Multiple Myeloma

Progression and Outcome Prediction

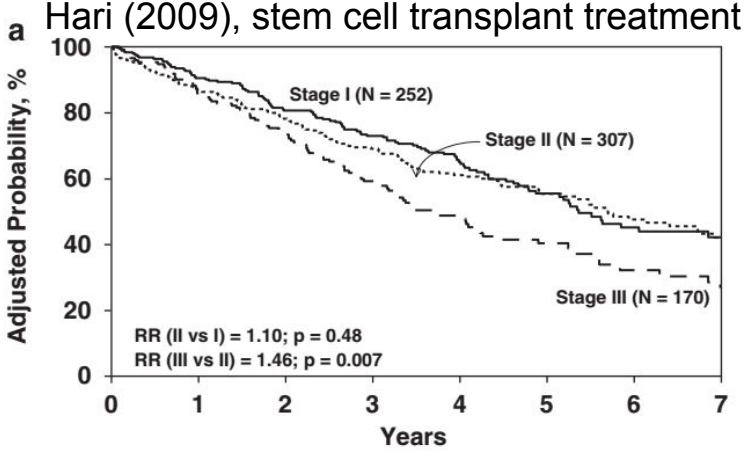
Kevin Shi, Ege Ozgirin, Dustin Doss

17 May 2017
Machine Learning in Healthcare 6.S897

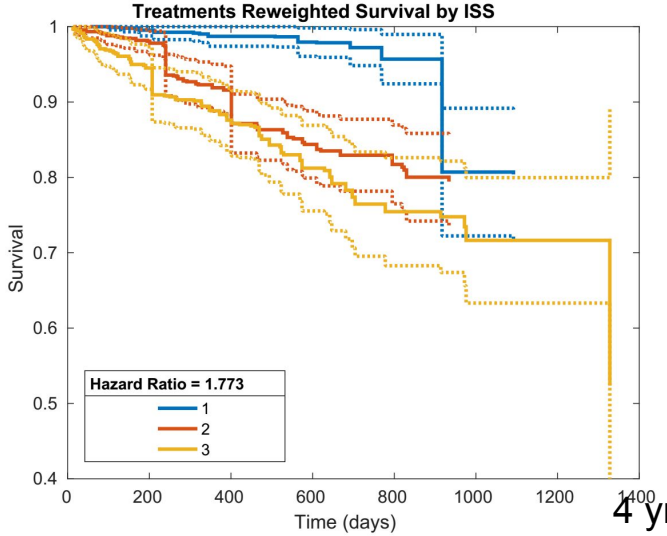
What is Multiple Myeloma?

- Cancer of malignant plasma cells in bone marrow
- Excess production of a clonal immunoglobulin leads to kidney/bone damage
- Average life expectancy is ~4 years (SEER 2013)
- Patient prognosis given by “stage”. International Staging System (ISS) defines 3 stages

Hari (2009), stem cell transplant treatment



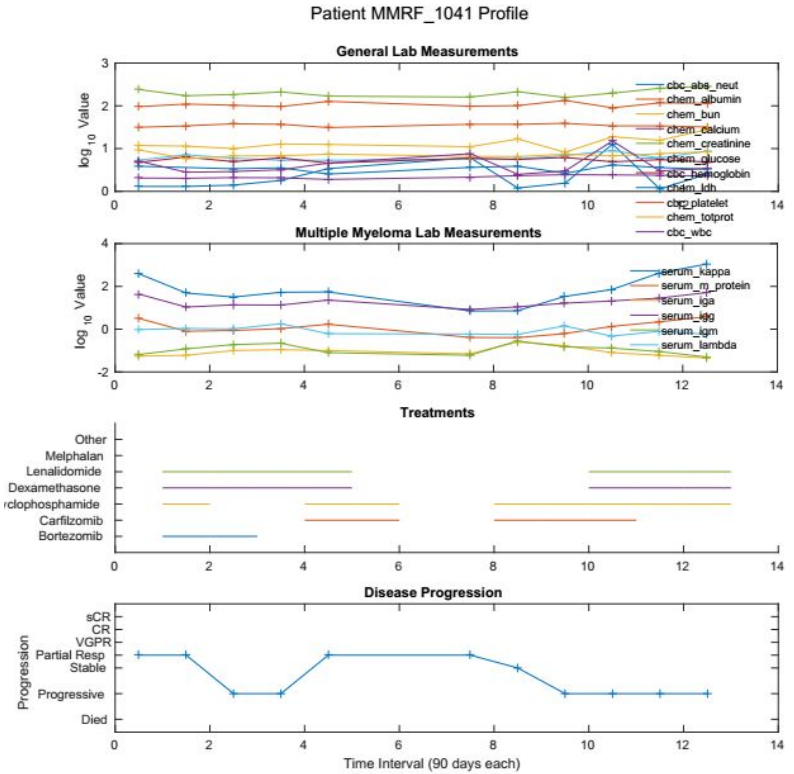
MMRF CoMMpass (ongoing), current treatments



