

우리나라 인구이동 데이터의 다양성과 상호보완성에 관한 연구 - 주민등록과 등록센서스 데이터의 비교를 중심으로 -

이상일*·조대현**

The Variety and Complementarity of Migration Data in South Korea: A Comparison of Resident Register and Register-Based Census Data

Sang-Il Lee* · Deaheon Cho**

요약 : 본 논문의 주된 연구목적은 우리나라의 주민등록 인구이동 데이터와 등록센서스 인구이동 데이터를 비교하여 상대적인 장단점과 상호보완성을 탐색하는 것이다. 이를 위해 우선적으로 인구이동 데이터 분류 범주에 대한 개념적 논의를 진행하였고, 분류 범주를 크게 수집 프레임워크, 시간 프레임워크, 공간 프레임워크, 속성 프레임워크에 의거해 정리하였다. 우리나라의 세 가지의 서로 다른 인구이동 데이터(전통센서스, 주민등록, 등록센서스)의 특성을 이러한 분류 범주에 의거해 비교하였다. 이들 중 주민등록 데이터와 등록센서스 데이터를 선정하여, 총이동량 및 총이동률, 이동자의 인구구조, 시공간 전출입 플로, 지역별 순이동의 측면에서 면밀한 비교 분석을 실시하였다. 이를 통해 인구추계에 있어서 등록센서스 데이터가 주민등록 데이터를 위한 보완적 혹은 대체적 원천으로서 어느 정도의 가능성이 있는지를 검토하고자 했다. 두 데이터 원천은 인구이동 포착성이라는 측면에서 상대적인 강점과 약점을 동시에 가지고 있으며 상호보완적인 관계에 있는 것으로 드러났다. 특히, 등록센서스 데이터는 주민등록 데이터가 가지고 있는 젊은 층의 이동성에 대한 과소추정 문제를 해결하는데 상당한 도움을 줄 수 있다는 측면에서, 인구추계를 위한 보완적 데이터로서의 가능성이 충분히 큰 것으로 판단되었다.

주요어 : 인구이동 데이터, 등록센서스, 주민등록, 인구추계

Abstract : The primary objective of this study is to explore the relative strengths, weaknesses, and complementarities of two migration data sources in South Korea: the resident register migration data and the register-based census migration data. To achieve this, we first conducted a conceptual discussion on the classification categories of migration data, organizing these categories into the collection, temporal, spatial, and attribute frameworks. We analyzed the characteristics of three different migration data sets in South Korea (traditional census, resident register, and register-based census) using these classification categories. For a detailed comparative analysis, we focused on the resident register data and the register-based census data, examining total migration volume and rate, demographic structure of migrants, inter-regional migration flows, and net migration by region. Our analysis aimed to assess the extent to which register-based census data can complement or serve as an alternative to resident register data for population projections. The study found that both data sources have unique strengths and weaknesses in capturing population mobility and can be complementary. Notably, register-based census data provides significant value as a supplementary source for population projections, particularly in addressing the underestimation of youth migration in resident register data.

Key Words : migration data, resident register, register-based census, population projection

이 논문은 2022년 통계청의 지원을 받아 수행된 『지역별 국내인구이동추계 작성방법 개선방안 연구』의 연구보고서(이상일 등, 2022) 일부를 수정·보완한 것이다.

* 서울대학교 지리교육과 교수(Professor, Department of Geography Education, Seoul National University, si_lee@snu.ac.kr)

** 경북대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Kyungpook National University, dlhcho@knu.ac.kr)

1. 서론

공간적 상호작용 데이터의 특수한 형태로서의 인구이동 데이터는 다른 공간적 데이터와는 구별되는 고유한 특성을 가지고 있으며(이상일, 2012; Bailey and Gatrell, 1995; Stillwell *et al.*, 2010), 그러한 특수성으로 말미암아 인구이동 데이터의 관리, 분석, 시각화 등의 측면에서 다양한 난제가 존재한다는 것은 널리 알려진 사실이다(이상일·김현미, 2021; 이상일·이소영, 2021). 그런데 이러한 특수성의 또다른 측면은 인구이동 데이터가 매우 다양한 형태로 존재한다는 점이다. 이러한 다양성의 궁극적인 이유는 인구이동이 다른 인구학적 요소들(출산력과 사망력)에 비해 개념과 실질적인 측정 사이의 간극이 비교적 넓기 때문이다. 인구이동은 보통 “인구의 통상적 거주지(usual residence) 상의 위치 변동” 정도로 정의되는데(Carmichael, 2016; Rees, 1977; Stillwell *et al.*, 2010), 거주지의 통상성을 어떻게 규정할지(Bell *et al.*, 2002), 어느 정도의 전위를 인구이동으로 간주할지(Stillwell *et al.*, 2010)와 같이 측정에 앞서 해소해야 할 개념적 모호성이 존재한다.

그러나 다양성의 보다 실질적인 이유는, 어떻게 측정할 것인가와 관련된 고려 사항 및 제약 조건이 상대적으로 복잡해 측정의 프레임워크를 표준화하기 어렵다는데 있다. 인구이동의 국제 비교를 행한 IMAGE(Internal Migration Around the Globe) 프로젝트는 인구이동 데이터가 국가 간에 매우 다양하게 존재함을 보고하고 있다(Bell *et al.*, 2002, 2015a; 2015b; 2020). 어떤 국가는 센서스의 조사 항목을 통해 인구이동 데이터를 얻지만, 어떤 국가는 행정자료나 전국 단위 서베이(survey)를 통해 인구이동 데이터를 얻는다. 또 어떤 국가는 인구이동이 발생한 때 시점을 포착하여 인구이동 데이터를 구축하지만, 어떤 국가는 특정 두 시점 사이(1년이나 5년 단위)의 거주지 변동을 기초로 하여 인구이동 데이터를 구축한다. 이외에 다양한 요소들이 전세계 국가별 인구이동 데이터의 이질성에 기여하고 있다.

그런데 인구이동 데이터의 국가간 다양성만큼 중요한 것이 단일 국가 내 인구이동 데이터의 다양성이다. IMAGE 프로젝트의 연구 결과에 따르면, 조사 대상이었던 179개국 중 61%에 해당하는 109개국이 다중 원천을 사용한다고 한다(Bell *et al.*, 2015a). 또한 많은 국가들이 인구이동의 발생 시점에 기반한 데이터 뿐만 아니라 두 시점 사이의 거주지 변동에 근거한 데이터도 수집하고 있다(Bell *et al.*, 2015a).

이런 하이브리드 방식의 인구이동 데이터 수집이 지속적으로 증가하고 있는 이유 중의 하나는 전통적인 센서스가 ‘등록센서스(register-based census)’나 ‘결합센서스(combined census)’로 전환되고 있기 때문이다(Bell *et al.*, 2015a). 등록센서스는 등록부(인구등록부, 건축물대장 등)를 포함한 다양한 행정자료를 활용하여 직접 조사 없이 센서스를 시행하는 것이고, 결합센서스는 전통적 센서스와 등록센서스를 절충하는 다양한 방식을 일컫는다(Baffour *et al.*, 2013; Coleman, 2013; Valente, 2015; UN, 2017; UNECE, 2018). 이러한 대안적 센서스의 성장은 인구이동 데이터 원천 간의 전통적인 경계를 모호하게 하고, 단일 국가 내 인구이동 데이터의 다양성을 지속적으로 강화하는 기제로 작용하고 있다.

단일 국가 내 인구이동 데이터의 다양성이라는 측면에서 우리나라는 다양성이 매우 강한 나라 중의 하나라고 볼 수 있다. 1970년 이후로 우리나라는 두 가지 인구이동 데이터의 원천이 존재했다(Lee and Kim, 2020). 하나는 ‘전통센서스’ 인구이동 데이터이고, 또 다른 하나는 ‘주민등록’ 인구이동 데이터이다. 전통센서스 데이터는 1970년 이래로 센서스의 표본부문에 조사된 데이터를 의미하는데, 2020년 센서스의 경우 20% 표본으로 1년전 거주지와 5년 거주지에 대한 데이터가 수집되어 있다(통계청, 2022). 주민등록 데이터는 전입신고서에 기반한 국내인구이동통계를 의미하는데(통계청, 2020), 현재 가장 활용성이 높은 인구이동 데이터로 인정받고 있다(Lee and Kim, 2020). 그런데, 2015년 우리나라가 등록센서스를 전면적으로 도입함에 따라 또 다른 인구이동 데이터의 원천이 생겨나게 되었다. 이를 ‘등록센서스’ 데이터라고 부를 수 있는데, 센서스 전수부문의 한 항목으로 1년 전 거주지에 대해 데이터가 수집되고 있는 것이다(통계청, 2022). 전수부문에 대한 등록센서스는 2015년 이후 매년 이루어지고 있기 때문에 현재 2016년 이후 매년 등록센서스에 의한 인구이동 데이터가 생성되고 있다.¹⁾

한 국가가 다양한 인구이동 데이터를 보유한 경우, 그들 간의 상대적인 장단점과 상호보완성을 면밀하게 따져보는 것은 매우 의미 있는 일이다. 특히, 인구이동의 강도 측정, 인구이동의 공간적 패턴 분석, 인구 추계 및 추정과 같은 보다 고차원적인 인구지리학적 연구에 중요한 함의를 제공할 수 있다. 이러한 중요성에도 불구하고 한 국가 내의 서로 다른 인구이동 데이터를 비교한 연구는 극히 드물다(예외로, Foley *et al.*, 2021). 이런 배경에서 본 논문의 주된 연구 목적은 기존의 주민등록 데이터와 새롭게 만들어지고 있

는 등록센서스 데이터를 비교하여 상대적인 장단점과 상호보완성을 탐색하는 것이다. 이를 위해 다양하게 존재하는 인구이동 데이터를 분류하기 위한 개념적인 논의를 진행하며, 이를 바탕으로 우리나라의 서로 다른 인구이동 데이터의 특성을 분석하고자 한다. 2016~2020년의 5개 연도에 대해 두 데이터 원천 간의 비교 분석을 실시하고자 하는데, 구체적으로 총이동량 및 총이동률, 이동자의 인구구조, 시공간 전출입 플로우(flow), 지역별 순이동에 집중하고자 한다. 비교 분석의 결과를 바탕으로 보다 합리적인 인구추계를 위한 두 데이터 원천 간의 상호보완성에 대한 토론을 제공하고자 한다.

2. 인구이동 데이터의 분류 범주와 우리나라 인구이동 데이터의 다양성

1) 인구이동 데이터의 분류 범주

몇몇 연구들이 인구이동 데이터를 분류하는 기준 혹은 분류 범주를 제시한 바 있다. 예를 들어, Bell *et al.* (2015a)은 데이터 원천, 데이터 유형, 인구이동 형태, 측정 간격, 지리적 구역 체계, 이동자의 특성 등 6가지의 범주를 제시한 바 있고, 다른 연구들도 이와 유사한 분류 범주를 제안한 바 있다(Bell *et al.*, 2020; Charles-Edwards *et al.*, 2019). 본 연구에서는 이러한 연구들을 바탕으로 인구이동 데이터의 분류 범주를 크게 네 가지로 구분하여 제시하고자 한다. 분류 범주는 일종의 프레임워크로 개념화될 수 있는데, '수집 프레임워크', '시간 프레임워크', '공간 프레임워크', '속성 프레임워크'로 명명하고자 한다.

