

# 소득분배 연구를 위한 건보공단 빅데이터의 의의와 한계: 서울시 사례연구를 중심으로\*

유 중 성\*\* · 김 민 혜\*\*\* · 김 승 연\*\*\*\* · 유 수 진\*\*\*\*\*

## 요약

소득분배지표의 측정을 비롯한 소득분배의 연구에 있어 기존의 가구조사 자료와 건강보험공단 빅데이터의 장단점을 비교, 고찰해보고자 한다. 건보공단 자료는 전국민의 시장소득과 공적연금소득 및 부동산 재산세 과표를 각각 2003년 및 2005년부터 패널로 구축하고 있는 유일한 자료로서 특히 지역 차원의 소득분배를 파악하는 데 유용하다. 그러나, 시장소득도 아직은 미비하고 가처분소득을 구하기가 어렵다는 한계가 존재한다. 다행히 시장소득의 파악률이 급격하게 상승하고 있으며, 조만간 거의 완벽한 수준에 이를 것으로 보인다. 현재도 사업소득의 경우 특고, 프리랜서 등 인적용역 사업소득까지 인별로 사업소득 정보가 제공되고 있다. 끝으로 가계금융복지조사를 기반으로 하는 소득분배 지표의 개선과 지역 단위의 소득분배 지표 측정을 위해 서베이 자료와 행정자료의 결합을 비롯한 개선방안을 제시한다.

**주요어:** 소득분배, 행정자료, 건보공단 빅데이터, 전인구 소득재산DB

\* 이 논문은 서울연구원과 가천대학교 불평등과 사회정책연구소의 협동연구인 “행정자료를 활용한 서울의 불평등과 빈곤연구(1차연도)”(서울연 2020-CR-15)의 결과에 기초함.

\*\* 제1저자, 가천대학교 가천리버럴아츠칼리지 교수 겸 불평등과 사회정책연구소 소장(jongsungyou@gmail.com)

\*\*\* 교신저자, 창원대학교 사회학과 조교수(minhyekim@changwon.ac.kr)

\*\*\*\* 공동저자, 서울연구원 도시사회연구실장(francies@si.re.kr)

\*\*\*\*\* 공동저자, 가천대학교 불평등과 사회정책연구소 데이터 연구원(suzinyou.sy@gmail.com)

## 1. 서론

전통적으로 사회과학 및 정책연구에 활용되어 온 서베이자료의 한계에 대한 인식과 함께 행정자료의 적극적인 활용이 대안으로 제시되고 있다(구인회 외, 2020; 유종성 외, 2020; 이현주 외, 2020; Chun et al., 2021; Cole et al., 2020; Penner & Dodge, 2019). 지난해 데이터 3법<sup>1)</sup>의 통과에 이어 개인정보보호위원회 산하에 가명정보결합종합지원시스템(<https://link.privacy.go.kr/nadac/index.do>)이 개통되었고, 올해 2월 25일 ‘한국 행정데이터 연구자 네트워크’(Administrative Data Research Network-Korea; ADRNK)가 발족하는 등 행정데이터의 구축과 활용에 대한 학계와 정부의 관심이 제고되고 있다.

특히 소득분배 지표의 측정 및 소득분배에 관한 연구에 있어 가계조사 자료의 한계가 노정되면서 행정자료의 활용이 모색되고 있다(Bee & Rothbaum, 2019; Medalia et al., 2019). 여러 나라에서 가계조사의 응답률 저하, 소득정보의 부정확성, 조사비용 증가의 문제 등이 제기되면서 소득세 자료 등 행정자료의 활용이 점차 증가하고 있다. 이에 따라 최근 북유럽 국가들을 중심으로 소득세와 사회보장 급여자료 등을 이용하여 가구소득을 작성하는 방법이 확산되고 있으며, 점점 더 많은 나라들이 가구조사 자료와 행정자료를 보완적으로 사용하는 추세이다(통계청, 2017). 한국에서도 가계조사의 문제점이 심각하게 제기됨에 따라 통계청이 2016년 소득부터는 가계동향조사에서 가계금융복지조사(약칭 가금복)로 소득분배 지표의 근거자료를 바꾸고 조사결과를 행정자료로 보정하고 있다(김낙년, 2020).

통계청이 가금복 자료 중 소득정보를 행정자료로 보정하여 과거보다 정확성을 높이고 있으나 가금복의 샘플 자체가 대한민국 인구와 가구 전체를 편향 없이 대표하지 못한다면, 이러한 샘플 편향(sampling bias)은 현재 방식의 행정자료 보정으로 해결되지 않는다. 실제로 가금복 샘플이 최상위 소득계층과 증가하는 1인가구를 과소대표하고 있는 것으로 보이는데(구인회 외, 2020; 김낙년, 2020), 이러한 문제는 단순한 행정자료 보정으로는 그대로 남는다. 또한, 가금복은 전국단위의 소득분배 지표 측정을 위한 목적으로는 충분한 샘플 수를 가지고 있지만, 지역단위의 소득분배 지표를 측정하기에는 지역별 샘플수가 불충분하다. 이에 따라 기초자치단체는 물론 광역자치단체들마저도 지역사회보장 기본계획의 수립 및 사회보장정책의 효과 평가에 가장 중요한 지표인 소득불평등이나 빈곤율 등의 지표가 부재한 실정이다(유종성, 2020).

한국에서 전 국민의 소득 정보를 담고 있는 자료는 국민건강보험공단(약칭 건보공단)의 소득재산 DB가 유일하다. 건보공단의 행정 빅데이터인 인구사회경제 및 소득재산 DB는 전국민

1) 데이터 3법에 대해서는 대한민국 정책브리핑, 정책위키의 “데이터 3법”(<https://www.korea.kr/special/policyCurationView.do?newsId=148867915>) 참조.

의 시장소득과 공적연금소득 및 부동산 재산세 과표를 각각 2003년 및 2005년부터 패넬로 구축하고 있다. 전 국민의 개인별, 세대별 소득정보를 포괄하고 있으므로 특히 지역 차원의 소득 분배를 파악하는 데 유용하다. 아쉽게도 건보공단 빅데이터의 경우 보건의료 분야의 연구 외에 소득분배나 사회정책에 관한 연구 목적으로는 최근까지 국책연구기관에서 매우 단편적인 접근만 하였을 뿐이다(계봉오 외, 2019; 이현주 외, 2020; 전병유, 2019).

필자들은 지난 해 건보공단 빅데이터를 사용하여 서울 지역의 소득불평등과 빈곤 등을 측정하고 소득이동성 및 부동산 자산분배에 관한 기초연구를 수행하였다. 필자들은 이 연구에 참여한 경험을 토대로 소득분배지표의 측정을 비롯한 소득분배의 연구에 있어 기존의 가계조사 자료와 현재 활용가능한 행정자료인 건보공단 빅데이터의 장단점을 비교, 고찰해보고자 한다. 건보공단 자료는 전국민 패넬데이터라는 장점을 가지고 있기는 하나 시장소득 정보가 불완전하고 가처분 소득 정보가 없다는 점에서 한계를 가지고 있다. 다행스러운 점은 시장소득의 파악률이 급격하게 상승하고 있으며, 조만간 거의 완벽한 수준에 이를 것이라는 점이다. 현재도 사업소득의 경우 특고, 프리랜서 등 인적용역 사업소득까지 인별로 사업소득 정보가 제공되어 유용성이 크다.

이 논문은 우선 소득분배지표의 측정과 관련 연구에서 가계조사의 한계 및 행정자료의 활용에 관한 국제적, 국내적 논의와 동향을 간단히 살펴보고, 건보공단 자료에 대한 소개 및 동 자료를 이용한 서울시 소득분배 연구를 토대로 그 유용성과 한계를 살펴볼 것이다. 또한 현재 가금복을 기반으로 하는 소득분배 지표의 개선과 지역 단위의 소득분배 지표의 측정을 위해 서베이 자료와 행정자료의 결합을 비롯한 개선방안을 제시한다.

## 2. 가계조사 자료 대 행정자료: 국제적 논의와 동향

소득불평등도와 빈곤율 등 소득분배 지표의 측정에는 전통적으로 가계조사 자료(household survey data)가 사용되어 왔다. 그러나, 근래에 들어와 과세 자료를 비롯한 행정자료가 활용되는 경우가 많아지고 있다. 대표적으로 토마 피케티 등이 소득세 자료를 사용하여 최상위 1% 내지 최상위 10%의 소득점유율이 서구 여러 나라에서 20세기 중반에는 떨어졌다가 1970년대 이후부터 다시 상승하는 것을 보여주었다(Piketty, 2014). 한국에서도 김낙년(2018), 홍민기(2020) 등이 소득세 자료를 사용해 최상위 1% 내지 10% 소득계층의 소득집중도의 변화추이를 측정하여 1990년대 중후반 이후 소득불평등도가 급격하게 상승한 것을 보여주었다. 가계조사에서는 최상위 소득계층에 대한 조사원의 접근성의 어려움으로 인해 이들이 과소 대표될 뿐만

아니라 이들이 소득을 축소 보고하는 경향이 있어 최상위층 소득점유율을 측정하는 데에는 과세 자료가 더 정확한 정보를 제공하기 때문이다.

최근 덴마크, 노르웨이, 핀란드 등 북유럽 국가들을 중심으로 소득세와 사회보장 급여 등 행정자료를 이용하여 가구소득을 작성하는 방법이 확산되고 있다(통계청, 2017). 이 방법은 가계 조사에 비하여 응답자의 부담과 조사비용을 줄이며, 소득을 축소 보고하는 경향이 있는 고소득층의 소득을 상세히 파악 가능하고, 표본이 아니라 전수집계가 가능한 장점이 있다. 아직은 행정자료만을 이용해 소득분배 통계를 작성하는 나라는 소수에 불과하지만, 점점 더 많은 나라들이 가구조사 자료와 행정자료를 보완적으로 사용하는 추세이다.

소득분배 및 소득 이동성과 관련된 연구에도 소득세 자료 등이 활용되고 있다. 라지 체티 등(Chetty et al., 2018)은 미국의 절대적 소득이동성(absolute income mobility)의 변화 추이를 연구하기 위해 행정자료를 사용하였다. 미국 인구총조사(Census), 인구현황조사(Current Population Survey; CPS), 국세청(Internal Revenue Services)의 소득세 신고(tax return) 자료 등이 이 연구에 사용되었다. 미국의 경우 개인이 소득이 있을 경우 대부분 종합소득세 신고를 하는데, 자녀 등 부양가족에 대해 세액공제를 받기 위해 그들의 이름과 사회보장번호를 기록한다. 이들은 국세청의 협력을 얻어 사회보장번호와 부모 자녀관계 정보를 추적하였다. 그 결과 1940년대생들과 1980년대생들 자신과 부모의 30세 당시 소득을 연결하여 부모-자녀 간 소득 결합 분포를 추정할 수 있었는데 모두 1천만이 넘는 사례를 포괄하였다.

여러 나라에서 행정빅데이터 구축 및 활용에 적극적인 입장을 취하게 된 것은 전통적인 서베이 데이터만으로는 많은 한계에 부딪혔기 때문이다. 미국에서는 서베이 조사에서 소득에 대한 무응답의 증가가 행정자료 활용의 중요한 동기가 되었다. 인구현황조사(CPS)는 미국에서 소득불평등을 측정하고 노동시장을 연구하기 위해 빈번히 사용되어 온 자료인데, 이 조사의 소득 관련 항목의 무응답률이 1983년에 14.3%에서 2008년에는 33.4%로 꾸준히 증가했다. 사회보장 급여에 대한 응답률의 경우도 지속적으로 감소하였다. OASDI(Old-age, Survivors, and Disability Insurance)는 미국의 대표적인 사회보장 프로그램인데 수급자가 응답 거부를 하는 경우가 1991년 21%이었다가 2013년에는 35%으로 크게 늘어났다(Meyer et al., 2015). 일반적으로 가구유형 및 소득계층에 따라 응답률에 차이가 나기 때문에 전반적인 응답률이 저하할수록 샘플의 편향이 커진다.

