From Quality Control to Disease Screening and Surveillance

Peihua Qiu

Department of Biostatistics
University of Florida
E-mail: pqiu@ufl.edu

October 2017
University of Cincinnati

Biostatistics!?

"That's what I want to say. See if you can find some statistics to prove it."

"I can prove it or disprove it! What do you want me to do?"

Biostat vs. Medicine

Biostatistics	Medicine
Data	Patients
MS-level biostatisticians	Nurses & Physician assistants
Ph.D.-level biostatisticians	Medical doctors
(Consultants, researchers)	(Physicians, researchers)
Biostatistical service	Medical service

BERD at UF CTSI

BERD Logo (The early bird gets the worm):

BERD at UF CTSI (con'd)

BERD Personnel:

- Peihua Qiu (Director, Biostat)
- Susmita Datta (Co-Director, Biostat)
- Baiming Zou (Biostat)
- Sam Wu (Biostat)
- Matt Robinson (MS biostatistician)
- Jim Chen (Epidemiology)
- Bruce Vogel (Health Outcome and Policy)
- Mary Ellen Young (Occupational Therapy)

UF Department of Biostatistics

- Faculty members: 20
- PhD students: 23
- MS students: 22 (on-campus), 25 (online)
- MPH in Biostatistics: 11
- $10.28 million extramural grant funding (yr 2017)
- Elected fellows: AAAS (2), ASA (4), IMS (2)
- Two centers: COG, CIDID

Quality of a Product

- Small items: clothes, shoes, watches, food, ...
- Larger items: bikes, cars, televisions, airplanes, ...
- Network systems: banking system, health care system, intenet system, ...
- CTSI: quality of consulting service
- quality of research

