Modeling & Simulation (Computational Immunology)

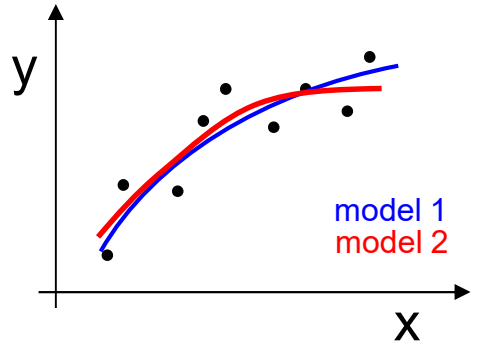
Steven H. Kleinstein

Departments of Pathology and Immunobiology
Yale School of Medicine

steven.kleinstein@yale.edu

March 30, 2020

Comparing Two Model Fits



- The number of data points, *N*, must exceed the number of model parameters, *M*, yielding the degrees of freedom (*DOF* = *N*-*M*)
- Increasing *M* using a more complex model will generally improve the quality of fit and reduce RSS

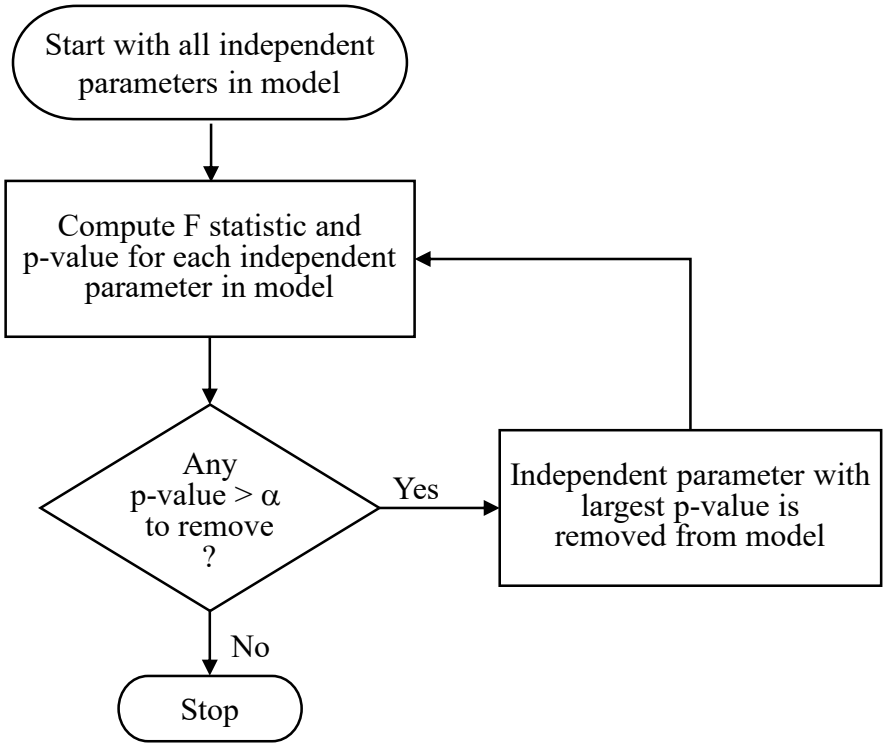
- An F-statistic can be computed to compare the results of two model fits
  - F ~ 1, the simpler model is adequate
  - F > 1, the more complex model is better, or random error led to a better fit with the complex model
  - P-value defines the probability of such a "false positive" result (lookup in F table)

(Costa, Kleinstein and Hershberg, Sci Signal. 2011)

Building models with variable selection

F statistic determines if variable added or deleted from model

Backward Elimination



Other Variations:

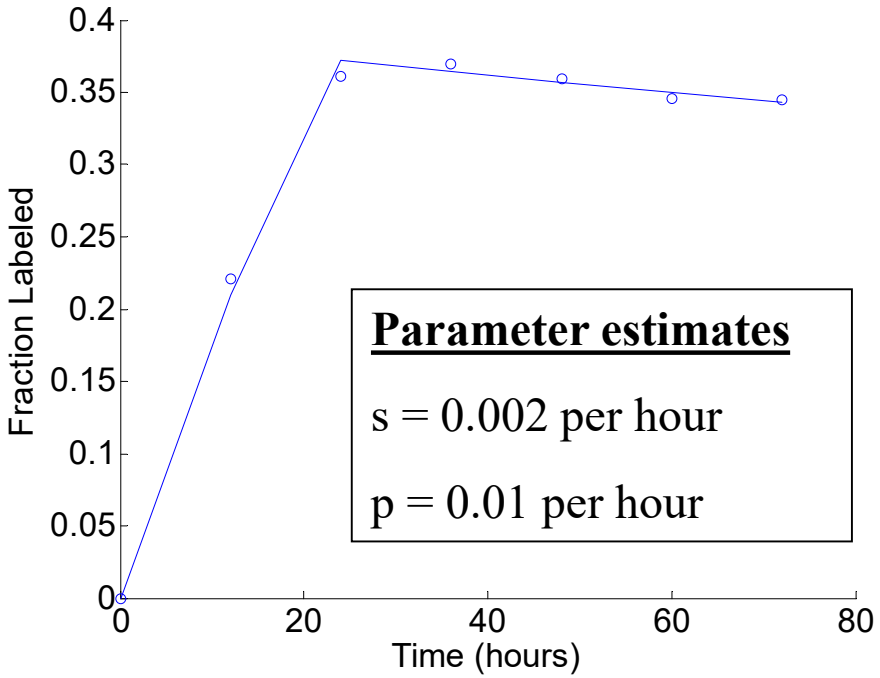
Forward selection: adds variables one at a time as long as significant F test.

Stepwise procedure: allows for removal of a parameter at each step

No guarantee that globally optimal model with be found (need all subsets, but prohibitive for large parameter space)

How much confidence to put in estimate?

Construct confidence intervals for model parameters

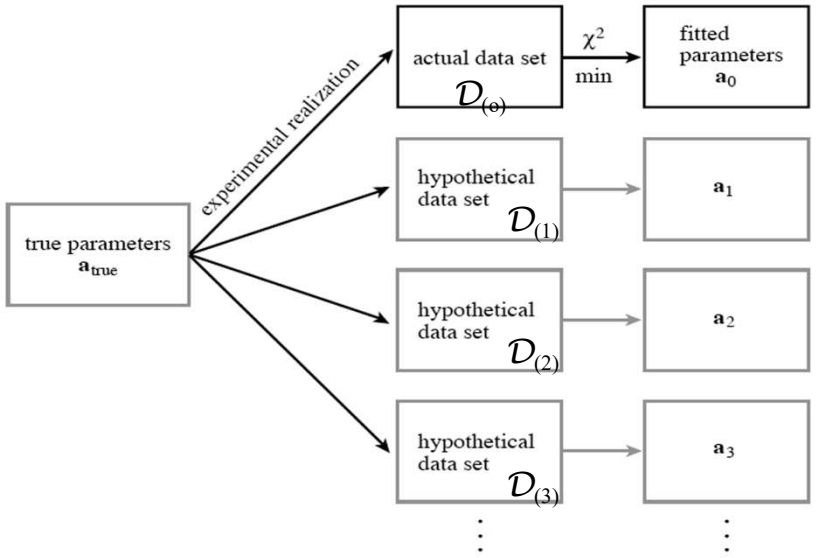


Estimate uncertainty given limited number of experimental observations

Accuracy of Estimated Model Parameters

Underlying true set of model parameters ($\mathbf{a}_{true}$) known to Mother Nature but hidden from the experimenter

- True parameters are statistically realized as measured data set $\mathcal{D}_{(o)}$

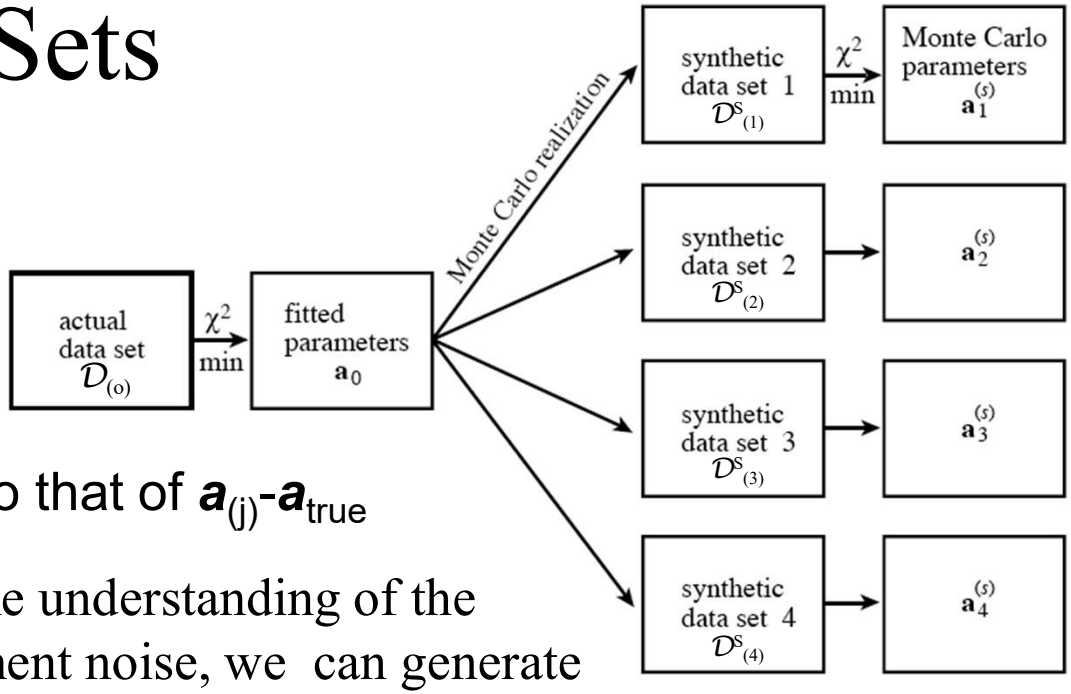


from Numerical Recipes online

- Fitting $\mathcal{D}_{(o)}$ yields estimated model parameters $\mathbf{a}_{(o)}$
- Other experiments could have resulted in data sets $\mathcal{D}_{(1)}$, $\mathcal{D}_{(2)}$, etc. which would have yielded model parameters $\mathbf{a}_{(1)}$, $\mathbf{a}_{(2)}$, etc.

