

# BOK 이슈노트

2026년 2월 9일

## 첨단 바이오헬스 산업 육성방안 - 바이오 데이터 활용 기반 구축을 중심으로

### 성원

한국은행 경제연구원 미시제도연구실 과장  
Tel. 02-759-5485  
sungwon31@bok.or.kr

### 최이슬

한국은행 경제연구원 미시제도연구실 과장  
Tel. 02-759-5435  
eschoi@bok.or.kr

### 이동원

한국은행 경제연구원 미시제도연구실 실장  
Tel. 02-759-5404  
dolee@bok.or.kr

- 1. 저성장 위기 속에서 바이오헬스 산업(제약·의료기기)이 신성장동력으로 주목받고 있다.** 고령화로 인한 의료 수요 증가와 생명공학-AI 융합 가속화로 향후 5년간 글로벌 바이오헬스 시장은 연평균 5.0% 성장할 전망이다. 이는 국내 주력 산업인 자동차 성장률(2.7%)의 약 두 배에 달하는 수치다.
- 2. 국내 바이오헬스 산업은 바이오시밀러(복제약)와 CDMO(위탁생산개발) 등 틈새시장을 중심으로 빠르게 성장했으나, 혁신생태계가 미흡하여 고부가가치를 창출하는 혁신 신약과 첨단 의료기기 분야에서는 선도국과 큰 격차를 보이고 있다. 세계 매출 상위 30 위 내에 포함된 국내 기업은 없으며, 국내 1위인 삼성바이오로직스의 매출(28억 달러)도 글로벌 상위 3개 제약사 평균(578억 달러)의 5% 수준에 불과하다.**
- 3. 이는 특히 등을 통해 고수익을 창출할 수 있는 원천기술의 부족에 기인한다.** 실제로 2016~2023년 미국 내 바이오헬스 특허출원건수는 세계 9위에 그쳐, 전체 분야 순위(4위)에 비해 낮은 수준에 머물러 있다. 2024년 의약품 개발 기술 수준도 미국과 3.6년의 격차가 있으며, 특히 중국에 대한 상대적 기술 우위가 희석된 상황이다.
- 4. 최근 AI 기술이 바이오헬스 산업의 경쟁 구도를 재편하는 '게임체인저'로 부상하면서, 우리나라는 기존 혁신생태계의 한계를 넘어설 기회를 맞이하고 있다.** AI는 신약 개발 기간을 30~50% 단축하는 등 R&D 비용을 획기적으로 절감하고 정밀 의료, 수술 보조 로봇 등 신시장을 창출함으로써, 우리나라와 같은 추격국이 선도국을 추월할 가능성을 제공한다.
- 5. 우리나라는 우수한 바이오 데이터 수집·연계 인프라를 보유하고 있어 이러한 기회를 실현할 충분한 잠재력을 갖추고 있다.** 단일 건강보험제도를 기반으로 수집하는 5천만 인구의 건강보험 및 병원 임상 데이터는 'AI 시대의 다이아몬드'라 불릴 만큼 세계적으로도 희소성이 높은 국가 전략 자산이 될 수 있다. 이에 대해 미국 출원 AI 특허와 FDA 승인 AI 의료기기 건수 모두 세계 4위를 기록해 글로벌 수준의 AI 기술력을 확보하고 있으며, 서울은 세계 2위 규모의 글로벌 임상시험 유치 도시로서 높은 의료 인프라를 입증하고 있다.

6. 이러한 데이터 수집·연계 인프라를 바탕으로 대규모 고품질 바이오 데이터를 구축하고 전략적으로 활용한다면, 글로벌 선도 기업과의 공동 개발은 물론 다국가 임상시험을 주도하는 글로벌 바이오헬스 R&D 허브로 자리매김할 수 있다.
7. 그러나 이 기회의 핵심인 바이오 데이터 활용은 저조하다. 2023년 의료·헬스케어 AI 기업의 81.4%가 '데이터 확보·품질 문제'를 주요 애로사항으로 지적했다. 이는 우수한 데이터 수집·연계 인프라를 갖추고도 정제, 공유 등 활용 단계에 병목 현상이 존재함을 시사한다.
8. [근본 원인] 바이오 데이터 활용이 저조한 데는 데이터 활용의 위험·비용은 정보주체(개인)와 수집관리자(병원)가 부담하는 반면 이익은 활용자(기업·연구자)와 사회 전체로 분산되는 '인센티브 불일치'가 작용하고 있다.
  - (정보주체: 개인) 공익적 활용 가치에는 공감하나 프라이버시 침해를 우려해 제공을 꺼린다. 설문 조사 결과, 데이터 제공 의향이 있다는 응답은 26.9%에 그쳤으며, 거부자의 52.5%는 해킹·유출로 인한 개인정보 침해를, 20.9%는 고용 차별, 보험 가입 거부 등 실질적 피해를 우려했다. 현행 사전 동의 및 비식별화 중심의 개인정보 보호 체계는 실질적 통제권을 담보하지 못해 신뢰를 구축하지 못하고 있다
  - (수집관리자: 병원) 데이터 정제 비용, 법적·평판 리스크 등 부담이 크지만 보상은 미미하다. 이에 따라 공유 기피 및 데이터 독점이 합리적 선택이 된다. '2020년 보건의료정보화 실태조사'에 따르면, 병원 데이터를 연구 목적으로 외부와 공유한 비율은 21%에 그쳤으며, 외부인이 단독으로 수행하는 연구의 경우 그 비율은 17.4%로 더욱 낮았다.
  - (활용자: 기업·연구자) 복잡한 접근·결합 절차, 법적 불확실성 등으로 적극적 활용이 어렵다. 2020년 10월~2023년 10월 3년간 보건의료 분야 데이터 결합·활용 건수는 78건에 그쳤다.
9. [정책 방안] 이러한 구조적 문제를 해소하고 데이터 활용을 활성화할 방안으로 '국가 승인형 바이오 데이터 개방 체계'(이하 국가 승인 체계)를 제안한다. 해당 체계는 사전 심사를 통해 법에서 정한 공익성 요건을 충족한 연구에 대해 데이터 활용을 승인하되, 승인된 연구에 대해서는 사전 동의 면제 등 규제를 완화하고 데이터 유통을 지원함으로써 인센티브 불일치를 완화하는 것을 목표로 한다.
  - (공적 승인과 통제권 담보) 사전 심의·승인을 통해 법에서 정한 공익적 활용을 보장하고, 정보주체에게 언제든 활용을 거부<sup>opt-out</sup>하거나 재허용<sup>opt-back-in</sup> 할 수 있는 통제권을 부여하여 신뢰 기반의 자발적 데이터 제공을 유도한다. 실증분석 결과, 이러한 '공익적 활용 보장'과 '정보 통제권 강화' 정책은 일반신체정보 및 정신건강정보 제공 의향을 각각 8.2%p, 15.5%p 제고하는 것으로 나타났다.

- **(승인된 연구에 대한 규제 완화)** 심사를 통해 공익성이 인정된 활용에 대해서는 사전 동의 면제, 포괄적 동의 허용 등 규제 완화를 통해 연구자의 법적·행정적 부담을 덜어준다.
- **(데이터 유통 생태계 활성화)** 데이터 탐색부터 결합, 심의, 제공까지 일괄 지원하는 통합 중개 허브를 구축하고 데이터 중개사를 육성하는 등 데이터 유통 생태계를 활성화함으로써, 데이터 접근 비용은 낮추고 민간 병원 등의 고품질 데이터 생산·공유에는 합리적 보상 체계를 제공한다. 이를 통해 공적 승인을 받은 연구가 실질적으로 수행될 수 있도록 정책적으로 지원한다.

**10. [거버넌스]** ‘국가 승인 체계’의 실효성을 확보하기 위해서는 EU의 ‘Health Data Access Bodies’와 같은 별도의 전담 기구 설립을 고려할 필요가 있다. 해당 기구는 독립성, 전문성, 재정적 지속가능성을 갖춰야 하며, 데이터의 직접 관리보다는 민간의 데이터 생태계를 지원하는 시장 조성자이자 신뢰 보증자 역할에 주력해야 한다.

**11. [기대 효과]** ‘국가 승인 체계’는 바이오 데이터의 전략 자산화를 통해 저성장 국면의 우리 경제에 새로운 성장동력을 제공할 것으로 기대된다. 이는 데이터 시장 확대와 R&D 효율성 제고를 통해 글로벌 바이오헬스 R&D 허브 도약과 양질의 일자리 창출이라는 선순환 구조를 이끌 것이다. 나아가 개인 맞춤형 건강관리부터 신약 개발, 정책 수립에 이르는 전 영역의 데이터 활용을 촉진함으로써 궁극적으로 국민 건강 증진에 크게 기여할 것이다.

- 본 자료의 내용은 한국은행의 공식견해가 아니라 집필자 개인의 견해라는 점을 밝힙니다. 따라서 본 자료의 내용을 보도하거나 인용할 경우에는 집필자명을 반드시 명시하여 주시기 바랍니다.
- 본 보고서 작성과 관련하여 유익한 논평을 해주신 국회입법조사처 김은정 입법조사관, 건양대 신지은 교수께 감사드립니다. 그리고 보고서 작성에 많은 도움을 주신 한국은행 경제연구원 이재원 원장, 나승호 부원장, 임호성 부원장, 조태형 前부원장, 이인로 차장, 그리고 원내세미나 참석자들께 감사드립니다. 아울러 기초 자료 수집과 정리에 큰 도움을 준 황현하 인턴에게도 감사를 전합니다.



## 1. 검토 배경

### 1. 저성장 위기를 극복하기 위해서는 신성장동력 산업의 발굴과 육성이 시급하다.

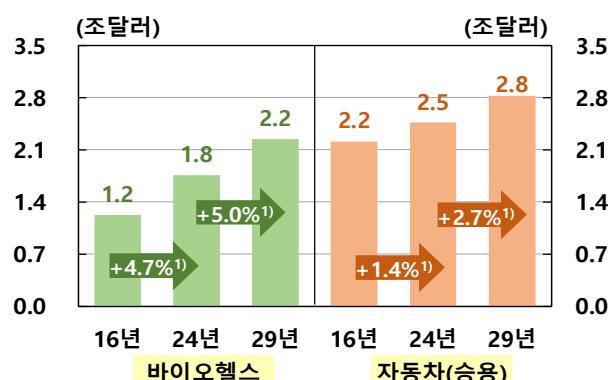
현재 우리 경제는 성장잠재력 약화라는 구조적 난관에 봉착해 있다. OECD(2025)에 따르면, 2011년부터 2025년까지 우리나라의 잠재성장률은 1.8%p 하락하여 OECD 회원국 중 세 번째로 큰 낙폭을 기록했다. 이는 급격한 고령화에 따른 노동 공급 감소와 주력 산업의 경쟁력 약화가 복합적으로 작용한 결과다. 특히 Google이나 Tesla와 같은 글로벌 혁신기업을 육성하여 차세대 주력 산업을 창출하지 못한 점은 우리 경제의 뼈아픈 현실이다.

### 2. 이러한 상황에서 바이오헬스 산업은 미래 성장동력의 핵심 대안으로 부상하고 있다.

전 세계적인 고령화로 의료 수요가 급증하고 생명공학 기술이 비약적으로 발전함에 따라, 바이오헬스 산업은 본격적인 구조적 성장기에 진입하였다. 제약과 의료기기를 중심으로 한 글로벌 바이오헬스 시장은 향후 5년간 연평균 5% 성장할 것으로 전망되는데, 이는 국내 주력 산업인 자동차 산업 성장률의 두 배에 달하는 수치이다(그림 1.1). 특히 유례없이 빠른 고령화가 진행 중인 우리나라에서 바이오헬스 산업은 급증하는 의료·돌봄 수요를 발판으로 삼아 양질의 일자리를 창출하는 새로운 성장 엔진이 될 잠재력이 충분하다.

바이오헬스 산업은 높은 성장세를 지속할 전망

[그림 1.1] 글로벌 바이오헬스·자동차 매출 실적·전망



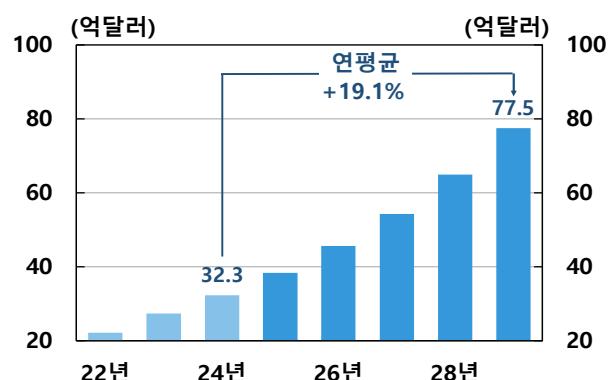
주: 1) 연평균 증가율(CAGR) 기준

2) 바이오헬스 산업은 제약 및 의료기기 산업을 의미

자료: Statista(25.12월, 25.6월)

생명공학-AI 융합이 빠르게 확산

[그림 1.2] 글로벌 생명공학 AI 매출 전망



자료: Markets and Markets(25.1월)

### 3. 최근 AI 기술의 혁신은 바이오헬스 산업에 중대한 전환점을 마련하고 있다.

AI와 생명공학의 융합은 R&D 패러다임을 기존의 실험 중심에서 데이터와 알고리즘 중심으로 전환하며, 신약 및 의료기기 개발에 소요되는 비용과 시간을 획기적으로 단축하고 있다. 이는 기술적 한계를 극복하고 신시장을 창출하는 기폭제가 되고 있으며, 실제로 글로벌 생명공학 AI 시장은 향후 5년간 연평균 19%의 가파른 성장세를 보일 것으로 예상된다(그림 1.2).

**4. AI 혁신이 가져올 기회를 선점하기 위한 열쇠는 '바이오 데이터'이다.** 데이터는 AI의 원천이며, 특히 생명과 직결된 바이오헬스 분야에서는 대표성을 갖춘 대규모 고품질 데이터의 확보가 경쟁력의 관건이다. 이에 바이오 데이터는 막대한 부가가치를 창출하는 'AI 시대의 다이아몬드'로 주목받고 있다. 따라서 바이오 데이터를 원활하게 활용할 수 있는 제도적·기술적 환경을 조성하는 것은 우리나라가 바이오헬스 선도국으로 도약하기 위한 필수 선결 과제이다.

**5. 이미 주요 선진국들은 바이오 데이터를 국가 핵심 자산으로 규정하고, 그 활용 기반을 확대하기 위한 치열한 경쟁에 돌입했다.** 스웨덴과 핀란드는 양질의 국가 단위 바이오 데이터를 앞세워 Roche, GSK, AstraZeneca 등 글로벌 기업의 R&D 투자를 성공적으로 유치하고 있다(Tony Blair Institute for Global Change 2024). 또한 EU의 'European Health Data Space'<sup>1)</sup> 법안과 미국의 'Web of Biological Data'<sup>2)</sup> 구상 역시 바이오 데이터 활용을 촉진하기 위한 제도적 기반 강화의 일환이다. 이러한 글로벌 흐름에 뒤처진다면 우리는 국민 건강 증진은 물론, 미래 성장동력 확보와 일자리 창출의 기회를 상실할 위험이 크다.

**6. 이에 본고는 AI 혁신이 제공하는 기회를 바탕으로 바이오헬스 선도국으로 도약하기 위한 핵심 과제로 바이오 데이터 활용 기반 확충을 제시한다.** 그리고 구체적인 실천 방안으로는 '국가 승인형 바이오 데이터 개방 체계'의 구축을 제안한다. 이는 사전 심사를 통해 법에서 정한 공익성 요건을 충족한 연구에 대해 데이터 활용을 승인하되, 승인된 연구에 대해서는 사전 동의 면제 등 규제를 완화하고 데이터 유통을 지원하는 것이 콜자다.

**7. 본고의 구성은 다음과 같다.** 2장에서는 국내외 바이오헬스 산업의 현황을 분석해 기회와 한계를 점검하고, 3장에서는 AI 기반 R&D 패러다임 전환과 국내 데이터의 전략적 가치를 진단한다. 4장에서는 국내 데이터 활용의 구조적 문제점을 규명하며, 5장에서는 해결책으로서 '국가 승인형 바이오 데이터 개방 체계'의 구축을 제안한다. 마지막 6장에서는 전체 내용을 종합·정리한다. 아울러 본고의 분석 범위는 진단과 치료라는 의료 분야의 기능을 수행하면서 글로벌 시장 진출을 직접 공략할 수 있는 제약 및 의료기기 부문으로 한정하며, 화장품과 의료서비스는 제외한다<sup>3)</sup>.

1) 「European Health Data Space」(EHDS)는 EU 시민이 국경을 넘어 자신의 건강 데이터를 활용해 의료서비스를 받고, 해당 데이터를 연구·혁신 목적으로 안전하게 재사용할 수 있도록 지원하는 EU 차원의 통합 보건 데이터 법안이다. 이 법안은 2024년 4월 유럽의회에서 채택되었으며, 2025년부터 단계적으로 시행되고 있다.

2) 2025년 4월 미국 의회 자문기구인 「National Security Commission on Emerging Biotechnology」(NSCEB)는 정부·학계·산업계 등 다양한 출처의 생물학 데이터를 통합해 AI 개발에 필요한 데이터에 단일 창구로 접근할 수 있도록 지원하는 중앙 데이터 인프라인 「Web of Biological Data」의 구축을 제안하였다. 세부 내용은 「Charting the Future of Biotechnology」 보고서(25.4월)를 참조하기 바란다.

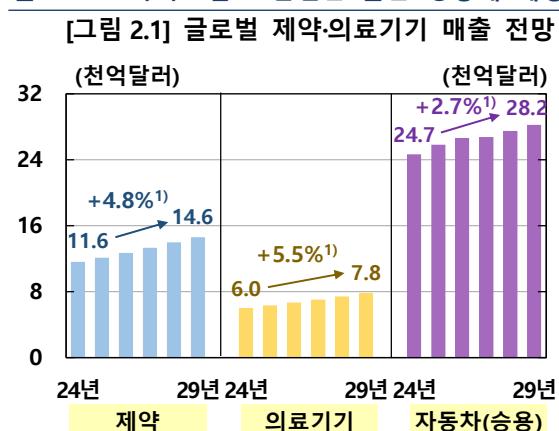
3) 「바이오헬스 산업 혁신전략」(관계기관 합동, 19.5월)은 바이오헬스 산업을 생명과학과 의약학 지식을 기반으로 인체용 제품을 생산하거나 서비스를 제공하는 산업으로 정의한다. 이에 따르면 해당 산업에는 의약품과 의료기기뿐만 아니라 화장품과 의료서비스(예: 병의원)도 포함된다.

## 2. 국내외 바이오헬스 산업 현황

### ■ 글로벌 바이오헬스 산업 현황: 미래 성장을 주도할 새로운 엔진

8. 최근 글로벌 바이오헬스 산업은 견조한 성장세를 지속하며 저성장 시대의 새로운 성장 엔진으로 부상하고 있다. 2016년부터 2024년까지 글로벌 제약 및 의료기기 산업 매출은 차세대 바이오의약품과 만성질환 의료기기를 중심으로 각각 연평균 4.4%, 5.2% 증가했다<sup>4)</sup>. 특히 의료 수요는 경기 변동에 둔감한 특성이 있어, 글로벌 불확실성이 확대되는 시기에도 바이오헬스 산업은 안정적인 미래 성장동력으로서 그 가치를 인정받고 있다. 이러한 흐름은 앞으로도 지속되어, 향후 5년간 글로벌 제약 및 의료기기 산업 매출은 각각 연평균 4.8%, 5.5% 성장할 것으로 전망된다(그림 2.1).

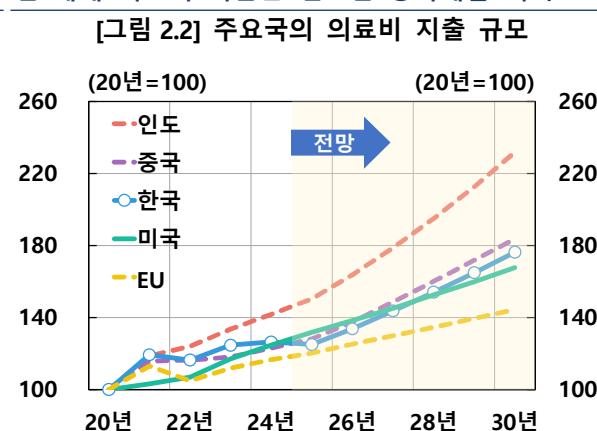
앞으로도 바이오헬스 산업은 높은 성장세 예상



주: 1) 연평균 증가율(CAGR) 기준

자료: Statista(25.12월, 25.6월)

전 세계 의료비 지출은 견조한 증가세를 지속



주: 1) 정부, 가계 및 민간기업의 지출을 모두 포함

자료: Statista(25.12월, 25.6월)

9. 수요 측면에서는 선진국의 인구 고령화와 신흥국의 의료 지출 확대가 산업 성장을 지속적으로 견인할 것이다(그림 2.2). 전 세계 65세 이상 고령인구 비중은 2020년 9.3%에서 2030년 11.7%로 높아진 데 이어, 2040년에는 14.3%까지 확대될 것으로 예상된다<sup>5)</sup>. 특히 우리나라를 포함한 주요 선진국은 2030년 고령인구 비중이 20~25%(우리나라는 2040년 30% 상회)에 육박하며 암, 심혈관 질환, 신경퇴행성 질환 등 만성질환 관리를 위한 의료 수요가 급증할 전망이다. 아울러 중국, 인도 등 신흥국에서도 소득 수준 향상과 의료 접근성 개선에 힘입어 정부 차원의 의료 지출이

4) 세부 품목별 글로벌 바이오헬스 산업 동향은 <참고 1>을 참조하기 바란다.

5) 주요국의 연도별 65세 이상 고령인구 비중은 다음과 같다.

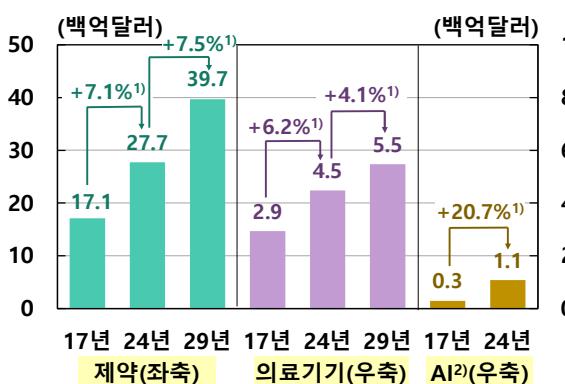
크게 늘어날 전망이다. 이에 따라 전 세계 의료비 지출은 2024년 11.2조 달러에서 2030년 15.1조 달러로 연평균 5.1%씩 견조하게 성장할 것으로 예측된다<sup>6)</sup>.

