

真实世界数据的发展，价值，应用

宣建伟, PhD.

中山大学药学院医药经济研究所所长，教授
美国佛罗里达大学临床教授，
复旦大学公共卫生学院顾问教授
国家药物政策与医药产业经济研究中心研究员

Shanghai. Oct. 21st, 2017



目录

- 第一部分：医疗大数据/真实世界数据定义
 - 第二部分：医疗大数据/真实世界数据库举例
 - 第三部分：医疗大数据/真实世界数据分析应用举例
-

什么是真实世界数据？

Big Data is more than “a lot of data”

Volume:

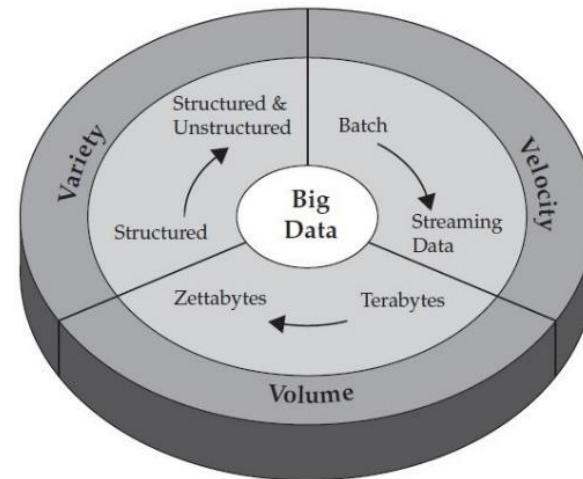
Exponential growth in data capture and compilation results in data size commonly exceeding terabytes

Velocity

Data capture frequency is changing from sparse batch reporting towards real time capture.

Variety

Structured data become “minority”. Huge unstructured information, such as medical images, videos, gene expression, social media, is also part of the data collection.



美国食品药品监督管理局（FDA）、国际药物经济学与结果研究协会

（ISPOR），将真实世界数据定义为除开传统临床试验以外的数据^{1, 2}

1.ISPOR Connection. Real-life data: A growing need. Available at: <http://www.ispor.org/News/articles/Oct07/RLD.asp>.

2.U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Draft Guidance for Industry and Food and Drug Administration Staff- Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices.

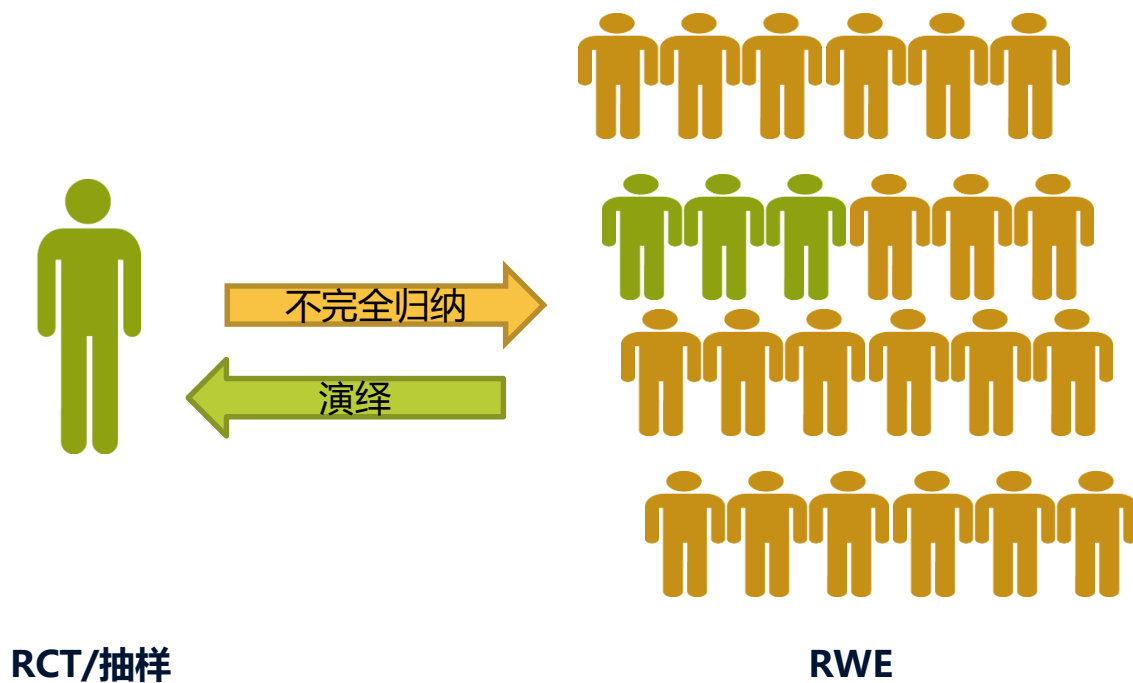
卫生决策中的重要数据源



真实世界大数据必须是来源于各个部门的真实数据比如医疗机构的临床数据

“小中见大” 和 “大中见小”

