OHDSI in Korea

Rae Woong Park, MD, PhD

Department of Biomedical Informatics,

Ajou University, School of Medicine,

Korea

Contents

- Korean OHDSI Network
- Why DRN and CDM is popular to Korea?
- Lesson learned from potential Data Owner
- Future plans in Korean OHDSI

Korean OHDSI network: 18

CDM Conversion Status

- Conversion completed: 4 institutions
  - Ajou University Hospital: 2.3M, 23 years EHR
  - Gachun Gil University Hospital: 2M, 10 years EHR
  - Kangwon National University Hospital: 0.5M, 10 years EHR
  - NHIS: 2M, 12 years, Claim + regular health exam (2018: 51M, 12 years)
- Conversion in progress 14 institutions
  - **By the end of 2017: 2**
  - Samsung Medical Center
  - Wonkwang University Hospital
  - Wonju Severance Hospital
  - **By the end of 2018: 14**
  - Chonbuk National University Hospital
  - Yonsei Unviersity Severance Hospital
  - Korea University Anam Hosital
  - Korea University Guro Hospital
  - Korea University Ansan Hospital
  - Gangdong Sacred Heart Hospital
  - Hanyang University Hospital
  - Ehwa Unviersity Mokdong Hospital
  - Ehwa Unviersity Magok Hospital
  - Cha University Hospital
  - HIRA/NHIS (51M, 12 years)

<MOU>

Activity in Korea

- Leadership meeting
  - Bimonthly
  - Leaders in charge of CDM for each hospital
  - Important decisions and policy-related issues

- Engineer meeting
  - Biweekly, TC (current 9th)
  - EHR experts from participating hospitals
  - Discuss all the technical issues during CDM conversion

Activity in Korea

- Open forum
  - Monthly, 3 hour lecture
  - Agenda
    - Introduction to OHDSI and CDM
    - OMOP CDM Structure
    - OMOP CDM Vocabulary/ vocabulary mapping
    - Tools for OMOP CDM
    - ETL process
    - Research Experience using OHDSI Network

CDM conversion Standard Operation Procedure (SOP)

	Step	Time-line	Human Resources From Data Partner	Role of Data Partner	Role of Supporting Organization	Contact
1	MOU		CDM director and engineer	Administrative affair	Administrative affair	shinda@ajou.ac.kr
2	Advance Preparation	-	CDM director and engineer	Budget for H/W, S/W, human resources	Orientation, lecture and consultation	
3	Vocabulary Extraction	1 wk	EMR expert	Frequency table of every drug, supply, procedure, diagnosis, lab test	Support	hidoyebi@ajou.ac.kr
4	Vocabulary Mapping	- Vocabulary customization: 2 mon - Code mapping: 3 mon	Medical expert x 2 (physician x1, Nurse x 1)	Support	Code mapping between local code and OMOP vocabulary (voca covering 99% of data)	
5	ETL Definition	2 mon	EMR expert x1 Medical expert x 1	ETL document	Support	enzo@ajou.ac.kr bacojun@ajou.ac.kr
6	ETL	2 - 4 weeks	DB expert x 1	ETL Query	Support	
7	OHDSI Tool application	1 - 4 weeks	R/JAVA expert x 1	Achilles, Atlas, etc installation	Support	jsh90612@gmail.com
8	DQM	4 -8 weeks	DB expert Medical expert	Run DQM codes	Provide DQM codes and evaluation and feedback	enzo@ajou.ac.kr bacojun@ajou.ac.kr

1) SOP for MOU

1. 협의사항

	단계	방법
MOU	1단계	신다혜 담당자를 통해 MOU 협약 내용 알림 * 전화번호 : 031-219-4473 * e-mail : shinda@ajou.ac.kr
	2단계	아주대병원 연구지원팀 이해경 선생님과 연결하여 협약 진행 * 전화번호 : 031-219-4414 * e-mail :sea8001@aumc.ac.kr

	단계	방법
용어매핑계약서	1단계	신다혜담당자를 통해 용어매핑사용계약서 가안파일을 전달 받기
	2단계	참여기관에서 계약서 내용 파악 및 수정후, 신다혜담당자에게 전달
	3단계	계약서 수정이 완료되면 아주대병원 연구지원팀 이해경 선생님께 전달하여 계약 진행

	협의사항	비고
1	착수 시간	변환 작업 착수일정 계획
2	작업 기간	실제 예상 작업 기간 협의
3	변환 소스 DB	EMR에서 직접 데이터를 추출할 것인지, 별도로 구축된 DW에서 데이터를 추출할 것인지 정의
4	소스 데이터 추출 형태	DUMP 파일 형태로 추출 권장
5	CDM 구축 범위	변환 대상 CDM 테이블: 전체 테이블 혹은 CDM 테이블들을 단계별로 구축계획 변환 대상 데이터의 기간 범위(예: 2010년 ~ 2015년)
6	CDM 업데이트 정책	1) 업데이트 주기 2) 버전 관리 방법 - 신규 DB 생성 혹은 기존 테이블에 추가
7	운용 장비/설비	참여기관 환경검토 예) DBMS 종류, 원격지원 가능여부, 보안정책, 라이센스 등
8	참여 인력 및 지원 인력	전담 인력, 서포트 인력의 수와 역할 전담 인력이 갖추어야 할 필수 자격/능력 요건
9	데이터/적용도구 공개 수준	Achilles 등의 툴 적용 후 공개수준(원내/원외/비공개)
10	인력 지원 방식	지원인력: 상주, 방문, 원격
11	인력 교육	시기, 횟수 및 방법(방문, 원격)
12	회의 일정	회의 주기, 회의 방법, 참석자

2) SOP for Advance Preparation

	구분	Spec	규격	비고
1	Model	DELL POWEREDGE R730XD	2U rack server	ETL수행 및 분석용 서버 1대 구성
2	Prcessors	INTEL XEON E5-2667V4 3.2GHZ	CPU * 2EA (8Core, 16Thread)	최소요구사항
3	Memory	768 GB, 24 x 32GB	PC4-17000P DDR4-2133 REGISTERED ECC MEMORY	320 GB 이상 추천
4	Hard Drive	18 X DELL 1.6TB	MLC SAS III SSD 2.5 INCH ENTERPRISE CLASS	4TB 이상 추천 (자체산정, CDM 및 원본 데이터 용량 고려) 빠른 연산 속도를 원하는 경우 intel SSD 1.2TB PCIE 추천
5	Raid Controller	PERC H730 1GB CACHE	12GB/S RAID CONTROLLER	최소요구사항
6	Management	IDRAC8 ENTERPRISE	Web base	웹 기반의 원격 서버 관리툴
7	Networking option	1GBE Network card	*4	고속 네트워킹 선택사항
8	Power Supply	Dual 1100W	-	최소요구사항

2.Server - Web

	구분	Spec	규격	비고
1	Model	HP ProLiant Server DL 380 Gen9	2U rack server	1개의 베어본 서버 Web 서버용
2	Prcessors	Intel Xeon E5-2640v3 (2.6GHz)	CPU *1EA (8Core, 16Thread)	-
3	Memory	16 GB (Max 768GB)	DDR4-2133R REGISTERED ECC MEMORY	ATLAS 운영을 위한 최소 요구 사항 256GB 이상
4	Hard Drive	NONE	Hot Plug 8 Bay	System OS 구성을 위한 SSD 512 GB 저장공간 할당을 위한 SAS HDD 1TB * 7EA
5	Raid Controller	SATA drives and supports Raid 0,1,1	Smart Array P440ar/2GB	HDD Raid용 컨트롤러
7	Networking option	4*1Gb Ethernet	331i Adapter	-
8	Power Supply	1 * 500W Flex Slot (94%+)	2U Rack from factor	-

3. Software

	구분	Spec	규격	비고
1	OS	Window server 2012	Enterprise license	권장사항
2	DB	SQL Server 2012	Standard or more	권장사항
3	Web service tool	Apache Tomcat 7	Free license	필수
4	Eniv.	Java Development Kit (JDK)	Over 7 version	-
5	Base application	R	Over 3.3 version	-
6	Base application	R studio	Lastest version	-
7	Base application	R tools	Lastest version	OS와 호환 여부 확인

