

# 행정자료를 이용한 소득 정보 구성: 국세 과세 미시자료의 활용 사례\*

박 형 존\*\*

## 요약

이 연구는 국세 과세 미시자료를 이용하여 사회정책 연구에 활용이 가능한 소득 정보를 구성하는 방법을 사례를 들어 고찰하는 것이 목적이다. <한국복지패널>과 <가계금융복지조사>의 소득 측정과 국세 과세자료의 세법상 소득 정의를 비교 검토함으로써 국세 과세 기반 소득자료의 특성을 파악하였고, 국세 과세자료를 이용한 2020년과 2021년 연구의 원천별 소득 및 소득세액의 산출 경험을 소개하였다. 그리고 국세 과세자료에서 생성한 소득과 <가계금융복지조사>의 소득을 몇 가지 대푯값을 기준으로 비교하였다. 연구 결과, 국세 과세자료와 사회조사 원천별 소득은 명칭 자체는 동일하나, 사회조사의 측정과 국세 자료의 세법상 정의는 일치하지 않는 부분이 상당하였고, 국세 자료를 이용한 선행연구에서 취한 소득 구성 방식에는 자료의 한계로 인해 검증할 수 없는 가정이 가미되어 있음을 알 수 있었다. 이는 사회조사의 소득과 국세 과세자료의 소득 대푯값 간 차이를 발생하게 하는 여러 이유 중 하나로 작용하였다. 다만 국세 과세자료 이용 경험에서 발견된 한계는 대부분 동 자료의 이용 범위를 넓히는 것으로 연구자가 다룰 수 있는 것이었고, 이에 향후 국세 과세자료를 이용한 소득 연구에서 자료 추출시 추가적으로 고려해야 할 점을 결론으로 정리하였다. 아울러 사회정책 연구를 위한 국세 과세자료의 가능성으로 비정형노동자를 특정할 수 있는 자료 이용 방안에 대해서도 제안하였다.

**주요어:** 국세청, 원천별 소득, 종합소득세, 원천징수, 사회조사

\* 이 논문은 2022년 한국사회정책학회 춘계학술대회에서 발표한 것을 보완한 것이다.

\*\* 세종시사회서비스원 정책연구부 연구위원(sheppig@sj.pass.or.kr)

## 1. 서론

이 연구는 국세 과세자료를 이용하여 사회과학 연구에 활용이 가능한 소득 정보를 구성하는 방법을 기술적인 차원에서 고찰하는 것을 목적으로 한다. 사회보장 및 관련 정책 연구 분야는 가구 및 개인 단위의 소득, 경제활동, 사회보장 수급 관련 정보를 파악할 필요가 있기 때문에 행정자료의 이용 수요가 높다. 이에 이현주 외(2020; 2018), 구인회 외(2021; 2020), 여유진 외(2020; 2017), 강신욱 외(2017; 2016), 한은희 외(2021) 등 국책연구기관을 위시하여 수행되어 온 정부의 연구용역은 행정자료를 꾸준히 활용해왔고, 앞으로 이러한 연구 수요는 학계의 개인 연구자를 포함함으로써 더욱 커질 것으로 예상된다. 행정자료는 사회보장의 거의 전 분야에서 수집·관리되고 있지만 그 중에서도 소득은 개인 및 가구의 인구사회적 속성만큼이나 기본이 되는 정보이자 그것만으로도 별도의 연구 영역이 되는 중요한 자료이다.

국세청의 국세 과세자료는 전 국민의 소득 정보를 얻기에 가장 적합한 행정자료일 것이다. 우리나라의 공식 소득분배지표 생산 근거 자료인 통계청의 <가계금융복지조사>도 가구의 소득을 조사한 후 국세청으로부터 연결한 소득 정보로 조사 자료의 일부를 대체함으로써 소득 파악의 정확성을 높이고 있다(통계청, 2019; 이원진, 정해식, 전지현, 2019). 또한 국세청은 2018년 6월 국세통계센터를 개소하여, 국세 미시자료를 국세법상 명시된 기관을 대상으로 제공하고 있었다. 2020년에는 그 이용 대상을 대학, 공공기관, 민간연구기관으로 확대하고, 2020년 12월 ‘데이터결합전문기관’<sup>1)</sup>으로 지정된 이후 2021년 2월부터는 지방공기업 지자체 출연연구원, 특수법인에도 국세 미시자료를 이용할 수 있도록 대상을 확대하였다(국세청, 2021a).

그동안 국세청의 국세 과세자료는 국민건강보험공단의 4대 사회보험 징수를 위해 연계되거나, 사회보장정보원의 사회복지통합관리망에 연계되어 자산조사를 해야 하는 사회보장제도의 수급 자격을 판정하는 등의 용도로 이용되어 왔으며, 전술한 바와 같이 일부 사회조사의 조사 정보를 보완하는 방식으로 이용되어 왔다. 최근의 사회보장위원회 사무국(보건복지부)의 연구용역에서는 국세청으로부터 자료를 협조받아 이른바 ‘국세청 소득자료’를 연구에 활용하기도 하였는데, 국세 과세자료를 바탕으로 사회조사의 원천별 소득의 형태로 소득 정보를 구성하고 여러 정보보유기관의 사회보장제도의 수급 관련 행정자료를 연결하여 공적이전의 재분배 효과를 살펴본 것이다(구인회 외, 2021; 이현주 외, 2020).

1) 서로 다른 정보보유기관의 개인정보를 가명처리하여 결합할 수 있도록 지정한 기관으로, 개인정보보호법 제 28조의3 제1항에 근거한다. 2022년 6월 기준 공공 부문의 결합전문기관은 개인정보보호위원회가 지정한 통계청, 국세청, 한국사회보장정보원, 한국지역정보개발원, 보건복지부가 지정한 건강보험심사평가원, 국민건강보험공단, 한국보건산업진흥원, 국토교통부의 한국도로공사, 과학기술정보통신부의 한국지능정보사회진흥원, 한국데이터산업진흥원, 교육부의 한국교육학술정보원, 산업통상자원부의 한전KDN, 행정안전부의 국가정보자원관리원이 있다.

이현주 외(2020)에서는 각종 사회보장 관련 행정데이터를 연결하였고, 국세 과세자료를 활용하여 국내 연구로서는 처음으로 행정자료만으로 소득 변수를 생성하여 분석에 활용하였다. 하지만, 국세기본법에 따라 국세 과세자료 및 생성한 소득 정보는 국세청 외부로 반출할 수 없었다. 기결합된 데이터 일체를 국세청 국세통계센터에 반입하여 과세자료와 결합하여 사용한 뒤, 일정 기간이 지나면 과세자료가 결합된 사회보장 행정데이터는 파기하였던 것이다. 구인회 외(2021)의 연구에서도 마찬가지였다. 이러한 점은 애써 결합한 데이터를 다른 연구에는 활용할 수 없다는 것 외에 더 큰 문제가 있었는데, 그것은 과세 자료를 활용하여 소득 정보를 구성하였던 선도적인 경험이 후속 연구에 전해질 수 없다는 점이다. 구인회 외(2021)와 이현주 외(2020)의 연구는 소득 정보를 구성하는 것 자체가 목적이 아니었다. 따라서 국세청의 과세 미시자료를 활용하여 소득 변수를 생성한 의미있는 사례였음에도 이 경험을 관련 연구 분야에 공유하는 것에 한계가 있었으며, 작업 과정과 소득 생성 결과의 타당성을 검증받을 기회를 갖지 못하였다.

이에 연구는 앞선 두 연구에서 수행한 국세 과세자료를 이용한 소득 정보 구성의 기술적인 측면을 공유하고자 한다. 행정자료를 이용한 정책 분석에서 소득은 사회과학의 거의 모든 분야에서 핵심 변수이다. 사회보장 부문 행정데이터를 이용하게 될 사용자들 간에 국세 미시자료를 바탕으로 한 소득의 구성과 활용에 대한 이해가 높아질 필요가 있다. 이는 향후 사회보장 부문에서 빈번해질 행정자료 연계와 활용을 사전적으로 대비하는 점에서도 필요하다. 이 연구가 행정자료를 이용한 소득 정보의 구성 방식을 서로 공유하고 참고할 수 있는 시작점이 되기를, 그리고 연구에서 취한 방법에 대해 면밀하고 비판적인 검토가 이루어지기를 기대한다.

글의 구성은 다음과 같다. 2장에서는 주요 사회조사와 국세청이 수집하고 있는 원천별 소득의 내용을 검토한다. 이는 연구에서 국세 과세자료를 이용하여 만들고자 하는 소득 정보가 학술 연구에서 통용되는 소득 개념과 이론적으로 어느 정도 부합하는지, 실무적으로는 사회조사에서 측정하고 있는 원천별 소득과 어떤 면에서 일치하고 차이는 무엇인지 알기 위함이다. 3장에서는 국세 과세 원자료를 바탕으로 소득 및 소득세 정보를 구성하는 방법을 구인회 외(2021)과 이현주 외(2020)의 경험을 기준으로 설명한다. 당시 작업에서 고려한 몇몇 가정과 한계 등도 포함한다. 4장에서는 국세 과세자료를 바탕으로 한 소득과 주요 사회조사의 소득을 몇 가지 대푯값을 기준으로 비교한다. 5장에서는 결론으로써 국세 과세자료를 이용한 소득 정보의 가능성과 한계, 앞으로 보완되어야 할 점을 논한다.

## 2. 사회조사 및 국세 과세자료의 원천별 소득 구분

### 1) 원천별 소득의 개념

원천별 소득의 개념적 정의에는 큰 이견이 없는 반면, 원천별 소득과 지출액을 조합하여 시장소득, 가처분소득과 같은 소득 개념을 산출하는 방식은 사용하는 자료와 시점에 따라 약간의 차이가 있다(여유진 외, 2005; OECD, 2012; OECD, 2017; 한국보건사회연구원·서울대사회복지연구소, 2021).<sup>2)</sup> 이 연구의 주제는 소득 개념 자체가 아니라, 기존 학술 연구에서 사용하는 원천별 소득 등을 국세 과세자료에서 어떻게 구현하는가라는 기술적 차원의 고찰이기 때문에 여러 연구자가 소개한 것 중 하나를 선택하여 논의를 이어가도록 한다.

[그림 1]은 한국보건사회연구원의 <한국복지패널조사>에서 준용하고 있는 여유진 외(2005)에서 소개한 소득 구분으로, 통계청의 방식과 OECD의 소득 개념을 살펴본 것이다. 근로소득, 사업 및 부업소득, 재산소득은 노동시장에서 벌어들이는 요소소득으로서, 이들을 합산하여 OECD 기준의 1차 소득으로 명명한다. <한국복지패널>과 <가계금융복지조사>와 같은 사회조사에서는 이처럼 대별한 원천 소득의 이하 내용을 근로소득은 상용근로와 임시일용근로, 사업 소득은 고용주 및 자영업자, 농림어업 등으로 나누어 파악하고, 재산소득은 금융소득, 임대소득 등으로 나누어서 파악하고 있다.

이처럼 소득 부문이 특화된 사회조사에서는 경상소득을 이루는 각 원천별 소득과 소득세, 사회보장부담금을 조사하고 있기 때문에, 총소득(통계청의 경상소득)과 가처분소득까지 산출할 수 있다. 하지만 국세 과세 단일 자료만으로 이들을 구현해 낼 수 없는데, 기본적으로 국세청에서 파악할 수 있는 소득 관련 자료는 아래의 그림을 기준으로 근로소득, 사업 및 부업소득, 재산소득, 소득세까지이기 때문이다. 국세 과세자료에 구인회 외(2021)와 이현주 외(2020)와 같이 다른 정보보유기관의 공적이전소득과 사회보장부담금 정보가 결합된다면 사적이전소득을 제외하고는 아래의 정의와 최대한 근접한 소득 개념을 구현해 낼 수 있다. 그렇기에 타 행정자료의 결합 전에 국세 과세자료에서 원천별 소득과 소득세액 정보를 타당하게 구성하는 것이 향후 행정자료 활용 연구의 관건이 될 것이다.

2) 일례로 OECD IDD(Income Distribution Database)의 2012년 정의에서는 본문의 [그림 1]과 같이 1차 소득을 정의하였지만(OECD, 2012), 2017년부터 변경된 OECD 정의에 의하면 기존의 1차 소득에 사적이전소득에서 사적이전지출을 차감한 금액을 더하여 1차 소득을 정의하고 있다(OECD, 2017). 국내의 2019년 이전의 빈곤 통계연보와 이전의 통계청 소득분배 지표도 OECD(2012)의 기준을 준용하다가 최근 새로운 정의에 맞추어 소득을 생성하고 있으며, <한국복지패널>에서 이용자 편의를 위해 제공하고 있는 생성 소득 변수는 기존의 정의를 그대로 이용하기도 한다. 사적이전은 수입과 지출 모두 국세 과세자료에서는 파악이 불가하기 때문에 논의의 편의를 위해 사적이전을 감안하지 않은 기존의 1차 소득 정의를 준용하였다.

[그림 1] 소득 구분

통계청			OECD				
소득	비경상 소득	경조소득 퇴직금 양도소득 기타					
		근로소득	Wage and Salary Income	1차소득 (Primary Income)	시장소득 (Market Income)	총소득 (Gross Income)	가처분소득 (Disposal Income)
	사업 및 부업소득	Gross Self-Employment Income					
	재산소득	Realized Property Income					
	이전 소득	사적	Occupational Pensions & other private cash				
공적		social insurance, social assistance, other cash benefit				소득세 + 사회보장 부담금	

자료: 박찬용·강석훈·김태완(2002); 유경준, 김대일(2003); 여유진, 김미곤, 김태완, 양시현, 최현수(2005)에서 재인용.

