iTHEMS @ RIKEN
RIKEN interdisciplinary Theoretical & Mathematical Sciences

Ryerson University
Department of Physics
Faculty of Science

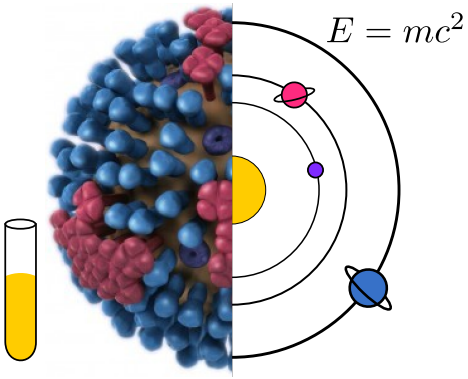
Lack of math: Side effects include dizziness and distrust

Nerd Nite Tokyo — 11 May 2018

Catherine Beauchemin
カトゥリン ・ ボシゥメン

Funded by: iTHEMS RIKEN interdisciplinary Theoretical & Mathematical Sciences; CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada; NSERC CRSNG; Ontario

Quick intro to me...



I was born/grew up in
Montréal, Canada (French)

My research is in a field I dubbed
ViroPhysics
(using physics' approach to study viruses)

1/3 of the time:
Senior Visiting Scientist
Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS) programme at RIKEN
Wako, Japan.

2/3 of the time:
Professor of Physics
Ryerson University
Toronto, Canada.

There is a crisis of trust in physicians

Public Trust in Physicians — U.S. Medicine in International Perspective

Robert J. Blendon, Sc.D., John M. Benson, M.A., and Joachim O. Hero, M.P.H.

The U.S. health care reform process i
phase, its emp
expanding h
improving our
ing patient c
question is w
cal professio
will play in
tional health
affect decisi
patient care.

Research
physicians to
role in such
there has to
level of public
sion's views
an examinati
opinion data

Social Survey Programme (ISSP), a

deal or quite a lot of confidence

2014). We
al profes-
seen as a

does not
any other
vel of pub-
s a group
nks near
els in the
tries sur-
closer ex-
nparisons
to those
s: individ-
with the
ived dur-
physician
ne decline

All things considered, Doctors in my country can be trusted (agree or strongly agree):

Switzerland	83% (1st)
Denmark	79% (2nd)
Australia	73% (10th)
Taiwan	72% (12th)
South Korea	62% (20th)
Japan	60% (23rd)
United States	58% (24th)
Russia	45% (28th)

N ENGL J MED 371;17 NEJM.ORG OCTOBER 23, 2014

The New England Journal of Medicine

Distrust of clinical practice somewhat justified

ORIGINAL ARTICLE

MAYO CLINIC

A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices

Vinay Prasad, MD; Andrae Vandross, MD; Caitlin Toomey, MD; Michael Cheung, MD; Jason Rho, MD; Steven Quinn, MD; Satish Jacob Chacko, MD; Durga Borkar, MD; Victor Gall, MD; Senthil Selvaraj, MD; Nancy Ho, MD; and Adam Cifu, MD

Of the 363 articles testing standard of care, 146 (40.2%) reversed that practice, whereas 138 (38.0%) reaffirmed it.

therapy. This study was conducted from August 1, 2011, through October 31, 2012.

Results: We reviewed 2044 original articles, 1344 of which concerned a medical practice. Of these, 981 articles (73.0%) examined a new medical practice, whereas 363 (27.0%) tested an established practice. A total of 947 studies (70.5%) had positive findings, whereas 397 (29.5%) reached a negative conclusion. A total of 756 articles addressing a medical practice constituted replacement, 165 were back to the drawing board, 146 were medical reversals, 138 were reaffirmations, and 139 were inconclusive. Of the 363 articles testing standard of care, 146 (40.2%) reversed that practice, whereas 138 (38.0%) reaffirmed it.

Conclusion: The reversal of established medical practice is common and occurs across all classes of medical practice. This investigation sheds light on low-value practices and patterns of medical research.

Published by Elsevier Inc on behalf of Mayo Foundation for Medical Education and Research ■ Mayo Clin Proc. 2013;88(8):790-798

Distrust of clinical practice somewhat justified

ORIGINAL ARTICLE

MAYO CLINIC

A Decade of Reversal: An Analysis of 146

E.g. The practice of implanting Gentamicin-collagen sponge to prevent infection following colorectal surgery, used in millions of patients worldwide since 1985...

A single-centre, randomized trial found a 70% decrease in surgical site infection with this practice.

In a larger, multi-centre, phase 3 trial it resulted in significantly more infections, more visits to emergency departments, and more hospitalization for resulting infection.

A problem with the institutional culture

In the shame/blame environment, where errors are seen as a form of personal moral failure that shatters the ***culture of infallibility*** *inculcated in physicians since the first day of professional training, the physician's ultimate fear — "losing face" in front of one's peers.*

Prof. Albert A. Wu, MD
Medical Error: The Second Victim.
British Medical Journal 320: pp.726–727, 2000

Burnout, compassion fatigue, and mental illness are enormous challenges to our profession, and ***working in a culture that demands infallibility*** *is a predisposing factor that we can and should address.*

Linda Fineman, DVM, DACVIM (Oncology)
To Err Is Human
American Veterinarian, January 19, 2018