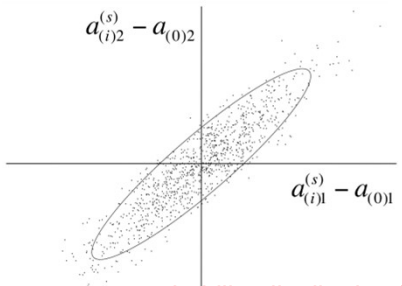
Estimate probability distribution of $\mathbf{a}_{(i)}$ - $\mathbf{a}_{true}$ without knowing $\mathbf{a}_{true}$

Monte Carlo Simulation of Synthetic Data Sets



from Numerical Recipes online

- Assume that if $\boldsymbol{a}_{(0)}$ is a reasonable estimate of $\boldsymbol{a}_{\text{true}}$, then the distribution of $\boldsymbol{a}_{(j)}$-$\boldsymbol{a}_{(0)}$ should be similar to that of $\boldsymbol{a}_{(j)}$-$\boldsymbol{a}_{\text{true}}$
- With the assumed $\boldsymbol{a}_{(0)}$, and some understanding of the characteristics of the measurement noise, we can generate "synthetic data sets" $\mathcal{D}^S_{(1)}$, $\mathcal{D}^S_{(2)}$,… at the same x_i values as the actual data set, $\mathcal{D}_{(0)}$, that have the same relationship to $\boldsymbol{a}_{(0)}$ as $\mathcal{D}_{(0)}$ has to $\boldsymbol{a}_{\text{true}}$
- For each $\mathcal{D}^S_{(j)}$, perform a model fit to obtain corresponding $\boldsymbol{a}^S_{(j)}$, yielding one point $\boldsymbol{a}^S_{(j)}$- $\boldsymbol{a}_{(0)}$ for simulating the desired M-dimensional probability distribution. **This is a very powerful technique!!**



2-parameter probability distribution for 1,000 Monte Carlo simulations

(Costa, Kleinstein and Hershberg, Sci Signal. 2011)

The Bootstrap Method

Estimating generalization error based on "resampling":
Randomly draw datasets with replacement from training data

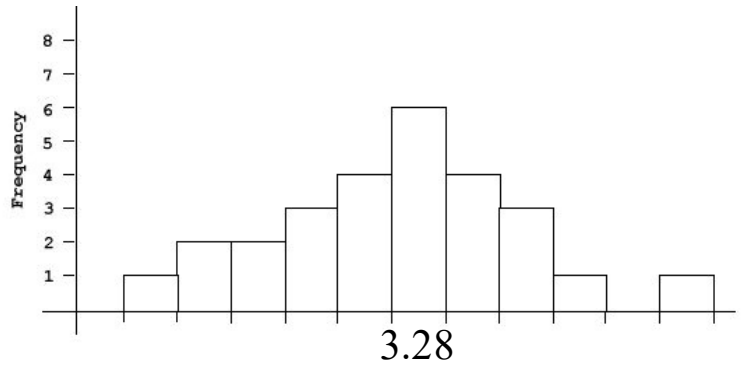
- If don't know enough about the measurement errors (i.e. cannot even say they are normally distributed) so Monte Carlo simulation cannot be used.
- Bootstrap Method uses actual data set $\mathcal{D}_{(o)}$, with its N data points, to generate synthetic data sets $\mathcal{D}^{S}_{(1)}$, $\mathcal{D}^{S}_{(2)}$,… also with N data points.
- Randomly select N data points from $\mathcal{D}_{(o)}$ *with replacement*, which makes $\mathcal{D}^{S}_{(j)}$ differ from $\mathcal{D}_{(o)}$ with a fraction of the original points replaced by *duplicated* original points.
- Fitting the $\mathcal{D}^{S}_{(j)}$ data yields model parameter sets $\mathbf{a}^{S}_{(j)}$ using actual measurement noise.

If sample is good approximation of population, bootstrap method will provide good approximation of sampling distribution of original statistic.

Bootstrap Methods

Randomly draw datasets with replacement from training data

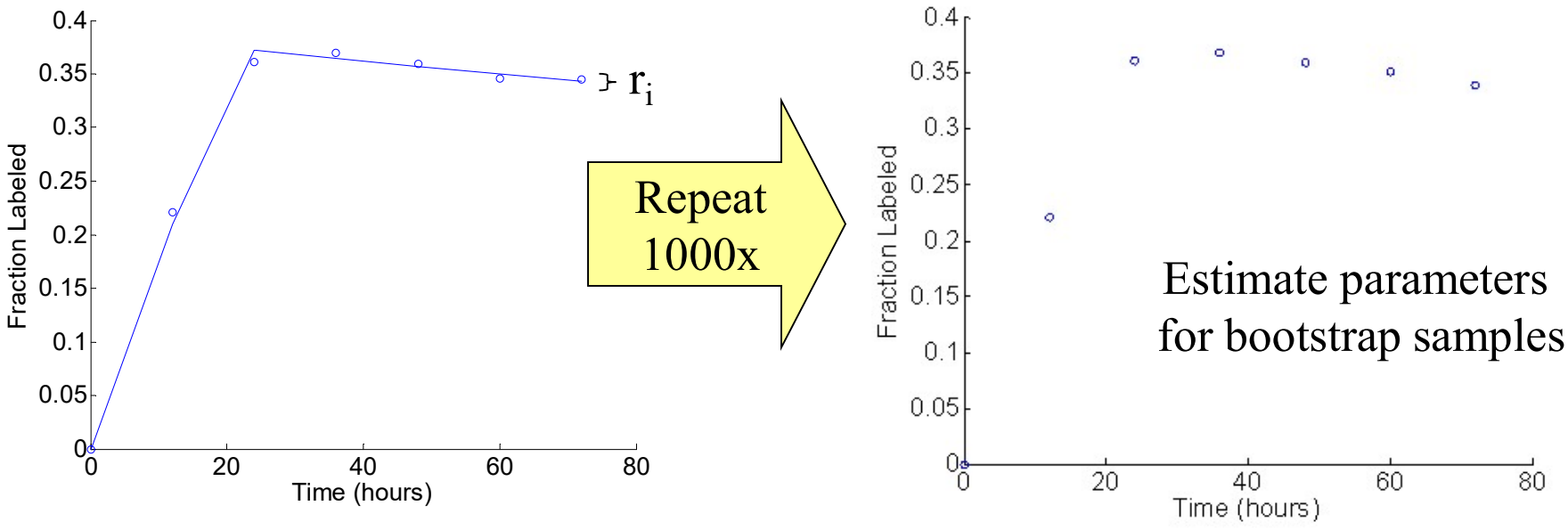
- D = [3.0, 2.8, 3.7, 3.4, 3.5] → average = 3.28
- Bootstrap samples D_N could be:
  - [2.8, 3.4, 3.7, 3.4, 3.5] → 3.36
  - [3.5, 3.0, 3.4, 2.8, 3.7] → 3.28
  - [3.5, 3.5, 3.4, 3.0, 2.8] → 3.24
  - ...



If sample is good approximation of population, bootstrap method will provide good approximation of sampling distribution of original statistic.

Bootstrapping Parameter Confidence Intervals

1) Fit model to data to obtain parameter estimates
2) Draw a bootstrap sample of the residuals (Fixed-X Bootstrapping)
3) Create bootstrap sample of observations by adding randomly sampled residual to predicted value of each observation



Bootstrapping observations also possible – asymptotically equivalent

Bootstrapping Parameter Confidence Intervals

Three commonly used methods: 1. Normal Theory Intervals, 2. Percentile Intervals, 3. Bias Corrected Percentile Intervals

Percentile Intervals

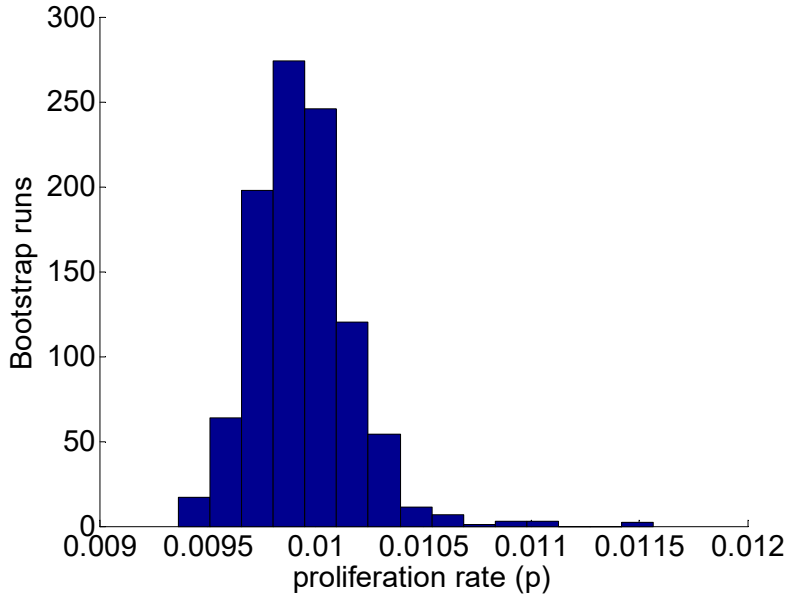
Calculate the parameter for each bootstrap sample and select α (e.g., 0.05)

LCL = α /2^{th} percentile.

UCL = $(1-\alpha/2)^{th}$ percentile.

Use MATLAB's prctile function:
= prctile(bootstrap estimates, 0.025)

Contains 95% of the estimates



Parameter estimates for synthetic data

Estimate of s = 0.0017 [0.0009,0.0030]
Estimate of p = 0.0099 [0.0095,0.0100]

May not have correct coverage when sampling distribution skewed

Practical reference for these kinds of methods

Numerical Recipes:
Includes source code for integration, optimization, etc.