4. License (연간계약)

	구분	요구사항		
1	용어	SNOMED CT		
2	용어	medical terms for insurance to OMOP concept ID mapping table)		

*비용은 라이센스 구매관련하여 아래와 같습니다.
- MS Window server standard (교육용) : 약 70 만원
- MS SQL standard 2016(교육용) : 약 130만원
- 용어매핑 저작권 사용 비용 : 1000만원 (연간 계약필요, 참여기관이 증가할수록 비용이 줄어들 예정)
- SNOMED 라이센스 : 약 200만원 (연간 계약필요)

3) SOP for Vocabulary Extraction

(1) 공통사항

- 각 테이블의 필드에서 '~ concept_id'로 표기된 필드에는 OMOP에서 제공하는 concept 테이블 내의 concept_id가 들어가야 함

- concept 테이블 내의 필드정보 중 다음의 조건을 만족하여야 함

필드명	조건
concept_name	유사수준이 가장 높은 concept_id를 찾아야 함
domain_id	각 OMOP CDM 테이블에 맞아야 함 (ex: drug_exposure 테이블의 drug_concept_id에 들어갈 concept_id의 domain_id는 'Drug')
standard_concept	S
invalid_reason	NULL 또는 공백

(2) 매핑과정

순번	항목	설명
1	Vocabulary 다운로드	데이터 신청양식을 채우고, 필요한 용어의 체크박스에 체크한 뒤 신청하면, 수 분 뒤 이메일을 통해 받은 링크에서 다운로드가 가능함 (http://athena.ohdsi.org/)
2	분리된 용어 통합	CPT4와 같이 EULA가 요구되는 용어를 체크하여 다운받은 경우 다운받은 파일 중 포함된 안내에 따라 vocabulary 파일에 통합함 (EULA는 athena 사이트 링크를 통해 설명을 숙지해야 함)
3	USAGI 설치	다음의 링크에서 파일을 다운받고 안내에 따라 설치함 (https://github.com/OHDSI/Usagi)
4	USAGI와 용어파일 통합	USAGI의 파일메뉴에서 vocabulary 파일들이 있는 위치를 지정
5	소스파일 정리	매핑하고자 하는 용어의 정보를 csv 파일로 정리함 (필수: 소스코드, 소스코드 영문설명; 추가: 추가설명 또는 빈도 등)
6	임시 매핑	USAGI에서 소스파일을 로드하여, 설명에 따라 sourcecode, description 및 additional info가 저장된 컬럼을 지정하고, 매핑조건(domain, vocabulary 등)을 지정하여 단어유사성 기반의 임시 매핑을 실행 (용량에 따라 수분에서 수십분이 소요될 수 있음)
7	본 매핑	임시매핑된 결과를 하나씩 확인하면서, 검색을 통해 새로운 concept_id로 수정하고, 매핑완료된 단어쌍에 대해 confirm하여 진행
8	교차비교	각 매핑은 2인에 의해 각각 이루어지며, 매핑결과를 비교하여 두사람이 다르게 매핑한 경우에 대해 논의하여 합의를 도출
9	3자검토	합의가 도출되지 못한 경우, 또는 추가 논의가 필요한 경우에 대하여 제 3자가 재검토를 통해 최종 매핑결과 도출

4) SOP for Vocabulary Mapping

3. 용어추출

* 지원기관에서 지원하는 EDI 코드:국제표준코드 매핑을 위해 참여기관의 코드와 빈도를 추출하여 주십시오.

*** 추출코드 및 필요사항**

	관련 테이블	추출내용	비고
1	Drug_exposure	LOCAL CODE, local코드명, EDICODE, EDI코드명, 처방빈도(빈도95%여부, 빈도99%여부)	* 약처방테이블에서의 코드 별 처방빈도 추출
2	Procedure_occurrence	LOCAL CODE, local코드명, EDICODE, EDI코드명, 처방빈도(빈도95%여부, 빈도99%여부)	* 진료관련 행위 처방테이블의 빈도 통합추출(수술, 처치, 마취, 검사 등) * EDICODE 6~8자리는 가격에 관련된 부분으로 실제 매핑은 5자리에 대해 이루어집니다.
3	Measurement	LOCAL CODE, local코드명, EDICODE, EDI코드명, 검체코드, 검체명, UNIT, 처방빈도(빈도95%여부, 빈도99%여부)	* LAB 검사결과 테이블의 빈도 추출 * 수치 또는 정형화된 text 결과값(p/n/+/++ etc..)을 갖는 검사가 이 테이블에 포함됩니다. * 각 'EDICODE + 검체' 항에 대해 LOINC 또는 SNOMED로 매핑됩니다.
4	Condition	-	* 전체 KCD코드의 매핑 제공
5	Device_exposure	LOCAL CODE, local코드명, EDICODE, EDI코드명, 처방빈도(빈도95%여부, 빈도99%여부)	* 치료재료 처방테이블에서의 코드 별 처방빈도 추출
6	전체	-	* 기관코드(local code) 단위로 추출해 주십시오. * 기관코드 분출이 어려운 경우, EDI코드 단위로 추출해 주시고, 기관코드:EDI코드 정보는 별도로 보관하여 주십시오. * 유효하지 않은 EDI코드는 제외하여 주십시오

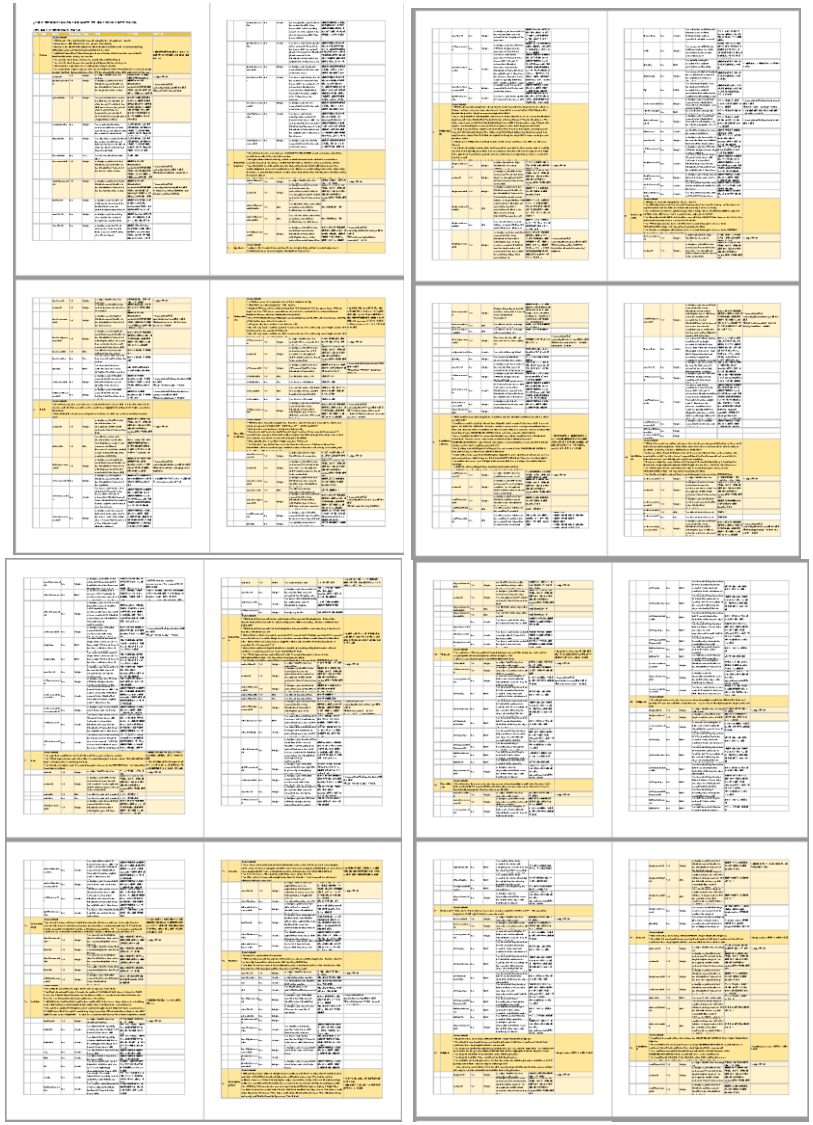
- Drug, procedure, device 코드는 EDI코드 기준으로 매핑됩니다.
- Measurement 코드는 랩원로컬코드+검체코드 쌍 기준으로 매핑됩니다.
- **처방빈도를 주시면 전체의 99% 정도는 아주대에서 계산하지만, 빈도 정보를 주기 어려우신 경우, 95%여부, 99% 여부를 'yes=1/no=0'으로 기재해 주시면 됩니다.**

