PLOS Science Wednesday: Hi Reddit, my name is Wan Yang, and I developed a forecast system that predicted the timing and magnitude of flu epidemics in Hong Kong – Ask Me Anything!

PLOSScienceWednesday [1] and r/Science AMAs[1]

[1]Affiliation not available

April 17, 2023

the WINNOWER

REDDIT

PLOS Science Wednesday: Hi Reddit, my name is Wan Yang, and I developed a forecast system that predicted the timing and magnitude of flu epidemics in Hong Kong -- Ask Me Anything!

PLOSSCIENCEWEDNESDAY R/SCIENCE

[removed]

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

DATE RECEIVED:
March 24, 2016

DOI:
10.15200/winn.145873.32408

ARCHIVED:
March 23, 2016

CITATION:
PLOSScienceWednesday , r/Science , PLOS Science Wednesday: Hi Reddit, my name is Wan Yang, and I developed a forecast system that predicted the timing and magnitude of flu epidemics in Hong Kong -- Ask Me Anything!, *The Winnower* 3:e145873.32408 , 2016 , DOI: 10.15200/winn.145873.32408

Where does the flu start? We hear about the flu coming from Asia every year, but is that true? Where in Asia and why Asia?

nate

Hi Nate. Thanks for the questions. Flu in humans is caused by three common type/subtypes: A/H1N1, A/H3N2, and type B. There are analyses suggesting that the A/H3N2 flu tends to start from East and Southeast Asia (e.g.: http://science.sciencemag.org/content/320/5874/340). However, A/H1N1 and B have complex global dynamics and for these two type/subtypes, East and Southeast Asia play a limited role in disseminating new variants. Here is recent study on this topic: http://www.nature.com/nature/journal/v523/n7559/full/nature14460.html You can also find a lot of related studies therein.

As to why Asia, there are many theories and we are still learning and trying to figure it out. But some possible reasons: (1) Asia has a large population and high population density, which would facilitate disease transmission. (2) Cultural practice, e.g., backyard poultry rearing and live poultry trading, may have also contributed to the mixing of human and animals (in particular, poultry, another host for flu). (3) Rice fields, common in Southeast Asia, may be another venue for the mixing of humans, pigs, domestic and wild fowls in this region. Such mixing may increase the chances of genetic mutation when co-infection of multiple flu strains in one host (e.g. pig) happens. Here is a paper on this if interested: http://www.ncbi.nlm.nih.gov/pubmed/9594271

What parameters go into a Hong Kong model that would be different from a model for another place?

Thanks for doing this ama. This is really interesting!

DerWasserspeier

Hi, German gargoyle (my colleague gave me the translation for your user name)! Great question. The model we used in this study is a simple susceptible-infected-recovered (SIR) epidemic model. Basically, it simulates how the flu spreads among the population by tracking the numbers of people in

three categories: (1) those who haven't been infected (i.e. susceptible), (2) those who are infected/infectious and are currently spreading the disease, and (3) those who had the disease and are immune (for a period of time).

The parameters are not much different from those used for another place (e.g. in the US); but the statistically methods we used are different, due to the more hectic epidemic pattern in Hong Kong. To give you an idea how the HK epidemic pattern differs from temperate regions such as the US: in temperate regions, flu typically surges only during wintertime once a year; however, in HK, flu can happen any time in the year and could have multiple epidemics in a year (e.g., during 1998-2013, over 16 years, we identified 44 epidemics in HK). To catch these more diverse epidemic dynamics in Hong Kong, we developed an algorithm termed space-reprobing and used it in conjunction with two data assimilation methods: the ensemble adjustment Kalman filter (EAKF) and a particle filter (PF). This algorithm is able to improve the two data assimilation methods in dealing with unexpected epidemic dynamics and make the predictions.

HongKonger here. We recently have an obvious lack of hospital resources to handle this year's influenza, and coupled with the prior government decision of reducing medical expenses/subsidy, quite a lot of debate was generated. This study seem to be able to help with some government decision.

