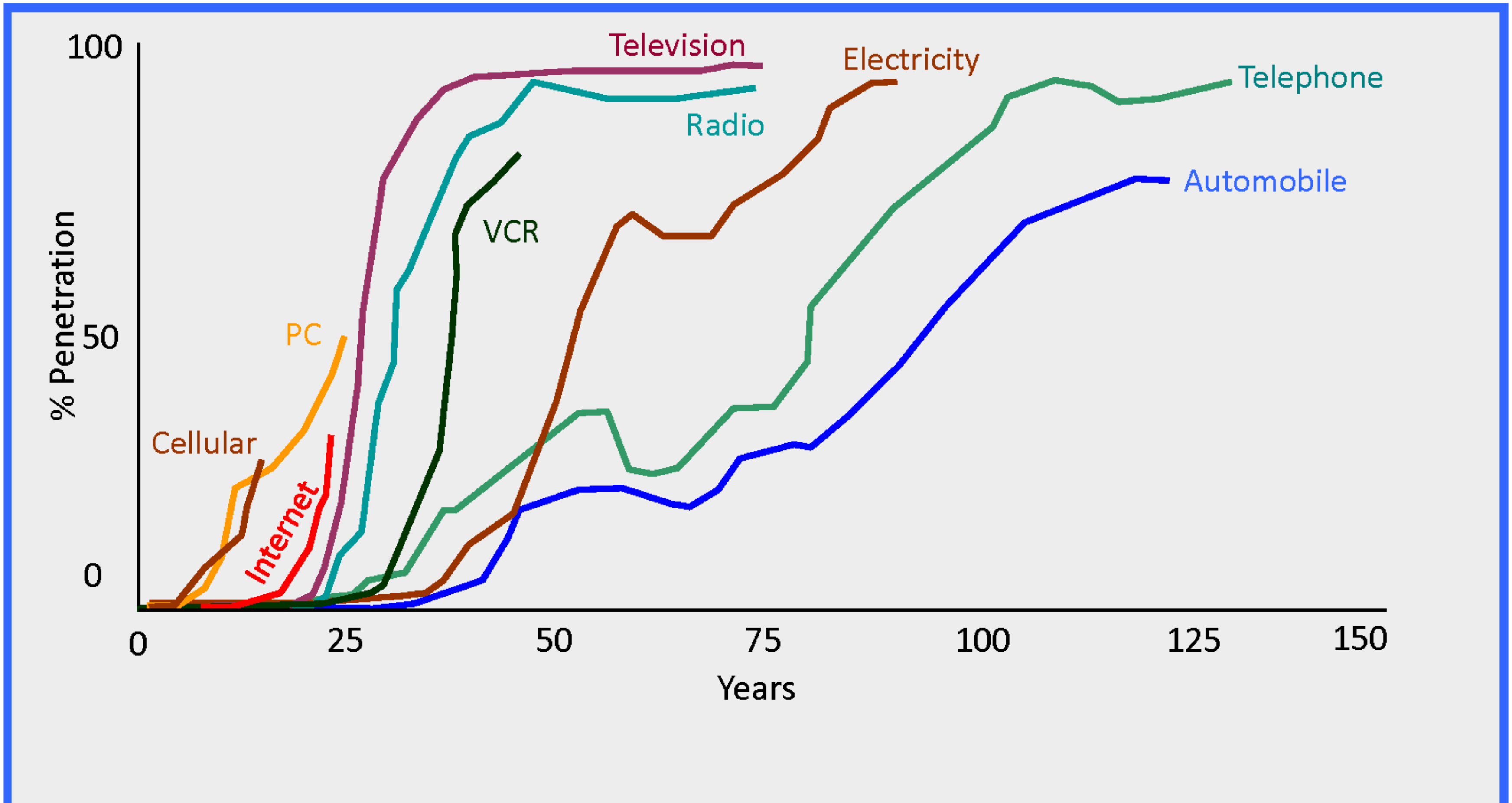
THE FUTURE OF OR

Dimitris Bertsimas
MIT

Pace of Innovation Accelerating

Newer technologies are taking hold at double or triple previous rates

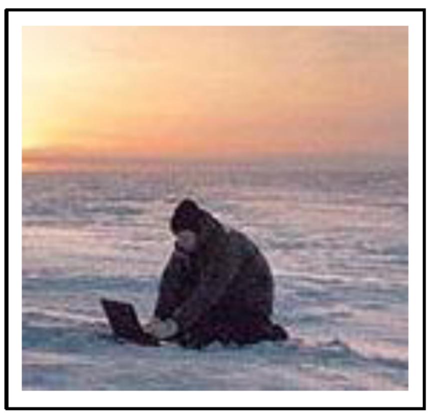


Growth of Machine-Generated Data

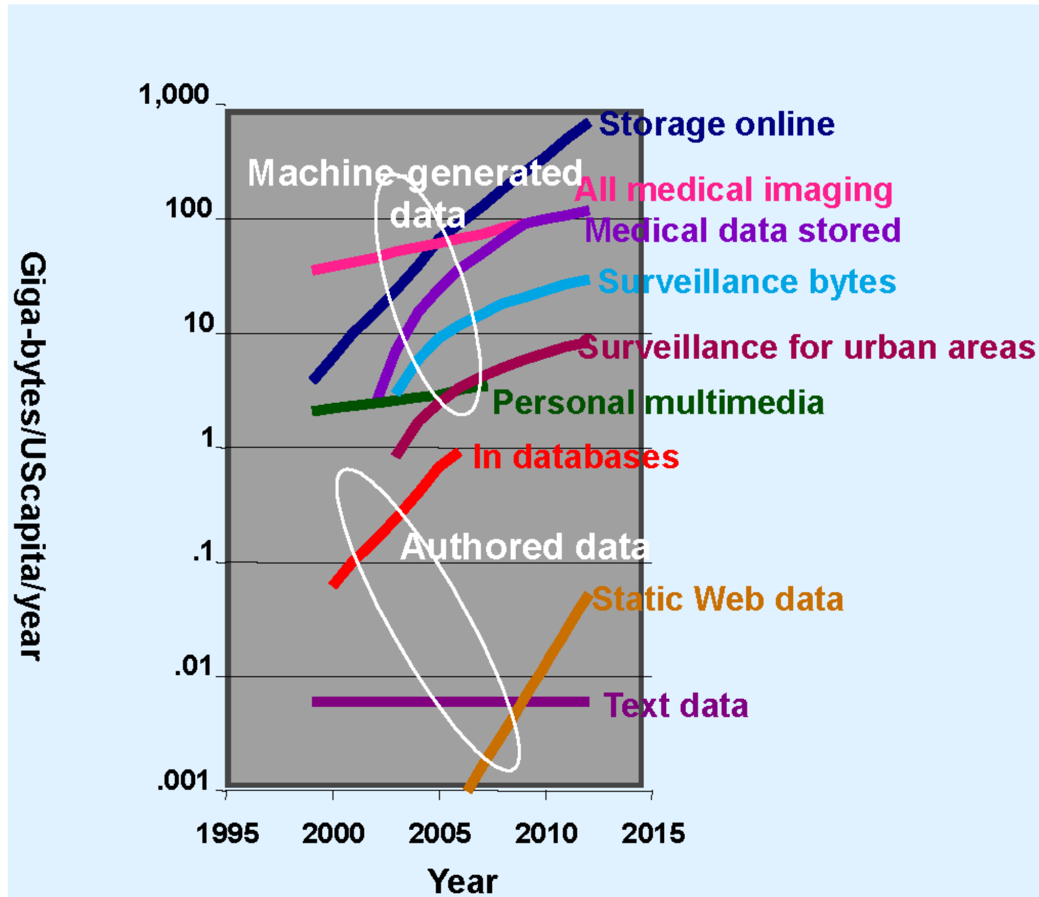
The growth of machine generated, time-based data from a variety of sources is changing the game