*** 처방빈도 95/99% 여부 계산법 (빈도정보 제공이 어려운 경우만 참조)**

0	EDI코드가 없거나, 잘못된 코드인 경우 제외(Measurement는 로컬코드 기준으로 확인)
1	전체 처방빈도의 합 계산
2	처방빈도 내림차순으로 정렬
3	각 행의 누적처방빈도 계산(각 행의 계산: 그 행까지의 처방빈도의 합) (ex: F2=E2, F3=F2+E3)
4	각 코드의 '누적처방빈도/전체처방빈도합*100' 계산
5	누적처방빈도<=95 이면 95%여부='1'
6	누적처방빈도<=99 이면 99%여부='1'

	B	C	D	E	F	G	H	I
1	LOCALCODENAME	EDICODE	EDICODENAME	FREQUENCY	cumul_freq	cumul_perce	p95	p99
2	★N/S 20ml/Amp(중외)	644902691	중외생리식염주사액(수출용	6,153,207	6,153,207	0.058	1	1
3	MUCOSTA 100mg/Tab(오츠카)	649900060	무코스타정(레바미피드)	1,823,888	7,977,095	0.0752	1	1
4	★N/S 50ml/BT(대한)	645100621	대한헬균생리식염수50밀리	1,746,410	9,723,505	0.0917	1	1
5	★N/S 1L/Bag(CJ)	640001091	씨제이0.9%생리식염주사액	1,459,923	11,183,428	0.1054	1	1
6	MVH 5ml/V(휴온)	657200570	엠브이에취주(바이알)	1,256,041	12,439,469	0.1173	1	1
7	Stillen 60mg/Tab(동아)	642502290	스티렌정(애엡95%에탄올인	1,126,958	13,566,427	0.1279	1	1
1592	TD-PUR INJ(노바티스)	650003060		3,236	104,995,649	0.9898	0	1
1593	Someto SC 320mg/Cap(함비오)	659900100	쏘메토320밀리그램연질캡슐	3,221	104,998,870	0.9898	0	1
1594	Petrisol 10% 500ml/BT(엠지랑)	642304060	페트리솔주 500밀리리터	3,202	105,002,072	0.9898	0	1
1595	★Mirtax 30mg/Tab(한국산도스)	668100030	미르탁스정30밀리그램(미르	3,184	105,005,256	0.9899	0	1
1596	MOPRAM 100mg/5ml/Amp(명문	649800990	명문모프람주사(염산독사조	3,174	105,008,430	0.9899	0	1
1597	Jarsin 300mg/Tab(한독)	652101010	야르신정(히페리시80%/에탄	3,174	105,011,604	0.9899	0	1

5) SOP for ETL documentation



GOVERNMENT'S INTEREST ON CDM

Korean government formally launched a task-force team for distributed bio-dig data sharing

Minister of Mistry of Trade, Industry and Energy of South Korea formally announced that they launch a TF team for distributed bio-dig data sharing to build a national biomedical data sharing platform and environment. This year, they will make a master plan for it and prepare budget to realize it.

dongA.com

신문보기 | 보이스뉴스 뉴스 오피니언 연예 스포츠 스튜디오 비즈N ENG | 中文 | 日文

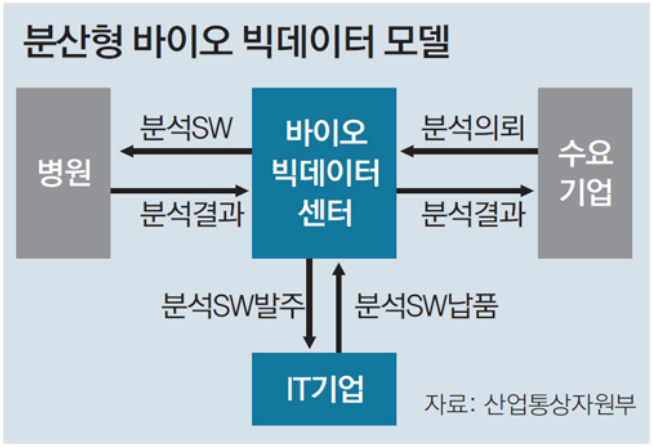
Apr 18, 2017

경제

병원 의료데이터 분석치, 제약사 등에 제공 길 열린다

박민우기자 입력 2017-04-18 03:00 수정 2017-04-18 03:00

정부, 맞춤형 의료서비스 제공위해 민관 바이오 빅데이터 구축하기로
개인정보法 따라 원본은 제공 안해



#

정부가 4차 산업혁명 시대에 걸맞은 맞춤형 의료 서비스를 제공하기 위해 바이오 빅데이터 시스템을 만드는 작업에 착수한다.

산업통상자원부는 17일 경기 성남시 판교 메리어트호텔에서 '바이오헬스 업계 간담회'를 열고 이 같은 내용을 포함하는 바이오헬스 산업 발전전략을 발표했다.

주요뉴스 1/3

"시진핑 부패 폭로" 자산19조 재벌, 당국과 ...

文 37.5%, 安 26.4%...격차 두자릿수로 벌어져

美 핵잠수함 미시간호, 25일 부산항 온다...北에 대한 강력한 경고?

美, 北 선제타격 못하는 이유가 한국 수도권 2500만 인구 때문이라고?

