

이달의 초점

사회보장분야 행정데이터 구축·연계와 활용 및 활성화 방안

사회보장 행정데이터 활용사례와 향후 과제

|유종성|

사회보장 행정데이터와 사회보장 정책의 근거 강화

|이현주|

사회보장정보시스템 행정데이터 활용 현황과 과제: 가명정보 결합 및 이용을 중심으로

|한은희|

노인 일자리 및 사회활동지원사업 실태조사의 행정데이터 연계 사례와 과제

|천재영|



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

사회보장 행정데이터와 사회보장 정책의 근거 강화

Integrated Administrative Data and Evidence-based Social Policy

이현주 한국보건사회연구원 빈곤불평등연구실 선임연구위원

이 글은 사회보장 행정데이터의 구축 배경과 특성을 소개하여 사회보장제도를 분석하기 위한 데이터의 한계를 극복하고 정책근거를 강화할 방안을 논의하고자 하였다. 우선 과거 사회보장제도 관련 데이터의 취약성과 사회보장 행정데이터 구축의 필요성, 사회보장기본법 개정과 사회보장 행정데이터 구축의 근거를 살펴보았다. 그리고 법에 근거하여 구축 중인 횡단면 형태 사회보장 행정 데이터의 강점과 개선 과제를 정리하였다. 향후 사회보장 분야 주요 정책 과제를 고려할 때 사회보장 행정데이터는 패널데이터, 집단시설 가구 거주자 행정데이터, 설문조사와 결합된 데이터로 데이터의 형태가 확충될 필요가 있음을 제안하였다. 한편 사회적 위험의 대리변수인 소득과 재산정보의 정확성이 높아져야 한다는 점을 확인하였다 이러한 과제들을 해결하기 위해서는 합리성을 담보한 체계적 협의 기구가 강화돼야 한다.

1 들어가며: 사회보장 행정데이터의 개요와 특성

가. 사회보장 행정데이터의 개요

2022년부터 사회보장기본법 제42조¹⁾에 의거, 사회보장 정책의 심의·조정 및 연구를 위하여 사회보장 행정데이터가 수집되고 있으며, 횡단면 형태의 통합행정데이터로 구축 중에 있다. 2022년에는 2020년 정보를, 2023년에는 2021년 정보를 추

1) 사회보장기본법 제42조는 사회보장 행정데이터의 제공 요청 목적과 수집 근거를 명시하고 있다. 해당 조항에 의하면 사회보장위원회는 '사회보장 정책의 심의·조정 및 연구를 위하여 관계 기관의 장에게 사회보장 행정데이터가 모집단의 대표성을 확보할 수 있는 범위에서 ... 사회보장 행정데이터의 제공을 요청할 수 있다. 이 경우 사회보장 행정데이터의 제공을 요청받은 관계 기관의 장은 특별한 사유 없으면 이에 따라야 한다.'

출, 연계하여 정책분석용 사회보장 행정데이터가 구축되고 있다. 향후 매년 횡단면 형태의 데이터가 구축될 예정이다.

사회보장 행정데이터에 대한 정의는 사회보장기본법 제3조에 명시되어 있다. '사회보장 행정데이터는 국가, 지방자치단체, 공공기관 및 법인이 법령에 따라 생성 또는 취득하여 관리하고 있는 자료 또는 정보로서 사회보장 정책 수행에 필요한 자료 또는 정보를 말한다.' 앞서 법상 정의에서와 같이 행정데이터는 통상 정책 집행, 행정을 목적으로 개인이나 가구, 사업체 등 조직에 대한 정보를 수집하여 관리하는 데이터이다. 행정 자료의 고유 목적은 효과적이고 효율적인 정책 집행일 것이다. 그러나 최근에는 여러 국가에서 정책 평가와 국민의 생활 여건을 정확하게 파악하기 위하여 행정데이터를 활용하는 경향이 강해지고 있다. 기술의 발전으로 전산으로 처리되는 행정데이터가 늘어나고 이러한 전산 정보가 체계적으로 관리되면서 정책기획의 근거를 강화하는 데 행정데이터를 활용하고자 하는 노력이 확산되고 있다.

사회보장제도를 정확하게 이해하기 위해서는 일정 수준 이상의 표본 크기와 정보 포괄성을 필요로 한다. 현재 구축하는 사회보장 행정데이터²⁾는 전가구의 약 20%를 표본으로 한다. 중요한 의미를 지니는 제도이지만 대상자 규모가 그리 크지 않은

사회보장제도가 적지 않다. 사회보장 행정데이터의 표본 규모는 이러한 점을 고려할 때 어느 정도 이상의 크기를 확보하여야 한다. 지역별 분석이나 소득·재산 분위별 분석을 할 때 안정성을 갖춘 결과를 기대한다면 필요한 표본의 크기를 어느 정도 추정해 볼 수 있을 것이다. 만약 특정 사회보장제도가 전국적으로 약 1만 가구를 정책 대상으로 할 때 20%의 표본 자료는 약 2000가구만을 해당 자료에 담아낼 수 있다. 분석하게 될 셀의 수를 고려하면 충분한 수의 표본을 확보할 수 있는 표본추출 확률을 낮게 정하기 쉽지 않다.

2022년 구축한 사회보장 행정데이터는 통계청의 인구·가구 DB를 활용하여 표본을 추출하였다. 개인 단위 정보를 추출, 연계하였지만 가구 아이디를 가구원이 공유하도록 하여 가구 단위로 정보를 구성할 수 있다. 사회보장제도 중 사회보험은 주로 개인을, 공공부조는 주로 가구를 지원의 단위로 하는 점을 고려할 때 개인과 가구 단위로 정보를 구성할 수 있어야 데이터의 유용성이 크다.

사회보장제도에 대한 심의·조정 등을 위한 근거를 마련하려면 주요 사회보장제도에 대한 분석을 종합적으로 할 수 있는 정보가 사회보장 행정데이터에 포함되어 있어야 한다. 구축 중인 사회보장 행정데이터는 통합행정데이터(integrated administrative data)이다. 사회보장 행정데이터를 구축하기 위하

2) 향후 사회보장 행정데이터는 다양한 형태의 데이터로 구축될 수 있을 것이다. 그럼에도 불구하고 이하에서는 대부분의 설명에서 현재 구축 중인 횡단면 자료 형태의 데이터를 '사회보장 행정데이터'로 명명하겠다.

