

연구보고서 2018-51

# 결혼 및 출산 관련 정책 수립과 평가를 위한 인구패널 구축 타당성 연구



김종훈  
박종서·이소영·최인선

**【연구책임자】**

김종훈 한국보건사회연구원 연구위원

**【공동연구원】**

박종서 한국보건사회연구원 연구위원

이소영 한국보건사회연구원 연구위원

최인선 한국보건사회연구원 연구위원

연구보고서 2018-51

**결혼 및 출산 관련 정책 수립과 평가를 위한  
인구패널 구축 타당성 연구**

발행일 2018년 12월

저자 김 종 훈

발행인 조 흥 식

발행처 한국보건사회연구원

주 소 [30147]세종특별자치시 시청대로 370  
세종국책연구단지 사회정책동(1~5층)

전 화 대표전화: 044)287-8000

홈페이지 <http://www.kihasa.re.kr>

등 록 1994년 7월 1일(제8-142호)

인 쇄 처 고려씨엔피

---

© 한국보건사회연구원 2018  
ISBN 978-89-6827-680-4 93330

## 발간사 <<

급격한 저출산과 고령화가 초래하는 인구 구조 변화의 사회경제적 영향과 충격이 더 이상 간과할 수 없는 수준에 이르렀다는 인식, 이에 대해 국가적 차원에서 정책 대응이 필요하다는 목소리가 일반화된 지도 상당 기간이 흘렀다. 그에 따라 중앙정부의 적극적이고 선제적인 정책 기반으로 인구학적 또는 인구정책적 목적에 충실한 종단형 실증 자료의 포괄적 구축과 체계적 관리가 제안되고 강조된 지도 10여 년이 지났다. 그럼에도 인구패널 구축에 관련된 논의가 실제로 구현되고 정책과 학술연구에 활용되지 못한 것은 아쉽고 의문스러운 일이다.

본 연구는 바로 이 점에 천착하여, 지금까지의 인구패널 논의를 정리하고 실제 구축으로 이어지지 못한 이유를 찾고자 하는 노력의 소산이다. 현시점에서 인구학적 연구 또는 인구정책을 주목적으로 하는 패널 자료의 새로운 구축이 꼭 필요한지, 기존의 다른 대안적 자료의 대응으로는 불충분한지를 명시적으로 묻고 답을 구하는 것은 새로운 패널 자료의 구축뿐만 아니라 기존 인구 관련 자료의 보다 유효하고 적절한 활용을 위해서도 필요한 일이다. 추상적인 논의 수준을 넘어 실제 인구 관련 패널 자료를 구축할 때 현실적으로 부각되는 문제를 제안 사례를 통해 구체적으로 짚은 부분은 이 연구의 미덕이다. 이는 동시에 향후 실자료 구축과 정책적 활용을 위한 유용한 지침으로 활용될 수 있을 것이다.

저출산·고령사회기본계획의 재개정과 본 연구원 주관 전국 출산력 및 가족보건·복지 실태 조사의 새로운 실시를 위한 준비 작업을 앞두고 이 연구가 정책과 학술 연구 안팎으로 적기에 역할을 하기를 기대한다.

본 연구를 위해 많은 관심을 기울이고 조언을 아끼지 않으신 경희대학교의 김현식 교수님, 원내 자문위원으로 큰 도움을 주신 이현주 연구위원께 감사드립니다. 보고서를 성심을 다해 검토, 검독해 주신 원내외 검독위원들과 편집, 교정 작업에 힘써 주신 원내 직원들께도 감사드립니다.

2018년 12월

한국보건사회연구원 원장

**조 흥 식**

# 목 차

Abstract .....	1
요 약 .....	3
<b>제1장 서 론 .....</b>	<b>5</b>
제1절 '인구패널' 논의의 배경 .....	7
제2절 '인구패널'에 대한 인식과 문제의 제기 .....	11
<b>제2장 패널 자료의 일반적 속성과 가치 .....</b>	<b>21</b>
제1절 자료 수집과 패널 구축 .....	23
제2절 표집 범위와 표본의 대표성 .....	30
제3절 패널 자료의 활용 가치 .....	35
제4절 자료 관측 단위, 주기 및 기간 .....	40
<b>제3장 국내외 주요 패널 및 인구 관련 자료 .....</b>	<b>47</b>
제1절 국내 주요 패널과 인구 관련 자료 .....	49
제2절 해외 주요 인구 관련 패널 자료 .....	67
제3절 '인구패널'을 위한 시사점 .....	77
<b>제4장 '인구패널'의 주요 요건 .....</b>	<b>83</b>
제1절 인구 관련 패널의 목표와 기본 성격 .....	85
제2절 인구 변동 패널화 관련 통계학적 고려 사항 .....	107

---

<b>제5장 인구패널 구체화에 대한 제언: 출산력 패널</b> .....	<b>119</b>
제1절 자료 구축 .....	121
제2절 출산력 패널의 활용도와 가치 .....	133
제3절 정책 과제 .....	135
<b>참고문헌</b> .....	<b>139</b>

## 표 목차

〈표 3-1〉 패널조사 계획 및 1차 조사 시작 시기 .....	50
〈표 3-2〉 주요 패널조사 개요(2018년 기준) .....	52
〈표 3-3〉 주요 패널조사의 조사 목적 및 대상 .....	53
〈표 3-4〉 연차별 각 총조사 개요 .....	61
〈표 3-5〉 전국 출산력조사 이력 .....	65
〈표 3-5〉 전국 출산력조사 이력(계속) .....	66

---

## 그림 목차

[그림 2-1] 패널 자료의 형태상 정의 .....	24
[그림 2-2] 패널 자료의 종장형(long-form) 또는 사전식(lexicographic) 표현 .....	26
[그림 2-3] 모집단에 대한 정보 대(對) 표본공간(sample space)의 분할(partition) .....	32
[그림 3-1] 한국노동패널의 원표본과 통합표본의 개요 .....	56
[그림 3-2] GGS의 데이터 관리 구조 .....	68
[그림 3-3] PSFD 조사 대상과 기간 .....	72
[그림 4-1] 주제별 인구패널의 직관적 분류 .....	101
[그림 4-2] 인구 변동의 개념적 분해: 직접-매개-환경 요인과 거시-미시 위계 .....	103
[그림 4-3] 주제별 인구학적 패널의 예: 출산 관련, 가족 관련, 고령화 관련 .....	104
[그림 4-4] Lexis diagram에서 표현된 총조사(2015년) 기반 가임여성(15~49세) 패널 .....	109
[그림 4-5] Lexis diagram에서의 불균등 가임여성 패널 .....	111
[그림 5-1] Lexis diagram에서 표현된 실질 가임여성(25~43세) 불균등패널 .....	128

---

## Abstract <<

### **A Study on Necessity and Feasibility of a New Panel Dataset for Population Policy**

Project Head: KIM, JongHoon

For over a decade, there have been calls on a new, comprehensive panel dataset focused on demographic research and population policy. The attempts so far have come short of materialization, even in the wake of current seismic demographic changes in Korean society and their impact. This paper looks into its causes with particular interest in 1) whether existing data sources are sufficient as alternatives, 2) whether a new panel dataset is inevitable, and 3) if at all, how we could construct a working panel dataset with desirable features.

It is shown that a new demographic panel dataset is essential in order to effectively capture demographic shifts and bursts and their socio-economic impacts, in the structural sense and at the micro-level. The paper suggests longer-than-usual longitudinal datasets separately per each important demographic factor or target population group in an unbalanced panel format as a viable way. A tryout case is outlined for a so-called marriage and fertility panel with its feasible implications on population policies and research.

---

Co-Researchers: PARK JongSuh, LEE SoYoung, CHOI InSun



### 1. 연구 배경 및 목적

인구 변동을 이해하고 설명할 실증 기반으로서 이른바 ‘인구패널’에 대한 수요는 지속적으로 제기되어 왔다. 그럼에도 논의는 초기 수준에 머무르고 있고 실제 패널 자료의 구축은 실현되지 못하고 있다. 이 연구는 그 이유가 무엇인지, 인구학적 연구와 인구정책을 위해 기존 인구 관련 자료의 대안적 활용은 불충분한지, 인구패널 자료의 구축이 필수적인지, 또한 패널 자료로서 인구패널은 어떻게 설계되어야 할지에 대해 문제를 제기하고 논의하는 것을 목적으로 기획되었다.

### 2. 주요 연구 결과

이 연구는, 인구패널 논의와 구축이 지지부진한 이유가 패널 구축의 이유와 목적을 패널 자료의 성격과 활용 가치를 기준으로 확립하지 못하고, 인구 변동의 장기 시계열적 고유 특성을 제대로 고려하지 못한 데에 있음을 밝힌다. 사회 현상의 일부로서 인구 구조 변화의 원인과 결과를 구조적이고 미시적인 수준에서 설명하려면 인구학적 패널 자료의 구축과 활용이 불가피함도 보인다. 또한, 인구학적 패널 자료의 설계에 반영되어야 할 실천적 문제와 통계적 난제들을 표본 대표성, 자료의 가변성을 중심으로 정리한다.

### 3. 결론 및 시사점

인구학적 패널은 목적 지향적인 미시-종단 증시형 패널 자료로 구축되어야 한다. 인구 현상 전체(출생-사망-인구 이동)를 아우르는 인구패널은 유용하기도 가능하기도 않으므로 주제별, 대상 인구집단별 패널을 구축하는 것이 적절하다. 실천적으로는 자료 관측 주기와 기간에 대해 명시적으로 고려한 불균등 패널이, 결혼·출산에 관한 패널 구축을 예시로 제안된다. 자료의 인구학적, 인구정책적 활용 가치도 논의된다.

\*주요 용어: 인구패널, 인구학적 패널, 임신·출산 관련 패널, 미시 패널, 횡단면 자료, 전수 자료, 표본 대표성, 자료의 가변성, 시계열 특성, 관측되지 않는 개체 특성, 선택편의, 동태-구조적 관계, 불균등 패널, 완결인구력

제 1 장

서론

제1절 '인구패널' 논의의 배경

제2절 '인구패널'에 대한 인식과 문제의 제기



## 제1절 ‘인구패널’ 논의의 배경

‘인구패널’에 대한 언급, 구축 필요성과 계획은 제1차 저출산·고령사회 기본계획(2006-2010년)의 초기 추진 과정에서 등장한다(이삼식, 정경희, 신인철, 김정연, 2006, p. 36). 우리 정부는 2006년 저출산·고령화 대응 정책의 기초 인프라 구축 방편으로 기본계획에 ‘인구패널’ 사업을 포함시키면서 2007년 인구패널 구축을 위한 준비사업, 2008년부터 실시할 본조사를 예고하였다(이삼식 외 2006, p. 36). 이를 실행하기 위한 기초 연구(이삼식 외(2006), 준비 사업으로 패널 구축 예비조사(보건복지가족부, 한국보건사회연구원, 2008)가 뒤따랐다.

정책 활용의 기초자료로서 인구패널 수요가 이 시기에 집중된 것으로 미루어 보건대, 인구패널에 대한 관심과 논의는 2000년 이전에 시작된 것으로 볼 수 있다. 우리나라가 경험적 연구와 실증 분석에서 패널 자료라는 구체적 형식의 필요성과 가치에 주목하게 된 것은 1990년대 중반 이후이다(아래 제3장 제1절 참조). 그에 따라 연구 주제와 대상별로 각종 패널 자료가 본격적으로 구축되고 학술, 정책 연구에 활용되기 시작한 것은 2000년대 초반(이희길, 2007 또는 아래 제3장 제1-1절 <표 3-1> 참조)이다. 인구패널의 경우, 같은 시기에 저출산과 고령화 현상이 가속화되고 ‘인구문제’가 사회적 이슈로 부각된 배경 요인도 크게 작용했다. 저출산·고령화의 원인과 전개 양상, 사회경제적 파급효과에 대한 관심과 우려가 정책적 해법에 대한 기대와 요구로 숨 가쁘게 이어지면서 인구정책

의 실증 기반 자료로서 '인구패널'에 대한 요구는 더더욱 탄력을 받았다.

그럼에도 불구하고, '인구패널'의 구축은 현재까지 실현되지 않고 있다. 인구통계 생산과 경험연구에 활용되는 인구 관련 자료는 이미 다수 존재한다. 이들은 공식적 인구통계 집계가 시작된 1970년, 멀게는 인구정책이 시작된 1960년대 초기부터 생산, 축적된 자료들로부터 시작되어, 정치, 행정 제도를 비롯한 유무형의 공공재나 국방, 교육, 교통, 보건의료 같은 공공서비스 제공의 기초로 구축된 초보적 인구통계로까지 이어진다.

인구 관련 조사-통계 자료 가운데 가장 기초적인 이 자료들을 인구 현황 추적과 인구 변동 관측을 주된 목적으로 하는 '인구 기초자료'라 볼 수 있다. 이들은 '인구학적 주제(인구동향, 인구구조 변화 실태)에 대해 정책적 활용(정책 수립 또는 정책 대응의 근거)을 목적'으로 구축된 1차적이고 직접적인 인구 관련 자료이다. 대표적인 인구 기초자료에는 통계청이 주관하는 인구주택총조사, 인구동향조사, 국내 인구이동통계, 출입국자 및 체류외국인 통계, 장래인구(가구)추계, 생명표, 주민등록인구현황, 경제활동인구조사가 있다.

이들은 단순 집계의 결과이자 집계 통계의 가공으로 산출된 결과물로서 국가 인구통계의 핵심 구성 요소이다. 이들 인구 기초자료를 통해 연령구조, 성별, 코호트 등 인구학적 분류 기준에 따라 시점별로 인구 현상의 정적인 윤곽을 관찰하고 파악할 수 있다. 더 나아가 동태적 인구 변동의 과정과 결과에 대한 분석도 인구학적 범위 내에서 어느 정도 가능하다. 출산력(사망력) 변화 추이에 따른 인구구조 변화 전망, 인구구조 변화에 대한 인구 변동요인(출생, 사망, 인구 이동)들의 기여도 분해 등이 그 예이다.

하지만 인구구조와 인구동태를 인구학적 요인들로 설명하는 것만으로는 인구 변동의 근본 동력과 기제를 이해하기에도 불충분하고 인구 관련

정책의 수립, 평가, 환류에 의미 있는 함의를 찾기도 미흡하다. 인구동태와 구조 변화는 사회, 경제 변화와 떼어 놓고 생각할 수 없고, 인구 변동(요인)의 배경 요인이자 매개가 되는 사회경제적 요소들, 예를 들어 결혼 행태, 가족 구성, 교육 수준, 경제활동과 여건, 사회문화적 가치관들과의 직간접적 상호작용을 떼어 놓고 인구 현상을 온전히 이해할 수 없기 때문이다.

대체로 인구학적 관점에 충실한 인구 기초자료들이 학술적으로든 정책적으로든 문제 해결에 충분하지 못하다는 인식이 여기에서 비롯된다. 인구와 관련된 많은 실증 연구에서 관심 주제가 인구학적 영역에 국한되지 않고, 그에 따라 다양한 탈(脫)인구학적 조사-통계자료들이 채용되고 있는 것이 그 방증이다.

인구 기초자료나 인구학적 목적이 일차적인 자료는 아니지만 인구적 요소와 함의를 담고 있는 이런 종류의 자료를 ‘인구학적 인접 자료’라고 부른다면, 이들은 더 느슨하고 넓은 의미에서 ‘인구 관련 자료’에 포함된다. 이러한 성격의 자료는 실태조사, 동향조사 형태로 다양하게 있는데, 인구 관련 연구에 실제 활용되는 대표적인 것들로, 전국 출산력 및 가족보건·복지 실태조사(한국보건사회연구원), 노인실태조사, 아동종합실태조사(보건복지부), 가족실태조사(여성가족부), 신혼부부통계·생활시간조사·가계금융·복지조사·가계동향조사·주택총조사(통계청), 주거실태조사(국토교통부), 사업체노동력조사[구(舊) 사업체임금근로시간조사]·고용형태별근로실태조사[구(舊) 임금구조기본통계조사](고용노동부), 교육기본통계(한국교육개발원), 건강보험통계(국민건강보험공단), 국민보건의료실태조사(보건복지부) 등을 들 수 있다.

이들 자료에서 간과된 시계열(time series)적 특성, 즉 동태적 변화도 함께 식별하기 위해 중단 방향(longitudinal)의 관측을 더한 것이 패널

(조사)자료이다. 패널 자료는 취지와 구성 자체로 인해 일반적 실태, 동향 조사자료들에 비해 대체로 '인구 관련성'이 더 강하고 실제로 인구 실증 연구에 더 적극적으로 자주 활용된다. 대표적인 예로는 재정패널조사(한국조세재정연구원), 한국노동패널조사(한국노동연구원), 한국복지패널(한국보건사회연구원-서울대학교 사회복지연구소), 여성·가족패널조사(한국여성정책연구원), 한국아동·청소년패널조사(한국청소년정책연구원), 국민노후보장패널조사(국민연금공단), 한국고령화연구패널조사(한국고용정보원)가 있다.

인구학적 활용도가 높은 패널 자료들 간에도 인구 인접성에서는 상대적으로 차이를 보인다. 관측의 단위가 기본적으로 가구 또는 개인이라는 점에 기인하는 인구학적 함의나 미시적 특성은 공통적이지만, 각 자료의 주제와 목적이 인구적 관점을 얼마나 적극적으로 반영하는지는 각각 다르다. 아예 생애주기상의 특정 기간(국면)이나 연령대별·성별 특정 인구 집단을 대상으로 하는 고령화패널, 여성가족패널, 아동·청소년패널, 노후보장패널 등이 적극적 인구지향형이라면, 재정패널, 노동패널, 복지패널 등의 목적의식에서의 인구학적 관점은 간접적, 부차적이다.<sup>1)</sup>

---

1) '인구패널'을 새로이 구축해야 할 자료로 보는 대신 기존의 패널 자료들을 인구학적 성향에 따라 분류, 연계한 느슨한 개념으로 보는 관점도 존재한다. 보건복지가족부·한국보건사회연구원(2008)은 이러한 관점을 부분적으로 반영하고 있다.

## 제2절 ‘인구패널’에 대한 인식과 문제의 제기

### 1. 기존 인구 관련 자료에 대한 판단

‘인구패널’을 새로이 구축해야 한다는 주장의 바탕에는, 기존의 인구 관련 자료들이 정책적 활용뿐만 아니라 인구 변동의 관찰과 분석에도 미흡하다는 인식이 깔려 있다.

인구동향조사, 인구주택총조사 등 현존하는 인구 기초자료들은 인구 현상에 대한 일차적인 관찰과 인구 기초통계의 생산에는 충실하게 구축되어 있지만, 인구학적 범위를 넘어선 사회경제적, 문화적 현상으로 인구 변동을 이해하는 데에는 부족한 것으로 이해된다. 인구구조 변화를 사회 변동, 경제성장과 경기변동, 문화와 가치관 변화와 이어 주는 주요 매개 요인과 변수들, 즉 결혼, 가족 구성, 교육 수준, 경제활동, 사회경제적 지위와 계층 차이, 사회적 가치관 등을 담지 못하기 때문이다.

또한, 기존에 활용되어 온 인구 관련 자료들은 인구학적 기반 자료로서 각각 한계가 뚜렷한 것으로 인식된다. 인구적 관점이 분명한 경우는 그 대상과 주제가 제한적이거나 인구 변동의 동태적 추적과 식별에 적합하지 않은 구조가 문제가 된다. 인구구조 변화가 주요한 관련 주제이더라도 부차적이거나 자료 구축의 주된 목표가 탈(脫)인구학적 주제에 초점이 맞추어진 경우도 있다. 인구 외적 주제가 중심인 자료들은 인구학적으로 중요하지만 관심 영역 밖인 주제와 변수들을 놓치는 한계가 있음이 흔히 지적된다.<sup>2)</sup>

2) 이 한계들이 이들 자료들의 확장 가능성을 배제할 정도로 본원적인지는 확실치 않다. 예를 들어 전국 출산력 및 가족 보건·복지 실태조사를 패널 형식으로 확장, 개선해서 인구 패널의 대안으로 삼는 것도 하나의 선택지일 수 있다. 이와 관련된 논의는 아래 제4장 제2-1절과 제5장 제1절을 참고할 수 있다.

이삼식 외(2006)는 ‘한국인구패널(KPDP: Korea Population Dynamic Panel)’의 새로운 구축이 필요한 배경 논리를 다루면서 이와 같은 인식을 분명히 적시하였다(pp. 99-108). 우리나라는 1990년대 중반에 이르러 양적 통제 위주 인구억제 정책을 공식적으로 폐기하는데 그 이후 10여 년 동안 저출산과 인구구조 고령화 현상이 본격적으로 심화하는 과정에서 인구정책 전환의 시점을 놓치고 전환 방향을 확립하지 못한 것으로 평가된다. 보고서는 그 주된 이유가 “인구 변동과 그 원인과 결과에 대한 지속적인 관심과 노력의 미흡”(이삼식 외, 2006, p. 101) 때문이라고 판단하는데 구체적으로 “인구 변동과 사회경제현상 간 불가분의 관계가 있음에도 불구하고 ... 인구 규모 및 구조에 직접적인 영향을 미치는 결혼, 출산, 사망, 국제이동 등 인구동태적 (배경) 요인의 변동에 대한 전망이 거의 이루어지지 못했으며 ... 장시간 동안 인구감시체계 (population surveillance system)가 전혀 작동되지 못하여 ... 급기야는 인구 변동과 그 원인 및 결과에 대한 데이터베이스의 고갈로 이어졌다.”(이삼식 외, 2006, p. 101)는 진단을 내린다.

이삼식 외(2006)는 가용한 인구 관련 자료를 인구주택총조사, 인구동태통계조사[현(現) 인구동향조사]와 전국 출산력 및 가족보건·복지 실태조사(이하 출산력조사), 전국 결혼 및 출산 동향조사<sup>3)</sup> 정도로 좁혀 보고, 이들 기존 자료로 인구 변동—특히 보고서에서 주목한 저출산 심화 현상—을 사회경제적 변화에 유기적으로 연계하여 설명하기에는 한계가 있음을 지적한다. 우선 이들 자료가 횡단면(cross-section) 조사로 인구구조와 분포 등 관측 시점의 현황(snapshot)을 포착하기에는 적절하지만, 인

3) 이삼식 외(2006) 연구가 작성된 시점에는 2005년 첫 조사가 완료된 상태였다. 이후 두 차례(2009년, 2012년) 더 조사가 실시되었지만 정례적인 (조사)자료로 자리 잡지는 못했는데, 이는 내용과 성격에서 출산력조사, 여성·가족패널 등 인접 (조사)자료와 중복되지 않으면서 차별화된 고유 가치와 위상을 확보하지 못했기 때문으로 판단된다.

구 동태, ‘인구 현상과 사회경제현상 간의 복잡하고 동적인’ 상호작용과 같은 시계열(time-series)적, 종단적(longitudinal)인 관계와 영향을 파악하기에 한계가 있다고 평가한다. 더불어 급격한 인구 변동의 구조와 원인 규명에 필요한 ‘사회경제적 변화와의 복합적인 인과관계’를 관찰하고 분석하는 데에도 이들 자료로는 부족하다고 본다.

이러한 횡단면 자료로서의 한계와 단점으로 인해 이들 자료들만으로는 인구 변동 요인의 변화와 원인 및 영향을 측정하는 것도 거의 불가능(p. 104)하고 중장기적 관점 ... 인구동학적 차원에서 인구 현상과 사회현상을 연계할 수 있는 기초 자료의 생성이 현실적으로 불가능하다(p. 104)고 진단하고, 새로운 인구 관련 패널 구축이 필요하다(p. 106)고 제안한다(이삼식 외, 2006).

인구 변동의 원인과 인구구조 변화로 인한 파급효과를 동태적, 종합적으로 살피기에 기존의 인구 관련 자료들로는 부족하다는 인식은 인구패널 논의 초기에서 크게 바뀌지 않았다. 인구패널 대신, 경험연구의 필요에 따라 다양한 인접 자료들을 보완적으로 활용하는 경로를 따르면서 인구 관련 자료의 종류가 크게 늘어나게 된 최근에도 이는 달라지지 않은 것으로 보인다. 어떤 인구학적 가설(청년 주거 부담이 줄면 결혼과 출산이 늘어날 것이다.)을 실증하려는데, 이에 필요한 개인의 동태적 선택과 행위(주거 특성이 반영된 결혼, 가족 형성)에 대한 시계열 추적 자료가 없다는 불만은 흔히 접하는 사례이다. 인구 관련 자료(재정패널)를 이용하여 인구적 함의(부부의 자산과 부채 수준 변화와 임신, 출산 수준 변화의 관계)를 얻고자 하는데 인구 기본 변수(임신 횟수, 출산 진도)가 빠져 있어 단순한 변화(출산 경험 유무)를 넘어선 심도 있는 분석에 한계를 보이는 경우도 대표적인 예이다.

하지만 새로운 인구패널 구축이 시급하고 불가피하다고 본, 초기의 질박함이나 적극성이 현재에도 여전히 유효한 견해인지는 확실하지 않다. 기존 자료들의 한계와 단점이 새로운 자료의 구축이 아니면 해소될 수 없는 종류인지, 또한 그러한 자료의 미비가 얼마나 위중한 문제인지 제대로 논의된 적이 없다. 이삼식 외(2006)는 ‘한국인구패널’의 기본 틀과 구성 요소에 대한 열개를 논의하였지만 그 내용을 뒷받침할 만한 논리적, 실용적 근거를 제시하지는 않았다. 이후 간헐적으로 학술적, 정책적 필요성이 제기되는 경우에도 인구패널 구축의 당위성과 타당성, 구축 방향에 대한 분석적 판단과 근거를 찾아보기는 어렵다. 이 때문에 인구패널 구축이 정말 필요한지, 필요하더라도 구축의 비용을 감당할 만큼 효용이 충분한지에 대한 의구심과 회의적 시각은 여전히 해소되지 못한 채로 남아 있다.

## 2. ‘인구패널’에 대한 인식

인구패널이 어떤 모습이어야 하는지, 즉 인구패널의 형태와 구조, 구성과 내용이 어떠해야 하는지는 인구패널의 구체적인 필요성과 활용 가치를 기준으로 결정되는 것이 타당하다. 하지만 정책담당자나 연구자들이 인구패널에 대해 가지고 있는 이미지, 인구패널에 대한 일반적인 인식이 그러한 엄밀한 근거와 이상적 조건에 맞게 형성되어 있는 것은 아니다. 이 연구는 그 간극을 살펴 문제를 제기하는 것으로부터 인구패널 논의를 시작한다. 그런 의미에서 이삼식 외(2006)에서 제안한 ‘한국인구패널’의 구조와 내용은 비교 기준이자 출발점으로 삼기에 적당하다. 최초로 제시된 인구패널의 구체적 방안으로서 인구패널에 대한 일반적인 인식의 기본 요소들을 고루 갖추었고 이들이 현재에도 대체로 유효해 보이기 때문이다. ‘한국인구패널’의 주요한 특징은 다음과 같다.<sup>4)</sup>

- 1) 매년 실시하는 가구 단위 패널조사이다.
- 2) 조사 주제가 출생에서 사망까지 전 생애주기에 걸친 인구 변동 과정 [결혼, 가족 형성, 출산(출생), 보육, 교육, 경제활동, 노후, 건강]을 아우른다.
- 3) 조사 규모가 방대할 수밖에 없어 패널화된 기본조사(핵심 주제)와 부정기적, 비패널 부가조사 또는 특별조사(기타 주제)로 구획한다.
- 4) 인구학적 편이에 따라 생애주기를 출산·양육에 관련된 전반부(성장과 교육, 취업과 경제활동, 결혼과 가족 형성, 임신·출산·양육)와 노후 생활에 연관된 후반부(노후소득보장, 노후 건강, 생활환경, 사회 참여)로 나누어 하위 주제별로 모듈화하고, 대체로 전반부를 기본조사 주제로, 연간 변동성이 완만한 출산력 관련 일부 주제나 인구 이동 관련 주제, 후반부는 대체로 부가·특별조사 주제로 배정한다.
- 5) 조사 대상 또한 인구학적 편이에 따라 가구, 기혼자, 미혼자, 노인, 청년으로 나누고, 이 분류대로 조사 자체를 개별적으로 시행한다.

패널조사 자료의 조사 주기를 연 단위로 삼은 것은 다른 패널 자료의 사례를 그대로 따른 것으로 이해된다. 가구 단위 조사와 기본조사-부가조사 구성을 제안한 것은 인구주택총조사와 개념적으로 조응하는 측면이라 할 수 있다. 전수조사(census)인 인구주택총조사와 달리 표본조사이고 조사 주기와 규모에서 분명히 차별되지만, 조사 대상자 수와 조사 분량(내용과 범위)이 여전히 방대한 규모일 가능성을 감안할 때 인구주택총조사의 선례를 차용하는 것은 일견 무리가 없는 선택으로 보인다.

한편, 이 인구패널은 정책적 활용, 즉 정책 발굴, 시행 평가, 효과성 평

---

4) 세부적인 방안이나 지침이 보건복지가족부, 한국보건사회연구원(2008)에서 구체화된 경우에는 후자의 내용을 평가하였다.

가 및 환류가 자료 구축의 주된 목적인 이른바 ‘정책패널’의 성격이 강한 점이 특징이다. 이는 패널 자료 구축의 핵심 목적과 비전에 명시되어 있다(이삼식 외, 2006, pp. 105-108). 무엇보다도, 인구패널 분석의 기본 틀과 핵심 조사 주제 분류 및 구성에서 정부의 종합 인구정책 또는 저출산·고령화 대응 정책의 기본 관점과 정책 체계 구성 원리<sup>5)</sup>를 그대로 따른 점에서 확연히 드러난다.

하지만 이러한 강한 정책 지향성에도 불구하고 ‘한국인구패널’ 실행안은 관련 정책의 기획, 시행 평가, 효과성 검증에 바로 사용할 수 있는 형태가 아니다. 조사 대상자(개인, 가구)별로 조사 항목들이 분석될 수 있도록 식별되고(identified), 변수화되어 있는 상태는 아니기 때문이다. 오히려 실증분석에 적용할 정책패널(2차 자료)로 재가공하기 이전의 1차 원자료에 더 가깝다. 이는, 인구패널 설계의 철학적 바탕에 ‘인구 변동 요인 분석을 위한 미시적 시계열 인구자료 생산’이라는 일차적 목표와 독립적으로 ‘인구 변동과 정책 수단에 관련된 조사 가능 항목을 모두 포괄하는 인구 기초자료 집계’라는 부차적, 암묵적 목적이 포함되었기 때문으로 판단된다. 이 부차적 목적은, 조사 규모의 부담이 실질적, 경제적인 제약으로 작용하는 경우 “쉽지 않은 대형 조사인데 당장 요긴하지는 않더라도 관련 정보는 일단 최대한 수집해 놓자.”-는 현실적 요구가 자연스럽게 반영된 것이다. 그 결과, ‘한국인구패널’의 설계에는 인구 변동에 관한 주요한 정보를 모두 관측하고 추적하려는 암묵적 동기가, 분석적 수요 충족이

---

5) 저출산·고령사회기본계획(1~3차)은 인구구조 변화가 사회·경제 변화의 원인이자 결과로서 미시-거시적으로 연계되어 있고 사회경제적 변화와의 복합적, 동태적, 순환적 상호작용을 통해서만 온전하게 이해할 수 있는 사회현상의 일부라는 기본 시각을 견지한다. 이를 바탕으로 한 정책적 체계는, 생애주기별로 전반부에 대한 정책-출산·양육을 중심으로 한 저출산 대책-과 후반부에 대한 정책-노후 생활을 대상으로 하는 고령화 대책-으로 우선 나누고 각각의 핵심 영역을 주요 해당 정책 과제 위주로 설정하는 방식을 따른다.

라는 본연의 명시적 목적보다 더 지배적으로 드러나게 되었다.

### 3. 문제 제기

이삼식 외(2006)의 ‘한국인구패널’ 안을 중심으로 인구패널에 대한 일반적 인식의 내용과 요소들을 살펴보면, 아래와 같은 기본적인 의문들이 제기된다.

- 1) 이러한 형태로 인구패널을 구축하면 인구 변동의 원인과 파급효과를 복합적, 동태적으로 설명하고 실효성 있는 정책 진단을 내릴 수 있는 기반 자료 역할을 할 수 있는가?
- 2) 기존 횡단면 위주 인구 관련 자료를 확장하고 개선하는 대신 새로운 패널 자료를 구축하는 것이 꼭 필요한가?
- 3) 새로운 자료의 구축 그리고 그 규모에 따른 비용과 노고의 부담을 정당화할 충분한 실질적 편익이 보장되는가?

인구학적 연구를 위해 이전과 다른 형식(패널)의 새로운 자료 구축이 불가피해지려면 기존의 인구 관련 자료로 불가능한 어떤 관측, 분석, 합의 도출이 요구되고 가능해져야 한다. 기존 자료의 확장이나 개편으로 극복하지 못하는 단점과 한계가 분명해야 하고 새로운 자료로 할 수 있는 일들이 구체적이고 요긴한 것이어야 한다. 방대한 인구 관련 주제와 영역들 가운데 중요한 부분이 패널화가 어렵거나 중단, 동태 분석이 필요하지 않다면 인구패널 구축을 정당화할 근거는 없어진다. 새로 구축한 패널 자료가 기존의 대형 표본조사나 전수조사 자료를 역사적으로 축적한 것과 큰 차이 없다면 기존 자료를 개선하고 보완 자료와 함께 사용하는 것이

더 나올 수도 있다. 새로운 인구패널의 시범 사업 단계에서 인구패널의 활용 목적과 구축 방향이 확립되지 못하고 출산력 또는 출산 관련 패널 구축 위주로 초점이 좁혀지는 모습(보건복지가족부, 한국보건사회연구원, 2008, pp. 65-66)은 명분과는 달리 인구패널의 구체적인 목표와 활용 가치에 대한 판단이 결여되어 있기 때문이다.

인구패널이 인구 변동을 추적, 관측하는 주기가 1년 단위여야 할 근거는 무엇인지, 인구 현상을 관측하는 단위가 가구일 필요성은 무엇인지, 조사 대상을 가구, 기혼자, 미혼자, 노인, 청년 등 집단별로 나누어야 할 이유는 무엇인지 불분명한 것도 먼저 그 근거에서부터 문제로 고민하지 않았기 때문이다. 패널 자료는 횡단면 자료와 다른 특성이 있고 그에 따라 자료 표본의 모집단 대표성 유지를 비롯한 자료 유지, 관리 방식에서 차이점도 적지 않을 텐데 그에 대한 고려가 미흡한 점도 마찬가지다.

