

# 의료명세서 빅데이터 분석

[Medical Claim Big Data]

2017. 12. 6.

# 의료 명세서 데이터.

(Medical Claim Data.)

# 의료 명세서 빅데이터 분석.

(MDwalks Data Services.)

# 의료 + 빅데이터

---

**Big Data.**

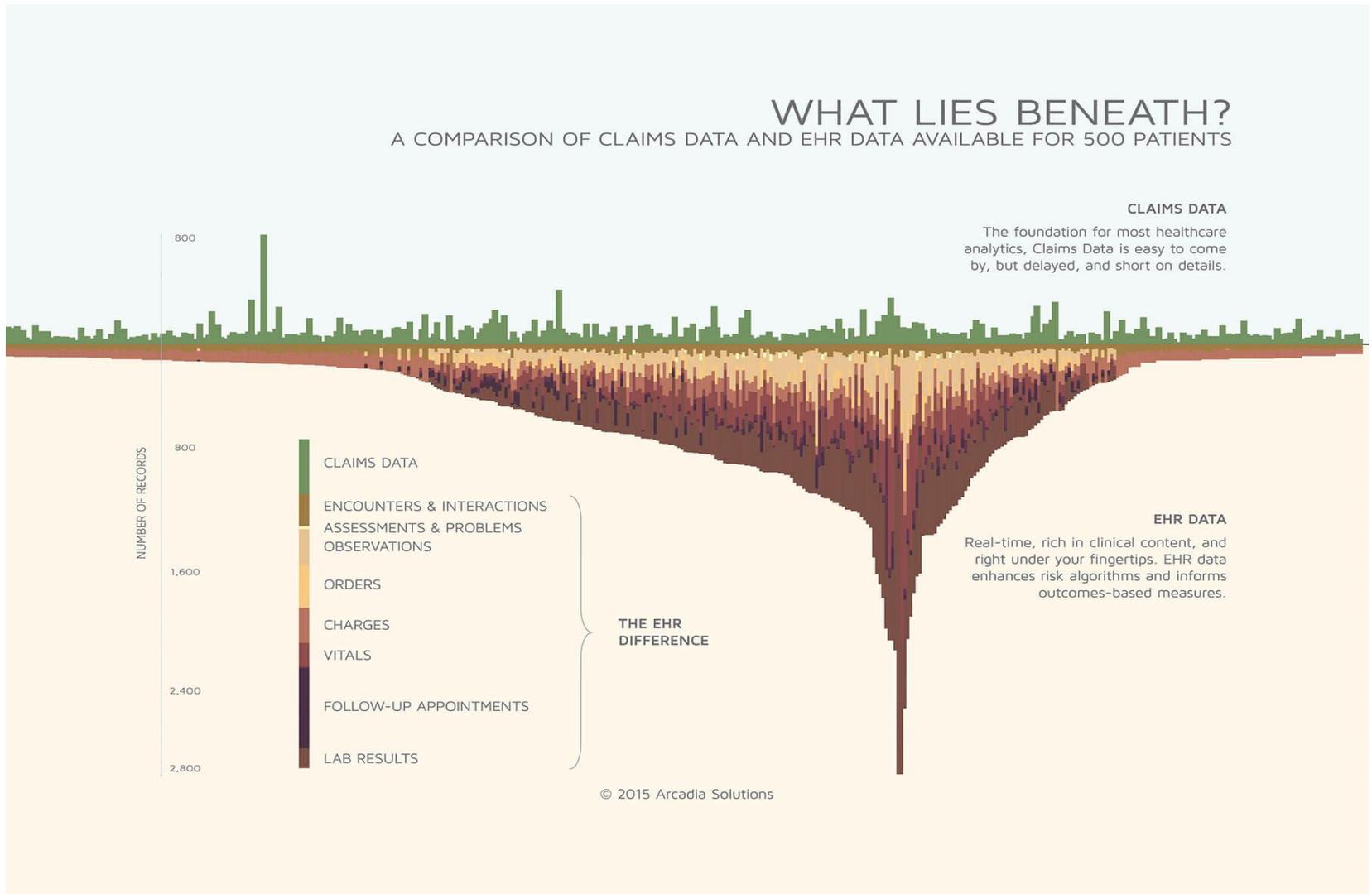
**Medical Data.**

**Precision Medicine.**

**Real World Evidence.**

'의료 데이터에서 패턴을 찾아내서 유용한 가치를 만든다!'

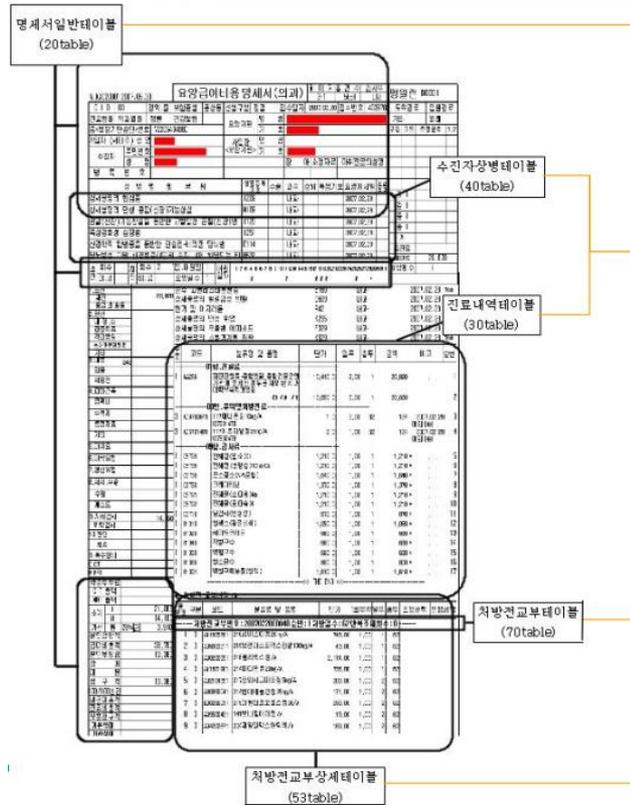
# Healthcare System Analytics



# 의료명세서 (Medical Statement)

우리나라는 전국민 건강보험을 시행하고 있으며 전국민 5천만 명의 의료보험청구명세서가 연간 14억 건씩 건강보험심사평가원으로 수집되고 있습니다.

- 상병명
- 환자 개인 신상 (이름, 주민번호, 보험자 등)
- 진찰료
- 입원료
- 투약 및 처방전료
- 주사료
- 마취료
- 이학요법료
- 정신요법료
- 이하 14가지 비용 및 진료 처리항목



- 상세 상병명
- 상병 분류기호
- 수술여부 항목
- 진료과목
- 상해요인
- 최초입원개시일
- 영양급여일수 등

- 의료행위코드
- 의약품 코드(일반, 제품명)
- 치료재료 코드
- 의료기기 코드
- 행위수가
- 1회 투약량
- 1일 투여횟수
- 총 투약일수

- 약품 구분 코드
- 수량
- 약품 코드
- 단가
- 1일 투여횟수
- 총 투약일수
- 처방일수 등

의료 명세서 데이터

# 의료명세서 데이터 구조

명세서일반정보 테이블 (200 table)

PAT_AGE	숫자	8수진자연령	2개월 이하:0.2, 3-12개월:0.9
SEX_TP_CD	문자	1성별구분코드	수진자 수급자의 성별구분 9:기타, 1:남, 2:여
agg	숫자	8연령군	5세단위 수진자 연령의 연령 군(5세단위별,16구간)
dif	문자	1입원, 외래 샘플구분	입원환자(외래진료포함), 외래환자(입원기록없는환자) 구분
DGRSLT_TP_CD	문자	1진료결과구분코드	요양급여비용명세서상 최종 진료일의 환자상태를 구분하여 기재 1: 계속 2: 이송 3: 회송 4: 사망 5: 기타 9: 퇴원
DGSBJT_CD	문자	2진료과목코드	요양기관에서 청구시 기재한 진료과목코드
DMD_TP_CD	문자	1청구구분코드	0:일반청구, 1:보완청구, 2:추가청구, 3:분리청구, 4:상한제 일반청구, 5:상한제 보완청구, 6: 상한제 추가청구, 7:상한제 분리청구(4-7 :2004.7.1부터사용)
FST_IPAT_DD	문자	8최초입원일자	요양기관에 그 상병진료를 위하여 그달에 최초 내원한 년/월/일.
INSUP_CD	문자	1보험자코드	4: 건강보험 5: 의료급여(본인일부부담금만 국비지원) 7: 보훈국비환자(진료비(약제비)전액을 전액 국비지원)
IPAT_ARIV_PTH_TP_CD	문자	2입원도착경로구분코드	도착경로(1:타요양기관경유, 2:응급구조대 후송, 3:기타) + 입원경로(1:응급실, 2:외래)
MSICK_CD	문자	6주상병코드	주된 상병분류기호(약국 증상분류기호, 한방상병코드)
RECU_DDCNT	숫자	8요양일수	수진자를 진료한 총일수(투약 일수 포함)
RECU_FR_DD	문자	8요양개시일자	수진자가 진료를 받기 시작한 일자, 조제투여일자, 최초내방일자
RVD_ADDC_RT	숫자	8심결가산율	진료개시일자 기준의 가산율, 공상가산율, 의료급여가산율
RVD_CT_TAMT	숫자	8심결CT총액	심사 결정된 CT 총요양급여비용
RVD_INSUP_BRAMT	숫자	8심결보험자부담금	심사결과 보험자가 부담해야 될 금액
RVD_MRI_TAMT	숫자	8심결MRI총액	심사 결정된 MRI 총요양급여비용
RVD_RPE_TAMT	숫자	8심결요양급여비용총액	심사결과 수진자 본인이 부담해야 될 금액과 보험자가 부담해야 될 금액을 합한 요양급여비용총액
RVD_SLF_BRAMT	숫자	8심결본인부담금	심사결과 수진자 본인이 부담해야 될 금액
SOPR_YN	문자	1수술여부	주된 상병명과 관련된 수술을 시행한 경우 0:비수술, 9:수술
SPC_SYM_TP_CD	문자	4특정기호구분코드	특정기호구분코드 ※표5. 특정기호구분코드 SHEET 참조
SSICK_CD	문자	6부상병코드	명세서부상병코드
VST_DDCNT	숫자	8내원일수	수진자가 진료를 받기위해 요양기관에 내원한 일수(초진+재진). 입원일수
MAIDCL_CD	문자	1의료급여종별코드	의료보호 대상자의 의료보호종별 구분 코드(1종, ~, 8종.) 1: 1종 2: 2종 4: 행려환자 5: 성병 6: 2종장애인 2차진료 8: 2종장애인 1차진료
RECU_TO_DD	문자	8요양만료일자	수진자가 진료 받기를 종료한 일자
RADT_DIAG_YN	문자	1방사선진단여부	방사선진단여부
RADT_TRET_YN	문자	1방사선치료여부	방사선치료여부
NO	숫자	8수진자고유번호	요양기관에서 진료받은 자의 주민등록번호 ※개인정보보호- 주민번호 사용하지 않고 임의의 연번 사용
YNO	숫자	8요양기관 고유번호	지역별, 종별 건강보험 지정을 인정한 요양기관 지정번호, 총 8자리로 구성 1~2자리:지역코드, 3자리:요양기관 종 별, 4자리:병원,보건기관 한방기관구분, 5~8자리:일련번호 ※개인정보보호- 요양기관기호 사용하지 않고 임의의 연번 사용
key	숫자	8명세서 연결코드	명세서 레벨의 팩트간 조인 Key