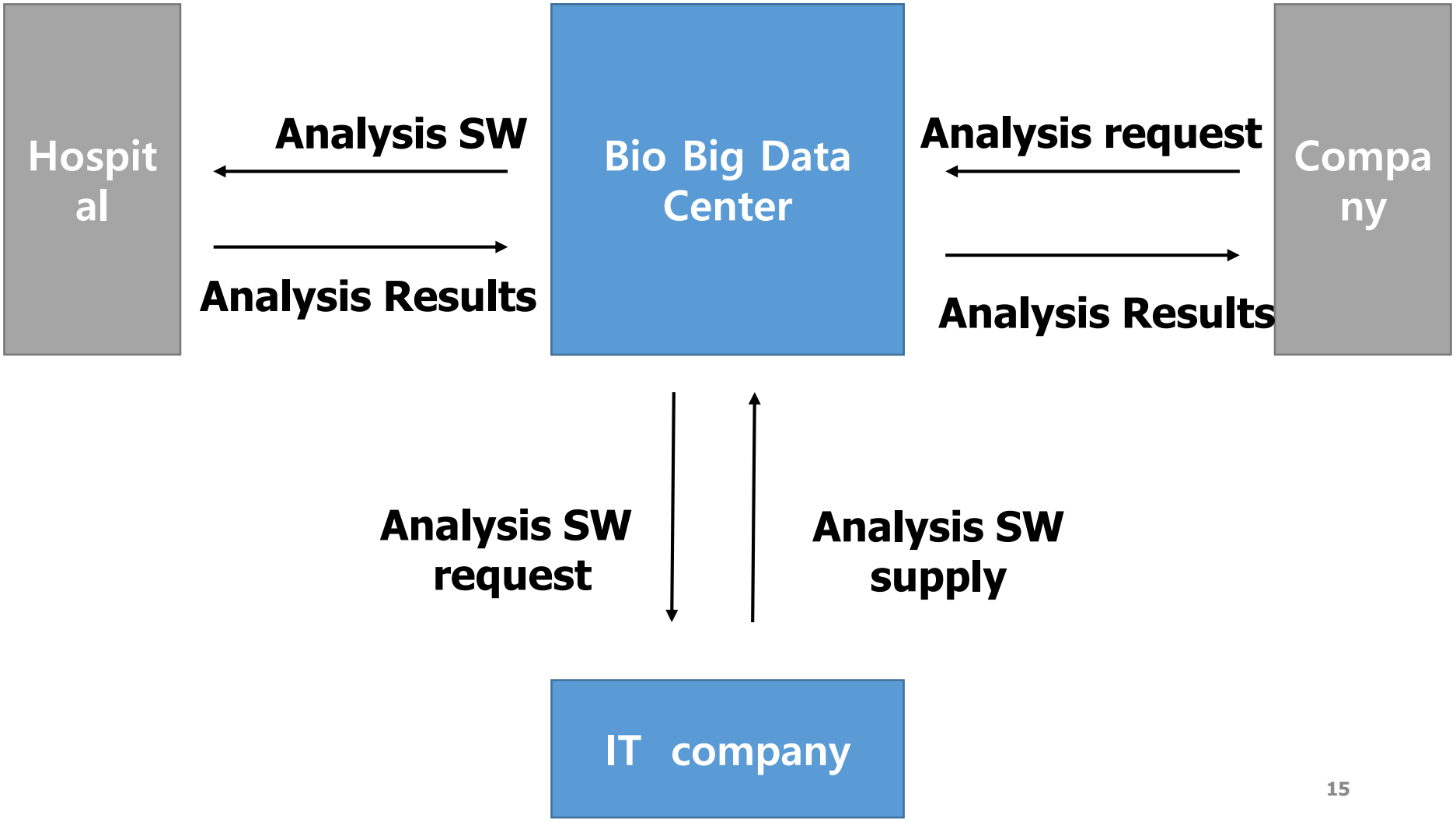
바른정당 긴급의총 돌입...유승민 후보 사퇴 여부 논의

장시호 "특검서 만난 崔, 삼성동 사저 돈 빼내 정유라 키우라고 해" 주장

칩거 깬 김한길 "문재인, 盧비서실장이라 지도자 추대...난센스"

김진태 "安, 변성기 중1 같아...더듬거리는 거 보면 때려주고파"

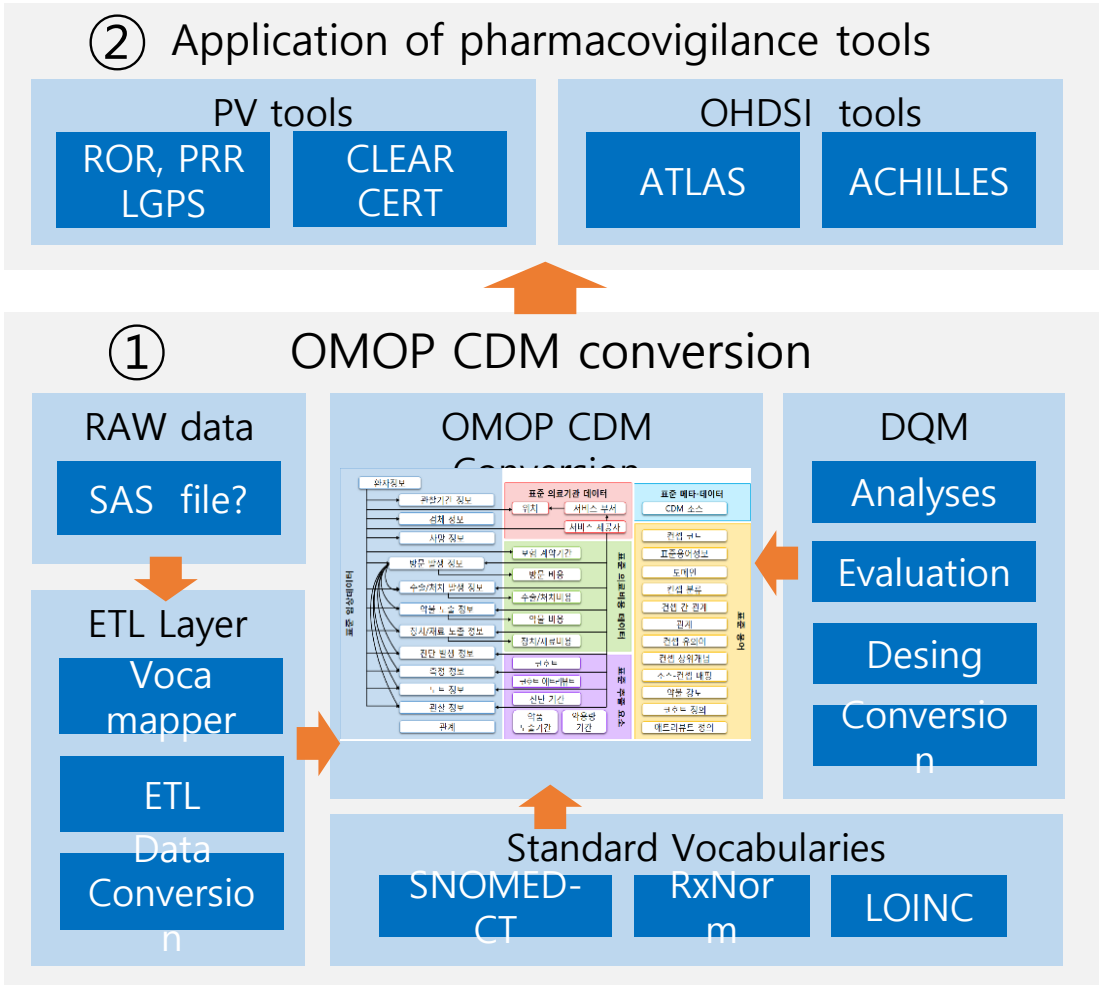
Distributed Bio Big Data Model



NHIS*, Development of a CDM-based Drug Safety Surveillance System

Three year project
2016-208

- 1st year: feasibility
  12 year of 1M pt data into CDM
- 2nd year; validation of usefulness of the CDM
- 3rd year : Full conversion
  12-years of 51 M patients



*National Health Insurance Service: Governmental National Health Insurer

Collaboration and competition in between government

- Ministry of Trade, Industry and Energy
- Ministry of Health and Welfare
- National Health Insurance Service
- Ministry of Food And Drug Safety
- Ministry of Science, ICT and future planning