**10. 공급 측면에서는 AI 기술과 생명공학의 융합이 산업 성장을 가속화하고 있다.**

2017년부터 2024년까지 글로벌 제약 및 의료기기 산업의 R&D 투자는 각각 연평균 7.1%, 6.2% 증가했다. 특히 AI가 바이오헬스 R&D의 기간과 비용을 획기적으로 단축하는 '게임체인저'로 부상하면서, 헬스케어 AI 분야에 대한 투자는 연평균 20.7% 급증했다(그림 2.3). 이러한 혁신은 차세대 바이오의약품<sup>7)</sup> 개발과 데이터 기반의 맞춤형 정밀 의료 등 새로운 고부가가치 수익 모델 창출로 이어지고 있다. 향후 5년 간 바이오의약품과 디지털 치료기기의 글로벌 매출은 각각 연평균 8.0%, 9.5%의 고성장을 기록할 것으로 전망된다(그림 2.4).

**바이오헬스 R&D 투자는 큰 폭 증가세 지속 전망**

[그림 2.3] 글로벌 바이오헬스 산업 R&D 투자

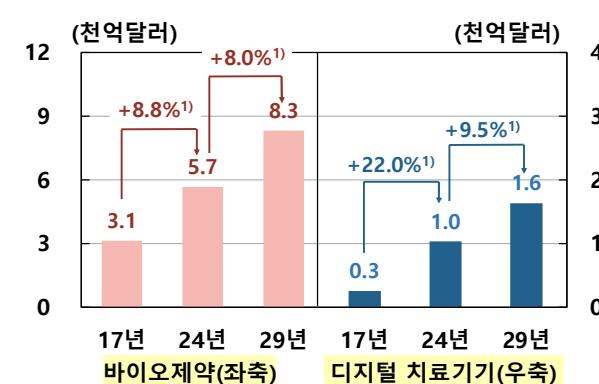


주: 1) 연평균 증가율(CAGR) 기준

2) Private investment in AI for medical & health care  
자료: Statista(25.12월, 25.6월), Quid via AI Index Report(2025)

**바이오의약품·디지털 치료기기 매출이 높은 성장세**

[그림 2.4] 글로벌 바이오제약·디지털 치료기기 매출



주: 1) 연평균 증가율(CAGR) 기준

자료: Evaluate Pharma(25.7월), Statista(25.12월, 25.6월)

	2010년	2015년	2020년	2025년 <sup>e)</sup>	2030년 <sup>e)</sup>	2040년 <sup>e)</sup>	(%)
▶ 전 세계	7.6	8.3	9.3	10.4	11.7	14.3	
▶ 미 국	12.8	14.3	16.1	18.4	20.4	22.0	
▶ E U	16.2	17.5	19.1	20.9	23.0	26.2	
▶ 한 국	11.0	13.0	15.8	20.3	25.1	33.8	

자료: UN World Population Prospects 2024

6) 연도별 전 세계 의료비 지출 실적 및 전망은 다음과 같다.

	2016년	2020년	2024년	2028년 <sup>e)</sup>	2030년 <sup>e)</sup>	(조달러)
▶ 의료비 지출	7.5	9.3	11.2	13.6	15.1	

자료: Statista(25.1월)

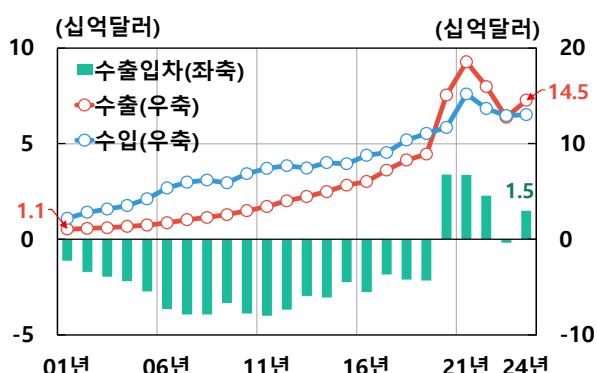
7) 바이오의약품은 생물체의 세포나 유전자를 원료로 하여 제조되는 의약품으로, 화학물질 기반의 합성의약품에 비해 독성이 낮고 부작용이 적다. 또한 특정 환자군을 표적으로 치료할 수 있어 희귀·난치성 질환 치료에 효과적이다.

## ■ 국내 바이오헬스 산업 현황: 틈새시장 공략의 성공, 그러나 원천기술의 벽

### ▶ (현황) 틈새시장을 중심으로 글로벌 입지 확대

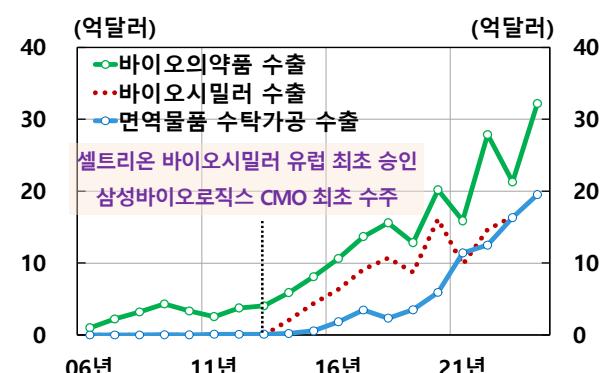
11. 우리나라 바이오헬스 산업은 짧은 역사에도 불구하고 비약적인 성장을 거듭하며 글로벌 시장에서의 입지를 강화하고 있다. 19세기부터 신약 개발을 주도해 온 유럽 등 선진국<sup>8)</sup>과 달리, 우리나라는 1999년에야 첫 국산 신약(SK케미칼, 선플라주)을 개발했고, 2003년에 이르러서야 미 FDA 승인을 받은 의약품(LG화학, 팩티브)을 배출했다. 의료기기 분야 역시 1950년대 개발된 초음파 영상진단기기를 1985년에야 국산화(삼성메디슨)하는 등 출발이 늦었다. 그러나 우리 산업은 후발주자의 한계를 극복하고 2001년부터 2024년까지 연평균 12.0%의 놀라운 수출 성장률을 기록했다. 그 결실로 만년 적자였던 바이오헬스 산업 무역수지는 2020년 사상 첫 흑자를 달성했으며, 2024년에도 15억 달러 규모의 흑자 기조를 유지하고 있다(그림 2.5).

바이오헬스 산업의 수출은 큰 폭 성장세를 지속  
[그림 2.5] 바이오헬스 산업 수출 및 수입



주: 1) 바이오헬스 산업은 제약 및 의료기기 산업을 지칭  
자료: 식약처

바이오의약품과 면역물품 수탁가공 수출이 크게 증가  
[그림 2.6] 바이오의약품 및 면역물품 수탁가공<sup>1)</sup> 수출



주: 1) 면역물품(HS3002) 수출거래 중 '국내 외국인투자업체가 외국으로부터 수탁받아 가공후 수출' 또는 '기타 일반업체가 수탁받아 가공 후 수출'로 신고된 건  
자료: 식약처, 관세청

12. 이러한 성과는 '빠른 추격자'<sup>fast follower</sup> 전략을 통해 바이오시밀러 바이오의약품 복제약과 CDMO 위탁개발생산 등 틈새시장을 효과적으로 공략한 결과다. 2001~2024년 중 제약 수출은 바이오시밀러와 CDMO를 필두로 연평균 12.5% 성장했으며, 의료기기 수출도 초음파 영상진단기기와 임플란트를 중심으로 연평균 11.3% 증가했다(그림 2.6).

- (바이오시밀러) 국내 바이오시밀러 산업은 글로벌 시장 환경 변화를 기회로 삼아 급성장했다. 2000년대 중반 주요국이 약가 인하를 위해 바이오시밀러의 승인 절차를 간소화<sup>9)</sup>하고, 2012년부터 블록버스터급(연매출 10억 달러 이상) 바이오의약

8) 일례로 최초의 합성의약품인 아스피린은 1897년에 개발되었다.

9) 바이오의약품은 세포 등 생물 유래 물질로 제조되어 구조가 매우 복잡하다. 이로 인해 생산 공정의 미세한 차이도 제품 특성에 영향을 미쳐, 합성의약품의 복제약(제네릭)처럼 완전히 동일한 제품을 만들기 어렵다. 이

품의 특허 만료가 이어지면서 시장 기회가 열렸다. 셀트리온<sup>10)</sup> 등 국내 기업은 이에 빨빠르게 대응하여 개발을 주도했고, 그 결과 2013~2024년 바이오의약품 수출은 연평균 20.7%라는 고속 성장을 이뤘다.

- **(CDMO)** 살아 있는 세포를 원료로 사용하는 바이오의약품은 생산 공정이 복잡하고 대규모 설비 투자가 필요해 전문 생산 기업에 대한 의존도가 높다. 삼성바이오로직스<sup>11)</sup> 등 국내 기업이 과감한 투자로 글로벌 제약사의 위탁 수요를 흡수하면서 우리나라는 CDMO 강국으로 도약하였다. 이에 따라 면역물품<sup>12)</sup> 수탁가공 수출은 2013년부터 2024년까지 연평균 60.6%의 높은 성장세를 기록했다.
- **(의료기기)** 2000년대 초반까지 영세했던 국내 의료기기 산업은 진단용 의료기기 국산화와 건강보험 급여 확대로 내수 기반을 다지며 성장했다. 특히 2004년 의료기기법 제정으로 품질 관리 체계가 선진화되면서 수출 경쟁력이 강화되었다. 이를 바탕으로 2000년대에는 초음파 영상진단기기(2003~2013년 연평균 16.5%)가, 2010년대 이후에는 치과용 임플란트(2010~2024년 연평균 20.9%)가 수출 주력 품목으로 자리 잡았다(그림 2.7). 코로나19 팬데믹 기간에는 진단키트 수출이 급증하며 산업 성장을 견인하기도 했다.

**13. 향후에도 이들 주력 분야는 글로벌 수요 확대를 바탕으로 견조한 성장세를 이어갈 전망이다**(그림 2.8). 바이오시밀러는 미국 등의 약가 인하 정책<sup>13)</sup>과 2030년까지 예정된 30여 종의 블록버스터급 바이오의약품 특허 만료에 힘입어 시장이 지속적으로 확대될 전망이다<sup>14)</sup>(Li et al. 2024). CDMO 분야 역시 고난도 공정이 요구되는 바이오의약품 특성상 외주 수요가 계속 증가하겠으며, 특히 미국 등의 '탈중국' 공급망 재편 움직임이 우리 기업에게 반사이익을 가져다줄 것으로 보인다<sup>15)</sup>. 치과기기 분야 또한 글로벌 고령화에 따른 임플란트 및 수복 치료 수요 증가로 안정적인 성장이 기대된다.

에 주요국은 오리지널 의약품과 효능·안전성이 동등하면서 구조·기능이 유사(similar)한 경우 시장 진입을 협용하는 별도의 절차를 마련했다.

10) 셀트리온은 「렘시마」로 2013년 세계 최초로 유럽에서 바이오시밀러 승인을 받았으며, 2016년에는 세계 최대 제약 시장인 미국에서도 승인을 획득했다.

11) 삼성바이오로직스는 2013년 첫 위탁생산 계약을 수주한 이래 빠르게 성장하여, 현재 CDMO 매출 규모와 생산 능력에서 글로벌 상위권을 차지하고 있다. 주요 CDMO 기업의 매출 규모 및 생산 능력은 다음과 같다.

(만리터, 백만달러)

구 분	순 위
▶ CDMO 부문 매출(24년) ① Lonza 4,174, ② WuXi Biologics 2,595, ③ 삼성바이오 2,207, ④ Catalent 1,952	
▶ 생산능력 ① 삼성바이오 78, ② Lonza 78, ③ WuXi Biologics 58, ④ Boehringer Ingelheim 49	

자료: 각 사 홈페이지 자료 취합

12) 면역물품(immunological products)은 감염성 질환에 대한 면역력을 제공하거나 면역 상태를 진단하는 데 사용되는 의약품으로, 항체의약품 등 바이오의약품을 포함한다.

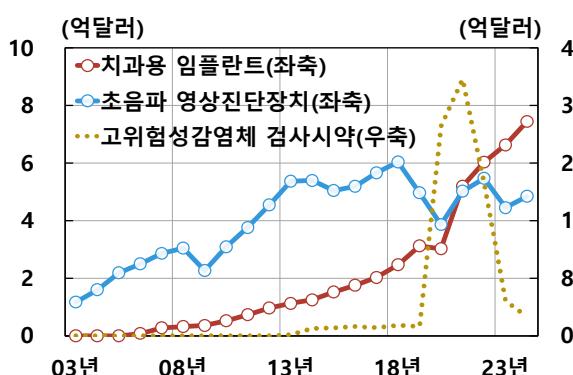
13) 최근 미국 정부는 「Lowering Drug Prices by Once Again Putting Americans First」(25.4월)와 「Delivering Most-Favored-Nation Prescription Drug Pricing to American Patients」(25.5월) 등의 정책을 통해 제네릭·바이오시밀러 개발 장려, 저비용 처방약 수입 확대, 유통시장 투명성 강화 등 약가 부담 완화 대책을 시행했다.

14) 일례로 세계 매출 1위 의약품인 「Keytruda」의 미국 특허는 2028년에 만료된다.

15) 2025년 12월 미 의회는 중국 등 적대국 기업의 미국 바이오 시장 진입을 선별적으로 제한하는 「Biosecure Act」을 통과시켰다.

최근 치과용 임플란트가 의료기기 수출을 주도

[그림 2.7] 주요 의료기기 품목의 수출



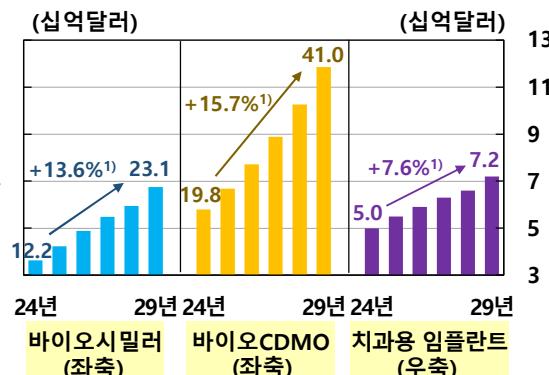
주: 1) 초음파영상 진단장치는 식약처 분류번호 A26380

2) 고위험성감염체 검사시약은 K05030.01 및 N05030.01

자료: 식약처

주력 바이오헬스 품목의 수요는 높은 증가세 전망

[그림 2.8] 주력 바이오헬스 품목의 글로벌 매출 전망



주: 1) 연평균 증가율(CAGR) 기준

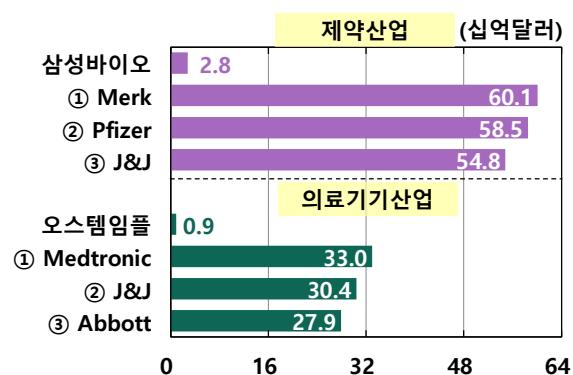
자료: market.us(24.4월, 25.4월, 24.10월)

## ▶ (한계) 원천기술 부족으로 선도국과의 격차 여전

14. 앞서 살펴본 것처럼 우리나라 바이오헬스 산업은 외형적 성장을 이루었지만 글로벌 선도국과의 격차는 여전히 크다. 2023년 기준, 제약 및 의료기기 양대 분야에서 세계 매출 상위 30위권에 이름을 올린 국내 기업은 전무하다. 반면 미국(제약 12개, 의료기기 19개), 일본(5개, 2개), 독일(각 3개) 등 주요 선진국은 상위권을 휩쓸며 시장 지배력을 과시하고 있다. 개별 기업의 매출 규모에서도 차이는 크다. 2023년 기준 국내 1위 제약 기업인 삼성바이오로직스(28억 달러)와 국내 1위 의료기기 기업인 오스템임플란트(9억 달러)의 매출은 글로벌 상위 3개사 평균(제약 578억 달러, 의료기기 304억 달러)의 각각 5%, 3% 수준에 불과했다(그림 2.9).

글로벌 상위기업과의 매출 격차가 뚜렷

[그림 2.9] 글로벌 바이오헬스 기업의 매출규모(23년)



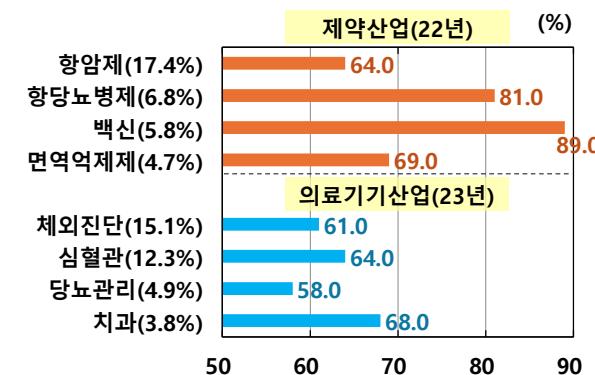
주: 1) 기업명 앞의 숫자는 글로벌 매출 순위를 의미

2) 매출규모는 각 기업의 제약·의료기기 사업부문으로 한정

자료: drugdiscoverytrends.com, meddionline.com

제약·의료기기 산업의 시장집중도는 매우 높음

[그림 2.10] 분야별 글로벌 상위 5개 기업의 시장점유율



주: 1) 매출액을 기준으로 상위 5개 기업을 선정

2) ( ) 내는 24년 기준 각 산업 내 매출비중

자료: Statista(25.12월, 25.6월)

**15. 이러한 격차의 근본 원인은 특허 등을 통한 고수익 창출의 핵심인 원천기술 부족에 있다.** 바이오헬스 산업은 생명과 직결되는 특성상 엄격한 규제가 적용되어 신제품 개발에 막대한 시간과 비용이 소요된다. 반면 인허가 과정에서 정보 공개가 불가피해 기술 기밀 유지가 어렵다는 특성이 있다. 이에 주요국은 기업의 R&D 투자를 유도하기 위해 특허<sup>16)</sup>나 자료보호제도<sup>17)</sup>를 통해 일정 기간 시장 독점권을 부여하고 있다. 이에 따라 치료 분야별로 상위 5개 기업이 전 세계 매출의 절반 이상을 점유하는 소수 혁신기업 중심의 ‘승자독식’ 구조가 고착화되어 있다(그림 2.10). 단적인 예로, 2024년 세계 매출 1위 의약품인 ‘Keytruda’의 단일 품목 매출은 295억 달러에 달해, 우리나라 바이오헬스 산업 전체 수출액(145억 달러)의 두 배를 상회했다<sup>18)</sup>. 반면 바이오시밀러 등 복제약은 가격 경쟁 심화로 수익성이 낮다. 실제로 2024년 글로벌 의약품 시장에서 복제약 매출 비중은 26%에 불과했다<sup>19)</sup>. 특히 미국 시장의 경우 2023년 복제약 처방 비중이 89.9%에 달했으나 금액 기준 점유율은 13.1%에 그쳤다(Association for Accessible Medicines 2024).

**16. 실제 우리나라 바이오헬스 산업의 원천기술 경쟁력은 주요 선도국에 비해 열위에 있다.** 원천기술의 양적 지표인 미국 내 바이오헬스 특허출원건수를 살펴보면, 우리나라는 2016~2023년 중 4.8천 건으로 세계 9위에 머물렀다<sup>20)</sup>. 이는 인구 규모<sup>21)</sup>가 훨씬 적은 스위스(8.7천 건)나 후발 주자인 중국(5.9천 건)보다도 낮은 수준이다(그림 2.11). 또한 질적 수준도 반영하는 ‘2024년 보건의료·산업 기술수준조사’에 따르면, 우리나라의 의약품 개발 기술은 최고 기술 보유국인 미국 대비 3.6년, 의료 기기 개발 기술은 2.8년 뒤처져 있으며, EU 및 일본과의 격차도 여전히 상존했다(그림 2.12). 특히 중국이 기술 격차를 빠르게 좁혀오는 동안, 우리나라는 상대적으로

16) 우리나라와 미국 등은 의약품에 대해 기본 특허 보호 기간인 20년에 더해 최대 5년의 추가 보호 기간을 인정한다. 이는 특허 설정 이후 허가·승인 절차로 실제 판매 기간이 단축되는 점을 보완하기 위한 조치이다.

17) 자료 독점권(data exclusivity)은 기업이 의약품 승인 신청 시 제출한 비임상·임상 자료를 일정 기간 독점적으로 보호해, 해당 기간 동안 복제약의 허가를 제한하는 제도이다. 이는 특허와는 별개의 보호 수단으로, 미국 FDA는 합성의약품에 5년, 바이오의약품에 12년의 자료 독점권을 부여한다.

18) 2024년 기준 주요 의약품의 글로벌 매출은 다음과 같다.

(억달러)					
의약품	① Keytruda	② Eliquis	③ Ozempic	④ Dupixent	⑤ Biktarvy
제조사	Merck & Co.	BMS	Novo Nordisk	Sanofi(프랑스)	Gilead
(소재국)	(미국)	(미국)	(덴마크)	Regeneron(미국)	(미국)
매출액	295	207	175	141	134

주: 의약품명 앞의 숫자는 매출 순위를 의미,

자료: drugdiscoverytrends.com

19) 2024년 글로벌 의약품 매출액은 1조 7,842억 달러이며, 이 중 제네릭은 4,454억 달러, 바이오시밀러는 122억 달러를 차지했다(자료: market.us). 한편 <그림 2.1> 등에서 제시된 글로벌 제약 매출액은 제조사 출고가 기준으로 산정되어, 도매상, 약국 등 유통업체 수익이 제외된 점에서 의약품 매출액보다 규모가 작다.

20) 반면 전체 기술 분야 기준으로 2016~2023년 우리나라의 특허출원건수는 세계 4위를 기록했다.