古之达人，推而通之，大而天地山河，细而秋毫微尘，此心无所不在，无所不见。是以小中见大，大中见小，一为千万，千万为一，皆心法尔。——宋·苏辙《洞山文长老语录》



常见的真实世界数据

观察性研究数据

- 患者调查
- 注册登记研究
(registry study)

临床实践数据

- 医院电子病历 (electronic medical record , EMR)
- 医保理赔数据库 (claims data)
- 公共卫生调查 (如糖尿病患病率调查)
- 公共健康监测 (如药品安全监测)
- 出生/死亡登记库

干预性研究数据

- 实效性临床试验
(pragmatic randomized controlled study , pRCT)

真实世界研究的优势

较好代表性，
反映真实世界

大样本量、
长跟踪时间

信息时代，
医疗大数据

提供传统RCT无法提供的证据（如真实诊疗中干预的有效性，长期用药的安全性，依从性等）

- ❑ 真实世界研究也存在一定局限，如混杂因素，数据的准确性、完整性等
 - ❑ 研究设计都有优势和不足，需要根据研究问题，选择合适的研究方法
-

与 RCTs 的区别与联系

RCTs

&

真实世界研究

各据优势和劣势，两类研究回答不同问题，不能相互取代，也不存在相互对立

- RCTs 在药物上市前提供有关安全性和有效性的信息，使审批部门和使用者的疗效和安全具备信心
- 真实世界研究在药物上市后开展，通过收集药物在临床使用中的实际效果数据，帮助医疗决策者了解新的药物在其所处医疗环境中所能发挥的真实效果（Risk Sharing etc.）

RCTs 是真实世界研究的基础前提，真实世界研究是 RCTs 的有益补充，只有把握好两者的关系，才能合理设计和实施相应的研究，获取更加有价值的证据

目录

- 第一部分：医疗大数据/真实世界数据定义
- 第二部分：医疗大数据/真实世界数据库举例
- 第三部分：医疗大数据/真实世界数据分析应用举例

Real World Data in China

Database Name	Database Type	Name of the Institute (Company)	Relationship Type
SuValue(溯直)EMR database	EMR database	SuValue Healthcare Scitific Co., Ltd.	Row data owner partner ship
The Shengkang hospital group database of Shanghai	EMR database	Wonder Group	Row data owner partner ship
EMR database of the First affiliated hospital of Sun Yat-sen University	EMR database	The First Affiliated Hospital of Sun-Yat-Sen University	Database owner
The Guangzhou health insurance database	Health insurance database Government claim	Health Economic Research Institute,school of pharmacy, Sun Yat-sen University	Row data owner partner ship
Hospital Purchase of Drug information Database of the Yangtze river	Drug Purchasing information database	Institute of science and technology information of SFDA	Database owner

