從個人到群體的流感傳播 (Influenza transmission from individuals to population)

Ta-Chien Chan

(Associate Research Fellow)

地理資訊科學研究專題中心
Center for GIS, RCHSS, Academia Sinica

Outline

Background

- How big data predict influenza epidemic

Influenza surveillance at different resolutions

- Macro level: Nationwide, Townships, Hospitals, Communities
- Micro level: Social network, Personal level

Challenges from those works

Future works on smart surveillance with rich and diverse big data

Background

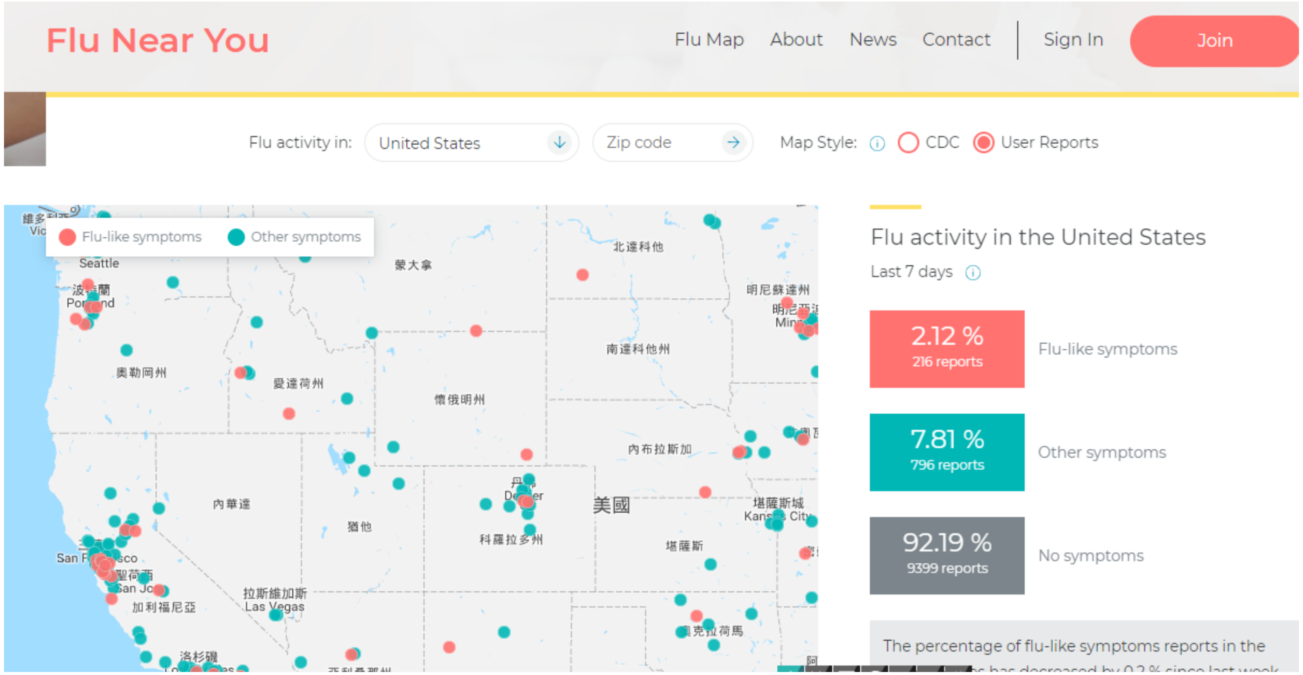
- High disease burden (morbidity and mortality)
  - WHO estimated 290 000 to 650 000 deaths every year are associated with respiratory diseases from seasonal influenza (Lee, 2018)
  - US FluSurv-Net:
    - 114,192–624,435 hospitalizations
    - 18,491–95,390 ICU admissions, and 4,915–27,174 deaths per year
    - 54–70% of hospitalizations and 71–85% of deaths occurred among adults aged 65+
- Preventable (pharmaceutical and non-pharmaceutical)
- Early detection and public health intervention

Types of traditional surveillance on influenza

- Morbidity
  - Sentinel physician
  - Syndromic surveillance
  - Health insurance data (Claims data)
  - EHR data
  - Notifiable infectious disease surveillance (Serious)
  - 學校傳染病通報系統
  - 人口密集機構傳染病監視
- Mortality
  - Pneumonia and influenza surveillance
- Laboratory surveillance
  - 實驗室自動通報系統(Laboratory Automated Reporting System, LARS)
  - 合約實驗室監視系統

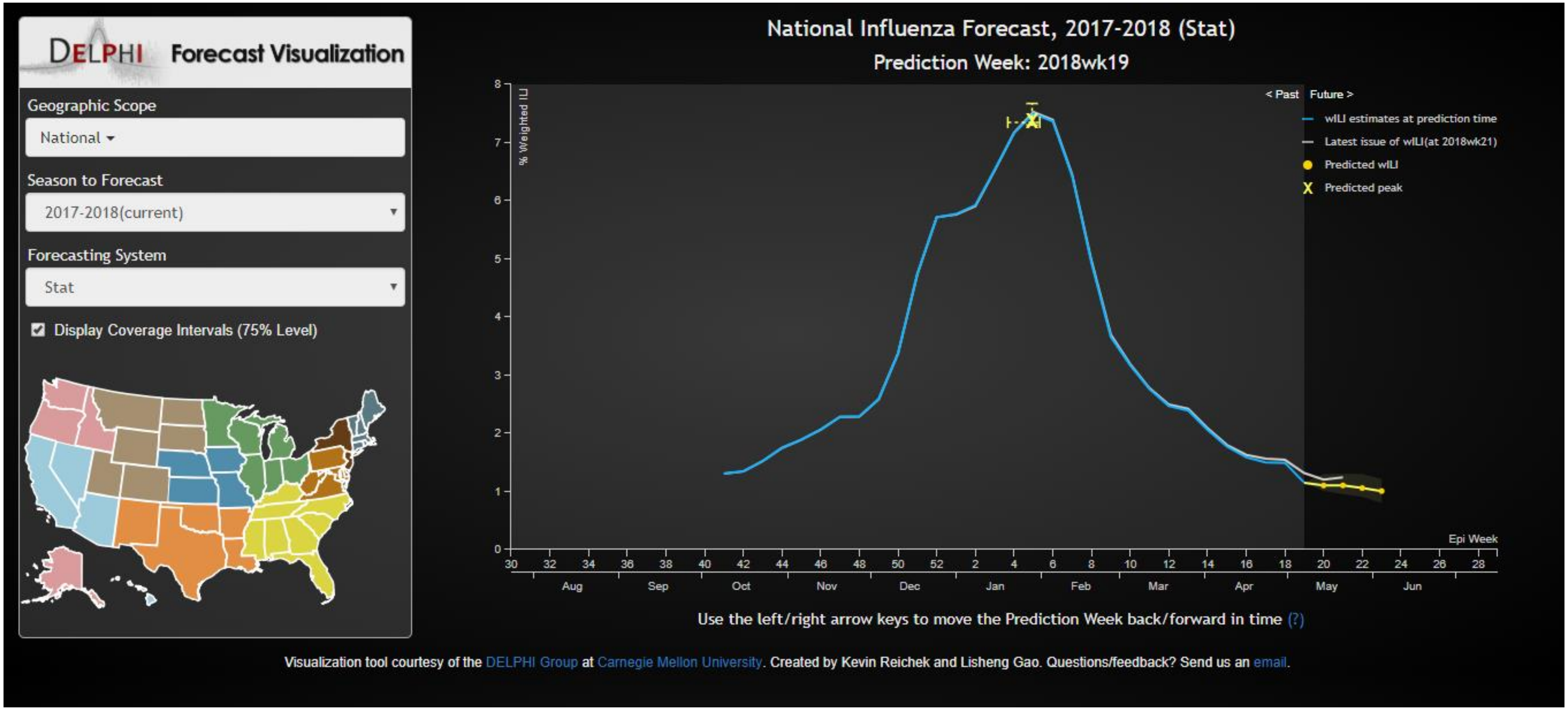
Types of informal surveillance on influenza