A problem with basic (lab) health research

Review

Reproducibility in Science

Improving the Standard for Basic and Preclinical Research

C. Glenn Begley, John P.A. Ioannidis

Abstract: **Medical and scientific advances are predicated on new knowledge that is robust and reliable and that serves as a solid foundation on which further advances can be built. In biomedical research, we are in the midst of a revolution with the generation of new data and scientific publications at a previously unprecedented rate. However, unfortunately, there is compelling evidence that the majority of these discoveries will not stand the test of time. To a large extent, this reproducibility crisis in basic and preclinical research may be as a result of failure to adhere to good scientific practice and the desperation to publish or perish. This is a multifaceted, multistakeholder problem. No single party is solely responsible, and no single solution will suffice. Here we review the reproducibility problems in basic and preclinical biomedical research, highlight some of the complexities, and discuss potential solutions that may help improve research quality and reproducibility. (*Circ Res*. 2015;116:116-126. DOI: 10.1161/CIRCRESAHA.114.303819.)**

Key Words: funding ■ journals ■ research integrity ■ universities

Problem

As physicians and scientists, we want to make a contribution that alters the course of human health. We all want to make

remarkably well with estimates of 85% for the proportion of biomedical research that is wasted at-large.[4–9] This irreproducibility is not unique to preclinical studies. It is seen across the

A problem with basic (lab) health research

Review

Reproducibility in Science

Improving the Standard for Basic and Preclinical Research

C. Glenn Begley, John P.A. Ioannidis

Table 1. Examples of Some Reported Reproducibility Concerns in Preclinical Studies (Modified - cut)

Author	Field	Reported Concerns
Ioannidis et al (2009)[22]	Microarray data	16/18 studies unable to be reproduced in principle from raw data
Sena et al (2010)[24]	Stroke animal studies	Overt publication bias: only 2% of the studies were negative
Prinz (2011)[1]	General biology	75% to 80% of 67 studies were not reproduced
Begley & Ellis (2012)[2]	Oncology	90% of 53 studies were not reproduced
Elliott et al (2006)[31]	Commercial antibodies	Commercial antibodies detect wrong antigens
Prassas et al (2013)[32]	Commercial ELISA	ELISA Kit identified wrong antigen

from http://circr

Problem

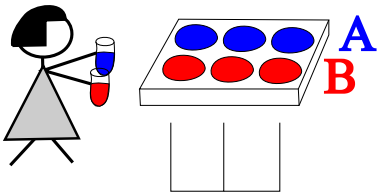
As physicians and scientists, we want to make a contribution that alters the course of human health. We all want to make

remarkably well with estimates of 85% for the proportion of biomedical research that is wasted at-large.[4–9] This irreproducibility is not unique to preclinical studies. It is seen across the

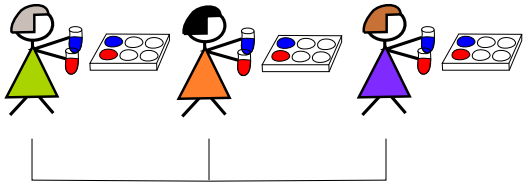
Why are health results not reproducible?

reproducible can be predicted to recur even when experimental conditions may vary to some degree (robust under change). → **Scientific req**

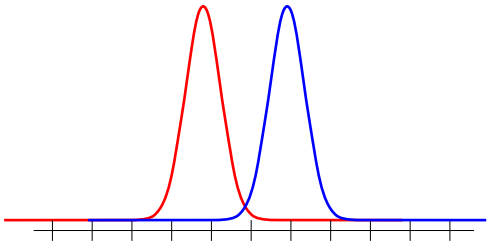
replicable ability to obtain an identical result when an experiment is performed under precisely identical conditions (requires no change).