이런 의문들이 제기되는 이유 그리고 그 답과 근거가 대체로 명확하지 않은 근본적 이유는 크게 두 가지이다. 첫째, 인구패널에 대한 일반적 인식 안에는 일관되지 않은 다양한 기대, 때로는 상충되는 견해들이 뒤섞여 있다. 무엇을 위해 패널을 구축하고 이 패널 자료로 실제 무엇을 할 수 있는지가 자료의 기획과 설계 단계에서부터 구체화되고 명시적으로 반영되어야 할 텐데 지금까지의 논의에서 이 부분은 보이지 않는다. 둘째, 기존의 인구 관련 자료들에 대한 판단이 빠져 있다. 횡단면 위주이고 인구 관점의 분석과 함의에 필요한 충분한 내용을 담지 못하고 있다는 단점과 한계가 지적되었을 뿐, 패널 형식의 인구 자료가 불가결한지, 기존 자료들이 '패널화'나 개편을 통해 문제점을 해결하는 것이 불가능한지, 새로운 인구패널을 구축할 필요성과 가치가 충분한지, 새로운 자료가 구축되면 기존 자료와의 관계는 어떻게 설정해야 하는지 등에 대한 판단은 결여되어 있다.

이 보고서는 위 두 주제를 중심으로 심도 있는 논의와 판단을 시도하는 것을 첫째 목표로 삼는다. 인구패널이 무엇이고 어떤 내용과 구성 요소를 갖추어야 하는지를 인구패널의 가치와 용도를 기준으로 분명하게 제시한다. 둘째 목표는 실용적 가치가 있는 인구패널이 갖추어야 할 전제 조건을 살펴보는 것이다. 그 조건을 기준으로 새로운 인구패널 자료를 구축하는 것이 필요한지, 효용성으로나 효율성으로 보아 타당한지, 기존의 인구 관련 자료들은 어떻게 활용해야 할지를 살펴본다. 셋째 목표는 구체적인 형식과 구조, 구성 요소들과 내용을 갖춘 인구패널의 예를 제안하는 것이다.

이를 위해 제2장에서는 ‘인구패널’에 패널 자료의 형식과 틀이 왜 필요한지와, ‘인구패널’이 패널 자료로서 갖추어야 할 최소한의 조건과 모습을 살핀다. 이를 근거로 주요 국내외 인구 관련 자료와 선행 패널 자료들의 특징과 성격을 제3장에서 정리하고 새로운 ‘인구패널’ 구축을 위한 시사점을 도출한다. 제4장에서는 인구학적 패널 자료가 본 목적과 취지에 부합하는 실용적 가치와 활용성을 발휘하기 위해 만족해야 할 조건들을 살펴본다. 인구학적 실증 연구를 목적으로 하는 자료에서 별도로 고려해야 할, 다른 분야의 자료와 달리 두드러지는 통계적, 계량경제학적 특징과 유의점도 짚어 본다. 제5장에서는 이러한 논의를 바탕으로 인구학적 패널을 실험적으로 구축해 본다. 이를 통해 구체적인 형태와 구조, 내용은 어떤 모습일지, 어떤 활용 방안을 기대할 수 있는지 살핀다.



## 제 2 장

# 패널 자료의 일반적 속성과 가치

제1절 자료 수집과 패널 구축

제2절 표집 범위와 표본의 대표성

제3절 패널 자료의 활용 가치

제4절 자료 관측 단위, 주기 및 기간



# 2

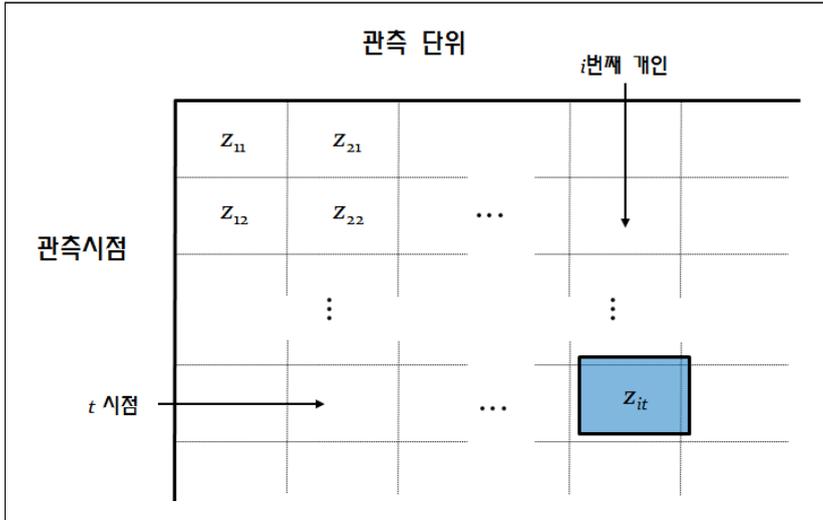
## 패널 자료의 일반적 속성과 << 가치

### 제1절 자료 수집과 패널 구축

인구패널에 대한 일반적인 통념은 횡단면 위주의 인구자료에 종단 방향의 관측을 더한 자료라는 인식이다. 이는 인구패널에 대한 가장 넓고 느슨한 관점으로서, 여러 시점에 걸쳐 동일한 조사를 반복해서 2차원 평면(panel) 형태로 ‘짚은’ 것 자체를 패널 자료(panel dataset)로 보는 것이다. 한 가지 조건은 관측 기간을 통틀어 관측의 대상(표본)이 동일하게 유지되어야 한다는 점이다. 기존의 인구 관련 각종 실태, 동향 조사들도 정기적으로 수집한 결과물의 역사적 축적이란 점에서는 다르지 않지만 동일한 관측대상(표본)을 시간의 경과를 따라 추적한 기간 간(period-to-period) 연속성(또는 시계열적 연속성)이 결여되어 있다는 점에서 패널의 요건을 충족하지 못한다.

일반적으로도, ‘관측 시점마다 수집한 횡단 자료(cross-sectional dataset)를 동일한 대상(표본), 구성과 내용을 유지하면서 일정 기간 주기적으로 반복해서 축적한 결과물’은 넓은 의미에서의 패널 자료의 정의(定義)이자 필요조건이 된다(Arellano, 2003, p. 1). 이에 따라 가장 기초적인 형태의 패널을 다음 [그림 2-1]과 같이 도식화할 수 있다. 여기에서  $z_{it}$ 는  $t$ 시점에서 관측단위(개인, 가구 등)  $i$ 의 모든 관측치를 담고 있는 (행)벡터를 의미한다.

[그림 2-1] 패널 자료의 형태상 정의



자료: 저자 작성.

하지만 이 정의에 따른 2차원 자료의 형성과 축적 자체를 패널 자료 구축의 목표이자 그 완결로 보는 것은 다른 문제이다. 언뜻 상식적이고 중립적으로 보이는 이 관점에 따라 패널 자료의 목적과 기본 방향이 결정되기 때문이다.

자료 구축 자체만이 목표라면 가장 자명하고 직관적인 패널 자료는 대규모 1차 원시자료 패널이다. 기초통계 생산이든 (정책)지표 산출이든 혹은 인과관계 실증이든, 활용 목적과 독립적으로, 연관성이 있는 미시적 기초 정보를 가능한 한 많이, 시간의 경과에 따라 지속적으로 축적한 결과(database) 자체를 패널로 보는 것이다. 특정 활용 목적에 국한하지 않은 개방적 구조를 전제하면, 이 패널은 변수(variable)로 가공되기 이전의 정보나 변수로 식별하기 쉽지 않은 내용까지 포함한 기초자료의 성격을 띠게 된다. 따라서 이 패널을 분석자료로 활용하려면 필요한 부분을

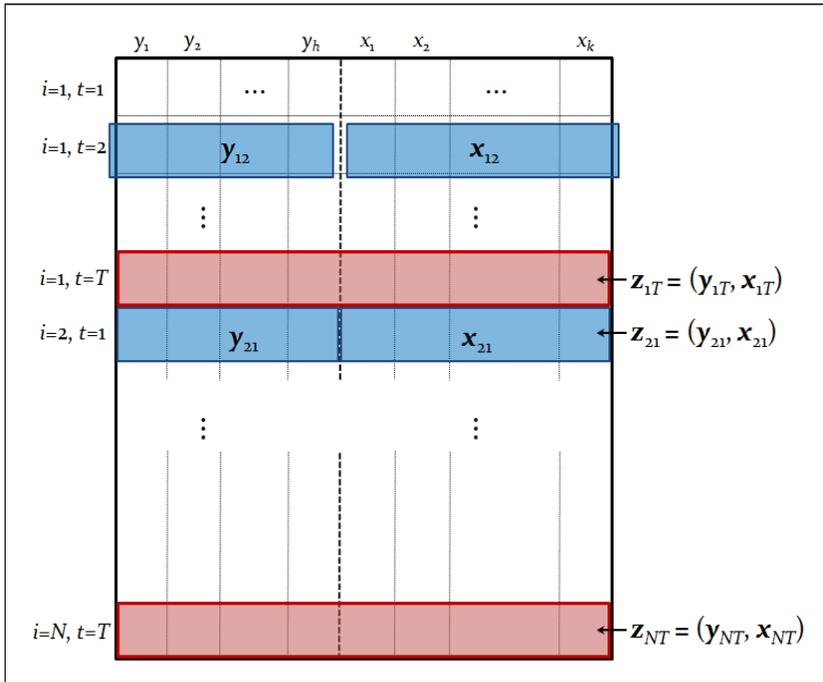
활용 목적에 맞도록 조작하여 변수화하고 계산이나 분석이 가능한 2차 '목적패널'로 재가공하는 것이 필요하다.

1차 (원시)자료 패널의 핵심 이슈는 조사(관측)의 방법과 절차, 패널 유지·관리의 비용과 효율성이다. 가능한 한 많은 정보와 내용을 담고자 하는 확장 성향 때문에 1차 자료는 표본의 크기, 조사 주제와 항목 수가 커지고 그에 따라 자료 전체의 규모가 커지는 것이 일반적이다. 이러한 자료의 질적 가치는 활용도보다 자료(표본)의 모집단 대표성을 기준으로 결정되기 쉽다. 가능하면 모집단 전체에 대한 관측(전수), 아니면 모집단의 확률분포를 최대한 모사(模寫)하는 분포 특성을 따르는 큰 표본이 선호된다. 따라서 1차 자료 패널은 대규모 자료의 관측과 축적, 특히 대규모 표본의 추출과 관리가 핵심이 되고 표본의 모집단 대표성이 가장 중요한 요건이 된다. 이는 대규모 횡단면 조사 자료들이 흔히 따르는 기본 관점과 같다.

패널 전체 자료(pooled dataset)를 분석의 편의를 위한 통상적인 형태, 즉 각 변수가 열벡터로 식별되도록 자료를 종장형(long-form)으로 재배치하면 [그림 2-2]와 같다. 패널 자료의 암묵적 체계(system)를 기준으로 내생적(endogenous) 또는 종속적 변수들과 외생적(exogenous) 변수들로 개념적으로 나누어 본다면, 그림에서  $y = (y^1, y^2, \dots, y^h)$ ,  $x = (x^1, x^2, \dots, x^k)$ 는 각각 내생변수와 외생변수들의 (행)벡터를 의미한다. 따라서 이 패널에 포함된 변수는 모두  $h + k$ 개이다. 또한, [그림 2-1]과의 관계는  $z = (y, x)$ 로 요약된다. 그러면, 확장 성향의 1차 자료 패널의 경우 조사 주제와 항목 수의 증가는 변수의 개수( $h + k$ ) 증가로 나타난다. 1차 자료 패널은 표본 규모( $N$ ) 또한 큰 모습을 띠 것이다. 전통적인 패널 자료에서 패널의 규모를 결정하는 것은 횡단의 크기, 즉 표본의 개체수( $N$ )이고 종단 크기인 관측 시점 개수( $T$ )는 비교적 작은 크기로 제한

되기 때문이다. 이렇게 본다면 1차 자료 성향의 패널은 대규모 횡단면 자료와 외형상 비슷한 규모 특성을 가지게 된다. 이 패널에서  $T=1$ 이면 횡단면 자료로 되돌아간다.

[그림 2-2] 패널 자료의 종장형(long-form) 또는 사전식(lexicographic) 표현



자료: 저자 작성.

1차 자료 성향 패널에서 종단 방향의 자료 축적을 정당화하는 명분은 횡단면에 치우쳤던 기존의 관점에서 종단분석(시계열분석)이 가능하도록 분석의 시야와 방법론이 확장될 수 있다는 논리로 대체로 귀결된다. 이때 흔히 상정되는 분석 방식은 여러 시점들을 대조하거나 비교하는 방법들이다. 통계적 기술(descriptive statistics)을 통한 비교 분석, 패널 전체

자료(pooled dataset)를 활용한 비교정태분석(comparative static analysis)이나 인구학적 요인분해를 비롯한 미실현-시나리오 분석(counter-factual analysis)이 유효한 기법으로 거론된다.

자료 구축 자체에 비중을 두고 활용 여지는 크게 염두에 두지 않는 이러한 성향은 현존하는 대표적인 국내 패널 자료들의 구성 원리와 활용 방식에서 공통적으로 나타난다. 이 자료들은 대규모 조사를 지향하는 점, 특히 조사의 주제와 범위를 결정할 때 사후적인 활용 목적에 구애받지 않고 가급적 많은 변수들을 포함하고자 하는 경향이 전형적이다.<sup>6)</sup> 인구패널에 관한 초기 논의에서도 이러한 목적의식은 분명해 보인다(이삼식 외, 2006 p. 110, <표 5-2>).

1차 자료의 구축이 목적이라면 그 형식이 패널이어야 할 이유는 뚜렷하지 않다. 현재 활용 중인 대규모 실태조사나 동향조사 자료들이 패널 자료 형식과 운영 방식을 따르지는 않지만, 통계적 기술 분석, 기초통계량 추정이나 지표 생산 등의 목적은 충족한다. 1차 자료를 바탕으로 한 중단 분석도 장기간 축적된 기존 조사 자료의 시계열 비교 요소를 활용하면 이미 어느 정도 가능하다(오영희, 김경래, 신창우, 배혜원, 2016).

역으로 패널이 1차 자료의 성향을 얼마나 지녀야 하는지, 이러한 성향이 패널의 필요조건이 되어야 하는지는 패널 자료의 성격과 목적에 대한 선택과 결정의 문제이다. 이는 결국 명시적으로 ‘패널 자료로 무엇을 하고자 하는지, 또 무엇을 할 수 있는지’를 묻는, 즉 활용과 효용의 측면을 따지는 것으로 귀결된다.

6) 패널 자료를 구축하고 활용하는 것이 대중화된 것은 (1) 대규모 자료 구축과 분석을 위한 계산 능력의 획기적 신장과 비용 제약 해소, (2) 자료의 미비로 불가능했던 수많은 실증 분석, 경험연구 주제의 존재 덕분이다. 그럼에도 경제적, 기술적 제약-대형 조사의 비용과 실현 가능성은 여전히 실질적인 부담으로 작용한다.

기존의 횡단면 실태조사, 예를 들어 출산력조사 자료를, 표본을 고정하고 시점마다 동일한 표본에서 조사 자료를 추출하여, 역사적으로 축적한다는 사전적 의미에 따라 ‘패널화’한다고 가정해 보자. 이 패널의 주된 활용 목적이 여전히 (1) 인구 기초통계량의 추정, 관련 지표의 생산과 이들에 대한 미래 전망(추계), (2) 이 추정치들을 공공·행정서비스 수요 파악 및 정책 기획에 활용하는 것에 있다면, 활용 측면에서 기존 자료에서 크게 달라진 점은 없다. 자료를 종단으로 누적하여 2차원 평면(panel) 형태로 만들고 시점들 간 비교 분석을 수행하는 것이 동태분석의 핵심이라면 패널 자료 고유의 활용 가치를 여기에서 찾기는 어렵다.

활용 측면에서, “패널 자료는 ‘구조적(structural)’<sup>7)</sup> 관계를 식별할 수 있게 해 주어서”(Arellano, 2003, p. 2) 다른 자료들과는 구별된다고 보는 계량경제학의 관점은 그래서 주목할 만하다. 패널 자료 구축과 분석 연구의 초기인 1950년대 말부터 주요한 관점 중 하나는 패널의 구성 자체가 관측하고자 하는 체제(system)의 ‘구조적 관계’를 반영한다는 것이었다. 이에 따르면, 패널 자료 구축의 의미와 가치는 1차원에서 살필 수 없는 구조적 관계와 현상에 대해 구조적 분석과 통계적 추론이 가능해지는 데 있다(Mundlak, 1961). 그리고 이 자료를 바탕으로 한 분석에서 ‘구조적인 요소’에 대한 고려는 필수적이다.

하나의 패널 자료를 상정했을 때,<sup>8)</sup> 조사 단위(개인, 가구 등)별로 관측되는 변수(observed variables)들은 각 개인의 개별적인 속성(individually attributable heterogeneity)을 포착하는데, 종단 차원의

7) 변수들 간의 구체적이고 상세한 관계가 생략되거나 상쇄되지 않고 시간의 진전이나 체제의 변화에도 안정적으로 유지되는 상태를 의미한다. 이러한 변수관계를 반영하는 분석 모형은 구조형(structural form)이라 하는데 결과적·평균적인 상관관계만을 요약하는 축약형(reduced form)에 반대되는 개념으로 사용된다.

8) 논의가 단순하도록 모든 조사 주제 영역과 항목이 변수화되어 있고, 관측에서 누락되어 있는 주제도 변수화할 수 있는 미관측 변수(unobserved variables)라고 가정한다.

축적을 따라가다 보면 각각의 변수에서 동태적으로 안정적이고 구조적(structural)인 부분과 그렇지 않은(idiosyncratic) 부분으로 구별할 수 있게 된다. 변수에 따라 개인의 개별적인 속성들 가운데 관측되지 못하고 감추어져 있던(unobservable) 부분이 시계열을 통해 드러날 수도 있다. 각 변수의 시계열을 장기적으로 따라가면 시간의 흐름에 따라 규칙이나 추세를 가지며 안정적으로 변화하는 구조적 요소(장기 추세)와 불규칙적이고 임의적으로 변동하는 무작위적 요소(단기 변동)가 구분될 수도 있다. 변수들 간의 관계에서도, 인과관계를 단정하기 어렵거나 변수들의(체계 내) 내생성(endogeneity)을 배제할 수 없어 변수들 간의 의미 있는 관계(structural relationship)를 식별하기 힘든 경우, 패널의 종단 구조가 문제를 해소하거나 완화해 줄 수 있다.

패널 자료로 구조적 분석을 할 수 있는 쉬운 예는 시점 간 차분(intertemporal differencing)이다. 이는 시간의 흐름에 따라 변화하는 개인 특성들(과 관계들)을 시간 불변인 요소들로부터 걸러 내서 시차에 따른 단순한 인과관계를 비롯한 체계의 구조적 특성을 파악하게 해 준다. 이를 일반화하면, 시점 간 비교에서도 시점별 실태 또는 현황(snapshot of states) 간의 정태적 비교에 그치지 않고 시점 간 전이과정(state-to-state transition)을 반영한 동태적인 분석까지 고려할 수 있게 된다. 인구학에서의 예로, 임신·출산과 이를 결정하는 요인들 간의 가장 중요한 매개변수가 결혼이라고 보는 인식(가설)을 보자. ‘혼인은 출산으로 이어진다.’라는 이면의 인과관계가 단순한 시차관계를 넘어 성립하는지, 이 관계가 구조적으로 안정적인지 확인돼야 혼인 장려·지원 정책이 출산장려 정책으로서 유효한지를 검증할 수 있다. 이 경우 인구학적 요인 분석 기법은 자료의 구조적인 활용에는 불충분하다. 패널 자료를 활용해 가구별 재산, 소득 수준(또는 소득 계층과 사회경제적 지위 격차) 변화가

다음 기 출산력에 차이를 가져오는지를 살펴본다고 하면, 시차에 따른 인과관계만을 단순하게 분석하기보다는 소득 수준과 출산력 모두에 영향을 미치거나 이들 모두로부터 영향을 받는 중간 요인들과의 상호작용을 식별하고 분석하는 것이 패널의 구조적 활용에 가깝다.

## 제2절 표집 범위와 표본의 대표성

앞서 살펴본 바와 같이 자료 구축 자체가 목표인 1차 자료형 패널은 확장 성향을 띤 대규모 자료가 될 가능성이 크다. 이는 실증분석을 위한 변수들로 정리된 형태보다는 미시적 기초 정보(preliminary micro-information)의 2차원 축적에 더 가까운 모습일 것임을 의미하기도 한다. 이러한 자료에서는 기초통계량 추정, 인구지표 생성, 단순 인과관계 분석과 같은 1차적 활용이 중시되고 특정 변수 관계나 이론의 검증, 분석은 부차적이 된다. 후자에 활용되기 위해서는 원패널을 특정 목적에 부합하는 2차 패널로 재가공하는 과정이 추가로 필요할 가능성이 크다. 그런 의미에서 1차 자료형 패널은 전수조사(census)의 결과로 얻어지는 원시 자료(raw dataset)를 역사적으로 누적한 것과 개념적으로 가까운 ‘전수조사 지향형’ 자료로 볼 수 있다.

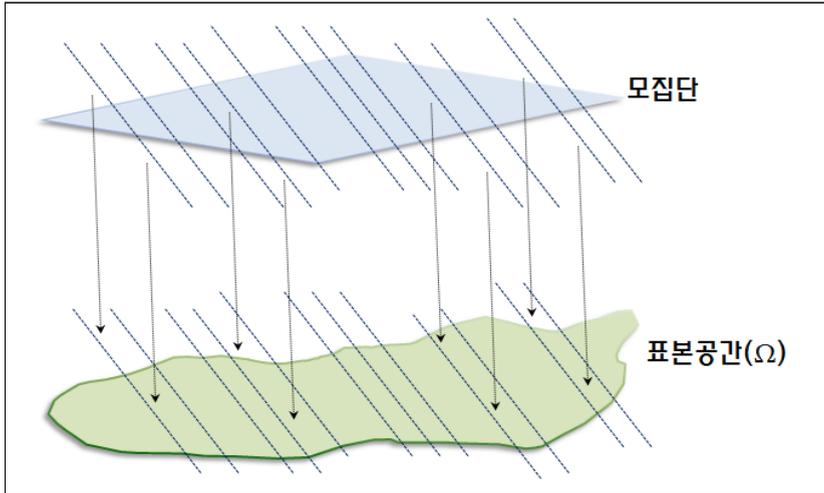
인구패널에 관한 논의에서 그러한 개념적 연계는 더 분명하게 드러난다. 대표적인 인구 기초 자료인 인구주택총조사(census)에서 총인구(population)는 자연스럽게 조사의 통계학적 모집단(population)이 된다. 인구학적 논의에서 출산율, 사망률 등과 같은 기초 인구지표들은 모집단의 분포특성(distribution)을 요약한 모수(population parameter)로 간주된다.

통계청 주관으로 진행되어 온 인구패널 구축 사업은 이런 맥락에서 기획된 패널 구축의 예라 할 수 있다. 지난 2년여 동안 기초 연구와 탐색 작업을 통해 추진하고 있는 방향은 5년 단위 본조사 주기 사이의 사회부가 조사를 패널화하려는 작업이라고 알려져 있다.<sup>9)</sup> 인구주택총조사가 인구등록센서스로 전환됨을 계기로 부족했던 인구학적 조사 항목들을 추가하여 패널을 구축하고, 여기에 주민등록통계 등 가용한 행정자료를 결합함으로써 조사 규모의 현실적 부담을 극복하려는 것을 주요 내용으로 한다. 그 결과는 기본 배경과 의도, 예상되는 활용 목적에 있어 전형적인 전수조사 지향형 패널 자료가 된다.

패널 자료를 1차 자료로 이해하고 조사에 가급적 많은 문항과 넓은 영역의 주제를 포괄하고자 하는 것은 개념적으로 패널을 전수조사에 가깝게 취급하는 것과 같다. 이는 다음과 같이 설명할 수 있다. 표본조사에서 각 표본은, 조사 문항별로 하나의 값을 가지는 벡터(vector)인 동시에, 표본의 기저에 있는 표본공간(sample space)을 이 벡터가 가질 수 있는 값들의 조합만큼으로 분할(partition)하는 확률변수(random variable)이다([그림 2-3] 참조). 그러면 표본을 실제 관측(조사)하는 것은 이 분할로 생긴 표본공간의 부분집합들 가운데 한 조각이 식별된 것이 된다. 표본이 의미하는 분할은 표본공간에서 유익한 정보를 걸러 내는 체(sieve)에 해당한다. 분할이 촘촘해(finier)질수록 표본공간이 더 잘게(finier) 나뉘져 그만큼 표본공간을 더 잘 알게 된다.

9) 아래 제4장 1-2절의 관련 내용 참조.

[그림 2-3] 모집단에 대한 정보 대(對) 표본공간(sample space)의 분할(partition)



자료: 저자 작성.

표본조사의 문항 수가 늘어날수록 분할은 더 세밀해진다. 다만 조사 주제와 영역을 더 넓히고 항목 수를 늘린다고 무조건 더 나은 것은 아니다. 개별 조사 항목이 분할의 유의미한 정보를 추가로 늘리려면 다른 항목, 주제들과의 차별성(stochastic independence) 그리고 전체 조사 자료의 활용 목적에 부합하는 합치성(statistical correlation)이 있어야 한다.<sup>10)</sup>

이렇게 보면, 표본조사는 분할의 일부가 (모집단의 확률분포 특성에 따

10) 따라서 더 많은 주제 영역에 대해서 더 많은 항목을 조사하는 것이 통계학적으로 항상 바람직하지는 않다. 통계학 정보이론(statistical information theory)의 개념을 빌리자면, 조사 항목을 더하는 일이 의미 있는 정보(signal) 대신 잡음(noise)만을 만들거나, 때로는 불필요한 잡음으로 필요한 정보를 가릴 수도 있기 때문이다. 그럼에도, 대규모 조사에서 비용 부담과 기회의 희소성이 오히려 조사의 범위와 항목 수를 가급적 늘리는 방향의 압력으로 작용하는 일은 흔히 볼 수 있다.

인구학적 분석 목적으로 보아 유용하지 않은 예로, 2015년 인구주택총조사 전수조사 항목 가운데 인구 분야(고유 항목)의 본관, 주택 분야(유엔 권고 항목)의 건축연도, 대지면적 등을 들 수 있다(통계청, 2018b).

라) 실현되는 경우, 전수조사는 분할된 표본공간 전체를 식별하는 경우가 된다. 전수조사는 모집단(population) 전체가 표본인 셈이기 때문이다. 전수조사에서 추가적인 주제 영역과 항목은 곧바로 분할의 세밀화로 이어지지만 표본조사에서는 표본의 크기가 충분하지 않으면 추가적인 정보가 식별되지 않을 수 있다. 1차 자료 성향의 패널이 표본의 규모를 중시하는 배경을 이렇게 이해할 수 있다. 더 나아가 표본이 충분히 크다는 전제 하에, 더 넓은 주제 영역과 더 많은 항목에 대한 관측을 추구하는 경향이 패널을 ‘전수조사에 가깝게’ 취급받게 하는 것이라는 점도 설명된다.

패널이 전수 자료에 충분히 가까운지는 패널의 표본이 얼마나 모집단에 가까운지에 달려 있다. 우선 표본이 충분히 커야 하겠지만 표본의 크기는 충분조건이 되지 못한다. 그보다는 표본이 모집단을 얼마나 잘 모사하는지, 특히 표본의 확률적 성격—즉, 표본 확률분포 특성—이 모집단의 분포 특성에 어느 정도 근사(近似)하는지가 기준이 된다. 표본 자료가 식별해 내는 분할에 대한 정보가 전수 자료와 비교해 질적으로 동일하고 세부 정보에서만 차이가 난다면, 분석을 위한 효용에 양자 간 큰 격차가 지지는 않을 것이기 때문이다. 표본의 대표성, 즉 표본의 모집단에 대한 확률적 유사성이 우선적 가치가 되는 이유가 여기에 있다.

대규모 횡단면 조사 자료나 패널 자료에서 표본의 모집단 대표성을 보장하고 유지하기 위해 다양한 방법들이 강구되는 배경에는 이들 자료의 전수 지향성이 있다. 대부분의 현용 패널 자료들에서 그리고 인구패널 구축에 관한 논의에서 패널의 활용도보다는 표본의 대표성이 패널의 질적 가치를 결정하는 핵심 기준으로 간주되는 것 또한 이와 무관하지 않다.

패널 자료에서 표본의 대표성은 횡단면 자료에서와 질적으로 다른 측면이 있다. 패널 자료를 정의하는 조건은 ‘동일한 표본의 시계열적 추적

과 관측'의 결과물이라는 점이다. 이 때문에 패널에 있어 전수 또는 1차 자료 지향적인 성향은 여타 자료와 조금 다른 함의를 지닌다. 패널의 표본은 시간의 경과에 따라 동일하게 유지되므로, 비록 대규모일지라도 종단 방향으로 표본 대표성이 일정하게 유지, 보장되기 어렵다. 패널 관측이 유지되는 기간이 짧다면 이는 큰 문제가 되지 않겠지만, 패널 유지 기간이 길어질수록 (고정된) 표본과 모집단의 확률적 특성 간 괴리가 커지는 현상은 피할 수 없다. 이는 패널의 노화(ageing)에 따른 불가피한 결과이다. 문제는 표본의 모집단 대표성 상실이, 이 표본을 바탕으로 한 통계적 추론(추정, 추계, 검정 등)의 통계적 신뢰성 상실로 이어질 수 있다는 점이다. 패널 자료를 암묵적으로 전수 자료처럼 여기는 것이 바람직하지 않을 수 있다. 인구학적 패널을 합계출산율 추정과 동태 추적에 쓰는 것이 인구주택총조사나 인구동향조사, 출산력조사를 쓰는 것과 비교할만 하지도 않고 결과를 같은 의미로 해석해서도 안 된다는 의미이다.<sup>11)</sup>

패널이 성숙할수록, 즉 패널 유지 기간이 길어질수록 표본의 대표성 문제는 더 실재적이 된다. 패널의 노화에 따른 문제들에 대처하기 위해 표본을 보강하고, 복수의 패널 세대(waves)를 중첩해서 유지하는 등 여러 장치들을 보완적으로 활용할 수 있지만, 패널의 종단 방향과 표본의 모집단 대표성 간의 긴장을 근본적으로 해소하는 것은 아니다. 표본의 모집단에 대한 대표성을 보장하기 위해 전형적으로 활용되는 조사 단계에서의 표본 층화와 표본 가중치 부여를 고려해 보자. 횡단면 조사에서 이 장치들은 표본의 분포 특성을 모집단에 더 가깝도록 보정해 주거나 표본조사를 바탕으로 한 분석 결과를 모집단으로 확장해서 해석하게 하는 주요 근거가 된다. 유지 기간이 긴 패널 자료의 경우 모집단의 분포 특성이 시간의 경과에 따라 변화하여 고정된 표본과의 괴리가 커지면, 표본 층화와

11) 이러한 목적의 인구학적 패널 자료 활용과 의의에 대한 논의는 아래 제4장 1절 참고.

표본 가중치 부여의 방법과 내용이 달라져야 할 수도 있다. 하지만 그렇다고 이들을 시점마다 조정하게 되면, 시점 간에 동일한 표본을 유지하고자 하는 패널의 근본 취지와 상충될 가능성이 있다.

1960-1970년대의 초기 패널 자료 구축과 분석 방법론 개발 과정에서 이러한 긴장이 문제시되지 않았던 것은 전형적 패널 자료를 세로가 짧고(종단 방향의 크기가 작고) 가로가 긴(횡단 방향으로 큰), 자료의 확장에서도 종단 방향으로는 고정되거나 제한적인 형태라고 암묵적으로 전제했기 때문이다. 한편으로 이는 패널 자료의 구축과 활용에서 표본의 모집단 대표성이 절대적인 가치는 아니란 것을 의미하기도 한다. 패널 자료에서 표본의 대표성은, 패널이 시작되고 처음 표본이 추출되는 시점에, 패널의 목적과 주제의식에 부합하는 특정 인구집단에 대한 대표성이다. 일반적인 패널 자료에서 여러 개의 패널 세대(waves)를 중첩하여 유지하는 것의 적극적 의미도 이러한 제한적 대표성에 있다.<sup>12)</sup>

이러한 맥락에서, 1차 자료 또는 전수 지향 패널이 대규모 횡단면 조사 자료의 역사적 축적과 구별되는 비교 우위가 있다고 보기는 어렵다.

### 제3절 패널 자료의 활용 가치

형식이 아닌 활용 측면에서 패널 자료 고유의 가치는, 패널이라면 기본적으로 다루어야 할 문제가 무엇인지, 또 패널이 아니라면 다루기 어려운 분석과 실증 문제가 어떤 것인지를 살피는 가운데서 찾을 수 있다. 패널 자료가 필요한 최소한의 이유는, 연관된 변수들 간의 전후방 관계가 관측

12) 인구학적 패널은 총인구(모집단, population)의 분포와 구조 변화, 이를 초래하는 요인을 관측하고 추적하는 것이 목표이기 때문에, 고정된 표본의 제한적 대표성으로는 충분하지 않다. 아래 제4장 2절 관련 논의 참조.

단위의 개별적(미시적) 특성에 기인한 것인지 공통적인 요인 때문인지, 또 시간의 경과에 따라 가변적이고 일시적으로만 관측되는 것인지 지속적으로 안정적인 성격인지 체계적으로 포착하게 해 준다는 점에 있다. 이러한 암묵적인 이해를 전제할 때, 패널은 목적과 주제의식이 분명할 뿐만 아니라 적합한 분석 방법까지 염두에 두고 구축된 자료라고 볼 수 있다.