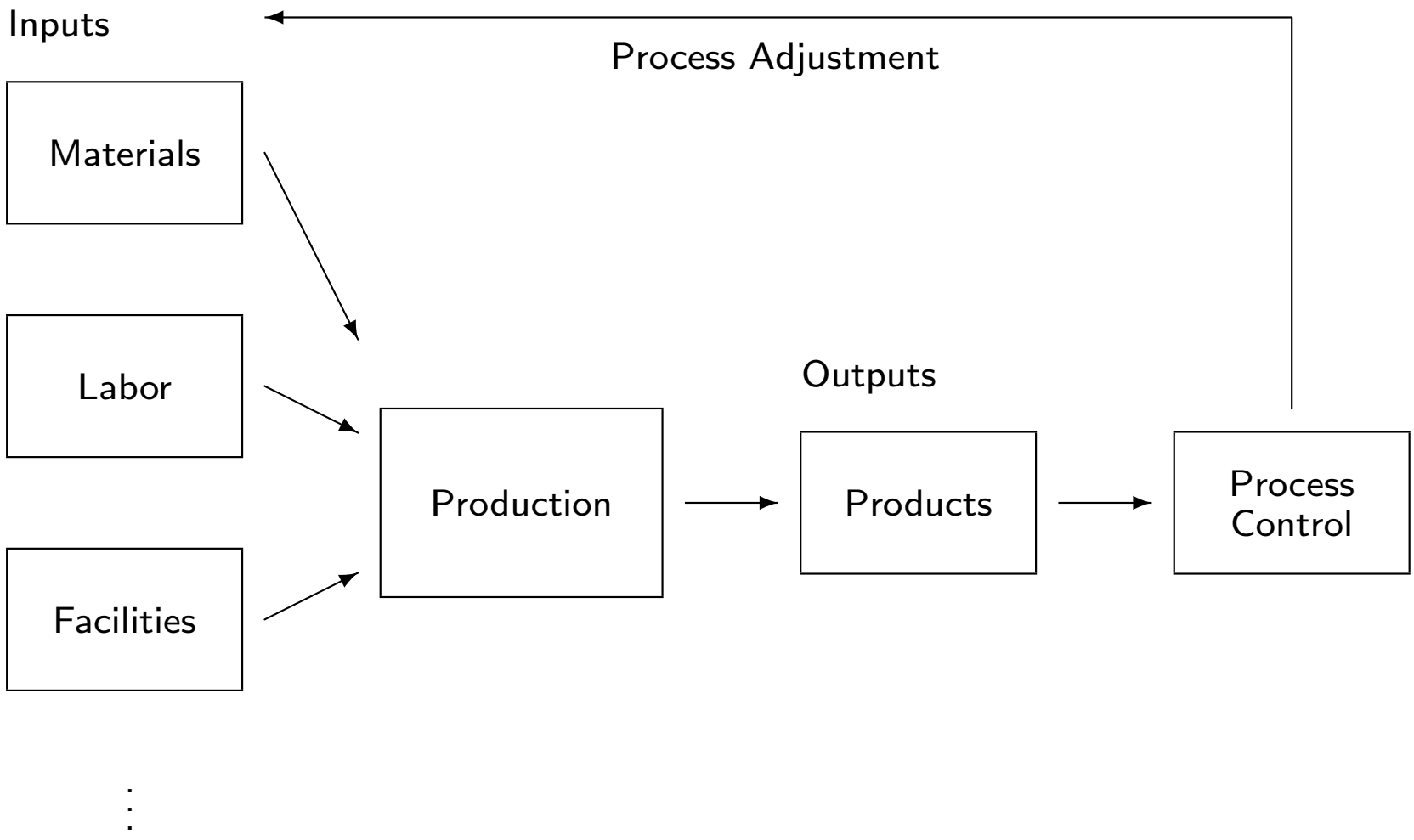
Quality Assessment

How can we assess the quality of a product?
Quality is a multi-faceted concept. Eight dimensions of quality:

- Performance
- Features
- Reliability: how often the product fails
- Conformance: meets its designed standard?
- Durability: life of the product
- Serviceability: maintenance of the product
- Aesthetics: visual appeal
- Perceived quality: reputation

Quality variables (indices)

Quality Management



Total Quality Management (TQM), Six-Sigma: project-by-project approach

Basic Terminologies

- Phase I: Properly adjust the process to make it run stably
- Phase II: Online monitor the process
- In-control (IC) versus Out-of-control (OC) status
- Common-cause variation versus special-cause variation

Basic Tools for Process Control

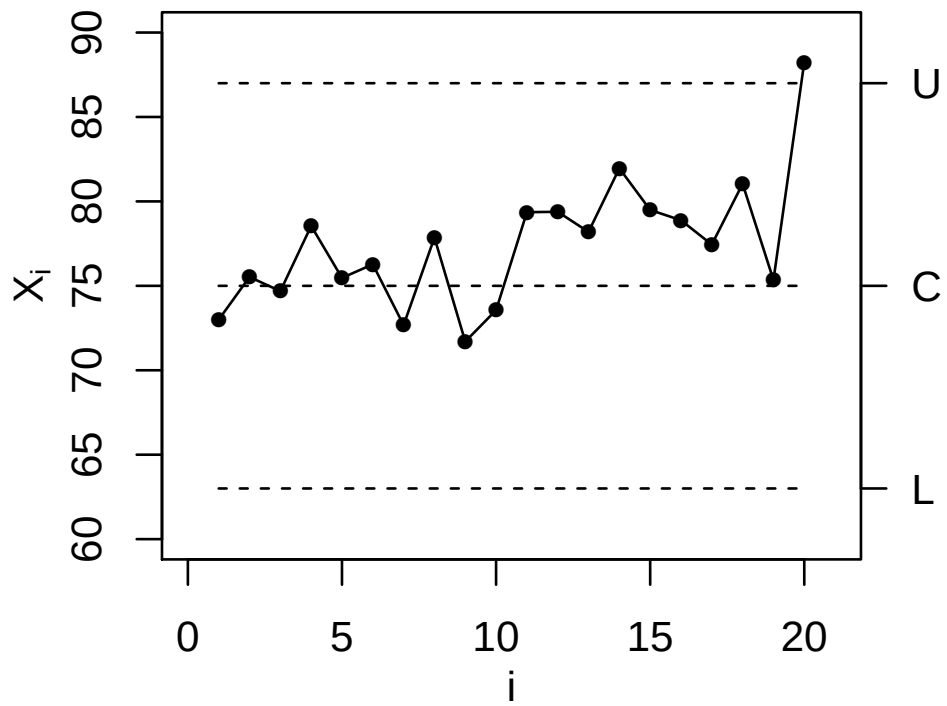
- Shewhart charts (Shewhart 1931)
- Cumulative sum (CUSUM) charts (Page 1954)
- Exponentially weighted moving average (EWMA) charts (Roberts 1959)
- Change-point detection (CPD) charts (Hawkins, Qiu and Kang 2003)

Major Goal: Distinguish special cause variation from common cause variation.