국세 과세자료에서도 상기 원천별 소득의 구분을 따르고, 명칭에서만은 동일하다. 하지만, 사회조사 자료에서 수집하고 있는 소득 정보를 기준으로 국세 과세자료의 소득 정보를 검토하기 위해서는 측정 방식의 차원으로 내려갈 필요가 있다. 구분과 명칭은 같더라도 사회조사와 국세 과세자료의 실제 측정은 다르기 때문이다.

## 2) 주요 사회조사와 국세 과세자료의 소득 측정

사회보장 부문 연구에서 널리 활용되고 있는 통계청의 <가계금융복지조사>와 <한국복지패널조사> 자료의 소득 측정 방식과, 국세 과세자료의 소득에 대한 세법과 국세통계연보의 정의를 검토한다. 세법과 국세통계연보 상의 정의는 국세청이 수집하는 개인별 과세자료의 정의와 일치하기 때문에 사회조사 자료의 조사 지침과 동일하게 간주할 수 있을 것이다. 그럼으로써 국세 과세자료에서 구현 가능한 소득 정보의 특성을 이해하고자 한다.

아래의 표는 <가계금융복지조사>의 원천별 소득 변수와 조사 지침의 내용을 정리한 것이다. <가계금융복지조사>는 이후 살펴볼 <한국복지패널>에 비해 상대적으로 단순한 근로소득, 사업소득, 재산소득의 세 가지 변수를 제공하고 있다. 재산소득 이하의 세부 변수는 금융소득, 임대소득, 기타로 나뉘어 있다. 근로소득은 세금과 사회보장부담금이 공제되기 전의 현금 및 현물

보수로 정의되어 있어, 앞서 살펴본 가처분소득 생성을 염두하고 있음을 알 수 있다. 사업소득과 재산소득은 수입금액에서 비용을 차감한 것을 소득으로 규정하고 있다. 국세 과세자료에서도 근로소득과 사업소득 공히 수입금액과 소득금액을 구분하고 있고, 사업소득에서의 ‘소득’은 경비를 차감한 것을 의미한다.

**[표 1] <가계금융복지조사>의 소득 측정**

구분	변수	조사 지침
근로소득	근로소득	근로의 대가로 받은 현금 및 현물 보수로서, 세금 또는 건강보험료·고용보험료 등 각종 부담금이 공제되기 전 금액
사업소득	사업소득	도소매업, 서비스업 등 사업경영을 통해 얻은 사업수입(매출액)에서 인건비, 재료비, 대출이자 등 비용을 차감한 금액. 자가소비용 생산품 등 현물소득 포함
재산소득		재산의 운영을 통해서 발생하는 소득(금융소득, 임대소득)에서 이자 등 비용을 차감한 금액. 현물소득 포함
	↳ 금융소득	채권·예금·적금 및 개인 간 거래 등에서 발생하는 이자소득 및 보험·신탁·주식 등에서 발생하는 배당소득에서 대출이자 등 비용을 차감한 금액
	↳ 임대소득	주택, 건물, 토지 임대에 따른 임대수입에서 임대용 부동산 유지·관리비, 대출이자 등 비용을 차감한 금액
	↳ 기타	이외 개인연금, 퇴직연금 및 유형자산(주택, 건물, 토지 이외)과 무형자산(저작권, 인세, 특허권 등)으로부터의 소득 등

자료: 2020년 <가계금융복지조사> 조사표를 바탕으로 저자 작성

<표 2>는 <한국복지패널>에서 수집하는 소득 정보와 제공 변수, 조사 지침을 정리한 것이다. <가계금융복지조사>와 동일하게 근로소득, 사업소득, 재산소득을 파악하고 있으나, 다른 점은 기타근로소득과 개인연금 및 퇴직연금을 재산소득과 분리하여 파악하고 있다는 점이다. <한국복지패널>은 <가계금융복지조사>에 비해 각 원천별 소득을 보다 세분하여 변수로 측정하고 있고, 표에는 제시되어 있지 않은 보조기입란까지 이용하여 상당히 많은 변수를 측정하고 있는데, 이는 각 원천별 소득을 쪼개어 물어봄으로써 연중 개인이 올리는 소득을 명목별로 정확하게 측정하기 위함이다. 기타근로소득은 근로소득으로, 개인연금과 퇴직연금은 재산소득으로 합산하여 사용한다면 <가계금융복지조사>의 근로, 사업, 재산소득의 분류와 차이점은 없다. <한국복지패널> 역시 근로소득은 세금 및 각종 사회보장부담금을 공제하기 전 금액으로 명시하고 있으며, 사업소득은 연간 총비용을 감하여 구한다고 명시한다.

[표 2] <한국복지패널조사>의 소득 측정

구분	변수	조사 지침
근로소득	상용근로 총급여액	고용 계약 기간이 1년 이상인 사람 또는 계속 정규직원으로 일하면서 받는 기본급여, 상여금, 초과근무수당, 성과급, 소득공제 환급분, 판공비, 복지카드 등을 모두 포함한 금액 세금 및 각종 사회보장부담금을 공제하기 전 금액
	임시일용근로 총급여액	임시근로자: 임금근로자중 상용이 아닌 사람으로 고용계약 기간이 1개월 이상 1년 미만인 사람 또는 일정한 사업완료(예컨대, 프로젝트)의 필요성에 의해서 고용된 사람 등 일용근로자: 임금 또는 봉급을 받고 고용되어 있으면서 고용계약기간이 1개월 미만인 사람 또는 일정한 사업장 없이 떠돌아 다니면서 일한 댓가를 받는 사람 등 세금 및 사회보장부담금 공제 전 금액이며, 자활근로, 공공근로, 노인 일자리사업 소득을 포함
사업소득	고용주 및 자영자 순소득	연간 총매출액에서 연간 총비용을 감하여 구함 세금과 사회보장부담금을 공제하기 전 금액
	농림축산업 순소득	판매수입과 기타잡수입에서 총 비용을 차감
	어업 순소득	판매수입과 기타잡수입에서 총 비용을 차감
기타 근로소득	기타근로소득	상기 응답한 근로 외 주 1시간 미만의 근로활동으로 벌어들인 소득
재산소득	이자(은행, 사채), 배당금	
	임대료(월세, 토지임대료 등)	
	기타(자격증 대여 등)	
민간보험	개인연금	일시금은 제외
	퇴직연금	

자료: 15차 <한국복지패널> 가구 조사표를 바탕으로 저자 작성

상기 사회조사에서의 소득 범위는 비과세소득을 포함한다. 비과세 소득은 면세점 이하의 소득과 면세점 이상 소득자의 과세 대상이 아닌 수입금액을 공히 포함한다. 국세 과세자료는 후자가 포함되어 파악되며, 면세점 이하의 비과세 소득도 인적 공제 등의 혜택을 받기 위해 원천징수의무자가 지급명세서를 신고한다면 원리적으로는 파악이 된다.

이상 살펴본 사회조사의 소득 변수들은 원천별로 조사되지만, 소득세는 하나의 변수로 측정된다. 반면, 국세 과세자료의 소득은 과세를 위해 신고된 수입금액 혹은 소득금액이다. 크게는 종합소득세 신고 부문과 원천세 부문으로 나뉘어지며, 국세청에 파악된 이들 수입·소득 금액이 양 부문에 상당 부분 중첩된다. 소득세액 또한 그러하다. 사회조사의 원리상 조사표의 각 소득은 조사문항 간 상호 배타이면서 이들을 단순 합산하면 가구의 소득을 모두 포괄하게끔 설정되어 있지만, 국세 과세자료는 조세 제도의 작동을 위해 수집된 자료이기 때문에 개인의 수입금액이 원천세 부문과 종합소득세 신고 부문 양쪽에 존재할 수 있으며, 심지어 명칭도 같다. 연구자는 동 자료를 이용할 때 각 자료 추출원의 성격을 이해하고 가공하여 사회조사의 원천별

소득 변수와 동일하게 상호 배타성과 포괄성을 담보해야만 한다.

아래의 <표 3>은 구인회 외(2021)와 이현주 외(2020)에서 활용하였던 국세 과세자료 변수를 바탕으로 국세청 홈페이지의 국세 신고 안내와 국세통계연보를 참조하여 사회조사의 <표 1>과 <표2>와 유사하게 정리한 것이다. <표 3>에는 당시 연구에서 활용하지 않았던 변수도 있지만, 소득 정보의 상호 배타와 포괄성을 설명하기 위하여 저자가 추가로 기입하였다.

항목명에 “\*” 표시가 된 것은 상호 배타의 원칙이 담보되지 않음을 의미한다. 즉, 해당 변수는 학술 연구의 원천별 소득의 성격을 두 가지 이상 갖는 것이다. 원천징수 대상 사업소득은 세 번째 컬럼의 설명처럼 인적 용역에 의한 소득으로써 근로 형태를 보면 사실상 근로소득에 해당한다. 대부분이 사업자등록이 되어 있지 않기도 하다. 만약 이 소득을 올리고 있는 개인이 <가계금융복지조사>나 <한국복지패널조사>에 응답한다면, 자신이 벌어들이는 소득의 성격을 무엇으로 응답할지 모호하다. 하지만 국세 과세자료에서 이들의 수입은 사업소득으로 데이터화 되어 있다.

사회조사와의 또 다른 차이점은 결이 다른 무수한 종류의 소득 금액들이 기타소득이라는 명칭으로 혼합되어 있다는 점이다. 기타소득 총지급금액에는 앞서 살펴본 <가계금융복지조사>에서 재산소득으로 조사되는 유무형 자산의 양도 및 대여의 대가, 저작권료 등이 있는가 하면, 강연료, 원고료와 같이 <한국복지패널>에서는 기타 근로소득으로 파악될 근로소득들도 있다. 상금과 복권 당첨금 등은 경상소득으로 볼 수 없다.

[표 3] 국세 과세자료의 소득과 세법상 정의

사회조사의 구분	국세통계연보 상 명칭(징수 부문)	세법상 정의
근로소득	근로소득 총수입금액(종합소득세)	특정 고용주에게 계속하여 고용되어 지급받는 급여
	연말정산 대상 근로소득(원천세 징수)	특정 고용주에게 계속하여 고용되어 지급받는 급여
	원천징수 대상 사업소득(원천세 징수)*	원천징수 대상 의료보전 용역, 원천징수 대상 인적용역 사업소득(분류업종 구분:저술가, 화가관련, 작곡가, 배우, 모델, 가수, 학원강사, 보험설계, 음료배달, 방문판매, 간병인, 대리운전, 기타자영업, 퀵서비스, 물품배달, 기타모집수당 등)
	연말정산 대상 사업소득(원천세 징수)*	간편장부대상자*로서 다음 어느 하나에 해당하는 사업자에게 사업소득을 지급하는 원천징수의무자는 해당 과세기간의 사업소득 금액에 대하여 연말정산하여 소득세를 징수(해당사업자: 보험모집인, 방문판매원, 음료배달판매원)
	일용근로소득(원천세 징수)	특정 고용주에게 계속하여 고용되어 있지 아니하고 일급 또는 시간급 등으로 받는 급여 연말정산 대상에 해당되지 아니함(지급시 원천징수로써 납세 의무 종결)
	기타소득(원천세 징수)*	상금, 복권 당첨금 등, 자산 등의 양도·대여·사용의 대가, 보상금 등 우발적인 소득, 인적용역소득, 서화·골동품 양도



		에 의한 소득, 종교인소득, 기타 *저작권 또는 저작권접권의 양도·사용의 대가, 강연료, 원고료 등이 포함됨
사업소득	사업소득, 부동산업 소득 (종합소득세 신고)	종합소득세 신고한 사업 및 부동산 소득
	원천징수 대상 사업소득(원천세 징수)*	상기 "원천징수 대상 사업소득(원천세 징수)*"과 동일
	사업자 연말정산(원천세 징수)* (선행연구 미활용)	상기 "연말정산 대상 사업소득(원천세 징수)*"과 동일
	기타소득(원천세 징수)*	상기 "기타소득(원천세 징수)*"과 동일
재산소득	부동산업 소득(종합소득세 신고)*	사업소득 중 부동산업에서 발생하는 소득
	기타소득 (원천세 징수) *	상기 "기타소득(원천세 징수)*"과 동일
	이자소득 (종합소득세 신고, 원천세 징수)	채권 또는 증권의 이자와 할인액, 예금의 이자 등
	배당소득 (종합소득세 신고, 원천세 징수)	법인으로부터 받는 이익이나 잉여여금의 배당 또는 분배금, 배당으로 간주 또는 처분되는 소득 등
	연금소득 (종합소득세 신고, 원천세 징수)	이 중 공적연금 이외의 연금(사적연금)

자료: 국세청 홈페이지 국세신고안내(<https://www.nts.go.kr>); TASIS 국세통계포털(<https://tasis.nts.go.kr>)을 재구성

이처럼 국제 과세자료는 첫째, 하나의 변수에 사회조사에서는 분리되어 측정될 여러 원천별 소득이 혼재되어 있으며, 둘째, 원천징수의무자가 신고한 원천징수 부문에도 소득이 있으면서도 종합소득세 신고 부문에도 소득이 있다는 특성이 있다. 첫 번째 특성은 구인회 외(2021)와 이현주 외(2020)에서 제공받은 국세청 자료의 범위 안에서 해결할 수 없었다. 이에 대해서는 연구의 마지막 장에서 국제 과세자료의 향후 가능성과 제언에서 다루고자 한다. 두 번째 특성은 자료를 대하는 연구자가 일정 정도 다룰 수 있는 내용으로, 국제 과세자료를 이용하여 소득 변수를 생성하는 작업의 대부분이 이와 관련되어 있다. 이에 대해서는 3장의 소득 생성 과정에서 다루었다.