TEACHING RESOURCE

COMPUTATIONAL BIOLOGY

Biomedical Model Fitting and Error Analysis

Kevin D. Costa,[1,*] Steven H. Kleinstein,[2,3] Uri Hershberg[4]

www.SCIENCESIGNALING.org 27 September 2011 Vol 4 Issue 192

Free NR versions online at http://www.nr.com/oldverswitcher.html

Hepatitis C Viral Dynamics and Interferon-α Therapy

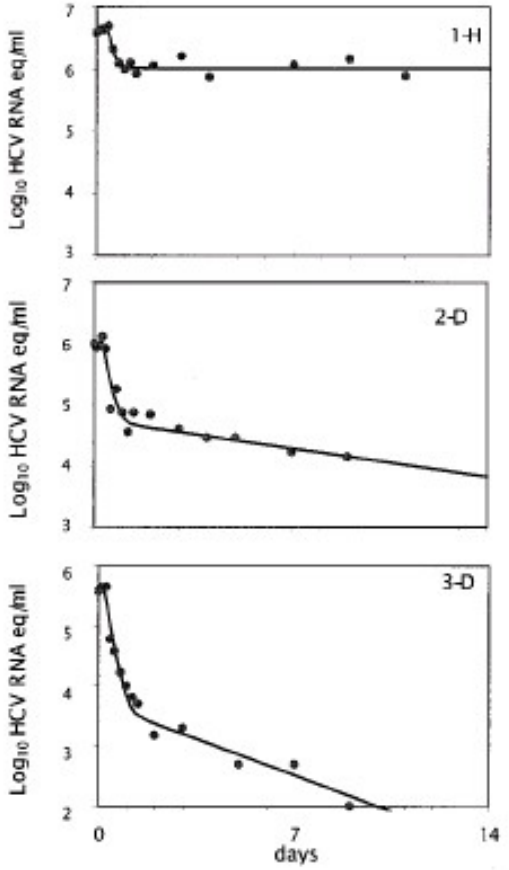
Modeling 23 patients during 14 days of therapy (daily doses)

Hepatitis C Viral Dynamics in Vivo and the Antiviral Efficacy of Interferon-α Therapy

Avidan U. Neumann,*† Nancy P. Lam,*‡ Harel Dahari, David R. Gretch, Thelma E. Wiley, Thomas J. Layden, Alan S. Perelson

SCIENCE VOL 282 2 OCTOBER 1998

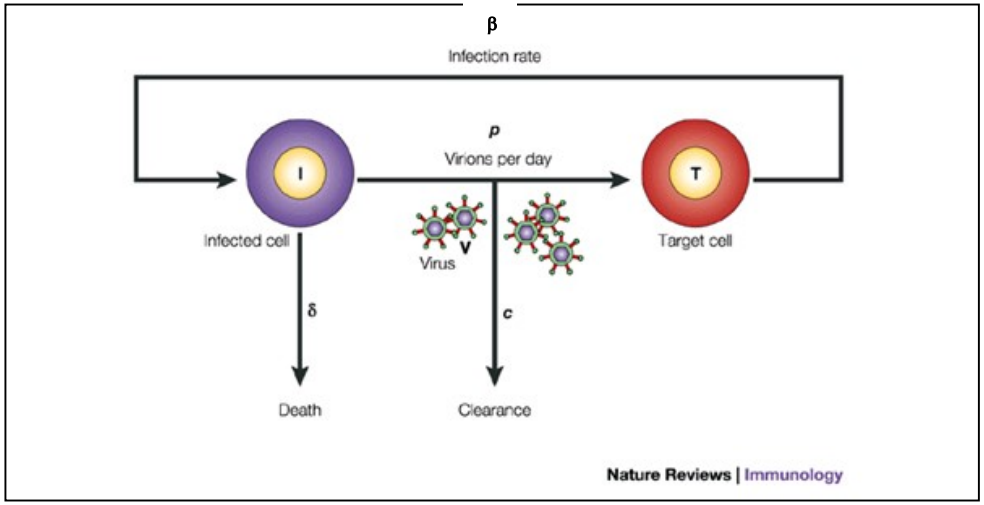
Viral loads exhibit short delay followed by biphasic decline in viral load



How does interferon therapy work?

Model of Hepatitis C Viral Dynamics

Includes virus along with target (T) and infected (I) cells



Target Cells $$dT/dt = s - dT \boxed{\ ?\ } \quad (1)$$

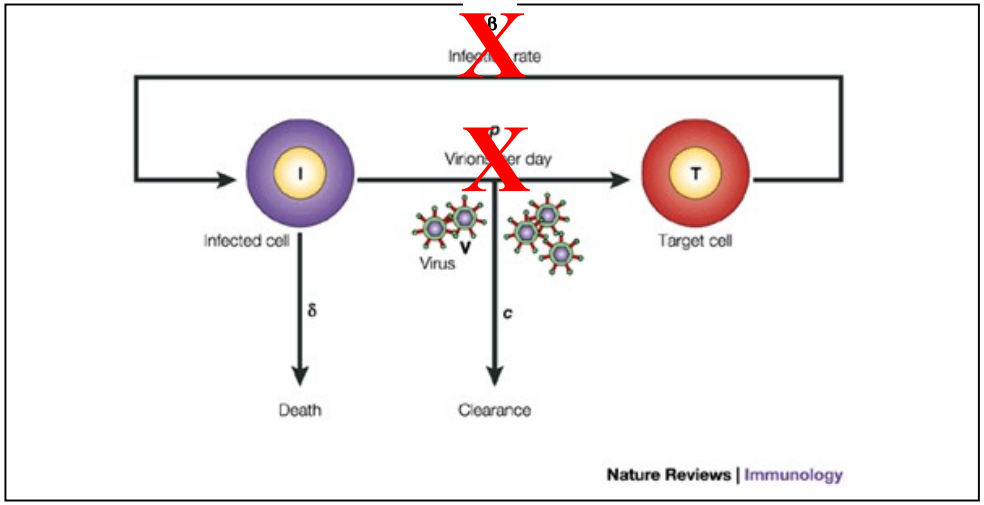
Infected Cells $$dI/dt = \boxed{\ ?\ } - \delta I \quad (2)$$

Virus (HCV RNA) $$dV/dt = pI - cV \quad (3)$$

Before therapy, virus load is approximately constant

Model of Interferon-α Therapy

Includes virus along with target (T) and infected (I) cells



Target Cells $$dT/dt = s - dT - \beta VT \quad (1)$$

Infected Cells $$dI/dt = \beta VT - \delta I \quad (2)$$

Virus (HCV RNA) $$dV/dt = pI - cV \quad (3)$$

Therapy can reduce the rate of infection, or production of virions

Hepatitis C Viral Dynamics and Interferon-α Therapy

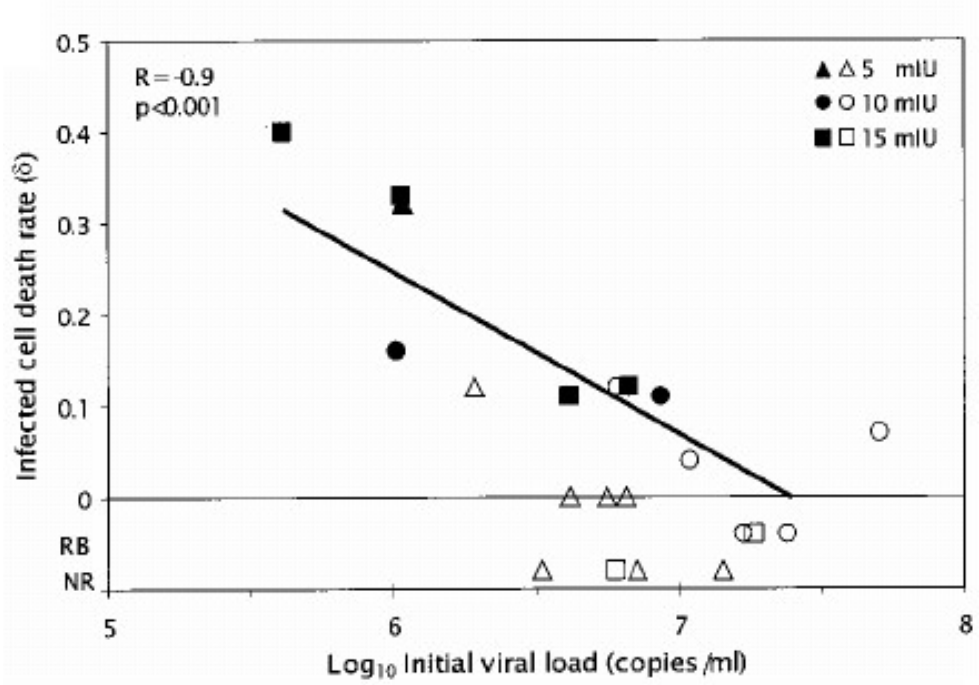
Modeling 23 patients during 14 days of therapy (daily doses)

Regimen	Patient	Initial VL (10^6 copies per milliliter)	Delay (hours)	Virion clearance (c) (1/day)	± error	Efficacy (ε) Percent	± error	Infected cell death (δ) (1/day)	± error	Production (10^9 copies per day)
1	A	5.6	8	5.9	1.1	79	4.0%	0	0.01	495
1	B	1.9	8	6.4	1.8	75	7.0%	0.12	0.02	290
1	C	14.2	NR	NR		NR		NR		NR
1	D	7.1	NR	NR		NR		NR		NR
1	E	1.1	11	7.0	0.6	86	0.1%	0.32	0.04	125
1	F	6.5	7	5.0	0.8	89	8.0%	0	0.01	601
1	G	3.3	NR	NR		NR		NR		NR
1	H	4.1	10	6.9	0.2	75	1.0%	0	0.01	498
1: Mean	±SD	5.5 ± 4.1	9 ± 1.5	6.2 ± 0.8		81 ± 8%		0.09 ± 0.14		402 ± 191
2	A	6.1	7	3.6	0.2	86	0.5%	0.12	0.01	410
2	B	16.7	9	6.0	0.3	98	0.4%	RB		1409
2	C	8.6	8	6.8	0.8	96	1.0%	0.11	0.03	1089
2	D	1.0	7	5.6	0.5	95	1.0%	0.16	0.04	92
2	E	59.0	10	11.2	0.6	99.7	0.01%	0.07	0.02	12191
2	F	10.9	7	4.4	0.1	96	0.9%	0.04	0.01	965
2	G	23.8	7	4.8	0.1	92	0.8%	RB		1780
2	H	2.7	9	7.9	1.0	99.3	0.2%	ND		324
2: Mean	±SD	16.1 ± 18.9	8 ± 1	6.3 ± 2.4		95 ± 4%		0.1 ± 0.05		2282 ± 4045
3	A	6.7	8	3.7	0.3	99.7	0.4%	0.12	0.04	405
3	B	4.1	11	9.5	3.7	91	2.0%	0.11	0.03	761
3	C	5.8	13	5.7	0.7	98	0.5%	ND		523
3	D	0.4	5	6.0	0.8	99.0	0.2%	0.4	0.05	42
3	E	18.3	7	6.0	0.9	97.5	1.6%	RB		2136
3	F	1.1	14	5.8	0.6	90	0.3%	0.33	0.03	112
3	G	6.0	NR	NR		NR		NR		NR
3: Mean	±SD	6.0 ± 5.9	9.5 ± 3.5	6.1 ± 1.9		96 ± 4%		0.24 ± 0.15		663 ± 769
All: Mean	±SD	9.4 ± 12.4	8.7 ± 2.3	6.2 ± 1.8		—		0.14 ± 0.13		1276 ± 498

Average virion production rate of 1.3×10^{12} virions per day

Hepatitis C Viral Dynamics and Interferon-α Therapy

Modeling 23 patients during 14 days of therapy (daily doses)



Suggests immune control has important role in lowering viral load

Patients with undetectable HCV after 3 months of therapy (filled symbols) had significantly faster cell death rates

Major impact on understanding HIV/AIDS

HIV-I protease inhibitor given to twenty infected patients in order to perturb the balance between virus production and clearance.

