연구보고서 2025-17

장래인구추계 방법론 개선 연구
: 출산력과 사망력 중심으로

송창길
류재린·오진호·손현섭·오다은

사람을
생각하는
사람들

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

연구진

연구책임자	**송창길**	한국보건사회연구원 부연구위원
공동연구진	**류재린**	한국보건사회연구원 부연구위원
	오진호	국립한밭대학교 교수
	손현섭	국민연금연구원 부연구위원
	오다은	한국보건사회연구원 전문연구원

연구보고서 2025-17

장래인구추계 방법론 개선 연구
: 출산력과 사망력 중심으로

발 행 일	2025년 12월
발 행 인	신 영 석
발 행 처	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	1999년 4월 27일(제2015-000007호)
인 쇄 처	㈜에이치에이엔컴퍼니

ⓒ 한국보건사회연구원 2025

ISBN 979-11-7252-092-2 [93510]
https://doi.org/10.23060/kihasa.a.2025.17

발|간|사

인구추계는 국가의 중장기 정책 수립에 핵심적인 기초자료를 제공하는 도구로서 정책 연구에서 매우 큰 가치를 지닌다. 특히 초저출산과 초고령 사회 진입이라는 전례 없는 인구변동 국면에서, 정책 변수를 명시적으로 반영할 수 있는 독자적인 인구추계 체계의 필요성이 더욱 높아지고 있다.

본 연구는 출산진도비(PPR) 방법을 도입하여 초혼율을 정책 변수로 명시화하고, 사망력 추계에서 생명표를 130세까지 확장하여 초고령 인구 분석의 정밀성을 높였다. 또한 혼인상태 전이에 기반한 인구-가구 통합 추계 모형을 구축하고, 마이크로시뮬레이션과의 연계 체계를 마련함으로써 다양한 정책 시나리오에 탄력적으로 대응할 수 있는 환경을 조성하였다는 점에서 의의가 있다.

이 연구는 송창길 부연구위원 책임하에, 국립한밭대학교 오진호 교수가 출산·사망력 전망의 이론적 배경을, 국민연금연구원의 손현섭 박사가 생명표 연장을, 본 연구원의 류재린 연구위원이 마이크로시뮬레이션 연계를, 오다은 연구원이 연구 전반을 지원하였다. 또한, 국민연금연구원의 김형수 박사와 한양대학교 이가은 교수의 자문으로 연구의 질을 한층 높일 수 있었다. 연구를 수행하는 과정에서 조언을 아끼지 않은 재정통계연구실, 인구정책실 박사님들께도 감사 인사를 드린다. 마지막으로 이 보고서의 내용은 한국보건사회연구원의 공식 견해가 아님을 밝힌다.

2025년 12월

한국보건사회연구원 원장

신 영 석

목 차

요약 ····· 1

제1장 서론 ····· 5

제1절 연구 배경 및 목적 ····· 7

제2절 연구 범위 및 구성 ····· 10

제2장 인구추계 방법론 현황 및 개선 방향 ····· 13

제1절 기존 인구추계 방법론 검토 ····· 15

제2절 인구·가구 추계의 통합적 접근 ····· 50

제3절 인구추계와 정책시뮬레이션 연계 체계 ····· 64

제3장 출산력 추계 방법론 개선 ····· 71

제1절 기존 출산력 추계 방법론 검토 ····· 73

제2절 출산력 추계 방법론 개선 방안 ····· 93

제3절 출산력 추계모형 구축 및 결과 검토 ····· 99

제4장 사망력 추계 방법론 개선 ····· 119

제1절 기존 사망력 추계 방법론 검토 ····· 121

제2절 사망력 추계 개선 방안 ····· 151

제3절 사망력 추계 개선 구현 및 결과 검토 ····· 157

제5장 통합 추계 체계 구축 및 활용 방안 ····· 171

제1절 인구·가구 통합 추계 모형 구조 및 설계 ····· 173

제2절 통합 추계 모형 결과 검토 ····· 185
제3절 KIHASA-SIM 연계 시뮬레이션 체계 설계 ····· 207

제6장 결론 ····· 237
제1절 주요 연구 결과 ····· 239
제2절 결론 및 향후 과제 ····· 243

참고문헌 ····· 247

Abstract ····· 257

표 목차

〈표 2-1〉 미국 연령별 출산율 6개 그룹 ····· 22
〈표 2-2〉 재현자료 생성 소프트웨어 ····· 66
〈표 3-1〉 2023~2072년 합계출산율 ····· 77
〈표 3-2〉 모의 연령별 출산율과 코호트 출산율 비교 ····· 80
〈표 3-3〉 출산율 예측모형 ····· 82
〈표 3-4〉 출산력 추계 방법론 구조 ····· 100
〈표 3-5〉 MaritalTransitionCalculator pseudo 코드 ····· 104
〈표 3-6〉 FertilityCalculator pseudo 코드 ····· 107
〈표 3-7〉 혼인상태별 이동 추정 결과 비교 ····· 112
〈표 3-8〉 출산진도비 추정 결과 ····· 114
〈표 3-9〉 합계출산율 비교 결과 ····· 116
〈표 4-1〉 통계청 사망력 가정별 기대수명 ····· 122
〈표 4-2〉 신, 구 생명표 개선사항 ····· 139
〈표 4-3〉 5세 이하의 ${}_{n}a_{x}$ 값 ····· 140
〈표 4-4〉 LifeTableCalculator pseudo 코드 ····· 158
〈표 4-5〉 Coale-Kisker 기반 한계연령 상향 조정 방법 pseudo 코드 ····· 160
〈표 4-6〉 Coherent Kannisto 기반 한계연령 상향 조정 방법 pseudo 코드 ····· 162
〈표 4-7〉 상한연령 상향 설정에 따른 생명표(Coale-Kisker vs Coherent Kannisto) ··· 165
〈표 4-8〉 사망력 가정별 기대수명 ····· 166
〈표 5-1〉 PopulationCalculator 주요 파라미터 ····· 181
〈표 5-2〉 인구추계 모형 실행 방법 ····· 184
〈표 5-3〉 출산율 시나리오별 주요 가정 ····· 187
〈표 5-4〉 총인구 추계 결과 ····· 191
〈표 5-5〉 인구성장률 및 성비 추계 결과 ····· 192
〈표 5-6〉 연령대별 인구 구성 추계 결과 ····· 193
〈표 5-7〉 가임기 여성의 혼인상태별 구성비 추계 결과 – 초혼율 유지 ····· 195
〈표 5-8〉 가임기 여성의 혼인상태별 구성비 추계 결과 – 초혼율 '+1.5%/연' ····· 196

〈표 5-9〉 가임기 여성의 혼인상태별 구성비 추계 결과 - 초혼율 '+3.0%/연' ····· 197
〈표 5-10〉 출생아수 및 합계출산율 추계 결과 ····· 198
〈표 5-11〉 연간 사망자수 추계 결과 ····· 199
〈표 5-12〉 기대수명 추계 결과 ····· 200
〈표 5-13〉 인구 추계 주요 가정 차이 ····· 202
〈표 5-14〉 총인구 추계 결과 비교 ····· 203
〈표 5-15〉 연령대별 인구 구성 추계 결과 비교 ····· 204
〈표 5-16〉 출생아수 추계 결과 비교 ····· 205
〈표 5-17〉 합계출산율 추계 결과 비교 ····· 205
〈표 5-18〉 기대수명 추계 결과 비교 ····· 206
〈표 5-19〉 2012~2022년 연령별 장애유병률 ····· 214
〈표 5-19〉 2012~2022년 연령별 장애유병률 〈계속〉 ····· 216
〈표 5-20〉 2012~2022년 연령별 건강수명 ····· 217
〈표 5-20〉 2012~2022년 연령별 건강수명 〈계속〉 ····· 218
〈표 5-20〉 2012~2022년 연령별 건강수명 〈계속〉 ····· 219
〈표 5-21〉 건강수명 추계 pseudo-code ····· 221

그림 목차

[그림 2-1] 장래 가정에 따른 인구추계 분류 ····· 18
[그림 2-2] 일본 장래인구추계 ····· 20
[그림 2-3] 미국 장래인구추계 주요 결과 ····· 24
[그림 2-4] 호주 장래인구추계와 시나리오 ····· 28
[그림 2-5] 캐나다 장래인구추계와 시나리오 ····· 32
[그림 2-6] 싱가포르 인구 2023 trend ····· 35
[그림 2-7] 싱가포르 장래인구추계 주요 결과 ····· 36
[그림 2-8] 대만 인구 추이 ····· 38
[그림 2-9] 대만 장래인구추계 주요 결과 ····· 39
[그림 2-10] 홍콩 인구 2020~2023 ····· 41
[그림 2-11] 홍콩 장래인구추계 주요 결과 ····· 42
[그림 2-12] 통계청(2024) 인구추계 과정 ····· 44
[그림 3-1] 출생 단기 및 장기 추계방법 ····· 73
[그림 3-2] 1970~2072년 합계출산율 추이 ····· 78
[그림 3-3] 2000~2072년 연령별 출산율 ····· 78
[그림 3-4] TFR 전이 3단계와 감소함수 궤적 ····· 89
[그림 4-1] 통계청 사망률 모형 변천 ····· 124
[그림 4-2] LLG 확장모형의 $B_{x,t}$ 변화 ····· 130
[그림 4-3] 성별 기대수명과 50세 기대여명 추이 ····· 137
[그림 4-4] UN-IGME의 5세 미만 사망력 ····· 140
[그림 4-5] 한계연령 상향 성별 사망확률(Coale-Kisker) ····· 163
[그림 4-6] 한계연령 상향 성별 사망확률(Coherent Kannisto) ····· 163
[그림 4-7] 100세 이상 인구 수(Coherent Kannisto) ····· 168
[그림 5-1] 인구피라미드 비교((좌) 통계청, (우) 본 연구: 2080년) ····· 201
[그림 5-2] 통합 인구추계 모형과 KIHASA SIM의 연계(안) ····· 209

요약

1. 연구의 배경 및 목적

한국은 2023년 기준 합계출산율 0.72명이라는 세계 최저 수준의 초저출산 현상과 급속한 고령화를 동시에 경험하고 있다. 이러한 인구 구조의 급격한 변화는 노동력 감소, 사회보장 재정 부담 증가, 지역 소멸 위기 등 다양한 사회경제적 문제를 초래하고 있어, 정밀하고 정책 연계성이 높은 인구추계 체계의 구축이 시급하다. 그러나 기존 인구추계 방법론은 여러 한계를 내포하고 있다. GLG 분포 등 모수 분포 기반 출산력 추계는 정책 경로가 불명확하고, 생명표 한계연령 100세는 초고령층 분석을 제약하며, 인구·가구 분리 추계 방식은 혼인·가구 정책과 출산력의 연계를 저해한다. 본 연구는 출산진도비(PPR) 방식 도입, 한계연령 130세 연장, 혼인상태 기반 통합추계 모형 구축, KIHASA-SIM 연계 체계 마련을 통해 정책 지향적이고 독자적인 인구추계 체계를 구축하고자 한다.

2. 주요 연구 내용

출산진도비(PPR) 방식을 도입하여 혼인상태별 인구 구성과 출산 의사결정을 명시적으로 모델링하였다. 통합형 PPR을 적용하는 PPR_SIMPLE 방식을 채택하고, 초혼율을 핵심 정책 변수로 설정하여 초혼율 유지, '+1.5%/연', '+3.0%/연'의 3개 시나리오를 구성하였다. 2023~2055년 PPR 기반 추계 후 역산된 연령별 출산율(ASFR)을 저장하여 재현 가능성을 확보하였으며, 이를 통해 혼인 지원 정책이 유배우 인구 증가를 거쳐 출산율로 연결되는 전체 경로를 추적할 수 있게 되었다.

생명표 한계연령을 기존 100세에서 130세로 연장하기 위해 Coale-Kisker

방법과 Coherent Kannisto 방법을 모두 구축하였다. 본 연구에서 구축한 두 방법론은 기존 Lee-Carter 계열 모형의 추계 결과와 연계하여 사용할 수 있도록 설계되었다. 이를 통해 사망력 추계 모형의 선택과 관계없이 한계연령 연장 방법론을 일관되게 적용할 수 있는 유연한 체계를 마련하였다. 한계연령을 130세로 연장한 결과, 0세 기대수명 증가 효과는 0.3~0.4세 수준으로 나타났으며, 이는 100세 이상 초고령층 인구를 연령별로 세분화하여 추적함으로써 향후 기대수명 지속 연장에 대비하고 요양·의료·돌봄 등 초고령층 대상 정책 수립에 필요한 정밀한 정보를 제공할 수 있게 되었음을 의미한다.

또한 3계층 데이터 구조(actuals/assumptions/projections)와 6개 핵심 모듈로 통합 추계 시스템을 구축하고, ProjectionPipeline 클래스를 통해 5단계 추계 프로세스를 통합 관리하였다. 2022년 기준 2122년까지 100년 추계 결과, 총인구는 초혼율 유지 시나리오 1,238만 명, ''+1.5%/연' 1,486만 명, '+3.0%/연' 1,803만 명으로 전망되었다. 시나리오 간 격차는 2050년 118만 명에서 2120년 565만 명으로 확대되어 초혼율 정책의 장기 누적 효과를 확인하였다.

65세 이상 인구 비율은 2023년 18.4%에서 2080년 51.4~57.4%로 급증하고, 합계출산율은 초혼율 유지 0.62, '+1.5%/연' 0.84, '+3.0%/연' 1.08로 수렴하였다. 기대수명은 2023년 83.9세에서 2090년 93.0세로 상승하며 성별 격차는 5.9세에서 2.7세로 축소되었다. 통계청 중위 시나리오 대비 본 연구의 '+1.5%/연' 시나리오는 2120년 총인구가 24.4% 낮게 전망되었으며, 주요 차이는 PPR 방식이 혼인 지연과 비혼 증가를 구조적으로 반영한 출산력 가정에서 발생하였다.

인구추계 모형의 혼인상태별 전이확률, 가구수, 출생·사망자수를 KIHASA-SIM에 외생변수로 투입하는 단방향 연계 체계를 설계하였다. Sullivan 방식 기반 건강수명 추계와 건강-고용 상태 동시 고려 건강노동

기대여명 산출 방법론을 제시하여, 향후 KIHASA-SIM 연계 시 고령화 정책의 통합적 효과 분석 기반을 마련하였다.

3. 결론 및 시사점

본 연구는 PPR 방식 도입으로 출산력 추계의 정책 연계성을 강화하고, 한계연령 130세 연장으로 초고령층 분석을 정밀화하였다. 혼인상태별 인구 전이 기반 통합 체계로 인구·가구·정책시뮬레이션의 일관성을 확보하고, 모듈화된 구조와 자동 아카이빙으로 재현 가능하고 확장 가능한 추계 시스템을 구현하였다.

초혼율 '+3.0%/연' 시나리오는 '+1.5%/연' 대비 2120년 총인구 348만 명(23.4%) 많고 합계출산율(TFR)은 0.24 높아, 혼인 지원 정책의 장기 효과를 확인하였다. 그러나 초혼율 증가가 출산으로 연결되기까지 10~20년의 시차가 존재하며, 2080년 고령 인구 비율은 시나리오 간 6.0%p 차이에 불과하여 고령화는 출산력 정책만으로 완화하기 어렵다. 따라서 저출산 정책은 장기 관점에서 지속 추진하고, 건강수명 연장과 노인부양 체계 개선 등 고령사회 적응 정책에 집중하는 것이 현실적이다.

방법론 측면에서 PPR 분리형 확장, 혼인 매칭 이론 도입, 혼인상태별 사망률 차이 반영이 필요하다. 데이터 측면에서 건강-고용 통합 패널 구축, 지표 표준화, 지역별 세분 데이터 확보가 요구된다. 시스템 측면에서 시·도별 추계 확장, 확률적 추계 도입, KIHASA-SIM 연계 강화가 필요하며, 정책 활용 측면에서 다양한 정책 시나리오 시뮬레이션과 불확실성 정량화를 통한 정책 리스크 관리 지원이 필요하다.

주요 용어: 인구추계, 마이크로시뮬레이션, 출산력, 출산진도비, 사망력, 한계연령

제 1 장

서론

제1절 연구 배경 및 목적
제2절 연구 범위 및 구성

제 1 장 서론

제1절 연구 배경 및 목적

1. 연구 배경

한국사회는 전례 없는 인구변동의 새로운 국면에 진입하고 있다. 2000년대 이후 지속되고 있는 초저출산과 급속한 고령화는 기존의 인구추계 모형이 상정하지 못했던 극단적 상황을 연출하고 있다. 2023년 합계출산율이 0.72명으로 역대 최저치를 기록하며, OECD 국가 중 유일하게 0점대 출산율을 지속하고 있는 상황이다. 동시에 65세 이상 고령인구 비율이 2025년 20%를 넘어서며 초고령사회로의 진입을 앞두고 있어, 인구구조의 급격한 변화가 예상된다.

이러한 극단적 인구변동은 기존 추계 모형의 예측력 한계를 드러내고 있다. 특히 코로나19 팬데믹 이후 건강에 대한 인식 변화, 결혼과 출산에 대한 가치관의 급변, 경제적 불확실성 증가 등 다양한 요인들이 복합적으로 작용하면서 인구변동 요인의 복잡성이 크게 증가하였다. 기존의 모수 분포 기반 추계 방식으로는 이러한 구조적 변화와 예측 불가능한 요소들을 충분히 반영하기 어려운 상황이다.

현재 통계청에서 5년 주기로 발표하는 장래인구추계는 100세로 제한된 한계연령, 모수 분포 중심의 출산력 추계 방식, 그리고 인구와 가구추계의 분리된 체계라는 구조적 한계를 가지고 있다.

먼저 한계연령 100세 설정은 기대수명이 지속적으로 연장되고 있는 현실을 충분히 반영하지 못하고 있다. 이는 연금추계를 비롯한 주요 정책

분석에서 요구되는 장기 전망의 정확성을 저해하는 요인으로 작용하고 있다. 또한 출산력 전망에서도 Generalized Log Gamma(GLG) 분포 등 모수 함수 기반의 연령별 출산율(ASFR) 추정 방식은 혼인·가구 구성과의 연계가 약하고, 정책 시나리오 설정이 자의적일 수 있다는 한계가 있다.

특히 인구추계와 가구추계가 분리된 현행 체계에서는 혼인상태 변화가 출산력과 가구 형성에 미치는 상호작용을 충분히 고려하기 어렵다. 미혼율 증가, 만혼 및 비혼 증가 등 혼인 패턴의 급격한 변화가 출산력에 직접적 영향을 미치고 있음에도 불구하고, 이러한 연관성을 체계적으로 분석할 수 있는 방법론적 기반이 부족한 실정이다.

인구구조 변화가 사회보장제도 전반에 미치는 영향이 심화되면서, 정책 개발과 사회보장제도 설계를 위한 정교한 인구추계 도구에 대한 수요가 증가하고 있다. 연금, 의료, 복지 등 각 영역별로 인구변동이 재정에 미치는 영향을 정확히 분석하고, 다양한 정책 시나리오에 따른 효과를 사전에 검토할 수 있는 맞춤형 분석 도구가 필요한 상황이다.

또한 건강수명과 노동수명의 중요성이 커지면서, 단순한 생존 기반의 인구추계를 넘어 건강상태와 경제활동 능력을 고려한 확장된 수명 지표에 대한 정책적 관심이 높아지고 있다. 이는 고령사회 정책 설계에서 핵심적인 기초 정보로 활용될 수 있으며, 노동력 전망, 의료비 추계, 돌봄 수요 예측 등 다양한 영역에서 정책 근거로 활용 가능하다.

2. 연구 목적

본 연구는 기존 출산력과 사망력 추계 방법론의 개선을 통해 급변하는 인구환경에 대응할 수 있는 한국보건사회연구원의 인구추계 체계를 구축

하는 것을 목적으로 한다.

첫째, 출산력 추계에서는 출산진도비(Parity Progression Ratio, PPR) 방식을 도입하여 혼인상태별 인구 구성과 출산 의사결정을 명시적으로 연계한다. 이를 통해 초혼율 증가와 같은 정책 변수가 출산으로 연결되는 경로를 추적하고, 저출산 대응 정책의 효과를 정량적으로 평가할 수 있는 분석 틀을 제공한다.