4 yr

Multiple Myeloma Research Foundation CoMMpass Study

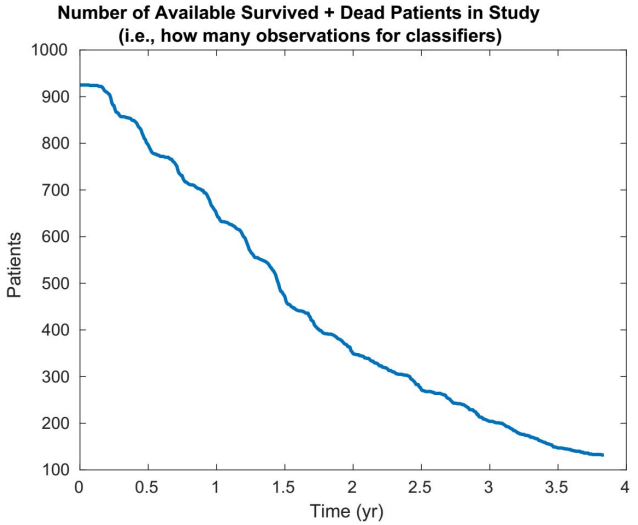
- For ~1000 patients, collect at multiple timepoints:
  - Demographics
  - Clinical measurements:
    - Bone damage
    - M protein levels
    - Usual lab work
  - Genomics (in tumor):
    - RNAseq expression data
    - Cytogenetics and mutation data
  - Progression status

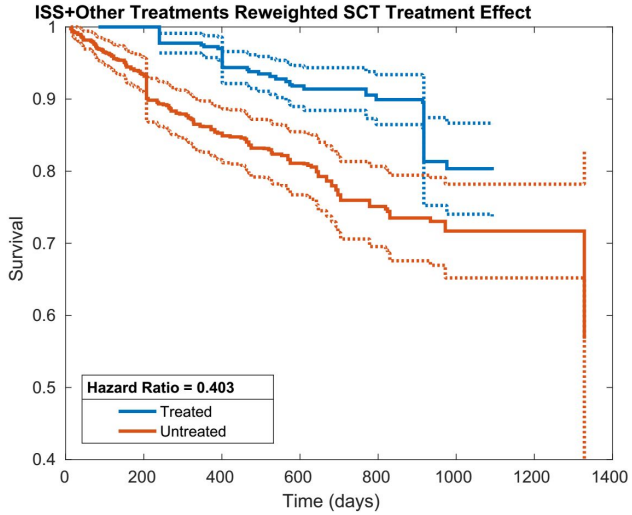


Subset of timeseries data for a patient with a relapse event

MMRF Dataset Statistics

- Behaves like an “observational” study
- ~1000 patients participating:
  - Median age 64
  - Fairly even split among ISS Stages I-III
  - ~13% morbidity, ~13% relapse over course of study
  - Over 60% of patients received more than 3 drugs as part of therapy
  - Around 58% of patients received a stem cell transplant (best treatment)

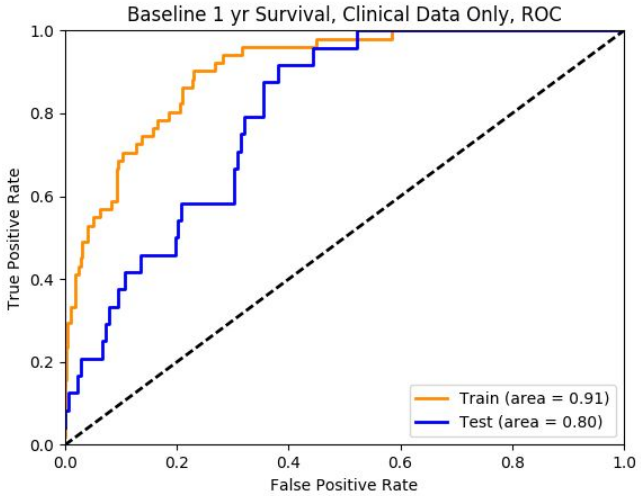




Aim 1: Baseline Survival Prediction

Logistic regression

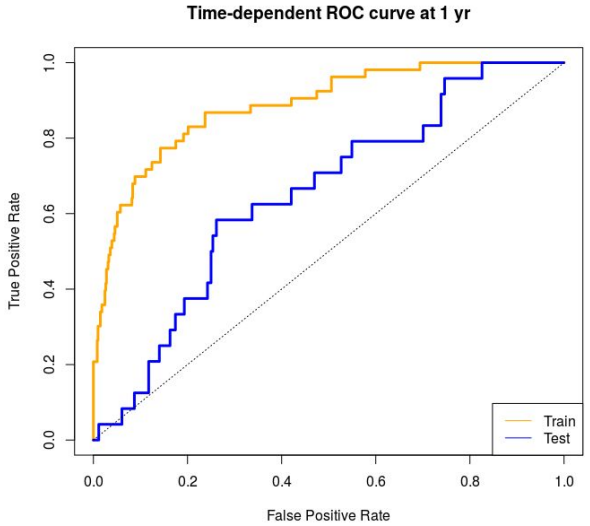
- Can we predict patient survival at 1 yr?