ARTICLES

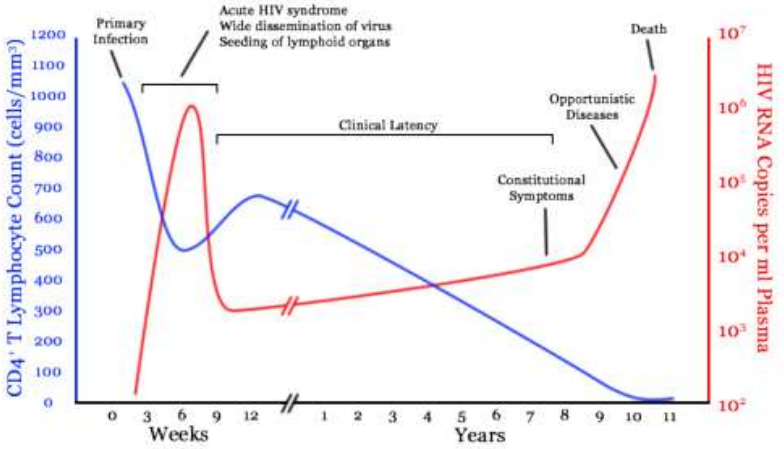
Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection

David D. Ho, Avidan U. Neumann*†, Alan S. Perelson†, Wen Chen, John M. Leonard‡ & Martin Markowitz

Aaron Diamond AIDS Research Center, NYU School of Medicine, 455 First Avenue, New York, New York 10016, USA
* Santa Fe Institute, Santa Fe, New Mexico 87501, USA
† Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
‡ Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

Treatment of infected patients with ABT-538, an inhibitor of the protease of human immunodeficiency virus type 1 (HIV-1), causes plasma HIV-1 levels to decrease exponentially (mean half-life, 2.1 ± 0.4 days) and CD4 lymphocyte counts to rise substantially. Minimum estimates of HIV-1 production and clearance and of CD4 lymphocyte turnover indicate that replication of HIV-1 *in vivo* is continuous and highly productive, driving the rapid turnover of CD4 lymphocytes.

NATURE · VOL 373 · 12 JANUARY 1995



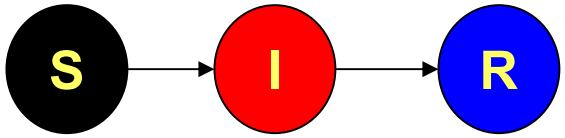
Discussion

We believe our new kinetic data have important implications for HIV-1 therapy and pathogenesis. It is self evident that, with rapid turnover of HIV-1, generation of viral diversity and the attendant increased opportunities for viral escape from therapeutic agents are unavoidable sequelae[19,20]. Treatment strategies, if they are to have a dramatic clinical impact, must therefore be initiated as early in the infection course as possible, perhaps even during seroconversion. The rapid turnover of HIV-1 in plasma also suggests that current protocols for monitoring the acute antiviral activity of novel compounds must be modified to focus on the first few days following drug initiation. Our interventional

Viral dynamics applied to a wide variety of systems

The SIR Model of Epidemics

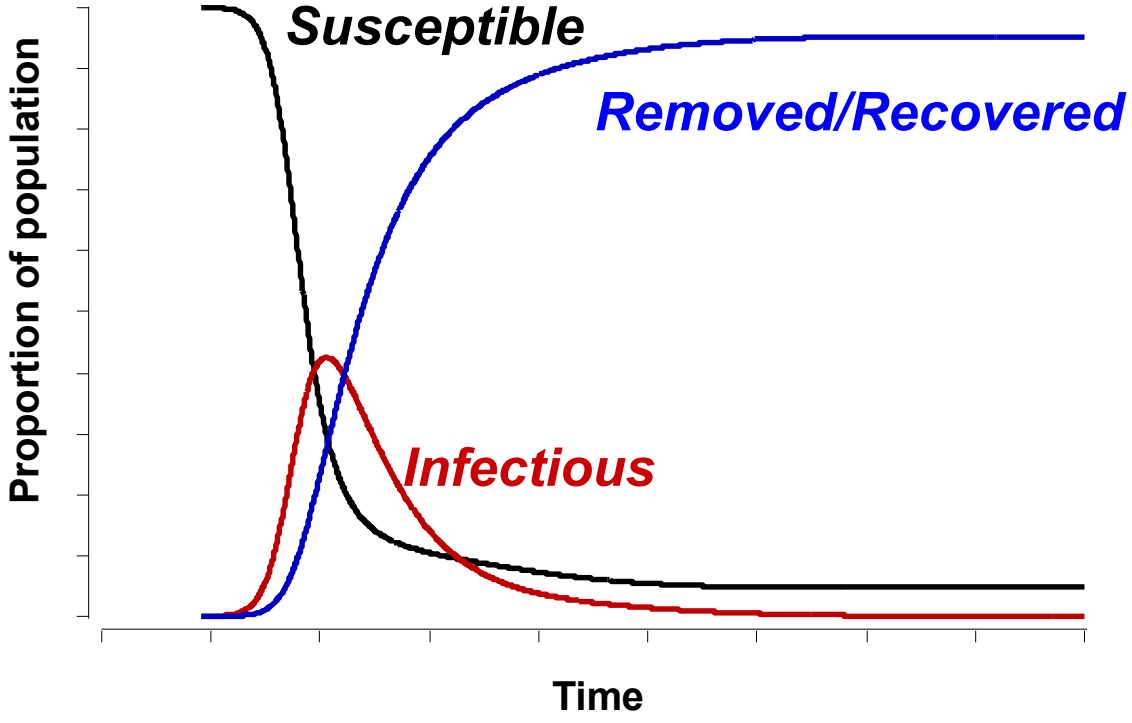
Model for many infectious diseases including measles



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \mu I$$

$$\frac{dR}{dt} = \mu I$$



Other versions allow recovered individual to be re-infected

Will the infection spread?

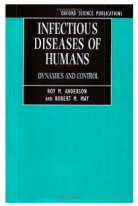
The basic reproductive ratio: R_0

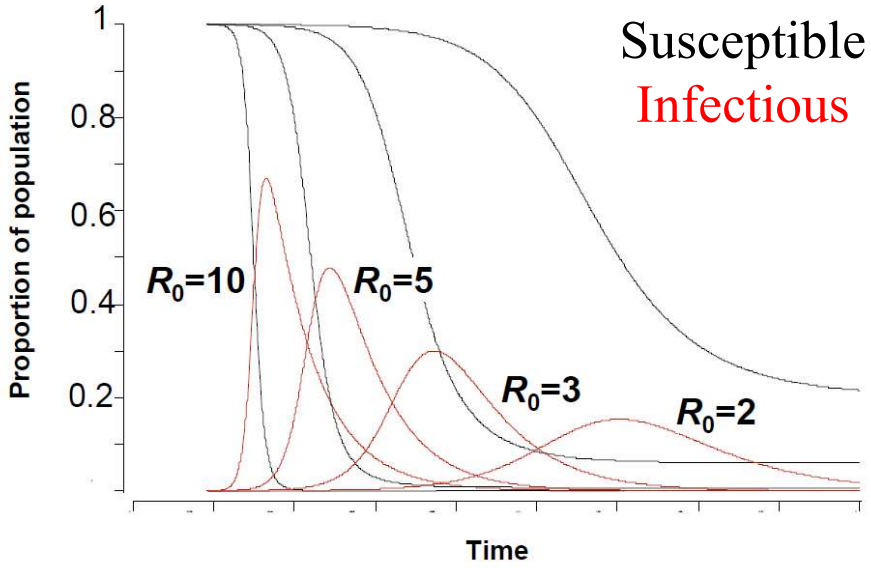
average number of secondary cases caused by an infectious individual in a totally susceptible population

$$R_0 = \frac{\beta}{\mu} \times S(0)$$

$R_0 < 1$: disease dies out

$R_0 > 1$: disease can invade

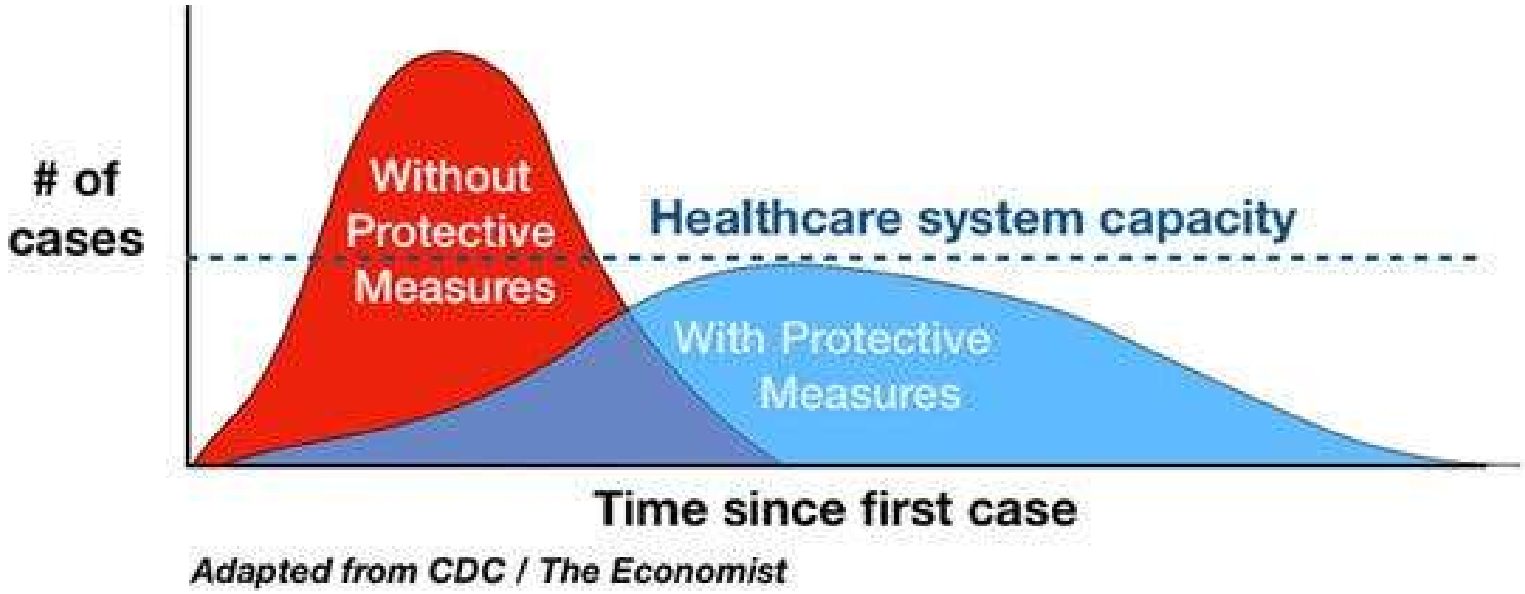




R_0 indicates whether population at risk from disease

“Flattening the Curve”