수집 프레임워크는 인구이동 데이터가 수집되는 총체적인 방식을 규정하며, 세부적으로 데이터의 '원천(source)'과 데이터의 '유형(type)'으로 나뉜다. 데이터의 원천은 수집 체계를 의미하는 것으로 센서스(census) 데이터, 레지스터(register) 데이터, 서베이(survey) 데이터의 세 유형으로 나뉜다. 센서스 데이터는 5년 혹은 10년 주기의 전통적인 센서스에서 거주지 이동과 관련된 조사 항목(예: 1년전 거주지)을 통해 수집된 것을 의미한다. 레지스터는 인구등록부나 관련 행정자료 상의 거주지 이동 정보에 기반한 데이터이다. 대표적인 것으로 영국(잉글랜드와 웨일스)의 NHSCR(National Health Service Central Register)(ONS,

2016)을 들 수 있다. 서베이는 전국 단위의 표본 조사를 통해 수집된 데이터를 의미하는데 대표적인 것에 미국의 ACS(American Community Survey)(US Census Bureau, 2023)가 있다.

수집 프레임워크의 두 번째 사항은 데이터의 유형이다. 데이터 유형은 데이터 포착 방식을 의미하는데, 크게 이벤트(event) 데이터와 트랜지션(transition) 데이터로 나뉜다(Stillwell *et al.*, 2010; Bell *et al.*, 2015a). 이벤트 데이터는 인구이동 발생의 모든 시점(혹은 등록 시점)을 기록함으로써 획득되는 데이터로, 주로 레지스터 원천과 관련된다. 이에 비해 트랜지션 데이터는 주어진 두 시점 사이의 거주지 변화에 근거한 데이터로, 주로 센서스 원천과 관련된다. 그러므로 데이터 유형은 실제 발생하는 인구이동을 얼마나 잘 포착하느냐와 관련되어 있다. 즉, 이벤트는 일시점 방식이므로 데이터 포착성이 높고, 후자는 이시점 비교 방식이어서 회귀 이동(return move), 반복 혹은 순환(repeat or circular) 이동 등이 누락될 수 있어서 상대적으로 데이터 포착성이 낮다.²⁾ 트랜지션 데이터는 측정 기간(observation interval)을 어떻게 설정하느냐에 따라 세 가지 유형으로 세분화된다(Bell *et al.*, 2015a). 고정기간(fixed interval) 이동은 가장 널리 사용되는 것으로, 1년 기간 혹은 5년 기간이 주로 사용된다. 현거주지와 1년전 혹은 5년전 거주지를 비교하는 것이다. 생애(lifetime) 이동은 현거주지와 출생지를 비교한다. 직전 이동(last move)은 현거주지와 가장 최근의 거주지를 비교하는데, 보통 현거주지에서의 거주 기간(duration)이 함께 조사된다.

시간 프레임워크는 데이터의 '시간범위' 및 '시간단위'와 관련되어 있다. 시간범위는 일관성 있는 인구이동 데이터를 얼마나 소급하여 구득가능함과 관련되어 있다(Duke-Williams and Stillwell, 2010). 즉, 넓은 시간범위는 높은 데이터 소급성을 의미한다. 시간단위는 데이터의 작성 및 공표 주기와 관련되어 있는 것으로 일관성 있는 인구이동 데이터가 얼마나 빈번하게 작성되고 공표되는가를 의미한다. 이것은 시간해상도 혹은 시간적 상세성 정도로 이해할 수 있다. 전통적인 센서스 원천의 데이터인 경우, 시간범위는 다른 데이터 원천에 비해 넓은 편이고, 시간단위는 센서스 주기에 의존적이다. 레지스터의 경우는 센서스에 비해 상대적으로 시간범위는 좁고 시간단위는 촘촘한 편이다. 시간단위는 위의 트랜지션 기간과도 관련되어 있다. 즉, 1년 기간이 5년 기간에 비해 시간해상도가 높다고 할 수 있다.

공간 프레임워크는 주로 데이터의 공간단위와 관련되어

있다. 공간단위의 크기는 측정 스케일(measurement scale)을 의미하는 것으로 공간해상도 혹은 공간 상세성을 의미하는 것으로 이해할 수 있다. 인구이동 데이터에서 측정 스케일은 두 가지 측면에서 매우 중요하다(이상일·이소영, 2021). 첫째, 인구이동을 규정하는 정의 스케일(defining scale)은 총이동의 규모와 이동의 강도를 기본적으로 규정하기 때문에 매우 중요하다. 미국의 경우 거주지 변경 모두를 인구이동을 간주하기 때문에 정의 스케일은 없다. 우리나라의 경우는 읍면동 수준이, 일본의 경우는 시구정촌(市区町村)이 정의 스케일의 기능을 한다(통계청, 2021). 정의 스케일의 공간 상세성이 좋을수록 최종적인 데이터의 공간 상세성이 양호할 가능성이 높다. 둘째, 데이터가 어느 정도의 공간 상세성으로 제공되느냐, 그리고 얼마나 다양한 종류의 지역 스케일로 제공되느냐와 관련된 것이다. 우리나라의 예로, 시도, 시군구, 읍면동의 세 수준 모두에서 데이터가 제공된다면 공간 상세성의 측면에서나 지역 스케일의 다양성 측면에서나 양호한 것으로 평가할 수 있다. 그런데 전입자수와 같은 지역별 통계와 OD 매트릭스 상의 플로와 같은 지역간 통계 간에 차이가 있는데(이상일·이소영, 2021; 2023), 보통 지역별 통계에 비해 지역간 통계의 공간적 상세성이 낮다.

마지막으로, 속성 프레임워크는 인구 포괄성, 인구구조 상세성, 유관 속성 이용가능성과 관련되어 있다. 인구 포괄성은 인구이동 데이터가 전인구를 포괄하느냐 아니면 특정 집단은 누락되느냐와 관련되어 있다. 예를 들어 센서스를 통해 주로 취득되는 5년 단위 트랜지션 데이터의 경우, 조사 시점 0~4세 인구에 해당 인구이동 정보는 누락될 수밖에 없다. 인구 구조 상세성은 성별 구분이 되어 있는지, 연령이 1세 단위인지, 5세 단위인지, 가장 높은 연령층은 어떻게 규정되는지와 관련되어 있다. 만일 성별, 1세별, 0~85세 및 85세 이상 인구에 대한 인구이동 데이터가 존재한다면 인구구조 상세성이 양호하다고 말할 수 있다. 관련 속성 이용가능성은 인구이동과 관련된 다양한 속성이 얼마나 풍부하게 제공되느냐와 관련되어 있다. 예를 들어 인구이동 사유가 제공되는지, 인구이동과 관련된 사회인구학적 속성이 얼마나 풍부하게 제공되는지가 유관 속성 이용가능성의 수준을 결정한다.

표 1은 데이터 원천별로 상대적 강점, 약점, 활용성을 평가한 것이다(Bell *et al.*, 2015a). 기본적으로는 세 가지 데이터 원천을 비교한 것이지만, 강점과 약점을 평가하는데 있어 시간 프레임워크, 공간 프레임워크, 속성 프레임워크와

관련된 내용이 함께 다루어 진다. 센서스와 레지스터간의 비교가 중요한데, 주목할 사항을 몇 가지로 요약하면 다음과 같다. 첫째, 센서스(전수조사의 경우)와 레지스터 모두 상당한 공간적 상세성을 갖춘 데이터를 생성하며, 인구이동 패턴에 대한 공간 분석, 인구이동 강도 측정, 인구추계 등의 분야에 공히 활용될 수 있다. 둘째, 레지스터는 센서스에 비해 시간적 상세성이 높은 데이터를 제공하고, 수집과 공표 사이의 지연 시간이 짧아 적시성이 높은 데이터를 제공한다. 셋째, 전통적인 센서스는 설문에 의존하기 때문에 회상 오류(recall error)나 무응답으로 인한 왜곡이 불가피하다. 특히 기억의 부정확성에 기인한 회상 오류의 경우 1년 기간 보다 5년 기간 데이터에서 보다 심각한 문제가 될 수 있다.

Bell *et al.*(2015a)은 데이터 유형별 활용성에 대한 평가도 제시하였다. 중요한 사항을 요약하면 다음과 같다. 첫째, 이벤트와 트랜지션의 상대적인 장단점은 데이터 포착성과 유관 속성 이용가능성에 의존한다. 즉, 이벤트는 데이터 포착성이 높는데 반해 유관 속성 이용가능성이 낮고, 트랜지션은 그 반대이다. 따라서 이동자 선별성에 대한 심도 깊은 연구에는 트랜지션 데이터가 상대적으로 유리하다. 둘째, 데이터 포착성이 상대적으로 높은 이벤트 데이터가 이동 강도 연구에서 상대적으로 유리하다. 그러나 1년 기간 트랜지션은 5년 기간 트랜지션에 비해 누락의 가능성이 낮기 때문에 이벤트와 유사한 수준의 이동 강도를 산출할 수 있다. 셋째, 공간 패턴 연구에서는 시간적 상세성이 낮은 데이터가 보다 명확한 결과를 보여줄 수 있다. 즉, 5년 기간 트랜지션 데이터가 1년 기간 트랜지션 데이터에 비해 안정적이고 분명한 공간적 패턴을 산출할 가능성이 높다. 개별 연도가 보여주는 이례적인 패턴이 유연화되어 신뢰할 만한 인구재분포 패턴이 드러날 수 있다. 그러나 앞에서 언급한 것처럼, 회상 오류의 가능성이 높아진다. 넷째, 인구추계 및 추정적 경우, 이벤트 데이터의 활용성이 트랜지션 데이터에 비해 월등히 높다. 그러나 1년 기간 트랜지션 데이터의 경우는 예외적으로 상당한 활용성을 보여줄 수 있다.

