



Big data and medical research in China

Luxia Zhang, MD, MPH

**PKU, Center for Data Science in Health and Medicine
Renal Division, Peking University First Hospital
Peking University Institute of Nephrology**



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

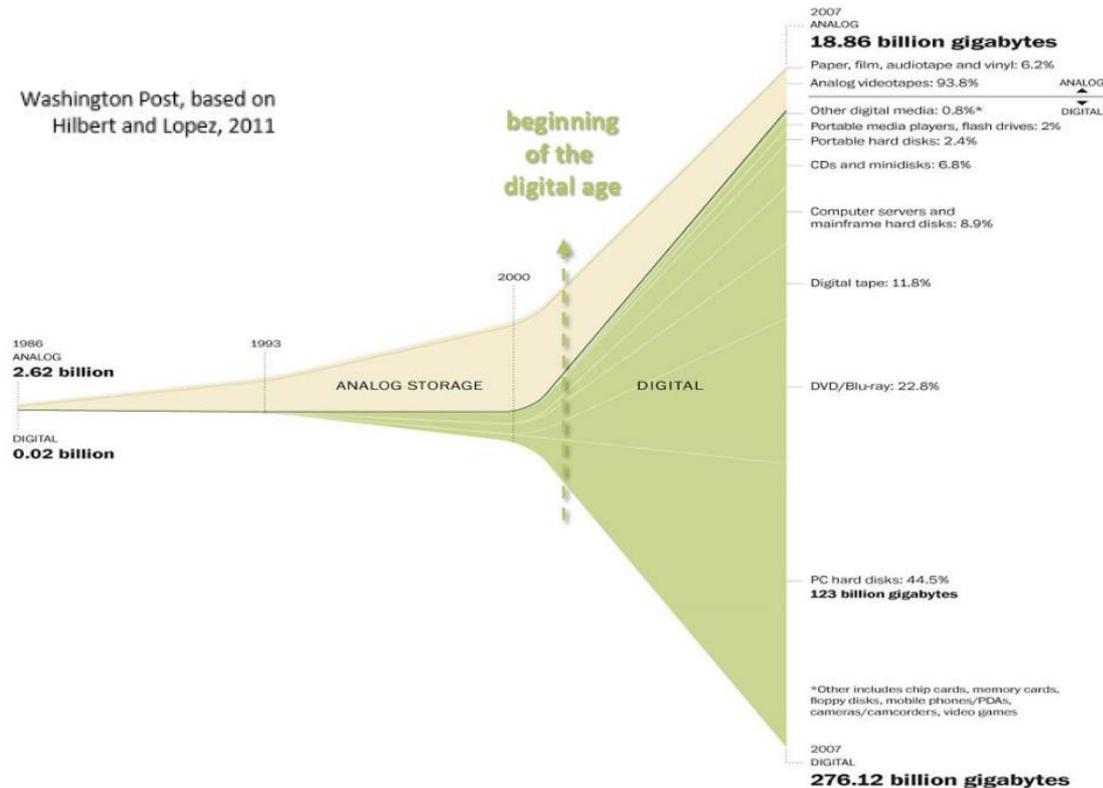
Outline

- **Big data in medicine - a national strategy in China;**
- **From kidney disease to big data center;**
- **Big data and “traditional medical research”.**



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

What is big data?

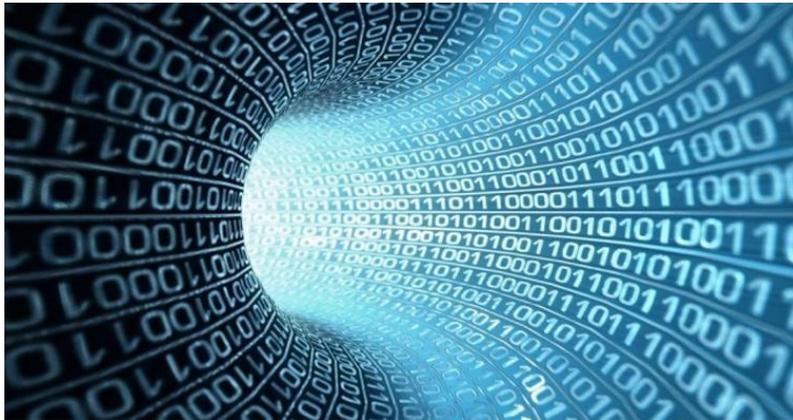


“There is little doubt that the quantities of data now available are indeed large, but that’s not the most relevant characteristic of this new data ecosystem.”