### 3. 국제 과세 미시자료를 이용한 소득 구성 사례

#### 1) 국제 과세자료의 특성

##### (1) 국세통계연보의 집계자료 및 국세 신고 분야

국세청이 매년 발간하는 국세통계연보는 다음의 세목으로 나누어 집계자료를 공표하고 있다(〈표 4〉). 최근에는 국세통계포털(TASIS)에서도 통계연보와 동일한 통계표를 열람하거나 내

려받을 수 있다.

**[표 4] 국세통계연보 집계자료의 분야: 2021년 연보 기준**

01 총괄	02 징수	03 종합소득세
04 원천세	05 양도세	06 상속증여세
07 종합부동산세	08 법인세	09 부가가치세
10 주세	11 소비자세	12 국제조세
13 세무조사	14 근로자녀장려금	15 기타

자료: 국세청(2021b).

이 중 개인의 경상소득을 구성하기 위해 우리가 관심을 가질 세목은 “03 종합소득세”, “04 원천세”이다.<sup>3)</sup> “03 종합소득세”에서는 사업·부동산소득, 근로소득, 연금소득, 기타소득, 이자소득, 배당소득이 집계되며<sup>4)</sup>, “04 원천세”는 다시 근로소득 연말정산 신고, 사업소득 연말정산 신고, 거주자의 연금소득 연말정산 신고, 퇴직소득 원천징수 신고, 거주자의 사업소득 원천징수 신고, 기타소득 원천징수 신고, 일용근로소득 지급명세서 제출 등으로 나뉘고, 사업소득, 근로소득, 연금소득, 기타소득이 집계된다. 원천세 분야와 종합소득세 신고 분야의 소득자 및 소득금액은 상호 배타가 아니다. 원천세 분야의 소득이 있는 개인이 종합소득세 신고를 할 수도 있기 때문이다. 한편 원천세 징수 분야의 소득만 있는 사람 중 일부는 소득세가 원천 징수됨으로써 신고 의무를 다하게 된다면 종합소득 신고를 할 필요가 없고, 그의 소득 자료는 원천세 부문에만 존재하게 된다. 따라서 국세 과세자료를 이용하여 개인별 소득 미시자료를 구성함에 있어 가장 관건은 서로 다른 과세자료의 출처에서 소득액과 세액이 상호 중첩되거나 어느 한편에만 정보가 있는 사례들을 일일이 구분하고 모종의 논리를 세워서 소득금액을 합산하는 것이 된다.

## (2) 국세통계센터의 미시자료 제공 목록: 국세통계센터 홈페이지

국세청은 수요자가 원하는 국세 관련 통계를 직접 분석하고 생산할 수 있도록 미시자료를 개방하고자 국세통계센터를 설립·운영하고 있다(국세청, 2021a). 상술한 바와 같이 개인정보보호법 상의 데이터결합전문기관이기도 한 국세통계센터는 구인회 외(2021)의 연구에서처럼 다기관 행정자료의 결합 작업을 수행함과 동시에 동시에 결합할 과세자료를 제공하기도 하며, 국세청이 보유·관리하는 자료를 수요가 있는 연구자에게 제공하기도 하는 것으로 알려져 있다.

국세통계센터 누리집 홈페이지(<https://datalab.nts.go.kr>)에서 제안하고 있는 미시자료 제공

3) 그 밖에 사회보장 부문에서 관심을 갖는 공적이전소득으로써 “14 근로자녀장려금”이 있겠으나, 이 연구에서는 다루지 않는다.

4) 국세통계연보 중 “3-1-2. 종합소득세 총수입금액 및 소득금액 신고 현황”에서 인원과 금액을 참고할 수 있다.

목록은 아래와 같다. 종합소득세, 원천세로 대별하고 있으며, 원천세 이하에서는 근로소득 연말정산 신고내역, 사업소득 연말정산 신고내역을 제공하고 있다. 국세통계연보에서 나타나지 않은 소득자 개인의 인구사회적 속성, 지역 정보와 함께 각 수입금액 및 소득 금액, 그리고 소득세액을 제공하고 있다. 하지만 이후 설명할 1차 소득 구성을 위해 필요한 원천별 소득 금액 몇몇이 누락되어 있기 때문에 연말정산을 하는 근로소득자의 근로소득과 사업자등록자의 사업소득을 개별적으로만 파악할 수 있는 형태이다. 2020년 기준 종합소득세는 785만여 명이, 원천세는 1,916만여 명이, 사업자 연말정산은 55만여 명이 신고하였다(국세청, 2021b).

[표 5] 국세통계센터의 자료제공목록: 국세통계센터 홈페이지 기준

구분1	구분2	연번	항목명	구분1	구분2	연번	항목명
종합 소득세	종합 소득세 신고내역	1	과세기간년도	원천세	근로 소득세 연말 정산 신고내역	1	귀속년도
		2	신고서기장의무구분코드			2	귀속년도시작일자
		3	신고서소득세신고유형코드			3	귀속년도종료일자
		4	거주자구분코드			4	신고구분상세코드
		5	내외국인구분코드			5	제출년월
		6	성별코드			6	징수의무자시도코드
		7	연령			7	소득자시도코드
		8	시도코드			8	소득자연령
		9	시군구코드			9	소득자성별코드
		10	주업종코드			10	과세대상급여금액
		11	업태코드			11	근로소득공제합계금액
		12	세무대리구분코드			12	과세대상소득금액
		13	총수입합계금액			13	기본공제합계금액
		14	부동산사업소득총수입금액			14	추가공제합계금액
		15	사업소득총수입금액			15	연금보험료공제합계금액
		16	부동산소득총수입금액			16	특별공제합계금액
		17	근로소득총수입금액			17	차감소득금액
		18	필요경비합계금액			18	종합소득세과세표준금액
		19	소득금액			19	근로소득산출세액
		20	부동산사업소득소득금액			20	세액감면합계금액
		21	사업소득금액			21	세액공제합계금액
		22	부동산소득금액			22	소득세결정세액
		23	근로소득금액			23	주민세결정세액
		24	종합소득금액			24	농어촌특별세결정세액
		25	과세소득금액			25	결정세액합계금액
		26	인적공제금액			26	징수의무자시군구코드
		27	기본공제금액			27	소득자시군구코드
		28	추가공제금액		사업 (연말 정산) 소득세	1	귀속년도
		29	국민연금보험료공제금액			2	사업소득구분코드
		30	기타연금보험료공제금액			3	시도코드
		31	특별공제합계금액			4	성별코드

	32	실제소득공제금액	신고 내역	5	연령
	33	과세표준금액		6	내외국인구분코드
	34	산출세액		7	지급금액
	35	소득세법세액공제금액		8	보험모집수입합계금액
	36	공제세액		9	방문판매수입합계금액
	37	감면세액		10	음료배달수입합계금액
	38	결정세액		11	사업소득금액
	39	가산세합계세액		12	종합소득공제합계금액
	40	추가납부세액		13	종합소득세과세표준금액
	41	총결정세액		14	사업소득연말정산산출세액
	42	기납부세액		15	합계지급금액
	43	차감환급예정세액		16	소득세결정세액
	44	차감납부예정세액		17	농어촌특별세결정세액
				18	주민세결정세액
		19	결정세액합계금액		
		20	소득세차감징수세액		
		21	농어촌특별세차감징수세액		
		22	주민세차감징수세액		
		23	차감징수세액합계금액		
		24	시군구코드		

자료: 국세통계센터 누리집 (<https://datalab.nts.go.kr>)

사회보장위원회의 사회보장 행정데이터 연구에서 사용한 국세 과세자료는 상기 제공목록에 더해 거주자의 사업소득, 일용근로소득, 기타소득이 추가되어 있었다(구인회 외, 2021). 대부분 원천세 영역에서 추가적으로 추출된 자료이며, 당시 연구진이 1차 소득을 구성하기 위해 필요한 원천별 소득 자료를 검토하여 요청한 결과이다. 2장에서 살펴본 바와 같이 근로소득과 사업소득 및 기타(부업)소득을 측정하기 위해서는 다양한 근로소득 금액을 파악할 필요가 있는데, <표 5>의 근로소득세연말정산 신고내역 10번, 과세대상급여금액은 연말정산을 하는 근로자의 근로소득으로써, 사회조사에서는 상용직 임금근로자로 분류되는 경제활동에 의한 소득금액이다.<sup>5)</sup> 때문에 그 외 근로자의 근로소득을 파악하기 위해서는 원천세 징수 영역의 일용근로소득과 원천징수대상 사업소득(거주자 사업소득), 기타소득 등이 필요하다.

### (3) 국세 과세자료의 ‘소득’ 개념과 유의할 점

다음으로 국세 과세자료를 활용함에 있어 유의하여야 할 점을 정리하였다. 이는 자료의 성격을 이해함과 동시에 이후 설명할 소득 및 소득세액 생성 작업의 사전 정보이다.

첫째, 국세 과세자료에서 통칭 ‘소득’이라 할 수 있는 정보는 ‘수입금액’, ‘지급금액’, 그리고

5) 또한 1차 소득을 구성하여 이후 재분배 효과를 분석하기 위해서는 과세대상 급여액에 더해 비과세 수입도 함께 파악할 필요가 있다.

‘소득금액’으로 표현되는데, 이들 중 원천소득별로 1차 소득 개념에 가까운 국세 과세자료의 변수가 다르다. 소득 분포뿐 아니라 재분배 효과에 관심이 있는 연구자 입장에서 소득 정보를 생성하기 위해서는 각 자료별 출처원의 특성을 감안하여 요소소득 개념에 가장 근접한 변수를 선택하여 사용해야 한다.

우선 연말정산 신고를 하는 근로자의 상시근로소득을 파악하기 위해서는 ‘근로소득 총 지급금액’이 필요하다. 근로자 연말정산에서 ‘지급금액’은 귀속 과세연도에 지급받은 세전 금액으로서 비과세소득을 포함한다. 국세청 홈페이지에 있는 <표 5>의 목록에는 ‘근로소득 총 지급금액’이 누락되어 있지만 이용이 가능한 것으로 알려져 있다. 종합소득세 신고 자료에서 근로소득을 파악하기 위해서는 ‘근로소득 총 수입금액’을 이용할 수 있다(<표 5>의 17). 근로소득을 파악하기 위해 ‘근로소득금액’을 사용하면 근로소득 공제 후의 금액을 사용하는 것으로, 세액이 부과되지 않았을 뿐 조세 제도의 재분배를 거친 후의 금액이 되며 사회조사에서 파악하고 있는 근로소득 개념과도 부합하지 않게 된다.

사업소득을 파악하기 위해서는 종합소득세 신고 자료에서 ‘사업소득금액’을 이용할 수 있다(<표 5>의 21). ‘사업소득금액’은 ‘사업소득 총 수입금액’에서 필요경비를 제한 뒤 산정하는 것으로 순소득의 개념과 거의 일치한다. 이는 사회조사에서 파악하는 사업소득에 준한다.

그 외 <표 5>의 목록에는 없지만 원천세 부분의 원천징수 대상 사업소득에 대한 간이지급명세서(거주자의 사업소득) 자료와 일용근로소득 지급명세서, 기타소득에 대한 원천징수 신고 자료에서는 원천징수의무자가 신고한 ‘총 지급금액’을 이용하면 과세 전의 수입금액을 파악할 수 있다.

둘째, 앞서 살펴본 바와 같이 원천별 소득의 구분이 사회조사의 그것과 다르다. 학술연구에서 원천별 소득은 흔히 근로소득, 사업소득, 재산소득, 기타소득으로 구분한다. 반면, 국세 과세자료의 항목 구분과 원천징수 자료는 소득세법을 따르기 때문에 학술연구의 원천별 소득 개념을 완전히 구현하기는 사실상 어렵고, 그와 근사한 수준으로 소득 정보를 구성하려면 다음과 같은 사항을 유의할 필요가 있다.

근로소득은 국세 과세자료의 근로자 연말정산 신고, 종합소득세 신고, 일용근로소득 지급명세서에서 정보를 얻을 수 있다. 연말정산 자료의 ‘근로소득 총지급액’은 연말정산을 하는 근로자의 소득만이 파악된다는 점을 유의할 필요가 있다. 선행연구에서 파악한 바, 이들 각 출처원의 근로소득 수입금액은 상호 배타가 아니며 중복되어 개인에게 부여되어 있다. 종합소득세 신고 자료에서는 ‘근로소득 총 수입금액’으로 상시적인 근로소득을 파악할 수 있으며, 연말정산만 하는 상시근로자의 경우 연말정산 신고 자료에서만 소득이 파악되지만 근로소득 외 다른 소득이 있어 종합소득을 신고하는 개인에게는 연말정산 신고 자료의 ‘근로소득 총 수입금액’과 종합소득세 신고 자료의 ‘근로소득 총 수입금액’이 동시에 존재한다. 그리고 이 경우 종합소득세

신고 자료의 총 수입금액이 연말정산의 그것을 포괄하며, 보다 정제된 자료라 할 수 있다. 만약 하나의 개체가 이 두 정보를 동시에 지니고 있다면 둘 중 하나만을 소득 구성에 활용해야 한다.