2) 우리나라 인구이동 데이터의 다양성

현재 이용가능한 우리나라 인구이동 데이터는 모두 세 가지인데, 각각 '표본센서스', '주민등록', '등록센서스' 인구이동 데이터로 부를 수 있다.³⁾

첫째, 표본센서스 인구이동 데이터는 센서스의 표본조사를 통해 획득되는 인구이동 데이터를 의미한다. 우리나라

표 1. 인구이동 데이터 원천별 강점, 약점, 활용성

데이터 원천	강점	약점	활용성
센서스	<ul style="list-style-type: none"> - 전인구를 대상으로 함(전수조사의 경우) - 작은 공간단위(높은 공간적 상세성) - 넓은 시간범위(높은 소급성) - 높은 유관 속성 이용가능성 - 센서스 문항 변경의 유연성 - 높은 국제 표준화 가능성 	<ul style="list-style-type: none"> - 큰 시간단위(낮은 시간적 상세성) - 데이터 수집과 공표 사이의 지연 - 상대적으로 낮은 인구 포괄성(영어, 사망자, 이민자 누락 가능) - 상대적으로 낮은 데이터 포착성(회귀 이동이나 반복 이동 누락 가능) - 높은 수집 비용 - 상대적으로 낮은 데이터 개방성 - 상대적으로 낮은 데이터 신뢰성(특히, 표본조사의 경우) - 상대적으로 높은 회상 오류 및 무응답 문제 	<ul style="list-style-type: none"> - 공간 패턴 분석 - 인구이동 강도 분석 - 이동자 선별성 분석 - 과거 경향성 분석 - 인구추계
레지스터	<ul style="list-style-type: none"> - 높은 데이터 포착성 - 작은 공간단위(높은 공간적 상세성) - 작은 시간단위(높은 시간적 상세성) - 시리즈 데이터 산출 가능(주로 연 단위) - 개인 식별자를 통한 다른 데이터 원천과의 연결 가능 - 종단자료 구축 가능 	<ul style="list-style-type: none"> - 데이터 이차성의 문제(부산물로서의 인구이동 데이터) - 특정 인구 누락 가능 - 국가별로 상이한 등록 프레임워크 - 전출입 매트릭스 이용상의 제한성 - 상대적으로 낮은 유관 속성 이용가능성 - 국가별로 상이한 데이터 신뢰성 	<ul style="list-style-type: none"> - 인구추정 - 공간 패턴 분석 - 인구이동 강도 분석 - 최근 경향성 분석 - 인구추계
서베이	<ul style="list-style-type: none"> - 상세한 인구이동 이력 데이터 구축 가능 - 높은 유관 속성 이용가능성 - 인과관계 조사 가능 - 설문 문항 변경의 유연성 	<ul style="list-style-type: none"> - 표본오차 상존 - 국가별로 상이한 데이터 포맷 - 큰 공간단위(낮은 공간적 상세성) 	<ul style="list-style-type: none"> - 인구이동 강도 분석 - 이동자 선별성 분석 - 인구이동 역동성 분석

출처: Bell *et al.*, 2015a, p.10, Table 5를 수정함.

는 1970년 센서스에 최초로 인구이동에 대해 10% 표본조사가 실시되었으며, 2015년 센서스부터 표본비율이 20%로 확대되었다(통계청, 2022). 우리나라의 센서스 주기는 5년이므로 그에 맞추어 이 데이터도 5년마다가 생성되며, 조사 시점은 11월 1일이다. 표본센서스 데이터는 기본적으로 트랜지션 데이터를 제공한다. 2020년의 경우 1년 기간, 5년 기간, 생애 트랜지션 데이터를 제공하며, 1년 기간 이동자, 즉 2019~2020년간 이동자는 8,284천명이고, 이동률은 16.6%이다(통계청, 2021). 앞에서 살펴본 것처럼, 트랜지션 데이터는 인구 포괄성의 문제로 인해 1년 기간의 경우는 0세, 5년 기간의 경우는 0~4세 인구가 누락된다. 또한 데이터 포착성의 문제로 인해 레지스터 데이터에 비해 이동자수나 이동률이 과소추정되는 경향이 있다. 시간 프레임워크의 관점에서 보면, 5년에 한번씩 데이터가 생성되므로 시간적 상세성은 높지 않다고 평가할 수 있다. 그러나 1970년부터 유사한 성격의 데이터가 생성되어 왔으므로 시간적 소급성은 상대적으로 높다고 할 수 있다. 공간 프레임워크의 관점에서 보면, 다른 데이터와 마찬가지로 정의의 스케일은 읍면동 단위이며, 공표의 공간단위는 시도와 시군구이다.⁴⁾ 속성 프레임워크의 관점에서 보면, 인구 포괄성과 인구구조 상세

성은 낮고, 성, 연령, 교육정도, 혼인상태, 경제활동 상태, 종사자 지위, 산업, 직업과 같은 속성이 제공되므로 유관 속성 이용가능성은 상대적으로 높다고 평가할 수 있다.

둘째, 주민등록 인구이동 데이터는 주민등록 전입신고에 의거한 국내인구이동통계를 의미하는 것으로(통계청, 2020), 전형적인 이벤트 데이터에 속한다. 이동자가 작성한 전입신고서는 전산망을 통해 중앙주민전산망센터에 보고되고, 월단위 집계를 통해 국내인구이동통계가 작성된다(통계청, 2020). 우리나라에서 가장 적시성이 높은 인구이동 데이터로 가장 널리 활용되고 있는 데이터라고 할 수 있다. 시간 프레임워크의 관점에서 보면, 매월 데이터가 생성 및 공표되므로 시간적 상세성은 매우 높다고 평가할 수 있다. 연단위 측정 시점은 매년 12월 31일이다. 2020년 한 해 동안 총 7,735,491명이 이동해 이동률은 15.1%였다. 그리고 시도 단위의 경우 1970년부터 유사한 성격의 데이터가 생성되어 왔으므로 시간적 소급성은 상대적으로 높다고 할 수 있다. 그러나 시군구 단위는 1995년부터 이용가능하므로 시간적 소급성은 하락한다. 공간 프레임워크의 관점에서 보면, 우선 정의의 스케일이 읍면동이라는 점을 명확히 할 필요가 있다. 주민등록 인구이동 데이터는 “읍면동

경계를 넘어 거주지를 변경한 경우”를 인구가동으로 간주한다(통계청, 2020). 지역 스케일로 보면, 읍면동, 시군구, 시도 단위 모두 제공되기 때문에, 공간적 상세성과 다양성이 모두 높다고 할 수 있다. 그런데, 지역별 데이터는 읍면동 수준까지 손쉽게 구득할 수 있지만, 지역간 데이터는 읍면동의 경우 마이크로데이터 서비스를 이용해야 한다(https://mdis.kostat.go.kr/). 속성 프레임워크의 관점에서 보면, 인구 포괄성과 인구구조 상세성은 높지만, 성, 연령 외에 전입 사유만 존재하기 때문에 유관 속성의 이용가능성은 낮다고 평가할 수 있다.

셋째, 등록센서스 인구가동 데이터는 센서스와 관련하여 가장 최근에 생산되기 시작한 데이터이다. 앞에서 살펴본 표본센서스 데이터는 전통적으로 표본조사 항목에 포함되어 있었던 1년 전 거주지와 5년 전 거주지에 대한 정보(방문조사 혹은 행정자료 기반)를 바탕으로 만들어진 것이라면, 등록센서스 데이터는 2015년 등록센서스의 전면적 도입과 함께 전수조사 항목으로 1년 전 거주지를 조사하기 시작하면서 만들어진 것이다. 2020년 기준의 우리나라 센서스 전수조사 항목은 모두 16개이고, 이 중 8개가 인구 관련 항목인데, 성명, 성별, 나이, 가구주와의 관계, 국적, 입국연월, 국적취득연도와 함께 1년전 거주지가 포함되어 있다(통계청, 2022). 등록센서스 데이터는 직접 조사없이 주

민등록부, 학적부, 기숙시설(대학)이용자명부 등과 같은 등록부나 명부, 그리고 여타의 행정자료에 기반해 생산된다(통계청, 2022). 표본센서스 데이터처럼 0세 인구에 대한 이동 정보는 누락된다. 시간 프레임워크의 관점에서 보면, 매년 데이터가 생성 및 공표되므로 시간적 상세성은 표본센서스 데이터보다는 높고, 주민등록 데이터보다는 낮다고 평가할 수 있다. 연단위 측정 시점은 매년 11월 1일이다. 2020년 기준으로 총 6,766,523명이 이동해 이동률은 13.6%였다(통계청, 2023). 데이터가 2016년부터 존재하기 때문에 시간적 소급성은 상대적으로 낮다고 할 수 있다. 공간 프레임워크의 관점에서 보면, 우선 정의 스케일이 주민등록 데이터와 마찬가지로 읍면동인데 “1년 전 거주지 기준으로 읍면동 경계를 벗어난 이동자”의 수를 측정하는 것이다(통계청, 2023). 지역 스케일로 보면, 기본적으로 시군구와 시도 단위가 제공되기 때문에 주민등록 데이터에 비해 공간적 상세성과 다양성이 모두 낮다고 할 수 있다. 그러나 5년 단위의 센서스 주기의 해에는 읍면동 단위의 데이터도 제공된다.

표 2는 세 가지 우리나라 인구가동 데이터의 특성을 앞에서 논의한 인구가동 데이터의 분류 범주에 기반하여 정리한 것이다. 여기서 가장 중요한 사항은 수집 프레임워크의 데이터 원천 부분이다. 표본센서스 데이터와 등록센서스

표 2. 우리나라 인구가동 데이터의 종류와 특성 비교

분류 범주		인구가동 데이터의 종류		
		표본센서스	주민등록	등록센서스
수집 프레임워크	데이터 원천	센서스와 레지스터의 혼합	레지스터	레지스터
	데이터 유형	트랜지션(1년, 5년, 생애)	이벤트	트랜지션(1년)
시간 프레임워크	시간범위*	1970~2020년	1970~2023년	2016~2022년
	시간단위	5년	월, 분기, 년	년
	측정 시점(연 단위의 경우)	11월 1일	12월 31일	11월 1일
공간 프레임워크	정의 스케일	읍면동	읍면동	읍면동
	지역 스케일 (공간 상세성과 다양성)	시군구, 시도	읍면동, 시군구, 시도	시군구, 시도(센서스 연도의 경우 읍면동 제공)
속성 프레임워크	인구 포괄성	- 1년 간격은 0세, 5년 간격은 0~4세 누락 - 내국인 + 외국인	- 전연령 - 내국인	- 0세 누락 - 내국인 + 외국인
	인구구조 상세성	보통(5세는 1~4세부터 75세 이상까지 16개 연령, 그러나 다른 속성과 결합시 상세성 하락)	높음(각세는 0세에서 100세 이상까지 101개 연령, 5세는 0~4세에서 80세 이상까지 17개 연령)	보통(5세는 1~4세부터 75세 이상까지 16개 연령)
	유관 속성 이용가능성	성, 연령 포함 다수	성, 연령, 사유에 한정	성, 연령, 이동자 유형

*시간범위: KOSIS(https://kosis.kr/)에서 2024년 8월 13일 확인.

데이터 모두 기본적으로 센서스의 산물이기 때문에 센서스 원천의 데이터로 볼 소지가 있지만, 데이터 생산의 과정에 행정자료가 많이 사용되기 때문에 레지스터 원천의 데이터로 볼 소지 역시 존재한다. 우선, 표본센서스 데이터는 센서스와 레지스터가 결합된 형태로 보는 것이 합당해 보인다. 표본센서스 데이터에서 생애 트랜지션(출생지)은 현장조사에 기반하므로 전통적인 센서스 원천의 데이터이지만, 1년 기간과 5년 기간 트랜지션 데이터는 2020년부터 행정자료 기반으로 바뀌면서 레지스터의 성격이 강하게 되었다(통계청, 2022). 등록센서스-기반 데이터는 순전히 레지스터 원천의 데이터로 간주하는 것이 합당해 보인다. 비록 데이터 유형이 전통적인 센서스 원천에서 주로 취득하는 트랜지션 형태의 데이터이지만, 데이터의 원천이 주민등록과 여타 행정자료에 완전히 의존하고 있기 때문이다. IMAGE 프로젝트에서도 등록센서스에 의거해 생산된 인구이동 데이터를 레지스터 원천의 데이터로 간주한다(Bell *et al.*, 2015a).