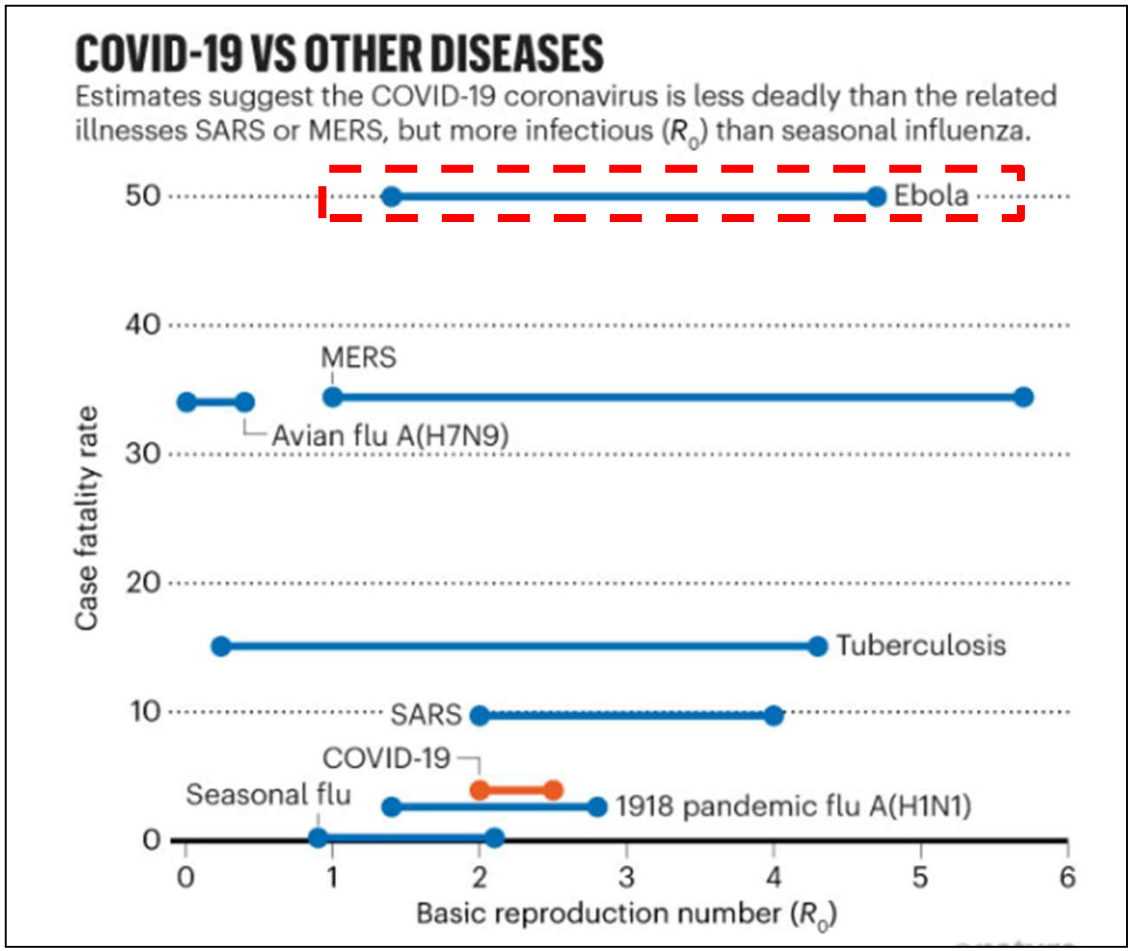
“Social distancing” increasing the physical space between people to avoid spreading illness (decreases R_0)



https://www.nytimes.com/2020/03/11/science/coronavirus-curve-mitigation-infection.html

Preventing a surge that would inundate the healthcare system

Examples of R_0 for Diseases

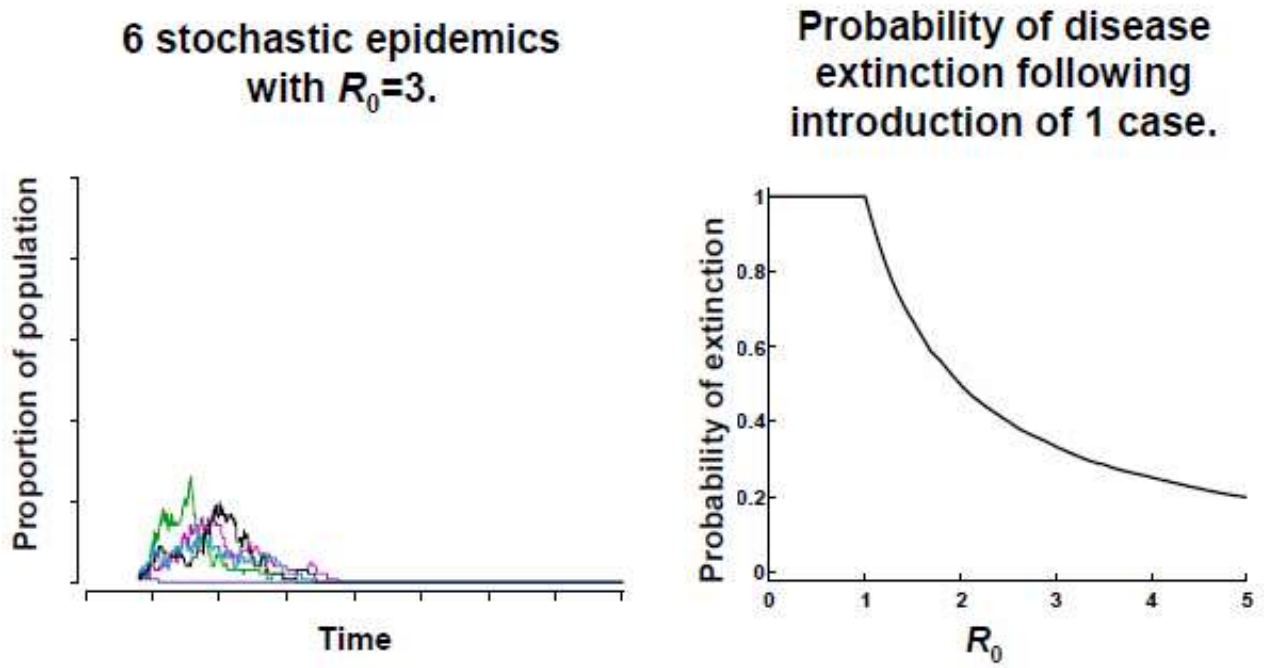


https://www.nature.com/articles/d41586-020-00758-2

Does $R_0 > 1$ guarantee spread of infection?

ODEs are deterministic

Predicts epidemic even with non-zero chance that disease dies out

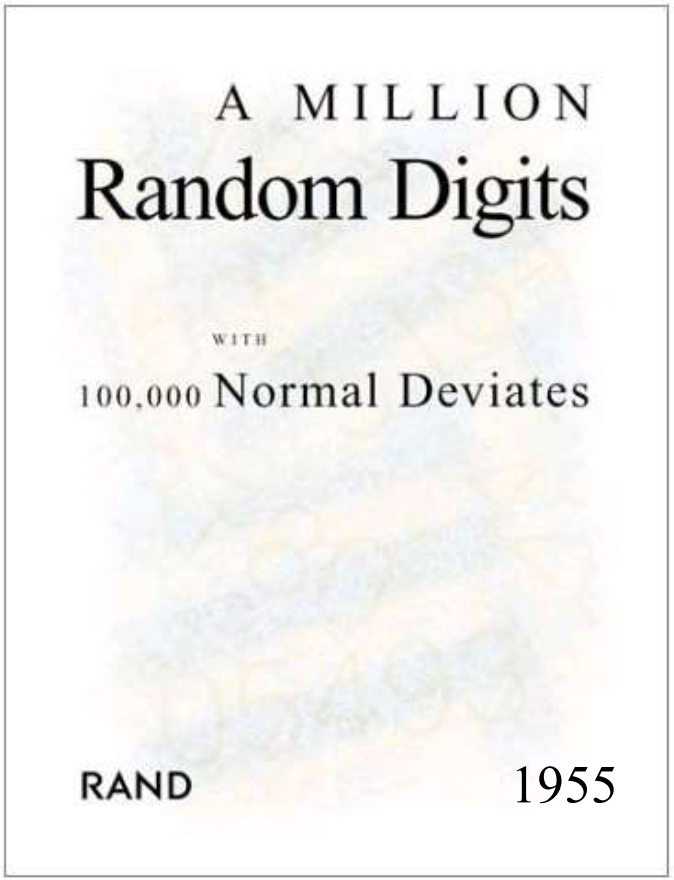


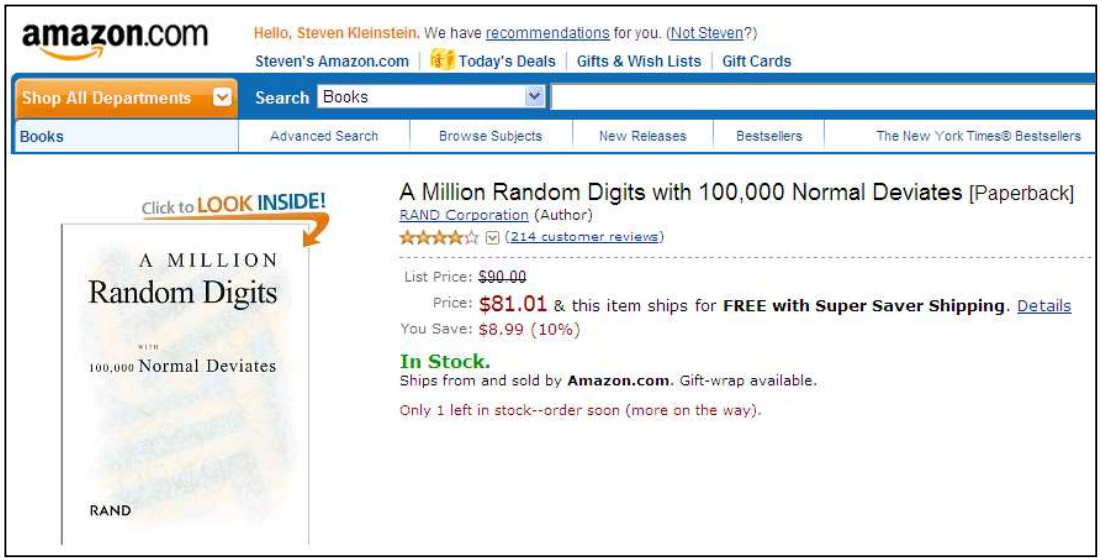
Stochasticity → risk of disease extinction when number of cases is small, even if R_0>1.

