Making decisions in high dimensions

DataEngConf, New York City

30 Oct 2017

Eric Novik

eric@generable.com

Vision

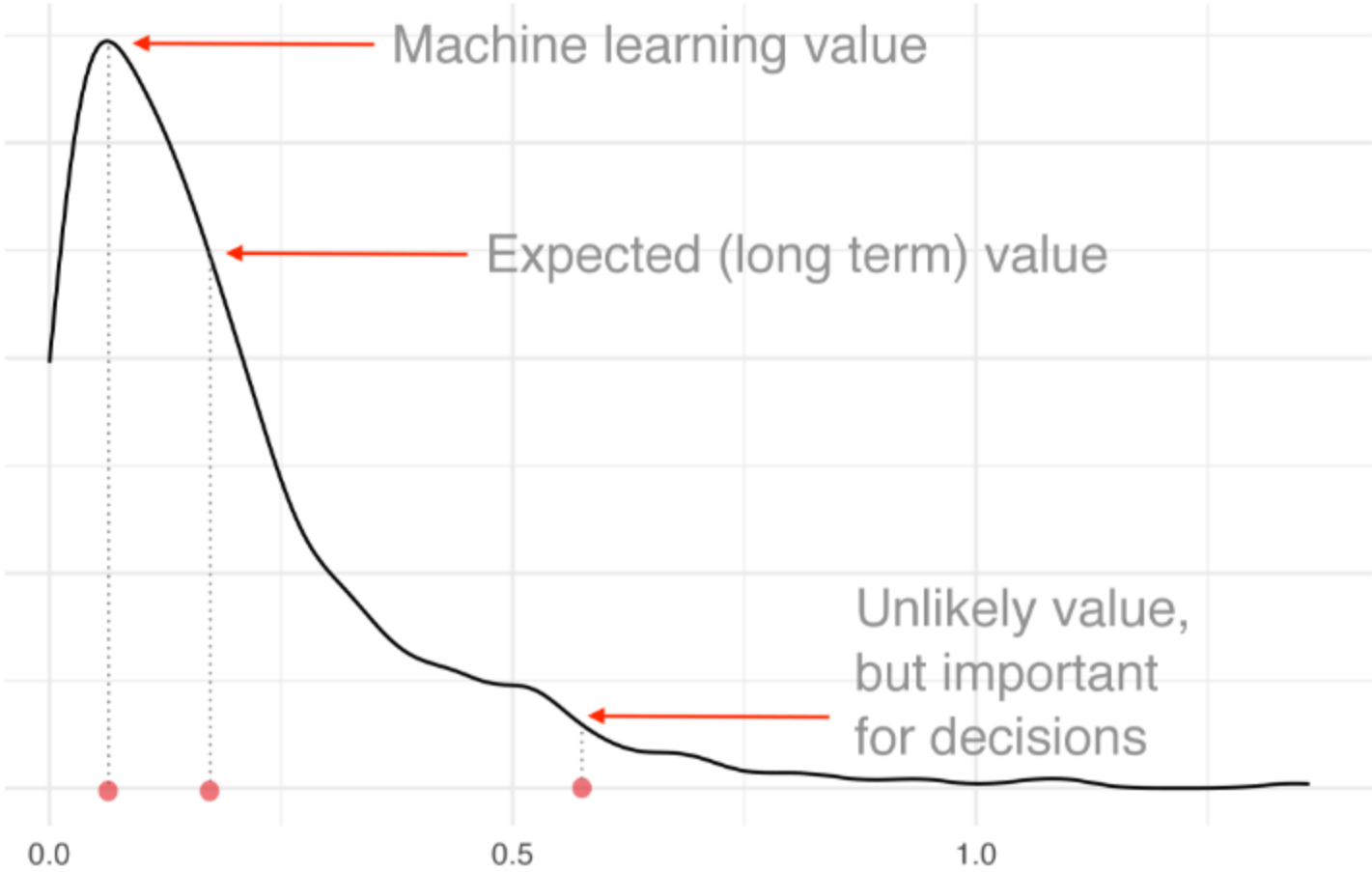
Generable is a Bayesian decision platform built on Stan