Brief Introduction of SuValue database

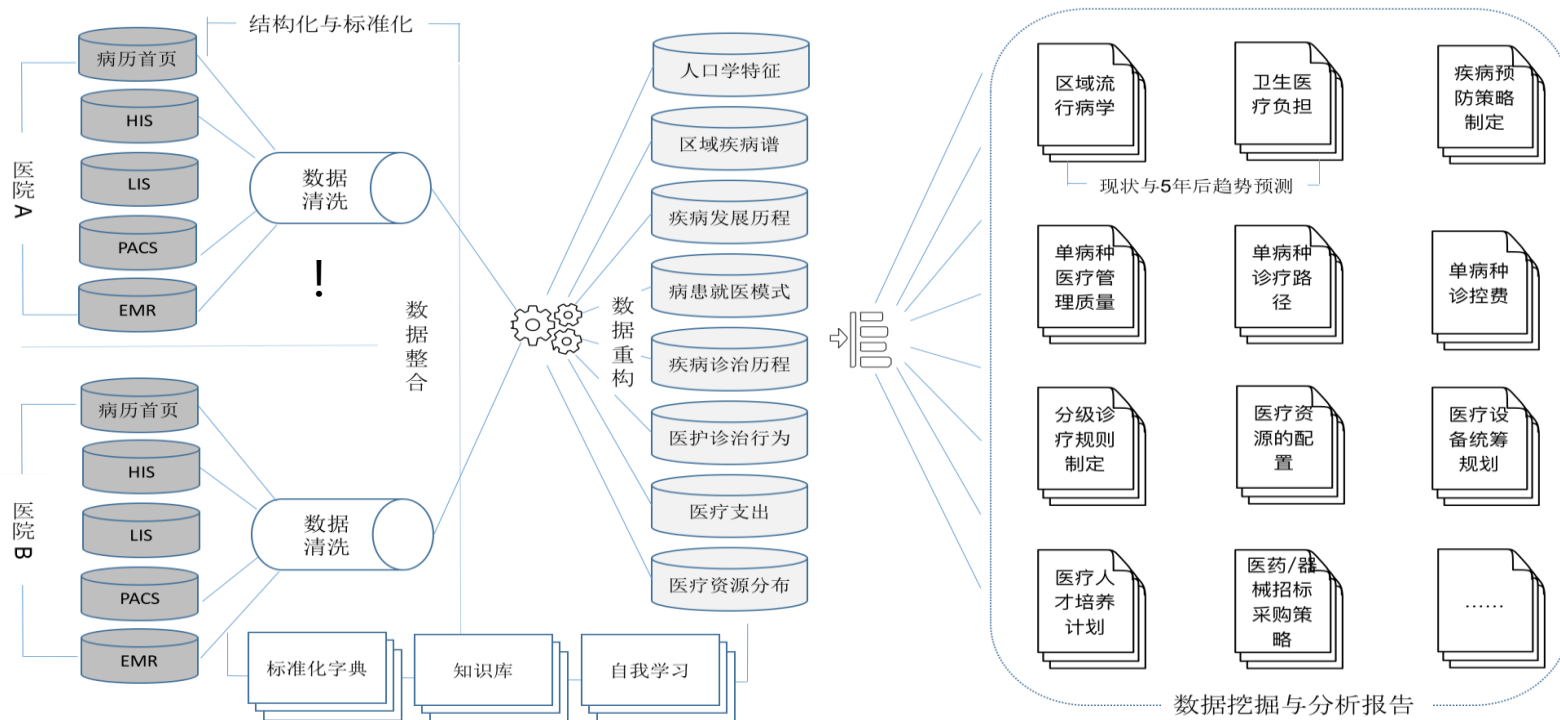


SuValue is a medical RWD database provider which obtains the *complete HIS/EMR data* from independent hospitals in various provinces/cities of China.

Currently contains more than **70+** grade II and above hospital data and more than 20 million patient information.

Geographic distribution: Beijing, Guangdong, Chongqing, Gansu, Henan, Hunan, Jiangx, Guizhou, Shanxi, etc.

数据重构与数据挖掘，为医院提供循证服务



Patient
Patient ID
Birth
Gender
Race
Blood Type
Insurance
Address

Register
Patient ID
Hospital ID
Regist time
Out patient
Hospital NO
Hospitalisation time
Discharge time
Age
Gender
Profession
Allerge
History
Race
Address

Diagnostic
Patient ID
Diagnostic time
Diagnostic type
Out patient diagnostic
O-P ICD 10
In Patient diagnostic
I-P ICD-10
Primary diagnostic
P- ICD10
Other diagnostic
Other-ICD10
Discharge diagnostic
D-ICD10

CPOE
Patient ID
Rx date
Rx type
Lab test type
Lab test time
Image type
Image tiem
Exe time
Exe type
CPOE type
Item type
CPOE name
Skin test
Drug type
CPOE duration
Dosage
Dosage unit
Number
Total number
Start time
End time
Drug duration
In take type
Drug usage

Medicine
Patient ID
CPOE ID
TCM
West medicine
ATC
Formulation
Dosage
In take type
Exe time
Drug name
Pack
Address
Dosage
Dosage unit
Usage
Drug duration
Amount
Unit
Skin test
Remark
Insurance
Code

Cost
Patient ID
CPOE ID
Drug cost
Consumable
Nursing cost
Hospital cost
Insurance type
Reimbursement

LIS
Patient ID
CPOE ID
Lab test name
Lab tes type
Test time
Test item
Test results
Ref
Unit