[표 1] 구축 중인 횡단면 형태 사회보장 행정데이터의 개요와 잠재력

특성	잠재력
<표본의 크기> - 전 가구의 약 20% 표본	- 연령별, 지역별로 구체적인 수준의 분석을 가능하게 한다. - 정책 대상이 작은 정책에 대한 안정적인 분석도 가능하게 한다.
<높은 대표성> - 통계청 인구·가구 DB 기반 표본추출	- 높은 대표성을 확보하며, 특히 개인 단위 대표성은 타 설문조사와 비교하여 비교우위라고 할 수 있다. 가구 단위의 정보를 구성할 수 있다.
<정보 포괄성> - 성, 연령, 거주지, 소득, 재산, 사회정책 수급 등 주요 정보의 연계 (30개 이상 기관의 약 600개 정보)	- 소득이나 재산, 성, 연령, 지역, 건강 등의 정보를 포괄하여 정책의 수요와 성과분석을 가능하게 한다. - 다양한 사회보장제도 수급 정보를 포괄하여 정책의 경계를 넘어 정책의 목표 달성 여부 등 종합적인 분석을 가능하게 한다.
<정보의 정확성> - 공적 이전(현금과 현물 이전) 소득의 정확성 - 세금과 사회보험료 등의 정확성 - 상위 소득 및 재산정보의 정확성	- 공적 이전 수급 효과에 대한 정확한 분석이 가능하다.

자료: 이현주, 강희정, 오욱찬, 이원진, 이상정, 이혜정, ..., 한겨레. (2023). 사회보장행정데이터 구축 및 기초분석. p. 389. <표 11-2>를 재구성.

여 매년 생년월일, 성, 이름을 가명 처리하여 결합 키를 생성하고 이 결합키를 활용하여 여러 행정데이터를 연계한다.

사회보장 행정데이터는 수십 개 기관으로부터 약 600개 정보를 수집, 연계하고 있다. 사회보장 행정데이터에는 제도 형태별로 사회보험, 공공부조와 수당을 포괄하면서 현금 지원과 사회서비스 급여정보를 담고 있다. 뿐만 아니라 사회보장제도 효과 분석의 기초정보가 되는 성, 연령, 가구 규모, 가구주 여부, 지역, 소득과 재산, 건강 등 관련 정보가 포함되어 있다. 이러한 기초정보는 사회보장제도의 지원을 필요로 하는지 여부를 파악하는 욕구의 대리 정보이자 정책 결과를 대리하는 정보일 수 있다. 이 정보들이 욕구와 정책 결과의 대리변수인 이유는 대부분의 사회보장제도가 저소득, 경제활동을 하기 어려운 여건의 연령이나 신체적 능력을 전제

로 대상을 선정하고 건강 개선, 소득 개선 등을 정책 목표로 하기 때문이다. 한편 사회보장 행정데이터는 데이터 특성상 정보의 정확성이 높다. 특히 공적이전소득 정보는 매우 정확하다. 이러한 정보의 정확성은 정책 평가에서 매우 긴요한 강점이라 할 수 있다. 다만 금융재산의 수집이 어렵고 조세행정으로 누락되는 소득정보가 있어서 재산과 소득의 정확성 개선은 과제로 남아 있다.

나. 사회보장 행정데이터 구축의 목적과 의의

사회보장기본법 제42조에 의하면 사회보장위원회는 '사회보장 정책의 심의·조정 및 연구를 위하여 관계 기관의 장에게 사회보장 행정데이터가 모집단의 대표성을 확보할 수 있는 범위에서 ... 사회보장 행정데이터의 제공을 요청할 수 있다. 이 경우 사회

보장 행정데이터의 제공을 요청받은 관계 기관의 장은 특별한 사유가 없으면 이에 따라야 한다.’ 사회보장위원회에서 구축하는 사회보장 행정데이터의 활용 목적이 사회보장 정책의 심의·조정 및 연구로 공공성을 갖는다고 할 수 있다. 국민의 세금으로 국민에 의하여 형성된 행정정보이다. 행정정보의 경우 비록 그 정보가 가명 처리를 한 정보라고 하여도 활용 목적에서는 공공성을 우선하는 것이 바람직할 것이다. 이러한 노력은 국민의 정보가 공익을 위하여 사용되도록 지향하는 의미일 뿐 아니라 행정자료 활용에 대한 시민사회의 신뢰를 높이고 정보 사용의 안전성을 높이는 의미를 지니기도 한다.

사회보장기본법 제20조는 사회보장위원회의 근거 조항으로 ‘사회보장에 관한 주요 시책을 심의·조정하기 위하여 국무총리 소속으로 사회보장위원회를 둔다’고 사회보장위원회의 주요 역할과 위상을 적시하고 있다. 이 법 제20조 2항에는 심의·조정 사항을 열거하고 있는데 해당 사항들에는 ‘사회보장 기본계획 및 주요 계획’, ‘사회보장제도 평가와 개선’ 등이 포함되어 있다. 사회보장 계획과 개선은 사회보장제도의 평가를 출발점으로 한다. 사회보장위원회의 역할을 효과적으로 수행하기 위해서는 정책 효과와 한계에 대한 정확한 파악이 전제되어야 한다. 사회보장 행정데이터의 필요성은 이러한 역할 수행과 긴밀하게 연관된다.

사회보장제도에 대한 정확한 이해가 갖는 유용성은 비단 사회보장위원회의 역할 수행에 국한되지 않는다. 개별 정책을 담당하는 정부 조직도 해당 정

책의 성과를 이해하고 해당 정책과 상보적, 또는 대체적 성격을 지니는 유관 제도에 대한 이해를 필요로 한다. 그동안 사회보장제도의 성과는 주로 투입(input) 평가, 산출(output) 평가 중심으로 관리되는 경향을 보였다. 이러한 평가 경향도 데이터의 부족과 무관하지 않다. 결과(outcome)나 영향(impact)을 평가할 만한 적절한 데이터를 확보하기 어려웠다. 사회보장제도에 대한 종합적인 평가가 가능한 데이터가 마련된다면 과거와 비교하여 정책 평가가 평가의 범위와 지표 구성에서 개선될 수 있을 것으로 기대된다.