Machine-generated data

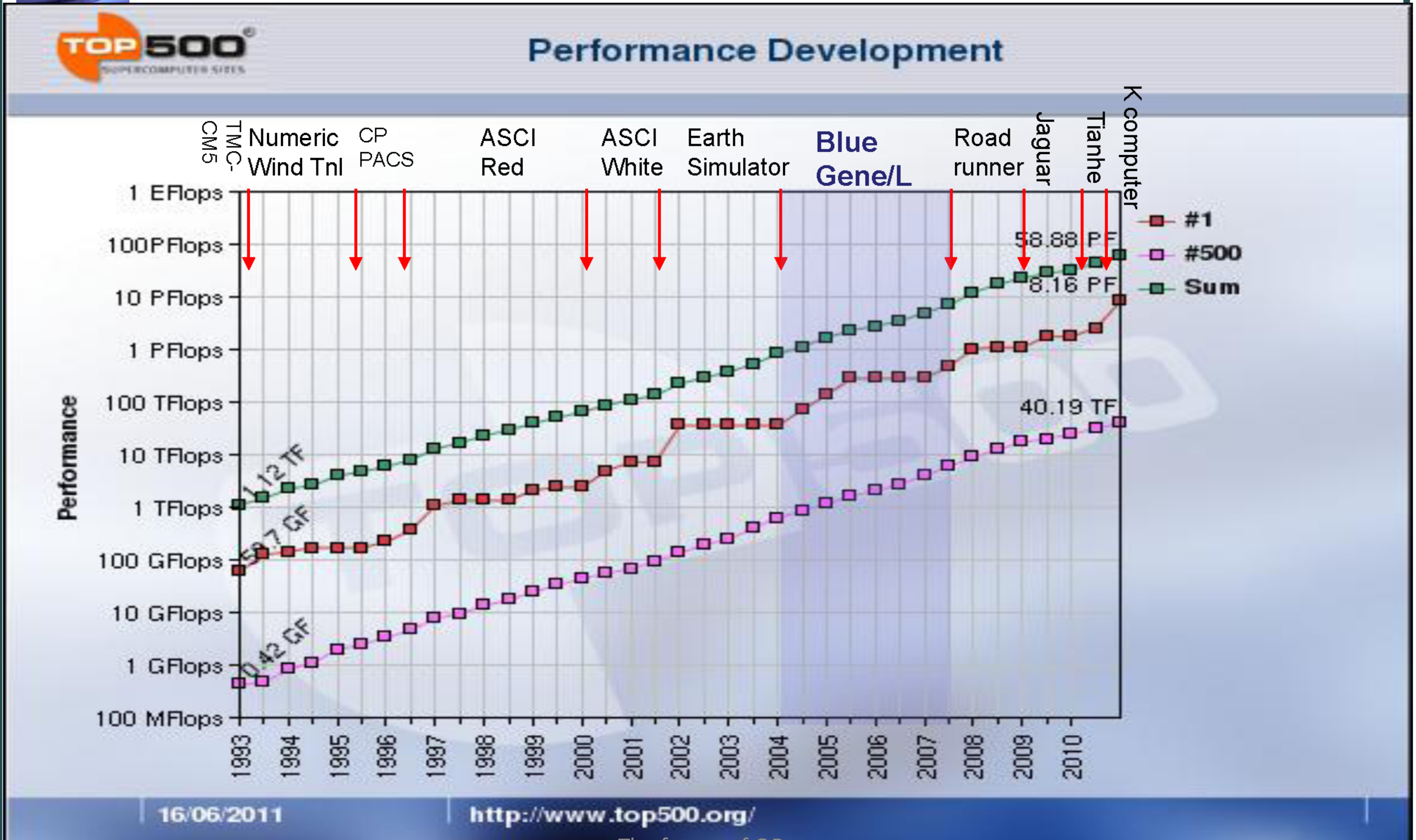
Authored Data



Machine-generated versus authored data



Computing Power grows exponentially



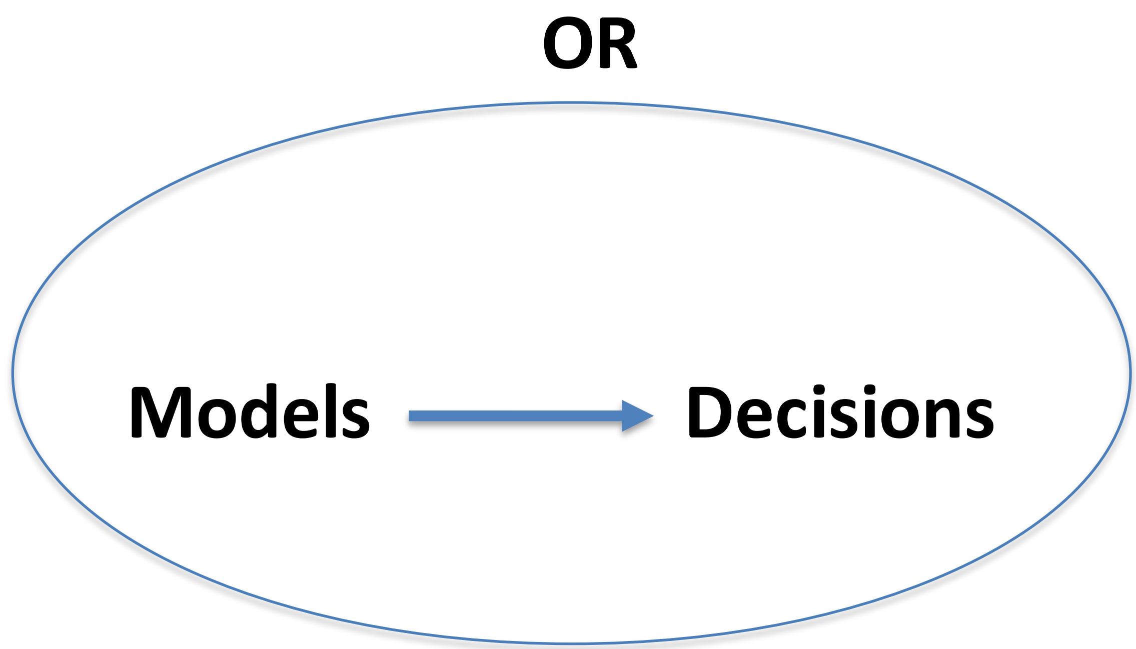