- Social Media (twitter, Facebook, wikipedia)
- Search engine (Google trend, Bidu)
- Participatory cohort (US: Flu near you)
- Over-the-counter surveillance
- …



Big data and AI methods on flu prediction

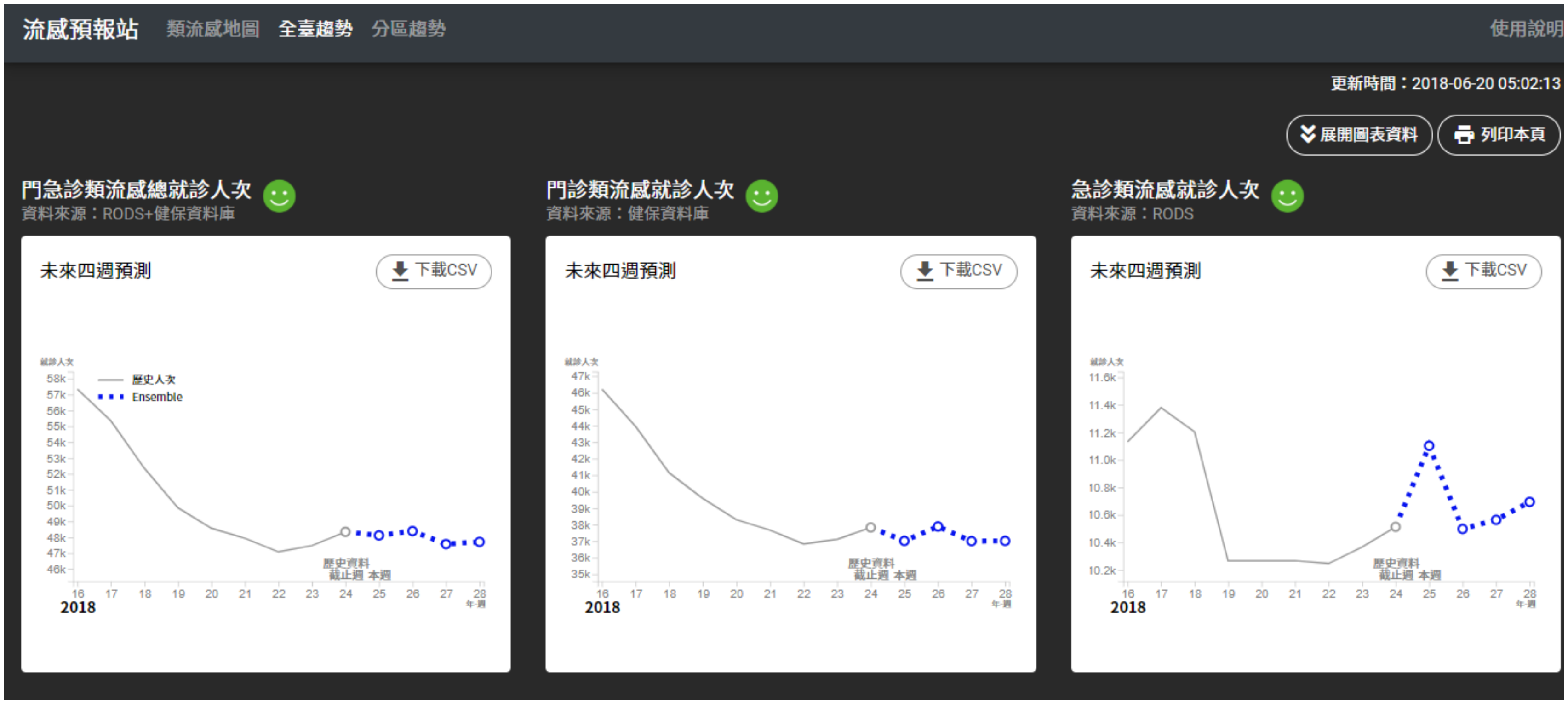
Data Mining + Official surveillance + Twitter



Professor Roni Rosenfeld, Carnegie Mellon University
2015-2016,2016-2017 Most accurate forecaster by US CDC

流感預報站 (TW CDC)

Data Mining & Statistics + Official surveillance



https://fluforecast.cdc.gov.tw/#/AllTaiwan

Influenza surveillance at different resolutions

- Influenza surveillance

全國
區域

醫院

個人

Post-SARS 或 Vaccine matching 對於P&I死亡率的影響 (PLoS One 2010)
運用健保資料進行小區域的預警，及時預警，門急診資料互補 (BMC Public Health 2015)
北市學童傳染病系統的監測 (PloS One 2015)

運用貝氏統計方法，針對醫院進行流感疫情的機率預報，類似天氣預報 (PLoS One 2010)
新型流感基因變異對疫情的影響、時空分布 (PLoS 2012)

個人是否施打流感疫苗的原因探討 (PLoS One 2014)
學齡前疫苗施打效力評估 (IJID 2015)
社會網絡與天氣因子對疾病傳染的影響 (EPIDEMIOLOGY AND INFECTION 2015)
健康行為、接觸、天氣對於流感傳播的線上追蹤研究 (*JMIR Public Health and Surveillance 2018*)