PACS
Patient ID
CPOE ID
Image name
Image type
Time
Details
Results

Procedure
Patient ID
Type
Cut type
Level
Anaesthesia type
Recovery

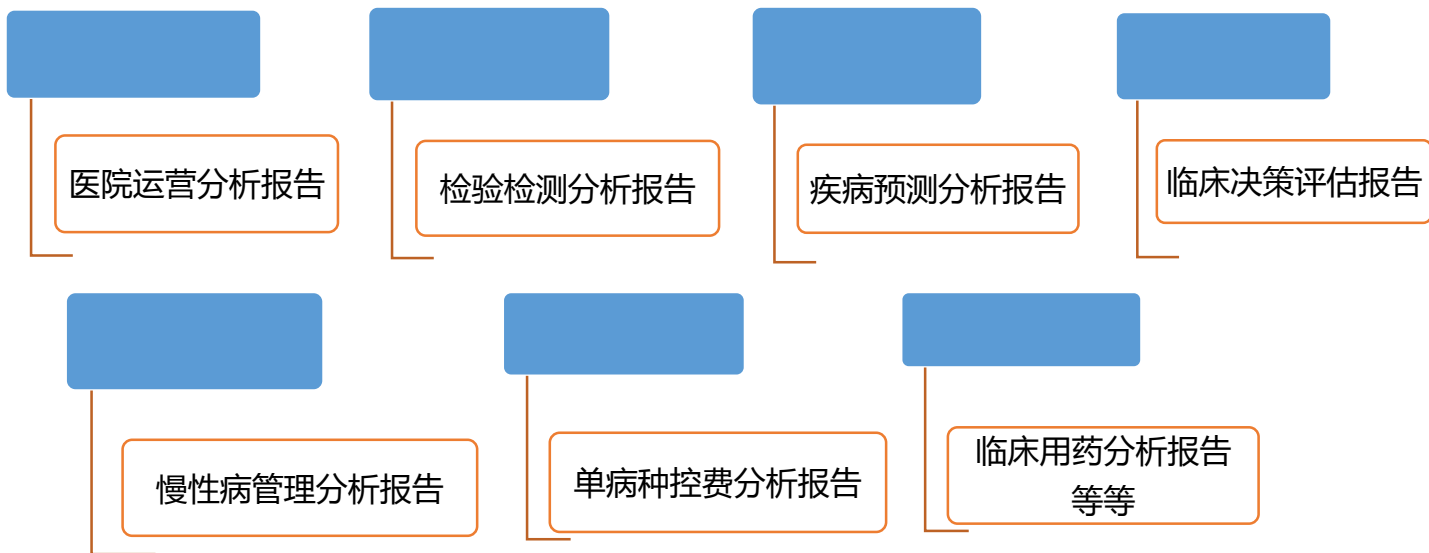
Data
Ready to use
Contain missing value
Standarization needed

LIS	Laboratory Information System
CPOE	Computerized physician order entry
ICD-10	Internation Code of Disease version 10
TCM	Traditional Chinese Medicine
ATC	Anatomical Therapeutic Chemical drug classification system
PACS	Picture Archiving and Communication Systems

目录

- 第一部分：医疗大数据/真实世界数据定义
 - 第二部分：医疗大数据/真实世界数据库举例
 - 第三部分：医疗大数据/真实世界数据分析应用举例
-

医疗大数据在医院管理及临床诊疗中的应用



◆ 大数据应用案例分享1

慢性病管理分析报告

Real World Diabetes Database

Age group	Total	<1 Year	>1 Year	>2 Year	>3 Year	>4 Year	>5 Year	>6 Year	>7 Year	>8 Year	>9 Year	>10 Year
0-10	3,696	3,601	95	47	21	8	5	2	-	-	-	-
10-20	2,638	2,431	207	94	41	16	7	4	2	-	-	-
20-30	22,434	21,119	1,315	539	242	109	49	28	14	7	2	1
30-40	38,472	34,264	4,208	1,988	1,066	496	261	125	57	9	2	2
40-50	79,622	65,489	14,133	7,652	4,492	2,382	1,392	716	337	77	6	4
50-60	120,346	98,016	22,330	12,835	8,058	4,653	2,967	1,699	925	226	29	15
60-70	142,431	114,852	27,579	16,928	11,253	6,859	4,665	2,802	1,625	434	67	37
70-80	99,770	79,779	19,991	12,460	8,416	5,358	3,695	2,310	1,413	365	65	34
80-90	34,850	27,704	7,146	4,260	2,861	1,734	1,190	766	421	120	24	10
Y>90	7,056	6,378	678	400	260	153	85	47	24	9	2	2
Subtotal	551,315	453,633	97,682	57,203	36,710	21,768	14,316	8,499	4,818	1,247	197	105

糖尿病人群数据概览

整体分析

	2013年	2014年	2015年	2016年
总人数	6,355,686	7,083,427	7,548,542	7,196,329
糖尿病人数	69,342	85,397	98,981	98,234
合并糖尿病人数	78,495	88,082	97,572	95,787
糖尿病总人数	147,837	173,479	196,553	194,021

