“Just Bring Me Everything”

Data Fusion in Mathematical Model Fitting and Parameterization

Eric Lofgren MSPH, PhD

Associate Professor, Paul G. Allen School for Global Health

Washington State University

Who Am I?

- PhD in Epidemiology from UNC Chapel Hill
- Postdoc at Network Dynamics and Simulation Science Lab, Virginia Tech
- Now an Associate Professor at the Paul G. Allen School for Global Health at Washington State University

Modeling is Inherently Data Fusion

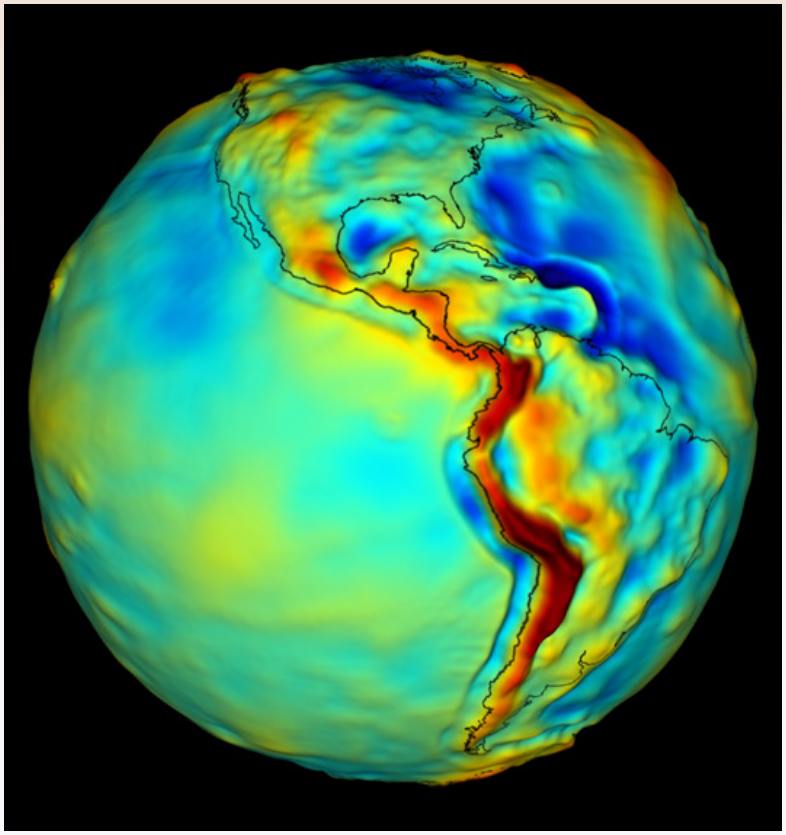
- Mathematical models are an attempt to describe how we think the world works
  - Admittedly with a degree of simplification
- This is inherently a data fusion process for several reasons
  - *Very* rarely has there been one data set so complete as to encapsulate *reality*
  - Even then it was probably on the wrong scale
  - No one is interested in *every* aspect of a process enough to collect data ranging from pharmacokinetics to medical staff time use
  - Good luck fitting all of that into three Specific Aims for your R01

An Example

- There are no studies of what happens when I drop an egg in the Mt. Hood room of the Portland Downtown Marriott
  - I would hazard a guess that the “Dropping Eggs from Speaking Podiums” literature generally is fairly sparse
- I am willing to bet however that none of you think this is a particularly good idea
- Why?
  - Because you all have a mental model, built from data fusion of your lived experiences, or the shared lived experiences of those you know
  - Eggs are fragile, hotel conference room floors usually don’t feature particularly deep carpet, nothing good has ever come from “I’m going to drop an egg…”, etc.

Models are Formalized Data Fusion

- Mathematical models are simply formalized version of this same process
- We can measure the height from my outstretched hand to the floor, the acceleration force to to gravity at this place on Earth, the resulting force on impact, the force an eggshell can withstand, etc.
- We may make some assumptions about the world
  - Eggs are perfect spheres, wind resistance is minimal, no one is going to make a heroic dive to catch the egg, etc.



Why This is Useful

- We can leverage knowledge from different fields, settings, etc. to provide quantitative estimates (with uncertainty estimates) about pragmatic questions
- Our data fusion process can be enhanced over time as new data comes out (a carpet manufacturer does a study on impact absorption of different products, trying to sell more carpets to owners of expensive smart phones…)
- We can create estimations through a data fusion process for *things that don't yet exist*

Fusing Fiction

- This same data fusion process can, in a modeling framework, credibly handle hypothetical data for which we have no studies
- What if Yao Ming were to drop the egg?
- What if we were to drop an egg in a conference room on Venus?
- What if was a *genetically engineered* egg with a stronger shell?
- What if some heroic SER member did make a diving save?