행정자료는 일반적으로 프로그램 대상자 전수에 대해 작성되므로 설문조사 자료처럼 응답률 차이에 따른 편향이 없을 뿐 아니라 설문조사 자료로 접근하기 힘든 소규모 집단과 지역별 분석이 가능하다. 또, 패널 조사에서는 샘플의 지속적인 유지가 매우 어려운데, 행정자료는 장기간에 걸친 패널 데이터의 생성이 쉽게 이루어진다. 북유럽 국가들이 행정자료의 구축, 활용

에 선구적으로 나선 데에는 전국민에 대한 장기간의 패널데이터가 정책의 단기 효과를 넘어서서 중장기 효과를 평가하는 데 유용하다는 점이었다. 이들 북유럽 복지국가들은 전국민이 등록되어 있는 행정자료를 구축하고 통계적으로 활용할 수 있도록 하는데, 인구나 가족뿐만 아니라 의료, 복지, 기업, 고용, 교육, 주거에 대한 내용을 포괄한다(유종성 외, 2020).

북유럽 국가들을 비롯해 여러 나라가 증거기반 정책수립을 위해 정책실험을 설계하고 실험 결과를 평가할 때에도 행정자료를 활용한다. 가령 핀란드의 기본소득 실험 설계시 미시모의실험에 활용된 자료는 행정등록자료와 설문조사 자료를 결합한 자료였다(Kangas, 2016). 실험결과 평가에는 전국민 행정등록자료가 주로 활용되었고, 서베이 데이터와 인터뷰 데이터 등이 추가로 활용되었다(Kangas et al., 2019).

미국의 소득 관련 행정 빅데이터 구축 및 타 데이터 연계 사례 중 벤치마킹할 만한 사례들을 몇 가지 소개한다. 미국 국세청(IRS)에 의해 구축된 전국민 개인별 과세자료의 패널 데이터(SOI Databank)를 첫째로 꼽을 수 있다. 국세청 소득통계자료의 경우 과거에는 소득세 신고를 하지 않은 이들을 누락하였다. 그러나 저소득자들이 종합소득세 신고를 하지 않는 경우의 수입 및 원천징수 세금을 고용주 혹은 사업자들의 신고 자료를 통해 파악할 수 있다. 고용주나 사업자들이 이들에게 임금이나 용역비 등을 지급하기 때문이다. 이 과정을 통해 SOI Databank라는 빅데이터가 만들어지는데, 전체 미국인에 대하여 1996년부터 각 연도별로 100개가 넘는 소득, 세금 관련 변수들 및 가족, 고용주와의 링크 등을 담고 있다(Chetty et al. 2018).

다음으로 국세청에서 파악하지 못하는 이전소득까지 포함하는 포괄적인 소득 데이터셋(Comprehensive Income Dataset)을 만드는 프로젝트이다(Medalia et al., 2019). 이것은 인구조사국의 카알라 메달리아 등 5명의 연구자들의 공동 프로젝트인데, 과세자료, 가계조사 자료, 복지급여 자료 등을 개인, 가족, 가구를 단위로 통합하여 정확한 시장소득과 가처분소득을 구하는 것이다. 가구조사 자료로는 인구현황조사(CPS), 소득 및 프로그램참여 조사(Survey of Income and Program Participation), 미국 공동체 조사(American Community Survey), 소비자 비용 지출 조사(Consumer Expenditure Survey) 자료들을 활용한다. 과세자료로는 종합소득세 신고(Forms 1040) 자료, 임금과 원천징수 세금지급액(W-2), 연금 및 보험금 등 지급액(1099-R), 실업보험 급여(Unemployment Insurance), 근로장려세 자료 등을 이용한다. 더불어 연방정부와 주정부에서 지급되는 각종 급여 프로그램 중 다수의 자료를 통합하고 있다. 이 자료들에는 각각 개인 식별 키(Personal Identification Key; PIK)가 있어서 자료간의 결합이 가능한데, 이 키는 사회보장번호를 익명화한 것이다. 과세자료 및 복지급여 프로그램의 경우 99%의 개인이 PIK가 있는 반면, 서베이 자료의 경우 90%에서 97%의 개인과 가구가 PIK를 가지고 있다. 이러한 이유로 연계가 어려운 개인에 대해서는 가중치를 낮추는 접근을 활용한다

(Medalia et al., 2019).

다음으로 스탠포드대학교의 데이비드 그러스키 교수 등이 인구조사국과의 협력 하에 소득 관련 정보뿐만 아니라 교육, 고용 등 많은 변수들에 대해 장기간의 패널 데이터를 구축하는 미국인의 기회 연구(The American Opportunity Study)라는 야심 찬 프로젝트가 있다(Grusky et al., 2019). 동 연구는 1950년부터의 인구총조사 데이터와 미국공동체조사 데이터를 여러 종류의 행정, 사회 보장, 실험 데이터와 연계하여 전체 미국인에 대한 70년에 걸친 대규모 패널 데이터를 만드는 것이다. 통합된 패널데이터에는 인구총조사와 미국 공동체 조사 자료가 새롭게 업데이트된다. 또한 과세, 근로소득, 복지급여 등 타 행정 자료와 서베이, 실험 자료가 연계될 계획이다. 이 데이터가 구축되면 세대내 이동성은 물론이고 세대간 이동성 역시 소득과 교육 등 다양한 측면에서 연구할 수 있게 된다. 또한 장기간의 노동시장 과정과 인구학적 변화과정의 분석과 프로그램 및 실험의 장기 효과에 대한 연구가 가능해질 것이다.

### 3. 한국의 가계조사 자료의 개선과 한계

김낙년, 김종일(2013)은 과거 가계동향조사가 고소득층을 과소표집하여 샘플의 모집단 대표성에 문제가 있으며, 고소득층일수록 소득을 과소 보고하는 경향이 있는데 특히 최상위층 집중도가 높은 금융소득의 경우 축소 보고의 문제가 매우 심함을 밝혀냈다. 2006년의 경우 가계동향조사 샘플에는 2억원 이상 소득자가 단 한 명도 없으며, 전체적으로 국민계정상 가계소득의 81.4%만을 파악한 것으로 나타났다. 소득 종류별로 보면, 임금 및 급여의 파악률은 82% 전후였지만, 금융소득(이자+배당) 파악률은 5.7%에 불과하였다. 이처럼 고소득층의 과소표집과 낮은 소득 파악률은 불평등 지표를 상당히 저평가했을 것으로 추정되었다.

이후 통계청은 2016년 소득부터 소득분배 지표의 근거 자료로 가계동향조사 대신 고소득층의 대표성을 보다 강화한 가계금융복지조사로 대체하였다. 그리고, 면접조사의 응답오차를 줄이기 위해 국세청의 과세 자료와 보건복지부의 사회수혜금 등 소득 관련 행정자료로 조사결과를 보정하고 있다(통계청, 2017).<sup>2)</sup> 2016년 소득의 경우 가구소득 1억원 이상의 고소득가구 비율이 4.9%에서 10.1%로 늘어났고, 다시 행정자료로 보정한 결과 12.7%로 증가했다. 근로소득이 없다고 응답한 가구 중 행정자료로 보완된 가구의 비율도 18%에 달했다. 이처럼 통계청이 소득분배 지표 작성에 사용하는 자료를 개선한 후 한국의 가처분소득 지니계수는 큰 폭으로 상

2) 한편 통계청은 2016년 소득부터 OECD의 소득분배 지표 작성기준 변경(Wave 6에서 Wave 7으로)에 따라 시장소득과 처분가능소득을 계산할 때 사적이전지출을 추가로 차감하도록 변경하였다. 즉, '순' 사적이전소득(net private transfer)만을 더하였다.

승했다. 2016년도 가처분소득 지니계수가 가계동향조사로는 0.304였는데, 가금복으로 대체한 후 0.342로 올랐고, 이를 다시 행정자료로 보정한 결과 0.357(이후 다시 0.355로 업데이트됨)로 상승하였다(통계청, 2017; 2019).

이처럼 지니계수가 개선된 자료의 사용에 의해 상승함에 따라 한국은 OECD 평균보다 소득 불평등이 낮은 편에 속했다가 평균보다 훨씬 높은 국가로 바뀌었다(OECD, 2020). OECD 소득 분배 데이터베이스에 따르면, 한국의 가처분소득 지니계수가 2012년 0.307에서 2015년 0.352, 2016년 0.355로 경충 뛰었고, 상대빈곤율(중위소득 50% 기준)도 2012년 14.6%에서 2015년 17.5%, 2016년 17.6%로 크게 올라 OECD 국가 중 미국 등과 함께 빈곤율이 가장 높은 국가군에 속하게 되었다. 2013년과 2014년은 한국의 소득분배 지표가 OECD 데이터베이스에는 공란으로 나와 있다. 2015년 소득분배지표는 통계청이 원래 가계동향조사를 근거로 발표했던 것을 이후 행정자료로 보정한 가금복 자료로 대체하여 OECD에 보고한 것으로 보인다.<sup>3)</sup>

그런데, 행정자료로 보정된 가금복 자료도 여전히 한계를 보이고 있다. 김낙년(2020)에 의하면 행정자료 보정에 따라 소득과악률이 많이 개선되었지만, 금융소득 포착율은 여전히 미흡하다. 행정자료 보정 이전의 가금복 조사결과는 한국은행에서 발간하는 자금순환표 자료의 총 금융소득의 13.9%에 불과하나 행정자료 보정 결과 금융소득 포착율이 70.6%로 대폭 증가했다. 그러나, 여전히 30%에 가까운 금융소득이 누락되고 있다. 더구나, 5억원 이상 소득구간의 금융소득 포착율은 가금복 조사결과로는 1%, 행정자료 보정 이후에도 10.4%에 불과하였다. 샘플에 포함된 가구는 행정자료에 의한 보정이 유효하지만, 샘플에서 누락된 경우에는 행정자료 보정이 불가능한 한계가 있다(김낙년, 2020).

최상위 금융소득층의 과소대표 못지 않게 심각한 문제는 갈수록 증가하고 있는 1인가구의 과소대표이다. 아래 [표 1]에 의하면, 가구원수별 가구의 구성비가 인구주택총조사(센서스)와 가계금융복지조사 간에 상당히 큰 차이를 보인다. 가금복에서 1인가구는 19.3%로 센서스의 29.3%에 비해 10% 포인트 과소대표되어 있다. 4인가구는 가금복에서 25.4%로 센서스의 17.0%에 비해 8.4% 포인트 과다대표되어 있다.

이처럼 센서스에 비해 가금복이 1인가구를 과소대표하는 데에는 센서스와 가구에 대한 정의가 다른 요인과 조사원의 1인가구 접근성 제약 문제가 중첩되어 있는 것으로 보인다. 센서스에서는 “1인 또는 2인 이상이 모여서 취사, 취침 등 생계를 같이 하는 생활단위”로 정의하고 있는 반면, 가금복에서는 “1인 또는 2인 이상이 모여 주거 또는 소득과 지출 등 생계를 같이 하는 생활단위”로 정의하여 주거가 다르더라도 생계를 같이 하면 같은 가구로 취급될 수 있다(‘등’과

3) 2015년 가처분소득 지니계수의 경우 가계동향조사에 의하면 0.295였으나 가금복 조사에 따라서 0.341로 오르고(통계청, 2016), 다시 행정자료 보정에 의해 0.354로 올랐다가(통계청, 2017), 이후 다시 0.352로 업데이트되었다(통계청, 2019).

‘또는’의 차이에 유의하도록 밀줄을 함). 이에 따라 가금복에서는 직장, 학업 등을 위해 일시적으로 떨어져 사는 가족을 포함하기 때문에 센서스에 비해 1인가구의 비중이 작을 수 있다. 또한, 조사원이 1인가구를 방문할 때 집에 없는 경우가 많아 재차 삼차 방문을 해도 면접조사가 어려워 1인가구가 샘플에 과소대표되는 정도가 심한 반면, 가구원수가 많으면 방문 접근이 그만큼 쉬운 요인도 작용할 수 있다. 구인회 외(2020: 23-24)가 가금복에서 가구주와의 동거여부 정보를 이용하여 취업과 학업으로 인한 비동거 가구원을 독립된 1인가구로 분류할 경우 1인가구가 19.3%에서 22.1%로 2.8% 포인트만 증가한 것을 볼 때 가구의 정의에 따른 차이보다 1인가구의 접근성 제약이 더 크게 작용한 것으로 보인다.