패널 자료는, 다른 자료(dataset)와 마찬가지로 자료가 대표하는 체제(system)가 어느 범위까지 현실을 포괄하도록 할지, 즉 (1) 어떤 변수들을 포함할지, (2) 이 변수들로 설명하고자 하는 현상과 변수들 간의 관계를 어느 범위까지 관측 대상으로 삼을지에 대한 견해를 담는다. 패널이 다른 자료와 다른 점은 그 판단과 관점이 더 명시적이라는 데 있다. 자료 자체에 분석과 설명이 필요한 관계들이 더욱 구체적으로 드러나고 이 관계들에 대한 기본적인 가설도 어느 정도 담겨 있기 때문이다. 그런 맥락에서 패널 자료는 목적지향형이고 구조적(structural)이다.

패널 자료는 이를 이용한 패널 분석 방법론과 밀접한 관계를 이루며 발전해 왔고 자료에 분석 방법론에 대한 고려를 의식적으로 반영한다는 점에서도 목적지향적이다. 초기 패널 연구자들의 관심이 자료의 횡단면에 집중되었다는 점은 앞에서 논의한 바 있는데, 특히 이들이 패널 자료에서 주목한 것은 (1) 자료에 관측되지 않은 개별적 특성(unobserved heterogeneity)이 상존하여 발생하는 통계적 편의(bias)를 완화하거나 해소하는 문제, (2) 일부 독립변수(통제변수 또는 외생변수)가 선별적(selective)이어서 관측 불가 부분이 잘려 나간(censored) 채로만 표본이 관측되는 표본 선택편의(sample selectivity bias)를 식별하고 우회하는 문제였다(Hsiao, 2003, pp. 8-10). 현재까지 개발된 패널 분석 방법론들은 패널 자료 안팎의 이러한 문제들에 대처하기 위해 고안된 기법

들이고, 패널 자료는 이들의 적용(또는 적용의 어려움)을 감안해서 설계, 구축된다.

패널 자료의 특성과 문제의식을 가장 정확하게 반영하는 분석 방법은 통계학적 분산분석(ANOVA: analysis of variance)이다. 패널 자료 분석의 기본 문법은 개별 관측 단위별[종(縱)으로]과 시점별[횡(橫)으로]로 ([그림 2-1] 참조) 공통적으로 나타나는 특성을 먼저 살펴서 관측되지 않은 개별 특성을 먼저 통제하고 나서 주 관심 변수들의 변화와 상호관계를 파악해 나가는 것인데, 이는 전체가 여러 개의 하위집단으로 구성되는 통계자료(dataset)에서 집단별로 통계적 특성(평균, 분산 등)을 파악하거나 이들 간 비교하는 것을 주목적으로 하는 분산분석의 현실적인 적용 사례와 다르지 않다. 이는 선형모형(linear model) 형태로 표현한 전형적 분산분석모형이 사실 패널 분석에서 가장 단순한 고정효과모형(fixed effect model) 또는 확률효과모형(random effect model)이라는 점에서 확인된다(Arellano, 2003, 2.2절, 3.1절 참조).

$$y_{it}^1 = \alpha D_{jt}^i + \beta D_{is}^t + \epsilon_{it}, \quad i = 1, \dots, N, t = 1, \dots, T, \quad (1)$$

$$D_{jt}^i = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}, \quad D_{is}^t = \begin{cases} 1, & s = t \\ 0, & s \neq t \end{cases}$$

[그림 2-2]의 변수 벡터  $z_{it} = (y_{it}^1, y_{it}^2, \dots, y_{it}^h; x_{it}^1, x_{it}^2, \dots, x_{it}^k)$ 에서 첫째 변수  $y^1$ 에 대한 분산분석모형을 구성해 보면 식(1)로 표현되는데, 이는 확률적 오차항(stochastic error or disturbance, 혹은 idiosyncratic error)을  $\epsilon_{it}$ 로 보는가 아니면 우변 전체( $D_{jt}^i + D_{is}^t + \epsilon_{it}$ )로 보는가에 따라 독립변수가 없는 (종속변수  $y^1$ 에 대한) 고정효과모형이거나 확률효과

모형이 된다. 여기에 종속변수를 포함한 기본적인 선형패널모형은 본질적으로 종단-횡단 구분에 따른 분산분석모형의 자연스러운 확장이다.

패널 자료의 특성을 가장 자연스럽게 활용하는 분석 방법은 자연실험 모형(natural experiment design)과 그 파생모형들이다. 자연실험모형은 시간 간격을 두고 변수들 간에 나타나는 시차상관 관계(across-time correlation 또는 intertemporal correlation)로부터 직접적 인과관계 함의를 이끌어 낼 수 있도록 구성된 방법이다. 가장 간단한 형태로 이를 두 기간에 걸친 패널(체제)에 적용한 사례가 이중차분(DID: differences-in-differences) 모형이다.<sup>13)</sup> 관측 기간에 설명변수(또는 외생변수)가 증감할 때, 종속변수(내생변수)의 증감은 그 영향을 받은 집단(treatment group)에서만 실제로 관측되기 때문에, 이들 간의 효과와 영향<sup>14)</sup>의 크기를 정확하게 추정하려면 설명변수 증감의 영향을 받지 않은 집단(control group)의 평균적인 변화분만큼의 왜곡을 보정해 주어야 한다.

가장 간단한 위 식(1)의 경우에서 다시 살펴보면, 종속변수  $y^1$ 에 영향을 미치는 것으로 짐작되는 정책( $P_i = 1$ )의 영향을 편의(bias) 없이 측정하려면, 전체 종속변수의 평균 변이분,  $E[y_i^1 | P_i = 1] - E[y_i^1 | P_i = 0]$ ,을 추정할 것이 아니라 영향을 안 받은 집단( $y_0^1$ )의 자연 변이분,  $E[y_{0i}^1 | P_i = 1] - E[y_{0i}^1 | P_i = 0]$  ( $= E[y_{0i2}^1 - y_{0i1}^1]$ ),를 빼 준 부분, 즉

13) 이중차분모형과 함께, 관측되지 않는 개별 특성을 걸러 내고 처치효과(treatment effect)를 식별하는 데 효과적인 도구가 성향점수(propensity score)이다. 이 때문에 이 둘은 패널 자료에 기반한 정책효과 분석에 유용한 도구로 활용된다. 이중차분모형을 더 일반화한 방법론이 회귀단절모형(RDD: regression discontinuity design) 또는 회귀굴곡모형(RKD: regression kink design)이다.

14) 통계적 변이(idiosyncratic error)를 배제한 평균 인과관계.

$$E[y_{1i}^1|P_i = 1] - E[y_{0i}^1|P_i = 0] = \{E[y_i^1|P_i = 1] - E[y_i^1|P_i = 1]\} - \{E[y_{0i}^1|P_i = 1] - E[y_{0i}^1|P_i = 0]\} \quad (2)$$

를 추정해야 한다. 이를 한목에 해결하는 방법은 식(1)에 정책 변수가 더해진

$$y_{it}^1 = \alpha D_{jt}^i + \beta D_{is}^t + \gamma P_{it} + \epsilon_{it}, \quad i = 1, \dots, N, t = 1, 2, \\ y_{0it}^1 = \alpha D_{jt}^i + \beta D_{is}^t + \epsilon_{it}, \quad y_{0it}^1 = y_{it}^1 \text{ when } P_{it} = 1,$$

를 각각 기간에 따라 차분해서 다시 그 차이를 구한 식(3)

$$\Delta y_i^1 - \Delta y_{0i}^1 = \gamma(P_{i2} - P_{i1}) + u_i, \quad \Delta y_i^1 = y_{i2}^1 - y_{i1}^1, \quad (3)$$

을 추정하는 것이다. 이를 다른 설명변수들의 존재와 상호작용을 포함하는 일반적 형태로 확장한 것이 준(準)실험모형(quasi-experimental design)에서의 이중차분 방법이 된다(Stock and Watson, 2003, 제11장).

패널 자료는 기본적으로 준(準)실험모형의 적용이 자연스러운 구조를 갖추고 있고 준실험모형의 구성 요소를 내포하는 특성이 있다. 패널 자료가 정책실험 또는 정책 효과의 추정에 적합하다는 인식, 패널 자료의 활용 가치를 정책성과 평가나 정책 파급효과 추정에 대한 경험연구로 보는 견해의 근거가 여기에 있다. 목적지향적인 패널 자료를 대개 정책패널로 이해하고 부르는 이유도 여기에 있다.

목적지향적인 패널 자료는 분석할 수 있는 주제와 대상, 포함되는 변수들에 대해 더 구체적이고 명시적일 수밖에 없다. 패널의 유지 기간이 길

어지면 패널에서 관측하고자 하는 주제와 변수들이 추가되거나 바뀔 수 있고, 패널의 주 관심사가 여러 영역으로 분화할 수 있다. 그 결과로 목적 지향적인 패널 자료가 오래 유지되고 규모가 커지면, 주제와 관심사에 따라 원패널로부터 더 작고 특화된 자(子)패널이나 손자(孫子)패널로 분화되는 경향이 나타난다. 대표적인 예가 미국의 의료비지출패널조사(MEPS: medical expenditure panel survey)이다.

## 제4절 자료 관측 단위, 주기 및 기간

### 1. 관측 단위

패널 자료에서 관측(조사) 단위는 어느 자료에서와 마찬가지로 중요하다. 관측 단위를 무엇으로 하는지에 따라 자료의 성격과 분석 방법, 결과의 해석이 달라지기 때문이다. 관측 단위를 결정하는 일반적 기준은 자료 구축의 목적, 주 관심 대상과 주제에 부합하는 수준이다. 여기에 자료의 가득성, 관측과 수집의 용이성, 축적과 관리의 비용과 부담에 의한 제약도 고려될 수밖에 없다.

관측 단위를 기준으로 구분하면 패널 자료도 일반적인 자료에서와 같이 크게 거시패널과 미시패널로 나뉜다(Arellano, 2003, pp. 1-2, p. 7). 여러 나라의 국민계정 자료를 모아 놓은 국가 패널(예: Penn World Table)은 국가를 단위로 하는 거시패널의 좋은 예다. 거시집계변수(macro-aggregate variables)를 관측 대상으로 삼아 국가 간 혹은 지역 간 제도, 정책 차이로부터 설명되는 이들의 변동요인을 포착하는 것을 목적으로 하는 국가(지역) 비교 패널들은 전형적 거시패널이다. 각종 국제

기구나 연구기관에서 집계, 발간하는 주요 자료들이 이에 해당한다. 거시패널은 다루는 변수들의 성격상 시계열적 통계 특성이 자연스럽게 부각되는 경우가 일반적이고, 집계(aggregate) 패널로서 통상 횡단과 종단의 규모가 비슷한 형태를 띤다.

관측 단위별 개별적인 특성(individual heteroskedasticity)을 식별할 수 있는 자료 제공은 패널 자료의 근본적인 존재 이유이다. 패널 자료의 분석 또한 개별 특성의 시계열적 변동 추적과 해석이 핵심이 되는 것은 그 때문이고 이는 거시패널에서도 마찬가지이다. 하지만 관측 단위인 '개체(individuals)'가 국가, 지역, 산업인 거시-집계패널보다 기업, 가구, 노동자, 여성인 미시패널이 그러한 패널 자료의 본령에 더 충실한 것은 어느 정도 자명하다. 대부분의 대표적인 국내외의 패널 자료들은 이 기준에서 전형적인 미시패널이다.

개인, 가구, 개별 기업 단위의 선호와 의사결정, 행위에 관련된 문제, 그 결과가 반영된 미시적 효과는 미시패널의 주 관측 대상이다. 패널 자료 분석의 의의—관측되지 않는 미시적, 구조적인 개체 개별 특성을 제어하는 것, 변수 변동을 요인별로 식별하고 변수 관계를 동태적으로 추적하는 것—는 미시패널에서 극대화된다. 이를 위한 자료 수집은 설문조사, 총조사, 등록 행정 기록, 회계 정보, 감사·평가 기록 등을 통해 이루어지는 것이 일반적이다. 이 때문에 미시패널은 대규모 횡단면 자료로부터 진화, 확장한 자료로 이해할 수 있다. 이는 전형적인 미시패널이 큰 수의 개체들을 비교적 단기간에 걸쳐 반복 관측 한 형태, 즉 횡단면 규모가 크고 종단 방향이 짧은 모습을 띠는 것에서 간접적으로 확인된다.

미시패널에서 특히, 관측 단위 결정은 패널 자료의 주제나 문제의식과 직접적으로 연계되고 일관되어야 한다. 주거와 주택 수요에 관한 자료라면 가구 단위가 자연스럽겠지만, 근로와 직업 선택, 취업에 관한 자료라

면 개인 단위가 더 적합하다. 임신과 출산에 대한 의사 결정과 행동 양식에 대한 관측의 기본 단위로는 행위 주체인 가임기 여성이 자연스럽다.<sup>15)</sup>

## 2. 관측 주기와 기간

앞서 언급되었듯이 패널 자료, 특히 미시패널에 대한 전통적인 관점은 자료가 종단 방향으로 짧고 규모가 제한적이라는 것이다. 자료의 확률적 특성과 통계적 추론의 전제(대표본이론)도 횡단 방향의 자료 크기에 전적으로 좌우되는 것으로 이해하는 것이 일반적이다. 그런 점에서 전통적 미시패널 자료는 횡단형 패널이라 할 수 있다.

하지만 자료를 축적하고 계산하는 능력이 급격하게 신장되고 비용이 줄어들면서 이 기본 전제는 점차 희석되고 있다. 패널이 종단 방향으로 커짐에 따라 종단 방향의 분석, 시계열적 해석이 의미를 지니게 되고 횡단 방향에 대한 상대적 규모가 무시할 수 없는 수준이 되어 종단이 전체 패널의 확률적 특성에도 영향을 미치는 경우에 대한 고려도 늘어 가고 있다.<sup>16)</sup>

종단이 증시되는 패널 자료에서 야기되는 첫 번째 문제는 앞서 살펴본 표본의 모집단 대표성 문제이다. 전통적 미시패널의 경우 종단의 크기인 관측 기간이 단기로 제한되거나 고정된 것으로 본다. 따라서 패널 관측 기간 동안 모집단이 급격하게 진화하거나 확률적 분포특성이 급변<sup>17)</sup>하지

- 
- 15) 미시패널이 인접 횡단면 자료와 밀접하게 연관되고 인구주택총조사를 비롯한 대표적 횡단면 조사 자료들이 대체로 가구를 조사의 기본 단위로 삼는다고 해서 패널 자료의 표본 추출과 자료 수집의 단위가 가구여야 할 필요는 없다. 다만 앞서 지적된 바대로 자료의 가득성(availability)이 제약이 될 수는 있다.
  - 16) 시계열적 패널 또는 패널의 시계열적 특성에 관한 논의에 대해서는 Arellano(2003) p. 2, Arellano and Honoré(2001), Geweke and Keane(2000), Hahn and Kuersteiner(2002), Phillips and Moon(1999), Blundell and Bond(1998) 참조.
  - 17) 구조적인 체제 변환(regime change or switch)을 예로 들 수 있다.

않았다는 가정하에, 동일한 표본과 변화하는 모집단의 괴리는 무시할 만한 수준에서 유지될 것으로 기대할 수 있다. 이때의 표본 대표성은 패널 시작 시점에서의 표본 추출에서 충분히 소화되는 것으로 간주해도 무리가 없다. 하지만 패널 유지 기간이 길어져 그러한 기본 전제를 더 이상 만족시키지 못할 임계점을 넘게 되면 앞서 논의한 것처럼 표본 대표성 또는 패널 특성 유지가 심각한 문제로 대두될 수 있다.

패널 자료의 종단 특성이 영향을 미칠 때 야기되는 두 번째 문제는 시간 경과에 따른 변수의 가변성(time variability)이다. 개별 변수의 가변성은 패널 관측 주기에 대한 상대적인 변화의 크기가 문제가 된다. 변수들 간의 가변성 정도 차이는 변수들의 시계(視界, time horizon) 차이로 인한 문제와 연결된다. 패널 자료의 구축과 패널 분석을 정당화하는 근거는 종단 방향으로 변수가 충분히 변동한다는 전제이다.

종단이 짧은 전통적 패널에서 변수의 가변성은 패널의 존재 근거 자체가 되는 반면, 관측 기간이 긴 종단 중시형 패널에서 변수의 가변성은 자료의 관측 주기와 더 밀접하게 연관된다. 패널 자료에서 관측되는 정보는 관측 주기(time frequency of observation)를 기준으로 포착 가능한 간격을 두고 발생한 사건, 패널 유지 기간에 관측 주기와 맞아 떨어지는 빈도(frequency)로 변화하는 추이로 제한된다. 관측 주기 사이에 발생하고 소멸되는, 즉 관측 주기보다 더 짧은 주기로 빈발하는 사건(high frequency events)은 관측에서 누락된다. 정반대로 관측 주기에 비해 너무 천천히 진행되는, 즉 관측 주기가 너무 짧아서 놓치는 장기 추이(low frequency events = long-term trends)도 패널에서 관측하지 못한다.<sup>18)</sup>

18) 이들을 포함해서, 일반적인 시계열 변수에서 특정 주기의 시계열 요소들이 관측 주기와 엇갈리기 때문에 관측에서 누락되거나 본래의 활률 특성이 왜곡된 형태로 관찰되고 해석되는 현상을 주기 착오 또는 주기 익명화(aliasing)에 따른 착시라고 부른다.

변수 간의 시계 차이가 과도하게 큰 경우 이들 간의 관계를 살펴보는 것이 시계열 측면에서 통계적 유의성이 있는지, 있다면 시계열적으로 어떻게 볼 것이며, 단기-장기의 구분으로 본다면 어떤 시계에 초점을 맞추어야 관측과 분석의 목적에 적합한지에 대한 판단이 요구된다. 시계열 변수는 단기적인 일시적 교란요인에서부터 규칙적인 단기순환변동, 장기적인 추세(trends) 요소에 이르기까지 다양한 파장과 주기의 요소들이 혼재되어 있는 것으로 이해된다. 이 요소들 간의 비중이 어떻게 발현되는지에 따라 개별 시계열 변수는 단기적 특성이 두드러질 수도 있고 장기 성향이 더 강하게 나타날 수도 있다. 시계열 특성이 유사하지 않은 변수들을 모아서 동태적 관계를 설명하려고 하면 그 관계가 가성적(假性的, spurious) 시계열 관계일 수 있다. 시계열적으로 가성적이라 함은 둘 이상의 시계열 변수가 실제로는 상호 안정적, 지속적인 관계가 없는 임의적인 묶음인데도 통계학적으로 유의한 관계가 식별되는 경우를 말한다(Hamilton, 1994, pp. 557-562).

이러한 종단 측면의 사정들이 실질적인 문제가 된다면 패널 자료의 구축 단계에서부터 이들에 대한 고려가 필요하다. 자료 구축에 이들을 반영하면 자료의 분석과 결과의 해석이 제한되거나 질적으로 달라질 수 있다.

정책의 효과에 대한 분석, 평가를 주목적으로 하는 정책패널의 경우 시계열적 관점이 특히 중요한데, 그 이유는 시계열적 고려가 무시되었던 이전과 상반되는 결과와 함의를 낳을 수 있기 때문이다(김중훈·이소영·이윤경·이지혜·김세진·변수정·조성호·백혜연·이상립, 2018, pp. 27-28). 인구 현상에 관한 패널 자료를 구축하는 경우를 생각해 보면 문제는 더 분명해진다. 인구 변동의 장기적 추이와 동태적 변화를 반영하려면 종단 측면이 강조되지 않을 수 없다. 그런데 예를 들어 5년 주기로 관측된 패널

---

Hamilton(1994), p. 161 참조.

이라면 관측 주기 사이에 발생하고 소멸되는 정책 변화 효과(예: 관측 1년 후 시행된 출산장려금 정책의 효과가 2년 동안만 지속되고 소멸되는 경우)는 관측에서 누락된다. 인공임신중절 합법화가 가지는 인구적 효과가 장기적이어서 기간 간 탄력성(intertemporal elasticity)—정책효과가 나타나기까지의 기간 또는 시차가 반영된 백분율 증가율—이 5년을 넘는다면 5년 주기 패널 자료를 통해 그 효과를 충분히 포착하기는 어렵다.<sup>19)</sup>

---

19) 인구구조 변화의 영향이 장기 추세적 변동을 통해서만 나타나는 경우는 많다. 인구구조와 구성비 변화에 따른 고등교육 취득 비중 증가, 노동시장 환경 불안정화, 가구 규모 다양화, 성평등 문화 급변 등은 인구 현상의 시계가 관측의 시계를 넘는 좋은 예이다. 이들을 정책의 시계와 비교했을 때의 시계 불일치도 다른 차원에서 정책적 문제가 된다.



# 제 3 장

## 국내외 주요 패널 및 인구 관련 자료

제1절 국내 주요 패널과 인구 관련 자료

제2절 해외 주요 인구 관련 패널 자료

제3절 '인구패널'을 위한 시사점



# 3

## 국내외 주요 패널 및 인구 관련 << 자료

### 제1절 국내 주요 패널과 인구 관련 자료

#### 1. 국내 주요 패널 사례

국내에서 최초로 시행된 패널조사는 1993년 대우경제연구소의 한국가  
구패널조사(Korea Household Panel Study)이다. 이 조사는 1997년  
외환위기의 영향으로 1998년에 7차 조사를 끝으로 중단되었다. 1998년  
에 시작된 한국노동연구원의 한국노동패널은 KHPS 운영 경험에 기초하  
여 구축되었다고 평가된다. 한국노동연구원이 패널조사를 도입함으로써  
공적 영역에서도 최초로 패널조사가 도입되었고, 한국노동패널이 성공적  
인 결과를 보이자 2000년 이후 공적 영역과 학계 등에서 패널조사가 급  
속히 확산되었다(이희길, 2007).

〈표 3-1〉을 살펴보면, 실제로 대부분의 패널조사가 2000년대 전반기  
에 계획되고 조사를 시작한 것을 알 수 있다. 2007년을 기준으로 볼 때,  
1998년 시작된 한국노동패널을 포함하여 총 17개 패널조사 중 12개 패  
널조사가 2005년 이전에 계획되었다. 그리고 거의 대부분의 패널조사는  
본조사 시작 전에 예비조사를 수행한 것을 알 수 있다. 2000년대에 패널  
조사가 크게 관심을 받으며 일부는 부실화의 우려를 지적받기도 하였다.  
패널조사는 특성상 한번 설계된 이후 계획이나 내용을 수정하기가 매우  
어려우므로 충분한 준비 기간을 갖고 검토되어야 하는데, 아래 표에서 알  
수 있듯이 계획부터, 예비조사 및 1차 본조사까지 1~2년 안에 완료되는

50 결혼 및 출산 관련 정책 수립과 평가를 위한 인구패널 구축 타당성 연구

경우도 있었기 때문이다.

〈표 3-1〉 패널조사 계획 및 1차 조사 시작 시기<sup>20)</sup>

번호	조사명		계획 연도	예비조사	1차 조사	진행 차수 (2007 기준)
1	청소년패널	중 2	2001	2003	2003	5
		초 5	2001	2004	2004	4
2	교육종단연구		2004	2004	2005	3
3	교육고용패널		2003	2003	2004	4
4	청년패널	YP2001	2000	2001	2001	7
		YP2007	구축 중			-
5	대졸자직업이동		2005	2006	2006	2
6	여성인력패널		2007	-	2007	1
7	고령화연구패널		2005	2006	2006	2
8	장애인고용패널		2007	2008	2008	0
9	아동패널		-	2007	2008	0
10	인구패널		2008	2008	-	-
11	노동패널		1998	1998	1998	10
12	여성가족패널		2006	2006	2007	1
13	복지패널		2005	2006	2006	2
14	의료패널		2006	2007	2007	1
15	노후보장패널		2004	2005	2005	3
16	사업체패널		2002	2002(5)	2002(5)	6(3)
17	인적자본기업패널		2004	2004	2005	2

자료: 이희길, 2007, p. 23.

20) 10번 항목의 인구패널은 2008년 예비조사를 진행하였지만, 이후 본조사는 진행하지 않았다.

여기서는 국내외의 패널조사 중 대표적인 패널조사와 인구 현상과 관련이 있는 패널조사를 중심으로 그 특성을 살펴보고 향후 인구패널 구축 가능성 탐색의 사례로 삼고자 한다. 우선 국내에서 가장 대표적인 패널로 평가되는 한국노동패널과 한국복지패널을 사례로 살펴보고, 주요 인구 현상을 내용으로 담은 여성가족패널도 중요한 사례로 검토한다.

### 가. 주요 패널조사 개요

한국노동패널은 1998년부터 시작되어 현재 한국에서 가장 대표적인 패널조사로 평가할 수 있다. 조사의 기본 단위는 가족이며 매년 1회 조사되고 있고, 1998년 시작된 표본 기준으로 2018년에 21차까지 조사가 완료되었다(한국노동연구원, 2018). 한국복지패널은 사회과학 분야에서 노동패널과 함께 널리 활용되고 있는 대표적인 패널로서 2006년에 1차 본조사가 시작되어 2018년 기준 13차 조사가 완료되었다. 가구를 조사단위로 하며, 매년 1회 조사한다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019). 여성가족패널은 2007년에 1차 본조사를 시작하여 2018년에 7차 조사를 완료하였다. 조사 단위는 가구를 기본으로 하며 2007년과 2008년에는 매년 조사하였지만, 그 이후부터는 격년 주기로 조사하고 있다. 세 가지 패널조사 모두 조사 단위는 가구이며, 실제로 가구 내의 개인을 별도로 조사하고 있기도 하다. 그리고 최근에는 종이조사 표보다는 컴퓨터 단말기를 활용하는 조사가 보편화되고 있다(한국여성정책연구원, 2018).

〈표 3-2〉 주요 패널조사 개요(2018년 기준)

조사명	본조사 시작 연도	조사 단위	조사 주기	현재 차수	조사 방법	수행 기관
한국노동패널	1998년	가구 (개인)	1년	(98표본) 21차	CAPI	한국노동연구원
한국복지패널	2006년	가구 (개인)	1년	13차	CAPI	한국보건사회연구원
여성가족패널	2007년	가구 (개인)	2년	7차	CAPI	한국여성정책연구원

자료: 각 패널조사 User's Guide를 참조함(한국노동연구원, 2018; 한국보건사회연구원, 서울대학교 사회복지연구소, 2019; 한국여성정책연구원, 2018).

## 나. 주요 패널조사의 목적 및 대상

한국노동패널조사는 경제활동 및 노동시장 이동, 소득활동 및 소비, 교육 및 직업훈련, 사회생활 등에 관하여 조사하는 종단면조사이다. 과거 초기 자료가 미흡한 문제를 해소하고 분석적인 노동시장 연구를 활성화하여 더욱 합리적이고 정확한 노동시장 및 고용정책을 수립하고 평가하기 위하여 실시하고 있다(한국노동연구원, 2018).

한국복지패널조사는 외환위기 이후 빈곤층, 근로빈곤층(working poor), 차상위층(near poor)의 가구 형태, 소득 수준, 취업 상태가 급격히 변화하고 있는 상황에서 시작하였다. 이들 계층의 규모 및 생활실태 변화를 동태적으로 파악하여 정책의 효과를 제고할 목적으로 실시하고 있다. 이를 통해 연령, 소득계층, 경제활동상태 등에 따른 다양한 인구집단별로 생활실태와 복지욕구 등을 역동적으로 파악하고 정책집행의 효과성을 평가하여 새로운 정책 입안과 제도 개선 등 정책 환류에 기여하고자 한다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019).

여성가족패널조사는 여성의 삶의 변화와 여성 정책 등을 정확히 포착하고 분석할 목적으로 실시하고 있다. 여성의 생애주기별 경제활동 지위

변화와 일자리 경험, 가족과 관련한 가치의 변화, 가족관계의 변화, 가족 형성 과정과 사건(event) 변화, 가족구조의 변화 등을 추적할 수 있도록 내용을 구성하였다. 이를 통해 여성의 경제활동 증가와 일-가족생활 조화(work-family life balance) 및 여성의 생애 전반에 걸친 일과 가족 관련 경험을 파악하여 가족의 현재와 미래에 대한 전망을 분석하고자 한다. 또한 저출산, 고령화, 보육, 일-가정 양립, 가족친화 수준 등에 대해 정보를 추적하여 여성 및 가족 정책 수립을 위한 기초자료로 활용되고 있다(한국여성정책연구원, 2018).

〈표 3-3〉 주요 패널조사의 조사 목적 및 대상

조사명	조사 목적	조사 대상
한국노동패널	<ul style="list-style-type: none"> <li>- 경제활동 및 노동시장 이동, 소득활동 및 소비, 교육 및 직업훈련, 사회생활 등에 관하여 추적 조사 하는 종단면조사임.</li> <li>- 노동시장 관련 조사 자료의 미흡 및 부족의 문제를 해소하고 나아가 분석적인 노동시장 연구를 활성화하여 더욱 합리적이고 정확한 노동시장 및 고용정책의 수립과 평가에 기여하고자 함.</li> </ul>	<ul style="list-style-type: none"> <li>- 98표본: 우리나라의 도시 거주 5000가구와 원가구에서 분가한 신규분가가구 그리고 해당 가구의 15세 이상 가구원</li> <li>- 통합표본(2009년 12차 추가) 원가구와 분가가구 그리고 15세 이상 가구원</li> </ul>
한국복지패널	<ul style="list-style-type: none"> <li>- 외환위기 이후 빈곤층, 근로빈곤층, 차상위층의 가구 형태, 소득 수준, 취업 상태가 급격히 변화하고 있는 상황에서 이들 계층의 규모 및 생활실태 변화를 동태적으로 파악하여 정책 형성에 기여함과 동시에 정책지원에 따른 효과성을 제고할 목적으로 실시함.</li> <li>- 연령, 소득계층, 경제활동상태 등에 따른 다양한 인구집단별로 생활실태와 복지욕구 등을 역동적으로 파악하고 정책집행의 효과성을 평가하여 새로운 정책의 형성과 제도적 개선 등 정책 환류에 기여하고자 함.</li> </ul>	<ul style="list-style-type: none"> <li>- 원표본 가구(2006년): 7072가구와 해당 가구의 15세 이상 가구원</li> <li>- 신규 추가 표본 가구: 2012년 7차 조사에서 1800가구가 추가됨(해당 가구의 15세 이상 가구원).</li> <li>- 중위소득 60% 미만 저소득층에 전체 표본의 약 50%를 할당함.</li> </ul>
여성가족패널	<ul style="list-style-type: none"> <li>- 여성의 변화와 여성 정책 등을 더 정확히 포착하고 분석할 목적으로 실시함</li> <li>- 여성의 생애주기별 경제활동 지위 변화와 일자리 경험, 가족과 관련한 가치의 변화,</li> </ul>	<ul style="list-style-type: none"> <li>- 원표본 9068가구 중 조사 불가 가구 등을 제외하고, 총 8877가구를 대상으로 함.</li> </ul>

54 결혼 및 출산 관련 정책 수립과 평가를 위한 인구패널 구축 타당성 연구

조사명	조사 목적	조사 대상
	<p>가족관계의 변화, 가족 형성 과정과 사건 변화, 가족구조의 변화 등을 추적할 수 있도록 내용을 구성함.</p> <ul style="list-style-type: none"> <li>- 여성의 경제활동 증가와 일-가족생활 조화 실상 및 여성의 생애 전반에 걸친 일과 가족 관련 경험을 파악하여 가족의 현재와 미래에 대해 분석하고 전망함.</li> <li>- 저출산, 고령화, 보육, 일-가정 양립, 가족 친화 수준 등에 대해 정보를 추적하여 여성 및 가족 정책 수립의 기초자료로 활용됨.</li> </ul>	- 해당 가구 내 만 19~64세 여성 전체(1차 연도 1만 446명)

자료: 각 패널조사 User's Guide를 참조함(한국노동연구원, 2018; 한국보건사회연구원, 서울대학교 사회복지연구소, 2019; 한국여성정책연구원, 2018).

각 패널조사의 조사 대상을 살펴보면, 먼저 한국노동패널은 두 개의 표본으로 구분할 수 있다. 1998년 최초 원표본에서 시작한 도시 거주 5000 가구와 원가구에서 분가한 신규분가가구를 대상으로 하며, 해당 가구에 거주하는 15세 이상 가구원을 대상으로 조사한다. 2009년에 신규로 추가한 표본도 역시 추가된 원가구와 분가가구 그리고 15세 이상 가구원을 대상으로 조사한다(한국노동연구원, 2018).

한국복지패널은 2006년 원표본 7072가구와 해당 가구의 15세 이상 가구원을 대상으로 조사한다. 그리고 2012년 7차 조사에서 추가한 1800 가구와 해당 가구의 15세 이상 가구원을 대상으로 조사한다. 다만 한국복지패널조사 대상의 특징은 중위소득 60% 미만 저소득층에 전체 표본의 약 50%를 할당한 점이다. 이는 저소득층 복지실태와 욕구 및 동태 파악이라는 조사 목적을 고려한 표본 할당이라 할 수 있다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019).

여성가족패널은 2007년 원표본 9068가구 중 조사 불가 가구 등 제외하고 총 8877가구를 대상으로 조사한다. 그리고 해당 가구 내 만 19~64세 여성 전체(1차 연도 1만 446명)를 대상으로 한다. 이렇게 가구와 가구

내 특정 여성만을 조사하는 것이 여성가족패널의 특징이라고 할 수 있다. 이 역시 여성가족패널의 고유한 목적을 위해 조사 대상을 여성으로 한정하는 것이라 할 수 있다(한국여성정책연구원, 2018).

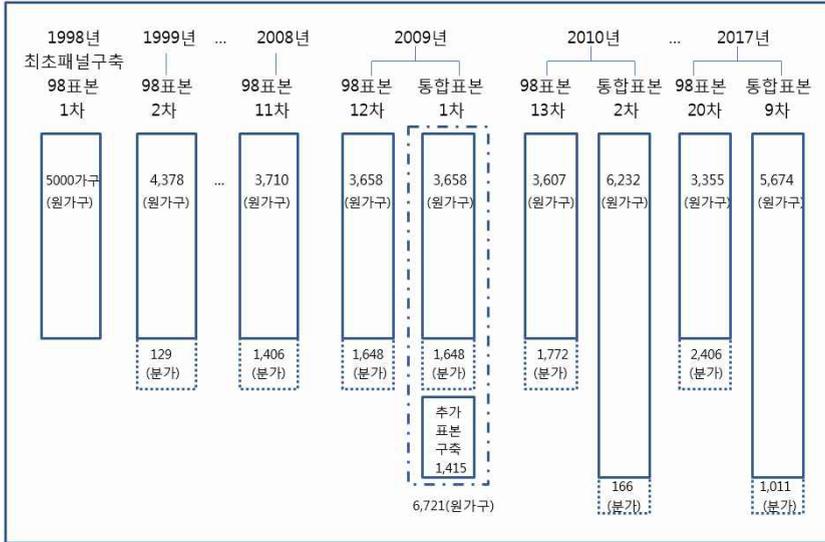
#### **다. 표본 추출과 패널 개선 노력**

패널조사는 일반적으로 최초 표본의 구축과 원표본의 유지가 중요하다. 횡단면 조사와 달리 최초 특정 표본을 계속 추적하여 동적인 자료를 구축하는 것이므로 원표본 유지 비율이 감소하면 패널의 의미가 점차 약화되기 때문이다.

