



보험업의 데이터 결합·활용 사례 및 시사점 : 의료데이터를 중심으로

박희우·이승주

본 보고서에 수록된 내용은 집필자 개인의 의견이며 우리원의 공식 의견이 아님을
밝혀 둔다.



목 차

• 요약	1
I. 서론	2
II. 보험업의 데이터 활용	4
1. 개요	4
2. 해외 보험회사의 외부데이터 결합·활용 사례	10
3. 국내 보험회사의 외부데이터 결합·활용 현황	23
III. 의료데이터의 이차적 활용	29
1. 개요	29
2. 국내 현황	31
3. 해외 사례	37
4. 시사점	44
IV. 결론	51
• 참고문헌	53

표 차례

〈표 II-1〉 보험회사 내부데이터 특징	8
〈표 II-2〉 보험회사가 결합·활용 가능한 외부데이터의 특징	9
〈표 II-3〉 LexisNexis Risk Classifier의 데이터 변수에 따른 사망률	12
〈표 II-4〉 LexisNexis와 ExamOne의 각 모델별 사용 데이터	14
〈표 II-5〉 LabPiQture의 데이터를 기반으로 한 보험 가입자의 사망률 분포	16
〈표 II-6〉 인텐트 데이터 분류	19
〈표 II-7〉 해외 보험회사의 외부데이터 활용 사례 요약	22
〈표 II-8〉 보험업권 공공데이터 활용 연혁	26
〈표 III-1〉 의료데이터의 종류	29
〈표 III-2〉 개인정보, 가명정보, 익명정보의 개념 및 활용 가능 범위	30
〈표 III-3〉 공공의료데이터 거버넌스 상위 국가의 주요 항목별 점수	34

그림 차례

〈그림 II-1〉 보험업 가치사슬에서 빅데이터 분석의 활용 여부	4
〈그림 II-2〉 보험업의 데이터 활용	6
〈그림 II-3〉 시간의 흐름에 따른 개별·집합 데이터 가치의 변화	8
〈그림 II-4〉 연체 잔액과 상대사망률(Relative A/E) 사이의 관계	13
〈그림 II-5〉 모델별 10분위수 그룹의 상대적 사망률	15
〈그림 II-6〉 보험회사의 수입보험료 중 인보험 비중	27
〈그림 III-1〉 국가별 공공의료데이터 가용성 및 거버넌스 점수 분포	33
〈그림 III-2〉 국가별 보험소비자의 긍정적·중립적·부정적 경험 비교(2015년)	36
〈그림 III-3〉 유럽 의료데이터의 이차적 활용: 국가별 점수 분포	38
〈그림 III-4〉 일본 차세대 의료기반법 체계	43
〈그림 III-5〉 노년부양비가 가장 높은 10개 국가(2019년, 2050년)	45
〈그림 III-6〉 의료데이터의 이차적 활용을 위한 필요 요건	45
〈그림 III-7〉 보험회사에 대한 소비자 신뢰도와 데이터 공유 선호도	47
〈그림 III-8〉 개인정보보호와 데이터 활용에 따른 효용	48

The Use of Data in the Insurance Industry

The use of data is the most important in the insurance industry. Insurers utilize internal data and external data, while the use of external data is not easy for the information subject to consent to the use of personal information. As legislative improvement is being promoted, expectations are growing that the insurance industry can utilize external data from different industries. In order to predict the benefits of using external data, overseas insurers' cases can be looked at, and overseas insurers are using external data to achieve various outcomes such as elaboration of premiums, improvement of underwriting efficiency, and prevention of insurance fraud.

Korean insurers are promoting the use of national health data as the first step to using external data. The availability and governance of national health data in Korea is evaluated as excellent in the world. On the other hand, data sharing with for-profit companies is not taking place, which seems to be due to the lack of social consensus on the use of national health data. In Finland, where social trust is high, health data is being used for innovation activities, but large-scale projects have failed in the UK, where social consensus is lacking.

As Korea is aging the fastest in the world, it is highly likely that the number of the elderly and the sick will increase in the insurance market, which will widen the insurance coverage gap. Utilization of national health data can serve as a cornerstone for approval of insurance subscriptions for the underprivileged, expansion of coverage, and discount on premiums. In Korea, the availability and governance of public health data is already in place, so it is necessary to form a social consensus.

보험업은 여러 분야에서 발생하는 위험을 통계적으로 분석하여 상품 혹은 서비스를 제공하기 때문에 데이터의 활용이 가장 중요한 산업이다. 보험회사는 영위하는 서비스로부터 직접 취득한 내부데이터와 다른 사업자로부터 취득한 외부데이터를 활용한다. 외부데이터는 개인정보 활용을 위한 정보주체의 선제적 동의가 쉽지 않아 활용에 제약이 있었지만, 제도적 개선이 추진됨에 따라 보험업권에서도 이중 업종 외부데이터의 결합과 활용을 통해 다양한 부가가치를 창출할 수 있다는 기대감이 커지고 있다. 보험회사가 외부데이터를 활용하여 기대할 수 있는 편익을 예상하기 위해서 해외 사례를 살펴볼 수 있으며, 해외 보험회사는 외부데이터를 활용하여 보험료 정교화, 언더라이팅 효율성 개선, 보험사기 방지 등 다양한 성과를 거두고 있다.

우리나라 보험회사는 외부데이터 활용을 위한 첫 단추로서 공공의료데이터의 활용을 추진 중이다. 우리나라 공공의료데이터의 가용성, 인프라 및 제도, 거버넌스는 세계적으로도 우수한 수준으로 평가받고 있지만, 데이터 활용에 대한 사회적 공감대가 부족하여 영리기업과의 공공의료데이터 공유는 이루어지지 않고 있다. 해외의 경우에도 사회적 신뢰가 높은 핀란드에서는 의료데이터를 공유하여 혁신 활동에 활용하고 있지만, 사회적 합의가 부족한 영국에서는 의료데이터 활용을 통한 대규모 프로젝트가 불발되기도 하였다.

한국은 세계에서 가장 빠르게 고령화가 진행되고 있기 때문에 보험시장에서 고령자·유병자가 증가하여 보험 보장 공백이 확대될 가능성이 높다. 공공의료데이터의 활용은 취약계층에 대한 보험 가입 승인과 보장 범위의 확대, 보험료 할인 등을 위한 초석이 될 수 있다. 한국에서는 이미 공공의료데이터의 가용성과 거버넌스가 갖추어졌기 때문에 사회적 공감대를 형성하여 데이터를 활용할 수 있다면 그 잠재가치가 높다고 판단된다. 사회적 공감대 형성은 쉽지 않을 것이지만, 정보보호 체계에 대한 홍보·안내를 강화하여 사회적 신뢰 제고를 위한 여건을 조성하거나 정보주체에 대한 이익 배분 방안을 고려할 수 있을 것이다.

4차 산업혁명이 본격적으로 도래하며 데이터는 경제, 금융, 의료, IT 등 모든 분야에서 핵심 가치로서 활용되고 있다. 데이터 활용 제도, 역량, 인프라 등은 기업, 산업, 국가 단위에서 경쟁력 확보를 위한 중요한 요소가 되었다.

보험업에서도 데이터의 활용을 통해 차별화된 상품과 서비스를 제공하고 있다. 보험회사는 직접 영위하는 보험 사업을 통해 수집한 내부데이터와 다른 사업자로부터 취득한 외부 데이터 등 다양한 데이터를 활용하고 있다. 최근 IoT 기술의 발달로 웨어러블, 텔레메틱스 등을 통해 데이터를 수집하기도 하며 해외의 경우 보험회사가 신용, 유통, 의료 등 이종 업종 데이터를 결합·활용하여 상품 및 서비스를 제공하기도 한다.

국내 보험회사의 경우 이종 업종 외부데이터의 결합·활용 사례가 많지 않다. 보험업에서는 외부데이터 중 특히 개인정보의 활용이 중요하지만, 개인정보 유출 시에 발생할 수 있는 피해가 매우 크기 때문에 보호를 위한 엄격한 규제가 적용되기 때문이다. 하지만 최근 개인정보 활용의 필요성에 대한 공감대가 형성되었고, 정부는 개정 데이터 3법과 후속 제도적 개선을 추진하여 가명정보를 정보주체의 동의 없이 활용할 수 있는 근거를 마련하였다. 이에 따라 우리나라 보험회사도 외부데이터의 결합·활용을 통한 가치 창출의 기반이 구축되었다.

보험회사는 외부데이터 활용의 첫 걸음으로 공공의료데이터의 사용을 추진 중이다. 보험회사는 공공의료데이터를 통해 내부 경험통계 등에서 제공하지 못하였던 데이터를 확보함으로써 보험 가입이 거절되거나 높은 보험료를 지불해야만 보험에 가입할 수 있었던 유병자·고령자의 보험 가입 문턱을 낮출 수 있다. 또한 의료데이터를 분석하여 희귀질환 보장 상품을 개발하거나 헬스케어 서비스를 제공하여 보장 확대 및 국민 건강 증진에도 기여할 수 있게 된다. 하지만 우리나라에서는 보험회사 등 영리기업의 개인정보 활용에 대한 사회적 인식이 좋지 않으며, 개인정보 유출 및 침해 등에 대한 우려가 높아 보험회사는 공공의료데이터를 본격적으로 활용하지 못하고 있는 것으로 보인다.

1) 본고에서 '정보'와 '데이터'의 의미는 혼용하여 사용함

보험업에서 공공의료데이터의 활용에는 여전한 제약이 존재하고 있지만, 이에 대한 원인을 객관적으로 살펴본 연구는 많지 않다. 보험회사의 공공의료데이터 활용을 위한 국내 제도에 모호성 혹은 미흡함이 존재하기 때문일 수도 있으며, 실질적인 활용을 위한 명시적인 근거가 부족하거나 혹은 제도 외에 사회적 신뢰 등 다른 요인 때문일 수도 있을 것이다. 제도적 근거가 부족하다면 데이터 활용에 따른 편익과 비용을 비교하여 개선 방안을 생각해볼 수 있을 것이며, 제도와는 다른 요인이 원인으로 지목된다면 이와 관련하여 시사점을 도출할 수 있다.

본고는 보험회사의 관점에서 내부·외부 등 데이터의 종류를 나누고 그 특징과 이슈사항을 종합적으로 정리하였다는 점에서 의미가 있다. 또한 우리나라 보험회사의 실질적인 외부데이터 결합·활용을 앞두고 향후 기대할 수 있는 편익을 해외 선진 보험회사의 사례를 통해 알아보았다. 이어서 공공의료데이터에 대한 보험업의 수요가 높은 이유를 정리하고, 우리나라 공공의료데이터의 질적 측면, 이차적 활용을 위한 제도적 현주소, 문제점 등을 국제적인 비교와 함께 종합적으로 평가하여 시사점을 도출하였다. 보험업의 외부데이터 활용 사례와 공공의료데이터를 실질적으로 활용하기 위한 시사점을 검토해 보았다는 점에서 본고는 기여점이 있을 것이다.

본고의 구성은 다음과 같다. II장에서는 보험회사가 활용하는 데이터를 내부데이터 및 외부데이터로 나눠서 각 특징을 소개할 것이다. 특히 외부데이터는 정보주체의 선제적 동의가 쉽지 않기 때문에 개인정보보호 및 활용과 관련한 제도적 이슈가 존재하여 국내에서는 실질적 활용이 이루어지지 못하였는데, 선진 해외 보험회사가 다양한 외부데이터를 활용하는 사례를 소개하여 향후 우리나라 보험회사가 외부데이터를 활용하여 얻을 수 있는 편익을 알아볼 것이다. 이와 더불어 우리나라 보험회사의 외부데이터 활용과 관련하여 제기되었던 논의 과정과 활용 연혁을 살펴보고, 특히 보험업권이 공공의료데이터의 활용에 대한 수요가 높은 요인에 대해서도 알아보고자 한다.

III장에서는 의료데이터의 이차적 활용과 관련하여 국내 현황과 해외 사례를 살펴볼 것이다. 국내 현황에서는 제도적 특징을 이해한 뒤 OECD 주요국과 비교하여 공공의료데이터의 가용성과 거버넌스의 측면에서 현주소를 파악하고자 한다. 또한 몇몇 국가에서는 사회적 신뢰 수준에 따라 의료데이터 활용의 성패가 결정되기도 하였는데 이러한 사례에 대해서도 소개할 것이다. 분석 내용을 바탕으로 우리나라에서 공공의료데이터의 실질적 활용에 제약이 존재하는 원인을 이해하고 보험회사가 공공의료데이터를 효율적으로 활용하기 위한 시사점을 도출하고자 한다.

II

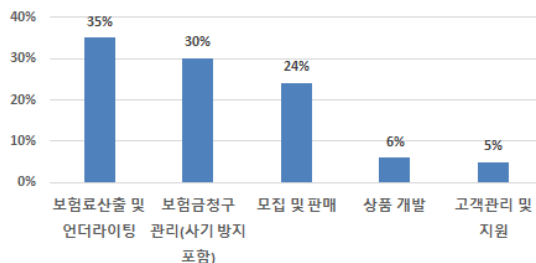
보험업의 데이터 활용

1. 개요

보험은 사망, 상해, 질병, 화재 등 불의의 사고로 인한 경제적 손실을 보전하기 위해 많은 사람이 합리적인 보험료를 모아 공동기금을 형성하고 사고를 당한 계약자에게 보험금을 지급하는 제도이다.²⁾ 이때 보험회사는 경험생명표 및 사고의 형태별 빈도·강도 등의 경험통계를 비롯한 다양한 데이터를 활용하여 수지상등의 원칙, 대수의 법칙을 기초로 보험료를 산출한다.

보험업은 여러 분야에서 발생하는 위험을 통계적으로 분석하여 새로운 상품을 개발하거나 보험료율을 산출하기 때문에 데이터의 활용이 가장 중요한 산업이며, 보험회사는 다양한 데이터를 활용함으로써 새로운 가치를 창출할 수 있다. 다시 말해 보험회사가 적합한 데이터를 갖추지 못할 경우 혁신적인 상품이나 서비스의 개발은 불가능할 것이다. 보험회사는 보험업의 전체 가치사슬인 상품개발, 모집 및 판매, 보험료 산출 및 언더라이팅, 서비스 및 고객관리, 보험금지급 및 보험사기 방지 등에서 다양한 데이터를 활용하고 있다.

〈그림 II-1〉 보험업 가치사슬에서 빅데이터 분석의 활용 여부



주: EU 28개국 222개 보험회사에 대한 설문조사 결과임
자료: EIOPA(2019)

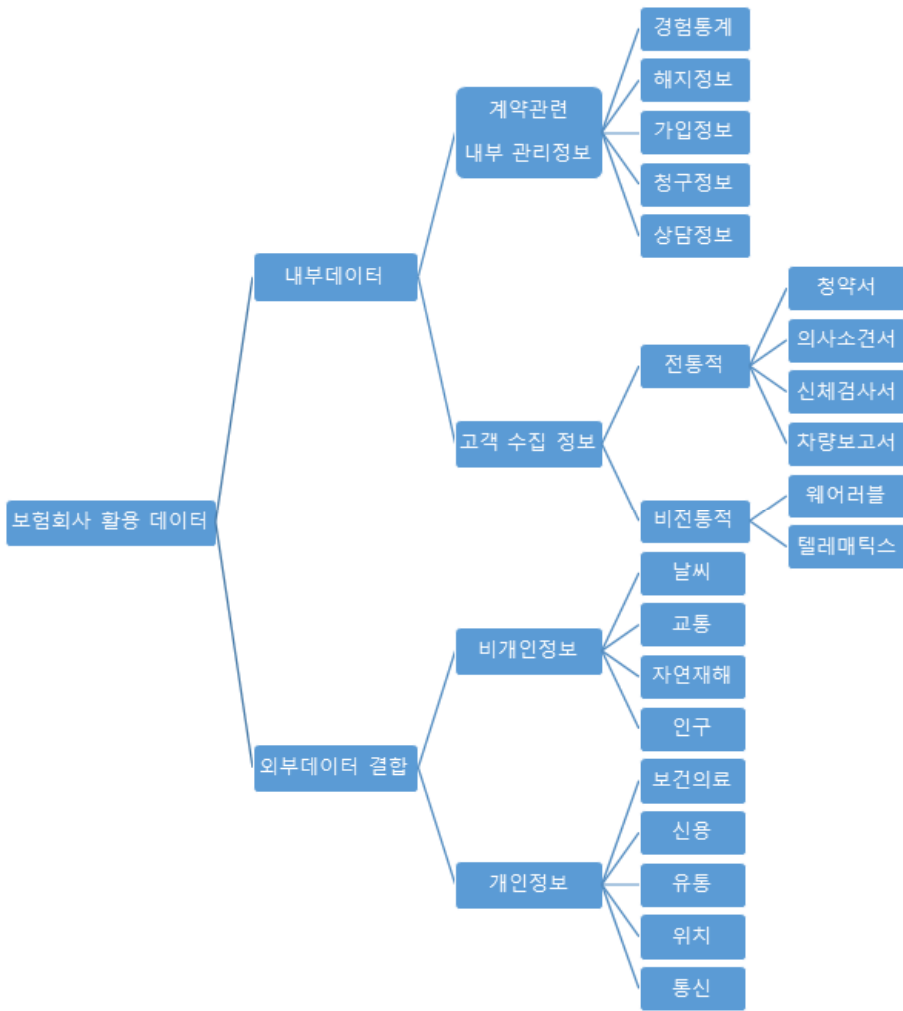
2) 생명보험협회(2022)

최근에는 보험산업에서도 디지털 전환이 이루어지며 웨어러블, 텔레메틱스 등 IoT 기술을 활용한 보험상품 및 서비스가 등장하고 있으며, 보험회사는 보건의료·신용·유통 등 이 중 업종 데이터를 결합하여 보험상품 개발, 보험료 산출 및 언더라이팅에 활용하기도 한다. 보험회사가 활용하고 있거나 활용할 수 있는 데이터는 <그림 II-2>와 같이 내부(Internal) 및 외부(External)데이터로 분류하여 나타낼 수 있다.

내부데이터는 보험회사가 영위하는 보험 사업 및 서비스로부터 직접 취득한 데이터를 의미하고, 외부데이터는 보험회사가 다른 사업자로부터 취득하는 데이터이다. 이때 내부데이터는 보험계약자 혹은 피보험자의 계약관련 내부 관리정보와 고객 수집 정보로 나눌 수 있다. 계약관련 내부 관리정보에는 나이, 성, 사망률 등 경험통계 정보와 해지, 가입, 청구 정보 등이 포함되며, 고객 수집 정보는 전통적으로 보험계약자가 보험 가입 시에 제출하는 서류인 청약서, 의사소견서, 신체검사서, 차량보고서 등을 비롯하여 IoT 기술을 활용하여 기존과는 다른 방식인 웨어러블, 텔레메틱스 등을 통해 고객으로부터 수집하는 정보가 있다.

외부데이터는 비개인정보와 개인정보로 나눌 수 있다. 비개인정보는 날씨, 교통, 자연재해, 인구 정보 등이 있고, 개인정보에는 보건의료, 신용, 유통, 위치, 통신정보 등의 업종에서 소비자 개인별 정보가 포함된다. 내부데이터는 보험회사가 보험계약자로부터 직접 정보 제공 및 이용에 대한 동의를 받는다는 점에서 외부데이터와 가장 큰 차이가 있다고 할 수 있다.

