

National Health Insurance Service

국민건강보험공단 빅데이터 활용 가이드

h·well

국민건강보험



BIG DATA

새로운 미래를 열다

목 차

National Health Insurance Service

01. 건강보험 빅데이터 소개	1p
02. 국민건강정보 DB 소개	2p
03. 건강보험 빅데이터 활용 분야.....	5p
04. 건강보험 빅데이터 활용 방법.....	6p
05. 의료이용지표 소개.....	9p
06. 의약품·의료기기·질병 통계 소개	10p
07. 공공데이터포털 소개.....	11p
08. 건강보험 빅데이터 관련 서비스 소개.....	12p

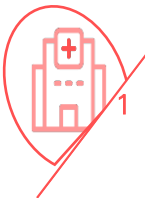


■ 건강보험 빅데이터란?

- 국민건강보험공단의 빅데이터는 4대 사회보험료(건강보험, 국민연금, 고용보험, 산재보험) 징수를 위한 관련 데이터 및 건강보험 급여 내역, 노인장기요양보험과 관련된 내역 등 보건 및 복지와 관련된 광범위한 자료로 구성되어 있습니다.



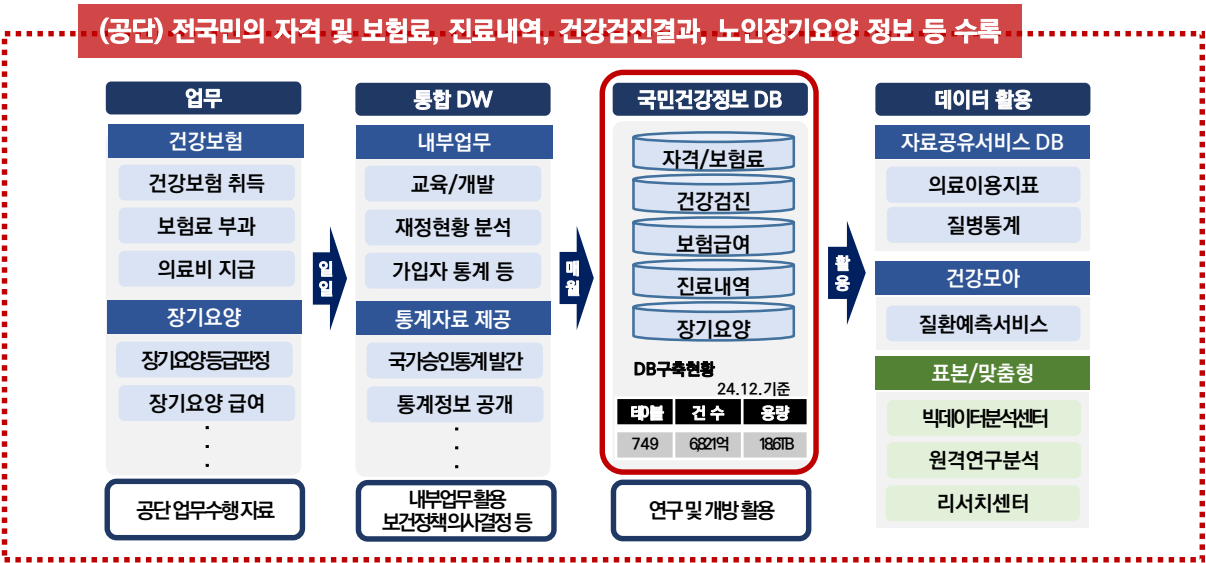
■ 건강보험 빅데이터 특징

	대표성 ▪ 단일보험자로서 전수 자료 보유 ▪ 전 국민의 사회경제적 수준 및 전체 의료공급자의 의료서비스 포함		현실성 ▪ 제한적 또는 실험적 정보가 아닌 현실이 반영된 자료
	완결성 ▪ 행위별 수가제에 기반한 상세 진료행위 및 처방내역 보유		계측성 ▪ 실제로 측정된 국가건강검진 자료 (예: BMI, 콜레스테롤, 혈압 등)
	연속성 ▪ 코호트자료를 구축하여 개인건강기록 및 관련 변수 조사 가능		효율성 ▪ 업무과정에서 발생하는 행정 자료이기 때문에 수집하고 구축하는 데 소요되는 비용 및 시간 절감



■ 국민건강정보 Data Base 구성

- 국민건강정보 DB란 공단이 관리하는 건강보험 및 장기요양보험 관련 정보를 연구 등에 이용할 수 있도록 전자적 방식으로 가명 처리한 자료로써, 이를 통해 건강정보 제공 및 표본·맞춤형 연구 DB 제공 등의 역할을 수행하고 있습니다.