# 의료명세서 데이터 구조

## 진료내역 테이블 (300 table)

key	숫자	8 명세서 연결코드	명세서 레벨의 팩트간 조인 Key
CZ_CD	문자	2 항코드	요양급여비용 명세서 항코드 01:진찰료, 02:입원료, 03:투약료, 04:주사료, 05:마취료, 06:이학요법료, 07:정신요법료, 08:처치 및 수술료, 09:검사료, 10:영상진단 및 방사선치료료, L:요양병원 정액, S:특수장비, T:특수재료, V:보훈 등 100분의 100 본인부담 W:비급여, X:정신과정액
ITEM_CD	문자	2 목코드	항코드 소분류 단위로 부여된 번호 기재 ※표3. 항코드·목코드 SHEET 참조
DIV_TY_CD	문자	1 분류유형코드	명세서의 분류 코드 구분(1:수가, 2:준용수가, 3:약가, 4:수입,원료,조제,제제, 7:협약, 8:치료재료) ※(한방 - A: 수가, B: 준용수가, C: 약가, H:치료재료)
DIV_CD	문자	9 분류코드	구분코드(DIV_CD)
I_II_TP_CD	문자	11_2구분코드	진료코드의 요양기관 종별 가산율 적용여부를 구분하기 위한 코드 I : 기본진료료, 약제, 치료재료, 위탁검사 및 혈액료 등 상대가치점수표상의 요양기관 종별가산율이 적용되지 아니한 항목 II : 요양기관 종별가산율이 적용되는 진료행위 항목
UNPRC	숫자	8 단가	진료내역 분류코드별 단가
DD_MDQTY_OR_EXEC_FQ	숫자	8 일일투여량또는실시횟수	1일 총 투약량(소수 셋째자리에서 4사5입하여 소수 둘째자리까지)을 기재
TOT_MDC_DDCNT_OR_EXEC_FQ	숫자	8 총투여일수또는실시횟수	총투여 일수 또는 실시 횟수
AMT	숫자	8 금액	단가 * (일일투여량 또는 실시횟수) * (총투여일수 또는 실시횟수) = 금액
LN_NO	숫자	8 줄번호	진료내역 줄번호
GNL_NM_CD_8	문자	9 일반명코드 8자리	약가 주성분 등재약, 수입약의 일반명 코드

# 의료명세서 데이터 구조

## 상병내역 테이블 (400 table)

key	숫자	8 명세서 연결코드	명세서 레벨의 팩트간 조인 Key
SNO	숫자	8 일련번호	일련번호
SICK_DGSBJT_CD	문자	2 상병진료과목코드	실제 진료를 받은 진료과목 또는 상병명에 해당되는 진료과목별 코드
DMD_SICK_SYM	문자	7 청구상병기호	통계청 고시 제1993-3호('93.11.20)에 의거 한국표준질병사인분류기호
DMD_DGSBJT_CD	문자	2 청구진료과목코드	진료를 받은 진료과목(병원급이상) 또는 상병명에 해당하는 진료과목(의원급) 을 기재하되, 진료과목이 2개 이상에 해당되는 경우 상병별로 모두 기재

# 의료명세서 데이터 구조

## 처방전상세내역 테이블 (530 table)

key	숫자	8명세서 연결코드	명세서 레벨의 팩트간 조인 Key
LN_NO	숫자	8줄번호	진료내역 줄번호
DIV_TY_CD	문자	1분류유형코드	명세서의 분류 코드 구분(1:수가, 2:준용수가, 3:약가, 4:수입, 원료, 조제, 제제, 7:협약, 8:치료재료) ※(한방 - A: 수가, B: 준용수가, C: 약가, H:치료재료)
UNPRC	숫자	8단가	진료내역 분류코드별 단가
FQ1_MDQTY	숫자	81회투약량	1회 환자에게 투약한량
DD1_MDQTY	숫자	81일투약량	1일 환자에게 투약한량
TOT_MDC_DDCNT_OR_EXEC_FQ	숫자	8총투여일수또는실시횟수	총투여 일수 또는 실시 횟수
AMT	숫자	8금액	단가 * (일일투여량 또는 실시횟수) * (총투여일수 또는 실시횟수) = 금액
GNL_NM_CD_8	문자	9일반명코드 8자리	약가 주성분 등재약, 수입약의 일반명 코드

# 의료명세서 데이터 화면

의료명세서에는 질병-의료행위-의약품-치료재료 등이 세세히 기록(Log)되어 있습니다.

key	fom_cd	no	pat_age	sex_tp_cd	msick_cd	ssick_cd	recu_fr_dd	recu_to_dd	dgsbjt_cd	gnl_nm_cd	div_cd
2453559	3	881294	67	1	C320	F058	20161210	20161210	13	412702ATB	AP502
2453560	3	908574	5	1	M519	F067	20161210	20161210	5	115302CLQ	J4949321
2453561	3	1210769	9	1	J209	M2556	20161210	20161222	13	100901ATB	B3030221
2453562	2	318406	49	1	I109	I219	20161210	20161210	11	111501ATB	B0394911
2453563	2	21593	58	2	K590	J029	20161210	20161231	1	100301ACH	H4040001
2453564	2	1284231	66	2	M751	M6521	20161210	20161215	15	111601BIJ	
2453565	3	554867	75	2	G479	I109	20161210	20161210	13	100901ATB	K3010102
2453566	2	225073	41	1	H109	H108	20161210	20161210	6	100901ATB	AA303911
2453567	2	20403	78	2	E10		20161210	20161210	1	616401ATB	C3710001
2453568	3	97409	22	2	I6309	F78	20161210	20161210	12	107601ATB	
2453569	4	696994	61	1	J029	J209	20161210	20161210	13	345232ACH	
2453570	2	1303365	65	1	I109	I129	20161210	20161210	2	111501ATB	E6541
2453571	13	1258	12	1	J209	J029	20161210	20161213	11	100901ATB	
2453572	3	1290485	53	2	J209	J300	20161210	20161210	1	100901ATB	
2453573	3	19224	18	2	J209	J310	20161210	20161211	1	100901ATB	
2453574	3	884592	2	2	F009	G20	20161211	20161211	11	665321ATC	K2223111
2453575	3	108583	43	2	J219	J302	20161211	20161217	3	114523ATB	L0928111
2453576	3	34526	81	2	K296	K30	20161211	20161212	1	201140ATB	K3411245

[의료명세서 데이터 화면]

건강보험심사평가원 표본 전체환자데이터셋(HIRA-NPS)의 데이터스키마 기준으로 임의생성한 화면입니다.

# 의료명세서 데이터 이용 현황

의료명세서 데이터는 의료 시스템 정책연구, 의약품 사용, 질병 발생에 대한 원천 자료이며, 수 많은 의료연구 기관과 제약회사 그리고 공공기관이 현재 데이터 분석을 수행 중입니다.

## 빅데이터분석 과제목록

건강보험심사평가원에 빅데이터분석을 신청한 현황입니다.

번호	분석과제명	세부과제명	신청기관명	주관기관명	연구책임자	접수일자	공개여부
593	국립암센터	국립암센터 발전...	연세대학교 산학...			2017-08-29	N
592	감기 한의임상진료지침 개...	한방의료기관 내...	부산대학교 약학...			2017-08-28	N
591	요통 환자에 대한 수술적...	추간판탈출증 및...	연세대학교 의과...	연세대학교 산학...		2017-08-28	N
590	성인 건선 환자의 의료이...	성인 건선 환자...	성균관대학교 산...			2017-08-24	N
589	류마티스 관절염의 상지수...	대한민국 류마티...	한양대학교 병원...	한양대학교 병원...		2017-08-24	N
588	(공동연구) 단장증후군	(공동연구) 한...	순천향대학교병원...		이*	2017-08-21	N
587	방사선의학 정책개발 및 ...	방사선의학 임상...	한국원자력의학원		김*	2017-08-17	N
586	퇴행성 요추질환의 치료양...	한국의 최근 1...	순천향대학교 서...		박*	2017-08-17	N
585	한국인 비판막성 심방세동...	한국인 비판막성...	서울대학교병원	서울대학교병원	최*	2017-08-10	N
584	투석 환자의 이중 항혈소...	투석 환자에 시...	서울대학교 병원			2017-08-09	N

# 의료명세서 데이터 분석의 문제

의료명세서 데이터를 활용하고자 하는 모든 사람은 다음과 같은 문제가 직면합니다.

key	fom_cd	no	pat_age	sex_tp_cd	msick_cd	ssick_cd	recu_fr_dd	recu_to_dd	dgsbjt_cd	gnl_nm_cd	div_cd
2453559	3	881294	67	1	C320	F058	20161210	20161210	13	412702ATB	AP502
2453560	3	908574	5	1	M519	F067	20161210	20161210	5	115302CLQ	J4949321
2453561	3	1210769	9	1	J209	M2556	20161210	20161222	13	100901ATB	B3030221
2453562	2	318406	49	1	I109	I219	20161210	20161210	11	111501ATB	B0394911
2453563	2	21593	58	2	K590	J029	20161210	20161231	1	100301ACH	H4040001
2453564	2	1284231	66	2	M751	M6521	20161210	20161215	15	111601BIJ	
2453565	3	554867	75	2	G479	I109	20161210	20161210	13	100901ATB	K3010102
2453566	2	225073	41	1	H109	H108	20161210	20161210	6	100901ATB	AA303911
2453567	2	20403	78	2	E10		20161210	20161210	1	616401ATB	C3710001
2453568	3	97108	22	1	J029	J029	20161210	20161210	12	07601ATB	
2453569	4	696194	61	1	J029	J029	20161210	20161210	13	145232ACH	
2453570	2	1303365	65	1	I109	I129	20161210	20161210	2	111501ATB	E6541
2453571	13	1258	12	1	J209	J029	20161210	20161213	11	100901ATB	
2453572	3	1290485	53	2	J209	J300	20161210	20161210	1	100901ATB	
2453573	3	19224	18	2	J209	J310	20161210	20161211	1	100901ATB	
2453574	3	884592	2	2	F009	G20	20161211	20161211	11	665321ATC	K2223111
2453575	3	108583	43	2	J219	J302	20161211	20161217	3	114523ATB	L0928111
2453576	3	34526	81	2	K296	K30	20161211	20161212	1	201140ATB	K3411245

데이터가 너무 많고 무슨 내용을 담고 있는지 알 수가 없다

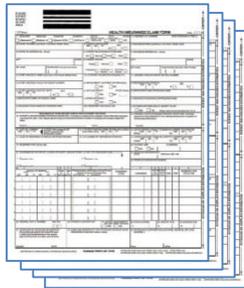
[의료명세서 데이터 화면]

# 의료명세서 데이터 레이크

데이터레이크(Datalake)는 대규모 데이터를 요소별로 분해해서 저장하며 활용 목적에 따라 필요한 데이터들을 추출하는 시스템입니다.

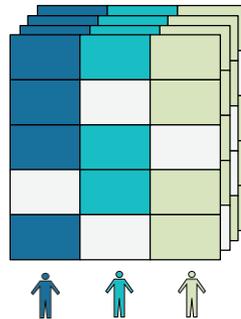
## Collect Medical Statements

- Demographic
- Payments
- Diseases
- Medicines
- Treatments
- Medical Devices
- Medical Materials
- Hospital Info

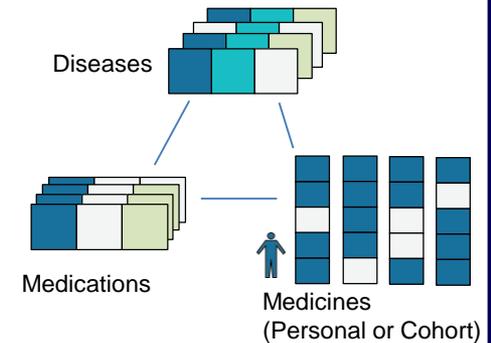


## Build Data lake system

Disease  
Medication  
Surgery  
Medicine  
Materials



## Retrieve Relational Data



# 의료명세서 데이터 레이크

데이터레이크를 사용하여 목적하는 데이터 마트(Data Mart)를 생성하고, 이를 탐색해 갈 수 있습니다.

