

美国临床医疗大数据分析简介

Healthcare Analytics using EHR in US

徐华

德州大学休斯顿生物医学信息学院

School of Biomedical Informatics, UTHealth

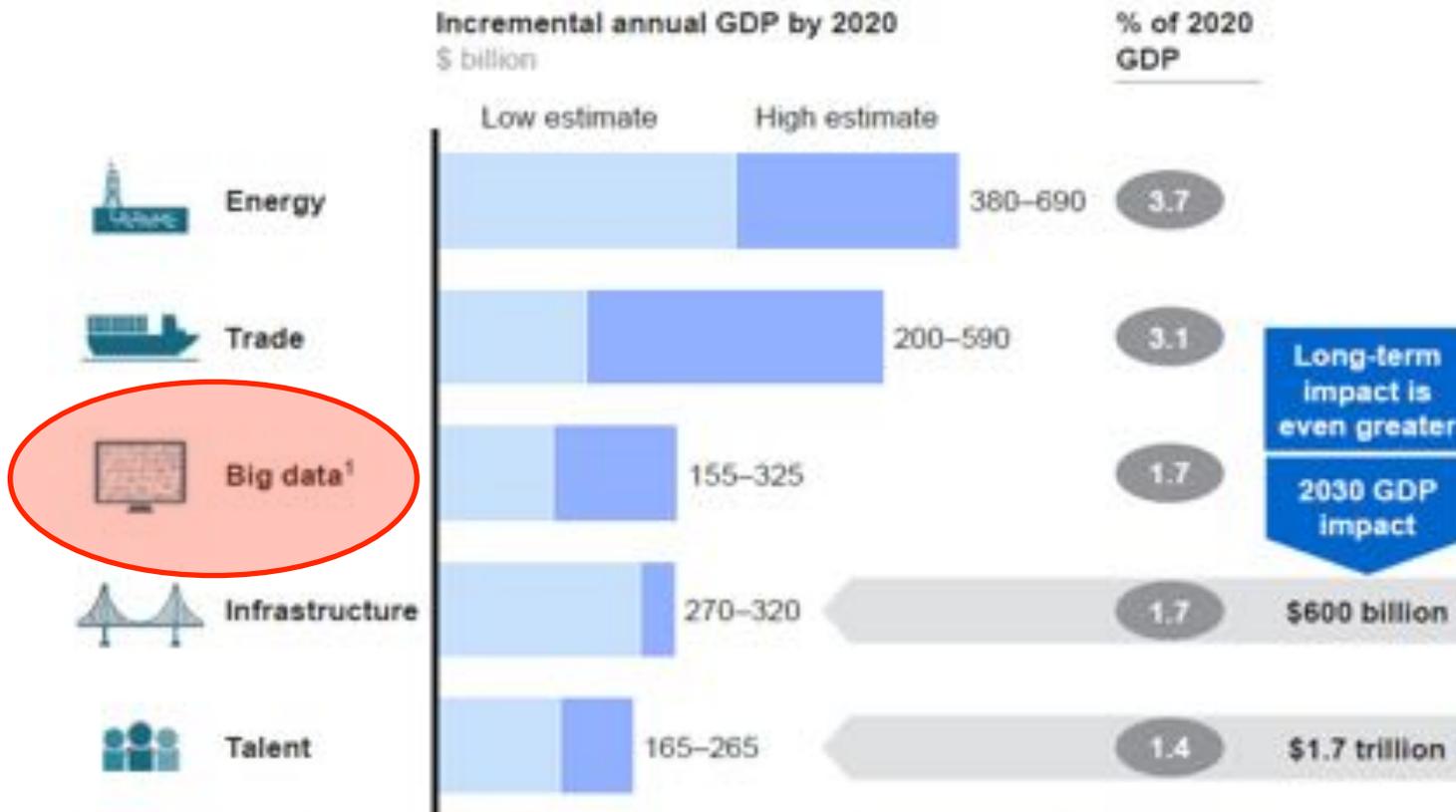
纲要

- 美国医疗大数据现状介绍
 - 医疗大数据背景介绍
 - 谁在促进医疗大数据产业
 - 临床医疗大数据分析的主要技术
- 美国医疗大数据公司介绍
- 促进中国临床医疗大数据分析—OHDSI中国介绍

医疗大数据背景介绍

影响美国GDP的五大因素

Each of the game changers could substantially raise US GDP by 2020



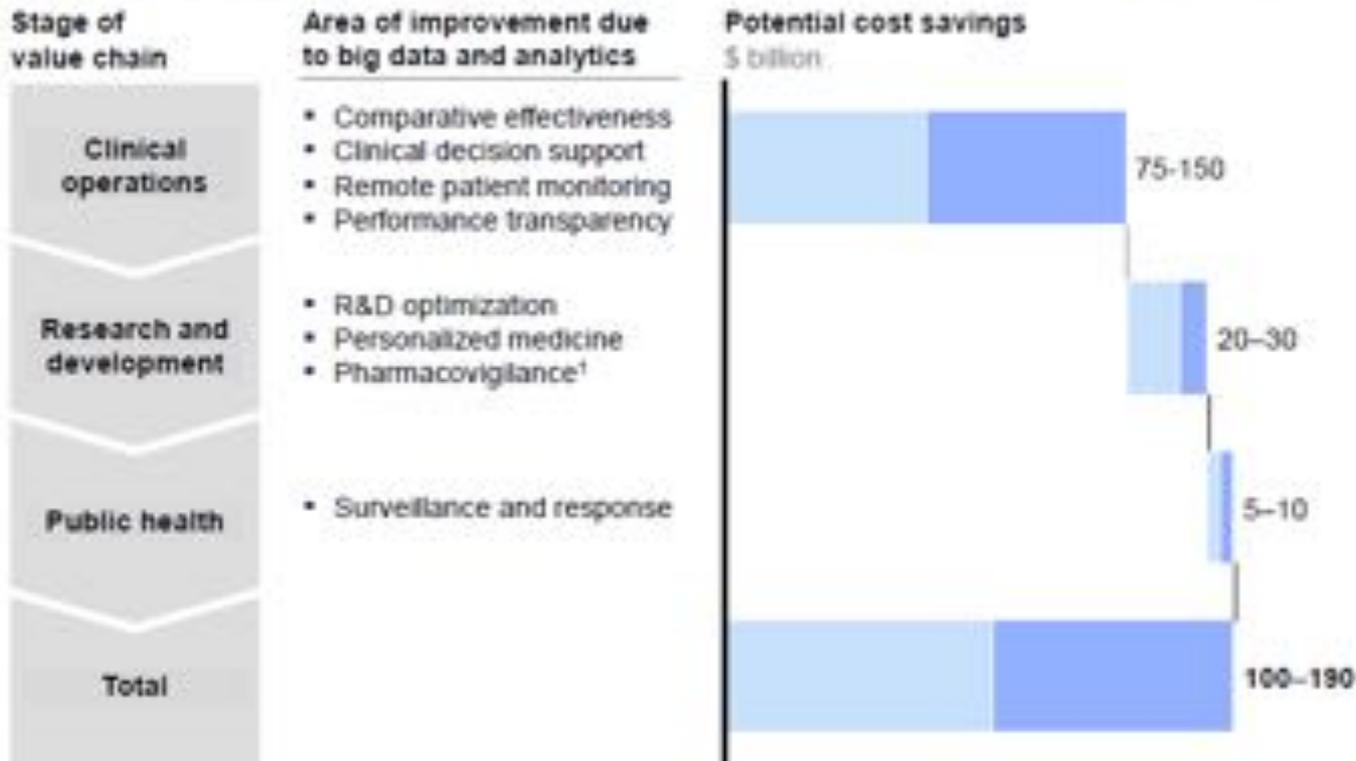
¹ Figures reflect additional GDP in retail and manufacturing sectors only. Big data could also produce cost savings in government services and health care (\$135 billion–\$285 billion), but these do not directly translate into additional GDP. NOTE: These figures are based on a partial-equilibrium analysis that estimates only first-order effects and therefore cannot be summed to calculate the full economic impact.

