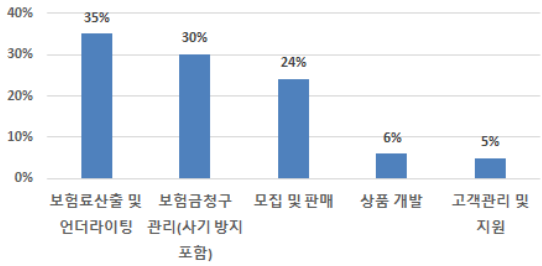
Ⅱ 보험업의 데이터 활용

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)

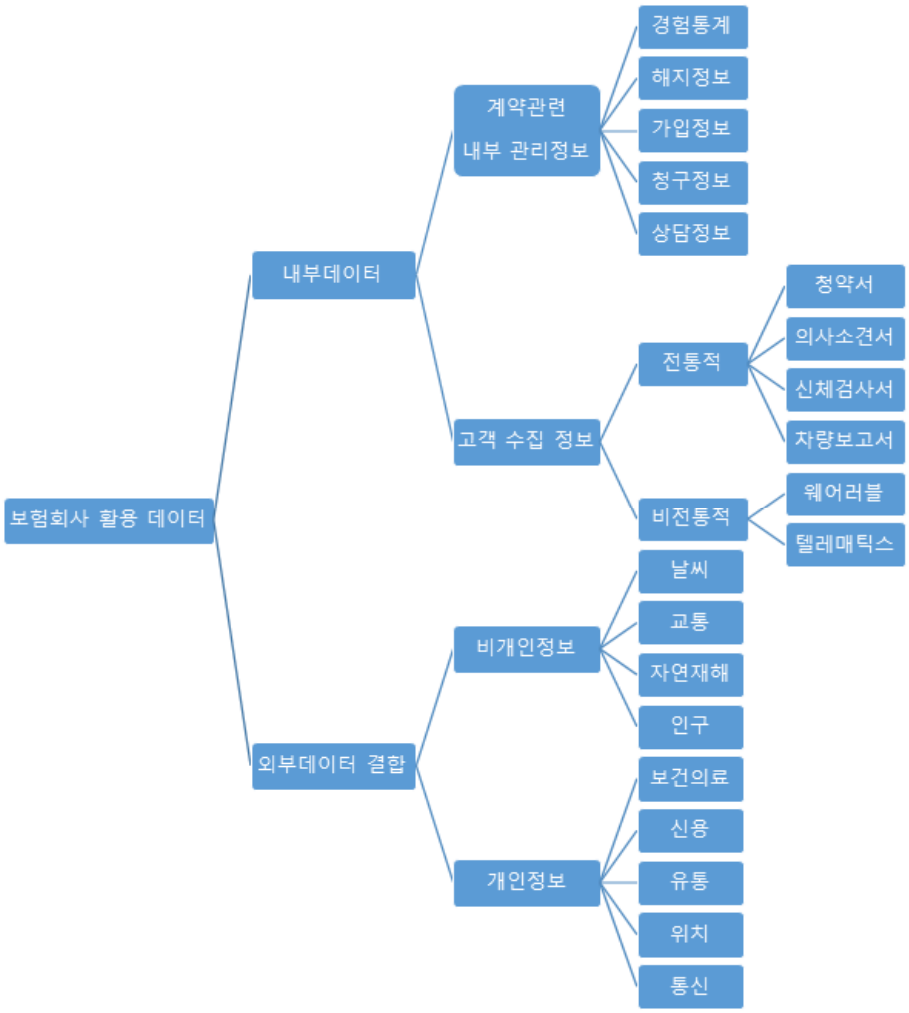
2) 생명보험협회(2022)

최근에는 보험산업에서도 디지털 전환이 이루어지며 웨어러블, 텔레매틱스 등 IoT 기술을 활용한 보험상품 및 서비스가 등장하고 있으며, 보험회사는 보건의료·신용·유통 등 이종 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, 텔레매틱스 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)

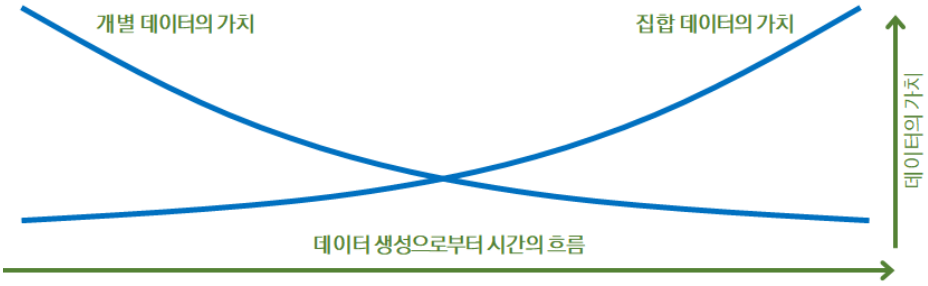
〈표 Ⅱ-1〉 보험회사 내부데이터 특징

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 경험통계 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 활 용 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 개인정보보호법, 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 이종 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 3법[7]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 국내외 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: FICO[9]