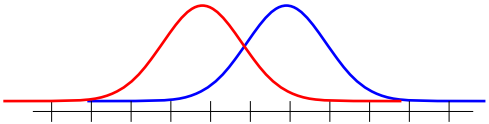
replicated 3x



reproduced 3x

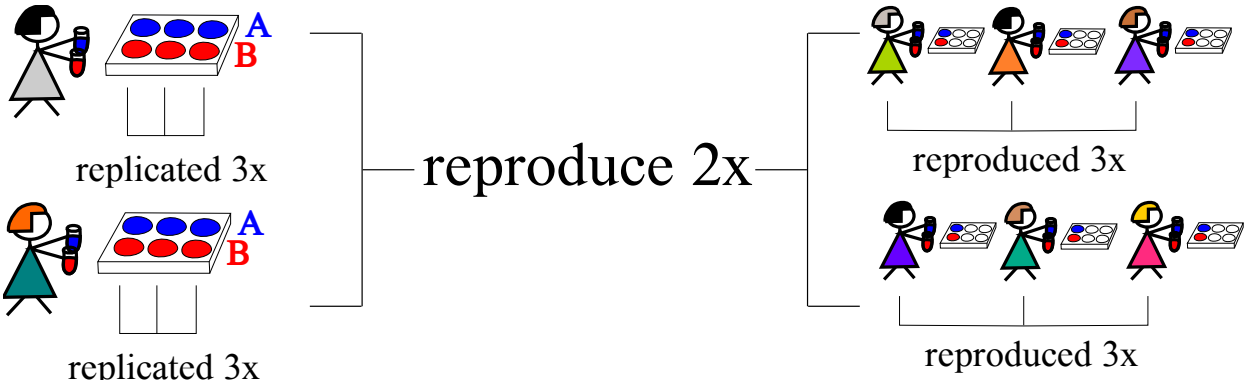


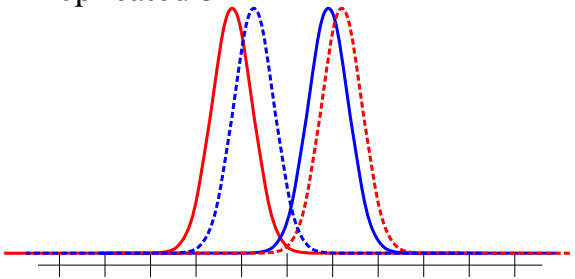
Replicated #1: **A** > **B**

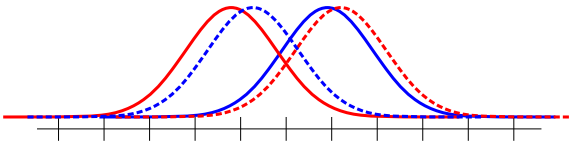


Reproduced #1: **A** ≈ **B**

Underestimating variability = bad science







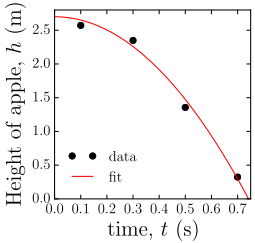
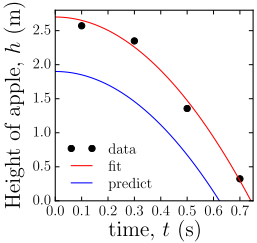
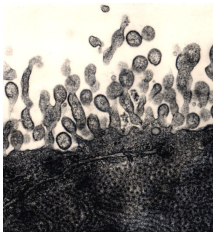
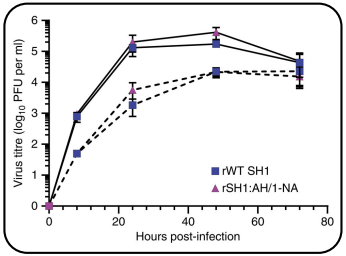
Replicated #1: **A** > **B**
Replicated #2: **A** < **B**

Reproduced #1: **A** ≈ **B**
Reproduced #2: **A** ≈ **B**

Results disagree (wrong)

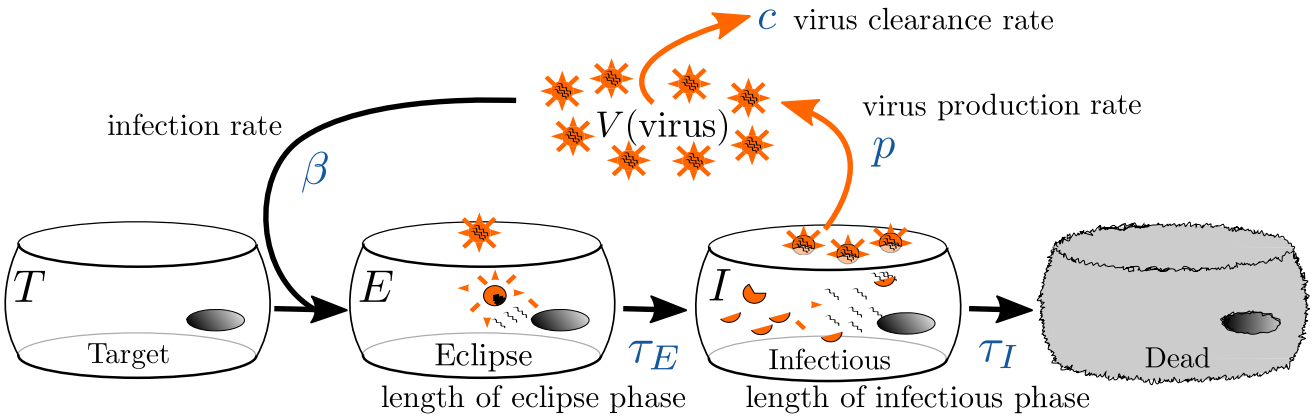
Results agree!

Health research suffers from lack of MATH!

	observe	analyze	explain	predict
Physics			$F = ma$ $a = g$ $v_f = v_i - gt$ $h_f = h_i - \frac{1}{2}gt^2$	
Virology			When cells are infected with influenza, approx. 10^6 pfu/mL virions are produced.	Additional funding is required in order to determine what happens when a cell is infected with 2× more virus...

- Experimental results **alone**, i.e. in the absence of a mathematical description, are of **limited scope** (non-predictive) and open to **mis-interpretation**.
- The qualitative (rather than quantitative) nature of experimental findings makes them **hard to validate or challenge**.
- Both experimental and theoretical investigations **together** are required to robustly advance a research field.

Virophysics: the laws of virus infection



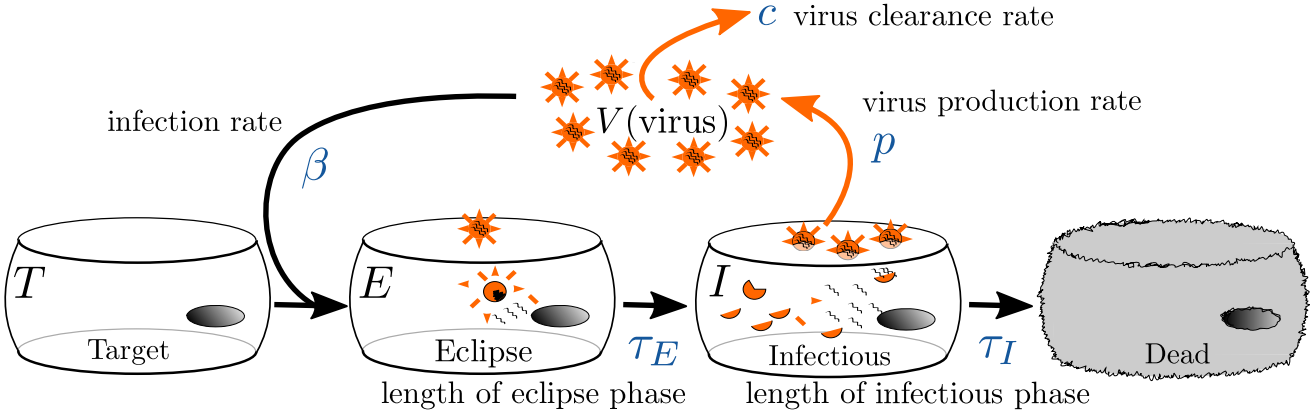
$$\text{Target} \quad \frac{\mathrm{d}T}{\mathrm{d}t} = -\beta T V_{\text{inf}}$$