解读：平均每家医院每年2,000-3,000例糖尿病患者，合并糖尿病的人数与糖尿病主诊人数的比例住院明显高于门诊，且呈上升趋势

合并糖尿病与主诊糖尿病的比值

	2013年	2014年	2015年	2016年
门诊	1.1	0.9	0.9	0.8
住院	1.6	1.8	1.9	2.4

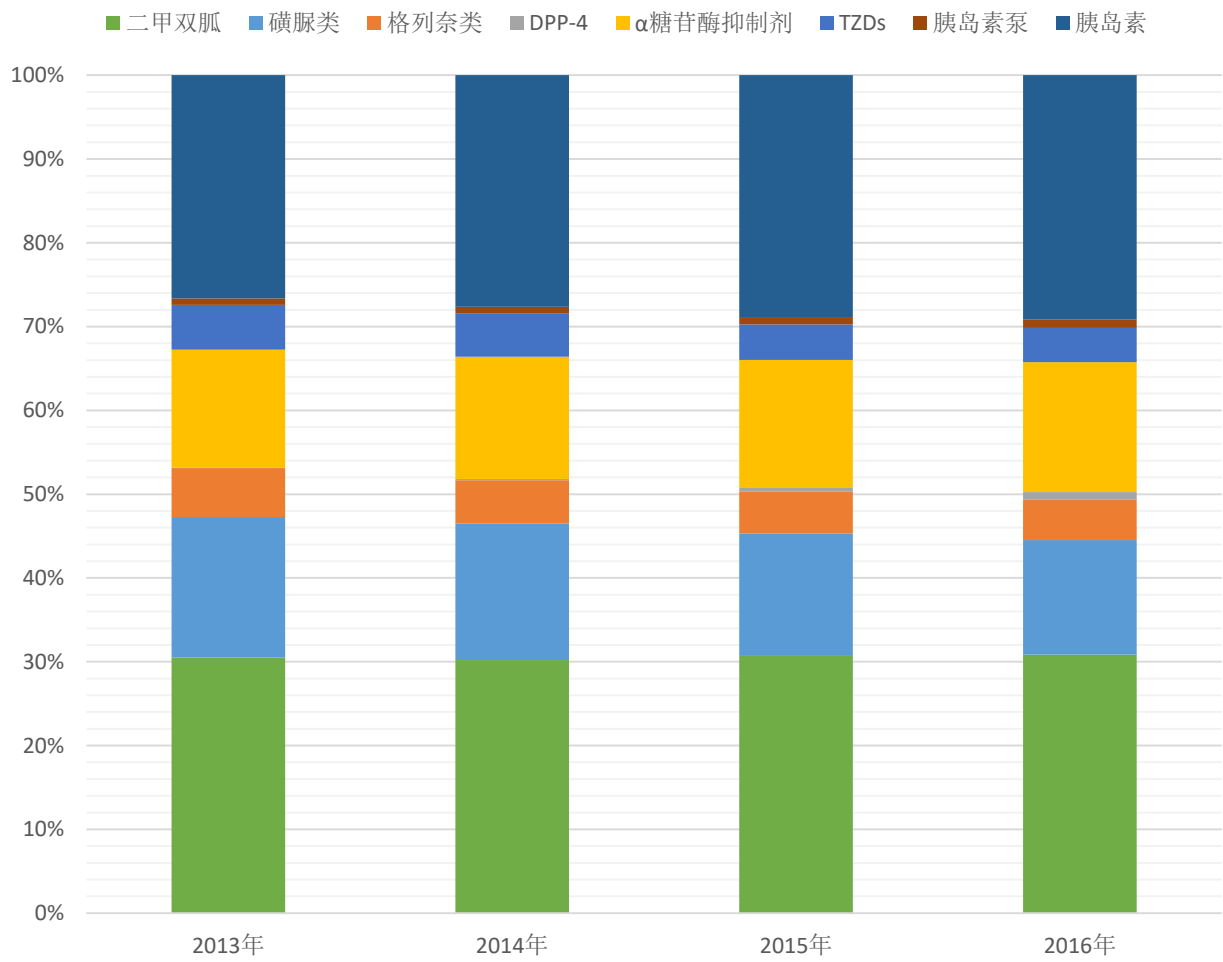
门诊分析

	2013年	2014年	2015年	2016年
总人数	5,751,069	6,394,646	6,817,187	6,529,854
糖尿病人数	60,621	75,292	87,188	87,810
合并糖尿病人数	65,072	71,209	77,047	73,427
糖尿病总人数	125,693	146,501	164,235	161,237