1인가구에 빈곤가구가 많은 것을 고려할 때 이는 저소득층의 과소대표로 빈곤율의 과소 측정을 초래하며, 아마 지니계수도 과소 추정되었을 가능성이 있다. 가금복이 가구구성의 변화를 제대로 포착하지 못하고 있는 것은 불평등과 빈곤 측정뿐만 아니라 각종 사회정책의 효과를 분석, 평가하는 데 있어서도 한계를 노정한다.

〈표 1〉 가계금융복지조사와 인구주택총조사에 나타난 가구원수별 가구비율(2018년)

가구규모	1인	2인	3인	4인	5인 이상
가계금융복지조사	19.3	26.2	21.2	25.4	7.9
인구총조사	29.3	27.3	21.0	17.0	5.4

자료: 통계청(2018a; 2018b)

가금복이 지역 차원의 소득분배 지표 측정을 위한 자료로는 사용될 수 없다는 점도 한계이다. 소득불평등과 빈곤율과 같은 지표가 기초자치단체는 커녕 광역자치단체 수준에서도 측정되지 않고 있다. 소득불평등을 나타내는 지니계수나 중위소득의 50% 미만을 기준으로 하는 상대빈곤율이 지역사회보장기본계획의 성과지표로 사용되고 있으나, 광역 및 기초자치단체들은 이와 같은 성과지표를 사용하지 못하고 투입지표와 산출지표 위주의 중기계획에 그치고 있다. 정책평가와 환류에서 중요한 것은 투입지표(input, 예: 기초연금 예산)나 산출지표(output, 예: 기초연금 수급자 수와 수급액)보다도 결과지표(outcome, 예: 노인 빈곤율과 빈곤갭의 완화 정도)인데, 이러한 핵심 결과지표의 측정 없는 4년 단위의 지역사회보장계획의 수립이 반복되고 있다.



#### 4. 건보공단 인구사회경제 및 소득재산 DB

북유럽 국가들은 일찌기 등록센서스에 소득 정보를 포함하고 있으며, 미국의 IRS는 전 인구에 대한 인별 소득-세금 패널데이터(SOI Databank)를 구축하였고, 영국과 핀란드 정부는 전국민에 대해 월 단위로 실시간 소득정보 시스템을 구축하고 있다(유종성 외, 2020). 그러나, 한국에서 소득과 재산에 대한 전국민 빅데이터를 구축, 보유하고 있는 곳은 국세청이나 통계청이 아닌 국민건강보험공단이 유일하다.

건강보험공단은 국민건강보험법에 따라 피보험자와 의료급여 수급권자의 자격관리 및 보험료 부과를 위해 행안부와 국세청 등 여러 부처로부터 인적 정보와 함께 소득 및 재산과 관련된 자료들을 이관 받고 있다. 소득을 파악하기 위해 국세청에서 보유한 종합소득세 자료, 건강보험공단이 가지고 있는 직장가입자 연말정산 자료 및 퇴직중간정산 자료와 5대 공적연금 자료가 활용된다. 재산의 경우 행정안전부에서 제공되는 재산세 자료와 국토교통부의 자동차 자료 등이 이용된다. 구체적으로 소득 정보는 소득원천별로 근로소득, 사업소득, 이자소득, 배당소득, 기타소득, 국민연금소득, 직역연금소득이 있다. 여기에는 일용근로소득, 2천만원 이하 금융소득과 금융자산과 같이 일부 소득과 자산 정보가 누락되어 있다. 그러나 여러 가지 정보원 사이에 일치하지 않는 소득 개념 및 정보 입수 시기 등은 조화된 상태에서 자료가 제공된다. 일부 불완전성에도 불구하고 현재로서는 2003년 이후 전국민을 아우르는 유일의 소득-재산 패널 데이터이다.

건보공단 행정자료가 전국민을 포괄할 수 있는 것은 국민의 97% 이상이 건강보험에 가입되어 있고, 나머지 3%에 가까운 기초생활보장 의료급여 수급권자에 대한 정보도 의료급여 사업을 맡고 있는 건보공단이 보유하고 있기 때문이다. 건보공단의 소득재산 DB가 유용하게 사용될 수 있는 데에는 단순히 인별로 소득과 재산 정보가 구축되어 있을 뿐만 아니라 출생과 사망, 거주지, 근로자 고용 및 사업자 등록, 장애 등록 및 가족관계 정보에 이르기까지 다양한 인구사회학적 정보를 담은 인구사회경제 DB와 연결되어 있기 때문이다. 특히, 건강보험공단의 소득-재산 데이터베이스의 경우 내부적으로 부모와 자녀 세대를 결합하는 세대 간 데이터베이스를 구축하고 있다. 이를 통해 소득 및 재산이 세대 간에 어떤 방식으로 전승되는지, 그 양상이 시간에 따라 어떻게 변화하는지를 분석할 수 있다(계봉오 외 2019).

[그림 1] 건강보험공단 인구사회경제 및 소득재산DB



자료: 박종현(2021)

건보공단 빅데이터가 가진 이러한 장점에도 불구하고 한계도 있다. 먼저 건보공단은 국세청이 보유한 경상소득 관련 정보 중 2천만 원 미만의 금융소득(종합과세 대상이 아님), 종합소득 신고에서 제외된 기타소득, 그리고 일용근로소득 정보를 최근까지 제공받지 못하였다. 다행히도 2022년부터 시행될 건강보험료 부과체계 제2단계 개편을 앞두고 2019년도 귀속 소득부터 국세청으로부터 이러한 정보들을 제공받기 시작하였다. 이에 따라 건보공단 자료가 앞으로 국세청이 파악하는 경상 시장소득 정보는 완벽하게 포착하게 될 것으로 보인다. 종합소득세와 별도로 분류과세되는 퇴직소득 및 양도소득과 같은 비경상소득은 일반적으로 소득분배 지표 작성에 사용되지 않는다. 그러나, 부동산과 주식 양도차익 등 자본이득이 갈수록 중요해지는 현실에서 향후 비경상소득을 포함한 소득분배의 연구도 필요하다고 볼 수 있다. 또한, 경상 시장소득 중에서 재배업소득과 같은 비과세소득, 조세 회피를 하는 비공식부문의 소득활동, 그리고 사적 이전소득 정보는 행정자료로는 파악할 수 없는 근본적인 한계를 가진다.

다음으로 건보공단 자료는 가처분소득을 구할 수 없는 한계가 있다. 즉, 종합소득신고 혹은 연말정산 자료에 있는 비소비지출(사회보장기여금, 세금 등)에 대한 데이터, 기초연금이나 기초생활보장 급여, 고용보험 급여와 같은 공적이전 소득 등은 포함하지 못한 상태이다. 소득분배 지표를 구할 때 시장소득보다 가처분소득을 기준으로 한 지표들이 보다 더 의미가 있고 널리 쓰이며, 시장소득의 불평등 또는 빈곤지표와 가처분소득의 그것들을 비교하여 복지국가의 재분배효과를 측정한다는 점에서 볼 때 이는 중요한 한계이다.

또한, 이 논문의 초점은 아니지만 건보공단의 재산 정보는 부동산에 한정되어 있고, 금융자산이나 사업자산과 같은 정보가 부재하다. 또한 부동산 정보도 시가나 공시가격이 아니라 재산세 과세표준이어서 시가와와는 상당한 괴리가 있다.

건강보험공단의 자료 공유 서비스 페이지에 일반 연구자들이 접근하면 여기서 소개한 소득과 재산 정보가 제공되지 않고 보험료 10분위만이 제공되고 있다. 가명정보라 하더라도 건보공단이 직접 생산하지 않고 타 기관으로부터 수집한 정보에 대해서는 사회보장기관이 아니면 제공할 수 없기 때문이다.<sup>4)</sup>

## 5. 건보공단 자료에 나타난 서울시의 소득분배

필자들이 참여한 연구의 주된 목적은 건보공단의 소득재산 DB를 이용하여 서울시의 소득불평등과 빈곤의 추이를 측정해보고 서울시와 구를 비롯한 지역 차원의 소득분배 지표 측정에 행정자료를 활용하는 가능성을 모색하는 데 있었다. 아울러서 건보공단의 재산세 과세표준 자료를 이용한 부동산 자산의 분배와 소득 이동성에 대한 탐색적 연구를 계획하였다.

이 연구는 연구진이 건보공단의 인구사회경제 및 소득재산DB에 포함된 개인 및 가구 단위의 소득과 재산 관련 정보를 인터넷이 차단된 건보공단의 빅데이터 분석센터에서 가명정보로 접근하여 분석한 후 분석결과에 대한 익명성 보장 심사를 거쳐 반출하는 절차로 진행되었다. 실제로는 자료에 접근해 분석하는 기간보다 자료를 준비하고 준비된 자료를 분석할 수 있는 컴퓨터 자리 배정을 받는 데 더 많은 시간이 소요되었다.<sup>5)</sup>

본 연구에서 활용한 자료는 2006년부터 2018년까지의 서울시민(주민등록 주소 기준) 전수 자료와 전국민 1% 주민등록 세대 샘플의 패널자료이다. 건보공단이 소득 자료는 2003년부터, 재산세 자료는 2005년부터 보유하고 있으나, 건보공단 빅데이터실 관계자들에 의하면 초기 자

4) 2020년 데이터 3법 통과에 따라 개정된 개인정보보호법은 과학적 연구에는 정보주체 동의 없이 가명정보를 결합해 사용할 수 있다고 규정하고 있으므로 이 법에 따르면 건보공단이 타 기관으로부터 제공받은 자료도 연구자에게 제공할 수 있어야 할 것이다. 아직은 기존의 관행이 유지되고 있는 것으로 보인다.

5) 사회보장기관인 서울시와 건강보험공단 간에 기관 MOU의 체결이 건보공단이 일반적인 자료공유 서비스를 통해 연구자들에게 제공하는 보험료 분위 외에 보험료 산정의 기초가 되는 소득과 재산에 관한 자료를 제공하는 데 필수요건이었고, 이를 기초로 서울연구원과 가천대 불평등과사회정책연구소 간에 학술연구 MOU를 체결하였다(2020. 4. 9). 이후 연구계획에 대해 가천대학교 생명윤리위원회로부터 심의면제 확인(2020. 5. 28)을 득하고 나서 건강보험공단 자료공유서비스의 자료제공 심의위원회의 승인(2020. 6. 24)을 받았다. 이후 본 연구진이 신청한 맞춤형 연구 데이터셋이 추출된 후 원주의 건보공단 빅데이터실을 직접 방문하여(2020. 9. 7) 확인 과정을 밟은 후 서울 분석센터에 자리를 배정받아(2020. 10. 21) 본격적인 자료 검토 및 분석에 착수했다. 이후 2021년 1월 31일까지 1차년도 연구를 완료했으며, 2차년도 연구는 지난 5월 6일 가천대학교 생명윤리위원회로부터 심의면제 확인을 받은 후 같은 절차를 진행 중이다.

료는 다소 불안정하므로 2006년부터 분석하는 것이 좋겠다고 하였다. 소득은 근로소득(상용근로자의 연말정산 근로소득으로 일용근로소득은 제외됨), 사업소득(부동산임대소득 포함), 금융소득(종합과세된 이자 및 배당소득으로 분리과세된 2천만원 이하의 금융소득은 제외됨), 기타소득(종합과세됨) 등 시장소득과 공적연금 수령액(국민연금과 지역연금 포함, 기초연금은 제외)을 포함하고 있다.