Macro level:
Township morbidity surveillance

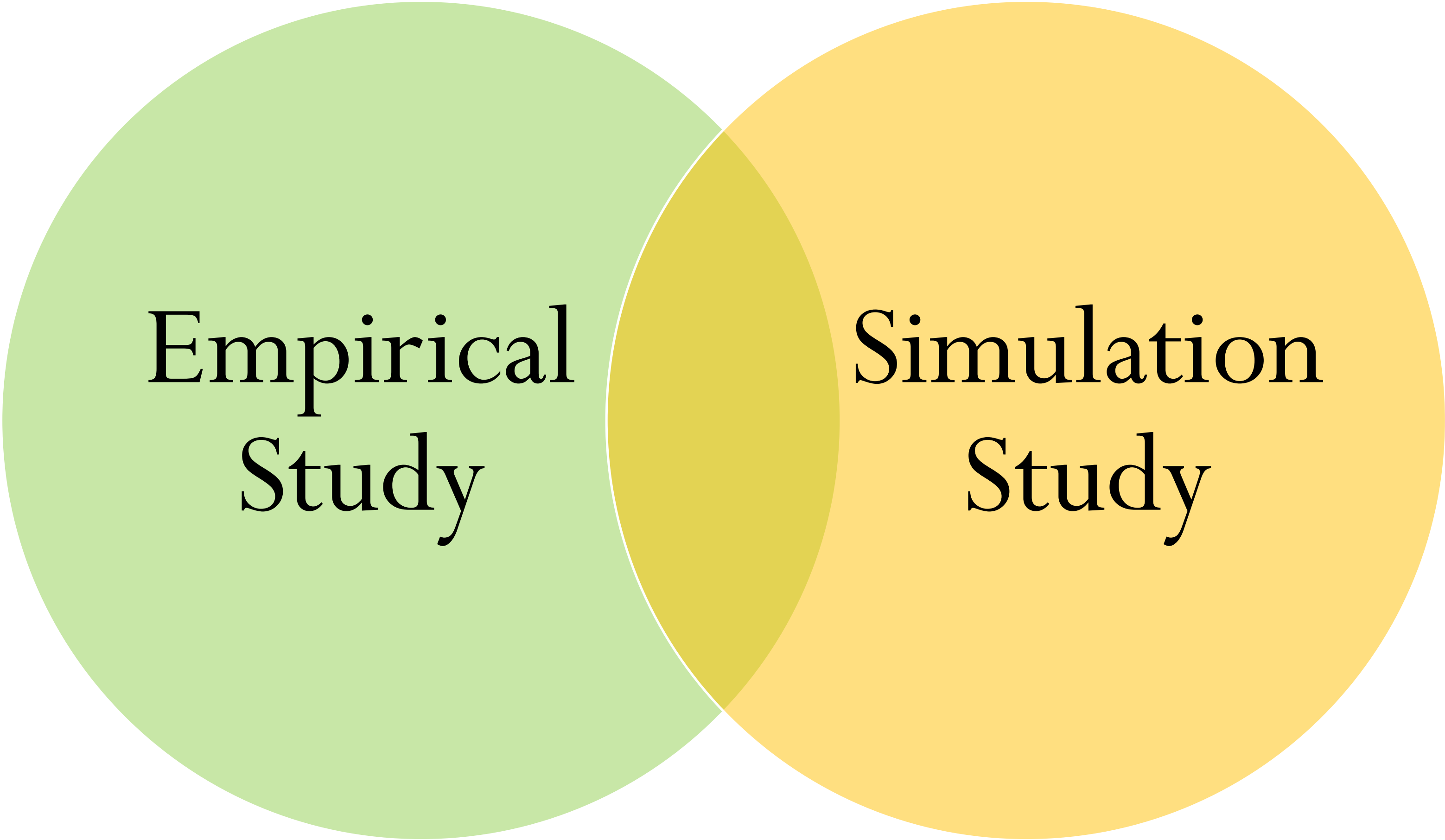
Ideas: What's new

台灣特有：
健保大數據

- Use the daily series of influenza-like illness (ILI) outpatient and ED visits in communities for outbreak detection in local areas
- Many available statistical methods for detecting aberrations
  - We used simple two-stage approach: NB regression and Pearson residuals
- Previous studies did not consider repeat clinical visits during the same infection course which might mask the true epidemic trend of ILI incidence

Repeated ILI visits within 14 days, we counted first visit

Two Approaches



Empirical study

Select predictors

- Negative binomial regression
- OPD & ED in three areas
- 2004-2007

Covariates we have: temperature, last week new ILI cases, day of the week, holidays, seasonal term (moving month of the year)

Compute residuals

- Using three years' temporal window to fit the NB regression
- Compute daily residuals

Aberration detection

- Monitor daily standardized Pearson residuals >1.96
- 2008-2009

Simulation Study

- We also conducted a simulation study to compare the performance between the proposed approach and modified CUSUM method

Simulate the baseline visits (include seasonal pattern and day-of-week pattern)

Inject the distribution of the daily new cases (controlled by signal-to-noise ratio)

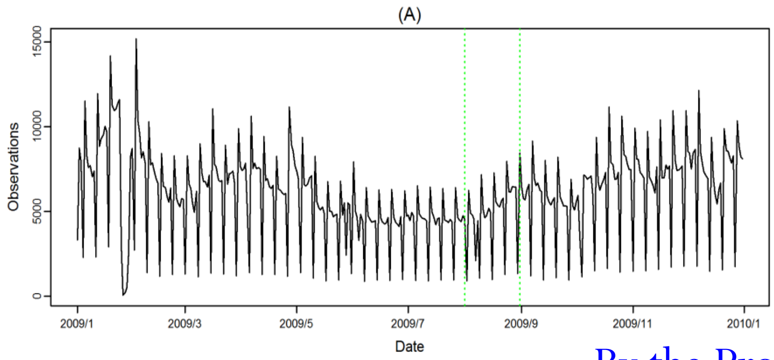
1,000 simulations

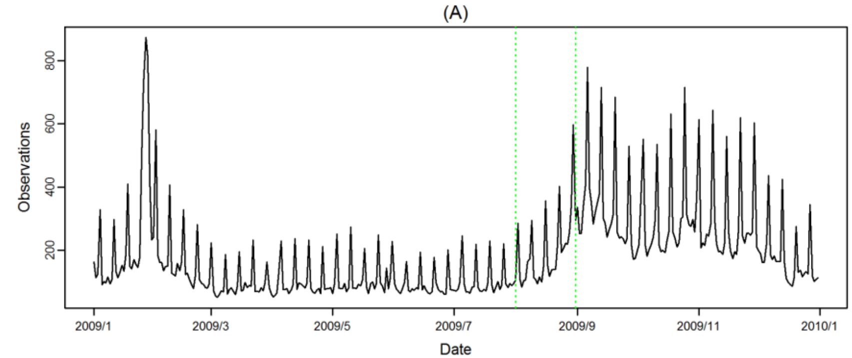
Compare the proposed method with C3 (Pearson residuals)

ILI aberration detection in northern Taiwan during 2009

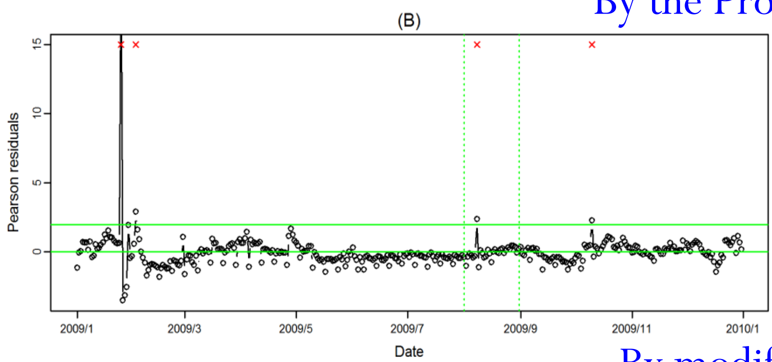
Outpatient visits

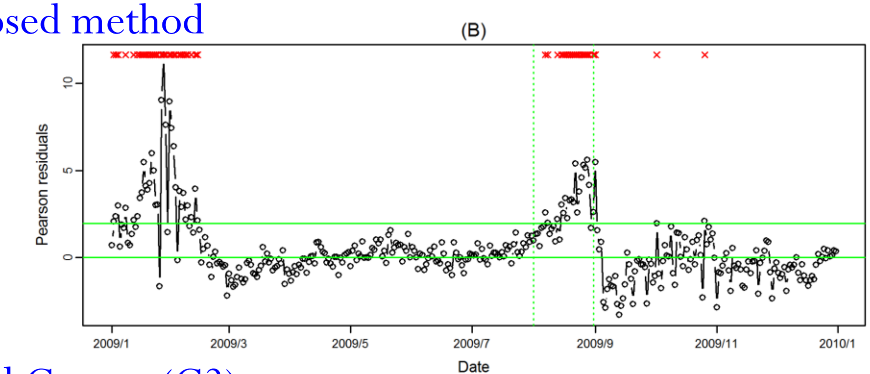
Emergency department visits



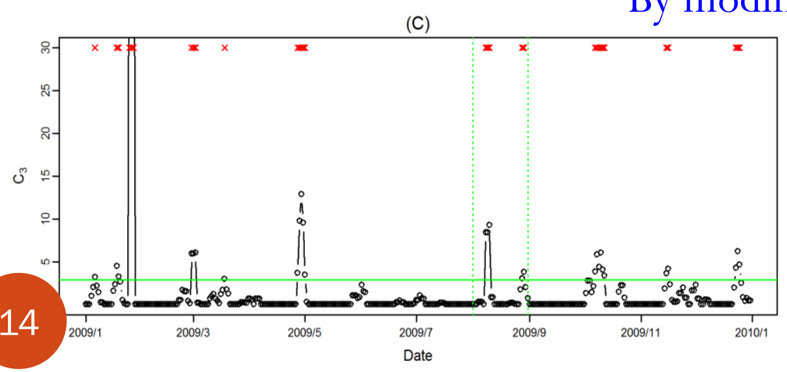


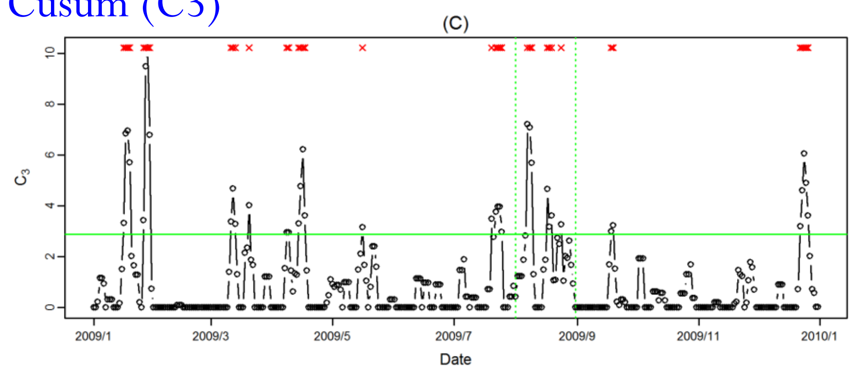
By the Proposed method





By modified Cusum (C3)





Summary

- By directly monitoring Pearson residuals of fitted negative binomial models to health insurance claims data, near real-time ILI aberration detection in communities is attainable
- Application of the approach to local health insurance claims data can improve the accuracy of outbreak detection in small-area-based ILI surveillance
- The complementary signals in both outpatient and ED visits were able to make ILI surveillance more comprehensive

Macro level

Hospital-based syndromic surveillance

THE BIG STORIES

Scientists create system to forecast the flu like weather

What if experts could predict when and where the flu will spread, like an oncoming storm? Columbia scientists have done just that.

