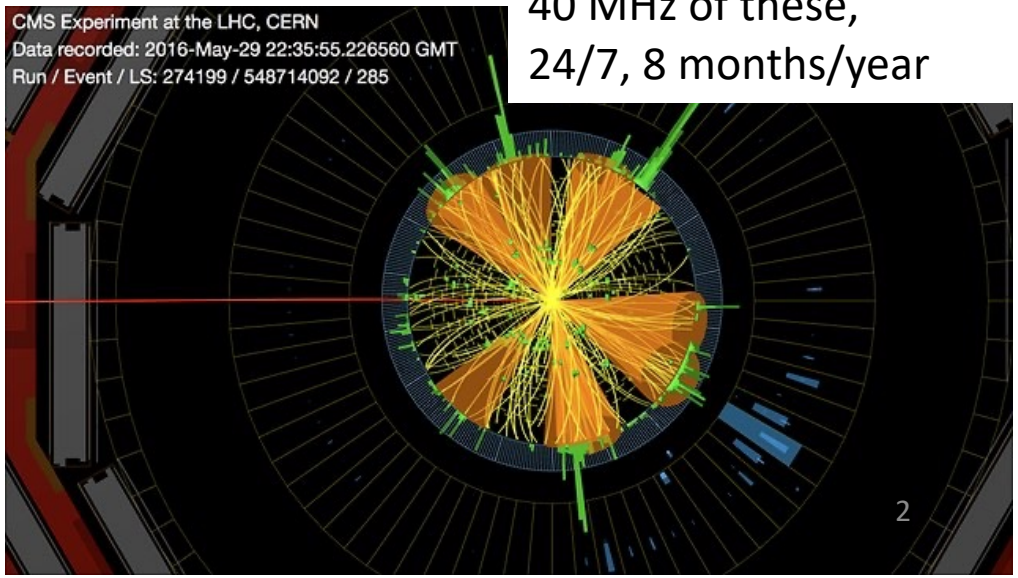
# ML in High Energy Physics (HEP)

- In HEP we have used “ML” for 30+ years
- We want to measure the rate and properties of various physical processes, e.g.,
  - How often does a proton-proton collision result in two top quarks?
  - Are the properties of these two-quark events consistent with theoretical expectations?
- Or, we want to discover new processes, e.g.,
  - The Higgs boson (!)
  - Or some other crazy (but rare!) thing
- Classification problem: “Signal” vs. “Background”
- Recently: ML also for improving measurement precision. In this talk, only classification.

A “detector”



