

# Clinical **RESEARCH** Data Warehouse in AMC

서울아산병원  
의생명정보학과  
신수용  
2013. 6. 13

@ KOSMI 2013 Spring



**ASAN**  
Medical Center

# 개요

1 연구와 관련된 국내 규제 강화

2 연구용 DW의 필요성

3 Clinical Research Data Warehouse의 특징

4 익명화

5 Pilot Programs

# 연구와 관련된 국내 규제 강화 (1/2)

## 개인정보보호법

- 2011년 9월 30일 시행
- 개인정보 이용 제한
  - 제18조 2항 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 **특정 개인을 알아볼 수 없는 형태로 개인정보를 제공**하는 경우” 예외로 인정
- 개인정보 수집 제한
  - 제15조 2항 “1. 개인정보의 수집, 이용 목적, 2. 수집하려는 개인정보의 항목, 3. 개인정보의 보유 및 이용 기간” 등을 반드시 통보하여야 함

## 생명 윤리 및 안전에 관한 법률

- 2013년 2월 2일 시행
- 개인정보 이용 제한
  - 제18조, 제38조, 제43조: **연구를 위해서 반드시 개별 동의를 받거나 또는 익명화 처리를 하여 기관위원회의 심의를 받아야 함**
- 인간대상연구
  - 사람을 대상으로 물리적으로 개입하는 연구
  - 설문조사, 행동관찰 등으로 자료를 얻는 연구
  - 개인을 식별할 수 있는 정보를 이용하는 연구
- 인체유래물연구

# 연구와 관련된 국내 규제 강화 (2/2)

연구의 경우는 진료 외 목적인 2차 이용이기 때문에  
반드시 개별 동의가 필요함

모든 연구(인간대상연구, 인체유래물연구, 설문조사 등)는  
IRB의 심의를 받아야함

※ 심의면제 대상도 포함 (신속 심의 가능)

대안은?

**익명화한다면 심의면제, 동의면제 대상**

# 진료와 연구는 같은가 ?

진료



환자, 입원/외래, 진료예약,  
수납, 청구삭감  
처방,검사,수납,수술,물류

의료법

연구

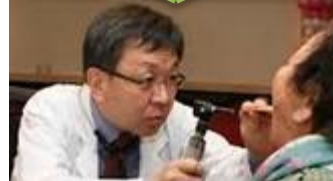


피험자, 전향적/후향적연구,  
임상시험, IRB승인,이상반응,  
Audit, 증례기록(CRF)