여기서는 주요 패널의 표본 구축의 특징과 원표본 유지 상태, 그리고 표본 개선 과정이 어떻게 진행되었는지를 검토한다.

한국노동패널조사는 1995년 인구주택총조사의 10% 표본조사구인 전국의 2만 1938조사구 중에서 제주도의 263조사구와 군부지역 2650조사구를 제외한 전국 도시지역 1만 9025조사구(시의 동부 1만 7353조사구와 시의 읍면부 1672조사구)를 표본틀로 삼았다. 표본의 모집단은 우리나라 도시 거주 가구와 15세 이상 개인이다. 제주도와 병역의무이행자, 시설거주가구(원)는 조사 대상에서 제외하였다(한국노동연구원, 2018).

[그림 3-1] 한국노동패널의 원표본과 통합표본의 개요



자료: 한국노동연구원, 2018, p. 10.

한국노동패널은 2008년 11차 조사에서 표본 유지율이 74.2%를 나타냈다. 1998년 5000가구로 출발하였고, 2008년까지 3710가구가 유지되었다. 감소한 원표본 가구를 보완하기 위하여 2009년에 1415가구를 추가하여 원표본 가구와 통합하여 다시 ‘통합표본’을 구축하였다. 따라서 통합표본은 2009년에 구축된 ‘원표본’으로 볼 수 있다. 2017년 기준으로 원표본(20차)은 3355가구, 원표본 분가가구 2406가구였으며, 통합표본(9차)은 원가구 5674가구, 분가가구 1011가구로 나타났다. 2009년 이후에는 원표본 가구가 통합표본 가구에 완전히 포함되지 않는 경우가 있을 수 있다는 점에 유의해야 한다(한국노동연구원, 2018).

한국복지패널조사는 2005년 인구주택총조사 90% 자료에서 섬·특수시설 조사구를 제외한 23만여 개 조사구를 추출틀로 하였다. 한국복지패널조사의 목표 모집단은 2017년 현재 전국에 거주하는 가구이며, 조사

모집단은 2005년 인구주택총조사 90% 조사구로 하였다. 표본은 2단계로 구분하여 추출하였다. 우선 인구주택총조사 90% 자료에서 517개 조사구를 표본으로 추출하여 가구의 소득 및 가구원의 경제활동 상태 등을 조사하였다. 다음으로, 1단계 조사 자료에서 일반 가구와 저소득층 가구를 각각 3500가구씩 총 7000가구를 표본가구로 추출하였다. 표본추출시 중위소득 60% 미만 저소득층에 전체 표본의 약 50%를 할당하여 국내 패널조사 중에서 가장 많은 저소득층 가구를 포함하고 있다는 점에서 저소득층 대상 정책이나 빈곤 연구에 적합함이 강조된다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019).

한국복지패널조사도 한국노동패널조사와 마찬가지로 중간에 표본을 추가하였다. 2006년 1차 원표본은 7072가구로 출발하였으며, 2012년 7차 신규 표본가구 1800가구를 포함하여 조사 대상 가구 규모는 전체 6709가구였으며, 이 중 2018년에 조사가 완료된 가구는 6474가구로 96.50%의 조사 완료율을 보였다. 2018년 조사 완료 가구 중 원표본 가구는 4266가구로 최초 7072가구 대비 원표본 가구 유지율은 60.32%로 나타났다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019).

여성가족패널은 2005년 인구주택총조사 결과의 약 26만 개의 일반조사구(ED: Enumeration District)를 1차 표본조사단위(PSU: Primary Sampling Units) 추출을 위한 기본 표집틀로 설정하였다. 1차 표본조사단위 추출을 위한 일반조사구의 표집틀에서는 시설단위 조사구 및 제주도 이외의 섬지역 조사구는 제외되었다. 여성가족패널조사의 모집단은 조사 시점인 2007년 현재 대한민국 영토 내에 거주하는 가구로 하였다. 군인 및 전투경찰대, 교도소, 소년원 구치소, 경찰서 보호소 등 시설, 외국군대의 병영막사, 조사 대상에서 제외되는 외국인의 거처 등 실질적으로 조사가 불가능한 경우는 모집단에서 제외하였다(한국여성정책연구원,

2018).

앞선 두 패널조사와 달리 여성가족패널은 아직까지 표본을 추가하지 않고 있다. 2016년 6차 웨이브 조사 결과 원표본 전체 9068가구에서 조사 불가 가구와 패널 탈퇴 요구 가구, 합가 소멸 가구를 제외한 8877가구 중 조사된 원표본은 6225가구로 원표본 유지율은 70.1%를 나타낸다(한국여성정책연구원, 2018).

## 라. 주요 패널의 조사 내용

한국노동패널조사는 가구와 개인의 경제활동과 노동시장의 동태적 변화를 관찰하기 위해 설계된 조사로서, 매년 동일한 내용의 설문을 반복적으로 조사하고 있다(한국노동연구원, 2018, p. 35).

가구용은 기본 가구원의 인적 사항, 가족관계, 주거 및 보육교육, 소득과 소비 관련 항목이 주를 이룬다. 개인 관련 내용은 주로 경제활동 상태와 근로조건, 구직활동, 교육 및 직업훈련 사항을 포함한다(한국노동연구원, 2018).

한국복지패널조사도 조사 내용을 크게 가구용과 가구원용을 구분할 수 있다. 가구 사항은 일반적인 내용으로 보통의 사회조사와 크게 다르지 않지만, 건강 및 의료 관련 사항이 포함되며, 각종 사회복지서비스 이용 관련 내용이 포함된 점이 특징이다. 가구원 사항에서도 사회보험과 퇴직금 및 개인연금 수급, 생활습관 및 정신건강 관련 세부 사항 등이 포함되어 있다(한국보건사회연구원, 서울대학교 사회복지연구소, 2019).

여성가족패널조사의 조사지는 크게 가구용, 여성 개인용, 일자리용 조사로 구분할 수 있다. 가구용 조사는 가구원의 일반적 사항, 주거상태, 가구소득, 가구소비, 자산과 부채 등의 영역으로 구성되며, 패널 가구의 전

반적인 특성을 파악할 수 있는 내용이 포함된다. 여성 개인용은 교육과 성장, 직장생활, 결혼과 결혼생활, 결혼 유형의 경험, 가족과의 관계, 부양가족, 가족 관련 가치관, 개인 특성·건강·여가, 노후생활 등의 영역으로 구성되어 있으며 여성 개인의 생활세계가 생애주기별, 주제별로 구성되어 있다. 일자리용은 현재의 경제활동 상태, 고용상태별 근로여건, 일만족도, 주된 일을 제외한 일, 현재 미취업 상태, 이전 일자리, 교육 및 직업훈련, 사회보험, 직장생활과 가정생활, 차별 사항, 모성보호 제도 등 여성의 일에 대한 경험이 생애주기 및 가족관계와 접목되어 종합적으로 구성되어 있다(한국여성정책연구원, 2018).

내용면에서 주요 패널이 인구 현상과 관련한 내용을 충분히 포괄하지 못하고 있다. 주요 인구 현상이라 하면 출산, 사망, 인구 이동 관련 사항, 가족 형성 및 변화 과정 관련 사항들이 해당할 것이다. 위의 세 가지 패널 조사 중에서 내용적으로 여성가족패널이 인구 현상의 많은 부분을 포괄하지만, 결정적으로 여성만을 대상으로 조사한다는 한계가 있다. 한국노동패널이나 한국복지패널은 당초 조사 목적이 인구 현상 파악에 있지 않기 때문에 내용의 포괄성에 한계가 있다.

## 2. 국내 주요 인구 관련 자료 사례

### 가. 인구주택총조사

유엔(UN·국제연합)의 정의에 따르면 인구총조사는 (1) 개인조사, (2) 일정 지역 내에서 조사의 보편성, (3) 조사의 동시성, (4) 한 조사와 다음 조사 사이의 일정한 간격이라는 네 가지 요건을 갖추어야 한다. 이런 개념에 맞는 총조사를 근대적 총조사라고 하는데, 우리나라에서 근대적 총

조사가 시작된 것은 1925년이다. 1960년 이후 거의 모든 국가에서는 이러한 개념에 따라 총조사를 실시하고 있다(권태환, 김두섭, 2007).

한국의 인구자료 생산체계는 빠르게 발전하였다. 1960년대에서 1980년대까지가 인구등록제도를 정비하는 단계였다면 1990년대에는 전산화를 통해 주요 인구등록자료와 인구동태통계를 전산화하였고, 2000년대에는 온라인을 통해 신속하고 정확하게 인구등록을 완료할 수 있게 되었다. 2015년에는 행정자료를 결합하여 총조사의 기본항목을 생산하는 인구등록센서스로 전환하게 되었다(박경숙, 2017).

인구등록센서스로 전환된 이후, 2015년 인구주택총조사는 정부 14개 기관 25종의 행정자료를 활용하여 현장조사 없이 통계를 생산하고 있다. 11개 기관은 행정자치부, 국토교통부, 교육부, 한국전력공사, 국민안전처, 경찰청, 외교부, 국방부, 보건복지부, 대법원, 법무부이다. 각 기관에서 25종의 행정자료를 통하여 12개 기본항목을 조사한다.<sup>21)</sup>

---

21) 관련 내용은 통계청(2020, [http://www.census.go.kr/cds/cdsCensusIndc.do?q\\_menu=2&q\\_sub=22](http://www.census.go.kr/cds/cdsCensusIndc.do?q_menu=2&q_sub=22))에서 2020. 2. 3. 인출, 정리하였다.

〈표 3-4〉 연차별 각 총조사 개요

조사 연도	조사 명칭	실시 주관 기관	기준 시점	주요 특징
1925	간이국제조사	조선 총독부	10. 1.	최초의 인구총조사
1930	조선국제조사	"	10. 1.	최초로 직업 등 경제활동조사
1935	조선국제조사	"	10. 1.	
1940	국제조사	"	10. 1.	
1944	간이국제조사	"	5. 1.	
1949	제1회 총인구조사	공보처 통계국	5. 1.	정부 수립 후 첫 총조사, 최초로 인구이동 사항 포함
1955	제1회 간이총인구조사	내무부 통계국	9. 1.	
1960	인구주택국세조사	경기원 통계국	12. 1.	처음으로 주택 사항 포함, 경제활동 및 출산력 표본 집계
1966	인구센서스	경기원 통계국	10. 1.	10% 표본조사 병행(경제활동 및 출산력)
1970	총인구 및 주택조사	"	10. 1.	10% 표본조사 병행(경제활동, 출산력, 이동)
1975	총인구 및 주택조사	"	10. 1.	5% 표본조사 병행(경제활동, 출산력, 이동)
1980	인구 및 주택조사	"	11. 1.	15% 표본조사 병행(경제활동, 출산력, 이동)
1985	인구 및 주택조사	"	11. 1.	전 항목 전수조사
1990	인구주택총조사	통계청	11. 1.	10% 표본조사 병행(경제활동, 출산력, 이동 등)
1995	인구주택총조사	"	11. 1.	10% 표본조사 병행
2000	인구주택총조사	"	11. 1.	"
2005	인구주택총조사	"	11. 1.	"
2010	인구주택총조사	"	11. 1.	"
2015	인구등록센서스	"	11. 15.	행정등록자료 결합조사, 인구기본항목 및 심층항목 20% 표본조사

자료: 구성렬, 1996, p. 13; 박경숙, 2017, p. 61.

기본항목은 인구 분야에서 성명, 성별, 나이, 가구주와의 관계, 국적, 입국 연월, 본관이며, 가구 관련 사항 그리고 주택 관련 항목은 거처의 종류, 주거용 연면적, 대지면적, 건축 연도이다. 그러나 행정자료를 통하여 파악되지 않는 심층항목에 대해서는 20% 가구 표본을 대상으로 기존과 같은 방식으로 현장조사를 실시하고 있다.<sup>22)</sup>

한편 인구등록센서스에 대해 의문이 제기되기도 한다. 즉 행정자료를 통해 구성된 인구가 정확히 무엇인가 하는 질문이 제기되기도 한다. 예를 들어 같은 주소이면서도 세대가 분리되었던 사례가 가족등록정보로 하나의 가구로 통합될 때에 통합된 가족 구성이 실제 가족관계를 반영한다고 단정할 수 없다는 것이다. 그리고 혼인 상태도 사실혼과 법적 혼인의 차이가 드러나지 않는다. 등록이라는 것이 법적인 지위를 획득하는 방법이 기 때문에 등록 자료에 기초하여 인구를 집계하는 상황에서 등록이 어떤 목적에서 어떻게 범주화되었는가는 중요한 쟁점이 될 수 있는 것이다(박경숙, 2017).

위와 같은 문제가 지적되지만 인구등록센서스는 인구 현상을 파악하기 위한 대부분의 항목을 포괄하기 때문에, 인구 현상을 분석하는 데에 횡단면 조사로서 중요한 자료로 활용되고 있다.

## 나. 인구동향조사

인구동태통계는 주로 사람의 생명현상 및 혼인과 관련된 인구학적 사건에 대한 통계를 말하며, 일반적으로 출생, 사망, 사산, 혼인, 이혼의 다섯 가지를 들 수 있다. 그리고 동태통계는 이 사건이 발생할 때마다 이루

22) 관련 내용은 통계청(2020, [http://www.census.go.kr/cds/cdsCensusIndc.do?q\\_me nu=2&q\\_sub=22](http://www.census.go.kr/cds/cdsCensusIndc.do?q_me nu=2&q_sub=22))에서 2020. 2. 3. 인출, 정리하였다.

어지는 신고에 기초해 작성된다. 총조사와 동태신고의 가장 근본적인 차이는 총조사는 동시조사인 데 비해 동태신고는 사건이 일어날 때마다 이루어지는 지속적인 조사라는 점이다.

해방 이후부터 1970년대까지 우리나라의 동태통계는 질이 낮아 연구 유용성이 떨어졌다. 그러나 1980년대 이후 신고제도의 개선과 행정가들에 대한 교육에 힘입어 출생과 사망에 관한 신고 자료는 그 완전성이 90% 이상이 되었으며 주요 인구자료로서 그 가치를 인정받게 되었다(권태환, 김두섭, 2007).

인구동향조사는 기본적으로 신고에 기초한다. 절차에 따라 신고된 출생, 사망, 혼인, 이혼 자료는 해당 지방자치단체의 인구동태입력시스템에 입력되고, 입력된 내용은 통계청의 검토와, 전산 자료의 중복·누락·입력 오류 확인을 거쳐 최종 통계로 작성된다(박경숙, 2017).

최근의 인구동향조사 자료는 정확하게 생산, 관리되고 있으며, 인구의 동태적 측면을 분석하는 데에 중요하고 유용한 자료로 활용되고 있다.

## 다. 출산력조사

출산력조사는 한국보건사회연구원에서 수행하는 횡단면조사로서 현재의 공식 명칭은 ‘전국 출산력 및 가족보건·복지 실태조사(이하 전국 출산력조사)’이다. 전국 출산력조사의 최초는 1964년 ‘전국 가족계획 실태조사’이며, 이것이 현재의 ‘전국 출산력 및 가족보건·복지 실태조사’로 이어지고 있다.

1964년 최초 조사는 합계출산율을 추정하거나 기혼 부인의 피임 경험률을 파악하는 데 활용되었다. 1960년대는 저조한 출생신고율 때문에 출생신고자료를 활용하여 연령별출산율 및 합계출산율을 구하기 어려웠다.

1960년대 출산율은 단지 인구 1000명당 40명 수준을 초과하는 것으로 추정하였을 뿐 자세한 자료를 제시하지는 못하였다(한국보건사회연구원, 1991). 이에 따라 1964년 전국 가족계획 실태조사 결과는 출산율을 추정하는 중요한 자료로 활용되었다.

2018년 조사를 기준으로, 전국 출산력조사는 15~49세 기혼 여성 가구 중 1만 1207명을 대상으로 표본조사를 실시하였다. 조사는 3년 주기로 한국보건사회연구원에서 조사하고 있다(이소영 외, 2018).

2018년 전국 출산력조사의 표본설계를 위한 목표 모집단은 2018년 현재 전국 17개 시·도의 일반 주거 시설에 거주하는 가구로 하였다. 조사 모집단은 2016년 인구주택총조사 조사구의 가구를 표본틀로 구성하였다. 전체 조사구 중 700개의 조사구를 추출하고 조사구당 15가구를 조사 완료 목표로 선정하였다. 2018년 최종 조사가 완료된 15~49세 기혼 여성은 1만 1207명이었다(이소영 외, 2018).

조사의 주요 내용은 결혼 이력, 임신 및 출산 이력, 산전후 관리, 자녀 양육 실태, 가족 관련 가치관, 결혼 및 출산 관련 정책에 관한 것이며, 이들을 조사 항목으로 다룬다(이소영 외, 2018).

전국 출산력조사는 인구 이동 사항을 제외하고 혼인, 출산, 입양 등의 인구 현황을 파악하기 위한 주요 변수 대부분을 포함한다. 이뿐만 아니라 인구동태의 관련 변인에 해당하는 자녀 양육 관련 사항, 경제활동 사항 등을 상세히 조사하여 학술적 정책적 연관분석의 활용도가 높게 평가되고 있다(이소영 외, 2018).

〈표 3-5〉 전국 출산력조사 이력

조사 연도	조사 명칭	조사 대상
1964	전국 가족계획 실태조사	전국 표본지역에 거주하는 15~54세의 기혼 부부(4000명)
1965	전국 가족계획 실태조사	전국 표본지역에 거주하는 45세 미만의 유배우 부인(3445명)
1966	전국 가족계획 실태조사	전국 표본지역에 거주하는 45세 미만의 유배우 부인(3368명)
1967	전국 가족계획 실태조사	전국 표본지역에 거주하는 50세 미만의 유배우 부인(3624명)
1968	전국 출산력 조사	전국 표본지역 내 조사 가구(8774가구)와 동 가구에서 50세 미만의 기혼 부인(7477명) 및 유배우 부인(6889명)
1971	전국 출산력 및 인공임신중절 조사	전국 표본지역 내 조사 가구(6800가구)와 동 가구에서 55세 미만의 기혼 부인(6285명) 및 유배우 부인(5629명)
1973	가족계획 종합실태 조사	전국 표본지역 내 조사 가구(3160가구)와 동 가구에서 50세 미만의 기혼 부인(3038명)과 18~27세의 미혼 여성
1974	한국 출산력 조사	전국 표본지역 내 조사 가구(2만 932가구)와 동 가구에서 3분의 1에 해당하는 표본가구에 거주하는 50세 미만의 기혼 부인(5417명)과 표본 조사구에 대한 지역 특성조사
1976	전국 출산력 및 가족계획 실태조사	전국 표본지역 내 조사 가구(8160가구)와 동 가구에서 50세 미만의 기혼 부인(6020명) 및 유배우 부인(15~44세: 5008명)
1978	가족계획 및 출산력 실태조사	전국 표본지역 내 조사 가구(5069가구)와 동 가구에서 표본추출에 의한 50세 미만의 유배우 부인(3603명)
1979	페임보급 실태조사	전국 표본지역 내 조사 가구(1만 9788가구)와 동 가구에서 표본추출에 의한 50세 미만의 기혼 부인(1만 4303명)
1982	전국 가족보건 실태조사	전국 표본지역 내 조사 가구(7375가구)와 동 가구에서 49세 이하 기혼 부인(5371명) 및 44세 이하 유배우 부인(4351명)
1985	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(7415가구) 및 15~49세 기혼 부인(8421명)
1988	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 1864가구) 및 15~44세 기혼 부인(7792명)

66 결혼 및 출산 관련 정책 수립과 평가를 위한 인구패널 구축 타당성 연구

〈표 3-5〉 전국 출산력조사 이력(계속)

조사 연도	조사 명칭	조사 대상
1991	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 1540가구) 및 15~49세 기혼 부인(7384명) 및 18~34세 사이의 미혼 남녀(3715명: 남성 1991명, 여성 1724명)
1994	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 613가구) 및 동 가구에 거주하는 15~49세 기혼 부인(6060명)
1997	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 1216가구) 및 동 가구에 거주하는 50세 미만(15~49세)의 기혼 부인(6450명: 임신·출산·피임·모자보건·자녀가치관·성선행호 분석 대상은 15~44세 유배우 부인) 및 65세 이상의 노인 조사
2000	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 1388가구) 및 동 가구에 거주하는 15~64세의 기혼 부인(8935명: 임신·출산·피임·모자보건 분석 대상은 15~44세 유배우 부인)
2003	전국 출산력 및 가족보건 실태조사	전국 표본지역 내 조사 가구(1만 5758가구) 및 동 가구에 거주하는 15~59세 기혼 가구(1만 1111명: 임신·출산·피임·모자보건 분석 대상은 15~44세 유배우 부인)
2006	전국 출산력 및 가족보건·복지 실태조사	전국 표본지역 내 조사 가구(1만 4497가구) 및 동 가구에 거주하는 15~59세 기혼 가구(1만 132명: 임신·출산·피임·모자보건 분석 대상은 15~44세 유배우 부인)
2009	전국 출산력 및 가족보건·복지 실태조사	전국 표본지역 내 조사 가구(1만 3883가구) 및 동 가구에 거주하는 15~64세 기혼 가구(9961명: 임신·출산·피임·모자보건 분석 대상은 15~44세 유배우 부인)
2012	전국 출산력 및 가족보건·복지 실태조사	전국 표본지역 내 조사 가구(1만 4970가구) 및 동 가구에 거주하는 15~64세 기혼 가구(1만 71명: 임신·출산·피임·모자보건 분석 대상은 15~44세 유배우 부인)
2015	전국 출산력 및 가족보건·복지 실태조사	15~49세 기혼 여성 가구(1만 1009명)
2018	전국 출산력 및 가족보건·복지 실태조사	15~49세 기혼 여성 가구(1만 1207명)

자료: 신창우 외, 2011, pp. 20-23; 이소영 외, 2018 보고서를 참조하여 재구성.

## 제2절 해외 주요 인구 관련 패널 자료

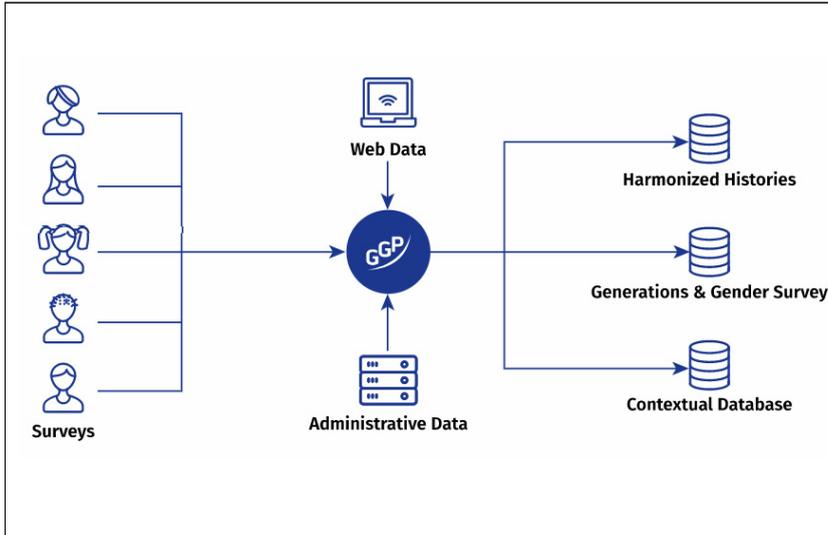
### 1. Generations and Gender Survey(GGS)<sup>23)</sup>

최근 유럽에도 저출산 현상이 나타나고 있으며, 남아메리카에도 저출산 현상이 심각해지고 있다. 이에 따라 저출산 현상을 이해하고 이에 대응한 정책을 수립하고자 하는 국제적인 공통의 목적으로 Generations and Gender Survey(GGS)가 적극적으로 수행되고 있다. 해당 조사는 1990년대 출산력과 가족 조사(Families and Fertility Survey)라는 명칭을 가지고 시작되었는데, 이후 세대와 성 조사(Generations and Gender Survey)로 명칭이 변경되었다. 조사 대상은 국가별 약 1만 명의 19~79세 남녀 또는 약 5000명의 19~49세 남녀 개인이다. 조사 목적은 인구와 가족 변동(dynamics)과 관련된 데이터를 생산하여 관련 현상을 이해하고 과학적 근거에 기반한 정책을 수립하는 것이다. 조사 방식은 횡단조사(cross-sectional)와 패널조사(panel)와 차별적인 종단(longitudinal)조사이다. 이를 통해 개인의 출산 의향과 계획이 실제로 이행되었는지를 파악할 수 있다. GGS의 데이터 관리 구조는 다음과 같다.

---

23) GGS와 관련된 본문의 내용은 관련 홈페이지의(UNECE(2020), <http://www.ggp-i.org>) 내용과 GGP(Generations and Gender Programme) 중앙본부 담당자와의 면담 내용을 토대로 구성하였다.

[그림 3-2] GGS의 데이터 관리 구조



자료: UNECE(2020). GGP. <https://www.ggp-i.org/>에서 2020.1.30. 인출.

이전 조사인 2004년 조사에는 유럽을 중심으로 19개 이상의 국가에서 18~79세의 20만 명 이상이 참여하였다. 그 결과로 GGS 첫 번째 웨이브에서 호주, 오스트리아, 벨라루스, 벨기에, 불가리아, 체코, 에스토니아, 프랑스, 조지아, 독일, 헝가리, 이탈리아, 일본, 리투아니아, 네덜란드, 노르웨이, 폴란드, 루마니아, 러시아, 스웨덴 등 20개 국가 자료가 현재 공개, 제공되고 있다. GGS 두 번째 웨이브는 호주, 오스트리아, 불가리아, 체코, 프랑스, 조지아, 독일, 헝가리, 이탈리아, 리투아니아, 네덜란드, 폴란드, 러시아, 스웨덴 등 14개 국가에 대한 것이 가용한 상태다. 조사 문항은 1000개 이상의 핵심 설문으로 이루어져 있으며 부모 자녀의 관계에 대한 부모 관점 및 자녀 관점, 배우자(파트너)와의 관계, 젠더 관점, 출산 이력, 인공임신중절 및 난임 치료, 가구 사항, 경제활동 상황, 소득 및 자산, 교육, 건강상태, 개인 사회망, 복지제도, 삶의 질, 가치관 등을 포함한다.

GGG(Generations and Gender Survey)는 유럽 및 선진국 사회 과학의 기초 지식을 제공하고 정책 수립에 기초 자료를 제공한다. GGS는 다음과 같은 특징이 있다.

첫째, GGS의 가장 큰 장점은 국제 비교가 가능하다는 것이다. 국제 비교를 위해 각 국가에서 공통적인 설문지와 방법론에 대한 지침을 기반으로 하여 조사가 이루어진다. 데이터 처리는 국가별 데이터가 중앙에서 합해지는 방식(harmonization)으로 이루어진다. 둘째, 조사 대상자는 만 18~79세로 연령대가 다양하다. 셋째, 종단 조사로서 3년 간격으로 동일한 사람을 면접하는 패널 디자인이다. 넷째, 표본 집단의 크기는 첫째 웨이브에서 국가당 평균 9000명이다. 다섯째, 조사 문항은 계획된 행동 이론 등 이론에 근거하며 다양한(multidisciplinary) 분야의 내용을 포함하므로, 수집된 데이터는 인구통계학자, 경제학자, 사회학자, 사회정책학자, 사회심리학자, 역학자에게 제공되어 이들의 정책 관련 연구에 기초 자료를 제공한다. 여섯째, 설문 조사 데이터를 GGP 데이터베이스에서 제공하는 매크로 데이터와 결합할 수 있으며 이를 통해 문화·경제·정치·사회·정책적 문맥에서 개인과 가족에 대한 분석을 할 수 있다.

## 2. The Netherlands Kinship Panel Study(NKPS)<sup>24)</sup>

The Netherlands Kinship Panel Study는 네덜란드에서 수행하는 GGP(Generation and Gender Programme) 조사이다. 따라서 조사의 주된 주제는 ‘가족 관계에서의 상호 친화감과 이러한 감정이 행동으로 표현되는 방식으로 정의되는 연대(solidarity)’이다. 해당 조사는 네덜란드

24) NKPS와 관련된 본문의 내용은 관련 홈페이지(NIDI, 2020, <http://www.nkps.nl/>)의 내용을 토대로 작성되었다.

의 인구연구소(The Netherlands Interdisciplinary Demographic Institute)와 주요 대학들(Utrecht University, Erasmus University Rotterdam, University of Amsterdam)이 주도하여 개발되었다. 네덜란드 국가 연구 재단(Dutch national research foundation)이 조사 예산을 지원한다. 2014년까지 네 번의 웨이브가 조사되었다(Wave 1 in 2002-2004, Wave 2 in 2006-2007, Wave 3 in 2010-2011, Wave 4 in 2014). 해당 패널 조사의 대상은 첫째 웨이브에서 9500명이었고 구조화된 면접(structured interviews)과 심층 면접(in-depth open interviews) 등 직접면접 방식으로 조사가 수행되었다.

### 3. Panel Study of Family Dynamic(PSFD)<sup>25)</sup>

The Panel Study of Family Dynamic(PSFD)은 1999년에 시작되었다. PSFD의 시작 단계에서 조사 대상은 대만에 거주하는 성인이었다. 패널 설문 조사 문항의 핵심 주제는 응답자와 그 가족의 상호 관계와 응답자와 가족의 행동이다. 2000년부터 조사 대상에 자녀가 포함되었다. 응답자와 응답자의 자녀에 대한 첫 번째 조사 및 후속 조사는 대면 조사로 이루어진다. PSFD를 주관하는 대만 국립중앙연구원 인문사회과학연구센터는 대만의 체제가 다소 안정된 후 중국 사회과학아카데미와 협력하여 설문 조사를 중국의 동남 해안선 지역까지 확대하였다. 해당 조사는 대만 해협에 걸쳐 데이터를 수집하여 중국의 가족에 대한 연구와 지역 비교 연구에 중요한 자료로서 가치가 있다.

---

25) PSFD와 관련된 본문의 내용은 관련 홈페이지(대만국립중앙연구원, 2020, [https://psfd.sinica.edu.tw/web/plan\\_02en.htm](https://psfd.sinica.edu.tw/web/plan_02en.htm))의 내용을 토대로 작성되었다.

### 가. 대만 조사(Taiwan Survey)

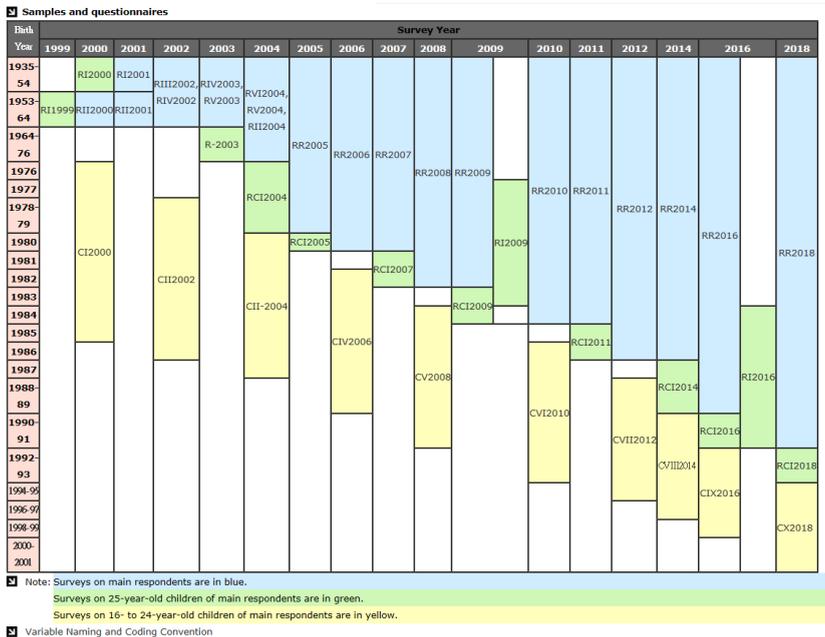
1999년에 시작된 PSFD의 첫 번째 조사 대상은 1953-1964년에 태어난 개인이다. 1935-1954년, 1964-1976년, 1977-1983년, 1984-1991년에 태어난 성인 응답자는 각각 2000년, 2003년, 2009년, 2016년에 처음으로 조사되었다. 이 다섯 집단의 주요 응답자들에 대한 후속 조사는 2012년 전까지는 매년 실시되었고 2012년 이후에는 2년마다 실시되었다. 후속 조사의 주요 설문 문항은 응답자의 인구 통계학적 특성, 업무 상태 및 직업 정보, 결혼 여부 및 배우자 정보, 부모 및 시부모의 인구 통계학적 특성, 가족과의 상호작용, 주택 및 주거 환경, 수입 및 지출, 출산 및 양육 정보 등을 포함한다.

응답자의 자녀인 아동의 경우 25세가 되면 주요 응답자로 간주되고 주요 응답자의 첫 번째 설문지(질문 RCI)를 사용하여 조사를 한다. 이후 주요 응답자(질문 R)와 동일한 설문지를 사용하여 다시 조사를 하고 매년 추가 조사를 실시한다. 조사에 대한 부담을 줄이기 위해 2012년부터 주요 응답자(질문 R)에 대한 후속 조사, 설문 C, 설문 RCI가 동시에 2년마다 수행된다. 주요 응답자의 5개 그룹 중 1999년, 2000년, 2003년, 2009년, 2016년에 실시된 1차 조사의 응답자 수는 각각 999명, 1960명, 1152명, 2182명, 1972명이었다. 2018년의 전체 응답자 수는 468명, 804명, 527명, 1418명, 1576명이었다. 같은 해에 아동 표본에 대한 전체 조사 횟수는 총 2153건으로 설문지 R은 1631건, 설문지 C는 1631건이었다.