I have 4 questions:

1. As I have said, this year influenza seems strong. Did you do any analysis of 2016 data and how accurate is it?
2. Does our government know your study, or does they actually have another but different model for prediction?
3. Assuming your model is accurate, would the knowledge obtained from your model be able to change the outcome? If so, would the model need to allow for the degree of public preparation?
4. Hong Kong people has a habit of wear surgical mask when they are sick (formed after the SARS in 2003). Do you find such habit to affect the model, or actually wearing of mask has no effect to the spread of influenza?

hinghenry

Thanks for your questions. To answer them to some extent:

(1). We haven't looked data for 2016. Observations always have biases, but they are indicator of the true epidemic dynamics.

(2). Our colleagues at the University of Hong Kong—two of the coauthors—Ben Cowling and Eric Lau, are working closely with the public health officials in Hong Kong. And we welcome collaborations with local public health agencies worldwide.

(3). That is the ultimate goal of our research, to aid public health interventions in preventing disease spread (so less cases). Our forecasts can provide information as to when the disease (e.g. the flu) is likely to surge (e.g. peak in 5 weeks) and how intensive the epidemic would be (e.g. how many cases to expect). This information could potentially be used for devising public health measures and logistics (e.g. how many vaccines/antivirals to stockpile and when to vaccinate people, etc.) And this information is particularly valuable, as in your situation—shortages in resources.

However, I'd like to point out that our HK flu forecast study is one of the first steps. We are still working on improving the forecast system.

(4). We haven't included the effect of self-adopted behavior changes in our model. But this is very

important and a future direction in our work. I think wearing face mask can to some extent reduce the spread of flu. Our colleague, Ben Cowling's group at the University of Hong Kong has some very interesting findings on this matter: http://www.nature.com/ncomms/2013/130604/ncomms2922/full/ncomms2922.html

You mention that the flu in Hong Kong is not seasonal. What are the key differences between places like the US and Hong Kong that causes the flu to be seasonal in one and not the other.

Robo-Connery

We are still learning. Climate is possibly an important factor. Studies showed that low humidity favors the flu virus's chance of survival and may therefore facilitate flu transmission. So in temperate regions like the US, flu typically circulates in wintertime when humidity is extremely low both indoors (due to heating) and outdoors (see e.g. http://www.pnas.org/content/106/9/3243.abstract). HK has a humid subtropical climate, humidity is lower in winter and thus conducive for flu transmission. During summer, humidity could be extremely high in HK. Interestingly, we found that extremely high humidity (close to 100%) is also good for the flu virus (see here: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0046789). Therefore, flu transmission could also be enhanced during humid summer days in HK, which is much less likely in the US.

Population density (i.e. whether it is a crowded environment) and travel patterns could be also contribute to the differences in flu circulation in different regions.

What were your thoughts about the now-discontinuedGoogle Flu Trends? Do you think a service like that could be even more powerful as more and more people join the Internet? Could localized search trends help fight the spread of easily-communicable diseases?

shiruken

I think GFT was a great service. We used GFT data to test our first forecast system. GFT provided near real-time estimates of flu intensity and thus a great supplement to traditional surveillance systems. It was not perfect, but none of the surveillance systems is. GFT is discontinued, but our group is collaborating with Google (http://googleresearch.blogspot.com/2015/08/the-next-chapter-for-flu-trends.html) to continue developing similar real-time estimates for our flu forecast website (http://cpid.iri.columbia.edu).

This kind of online surveillance explicitly relies on the internet traffic, so high internet access rate is definitely helpful for this system. However, whether the signals coming from internet searches are representative of true disease cases will need sophisticated "nowcast" (i.e. estimating the current disease intensity) algorithms to tease that out. Localized search trends, if utilized in a timely manner and effectively incorporated into decision making could potentially help fight infectious disease spread.

I live in Hong Kong! What is the most important point you want me to gain from your study? Thanks !

chesterdh

Hong Kong experiences very diverse flu epidemics. However, our study showed that accurate forecast of such erratic epidemics is still possible. We are working on building real-time forecasts for HK. In the meantime, it is always wise to get vaccinated and maintain good personal hygiene. So stay healthy!

Will your model for predicting likely outbreaks and pandemics be applicable to other countries?

If so, what is the next step to test this model across other similar countries with seemingly unpredictable events?