Optimization Power

- Improvements in Software and hardware have accelerated Mixed Integer Optimization

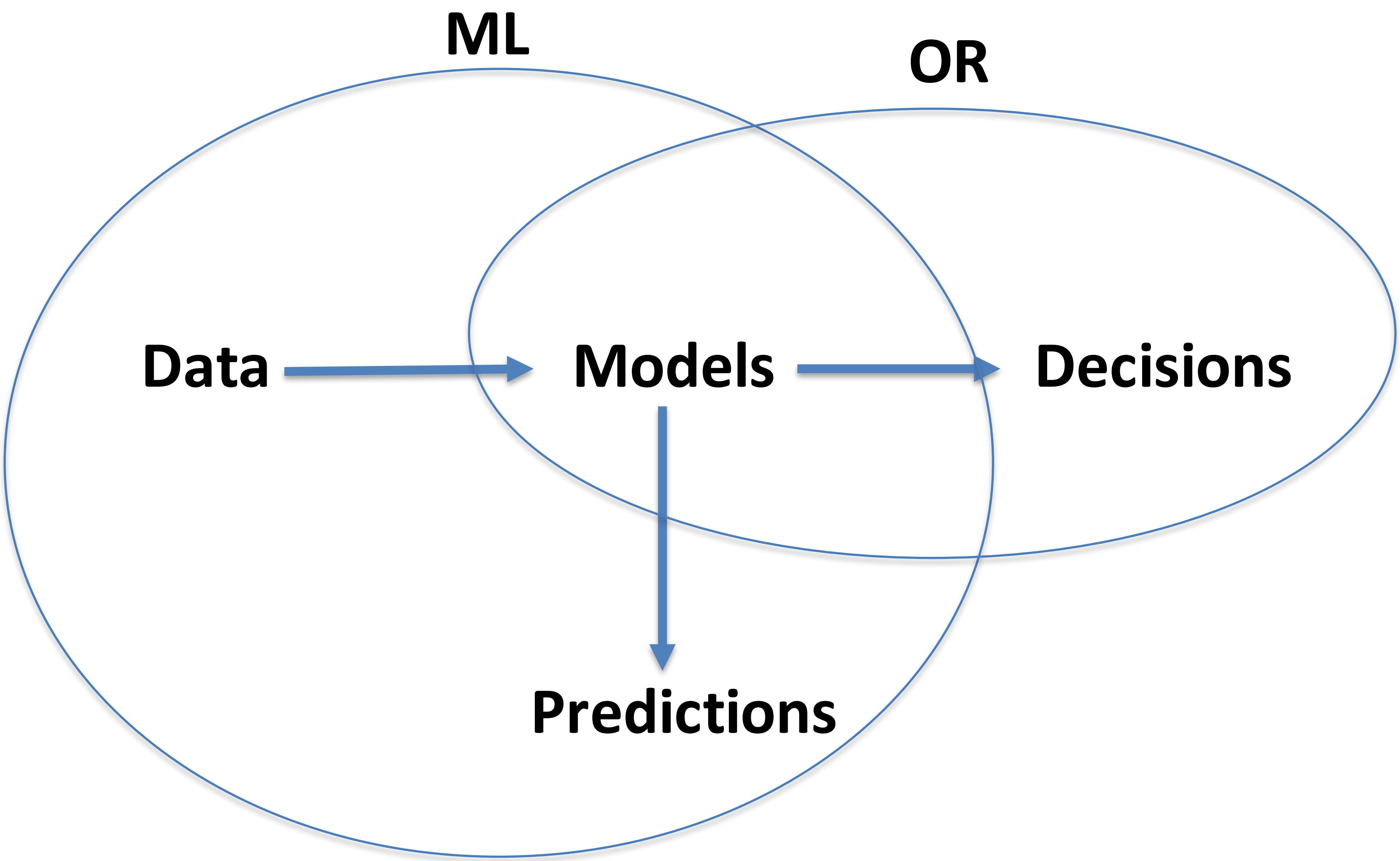
2.2 Trillion times!