생명윤리 및 안전에 관한 법률

진료정보

연구정보



진료환자 = 피험자  
될 수 있지만,

관련 법, 인프라, 정보관리, 지원행정 등은

**진료-연구 엄격히 구분됨**

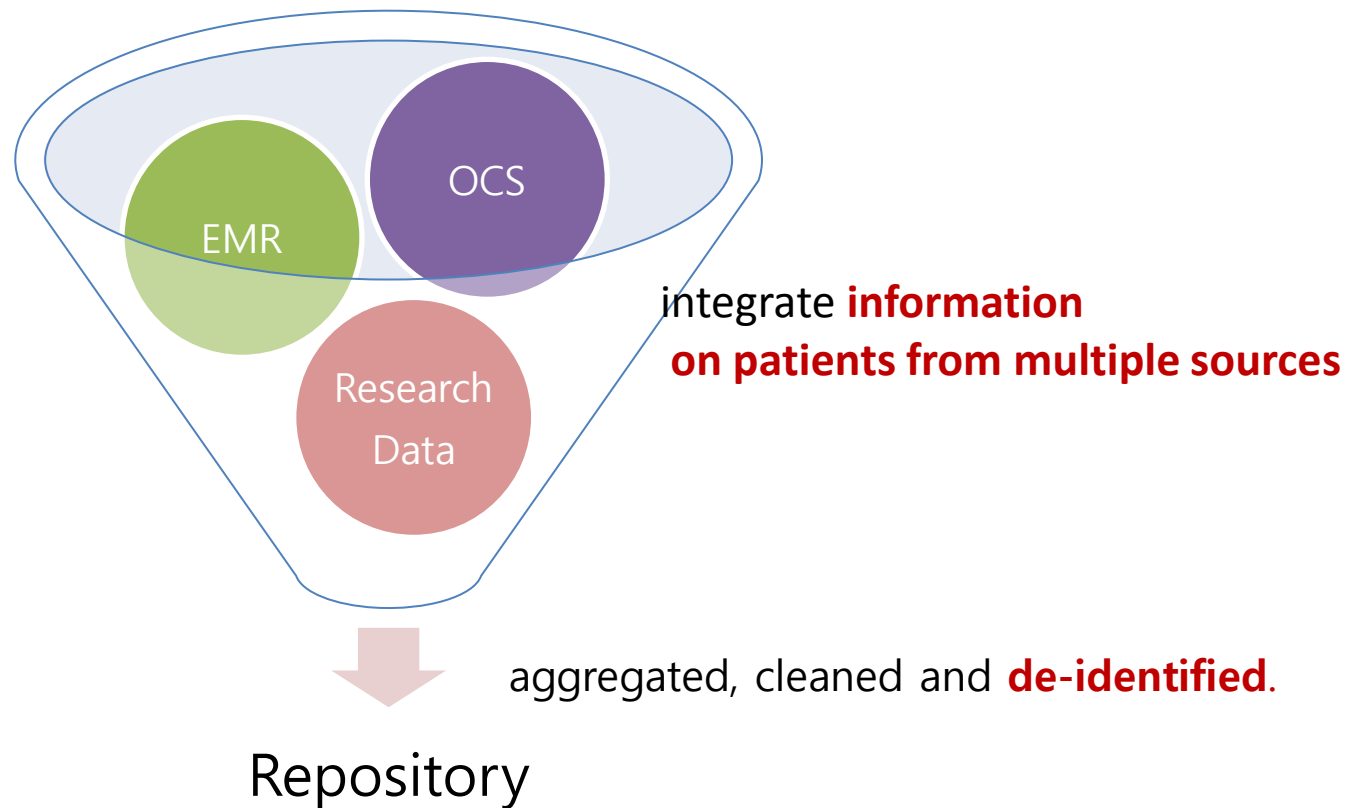
# Clinical Research Data Warehouse

## WHY?

- 연구를 위해서는 (돈과 관련된) 원무 정보는 크게 필요하지 않다.
- EMR외에 기존 연구관련 정보 (연구자 DB, 각종 질병 registry 등)등을 통합해서 활용해야 편리하다.
- 중개연구를 위해서는 각종 다양한 유전정보와 같은 BIO DB 들이 연동되어야 한다.

즉, 기존의 경영정보를 위한 Enterprise DW와는 다른  
연구용 DW가 필요하다.

# Clinical Research Data Warehouse



# Uniqueness of CDW

Conventional DW	Clinical DW (for Research)
Providing decision <b>support</b> <b>for queries of a financial</b> nature	Focused on tasks like <b>cohort identification</b> and <b>hypothesis</b> <b>generation/analysis</b>



Categories	CRDW
Aim	Clinical & Translational Research
Needs	<b>Data Extraction &amp; Review,</b> <i>De-Identification</i>
Systems	<b>Data Extraction System</b> <b>Chart Review System</b> Database
Interface	<i>IRB approval process</i> Biologic Specimen search
<b>Essential functions</b>	<b>Research Design (Cohort)</b> <b>Chart Review (De-Identification)</b> <b>Data Extraction (IRB approval)</b>
Subject Area	Research Data ex> Disease, Lab., Medication
Users	Researchers
Internal Data	EMR data, Bio Specimen data
Outside Data	e-CRF data
Metadata	+
Structured	Text & structured (pathology)
<b>De-Identification</b>	<b>De-Identification(design/chart review)</b> <b>Identified (IRB approval)</b>

# Key Elements of CRDW

Key Elements	Explanation	Roles
<b>Honest Broker</b>	De-identification of ID or indefinable. For Data retrieval, Chart review	De-identified
<b>Data Extraction</b>	Researcher: - direct query, OLAP, research design DRM module	- Cohort Discovery, Clinical Data Extraction - Access Control
<b>Chart Review</b>	De-identification EMR chart review	Clinical Data Review
<b>IRB</b>	Research approval, Wavier.	Approval
<b>Privacy &amp; Security</b>	Hospital policy, HIPAA compliance	Security Regulation

# CRDW vs. EDW

구분	항목	CRDW(연구정보)	EDW(진료정보)
도입	목적	연구용 - 연구 설계 - 연구 자료 추출	병원 경영 - 경영 통계 추출
	요구사항	- 자료추출/리뷰 - 익명화	- 통합 보고서
아키텍처	시스템구성	- 자료추출 시스템 - 검색 시스템 - 리뷰 시스템 - Database	- OLAP tool (자료추출) - Database
	인터페이스	- CDW - IRB - CTMS(Clinical Trial Management System) - BRC - eCRF	- EDW - EMR, OCS - 서식생성기
사용자 프로그램	client	자체 개발 도구	사용 도구
	사용자	연구자	행정 담당자
	사용권한	모든 연구자	특정 사용자
Data	내부 Data	- 환자의무기록 (EMR, OCS, PACS 등) - 검제정보 - 원무 정보 일부분	- EMR, OCS - 원무정보 (PACS 불필요)
	외부 Data	- 과별/개인별 연구DB - public research DB - e-CRF - 논문정보 - 기타 연구에 필요한 외부 정보	- 심사평가원 등
	Metadata	연구에 관련된 용어 전부 관리 필요	행정 용어 위주
	구조화	구조화 필수	구조화 선택 사항
	Text	코드화 필수	불필요
	익명성	익명화 필수	익명화 불필요
정보보호	법적 규제	생명윤리 및 안전에 관한 법률	개인정보보호법