같은 논리로 일용근로소득만을 올리는 근로자는 원천징수의무자(고용주)가 신고한 수입금액만이 국세 과세자료에 존재하겠지만, 다른 소득이 있거나 연중 여러 사업장에 취업한 경우 연말정산 신고 자료 혹은 종합소득세 신고 자료에도 수입금액이 존재하게 된다. 앞서 설명한 사례와 다른 점은, 일용근로소득 지급명세 자료의 수입금액은 연말정산 신고 혹은 종합소득세 신고 출처원과 중복되지 않는다는 점이다. 과세소득과 비과세소득이 모두 잡히고 있으므로 이를 합산하여 사용할 수 있었다. 또한 비과세구간의 수입금액이라 하더라도 원천징수 의무자가 필요경비 인정 등의 이유로 국세청에 지급명세를 신고한다면 국세 과세자료에서 이들의 소득이 파악된다.

원천징수 대상 사업소득은 부가가치세가 면제되는 인적용역 등을 제공하여 얻는 소득으로, 사실상 근로소득의 개념과 가깝다. 대부분의 사회조사에서는 조사 지침상 이들의 소득을 근로소득으로 파악하고 있을 것이다. 만약 원천징수 대상 사업소득(거주자 사업소득)을 올리는 개인이 종합소득세 신고를 한다면 종합소득세 신고 자료의 '사업소득 금액'에 합산된다. 구인회 외(2021)과 이현주 외(2020)에서는 이들의 소득을 근로소득처럼 이용하지 못하고, 후술할 논리에 의해서 종합소득세 신고 부문과 대조한 뒤 결과적으로 사업소득으로 합산하였다. 만약 종합소득세 신고 자료의 사업소득을 사업자등록 여부에 따른 사업소득과 그렇지 않은 사업소득으로 구분할 수 있고, 기장의무(복식부기/간편장부)를 구분할 수 있으며, 원천징수 대상 사업소득인지 여부를 구분할 수 있다면 이들 양 소득을 분리할 수 있을 것이지만 그것이 국세청의 자료 제공 정책과 부합하는지, 혹은 기술적으로 가능한지는 명확하지 않다.

기타소득 원천징수 자료의 '총 지급금액' 또한 사회조사의 근로소득과 일정 부분 일치할 수 있다. 일례로 <한국복지패널>에서 조사하고 있는 기타근로소득이 이에 해당될 수 있으며, 반면 저작권료와 같은 지적재산에 의한 소득은 <한국복지패널>에서는 지침이 불명확하지만 <가계금융복지조사>에서는 재산소득으로써 명시하고 있다. 즉 국세청의 기타소득 총 지급액은 상기 언급한 원천징수 대상 사업소득과 마찬가지로 사회조사에서 응답자의 판단에 따라 다른 원천으로 조사·파악될 수 있다.

## 2) 국세청 과세자료로부터 원천별 소득 및 소득세액 정보 구성

이하에서는 선행연구에서 제공받은 국세 과세자료 명세의 변수를 기준으로 원천별 소득 변수와 소득세액을 생성한 방식을 소개한다. 이들 내용은 구인회 외(2021)와 이현주 외(2020)의

작업 사례를 기본으로 하며, 당시 연구의 주제와 보고서 지면 상의 이유로 보고하지 않은 내용이다.

### (1) 사회보장위원회의 사회보장 행정데이터에서 사용한 국세 과세자료

〈표 5〉의 목록처럼 신고 및 징수 영역으로 대별하면 종합소득세와 원천세로 구분되지만, 영역 이하의 세부 범주, 특히 원천세 이하의 범주를 구분하기 위해서는 용어를 달리하는 것이 이해를 도울 수 있다. 이후부터는 원천세 징수 영역 이하의 세부 분야는 DB(Database)라는 표현을 사용함으로써 각 세목을 구분하여 서술한다.

〈표 6〉은 구인회 외(2021)의 연구 당시 국세통계센터가 추출·제공한 국세 과세자료의 명세이다. 표에 기재되어 있는 변수의 설명은 국세통계센터의 기입 방식과 표현을 원용하였으며, 추출 DB의 명칭은 국세통계센터 담당자로부터 원자료의 추출 원천(source)을 국세통계연보의 구성과 세목을 기준으로 파악하여 저자가 기재한 것이다.<sup>6)</sup> 변수명은 저자가 설정한 것으로, 이해와 지칭하는 바를 명확하기 위하여 본문 중 주요 부분에서는 변수 설명의 명칭과 변수명을 함께 병기하기로 한다. 향후에도 소득을 구성하기 위한 목적으로 국세청의 과세자료를 추출한다면 아래의 자료 명세에서 큰 변화가 없을 것으로 판단한다.<sup>7)</sup>

구인회 외(2021)와 이현주 외(2020)의 연구에서 결합한 국세청의 과세자료는 〈표 6〉과 같이 종합소득세신고DB, 연말정산DB, 원천징수대상 사업소득DB, 장려세제DB, 일용소득DB, 기타소득DB로 영역이 구분되어, 각 영역 이하의 각종 수입금액, 소득금액, 세액 등이 변수로 추출되어 있었다.

아래의 자료명세 중 주요 사용 변수와 각 변수의 상세 정의를 설명하면 다음과 같다. 첫 번째로 사업소득 구성을 위해 “종합 소득\_사업소득금액(nts4)”와 “거주자사업소득\_총지급액(nts40)”을 이용하였다. “종합 소득\_사업소득금액(nts4)”은 “종합 소득\_사업소득총수입금액(nts1)”에서 필요 경비를 차감한 금액으로, 사회보장 부문의 주요 사회조사 자료 분석에서 사용하는 순소득 개념의 사업소득과 부합한다(통계청, 2021; 한국보건사회연구원, 서울대학교 사회복지연구소,

6) 당초 국세통계센터에서 제공한 원자료에서의 영역별 명칭은 “종합소득”, “근로소득”, “사업소득”, “일용 소득”, “기타 소득”이었다. 일견 국세통계센터의 명칭도 크게 무리는 없으나, 본문의 표를 기준으로 “종합소득세신고DB”의 근로소득과 “연말정산DB”의 근로소득, “종합소득세신고DB”의 사업소득과 “원천징수 대상 사업소득DB”의 사업소득이 혼동될 여지가 있고, 본문의 소득 합산 방식에 있어 변수의 출처에 따른 소득 금액이 분별될 필요가 있기에 저자가 별도 파악한 국세통계연보의 세목을 준용하여 “…… DB” 로써 지칭하였다.

7) 첫 번째 이유는 대부분의 정보보유기관이 그러하듯 정보관리 테이블에서 자료의 추출이 이루어지면 처음 추출하였던 프로세스를 보관하고 있다가 후속 작업에서 사용하는 경우가 많기 때문이다. 두 번째 이유는 당시 통상적으로 추출하지 않던 원천세 영역의 자료들에 대해서는 국세통계센터가 국세청 소관 부서의 내부적인 검토와 승인을 거쳤기에 또 다른 검토·승인이 전제되지 않는 한 다른 형태로 추출할 가능성이 거의 없기 때문이다.

2021).<sup>8)</sup> “거주자사업소득\_총지급액(nts40)”은 세금을 원천징수하기 전의 수입금액이 1차 소득 개념에 부합한다.

두 번째로 근로소득 구성을 위해서는 “근로 소득\_총지급금액(nts12)”과 “종합 소득\_근로소득총수입금액(nts3)”를 이용하였다. 국제 과세자료의 근로소득 부문에서 ‘소득’이라 함은 세율을 적용하기 위한 각종 공제를 거친 후의 과세표준액이기 때문에 1차 소득 개념에 부합하는 ‘수입’금액을 이용하였다. 후술하겠지만 종합소득세신고DB와 연말정산DB에서 중첩되는 소득금액은 중복을 최대한 지양하여 합산하였다.

세 번째로 “종합 소득\_부동산소득금액(nts5)”은 부동산사업소득으로, “일용 소득\_총지급금액(nts49)”은 일용근로소득으로 1차 소득 구성에 합산하였다. 부동산사업소득도 사업소득과 마찬가지로 필요경비를 제외한 개념의 금액을 소득으로 간주하였고, 일용근로소득 지급명세서 상의 원천징수의무자가 제출한 일용근로소득 지급명세서 상의 지급금액을 일용근로소득으로 간주하였다. “일용 소득\_총지급금액(nts49)”은 “일용 소득\_비과세금액(nts50)”을 포함한다.

네 번째로 “종합 소득\_총결정세액(nts7)”, “근로 소득\_결정세액합계금액(nts27)”, “거주자사업소득\_소득세(nts41)”, “거주자사업소득\_지방소득세(nts42)”는 해당 소득의 유무를 파악하여 소득세액 구성에 사용하였다. 또한 “일용 소득\_총부담세액(nts51)”과 “기타 소득\_총 합계세액(nts56)”을 소득세로 합산하였다. 합산한 소득세액은 선행연구에서 공적이전소득을 합산한 후 차감하여 처분가능소득을 생성하는 것에 활용하였다(구인회 외, 2021; 이현주 외, 2020).

**[표 6] 국제 과세자료 명세: 사회보장위원회의 사회보장 행정데이터 구축 당시 기준**

추출 DB	연번	변수명	변수 설명	추출 DB	연번	변수명	변수 설명
종합소득 신고 DB	1	nts1	종합 소득_사업소득총수입금액	연말정산 (근로자) DB	29	nts29	근로 소득_주현주민세기납부세액
	2	nts2	종합 소득_부동산소득총수입금액		30	nts30	근로 소득_주현농어촌특별세기납부세액
	3	nts3	종합 소득_근로소득총수입금액		31	nts31	근로 소득_주현기납부세액합계금액
	4	nts4	종합 소득_사업소득금액		32	nts32	근로 소득_종전소득세기납부세액
	5	nts5	종합 소득_부동산소득금액		33	nts33	근로 소득_종전주민세기납부세액
	6	nts6	종합 소득_근로소득금액		34	nts34	근로 소득_종전농어촌특별세기납부세액
	7	nts7	종합 소득_총결정세액		35	nts35	근로 소득_종전기납부세액합계금액
	8	nts8	종합 소득_지방소득세산출세액		36	nts36	근로 소득_소득세차감징수세액
	9	nts9	종합 소득_농어촌특별세총결정세액		37	nts37	근로 소득_주민세차감징수세액
	10	nts10	종합 소득_기납부세액		38	nts38	근로 소득_농어촌특별세차감징수세액
	11	nts11	종합 소득_지방소득세기납부세액		39	nts39	근로 소득_차감징수세액합계금액
연말정산	12	nts12	근로 소득_총지급금액	원천징수	40	nts40	거주자사업소득_총지급액

8) 통계청의 <가계금융복지조사>에서 사업소득은 애초에 “사업수입(총매출액)에서 비용(재료비, 인건비, 임대료, 이자 등)을 차감한 금액”으로 조사하고 있으며, 한국보건사회연구원 <한국복지패널조사>에서는 고용자 및 자영업자의 경우 매출액과 순소득을, 농림어업의 경우 판매수입과 비용, 순소득을 조사하면서 사업소득 합산에는 수입(매출액)에서 비용을 차감한 순소득을 이용하고 있다.



(근로자) DB	13	nts13	근로 소득_근로소득합계금액	대상사업 소득 DB	41	nts41	거주자사업소득_소득세
	14	nts14	근로 소득_종전근로소득합계금액		42	nts42	거주자사업소득_지방소득세
	15	nts15	근로 소득_비과세소득합계금액	장려 세제	43	nts43	장려세제_근로장려심사금액
	16	nts16	근로 소득_감면소득합계금액		44	nts44	장려세제_근로장려수급 대상여부
	17	nts17	근로 소득_과세대상급여금액		45	nts45	장려세제_가구원구성코드
	18	nts18	근로 소득_근로소득공제합계금액		46	nts46	장려세제_부양자녀수
	19	nts19	근로 소득_과세대상소득금액		47	nts47	장려세제_자녀장려심사금액
	20	nts20	근로 소득_특별공제합계금액		48	nts48	장려세제_자녀장려수급 대상여부
	21	nts21	근로 소득_특별세액공제합계금액	일용소득 DB	49	nts49	일용 소득_총지급금액
	22	nts22	근로 소득_특별공제표준세액공제금액		50	nts50	일용 소득_비과세금액
	23	nts23	근로 소득_세액공제합계금액		51	nts51	일용 소득_총부담세액
	24	nts24	근로 소득_소득세결정세액	기타소득 DB	52	nts52	기타 소득_총지급금액
	25	nts25	근로 소득_주민세결정세액		53	nts53	기타 소득_소득세
	26	nts26	근로 소득_농어촌특별세결정세액		54	nts54	기타 소득_지방소득세
	27	nts27	근로 소득_결정세액합계금액		55	nts55	기타 소득_농어촌특별세
	28	nts28	근로 소득_주현소득세기납부세액		56	nts56	기타 소득_총 합계세액

자료: 구인회 외(2021).