## 명세서일반정보 테이블 (200 table)

```
2502332, 0, 3, 031, 1454413, 16, 0.0300001145, 33.333206114, 2, 80.0, 4, , E116, I10, 1, 01, 05, , 999000, , , 0, 1, 1, 20160127, 20160127,
3327735, 0, 3, 031, 365245, 9, 0.0300003033, 33.332996304, 2, 42.0, 4, , E116, , 1, 23, 23, , 999000, , , 0, 1, 1, 20160216, 20160216, 2660
23896924, 0, 3, 031, 1454413, 16, 0.0300001145, 33.333206114, 2, 80.0, 4, , E116, I10, 1, 01, 05, , 999000, , , 0, 1, 1, 20160226, 20160226,
```

## 진료내역 테이블 (300 table)

```
2502332, 2, 06011, 1, MM080, 2, 1.0, 1.0, 1.0, 1, 1.0, 3200, 3200, 3680,
2502332, 3, 09011, 1, C3710, 2, 1.0, 1.0, 1.0, 1, 1.0, 1000, 1000, 1150,
2502332, 1, 01021, 1, AA254010, 1, 1.0, 1.0, 1.0, 1, 1.0, 12550, 12550, 12550,
```

MM080 – 간섭파전류치료(ICT)

C3710 – 당검사

AA254010 – 재진진찰

## 상병내역 테이블 (400 table)

```
2502332, 3, 05, M511, 05
2502332, 2, 01, I10, 01
2502332, 1, 01, E116, 01
```

M511 – 신경병증을 동반한 요추 및 추간판 장애

I10 – 본태성(일차성)고혈압

E116 – 인슐린 비의존 당뇨병

## 처방전상세내역 테이블 (530 table)

```
2502332, 1, 2016012700003, 1.0, 2.0, 30, 60.0, 525, 31500, 513700AT
2502332, 4, 2016012700003, 1.0, 1.0, 7, 7.0, 414, 2898, 100903AT
2502332, 3, 2016012700003, 0.5, 1.0, 30, 15.0, 10, 150, 170801AT
2502332, 5, 2016012700003, 1.0, 1.0, 7, 7.0, 159, 1113, 222805AT
2502332, 2, 2016012700003, 1.0, 1.0, 30, 30.0, 367, 11010, 107601AT
```

513700AT – sitagliptin 50mg 당뇨병용제

100903AT – aceclofenac 0.2g 해열, 진통, 소염제

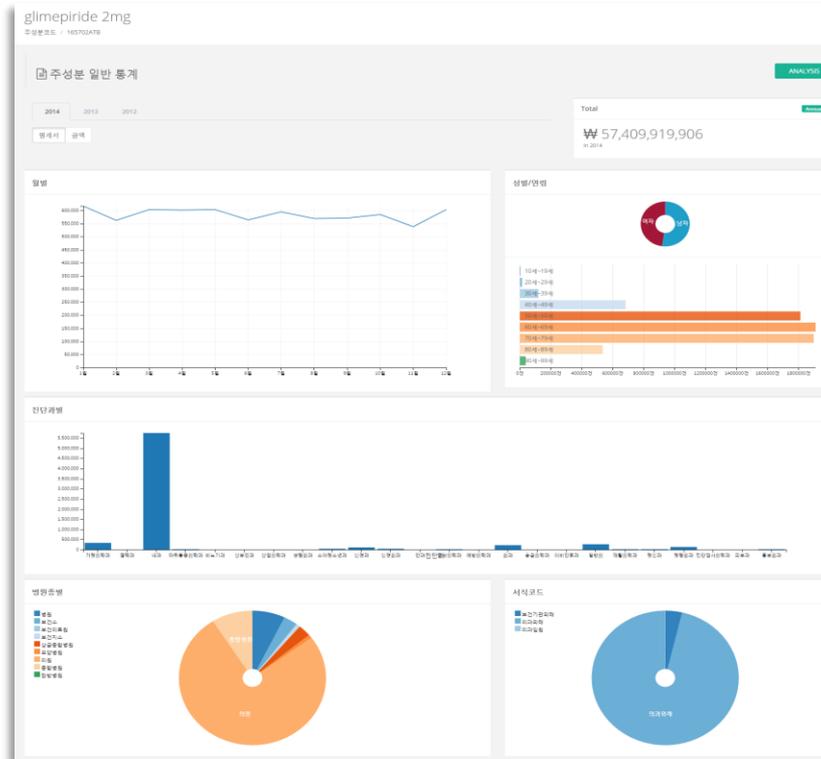
170801AT – hydrochlorothiazide 25mg 이뇨제

222805AT – ranitidine 75mg 소화성궤양용제

107601AT – amlodipine besylate 5mg 혈압강하제

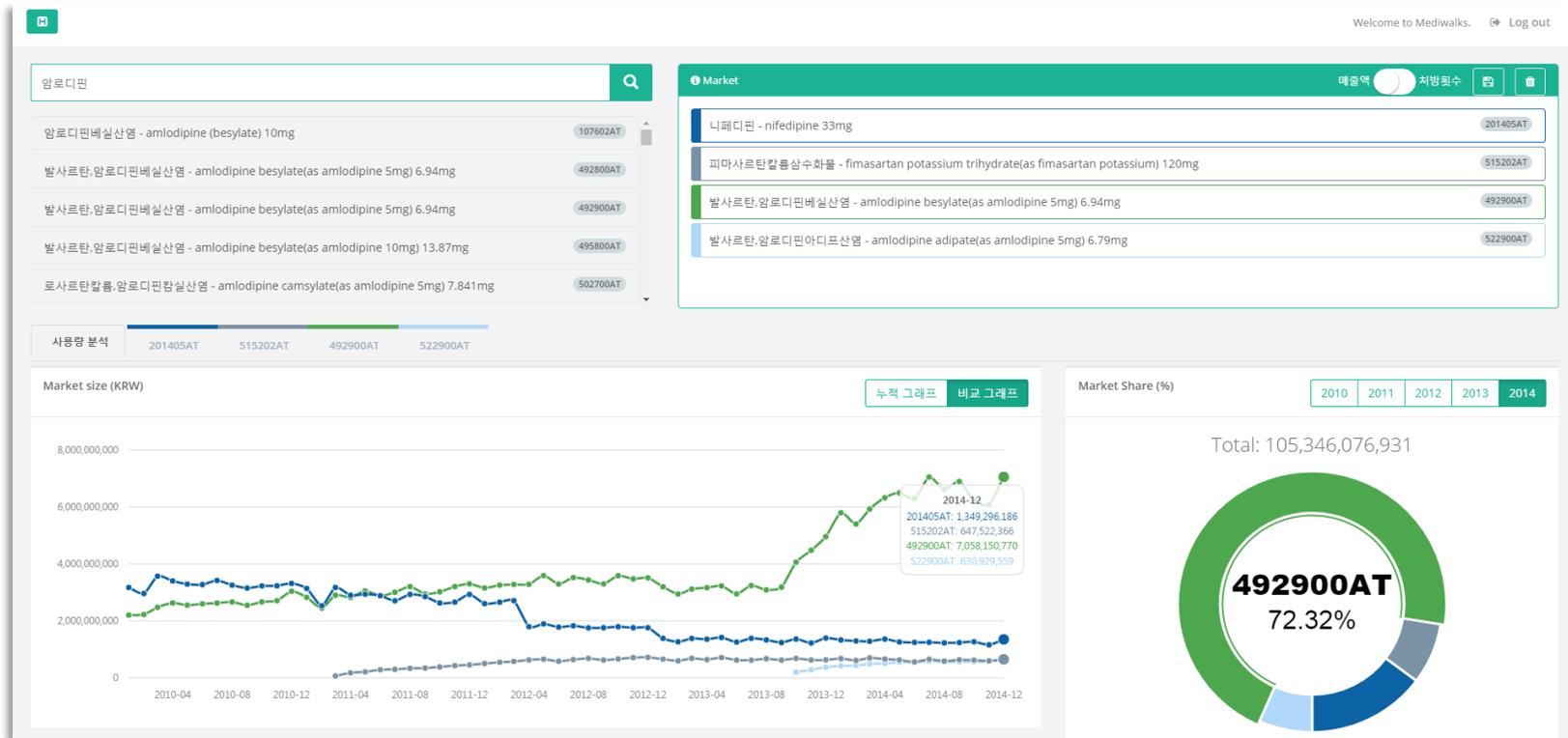
# 의약품 기준 데이터 탐색 (Pharmacy EDA)

의료명세서 모든 의약품을 기준으로 코호트 인포그래픽을 작성하였습니다.

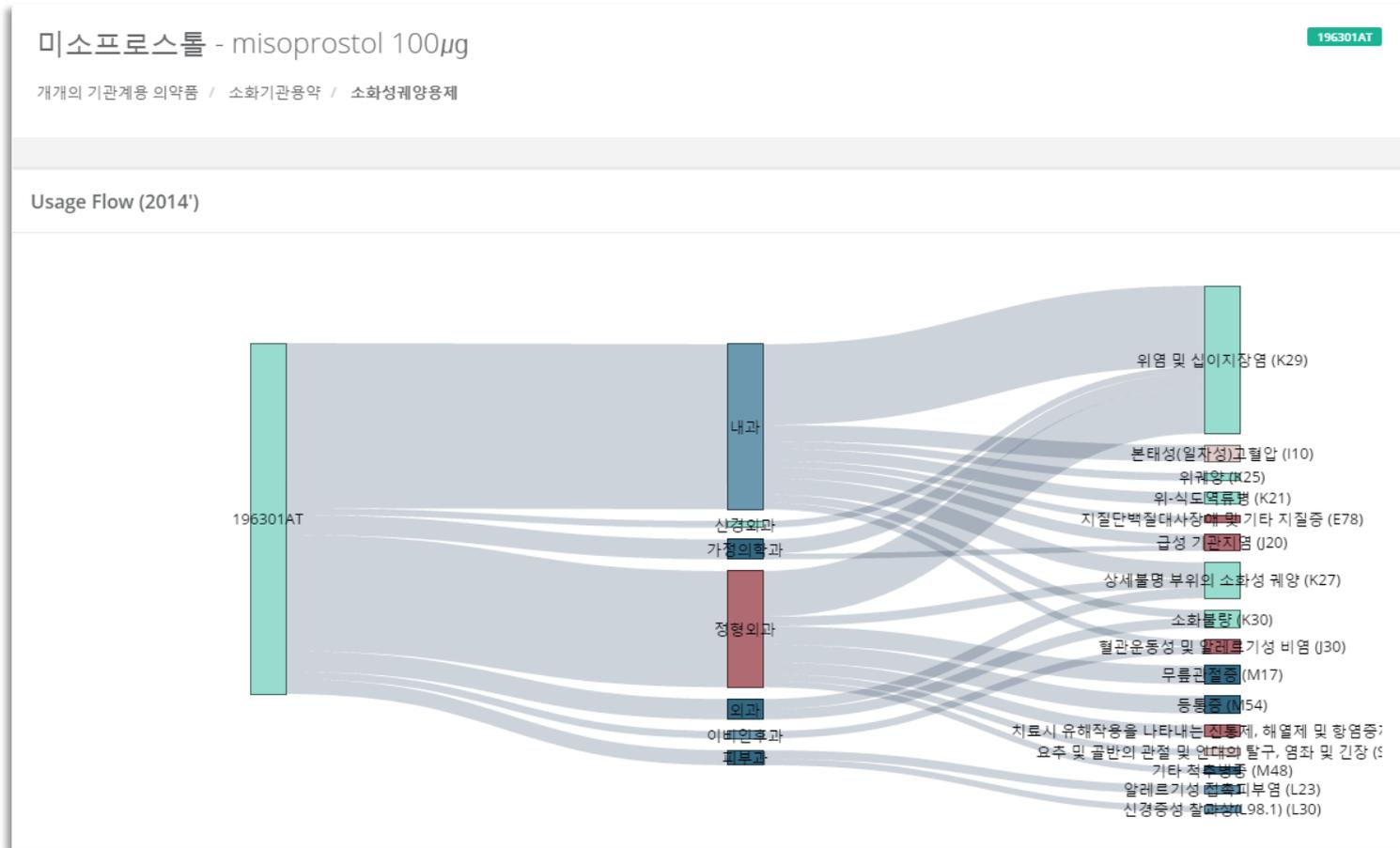


의약품 주성분별 데이터 그래프 → 1. 의료연구 : 의약품-질병 사용성 분석  
 2. 산업계 : 의약품 총사용량 · 시장 크기 분석  
 3. 산업계 : 복합제 · 신약 기반 데이터 분석

의약품의 사용 추세(trends)와 연결성 분석을 제공하는 서비스 입니다.