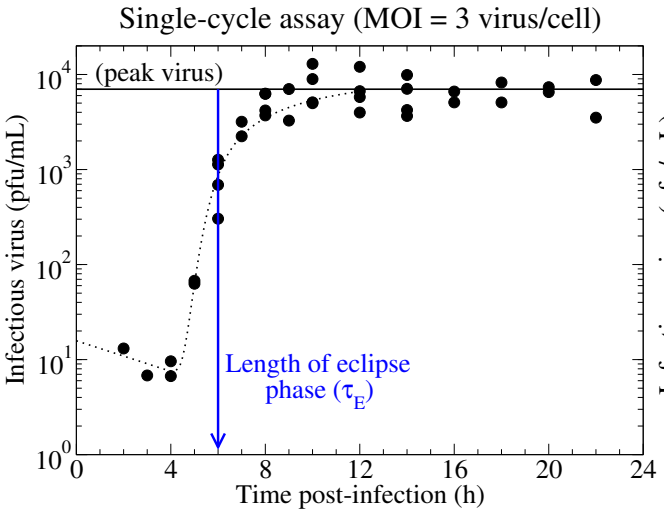
$$\text{Eclipse} \quad \frac{\mathrm{d}E_1}{\mathrm{d}t} = \beta T V_{\text{inf}} - \frac{n_E}{\tau_E} E_1 \qquad \frac{\mathrm{d}E_{i=2,3,\ldots,n_E}}{\mathrm{d}t} = \frac{n_E}{\tau_E} E_{i-1} - \frac{n_E}{\tau_E} E_i$$

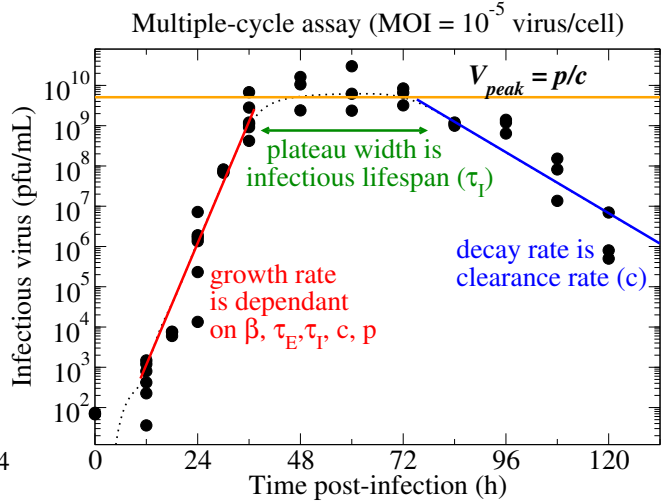
$$\text{Infectious} \quad \frac{\mathrm{d}I_1}{\mathrm{d}t} = \frac{n_E}{\tau_E} E_{n_E} - \frac{n_I}{\tau_I} I_1 \qquad \frac{\mathrm{d}I_{j=2,3,\ldots,n_I}}{\mathrm{d}t} = \frac{n_I}{\tau_I} I_{j-1} - \frac{n_I}{\tau_I} I_j$$

$$\text{Virus} \quad \frac{\mathrm{d}V_{\text{inf}}}{\mathrm{d}t} = \underbrace{p_{\text{inf}} \sum_{j=1}^{n_I} I_j}_{\text{production}} - \underbrace{c_{\text{inf}} V_{\text{inf}}}_{\text{infectivity decay}} \qquad \frac{\mathrm{d}V_{\text{tot}}}{\mathrm{d}t} = p_{\text{tot}} \sum_{j=1}^{n_I} I_j - c_{\text{tot}} V_{\text{tot}}$$

