Challenges of Epidemiological Forecasting in a Pandemic

Ryan Tibshirani

Depts. of Statistics & Machine Learning
Carnegie Mellon University

DELPHI

CDC (Center of Excellence)

Amazon Scholar (AWS AI)
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
- Since 2013-14, CDC has held forecasting challenge each year
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
- Since 2013-14, CDC has held forecasting challenge each year (about 40 systems participating in recent years)
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
- Since 2013-14, CDC has held forecasting challenge each year (about 40 systems participating in recent years)
- Why? Reliable forecasts could help them with preparedness and countermeasures
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
- Since 2013-14, CDC has held forecasting challenge each year (about 40 systems participating in recent years)
- Why? Reliable forecasts could help them with preparedness and countermeasures
- Delphi had the highest accuracy forecasts every year except last (second highest)
Flu background

- **Seasonal influenza** is associated with 250,000–500,000 deaths each year (WHO estimates)
- About 80,000 flu-related deaths in US in 2017-18 season (CDC)
- In 2013, Roni Rosenfeld and I started Delphi group to model and forecast seasonal flu
- Since 2013-14, CDC has held forecasting challenge each year (about 40 systems participating in recent years)
- Why? Reliable forecasts could help them with **preparedness and countermeasures**
- Delphi had the **highest accuracy** forecasts every year except last (second highest)
- We were awarded **CDC Center of Excellence** in September 2019
Flu prevalence data

Most basic target of interest: % ILI = influenza-like illness (fever of 100 or higher, along with sore throat or cough), reported by network of hospitals and doctors in the US
Flu prevalence data

Most basic target of interest: % ILI = influenza-like illness (fever of 100 or higher, along with sore throat or cough), reported by network of hospitals and doctors in the US

Why? Believed to be leading indicator for flu burden
Flu prevalence data

Most basic target of interest: $\% \text{ILI} = \text{influenza-like illness}$ (fever of 100 or higher, along with sore throat or cough), reported by network of hospitals and doctors in the US

Why? Believed to be leading indicator for flu burden
Flu prevalence data

Most basic target of interest: $\% \text{ILI} = \text{influenza-like illness}$ (fever of 100 or higher, along with sore throat or cough), reported by network of hospitals and doctors in the US

Why? Believed to be leading indicator for flu burden

- CDC releases state ILI data weekly: but always one-week-old
Flu prevalence data

Most basic target of interest: $\% \text{ILI} = \text{influenza-like illness}$ (fever of 100 or higher, along with sore throat or cough), reported by network of hospitals and doctors in the US

Why? Believed to be leading indicator for flu burden

- CDC releases state ILI data weekly: but always one-week-old
- Data is subject to revisions, sometimes significant ones
Severity pyramid

- **Deaths**: 12,000 - 61,000*
- **Hospitalizations**: 140,000 - 810,000*
- **Illnesses**: 9,300,000 – 45,000,000*
How to forecast a seasonal epidemic

Before how, what:
How to forecast a seasonal epidemic

Before how, what:

- Short-term forecasts: % ILI for each of the next 4 weeks
How to forecast a seasonal epidemic

Before how, what:

- **Short-term** forecasts: % ILI for each of the next 4 weeks
- **Long-term** forecasts: peak week, and peak % ILI, and duration
How to forecast a seasonal epidemic

Before how, what:

- **Short-term forecasts**: % ILI for each of the next 4 weeks
- **Long-term forecasts**: peak week, and peak % ILI, and duration
- Predictions repeated each week, as new data comes in
How to forecast a seasonal epidemic

Before how, what:

- **Short-term forecasts:** % ILI for each of the next 4 weeks
- **Long-term forecasts:** peak week, and peak % ILI, and duration
- **Predictions repeated** each week, as new data comes in

Same for hospitalizations
How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season
How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of how:
How to forecast a seasonal epidemic

Before how, **what**: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of **how**:

1. Choose a type of model: *mechanistic* or *statistical*
How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of how:

1. Choose a type of model: mechanistic or statistical
2. For statistical models, collect lots of data:
How to forecast a seasonal epidemic

Before how, **what**: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of **how**:

1. Choose a type of model: mechanistic or statistical
2. For statistical models, collect **lots of data**: proxies like Google search queries, Wikipedia page hits, thermometer sales, etc.
How to forecast a seasonal epidemic

Before how, **what**: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of **how**:

1. Choose a type of model: *mechanistic* or *statistical*

2. For statistical models, collect **lots of data**: proxies like Google search queries, Wikipedia page hits, thermometer sales, etc.