### 나. 중국 대륙 조사(China Survey)

중국 대륙 조사는 기존에 수행되었던 대만에서 조사 지역을 확장하여 중국 전체 가족에 대해 분석하고자 실시되었다. 설문 문항은 대체로 기존 대만 조사의 설문 문항과 동일하게 구성되었다. 다만, 대만과 중국 대륙 간의 언어적 차이를 반영하여 약간의 차이가 있는 문항이 있고 조사의 주체인 중국 사회과학아카데미(IPLE-CASS)의 연구 관심사를 반영하여 추가된 문항이 있다. 전체 응답자 수는 4684명이었다. 중국 본토의 후속 조사는 2006년, 2011년, 2013년, 2017년에 수행되었다. 2017년에 완료된 인터뷰는 3535건이며, 후속 사후 조사는 2019년으로 예정되어 있다.

[그림 3-3] PSFD 조사 대상과 기간



자료: 대만국립중앙연구원(2020). PSFD Structure. <https://psfd.sinica.edu.tw/web/checken.htm>에서 2020. 1. 31. 인출.

#### 4. Longitudinal Study of Generations, California(LSOG) 1971, 1985, 1988, 1991, 1994, 1997, 2000, 2005 (ICPSR 22100)<sup>26)</sup>

가족 내 세대 간 관계를 주제로 하는 조사인 The Longitudinal Study of Generations(LSOG)는 1971년에 시작되었다. 조사 대상은 60대 조부모, 40대 초반 중년 부모, 15~26세 손자의 3세대가 있는 캘리포니아에 거주하는 300가구이다. 조사 대상은 점차 확대되어 1991년부터 같은 가구의 증손자인 4세대를 포함한다. 이 조사는 세대 간 시간적 흐름을 정교하게 파악할 수 있도록 설계되어 있다. 따라서 동일한 생애주기에 놓여 있으면서도 다른 세대인 고령의 부모와 자녀를 비교할 수 있다. 이러한 비교를 통해 35년간, 4세대에 걸친 세대 간 연대 또는 갈등에 대한 사회적 변화의 영향을 분석할 수 있다. 그뿐만 아니라 스트레스가 큰 삶의 전환(예: 노화, 이혼, 재혼, 높은 여성 노동력 참여, 노동 및 경제의 변화, 가족의 의무 규범 약화)을 완화하는 가족의 사회적 변화에 대한 영향을 분석할 수 있다. 또한 사회 변화가 세대 간 가치, 자원, 행동이 전달되는 데 미치는 영향에 대해서도 분석이 가능하다. 이는 LSOG 데이터에 인구 통계학적 변수로 연령, 성별, 소득, 고용 상태, 결혼 상태, 사회경제적 이력, 교육, 종교, 인종, 병역이 포함되어 있을 뿐만 아니라 가족 구조, 가구 구성, 연대 및 갈등, 가치, 태도, 행동, 역할의 중요성, 결혼 관계, 건강 및 운동 상태, 정신건강 및 복지, 돌봄, 여가 활동 및 생애 사건 및 걱정 등에 대한 정보도 포함되어 있기 때문이다.

26) ICPSR(2020), <https://www.icpsr.umich.edu/icpsrweb/NACDA/studies/22100/datadocumentation> 참조.

## 5. The British Household Panel Survey(BHPS)<sup>27)</sup>

에식스대학교의 사회 경제 연구소에서 수행된 영국가구패널조사(BHPS)는 사회 경제 연구를 위한 패널조사로서 첫 번째 조사는 1991년에 시작되었고 이때 패널로 모집된 조사 대상은 영국 전역에 거주하는 1만 300명이다. 이 조사는 2008년부터 Understanding Society라고 불리는 UKHLS에 통합되었다. BHPS 데이터는 유럽 Community Household 패널과 호주, 캐나다, 독일, 영국, 미국의 패널 데이터를 포함하는 CNEF(Cross National Equivalent File)에 통합되었다.

### 가. 조사 대상<sup>28)</sup>

BHPS는 1991년에 모집된 패널 약 5500가구의 전국 대표 표본으로 구성된 설문 조사로, 약 1만 명의 인터뷰 대상자가 포함되어 있다. 조사 대상은 우편 번호 주소 파일에서 가져온 계층화된 클러스터 디자인에 의해 표집되었으며 설문 조사의 첫 번째 웨이브(wave)에 해당 주소지에 거주하는 모든 사람을 패널로 구성하였다. 이렇게 구성된 패널을 대상으로 매년 조사가 이루어졌고, 원세대에서 분리되어 새로운 세대를 구성하게 되면 이들을 추적하여 이들을 대상으로 조사하는 방식으로 수행되었다. 또한 새롭게 패널 표본으로 선정된 경우와 아동이 16세가 되는 경우에는 패널로서 처음 조사에 응답하게 된다. 1999년에는 BHPS의 조사 대상(패널)에 스코틀랜드와 웨일스에 있는 1500가구까지 확장되어 영국 내 각

27) ISER(2020), LSOG와 관련된 본문의 내용은 관련 홈페이지(<https://www.iser.essex.ac.uk/bhps/>)에서 조사 항목(<https://www.iser.essex.ac.uk/bhps/about/questionnaire-content>) 각 항목별 내용을 포괄적으로 인용하였다.

28) 조사 대상은 관련 웹페이지(ISER, 2020, <https://www.iser.essex.ac.uk/bhps/faqs/sample>)를 저자가 번역하여 작성하였다.

지역을 독립적으로 분석할 수 있게 되었다. 2001년에는 북아일랜드에 거주하는 2000가구 패널이 추가되었다. 최종적으로 BHPS 패널의 전체 크기는 확장된 패널을 포함하여 영국에 거주하는 약 1만 가구이다.

#### 나. 주요 조사 문항<sup>29)</sup>

조사 대상은 조사 대상 가구에 속한 16세 이상 가구원으로 기본적인 조사의 대상자이다. 또한 11~15세 사이의 어린이는 1994년에 도입된 자기 기입식 청소년 설문지를 작성한다. 성인 개인을 대상으로 하는 조사는 약 45분이 소요되고, 가구원 중 한 사람을 대상으로 추가 가구조사를 실시한다. 주요 문항은 다음 사항을 포함하여 사회과학과 정책 분야의 문항을 폭넓게 포함한다. 즉, 가구 구성, 주거 상황, 거주 이동, 교육 및 훈련, 건강 및 의료서비스의 이용, 고용시장 참여 행태, 사회경제적 가치, 소득, 연금 등 공적 보험 등을 포함한다. 또한 매년 조사하지 않는 부가적인 문항으로 정책 및 연구 문제의 변화로 인해 발생하는 새로운 질문 및 첫 번째 인터뷰 전에 패널의 생애사에 대한 회고적인 데이터를 도출하기 위한 문항이 포함된다. 또한 결혼·동거·출산 이력, 직업 이력, 부와 자산, 추가적인 건강 지표, 노화, 은퇴, 삶의 질, 자녀 및 양육, 이웃과 사회 관계망 등이 포함된다.

29) 조사 문항은 관련 웹페이지(I SER, 2020, <https://www.iser.essex.ac.uk/bhps/about/questionnaire-content>)를 저자가 번역하여 작성하였다.

## 6. China Family Panel Studies(CFPS)<sup>30)</sup>

중국 가족패널 연구(CFPS: China Family Panel Studies)는 중국 북경대학교의 사회과학 조사 연구소(ISSS)가 2010년부터 수행하고 있는 중국 사회, 가족 및 개인에 대한 전국을 대표하는 패널조사이다. CFPS는 중국의 개인, 가족 및 지역 차원의 종단 데이터를 수집하기 위해 설계되었다. 이 조사는 경제활동, 교육, 가족 역할 및 관계, 이주 및 건강 등의 주제를 포함한 정보뿐만 아니라 중국 인구의 경제활동 및 비경제활동과 복지(well-being)를 다룬다. 조사는 25개 성(省)·시·지역의 1만 6000가구의 모든 가족 구성원을 대상으로 실시하고 있다. 조사 예산은 중국 정부가 베이징 대학을 통해 지원하고 있다. CFPS의 주요 특징은 다음과 같이 요약될 수 있다. 조사 대상은 표본 가구에 거주하는 9세 이상의 모든 가구 구성원이다. 각 가구원은 CFPS의 핵심(core) 패널(표본)이며 이들은 사망할 때까지 이 조사의 대상이 된다. CFPS의 모든 핵심 패널의 부가적인 조사는 매년 수행되도록 설계되어 있다. 표집의 경우 지역을 비교할 수 있도록 시작 단계에서 5개 주(각 1600가구)에 과표집을 한다. CFPS의 나머지 표본(8000가구)은 멀리 떨어져 있는 몇 개의 예외적인 지역을 제외하고 다른 지방에서 추출하여 가중치를 통해 국가 전체를 대표하도록 설계되어 있다. 2010년 초기 조사의 표본은 설문 조사의 운영비를 줄이고 사회적인 맥락을 분석하는 것이 가능하도록 계층화 다단계 확률 표집을 실시하였다. CFPS 연구의 각 하위 표본 집단은 지역사회(county), 마을(village), 가구의 세 단계로 구성되어 있다. 조사는 미시간 대학교(University of Michigan)의 조사센터(Survey Research Center)에서

30) CFPS와 관련된 본문의 내용은 관련 홈페이지(ISSS, 2020, <https://opendata.pku.edu.cn/dataverse/CFPS>)의 내용을 저자가 번역하여 구성하였다.

제공하는 CAPI(Computer Assisted Personal Interviewing) 방법을 활용하여 실시되고 있다. 2010년 조사에서 응답자는 약 1만 5000가구 3만 명의 가구원이었으며, 약 79%의 응답률을 보였다.

### 제3절 ‘인구패널’을 위한 시사점

앞서 국내 주요 패널조사 사례 및 인구 관련 조사의 주요 특징을 살펴 보았는데, 인구패널의 특성 및 도입의 측면에서 몇 가지 시사점을 찾을 수 있다.

우선 패널조사의 중요한 속성 중 하나는 원표본을 계속 추적하여 변화를 관찰하는 것이다. 원표본이 조사에서 이탈하지 않고 안정적으로 유지되어야 패널조사의 의미가 퇴색하지 않는다. 따라서 패널조사의 구축 및 관리에서 원표본 유지율에 관심을 갖게 된다. 특히 현대 사회가 변화하면서 개인의 정보보호 인식이 강화되고 사적 영역에 대한 보호 인식도 강화되면서 조사 환경이 더 어렵게 조성되고 있다. 이러한 사회 변화는 패널 조사에서 원표본 유지율을 약화하고 있다.

앞서 살펴본 사례에 따르면 국내 주요 패널조사의 경우 대체로 원표본 유지율이 떨어져 표본 개선 노력을 하고 있다. 한국노동패널은 최초 1998년 표본과, 2009년 추가 표본을 합한 통합표본으로 패널을 양분할 수 있다. 한국복지패널조사도 2006년 1차 표본과 2012년 추가한 통합표본으로 구분할 수 있다. 이렇게 표본이 추가되면 분석의 필요에 따라 원표본만을 분석할 수도 있고, 통합표본을 분석할 수도 있다. 그러나 원표본과 통합표본은 분명히 다른 표본일 수밖에 없다.

이런 점을 고려하면 인구패널을 구축하면서 패널을 안정적으로 유지하

고 지속할 수 있는 방안을 충분히 고려해야 하며, 향후 원표본 유지율이 떨어질 때의 보완책도 사전에 준비할 필요가 있다.

둘째로 기존 패널조사의 내용과 목적에 관한 시사점을 살펴볼 수 있다. 기존 패널조사는 목적이 명확하고 그에 따라 설계되어 있다. 즉, 조사 내용이 당초 목적에 따라 명확히 규정되어 있기 때문에 다른 목적으로 활용하는 데 한계가 있다. 한국복지패널이나 여성가족패널이 인구 사건(events)을 일부 포함하고 있지만, 기존 패널 자료가 인구 특성 분석을 위해 충분한 조건을 갖추고 있지는 못하다. 한국노동패널이 출산 사항을 포함하지만, 추적관찰을 위해 중단적으로 인구 사건을 추출해 보면 분석을 위한 충분한 관측치를 확보하기 어렵다. 인구 현상 분석 관점에서 볼 때, 이런 한계는 기존의 모든 패널에서 공통으로 나타난다. 그리고 인구 변동의 3요소에 해당하는 출산, 사망, 이동의 측면에서 주요 항목이 다 포함되어 있지 않은 점도 문제로 지적될 수 있다. 특히 이동 관련 항목은 거의 찾아보기 힘들다. 이런 점을 고려하면 기존 주요 패널조사자료를 인구분석에 직접, 즉시 활용하기 어렵기 때문에 새로운 인구패널의 구축을 검토할 필요성이 인정된다.

셋째로 인구 관련 조사의 시사점을 살펴볼 수 있다. 우선 인구주택총조사와 인구동향조사는 기본적으로 표본조사가 아닌 전수조사로 볼 수 있다. 내용적으로도 인구 사건의 주요한 내용을 포함하고 있어 핵심적인 분석자료에 해당한다. 그러나 개인단위 또는 가구단위 변화를 추적하여 관찰하기 위해서는 많은 검토와 사회적 논의가 필요하다.

행정적으로 개인 식별이 가능할 수 있지만, 인권의 관점에서 개인정보 보호 필요성이 제기되고 있고, 실제로 개인정보보호 관련 법률의 제한도 작용하고 있는 실정이다.

한편, 개인식별정보를 이용하여 자료의 패널적 성격을 확보하되, 개인

정보는 보호할 수 있는 방안을 찾을 수 있다는 의견도 있다. 개인식별정보를 재부호화하거나 암호화하여 자료의 연계성만 확보하고, 실제로 개인의 정보는 노출하지 않을 수 있다는 것이다. 이런 과정이 제도화되기 위해서는 관련 법률의 개정과 사회적 논의가 충분히 진행되어야 할 것이다. 따라서 즉시 패널 자료로 전환하여 활용하기에는 한계가 있을 것으로 판단된다.

그리고 출산력조사는 오랜 전통을 가진 주기적 횡단조사로 지속되고 있다. 내용적으로도 인구 사건의 많은 내용을 포함한다. 특히 혼인 이력과 출산 이력은 상세한 내용을 담고 있다. 그리고 피임 관련 사항, 모자보건 등의 내용과 혼인 출산에 영향을 미치는 사회경제적 주요 사항을 조사하고 있다. 따라서 내용적인 측면을 보면 출산력조사를 인구패널로 전환하는 것을 검토해 볼 수 있을 것이다.

해외 사례의 경우, 앞에서 살펴본 네덜란드의 The Netherlands Kinship Panel Study(NKPS) 및 Generations and Gender Survey(GGS), 미국 캘리포니아주의 Longitudinal Study of Generations, California(LSOG), 영국의 The British Household Panel Survey(BHPS), 대만의 Panel Study of Family Dynamic(PSFD), 중국의 China Family Panel Studies(CFPS)와 같이 결혼, 출산, 가족 구성원 간의 역동(dynamics)을 관찰할 수 있는 패널조사가 오랫동안 실시되어 왔다. 이러한 국외 패널조사는 대체로 가족에 대한 내용을 포괄적으로 다루기 때문에 가족의 형성과 변화를 파악하기 위해 일회성의 횡단조사의 방식을 취하지 않고 변화를 추적할 수 있는 패널조사의 방식을 취했다고 볼 수 있다. 이러한 패널조사들은 가족을 중심으로 인구의 변화를 파악하는 데 유용한 자료로 활용되고 있다.

특히 인구와 가족 변동(dynamics)과 관련된 데이터를 생산하여 관련 현상을 이해하고 과학적 근거에 기반한 정책 수립을 목적으로 하는 Generations and Gender Survey(GGS)는 인구패널의 대표적인 예로 볼 수 있다. 따라서 GGS가 주는 시사점을 살펴보는 것은 의미가 있다고 하겠다. 우선, 조사 방식에서 이전까지 횡단 조사방식을 취하다가 2004년부터 패널조사로 그 방식을 변경하였다는 점에서 한국의 인구패널 구축에 주는 시사점이 크다고 볼 수 있다. GGS는 종단조사이면서 3년 간격으로 동일한 사람을 면접하는 패널 디자인이 적용되어 있기 때문에 결혼 및 출산에 관련된 의향, 가치관, 태도가 실제 행동으로 실현되었는지를 확인할 수 있는 중요한 자료가 된다. 즉, 일회성의 횡단조사인 기존의 실태조사는 결혼과 출산을 포함한 생애주기를 반영한 생애의 이벤트를 추적할 수 없다는 한계가 있고, 인구의 변동을 파악하기 위한 핵심이 되는 결혼과 출산이라는 생애 사건을 추적하기 위해서는 인구패널이 필요하다는 것이다. 따라서 첫 번째 조사에서는 가구 조사를 포함하여 다양한 내용을 조사하고 두 번째 또는 세 번째 조사에서는 첫 번째 조사에서 질문한 출산 관련 의향과 계획이 어떻게 이행되었는지에 초점을 두고 조사하여 조사 내용을 축소하는 방식으로 진행할 수 있다.

또한, '전국 출산력 및 가족보건·복지 실태조사'에 관해 조사 명칭에 있는 '출산력'이라는 용어에 대한 문제와 여성의 전통적 성역할에 관한 문항 등 조사 내용에 대한 문제가 제기된 바 있다. 이는 조사에 젠더의식을 반영할 필요가 있다는 의미라고 볼 수 있다. 이러한 문제의식에 대해 GGS가 주는 시사점이 있다. GGS의 경우 1990년대에 '출산력과 가족 조사(Families and Fertility Survey)'에서 '세대와 성 조사(Generations and Gender Survey)'로 명칭을 변경한 바가 있다. 결혼과 출산이라는 좁은 영역에서 성(gender)을 넓게 포괄하는 내용을 담았고 생애주기 전

반과 가족 관련 내용으로 확대하였으며 조사 대상자에 여성뿐만 아니라 남성도 포함하고, 연령에서도 가임기를 반영한 49세에서 79세로 확장하였다. GGS에서도 성역할(gender role)을 이해하는 것은 출산 결정에서 중요하다고 보아 관련 가치관 등에 관한 문항이 존재한다. 이는 인구패널의 구축에도 젠더의식과 관련된 내용을 조사에 포함할지, 조사 대상과 범위를 설정할 때 이를 반영할지를 판단할 필요가 있음을 시사한다.

추가적으로 GGS의 경우와 같이 시계열 유지를 위해 가능한 한 문항을 유지하여 지표(updated list of indicators)를 지속적으로 생산하되, 간단하게 하는 것(simplified list of indicators)을 원칙으로 조사 대상자가 대답할 만한(tolerable) 수준의 문항을 개발할 필요가 있다. GGS의 경우 많은 국가들이 참여하고 있고, 이에 따라 국제 비교가 가능한 만큼 인구패널의 한 형태 또는 인구패널에 부가적인 자료로서 우리나라도 이와 유사한 조사를 수행하는 것을 고려해 볼 수가 있다.



# 제 4 장

## ‘인구패널’의 주요 요건

제1절 인구 관련 패널의 목표와 기본 성격

제2절 인구 변동 패널화 관련 통계학적 고려 사항



# 4

## ‘인구패널’의 주요 요건 <<

### 제1절 인구 관련 패널의 목표와 기본 성격

#### 1. ‘인구패널’의 목표와 역할

인구패널 또는 인구학적 패널을 새로 구축하는 것이 타당하려면 적어도 (1) 어떤 목적에서 무엇에 쓰고자 구축하는지가 분명해야 하고, (2) 이 자료로 무엇을 할 수 있는지, 특히 패널이 아닌 자료로는 할 수 없는 무엇이 가능한지가 뚜렷하게 드러나야 한다.

기존 논의에서 공통되는 요지만 간추려 본다면, 인구패널의 목적은 “인구 변동의 복합적인 원인과 영향을 정확하게 진단하고, 이를 기초로 효율적인 대응정책을 수립하기 위한 인구 동학적인 자료(의 제공)”(이삼식 외, 2006, p. 106)라 할 수 있다. 이러한 목적을 위해 인구패널에 기대할 수 있는 구체적인 역할, 즉 인구패널 활용의 실천적 목표는 대체로 다음과 같이 정리된다.

- 1) 인구 관련 지표 산출과 기초통계량(또는 모수) 추정의 보조 자료
- 2) 실증분석과 통계적 추론(추계, 검정, 의사결정)을 위한 기초 자료
- 3) 정책 기획, 평가, 환류를 위한 실증(정책실험, 정책 평가와 검증)의 근거 자료

하지만 인구패널에 기대하는 이러한 목적과 역할이 기존의 인구 관련

자료와 크게 다른 것은 아니다. 인구 변동의 원인과 영향을 진단하기 위해 구축한 기초자료라는 점은 인구주택총조사나 출산력조사 같은 기존의 인구 관련 주요 횡단면 조사도 마찬가지다. 정책적 활용(정책 발굴과 기획, 평가와 환류)이 핵심이라는 점은 대표적 인구 관련 횡단면 자료들의 1차적 목적과 크게 다르지 않다. 예를 들어, 출산력조사는 명시적으로 ‘인구정책의 수립과 평가를 위한 기초자료의 제공’이 주된 목표이다(이소영, 김은정, 박종서, 변수정, 오미애, 이상림, 이지혜, 2018, pp. 13-15).

인구패널의 차별성에 대해 지금까지 강조되어 온 일반적인 논거는 피상적이고 소극적인 수준에 머물러 있다. 기존의 횡단적 인구 관련 자료는 시점별 인구 현황의 정태적 포착(snapshot)을 넘어서 (급격한) 인구 변동의 사회·경제적 원인, 가치관·문화·생활환경 같은 환경요인의 영향을 밝히는 데 부족하고 한계가 있음이 주로 지적되어 왔다. 이들 횡단면 자료로는 인구 현상과 사회현상 간의 복합적인 구조와 관계를 파악하고 인구 변동을 동태적으로 분석하는 데에 한계가 있다는 것이 주된 내용이었다(이삼식 외, 2006, pp. 102-104).

반면에 인구패널의 필요성에 대한 적극적인 논거는 의외로 뚜렷한 주장을 찾아보기 힘든데, 대개 패널의 종단 또는 시계열 측면이 동태적 분석을 가능하게 해 준다거나 패널 자료를 이용하면 횡단과 종단을 아우른 통합적 회귀분석(pooled regression analysis)이 가능해진다는 정도가 제시된다.

하지만 “인구 변동의 복합적인 원인과 영향을 정확하게 진단하고... 인구동학적(으로)...분석”(이삼식 외, 2006, p. 106)하는 데 인구패널이 아니면 안 되는 이유, 인구패널의 차별성과 비교우위가 어디에 있는지에 대한 구체적 설명은 찾아보기 힘들다. 인구패널의 한계와 단점까지 감안하

여 그럼에도 구축이 필요한지 판단하는 과정도 생략되어 왔다. 인구패널에 대한 회의적 의견이 존재하고 이 의견이 적지 않게 공감을 사는 이유의 상당 부분이 여기에 있다. 새로운 대규모 패널 자료의 구축 대신, 기존의 인구주택총조사, 인구동향조사 같은 인구 기초자료에다 주제와 연구 목적에 따라 복지패널, 여성·가족패널, 재정패널 등의 인구 관련 자료를 보조적으로 활용하면 자료 구축의 목적을 달성하는 데 충분하다는 반론이 가능한 지점 역시 여기이다.<sup>31)</sup>

인구 변동은 기본적으로 계열상관성(serial correlation)이 뚜렷한 시계열 현상이다. 출생, 사망, 인구 이동은 시차와 강한 자기상관성(auto-correlation)이 두드러진 사회현상이고 그 바탕에 있는 모집단(population)의 구성과 확률적 분포 특성 역시 전체 인구(population)의 진화 과정을 요약하는 시계열(time series) 특성이다. 출생, 사망 기본 통계는 계열상관성이 뚜렷한 시계열 변수(time series)들이고, 이들을 가공해서 얻어지는 인구 기초율(혹은 모수)인 출산율, 사망률, 인구구성비, 부양비, 출산진도비 등은 경제성장률, 실업률, 물가상승률과 마찬가지로 거시-집계변수 시계열(aggregate macro-time series)의 전형적 형태이다.

이들 인구 기본 통계와 기초율 변수들을 생산하기 위한 기반 자료가 인구동향조사나 인구주택총조사와 같은 횡단면(cross-sectional) 자료라서 문제가 있는 것은 아니다. 출산력조사가 더 이상 합계출산율을 추정하기 위한 핵심 자료 역할을 하지 않는 것은 전수조사인 인구주택총조사가

31) 패널 자료에 대한 회의론 중 하나는 5년 주기의 인구주택총조사보다 더 짧은 주기의 패널조사가 인구학적으로 의미가 있는지에 대한 의문이다. 이는 인구 변동의 핵심은 장기적 움직임과 추세이고 단기 현상은 미미하거나 큰 의미가 없다는 인식에 따른 것이다. 그럼에도 인구패널의 필요성에 대한 이전의 논의 가운데에는 기존 횡단면 자료의 조사 간격이 장기적임을 한계로 지적하는 경우도 있다(이삼식 외, 2006, p. 103, <표 5-1>).

있기 때문이지 출산력조사가 패널이 아니기 때문은 아니다. 제2장에서 살펴본 바와 같이 인구 기초 통계나 기초율을 생산하고 추정하는 데에 패널 자료가 불가결한 것도, 최적의 자료 형태도 아니다.

거시-집계변수로서 인구 기초지표나 인구 관련 변수들의 시계열적 특성을 포착하고 분석하고자 할 때 패널 자료 형식이 불가피한 것도 아니다. 예를 들어, 출산율과 실업률(혹은 고용률), 경제성장률 간의 관계를 장·단기에 걸쳐 시계열적으로 분석해 보고자 한다면 일반적인 거시 시계열 분석 모형과 방법론으로 가능하다. 이들 거시-집계변수를 생산하고 분석하는 데에 패널 자료가 필요한 것은 아니다.<sup>32)</sup>

인구 변동을 집계 수준(aggregate level)에서 분석하고자 할 경우에도 패널 자료의 필요성은 불분명하다. 기간 간 출산율 증감을 인구학적 요인 분해 방법으로 설명한다고 할 때 개인이나 가구별 개체 특성(individual heteroskedasticity)이 개입할 여지는 별로 없다. 최근의 저출산 추이를 설명하는 요인으로 인구모멘텀 효과와 결혼 행태, 출산 행태 가운데 어느 것이 더 중요한지를 살펴본다면 개인 특성은 '일반화된 개인(representative agent)'의 행태로만 반영해도 충분하다. 경제활동인구의 연령 구조와 분포가 고령화되는 현상과 노동생산성의 변화와의 관계를 분석하고자 한다면 개인 수준의 변화보다는 집계 수준에서의 평균적 변화가 초점이 되는 것이 일반적이다. 이상의 예들에서 패널 자료가 필수적이지는 않다.

하지만 인구 변동을 설명하고 이해하려는 노력과 시도는 점차 인구 현상의 미시적, 구조적(structural) 측면으로 초점이 옮겨가고 있다(제2장

32) 거시경제학에서 가장 흔한 예로 물가(인플레이션율)와 실업(실업률)과 경제성장률의 관계에 대한 실증분석을 수행할 때, 패널 자료가 요구되지는 않는다.

제1절, p. 26, 각주 8 참조). 최근 우리가 경험하고 있는 급격한 저출산·고령화는 집계 수준에서 인구 변동 요인을 분석하는 데에서 더 나아가 인구 현상을 이루는 개별적, 개인적 요소들의 움직임과 경향을 집계 이전 과정에서부터 파악할 것을 요구하는 단계에 이르렀다. 저출산·고령화 현상의 심화에 정책적 대응이 시급한 상황에서 대부분의 저출산 대책, 고령화 대책은 대상과 정책 수단 모두에서 개인과 가족 단위로 기획, 집행되고 그 성과가 평가된다. 정책적 능력 또한 정책이 구현되는 단위와 정책을 구사하는 도구에서 개인·가족 수준까지 식별할 정도로 세밀하고 미시적인 수준으로 진화했다.

인구 변동은 직접적인 인구 변동 요인(출생, 사망, 인구 이동)에 따른 분해, 분석만으로 충분치 않을 만큼 복잡하고 종합적인 사회현상이다.<sup>33)</sup> 인구 변동과 인구구조 변화를 이끄는 직접적 변동 요인의 바탕에는 사회·경제·문화적인 환경요인이 존재하고, 결혼·가족·주거·노동 등 주요 사회제도와 체제가 이들을 연결하는 매개요인으로 작용한다. 인구 변동의 기제를 온전히 이해하고 적절히 대처하기 위해서는 이들 직접요인(인구학적 변수), 환경요인(사회·경제적 변수), 매개 요인(제도·체제 변수)들이 상호 원인과 결과로 작용하는 과정을 복합적으로 고려하여 이들 간의 관계를 포착해 내는 것이 불가피하다.

이를 지향하는 경험연구에서 인구 변동과 그 요인은 인구 현상의 최소 단위인 개인과 가족 수준에서 관측, 파악되어야 한다. 개인과 가족이 처한 사회경제적 조건과 문화적 환경, 개인의 선호와 가치관, 가족 단위의 의사결정과 행태가 관측되어야 개별적 특성이 개별적 인구 현상을 낳는 미시적 경로를 식별할 수 있고, 어떤 집계 과정을 거쳐 어떻게 사회적, 거

33) 이삼식, 최효진, 계봉오, 김경근, 김동식, 서문희, 윤자영, 이상협, 이윤석, 천현숙 (2016); 제1, 3차 저출산·고령사회기본계획(대한민국정부, 2005, 2015) 참조.

시적 인구 변동으로 나타나는지를 파악할 수 있다.

개인의 교육 수준 또는 학력 향상이 출산력에 미치는 영향을 집계 수준에서 살펴보는 것은 평균적인 수준 간의 관계에 주목하는 것과 같다. 여기서 더 깊이 들어가서 개인의 학력 신장이 사회·경제적 배경 요인의 변화와 어떤 관계를 이루고 있는지, 그로 인해 출산력에 연계되는 결혼·주거·근로 등의 매개요인들 대한 개인의 선택이 어떻게 바뀌는지 식별하려면 개인적, 미시적 수준의 관측이 필요하다.

인구 고령화에 따른 노인 돌봄·의료·요양 서비스에 대한 수요-공급 관계를 전망하고자 할 때 ‘고령인구 증가 → 노인빈곤 심화, 고령자 돌봄·의료·요양 수요 증가 → 공공 서비스 부담 증가’라는 평균적, 추세적 도식은 자명하면서도 불충분하다. 집계 이전에 개인과 가족 수준에서 노인 경제력, 노인 건강과 의료비 지출, 부양과 사적 이전 행태 등 개인적, 미시적 특성과 조건으로 포착되는 관계나 조정 과정이 유효한 정보임에도 거시-집계 수준에서는 평균으로 상쇄되어 파악이 불가하기 때문이다.

변수들 간의 인과관계를 좀 더 명확히 밝히고자 하는 노력도 미시적, 구조적 성격의 자료에 대한 요구를 강화하고 있다. 거시-집계 수준에서 파악하는 인과관계는 거시시계열 특성에 의존한 시차상관관계나 확률적 인과관계<sup>34)</sup>와 같은 평균적, 결과적 수준에 머물러 실질적으로 의미 있는 학술적, 정책적 함의까지는 이끌어 내지 못하기 쉽다. 여성의 경제활동 참여와 출산력의 관계에 대한 경험적 연구가 흔히 불분명하고 때로는 상반되기까지 하는 결과를 낳는 것은 거시-집계 변수들 간의 관계를 위주로 분석했기 때문일 수 있다.<sup>35)</sup> 혹은 여성의 경제활동 참여가 출산력에 긍정

34) 대표적인 예로는 Granger-causality가 있다. Stock and Watson(2003, pp. 448-450); Hamilton(1994, pp. 302-309) 참조.

35) 분석 자료의 관측 기간이 이른바 ‘여성의 사회적 지위를 기준으로 본 사회 발전 단

적이거나 부정적인 요소를 모두 가지고 있어 실제 자료상에 어느 쪽이 더 강하게 발현되었는지 여부에 따라 결과가 달라지거나, 이들 변수들이 서로 영향을 주고받아 인과관계가 한 방향으로 성립하지 않는 것이 원인일 수도 있다. 이때 여성의 사회경제적 지위와 연관된 사회·문화 환경요인이나 이들 변수를 매개하는 가족제도 등의 변천을 개인적, 미시적 수준에서 추적하는 것이 드러나지 않은 인과관계를 식별하는 데 도움이 될 수 있다.

또한, 인구 변동은 사회현상의 일부이기 때문에 인구 변동과 사회·경제·문화적 배경과의 관계, 또는 이 둘을 매개하는 사회 제도와 규범들(결혼, 가족, 세대 이전 등)과의 관계에서 기본적으로 구조적(structural), 내생적(endogenous)이다. 특정 인구 현상의 원인이나 그로 인한 인구 변동의 파급효과는 단선적, 일방적이기보다는 주목하고자 하는 관계나 국면, 관측의 기간[=시계(視界)]과 시차 등에 따라 원인과 결과가 뒤바뀔 수도 있고 복잡하고 다양한 상호작용이 뒤섞여 인과관계를 특별히 정하는 것이 현실적으로 불가하거나 무의미할 수 있다. 이때 학술적, 정책적 함의를 지니는 직관적 인과관계를 도출하려는 경험연구는 변수의 내생성을 간과한 오류(endogeneity bias)를 범할 가능성이 크다.

여성의 교육 수준이 장기간에 걸쳐 향상된 것이 출산력의 저하를 인과적으로 얼마나 설명하는지 한계효과를 추정하려 한다면 여성 개인이 일정 학력 수준을 획득한 이후 출산을 완료할 때까지의 행태, 즉 (학력 수준 조건부) 완결출산력(completed fertility)을 보아야 한다. 학력 수준에 따른 여성의 계층별 출산력 차이를 살피고자 할 때도 마찬가지다(Choe and Retherford, 2009; Yoo, 2014; 우해봉, 장인수, 2018). 이때 출산력을 대리하는 변수로 합계출산율(period fertility rate)을 사용하는 것

---

계'(Esping-Anderson and Billari, 2015)의 어느 부분에 속하는지에 따라 다른 결론을 낳을 수 있다는 이론적 해석도 유력하다.