〈그림 II-2〉 보험업의 데이터 활용



가. 내부데이터

내부데이터는 보험회사가 보험계약자로부터 직접 취득하는 데이터이다. 보험계약자가 가입한 보험상품으로부터 얻어진 정보인 계약 관련 내부 관리 정보에는 나이, 성별, 사망률, 질병률 등 경험통계를 비롯하여 계약철회, 해지, 청구, 가입 정보 등이 포함된다. Willis Towers Wire의 설문조사에 따르면 보험회사는 계약 관련 내부 관리 정보로 청구이력 정보

(77%), 계약자 관련 정보(55%), 가입 정보(46%) 등을 이용한다.³⁾

고객 수집 정보는 보험계약자가 보험 가입 시에 보험회사에 추가로 제출하는 정보이며, 전통적 정보에는 건강·자동차 보험사고 발생 가능성과 관련이 높은 것으로 알려진 과거 병력 관련 정보 혹은 차량 관련 정보, 사고이력 등이 포함된다. 이와 더불어 IoT 기술이 발전함에 따라 보험회사는 전통적 정보 외에도 웨어러블 기기, 텔레매틱스 등을 활용하여 신체활동 관련 정보(걸음 수, 소모 칼로리, 심박수 등), 차량 주행 정보(주행거리, 주행시간 등), 운전 습관 정보(차로 이탈, 과속, 급제동 등)를 활용하여 보험계약자의 보험료를 할인해주거나 맞춤형 상품 혹은 부가서비스를 제공하기도 한다. 예를 들어, 자동차보험에서는 텔레매틱스를 통해 실시간 데이터를 활용하여 주행거리·시간, 가속, 과속 등을 보험료 산정 및 언더라이팅에 반영한 UBI(Usage-Based Insurance)보험이 판매되고 있으며, 운전자의 부주의한 운전습관 등을 반영한 BBI(Behaviour-Based Insurance)보험의 활성화도 기대되고 있다.⁴⁾ 이러한 보험상품은 기존의 평가 방식에서 고위험군으로 분류한 고객들의 운전행태를 평가하여 보험료를 할인해주거나 다양한 부가서비스(운전습관 피드백 서비스, 자동 사고신고 서비스 등) 제공을 통해 상품을 차별화하기도 한다.

내부데이터는 정보주체인 보험계약자가 상법 제651조 및 제651조의2에 근거하여 계약 전 알릴 의무(고지의무)에 따르거나 보험료 할인 등을 위해 보험계약 체결을 위한 개인정보 활용에 동의한 후 수집되게 된다. 보험회사는 피보험자의 사고위험 평가를 통한 적정 보험료 및 보험금 산정, 인수심사, 보험금 청구 및 지급관리 등 보험업무를 목적으로 보험계약자 및 피보험자의 정보를 활용할 수 있다.

보험회사는 자체적으로 수집하는 데이터의 범위를 넓힘으로써 보험계약자 및 피보험자의 적정 위험 수준을 반영하여 보험료를 산정하거나 사업비 절감 혹은 고객 특성과 니즈에 맞춘 신상품 개발 등의 기대 효과를 누릴 수 있으며, 보험금 청구 및 지급 과정을 간소화하거나 보험사기를 방지하여 보험소비자와 사회 전체의 편익도 높일 수 있다. 하지만 청약서, 건강진단서 등에 수록된 건강정보를 비롯하여 웨어러블 기기, 텔레매틱스 등에는 민감한 개인정보가 포함되어 있기 때문에 보험회사가 데이터를 활용하는 데 있어 개인정보 유출 방지 및 보호를 위해 각별히 신경써야 한다는 우려 또한 존재한다.⁵⁾

3) 금융감독원(2019)

4) 안소영(2020)

5) 손지영(2021)

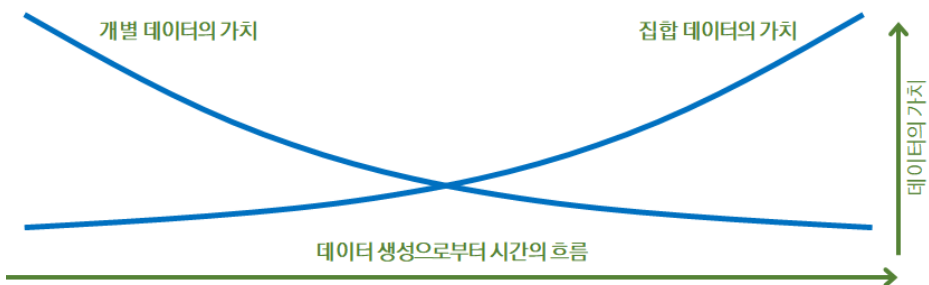
〈표 II-1〉 보험회사 내부데이터 특징

구분	특징
계약 관련 내부 관리정보	● 보험계약자가 가입한 보험상품으로부터 얻어진 정보로 경험통계, 청구, 해지, 상담 정보 등이 있음 ● 보험상품 개발, 보험료 산출, 보험금 청구 및 지급(보험사기 포함) 관리를 위해 경험 통계 등의 정보를 활용할 수 있음
고객 수집 정보	● 보험계약자가 보험계약 체결 혹은 보험료 할인을 위해 보험회사에 제공하는 정보임 ● 정보주체가 개인정보 활용에 동의한 후 정보를 활용하기 때문에 개인정보보호 및 활용 관련 이슈는 많지 않음 ● 건강·행태 등 민감한 정보가 포함되기 때문에 보안 강화 및 정보 유출 방지에 대한 주의가 필요함

나. 외부데이터

보험회사는 다른 사업자가 수집한 외부데이터를 내부데이터와 결합하여 활용하기도 한다. 보험업은 다양한 분야에서 발생하는 사고의 정도와 빈도 등에 대한 예측이 필요하여 보험회사 내부데이터 외에도 이종 업종 데이터에 대한 수요가 높기 때문이다. 〈그림 II-3〉과 같이 다양한 업종의 데이터를 결합한 집합 데이터는 데이터 생성으로부터 시간이 경과하여도 일반적으로 개별 데이터에 비해 그 가치가 높고, 다른 데이터와 결합해야만 가치가 드러나는 데이터도 존재한다.⁶⁾

〈그림 II-3〉 시간의 흐름에 따른 개별·집합 데이터 가치의 변화



자료: Gil Press(2013)

6) 최상아(2019)

보험회사가 결합하여 활용하는 외부데이터는 비개인정보와 개인정보로 나눌 수 있으며, 비개인정보에는 날씨, 교통, 인구, 자연재해 정보 등이 포함된다. 비개인정보는 개인정보 보호법, 신용정보법 등에서의 개인정보보호 조항과는 관련이 없기 때문에 보험회사는 자체적으로 데이터를 수집하거나 데이터 보유 기관과의 협력을 통해 비개인정보를 수집·활용할 수 있다. 개인정보는 개인을 알아볼 수 있는 정보 혹은 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보로 정의되며, 보험회사는 보험 모집단계부터 보험금 지급단계에 이르기까지 전 가치사슬에서 개인정보를 다양하게 활용하고 있다.

한편 보험회사는 외부 및 내부 데이터 모두에서 개인정보를 활용하고 있지만, 각 경우 직면하게 되는 개인정보의 보호 및 활용과 관련한 법적 이슈에서 차이가 있다. 내부데이터-개인정보는 외부데이터-개인정보와는 다르게 정보주체인 보험계약자 및 피보험자의 정보이용 동의를 받은 뒤 그 범위 내에서 정보를 활용한다는 데 특징이 있다. 보험계약자는 보험 가입을 위해서 개인정보 활용에 필수적으로 동의를 해야 하므로 내부데이터-개인정보는 보험회사가 목적에 맞게 활용할 수 있다. 하지만 외부데이터는 개인정보 활용을 위한 정보주체의 선제적 동의가 쉽지 않고, 정보주체의 동의가 동반되지 않을 경우 개인정보보호 관련 법률 등에 따라 활용에 여러 가지 제약이 따른다. 우리나라 보험회사는 보건의료, 신용정보 등 대부분 이중 업종 외부데이터를 자유롭게 활용하지 못하고 있다.

〈표 II-2〉 보험회사가 결합·활용 가능한 외부데이터의 특징

구분	특징
비개인정보	● 날씨, 교통, 인구, 자연재해 정보 등이 있음 ● 개인정보보호 법률과는 관련이 없기 때문에 보험회사는 자체적으로 데이터를 수집하거나 데이터 보유 기관과의 협력을 통해 비개인정보를 수집·활용할 수 있음
개인정보	● 보건의료, 신용, 유통, 위치, 통신 정보 등이 있음 ● 정보주체의 선제적 동의가 쉽지 않고, 정보주체의 동의가 동반되지 않을 경우 개인정보보호 관련 법률 등에 따라 활용에 여러 가지 제약이 따름

정부는 개인정보의 활용을 촉진하기 위해 「개인정보보호법」, 「신용정보법」, 「정보통신망법」 등 데이터 3법⁷⁾을 개정하였으며, 개정 데이터 3법이 2020년 8월 시행됨에 따라 보험회사의 외부데이터 결합과 활용에 대한 근거가 마련되었다. 특히 개정 데이터 3법에서는 가명

7) 데이터 3법은 신용정보의 이용 및 보호에 관한 법률, 개인정보보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률을 의미하며, 2020년 1월 개정안이 국회 본회의를 통과하였고 2020년 8월 5일부터 시행되었음

정보가 새로이 정의되어 정보주체의 선제적인 정보동의 절차를 거치지 않아도 과학적 연구, 통계 작성, 공익적 기록보존 목적을 위해서 데이터의 활용 및 결합이 가능하게 되었다. 이어서 정부는 ‘가명정보 결합’ 제도의 활성화를 위해 규제 완화 및 제도 개선안을 순차적으로 발표하였다.⁸⁾ 복잡한 활용절차, 인프라 등을 개선하여 데이터 결합이 편리하고 효율적으로 이루어지도록 하는 취지이다. 법적 기반이 마련되고 제도 개선이 추진됨에 따라 보험업권에서도 이종 업종 외부데이터의 결합과 활용을 통해 다양한 부가가치를 창출할 수 있다는 기대감이 커지고 있다.

보험회사가 외부데이터를 활용하여 기대할 수 있는 성과를 이해하기 위해서는 국내의 보험회사의 외부데이터 활용 사례를 살펴볼 필요가 있다. 해외 주요 보험회사들은 자국의 법적 제도와 사회환경에 맞춰 내부데이터와 외부데이터를 결합해 제공 서비스의 차별성을 확보하고 있으며, 우리나라 보험회사도 외부데이터를 활용하여 상품 개발, 언더라이팅에 활용한 바 있다. 다음 장에서는 해외 보험회사의 외부데이터 활용 사례를 살펴본 뒤 국내 보험회사의 신용정보, 공공의료데이터 활용 연혁을 살펴보고자 한다. 국내 보험회사의 경우 공공의료데이터 활용과 관련하여 보다 집중해서 알아볼 것이다.

2. 해외 보험회사의 외부데이터 결합·활용 사례

가. 신용기반 보험점수: FICO⁹⁾

미국에서는 보험회사가 신용정보를 이용하여 사고율을 예측하고 언더라이팅에 반영하고 있다. 소비자 보고기관(Consumer Reporting Agencies)은 공정 신용 보고법(Fair Credit Reporting Act)을 근거로 소비자의 부채상환 및 연체 기록, 신용한도 대비 부채율, 신용거래 유지 기간 등에 대한 신용정보를 수집할 수 있으며,¹⁰⁾ 소비자 보고기관의 종류 중 하나인 신용평가기관은 해당 정보를 수집하고 고유의 산출 모델을 이용하여 소비자 개인의 신

8) 금융위원회 보도자료(2021), “빅데이터 인프라 확충과 데이터 산업 선도를 위하여 금융분야 가명정보 결합·활용을 가속화 하였습니다”; 관계부처합동 보도자료(2021), “가명정보 활용성과 및 확산 방안”; 금융위원회 보도자료(2022), “이종산업 간 데이터 결합·활용 활성화를 위해 데이터 결합 제도를 합리적으로 개선하겠습니다”

9) Experian(2020); <https://content.naic.org/cipr-topics/credit-based-insurance-scores>를 참고하여 작성함

10) 공정 신용 보고법에서는 신용평가기관이 개별 신용정보와 관련된 정보 수집은 허용하지만 인종, 민족, 성적 지향, 종교 및 정치색, 혼인 여부, 직위, 주소 등 심사 과정에서 차별을 야기할 수 있는 정보의 수집은 금지함

용점수를 산출한다. 산출된 개인 신용점수는 미국의 은행, 보험회사 등 금융기관이 대출 또는 보험 인수를 심사할 때 쓰인다.

개인 신용점수에는 소비자의 신용평가를 위한 신용점수와 보험회사의 언더라이팅을 위한 신용기반 보험점수(Credit-Based Insurance Score)가 있다. 신용점수는 FICO가 자체 개발한 모델을 통하여 산출된 FICO 스코어가 3대 신용평가기관¹¹⁾의 공식 신용점수로 채택되어 각 금융기관이 대출 심사 시에 참고한다. 신용기반 보험점수는 보험회사가 보험계약 인수심사 시에 사용하며, 이때 보험회사는 소비자의 신용평가를 위한 신용점수를 단독으로 활용할 수 없다. 신용기반 보험점수를 산출하는 대표적인 회사는 FICO와 LexisNexis가 있다.

신용기반 보험점수는 산출하는 데 필요한 항목이 연체 기록, 미상환 부채, 신용거래 이력, 결제내역, 채권 추심 및 파산 이력 등으로 신용점수와 유사하거나 동일하지만, 점수를 산출하는 목적이 다르다. 신용점수는 소비자가 채무 의무를 성실히 이행할 수 있는지를 위한 평가에 중점을 뒀다면, 신용기반 보험점수는 소비자가 보험금을 청구할 가능성과 청구 정도를 파악하여 미래 손해율을 예측하기 위해 설계됐다. FICO 신용점수는 300~850점 사이에서 산출되며, FICO 신용기반 보험 점수는 100~900점으로 각각 다른 범위 내에서 점수가 산출된다. 2021년 FICO의 추산에 따르면 미국 내 신용기반 보험점수 활용이 허가된 주에서 95%의 자동차보험회사 및 85%의 주택보험회사가 보험계약 언더라이팅 및 보험료 산출에 신용기반 보험점수를 활용하고 있다.

나. 신용정보, 공공기록 및 운전기록을 이용한 사망률 측정 고도화: LexisNexis의 Risk Classifier¹²⁾

LexisNexis는 여러 산업 분야에서 빅데이터를 처리하고 이에 따른 분석 결과를 제공하는 미국의 데이터분석 전문 기업이다. LexisNexis는 고객사에 데이터 보고서를 제공하며, 보고서에는 개인의 파산, 범죄 기록, 차압 기록 및 재판 판결 결과, 라이프스타일 및 행동, 전문 라이선스, 공공 기록, 부동산 소유권 등의 정보를 담고 있다. LexisNexis의 서비스를 이용하는 고객사의 산업 분야는 금융, 학술, 보건, 법, 미디어 등으로 다양하다.

11) Experian, Equifax, TransUnion

12) MunichRe(2018); LexisNexis(2016)를 참고하여 작성함

LexisNexis가 개발한 Risk Classifier는 주 고객층이 생명보험회사이며, 비의료데이터를 사용하여 사망률 예측을 고도화하기 위한 서비스이다. Risk Classifier에서는 보유 신용카드의 개수나 90일 이상 지불 연체 기록, 부채 기록, 부동산 및 기타 대출 기록 등 신용정보와 파산 기록, 중죄 및 유죄 판결 기록, 전문 라이선스 및 무기 면허증 소유 여부 등을 담은 공공기록, 그리고 사고기록, 속도위반이나 음주운전 이력 등의 운전기록 데이터를 종합한 후 스코어를 산출하여 사망률을 계층화한다.

〈표 II-3〉 LexisNexis Risk Classifier의 데이터 변수에 따른 사망률

더 높은 속성값	▲ = 사망률 악화 ▼ = 사망률 개선
연체 잔액 총액	▲
공공 연체 관련 정보 (bankruptcies, liens and judgments)	▲
중점죄 유죄판결 횟수	▲
형사 유죄판결 횟수	▲
전문 라이선스 보유	▼

자료: MunichRe(2018)

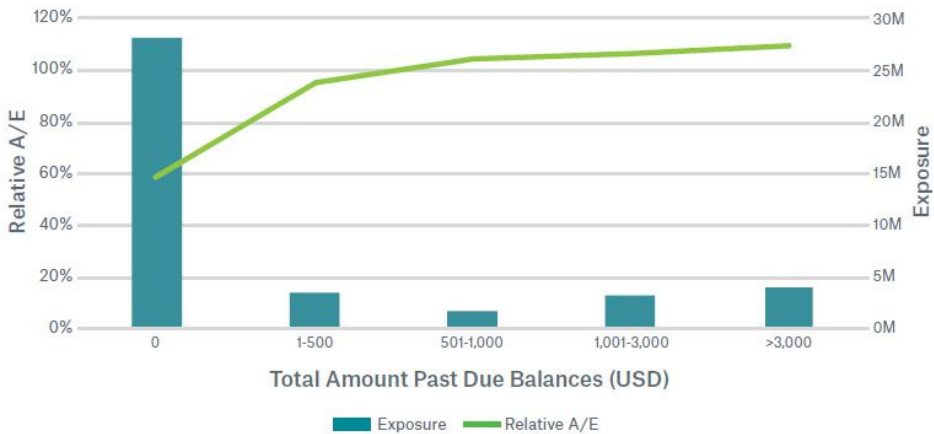
Risk Classifier의 정합성을 확인하기 위해 MunichRe(2018)는 LexisNexis를 통한 보험 가입자 8백만 명을 샘플링한 데이터를 이용하여 자산,¹³⁾ 공공기록, 신용정보 등과 사망률 사이의 관계를 분석하였다. 분석 결과 다양한 정보들이 사망률에 영향을 미침이 통계적으로 확인되었다. 사망률에 영향을 주는 정보는 개인이 보유한 신용카드의 개수 혹은 자산 수준이 있다. 보유한 신용카드 개수가 적거나 자산 수준이 낮은 사람은 경제적으로 어려운 상황에 처할 가능성이 높고 경제적으로 어려운 사람은 사망의 가능성도 높기 때문에 보유 카드 개수가 적거나 보유 자산 수준이 낮을수록 사망률이 높아지게 된다. 〈그림 II-4〉에서는 연체 잔액¹⁴⁾과 상대사망률¹⁵⁾ 사이의 관계를 나타내었으며, 연체 잔액이 0이 아닌 그룹은 연체 잔액이 없는 그룹에 비해 상대사망률이 높은 것을 확인할 수 있다.