Volume + Variety, Velocity, Variability, Veracity



健康医疗大数据研究中心
Center for Data Science in Health and Medicine



**Data is not the purpose
- To extract value from data**



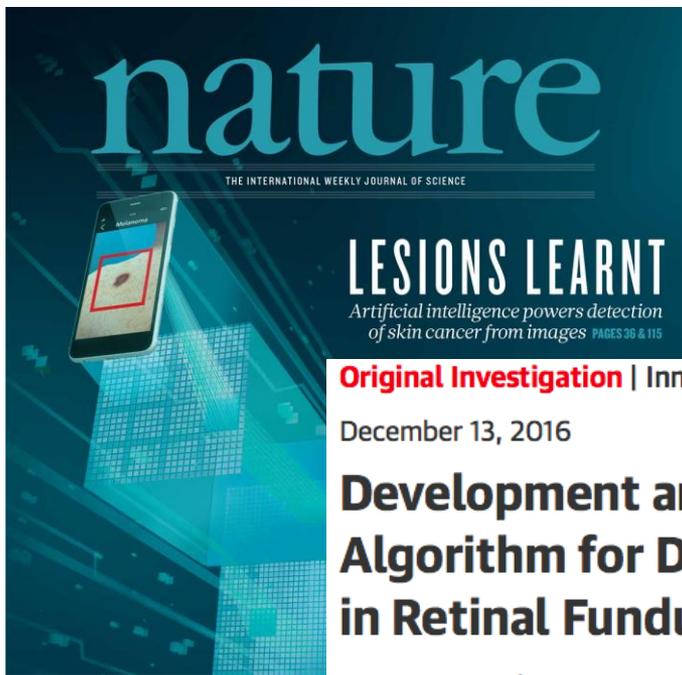
The NEW ENGLAND JOURNAL of MEDICINE

Predicting the Future — Big Data, Machine Learning, and Clinical Medicine

Ziad Obermeyer, M.D., and Ezekiel J. Emanuel, M.D., Ph.D.

By now, it's almost old news: big data will transform medicine.

Machine learning has become ubiquitous and indispensable for solving complex problems in most sciences. The same methods will open up vast new possibilities in medicine.



Original Investigation | Innovations in Health Care Delivery

December 13, 2016

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD¹; Lily Peng, MD, PhD¹; Marc Coram, PhD¹; [et al](#)

» [Author Affiliations](#)

JAMA. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

Opinion

INNOVATIONS IN HEALTH CARE DELIVERY

Machine Learning and the Profession of Medicine

Must a physician be human?

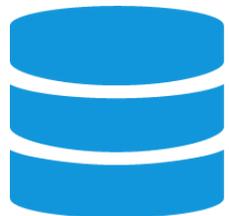


“ After 30 yrs, physicians will loose job, hospital will disappear in China. ”

“ There is one thing for sure though, which is Jack Ma will disappear after 50 yrs.”