## (2) 원천별 소득 변수 생성

다음으로 원천별 소득을 구성하였던 방식을 소개한다. <표 6>의 국세 과세자료에서 상시근로소득, 사업소득, 부동산임대사업소득, 거주자사업소득, 일용근로소득, 기타소득을 파악·생성하여 합산할 수 있었다. 금융소득(이자소득, 배당소득)은 국세청의 협조 불가 방침에 따라 활용하지 못하였다.<sup>9)</sup> 최종적으로 원천별 소득은 상시근로소득, 사업소득, 부동산임대사업<sup>10)</sup>소득, 일용근로소득, 기타소득을 생성하였고, 이들을 합산한 재산소득이 없는 1차 소득을 구인회 외(2021)과 이현주 외(2021)의 연구에서 공적이전의 재분배 효과를 분석하기 위해 사용하였다.

<표 6>의 국세 과세자료를 바탕으로 선행연구에서 생성한 원천별 소득은 <표 7>의 내역과 같다. 2장에서 언급한 바와 같이, 자료의 한계 상 학술적 개념의 원천별 소득과는 명칭만 동일하고 그 내용은 다르다. 각 개념에 부합하는 수입(소득) 금액은 대부분 2개 이상의 DB를 출처로 하고 있다. 종합소득 확정 신고와 원천징수 부문에서 동시에 수입(소득)이 포착되는 한편, 그렇다고 모든 소득자의 해당 소득이 양 DB에 있는 것도 아니다. 따라서 각 출처 DB의 수입(소득) 금액의 성격과 신고하는 소득자의 중첩과 상호 배타를 파악하여 일련의 논리를 세워 금액을 합산하였다.

9) 2020년 기준 금융소득 종합과세 신고자는 178,953명이다(국세청, 2021b).

10) 국세청의 추출 자료 명세의 표현은 “부동산사업”으로 되어 있으나, 의미 전달을 명확하게 하기 위해 부동산 임대사업소득으로 기술하였다. 부동산임대사업자등록이 있는 사업자의 해당 사업소득이다.

**[표 7] 생성한 원천별 소득과 활용한 국세 과세자료**

사회조사 구분	연구에서 생성한 원천별 소득	국세 과세자료		
		수입(소득)	관련 세액	DB
근로소득	상시근로소득	근로 소득_총지급금액(nts12)	근로 소득_결정세액합계금액(nts27)	연말정산DB
	일용근로소득	종합 소득_근로소득총수입금액(nts3)	종합 소득_총결정세액(nts7)	종합소득DB
사업소득	사업소득	일용 소득_총지급금액(nts49)	일용 소득_총부담세액(nts51)	일용소득DB
		종합 소득_사업소득금액(nts4)	종합 소득_총결정세액(nts7)	종합소득DB
		종합 소득_부동산소득금액(nts5)	종합 소득_총결정세액(nts7)	종합소득DB
		거주자사업소득_총지급금액(nts40)	거주자사업소득_소득세(nts41) 거주자사업소득_지방소득세(nts42)	원천징수DB
		<i>연말정산 신고 사업자의 사업소득=수입금액×소득률 (소득률=1-단순경비율)</i>	<i>결정세액합계</i>	<i>사업자 연말정산DB</i>
기타소득	기타소득	기타 소득_총지급금액(nts52)	기타 소득_총 합계세액(nts56)	기타소득DB

주 1) 음영 표시는 당시 연구에서 활용하지 못한 것을 의미함  
 2) 재산소득은 주택임대소득, 기타소득 등에 산재되어 있고, 금융소득은 제공받지 못하여 구성하지 못하였음.

상시근로소득을 생성하였던 과정과 논리는 다음과 같다. 상시근로소득의 생성에는 <표 6>의 연말정산DB의 ‘근로소득 총지급금액(nts12)’, 종합소득세신고DB의 ‘근로소득 총수입금액(nts3)’ 변수를 사용하였다. 우선, 종합소득세신고DB와 연말정산DB 양쪽에 금액이 존재하는 경우 종합소득세신고DB의 금액만을 사용하고, 두 번째 경우로 종합소득세신고DB에만 금액이 존재하고 연말정산DB에 금액이 존재하지 않는 경우 종합소득세신고DB의 ‘근로소득 총수입금액(nts3)’을 사용하였다. 세 번째 경우에는 연말정산DB에만 금액이 존재하고 종합소득세신고DB에 금액이 존재하지 않는 것인데, 연말정산DB의 ‘근로소득 총지급금액(nts12)’을 사용하였다. 연말정산만을 하고 종합소득세를 신고하지 않는 근로자는 세 번째 경우에 해당한다. 상기 논리는 기본적으로 종합소득세 신고 자료에는 연말정산 신고를 한 근로소득이 연동되어 들어온다는 점에 착안, 종합소득세신고DB의 ‘근로소득 총수입금액(nts3)’을 우선하는 방식이다.

**[표 8] 상시근로소득 생성 로직**

상시근로소득 생성	소득(수입)금액 존재 여부	
	종합소득세신고DB 근로소득 총수입금액(nts3)	연말정산DB 근로소득 총지급금액(nts12)
상시근로소득 = 0	0	0
상시근로소득 = 0	0	.
상시근로소득 = 0	.	0
상시근로소득 = 종소DB근로소득총수입금액(nts3)	1	1

상시근로소득 = 종소DB근로소득총수입금액(nts3)	1	0
상시근로소득 = 종소DB근로소득총수입금액(nts3)	1	.
상시근로소득 = 연말DB총지급금액(nts12)	0	1
상시근로소득 = 연말DB총지급금액(nts12)	.	1

주: 0은 0원, 1은 금액이 있음, "."은 결측을 의미

사업소득은 종합소득세신고DB의 '사업소득금액(nts4)'과 원천징수대상 사업소득DB의 '거주자사업소득<sup>11)</sup> 총지급액(nts40)' 변수를 사용하였다. 그 전에 종합소득세 신고 DB의 사업소득 금액은 "사업소득총수입금액(nts1) - 필요경비" 개념인 것을 사전에 확인하였다. 기본적으로 사업소득 생성 로직에서도 종합소득세신고DB의 것을 우선하였다. 첫 번째 경우로, 종합소득세신고DB와 원천징수대상 사업소득DB에 금액이 중복 존재하는 경우 종합소득세신고DB의 금액만을 사용하였다. 두 번째 경우로 종합소득세신고DB에만 금액이 존재하고 원천징수대상 사업소득DB에 금액이 존재하지 않는 경우 종합소득세신고DB의 '사업소득금액(nts4)'을 사용하였다. 마지막 세 번째 경우로, 원천징수대상 사업소득DB에만 금액이 존재하고 종합소득세신고DB에 금액이 존재하지 않으면 원천징수대상 사업소득DB의 '총지급액(nts40)'을 사업소득 변수로 생성하였다.

[표 9] 사업소득 생성 로직

사업소득 생성	소득(수입)금액 존재 여부	
	종합소득세신고DB 사업소득금액(nts4)	원천징수대상사업소득DB 총지급액(nts40)
사업소득 = 0	0	0
사업소득 = 0	0	.
사업소득 = 0	.	0
사업소득 = 종소DB사업소득금액(nts4)	1	1
사업소득 = 종소DB사업소득금액(nts4)	1	0
사업소득 = 종소DB사업소득금액(nts4)	1	.
사업소득 = 원천징수대상사업소득DB총지급액(nts40)	0	1
사업소득 = 원천징수대상사업소득DB총지급액(nts40)	.	1

주: 0은 0원, 1은 금액이 있음, "."은 결측을 의미

11) 부가가치세가 면제되는 일정한 인적용역을 제공하는 자의 원천징수 대상 사업소득이다. 「소득세법 시행령」 제184조의 원천징수대상 사업소득으로, 「부가가치세법 시행령」 제35조제4호 및 제42조제1호바목에 따른 용역의 공급으로 발생하는 소득을 의미한다. 사업자등록이 없으며 고용관계 없이 독립된 자격으로 용역을 제공하고 일의 성과에 따라 지급받는 금액으로 사업자, 법인세 납세의무자 등 원천징수 의무자에 의해 신고가 되며, 소득을 올리는 사람이 종합소득신고를 할 수도 있고 안할 수도 있다.

그 밖에 부동산임대사업소득은 종합소득세신고DB의 '부동산소득금액(nts5)' 변수를 사용하였다. '부동산소득금액(nts5)' 또한 '부동산소득총수입금액(nts2)'에서 필요 경비를 차감한 순소득 개념임을 사전에 확인하였다. 일용소득DB의 '총지급금액(nts49)'은 앞서 설명한 바와 같이 연말정산DB와 종합소득세신고DB와 중첩되지 않기 때문에 일용근로소득 변수로써 1차 소득에 합산하였으며, 기타소득DB의 '총지급액(nts52)' 변수 또한 기타소득으로써 1차 소득 생성에 사용하였다.

### (3) 소득세액 변수 생성

소득세액은 재분배 효과를 분석하기 위해 필수적인 정보이다. 구인회 외(2021)와 이현주 외(2020)에서는 생성한 1차 소득에서 각 사회보장 정보보유기관으로부터 추출한 공적이전급여를 합산한 뒤, 제한적으로 파악한 사회보험료와 국세 과세자료로부터 생성한 소득세액을 차감하여 처분가능소득을 구성한 바 있다.

소득세액의 생성 또한 국세 과세자료의 각 DB에 있는 총결정세액을 DB간 중복을 조정하여 합산하는 방식으로 이루어졌다. DB 간 소득세액의 중복 여부는 상기 원천별 소득을 구성할 때 사용한 각 DB별 수입/소득/지급금액의 유무를 파악한 결과, 총 32개의 경우의 수가 발견되었다. 이들을 크게 네 가지의 경우로 분류한 다음 하위 케이스까지 <표10>의 로직대로 세액 변수를 생성함으로써 소득세액의 중복 합산을 지양하였다.

첫 번째, 주어진 데이터에서 종합소득세신고DB의 '근로소득 총수입금액(nts3)'이 존재하는 경우는 모두 16개의 형태를 보였다. 이들 케이스들은 '근로소득 총수입금액(nts3)'만 존재하더라도 기본적으로 종합소득세신고DB의 '총결정세액(nts7)'을 합산해야 했다. 종합소득세신고DB의 '사업소득금액(nts4)'과 '부동산(임대사업)소득금액(nts5)'이 존재하는 케이스들도 종합소득세신고DB의 '총결정세액(nts7)'이 이들 수입에 대한 결정세액을 모두 포함하기 때문에 '총결정세액(nts7)'만을 이용하면 되었다. 다만 <표 10> 첫 번째 분류의 마지막 두 개 유형과 같이, 원천징수대상 사업소득DB의 '총지급액(nts40)'이 있지만 종합소득세신고DB의 '사업소득금액(nts4)'이 없는 경우는 종합소득세 신고가 되지 않아 종합소득세 신고DB의 '총결정세액(nts7)'에 이들 수입에 대한 세액이 포함되지 않았음을 의미한다. 따라서 이들 케이스들은 종합소득세신고DB의 '총결정세액(nts7)'과 원천징수대상 사업소득DB의 세액(nts41+nts42)을 합산하였다.<sup>12)</sup>

두 번째, 종합소득세신고DB의 '근로소득 총수입금액(nts3)'이 존재하지 않고 동 DB의 '사업

12) 원천징수대상 사업소득DB의 세액은 소득세(nts41)와 지방소득세(nts42)의 합산액이다. 당시 이용한 자료에서 지방소득세액은 기타소득DB와 함께 해당 소득DB가 유일하고, 다른 DB에서는 지방소득세액이 합산된 것인지 국세만이 추출된 것인지 알 수 없었다. 정보의 한계가 있었지만 파악된 세액은 모두 합산하는 방향으로 주어진 소득세액 변수를 모두 활용하기로 하였다.

소득금액(nts4)이 존재하는 경우는 하위 케이스까지 모두 8개의 형태이다. 원천징수대상 사업 소득DB의 '총지급액(nts40)'과 종합소득세신고DB의 '부동산(임대사업)소득금액(nts5)'이 있는 개인이 종합소득세 신고를 한다면 이들 수입에 대한 세액은 종합소득세신고DB의 '총결정세액(nts7)'에 합산된다. 따라서 이들 양 소득이 있는 케이스들에서 연말정산DB의 '총지급금액(nts12)'가 존재하지 않는다면 종합소득세신고DB의 '총결정세액(nts7)'만을 소득세액으로 합산하였다. 만약, 연말정산DB의 '총지급금액(nts12)'이 존재한다면 사업소득과 부동산임대사업소득에 대해서 종합소득세 신고를 하고, 모종의 이유로 근로소득에 대해서는 신고를 하지 않은 경우로 볼 수 있기 때문에 종합소득세신고DB의 '총결정세액(nts7)'과 연말정산 DB의 '결정세액(nts27)'을 합산하여 소득세액을 생성하였다.

세 번째는 종합소득세신고DB의 '근로소득 총수입금액(nts3)'과 '사업소득금액(nts4)'이 존재하지 않고, 연말정산 DB의 '총지급금액(nts12)'가 존재하는 유형이다. 이들 중 첫 번째 유형, 다른 수입은 없이 연말정산 신고만 한 근로소득자는 연말정산 DB의 '결정세액(nts27)'만을 합산하였다. 다음으로 원천징수 사업소득DB의 '총지급액(nts40)'이 있다면 원천징수대상 사업소득DB의 세액(nts41+nts42)을 합산하고, 종합소득세신고DB의 '부동산(임대사업)소득금액(nts5)'이 있으면 종합소득세신고DB의 '총결정세액(nts7)'을 합산하였다. 또한 원천징수대상 사업소득DB의 '총지급액(nts40)'과 종합소득세신고DB의 '부동산(임대사업)소득금액(nts5)'이 모두 존재하는 경우 해당 DB의 세액을 모두 합산해주었다.