Where This is Dangerous

- Your assumptions are wrong
  - You're modeling a world that is *mostly* like ours
  - Sometimes this matters
  - This is, in essence, an external validity problem
    - Hot take: Mathematical models are *perfectly* internally valid in a way RCTs can only manage a pale imitation of
  - This can be thought about in two large categories
    - Parameter sensitivity (you got a specific parameter value wrong)
    - Structural sensitivity (you misspecified how the world works)
    - Both are a threat, the former gets talked about a *lot* more

Practical Examples

Eric, please stop talking about dropping eggs, you're making everyone nervous.

Hospital Epidemiology

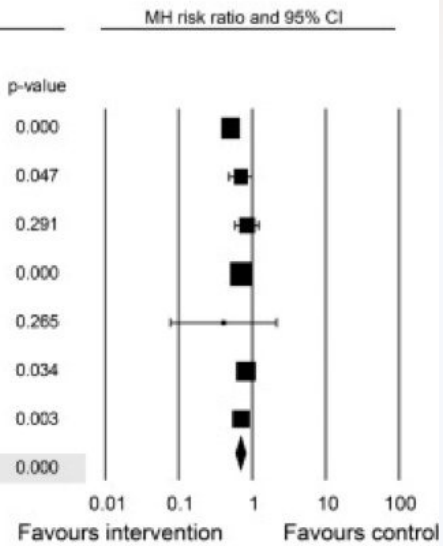
- We'd like to know the *per-use* efficacy of Chlorhexidine (CHG) in decolonizing patients with Methicillin-resistant *Staphylococcus aureus*
- This is important for studying CHG resistance, trying to figure out effect estimates from studies with multiple interventions, etc.
- But this hasn't been widely estimated
  - It's a pain
  - Doing it right would get in the way of the patient-care mission
  - It's implemented in the real world as a ward- or hospital-level intervention

What We Do Have

- Close to 390,000 patients from cluster randomized trials of CHG

Study	Statistics for each study					
	Control		Intervention		Risk ratio	
	Events	Total	Events	Total	M-H, random, [95% CI]	p-value
Gould, I. M. 2007 [30]	193	6899	92	6664	0.493 [0.386, 0.631]	0.000
Climo, M. W. 2009 [25]	67	13300	45	13096	0.682 [0.468, 0.994]	0.047
Climo, M. W. 2013 [3]	58	24983	47	24902	0.813 [0.554, 1.194]	0.291
Huang, S. S. 2013 [31]	279	88222	217	101603	0.675 [0.566, 0.806]	0.000
Camus, C. 2014 [32]	5	1961	2	1991	0.394 [0.077, 2.028]	0.265
Viray, M. A. 2014 [33]	180	14279	164	16327	0.797 [0.646, 0.983]	0.034
Petlin, A. 2014 [34]	132	34333	109	41376	0.685 [0.532, 0.883]	0.003
Subtotal	914	183977	676	205959	0.673[0.586, 0.774]	0.000

Test for heterogeneity: τ^2=0.013, I^2=39.70, P = .127
Test for overall effect: z=-5.563, P < .001



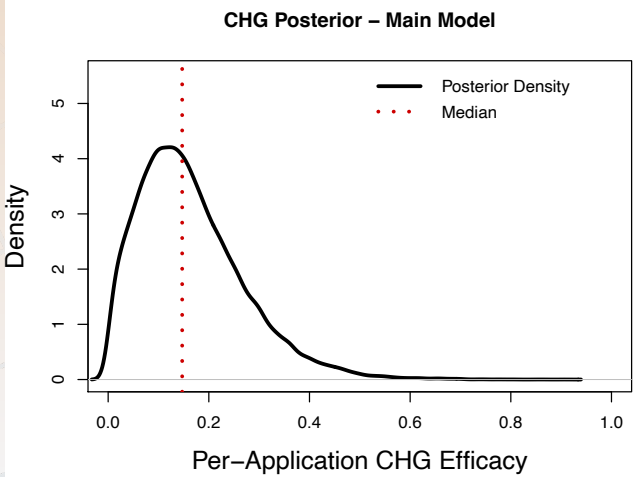
Kim, HY *et al.* Journal of Critical Care, 2016