We are productizing generative models

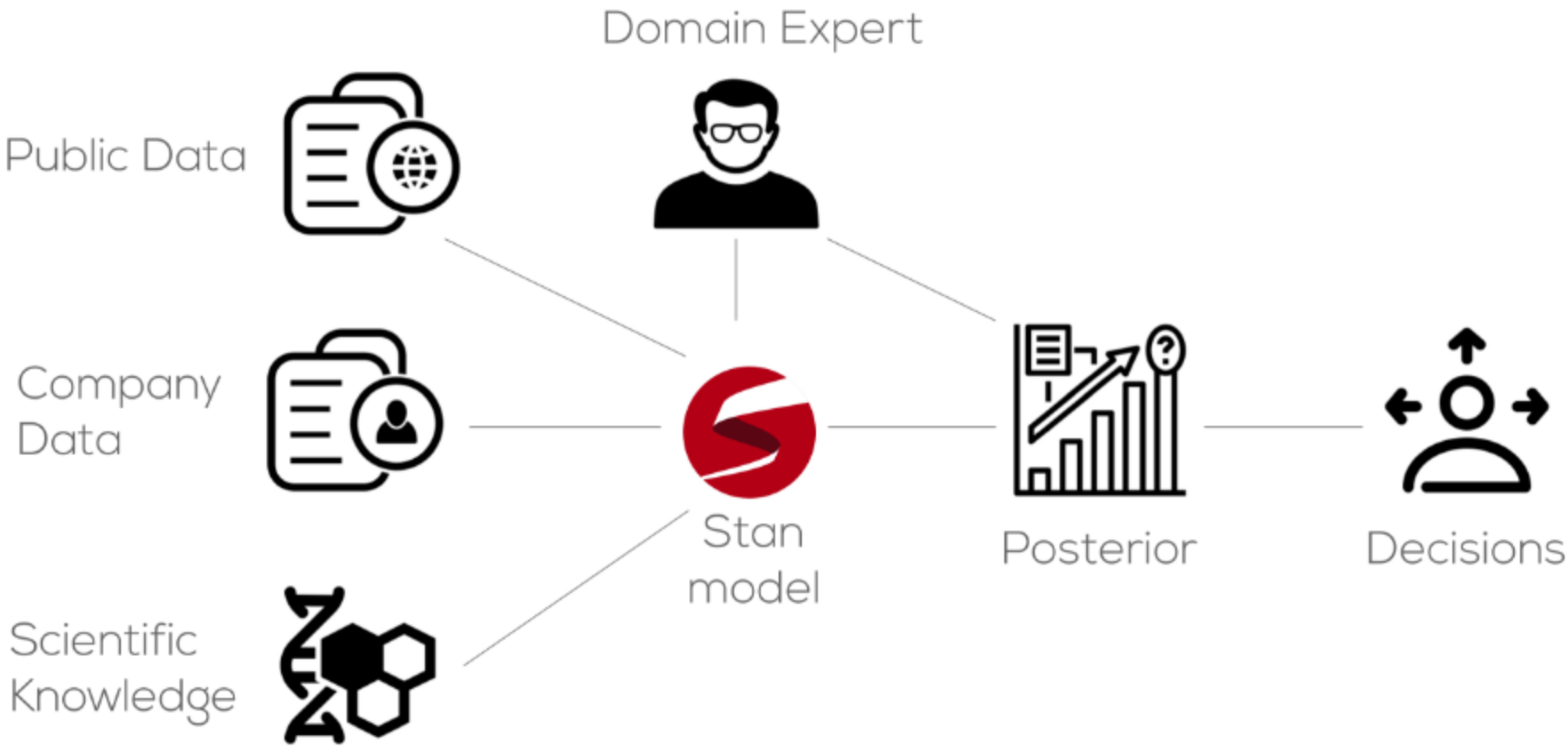
Problem

Real world systems can not be learned from data alone

Why Bother With Bayes



Use All Available Information



Statistical Model as a (Stan) Program

Knowns: X, y

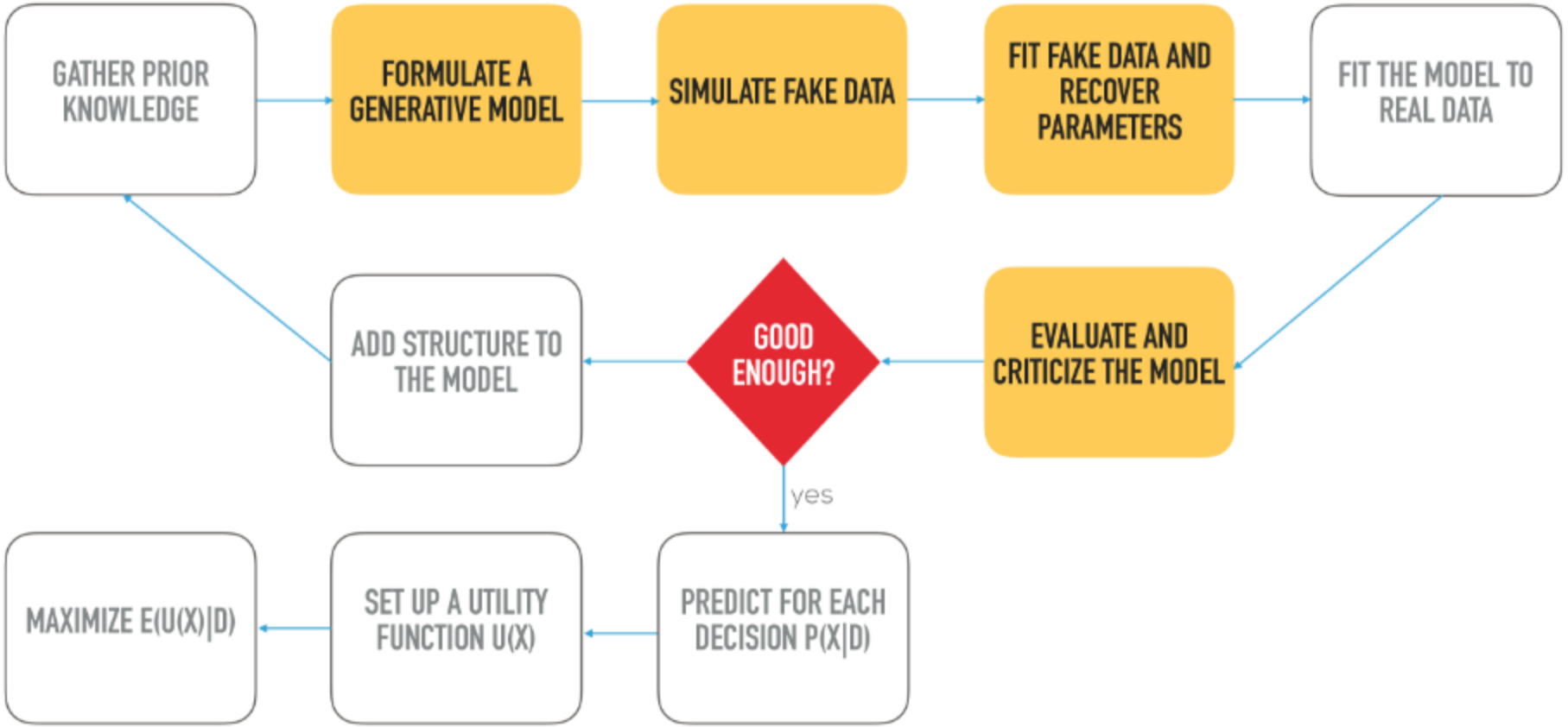
Unknowns: θ

Model: $\log(p(y|X,\theta)) + \log(p(\theta))$

Predictions: $p(\tilde{y}|y)$

```
data {
  int<lower=1>  N;
  int<lower=1>  K;
  matrix[N, K] X;
  vector[N] y;
}
parameters {
  real alpha;
  vector[K] beta;
  real<lower=0> sigma;
}
model {
  y ~ normal(X * beta + alpha, sigma);
  alpha ~ normal(0,10);
  beta ~ normal(0,10);
  sigma ~ cauchy(0,10);
}
generated quantities {
  vector[N] y_rep;
  for (n in 1:N)
    y_rep[n] = normal_rng(X[n,] * beta +
                          alpha, sigma);
}
```

How Do We Find a Good Model?