미국에서는 보험회사가 신용정보를 이용하여 사고율을 예측하고 언더라이팅에 반영하고 있다. 소비자 보고기관(Consumer Reporting Agencies)은 공정 신용 보고법(Fair Credit Reporting Act)를 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 하겠습니다"; 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-scores를 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 신용평가기관[11]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 Classifier[12]

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)

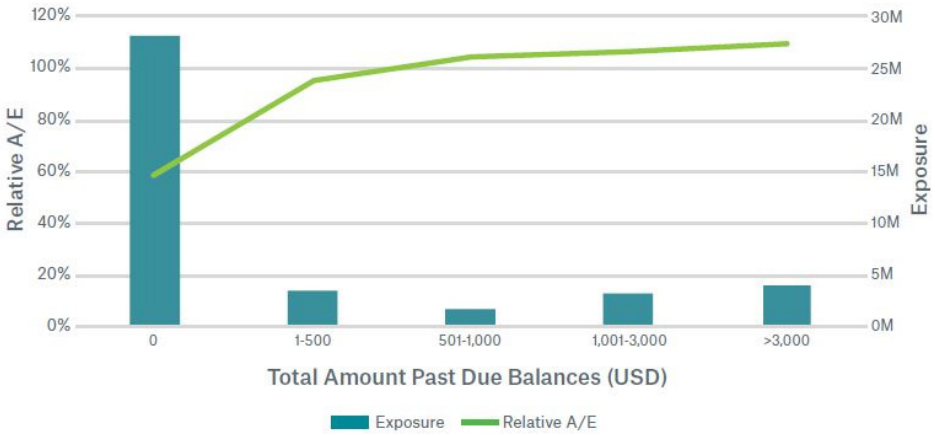
Risk Classifier의 정합성을 확인하기 위해 MunichRe(2018)는 LexisNexis를 통한 보험가입자 8백만 명을 샘플링한 데이터를 이용하여 자산,[13] 공공기록, 신용정보 등과 사망률 사이의 관계를 분석하였다. 분석 결과 다양한 정보들이 사망률에 영향을 미침이 통계적으로 확인되었다. 사망률에 영향을 주는 정보는 개인이 보유한 신용카드의 개수 혹은 자산 수준이 있다. 보유한 신용카드 개수가 적거나 자산 수준이 낮은 사람은 경제적으로 어려운 상황에 처할 가능성이 높고 경제적으로 어려운 사람은 사망의 가능성도 높기 때문에 보유 카드 개수가 적거나 보유 자산 수준이 낮을수록 사망률이 높아지게 된다. 〈그림 Ⅱ-4〉에서는 연체 잔액[14]과 상대사망률[15] 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 Data[16)]

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

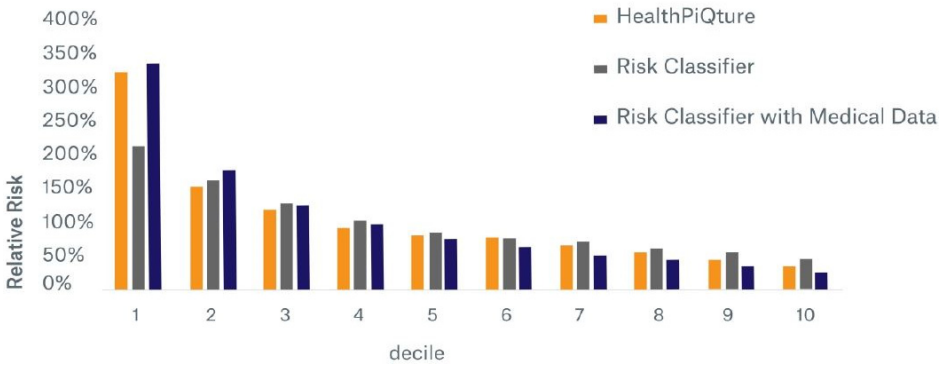
자료: PartnerRe(2021); MunichRe(2021)

PartnerRe(2021)와 MunichRe(2021)는 비식별 처리된 2백만 명의 데이터[18]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 LabPiQture[20)]

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 보험회사의 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 Tennessee[25]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 Corss 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 점수[29]

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 Sainsbury[31]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 정규수분지수(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 Insurance[37]

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 Atract가 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 보험료율 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·여윤희(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 보건의료 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 공공데이터법[44] 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 최소화의 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

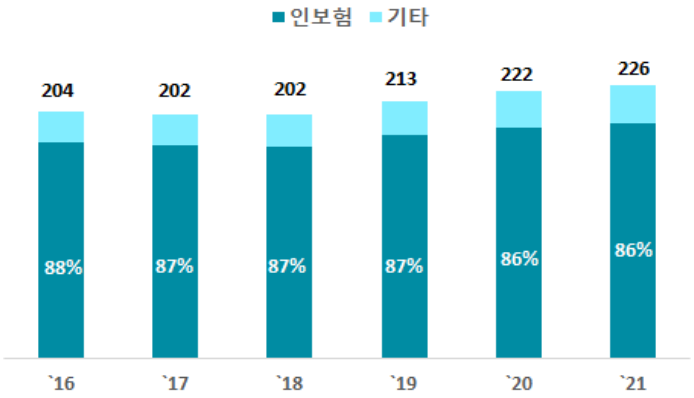
2) 수요 요인

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 희귀질환, 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 보험료율 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 진료기록이 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 3법[51]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 민감정보[53]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