# 익명화 대상

- 직접적 개인식별정보 (Direct Identifier)
  - 삭제 or Masking
- 간접적 개인식별정보 (Indirect Identifier)
  - 익명화 대상 아님
- 조합을 통해 개인식별이 가능한 정보 (Quasi Identifier)
  - Ex) 희귀 질환

# 익명화 참고 방법

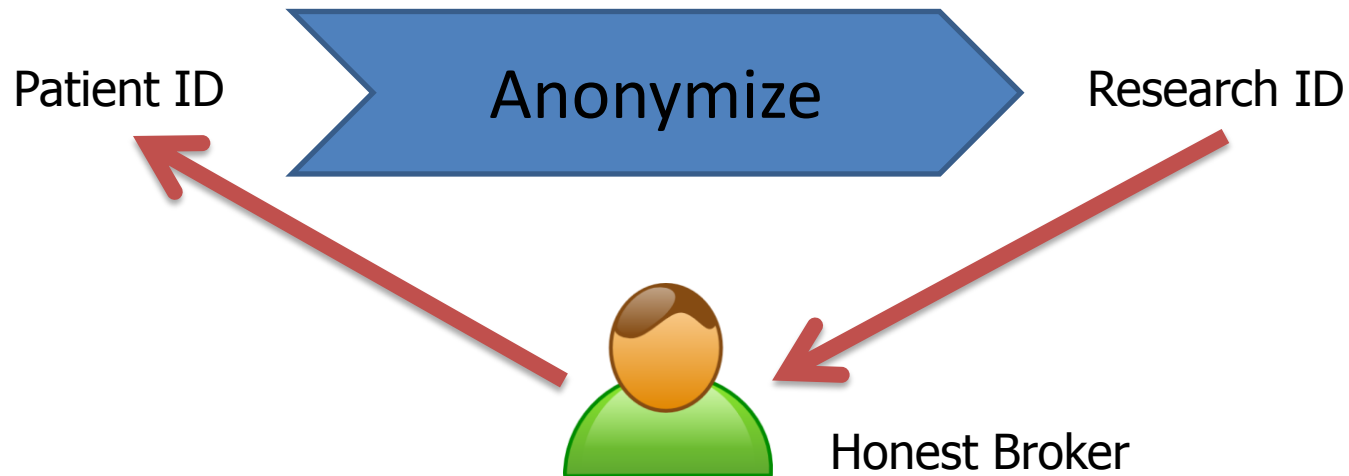
- **HIPAA Expert Determination Methods**
  - HIPAA에서 제시한 18 PHI들을 최대한 제거하고 해당 분야 **전문가가 신뢰성을 판단**
- **Safe Harbor Methods**
  - 모든 PHI를 제거
  - 현실적으로 불가능

# 개인식별정보

- 개인식별정보의 정의가 필요함
  - Ex) HIPAA 18 PHI
- 서울아산병원의 경우 20개를 개인식별정보로 정의함

# 등록번호 제거방법

**Generating new research ID for each patient**



: developed in-house anonymization function for re-identification

# Quasi ID 제거 방법

간접적인 개인 식별 정보 (Indirect Identifier)는?