SOURCE: Economist Intelligence Unit, IHS Global Insight, McKinsey Global Institute analysis

医疗大数据分析应用领域

Big data analytics can generate up to \$190 billion annually in health-care cost savings by 2020

Low estimate
High estimate

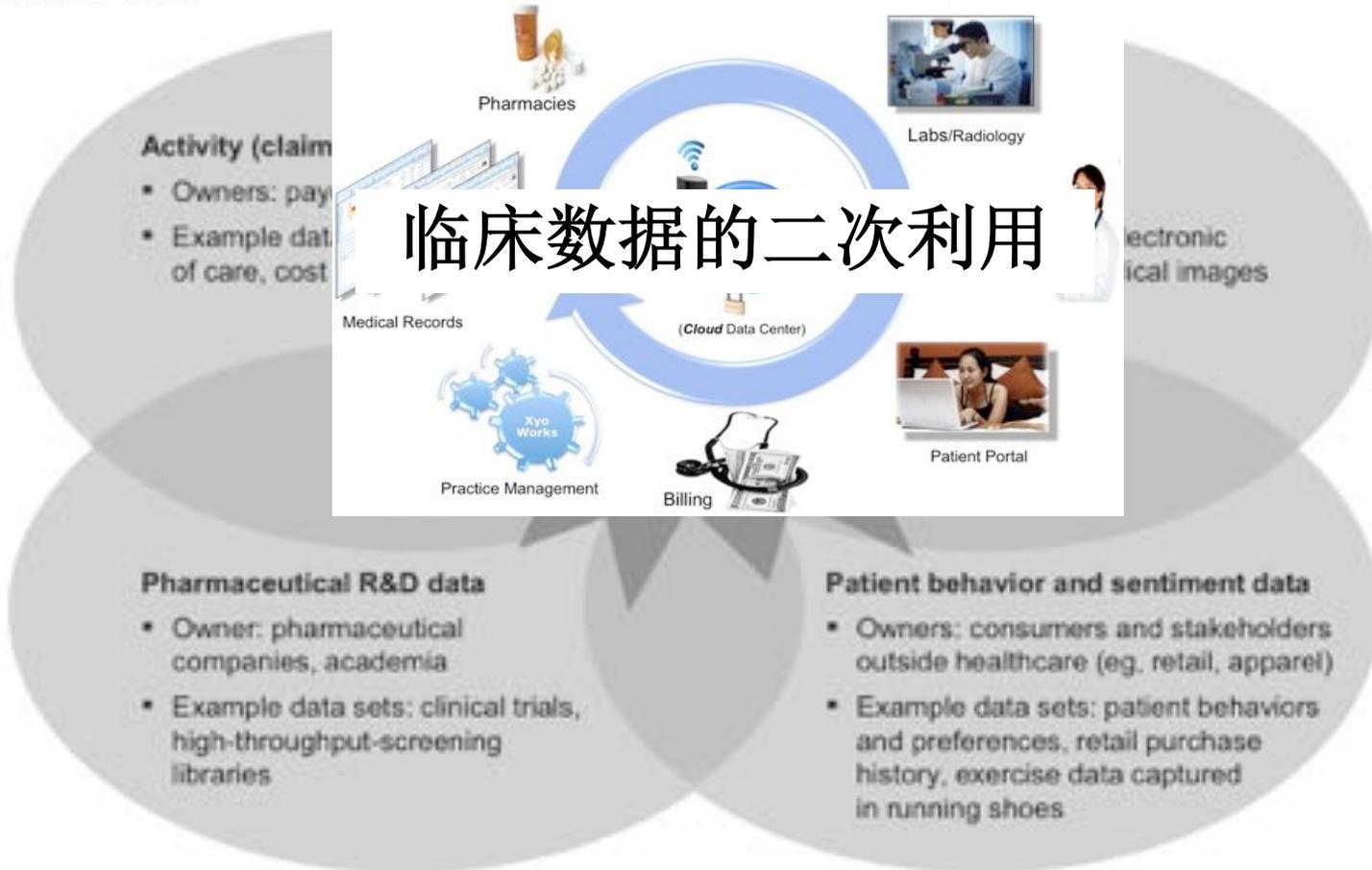


¹ Also known as drug safety, "pharmacovigilance" is defined by the World Health Organization as the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem.

SOURCE: World Health Organization; McKinsey Global Institute analysis

医疗大数据的组成部分

Exhibit 2: Primary data pools are at the heart of the big-data revolution in healthcare.



相关报告

Center for US Health System Reform
Business Technology Office



The 'big data'
revolution in healthcare

Accelerating value and innovation

McKinsey & Company

Artificial Intelligence for Health and Health Care

December 2017

JSR-17-Task-002

JASON
The MITRE Corporation

谁在促进医疗大数据产业？

政府 — 机构和法案

- 美国复苏与再投资法案/American Recovery and Reinvestment Act (ARRA) 2009
 - 经济与临床健康信息技术法案 / Health Information Technology for Economic and Clinical Health Act (HITECH)
- 患者保护与平价医疗法案
 - Patient Protection and Affordable Care Act of 2010
- 国家HIT协调员办公室
 - Office of the National Coordinator for Health IT(ONC)
- 美国医疗保健研究与质量局
 - Agency for Healthcare Research and Quality (AHRQ)
- 联邦医疗保险和联邦医疗辅助计划服务中心
 - Centers for Medicare and Medicaid Services (CMS)

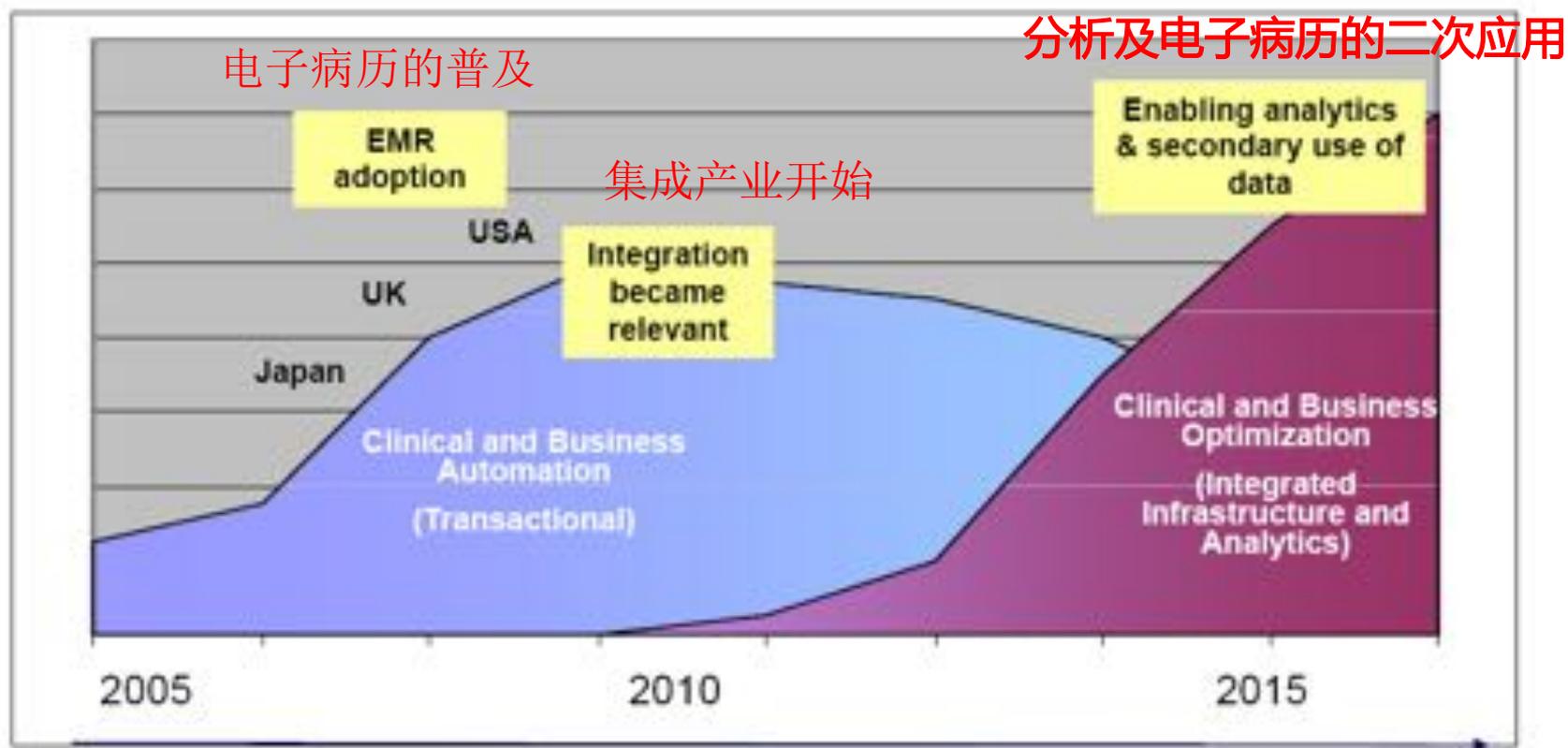
Centers for Medicare and Medicaid Services (CMS)