■ 국민건강정보 세부 DB정보(자격/보험료)

- (자격 DB) 성별, 나이, 출생/사망, 거주지역 등으로 구성되며, 건강보험 및 노인장기요양보험의 자격 관리에 사용됨
- (보험료 DB) 보험료, 보험료 분위 등으로 구성됨



성별
■남/녀



나이
■수진시점/연초



출생
■출생년도



사망
■사망일자



장애
■장애유형(15개)
■장애등급(중증/경증)



거주지역
■환자, 의료기관, 사업장
■시군구 수준/지오코딩



가입자 유형
■직장가입자 및 피부양자
■지역세대주 및 세대원
■의료급여수급권자



보험료
■금액
■분위



사업장 유형
■산업 분야
■사업장 규모

BIG DATA



■ 국민건강정보 세부 DB정보(진료내역)

- (진료내역 DB) 요양기관 등에서 진료받은 환자의 의료 이용 정보로 구성
- (보험급여 DB) 보험적용을 받는 국민들의 의료이용을 위해 제공한 현금급여(요양급여, 건강검진) 및 현금급여(요양비, 본인부담상한제 환급금 등) 정보로 구성



■ 국민건강정보 세부 DB정보(건강검진)

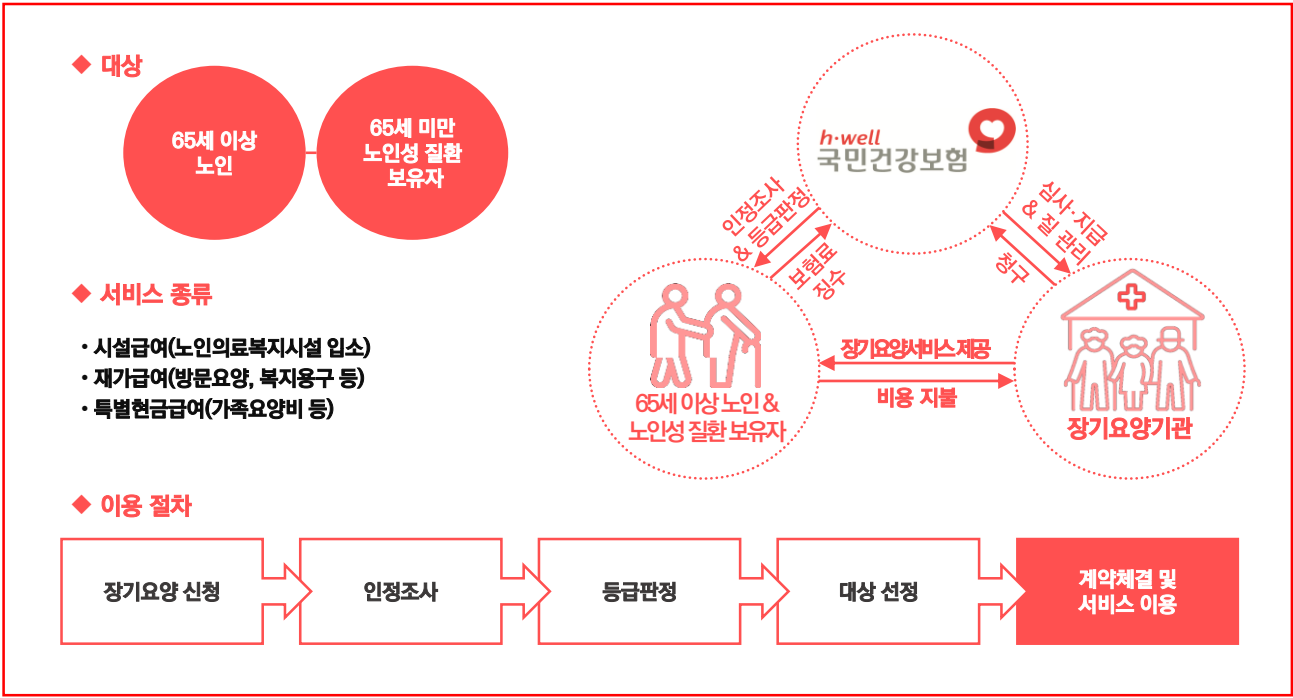
- (건강검진 DB) 공단에서 실시하는 건강검진(일반건강검진, 암검진, 영유아건강검진, 생애전환기검진 등) 관련 정보로 구성



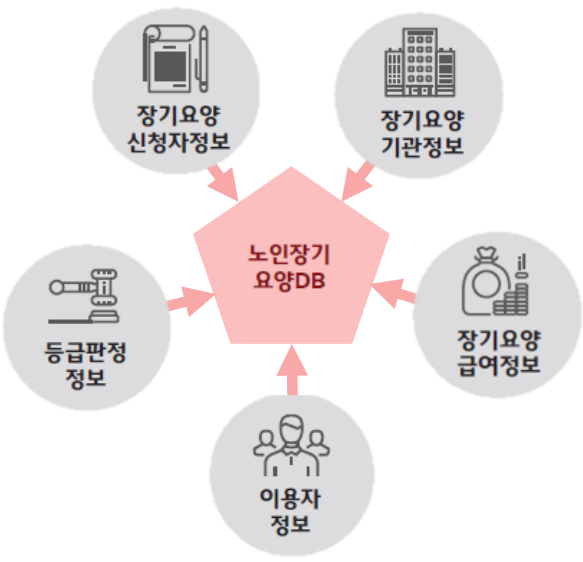


■ 국민건강정보 세부 DB정보(장기요양)

● (장기요양 DB) 노인장기요양보험 제도 운영을 통해 축적된 정보로 구성



● 장기요양 DB는 다음과 같은 세부 구성요소로 이루어져 있습니다.