둘째, 사망력 추계에서는 한계연령을 130세로 확장하고 생명표 연장 방법론(Coale-Kisker, Coherent Kannisto)을 적용하여 초고령층 인구 분석의 정밀도를 높인다. 또한 Lee-Li-Gerland Extended(LLG-E) 모형을 통해 성별 간 사망률 개선의 일관성을 확보하고 불확실성을 정량화한다.

셋째, 인구·가구 통합추계 모형을 개발하여 혼인상태별 인구이동과 가구 형성의 상호작용을 체계적으로 분석할 수 있는 종합적 인구변동 분석 체계를 구축한다. 혼인상태 전이 모형을 통해 미혼, 유배우, 사별, 이혼 간의 이동을 추적하고, 이를 가구 추계와 연계한다.

넷째, 인구추계의 정책 활용성을 극대화하기 위해 마이크로시뮬레이션 모형(KIHASA-SIM)과의 연계 체계를 설계한다. 혼인상태별 전이확률, 가구주의 혼인상태별 가구수, 출생아수, 사망자수를 KIHASA-SIM에 외생변수로 투입하여 동일한 인구·가구 시나리오 하에서 노동, 소득, 연금 등의 정책 효과를 일관되게 분석할 수 있도록 한다. 또한 건강수명과 노동수명 추계 방법론을 검토하여 향후 연계 확장 방안을 제시한다.

이를 통해 한국보건사회연구원만의 독자적 인구추계 모형을 개발하여 정책연구의 자율성과 전문성을 제고하고, 정책 의사결정 지원을 위한 다양한 시나리오 분석체계를 구축하여 연구원의 정책연구 기반을 강화하고자 한다.

제2절 연구 범위 및 구성

1. 연구 범위

기준 인구는 통계청의 「장래인구추계: 20222072년」과 마찬가지로 2022년 11월 1일 인구총조사(등록센서스 방식)를 기준으로 하되, 2022년 7~10월 사이 발생한 인구변동요인을 가감한 연앙인구를 활용한다. 본 연구는 전국 단위 인구추계를 중심으로 하며, 내외국인을 포괄하는 인구를 대상으로 한다.

추계 기간은 2023년부터 2122년까지 100년으로 설정하여, 중장기 정책분석에 필요한 충분한 전망 기간을 확보한다. 이는 현재 은퇴 연령층이 완전히 생애를 마무리하는 시점까지를 포괄하여 연금 등 사회보장제도의 장기 재정분석에 활용할 수 있도록 한다.

연구의 핵심은 출산력과 사망력 추계 방법론 개선에 있으며, 이를 바탕으로 인구·가구 통합추계 모형을 구축한다. 출산력 추계에서는 PPR 방식을 적용하여 혼인상태별 인구 구성과 출산을 연계하고, 사망력 추계에서는 LLG-E 모형과 한계연령 130세 연장을 통해 초고령층 분석을 강화한다. 또한 혼인상태 전이 모형을 통해 인구추계와 가구추계의 통합적 분석이 가능하도록 설계한다.

KIHASA-SIM과의 연계를 통해 개발된 추계 모형의 정책 활용성을 검토하고, 사회보장제도 영향분석을 위한 시뮬레이션 체계를 제시한다. 또한 건강수명과 노동수명 추계 방법론을 검토하여 향후 확장 방향을 제시한다. 이를 통해 이론적 방법론 개선에서 정책 활용까지의 전 과정을 포괄하는 연구를 수행한다.

2. 연구 구성

본 연구는 총 6개 장으로 구성된다. 제2장에서는 국내외 인구추계 방법론을 검토하고, 인구·가구 통합추계의 필요성과 정책시뮬레이션 연계 방향을 제시한다. 제3장에서는 출산력 추계 방법론 개선 방안으로 출산진도비(PPR) 방식의 이론적 배경과 구현 방안을 다룬다. 제4장에서는 사망력 추계 방법론 개선과 생명표 한계연령 연장 방안을 제시한다. 제5장에서는 통합 추계 체계를 구축하고 시나리오별 추계 결과를 검토한다. 또한 KIHASA-SIM과의 연계 방안과 건강수명·노동수명 추계 방법론을 검토한다. 제6장에서는 주요 연구 결과를 종합하고 정책적 시사점 및 향후 과제를 제시한다.

제2장

인구추계 방법론 현황 및 개선 방향

제1절 기존 인구추계 방법론 검토
제2절 인구·가구 추계의 통합적 접근
제3절 인구추계와 정책시뮬레이션 연계 체계

제 2 장 인구추계 방법론 현황 및 개선 방향

제1절 기존 인구추계 방법론 검토

1. 인구추계 방법론 개요

인구추계(Population projection)는 재정과 사회보장 정책 등 국가 중장기 경제, 사회 발전계획 수립의 기반이 될 뿐만 아니라 노동, 교육, 보건, 복지, 산업, 환경, 주택 등 다양한 분야에서 광범위하게 활용되고 있다(통계청, 2024). 이런 인구추계는 인구총조사로 집계한 총조사인구(11월)에 7~10월 사이 발생한 인구변동요인을 가감하여 성, 연령별 기준 인구(base population)를 확정하고, 출생, 사망, 국제이동과 같은 인구변동요인을 분석하고 장래가정을 설정한 다음, 이 들 값을 식 (1)과 같은 인구균형방정식(demographic balancing equation)에 대입하여 코호트요인법[1](cohort component method)에 따라 미래 연도별 인구를 추계한다.

$$P_t = P_{t-1} + B_{t-1} - D_{t-1} + NM_{t-1} \tag{1}$$

$$\begin{cases} P_t : t\text{년 인구},\ B_{t-1} : t-1\text{년 출생아수} \\ D_{t-1} : t-1\text{년 사망자수},\ NM_{t-1} : t-1\text{년 국제순이동자수} \end{cases}$$

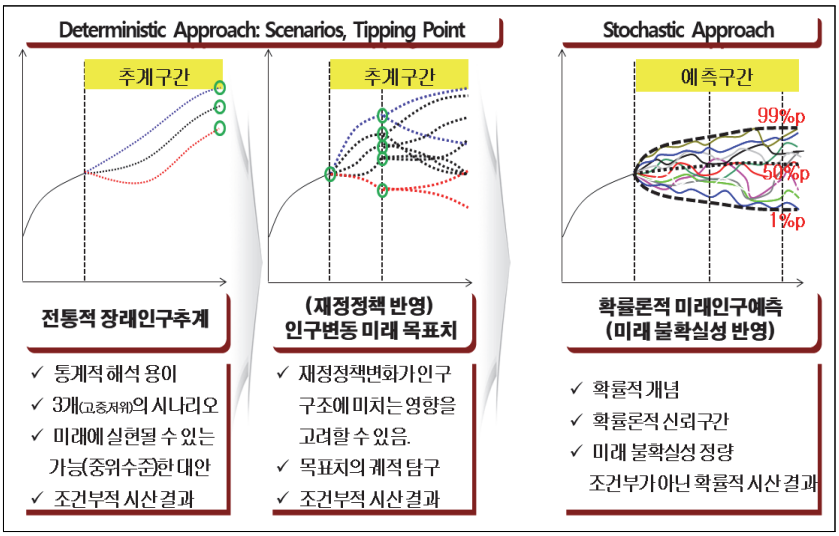
1) 인구변동요인별(출생, 사망, 국제이동) 미래 수준을 각각 예측한 후, 추계의 시작점이 되는 기준 인구에 출생아수와 국제순이동자수는 더하고, 사망자수는 빼는 인구균형방정식을 적용하여 다음해 인구를 반복적으로 산출하는 인구추계 방법.

인구변동요인 분석 가운데 출생 추이 분석은 출생아수, 출산연령, 합계출산율, 출생성비와 같은 인구동향조사 출생통계와 관련된 기간 출산 추이 분석과 출생 코호트별 완결, 미완결 추이 분석, 초혼연령, 연령별 혼인율과 특정 연령까지의 누적혼인율 그리고 출생아 모의 평균 결혼기간 등을 분석하는 혼인과 출생 추이 분석으로 이루어진다. 사망 추이분석은 사망자수, 사망자수와 주민등록연안인구로 나눈 사망률과 사망률을 생명표에 대입하여 연령별 기대여명 추이 분석으로 구성된다. 국제이동 추이 분석은 입국과 출국자수를 비교하는 순유입을 정의하고, 내국인과 외국인으로 분리하여 국제 순이동 추이를 살펴본다.

장래 인구 예측은 여러 방법이 있다. 먼저 전통적 장래인구추계 방식인 조건부 시나리오에 기반을 둔 결정적 추계(deterministic population projection)방식이다. 즉, 전통적 장래인구추계는 결정적 시나리오(scenario) 기반 장래인구추계이다. 따라서 미래의 조건부적 시산으로 실현될 수 있는 여러 가능한 대안을 제시하는 것이고, 통계적 해석이 용이하여 여러 국가 통계청에서 이 방식을 채택하고 있다. 다음으로 미래 목표값(target value)을 선정하여, 이를 도달하는 가상 궤적(trajectory)을 살펴보는 것이다. 인구변동요인의 특정 연도 목표치를 설정한 다음 목표치의 궤적을 탐구하는 것으로, 이 방식은 기획한 정책이 미래에 연착륙 될 때 인구구조가 어떻게 변할지를 조명하는 것이다. 이 방식도 재정정책에 따른 인구변화를 살펴보는 것으로 조건부적 시산 결과로 볼 수 있다. 끝으로 미래는 불확실성을 내포하고 있으므로 이를 대변하기 위해 출생, 사망, 국제이동 추이에 랜덤분포를 가산하여 다양한 미래 궤적(trajectories)을 생성하여 확률적인 관점에서 살펴보는 것으로 확률적 인구추계(stochastic population projection)이다. 인구변동요인의 미래 불확실성을 나타내기 위해 랜덤 추이를 가미하여 조건부 시산이 아닌 확률 기반 분위수(percentile) 조명 방식이다. 따라서 확률적 인구추

계는 미래 인구를 단일한 값이 아닌 확률 분포로 나타내는 방식이다. 즉, 불확실성을 반영하여 다양한 가능성을 제공하며, 결과값은 보통 분위수(percentile)로 표현된다. 예를 들어 2050년의 A국 인구를 확률적으로 추정한 결과, 10분위수(10th percentile): 2,500만 명; 25분위수(25th percentile): 3,000만 명, 50분위수(50th percentile, 중앙값): 3,500만 명, 75분위수(75th percentile): 4,000만 명; 90분위수(90th percentile): 4,500만 명이라고 하자. 여기서 중앙값(50분위수, Median: 3,500만 명)은 미래 인구가 3,500만 명 이하일 확률이 50%이고, 3,500만 명 이상일 확률도 50%이다. 즉, 가장 대표적인 예상값이다. 10분위수(2,500만 명)는 미래 인구가 2,500만 명보다 적을 확률이 10%에 불과함을 의미한다. 즉, 아주 낮은 확률로 이보다 적은 인구가 될 수도 있지만 가능성은 낮다. 90분위수(4,500만 명)는 반대로 미래 인구가 4,500만 명보다 많을 확률은 10%밖에 없다는 것을 의미한다. 즉, 인구가 4,500만 명을 초과할 가능성은 낮다. 25분위수(3,000만 명), 75분위수(4,000만 명)는 인구가 3,000만 명보다 많을 확률은 75%, 인구가 4,000만 명보다 많을 확률은 25%임을 의미한다. 결국 이를 통해 미래 인구는 대략 3,000만~4,000만 명 사이에 있을 가능성이 높다는 결론을 내릴 수 있다. 따라서 확률적 인구추계는 단순히 '2050년 인구는 00명이다'라고 단정 짓는 것이 아니라, 다양한 가능성을 수치화하여 정책 수립과 의사결정에 활용할 수 있도록 돕는 강력한 도구이다. 특히 미래 정책을 수립할 때 불확실성을 고려할 수 있다. 예를 들어, 정부가 2050년을 대비한 도시계획을 세운다면, 인구가 3,500만 명이 될 것을 기본 시나리오로 고려하지만, 10% 이하의 확률로 발생하는 2,500만 명 이하로 감소할 경우의 대비책(노령화 대응, 복지 시스템 조정)도 준비하고, 4,500만 명 이상으로 증가할 가능성(이민 증가, 출산율 반등 등)에 대비한 대책(주택, 교통, 교육 인프라 확충)도 마련할 수 있다. 〔그림 2-1〕은 이들 3가지의 인구 추이를 보여준다.

〔그림 2-1〕 장래 가정에 따른 인구추계 분류



자료: 저자 작성.

선진국 통계청(일본, 미국, 영국, 호주, 캐나다 등)과 통계청은 첫 번째 시나리오 방식을 채택하여 미래를 조명한다. 통계청(2024)은 출생, 사망, 국제이동의 장래 수준은 중위, 고위, 저위로 설정한 후 요인별 수준을 조합하고 특별 시나리오 4개(국제 무이동(zero) 시나리오, 출산율 2022년 출산율 지속 시나리오, 출산율 OECD 평균 시나리오, 사망력 코로나19 장기 영향 시나리오)를 추가하여 총 31개의 장래인구 시나리오를 작성하여 제공한다.

2. 주요국 인구추계 현황

이번 절에서는 일본, 미국, 호주, 캐나다와 같은 선진국의 장래인구추계 모형과 전망을 검토하고, 초저출산 수준(TFR 1.0명 이하)을 보이는 홍콩, 대만, 싱가포르 등의 향후 출산율 추이와 추계 방법도 소개한다.

가. 선진국 사례(일본, 미국, 호주, 캐나다)

1) 일본

먼저 일본의 경우, 후생노동성 산하 국립사회보장·인구문제연구소(National Institute of Population and Social Security Research)[2)]가 인구센서스와 추계를 담당하며 결과를 공표한다. 최근 공표된 2023년 자료에 따르면, 출산율, 사망률, 국제 이동 추세와 남녀 성비를 활용한 코호트요인법을 사용하여 2021년부터 2070년까지 성별 및 연령별 미래 인구를 산출하였다. 주요 결과에 따르면, 50년 후 일본의 총인구는 현재 수준의 약 70%로 감소할 것으로 보이며, 65세 이상 인구는 약 40%를 차지할 것으로 예상된다. 2020년 센서스 기준 총인구는 1억 2,615만 명인데, 2070년에는 8,700만 명으로 감소할 것으로 전망된다.

출산율과 관련해서는 일본이 1.0명 이하 수준인 우리나라와는 달리, 1.26명으로 보고되고 있다. 그러나 일본 정부는 이런 출산율 수준을 국가적 위기로 인식하고, 대규모 출산장려정책을 통해 출산율 반등을 도모하고 있다. 과거 2017년 발표에서는 2065년 기준 TFR이 1.44명으로 전망되었으나, 2023년 발표에서는 2070년 기준 중위 시나리오로 1.36명(고위 1.64명, 저위 1.13명)으로 하향 조정되었다. 코로나19 팬데믹 이전부터 이미 진행된 출산율 감소를 반영한 결과이며, 팬데믹 기간 동안 결혼 감소로 인해 단기적으로 출산율이 낮은 수준을 유지할 것으로 보인다.

사망률에 따른 기대수명은 2020년 남성 81.58년, 여성 87.72년에서 2070년 중위 수준 기준 남성은 85.89년, 여성은 91.94년으로 증가할 것으로 예상된다. 이는 2017년 개정에서 예측된 2065년 기대수명(남성

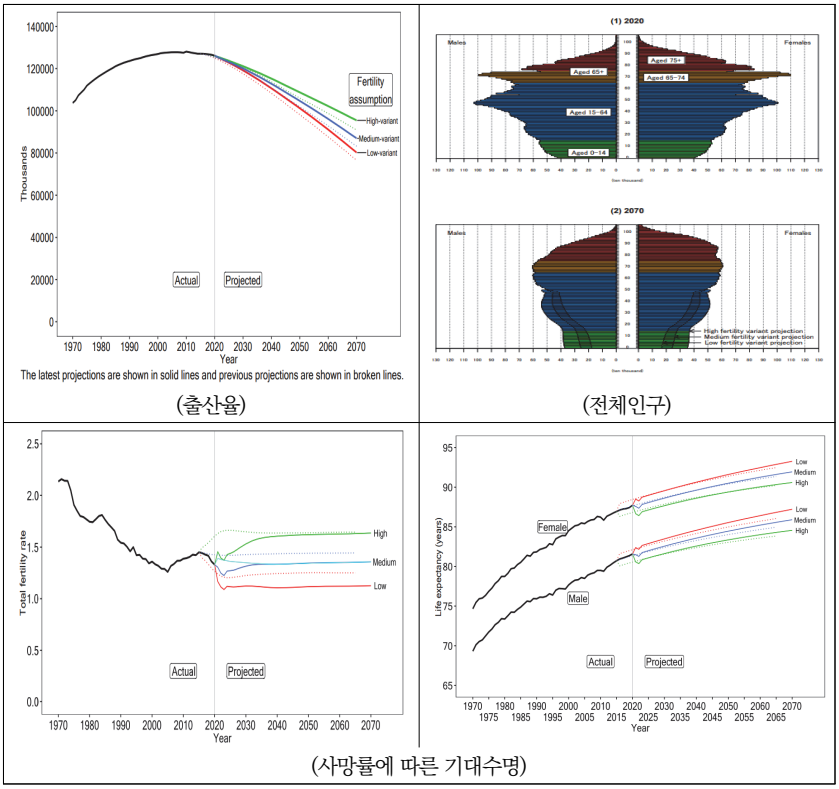
2) https://www.ipss.go.jp/index-e.asp.

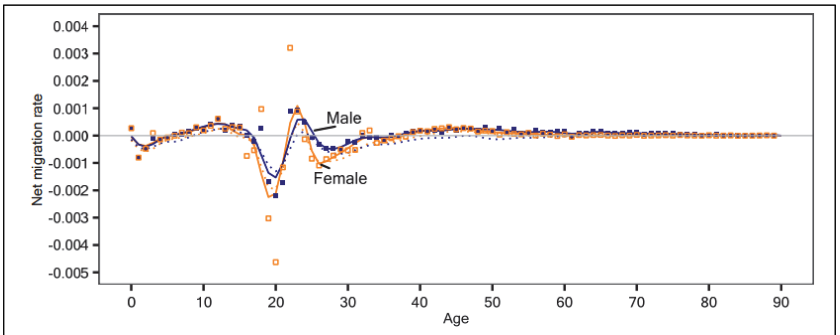
84.95년, 여성 91.35년)보다 다소 높은 수치이다.

국제 이동과 관련해서는 순이동 감소가 약화된 것으로 나타났으며, 이러한 추세는 가정에 반영되었다. 팬데믹 이전에 국제 이주민 수가 증가한 점을 고려하여, 2017년 개정에서는 연간 순이동이 2035년 69,000명에서 2040년 164,000명으로 증가하도록 상향 조정되었다. 이와 같은 결과는 〔그림 2-2〕에서 확인할 수 있다.

출산율 그래프에서 실선은 2023년 추계 결과를, 점선은 2017년 추계 결과를 나타낸다.

〔그림 2-2〕 일본 장래인구추계





(연령별 국제 순이동)

자료: www.ipss.go.jp/pp-zenkoku/e/zenkoku_e2023/pp2023e_PressRelease.pdf.

2) 미국

미국은 센서스국(U.S. Census Bureau)이 정기적으로 인구 센서스와 추계 결과를 공표하고 있다. 최근 2023년 센서스국은 2022년 7월 1일 기준 인구 추정을 시작으로 코호트요인법을 활용하여 2023년부터 2100년까지의 장래인구추계를 제공[3]하고 있다. 2023년 장래인구추계는 두 기간으로 이루어져 있다. 2060년까지는 연령, 성별, 인종, 히스패닉계 출신, 출생지, 그리고 출생지에 따른 거주 인구를 포함하고, 2061년부터 2100년까지는 연령, 성별, 출생지에 따른 거주 인구를 포함한다. 2023년 기준 인구는 센서스국의 2022년 빈티지 추정치(the Census Bureau's Vintage 2022 Estimates)를 기반으로 도출되었다. 이 추정치는 연령(0~99세, 100세 이상), 성별, 인종(31개 그룹), 히스패닉계 출신을 포함하지만, 출생지에 대한 세부 정보는 포함하지 않는다.