Visualization of one “event”
40 MHz of these,
24/7, 8 months/year

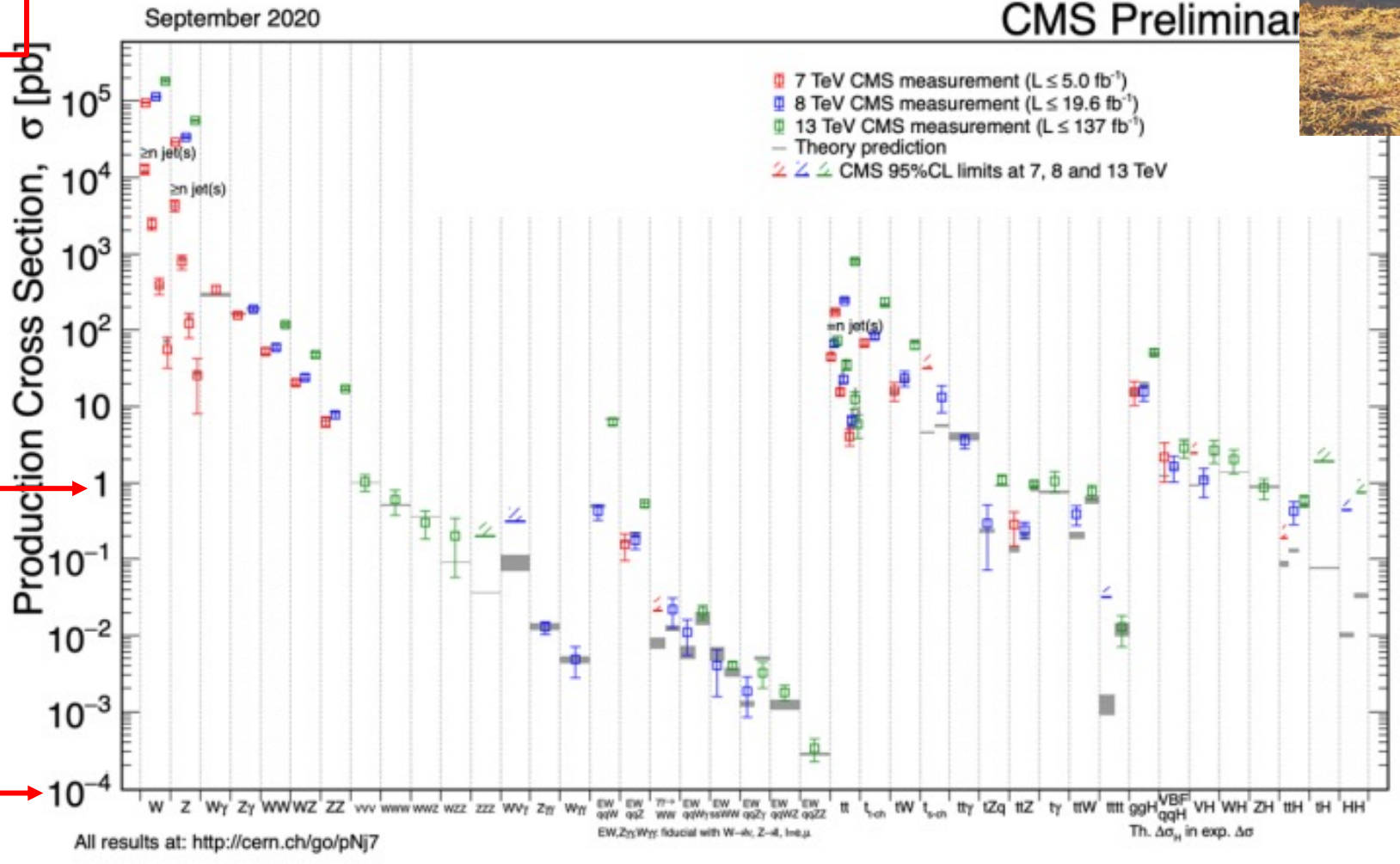


# We are interested in rare processes

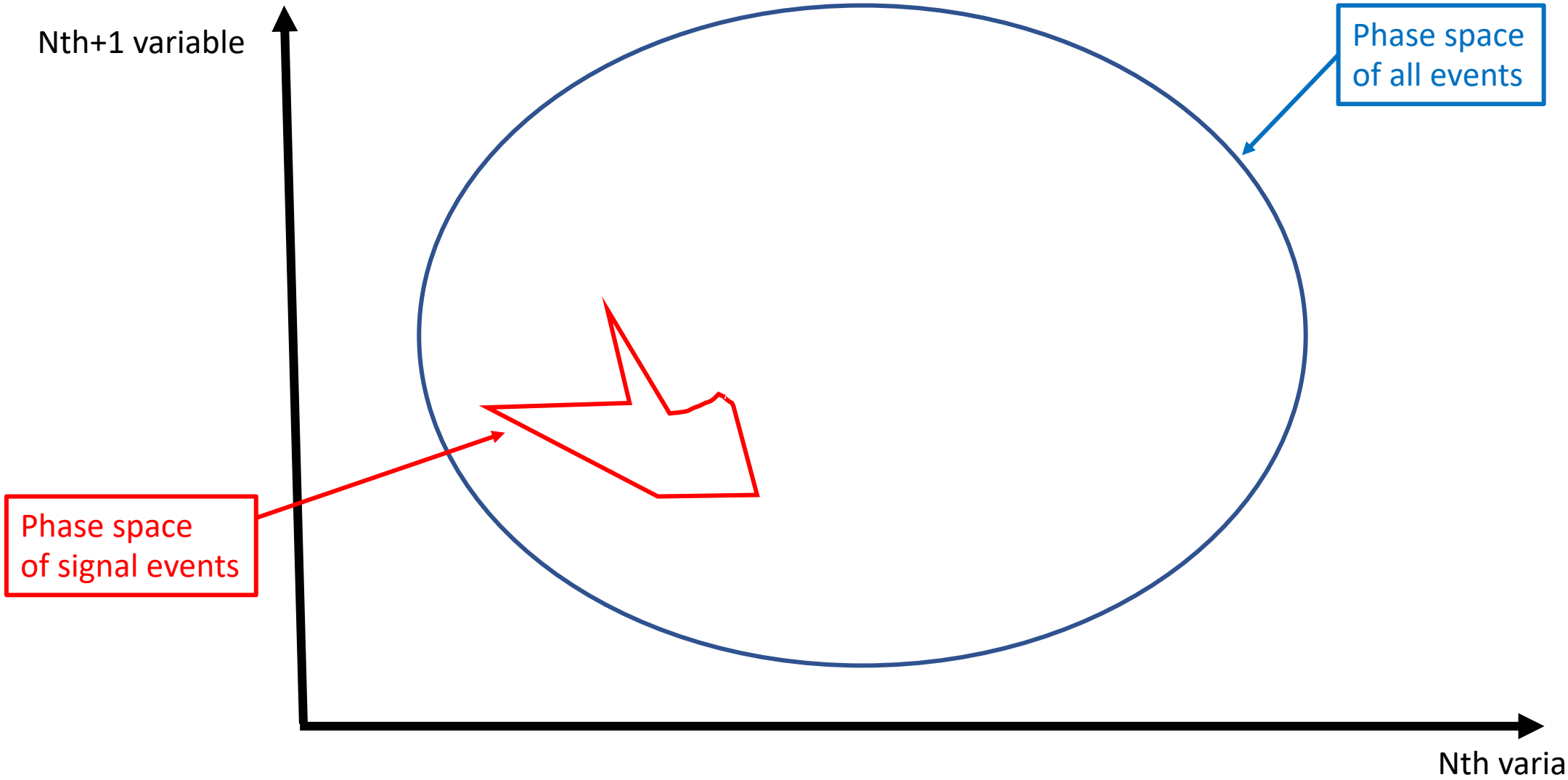
**Total event rate: 40 MHz**

**10 mHZ**

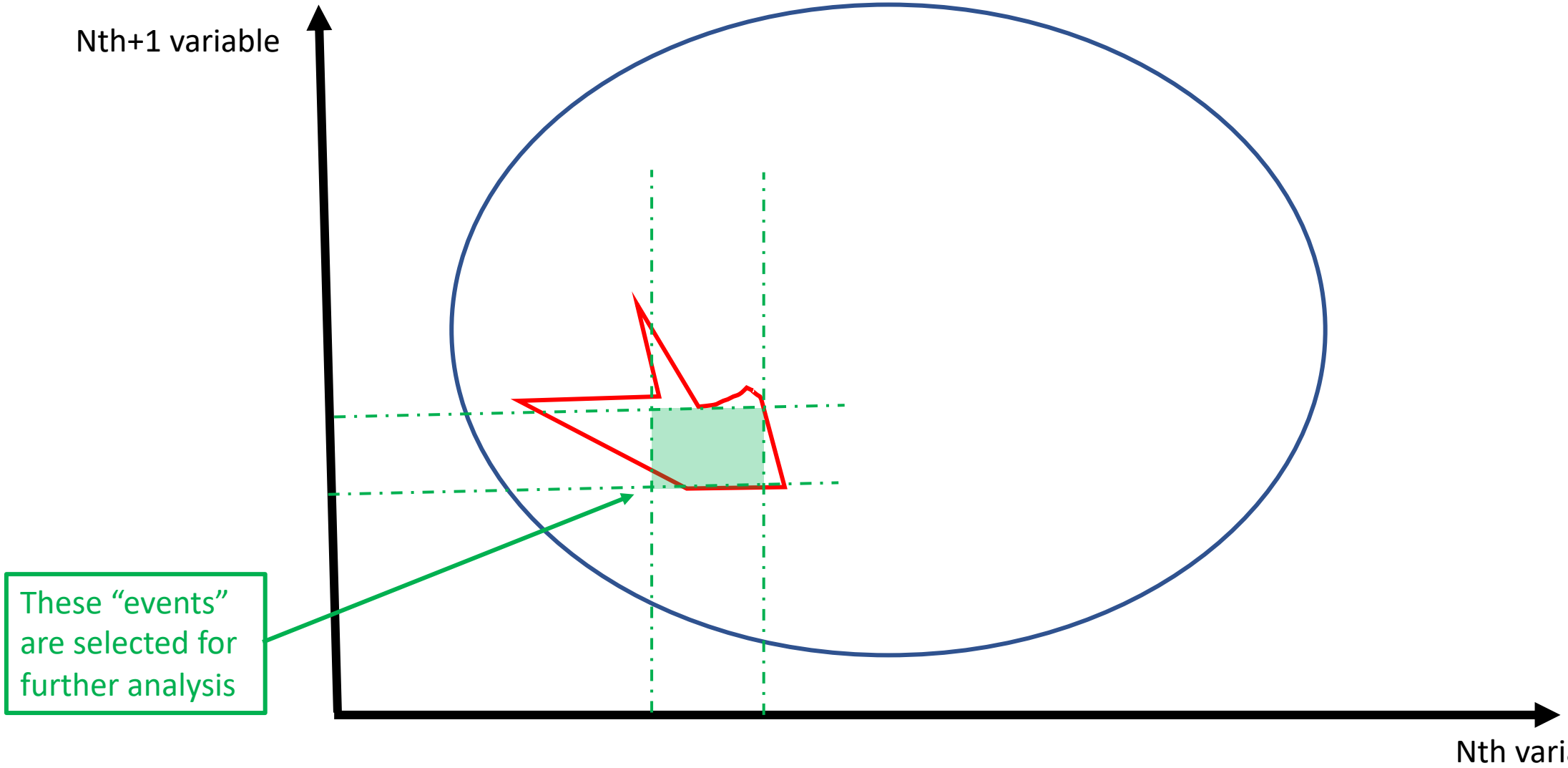
**10/year**



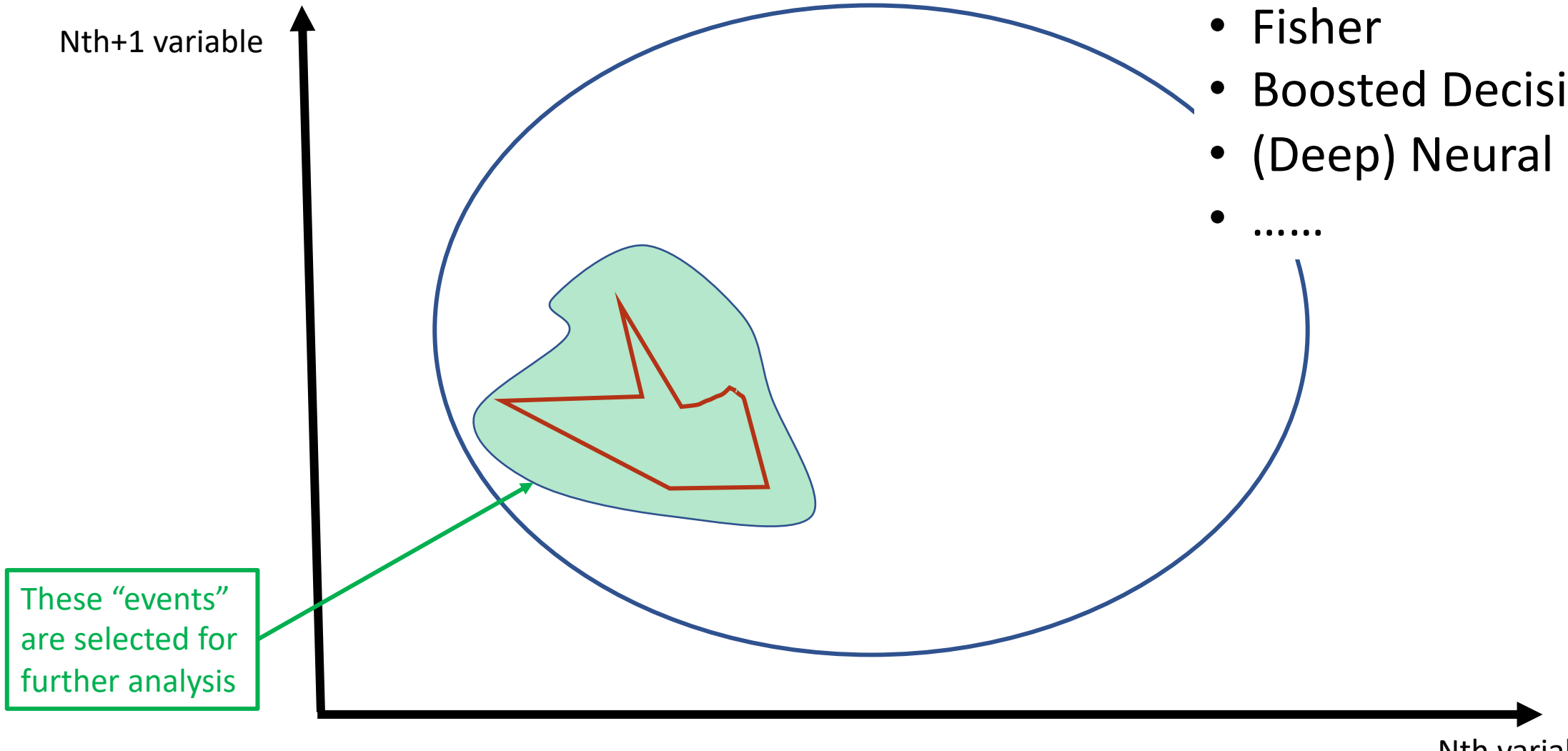