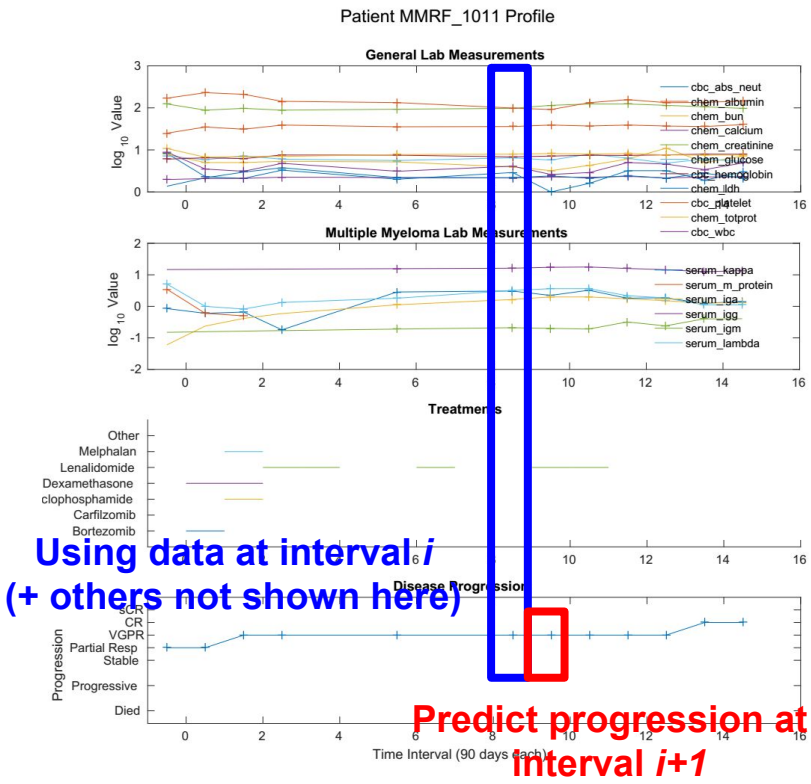
- Including genomic data usually results in overfitting

Cox proportional hazards regression

- Can we predict hazard, the (increased) chance of dying?
- Overall c-index: Train: 0.90, Test: 0.66



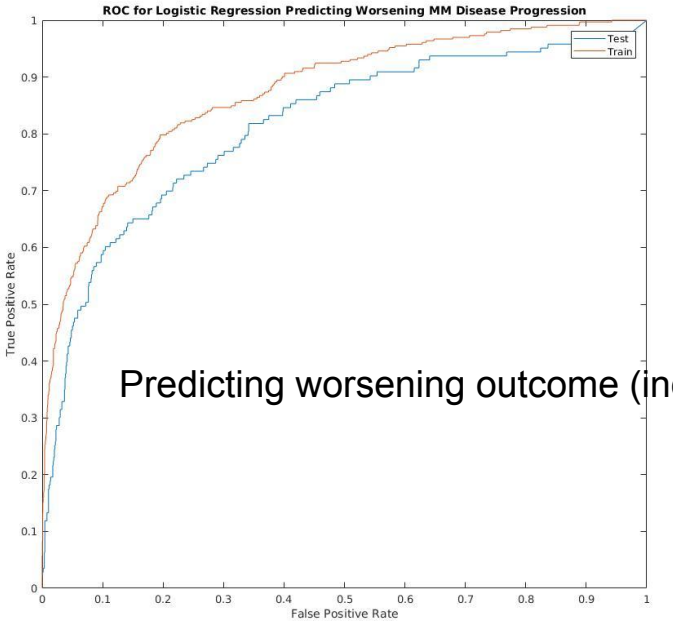
Aim 2: Time-series Progression Prediction



Using data at interval *i*
(+ others not shown here)

Predict progression at interval *i+1*

Given patient state in the current time interval, can we predict disease progression in the next time interval?



Predicting worsening outcome (including death)

Discussion and Future Work

- In general, it’s hard to tease out particular treatment effects
  - CoMMpass is not a randomized controlled trial
  - Focus on finding predictors of prognosis
- We saw counterintuitive results in regression features
  - Some medications individually negatively correlated with survival--confounding effects?
- But we can still get quite reasonable performance
- Simple methods used here, but future work can do better:
  - Timeseries imputation: simple imputation -> Gaussian process
  - Linear/logistic regression -> any kind of classifier
  - More principled hazards regression - regularization
  - More principled incorporation of genomics data