Class slides for Thursday, November 5: Disease empirics

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University

TheUpshot

What Trump Needs to Win: A Polling Error Much Bigger Than 2016's

Several factors that led to the misfire last time are no longer in play.

By Nate Cohn

President Trump needs a very large polling error to have a hope of winning the White House. Joe Biden would win even if polls were off by as much as they were in 2016.

Polling averages as of 10 p.m. on Nov. 1, 2020

	POLLING LEADER	IF POLLS ARE AS WRONG AS THEY WERE IN... 2016	2012
U.S.	+9 **Biden**	+7	+12
N.H.	+11 **Biden**	+8	+15
Wis.	+10 **Biden**	+4	+14
Minn.	+10 **Biden**	+4	+12
Mich.	+8 **Biden**	+4	+14
Nev.	+6 **Biden**	+8	+9
Pa.	+6 **Biden**	+1	+7
Neb. 2*	+5 **Biden**	+9	<1
Maine 2*	+4 **Biden**	+9	+9
Ariz.	+4 **Biden**	+2	+2
Fla.	+2 **Biden**	<1	+4
N.C.	+2 **Biden**	+3	+3
Ga.	+2 **Biden**	<1	+2
Ohio	<1 **Trump**	+6	<1
Iowa	+2 **Trump**	+6	+3
Texas	+2 **Trump**	+4	+1

Electoral votes if polling leads translate perfectly to results (they won't):

	TOTALS BASED ON 2020 POLLS	IF POLLS ARE AS WRONG AS THEY WERE IN...	
Electoral votes	351 **Biden** ✓	335 ✓	412 ✓

President Trump needs a very large polling error to have a hope of winning the White House. Joe Biden would win even if polls were off by as much as they were in 2016.

Polling averages as of 10 p.m. on Nov. 1, 2020

	POLLING LEADER	IF POLLS ARE AS WRONG AS THEY WERE IN... 2016	2012
U.S.	+9 **Biden**	+7	+12
N.H.	+11 **Biden**	+8	+15
Wis.	+10 **Biden**	+4	+14
Minn.	+10 **Biden**	+4	+12
Mich.	+8 **Biden**	+4	+14
Nev.	+6 **Biden**	+8	+9
Pa.	+6 **Biden**	+1	+7
Neb. 2*	+5 **Biden**	+9	<1
Maine 2*	+4 **Biden**	+9	+9
Ariz.	+4 **Biden**	+2	+2
Fla.	+2 **Biden**	<1	+4
N.C.	+2 **Biden**	+3	+3
Ga.	+2 **Biden**	<1	+2
Ohio	<1 **Trump**	+6	<1
Iowa	+2 **Trump**	+6	+3
Texas	+2 **Trump**	+4	+1

Electoral votes if polling leads translate perfectly to results (they won't):

	TOTALS BASED ON 2020 POLLS	IF POLLS ARE AS WRONG AS THEY WERE IN...	
Electoral votes	351 **Biden** ✓	335 ✓	412 ✓

Nate Silver: How likely is a big polling error? We don't have much data. 1948 was biggest poll error on record, but that was different; 2020 is different from 1992.

https://www.vox.com/21538214/nate-silver-538-2020-forecast-2016-trump-biden-election-podcast

Limits to Prediction

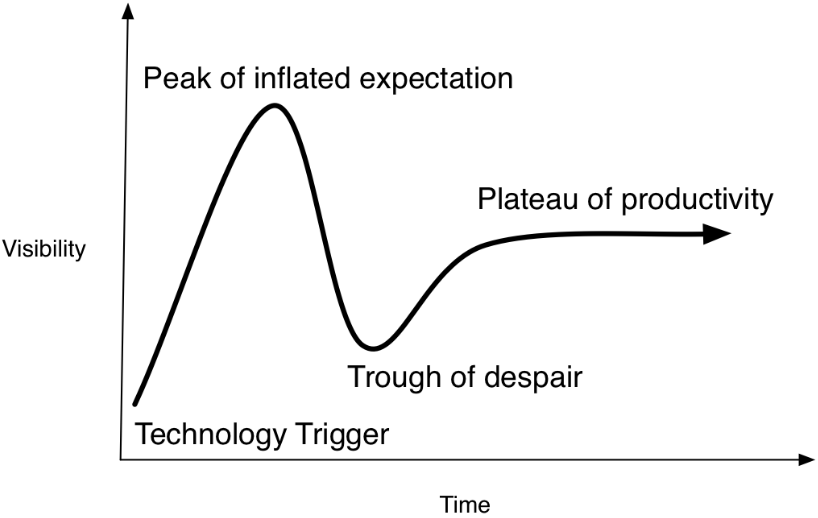
With enough data and the right algorithms does everything become predictable?