# Cartoon of an HEP event selection



# Old fashioned approach ”square cuts”

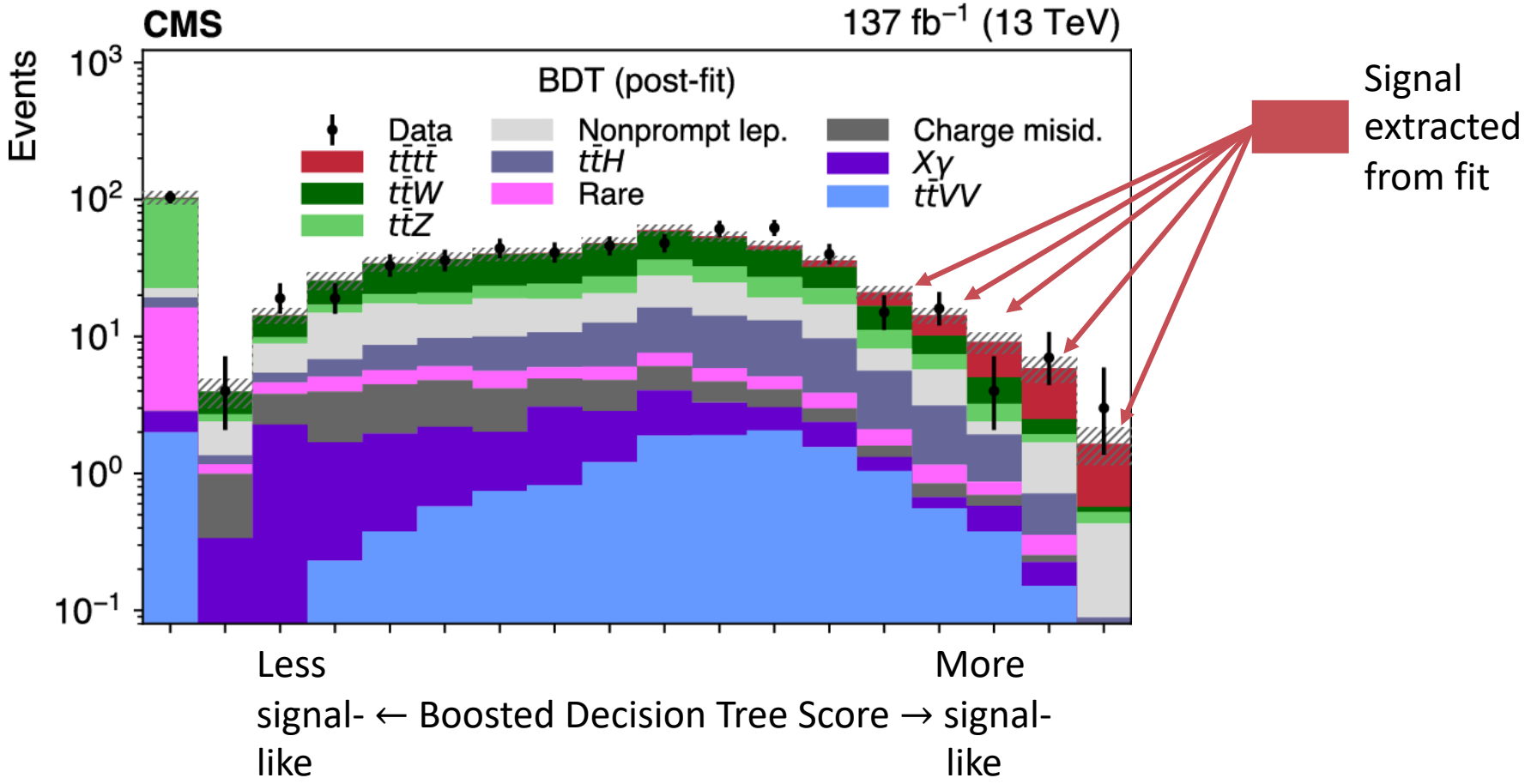


# Multivariate (ML) HEP event selection




- Likelihood
- Fisher
- Boosted Decision Trees
- (Deep) Neural Nets
- ……

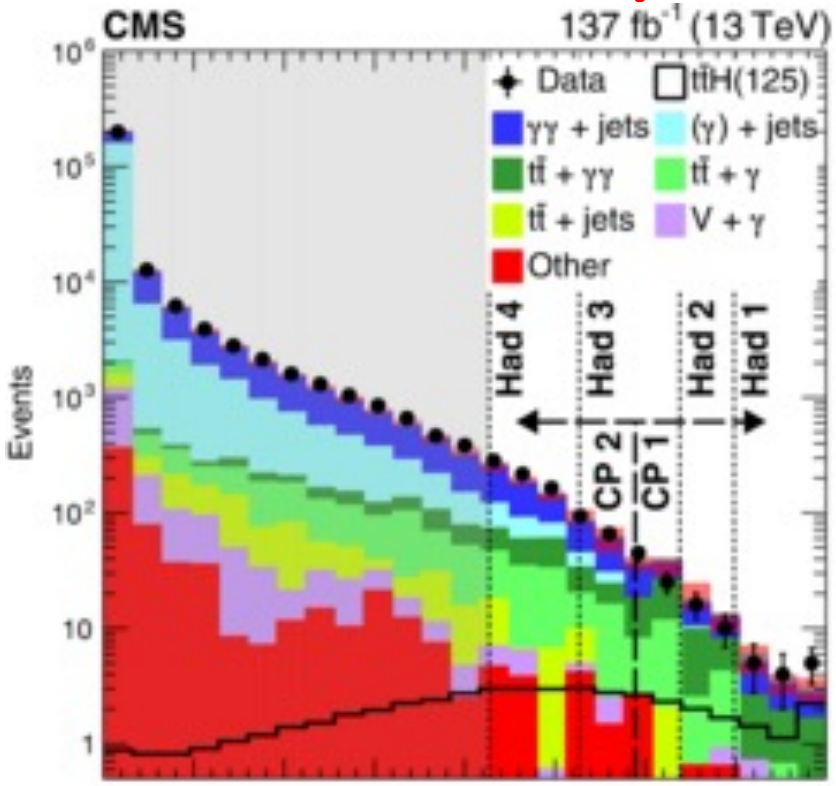
# An example from my HEP research.
# Full ML analysis