13) LexisNexis는 개인의 자산 추정치를 제공하며, 자산은 소비자가 보유하고 있는 부동산, 선박, 항공기 등의 가치를 나타냄

14) 연체 잔액은 개인별 신용보고서에 포함된 모든 계좌의 누적 미결제 잔액을 나타냄

15) 상대사망률은 환자의 실제사망률을 동일한 성·연령대를 가진 사람의 기대사망률로 나눈 값임

〈그림 II-4〉 연체 잔액과 상대사망률(Relative A/E) 사이의 관계



자료: MunichRe(2018)

현재 미국에서는 600개 이상의 생명보험회사가 보험 인수 과정에서 Risk Classifier를 이용하고 있다. Risk Classifier가 제시하는 소비자의 보험 점수는 보험회사가 보험계약의 인수 여부를 결정하는 과정에서 단독으로 활용할 수는 없지만 중요한 요소로 활용이 가능하며, 보험회사는 Risk Classifier를 통해 고객의 사망률을 효과적으로 예측하고 인수심사 프로세스와 보험료 산출을 고도화할 수 있다. 한편 Risk Classifier에서 스코어를 산출할 때 인종, 종교, 국적, 결혼 여부, 성적 지향, 장애 유무, 직위, 고용 기록 등 공정 신용 보고법에 반하는 데이터를 포함하지 않는다.

다. 신용정보, 공공기록, 운전기록 및 의료데이터를 이용한 사망률 측정 고도화 : LexisNexis의 Risk Classifier with Medical Data¹⁶⁾

LexisNexis는 공공정보, 신용정보, 운전기록을 사용하는 Risk Classifier에 진료·처방정보, 임상정보, 의료 진단코드 등 의료데이터를 사용하는 ExamOne의 HealthPiQture를 결합함으로써 사망률 예측을 고도화할 수 있는 통합 모델 Risk Classifier with Medical Data를 개발하였다.¹⁷⁾ 서로 다른 종류의 데이터를 결합함으로써 활용 가치를 높인 사례

16) PartnerRe(2021); MunichRe(2021)를 참고하여 작성함

17) <https://risk.lexisnexis.com/products/risk-classifier-with-medical-data-next-generation-mortality-model>

이다. 보험회사는 LexisNexis의 통합 모델을 통해 외부데이터를 활용한다.

〈표 II-4〉 LexisNexis와 ExamOne의 각 모델별 사용 데이터

개발사	모델명	사용 데이터
LexisNexis	Risk Classifier	공공정보, 신용정보, 운전기록 등
ExamOne	HealthPiQture	진료·처방 정보, 임상정보, 의료청구 진단코드 등
LexisNexis	Risk Classifier with Medical Data	위 두 모델의 데이터를 통합하여 활용

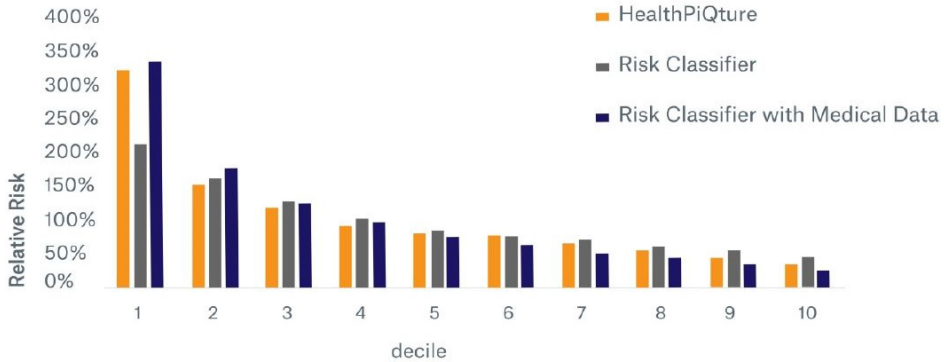
자료: PartnerRe(2021); MunichRe(2021)

PartnerRe(2021)와 MunichRe(2021)는 비식별 처리된 2백만 명의 데이터¹⁸⁾를 이용하여 결합 모델인 Risk Classifier with Medical Data의 사망 예측력을 분석하였으며, 분석 결과 결합 모델의 사망률 예측력이 단독 모델인 Risk Classifier, HealthPiQture보다 더 높게 나타났다. Risk Classifier와 HealthPiQture는 모두 사망률과 관련된 정보를 포함하고 있지만 내포하는 정보에는 차이가 있을 것이며, Risk Classifier with Medical Data는 두 모델에서 중복되지 않은 정보를 활용할 수 있기 때문에 사망률 예측력을 개선할 수 있는 것이다. 실제로 Risk Classifier와 HealthPiQture에서 모두 고위험으로 평가하는 고객군은 전체에서 18.1%이지만 개별 모델에서만 고위험으로 평가하는 고객은 81.9%로 나타나기도 하였다.

〈그림 II-5〉와 같이 모델별로 산출된 스코어에 따라 사망률이 높은 순으로 순위를 매겨 10개의 그룹으로 나눈 결과 Risk Classifier with Medical Data에서 10개의 그룹별 사망률 차이가 가장 크다는 것을 알 수 있다. 즉, 결합 모델에서는 저위험군으로 분류된 사람은 사망 위험이 더 낮고 고위험군으로 분류된 사람은 사망률이 더 높다는 것을 의미한다. 신용정보 및 의료데이터를 동시에 활용한 Risk Classifier with Medical Data는 신용정보를 활용한 Risk Classifier 및 의료데이터를 활용한 HealthPiQture와 비교하여 우수한 성능을 나타내며, 여러 유형의 데이터를 결합할 경우 시너지가 발생하는 것을 의미한다.

18) 분석 데이터는 모델 개발 목적으로 LexisNexis와 ExamOne가 보유한 데이터를 샘플링하여 생성함

〈그림 II-5〉 모델별 10분위수 그룹의 상대적 사망률



자료: MunichRe(2021)

보험회사는 Risk Classifier with Medical Data를 활용하여 사망률 예측 모델의 성능을 개선하고 언더라이팅 프로세스를 간소화하여 업무 효율성을 높일 수 있다. 즉, 데이터 분석 및 제공 전문 업체의 다양한 외부데이터를 활용하여 보험회사는 보험료 산출 및 보장 범위를 정교하게 채택하고 결과적으로 보험 가입 대상자를 확대할 수 있다.¹⁹⁾

라. 건강검진 기록 활용을 통한 언더라이팅 효율성 제고

: ExamOne의 LabPiQture²⁰⁾

ExamOne은 미국 최대 임상검사 전문기관 Quest Diagnostics의 자회사로 Quest Diagnostics가 보유하고 있는 임상검사 데이터와 LabCorp의 건강검진 데이터를 활용한 언더라이팅 툴 LabPiQture를 개발하였다. ExamOne의 LabPiQture는 최대 7년간의 임상검사 및 건강검진 데이터를 포함하며 생명보험회사에게 서비스를 제공한다.

생명보험회사는 크게 두 가지 측면에서 LabPiQture를 활용할 수 있다. 첫 번째는 신속한 언더라이팅이다. 미국에서는 보험소비자가 생명보험 가입을 희망하는 경우 혈액 및 소변 검사 결과를 보험회사에 제출해야 하는 경우가 존재한다. 그러나 이전에 받은 피보험자의 혈액 및 소변 검사 결과의 일부가 언더라이팅 심사가 진행되는 시점에 유실되는 경우, 피

19) LexisNexis의 Risk Classifier with Medical Data는 2021년 7월 출시되었으며, 이를 실제로 이용 중인 보험 회사의 수는 확인하지 못함

20) MunichRe(2020); GenRe(2020)를 참고하여 작성함

보험자는 다시 검사를 받아야 하는 번거로움이 있었다. 생명보험회사는 LabPiQture에서 제공하는 임상검사 및 건강검진 기록을 활용함으로써 보험소비자가 추가 검진을 받지 않도록 할 수 있다.²¹⁾ 특히 LabPiQture는 코로나19로 인한 비대면 언더라이팅 프로세스의 중요성이 부각되면서 주목을 받았다.

두 번째는 사망률 정교화를 위한 활용이다. LabPiQture가 제공하는 데이터에는 일반 혈액검사 결과(Complete Blood Count; CBC), 갑상선검사, 호르몬검사 등 일반적인 생명보험 가입을 위한 검진에서 요구되지 않는 검사 항목 결과를 포함하고 있다. 이를 통해 보험회사는 피보험자의 건강 상태에 대한 이해도를 높이거나 임상검사 데이터 결과를 기반으로 사망률을 조정하여 언더라이팅 프로세스를 정교화 할 수 있다.

이에 대한 유의미성을 검증하기 위해 GenRe(2020)는 비식별 처리된 10만 명의 보험 가입 신청자의 인수심사 점수와 LabPiQture를 이용한 점수를 비교하였다. 사망률을 기반으로 하는 인수심사 점수의 변화 여부에 따라 표준(Standard), 표준미달(Substandard), 인수거절(Decline/Uninsurable) 등 3단계로 나누어 분석한 결과, 84.4%의 보험 가입자는 LabPiQture가 추가적으로 적용하는 데이터로 사망률을 산출하여도 표준 수준에 머물렀으며, 9.9%는 표준미달, 5.7%는 인수거절로 분류되었다. 이렇게 인수심사 점수가 변화한 이유는 생명보험 가입을 위해 요구하는 검진 항목 외의 임상검사 및 건강검진 데이터에서 극단적인 결과 값이 도출되었기 때문이다.

〈표 II-5〉 LabPiQture의 데이터를 기반으로 한 보험 가입자의 사망률 분포 (단위: %)

분류	남성	여성	합계
표준(Standard)	80.5	89.1	84.4
표준미달(Substandard)	13.4	5.7	9.9
인수거절(Decline/Uninsurable)	6.1	5.2	5.7

자료: GenRe(2020)

위와 같이 생명보험회사는 보험 가입 절차에는 요구되지 않으나 LabPiQture가 제공하는 임상검사와 건강검진 데이터를 보충적으로 활용하여 언더라이팅의 효율성을 제고하고 보험 가입자의 사망률을 고도화하기 위한 새로운 인사이트를 얻을 수 있다.

21) MunichRe(2020)와 GenRe(2020)에 따르면 LabPiQture가 제공하는 데이터는 보험회사가 요구하는 검진 결과 데이터와 높은 확률로 부합하였음

마. 오피오이드(Opioid) 남용방지를 위한 약물 중독 위험군 식별: Blue Cross and Blue Shield

전 세계 선진국의 기대수명은 의료기술의 발달, 생활 수준의 향상 등으로 인해 꾸준히 증가하고 있다. 미국의 경우에도 1950년부터 기대수명이 증가 추세를 보여왔다. 하지만 2014년부터 2018년까지 기대수명이 0.03년씩 줄어드는 특이한 현상이 포착됐다.²²⁾ 이에 대한 원인은 미국 내에서 심각한 사회 문제로 부상했는데, 기대수명 감소가 마약성 진통제 오피오이드(Opioid) 처방 및 복용 남용에 기인한다는 학계 전문가들의 의견 때문이었다.²³⁾ 이와 더불어 미국 국립안전위원회(National Safety Council; NSC)의 분석 결과에 따르면 오피오이드 과다 복용으로 인한 사망률이 교통사고로 인한 사망률을 능가하는 것으로 나타났다.²⁴⁾

오피오이드 남용이 사회적 이슈로 부상한 가운데 Blue Cross and Blue Shield of Tennessee²⁵⁾는 빅데이터 분석회사 Fuzzy Logix와 협력하여 오피오이드 남용 잠재 환자를 식별하는 작업을 실시했다.²⁶⁾ Blue Cross and Blue Shield는 보험계약자의 과거 및 현재 의료기록, 보험금 청구데이터, 처방 기록 및 위치와 인구 통계정보를 분석하였고, 환자가 각각 다른 병원에서 처방받은 오피오이드 용량, 주기 등을 파악해 오피오이드 남용 환자와 오피오이드를 복용하지만 남용하지 않는 환자를 분류하여 그 특징을 파악하였다. 분석 결과에 의하면 오피오이드 남용자는 연간 5만 달러 이상 고액의 보험금을 청구할 가능성이 비남용자보다 59% 높은 것으로 나타났다. 즉, 환자가 오피오이드 중독에 빠지기 전에 선제적으로 방지할 경우 보험금 누수를 막는 효과를 기대할 수 있다. Blue Cross and Blue Shield는 의료인의 직접적인 개입이나 교육 캠페인을 통해 오피오이드 남용을 막아 보험금을 감소시키고, 이에 따라 모든 계약자의 보험료를 낮추며 오피오이드 중독에 따른 사회적 문제도 해결하여 사회적 후생을 향상하는 데 그 목표를 두고 있다.

또한 오피오이드 남용은 사망률뿐만 아니라 신생아에게도 심각한 영향을 끼치는 것으로

22) <http://www.macrotrends.net>

23) The New England Journal of Medicine(2015), "Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States"

24) NSC(2019)

25) Blue Cross and Blue Shield Association(BCBSA)은 미국 35개 주의 개별 건강보험회사로 이루어진 비영리 민영의료보험회사 연합체임. BCBSA의 라이선스를 받은 보험회사만이 "Blue Cross and Blue Shield"라는 상호를 사용할 수 있으며, 동 보고서에서 소개되는 BCBS 테네시, 앨라바마 지사는 해당 주를 대표하는 가장 큰 의료보험회사임

26) Forbes(2017), "How Big Data Helps To Tackle The No 1 Cause Of Accidental Death In the U.S."

드러났다. 임신기간 중 산모의 오피오이드 등 마약성 진통제 복용은 신생아에게 신생아 금단 증후군(Neonatal Abstinence Syndrome: NAS)²⁷⁾을 일으킬 수 있는 것으로 나타났다. Blue Cross and Blue Shield of Alabama는 의료 서비스 솔루션 제공업체 AxialHealthcare와 협력을 통해 앨라배마주의 15세에서 49세 사이 여성의 오피오이드 및 중독 위험을 일으킬 수 있는 약물 처방 기록을 분석하여 잠재적으로 오피오이드 남용이 예상되는 가임기 여성을 식별했다.²⁸⁾ AxialHealthcare는 주치의로 하여금 해당 환자의 오피오이드 처방량을 줄이도록 정보를 제공할 수 있으며, 주치의는 가임여성이나 현재 임신 여성에게 약물 이용법에 관해 설명할 수 있는 코칭을 받을 수 있다.

즉, Blue Cross and Blue Shield는 의료데이터를 활용하여 오피오이드 남용이 예상되는 환자를 식별할 수 있었으며, 이에 따라 사회적 문제를 해결하고 국민 건강을 증진하는 효과를 기대할 수 있다.

바. 실시간 행동 분석을 통한 언더라이팅 가속화 및 보험사기 방지: ForMotiv의 인텐트 점수²⁹⁾

ForMotiv는 연간 3억 개 이상의 온라인 보험 계약을 분석하는 디지털 행동 지능(Digital Behavior Intelligence) 전문 플랫폼 기업이다. ForMotiv는 기계학습을 통해 익명화된 행동 데이터를 파악한다. 즉, 보험 가입자가 온라인 보험신청서를 작성하면서 보여준 마우스의 움직임, 키 입력(Keystrokes) 소요 시간, 스크롤링, 복사 및 붙여넣기 등 약 150가지 행동을 결합하여 인텐트 점수(Intent Scores)를 산출한다. 예를 들어, 소비자가 온라인상에서 보험에 가입할 때 마우스 이동 패턴을 분석하거나 흡연, 의료 및 운전기록 등 보험 가입에 있어 민감한 정보를 기재할 당시 답변을 망설였다가 변경하는 등 일련의 행동을 분석하여 점수를 산출하는 것이다. 이를 통해 보험회사는 해당 결과에 따라 언더라이팅 속도를 향상하거나 보험사기를 방지할 수 있다.³⁰⁾

27) 신생아 금단 증후군(Neonatal Abstinence Syndrome: NAS)이란 신생아가 출산전 중독성 약물이나 처방 약물에 노출된 경우 유발되는 중증장애로 태아기 동안 태아가 마약성 진통제에 신체적으로 의존하게 되고 이후 엄마와 분리로 인해 출산 시 급작스럽게 이 같은 약물을 끊게 될 때 유발되는 장애임

28) Axialhealthcare(2018), "Can big data help mothers and babies in Alabama?"

29) <https://formotiv.com/blog/what-is-intent-data/>

30) ForMotiv(2022)

ForMotiv는 인텐트 점수를 산출하기 위해 제1자(First-Party), 제2자(Second-Party), 제3자(Third-Party) 데이터 등 세 가지 종류의 데이터를 수집한다. 제1자 데이터는 고객기업으로부터 직접 수집하는 데이터이며, 제2자 및 제3자 데이터는 파트너사와 외부 데이터 판매기업으로부터 수집한 데이터를 의미한다. 제1자 데이터 수집의 예를 들면, A 보험회사가 자신의 웹사이트에서 수집한 사용자 행동 데이터를 ForMotiv에 제공하여 잠재 고객의 디지털 행동 분석을 의뢰하는 방식이다. 따라서 ForMotiv는 인텐트 점수를 산출할 때, 보험회사가 직접 제공하며 연관성도 깊은 제1자 데이터를 우선으로 활용한다.

제2자 데이터는 데이터 공유에 동의한 파트너사의 제1자 데이터를 의미한다. 고객기업과 파트너사는 ForMotiv를 매개로 서로의 제1자 데이터를 공유함으로써 더 넓은 범위의 고객 행동을 분석할 수 있다. 파트너사 업종은 고객기업과 상이할 수 있으며, 파트너사로부터 수집한 제2자 데이터는 고객기업과의 직접적인 관련성은 다소 떨어질 수 있다. 마지막으로 제3자 데이터는 일정한 요건을 갖추어 수집한 개인정보를 정보주체 동의 없이 제3자에게 제공해 활용할 수 있는 데이터이다. 제3자 데이터는 주로 제3자 기업으로부터 구입하여 분석에 활용되며 다양한 목적을 위해 쓰일 수 있다. 한편 제3자 데이터는 정보주체의 사전적인 승인 절차를 거치지 않고 수집되기 때문에 판매 및 구입에 대해 개인정보 침해 논란이 제기되고 있기도 하다.

〈표 II-6〉 인텐트 데이터 분류

구분	제1자 데이터	제2자 데이터	제3자 데이터
데이터 주요 수집원	고객으로부터 직접 수집	유사한 성질의 고객을 보유한 파트너사	그 외 다수 수집원
활용 사례	현 고객 리타겟팅	파트너사 고객 타겟팅	임의의 고객 타겟팅
데이터 품질	높음	중간	경우에 따라 변동적임
타겟팅 범위	명확함	좁음	광범위함

자료: <https://formotiv.com/blog/what-is-intent-data/>

사. 유통정보 활용을 통한 보험료 할인: Tesco와 Sainsbury

Tesco와 Sainsbury³¹⁾는 소비자의 쇼핑 습관 정보를 반영하여 자동차보험료를 할인해주고 있다. Tesco의 경우 Clubcard loyalty를 통해 8백만 명 이상의 고객 정보를 수집하고

있으며, Sainsbury는 Nectar card를 통해 고객 정보를 수집하여 주택, 자동차, 펫 보험에 대해 상품별로 최대 12.5~25%의 보험료 할인을 제공한다. 특히 운전자들은 주기적으로 일정 금액 이상의 쇼핑을 하는 경우 저렴한 가격으로 자동차보험에 가입할 수 있다. 매주 슈퍼마켓을 정기적으로 방문하여 쇼핑하는 운전자는 계획적으로 생활할 가능성이 높으며, 운전을 할 때도 더욱 신중하고 안전하게 운전할 가능성이 높다는 논리이다. Tesco와 Sainsbury는 유통 정보를 자동차보험 할인 시에만 사용하고 있으며, 보험료를 할증하고 있지는 않다.³²⁾

하지만 자동차 사고 확률과 쇼핑 습관은 직접적인 인과관계가 없다는 지적이 있으며, 사생활 침해에 대한 논란이 지속되고 있다. 금융감독원(Financial Conduct Authority; FCA)은 보험회사가 쇼핑을 하지 않은 운전자가 보험을 갱신할 때 과도한 비용을 청구할 수 있다고 우려를 나타내기도 했다.³³⁾

아. 인공위성 정보를 활용한 위험평가, 보험사기 탐지 및 보험금 자동 지불 서비스

보험회사는 인공위성 이미지와 분석 솔루션 데이터를 활용하여 위험평가 및 모델링, 보험사기 탐지 및 피보험 재산 증명, 보험금 청구 등 보험 인수에서 보험금 지급까지 이어지는 프로세스를 개선하고 있다.