- COS 597E/SOC 555, Princeton University, Fall 2020
- Arvind Narayanan and Matthew Salganik
- Tuesday and Thursday, 11:00am - 12:20pm
- GitHub, Piazza, Canvas

Class slides for Thursday, November 5: Disease empirics

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University



https://en.wikipedia.org/wiki/Hype_cycle

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

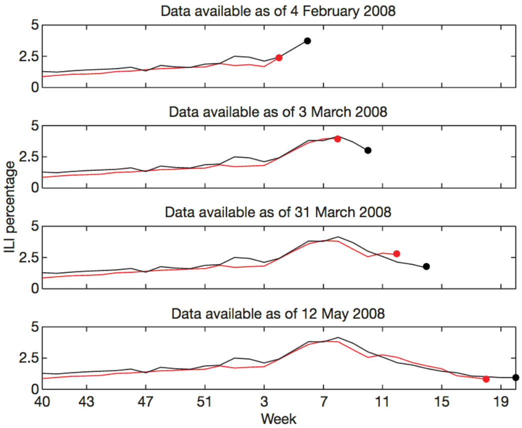


Figure 3 | ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season. During week 5 we detected a sharply increasing ILI percentage in the mid-Atlantic region; similarly, on 3 March our model indicated that the peak ILI percentage had been reached during week 8, with sharp declines in weeks 9 and 10. Both results were later confirmed by CDC ILI data.

- Trade-off between manual vs automatic selection of query terms

- Trade-off between manual vs automatic selection of query terms
- Very simple model $logit(I(t)) = \alpha logit(Q(t)) + \epsilon$ (where is the magic coming from?)

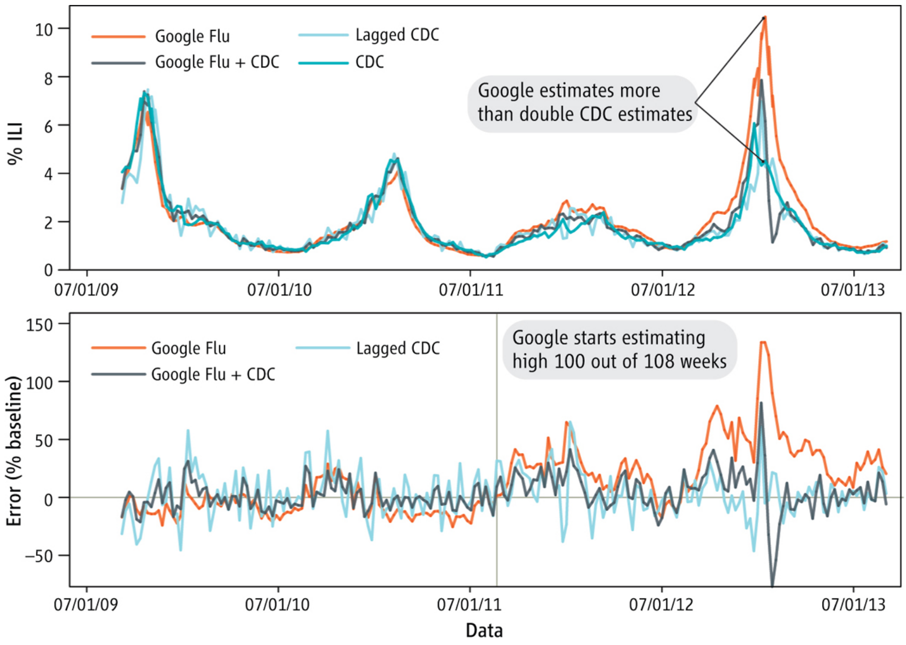
- Trade-off between manual vs automatic selection of query terms
- Very simple model $logit(I(t)) = \alpha logit(Q(t)) + \epsilon$ (where is the magic coming from?)
- Performance metric: correlation between CDC and estimates