2 사회보장 행정데이터의 구축 배경: 제도의 발전과 정책의 근거 강화

가. 사회보장제도 이해를 위한 데이터의 한계

우리나라는 지난 40여 년 동안 매우 빠른 속도로 사회보장제도가 확충되어 왔다. 국내총생산(GDP) 대비 공적사회지출 비율은 1990년 2.6%에서 2015년 9.6%로, 2019년 12.3%로 증가하였다. 언급하는 주체에 따라 사회보장제도의 수가 다소 상이하지만 2020년대 들어 그 수는 몇백 종을 넘어선다는 설명에는 별 이견이 없다. <표 3>은 개인이나 가구 단위로 지원하는 주요 사회보장제도를 범주별로 단순하게 예시, 나열해 본 것이다. 주요 제도만 나열하여도 그 종류와 수가 적지 않은 것을 쉽게 알 수 있다.

[표 2] GDP 중 공적사회지출 비율의 변화

(단위: %)

연도	1980	1990	2000	2005	2010	2015	2019	2020	2021	2022
프랑스	20.1	24.4	27.7	28.8	31	31.8	30.7	34.9	32.7	31.6
독일	21.8	21.4	25.5	26.4	26.1	25.1	25.6	27.9	27.6	26.7
한국	..	2.6	4.4	5.9	7.9	9.6	12.3	14.4	14.9	14.8
스웨덴	24.5	26.9	26.4	27.1	25.8	26.1	25.1	25.9	24.9	23.7
영국	15.6	14.9	16.8	19.1	23.1	21.3	19.5	22.5	22.1	..
미국	12.9	13.2	14.1	15.5	19	18.5	18.3	23.9	22.7	..
경제협력개발기구 (OECD)	14.4	16.4	17.3	18.1	20.4	20.1	20.1	23	22	21.1

자료: OECD. *SOCX stat*. <http://stats.oecd.org>에서 2023. 9. 24. 인출.

[표 3] 2020년대 사회보장제도의 예

제도의 범주	제도의 예
사회보험	건강보험, 고용보험, 국민연금(과 지역연금 등), 산재보험, 요양보험
소득보장	기초연금, 취업 성공 패키지 등 취업 지원, 청년구직촉진수당, 청년내일채움공제, 근로·자녀장려세제, 아동수당, 모성보호육아 지원, 가정양육수당, 아동발달지원계좌, 한부모가족지원, 장애수당, 장애인연금, 예술인 생활안정자금
사회서비스	모성보호육아 지원, 보육료 지원, 노인맞춤 돌봄서비스, 재정일자리 지원, 노인일자리사업, 장애인일자리, 차상위 건강보험 본인부담 경감, 저소득층 장학사업, 국가장학금, 장애인활동 지원, 임대주택, 산모산생아 건강관리 지원, 영양플러스사업, 기저귀조제분유 지원, 치매검진 지원, 치매치료관리비 지원, 장애인활동 지원, 발달재활서비스, 발달장애인 주간활동 지원, 발달장애인 방과후활동 지원, 여성장애인 출산비용 지원, 중증장애인 직업재활 지원, 중증장애인 근로지원인 지원, 중증장애인 고용 지원, 통합문화이용권, 스포츠 강좌 이용권, 지역사회서비스 투자
공공부조	생계급여, 교육급여(저소득층 학비 지원), 주거급여, 의료급여, 자활지원, 긴급복지지원제도

자료: 저자 작성

다종다양한 사회보장제도의 역할을 고려한 정책 기획이나 개선안을 마련하기 위해서는 정교한 평가가 이루어져야 한다. 국민을 사회적 위험으로부터 효과적으로 보호하기 위해서는 누가 이러한 사회보장제도에서 배제되어 사회적 위험에 직면하게 되는지, 또는 누가 사회적 위험 대비 과한 사회적 보호를 누려 오남용의 위험에 놓이는지를 파악하여야

한다. 더불어 공동의 정책 목표를 가진 정책들이 실제 해당 목표를 달성하는 데 기여하고 있는지, 역할 관계가 어떻게 작동하는지를 개별 정책의 경계를 넘어 정책군 평가가 가능해야 한다. 예를 들어 소득빈곤의 위험 완화를 목표로 하는 여러 제도들이 정책 목표 달성에 얼마나 기여하는지를 파악하려면 해당 제도들의 정보가 포함된 데이터를 필요로 한다.

[표 4] 주요 조사 자료 중 사회보장제도 정보

가계금융복지 조사 자료	한국복지패널 조사 자료
공적연금, 기초연금, 양육수당, 장애인연금, 현금성 공공부조, 근로장려금, 자녀장려금, 기타 정부 보조금	공공부조, 장애수당, 연금 가입과 수급, 기초연금, 고용보험 가입과 수급, 건강보험 가입과 의료지원, 보육료 지원, 아동수당, 양육수당, 장애인·노인 돌봄서비스 이용, 한부모 지원, 학비 지원, 주거 보장, 긴급복지, 근로장려세제, 에너지요금 감면, 통신비 보조 등 보조금, 기타 바우처

자료: 이현주 외. (2023). 사회보장행정데이터 구축 및 기초분석. p. 41 (표 1-1)을 기초로 재구성.