DRN, OMOP-CDM
EMR + Omics + life-log

merge, HL7 CDA?
OMOP-CDM?
EMR + Claim

DRN, OMOP-CDM
Claim + Health Exam

DRN, K-CDM (?)
EMR

HIS with CDM

Characteristics of Korean OHDSI

Korean OHDSI

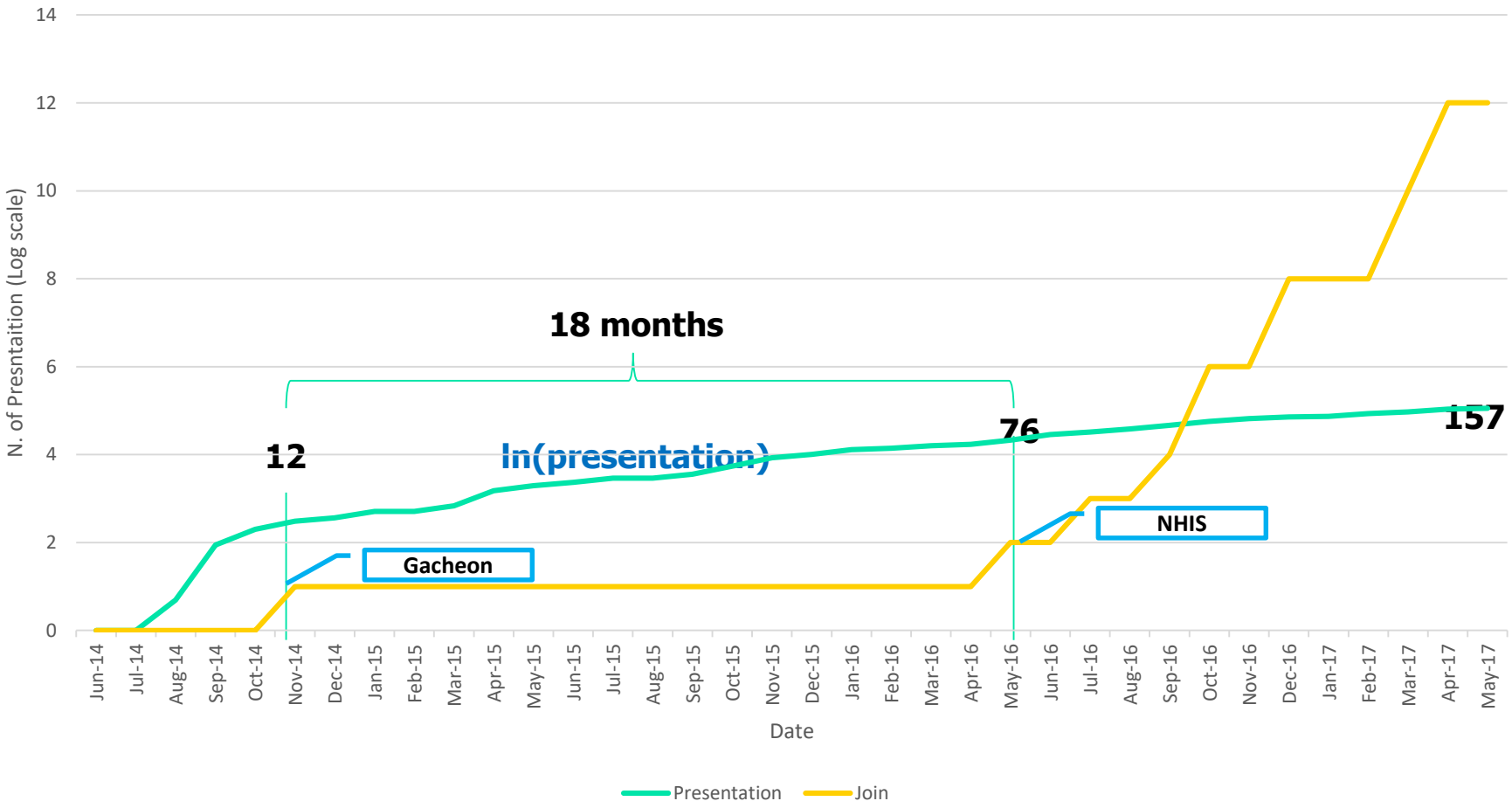
- Data partners:
  - Major tertiary teaching hospitals
    - Detailed time stamp
    - Test results
    - Outcome data
- National Health Insurance Data
  - Compulsory health insurance
  - Claim data + socioeconomic data + regular Health exam data (includes lab tests)
  - 12-year of observation period
  - Covers all the citizens (51M)

WHY DRN AND CDM IS POPULAR TO KOREA?

Invited Talks

- 157 invited talks during past 33 months since July 2014.
  - 2014: 13 times
  - 2015: 42 times
  - 2016: 74 times
  - 2017: 28 times

No. of Presentations and Data Partner Join



Lesson learned from potential Data Owner

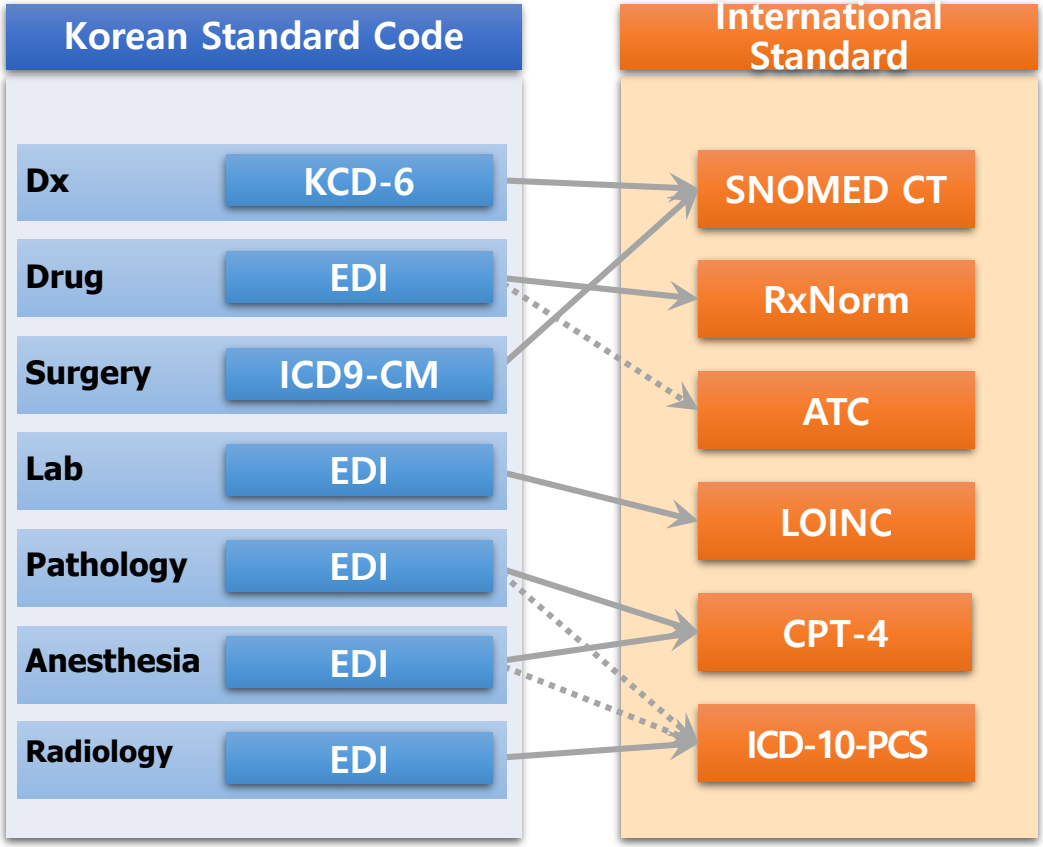
- Quick-prototyping
- Live demonstration
- Success story
- Focusing on clinicians

Lesson learned from potential Data Owner

- Quick-prototyping
  - Launch Achilles ASAP!
  - Need vocabulary mapping!

Code Mapping

>170K codes to map



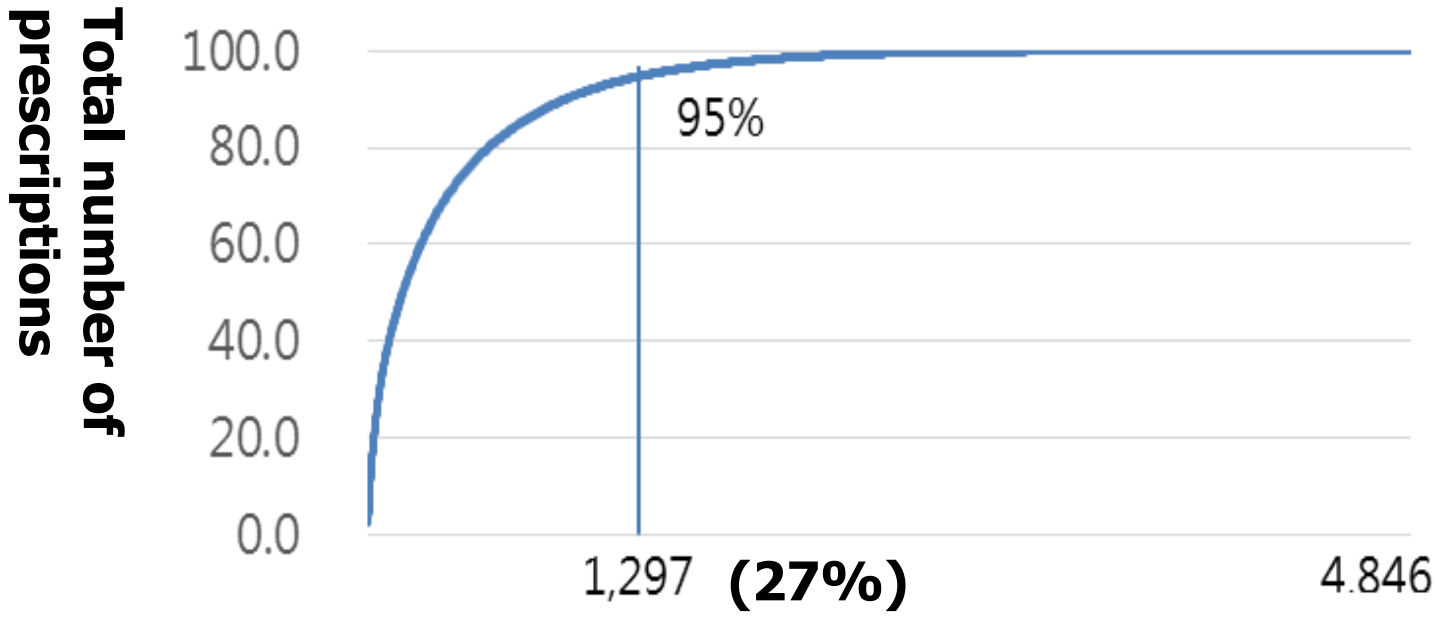
- EDI (Electronic Data Interchange) code in Korea
  - National standard for national health insurance claim
  - Managed by Health Insurance Review & Assessment Service (HIRA)