- This forces to rethink what is tractable

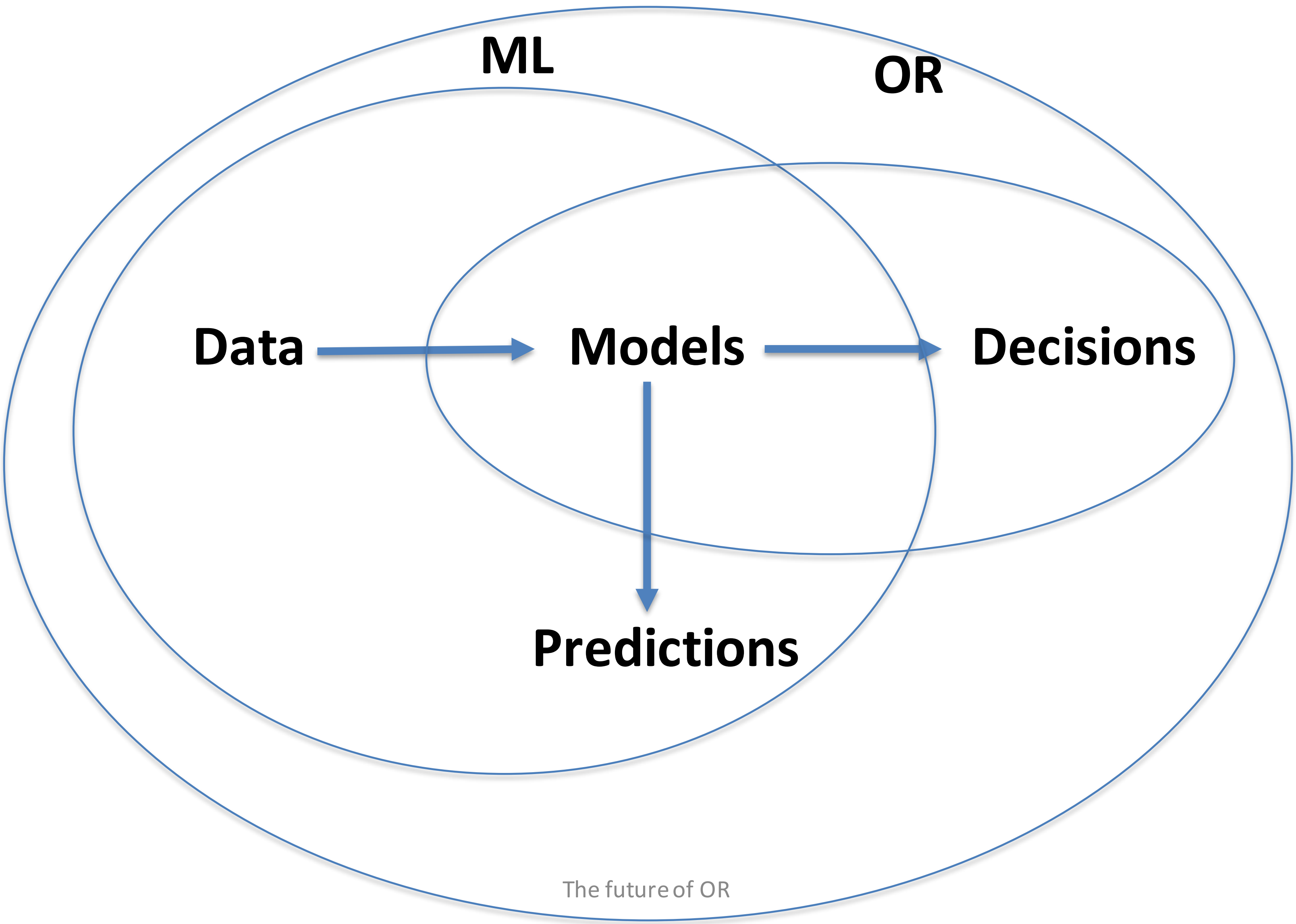
OR Now



OR and ML now



The Future: Analytics



What is Analytics?

An approach to solving problems that starts with **Data**, builds **Models** to arrive at **Decisions** that create **Value**.

The Role of Models

- Models, in my experience, exist in our imagination.

- Old saying: All models are wrong, some are useful.

- Data represent an objective reality.

Impact in Research

- Use of Optimization to Solve classical Problems of ML
- From Predictive to Prescriptive Analytics
- De-emphasize Models, Emphasize Data
- Interpretability Matters
- I believe ML/Statistics will be more linked with Optimization rather than Probability

Impact in Education

- Courses at all levels should start with data
- For example, LO, NLO, RO
- Emphasis on impact and value
- Emphasis on Action Learning
- An example of a MOOC: The Analytics Edge

  that attracts about 100,000 students per time.
- A PhD class in ML at MIT attracts 600-700 students a year!