The rapid development of HIS in China

Everything is data



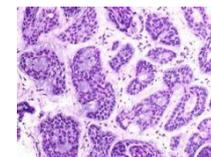
EHR



PACS



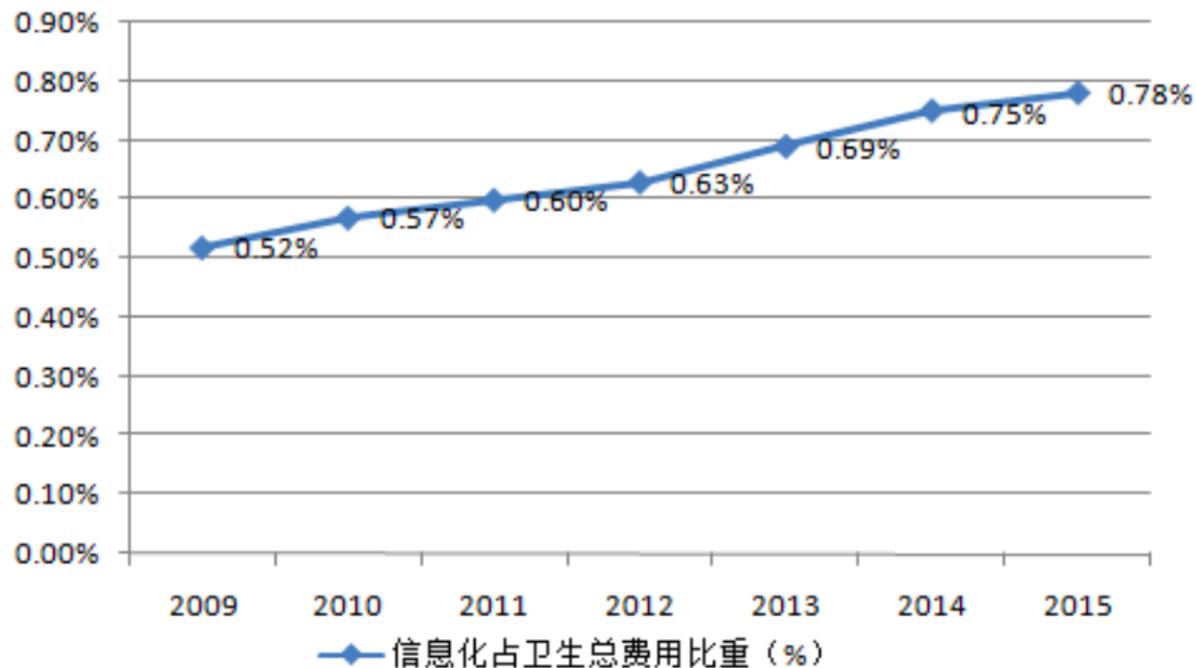
ECG



Pathology



Multi-omics



Currently, above 90% of hospitals in China are utilizing integrated hospital information system for both in-patient and out-patient care.

Big data in medicine – a national strategy of China

2014.3

Big data was prominently highlighted in the Report on the Work of Government

2016.6

“The application and development of big data in the healthcare sectors” by the State Council

2016.8-10

Promoting the application of big data in health and medicine -National Health Conference, Healthy China 2030

2016.10

A pilot program of national and regional centers and industrial parks of big data in health and medicine



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

However, currently in China ...

- **More discussion, less practice;**
- **More enthusiasm about building data warehouse and data analytics methods, less focused on application in health and medicine.**





健康医疗大数据研究中心
Center for Data Science in Health and Medicine

Outline

- **Big data in medicine - a national strategy in China;**
- **From kidney disease to big data center;**
- **Big data and “traditional medical research”.**