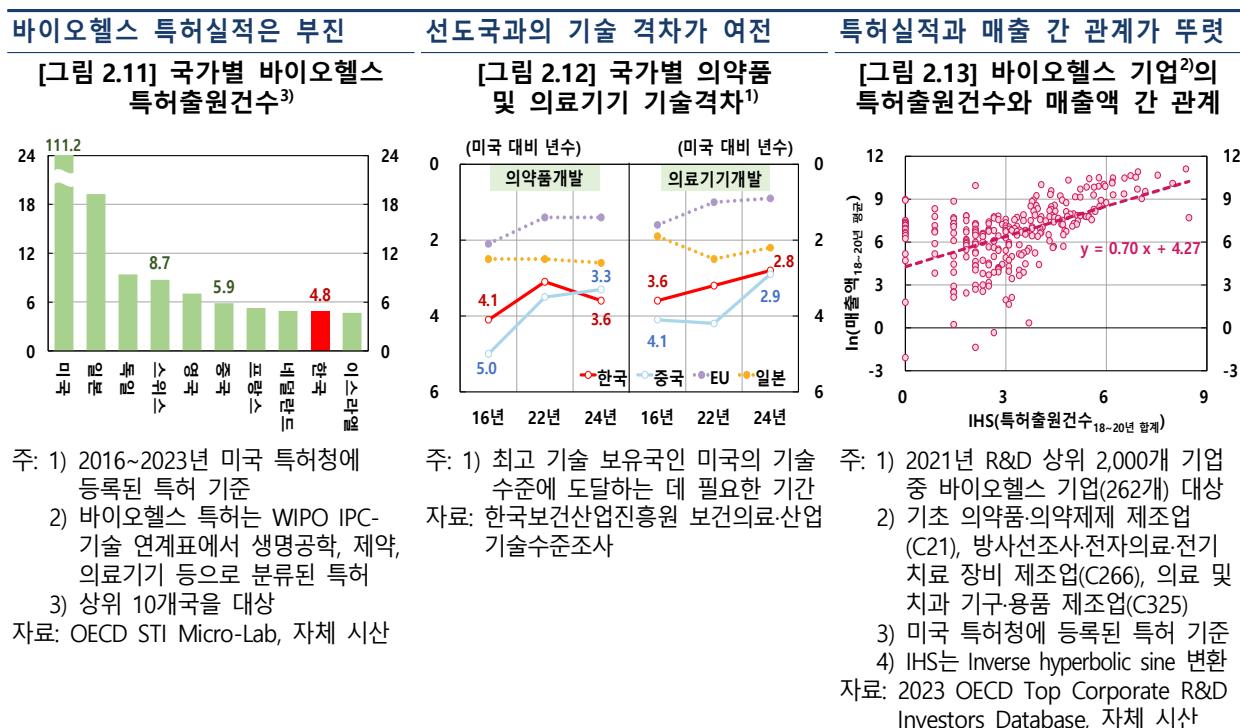
(천건)										
국 가	① 미국	② 일본	③ 중국	④ 한국	⑤ 독일	⑥ 대만	⑦ 프랑스	⑧ 영국	⑨ 캐나다	⑩ 스위스
특허출원건수	1,363	456	215	212	136	104	51	49	43	43

주: 1) 미국 특허청 기준, 2) 국가명 앞의 숫자는 전체 분야 특허출원건수 기준 순위를 의미

자료: OECD STI Micro-Lab, 자체 시산

21) 2023년 기준 스위스의 인구는 0.9천만 명으로, 우리나라 인구(5.2천만 명)의 약 17% 수준이다.

더딘 개선에 그친 상황이다. 원천기술 확보와 상업적 성과 간의 연관성은 뚜렷한데, 글로벌 바이오헬스 기업의 특허출원건수가 10% 증가할수록 매출액은 약 7% 늘어나는 것으로 분석된다(그림 2.13). 이는 국내 기업의 낮은 글로벌 시장 점유율이 주로 원천기술의 열위에서 기인함을 시사한다.



17. 이러한 원천기술의 부족은 자금 공급, 기초연구, 전문인력 양성 등 바이오헬스 산업의 혁신 생태계 전반이 구조적으로 미흡한 데에서 비롯된다<sup>22)</sup>. 바이오헬스 산업은 매출 대비 연구개발비 비중이 2023년 기준 13.3%에 달해 ICT 기기(8.2%)나 자동차(4.8%)보다 월등히 높은 대표적인 기술집약 산업인 만큼, 혁신 생태계의 뒷받침이 필수적이다<sup>23)</sup>. 그러나 현재 자금공급, 기초연구, 전문인력 등 생태계의 핵심 요소들은 모두 미흡한 실정이다.

- (자금공급) 혁신 신약 개발은 10년 이상의 장기간에 걸친 안정적이고 대규모의 투자를 필요로 하며, 자금 조달의 어려움이 R&D 투자의 주요 결림돌로 작용하

22) 바이오헬스 산업을 대상으로 국가패널분석을 실시한 결과, 혁신 생태계를 대변하는 지표인 ① 기업 R&D 지출, ② 피인용 상위 10% 우수 논문 수, ③ 연구개발인력 규모는 기술 역량(특허출원건수) 및 상업적 성과(매출)에 모두 통계적으로 유의한 정(+)의 영향을 미치는 것으로 나타났다. 세부 내용은 <별첨 1>을 참조하기 바란다.

23) EU가 선정한 R&D 상위 2,000개 기업을 대상으로 2023년 R&D 집약도(연구개발비 ÷ 매출액)를 분석한 결과, 바이오헬스 산업은 「Industry Classification Benchmark(ICB) level 3」 기준 산업 중 가장 높은 수준을 기록했다.

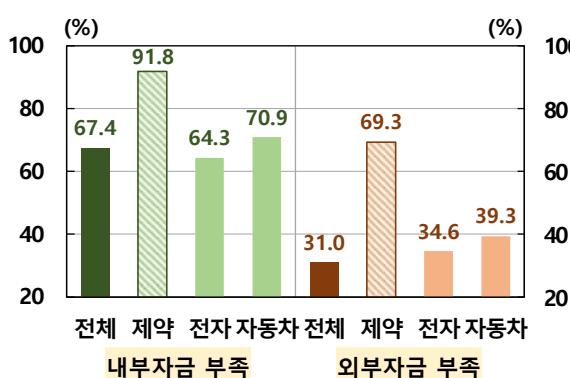
산 업	바이오 헬스	ICT 소프트웨어	ICT 기기	자동차	항공우주 및 방위	화 학	(%)
R&D 집약도	13.3	10.9	8.2	4.8	4.1	2.4	5.1

자료: EU Industrial R&D Investment Scoreboard, 자체 시산

고 있다. '한국기업혁신조사'에 따르면, 2023년 제약 기업의 91.8%가 '내부 자금 부족'을, 69.3%가 '외부 자금 부족'을 혁신 저해 요인으로 꼽아 전체 제조업 평균(각각 67.4%, 31.0%)을 크게 상회하였다(그림 2.14). 이러한 자금 부족으로 인해 2023년 국내 바이오헬스 기업의 총 R&D 지출액은 27억 달러에 그쳐, 미국(1,482 억 달러), 스위스(264억 달러) 등과 현격한 격차를 보였다<sup>24)</sup>(그림 2.15). 이는 글로벌 제약사 Merck & Co. 1개 기업의 R&D 투자액(305억 달러)에도 크게 미치지 못하는 규모다<sup>25)</sup>.

자금부족은 제약기업 R&D 투자를 저해

[그림 2.14] 주요 제조업별 혁신활동 저해 요인 응답비중(23년)

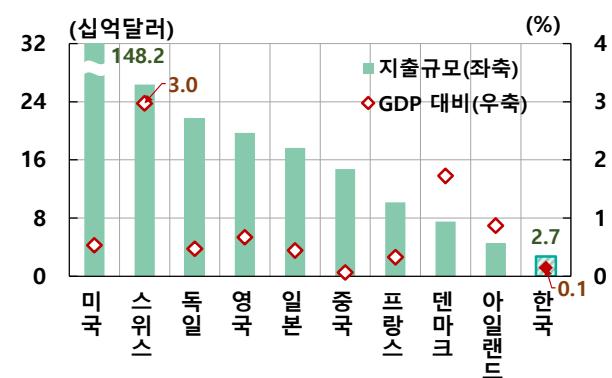


주: 1) 제약은 의료용 물질 및 의약품 제조업을, 전자는 전자 부품, 컴퓨터, 영상, 음향 및 통신장비 제조업, 자동차는 자동차 및 트레일러 제조업을 각각 의미

자료: 한국기업혁신조사

한국 바이오헬스 기업의 R&D 투자가 부진

[그림 2.15] 국가별 전체 바이오헬스 기업의 R&D 지출액(23년)



주: 1) 한국은 연구개발활동조사에서 의료용 물질 및 의약품 제조업과 의료용 기기 제조업 기업을 기준  
2) 다른 국가는 2024년 EU가 선정한 R&D 상위 2,000개 기업 중 Health industries 기업(총 439개)의 수치이며, 상위 9개국을 선정

자료: 국가과학기술지식정보서비스, EU Industrial R&D Investment Scoreboard, 자체 시산

- **(기초연구)** 혁신적인 진단 및 치료법 개발을 위해서는 질병의 근본 원리를 규명하는 기초연구<sup>26)</sup>가 필수적이다(Toole 2007, Blume-Kohout 2012). 그러나 기초연구의 질적 수준을 나타내는 피인용 상위 10% 우수 논문 수에서 우리나라 2016 ~2023년 누적 1.1만 건에 머물러, 미국(15.7만 건), 중국(7.6만 건) 등을 크게 하

24) 여타 국가의 수치는 2024년 「EU Industrial R&D Investment Scoreboard」에 포함된 R&D 상위 2,000개 기업 중 「Health industries」 분야 기업(총 439개)을 대상으로 산출하였다. 해당 Scoreboard는 전 세계 민간 R&D 지출의 80~95%를 포괄하여 높은 대표성을 갖는 것으로 평가된다(Nindl et al. 2024).

25) 2023년 주요 글로벌 제약 기업의 R&D 지출 규모는 다음과 같다.

기업명 (소재국가)	Merck & Co. (미국)	Roche (스위스)	Johnson & Johnson (미국)	Novartis (스위스)	AstraZeneca (영국)
R&D 지출	305	147	120	137	109

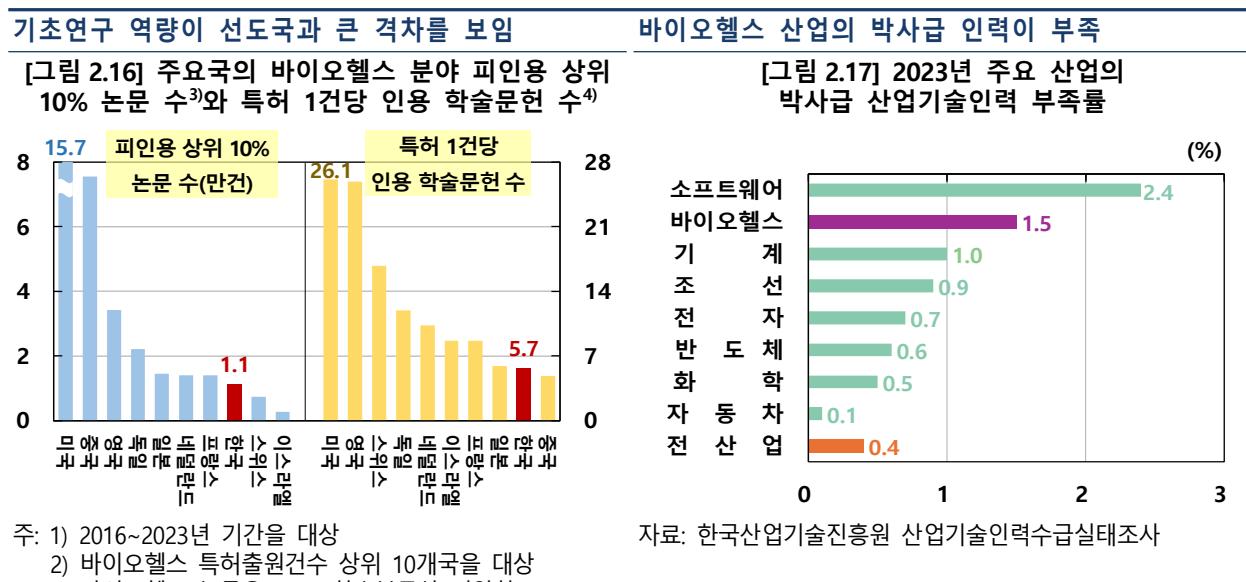
주: 제약부문 R&D에 한정,

자료: drugdiscoverytrends.com

26) 기업의 연구개발활동은 크게 기초 · 응용 · 개발연구로 구분된다. 이 중 기초연구는 특정 상업적 목적 없이 자연 현상에 대한 이해 증진과 새로운 지식 창출을 목표로 한다(과학기술정보통신부 연구개발활동조사, OECD Frascati Manual).

회했다(그림 2.16). 또한 기초연구 성과의 산업적 파급력을 보여주는 특허 1건당 인용 학술문헌 수도 5.7건으로, 특허출원건수 상위 10개국 중 중국(4.8건) 다음으로 낮은 수준이다. 이는 국내 기업의 연구개발이 기초연구에 기반한 파괴적 혁신보다는 기존 제품의 점진적 개선에 치중해 있음을 시사한다.

- (전문인력) 혁신적 R&D를 주도할 전문인력의 확보 또한 원활하지 않다. '산업기술인력수급실태조사'에 따르면, 2023년 바이오헬스 산업의 박사급 인력 부족률은 1.5%로 소프트웨어 산업(2.4%)에 이어 두 번째로 높았으며, 전체 산업 평균(0.4%)을 크게 상회하였다(그림 2.17). 전체 학력을 포함한 바이오헬스 산업 인력 부족률 역시 3.3%로 전 산업 평균(2.2%)보다 높았다. 아울러 국가별 민간 기업 연구개발인력 중 바이오헬스 산업 비중을 살펴보면, 우리나라 4.2%로 OECD 33개국 평균(5.1%)에 미치지 못하는 17위 수준에 그쳤다.



주: 1) 2016~2023년 기간을 대상  
2) 바이오헬스 특허출원건수 상위 10개국을 대상  
3) 바이오헬스 논문은 ASJC 학술분류상 의약학 분야(대분류코드 27, 30, 35, 36)에 해당하는 논문  
4) 바이오헬스 특허는 WIPO IPC-기술 연계표에서 생명공학, 제약, 의료기기 등 기술로 분류된 미국 특허청 등록 특허를 의미

자료: OECD Bibliometric Indicators, OECD STI Micro-Lab

자료: 한국산업기술진흥원 산업기술인력수급실태조사

18. 이와 같이 미흡한 혁신 생태계를 감안할 때, 기존의 '빠른 추격자' 전략에서 벗어나 새로운 도약의 기회를 모색하는 전략적 전환이 시급하다. 틈새시장을 공략하며 선도국을 단순히 추격하는 기존 방식에 머무른다면, 원천기술 등 혁신 역량의 격차를 단기간 내에 해소하기는 요원할 것이다.

### 3. 바이오헬스 산업의 새로운 전환점: AI 시대와 바이오 데이터

19. 최근 AI 기술이 바이오헬스 산업의 경쟁 구도를 재편하는 '게임체인저'로 부상함에 따라, 우리나라는 우수한 바이오 데이터 수집·연계 인프라를 발판으로 글로벌 바이오헬스 R&D 허브로 도약할 새로운 기회를 맞이하고 있다.

#### ■ AI 혁신: 바이오헬스 산업의 「게임체인저」

20. 최근 AI 혁신은 생명공학의 비약적인 발전을 견인하며, 연구자가 생명 시스템을 직접 프로그래밍하는 '생명 설계' *biology by design*의 시대를 열고 있다. 수십억 년 진화의 산물인 생명체의 복잡성으로 인해 기존에는 단순한 미생물조차 정밀하게 제어하기 어려웠으나, AI는 이러한 한계를 극복하는 결정적인 돌파구를 제시하고 있다. 대표적인 사례인 구글 딥마인드의 'AlphaFold'는 2024년 노벨화학상의 토대가 된 기술로, 방대한 단백질 구조 데이터를 학습하여 수년이 걸리던 수억 개의 단백질 구조 규명을 단 며칠 만에 원자 수준의 정확도로 예측해냈다. 이 같은 AI의 분석력은 2020년 노벨화학상 수상으로 이어진 '유전자 가위' 등 유전체 편집<sup>27)</sup> 기술과 결합하여, 인류 건강 증진을 목적으로 생명체를 설계하고 제어하는 시대를 앞당기고 있다.

AI는 바이오헬스 R&D 패러다임을 전환

[그림 3.1] AI에 따른 바이오헬스 R&D의 변화



자료: NSCEB(2025)

AI는 신약 개발 기간을 30~50% 단축

[그림 3.2] AI에 따른 신약 개발 기간의 단축



주: Saini et al.(2025)를 재구성

21. 이러한 AI와 생명공학의 융합은 바이오헬스 R&D의 패러다임을 기존의 '실험과 관찰' 중심에서 '데이터와 알고리즘' 중심으로 전환하며, 개발 비용과 시간을 획기적으로 절감하고 있다. 과거의 바이오헬스 R&D는 대규모 실험 인프라와 축적된 임상

27) 유전체 편집은 생명체의 유전체에서 특정 DNA 서열을 정밀하게 변경·수정하는 기술로, 유전 질환 치료와 신약 개발 등 다양한 분야에 활용되고 있다.

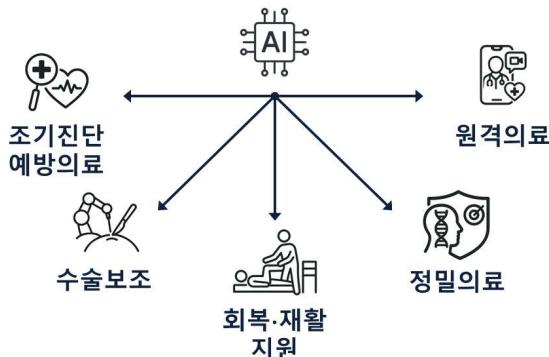
경험에 크게 의존하여 대규모 자금과 인력이 소요되었다. 실제로 신약 개발은 FDA 승인까지 평균 10~15년이 걸리고, 개당 개발 비용도 26억 달러에 달해 우리나라 전체 바이오헬스 기업의 연간 R&D 지출액(2023년 27억 달러)과 맞먹는 수준이었다 (DiMasi et al. 2016, Sun et al. 2022). 그러나 이제는 AI가 고성능 컴퓨팅을 기반으로 방대한 데이터를 분석하고 자율 실험까지 수행하면서, R&D 프로세스가 더욱 빠르고 정밀한 목표 지향적 형태로 혁신되고 있다(그림 3.1). 특히 신약 개발에 AI가 도입되면서 신약 후보 물질 발굴, 임상시험 설계, 임상 대상자 선정 등 전 과정의 효율성이 제고되어 개발 기간과 비용이 약 30~50% 절감될 것으로 전망된다 (Rodriguez et al. 2023)(그림 3.2).

**22. 아울러 AI는 기존의 기술적 한계를 넘어 새로운 바이오헬스 시장을 창출하는 기회의 창이 되고 있다**(그림 3.3). AI가 주도하는 신규 시장들은 연평균 20~30%의 높은 성장세를 보일 것으로 예측된다(그림 3.4).

- **(조기 진단·예방 의료)** AI가 유전체 정보, 웨어러블 기기 데이터, 의료 영상 등을 정밀 분석하여 질병 발생 위험을 조기에 예측함으로써, 의료 패러다임을 '사후 치료' 중심에서 '사전 예방' 중심으로 전환한다.
- **(수술 보조)** AI가 수술 로봇과 결합하여 절개 부위 최적화, 봉합 패턴 조정 등을 자율적으로 수행함으로써 수술의 정밀도를 극대화하고 인적 오류를 최소화한다.
- **(회복·재활 지원)** 뇌졸중이나 부상 환자에게 최적화된 회복 프로그램을 AI가 설계하고, 로봇 보조 장치를 통해 정밀한 재활 훈련을 지원하여 회복 효과를 높인다.
- **(정밀 의료)** 유전 정보, 생활 습관, 병력 등 다차원 데이터를 AI가 종합 분석하여 개인에게 최적화된 치료법을 제시하는 '개인 맞춤형 의료'를 실현한다.
- **(원격 의료)** 웨어러블 기기 데이터를 AI가 실시간 모니터링하여 원격 진료 및 디지털 돌봄 서비스를 구현함으로써 의료 접근성을 대폭 향상시킨다.

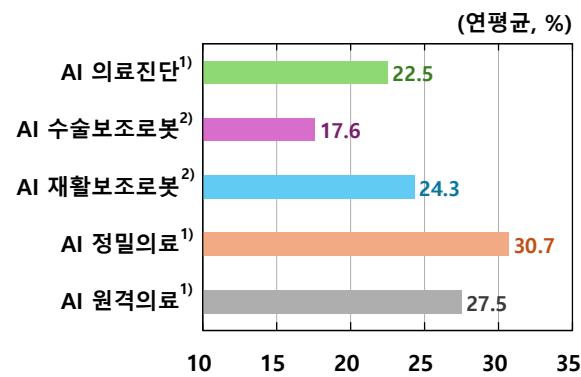
AI는 새로운 바이오헬스 시장을 창출

[그림 3.3] AI가 창출하는 바이오헬스 신규 시장



AI 기반 바이오헬스 신규시장이 고성장을 보일 전망

[그림 3.4] AI 기반 바이오헬스 신규시장의 매출 전망



자료: Markets and Markets, Research and Markets

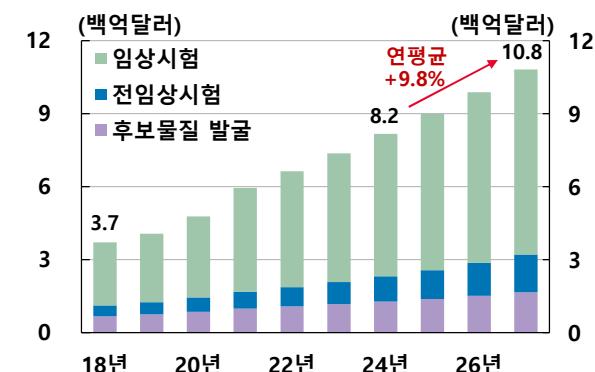
**23. 이러한 환경 변화에 대응하여 글로벌 바이오헬스 선도기업들은 데이터 및 AI 전문 기관 등과의 글로벌 R&D 협력을 확대하고 있다.** 과거 대형 제약사들은 R&D 전 과정을 내부에서 소화하는 폐쇄적 모델을 고수했으나, 신약 개발의 난이도와 비용 급증으로 생산성 저하 위기에 직면했다(Myers·Pauly 2019). 이에 따라 라이선스 계약을 통한 외부 기술 도입으로 신약 후보 물질을 확보하고, 비용이 많이 드는 임상시험은 CRO<sup>28)</sup>Contract Research Organization, 위탁연구기관에 아웃소싱하는 방식으로 변화해 왔다(그림 3.5, 3.6). 최근 AI 기술 발전은 이러한 ‘개방형 혁신’Open Innovation을 더욱 가속화하고 있다. 대형 제약사들은 생물학·화학 중심의 인력 구조로 인해 데이터 및 AI 전문성이 부족하고, AI 모델 학습에 필수적인 유전체 데이터나 실사용 증거<sup>29)</sup>Real World Evidence, RWE 확보에 어려움을 겪고 있다. 이러한 한계를 극복하기 위해 바이오 벤처, 전문 데이터 기관, AI 스타트업 등과의 파트너십을 강화하며 글로벌 R&D 분야 모델로 빠르게 전환하고 있는 것이다.