Analytics Degrees

- At MIT, we started a master of Business Analytics in 2016
- 2016: 300 applicants, a class of 16
- 2017: 950 applicants, a class of 30
- In the US, MBAn applications will exceeed MBA applications.

A Vision for Personalized Medicine

Motivation

- Medicine today **is not personalized**. Patients with similar symptoms diagnosed with the same disease are given the same treatment, independent on their history, biological variations and hereditary factors.
- Medical education **has not changed** since the 1920s.
- Can we improve outcomes for patients by personalizing treatment?
- How should we educate the new generation of doctors?

A Vision for 21st Century Medicine

- John, a patient, visits a doctor and gives her access to his medical history in the cloud.
- The doctor inputs specific details from her discussion with John.
- An algorithm uses this data to propose alternative diagnoses and treatments personalized to John together with an explanation of why.
- A report is produced on the likelihood of success of treating the patient and the possible side effects.
- This is **a vision of personalized, analytics driven medicine**.

A Vision for 21st Century Analytics

- OR departments offer degrees of Analytics in Medicine in collaboration with Medical schools.
- Medical schools train doctors in Analytics.
- Electronic medical records, the maping of the human genome provide an unprecedented opportunity for our field to affect medicine, the most important opportunity in my career.
- Conversely, medicine can help transform our field to using much more data than is using today.

Personalized Diabetes management

joint work with

Nathan Kallus, Alex Weinstein, Daisy Zhou

Diabetes Care 2016

Data

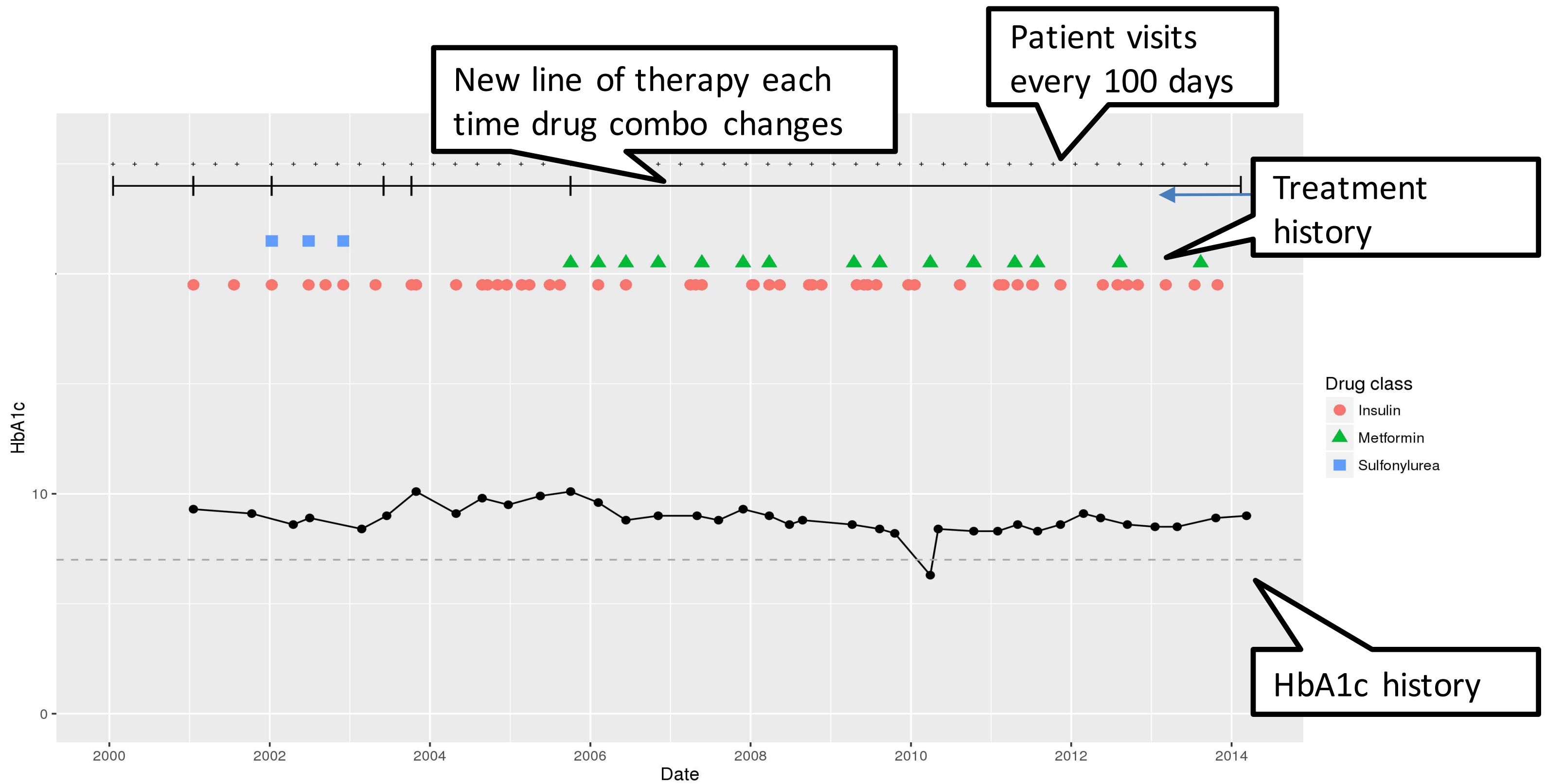
- EMR for > 1.1 million patients from Boston Medical Center from 1999-2014
  - 10,086 type 2 diabetes patients with sufficient data.
- Patient characteristics
  - *Demographic*: age, sex, race/ethnicity, language, religion, marital status.
  - *Medical history*: records for BMI, HbA1c, serum creatinine levels.
  - *Treatment history*: medication records.