2021년 미국 커뮤니티 조사(American Community Survey, ACS)

3) https://www2.census.gov/programs-surveys/popproj/technical-documentation/methodology/methodstatement23.pdf.

는 각 연령, 성별, 인종 및 히스패닉계 출신 그룹의 원주민 비율을 계산하는 데 활용되었다. 이를 통해 출생률이 추가되었고, 해당 비율은 2022년 빈티지 추정치에 적용되었다. 원주민 인구는 반올림된 값을 사용해 생성되었으며, 빈티지 2022 추정치에서 이를 제외하여 외국 태생 인구를 산출하였다.

장래인구추계를 위한 주요 인구 변동 요인으로 연령별 출산율(ASFR)은 6개의 출생, 인종, 히스패닉계 출신 그룹을 기준으로 계산 및 예측되었다(〈표 2-1〉 참조).

〈표 2-1〉 미국 연령별 출산율 6개 그룹

Foreign-Born	Native[5]
Hispanic	Asian or Pacific Islander
Non-Hispanic Asian or Pacific Islander	White
Other Non-Hispanic Foreign-Born	Other Native-born

주: 5) 원주민 그룹에는 히스패닉계와 비히스패닉계 모(母)가 모두 포함.
자료: American Community Survey, ACS.

모의 출생을 설명하기 위해 이분법 변수가 사용되었으며, 미국 또는 미국 영토에서 태어난 원주민 어머니와 외국에서 태어난 어머니를 구분하였다. 비거주자의 출산은 분석에서 제외되었다. 또한, 유사한 출산율과 추세를 보이는 4개 그룹은 집계되었으며, 외국 태생 3개 그룹과 원주민 3개 그룹에 대한 비율이 생성되었다.

세부적인 연령별 출산율 계산 과정에서는 매년 출생아수를 예상 여성 인구와 예상 출산율을 곱하여 산출하고, 출생아의 성별은 출생지, 인종, 히스패닉계 출신 모 그룹 내에서 예상 출생에 따라 할당되었다. 미래 출생 성비는 2010년부터 2020년까지의 평균 출생 성비를 기준으로 설정되었으며, 모든 예상 출생은 원주민으로 간주되었다. 예상 출생의 인종 및 히스패닉계 출신은 모의 인종, 예상 인구 내 남성의 인종 구성, 그리고

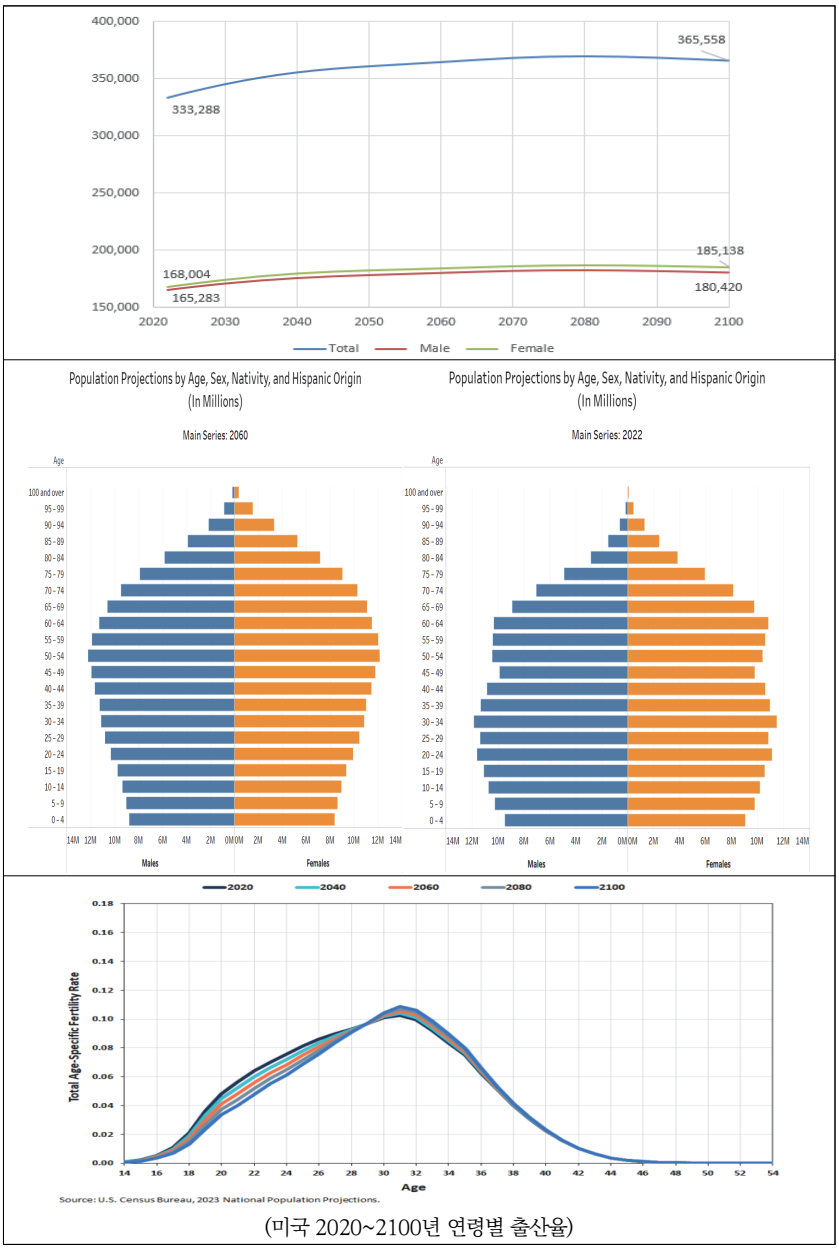
2010년 센서스에서 자녀가 있는 가구 내 여성과 남성의 인종 및 민족 분포를 기반으로 할당되었다.

사망률은 3단계 과정을 통해 산출되었다. 첫째, 성별에 따른 출생 시 기대수명을 2123년까지 예측하여 이를 기반으로 사망률 예측에 사용할 모델 생명표를 결정하였다. 둘째, 2100년까지의 사망률을 예측하였다. 마지막으로, 예측된 사망률을 활용해 2020년부터 2100년까지의 생명표를 작성하였다.

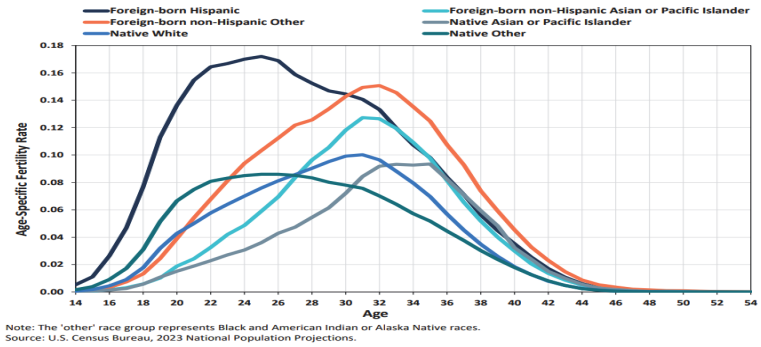
국제 순이동 예측은 외국 태생 입국(Foreign-born immigration), 외국 태생 출국(Foreign-born emigration), 그리고 내국인 국제이동(Net native migration)으로 구성하여 이루어졌음을 밝히고 있다.

〔그림 2-3〕은 미국의 장래인구추계와 인구변동요인의 시나리오를 보여준다. 미국의 총인구는 2022년 3억 3천만 명에서 증가하여 2100년에는 약 3억 6천만 명에 이를 것으로 전망된다. 여성 인구는 2022년 1억 6,800만 명에서 2100년 1억 8,500만 명으로, 남성 인구는 1억 6,500만 명에서 1억 8,000만 명으로 증가할 것으로 예상된다. 또한, [그림 2-3] 하단에는 인종별 및 성별에 따른 출산, 사망, 국제 이동의 추이를 시각적으로 나타내고 있다.

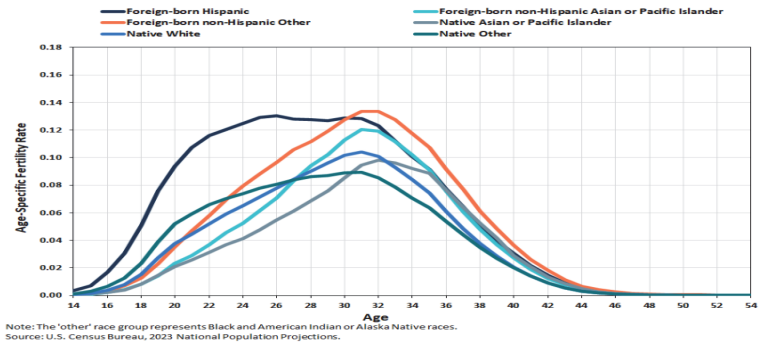
〔그림 2-3〕 미국 장래인구추계 주요 결과



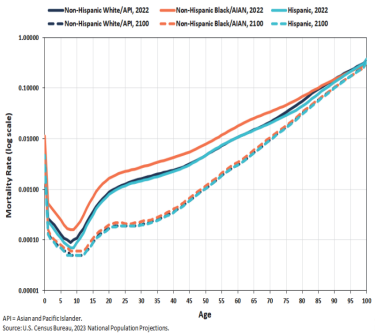
(미국 2020~2100년 연령별 출산율)

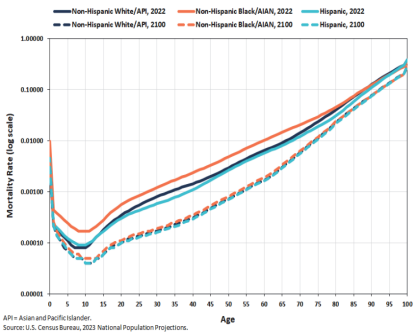


(미국: Age-Specific Fertility Rates by Nativity, Race, and Hispanic Origin: 2020년)

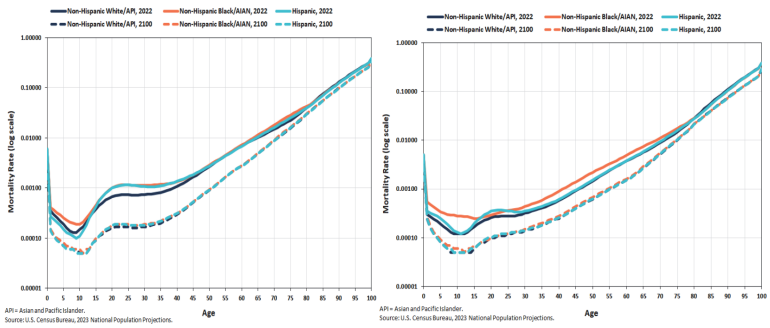


(미국: 예측 Age-Specific Fertility Rates by Nativity, Race, and Hispanic Origin: 2060년)

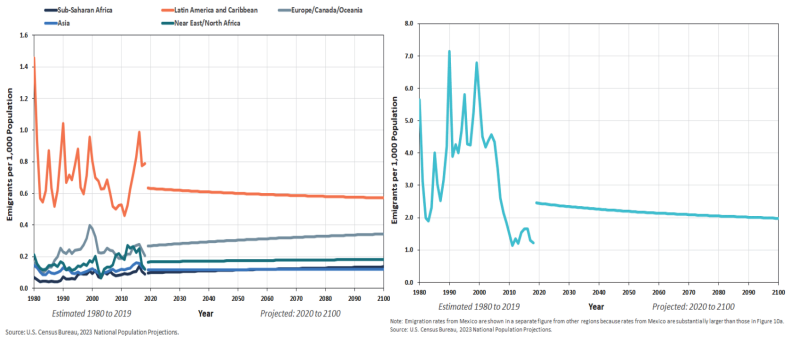




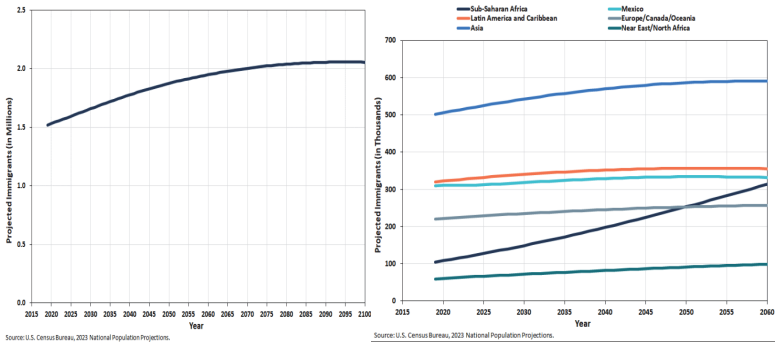
(좌) Mortality Rates by Age, Race, and Hispanic Origin for native Males:2022~2100년)
(우) Mortality Rates by Age, Race, and Hispanic Origin for native Female:2022~2100년)



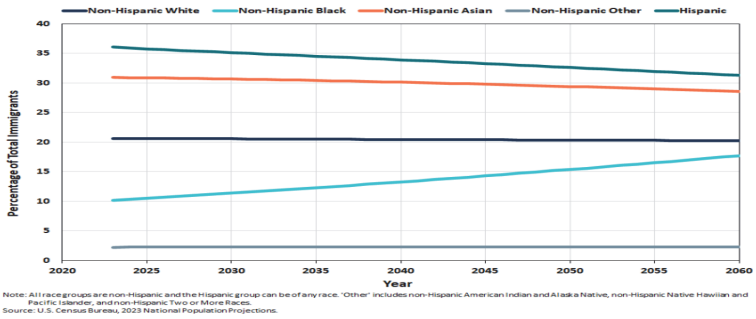
(좌) Mortality Rates by Age, Race, and Hispanic Origin for Foreign-born Males:2022~2100년)
(우) Mortality Rates by Age, Race, and Hispanic Origin for Foreign-born Female:2022~2100년)



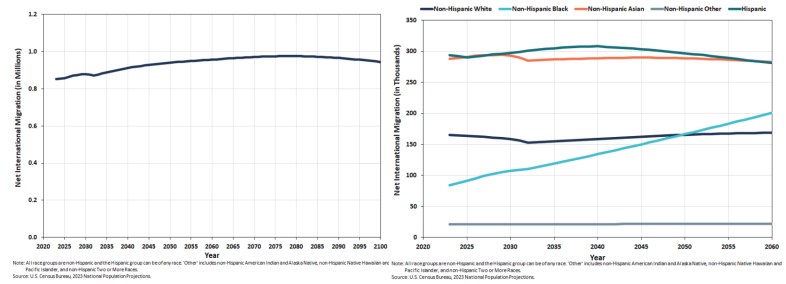
(좌) 입국률(Sending Region to United States): 1980~2100년
(우) 입국률(Mexico to United States): 1980~2100년



(좌) projections of foreign-born immigration to the United states : 2019~2100년
(우) projections of foreign-born immigration to the United states by Sending Regions : 2019~2060년



인종(Race) 히스패닉(Hispanic Origin) 외국인별 입국자 수 예측 분포(2023~2060년)



(좌) 2023~2100년 순 이동자 수 예측
(우) 2023~2100년 인종과 Hispanic origin별 순 이동자 수

자료: https://www2.census.gov/programs-surveys/popproj/technical-documentation/.

3) 호주

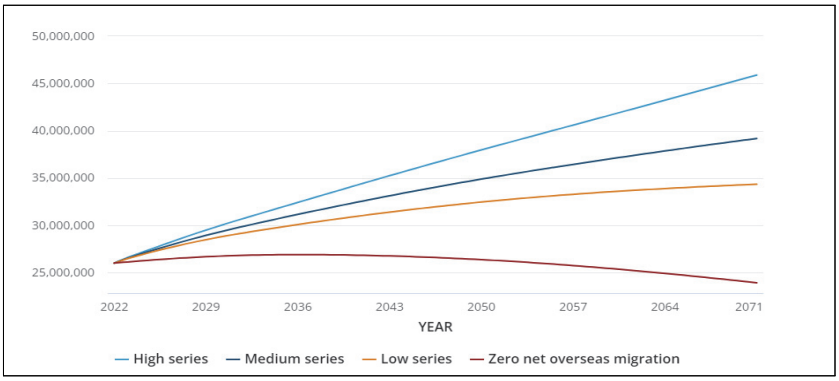
호주는 통계청(Australian Bureau of Statistics, ABS)이 인구센서스와 추계 결과를 정기적으로 공표하고 있다. 최근 2022년 ABS에서 발표한 통계는 2071년까지의 장래인구추계를 제공한다. ABS는 코호트요인법을 활용해 인구 예측을 수행하며, 추계 기준일은 2022년 6월 30일이다. 예측 범위는 호주 전체, 각 주와 영토, 주요 도시, 그리고 기타 지역을 포함하여 2023년 6월 30일부터 2071년 6월 30일까지의 기간을 포괄한다. 기준 인구는 2021년 인구 및 주택 조사 결과를 반영해 2022년 6월

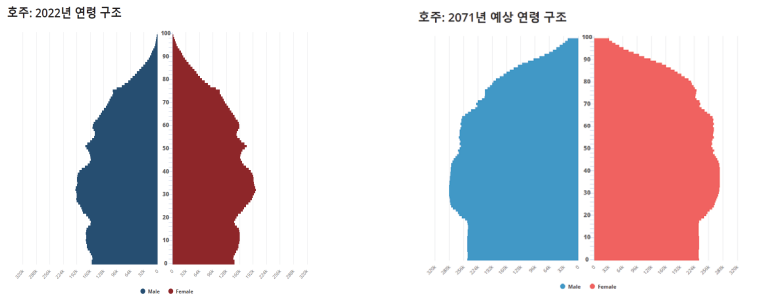
30일 기준으로 추정된 거주 인구이다.

주요 결과에 따르면, 2022년 호주의 인구는 약 2,600만 명이며, 2071년까지 3,430만~4,590만 명으로 증가할 것으로 예상된다. 현재 10년 평균 연평균 성장률인 1.4%는 장기적으로 0.2%~0.9%로 감소할 것으로 보인다. 중위 연령은 2022년 38.5세에서 2071년 43.8세~47.6세로 증가할 것으로 전망된다. 2032년 6월까지 연평균 1.2%~1.7%의 성장률로 인구가 증가하며, 2032년까지는 2,920만~3,080만 명, 2071년까지는 3,430만~4,590만 명에 이를 것으로 보인다.

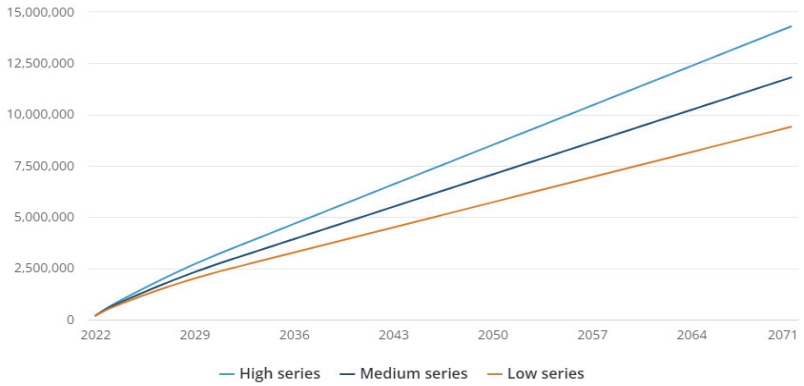
ABS의 장래인구추계 시나리오는 인구 변동 요인별로 단계적으로 적용하는 것이 특징이다. 예를 들어, 합계출산율은 2026~2027년, 출생 시 성별 기대수명은 2070~2071년, 순 해외 이주는 2031~2032년을 기준으로 추정된다. 특히, 호주는 국제 이동의 영향을 크게 받는 국가로, 모든 시나리오에서 입국자가 출국자를 초과할 것으로 가정하고 있다. 이에 따라, 2022년부터 2071년까지 총 해외 순이동은 920만 명에서 1,430만 명 사이로 증가할 것으로 예측된다(〔그림 2-4〕 참조).