조합을 통해 개인 식별이 가능하다고 판단되는  
정보 (quasi-identifier) 제거 필요



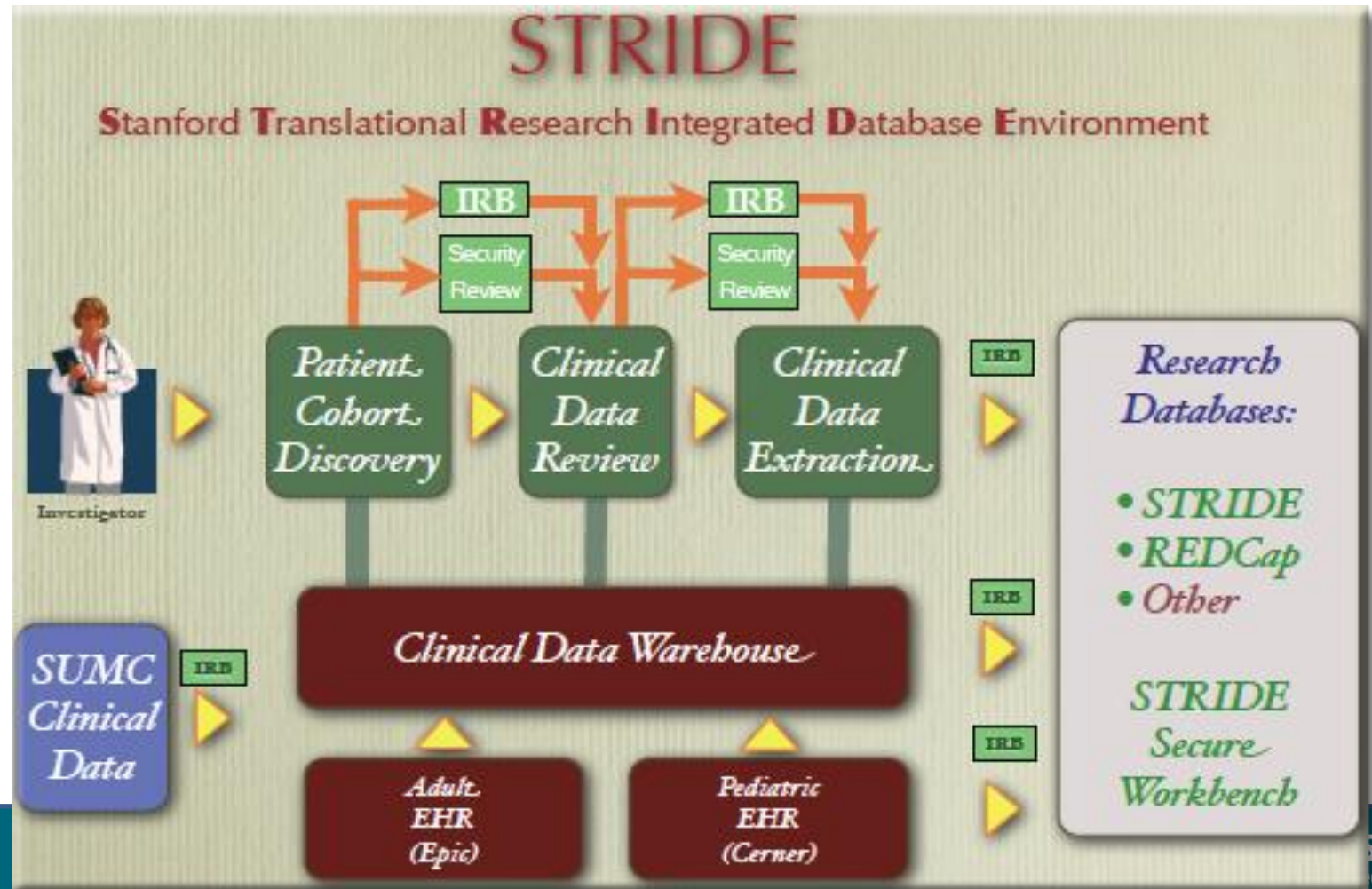
# 기존 프로그램 성능

## Automatic de-identification systems (2012년 Cincinnati Children's Hospital Medical Center Paper)

Name	Document Types	PHI	Precision	Recall
MIST1	22 note types (incl. ER, OR, inpatient, outpatient) 3503 clinical notes	18 PHIs	92.05	91.02
MIST2			92.79	92.81
MCRF1			95.25	89.86
MCRF2			95.08	91.92
Human Annotator 1	22 note types, 10% of 3503 clinical notes		93.95	92.15
Human Annotator 2			88.45	94.55
MCRF2			95.73	92.91