3. For statistical models, rely on long history of training data to learn predictive relationships (carefully handle *nonstationarity*)
How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of how:

1. Choose a type of model: mechanistic or statistical
2. For statistical models, collect lots of data: proxies like Google search queries, Wikipedia page hits, thermometer sales, etc.
3. For statistical models, rely on long history of training data to learn predictive relationships (carefully handle nonstationarity)
4. For either type, should model revisions (backcasting!)
How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of how:

1. Choose a type of model: mechanistic or statistical
2. For statistical models, collect lots of data: proxies like Google search queries, Wikipedia page hits, thermometer sales, etc.
3. For statistical models, rely on long history of training data to learn predictive relationships (carefully handle nonstationarity)
4. For either type, should model revisions (backcasting!)
5. For either type, probabilistic forecasts are key; as are ensembles

Our system is an ensemble of 6 separate forecasting models


How to forecast a seasonal epidemic

Before how, what: short- and long-term characteristics of % ILI (also hospitalizations), for rest of season

Now, key elements of how:

1. Choose a type of model: mechanistic or statistical
2. For statistical models, collect lots of data: proxies like Google search queries, Wikipedia page hits, thermometer sales, etc.
3. For statistical models, rely on long history of training data to learn predictive relationships (carefully handle nonstationarity)
4. For either type, should model revisions (backcasting!)
5. For either type, probabilistic forecasts are key; as are ensembles

Our system is an ensemble of 6 separate forecasting models
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

![Graph showing National ILI: 2019-2020 with US having 100th COVID-19 case marked with a vertical line.](image)
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

• Digital surveillance sensors are going haywire
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

- Digital surveillance sensors are going haywire
- Without these sensors most models are predicting decrease in % ILI
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

- Digital surveillance sensors are going haywire
- Without these sensors most models are predicting decrease in % ILI

CDC ended 2019-2020 flu forecasting challenge.
How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

- Digital surveillance sensors are going haywire
- Without these sensors most models are predicting decrease in % ILI

How NOT to forecast a pandemic

What NOT to do: use your system for forecasting seasonal epidemic

- Digital surveillance sensors are going haywire
- Without these sensors most models are predicting decrease in % ILI

Current forecasting landscape

- About 20 teams participating in CDC COVID-19 forecasting challenge.
Current forecasting landscape

- About 20 teams participating in CDC COVID-19 forecasting challenge. *Far from an academic exercise!*
Current forecasting landscape

- About 20 teams participating in CDC COVID-19 forecasting challenge. Far from an academic exercise!
- Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)
Current forecasting landscape

• About 20 teams participating in CDC COVID-19 forecasting challenge. *Far from an academic exercise!*

• Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)

• Statistical models must be completely revamped.
Current forecasting landscape

• About 20 teams participating in CDC COVID-19 forecasting challenge. *Far from an academic exercise!*

• Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)

• Statistical models must be completely revamped. Mechanistic models easier to fit, but trust uncertainty?
Current forecasting landscape

• About 20 teams participating in CDC COVID-19 forecasting challenge. Far from an academic exercise!

• Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)

• Statistical models must be completely revamped. Mechanistic models easier to fit, but trust uncertainty?

• Severity pyramid more important than ever: we are observing some very weird data streams
Current forecasting landscape

• About 20 teams participating in CDC COVID-19 forecasting challenge. Far from an academic exercise!

• Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)

• Statistical models must be completely revamped. Mechanistic models easier to fit, but trust uncertainty?

• Severity pyramid more important than ever: we are observing some very weird data streams

• Brand new challenge: counterfactual reasoning.
• About 20 teams participating in CDC COVID-19 forecasting challenge. Far from an academic exercise!

• Delphi (CMU) and Reich Lab (UMass), as two CDC Centers of Excellence, helping in advisory role (e.g., building community ensemble)

• Statistical models must be completely revamped. Mechanistic models easier to fit, but trust uncertainty?

• Severity pyramid more important than ever: we are observing some very weird data streams

• Brand new challenge: counterfactual reasoning. Can models escape confounders, produce “scenario projections”? 
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

National ILI: 2019–2020

US has 100th COVID–19 case

What you can do:
1. Take symptom surveys.
2. Join our (separate) crowdcasting effort: https://delphi.cmu.edu/crowdcast/
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

- We are ambitiously building big data sets

![National ILI: 2019–2020](image)

US has 100th COVID−19 case

What you can do:

1. Take symptom surveys.
2. Join our (separate) crowdcasting effort: [https://delphi.cmu.edu/crowdcast/](https://delphi.cmu.edu/crowdcast/)
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

![Graph showing National ILI: 2019–2020]

- We are ambitiously building big data sets
- Plan: very soon, we will be making county-level nowcasts + forecasts
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

- We are ambitiously building big data sets
- Plan: very soon, we will be making county-level nowcasts + forecasts
- Everything that can be made public, will

![Graph showing National ILI: 2019–2020]

- US has 100th COVID–19 case

What you can do:
1. Take symptom surveys.
2. Join our (separate) crowdcasting effort: [https://delphi.cmu.edu/crowdcast/](https://delphi.cmu.edu/crowdcast/)
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

• We are ambitiously building big data sets
• Plan: very soon, we will be making county-level nowcasts + forecasts
• Everything that can be made public, will

What you can do:

US has 100th COVID−19 case
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful.

- We are ambitiously building big data sets.
- Plan: very soon, we will be making county-level nowcasts + forecasts.
- Everything that can be made public, will

What you can do: 1. Take symptom surveys.

National ILI: 2019–2020

US has 100th COVID-19 case
We are trying

We (like many, many groups, not just those involved with the CDC challenge) are trying to do something useful

- We are ambitiously building big data sets
- Plan: very soon, we will be making county-level nowcasts + forecasts
- Everything that can be made public, will

What you can do: 1. Take symptom surveys. 2. Join our (separate) crowdcasting effort: https://delphi.cmu.edu/crowdcast/