〔그림 2-4〕 호주 장래인구추계와 시나리오





(2022~2071 장래인구추계)



(2022~2071 국제 이동)

	최근 관찰된 평균	낮은 가정	중간 가정	높은 가정	단계적으로 도입
합계출산율(a)	1.64	1.45	1.60	1.75	2026-27
출생 시 남성의 기대수명(a)	81.2	83.9	83.9	88.4	2070-71
출생 시 여성의 기대수명(a)	85.3	86.8	86.8	91.1	2070-71
순 해외이주(a)	103,300	175,000	225,000	275,000	2031-32
순 주간이주	..	작은 흐름	중간 흐름	큰 흐름	2026-27

.. 해당 없음

a. 최근 관찰된 평균은 2020-2022년 달력 연도입니다.

(단계별 인구변동요인별, 단계별 시나리오)

자료: https://www.abs.gov.au/statistics/people/population/household-and-family-projections-australia.

4) 캐나다

캐나다는 통계청(Canada's national statistical agency)을 통해 인구센서스와 추계 결과를 정기적으로 발표하고 있다. 가장 최근의 장래인구추계는 2023년 인구 추정치를 기준으로 하며, 2023년부터 2048년까지는 주와 준주 규모로, 2023년부터 2073년까지는 국가 차원의 추계를 제공한다. 특히, 2021년 인구 조사에서는 처음으로 성별 개념을 통합하여, 성별을 남성, 여성 또는 비바이너리(non-binary person, 남성이나 여성에 국한되지 않는 사람)로 구분하고, 개인의 사회적 및 개인적 정체성을 반영해 분석하였다.

다양한 시나리오에 따르면, 2023년 약 4,010만 명으로 추산된 캐나다의 인구는 지속적으로 증가하여 2073년까지 4,710만 명(저위 시나리오, LG)에서 8,720만 명(고위 시나리오, HG) 사이에 이를 것으로 예측된다. 중위 시나리오(M1)에서는 2073년 캐나다 인구가 6,280만 명에 도달할 것으로 전망된다. 과거 30년간 연평균 1.12%였던 인구 성장률은 중위 시나리오에 따르면 2072~2073년까지 0.79%로 감소할 것으로 보인다. 반면, 고위 시나리오에서는 1.59%, 저위 시나리오에서는 0.07%의 성장률을 기록할 것으로 예측된다. 모든 시나리오에서 국제이동은 캐나다 인구 증가의 주요 동력으로 작용할 것이며, 이는 1990년대 초에 시작된 국제이동 증가 패턴을 이어갈 것으로 보인다.

인구 구조를 살펴보면, 65세 이상 인구의 비율은 2023년 18.9%에서 2073년 21.9%(저속 고령화 시나리오, SA)에서 32.3%(고속 고령화 시나리오, FA)까지 증가할 것으로 보인다. 그러나 베이비붐 세대가 2030년 이후 이 연령대를 통과하면, 65세 이상 인구 비율 증가는 점차 완화될 것으로 예상된다. 유소년층(0~14세)의 비율은 1962년 34.0%로 최고점을

기록한 이후 감소세를 보였으며, 2023년 15.4%로 추정된다. 모든 시나리오에서 유소년층 비율은 감소할 것으로 보이며, 이는 느린 고령화(SA) 및 고성장(HG) 시나리오를 제외한 결과이다.

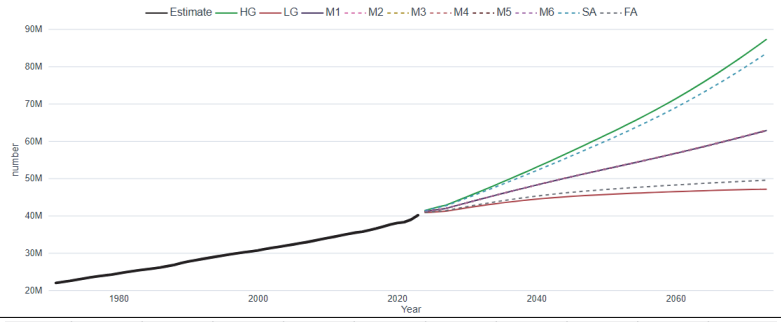
초고령층(85세 이상 인구)은 향후 몇 년간 빠르게 증가할 전망이며, 특히 2031년에서 2050년 사이 베이비붐 세대가 이 연령대에 진입함에 따라 건강관리와 서비스 수요가 크게 증가할 것이다. 예측 시나리오에 따르면, 85세 이상 인구는 2023년 약 896,600명에서 2073년 330만 명(저위 시나리오, LG)에서 430만 명(고위 시나리오, HG)으로 증가할 것으로 보인다.

마지막으로, 캐나다 인구의 중위 연령은 2023년 41.6세에서 2073년에는 42.6세(시나리오 SA)에서 50.1세(시나리오 FA)로 증가할 것으로 전망된다. 〔그림 2-5〕는 이러한 캐나다 장래인구추계의 주요 결과를 시각적으로 보여준다.

〔그림 2-5〕 캐나다 장래인구추계와 시나리오

• By 2073, the total population in Canada could range between 47,084,500 (Low-growth scenario (LG)) and 87,238,400 (High-growth scenario (HG)).

• Under the medium-growth scenario (M1), the total population in Canada could reach 62,796,600 by 2073, a 56.6% increase compared with 2023 (40,097,761).



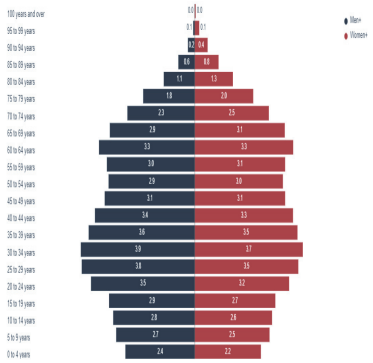
Projection scenario[1]	Projection scenario LG: low-growth	Projection scenario M1: medium-growth	Projection scenario M2: medium-growth	Projection scenario M3: medium-growth	Projection scenario M4: medium-growth	Projection scenario M5: medium-growth	Projection scenario M6: medium-growth	Projection scenario HG: high-growth	Projection scenario SA: slow-aging	Projection scenario FA: fast-aging

Summary of the projection scenarios

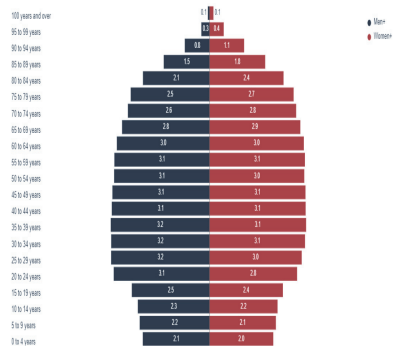
Scenario	Fertility	Mortality	Immigration	Emigration and returning emigration	Non-permanent residents	Internal migration
M1	Medium	Medium	Medium	Medium	Medium	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023
M2	Medium	Medium	Medium	Medium	Medium	1995/1996 to 2010/2011
M3	Medium	Medium	Medium	Medium	Medium	2003/2004 to 2008/2009
M4	Medium	Medium	Medium	Medium	Medium	2009/2010 to 2016/2017
M5	Medium	Medium	Medium	Medium	Medium	2014/2015 to 2016/2017
M6	Medium	Medium	Medium	Medium	Medium	2020/2021 to 2022/2023
LG	Low	High	Low	High	Low	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023
HG	High	Low	High	Low	High	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023
SA	High	High	High	Medium	High	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023
FA	Low	Low	Low	Medium	Low	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023

Notes: LG (low growth), HG (high growth), SA (slow aging) and FA (fast aging).
Source: Statistics Canada, Centre for Demography.

(시나리오 개요)



(2024년)



(2073년)

Detailed summary of projection scenarios

Component / Temporal horizon	Scenario: Low growth LG	Medium growth M1	M2	M3	M4	M5	M6	High growth HG	Slow aging SA	Fast aging FA
Fertility (period total fertility rate (number of children per woman))										
2027/2028	1.13	1.24	1.24	1.24	1.24	1.24	1.24	1.36	1.36	1.13
2047/2048	1.15	1.37	1.37	1.37	1.37	1.37	1.37	1.61	1.61	1.15
2072/2073	1.15	1.37	1.37	1.37	1.37	1.37	1.37	1.61	1.61	1.15
Immigration (rate per thousand)										
2027/2028	10.3	11.9	11.9	11.9	11.9	11.9	11.9	13.8	13.8	10.3
2047/2048	7.0	9.3	9.3	9.3	9.3	9.3	9.3	12.0	12.0	7.0
2072/2073	7.0	9.3	9.3	9.3	9.3	9.3	9.3	12.0	12.0	7.0
Life expectancy at birth, males (in years)										
2027/2028	80.3	80.6	80.6	80.6	80.6	80.6	80.6	80.9	80.3	80.9
2047/2048	83.6	84.3	84.3	84.3	84.3	84.3	84.3	85.0	83.6	85.0
2072/2073	86.5	87.3	87.3	87.3	87.3	87.3	87.3	88.1	86.5	88.1
Life expectancy at birth, females (in years)										
2027/2028	84.5	84.8	84.8	84.8	84.8	84.8	84.8	85.0	84.5	85.0
2047/2048	87.1	87.8	87.8	87.8	87.8	87.8	87.8	88.4	87.1	88.4
2072/2073	89.6	90.3	90.3	90.3	90.3	90.3	90.3	91.0	89.6	91.0
Proportion of non-permanent residents (%)										
2028	4.1	5.0	5.0	5.0	5.0	5.0	5.0	6.1	6.1	4.1
2048	3.1	4.7	4.7	4.7	4.7	4.7	4.7	6.5	6.5	3.1
2073	3.1	4.7	4.7	4.7	4.7	4.7	4.7	6.5	6.5	3.1
Emigration (gross migraproduction rate per thousand)										
2027/2028	2.4	2.0	2.0	2.0	2.0	2.0	2.0	1.7	2.0	2.0
2047/2048	3.0	2.2	2.2	2.2	2.2	2.2	2.2	1.6	2.2	2.2
2072/2073	3.0	2.2	2.2	2.2	2.2	2.2	2.2	1.6	2.2	2.2
Return emigration (gross migraproduction rate per thousand)										
2027/2028	1.7	1.4	1.4	1.4	1.4	1.4	1.4	1.2	1.4	1.4
2047/2048	1.8	1.4	1.4	1.4	1.4	1.4	1.4	1.0	1.4	1.4
2072/2073	1.8	1.4	1.4	1.4	1.4	1.4	1.4	1.0	1.4	1.4
Interprovincial migration										
Reference period	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023	1995/1996 to 2010/2011	2003/2004 to 2008/2009	2009/2010 to 2016/2017	2014/2015 to 2016/2017	2020/2021 to 2022/2023	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023	Recent trends (2020/2021 to 2022/2023) transitioning linearly in 10 years to the average of the period 1991/1992 to 2022/2023

Note: The medium growth scenarios M2, M3, M4, M5 and M6 were created in order to reflect distinct interprovincial migration assumptions in comparison with the medium growth scenario M1. For more details, see the section on interprovincial migration in *Population Projections for Canada (2023 to 2073), Provinces and Territories (2022 to 2046): Technical Report on Methodology and Assumptions*, Statistics Canada catalogue number 91-620.
Source: Statistics Canada, Centre for Demography.

(상세한 시나리오 설명)

자료: https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2022015-eng.htm.

나. 초저출산 국가(싱가포르, 홍콩, 대만)

최근 국가별 합계출산율 공표에 따르면, TFR이 1.0명 이하인 국가는 한국(2023년 인구 5,133만 명, TFR 0.72명), 홍콩(2021년 인구 734만 명, TFR 0.77명), 대만(2022년 인구 2,330만 명, TFR 0.87명), 싱가포르(2023년 인구 590만 명, TFR 0.97명)이다. 이러한 현상은 언론 매체에서 '동아시아 인구 절벽'이라는 표현으로 연일 보도되고 있다. 특히, 이들 국가의 향후 출산율과 인구 추이는 주목할 만하다.

그러나 2024년 8월 기준으로, 싱가포르, 홍콩, 대만의 통계청은 최근 인구통계[4]만 제공하고 있으며, 장래인구추계에 대한 정보는 제공하지 않고 있다. 따라서 본 연구는 UN의 World Population Prospects (WPPs) 2024[5]를 참고하여 이들 국가의 인구 변동 요인과 인구 전망을 분석하기로 한다.

1) 싱가포르

싱가포르 통계청에 따르면, 2023년 공식 인구는 592만 명이며, TFR은 0.97명, 기대수명은 남성 80.7년, 여성 85.2년, 전체 평균 83.0년이다. TFR은 1990년 1.83명에서 지속적으로 하락하여 2022년 1.04명, 2023년에는 1.0명 아래로 떨어져 0.97명을 기록하고 있다. 〔그림 2-6〕

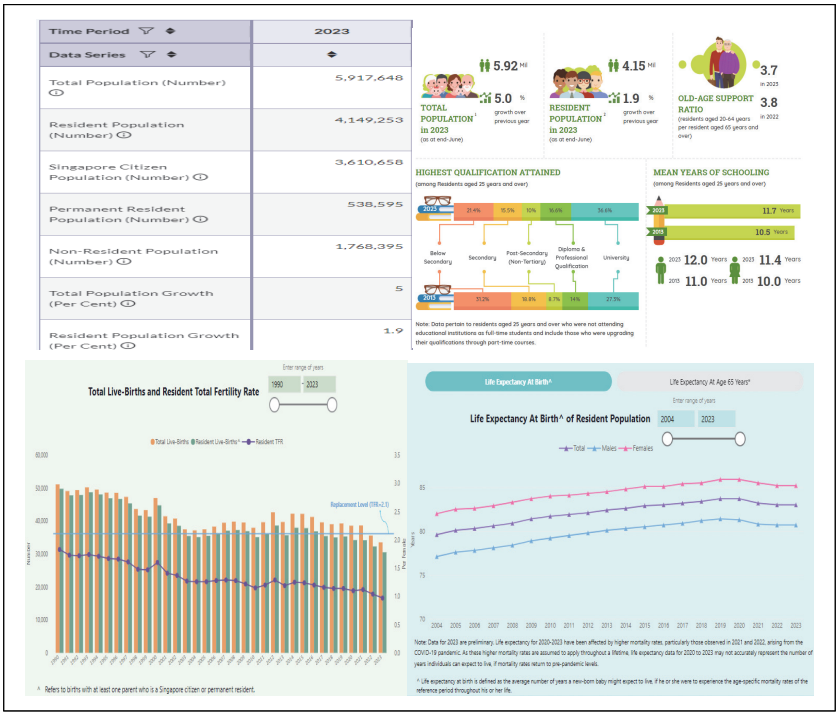
---

4) 홍콩: https://www.censtatd.gov.hk/en/, https://www.censtatd.gov.hk/en/data/stat_report/product/FA100090/att/B72302FA2023XXXXB0100.pdf.
대만: https://eng.stat.gov.tw/Default.aspx, https://eng.stat.gov.tw/public/data/dgbas03/bs2/yearbook_eng/y005.pdf.
싱가포르: https://www.singstat.gov.sg/, https://www.singstat.gov.sg/find-data/search-by-theme/population/births-and-fertility/latest-data, https://www.singstat.gov.sg/publications/population/population-trends.

5) https://population.un.org/wpp/.

은 싱가포르 통계청이 발표한 2023년 인구 트렌드 주요 결과를 시각적으로 보여준다. 싱가포르의 장래 인구를 살펴보기 위해 UN WPPs 2024 자료를 참고하면, 2025년 이후 싱가포르의 인구는 지속적으로 증가하여 2035~2040년에 약 600만 명으로 정점을 찍을 것으로 보인다. 이후에는 인구 감소가 예상된다.

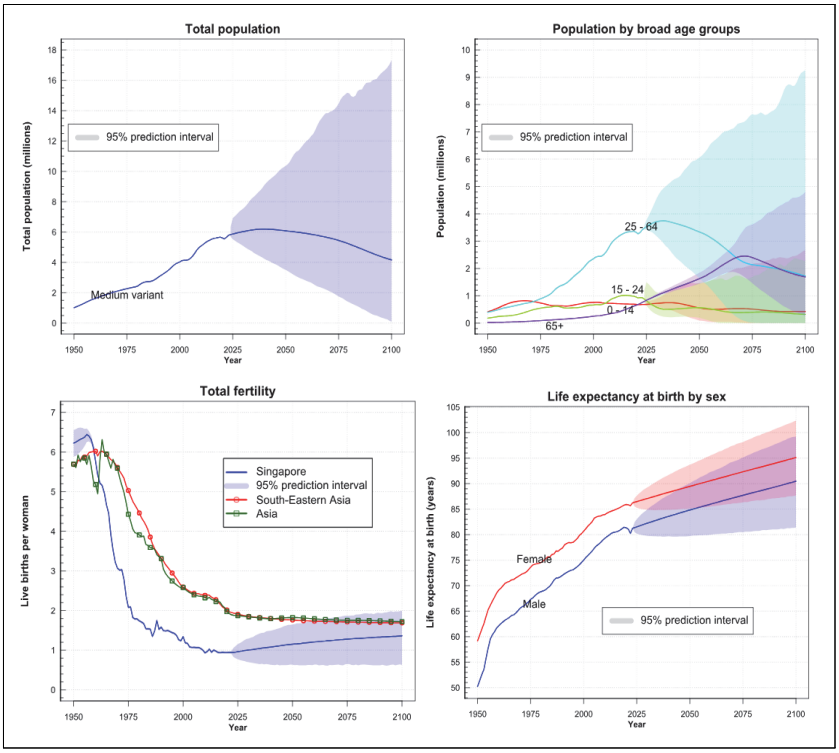
[그림 2-6] 싱가포르 인구 2023 trend

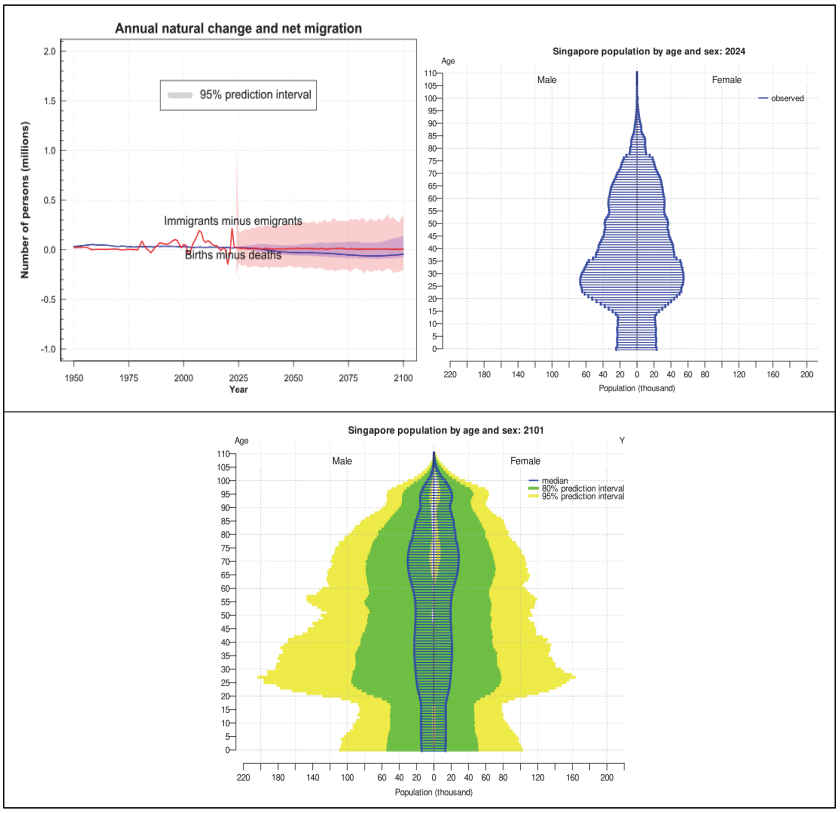


자료: https://tablebuilder.singstat.gov.sg/table/TS/M810001.
https://www.singstat.gov.sg/modules/infographics/population.
www.singstat.gov.sg/find-data/search-by-theme/population/death-and-life-expectancy/visualising-data/mortality-dashboard.
https://www.singstat.gov.sg/find-data/search-by-theme/population/births-and-fertility/visualising-data/fertility-dashboard.

인구 구조 변화를 살펴보면, 유소년 및 생산가능인구는 지속적으로 감소하는 반면, 65세 이상 인구는 점진적으로 증가하는 추세를 보인다. 특히, UN WPPs 2024에 따르면, 2023년 0.97명 수준이었던 싱가포르의 TFR은 지속적인 감소세에서 반등하여 1.3~1.5명 수준으로 회복될 것으로 전망된다. [그림 2-7]은 싱가포르 장래인구추계의 주요 결과를 보여준다.