# Stanford Translational Research Integrated Database Environment (STRIDE)

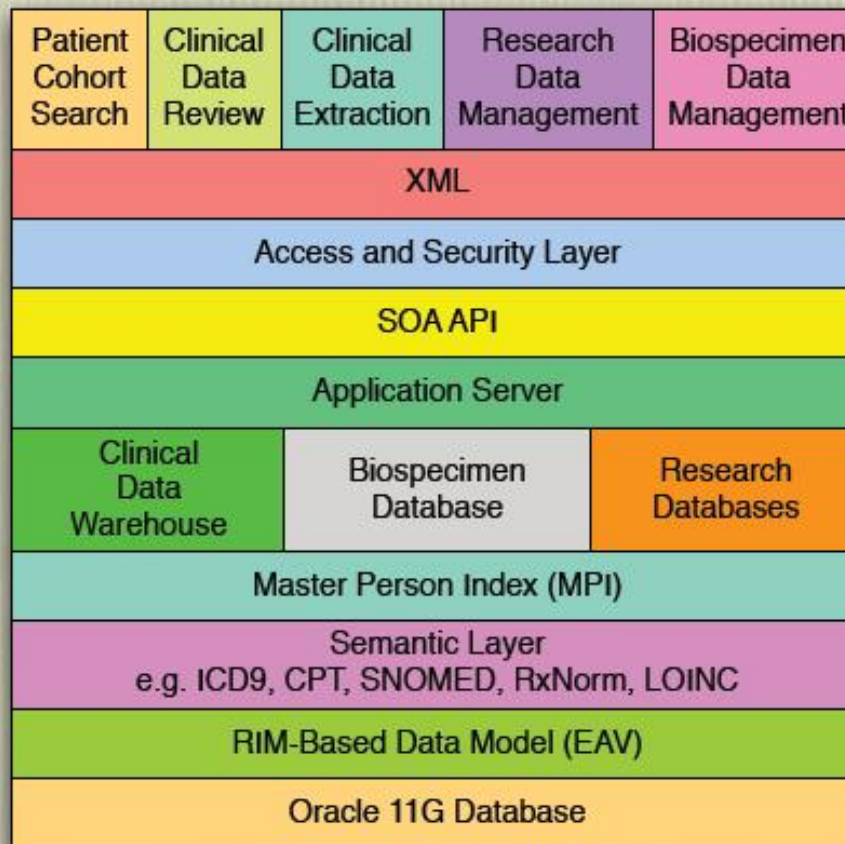
A secure, HIPAA-compliant, integrated repository of clinical and research information