- Trade-off between manual vs automatic selection of query terms
- Very simple model $logit(I(t)) = \alpha logit(Q(t)) + \epsilon$ (where is the magic coming from?)
- Performance metric: correlation between CDC and estimates
- Using digital trace data to measure/predict other things

Lots of hype and a bit of criticism

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[5,6,3]



Google Flu
Lagged CDC
Google Flu + CDC
CDC
Google estimates more than double CDC estimates
% ILI
0
2
4
6
8
10
07/01/09
07/01/10
07/01/11
07/01/12
07/01/13
Google Flu
Lagged CDC
Google Flu + CDC
Google starts estimating high 100 out of 108 weeks
Error (% baseline)
150
100
50
0
−50
07/01/09
07/01/10
07/01/11
07/01/12
07/01/13
Data

Google

Thank you for stopping by.

Google Flu Trends and Google Dengue Trends are no longer publishing current estimates of Flu and Dengue fever based on search patterns. The historic estimates produced by Google Flu Trends and Google Dengue Trends are available below. It is still early days for nowcasting and similar tools for understanding the spread of diseases like flu and dengue – we're excited to see what comes next. Academic research groups interested in working with us should fill out this form.

Sincerely,

The Google Flu and Dengue Trends Team.

Nice history: https://www.theatlantic.com/technology/archive/2014/03/in-defense-of-google-flu-trends/359688/

Three lessons from the parable

- Transparency and replicability

Three lessons from the parable

- Transparency and replicability
- Use big data to understand the unknown (What can Google Flu Trends do beyond speeding up an existing system? Small area estimation)

Three lessons from the parable

- Transparency and replicability
- Use big data to understand the unknown (What can Google Flu Trends do beyond speeding up an existing system? Small area estimation)
- It is not just about the size of the data

Three lessons from the parable

- Transparency and replicability
- Use big data to understand the unknown (What can Google Flu Trends do beyond speeding up an existing system? Small area estimation)
- It is not just about the size of the data

Other issue: accountability for Google

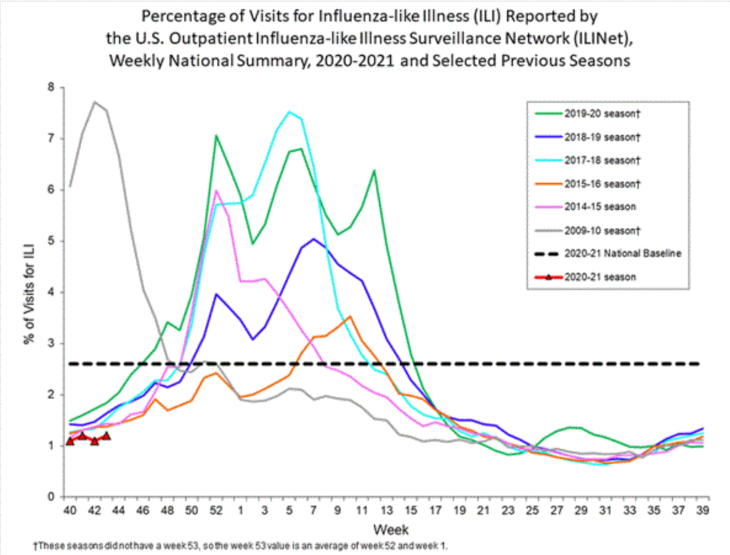
A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States

Nicholas G. Reich[a,1], Logan C. Brooks[b], Spencer J. Fox[c], Sasikiran Kandula[d], Craig J. McGowan[e], Evan Moore[a], Dave Osthus[f], Evan L. Ray[g], Abhinav Tushar[a], Teresa K. Yamana[d], Matthew Biggerstaff[e], Michael A. Johansson[h], Roni Rosenfeld[i], and Jeffrey Shaman[d]

- Outcomes related to: % of patient visits reported through the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet) were due to influenza-like illness (ILI).