3. 주민등록 인구이동 데이터와 등록센서스 인구이동 데이터의 비교

이 장에서는 앞에서 살펴본 세 가지 인구이동 데이터 중 주민등록 데이터와 등록센서스 데이터를 비교하고자 한다. 비교는 2016~2020년의 5년간 개별 연도 데이터를 대상으로 하며, 지역 스케일은 시도와 시군구이다. 비교의 가장 중요한 목적은 인구추계에 주민등록 데이터가 지배적으로 사용되어 왔는데(통계청, 2024a; 2024b), 등록센서스 데이터가 새로운 데이터 원천으로서 어느 정도의 보완성을 가지는지를 타진해 보는 것이다. 두 데이터 모두 성별·연령별 시군구간 O-D 매트릭스를 분석의 기본으로 하였다.⁵⁾ 연령은 5세 간격의 연령 그룹을 사용하였는데, 0~4세에서 80세 이상까지의 17개 그룹이다.

1) 비교의 선행 과정

두 데이터의 상대적인 장단점과 상호보완성을 보다 정확히 평가하기 위해서는 두 데이터를 최대한 공통의 포맷으로 전환할 필요가 있다. 본 연구의 전략은 주민등록 데이터를 기준으로 하여 등록센서스 데이터를 최대한 유사하게 변형

하는 것이다. 이를 위해 다음의 세 가지 절차를 거쳤다.

첫째, 표 2에서도 살펴본 것처럼, 주민등록 데이터는 내국인만을 대상으로 하지만, 등록센서스 데이터는 내국인과 외국인 모두를 합산한 것이다. 등록센서스 데이터를 주민등록 데이터에 맞추는 것이 기본 과정이기도 하지만, 등록센서스 데이터에 포함된 외국인 이동자의 수가 매우 적어 굳이 이를 고려해야 할 특별한 이유가 없기 때문이기도 하다. 예를 들어, 2020년 등록센서스 기준으로 시군구 경계를 넘어 이동한 외국인은 총 142,303명인데, 동일 시군구 쌍을 제외한 총 52,441 시군구쌍 가운데 32,031쌍에 대해서는 1명의 이동도 없는 정도이다. 그래서 등록센서스 데이터에서 내국인 이동만 추출하여 사용하였다.

둘째, 등록센서스 데이터에는 1년 기간 트랜지션 데이터라는 점에서 당연히 0세 인구가 누락되어 있다. 따라서 1~4세 인구를 0~4세 인구로 전환하는 방법을 고안할 필요가 있다. 이를 위해 주민등록 데이터에서 1세 단위 OD 매트릭스 추출이 가능한 통계청 마이크로 데이터(MDIS)를 활용하였다. 절차는 다음과 같다. 우선, 마이크로 데이터에서 1~4세가 0보다 큰 경우와 0인 경우로 구분하였다. 전자의 경우에는, 마이크로 데이터에서 0~4세 이동자를 1~4세 이동자로 나눈 값을 일종의 확장계수로 활용하였다. 즉, 이 확장계수를 등록센서스 데이터의 1~4세 인구에 곱하여 0~4세 인구를 추정하였다. 후자의 경우는 좀 더 복잡하다. 마이크로 데이터에서 1~4세 이동자는 0이지만, 0세 이동자는 0일 수도 있고 0이 아닐 수도 있기 때문이다. 0인 경우는 등록센서스 데이터에서도 0세 이동자가 없는 것으로 간주하여 1~4세 이동자로 0~4세 이동자를 추정하였다. 마이크로 데이터의 0세 이동자가 0이 아닌 경우는 부모 세대에서의 두 데이터 셋간 비율을 이용해 추정하였다. 즉, 등록센서스 데이터의 15~49세 이동자수를 마이크로 데이터에서의 15~49세 이동자수로 나누어 비를 구하고, 그 비에 마이크로 데이터에서의 0세 이동자수를 곱하여 등록센서스 데이터의 0세 이동자를 추정하였다. 그 값에 1~4세 이동자수를 더하여 최종적인 0~4세 이동자수를 추정하였다.

셋째, 연앙인구는 인구이동 관련 비율 측도를 계산하는데 필수적이다. 주민등록 데이터의 경우는 6월 30일(혹은 7월 1일) 인구에 해당하는 연앙인구가 이미 KOSIS 상에 존재한다. 이와 유사한 개념의 인구를 등록센서스 데이터에 대해서도 계산할 필요가 있다. 등록센서스 인구는 11월 1일 기준이므로, 매년의 연앙인구는 4월 30일(5월 1일) 인구여야 하는데, 해당 월일의 인구를 정확히 계산하기 어렵기

때문에 등록센서스 총인구 기준으로 직전 연도의 인구나 해당 연도의 인구를 평균한 것을 연앙인구로 산정하였다. 각 전출지 기준의 비이동자수를 산정하는 것도 매우 중요한데, 인구추계에 비이동률이 사용되기 때문이다. 각 전출지별 해당 연도의 연앙인구에서 다른 전입지로 이동한 총 전출자를 제외한 인구를 비이동자수로 산정하였다.

2) 비교 결과

(1) 총이동 및 총이동률

표 3은 공간단위 수준별 데이터 원천 간 총이동(gross migration)과 총이동률을 비교한 결과를 보여주고 있다. 여기서 총이동과 총이동률은 전역 스케일(우리나라 전체)에서의 인구이동 규모와 인구이동 강도를 측정하는 것인데, 이러한 전역 스케일의 측도는 주어진 지역 스케일에 따라 다양하게 산출될 수 있다(이상일·이소영, 2021). 즉, 전역 스케일의 인구이동에 대한 측도값은 주어진 지역 스케일(시도 혹은 시군구)에 의존적이다. 표 1에 나타난 사항 중 중요한 것을 정리하면 다음과 같다.⁶⁾ 첫째, 두 공간단위 수준 모두에서, 등록센서스 데이터가 주민등록 데이터에 비해 총이동량의 규모가 작았다. 시도 수준에서는 주민등록 데이터가 약 240~255만 정도였지만 등록센서스의 경우는 200~215만 정도로 연간 대략 40만명 정도 작았다. 시군구 수준에서도 주민등록 데이터는 440~480만명 정도였지만 등록센서스의 경우는 370~400만명 정도로 대략 70~80만명 정도 작은 규모로 나타났다. 둘째, 총이동률에서도 공간

단위 수준에 관계없이 등록센서스가 주민등록에 비해 낮은 값을 나타냈다. 시도 수준의 경우는 대략 0.55~0.65% 정도의 차이를, 시군구 수준에서는 1.10~1.35% 정도의 차이를 나타냈다.

총이동과 총이동률에서의 이러한 두 원천간 차이를 상이한 데이터 유형(트랜지션과 이벤트)에 따른 데이터 포착성에서의 차이라고 간단히 말하는 것은 적절치 않다. 그러한 설명은 우리나라 주민등록 데이터와 2015년 이전의 현장 조사 기반의 표본센서스 데이터를 비교하는 경우라면 적절할 수 있다. 앞 절에서 얘기한 것처럼, 두 데이터 원천 모두 레지스터 원천에 속하며, 따라서 다른 방식으로 데이터 포착성의 장단점을 논의해야만 한다. 두 데이터 원천은 서로 다른 근거에서 데이터 포착성이 상대적으로 높다. 우선 주민등록 데이터는 기본적으로 이벤트 데이터이므로 트랜지션 데이터인 등록센서스에 비해 데이터 포착성이 양호하다. 반면에 등록센서스 데이터는 다양한 행정 데이터를 사용하기 때문에 보다 실질적인 거주 이동의 양상을 파악할 수 있다는 측면에서 주민의 신고에만 전적으로 의존하는 주민등록 데이터에 비해 데이터 포착성이 양호하다. 결국 문제는 상대적으로 양호한 포착성의 규모 간의 차이이다. 표 3이 보여주고 있는 것은, 등록센서스가 더 포착해 낸 인구이동의 규모를 상쇄하고도 남을 만큼의 더 큰 규모의 인구이동을 주민등록 데이터가 포착해 냈다는 점이다.

시도별로 총이동과 총이동률을 계산하여 데이터 원천간 편차를 분석하였다. 여기서 총이동과 총이동률은 표 3에 나타나 있는 것과 같은 전역적 수준의 측도가 아니라, 시도

표 3. 공간단위 수준별 총이동 및 총이동률 비교

연도	시도			시군구		
	주민등록 인구이동	등록센서스 인구이동	차이	주민등록 인구이동	등록센서스 인구이동	차이
2016년	2,453,342 (4.80)	1,820,710 (3.66)	632,632 (1.14)	4,538,135 (8.88)	3,400,037 (6.83)	1,138,098 (2.05)
2017년	2,410,930 (4.71)	2,069,206 (4.15)	341,724 (0.56)	4,452,826 (8.69)	3,790,343 (7.60)	662,483 (1.10)
2018년	2,429,184 (4.74)	2,042,538 (4.09)	386,646 (0.65)	4,520,418 (8.81)	3,736,414 (7.48)	784,004 (1.33)
2019년	2,384,948 (4.65)	2,032,520 (4.07)	352,428 (0.58)	4,444,572 (8.66)	3,747,248 (7.50)	697,324 (1.16)
2020년	2,534,114 (4.94)	2,148,356 (4.29)	385,758 (0.64)	4,809,085 (9.37)	4,033,141 (8.06)	775,944 (1.31)

주: 괄호 속의 값은 총이동률(%)임.

별로 따로 계산한 지역적 스케일의 측도이다. 총이동은 시도별 전입과 전출의 규모를 합한 것이고, 총이동을 연방인구로 나눈 것이 총이동률이다(이상일·이소영, 2021; Rowland, 2003). 17개 시도 모두 표 3에 나타난 것과 유사한 경향을 보여주었다. 즉, 모든 연도에 대해, 등록센서스가 주민등록에 비해 총이동과 총이동률 모두에서 낮은 값을 나타냈다. 그런데, 2020년 한 해를 대상으로, 두 데이터 원천 간의 총이동에서의 편차를 시도별로 계산하고, 그 편차의 시도간 차이를 살펴본 결과 흥미로운 결과가 나타났다. 제주(24.9), 전라남도(22.1), 경기도(18.7), 대구(18.2), 서울(17.8), 세종(15.8), 경남(15.7)은 등록센서스의 총이동이 주민등록의 총이동에 비해 15% 이상 적었지만, 충청남도(3.3), 강원도(5.9), 충청북도(7.1), 대전(8.4)는 그 편차가 10% 미만이었다. 이는 두 데이터 원천 간의 차이, 특히 데이터 포착성에서의 차이가 모든 지역에 대해 동일하게 나타나는 것을 아니라는 점을 명확히 보여준다.