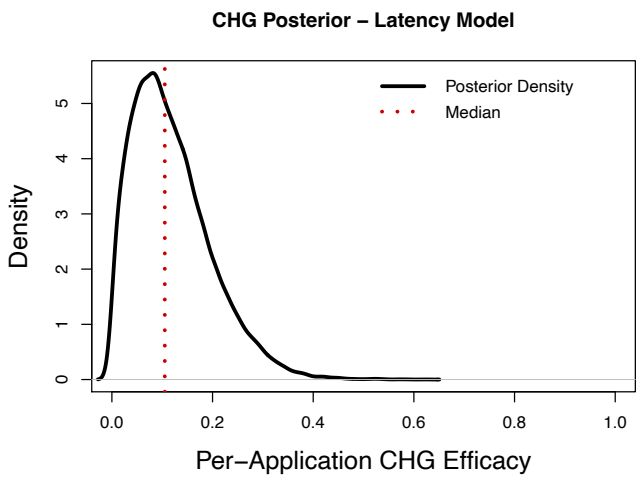
What We Also Have

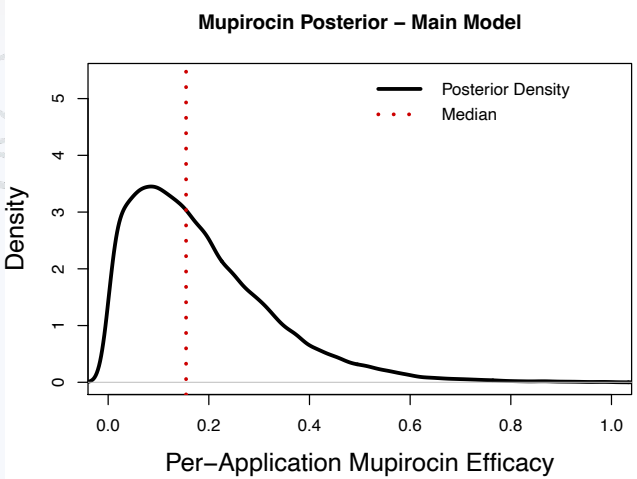
- Some assumptions!
- The relative risk in this context approximates an incidence rate ratio
  - I know that's backwards from how we normally think about it
- That a mix of RCTs and observational studies can both estimate the effect of CHG on MRSA acquisition
  - If you don't think this is true, you may be in the wrong conference…

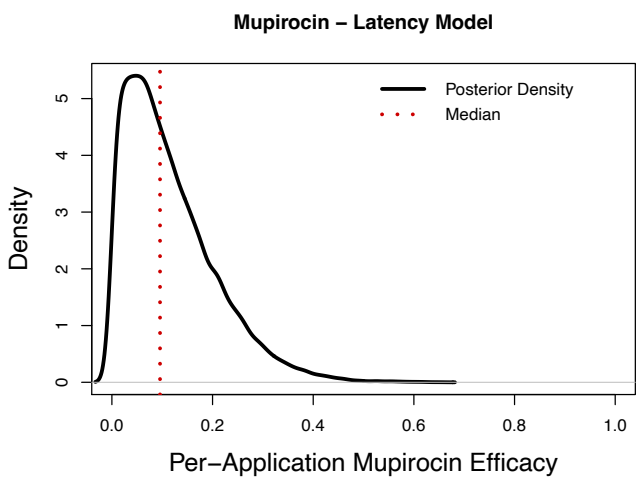
Let the Fusion Begin

- Add a decolonization process to an existing model of MRSA colonization
- Use Approximate Bayesian Computation to ask what value for per-use decolonization, when put into the model, gives us back the estimates from the meta-analysis
- That's our estimate of interest!


CHG Posterior – Main Model
Posterior Density
Median
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Per–Application CHG Efficacy


CHG Posterior – Latency Model
Posterior Density
Median
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Per–Application CHG Efficacy