Jake_The_Muss_Heke

Yes, we believe so. We started the forecast work in the US for flu (check out our real-time flu forecast website here: http://cpid.iri.columbia.edu). We have also generated real-time forecasts for the recent Ebola epidemic in Guinea, Liberia, and Sierra Leone.

Our disease forecast system is highly adaptive. It includes three components: (1) an epidemic model to simulate how the disease spreads; (2) observational data (e.g. how many flu cases recoded for each week); and (3) a data assimilation method to incorporate to the observations into the model. With these three components, our forecast system first “trains” the model using past observations up to the most recent one. This “training” process thus optimizes the model based on local observations (so it can be adapted to other countries/regions given the local observations). It then uses this optimized model to generate a forecast.

In addition, as the model is also an independent component, we can develop different models for different infectious diseases in different countries/regions and therefore adapt the same forecast framework to other infectious diseases. Currently, our research group at Columbia University (http://blogs.cuit.columbia.edu/jls106/) is also working on forecast of other infectious diseases (e.g. respiratory syncytial virus, dengue, and West Nile virus). For dengue, the forecasts are for Iquitos, Peru and San Juan, Puerto Rico.

How different is forecasting Influenza epidemics compared to current and past epidemics (Zika, Ebola, H1/N1, etc..)? Can your models be adjusted to get a better understanding of future pandemics? What about non "airbone" viruses (HIV)?

salazarb

To answer your question, let me first explain how our disease forecast system works. It typically includes three components: (1) an epidemic model to simulate how the disease spreads; (2) observational data (e.g. how many flu cases recoded for each week); and (3) a data assimilation method to incorporate to the observations into the model. With these three components, our forecast system first “trains” the model using past observations up to the most recent one. This “training” process thus optimizes the model based on local observations (so it can be adapted to other countries/regions given the local observations). It then uses this optimized model to generate a forecast.

The diseases are very different, so different mechanisms may be needed for different diseases. However, as I mentioned above, the epidemic model is an independent component, we can develop different models for different infectious diseases in different countries/regions and therefore adapt the same forecast framework to other infectious diseases. Currently, in addition to the flu, our research group at Columbia University (http://blogs.cuit.columbia.edu/jls106/) is also working on forecast of other infectious diseases (e.g. respiratory syncytial virus, dengue, and West Nile virus). We have also generated real-time forecasts for Ebola in Guinea, Liberia, and Sierra Leone during the recent Ebola epidemic.

Would this tool be viable to use in getting a head start on any sort of prevention such as vaccination?

TheCrispyDud

That is the ultimate goal of our research, to aid public health interventions in preventing disease spread. Our forecasts can provide information as to when the disease (e.g. the flu) is likely to surge (e.g. peak in 5 weeks) and how intensive the epidemic would be (e.g. how many cases to expect). This information could potentially be used for devising public health measures and logistics (e.g. how many vaccines/antivirals to stockpile and when to vaccinate people, etc.) Of course, vaccination is always as early as possible! However, this information may still help planing, e.g. deploying resources to places that are expected to have increased cases well in advance, in particular, when there is a resource shortage.

With what method did you all create the model? Was machine learning involved? Did it involve math involving chaotic systems?

What do you mean by "irregular"? Irregular as in no pattern can be discerned? Or as in occurring with irregular intervals?

Thank you!

DarkSkyKnight

Hi there! Our disease forecast system includes three components: (1) an epidemic model to simulate how the disease spreads; (2) observational data (e.g. how many flu cases recoded for each week); and (3) a data assimilation method to incorporate to the observations into the model. With these three components, our forecast system first “trains” the model using past observations up to the most recent one. This “training” process optimizes the model prior to generating a forecast. It then uses this optimized model to generate a forecast.

The epidemic model we used for the HK forecast is a simple susceptible-infected-recovered (SIR) epidemic model. Basically, it simulates how the flu spreads among the population by tracking the numbers of people in three categories: (1) those who haven’t been infected (i.e. susceptible), (2) those who are infected/infectious and are currently spreading the disease, and (3) those who had the disease and are immune (for a period of time). It is a 2 variable oscillator, so the system is slightly chaotic.

The data assimilation methods are broadly speaking machine learning algorithms.