사회보장제도의 성과를 평가할 수 있는 데이터는 충분하지 않았다. 사회보장제도를 평가한 많은 연구에서 활용된 대표적인 데이터를 꼽으라면 가계금융복지 조사 자료와 한국복지패널 자료일 것이다. 이와 더불어 고용 관련 정책 평가에서는 한국노동패널이, 노인정책에 대한 평가에서는 고령자패널이 자주 활용된 데이터라 할 수 있다. 기타 개별 정책 영역에서 노인, 장애인 등 특정 인구집단의 사회적 위험과 정책 수급을 조사한 설문조사 자료들이 분석에 활용된 바 있다. 설문조사는 최근으로 올수록 표본인 응답자를 만나는 것이 더 어려워지고 설사 응답 대상을 만나도 응답 기피가 심화되어서 시간과 금전적 비용 부담은 커지고 대표성 확보는 더 어려워지는 상황에 놓여 있다. 이러한 어려움을 차치하고라도 설문조사 자료는 정책분석에서 근본적인 한계를 지니고 있었다.

설문조사 자료의 대부분은 크게 두 가지 문제에서 자유롭지 못했다. 첫째는 필요한 정책정보 중 극히 일부 정보만을 포함하여 정책분석의 범위를 제한한다는 점이다. 다른 하나는 설문으로 조사된 정보들이 얼마나 정확한 것인가에 대한 의문이 적지

않다. 아래는 가계금융복지 조사 자료와 한국복지패널 조사 자료에 포함된 사회보장제도 정보이다.

앞서 언급한 주요 사회보장제도와 비교하여 상기 조사 자료에 포함된 사회보장제도의 정보가 매우 부족함을 알 수 있다. 특히 소득분배 분석에서 공식 통계 자료인 가계금융복지 조사 자료의 경우 횡단면 자료로 표본 안정성을 갖춘 자료이다. 각 시점의 정책 효과를 보다 정확하게 이해할 수 있는 이점이 있으나 포함된 사회보장 정보는 매우 제한되어 있음을 알 수 있다. 오랜 시간 사회보장제도 분석에 활용할 수 있는 적절한 데이터를 확보하는 것이 쉽지 않았다.

정책 연구 중 일부 연구는 선도적으로 대안적 자료 확보를 시도하였다. 분석하고자 하는 제도의 행정정보를 추출하여 이 자료를 분석한 연구들이 일부 추진된 바 있다. 이 연구들은 행정데이터를 활용하여 제도 수급자가 어떤 특성을 지니는 집단인지를 보다 큰 표본 규모의 정보를 활용하여 안정적으로 파악할 수 있는 기회를 얻게 되었다. 다만 이 경우에는 해당 정보가 특정 정책 수급자 정보로 국한되는 경우가 다반사여서 수급자가 전 인구, 전 가구

[표 5] 설문조사 자료와 개별 정책 행정데이터의 한계

자료	한계	영향
설문조사 자료	정보 포괄성 부족	일부 제도에 한정된 정책분석만 가능 정책 목표 중심의 정책군 분석 불가 제도 변화를 반영한 정책분석 불가
	정보 정확성 부족	제도 효과 분석의 신뢰성 저하 특히 서비스 수급의 경우 서비스 효과 분석 불명확 수급 여부와 만족도 중심의 분석 편향 초래
(통합되지 않은)제도별 행정데이터	표본의 대표성	전 국민, 전 인구 중 수급자의 사회경제적 지위 및 변화 파악 불가 대조군 없는 정책 평가

자료: 저자 작성

중 어떠한 수준의 사회경제적 지위에 있는 집단인지를 이해하기 어려웠고 결국 정확한 정책 효과를 파악할 수 없었다. 패널 형태가 아니라면 정책 수급의 효과를 전후 비교를 통하여 확인하지도 못하였다. 통합되지 않은 행정데이터는 대개 수급자 외 다른 인구집단 정보가 포함되지 않은 데이터여서 대조군을 갖춘 분석을 하기도 어려웠다.

설문조사에서 가장 파악하기 어려운 것은 사회서비스의 수급으로 인한 경제적 이익과 효과들이다. 사회서비스를 수급한 사람들 중 적지 않은 수의 사람들은 이용한 서비스의 (본인 부담을 포함한 전체) 가격을 모르고 있다. 결과적으로 서비스 이용으로 인한 경제적인 또는 실질적인 이익을 정확하게 답변하기 어렵다. 사회서비스에 대한 조사에서는 많은 문항이 수급 여부, 그리고 서비스에 대한 만족도를 중심으로 구성돼 있다. 수급 효과 분석에서도 수급 여부와 만족도를 중심으로 정책 평가가 이루어

지는 편향을 초래되기도 한다.

나. 사회보장 행정데이터 구축의 기반

사회보장제도가 빠르게 발전하여 예산과 수급자 규모에서 크게 확충되었지만 사회보장제도를 정확하게 분석할 수 있는 데이터는 마련되지 않은 상황이 오래 지속되었다. 사회보장제도와 관련된 정책 이슈가 계속 이어져 사회적으로 큰 관심사가 되고, 정책 개선의 근거 강화를 위한 사회적 압력이 커졌지만, 이에 대응할 만한 데이터 기반이 갖추어지지 못한 것이다.

2010년대 이후 통합된 사회보장 행정데이터는 사회보장제도를 보다 넓게 그리고 정확하게 이해하는 데 중요한 대안으로 거론되기 시작하였다. 사회보장 행정데이터에 대한 심층적 논의가 본격화된 것은 2010년 이후³⁾ 사회보장제도 신청 및 수급 정

3) 1980~90년대에는 국제 업무의 전산화가 추진되었고 2010년에는 ‘행복e음’으로 불리는 사회보장 정보 시스템이 개통되었다.

보가 전산화되기 시작하고 해당 정보가 체계적으로 관리되기 시작하면서라고 할 수 있다.

2021년 사회보장기본법이 개정되기 이전 행정 데이터 등 각종 데이터 이용에서 획기적인 변화를 맞이하게 된 것은 개인정보보호법의 개정이었다. 2019년 데이터 이용을 활성화하기 위한 목적으로 「개인정보보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」 등 속칭 데이터 3법 개정안이 발의되었다. 데이터 이용에 관한 규제 혁신과 개인정보 보호 협치 체계 구축을 위한 시도로 이 개정안은 2020년 1월 9일 국회 본회의를 통과하였다. 개인정보보호법 개정으로 '가명정보'라는 개념이 도입되었고 개인정보 보호를 위한 유사, 중복 법률을 정비하여 추진체를 일원화하는 등 주요 개정 내용 담겨 있었다.⁴⁾ 이후 행정데이터도 가명 처리를 하고 결합하여 정책분석이 가능한 환경이 조성되었다.