Shewhart chart

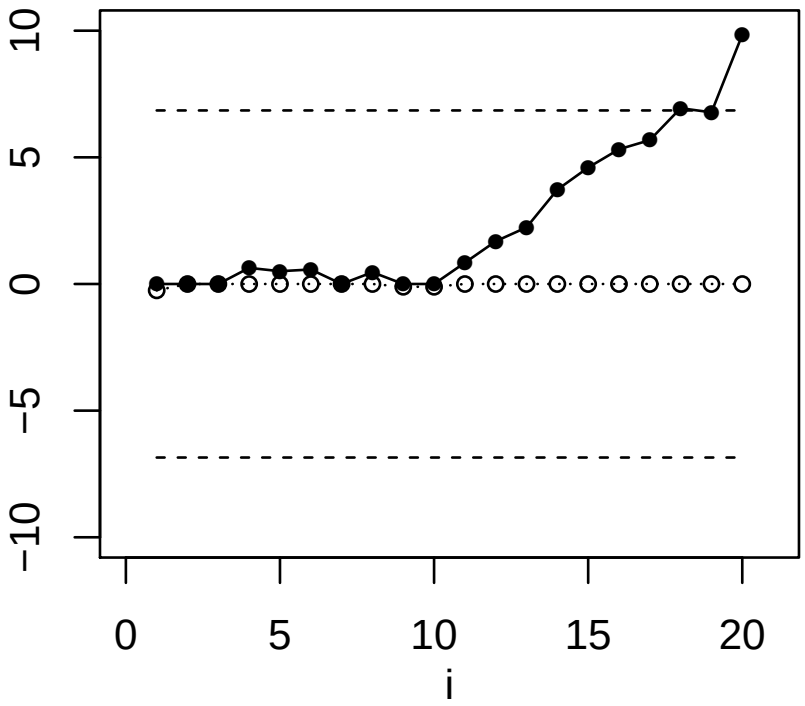
Assume that the diastolic blood presure of a healthy person follows the distribution $N(\mu_0, \sigma^2) = N(75, 4^2)$.

$$U = \mu_0 + 3\sigma = 87, \quad C = \mu_0 = 75, \quad L = \mu_0 - 3\sigma = 63.$$



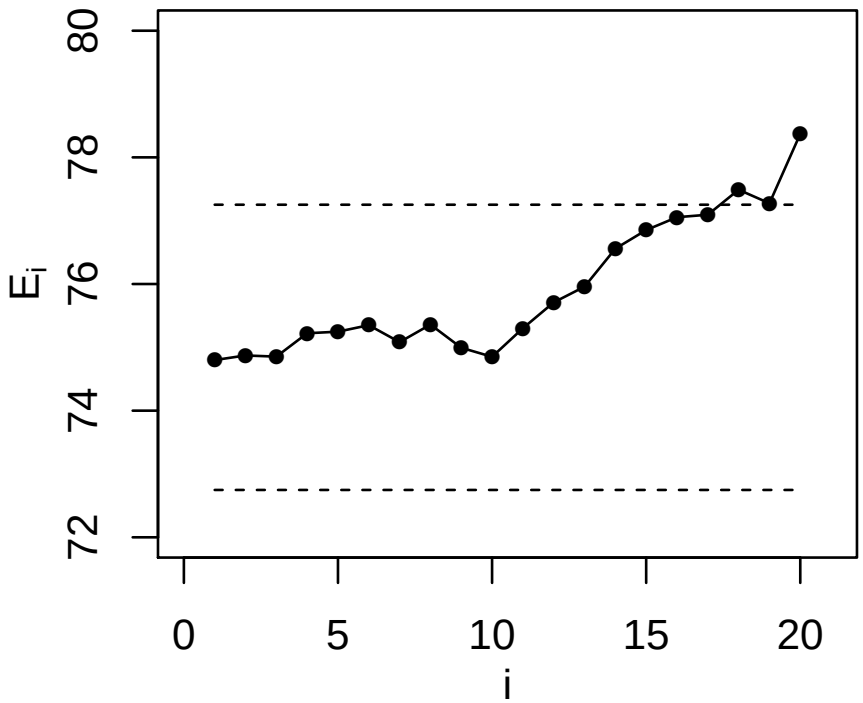
CUSUM Chart

- $$\begin{aligned} C_i^+ &= \max\left(0, C_{i-1}^+ + (X_i - \mu_0)/\sigma - k\right) \\ C_i^- &= \min\left(0, C_{i-1}^+ + (X_i - \mu_0)/\sigma + k\right) \end{aligned}$$
- It gives a signal if $C_i^+ > h$ or $C_i^- < -h$.



EWMA Chart

- $E_i = \lambda X_i + (1 - \lambda) E_{i-1}$
- It gives a signal if $E_i > \mu_0 + \rho\sigma\sqrt{\lambda/(1-\lambda)}$ or $E_i < \mu_0 - \rho\sigma\sqrt{\lambda/(1-\lambda)}$.