住院分析

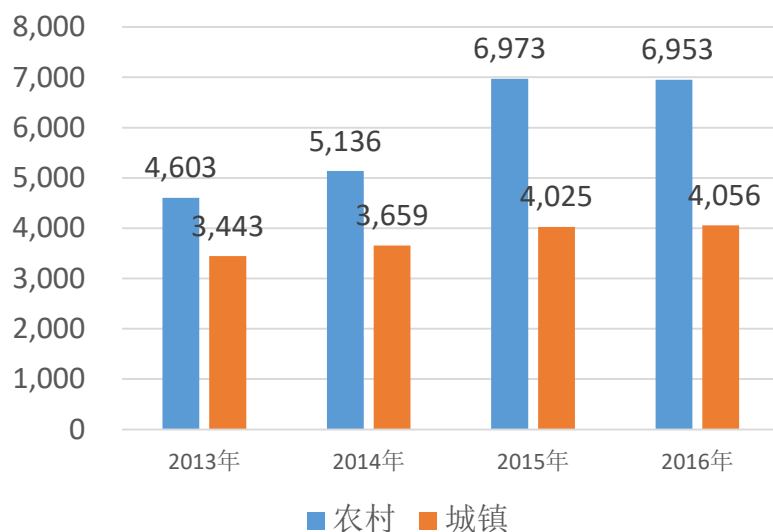
	2013年	2014年	2015年	2016年
总人数	633,329	736,304	809,295	770,582
糖尿病人数	9,093	10,616	12,484	11,416
合并糖尿病人数	14,406	19,007	24,008	26,872
糖尿病总人数	23,499	29,623	36,492	38,288

治疗方案构成



糖尿病患者的疾病负担

糖尿病病人的人均年医疗费用



医保类型	类别	2013年	2014年	2015年	2016年
农村	糖尿病患者	7,718	8,607	9,239	10,397
	医保糖尿病人人均费用	4,603	5,136	6,973	6,953
	门诊人次	31,031	34,653	37,323	32,277
	住院人次	2,027	2,241	2,850	3,242
	住院/门诊	0.07	0.06	0.08	0.10
	药占比	56%	54%	49%	45%
城镇	糖尿病患者	55,272	65,197	77,477	82,585
	医保糖尿病人人均费用	3,443	3,659	4,025	4,056
	门诊人次	560,609	598,202	631,851	646,804
	住院人次	4,654	6,295	7,747	8,198
	住院/门诊	0.01	0.01	0.01	0.01
	药占比	69%	65%	64%	63%

农村糖尿病患者的人均年医疗费用高可能与该部分人群住院率高有关，农村糖尿病患者的住院率在6-10%之间，而城镇糖尿病患者的住院率却只有1%。

远期并发症不良事件发生率

糖尿病患者远期不良事件的发生率	有后续记录人数	就诊记录达1年人数（任何科室）	1年不良事件发生率	就诊记录达3年人数（任何科室）	3年不良事件发生率
总人数	206,486	192,284		103,886	
首诊后有绞痛记录人数	991	448	0.23%	812	0.40%
首诊后有心肌梗死记录人数	368	180	0.09%	307	0.15%
首诊后有脑卒中记录人数	558	347	0.18%	473	0.23%
首诊后有眼底视网膜病变记录人数	3	2	0.00%	3	0.00%
首诊后有糖尿病肾病记录人数	6,134	3,937	2.05%	5,484	2.69%
首诊后有糖尿病足/坏疽记录人数	1,104	645	0.34%	921	0.45%
首诊后死亡记录人数	无	无	无	无	无

孕产妇的糖尿病发病率

孕妇人群分析				
有明确检查信息的孕妇人数	孕前患有糖尿病人数	孕前糖尿病发病率	妊娠期糖尿病人数	妊娠期糖尿病发病率 *
438,588	173	0.04%	68,590	16%

*注：由于 HIS 系统中明确注明有妊娠期糖尿病诊断者非常少，因此，我们采用了追溯空腹血糖检查的方法，妊娠期糖尿病诊断标准空腹血糖 $\geq 7.0\text{mmol/L}$

◆ 大数据应用案例分享2

单病种控费分析报告（寻找差据，对照同类医院，以更有效地管理疾病，提高效益）

    -  + 100%

◆ 大数据应用案例分享3

药品分析报告

结合药物经济学评价结果对药品实施差异化管理，帮助科学合理费用管控

结合药物经济学评估结果进行科学控费

以排名靠前的药品为切入点：

企业可主动提交药物经济学结果，为医院提供参考帮助

排名靠前的药品

销量排名	产品
1	氯化钠
2	葡萄糖
3	人血清白蛋白
4	血栓通
5	波立维
6	奥德金
7	倍通
8	立普妥
9	丹参多酚酸
10	舒血宁
11	兰索拉唑
12	醒脑静
13	疏血通
14	泮托拉唑钠
15	康艾