마지막으로, 종합소득세신고DB의 세 가지 수입금액이 없고, 연말정산DB에도 지급 금액이 없는 유형이 남는다. 어디에도 수입금액이 있지 않은 경우는 해당 DB의 소득세액을 합산하지 않았다. 원천징수대상 사업소득DB의 '총지급액(nts40)'이 있거나 종합소득세신고DB의 '부동산(임대사업)소득금액(nts5)'이 존재하는 경우는 동 DB의 세액을 합산하고, 양 DB 모두 수입이 있는 경우에는 양 DB의 세액을 합산하였다. 지금까지의 사례별 로직을 표로 정리하면 다음과 같다.

[표 10] 소득세액 합산 로직

분류	소득세액 생성	소득(수입)금액 존재 여부				
		중소DB 근로소득 총수입 금액 (nts3)	중소DB 사업소득 금액 (nts4)	연말DB 총지급 금액 (nts12)	원천징수 대상사업 소득DB 총지급액 (nts40)	중소DB 부동산 소득금액 (nts5)
1	소득세액 = 중소기업총결정세액(nts7)	1	1	1	1	1
		1	1	1	1	0
		1	1	1	0	1
		1	1	1	0	0
		1	1	0	1	1
		1	1	0	1	0
		1	1	0	0	1
		1	1	0	0	0
		1	0	1	1	1
		1	0	1	1	0
		1	0	1	0	1
		1	0	0	0	1
		1	0	0	0	0
1	소득세액 = 중소기업총결정세액(nts7) + 원천징수대상사업소득DB세액(nts41+nts42)	1	0	0	1	1
		1	0	0	1	0
2	소득세액 = 중소기업총결정세액(nts7)	0	1	0	1	1
		0	1	0	1	0
		0	1	0	0	1
		0	1	0	0	0
	소득세액 = 중소기업총결정세액(nts7) + 연말정산DB결정세액합계금액(nts27)	0	1	1	1	1
		0	1	1	1	0
		0	1	1	0	1
		0	1	1	0	0
3	소득세액 = 연말정산DB결정세액합계금액(nts27)	0	0	1	0	0
	소득세액 = 연말정산DB결정세액합계금액(nts27) + 원천징수대상사업소득DB세액(nts41+nts42)	0	0	1	1	0
	소득세액 = 연말정산DB결정세액합계금액(nts27) + 중소기업총결정세액(nts7)	0	0	1	0	1
	소득세액 = 연말정산DB결정세액합계금액(nts27) + 원천징수대상사업소득DB세액(nts41+nts42) + 중소기업총결정세액(nts7)	0	0	1	1	1
4	소득세액 = 0	0	0	0	0	0
	소득세액 = 원천징수대상사업소득DB세액(nts41+nts42)	0	0	0	1	0
	소득세액 = 중소기업총결정세액(nts7)	0	0	0	0	1
	소득세액 = 중소기업총결정세액(nts7) + 원천징수대상사업소득DB세액(nts41+nts42)	0	0	0	1	1

주: 0은 0원, 1은 금액이 있음을 의미

마지막으로 종합소득신고 의무가 없는 일용근로소득의 경우 원천징수로서 납세 의무가 종결되기 때문에 일용소득DB의 '총부담세액(nts51)'은 상기 과정에서 만들어진 세액에 더해 추가로 차감하는 방식으로 해당 소득에 대한 세액을 반영하였다. 또한 결정세액 중 마이너스 세액은 환급을 의미함을 사전 확인 후 처분가능소득 생성 시 그대로 더해주었다.

#### (4) 기타소득 관련 세액 차감과 소득세액 생성의 한계

소개한 소득세액 생성 작업에서 기타소득에 대한 세액은 언급이 없는데, 기타소득에 대한 세액은 자료의 한계로 인해 상기의 논리대로 차감이 불가하였기 때문이다. 당시 추출한 종합소득세신고DB 자료에는 기타소득(수입)금액이 포함되지 않았고, 기타소득 총지급액은 기타소득DB에만 존재하였다. 반면 종합소득세신고DB의 '총결정세액(nts7)'에는 기타소득에 대한 세액도 포함되어 있다. 앞서 설명한 논리대로라면 기타소득(수입)금액이 기타소득DB와 종합소득세신고DB에 있는지 확인한 후 양 DB에 모두 존재한다면 종합소득세신고DB의 세액을 우선하여 합산하고 종합소득세신고DB에 해당 소득(수입)금액이 없다면 기타소득DB의 원천징수된 세액을 합산해야 한다. 하지만 종합소득세신고DB에 기타소득(수입)금액이 추출되어 있지 않아 이 방식은 불가하였다. 종합소득세신고DB의 기타소득 유무를 알 수 없기에 세액을 합산함에 있어 한가지 방법을 선택을 하였다.

그것은 기타소득DB에서 소득이 잡힌 개인이 해당 소득을 종합소득세 신고에 포함하였을 것이라 가정하고 종합소득세신고DB의 결정세액만을 합산하고, 기타소득DB의 세액은 합산하지 않는 것이었다. 다만, 종합소득세신고DB의 근로, 사업, 부동산임대 수입금액이 없고 원천징수 대상 사업소득DB의 총지급액이나 연말정산DB의 근로소득총지급액이 있는 경우에는 기타소득DB의 원천징수세액을 합산해주었다. 이들 케이스들은 기타소득의 합계액이 연 300만원 이하이면서 다른 종류의 소득이 없거나 다른 종류의 소득에서도 종합소득 신고 의무가 없기 때문에 종합소득세신고DB에 수입금액이 없는 것으로 간주할 수 있기 때문이다.<sup>13)</sup>

다른 방법으로 기타소득을 올리는 개인이 종합소득세 신고를 하지 않았을 것이라 가정하고 기타소득DB의 원천징수된 세액만을 합산하는 것을 선택할 수도 있다. 이 경우 연간 기타소득이 300만원을 초과하거나 다른 종류의 소득이 있어 종합소득 신고를 하는 개인은 <표 6>의 자

13) 무조건 분리과세·종합과세 대상을 제외한 기타소득금액의 합계액이 300만원 이하이면서 원천징수된 경우 종합소득 과세표준에 합산할 것인지 분리과세로 납세의무를 종결할 것인지 선택 가능하다(국세청 홈페이지 국세신고안내). 이는 필요경비를 80% 기준으로 했을 때, 연 기타소득 총수입금액이 1,500만원에 해당하는데, 만약 그보다 적은 기타소득 수입을 올리는 경우라도 절세 등을 이유로 이 수입에 대해 종합소득세 신고를 한다면 과세자료에서 파악이 되지만, 신고 의무가 있는 소득자에 비해 종합소득세신고DB에서 파악될 가능성이 낮다.

료명세에서 종합소득 DB의 결정세액에서 이미 합산된 기타소득에 대한 세액이 중복 합산된다. 이 경우 매우 많은 기타소득이 있으면서 종합소득세신고를 하는 극소수의 개인 관측치에서 1차 소득에 이전소득을 합산하고 소득세를 차감한 처분가능소득을 생성할 때 소득이 마이너스가 되었고, 평균 등 처분가능소득의 대푯값에 영향을 주었다.

향후 국세 과세자료를 이용할 때, 처분가능소득을 생성하는 것을 염두한다면 종합소득세신고DB에서도 기타소득금액 자료를 추출하여 자료의 한계를 극복하는 것이 필요하다. 구인회 외(2021)와 이현주 외(2020)에서는 소득 구성을 우선하였기 때문에 소득세액 산출에 있어 사전에 면밀한 검토를 하지 못하였다는 한계를 지닌다.

#### 4. 국세 과세자료를 바탕으로 한 소득과 사회조사의 소득 비교

이번 장에서는 국세 과세자료를 바탕으로 생성한 소득 변수의 기술통계를 제시하고, 사회조사에서 측정한 소득 변수의 기술통계와 비교한다. 본 장에서 제시한 국세 과세자료의 소득 대푯값은 두 개의 시점을 기준으로 한다. 이현주 외(2020)에서 인용한 기술통계는 2017년, 구인회 외(2021)의 기술통계는 2019년이 기준 시점이다. 사회조사의 기술통계는 동일한 기준 연도의 <가계금융복지조사> 원자료를 분석한 결과이다.

<표 11>와 <표 12>에서 상단의 표는 국세 과세자료의 소득 대푯값이고, 아래는 <가계금융복지조사>의 것이다. 원천별 소득의 대푯값은 자료에서 가능한 것만을 제시하였는데, 국세 자료에서는 재산소득이라는 항목으로 소득 파악이 불가하기에 산출할 수 없었고, <가계금융복지조사>에서는 기타소득이라는 항목이 존재하지 않기에 산출하지 않았다. 물론 2장에서 설명한 바와 같이, 국세 자료에는 사업소득 안에 비정형 노동자의 근로소득이 포함되어 있고, 기타소득에는 <가계금융복지조사>에서 재산소득으로 조사하는 저작권료, 유무형 자산의 제공으로 인한 소득, 비정기적 근로소득 등이 포함되어 있다. 국세 자료의 1차 소득은 근로소득과 사업소득, 기타소득을 3장에서 설명한 과정을 거쳐 합산한 것이다. <가계금융복지조사>의 1차 소득은 근로소득과 사업소득, 재산소득을 합산한 것이다.

표의 <가계금융복지조사> 근로·사업소득은 사실상 국세 과세자료에서 만들어진 1차 소득과 측정 차원에서 가장 근접하다 판단하여 대푯값을 제시하였다. 국세 과세자료의 1차 소득이 근로소득과 사업소득의 합산액이라는 점에서는 동일하고, 기타소득에는 <가계금융복지조사>의 재산소득에 해당하는 소득이 포함되어 있기는 하지만 비정기적 근로소득까지 포함되어 있기 때문이다. 즉, 국세 자료의 1차 소득은 항목 면에서는 <가계금융복지조사>의 근로·사업소득 보



다 재산소득으로 분류되는 일부 유형의 소득을 더 많이 포괄하지만, <가계금융복지조사>의 1차 소득에 비해서는 금융자산에 의한 소득 및 과세 대상이 아닌 임대료 수입 등이 제외되어 더 적은 유형의 소득을 포괄한다.

2017년 기준 국세 자료의 1차 소득이 있는 가구의 비율을 살펴보면, 전체 표본 가구 중 약 79%가 해당 소득이 있는 것으로 나타나는 반면, 포괄 범위가 다소 작은 <가계금융복지조사>의 근로·사업소득이 있는 가구 비율은 약 88%에 이른다(<표 11>). 국세 자료의 1차 소득이 보다 많은 유형의 소득을 포괄한다는 것을 고려한다면, 적어도 9% 보다는 많은 가구의 소득이 국세 자료에서는 포착되지 않는 것으로 판단할 수 있다.<sup>14)</sup> 2019년 기준 국세 자료의 1차 소득이 있는 가구의 비율은 약 80.1%, <가계금융복지조사>의 근로·사업소득이 있는 가구 비율은 약 87.8%로 차이가 약간 작아진 것을 알 수 있다.

2017년 국세 자료의 1차 소득 평균은 연 3,686만원, 중위값은 2,232만원으로 나타났다. <가계금융복지조사>의 근로·사업소득 평균은 4,883만원, 중위값은 3,810만원으로 국세 자료의 1차 소득보다 평균은 1,197만원, 중위값은 1,578만원이 높았다. 2019년 기준 국세 자료의 1차 소득과 <가계금융복지조사> 자료의 근로·사업소득 간 차이는 평균값이 946만원, 중위값은 1,343만원으로 약간 줄어들었다. 구인회 외(2021)에서는 0을 초과하는 가구로만 한정하여 대푯값을 산출하였는데, 그에 맞추어 <가계금융복지조사> 자료를 분석한 결과, 국세 자료의 1차 소득과 <가계금융복지조사>의 근로·사업소득의 평균 차이는 640만원, 중위값은 1,024만원으로 나타났다.<sup>15)</sup>

14) 물론 이러한 차이는 4장 초반에 설명한 가구 구성의 차이와 소득 신고가 되지 않은 비공식 부문의 근로·사업소득, 농어업 소득 비과세 등 영향이 작용할 것이다. 하지만 이들 요인으로 인한 양 자료의 차이가 어느 정도인지는 향후 국세자료의 활용 범위가 보다 넓어지면 알 수 있을 것이라 기대한다.

15) 그 밖의 원천별 소득과 <가계금융복지조사>의 1차 소득 대푯값은 양 자료가 포괄하고 있는 세부 유형별 소득이 너무나 다르고, 빈곤 지표의 차이 또한 구체적으로 의미를 논하는 것이 불가하였다. 때문에 표에는 제시하였지만, 본문에는 서술하지 않는다. 표에 제시한 이들 수치는 단순 참고 이상의 용도는 가지지 못할 것으로 판단하였다.