의약품 주성분 - 진단과 - 질병을 연결한 그래프와 데이터를 제공합니다.



[엠디웍스 트렌즈: 미소프로스톨 Usage Flow 분석화면]

의약품이 사용되는 **상병**과 동시에 빈번하게 같이 처방되는 **의약품** (Frequent-Pattern) 네트워크 분석을 제공합니다.



[엠디웍스 트렌즈: 미소프로스톨 Usage Flow 분석화면]

질병 기준 처방 의약품 집합을 추출합니다.

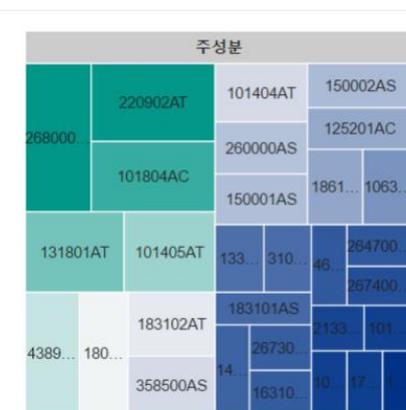
질병 주성분 빈도수 테이블

질병	주성분
1 E10 인슐린-의존 당뇨병	461801BI insulin glargine(as human insulin 300I.U) 10.9134mg
2 E10 인슐린-의존 당뇨병	191502AT metformin hydrochloride 0.5g
3 E10 인슐린-의존 당뇨병	170101BI human Insulin 1000I.U
4 E10 인슐린-의존 당뇨병	441303BI Insulin aspart(insulin aspart protamine/Insulin aspart 70%/30%) 300I.U
5 E10 인슐린-의존 당뇨병	441301BI Insulin aspart 300I.U 461801BI insulin glargine(as human insulin 300I.U) 10.9134mg
6 E10 인슐린-의존 당뇨병	175302BI Insulin lispro 300I.U 461801BI insulin glargine(as human insulin 300I.U) 10.9134mg
7 E10 인슐린-의존 당뇨병	461801BI Insulin glargine(as human insulin 300I.U) 10.9134mg 484901BI Insulin glulisine 10.47mg
8 E10 인슐린-의존 당뇨병	165702AT glimepiride 2mg 191502AT metformin hydrochloride 0.5g
9 E10 인슐린-의존 당뇨병 K29 위염 및 심이장염 M17 무릎관절증 M48 기타 척추병증 M54 등통증 M75 어깨 병변	184101AT limaprost α-cyclodextrin (as limaprost 5μg) 186101AT loxoprofen sodium hydrate (as loxoprofen sodium 60mg) 205203AT orphenadrine citrate 50mg 243203AT triamcinolone 4mg 438901AT streptokinase-streptodornase (streptokinase 10KI.U, streptodornase 2.5KI.U) 12.5KI.U
10 E10 인슐린-의존 당뇨병	175301BI Insulin lispro 1000I.U
11 E10 인슐린-의존 당뇨병 G63 달리 분류된 질환에서의 다발신경 병증 K29 위염 및 심이장염 M47 척추증	164202AC gabapentin 0.3g

동반 진단 질병 분석

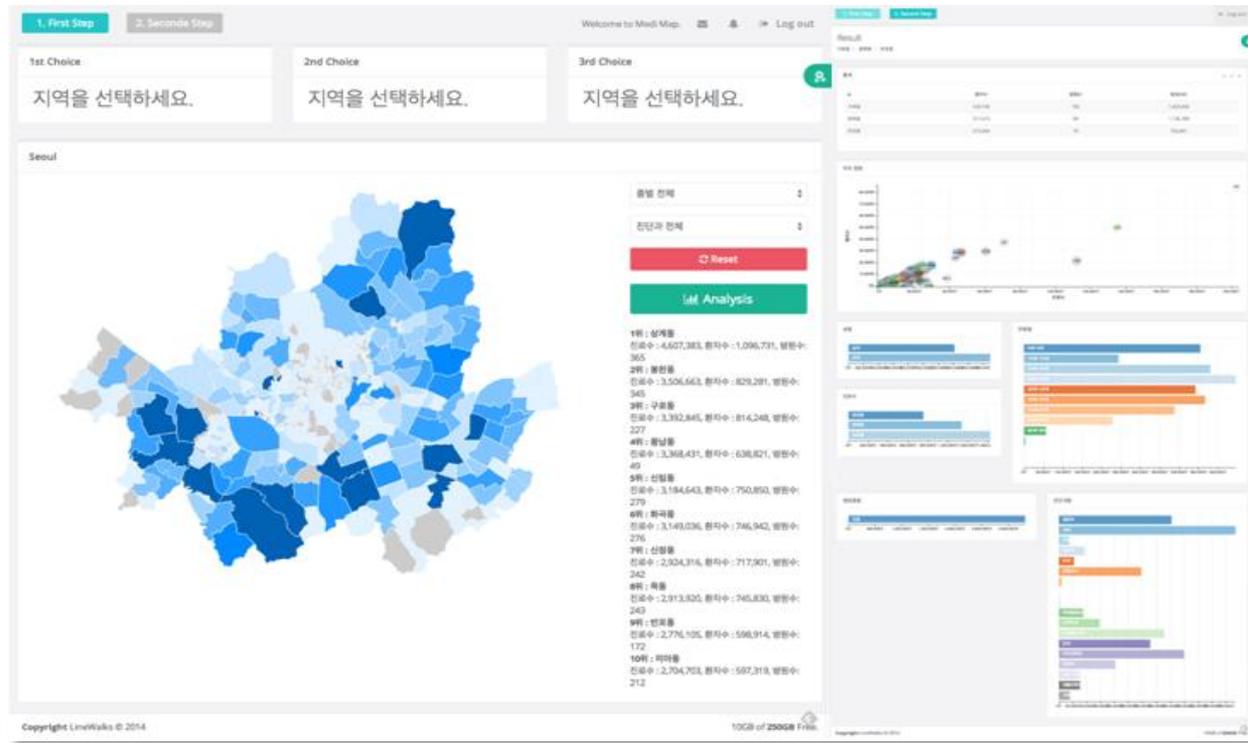


주성분 처방 비율



# 지역기준 의료기관 데이터 탐색 (Hospital EDA)

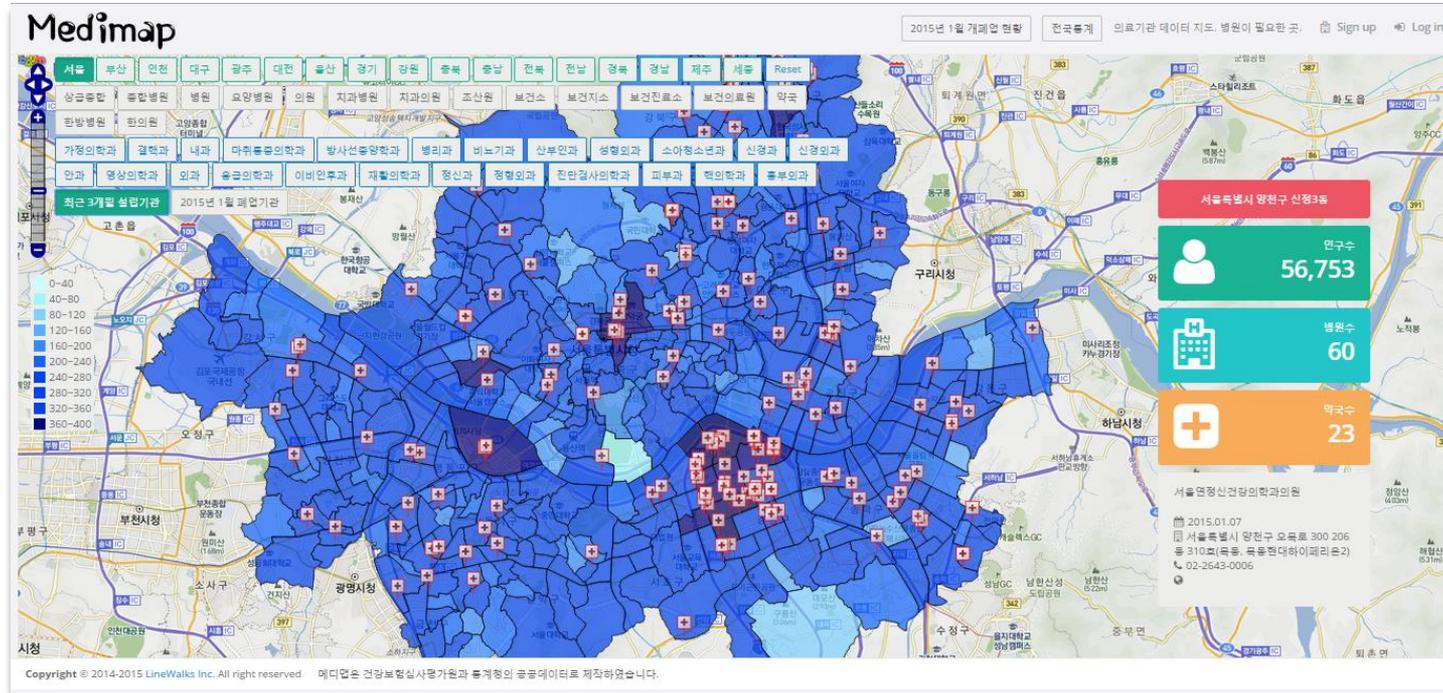
의료 명세서 데이터를 전국지역에 분산하여 시각화 분석을 하였습니다.



전국 행정동 단위 별 명세서 분포 시각화 → 1. 공공정책 : 의료 수요 분석 / 의료사각지대 분석  
2. 공공정책 : 의료 이용 행태 빅데이터 분석  
3. 산업계 : 개원/개국 지역 분석 서비스

메디맵은 인구수 대비 의료기관 밀집도 데이터를 제공합니다.

