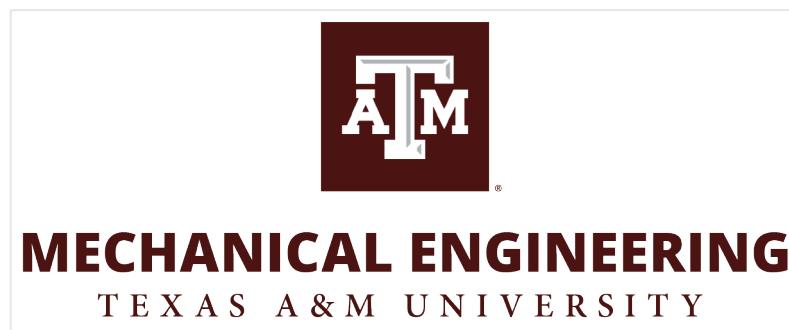


Bad Blood: Combining Data Analytics And Chemical Kinetics to Study Human Blood Coagulation System in Certain Diseases

Jayavel Arumugam

Department of Mechanical Engineering
Texas A&M University



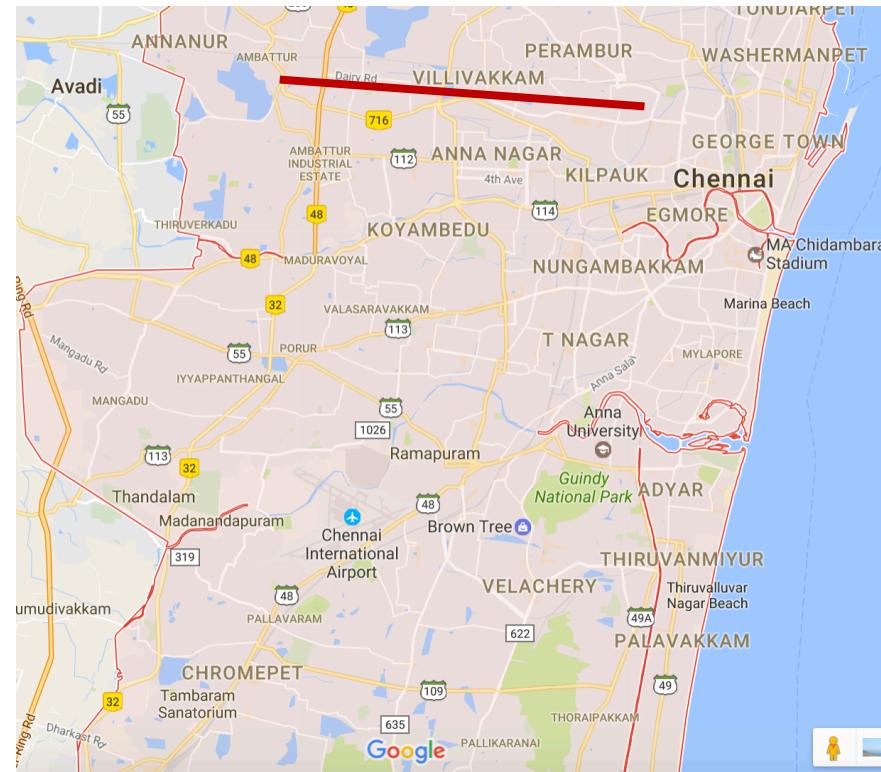
Dr. Arun Srinivasa

Dr. J.N. Reddy

Dr. Krishna Narayanan

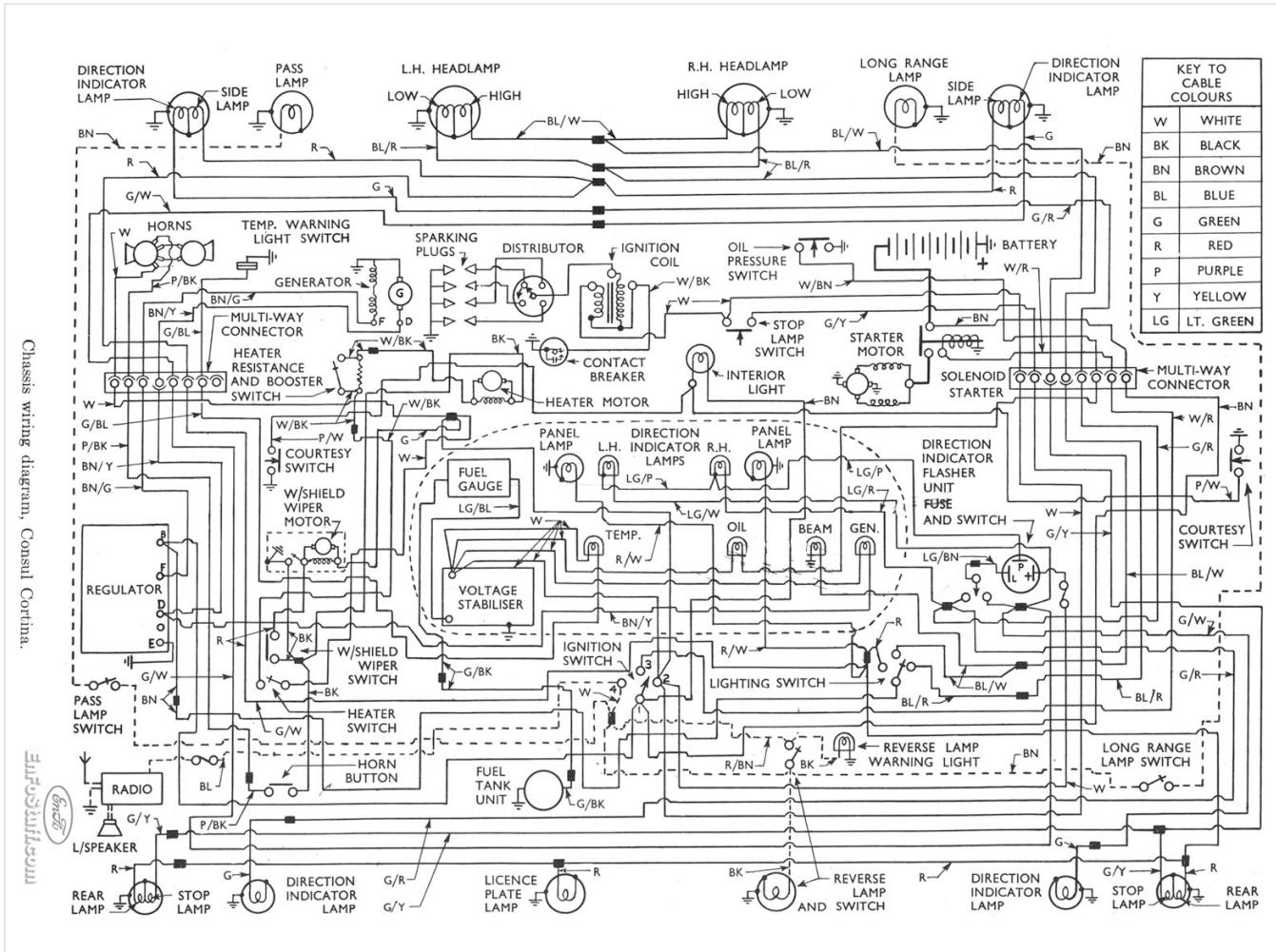
Dr. Alan Freed

I am From India



Ambattur, Villiwakkam, ICF, Ayanavaram, and A&M.

Problem Analogy: Car

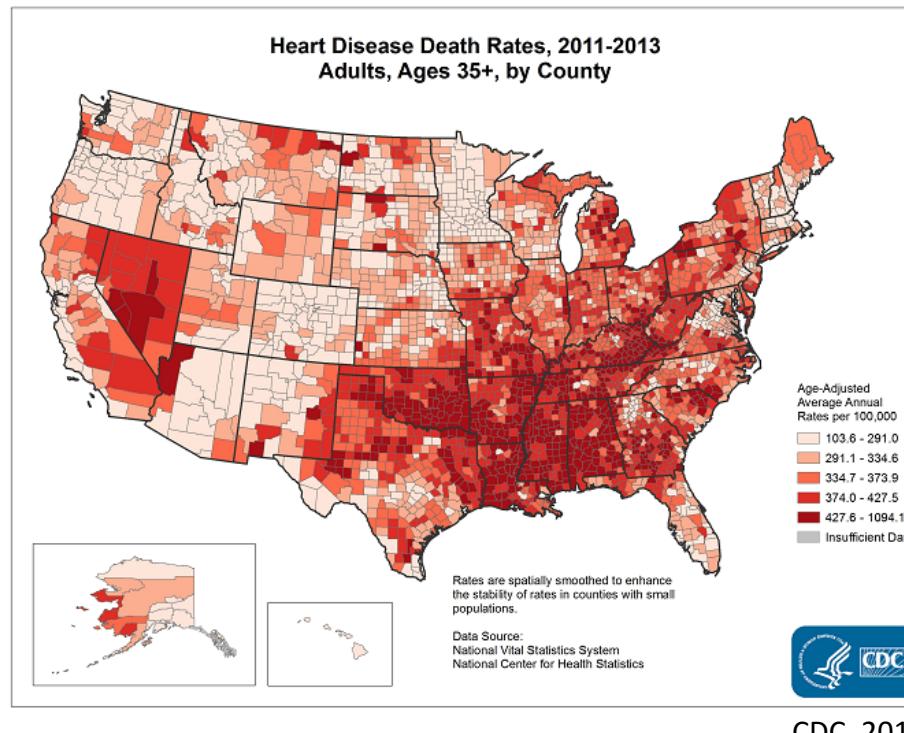


We need to organize information in order to use, diagnose and correct problems.

Introduce the Problem

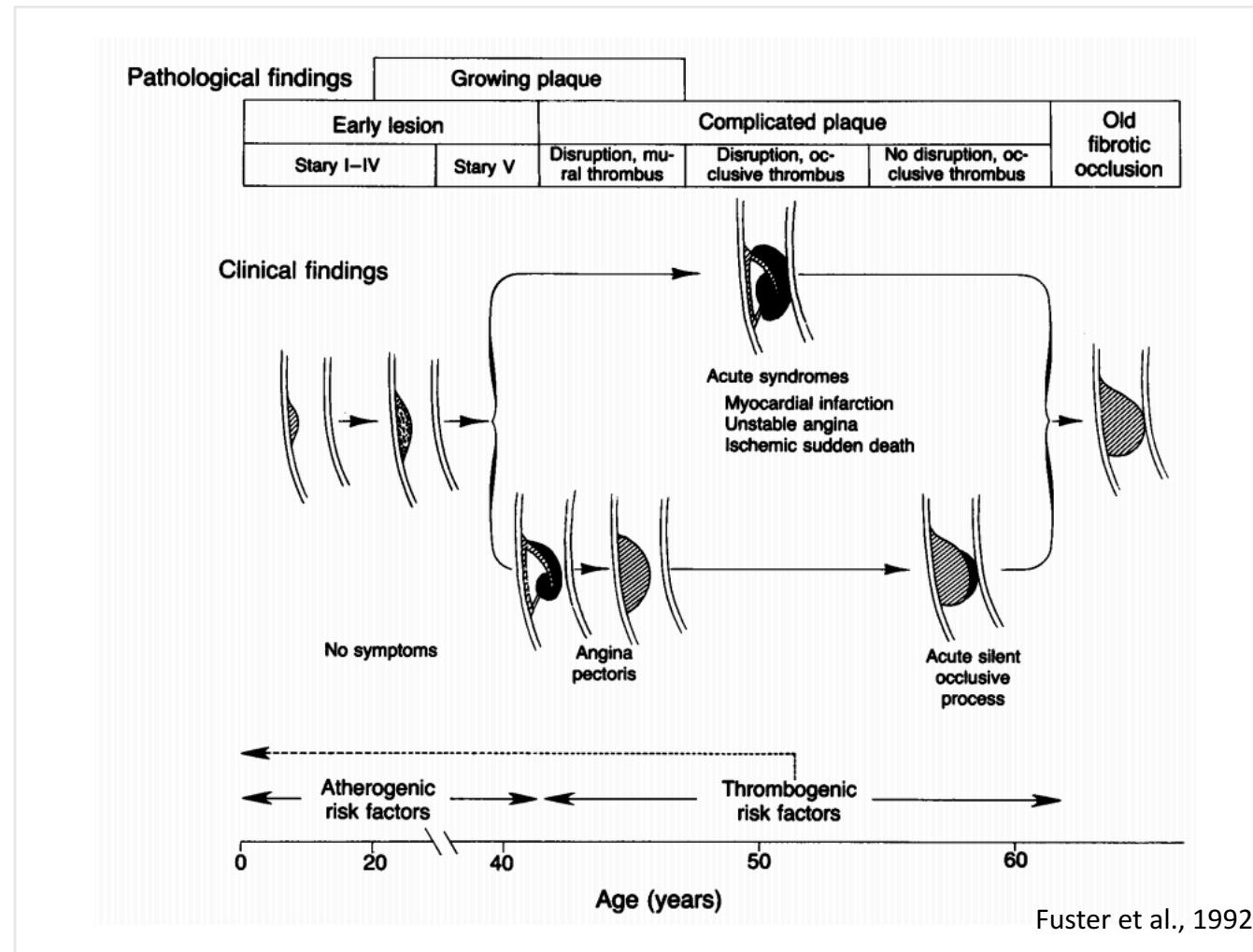
Heart diseases cost \$108.9 billion each year (CDC).

- 610,000 Americans die from heart disease each year (1 in 4 deaths)



Cardiovascular diseases remain a challenging problem.

Introduce the Problem



Acute syndromes are caused by occlusion of heart blood vessels due to clotting.

Extensive Experiments from 19th Century

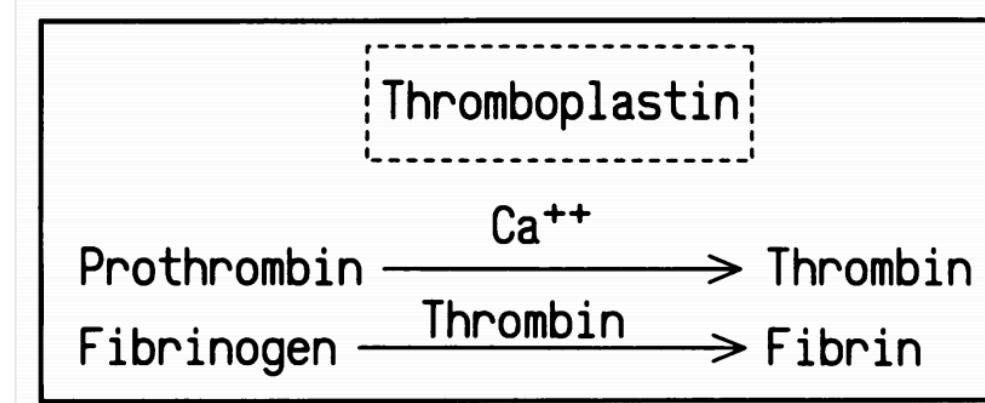
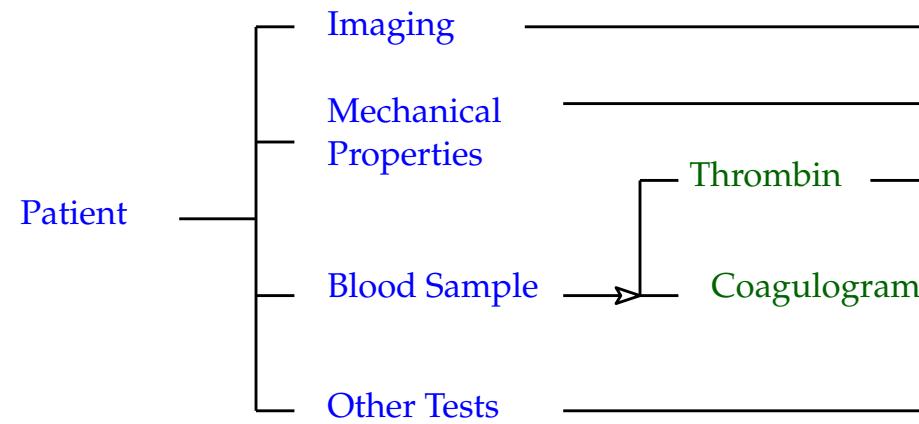


Figure 1.—In the classical theory of blood coagulation, prothrombin is converted to thrombin by the action of thromboplastin (tissue factor) in the presence of calcium ions. Thrombin then acts on the soluble fibrinogen of plasma, converting it to insoluble fibrin.

Rapaport, 1992.
Maurowitz, 1905.

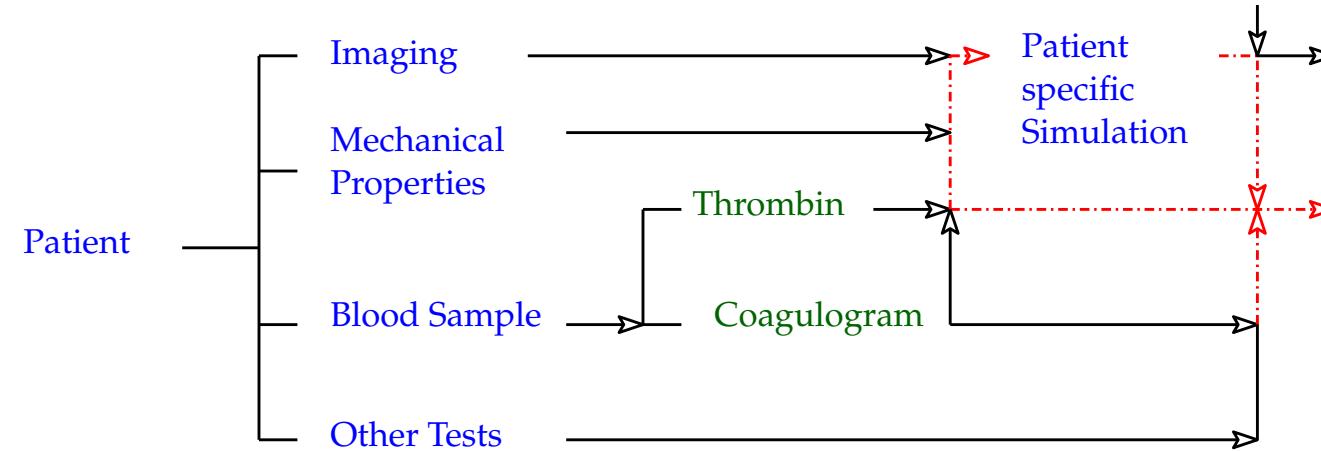
Clotting was recognized as a chemical phenomena.