건보공단 자료 중 가장 유용한 부분은 사업소득 자료로 보인다. 근로소득과 금융소득의 불완전성에 비해 사업소득 자료는 종합소득 신고가 면제되는 연말정산 사업소득(보험모집, 방문판매, 음료배달 중 간편장부 대상자)은 물론 원칙적으로 종합소득 신고를 하게 되어 있으나 실제로는 하지 않는 특고, 프리랜서 등 다수의 저소득 인적용역 사업소득자의 원천징수 사업소득까지도 포함하고 있다.<sup>6)</sup>

본 연구에서는 서울시 성인(20세 이상)의 개인소득과 가구소득 분포를 아래 <표 2>와 같이 근로소득, 사업소득, 근로소득과 사업소득을 합한 노동소득, 금융소득과 기타소득, 공적연금소득 등 소득원천별로, 그리고 시장소득(노동소득+금융소득+기타소득) 및 총소득(시장소득+공적연금소득)을 기준으로 살펴보고, 2006년부터 2018년까지의 추이와 함께 전국 평균과 비교해보았다. 서울시 소득분배에 관한 주요 발견을 아래에 간략히 소개한다.<sup>7)</sup>

<표 2> 소득 구분

소득재산DB 제공 소득변수		본 연구에서 추가로 분석하는 소득변수	
근로소득		노동소득	
사업소득			
이자소득		금융소득	시장소득
배당소득			총소득
기타소득			
국민연금소득		공적연금소득	
지역연금소득			

6) 원천징수 대상 사업소득에는 의료보건용역(수의사의 용역 포함)과 저술가·작곡가 등이 직업상 제공하는 인적용역이 포함된다. 의료보건용역 외에 원천징수 대상인 인적용역 사업소득자로는 물적 시설 없이 근로자를 고용하지 않고 독립적으로 용역을 제공하고 대가를 받는 저술가, 화가, 작곡가, 배우, 모델, 가수, 성악가, 연예보조, 자문 고문, 학원강사, 직업운동가, 보험설계, 음료배달, 방문외판, 다단계판매, 기타 모집수당, 간병인, 대리운전, 행사도우미, 물품배달, 기타 자영업 등이 있다. 인적용역 사업자의 경우에도 원칙적으로 종합소득세 신고를 해야 하지만, 간편장부 대상자(해당 과세기간 신규 사업 개시 또는 해당 사업소득만 있는 자로서 직전 과세기간 수입금액 7,500만원 미만)인 보험모집인, 방문판매원, 음료배달판매원에 대하여는 원천징수의무자가 해당소득에 대한 연말정산을 함으로써 납세의무를 종결한다. 또 상당수의 저소득 인적용역 사업자는 종합소득세 신고를 하지 않는 것으로 보인다.

7) 연구보고서(김승연 외, 2020) 전문은 가천대 불평등과 사회정책연구소 홈페이지([http://www.kirl.re.kr/bbs/board.php?tbl=ga\\_bbs32](http://www.kirl.re.kr/bbs/board.php?tbl=ga_bbs32))에서 볼 수 있다.

## 1) 서울시의 개인소득 분배

건보공단 자료에 의하면 2018년 서울시 거주 20세 이상 성인 중 67.6%가 소득자로서 2006년의 소득자 비율(47.9%)보다 20% 포인트 가까이 증가한 것으로 나타난다. 이는 이 기간 동안에 소득활동을 한 성인의 비율이 이처럼 큰 폭으로 증가했다기 보다는 국세청의 소득파악 능력의 향상에 기인한 것이라고 보인다.

아래 <표 3>를 보면, 건보공단 자료상 서울시의 근로소득자는 2006년 성인 인구의 32.4%에서 2018년 44.5%로 증가했고, 사업소득자는 같은 기간에 성인 인구의 14.9%에서 26.2%로 거의 두 배 가까이 증가하였다. 이를 통계청 경제활동인구조사의 종사상 지위별 취업자 통계와 비교해보면 건보공단 자료가 동 기간 중 서울시 임금근로자의 71%에서 92.2%로 파락률을 크게 높인 것으로 나타나며, 건보공단 자료상 서울시 사업소득자 수는 통계청의 서울시 자영업자 수의 105.6%에서 248.6%로 증가한 것으로 나타난다.<sup>8)</sup> 건보공단 자료에서 사업소득과 근로소득이 다 있는 사람들을 빼고 사업소득만 있는 순수 사업소득자 수는 통계청 자영업자 수 대비 88.7%에서 171%로 증가하였다. 건보공단 자료상 근로소득이나 사업소득이 있는 노동소득자는 2006년에서 2018년 사이에 서울시 성인인구의 44.9%에서 62.5%로 증가했는데, 이는 통계청의 서울시 취업자 수 대비 72%에서 103.9%로 증가한 것이다.

<표 3> 2006년과 2018년 서울시 성인 인구중 건보자료와 통계청 취업자 통계의 비교

연도	건보/통계청구분	건보 인원수	비율	통계청(천명)	건보/통계청
2006	근로소득자/임금근로자	2,605,862	32.4%	3,670	71.0%
	사업소득자/비임금근로자	1,198,337	14.9%	1,135	105.6%
	순수사업소득자/비임금근로자	1,006,183	12.5%	1,135	88.7%
	노동소득자/취업자	3,612,045	44.9%	5,015	72.0%
	성인 인구	8,050,771			
2018	근로소득자/임금근로자	3,756,937	44.5%	4,075	92.2%
	사업소득자/비임금근로자	2,210,320	26.2%	889	248.6%
	순수사업소득자/비임금근로자	1,520,094	18.0%	889	171.0%
	노동소득자/취업자	5,277,031	62.5%	5,080	103.9%
	성인 인구	8,445,465			

2006년부터 2018년 사이에 매년 서울시 사업소득자수가 증가했지만, 특히 큰 폭의 증가는 2006년(성인의 14.9%)에서 2008년(19.3%) 사이, 그리고 2013년(20.7%)에서 2018년(26.2%) 사

8) 전국적인 차원에서는 건보의 사업소득자 수가 통계청 경찰조사의 자영업자 수의 80.9%(2006년)에서 175.5%(2018년)으로 증가한 것으로 나타난다.

이에 나타났다. 2006년부터 2008년까지의 급속한 증가는 국세청의 소득 파악노력으로 연말정산 및 원천징수 사업소득자가 크게 늘어난 때문으로 보인다. 국세청은 2006년 귀속소득부터 사업소득 연말정산 신고현황을, 2008년 귀속소득부터 거주자의 사업소득 원천징수 신고현황을 국세통계연보에 포함시켜 발표하기 시작했다. 2013년 이후는 특고, 프리랜서 등 저소득 불안정 취업자의 지속적인 증가 및 국세청의 강화된 과세행정이 주된 원인인 것으로 추측된다. 이는 인적용역 사업소득자 중 국세청 업종 코드로 분류하기 어려운 ‘기타 자영업자’가 지난 수년간 급증한 데서 알 수 있다. 국세청이 장혜영 의원실에 제출한 자료에 따르면, 원천징수 사업소득자 중 ‘기타 자영업자’가 2014년 101.9만명에서 2018년 264.7만명으로 4년 동안 162.8만명 이상 급격히 증가했고, 2019년에는 315.2만명으로 2014년에 비해 5년간 3배 넘게 급증했다(박해진, 2020).

2018년에도 건보공단 자료상 근로소득자 수가 통계청 경제활동인구조사의 임금근로자 통계보다 적게 나온 것은 일용근로소득이 누락된 때문으로 이해할 수 있다. 건보공단은 연말정산을 하는 상용근로자만 직장 가입자로 취급하고, 일용근로자는 지역가입자로 분류되고 있는데, 이들의 건보료 산정을 위해 사업소득과 금융소득, 기타소득, 공적 연금은 물론 부동산 재산세 자료까지 사용하면서 정작 일용근로소득은 간과해온 것으로 보인다. 건보공단이 소득 중심의 건보료 부과체계 개편 계획에 따라 2019년 소득부터 일용근로소득 정보를 국세청으로부터 제공받기 시작했으므로 이 문제는 금명간 개선될 것으로 생각된다.

건보공단의 사업소득자와 통계청의 자영업자 사이에 큰 차이가 나타난 것에는 분류 기준이 다르기 때문으로 보인다. 건보공단의 사업소득자는 물적 시설 없이 인적 용역을 제공하는 특고, 프리랜서 등 인적 용역 사업소득자를 포괄하고 있다. 국세청이 저소득 불안정 노동계층에 속하는 인적용역 제공자들로부터 연말정산 사업소득과 원천징수 사업소득의 징수를 늘려왔지만, 이들 중 다수가 통계청의 경제활동인구 조사에서는 자영업자가 아닌 임금근로자나 비경제활동인구로 분류되기 때문으로 보인다.

아래 <표 4>을 보면 서울시 성인 인구 중 2006년에서 2018년 사이에 순수사업소득자가 12.5%에서 18%로 증가했고, 근로소득과 사업소득을 동시에 올린 경우가 2.4%에서 8.2%로 급증했다. 특히 후자의 경우는 대부분 종합소득세 신고를 하지 않은 연말정산 또는 원천징수의 인적용역 사업자로 추정되며, 이들은 대부분 임금근로자로 분류되었을 것으로 생각된다. 또한 특고, 프리랜서 중 상당수는 통계청의 조사에 답할 때 스스로를 자영업자로 인식하지 않고 임금근로자로 답하여 근로소득 없이 사업소득만 있는 경우에도 임금근로자로 분류되는 경우가 늘어나고 있는 것으로 추측된다.

〈표 4〉 ‘순수 사업소득자’와 ‘사업+근로소득자’의 증가 추이(서울, 2006, 2018년)

(단위: %)

2006년	근로소득자	사업소득자	근로+사업	노동소득자
20대	39.0	4.8	1.3	45.1
30대	41.3	11.9	3.0	56.2
40대	27.2	17.5	3.0	47.6
50대	22.5	18.0	2.9	43.5
60대	15.2	14.2	2.2	31.7
70세 이상	3.4	7.6	0.8	11.8
계	30.0	12.5	2.4	44.9
2018년	근로소득자	사업소득자	근로+사업	노동소득자
20대	40.5	14.4	9.4	64.2
30대	53.7	14.9	9.9	78.5
40대	43.2	20.0	10.4	73.5
50대	33.9	23.6	8.7	66.2
60대	23.7	21.3	5.7	50.6
70세 이상	8.3	12.0	1.9	22.2
계	36.3	18.0	8.2	62.5

이상을 통해 알 수 있는 것은 건보공단 자료가 파악하는 노동소득자 수는 통계청 경제활동 인구 조사가 파악하고 있는 취업자 수보다 훨씬 더 크다는 것이다. 〈표 3〉의 2018년 서울시 근로소득자 357.7만명에 각각 서울시의 순수 일용근로소득자(상용근로소득이나 사업소득 없이 일용근로소득만 있는 경우) 추정 인원 80만명을 더하면,<sup>9)</sup> 근로소득자 수는 서울시의 성인 인구 중 54%에 해당하는 456만명 내외에 달해 통계청의 임금근로자 407.5만명에 비해 12% 가량 더 많은 것으로 나타난다. 이는 일용근로자 중 상당수가 경제활동인구 조사기간 중에는 일을 하지 않아 비경제활동인구로 구분된 때문으로 추정된다. 또한 노동소득자 527.7만명에 80만명을 더 하면 서울시의 성인 인구 중 72%에 해당하는 608만명 내외가 근로소득 또는 사업소득을 올린 것으로 보인다. 통계청의 취업자 통계에 잡힌 508만명(무급가족종사자 11만6천명 포함)보다 100만명이 더 많은 수치이다. 취업자 통계에서 빠진 100만명 중 대다수는 비경제활동인구로 분류되고, 일부 실업자로 잡힌 경우도 있을 것이다.

다음으로 건보공단 자료는 지난 10여년간 노동소득자를 포함한 소득자수의 증가뿐만 아니라 소득 파악률도 상당히 큰 폭으로 증가하였다. 건보공단 자료에 의하면 서울시 인구 1인당 평

9) 2019년 국세통계연보(2018년 귀속 소득에 관한) 4-4-11에 의하면 상용근로소득이나 사업소득이 있는 경우를 포함한 전체 일용근로소득자 777만명 중 17.5%에 달하는 136만명이 서울시 거주자이며, 국세청이 윤후덕 의원실에 제공한 자료에 의하면 전체 일용근로소득자 중 454만명이 상용근로소득이나 사업소득이 없는 순수일용근로자이다. 따라서 서울의 순수 일용근로소득자 수를 454만명의 17.5%인 80만명 가량으로 추정한다.