신청	성별 / 나이 / 신청일자 / 대상자 구분 / 사망 여부
인정조사	신체기능 / 인지기능 / 행동변화 / 간호처치 / 재활 / 주거상태 / 현재 희망급여
등급판정	등급 판정일 / 판정결과 / 인정유효기간 / 의사 진단서
질병 및 증상	주요 질병 및 증상
급여유형	급여 종류 / 급여제공횟수 / 복지용구 품목 / 서비스 유형
청구	청구일자 / 급여비용 / 본인일부부담금
기관정보	기관 종류 / 시설 형태 / 설립 구분 / 소재지 / 인력 현황 등



■ 건강보험 빅데이터 활용 분야 소개

- 국민건강보험은 국민의 질병예방과 건강증진, 보건의료정책 수립 등에 필요한 정보 생산을 위해 다양한 분야에서 구축된 자료를 활용하고 있으며, 주된 활용분야는 크게 5가지로 구분할 수 있습니다.

01★ 빅데이터 개방

- 정책 근거 마련을 위한 연구 자료 제공
- 학술 연구 및 학위논문 등에 연구 자료 제공
- 공공데이터포털을 통한 통계 자료 제공

02★ 가명정보 결합

- 통계작성, 과학적연구, 공익적 기록보존 목적의 결합신청 가능
- 가명정보 결합제도 교육, 컨설팅 등 큐레이팅 수행

03 지식 기반 공동연구

- 학회 및 기관 연구 지원(공동연구 협약 체결)
- 지식 기반 네트워크 구축

04 근거 기반 정책 수립 지원

- 공단 내 주요 사업 지원
- 정책연구, 학술연구, 공동연구
 - 질환연구, Fact sheet 발간, 모니터링 지표 정의
- 국가승인통계 생산
- 한국인 건강지수 참조표준

05★ 맞춤형 건강 서비스 제공

- 개인건강기록(PHR) 기반 맞춤형 건강서비스 제공
- 건강보험 의료이용지표
- 의약품/의료기기/질병 통계 및 데이터 제공

건강보험 빅데이터 활용 방법



■ 건강보험 빅데이터 플랫폼(<https://nhiss.nhis.or.kr>)



- 공단에서 자체 운영 중인 빅데이터로 정책연구, 기술개발, 제도 개선 등 연구자 및 예비 창업(개발)자 등을 지원하고자 표본 및 맞춤형 데이터 제공 및 데이터 결합 서비스, 의료이용지표 정보 등을 제공합니다.

※ 빅데이터 플랫폼 활용메뉴얼 및 구체적인 안내 자료는 해당사이트-고객지원-자료실에 게시



■ 가명정보 결합서비스 운영(<https://nhiscs.nhis.or.kr>)



- 보건복지부가 지정한 결합전문기관으로서 보건의료분야에 특화된 가명정보 간 결합 뿐 아니라 이종간 결합이 가능하며, 결합적정성 검토, 결합 및 추가가명처리, 반출심사를 통해 안전한 결합데이터 반출을 지원합니다.

※ 가명정보 결합서비스의 구체적인 안내자료는 해당사이트-고객센터-자료실에 게시

BIG DATA

새로운 미래를 열다



■ 표본 DB란?

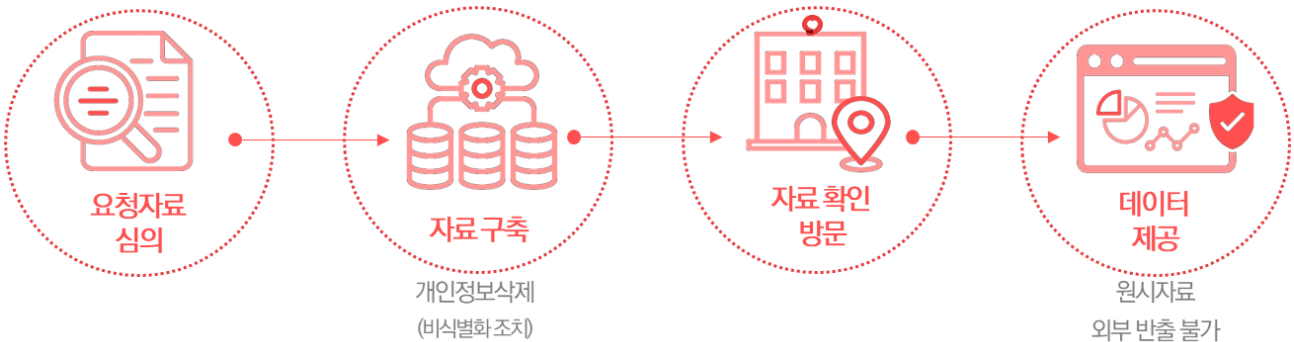
- 표본 DB란 전국민의 건강보험 빅데이터에서 연구수요가 많은 데이터를 선정하여 표본으로 추출, 비식별 조치 후 연구용으로 제공하는 규격화된 데이터셋을 말합니다.