What We Would Like To Do



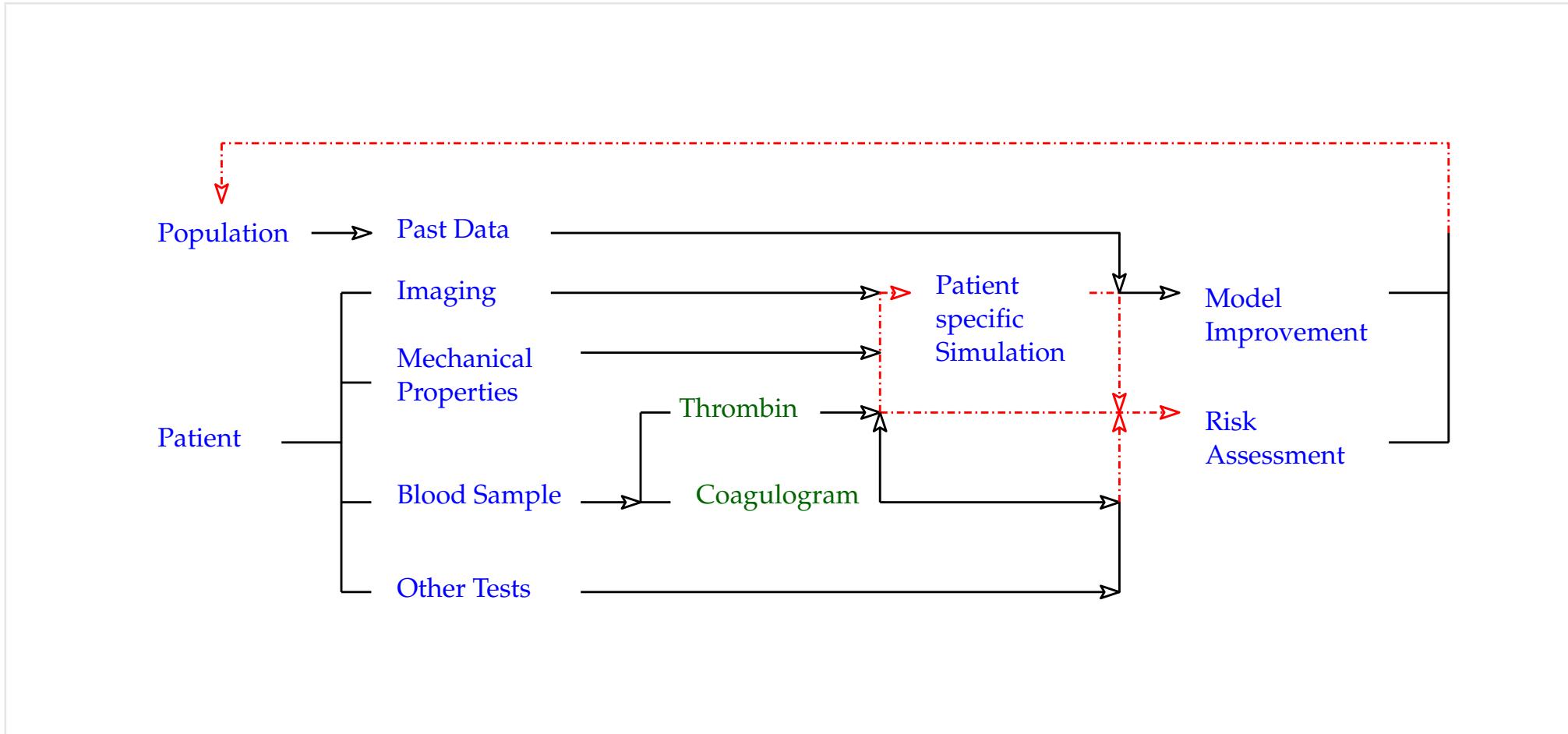
Chemical kinetics poses many questions and provides many answers.

Missing Links



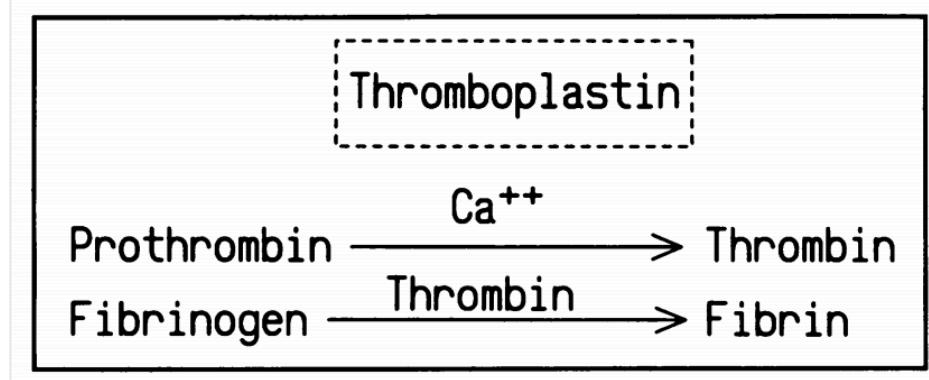
Realistic patient specific simulations is an open problem (Taylor and Humphrey, 2009).

Assess Risk and Screen Patients

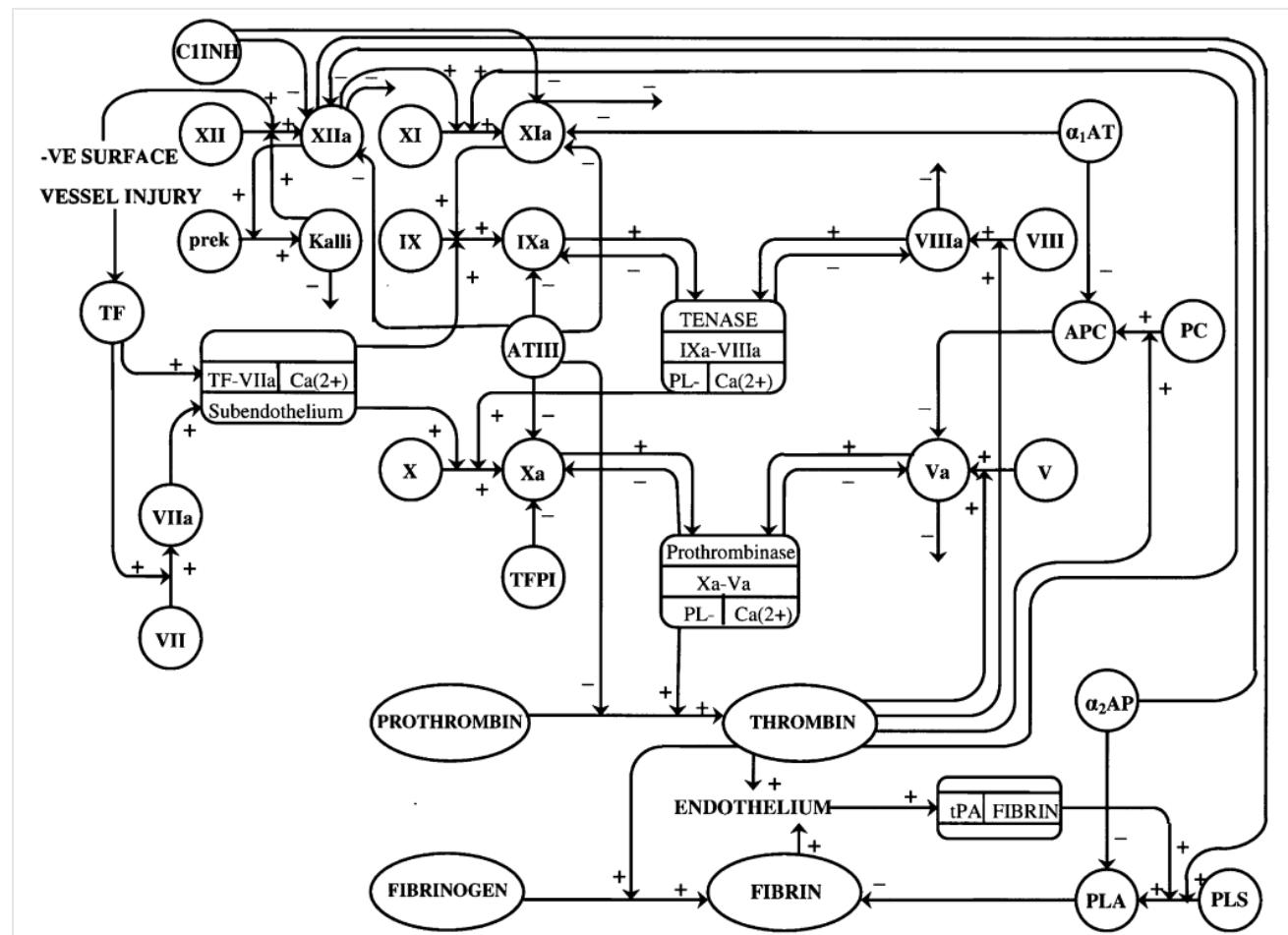


Gaining insights, developing and correcting causal explanations is also important.

But the problem is ...



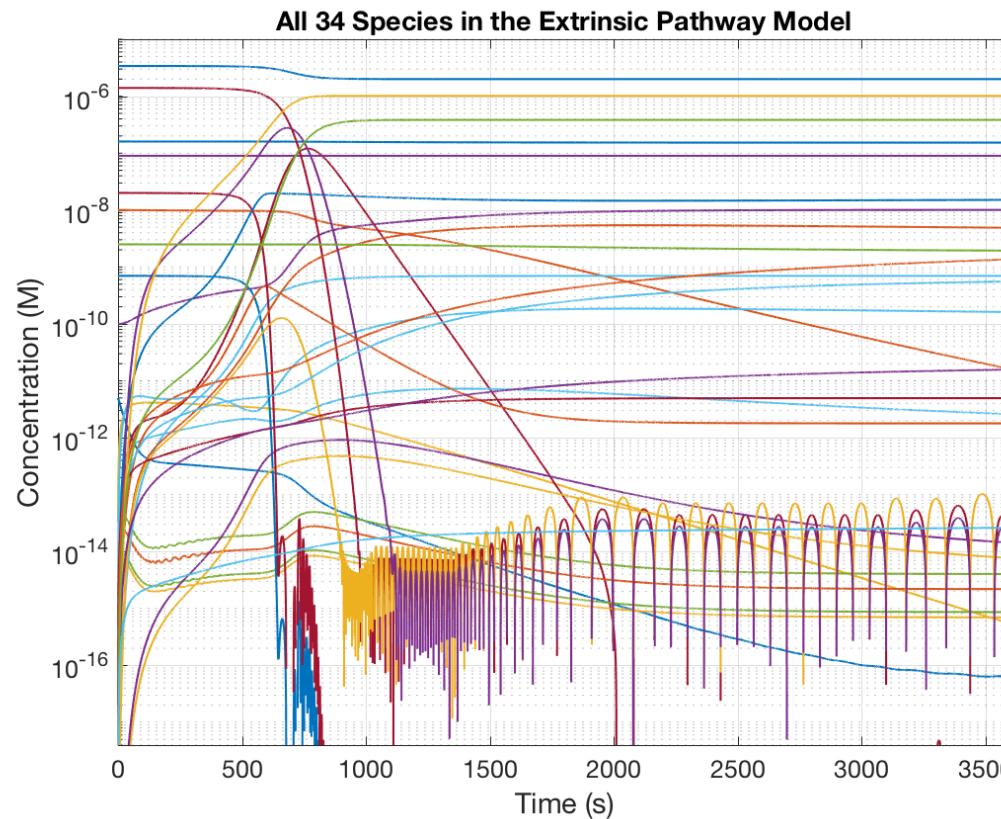
Rapaport, 1992.
Morawitz, 1905.



Lacroix and Anand, 2011.

Simple models evolved to more complex models.

What is challenging

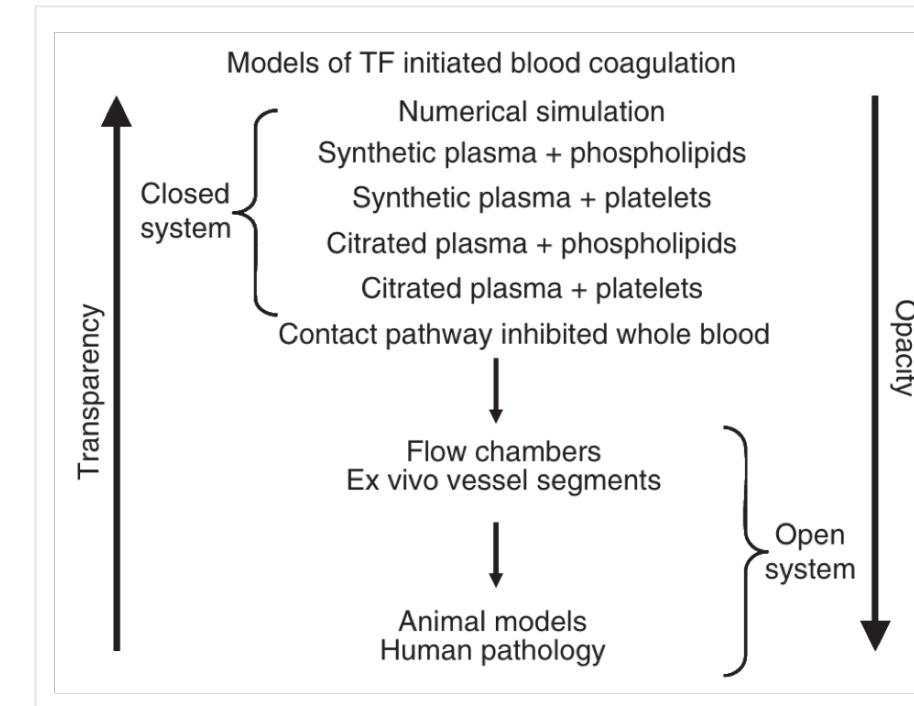


5 pM of tigger creates a macroscopic response.

What is challenging

Hematologists are looking for (Panteleev and Hemker, 2015):

- Better markers for many diseases
- Better ways to characterize abnormal clotting
- Better causal explanations



Mann, 2012.

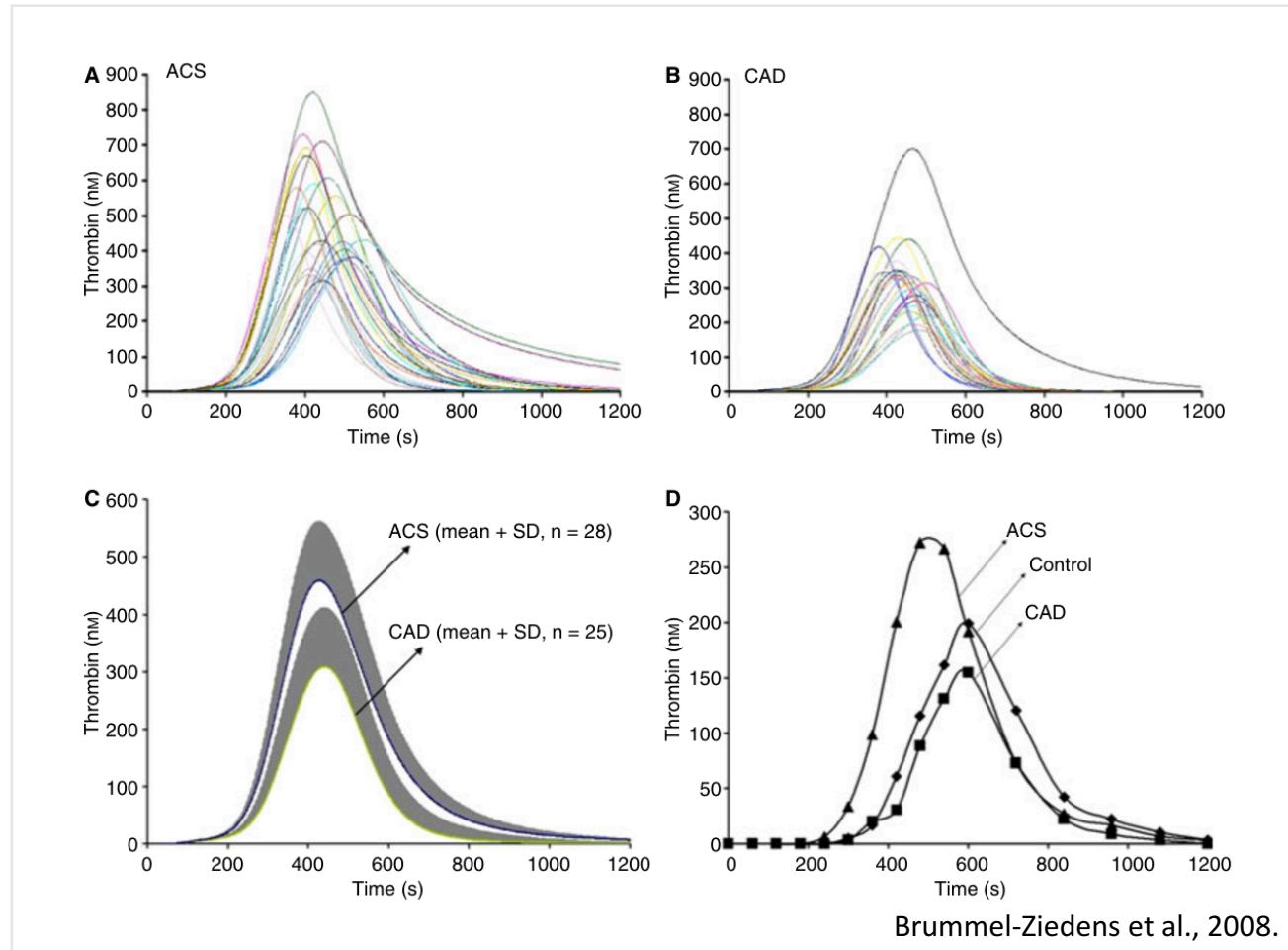
Existing markers and understanding does not suffice to characterize abnormalities.

Abnormalities

1. Protein is missing or present
2. Stoichiometry is abnormal
3. Rates are abnormal
4. Effect of abnormal rates

Stoichiometry affect rates. Rates could have further effects.

Thrombin is the Center of interest



Significant difference in thrombin generation profiles in ACS and CAD populations.

Introduction Summary

- Blood coagulation disorders still remain a challenging problem
- Better sensitive markers are required for disease characterization
- Understanding and improving chemical kinetics models is a critical requirement
- Thrombin generation experiments and models show promise towards individualized diagnosis and treatment
 - Venous thrombosis, Acute Coronary Syndromes, Chronic obstructive pulmonary disease, Acute cerebrovascular diseases, Rheumatoid arthritis

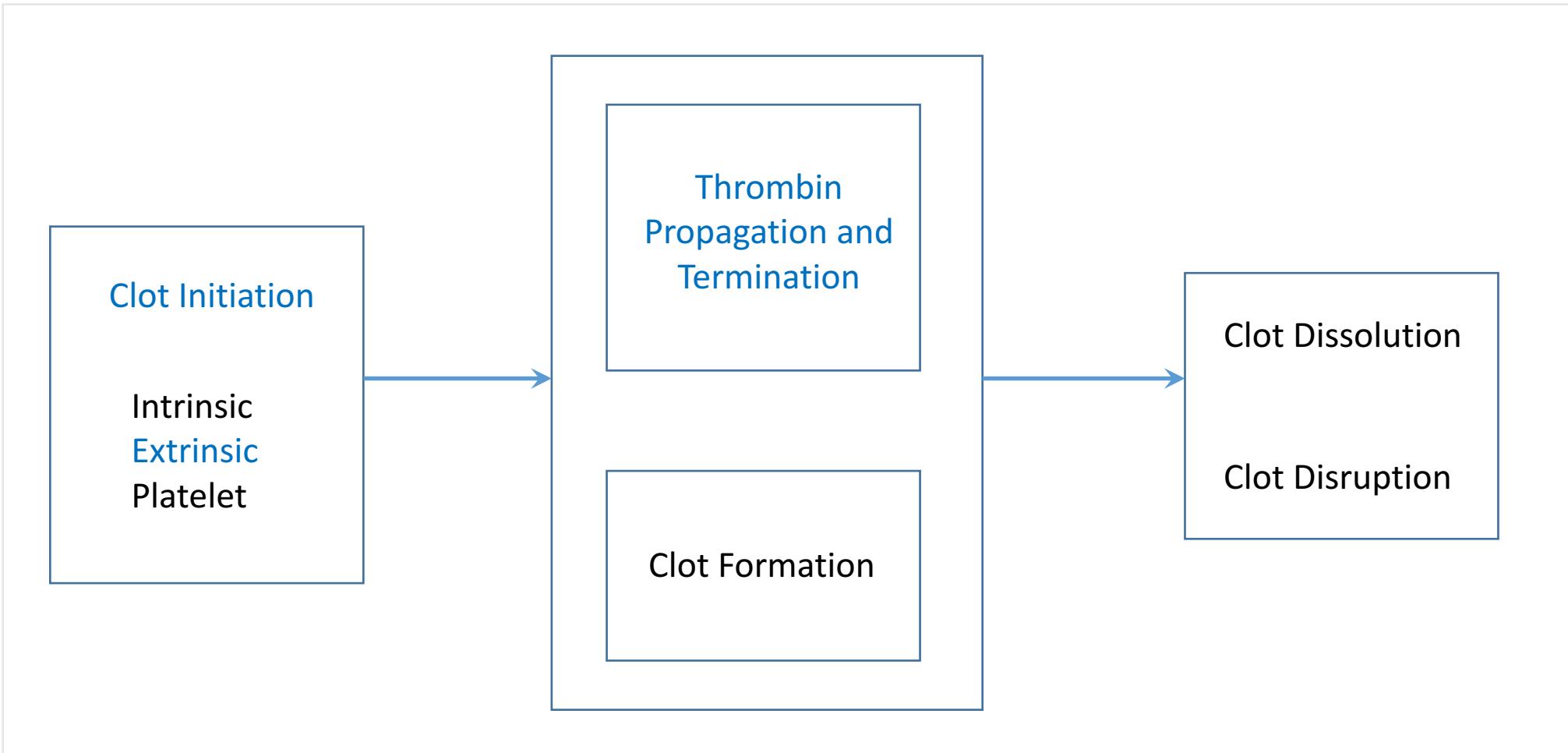
We will start focusing more on thrombin generation.