미국의 민간 인공위성 기업 Planet Labs는 피보험재산의 위치를 파악하고 변경사항을 감지할 수 있는 인공위성 이미지와 분석 솔루션 데이터를 보험회사에 제공한다. 인공위성 이미지를 활용한 위험평가의 일례는 홍수보험이다. 보험회사는 허리케인이 발생한 이후 미국 연방재난관리청(Federal Emergency Management Agency; FEMA)이 지정한 특별 홍수 위험지역과 관측 위성 기술을 통해 실제 대지의 수분을 탐지했을 때 표시되는 정규수 분지수(Normalized Difference Water Index: NDWI)를 대조하여 향후 해당 지역에서 허리케인이 발생할 때 홍수 위험을 보다 정확히 식별할 수 있었다.³⁴⁾

또한 인공위성 데이터는 높은 사회적 비용을 초래할 수 있는 자연재해 관련 보험사기를 예

31) Tesco와 Sainsbury는 영국 소매 유통 시장의 45% 이상을 차지하는 영국에서 가장 큰 소매업체이며, 슈퍼마켓과 할인마트 등의 점포를 활용하여 주로 자동차보험과 주택보험을 판매하고 있음

32) Joe Finnerty(2017)

33) Thomson Reuters(2017)

34) Planet(2017)

방하는 데 활용될 수 있다. 실제로 미국 연방수사국(Federal Bureau of Investigation; FBI)은 2005년 허리케인 카트리나가 미국 남동부를 강타한 이후 도시 재건을 위해 쓰인 정부 자금의 800억 달러 중 보험사기가 60억 달러를 차지할 것으로 추정한 바 있으며, 미국 연방 검찰이 재해보험사기로 기소한 사건의 수는 1,300건 이상에 이르기도 한다. 이렇게 보험사기의 규모가 큰 이유는 자연재해 발생 시 보험계약자가 주장하는 피보험 재산의 피해 규모를 직접 확인하는 데 많은 시간과 비용이 들기 때문이다. 그러나 보험회사는 피해 발생 전후 촬영된 인공위성 이미지와 기계학습을 통해 표기된 이미지상의 변경 사항을 활용하여 피보험 재산의 피해규모를 더욱 명확하게 파악하여 보험금 청구 사기를 방지하거나, 보험금 지급 이후 인공위성 데이터를 확인하여 피해 발생의 진실 여부를 조사할 수 있다.

이와 더불어 보험회사는 인공위성 이미지 데이터를 활용하여 보험금 산출의 정확도와 지급 속도를 향상할 수 있다. 이는 특히 드론비행이 금지돼 있는 지역에서 피해가 발생했을 시, 인공위성 데이터는 빠른 시간 내에 피해규모를 파악하는 데 유용하게 쓰일 수 있다. 실제로 인도 국영 보험회사(Agriculture Insurance Company of India)는 유럽우주국(European Space Agency; ESA)이 개발 및 발사한 인공위성 Sentinel-1이 제공하는 이미지 데이터를 활용하여 140년 이래 발생한 최악의 가뭄 피해를 입은 농작물 보험계약자를 식별하고 신속하게 보험금을 지불했다.³⁵⁾ 해당 데이터를 활용하기 이전에는 청구 건의 17%만이 농작물이 성장하는 시기에 보험금을 받을 수 있었다. 이와 비슷한 사례로, 독일 정부와 홍수 모델링 전문 스타트업 Cloud to Street는 Planet Labs로부터 받은 인공지능 이미지 데이터를 분석하여 홍수가 일어났을 때 보험회사가 보험금을 자동으로 지불할 수 있는 홍수 감지 체계(Flood detection scheme)를 개발하기도 했다.³⁶⁾

자. 기업 신용정보를 활용한 중소기업 보험인수 리스크 측정: LexisNexis의 Attract™ for Commercial Insurance³⁷⁾

LexisNexis는 2021년 SBFE(Small Business Financial Exchange)³⁸⁾와의 협력을 통해 보험회사가 중소기업의 리스크를 가시화하고 측정하는 플랫폼 Attract™ for Commercial

35) ESA(2017)

36) Pacific Standard(2018)

37) LexisNexis(2021)를 참고하여 작성함

38) SBFE는 미국의 중소기업 비영리 협회로 은행, 신용기관(American Express, Capital One, PNC, Sun Trust 등)으로부터 중소기업의 신용카드 결제내역, 임대료 결제내역, 대출 기록 등 신용데이터를 수집함

Insurance를 출시했다. Attract™ for Commercial Insurance는 SBFE가 수집한 중소기업의 신용정보, 대출이력 등을 활용할 수 있다. 보험회사는 Attract™ for Commercial Insurance를 이용하여 중소기업 보험을 인수하거나 갱신하는 시점의 기업의 손실 경향(loss propensity)을 평가하여 해당 기업의 리스크를 측정하고 보험료를 책정할 수 있다. 리스크 책정에는 SBFE가 제공하는 기업 신용데이터를 LexisNexis Attract의 점수 산출 모델에 적용하여 이루어진다. 이를 통해 보험회사는 SBFE가 제공하는 3천 9백만 개의 중소기업 신용 데이터와 LexisNexis Attract가 제공하는 천만 개 이상의 상업 보험 약관의 손실 데이터를 기반으로 산출된 미래 보험 계약자의 보험 점수를 확인할 수 있다.

〈표 II-7〉 해외 보험회사의 외부데이터 활용 사례 요약

사례명	활용 데이터	데이터 제공사	활용 분야	특징
신용기반 보험점수	부채상환, 연체기록 등 신용정보	FICO	언더라이팅 및 보험료 산출	미국 95%의 자동차보험 회사가 사용 중이며, 단독 사용은 불가능함
Risk Classifier	신용정보, 공공기록, 운전기록 등	LexisNexis	언더라이팅 및 보험료 산출	미국 600개 이상의 생명 보험회사가 사용 중임
Risk Classifier with Medical Data	신용정보, 운전기록, 의료데이터 등	LexisNexis	언더라이팅 및 보험료 산출	신용, 의료 등 다양한 분야의 데이터를 인수 심사에 활용함
LabPiQture	임상검사, 건강검진 데이터	ExamOne	언더라이팅	코로나19로 인해 비대면 언더라이팅 프로세스의 중요성이 부각되며 주목 받음
오피오이드 위험군 식별	의료데이터 및 오피오이드 처방 내역	Blue Cross and Blue Shield	약물 남용 방지 및 보험금 감소	비영리 민영 보험회사가 데이터 분석을 통해 사회적 후생을 향상시킴
인텐트 점수	온라인 행동 데이터	ForMotiv	언더라이팅 및 보험사기 방지	제1,2,3자 인텐트 데이터에 따라 품질과 범위가 다름
유통정보 활용을 통한 보험료 할인	유통정보	Tesco	보험료 할인	유통정보를 보험료 할증에는 이용하지 않음
보험업의 인공위성 정보 활용	인공위성 정보	Planet Labs	위험평가, 보험사기 방지 등	비개인정보인 인공위성 정보를 다양하게 활용함
Attract™ for Commercial Insurance	기업 신용정보	LexisNexis	언더라이팅 및 보험료 산출	기업 신용정보를 보험 인수에 활용하기도 함

3. 국내 보험회사의 외부데이터 결합·활용 현황

가. 신용정보

2006년 일부 생명보험회사는 언더라이팅 시에 개인 신용정보를 활용하였다. 보험회사는 보험 가입 심사 시에 연령, 과거 질환 여부 등을 감안해 보험 가입을 거절하거나 보험료에 차등을 두고 있는데, 여기에 가입자의 신용등급을 반영하는 방식이다. 실제로 한 생명보험회사의 경우 신용등급이 가장 낮은 10등급의 가입자는 사망보험금 기준 보험 가입 금액을 3천만 원으로 제한하기도 하였다.³⁹⁾ 신용등급이 낮은 소비자는 재무 상황이 좋지 않아 보험료를 지불할 능력이 떨어지고 보험사기에 빠질 개연성도 있기 때문에 보험계약을 중도 해지하거나 실효상태에 빠질 가능성 혹은 보험사고 발생의 가능성이 높다는 이유에서다.

하지만 생명보험회사의 개인신용정보 반영 정책은 오래 지속되지 못하였다. 시민단체의 반대와 더불어 금융감독원이 생명보험회사들에게 신용등급이 낮은 사람의 보험 가입을 제한하지 못하도록 지도하였기 때문이다.⁴⁰⁾ 신용등급은 개인의 신용 상태와 채무 이행 능력을 판단하기 위한 지표이며, 보험사고 발생과는 구체적 인과관계가 밝혀지지 않았고 보험료를 산출의 기초자료인 생명표에도 반영되어 있지 않기 때문이라는 논리였다. 또한 신용등급이 낮은 소비자의 보험 가입을 거절할 경우 이는 부당한 차별에 해당될 수 있다는 우려도 제기되었다. 다만 금융감독원은 생명보험회사들이 보험 인수 심사 시에 신용등급을 단독 판단 기준으로 활용하여 보험 가입 거절이나 보험 가입 금액 한도를 제한해서는 안 되지만 가입자의 보험료 부담 가능성 등 보험 가입의 적합성 여부를 판단하는 참고자료로만 선별적으로 이용할 것을 지시하였다. 즉, 신용등급이 낮아 보험료 납부 능력이 떨어지는 가입자의 경우 담보를 조정하거나 보험료가 저렴한 상품을 안내하는 방식을 권장하였다. 결론적으로 생명보험회사들이 신용등급을 이용하여 보험 가입 여부 및 가입 금액 한도를 제한하려는 정책은 중단되었다.

이후 학술적으로는 개인 신용정보가 보험사고 확률 혹은 손해율과 관계가 있다는 연구결과가 발표되기도 하였다. 정재욱·손승호(2009)에 따르면 신용등급과 자동차보험 사고 발

39) 이데일리(2006. 9. 4), “빚 많으면 보험 가입도 맘대로 못해?”

40) 김세환(2009)

생 확률 및 손해율과의 상관관계가 존재하였으며, 신용등급이 낮을수록 사고 발생 확률과 손해율이 높아지는 것으로 나타났다. 또한, 순수한 보험정보만을 활용한 모형의 보험사고 발생 예측력에 비해 신용정보를 추가한 모형의 예측력이 더 우수한 것으로 나타났다. 정재욱·여유희(2011)은 생명보험상품의 언더라이팅 시에 개인신용정보의 활용 효과를 분석하였는데 신용등급이 낮을수록 보험사고율이 높았으며, 신용정보를 활용한 보험사고 예측모형의 예측력이 상대적으로 높은 것으로 나타났다. 국내에서 개인신용정보와 보험사고 발생 확률 혹은 보험계약 인수 심사, 보험료 산출과 관련된 연구는 많지 않은 것으로 보인다. 미국, 캐나다 등 해외의 경우 보험계약 인수 심사 시에 개인 신용정보를 활용하는 것이 일반적이지만, 한국의 경우 규제의 방식 혹은 사회적 양극화 가중에 대한 우려 등 부정적 인식 또한 크기 때문에 이를 적극적으로 반영하기가 쉽지 않은 것으로 판단된다.

나. 공공의료데이터

1) 활용 연혁

우리나라 보험업권은 2014~2017년에 공공의료데이터를 활용하여 보험상품을 개발하고 보험료율을 산출하였다. 건강보험심사평가원은 2013년 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’을 근거로 보건 의료 빅데이터센터를 통해 의료데이터를 개방하였고, ‘개인정보 비식별조치 가이드라인’에 따라 비식별조치를 거친 뒤 데이터를 제공하였다. 보험회사는 개방된 공공의료데이터를 보험상품 개발에 활용하였고, 중기간질환, 중기폐질환, 중기심질환 등을 보장하는 상품을 출시할 수 있었다.⁴¹⁾ 하지만 2017년 10월 국정감사에서는 비식별조치가 완료된 데이터도 재식별을 통해 개인정보가 노출될 위험이 있고, 보험회사가 유병자 등을 보험 가입에서 차별할 수 있다는 지적이 제기되었다. 또한 시민단체도 공공의료데이터를 영리단체인 보험회사에 제공한다는 점에서 문제를 제기하였고, 이에 따라 건강보험심사평가원은 의료데이터 제공을 중단하였다.

건강보험심사평가원의 데이터 사용이 불가능해짐에 따라 보험회사는 호주, 캐나다, 일본 등 국가의 의료데이터 통계 등을 활용하여 신상품 개발에 착수하였다. 하지만 한국인과 인종, 식습관 등이 다른 국가의 데이터를 이용해 보험상품을 개발할 경우, 우리나라 소비자에게 맞는 사고 발생률을 정교하게 예측할 수 없어 보장 범위와 보험료 산출을 보수적

41) 김석영(2017)

으로 하여 상품을 개발할 수밖에 없었다.

이후 데이터 3법이 시행됨에 따라 가명정보를 활용하여 산업적 목적을 포함한 과학적 연구, 통계 작성 등을 할 수 있는 근거가 마련되었고, 건강보험심사평가원의 의료데이터에 대한 보험회사의 이용이 2021년 7월 승인되었다.⁴²⁾ 보험회사는 IRB(Institutional Review Board)⁴³⁾ 심사를 거쳐 공공데이터법⁴⁴⁾ 및 개인정보보호법 등에 따라 연구, 모델 개발 등을 위해 데이터를 활용할 계획이다. 향후 보험회사는 공공의료데이터를 분석하여 유병자나 고령자와 같은 기존 보험시장에서 사각지대에 놓인 취약계층을 위한 모델을 개발하고 전용 상품을 출시하여 이들에 대한 보장 범위가 확대되고 보험료가 정교화될 것으로 기대된다.

하지만 개인정보 보안 및 보험회사의 과도한 사익 추구에 대한 우려는 끊이지 않고 있다. 공공의료데이터의 개방으로 인해 가명정보의 재식별을 통한 특정 개인의 정보가 노출되어 보험회사가 소비자 가입 차별 등에 이용하거나, 공공데이터를 통해 정보주체인 국민 편익보다 보험회사 이익 극대화에 초점을 맞출 수 있다는 것이다. 이에 건강보험심사평가원은 ‘보건의료데이터 활용 가이드라인’에 따라 가명처리 후 데이터를 제공하여 개인추적 및 특정의 가능성을 없앴으며, 지정된 이용자의 방문 분석을 통해서만 분석을 허용하여 정보주체가 특정될 수 있는 위험을 최소화하고 있다.⁴⁵⁾ 하지만 보험회사가 공공의료데이터 활용을 통해 정보주체와 사회 공익을 극대화 할 수 있는지에 대한 논의는 합의된 결론에 이르지 못하고 있다.⁴⁶⁾

보험업권은 국민건강보험공단의 공공의료데이터 사용에 대한 심의도 진행하고 있다. 국민건강보험공단의 의료정보에는 건강보험심사평가원과는 다르게 표본코호트DB를 제공하여 환자의 진료내역, 자격정보, 검진정보 등을 추적할 수 있기 때문에 1년 단위의 진료내역 정보 등을 포함하는 건강보험심사평가원의 데이터와는 다른 특징을 갖고 있다. 즉, 환자가 약물을 복용하면서 병세가 달라지는 추이 등을 분석할 수 있다는 장점이 있다. 공공데이터법 및 개인정보보호법에 근거하여 공공기관은 특별한 사유 없이 공공데이터 신청 및 제공을

42) 금융위원회 보도자료(2021), “6개 보험사, 건강보험심사평가원의 공공의료데이터 이용을 위한 최종 승인 획득”

43) IRB란 생명윤리법에 따라 특정 연구가 윤리적·과학적으로 타당한지 여부 등을 심의하는 절차임

44) 공공데이터의 제공 및 이용 활성화에 관한 법률

45) 건강보험심사평가원(2022)

46) 보험업권은 공공의료데이터를 활용하여 희귀질환자, 고령자·유병자, 취약계층 등에 보장을 확대하거나 전용 상품을 개발할 수 있다고 주장하고 있지만, 시민단체와 의료계는 피보험자의 가입제한을 위해 공공의료데이터를 활용할 수 있다는 우려를 제기하고 있음

거부할 수 없기 때문에 국민건강보험공단에서는 이에 대해 심의하였지만, ‘국민 이익 침해 가능성’, ‘과학적 연구 기준 부합’, ‘자료제공 최소화 원칙 부합’ 등에 대한 이유로 2021년 9월 이를 반려하였다. 향후 추가적인 심의 절차가 진행될 것으로 예상된다.⁴⁷⁾

〈표 II-8〉 보험업권 공공데이터 활용 연혁

시기	내용
2013년	공공데이터 제공 및 이용 활성화에 관한 법률이 제정·시행됨
2014년	건강보험심사평가원은 ‘개인정보 비식별조치 가이드라인’에 따라 보험업권에 비식별화한 의료정보를 제공함
2017년	국정감사에서 공공의료정보의 재식별 가능성, 영리 추구, 가입자 차별 등의 문제가 제기되고 정보 이용에 대한 법적 근거가 명확하지 않다는 지적으로 인해 건강보험심사평가원은 보험업권에 의료정보 제공을 전면 중단함
2018년	데이터 경제 활성화를 위한 데이터 3법의 개정안이 발의됨
2020년	데이터 3법이 시행되어 산업적·상업적 가명정보 이용에 대한 법적 근거가 마련됨
2021년	보험업권의 건강보험심사평가원 의료정보 이용이 승인됨

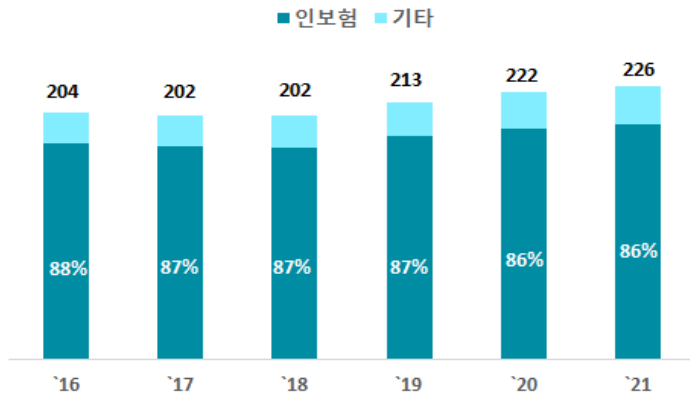
2) 수요 요인

한편 국내 보험회사는 외부데이터 결합·활용 시에 공공의료데이터에 대한 수요가 가장 높은 것으로 보인다. 이에 대한 이유로는 크게 4가지를 생각해볼 수 있다. 첫 번째는 인보험이 우리나라 보험시장의 대부분을 차지하고 있기 때문이다. 〈그림 II-6〉과 같이 우리나라 보험시장은 '21년 기준 수입보험료의 86%가 인보험(생명보험+손해보험의 연금 및 장기손해보험)에서 발생하였다. 상법에서 인보험은 ‘피보험자의 생명이나 신체에 보험사고가 발생할 경우 보험회사가 보험계약으로 정하는 바에 따라 보험금이나 그 밖의 급여를 지급하는 보험’으로 정의하는데, 인보험상품은 개발 및 언더라이팅, 보험료 산출에 있어서 피보험자의 사망률, 입원율 등 의료데이터가 무엇보다 중요하기 때문이다. 특히 희귀 질환, 유병자 등에 대한 진료·처방 내역, 입·통원 내역, 건강상태 등의 데이터는 보험회사의 경험통계에서 산출하기 어려우며, 이에 대한 정보를 보험회사가 공공의료데이터를 통해 얻을 수 있다면 새로운 유병자 전용 보험상품을 개발하거나 해외 통계를 활용한 보수

47) 국민건강보험공단 보도자료(2021), “민간보험사 연구계획 과학적 연구 기준 미충족, 연구결과 공개 및 검증 절차 결여”

적인 보험료를 체계를 사용하지 않고 요율체계를 세분화할 수 있다.⁴⁸⁾ 즉, 유병자·고령자 등 취약계층을 위한 보험상품과 희귀질환을 보장할 수 있는 상품을 개발하여 보험사각지대를 해소하고 다양한 소비자의 니즈를 충족시킬 수 있다.