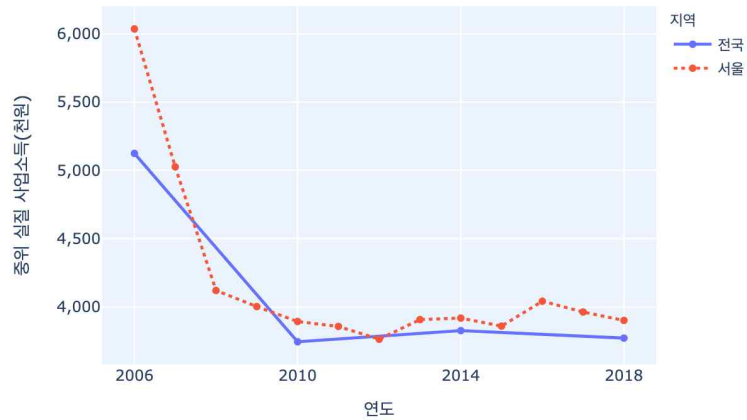
균 총소득은 명목소득으로 2006년 957만원에서 2018년 2,058만원으로 증가했고, 같은 기간 실질소득(기준년도 2015년)으로는 1,201만원에서 1,968만원으로 증가했다. 이 기간 중 통계청의 지역소득 증가 추이를 보면 서울시의 1인당 지역총소득은 77.1% 증가, 1인당 개인소득(처분가능소득)은 66.5% 증가하였는데, 건보 자료상 서울시의 1인당 소득은 115.1% 증가하였다. 전국적으로 건보공단의 소득 파악률이 국민계정의 가계본원소득에 비추어 2006년 66.6%에서 2010년 73.6%, 2014년 83.7%, 그리고 2018년에는 91.3%로 증가했는데, 서울의 소득 파악률은 전국보다 다소 높을 것으로 추측된다.<sup>10)</sup> 앞으로 일용근로소득과 분리과세된 금융소득과 기타 소득 등의 자료가 포함됨에 따라 건보공단의 시장소득 파악률은 비과세 소득과 비공식부문 및 사적 이전소득을 제외하면 거의 완벽해지리라 기대된다.

흥미로운 것은 2006년부터 2018년 사이에 서울시 성인의 실질 근로소득과 실질 총소득은 상당히 증가했지만, 실질 사업소득은 감소했다는 것이다. 2015년 소비자물가를 기준으로 한 실질소득 기준으로 평균 사업소득은 2006년 1,797만원에서 2018년 1,551만원으로, 중위소득은 같은 기간에 604만원에서 390만원으로 떨어졌다. 이 기간 중 평균 사업소득이 명목소득 기준으로 약 1.4배 올랐지만(1,432만원에서 1,622만원으로) 중위 사업소득은 명목 기준으로도 감소했다(481만원에서 408만원으로). 이 기간 중 사업소득의 지니계수는 0.72에서 0.77로 상승했다. 이처럼 평균과 중위 실질 사업소득이 모두 하락한 것은 저소득 인적용역 사업소득자들의 소득이 잡히지 않다가 이들의 사업소득이 국세청에 포착됨에 기인한 것으로 보인다. 국세청이 연말정산과 원천징수 사업소득 통계를 새로이 발표하고 사업소득자 수가 급증한 2006년에서 2008년 사이에 평균과 중위 사업소득 모두가 명목기준으로도 상당폭 하락하고 실질 기준으로는 대폭 하락한 것으로 나타난 것이 이러한 해석을 뒷받침해준다. 또한 2010년대 내내 사업소득자의 지속적 증가와 함께 실질 사업소득이 정체한 것은 '기타 자영업자' 등 전통적인 직업분류로 구분하기 어려운 새로운 형태의 특고와 프리랜서 등 저소득 불안정 근로자성 자영업자들의 지속적인 증가 때문으로 추정된다. 이러한 현상은 서울뿐 아니라 전국적으로 비슷하게 나타난 것으로 보인다([그림 2], [그림 3] 참고).

10) 통계청의 지역별 개인소득 통계는 처분가능소득을 나타내는 것이므로 건보자료의 개인소득(시장소득)과 직접 비교하기에 적합하지는 않다. 국민계정상 가계본원소득은 지역별 통계가 없어 전국 단위에서 건보자료의 소득 파악률을 구하였다. 서울은 비과세 소득인 재배업 소득의 비중이 작으므로 서울의 소득 파악률이 전국 평균보다 다소 높을 것으로 추측할 수 있다.



[그림 2] 성인 사업소득자의 중위 사업소득 시계열 추이(서울과 전국, 실질소득)



[그림 3] 성인 사업소득자의 평균 사업소득 시계열 추이(서울과 전국, 실질소득)



사업소득과 달리 근로소득과 총소득은 2006년과 2018년 사이에 명목은 물론 실질 기준으로 상당히 증가한 것으로 나타나며, 소득불평등을 나타내는 지니계수도 약간 하락한 것으로 나타난다. 이 기간 중 서울시 성인 근로소득자의 수가 증가했을뿐 아니라 근로소득자의 평균 및 중위소득이 실질 기준으로도 상승했다. 근로소득자의 명목 평균 근로소득이 2,906만원(실질 3,647만원)에서 4,145만원(실질 3,963만원)으로, 명목 중위 근로소득이 1,990만원(실질 2,497만원)에서 2,769만원(실질 2,648만원)으로 증가했다. 또한, 이 기간 중 총소득자의 수가 증가했을 뿐만 아니라 이들의 평균 및 중위 총소득도 실질 기준으로 증가했다. 명목 평균소득은 2,581만원(실질 3,239만원)에서 3,612만원(실질 3,454만원)으로, 명목 중위소득은 1,421만원

(실질 1,783만원)에서 2,125만원(실질 2,032만원)으로 각각 상승했다. 또한 서울시 성인 근로소득자의 근로소득 지니계수는 0.50에서 0.49로 약간 줄었고, 성인 총소득자의 총소득 지니계수도 0.60에서 0.59로 소폭 하락하였다. 그러나, 이러한 실질소득의 증가와 지니계수의 하락 중 어느 정도가 실제를 반영한 것이고 어느 정도가 국세청의 소득 파악 증가에 따른 것인지는 알 수가 없다. 따라서, 현재까지의 건보공단 자료로는 종단분석에 한계가 있다.

2018년 성인 노동소득자(근로소득 또는 사업소득이 있는 자)의 노동소득, 유소득자의 총소득, 그리고 무소득자를 포함한 성인 1인당 평균 및 중위소득을 비교하면 아래 <표 5>와 같다. 서울시 전국에 비해 소득자 비율도 더 높고 평균 및 중위소득도 더 높음을 보여준다. 성인 노동소득자 및 유소득자 안에서 평균과 중위소득 간에 큰 차이가 있는데, 평균 값보다는 중위 값이 전형적인 노동소득자나 유소득자의 소득수준을 파악하는 데 더 유용하다. 다만, 성인 1인당 소득을 구할 때에는 무소득자 때문에 중위값이 매우 낮게 나오게 되므로 평균값을 보는 것이 보다 유용하다고 할 것이다.

<표 5> 2018년 노동소득자, 유소득자 및 성인 1인당 평균 및 중위소득

(단위: 만원/연)

		소득자비율	평균	중위
노동소득자 (근로+사업소득)	전국	59.0%	3,159	2,129
	서울	62.5%	3,630	2,196
유소득자 (총소득)	전국	65.0%	3,042	2,004
	서울	67.6%	3,612	2,125
성인1인당소득	전국		1,864	367
	서울		2,268	553

\*주: 노동소득과 총소득은 소득자 가운데서의 평균과 중위값을 나타내며, 성인 1인당 소득은 무소득자를 포함한 성인 전체를 대상으로 평균과 중위값을 나타낸다.

앞에서 통계청 경제활동인구조사(경찰)의 자영업자 수와 건보 자료의 사업소득자 수 사이에 큰 차이가 있지만, 경찰의 임금근로자 수와 건보의 근로소득자 수는 차이가 크지 않음을 보여준 바 있다. 그러나, 경찰 임금근로자의 근로소득 분포와 건보 근로소득자의 근로소득 분포에는 매우 큰 차이가 나타난다. <표 6>을 보면 건보의 평균 근로소득은 3,611만원, 경찰의 평균 근로소득은 3,090만원으로 경찰 소득이 건보 소득의 85.6%에 그친다. 그런데, 1분위부터 5분위까지는 경찰 소득이 건보 소득보다 높으며, 6분위 이상은 반대로 나온다. 저소득층으로 갈수록 경찰의 근로소득이 건보의 근로소득보다 높으며(1분위는 경찰이 건보의 184.2%), 고소득층으로 갈수록 낮게 나온다(10분위는 경찰이 건보의 66.1%). 특히 최상위 1%의 평균 근로소득은 건보 2억 5,782만원, 경찰 1억 3,197만원으로 경찰 소득은 건보의 51.2%에 불과하다. 이는 경

제활동인구조사가 고소득 근로자와 저소득 근로자를 제대로 포착하지 못하고 있을 가능성이 큼을 시사하며, 고소득층과 저소득층의 파악에 행정자료의 유용성을 보여주기도 한다.<sup>11)</sup>

〈표 6〉 근로소득 10분위 및 최상위 1%의 평균 근로소득(2018년): 건보와 경찰자료 비교

(단위: 천원)

분위	건보	경찰	경찰/건보
1	3,367	6,201	184.2%
2	9,405	14,174	150.7%
3	15,219	18,985	124.7%
4	19,739	21,889	110.9%
5	23,871	24,231	101.5%
6	29,260	28,723	98.2%
7	36,397	32,820	90.2%
8	46,698	38,246	81.9%
9	63,230	48,680	77.0%
10	113,894	75,247	66.1%
top 1%	257,819	131,967	51.2%
Total	36,108	30,904	85.6%

자료: 통계청 경제활동인구 부가조사(2018년 8월)와 건보공단 소득재산DB

건보공단 자료 상으로 20세 이상 성인인구 중 최상위 소득자 1%의 소득점유율(건보자료로 파악한 소득 총액 기준)의 추이를 보니 2006년부터 2018년 사이에 서울은 21.3%에서 17.9%로, 전국은 17.0%에서 13.8%로 낮아졌다. 최상위 소득자 10%의 소득점유율도 서울은 62.6%에서 53.0%로, 전국은 61.6%에서 49.6%로 큰 폭으로 떨어졌다. 그러나, 김낙년(2018)과 홍민기(2020)에 의하면, 위 기간 동안에 상위 1%와 상위 10%의 총소득 점유율이 증가한 것으로 나와 건보자료의 결과와 다르게 나타난 바 있다. 전국을 대상으로 상위 1% 소득점유율의 경우 김낙년(2018)은 2006년 10.6%에서 2016년 12.1%로 상승, 홍민기(2020)는 2006년 12.0%에서 2018년 14.7%로 상승한 것으로 추정하였다.

이는 방법론 상의 차이와 아울러 건보공단의 자료가 기초하고 있는 국세청의 소득파악률 개선에 따른 효과가 크게 반영된 것으로 보여, 이 기간동안 실제로 소득분배가 개선되었다고 판단하기 어렵다. 김낙년(2018)과 홍민기(2015; 2020)는 모두 성인(20세 이상) 인구를 기준으로

11) 미국에서는 서베이 조사에서 저소득 계층이 실제보다 소득을 높여서 응답하고, 고소득 계층의 경우 소득을 낮춰서 대답하는 경향이 나타났다고 한다 (Kim & Tamborini, 2014). 경찰자료와 건보자료간의 차이가 저소득 및 고소득 근로자의 과소표집, 소득 확대 및 축소 보고 등에 기인하는 정도에 대해 추가적인 연구가 필요하며, 다른 가계조사 자료와의 비교 연구도 요청된다. 한편, 행정자료가 저소득층의 소득을 과소 파악하고 있을 수도 있다.