구분	표본수	내용
표본 코호트DB	1,000천명	2006년 기준 전국민을 대표하는 약 100만명의 연구용 DB (전국민의약 2%)
건강검진 코호트DB	515천명	만 40~79세의 건강검진수검자 중심으로 의료이용, 검진 결과 등을 분석 가능 (모집단의 10%)
노인 코호트DB	511천명	2008년 기준 만 60세~80세 노령층을 중심으로 구축한 연구용 DB (모집단의 8%)

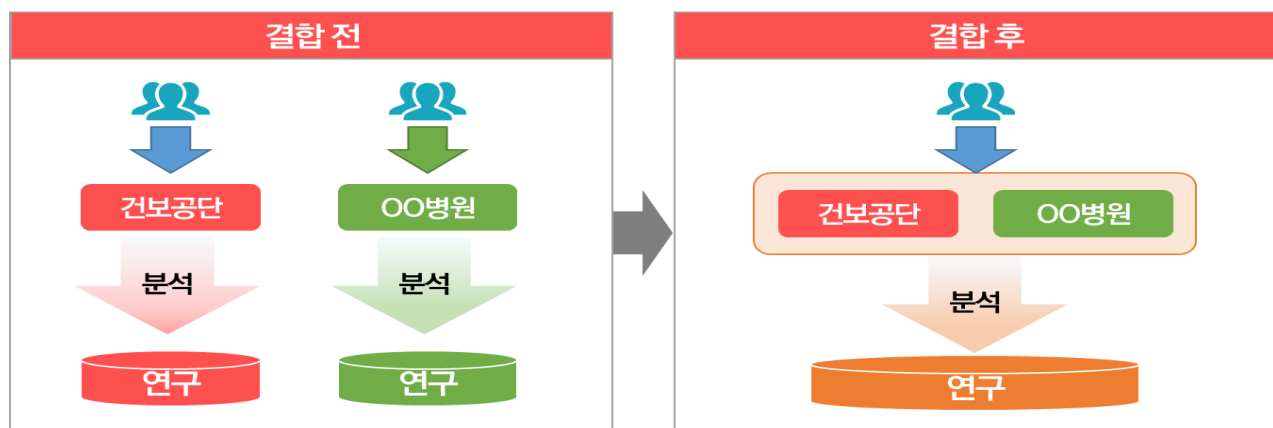


■ 맞춤형 DB란?

- 표본 DB로는 연구할 수 없는 데이터가 필요한 연구를 위해 연구자 신청에 맞춰 구축·제공하는 DB를 말합니다.
(예: 희귀난치성 질환 연구, 특정 지역/특정 계층별 연구분야 등)
- 맞춤형 연구 DB는 해당 데이터를 열람 및 연구 분석할 수 있는 PC가 설치된 공단 빅데이터 분석센터에서 사용 가능합니다.



- 데이터 3법 개정 후 근거법 없이도 연구자의 데이터 결합 연구가 가능합니다.
- 현재, 복수 기관간의 자료를 결합하여 연구를 수행하고자 하는 수요는 많고, 이를 담당할 공단의 인력, 조직도 갖춰진 상태지만 결합절차 인식의 부족 및 외부 자료보유기관의 체계가 덜 갖춰져 있는 상황으로 공단에서는 이를 위해 큐레이팅 (상담, 설명회), 찾아가는 결합프로세스 컨설팅 등을 추진 중입니다.