**글로벌 바이오헬스 라이선스 거래가 증가**  
[그림 3.5] 글로벌 헬스케어·생명과학 라이선스 거래가치



자료: DealForma

**바이오헬스 CRO 시장이 높은 성장세를 지속**  
[그림 3.6] 글로벌 CRO 매출 실적 및 전망



자료: Frost & Sullivan(22.11월)

**24. 이와 같이 AI 혁신이 바이오헬스 산업의 ‘게임체인저’로 부상함에 따라, 우리나라와 같은 추격국이 선도국을 추월할 수 있는 새로운 기회가 열리고 있다.** AI 기술은 R&D 기간과 비용을 획기적으로 절감함으로써 대규모 실험 인프라가 부족하더라도 글로벌 제약사와 대등한 조건에서 R&D를 추진할 수 있는 환경을 제공한다. 특히 자본력은 제한적이지만 혁신적인 아이디어를 보유한 바이오 벤처와 스타트업이 AI를 활용해 시장의 주역으로 부상할 수 있는 토대가 마련되고 있다. 또한 기존 선도 기업과의 직접적인 정면 대결에서 벗어나, AI가 창출하는 신규 시장을 선점함으

28) CRO(Contract Research Organization)는 제약사 및 바이오헬스 기업을 대상으로 전임상, 임상시험, 인허가 지원 등 연구개발 전 과정을 대행·지원하는 전문기관으로, 기업들은 비용 절감과 내부 역량 보완을 위해 CRO 활용을 확대하고 있다.

29) 실사용 증거는 통제된 임상시험의 아닌 실제 진료 현장에서 수집된 환자 데이터(진료 기록, 보험 청구 내역, 웨어러블 기기 데이터 등)를 활용해 의약품·의료기기의 효능과 안전성을 입증하는 자료를 말한다. 최근 FDA 등 주요 규제기관은 신약 허가 및 적응증 확대 심사에 실사용 증거를 적극 활용하고 있다.

로써 글로벌 리더로 도약하는 전략적 선택도 가능해졌다. 나아가 개방형 혁신 기조에 맞춰 대규모 바이오 데이터를 전략 자산으로 구축하고 AI 기반 R&D 생태계를 조성한다면, 글로벌 자본과 인재가 모여드는 매력적인 ‘글로벌 바이오헬스 R&D 허브’로 거듭날 수 있다.

**25. 이처럼 AI 혁신은 산업 판도를 뒤바꿀 전례 없는 기회인 동시에, 실기할 경우 기술 종속국으로 전락할 수 있다는 엄중한 경고이기도 하다.** 이미 주요국들은 바이오헬스 산업의 글로벌 주도권을 차지하기 위한 총력전에 돌입했다. 미국은 ‘National Biotechnology & Biomanufacturing Initiative’(22.9월) 등을 통해 자국 중심의 공급망 재편과 데이터 주권 강화를 가속화하고 있으며, 중국 또한 ‘제14차 5개년 계획’(20.10월)을 통해 바이오 경제를 국가 전략 산업으로 지정하며 기술 굴기를 추진 중이다. 이러한 글로벌 속도전 속에서 지금의 기회를 놓친다면 선도국과의 기술 격차는 회복 불가능한 수준으로 고착화될 위험이 크다. 따라서 AI 혁신이라는 거대한 전환점을 적극적으로 활용하는 것은 단순한 선택의 문제가 아니라, 국가의 미래 성장 경로를 결정짓는 필수 생존 전략이다.

## ■ 바이오 데이터의 중요성: 「AI 시대의 다이아몬드」로 부상

**26. AI 혁신이 제공하는 새로운 기회를 실현하기 위한 핵심 요소는 AI 기술의 원료라 할 수 있는 대규모 데이터이다.** 최근 AI가 비약적으로 발전할 수 있었던 결정적 계기는 인터넷의 등장으로 방대한 데이터를 확보할 수 있게 된 데 있다. 사실 딥러닝의 중심인 ‘역전파 알고리즘’<sup>30)</sup> 등 AI의 기본 이론은 이미 1980년대에 정립되었다. 하지만 이 알고리즘의 선구자인 Hinton(2024년 노벨물리학상)이 회고했듯이, 당시에는 라벨링된 데이터셋의 규모가 현재보다 수천 배 작았고, 컴퓨팅 성능 역시 수백만 배 느려 이론을 실제로 구현할 기술적 환경이 미비했다(Christian, 2020). 그러나 2000년대 들어 고성능 GPU 그래픽처리장치가 보급되고 인터넷의 성장으로 방대한 데이터가 생성되면서, ChatGPT와 같은 최신 모델이 탄생하고 오늘날 우리가 체감하는 놀라운 성능을 보여주게 되었다.

**27. 특히 생명과 직결되는 바이오헬스 분야에서는 데이터의 양뿐만 아니라 대표성을 갖춘 고품질 데이터의 확보가 AI 모델의 안전성과 신뢰성을 결정짓는 중요 변수다.** 다양한 인구 집단과 희귀 사례를 포괄하는 데이터셋은 AI 모델이 특정 조건에만 최적화되는 편향을 방지하고 통계적으로 강건한 예측 성능을 확보하는 기반이 된다. 반대로 학습 데이터가 특정 집단에 편중되거나 품질이 낮을 경우, AI의 효용성은 급격히 저하되며 잘못된 판단으로 생명을 위협하는 심각한 결과를 초래할 수

30) 역전파(backpropagation) 알고리즘은 인공신경망의 예측값과 실제 정답 간 오차를 바탕으로, 신경망의 끝(출력층)에서부터 처음(입력층)으로 오차의 원인을 역추적하며 가중치를 조정하는 딥러닝의 학습 메커니즘이다.

있다. 실제로 2019년 구글 딥마인드가 발표한 급성 신장 손상 예측 AI는 미국 재향 군인부 환자 70만여 명의 데이터를 학습했으나, 데이터의 94%가 남성으로 구성되어 여성 환자에 대해서는 예측 정확도가 현저히 낮아지는 한계를 보였다(Cao et al. 2022). 이 사례는 보편적으로 적용 가능한 신뢰도 높은 바이오헬스 AI를 개발하기 위해 전체 인구를 대표하는 고품질 데이터 확보가 필수적임을 보여준다.

**28. 그러나 전 세계적으로 고품질 바이오 데이터에 대한 접근성 부족은 바이오헬스 AI 개발의 핵심 장애물로 지적되고 있다.** 현대 의료 환경에서는 임상 데이터, 건강보험 청구 내역, 개인 생성 생활 데이터, 공공 연구 데이터 등 방대하고 다양한 바이오 데이터가 생성되고 있다(그림 3.7). 그럼에도 불구하고 북유럽 국가나 이스라엘 등 일부 선도 사례를 제외하면, 이러한 데이터를 서로 연계하여 통합적으로 활용하기는 매우 어려운 설정이다(Sudlow 2024). 실제 제약업계 전문가 설문조사에서도 AI 도입의 가장 큰 걸림돌로 '고품질 데이터 부족'이 가장 많이 꼽혔다(그림 3.8). 나아가 병원에서 생성되는 데이터의 97%가 활용되지 못한 채 사장된다는 분석도 있다(WEF 2019).

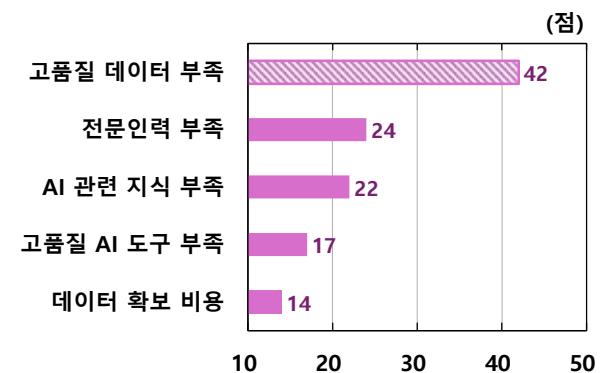
다양하고 방대한 바이오 데이터가 생산

[그림 3.7] 생산기관별 바이오 데이터 종류



AI 도입의 가장 큰 장벽은 고품질 데이터 부족

[그림 3.8] 신약개발 AI의 도입 장벽에 관한 설문 결과



주: 1) "신약 개발 분야에서 AI 도구의 활용을 확대하는 데 있어 현재의 장벽은 무엇입니까?"에 대한 응답 결과  
 2) 학계, 산업계, 투자기관 등 선진국 관계자 66명을 대상  
 3) 응답자는 최대 3개까지 선택 가능

자료: Wellcome Trust & BCG(2023)

**29. 이처럼 대표성을 갖춘 대규모 바이오 데이터는 단순한 보건 정보를 넘어, 막대한 부가가치를 창출하는 'AI 시대의 다이아몬드'로 부상한 상황이다.** 이러한 데이터는 AI 개발을 위한 핵심 투입 요소인 동시에 그 자체로 높은 가치를 지닌 전략적 자산이다. 따라서 바이오 데이터의 안전하고 원활한 활용을 뒷받침하는 제도적·기술적 환경을 조성하는 것은 미래 시장의 주도권을 확보하기 위한 필수 경쟁력이 될 것이다.

## ■ 한국의 잠재력: 우수한 데이터 수집 · 연계 인프라

### 30. 우리나라를 전 국민의 바이오 데이터를 체계적으로 수집하고 연계할 수 있는 세계 최고 수준의 인프라를 보유하고 있다.

- **(수집)** 단일 건강보험 제도를 기반으로 모든 국민의 진료 정보, 투약 이력, 보험 청구 내역 등이 중앙 관리 체계하에 통합적으로 수집되고 있다. 또한 우수한 의료 접근성과 전 국민 건강검진 제도, 그리고 전면 보급된 전자의무기록(EMR)<sup>31)</sup>은 의료 현장의 방대한 임상 데이터 축적을 뒷받침한다(표 3.1). 이러한 역량을 인정 받아 우리나라는 OECD(Oderkirk 2021)로부터 데이터 가용성 부문 세계 최상위권 국가로 평가받았으며, ‘핵심 공공 보건 데이터셋’<sup>32)</sup>의 개선 주기와 대표성 측면에서도 글로벌 선도 수준을 나타냈다(그림 3.9).
- **(연계)** 주민등록번호라는 단일 식별체계를 통해 서로 다른 데이터 소스를 정밀하게 결합함으로써, 개별 국민의 평생 의료 기록을 추적하는 ‘종단적 데이터’ 구축이 가능하다. OECD(Oderkirk 2021)에 따르면, ‘핵심 공공 보건 데이터셋’ 전체에 단일 식별자를 사용하는 국가는 전 세계에서 우리나라, 덴마크, 핀란드 등 단 3개국에 불과하다(그림 3.9).

전 국민의 바이오 데이터가 체계적으로 수집

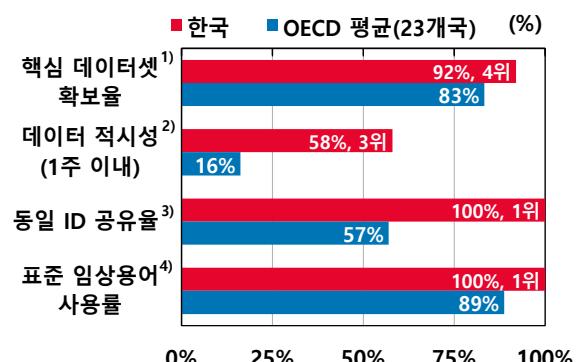
[표 3.1] 우리나라 바이오 데이터 현황

기 관	데이터
▶ 건강보험 심사평가원	- 진료 · 투약, 의약품 유통
▶ 국민건강 보험공단	- 보험 자격 · 청구, 진료 · 투약, 건강검진 결과, 장기요양
▶ 질병관리청	- 예방접종, 감염병, 국민건강영양조사, 한국인유전체역학조사
▶ 국립암센터	- 암 등록 정보
▶ 국가데이터처	- 사망 통계
▶ 의료기관 (임상)	- 전자의무기록, 의료영상, 문진 · 설문 내역
▶ 기업 · 개인	- 의약품 정보, 웨어러블 기기 데이터

자료: 보건복지부(2021)

우리나라의 바이오 데이터 수집 · 연계 인프라는 우수

[그림 3.9] 의료 데이터의 수집 및 연계 역량 평가



주: 1) 13개 핵심 공공 보건 데이터셋 중 실제 활용 가능한 데이터셋이 차지하는 비율

2) 활용 가능한 핵심 공공 데이터셋 중 기록 생성 후 1주일 이내에 반영되는 비율

3) 활용 가능한 핵심 공공 데이터셋 중 동일한 환자 식별자를 공유하는 비율

4) 활용 가능한 핵심 공공 데이터셋 중 표준화된 임상용어를 사용하는 비율

자료: Oderkirk(2021)

31) 2021년 기준 우리나라의 전자의무기록 보급률은 약 85%였다(Slawomirski et al. 2023).

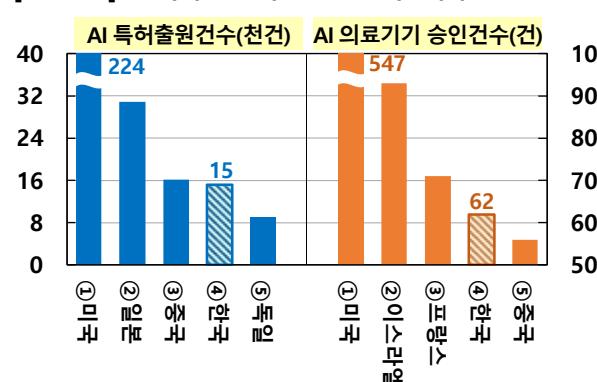
32) 공공 보건 데이터란 공공기관이 체계적으로 관리하는 보건 정보를 의미한다. 핵심 분야는 총 13개로 다음과 같다: 병원 입원환자 데이터, 정신병원 입원환자 데이터, 응급의료 데이터, 일차의료 데이터, 처방 의약품 데이터, 암 등록 데이터, 당뇨병 등록 데이터, 심혈관 질환 등록 데이터, 사망률 데이터, 공공 장기요양 데이터, 환자 경험 설문 데이터, 국민 건강 설문 데이터, 인구 총조사 또는 등록 데이터. 우리나라는 이 중 심혈관 질환 등록 데이터를 제외한 12개 분야의 데이터를 보유하고 있다.

**31. 이러한 인프라를 통해 고품질 바이오 데이터를 구축할 수 있다면, 이는 세계적으로 희소한 국가 단위의 대규모 데이터 자산이 될 것이다.** 바이오 데이터 강국인 북유럽 국가나 이스라엘의 인구 규모가 최대 1천만 명<sup>33)</sup> 수준인 반면, 우리나라는 5천만 명의 인구를 바탕으로 타 국가 대비 대규모 데이터셋을 구축할 잠재력을 보유하고 있다. 이는 데이터의 대표성 측면에서 통계적 신뢰도를 제고하는 독보적인 질적 경쟁우위를 제공한다. 특히 급성장하는 아시아 바이오헬스 시장의 흐름에 맞춰 아시아 인종 데이터에 대한 글로벌 수요가 크게 증가하고 있는 점은 우리나라 바이오 데이터의 전략적 가치를 더욱 높이는 요인이다.

**32. 우수한 데이터 수집·연계 인프라에 더해, 우리나라는 AI 기술력, 임상시험 역량, 그리고 유망한 시장 환경 등 글로벌 R&D 협력을 이끌어낼 여러 강점도 보유하고 있다.** 우리나라는 AI 분야 미국 내 특허출원건수(2016~2023년)와 미 FDA 승인 AI 의료기기 승인건수(2016.1월~2025.5월)에서 모두 세계 4위를 기록하며, 글로벌 수준의 AI 기술력을 입증하고 있다(그림 3.10). 또한 서울은 세계 2위 규모의 글로벌 임상시험 유치 도시로, 고도로 숙련된 의료 인력과 높은 환자 참여도를 바탕으로 우수한 임상시험 역량을 갖추고 있다(그림 3.11). 아울러 급속한 고령화에 따른 내수 시장 확대 역시 중요한 동력이다. Acemoglu·Linn(2004)에 따르면 고령화로 의약품 시장 규모가 1% 확대될 때 신약 개발 건수가 4~6% 증가하여, 시장 규모의 확대가 기술 혁신을 촉진하는 촉매제인 것으로 분석된다. 이는 내수 시장의 성장이 바이오헬스 기술 혁신의 중요한 기반이 될 수 있음을 시사한다.

AI 기술 분야에서 경쟁력을 유지

[그림 3.10] AI 특허출원건수<sup>12)</sup> 및 AI 의료기기 FDA 승인건수<sup>3)</sup>



주: 1) 2016~2023년 미국 특허청 등록 기준

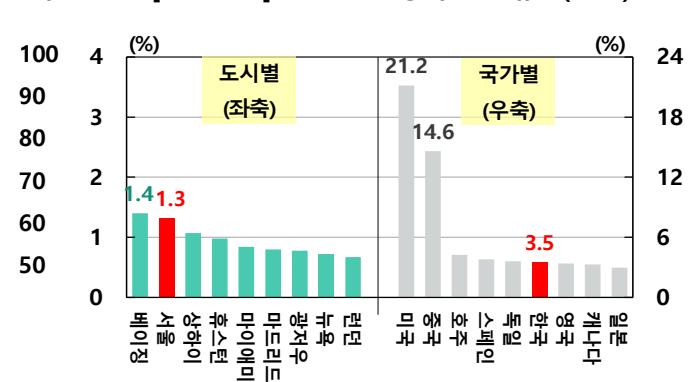
2) AI 특허 식별은 Pairolo et al.(2025)의 분류 기준을 활용

3) 16.1월~25.5월 FDA 승인 기준

자료: Pairolo et al.(2025), OECD STI Micro-Lab, FDA, 자체 시산

서울의 임상시험 점유율이 세계 2위

[그림 3.11] 글로벌 임상시험 점유율(24년)



자료: 국가임상시험지원재단

33) 2023년 기준 덴마크, 핀란드, 노르웨이의 인구는 각각 0.6천만 명이고, 이스라엘은 1.0천만 명이다.

**33. 따라서 우수한 데이터 수집·연계 인프라를 바탕으로 대규모 고품질 바이오 데이터를 구축하고 전략적으로 활용한다면, AI 혁신이 제공하는 추격의 기회를 선점해 글로벌 바이오헬스 R&D 허브로 도약할 수 있을 것이다.** 바이오헬스 산업은 대규모 개발 비용과 복잡한 인허가 규제, 그리고 의료계의 신뢰 장벽 등으로 인해 후발국이 선도국을 단기간에 따라잡기 어려운 분야이다. 그러나 대규모 바이오 데이터 구축을 통해 AI 기반 R&D 생태계를 조성한다면, 글로벌 선도 기업의 대규모 투자를 유치할 수 있다. 이러한 파트너십을 통해 이들 기업의 자본과 경험으로 혁신 신약과 침단 의료기기를 공동 개발하고 다국가 임상시험을 주도함으로써, 기존의 진입 장벽을 넘어 글로벌 바이오헬스 R&D 허브로 자리매김할 수 있을 것이다.

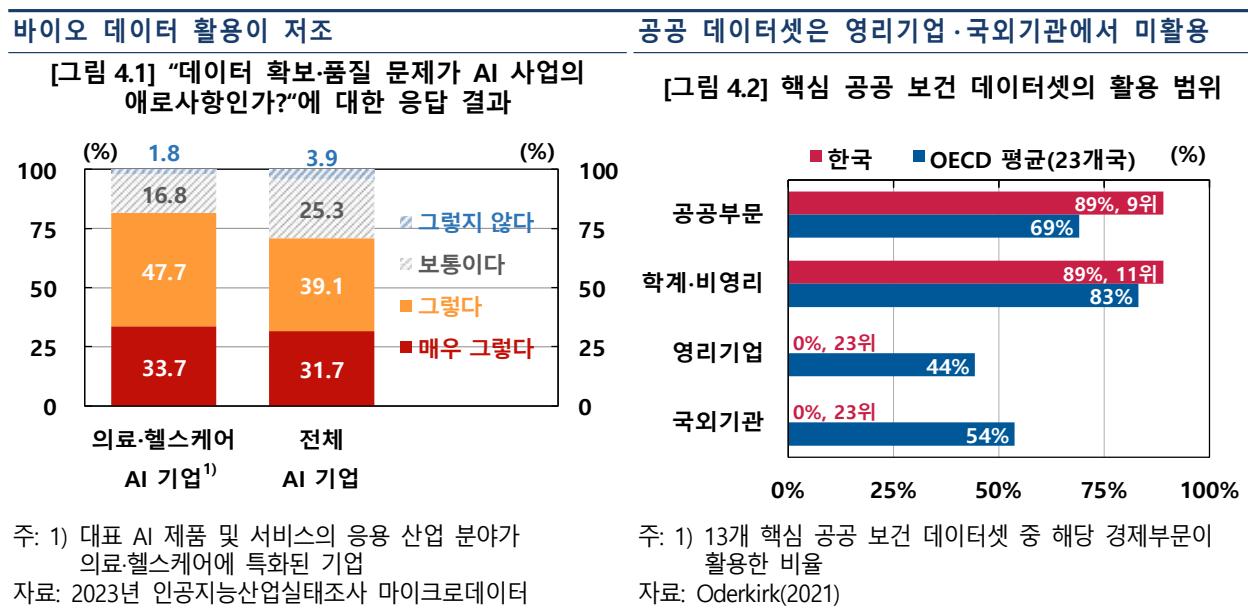
## 4. 바이오헬스 산업 도약의 문제점: 바이오 데이터 활용의 한계

### ■ 국내 바이오 데이터 활용 현황

**34. 우리나라의 우수한 데이터 수집·연계 인프라가 실질적인 가치 창출로 이어지기 위해서는 원시 데이터를 연구 및 산업 현장에서 즉시 활용 가능한 자산으로 전환하는 데이터 정제<sub>curation</sub> 과정과 안전한 활용을 보장하는 공유 체계 구축이 필수적이다.** 의료 데이터는 대부분 진료나 행정 목적으로 생성된 원본 자료이므로, 이를 연구나 AI 개발에 적용하려면 반드시 정제 과정을 거쳐야 한다. 예를 들어, AI 모델이 오진을 학습하는 것을 방지하려면 추정 진단이 아닌 실제 치료 결과를 기준으로 데이터를 라벨링해야 한다. 이 과정은 고도의 의학적 전문성이 요구되며, 전체 프로젝트 기간의 약 80%를 차지할 정도로 많은 시간과 비용이 소요된다(Goldacre·Morley 2022). 또한 데이터는 공유 범위가 넓을수록 혁신을 촉진해 가치가 증대되는 비경합적 특성을 지녔으나, 동시에 민감한 개인정보 유출이라는 사회적 위험도 내포한다. 따라서 데이터의 오남용을 막으면서도 혁신을 가속화할 수 있는 안전한 공유 체계가 병행되어야 한다. 이러한 정제 및 공유 기반의 부재는, 마치 다이아몬드 원석을 채굴할 광산은 풍부하나 이를 고부가가치 보석으로 가공하고 유통할 공장과 시장이 없는 상황과 같다.