Research Questions

- How does one systematically characterize abnormalities in stoichiometry and rate information?
- How does one find useful experiments?
- How does one factor out and simplify certain aspects of clotting?

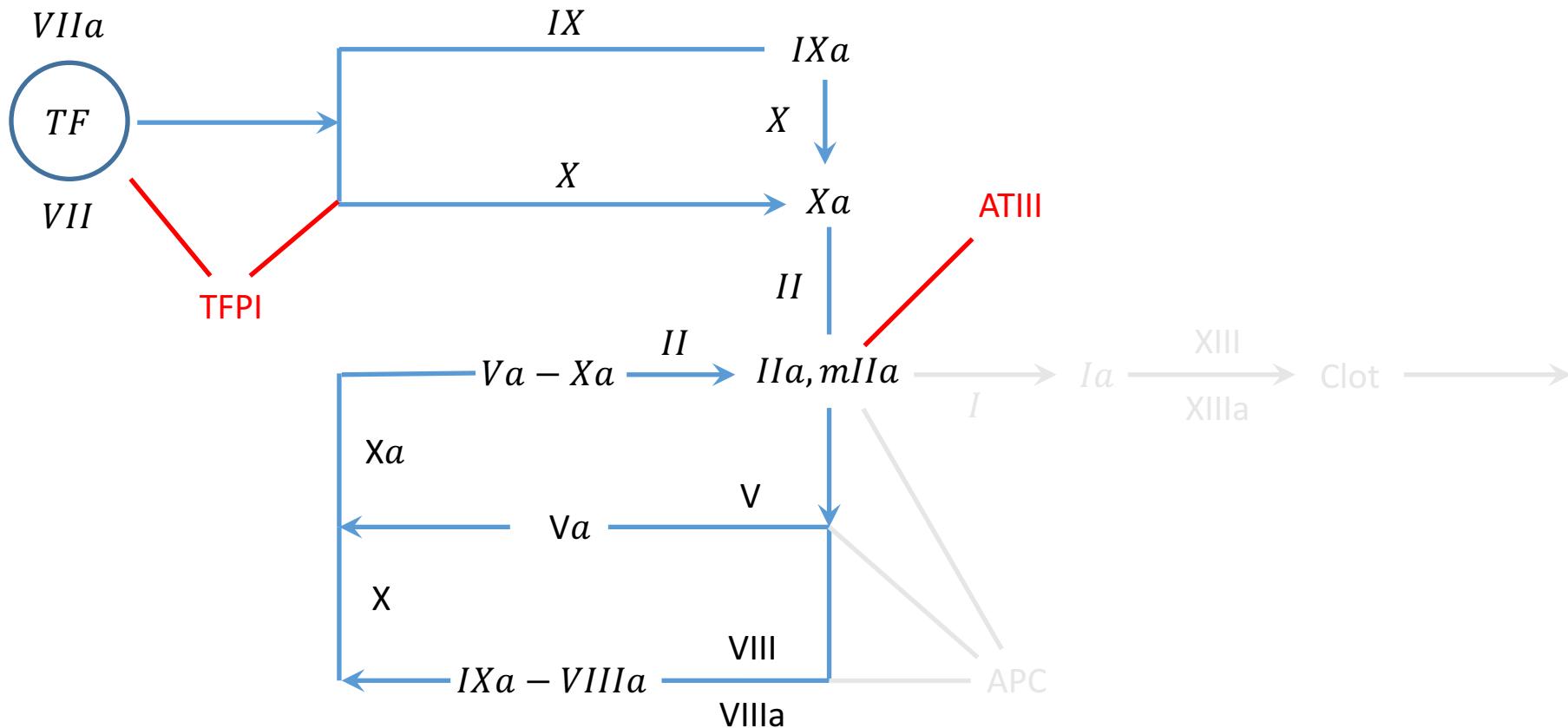
Patient specificity is a key aspect.

Major Steps in Clotting



Scope is studying thrombin dynamics in chemical kinetics.

Extrinsic Pathway Used



A specific model for thrombin generation (Hocking et al., 2002).

Initial Conditions vary from person to person

Table S2: Typical initial coagulation factor concentrations and their normal ranges.

Protein Factor	Variable	Initial Concentration (M)	Normal Range* (%)	LETS Population Range# (%)	In house Population‡ (%)
TF	x_1	5×10^{-12}	Undefined	Undefined	Undefined
VII	x_2	1×10^{-8}	60 - 140	41 -171	76-147
TF=VII	x_3	0			
VIIa	x_4	1×10^{-10}	60 - 140		
TF=VIIa	x_5	0			
Xa	x_6	0			
Ila	x_7	0			
X	x_8	1.6×10^{-7}	60 - 140	46 -163	83-184
TF=VIIa=X	x_9	0			
TF=VIIa=Xa	x_{10}	0			
IX	x_{11}	9×10^{-8}	69 - 151	52 -188	74-151
TF=VIIa=IX	x_{12}	0			
IXa	x_{13}	0			
II	x_{14}	1.4×10^{-6}	60 - 140	55 -153	89-152
VIII	x_{15}	7×10^{-10}	64 - 232	49 -232	99-193
VIIIa	x_{16}	0			
IXa=VIIIa	x_{17}	0			
IXa=VIIIa=X	x_{18}	0			
VIII.1ca1	x_{19}	0			
VIII.a2	x_{20}	0			
V	x_{21}	2×10^{-8}	60 - 140	47 - 302	86-138
Va	x_{22}	0			
Xa=Va	x_{23}	0			
Xa=Va=II	x_{24}	0			
mIIa	x_{25}	0			
TFPI	x_{26}	2.5×10^{-9}	46 - 171	46 -170	88-148
Xa=TFPI	x_{27}	0			
TF=VIIa=Xa=TFPI	x_{28}	0			
AT	x_{29}	3.4×10^{-6}	88 - 174	63 -125	74-131
Xa=AT	x_{30}	0			
mIIa=AT	x_{31}	0			
IXa=AT	x_{32}	0			
Ila=At	x_{33}	0			
TF=VIIa=AT	x_{34}	0			

*: The normal ranges (Reference Intervals) were determined in 2006/2007 by The Fletcher Allen Health Care Special Coagulation/Hematology Laboratory (FAHC, Burlington, Vermont). They tested equal numbers of normal males and females (n=75 plasmas/gender) with 50 donors supplied from Precision Biologic and 25 donors drawn at FAHC. The reference ranges were calculated using the mean +/- 2SD.

#: van der Meer et al (1997) Thromb Haemost 78: 631-5.

‡: Range of thirty-two apparently healthy individuals recruited from hospital and university staff at Jagiellonian University Medical College, Krakow, Poland.

Danforth et al., 2012.

Variations within normal range could give abnormal clotting.

Reaction Rates

TABLE S1: Ordinary differential equations comprising the model.

$\frac{d[TF]}{dt} = -k_2[TF][VII] + k_1[TF = VII] - k_4[TF][VIIa] + k_3[TF = VIIa]$	(1)	$\frac{d[VII]}{dt} = -k_2[TF][VII] + k_1[TF = VII] - k_5[TF = VIIa][VII] - k_6[Xa][VII] - k_7[IIa][VII]$	(2)	$\frac{d[IXa = VIIa]}{dt} = k_{19}[VIIa][IXa] - k_{18}[IXa = VIIa] - k_{21}[IXa = VIIa][X] + k_{20}[IXa = VIIa = X] + k_{22}[IXa = VIIa = X] - k_{25}[IXa = VIIa]$	(17)
$\frac{d[TF = VII]}{dt} = -k_1[TF = VII] + k_2[TF][VII]$	(3)	$\frac{d[VIIa]}{dt} = -k_4[TF][VIIa] + k_3[TF = VIIa] + k_5[TF = VIIa][VII] + k_6[Xa][VII] + k_7[IIa][VII]$	(4)	$\frac{d[VIII.lcal]}{dt} = k_{21}[IXa = VIIa][X] - k_{20}[IXa = VIIa = X] - k_{22}[IXa = VIIa = X] - k_{25}[IXa = VIIa = X]$	(18)
$\frac{d[VIIa]}{dt} = -k_4[TF][VIIa] + k_3[TF = VIIa] + k_5[TF = VIIa][VII] + k_6[Xa][VII] + k_7[IIa][VII]$	(5)	$\frac{d[VIII.a2]}{dt} = k_{24}[VIIa] + k_{25}[IXa = VIIa = X] + k_{25}[IXa = VIIa] - k_{23}[VIII.lcal][VIII.a2]$	(19)	$\frac{d[VIII.a2]}{dt} = k_{24}[VIIa] + k_{25}[IXa = VIIa = X] + k_{25}[IXa = VIIa] - k_{23}[VIII.lcal][VIII.a2]$	(20)
$\frac{d[TF = VIIa]}{dt} = -k_3[TF = VIIa] + k_4[TF][VIIa] - k_9[TF = VIIa][X] + k_8[TF = VIIa = X] - k_{12}[TF = VIIa][Xa] + k_{11}[TF = VIIa = Xa]$ $- k_{14}[TF = VIIa][IX] + k_{13}[TF = VIIa = IX] + k_{15}[TF = VIIa = IX] - k_{37}[TF = VIIa][Xa = TFP] - k_{42}[TF = VIIa][AT]$	(6)	$\frac{d[V]}{dt} = -k_{26}[IIa][V] - k_{44}[mIIa][V]$	(21)	$\frac{d[Va]}{dt} = k_{26}[IIa][V] - k_{28}[Xa][Va] + k_{27}[Xa = Va] + k_{44}[mIIa][V]$	(22)
$\frac{d[Xa]}{dt} = -k_{12}[TF = VIIa][Xa] + k_{11}[TF = VIIa = Xa] + k_{22}[IXa = VIIa = X] - k_{28}[Xa][Va] + k_{27}[Xa = Va] - k_{34}[Xa][TFP]$ $+ k_{33}[Xa = TFP] - k_{38}[Xa][AT] + k_{43}[IXa][X]$	(7)	$\frac{d[Xa = Va]}{dt} = k_{28}[Xa][Va] - k_{27}[Xa = Va] - k_{30}[Xa = Va][II] + k_{29}[Xa = Va = II] + k_{31}[Xa = Va = II]$	(23)	$\frac{d[Xa = Va = II]}{dt} = k_{30}[Xa = Va][II] - k_{29}[Xa = Va = II] - k_{31}[Xa = Va = II]$	(24)
$\frac{d[IIa]}{dt} = k_{16}[Xa][II] + k_{32}[mIIa][Xa = Va] - k_{41}[IIa][AT]$	(8)	$\frac{d[mIIa]}{dt} = k_{31}[Xa = Va = II] - k_{32}[mIIa][Xa = Va] - k_{39}[mIIa][AT]$	(25)	$\frac{d[TFP]}{dt} = -k_{34}[Xa][TFP] + k_{33}[Xa = TFP] - k_{36}[TF = VIIa = Xa][TFP] + k_{35}[TF = VIIa = Xa = TFP]$	(26)
$\frac{d[X]}{dt} = -k_9[TF = VIIa][X] + k_8[TF = VIIa = X] - k_{21}[IXa = VIIa][X] + k_{20}[IXa = VIIa = X] + k_{25}[IXa = VIIa = X] - k_{43}[IXa][X]$	(9)	$\frac{d[Xa = TFP]}{dt} = k_{34}[Xa][TFP] - k_{33}[Xa = TFP] - k_{37}[TF = VIIa][Xa = TFP]$	(27)	$\frac{d[Xa = TFP]}{dt} = k_{36}[TF = VIIa = Xa][TFP] - k_{35}[TF = VIIa = Xa = TFP] + k_{37}[TF = VIIa][Xa = TFP]$	(28)
$\frac{d[TF = VIIa = X]}{dt} = k_9[TF = VIIa][X] - k_{10}[TF = VIIa = X] - k_8[TF = VIIa = X]$	(10)	$\frac{d[Xa = AT]}{dt} = -k_{38}[Xa][AT] - k_{39}[mIIa][AT] - k_{40}[IXa][AT] - k_{41}[IIa][AT] - k_{42}[TF = VIIa][AT]$	(29)	$\frac{d[Xa = AT]}{dt} = k_{38}[Xa][AT]$	(30)
$\frac{d[TF = VIIa = Xa]}{dt} = k_{10}[TF = VIIa = X] + k_{12}[TF = VIIa][Xa] - k_{11}[TF = VIIa = Xa] - k_{36}[TF = VIIa = Xa][TFP] + k_{35}[TF = VIIa = Xa = TFP]$	(11)	$\frac{d[mIIa = AT]}{dt} = k_{39}[mIIa][AT]$	(31)	$\frac{d[IXa = AT]}{dt} = k_{40}[IXa][AT]$	(32)
$\frac{d[IX]}{dt} = -k_{14}[TF = VIIa][IX] + k_{13}[TF = VIIa = IX]$	(12)	$\frac{d[IXa = AT]}{dt} = k_{40}[IXa][AT]$	(33)	$\frac{d[IIa = AT]}{dt} = k_{41}[IIa][AT]$	(34)
$\frac{d[TF = VIIa = IX]}{dt} = k_{14}[TF = VIIa][IX] - k_{13}[TF = VIIa = IX] - k_{15}[TF = VIIa = IX]$	(13)	$\frac{d[TF = VIIa = AT]}{dt} = k_{42}[TF = VIIa][AT]$			
$\frac{d[IXa]}{dt} = k_{15}[TF = VIIa = IX] - k_{19}[VIIa][IXa] + k_{18}[IXa = VIIa] + k_{25}[IXa = VIIa = X] + k_{25}[IXa = VIIa] - k_{40}[IXa][AT]$	(14)				
$\frac{d[II]}{dt} = -k_{16}[Xa][II] - k_{30}[Xa = Va][II] + k_{29}[Xa = Va = II]$	(15)				
$\frac{d[VII]}{dt} = -k_{17}[IIa][VII]$	(16)				
$\frac{d[VIIa]}{dt} = k_{17}[IIa][VII] - k_{19}[VIIa][IXa] + k_{18}[IXa = VIIa] - k_{24}[VIIa] + k_{23}[VIII.lcal][VIII.a2]$	(17)				
$\frac{d[IXa = VIIa]}{dt} = k_{19}[VIIa][IXa] - k_{18}[IXa = VIIa] - k_{21}[IXa = VIIa][X] + k_{20}[IXa = VIIa = X] + k_{22}[IXa = VIIa = X] - k_{25}[IXa = VIIa]$					

Danforth et al., 2012.

One of those rare models where stoichiometry could be verified.

Structure

- **Sampling**
 - How do we do patient specific simulation?
- **Density Estimation**
 - How do we estimate likelihoods of various clotting features?
- **Classification**
 - How do we classify using all variables?
- **Significance Selection**
 - How do we find important features
- **Simplification**
 - Is there a simplified model for the dynamics of thrombin generation?

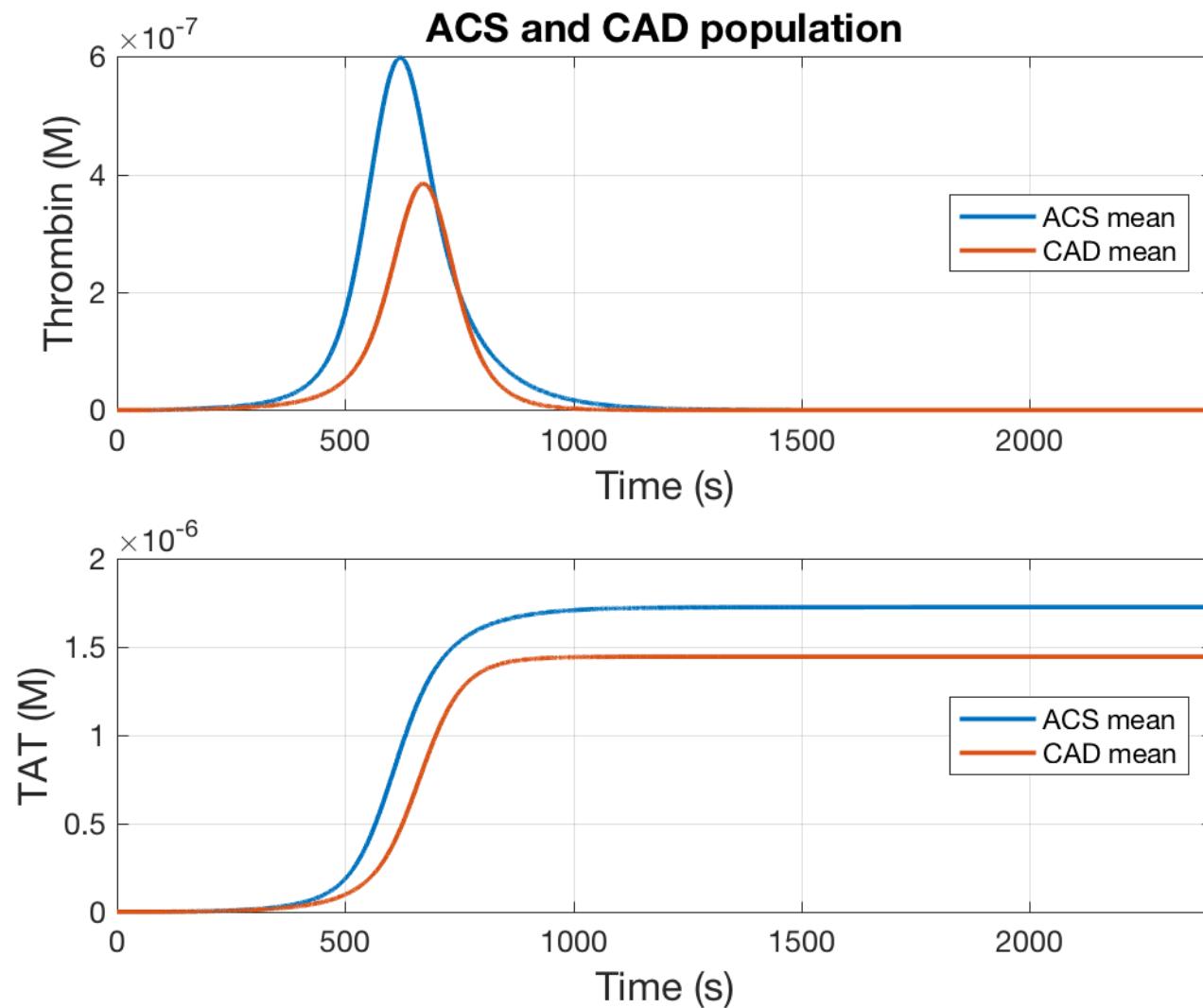
Idea is to ask progressively harder questions.