(2) 이동자의 인구구조

등록센서스의 상대적인 과소 포착의 경향이 연령별로 다른 양상을 보이는지의 여부를 평가하기 위해 일종의 인구이동 피라미드를 제작하였다(김감영, 2010; Plane, 1992). 2016년을 제외하고 나머지 2017~2020년의 4개년 모두 유사한 패턴이 나타났는데, 그림 1은 2020년의 결과를 나타낸 것이다. 그림 1(a)에 나타나 있는 시도간 이동을 살펴보

면, 15~19세와 20~24세를 제외한 나머지 모든 연령층에서 주민등록이 등록센서스에 비해 더 많은 이동량을 나타냈다. 그리고 연령별 이동량에서의 편차는 거의 이동 규모에 비례하는 것으로 보인다. 즉, 이동량이 가장 많은 25~29세에서 두 데이터 원천 간의 차이가 가장 크게 나타난다. 그림 1(a)에서 가장 주목해야 하는 사실은 15~19세와 20~24세에서는 등록센서스가 주민등록보다 오히려 더 큰 이동량을 기록했다는 점이다. 특히 주민등록 데이터의 경우, 25~29세가 20~24세보다 월등히 많은 이동자수를 기록한 반면, 등록센서스에서는 20~24세가 25~29세보다 오히려 더 큰 값을 보여주고 있다. 이것은 데이터 포착성에서의 차이가 15~24세 인구에도 적용된다는 점을 감안할 때 매우 이례적인 결과가 아닐 수 없다. 이러한 경향성은 그림 1(b)에 나타나 있는 시군구 수준의 데이터에서도 거의 동일하게 드러난다. 즉, 15~24세의 젊은층에서 등록센서스가 주민등록에 비해 더 많은 이동자를 기록하고 있는 것이다. 차이 점이려면 두 데이터 원천 모두에서 25~29세 이동자수가 20~24세 이동자수보다 많다는 것이다. 하지만 등록센서스의 20~24세 이동자수는 주민등록과는 달리 20~34세 이동자수보다 많다. 덧붙여 80세 이상에서도 등록센서스가 더 많은 이동자를 기록하고 있다.

트랜지션 데이터인 등록센서스가 유독 15~24세 인구에서 예외적으로 높은 데이터 포착성을 보이는 이유는 무엇보다도 등록센서스가 주민등록을 포함한 다양한 행정자료

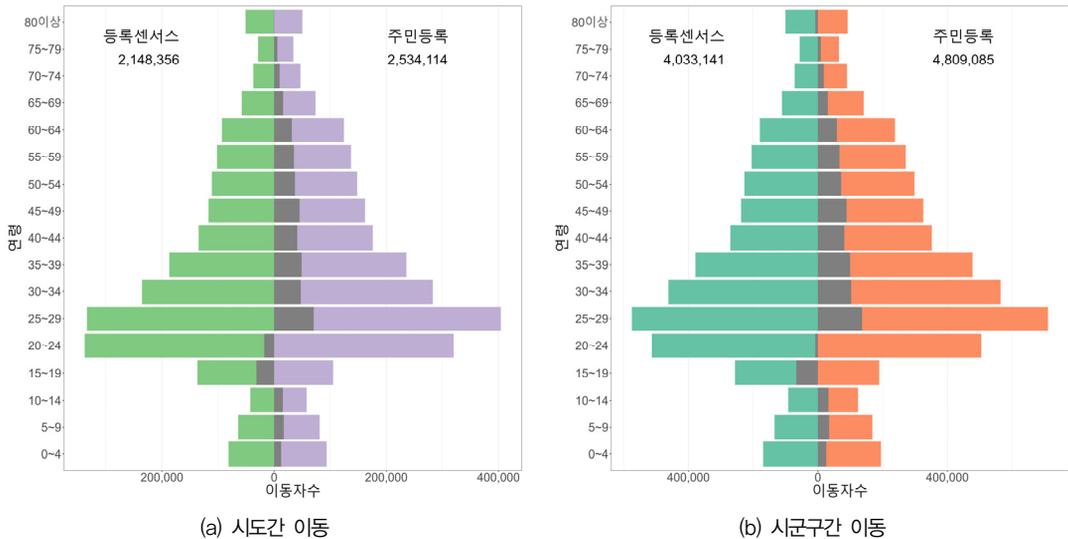


그림 1. 이동자의 연령별 인구이동량 비교(2020년)

를 활용하여 수집된 것이기 때문이다. 우리나라 등록센서스에는 모두 24층(2023년 11월 기준)의 행정자료가 사용되는데, 이들 중 학적부(대학), 기숙시설(대학)이용자명부, 출입국자료 등 젊은 층의 인구이동 양상을 잘 포착하는 정보의 사용이 가장 큰 기여를 한 것으로 보인다(통계청, 2022). 대학 진학 후 기숙시설 등에 거주하는 경우 부모의 주민등록지를 그대로 유지하는 경우가 많은 현실을 감안할 때 이러한 특성을 어렵지 않게 이해할 수 있다. 유사한 분석을 시도별 총이동자의 인구구조에도 적용했는데, 대부분의 시도가 전국적인 경향성과 유사한 패턴을 보였다. 그러나 앞에서 두 데이터 원천 간의 총이동량의 편차가 상대적으로 컸던 지역(충청남도, 강원도, 충청북도, 대전 등)이 15~24세 연령층에서 가장 큰 역전을 보인 지역이기도 했다.

그림 1에 나타나 있는 이동자의 연령 구조는 절대적인 이동자수에서의 연령별 상대적 비중을 파악할 수 있다는 장점이 있지만, 연령별 인구이동 확률에 대한 정보는 제공해주지 못한다. 즉, 특정 연령의 인구이동량이 더 많다고 해서 해당 연령의 이동확률이 더 높다고 말할 수는 없는 것이다. 그림 2는 이러한 측면을 검토하기 위해 연령층별 이동률(ASMR, age-specific migration rate)을 나타낸 것이다. 여기서 연령층별 이동률은 해당 연령층에 포함된 인구가 주어진 공간단위의 경계를 넘어 인구이동을 감행할 실질적인 확률을 의미한다. 그림 2(a)는 시도간 이동을 나타낸

것으로 15~24세의 이동확률이 등록센서스가 주민등록에 비해 더 높음을 명확히 보여주고 있다. 또한 주민등록의 경우는 25~29세의 이동확률이 가장 높게 나타나지만, 등록센서스의 경우는 20~24세의 이동확률이 더 높게 나타남을 알 수 있다. 그림 2(b)는 시군구간 이동의 경우를 보여주고 있는데, 15~24세와 80세 이상에서 등록센서스의 이동확률이 상대적으로 더 높음을 알 수 있다. 결국 등록센서스는 주민등록에 비해 젊은 층에서 인구이동량 뿐만 아니라 이동확률에서도 더 높은 값을 나타낸다.

그림 1이 이동자의 연령 구조에 집중했다면 그림 3은 이동자의 성별 구조와 연령별 구조를 동시에 파악하기 위해 완전한 인구 피라미드 형태로 시각화한 것이다. 그림 3(a)에는 시도간 이동자의 성별·연령별 구조가 나타나 있다. 주민등록이 등록센서스에 비해 상대적으로 비중이 높은 대부분의 연령대에서는 남성과 여성의 차이가 그리 크지 않다. 그런데 등록센서스의 비중이 상대적으로 높은 15~24세에는 남성의 비중이 다소 높은 것으로 나타났다. 특히 20~24세의 경우에 이러한 경향이 두드러진다. 실질적으로 20~24세의 경우 주민등록의 경우는 여성이 남성보다 많지만, 등록센서스의 경우는 남성이 여성보다 많다. 이것은 행정자료의 보완으로 인한 등록센서스의 데이터 포착성의 상대적인 향상이 여성보다는 남성에서 더 많이 이루어졌음을 의미하는 것이다. 80세 이상에서는 그 반대의 경향이

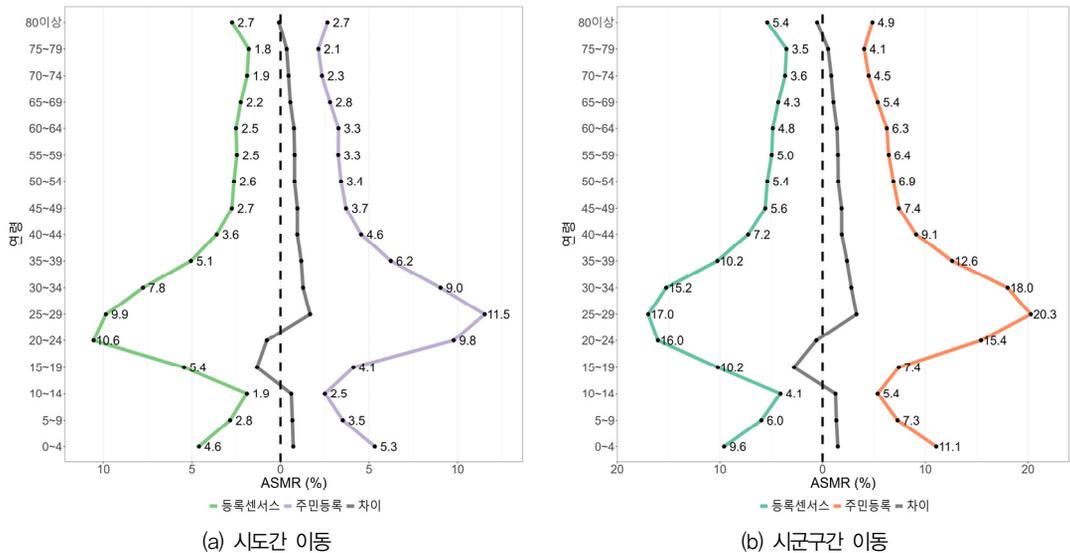


그림 2. 이동자의 연령별 이동률 비교(2020년)



그림 3. 이동자의 성별·연령별 인구구조 비교(2020년)

나타난다. 그림 3(b)는 시군구간 이동에 대한 성별·연령별 인구구조를 비교하고 있다. 그림 3(a)과 거의 동일한 패턴이 나타남을 확인할 수 있다.

(3) 시도간 전출입 플로

지금까지는 우리나라 전체와 시도별 분석에 집중했다. 여기서는 시도간 전출입 매트릭스의 분석에 집중하고자 한다. 시도간 OD 매트릭스 분석에서 중요한 것은 절대적인 이동량을 그대로 비교하는 것이 아니라 전체 이동량 중 각 지역쌍이 차지하는 상대적인 이동량을 통해 비교하는 것이다. 왜냐하면 절대적 이동량의 경우 모든 지역간 플로에서 데이터 포착성이 높은 주민등록 데이터가 등록센서스에 비해 높은 값을 보일 가능성이 높기 때문이다. 실질적으로 2020년을 대상으로 시도간 전출입에 대해 두 데이터 원천을 비교한 결과 모든 시도간 플로에서 주민등록이 높은 값을 나타냈다. 따라서 여기에서는 상대적 이동량으로 구성된 두 개의 OD 매트릭스를 차감함으로써 어떤 데이터 원천에서 어떤 플로가 상대적으로 더 두드러지는지를 검토하고자 한다. 더 나아가 OD 매트릭스의 차이는 연령 집단에 따라 달리 나타날 가능성이 큰 데 등록센서스에서 상대적 데이터 포착성이 높게 나타난 15~19세 연령층에 대해서도 동일한 분석을 실시하였다.