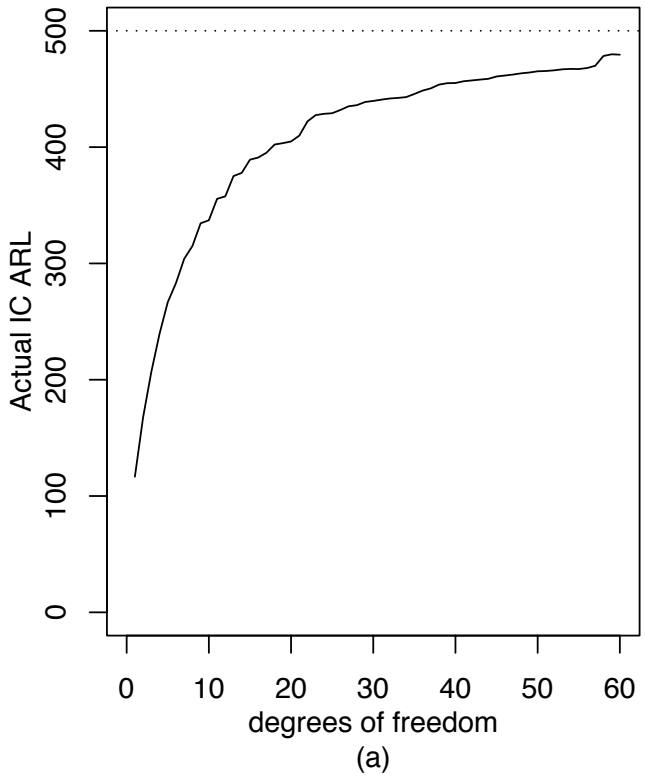


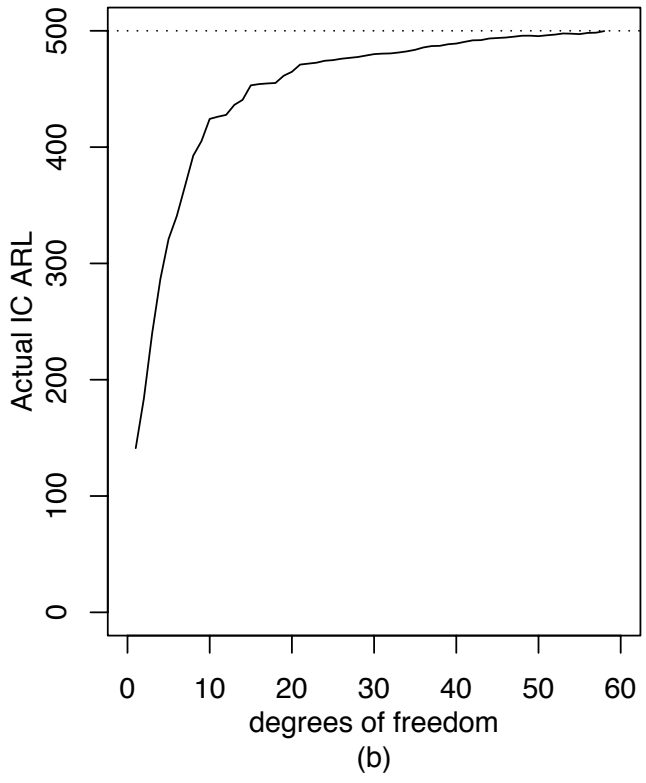
Performance Measures

- IC average run length ARL_0: average number of time points from the beginning of monitoring to a signal when the process is IC.
- OC average run length ARL_1: average number of time points from a process shift to a signal after the process becomes OC.
- ARL_0 is similar to the Type-I error probablity
- ARL_1 is related to the power of a test.
- Usually, ARL_0 is pre-specified at a level (e.g., 200). Then, a chart performs better if its ARL_1 is smaller for detecting a shift of a given size.

Assumptions

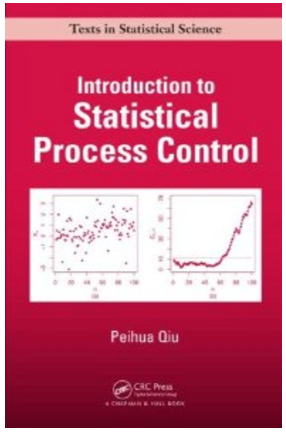
- Observations are independent and identically normally distributed
- Assume $ARL_0 = 500$ and the IC distribution is $N(0, 1)$, but the true distribution is the standardized version of the chi-square (plot (a)) or t (plot (b)).





Recent SPC Methods

- Univariate/multivariate nonparametric SPC
  - charts using ranks (Qiu and Hawkins 2001,2003, Qiu and Li 2011a,b, Qiu and Zhang 2015)
  - charts by log-linear modeling (Qiu 2008)
  - Profile monitoring (Qiu, Zou and Wang 2010, Paynabar, Qiu and Zou 2016)
  - Overviews (Qiu 2017)
- Qiu, P. (2014), *Introduction to Statistical Process Control*, Chapman & Hall/CRC.