〈그림 II-6〉 보험회사의 수입보험료 중 인보험 비중



주: 인보험은 생명보험+손해보험의 연금 및 장기손해임
 자료: 보험회사, 「업무보고서」

두 번째로 헬스케어 서비스의 고도화에도 공공의료데이터를 활용할 수 있다. 보험회사는 공공의료데이터를 기초로 건강증진형 보험상품의 개발을 활성화하여 국민 건강 증진을 도모하고 보험소비자에 대한 보험료 할인 혜택을 제공할 수 있다. 예를 들어, 골다공증 환자의 골절 발생률, 진료내역 등을 분석하여 골다공증 관리 서비스를 제공하거나 중대질병의 발생 이력을 추적하여 질병 발생을 사전적으로 예방할 수 있는 맞춤형 헬스케어 서비스를 제공할 수 있을 것이다.⁴⁹⁾

세 번째로 우리나라에서는 전 국민이 단일체계인 국민건강보험에 가입하여 양질의 진료 기록이 축적되고 있기 때문이다. 2019년 국민건강보험공단과 건강보험심사평가원이 보유하고 있는 의료데이터는 각각 3조 4천억 건과 3조 건에 이른다.⁵⁰⁾ 이와 더불어 공공의료데이터는 표준화 정도도 우수하여 민간 의료데이터의 활용에 앞선 의료데이터 활용의

48) 유주선(2020)

49) 손해보험협회 보도자료(2021), “「데이터 경제 시대의 보험산업 혁신방안」 세미나 개최”

50) 보건복지부 보도자료(2020), “안전한 보건의료 가명정보 결합·활용의 현장 조기 안착을 위해 나선다”

첫 단추로서 주목받는 것이다.

마지막으로는 개정 데이터 3법⁵¹⁾과 공공데이터법에 따른 제도적 근거도 마련되었기 때문이다. 공공데이터법에 따르면 공공기관은 공공데이터의 영리적 이용인 경우에도 이를 제공할 의무가 있으며,⁵²⁾ 국민건강보험공단, 건강보험심사평가원은 공공의료데이터를 개방하고 있어 제도적으로 의료데이터를 활용할 수 있는 근거는 마련되었다고 할 수 있다.

위와 같이 (공공)의료데이터는 보험업에서 가장 수요가 높은 종류의 데이터이며, 다방면에서 효용을 기대할 수 있다. 하지만 의료데이터는 개인정보 중에서도 민감정보⁵³⁾로 분류되어 활용을 위한 규제가 엄격하다. 즉, 의료데이터의 안정적인 활용을 위해서는 개인정보보호와 활용의 균형을 위한 제도적 측면의 고민이 무엇보다 중요하다. 다음 장에서는 의료데이터의 이차적 활용과 관련한 우리나라의 제도적 현황을 먼저 살펴보고 유럽, 미국, 일본 등 해외 주요 국가들의 제도적 현황과 비교하여 우리나라 제도의 현 주소와 시사점을 얻고자 한다.

51) 본고 p. 10에서 관련 이슈사항을 설명함

52) 공공데이터의 제공 및 이용 활성화에 관한 법률 제3조 제4항에 따르면 공공기관은 다른 법률에 특별한 규정이 있는 경우 또는 제28조 제1항 각호의 경우를 제외하고는 공공데이터의 영리적 이용인 경우에도 이를 금지 또는 제한하여서는 아니 됨. 하지만 제28조 제1항에 따르면 공공데이터의 이용이 제3자의 권리를 현저하게 침해하는 경우 공공데이터의 제공이 금지될 수 있음

53) 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보를 의미함

III

의료데이터의 이차적 활용

1. 개요

의료데이터는 사람으로부터 생성되는 데이터를 의미하며, 환자의 진료·처방 정보로부터 개인이 일상생활에서 생성하거나 유전적으로 확보되는 정보까지 다양한 범위를 포함한다. 의료데이터는 숫자, 문자로 이루어진 정형화된 데이터부터 영상, 그림, 음성 등 비정형화된 데이터까지 다양한 형태로 구성되어 분석이나 비식별화 조치가 어렵다는 특징이 있다. 다양한 형태의 의료데이터가 존재하기 때문에 의료기관별로 집적하는 정보는 표준화되어있지 않은 경우가 많아 연계하여 활용하기 어려운 복잡성을 나타내기도 한다.

〈표 III-1〉 의료데이터의 종류

구분	설명
개인유전정보	● 1인당 약 30억 개의 유전자 염기서열 정보 존재
개인건강정보	● 스마트폰 앱 또는 IoT 디바이스로 수집되는 Life-log(예: 수면패턴 등)
전자의무기록	● 의료기관에서 관리하며 환자의 모든 진료정보를 전산화하여 입력, 저장, 관리하는 형태(진단정보, 처방자료, 처방결과 등)
국민건강정보 (공공의료데이터)	● 공공기관에서 관리하며 자격 및 보험료, 진료내역, 건강검진 결과, 의료급여 등

자료: 삼정KPMG경제연구원(2020)

의료데이터의 이차적 활용(Secondary Use of Health Data)은 의료데이터의 일차적 수집 목적과는 다른 목적으로 사용하여 의료데이터의 가치를 높이는 것을 의미한다. 즉, 건강 보험청구, 진료기록, 임상 시험 등을 통해 수집된 의료데이터가 과학적 연구, 신약 개발, 안전성 모니터링, 개인별 맞춤 케어 플랜 등을 위해 사용되는 것이다. 예를 들어, 임상시험 데이터는 질병 및 보험금 청구 기록과 같은 데이터와 결합하여 데이터의 비식별화 조치를 거친 후 분석하여 환자 건강과 치료에 대한 이해를 높이고 혁신 상품 및 맞춤형 의료

서비스를 제공하거나 효율적인 의료 시스템을 구축하는 데 활용할 수 있다.

의료데이터를 이차적으로 활용하기 위해서는 다양한 이해관계자의 요구사항을 만족시킬 필요가 있다. 이해관계자에는 가장 먼저 환자와 같은 일반 정보주체가 있으며, 의료기관, 데이터 이용자, 규제 당국, 입법기관 등이 있을 수 있다. 일반 정보주체는 데이터 활용 목적, 데이터 활용 시 위험요인의 종류, 데이터 활용에 대한 동의 수단, 데이터 활용 목적의 공공성 등을 요구할 것이고, 의료기관은 일차적 데이터 가공 및 수집에 대한 능력이, 데이터 이용자는 데이터 수집·활용·분석·관리할 수 있는 능력, 개인별 맞춤형 서비스 제공 능력, 신상품 및 서비스 개발 능력 등이 필요할 것이다. 규제 당국은 데이터 활용 목적이 적합한지에 대한 감독, 개인정보보호에 대한 시스템 모니터링, 의료이용 불평등 및 가치의 공정한 분배에 대한 점검 등에 신경 쓰고 입법기관은 개인정보의 보호와 활용 사이의 균형을 이룰 수 있는 정책 목적을 달성할 수 있게 고민이 필요할 것이다.⁵⁴⁾

의료데이터는 개인에 관한 정보이기 때문에 데이터 유출 시에 발생하는 피해가 매우 크고, 사회적 불이익 혹은 불합리한 차별로 악용될 수도 있다. 따라서 의료데이터는 일반 데이터와는 다르게 활용이 법적으로 엄격하게 규제되고 있으며, 비식별화 조치와 활용 가능한 범위의 설정이 중요하다. 즉, 의료데이터는 비식별화 조치에 따라 가명정보와 익명정보로 나눌 수 있으며, 이에 대한 개념 및 활용가능 범위는 <표 Ⅲ-2>와 같다. 다음 장에서는 의료데이터의 이차적 활용을 위한 국내 제도적 현황을 OECD 주요국과 비교하여 살펴보고, 해외 주요 국가의 특징과 사례를 소개하고자 한다.

<표 Ⅲ-2> 개인정보, 가명정보, 익명정보의 개념 및 활용 가능 범위

구분	개념	활용 가능 범위
개인정보	특정 개인에 대한 정보 혹은 개인을 알아볼 수 있게 하는 정보	사전에 구체적인 동의를 받은 범위 내에서 활용 가능
가명정보	추가정보 없이는 특정 개인을 알아볼 수 없도록 처리한 정보	통계작성, 과학적 연구, 공익적 기록보존 목적에 있어서는 동의 없이 활용 가능
익명정보	더 이상 개인을 알아볼 수 없는 정보	개인정보가 아니기 때문에 제한 없이 자유롭게 활용 가능

자료: 보건복지부(2021)

54) ODI(2021)

2. 국내 현황

가. 제도적 기반

우리나라에서는 2020년 8월 개정 데이터 3법이 시행되며 의료데이터의 이차적 활용 관련 제도 또한 개선되었다. 개정 데이터 3법에서는 가명정보와 익명정보의 개념이 도입되어 정보주체의 동의 없이도 개인정보를 활용할 수 있는 근거가 마련되었다. 가명정보는 과학적 연구, 통계 작성, 공익적 기록보존을 목적으로 정보주체의 사전 동의 없이 처리가 가능하며, 익명정보는 개인정보가 아니기 때문에 개인정보보호법의 적용을 받지 않고 처리가 가능하다. 또한 신용정보법⁵⁵⁾과 보건복지부(2021a)의 ‘보건의료데이터 활용에 대한 가이드라인’에서는 과학적 연구에 산업적 목적의 연구⁵⁶⁾가 포함되면서 향후 의료데이터를 산업적 목적으로 연구·활용할 수 있는 근거가 마련되었다.

위와 같은 법적 근거의 마련에도 불구하고 개인 의료데이터는 민감정보로 분류되기 때문에 가명정보 처리에 대한 특례를 적용할 수 있는지에 대한 논란이 제기되기도 하였다.⁵⁷⁾ 개인정보보호법 제23조에 따르면 정보주체로부터 별도의 동의가 있거나 법령에서 민감정보의 처리를 요구·허용하는 등의 경우가 아닐 경우 원칙적으로 민감정보의 처리가 금지되고, 신용정보법 제33조 제2항에 따르면 개인의 질병·상해 등의 정보를 활용하려면 미리 정보주체의 동의를 받아야 하기 때문이다. 하지만 이러한 논란에 대한 법제처와 금융위원회의 유권해석⁵⁸⁾에 따르면 민감정보도 가명정보처리에 대한 특례를 적용할 수 있다. 가명처리된 질병정보 등 민감정보의 경우 특정 개인을 알아볼 수 없게 처리된 정보이기 때문에 민감성이 낮아 개인정보 침해가 발생할 우려가 낮으며, 정보주체를 특정하여 동의받기가 불가능하기 때문이다. 가명처리된 민감정보는 정보주체의 선제적 동의 없이 과학적 연구, 통계작성 등에 활용할 수 있다는 의미이다.

55) 신용정보법 제32조 제9의 2항

56) 산업적 목적의 연구는 새로운 기술·제품·서비스의 연구개발 및 개선 등이 있음

57) 황현아(2020)

58) 법제처 법령해석 사례 안건번호 20-0569; 금융위원회 법령해석 일련번호 200258

나. OECD 주요국과의 비교

우리나라에서 의료데이터의 이차적 활용을 위한 인프라와 제도는 세계적으로도 우수한 수준이다. OECD(2022)에 따르면 한국은 OECD 내에서도 국민 건강 증진 목적의 연구와 혁신을 가능하게 하는 의료데이터 활용 환경이 세계 최고 수준이며, 고성능 의료 시스템을 구축하기에 적합한 국가라고 평가받고 있다.

우리나라 (공공)의료데이터의 가용성과 거버넌스 현황에 대한 보다 구체적인 비교·분석은 Jillian Oderkirk(2021)를 참고할 수 있다. Jillian Oderkirk(2021)는 2019/20년 설문조사를 통해 OECD 23개국⁵⁹⁾의 국가별 주요 공공의료데이터⁶⁰⁾에 대한 가용성(Availability, maturity, and use) 및 거버넌스(Governance) 수준을 비교하여 살펴보았다. 의료데이터를 효율적으로 활용하기 위해서는 데이터의 질적 성숙도, 데이터 활용 인프라 등 가용성이 먼저 갖춰져야 하며, 가용성이 우수한 국가에서는 개인정보보호 및 보안 제도 등 거버넌스가 갖춰져야 할 것이다. 가용성과 거버넌스가 우수한 국가에서는 공공의료데이터의 사용을 장려할 수 있는 근거가 마련되었다고 할 수 있다.

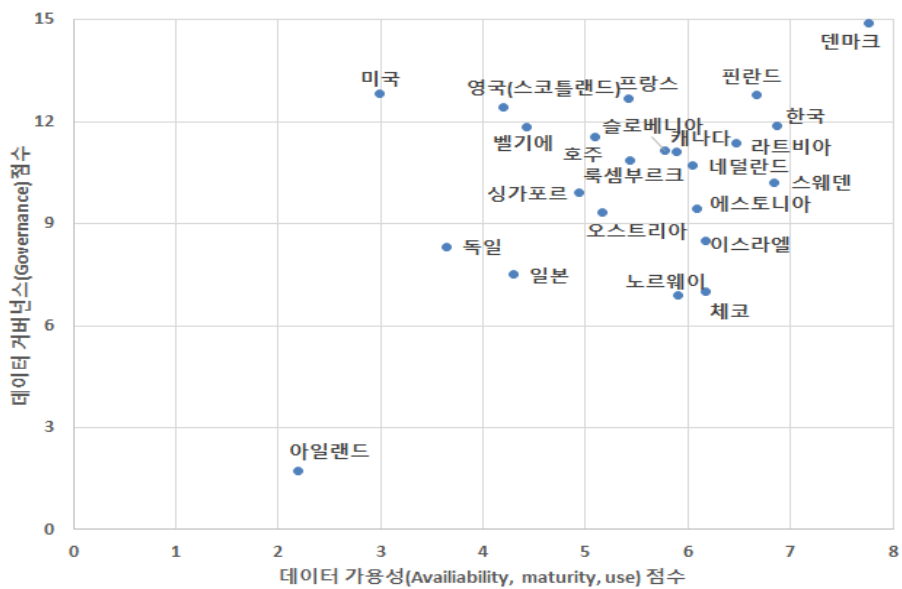
데이터 가용성에는 8개의 항목이 포함되며, 데이터 가용성(주요 공공의료데이터에 대한 가용 비율), 커버리지(주요 공공의료데이터의 총인구에 대한 커버리지 비율), 자동화(주요 공공의료데이터의 자동 추출 사용 가능 비율), 적시성(주요 공공의료데이터에서 기록 생성과 데이터 생성 사이의 시간이 1주일 이하인 비율), 단일 식별자(주요 공공의료데이터에서 환자 고유의 식별ID를 사용할 수 있는 비율), 코딩(주요 공공의료데이터에서 표준화된 코드가 사용 가능한 비율), 연계성(연구, 통계, 감독을 위해 연계 가능한 주요 공공의료데이터 비율), 정기적 보고(품질 및 시스템 성과에 대한 정기적 보고가 가능한 공공 의료데이터 비율)에 대한 국가별 점수를 산출한다. 거버넌스에는 15개의 항목이 포함되며, 데이터 보호 및 활용을 위한 법률, 개인정보보호 책임자, 비식별 조치 여부, 재식별 가능성 테스트, 공공·연구기관·영리기업 등과의 데이터 공유 여부, 원격 데이터셋 분석 시스템 등에 대하여 점수를 평가한다.

59) 호주, 오스트리아, 벨기에, 캐나다, 체코, 덴마크, 독일, 에스토니아, 핀란드, 프랑스, 아일랜드, 이스라엘, 일본, 한국, 라트비아, 룩셈부르크, 네덜란드, 노르웨이, 싱가포르, 슬로베니아, 스웨덴, 영국(스코틀랜드), 미국임

60) 병원 입원 데이터, 정신병원 입원 데이터, 응급의료 데이터, 일차진료 데이터, 처방 데이터, 암 발병 데이터, 당뇨병 발병 데이터, 심혈관 질환 발병 데이터, 사망률 데이터, 장기 요양 데이터 등 10개의 헬스케어 데이터와 환자 설문조사 데이터, 인구 설문조사 데이터 및 인구 등록 데이터 등 3개의 데이터로 구성됨

〈그림 III-1〉에서는 공공의료데이터에 대한 가용성과 거버넌스를 국가별로 비교하여 나타내었다. Jillian Oderkirk(2021)에 따르면 공공의료데이터의 가용성과 거버넌스를 종합하여 평가하였을 때 덴마크, 핀란드, 한국이 최상위 국가에 속하였으며, 프랑스, 영국(스코틀랜드), 미국의 경우 데이터 거버넌스에서는 높은 점수를 받았지만 가용성에서 상대적으로 낮은 평가를 받았다.

〈그림 III-1〉 국가별 공공의료데이터 가용성 및 거버넌스 점수 분포



자료: Jillian Oderkirk(2021)

특히 한국은 덴마크, 에스토니아, 라트비아와 함께 데이터 적시성에서 높은 점수를 받았으며, 체코, 핀란드, 이스라엘, 노르웨이 등과 함께 환자의 단일 식별자가 90% 이상의 데이터셋에서 존재하여 활용성이 우수한 것으로 나타났다. 또한 한국에서는 덴마크, 프랑스, 싱가포르, 미국과 유사하게 대부분의 의료데이터에 대해 현실적인 비용과 시간을 고려하였을 때 재식별 시도가 성공할 가능성이 낮다. 한국의 공공기관, 연구기관 혹은 헬스케어 서비스 제공자는 비식별화된 공공의료데이터에 접근이 가능하지만 영리기업, 해외 연구기관에 속한 이용자는 공공의료데이터를 제공받지 못하고 있다.⁶¹⁾ 이에 반해 덴마크,

61) 보건복지부 보도자료(2021), “보건의료데이터 활용 물꼬 트기 위한 국가 전략 수립 본격화”

핀란드, 핀란드, 미국 등의 국가에서는 영리기업에 속한 이용자가 비식별 개인 의료데이터에 대한 접근을 신청하고 승인을 받을 수 있다. <표 Ⅲ-3>에서 확인할 수 있듯이 데이터 거버넌스가 우수한 8개국 중 영리기관과 데이터 공유가 불가능한 국가는 한국과 벨기에가 유일하다.