Structure

- Sampling
 - How do we do patient specific simulation?
- Density Estimation
 - How do we estimate likelihoods of various clotting features?
- Classification
 - How do we classify using all variables?
- Significance Selection
 - How do we find important features
- Simplification
 - Is there a simplified model for the dynamics of thrombin generation?

How do we sample necessary data?

Simulation



This is one of the known differences that is useful.

Plasma Factor Composition

Table 1 Plasma composition

Protein	ACS, n = 28 (%, mean ± SD)	CAD, n = 25 (%, mean ± SD)	Healthy range (%)
Factor II	123 ± 21*	103 ± 20	60–140
Factor V	121 ± 28	131 ± 41	60–140
Factor VII	116 ± 29	121 ± 33	60–140
Factor VIII	162 ± 34*	123 ± 30	64–232
Factor IX	117 ± 29	121 ± 32	69–151
Factor X	111 ± 24	114 ± 22	60–140
TFPI	114 ± 19*	102 ± 12	46–171
Antithrombin	89 ± 12*	113 ± 15	86–128

ACS, acute coronary syndrome; CAD, coronary artery disease; TFPI, tissue factor pathway inhibitor. *P < 0.01 when compared with CAD.

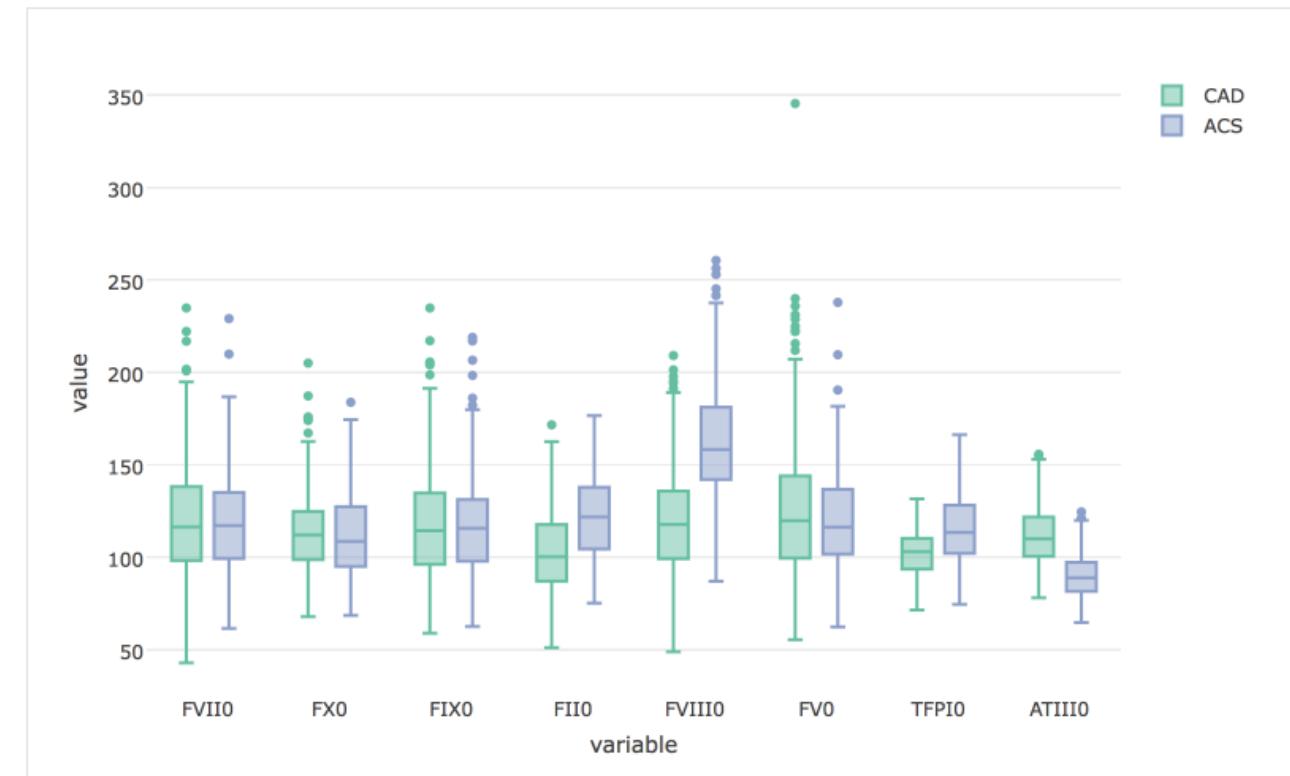
Brummel-Ziedens et al., 2008.

Population data is given. How do we do patient specific simulations?

Maximum Entropy

$$\hat{p}(x) = \arg \min_{p(x)} - \int p(x) \log p(x)$$

$$\text{s.t. } f_i(p(x)) = 0, i = 1, \dots, N$$



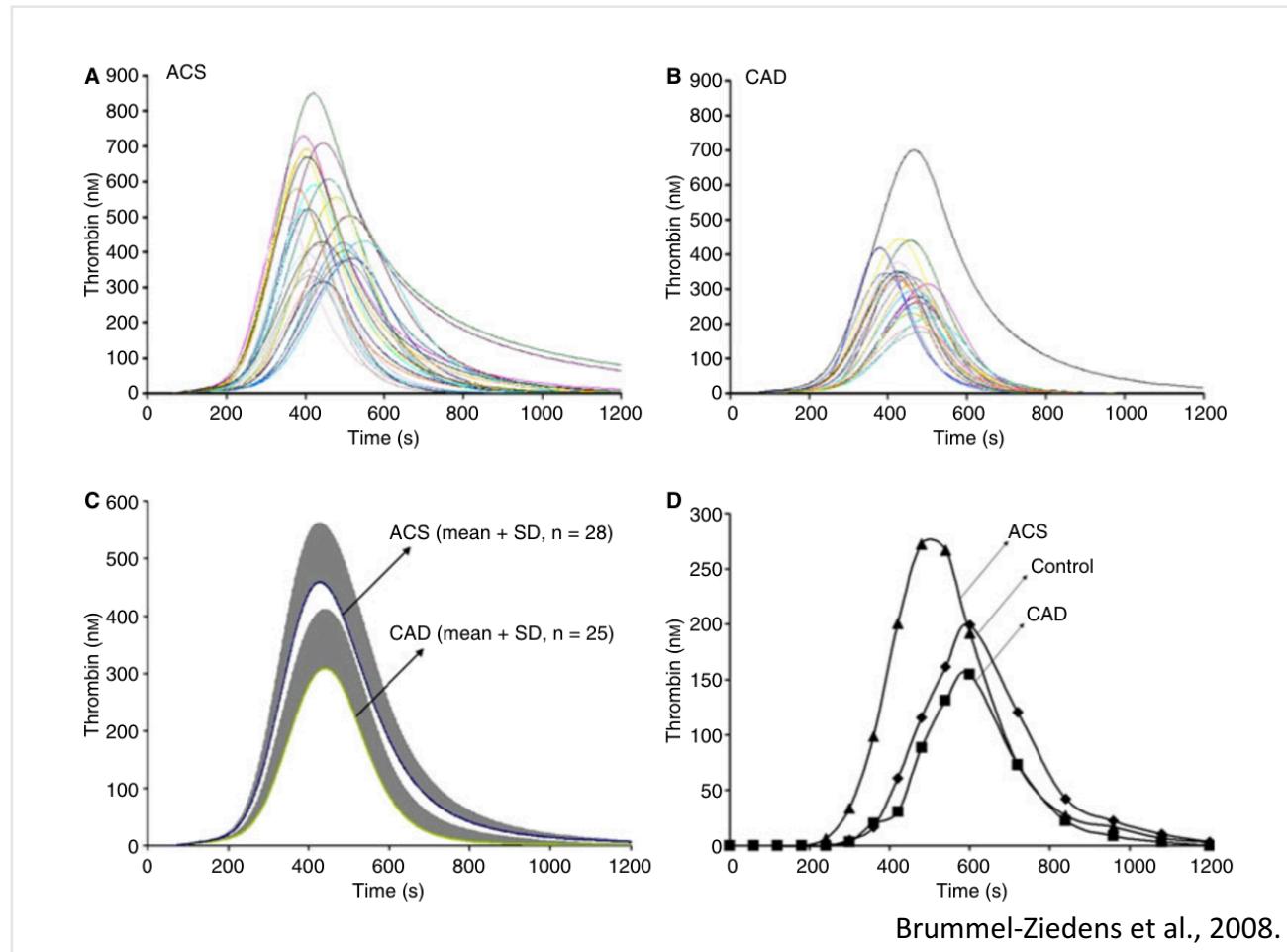
Maximize ignorance subject to known constraints.

Structure

- Sampling
 - How do we do patient specific simulation?
- Density Estimation
 - How do we estimate likelihoods of various clotting features?
- Classification
 - How do we classify using all variables?
- Significance Selection
 - How do we find important features
- Simplification
 - Is there a simplified model for the dynamics of thrombin generation?

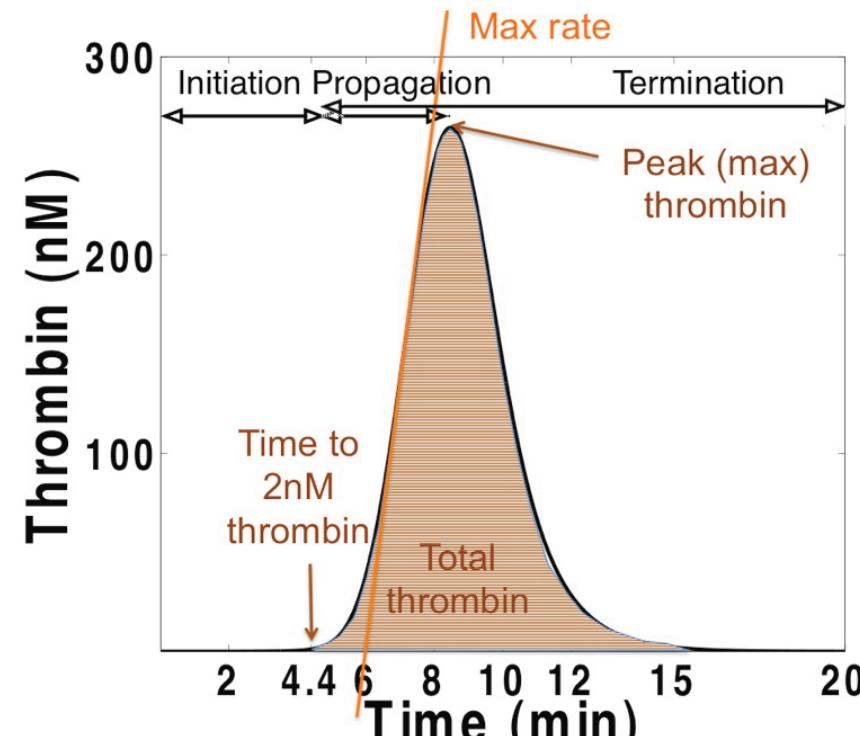
How do we find patterns in the data?

Classification Problem

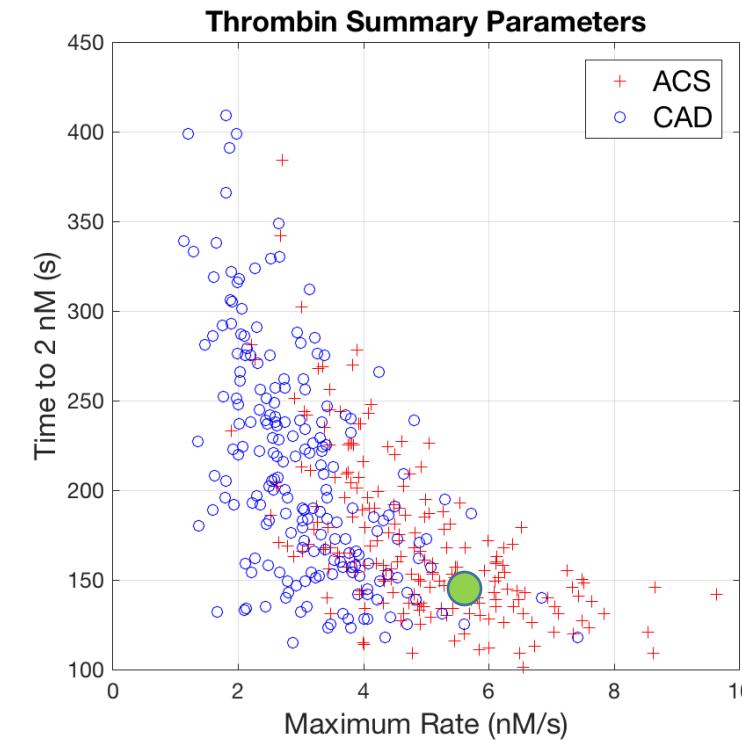


Given a new thrombin generation curve, can we classify it?

Density Estimation



Danforth et al., 2012.



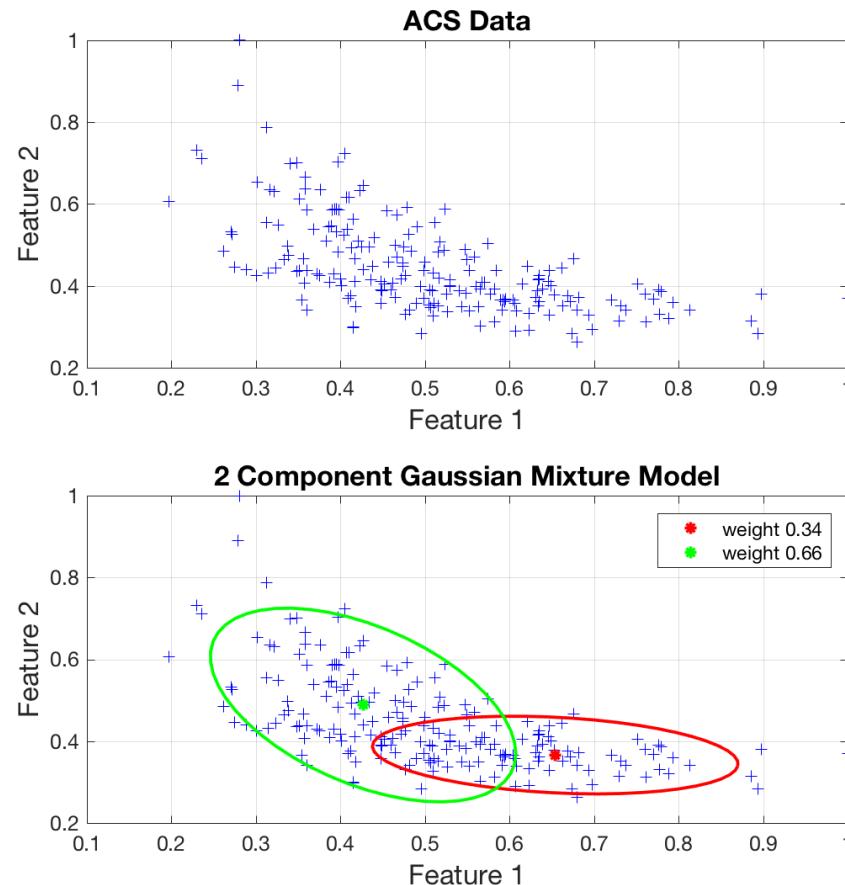
One way to classify is to find the respective probability distributions.

Gaussian Mixture Model

Likelihood for a given class,

$$p(x | \theta) = \sum_{i=1}^K \alpha_i \mathcal{N}(\mu_i, \Sigma_i)$$

Find $\theta = \{\alpha_i, \mu_i, \Sigma_i\}$ by maximizing likelihood.



Finding the parameters is a hard problem.

Expectation Maximization

1. Initialize parameters θ_i .

2. Find Expectation

$$Q(\theta, \theta_i) = E[\log p(x, y|\theta) | y, \theta_i]$$

3. Maximize Expectation

$$\theta_{i+1} = \arg \max_{\theta} Q(\theta, \theta_i)$$

4. Iterate 2 and 3 until convergence.

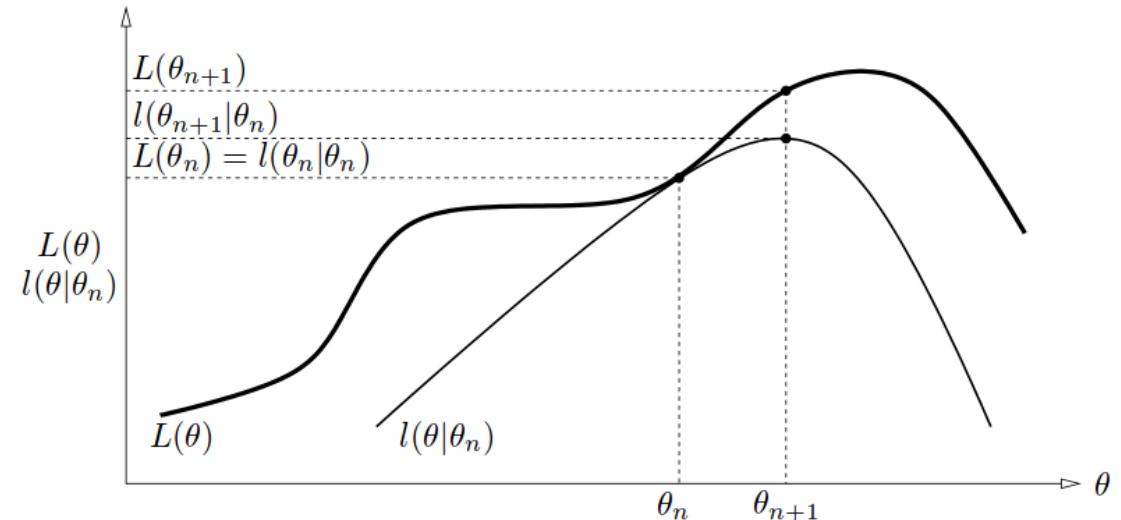
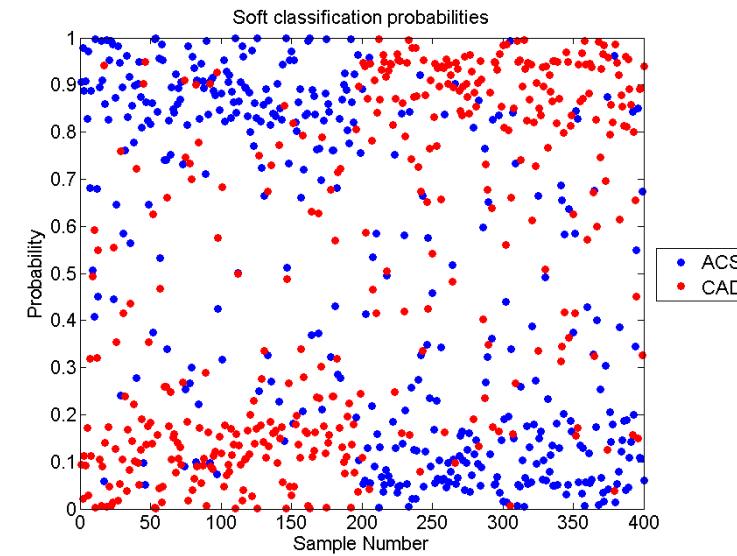
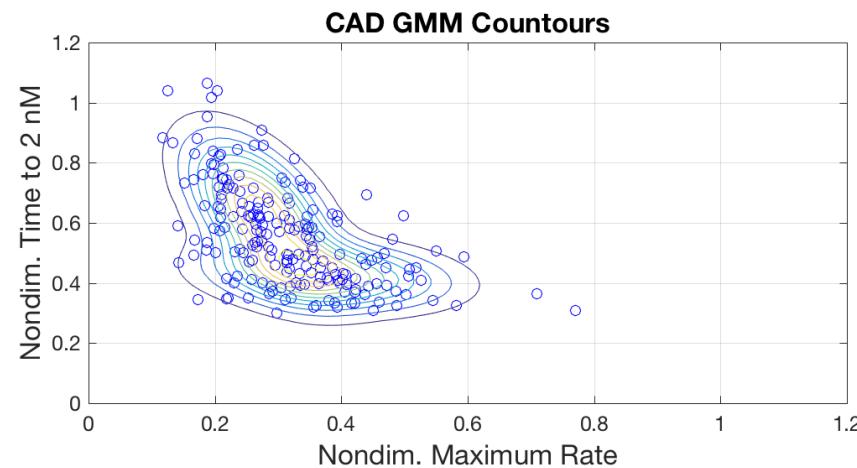
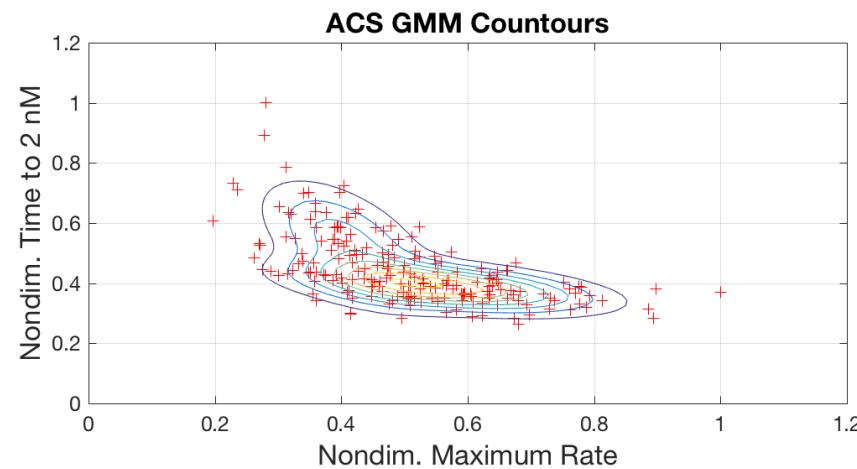


Figure 2: Graphical interpretation of a single iteration of the EM algorithm: The function $L(\theta|\theta_n)$ is upper-bounded by the likelihood function $L(\theta)$. The functions are equal at $\theta = \theta_n$. The EM algorithm chooses θ_{n+1} as the value of θ for which $l(\theta|\theta_n)$ is a maximum. Since $L(\theta) \geq l(\theta|\theta_n)$ increasing $l(\theta|\theta_n)$ ensures that the value of the likelihood function $L(\theta)$ is increased at each step.