Strategy for Standard Vocabulary Mapping

- Sort codes by frequency of usage
- Number of codes required to cover

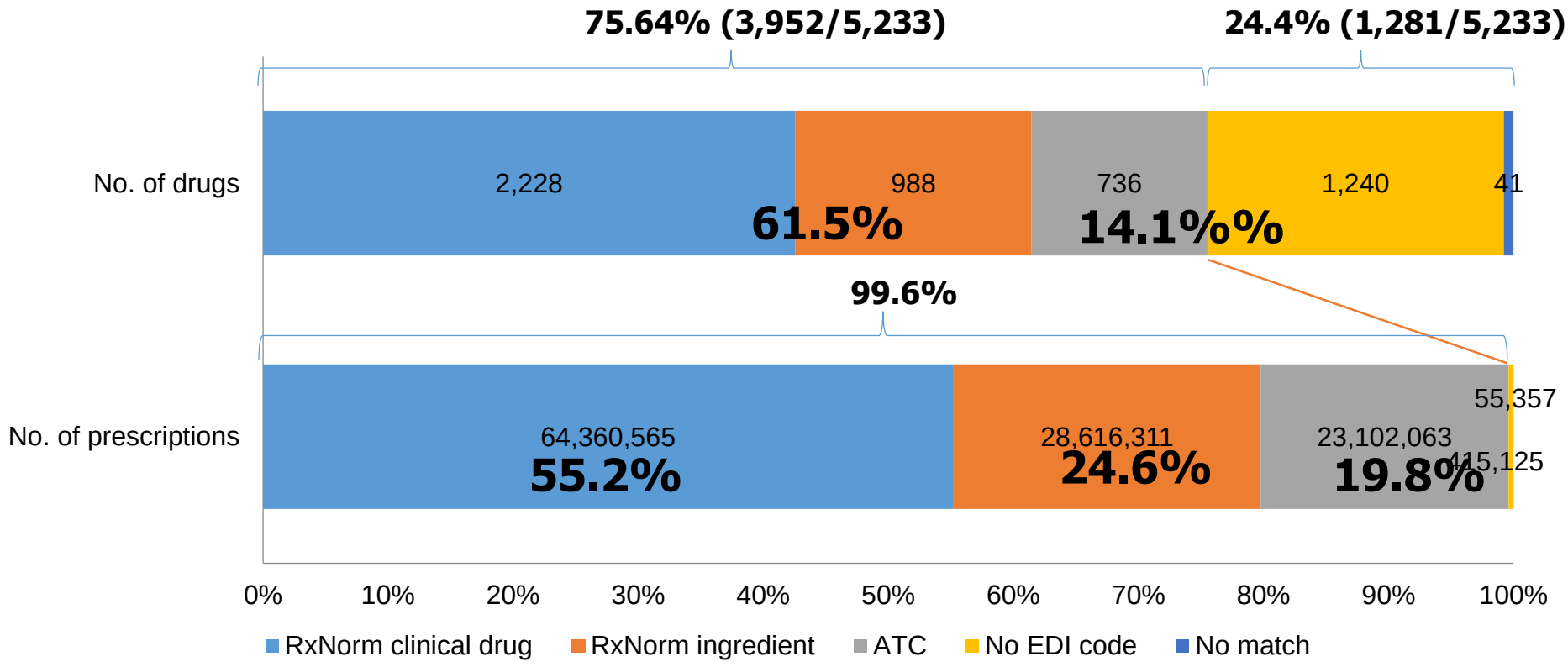
95 % of transaction data → 99% → 100%

20-30% of codes 60-70% of codes



DRUG code mapping (Ajou Univ.)

Source code: local code of Ajou university hospital (mapped with EDI, partially)
Standard code: RxNorm, ATC



Lesson learned from potential Data Owner

- Live demonstration using RWD

Achilles: data characterization

Openness: http://ami.ajou.ac.kr:8080

Malignant tumor, Breast
Ajou Univ. Hospital 2.3M 22years

Treemap | Table

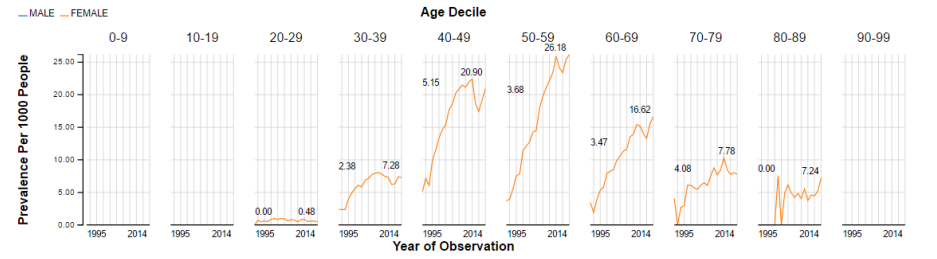
Search: breast | Show / hide columns

SOC	HLT	SNOMED	Person Count	Prevalence	Records per Person
NA	NA	Benign tumor of breast	27,261	1.20%	3.61
NA	NA	Primary malignant neoplasm of breast	5,920	0.26%	28.90
NA	NA	Fibrocystic disease of breast	2,968	0.13%	2.76
NA	NA	Breast finding	1,671	0.07%	2.76
NA	NA	Primary malignant neoplasm of breast upper outer quadrant	1,539	0.07%	5.44

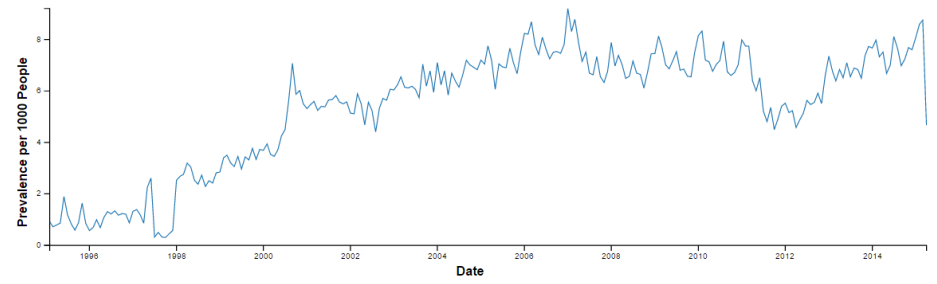
Showing 1 to 5 of 34 entries (filtered from 5,156 total entries) Previous 1 2 3 4 5 6 7 Next