By Kristin Toussaint Published : March 05, 2018 | Updated : March 05, 2018

Ideas: What's New

- Can we make flu forecast like weather forecast ?
  - I have proposed it in 2010 (Chan et al., 2010)

Tradition: Binary 0 or 1

Probability: 0 ~ 100%

地區名稱	天氣預測	氣溫	降雨機率
台北市	多雲午後短暫陣雨	28~32°C	50%
基隆北海岸	多雲午後短暫陣雨	28~32°C	30%
台北地區	多雲午後短暫陣雨	28~32°C	50%

- Syndromic ILI data in hospitals
- Considering spatial interaction among hospitals

Data and Software in Probability Prediction

- Data Sources
  - 5 emergency rooms of community hospitals in Taipei City
  - Daily data
  - 2 years’ training data and 2 months’ validation data
  - ILI Diagnosis by composite ICD-9 codes
  - The daily meteorological data were from Taiwan Central Weather Bureau (TWCWB)
- Software:
  - R and WinBUGs

Gibbs sampling algorithm

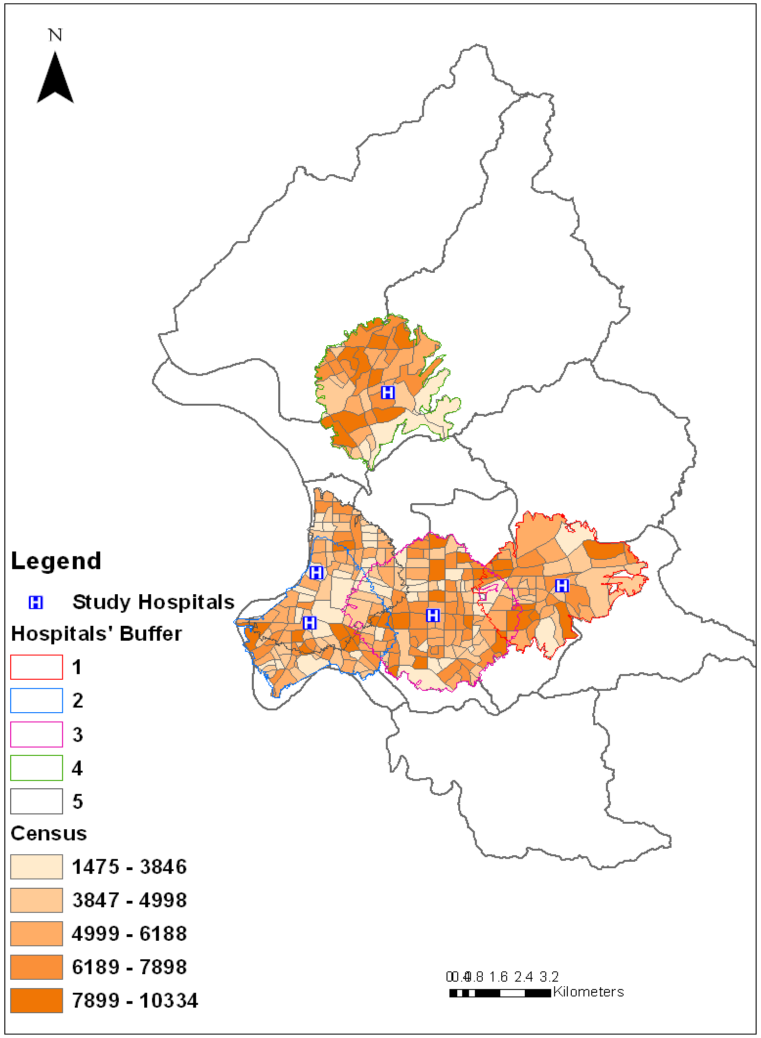
1. Prediction and Graph
2. Gelman and Rubin statistic for convergence diagnosis

Bayesian Hierarchical Model

- Outcome (Dependent) variable
  - ILI daily visits in each hospital
  - Poisson distribution
  - Population at risk within each buffer (census data)
- Explanatory (Independent) variables
  - Previous day's ILI visits
  - Previous day's mean temperature
  - Previous day's mean vapor pressure
  - Conditional Autoregressive Model (CAR) for spatial interaction
  - Weekend and holiday effect
  - Linear temporal trend and seasonal effect
- Probability Exceeding the Threshold
  - Dynamic threshold: the maximum value of past 7 days

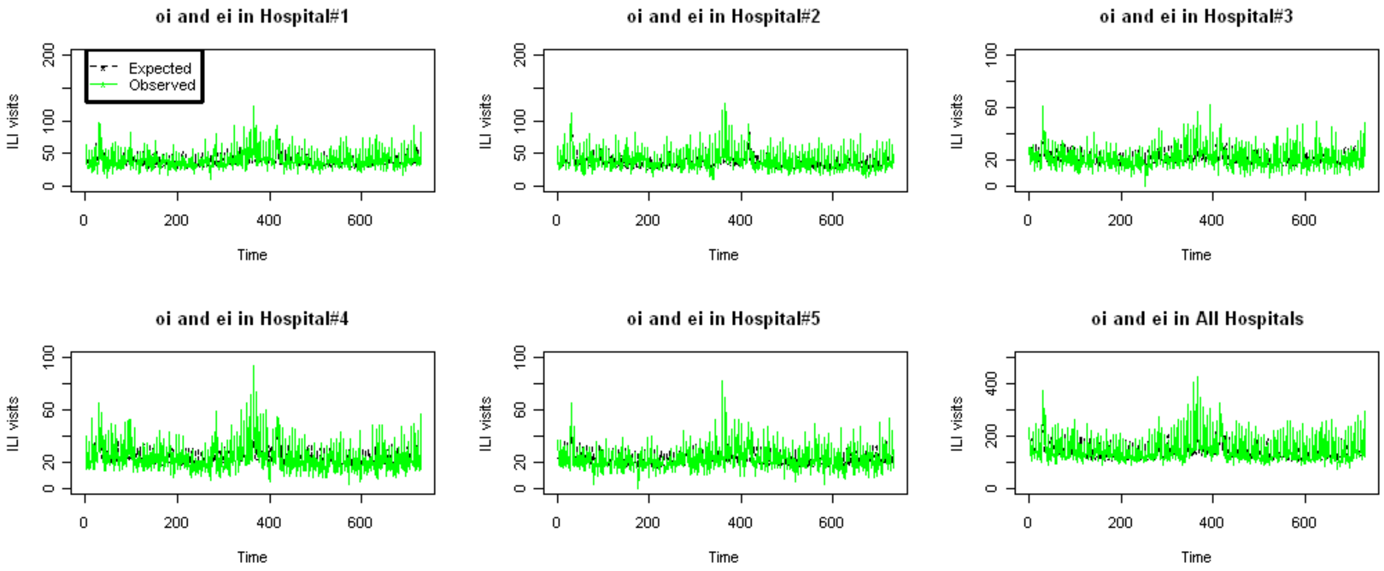
Spatial correlation

- The buffer was calculated by the real road network
- The spatial neighboring relationship was defined with a 3 km network buffer surrounding each hospital.
- E.g., hospital 3 vs. hospital 1,2,5
- Calculate by ArcGIS, Network Extension