Dynamic Disease Screening System (DDSS)

Motivating Example

- SHARe Framingham Heart Study of NHLBI.
- Many residents at Framingham MA were involved.
- Major risk factors of cardiovascular diseases: blood pressure, total cholesterol level (TCL), smoking, obesity, ...
- Identify patients with irregular longitudinal patterns of the disease risk factors as early as possible.
- Disease early detection and prevention
- Dynamic screening (DS) problem

DS Problem

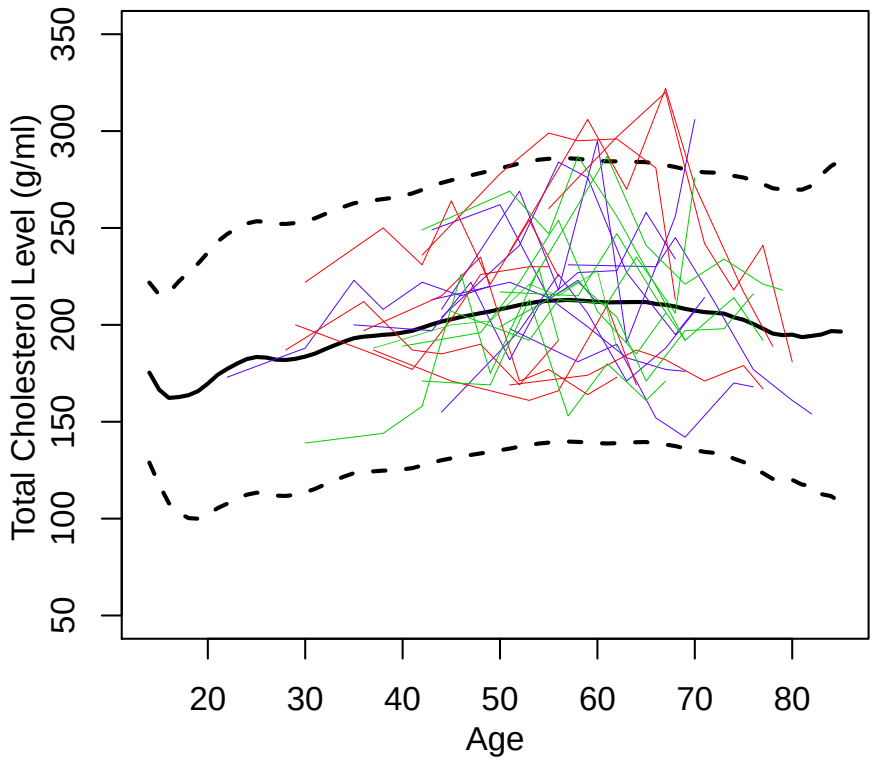
- DS problem is popular
- Most products (e.g., airplanes, cars) are checked regularly or occasionally about certain variables related to their quality and/or performance.
- If the observed values of a product are significantly worse than the values of a typical well-functioning product of the same age, then some adjustments or interventions should be made to avoid unpleasant consequences.
- No existing statistical methods for solving the DS problem.