[표 11] 국세 과세자료를 바탕으로 한 소득과 <가계금융복지조사>의 소득 기술통계(2017년 기준)

(단위: %, 만원/연)

구분	국세 과세자료					
	소득가구 비율	평균	중위값	빈곤율 (중위25%)	빈곤율 (중위 50%)	지니계수
근로소득	68.27	3,067	1,623			
사업소득	33.50	595	0			
재산소득	-	-	-			
기타소득	11.87	46	0			
1차소득 <sup>1)</sup>	79.01	3,686	2,232			
균등화 1차소득	79.01	2,293	1,536	32.36	39.46	0.591

  

구분	가계금융복지조사								
	소득가구 비율	평균	중위값	빈곤율 (중위25%)	빈곤율 (중위 50%)	지니계수			
근로소득	75.34	3,640	2,373						
사업소득	41.37	1,243	0						
재산소득	85.82	380	28						
기타소득	-	-	-						
근로·사업소득	88.36	4,883	3,810						
1차소득	96.93	5,263	4,059						
균등화 근로·사업소득	88.36	2,800	2,339				19.40	28.19	0.473
균등화 1차소득	96.93	3,034	2,471				16.96	26.77	0.463

주: 1) 이현주 외(2020)의 본문에는 “시장소득”으로 명명하고 있다.

2) 가구 소득 기준이다.

3) 소득 가구 비율은 전체 가구 중 해당 소득이 0을 초과하는 가구의 비율을 의미한다.

자료: 이현주 외(2020: 83); 통계청. 2018년 <가계금융복지조사> 원자료.

〈표 12〉 국세 과세자료를 바탕으로 한 소득과 〈가계금융복지조사〉의 소득 기술통계(2019년 기준)

(단위: %, 만원/연)

구분	국세 과세자료												
	소득가구 비율	평균	중위값	평균 (0초과)	중위값 (0초과)	인구빈곤율 (중위 40%)	인구빈곤율 (중위 50%)	가구빈곤율 (중위 40%)	가구빈곤율 (중위 50%)				
근로소득	69.04	3,301	1,920	4,781	3,563								
사업소득	32.38	539	0	1,665	527								
부동산사업소득	4.10	87	0	2,132	1,019								
재산소득	-	-	-	-	-								
기타소득	14.45	69	0	475	50								
1차소득 <sup>1)</sup>	80.08	3,996	2,497	4,991	3,540								
균등화 1차소득	80.08	2,556	1,798	3,192	2,471	28.43	31.76	34.65	37.15				
구분	가계금융복지조사												
	소득가구 비율	평균	중위값	평균 (0초과)	중위값 (0초과)	인구빈곤율 (중위 40%)	인구빈곤율 (중위 50%)	가구빈곤율 (중위 40%)	가구빈곤율 (중위 50%)				
근로소득	75.14	3,791	2,400	5,045	3,878								
사업소득	42.52	1,151	0	2,707	1,500								
부동산사업소득	-	-	-	-	-								
재산소득	86.28	417	35	484	63								
기타소득	-	-	-	-	-								
근로·사업소득	87.76	4,942	3,840	5,631	4,564								
1차소득	96.79	5,359	4,119	5,537	4,314								
균등화 근로·사업소득	87.76	2,882	2,403	3,284	2,771					19.47	23.72	25.94	29.55
균등화 1차소득	96.79	3,144	2,585	3,248	2,678					17.89	22.26	23.95	28.00

주: 1) 구인회 외(2021)에서는 "시장소득"으로 명명하고 있다.

2) 가구소득 기준이다.

3) 소득 가구 비율은 전체 가구 중 해당 소득이 0을 초과하는 가구의 비율을 의미한다.

4) "0 초과"는 해당 소득이 0을 초과하는 가구만을 대상으로 산출한 결과를 의미한다.

자료: 구인회 외(2021: 29); 통계청, 2020년 〈가계금융복지조사〉 원자료.

지난 2년 동안 국세 자료와 〈가계금융복지조사〉 자료의 대푯값 차이는 줄어들었다. 구인회 외(2021)과 이현주 외(2020)에서 국세 자료의 소득 정보 구성 방식은 완전히 같기 때문에, 적어도 국세 자료의 소득이 〈가계금융복지조사〉에서 파악된 소득과 근접해진 것으로 평가할 수 있다. 이는 기간 동안의 조세 제도의 과세 기반 확대, 혹은 국세청의 소득 파악의 개선 등에 의한 것으로 조심스럽게 짐작해볼 수 있겠다. 유종성 외(2021)에서도 국세청 자료에 근거한 건강보험공단의 소득 파악률이 2006년에서 2018년 사이 상당한 정도로 향상된 것을 보고하고 있다.

사실 양 자료의 결과를 비교하는 것으로써 특정한 함의를 기대하기는 어렵다. 이유는 측정과 관련된 세 가지 측면 때문이다. 첫째, 양 자료 모두 근로소득, 사업소득 등의 명칭을 사용하지

만 해당 변수가 포괄하는 실제 소득자와 금액은 앞서 2장에서 설명한 바와 같이 다르다. 둘째, 사회조사는 응답자가 조사원에게 의도적으로 숨기지 않는 한 소득 신고를 하지 않은 소득도 조사되는 반면, 국세청의 자료는 그렇지 않다. 비공식 근로 혹은 면세점 이하의 수입, 현금 거래, 의도적인 신고 기피 등에 의한 소득은 파악되지 않는다. 셋째, 가구의 구분이 다르다. <한국복지패널>과 <가계금융복지조사>와 같은 사회조사는 “생계를 함께 하는”지 여부로 가구를 규정하기 때문에 주민등록상 거주지와 완전히 일치하지 않는 반면, 구인회 외(2021)과 이현주 외(2020)의 가구는 행정안전부의 주민등록상 세대로써 가구를 규정하였다.<sup>16)</sup> 이들이 복합적으로 작용하여 양 자료 간 결과의 차이를 야기한다. 당시의 원자료를 사용할 수 없는 상황에서 각 자료를 구분하여 파악하는 것은 이 연구에서 불가능하다.

그럼에도 불구하고, 양 자료의 소득을 비교하는 것은 현재의 행정자료를 이용한 소득 정보의 위상을 기록하는 차원에서 의미가 있을 것으로 판단한다. 본 연구에서 적시한 측면들의 쟁점이 향후 연구를 거듭하며 하나씩 정리되는 동안 지금과 같이 차이를 살펴보는 작업을 반복한다면 개선된 자료의 성격과 특성을 더욱 잘 이해할 수 있을 것이기 때문이다.

## 5. 결론: 국세 과세자료를 이용한 소득 정보의 활용 가능성과 제언

이 연구는 사회과학 연구에서 필수적인 소득 정보를 최근 주목받고 있는 행정자료를 이용하여 구성하는 방법을 고찰하였다. 국세청의 국세 과세자료를 이용한 소득 정보 생성 작업의 사례를 공유하는 것을 주된 목적으로 이 연구는 다음의 내용을 다루었다.

우선, 기존 학술 연구에서 준용하는 원천별 소득의 개념을 바탕으로 국세 자료에서 얻을 수 있는 소득 정보의 특성을 검토하였다. 이를 통해 국세 과세 원자료 자체와 선행연구에서 취하였던 소득 구성 방식의 이해를 도모하였다. 학술적 개념을 준용한 사회조사의 원천별 소득 측정과 국세 과세자료가 집계하는 소득은 그 명칭은 같으나, 내용적으로 일치하지 않는 것을 검토 결과 알 수 있었다. 따라서 국세 과세자료의 원자료를 이용하거나, 원자료로부터 생성한 소득 변수를 이용할 때 이들 자료의 특성을 사전에 이해함이 필요하다.

다음으로는 국세 과세자료로부터 소득과 소득세액을 구성한 선행연구의 작업을 소개하였다. 통상적으로 사용하던 사회조사의 원천별 소득과 차이가 있음에도 불구하고, 국세청의 각 DB에서 추출한 과세 정보의 중복을 최대한 지양하여 국세 자료만의 원천별 소득을 생성하고

16) 구인회 외(2021)와 이현주 외(2020)의 행정자료 추출 당시에는 개인 관측치들을 사회조사에서 정의하는 가구 개념대로 묶을 수 없었다. 표본 추출 단위가 가구여야 가능한데, 이를 충족하는 행정자료가 없었기에 가장 근접한 추출 단위인 세대를 사용하였던 것이다.

그를 합산하여 결과적으로는 사회조사의 근로 및 사업소득, 1차 소득의 개념과 부합하는 소득 변수를 생성하는 과정을 공유하였다. 또한 재분배 효과를 알기 위해 필수적인 소득세액 변수 산출의 실무적 과정도 소개하였다. 그리고 선행연구에서 만들어진 국세 자료 기반의 소득 정보와 사회조사 자료의 소득 정보의 기술통계치를 비교하였다.

이상의 고찰 과정에서 행정자료를 이용한 소득 정보 생성의 가능성과 한계가 동시에 드러났다. 지금까지 활용한 국세 과세자료만을 이용한다면, 한계로써 제기된 쟁점들은 해결이 어렵다. 다행히도 국세 과세자료의 한계를 극복하고 활용 가능성을 보다 높이기 위해서 필요한 조치는 이용 국세 자료의 범위를 넓히는 것으로 일정 정도 가능한데, 이에 대해 간략하게 논의함으로써 결론을 대신하고자 한다.

대다수의 정보보유기관은 자신들이 공표하는 집계자료 항목에 한하여 미시자료를 제공하는 경향이 있는데, 국세청도 크게 다르지 않을 것으로 판단한다. 하지만 향후 사회보장 행정데이터의 소득 정보의 정확성과 연구 활용도를 높이기 위해서는 연구자 스스로 자료를 충분히 이해한 뒤 추가 요청하는 자료 항목마다 구체적인 이용 목적과 활용의 가능성을 제시하면서 적극적으로 협의할 필요가 있다. 본 연구에서 다른 선행연구의 경험을 토대로, 향후 국세 과세자료 활용을 위한 제언을 첫째, 생성한 소득 정보의 정확성 제고의 차원, 둘째, 보다 풍부한 연구를 위한 가능성의 차원으로 나누었다.

생성 소득 정보의 정확성 제고 차원에서 첫째, 가급적 모든 자료에서 원천세 이하의 DB와 종합소득세신고DB의 수입(소득) 금액과 세액이 중복으로 추출될 필요가 있다. 원천세 이하 DB에 관련 금액이 있다 하여 종합소득세신고DB에서 추출하지 않는 것을 지양해야 한다. 연구의 3장에서 논의한 것처럼 소득세액 산출을 위해서는 각 세부 소득의 존재 여부로 세액 차감의 논리를 완성해야 했는데, 추출 작업의 효율을 우선하여 종합소득세신고DB에 각 원천별 소득의 모든 내역이 포함되지 않았고, 이는 특히 기타소득의 세액에서 문제가 되었다. 국세 과세자료 안의 소득 금액은 DB간 상호배타성이 확보되지 않기 때문에 양쪽의 DB로부터 금액을 추출하여 연구자가 상호 대조하면서 소득(세) 합산을 해야 한다. 선행연구에서는 이 점이 완벽하지 않았고, 기타소득에 의한 세액의 합산은 결국 검증할 수 없는 가정하에 이루어질 수 밖에 없었다. 종합소득세신고DB와 원천세 이하 DB의 해당 소득이 중복 추출되어 있지 않으면 발생할 수 있는 사안에 대해서는 각주로 부연하였다.<sup>17)</sup>

두 번째, 사업자 연말정산 신고 관련 자료를 추가로 이용해야 한다. 연말정산 대상 사업소득자는 간편장부기입 대상 사업소득자<sup>18)</sup> 중 보험모집인, 방문판매, 음료배달판매원 등 독립된 자격으로 용역을 제공하고 실적에 따라 수당을 받는 사업자의 수입금액에 대하여 원천징수의무

자가 신고하는 것으로 2020년 기준 55만여 명이 대상이었다(국세청, 2021b). 지난 두 차례 선행연구의 자료 구축 당시 누락된 정보이다. 만약 사업소득 연말정산 신고 과세 자료를 이용하지 못한다면 해당 사업소득만 올리면서 종합소득세 신고를 하지 않는 비정형노동자의 소득은 체계적으로 누락될 가능성이 있다.

세 번째, 종합소득세 신고 부문과 원천세 부문에서 금융(이자·배당)소득 관련 자료를 이용해야 한다. 이는 사회조사에서는 재산소득에 해당하며, 2020년 종합소득세 신고 현황에 따르면 이자소득을 신고한 인원은 17만여 명, 배당소득은 16.5만여 명이었다. 1차소득을 최대한 구현하기 위해서 필요한 정보이다. 마지막으로, 종합소득세 신고 연금소득과 퇴직소득 관련 금액을 추가로 받아 타 정보보유기관의 자료를 결합하여 공적연금 수급액과 교차 확인을 하고, 소득세액 계산의 정확성을 높일 필요가 있다.