그림 4는 주민등록의 상대적 플로 매트릭스에서 등록센서스의 상대적 플로 매트릭스를 뺀 값을 시각화한 것이다. 짙은 붉은색은 높은 양의 값, 짙은 푸른색은 높은 음의 값을 나타낸다. 그림 4(a)는 전연령에 대한 것인데, 뚜렷한 패턴 없이 매우 다양한 양상으로 나타남을 알 수 있다. 여기서 가장 중요한 것은 경기에서 서울, 서울에서 경기로의 플로가 모두 음수값을 나타낸다는 점이다. 이것은 두 개의 가장 큰 시도 수준의 플로가 주민등록보다는 등록센서스로 측정했을 때 상대적으로 보다 큰 흐름으로 포착된다는 것을 의미한다. 이것은 더 나아가 행정자료를 통해 주민등록에 의거한 인구가동을 보정한 효과가 가장 거대한 두 플로에서 가장 잘 나타난다는 것을 함축한다. 그림 4(b)는 이러한 측면을 한번 더 강조해준다. 15~19세로 한정하여 비중의 차이를 구한 결과 경기에서 서울과 서울에서 경기로의 플로에서 등록센서스의 상대적인 포착성이 두드러지게 나타난다.

다음으로 전이확률 매트릭스를 비교하고자 한다. 절대적 이동 플로 매트릭스의 모든 값을 해당 전출지의 인구로 나누면 전이확률(transition probability) 매트릭스를 구할 수 있다. 전이확률은 전출지의 인구가 전입지로 인구가동을 행할 가능성에 대한 측도이기 때문에 인구 추계와 같은 고급 인구분석에서 매우 중요하다. 그림 5는 주민등록의 전이확률 매트릭스에서 등록센서스의 전이확률 매트릭스

를 뺀 값을 시각화한 것이다. 그림 4와 마찬가지로 짙은 붉은색은 높은 양의 값을, 짙은 푸른색은 높은 음의 값을 나타낸다. 그림 5(a)는 전 연령에 대한 것인데, 대부분의 플로에서 주민등록 데이터에 기반한 전이확률이 등록센서스에 기반한 전이확률에 비해 높다. 이것은 절대적 이동 플로가 주민등록이 더 크고, 그것을 전출지 인구로 나누어도 큰 값

이 나타나기 때문에 매우 자연스러운 결과이다. 양의 값의 편차는 거의 이동량에 비례한다. 그림 5(b)는 15~19세에 대한 것으로 등록센서스의 인구이동 포착성이 높은 연령대이므로 대부분의 플로에서 음의 값이 나타나고 있다. 그런데 여기서 주목해야 하는 것은 몇 개의 플로(서울에서 경기, 경기에서 서울, 대전에서 세종, 인천에서 경기)에서는 주민

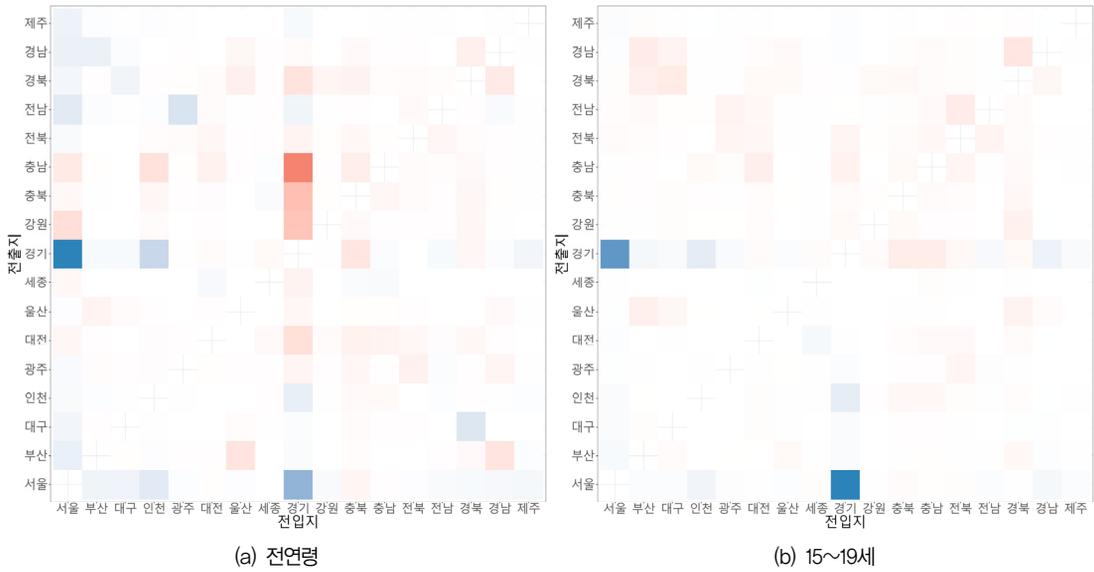


그림 4. 시도간 이동 플로의 비교(2020년)

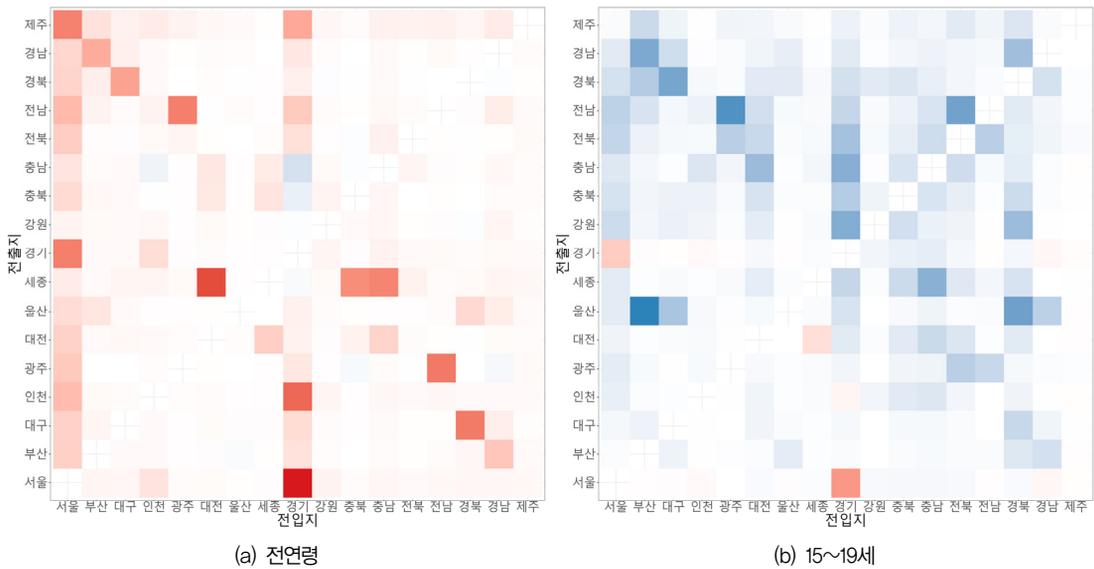


그림 5. 시도간 전이확률의 비교(2020년)

등록의 전이확률이 등록센서스의 전이확률보다 높게 나타난다는 사실이다. 이것은 젊은 층에서의 데이터 포착성이 등록센서스가 뛰어나다는 점을 염두에 둔다면, 이러한 플로에서 주민등록에 기반한 전이확률이 실질적으로 과장된 것임을 함축하고 있다.

(4) 지역별 순이동

2020년에 대해 시도별 순이동이 데이터 원천에 따라 어느 정도의 차이를 보이는 지를 분석하였다. 표 4는 전연령과 15~19세에 대한 시도별 순이동의 차이를 보여주고 있다. 우선 전연령에 대해 살펴보면, 대부분의 시도가 순이동의 절대적인 규모에서 두 데이터 원천 간에 차이를 보이고 있지만 부호는 강원과 충남을 제외하고는 동일함을 알 수 있다. 여기서 중요한 것은 서울과 경기인데, 서울은 등록센서스에서 음의 순이동이 더 작고, 경기는 등록센서스에서 양의 순이동이 더 크게 나타난다. 이 두 시도를 포함해서 전남과 경남이 등록센서스의 순이동이 주민등록의 순이동에 비해 4,000명 이상 큰 값을 보여주었다. 이에 반해 충남, 강원, 광주는 주민등록이 등록센서스 보다 순이동이 4,000명

이상 큰 지역들로 나타났다. 특히 충남은 두 데이터 원천 간에서 순이동량의 차이가 가장 크게 드러난 곳이다. 15~19세의 경우로 한정하면 좀 다른 양상이 드러난다. 가장 중요한 것은 서울, 대전, 부산, 대구와 같은 대도시의 경우 등록센서스의 순이동이 주민등록에 비해 1,000명 이상 크게 나타난다는 사실이다. 이에 반해, 경남, 전남, 울산은 해당 연령층에서 등록센서스의 순이동 규모가 주민등록에 비해 1,000명 이상 작게 나타난다.

2020년에 대하여 시군구별 순이동률(NMR, net migration rate)을 계산하여 두 데이터 원천을 비교하였는데, 주민등록 데이터의 순이동률에서 등록센서스의 순이동률을 차감한 값으로 지도를 제작하였다. 그림 6의 지도에서 큰 음의 값을 보인 시군구는 주민등록 데이터에 비해 등록센서스 데이터의 순이동률이 더 큰 경우이다. 그림 6(a)에 나타나 있는 전연령의 경우, 두 순이동률 사이에는 큰 차이가 없고, 뚜렷한 패턴도 탐지되지 않는다. 111개의 시군구가 양의 값을, 118개 시군구가 음의 값을 나타냈다. 그런데도 그림 6(b)에 나타나 있는 15~19세의 경우, 두 데이터 원천 간의 순이동률 편차는 규모도 커지고, 패턴도 뚜렷하게 나타난다

표 4. 데이터 원천에 따른 시도별 순이동의 차이(2020년)

시도	전연령			15~19세		
	주민등록 인구이동 (A)	등록센서스 인구이동 (B)	차이(A-B)	주민등록 인구이동 (A)	등록센서스 인구이동 (B)	차이(A-B)
서울	-64,850	-58,153	-6,697	3,436	6,144	-2,678
부산	-14,347	-16,926	2,579	-43	1,500	-1,543
대구	-16,835	-14,140	-2,695	-996	119	-1,115
인천	-15,748	-14,178	-1,570	-170	-983	813
광주	-6,083	-10,831	4,748	-566	-380	-186
대전	-11,097	-13,603	2,506	-291	1,426	-1,717
울산	-13,584	-10,189	-3,395	-1,082	-2,331	1,249
세종	13,025	11,706	1,319	331	576	-245
경기	168,373	172,529	-4,156	5,133	5,911	-778
강원	5,457	-1,935	7,392	-272	-1,061	789
충북	3,454	3,817	-363	-415	616	-1,031
충남	741	-12,451	13,192	-775	-765	-10
전북	-8,494	-7,808	-686	-685	-908	223
전남	-9,754	-1,864	-7,890	-1,251	-3,724	2,473
경북	-16,978	-18,524	1,546	-827	-692	-135
경남	-16,658	-10,709	-5,949	-1,440	-4,740	3,300
제주	3,378	3,259	119	-87	-678	591