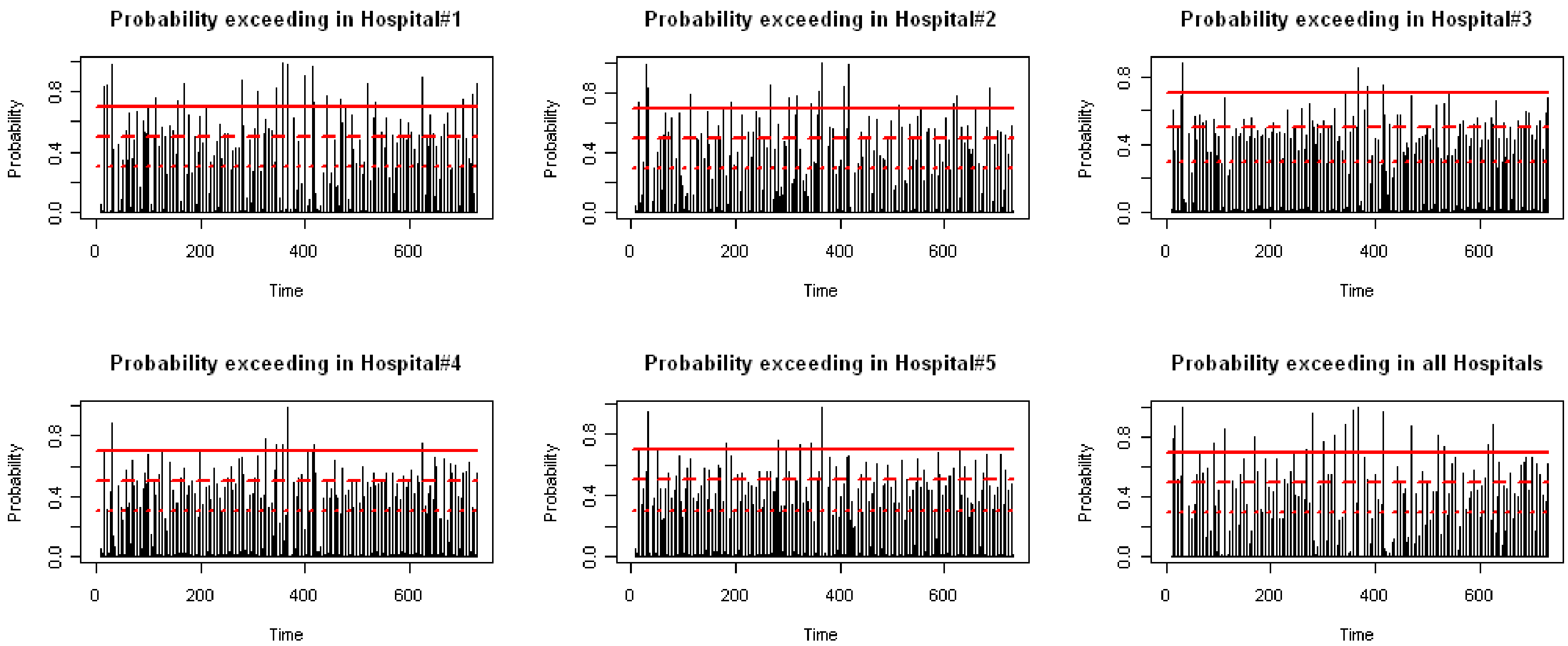
Probabilistic prediction in Taipei City

Temporal Patterns of Observed (oi) and Expected ILI (ei) Visits during 2006-2007



oi and ei close together, overall r=0.8 (p<0.0001)

Probability of Alert at the Stage of Model Fitting



Around 1 ~ 3% of days exceeding 70%

Top line is 70%; Middle one is 50%; Bottom one is 30%.

Prediction accuracy for validation with two time scales

Weekly Prediction				Monthly Prediction			
Days for updating model	Days for validation	APE	ARRMSE	Days for updating model	Days for validation	APE	ARRMSE
1–730	731–737	14.73	0.25	1–730	731–758	3.00	0.40
1–737	738–744	10.39	0.34	1–737	738–765	−6.67	0.43
1–744	745–751	−5.34	0.57	1–744	745–772	−10.65	0.42
1–751	752–758	−10.55	0.38	1–751	752–779	−8.34	0.31
1–758	759–765	−21.39	0.42	1–758	759–786	−2.50	0.30
1–765	766–772	4.52	0.26	-	-	-	-
1–772	773–779	−1.47	0.09	-	-	-	-
1–779	780–786	12.63	0.32	-	-	-	-
Average =		0.44	0.33	Average =		−5.03	0.37

Summary

- The Bayesian model is able to detect the signals 1-2 days prior to the rise of ILI visits

- Intensive computation issues can be solved by R-INAL recently
  - Especially for parameter updating
  - Multi-stream big data such as weather info.
  - Including weather forecasts as the model's inputs

- The decision rule for public health action needs to be further examined with laboratory isolation data

Micro level

Participatory cohort study

Personal risk factors and social network effects

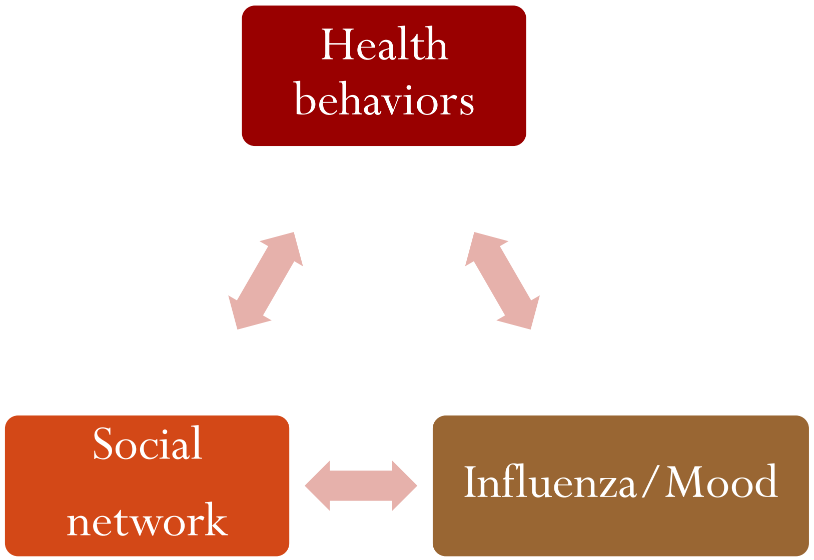
Social network and health

http://cdiary.tw or http://點日記.tw

目前已經進入點日記2.0
問的問題與介面已經跟1.0不一樣

Where did ClickDiary from ?

- Three years research project funded by Academia Sinica (2014-2016) : Social Media and Contingent Network Structures: Events, Boundaries, and Diffusion
- Joint groups from statistics, sociology, epidemiology
- ClickDiary is one of the sub-project



Features of Clickdiary project (2014-2016)

Basic Info.

- Demographics
- Personality traits
- Emotion
- Health status
- Contact Network

Health Behavior

- Sleep
- Diet
- Exercise
- Emotion
- Weight/ waistline (Option)
- Blood pressure/ heartbeat (Option)

Social Network

- Number of Contacts
- Personal network
- Types of contacts
- Contact intensity
- Purposes of contacts
- Contact emotion

Disease Surveillance

- Ego with ILI symptoms
- Alters with ILI symptoms

Ideas

- Inter-personal Influenza transmission through social contact
  - Aerosol transmission accounts for more than 50% of transmission
  - Household or school transmission is the major research setting
  - Few prospective cohort studies
- Traditionally, the survey will let the participants to recall a long period of your health behaviors, flu symptoms
  - Such as past two weeks, one month, one year…
  - Here we used **diary-based study design**
- External exposure also plays some roles on transmission and personal immunity
  - Air pollutants, weather conditions

Aims of the study

- We try to answer:
  - Examine the risk of getting influenza infection after contacts with infected persons
  - Examine how personal health behaviors, weather and air pollution affect the chance of getting influenza infection