〈표 Ⅲ-3〉 공공의료데이터 거버넌스 상위 국가의 주요 항목별 점수

국가	데이터 보호 및 활용에 대한 법률	개인 정보 보호 책임자	비식별 조치	재식별 가능성 테스트	공공기관 과의 데이터 공유	비영리 연구기관 과의 데이터 공유	영리기업 과의 데이터 공유	거버넌스 종합 점수
호주	67%	100%	100%	0%	78%	89%	89%	11.56
벨기에	100%	100%	100%	43%	57%	100%	0%	11.86
덴마크	100%	100%	100%	90%	100%	100%	100%	14.9
핀란드	100%	100%	100%	0%	100%	100%	100%	12.78
프랑스	78%	100%	100%	78%	100%	100%	78%	12.67
한국	100%	100%	100%	89%	89%	89%	0%	11.89
영국 (스코틀랜드)	89%	89%	89%	89%	89%	89%	89%	12.44
미국	100%	100%	100%	100%	83%	100%	100%	12.83

주: 1) 거버넌스 종합 점수 상위 8개국의 주요 7개 거버넌스 항목(총 15개)에 대한 점수이며, 점수는 주요 공공의료 데이터에 대한 각 거버넌스 항목의 여부 비율임

2) 거버넌스 종합 점수는 15개 모든 거버넌스 항목에 대한 점수의 합계임

자료: Jillian Oderkirk(2021)

정리하면 한국은 주요국과 비교할 때 공공의료데이터의 가용성이 우수하고, 데이터 거버넌스 체계도 잘 갖춰져 있다고 할 수 있다. 또한 비식별 조치를 충분히 취하고 있어서 주요 공공의료데이터에 대한 재식별 가능성 테스트를 높은 수준으로 통과하였다. 하지만 데이터 공유는 공공기관 및 비영리 연구기관의 이용자와만 가능하며 영리기관의 이용자는 공공의료데이터 활용이 불가능하다. 즉, 한국에서는 의료데이터의 이차적 활용에 대한 잠

재 가치를 충분히 갖고 있지만, 실제 활용은 공공기관과 비영리 연구기관에 대해서만 제한적으로 이루어지고 있다.

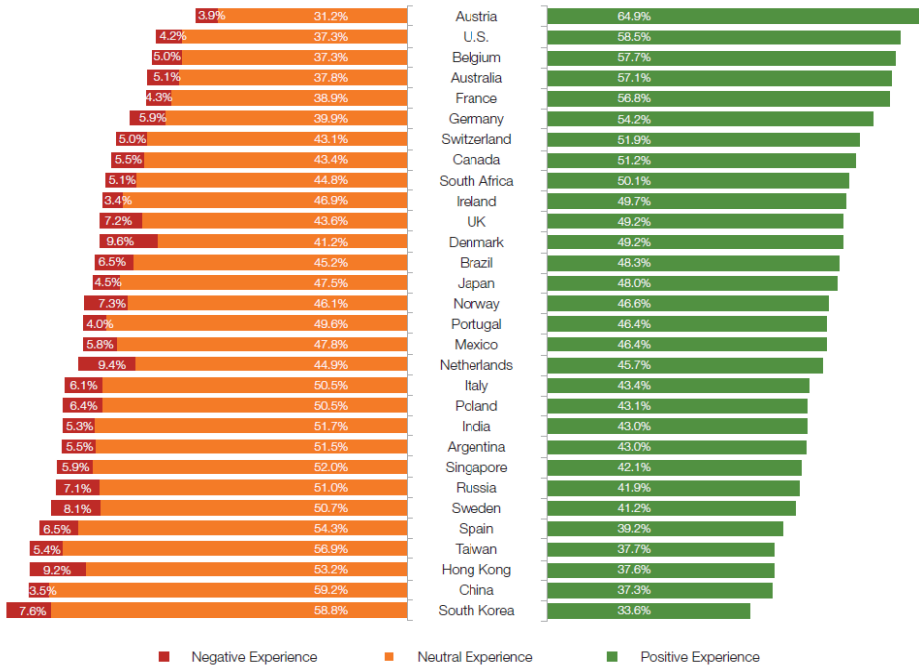
한국에서 법적으로 의료데이터의 활용이 가능하지만 제한적으로 활용되고 있는 이유는 법률 해석이 까다롭고 주요 이해관계자 간의 신뢰와 협력이 부족하기 때문인 것으로 보인다.⁶²⁾ 특히 개인정보가 유출되거나 재식별을 통해 개인이 특정되어 불이익이 발생하고 특정 집단의 영리목적으로 활용될 수 있는 사회적 우려가 크다. 특히 2013년에는 역대 최대 규모의 카드사 개인정보 유출 사건이 발생하면서 개인정보 유출에 대한 사회적 우려가 확대되기도 하였다. 또한 개인정보의 활용이 사회적 가치의 창출보다는 마케팅이나 피싱, 사기 등에 활용된다는 우려 또한 신뢰도 하락의 요인으로 지적되고 있다.⁶³⁾

이와 더불어 한국에서는 보험시장에 대한 신뢰도 부족한 편이다. Capgemini(2016)가 2015년에 국가별 보험소비자를 대상으로 실시한 설문조사에 따르면 보험회사에 대해 긍정적 경험(Positive experience)을 했다고 응답한 소비자 비율은 우리나라에서 33.6%로 30개국 중 가장 낮았으며, 부정적 및 중립적 경험을 했다고 응답한 비율은 66.4%로 가장 높았다. 물론 단편적인 설문조사로 사회 전체의 분위기를 대변하는 데에는 한계가 있고 과거에 실행한 설문조사 결과이기도 하지만, 우리나라에서는 보험시장에 대한 부족한 신뢰가 보험회사의 의료데이터 활용에 대한 제약 요인으로 작용할 수 있다.

62) OECD(2022); Jillian Oderkirk(2021)

63) 김용민(2019)

〈그림 III-2〉 국가별 보험소비자의 긍정적·중립적·부정적 경험 비교(2015년)



자료: Capgemini(2016)

한편 해외에서도 국가별 사회적 신뢰도 차이로 인해 의료데이터 활용의 활성화 여부가 달라지기도 한다. 대표적 사례로는 핀란드와 영국을 꼽을 수 있다. 뒤에서 언급하겠지만 핀란드, 영국은 유럽에서 의료데이터 활용을 위한 인프라, 제도적 기반이 가장 발달하였지만 사회적 신뢰에 따라 핵심 프로젝트의 성패가 달라진 바 있다. 핀란드의 경우 사회적 신뢰를 바탕으로 핀젠프로젝트가 별다른 반대 없이 활발하게 진척되고 헬스케어 산업에서 상당한 무역 흑자가 발생하는 등 성과를 보이고 있지만, 영국의 경우에는 사회적 신뢰가 이루어지지 않으며 굴지의 케어닷프로젝트가 영구히 중단되는 등 진통을 겪고 있다. 우리나라도 의료데이터 활용을 위한 환경이 세계적으로 우수하지만 사회적 신뢰가 구축되지 않은 상황으로 위 두 국가의 사례는 참고할만하다고 할 수 있다. 한편 세계에서 가장 빠른 고령화로 의료데이터 활용 정책을 국가적 최우선 과제로 추진 중인 일본의 사례도 참고할 필요가 있다.⁶⁴⁾ 일본은 의료데이터 활용을 위한 특별법을 제정하여 연구개발 및 신산업

64) Jillian Oderkirk(2021)이 2019/20년 분석한 자료에 따르면, 일본은 2013년에 비해 OECD에서 가장 빠르게 공공의료데이터 가용성이 개선된 국가로 평가받음

창출을 포함한 활용 방안을 적극적으로 모색 중이다. 다음 장에서는 유럽 국가별 의료데이터 활용 관련 정책적 현황을 간단히 비교하고 핀란드·영국의 사례에 이어 일본의 사례를 살펴보고자 한다.

3. 해외 사례

가. 유럽 국가별 비교

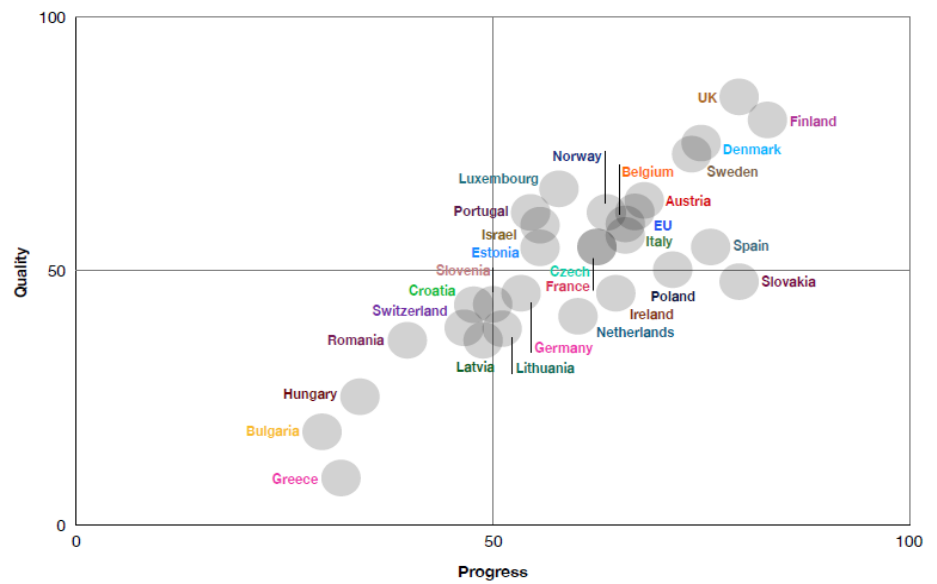
유럽 국가별 의료데이터 이차적 활용 현황은 주로 ODI(2021)의 조사를 참고하였으며, ODI(2021)는 OECD 국가들의 공공의료데이터 현황을 비교한 Jillian Oderkirk(2021)와는 달리 의료데이터 전반에 걸친 이차적 활용 현황을 살펴보았다. ODI(2021)는 유럽 국가별 의료데이터의 이차적 활용 생태계 구축을 위한 정책적 요소를 평가하기 위해 인프라(Infrastructure), 역량(Capability), 혁신(Innovation), 공정(Equity), 윤리(Ethics), 협력(Engagement) 등 6가지 항목을 고려하였다. 인프라는 의료데이터 공유 및 활용을 위한 기반시설의 구축 혹은 투자 여부, 역량은 의료데이터 분석 기술 및 경험, 혁신은 국가 내 이해관계자의 의료데이터 이차적 활용에 대한 가치 부여 정도, 공정은 의료데이터 활용으로 인한 편익의 공평한 분배 혹은 데이터 활용을 위한 개인정보 규정의 존재 여부, 윤리는 의료데이터 이차적 활용에 대한 정보주체의 자발적인 동의 여부 혹은 책임과 의무 그리고 보상 방안의 명확성, 협력은 공사의 협력을 통한 데이터 활용과 접근의 가능성 등에 대해서 평가한다. ODI(2021)는 국가별 정책당국과 유관 기관의 협조를 받아 6가지 항목에 대해 현재 정책 구현의 효율성(Quality) 및 향후 정책 구현의 진전(Progress) 등 두 가지 기준에서 점수를 부여하였다. 이를 통해 국가별 의료데이터 이차적 활용 정책에 대한 점수와 순위를 평가하였다.

효율성(Quality)과 진전(Progress) 측면에서 유럽 국가별 이차적 활용에 대한 점수는 <그림 III-3>으로 요약할 수 있다. 선두 그룹인 1사분면에는 영국, 핀란드, 스웨덴, 덴마크 등이 포함되었으며, 이들 국가는 의료데이터의 이차적 활용을 통한 혁신, 맞춤형 헬스케어 등의 가치에 대해서 잘 인지하고 인프라 구축을 통해 실제 의료데이터(Real-world data)를 의료 시스템에서 활용하고 있다. 영국, 핀란드 등의 국가에서는 의료데이터의 이차적 활용과 기회에 대하여 명문화하고 있으며, 특히 핀란드에서는 법 개정을 통해 이차적 활용

이 가능성을 분명히 하고 있다. 1사분면에 속한 대부분의 국가에서는 이차적 활용을 통해 개인 맞춤형 진료, 진단 및 예방의 개선 등이 가능한 근거가 마련되었다.

2사분면은 정책 추진 동력이 부족한 국가들로 의료데이터 이차적 활용의 가치는 인지하고 있지만 구체적인 시행 전략과 계획이 부족한 국가들이다. 2사분면에 속한 국가는 부재하였다. 3사분면은 준비가 미흡한 국가로 그리스, 불가리아, 헝가리 등이 있다. 4사분면은 비전(Vision)이 부족한 국가로 독일, 네덜란드, 폴란드 등이 포함되었으며, 의료데이터의 정책적 인프라와 상호운영성 등 질적 효율성이 부족하지만, 정책적 시행 과정과 향후 활용 계획이 존재하는 국가들이다.

〈그림 III-3〉 유럽 의료데이터의 이차적 활용: 국가별 점수 분포



자료: ODI(2021)

이어서 효율성(Quality)과 진전(Progress) 기준을 모두 고려하였을 때 점수가 가장 높은 핀란드, 영국의 제도적 특징을 간략히 알아보고, 각국의 사회문화적 차이로 인한 데이터 활용 프로젝트 성패의 사례도 살펴보고자 한다.

나. 핀란드

인구 고령화 현상이 가장 빠르게 진행되고 있는 국가 중 하나인 핀란드는 이에 대한 대응 방안으로 헬스케어 산업을 집중적으로 육성하고 있다. 핀란드에서는 2019년 보건의료데이터의 이차적 활용에 대한 법률(Act on the Secondary Use of Health and Social Data)이 제정되어 국민의 공공의료 및 사회보장 데이터를 자유롭게 활용할 수 있게 근거를 마련하였다. 법안은 GDPR을 보완하는 형태로 제정되었으며, 의료데이터를 이차적으로 활용하기 위해서는 정보주체의 사전적 동의를 받거나 비식별화(De-identified) 처리를 거친 후 선제적 동의 없이 아래의 허용된 경우에 한해 사용할 수 있다. 가명처리(Pseudonymised)된 데이터는 과학적 연구에 사용이 가능하며, 익명처리(Anonymised)되거나 통합 집계된(Aggregated-level) 데이터는 과학적 연구, 신기술 개발 및 혁신 활동, 교육, 규제적 목적 및 기타를 위해 사용할 수 있다.⁶⁵⁾ 핀란드 내 민간기업과 연구소, 정부 기관들은 국민의 의료데이터 및 사회보장 정보를 활용할 수 있게 되었다. 정보주체는 개인정보의 사용을 원하지 않는 경우 옵트아웃⁶⁶⁾(Opt-out)방식으로 사용을 거부할 수 있다.

핀란드 정부는 2007년부터 '칸타(Kanta)시스템'에 국민 의료기록을 중앙화하여 저장하고 있다. 칸타시스템에는 환자들의 진료기록, 처방전, 영상검사 결과 등 모든 의료데이터가 저장되며, 이렇게 저장된 의료데이터는 칸타서비스의 형태로 환자, 의료전문가, 시스템 개발자에게 제공된다. 환자는 본인의 의료 내역을 한 눈에 확인할 수 있으며, 의료서비스 제공자는 환자의 동의하에 정보를 열람할 수 있다. 연구기관, 제약사 등은 2020년부터 데이터 활용 허가기관인 핀데이터(Findata)를 통해 여러 기관에 산재해 있는 의료 및 사회보장 데이터를 연계하여 제공받을 수 있다.⁶⁷⁾

이와 동시에 핀란드 정부는 2013년에 민간기업의 의료데이터 수집과 활용을 허용하는 「바이오뱅크 법」을 제정해 민간기업의 바이오 연구를 장려하였다. 바이오뱅크 법이 통과되면서 혈액, 조직 등 인체에서 채취한 정보와 유전정보를 체계적으로 활용하여 연구할 수 있는 근거가 마련된 셈이다. 최종적으로 핀란드 정부는 2017년 '핀젠(FinnGen)' 프로젝트를 진행하여 칸타시스템에 저장된 환자의 의료데이터와 바이오뱅크에서 수집하는 유전

65) Act on the Secondary Use of Health and Social Data: Section 2, 3

66) 옵트아웃(Opt-out)방식이란 정보주체에게 개인정보 수집, 이용, 제공에 대해 사전동의를 아닌 고지하는 방식으로 정보주체가 반대의를 표시하는 경우(자기정보 결정권 행사)에 개인정보 수집, 이용, 제공이 제한됨. 이와 반대로 사전동의를 필요한 방식은 옵트인(Opt-in)임

67) Sitra(2019b)

자 정보의 결합을 통해 새로운 의료데이터를 생성하고 맞춤형 의료서비스를 제공하고자 하였다. 해당 프로젝트는 2023년까지 전 국민의 10%에 달하는 50만 명의 정보 수집을 목표로 특정 질병 사이의 상관관계를 연구하고 있다.⁶⁸⁾

핀란드는 헬스케어 산업에서 수출이 크게 늘어 2018년 수출액이 전년보다 3.4% 증가한 약 23억 유로를 기록하여 10억 유로 이상의 흑자를 발생시켰다.⁶⁹⁾ 핀란드에서 헬스케어 산업 규모는 GDP 대비 4% 수준에 불과하지만 지속해서 성장 중이며, GE헬스케어와 퍼킨 엘머, 바이에르 MSD 등 글로벌 헬스케어 기업들이 핀란드로 진출하고 있다. 대부분의 국가에서 개인정보 유출 및 프라이버시 침해의 이유로 의료데이터의 수집 및 활용에 어려움을 겪고 있으며, 산업적 목적으로 이차적 활용을 위한 별도의 법안을 제정하고 있는 나라가 거의 존재하지 않는다는 점에서 핀란드에서 진행 중인 보건의료 정책과 법률 제정은 매우 혁신적인 것으로 평가할 수 있다.

핀란드에서 의료데이터의 이차적 활용에 대한 법률 제정이 가능했던 첫 번째 배경에는 고령화 문제와 국민적 신뢰가 있다. 핀란드에서는 고령화⁷⁰⁾로 인해 중증 질병이나 만성 질병을 앓는 인구가 급증하여 더 이상 공공 의료서비스 체계를 유지하기 힘들다는 위기감이 커졌기 때문이다. 이러한 상황에서 의료데이터의 공개와 활용은 의료서비스의 질과 효율성을 높여주고 헬스케어 산업을 육성할 수 있는 계기가 될 수 있다.