개인정보보호법의 개정으로 사회보장 관련 행정 데이터를 가명 처리하여 수집하고 이를 정책분석에 활용하고자 하는 노력이 더욱 본격화되었다. 그러나 실제 행정데이터를 가명 처리한다고 하더라도 행정데이터를 관리하는 주체별로 법 조항의 해석과

협력을 위한 태도가 상이하여 안정적으로 행정데이터를 추출, 연계하는 것이 쉽지 않았다. 결국 법령 수준의 근거 조항을 필요로 하게 되었다. 이러한 배경에서 2021년 사회보장기본법을 개정하게 된 것이다. 앞서 언급한 바와 같이 이 법 제42조에 사회보장 행정데이터의 제공 요청 근거를 마련하고, 제43조에는 이 데이터의 분석, 활용을 위한 분석센터 설치, 운영 조항을 두었다. 사회보장기본법의 개정으로 별도의 법적 근거를 마련하여 사회보장 행정데이터를 구축하고, 이 데이터를 정책기획을 위한 자료로 활용하게 되었지만 개인정보의 가명 처리, 가명 정보의 활용에 대한 수용적인 환경 조성도 사회보장 행정데이터 구축에 긍정적 영향을 주었다고 판단된다.

3 주요 정책 과제와 행정데이터의 수요 전망

사회보장 행정데이터 구축의 목적을 고려하면 사회보장제도의 개선 과제 또는 정책 이슈와 데이터 활용을 위한 향후 과제가 긴밀하게 연관될 것이다. 미래 사회보장제도의 발전을 위한 정책 과제가

4) 데이터 3법의 개정 이후 이전과 비교하여 관련 법령이 체계화되었지만 신용정보법의 경우 통합개정된 것은 아니어서 개인정보보호법과 유사한 규정을 남겨 두고 있다. 감독기관, 신용정보 관리 보호인, 신용정보 활용 체제 등은 개인정보보호법과 거의 중복되는 내용으로 한계를 남겨두고 있다는 평가를 받는다. 즉 개인정보 보호 관련 법, 보호 체계가 통합 일원화되는 데 아직 개선해야 할 과제가 남아 있다. 관련 내용은 다음을 참조하기 바란다. 김현정. (2023). 개정 「개인정보 보호법」의 의미와 과제. KISO 저널 제51호. 대한민국 정책브리핑. (2021). 데이터 3법.

한편 2023년 개인정보보호법이 개정되어 이 법 제11조의 2항에 보호위원회가 공공기관 중 중앙행정기관 및 그 소속기관, 지방자치단체, 그 밖의 대통령령으로 정하는 기관을 대상으로 매년 개인정보 보호 수준의 평가를 하도록 규정하고 있다. 필요한 평가로 판단되지만 해당 평가를 공공기관 중 일부로 한정하고 민간기관에 유사한 의무를 부과하지 않은 점에 대해서는 이견이 있을 수 있다. 한편 이번 법 개정으로 평가 대상에 포함되는 일부 공공기관의 경우 2024년 이전까지 이러한 평가에 대비하는 준비가 필요한 상황이 있다.

무엇인가에 대하여 많은 제안이 있을 수 있다. 주요 정책 과제 중 몇 가지만 예시하면 <표 6>과 같다. 저출산 대응이 제대로 평가되고 개선되려면 수많은 관련 정책을 종합적으로 분석할 수 있어야 한다. 노인 빈곤을 완화하고자 하는 정책기획에서도 마찬가지이다. 저출산 대응 정책으로 출산지원, 육아휴직 뿐 아니라 관련된 각종 고용지원, 주거지원, 교육지원, 조세지원, 금융지원 등 여러 제도 정보가 하나의 데이터로 구축되어야 한다. 더욱이 이러한 데이터가 여러 해 동안 지속적으로 구축, 활용될 수 있어야 한다. 노인 빈곤을 완화하는 정책기획에서도 공공부조 제도뿐 아니라 국민연금과 직역연금, 기초연금, 노인의 경제활동 및 근로소득 정보, 가구 구성과 가구원의 소득정보가 공히 포함된 데이터를 필요로 한다.

노동시장의 변화에 따라 일 방식이 변화되면서 사회보장제도의 설계에서 새로운 고안을 필요로 하고 있다. 비단 고용보험뿐 아니라 여타의 사회보험과 각종 취업지원 등의 정보가 기존 사회보장제도

의 대응력을 평가하고 개선안을 마련하는 데 필요하다. 우리 사회에서 사회 이동성이 낮아지고 있다는 불편한 추정이 이어지고 있다. 하지만 이러한 분석을 정확하게 할 수 있는 데이터도 역시 부족하다. 가족관계, 부모와 자녀의 학력, 취업과 소득 등이 포함된 데이터가 별로 없기 때문이다. 표본 수가 그리 많지 않은 일부 패널 자료가 이러한 분석에 활용되고 있는 실정이다. 생활시설이나 병원보다 지역에서 오래 머물 수 있는 서비스 환경을 조성하는 것이 이상적인 돌봄 제공이라는 연구 결과가 적지 않다. 지역사회 기반 돌봄서비스를 강화하자면 지역 기반 서비스와 관련된 요인을 파악하여야 한다. 이를 위해서는 생활시설이나 병원을 이용하는 사람과 재가서비스를 이용하는 사람에 대한 정확한 정보가 필요하다.