- Outcomes related to: % of patient visits reported through the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet) were due to influenza-like illness (ILI).
- Target outcomes selected by the CDC, not the researchers

- Outcomes related to: % of patient visits reported through the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet) were due to influenza-like illness (ILI).
- Target outcomes selected by the CDC, not the researchers
- Composite outcome

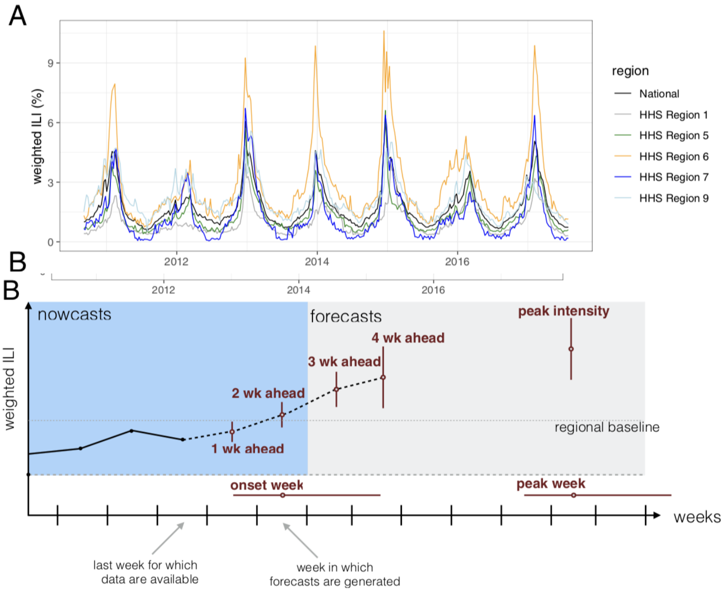


https://www.cdc.gov/flu/weekly/index.htm#ILINet

H7: Ill-conceived target variable

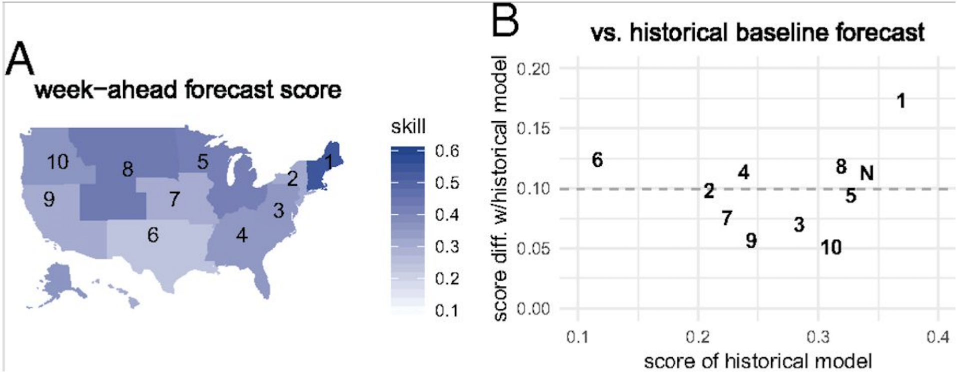
Some target variables will be hard to predict because they are poorly measured or unstable. For example, imagine trying to predict a student's performance on a math test. As the number of questions on the test increases, the measurement of math performance should happen with less error (assuming that it is a well-designed test). However, we also hypothesize that there will be diminishing returns to improved measurement and that predictive performance will plateau well below perfection.

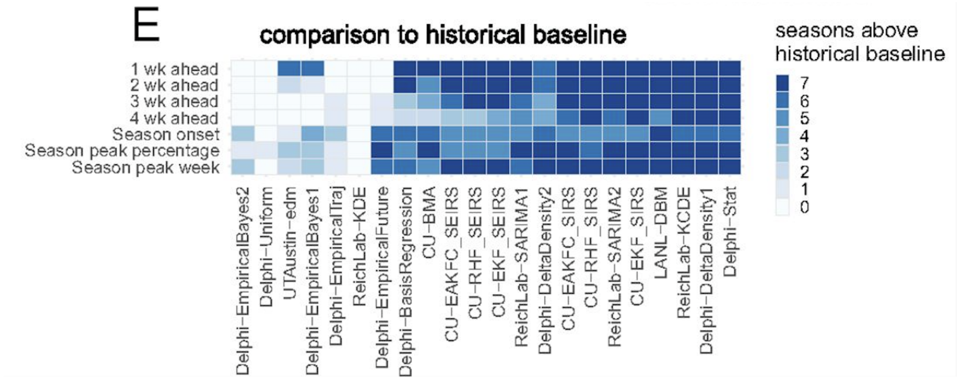
Please put other composite outcomes in the chat