Signal extracted from fit

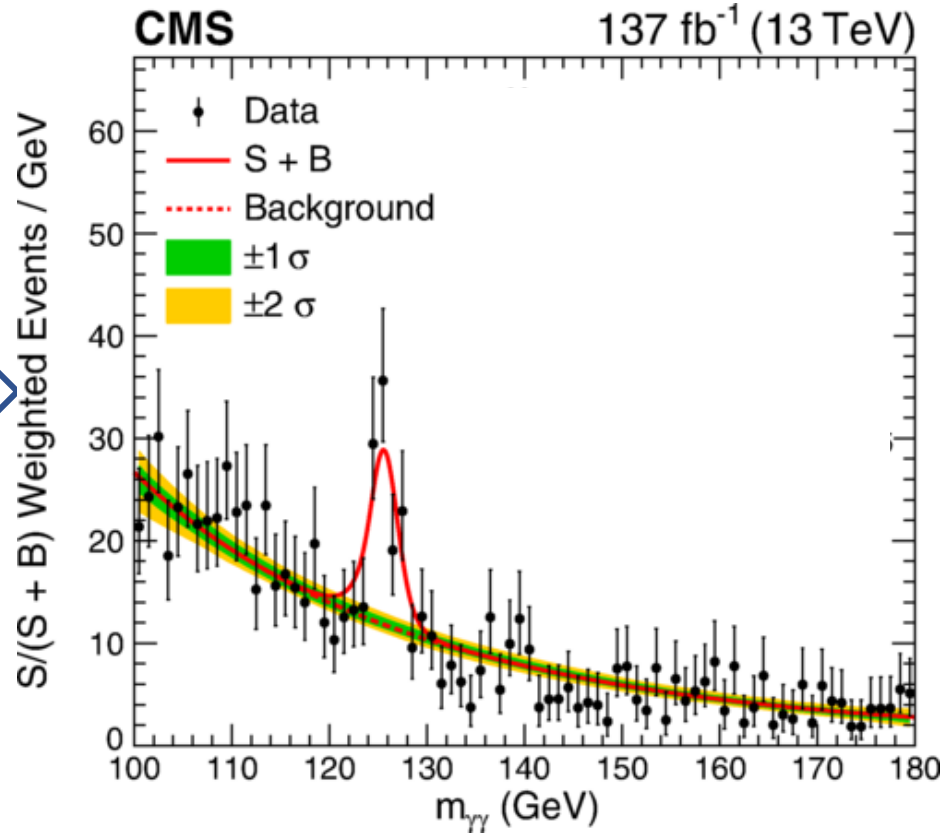
Eur. Phys. J. C80 (2020) no.2, 75

# Another example from my HEP work
# Partial ML analysis (more robust)




Less signal like ← Boosted Decision Tree Score → More signal like

Cut on BDT, then look at most discriminating variable




Phys. Rev. Lett. 125 (2020) 6, 061801

# Multivariate (ML) vs “square cuts”

## Multivariate

- More efficient
- More information (more variables)
- Naturally: each selected event has a “weight”
  - Signal-like vs Background-like
- More opaque
  - Garbage-in-garbage-out

## Square Cuts

- Less efficient
- Not as much information
- Weight of each selected event
  - Not automatic
  - But more under control
- Less opaque

**Bottom line: multivariate approaches ~ 10% to factor 2 better**

# ML and Cardiology

Menu

UC San Diego

UC SAN DIEGO **TODAY**

# Artificial Intelligence Tool Predicts Life Expectancy in Heart Failure Patients

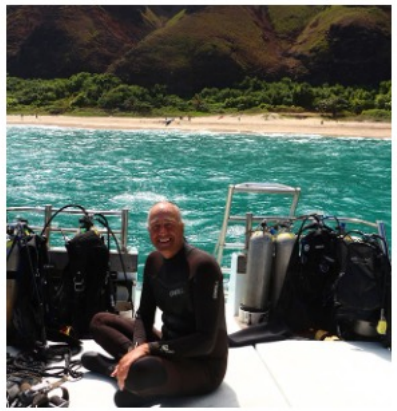
Algorithm developed by physicists and cardiologists achieved 88 percent success rate

**By:**

Michelle Brubaker

**November 13, 2019**

When Avi Yagil, PhD, Distinguished Professor of Physics at University of California San Diego flew home from Europe in 2012, he thought he had caught a cold from his travels. When a "collection of pills" did not improve his symptoms, his wife encouraged him to see a doctor.



*Avi Yagil, PhD, Distinguished Professor of Physics at University of California San Diego, back to his hobbies after a heart transplant.*

Further tests revealed something far more life-threatening to Yagil than the common cold. "A chest X-Ray showed my lungs were flooded with fluid, and a subsequent echocardiogram found I had damage to my heart."

Yagil was diagnosed with heart failure. "UC San Diego Health cardiologists tried to manage my condition with medication, but all systems were failing as my heart struggled to keep me alive."

In June 2016, Yagil received a heart transplant. "I consider June 17 my second birthday."

While Yagil recovered from surgery, he began thinking about how he could improve the process for patients like him.

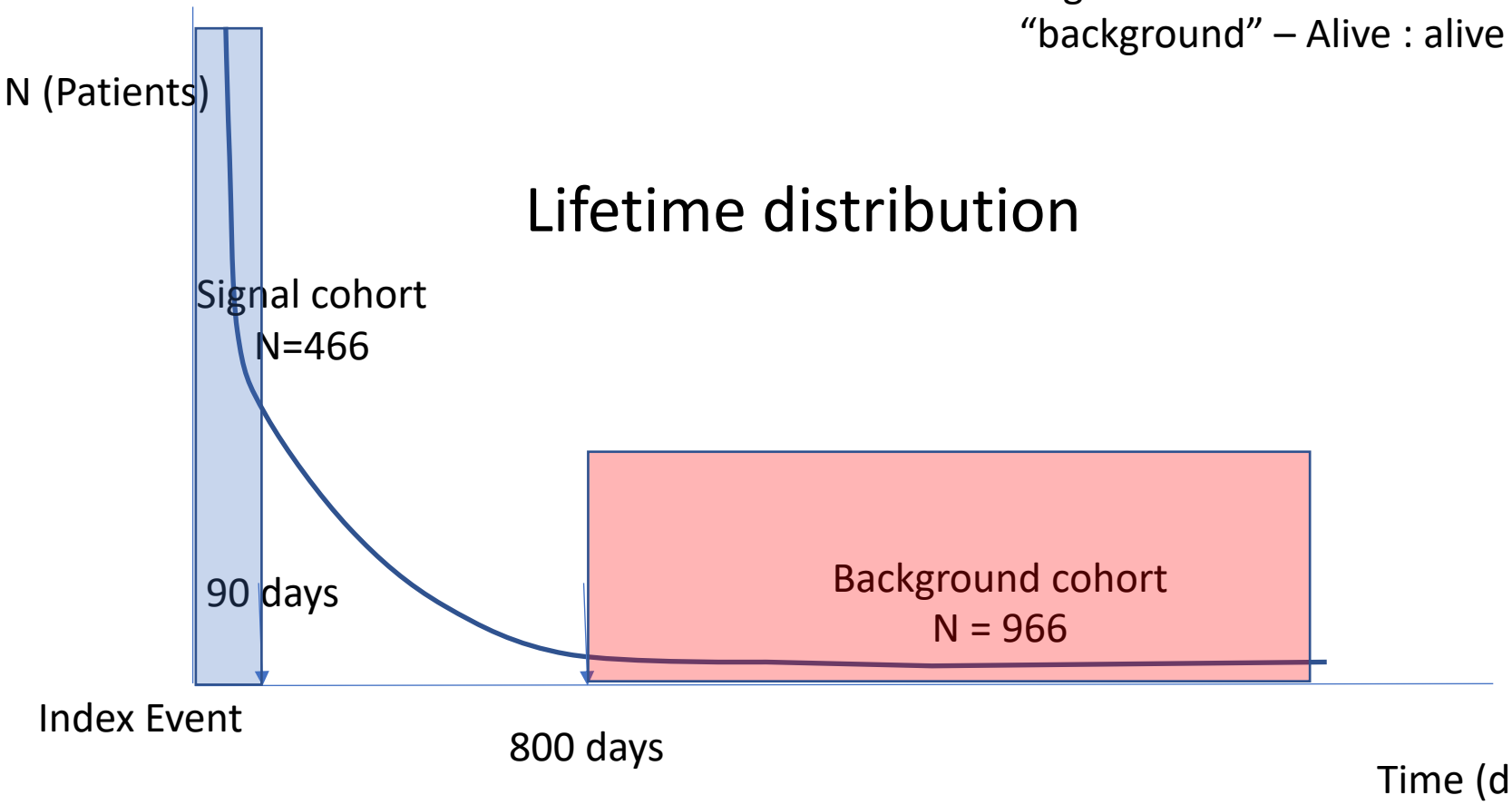
"In my day job, I use machine learning to understand a vast amount of information and measurements of particles and how they interact," he said. "The human body is even more complex, but the

# Marker-HF: a risk score for Heart Failure (HF) patients

- Boosted Decision Tree (AdaBoost)
- Based on Electronic Health Record (EHR) of UCSD Medicine
  - Challenging
- Precise definition of outcomes
- No imputation in algorithm design
- Small number of inputs (8)
  - Ease of use
  - Not enough patient statistics to do anything super-sophisticated
- Strict temporal requirements on data collection

# Cartoon of samples definitions:

"signal" – Dead: died less then 90 days after index event
"background" – Alive : alive more than 2 years after index event




**Variables Used:**

Diastolic blood pressure
Creatinine
Blood Urea Nitrogen
Hemoglobin
White blood cell count
Platelets
Albumin
Red Blood Cell Distribution Width

# Input Variables

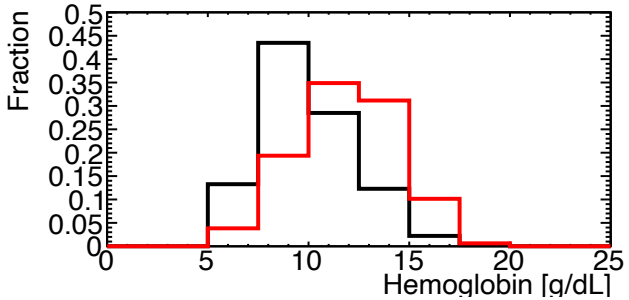
Red - Low risk cohort

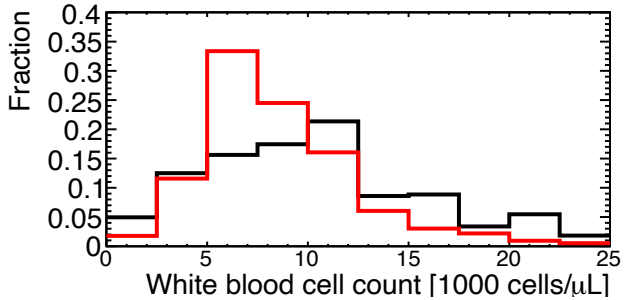
Black - High risk cohort

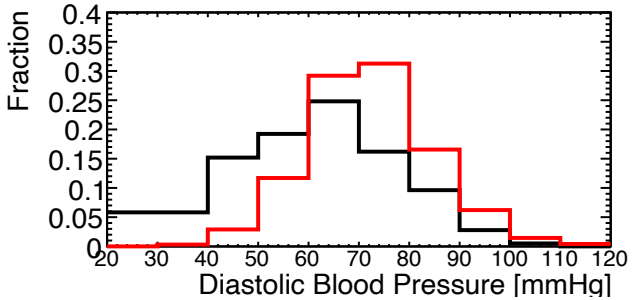
No ”silver bullet”:

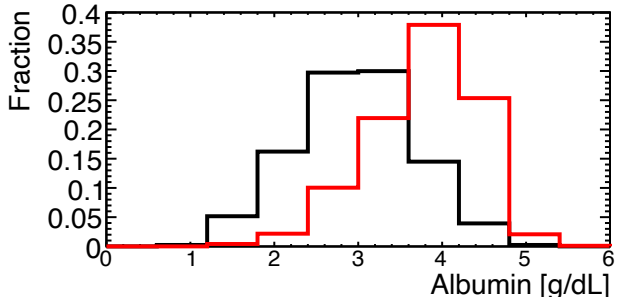
- No single great discriminating variable.
- Each shows some separation.