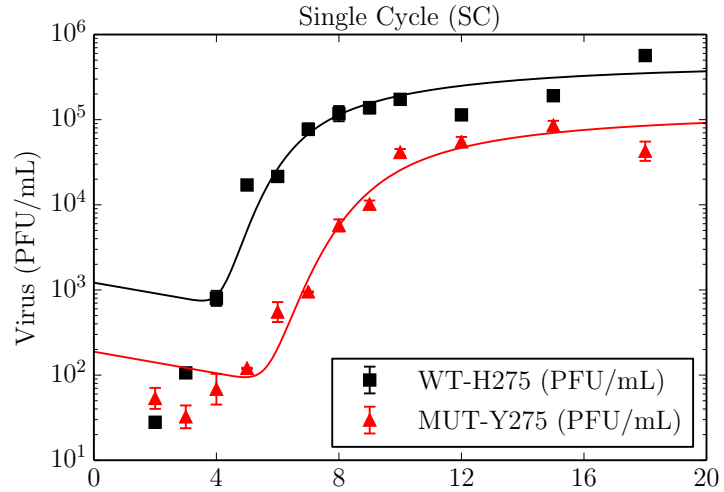
Matching math model with experiments

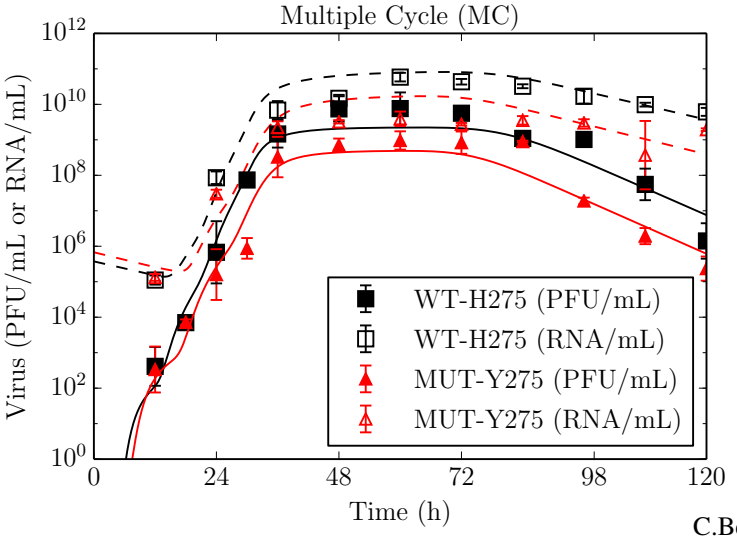


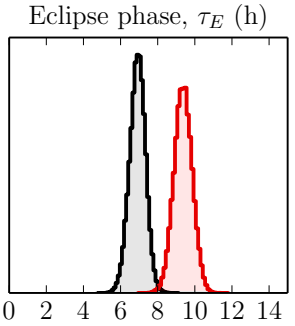


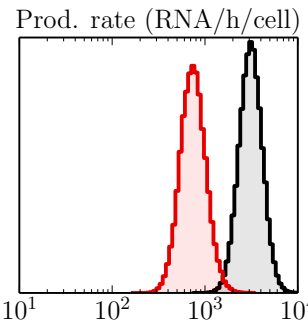


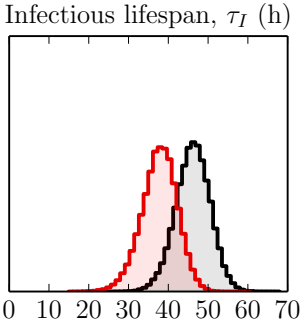
Differences between wild-type & mutant strain