- 有效利用电子病历法案 — “Meaningful Use” incentive program by CMS — Utilizing EHRs to improve health outcomes, three Stages
 - Stage 1 – data capture and sharing
 - Stage 2 – advanced clinical processes
 - Stage 3 - improved outcome
- 基于医疗价值的报销模式 — Value-based program by CMS
 - Transfer from Fee-For-Service payment models (quantity) to Value-Based Reimbursement models (quality)
 - Better care for individuals and populations, with lower cost
 - Example: Readmission rates (再次住院率)
 - 30% percent of VBR by 2016; 50% percent of VBR by 2018

Healthdata.gov

- 向研究机构，开发者和研究者开放对联邦数据库的访问权限
- 全国住院病人数据库 — National Inpatient Sample (NIS)
 - 所有美国住院病人抽样
 - 700万人次 / 年
 - 大于20年数据
 - 获取要求： 15分钟网络课程 + DUA
 - 费用： \$50-350/year

工业界 - 分析服务是行业发展趋势



"CDOs have already made substantial investments in information systems. Although investment in clinical systems will continue, it is essential for CDOs to work on optimizing the value from them"

美国十大医疗数据分析公司

- IBM
- Optum
- Oracle
- Verisk Analytics
- MEDai's Health
- MedeAnalytics
- McKesson
- Truven Health Analytics
- Allscripts Healthcare Solutions
- Cerner

可以购买的商业化医疗数据集

- 大规模商业化医疗数据集
 - OPTUM (3800万病人, 从2005年开始)
 - Humedica (2500万病人, 2008-2013)
 - Cerner Health Facts (5000万病人, 从2000年开始)
 - Truven MarketScan Commercial Claims and Encounters (CCAЕ) (1亿1700万病人, 从2000年开始)
- 去除敏感信息, 根据隐私保护法案HIPAA (Health Insurance Portability and Accountability Act)
 - 18 PHI: 姓名, 地址、城市、区、邮编等 (所有省级以下), 日期, 电话, 传真, 电子邮箱, 社会安全号, 病历号, 医疗保险号, 账号, 证书号, 车牌号, 仪器编号, 网址, IP地址, 生物识别 (指纹 / 声音), 脸部照片, 其它唯一识别号

学术界 — 大量的科研经费

- 电子病历和基因组学研究， eMERGE – electronic medical records and genomics
 - 2007开始至今9年
 - 大于3亿美元的投入
- 精准医疗计划， Precision Medicine Initiative
 - 2016年开始
 - 4-6亿美元投入
- 全美临床数据网络， PCORNet
 - 2013年开始
 - 大于2亿美元投入
 - 超过1亿人口的临床数据
- OHDSI – The Observational Health Data Sciences and Informatics
 - >20 国家
 - >6.6亿人口数据

重量级文章关于临床电子病历数据的二次利用

Significant papers about secondary use of EHRs

nature
REVIEWS GENETICS

Mining electronic health records: towards better research applications and clinical care

Peter B. Jensen¹, Lars J. Jensen¹ and Soren Brunak^{1,2}

Abstract | Clinical data describing the phenotypes and treatment of patients represents an underused data source that has much greater research potential than is currently realized. Mining of electronic health records (EHRs) has the potential for establishing new patient-stratification principles and for revealing unknown disease correlations. Integrating EHR data with genetic data will also give a finer understanding of genotype-phenotype relationships. However, a broad range of ethical, legal and technical reasons currently hinder the systematic deposition of these data in EHRs and their mining. Here, we consider the potential for furthering medical research and clinical care using EHR data and the challenges that must be overcome before this is a reality.

JAMA The Journal of the
American Medical Association

Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing

Harvey J. Murff, MD, MPH

Fern FitzHenry, RN, PhD

Michael E. Matheny, MD, MSc, MPH

Nancy Gentry, RN

Kristen L. Kotter, MS

Context Currently most automated methods to identify patient safety occurrences rely on administrative data codes; however, free-text searches of electronic medical records could represent an additional surveillance approach.

Objective To evaluate a natural language processing search-approach to identify postoperative surgical complications within a comprehensive electronic medical record.



The NEW ENGLAND
JOURNAL of MEDICINE

Azithromycin and the Risk of Cardiovascular Death

Wayne A. Ray, Ph.D., Katherine T. Murray, M.D., Kathi Hall, B.S.,
Patrick G. Arbogast, Ph.D., and C. Michael Stein, M.B., Ch.B.

ABSTRACT

BACKGROUND

Although several macrolide antibiotics are proarrhythmic and associated with an increased risk of sudden cardiac death, azithromycin is thought to have minimal cardiotoxicity. However, published reports of arrhythmias suggest that azithromycin may increase the risk of cardiovascular death.

JAMA The Journal of the
American Medical Association

Original Investigation

Association Between Blood Pressure Control and Risk of Recurrent Intracerebral Hemorrhage

Alessandro Biffi, MD, Christopher D. Anderson, MD, MMSc, Thomas W. K. Battey, BS, Alison M. Ayres, BA,
Steven M. Greenberg, MD, PhD, Anand Viswanathan, MD, PhD, Jonathan Rosand, MD, MSc

IMPORTANCE Intracerebral hemorrhage (ICH) is the most severe form of stroke. Survivors are at high risk of recurrence, death, and worsening functional disability.

OBJECTIVE To investigate the association between blood pressure (BP) after index ICH and risk of recurrent ICH.

DESIGN, SETTING, AND PARTICIPANTS Single-site, tertiary care referral center observational study of 1145 of 2197 consecutive patients with ICH presenting from July 1994 to December 2013. A total of 1145 patients with ICH survived at least 90 days and were followed up through December 2013 (median follow-up of 36.8 months [minimum, 9.8 months]).