Statistical Process Screening System

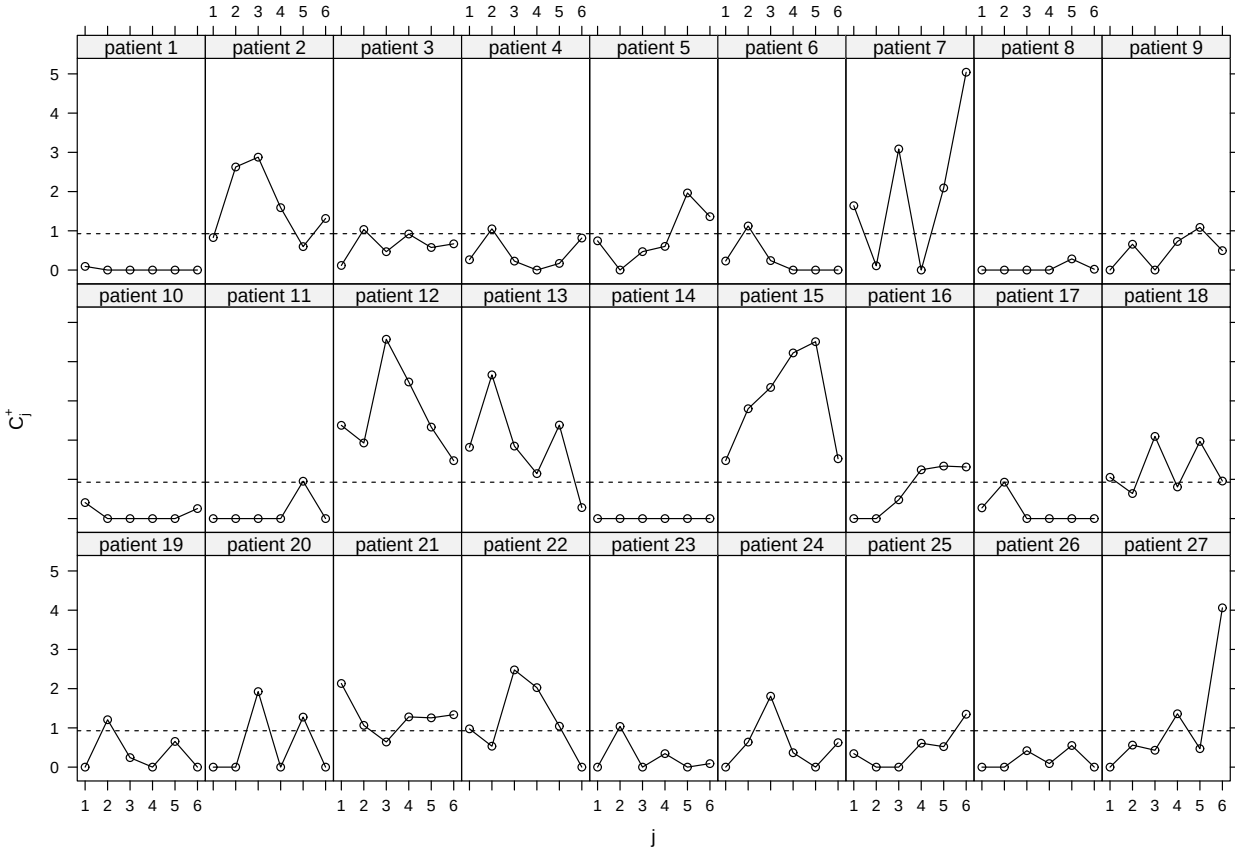
- Estimate regular longitudinal pattern from an IC dataset
- Standardize observations of a new subject to monitor
- Monitor the standardized observations by a control chart

SHARe Framingham Heart Study

- $m = 1028$ non-stroke patients (IC data)
- 27 stroke patients (new subjects)
- each patient was followed 7 times (i.e., $J = 7$)
- y is the total cholesterol level (in mg/100ml)



SHARe Framingham Heart Study (con'd)