Data

Table 1 Demographics, medical history, and treatment history of patients ($N = 10,806$)

Feature	**Mean (SD)**
*Age (years)**	59.7 (13.6)
% Male	42.4%
% Black	58.5%
% Hispanic	15.1%
% White	16.6%
*BMI (kg/m2)**	33.1 (8.1)
*HbA1c (%)**	7.9 (1.8)
Years since first treatment in EMR	3.52 (3.66)
Current prescription for metformin†	45.6%
Current prescription for insulin†	30.2%
Contraindicated to metformin‡	17.4%

Modeling lines of therapy and visits



Decisions and Outcomes

- Decisions and outcomes are defined relative to each patient visit:
  - 48,140 unique patient visits.
- Outcome of interest:
  - Average post-treatment HbA1c in period 75-200 days after each visit.
- At each visit, we observe ground-truth “standard of care” treatment:
  - For most visits, provider prescribed continuation of current line of therapy.

Decisions and outcomes

Table 2 Pharmacological regimens

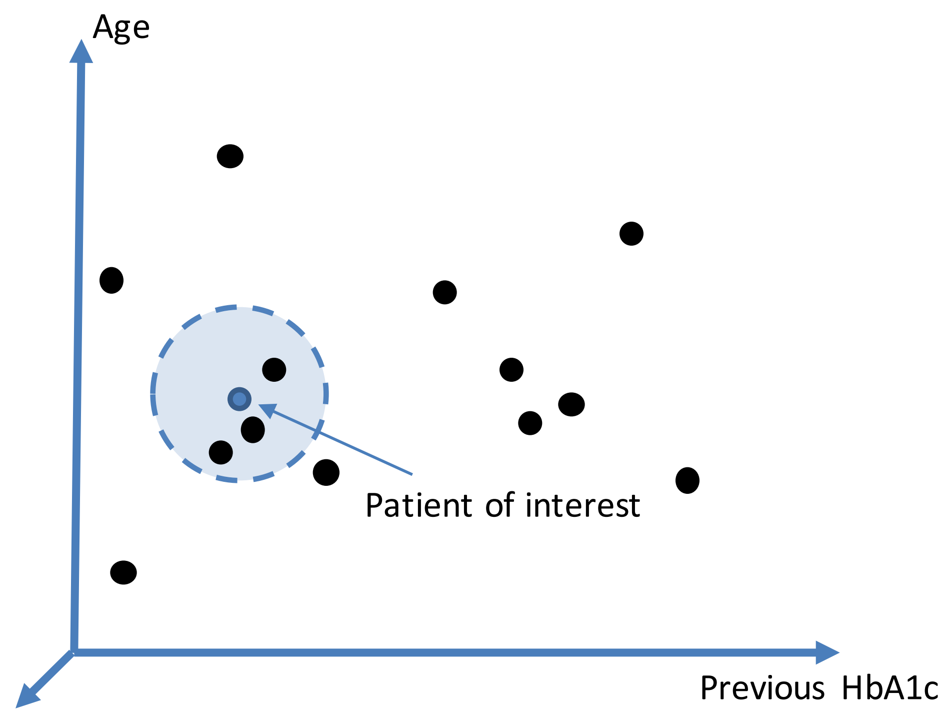
Observed standard of care regimen (abbreviation)	**Number of patient visits**
No regimen prescribed, new patient (NEWPT)	5,449
No regimen prescribed, existing patient (NORX)	2,137
Metformin monotherapy (MET0)	9,649
Non-metformin oral monotherapy (ORAL0)	4,671
Insulin monotherapy (INS0)	7,539
Metformin combined with one other oral agent (MET1)	6,959
Metformin combined with insulin (METINS0)	3,977
Insulin combined with one non-metformin oral agent (INS1)	2,139
Combination of two non-metformin oral agents (ORAL1)	1,047
Metformin comb. with two other oral agents (MET2)	1,749
Metformin comb. with insulin and one other oral agent (METINS1)	2,005
Insulin combined with two non-metformin oral agents (INS2)	249
All other multi-drug (3+) combinations (MULTI)	570
Total	48,140

Estimating potential outcome via *k*NN

To estimate a patient's potential outcome under treatment *T*, we

- search the EMR database for the *k* most similar patient visits receiving treatment *T*, and
- take average of neighbors' outcomes.

Similarity defined as weighted distance among patient demographic, medical history, and treatment history characteristics



Defining similarity metric

- Goal: **factors most predictive** of HbA1c outcome have **larger weight** when finding neighbors.
  - Train separate linear regression model for each treatment regimen.
  - Predict HbA1c outcomes for patients receiving that treatment based on demographic, medical history, and treatment history features.
  - Use magnitude of coefficients from each regression model as weights in distance metric.

E.g. when finding similar neighbors who received metformin monotherapy, the most predictive factors were:

- Previous HbA1c measurement (weight=0.22)
- Patient currently on insulin (0.11)
- Mean BMI last 100 days (0.11)
- Various other BMI and HbA1c measurements (weights ranging from 0.03 to 0.10)

*k*NN yields accurate predictions

- We calculate out-of-sample R^2 of *k*NN HbA1c predictions
  - Among unseen patients who *actually received* each treatment.
  - R^2 differs by model but fairly predictive for all treatments.
- Compare with lasso and random forest predictive models
  - Similar accuracy, but more interpretable

	kNN	Lasso	Random forest
Average R^2	0.40	0.39	0.41
Min. R^2	0.20	0.33	0.24
Max. R^2	0.54	0.53	0.53