Clinical Decision Making and Personalized Medicine

In collaboration with:

AstraZeneca

Sam Brilleman, Monash University

Jacki Buros Novik, Hammer Lab

Generable Groups Individuals

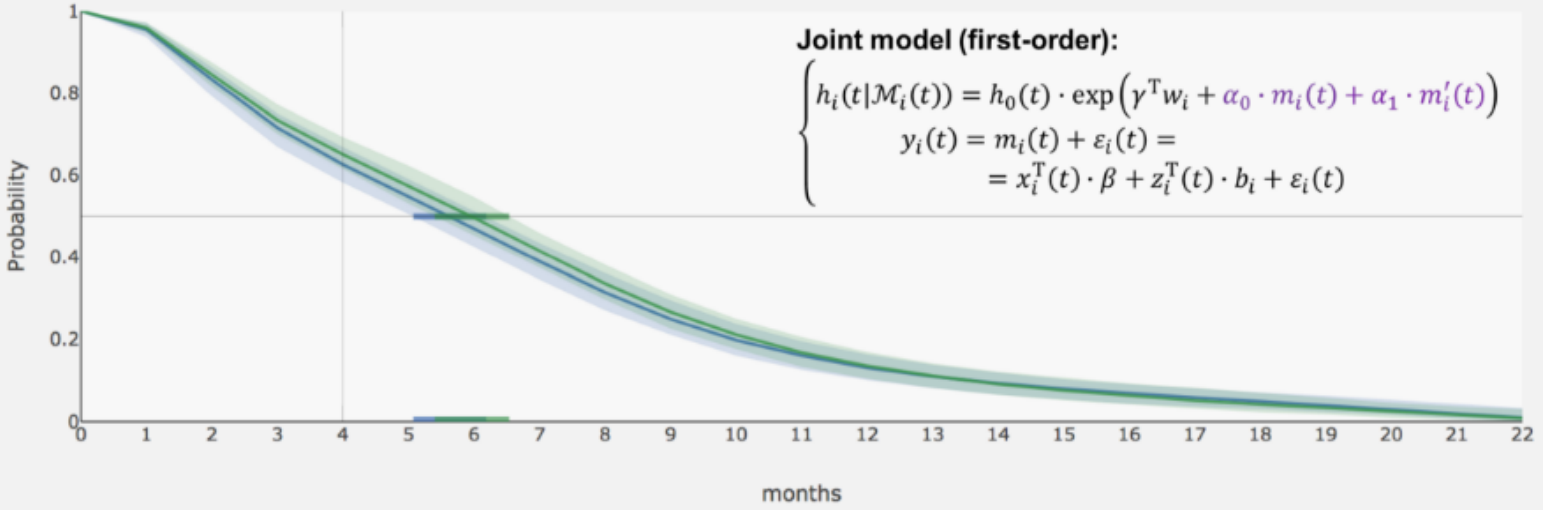
Groups

- Arm
- Mutation
- Arm + Mutation

Compare

A B

Survival Probability



Joint model (first-order):