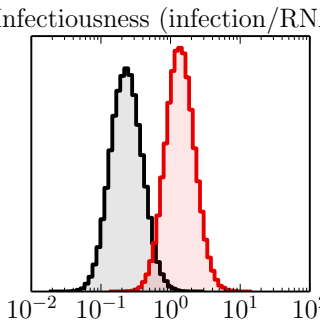


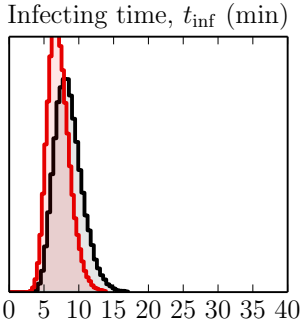


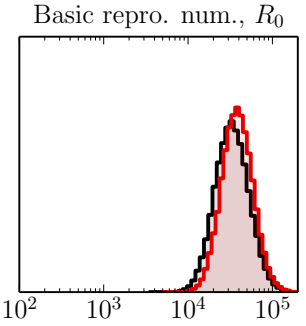




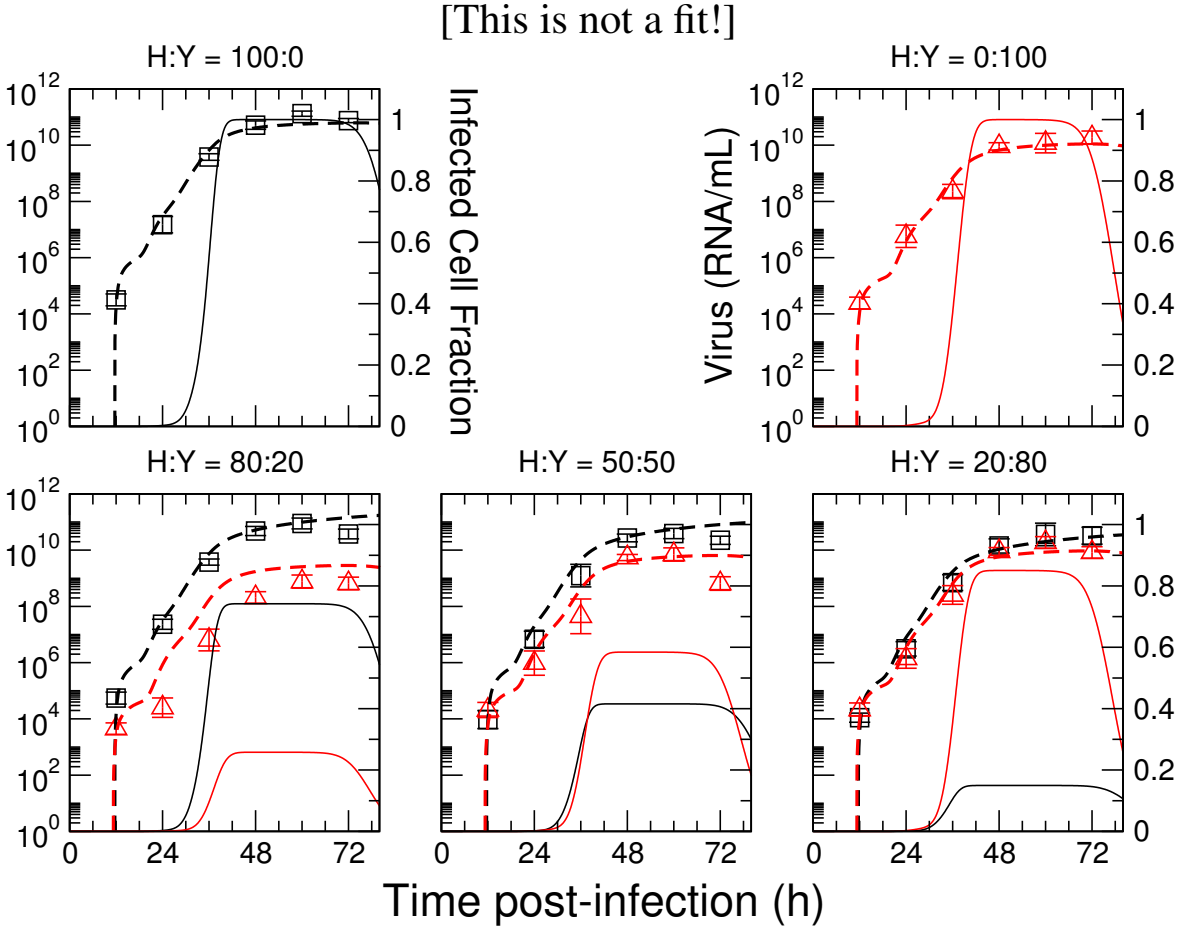






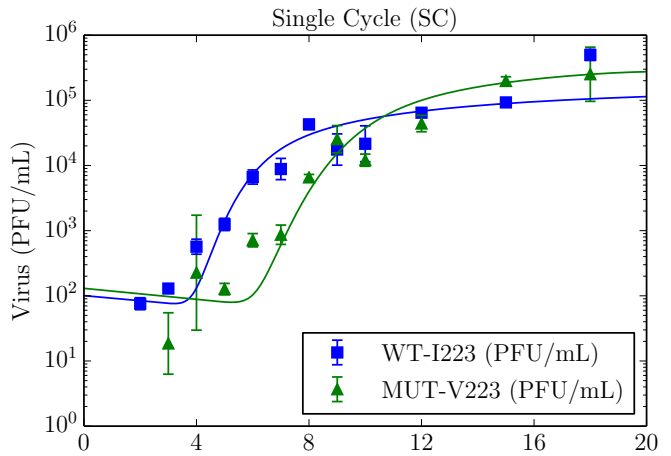


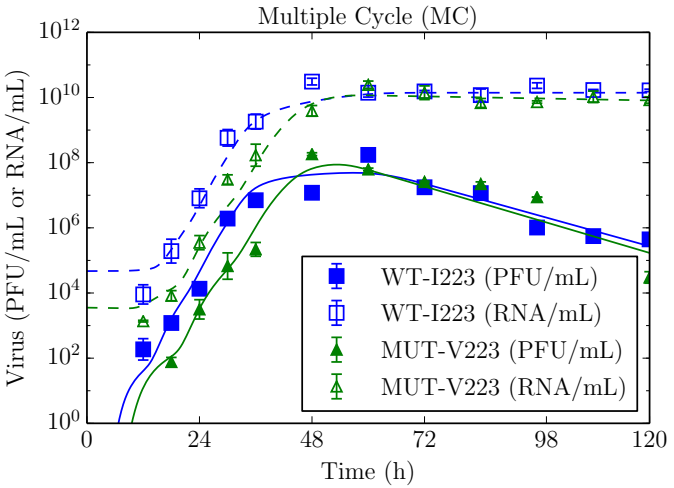
Math model predicts experimental outcomes