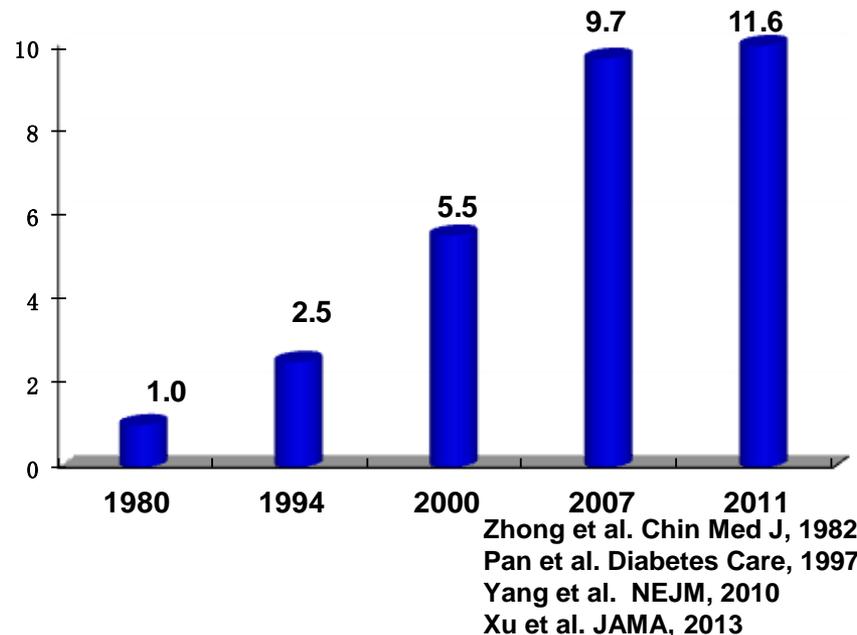


The debut of big data in nephrology in China

- Globally, diabetes is the No1. cause of ESRD (1/3)
- 1999 Dialysis registry in China : GN 49.9% , DKD 13.3% ;
- 2012 Dialysis registry in China : GN 57.3% , DKD 16.4%.

Chin J Nephrol, 1999
Clinical Nephrology, 2010 (S20-S22)

Prevalence of DM in China



- **Lack of population-based studies, no national surveillance system regarding kidney disease;**
- **Current available data – GN or ESRD**
 - **ESRD : <1% of CKD ;**
 - **rapidly increased prevalence of diabetes?**

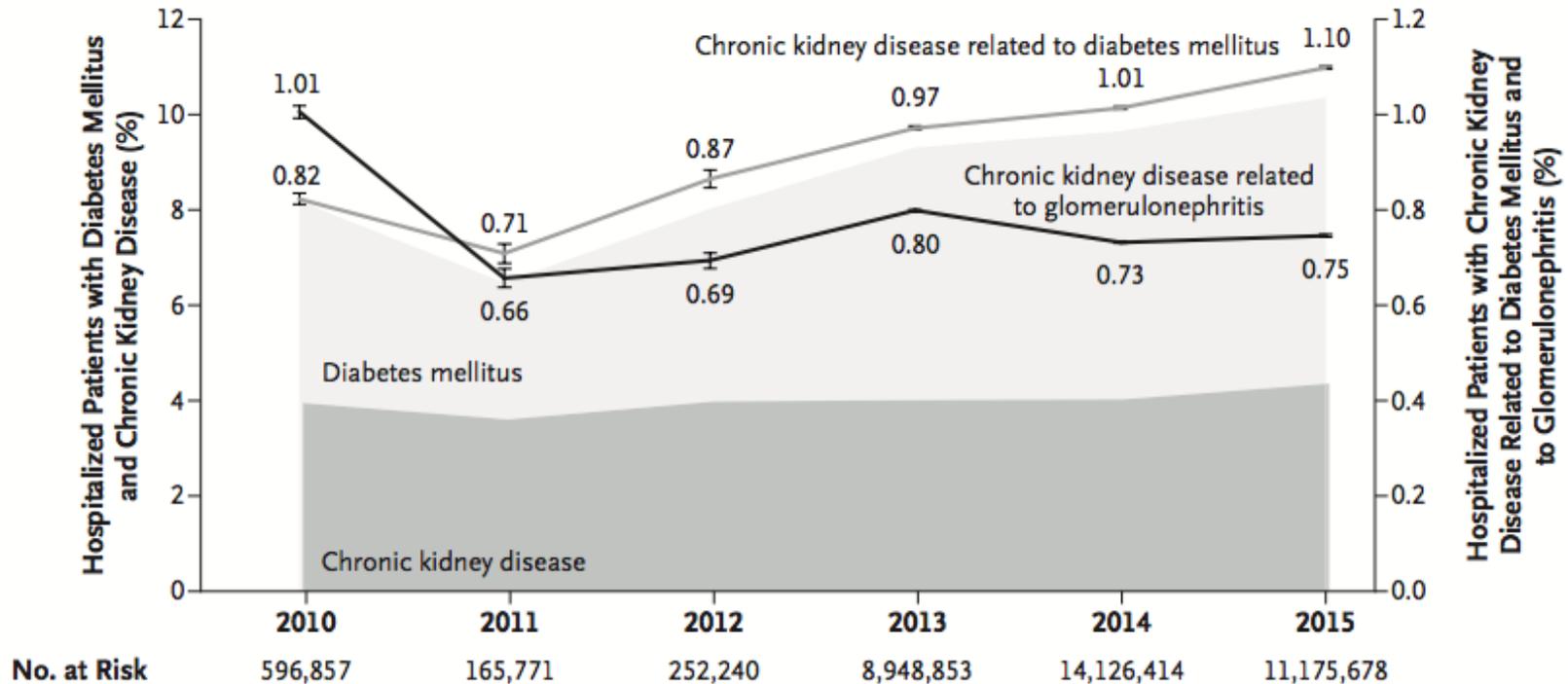




China-WHO Biennial Collaborative Projects 2014-2015 Research on the Monitoring, Evaluation and Improvement of National Medical Service in China