Supplement
jama.com

CME Quiz a
jamanetw
CME Questi

nature
biotechnology

Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data

Joshua C Denny^{1,2}, Lisa Bastarache³, Maryryn D Ritchie⁴, Robert J Carroll⁵, Raquel Zink⁶, Jonathan D Mosley⁴, Julie R Field⁴, Jill M Palley^{4,5}, Andres H Ramirez⁷, Erica Bowton⁴, Melissa A Bassford⁴, David S Carnethon⁸, Peggy L Peissig⁷, Abel N Kho⁹, Jennifer A Pacheco⁹, Luke V Rasmussen¹⁰, David R Crosslin¹¹, Paul K Crane¹², Jyotishman Pathak¹³, Suzette J Bielski¹⁴, Sarah A Pendergrass¹, Hua Xu¹⁵, Lucia A Hindorf¹⁶, Rongling Li¹⁶, Teri A Manolio¹⁶, Christopher G Chute¹³, Rex I Chisholm¹⁷, Eric B Larson⁶, Gail P Jarvik^{11,12}, Murray H Brilliant¹⁸, Catherine A McCarty¹⁹, Iftekhar J Kullo²⁰, Jonathan L Haines²¹, Dana C Crawford²¹, Daniel R Massy²² & Dan M Roden^{1,2,3}

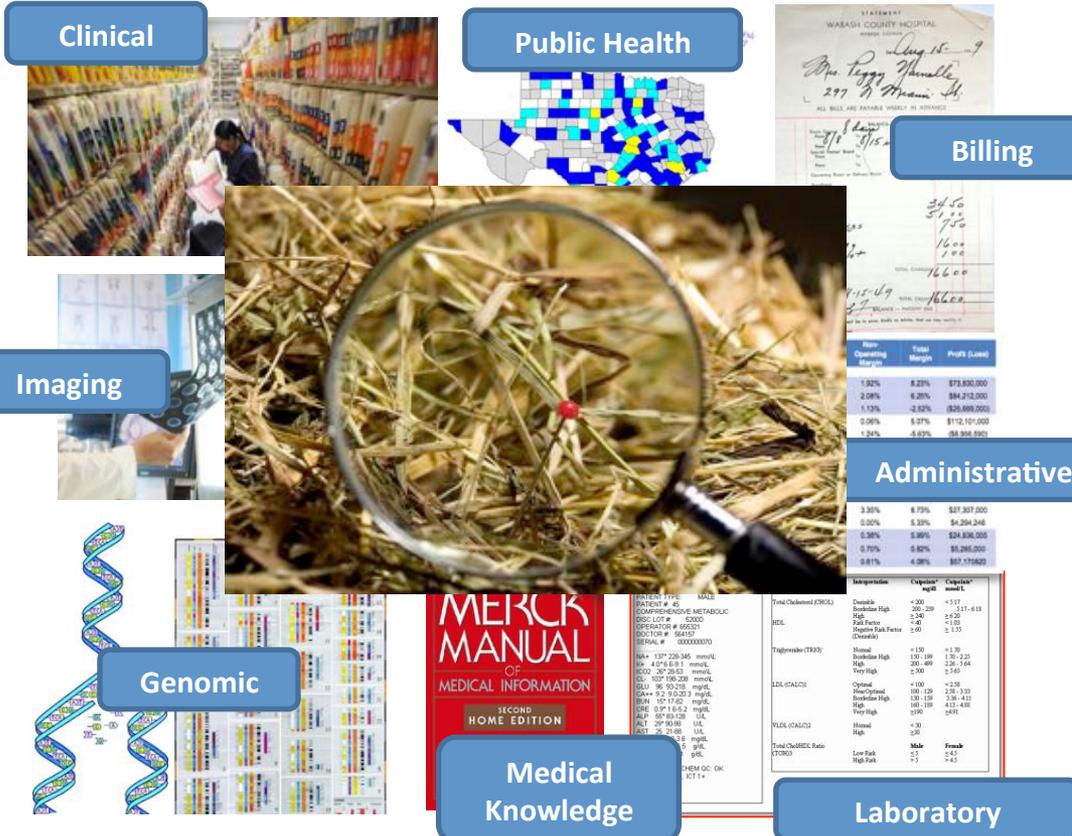
Candidate gene and genome-wide association studies (GWAS) have identified genetic variants that modulate risk for human disease; many of these associations require further study to replicate the results. Here we report the first large-scale application of the phenome-wide association study (PheWAS) paradigm within electronic medical records (EMRs), an unbiased approach to replication and discovery

large number of single variant-phenotype associations has led to the serendipitous identification of single loci associated with multiple diseases, or pleiotropy. Notable examples include variants at 9p21.3, which were associated initially with early myocardial infarction²³ and subsequently with intracranial aneurysm and abdominal aortic aneurysms²⁴; variants in the human leukocyte antigen (HLA) region and IL23A, which were associated initially with inflammatory bowel di-

临床医疗大数据分析的主要技术

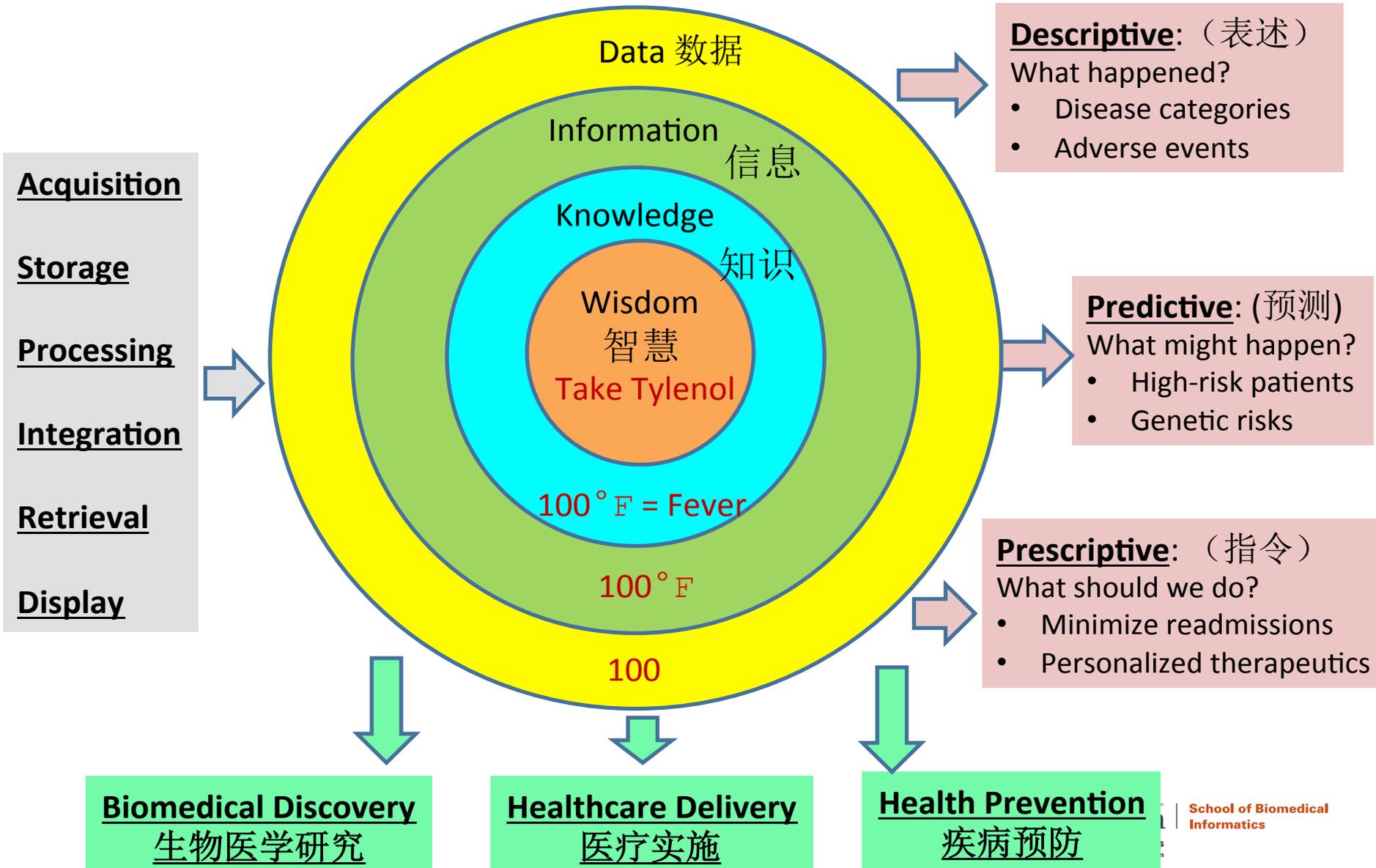
临床医疗数据分析的挑战

Challenges of EHR Analytics



- 不是为数据分析设计的
Not designed for research/analytics
- 数据问题: 复杂, 异构, 缺失, 互相矛盾...
EHR data: Complex, Heterogeneous, Incomplete, Discrepancy
- 知识域广
Require expertise in multiple disciplines such as medicine, biostatistics, epidemiology, and informatics

数据 —》信息 —》知识 —》智慧



生物医学信息学

Biomedical Informatics

Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health.