Personalized recommendation algorithm

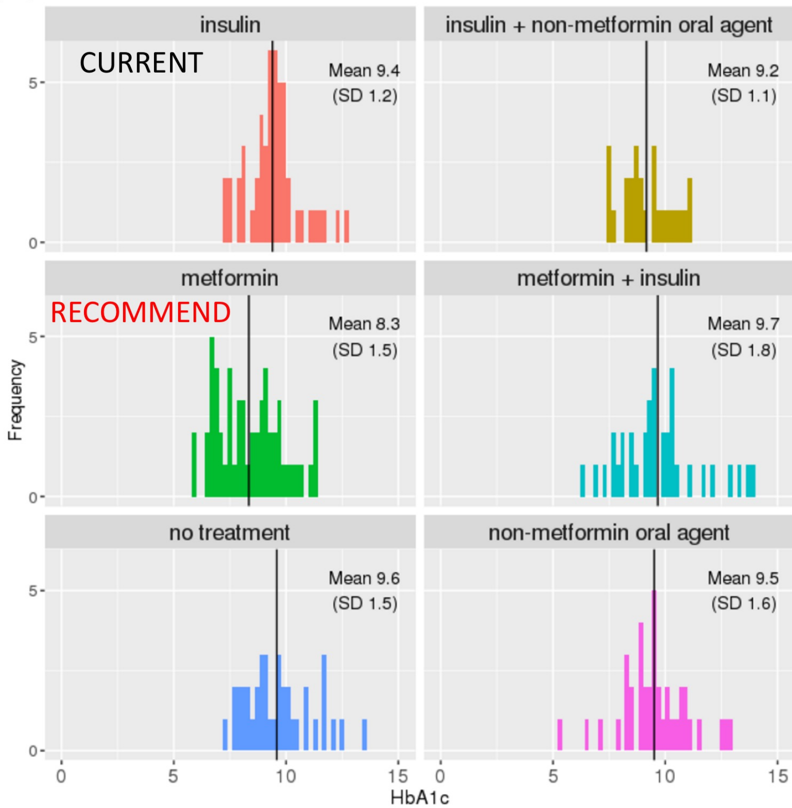
For any given patient at any given visit:

1. Generate menu of available treatment options.
   - Menu includes current treatment and natural deviations from current treatment; incorporates contraindications to metformin.
2. Use *k* nearest neighbor regression to predict potential outcome under each treatment option.
3. Reject any non-current treatment option with predicted outcome above pre-specified HbA1c threshold.
   - Threshold: HbA1c at least **0.8%** better than continuing current treatment.
4. Recommend remaining option with best predicted outcome.

(a) **Recommendation:** Switch from insulin monotherapy to metformin monotherapy

Predicted HbA1c (%): 8.3

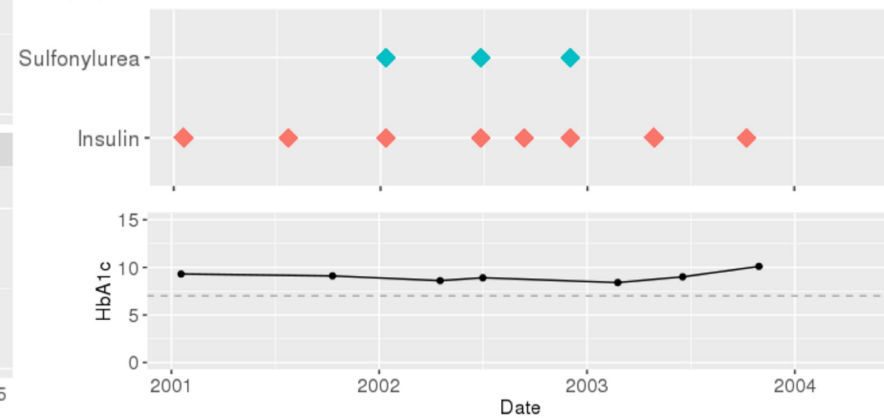
(b) **Outcomes for similar patients who were prescribed…**



(c)

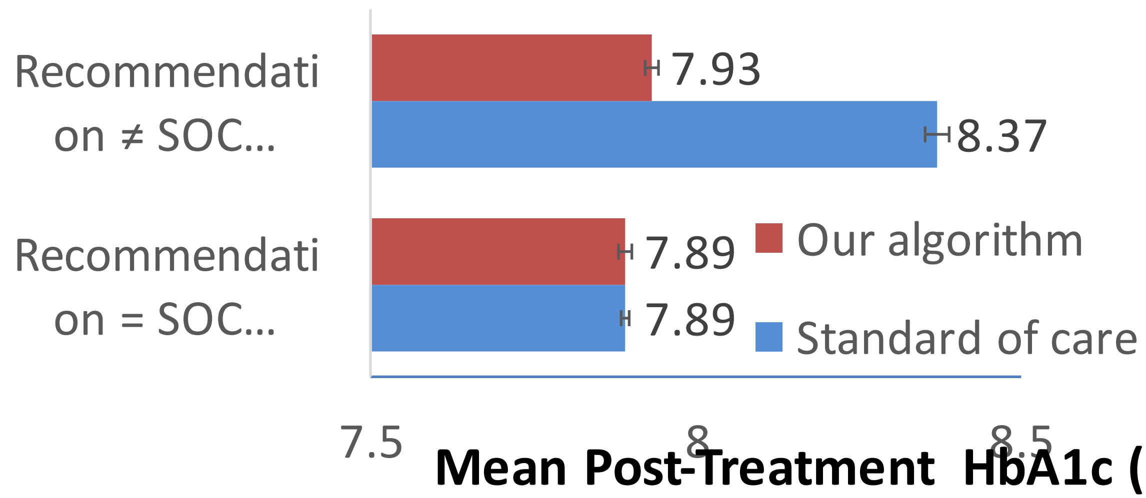
PATIENT ID	12XXXXX
AGE (Years)	61.9
SEX	F
RACE/ETHNICITY	Black
CURRENT HbA1c (%)	10.1
CURRENT REGIMEN	Insulin