하여 상위 소득층의 소득을 국세청 자료를 기초로 하되 국세통계에서 소득구간에 따른 분포를 보여주지 않는 일용근로소득, 2천만원 이하 금융소득 등을 일정한 가정에 따라 반영하였다. 총 소득의 분모에 대해서는 둘 다 국민계정의 개인소득 통계를 사용하여 추정하였다. 둘 사이에 추정치에 차이는 있지만(홍민기가 다소 과대 추정한 것으로 보임), 둘 다 상위층 소득집중도의 강화추세를 보여주고 있다.<sup>12)</sup> 따라서 건보공단 자료를 근거로 2006년부터 2018년 사이에 개인소득 분배가 개선되었다고 결론을 내릴 수는 없으며, 국세청의 소득파악 능력 개선의 효과가 더 크게 나타난 것으로 생각된다.

그러나, 아래 <표 7>이 보여주는 것과 같이 건보공단 자료의 소득파악률이 높아짐에 따라 상위 1%와 상위 10% 소득점유율이 김낙년(2018) 및 홍민기(2020)의 추정과의 차이가 줄어들고 있다. 2006년의 경우 건보공단 자료에 근거한 상위 1%의 소득점유율 17.0%는 김낙년의 10.6%, 홍민기의 12.0%보다 각각 6.4%, 5.0% 높았는데, 2018년의 경우 건보공단 자료상 상위 1%의 소득점유율 13.8%는 김낙년의 12.1%, 홍민기의 14.7%에 비해 각각 +1.7%, -0.9%로 차이가 감소했다. 즉, 최근 자료에서는 최상위 소득자 1%의 소득점유율이 어느 정도 유사하게 추정되는 것이다. 상위 10% 점유율도 2006년에는 61.6%로 김낙년의 41.8%, 홍민기의 46.0%에 비해 19.8%, 15.6% 차이가 났으나, 2018년에는 49.6%로 김낙년의 43.2%, 홍민기의 48.9%에 비해 6.4%, 0.7%로 차이가 감소했다.

<표 7> 상위 1% 및 상위 10%의 소득점유율 추정 비교

연도	상위 1% 소득점유율 (%)			상위 10% 소득점유율 (%)		
	건보자료	김낙년	홍민기	건보자료	김낙년	홍민기
2006	15.6	10.6	12.0	63.0	41.8	46.0
2010	17.0	11.7	12.4	61.6	43.0	45.4
2014	16.1	11.8	13.5	57.6	42.7	47.4
2018*	14.8	12.1	14.7	53.8	43.2	48.9

\* 김낙년은 2018년이 아닌 2016년의 통계임.

자료: 김낙년(2018), 홍민기(2020)

건보공단 자료가 국세통계 자료에 비해 유용성을 보이는 분야로 공적연금 소득을 들 수 있다. 국세통계 자료는 국민연금과 직역연금을 구분하지 않는 반면 건보공단 자료는 이를 구분해 총연금 수령액을 보여준다. 이에 따라 우리 나라 노인이 직역연금 수급자, 국민연금 수급자, 공적연금(기초연금 제외) 비수급자의 세 계급으로 뚜렷하게 구분됨을 알 수 있다.

12) 김낙년(2018)은 홍민기(2015)가 이자소득에 대해 분모에서는 이자 지급을 제하여 과소 설정을 하였고, 분자와 분모의 소득이 서로 조응하지 않게 되었으며, 이에 따라 상위 1% 소득점유율을 과대 추정했음을 지적하고 있다.

2018년 서울 거주 62세(현재 국민연금 노령연금 수급 개시 연령) 이상 연령층의 4.6%가 공무원, 군인, 사학연금 등 직역연금 수급자로서 이들의 평균 연금수령액은 연 3,295만원(월 275만원), 중위값은 연 3,193만원(월 267만원)이다. 국민연금 수급자는 62세 이상 연령층의 36.6%로서 평균 연금수령액은 연 499만원(월 42만원), 중위값은 연 374만원(월 31만원)에 불과하다. 공적연금 비수급자 중 다수와 국민연금 수급자 일부가 기초연금을 수령할 것으로 추측되나, 건보공단 자료에는 기초연금 정보가 없다.

직역연금과 국민연금 수급자 중에는 각각 40.9% 또는 53.2%가 시장소득을 올리고 있어 이들의 총소득은 직역연금 수급자는 평균 4,749만원(중위소득은 3,853만원), 국민연금 수급자는 평균 2,346만원(중위소득은 761만원)에 달하며, 비수급자 중에는 23.6%만이 시장소득을 올리고 있고 시장소득의 평균과 중위값도 낮아 이들의 총소득 평균이 646만원(중위소득은 0)에 불과하다.

〈표 8〉 2018년 서울 62세 이상 노인의 공적연금 수급 구분에 따른 공적연금소득, 시장소득과 총소득

(단위: %, 천원)

수급자 구분	수급율	평균	중위	시장소득자			총소득	
				비율	평균	중위	평균	중위
직역연금	4.6	32,952	31,933	40.9	34,957	18,885	47,488	38,526
국민연금	36.6	4,992	3,743	53.2	34,444	15,552	23,463	7,606
비수급자	59.1	0	0	23.6	27,390	10,339	6,462	0

\* 공적연금 소득과 시장소득의 평균과 중위는 소득자 중에서의 값이며, 총소득은 62세 이상 노인 전체 가운데서 평균과 중위값임.

다음으로 아래 〈표 9〉는 2018년 서울시의 62세 이상자들을 시장소득 기준으로 무소득자(64.8%)와 유소득자(35.2%)로 나누고 유소득자를 다시 4분위(각 분위당 62세 이상 인구의 8.8% 대표)로 나누어 이들의 시장소득과 공적연금소득, 그리고 총소득의 평균과 중위값을 비교해 보여준다. 시장소득이 없는 64.8%의 노인 중 공적연금 소득자가 30.4%로서 시장소득이 있는 35.2%의 노인 중 공적연금 소득자의 비율인 60.4%의 절반에 불과하여 공적연금이 시장소득의 불평등을 완화하기보다는 확대 강화하는 것을 알 수 있다. 무시장소득자의 총소득 평균이 연 244만원에 불과하며 중위소득은 0이어서 비록 기초연금과 기초생활보장급여 등이 누락되긴 했지만, 다수의 노인들이 빈곤한 상태에 있음을 알 수 있다. 유시장소득자 중에서도 4분위중 1분위와 2분위 노인들의 총소득 평균이 각각 연 512만원과 1,259만원에 불과하며, 3분위와 4분위에 속한 17.6%의 노인을 제외하면 노인들의 개인소득이 대략 연 2천만원 미만 수준에 불과함을 보여준다. 또한, 하위 78%의 노인들이 시장소득 연 840만원(월 70만원), 공적연금을 더한 총소득 연 1,099만원(월 91.6만원) 이하이며, 하위 69.2%의 노인들이 시장소득 연 160만원(월

13.3만원), 총소득 연 339만원(월 28.3만원) 이하이다. 기초연금 수급 대상을 하위 70%에서 적어도 80% 이상으로 확대할 필요가 있음을 보여준다.

〈표 9〉 2018년 서울 62세 이상 노인의 시장소득 분위에 따른 시장소득, 공적연금소득과 총소득

(단위: 천원)

시장소득분위	비중	시장소득 평균	시장소득 중위	공적연금 소득자	연금소득 평균	연금소득 중위	총소득 평균	총소득 중위
무시장소득자	64.8%	0	0	30.4%	8,029	3,552	2,439	0
유시장소득자	35.2%	31,683	13,540	60.4%	8,232	4,716	36,653	17,400
1분위	8.8%	1,664	1,597	51.7%	6,684	3,884	5,121	3,390
2분위	8.8%	8,547	8,400	55.8%	7,244	4,067	12,586	10,993
3분위	8.8%	18,798	18,910	65.8%	7,727	4,523	23,885	22,071
4분위	8.8%	97,741	43,350	68.2%	10,700	7,178	105,041	52,324

\* 시장소득과 공적연금소득의 평균과 중위는 소득자 중에서의 값이며, 총소득은 62세 이상 노인 전체 가운데서 평균과 중위값임.

## 2) 서울시의 가구소득 분배

다음으로 본 연구는 가구소득의 분배에 대한 분석을 시도하였다. 건보공단 자료는 건강보험 자격을 기준으로 한 세대(직장가입자와 피부양자, 지역 세대주와 세대원, 의료급여 세대주와 세대원)와 함께 주민등록세대 정보를 제공하고 있으나, 가구 정보를 제공하지 않고 있다. 주민등록 세대는 “주거 및 생계를 같이 하는 집단”(주민등록 사무편람)으로 정의되어 인구주택총조사의 가구(취사, 취침 등 생계를 같이 하는 생활단위)와 큰 차이가 없으나, 다양한 이유로 동일 가구 내에서 세대 분리를 한 경우가 많다. 또한 가계금융복지조사의 경우 직장, 학업 등을 위해 일시적으로 떨어져 사는 가족을 별도의 가구로 분리하지 않는데, 이들은 대부분 주민등록을 별도로 하기 때문에 주민등록 상으로는 별도의 세대로 구분된다.

건보공단 자료 상 주민등록 세대를 기준으로 1인 세대의 비율이 2018년 서울시는 37.6%, 전국은 35.9%로서 가계금융복지조사(가금복)의 1인 가구 비율 19.3%보다 거의 두 배이고, 인구주택총조사(센서스)의 서울 32.0%, 전국 29.3%에 비해서도 매우 높다. 주민등록 세대와 가구 간에 차이가 매우 큼을 보여준다. 따라서 본 연구는 주민등록 세대를 기준으로 한 분석과 별도로 주민등록 1인세대 중 건강보험 세대원인 경우는 세대주의 가구에 편입시켜 가금복의 가구와 비슷하게 재편하였다.<sup>13)</sup> 다만, 이러한 가구 재편은 전수 자료를 접근한 서울에 대해서만 가

13) 센서스의 가구 개념에 근접시키고자 한다면, 같은 주소지에 살고 있는 가구원으로서 세대분리를 한 가족이나 친족의 경우에는 주소지 정보를 이용해 가구 통합을 할 수도 있겠지만, 본 연구진이 접근한 자료는 주소 정보가 동 단위까지만 제공되어 불가능했다. 향후 건보공단 연구진의 협력을 얻어 센서스의 가구 개념에 따

능했고, 전국 샘플에서는 건보 세대원의 세대주를 찾을 수 없어 불가능했다. 또한, 건보세대원이 서울에 거주하는 경우에도 세대주가 지방에 거주하는 경우에는 제외할 수밖에 없었다. 이처럼 2018년도 서울시의 주민등록세대를 가구로 재편한 결과 가구원수별 가구유형의 비율이 아래 <표 10>과 같이 변화되었다. 주민등록세대 기준으로는 1인세대가 37.6%에 달했으나, 재편된 가구 기준으로는 1인가구가 32.2%로 줄어들었다.

<표 10> 2018년 서울시의 가구원수별 가구유형 비율

(단위: %)		
가구수 비율	주민등록세대 기준	재편된 가구 기준
1인	37.6	32.2
2인	20.4	21.9
3인	18.6	19.8
4인	17.7	19.0
5인 이상	5.7	7.1
합계	100.0	100.0

재편된 가구 기준으로 서울시의 1인가구는 2006년 27.3%에서 32.2%로 증가했고, 인구 비율로는 9.9%에서 12.9%로 증가했다. 2인 이상 가구는 각각 1.2% 내지 1.5% 포인트만큼 증가했다.<sup>14)</sup> 이들의 연령별 분포를 보면, 50대 이상의 중고령층에서 1인가구가 크게 증가한 것으로 나타난다. 20대와 30대에서는 1인가구의 증가가 없었고, 40대에서도 미미한 증가에 그쳤으나, 50대 1인가구는 2006년 12만 5천 에서 2018년 22만 9천으로 급증했고, 60대 1인가구는 7만 5천에서 16만 2천으로, 70세 이상 1인가구는 5만 9천에서 12만으로 두 배 이상 늘었다. 1인가구 중 소득자 비율은 남녀 모두 73%로 비슷하나, 소득자의 평균소득은 남성은 3,602만원, 여성은 2,917만원으로 차이가 났다. 1인가구 중 27%가 시장소득도 공적연금 소득도 없다는 것은 1인가구의 빈곤이 심각함을 시사한다.