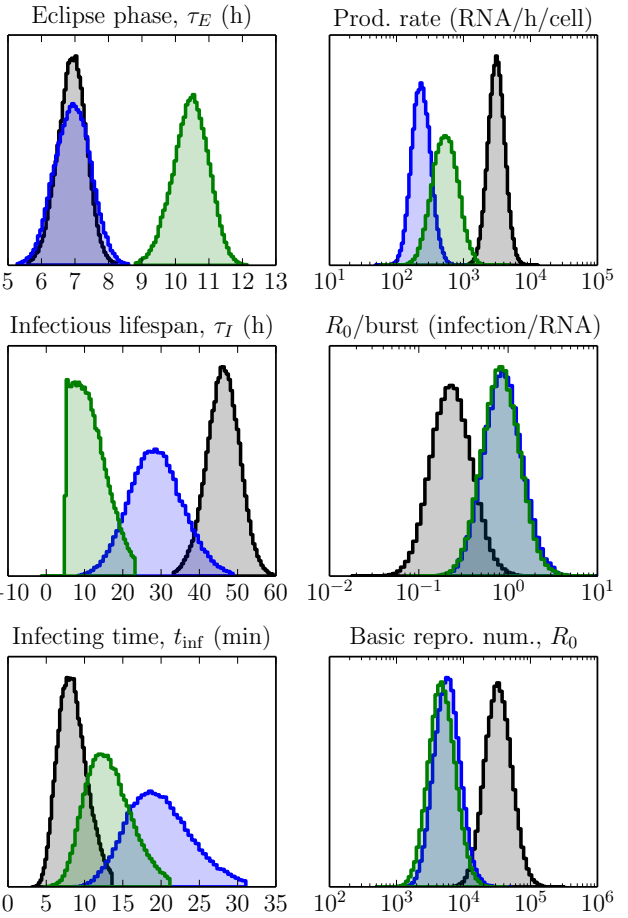


Yay! Let's do it again for a new WT and MUT pair... oups

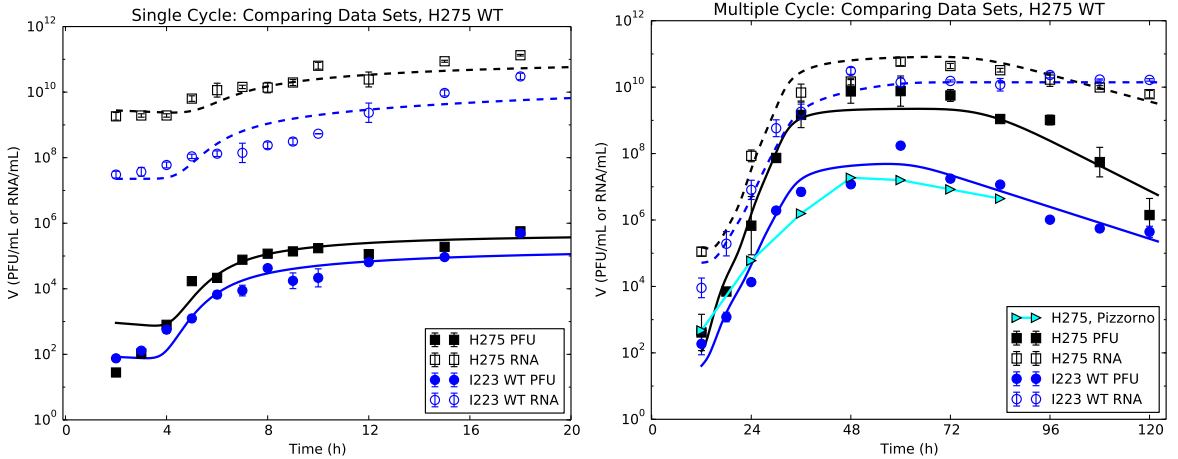
NEW WT params don't all match OLD ones







Problems come from experiment, not analysis



Our math analysis found that between **OLD** $\rightarrow$ **NEW** experiments:

A lower virus production rate (p) $\rightarrow$ lower peak virus in MC.

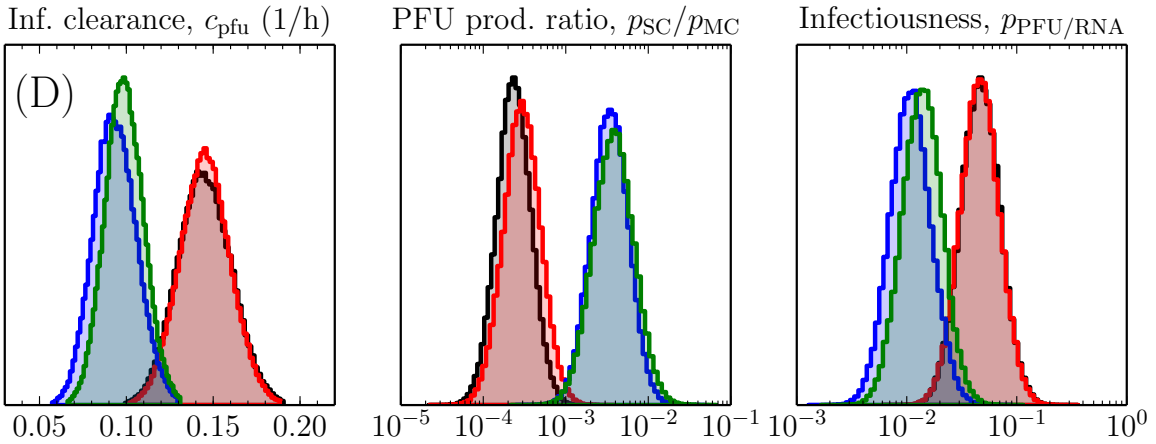
A shorter infectious lifespan (τ_I) $\rightarrow$ shorter virus plateau width in MC.

A similar eclipse phase length (τ_E) $\rightarrow$ same time of saturation in SC.

The parameter changes we found are all echoed in the data...

If the properties of a strain are experiment-specific, aren't experiments just producing random answers? Well... euh... yeah. But...

Parameters consistent within an experiment

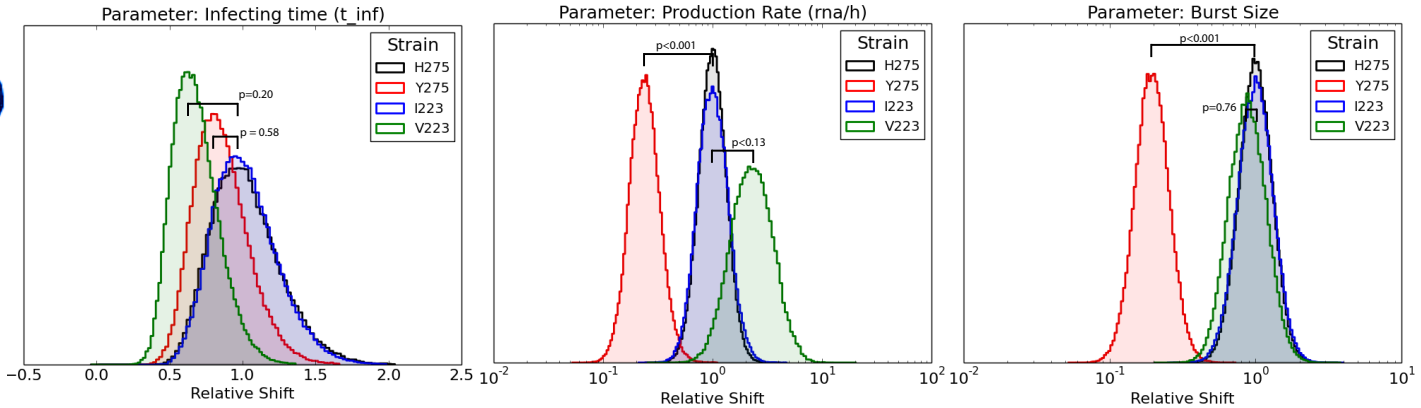


- While parameters do vary from experiment-to-experiment (WT $\neq$ WT)
- Parameters that are **not** strain specific are consistent and robust within a given experiment [(WT = MUT), (WT = MUT)].