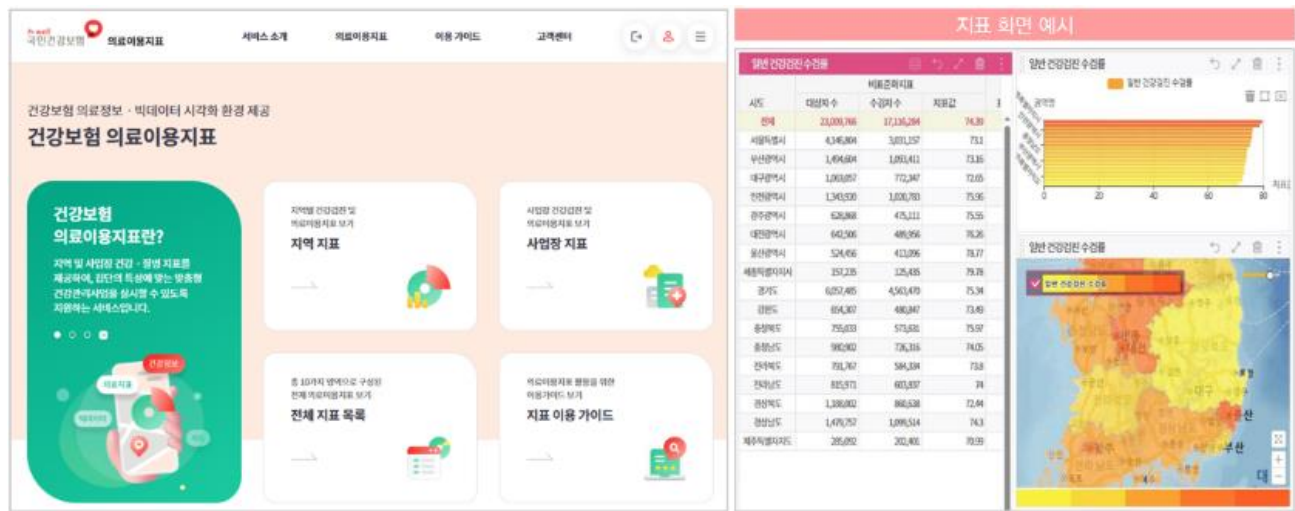


의료이용지표 소개



■ 의료이용지표란?

- 공단에서 운영 중인 의료이용지표 플랫폼(<https://nhiss.nhis.or.kr:4443>)으로써 건강보험 빅데이터 기반 지역 및 사업장의 건강·질병 지표를 제공하여 집단 특성에 맞는 맞춤형 건강관리사업 및 정책 의사결정의 객관적 근거 제공, 국민의 합리적 의료 선택을 지원하고 있습니다.



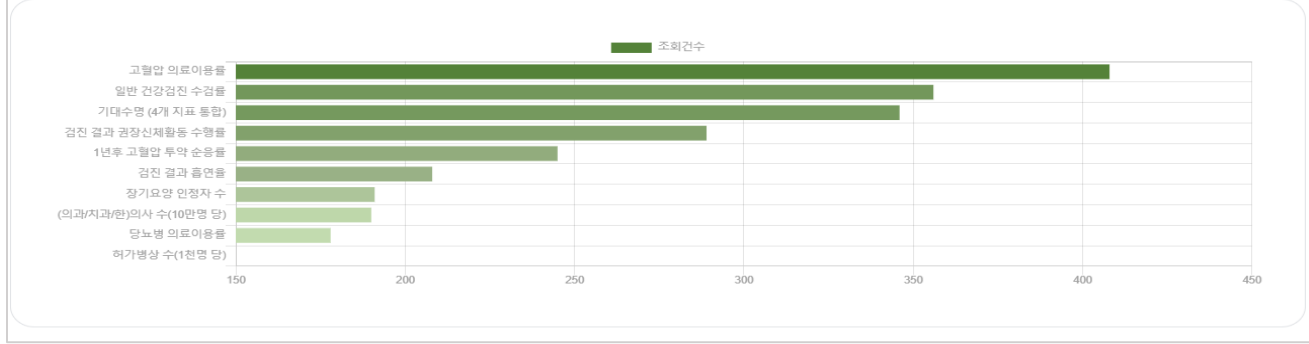
- 구성 지표는 ‘기본특성, 의료이용, 의료자원, 환경성질환, 건강검진, 만성질환 및 직무관련질환, 장기요양, 건강결과’의 8가지 영역(124종)으로 구성되어 있습니다.

지표 이용현황

전월 조회건수 기준 다빈도 지역 지표 이용현황 및 사업장 지표 이용현황입니다.