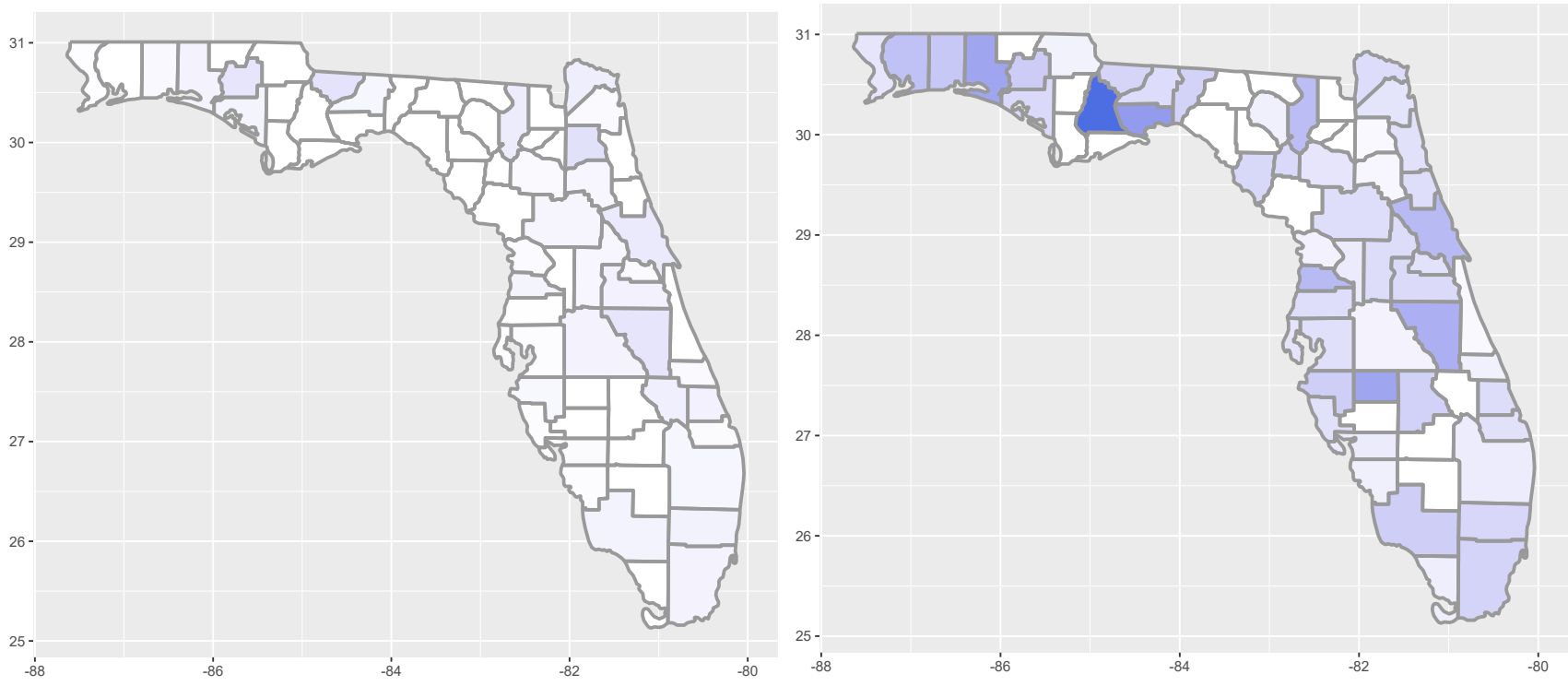


- CI approach: 6 out of 27 stroke patients got signals
- DDSS approach: 22 out of 27 stroke patients got signals; 131 out of 1028 non-stroke patients got signals

References

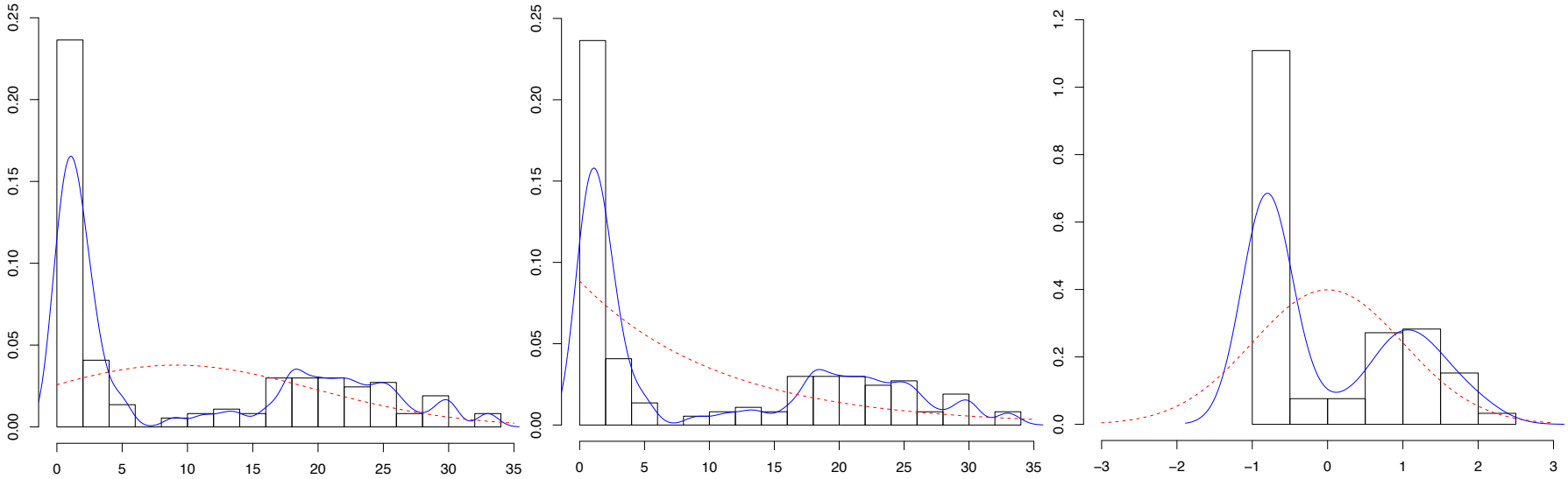
- Qiu, P., and Xiang, D. (2014), "Univariate Dynamic Screening System: An Approach For Identifying Individuals With Irregular Longitudinal Behavior," *Technometrics*, **56**, 248–260.
- Qiu, P., and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system," *Statistics in Medicine*, **34**, 2204–2221.
- Li, J., and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data," *IIE Transactions*, **48**, 772–786.
- Qiu, P., Zi, X., and Zou, C. (2017), "Nonparametric Dynamic Curve Monitoring," *Technometrics*, in press.