The term “irregular” is referring to the non-seasonal flu epidemics in Hong Kong, in comparison to the wintertime seasonal epidemics (so more ‘regular’) in temperate regions like the US.

What statistical methods do you use for your predictions?

I read a paper recently that compared the predictive value of logistic regression to regression trees in predicting 28 day mortality in patients admitted with heart failure. I don't have the reference to hand as I am on my phone but my question would be what methods have you used that are the most useful predictors in your field

lasagnwich

We used two data assimilation methods: the ensemble adjustment Kalman filter (EAKF, see the original paper here: http://www.gfdl.noaa.gov/bibliography/related_files/jla0101.pdf) and a particle filter (PF, see for example: http://www.eecs.berkeley.edu/~pabbeel/cs287-fa12/optreadings/Arulampalam_etal_2002.pdf). In addition, we developed an algorithm termed space-reprobing (see detail here: http://arxiv.org/abs/1403.6804) and used it in conjunction with the two data assimilation methods, to catch the more diverse epidemic dynamics in Hong Kong. This algorithm is

able to improve the two data assimilation methods in dealing with unexpected epidemic dynamics and make the predictions.

How did you get into working in such an interesting field? Also, have the HK government been as interested in your research as I'd imagine they would be?

Eriot

Thanks for the questions. My background is in Environmental Engineering, and for my Ph.D., I studied how environmental factors (e.g. humidity) affect the transmission of flu virus. I found it extremely intriguing; so after graduation, I joined Dr. Jeffrey Shaman's group at Columbia and continue to work on infectious diseases including the flu, Ebola, and measles.

Infectious disease forecasting is a very interesting and important research direction. Our research group at Columbia University (http://blogs.cuit.columbia.edu/jls106/) started the flu forecast work a few years back and have been generating real-time flu forecasts for over 100 cities in the US since 2013 (check out our flu forecast website here: http://cpid.iri.columbia.edu).

Our colleagues at the University of Hong Kong—two of the coauthors—Ben Cowling and Eric Lau, are working closely with the public health officials in Hong Kong. And we welcome collaborations with local public health agencies!

What were the primary algorithms you used? Why? Where they're any surprising KPIs? What languages/software packages did you use? Where did you get all the data? Were there any exceptions?

Sorry for formatting, at work. Thanks for doing an AMA!

AwePhox

We used two data assimilation methods: the ensemble adjustment Kalman filter (EAKF, see the original paper here: http://www.gfdl.noaa.gov/bibliography/related_files/jla0101.pdf) and a particle filter (PF, see for example: http://www.eecs.berkeley.edu/~pabbeel/cs287-fa12/optreadings/Arulampalam_etal_2002.pdf). In addition, we developed a new algorithm termed space-reprobing (see detail here: http://arxiv.org/abs/1403.6804) and used it in conjunction with the two data assimilation methods, to catch the more diverse epidemic dynamics in Hong Kong. This algorithm is able to improve the two data assimilation methods in dealing with unexpected epidemic dynamics and make the predictions.

We wrote the code in R (https://www.r-project.org) and published some of the code in another paper (available here: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003583)

The data for the HK flu forecast were obtained from the Department of Health in Hong Kong.

1) what input data goes into the model? Personal (anonimized) info or just total amount of flu diagnoses per day?

2) what type of model? Is this graph based?

OP isn't answering much, I think they got the flu.

extracoffeeplease

(1) The data we used are weekly influenza-like illness records (a measure for the flu but it is not very specific), combined with concurrent flu positivity rates (i.e. the percentage of sample tested positive for the flu virus). So the data are population based, i.e. no personal info involved.

(2) The model we used in this study is a simple susceptible-infected-recovered (SIR) epidemic model. Basically, it simulates how the flu spreads among the population by tracking the numbers of people in three categories: (a) those who haven't been infected (i.e. susceptible), (b) those who are infected/infectious and are currently spreading the disease, and (c) those who had the disease and are immune (for a period of time). The reason we used this simple model is that mechanisms underlying the flu epidemics in Hong Kong are still unclear. So we opted for this simpler model and used Bayesian statistical inference methods in conjunction with the model and observational data for the forecast.