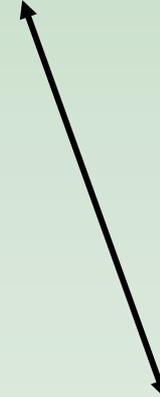


Biomedical Informatics in Perspective

生物医学信息学分类

Basic Research

Biomedical Informatics Methods,
Techniques, and Theories



Bioinformatics

Imaging
Informatics

Clinical
Informatics

Public Health
Informatics

Applied Research
And Practice

Molecular and
Cellular
Processes

Tissues and
Organs

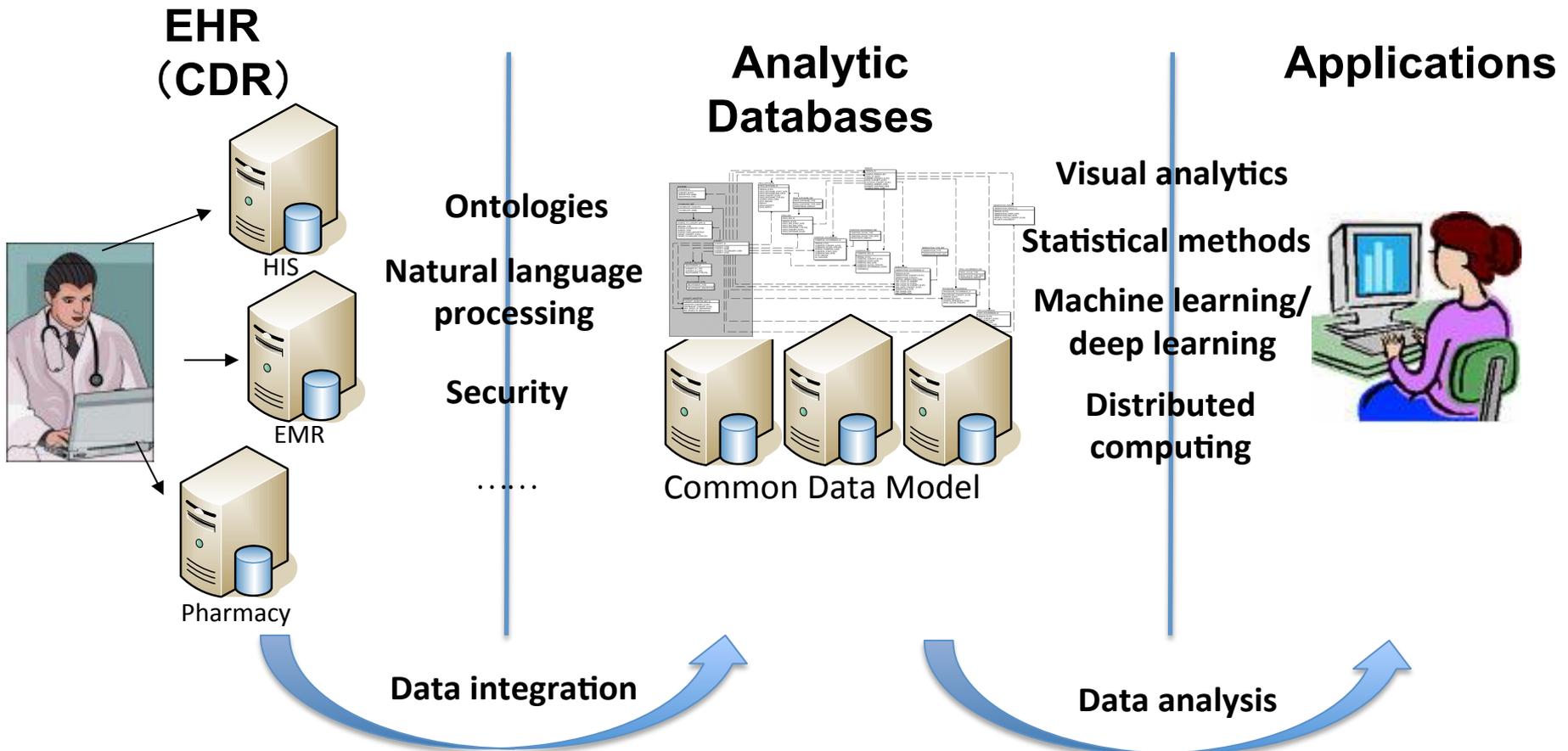
Individuals
(Patients)

Populations
And Society



基于生物医学信息学的临床数据分析技术

Biomedical informatics technologies used for EHR analytics



美国临床医疗大数据公司

Cerner

- 美国最大电子病历系统供应商
 - 2017 营收 ~\$5.1 B
 - 员工~24, 000
- 营收以服务为主

2017 Revenue	\$	%
System sales	1,355,172,000	26%
Support and maintenance	1,046,656,000	21%
Services	2,638,981,000	51%
Reimbursed travel	101,463,000	2%
Total	5,142,272,000	100%

- HealthIntent平台（population health management）

- Optum是一家为广泛的医疗保健市场提供服务的公司，其中包括：
 - 需要照顾的人：病人 - 需要正确支持，消息，资源和产品以实现其健康目标的消费者。
 - 提供护理服务的机构：药店，医院，医生，医疗从业者和其他医疗保健机构，希望实现卫生系统的现代化，并支持最佳的病人护理和体验。
 - 为护理付费的机构：雇主，保险，州政府，联邦政府和市政机构能够得到高质量的护理，有效且高效地进行管理和交付。
 - 为护理创新的机构：药厂或是生命科学研究机构 - 致力于开发更有效的护理方法，实现改善护理服务和健康结果的技术和药物。

- Optum利用数据和分析，药房护理服务，人口健康，医疗保健服务和医疗保健服务等独特功能开展三个业务部门的业务：
 - OptumHealth 专注于护理提供，护理管理，健康和消费者参与以及健康金融服务；
 - OptumInsight 专注于数据和分析以及其他医疗保健信息技术服务，并提供运营服务和支持
 - OptumRx 提供药房护理服务。
- 母公司为上市公司United Health Group (全球财富500强位列第6)



ACHIEVING POPULATION HEALTH MANAGEMENT

Optum serves the unique needs of a population by delivering effective and sustained consumer engagement with our comprehensive clinical and financial services, targeted through deep population health analytics. This results in the right care provided at the right time and in the setting where the consumer is most receptive.

70% INCREASE
in care gaps closed for a
Fortune 500 employer

LARGEST HEALTH SAVINGS
account provider in
the United States

10% FEWER ER VISITS
through physician-led outreach to
a diabetic employee population

34 MILLION PEOPLE
access Optum Behavioral
Health services

- 深度人群健康分析
- 面向糖尿病人群的医生导向服务
 - 减少10%急诊的访问量
- Optum 行为健康服务
 - 3400万人受益



LEVERAGING OUR FOUNDATION IN DATA AND ANALYTICS

With our industry-leading health care data — 1.8 billion lives of claims and 85 million lives of clinical data — and technology and analytics expertise, Optum is delivering predictive analytics that help customers take action.