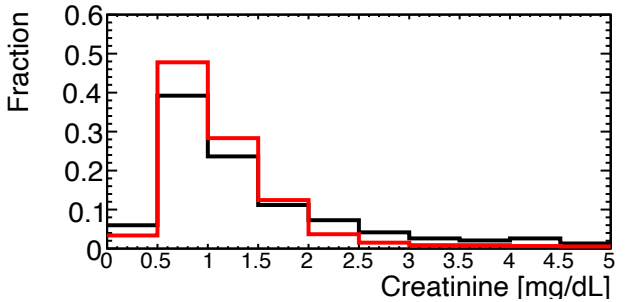
Key is the combination and correlations between the whole set

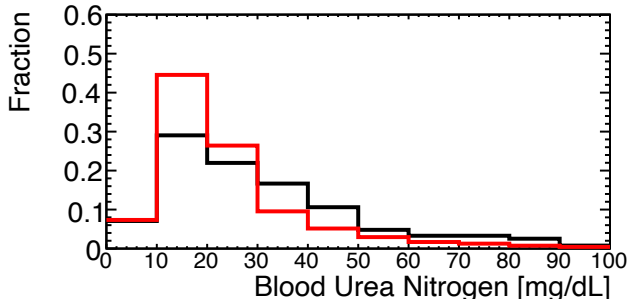


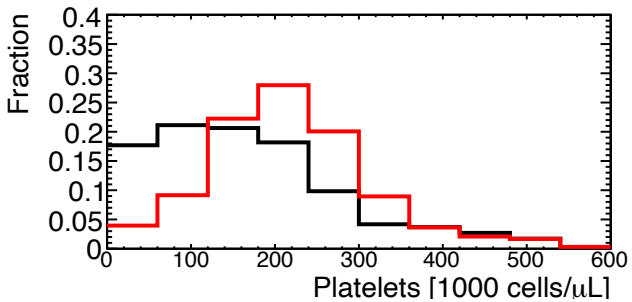



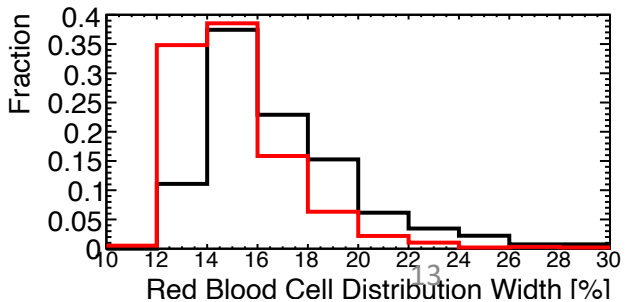



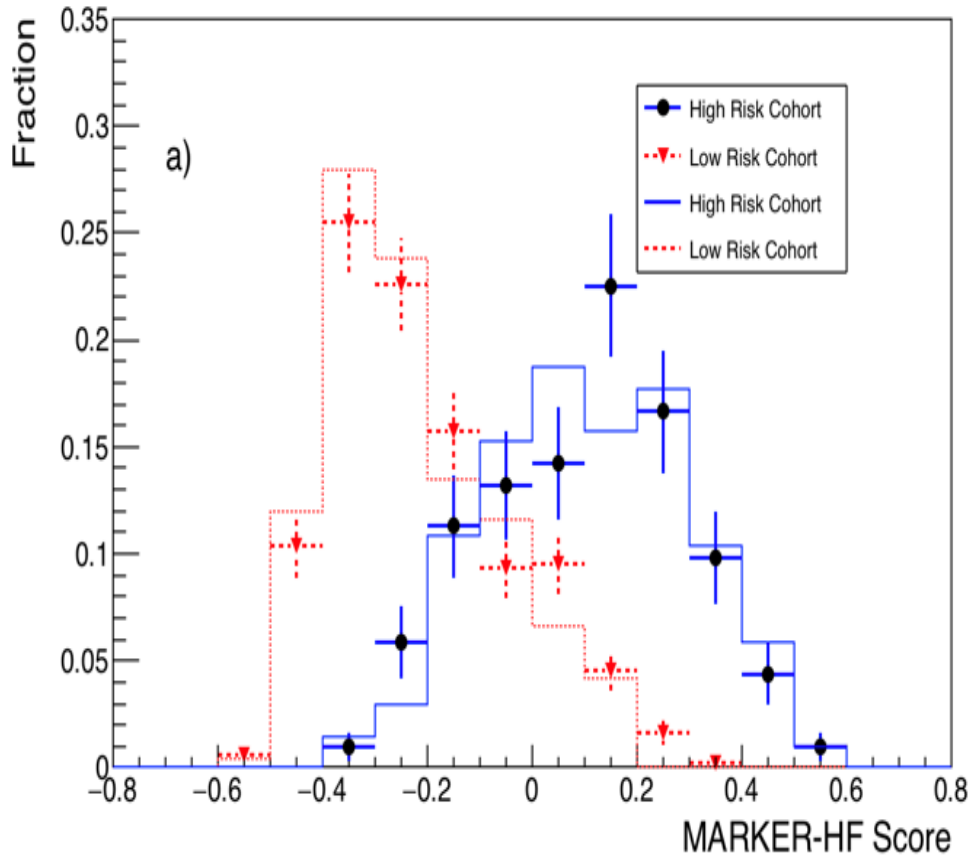



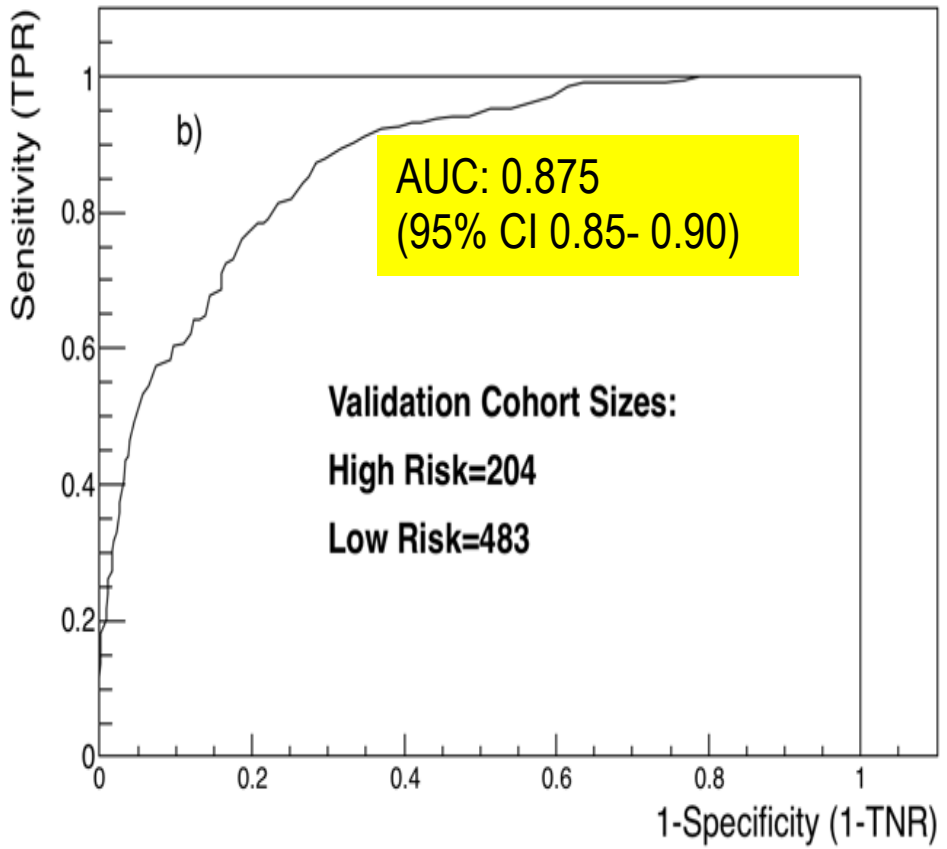



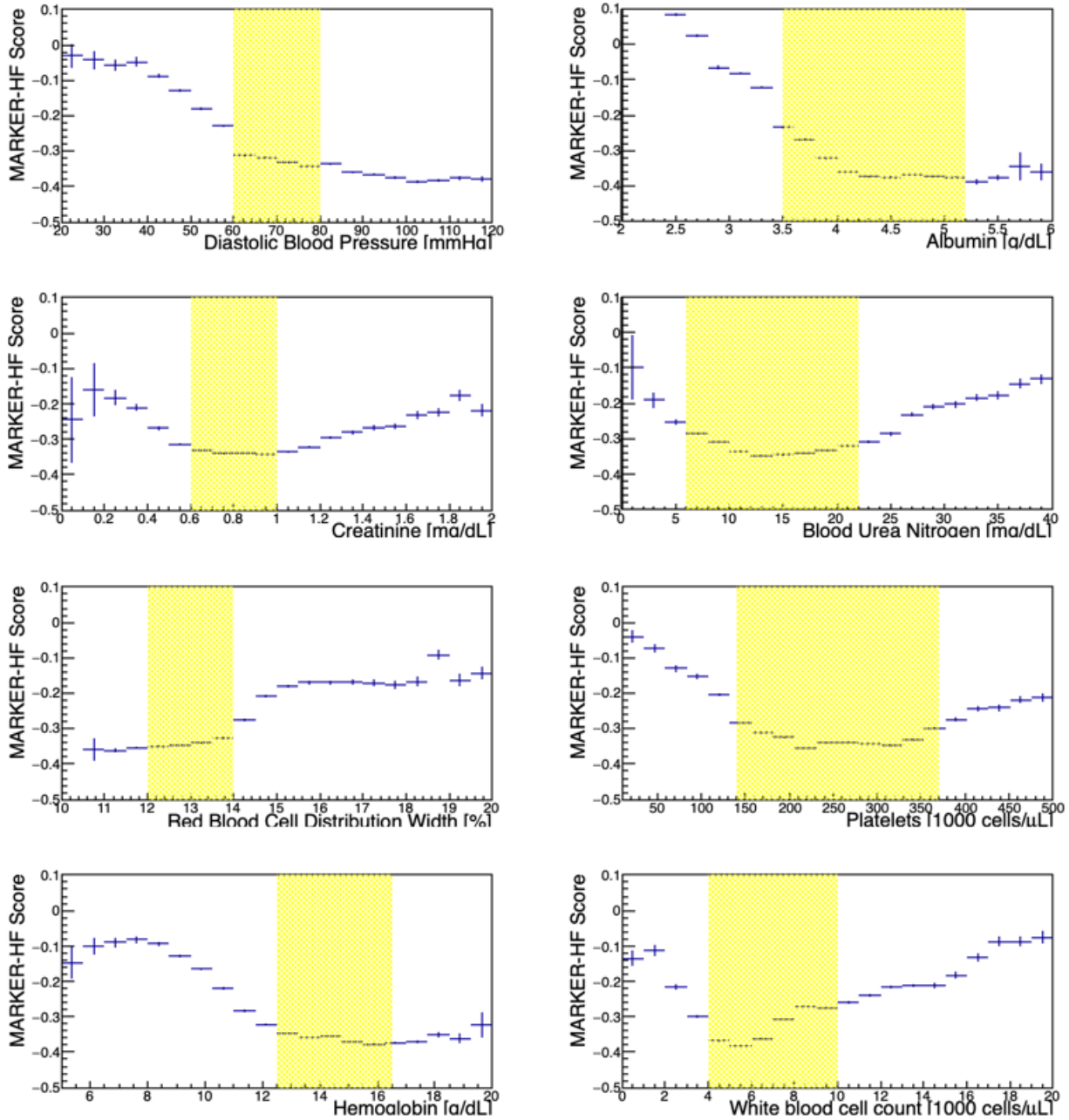
# MARKER-HF Training and Performance

A Boosted Decision Tree algorithm (Anaboost, 200 trees, maximum depth of 2) to derive a model and relate variables to the known outcome using the training subset of the sample only. Similar results obtained with NN.  Even with Fisher Discriminant (!)

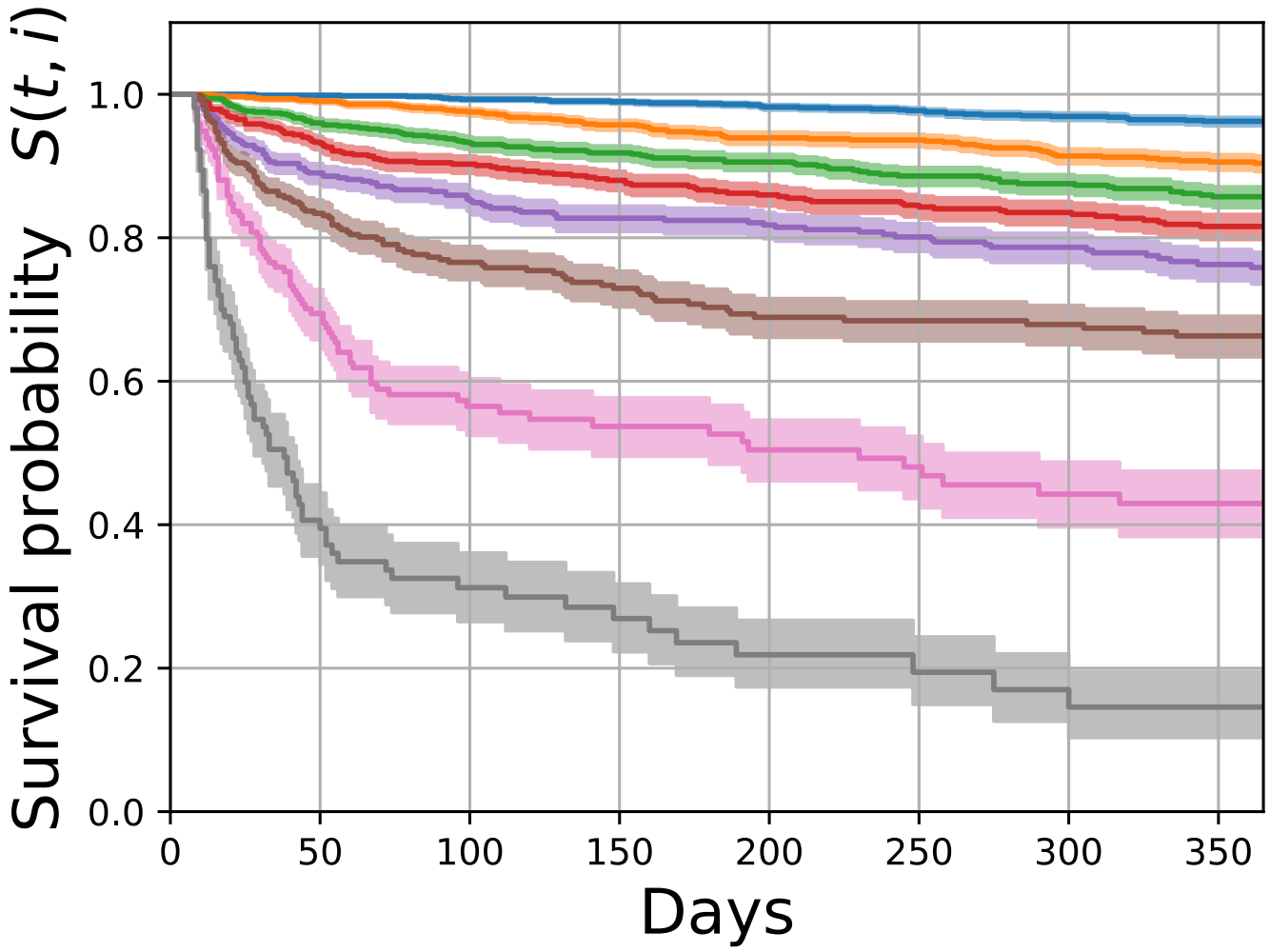


The algorithm figures out automatically what the “healthy” ranges of the covariates are.

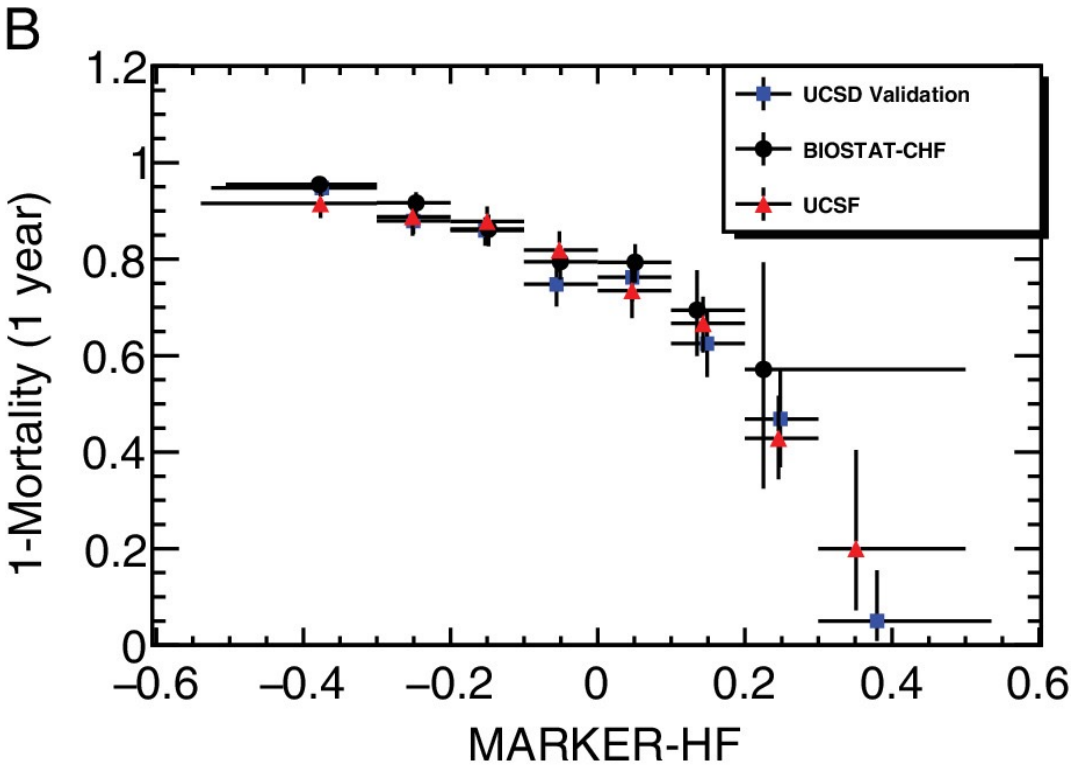


# Mortality in Marker-HF strata



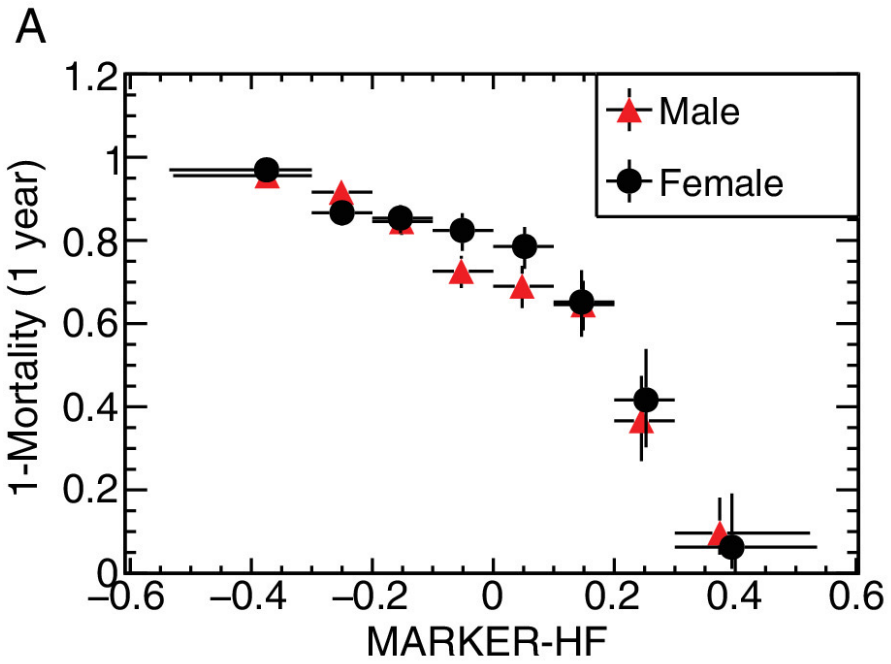

| | |
|---|---|
| bdt bin 0 | marker < -0.3 |
| bdt bin 1 | -0.3<marker<-0.2 |
| bdt bin 2 | -0.2<marker<-0.1 |
| bdt bin 3 | -0.1<marker<0.0 |
| bdt bin 4 | 0.0<marker<0.1 |
| bdt bin 5 | 0.1<marker<0.2 |
| bdt bin 6 | 0.2<marker<0.3 |
| bdt bin 7 | marker>0.3 |