A
weighted ILI (%)
0
3
6
9
2012
2014
2016
region
National
HHS Region 1
HHS Region 5
HHS Region 6
HHS Region 7
HHS Region 9
B
weighted ILI
nowcasts
forecasts
peak intensity
4 wk ahead
3 wk ahead
2 wk ahead
1 wk ahead
regional baseline
onset week
peak week
weeks
last week for which data are available
week in which forecasts are generated

Predictability for a target can be broken into two components: 1) baseline from historical average 2) improvement over baseline. (prediction vs skillful prediction)

Predictability for a target can be broken into two components: 1) baseline from historical average 2) improvement over baseline. (prediction vs skillful prediction)





Doing better than historical baseline is not easy

How does forecast score decay as time window increases?

How does forecast score decay as time window increases?
`CU-EKF_SIRS` (best): 0.55 (1 week ahead), 0.44 (2 weeks ahead), 0.36 (3 weeks ahead), 0.31 (4 weeks ahead)

Scores were lower for seasonal targets than week-ahead targets, although model showed greater relative improvement compared to baseline. In other words, it the details of the outcome matter.

Target	Score: Statistical model	Compartmental model	Difference
1 wk ahead	0.49	0.43	0.06
2 wk ahead	0.40	0.41	−0.01
3 wk ahead	0.35	0.34	0.00
4 wk ahead	0.32	0.30	0.02
Season onset	0.23	0.22	0.01
Season peak percentage	0.32	0.27	0.05
Season peak week	0.34	0.32	0.02

The difference column represents the difference in the average probability assigned to the eventual outcome for the target in each row. Positive values indicate the top statistical models showed higher average score than the top compartmental models.

- Compartmental model

	Score		
Target	Statistical model	Compartmental model	Difference
1 wk ahead	0.49	0.43	0.06
2 wk ahead	0.40	0.41	−0.01
3 wk ahead	0.35	0.34	0.00
4 wk ahead	0.32	0.30	0.02
Season onset	0.23	0.22	0.01
Season peak percentage	0.32	0.27	0.05
Season peak week	0.34	0.32	0.02

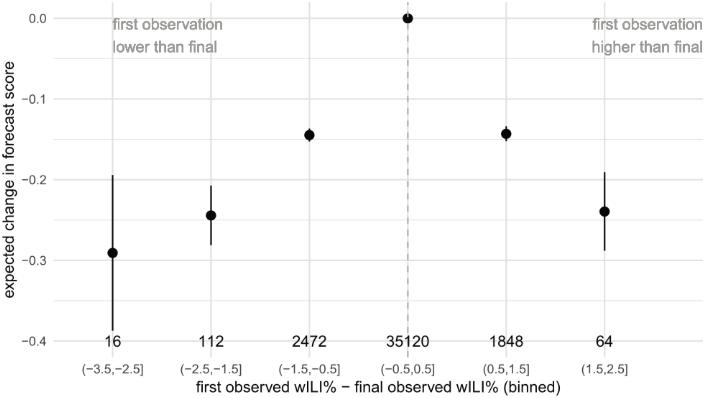
The difference column represents the difference in the average probability assigned to the eventual outcome for the target in each row. Positive values indicate the top statistical models showed higher average score than the top compartmental models.