두 번째 배경으로는 핀란드 국민들의 높은 사회적 신뢰를 꼽을 수 있다. 핀란드는 세계적으로도 신뢰가 높은 사회로 평가받고 있으며, 2019년 OECD가 실시한 설문조사에 따르면 핀란드 국민들은 64%가 정부를 신뢰한다고 응답하여 OECD 평균인 45%와 비교하여 상당히 높은 수준이다.⁷¹⁾ OECD(2021)에 따르면 핀란드에서는 역사적으로 신뢰 기반의 거버넌스 시스템이 자리잡고 있기 때문에 사회적 신뢰가 높다. 이렇게 핀란드에서는 높은 사회적 신뢰와 정부에 대한 국민의 믿음을 바탕으로 새로운 입법 모델이 성공할 수 있는 계기가 된 것이다. 또한 핀란드에서는 연구기관, 기업가, 환자 등 다양한 이해관계자가 만나서 지속적인 의견수렴 과정을 거쳤으며, 국민들은 의료데이터 활용 시에 수집되는 자신의 데이터 종류와 데이터가 사용되는 목적, 조건 및 권한에 관심이 높다는 응답이 많았기

68) https://www.finnngen.fi/en/for_researchers

69) HealthTech Finland(2019)

70) UN(2019)에 따르면 핀란드의 2019년 기준 65세 이상 인구 비중은 22.1%를 기록하여 일본(28.0%), 이탈리아(23%), 포르투갈(22.4%)에 이어 전 세계에서 네 번째로 높음

71) OECD(2021)

때문에 단일 데이터 허가기관을 통해 데이터 사용 승인 및 관리를 전담하게 하였다.⁷²⁾ 지난 십여 년 동안 핀란드 정부는 국민들에게 의료데이터를 안전하게 보호하며 사용한다는 믿음을 부여하고 제도적 절차를 마련해왔으며, 최신 정보보호 기술에 대한 활용과 정보주권 부여에 따른 개인정보 활용 과정에 대한 투명성도 지켜왔다. 핀란드의 국민들은 개인 의료데이터 제공에 대한 대가로서 의료서비스의 질적 개선과 사회보장 서비스를 받을 수 있다는 강한 신뢰를 갖고 있다.

다. 영국

영국은 브렉시트 이후 GDPR의 직접적인 구속력을 받을 필요가 없었지만, 유럽연합과 동일한 수준의 개인정보보호 관련 규제를 유지하겠다는 방침을 보이고 있다. 예를 들어, 영국의 데이터보호법에서는 개인정보처리와 관련하여 GDPR 규정에 따른다는 조항이 다수 포함되어 있으며, 원칙적으로 GDPR과 같이 환자의 정보를 제3자와 공유하기 위해서 정보주체의 동의가 필요하다. 영국에서는 케어닷프로젝트, 데이터보호법(Data Protection Act 2018), 국가 데이터 전략(National Data Strategy 2021) 등을 발표하며 국가 차원에서 의료데이터를 활용한 가치 창출의 목표를 제시하고 있다.⁷³⁾

영국은 NHS에서 전 영국인의 의료데이터를 보유하고 있기 때문에 의료데이터의 양적인 측면에서 세계적으로도 선도적인 모습을 보여주고 있다. 하지만 NHS는 규모가 워낙 방대하고 트러스트별로 독립적으로 운영되어 의료데이터 통합이 불가능하다는 한계에 봉착한 바 있다. 이러한 불리한 상황의 돌파를 위해 영국에서는 의료데이터 구축 사업으로 알려진 케어닷프로젝트를 추진한 바 있다. 케어닷프로젝트는 핀란드의 핀젠프로젝트와 마찬가지로 유전정보를 의료데이터와 결합하고 개방하여 중앙 데이터베이스를 구축하고 연구에 활용함으로써 영국 의료 서비스의 질적 제고를 이룬다는 목표로 영국 정부가 야심차게 추진한 프로젝트이다.

하지만 영국에서는 사회적 합의의 불발로 케어닷프로젝트의 최종 시행을 앞둔 2014년 2월 보류되었고, 2016년 7월에는 전면 중단에 이르렀다. 대규모 의료데이터 활용 프로젝트가 중단된 원인은 환자의 정보가 NHS를 벗어나 민간 기업에 공유될 경우 국민 건강 증

72) Sitra(2019a)

73) ODI(2021)

진의 목적이 아닌 영리 목적으로 활용될 가능성, 민감한 개인건강정보 유출에 대한 우려, 관리부실, 환자 정보의 재식별 가능성, 의사소통의 부재 등이 있었으며, 지역주치의, 프라이버시 운동가들을 포함한 빅데이터에 대해 우호적인 입장의 단체들을 포함한 모두가 반대하였다. 의료데이터의 활용을 위해서는 우수한 인프라의 구축뿐만이 아니라 사회적 신뢰가 무엇보다 중요하다는 것을 보여주는 사례이다.⁷⁴⁾

라. 일본

일본은 세계에서 가장 먼저 초고령국가에 진입하였으며 2019년 기준으로 세계에서 가장 높은 수준의 65세 이상 인구 비중(28%)을 기록하고 있다. 이와 더불어 평균 수명이 길어어짐에 따라 일본에서는 의료 서비스의 개선과 의료비 적정화가 중요한 과제로 떠올랐다. 일본 정부는 이에 대한 대응 방안의 일환으로 개인 의료데이터의 해석 및 활용의 촉진 이 중요하다고 판단하였고 효율적인 보건의료 정책 수립을 위해 법률 개정을 추진하였다. Jillian Oderkirk(2021)에 따르면 일본은 공공의료데이터 가용성 및 거버넌스 수준이 OECD 국가들에 비해 높은 편은 아니지만, 고령화에 따른 국가적 의료데이터 활성화 정책과 함께 의료데이터 활용 환경이 빠르게 개선되고 있다.

일본은 2015년, 2017년 전반적인 개인정보에 대하여 다루고 있는 개인정보보호법을 대대적으로 개정하여 익명가공정보⁷⁵⁾ 개념을 도입하고 익명가공처리된 개인정보를 활용할 수 있게 하였다. 또한 중개업자(Data broker)를 통해 익명가공정보가 유통될 수 있는 법적 환경을 마련하였다. 개인정보보호법에서는 개인의 의료데이터를 우리나라의 민감정보와 유사하게 ‘배려를 요하는 개인정보’⁷⁶⁾로 정의하여 이에 대한 정의와 활용을 위한 조건을 명시하였다. 배려를 요하는 개인정보의 정의는 의료데이터에 대한 관점을 기존의 ‘보호 중심’에서 ‘활용 중심’으로 전환되는 근거가 될 수 있었다.

그렇지만 배려를 요하는 개인정보를 수집하고 이차적 활용을 위해서는 정보주체가 사전에 동의(Opt-in)하는 것을 원칙으로 하여 보수적인 접근 절차가 존재하며,⁷⁷⁾ 이를 통한

74) 제갈춘기(2017)

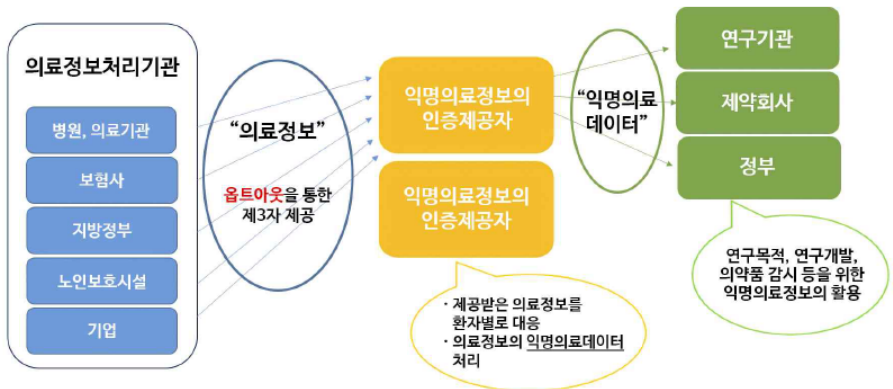
75) 익명가공정보는 특정 개인을 식별할 수 없도록 복원이 불가능하게 개인정보를 가공하여 만든 정보로 우리나라의 익명정보에 가까움

76) 본인의 인종, 신조, 사회적 신분, 병력, 범죄의 경력, 범죄로 인해 피해를 입은 사실 및 그 밖에 본인에 대한 부당한 차별, 편견 및 그 밖의 불이익이 생기지 않도록 그 취급에 특별한 배려를 요하는 개인정보를 의미함

77) 일본 개인정보보호법 제1장 제2조

의료분야의 연구·개발은 현실적으로 어렵다는 지적이 제기되었다. 일본 정부는 이러한 한계를 극복하기 위해 2017년 5월 「차세대의료기반법」을 제정하여 정보주체의 사전적 동의 없이도 익명가공정보를 통한 의료데이터의 활용·거래가 가능하도록 근거를 마련하였다. 익명가공정보를 활용하여 의료분야 연구개발, 신산업 창출 및 건강한 장수사회를 만들고자 하였으며, 개인정보를 보호하면서도 제약사, 연구기관, 공공기관 등의 데이터 활용성을 높이하고자 하였다. 일본 정부는 익명가공처리 기술과 높은 정보보안 기술을 보유한 사업자에 대해 인증사업자 제도를 도입하여 의료데이터를 첨단연구개발 및 신산업 창출 촉진 등의 목적으로 활용할 수 있도록 하였다. 차세대의료기반법 체계에서는 <그림 III-4>와 같이 병원, 의료기관, 보험회사, 지방정부 등 의료정보처리기관이 제공한 의료데이터를 익명의료정보의 인증제공자가 익명가공처리하여 연구기관, 제약회사 등에게 전달할 수 있다. 예를 들어, 일본의료데이터센터(Japan Medical Data Center; JMDC)는 일본 건강보험조합, 병원 등으로부터 비식별화된 건강검진데이터, 진료보수명세서 등을 받아서 분석·가공하여 보험회사 등에게 데이터분석 서비스를 제공하기도 한다.⁷⁸⁾

<그림 III-4> 일본 차세대 의료기반법 체계



자료: Akimoto Naoko(2021)

일본에서는 우리나라의 국민건강보험과 같은 단일 건강보험에 가입하지 않기 때문에 민간기업을 중심으로 의료데이터를 수집하고 활용하고 있다. 한편 정보주체는 제3자에 대한 본인의 정보제공을 희망하지 않을 경우 제공에 대한 거부권(opt-out)을 행사할 수 있

78) 정애령(2020); 손경한(2020)

다. 익명가공정보는 재식별이 가능한 가명정보와 비교할 때 재식별이 불가능하여 데이터의 활용 가치가 다소 떨어질 수 있다. 하지만 자기정보에 대한 주권 의식이 약하고 개인정보 유출에 대한 우려가 강한 일본의 사회적 상황을 고려하여 차세대의료기반법에서는 인증사업자 제도와 익명가공정보가 도입된 것으로 보인다.⁷⁹⁾

4. 시사점

가. 공공의료데이터 활용의 필요성과 필요 요건

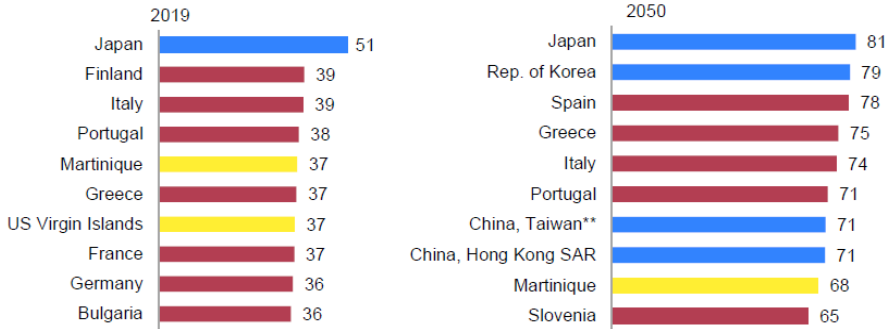
핀란드와 일본에서 의료데이터 활용을 위한 정책 추진과 사회적 공감대가 진전되고 있는 가장 큰 이유는 빠르게 진행되는 고령화이다. 일본은 세계에서 가장 빠르게 초고령사회에 진입하였으며, 핀란드 또한 초고령 국가 중 하나이다. 이들 국가는 고령화로 인한 환자와 진료비 증가에 대응하기 위해서 의료데이터 활용을 통한 의료시스템 개선을 최우선 과제로 채택하였다.

현재 한국은 고령사회⁸⁰⁾로 아직 초고령사회에 진입하지 않았지만, 출산율 저하 등으로 인해 세계에서 가장 빠르게 고령화가 진행되고 있는 나라이다. UN(2019)에 따르면 2019년 대비 2050년 65세 이상 인구 비중 증가폭은 한국이 23.0%p로 전 세계에서 가장 높으며, 2050년 노년부양비(65세 이상 인구/20~64세 인구)는 한국이 79로 일본(81)에 이어 높은 수준이다. 한국에서도 핀란드, 일본과 같이 고령화에 선제적으로 대처하기 위한 정책 과제가 최우선 순위로 검토될 필요가 있다.

79) 정승모(2019)

80) 한국의 2021년 기준 65세 이상 인구는 전체 인구의 16.5%임

〈그림 III-5〉 노년부양비가 가장 높은 10개 국가(2019년, 2050년)

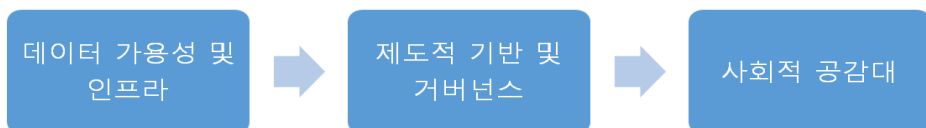


주: 노년부양비는 20~64세 인구 100명에 대한 고령(65세 이상)인구의 비(65+/20~64)를 의미함
 자료: UN(2019)

고령화가 진행될수록 국내 보험시장에서는 보험 가입에 거절되거나 높은 보험료를 지불해야 하는 고령자·유병자가 증가하여 보험 보장 공백이 확대될 가능성이 높다. 특히 희귀질환자들이 늘어날수록 보험회사의 경험통계로는 이러한 질병들에 대한 발병률, 사망률 등을 정교하게 반영하기 어려울 것이며, 이를 보장하는 보험상품의 보험료는 보수적인 방법으로 산출할 수밖에 없다. 공공의료데이터의 활용은 고령자, 희귀질환자 등 취약계층에 대한 보험 가입 승인, 가입 금액과 보장 범위의 확대, 보험료 할인 등을 위한 한 가지 방안이 될 수 있는 것이다.

의료데이터의 이차적 활용을 산업적으로도 실행하기 위해서는 단계적으로 필요 요건이 갖춰져야 한다. 첫 번째로 데이터의 표준화, 전자화, 적시성 등 가용성과 인프라가 구축되어야 한다. 두 번째로 개인정보보호와 활용을 위한 법적 근거가 정립되는 등 제도적 기반과 거버넌스가 마련되어야 한다. 마지막으로 인프라와 제도적 기반이 마련되어도 개인 의료데이터는 사람의 건강이나 신체에 대한 정보를 포함하기 때문에 실질적 활용을 위해서는 사회적 공감대가 뒷받침되어야 한다.

〈그림 III-6〉 의료데이터의 이차적 활용을 위한 필요 요건



우리나라에서는 공공의료데이터가 의료데이터의 산업적 활용의 첫 단계로서 논의되고 있으며, 위 필요요건 중 앞의 두 단계인 가용성과 거버넌스는 이미 갖추어졌다고 볼 수 있다. 다시 말해 공공의료데이터는 단일 건강보험체계로 인한 높은 커버리지 및 표준화 정도 등 우수한 가용성과 인프라를 구축하고 있으며 과학적, 산업적 목적으로의 활용을 위한 제도적 근거와 거버넌스 체계도 마련되어 있다. 즉, 우리나라는 공공의료데이터의 활용을 위한 인프라, 제도적 기반은 마련되었지만, 실질적 활용에 대한 사회적 공감대가 부족하여 제약이 생긴 상황이라고 볼 수 있다. 공공의료데이터의 이차적 활용을 위한 사회적 공감대를 형성하기 위해서는 사회적 신뢰도 제고를 위한 방안을 모색하고, 정보주체에 대한 이익 배분 방안을 고려하는 것도 방안이 될 수 있다.

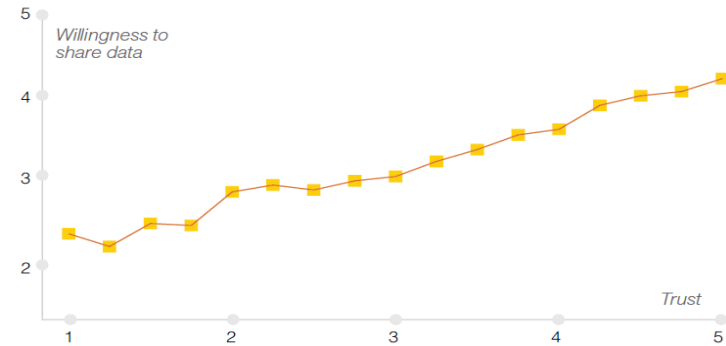
나. 사회적 신뢰도 제고를 위한 여건의 조성

우리나라에서 보험회사는 소비자로부터 높은 신뢰를 얻지 못했기 때문에 의료데이터의 활용에 대해서도 사회적 공감대가 쉽게 형성되지 않는 것으로 보인다. 이는 <그림 III-7>과 같이 보험회사에 대한 소비자의 신뢰도가 상승할수록 데이터 공유에 대한 선호도가 높아진다는 설문조사 결과에서도 확인할 수 있다.⁸¹⁾ 물론 사회적 신뢰는 단기간에 극복하기 어려우며, 핀란드에서의 높은 사회적 신뢰도 수십년에 걸쳐 형성된 것으로 보인다.⁸²⁾ 사회적 신뢰도를 제고하기 위한 단편적인 방안은 쉽게 찾을 수 없을 것이지만, 여기서는 의료데이터 활용과 관련한 IpsosMORI의 설문조사 결과를 소개하고 시사점을 생각해보고자 한다.

81) IBM Institute for Business Value(2017)가 16,000여 명의 소비자를 대상으로 실시한 설문조사 결과에 따르면 보험회사에 대해 신뢰도가 높다고 응답한 소비자일수록 높은 데이터 공유 성향을 나타냄

82) OECD(2021)

〈그림 III-7〉 보험회사에 대한 소비자 신뢰도와 데이터 공유 선호도



자료: IBM Institute for Business Value(2017)

영국의 글로벌 여론조사기관 IpsosMORI(2016)가 국민 2,000여 명을 대상으로 실시한 설문조사에 따르면 응답자들은 의료데이터 수집 및 활용으로 인한 가장 큰 우려 사항으로 개인정보 식별에 대한 가능성을 꼽았다. 소비자(응답자)들은 개인정보 보안의 우수성이 전제되어야지만 의료데이터의 활용에 동의하는 것이다. 그런데 재미있는 사실은 소비자들은 개인정보를 보호하기 위한 기술, 제도, 시스템 등의 구체적인 특징에 따라 응답이 달라지기보다는 보호 제도의 마련 여부를 인지하였을 때 데이터 활용에 대해 긍정적인 응답을 보였다.