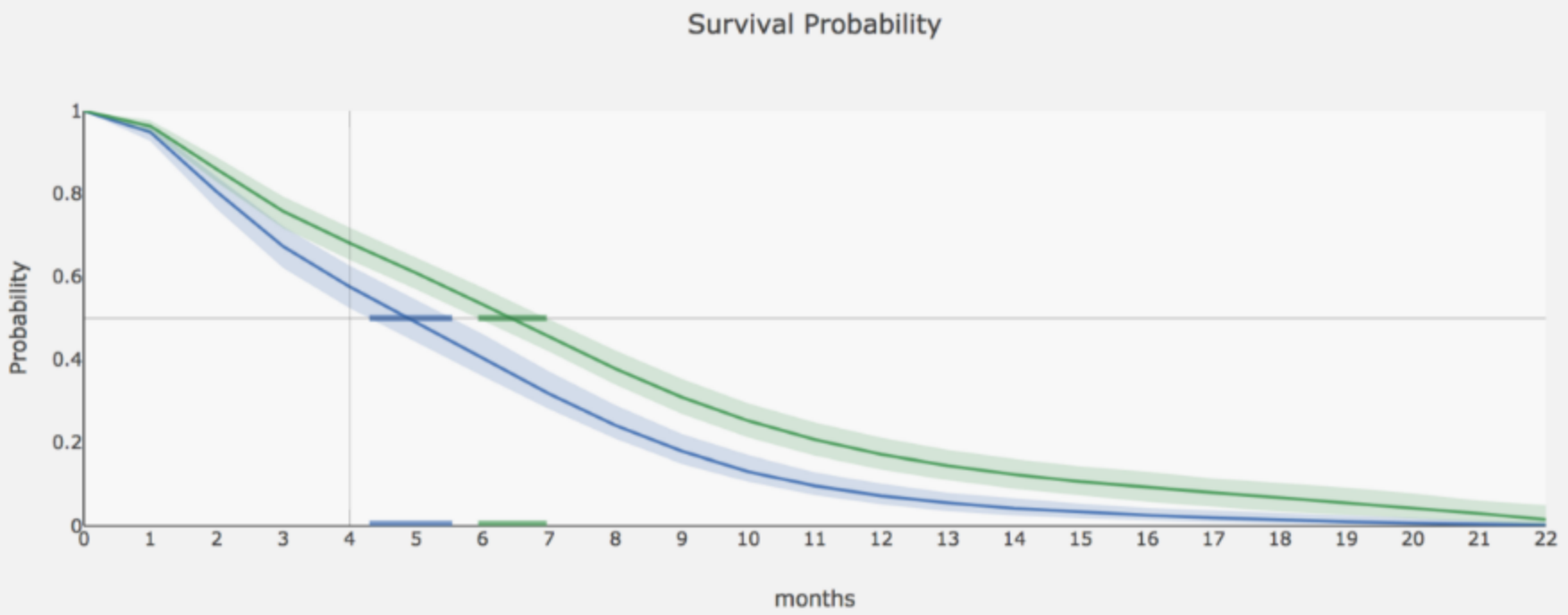
$$\begin{cases} h_i(t|\mathcal{M}_i(t)) = h_0(t) \cdot \exp\left(\gamma^{\mathrm{T}} w_i + \alpha_0 \cdot m_i(t) + \alpha_1 \cdot m_i'(t)\right) \\ y_i(t) = m_i(t) + \varepsilon_i(t) = \\ \quad = x_i^{\mathrm{T}}(t) \cdot \beta + z_i^{\mathrm{T}}(t) \cdot b_i + \varepsilon_i(t) \end{cases}$$