<표 11> 가구유형별 소득원천별 가구소득 평균(2018년)

(단위: 천원)							
	근로 소득	사업 소득	금융 소득	기타소득	공적연금 소득	가구소득 평균	가구소득 중위
가구원수							
1인	19,149	3,795	555	79	386	23,963	13,101
2인	25,454	7,433	1,814	165	3,023	37,889	20,720
3인	48,268	10,626	1,912	272	2,411	63,489	43,000

른 가구 재편도 시도해볼 여지가 있겠다.

14) 이는 구인회 외(2020)가 가금복에서 가구주와의 동거여부 정보를 이용하여 취업과 학업으로 인한 비동거 가구원을 독립된 1인가구로 분류할 경우 1인가구가 19.3%에서 22.1%로 늘고 2인 이상 가구들은 다소 감소한 것과 일맥상통한다.

4인	68,294	14,798	2,607	397	1,518	87,614	62,573
5인 이상	72,895	18,210	3,683	449	2,548	97,786	68,877
1인가구중							
20세 미만	2,178	913	0	1	0	3,093	0
20~29세	23,736	1,544	106	53	0	25,439	23,803
30~39세	30,381	3,395	269	84	0	34,130	29,555
40~49세	22,349	5,196	565	109	5	28,223	13,978
50~59세	12,068	4,270	615	93	96	17,143	2,730
60~69세	7,952	4,286	850	72	1,829	14,988	3,370
70세 이상	3,475	4,058	1,512	29	1,417	10,491	0

〈표 11〉의 서울시 가구유형별 소득원천별 가구소득(2018년)을 보면, 1인가구의 평균소득이 2,396만원, 중위소득은 1,310만원으로 4인가구의 평균소득 8,761만원과 중위소득 6,257만원의 1/4 내외에 불과하다. 이는 균등화소득 기준으로도 두 배 정도의 차이가 남을 의미한다. 다음으로 1인가구 중 연령별로 구분해보면 연령대가 올라갈수록 평균과 중위소득이 떨어진다. 특히 지난 10여년간 급증한 중고령층 1인가구에 빈곤이 집중되어 있음을 보여준다. 1인가구중 50대의 평균소득은 1,714만원, 중위소득은 273만원, 60대의 평균소득은 1,499만원, 중위소득은 337만원, 70세 이상의 평균소득은 1,049만원, 중위소득은 0로 나타난다. 60대와 70대 1인가구의 경우 공적연금소득 평균이 각각 183만원(월 15만원), 142만원(월 12만원)에 불과하여 노후소득보장은 커녕 빈곤완화에도 거의 역할을 하지 못함을 보여준다.

균등화소득 기준으로 지니계수와 상대빈곤율을 구해보면, 2006년부터 2018년 사이에 서울시의 소득분배는 크게 개선된 것처럼 보인다. 즉, 이 기간 중 지니계수가 주민등록 세대 기준으로 0.65에서 0.57로, 재편된 가구 기준으로 0.64에서 0.55로 개선되었고, 상대빈곤율(중위소득 50% 기준)도 주민등록 세대 기준으로 38.5%에서 31.1%로, 재편된 가구 기준으로 37.8%에서 29.9%로 낮아졌다. 그러나, 이를 두고 소득분배가 개선되었다고 할 수는 없다. 건보공단 자료의 소득파악률 개선에 크게 영향을 받았기 때문이다.

서울시의 1인가구와 2인이상 가구의 상대빈곤율 추이를 보면 1인가구는 2006년 57%에서 2018년 49%로, 2인이상 가구는 35.7%에서 27%로 떨어진 것으로 나타나는데, 1인가구의 빈곤율이 2018년에도 여전히 높게 나타난다. 또한, 2018년 서울시 1인 가구의 48.6%와 2인 가구의 38.9%가 월소득 100만원 미만이며, 중위 균등화소득이 1인 가구는 1,310만원, 2인 가구는 1,465만원으로 3인 가구의 2,483만원, 4인 가구의 3,129만원에 비해 매우 낮아 1인가구뿐만 아니라, 2인가구 중에도 많은 가구가 빈곤을 겪고 있음을 나타낸다.

65세 이상의 노인빈곤율도 지난 10여년간 거의 개선되지 않은 것으로 나타난다. 주민등록세대 기준으로는 53.8%에서 53.3%로, 재편된 가구 기준으로는 48.8%에서 47.8%로 약간 낮아진 했지만, 공적연금소득을 포함한 건보자료 상으로도 노인빈곤 문제는 거의 완화되지 않고 있



다. 아래 <표 12>에서 국민연금 수급개시 연령인 62세 이상 노인의 빈곤율은 49.5%인데, 가구 유형별로 보면 독거노인과 노인부부의 빈곤율이 각각 85.4%, 62.7%로서 성인동거 노인의 빈곤율 36.2%에 비해 매우 높게 나타났다.

<표 12> 2018년 서울 62세 이상 노인가구유형별 균등화시장소득, 균등화연금소득과 빈곤율

(단위: 천원, %)

	균등화시장소득(실질)		균등화연금소득(실질)		빈곤율			
	평균	중위	평균	중위	시장소득	총소득	시장소득 (전국)	총소득 (전국)
독거노인	7,030	0	1,720	0	85.1	82.3	87.9	85.0
노인부부	16,729	1,068	5,285	2,379	71.1	58.9	78.2	65.2
성인동거	27,173	16,441	3,179	1,319	36.9	32.3	39.3	34.3
전체노인	20,311	7,352	3,406	1,309	55.7	49.5	61.7	55.1

2인 이상 가구의 가구주 중 무소득자 비율이 여성은 평균 31.4%로서 남성의 15.7%에 비해 두 배에 달하며, 소득자의 평균소득은 남성이 5,731만원, 여성이 2,952만원으로 여성 가구주의 소득이 남성 가구주 소득의 51.5%에 불과하다. 또, 가구소득 월 100만원 미만에 남성 가구주 가구가 26.2%, 여성 가구주 가구가 39.5%에 달해 여성가구주 가구의 빈곤율이 높다. 장애인 가구주의 경우 가구소득 평균은 2,893만원, 중위값은 1,074만원에 불과해 다수의 장애인가구가 빈곤에 처해 있음을 알 수 있다.

주민등록 세대 기준으로 균등화소득을 시도별로 비교하면 평균소득은 서울이 가장 높고, 중위소득은 세종시, 울산시에 이어 3위로 나타났다. 서울은 평균소득 대비 중위소득 비율이 0.68로서 전남(0.64)에 이어 두 번째로 평균소득과 중위소득의 격차가 커서 소득불평등이 높음을 보여준다.

끝으로 균등화소득 기준으로 자치구별 지니계수를 구해본 결과 상당히 큰 편차가 나타났다. 2018년 현재 지니계수가 가장 낮은 5개 구는 노원구(0.48), 강서구(0.48), 도봉구(0.49), 강동구(0.49), 은평구(0.49)이며, 가장 높은 5개 구는 용산구(0.67), 종로구(0.63), 중구(0.61), 강남구(0.61), 서초구(0.58)로 나타났다. 상대빈곤율이 높은 구는 강북구(40.7%), 금천구(37.4%), 중랑구(37.0%), 중구(36.7%), 동대문구(35.4%)이며, 낮은 구는 서초구(19.0%), 강남구(21.0%), 송파구(22.9%), 양천구(25.3%), 마포구(27.3%) 순으로 나타났다.

## 6. 결론: 건보자료의 의의와 한계 및 발전방안

이상에서 건보공단의 전국민 소득재산 DB에 대한 소개와 함께 동 자료를 이용한 서울시의 소득분배 분석결과를 간단히 살펴보았다. 이를 통해 건보자료가 가지는 의의와 한계가 드러났다.

먼저 건보자료는 지역 차원의 소득분배를 파악할 수 있는 정보를 제공한다는 점에서 그 의의가 크다. 그동안 불평등과 빈곤 등 소득분배 지표가 서울시나 구와 같은 지역 차원에서는 측정되지 않았다. 따라서 지역사회보장계획의 수립과 사회보장 정책효과의 평가가 불평등과 빈곤과 같은 결과지표를 사용하지 못하고 산출지표 위주로 이루어지는 한계를 노정했다.

건보공단 자료는 전인구를 포괄하기 때문에 광역은 물론 기초 자치단체 수준의 소득분배도 파악할 수 있다. 더 나아가서 읍·면·동 단위의 지표도 산출할 수 있으며, 연령별, 성별은 물론 건보 자격, 장애인등록 등에 따른 집단별 분석도 가능하다. 또한, 인구 전체를 사망시까지 계속 추적하기 때문에 세대 내 이동성 분석이 가능하며, 부모 자녀간 관계 정보를 활용하여 세대 간 이동성 분석도 가능하다.

아직까지는 건보의 소득 정보의 불완전성 및 소득 파악률의 변화로 말미암아 건보자료로 공식적인 소득분배지표를 측정, 발표하기는 곤란하며 시계열 분석을 하는 데에도 한계가 있으나, 건보공단 자료의 소득파악률이 빠르게 개선되고 있다. 국민계정의 개인소득(가계분원소득) 통계에 비추어 건보공단의 소득파악률은 2006년 66.6%에서 2018년 91.3%로 대폭 상승했다. 특히 사업소득자 수는 통계청 경제활동인구조사의 자영업자 수의 두 배가 넘게 잡혀 전통적인 자영업자 외에 인적 용역에 대한 원천징수 또는 연말정산 사업소득을 올리는 특수고용근로자와 프리랜서 등의 소득까지 파악하는 국세행정의 개선이 반영된 것으로 보인다. 조만간 일용근로 소득과 2천만원 이하의 금융소득 등의 정보를 포함해 거의 완벽한 시장소득 정보가 구축되고 나면, 건보자료의 활용도는 매우 커질 것으로 전망된다. 또한 경찰조사와의 근로소득 분포 비교가 보여주듯 행정자료는 서베이 자료에 비해 최상위 소득층을 비롯한 고소득층과 저소득층의 파악에 유용함을 알 수 있다.

다만, 자영업자의 소득 축소신고와 재배업 등 비과세 소득, 비공식 부문 및 사적 이전소득 정보의 부재로 상대적으로 저소득층의 소득이 적게 파악되는 행정자료의 근본적인 한계는 최상위층과 1인가구 등의 모집단 대표성 확보가 어려운 가구조사의 근본적인 한계와 마찬가지로 해결하기 쉽지 않다. 또한, 세금과 사회보험료 등 비소비지출과 각종 공적이전소득 정보의 부재로 인해 가처분소득을 구할 수 없는 것은 소득분배 지표의 측정과 연구에 근본적인 한계를 노정한다. 따라서 행정자료와 조사자료의 결합을 통해 상호 장점의 활용과 단점의 보완을 통한 소득분배 지표의 개발이 필요하다.

시장소득 정보의 불완전성과 가처분소득의 부재 외에 가구 정보의 불완전성도 건보자료의 한계로 지적할 수 있다. 비록 주민등록 행정자료를 이용해 건강보험 자격에 따른 세대 외에 주민등록 세대 단위로 전인구의 소득을 파악할 수 있도록 하고 있으나, 인구주택총조사나 가계금융복지조사의 가구 개념과는 차이가 있다.