[그림 2-7] 싱가포르 장래인구추계 주요 결과





자료: https://population.un.org/wpp/Graphs/DemographicProfiles/Line/702.
https://population.un.org/wpp/Graphs/DemographicProfiles/Pyramid/702.

2) 대만

다음으로, 대만 통계청에 따르면 2020년 인구총조사를 실시하였으며, 2024년 6월 기준 인구는 2,340만 명, 2021년 기준[6] TFR은 0.975명이다. 기대수명은 2020년 기준 남성 78.11년, 여성 84.75년으로 공표되었다. 대만의 TFR 추이는 싱가포르와 유사하게 지속적인 감소세를 보이며,

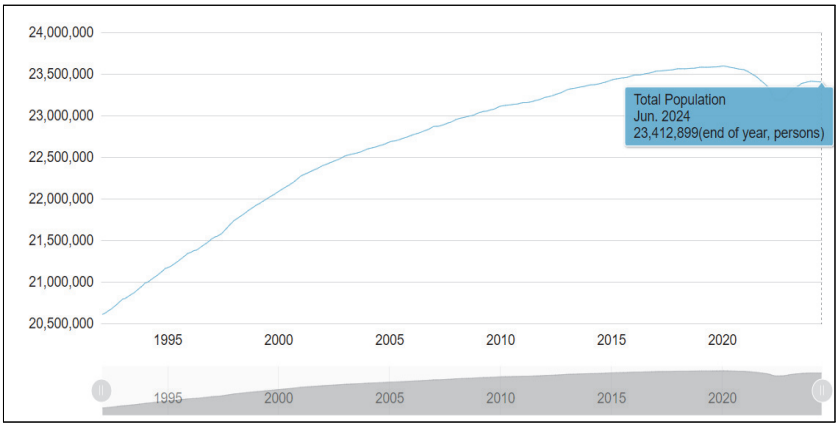
6) https://ws.dgbas.gov.tw/public/data/dgbas03/bs2/yearbook_eng/yearbook2021.pdf.

1991년 1.72명, 2015년 1.175명에서 하락하여 2020년 1.0명 아래인 0.99명, 2021년 0.975명을 기록하였다.

〔그림 2-8〕은 대만 통계청에서 제시한 인구 추이를 나타낸다. 1995년부터 인구 증가세를 유지하다가 2020년 이후 감소세로 전환되는 경향이 확인된다.

대만의 장래 인구를 분석하기 위해 UN WPPs 2024를 참고하면, 2025년 이후 대만의 인구는 증가세를 보이며 2035~2040년경 약 2,300만 명 수준에서 정점을 기록한 후 지속적으로 감소할 것으로 전망된다. 인구 구조 변화에서는 유소년 및 생산가능인구의 지속적인 감소와 65세 이상 인구의 증가 경향이 나타난다.

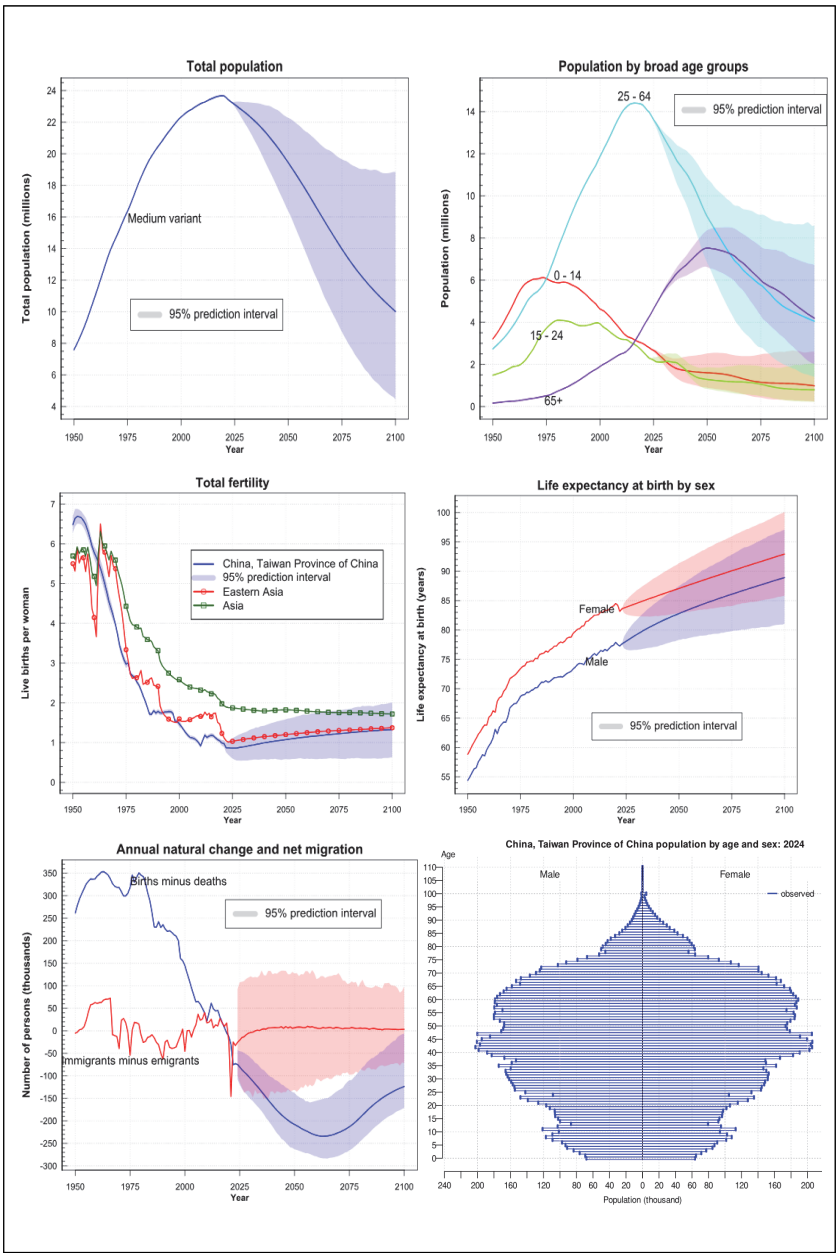
〔그림 2-8〕 대만 인구 추이

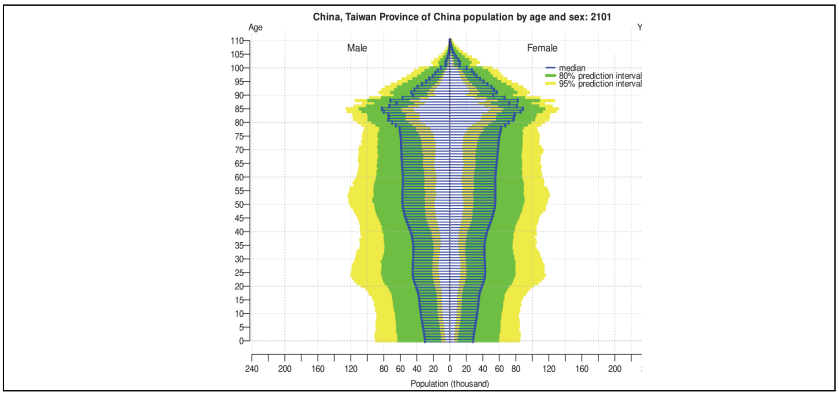


자료: https://eng.stat.gov.tw/Point.aspx?sid=t.9&n=4208&sms=11713.

특히, UN WPPs 2024는 2023년 TFR이 0.97명 수준에서 감소세를 멈추고 상승세로 전환되어, 1.3~1.5명 수준으로 회복할 것으로 전망하고 있다. 〔그림 2-9〕는 대만의 장래인구추계 주요 결과를 시각적으로 보여준다.

〔그림 2-9〕 대만 장래인구추계 주요 결과





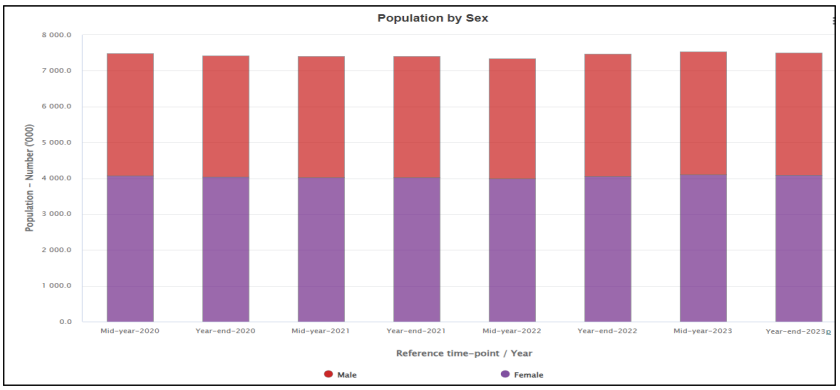
자료: https://population.un.org/wpp/Graphs/DemographicProfiles/Line/158.
https://population.un.org/wpp/Graphs/DemographicProfiles/Pyramid/158.

3) 홍콩

끝으로, 홍콩 통계청에 따르면 2023년 6월 기준 홍콩의 인구는 753만 명이며, 2021년 기준[7] 합계출산율(TFR)은 0.772명이다. 기대수명[8]은 2020년 기준 남성 83.35년, 여성 87.75년으로 공표되었다. 홍콩의 TFR 추이는 싱가포르 및 대만과 유사하게 1991년 1.281명, 2012년 1.285명 수준으로 유지되다가 2019년 1.064명, 2020년 0.883명으로 하락하였으며, 2021년에는 0.772명으로 1.0명 이하 수준을 기록하였다. 〔그림 2-10〕은 홍콩 통계청에서 제시한 인구 추이를 나타낸다. 2020년 이후 홍콩의 총인구는 750만 명 수준을 유지하고 있으며, 남성보다 여성의 비중이 높은 것으로 나타났다.

7) www.censtatd.gov.hk/en/data/stat_report/product/FA100090/att/B72302FA2023XXXXB0100.pdf.
8) https://www.censtatd.gov.hk/en/web_table.html?id=115-01021.

〔그림 2-10〕 홍콩 인구 2020~2023

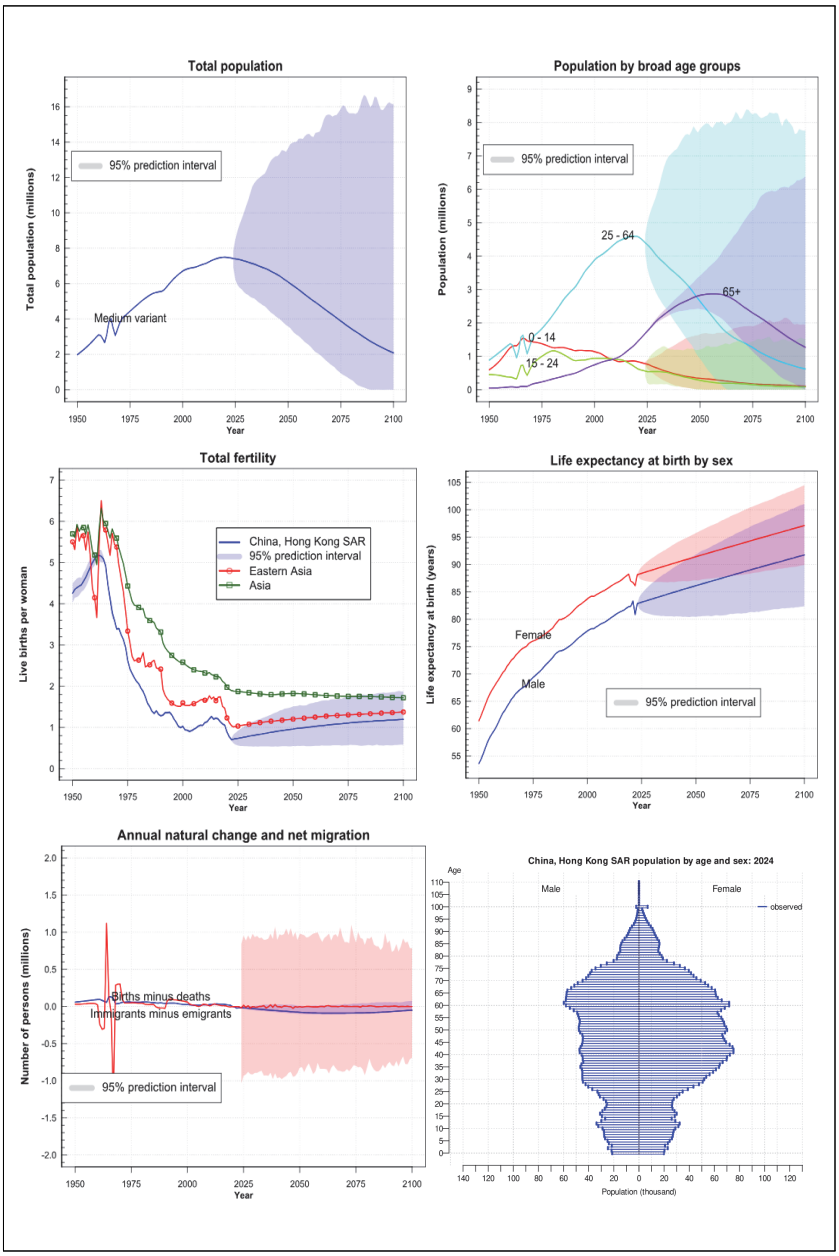


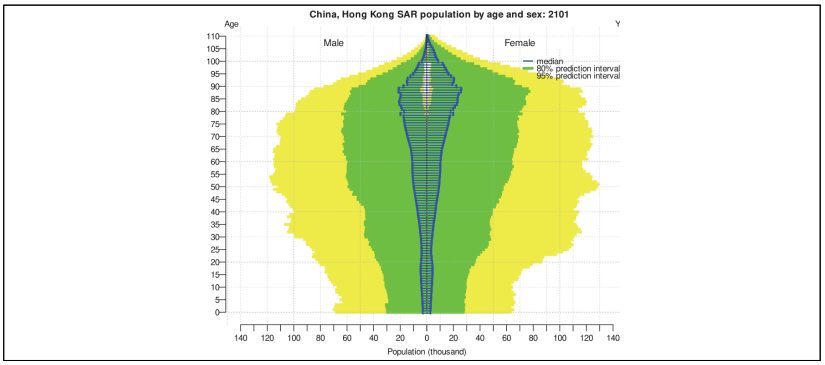
자료: https://www.censtatd.gov.hk/en/web_table.html?id=110-01001#.

홍콩의 장래인구추계를 살펴보기 위해 UN WPPs 2024를 참고하면, 2023년까지 750만 명 수준에서 정점을 기록한 후, 이후 지속적으로 감소할 것으로 전망된다. 2101년에는 인구가 1950년 수준인 200만 명까지 감소할 것으로 예상된다. 인구 구조를 보면, 유소년 및 생산가능인구는 지속적으로 감소하는 반면, 65세 이상 인구는 증가하는 경향을 보인다. 다만, 고령 인구도 2060년 이후에는 감소세로 전환될 것으로 전망된다.

특히, UN WPPs 2024는 2021년 0.772명 수준이었던 TFR이 지속적인 감소세에서 반등하여 1.2~1.3명 수준으로 회복될 것으로 예측하고 있다. 〔그림 2-11〕은 홍콩의 장래인구추계 주요 결과를 시각적으로 보여준다.

〔그림 2-11〕 홍콩 장래인구추계 주요 결과





자료: https://population.un.org/wpp/Graphs/DemographicProfiles/Pyramid/344.
https://population.un.org/wpp/Graphs/DemographicProfiles/Line/344.

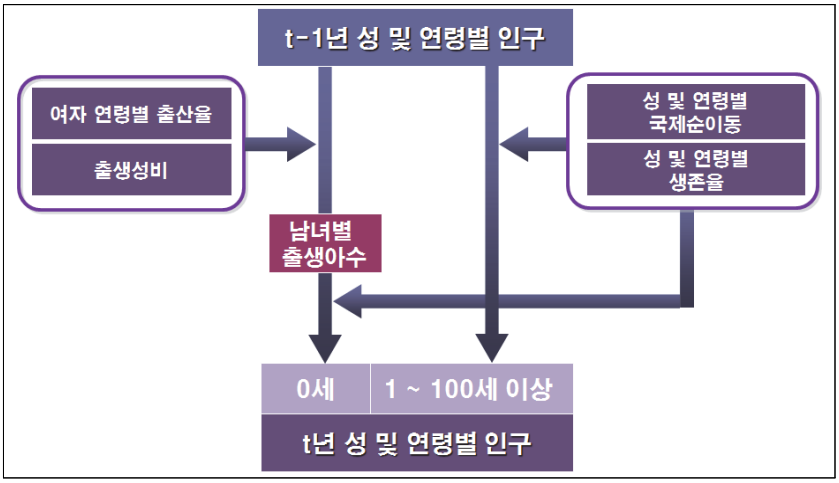
지금까지 싱가포르, 대만, 홍콩 통계청의 최신 인구통계와 UN WPPs 2024를 바탕으로 이들 국가의 인구 변동 요인, 인구 구조 변화, 그리고 장래인구추이를 분석하였다.

그러나 여기서 유의해야 할 점은, UN WPPs 2024는 선진국의 출산율 사전 분포(prior distribution)와 개발도상국 또는 후진국의 출산율 사후 분포(posterior distribution)를 생성하여 베이지안 방법에 적용하여 도출한 결과이다. 특히, 출산율은 인구 대체 수준인 2.1명으로 회귀(regression)하거나 그 이하 수준에 머물 것이라는 가정하에 예측되었다. 따라서 지속적으로 TFR이 하락하고 있는 이들 국가의 최근 출산율 변화 양상이 충분히 반영되지 않았을 가능성이 있다. 결과 해석 시 이런 점을 주의할 필요가 있다.

통계청(2024) 장래인구추계는 등록센서스 방식의 2022년 인구총조사와 최근까지의 인구동태(출생, 사망)통계와 국제인구이동통계를 활용하여 코호트요인법에 의해 2022~2072년까지의 인구규모와 성과 연령별 인구구조를 추계한다. 코호트요인법은 출생, 사망, 국제이동의 인구변동 요인별 미래 수준을 각각 예측한 후, 추계의 시작점이 되는 기준 인구

(base population, 장래인구추계의 출발점이 되는 2022년 7월 1일 시점의 인구)에 출생아수와 국제순이동은 더하고 사망자수는 빼는 인구균형방정식(demographic balancing equation)을 적용하여 다음해 인구를 반복적으로 산출해 나가는 인구추계 방법이다. 〔그림 2-12〕는 코호트요인법에 따른 통계청(2024)의 인구추계 과정을 보여준다.

〔그림 2-12〕 통계청(2024) 인구추계 과정



자료: 통계청(2024) 장래인구추계2022~2072.

3. 현 인구추계 주요 쟁점

가. 현 인구추계 모형의 특징과 한계

현재 통계청을 비롯한 주요국 통계기관이 채택하고 있는 코호트요인법 기반의 결정적 추계는 여러 장점에도 불구하고 급변하는 인구변동 환경에서 몇 가지 중요한 한계를 드러내고 있다.

첫째, 시나리오 기반 추계의 해석상 한계이다. 통계청(2024)이 제공하는 31개 시나리오는 다양한 가능성을 제시하지만, 각 시나리오가 실현될 확률적 정보는 제공하지 않는다. 이는 정책 입안자들이 “어떤 시나리오를 기준으로 정책을 설계해야 하는가?”라는 근본적 질문에 직면하게 만든다. 확률적 인구추계의 필요성이 제기되는 이유이기도 하다.

둘째, 한계연령 100세 제한으로 인한 정책분석 제약이다. 현행 추계는 100세를 한계연령으로 설정하고 있어, 기대수명 연장과 초고령사회 진입에 따른 연금 재정추계, 의료비 지출 전망 등 주요 정책분석에서 과소추정의 위험이 있다. 특히 우리나라의 급속한 기대수명 증가(남성 81.04년, 여성 87.33년, 2023년 기준)를 고려할 때, 한계연령 확장이 필요한 상황이다.