Groups

- Arm
- (•) Mutation
- Arm + Mutation

Compare

Negative × ▾ Positive × ▾

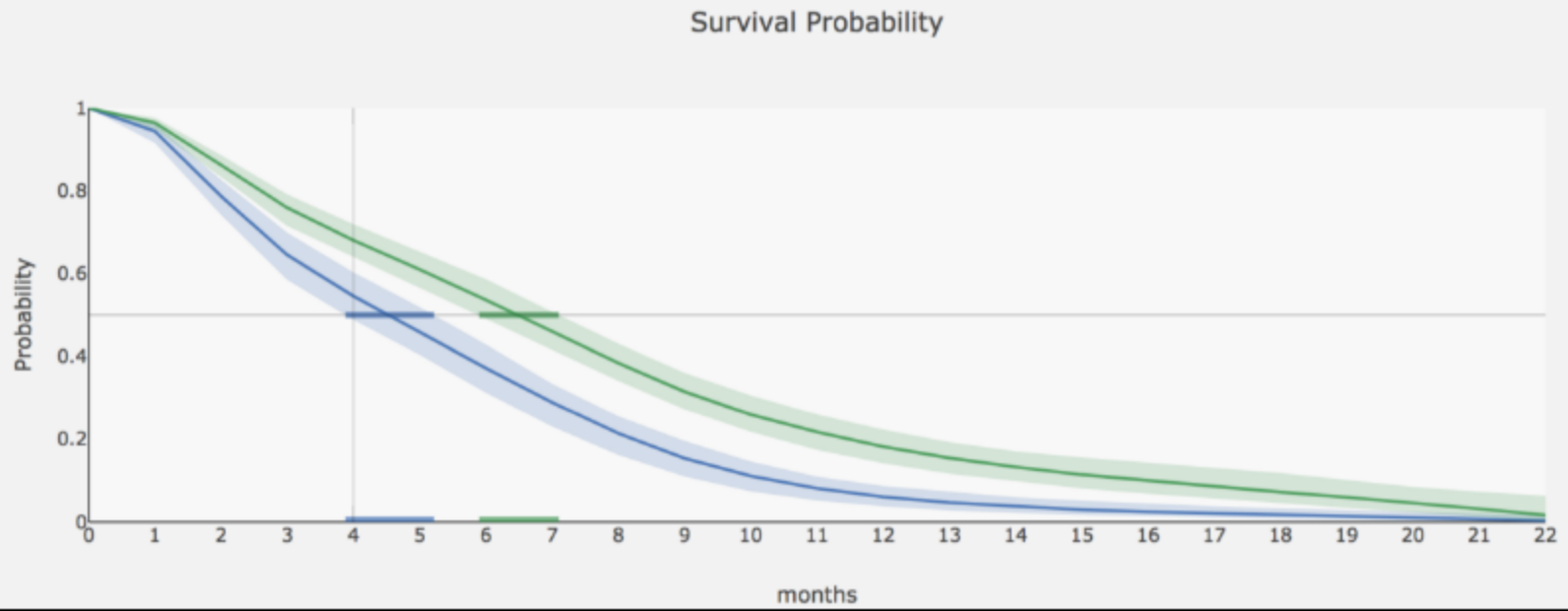


Groups

- Arm
- Mutation
- Arm + Mutation

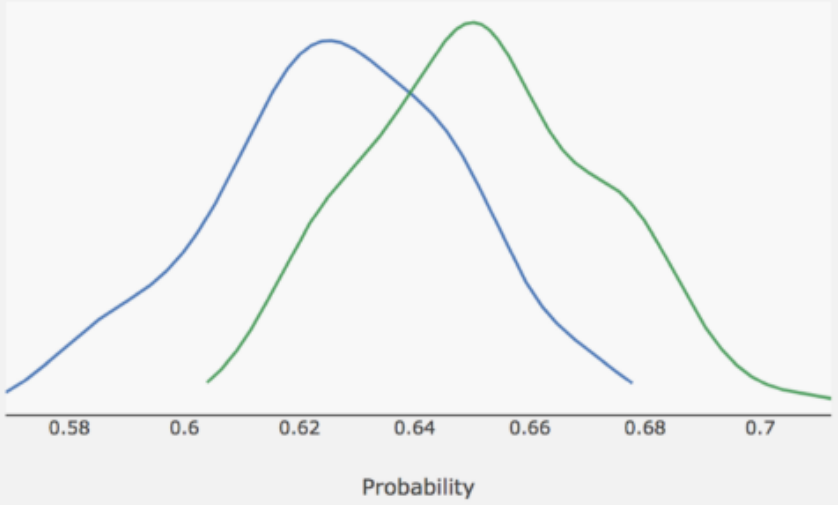
Compare

A + Negative × ▾ A + Positive × ▾

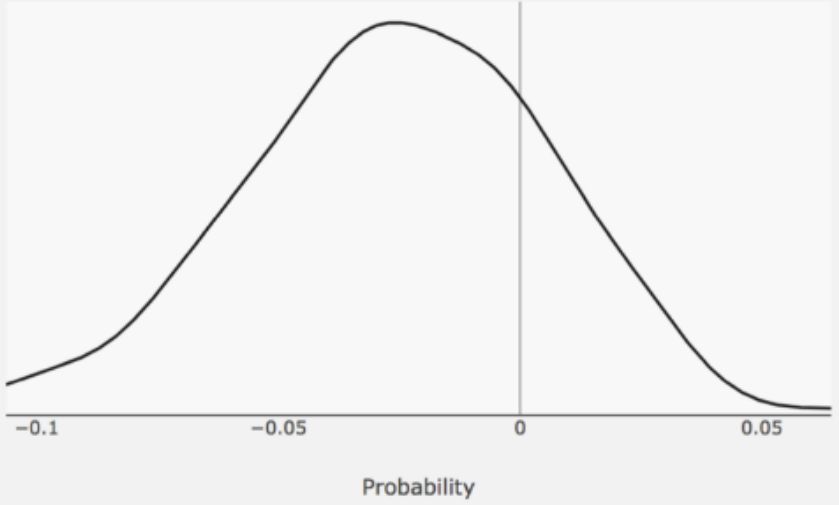


At Month 4

Survival Probability for Groups



Difference in Groups



Individuals

Filter Rows

	uid	arm	mutation	activity	sex	SLD (baseline)	SLD (last)
	Search	Search	Search	Nor	M	<30	Search
■	[redacted]	A	Positive	Normal Activity	M	22	23
	[redacted]	A	Positive	Normal Activity	M	29	28
■	[redacted]	B	Negative	Normal Activity	M	22	20