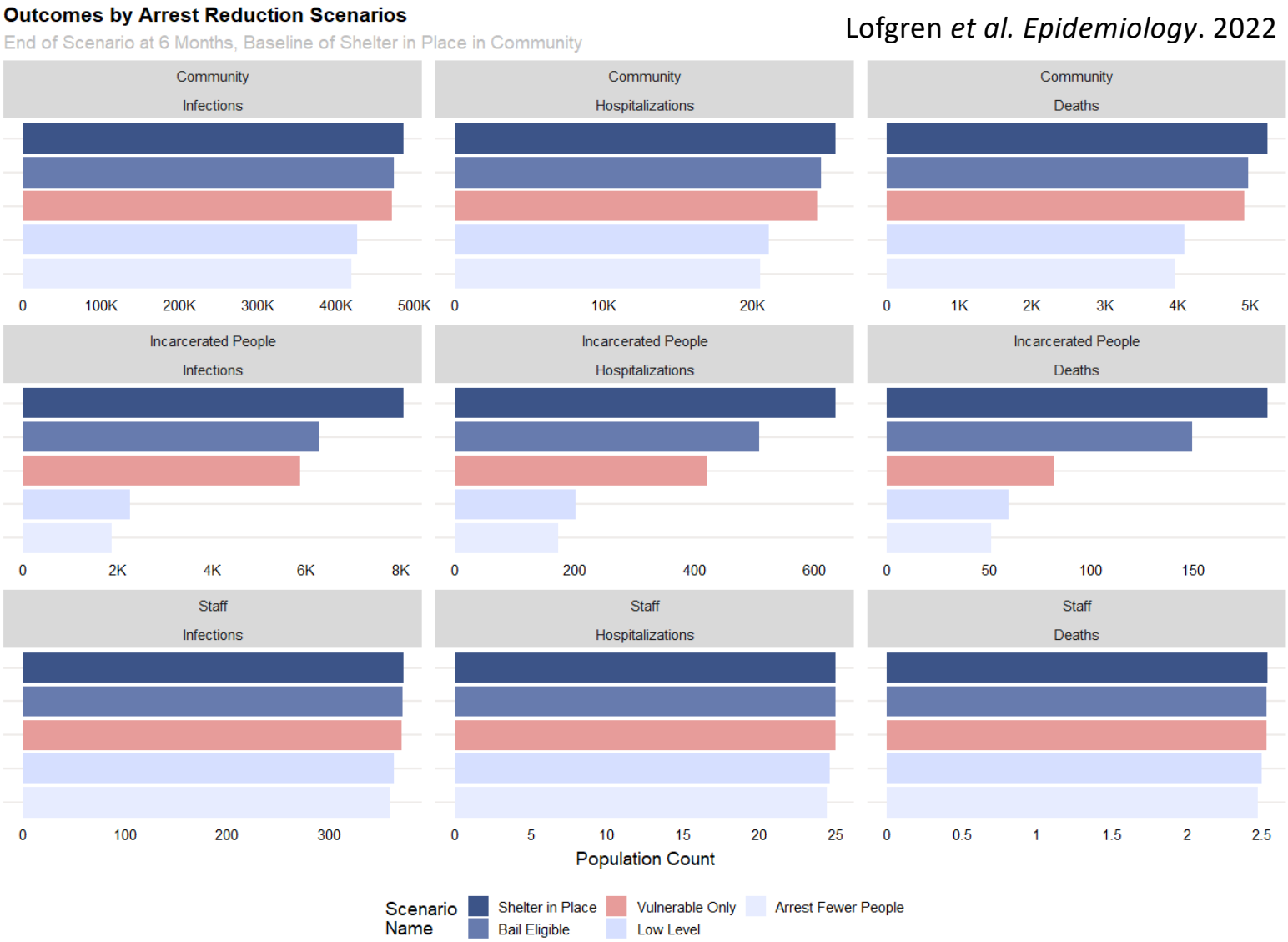

Mupirocin Posterior – Main Model
Posterior Density
Median
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Per–Application Mupirocin Efficacy


Mupirocin – Latency Model
Posterior Density
Median
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Per–Application Mupirocin Efficacy

COVID-19 and Jails

- Worked with the ACLU to look at hypothetical interventions
- A data fusion process between what we know about the justice system and what we know about COVID
- *Wildly* impractical to have an observational study about, and especially not an RCT
- In this example, focusing on a very practical intervention – releasing folks from jail to home to await trial

Lofgren *et al. Epidemiology*. 2022



The Limits of Fusion

- Knowing the limits of modeling
  - The jail model gives a broad strokes answer for a county *like* Allegheny Co, Pennsylvania, for a jail without much structure to it, in Spring of 2020 where the surrounding community is sheltering in place
  - Huge demand for very specific forecasts, which we said the model couldn't do, especially not for very specific settings
- There needs to be more emphasis on *qualitative* results from models
  - "Have any of these results convinced you you that HAART is a bad idea?" – Steve Cole

The Value of Theory

- One of the benefits of looking at data fusion from a modeling lens is that we can pull information from *many* different sources and many different fields
- I'd assert that modeling is more comfortable with this because we draw a lot from fields that are more comfortable with general theory – notably Ecology
- There are of course limits to this
  - Is an ICU in Canada a good source for a model of an American ICU? What about one in Japan? Kenya?
- There is however a lot of value in this as well, especially when making decisions in times of uncertainty, or without clean data
  - "We don't need an RCT, we have math" – A paraphrased engineer, on Evidence Based Bridge Building

Thank You and Funding

Lofgren Lab

The Paul G. Allen School for Global Health

Washington State University Resistance Epidemiology Modeling Initiative

- Contact Information:
  - Eric.Lofgren@wsu.edu
  - Twitter: @GermsAndNumbers
- Funding:
  - NIH R35GM147013 and CDC U01CK000673
- We're hiring!
  - Actively recruiting *at least* one postdoc and graduate students
  - Staff position in data management to work on some very cool global health projects
  - Interested in helping lead and develop an undergraduate program in public health? Come talk to me