셋째, 초저출산 지속에 따른 기존 모형의 예측력 한계이다. 우리나라의 합계출산율이 2023년 0.72명으로 세계 최저 수준을 기록하면서, 기존 출산력 전망 모형의 적합성에 대한 의문이 제기되고 있다. 앞서 살펴본 초저출산 국가들의 UN 전망치가 ‘인구 대체 수준인 2.1명으로 회귀하거나 그 이하 수준에 머물 것’이라는 사전 가정하에 도출된 것처럼, 현실의 급격한 변화를 충분히 반영하지 못할 가능성이 있다.

넷째, 코로나19 이후 나타난 새로운 인구변동 패턴에 대한 대응 한계이다. 팬데믹은 결혼·출산 가치관의 급변, 국제이동 패턴의 변화, 사망력의 일시적 증가 등 기존 추계 모형이 상정하지 못한 복합적 변화를 가져왔다. 이러한 구조적 변화를 반영할 수 있는 보다 유연한 추계 도구의 필요성이 대두되고 있다.

나. 인구·가구 통합추계 접근의 필요성

최근 한국 사회는 저출산·고령화의 가속화, 혼인과 가족형태의 다양화, 1인가구의 급증 등 급격한 인구사회적 구조 변화를 겪고 있다. 이러한 인구구조의 변화는 가구구조의 변화를 동반하며, 주거정책, 복지정책, 노동 및 보건의료체계에 이르기까지 다양한 정책 영역에 중대한 영향을 미친다. 이에 따라 단순한 인구추계를 넘어, 가구 단위의 변화 추이를 종합적으로 파악하는 추계 모델의 필요성이 증가하고 있다.

통계청의 장래가구추계(2024)는 이러한 인구-가구 연계를 반영하기 위한 대표적인 시도로 볼 수 있다. 해당 추계는 인구추계 결과를 기반으로 대상인구를 도출하고, 이를 연령별, 성별, 혼인상태별로 세분화한 후, 가구주가 될 확률(가구주율)을 적용하여 장래 가구수, 가구원수와 가구유형을 예측한다. 특히 혼인상태별 전이확률을 고려한 가구주율법(Headship Rate Method)을 활용함으로써, 혼인, 이혼, 사별, 사망 등의 인구학적 사건이 가구형태 변화에 미치는 구조적 연결성을 반영하고 있다.

그러나 현행 추계 방식의 한계도 존재한다. 기존의 장래가구추계는 인구추계의 결과만을 활용할 뿐 인구추계와 독립적인 절차로 운영되어 왔으며, 연령별 단순 가구주율이나 고정된 가구유형 분포를 적용하는 방식이다. 이로 인해 인구추계에서 전망한 인구와 가구추계에서 가정한 혼인상태 및 연령대별 가구주율 간의 연계가 체계적으로 이루어지지 않아, 가구 형성 과정에서의 논리적 모순이 발생할 수 있다.

특히 혼인상태별 접근의 중요성이 더욱 부각되고 있다. 미혼율 증가와 고령층의 가족분화는 1인가구와 독거노인 가구의 급증으로 이어지고 있으며, 이러한 변화는 단순한 인구수 변화보다 가구구조 변화의 속도와 영향이 더욱 클 수 있음을 시사한다. 혼인상태는 출산력과 가구 형성의 핵

심 변수임에도 불구하고, 현행 추계에서는 이를 체계적으로 고려하지 못하고 있어, 미혼율 증가, 초혼 연령 상승, 이혼율 변화 등이 출산력에 미치는 영향을 정확히 포착하기 어렵다.

정책적 활용도 측면에서도 통합적 접근의 필요성이 크다. 고령자 복지, 주택 수급, 돌봄정책 등은 단순한 연령구조보다 가구유형과 구성원 간 관계에 따라 정책 수요가 달라지기 때문에, 보다 세분화된 예측이 필요하다. 예를 들어, 단순 인구고령화 추계보다 고령 1인가구 증가율 추계가 복지, 돌봄, 주택정책 설계에 훨씬 유용할 수 있다.

국제적 동향을 살펴보면, 주요 기관들도 인구-가구 연계 추계를 강화하고 있다. UN[9]의 'Household Size and Composition 2022', Eurostat(2023)[10]의 Household-Based Population Projection, Statistics Canada(2025)[11]의 Integrated Demographic Microsimulation(Demosim) 등이 대표적 사례이다. 이는 인구추계 → 대상인구 도출 → 혼인상태별 전이율, 전이확률 → 가구유형 분류로 이어지는 일련의 추계과정을 하나의 통합 모델링 프로세스로 인식하는 접근이다.

다. 정책분석 도구로서의 연계 체계로 확장 필요성

현행 인구추계가 단순한 인구 규모와 구조 전망을 넘어서 실질적인 정책분석 도구로 기능하기 위해서는 다양한 정책 영역과의 연계 체계 구축이 필요하다는 판단이다. 특히 정책연구 수행을 위해서는 기존 인구추계

9) https://www.un.org/development/desa/pd/data/household-size-and-composition?utm_source=chatgpt.com.
10) https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Household_composition_statistics.
11) https://www150.statcan.gc.ca/n1/pub/17-20-0001/172000012025001-eng.htm.

보다 확장성 있는 방법론을 다루기 위한 통계 생산과 분석 시스템 구축이 필요하다.

첫째, 인구추계용 통계생산 체계 강화가 필요하다. 출생·사망 통계뿐만 아니라 결혼·이혼, 국제 이동, 건강 상태, 경제활동 참가율 등을 포함하는 종합적인 인구통계 시스템을 구축해야 한다. 또한, 보건복지통계와 패널 자료, 행정 데이터와 인구 조사 데이터를 결합하여 시계열적 일관성을 확보하고, 실시간 인구 변화를 반영하는 동태적 통계를 제공할 수 있도록 해야 한다.

둘째, 건강수명 및 노동수명 연계 추계 체계 구축이다. 고령화가 진행되면서 단순한 기대수명보다는 건강하게 살 수 있는 기간과 경제활동이 가능한 기간에 대한 전망이 더욱 중요해지고 있다. WHO의 건강수명(HALE) 방식이나 Garfinkle-Pollard 방식의 노동생명표를 활용한 수명 지표의 통합적 추계가 필요한 상황이다. 현행 추계로는 "얼마나 많은 사람이 살 것인가?"는 답할 수 있지만, "그들이 얼마나 건강하게, 얼마나 오래 일할 수 있을 것인가?"에는 답하기 어렵다.

셋째, 정책 연구를 위한 시나리오 분석 툴 개발이 필요하다. 미래 인구 변화를 다양한 사회적·경제적 요인과 연계하여 분석할 수 있는 시뮬레이션 시스템을 구축함으로써, 정책 대안별 효과를 비교·평가할 수 있어야 한다. 예를 들어, 출산율 제고 정책(보육 지원 확대, 여성 고용 안정 등), 고령층 경제활동 연장 정책(정년 연장, 노인 일자리 확대 등), 국제 이주 정책(이민 확대 또는 제한) 등 다양한 정책이 인구 구조와 노동 시장에 미치는 영향을 분석하는 모델이 필요하다.

넷째, 마이크로시뮬레이션과의 연계를 통한 정책효과 분석 체계 구축이다. 출산지원정책, 고령자 고용정책, 사회보장급여 조정 등 구체적인 정책 개입이 인구구조에 미치는 영향을 분석하고, 반대로 인구구조 변화

가 다양한 정책 영역에 미치는 파급효과를 종합적으로 평가할 수 있는 시뮬레이션 체계가 필요하다. 이를 위해 마이크로시뮬레이션과 거시 인구추계를 결합한 하이브리드 모델을 개발하는 것이 효과적인 접근이 될 수 있다.

다섯째, 정책연구 기관 고유의 추계 역량 확보이다. 현재 대부분의 정책연구는 통계청의 공식 추계를 전제로 수행되고 있으나, 정책연구의 특성상 다양한 가정과 시나리오에 따른 민감도 분석이나 정책 효과 검증을 위한 독자적 추계 도구가 필요하다. 특히 한국보건사회연구원과 같은 사회정책 전문기관의 경우, 보건·복지 정책과 밀접하게 연계된 특화된 인구추계 도구의 확보가 연구 역량 강화에 필수적이다.

이와 같은 개선 방안을 통해 장래인구추계의 신뢰성을 높이고, 정책적 활용도를 극대화할 수 있을 것이다. 특히, 한국보건사회연구원이 주도하는 인구정책 연구에서 이러한 통계 및 시뮬레이션 시스템을 적극적으로 활용할 경우, 보다 정밀한 인구 전망과 효과적인 정책 수립이 가능할 것이다.

이러한 쟁점들은 결국 기존 인구추계 방법론의 개선과 새로운 연계 체계 구축의 필요성을 시사한다. 다음 절에서는 이러한 한계를 극복하기 위한 인구·가구 통합추계의 구체적 접근 방안을 살펴보고, 이어서 정책시뮬레이션과의 연계 체계 구축 방안을 검토하고자 한다.

제2절 인구·가구 추계의 통합적 접근

1. 인구·가구 연동 메커니즘

인구와 가구는 상호 밀접한 관련성을 가지며, 인구변동이 가구구조에 미치는 영향과 가구 형성 행태가 인구변동에 미치는 역방향 효과가 동시에 존재한다. 이러한 양방향 연동 메커니즘을 이해하는 것이 통합추계 접근법의 핵심이다.

가. 인구변동 → 가구구조 변화의 주요 경로

첫째, 출산력 변화의 영향이다. 출산력 감소는 자녀 수 감소를 통해 가구원수 축소와 가구구조 단순화를 야기한다. 우리나라의 경우 합계출산율이 2023년 0.72명까지 하락하면서, 전통적인 4인 가구(부모+자녀 2명) 구조가 급속히 해체되고 있다. 이는 단순히 가구원수 감소를 넘어서 가구 내 세대 구성, 부양 관계, 경제적 의존성 등 가구의 기본적 특성을 근본적으로 변화시키고 있다.

둘째, 사망력 개선에 따른 구조적 변화이다. 평균수명 연장은 고령층 가구의 지속기간을 연장시키고, 특히 성별 기대수명 차이(2024년 기준 5.8세(남성 80.6세, 여성 86.4년))는 고령 여성 1인가구 증가로 이어진다. 이러한 변화는 단순한 가구수 증가를 넘어서 돌봄, 주거, 의료 등 다양한 정책 수요의 질적 변화를 동반한다.

셋째, 혼인 행태 변화의 파급효과이다. 만혼·비혼 증가는 1인가구 증가와 가구 형성 시기 지연을 초래한다. 2024년 기준 평균 초혼연령이 남성 33.9세, 여성 31.6세로 지속적으로 상승하면서, 전통적인 가구 형성 패

턴이 근본적으로 변화하고 있다. 특히 30대 1인가구 비율의 급증은 주택시장, 소비패턴, 사회보장제도 등에 광범위한 영향을 미치고 있다.

나. 가구 형성 행태 → 인구변동의 역방향 효과

가구 형성 비용과 출산 결정의 연계도 중요한 메커니즘이다. 가구 형성 비용 증가와 주거비 부담은 독립가구 형성을 지연시키고, 이는 결혼 연령 상승과 출산 지연으로 이어진다. 특히 서울·수도권의 주택가격 상승은 청년층의 가구 독립과 결혼을 지연시키는 주요 요인으로 작용하고 있다.

1인가구 증가의 출산력 억제 효과도 주목할 필요가 있다. 1인가구 비율 증가는 결혼 기회 감소와 출산 의향 저하를 통해 저출산을 심화시키는 요인으로 작용한다. 2023년 기준 전체 가구의 34.5%를 차지하는 1인가구는 혼인시장의 구조적 변화와 출산 동기 약화를 통해 인구변동에 직접적 영향을 미치고 있다.

다. 혼인상태별 접근의 역할

가구추계에서 혼인상태별 인구이동이 핵심적 역할을 하는 이유는 혼인상태가 가구 형성의 가장 중요한 결정요인이기 때문이다. 미혼에서 유배우로의 전이는 새로운 부부가구 형성을 의미하며, 유배우에서 사별이나 이혼으로의 전이는 1인가구 증가나 가구 해체로 이어진다.

특히 한국의 경우 법적 혼인과 출산의 강한 연계성으로 인해 혼인상태 변화가 출산력에 직접적 영향을 미친다. 2022년 기준 혼외출생아 비율이 2.9%에 불과한 우리나라의 특성상, 혼인 패턴의 변화는 곧 출산력 변화로 직결된다. 따라서 혼인상태별 전이확률의 정확한 추정과 전망이 인구·

가구 통합추계의 성공을 좌우한다고 할 수 있다.

이러한 복합적 연동 메커니즘을 종합적으로 고려하지 않고서는 미래 인구구조와 가구구조의 변화를 정확히 전망하기 어렵다. 특히 코로나19 이후 나타난 비대면 문화 확산, 재택근무 증가, 라이프스타일 변화 등은 기존 연동 메커니즘에 새로운 변수를 추가하고 있어, 보다 정교한 통합적 접근이 필요한 상황이다.

2. 혼인상태별 인구·가구추계

가. 혼인상태 접근의 필요성

혼인상태별 접근은 단순히 통계적 편의를 위한 분류가 아니라, 인구변동과 가구 형성의 핵심 동력을 이해하고 예측하기 위한 필수적인 도구라고 할 수 있다.

먼저, 혼인상태 접근이 필요한 이유 중 하나는 출산력 예측의 합리성을 제고하기 위함이다. 우리나라는 혼외 출생 비율이 극히 낮아 출산이 혼인 여부에 강하게 의존하는 구조이다. 따라서 전체 합계출산율(TFR)을 단순히 연령별로 분해하는 기존 방식으로는 만혼·비혼 증가가 출산력에 미치는 복합적 영향을 정확히 포착하기 어렵다. 혼인상태별 출산율(Marital Status-specific Fertility Rate, MSFR) 접근을 통해서만 혼인 행태 변화와 출산력 변화 간의 구조적 관계를 명확히 분석할 수 있으며, 이를 바탕으로 보다 정교한 출산력 가정을 수립할 수 있다.

둘째, 혼인상태를 활용한 접근은 가구 형성 과정의 세밀한 모델링을 가능하게 한다는 점이다. 가구 형성은 본질적으로 혼인상태 변화와 직결되어 있다. 미혼 → 유배우 전이는 신혼가구 형성을, 유배우 → 사별·이혼

전이는 가구 해체와 1인가구 증가를 의미한다. 혼인상태를 고려하지 않은 단순한 연령별 가구주율 적용으로는 이러한 동태적 과정을 적절히 반영할 수 없다.

셋째, 정책 대상 집단의 정확한 식별이다. 저출산 대응, 고령자 돌봄, 주거복지 등 주요 사회정책의 대상은 연령보다는 혼인상태에 따라 더 명확히 구분된다. 예를 들어, 동일한 30대 내에서도 미혼과 유배우의 정책 수요는 전혀 다르며, 같은 고령층에서도 유배우와 사별자의 돌봄 필요도는 현저한 차이를 보인다.

넷째, 사회경제적 변화의 인구학적 파급효과 분석이다. 교육기간 연장, 여성 경제활동 증가, 주거비 상승 등 거시적 사회경제 변화가 인구에 미치는 영향은 주로 혼인 행태 변화를 통해 매개된다. 따라서 혼인상태별 분석 없이는 이러한 구조적 변화가 장기 인구동향에 미치는 영향을 정확히 예측하기 어렵다.

나. 혼인상태 전이와 가구 형성 과정

통계청의 「2024년 장래가구추계」는 혼인상태별 전이를 체계적으로 고려한 정교한 방법론적 체계를 구축하고 있다. 전체 작성과정은 다음과 같은 단계별 절차를 따른다.

먼저, 1단계로 기준 인구 설정 및 대상인구를 정의한다. 인구총조사 자료를 기준시점인 7월 1일로 환산하여 기준 인구를 설정하고, 이를 바탕으로 장래인구추계를 산출한다. 이후 가구추계의 대상인구를 정의하기 위해, 추계인구에서 집단가구와 외국인 가구 인구 등 일반가구에 포함되지 않는 인구(제외 인구)를 제외한다. 즉, 가구추계는 일반가구만을 대상으로 하므로, 집단과 외국인 가구는 포함하지 않는다.

2단계는 혼인상태별 분류 및 전이율 추정이다. 대상인구를 혼인상태(미혼, 유배우, 사별, 이혼)별로 분류하고, 과거 자료를 바탕으로 혼인상태 간 전이율(전환건수/혼인상태별 인구)을 추정한다. 이 과정에서 곰페르츠(Gompertz) 모형과 로지스틱(Logistic) 모형을 적용하여 전이율을 보정한 후, 이를 전이확률로 변환한다.

3단계는 혼인상태별 인구 예측이다. 산출된 전이확률을 대상인구에 적용하여 미래의 혼인상태별 인구를 예측한다. 이 과정에서 9가지 전이 유형(미혼 → 초혼, 사망; 유배우 → 사별, 이혼, 사망; 사별 → 재혼, 사망; 이혼 → 재혼, 사망)이 모두 고려된다.

4단계는 가구주율 추계 및 가구수의 산출이다. 성, 연령, 혼인상태별로 수정지수모형을 적용하여 가구주율을 추계하고, 이를 바탕으로 최종적으로 성, 연령, 혼인상태별 가구수를 산출한다.

통계청(2024)은 장래가구추계를 위한 혼인상태별 전이율을 추정함에 있어, 전국 단위의 추계에는 Lee-Carter(LC) 모형을, 시도 단위의 추계에는 Lee-Li(LL) 모형을 각각 적용하고 있다. 혼인상태 간 전이확률은 이들 전이율을 바탕으로 전환공식을 통해 $T_{x,t}^{ij} = t_{x,t}^{ij} / (1 + 0.5 \sum_{i \neq j}^{k+1} t_{x,t}^{ij})$와 같이 산출된다. 이 방식은 복잡한 행렬 지수(matrix exponential) 계산 없이 간단한 근사로 전이확률을 도출할 수 있으며, 전이율의 총합이 1보다 훨씬 작은 경우에도 근사 오차가 비교적 작아 정확도가 높다는 장점을 가진다. 또한 전이확률 행렬을 구성할 때 각 행(row)의 합이 자동으로 1이 되도록 조정되는 수학적 성질을 지니고 있어, 확률 행렬의 기본 조건을 자연스럽게 충족시킨다.

다. 현행 체계의 한계와 개선 방향

지금까지의 검토를 통해 볼 때, 통계청의 현행 혼인상태별 가구추계 체계는 상당한 방법론적 정교함을 갖추고 있다. 그러나 통합추계의 관점에서 몇 가지 중요한 한계가 드러난다.

먼저, 일방향적 연계의 한계이다. 현행 체계에서는 혼인상태별 전이율 전망이 가구추계에만 활용될 뿐, 정작 인구추계 자체에는 적용되지 않는다는 점이다. 인구추계에서는 여전히 전체 출산율을 기준으로 추계가 이루어지고 있어, 혼인상태 변화가 출산력에 미치는 직접적 영향이 충분히 반영되지 못하고 있다.

둘째, 구조적 변화 대응의 한계이다. 현행 방식은 혼인행태의 급격한 변화를 반영한 다양한 시나리오 구성이 어렵다는 한계가 있다. 저위-중위-고위의 전통적 범위 설정으로는 만혼·비혼 증가 등 구조적 변화가 출산력에 미치는 복합적 영향을 충분히 탐색하기 어렵다.

셋째, 정책 변수 반영의 부족이다. 출산지원정책, 주거정책, 고용정책 등이 혼인 행태에 미치는 영향을 체계적으로 고려할 수 있는 메커니즘이 부족하다. 이로 인해 정책 효과 분석이나 정책 시나리오별 전망이 제한적일 수밖에 없다.