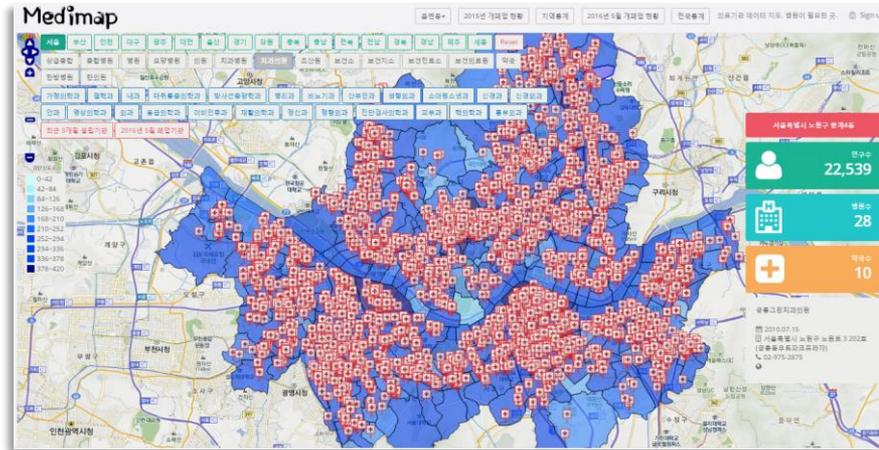
<http://medimap.net>



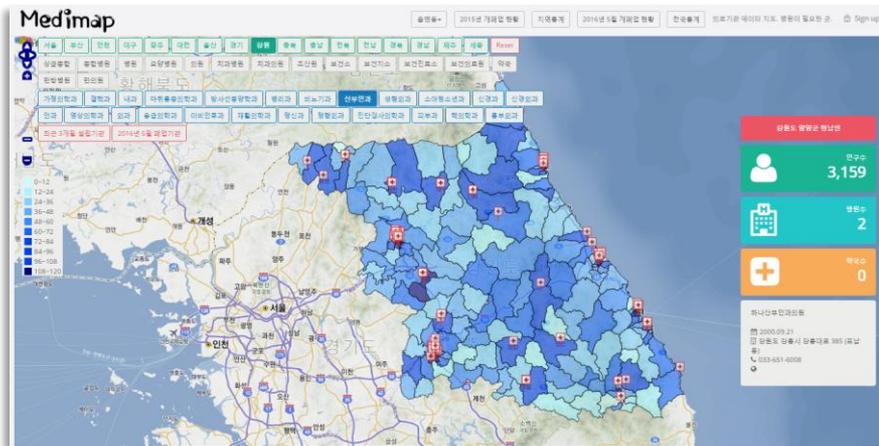
'인구수 데이터 수집' + '인구수 대비 병원 밀집도 추출' + '질환별 수요·공급 모델 개발'  
= 지역별 병원/약국 개폐업 분석

메디맵은 인구수 대비 의료기관 밀집도 데이터를 제공합니다.

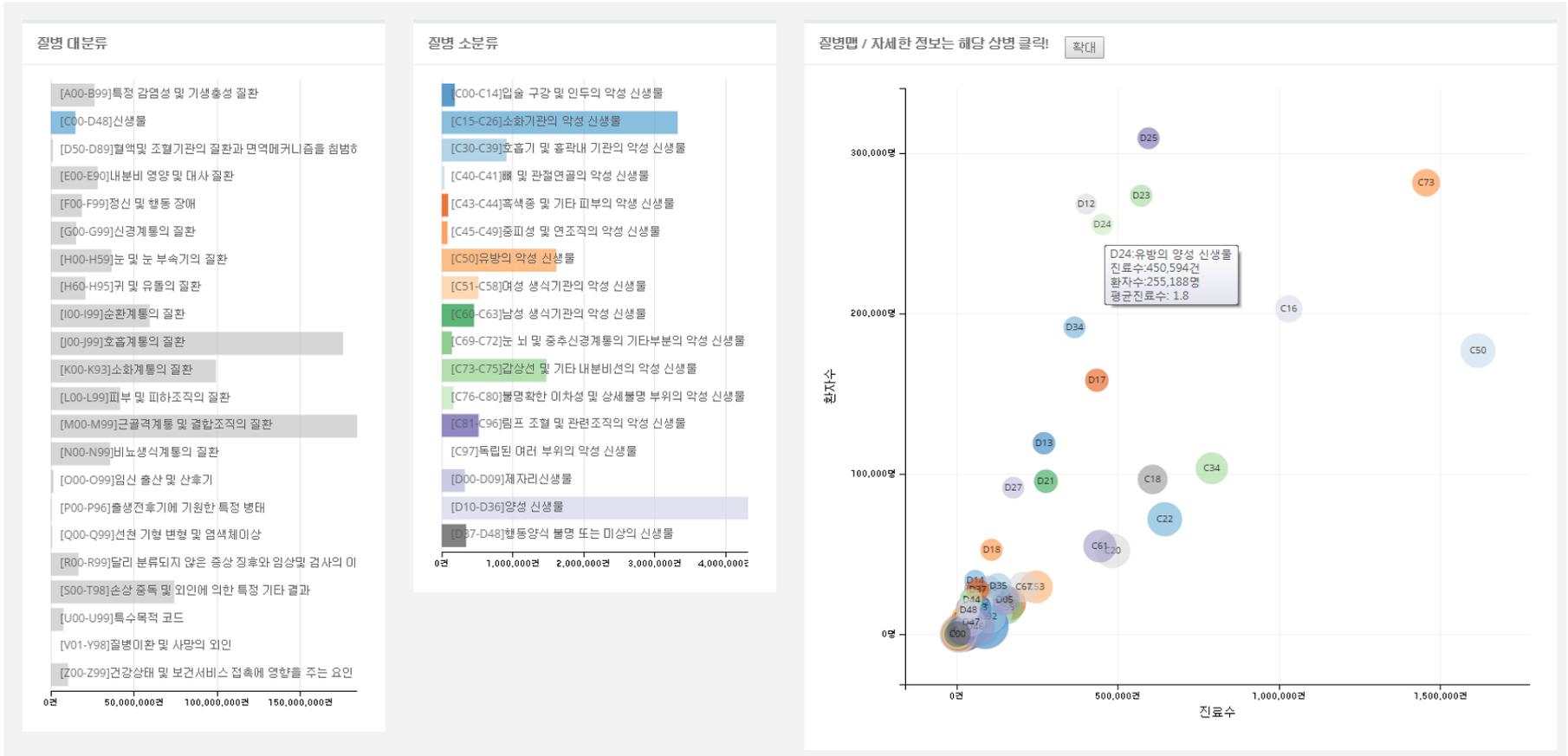
## 서울시 치과의원 밀집도



## 강원도 산부인과 밀집도



# 질병 통계 기준 데이터 탐색 (Disease - EDA)



전국 단위 질병 분포 시각화

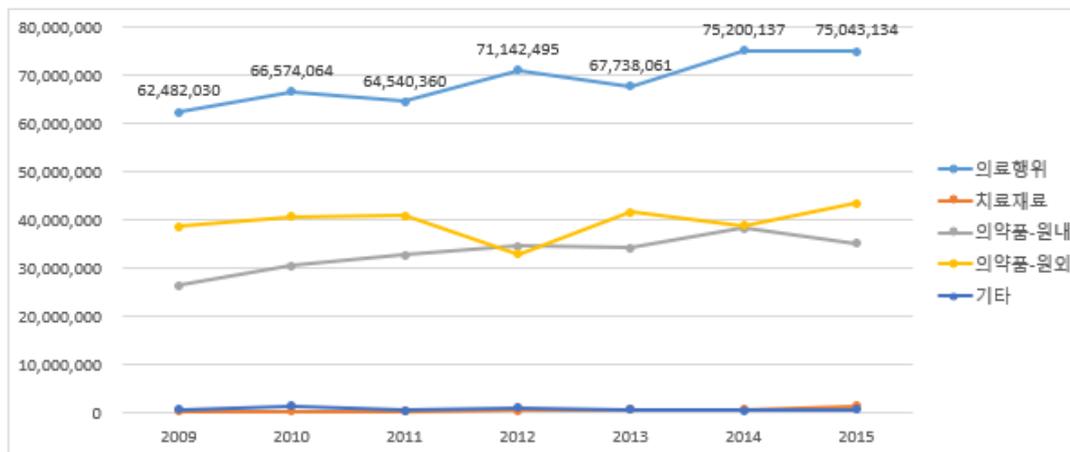


1. 공공정책 : 질병 통계 분석
2. 공공정책 : 질병 부담도(Burden of Disease) 분석
3. 병원 진료-경영 분석 서비스

질병부담도(Burden of Diseases) 데이터를 추출 분석하고 있습니다.

## Summary : 질병 의료 비용 분석 - C16 (위의 악성 신생물)

Table 1-1-1) 항목별 의료비용 - 외래



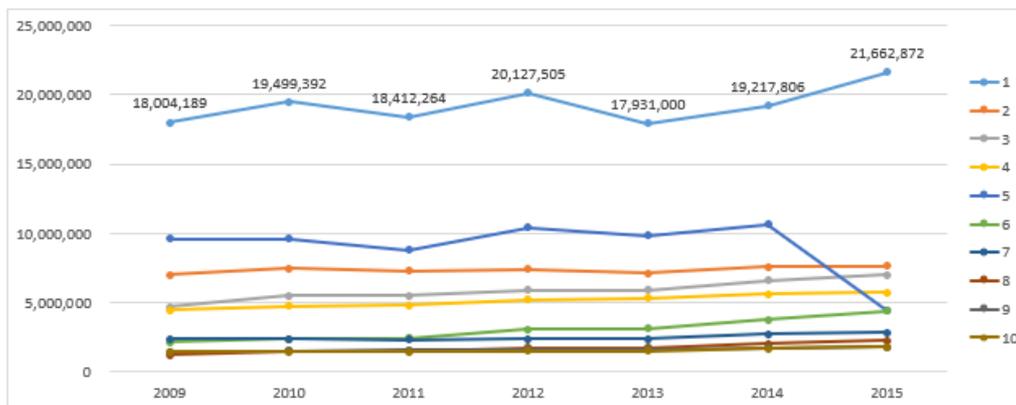
(단위: 천원)

구분	2009	2010	2011	2012	2013	2014	2015
의료행위	62,482,030	66,574,064	64,540,360	71,142,495	67,738,061	75,200,137	75,043,134
치료재료	259,513	395,968	346,293	534,733	621,683	674,269	1,356,088
의약품-원내	26,377,279	30,500,230	32,749,543	34,584,114	34,155,340	38,305,959	35,139,779
의약품-원외	38,697,794	40,667,577	40,993,679	32,889,957	41,606,982	38,783,072	43,525,772
기타	743,805	1,396,794	521,448	1,004,074	664,161	610,360	657,115
합계	128,560,420	139,534,633	139,151,323	140,155,373	144,786,226	153,573,797	155,721,888

질병부담도(Burden of Diseases) 데이터를 추출 분석하고 있습니다.