Perhaps: the properties of one strain **relative** to another ($A = 3\times$standard strain behaviour), rather than the **absolute** properties ($A = 5$), are preserved between experiments.

Where can we go from here?

- Parameter estimation and predictions work well but...
- Inter-experimental variability is often greater than changes studied. We should:
  - Express parameters (strain properties) **relative to a reference strain**. Absolute parameter values are meaningless; and/or
  - Isolate main cause(s) of variability (e.g., cytokine-competency of cells, serum stock) and **account for it** (i.e., characterize action and incorporate into model).
- Either way, we'll need math models to do this.



Butting against a flawed institutional culture...

→ Quote from a reviewer (who rejected our paper):

There should be little to no inter-experimental variation, if proper techniques are used. Were two different people performing these experiments?

→ Translation:

variability in biology = bad/not trustworthy → reject paper.

proper techniques = same person, day, equipment → redefining variability = redefine results significance

Wait... WHAT?!

→ But if experimental variability is too heavily controlled (e.g., same person, same day, equipment, reagents), "experimental results" are unlikely to hold when **repeated**, and even less likely to hold in a person = useless!

Take home message...

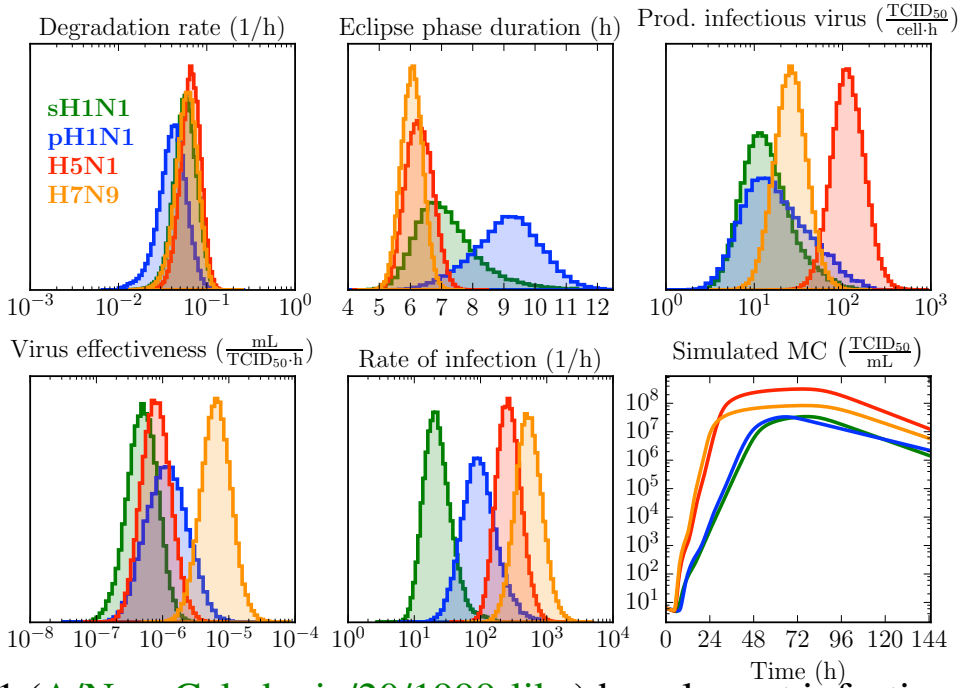
- Bad "science" in health research gives science a bad reputation!
- You should be skeptical of medical/health research.
- **Math description is required to tackle the issues**, i.e. address/study reproducibility/variability; quantify info burried in data.
- Solid medical results exist, e.g. many vaccines (mumps, rubella, etc.) and antibiotics can save your life, limb re-attachment and cast for broken bones are awesome!
- Messaging in health research must improve. E.g., not all vaccine are created equal (e.g., mumps vs influenza); Drs should communicate degree of uncertainty in treatment with patient and involve them in decision-making.

The END.

Lack of math: Side effects include dizziness and distrust

Catherine Beauchemin (**cbeau@ryerson.ca**)
[URL: **http://phymbie.physics.ryerson.ca/~cbeau**]

Another application: human flu vs bird flu



- Seasonal H1N1 (A/New Caledonia/20/1999-like) has slowest infection rate.
- Pandemic H1N1 (A/Mexico/INDRE4487/2009) has a faster rate of infection, but a much longer eclipse phase → similar to sH1N1.
- Avian (H5N1 and H7N9) have shorter eclipse phase and faster infection rate.
- For H5N1 (A/Indonesia/05/2005) it is due to higher virus production rate; for H7N9 (A/Anhui/1/2013) it is due to higher virus infectivity.

But this is a dangerous position to take...

Quotes from a reviewer:

Further, a significant portion of the manuscript examines the issue of inter-experimental variability. I find this to be a major limitation of this work since ***this type of variability should not exist if proper techniques are used****. In general, variability of this nature in biological systems makes it difficult to believe the results.*

There should be little to no inter-experimental variation, if proper techniques are used. ***Were two different people performing these experiments?*** *Was the virus stock the same? Were each identical in their sequences? These questions and others about the experimental setup (reagents, cell types, etc.) should be controlled for and detailed included in this manuscript.* ***It is unlikely that the variation is true or biologically interesting****. Repeated experiments with each virus are necessary to distinguish between poor experimental techniques and biological relevance.*

But if results are not robust (e.g., within error/variability, both $A > B$ and $A < B$ is possible) to the use of a different experimentor or batch of cells or reagents between 2 experiments, then is it really a result? And how can we expect it to translate **in vivo in a person**?