지방자치제가 강화되면서 지방자치단체 사회보장제도의 영향도 커지고 있다. 사회보장제도의 구성과 효과를 제대로 이해하려면 지방자치단체의 행정정보가 더 중요해진다. 한편 복지정책의 설계와

[표 6] 주요 정책 과제 예시와 데이터 수요 전망

주요 정책 과제	데이터 수요 전망
저출산 대응, 노인 빈곤 완화, 노동시장 변화에 대응하는 사회보장제도 개편	통합행정데이터
사회 이동성의 강화	패널화된 통합행정데이터
지역사회 기반 돌봄서비스 강화	시설, 가구, 개인 단위 연계 통합행정데이터
서비스 이용의 지역 격차 완화, 지방자치단체의 특성을 고려한 서비스 기획	지방자치단체, 중앙정부 연계 통합행정데이터
사회보장제도와 관련된 국민의 인식 변화	설문조사와 결합된 통합행정데이터

자료: 저자 작성

복지정치와 관련하여 일에 대한 태도, 사회보장제도의 확충에 대한 기대와 국민 부담에 대한 태도를 비롯하여 사회 구성원의 인식에 대한 관심이 높아지고 있다. 이러한 국민 인식은 설문조사를 통하여 정보를 수집할 수밖에 없다. 사회보장제도의 수급과 인식의 관계를 이해하자면 설문조사 자료와 행정데이터의 결합이 필요하다.

〈표 6〉은 사회보장제도의 주요 정책 과제를 고려할 때 필요한 데이터 수요를 정리한 것이다. 현재 매년 구축 중인 횡단면 형태의 사회보장 행정데이터는 저출산 대응 정책 종합분석, 노인 빈곤에 대한 각종 정책의 영향 분석 등을 가능하게 할 것으로 기대된다. 앞서 언급하였듯이 해당 데이터가 개인과 가구 단위의 각종 사회보장 수급 정보를 포괄하고 있기 때문이다.

표본의 규모, 정보 포괄성, 정보 정확성에서 상대적으로 강점을 가지는 행정데이터이지만 아직 정책 과제별로 정책의 근거를 강화하기 위해서는 데이터 형태와 정보의 구성에서 보완을 해야 한다. 사회보장 행정데이터는 형태와 내용이 고정된 것이라 할 수 없다. 오히려 사회보장 정책의 심의, 조정 연구를 위하여 필요한 데이터라 보는 것이 더 적절하다.

향후 패널데이터, 시설의 이용자 정보와 연계된 통합데이터, 지방자치단체 정책정보와 연계된 통합데이터, 설문조사와 연계된 통합데이터 등 형태 다변화가 이루어져야 한다. 데이터 생산 및 확충의 우선순위, 데이터별 생산 주기와 방식 등에 대한 논의와 협력이 순차적으로 진행되어야 할 것이다.

4 사회보장 행정데이터의 개선 과제

사회보장제도의 주요 정책 과제를 고려한 사회보장 행정데이터의 확충을 위해서는 몇 가지 개선이 필요하다. 개선 과제들을 크게 구분하여 보자면 결합키의 정확성 향상, 표본추출의 방식 변화, 정보 정확성을 높이기 위한 정보 포괄성 보완으로 나누어진다. 그리고 개선 과제들이 체계적으로 해결되려면 공통적으로 협력을 위한 체계적 논의 구조가 강화되어야 한다.

사회보장 행정데이터가 향후 패널 형태의 데이터로 추가, 구축되거나 집단시설 거주 가구 거주자 정보 또는 설문조사 정보와 연계되어 새로운 통합 행정데이터로 구축되려면 결합키가 정교하여질 필요가 있다. 현재 사회보장 행정데이터를 구축하면서 성, 생년월일, 이름을 가명 처리한 상태로 결합키를 생성해 활용한다. 성, 생년월일, 이름이 같은 표본이 적지 않아 표본추출 후 상당수의 중복 표본을 제거하여야 한다. 2017년 주민등록 DB에서 전 가구의 7% 표본추출을 목표로 가구주 중심으로 1차 표본추출, 가구원 개인 표본추출 순으로 표본을 추출한 연구⁵⁾에서도 결합키 중복으로 인한 표본 삭제 비율이 33.1%였다(이현주, 오욱찬, 이윤경, 이원진, 성재민, 이길제, ..., 이병재, 2020, p. 77.). 표본 중 중복 개인은 12.6%, 중복 개인이 속한 가구는 가구 중 25.7%, 중복 개인이 속한 가구에 속한 가구원 전체는 개인의 33.1%였다. 2020년 기준으로 통계청의 인구·가구 DB를 기초로 표본을

추출한 경우에도 중복 가구원은 30.6%였다(이현주 외, 2023, p. 59.). 결합키의 중복으로 제거되는 표본은 일종의 체계적인 특징을 지닌다. 예를 들어 연령, 거주 지역, 가구 규모에서 편향성이 존재한다.⁶⁾ 따라서 가중치 조정으로 대표성을 확보하여야 한다.

이러한 결합키의 중복으로 인한 부정확성은 향후 데이터 확충에 여러 가지 제약을 낳게 된다. 첫째, 패널데이터를 구축할 때 분리 가구의 생성 가구 정보를 완성하기 어렵다. 예를 들어 생성 가구의 가구원 중 중복된 연계키의 가구원이 존재할 수 있다. 또는 기존 가구의 가구원이라 하여도 새로운 가구원의 연계키가 중복 연계키가 될 수도 있다. 설문조사와 집단시설 거주 가구 거주자 정보와 연계할 때는 결합키의 중복으로 인한 표본 크기의 제약 정도가 심각할 수 있다. 대부분의 설문조사 표본은 크지 않아서 표본이 몇 천 개에 불과하고 큰 경우 2만 개 사례를 포괄하는 수준에 그친다. 이러한 표본의 경우 설문조사에서 성, 생년월일, 이름을 응답 내용으로 포괄하는 경우도 흔치 않고 설사 해당 내용을 가명 처리한 상태로 확보하여 결합키를 마련한다 하여도 설문 응답 사례 중 약 30%를 중복으로 삭제하여야 한다. 결국 표본의 크기가 매우 작아질 위험이 있다. 집단거주시설 거주자의 경우 대개 그 수가

많지 않다. 아동양육시설의 보호 아동 수는 2021년 1만 72명이고(보건복지부·한국보건사회연구원, 2022, p. 96.), 장애인거주시설 거주 장애인의 수는 2만 8565명이다(보건복지부·한국보건사회연구원, 2022, p. 145.). 집단시설 거주 가구 거주자에 대한 분석에서도 결합키의 부정확성, 중복 발생은 적정 수준의 표본 확보를 어렵게 할 수 있다.