- Compartmental model
- Value of the multiple analyst design

	Score		
Target	Statistical model	Compartmental model	Difference
1 wk ahead	0.49	0.43	0.06
2 wk ahead	0.40	0.41	−0.01
3 wk ahead	0.35	0.34	0.00
4 wk ahead	0.32	0.30	0.02
Season onset	0.23	0.22	0.01
Season peak percentage	0.32	0.27	0.05
Season peak week	0.34	0.32	0.02

The difference column represents the difference in the average probability assigned to the eventual outcome for the target in each row. Positive values indicate the top statistical models showed higher average score than the top compartmental models.

- Compartmental model
- Value of the multiple analyst design
- Can we quantify the value of the multiple analyst design?



Bad initial measurements hurt performance. Measuring the present helps us predict the future

On the multibin logarithmic score used in the FluSight competitions

Johannes Bracher[a,1]

- proper scoring rules encourage accurate reporting, but how important is that here?

On the multibin logarithmic score used in the FluSight competitions

Johannes Bracher[a,1]

- proper scoring rules encourage accurate reporting, but how important is that here?
- important question for the design of prediction tasks

Box 1. Toward a principled understanding of infectious disease forecasts: Key questions

1. How do prediction skills decrease with time horizon? Are there inherent limits to skilled predictions (currently 2 to 4 wk)?
2. How do prediction skills scale with data accuracy and quantity? What is the value added by different datasets and disease parameters (age-specific surveillance data, digital data, social behavior, vaccination, strain composition, background population immunity)? How long and how granular does each dataset need to be?
3. How should ensemble predictions be optimized? How many models? How many approaches? How should the weights of each approach be optimized?
4. What is the right spatial scale for transmission, and hence for forecasts? How does it relate to the spatial scale of disease surveillance and public health interventions?
5. The above questions should be studied or simulated across a variety of pathogens, populations, and data situations.

Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast

Kelly R. Moran,[1] Geoffrey Fairchild,[1] Nicholas Generous,[1] Kyle Hickmann,[2] Dave Osthus,[3] Reid Priedhorsky,[4] James Hyman,[2,5] and Sara Y. Del Valle[1]

[1]Analytics, Intelligence, and Technology Division, [2]Theoretical Division, [3]Computer, Computational & Statistical Sciences Division, [4]High Performance Computing Division, Los Alamos National Laboratory, New Mexico; and [5]Department of Mathematics, Tulane University, New Orleans, Louisiana

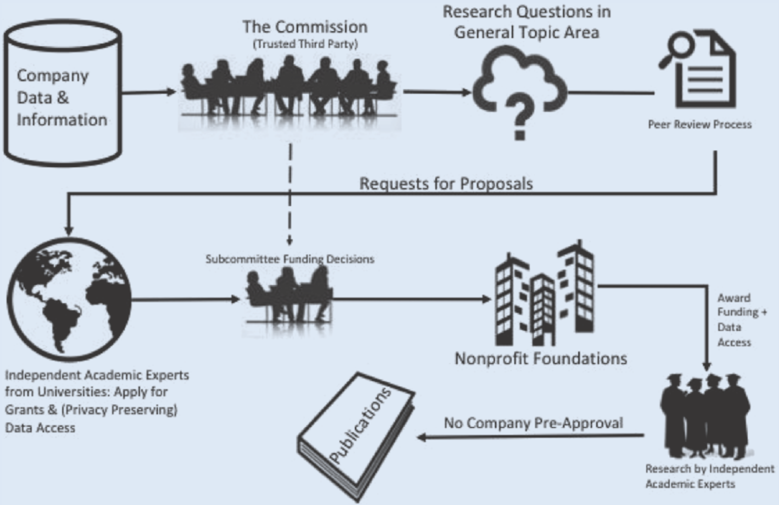
Data access

A New Model for Industry–Academic Partnerships

Gary King, *Harvard University*
Nathaniel Persily, *Stanford University*

Figure 1

Outline of Industry–Academic Partnership Model



My 3 takeaways:

- Data access as a limit to prediction

My 3 takeaways:

- Data access as a limit to prediction
- Composite outcomes as a limit to prediction (prevents isolation of one system)

My 3 takeaways:

- Data access as a limit to prediction
- Composite outcomes as a limit to prediction (prevents isolation of one system)
- How much data do we really have?