2016 COMPANY OF THE YEAR
for Population Health Management
by Frost & Sullivan

138,000 PROVIDERS
nationwide rely on Optum One
predictive analytics

**20 OF THE TOP 25
US HEALTH PLANS**
leverage Optum payer analytics

30 RESEARCH PARTNERS
collaboratively researching and
innovating with OptumLabs™

- 业界领先的健康数据储备：
 - 1.8亿保险索赔数据（claims）
 - 8500万临床数据（clinical data）
- 帮助客户进行预测分析（predictive analytics）
 - 13.8万供应商依靠Optum One的预测分析
 - 美国前25的健康方案（Health plans）中20个采用Optum的分析结果
 - 30家科研合作机构与OptumLabs合作



DELIVERING PHARMACY CARE SERVICES

OptumRx® goes beyond traditional pharmacy benefit management, leveraging analytics and the most frequent health care touch point to impact overall health – especially where it matters most – within complex conditions.

>65 MILLION CONSUMERS
served by OptumRx

>1 BILLION SCRIPTS
processed annually

60% ENGAGEMENT
in care management when
connected through OptumRx

\$11-16 PMPM SAVINGS
on medical costs, incremental
to pharmacy savings

- OptumRx通过其超过67,000家零售药店，多个送货上门和专业药房以及提供家庭输液服务的网络，向美国超过6500万人提供全方位的药房护理服务。
- OptumRx综合全面的药房护理服务方法集成了人口统计学，医学，实验，制药和其他临床数据，并将分析应用于推动临床护理洞察，以支持护理治疗和合规性，通过增强服务，提高临床质量使客户和个人受益。
- 每个成员每月 (Per Member Per Month); 节省11-16美金

IBM Watson Health

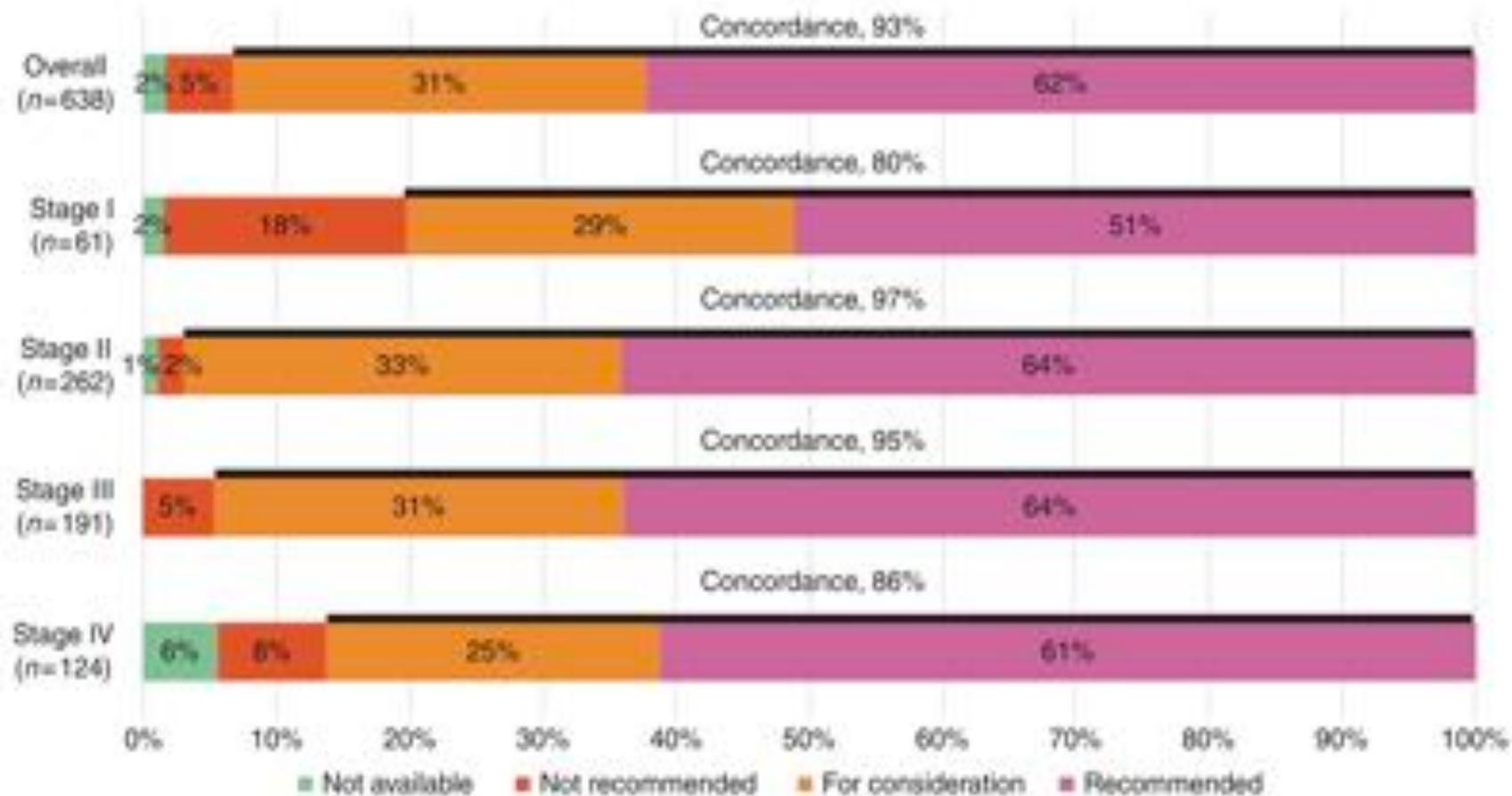


Figure 1. Treatment concordance between WFO and the MMDT overall and by stage. MMDT, Manipal multidisciplinary tumor board; WFO, Watson for Oncology.

Flatiron Health

- 肿瘤大数据和技术平台
 - For community oncology (OncoCloud)
 - For life sciences
 - For academics and hospitals
- 融资、经营情况
 - 2013 , \$8M, Google Ventures
 - 2014, \$130M, Google Ventures
 - 2016, \$175M, Roche
 - 2018, \$1.9B, Acquired by Roche
- 成功原因
 - 高质量数据 (FDA)
 - 肿瘤医生网络

More Health

- 远程医疗公司 – Telemedicine
 - 中美顶级专家联合远程会诊
 - 独立第二诊疗意见
 - 合法提供最先进特效药
 - 赴美医疗绿色通道
- 融资、经营情况
 - 2012, 自筹
 - 2015, \$3M, New Enterprise Associates
 - 2018, 收支平衡
- 成功原因
 - 技术
 - 医疗资源
 - 政策法规

Pieces Technologies Inc.