药物经济学
评价

A 有药物经济学评价且具有成本效果优势*：

- 以满足临床需求为前提，设计相对宽松的控制使用措施

B 不具有成本效果优势或无药物经济学评价：

- 加强监督管理，制定严格的控制使用措施



注释：判断成本效果优势一般采用增量成本效果比（ICER）；若 $ICER < 1$ 人均GDP，则具有很高的成本效果优势，若 $1 < ICER < 3$ 人均GDP，则具有成本效果优势，若 $ICER > 3$ 人均GDP则不具有成本效果优势

◆ 大数据应用案例分享4

药品安全性分析

病人报告，医生观察某个X药品可能会导致使用病人水储留（如何进行确认或否定）

大数据人群	病人人数	产生水储留病人数	发生比例	结论
所有该疾病病人	N	n	n/N	运用 Disproportionate Analysis 比较发生 比例，找出阈值， 得出结论
所有该疾病病人 使用了X 药品	N1	n1	$N1/N1$	
所有该疾病病人 使用了X 药品以 外的任何药品	N2	n2	$N2/N2$	

◆ 大数据应用案例分享5

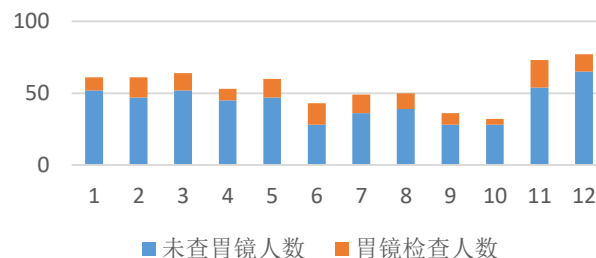
医院运营分析管理

细分产品示例四：医院运营分析报告摘要展示

22654	内二科	胆汁返流性胃炎	298
23039	内二科	腹痛查因：急性胃炎？	53
23585	内二科	急性肠胃炎	5
23672	内二科	急性胃炎	814
23673	内二科	急性胃炎？	20
24010	内二科	慢性胃炎	1157
24192	内二科	呕吐查因：急性胃炎？	22
24908	内二科	胃炎	1502
26104	内一科	腹痛查因：急性胃炎？	2
26474	内一科	胆汁返流性胃炎	118
26857	内一科	腹痛查因：急性胃炎？	4
27301	内一科	急性肠胃炎	7
27389	内一科	急性胃炎	503
27390	内一科	急性胃炎？	13
27741	内一科	慢性胃炎	1336
28981	内一科	胃炎	1831