**35. 그러나 국내 바이오 데이터의 실질적 활용 수준은 저조하여 정제, 공유 등 활용 단계에 구조적 병목 현상이 존재함을 시사한다.** ‘2023년 인공지능산업 실태조사’에 따르면, 의료·헬스케어 AI 기업의 81.4%가 ‘데이터 확보 및 품질 문제’를 주요 경영 애로사항으로 꼽았다(그림 4.1). 이는 전체 AI 기업 평균인 70.8%를 10%p 이상 상회하는 수치로, 바이오헬스 분야의 데이터 품질 및 접근성 문제가 상대적으로

크게 인식되고 있음을 보여준다. 국가가 관리하는 ‘핵심 공공 보건 데이터셋’ 또한 활용 범위가 제한적이다. OECD에 따르면 우리나라의 ‘핵심 공공 보건 데이터셋’이 영리 기업이나 국외 기관에 활용된 사례는 전무한 것으로 나타났는데, 이는 공공 데이터의 상업적 활용과 국제 공동 연구가 활발한 북유럽 국가나 싱가포르와 뚜렷하게 대조된다(Oderkirk 2021)(그림 4.2). 결론적으로 우리나라는 우수한 데이터 수집·연계 인프라를 보유하고 있음에도 불구하고, 이를 실질적인 가치 창출로 연결하는 활용 단계에서 큰 장애물에 직면해 있다.



## ■ 현행 바이오 데이터 활용 체계

**36. 국내에서 바이오 데이터는 개인정보보호법상 ‘민감정보’로 분류되어 일반 개인정보보다 엄격하게 보호된다.** 질병 이력, 유전자 정보, 신체적 특성 등의 바이오 데이터가 유출될 경우, 보험 가입 거절이나 채용 차별과 같은 경제적 불이익은 물론 사회적 낙인, 신원 도용 등 심각한 사회적 피해로 이어질 수 있다. 이러한 위험성을 고려해 ‘개인정보보호법’은 바이오 데이터 처리<sup>34)</sup> 시 원칙적으로 정보주체의 별도 동의를 요구한다. 아울러 ‘의료법’상의 비밀 유지 의무와 ‘생명윤리 및 안전에 관한 법률’에 따른 인간 대상 연구 규제도 준수해야 하므로, 바이오 데이터를 활용하기 위해서는 복합적인 법적 요건을 동시에 충족해야 하는 구조가 형성되어 있다.

**37. 현행 법령상 바이오 데이터 활용 방식은 크게 ① 정보주체 동의 기반 이용, ② 익명정보 이용, ③ 가명정보 처리 후 이용의 세 가지로 구분된다**(표 4.1). 우선, ‘정

34) 개인정보보호법상 개인정보 처리는 수집, 저장, 이용, 제공, 파기 등 정보주체의 개인정보를 다루는 전 과정을 의미한다.

보주체 동의 기반 이용'이란 개인정보의 주체인 당사자에게 수집·이용 목적과 항목 등을 구체적으로 알리고 직접 동의를 받아 데이터를 활용하는 방식을 의미한다. 다음으로, '익명정보'는 더 이상 개인을 식별할 수 없도록 처리되어 개인정보보호법의 적용을 받지 않는 정보를 말한다. 다만, 개인 식별을 원천 차단하기 위해 정보를 삭제하거나 범주화하는 과정에서 데이터의 가치 및 유용성이 급격한 저하된다는 한계가 있다. 마지막으로, '가명정보'란 추가 정보 없이는 특정 개인을 식별하기 어려운 정도로만 가공하여, 익명처리에 따른 데이터 가치 손실을 보완한 정보를 말한다. 가명정보 제도는 2020년 '데이터 3법'<sup>35)</sup> 개정을 통해 도입되었으며, 이를 통해 의료기관과 기업은 정보주체의 별도 동의 없이도 신약 개발이나 AI 학습 등 과학적 연구(산업적 활용 포함) 등의 목적으로 정보 가치가 어느 정도 보존된 데이터를 활용할 수 있게 되었다<sup>36)</sup>.

#### 사전 동의, 익명 처리, 가명 처리를 통해 AI 개발에 활용

[표 4.1] AI 개발을 위한 바이오 데이터의 활용 경로

	개념	데이터 성격	활용범위	국외반출
	사전 동의	환자에게 직접 동의 얻음	개인정보 (원본)	동의 받은 목적으로만 동의 내용에 포함시 가능
	익명 처리	완전히 식별 불가능하게 가공	개인정보 아님	제한 없음 가능 (개인정보 아님)
	가명 처리	추가 정보 없이는 식별 불가능하게 가공	개인정보 (재식별 가능)	과학적 연구 등 제한된 목적으로만 불가능 (개인정보)

#### 여러 통합 플랫폼 구축이 추진

[표 4.2] 보건의료 데이터 통합 플랫폼

구 분	통합 데이터
▶ 보건의료 빅데이터 플랫폼(복지부)	10개 공공기관 보건의료 데이터
▶ 헬스케어 빅데이터 플랫폼 및 센터 (과기부)	국립암센터 · 10개 민간병원 암종별 임상데이터
▶ K-CURE 라이브러리 (복지부)	18개 병원 임상데이터 및 5개 공공기관 암데이터
▶ K-Meloddy (과기부, 복지부)	신약 개발 데이터 (AI 연합학습 플랫폼)

**38. 다만 가명정보를 활용하더라도 실제 이용 과정에서는 기획 단계부터 데이터 파기까지 엄격한 안전 절차를 준수해야 한다.** 연구자는 우선 '생명윤리 및 안전에 관한 법률'에 따라 기관생명윤리위원회IRB의 윤리적 검증을 받아야 하며, 데이터 보유 기관의 데이터심의위원회DRB로부터 가명 처리 수준과 활용 적정성에 대한 승인을 추가로 받아야 한다. 실제 데이터 분석은 외부와 차단된 폐쇄망이나 안심구역에서 수행하는 것이 원칙이며, 분석 결과를 외부로 반출할 때에도 재식별 가능성에 대한 엄격한 심사를 거쳐야 한다. 더 나아가 서로 다른 기관의 데이터를 결합해야 할 경우에는 정부가 지정한 '결합전문기관'<sup>37)</sup>을 통해야 하는 등 모든 과정에서 개인정보 침해 위험을 최소화하기 위한 통제가 중층적으로 작동하고 있다.

35) 데이터 3법은 「개인정보보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 및 「신용정보의 이용 및 보호에 관한 법률」을 통칭한다.

36) 미국, EU 등 주요국도 개인정보를 안전하게 보호하는 동시에 데이터 활용을 촉진하는 방향으로 바이오 데이터 보호 및 활용 제도를 발전시켜 왔다. 세부 내용은 <참고 2>을 참조하기 바란다.

37) 현재 결합전문기관은 총 21개이며, 이 중 보건의료 분야 결합전문기관은 국민건강보험공단, 건강보험심사평가원, 국립암센터 등 3곳이다.

**39. 한편 정부는 바이오 데이터의 안전하고 책임 있는 활용을 촉진하기 위해 제도 정비와 인프라 구축을 중심으로 다양한 정책적 노력을 지속해 왔다.** 우선 법적 불확실성을 완화하기 위해 ‘보건의료데이터 활용 가이드라인’을 마련하여 바이오 데이터의 특수성을 반영한 활용 기준과 절차를 제시하였다. 또한 ‘전자의무기록시스템 인증제’ 시행<sup>38)</sup>과 임상용어 통일을 위한 ‘공통 데이터 모델’(Common Data Model 도입<sup>39)</sup>) 등 데이터 표준화를 통한 상호운용성 확보 노력도 지속해 왔다. 아울러 건강보험심사평가원, 국민건강보험공단 등의 의료 데이터를 연계하는 통합 플랫폼 구축 사업도 병행되었다(표 4.2). 다만 이러한 통합 플랫폼 구축 정책은 여러 개별 사업으로 나뉘어 추진되고 있고, 데이터 제공 범위와 이용 자격이 엄격한 조건 하에 제한되는 등 활용에 일정한 제약이 있는 것으로 평가된다<sup>40)</sup>.

### ■ 바이오 데이터 활용을 둘러싼 인센티브 불일치

**40. 그동안의 정책적 노력에도 불구하고 실제 데이터 활용이 담보 상태인 배경에는, 비용과 위험을 부담하는 주체와 편익을 누리는 주체가 서로 다른 ‘인센티브 불일치’ 문제가 자리하고 있다.** 정보주체(개인)와 데이터 수집관리자(병원)는 프라이버시 침해 위험, 관리 비용, 법적 책임 등을 부담하는 반면, 신약 개발 등 데이터 활용의 편익은 대체로 데이터 활용자(기업)와 사회 전체로 분산된다(표 4.3). 여기에 더해 사전 동의 확보와 비식별 처리에 대한 책임까지 각 이해관계자에게 개별적으로 부과되면서, 부담과 편익 간의 불균형은 더욱 심화되고 있다.

바이오 데이터 생태계의 이해관계자 간 인센티브가 불일치

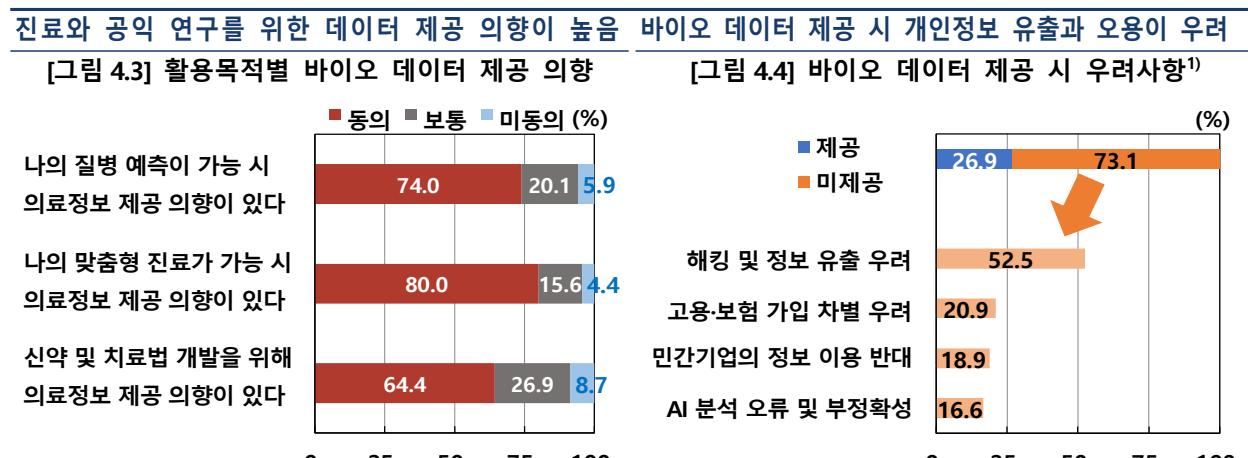
[표 4.3] 바이오 데이터 생태계의 인센티브 구조

	핵심 관심사	인센티브
 정보주체 (개인)	- 프라이버시 침해 위험(보험료 할증 등) - 개인정보 통제권 상실	데이터 제공 주체
 수집관리자 (병원)	- 데이터 처리 비용 부담(정제, 비식별화 등) - 데이터 유출 시 법적·평판 리스크 - 데이터 독점을 통한 경쟁우위 확보	데이터 공유 기피
 활용자 (기업)	- 데이터 접근·결합 비용 부담 - 법적 불확실성에 따른 수익모델 부재	데이터 활용 소극적

- 38) 보건복지부와 한국보건의료정보원은 2020년부터 의료 데이터의 상호운용성 강화를 위해 「전자의무기록시스템 인증제」를 시행하고 있다. 해당 제도는 전자의무기록(EHR) 소프트웨어의 국가 표준 준수 여부를 인증함으로써 의료기관 간 데이터 연계와 호환성 제고를 목표로 한다.
- 39) 보건복지부와 산업통상자원부는 기관 간 데이터 연계를 지원하기 위해 「공통 데이터 모델」을 기반으로 건강보험심사평가원의 청구 자료와 43개 의료 데이터 중심병원의 임상 자료를 표준화된 용어와 구조로 변환해 데이터베이스화하였다.
- 40) 일례로 보건복지부의 「보건의료 빅데이터 플랫폼」은 데이터 제공 범위가 10개 공공기관 데이터로 한정되어 민간 병의원의 임상 정보와 행정 데이터, 유전체 데이터 등을 포함되지 않으며, 이용 자격도 공공기관, 학계, 연구기관 소속으로 제한된다.

## ▶ 정보주체: 개인정보 보호 신뢰 부족으로 데이터 제공을 주저

**41. 정보주체는 데이터에 대한 실질적 통제권을 충분히 담보받지 못해 개인정보 보호 체계에 대한 신뢰가 낮은 상황이다.** 현행 사전 동의 절차는 지나치게 복잡하고 형식적으로 운영되어, 이용자가 내용을 충분히 이해하지 않은 채 ‘읽지 않고 동의’하거나 서비스 이용을 위해 사실상 수밖에 없는 ‘강제적 동의’ take-it-or-leave-it의 구조를 띠고 있다. 이는 정보주체의 통제권을 무력화하여 데이터 활용 전반에 대한 불신을 확대시키는 요인으로 작용한다. 또한 AI 기술의 발전으로 데이터 결합과 고도화된 패턴 분석이 용이해지면서 가명·익명정보의 재식별 위험도 지속적으로 증가하고 있다. 이는 비식별화 조치만으로는 개인정보 침해를 완전히 예방하는 데 한계가 있음을 시사한다. 결국 한 번 공유된 데이터는 개인의 통제 범위를 벗어나게 되고, 정보주체는 자신의 데이터가 언제, 누구에게, 어떤 목적으로 활용되는지 알기 어려운 ‘정보 비대칭’ 상황에 놓이게 된다. 그 결과, 데이터 오남용이나 유출 사고가 발생하더라도 수집·활용 기관에 책임을 묻기 어려운 구조가 고착화되고 있다.



자료: 4차 산업혁명위원회(2020)

주: 1) 개인건강정보 제공을 꺼리는 주요 이유(1·2순위 복수 응답 결과)

자료: 한국생명공학연구원(2024)

**42. 이러한 구조적 한계로 인해 정보주체는 데이터 활용의 공익적 필요성에는 공감하면서도, 유출 및 오용에 대한 우려로 실제 데이터 제공은 주저하고 있다.** 4차 산업혁명위원회(2020) 조사에 따르면, 응답자의 70% 이상이 질병 예측 및 맞춤형 진료를 위해, 64.4%는 신약 개발 등 공익 연구를 위해 의료정보를 제공할 의향이 있다고 응답했다(그림 4.3). 반면 본 연구의 설문조사에서 구체적인 활용 목적을 제시하지 않고 데이터 제공 의향을 질의했을 때에는 긍정 응답이 26.9%에 그쳤다. 정보 제공을 거부한 주된 이유로 응답자의 52.5%는 해킹이나 유출로 인한 개인정보 침해를, 20.9%는 고용이나 보험 가입 과정에서의 차별 등 실질적 피해를 지적했다(그림 4.4). 이는 사회적 신뢰 형성에 한계를 보이며, 데이터 활용에 대한 사회적 수용성을 제약하고 있음을 시사한다.

## ▶ 데이터 수집관리자: 이익 대비 비용이 커 데이터 공유를 기피

**43. 데이터 수집관리자 입장에서는 데이터 공유는 기대되는 이익에 비해 감수해야 할 비용과 위험이 현저히 크다.** 데이터를 공유하려면 정제, 비식별 처리, 보안 인프라 구축 등을 위해 상당한 기술적·행정적 비용을 부담해야 한다. 데이터 유출 시에는 법적 책임과 평판 훼손이라는 중대한 위험도 감수해야 한다. 또한 경쟁 의료기관과의 데이터 공유는 환자 이탈이나 진료 노하우 유출로 이어질 수 있고, 데이터 독점을 통한 경쟁 우위를 상실할 가능성도 내포하고 있다. 반면 이에 대한 보상은 현저히 부족하다. 현행 유권해석<sup>41)</sup>은 가명정보 제공 대가를 ‘제공에 소요된 비용’ 즉 실비 수준으로 한정하고 있어, 데이터 정제를 통해 창출되는 부가가치나 잠재적 법적·평판 위험을 충분히 보상하지 못하고 있다. 연구 협력을 통한 평판 제고나 진료 역량 강화와 같은 비재무적 이익 역시 데이터 독점을 통해 유지되는 경쟁 우위를 포기할 만큼 유의미한 유인으로 작용하기는 어렵다.

**44. 이러한 비용·위험 대비 보상의 불균형은 데이터 수집관리자의 데이터 공유 기피로 이어지고 있다.** ‘2020년 보건의료정보화 실태조사’에 따르면, 병원 데이터를 연구 목적으로 외부와 공유한 비율은 21%에 그쳤으며, 외부인이 단독으로 수행하는 연구의 경우 그 비율은 17.4%로 더욱 낮았다. 이는 병원과 직접적인 협력 관계가 없는 외부 연구자나 기업이 의료 현장의 임상 데이터에 접근하기가 사실상 어렵다는 현실을 보여준다(그림 4.5). 아울러 ‘2024 개인정보보호 및 활용조사’에서도 최근 1년간 가명정보를 활용(내부 활용 포함)하거나 제공한 보건·복지 서비스 기관은 9.8%에 불과했다. 이를 기관은 가명정보 제공·활용 과정에서의 주된 애로사항으로 예산 부족(35.0%)과 교육·지원 부족(27.8%)을 꼽았다. 데이터 공유의 핵심 기반인 ‘전자의무기록시스템 인증제’ 역시 병원급 의료기관의 참여 저조로 전체 인증률이 약 10% 수준에 머물러 있다<sup>42)</sup>. 나아가 법률을 지나치게 보수적으로 해석하거나 절차를 복잡하게 설계함으로써 공유를 회피하는 행태도 관찰된다. 실제로 미국에서는 의료정보 보호법을 도입한 주가 그렇지 않은 주에 비해 전자의무기록 도입률이 약 24% 낮다는 연구 결과가 보고된 바 있다(Miller·Tucker 2009).

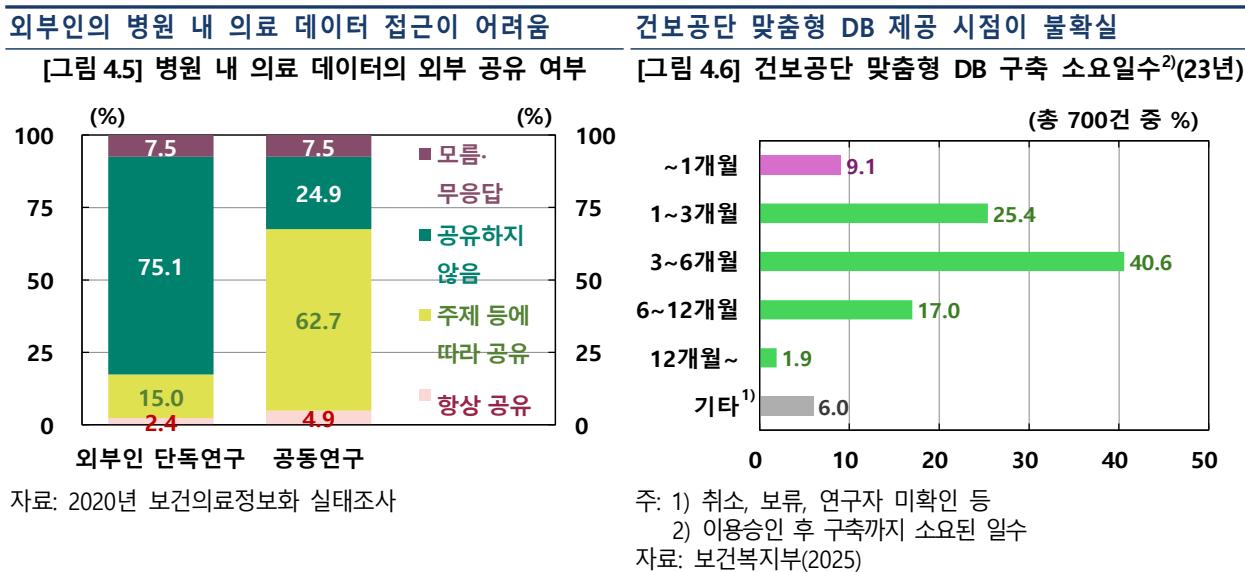
41) 「개인정보 질의응답 모음집」(24.12월), 「보건의료데이터 활용 가이드라인」(24.12월)

42) 전자의무기록시스템 인증은 의무가 아닌 자율 참여 방식으로 운영된다(의료법 제23조의2). 의료기관은 이를 통해 의료서비스의 질 제고나 홍보 효과를 기대할 수 있다. 의료기관 유형별 전자의무기록시스템 인증률은 다음과 같다.