다빈도 지역 지표 이용현황

※ 전월 조회건수 기준



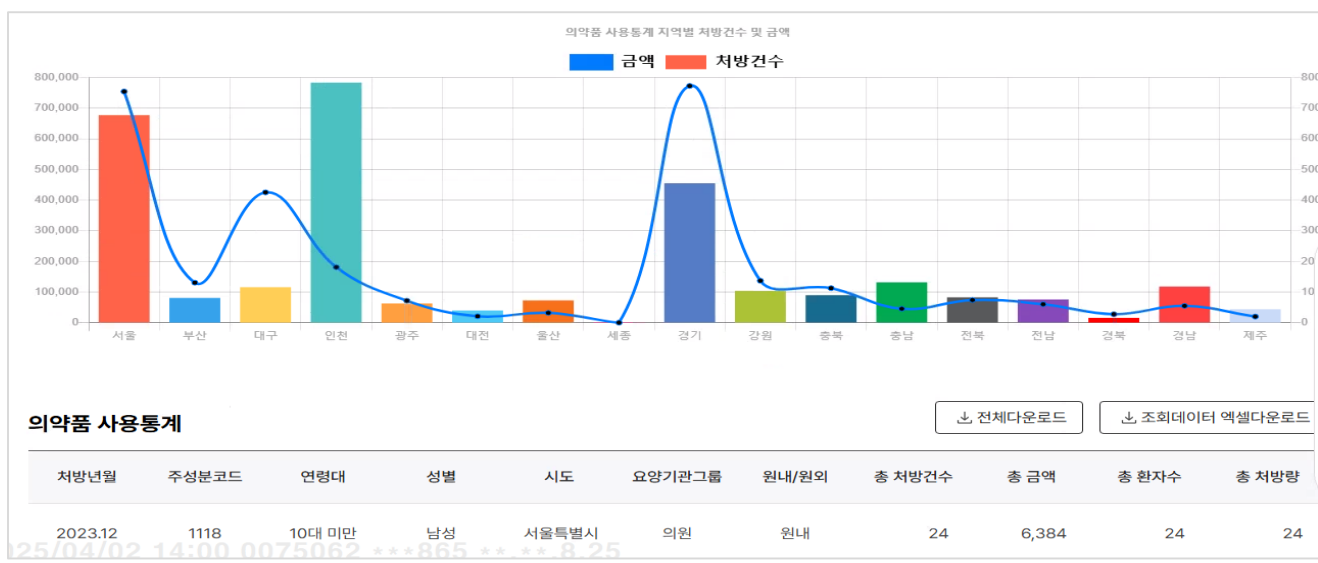
BIG DATA

새로운 미래를 열다



■ 의약품·의료기기 사용 통계 및 질병 통계란?

- 해당 통계 및 데이터는 4차 산업시대의 중요성이 커지고 있는 AI·신기술 발전에 대응하고자 AI의 주요 특징인 시계열성/완결성/신뢰성/범용성을 준수하여 정제된 데이터로서, 이를 통해 바이오·헬스산업 및 관련 연구기관의 발전을 지원하고자 합니다(건강보험 빅데이터 플랫폼을 통해 활용 가능).



- 구성 통계로는 ‘의약품 사용통계’, ‘의료기기 사용통계’, ‘질병 통계’가 있으며, 이용자의 수요를 반영한 더 많은 통계 지표를 개발할 예정입니다.

사용 통계별 구성 소개

의약품 사용통계

[기간] 2018.1월~, 월별 자료 ※ 매월 업데이트
[항목] 의약품별 청구건수, 환자 수, 총 금액, 총 처방량
[구분] 주성분코드(4자리), 사용년월, 연령대(10세 단위), 성별, 지역(시도), 요양기관그룹별, 원내·원외처방

의료기기 사용통계

[기간] 2018.1월~, 월별 자료 ※ 매월 업데이트
[항목] 의료기기별 청구건수, 환자 수, 총 금액, 총 처방량
[구분] 치료재료코드(5자리), 사용년월, 연령대(10세 단위), 성별, 지역(시도), 요양기관그룹별, 입원·외래

질병 통계

[기간] 2019 ~ 2023년 ※ 매년 업데이트
[항목] 질병세분류(3단 상병)별 실 인원수, 입·내원 환자 수 및 이용일수, 신 환자수, 사망자수
[구분] 연령대(5세 단위), 성별, 요양기관종별



■ 공공데이터포털이란?

- 행정안전부에서 운영 중인 공공데이터 공개 플랫폼(DATA.GO.KR)으로써 국민의 알 권리를 위하여 다양한 주제의 데이터를 공개하고 있으며, 목록 외에도 기초적인 데이터에 한하여, 신청을 통해 추가 제공하고 있습니다.

테마별

카테고리별

국가중점데이터별

제공기관유형별

<

교육

국토관리

공공행정

재정금융

산업고용

사회복지

식품건강

문화관광

>

인기 데이터

누적조회순

더보기 >

최신 데이터

수정일자순

더보기 >

소식지

1 / 2

II

파일 데이터

오른 API

1 소상공인시장진흥공단_상가(상권)정보

2 국민연금공단_국민연금 가입 사업장 내역

3 기상청_단기예보

4 특허청_KIPRIS_지식재산정보 검색서비스

5 경찰청_범죄 발생 지역별 통계

파일 데이터

오른 API

한국중부발전(주) 화학물질 사용 정보

한국중부발전(주)_바이오매스 사용량

한국중부발전(주) 해외 발전사업 및 해외 CSV 활동

한국중부발전(주)_사전협의 법제도 검토 기준

한국중부발전(주)_발전연료 입찰 규칙

데이터 제공신청

찾는 데이터가 없을 경우 데이터 제공을 신청해보세요.