Borman, 2004.

Expectation maximization (Bilmes, 1998) cleverly solves a difficult problem.

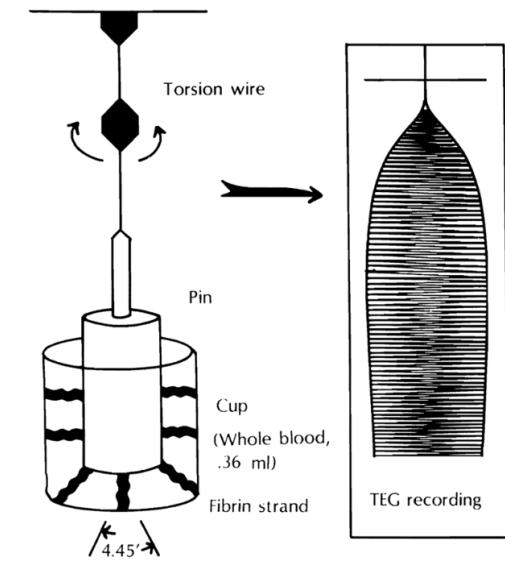
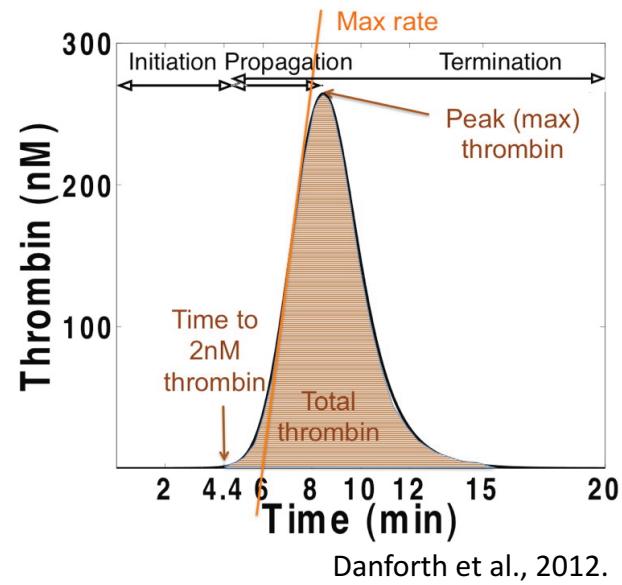
Gaussian Mixture Results



Gaussian Mixture Models helped calculate likelihoods. We used it to classify.

Density Estimation Applications

- Plasma factor composition
- Thrombin generation parameters
- Thromboelastography



Kang, 1985.

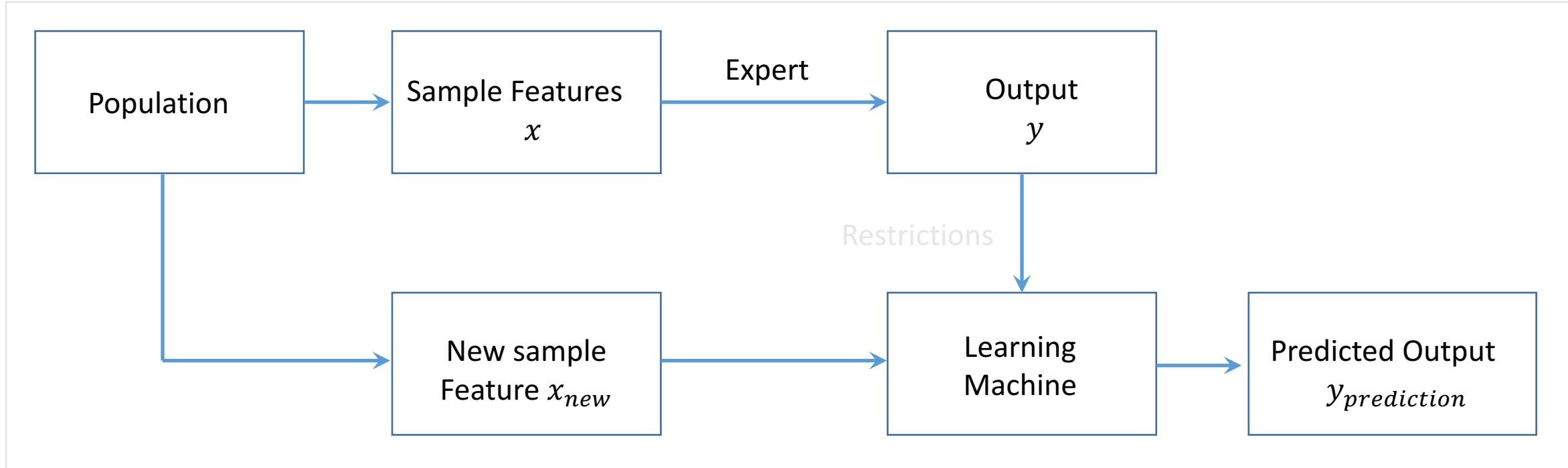
Many relevant applications.

Structure

- Sampling
 - How do we do patient specific simulation?
- Density Estimation
 - How do we estimate likelihoods of various clotting features?
- Classification
 - How do we classify using all variables?
- Significance Selection
 - How do we find important features
- Simplification
 - Is there a simplified model for the dynamics of thrombin generation?

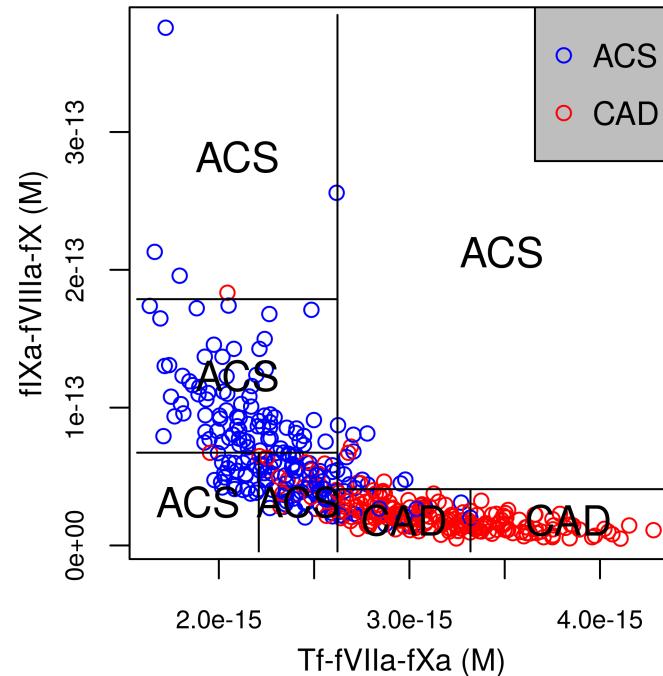
How do we classify data?

Classification Problem



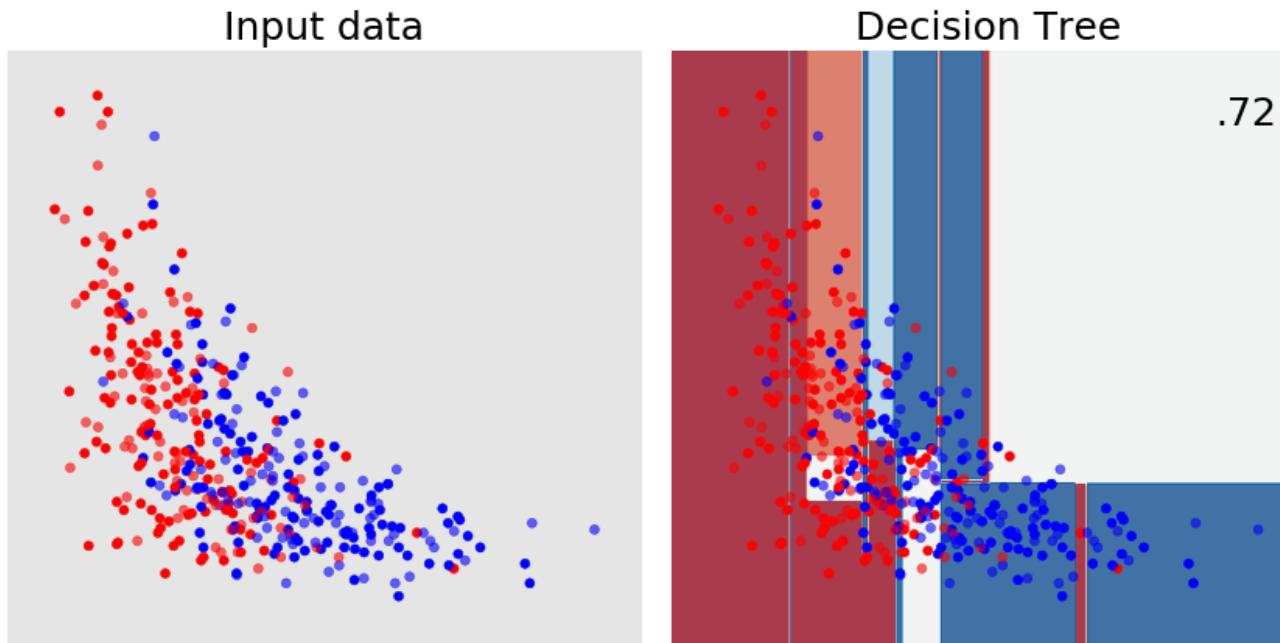
Learning automates decisions. Learning is about sound generalizations.

Decision Tree



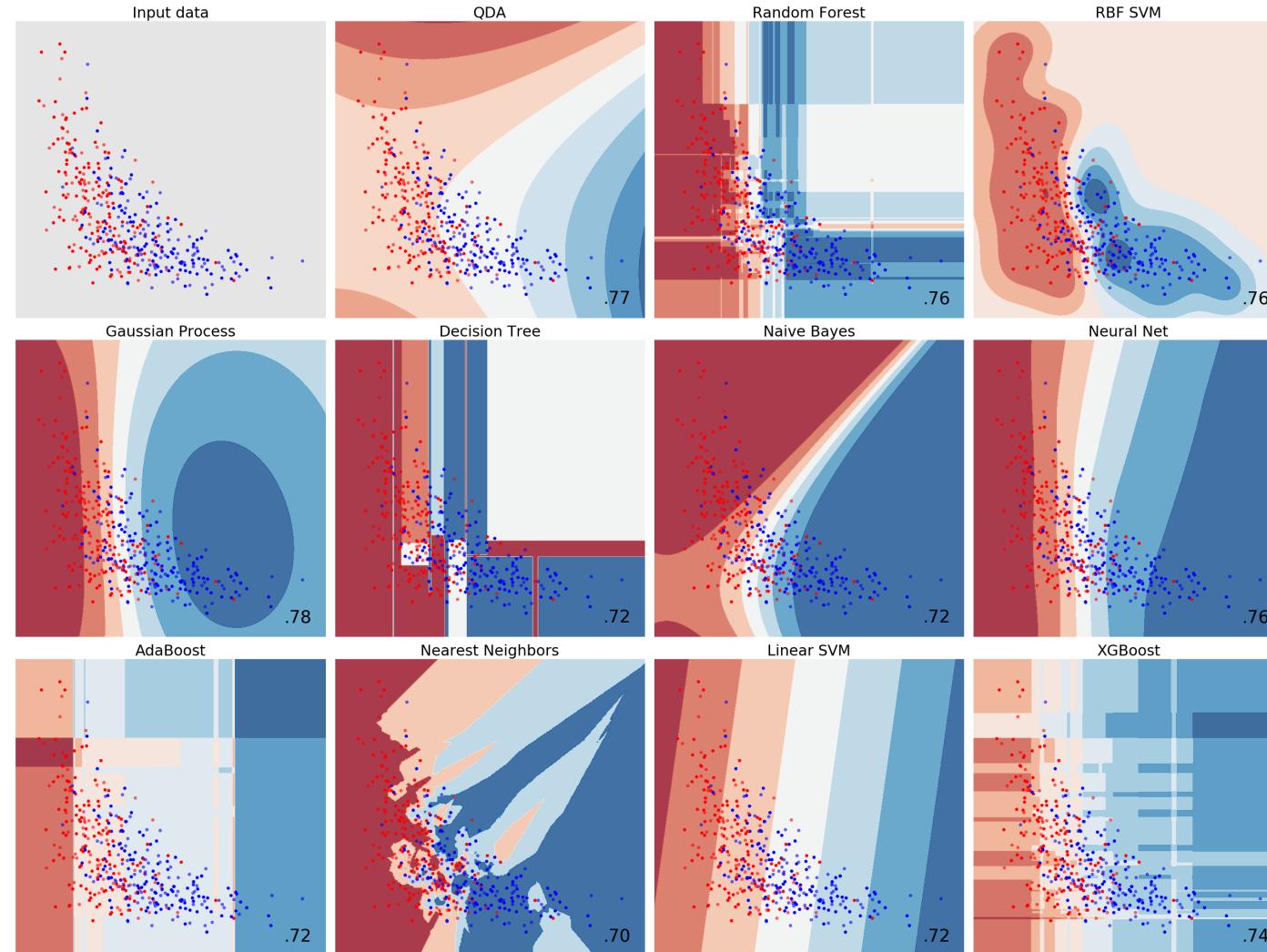
Does a recursive binary partitioning of feature space.

Decision Boundary Color Scheme



Decision Boundaries.

Classification Methods



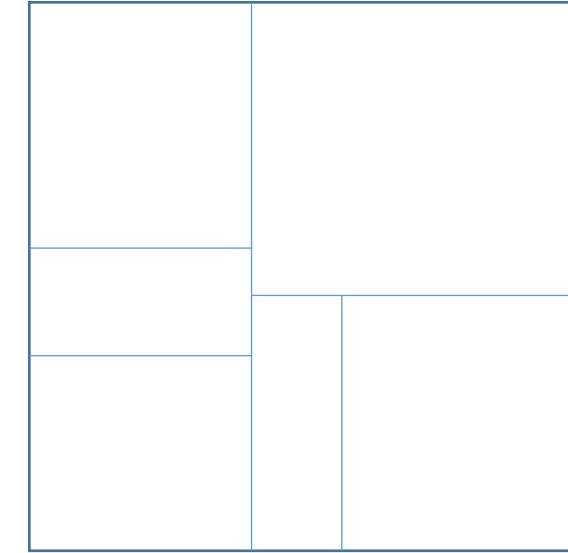
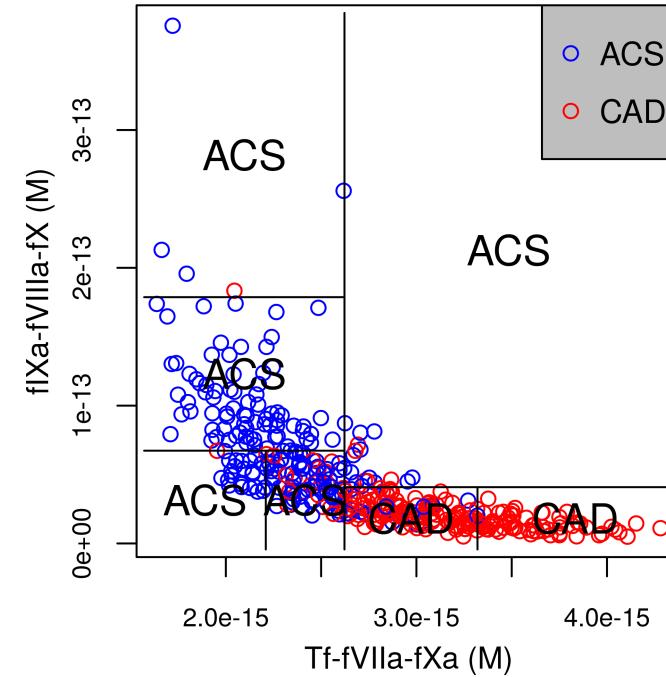
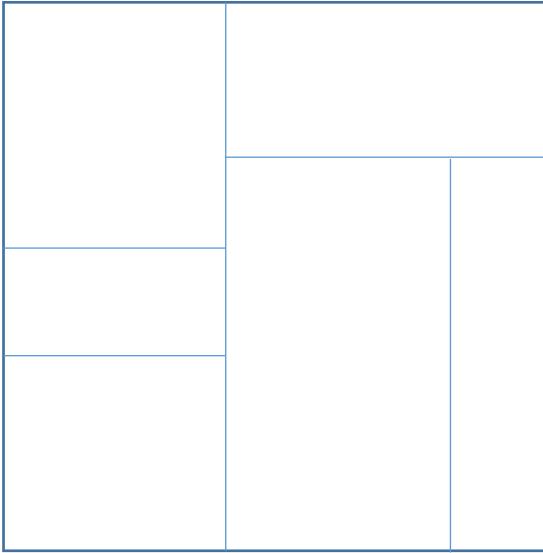
Many possibilities exists.