Simulate using stochastic approach – Gillepsie Method

Simulating Stochastic Models

How can we generate random number stream(s)?





almost perfect

Such a terrific reference work! But with so many terrific random digits, it's a shame they didn't sort them, to make it easier to find the one you're looking for.

Published on October 26, 2006 by a curious reader

Now we have algorithms to generate random # streams

Pseudo-Random Number Generators (PRNGs)

Starting with the same seed will give you equivalent stream

Uniform deviates: [0,1)

Linear congruential generator

$$I_{j+1} = aI_j + c \pmod{m}$$

Fast, but sequential calls can be correlated, so not used much

I_0 is the seed (common to use system clock)

$I_{j+1} = 3I_j + 7 \pmod{10}$

Produces: 6,5,2,3

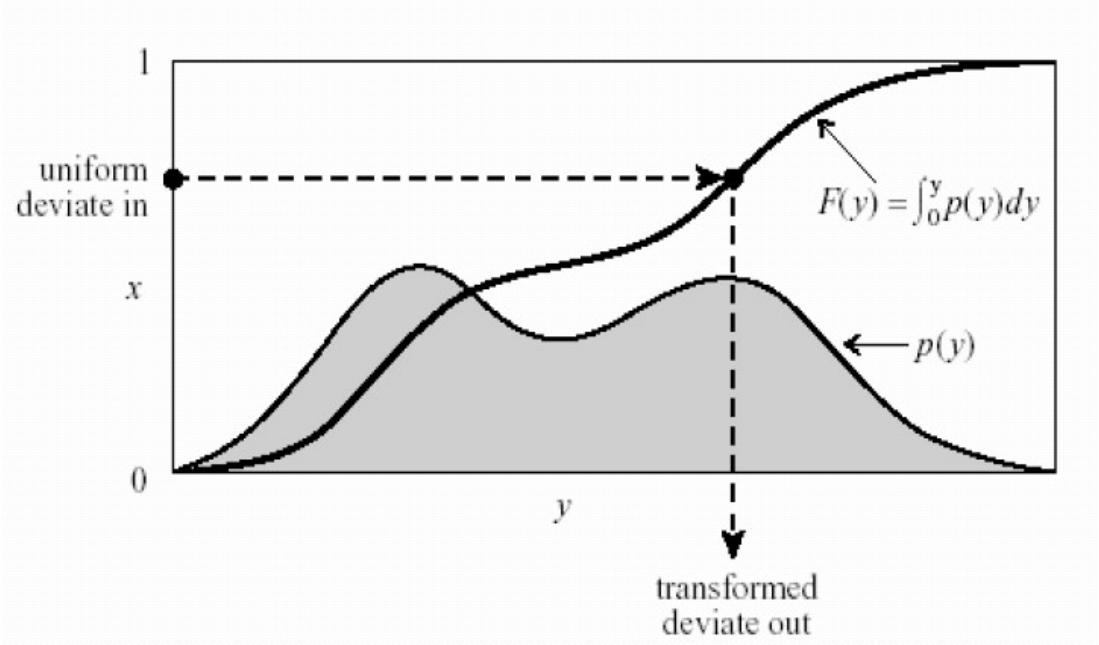
Period: time before stream repeats itself (maximum m)

Better approach
Mersenne Twister
(period 2^{19937}-1)

Be careful on computer clusters (streams can be correlated):
Check out the "parallel" package in R

Simulating from other distributions

Transformation Method: indefinite integral of p(y) must be known and invertible



Transformation to generate exponential distribution (Poisson process)

$$\text{Exponential}(\alpha) = -\frac{1}{\alpha}\ln\left[\text{Uniform}(0,1)\right]$$

Methods based on underlying ability to generate uniform distributions

Boolean Network Models

Logical modeling

Can be useful where kinetic parameters are not sufficiently known

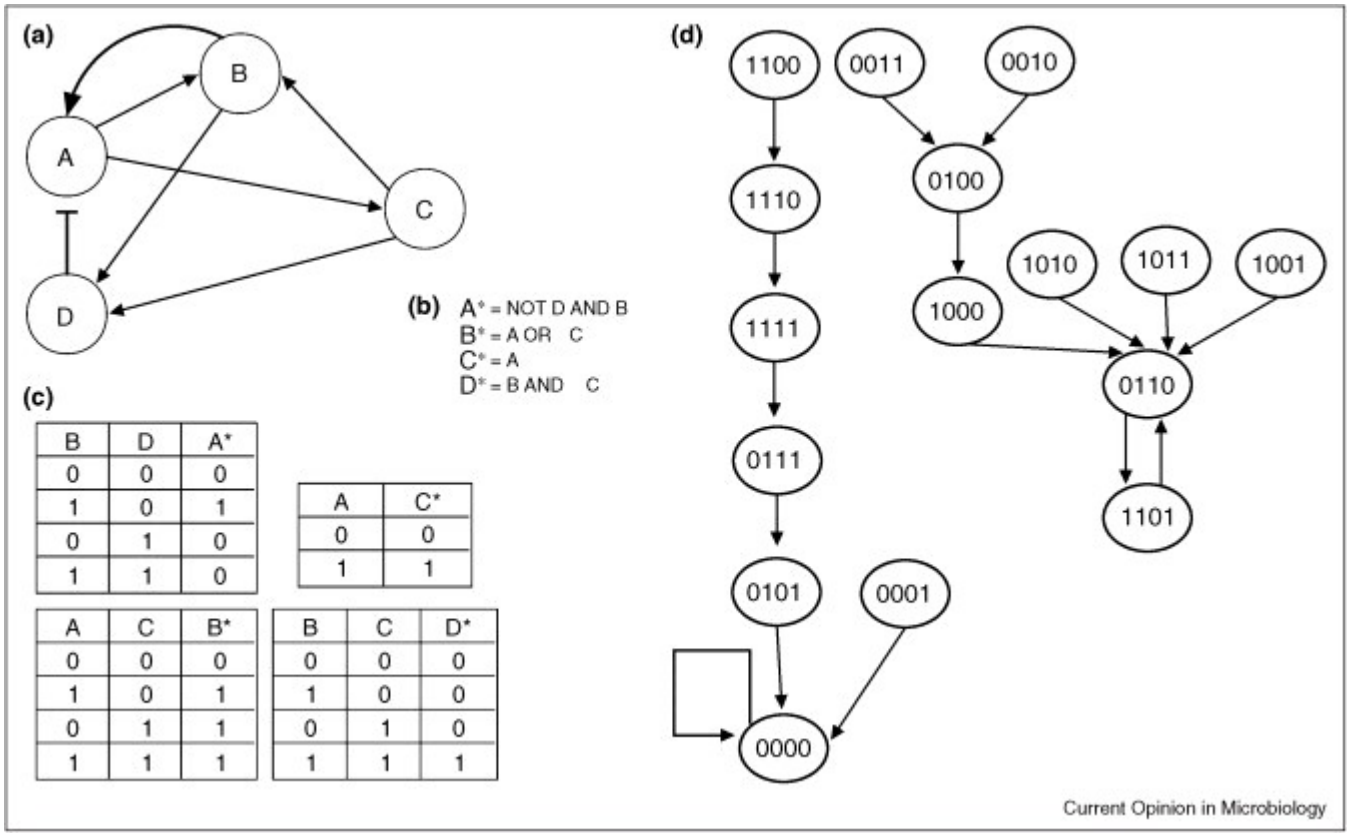
- A directed graph (network)
  - Nodes represent the elements of a system
  - Edges represent regulatory relationships between elements
- Nodes characterized by True/False state
  - Network with N nodes will have 2^N possible states
- As time passes, node state determined by the states of neighbors, through a rule called a **transfer function**
  - Eg, logical function using the operators NOT, AND, OR
  - Output of transfer function determines state of the node

Often matches biological intuition: eg, genes are on/off.

Boolean Network Models

Qualitative approach

Can be useful where kinetic parameters are not sufficiently known



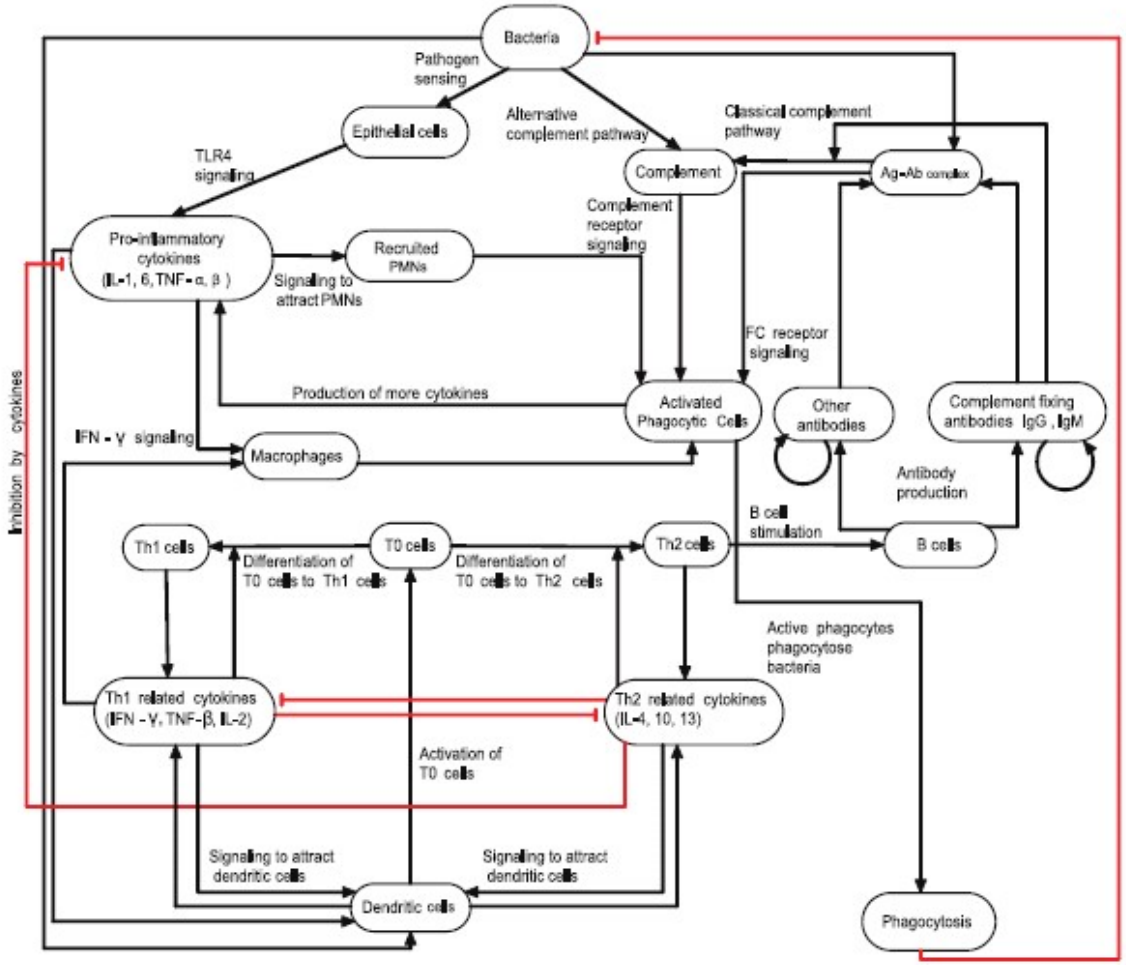
(Thakar and Alberta, 2010)

time is discrete, specifying instances in which the state of the nodes may change