은 흔히 해석의 오류로 이어지기 쉽다. 합계출산율은 가상의 대표 가임여성 상정하는 기간출산력 지표이다. 따라서 합계출산율을 보게 되면 가임여성 인구집단의 전반적 교육 수준 변화 또는 교육 수준별 계층 분포의 변화가 출산 시점을 미루는 기간효과(tempo effect)의 변화를 설명하는 부분을 주로 포착하게 된다. 게다가 교육 수준 변화를 기간출산력 변화와 직접 연계하게 되면 학력 축적과 출산력 축적의 시차가 뒤섞이거나 후자가 전자에 영향을 미치는 내생적 관계를 배제하지 못해 인과관계의 식별이 불가능하게 된다.

완결출산력을 보는 경우에도 거시-집계 수준에서는 내생성의 문제에서 완전히 자유롭지 못하고 인과적 함의는 그만큼 근거가 약해질 수밖에 없다. 여성의 경제활동 참여와 출산력의 인과관계를 실증하고자 하는 경우가 좋은 예다. 횡단면 자료의 역사적 비교에 근거한 경험연구들에서 흔히 여성 경제활동 참여가 출산을 억제하는 효과가 과추정되는 경향은 어느 정도 알려진 사실인데 그 배경에는 내생성의 문제 그리고 개인 수준에서 두 현상을 연결하는 매개요인들 그리고 그 배경에 놓인 환경요인들과의 복합적인 관계가 식별되지 않음으로 인한 문제(unobserved heterogeneity)가 자리 잡고 있다. 이럴 때, 패널 자료의 미시적, 구조적 특성은 시계열적 단순 인과관계의 한계효과<sup>36)</sup>를 과추정하는 것을 보정해 주거나 적어도 인과관계에 관련된 변수들 간의 간접 효과를 식별하게 해 주는 역할을 할 수 있다.

인구 변동의 원인과 결과를 다양한 요인과 변수들 간의 복합적인 상호관계에 대한 실증<sup>37)</sup>을 통해 설명하고 이해하려면 미시-개인 수준에서 구

36) 혹은 축약형(reduced form) 인과관계의 한계효과.

37) 추정, 가설 검정, 추계와 통계적 의사결정 등 통계적 추론.

조적 관계를 관측, 추적, 분석할 수 있는 도구가 필요하다. 패널 자료는 이러한 관측과 추적이 가능한 형태로 구축된 자료이다(제2장 제1절, pp. 28-30 참조). 이를 바탕으로 한 실증분석을 뒷받침하는 방법론적 토대는 계량경제학에서 축적되어 온 패널 분석 이론에 의해 제공된다.<sup>38)</sup>

정책적 함의를 도출하고 정책 발굴, 사업 시행 평가, 정책 효과성 평가 및 환류에 직접 활용하고자 하는 목적의 경험연구는 흔히 방법론적 한계와 분석 결과의 신뢰성 문제들에 봉착하는데(Gauthier, 2007; 이상협, 이철희, 홍석철, 2016, pp. 6-7), 패널 자료의 활용과 패널 분석 방법론의 채용은 이를 완화하거나 해소하는 데에도 유효한 출발점이 될 수 있다.

## 2. '인구패널'의 필요충분조건

인구학적 패널은 인구 변동의 원인과 영향에 대한 복합적인 분석을 가능하게 해 주는 동시에 학술적, 정책적 가치가 분명한 해석과 함의 도출에 도움이 되는 것을 주목적으로 구축되어야 한다. 이러한 목적의 경험연구에서 인구학적 고려가 불충분하거나 패널 형식이 아닌 인구 관련 자료를 활용할 때 드러나는 한계와 문제점을 보면 인구학적 패널이 지향해야 할 목표는 더 확연해진다.

앞서 출산력 요인 분석의 예에서 보았듯이, 집계 이전에 개인과 가족 단위에서(식별되는 변수들 간의) 미시적 상호관계로 관측되는 인구 현상은 개념적으로 개인과 가족 단위에서 경험하는 완결인구력(completed

38) 선형 회귀모형을 기본 전제로 삼는 전통적 패널 분석 이론이 주를 이루는데 대표적인 문헌으로 Chamberlain(1984), Hsiao(2003), Arellano(2003) 등이 있다. 계산 능력의 획기적 신장, 수치해석 기법의 발달과 더불어 분석 모형은 변수들 간의 미시-구조적 관계를 더 현실적인 형태로 포착하도록 일반화되는데, 동태행위모형(dynamic programming)으로 대표되는 이들 동태구조 모형에서도 패널이 자료의 기본 형태가 된다. 제5장 제2절의 관련 논의 참조.

cohort demographics)이고, 한 시점에서 동시대의 전부를 평균적인 하나의 대상으로 간주하는 기간력(period demographics)과 구별된다. 교육 수준이나 소득계층이 출산력에 미치는 영향을 미시적으로 추적, 분석하려면 기본적으로 완결출산력을 보아야 한다는 뜻이다. 여기에 기간출산력 지표(합계출산율)를 대신 사용하면 어떤 오류를 범할 수 있는지는 이미 살펴본 바와 같다. 한편, 횡단면 자료는 기간력에 특화된 자료 형태인 반면 패널 자료는 완결력 분석에 적합한 형태이다. 이런 맥락에서 미시-개인 수준에서 인구 변동의 직접-환경-매개 요인들 간의 상호작용을 관측하고 해석하기 위해서는 패널 자료가 최적이다.

횡단면 성격의 인구 관련 자료를 바탕으로 삼는다고 완결력을 분석할 수 없는 것은 아니다. 같은 출산력 요인 분석 경험연구의 예에서 Choe and Retherford(2009)나 Yoo(2014)는 인구주택총조사를 바탕으로 추정된 완결출산율(cohort fertility rate)을 출산력의 주요 지표로 삼고 분석을 수행하였다. 한편, 여성의 학력 수준별 출산력 차이를 핵심 주제로 분석한 우해봉, 장인수(2018)의 경우 여성가족패널(2007~2014년)을 기반으로 산출한 완결출산율을 분석했는데, 산출 과정에서 출산진도비가 필요하고 이를 생명표 분석을 통해 추정한다는 점에서는 앞선 두 경우와 방법론적으로 동일하다. 인구학적 특성을 주안점으로 하지 않는 인접 패널 자료를 활용한 유사한 연구들에서 완결력의 추정에 비슷한 단계가 필요하다는 점은 주목할 필요가 있다.

Choe and Retherford(2009), Yoo(2014)와 우해봉, 장인수(2018)은 장기간 간격을 둔 코호트들의 출산율 차이를 교육 수준에 따른 계층의 분포 변화와 계층별 출산력 변화로 요인분해 하여 설명하는 것을 주요 분석 방법론으로 삼는다는 점에서도 유사하다. 특정 인구 변동 현상이 인구 구성(분포) 변화 때문인지 아니면 인구학적 기초율(행태) 변화가 주된 요

인인지 간명하게 파악할 수 있다는 점에서 이러한 인구학적 표준화와 요인 분해는 유용하다.

하지만 이렇게 단순화한 요인 이분화(dichotomy)는 다른 직간접 요인들이 기간 간 변화 없이 그대로 유지된다는 가정을 암묵적으로 전제하고 있어 비현실적이다. 분해 대상 요인과 독립적인 다른 요인들을 임의로 고정해서 이들이 간접적으로 출산력에 영향을 미치는 경로가 배제되기 때문이다. 이는 일반적인 비교정태(comparative static) 가정보다 더 강한 단순화를 전제한다. 또한 구성비나 기초율의 표준화는 현실에서 실현되지 않은 실험적 시나리오(counter-factual experiment)이고 실제로 이들이 변화하는 동태적 과정을 무시하고 있어 실제 현상과의 괴리가 클 수 있다.

학력 수준 또는 이로써 대표되는 사회경제적 지위가 여성의 (완결)출산력에 미치는 영향은 개인 수준에서 결정되고 행위로 나타나는 미시-개인적 인구 현상이다. 이를 거시-평균적 자료와 분석 방법으로 다루게 되면 자료 수집 단계에서 암묵적 평균화로 인해 중요한 많은 정보를 놓치게 된다. 패널 자료를 활용하면 인구 현상을 개인적 수준에서 추적하고 미시적 수준에서 분석하는 것이 자연스럽게 이루어진다. 완결력을 직접 추적할 수 있게 되고(아래 제4장 제2-1절 참조), 요인분해 방법보다 일반화된 준실험적 방법론(quasi-experimental design)들을 쉽게 적용할 수 있게 된다(제2장 제3절, pp. 36-38 참조). 패널 자료에 특화된 분석 도구들을 활용하여 횡단면 자료에서 미처 파악하지 못했던 '관측되지 않은 개인 특성(unobserved individual heteroskedasticity)'과 '누락된 변수에 기인한 선택 편의(selection bias)'에 대한 의식적, 적극적 고려가 가능해진다(제2장 제1절, pp. 26-27; 제2장 제3절, pp. 34-35 참조).

기존의 인구 기초 자료나 인구 관련 자료를 보완하는 것으로, 특히 기존의 횡단면 인구 자료를 패널화하는 것으로 패널 자료에 대한 수요를 대체할 수는 없는 것일까? 기존 자료를 활용한다고 하면 대표적인 인구 기초 자료인 인구주택총조사나 인구동향조사, 국내 인구가동통계 또는 대표적 인구 관련 횡단면 조사인 출산력조사, 노인실태조사, 가족실태조사 등을 패널 자료화하는 것을 생각해 볼 수 있다.

이중 가장 유력하고 활발하게 논의되어 온 것이 통계청을 중심으로 진행되어 온 인구패널이다. 전수조사인 인구주택총조사 또는 부속 대규모 표본조사를 패널화하는 것을 목표로 타당성이 검토되어 왔는데, 본격적인 계기는 2015년 인구주택총조사부터 도입된 인구등록센서스의 확대이다. 인구주택총조사의 기본적인 전수 항목을 가용한 행정자료의 표준화와 연계를 통해 수집하고 행정자료로 파악이 곤란한 항목에 대해서 심층 항목 현장조사(2015년의 경우 20% 가구에 대한 표본조사)를 병행하는 방식이 주로 고려되고 있다. 총조사와 행정자료의 연계는 자연스럽게 행정패널 성격의 전수형 인구패널을 상정하게 한다.

통계청 인구패널은 개념적으로 통계청에서 실시하는 격년제 사회조사와 인구주택총조사 표본조사의 자료를 연계하려는 노력과 논의에도 뿌리가 닿아 있다. 통계청의 사회조사는 가구조사로서 총 10개 부문(보건, 교육, 안전, 가족, 환경, 복지, 사회참여, 문화·여가, 소득·소비, 노동)을 5개 부문씩 격년으로 실시하고 있는데, 이를 다양한 관련 조사 자료와 더불어) 시계열적으로 연계하고자 하는 시도가 2015년 인구주택총조사를 전후해 등장했다(정미옥, 최필근, 2014; 정미옥, 2015). 사회조사가 인구주택총조사의 표본조사구에서 표본을 추출하는 관계로, 여기에서의 주목 대상은 인구주택총조사와의 연계와 활용성이다(정미옥, 백지선, 김현식, 최은영, 2016, p. 1).

인구주택총조사를 바탕으로 한 인구패널에 대한 논의는 2020년 인구주택총조사 실시와 함께 본격적으로 구체화할 것이다. 하지만 그것이 행정자료와 연계된 전수조사형 행정패널을 지향할 것인지, 아니면 사회조사와 20% 심층항목 표본조사 자료의 연계 맥락에서 기존 사회조사의 인구 관련 부분에 모자란 항목들을 보강한 인구학적 조사 시계열화를 염두에 둘 것인지는 아직 불확실하다. 인구패널의 규모, 관측 주기, 조사 세부항목과 내용에 대해서도 방향이 정해진 바는 없는 것으로 보인다.

이러한 횡단 인구자료의 시계열화는 패널의 외형적 요건을 갖춘 자료를 생성하는 한 방편이 된다. 기준을 뛰어넘는 규모, 기존 자료(행정자료 또는 기존 사회조사 자료) 연계와 복합을 통한 축적·유지 부담 감소 등에 상되는 이점만으로도 기대와 주목을 받고 있기도 하다. 하지만 이러한 방식으로 구축된 패널이 자료의 성격과 활용 가치 측면에서 타당성을 확보할지는 미지수다. 인구 변동의 원인과 영향에 대한 미시-구조적 분석과 그로부터의 실질적 함의 도출이라는 목적에 행정패널이 적합한 형태인지는 불확실하다. 당장 가용한 행정자료로 노동, 소득과 소비, 가족력 등과 같은 주요 인구학적 매개요인, 환경요인들을 정확하게 파악하기는 쉽지 않다. 일부 핵심 인구지표들(예: 개인별 혼인력, 출산 진도, 동거 상태 등)의 경우 행정자료(주로 주민등록통계)의 정확도와 신뢰성이 의문시된다.

사회조사를 인구학적 목적에 따라 개편하고 인구주택총조사 표본조사와 연계해 시계열화한 것이 더 합목적적인 패널 자료에 가깝게 될 것이다. 하지만 일종의 순환패널(rotating panel) 형태를 띠는 시계열 연계 자료가 앞서 논의한 인구학적 인과관계의 미시-구조 분석에 충분한 기능과 역할을 보장할지는 분명치 않다. 자료의 규모와 범위를 어떻게 정하는지에 따라 기존 사회조사의 인구학적 재편과 보강, 시계열 연계 방법에 필요한 작업 부담이 인구학적 패널 자료를 새로 구축하는 것과 크게

다를 바가 없을 수도 있다.

인구패널의 새로운 구축에 대해 근본적으로 회의적인 입장도 짚어 볼 필요가 있다. 이 관점의 요지는 ‘인구 변동의 전 과정을 관측하고 통찰할 수 있는 인구 자료로서 출생에서 사망까지 전 생애주기에 걸쳐 인구 현상과 관계를 이루는 다양한 분야의 변수와 지표들이 인구패널 안에 포괄되어야 할 텐데 현실성도 없고 비용 대비 실익도 없다. 차라리 분석 주제와 대상에 따라 필요한 인구 관련 자료들을 적절하게 재가공해서 사용하는 것이 현실적’이라는 인식이다. 이러한 비판은 인구패널 구축에 관한 논의의 거의 무력화할 만큼 강력하고 유효한데, 그 바탕에는 인구패널의 구성과 포괄 범위에 대한 당위론적 요구가 암묵적으로 전제되어 있다. 인구패널이라 부르려면 모름지기 (1) 그 틀 안에 출생-사망-인구 이동에 관련된 인구학적 결정 요인들이 담겨야 하고, (2) 전 연령대와 세대를 통틀어 각각의 개인(또는 가구)을 대상으로 인구 관련 주요 환경요인과 매개요인들이 변수(항목)로 식별, 관측, 수집되어야 한다는 생각이 그것이다. 인구패널에 대한 초기 논의에서도 이러한 암묵적 인식과 전제를 엿볼 수 있고, 그 때문에 초래되는 실질적 문제들로 고민한 흔적이 자료 구축 단계 여러 곳에서 발견된다. 인구패널의 개념적 논거에 집중했던 초기(이삼식 외, 2006, pp. 109-136)에는 인식하지 못했지만 패널조사 대상, 주제, 내용과 범위, 기간 등을 설정하는 작업은 후속 사업(보건복지가족부, 한국보건사회연구원, 2008)에서 인구패널의 이상적 요건과 부딪치는 실질적 문제로 부각된다. 보고서는 조사 대상을 가구, 청소년, 미혼자, 기혼자, 노인으로 구분하는 방식을 제안한다. 그리고 각 대상에 따라 기본조사 주제(모듈)와 조사영역이 별개로 설정되고 조사표, 조사 규모 등이 따로 선정되어 실질적으로 다섯 개의 독립적인 조사가 설계된다. 결과적으로 여러

개의 패널조사를 제안하고 이름으로만 인구패널로 묶어 놓은 셈이 되는데, 이중 임신·출산·결혼에 관련된 부분은 출산력조사를 확장한 형태로 집중적으로 논의되는 반면 고령세대를 대상으로 한 부분은 기존의 패널 자료(노후보장패널)를 차용할 것을 제안하는 등 혼란스러운 모습을 띤다.<sup>39)</sup>

인구학적 실증 연구와 분석을 위한 자료로서의 활용 가치가 궁극적 목적이라면 '인구패널'이라는 외연에 집착할 이유는 없다. 현실적으로 보아도 자료의 미시적 성격상 규모, 대상, 주제의 범위를 설정하는 데에 운용·유지나 비용 대비 편익을 고려한 적절한 균형을 찾는 것이 불가피하다.

인구 변동에 관련한 모든 핵심 변수들과 주요 관계를 하나의 패널에 담는 것은 가능하지도 바람직하지도 않다. 인구 변동을 구성하는 출생-사망-인구 이동은 생애주기에 걸쳐 고르게 경험하기보다는 특정 연령대와 세대에 집중된다. 이들 각각을 결정하는 배경요인들과 비중은 같지 않고, 각각 영향을 미치는 사회경제 현상과 분야가 일치하지 않는다. 출산과 가족 형성이 대체로 개인의 자발적 선택과 선호의 결과라면, 늙고 죽는 현상은 수동적으로 경험하고 받아들이는 과정에 가깝다. 개인 차원에서 인구 이동은 미리 예상하거나 계획되기보다는 주변 환경이나 여건 변화에 대응하는 행태의 경향이 강한 비교적 단기 시계의 인구 현상인 반면, 결혼과 출산, 가족의 형성과 성장은 더 긴 시간을 거쳐 실현되고 지속되는 현상이고, 고령화와 사망은 그 이상의 점진적인 과정을 거치는 장기 시계의 현상이다.

따라서 인구학적 패널을 구축하고자 한다면, 주 관심 대상이 어떤 인구

39) 보건복지가족부, 한국보건사회연구원(2008), pp. 60-61, pp. 81-83 참조. 인구패널의 이상적 외형에 치중하다 보니 결과적으로는 패널 구축의 목적에 대한 인식이 흐려지고 전반적인 체계의 일관성이 흔들리는 모습(pp. 65-66)도 관찰된다.

현상인가에 따라 적합한 내용과 범위, 관측 기간과 주기가 달라지는 점을 반영하는 것이 합리적이다. 관심 주제별로 주요 대상 인구집단이 달라지므로 자료 관측의 단위, 수집 규모에서부터 해당 인구 현상과 주제와의 정합성이 고려되어야 한다.

그렇다면 인구학적 패널의 형태는 관심 대상 인구 주제별로 또는 분석하고자 하는 인구 현상에 따라 별개로 구축된 인구 주제별 패널이 논리적으로 자연스럽다. 인구 변동의 어떤 측면에 주목하는가에 따라 연관된 사회·경제·문화 현상과 배경의 내용과 범위가 정해지고, 이들 간의 상호작용과 관계 가운데 핵심이 되는 부분이 어떤 시계에서 얼마만큼의 시차를 두고 관측되어야 유효한지가 자료 수집의 주기와 기간 선택에 반영되는 것이 합리적이다.<sup>40)</sup> 결혼과 출산, 출산력에 결부된 인구학적 요인과 사회경제적 요인들 간의 인과관계를 규명하려는 목적의 패널 자료라면 가임 여성을 주 관측 대상(이자 관측 단위)으로 삼고 이들의 결혼·임신·출산 결정과 행태를 설명하는 요인들의 관측에 주목하는 것이 마땅하다. 가족의 형성과 성장 및 분화를 인구학적으로 살피는 목적의 패널이라면 결혼과 출산은 목적변수가 아닌 설명변수 역할을 하도록 자료의 구성에 반영되는 것이 맞는다.<sup>41)</sup>

인구학적 주제 또는 핵심 인구 현상을 중심으로 패널을 구축하려면 핵심 주제와 현상을 어떻게 분류하고 설정할 것인지에 대한 근거와 기준을 마련하는 것이 관건이다. 가장 직관적인 기준은 인구 변동의 직접적 요인

40) 이삼식 외(2006)에서도 이러한 측면에 대한 고민이 드러나는데, 인구패널 조사 주제의 내용과 범위를 설정하고 '기본조사주제와 부가조사주제로 모듈화'하는 논의(pp. 109-113, pp. 114-116)가 그 예이다. 다만, 이를 패널 설계에서 소화하는 방식은 다분히 소극적이고 기계적이다.

41) 인구 주제별 자료 구축이라는 관점에서 본다면 출산력조사는 자료의 목적의식에서 출산력과 가족력 간의 구분이나 우선순위가 모호한 점이 문제로 적시된다.

을 중심으로 출생력, 사망력, 인구 이동력 패널로 구분해서 구축하는 것이다. 각 인구 변동 요인의 주체가 되는 연령, 세대를 주 관측 대상으로 삼고 이들의 인구학적 행태와 밀접하게 연관된 사회·경제·문화적 배경요인들을 중심으로 인과관계를 미시적, 구조적으로 추적하는 것이다.

[그림 4-1] 주제별 인구패널의 직관적 분류

	출생	사망	인구 이동
영유아·아동			
청소년			
청년	출산력 패널 가임기 여성		진학·취업 이동력 패널
중·장년			은퇴·분가
고령자		고령화 패널	

자료: 저자 작성.

출생과 직접 연관된 결혼·임신·출산의 주 행동 주체는 청년, 특히 핵심 연령대 가임기 여성이고, 이들의 선택과 행동에 연결되는 주요 배경요인은 통용되어 온 기본 개념(이삼식 외, 2006, p. 114, [그림 5-3]; 제3차 저출산·고령사회기본계획(대한민국정부, 2015), p. 37, p. 41; 이삼식 외, 2016, pp. 287-288, [그림 11-1], [그림 11-2])을 준용해 교육, 취업과 노동, 결혼·임신·출산, 보육·양육, 환경(주거, 일·생활균형, 가치관,

보건) 등의 주제로 구성할 수 있다.

사망력 패널은 고령자 중심 또는 중장년 이후에 초점을 맞추어 노후소득보장, 노후건강·돌봄, 고령자 사회참여, 고령자 생활환경 등의 주제에 맞게 변수와 항목을 구성할 수 있을 것이다(이삼식 외, 2006, p. 114, [그림 5-3]; 제3차 저출산·고령사회기본계획, p. 37, p. 41).

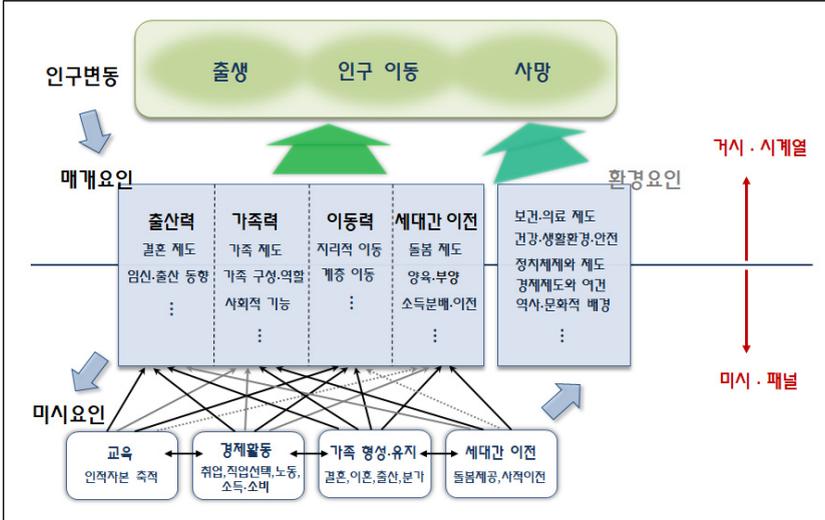
다만, 인구 변동을 출생-사망-인구 이동으로 대별할 때는 거시-집계 수준에서의 인구 변동 요소의 개념들이 되는 반면, 위와 같은 구성의 출생력, 사망력 패널 자료에서의 출생과 사망은 개인적 수준에서 관측되는 미시 현상인 점은 문제이다. 이 둘을 매개하는 중간 과정이 없이 직접 인과관계로 묶는 통상적인 도식화가 출생력, 사망력 분석과 해석에 유효한 접근 방식인지 의문이기 때문이다.<sup>42)</sup>

더 유력한 구성은 개인(과 가구) 수준에서 출생력-사망력-이동력과 직접적으로 연계된 주요 환경요인과 매개요인의 주제에 따라 패널을 분류하여 설계하는 것이다. 개인의 선택과 행동이 미시적 인구 행태로 나타난 연후에 출생-사망-인구 이동의 거시적 인구 변동 요소로 집계되는 과정은 주요한 사회 제도, 사회 행태나 관습 기제를 매개(매개요인)로 이루어 지는데, 이들을 출산력과 출산 동향, 가족력과 가족 제도, 세대 이전 및 계층 이동 기제 등의 인구적 범주와 그 외의 사회적·자연적 여건과 한계를 구성하는 환경적 범주(보건의료 제도와 건강, 생활 환경과 안전, 정치 체제와 제도, 경제 제도와 여건, 역사·문화적 배경과 가치관 등)로 분류할 수 있다.<sup>43)</sup>

42) 이삼식 외(2016)와 제3차 저출산·고령사회기본계획(2015)의 핵심 이론 틀인 이른바 ‘거시-미시 연계 패러다임’은 자료 구축 단계에서의 이러한 통계적 이슈까지 감안하지는 못한 것으로 보인다.

43) 미시적 인구 현상이 거시 인구 변동 요인으로 집계·전환되는 매개 경로를 이렇게 직접적 성격의 요소(인구적 범주)와 간접적 요인(환경적 범주)으로 나누고, 인구적 범주를 출산력, 가족력, 이동력 등으로 세분하는 것이 절대적 형태는 아니다. 개념적으로 이러한

[그림 4-2] 인구 변동의 개념적 분해: 직접-매개-환경 요인과 거시-미시 위계



자료: 저자 작성.

개인과 가족 수준의 미시 인구 행태는 출산력, 가족제도, 사회적 계층 이동, 세대 간 이전 메커니즘 등 인구적 범주의 매개변수들을 결정하는 것으로 이해할 수 있고, 그 과정을 설명하는 (인과)관계를 추적하고 식별해서 분석하는 것이 패널 자료의 목적이라고 볼 수 있다. 이렇게 나누어 보면, 출생-사망-인구 이동의 거시적 추이는 이들 매개 범주들 가운데 관련된 요소들의 조합과 상호작용을 통해 발현되는 거시-집계 수준의 현상으로 따로 다루어 분석하고, 미시-개인 수준에서 인과적으로 설명하려고 시도하지 않는 것이 분석적으로 더 유효하고 개념적으로 더 간명하다.

매개 요인들은 개인(가족)의 관점에서 보면 의지대로 선택하고 바꿀 수 없는, 주어진 여건 또는 상태(state variables)라 할 수 있다. 동시에 이들은 개인의 미시적 인구 행태 여러 유형들이 모여 만들어지는 사회적, 평

틀에 부합하는 분류와 구성도 경직적이거나 고정적일 이유는 없다.

균적 인구 행태와 동향이다. 반면에, 개인의 인구 행태(결혼, 출산, 가족 형성, 지리적 이동 등)는 개인의 선호와 선택에 의해 결정되고 행동으로 나타나는 미시적 사회경제활동의 결과들(choice variables)이다. 이들은 인구학적으로 중요한 주제 영역들로 개념적으로 분류하고 변수화할 수 있는데, 한 가지 방안은 [그림 4-3]과 같이 ‘교육(학력)과 인적자본’ 축적, 취업과 직업 선택, 노동, (근로·자산)소득과 소비를 아우르는 ‘경제활동’, 결혼과 이혼, 출산, 분가 등에 기인하는 ‘가족의 형성과 유지’, 돌봄이나 부양 제공, 사적 소득이전이 주가 되는 ‘세대 간 이전’ 등으로 구분하는 것이다.<sup>44)</sup>

[그림 4-3] 주제별 인구학적 패널의 예: 출산 관련, 가족 관련, 고령화 관련

	교육과 인적자본	경제활동	가족형성 및 유지	세대간이전	기타
영유아·아동					
청소년					
청년	출산 관련 주요 영역			가족력 관련 주요 영역	
중·장년					
고령자				고령화 관련 주요 영역	

자료: 저자 작성.

44) 이 분류 또한 절대적 기준일 수 없다. 예를 들어 개인의 사회경제적 계층 이동, 가구의 지리적 이동, 건강·의료서비스 소비 등을 별개의 영역으로 따로 떼어 내 볼 수도 있다.

인구 매개 요인들의 범주 구분에 따라 구축한 패널은 각 주제별로 인구 행태 변수와 사회경제적 행태 변수들 간의 (인과)관계를 미시-개인 수준에서 설명하는 것이 주목적이 된다. 그리고 이로부터 얻어지는 특정 (인과)관계가 구조적으로 식별되고 다수의 또는 대표적 개인을 통해 사회적 인구 행태를 설명하는 하나의 현상으로 유형화하는 과정을 보이는 것이 이 자료를 이용한 분석의 궁극적 목표가 된다. 주제별 패널은 자료 구축의 목적과 활용 가치에 부합하는 자연스러운 형태라고 볼 수 있는데, 핵심 주제의 개념과 범주 설정에 따라 출산력 또는 출산 관련 패널, 가족력 또는 가족 형성·성장·분화 관련 패널, 세대 이전과 고령화 관련 패널, 사회적·지리적 인구 이동과 계층 변화 관련 패널 등으로 묶는 것이 한 예가 될 수 있다(제2장 pp. 28, 36 참조).

주요 주제별로 인구학적 패널을 따로 구축하는 방식에 대해서는 몇 가지 우려가 제기될 수 있는데 이러한 새로운 접근이 무의미하게 될 정도로 큰 문제는 아니다. 우선 제기되는 비판은 대상 인구 집단과 인구 현상을 기준으로 임의로 나눈 분절적 자료라는 것이다. 패널별로 해당 주제와 영역에만 집중하여 관측·수집되다 보니 자료에 포함되지 않은 인구 집단, 자료의 범위를 넘어서는 변수와 주제 영역이 관련된 인과관계를 놓칠 가능성이 있다. 자칫 특정 인구 집단에만 해당되는 자료로 제한되면 이를 기반으로 한 경험연구의 결과는 전체 인구 변동을 설명하는 하나의 유형으로 일반화하기 어려울 수도 있다.

위에서 개념화한 주제별 패널에서 이런 문제들이 심각한 것은 아니다. [그림 4-2]와 [그림 4-3]의 개념적 근거는 인구 집단이나 인구 현상의 배타적, 수평적 분할과는 무관하고, 주제별로 특정 인구 집단의 배제 관련 변수 영역의 제한을 전제로 하고 있지도 않다. 예를 들어 출산력 패널을

구축하는 데 49세 이상의 인구 집단이나 남성을 자료의 관측 대상에서 제외한다고 해서 이들이 가족 내에서 혹은 부부 간 관계를 매개로 출산력에 관련된 결정과 행동에 영향을 미칠 가능성을 배제하는 것은 아니다. 오히려 출산 관련 인구행태 가운데 거시적 인구 현상으로 일반화, 유형화할 수 있는 핵심 경로에 우선 초점을 맞추고자 하는 의도가 자료 구축에 명시적으로 반영된 것으로 보는 것이 맞는다. 가임여성이 관측의 기본 단위가 되는 출산력 패널이라도 개별 여성의 가족력(가족 관계, 혼인 관계 등)을 통해 친지나 배우자의 영향이 반영될 수 있다. 필요하다면 출산력 패널의 관측 단위를 핵심 연령대 여성과 남성으로 확장하는 유연한 대안도 생각해 볼 수 있을 것이다.

다만, 패널의 주제를 어떤 기준과 범위에서 설정하는지, 다시 말해 인구 매개 요인의 범주 분류를 어떻게 하는지가 중요해진다. 패널 자료의 활용 범위는 패널의 주제에 한정될 수밖에 없기 때문이다. 출산력에 관한 패널은 개인-미시 수준에서 가족력의 결정 요인을 분석하는 데 활용하기에 적합하지 않다. 주제에 대한 절대적이고 완벽한 분류 기준이 없기 때문에 패널의 주제 선정과 분류가 자료의 가치와 활용도에 절대적일 수 있다.

주제별 인구패널의 대안으로 기존 인구 관련 패널들에 비견할 수 있는 인구 대상별 인구학적 패널 구축을 생각해 볼 수도 있다. 인구 현상의 주체인 인구집단(연령, 세대)별로 아동·청소년 패널, 청년 패널, 중장년 패널, 고령자 패널, 여성 패널 등을 각각 구축하는 것이다. 하지만 기존의 해당 패널 자료들에서 드러나듯이, 인구 대상별 패널은 인구학적 패널의 요건에 부합하지 않는 수평적, 배타적 분할 자료가 될 가능성이 크다. 인구 변동 요인과 특징적 현상들이 연령 또는 세대별 인구집단으로 정확히 구별되지는 않기 때문이다. 예를 들어, ‘청년 패널’을 출산력과 이동력 추

적을 위해 구축할 수 있겠지만 이를 바탕으로 관측된 출산력 관련 주요 인과관계를 전 인구로 일반화하는 것은 간단치 않다. 연령대별로 구분된 자료는 사회정책이나 복지서비스에서처럼 대상이 명확히 구분되는 주제에는 그 나름의 장점이 있지만 이마저도 인과관계와 정책 효과가 연령대를 초월하는 시계와 시차를 가지는 경우에는 마찬가지로 문제가 된다.

## 제2절 인구 변동 패널화 관련 통계학적 고려 사항

### 1. 표본추출과 표본 대표성

제2장에서 논의한 바와 같이 패널 자료의 표본 대표성은 기본적으로 자료 시작 시점의 모집단(총인구)에 대한 대표성을 의미한다. 따라서 패널이 종단 방향으로 커지게 되면, 즉 패널 자료가 포괄하는 기간이 늘어날수록 모집단 대표성은 희석될 수밖에 없다. 패널의 종단 규모가 커지는 것은 패널 유지기간이 길어지기 때문이기도 하지만 자료의 관측 주기가 길어지는 경우에도 불가피하게 발생한다.