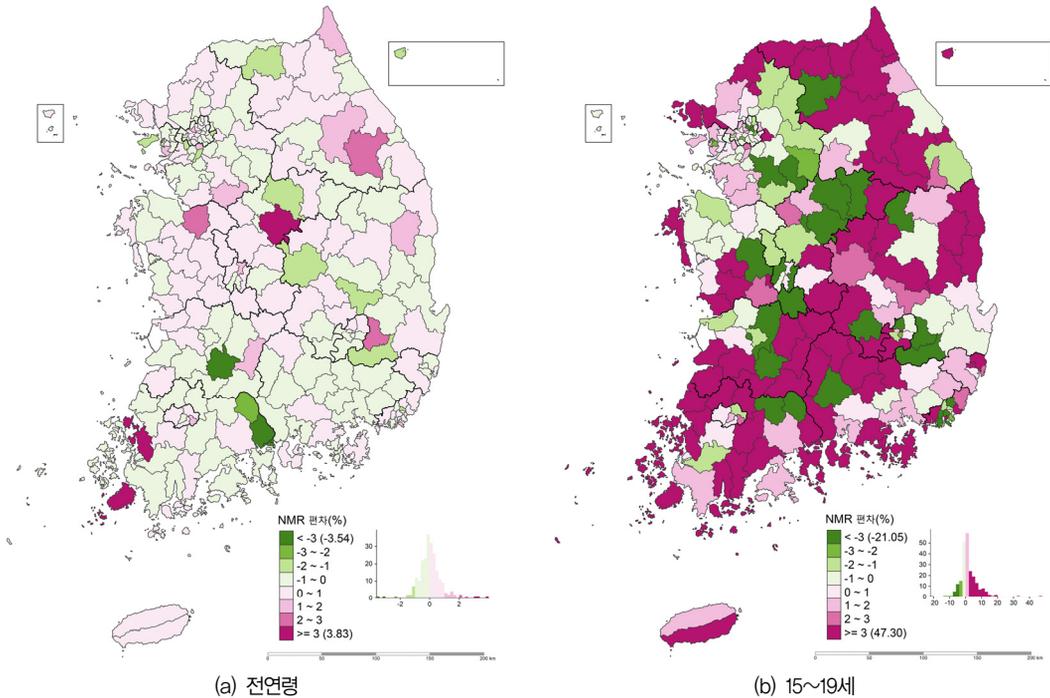


그림 6. 데이터 원천 간 시군구별 순이동률의 비교(2020년)

다. 즉, 서울의 일부 지역을 포함한 수도권, 대도시의 일부 지역 및 그 주변 지역에서 비교적 높은 음의 값이 나타나고 있다. 즉, 이들 지역에서는 주민등록에 비해 등록센서스에서 더 우호적인 인구 교환이 발생하고 있음을 보여준다. 경기의 이천시와 용인시, 서울의 성북구와 동대문구, 부산의 사상구, 부산진구, 영도구, 금정구, 사하구, 중구, 대전의 동구, 유성구, 대구의 북구 등이 3% 이상의 순이동률 차이를 보였다. 등록센서스에 비해 주민등록에서 더 우호적인 값으로 나타나는 지역은 대체로 농어촌 및 산간지역에 해당하는데, 경북 영양군, 전북 장수군, 경북 청송군, 전남 보성군, 장성군, 경남 고성군, 전남 진도군이 15%이상의 편차를 나타냈다.

4. 토론 및 결론

본 논문의 주된 연구목적은 우리나라의 주민등록 인구이동 데이터와 등록센서스 인구이동 데이터를 비교하여 상대적인 장단점과 상호보완성을 탐색하는 것이었다. 이를

위해 우선적으로 인구이동 데이터 분류 범주에 대한 개념적인 논의를 진행하였다. 분류 범주를 크게 수집 프레임워크(데이터 원천과 데이터 유형), 시간 프레임워크(시간범위, 시간단위, 측정 시점), 공간 프레임워크(정의 스케일, 지역 스케일), 속성 프레임워크(인구 포괄성, 인구구조 상세성, 유관 속성 이용가능성)로 나누어 정리하였다.

우리나라는 결합센서스를 시행하고 있는 나라이다. 표본항목에 대해서는 여전히 5년 주기의 센서스를 시행하고 있을 뿐만 아니라 많은 항목을 직접 조사에 의존하고 있기 때문에 전통적인 센서스의 특징을 보유하고 있다. 그러나 전수항목에 대해서는 2015년 등록센서스를 전면적으로 도입하여 매년 행정자료에 기반한 데이터 수집을 행하고 있을 뿐만 아니라 5년 주기 센서스의 몇몇 항목 역시 전적으로 행정자료에 의존하고 있다(통계청, 2022). 결국 우리나라는 전통적 센서스와 등록센서스를 절충한 결합센서스의 시스템을 구축하고 있는 것이다. 하이브리드 센서스의 시행은 우리나라 인구데이터의 다양성에 기여했으며, 현재 세 가지의 서로 다른 인구이동 데이터(전통센서스, 주민등록, 등록센서스)가 존재한다. 인구이동 데이터의 분류 범주에 기반하여 이 세 가지 인구이동 데이터의 특성을 정리

하였다.

본 연구는 주민등록 데이터와 등록센서스의 비교 연구에 집중했다. 둘 다 정의 스케일이 읍면동으로 동일하고, 고차적 인구지리학적 분석에 투입할 만큼 충분한 시간범위를 가지고 있고, 비슷한 수준의 공간 상세성과 다양성을 보여주고 있기 때문이다. 이러한 비교 분석에서 주목할 만한 부분은 인구의 변화, 특히 미래 인구추계에 있어서 등록센서스 데이터가 얼마나 보완적 혹은 대체적 기능을 할 수 있는지를 검토하는 것이었다. 우리나라 인구추계는 기본적으로 총이동 코호트-요인법(gross migration cohort-component method)(이상일·조대현, 2012; Smith *et al.*, 2001)에 기반하고 있다. 그 중 인구이동 컴포넌트에 대해 변동전이 확률모형을 적용하고 있는데, 기본 전이확률 계산을 위해 최근 몇 년간의 주민등록 데이터를 사용한다(이상일 등, 2022; 전광희 등, 2014; 통계청, 2024a, 2024b). 인구추계라는 측면에서 두 데이터를 비교한다고 했을 때 가장 중요한 비교 범주는 데이터 포착성이다. 주민등록 데이터는 이벤트 데이터라는 측면에서 트랜지션 데이터인 등록센서스 데이터에 비해 데이터 포착성이 양호하다. 이에 반해 등록센서스 데이터는 주민등록 외에 다양한 행정자료를 활용한다는 측면에서 데이터 포착성이 양호하다. 결국 두 데이터는 데이터 포착성의 양호/불량의 측면에서 차이가 나는 것이다. 이러한 차이가 인구추계를 위한 기본 데이터로서 어떻게 상호보완적인 역할을 할 수 있을 것인가를 검토하는 것이 본 논문의 요체이다.

2016~2020년의 5개 연도에 대해 두 데이터 원천 간의 비교 분석을 실시하였는데, 크게 총이동량 및 총이동률, 이동자의 인구구조, 시도간 전출입 플로, 지역별 순이동을 중심으로 비교하였다. 비교 결과를 인구추계에의 함의라는 측면에서 정리하면 다음과 같다.

첫째, 시도 수준과 시군구 수준 모두에서, 등록센서스 데이터가 주민등록 데이터에 비해 총이동량이 적고 총이동률이 낮았다. 주민등록 데이터는 이벤트 데이터이므로 상대적으로 데이터 포착성이 양호하고, 등록센서스 데이터는 다양한 행정 데이터를 사용하기 때문에 상대적 데이터 포착성이 양호하다. 총이동량과 총이동률 모두에서 주민등록 데이터가 등록센서스 데이터 보다 더 높은 값을 보인 것은, 후자가 더 포착해 낸 인구이동의 규모를 상쇄하고도 남을 만큼의 더 큰 규모의 인구이동을 전자가 포착해 냈다는 것을 의미한다.

둘째, 젊은 층(15~24세)에 대해서는 절대적인 이동량, 상

대적인 비중, 이동확률 등 모든 측면에서 등록센서스가 주민등록에 비해 상대적으로 높은 값을 나타냈다. 이것은 젊은 연령층의 인구이동을 보다 정확하게 파악할 수 있는 다양한 행정자료를 활용했기 때문이다. 결국 트랜지션 데이터의 한계로 인한 데이터 포착성의 하락에 비해 행정자료의 적극적 활용에 따른 데이터 포착성의 향상이 해당 연령에서 월등히 컸다는 해석이 가능하다. 이러한 사실은, 주민등록 데이터에 기반한 기존의 인구추계는 젊은 층의 인구이동의 역동성을 과소추정했을 수도 있음을 함축한다. 이러한 측면에서 등록센서스 데이터는 인구추계를 위한 보완적 데이터로서의 가치가 충분하며, 주민등록 데이터가 가지고 있는 젊은 층의 이동성의 과소추정 문제를 해결하는데 큰 도움을 줄 수 있을 것으로 평가한다.

셋째, 젊은 층(15~24세)의 경우, 가장 큰 시도간 흐름(서울에서 경기와 경기에서 서울)의 상대적 비중이 주민등록에 비해 등록센서스에서 좀 더 크게 나타났다. 이것은 두 개의 가장 큰 시도 수준의 플로가 주민등록보다는 등록센서스로 측정했을 때 상대적으로 보다 큰 흐름으로 포착된다는 것을 의미한다. 이것은, 인구추계에서 등록센서스 데이터를 사용할 경우, 주민등록 데이터를 사용한 경우에 비해 가장 큰 시도간 플로(서울에서 경기, 경기에서 서울)의 지배성이 좀 더 두드러지게 나타날 가능성이 있음을 함축한다. 또한 젊은 층의 시도간 전이확률에서 규모가 큰 몇 개의 플로(서울에서 경기, 경기에서 서울, 대전에서 세종, 인천에서 경기)에서는 주민등록의 전이확률이 등록센서스의 전이확률보다 높게 나타났다. 젊은 층에서의 데이터 포착성이 등록센서스가 뛰어나다는 점을 염두에 둔다면, 이러한 사실은 인구추계의 측면에서 중요한 시사점을 제공한다. 즉, 기존의 주민등록 데이터를 기반으로 계산된 젊은 층의 전이확률 중 규모가 큰 흐름의 경우 인구이동량이 과대추정될 가능성이 높은 것으로 보인다. 다시 말해, 규모가 큰 시도간 흐름의 경우, 등록센서스 데이터를 사용해 미래 인구이동을 추정하는 경우 좀 더 안정적인 지역간 플로가 산출될 수 있다.

넷째, 시도 단위의 순이동의 경우, 등록센서스가 보다 우호적인 인구 교환을 보여주는 시도는 서울과 경기이며, 반대의 경향이 가장 뚜렷한 것은 충남이었다. 젊은 층의 순이동의 경우, 등록센서스를 사용할 경우 순이동의 값이 크게 나타나는 시군구는 수도권과 대도시권의 일부 지역을 포함하는 반면, 등록센서스를 사용할 경우 순이동의 값이 작게 나타나는 시군구는 대체로 농어촌과 산간지역에 집중 분포

한다. 이것은 인구추계에서 등록센서스 데이터를 사용할 경우, 서울의 인구가동에 의한 인구감소량은 줄고, 경기의 인구가동에 의한 인구증가는 늘어날 것임을 함축한다. 또한, 수도권과 대도시권의 일부 지역에서 젊은 층의 인구 유입 양상이 좀 더 두드러지고, 농촌성이 두드러진 지역에서는 그 반대의 양상이 나타날 가능성이 높음을 시사한다.