Method Selection

- Scarce data
- Effective in high-dimensions
- Avoid overfitting
- Select important features
- Less parameter tuning

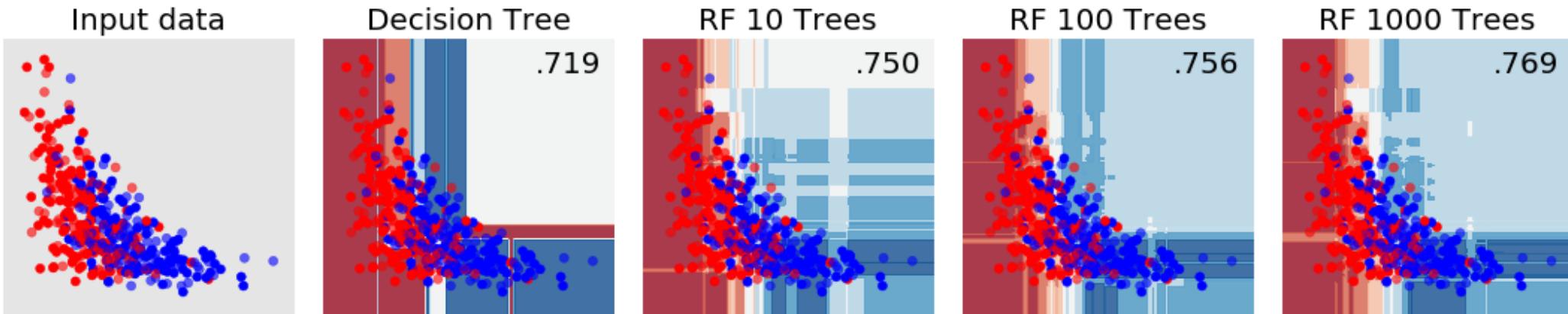
We use Random Forests.

Problem with Decision Trees



Decision trees easily overfit. Random Forests uses ensembles to avoid that.

Random Forests



Random Forests are ensembles of decision trees that do not overfit.

Discussion on PLOS One Paper

(PMID:27171403 PMCID:pmc4865224)

Full Text

Citations

BioEntities

Related Articles

External Links



ONE A Peer-Reviewed, Open Access Journal

[View this Article](#) [Submit to PLOS](#) [Get E-Mail Alerts](#) [Contact Us](#)

PLoS One. 2016; 11(5): e0153776.

PMCID: PMC4865224

Published online 2016 May 12. doi: [10.1371/journal.pone.0153776](https://doi.org/10.1371/journal.pone.0153776)

Random Forests Are Able to Identify Differences in Clotting Dynamics from Kinetic Models of Thrombin Generation

[Jayavel Arumugam](#),^{#1} [Satish T. S. Bukkapatnam](#),^{#2} [Krishna R. Narayanan](#),^{#3} and [Arun R. Srinivasa](#)^{#1,*}

Yoshihiro Fukumoto, Editor

[Author information ▶](#) [Article notes ▶](#) [Copyright and License information ▶](#)

Abstract

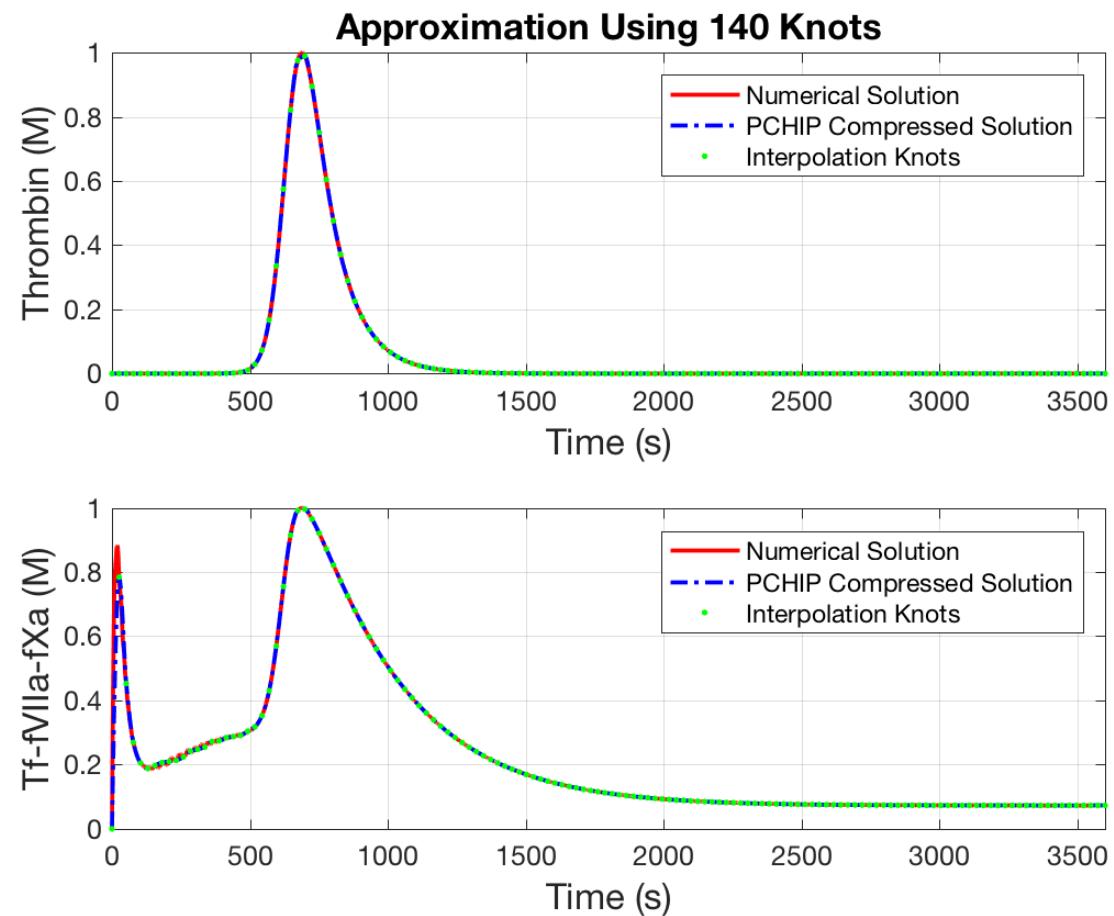
Go to: ▾

Current methods for distinguishing acute coronary syndromes such as heart attack from stable

Will try to highlight interesting aspects.

Feature Extraction

1. PCHIP features (18904)
2. Plasma factor composition (8)
3. Conventional features (11)
4. Concentration moving average (612)



Represent information in all variables as a high-dimensional (~19500) vector.

Hard Question

Random Forest Classifier		Mean (SD)
PCHIP Features		
	All PCHIP Coefficients	88.59 (0.36)
Plasma Factor Composition		
	8 Initial Conditions	88.13 (0.49)
Conventional Features		
	fXa	82.58 (0.53)
	Active Thrombin	81.04 (0.46)
Moving Averages		
	All 200s-MA	88.78 (0.32)

We can use all variables to classify with an accuracy of ~88.7 %. We can screen 50000 patients better.

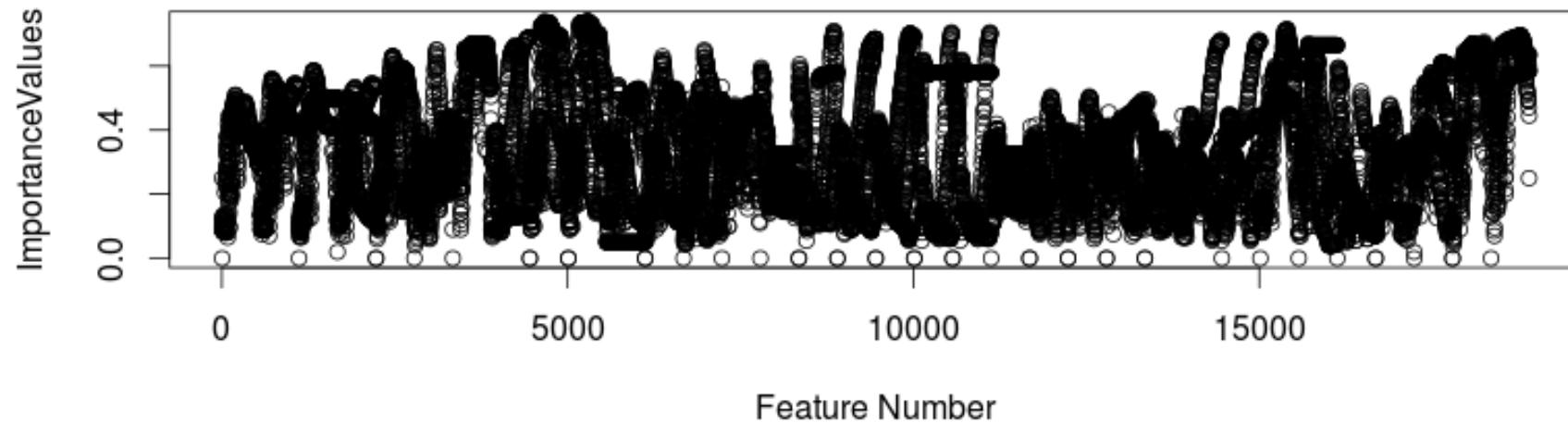
Structure

- Sampling
 - How do we do patient specific simulation?
- Density Estimation
 - How do we estimate likelihoods of various clotting features?
- Classification
 - How do we classify using all variables?
- Significance Selection
 - How do we find important features
- Simplification
 - Is there a simplified model for the dynamics of thrombin generation?

If we have to design an experiment to diagnose better, how would we do?

Difficulty of Finding Significant Features

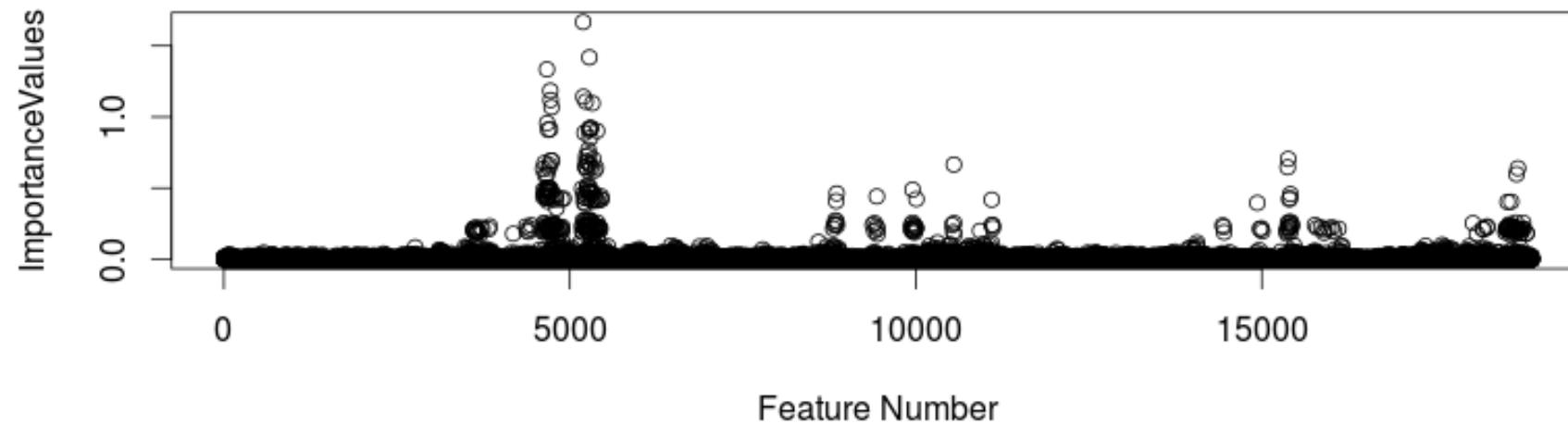
Gini Index Values the spline Coefficients



Say, given 18000 students, we try to rank them.

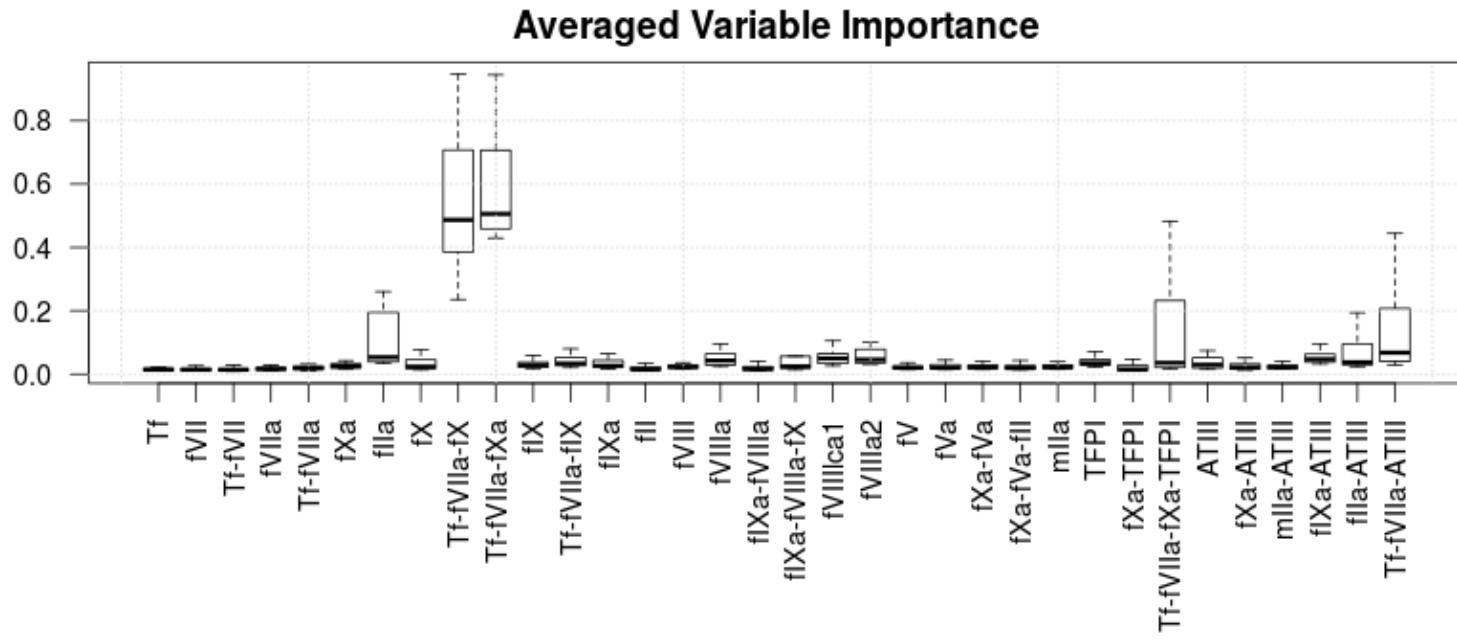
Mean Decrease in Gini Index

Gini Index Values the spline Coefficients



Finds importance by random perturbation of variables and quantifying its effect.

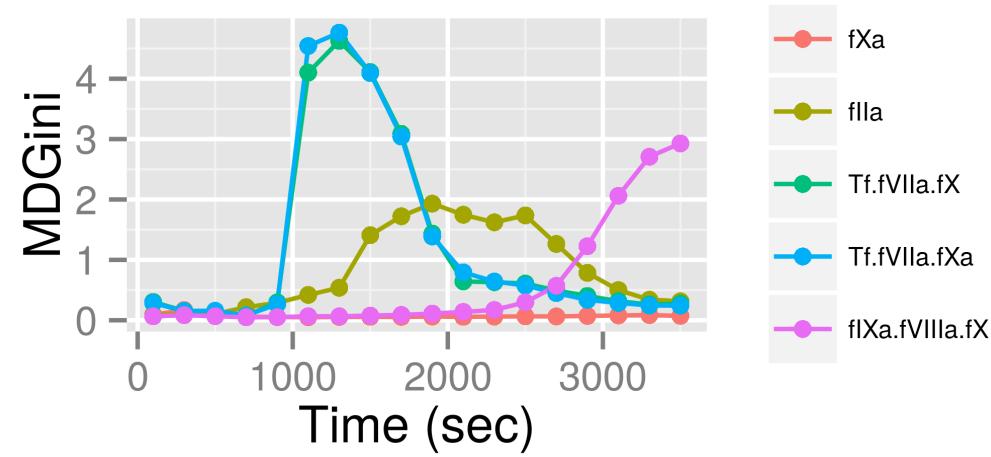
Selection of Important Variables



Greedily select variables.

Probing Importance over Time

MDGini for Selected Variables



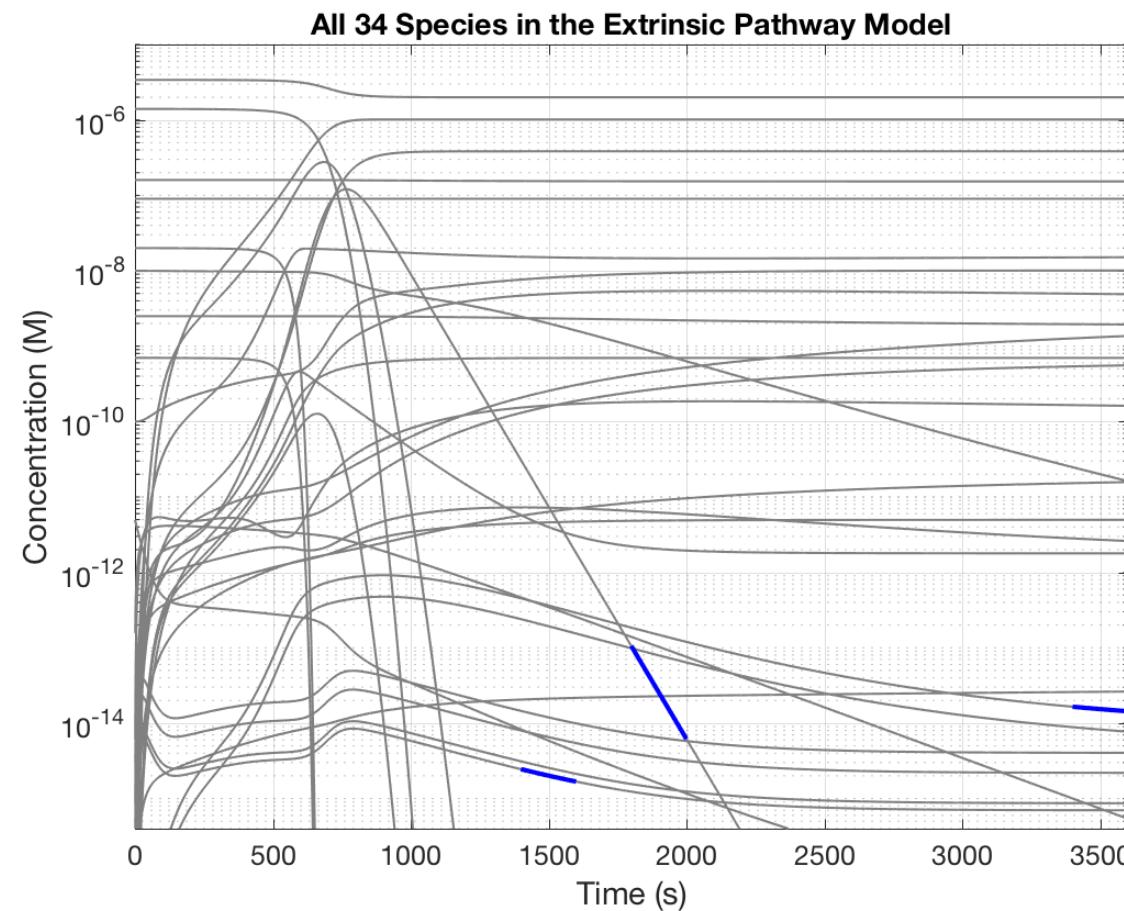
Study significance as a function of time.

Harder Question

Combination	Mean (SD)
Tf-fVIIa-fXa, IIa, and fIXa-fVIIIa-fX	87.16 (0.39)
Tf-fVIIa-fX, IIa, and fIXa-fVIIIa-fX	87.03 (0.40)
Tf-fVIIa-fXa, Tf-fVIIa-fX, and fIXa-fVIIIa-fX	85.58 (0.39)
Tf-fVIIa-fXa, Tf-fVIIa-fX, and IIa	84.90 (0.50)

We could get away by measuring three variables.