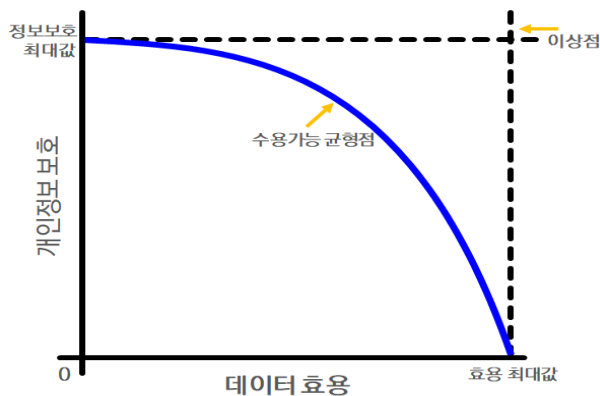
설문조사에서는 옵트아웃 옵션, 익명 처리, 접근 제한, 데이터 거버넌스 구축 등 네 가지 개인정보보호 방안이 각각 존재할 때 의료데이터의 상업적 활용을 허용할 것인지에 대한 찬반 여부를 조사하였다. 조사 결과 옵트아웃 옵션 61%, 익명 처리 56%, 접근 제한 64%, 데이터 거버넌스 구축 60%의 찬성 비율이 나타나 개인정보보호 방안에 따른 소비자의 선호도가 크게 다르지는 않았다. 하지만 소비자는 개인정보보호 조치가 존재하지 않을 경우 의료데이터 활용에 49%가 동의하여 보호 조치가 존재하는 경우보다 가장 낮은 찬성률을 보였다.

영국과 우리나라의 상황은 다를 수 있지만, 설문조사 결과로부터 얻을 수 있는 시사점은 소비자들에게 개인정보보호 방안에 대해 꾸준히 안내·홍보하여 소비자가 보호 방안을 이해하고 데이터 활용에 대한 신뢰를 높일 수 있는 여건의 조성이 필요하다는 점이다. 실제로 우리나라에서 개인정보보호 장치와 유출 및 재식별 시에 제재 조치는 높은 수준이다. 공공의료데이터는 무작위 표본 추출을 거친 후 가명처리하고, 승인된 연구자가 폐쇄 분석

공간에서 분석 결과 통계수치만을 반출하는 형태로 제공되기 때문에 합리적인 비용과 시간을 고려할 때 개인정보 유출이나 재식별의 가능성이 낮다. 이와 더불어 혹여나 기업의 정보처리자가 특정 개인을 알아보기 위한 목적으로 정보를 처리한 경우 전체 매출액의 3% 이하에 해당하는 금액의 과징금이 부과⁸³⁾될 수 있기 때문에 의도적인 재식별이 일어나기는 쉽지 않다. 과징금의 대략적인 수준을 보험산업에 대해 예를 들어 생각해보면, 2021년 보험회사 전체의 매출(수입보험료)이 224.9조 원⁸⁴⁾이므로 과징금은 최대 6.7조 원까지 부과될 수 있다. 2021년 보험회사 전체의 당기순이익은 8.3조 원이므로 과징금은 당기순이익의 80% 수준까지 부과될 수 있으므로 상당한 액수로 보인다. 즉, 이러한 데이터 공유 프로세스와 개인정보보호 장치 및 제재 조치에 대해 소비자가 인식 및 이해하는 정도가 높아질수록 데이터 활용에 대한 사회적 공감대 형성에 가까워질 수 있을 것이다.

물론 구체적인 개인정보보호 방안에 대해서도 지속적인 검토가 필요하다. 비식별화 조치의 기술이 발전하는 만큼 재식별화 기술이 발전하고 있는 것으로 보아 현재의 비식별화 처리된 데이터가 미래에는 재식별이 이뤄질 가능성이 전무하다고 할 수는 없기 때문이다. 반면 <그림 Ⅲ-8>과 같이 개인정보보호를 지나치게 강조하여 비식별화 조치의 수준을 높일 경우 데이터 활용에 따른 효용은 감소하게 되어 개인정보보호의 합리적인 수준을 고민할 필요가 있다.

<그림 Ⅲ-8> 개인정보보호와 데이터 활용에 따른 효용



자료: Gregory S. Nelson(2015)

83) 개인정보보호법 제28조의6

84) 금융감독원 보도자료(2022), “'21년 보험회사 경영실적(잠정치)”

다. 정보주체에 대한 이익 배분

의료데이터 활용에 대한 사회적 공감대를 형성하기 위해서 정보주체에 대한 이익 배분 방안을 생각해볼 수 있다. 즉, 의료데이터는 정보주체가 환자 개개인이며 이를 이용하여 얻은 수익 중 일부는 정보주체에게 배분되어야 할 필요가 있다.⁸⁵⁾ 미국에서는 인간대상 연구에 관한 법률인 Common Rule에서 개인의 생체시료(Biospecimens)가 상업적 이익을 위해 사용되었다면 정보주체가 상업적 이익의 공유 여부를 선택할 수 있도록 명시하고 있기도 하다.⁸⁶⁾

공공의료데이터는 전 국민으로부터 얻어지는 정보이다. 데이터 활용으로 인한 이익이 모든 국민들에게 배분되어야 하며 특정 단체, 기업이 독점하여서는 안 된다. 보험회사의 경우에도 마찬가지이다. 물론 보험회사는 공공의료데이터의 활용으로부터 유병자·고령자에 대한 보장 확대, 희귀질환 보장 상품 출시, 보험료 할인, 헬스케어 서비스 제공 등을 통해 국민들에게 일정 혜택을 돌려줄 수 있겠지만 보다 명시적인 이익 배분 방안도 필요하다. 민간기업이 공공의료데이터를 활용하여 얻을 수 있는 이익 중 일부를 시스템적으로 전 국민에게 배분한다면 의료데이터 활용과 관련한 논란이 보다 줄어들 수 있을 것으로 판단된다.

직관적인 방안 중 하나는 공공의료데이터 활용에 따른 수익의 일부를 국민건강보험금에 귀속시키는 방식이다. 공공의료데이터의 정보주체는 국민건강보험의 가입자이기 때문이다. 대만의 경우에도 건강보험청(NHI)의 일부 데이터를 보험회사에게 공개하고 있는데, 제공된 데이터를 통한 수익의 일부를 건강보험기금의 형태로 국민에게 반환하는 방안을 추진 중이다.⁸⁷⁾ 건강보험심사평가원 또한 의료데이터 개방 시에 수취하는 수수료를 정책 연구 등 공공영역에 대해서는 낮추고, 보험회사 등 민간기업에 대해서는 높이는 방식으로 적정수수료 부과체계를 검토 중인 것으로 보인다.⁸⁸⁾

물론 의료데이터의 이차적 활용으로 발생한 정확한 이익과 정보주체에게 배분해야 하는 비율을 결정하기는 쉽지 않다. 정보주체에게 배분해야 하는 이익 비율이 너무 높을 경우

85) 이경권(2019)

86) Protection of Human Subjects, 45CFR46(2018)

87) 손재희·이소양(2021)

88) 히트뉴스(2021. 10. 27), “공단·심평원 공공의료데이터 활용 기준 ‘제각각’”;
<http://www.hitnews.co.kr/news/articleView.html?idxno=36605>

에도 실질적으로 의료데이터의 활용이 이루어지지 않을 것이며, 보험회사의 이익 중 공공 의료데이터로 인한 부분을 정확히 산출하기도 어려울 것이다. 향후 지속적으로 논의를 거쳐 구체적인 방안을 찾을 필요가 있다.

보험업은 데이터의 활용이 가장 중요한 산업이며, 보험회사는 다양한 데이터를 활용하여 새로운 가치를 창출하고 있다. 반면 보험회사의 외부데이터는 정보주체의 선제적 동의가 쉽지 않아 적극적인 활용이 어려웠다. 이에 우리나라에서는 데이터 활용의 중요성에 대한 인식이 확산되며 데이터 3법이 개정되는 등 데이터 활용을 촉진하는 제도적 개선이 이루어졌다. 보험업권은 외부데이터 활용의 첫 단계로 데이터의 질적 수준이 우수하고 수요가 높은 공공의료데이터의 활용을 추진하고 있다. 하지만 실제 공공의료데이터에 대한 실질적 활용 성과는 미흡한 편이며, 이는 보험회사와 같은 영리기업의 공공의료데이터 활용에 대한 사회적 공감대가 형성되지 않은 것이 가장 큰 원인으로 보인다. 개인정보 유출 및 침해에 대한 우려가 높고, 특히 보험업권에 대한 신뢰가 형성되지 않아 보험업의 공공의료데이터 활용이 이루어지지 않고 있다.

우리나라에서는 공공의료데이터의 이차적 활용을 위한 데이터 인프라와 가용성 및 제도적 기반과 거버넌스가 해외 선진국과 비교하여 최상위 수준이다. 공공의료데이터를 활용한 가치창출에 있어서 세계적으로 선도국이 될 수 있는 잠재력을 갖췄다고 할 수 있다. 또한 제도적으로 추진되고 있는 가명정보의 공유와 활용은 익명정보와 비교하여 산업적으로 활용도가 더 높다. 의료데이터를 활용하여 상당 규모의 흑자를 달성하고 있는 핀란드와 같이 우리나라에서도 의료데이터에 대한 분석 역량을 선제적으로 갖춘다면 다양한 산업에서 가치를 창출할 수 있을 것이다.

물론 의료데이터의 활용에 있어서 보안절차, 활용범위 등은 매우 중요하기 때문에 개인정보보호를 위한 장치는 지속적인 검토가 필요하다. 현재 우리나라에서 공공의료데이터의 개방에 대한 보안 체계와 사고 발생 시에 취할 수 있는 제재는 높은 수준이다. 하지만 재식별화 기술도 빠르게 발전하고 있는 만큼 개인정보보호를 위한 절차와 방안에 대해서는 지속적인 개선이 필요하다. 본고에서는 개인정보(의료데이터) 보호를 위한 비식별조치, 보안절차 등에 대한 구체적인 검토를 포함하지 않고 있으며, 향후 연구에서는 의료데이터의 가명·익명처리 수준에 대해 국내와 해외의 현황을 비교하여 시사점을 도출하는 것도 의미 있을 것이다.

개정 데이터 3법을 비롯하여 가명정보 결합·활용 방안, 의료데이터 활용 가이드라인 등 개인정보 활용을 위한 제도적 기반이 사회적 공감대와 함께 마련된 만큼 이제는 실질적 활용이 이루어질 단계라고 생각된다. 우려와는 달리 보험회사가 공공의료데이터의 활용을 통해 신상품 개발, 보장 공백 축소, 보험료 할인 등 사회적 효용을 증진할 경우 이는 다시 사회적 신뢰도를 높이는 등 선순환이 이루어질 수도 있다. 물론 공공의료데이터의 활용에 있어서 사회구성원과의 소통은 지속해야 하며, 정보주체인 국민에 대한 이익 배분 방안도 추가적인 검토가 이루어져야 한다.

사회적 신뢰도는 확보하기 위해서 오랜 시간이 걸리지만 사소한 실수나 부주의에도 쉽게 사라진다. 보험회사가 의료데이터를 영리 추구의 목적으로만 사용하거나 개인정보가 유출되는 사건이 발생할 경우 데이터 공유에 대한 사회적 신뢰는 다시 회복할 수 없다. 향후 공공의료데이터의 본격적인 활용이 추진될 경우에도 다양한 시행착오가 예상되는 만큼 보험회사가 공공의료데이터를 통해 실제로 국민에게 혜택을 줄 수 있는 상품을 만들고 있는지, 과도하게 상업적으로 활용하지는 않는지에 대해 정보주체를 비롯하여 규제당국, 시민단체 등 사회구성원의 지속적인 모니터링이 필요할 것이다.

참고문헌

- 건강보험심사평가원(2022), 「HIRA 결합자료 신청자 이용 가이드」
- 금융감독원(2019), “미국의 보험산업 및 규제동향 Big Data의 활용과 감독상 이슈”, 조사
연구자료
- 김석영(2017), 「건강데이터 활용과 보험상품 개발」, 『KIRI 리포트』, 보험연구원
- 김세환(2009), 「미국 보험평점제도 관련 동향 및 시사점」, 『KIRI Weekly』, 보험연구원
- 김용민(2019), 「보건의료 빅데이터 활성화 방안에 관한 연구 - 핀란드의 사례를 중심으로 -」,
『의생명과학과법』, 제22권
- 보건복지부(2021), “보건의료 데이터 활용 가이드라인”
- 삼정KPMG경제연구원(2020), 「데이터 3법 통과: 의료데이터, 개방을 넘어 활용으로」
- 생명보험협회(2022), 『생명보험이란 무엇인가』
- 손경한(2020), 「데이터 거래·활용 활성화를 위한 법제도 개선 방안」, 국회입법조사처 정
책연구용역보고서
- 손재희·이소양(2021), 「공·사 보건의료 데이터 공유 사례와 시사점」, 『KIRI 리포트』, 보험
연구원
- 손지영(2021), 「보험산업의 웨어러블 데이터 활용」, 『KIRI 리포트』, 보험연구원
- 안소영(2020), 「운전행태 기반 자동차보험(BBI)」, 『KIRI 리포트』, 보험연구원
- 유주선(2020), 「데이터 3법 개정과 보험업에 관한 법적 연구」, 『상사판례연구』, 제33권 제
3호
- 이경권(2019), “21세기 붐이 김선달”, 메디칼타임즈 오피니언
- 정승모(2019), 「일본의 개인 의료데이터 현황 및 정책적 시사점」, 『방송통신정책』, 제31
권 3호
- 정애령(2020), 「보건의료데이터 활용의 헌법적 근거와 해외입법동향」, 『한국부패학회보』,
제25권 제1호
- 정재욱·손승호(2009), 「개인신용정보와 개인용자동차보험 상고발생확률 및 손해율 간의
상관관계 연구」, 『리스크관리연구』, 제20권 제2호

- 정재욱·여유희(2011), 「생명보험 언더라이팅 시 개인신용정보의 활용 효과 분석」, 『금융연구』, 제25권 제1호
- 제갈춘기(2017), 「영국의 케어닷데이터 프로그램」, 『의료와 사회』
- 최상아(2019), 「신용정보보호 및 빅데이터 활용을 위한 정책사례 및 국내 적용방안 연구」, 금융위원회
- 황현아(2020), 「가명정보 활용의 쟁점과 과제: 가명처리된 민감정보의 활용을 중심으로」, 『보험법리뷰』, 보험연구원
- Akimoto Naoko(2021), “일본 의료빅데이터의 활용과 과제 - 일본 차세대의료기반법을 중심으로 -”, 2021년 제4차 보건의료데이터 혁신 포럼
- Capgemini(2016), “World Insurance Report 2016”
- EIOPA(2019), “BIG DATA ANALYTICS IN MOTOR AND HEALTH INSURANCE: A THEMATIC REVIEW”
- ESA(2017), “Sentinel-1 speeds up crop insurance payouts”
- Experian(2020), “What Is the Difference Between Credit-Based Insurance Scores and Credit Scores?”
- ForMotiv(2022), “Behavioral Analytics: The Digital Transformation’s Impact on Banking and Insurance”
- GenRe(2020), "Evaluating ExamOne’s LabPiQture"
- Gil Press(2013), “The Big Data Landscape Revisited”, Forbes
- Gregory S. Nelson(2015), “Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification”
- HelthTech Finland(2019), “Healthtech industry in Finland”
- IBM Institute for Business Value(2017), “Data: gold or kryptonite? An insurer’s guide to the resource of the future”
- IpsosMORI(2016), “The One-Way Mirror: Public attitudes to commercial access to health data”

- Jillian Oderkirk(2021), "Survey results: National health data infrastructure and governance", OECD Health Working Papers No. 127
- Joe Finnerty(2017), "How Tesco and Sainsbury's are spying on your shopping bill to set your car insurance premium", *The Sun*
- LexisNexis(2016), "New Life Insurance Underwriting Tools Enable Insurers to Meet Consumer Expectations"
- _____ (2021), "LexisNexis Risk Solutions Can Provide up to 82% Scorable Rate Coverage on Small Businesses with SBFE and New Predictive Model"
- MunichRe(2018), "LexisNexis® Risk Classifier - Stratifying Mortality Risk Using Alternative Data Sources"
- _____ (2020), "ExamOne LabPiQtire™: An In-depth Exploration of Clinical Lab Histories"
- _____ (2021), "LexisNexis Risk Classifier with Medical Data- Stratifying mortality risk by combining behavioral attributes and medical data"
- NSC(2019), "For the First Time, We're More Likely to Die From Accidental Opioid Overdose Than Motor Vehicle Crash"
- OECD(2021), "Drivers of Trust in Public Institutions in Finland", OECD Publishing, Paris
- _____ (2022), "Towards an Integrated Health Information System in Korea", OECD Publishing, Paris
- Open Data Institute(ODI)(2021), "Secondary use of health data in Europe"
- Pacific Standard(2018), "HOW SATELLITES CAN HELP CATCH DISASTER INSURANCE FRAUDSTERS"
- PartnerRe(2021), "PartnerRe Review of LexisNexis Risk Classifier with Medical Data Combined Model Indicates Greater Predictive Value", *PartnerRe Insights*
- Planet(2017), "INSURTECH'S FUTURE WITH SATELLITE IMAGERY INCREASING CLARITY IN UNDERWRITING, RISK ANALYSIS, AND CLAIMS"

Sitra(2019a), “A Finnish Model for the Secure and Effective Use of Data: Innovating and promoting the secondary use of social and health data”, *Sitra Studies*, 153
_____(2019b), “Innovating the secondary use of health data - the Finnish model”
Thomson Reuters(2017), “Want cheaper British car insurance? Mind how you shop”,
INSIGHT
UN(2019), “World Population Ageing 2019: Highlights”

도서회원 가입안내

회원	연회비	제공자료	
법인 회원	₩300,000원	- 연구보고서 - 기타보고서 - 연속간행물 · 보험금융연구 · 보험동향 · 해외 보험동향 · KOREA INSURANCE INDUSTRY	영문 연차보고서 추가 제공
특별 회원	₩150,000원		
개인 회원	₩150,000원		

* 특별회원 가입대상 : 도서관 및 독서진흥법에 의하여 설립된 공공도서관 및 대학도서관



가입 문의

보험연구원 도서회원 담당

전화 : (02)3775-9113 | 팩스 : (02)3775-9102



회비 납입 방법

무통장입금

- 계좌번호 : 국민은행 (400401-01-125198) | 예금주: 보험연구원



자료 구입처

서울 : 보험연구원 자료실(02-3775-9113 | lsy@kiri.or.kr)

| 저자약력

박 희 우 KAIST 경영공학 박사 / 연구위원
E-mail : hwpark@kiri.or.kr

이 승 주 한국외국어대학교 통번역학 석사 / 연구원
E-mail : sj_lee@kiri.or.kr

연구보고서 2022- 10

보험업의 데이터 결합·활용 사례 및 시사점: 의료데이터를 중심으로

발 행 일 2022년 7월
발 행 인 안 철 경
발 행 처 보험연구원
주 소 서울특별시 영등포구 국제금융로 6길 38 화재보험협회빌딩
인 쇄 소 고려씨앤피

ISBN 979-11-89741-82-2
979-11-85691-50-3(세트)

(정가 10,000원)