다음으로 국세 과세자료의 가능성 차원의 제언이다. 사회조사의 소득 정보도 그러하지만, 국세 과세자료의 소득 정보는 개인의 경제활동을 가늠할 수 있다는 장점이 있다. 근로소득과 일용근로소득, 사업소득 등으로 경제활동 상태를 알 수 있는 것은 당연하며, 명칭으로는 사업소득이지만 인적용역을 제공하여 벌어들이는 '원천징수 대상 사업소득'과 관련 속성 정보를 통해 이른바 비정형 노동자를 특정할 수 있다.<sup>19)</sup> 물론 국세 자료 중 기타소득에도 이러한 노동자의 수입금액이 포함되기 때문에, 기타소득의 업종 정보와 원천징수 대상 사업소득의 업종 정보를 속성자로 함께 제공받고, 사업자등록 여부 정보 등을 함께 제공 받아 가공하면 사회조사 이상으로 비정형노동자를 특정하여 분석할 수 있다. 이렇게 특정한 자료에 구인회 외(2021)과 이현주 외(2020)의 연구처럼 각종 사회보장제도의 수급 및 급여 정보가 제공된다면 비정형노동자의 사회보장 수급 지위나 노후소득보장 현황과 같은 분석이 가능할 것이다.

17) 아래의 표는 종합소득세신고DB에 세부 소득이 구분되어 추출되지 않는 경우를 상정하여 발생할 수 있는 문제를 요약한 것이다. 종합소득세 부문에서 결정세액은 모든 종합소득에 대해 합계액으로 산출된다. 따라서 원천세 부문에서 개별 소득에 대해 징수된 세액이 있어야 소득세의 중복 차감을 방지할 수 있다.

종합소득세신고DB	원천세 이하 DB	판단
(자료에서는 모르지만) 1	1	세액을 어디에서 차감해야 하는지 판단 불가 - 만약 종소 것을 취한다면 제대로 세액 차감 - 원천세 것을 취한다면 세액의 중복 차감
(자료에서는 모르지만) 1	0	관련 소득이 잡히지 않음 종소 DB의 결정세액을 차감하므로 소득 없이 세액만 차감
(자료에서는 모르지만) 0	1	세액을 어디에서 차감해야 하는지 판단 불가 - 만약 종소 것을 취한다면 세액 차감이 안됨 - 원천세 것을 취한다면 제대로 세액 차감
(자료에서는 모르지만) 0	0	판단 불필요

주: 0은 0원, 1은 금액이 있음을 의미

18) 해당 과세기간 신규 사업 개시 또는 해당 사업소득만 있는 자로서 직전 과세기간 수입금액이 업종에 따라 3억 원 혹은 1억 5천만 원 혹은 7천 500만 원 미만인 자(국세청 홈페이지 참조)

19) 장혜영 의원실에서는 국세청으로부터 '2020년 귀속 국세청 인적용역 업종별 사업소득 원천징수 현황' 자료를 제공받아 비임금 노동자 수를 산출하기도 하였다.(경향신문, 2022.2.7.)

그 밖의 어쩔 수 없는 한계로써 국세 과세자료의 특성이 있다. 국세 과세자료는 면세점 이하 혹은 비공식 근로에 의한 소득은 파악되지 않는다. 또한 현행 세법상 농어업 사업소득의 비과세범위가 상당히 크고, 해당 수입금액 신고에 대한 규정이 없어 수입금액이 신고되지 않는 경우가 많다(박주철 외, 2021). 이러한 이유로 누락되는 소득이 적지 않기 때문에 소득분배지표 생산과 그의 해석에 주의가 필요하다. 구인회 외(2021)와 이현주 외(2020)에서 시장소득 10분위로 표본을 구분하였을 때, 소득 하위 1, 2분위가 구분되지 않는 현상이 발생하였고, 빈곤율 또한 사회조사의 그것과 매우 큰 차이를 보이며 높게 나타났다. 이는 본문에서도 설명하였지만, 당시 사용한 과세자료의 범위의 한계가 상당히 작용하였다. 향후 이용할 수 있는 국세 과세자료의 범위가 국세청이 보유하고 있는 모든 과세자료로 확대된다면 기존의 차이가 감소함과 동시에 남아 있는 차이의 원인 또한 명확해질 것으로 기대한다. 아울러 국세청에서 선제적으로 전 국민 과세자료를 바탕으로 보다 활용이 편리한 형태로 소득 정보를 체계적으로 구축하는 것이 중요하다. 또한 가구 소득 구성은 국세청의 단일 자료로는 불가하다는 한계가 있다. 따라서 국세 과세자료 단일자료만을 이용하는 것을 넘어 행정자료간 연계가 반드시 필요하다. 국세 과세자료의 바람직한 활용을 위해서는 타 정보보유기관과의 연계 및 결합에 보다 전향적인 자세가 중요할 것으로 보인다.<sup>20)</sup>

정부가 수집·관리하는 행정자료 중 전 국민의 소득을 파악할 수 있는 원천으로는 국세청이 유일하다. 처음부터 연구 활용을 상정하고 수집된 사회조사 자료와 달리, 행정자료는 정보보유기관이 그것을 보유·관리해야만 하는 본래의 용도에 따라 범위와 형태가 결정된다. 국세청은 소득 관련 자료를 4대 보험 징수 혹은 자산조사 등의 목적으로 다른 정보보유기관으로 연계·제공하기는 하지만<sup>21)</sup>, 제공받는 기관의 용처에 따라 국세 과세자료 중 일부만을 제공한다.<sup>22)</sup> 정보보유기관의 본래 목적에 의해 연구자가 이용할 수 있는 행정자료의 범위와 형태도 결정되기에, 전 국민을 대상으로 소득을 파악하고 과세를 하기 위한 목적을 지닌 국세청의 과세자료는 전 국민의 소득을 알 수 있는 사실상 유일한 행정자료 원천이다. 즉, 국세청의 과세자료의 수집과 관리는 사실상 전 국민을 모집단으로 하여 소득 정보를 알고자 하는 사회조사와 그 목적에

20) 가령 미국 국세청의 전국민 소득-세금 패널데이터(SOI Databank) 구축(Chetty et al., 2018)이나 스웨덴 통계청의 소득 및 과세등록부 자료 연계의 사례처럼 한국의 국세청도 시점별 데이터 구축과 타 자료와의 연계에 더욱 적극적으로 임할 필요가 있을 것이다. 해외 과세자료의 구축과 자료 연계 사례는 유종성(2022)의 연구에 자세히 소개되어 있다.

21) 대표적으로 4대보험 통합 징수를 하는 국민건강보험공단, 사회복지급여의 수급 자격을 관리하는 한국사회보장정보원이 있다. 특히 국민건강보험공단은 국세청으로부터 연계·제공된 일부 과세자료로부터 2003년 귀속 소득부터 전 국민의 소득 정보를 가공한 것으로 알려져 있다. 이에 대한 자세한 내용은 유종성 외(2021)를 참고할 수 있다.

22) 예를 들어 사회보험료 징수를 위한 목적이기 때문에 비과세소득과 일용근로소득은 제공되지 않는다거나, 제도의 수급 자격을 판별하기 위한 것이기 때문에 전 국민 중 수급 대상 인구에 한하여 정보가 제공되는 식이다.

서 동일하다. 그리고 국세 과세자료의 활용 가능성은 선행연구에서 보여준 것 이상이다. 향후 동 자료에 대한 학계의 이해도가 높아짐을 바탕으로 정보보유기관에게 구체적인 연구 수요가 전달되고, 보다 활발한 자료 활용이 이루어지기를 기대한다.



## ■ 참고문헌 ■

- 강신욱, 노대명, 류정희, 이현주, 정해식, 황도경, 박형준. (2016). 맞춤형 급여제도 개편에 따른 차상위계층 지원제도 개선 방안. 보건복지부·한국보건사회연구원.
- 강신욱, 임완섭, 정해식, 강병구, 박찬임, 이건범, 박형준, 이병재. (2017). 사회보장제도 생계지원분야 기본 평가. 보건복지부·한국보건사회연구원.
- 구인회, 강상경, 김용득, 홍백의, 정찬미. (2020). 사회보장 행정데이터의 효과적 활용 및 발전 방안 연구. 보건복지부·서울대학교 산학협력단.
- 구인회, 강성호, 박형준, 손병돈, 우해봉, 이원진, 최옥금, 함선유, 한경훈, 이지완. (2021). 행정데이터를 활용한 다층노후소득보장체계 심층분석. 보건복지부·서울대학교 산학협력단.
- 국세청. (2021a). 국세 데이터 활용 안내서. 국세청.
- \_\_\_\_\_. (2021b). 2020년 국세통계연보. 국세청.
- 국세청 국세신고안내 <https://www.nts.go.kr>
- 국세청 국세통계포털(TASIS) <https://tasis.nts.go.kr>
- 국세통계센터 누리집 <https://datalab.nts.go.kr>
- 박주철, 한정희, 천정애, 김용관, 이영훈. (2021). 농어업 사업소득 비과세제도의 문제점과 개선방안. 세무회계연구 68: 135-152.
- 박찬용, 강석훈, 김태완. (2002). 소득분배와 빈곤동향 및 변화요인분석. 한국보건사회연구원.
- 여유진, 김미곤, 김태완, 신현웅, 김문길, 정해식, 여나금, 이원진, 우선희, 유아마 아쓰시. (2020). 제2차 국민기초생활보장제도 기본계획 및 종합계획 수립 연구. 보건복지부·한국보건사회연구원.
- 여유진, 김미곤, 김태완, 신현웅, 정해식, 김문길, 황도경, 이주미, 최준영, 박형준, 여나금. (2017). 제1차 국민기초생활보장제도 기본계획 및 종합계획 수립 연구. 보건복지부·한국보건사회연구원.
- 여유진, 김미곤, 김태완, 양시현, 최현수. (2005). 빈곤과 불평등의 동향 및 요인분해. 한국보건사회연구원.
- 유경준, 김대일. (2003). 소득분배 국제비교와 빈곤연구. 한국개발연구원.
- 유종성, 김민혜, 김승연, 유수진. (2021). 소득분배 연구를 위한 건보공단 빅데이터의 의의와 한계: 서울시 사례연구를 중심으로. 한국사회정책 28(3): 75-105.
- 유종성. (2022). 사회보장 행정데이터 활용연구의 초기성과와 한계 및 과제. 한국사회정책학회 춘계학술대회 자료집. 48-63.
- 이원진, 정해식, 전지현. (2019). 소득조사 마이크로데이터 비교 분석 - <가계동향조사>와 <가계금융·복지조사>를 중심으로. 한국보건사회연구원.
- 이현주, 강희정, 류정희, 손병돈, 이재원, 박형준, 김성아. (2018). 사회보장제도의 소득·서비스보장 현황과 정책과제 연구. 보건복지부·한국보건사회연구원.
- 이현주, 오욱찬, 이윤경, 이원진, 성재민, 이길재, 박형준, 이병재. (2020). 사회보장정책 효과성 분석을 위한 행정데이터 연계·활용 방안. 보건복지부·한국보건사회연구원.
- 통계청. (2019). 가계금융복지조사 통계정보 보고서. 통계청.
- \_\_\_\_\_. (2021). 2021년 가계금융복지조사 결과 보도자료. 통계청.
- \_\_\_\_\_. 2018년, 2020년 가계금융복지조사 원자료.

\_\_\_\_\_. 2020년 가계금융복지조사 조사표.

한국보건사회연구원, 서울대학교 사회복지연구소 (2021). 한국복지패널 15차년도 조사자료 유저가이드.

한국보건사회연구원. 제15차 한국복지패널 가구 조사표.

한은희, 윤상용, 정은희, 김정현, 박형준, 이우식, 김인수, 최솔지, 박규범, 권하늬. (2020). 사회보장정보 시스템을 활용한 가구유형별 복지사각지대 발굴 방안. 보건복지부·한국사회보장정보원.

“코로나로 플랫폼·특수고용 등 급증…노동법 밖 ‘비임금 노동자’ 700만명”. 경향신문. (2022.2.7.)

Chetty, R., Friedman, J. N., Saez, E. & Yagan, D. (2018). The SOI Databank: 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

OECD (2017). *Terms of Reference: OECD Project on The Distribution of Household Incomes*. OECD Income Distribution Database (<http://oe.cd/idd>).

OECD (2012). *Terms of Reference: OECD Project on The Distribution of Household Incomes*. OECD Income Distribution Database (<http://oe.cd/idd>).

Abstract

## Using Administrative Data to Compose Income Information: Focused on National Tax Microdata

Hyungjohn Park\*

This study aims to examine the method of composing income information that could be used for social policy research by using microdata on national taxation. As a research method, the characteristics of national tax microdata were identified by comparing income measurement of the Korea Welfare Panel Study (KOWEPS) and the Survey of Household Finances and Living Conditions (SFLC) with the definition of income under the tax law. The experience of calculating income and income tax in previous studies using national tax microdata was introduced. In addition, income from national tax data and income from SFLC were compared based on descriptive statistics. As a result, although the name itself is the same for income by source of national tax microdata and social survey, there was some difference between the measurement of social survey and the tax law definition of national tax microdata. And it was found that the previous study created some income variables on the assumptions that could not be verified due to data limitations. This acted as one of the reasons for the difference in descriptive statistics between income from social surveys and income from national tax data. However, most of the limitations found in the experience of using national tax microdata could be dealt with by the researcher by broadening the scope of use of national tax data. Therefore, I discussed the data that should be additionally extracted and used for income studies using national tax microdata in the future. In addition, a method for using data that could identify irregular workers as the possibility of national taxation data for social policy research was suggested.

**Keywords:** national tax service, income by source, income tax, withholding tax, social survey

◆ 2022. 4. 28. 접수 / 2022. 6. 8. 1차수정 / 2022. 6. 21. 게재확정

---

\* Research Fellow, Policy Research Department, Sejong Public Agency for Social Service (sheeppig@sj.pass.or.kr)