Interpretation of results



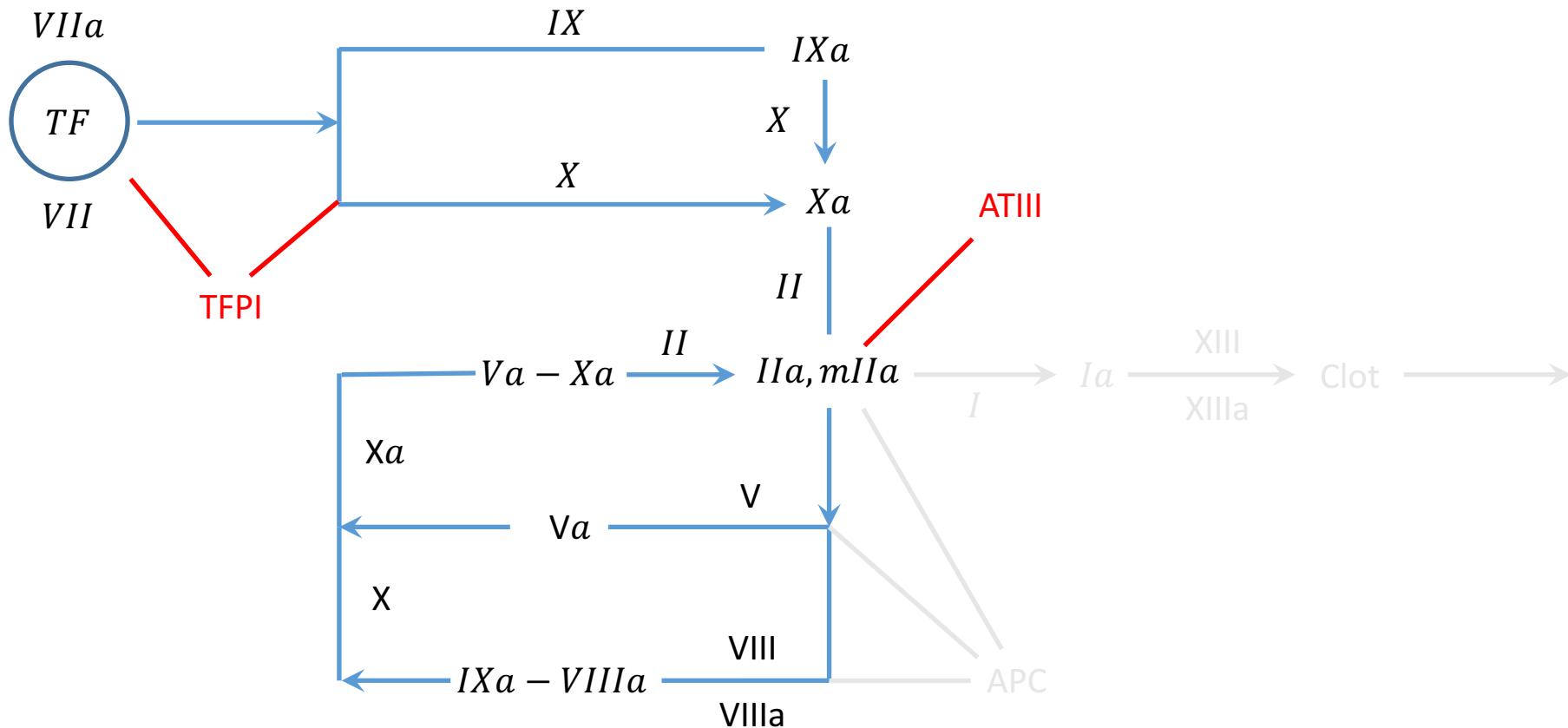
Intelligent experimental sampling. Blue marks has the most classification information.

Structure

- Sampling
 - How do we do patient specific simulation?
- Density Estimation
 - How do we estimate likelihoods of various clotting features?
- Classification
 - How do we classify using all variables?
- Significance Selection
 - How do we find important features
- Simplification
 - Is there a simplified model for the dynamics of thrombin generation?

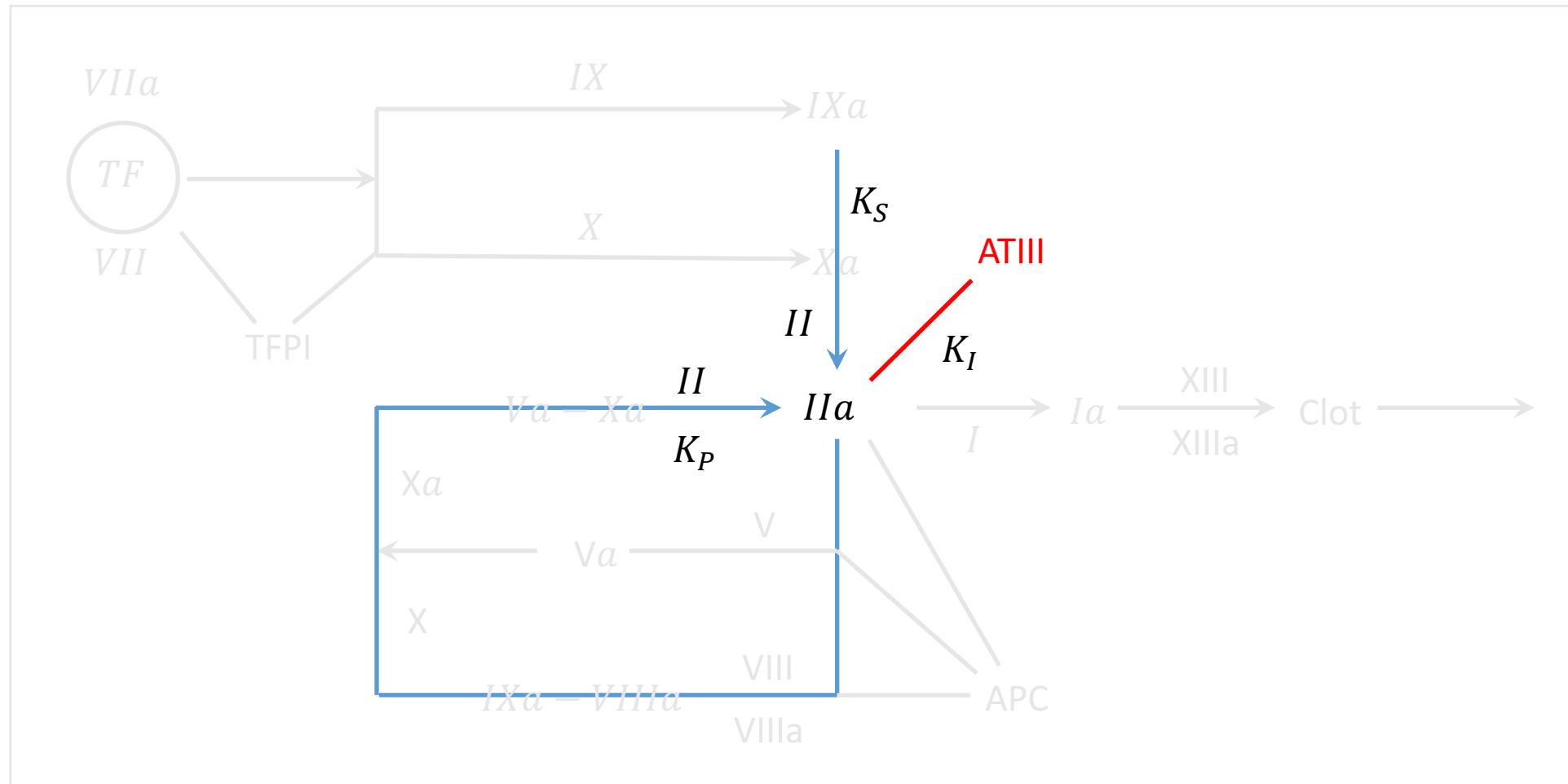
A good simplified model should have predictive capabilities.

Extrinsic Pathway Used



A specific model for thrombin generation (Hocking et al., 2002).

Simplified Model



Use notion of initiation, propagation, and termination.

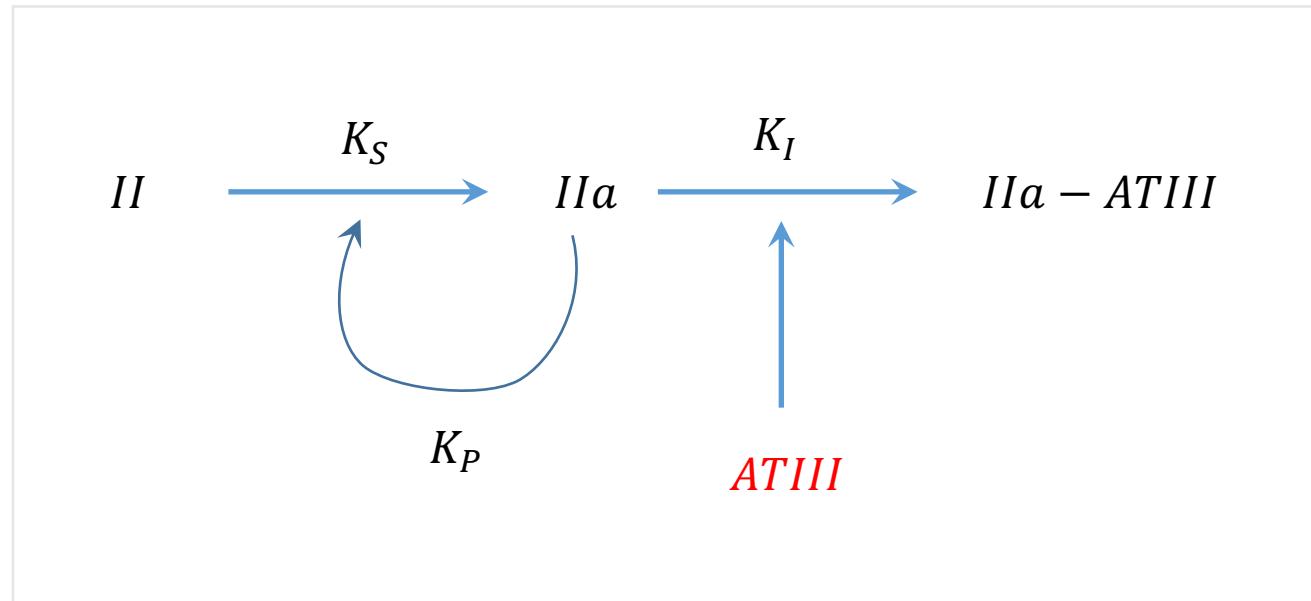
Simplified Model Dynamics

$$\frac{d}{dt}[II] = -K_S - K_P[II][IIa]$$

$$\frac{d}{dt}[IIa] = K_S + K_P[II][IIa] - K_I[IIa][AT]$$

$$\frac{d}{dt}[AT] = -K_I[IIa][IIa]$$

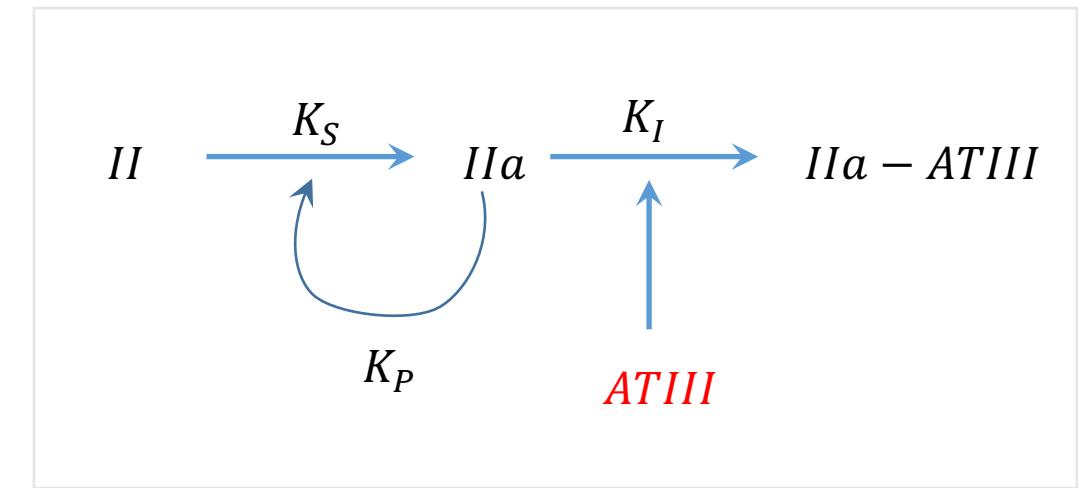
$$\frac{d}{dt}[IIa - AT] = K_I[IIa][AT]$$



Clot initiation is akin to kick-starting.

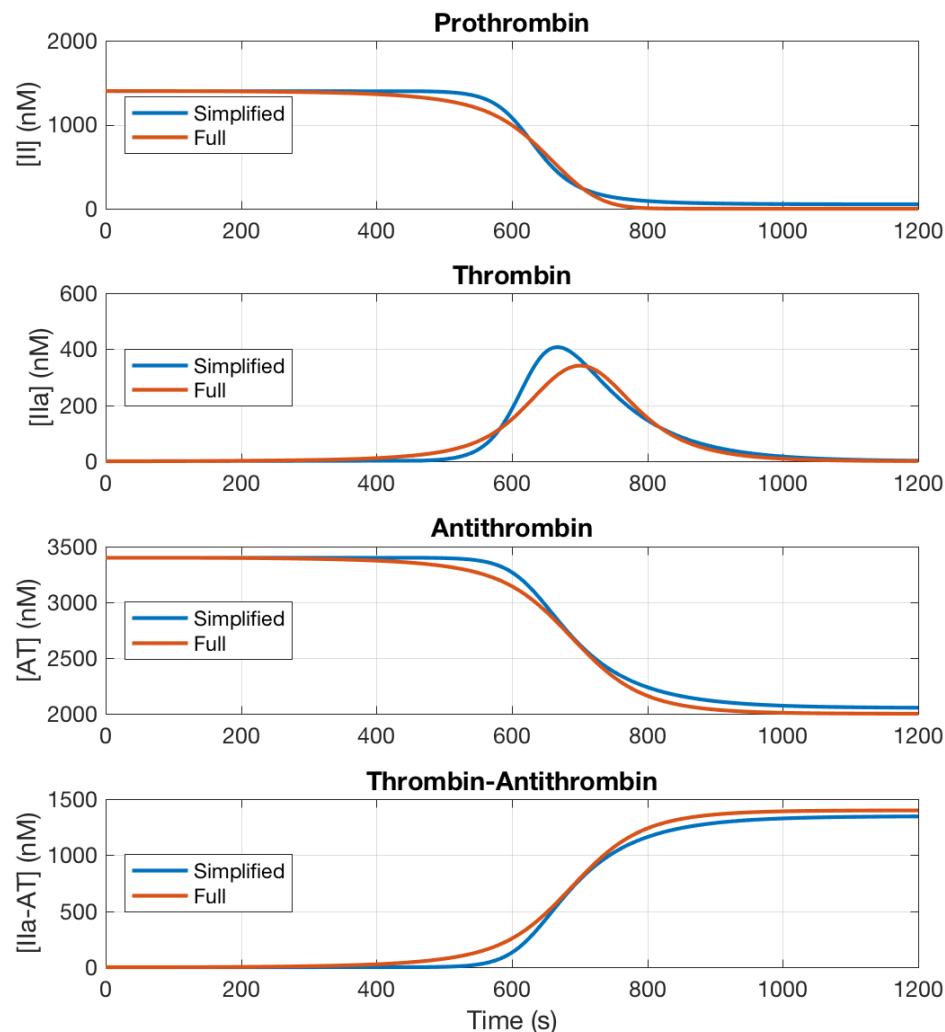
Switching in Simplified Model

	$[IIa] < 2 \text{ nM}$	$[IIa] \geq 2 \text{ nM}$
K_S	$k_s > 0$	0
K_I	$k_{i2} > 0$	$k_{i1} > 0$
K_P	0	$k_p > 0$



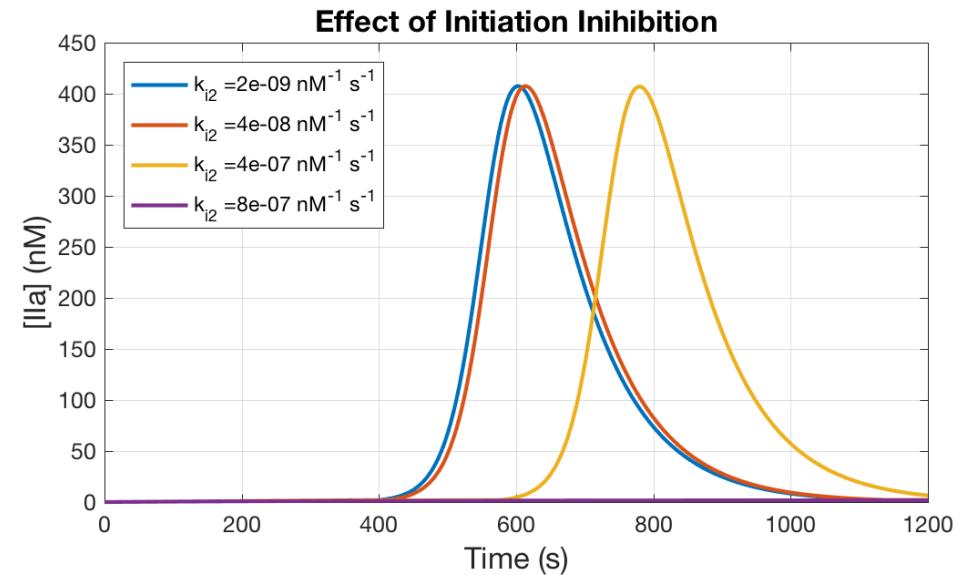
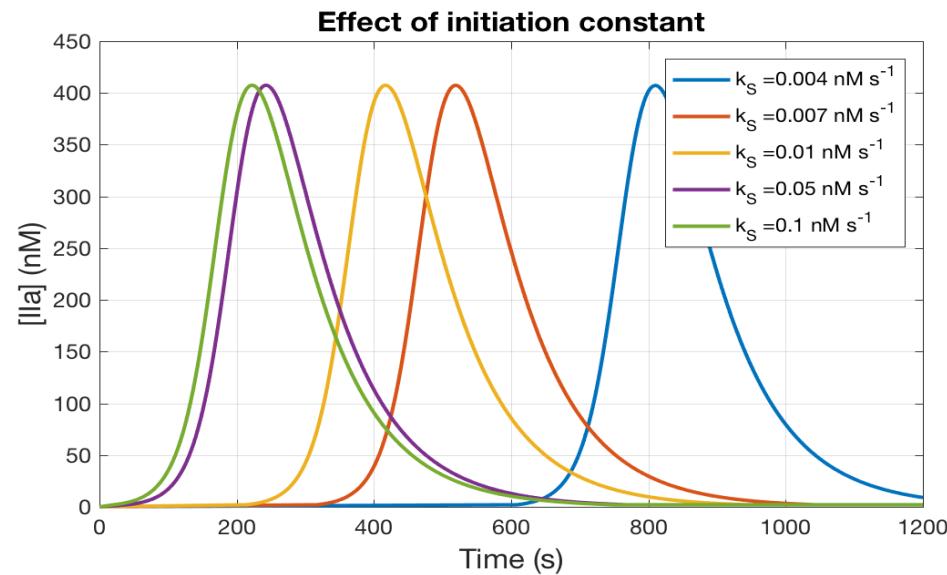
One change: Clot initiation is akin to jump starting.