## C16 (위의 악성 신생물) 의료 비용 분석 - 의료행위 - 외래

Table 2-1-1) 처방금액을 기준으로 한 의료행위 - 외래



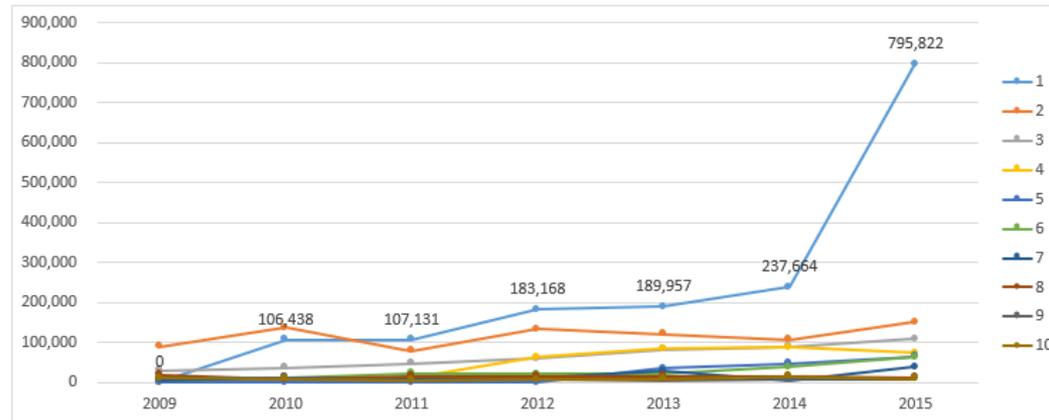
(단위: 천원)

의료행위 중분류	의료행위 소분류	2009	2010	2011	2012	2013	2014	2015
1 방사선 특수영상진단료	전산화단층영상진단	18,004,189	19,499,392	18,412,264	20,127,505	17,931,000	19,217,806	21,662,872
2 외래환자 진찰료	재진 진찰료	7,006,273	7,485,872	7,287,304	7,378,320	7,131,808	7,564,062	7,615,947
3 내시경, 천자 및 생검, 초음파 검사료	내시경	4,662,079	5,492,974	5,469,520	5,893,144	5,903,645	6,570,659	7,009,451
4 검체 검사료	종양표지자 검사	4,467,086	4,716,670	4,790,809	5,188,512	5,323,407	5,580,994	5,731,244
5 핵의학영상진단 및 골밀도검사료	핵의학영상진단 및 골밀도검사료	9,622,314	9,610,879	8,813,463	10,400,801	9,815,096	10,619,930	4,411,185
6 검체 검사료	지질, 영양 관련 검사	2,138,471	2,344,509	2,445,893	3,080,313	3,137,413	3,760,176	4,384,857
7 검체 검사료	일반화학검사	2,386,249	2,340,662	2,312,670	2,387,420	2,387,665	2,733,330	2,827,472
8 병리 검사료	조직병리검사	1,194,023	1,481,655	1,495,548	1,716,805	1,728,865	2,033,939	2,265,607
9 검체 검사료	혈액학 검사	1,450,508	1,498,797	1,551,753	1,633,100	1,606,038	1,728,645	1,788,346
10 검체 검사료	효소검사	1,444,776	1,442,043	1,421,249	1,471,019	1,470,997	1,670,504	1,771,663
기타	기타	10,106,062	10,660,612	10,539,886	11,865,556	11,302,127	13,720,091	15,574,489
총합계		62,482,030	66,574,064	64,540,360	71,142,495	67,738,061	75,200,137	75,043,134

질병부담도(Burden of Diseases) 데이터를 추출 분석하고 있습니다.

## C16 (위의 악성 신생물) 의료 비용 분석 - 치료재료 - 외래

Table 3-1-1) 처방금액을 기준으로 한 치료재료 - 외래



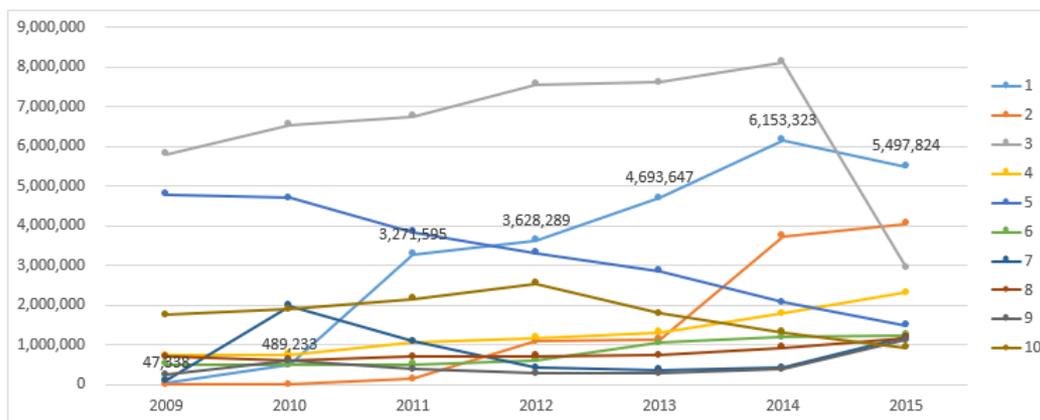
(단위: 천원)

치료재료 군	치료재료 대분류	2009	2010	2011	2012	2013	2014	2015
1	정액수가	0	106,438	107,131	183,168	189,957	237,664	795,822
2	중재적 시술용 군	89,188	137,213	79,199	133,449	120,764	107,625	150,976
3	일반재료군 II	29,350	36,983	46,467	58,999	82,663	89,070	109,012
4	중재적 시술용 군	580	1,016	12,555	62,541	84,493	88,080	73,605
5	중재적 시술용 군	0	0	0	0	34,320	47,275	65,235
6	일반재료군 III	4,455	11,242	23,289	20,529	21,829	38,582	62,896
7	일반재료군 II	4,387	11,832	12,348	12,497	28,351	5,746	38,725
8	중재적 시술용 군	19,560	6,861	15,367	15,231	16,030	13,831	11,772
9	일반재료군 II	9,935	12,247	3,964	6,171	5,265	6,701	9,151
10	중재적 시술용 군	10,002	8,927	5,758	6,405	8,288	15,083	8,134
	기타	92,055	63,207	40,216	35,743	29,723	24,614	30,759
	총합계	259,513	395,968	346,293	534,733	621,683	674,269	1,356,088

질병부담도(Burden of Diseases) 데이터를 추출 분석하고 있습니다.

## C16 (위의 악성 신생물) 의료 비용 분석 - 의약품(원내/원외처방) - 외래

Table 4-1-1) 처방금액을 기준으로 한 의약품(원내처방) - 외래

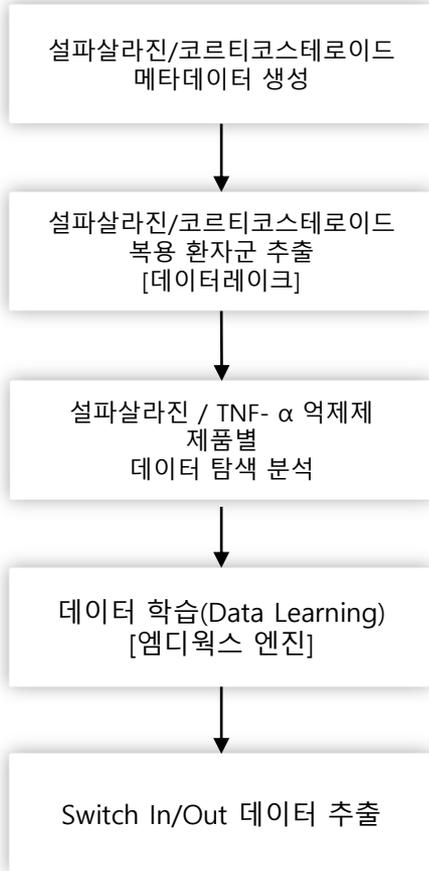


(단위: 천원)

의약품주성분 소분류	의약품주성분 명칭	2009	2010	2011	2012	2013	2014	2015
1 항악성종양제	oxaliplatin 100mg	47,338	489,233	3,271,595	3,628,289	4,693,647	6,153,323	5,497,824
2 항악성종양제	trastuzumab 0.15g	0	0	140,599	1,080,093	1,144,875	3,728,466	4,044,285
3 방사성 의약품	2-deoxy-2-fluoro-D-glucose (18F) inject	5,804,285	6,548,629	6,746,775	7,565,829	7,617,953	8,127,033	2,941,645
4 X선조영제	iohexol 755mg	732,928	760,262	1,071,241	1,174,187	1,309,319	1,794,128	2,312,437
5 X선조영제	iopromide 623mg	4,795,239	4,705,540	3,839,881	3,318,433	2,863,823	2,076,180	1,496,412
6 X선조영제	iopamidol 612mg	511,264	487,243	515,171	613,099	1,045,875	1,194,612	1,243,290
7 항악성종양제	imatinib mesylate (as imatinib 0.1g)	110,365	1,987,221	1,091,675	427,971	369,042	417,930	1,184,918
8 X선조영제	ioversol 678mg	692,948	614,511	698,788	716,634	731,522	929,375	1,179,284
9 항악성종양제	paclitaxel 30mg	251,563	601,738	388,025	271,336	284,828	385,007	1,122,718
10 X선조영제	iohexol 647mg	1,754,469	1,888,375	2,155,180	2,536,936	1,793,242	1,322,695	937,273
기타	기타	11,676,881	12,417,479	12,830,612	13,251,306	12,301,214	12,177,210	13,179,692
총합계		26,377,279	30,500,230	32,749,543	34,584,114	34,155,340	38,305,959	35,139,779

# 의약품 사용성 분석: 설파살라진-코르티코스테로이드 -> TNF- $\alpha$ 억제제

항염증약(설파살라진/코르티코스테로이드)에서 TNF- $\alpha$  억제제 (종양 괴사 인자 알파 억제제) 변화 데이터 추출



## 설파살라진/코르티코스테로이드 종류

주성분명	영문명	주성분 코드	ATC 코드	ATC cat3
설파살라진	sulfasalazine	232801ATE	A07EC01	Aminosalicylic acid and similar agents
메살라진	mesalazine	190801ATB, etc.	A07EC02	Aminosalicylic acid and similar agents
발살라지드 이나트륨 이수화물	balsalazide disodium dihydrate	475001ACH	A07EC04	Aminosalicylic acid and similar agents
부데소니드	budesonide	119406ACH, etc.	A07EA06	Corticosteroids acting locally
베클로메타손	beclometasone	114511ATE	A07EA07	Corticosteroids acting locally
덱사메타손	dexamethasone	141901ATB, etc.	H02AB02	Glucocorticoids
히드로코르티손	hydrocortisone	170901ATB, etc.	H02AB09	Glucocorticoids
메틸프레드니솔론	methylprednisolone	193301ATB, etc.	H02AB04	Glucocorticoids
프레드니솔론	prednisolone	217034ASY, etc.	H02AB06	Glucocorticoids
프레드니손	prednisone		H02AB07	Glucocorticoids

## TNF- $\alpha$ 억제제 종류

주성분명	영문명	주성분 코드	ATC 코드	ATC cat3	주성분 수(8자리 코드 기준)	의약품
에타너셉트	etanercept	455801BU	L04AB01	Tumor necrosis factor alpha (TNF- $\alpha$ ) inhibitors	5	엔브렐, 브렌시스
인플릭시맙	infliximab	383501BU	L04AB02	Tumor necrosis factor alpha (TNF- $\alpha$ ) inhibitors	1	레미케이드, 렘시마, 렘플렉시스
아달리무맙	adalimumab	488401BU	L04AB04	Tumor necrosis factor alpha (TNF- $\alpha$ ) inhibitors	2	휴미라
서톨리주맙페골	certolizumab pegol	625301BU	L04AB05	Tumor necrosis factor alpha (TNF- $\alpha$ ) inhibitors	2	퍼스티맙프리필드주
골리무맙	golimumab	621201BU	L04AB06	Tumor necrosis factor alpha (TNF- $\alpha$ ) inhibitors	6	심퍼니

- 오리지널: 휴미라, 엔브렐, 레미케이드 등
- 바이오시밀러: 셀트리온의 렘시마, 삼성바이오에피스의 브렌시스, 렘플렉시스

# 의약품 사용성 분석: 설파살라진-코르티코스테로이드 -> TNF- $\alpha$ 억제제

항염증약(설파살라진/코르티코스테로이드)에서 TNF- $\alpha$ 억제제 (종양 괴사 인자 알파 억제제) 변화 데이터 추출

명세서 데이터 / 환자 데이터 추출

크론병(K50), 궤양성 결장암(K51)

year	claim count	patient count	freq	percent (freq/claim count)
2009	5855	1070	1324	0.23
2010	5796	1113	1282	0.22
2011	7182	1280	1619	0.23
2012	9137	1466	2269	0.25
2013	9418	1530	2349	0.25
2014	10039	1598	2561	0.26
2015	9576	1664	2212	0.23

Switch Data Top 10.

