Improving risk prediction in heart failure using multivariate analysis techniques from high energy physics

Eric D. Adler MD1, Adriaan A. Voors MD2, Liviu Klein MD, MS8, Fima Macheret MD7, Oscar O. Braun MD3, Marcus A. Urey MD1, Wenhong Zhu PhD6, Iziah Sama PhD2, Matevz Tadel PhD5, Claudio Campagnari PhD4, Barry Greenberg MD1, Avi Yagil PhD1,5

1 Department of Medicine, Division of Cardiology, UC San Diego, USA
2 University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
3 Departments of Clinical Sciences, Cardiology, Lund University and Skåne University Hospital, Lund, Sweden
4 Physics Department, UC Santa Barbara, USA
5 Physics Department, UC San Diego, USA
6 Altman Clinical and Translational Research Institute (ACTRI), UC San Diego, USA
7 Department of Medicine, Divisions of Hospital Medicine and Biomedical Informatics, UC San Diego, USA
8 Department of Medicine, Division of Cardiology, UC San Francisco, USA
How did this start

Conversations between exp high energy physicists and clinical cardiologists:

• Attempts to predict heart failure (HF) mortality using clinical variables or risk scores have not been very successful
• Fails in one or more of the following ways:
  • limited predictive power (poor AUC, typically < 0.7)
  • loss of accuracy when applied to other cohorts or populations
  • dependence on variables that are subjective or not readily available

There are many many risk-scores on the market to the point that “risk-score fatigue” is prevalent
Today’s buzzword…Machine Learning (ML)

• In HEP have been using (simple) ML for > 20 years to categorize “events” or “objects” as “signal” vs. “background”
  • Event = final state of decay or of interaction
  • Object = electrons, photons, bottom quarks, etc

• Can our experience be brought to bear?
  • Signal/Background → Early death/Long term Survival
Primary Objective

Define a risk score aiming to avoid common pitfalls by:

• **Using strict data collection and cleanup methodology** that ensures maintaining the correlations that define the physical state of the system (patient) within a relatively limited period of time.

• **Having precise definitions of outcomes.**

• **Avoiding imputation** by requiring all covariates used in the creation of the model to be present.

• **Limiting the number of inputs to a small number of widely available covariates** that are checked routinely in HF patients.

• **Capturing the multi-dimensional correlations** between the covariates and the outcomes.
Methods and Study definition

• Retrospectively extracted de-identified EMR of patients in the UCSD health system with 1st recorded diagnosis of Heart Failure.
• Used an iterative process to select a minimal, most common, and discriminating set of variables (only 8)
  • HEP software, TMVA in CERN Root package
• For patient to be included in the analysis all variables needed to be present using data collected over a narrow time window (<7 days)
• Excluded patients: over 80, with an ICD, with indication of Sepsis
  • Minimize dilution of outcomes
Statistics

Comments:

• Low statistics compared to typical HEP problem

• Dirty data, big loss from missing variables
  • Systematics?

• Extracting data from the EMR was a bit of an “adventure”
Cartoon of samples definition

Lifetime distribution

High Risk cohort
N = 407

Low Risk Cohort
N = 800

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

Important: No ”silver bullet”:
- No single discriminating variable.
- Each shows some separation.
- But poor AUC, individually.

Key is the combination and correlations between the whole set
A Boosted Decision Tree algorithm was used (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 ANN*

AUC: 0.875 (95% CI 0.85-0.90)

Validation Cohort Sizes:
- High Risk = 204
- Low Risk = 483
Does it make sense clinically?

Does a high score corresponds to abnormal values of the measured inputs??

- Input variables plotted against the mean score
- The yellow bands represent the normal ranges.

➤ NOT an explicit input to the model. “Learned” by the training
Comparisons with similar risk scores


Reproducibility with patients not from UCSD

• “Local” bias of studies major concern in medical field

• Applied model derived on UC San Diego patients on patients from
  • UC San Francisco EMR
  • BIOSTAT—CHF
    • A study from 69 centers in 11 EU countries

• AUC is statistically consistent in the 3 cohorts of patients
Applicability to all patients, i.e., not just the “high risk” and “low risk”

Consistency between the 3 cohorts
Comparing NT-proBNP and Marker-HF

Amino-terminal pro-B-type natriuretic peptide, (NT-proBNP) is a well validated biomarker associated with HF
Other features that did not have time to cover

Consistent performance in
• Men vs. Women
• As a function of ethnicity
• As a function of age
• For in- vs. out-patients (proxy for health status)
• Patients with pulmonary edema vs. other HF diagnoses
• No detectable bias due to excluding patients with missing variables
Conclusion

• Succeeded in designing predictive risk score algorithm for HF using ML
• Key features, some based on HEP experience
  • Tight data collection requirements → Clean training/testing samples
  • Clear definition of outcomes using the extrema
  • Automatic inclusion of correlations (potentially non-linear)
• Paper submitted to medical journal
• Starting to think about further studies, still in cardiology, e.g., can we predict which patients will suffer major bleeding with anticoagulants?