패널의 성숙 또는 노화에 따른 문제를 우려해서 표본을 보강하고 복수의 패널 세대(waves)를 중첩해서 유지하는 등 보완적 장치들을 활용할 수 있지만, 패널의 종단 방향과 표본 대표성 간의 긴장을 근본적으로 해소하는 것은 아니다. 자료 관측 주기가 모집단 대표성을 우려할 만큼 길어지는 경우에는 통상의 장치들로 문제를 완화할 수 있는지도 확실치 않다.

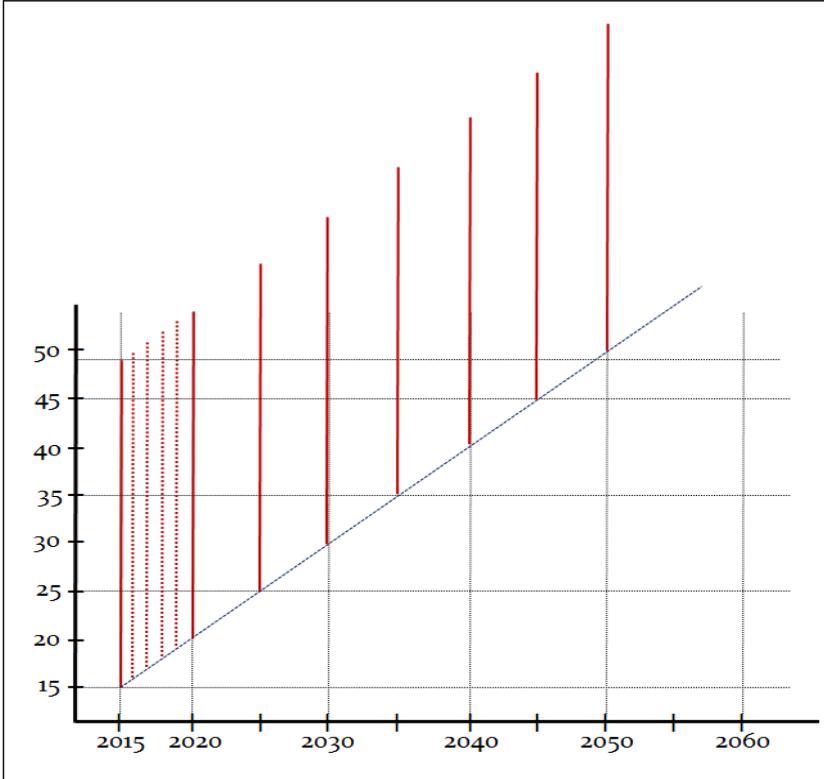
인구학적 패널의 경우 이는 상당히 심각한 문제가 될 수 있다. 인구 변동과 인구구조 변화 자체를 주제로 삼기 때문에 표본의 인구학적 특성과 표본 내에서의 인구 동태가 모집단을 얼마나 잘 반영하는지가 더 중요할

수밖에 없다. 그림에도 인구 현상은 장기에 걸쳐 점진적으로 일어나고 시작부터 끝까지의 시차가 긴 것이 일반적이므로 자료 관측의 주기와 기간이 상대적으로 길어야 한다.

또한, 앞 절에서 살펴본 바와 같이, 인구학적 패널은 기본적으로 개별 관측 대상(개인 또는 가구)의 인구경험을 시작부터 완료까지 전 기간에 걸쳐 따라가는 완결인구력(cohort demographics)에 초점을 맞춘 자료이다. 이러한 코호트 증시 성향은 패널 표본의 모집단 대표성과 상충할 여지가 크다. 상대적으로 시점별 인구 특성, 즉 기간인구력(period demographics) 관점에 더 충실한 횡단면 자료가 표본의 모집단 대표성에 더 적극적이고 또 그럴 수밖에 없는 것과 비교하면 분명해진다.

출산력 또는 임신·출산 관련 패널 자료를 구축하는 경우를 상정해 보자. 관측의 대상을 가임여성(15~49세)으로 특정하고 인구주택총조사를 기반으로 패널을 축적하면 표본이 어떻게 성숙하는지를 [그림 4-4]의 lexis diagram상에서 살펴볼 수 있다. 2015년 조사 시점의 가임여성 전체를 표본으로 삼고 5년 단위 조사 주기에 따라 패널을 유지한다고 하면 패널 표본은 그림에서 45도 선 위쪽의 붉은 선분들로 표현된다. 관측 주기를 일반적인 패널 자료에서처럼 연간으로 좁히면 표본은 점선 선분들로 표시된다.

[그림 4-4] Lexis diagram에서 표현된 총조사(2015년) 기반 가임여성(15~49세) 패널



자료: 저자 작성.

그림에서 보듯 이 표본 코호트(2015년 가임여성)를 따라가면 2050년에 전체 표본의 완결출산력이 관측되는데, 동시에 표본 내에서 49세 이하의 가임여성에 대한 실효 관측치는 표본의 노화와 더불어 비례적으로 감소한다. 이 기간에 표본의 인구분포와 특성은 2015년 코호트에 고정되어 시점별 가임여성(15~49세) 모집단의 인구특성과 괴리가 커진다. 이는 실효적 표본의 탈락이 일상적인 다년간 조사와 비교할 수 없이 큰 규모로 구조적으로 발생하기 때문이다. 패널 성숙의 중간 단계인 2030년에 이미

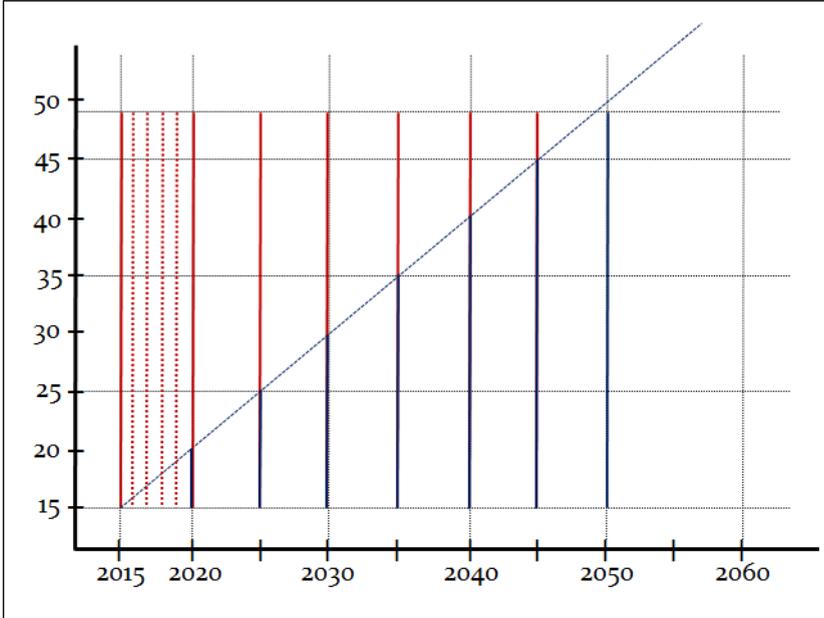
실효적 표본의 마모에 따른 모집단과의 괴리가 상당히 큰 수준에 이르는 점을 그림에서 분명히 볼 수 있다.

인구학적 패널을 불균등패널(unbalanced panel)로 구축하면 이 문제를 의외로 간단하게 해소할 수 있다. 일반적인 패널 자료는 위에서 본 것처럼 각 표본이 패널 유지 전 기간(혹은 전 생애기간)에 관측되는 균등패널(balanced panel)을 기본으로 한다. 이는 모집단(총인구)의 동태적 구조 변화를 반영하지 못한다. 개인의 생애과정은 연속적이지만 모집단 진입(출생)과 퇴장(사망) 시점에서 단절되는데, 매 시점 이를 연령 코호트별로 불균등하게 집계한 결과가 모집단이기 때문이다. 연령의 범위를 기준으로 연령 코호트별로 진입과 탈락이 순차적으로 중첩되도록 관측하게 되면 자연스럽게 불균등패널이 되고 모집단의 동태 구조 변화를 따라가게 된다. 불균등패널의 분석기법은 근본적으로 균등패널에서와 다르지 않으므로<sup>45)</sup> 인구학적 패널의 설계에 굳이 균등패널 표본을 고집할 이유는 없다.

위 출산력 패널의 경우 49세 이후의 표본은 패널에서 탈락하는 것이 자연스럽게 새로 15~20세가 되는 연령대 표본이 (가임여성 전체 인구분포 특성에 맞는 규모로) 패널에 추가되는 것이 맞는다. 그 결과는 일종의 불균등패널이 된다([그림 4-5] 참조). 이는 개념적으로 조사 시점마다 새로운 표본 세대(wave)를 부분적으로 추가하는 것과 마찬가지로이다.

45) 이것이 성립하려면 관측 주기 대비 관측 기간이 충분히 길어야 한다. 그렇지 않으면 통계적 추론을 위해 패널을 균등화할 때 자료 모서리의 탈락이 너무 커서 추론의 기본 전제에 위배될 수도 있기 때문이다.

[그림 4-5] Lexis diagram에서의 불균등 가임여성 패널



자료: 저자 작성.

이 자료가 충분히 성숙되면 특정 연령대 코호트의 완결출산력 추적이 가능해진다. 필요에 따라 균등패널의 상단이 45도 방향으로 절단된 형태의 단절균형패널(censored balanced panel)로 분석할 수도 있다.

이러한 방식을 따르면, 관측 2기 이후부터는 표본을 추출하고 추가하는 비용이 새로운 표본(15~20세)으로 국한되어 매기 전체 표본을 새로 구성하는 횡단면 조사에 비해 작은 수준으로 유지될 수 있다. 패널 관측 기간이 긴 인구학적 패널에서 표본을 장기간 유지하는 것은 상당한 부담이 될 것이다. 일반적인 패널에서의 표본 탈락과 마모가 더 오래 지속되는 것을 의미하기 때문에 패널의 유지·관리가 그만큼 어려운 문제가 될 것이다. 이때 불균등패널이 그러한 문제를 일부 완화하는 방향으로 도움

이 될 수 있다.<sup>46)</sup>

불균등패널의 표본은 매 시점 일부만 교체되고 모집단의 연령 구조와 분포를 더 자연스럽게 따라가기 때문에, 시점별로 표본의 모집단 대표성에 대한 부담이 완화될 수 있다. 특히, 표본 추출 과정에서의 통상적인 표본 층화와 가중치 산출 작업과 근거 마련 부담의 경감이 기대된다.

## 2. 인구를 주제로 한 패널의 시계열적 특성

인구학적 주제의 패널은 장기 인구 현상을 주제로 하기 때문에 종단 측면이 중시될 수밖에 없다. 따라서 제2장에서 살펴본 바와 같이 자료의 시계열 특성에 대한 별도의 고려가 필요한데, 특히 두드러지는 다음과 같은 문제들은 자료 설계 단계에서부터 감안되어야 한다.

### 가. 중장기 시계와 시차, 동태구조적 변수 관계

인구학적 패널에서의 변수들은 시계열적이다. 관측 주기가 상대적으로 길다고 가정하면, 종단의 크기가 시계열분석이나 일반적인 대표본 확률 이론들을 적용할 만큼 충분히 크지는 않더라도 관측 기간이 장기에 걸쳐 있고 주 관심 대상도 변수들 간의 장기-구조적 관계일 가능성이 크다. 하지만 이를 개인 수준에서 추적, 관측하다 보면 변수마다의 시계(time horizon)가 다르고 동태적 변화의 양상이 같은 수준에서 비교하기 어려운 경우가 적지 않을 수 있다. 앞에서의 출산력 패널을 예로 든다면, 관측 단위인 여성 개인별로 패널의 각의 변수들은 15~49세 기간 동안 5년 주

46) 불균등패널 유지·관리의 전체 비용과 부담은 매기 탈락, 추가되는 표본 규모, 표본 유지 기간, 관측 주기 등 여러 가지 요소에 따라 달라지므로 횡단면 자료나 균등패널과 일률적으로 비교하여 판단하기는 어렵다. 아래 주 68)의 관련 논의 참조.

기로 7년에 걸쳐 관측된다. 관측치 규모는 작지만 관측 기간이 30년을 상회하는 장기가 된다. 이 기간에 개인의 결혼·임신·출산에 영향을 주고받는 주요 변수로 교육 수준, 소득 수준, 취업 여부만 보아도 이들 사이의 장단기 특성과 변동 양상은 제각각이다.

인구학적 변수들 간의 시계열 특성 차이는 가성적(假性的) 관계를 이루어 실제로는 무관한데 실증적으로는 연관성이 입증되는 착시현상으로 이어질 수 있다(제2장 제4-2절, p. 46 참조). 가성적이 아니더라도 패널 관측 기간을 넘어서는 장기적 주기나 시차(時差) 특성을 가지는 변수들의 경우 패널 자료의 시계(視界) 내에서 유의미한 관계를 포착하지 못하는 경우(aliasing)가 발생할 가능성도 크다(제2장 제4-2절, p. 43, 각주 19) 참조). 여성 개인별로 20대 이전의 교육과 학력 수준 축적이 결혼 성향과 가능성을 변화시켜 자녀 출생 여부 또는 출생아 수 결정에 기여하는 효과는 단기에 관측할 수 없다. 부와 소득의 세대 간 이전 형태가 공적 제도 중심으로 변화하는 것이 개인 수준에서 노후소득보장에 어떤 작용을 하는지도 단기적으로 드러나지 않는다.

인구 변동과 같은 장기 현상을 대상으로 하는 정책을 기획하고 성과를 평가할 때 특히, 시계열적 가성회귀(spurious regression) 관계나 정책 기획-효과 시차와 자료 관측 주기 불일치로 인한 착시(aliasing)는 심각한 문제가 될 수 있다. 기간출산력에 초점을 둔 출산 장려 정책(출산장려금)이 완결출산력의 변화를 목표로 하거나, 완결출산력에 대한 함의는 상반되는 정책(일·생활균형)을 기간출산력 목표에만 맞추게 되면 기획 의도에도 어긋날 뿐 아니라 정책의 효과성도 가성적일 가능성이 크다.

여성의 대학교육이 (합계)출산율에 미치는 효과를 분석한 손호성(2017)은 회귀굴곡모형(RKD)을 적용하여 여성의 고학력과 출산력의 인과관계를 실증적으로 규명했지만, 그 관계가 시계열적으로 가성적일 가

능성을 배제할 수는 없다. 분석 기법에서도, 여성 고학력(대졸 확률)과 출산율의 관계가 1973년 출생코호트를 기점으로 전후 세대 사이에 구조적으로 다르다는 핵심 가설은 시계열적으로 설정 오류(misspecified)일 가능성이 있다. 1973년 코호트의 대학 진학 시점(1993년)에 급변한 대학 관련 정책에 기인한 대학 진학 효과의 단절은 식별했지만, 교육 수준의 출산에 대한 한계효과를 장기탄력성으로 측정해야 하는 점은 간과한 것으로 보이기 때문이다.

출산 지원(장려) 정책에 대한 실증분석 연구들을 폭넓게 수집하여 비판적으로 검토한 Gauthier(2007)는 출산정책의 효과성을 경험적으로 입증하는 것이 쉽지 않다는 것을 보였다. 그 핵심적인 원인의 하나로 효과성 분석을 위한 자료가 대부분 한계가 뚜렷하다는 점을 지목하였는데 기본적으로 이는 관련 변수들의 시계열 특성이 자료와 분석 방법에서 충분히 감안되지 못했다는 사실과 연결된다.<sup>47)</sup>

인구학적 패널에서 변수들의 시계열적 특성을 최대한 반영하는 방법은 인구 현상이나 정책 효과의 시계에 적합하도록 관측 기간을 충분히 길게 잡는 것이다. 관측 주기 불일치로 인한 착시를 줄이기 위해서는 주기를 최대한 짧게 정하는 것이 상식적지만, 주요 인구 현상의 핵심 주기(prime frequencies)에 호응하는 수준에서의 균형이 요구된다. 여기에 자료의 규모에 따른 비용과 유지의 현실적 문제는 제약으로 작용한다.

인구학적 패널 자료의 관측은 장기에 걸쳐 이루어져야 한다. 인과관계의 장기탄력성에 초점을 맞추어야 인구 현상의 시계에 맞는 적합한 한계효과나 정책효과성 분석이 가능해진다.

47) Gauthier(2007)가 지목하는 또 다른 핵심 원인은 효과분석 모형들의 근본적 내생성(endogeneity) 문제인데, 이는 패널 자료와 패널 분석의 분석적 가치와도 연관된다. 본 보고서 제2장 제1절, p. 28 참조.

## 나. 자료의 가변성과 장기적 추적의 필요

패널 자료의 구축과 패널 분석 방법론을 정당화하는 근거는 종단 방향으로의 변수의 가변성(variability)이다. 자료의 내용에 포함되는 주제 영역과 변수의 선정에서도 패널 유지 기간에 다른 변수들의 움직임을 설명하는 데 충분한 변이(variation)가 기본 전제가 된다(제2장 제2절, p. 44 참조).

인구 변동은 다른 사회·경제적 배경 요인이나 이들의 매개 요인들로부터 영향을 주고받는 과정이 대체로 길고 점진적이다. 사회 제도나 산업 구조, 경제 체제 등과 같은 환경 요인들은 단기적인 변화의 관측이 무의미할 정도로 변화가 더디거나 희소할 수 있다. 실제로 인구 동태를 설명하는 데 이러한 장기적인 변화가 대부분을 차지한다면, 단기 변동은 미미하거나 아예 부재할 수 있다. 패널 자료에서 일반적인 분기별, 혹은 연간 관측 주기에도 분석의 의의가 있는 인구학적 변화는 드물 수 있다. 인구주택총조사(5년)보다 짧은 주기의 인구학적 패널이나 단기적 인구 변동 관련 조사는 실질적으로 유효성이 없는 비용 낭비라고 보는 비판적 견해는 대체로 이런 관점을 배경으로 한다.

앞서 사례로 든, 손호성(2017)과 Gauthier(2007)에서 통계적으로 유의하지 않은 결과들의 상당 부분은 출산율의 가변성이 자료의 관측 주기에 비추어 충분히 크지 않은 점에 기인한 것일 수 있다. 이상협 외(2016)에서 전반적인 출산장려정책의 효과가 통계적으로 유의하게 포착되지 않은 것은 자료의 기간이 7년(2004~2010년)으로 짧은데 그 기간에 출산율은 뚜렷한 단기 변동 없이 추세적으로 하락했기 때문으로 설명된다. 같은 연구에서 부가적으로 수행한, 영유아 보육료와 양육수당 지원 정책의 출산율 제고 효과 분석 역시 결과의 상당 부분이 2000~2015년이라는 기간

의 길이와 특수성에 기인한다. 보육·양육 현금 지원 정책들이 보편적 지원으로 확대, 전환된 2012년과 2013년 이후의 정책 효과 변화를 추정하는 것이 초점인 데 반해, 지표인 출산율은 2005년을 기점으로 이전의 안정적 등락과 이후의 추세 하락으로 확연히 갈리는 모습이 관찰된다. 2012년과 2013년 이후의 정책 효과를 관측하기에 2015년까지의 자료는 턱없이 미흡한 점도 문제다.

이러한 자료의 가변성 문제 때문에 인구학적 패널에서 관측 주기의 선정은 특히 주의가 필요하다. 더 구체적이고 실증적인 근거가 필요한 전제이겠으나, 대부분의 인구 현상이 1년 이내에 추세를 벗어난 유의미한 변동으로 나타나기 어렵다고 본다면 연간 주기로 자료를 관측하는 것은 의문이다. 인구패널의 초기 논의에서 기존의 다른 패널 사례를 준용하여 연간 혹은 격년 주기의 조사·관측을 기본으로 간주한 관행(이삼식 외, 2006, pp. 122-123)은 재고할 필요가 있다.

마찬가지로 5년 주기의 인구주택총조사가 주요 인구 현상의 주기에 부합하는 것인지는 확실치 않다. 다만 장기간에 걸쳐 받아들여진 5년 주기가 단순한 관행이기보다는 어느 정도 경험적 타당성을 반영하고 있는 것은 분명해 보인다. 출산력조사의 주기가 3년이라는 점도 출산력 패널이라면 유념할 만하다. 딱히 5년 또는 3년 간격의 이유가 불분명하더라도 보수적으로 볼 때 이보다 짧은 주기를 고려하는 것은 충분한 근거가 필요하다.

패널의 주제와 관련 인구 현상의 특성에 따라 자료 관측의 주기는 적합하게 조정되어야 한다. 이를 위해서는 해당 인구 주제별로 주기 특성을 포함한 시계열적 성격에 대한 분석과 실증이 필요하다. 관측 대상이 개인인 경우 관측 연령 간격에 대한 고려도 요구된다. 앞 절에서 살펴보았듯

이 주제에 따라 주 관측 대상 연령대, 세대 등의 범위가 설정될 필요가 있다. 동시에 해당 범위 내의 모든 연령대에 걸쳐 촘촘한 표본 구성이 필요 한지도 살펴보아야 한다. 제3장 제2-1절에서 해외 사례로 거론된 GGS 패널의 관측 주기와 표본 구성은 참고할 만하다.

#### 다. 자료 가변성과 표본 대표성의 충돌

인구학적 패널에서 자료 가변성의 문제는 자료 설계와 관련된 기술적인 문제로도 부각될 수 있다. 패널의 최초 표본은 모집단 대표성을 위한 설계—모집단 층화와 층별 표본 배분—로부터 설정된 표본 조사구를 단위로 추출되는데, 패널 주제인 인구 현상의 빈도와 주기에 비해 조사구가 작고 촘촘하게 설계되는 경우가 문제가 된다. 횡단면 조사인 2018년 출산력조사의 예를 보면, 3년 주기의 1만 개 표본 규모에 맞추어 가구 단위 표본 조사구가 700개로 설정되고 각 조사구에서 15개 내외의 표본을 할당추출 하도록 설계되었다(이소영 외, 2018, p. 39). 이 경우 조사구 단위로 볼 때, 관심 인구 변동(출산, 이동, 사망의 변화)은 15개 내외의 표본 가운데 극히 소수에서만 관측되고 변수들 간의 유의미한 관계는 이들 소수의 실측치에 좌우될 수 있다. 그러면 이를 활용한 생존모형분석(duration analysis)이나 사건사분석(event history analysis)과 같은 전형적인 분석 방법(limited dependent variables or discrete choice models)에서 perfect classifier의 왜곡으로 이어질 수 있다(Davidson and MacKinnon, 1993, pp. 520-521).

자료 설계 단계에서 이러한 문제의 가능성을 완화하는 방법은 두 가지이다. 하나는 표본의 크기를 늘리는 것인데, 자료 수집의 비용과 유지 부담이 제약이 될 수 있고 표본 조사구가 커지지 않는 한 조사구 단위에서

의 문제가 해소되지 않는 한계가 있다. 다른 하나는 관측 주기를 길게 정하는 것이다.

인구학적 패널의 경우 앞선 논의들을 감안해 보면 관측 주기를 충분히 늘려 인구 현상의 주기와 빈도에 맞추는 것이 자료 가변성에 대한 적합한 대처로 보인다. 이때 통상적인 균형패널 형식을 취하면 모집단 대표성의 문제가 더 빠르게 악화될 가능성이 있다.

제 5 장

## 인구패널 구체화에 대한 제언: 출산력 패널

제1절 자료 구축

제2절 출산력 패널의 활용도와 가치

제3절 정책 과제



# 5

## 인구패널 구체화에 대한 제언: <<출산력 패널>>

### 제1절 자료 구축

이 장에서는 제3장과 제4장에서의 논의를 바탕으로 인구학적 주제의 패널을 실제로 설계해 본다. 그 대표적인 예로 출산력 또는 임신·출산에 관한 패널을 구축한다면 어떤 형식과 구성, 내용을 갖추게 될지 살펴보고 이로부터 실현 가능한 ‘출산력 패널’ 구축안을 제시하고 학술적, 정책적 활용도까지 구체적으로 검토해 본다. 이로부터 인구학적 주제의 패널 자료가 현실적으로 구현 가능한지, 그 결과는 기존의 전형적 패널과 비교하여 구체적 형태와 내용, 활용 가치에서 어떻게 다른지 확인할 수 있다.

#### 1. 자료 구축 대상

인구 변동의 직접 요인 가운데 출생에 관한 패널 자료를 구축하고자 한다면 가장 핵심적인 주제는 출산력(fertility)이 될 것이다.<sup>48)</sup> 인구학적으로 출생의 포착은 여성의 임신과 출산이라는 물리적 조건에서 벗어나지 못하기 때문이다. 출산력을 결정하는 요인, 출산력의 변동을 설명하는 인과관계, 출산력 변동의 영향을 개인 수준에서 미시적으로 파악하고 분석하고자 한다면 제4장 제1-2절의 개념적 논의([그림 4-2], p. 106)를 근

48) 출생력(procreativity)은 한 사회나 국가가 다음 세대의 구성원을 창출하고 키워 내서 구성원 전체를 뜻하는 인구를, 역외로부터의 유입을 제외한 방식으로 재생산할 수 있는 역량과 능력의 총체를 의미한다. 이는 인구학적 또는 생물학적 관점에서의 재생산 능력(reproductivity, fecundity) 또는 이의 사회적 총합을 의미하는 출산력(fertility)을 포함하는 더 넓은 개념으로 볼 수 있다.

거로 삼을 수 있다. 이에 따르면 임신·출산을 결정하는 개인(또는 가구)의 선택과 행동 양식의 변화를 사회경제적 매개 요인, 환경 요인들의 개인별 차이나 변화와의 관계를 통해 설명하는 미시적 분석의 영역과 이를 거시-집계 수준으로 유형화, 일반화하는 거시적 분석의 영역으로 나뉘게 된다. 출산력 패널은 앞부분을 위한 기반 자료 역할을 주목적으로 구축되고 이를 위해서는 임신·출산에 관한 인구 변수들과 관련 주요 사회·경제적 배경 요인들<sup>49)</sup>을 관측, 추적하는 자료가 필요하다(이삼식 외, 2006, pp. 109-114, p. 110 <표 5-2>, p. 114 [그림 5-3]; 이삼식 외, 2016, pp. 17-24).

이러한 기획에 부합하는 자료 관측, 수집의 단위는 개인, 그중에서도 가임여성이다. 임신·출산에 대한 의사 결정과 행동은 가구나 가족이 아니라 행위 주체인 여성 개개인을 단위로 이루어지고 발현된다. 그 배경이 되는 사회경제적 매개·환경 변수들의 변화와 영향도 가구나 가족을 단위로 하기보다는 여성 개개인 수준에서 파악되는 것이 적확하다. 교육 수준, 취업 여부, 건강 상태와 같이 개인에게 체화되는 변수들뿐만 아니라 가족 또는 가구 단위에서 공유되는 세대 관계, 가치관, 소득 계층과 같은 변수들도 임신·출산에 영향을 미치는 경로의 기착점은 여성 개인이다.

이는 기존의 출산력조사가 기혼 여성에 초점을 맞춰 이뤄져 왔던 바탕의 암묵적 인식과도 일치한다. 혼외 출산의 비중이 미미한 최근까지의 우리나라의 특성상 출산의 행위 주체가 되는 기혼 여성은 가임여성과 다름 없기 때문이다. 다만, 결혼이 출산의 절대 불가분인 전제가 아니고 임신·출산에 대한 여성의 선택과 행동이 생애주기적 판단으로서 혼인 이후에 국한되어야 할 근거는 박약하므로 가임여성 전체를 모집단으로 삼는 것

49) 제4장 제1-2절을 따르면 교육(학력), 경제활동(직업 선택, 취업, 노동, 소득·소비·투자), 가족력(결혼, 가족 구성), 세대 관계, 건강, 환경, 가치관 등이 포함된다.

이 타당하다.

임신·출산에 집중하더라도 출산력에 대한 총체적인 자료라면 배경과 여건으로 빼놓을 수 없는 배우자, 남성, 부모와 친지 등에 대한 변수, 가족 단위에서 파악해야 할 임신·출산 변수들도 포함해서 관측해야 할 텐데, 가임여성에 국한된 자료는 이 부분을 간과하는 한계가 분명하다는 비판을 제기할 수 있다. 이러한 우려는 자료 관측 단위의 문제는 아니다. 가임여성 개인 단위로 구축된 자료가 개인이 속한 가족과 친지에 대한 내용을 포함하지 못할 이유는 없다. 임신과 출산에 대한 의사와 행동을 결정하는 주체를 여성 개인으로만 좁혀 볼 것인지, 다른 성, 세대, 연령의 인구집단까지 포함할 것인지는 자료 설계 이전의 기초에 관한 선택의 문제이다. 출산력을 미시-개인 수준에서 분석하기 위해 여성 이외의 다양한 행위 주체의 행태까지 함께 살펴야 하는지는 더 깊은 논의가 필요하다.

대안으로서 가구를 기본 단위로 자료를 구축하는 것도 생각해 볼 수 있다. 가구가 관측의 기본 단위가 되면 가구원 개인에 대한 조사는 필요에 따른 부차적인 요소가 된다. 이는 일단 가구(또는 가족) 단위에서 미시 수준에서 임신·출산을 결정하는 요인과 작동 기제를 식별하고 이해하겠다는 것을 뜻한다. ‘개인’의 선택과 행위(임신·출산)에서 비롯되는 사회·경제 현상으로서의 인구 현상(출산력과 가족력을 중심으로 한 인구력)을, ‘개인’의 행위 결정 과정에 초점을 맞춰 설명하려는 것이 미시적 접근이고 이를 위한 근거 자료가 패널이라면 자료 관측은 ‘개인’을 기본 단위로 식별해야 타당하다.

가구 단위의 출산력 패널은 이 ‘개인’이 가구 또는 가구의 구성원이라고 간주하는 것과 같다. 깊은 철학적인 논의가 있어야 하겠지만, 행위 결정의 주체인 ‘개인’을 가임여성으로 보는 것보다 가구로 보는 것이 미시 실증 분석을 위해 더 나은 선택인지는 의문스럽다. 가구 단위를 따르게

되면, 가임여성이 복수인 가구의 경우 개별 가구원 부가 조사에 포함되지 못한 개인 특성—예를 들어, 가구원(여성) 단위의 소득, 소비, 노동에 관한 정보, 부가 조사에서 수집하지 못한 개인적 정보—은 관측이 어려울 수도 있다. 가구 단위에서 출생에 관련된 자료가 수집되면 임신과 출산에 대해 여성의 선택과 결정이 행동으로 옮겨지는 과정이 정확하게 식별되지 못할 수도 있다.<sup>50)</sup>

결혼과 출산이 인구 변동(출산력)에 이르는 경로를 가족력을 매개로 찾는다면 가구 단위 접근이 타당할 수 있다. 반대로 가족력을 주제로 하는 인구학적 패널의 관측 기본 단위를 가구 대신 가족 형성과 분화를 주도하는 가구원(개인 또는 부부)으로 선택할 수도 있다. 모집단과 표본, 관측의 단위를 개인으로 볼 것인지, 가구(또는 가족)으로 볼 것인지는 서로 배타적이지 않다. 가구조사에 가구원 개별 조사가 부분적으로 포함되는 흔한 예로 보아도 그렇고, 개인 단위 관측 자료이더라도 가구원으로서 가구 내에서의 위치—가구 단위의 의사결정을 주도하는 주체(가구주 또는 배우자)인지 아닌지(자녀, 피부양자 등)—에 따라 가구 성격을 반영한 내용을 담을 여지가 있는 점으로도 그렇다. 하지만 어느 쪽을 기본으로 삼는지에 따라 패널 자료에서 관측하고자 하는 주요 미시적 행태와 인과관계의 성격과 함의는 달라질 수밖에 없다.

## 2. 자료 구축 방법

자료 관측의 기본 단위를 가임여성으로 본다면 이는 엄밀하게 표본 설계와 추출 단계에서부터 적용하는 것이 타당하다. 출산력조사와 같은 대

50) 반대로 개인 단위의 자료에서 출산력에 관계된 가구와 가족 정보를 포착하는 것은 상대적으로 용이하다.

표적인 횡단면 조사 자료에서의 통상적 관행은 인구주택총조사를 표본추출의 기본틀로 삼고 전국 일반 가구의 분포 특성(가구 지역 특성, 가구당 여성 수, 가구원 수)에 따른 층화를 거쳐 표준화된 조사구를 추출하여 각 조사구마다 일정 수의 표본을 가구 단위로 추출하는 방식을 따르는 것이다(이소영 외, 2018, pp. 27-42). 출산력 패널을 위한 가임여성 표본 역시 동일한 조사구 방식으로 구성하여 모집단 대표성을 확보할 수 있다. 이때 모집단, 조사구 설정과 표본 추출의 기본 단위를 가임여성으로 삼으면 된다. 혹은, 절충적으로 출산력조사에 조사구 방식을 따르되 표본 추출을 조사구 내의 가임여성 집락에서 시행하는 것도 방법이 될 수 있다.

모집단과 표본 추출, 관측의 기본 단위가 어떤 인구학적 주제인지 따라 패널마다 다른 것이 자연스럽다는 점은 현용 패널 자료의 다양한 실례를 통해 이미 확인된 바 있다(이삼식 외, 2006, p. 133). 같은 사례 검토에서 청년패널, 노동패널, 복지패널, 여성·가족패널 등 우리나라의 대표적인 패널 자료들은 “전국 일반가구를 모집단으로 ... 가구를 단위로 표본이 표출되는 가구패널”이라는 점도 적시되지만, 이들의 경우 공통된 주제가 가족의 형성, 성장 및 분화이거나 가족을 매개로 하는 부양·돌봄과 부와 소득의 세대 간 이전과 같은 가족력이라는 점은 간과되고 있다. 가임여성 개인을 기본 단위로 출산력 패널 자료를 구축한다면 인구학에서의 가임기간인 15~49세 연령대 전 여성을 모집단으로 보는 것이다. 여기에 조사구 방식을 대입한다면 가임여성의 분포 특성을 반영하기 위한 층화(예: 지역, 연령, 교육 수준)를 거쳐 조사구를 추출하고 가중치 부여를 통해 표준화하는 표본 설계가 가능하다. 앞서 제4장 제2-1절의 제안을 따른다면, 패널 자료 시작 시점에서 조사구 방식의 표본을 추출하되 다음 관측 주기부터는 모집단(전체 가임여성)의 인구분포와 동태적 변화를 1차적으로 반영하여 표본 가운데 가임기(15~49세)를 벗어난 연령대 코호트를 탈락

시키고 새로 가입기에 진입한 연령대 코호트에서 표본을 추가하는 것이다. 그 결과는 앞에는 논의한 바와 같이 불균등패널로 귀결된다(그림 4-5 참조).