(d) **Patient Treatment & HbA1c History**



Effectiveness of algorithm

- The algorithm is tuned to be conservative; it only recommends a change if the predicted benefit is large
  - In 31.8% of patient visits, the algorithm recommends a treatment different from standard of care
  - Among those visits, mean HbA1c % under algorithm was lower than SOC by **0.44** ($p<0.001$)

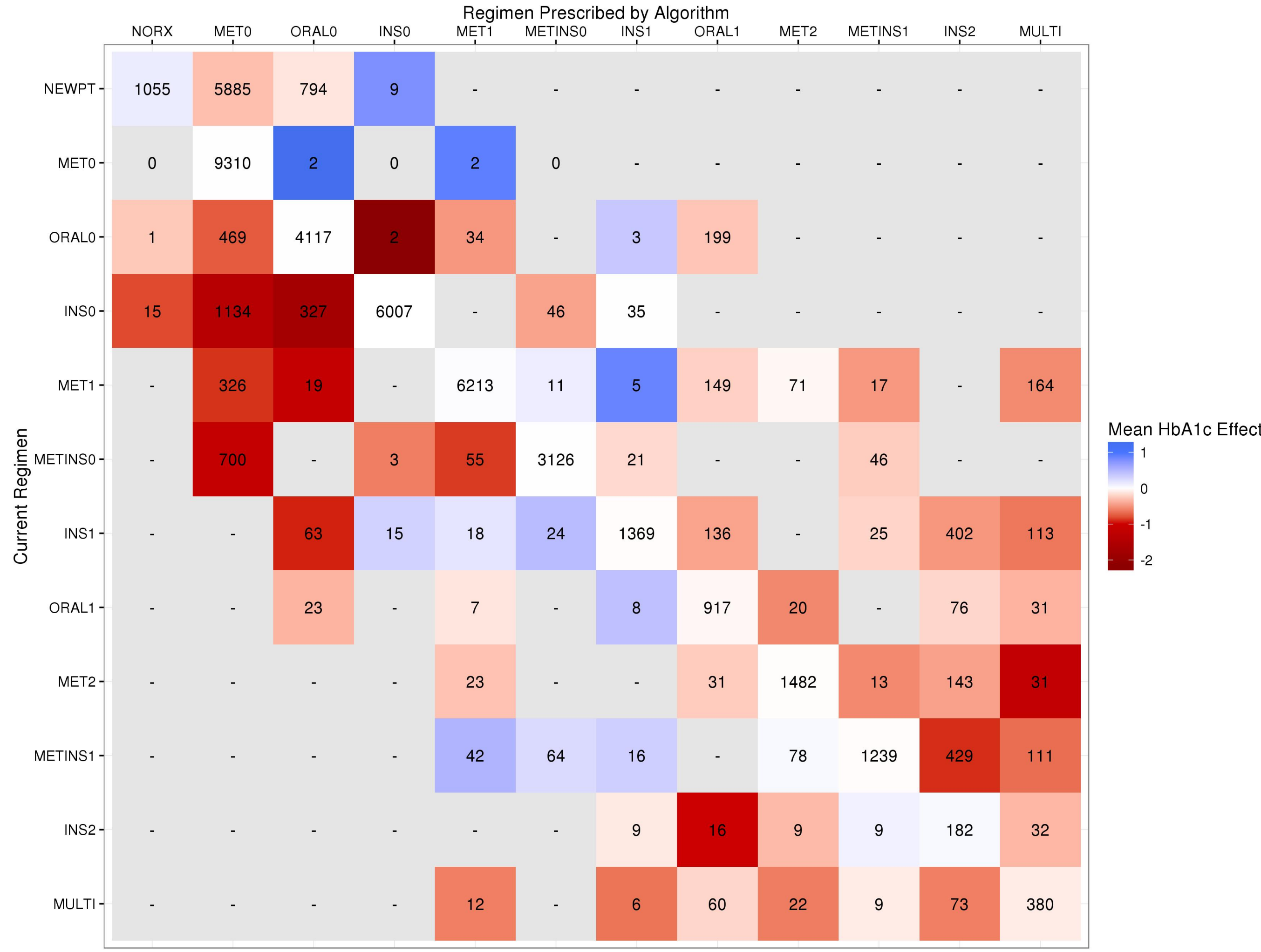


Color = mean benefit when algorithm recommends switching from [row] to [col].

Red = better than SOC

Blue = worse than SOC

Label = # of patient visits for which algorithm recommends switching from [row] to [col].



Subgroup analysis

		Recommendation differs from SoC		HbA1c benefit relative to SoC	
Subgroup		Number of instances	% of instances in subgroup	Mean (% pts)	SE (% pts)
Sex	Male	6,363	31.5%	-0.44	0.02
	Female	8,960	32.1%	-0.44	0.02
Ethnicity	Black	9,103	31.3%	-0.45	0.02
	White	2,309	31.0%	-0.29	0.03
	Hispanic	2,400	35.7%	**-0.61**	0.03
	Other	1,511	31.2%	-0.34	0.04
Age	<60	8,783	37.1%	**-0.55**	0.02
	60+	6,540	26.8%	-0.30	0.02
Glycemic control	HbA1c <=7	4,438	24.4%	-0.20	0.02
	HbA1c > 7	10,885	36.3%	**-0.54**	0.02

The future

- **Personalized medicine for**

  Various forms of Cancer

  Cardiovascular disease

  Blood Pressure

  Diabetes and Blood Pressure

- **Personalized screening for**

  Breast Cancer

  Prostate cancer

  Colon cancer

Conclusions

- This is a time of great opportunity
- This is a time of great challenge; huge demand
- OR should Embrace ML, adapt
- The best time for the field during the 30+ years I have been a professor.