구 분	상급종합병원 (병상 500개 이상)	종합병원 (병상 100~500개)	병 원 (병상 30~100개)	의 원 (병상 30개 미만)	전 체 (%)
▶ 인증률 <sup>1)</sup>	93.6	52.3	3.5	10.2	10.4

주: 1) 사용인증 기관수(25.9월)÷전체 기관수(25.2/4분기)

자료: 전자의무기록시스템인증 홈페이지, 건강보험통계



#### ▶ 데이터 활용자: 높은 접근 비용과 법적 불확실성으로 데이터 활용에 소극적

**45. 데이터 활용자는 과도한 접근 비용과 법적 불확실성이라는 이중의 제약에 직면해 있다.** 우선 의료 데이터가 여러 기관에 분산되어 있고 표준화 수준도 미흡한 데다, 결합·활용 절차가 복잡하여 데이터를 확보하는 데 많은 시간과 비용이 소요된다. 특히 심의를 일괄 처리하는 통합 원스톱 서비스 one-stop service가 부재하여, 연구자는 생명윤리위원회 IRB와 각 데이터 제공 기관의 심의를 순차적으로 거쳐야 하는 부담을 안고 있다. 데이터 제공 소요 시간 또한 예측 가능성이 낮다. 일례로 2023년 국민건강보험공단 맞춤형 DB<sup>43)</sup> 사례를 보면, 승인 후 1개월 이내에 제공된 비율은 9.1%에 불과한 반면, 6개월 이상 소요된 경우는 18.9%에 달했다(그림 4.6). 더불어 데이터 품질을 사전에 검증하기 어려워 확보한 데이터가 연구 목적에 부합하지 않거나 특정 집단에 편향될 위험도 존재한다. 재식별 우려로 인한 과도한 비식별 조치는 유전체 정보 등 핵심 데이터의 정보 가치를 훼손해 연구 실효성을 저해하기도 하며, 데이터의 지속적 제공이 보장되지 않아 연구 결과의 재현성 확보에도 한계가 있다. 법적 불확실성 또한 중요한 제약 요인이다. 현행 ‘보건의료데이터 활용 가이드라인’은 산업적 목적의 가명정보 활용을 허용<sup>44)</sup>하고 있으나, 영리적 활용에 대한 사회적 반감이 여전히 존재하는 가운데 향후 규제가 보수적으로 해석될 가능성에 대한 법적 리스크가 상존하고 있다.

43) 건강보험공단 맞춤형 DB는 건강보험 및 장기요양보험 원자료를 연구 목적에 맞게 추출·가공한 후, 개인 식별이 불가능하도록 조치해 제공하는 데이터셋이다.

44) 「보건의료데이터 활용 가이드라인」은 가명정보의 활용 가능 범위에 새로운 기술·제품·서비스의 연구개발 및 개선 등 산업적 목적의 연구를 과학적 연구로 포함하고 있음을 명시하고 있다.

**46. 이러한 제약으로 데이터 활용자는 AI 기반 신약 개발이나 질병 예측 모델링 등 혁신을 위해 고품질 데이터를 필요로 하면서도, 실제 데이터 활용에는 소극적인 태도를 보이고 있다.** 국내 의료 데이터는 개별 병원뿐만 아니라 건강보험심사평가원, 국민건강보험공단 등 여러 기관에 분산되어 있어, 의미 있는 혁신을 위해서는 데이터 연계와 결합이 필수적이다. 그럼에도 불구하고 보건의료 분야 결합전문기관을 통한 데이터 결합 건수는 2020년 10월부터 2023년 10월까지 3년간 총 78건에 그쳤다(보건복지부 2023). 이는 바이오 데이터의 제도적 활용이 여전히 제한적임을 보여준다. 결과적으로 많은 국내 기업들은 데이터 접근에 따른 비용과 법적 불확실성을 회피하기 위해, 병원과의 공동연구 업무협약MOU이나 소프트웨어 공급 계약 등을 통해 자체적으로 데이터를 확보하는 방식을 선택하고 있다(OECD 2022).

**47. 결론적으로, 바이오 데이터 생태계 전반에서 이해관계자 간 인센티브 불일치는 데이터 활용을 저해하는 구조적 '시장 실패'로 귀결되고 있다.** 비용과 위험을 부담하는 주체와 편익을 누리는 주체가 서로 다른 구조 속에서 각 이해관계자는 데이터를 공개하지 않거나 독점하는 것을 합리적인 선택으로 인식하게 된다. 그 결과 데이터는 기관별로 파편화되고, 품질은 정제되지 않은 채 유지되며, 활용 절차의 비효율성이 고착화되고 있다. 이는 데이터 공유가 의학적 혁신과 글로벌 R&D 협력을 촉진함으로써 사회 전체의 편익을 크게 증대시킬 잠재력을 지니고 있음에도 불구하고, 이를 충분히 실현하지 못하는 전형적인 시장 실패의 모습이라 할 수 있다.

## 5. 바이오 데이터 활용 기반 확충을 위한 정책 방안

### ■ 인센티브 재설계 방안: 「국가 승인형 바이오 데이터 개방 체계」의 구축

**48. 각 주체가 직면한 비용-편익 인센티브 왜곡을 해소하기 위한 구체적 방안으로 「국가 승인형 바이오 데이터 개방 체계」(이하 국가 승인 체계)의 구축을 제안한다.** 바이오 데이터의 실질적 활용을 촉진하려면, 각 주체가 부담하는 위험과 비용은 최소화하고 합당한 보상은 명확히 보장하는 방향으로 인센티브 구조를 재설계해야 한다. 이를 위해 해당 체계는 사전 심사를 통해 법에서 정한 공익성 요건을 충족한 연구에 한해 데이터 활용을 승인하되, 승인된 연구에 대해서는 사전 동의 면제 등 규제를 완화하고 데이터 유통을 지원함으로써 인센티브 불일치를 완화하는 데 초점을 둔다. 이는 개인의 자발적 협조에만 의존하기보다, 각 주체가 합리적 선택에 의해 데이터 생태계에 참여하도록 유도하는 제도적 기반을 마련하는 데 목적이 있다. 「국가 승인 체계」는 ① 국가의 엄격한 승인과 통제권 담보, ② 이를 전제로 한 규제 완화, ③ 데이터 활용 지원을 위한 유통 생태계 활성화를 3대 축으로 한다.

## ▶ 신뢰 구축을 위한 공적 승인과 통제권 담보

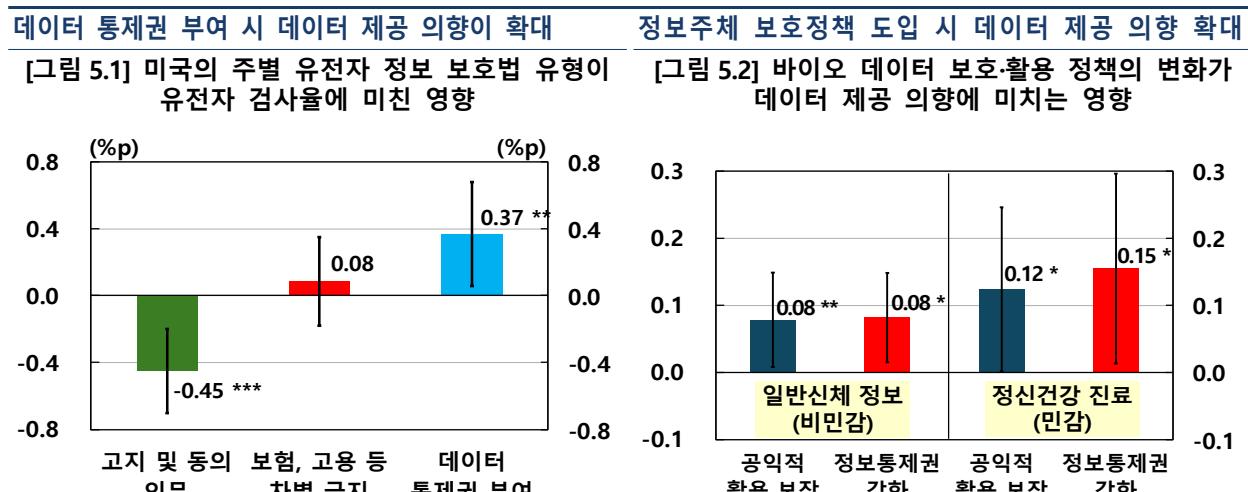
49. ‘국가 승인 체계’에서 국가는 공익적 활용을 보장함으로써 데이터 제공에 수반되는 위험과 비용을 낮추고 정보주체의 권리를 담보하는 데이터 보호의 수문장 역할을 수행한다. 현재의 형식적인 사전 동의나 기술적 비식별화만으로는 보험 가입 거절, 채용 차별 등 정보주체가 우려하는 잠재적 피해를 예방하는 데 한계가 있다. 이에 ‘국가 승인 체계’는 다음과 같은 안전장치를 통해 신뢰를 회복하고 자발적 데이터 제공을 유도한다.

- (활용 목적 중심의 규제 전환) EU의 ‘European Health Data Space’ 법안<sup>45)</sup>과 같이 데이터의 ‘수집·공유’ 자체보다 ‘활용 목적’에 대한 규제를 강화한다. ‘수집 목적 외 활용’<sup>secondary use</sup>에 대해서는 사전 심의·승인을 의무화하여, 공익적 활용은 보장하되 정보를 제공한 개인에게 불이익을 주거나 사회적 윤리에 반하는 활용은 원천 차단한다.
- (실질적 통제권 담보) 정보주체가 자신의 데이터 활용 현황을 투명하게 확인하고, 언제든 거부<sup>opt-out</sup>하거나 다시 허용<sup>opt-back-in</sup>할 수 있는 통제권을 부여한다. 이는 ‘프라이버시 우려 완화→신뢰 형성→데이터 제공 확대’의 선순환을 가능하게 한다. Miller·Tucker(2018)의 유전자 검사 동의 실험에 따르면, 단순 위험 고지 시에는 동의율이 감소했으나 실질적 통제권을 부여했을 때는 동의율이 0.37%p 유의하게 증가하였다(그림 5.1).
- (활용 과정 추적·감독) 데이터 접근부터 분석까지의 전 과정을 체계적으로 추적·감독하고 심사 결과를 투명하게 공개한다. 오남용 발생 시 즉각적인 승인 취소와 법적 제재를 부과함으로써 사후 관리 체계에 대한 사회적 신뢰를 확립한다.

50. 실제로 ‘공익적 활용 보장’과 ‘정보 통제권 강화’ 정책이 정보주체의 데이터 제공 확대로 이어질 수 있는 점은 설문 실험을 통해 확인할 수 있다. 정책 도입의 인과적 효과를 분석하기 위해, 설문 응답자를 무작위 그룹으로 나눈 뒤 서로 다른 정책 시나리오를 제시하는 ‘무작위 정보 처치 실험’<sup>Randomized Information Treatment</sup>을 수행하였다<sup>46)</sup>. 그 결과, 두 정책은 모두 개인건강정보 제공 의향을 통계적으로 유의하게 높이는 것으로 나타났다(그림 5.2). 비민감 정보인 일반 신체 정보(키, 체중 등)의 경우 미도입 시 대비 제공 의향이 각각 7.9%p, 8.2%p 증가했다. 특히 민감 정보인 정신건강 진료기록의 경우 제공 의향이 각각 12.4%p, 15.5%p 증가하여, 정책 도입의 효과가 민감 정보에서 더 큰 것으로 나타났다.

45) 「European Health Data Space」는 건강 데이터 활용을 공익적 목적으로 제한하며, 다음의 활용을 명시적으로 금지한다: (a) 개인에게 불이익을 초래하는 결정, (b) 채용 또는 상품·서비스 제공(보험·신용 계약 포함)과 관련한 차별적 조건 설정, (c) 마케팅 활동, (d) 중독성 물질 등 개인·공중 보건 또는 사회 전체에 해를 끼칠 수 있는 제품 개발, (e) 국내법상 윤리 규정과 상충되는 방식의 사용.

46) 바이오 데이터 보호·활용 정책이 개인의 데이터 제공 의향에 미치는 영향에 대한 세부 분석 방법은 <별첨 2>를 참조하기 바란다.



주: 1) 평균 유전자 검사율(0.54%)을 기준으로 한 변화율임  
2) 막대는 효과의 평균 추정치를, 세로선은 95% 신뢰구간을 나타냄

자료: Miller-Tucker(2018)

주: 1) 막대는 각 정책 효과의 평균 추정치를, 세로선은 90% 신뢰구간을 나타냄

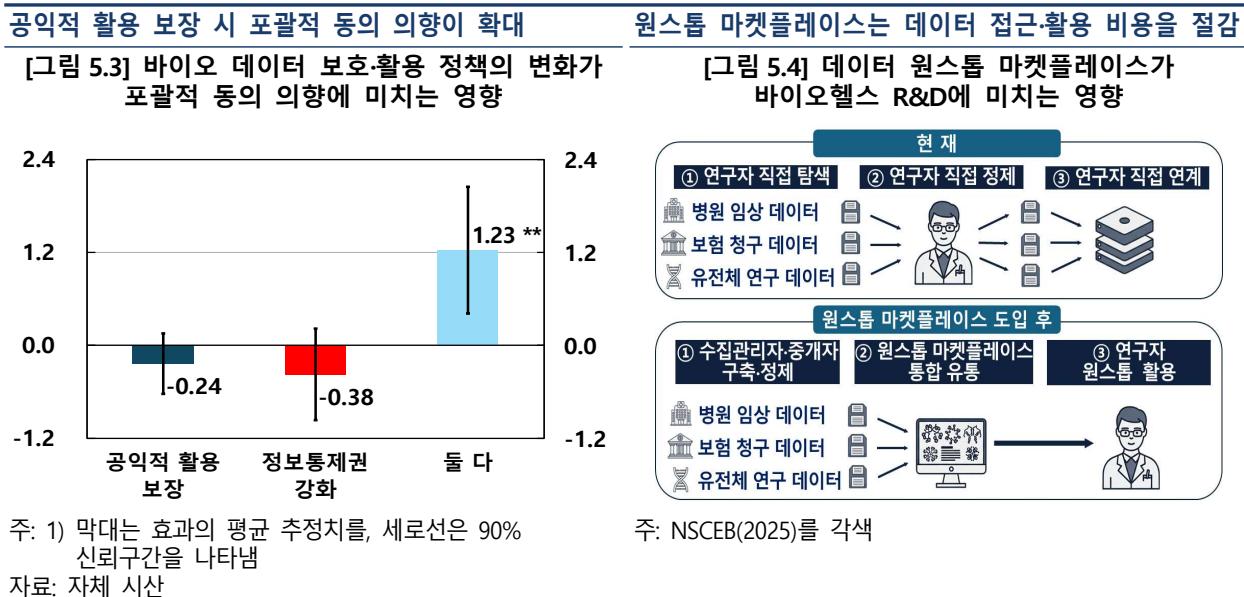
자료: 자체 시산

## ▶ 승인된 연구에 대한 규제 완화

**51. 국가의 엄격한 심사를 통과하여 공익성이 입증된 연구에 대해서는 법적 불확실성과 행정 비용을 획기적으로 경감하는 규제 완화를 적용한다.** 모호한 규제와 중복된 심사 절차는 데이터 활용자의 투자 회수 가능성을 낮추고 R&D를 위축시키는 주요 요인이다. 이에 영리 기업의 연구라 할지라도 신약 개발 등 사회적 후생 증진 효과가 명확한 경우에는 다음과 같은 규제 특례를 부여한다.

- **(사전 동의 면제 및 포괄적 동의 허용)** 심사·승인 과정을 통해 공익성이 인정된 '수집 목적 외 활용'에 대해서는 건별 사전 동의를 면제하거나, 활용 목적을 사전에 특정하지 않는 포괄적 동의<sup>47)</sup>를 허용한다. 이를 통해 연구자는 동의 획득의 난관과 법적 리스크에서 벗어나 본연의 연구에 집중할 수 있다. 설문 실험 결과, '공익적 활용 보장'과 '정보 통제권 강화'가 동시에 강화될 경우, 포괄적 동의에 대한 선호도가 정책 미시행 대비 1.23점 유의하게 상승하는 것으로 나타났다. 이는 국가의 엄격한 관리가 전제된다면 규제 완화에 대한 사회적 합의와 수용성 확보가 충분히 가능함을 시사한다(그림 5.3).
- **(원스톱 심사·승인)** 개인정보보호법, 생명윤리법 등 다층화된 법적 요구사항을 국가가 일괄 검토·승인함으로써, 연구자가 겪는 중복 행정 절차의 비효율을 해소한다.

47) 포괄적 동의란 특정 연구 목적을 사전에 한정하지 않고, 향후 수행될 여러 연구를 위해 개인정보를 저장·이용하는 것에 대한 동의를 의미한다.



#### ▶ 데이터 활용 지원을 위한 유통 생태계 활성화

52. 공적 승인을 받은 연구가 실질적으로 수행될 수 있도록 데이터 유통 생태계를 활성화하여 과도한 데이터 접근 비용은 낮추고 고품질 데이터 생산·공유에는 합당한 보상 체계를 구축한다.

- **(국가 통합 데이터 중개 허브 구축)** 데이터 공급자인 수집관리자와 수요자인 활용자를 매칭하는 국가 통합 데이터 중개 허브를 구축한다. 연구자나 기업은 개별 기관을 일일이 접촉할 필요 없이, 단 한 번의 신청으로 다수 기관 데이터의 접근, 결합, 비식별 조치 등을 일괄 진행할 수 있다(그림 5.4). 또한 데이터셋의 종류, 변수 구성, 보유 기관 등의 정보를 담은 통합 데이터 카탈로그를 제공하여 검색 편의성을 높이고, 데이터 탐색부터 법적 검토까지 전 과정에 전문 상담을 지원하여 시행착오를 최소화한다.
- **(데이터 중개자 육성)** 중립적인 제3자인 ‘데이터 중개자’를 육성하여 정제, 보안, 분석 지원 등 전문 과업을 대행하게 함으로써 기술적 진입 장벽을 낮춘다<sup>48)</sup>. 이는 수집관리자의 관리 비용 절감과 활용자의 고품질 데이터 확보를 동시에 달성하여 거래 효율성을 제고한다. 특히 중개자는 데이터 연계 시 발생할 수 있는 프라이버시 침해나 영업기밀 유출과 같은 부정적 외부효과를 차단하여 신뢰 기반의 거래를 촉진한다(Acemoglu et al. 2022). 단, 중개자의 데이터 독점이나 이해 상충을 방지하기 위해 EU의 ‘Data Governance Act’와 같은 중개자 등록제, 데이터 자가 이용 금지, 남용 방지 조치 의무화 등 엄격한 규제를 병행해야 한다.
- **(생산·공유에 대한 보상)** 데이터 임대 및 중개 수수료, 데이터 품질에 비례한 가격제 등 다양한 수익 모델<sup>49)</sup>을 허용하여 고품질 데이터 생산·공유에 따르는 비용

48) 정부는 데이터 시장 활성화를 위해 2023년부터 데이터 자문과 함께 유통·거래·증개를 담당하는 「데이터 거래사」 제도를 운영하고 있다.

을 보전하고 경제적 유인을 제공한다. 아울러 민간 병원 등의 실질적 참여를 촉진하기 위해 초기 표준화 및 시스템 구축 비용 지원, 기여 의료진에 대한 성과 인정, 법적 책임 소재의 명확화 등 제도적 지원도 병행할 필요가 있다. 데이터가 안정적 수익을 창출하는 자산으로 인식될 때, 수집관리자는 자발적으로 보안과 품질 관리에 투자하게 된다. 특히 이는 정보주체가 보안을 우려하면서도 편의를 위해 정보를 쉽게 제공하는 ‘프라이버시 역설’<sup>privacy paradox<sup>50)</sup></sup>

 상황에서 수집관리자가 가질 수 있는 보안 불감증을 교정하는 기제로 작용할 수 있다(Athey et al. 2017).

### 53. 종합하면, ‘국가 승인 체계’는 바이오 데이터 활용의 병목 원인이었던 이해관계자별 인센티브 불일치를 다음과 같이 완화함으로써, 사회적 신뢰에 기반한 데이터 제공·공유와 혁신 수요가 맞물려 돌아가는 선순환 구조를 이끌 것이다.

- (정보주체) 자신의 데이터가 안전하게 관리·활용되는 과정에 대해 충분한 정보를 제공받고, 활용 범위와 방식에 대해 주체적으로 선택하며 실질적인 통제권을 행사할 수 있다.
- (데이터 수집관리자) 사회적 신뢰를 바탕으로 고품질 데이터를 안전하게 생산·공유하고 합당한 수익 창출 기회를 통해 운영의 지속 가능성 확보할 수 있다.
- (데이터 활용자) 제도적 예측 가능성 아래 양질의 데이터에 원활하게 접근함으로써, AI 개발 등 혁신을 가속화할 수 있다.

### ▶ 「국가 승인형 바이오 데이터 개방 체계」를 위한 거버넌스

### 54. ‘국가 승인 체계’가 실효성 있게 작동하기 위해서는 공적 승인과 사후 감시 등 국가의 핵심 역할을 전담할 별도의 기구 설립을 고려할 필요가 있다. 이미 EU의 ‘Health Data Access Bodies’나 핀란드의 ‘Findata’ 등 선진국들은 유사한 기능을 수행하기 위해 독립된 전담 기구를 운영하고 있다.

### 55. 해당 전담 기구가 본연의 기능을 효과적으로 수행하려면 독립성, 전문성, 재정적 지속가능성을 갖춘 운영 체계를 갖출 필요가 있다.

- (독립성) 기존 공공기관 산하 조직 형태는 소극적 행정이나 부처 간 이해 상충 문제에 직면할 소지가 크다<sup>51)</sup>. 따라서 데이터 활용을 핵심 임무<sub>mandate</sub>로 명시한

49) 데이터는 정보자처럼 초기 구축에는 많은 비용이 들지만, 일단 구축되면 복제와 배포 비용이 거의 발생하지 않는다. 이러한 특성 때문에 희소성을 전제로 작동하는 전통적인 시장의 수요-공급 원리를 적용하기 어렵다. 이에 따라 데이터 거래에는 음악 스트리밍 서비스처럼 정기 구독료를 받는 「임대 모델」이나, 소프트웨어처럼 버전별로 품질과 갱신 수준을 차별화하는 「버전 관리 모델」을 적용할 수 있다.

50) Athey et al.(2017)은 현장 실험을 통해 개인정보 보호에 대한 우려와 실제 행동 간 괴리가 존재하는 이른바 「프라이버시 역설」을 확인했다. 소비자들은 소액의 금전적 보상(예: 피자 한 조각 쿠폰)에도 개인정보를 쉽게 제공했으며, 몇 차례의 클릭과 같은 사소한 번거로움만으로도 필요한 보호 조치를 취하지 않는 경향으로 보였다.