이러한 한계를 극복하기 위해서는 다음과 같은 방법론적 개선이 필요하다. 먼저, PPR(Parity Progression Ratio) 기반 통합 접근을 통해 성, 연령별 가구주의 가구원수별 출산진도비 접근을 통해 혼인 패턴 변화가 출산력에 미치는 영향을 보다 세밀하게 모델링할 필요가 있다. 이를 통해 혼인상태 변화와 출산력 변화를 동시에 고려한 통합적 전망이 가능해진다. 또한 정책 시나리오 반영 체계를 활용하여 주거정책, 육아지원정책, 고용정책 등이 혼인 행태에 미치는 영향을 수량화하여 모형에 반영할 수 있는

체계를 구축해야 한다. 이를 통해 정책 변화에 따른 인구·가구 구조 변화를 사전에 분석할 수 있다. 그 외로 동적 피드백 메커니즘도 필요한 요소이다. 동적 피드백 메커니즘의 도입을 통해 인구구조 변화가 혼인시장에 미치는 영향(성비 불균형, 교육수준별 매칭 등)을 보다 현실적으로 포착하고 이를 연장하여 미래 수준을 가늠케 한다. 특히 만혼과 비혼 증가, 1인가구 급증 등 최근의 급격한 사회변화가 출산력과 가구 형성에 미치는 복합적 영향을 고려할 때, 인구추계와 가구추계를 보다 유기적으로 연결하는 통합적 접근이 필요한 시점이다. 이러한 인식을 바탕으로 다음 절에서는 인구·가구 통합추계를 위한 구체적인 고려사항들을 살펴보고자 한다.

3. 인구·가구 통합추계를 위한 주요 고려사항

가. 인구·가구추계 범위 설정

인구·가구 통합추계 체계를 구축할 때 가장 먼저 결정해야 할 사항은 추계 대상의 범위이다. 이는 단순한 기술적 문제를 넘어서 추계의 정확성과 정책적 활용도에 직접적인 영향을 미치는 핵심적 선택이다.

내국인 중심 vs. 외국인 포함 접근의 선택이 첫 번째 쟁점이다. 현재 통계청의 가구추계는 집단가구와 외국인 가구를 제외한 내국인 일반가구만을 대상으로 하고 있다. 이는 혼인상태별 전이 데이터의 가용성과 신뢰성, 그리고 분석의 복잡성을 고려한 현실적 선택이다. 내국인의 혼인상태별 이동은 상대적으로 안정적인 패턴을 보이며, 혼인신고, 이혼신고 등을 통한 행정통계 기반의 정확한 추적이 가능하다.

반면 외국인을 포함하는 접근은 보다 포괄적인 인구 현실을 반영할 수 있다는 장점이 있다. 특히 국제결혼의 증가, 외국인 근로자의 정착, 다문

화가정의 확산 등을 고려할 때, 외국인을 배제한 추계는 점차 현실 설명력이 떨어질 수 있다. 그러나 외국인의 혼인상태 변화는 내국인과 다른 패턴을 보이며, 본국과의 이주, 체류자격 변경 등 복잡한 요인들이 개입하여 예측의 불확실성이 높다.

본 연구에서는 방법론적 확장성을 고려한 포괄적 접근 전략을 제안한다. 즉, 외국인을 포함하는 모델 구조를 구축하되, 현 단계에서는 외국인의 혼인상태별 이동을 내국인과 동일하다고 가정하는 방식이다. 이는 외국인에 대한 별도의 혼인상태 전이 데이터가 충분히 축적되지 않은 현실적 제약을 인정하면서도, 향후 외국인 대상의 혼인상태별 이동 모델링이 완성되면 즉시 적용할 수 있는 확장 가능한 모델 구조를 미리 확보하는 접근법이다.

이러한 접근법의 장점은 다음과 같다. 첫째, 즉시 적용 가능성이다. 내국인 대상으로 검증된 혼인상태 전이 모델을 외국인에게도 임시 적용함으로써, 현재 단계에서도 외국인을 포함한 전체 인구의 통합추계가 가능하다. 둘째, 향후 개선의 용이성이다. 외국인의 혼인 행태에 대한 충분한 데이터가 축적되면, 기존 모델 구조를 유지하면서 외국인 특화 파라미터만 교체하면 되므로 시스템 전체를 재구축할 필요가 없다. 셋째, 정책적 포괄성이다. 다문화가정, 외국인 근로자 등을 고려한 정책 수요 분석이 가능해진다.

물론 이 방식은 외국인과 내국인의 혼인 행태 차이를 완전히 반영하지 못한다는 한계가 있다. 그러나 이는 방법론 구현 차원의 현실적 선택이며, 데이터 축적과 함께 단계적으로 개선해 나갈 수 있는 발전적 한계로 이해할 수 있다. 이는 구현 가능성과 포괄성, 그리고 향후 확장성 간의 균형을 추구하기 위한 현실적 접근법이라 할 수 있다.

나. 통합 추계 방법론의 설계

1) 동시 추계 vs. 순차 추계의 선택

인구·가구 통합추계 모형 설계에서 고려해야 할 선택 중 하나는 동시 추계(simultaneous projection)와 순차 추계(sequential projection) 중 어느 방식을 택할 것인가이다.

동시 추계 방식은 인구변동과 가구변동을 하나의 통합 시스템 내에서 동시에 모델링하는 접근법이다. 이 방식의 장점은 인구와 가구 간 상호작용을 실시간으로 반영할 수 있어 이론적으로 가장 정확한 결과를 제공한다는 점이다. 예를 들어, 1인가구 증가가 결혼 기회 감소로 이어지고, 이것이 다시 출산력 저하와 가구 형성 패턴 변화로 순환되는 복잡한 피드백 효과를 동시에 고려할 수 있다.

그러나 모형의 복잡성이 높아 구현이 어렵고, 파라미터 추정과 검증이 까다로우며, 수렴성(convergence) 문제가 발생할 수 있다는 단점이 있다. 특히 혼인상태별 9가지 전이 경로와 출산진도비, 사망률 등이 동시에 상호작용하는 시스템을 안정적으로 운영하기 위해서는 상당한 기술적 복잡성을 감수해야 한다.

순차 추계 방식은 먼저 인구추계를 수행한 후, 그 결과를 바탕으로 가구추계를 실시하는 접근법이다. 현재 통계청이 채택하고 있는 방식으로, 구현이 상대적으로 용이하고 각 단계별 검증이 가능하다는 장점이 있다. 또한 기존 인구추계 체계와의 연속성을 유지할 수 있어 실무적 적용이 수월하다. 다만 인구변동이 가구 형성에 미치는 영향은 반영할 수 있지만, 가구 형성 행태가 인구변동에 미치는 역방향 효과는 충분히 고려하기 어렵다는 한계가 있다. 예를 들어, 주거비 상승으로 인한 독립가구 형성 지

연이 결혼 연령 상승으로 이어지고, 이것이 다시 출산력에 영향을 미치는 역방향 피드백은 완전히 포착하기 어렵다.

본 연구에서는 현실적 고려사항을 종합하여, 단계적 개선 전략으로 순차 추계 방식을 기반으로 하되, 주요 피드백 효과를 부분적으로 반영할 수 있는 방안을 제시하고자 한다. 구체적으로는 가구구성을 출발점으로 하는 접근법을 채택한다. 즉, 혼인상태별, 가구원 수별 가구 비중을 미리 구성해놓고, 이러한 가구구성에서 발생하는 출산을 통해 출산율 변화를 모델링하는 방식이다. 이는 가구 특성(가구주 연령, 혼인상태, 기존 자녀 수 등)이 출산 결정에 미치는 직접적 영향을 체계적으로 반영할 수 있게 해준다.

이 접근법의 장점은 다음과 같다. 첫째, 구현의 현실성이다. 복잡한 동시방정식 체계를 구축하지 않고도 가구구성과 출산력 간의 핵심적 연계를 모델링할 수 있다. 둘째, 정책적 활용도이다. 가구 유형별 출산율 차이를 명확히 파악함으로써, 가구 특성에 따른 차별화된 출산지원정책 설계가 가능해진다. 셋째, 향후 확장 가능성이다. 출산이 다시 가구구성에 미치는 역방향 피드백 효과는 기술적 복잡성을 고려하여 현 단계에서는 제외하되, 향후 가구추계 개선 시 이러한 피드백 메커니즘을 단계적으로 추가할 수 있는 구조를 유지한다.

이러한 방식은 “가구 → 출산”의 일방향 연계를 우선 완성하여 방법론적 안정성을 확보한 후, 장기적으로 “출산 → 가구”의 역방향 피드백까지 포함하는 완전한 통합 모델로 발전시켜 나가는 단계적 개선 전략이며, 모형의 구현 가능성과 정확성, 그리고 향후 발전 가능성 간의 균형을 추구하는 현실적 접근법이라고 할 수 있다.

2) 통합추계 핵심 변수: 출산진도비와 사망률

인구·가구 통합추계에서 출산진도비(Parity Progression Ratio, PPR) 접근법의 활용은 출산력과 가구구조 변화를 연결하는 핵심 고리 역할을 한다. PPR은 특정 출산순위에서 다음 순위로 진행할 확률을 의미하며, 이는 가구원수 변화를 직접적으로 결정하는 요인이다.

성, 연령별 가구주의 가구원수별 PPR 적용을 통해 혼인 타이밍이 출산 패턴에 미치는 영향을 정교하게 모델링할 수 있다. 예를 들어, 초혼 연령이 늦어질수록 첫째 자녀 출산 확률이 감소하고, 둘째 자녀로의 진행 확률도 함께 낮아지는 패턴을 가구주 연령별 PPR 변화로 반영할 수 있다.

PPR 접근법의 주요 장점은 먼저, 혼인상태와 출산력의 연동 모델링이다. 만혼·비혼 증가가 출산력에 미치는 영향을 직접적으로 모델링할 수 있어, 단순한 TFR 기반 접근보다 훨씬 정교한 분석이 가능하다. 두 번째로, 가구유형별 추계 제공이다. 무자녀 가구, 1자녀 가구, 2자녀 가구 등의 구체적 전망을 통해 보육정책이나 육아지원정책의 수요 예측에 직접 활용할 수 있는 정보를 제공한다. 일반적으로 출산력 평가는 주로 합계출산율(TFR)과 연령별 출산율(ASFR)에 기반하여 이루어진다. 이러한 지표들은 특정 시점의 출산행태를 간명하게 요약할 수 있는 장점을 가지며, 정책 지표로도 널리 활용되어 왔다. 그러나 이들은 출산의 시간적 분포(tempo)에는 민감하나, 출산의 양(quantum)이나 출산 순위별 구조와 같은 본질적인 출산행태의 이질성을 충분히 반영하지 못한다는 한계를 지닌다. 특히 출산이 특정 계층에 편중되는 경향이 심화되고 있는 현실에서는 평균적인 지표만으로는 사회 전반의 가구구조 변화나 향후 인구 동태를 정확히 파악하기 어렵다. 그에 비해 PPR은 상대적으로 보다 정밀한 출산력 분석을 가능케 하는 대안적 접근이다. PPR은 여성의 출산 순위별

이행률, 즉 무자녀 상태에서 첫째 자녀로, 첫째에서 둘째 자녀로의 이행 확률을 구조적으로 포착한다. 이는 단순히 자녀 수의 평균값을 산출하는 것이 아니라, 출산 분포의 누적 경로를 계량화함으로써 개인의 생애 출산 궤적을 보다 세밀하게 분석할 수 있게 해준다. 특히 생존모형(discrete-time survival model)을 활용한 PPR 분석은 출산력 예측을 단순한 시계열적 외삽에서 벗어나 인과적 추정과 누적 이행 경로를 바탕으로 한 구조적 분석으로 전환시킨다. 예컨대 complementary log-log 링크를 도입한 이산형 생존모형을 통해 각 출산 순위별 이행률을 추정하고 이를 곱함으로써 TFR을 재구성할 수 있다.

정책적으로도 PPR 기반 분석은 중요한 함의를 갖는다. 예를 들어 전체 출산율은 정체되어 있으나, PPR_0(무자녀→첫째 출산 이행률)의 하락이 주요 원인이라면 출산 정책은 첫째 자녀 출산 진입을 위한 제도적 장벽 해소나 경제적 지원에 집중해야 할 것이다. 반면 PPR_1, PPR_2 등의 하락이 주원인이라면 다자녀 유지에 대한 장기적 지원이나 일-가정 양립 정책이 보다 효과적일 수 있다. 즉, 출산진도비는 단일 합계출산율, 연령별 출산율과 같은 평균값 지표가 가질 수밖에 없는 구조적 정보 상실을 보완하여, 실질적이고 차별화된 출산정책 수립을 가능케 하는 지표이다. 나아가 PPR 분석은 단순한 인구예측 도구를 넘어, 가구구조 예측모형으로의 확장이 가능하다. 자녀 수에 따라 가구 유형이 달라지고, 이는 사회 전반의 1인가구, 한부모 가구, 다자녀 가구 등의 변화에 결정적 영향을 미친다. 이에 따라 출산진도비 기반의 출산력 전망은 향후 인구-가구 통합 시뮬레이션 모델의 주요한 구성요소로 기능할 수 있다.

다. 데이터 측면

정확한 장래인구추계를 위해서는 고품질의 인구 변동 요인에 대한 인구동태통계(vital statistics) 생산과 체계적인 관리가 필수적이다. 현재 인구추계는 통계청의 인구센서스(등록센서스), 출생·사망 통계, 국제 이동 자료 등을 기반으로 이루어지지만, 보다 정밀한 분석을 위해 인구 속성과 경제·사회 분야별로 세분화된 미시(micro) 자료 확보가 요구된다.

1) 통합 데이터 구축의 필요성

특히, 연령별·지역별 인구 이동 데이터, 출산 및 사망의 사회경제적 요인, 결혼·가구 형태 변화에 대한 장기 시계열 통계를 구축할 필요가 있다. 이를 위해 행정자료, 패널자료, 대규모 설문조사 데이터를 연계·통합하여 활용하는 방안이 효과적이다. 이러한 통계자료의 대표적 사례가 통계등록부(statistical register)이다. 통계등록부는 행정자료와 조사자료를 연계하여 만든 개인·기업 등의 명부 및 정보를 의미하고, 청년통계, 인구통계, 기업통계 등과 관련된 등록부가 있다. 청년통계등록부는 15~34세 인구의 지역, 성별, 연령, 취업 여부, 임금소득, 1인가구 여부, 배우자 유무, 거처 형태 등의 특성을 포함한다. 이는 만혼·비혼 증가의 핵심 연령대에 대한 종합적 정보를 제공하여 혼인상태 전이 모델링에 직접적으로 활용할 수 있다. 인구통계등록부는 생애주기와 취약계층 맞춤형 정책 분석을 지원하기 위해 맞벌이, 다문화, 한부모 가구, 주택 소유 여부 등 관련 핵심 항목을 통합한 데이터이다. 이는 가구 형성과 해체의 사회경제적 배경을 분석하는 데 핵심적 자료가 된다. 기업통계등록부는 다양한 기관의 행정자료와 조사자료를 연계하여 구축된 기업 및 사업체 단위의 모집단 자

료로, 사업자 등록번호, 대표자 성별, 주소, 종사자 수, 매출액, 산업 분류 등의 정보를 포함한다. 이는 고용 안정성과 소득 수준이 혼인 결정과 출산 타이밍에 미치는 영향을 분석하는 데 중요한 자료이다.

2) 데이터 활용 체계 구축

이러한 데이터는 통계데이터센터(Statistical Data Center, SDC)[12]에서 제공되고 있으며, 보다 효율적인 인구추계를 위해 적극적으로 활용될 필요가 있다. 특히 개인정보 보호와 연구 활용 간의 균형을 확보하면서, 연구자들이 안전하게 접근할 수 있는 환경을 조성하는 것이 중요하다. 데이터 연계의 기술적 과제로는 서로 다른 출처의 데이터 간 식별체계 통일, 시점 불일치 조정, 누락값 처리 등이 있다. 이를 해결하기 위해서는 표준화된 데이터 처리 및 품질 관리 체계의 구축이 필요하다.

향후 출산력 전망체계는 합계출산율, 연령별 출산율과 같은 단일 평균 지표에서부터 시작해 출산 분포의 이행 경로와 구조적 변동성까지 포착할 수 있는 출산진도비 기반의 정교한 분석까지 포함되어야 한다. 이를 위해서는 통계청과 같은 국가통계기관이 합계출산율과 연령별 출산율 중심의 공식 통계 외에도 출산진도비(PPR) 산출 체계를 공식화하고, 이를 기반으로 정책 목표와 시뮬레이션 예측에 활용할 수 있는 체계를 구축하는 것이 필요하다.

출산의 '누적성과', '진입경로', '이행률'에 대한 이해 없이는 향후 지속가능한 출산정책의 설계가 불가능하다는 점에서, 출산진도비는 향후 출산력 분석과 정책 설계의 새로운 표준이 될 수 있다.

---

12) 통계데이터센터는 외부와 차단된 공간에서 통계등록부와 통계자료 등의 데이터를 분석할 수 있는 '데이터 안심존'으로, 누구나 신청 후 이용할 수 있다. 현재 전국 13개소에서 운영 중이다.

제3절 인구추계와 정책시뮬레이션 연계 체계

1. 연계의 필요성과 마이크로시뮬레이션 모형의 특성

기존의 장래인구추계는 주로 거시적 접근 방식의 코호트요인법을 기반으로 하지만, 개별 행위자의 인구학적·사회적 특성을 반영하는 마이크로시뮬레이션(micro simulation) 기법을 도입함으로써 보다 정교한 인구 전망이 가능하다.

마이크로시뮬레이션은 개별 개인과 가구 단위에서 출생, 사망, 혼인, 이혼, 이동, 경제활동 참여 등의 생애 이벤트 변수를 모의 실험하여 보다 세밀한 인구 변화 양상을 예측할 수 있도록 한다. 이를 통해 특정 인구 집단(예: 1인가구 증가, 고령층의 경제활동 지속 여부)의 변화가 장기적으로 인구 구조에 미치는 영향을 보다 정확하게 분석할 수 있다. 또한, 이러한 기법을 활용하면 지역별, 연령별, 사회경제적 계층별 인구 변화를 보다 정교하게 예측할 수 있어 정책적 활용도가 높아진다. 앞서 1절에서 살펴본 현 인구추계의 한계들(한계연령 100세 제한, 초저출산 대응 부족, 정책분석 도구로서의 제약)을 극복하기 위해서는 단순한 인구 전망을 넘어선 정책효과 분석이 가능한 시뮬레이션 체계가 필요하다. 특히 사회보장제도의 지속가능성 분석을 위해서는 단순한 인구수 변화만으로는 불충분하다. 연금 재정추계, 의료비 지출 전망, 돌봄서비스 수요 예측 등을 위해서는 건강상태, 경제활동 참여, 소득수준 등을 종합적으로 고려한 미시적 분석이 필수적이다. 또한 출산지원정책, 고령자 고용정책, 이민정책 등 구체적인 정책 개입이 인구구조에 미치는 영향을 사전에 분석하고, 반대로 인구구조 변화가 다양한 정책 영역에 미치는 파급효과를 종합적으로 평가할 수 있는 정책시뮬레이션 도구가 필요한 상황이다.

다만 이런 시뮬레이션을 구현하기 위해서는 인구 관련 빅데이터가 필수적인 반면 개인정보보호로 자료 연계와 활용이 제한적이다. 이런 환경에서 고려해 볼 수 있는 자료는 재현자료(synthetic data)이다. 인구추계에서는 인구센서스, 행정자료, 설문조사 데이터 등이 활용된다. 그러나 개인정보보호 문제로 인해 원 데이터를 직접 사용하기 어려운 경우가 많다. 이러한 한계를 극복하기 위해, 원 데이터의 통계적 특성을 유지하면서도 특정 개인을 식별할 수 없는 재현자료를 생성하여 활용할 수 있다. 재현자료 활용의 장점으로는 원 데이터를 직접 사용하지 않으면서도 동일한 분석 결과를 도출할 수 있어 개인정보보호가 가능하며, 연구자들이 민감한 행정자료에 접근하지 않고도 공공기관에서 제공하는 재현자료를 이용하여 인구추계를 수행할 수 있다. 또한 원 데이터를 기반으로 가상의 개별 행위자(agent)를 생성하여 출산, 사망, 국제 이동, 혼인, 이혼 등의 인구학적 사건을 모의 실험(시뮬레이션)할 수 있다.