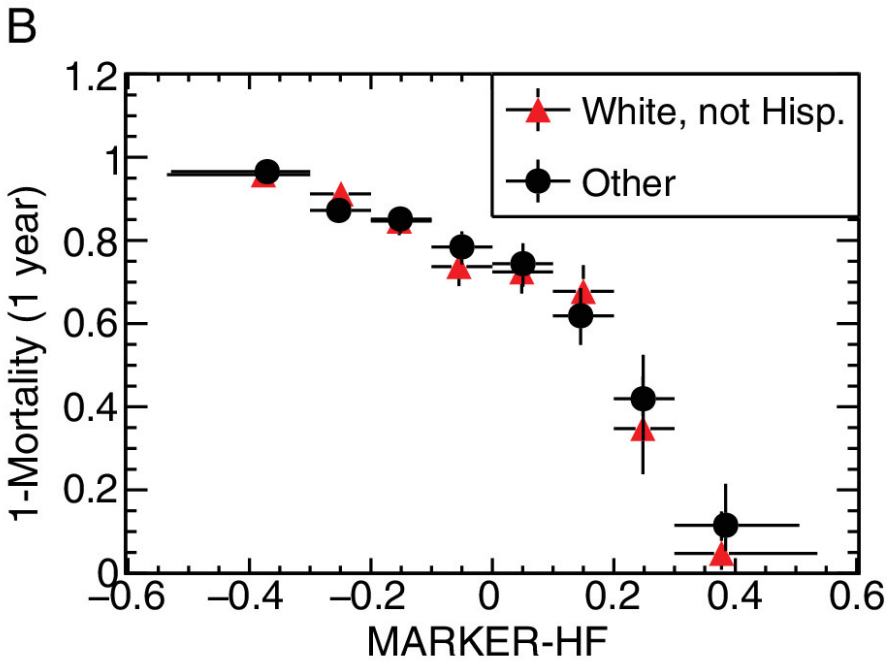
# External (outside UCSD) validation

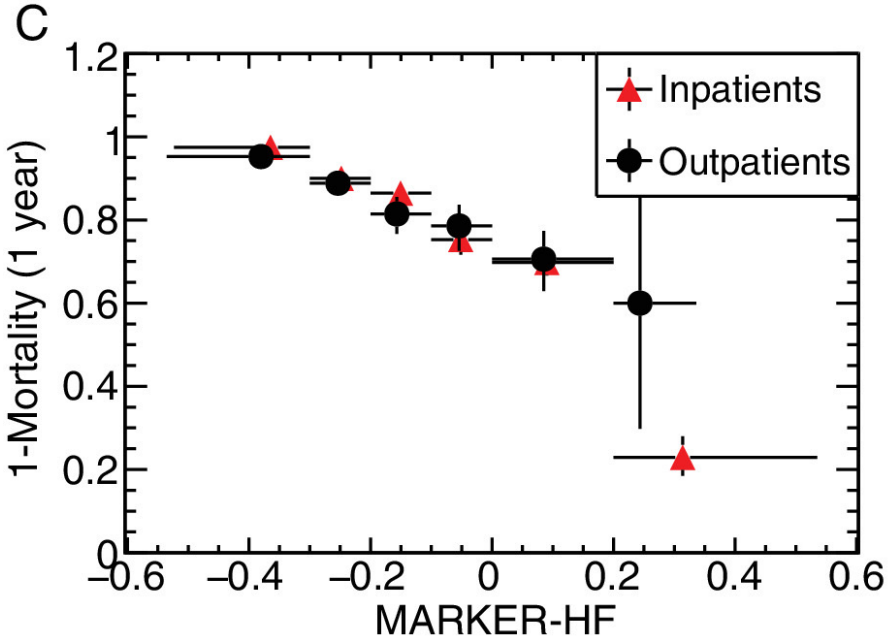


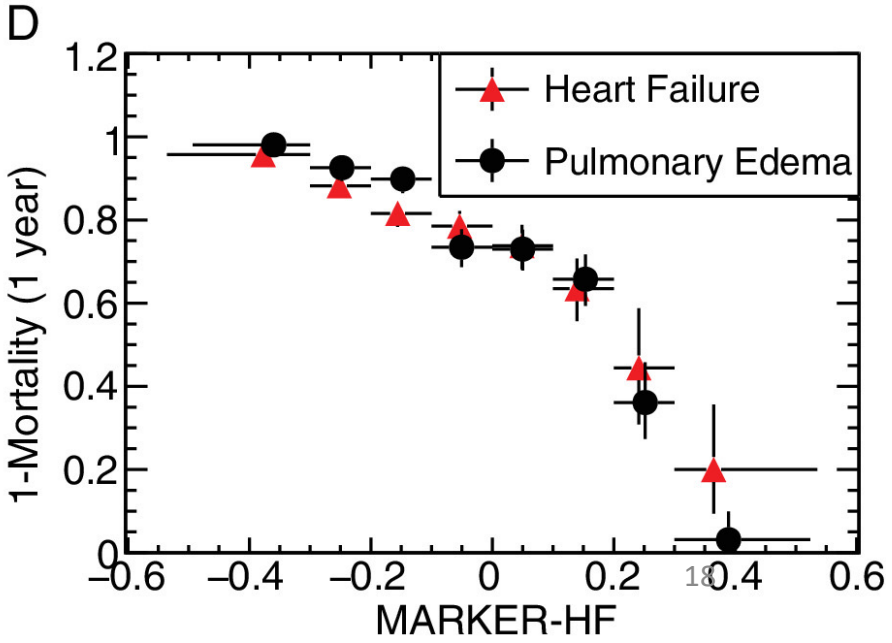

| Cohort | High-risk, *n* | Low-risk, *n* | AUC | 95% CI |
|---|---|---|---|---|
| UCSD (all variables) | 204 | 483 | 0.88 | 0.85–0.90 |
| UCSD (RDW imputed) | 204 | 483 | 0.87 | 0.84–0.89 |
| UCSF[a] | 135 | 330 | 0.81 | 0.77–0.86 |
| BIOSTAT-CHF[a] | 35 | 228 | 0.84 | 0.78–0.90 |

# Demographics



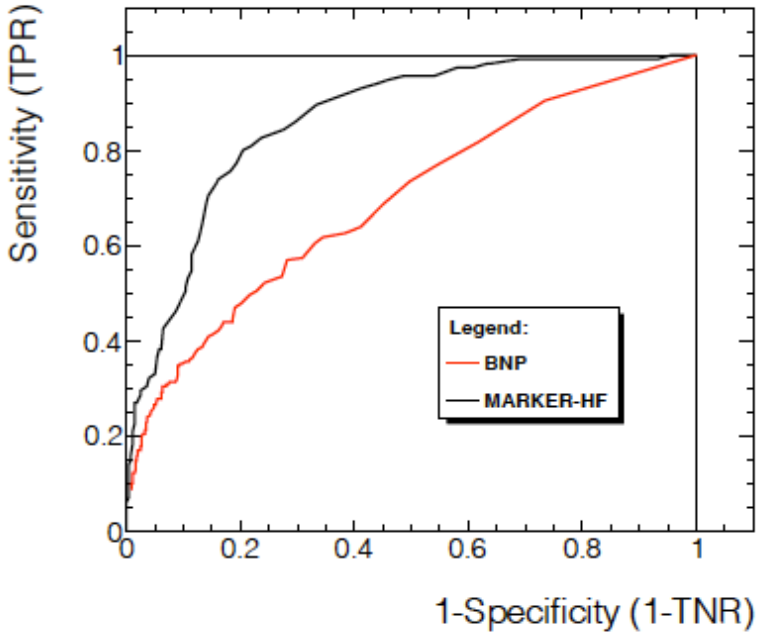



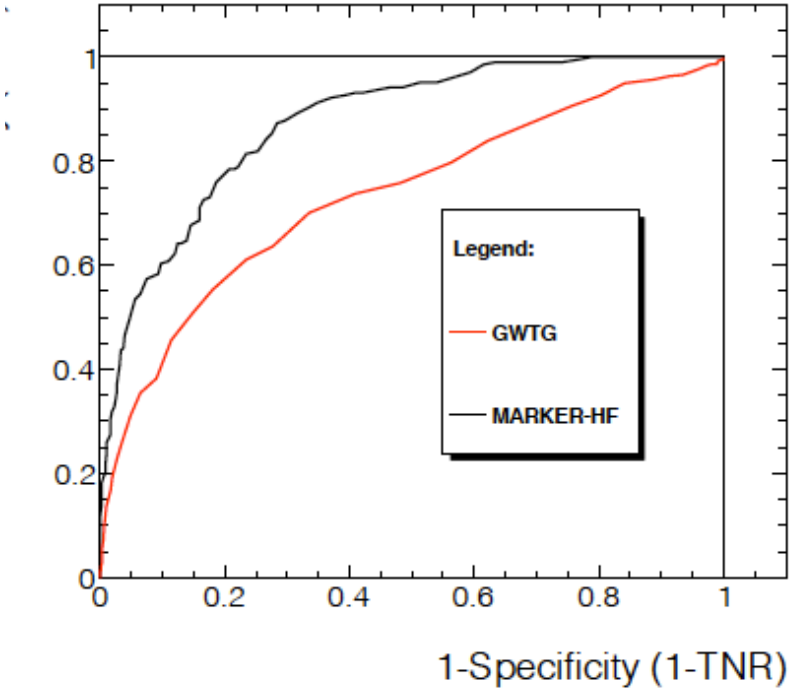
# Comparison with other Risk Scores/Markers

N-terminal pro-hormone BNP



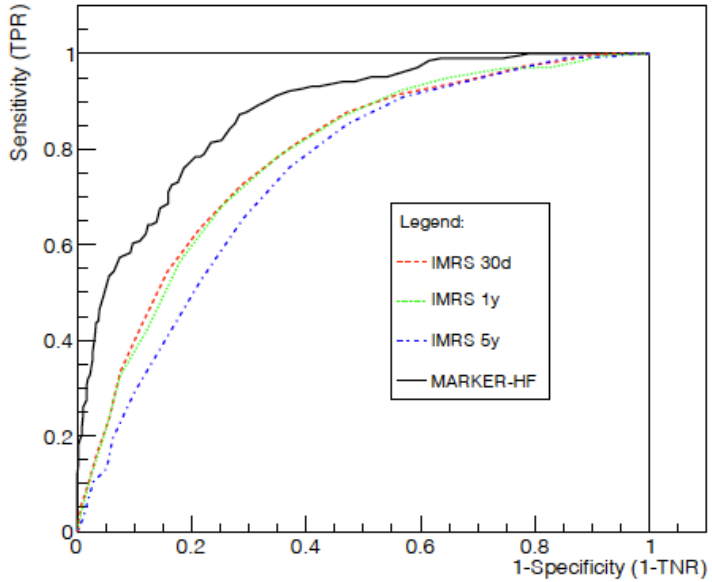

AUC(NT-BNP) = 0.69

Get-With-The-Guideline Score




AUC(GTWG) = 0.73

Inter Mountain Risk Score




AUC(IMRS) = 0.75 – 0.78

AUC(MARKER-HF) = 0.88

# Why is Marker-HF "better"

- IMRS score: 15 variables vs. 8 for MARKER-HF
- IMRS: very simple algorithm
- No correlations!