공공데이터 제공신청

바로가기 >

분쟁조정신청

바로가기 >

개방현황

2025-07-29 현재

개방기관

파일데이터

오픈 API

표준데이터셋

1,083개

75,776건

11,497건

11,951건

- 공단은 해당 포털을 통해 **약 593건('25.7.29. 기준)의 데이터를 개방** 중이며, 검진기관 정보 및 장기요양기관 정보 등의 데이터 등도 **API형태로 제공**하고 있습니다.

■ 국민건강보험공단 공공데이터 다운로드 수 상위 5개 목록 ■

(단위 : 건)

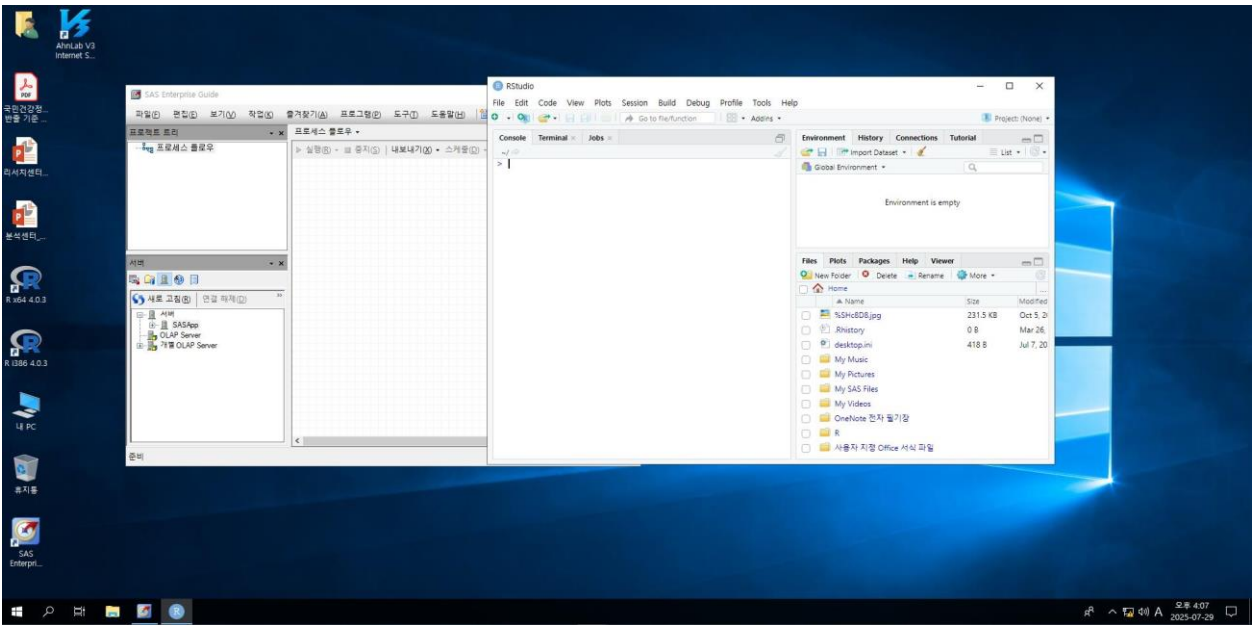
연번	목록명	제공주기	다운로드 수
1	장기요양기관 시설별 현황	수시	2,000
2	우울증 조울증 <u>조현병</u> 진료 통계	수시	1,853
3	연령대별 요양보호사 수(자격증 취득자 및 종사자)	수시	1,129
4	특정상병코드별_진료인원수	수시	802
5	간호간병통합서비스 운영현황 및 입원료구분 수가코드	수시	768



가상화룸 서비스 소개

- 국민건강보험공단이 수집/보유/관리하는 건강보험 및 장기요양보험 관련 자료를 정보 주체를 알아볼 수 없도록 비식별조치한 후 인터넷을 활용하여 제공하는 서비스 (이용자 PC를 활용, 원격 접속을 통하여 사용 인증 후 분석 서버에서 자료를 분석 가능)

가상화룸 서비스 사용 화면 예시



건강보험 빅데이터 분석센터 소개

- 연구자가 요청한 맞춤형(결합)데이터를 보안이 유지된 상태에서 활용할 수 있도록 공단에서 운영중인 분석 센터로 이용자 편의를 위해 전국 14개소(305석)에서 언제든지 분석에 활용할 수 있도록 운영 중입니다.

지역별 빅데이터 분석센터 좌석수 현황

구분	계	본부	서울	부산	대구	광주	전주	대전	청주	경인	일산 병원	서울대 병원	연세 의료원	서울 대학교	고려대 의료원
좌석 수	305	22	110	8	6	3	2	14	2	12	19	20	28	39	20

※ [운영시간] 월요일 ~ 금요일 9시 ~18시(단, 근로자의 날, 공휴일, 법정 공휴일은 휴무입니다.)