끝으로 건강보험공단의 소득자료 입수가 실시간으로 이루어지지 못하고, 더구나 연구용으로 제공되는 데에는 더 많은 시간이 소요되는 한계가 있다. 직장가입자인 근로자의 경우는 매년 3월에 전년도 보수의 연말정산 자료를 입수해 4월부터 보험료에 적용하고 있으나, 지역가입자의 경우 전전년도 귀속 소득을 사용한다. 이에 따라 2020년에 이루어진 건보자료 이용 서울시 소득분배 연구는 2018년도 귀속 소득까지만 접근할 수 있었고, 2019년도 소득정보는 볼 수가 없었다. 이 문제는 향후 정부가 실시간 소득파악 시스템을 구축하는 계획이 실현되면 개선될 수 있지 않을까 기대한다. 지난 3월 정부의 소득중심 고용보험 추진 방침에 따라 국세청에 소득자료관리준비단을 설치하여 소득자료를 월별로 수집하여 근로복지공단에 제공할 수 있는 실시간 소득파악 시스템을 구축할 계획이라고 한다. 올해 7월부터 일용근로자, 인적용역형 사업소득자에 대한 소득자료 제출주기가 월별로 단축될 예정이고, 플랫폼 종사자에 대해서는 기존의 연 단위 제출주기를 단축하는 문제가 국회에서 추가적으로 논의될 예정이다(국세청, 2021). 이에 따라 조만간 우리 국세청도 영국이나 핀란드와 같이 전국민 실시간 소득정보 시스템을 갖추게 될 수 있을 것으로 보인다. 그렇게 되면, 건보공단도 실시간 소득정보를 받아서 건보료 산정에 활용할 수 있게 되고, 연구용 자료 제공도 보다 신속하게 이루어질 수 있을 것으로 전망된다.

정교한 정책 및 학술 연구를 위한 정보에 대한 접근성이 떨어진다는 점도 언급될 필요가 있다. 건강보험공단에서는 연구계획 심사 및 안전 공간 접속 제한을 활용하여 연구 데이터베이스를 외부 연구자들이 활용할 수 있도록 하고 있다. 그런데 건강보험 자료공유서비스를 통해 일반적으로 접근할 수 있는 정보는 포괄적이지 않다. 건강보험공단에서 보유하고 있는 전국민 소득-재산 데이터베이스에 포함된 변수들이 자료공유서비스 사이트에서는 접근이 되지 않는 상태에서 보험료 기준의 10분위 자료만 제공되는 것이다. 소득 및 재산과 관련하여 보다 상세한 자료에 접근하기 위해서는 특별한 절차를 밟아야 한다.<sup>15)</sup>

본 연구진은 현재 가금복 자료와 건보공단 자료의 결합을 추진 중이며, 향후 고용보험 자료와 건보자료의 결합을 추진할 계획이다. 가금복 자료와의 결합은 건보자료의 시장소득 불완전 및 가처분소득 부재를 보정하는 데 도움이 될 것이며 또한 가금복 샘플이 전체 인구와 가구(주

15) 전국민 소득-재산 데이터베이스를 이용하기 위해서는 건강보험 자료공유서비스에서 제시하는 절차를 따라 소속기관 IRB 승인 및 건보공단 자료제공심의위원회의 승인을 받아야 한다. 건강보험공단 본부의 빅데이터 실 내에서만 자료에 접근하고 분석을 할 수 있으며 별도의 승인을 받은 후 분석결과만을 반출할 수 있다.

민등록 세대) 구성을 정확하게 대표하지 못하고 있는 문제를 보정할 수 있는 정보를 얻을 수 있을 것으로 기대한다. 또, 건보자료가 부동산 재산세 과세표준 외에 금융자산이나 사업자산 정보를 가지고 있지 못한데, 2천만원 이하 금융소득 정보의 입수와 함께 가금복의 금융자산 정보와의 결합은 금융자산을 포함한 자산분배를 파악하는 데에도 도움이 될 것이다.

또한, 건보공단 자료와 사회보장 등 타 행정자료 및 서베이자료와의 결합을 탐색 중인데, 증거기반 정책을 위해 가명정보 결합을 촉진시키고자 하는 정부의 방침에도 불구하고 행정자료를 보유한 개별 기관들은 아직 자료 공유에 소극적인 경향이 있어 지난한 노력이 소요될 것으로 보인다. 가명정보 결합을 통한 사회과학 및 정책연구의 활성화를 위해 정부의 적극적인 정책의지 뿐 아니라 공무원들의 인식 제고, 또한 학계의 적극적인 관심과 노력이 필요하다고 본다.

## ■ 참고문헌 ■

- 계봉오, 황선재, 최을 (2019). 행정자료를 이용한 세대 간 분위소득 이동 분석. 정해식 외(편). 소득불평등 심화의 원인과 정책적 대응 효과 연구 2. 서울: 경제인문사회연구회 협동연구총서 18-34-01.
- 구인회, 강상경, 김용득, 홍백의, 정찬미 (2020). 사회보장 행정데이터의 효과적 활용 및 발전 방안 연구. 보건복지부·서울대학교 산학협력단.
- 국세청 (2021). 소득자료관리준비단 출범 보도자료 (3월 11일).
- 김낙년 (2018). 한국의 소득집중도: Update, 1933-2016. 한국경제포럼. 11(1): 1-32.
- \_\_\_\_\_ (2020). 가계조사의 행정자료에 의한 보정: 2016년 가계금융복지조사를 중심으로. 한국사회정책. 27(1). 39-61.
- 김낙년, 김종일 (2013). 한국 소득분배 지표의 재검토. 한국경제의 분석. 19(2): 1-50.
- 김승연, 유종성, 최광은, 이해림, 전병유, 정준호, 김민혜, 이승주, 유수진, 박종현, 배세진 (2020). 행정 자료를 활용한 서울의 불평등과 빈곤연구(1차연도): 건강보험데이터를 활용한 서울시의 소득 및 자산분배. 서울연 2020-CR-15. [http://www.kirl.re.kr/bbs/board.php?tbl=ga\\_bbs32](http://www.kirl.re.kr/bbs/board.php?tbl=ga_bbs32)
- 박종현 (2021). 건강보험공단 인구사회경제 및 소득재산DB. 행정자료를 활용한 서울시의 소득 및 자산 분배 기초연구 보고회 자료집(2021. 2. 25). <http://www.si.re.kr/node/64468>
- 박세진 (2020). 국세청 업무코드 분류 어려운 '기타 자영업자' 315만명. 한국세정신문. (2020.12.31.) <http://www.taxtimes.co.kr/news/article.html?no=247838>
- 유종성 (2020). 행정빅데이터 활용 성과관리체계 운영 방안 검토. 이채진, 정혜진, 김지원, 유종성(편). 정부 중장기계획 평가방안 연구. 국회미래연구원, 95-136.
- 유종성, 전병유, 신광영, 이도훈, 최성수 (2020). 증거기반 정책연구를 위한 행정자료의 활용. 한국사회 정책. 27(1). 5-37
- 이현주, 오욱찬, 이윤경, 이원진, 성재민, 이길제, 박형준, 이병재 (2020). 사회보장정책 효과성 분석을 위한 행정데이터 연계·활용 방안. 한국보건사회연구원.
- 전병유 (2019). 소득 관련 통계 인프라 실태 파악과 개선방안. 정해식 외(편). 소득불평등 심화의 원인과 정책적 대응 효과 연구 2. 서울: 경제인문사회연구회 협동연구총서 18-34-01.
- 통계청 (2016). 2016년 가계금융·복지조사 결과 보도자료.
- \_\_\_\_\_ (2017). 2017년 가계금융·복지조사 결과 보도자료.
- \_\_\_\_\_ (2018a). 2018년 가계금융·복지조사 결과 보도자료.
- \_\_\_\_\_ (2018b). 2018년 인구주택총조사, 등록센서스 방식 집계결과: 보도자료.
- \_\_\_\_\_ (2019). 2019년 가계금융·복지조사 결과 보도자료.
- 홍민기 (2015). 최상위 소득비중의 장기 추세(1958-2013). 경제발전연구. 21(4). 1-34.
- \_\_\_\_\_ (2020). 최상위 소득 비중(~2018년). 노동리뷰(2월호). 87-89.
- Bee, A. & Rothbaum J. (2019). The Administrative Income Statistics (AIS) Project: Research on the use of administrative records to improve income and resource estimates. SEHSD Working Paper Number 2019-36.
- Chetty, R., Friedman, J. N., Saez, E. & Yagan, D. (2018). The SOI batabank: A case study in leveraging administrative data in support of evidence-based policymaking. Statistical

- Journal of the IAOS. 34. 99-103. DOI: 10.3233/SJI-170418.
- Chun, A. Y., Michael D. Larsen, G. D. & Jerome P. R. (eds). (2021). *Administrative records for survey methodology*. Hoboken, NJ: Wiley.
- Cole, S., Iqbal D., Anja S. & Lars V. (2020). *The handbook on using administrative data for research and evidence-based policy*. Abdul Latif Jameel Poverty Action Lab (J-PAL).
- Grusky, D., Hout, M., Smeeding, T. M. & Snipp, C. M. (2019). The american opportunity study: A new infrastructure for monitoring outcomes, evaluating policy, and advancing basic science. RSF: The Russell Sage Foundation Journal of the Social Sciences. 5(2). 20-39.
- Kangas, O. (2016). "From idea to experiment" Report on universal basic income experiment in Finland. Kela Working Papers 106-2016. Kela, Helsinki.
- Kangas, O., Signe J., Miska S., Minna Y. (eds). (2019). *The basic income experiment 2017-2018 in Finland. Preliminary results*. Ministry of Social Affairs and Health, Helsinki.
- Kim, C. H. & Tamborini, C. R. (2014). Response errors in earnings: An analysis of the survey of income and program participation matched with administrative data. Sociological Research & Methods. 43(1). 39-72.
- Medalia, C., Meyer, B., O'Hara, A. & Wu, D. (2019). Linking survey and administrative data to measure income, inequality, and mobility. International Journal of Population Data Science. 4(1). 04.
- Meyer, B., Mok, W. K. C. & Sullivan, J. X. (2015). Household surveys in crisis. Journal of Economic Perspectives. 29(4). 199-226.
- OECD (2020). OECD Income distribution database (IDD): Gini, poverty, income, methods and concepts. [www.oecd.org/social/income-distribution-database.htm](http://www.oecd.org/social/income-distribution-database.htm)
- Penner, A. M. & Dodge, K. A. (2019). Using administrative data for social science and policy. RSF: The Russell Sage Foundation Journal of the Social Sciences. 5(3). 1-18.
- Piketty, T. (2014). *Capital in the twenty-first century*. Translated by Arthur Goldhammer. Cambridge, MA: Harvard University Press.

Abstract

## The Merits and Limitations of the NHIS Big Data in the Study of Income Distribution

Jong-sung You\* · Minhye Kim\*\* · Seung-Yeon Kim\*\*\* · Suzin You\*\*\*\*

We examine the merits and limitations of the National Health Insurance Service (NHIS) big data as well as the traditional household survey data in the study of income distribution, including the measurement of income distribution indicators. The NHIS has compiled a full-population panel dataset for market income, public pension income, and properties tax for the years starting from 2003 or 2005. The data is particularly useful for the study of income distribution at sub-national levels. However, the information on market income in the data is not perfect and there is lack of information to construct disposable income. It is fortunate that the data's capturing of market income has been rapidly increasing and is expected to be almost perfect in the near future. In the case of business income of the self-employed, the data provide full information on business income of individuals, including personal service providers such as special employees and freelancers. We also suggest ways to improve measurement of income distribution indicators based on the Survey of Household Finance and Living Conditions (SHFLC) and new strategies to merge survey data and administrative data to measure income inequality at local level.

**Keywords:** income distribution, administrative data, NHIS big data, full-population income and wealth DB

◆ 2021. 7. 31. 접수 / 2021. 9. 1. 1차수정 / 2021. 9. 15. 게재확정

\* First Author, Professor and Director, Inequality and Social Policy Institute, Gachon University (jongsungyou@gmail.com)

\*\* Corresponding Author, Assistant Professor, Department of Sociology, Changwon National University (minhyekim@changwon.ac.kr)

\*\*\* Coauthor, Director, Department of Urban Society Research, Seoul Institute (francies@si.re.kr)

\*\*\*\* Coauthor, Data Analyst, Inequality and Social Policy Institute, Gachon University (suzinyou.sy@gmail.com)