**Table 2** Sex-specific Values[a] Are Used to Calculate the Intermountain Risk Score as the Sum of an Individual's Corresponding Values from Each Component at a Given Time Point

| | Female | | | Male | | |
|---|---|---|---|---|---|---|
| Component | 30-d | 1-y | 5-y | 30-d | 1-y | 5-y |
| Hematocrit ≤ 34.6 | 1 | 1 | 2 | 2 | 3 | 3 |
| 34.7-38.2 | 0 | 0 | 1 | 2 | 2 | 3 |
| 38.3-41.0 | 0 | 0 | 0 | 1 | 1 | 2 |
| 41.1-44.1 | 0 | 0 | 0 | 0 | 1 | 1 |
| ≥44.2 | 0 | 0 | 1 | 0 | 0 | 0 |
| White blood cell count | | | | | | |
| ≤5.9 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6.0-7.3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7.4-8.9 | 1 | 0 | 0 | 0 | 1 | 1 |
| 9.0-11.2 | 2 | 1 | 1 | 2 | 2 | 1 |
| ≥11.3 | 4 | 3 | 2 | 4 | 3 | 2 |
| Platelet count ≤ 183 | 2 | 1 | 2 | 2 | 1 | 1 |
| 184-220 | 1 | 0 | 0 | 1 | 0 | 0 |
| 221-254 | 1 | 0 | 1 | 0 | 0 | 0 |
| 255-300 | 0 | 0 | 1 | 1 | 1 | 0 |
| ≥301 | 0 | 0 | 1 | 1 | 1 | 1 |
| Mean corpuscular volume | | | | | | |
| ≤86.3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 86.4-89.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 89.2-91.4 | 1 | 0 | 0 | 0 | 0 | 0 |
| 91.5-94.0 | 0 | 0 | 1 | 0 | 0 | 0 |
| ≥94.1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Mean corpuscular hemoglobin concentration | | | | | | |
| ≤33.3 | 1 | 1 | 0 | 1 | 1 | 0 |
| 33.4-33.8 | 0 | 0 | 0 | 0 | 1 | 0 |
| 33.9-34.2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 34.3-34.6 | 0 | 0 | 0 | 0 | 0 | 1 |
| ≥34.7 | 0 | 0 | 0 | 0 | 0 | 1 |
| Red cell distribution width | | | | | | |
| ≤12.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12.6-13.0 | 2 | 1 | 1 | 1 | 0 | 0 |
| 13.1-13.5 | 1 | 1 | 2 | 1 | 1 | 2 |
| 13.6-14.3 | 3 | 2 | 2 | 2 | 2 | 3 |
| ≥14.4 | 4 | 4 | 5 | 3 | 3 | 4 |
| Mean platelet volume | | | | | | |
| ≤7.5 | 1 | 1 | 1 | 1 | 1 | 0 |
| 7.6-8.0 | 1 | 0 | 1 | 1 | 0 | 0 |
| 8.1-8.4 | 1 | 0 | 0 | 2 | 0 | 0 |
| 8.5-9.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| ≥9.2 | 0 | 0 | 0 | 1 | 0 | 0 |
| Sodium ≤ 138 | 1 | 1 | 2 | 1 | 1 | 2 |
| 139 | 0 | 0 | 1 | 1 | 0 | 0 |
| 140-141 | 0 | 0 | 1 | 0 | 0 | 0 |
| 142 | 0 | 0 | 0 | 1 | 0 | 0 |
| ≥143 | 1 | 1 | 0 | 2 | 1 | 0 |
| Potassium ≤ 3.7 | 1 | 1 | 1 | 2 | 0 | 0 |
| 3.8-3.9 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4.0-4.1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 4.2-4.4 | 0 | 0 | 0 | 0 | 0 | 0 |
| ≥4.5 | 1 | 0 | 1 | 1 | 0 | 0 |
| Bicarbonate ≤ 23 | 3 | 1 | 1 | 4 | 2 | 1 |
| 24-25 | 1 | 0 | 0 | 2 | 0 | 0 |
| 26 | 1 | 0 | 0 | 1 | 0 | 0 |
| 27-28 | 0 | 0 | 0 | 0 | 0 | 0 |
| ≥29 | 2 | 1 | 1 | 1 | 1 | 1 |
| Calcium ≤ 8.5 | 4 | 3 | 3 | 1 | 2 | 2 |
| 8.6-8.9 | 2 | 2 | 2 | 0 | 1 | 2 |
| 9.0-9.2 | 2 | 1 | 1 | 0 | 0 | 1 |
| 9.3-9.5 | 0 | 0 | 0 | 0 | 1 | 0 |
| ≥9.6 | 1 | 1 | 0 | 0 | 0 | 0 |
| Glucose ≤ 85 | 1 | 0 | 0 | 1 | 1 | 0 |
| 86-94 | 0 | 0 | 0 | 0 | 0 | 0 |
| 95-104 | 1 | 0 | 1 | 1 | 1 | 0 |
| 105-125 | 1 | 1 | 1 | 2 | 1 | 1 |
| ≥126 | 3 | 2 | 2 | 3 | 2 | 1 |
| Creatinine ≤ 0.8 | 0 | 1 | 1 | 2 | 3 | 2 |
| 0.9 | 0 | 0 | 1 | 1 | 1 | 1 |
| 1.0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1.1-1.2 | 1 | 1 | 1 | 0 | 0 | 0 |
| ≥1.3 | 2 | 2 | 3 | 2 | 2 | 1 |
| Age (y) | | | | | | |
| 18-29 | −3 | −5 | −5 | 1 | 0 | 0 |
| 30-39 | −2 | −1 | −1 | 1 | −1 | 0 |
| 40-49 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50-59 | 1 | 1 | 1 | 1 | 1 | 1 |
| 60-69 | 2 | 2 | 3 | 1 | 1 | 2 |
| 70-79 | 2 | 3 | 4 | 2 | 2 | 3 |
| ≥80 | 5 | 6 | 8 | 4 | 5 | 7 |
| Sex | | | | | | |
| Female | 0 | 0 | 0 | — | — | — |
| Male | — | — | — | 0 | 0 | 0 |

[a]Risk models and component values are Copyright © 2006-2008, IHC Health Services, Inc (freely available for academic use).

ESC European Society of Cardiology

European Journal of Heart Failure (2020) **22**, 139–147
doi:10.1002/ejhf.1628

**RESEARCH ARTICLE**

# Improving risk prediction in heart failure using machine learning

**Eric D. Adler[1], Adriaan A. Voors[2], Liviu Klein[3], Fima Macheret[4], Oscar O. Braun[5], Marcus A. Urey[1], Wenhong Zhu[4], Iziah Sama[2], Matevz Tadel[6], Claudio Campagnari[7†], Barry Greenberg[1*†], and Avi Yagil[1,6]**

[1]Division of Cardiology, Department of Medicine, UC San Diego, La Jolla, CA, USA; [2]University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [3]Division of Cardiology, Department of Medicine, UC San Francisco, San Francisco, CA, USA; [4]Altman Clinical and Translational Research Institute (ACTRI), UC San Diego, La Jolla, CA, USA; [5]Cardiology, Department of Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden; [6]Physics Department, UC San Diego, La Jolla, CA, USA; and [7]Physics Department, UC Santa Barbara, Santa Barbara, CA, USA


# MARKER-HF™ Calculator

## Terms for Licensed Healthcare Professionals

I, Enter Your Name (hereinafter also referred to as "You"), represent and warrant that I am currently a licensed healthcare professional in -- Select Location -- .

In exchange for the representations, warranties, and promises contained herein, Eric Adler, Claudio Campagnari, Barry Greenberg, and Avi Yagil, (collectively "Developers") and UC San Diego give You permission to use the MARKER-HF Calculator on UC San Diego's website.

You understand and agree that the MARKER-HF Calculator is intended for use only by licensed healthcare professionals knowledgeable in the field of medicine. You also understand and agree that the information, data, results, and output of the MARKER- HF Calculator are not intended to be used as a decision-making tool and/or as a replacement for your professional expertise and/or judgment. UC San Diego and Developers make no representations nor warranty with regard to the accuracy, validity, completeness, and/or availability of information or results obtained from the MARKER-HF Calculator, nor should it be construed to indicate that any treatment, treatment combination, and/or outcome is safe, expected, appropriate, and/or warranted for any given patient.

The MARKER-HF Calculator is distributed "as is" as a public service, without any additional service, support, and/or licensing by UC San Diego and/or Developers. UC San Diego and Developers assume no responsibility for any errors or omissions relating to the MARKER- HF Calculator and You are expected to exercise your own full, reasonable, and independent professional judgment in connection with any use, by YOU, of the MARKER- HF Calculator.

IN NO EVENT SHALL UC SAN DIEGO AND/OR DEVELOPERS BE LIABLE TO YOU AND/OR ANY OTHER PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, AND/OR CONSEQUENTIAL DAMAGES, INCLUDING LOST PROFITS, PROPERTY DAMAGE, AND/OR INJURIES (MINOR, MAJOR, OR CATASTROPHIC) OR DEATH, ARISING OUT OF AND/OR RELATING TO THE USE OF THE MARKER-HF CALCULATOR.

UC SAN DIEGO AND DEVELOPERS SPECIFICALLY DISCLAIM ANY AND ALL WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. UC SAN DIEGO AND DEVELOPERS ALSO DO NOT REPRESENT OR WARRANT THAT THE PROVIDED CONTENT DOES NOT INFRINGE THE INTELLECTUAL PROPERTY, PROPRIETARY, OR CONTRACTUAL RIGHTS OF THIRD PARTIES. THE MARKER-HF CALCULATOR PROVIDED ON THIS WEBSITE IS PROVIDED ON AN "AS IS" BASIS, AND UC SAN DIEGO AND DEVELOPERS HAVE NO OBLIGATION TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS.

This Agreement and any disputes that may result from, or involve, YOUR use of the MARKER-HF calculator shall be governed by California law, and the exclusive jurisdiction and venue for any and all actions arising out of, or brought pursuant to, this Agreement and/or YOUR use of the MARKER-HF calculator shall be in a court of competent jurisdiction within the County of San Diego.

THE SITE CONTENT IS SUBJECT TO CHANGE WITHOUT NOTICE.

I Accept

UC San Diego
Altman Clinical and Translational Research Institute

# MARKER-HF™ Calculator

Enter values to calculate the MARKER-HF™ score based on *Improving risk prediction in heart failure using machine learning* Eric D. Adler *et al.*, published in *European Journal of Heart Failure*

| | | |
|---|---|---|
| Diastolic pressure (mm Hg): | | (20-120) |
| Creatinine (mg/dL): | | (0-25) |
| Blood Urea Nitrogen (mg/dL): | | (0-160) |
| Hemoglobin (g/dL): | | (2-20) |
| White Blood Cell Count ($10^3$µL): | | (0-40) |
| Platelets ($10^3$µL): | | (0-1500) |
| Albumin (g/dL): | | (0-6) |
| Red Blood Cell Distribution Width (%): | | (10-30) |

Calculate MARKER-HF | Clear fields

UC San Diego
Altman Clinical and Translational Research Institute

# MARKER-HF™ Results

## Your Inputs:

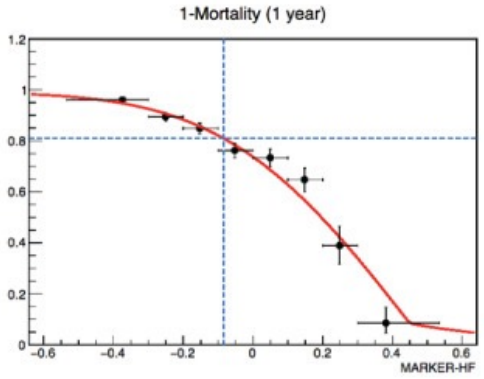
| | |
|---|---|
| Diastolic pressure: | 68 |
| Creatinine: | 1.47 |
| Blood Urea Nitrogen: | 38 |
| Hemoglobin: | 12.8 |
| White Blood Cell Count: | 6.7 |
| Platelets: | 147 |
| Albumin: | 2.9 |
| Red Blood Cell Distribution Width: | 14 |

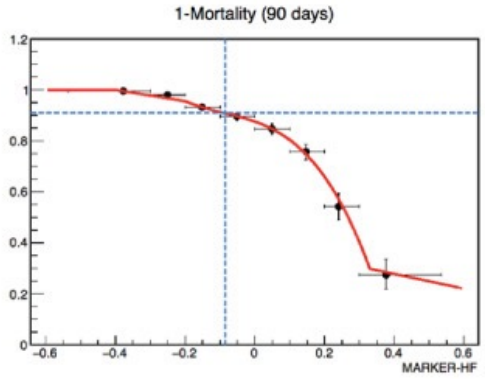
I entered some random numbers (!)