Statistical Method

- Two stages modeling
  - Logistic regression model (First stage)
  - Mixed-effects logistic regression model (Second stage)
- Logistic regression model
  - To find influential variables of health behaviors and weather, and air pollution
- Mixed-effects logistic regression model
  - Use the variables selected from the first stage
  - Add random component for subject-to-subject variation
  - Add random component for the food items
  - Consider the correlation between a pair of the responses of a subject

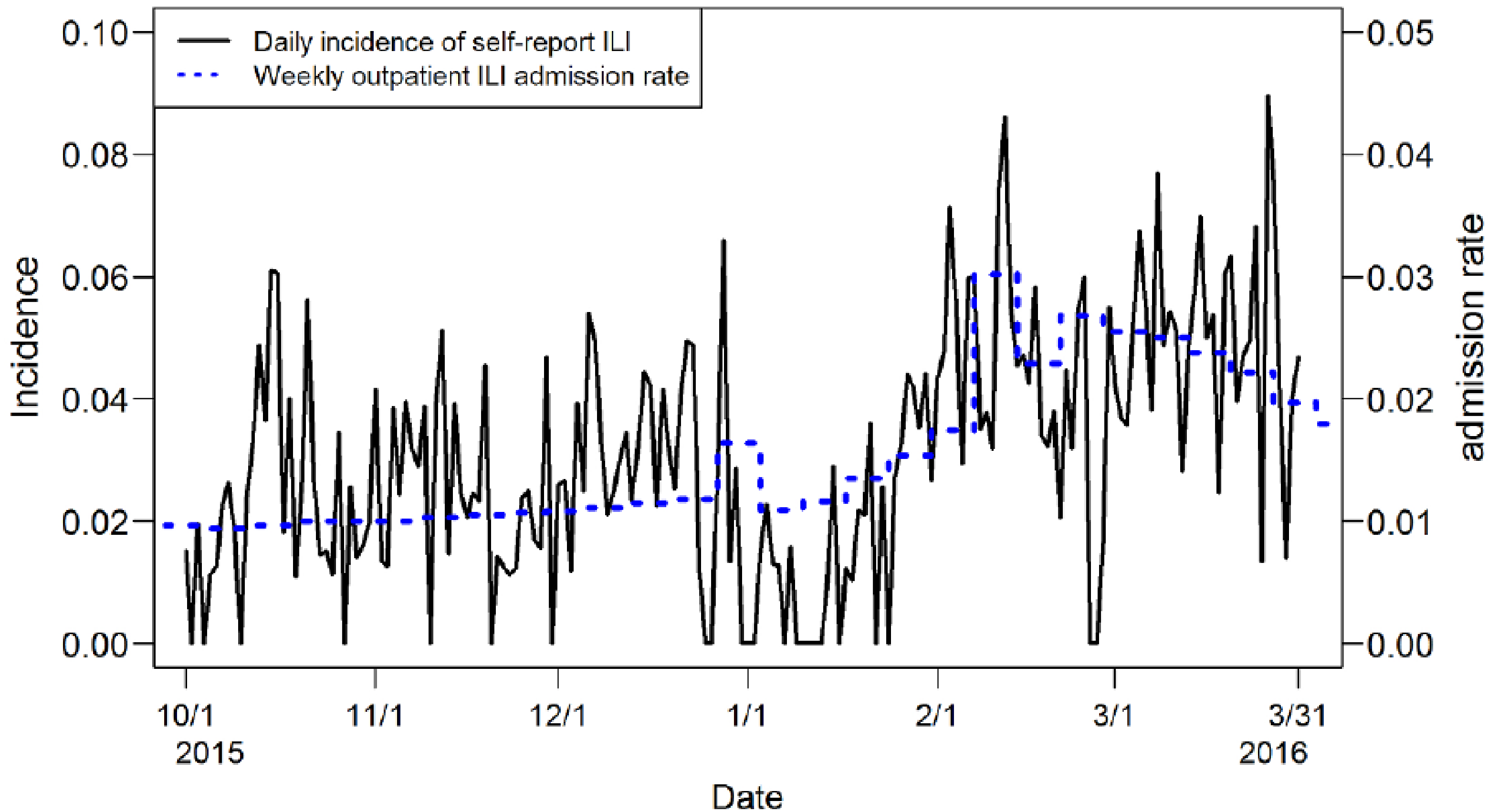
Results

- 160 participants (small samples)
  - 14,317 person-days
  - 124,222 face-to-face contacts

Descriptive statistics of number of days when participants recorded health diary and face-to-face contact diary, from October 1, 2015 to March 31, 2016.

Type of diary	Number of days					
	Minimum	25%	50%	75%	Maximum	Mean
Health diary	29	63	76	113	183	89
Contact diary	6	64	71	116	183	91
Either diary	29	66	91	128	183	99

Temporal trends of daily and weekly incidence rate of self-report influenza-like illness (ILI) and weekly outpatient ILI admission rate



Estimates of the odd ratios in the mixed-effects logistic regression models

Variables	IQR	OR (95% CI)
Binary variables		
Free of ILI and contact with infected persons[a]		1.87 (1.40-2.50)
Self-reporting ILI and no contact with infected persons[a]		55.79 (45.26-68.77)
Self-reporting ILI and contact with infected persons[a]		59.97 (44.32-81.14)
Age >60[b]		0.06 (0.0005-8.23)
Male[c]		0.30 (0.05-1.76)
Late bedtime[d]		1.43 (1.11-1.84)
Continuous variables		
Vegetables	1.0	0.92 (0.64-1.33)
Fruits	1.5	0.37 (0.19-0.75)
Cereals	1.25	0.99 (0.70-1.40)
Beans and pulses	1.0	0.42 (0.20-0.87)
Meats and eggs	2.17	1.09 (0.67-1.77)
Dairy products	0.67	0.31 (0.14-0.69)
Sleep duration (h)	1.67	0.97 (0.84-1.12)
Exercise time	30.5	0.73 (0.63-0.84)
Temperature deviation	1.37	1.25 (1.13-1.39)
log ($PM_{2.5}$)[e]	0.68	1.13 (0.99-1.30)
O_3	12.66	1.33 (1.20-1.49)
For two continuous variables		
log ($PM_{2.5}$) and O_3	0.68 and 12.66	1.51 (1.29-1.76)