51) 개인정보보호위원회의 「공공분야 가명정보 제공 실무안내서」(21.1월)에 따르면, 공공기관의 데이터 제공은 의무 사항이 아니며, 가명 처리에 필요한 비용이나 기관 본연의 업무에 미치는 영향 등 내부 여건을 고려해 재량적으로 결정할 수 있다.

별도의 법정 기구를 신설하거나, '한국보건의료정보원'<sup>52)</sup> 등 유관 기관을 독립성이 보장되는 형태로 확대 개편하는 방안을 검토할 필요가 있다. 이를 통해 개별 기관의 이해관계를 초월하여 일관되고 신속한 의사결정을 수행할 수 있도록 해야 한다.

- **(전문성·중립성)** 데이터 전문가, 의사 과학자 등 전문 인력을 중심으로 거버넌스를 구축할 필요가 있다. 이사회는 공익성을 최우선 가치로 두되, 시장 변화에 대응하는 민첩성을 갖춘 구조로 설계되어야 한다. 아울러 시민 자문 그룹을 제도화하여 사회적 신뢰를 확보하고, 이해관계자 간의 갈등을 효과적으로 중재하는 장치를 마련해야 한다.
- **(재정적 지속가능성)** 정부 예산에만 전적으로 의존하는 구조를 지양하고, 데이터 중개 등 서비스 제공에 따른 합리적 수수료 체계를 도입하여 자체 수익 모델을 구축할 필요가 있다. 이는 중장기적 인프라 투자와 서비스 고도화를 가능케 하며, 정책 환경의 변화에도 흔들리지 않는 안정적인 운영 기반이 될 것이다.

**56. 단, 전담 기구가 데이터의 수집부터 정제, 제공까지 모든 과정을 직접 수행하는 것은 기술적 전문성, 장기 투자 역량, 시장 대응력 측면에서 한계를 가질 수 있음을 유의해야 한다.** 일례로 공공 주도 모델인 핀란드의 'Findata'는 과도한 행정 절차와 기술적 복잡성으로 인해 데이터 접근 승인에만 6~10개월이 소요되고 있으며, 지속적인 데이터 품질 이슈와 상업적 활용(AI 개발 등)의 유연성 저하 문제가 제기되고 있다(Eronen et al. 2025). 따라서 전담 기구는 모든 것을 직접 수행하기보다는, 민간이 자율적으로 고품질 데이터 생태계를 생산·공유할 수 있도록 지원하는 시장 조성자이자 신뢰 보증자로서의 역할에 집중할 필요가 있다.

## ■ 기대효과

**57. '국가 승인 체계'는 고품질 바이오 데이터를 국가 핵심 전략 자산으로 전환하여, 저성장 국면의 우리 경제에 새로운 성장동력을 제공할 것이다.** 안전하고 신뢰할 수 있는 데이터 활용 기반이 구축된다면, 바이오 데이터는 고부가가치 산업 자산으로서 다음과 같은 경제적 파급효과를 창출할 수 있다.

- **(데이터 시장 확대)** 대표성을 갖춘 데이터를 고부가가치 상품으로 개발하고, 글로벌 제약사와 의료기기 기업에 안전한 공유 서비스를 제공함으로써 국제 공동 연구와 R&D 투자를 유치하여 새로운 국부를 창출할 수 있다. 이는 바이오 데이터의 가치와 시장 규모를 확대하는 효과를 가져온다.
- **(R&D 효율성 향상)** 여러 기관에 흩어진 데이터 접근 창구를 하나로 통합하면 연구 속도와 효율이 크게 향상되어 혁신 신약 및 첨단 의료기기 개발 비용의 절감과 성공률 제고로 이어진다. 이는 바이오헬스 산업 전반의 R&D 생산성과 경쟁력

52) 한국보건의료정보원은 보건의료 데이터의 안전한 활용과 가치 창출을 지원하기 위해 2019년에 설립된 공공 기관으로, 보건의료 데이터 표준화, 전자의무기록시스템 인증, 진료정보교류 및 본인진료기록열람시스템 운영, 의료데이터중심병원 추진, 보건의료 빅데이터 플랫폼 운영 등의 기능을 수행하고 있다.

을 강화하는 계기가 될 것이다.

- (글로벌 R&D 협력 도약) 아울러 국내 기업은 통합 데이터를 활용해 신약 후보물질 발굴, 디지털 헬스케어, AI 기반 진단·치료 솔루션 등 혁신 제품을 개발할 실질적 기반을 확보하게 된다. 이를 토대로 글로벌 기업과의 공동 연구개발이 활성화되어 우리나라가 글로벌 바이오헬스 R&D 협력으로 도약할 수 있는 발판이 마련될 것이다. 아울러 데이터 과학자, AI 전문가, 데이터 중개인 등 전문 인력 수요 증대로 양질의 일자리가 창출되는 성장의 선순환 구조가 형성될 것이다.

## 58. 이러한 경제적 효과에 더해 '국가 승인 체계'는 개인, 의료기관, 산업체, 정부 등 전 영역의 데이터 활용을 촉진하여 국민 건강 증진에도 기여할 것이다(표 5.1). 데이터의 안전한 활용이 촉진될 경우 각 주체는 다음과 같은 실질적 편익을 누릴 수 있다.

- (개인) 자신의 유전체, 임상 정보, 생활습관 데이터를 기반으로 맞춤형 식단과 운동 계획을 제공받는 등 건강 관리 서비스의 질적 개선이라는 직접적인 편익을 누릴 수 있으며, 나아가 신약 개발 등 데이터 활용을 통한 혁신의 성과가 개인에게도 실질적인 효익으로 귀속된다.
- (의료기관) 분산된 환자 정보를 신속히 파악해 진단과 치료의 정밀성을 높일 수 있다. 또한 데이터 기반 행정 효율화를 통해 의료 자원을 보다 효율적으로 배분함으로써 의료서비스의 질을 향상시킬 수 있다.
- (산업계) 양질의 데이터를 AI 등 첨단 기술과 결합해 혁신 신약 및 첨단 의료기기 개발을 가속할 수 있다. 나아가 질병의 예방부터 치료, 관리에 이르는 헬스케어 전 주기를 혁신하고 새로운 비즈니스 모델을 창출할 수 있다.
- (정부) 방대한 데이터를 분석하여 질병 통계 생산, 감염병 확산 예측 등 과학적 근거에 기반한 보건정책을 수립할 수 있다. 이는 정책의 정확성과 대응 속도를 제고하는 동시에, 한정된 재원을 보다 효과적으로 배분하는 데 기여한다.

### '국가 승인 체계'는 전 영역의 데이터 활용을 촉진해 국민 건강 증진에 크게 기여

[표 5.1] 바이오 데이터가 국민 건강 증진에 기여하는 경로

영역	활용 분야	주요 내용	기대 효과
개인	▶ 개인 맞춤형 치료	- 개인의 유전체, 임상, 생활습관 데이터를 통합 분석해 최적 관리 · 치료법 적용	- 관리 · 치료 효과 극대화, 부작용 최소화
	▶ 질병의 조기 예측 및 예방 관리	- 바이오 데이터를 학습한 AI 모델을 통해 질병 위험 예측 및 고위험군 관리	- 중증질환 조기 대응, 질병 발생 억제
병원	▶ 환자 치료 및 의료서비스 지원	- 환자 정보 공유를 통한 정확한 진단 · 모니터링	- 의료 오류 감소, 진료 효율성 제고
	▶ 의료시스템 개선 및 행정 효율화	- 대기 · 청구 · 퇴원관리 자동화, 기관 간 정보 연계	- 종복 검사 감소, 의료비 절감, 의료서비스 질 향상
산업	▶ 신약 및 새로운 치료법 개발	- AI · 빅데이터 기반 후보물질 탐색, 독성 예측, 임상시험 최적화	- 신약 개발 기간 단축, 희귀 · 난치 질환 치료 기회 확대
국가	▶ 보건정책 수립 및 공중보건 강화	- 국가 단위 데이터 분석을 통한 보건통계 산출 및 정책 설계	- 과학적 근거 기반 정책 수립, 감염병 대응력 강화

## 6. 맷음말

**59. AI 혁신은 우리나라가 글로벌 바이오헬스 R&D 허브로 도약할 새로운 기회를 제공하고 있다.** 그동안 우리나라는 ‘빠른 추격자’ 전략을 통해 바이오시밀러, CDMO 등 틈새시장을 중심으로 성과를 거두었으나, 고부가가치를 창출하는 혁신 신약 및 첨단 의료기기 분야에서는 선도국과 큰 격차를 보이고 있다. 그러나 AI가 바이오헬스 산업의 경쟁 구도를 재편하는 ‘게임체인저’로 부상하면서, 우수한 바이오 데이터 수집·연계 인프라를 가진 우리나라에게 새로운 기회의 창이 열리고 있다. 대규모 고품질 바이오 데이터 구축을 통해 AI 기반의 R&D 환경을 조성한다면 글로벌 자본과 기술, 경험과의 전략적 파트너십을 통해 혁신을 주도하는 바이오헬스 강국으로 발돋움할 수 있다.

**60. 이러한 기회를 현실로 만들기 위해서는 바이오 데이터 활용 기반의 구축이 필수적이다.** 우리나라는 단일 건강보험 체계와 높은 의료 접근성 덕분에 세계적으로도 희소한 대규모 데이터를 구축할 잠재력을 보유하고 있다. 그럼에도 불구하고 실제 데이터 공유와 활용은 저조하다. 이러한 한계의 근본 원인은 데이터 활용을 둘러싼 이해관계자 간 인센티브 불일치에 있다. 정보주체는 프라이버시 침해를, 데이터 수집관리자는 처리 비용과 법적·평판 리스크 부담을, 데이터 활용자는 높은 접근 비용과 법적 불확실성을 우려하며, 그 결과 데이터 활용이 위축되는 시장 실패가 발생하고 있다.

**61. 이에 본고는 인센티브 불일치를 완화하고 데이터 기반 혁신을 촉진하기 위한 해법으로 ‘국가 승인형 바이오 데이터 개방 체계’의 구축을 제안한다.** 해당 체계는 첫째, 공적 승인을 통해 데이터 활용의 공익성을 담보하고 정보주체에게 실질적인 데이터 통제권을 부여한다. 이를 통해 데이터 제공에 따르는 위험과 비용을 낮춤으로써 신뢰를 회복하고 자발적인 데이터 제공을 유도한다. 둘째, 국가의 엄격한 심사를 통과하여 공익성이 입증된 연구에 대해서는 사전 동의 면제 등 관련 규제를 완화하여, 연구자가 직면한 행정적 부담과 법적 불확실성을 완화한다. 셋째, 국가 단일 데이터 중개 허브 구축 등을 통해 유통 생태계를 활성화한다. 이를 통해 과도한 데이터 접근 비용은 낮추고, 고품질 데이터의 생산과 공유에는 합리적 보상을 제공하는 선순환 구조를 만든다.

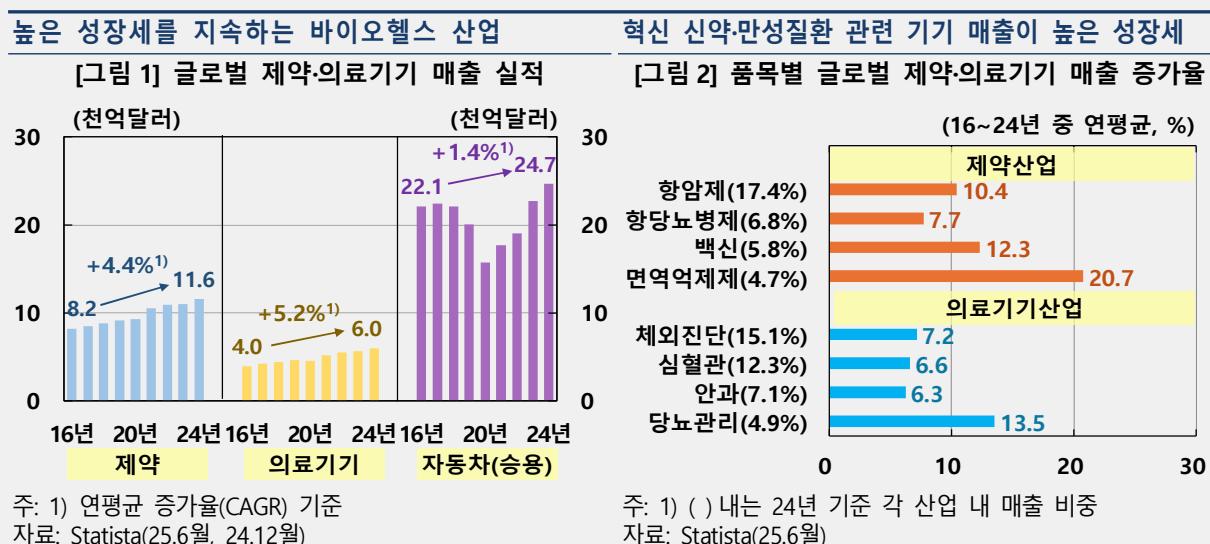
**62. AI 혁신으로 바이오헬스 산업이 중대한 전환점을 맞이한 지금, 데이터 오용이나 유출의 위험뿐만 아니라 데이터가 활용되지 못함으로써 발생하는 기회비용의 위험 역시 중대하게 인식해야 한다.** 개인정보 보호와 데이터 활용은 상충하는 가치가 아니라, 신뢰를 기반으로 상호보완적으로 작동하는 가치로 보아야 한다. 철저한 보

호를 통해 신뢰를 구축할 때에만 안전한 활용과 지속적인 혁신이 가능하다. ‘국가 승인형 바이오 데이터 개방 체계’는 이러한 관점을 제도적으로 구현하는 핵심 장치로서, 개인의 신뢰를 바탕으로 고품질 바이오 데이터를 국가 핵심 전략 자산으로 전환해 민간 혁신을 촉진할 것이다. 이를 통해 우리나라는 글로벌 바이오헬스 선도 국으로 도약하여 저성장 국면을 돌파할 새로운 성장동력을 확보하고 국민 건강 증진에도 기여할 수 있을 것이다.

## 참고 1 세부 품목별 글로벌 바이오헬스 산업 동향

최근 글로벌 바이오헬스 산업은 차세대 바이오의약품 및 만성질환 의료기기를 중심으로 높은 성장세를 보이고 있다(그림 1, 2).

- (제약)** 최근 유전자 재조합<sup>53)</sup>, 항체 생산 기술(Hybridoma)<sup>54)</sup> 등 생명공학 기술이 발전하며 면역항암제, 항당뇨병제, 면역억제제 등 차세대 바이오의약품이 성장을 주도하고 있다. 이에 따라 암, 당뇨, 자가면역질환 등 치료를 위한 바이오의약품 개발이 증가하며 매출도 빠르게 성장하고 있다. 특히 코로나19 팬데믹 기간에는 mRNA 기술<sup>55)</sup>을 적용한 백신 매출이 급증했다.
- (의료기기)** 심혈관기기, 안과기기, 당뇨관리기기 등 만성질환 관련 제품이 성장을 견인하고 있다. 세계적인 고령화 추세로 고혈압, 백내장, 당뇨 등 만성질환자가 증가하면서 관련 의료기기 수요가 급증하고 있다. 또한 최소침습수술<sup>56)</sup>, 레이저, 디지털 기술 등을 적용한 첨단 의료기기 개발 역시 매출 증가에 기여하고 있다. 한편, 체외진단기는 코로나19 팬데믹 동안 진단키트 판매 호조로 전례 없는 성장을 기록했다.



53) 유전자 재조합은 생물체의 유전자(DNA)를 인위적으로 절단·결합해 새로운 유전자 조합을 만들고, 이를 통해 기존에 없던 형질을 발현시키는 기술이다. 이 기술로 개발된 대표적인 의약품으로 인슐린이 있다.

54) Hybridoma 기술은 특정 항체(antibody)를 대량 생산하기 위해 세포를 융합하는 기술이다. 항체는 면역계에서 항원을 인식·결합하는 단백질이며, 항체의약품은 이를 특정 질병 표적(예: 암세포)을 공격하도록 설계한 의약품이다. 현재 글로벌 매출 상위 바이오의약품의 상당수가 항체의약품이다.

55) mRNA 백신은 인공 RNA를 인체에 전달해 바이러스 유전 정보를 발현시키고 항체 생성을 유도하는 방식으로, 약화된 바이러스를 사용하던 기존 백신에 비해 연구·생산 기간을 단축하면서도 높은 효능을 보여 코로나19 조기 극복에 크게 기여했다.

56) 최소침습수술(minimally invasive surgery)은 절개 부위를 최소화하는 수술법으로, 복강경 수술과 로봇 수술 등이 이에 해당한다. 통증 감소, 회복 속도 향상, 정밀도 개선 등의 장점이 있다.

## 참고 2 미국 및 EU의 의료 데이터 보호 및 활용 제도

미국과 EU는 모두 민감한 환자 정보를 철저히 보호하여 사회적 신뢰를 구축하고, 이를 토대로 데이터 활용을 촉진하는 방향으로 의료 데이터 제도를 발전시켜 왔다. 그러나 미국은 사전 동의와 비식별화에 의존하는 민간 중심의 분산형 데이터 활용 체계를 갖춘 반면, EU는 European Health Data Space(EHDS)를 통해 공공기관이 데이터 접근·허가를 통합 관리하는 중앙집권적 체계를 구축하고 있다.

미국의 의료 데이터 제도는 Health Insurance Portability and Accountability Act(HIPAA)를 기반으로 의료기관이 수집한 환자 개인정보를 엄격히 보호하는 동시에, 특정 조건 하에서 데이터 활용을 허용한다. 개인 식별이 가능한 모든 건강 정보, 즉 보호의료정보(Protected Health Information)는 원칙적으로 공유가 제한되지만, 다음과 같은 경우에는 예외적으로 활용할 수 있다.

- **(사전 동의)** 정보의 사용 목적, 공개 대상, 유효 기간 등에 대해 환자의 명시적 동의(authorization)를 받은 경우
- **(비식별화)** ① 이름, 주소, 생년월일 등 18개 식별 정보를 제거하거나 ② 통계 전문가로부터 재식별 위험이 매우 낮다는 서면 평가를 받은 경우. 이는 우리나라의 익명 처리와 유사하며, 데이터 브로커, 헬스 데이터 네트워크 등을 통한 의료 데이터 거래에 가장 널리 활용되는 방식이다.
- **(제한된 데이터셋)** 연구, 공중 보건, 의료기관 운영 목적으로 지역, 생년월일 등 일부 식별 정보가 포함된 제한된 데이터셋(Limited Data Set)을 사용하는 경우. 단, 재식별 금지, 계약 목적 외 사용 금지 등을 명시한 데이터 사용 계약(Data Use Agreement)을 체결해야 한다. 이 방식은 우리나라의 가명 처리와 유사하지만 활용 범위가 더 제한적이다.
- **(포괄적 동의)** Federal Policy for the Protection of Human Subjects(Common Rule)에 의거, 특정 연구 목적을 명시하지 않고, 향후 수행될 여러 연구를 위해 개인정보를 저장하고 사용하는 것에 대해 포괄적 동의(Broad Consent)를 받은 경우. 단, 개별 연구를 시작할 때마다 생명윤리위원회(IRB)의 심사를 받아야 하며, 동의 이력 관리 및 철회 지원 등 추가적인 보호 장치를 마련해야 한다. 데이터 형태를 바꾸는 우리나라의 가명 처리와 달리, 포괄적 동의는 동의 절차의 유연성을 높이는 방식이다.
- **(치료·지불·운영 목적)** 치료, 지불, 운영을 위해 다른 의료기관과 데이터를 교환하는 경우. 한편 21st Century Cures Act에서는 환자의 전자건강정보의 접근, 교환, 사용을 방해하는 정보차단행위(Information Blocking)를 금지하고 있다.
- **(공중 보건 목적)** 전염병 보고, 아동 학대 신고 등 법률에서 정한 공중 보건 활동에 필요한 경우

EU의 의료 데이터 제도는 기본법인 General Data Protection Regulation(GDPR)을 기반으로 데이터를 엄격하게 보호하는 동시에, 올해 3월 발효되어 단계적으로 시행될 European Health Data Space(EHDS)를 통해 데이터 활용을 촉진하고 있다. GDPR은 의료 데이터를 민감정보로 분류하고, 명시적 동의나 공중 보건 목적이 있을 때만 수집 및 활용을 허용한다. EHDS는 여기에 더해 각 회원국에 설치된 Health Data Access Bodies의 허가를 받으면 수집 목적 외의 2차 목적(Secondary Use)에도 데이터를 활용할 수 있도록 했다. 이는 데이터 보호와 활용을 중앙집권적으로 관리하는 방식으로, 개별 활용자가 사전 동의 및 비식별화 조치를 직접 수행하는 우리나라 제도와는 뚜렷한 차이를 보인다. EHDS의 2차 목적 데이터 활용은 다음과 같이 관리된다.

- **(활용 목적 제한)** 연구, 혁신, 공중 보건 등 공익적 목적으로만 활용이 제한되며, 보험료 산정이나 유해 제품 개발과 같이 개인에게 불이익을 주는 활용은 금지한다.
- **(보안 환경 제공)** 데이터는 외부와 단절된 보안 처리 환경(Secure Processing Environment)에서만 접근할 수 있으며, 익명 또는 가명 처리된 상태로 제공된다.
- **(데이터 수집·관리)** 데이터 보유자는 2차 목적 활용을 위해 데이터를 의무적으로 공유해야 하며, Health Data Access Bodies가 대표성, 정확성 등 데이터 품질을 검증한다.
- **(개인 선택권 보장)** 정보주체는 자신의 데이터 접근 내역을 확인할 수 있고, 특정 2차 목적 활용을 거부(opt-out)할 수 있다.

EHDS가 2차 목적 활용에 중앙집권적 관리 방식을 도입한 것은 기존 사전 동의 방식이 가진 다음과 같은 한계를 보완하기 위함이다.

- 모든 연구에서 환자의 명시적 동의를 일일이 받는 것은 시간과 비용이 과도하게 소요되어 데이터 부족과 활용 위축을 초래한다.
- 연령, 사회경제계층, 질병군 등에 따라 동의율이 달라지면 확보된 데이터가 전체 인구를 대표하지 못하는 편향 문제가 발생할 수 있다.
- 환자가 연구 목적의 복잡한 기술적·윤리적 함의를 완벽히 이해하기 어려울 수 있으며, 진료를 받기 위해 의료기관이나 기업에 어쩔 수 없이 동의해야 하는 권력 불균형 문제도 존재한다.