- 位于达拉斯的医疗人工智能公司
 - Hospital systems and health plans
 - Social service providers
 - 50人左右
- 融资、经营情况
 - 2016, \$21.6M, Jump Capital等6家
 - 2018, 寻求B轮
- 特色
 - 本地医疗资源
 - 针对性产品

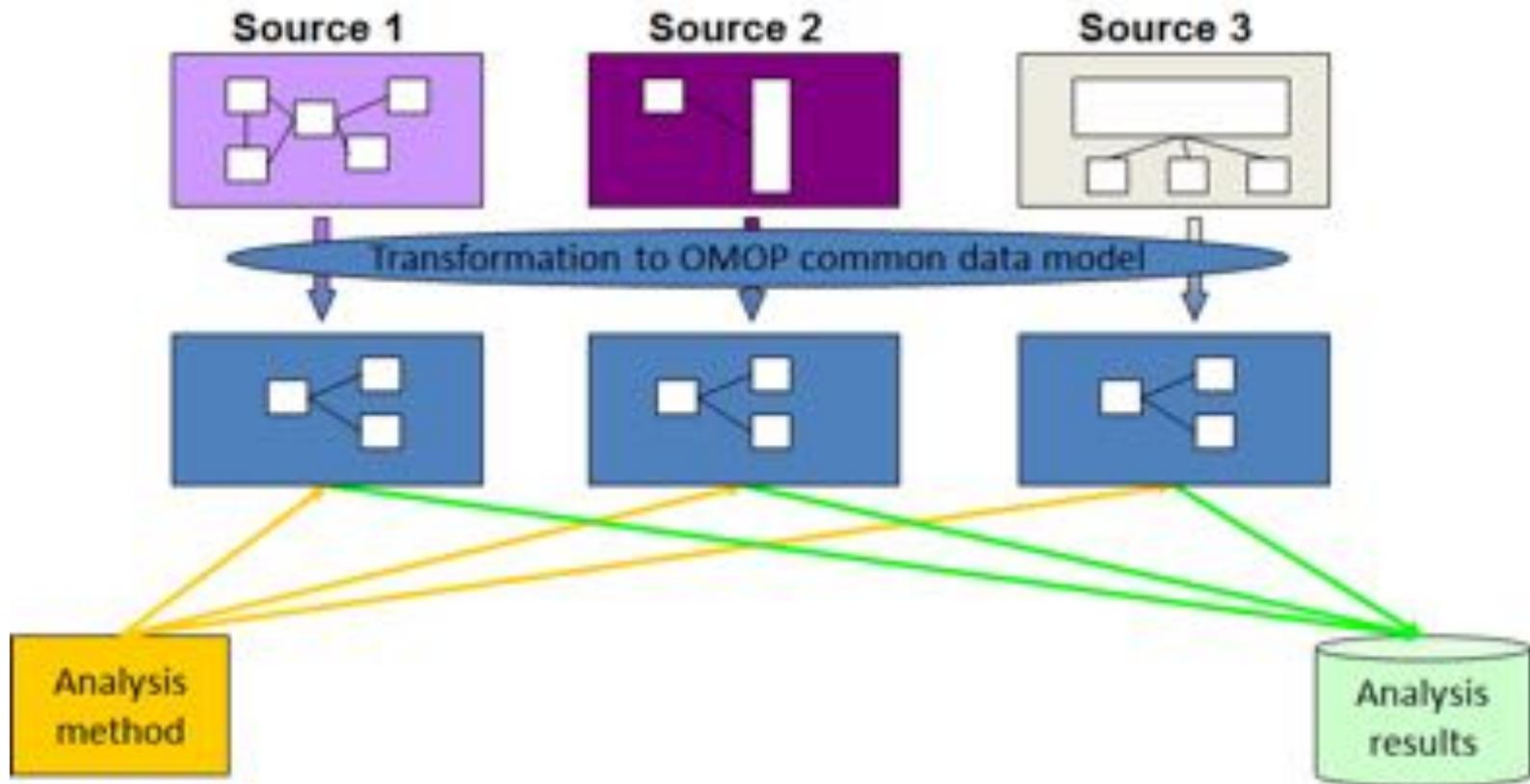
促进中国临床医疗大数据分析 OHDSI中国介绍

开放性临床数据联盟 — OHDSI

- **The Observational Health Data Sciences and Informatics (OHDSI)** program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics
- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

<http://ohdsi.org>

分享数据统计结果（不是原始数据）

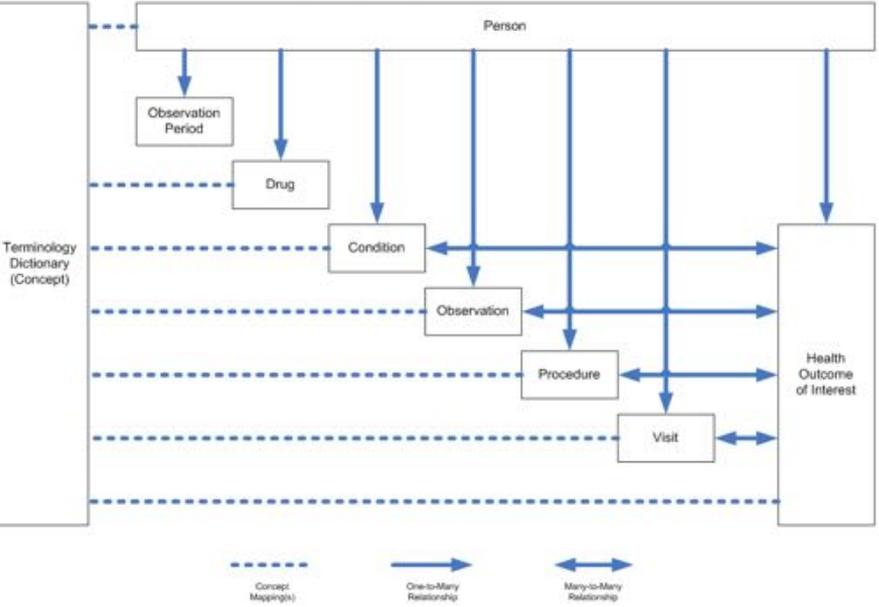


OMOP数据模型经历了多次改编，现有版本5.0

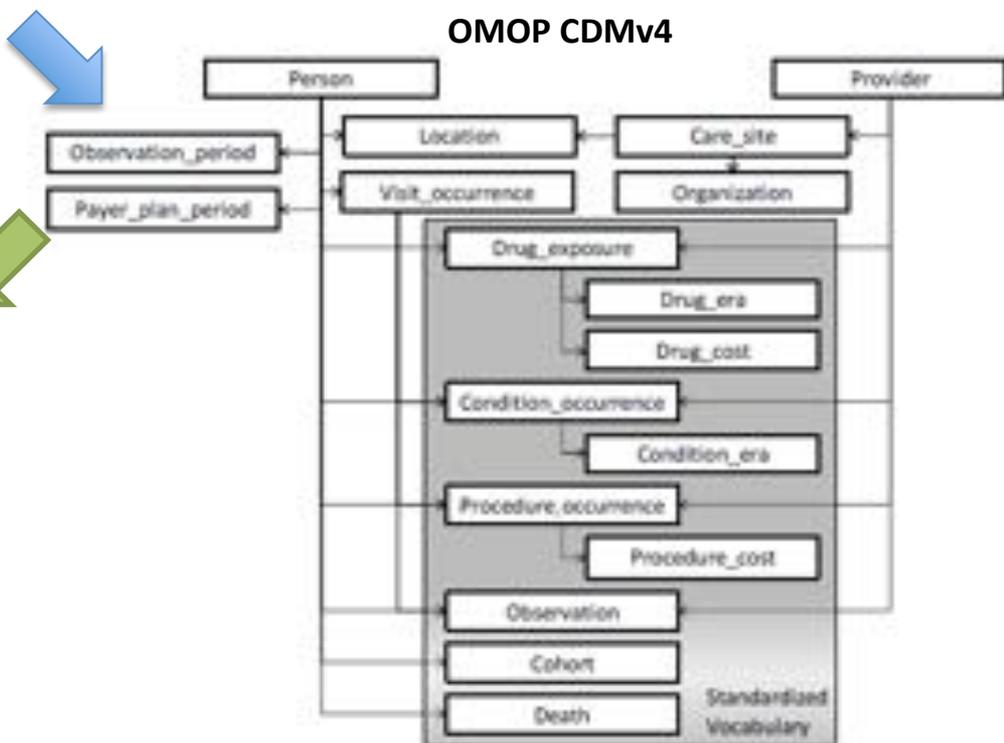
The journey of the OMOP Common data model

OMOP CDM now Version 5, following multiple iterations of implementation, testing, modifications, and expansion based on the experiences of the community who bring on a growing landscape of research use cases.