Possible transmission

Bedtime

Fruits

Beans

Diary products

Exercise

Temperature

Air pollution

Summary

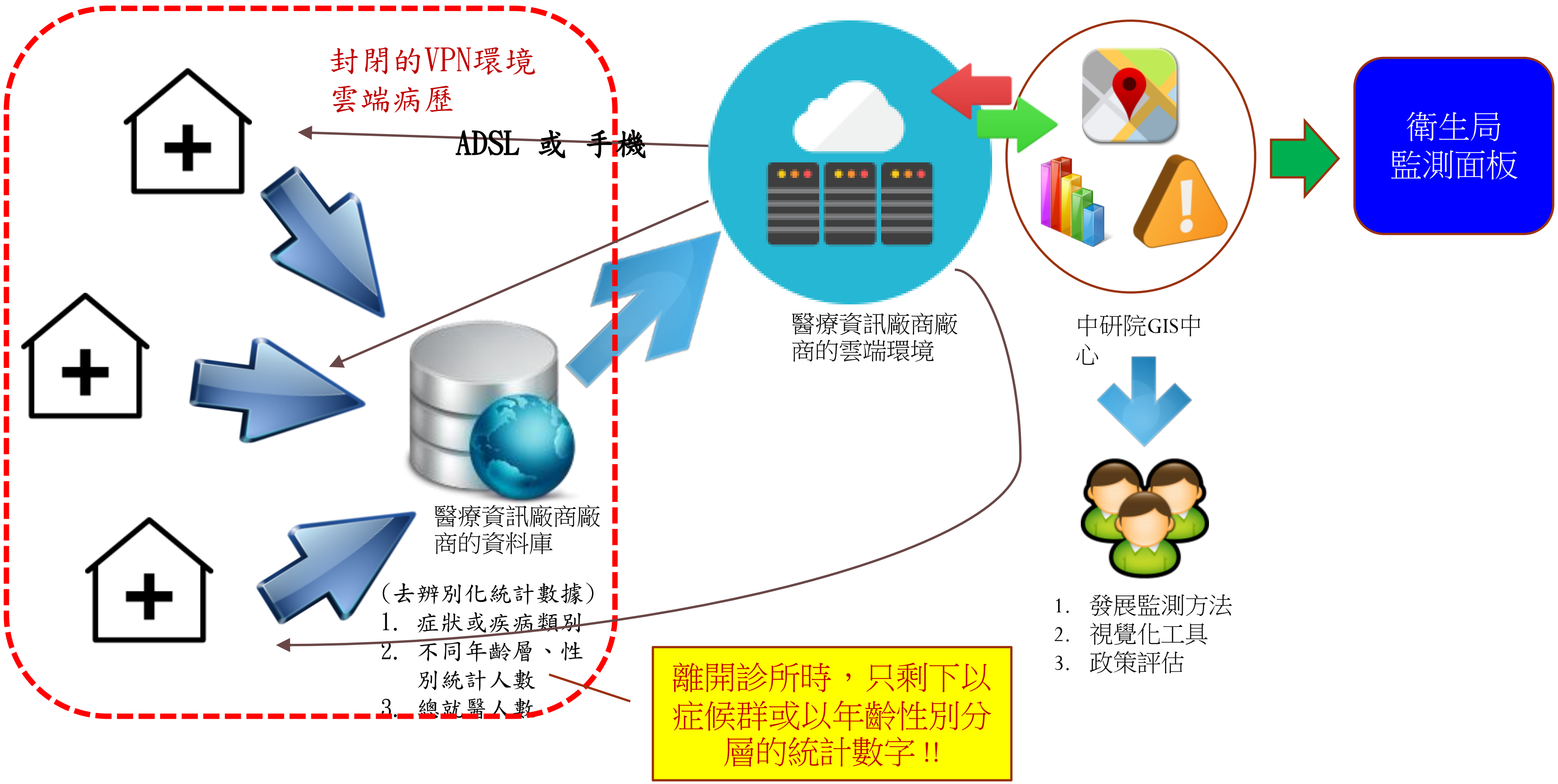
- Our study shows that keeping a healthier lifestyle, including having a nutritious diet, sleeping earlier, and doing longer physical exercise reduces the risk of influenza infection.
- Self-protection and avoiding contact with infected persons, as well as keeping alert to temperature changes and air quality can also reduce the risk of influenza infection.

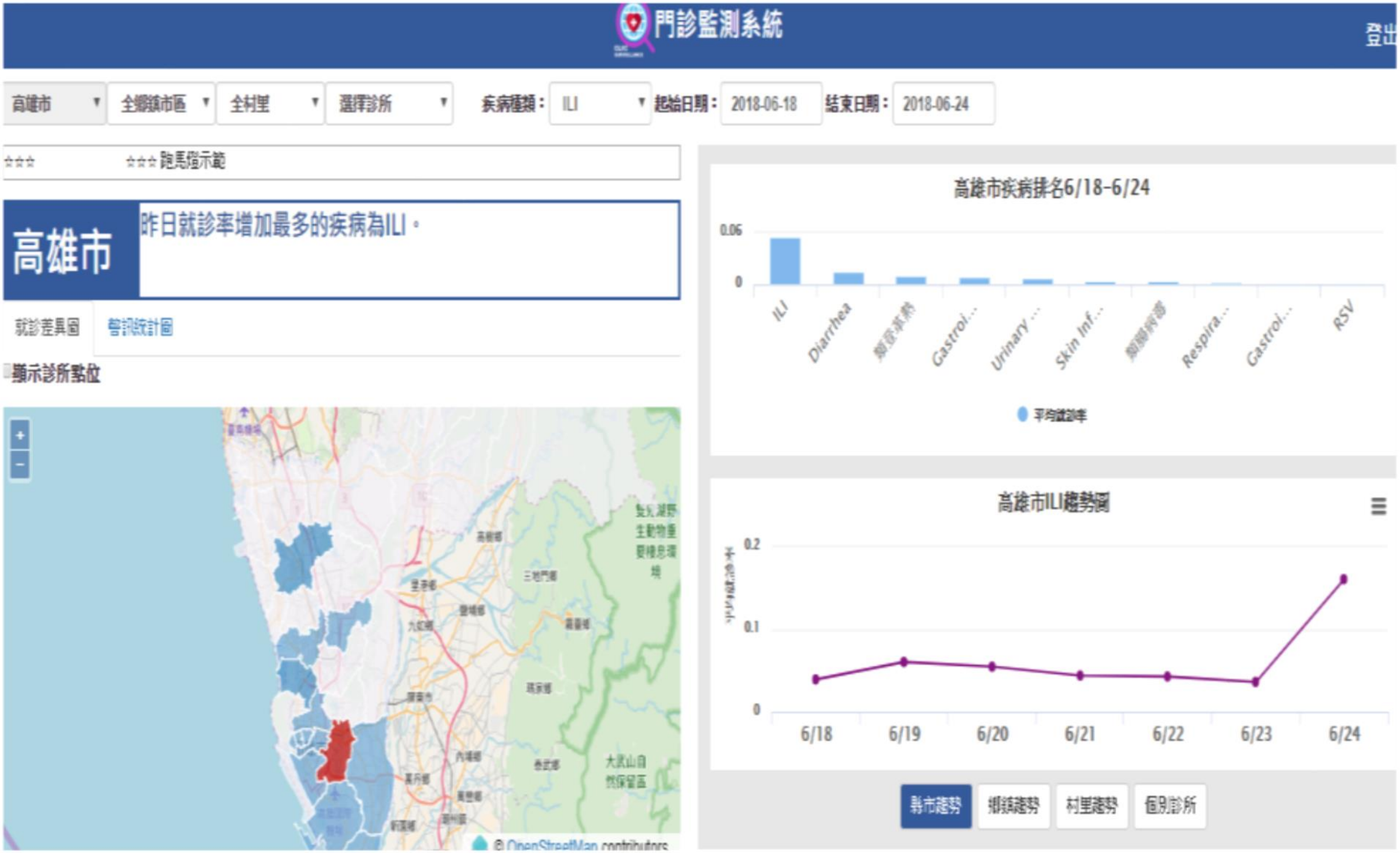
Challenges from those works

- Can we integrate different resolutions' data for prediction ?
- Can we improve our estimation on population at risk and human movement ?
- Can we use participatory cohort approach to enhance the understanding of personal behaviors, attitudes, morbidity, vaccine effectiveness and virus change ?
  - UK Flu Watch (Int J Epidemiol. 2017 Apr; 46(2): e18)
  - US Flu Near You (Scientific Reports 2015;5:9540)

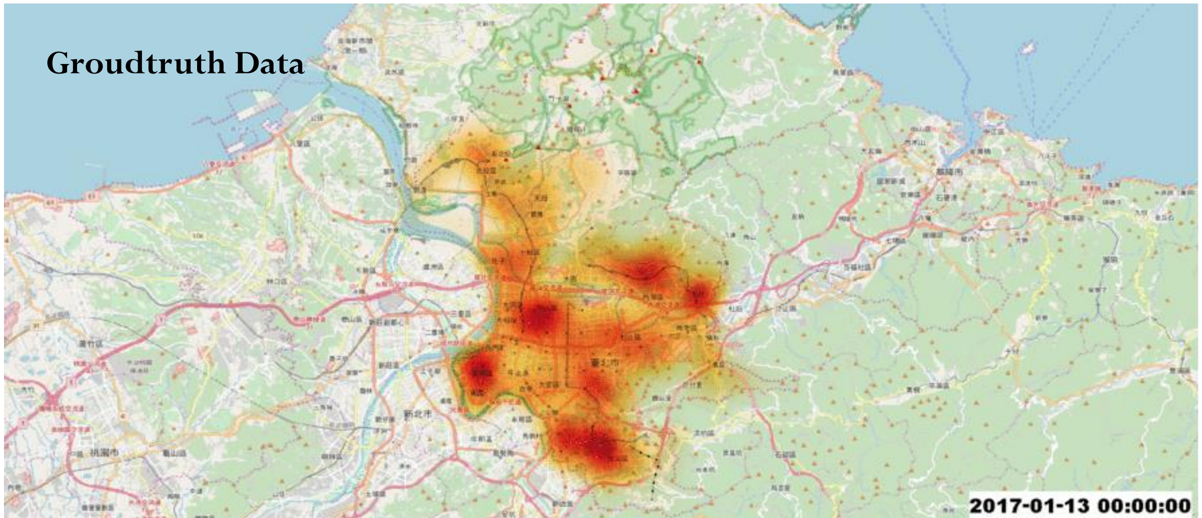
Future works on smart surveillance with rich and diverse big data