Easy to model combinatorial regulatory relationships

Boolean Interaction Network of Immune Response

Can be useful where kinetic parameters are not sufficiently known



Thakar J., *et.al.* (2007) PloS CB

Future states of each node decided by transition rules using Boolean operators

ODEs Neglect Spatial Structure

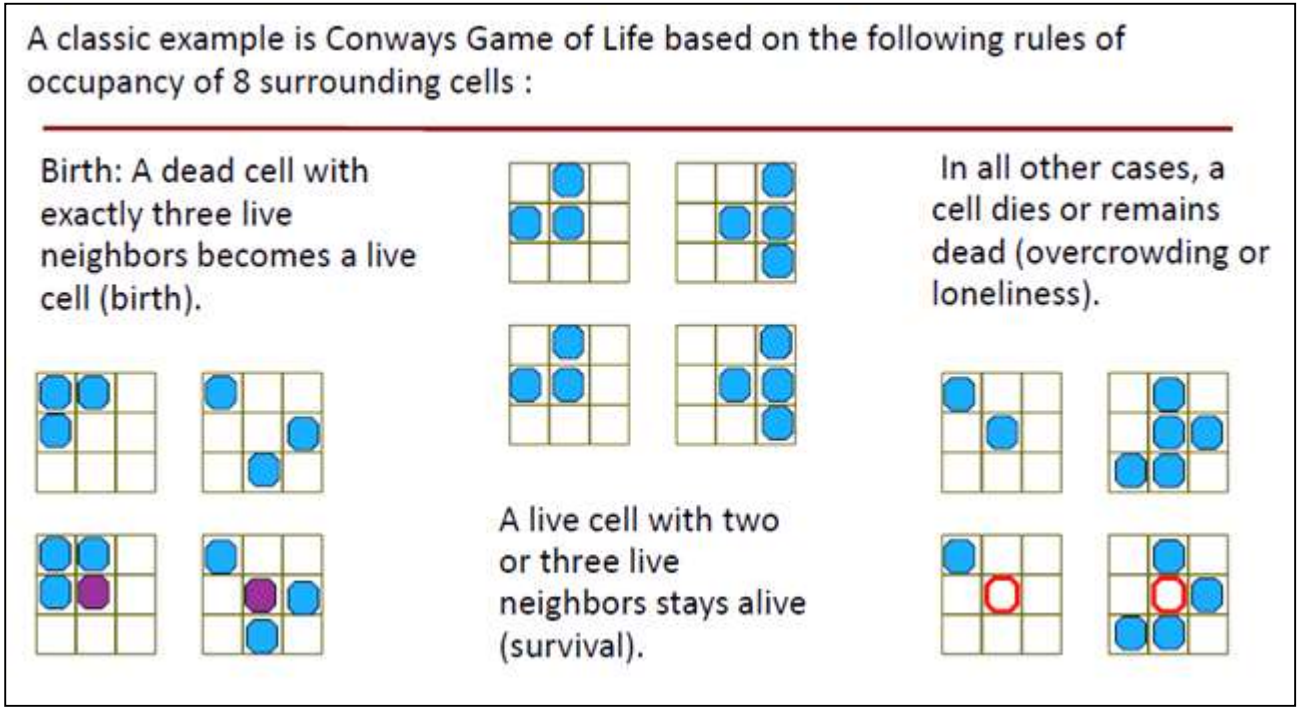
Several approaches to including spatial effects

- **Partial Differential Equations (PDEs)**
  - Allows quantities to vary over both space and time
  - Continuous and deterministic
- **Compartment Modeling**
  - Compartments assumed to be well-mixed
  - Elements present in each compartment tracked using ODEs.
  - ODEs incorporate coupling between compartments
- **Agent-Based Modeling (ABM)**
  - object-oriented, discrete-event, rule-based, stochastic
  - views system as an aggregation of components (agents) that follow intrinsic rules of behavior (agent-rules)

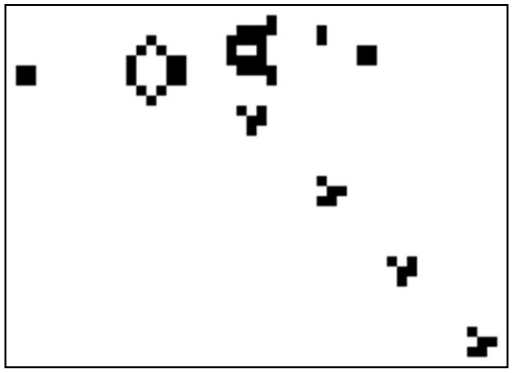
"Right" approach depends on question, and available data

Cellular Automata Models

A regular grid of cells, each in one of a finite number of states



(John Parkinson)



Gosper's Glider Gun

A new generation is created (advancing t by 1), according to some fixed rule (generally, a mathematical function) that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood.

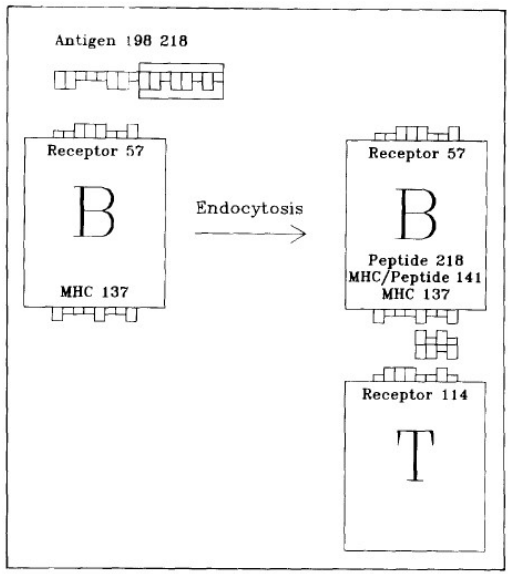
Agent-based Models (ABMs): IMMSIM

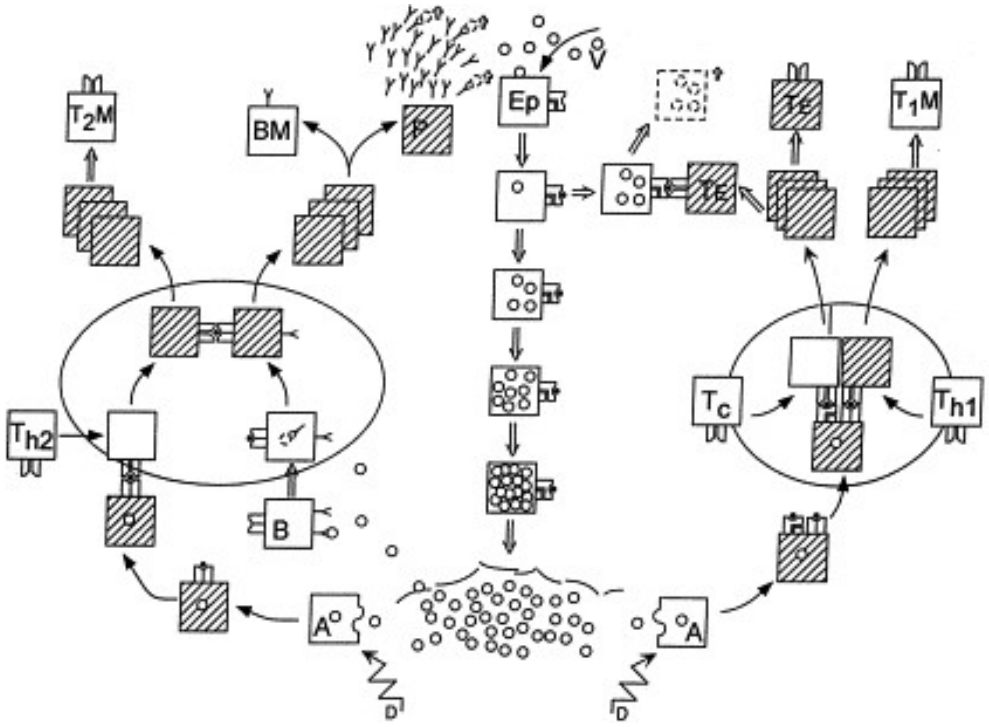
Individual cells given unique properties: receptors and internal state

A computer model of cellular interactions in the immune system

Franco Celada and Philip E. Seiden

Immunology Today 56 Vol. 13 No. 2 1992



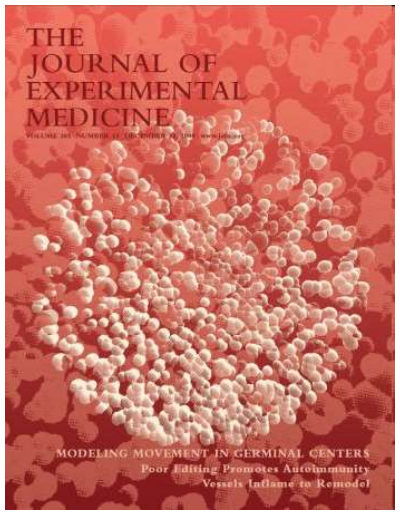


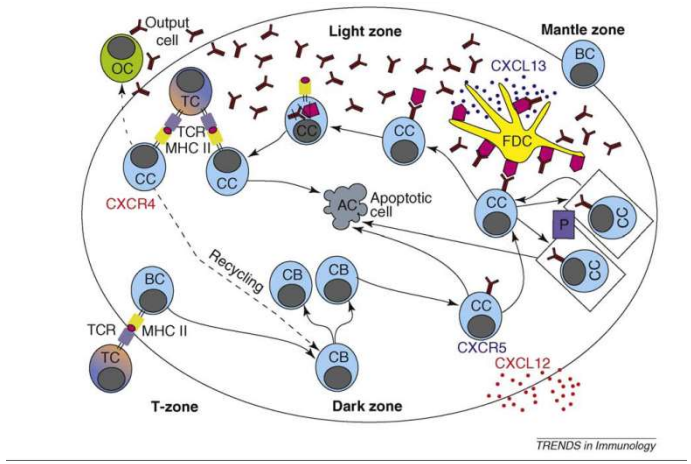
(Kohler et al, 2000)