Spatio-Temporal Disease Surveillance



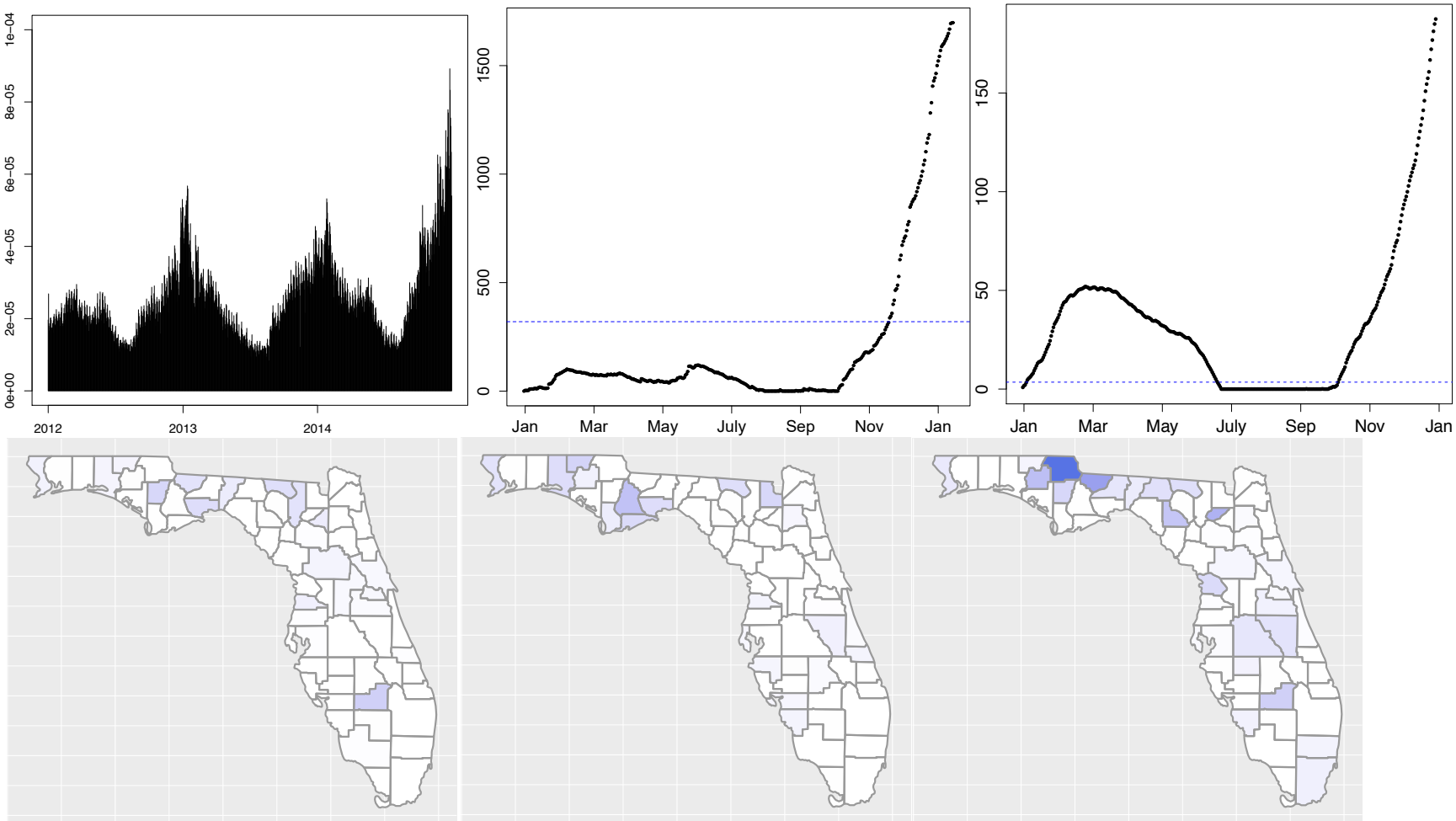
- Observed influenza-like illness (ILI) incidence rates in Florida on 06/01/2012 (left) and 12/01/2012 (right).

Spatio-Temporal Disease Surveillance (con'd)



- Density histogram of the ILI incidence in the Lee county of Florida in 2013.
- Seasonality, spatio-temporal correlation, non-normality

Spatio-Temporal Disease Surveillance (con'd)



- Top panels: Observed ILI incidence rates in Florida during 2012-2014, our chart, conventional CUSUM chart.
- Bottom panels: Observed ILI incidence rates in Florida at the county level on November 20 in year 2012, 2013, and 2014.

Concluding Remarks

- Statistical process control charts provide an important tool for monitoring processes with big data.
- Statistics overall plays a key role in scientific research.
- But, we need to use statistics properly. Otherwise, statistics can provide misleading results.
- Current statistical knowledge is very limited. New statistical methods are needed.