결합키의 정확성을 높이는 대안이 필요하다. 주민등록번호는 현재 개인정보보호법상 가명 처리를 한다고 하여도 이를 연계키 개선을 위하여 활용하기 어렵다. 우리나라 사회 구성원에게 주민등록번호와 무관한 별도의 새로운 아이디를 부여하자는 안부터 지역정보와 같이 공유 가능성이 높은 정보를 표준화하고 이를 수집하여 행정데이터 결합키 생성에 추가 활용하자는 안까지 여러 대안이 논의된 바 있다. 하지만 아직 뚜렷한 대안을 마련하지 못하고 있다. 북유럽 국가 사례를 참고하자면 스웨덴은 통계청에서 개인 식별 아이디를 이용하여 행정데이터를 결합하고 비식별 조치 후 승인된 연구자들이 이 데이터를 분석할 수 있도록 지원하고 있다(이현주 외, 2020, p. 374.). 개인 식별 아이디는 우리나라의 주민번호와 유사한 성격을 지니는 것이다. 정부와 공공 부문에서 통합행정데이터를 구축할 때는 민간의 결합과 달리 결합의 정확성을 높

5) 해당 연구는 아래와 같다.

이현주, 오욱찬, 이윤경, 이원진, 성재민, 이길제, ..., 이병재. (2020). 사회보장정책 효과성 분석을 위한 행정데이터 연계·활용 방안. 세종: 한국보건사회연구원.

상기 연구는 사회보장기본법 개정 전 행정안전부의 가이드라인에 따라 사회보장 행정데이터를 결합한 실험연구로 진행되었다.

6) 예를 들어 3인 이상 가구에서 중복으로 인한 표본 제거 비율이 높은 편이다.

[표 7] 사회보장 행정데이터의 개선 과제

수요 예상 사회보장 행정데이터	사회보장 행정데이터의 과제	행정데이터 개선 기반
패널행정데이터, 설문조사 연계 행정데이터, 집단시설 거주 가구 거주자 정보 연계 행정데이터	정확한 연계키의 확보	
집단시설 가구 거주자 포괄 행정데이터	표본의 확장, 시설정보의 목록화 및 표준화	⇒ 협력을 위한 거버넌스 강화
양질의 통합사회보장 행정데이터 지방자치단체 행정데이터 연계 행정데이터	정보 포괄성 보강	

자료: 저자 작성

이는 보다 적극적인 대안을 고려할 수 있을 것이다.

구축 중인 사회보장 행정데이터는 통계청 개인 가구 DB 중 일반 가구를 기반으로 표본추출 틀을 구성한다. 기숙사나 노인요양시설, 보육원 등 사회 시설에 집단으로 살고 있는 가구인 집단시설 가구나 외국인으로 구성된 외국인 가구는 표본으로 포함하지 못하고 있다(이현주 외, 2023, p. 57.). 설문조사 자료의 대부분도 집단시설 가구 거주자나 외국인 가구의 외국인을 표본으로 포괄하지 못하고 있다. 집단시설에 거주하는 국민은 사회보장제도의 보호를 필요로 하는 대표적인 취약집단이며, 사회보장제도에서 우선하여 사회적 보호를 강화하여야 하는 집단이기도 하다. 해당 집단에 대한 정보는 사회보장 행정데이터에서 매우 중요하다. 향후 해당 집단을 표본으로 포괄하고 해당 집단이 보다 이상적인 서비스를 지원받을 수 있는 정책 설계에 활용될 데이터 구축이 필요하다. 물론 이러한 데이터 구축을 위해서는 표본의 확장뿐 아니라 시설정보 중 수집한 정보를 선별하는 준비 작업들이 함께 진행

되어야 한다.

정보의 포괄성을 높이는 노력도 지속되어야 한다. 앞서 언급한 것과 같이 소득과 재산 정보 중 누락된 정보를 보강하기 위한 협력을 추진하여야 한다. 금융재산정보는 사회보장제도의 대상자 선정 등에서 중요한 정보이다. 즉 정책분석에서 기초적인 정보이기도 하다. 하지만 수집과 결합은 매우 어려운 상황이다. 2021년 한국데이터산업진흥원에서 지원한 연구 보고서에 따르면 가명 정보 관련 국내법이 개인정보보호법과 특별법인 신용정보법이라는 이중적인 체계로 구성되어 있고, 특별법인 신용정보법이 우선 적용돼 법체계의 복잡한 구조가 존재한다고 평가한 바 있다(부산대학교, 2020, pp. 29-34.). 2021년 7월 관계 부처 합동으로 발표한 자료에 따르면 가명 정보 결합 건수는 105건이며, 이 중 73%가 신용정보법 기반이었다(부산대학교, 2020, pp. 51-2.). 신용정보법에 기반하여 가명 정보를 결합한 사례 중 공공은 13%, 나머지는 개인신용조회회사, 유통, 통신 등이었다. 신용정

보법 적용으로 가명 처리된 금융정보의 활용 상황을 고려할 때 사회보장 정책분석을 목적으로 하는 금융 재산정보의 수집은 조금 더 전향적으로 검토될 필요가 있다. 사회보장제도의 적지 않은 제도는 경제력을 조사하여 지원 대상을 선정한다. 경제력은 소득 뿐 아니라 금융재산, 부채를 포함하여 평가된다. 정책의 효과나 운영의 적절성은 금융재산과 부채를 파악하여야 정확하게 평가될 수 있다. 끝으로 행정정보 중 관리되지 못하는 정보의 확보를 위한 점검을 할 필요가 있다. 예를 들어 사회서비스의 본인 부담금 정보나 조세 행정으로 누락되는 소득정보의 확보 등은 중장기적 과제가 될 수 있을 것이다.


5 나가며

현재 구축 중인 사회보장 행정데이터는 횡단면 형태의 통합행정데이터로서 개인·가구 단위의 사회보장제도 수급 효과를 종합적으로 분석할 수 있는 기회를 제공하고 있다. 향후 패널 형태 데이터, 집단시설 가구 거주자 정보, 설문조사 자료 및 지방자치단체의 행정데이터와 결합한 데이터로 확충된다면 주요 사회보장제도의 정책 근거가 더 강화될 수 있을 것으로 기대된다.