Parameter Estimation for the Physiological Mean Composition



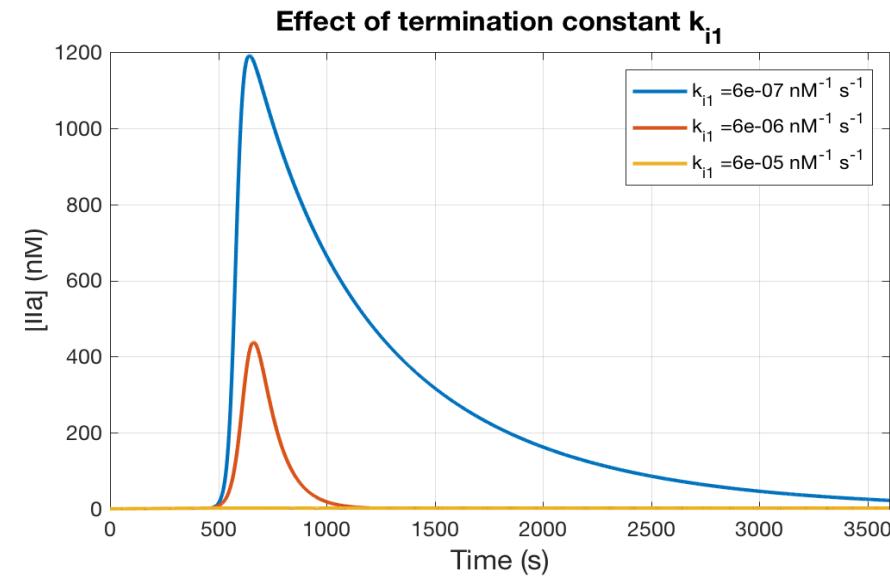
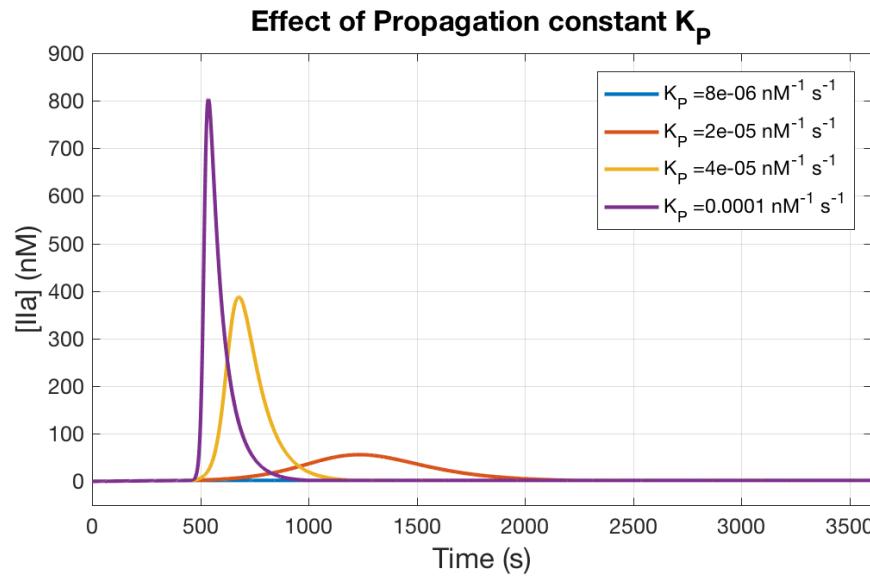
Reasonable fit based on easily measureable variables.

Parameter Study: Initiation Parameters



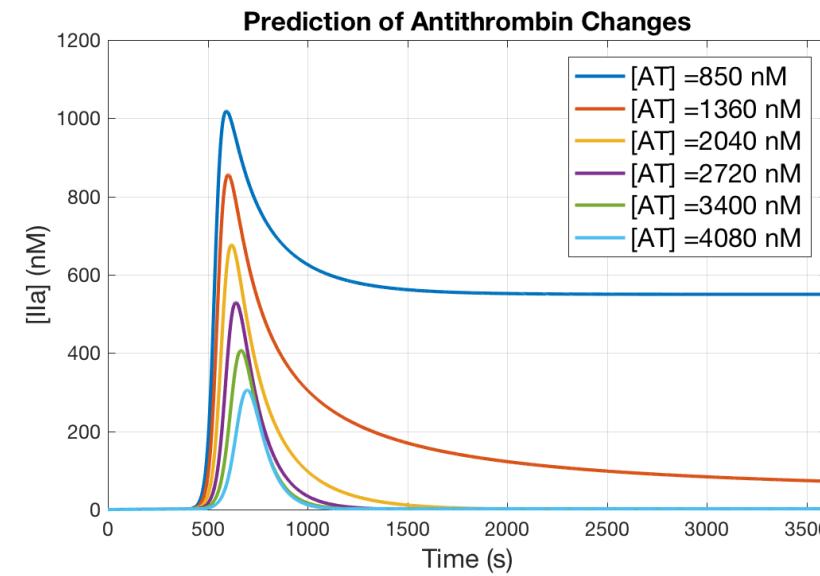
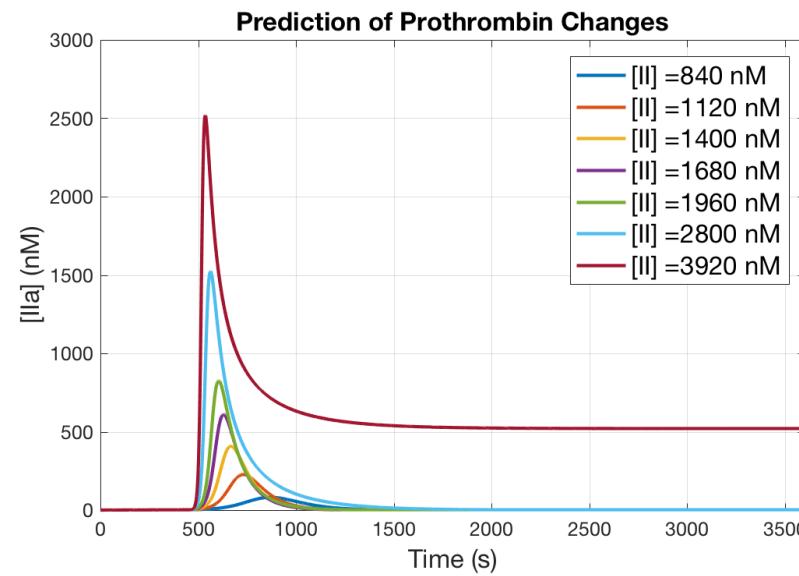
The head and tail are separated. Now we can do magic.

Parameter Study: Propagation and Termination



Further, easy control over response.

Model Prediction



Interesting and most significant prediction.

Experimental Evidence

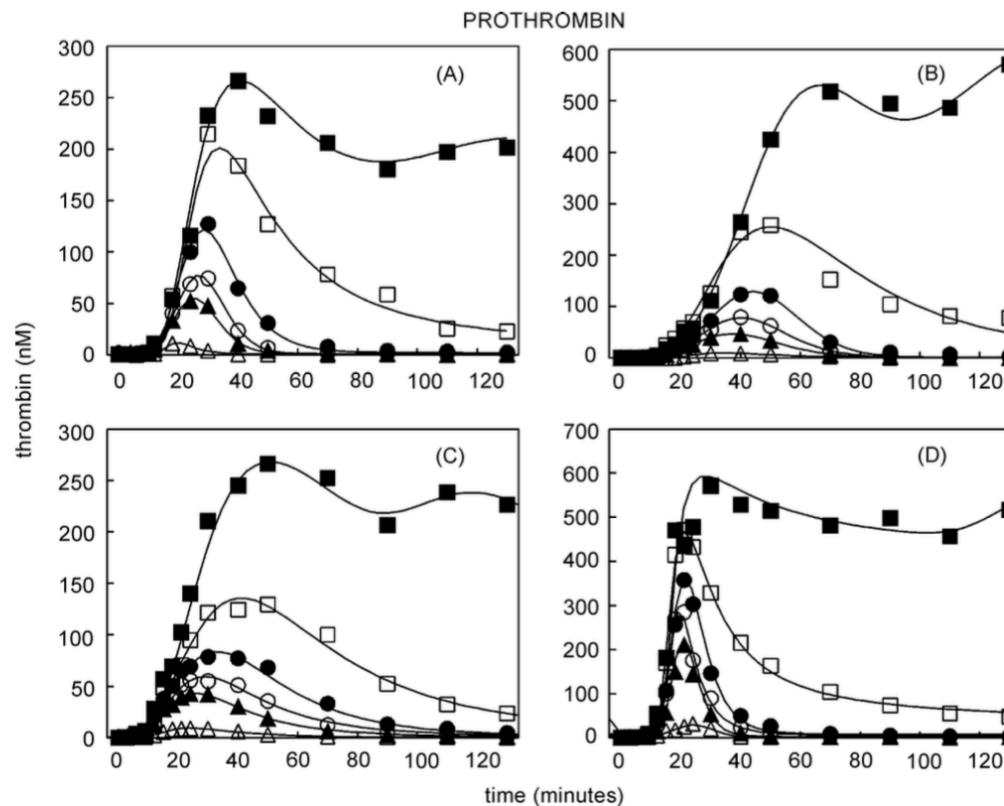
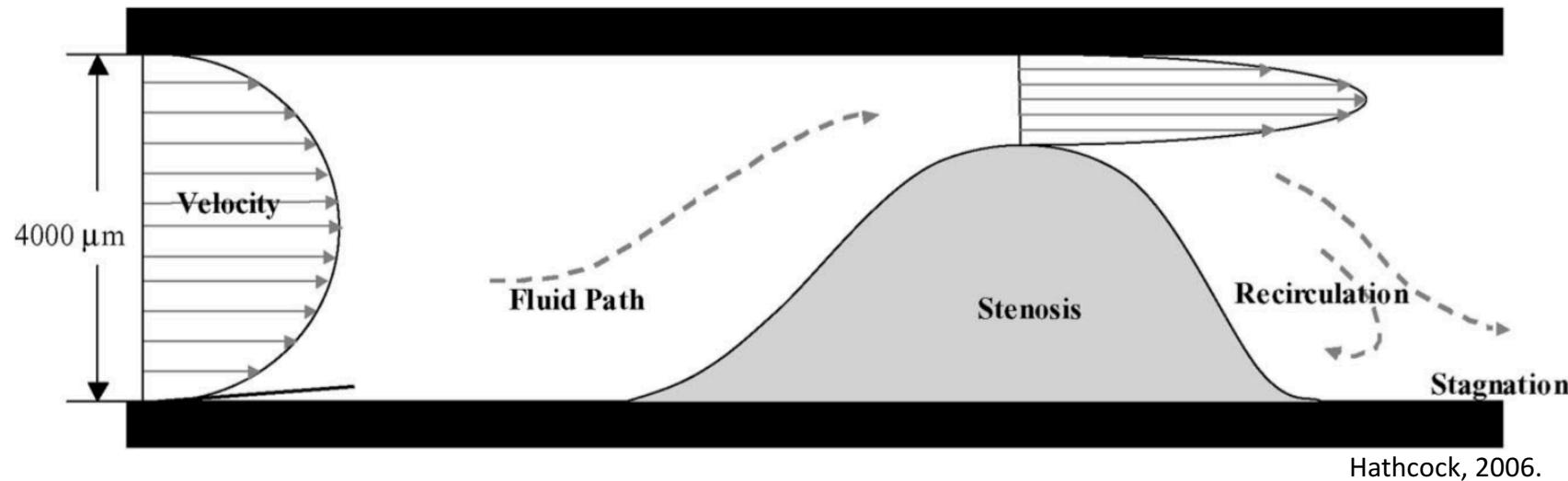


Fig. 9. Thrombin production with varied prothrombin levels. In separate assays using platelets from four healthy, normal subjects (designated A, B, C, D), levels of prothrombin were varied (10%; 50%; 75%; 100%; 150%; 200% of pooled plasma levels) in a model system of coagulation with otherwise normal plasma levels of factors V, VIII, IX, X and XI, antithrombin and tissue factor pathway inhibitor, catalytic amounts of factor VIIa, tissue factor-bearing monocytes and unactivated platelets. Timed samples were taken for determination of thrombin generation. $\Delta = 10\%$, $\blacktriangle = 50\%$, $\circ = 75\%$, $\bullet = 100\%$, $\square = 150\%$, $\blacksquare = 200\%$.

Allen et al., 2003

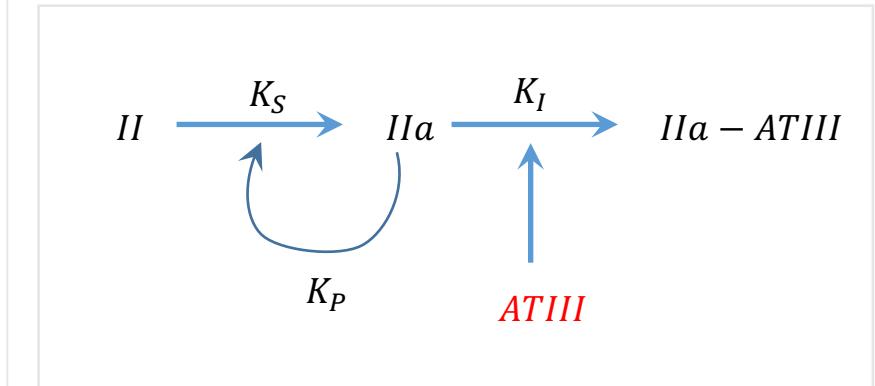
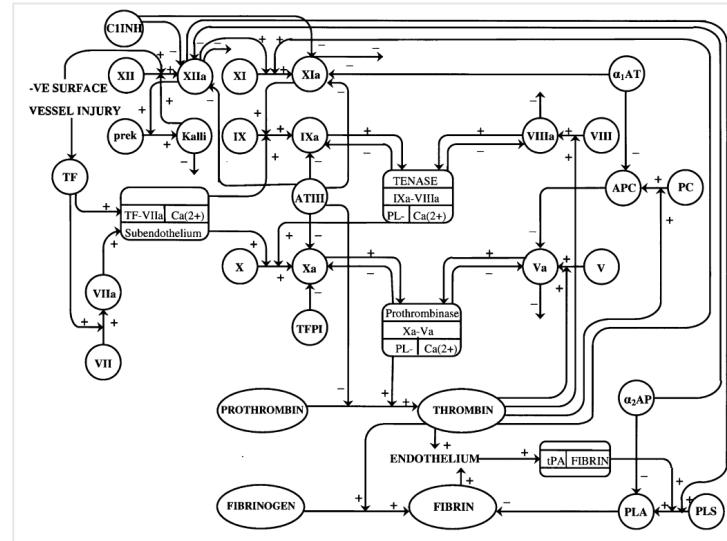
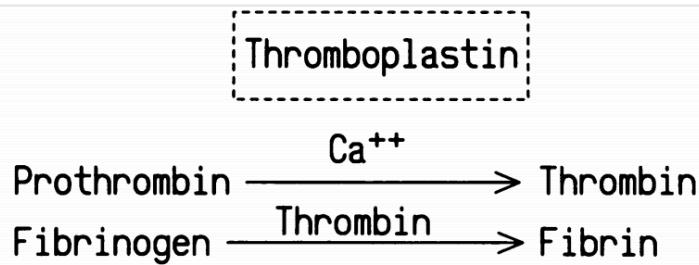
Thrombin runs out of inhibitor.

Implications



New supply of inhibitors is based on transport.

To Summarize

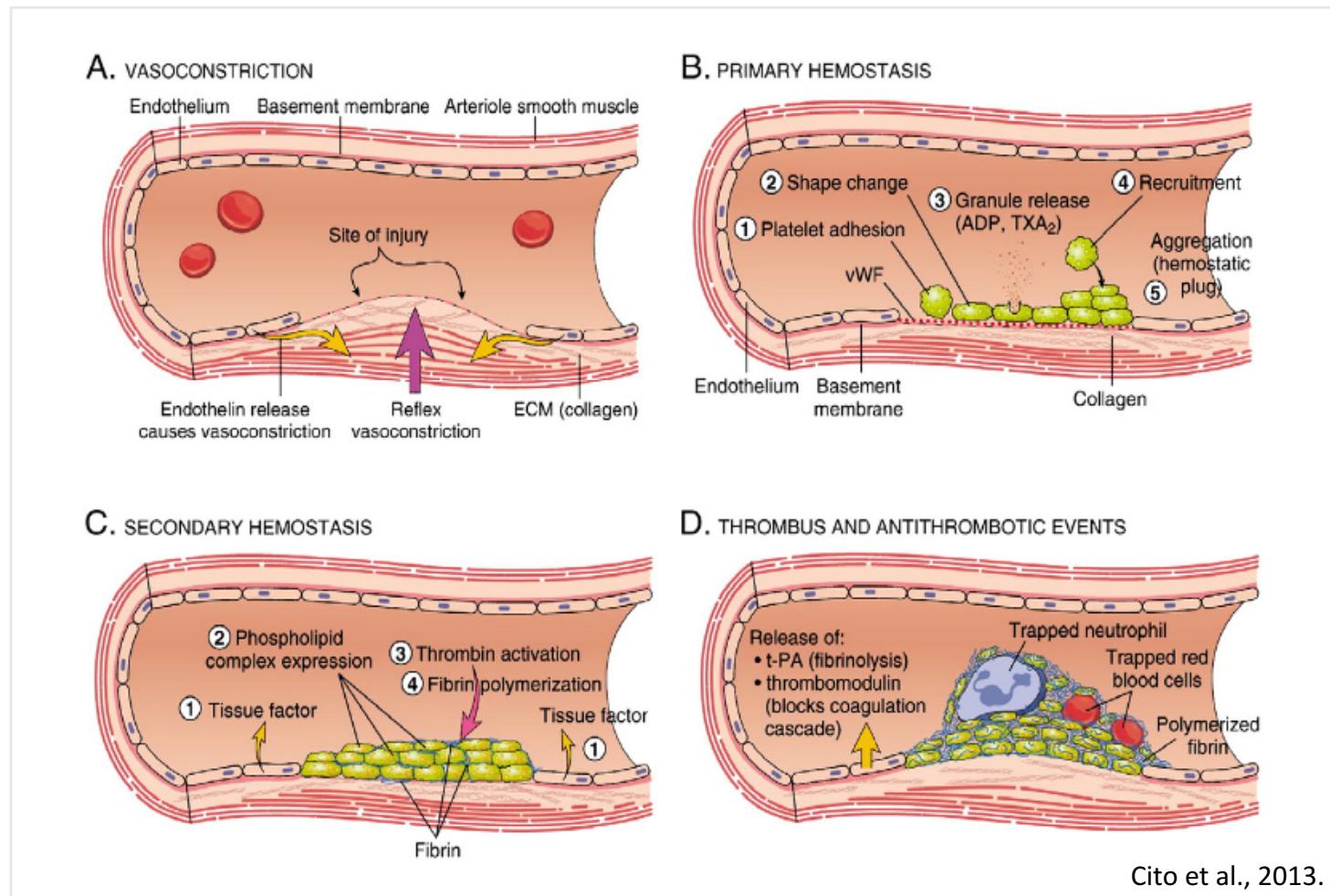


We have drastically simplified a part of the cascade while identifying and retaining essential functionality.

Summary of Work

- How do we do patient specific simulation? MaxEnt
 - ✓ Sample using log-normal for positive random variables.
- How do we estimate likelihoods of various clotting features? EMGM
 - ✓ Expectation propagation of Gaussian Mixture Models. Using 2 features, we could classify with accuracies close to 77 %.
- How do we classify using all variables? Random Forests
 - ✓ Using information from all variables could classify with an accuracy of 88.7 %.
- Significance Selection? Mean Decrease in Gini Index
 - ✓ 3 proteins at specific times could classify with an accuracy of 87 %.
- Simplification? Switching model
 - ✓ Model predicts sustained thrombin activity in certain plasma factor compositions.

Closer Look at Clotting



Clotting is a complicated phenomena traditionally modeled in 4 stages.

Future Directions

- Study and account for the effect of flow
- Augment with data-driven models
- Platelet aggregation model
- Study more mechanically related risk factors

There is so much to do ...

Thank you all



And so little time we have.

Back Slides