■ 빅데이터 활용 교육 소개

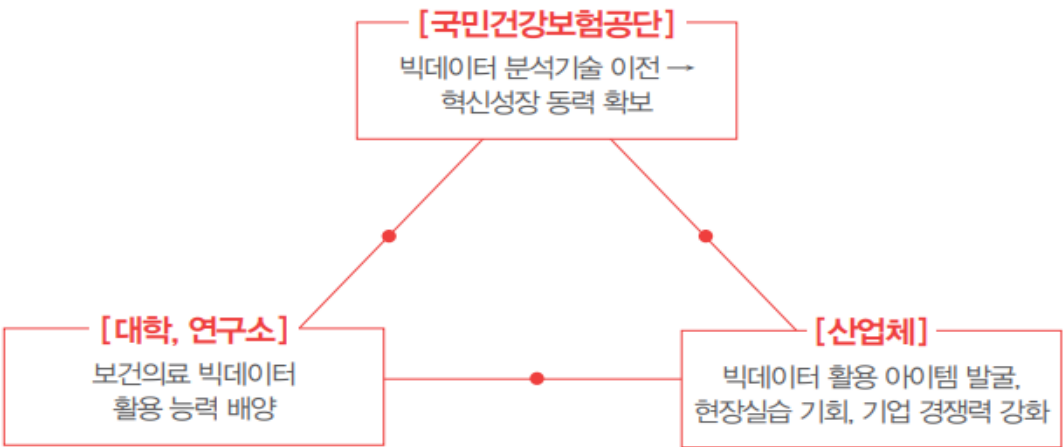
- 국민건강보험공단은 연구자의 연구개발역량 향상을 위하여 전문가 교육을 실시하고 있습니다. 신청방법은 건강보험 빅데이터 플랫폼에 접속 후 **고객지원 → 행사 및 교육안내 → 교육리스트**를 통해서 신청 부탁드립니다.

행사구분	제목	진행단계	차수	정원	대기인원	신청인원	신청접수기간	행사기간
교육	2025년 2분기 건강보험 빅데이터 활용 교육	종료	3 4	150명 150명	10명 10명	215명	2025-05-30 ~ 2025-06-20	2025-06-26 ~ 2025-06-27
교육	2025년 1분기 건강보험 빅데이터 활용 교육 추가 모집	종료	1 2	50명 50명	10명 10명	33명	2025-03-17 ~ 2025-03-23	2025-03-27 ~ 2025-03-28
교육	2025년 1분기 건강보험 빅데이터 활용 교육	종료	1 2	99명 99명	10명 10명	44명	2025-03-04 ~ 2025-03-14	2025-03-27 ~ 2025-03-28

■ 빅데이터 산학협력 아카데미 소개

- 국민건강보험공단은 지역 대학과 협력하여 건강보험 빅데이터 활용 능력을 배양하고 데이터분석 기술 이전을 위한 목적으로 ‘건강보험 빅데이터 산학협력 아카데미’를 운영하고 있습니다

산·학·연 협력체계 강화





■ 가명정보 결합 큐레이팅 소개



- 국민건강보험공단은 보건분야의 연구 및 발전을 위하여 서로 다른 가명정보를 결합하여 제공하는 가명정보 결합·반출 서비스를 제공하고 있으며, 서비스 활성화를 위하여 빅데이터 결합서비스 큐레이팅을 진행하고 있습니다

큐레이팅이란?

☞ 이용자가 체감하는 결합절차의 복잡성, 추진 과정의 어려움을 해소하고, 결합연구 활성화를 장려하기 위해 교육, 컨설팅, 홍보 등을 융합하여 수행하는 활동

■ 가명정보 결합 큐레이팅의 종류

제도이해형 큐레이팅

- 결합연구에 관심이 있는 기관, 연구자 등을 대상으로 가명정보 결합제도·절차 교육 실시(반기별 1회)

사례중심형 큐레이팅

- 결합연구 희망기관 및 연구자를 대상으로 단계별 교육, 안내, 컨설팅 등 1:1 맞춤 큐레이팅 실시(수시)

■ 가명정보 결합 큐레이팅 신청절차

- 국민건강보험 가명정보 결합서비스 홈페이지 접속(<https://nhiscs.nhis.or.kr>) 후 홈페이지 내 **고객센터 메뉴** 하단의 **교육 및 심포지엄 메뉴**를 통해 신청하거나 **결합전문기관 담당자에게 문의**하시면 됩니다

h·well 국민건강보험



제 목 : 국민건강보험공단 빅데이터 활용가이드

발행부서: 국민건강보험공단 빅데이터사업실

발 행 일 : 2025. 8. 11.

발 행 처 : 국민건강보험공단

발행번호 : 2025 - 01

※ 해당 자료에 대한 무단 도용 및 수정을 엄금하며, 이를 어길시 법적 처벌을 받을 수 있습니다.