# Stanford Translational Research Integrated Database Environment (STRIDE)

## STRIDE Architecture Stack

3-tier infra with SSL encryption



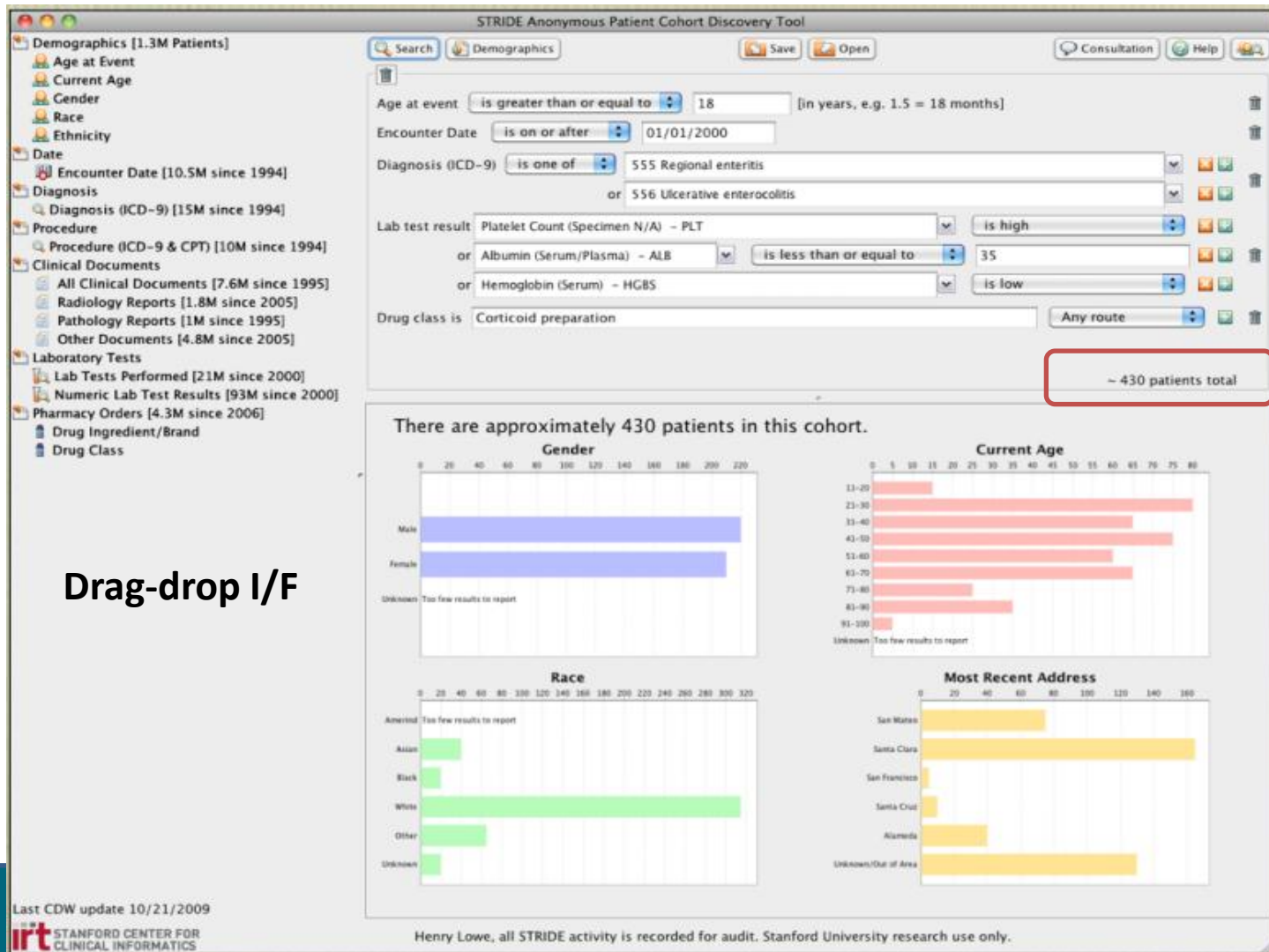
Swing 사용, cross-platform

JSON, Java

HL7 version 3 RIM

# STRIDE Cohort Discovery Tool

연구에 활용될 수 있는 환자군 크기 파악

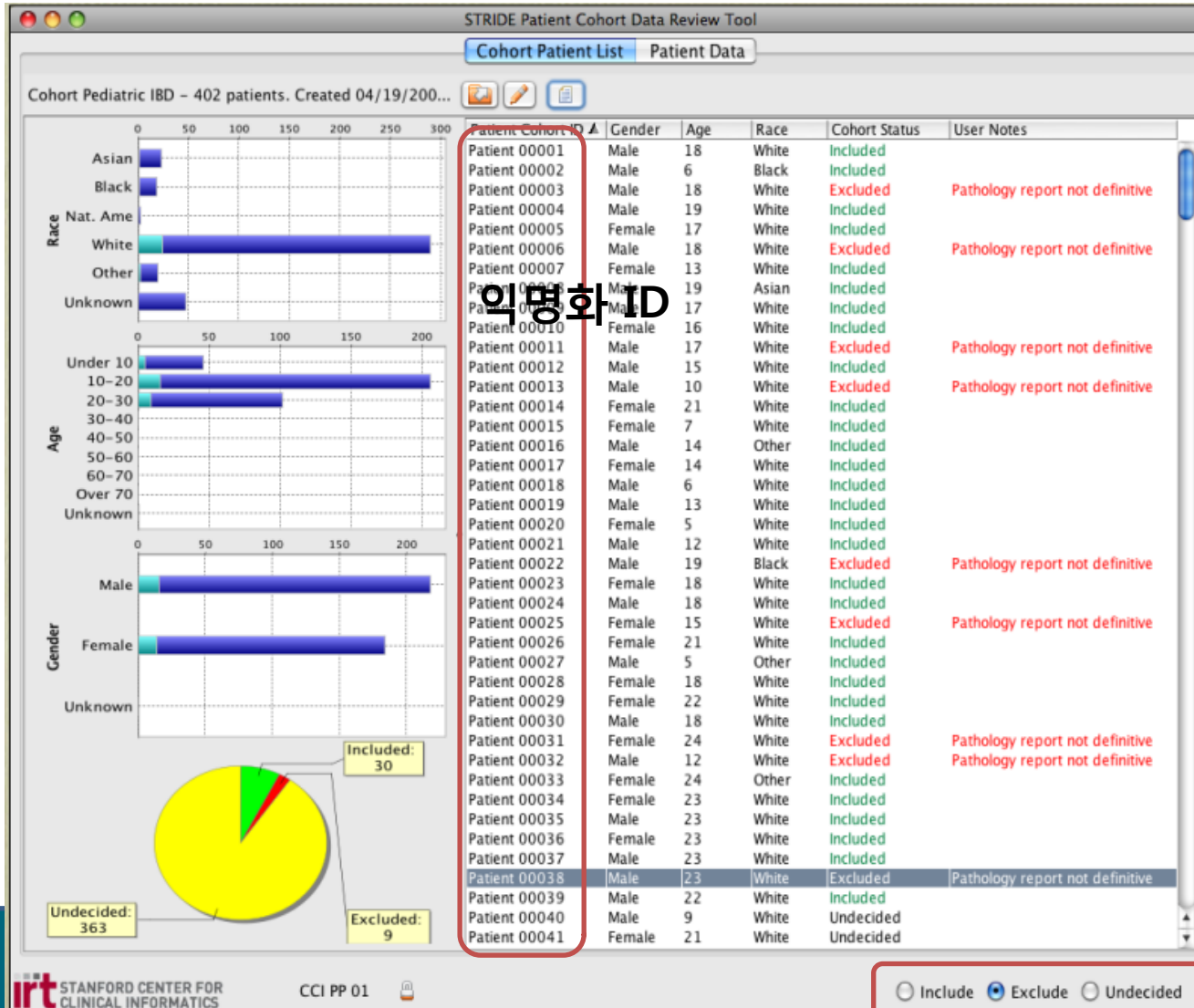


최종 숫자만 제공

Drag-drop I/F

# STRIDE Data Review Tool

개별 data review를 통한 구군 최종 선정 도구





# STRIDE Data Review Tool

STRIDE Patient Cohort Data Review Tool

Cohort Patient List Patient Data

Patient 00004 - [REDACTED]

Search

Data Element

- Visit (23)
  - Inpatient (12)
  - Outpatient (10)
  - Pre-admit (1)
- Diagnosis (28)
  - Abdominal or pelvic swelling mass or lump right lower qua...
  - Anemia unspecified (ICD9CM 285.9) (1)
  - Benign neoplasm of colon (ICD9CM 211.3) (1)
  - Diarrhea (ICD9CM 787.91) (1)
  - Iron deficiency anemia unspecified (ICD9CM 280.9) (1)
  - Loss of weight (ICD9CM 783.21) (1)
  - Other and unspecified noninfectious gastroenteritis and co...
  - Other symptoms concerning nutrition metabolism and dev...
  - Splenomegaly (ICD9CM 789.2) (1)
  - Ulcerative colitis unspecified (ICD9CM 556.9) (5)
  - Universal ulcerative (chronic) colitis (ICD9CM 556.6) (1)
- Procedure (10)
  - Closed [endoscopic] biopsy of large intestine (ICD9CM 45.2)
  - Closed [endoscopic] biopsy of rectum (ICD9CM 48.24) (2)
  - Esophagogastroduodenoscopy [EGD] with closed biopsy (K...
  - LVL IV-SURG PATH GROSS&MCRCSCP XM (CPT 88305) (2)
  - Need for prophylactic vaccination with tuberculosis (bcg) v...
  - OFFICE OUTPT EST 25 MIN (CPT 99214) (1)
  - SCOPE OF COLON WITH BIOPSY (CPT 45380) (1)
- Laboratory Test Result (40)
  - Pathology report (2)
  - Radiology report (4)
  - Transcription (14)
  - Pharmacy (7)