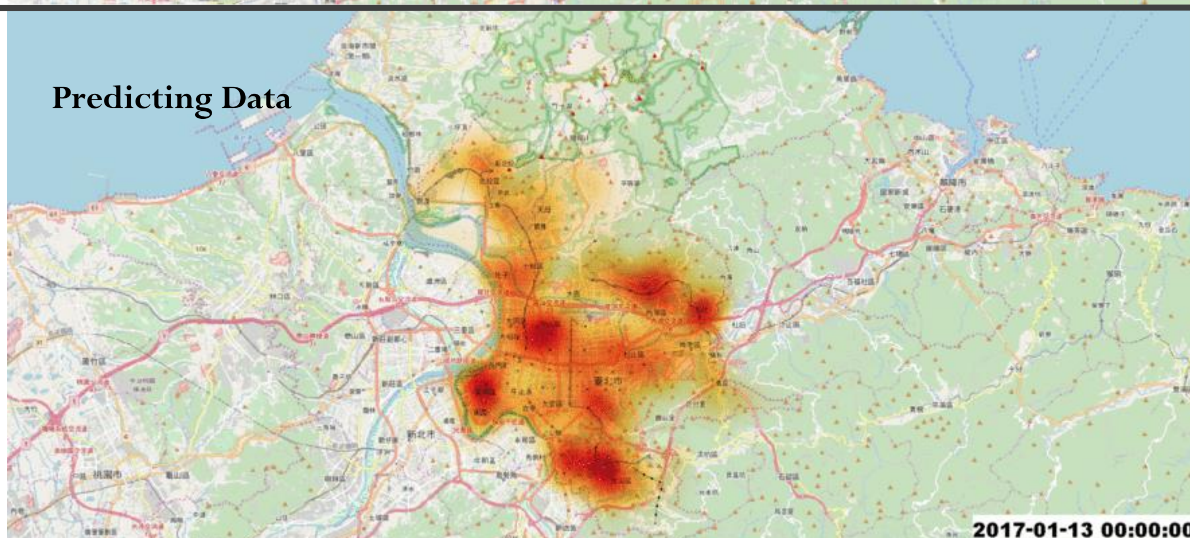
Enhanced sentinel surveillance in communities




門診監測系統
登出
高雄市
全鄉鎮市區
全村里
選擇診所
疾病種類：
ILI
起始日期：
2018-06-18
結束日期：
2018-06-24
跑馬燈示範
高雄市
昨日就診率增加最多的疾病為ILI。
就診差異圖
警訊統計圖
顯示診所點位
© OpenStreetMap contributors
高雄市疾病排名6/18-6/24
0.06
0
ILI
Diarrhea
Gastroi...
Urinary ...
Skin Inf...
Respira...
Gastroi...
RSV
平均就診率
高雄市ILI趨勢圖
0.2
0.1
0
6/18
6/19
6/20
6/21
6/22
6/23
6/24
縣市趨勢
鄉鎮趨勢
村里趨勢
個別診所

Human movement and population at risk from mobile phone data →Estimate Transmission





Thank you for your attention

E-mail: tachien@gate.sinica.edu.tw

Contact tree

Average mood status after your contact

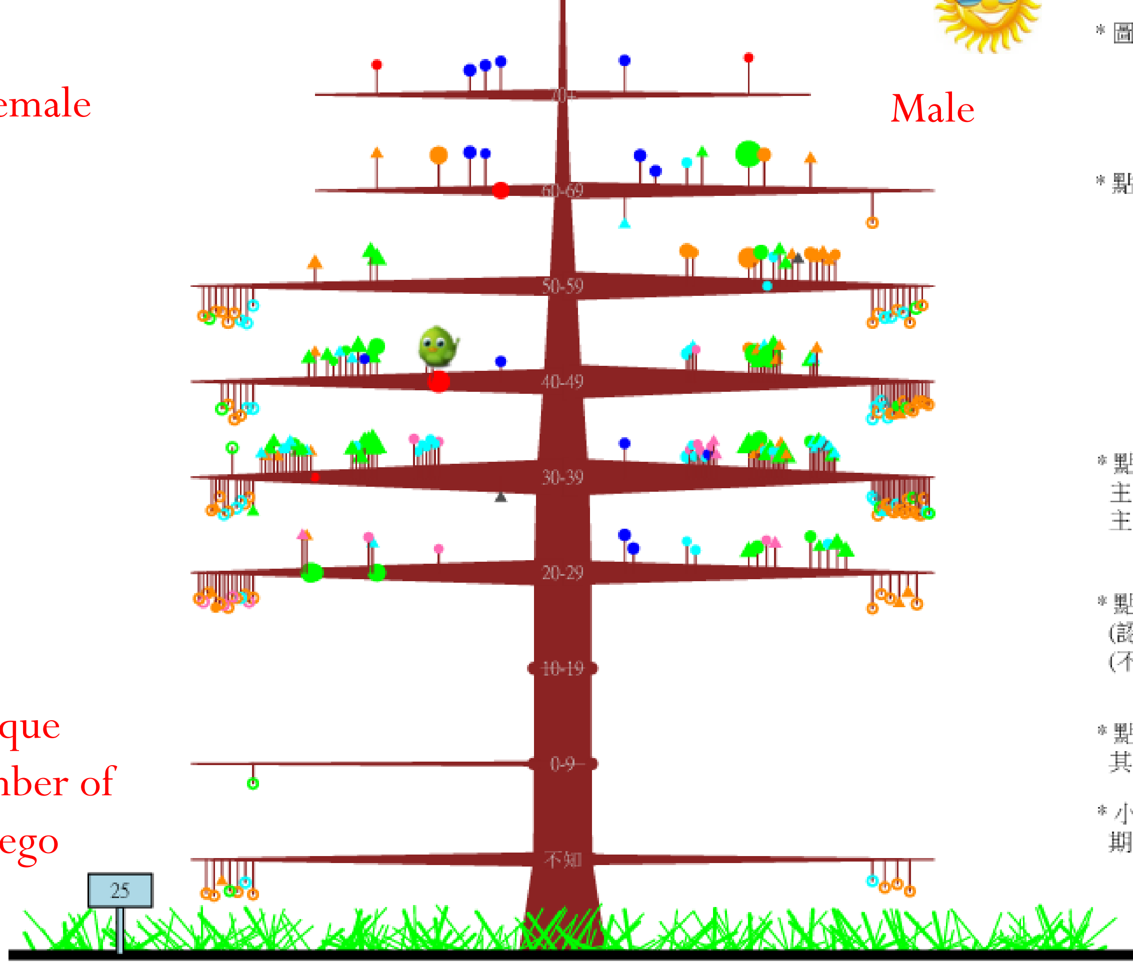
* 右上角圖表示最近一天與人接觸的心情

* 圖上的點表示接觸者與我的熟悉程度:
  - ● 很熟
  - ▲ 認識不熟
  - ○ 不認識

1. Familiar
2. Known but unfamiliar
3. Unknown

Female

Male



Unique number of the ego

* 點的顏色:

家人
親戚
同事
鄰居
朋友
同學
其他

Color represented different kinds of relationships

* 點在主枝幹上表示之前常聯絡,
  主枝幹上方表示不常聯絡,
  主枝幹下方表示從未聯絡

* 點離主幹的距離由近到遠依序表示
  (認識20年以上),(5-19年),(1-4年),
  (不到1年),(不認識)

* 點越大表示該接觸者認識越多
  其他接觸者

* 小鳥停駐的位置表示點日記
  期間這位接觸者與我最常連絡