매 관측 시점마다 탈락되고 추가되는 표본의 크기는 전체 표본의 규모와 관측 주기, 가입여성의 분포 특성에 의해 결정된다. 이 방식에 따라 매 시점 새로 진입하는 표본은 그 시점의 모집단 분포 특성과 남아 있는 표본 코호트의 특성을 함께 반영하여 해당 연령대 여성을 부분적인 모집단으로 삼아 추출하도록 설계한다. 여기에도 조사구 방식을 적용할 수 있다.

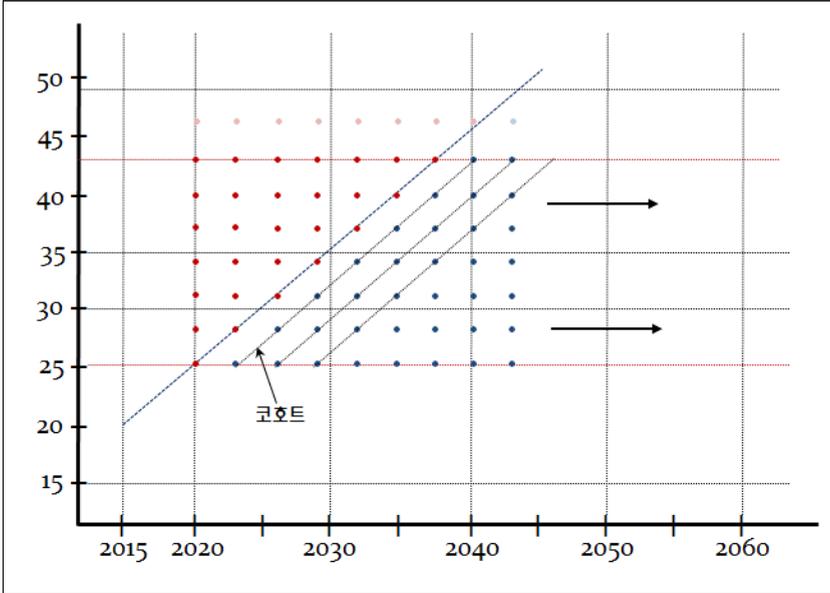
이와 같은 표본 구성과 불균등패널은 각 가입여성을 연령별 코호트를 기준으로 가입기 동안 추적하여 임신·출산의 완결력을 포착하도록 해주는데, 그로 인해 패널을 유지하는 것이 현실적으로 큰 부담이 될 수 있다. 표본 코호트별로 패널에 머무는 기간(35년)이 길어야 하고 그에 따라 장기 유지에 따른 비용, 탈락과 마모의 확률이 급격하게 커질 수 있기 때문이다.

이러한 현실적 부담을 완화하고 감당할 수 있는 범위 내로 제어할 수 있는 방법은 표본의 규모를 줄이는 것인데, 이는 두 방향에서 가능하다. 첫째는 표본의 연령 간격과 관측 주기를 늘리는 것이다. 현실적인 고려가 아니더라도 이는 합리적이고 효율적인 선택일 수 있다. 인구 변동이 장기적이고 점진적인 성격을 띠는 것은 이미 살펴본 바와 같고, 임신·출산의 동태적 변화 가운데 출생력과 전체 인구 변동에 기여하는 중요한 부분은 시계열 특성으로 보아 장기지속적(persistent)이거나 장기-추세적(long-term and trending)인 요소이다. 미시-개인 수준의 임신·출산 행태도 단기·일시적 변화보다는 중장기 추이가 핵심이다. 단기·일시적인 임신·출산 변동은 규모는 작고 빈도는 낮아 가변성도 작다. 출산력 변동

으로 유형화, 일반화할 만큼 안정적, 지속적이지도 못하다. 인구주택총조사(5년)에서 관측되는 변동만으로도 유의한 출산력 변화는 충분히 포착할 수 있다는 견해, 출산력조사를 3년 주기로 시행하는 관행의 바탕에는 이러한 개념적 판단이 깔려 있다. 그렇다면 출산력 패널에서도 예를 들어 관측의 연령 간격과 주기를 3년으로 정할 수 있다. 그렇게 15세부터 가임여성의 표본을 구성하면 연령별 표집 포인트는 12개가 된다. 둘째, 인구학적 가임기간 15~49세 전체에 걸친 표본 구성도 재고의 여지가 있다. 임신·출산을 비롯한 여러 인구 행태의 연령 구조는 특정 연령대에 집중되는 분포 특성이 전형적이다([그림 4-3], p. 108). 여성 개인의 임신·출산에 관한 의사결정과 행태를 설명하는 경로와 인과관계가 거시·집계 수준의 출산력 변화로 일반화하기 위해서는 유형화할 수 있을 만큼 다수에서 전형적으로 관찰되어야 한다. 이런 조건에 맞는 미시 행태가 임신·출산의 집중 연령대 밖에서 발생할 가능성은 희박하다. 정책적 목적 등의 이유로 특정 연령대나 계층 여성들의 출산력에 주목하는 경우도 있겠지만, 연령 구간이 집중 연령대 밖이라면 인구학적 관점에서 일반화될 정도로 빈번하기는 쉽지 않다.

예컨대 관측 연령 구간을 최근 우리나라 여성 임신·출산의 주 연령대를 포함하는 25세~43세로 좁혀서 표본을 구성하면 3년 간격의 연령별 표집 포인트는 7개(25세, 28세, 31세, 34세, 37세, 40세, 43세)로 줄어든다. 이를 관측 주기 3년과 연결하면, 표본의 규모와 유지 기간의 부담에 기인하는 현실적 문제를 상당히 완화할 수 있을 것이다. 왜냐하면, 이렇게 구축한 출산력 패널이 최대 크기에 도달했을 때 전체 표집 포인트는 42개로 한정되기 때문이다. 각 개인이 25세부터 시작해서 43세까지의 개인별(실질)가임기를 완결하고 패널에서 이탈하는 시점까지 걸리는 시간은 최초 관측 이후 15년, 관측 주기로는 6회차가 된다([그림 5-1] 참조).

[그림 5-1] Lexis diagram에서 표현된 실질 가임여성(25~43세) 불균등패널



자료: 저자 작성.

출산력 패널의 적정 규모 또는 유효 최소 규모에 대한 이론적인 근거는 뚜렷하지 않다. 표본의 모집단 대표성, 자료의 가변성을 저해하지 않고 이들 간의 균형을 맞출 수 있는 규모라는 개략적인 기준은 반영이 되어야 하지만 이를 넘어 구체적인 수치에 대한 판단 기준을 찾기는 어렵다. 패널 구축 및 자료 수집 비용, 장기간 유지·관리 부담, 패널 마모와 탈락에 대한 대처 등 현실적 문제들이 패널의 규모를 제한하는 실질적인 제약조건이 된다.

지금까지의 구상에 따른 출산력 패널을 표본 수 1만 명 수준에서 구축, 운용, 유지하려면 패널 시작 시점에 이 규모의 표본 추출과 관측이 필요하다. 이후에는 매 주기 새로 편입되는 표본을 추가로 추출해서 패널에 더하는 일, 관측 주기 간 패널에 잔류하는 표본을 유지·관리하는 일이 실

질적인 패널 관리 부담이 된다. 실질 가임여성 모집단의 연령 구조에 달려 있기는 하지만, 3세 간격의 연령별로 가임여성 수가 비슷하거나 완만하게 줄어드는 형태라고 본다면 관측 시점마다 1400명 내외의 표본이 탈락하고 이보다 다소 적은 수의 표본이 새로 추가될 것이다.

연령별로 표본이 패널에 잔류하는 기간이 15년이므로 설계상 반영되지 않은 표본 탈락과 마모까지 감안하면 패널의 전체 규모는 상당히 빠르게 줄어들 수 있다. 패널의 의도하지 않은 탈락률(또는 유지율), 실질 가임여성 모집단의 설정 범위, 모집단의 연령 구조와 분포 특성, 표본 관측점의 간격—즉, 표집 체(sieve)의 조밀성—이 모두 패널이 어느 규모로 설정되고 얼마나 마모되고 어떻게 유지될 수 있는지에 영향을 미칠 것이다.

표본 수가 1만을 상회하는 3년 주기 횡단면 조사인 출산력조사 자료와 표본 수 8000 이상의 연간 주기 가구패널인 복지패널의 선례를 감안하면 지금까지의 구상에 따른 표본 수 1만 명 내외의 출산력 패널을 구축, 운용, 유지하려면 연평균으로 5억 원 이상은 소요될 것으로 판단된다.<sup>51)</sup>

### 3. 자료의 주요 내용

앞 장 제1-2절에서의 논의에 따르자면, 출산력 패널의 구성은 개인(가임여성) 수준에서 임신·출산과 관련된 선택과 행동 양식 그리고 이들과 인과관계를 맺는 미시 요인들을 담아야 한다. 여기에는 임신과 출산 관련 인구 변수, 가족력과 가족관계(혼인 관계, 가족 구성, 가구 내 지위, 가족

51) 연간 주기인 복지패널과 비교해 자료 수집의 빈도는 3분의 1로 줄지만 관측 주기 간 패널 유지 비용은 더 많이 들 것이라는 가정과 복지패널에 비해 변수 개수와 자료 수집 항목을 간소화한다는 방침을 전제로 개략적으로 판단한 결과이다. 3년 주기의 표본 부분 추가 표집은 출산력조사보다 규모의 부담이 훨씬 덜하다는 점도 감안하였으나 잔류하는 표본의 유지, 관리가 그 못지않게 부담이 되는 경우를 가정했다. 복지패널 및 출산력조사 연구자와 이 보고서 검토자들의 의견을 토대로 하였다.

내 세대 관계)를 비롯한 인구학적 매개 변수, 개인적 특성을 드러내는 사회경제적 매개 변수들(교육 및 학력 수준, 취업 여부와 직종, 근로와 시간 활용, 자산과 부채, 소득과 지출, 주거 형태와 환경 등)이 포함된다.

보육 및 돌봄 여건, 임신·출산에 친화적인 사회·경제적 구조와 조건, 임신·출산에 대한 역사적·문화적 토양과 가치관과 같은 환경 요인은 개인 수준에서 그 특성이 식별되고 임신·출산에 관한 의사결정과 행동에 직접적으로 관계된 요소만 패널의 관심 내용이 된다. 패널의 1차적 목적이 미시-개인 수준에서의 출산력 결정 요인과 동태적 변화 추적이기 때문이다. 개인별 특성과 개인 간 차이가 상쇄되는 집계 변수들은 개인별로 따로 관측할 이유는 없다. 미시 분석에서 공통적인 배경요인으로 유효한 거시 변수라도 패널 자료의 추적 과정에서 고려할 이유는 없다. 따라서 출산력 패널의 구축 단계에서 거시-집계 변수들은 자료의 구성과 내용에 포함될 필요가 없다.

패널에 포함될 변수를 구분하고 선정함에 있어 출산력을 설명하는 중요성과 비중이 맞도록 매개·환경 요인들을 주제 영역으로 나누고 영역 간 위계를 설정하는 것이 자연스럽다. 이러한 개념 구성은 패널 자료의 분석을 위한 기본적인 관점을 제공하기도 한다. 하지만 패널 구축 단계에서 설정한 주제 영역의 체계가 확정적일 이유는 없다. 패널의 각 주제 영역별로 몇 개의 변수들이 어떤 비중과 중요성을 띠고 포함될 것인지에 대한 판단도 경직적일 필요가 없다. 패널 자료가 유지되는 과정에서 새롭게 중요성이 인지되는 요인과 주제 영역을 추가하거나 보강해야 할 필요가 생길 수도 있다. 미시적 임신·출산 행태를 설명하는 구조적 요인들이 출산력과 사회경제적 배경요인의 동태적 변화로 인해 바뀔 수도 있다.

그런 맥락에서 처음부터 가능한 모든 변수와 주제 영역들을 포함시켜 출산력 패널을 구성하려는 포괄적 접근보다는, 가장 핵심적이고 우선적

인 영역과 변수들에 먼저 초점을 맞추고 관심 영역과 변수들을 점진적으로 늘려가는 접근이 더 적절할 것이다. 출산력의 결정 요인을 규명하는 데에서도 가장 중요한 관계부터 주목하고 차츰 부차적인 부분으로 확장해 가는 것이 맞는다. 자료 구축에서 이런 보수적(parsimonious)인 접근 방식은 자료 규모와 유지 기간의 부담을 완화해 주는 부수적 이득도 가져다줄 수 있다.

주제 영역과 변수 선정이, 패널의 장기적 시계에 비추어 경직적일 필요가 없다는 입장은 출산력 패널이 관련 정책의 체계와 구조<sup>52)</sup>를 따라 구축돼야 할 이유가 없다는 정책적 함의를 지닌다. 패널의 내용과 구성을 결정하는 데에서도, ‘정책목표에 따라 조사 영역을 설정하고, 관련 정책지표를 개발한 이후에 조사 항목과 변수를 여기에 맞춰 선정하는 방식’(보건복지가족부, 한국보건사회연구원, 2008, pp. 67-82)이 기본 원칙이어야 할 이유는 없다. 출산 관련 정책의 기획-집행-평가에 활용하는 것만을 목적으로 하는 정책패널이라 하더라도 내용과 구성부터 정책목표와 정책지표를 기준으로 삼은 자료가 정책의 성과와 효과성을 적절히 평가하고 환류하는 데 필요한 내용과 정보를 제공할 수 있을지는 의문이다. 이런 의미의 정책 자료라면 굳이 패널 형식으로 새로 구축할 필요가 있는지도 불분명하다. 기존의 저출산·고령사회기본계획의 틀 안에서 시행되어 온 연 단위 저출산 대책 성과 평가의 방식과 자료를 개선해서 활용하는 것이 더 적절할 수도 있다.

출산력 패널 자료에 임신·출산에 관한 개인의 인식과 태도, 의향 등을 포함해야 하는지는 의문스럽다. 임신과 출산, 결혼에 대한 개인의 가치관

52) 예를 들어, 제3차 저출산·고령사회기본계획의 추진 전략 체계, 대한민국정부(2015), p. 41.

과 태도, 사회적 인식과 관습 등이 출산력에 중요한 영향을 미치는 배경 요인인 것은 분명하지만, 이를 정확하게 측정하고 분석할 수 있는 독립된 변수로 식별하는 것은 매우 어려운 일이기 때문이다.

인식과 의향을 직접 측정해서 변수화하거나 간접적인 조사 문항을 통해 지수화하는 것이 일반적인 방법이지만, 측정 과정에서의 통계적 잡음(noise), 변수로서의 안정성, 지수화의 임의성 등으로 인해 이들로부터 얻어지는 정보의 신뢰성은 부족하다. 설령 인식과 태도를 적절히 포착할 수 있다고 하더라도 이에 관한 변수가 실제 임신·출산 행위를 설명하는 경로와 과정이 식별할 수 있고 구조적 관계로 설명되는지도 확실치 않다. 직접 관측이 불가능한 인식, 의도 같은 부분은 가능하면 관측이 가능한 '사실로 구현되는 선호와 지향(revealed preference)'을 통해 간접적으로 파악하는 것이 그나마 신뢰성 있는 통계적 추론과 함의를 보장한다는 점 또한 잘 알려져 있다.

출산력 패널에서 인식과 의향, 가치관 등 관측 가능하지 않은 요소들을 변수화하는 것은 분석적 가치를 위해 지양해야 한다. 같은 맥락에서, 대면·설문 조사와 같은 직접적 자료 추출 방식보다는 가능하면 행동과 동태의 결과를 추적하는 간접적 자료 관측 방식을 최대한 활용하는 것이 자료의 통계적 특성과 신뢰성을 위해서 더 바람직하다. 하지만 미시·개인 수준의 자료에서 직접 자료 추출 방식이 불가능한 부분은 대면·설문 조사가 불가피하다.<sup>53)</sup>

53) 정책적 목적에서 인식과 의도, 가치관에 대한 주기적인 관찰은 필요할 수 있다. 이를 포함한 인식 및 의향에 관한 조사는 이미 존재하는 별도의 인식·태도 조사로 소화할 수 있고, 굳이 패널 자료에 통합하지 않는 것이 간명하다. 앞서 논의한 3년 주기의 출산력 패널이라면, 패널 관측 주기 사이에 인식 및 의향 조사를 부가적으로 수행하는 것은 패널 표본 관리 차원에서 유용할 수 있다. 다만, 표본의 모집단 대표성을 감안할 때 패널 표본에 대한 부가조사가 합목적적인지는 불분명하다.

## 제2절 출산력 패널의 활용도와 가치

출산력 패널의 1차적 활용 목표는 출산력 변동과 그 동인들 간의 상호 작용을 미시-개인 수준에서 식별할 수 있는 기반 자료 역할에 있다. 제4장 제1-1절에서 논의한 바와 같이, 고정된 표본(가임여성)의 장기 종단 추적이라는 자료 특성상 출산력 패널은 인구 관련 지표 산출과 인구통계량 추정의 기초 자료로 쓰기에 최적은 아니다. 예를 들어 출산력 패널 자료로 추정된 합계출산율이 출산력조사나 인구주택총조사 자료를 기초로 추정한 합계출산율보다 더 나올 수는 없다. 앞(제4장 제1절)에서 제안한 불균등패널 형식을 취하더라도 이로부터 산출·추정한 통계량이 대규모 횡단 자료나 전수 자료에 기초한 통계량보다 우월할 수는 없다. 다만, 다른 자료로부터 산출된 임신·출산 지표와 통계량과 비교하여 검증할 대상(benchmark)으로 활용할 수는 있을 것이다.

출산력 패널을 활용해 출산력 변동을 추적하고 이를 결정하는 배경 요인과의 인과관계를 분석하는 전형적인 방법은 패널회귀분석(panel regression analysis)이다. 특히 패널고정효과모형(panel fixed effect model)과 확률효과모형(random effect model)은 횡단과 종단에 걸쳐 다양한 기준으로 가임여성 모집단을 분류하고 그에 따른 계층별 특성이 임신·출산의 변동에 미치는 영향을 분석할 수 있는 기본적 틀이 된다. 이는 인구학적 요인분해와 표준화를 통한 구성효과 분석을 다차원으로 확장한 것이다(제2장 제3절 참조).

출산력 패널이 제언대로 장기에 걸쳐 구축, 유지되어 종단 특성을 충분히 활용하게 되면 기본적인 정책실험과 평가를 개선하고 더 풍부하게 만들 수 있다.<sup>54)</sup> 이는 패널 자료 구축의 개념적 근거 중 하나가 시차를 바탕

으로 한 인과관계 분석과 이를 응용한 정책실험이라는 점에서 자연스럽게 예상된다. 패널 기반 준(準)실험모형(quasi-experimental design)을 적용한 정책실험과 평가도 더 용이해진다. 출산력 패널은 단순한 선형 이중차분법 뿐만 아니라 보다 확장된 최근의 준실험방법론들—회귀단절모형(RDD), 회귀굴곡모형(RKD)—을 직접 적용할 수 있는 자료라는 이 점을 지닌다(p. 39, 각주 13 참조). 일반적 패널과 비교해서 훨씬 장기에 걸쳐 관측, 수집된 자료를 활용하면 기존의 분석 결과가 단기 한계효과(marginal effect) 또는 한계탄력성(marginal elasticity)으로 제약되는 한계를 넘어 장기 탄력성과 중장기 효과에 대한 함의를 이끌어 내는 것도 가능해진다.<sup>55)</sup>

출산력을 결정하는 미시-구조적 관계와 역학을 모형화하는 가장 일반적인 방법론은 임신·출산에 관한 동태구조모형(dynamic programming model)이다. 개인(가임여성)의 선호와 자발적 선택의 결과로 나타나는 인구행태(임신·출산)는 이를 둘러싼 사회경제적, 제도적, 문화적, 물리적 환경요인(state variables)에 의해 주어지는 여건하에서 행위 주체(여성)의 의사결정에 의해 교육, 취업, 노동, 결혼, 주거와 같은 매개요인들을 어떻게 결정(choice variables)하는지에 따라 동태적, 확률적 과정을 거쳐 나타나는 인구적 결과(dependent demographic variables)로 재해석할 수 있다. 이런 개념을 비선형, 확률 미시모형으로 구체화한 것이 동태구조모형인데, 위에서 제안한 출산력 패널은 개념적으로 여기에 정확하게 부합한다(제4장 제1-2절, pp. 107-112 참조).

54) 생존모형이나 사건사분석류의 단순한 평가 모형에 시차를 반영한 조건부 확률(conditional probability)을 적극적으로 도입할 수 있다. 이를 일반화하면 성향점수(PS: propensity score) 기반 평가방법론으로의 연계도 가능할 것이다.

55) 제4장 제2-2-가절, pp. 117-118, 손호성(2017)의 회귀굴곡모형과 실증분석에 대한 평가 참조.

동태행위모형의 두 축을 이루는 동태적 수리해(numerical dynamic solution) 계산과, 모형의 구조 모수(structural parameters)에 대한 추정 및 미세조정(estimation and calibration)에 필요한 전형적 표본 형태는 패널이다(Keane and Wolpin, 1997; Wolpin, 2007). 인구구조 변화에 따른 사회보장재정 전망과 분석에 동태행위모형을 활용한 원종욱, 이연희(2018)에서는 고령화연구패널이, 기혼 여성의 출산과 노동공급의 관계를 살피는 데 이 모형을 적용한 조운영(2006)에서는 노동패널이 활용되었다. 하지만 인구학적 함의가 큰 이들 연구 모두 인구학적 패널을 활용하지 못한 데 따른 분석의 한계는 분명하다. 출산 지연, 출산 진도와 같은 주요 인구 요인이 반영되지 못하거나(조운영, 2006), 패널이 허용하는 짧은 기간의 연간 자료에만 의존할 수밖에 없는 제약—원종욱, 이연희(2018)의 경우 2006-2012년, 조운영(2006)의 경우 1998-2003년—이 그 예다. 출산력 패널이 구축되면 출산력 분석을 위한 보다 개선된 동태행위모형의 개발과 활용이 가능해질 것이다. 동태행위모형에 기반한 출산 관련 정책 분석·평가 방법론 개발도 앞당길 수 있게 된다.

### 제3절 정책 과제

출산력 패널을 구축하고 자료를 축적하게 되면 패널회귀분석이나 패널 기반 준실험모형을 활용한 출산 관련 정책 평가에 1차적으로 활용될 수 있다. 제4장을 중심으로 살펴본 여러 관련 경험연구들은 기존의 인구 관련 자료들을 대안으로 활용하여 평가 방법론을 적용한 경우가 많은데, 출산력 패널을 활용하게 되면 이러한 기존 실증 분석 및 정책 평가 결과들을 재평가할 수 있는 비교 검증 대상(benchmark) 역할을 할 수 있을 것

으로 기대된다. 결혼과 출산에 대한 정책 분석·평가 연구 중 미시 수준의 모형을 채택한 많은 경우 적합한 출산력 패널이 부재하다 보니 자료를 따로 구축하거나 노동패널과 같은 인접 패널을 임시방편으로 대응하는 모습이 흔히 관찰되는데, 출산력 패널은 이러한 정책 평가용 자료에 대한 갈증을 해소하는 역할을 할 것이다.

이 보고서에서 다룬 인구학적 패널의 특성과 목적, 활용 가치에 대한 논의는 기존 정책 관련 실증분석 방법론에 대한 새로운 검토 도구로 활용될 수 있을 것이다. 예를 들어 출산 관련 정책의 효과성 평가에서 합계출산율을 종속변수로 정책효과를 검증하는 모형들 가운데 거시-집계 수준의 변수(유배우출산율)의 설명력을 보는 경우와 미시 수준의 정책 변수(보육 지원 정책)의 영향을 보는 것(이상협 외, 2016, 3장, 4장)이 적합성과 분석 타당성에서 어떻게 다른지를 이 보고서의 논의를 기초로 판단할 수 있다.

출산력 패널이 실제 구축되고 자료가 축적되면, 그 과정에서 기존의 출산력조사 어떤 부분을 보완하거나 대체할 수 있을지 논의를 시작할 수 있다. 출산력조사의 향후 개선, 발전 전망과 방안 마련에서도 출산력 패널이 주요한 비교 대상이나 대안으로 고려될 수 있을 것이다.

현재의 저출산·고령사회기본계획의 구조와 틀 안에서 (저)출산 관련 정책에 대한 인구영향평가제도 및 방법론에 대한 논의는 지속되어 왔는데 성과는 거의 부재인 상태이다. 평가에 유효한 임신·출산 또는 출산력 관련 자료의 부재가 그 주된 이유 중 하나다. 출산 관련 정책의 상당 부분이 개인(여성)이 정책 대상이자 단위인 미시 성격의 정책이라는 점에서 출산

력 패널은 부족한 정책 인프라의 핵심적인 부분을 메워 주는 도구가 될 것이다.

더불어 이 보고서의 제언은 새로운 정책 평가 방법론 구축 과정의 일부로 활용될 수 있다. 특히, 인구학적 패널은 정책 분석·평가에 활용할 수 있는 미시 계량 모형의 개선과 개발에 기초가 될 수 있다.



### <국문 자료>

구성렬(1996). 인구경제론, 박영사.

권태환·김두섭(2007). 인구의 이해, 서울대학교출판부.

김종훈·이소영·이윤경·이지혜·김세진·변수정·조성호·백혜연·이상립(2018). 2  
제3차 저출산·고령사회 기본계획 재구조화 연구, 저출산고령사회위원회·  
한국보건사회연구원, 정책보고서.

박경숙(2017). 인구학 방법: 인구동태의 측정과 모형, 서울대학교출판문화원.

손호성(2017). 대학 교육이 출산에 미치는 인과적 효과 및 메커니즘에 대한 분  
석, 한국보건사회연구원, 미발간 연구보고서.

오영희·김경래·신창우·배혜원(2016). 출산력 조사(1974~2012)를 활용한 한국  
의 출산력 변천과정 연구, 한국보건사회연구원, 연구보고서.

우해봉·장인수(2018). 생존모형을 활용한 한국의 출산력 변동 분석, 통계연구,  
23(2), pp.1-26.

원종욱·이연희(2018). 동태행위모형을 이용한 사회보장제도 정책평가모형 개발  
연구, 한국보건사회연구원, 연구보고서.

이삼식·정경희·신인철·김정연(2006). 인구패널 구축을 위한 기초연구, 한국보건  
사회연구원, 연구보고서.

이삼식·최효진·계봉오·김경근·김동식·서문희·윤자영·이상협·이윤석·천현숙  
(2016). 결혼·출산 행태 변화와 저출산 대책의 패러다임 전환, 한국보건사  
회연구원, 연구보고서.

이상협·이철희·홍석철(2016). 저출산대책의 효과성 평가, 한국보건사회연구원-  
서울대학교 경제추적연구소, 연구보고서.

이소영·김은정·박종서·변수정·오미애·이상립·이지혜(2018). 2018년 전국 출  
산력 및 가족보건·복지 실태조사, 한국보건사회연구원, 연구보고서 .

이희길(2007). 국내 패널 조사의 현황 분석, 통계개발원.

- 정미옥·최필근(2014). 사회조사 자료연계 방법 연구, 통계개발원, 2014년 하반기 연구보고서III, pp.34~72.
- 정미옥(2015). 사회조사와 인구·주택총조사 표본자료의 연계방안 연구, 통계개발원.
- 정미옥·백지선·김현식·최은영(2016). 사회조사와 인총 표본조사 연계자료 분석 및 활용, 통계개발원, 2016년 상반기 연구보고서, pp.65~145.

〈영문 자료〉

- Arellano, M. (2003). *Paanel Data Econometrics*. Oxford: Oxford University Press.
- Arellano, M. and Honoré, B. (2001). Panel Data Models: Some Recent Developments, in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland, ch 53, pp.3229~3296.
- Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87(1), pp.115-143.
- Chamberlain, G. (1984). Panel Data, in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: Elsevier Science B.V., pp.1247~1318.
- Choe, M. K. and Retherford, R. D. (2009). The Contribution of Education to South Korea's Fertility Decline to "Lowest-Low" Level, *Asian Population Studies*, 5(3), pp.267-288.
- Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Esping-Anderson, G. and Billari, F. (2015). Re-Theorizing Family Demographics, *Population and Development Review*, 41(1), pp.1~31.

- Gauthier, A. (2007). The Impact of Family Policies on Fertility in Industrialized Countries: A Review of the Literature, *Population Research and Policy Review*, 26(3), pp.323~346.
- Geweke, J. and Keane, M. P. (2000). An Empirical Analysis of Earnings Dynamics among Men in the PSID: 1968-1989, *Journal of Econometrics*, 96(2), pp.293-356.
- Hahn, J. and Kuersteiner, G. (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both  $n$  and  $T$  Are Large, *Econometrica*, 70(4), pp.1639-1657.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hsiao, C. (2003). *Analysis of Panel Data*. 2nd ed. NY: Cambridge University Press.
- Keane, M. and Wolpin, K. I. (1997). The Career Decisions of Young Men, *Journal of Political Economy*, 105, No. 3, pp.473~522.
- Mundlak, Y. (1961). Empirical Production Function Free of Management Bias, *Journal of Farm Economics*, 43, No. 1, pp.44~56.
- Phillips, P. C. B. and Moon, H. R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67(5), pp.1057-1111.
- Stock, J. H. and Watson, M. W. (2003). *Introduction to Econometrics*. NY: Addison Wesley.
- Wolpin, K. I. (2007). Ex Ante Policy Evaluation, Structural Estimation, and Model Selection, *American Economic Review*, 97, No. 2, pp.48~52.
- Yoo, S. H. (2014). Educational Differentials in Cohort Fertility during the Fertility Transition in South Korea, *Demographic Research*, 30(53), 1463-1494.

〈참고 자료〉

- 김승권·박종서·김유경·김연우·최영준·손창균·윤아름(2012). 2012년 전국 결혼 및 출산 동향 조사, 한국보건사회연구원, 정책보고서 2012-88.
- 장진희(2017). 서울시 여성가족패널 구축방안 연구, 서울시여성가족재단.
- 조운영(2006). 기혼여성의 출산과 노동공급: 생애주기모형, 한국개발연구원, 정책 연구시리즈.
- 보건복지부·한국보건사회연구원(2009). 2009년도 전국 결혼 및 출산 동향 조사.
- 통계청(2018a). “2018 인구·주택총조사” 보도자료, 통계청.
- 한국보건사회연구원·저출산고령사회위원회(2005). 2005년 전국 결혼 및 출산 동향조사.

〈정부 및 연구기관 발행 자료〉

- 대한민국정부(2005). 제1차 저출산고령사회 기본계획.
- 대한민국정부(2015). 제3차 저출산고령사회 기본계획.
- 보건복지가족부·한국보건사회연구원(2008). 인구패널 및 DB구축을 위한 예비 조사, 한국보건사회연구원.
- 통계청(2018b). 내부 회의 자료, 통계청.
- 한국노동연구원(2018). 한국노동패널 1~20차년도 조사자료 User's Guide, 한국노동연구원.
- 한국보건사회연구원(1991). 인구정책 30년, 한국보건사회연구원.
- 한국보건사회연구원·서울대학교사회복지연구소(2019). 한국복지패널 13차년도 조사자료 User's Guide, 한국보건사회연구원·서울대학교사회복지연구소.
- 한국여성정책연구원(2018). 여성가족패널 1~6차 웨이브 조사자료 User's Guide, 한국여성정책연구원.

## 〈웹 자료〉

- 대만 국립중앙연구원(2020), PSFD. <https://psfd.sinica.edu.tw/web/plan02en.htm>에서 2020.1.31. 인출.
- 중국 사회과학조사연구소(ISSS)(2020) CFPS. <https://opendata.pku.edu.cn/dataverse/CFPS>에서 2020.2.4. 인출.
- 통계청(2020). 인구주택총조사. [http://www.census.go.kr/cds/cdsCensusI.ndc.do?q\\_menu=2&q\\_sub=22](http://www.census.go.kr/cds/cdsCensusI.ndc.do?q_menu=2&q_sub=22)에서 2020. 2.3. 인출.
- 한국노동연구원(2019). 한국노동패널. <https://www.kli.re.kr/klips/content.s.do?key=137>에서 2020.2.3. 인출.
- ICPSR(2020). <https://www.icpsr.umich.edu/icpsrweb/NACDA/studies/22100/datadocumentation>에서 2020.2.4. 인출.
- NIDI(2020). GGP. <http://www.nkps.nl/dptj> 2020.1.31. 인출.
- UNECE(2020). Generation and Gender Survey. <http://www.ggp-i.org/data/dptj>에서 2020.1.31. 인출.
- ISER(2020). BHPS. <https://www.iser.essex.ac.uk/bhps/about/questionaire-content>에서 2020.2.4. 인출.



## 간행물 회원제 안내

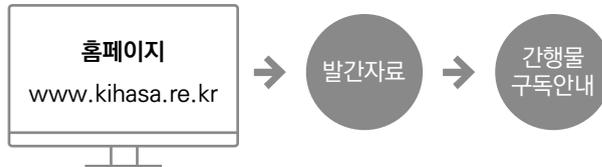
### 회원제에 대한 특전

- 본 연구원이 발행하는 판매용 보고서는 물론 「보건복지포럼」, 「보건사회연구」도 무료로 받아보실 수 있으며 일반 서점에서 구입할 수 없는 비매용 간행물은 실비로 제공합니다.
- 가입기간 중 회비가 인상되는 경우라도 추가 부담이 없습니다.

### 회원 종류

전체 간행물 회원 120,000원	보건 분야 간행물 회원 75,000원
사회 분야 간행물 회원 75,000원	정기 간행물 회원 35,000원

### 가입방법



### 문의처

- (30147) 세종특별자치시 시청대로 370 세종국책연구단지  
사회정책동 1~5F  
간행물 담당자 (Tel: 044-287-8157)

## KIHASA 도서 판매처

- 한국경제서적(총판) 02-737-7498
- 영풍문고(종로점) 02-399-5600
- Yes24 <http://www.yes24.com>
- 교보문고(광화문점) 1544-1900
- 알라딘 <http://www.aladdin.co.kr>