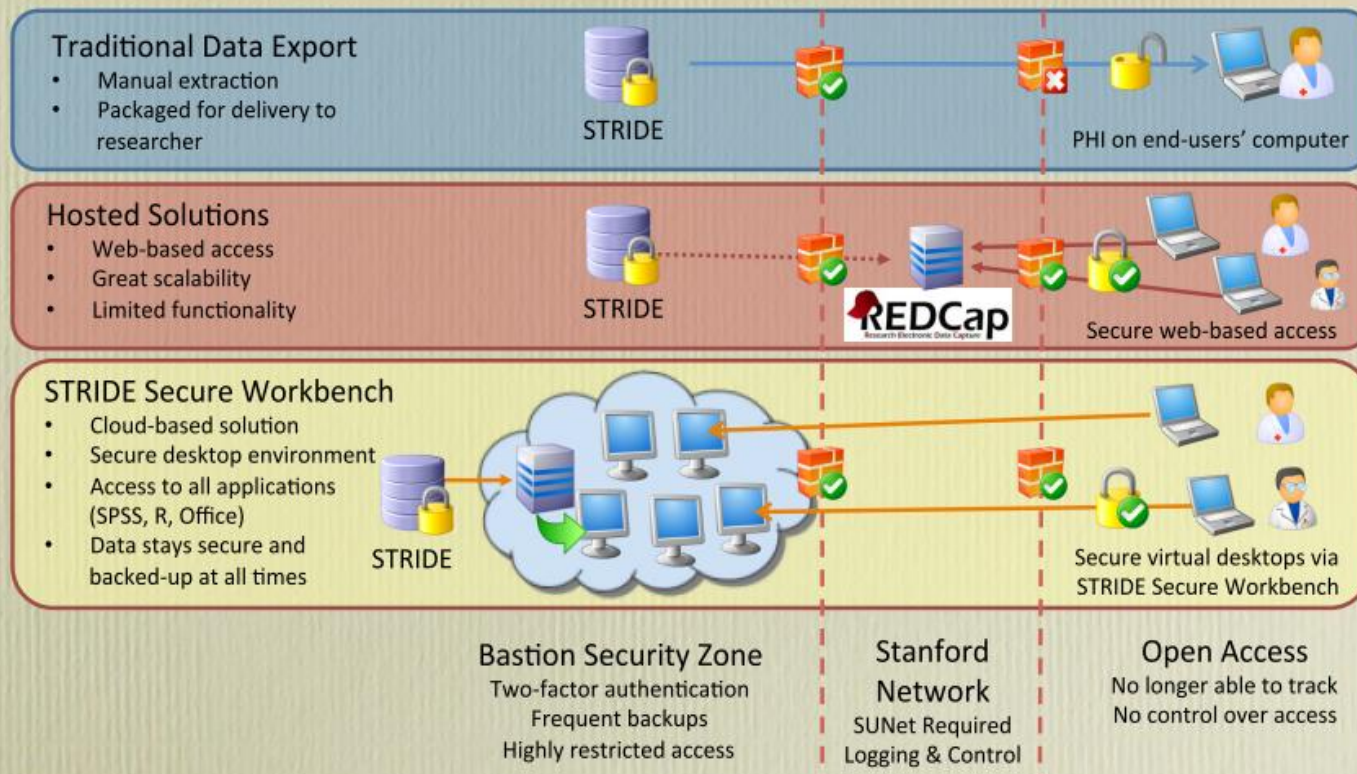
Type	Subtype	Date
Visit	Inpatient	
Transcription	10	
Laboratory Test Result	WBC (Specimen N/A) - WBC	
Laboratory Test Result	Protein, Total (Serum/Plasma) - TP	
Visit	Inpatient	
Diagnosis	Other and unspecified noninfectious gastroenteritis and colitis (ICD9C...	
Laboratory Test Result	Protein, Total (Serum/Plasma) - TP	
Laboratory Test Result	WBC (Specimen N/A) - WBC	
Transcription	13	
Diagnosis	Other and unspecified noninfectious gastroenteritis and colitis (ICD9C...	
Visit	Inpatient	
Diagnosis	Other and unspecified noninfectious gastroenteritis and colitis (ICD9C...	
Visit	Inpatient	
Transcription	13	
Laboratory Test Result	Protein, Total (Serum/Plasma) - TP	
Laboratory Test Result	WBC (Specimen N/A) - WBC	
Diagnosis	Ulcerative colitis unspecified (ICD9CM 556.9)	
Visit	Outpatient	
Visit	Inpatient	
Transcription	13	
Diagnosis	Ulcerative colitis unspecified (ICD9CM 556.9)	
Laboratory Test Result	Protein, Total (Serum/Plasma) - TP	
Laboratory Test Result	WBC (Specimen N/A) - WBC	
Visit	Inpatient	
Diagnosis	Ulcerative colitis unspecified (ICD9CM 556.9)	
Diagnosis	Ulcerative colitis unspecified (ICD9CM 556.9)	
RBC (Specimen N/A) - RBC	4.2 M/UL	[4.5-5.9] Abnormally low
Hemoglobin (Specimen N/A) - HGB	12.6	[13.5-17.0] Abnormally low
Hematocrit (Specimen N/A) - HCT	36.1 %	[40-52] Abnormally low
Mean Corpuscular Volume (Specimen N/A) - MCV	86.9 FL	[81-97]
Mean Corpuscular Hemoglobin (Specimen N/A) - MCH	30.3 PG	[26.5-33]
Mean Corpuscular Hemoglobin Concentration (Specimen N/A) - MCHC	34.8 %	[32-36]
Red Cell Distribution Width (Specimen N/A) - RDW	12.8 %	[11.5-14.5]
Platelet Count (Specimen N/A) - PLT	179.0 K/UL	[150-450]
Neut ABS (Specimen N/A) - ANEUT	1.3 K/uL	[no ref rang]
Lymphocytes ABS (Specimen N/A) - ALYM	1.2 #	
Monocytes ABS (Specimen N/A) - AMON	.4 #	
Eosinophils, ABS (Specimen N/A) - AEOS	.3 #	
Basophil, ABS (Specimen N/A) - ABAS	.0 #	
Neut % (Specimen N/A) - NEUTP	40.1 %	
Lymphocytes % (Specimen N/A) - LYMP	38.1 %	

Include Exclude Undecided

환자 개별 review 화면

# STRIDE

## STRIDE Secure Workbench





sooyong.shin@amc.seoul.kr

 @likesky3



**ASAN**  
Medical Center