EHDS는 2차 목적 활용을 엄격히 통제하고 정보주체의 사후 통제권을 강화함으로써, 이러한 한계를 극복하고 의료 데이터의 보호와 활용을 동시에 촉진하고자 한다.

## 별첨 1 바이오헬스 산업 혁신 생태계와 특허출원건수·매출 간 관계 분석

바이오헬스 산업에서 자금공급, 기초연구, 전문인력 등 혁신 생태계가 기업의 기술 역량에 미치는 영향을 분석하기 위해 다음의 국가수준 패널모형으로 추정하였다.

$$\ln Y_{it} = \beta_1 \ln RND_{it-1} + \beta_2 \ln TCP_{it-1} + \beta_3 \ln TRP_{it-1} + Z_{it} \gamma + \tau_t + \epsilon_{it}$$

여기서 아래첨자  $i$ 는 국가,  $t$ 는 연도를 나타낸다.  $Y_{it}$ 는 종속변수로, 특정 국가의 바이오헬스 특허출원건수 또는 바이오헬스 기업 매출 규모를 각각 사용했다. 설명변수 중에서  $RND_{i,t-1}$ 는 바이오헬스 기업의 R&D 지출,  $TCP_{i,t-1}$ 는 바이오헬스 분야 피인용 상위 10% 우수 논문 수(Top-Cited Publications),  $TRP_{i,t-1}$ 는 바이오헬스 기업의 연구개발인력 규모(Total R&D Personnel)를 각각 의미한다.  $Z_{it}$ 는 통제변수로, GDP 캡, 고령인구(55세 이상) 규모 등을 포함했다. 마지막으로  $\tau_t$ 는 연도고정효과,  $\epsilon_{it}$ 는 오차항을 의미한다.

분석기간은 2011~2022년이며, 분석대상은 데이터 확보가 가능한 총 24개국<sup>57)</sup>(한국 포함, 미국 제외)이다. 각 변수의 정의와 출처는 다음과 같다.

[표 1] 국가패널분석 주요 변수 및 데이터 출처

구분	분석변수	데이터 출처	세부 내용
종속 변수	▶ 바이오헬스 특허출원건수	- OECD STI Micro-Lab	- 미국 특허청에 등록된 특허 기준 - WIPO IPC-기술 연계표에서 생명공학, 제약, 의료기기 등 기술로 분류된 특허의 출원건수를 국가별로 합산
	▶ 바이오헬스 기업 매출 규모	- EU Industrial R&D Investment Scoreboard	- 매년 EU가 선정한 R&D 상위 2,000개 기업 중 Health industries 기업의 수치를 국가별로 합산 - PPP 환율을 적용한 실질 기준
	▶ 바이오헬스 기업 R&D 지출	상 동	상 동
설명 변수	▶ 바이오헬스 분야 피인용 상위 10% 우수 논문 수	- OECD Bibliometric Indicators	- ASJC 학술분류상 의약학 분야(대분류코드 27, 30, 35, 36) 발간 논문을 국가별로 합산
	▶ 바이오헬스 기업 연구개발인력 규모	- OECD Research & Development Statistics	- 바이오헬스 기업은 국제표준산업분류(ISIC)상 기초 의약품 및 의약 제제 제조업(C21), 방사선조사·전자의료·전기치료 장비 제조업(C266), 의료 및 치과 기구·용품 제조업(C325)에 해당하는 기업
통제 변수	▶ GDP 캡	- Eurostat	- 인력수치는 전일제 환산(Full-Time Equivalent) 기준
	▶ 고령인구 규모(55세 이상)	- 연구개발 활동조사	- OECD 자료를 Eurostat 자료(EU 회원국)와 연구개발활동조사(한국)로 업데이트 - 일부 연도에 관측치가 없는 국가가 있음
		- OECD	- HP-filter로 산출
		- OECD	

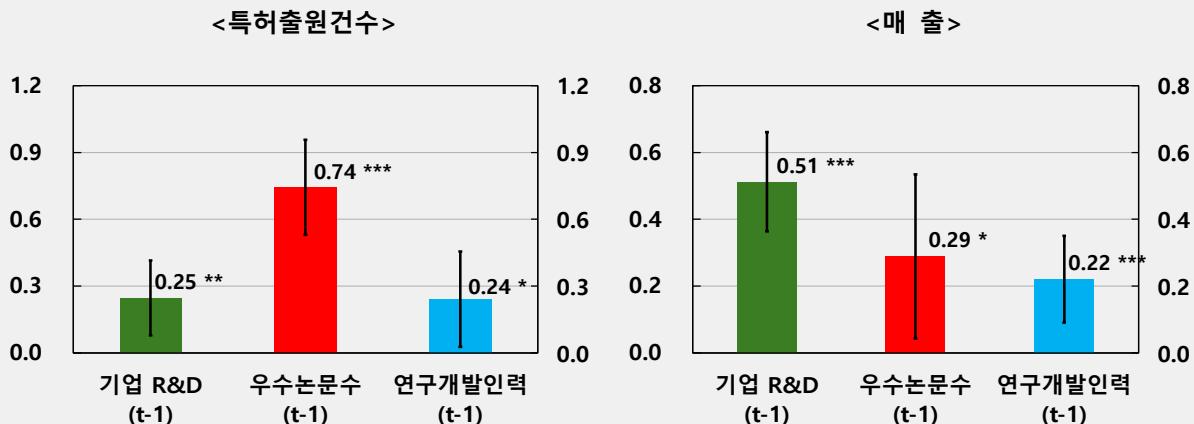
57) 24개국은 다음과 같다: 그리스, 네덜란드, 뉴질랜드, 덴마크, 독일, 벨기에, 스웨덴, 스위스, 슬로베니아, 스페인, 아일랜드, 오스트리아, 영국, 이스라엘, 이탈리아, 일본, 중국, 캐나다, 포르투갈, 핀란드, 프랑스, 헝가리, 한국, 호주.

분석 결과, 혁신 생태계를 대변하는 지표인 ① 기업 R&D 지출, ② 피인용 상위 10% 우수 논문 수, ③ 연구개발인력 규모가 기술 역량(특허출원건수) 및 상업적 성과(매출)에 모두 통계적으로 유의한 정(+)의 영향을 미치는 것으로 나타났다(그림 1).

- (기술역량) 기업 R&D 지출과 연구개발인력 규모가 각각 10% 증가할 때, 특허출원 건수는 2.5%, 2.4% 증가했다. 특히, 피인용 상위 10% 우수 논문 수가 10% 늘어 날 경우 특허출원건수가 7.4% 증가해, 탄탄한 기초연구가 바이오헬스 산업 혁신의 핵심 원천임을 보여준다.
- (상업적 성과) 혁신 생태계 확충은 매출 성장으로도 이어진다. 기업 R&D 지출이 10% 증가하면 매출이 5.1% 성장해, 자금공급이 상업적 성과에 큰 영향을 미침을 알 수 있다. 또한 우수 논문 수와 연구개발인력 규모가 각각 10% 증가할 때도 매출을 2.9%, 2.2% 늘리는 효과가 있었다.

#### 바이오헬스 산업에서 자금공급, 기초연구, 전문인력 등은 특허출원건수와 매출을 증가시키는 주요 요인

[그림 1] 바이오헬스 산업 혁신 생태계 구성요소와 특허출원건수·매출 간 관계  
- 국가패널분석 추정결과



주: 1) 데이터 확보가 가능한 2011~2022년 24개국(한국 포함, 미국 제외)을 대상으로 분석

2) 모든 변수는 로그 변환하여 분석, 3) 막대는 계수 추정치, 수직선은 90% 신뢰수준을 의미

자료: 자체 시산

## 별첨 2 바이오 데이터 보호·활용 정책과 개인의 데이터 제공 의향 간 관계 분석

### ▶ 연구 개요

건강 데이터 보호 및 활용 정책의 변화가 개인의 데이터 제공 의향에 미치는 영향을 분석하기 위해 정보 제공 실험(information provision experiment)을 실시하였다. 구체적으로, 공익적 활용 보장과 정보 통제권 강화라는 두 가지 정책 정보를 제공하고, 이에 따른 개인의 태도 변화를 측정했다.

### ▶ 실험 절차

본 실험은 그룹 간 비교 설계(between-subject design)를 적용하여, 다른 종류의 정책 정보를 고지받은 그룹 간 데이터 제공 의향 차이를 분석하였다. 모든 응답자는 본인의 건강 데이터가 의료 AI 개발에 활용되는 상황을 가정하고 실험에 참여하였다. 실험 절차는 아래 3단계로 구성된다.

- ① (공통 안내문 제시) 먼저 모든 응답자에게 건강 데이터 활용 주체(병원, 대학, 정부·공공기관, 민간 기업)와 개인이 얻을 수 있는 혜택(맞춤형 건강관리, 의료비 절감 등)을 설명하였다.
- ② (정보 처치) 응답자를 네 개 그룹 중 하나에 무작위로 배정하였다. 처치 그룹에는 세 가지 정보 중 하나를 제시하였고 대조군은 새로운 정책 정보를 받지 않았다.  
(그룹 ①) 개인정보 보호 및 데이터 통제권 강화 정책 시행  
(그룹 ②) 데이터의 공익적 활용 보장 정책 시행  
(그룹 ③) 위 두 정책 모두 시행 등  
(그룹 ④) 새로운 정책 정보 없음(대조군)
- ③ (정보 제공 의향 측정) 정보를 확인한 응답자에게 민감도 수준이 서로 다른 10개 유형의 건강 데이터<sup>58)</sup>(예: 신체 정보, 만성 질환 기록, 성매개 감염병 기록, 정신 건강 진료 기록 등)에 대한 제공 의향을 측정하였다(표 1).

[표 1] 그룹별 정보 처치 내용

그룹	처치 내용
	개인정보 보호와 데이터 통제권을 강화하는 정책이 시행된다는 정보를 제공
① 개인정보 보호 강화	- 데이터 활용 현황 공개 및 활용 거부권 보장 - 안전한 데이터 보관 및 비식별화 의무화
② 공익적 활용 보장	데이터의 활용 목적의 공익성을 보장하는 정책이 시행된다는 정보를 제공 - 신약 개발, 난치병 치료 등 의료·바이오 발전 - 국민 건강 위험 예측(만성질환 예방, 건강격차 해소) 등 사회 전체의 건강 향상
③ 두 정책 모두 시행	- 위 ①, ②의 정보를 모두 제공

## ▶ 분석 모형

본 실험 설계를 통해 각 처치 그룹에 제공된 추가 정보가 개인의 데이터 제공 의향 변화에 미치는 인과효과(causal effect)를 아래의 계량모형으로 추정할 수 있다.

$$Y_i = \alpha + \beta^k Treatment^k + X_i + \epsilon_i$$

여기서 아래 첨자  $i$ 는 응답자를,  $k$ 는 정보 처치에 따른 그룹( $k = \{1, 2, 3\}$ )을 나타낸다. 종속변수  $Y_i$ 는 정보 제공 여부 더미변수이며, 설명변수  $Treatment^k$ 는 각 정보 처치를 받았는지 여부를 보여주는 더미변수이다.  $\beta^k$ 는 본 연구의 관심 계수로, 그룹  $k$ 에 제공된 특정 정보가 개인의 데이터 제공 의향에 미친 평균적인 그룹간 효과(between-subject effect)를 보여준다. 즉, 새로운 정책에 대한 정보를 받지 못한 대조군에 비해 새로운 정책 정보를 받은 처치군 간에 데이터 제공 의향이 얼마나 차이를 보이는지 측정한다.  $X_i$ 는 통제변수 벡터로, 프라이버시 민감도, AI 기술 신뢰도와 인구사회학적(직업, 소득, 가족관계 등) 개인 특성을 나타내며,  $\epsilon_i$ 는 오차항이다.

## ▶ 설문조사 개요

본 연구를 위한 설문조사는 다음과 같이 실시하였다.

- (조사 대상) 전국 만 19세 이상 성인 남녀
- (조사 기간) 2025년 11월 11일 ~ 2025년 11월 13일
- (조사 방법) 온라인 설문조사
- (표본 크기) 총 300명<sup>59)</sup>
- (조사 기관) (주)한국리서치

58) 10개 유형의 건강 데이터는 다음과 같다: 신체 정보(키, 몸무게 등), 일상 생활 습관(운동, 식단, 수면 기록), 건강 습관(예방접종, 흡연 · 음주 이력), 경증 질환 기록(감기 등), 만성 질환 기록(고혈압, 당뇨 등), 중증 질환 기록(암 등), 유전체 정보, 생식 관련 질환 기록(불임 등), 성매개 감염병 기록, 정신건강 진료 기록.

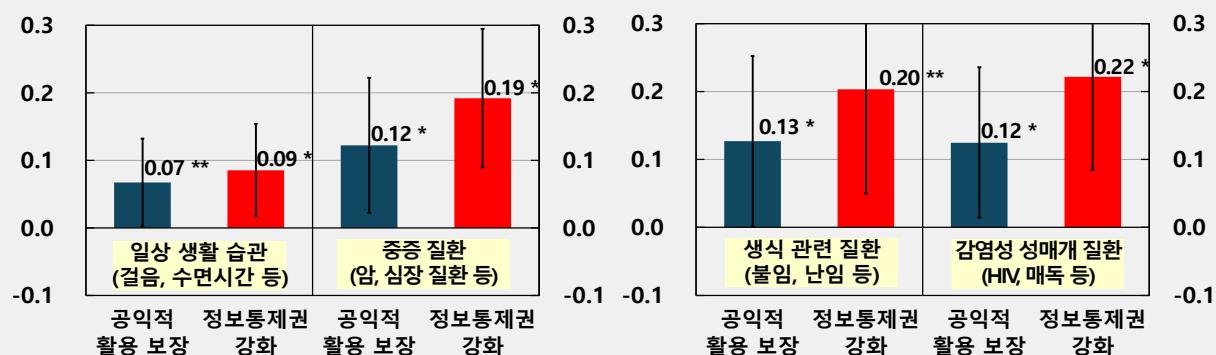
59) 본 결과는 예비조사 단계에서 집계된 것이며, 약 4천 명을 대상으로 실시될 최종 조사 결과는 성원 · 최이슬 (2026)에 포함될 예정이다.

## ▶ 강건성 검증

분석 결과의 강건성을 검증하기 위해 다양한 유형의 건강 정보를 대상으로 정보 제공 의향을 분석하였다. 그 결과, 상대적으로 덜 민감한 일상생활습관 정보(예: 걸음수, 수면시간)와 일반 질환 정보(예: 암, 심장질환)의 경우, 공익적 활용 보장 정책 및 정보 통제권 강화 정책에 대한 정보만 제공받은 그룹 모두에서 정보 제공 의향이 통계적으로 유의하게 증가하였다(그림 1). 이와 유사하게, 상대적으로 민감한 정보에 해당하는 생식 관련 질환(예: 불임, 난임)과 감염성 성매개 질환(예: HIV, 매독)에 대해서도 두 정책 정보만 제공받은 그룹 모두에서 정보 제공 의향이 통계적으로 유의하게 증가하였다(그림 2). 한편, 민감 정보의 경우 공익적 활용 보장보다 정보 통제권 강화가 정보 제공 의향에 더 크게 영향을 미쳐 정책 효과가 보다 우수한 것으로 분석되었다.

### 민감 정보의 경우 공익적 활용 보장보다 정보 통제권 강화가 정보 제공 의향 더 크게 영향

[그림 1] 바이오 데이터 보호·활용 정책의 변화가 데이터 제공 의향에 미치는 영향



주: 1) 막대는 효과의 평균 추정치를, 세로선은 90% 신뢰구간을 나타냄

자료: 자체 시사

한편, 응답자를 대상으로 건강정보의 금전적 가치를 5천 원 미만에서 10만 원 사이의 구간으로 제시해 설문한 결과, 민감 정보에 해당하는 정신건강 진료 기록의 인지된 금전적 가치는 신체 정보에 비해 약 12.1% 높게 나타났다(표 2).

[표 2] 건강정보의 금전적 가치

구분	정보 예시	금전적 가치
1. 신체 정보	키, 몸무게 등	56,875원
2. 일상적 생활습관	걸음수 등 운동기록, 식단기록 등	59,688원
3. 중증 질환	암, 심장질환, 뇌혈관질환 등	57,589원
4. 감염성 성매개 감염 검사·진단	HIV, 매독 등 검사 결과	62,500원
5. 정신건강 진료 기록	우울증, 불안장애, 조현병 등	63,800원

## 참고문헌

- 보건복지부. (2021). 보건의료 데이터·인공지능 혁신전략 [보도자료]. 2021년 6월 3일.
- 보건복지부. (2023). 보건의료 분야 결합전문기관 재지정으로 보건의료데이터 활용 촉진 [보도자료]. 2023년 11월 1일.
- 보건복지부. (2025). 국민건강보험공단 종합감사 처분요구서.
- 이동원, 성원, 정종우, 최이슬, 김동재, & 조태형. (2024). 혁신과 경제성장: 우리나라 기업의 혁신활동 분석 및 평가. 한국은행 경제전망보고서 중장기 심층연구. 한국은행.
- 이상현, 이후석. (2022). 의료 마이데이터 도입이 보건의료 데이터산업에 미치는 영향 - 경제적 파급효과 분석을 중심으로. 보건산업브리프 Vol. 367. 한국보건산업진흥원. 2022년 10월.
- 이은경, 천동민. (2024). 우리 경제의 잠재성장을 과 향후 전망. BOK 이슈노트. 한국은행.
- 한국데이터산업진흥원. (2021~2024). 데이터산업 현황조사.
- Acemoglu, D., & Linn, J. (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *The Quarterly Journal of Economics*, 119(3), 1049 - 1090.
- Acemoglu, D., Cheema, A., Khwaja, A. I., & Robinson, J. A. (2020). Trust in state and non-state actors: Evidence from dispute resolution in Pakistan. *Journal of Political Economy*, 128(8), 3090 - 3147.
- Acemoglu, D., Makhdoomi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4), 218 - 256.
- Association for Accessible Medicines. (2024). The U.S. generic & biosimilar medicines savings report. September 2024.
- Athey, S., Catalini, C., & Tucker, C. E. (2017). The digital privacy paradox: Small money, small costs, small talk. NBER Working Paper No. 23488. National Bureau of Economic Research.
- Blume-Kohout, M. E. (2012). Does targeted, disease-specific public research funding influence pharmaceutical innovation? *Journal of Policy Analysis and Management*, 31(3), 641 - 660.
- Cao, J., Zhang, X., Shahinian, V., Goutman, S. A., Tsoi, L. C., Wiens, J., & Jung, K. (2022). Generalizability of an acute kidney injury prediction model across health systems. *Nature Machine Intelligence*, 4, 1121 - 1129.

- Christian, B. (2020). The alignment problem: Machine learning and human values. W. W. Norton & Company.
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47, 20 - 33.
- Eronen, E., Värri, A. O., & Järvinen, M. (2025). The secondary use of the Finnish national health data repository Kanta - opportunities and obstacles. *Finnish Journal of EHealth and EWelfare*, 17(2), 188 - 199.
- European Commission. (2022). Study to support an Impact Assessment on enhancing the use of data in Europe.
- European Commission. (2025). Reuse of health data.
- Goldacre, B., & Morley, J. (2022). Better, broader, safer: Using health data for research and analysis. Department of Health and Social Care.
- Jayatunga, M. K. P., Ayers, M., Bruens, L., Jayanth, D., & Meier, C. (2024). How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. *Drug Discovery Today*, 29(6), 103986.
- Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9), 2819-2858.
- Li, S., Drew, J., & Chancellor, D. (2024). The \$120bn biosimilar opportunity. Evaluate Pharma. October 2024.
- Miller, A. R., & Tucker, C. (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science*, 55(7), 1077 - 1093.
- Miller, A. R., & Tucker, C. (2018). Privacy protection, personalized medicine, and genetic testing. *Management Science*, 64(10), 4648 - 4668.
- Myers, K., & Pauly, M. (2019). Endogenous productivity of demand-induced R&D: Evidence from pharmaceuticals. *The RAND Journal of Economics*, 50(3), 591 - 614.
- National Security Commission on Emerging Biotechnology. (2025). Charting the future of biotechnology: An action plan for American security and prosperity.
- Nindl, E., Napolitano, L., Confraria, H., Rentocchini, F., Fako, P., Gavigan, J., & Tubke, A. (2024). The 2024 EU industrial R&D investment scoreboard. Publications Office of the European Union.
- Oderkirk, J. (2021). Survey results: National health data infrastructure and governance (OECD Health Working Papers, No. 127). OECD Publishing.

- OECD. (2022). Towards an integrated health information system in Korea. OECD Publishing.
- OECD. (2025). OECD Economic Outlook. Volume 2025 Issue 1: Tackling Uncertainty, Reviving Growth. OECD Publishing.
- Pairolero, N. A., Giczy, A. V., Torres, G., Islam Erana, T., Finlayson, M. A., & Toole, A. A. (2025). The artificial intelligence patent dataset (AIPD) 2023 update. *Journal of Technology Transfer*.
- Rodriguez, A., Meier, C., Serazin, E., Gooch, J., Aggarwal, P., Steene, A., Jayatunga, M., Nayak, S., & Randall, W. (2023). Unlocking the potential of AI in drug discovery. Boston Consulting Group.
- Saini, J., Thakur, A., & Yadav, D. (2025). AI-driven innovations in pharmaceuticals: Optimizing drug discovery and industry operations. *RSC Pharmaceutics*, 2(3), 437 – 454.
- Slawomirski, L., de Looper, M., Santaeulàlia, M. J., & Yoshida, Y. (2023). Progress on implementing and using electronic health record systems: Developments in OECD countries as of 2021. *OECD Health Working Papers*, No. 160. OECD Publishing.
- Stern, A. D. (2017). Innovation under regulatory uncertainty: Evidence from medical technology. *Journal of Public Economics*, 145, 181 – 200.
- Sudlow, C. (2024). Uniting the UKs health data: A huge opportunity for society. Zenodo.
- Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7), 3049 – 3062.
- Tony Blair Institute for Global Change. (2024). A new national purpose: Harnessing data for health.
- Toole, A. A. (2007). Does public scientific research complement private investment in research and development in the pharmaceutical industry? *Journal of Law and Economics*, 50(1), 81 – 104.
- Wellcome Trust, & BCG. (2023). Leading barriers for AI adoption in drug discovery in high-income countries worldwide as of 2023.
- World Economic Forum. (2019). How data will improve healthcare without adding staff or beds. In *Global innovation index 2019: Creating healthy lives - The future of medical innovation* (Chap. 8).