Trends in Chronic Kidney Disease in China

- **General population : a representative sample of adult population in China; n=47 204; 2009**
- **Hospitalized population: 35 million patient from a national in-patient database were used; 2010-2015.**

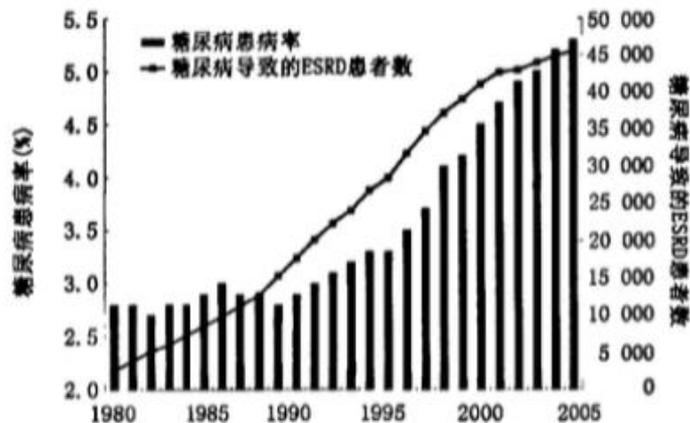


- Study report – Ministry of Health, WHO
- **CALL FOR ACT**
 - to mitigate the epidemic of advanced diabetic target organ damage, after the failure to control the epidemic of diabetes;
 - to promote education, clinical training and research regarding DKD in China.

从流行病学的角度探讨我国糖尿病肾病的发病趋势及对策

张路霞 王海燕

糖尿病肾病已成为全球慢性肾脏病 (CKD) 及其导致的终末期肾病 (ESRD) 的主要构成原因。每年三月份的第二个星期四是“世界肾脏日”，继 2009 年“肾脏与高血压”的主题之后，今年将“糖尿病肾病”作为活动主题。随着亚洲国家生活方式西方化，糖尿病发病率呈现显著上升趋势，我国亦不例外。1980 年我国 14 省市的调查显示糖尿病患病率约为 1%，1994 - 1995 年的结果显示糖尿病患病率为 2.5%，到 2000 - 2001 年，该数字攀升至 5.5%^[1]。值得注意的是，前两次调查筛查手段为空腹血糖结合口服葡萄糖耐量试验 (OGTT)，而第三次调查仅检测了空腹血糖，因此对于患病率



Big data

- to provide answers difficult to be provided by traditional studies
- excellent to show “trends”



中国肾脏疾病数据网络
China Kidney Disease Network

China Kidney Disease Network CK-NET

- **To integrate various source of data in renal disease;**
- **To be combined with cutting-edge analytics;**
- **To provide data-based evidence for healthcare policy, to strengthen academic research, and to promote effective disease management, in the field of nephrology.**



Safety of IV Iron in Hemodialysis
Cognitive Decline in Hemodialysis
Safety of Warfarin in Elderly Patients
China Kidney Disease Network Annual Data Report (e-Supplement)

The China Kidney Disease Network (CK-NET): “Big Data—Big Dreams”



Related Content: June e-supplement

In an e-supplement to this issue of *AJKD* appears the first report of the newly established China Kidney Disease Network (CK-NET),¹ which intends to become a comprehensive chronic kidney disease (CKD) surveillance system for that country. In addition to being the most populous country in the world

evolving to keep up with changing patterns of and treatments for kidney disease. Uniquely, it also supplies de-identified datasets and a data-merging service to researchers and organizations through formal data request procedures. Other prominent examples of well-developed surveillance systems in the United States include the SEER Cancer Registry⁷ and the National CDC-CKD Surveillance Sys-