## Results:

| | |
|---|---|
| Marker-HF™: | -0.085 |
| One-year Survival Probability (i.e., 1-Mortality): | 0.81 |
| 90-day Survival Probability (i.e., 1-Mortality): | 0.91 |

The values of Survival Probability (i.e., 1-Mortality) are calculated from the value of MARKER-HF and the red curves shown below. The curves are (rough) fits to data from the MARKER-HF paper (link)




Back

Now integrated on an “experimental” basis in the UCSD and Northwestern hospital systems.

Helps to triage HF patients to advanced care

# ML: HEP vs. Cardiology

## High Energy Physics

- Train with simulated data (mosty)
- High Statistics training sets.  Easily millions of events
- Can go “deep”
- Crisp definition of outcomes
  - Signal vs Backgroud
- At what level can you really trust the simulation of backgrounds?
  - The underlying physics processes
  - The simulation of all the detector hydiosincracies
  - Garbage-in-garbage-out
- Use control samples, be smart

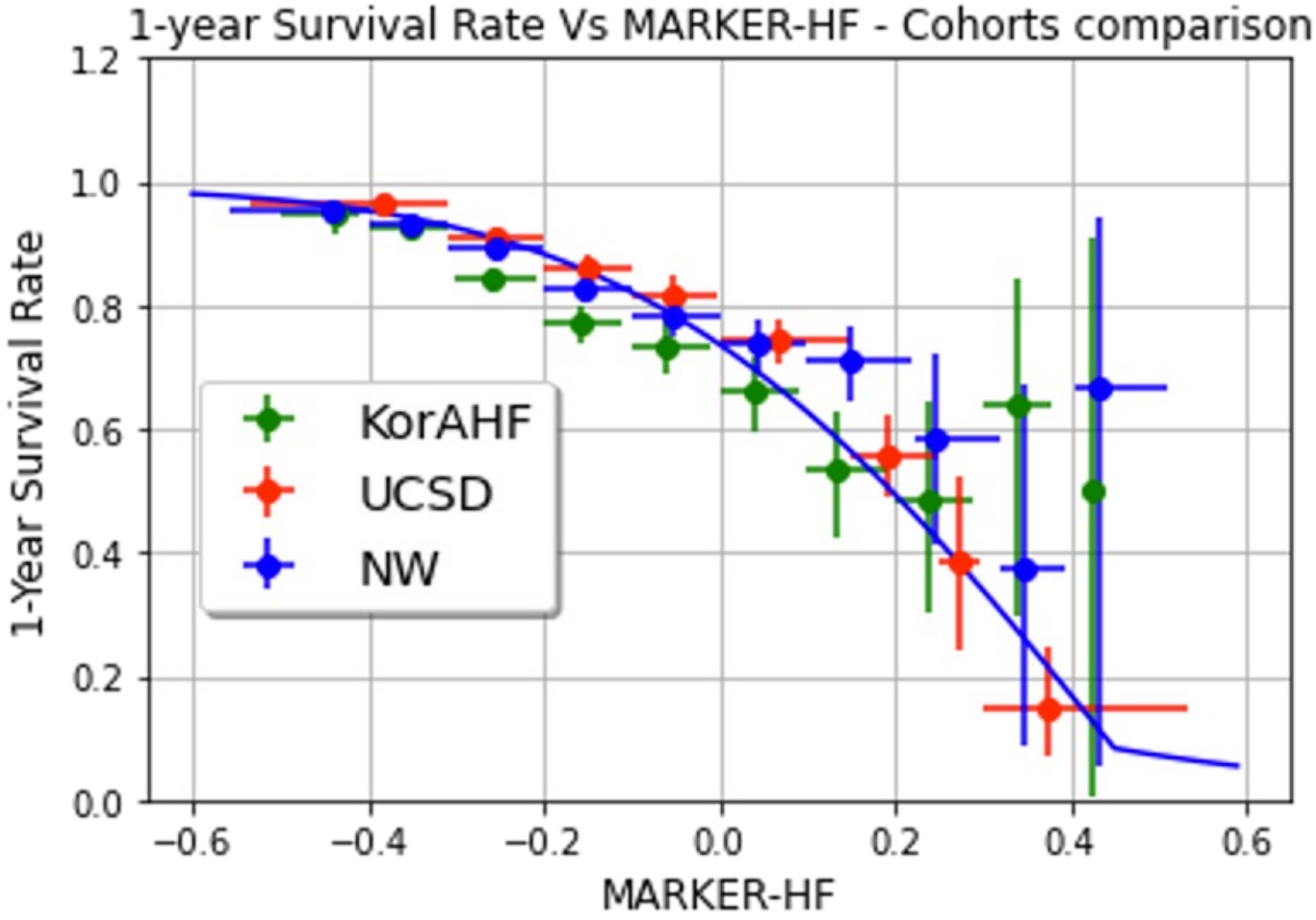
## Cardiology

- Train with real patient data
- Limited statistics (small N)
- Algorithms cannot be too complicated
- Continuos range of outcomes
- Getting reliable EHR data is painful
  - Bureaucracy
  - Poor data quality.  Needs a lot of attention
- A-posteriori clinical trials are not much better
- Censoring issues
- Biased samples, validation

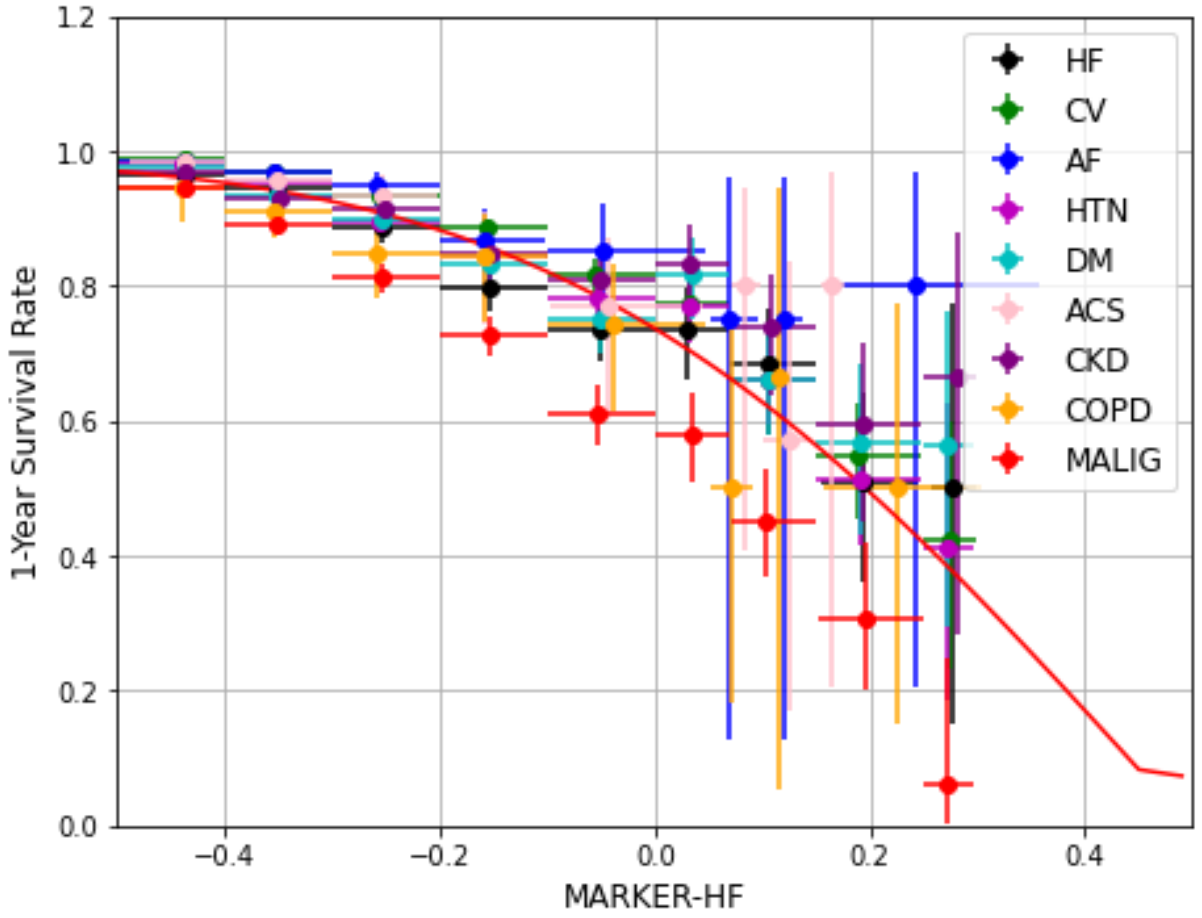
# Since the publication of the MARKER-HF paper….

- With a group from Northwestern:
  - Verified that MARKER-HF works on their patients
  - Compared with two other more state-of-the-art scores
    - Seattle Heart Failure Model (SFHM)
    - Meta-Analysis Global Group in Chornic (MAGGIC) HF Score
    - We couldn't do that originally because these scores were too complicated to calculate
  - Found that MARKER-HF works just as well, and it is much easier to deal with
  - Paper has been submitted
  - **MARKER-HF now incorporated in Northwestern clinical practice**
- With a group from Brigham and Women/Harvard:
  - Verified that MARKER-HF works on Clinical Trials patients
  - Studied the effect of using MARKER-HF as a tool to select patients to improve efficiency and lower cost of Clinical Trials
    - Eur J Heart Fail. doi:10.1002/ejhf2155 (2022).
- With two Korean groups:
  - Verified that MARKER-HF works on a Korean HF population
  - Found that it also works for conditions beyond HF
  - Paper in preparation

# MARKER-HF vs 1-year mortality.
# UCSD, Northwestern, Korean cohorts.



# MARKER-HF vs 1-year mortality. Korean patients, different diagnoses




HF = Heart Failure
CV = Cardiovascular Disease
AF = Atrial Fibrillation
HTN = Hypertension
DM = Diabetes Mellitus
ACS = Acute Coronary Syndrome
CKD = Chronic Kidney Disease
COPD = Chronic obstructive pulmonary disease
MALIG = Malignancy

Doesn't work for cancer v. well
Not too surprising.

# Beyond Marker-HF

- Was supposed to be a “proof of principle”, took on a life of its own
- Our cardiology colleagues have several ideas for problems to pursue
- Getting our hands on good data sets has proven difficult
  - Even though the cardiologists in our group are well positioned in their community
  - PIs of large Clinical Trials
- Few irons in the fire.  Most interesting is the *Sudden Cardiac Death* project
- Briefly:
  - Existing guidelines to install de-fribillators in people are not optimal
  - Many patients that do not need it, get an implant.
    - Invasive, not risk-free
  - Some patients that could have been saved by the implant are excluded
  - Can a ML algorithm help?  Looks promising.

# Concluding Remarks

- For ML in medicine, distinguish image analysis vs. simple “numbers-based” problems
- Layman impression: image tools are quite advanced (Google!)
- For the other type of problems, low hanging fruit?
- Newbies like Avi and I (with clinical advice from MDs) can develop algorithms at or beyond the state of the art
- The tools that we use are not particularly sophisticated. Do not need super-expert understanding
  - We used software that has been around for 20+ years in HEP
  - Equivalent or better toolkits are available elsewhere
- The challenge has not been the technology but rather
  - Formulating interesting/important problems
  - Getting ahold of decent data sets