서로 다른 데이터의 원천은 서로 경쟁 관계에 있는 것이 아니라 상호보완적인 관계에 있다는 점을 인식하는 것이 매우 중요하다(Bell *et al.*, 2015a). 위에서 살펴본 것처럼, 우리나라의 주민등록 데이터와 등록센서스 데이터는 인구추계에 다양한 함의를 제공하는 상호보완성이 높은 데이터임이 밝혀졌다. 특히 주민등록 데이터의 젊은 층에 대한 과소추정의 문제를 해결하는데 중요한 역할을 할 수 있을 것으로 평가되었다. 그러나 이러한 상호보완성이 구체적인 인구추계에서 어떠한 결과로 드러나게 될지는 또 다른 독립적인 연구를 요하는 연구과제이다.

주

- 1) 1년전 거주지에 대한 조사에 기반하고 있으므로 데이터는 2016년부터 존재한다.
- 2) 이것은 일반적인 설명임을 이해할 필요가 있다. 이벤트 데이터는 전적으로 신고에 의존하기 때문에 신고 누락의 가능성이 충분히 있다. 또한 이벤트 데이터는 오로지 이동자만을 대상으로 하는 반면 트랜지션 데이터(특히 전수조사의 경우)는 기본적으로 전 인구의 인구이동 여부를 따져보는 것이다. 이런 측면에서 보면 트랜지션이 이벤트에 비해 데이터 포착성이 더 양호하다고 얘기할 수도 있다. 그러나, 일반적으로 의미에서는 이벤트의 데이터 포착성이 트랜지션보다 양호하다고 충분히 말할 수 있다.
- 3) 이 세 가지 데이터 외에 인구이동이 부차적으로 수집된 서베이 원천의 데이터가 있을 수 있다. 대표적인 것에 대졸자직업이동경로조사를 통해 수집된 GOMS(Graduates Occupational Mobility Survey) 데이터(<https://goms.hrc-global.com/>)가 있다. GOMS 데이터에 대한 자세한 설명과 이 데이터를 활용한 인구이동 연구로 구형모(2018)를 참조할 수 있다.
- 4) 표본센서스 인구이동 데이터의 공간 프레임워크에 대해서는 다소간의 모호함이 존재한다. 다른 두 데이터 원천과 달리, 정의 스케일이 읍면동 수준이라는 사실이 통계청의 관련 문건(2021; 2022)에 명확히 기술되어 있지는 않다. 그런데 통계청의 담당 부서에 문의한 결과 내부지침상 정의 스케

일이 읍면동 수준이라는 답변을 들을 수 있었다. 단지, 2019~2020년 한 해 동안의 이동자수가 8,284천명으로, 등록센서스의 2020년 한 해 이동자 6,766,523명과 상당한 편차를 보이는 것은 해당 문건(통계청, 2021)의 이동량에는 읍면동 내부 이동까지 포함되어 있기 때문이라고 한다.

- 5) 주민등록-기반 OD 매트릭스 데이터는 KOSIS의 [온라인간행물]-[인구]-[국내인구가동통계] 항목에서 다운로드하였다. 등록센서스-기반 데이터는 통계청으로부터 직접 제공받았다. KOSIS에서 해당 통계는 [주제별 통계]-[인구]-[인구총조사]-[인구부문]-[총조사인구(2015년 이후)]-[전수부문(등록센서스, 2015년 이후)]-[국내인구가동]에서 다운로드할 수 있는데, 여기에는 외국인도 포함되어 있고, 결정적으로 OD 매트릭스 형태의 데이터가 제공되지 않는다.
- 6) 2016년은 다른 해들에 비해 두 데이터 원천간 편차가 월등히 크므로 다소 예외적인 경우로 판단하여 2017~2020년을 중심으로 정리하고자 한다.

참고문헌

- 구형모, 2018, “대학 진학 이동과 대학 졸업자 취업 이동의 공간적 특성 탐색,” 한국도시지리학회지, 21(3), 109-122. doi: 10.21189/JKUGS.21.3.8.
- 김감영, 2010, “연령별 인구이동 특성에 대한 탐색적 공간 데이터 분석(ESDA): 대구시를 사례로,” 한국지역지리학회지, 16(5), 590-609.
- 이상일, 2012, “공간적 상호작용론의 본질과 연구 영역: 인문지리학에 대한 통섭적 접근,” 한국지리학회지, 1(1), 137-151.
- 이상일·김현미, 2021, “인구가동 플로의 연령-특수적 패턴 분석을 위한 방법론 연구 - 우리나라 시군구 단위 인구 이동에의 적용 -,” 대한지리학회지, 56(5), 537-550. doi: 10.22776/kgs.2021.56.5.537.
- 이상일·이소영, 2021, “인구가동 플로의 지리적 시각화를 위한 개념적 명료화: 우리나라 2020년 인구이동에 대한 주제도 제작,” 한국지도학회지, 21(3), 23-42.
- 이상일·이소영, 2023, “인구가동이 인구재분포에 미치는 영향력의 시공간적 역동성 탐색: 우리나라 국내 인구이동에의 적용,” 한국지도학회지, 23(1), 1-19. doi: 10.16879/jkca.2023.23.1.001.
- 이상일·조대현, 2012, “지역간 인구이동의 예측을 통한 우리나라 시도별 장래 인구 추계: 다지역 코호트-요인법의 적용,” 대한지리학회지, 47(1), 98-120.

- 이상일·조대현·이소영·김세창, 2022, 지역별 국내인구이동 추계 작성방법 개선방안 연구, 통계청, 대전.
- 전광희·이상일·박민규·계봉호, 2014, 등록센서스 기반 주요 인구통계의 일관성과 시의성 확보방안 연구, 통계청, 대전.
- 통계청, 2020, 국내인구이동통계, 통계정보보고서(작성일자: 2021년 11월 30일)
- 통계청, 2021, 2020 인구주택총조사 표본 집계 결과: 인구 특성 항목(여성, 출산력, 아동, 인구이동, 통근·통학), 보도자료(보도시점: 2021년 11월 29일).
- 통계청, 2022, 인구총조사, 통계정보보고서(작성일자: 2024년 3월 7일).
- 통계청, 2023, 2022년 인구주택총조사 결과 <등록센서스 방식>, 보도자료(보도시점: 2023년 6월 27일).
- 통계청, 2024a, 장래인구추계, 통계정보보고서(작성일자: 2024년 3월 22일).
- 통계청, 2024b, 장래인구추계(시도편): 2022~2052년, 보도자료(보도시점: 2024년 5월 28일).
- Baffour, B., King, T. and Valente, P., 2013, The modern census: Evolution, examples and evaluation, *International Statistical Review*, 81(3), 407-425. <https://doi.org/10.1111/insr.12036>.
- Bailey, T. C. and Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Longman, Harlow, Essex.
- Bell, M., Bernard, A., Charles-Edwards, E. and Ke, W., 2020, Comparative measures of internal migration, in Bell, M., Bernard, A., Charles-Edwards, E., and Zhu, Y. (eds.), *Internal Migration in the Countries of Asia: A Cross-national Comparison*, Springer, New York. Doi: 10.1007/978-3-030-44010-7_3.
- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J. and Hugo, G., 2002, Cross-national comparison of internal migration: issues and measures, *Journal of the Royal Statistical Society Series A: Statistics in Society*, 165(3), 435-464. doi: 10.1111/1467-985X.t01-1-00247.
- Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J. and Zhu, Y., 2015a, Internal migration data around the world: assessing contemporary practice, *Population, Space and Place*, 21(1), 1-17. doi: 10.1002/psp.1848.
- Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M. and Kupiszewska, D., 2015b, Internal migration and development: comparing migration intensities around the world, *Population and Development Review*, 41(1), 33-58. doi: 10.1111/j.1728-4457.2015.00025.x
- Carmichael, G. A., 2016, *Fundamentals of Demographic Analysis: Concepts, Measures, and Methods*, Springer, New York.
- Charles-Edwards, E., Bell, M., Bernard, A. and Zhu, Y., 2019, Internal migration in the countries of Asia: levels, ages and spatial impacts, *Asian Population Studies*, 15(2), 150-171. doi: 10.1080/17441730.2019.1619256.
- Coleman, D., 2013, The twilight of the census, *Population and Development Review*, 38(supplementary), 334-351. doi: 10.1111/j.1728-4457.2013.00568.x.
- Duke-Williams, O. and Stillwell, J., 2010, Temporal and spatial consistency, in Stillwell, J., Duke-Williams, O. and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*, Business Science Reference, New York. doi: 10.4018/978-1-61520-755-8.ch005.
- Foley, B., Champion, T. and Shuttleworth, I., 2021, How does administrative data compare with census data for the measurement of internal migration? The case of Northern Ireland, 2001-2011, *Applied Spatial Analysis and Policy*, 14, 659-677. doi: 10.1007/s12061-020-09369-w.
- Lee, Y. and Kim, D., 2020, Internal migration in South Korea, in Bell, M., Bernard, A., Charles-Edwards, E. and Zhu, Y. (eds.), *Internal Migration in the Countries of Asia: A Cross-national Comparison*, Springer, New York. doi: 10.1007/978-3-030-44010-7_6.
- ONS (Office for National Statistics), 2016, Methodology used to produce the 2014-based subnational population projections for England, <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/methodologyusedtoproducethesubnationalpopulationprojectionsforengland>, 2024년 8월 13일 접속
- Plane, D. A., 1992, Age-composition change and the geographical dynamics of interregional migration in the U.S., *Annals of the Association of American Geo-*

- graphs*, 82(1), 64-85. doi: 10.1111/j.1467-8306.1992.tb01898.x.
- Rees, P. H., 1977, The measurement of migration, from census data and other sources, *Environment and Planning A*, 9(3), 247-272. doi: 10.1068/a090247.
- Rowland, D. T., 2003, *Demographic Methods and Concepts*, Oxford University Press, Oxford.
- Smith, S. K., Tayman, J. and Swanson, D. A., 2001, *State and Local Population Projections: Methodology and Analysis*, Kluwer Academic/Plenum Publishers, New York.
- Stillwell, J., Dennett, A., and Duke-Williams, O., 2010, Interaction data: Definition, concepts and sources, in Stillwell, J., Duke-Williams, O. and Dennett, A. (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*, Business Science Reference, New York. doi: 10.4018/978-1-61520-755-8.ch001.
- UN (United Nations), 2017, *Principles and Recommendations for Population and Housing Censuses, Revision 3*, Department of Economic and Social Affairs, Statistics Division, New York.
- UNECE (United Nations Economic Commission for Europe), 2018, *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses*, United Nations, New York and Geneva.
- US Census Bureau, 2023, American Community Survey Migration Flows, <https://www.census.gov/data/developers/data-sets/acs-migration-flows.html>, 2024년 8월 13일 접속.
- Valente, P., 2015, From the 2010 to the 2020 census round in the UNECE region - Plans by countries on census methodology and technology, *Conference of European Statisticians, Economic Commission for Europe*, Working paper 24, Geneva.
- 교신 : 조대현, 41566 대구광역시 북구 대학로 80 경북대학교 사범대학 지리교육과(이메일: dhcho@knu.ac.kr, 전화: 053-950-5861)
Correspondence: Daeheon Cho, Department of Geography Education, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Korea (e-mail: dhcho@knu.ac.kr, Phone: +82-53-950-5861)

최초투고일 2024. 7. 4

수정일 2024. 7. 27

최종접수일 2024. 7. 29