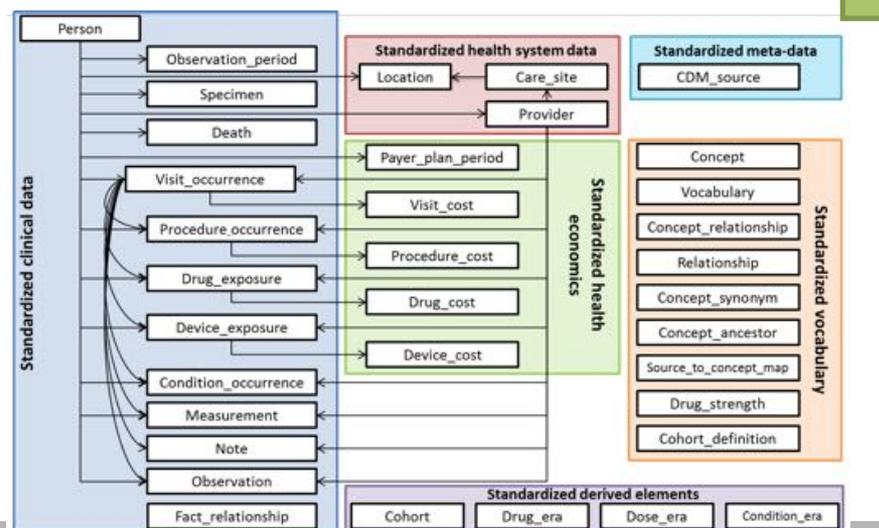
OMOP CDMv2



OMOP CDMv4



OMOP CDMv5



OHDSI网络

OHDSI community in action



OHDSI Collaborators:

- >140 researchers in academia, industry, government, health systems
- >20 countries
- Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer science, machine learning, clinical sciences

Standardized process for network analyses!

Ask clinical question

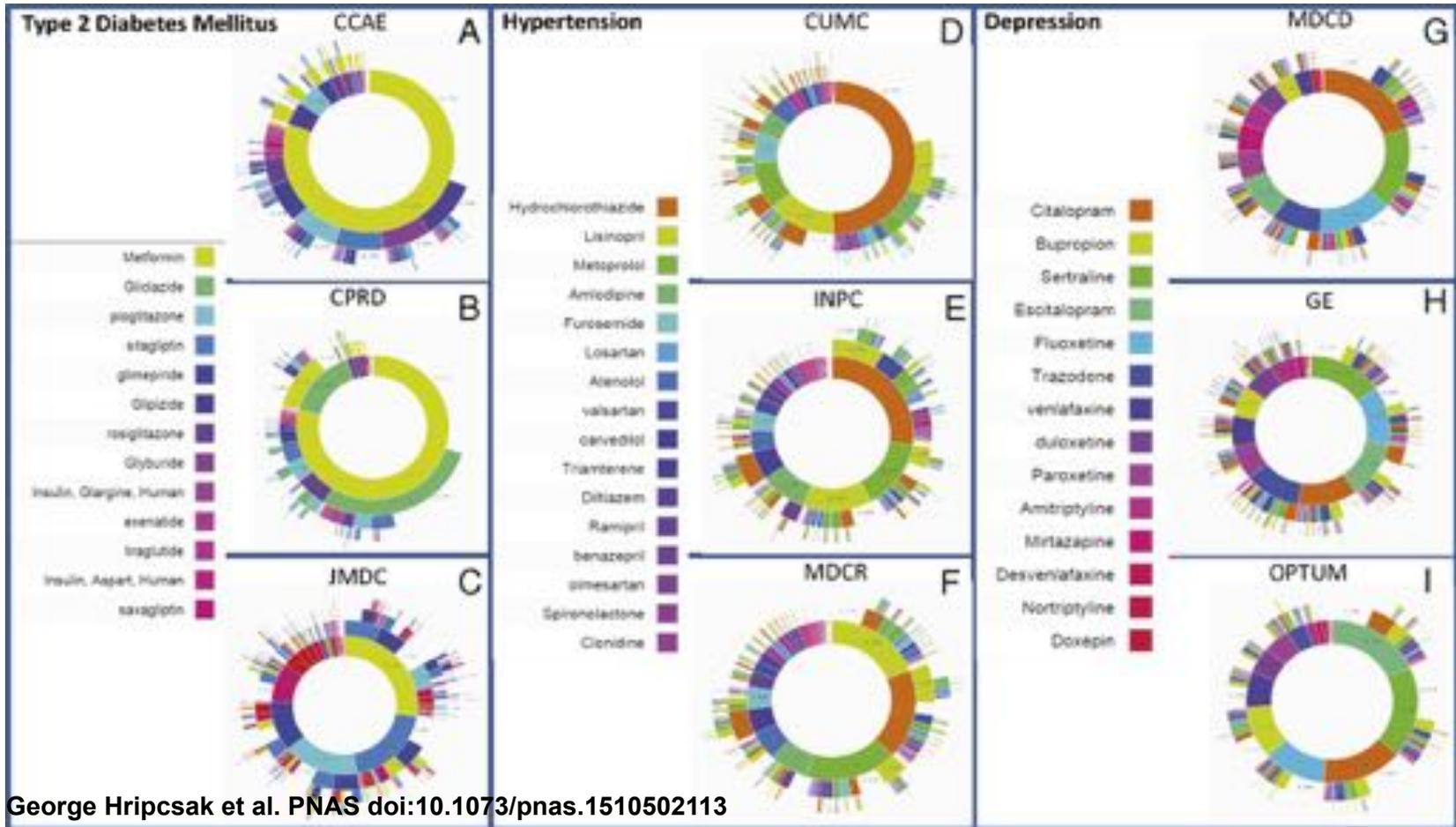
Design protocol

Develop standardized analytics

Generate and disseminate evidence

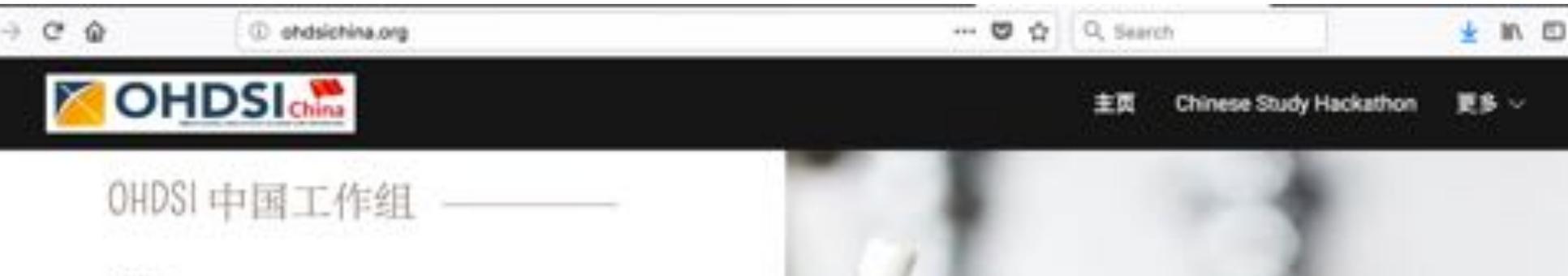
OHDSI研究 — 慢性病药物使用路径

Characterizing treatment pathways at scale using the OHDSI network



George Hripcsak et al. PNAS doi:10.1073/pnas.1510502113

OHDSI China



**2018 中美医学信息大会（暨OHDSI中国年会）
于6月29日-7月1日在广州召开！**

“和世界一流专家学者交流医疗大数据技术和应用

- 促进生物医学信息学领域的国际合作与教育培训工作。
- OHDSI 中国工作组主页：
<http://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:china:wg>

谢谢!
Thank you!

For questions, please contact:

Hua.Xu@uth.tmc.edu