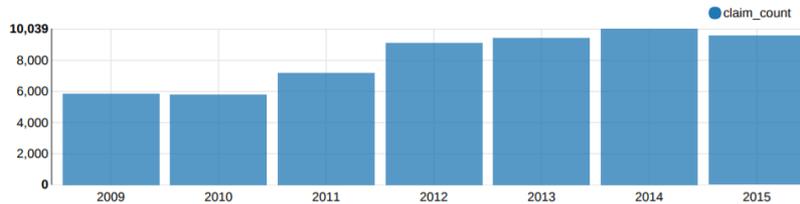
rank	switch(src.dst)	freq
0	(Glucocorticoids, Aminosaliclyic acid etc.)	4398
1	(Aminosaliclyic acid etc., Glucocorticoids)	4276
2	(Aminosaliclyic acid etc., TNF-a inhibitors)	1273
3	(TNF-a inhibitors, Aminosaliclyic acid etc.)	1263
4	(Glucocorticoids, TNF-a inhibitors)	582
5	(TNF-a inhibitors, Glucocorticoids)	545
6	(Aminosaliclyic acid etc., Corticosteroids)	484
7	(Corticosteroids, Aminosaliclyic acid etc.)	451
8	(Corticosteroids, Glucocorticoids)	150
9	(Glucocorticoids, Corticosteroids)	131
10	(TNF-a inhibitors, Corticosteroids)	34

# 의약품 사용성 분석: 설파살라진-코르티코스테로이드 -> TNF- $\alpha$ 억제제

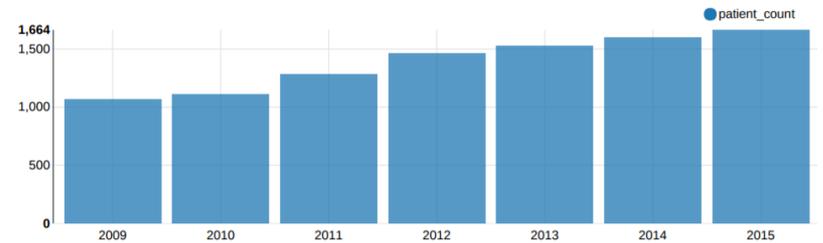
항염증약(설파살라진/코르티코스테로이드)에서 TNF- $\alpha$  억제제 (종양 괴사 인자 알파 억제제) 변화 데이터 추출

## 데이터 추출 및 인포그래픽 생성

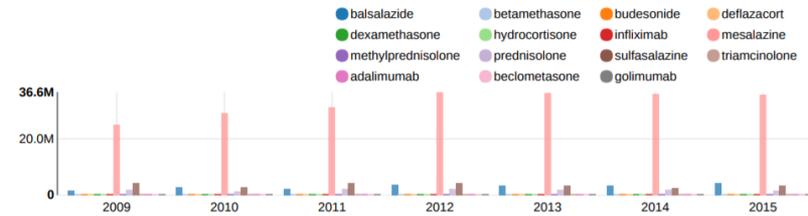
### 연도별 명세서수



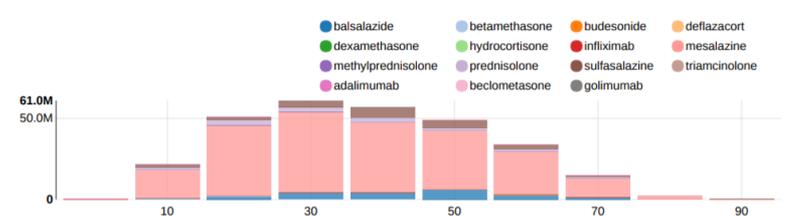
### 연도별 환자수



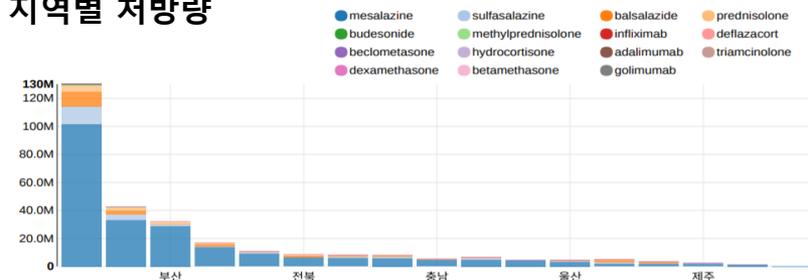
### 연도별 처방량



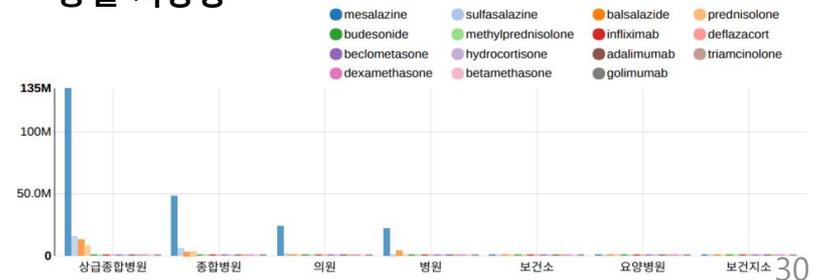
### 연령별 처방량



### 지역별 처방량



### 종별 처방량



# 의약품 사용성 분석: HMG-CoA 환원효소 억제제(reductase inhibitor) aka 스타틴

스타틴 의약품 내에서의 사용 변화 데이터 추출

메타데이터 생성

## HMG-CoA 환원효소 억제제(reductase inhibitor) 종류

코드	이름	주성분 수(8자리 코드 기준)	의약품 예
C10AA	HMG CoA reductase inhibitors	27	
C10AA01	simvastatin	5	심바스타틴정 등
C10AA02	lovastatin	1	로바스타틴정 등
C10AA03	pravastatin	4	메바로친정 등
C10AA04	fluvastatin	3	레스콜엑스엘서방정 등
C10AA05	atorvastatin	8	뉴아토르타정 등
C10AA07	rosuvastatin	3	인스토정 등
C10AA08	pitavastatin		피타스타정 등

# 의약품 사용성 분석: HMG-CoA 환원효소 억제제(reductase inhibitor) aka 스타틴

스타틴 의약품 내에서의 사용 변화 데이터 추출

명세서 데이터 / 환자 데이터 추출

Switch Data Top 10.

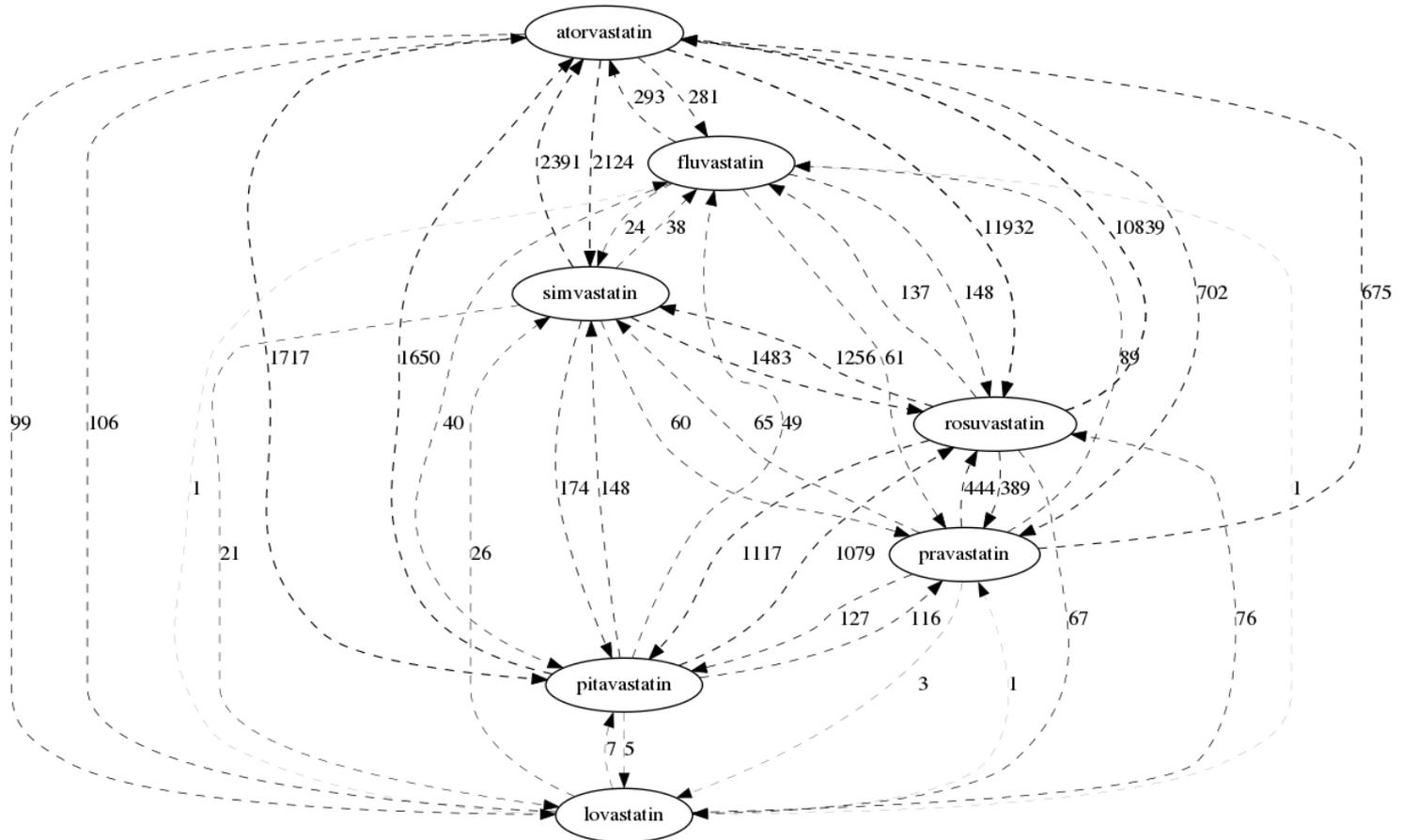
year	claim count	patient count	freq	percent (freq/claim count)
2009	587229	106530	41475	0.07
2010	674398	121679	41092	0.06
2011	753443	134415	41197	0.05
2012	866448	149100	45364	0.05
2013	949215	161838	46273	0.05
2014	1046091	179517	75946	0.07
2015	1134423	194876	40061	0.04

rank	switch	freq
0	(atorvastatin, rosuvastatin)	55900
1	(rosuvastatin, atorvastatin)	54700
2	(simvastatin, atorvastatin)	47146
3	(atorvastatin, simvastatin)	47116
4	(atorvastatin, pitavastatin)	15770
5	(pitavastatin, atorvastatin)	15747
6	(pravastatin, atorvastatin)	10090
7	(atorvastatin, pravastatin)	10085
8	(simvastatin, rosuvastatin)	9556
9	(rosuvastatin, simvastatin)	9361