Selected

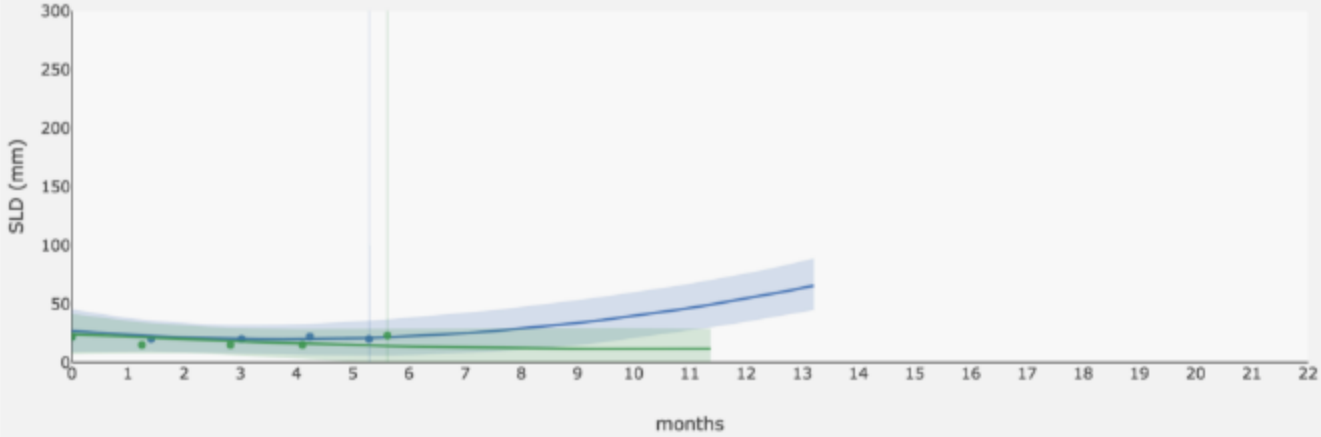
[redacted]

- Arm: B
- Mutation: Negative
- Sex: M
- Activity: Normal Activity
- Baseline SLD: 22
- Last SLD: 20
- Last SLD time: 5.3

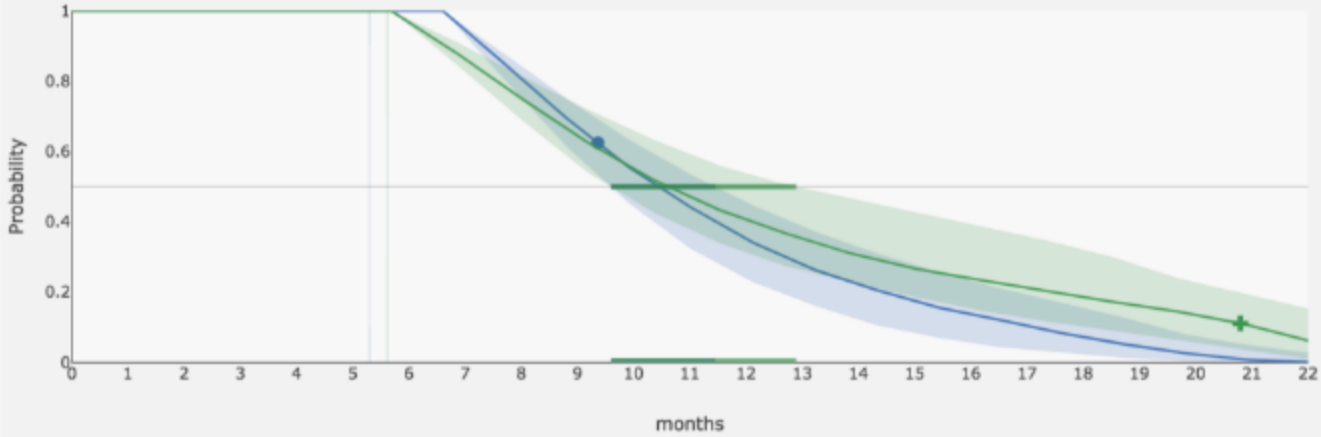
[redacted]

- Arm: A
- Mutation: Positive
- Sex: M
- Activity: Normal Activity
- Baseline SLD: 22
- Last SLD: 23
- Last SLD time: 5.6

Tumor Size



Survival Probability



Thank You!

- eric@generable.com
- @ericnovik, @GenerableHQ
- www.linkedin.com/in/enovik

Data Science Track 2 pm: Productizing Structural Models.
Jim Savage, Lendable