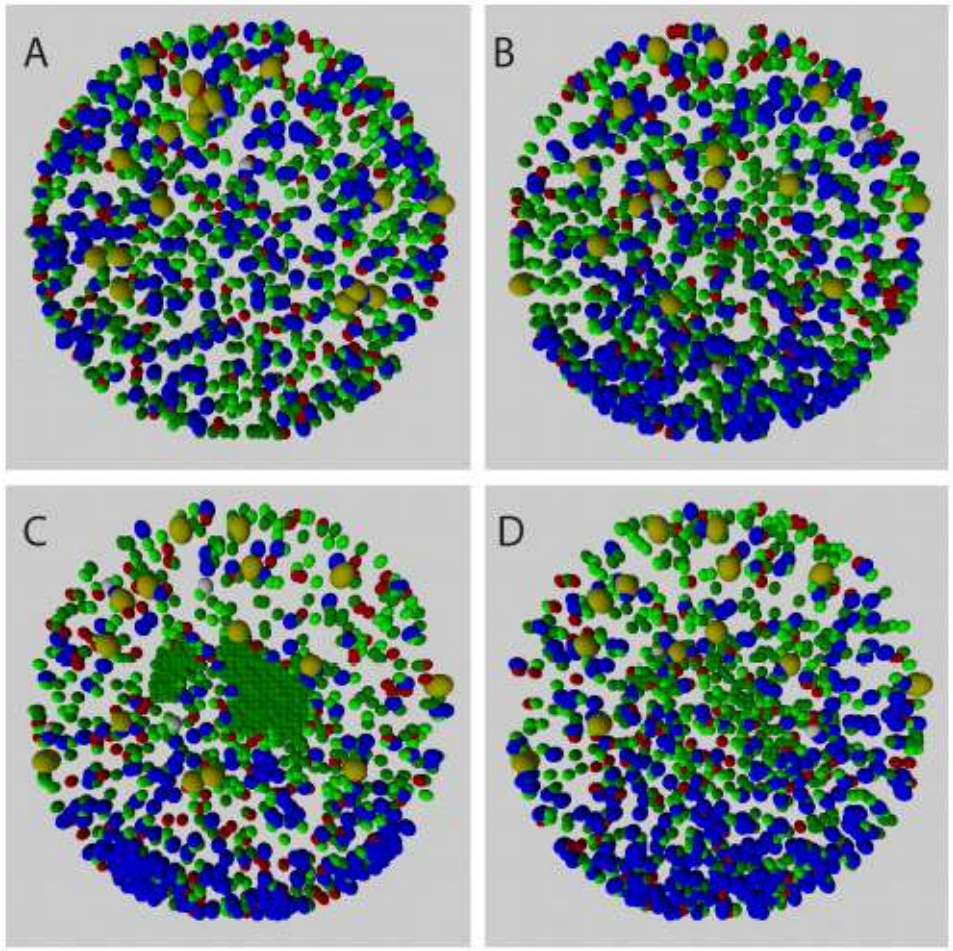
study immune receptor signal–based cellular behavior with a bit-string representation for receptor specificities

Detailed spatial pattern formation

Realistic models of cell diffusion and response to chemokines







New intravital imaging techniques provide underlying data

Range of Current Modeling Frameworks

Various types of computational models can be built

Table 1 Computational approaches and tools for systems biology

Modeling approach	Typical applications	Limitations	Tools
Individual particle-based stochastic	Small subcellular signaling processes, aspects of bacterial biochemistry	Applies only to small systems (in terms of space and chemical complexity)	MCell (32), Smoldyn (314), ChemCell (315), GetBonNie (nonspatial) (49)
Particle number stochastic	Signaling processes with important stochastic aspects (due to small system size or high sensitivity)	Applies only to small systems (in terms of space and chemical complexity), has less detail than individual particle simulation	MesoRD (35), SmartCell (33), GetBonNie (nonspatial)
Concentration-based spatial, nonstochastic	Cellular signaling processes with important spatial aspects	Provides either high spatial resolution or biochemical complexity, has no stochasticity	Virtual Cell (37), Simmune (36)
Concentration-based, nonspatial, nonstochastic	Cellular signaling processes without spatial aspects	Assumes global biochemical homogeneity in the simulated system	Copasi (46), E-cell (44), Cellware (45), Systems Biology Workbench (47), GetBonNie

(Germain et al, 2010)

Each method has advantages and limitations – no one right approach.

Interchange format for computer models

XML encoding: wide variety of models can be described

SBML.org The Systems Biology Markup Language

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <sbml xmlns="http://www.sbml.org/sbml/level1"
3        level="1" version="2">
4    <model name="gene_network_model">
5      <listOfUnitDefinitions>
6          ...
7      </listOfUnitDefinitions>
8      <listOfCompartments>
9          ...
10     </listOfCompartments>
11     <listOfSpecies>
12         ...
13     </listOfSpecies>
14     <listOfParameters>
15         ...
16     </listOfParameters>
17     <listOfRules>
18         ...
19     </listOfRules>
20     <listOfReactions>
21         ...
22     </listOfReactions>
23   </model>
24 </sbml>
```

A software package can read in a model expressed in SBML and translate it into its own internal format for model analysis.

```
<listOfReactions>
 <reaction name="R1" reversible="false">
   <listOfReactants>
     <species Reference species="src" />
   </listOfReactants>
   <listOfProducts>
     <species Reference species="RNAP"/>
   </listOfProducts>
   <kineticLaw formula="Vi/(1+P/Ki)" />
 </reaction>
 ...

</listOfReactions>
```

Still, most researchers develop models from scratch for every project

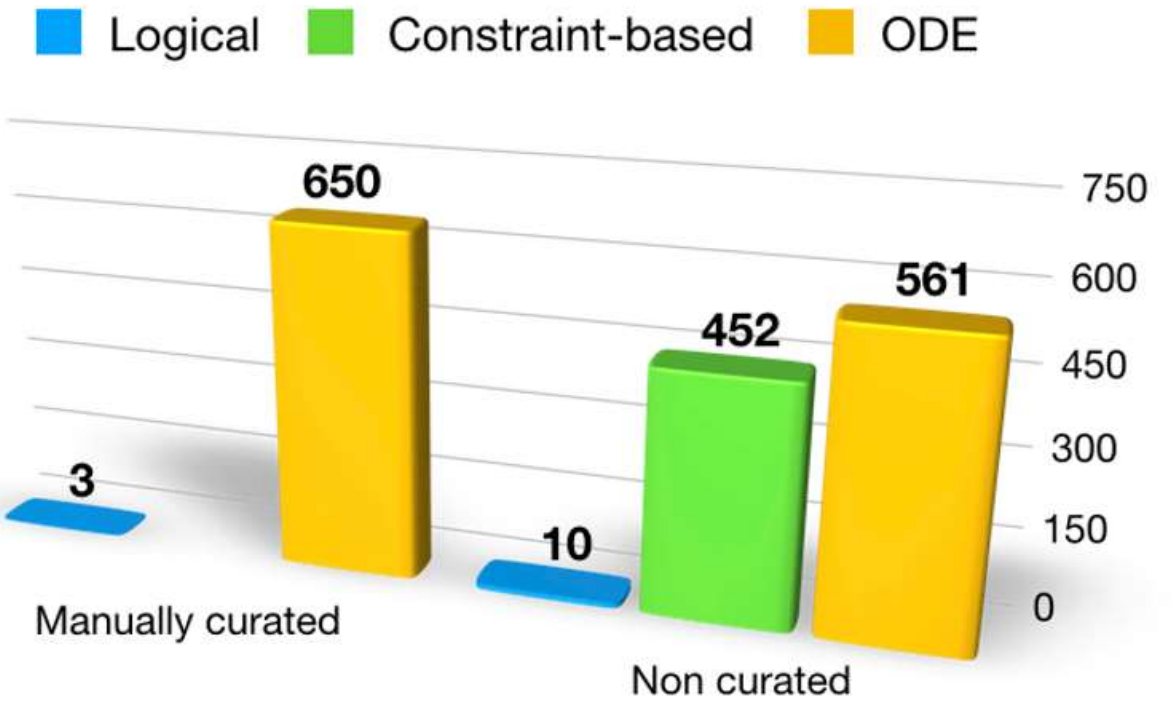
Repository of mathematical models

BioModels (http://www.biomodels.org)

Two branches
- ☑ Manually curated
- ☑ Non curated

Model formats
- ☑ SBML
- ☑ CellML
- ☑ Matlab
- ☐ …

Modelling approaches
- ☑ Ordinary Differential Equation
- ☑ Logical
- ☑ Constraint-based
- ☐ …



Still, most researchers develop models from scratch for every project

For more information…

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Message from ISCB

Getting Started in Computational Immunology

Steven H. Kleinstein*

Interdepartmental Program in Computational Biology and Bioinformatics, and Department of Pathology, Yale University School of Medicine, New Haven, Connecticut, United States of America

TEACHING RESOURCE

COMPUTATIONAL BIOLOGY

Biomedical Model Fitting and Error Analysis

Kevin D. Costa,[1,*] Steven H. Kleinstein,[2,3] Uri Hershberg[4]

www.SCIENCESIGNALING.org 27 September 2011 Vol 4 Issue 192

Trends in Immunology

CellPress

Review

Solving Immunology?

Yoram Vodovotz,[1] Ashley Xia,[2,17] Elizabeth L. Read,[3] Josep Bassaganya-Riera,[4] David A. Hafler,[5] Eduardo Sontag,[6] Jin Wang,[7,8] John S. Tsang,[9] Judy D. Day,[10] Steven H. Kleinstein,[11,12] Atul J. Butte,[13] Matthew C. Altman,[14] Ross Hammond,[15] and Stuart C. Sealfon[16,*]

Feel free to email me with questions:

steven.kleinstein@yale.edu