# 의약품 사용성 분석: HMG-CoA 환원효소 억제제(reductase inhibitor) aka 스타틴

스타틴 의약품 내에서의 사용 변화 데이터 추출

2015년도 Statin Switch InOut 그래프

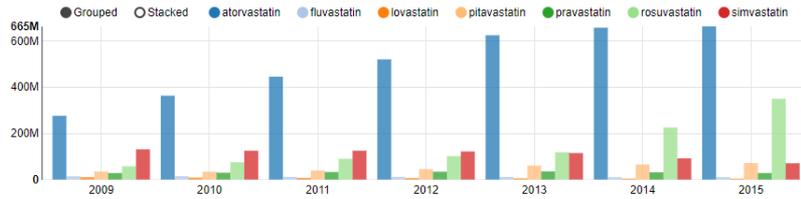


# 의약품 사용성 분석: HMG-CoA 환원효소 억제제(reductase inhibitor) aka 스타틴

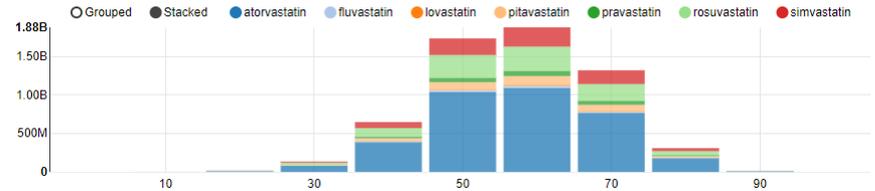
스타틴 의약품 내에서의 사용 변화 데이터 추출

데이터 추출 및 인포그래픽 생성

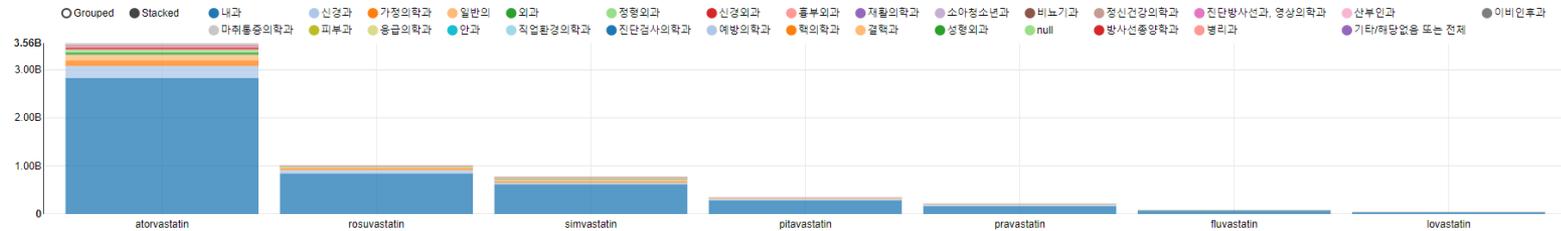
## 연도별 처방량



## 연령별 처방량

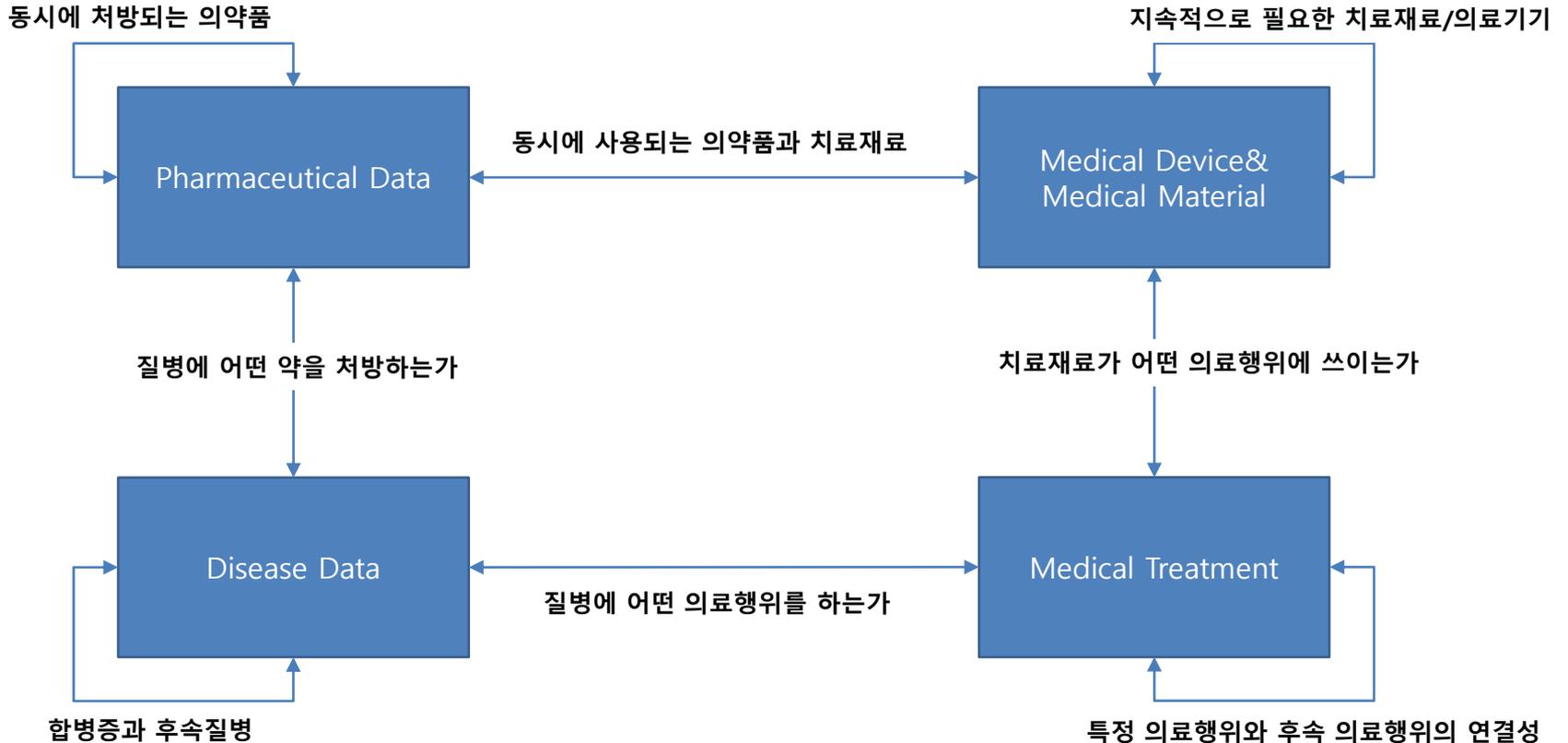


## 진단과별 처방량



# 의료명세서 내의 데이터 연결분석

의료명세서 데이터 탐색을 위한 시스템을 구축하면 목적에 따라 데이터를 연결하여 분석할 수 있습니다.



# 의료 + 빅데이터

---

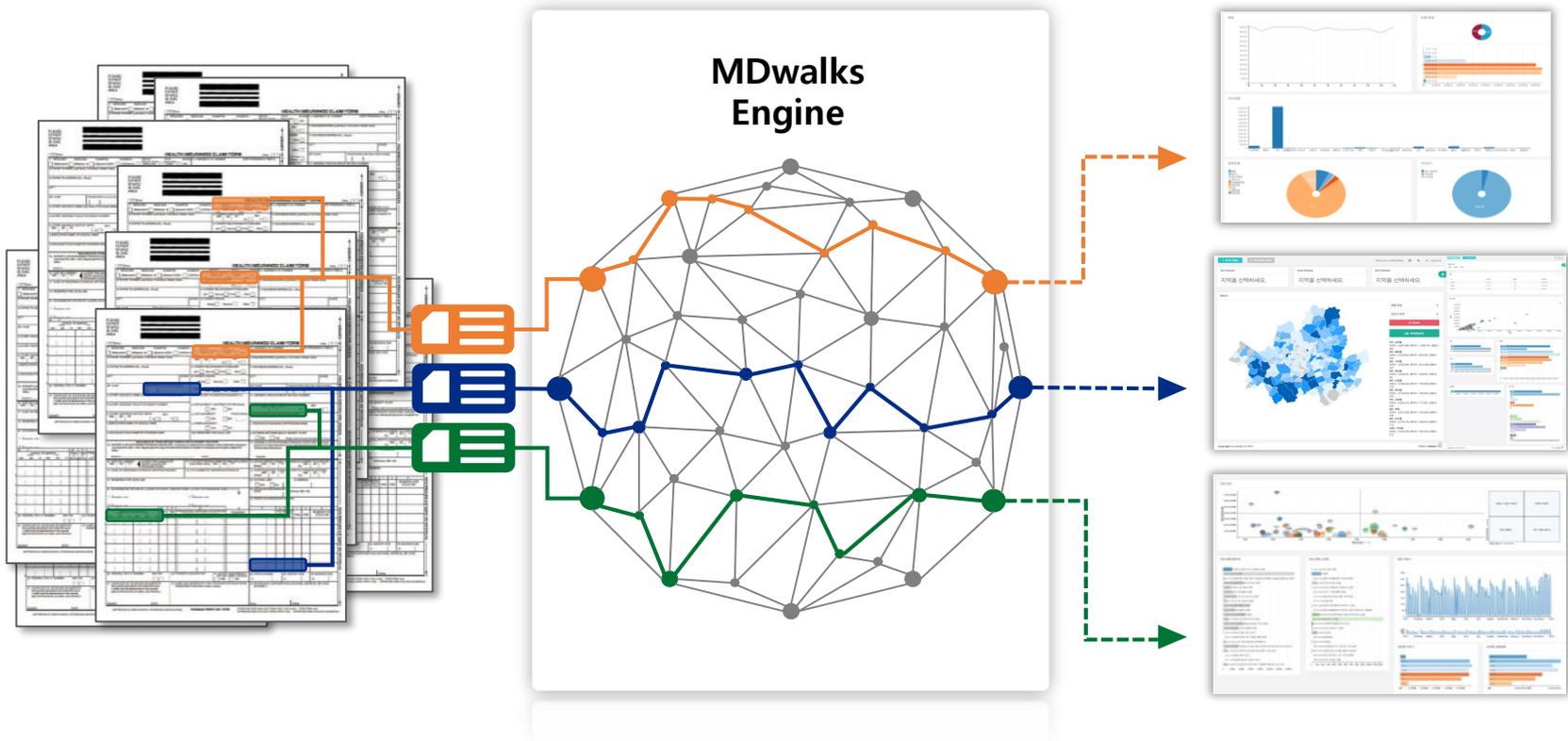
**Big Data.**  
**Medical Data.**  
**Precision Medicine.**  
**Real World Evidence.**

'의료 데이터 + 패턴을 찾아 + 유용한 가치'



# MDwalks Engine.

의료 전문 지식으로 필요한 데이터를 정의(Define)하고, 이를 데이터 엔진에 학습시켜서 지속적으로 데이터를 분석 추출하는 시스템으로 발전하고 있습니다.





# 마치며..

---

- 1.** 의료명세서 빅데이터는 매우 큰 가치가 있습니다.  
헬스케어 시스템의 환자수요와 의료공급 데이터를 분석할 수 있습니다.
- 2.** 의료명세서 데이터를 탐색 분석 (Exploratory Data Analysis)를 수행할 때 더 의미 있는 패턴을 찾아갈 수 있습니다.
- 3.** 빅데이터는 앞으로의 일을 예측하는 것이 아니라 현재 어떤 일이 일어나고 있는지 알아채고, 가설을 빠르게 검증하고 변화에 만들어 갈 수 있게 하는 지침입니다.

감사합니다.

---

linewalks

Big Data

Data Design

Contact : [yonghyun@linewalks.com](mailto:yonghyun@linewalks.com)  
[web@linewalks.com](mailto:web@linewalks.com)