Split into groups

	Weather	Disease
Knowledge of fund laws		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms
Adjust to phenomena		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms
Adjust to phenomena	None	Some from people, gov, firms

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms
Adjust to phenomena	None	Some from people, gov, firms
Goal		

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms
Adjust to phenomena	None	Some from people, gov, firms
Goal	Prediction for accommodation	Prediction for intervention

	Weather	Disease
Knowledge of fund laws	high	high for medical, low for social
Global cooperation	Possible	creepy (& potentially dangerous)
Openness of inp, pred & outs	Public & standardized	problems from gov (interference) & firms (inaccessible, bad measurement)
Reactivity to predictions	Low	Some from people, gov, firms
Adjust to phenomena	None	Some from people, gov, firms
Goal	Prediction for accommodation	Prediction for intervention

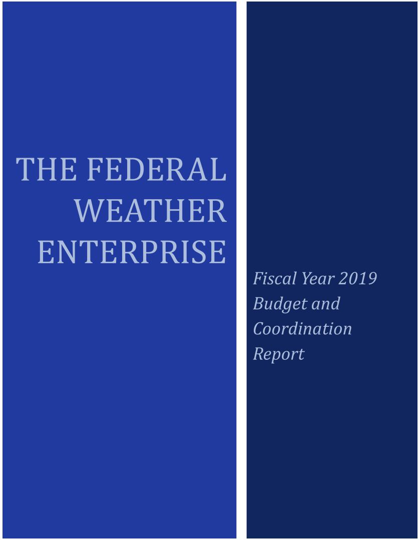
Provocation: For weather forecasting fundamental limits come from nature, and for disease forecasting fundamental limits comes from social organization

A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States

Nicholas G. Reich[a,1], Logan C. Brooks[b], Spencer J. Fox[c], Sasikiran Kandula[d], Craig J. McGowan[e], Evan Moore[a], Dave Osthus[f], Evan L. Ray[g], Abhinav Tushar[a], Teresa K. Yamana[d], Matthew Biggerstaff[e], Michael A. Johansson[h], Roni Rosenfeld[i], and Jeffrey Shaman[d]

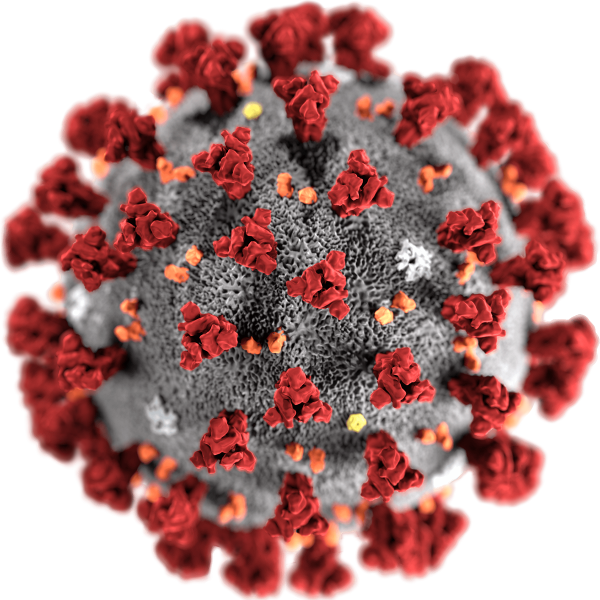
A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States

Nicholas G. Reich[a,1], Logan C. Brooks[b], Spencer J. Fox[c], Sasikiran Kandula[d], Craig J. McGowan[e], Evan Moore[a], Dave Osthus[f], Evan L. Ray[g], Abhinav Tushar[a], Teresa K. Yamana[d], Matthew Biggerstaff[e], Michael A. Johansson[h], Roni Rosenfeld[i], and Jeffrey Shaman[d]



Total budget about 4 billion USD

https://www.ofcm.gov/publications/



https://phil.cdc.gov/Details.aspx?pid=23312

Next steps: Predictability of life trajectories

- Tuesday: Fragile Families Challenge (little reading)
- Thursday: "Dark matter" interviews (lots of reading, potentially emotionally difficult)

Send me your completed and signed data agreement by today (Thursday) at 5pm

Class slides for Thursday, November 5: Disease empirics

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University