“Big data, big dreams”

- I. Pts with CKD in hospitalized population
- II. Treatment of pts with CKD
- III. Disease causes and geographic distribution of pts with CKD
- IV. Comorbid conditions of pts with CKD
- V. Cardiovascular disease in pts with CKD
- VI. In-hospital mortality of pts with CKD

- VII. Length-of-stay of pts with CKD
- VIII. Costs of pts with CKD
- IX. Other aspects of pts with CKD
- X. AKI
- XI. Kidney transplant waiting list

Over 10-year's experience of population-based study in nephrology, based on huge amounts of data, using descriptive statistics — the next step ?

- Questions to be answered: factors contributed to the changing spectrum of kidney disease? Medical migration for pts with kidney disease?**

“Out”

Gansu	33.8%
Tianjin	15.4%
Hebei	14.5%



“In”

Beijing	25.5%
Shanghai	21.0%
Shanxi	9.7%

- Team : nephrologists, epidemiologists, data scientists**



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

Ongoing projects

- **Socio-economic, and environmental factors vs. spectrum of kidney disease: Tensor Decompositions**
- **Medical migration of pts with kidney disease: Network Diffusion Model.**
- **Urinary sediments analyses: CNN**



Expand to other specialties, or to develop a multidisciplinary platform @ PKU?

Peking University, Center for **Data Science** in **Health and Medicine**

August 28, 2016



Peking University, Center for Data Science in Health and Medicine

- **To integrate resources in health and medicine, data science, computer science, humanities and social sciences in Peking University;**
- **To facilitate the efforts of Ministry of Health, to translate cutting-edge technology to clinical practice, to launch high-quality researches, to establish interdisciplinary education and training.**

National Institute of Big Data in Health and Medicine @ PKU

Question-based strategy, close collaborations with affiliated hospitals of PKU



中国队列共享平台
China Cohort Consortium

Center for Data Science in Health and Medicine

China Medical Data Sharing System (C-MEDs)



- **Two data-sharing platform to be released next month;**
- **Launched by School of Public Health and DSHM;**
- **Various sources of data**
 - **epidemiological studies;**
 - **data from pharmaceutical company;**
 - **basic health insurance, regulatory data;**
 - **real world data (based on specialties)**



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

Outline

- **Big data in medicine - a national strategy in China;**
- **From kidney disease to big data center;**
- **Big data and “traditional medical research”.**



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

Statistics	Computer Science	Meaning
estimation	learning	using data to estimate an unknown quantity
classification	supervised learning	predicting a discrete Y from $X \in \mathcal{X}$
clustering	unsupervised learning	putting data into groups
data	training sample	$(X_1, Y_1), \dots, (X_n, Y_n)$
covariates	features	the X_i 's
classifier	hypothesis	a map from covariates to outcomes
hypothesis	—	subset of a parameter space Θ
confidence interval	—	interval that contains unknown quantity with a prescribed frequency

SPRINGER TEXTS IN STATISTICS

All of Statistics

A Concise Course
in Statistical
Inference

Larry Wasserman



- **Statisticians and computer scientists often use different language for the same thing.**
- **“Machine learning is for Computer Science majors who couldn’t pass a Statistics course.”**
- **It was the in-depth collaboration between statisticians and computer scientists brought the considerable development in machine learning in the last decades.**

The harmony is only achieved when the two forces combined in balance





- **Big data in health and medicine: improving the overall health and wellness of patients and the population**
 - not different with the aim of “traditional research”;
 - provide real-world evidence, and maximize efficiency;
- **Basic concepts of epidemiology should be valued: eg. bias ;**
- **The essence of big data in medicine is medicine**
 - question-driven, instead of methodology-driven
 - seeking for causality instead of correlation.



健康医疗大数据研究中心
Center for Data Science in Health and Medicine

Big data in health and medicine, we need...

- **To initiate discussion regarding law and ethics;**
- **To build standard clinical health terminology in Chinese;**
- **More involvement of health professionals in the area**
 - **to solve problems in the real world.**