胃十二指肠炎症/溃疡患者男性且大于40岁者胃镜检查比例



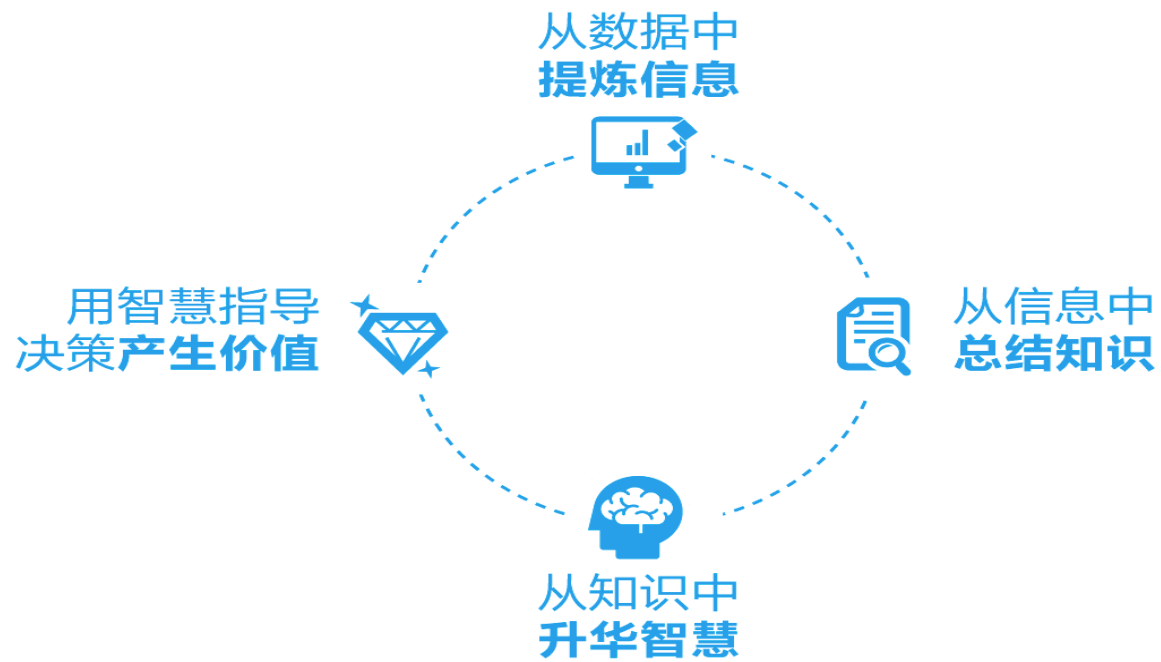
要点解析：

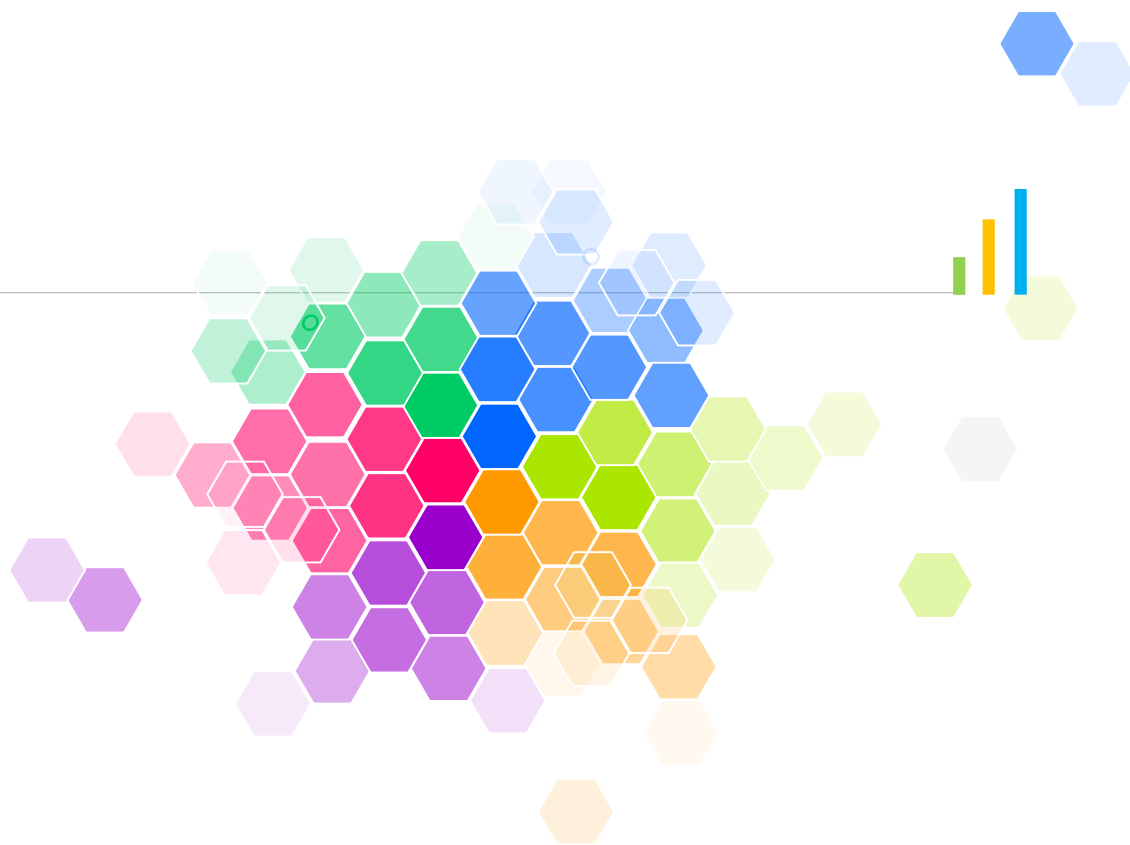
- 1、二级医院在胃癌高发人群中胃镜检查的比例不足30%，这就会造成大量胃癌患者被漏诊；
- 2、在没有明确病情的情况下给予治疗，就会造成盲目开药，乱开药，使占比不合理的增高；
- 3、在分级诊疗的实施后，会有更多的类似患者下沉到二级医院，今后的潜在风险和医药管理压力会更大。

挑战与机遇：

- 1、对有适应症的患者开展必须检查，既可以降低漏诊误诊的风险，还可以控制盲目开药，避免不必要的治疗，使医疗收入结构更合理；
- 2、对于癌症高危人群，早期筛查，预防为主，也是符合现代化疾病管理理念和国家分级诊疗的策略。
- 3、增加合理收入，假设每年增加30%的检查率，合适人数1万人，可以使合理营收增加近100万元。

理念





谢谢！

2017.09