Primary malignant neoplasm of breast

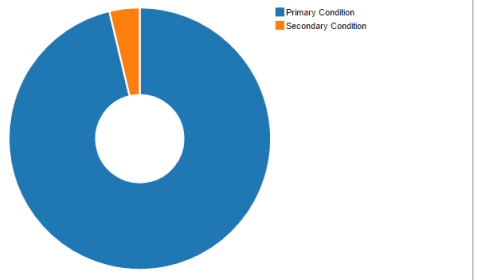
Condition Prevalence



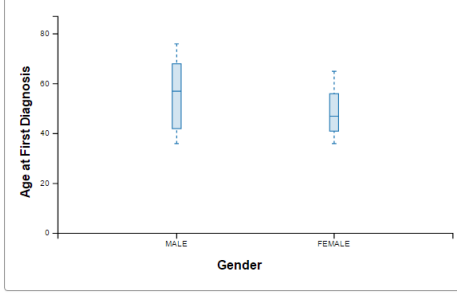
Condition Prevalence by Month



Conditions by Type



Age at First Diagnosis



Malignant tumor, Breast
NHISS, 1M sample cohort, 10 years

Conditions

Treemap | Table

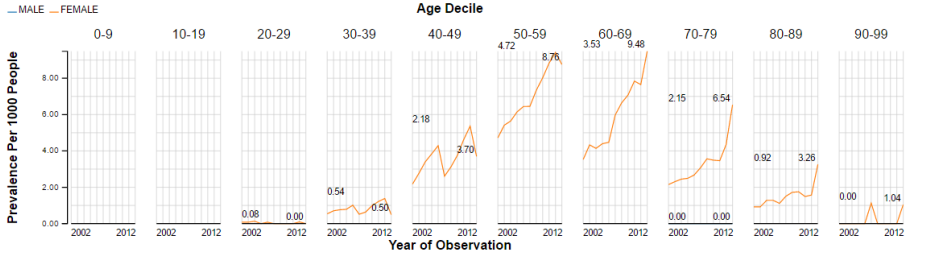
Search: breast | Show / hide columns

SOC	HLT	SNOMED	Person Count	Prevalence	Records per Person
NA	NA	Inflammatory disorder of breast	11,296	1.00%	2.03
NA	NA	Fibrocystic disease of breast	7,569	0.67%	2.06
NA	NA	Fibroadenosis of breast	7,164	0.64%	1.65
NA	NA	Primary malignant neoplasm of breast	4,701	0.42%	29.01
NA	NA	Solitary cyst of breast	3,405	0.30%	1.49

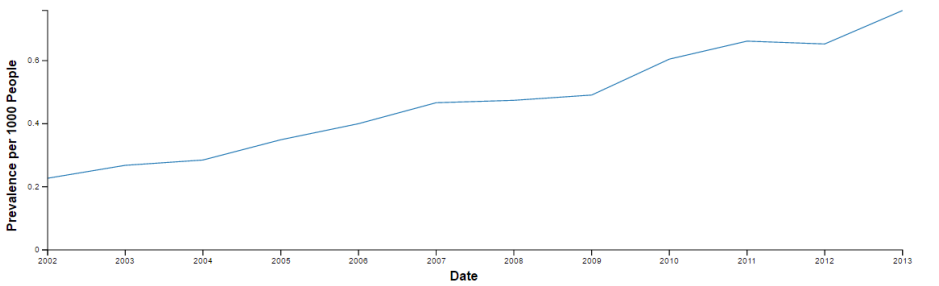
Showing 6 to 10 of 39 entries (filtered from 5,989 total entries) Previous 1 2 3 4 5 … 8 Next

Primary malignant neoplasm of breast

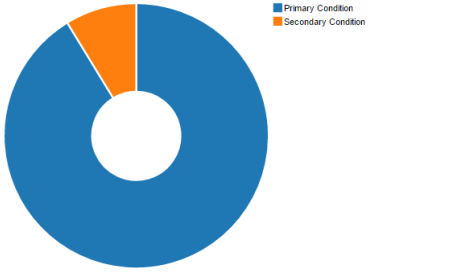
Condition Prevalence



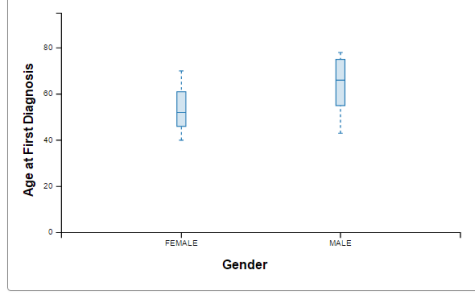
Condition Prevalence by Month



Conditions by Type



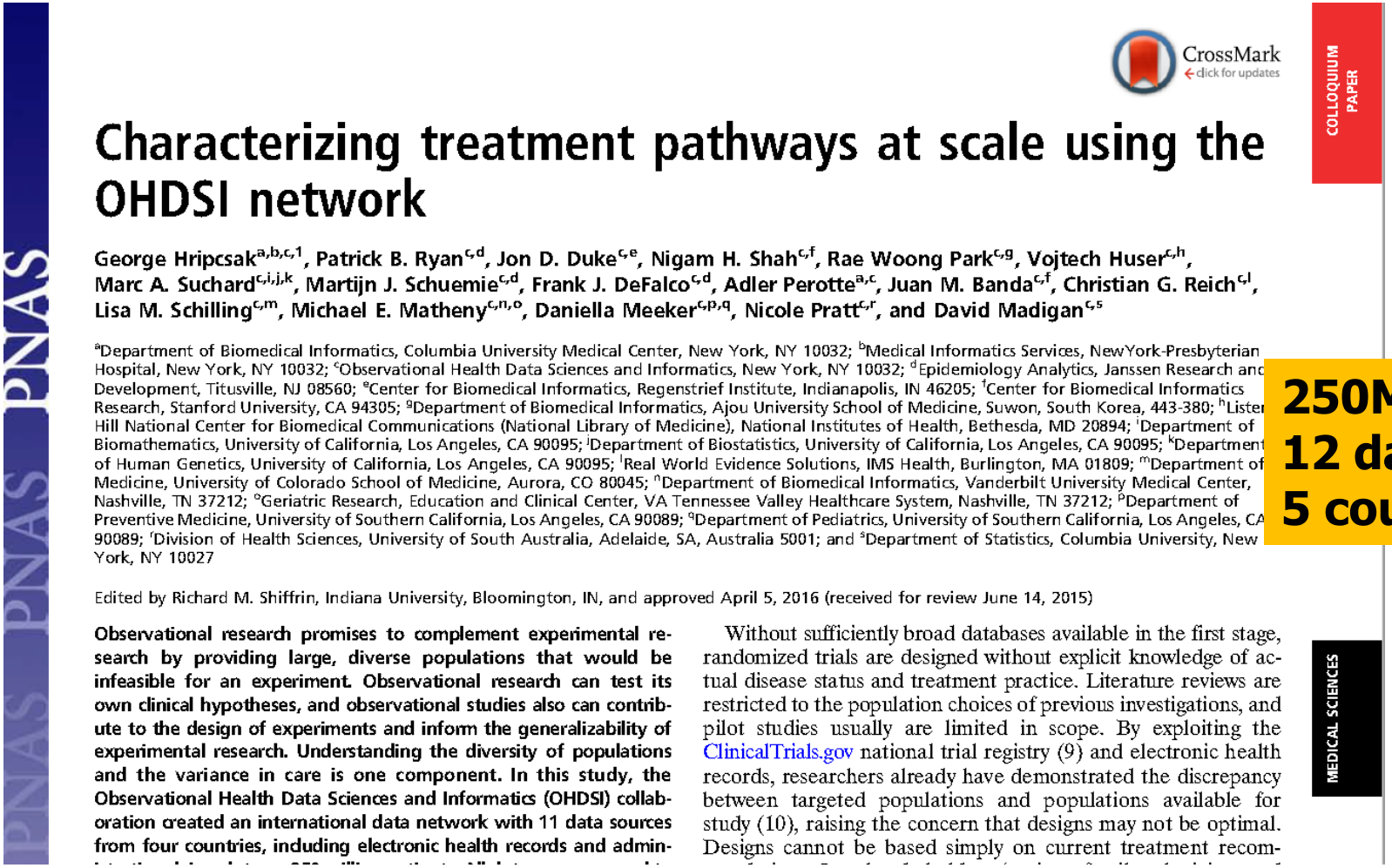
Age at First Diagnosis



Lesson learned from potential Data Owner

- Using a Success story and role model

Examples of clinical researches using big data



PNAS PNAS PNAS

CrossMark click for updates

COLLOQUIUM PAPER

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

[a]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; [b]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; [c]Observational Health Data Sciences and Informatics, New York, NY 10032; [d]Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; [e]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; [f]Center for Biomedical Informatics Research, Stanford University, CA 94305; [g]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; [h]Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; [i]Department of Biomathematics, University of California, Los Angeles, CA 90095; [j]Department of Biostatistics, University of California, Los Angeles, CA 90095; [k]Department of Human Genetics, University of California, Los Angeles, CA 90095; [l]Real World Evidence Solutions, IMS Health, Burlington, MA 01809; [m]Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; [n]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; [o]Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; [p]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; [q]Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; [r]Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and [s]Department of Statistics, Columbia University, New York, NY 10027