앞서 언급한 바와 같이 사회보장 행정데이터의 개선은 집단시설 가구 거주자 등으로 표본을 확장하고 금융재산과 같은 누락 정보를 추가하여 정보의 포괄성을 높이면서 결합키의 정확성을 보완하는 방향으로 추진되어야 한다. 사회보장 행정데이터

개선을 위한 과제 각각은 모두 심층적인 논의와 협력을 필요로 한다.

사회보장 행정데이터 구축과 관련된 여러 주체의 입장을 고려할 때 데이터 개선 노력은 합리적이고 체계적인 논의체, 거버넌스의 강화와 분리하여 생각하기 어렵다. 관련 주체들의 다양한 입장을 보다 적극적으로, 합리적으로 조정할 수 있는 체계적인 논의체가 강화되어야 관련 쟁점들이 제대로 정리되고 이후 데이터 확충 노력이 효과적이고 효율적으로 진행될 수 있다. 북유럽 국가들과 비교하여 다소 늦게 정부가 행정데이터 활용에서 체계적인 정책 추진을 한 미국과 영국처럼 한시적인 논의 구조를 갖추어 운영할 수도 있을 것이다. 미국의 경우 2019년 제정된 증거기반정책법에 의거해 데이터 자문위원회(ACDEB: Advisory Committee on Data for Evidence Building)를 설치한 바 있다. 이 위원회에는 국가통계 책임자, 증거 구축 관련 정부 관계자, 개인정보 보호 책임자 등이 포함되었다(윤건, 김철우, 2021. pp. 95-6.). 이 위원회는 연방 데이터 사용과 관련된 평가와 분석, 권고의 기능을 수행하며 구체적으로 데이터 공유와 연계, 개인정보 보호 기술개발 활성화 방안을 제시하고, 증거 생산을 위한 데이터 공유나 데이터 이용 관련 조항이 잘 지켜지고 있는지를 평가하는 기능을 수행하였다. 개인정보 보호는 매우 중요하다. 개인정보 보호를 목적으로 하는 논의체와 더불어 행정데이터의 안전한 활용으로 정책의 근거를 강화하기 위한 논의체를 구성하는 것을 적극 고민하여 볼 수 있을 것

이다. 행정데이터를 활용하여 정책의 근거를 강화하기 위하여 필요한 여러 이슈를 함께 논의하고 대안을 모색할 수 있는 구성이 되도록 이 논의체에는 개인정보 보호 전문가뿐 아니라 정책통계 전문가, 정책 평가 등 정책분석 연구자의 참여도 고려되어야 하겠다. 데이터와 관련하여 정부에 설치된 위원회 등의 조직은 적지 않다. 초점으로 하는 데이터의 성격과 범위가 조금씩 상이하고 위원회의 구성과 역할도 동일하지는 않다. 여러 위원회가 존재하지만 행정데이터를 활용한 정책 근거를 강화한다고 할 때 각종 쟁점을 종합적으로 살펴보는 위원회가 분명하게 떠오르지는 않는다. 현재 산재한 논의체, 데이터 플랫폼의 옥상옥이 되지 않도록 하면서 실질적 조정 기제를 가진 논의 구조를 신설, 또는 강화하는 것이 사회보장 행정데이터 개선에서 가장 핵심적인 출발일 수 있다. 

참고문헌

- 김현정. (2023). 개정 「개인정보 보호법」의 의미와 과제. **KISO저널 제51호**. <https://journal.kiso.or.kr/?p=12195>에서 2023. 9. 27. 인출.
- 보건복지부·한국보건사회연구원. (2022). **통계로 보는 사회보장 2022**.
- 법제처 국가법령정보센터. (n.d.). 사회보장기본법. <https://www.law.go.kr/법령/사회보장기본법>.에서 2023. 9. 24. 인출
- 부산대학교. (2020). **가명정보 활용 지원 방안 연구**. 한국데이터산업진흥원. https://www.kdata.or.kr/kr/board/info_01/boardView.do?pageIn dex=1&bbsIdx=33307&searchCondition=all&searchKeyword=%EA%B0%80%EB%A A%85%EC%A0%95%EB%B3%B4에서 2023. 10. 3. 인출.
- 윤건, 김철우. (2021). 증거기반정책과 데이터기반행정 법: 미국과 영국의 법·제도 비교 및 합의. **가천법학 vol.14**, no. 3, 통권 48호, pp. 83-112.
- 이현주, 강희정, 오욱찬, 이원진, 이상정, 이혜정, 성재민, 이길제, 이재원, 이병재, 한겨레. (2023). **사회보장행정데이터 구축 및 기초분석**. 세종: 보건복지부, 한국보건사회연구원.
- 이현주, 오욱찬, 이윤경, 이원진, 성재민, 이길제, 박형준, 이병재. (2020). **사회보장정책 효과성 분석을 위한 행정데이터 연계 활용 방안**. 세종: 한국보건사회연구원.
- 대한민국 정책브리핑. (2021). **데이터 3법**. <https://www.korea.kr/special/policyCurationView.do?newsId=148867915>에서 2023. 9. 27. 인출.
- OECD. (n.d.). *SOCX stat*. Retrieved from <http://stats.oecd.org>. 2023. 9. 24.

Integrated Administrative Data and Evidence-based Social Policy

Lee, Hyonjoo

(Korea Institute for Health and Social Affairs)

This article examines social security administrative data in terms of its characteristics and the background of its formation and seeks ways to overcome its current limitations to improve its use in research. In particular, we examine past deficiencies in social security data, the subsequent need for a social security administrative database, and recent amendments to the Framework Act on Social Security that supported its creation. This study highlights the advantages of having a cross-sectional database of social security administrative records, currently under construction as mandated by law, and discusses necessary improvements. Given the policy challenges in the field of social security, social security administrative records should also be structured into panel data in a way that combines with administrative data on facility residents and survey findings. Furthermore, there is a need for a committee to oversee the suggested improvements.