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

Observational research promises to complement experimental research by providing large, diverse populations that would be infeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also can contribute to the design of experiments and inform the generalizability of experimental research. Understanding the diversity of populations and the variance in care is one component. In this study, the Observational Health Data Sciences and Informatics (OHDSI) collaboration created an international data network with 11 data sources from four countries, including electronic health records and admin-

Without sufficiently broad databases available in the first stage, randomized trials are designed without explicit knowledge of actual disease status and treatment practice. Literature reviews are restricted to the population choices of previous investigations, and pilot studies usually are limited in scope. By exploiting the ClinicalTrials.gov national trial registry (9) and electronic health records, researchers already have demonstrated the discrepancy between targeted populations and populations available for study (10), raising the concern that designs may not be optimal. Designs cannot be based simply on current treatment recom-

MEDICAL SCIENCES

250M patients
12 database
5 countries

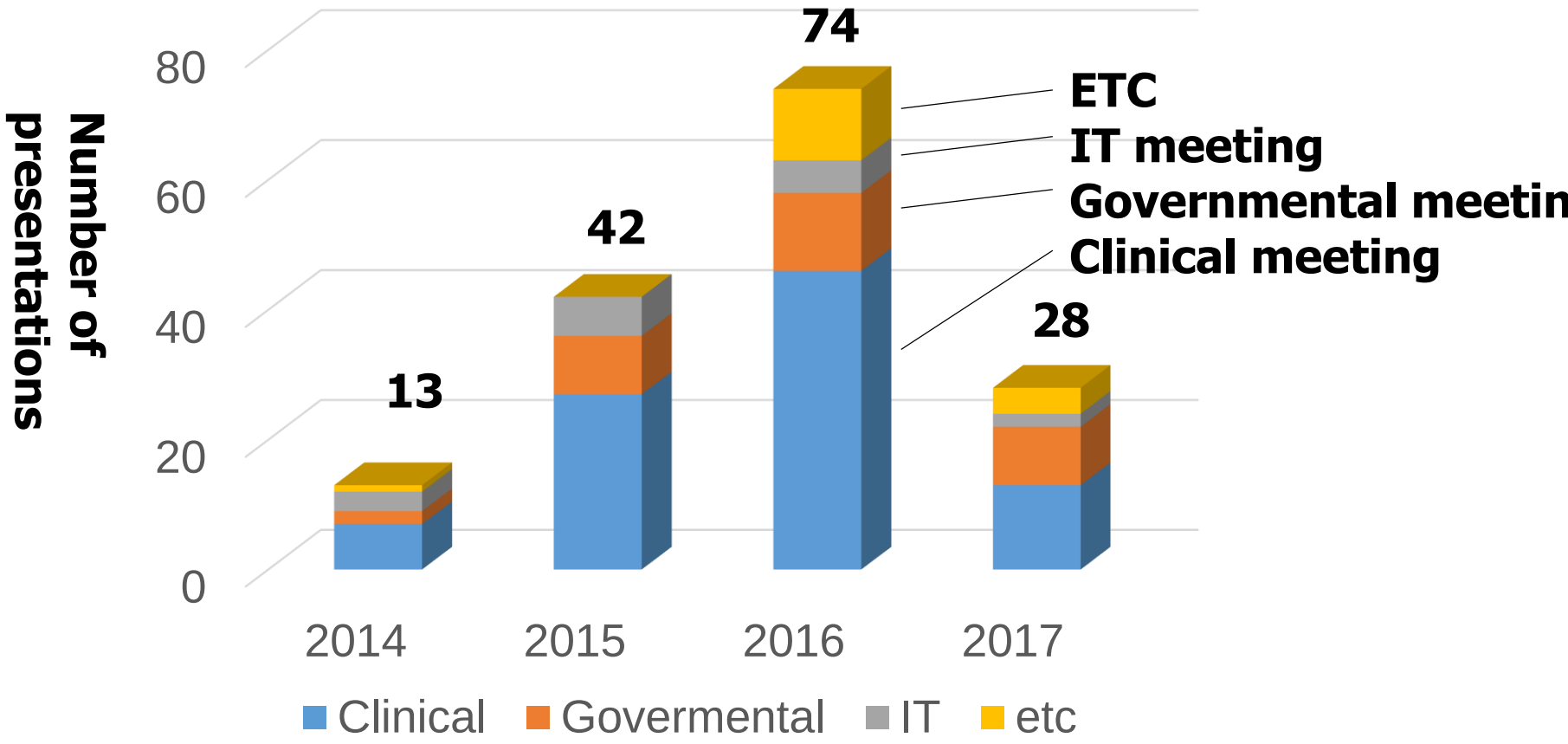
Lesson learned from potential Data Owner

- Focusing on Clinicians

Focusing on clinicians

- Clinicians, rather than informaticians: major decision makers in a hospital are usually clinicians.
- Young clinicians (assistant – associate professors) : highest interest on OMOP CDM, because they do not have enough fund, resources and data for their research.
- Inter-disciplinary clinical meeting
  - then homogenous clinical meeting at initial stage

Invited Talks by domain



HOW ABOUT PROVIDE FUND OR INCENTIVES TO HOSPITAL?

Not successful

Challenges and plans in Korean OHDSI

- Expanding the data partners
  - All the tertiary teaching hospitals
  - Most of general hospitals
  - Some of private clinics
  - Most of pharmacies
  - All the claim data covering all the Korean population
- Governance for data sharing
  - No IRB for pre-defined and verified analyses
  - Open all the Achilles website of Korean data partner to public
  - Maintain regular leadership meeting
- Free sample data
  - For training
  - To test analytic code
  - For Feasibility test
- Real-time CDM
  - Up to date information
  - Real-time eligibility screening
  - CDSS
- CDM-based PHR
- Expansion of CDM model for
  - bio-signal data
  - Genomic data
  - Image data

Summary

- About South Korea
- Korean OHDSI Network
- Why DRN and CDM is popular to Korea?
- Lesson learned from potential Data Owner
  - Quick-prototyping
  - Live demonstration
  - Success story
  - Focusing on clinicians
- Projects/Tools
- Future plans in Korean OHDSI

Acknowledgements

Dukyong Yoon, MD, PhD
Hyun Wook Han, MD, PhD
Song Vogue Ahn, MD, PhD
Soo Yeon Cho, MPH
DaHye Shin, BS
JungHyun Byun, BS
Hojun Park, BS
MinSeok Jeon, BE
Sungjae Jung, BE
Doyeop Kim, BE
Sanghyung Jin, MS
Tae Young Kim, BE
Seojeong Shin, MS
Jaehyeong Cho, BS
Lee hye jin, MBA
Eun Sung Kim BS
Jaehee Han, BN
So Young Eo, BN
Hyo Hung Kim,BN
Eugene Jeong, BS
Seungbin Oh, BPharm
Yourim Lee, BS
Seng Chan You, MD, MS
Haze Lee
Jang ho Lee, BE
Taehwan Kim, BE
A Rum Lee, BS
EunBi Kim, BS
Yeo Kyung Lee, BS
Seong-yun Oh
Hyukjun Cho, MS
EunKang Kim
Yun Yeon Ju
Yujung Kwon
Soojung Cho, MPH