재현자료 기반 인구추계를 구축하기 위해서는 먼저, 원 데이터 확보와 자료 처리가 필요하다. 인구센서스, 행정자료, 사회조사 데이터 등 원 데이터를 기반으로 인구학적 특성을 정리하고, 출생, 사망, 국제 이동, 결혼, 이혼 등 주요 인구 변동 요인을 반영할 변수를 선택한다. 두 번째로, 재현자료의 생성이다. 원 데이터의 분포를 반영하여 통계적 기법(기계 학습, 다변량 분포 모델 등)을 활용하여 가상의 개체(individual)를 생성한다. 대표적인 방법으로 다변량 회귀 분석, 베이지안 네트워크 모델, 의사결정 트리 기반 방법이 활용될 수 있다. 세 번째 단계로는 개별 행위 기반 마이크로시뮬레이션 적용이다. 재현된 개체별로 미래의 출산, 사망, 이주 등의 인구 변동을 모의 실험(시뮬레이션)한다. 예를 들어 각 개체의 인구학적 상태 변화(예: 20대 여성의 결혼 가능성, 65세 이상 인구의 기대여명 변화 등)를 확률적으로 예측할 수 있다. 마지막으로, 시나리오 분석 및

정책 평가이다. 인구 변동 요인의 변화에 따른 인구 구조 변화를 예측하고, 출산율 증가, 고령화 속도, 국제 이주 확대 등의 정책적 변수가 반영된 다양한 시나리오를 적용하여 정책 효과 평가가 가능하다. 현재 다양한 재현자료 생성 소프트웨어가 개발되어 있어 실무적 활용이 가능한 상황이다(〈표 2-2〉 참조).

특히 한국보건사회연구원은 인구 및 사회정책 연구를 수행하는 기관으로, 인구추계용 통계 생산과 정책 연구를 위한 시나리오 분석 툴 개발이 필요하다. 이를 위해 재현자료를 활용한 인구추계 기법을 도입할 경우 다음과 같은 이점이 있다.

첫째 인구 정책 연구의 정밀성 강화 차원에서 개별 행위 기반 시뮬레이션을 활용하여 출산율 변동, 국제 이주, 고령화 등의 요인이 인구 구조에 미치는 영향을 보다 정밀하게 분석할 수 있다.

둘째 정책 시나리오 분석 시스템 구축이다. 출산 장려정책, 정년 연장 정책, 이민 확대정책 등의 효과를 다양한 가정하에서 평가할 수 있는 인구 시뮬레이션 모델을 개발할 수 있다.

셋째 보건·복지 정책 연구에 활용이다. 의료 서비스 수요, 노인 돌봄 수요, 연금 재정 지속 가능성 등의 장기적인 사회적 영향을 분석하는 데 활용할 수 있다.

〈표 2-2〉 재현자료 생성 소프트웨어

소프트웨어	내용	홈페이지 URL
Data -Synthesizer	차등보호(Differential privacy) 기술을 활용한 재현데이터 생성 오픈소스코드로, Django를 활용한 UI 앱도 제공함	https://github.com /DataResponsibly/ DataSynthesizer
Gretel	미국 샌디에이고에 위치한 재현데이터 생성 스타트업으로 LSTM, DGAN, Diffusion 등의 딥러닝 방법을 활용하여 관계형, 시계열, 비정형, 이미지 재현데이터를 생성하는 서비스를 제공함.	https://gretel.ai/

소프트웨어	내용	홈페이지 URL
PopGen	Arizona State University의 SimTRAVEL Research Initiative에서 개발되었으며, 상대적으로 전수 정확도가 높은 반복비율갱신(Iterative Proportional Updating) 알고리즘으로 전수 인구 데이터 생성이 가능	https://www.mobilityanalytics.org/popgen.html
SDV	datacebo에서 관리하는 재현데이터 생성 공개소스 코드로, 개별, 관계형, 시계열 데이터에 대한 재현데이터를 생성함. 생성된 데이터에 대한 평가 및 시각화 모듈도 제공함. 소스코드 공유 사이트 github에서 1,500개 이상의 star를 받을 정도로 높은 대중적 인지도를 가지고 있음	https://sdv.dev/
simPop	주체가 가진 속성값에 따라 다르게 적용되는 정책의 거시적인 효과를 예측하기 위한 복잡한 구조의 데이터 재현에 매우 유용함. 가구와 가구 구성원 정보 등 계층적구조 처리가 가능함. IPF와 SA를 사용한 통계량 조정, 로지스틱 회귀를 통한 모델링 기능을 제공함	https://cran.r-project.org/web/packages/simPop/index.html
sms	주어진 영역 내 매크로 데이터로부터 마이크로 데이터를 시뮬레이션하는 기능을 제공함. 계층적 구조의 데이터 처리는 불가능하나, Simulated Annealing을 단순화하여 제한된 영역에 대한 설명을 최적화하는 기능이 존재함	https://cran.r-project.org/web/packages/sms/index.html
Synthia	비영리 연구기관인 RTI에서 개발한 웹 기반 재현데이터 생성 프로그램으로, 사용자 정의 변수를 사용하여 사용자가 정의한 학습 영역에 대한 재현데이터를 생성함	https://synthia-dataset.net/
Synthpop	분류 회귀모형을 사용하여 재현데이터에 대한 변수를 생성함. 정교한 샘플링 디자인을 필요로 하거나 가구 및 구성원 정보 등과 같은 계층 또는 클러스터 구조를 가진 데이터를 처리하는 기능은 없으나 편의성이 높음	https://www.synthpop.org.uk/
TRANSIMS	미국 Los Alamos National Laboratory의 연구원이 개발한 운송 분석 시뮬레이션 시스템임. 인구조사 마이크로 데이터를 기반으로 재현데이터를 생성함	https://sourceforge.net/projects/transims/

자료: 김현태, 장가영 (2023);김승현 (2020) 활용 저자 정리.

2. 정책시뮬레이션 연계 체계 구축을 위한 주요 고려사항

가. 정책시뮬레이션 연계 범위 설정

인구추계와 정책시뮬레이션의 연계 범위를 설정할 때는 기술적 구현 가능성과 정책적 필요성 간의 균형을 고려해야 한다.

단계별 연계 범위 확장 전략을 통해 초기에는 핵심 기능에 집중하되, 점진적으로 연계 범위를 확대해 나가는 접근이 바람직하다. 구체적인 우선순위는 다음과 같다.

1단계: 인구·가구 기본 구조 연계에서는 성별, 연령별, 혼인상태별 인구 구조와 가구 유형, 가구원 수, 가구주 특성을 중심으로 한 연계를 우선 구축한다. 출생, 사망, 혼인상태 변화 등 기본적 인구학적 사건의 연계도 이 단계에 포함된다. 이는 양 시스템 간 연계의 기본 토대를 구축하는 단계로, 가장 안정적이고 검증 가능한 영역이다.

2단계: 건강 및 노동 관련 확장에서는 건강수명, 건강상태별 인구 분포, 경제활동 참여율, 노동수명을 포함한다. 또한 교육수준, 소득분위별 특성도 이 단계에서 연계한다. 이는 사회보장정책 분석의 핵심 요소들로, 1단계가 안정화된 후 확장하는 것이 적절하다.

3단계: 정책 효과 분석 모듈에서는 출산지원정책, 고용정책, 사회보장정책이 인구구조에 미치는 영향 분석과 정책 시나리오별 장기 파급효과 추정을 다룬다. 이는 가장 복잡하고 고도화된 분석 기능으로, 앞선 단계들의 안정적 운영을 바탕으로 구현되어야 한다.

나. 연계 방법론의 설계와 주요 변수

KIHASA-SIM과의 실질적 연계를 위해서는 체계적인 기술적 원칙과 단계적 발전 전략이 필요하다.

기술적 연계 원칙으로는 먼저 시간 및 연령 단위의 일치가 핵심이다. 양 모형 모두 1년 단위, 1세 구간의 연령 코호트를 기준으로 설정하여 자료 변환의 필요성을 최소화해야 한다. 집계값 정합성 유지도 중요한데, 연계 이후 KIHASA-SIM의 결과가 매년 추계된 총인구나 가구수와 불일치할 경우 집계값을 조정함으로써 정합성을 유지해야 한다.

모듈 독립성 보장 측면에서는 외생 입력값이 각 모듈의 로직과 알고리즘을 가능한 한 보존하는 방식으로 주입되어야 하며, 연계 과정이 각 모듈의 시뮬레이션 알고리즘에 영향을 미치지 않도록 설계되어야 한다. 재현가능성 확보를 위해서는 입력자료, 난수 시드(seed) 등 모형 간 연계에 사용된 모든 요소를 통제하고, 버전 관리 체계를 확립하여 시나리오 결과의 재현 가능성을 확보할 필요가 있다.

연계 구조의 단계적 발전에서는 초기에 인구추계 → KIHASA-SIM으로의 단방향 흐름으로 시작하되, 일정 기간의 검증을 거친 이후 KIHASA-SIM에서 도출된 노동 공급, 소득분포 등의 결과를 인구추계 모형의 사회경제적 변수로 환류시키는 순환형 피드백 구조로 확장하는 것이 바람직하다. 초기 단계에서는 인구 및 가구만 연계하지만, 추후 건강기대여명, 건강노동기대여명 등의 산출물을 점진적으로 다른 모듈과 연계하여 적용 범위를 확대할 수 있다.

기술적 과제와 대응 방안을 살펴보면, 이러한 연계는 단순한 산출물 주입 수준이 아닌 고도의 기술적 설계와 검증을 요구하는 복합적 과정이다. 데이터 구조의 차이 문제가 가장 핵심적인데, 인구추계 모형의 간단한 연

령×성별 행렬 구조와 KIHASA-SIM의 복잡한 시뮬레이션 로직 간 통계적 매핑과 스케일링의 정합성 확보가 중요하다.

시스템 상호의존성 측면에서는 KIHASA-SIM의 모듈 간 높은 상호의존성으로 인해 출생률이나 사망률의 조정만으로도 전체 시뮬레이션 결과가 크게 변동할 수 있다는 점에 유의해야 한다.

핵심 연계 변수의 체계적 설계가 필요한데, 두 시스템 간 연계를 위해서는 핵심 연계 변수를 명확히 정의해야 한다. 인구학적 기본 변수(성별, 연령, 혼인상태, 가구 내 지위), 사회경제적 변수(교육수준, 경제활동상태, 소득분위), 건강 관련 변수(건강상태, 의료서비스 이용 여부), 정책 관련 변수(사회보험 가입 여부, 복지서비스 수급 여부) 등이 포함된다.

제3장

출산력 추계 방법론 개선

제1절 기존 출산력 추계 방법론 검토
제2절 출산력 추계 방법론 개선 방안
제3절 출산력 추계모형 구축 및 결과 검토

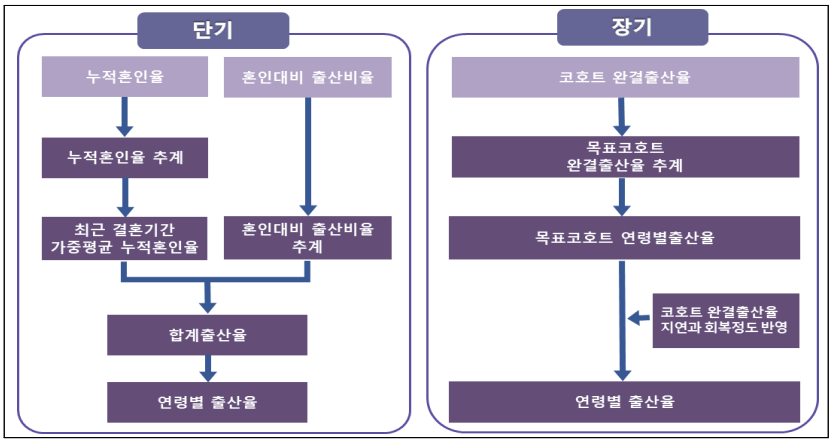
제3장 출산력 추계 방법론 개선

제1절 기존 출산력 추계 방법론 검토

1. 통계청 방법론 분석

통계청(2024)에서 공표하는 출생 추계 방법은 먼저 인구동향조사 출생 통계 자료에 기반을 둔 기간 출산 분석, 코호트 출산추이 분석, 혼인과 출생 추이 분석을 한 후 단기적으로 혼인 추세를 반영하고 장기적으로 코호트출산율 회복을 반영해 추계한다. 〔그림 3-1〕은 출생 단기와 장기 추계 방법을 보여준다.

〔그림 3-1〕 출생 단기 및 장기 추계방법



자료: 통계청(2024) 장래인구추계, 2022~2072.

출생 추계의 기초자료는 다음과 같다. 통계청은 2014~2023년 15~39세 연령별 혼인율, 2014~2023년 가중평균 만 39세까지 누적혼인율 대비 출산순위별(1아, 2아, 3아 이상) 출산비율, 2013~2022년 출산순위별 모의 결혼기간 비중, 1948~1987년생 코호트완결출산율, 그리고 1955~1987년생 평균출산연령이다. 이 자료를 기반으로 통계청(2024)은 최근 혼인율과 혼인 대비 출산비율의 추세를 활용하여 단기적으로 합계출산율을 식 (2)에 따라 예측한다.

$$TFR_t = \sum_i M_{i,t} b_{i,t} + \epsilon_t \tag{2}$$

$$\begin{cases} TFR_t : t\text{시점 합계출산율} \\ M_{i,t} : t\text{시점 출산순위 } i\text{별 최근 혼인율} \\ b_{i,t} : \text{혼인한 여성1명당 } t\text{시점의 출산순위}i\text{번째 아이 출산비율} \end{cases}$$

혼인율은 연도별 39세까지의 누적혼인율을 기반으로 산출하며, 출산순위별(1아는 최근 3년, 2아는 최근 2~6년, 3아 이상은 최근 4~9년) 결혼기간별 비중을 가중 평균하여 계산한다. 혼인 대비 출산비율은 '출산순위별 출산율을 출산순위별 최근 결혼기간 가중평균 누적혼인율로 나눈 값'으로 정의한다. 통계청의 출생 추계는 다음의 5단계 절차에 따라 이루어진다. 첫째, 연도별 누적혼인율은 Gompit 변환을 적용한 후 Lee-Carter(LC, 1992) 모형을 이용하여 예측한다. 여기서 Gompit 변환은 $-\ln(-\ln(M_x))$, M_x는 x세까지의 연령별 누적혼인율에 대한 변환 방식이다. 둘째, 출산순위별 결혼기간별 비중을 가중치로 적용하여 만 39세까지의 가중평균 누적혼인율을 출산순위별로 산출한다. 셋째, 출산순위별 혼인 대비 출산비율은 LC 모형을 이용하여 예측한다. 넷째, 식 (2)를 사용하여 누적혼인율과 혼인 대비 출산비율의 곱을 합산하여 합계출산율, TFR_t를 산출한다. 마지막으로, 연령별 출산율 분포는 식 (3)과 같은 일반화 로그감마(Generalized log gamma)함수를 이용해 산출한다.

$$f_i(x) = \frac{C_i\,|\lambda|}{b\Gamma(1/\lambda^2)}\left(\frac{1}{\lambda^2}\right)^{\lambda^{-2}} \exp\left[\frac{1}{\lambda}\left(\frac{x-u}{b}\right) - \frac{1}{\lambda^2}\exp\lambda\left(\frac{x-u}{b}\right)\right] \quad (3)$$

· C : 출산순위별 출산확률 (출산순위을 고려하지 않는 경우는 합계출산율과 동일함)
· u : 평균출산연령[13)]
· b : 평균출산연령의 표준편차
· λ : 분포 양상

$f_i(x)$는 출산순위(i)별 연령별 출산율이다. 이 함수는 C, μ, b, λ과 같은 모수 4개를 포함한다. 그리고 15~49세까지 출산순위별 연령별 출산율은 이 함수에 대입하여 산출한다.

장기적으로는 2022년 기준 15세인 목표 코호트(2007년생)의 완결출산율을 식 (4)에서 제시된 1차 자기회귀 시계열 모형으로 예측하고, 출산 지연과 회복 과정을 반영한 코호트출산율 회복을 포함한다. 이 과정은 여섯 단계의 절차를 통해 이루어진다.

$$CFR_t = \beta_0 + \beta_1 \ln(t) + \epsilon_t, \quad \epsilon_t = \phi_1 \epsilon_{t-1} + a_t \quad (4)$$

첫째, 목표 코호트인 2007년생의 완결출산율은 1948~1987년생의 완결출산율을 기초로 1차 자기회귀 시계열 모형을 사용하여 예측한다. 둘째, 목표 코호트의 연령별 출산율 분포는 식 (3)을 이용하여 산출한다. 목표 코호트 이후에 태어난 출생 코호트의 연령별 출산율은 2007년생 코호트와 동일하다는 가정을 적용한다. 셋째, 출산 지연과 회복 정도를 계량화하기 위해 1988~1992년생 코호트를 대상으로 준거, 표준, 특정의 세 가지 코호트로 구분하여 완결출산율을 예측한다. 준거코호트는 1969년생 이후 코호트 중 출산의 감소 및 지연 정도가 일관되고 명확히 나타나

13) 통계청(2024)에 따르면, 단기모형의 평균출산연령은 최근 증감추이를 적용하고, 장기모형의 목표코호트 평균출산연령은 1955~1987년생의 연평균 증가속도(0.10세)*가 지속될 것으로 가정, 저위는 1955~1987년생의 연평균 증가속도(0.14세)**를 반영하여 가정, 고위는 중위와 저위의 차이 범위(0.77세)로 설정.
*전년 대비 0.2세 이상 증가 제외, **전년 대비 0.3 이상 제외

는 코호트를 의미한다. 표준코호트는 1975년생과 1980년생을 지칭하며, 이 코호트는 2020년에 각각 40세에 도달하였다. 특정코호트는 식 (5)를 활용하여 산출한다. 이와 같은 방법론을 통해 목표 코호트의 장기 출산율 예측이 이루어진다.

$$CFR_c(49\text{세}) = \{CFR_b(40\text{세}) + d_c(40\text{세})\} \times \frac{CFR_{last}(49\text{세})}{CFR_{last}(40\text{세})} \quad (5)$$

식 (5)의 $CFR_c(49\text{세})$는 특정코호트 완결출산율, $CFR_b(40\text{세})$는 준거코호트 40세 누적출산율, $d_c(40\text{세})$는 특정코호트 40세에서의 출산 지연 수준이다. 그리고 $CFR_{last}(40\text{세})$, $CFR_{last}(49\text{세})$는 기준연도 40, 49세의 출산율이 관찰되는 코호트 중 가장 최근(last) 코호트의 해당 연령 누적출산율이다. 또한 $d_c(40\text{세})$는 식 (6)에 의해서 도출한다.

$$d_c(40\text{세}) = d_n(m)PR_c(m) + r_n(40\text{세})PR_c(40\text{세}) \quad (6)$$

여기서 m은 저점연령으로 준거코호트와 특정코호트의 누적출산율 차이가 최대가 되는 연령이다. $PR_c(m) = d_c(m)/d_n(m)$은 지연비로 저점연령 m세에서 표준코호트와 준거코호트 간 누적출산율 차이 대비 특정코호트와 준거코호트 간 누적출산율 차이이다. $PR_c(40\text{세}) = r_c(40\text{세})/r_n(40\text{세})$은 회복비로 표준코호트와 준거코호트 간 누적출산율 차이의 회복 수준 대비 특정코호트와 준거코호트 간 누적출산율 차이의 회복수준이다. 그리고 $r_c(40\text{세})$는 지연된 출산의 40세 회복수준으로 $d_c(40\text{세}) - d_c(m)$로 계산된다. 통계청에 따르면 지연비 추계는 직교다항회귀식, 회복비 추계는 홀트지수평활방법으로 추세를 연장한다고 밝히고 있다.

넷째, 1988~1992년생 코호트의 연령별 출산율 분포는 식 (3)을 사용하여 산출하고 이를 반영한다. 다섯째, 1992년생 코호트와 목표코호트

(2007년생)의 연령별 출산율은 지수보간법을 통해 완결출산율을 추정한다. 이후, 완결 코호트의 연령별 출산율을 기간 연령별 출산율로 전환한다. 여섯째, 2037~2041년 기간에는 1988~1992년생 코호트의 출산 회복 정도를 반영한 연령별 출산율을 적용한다. 지금까지 단기 및 장기 추계 방법의 세부 단계 과정을 검토하였다. 마지막으로, 통계청(2024)은 단기와 장기 추계를 연결하기 위해 식 (7)에서 제시된 방식을 사용했음을 밝히고 있다.

$$TFR_t = \sum_{i=15}^{49}(1-w_t)ASFR_{s,i} + \sum_{i=15}^{49} w_t ASFR_{l,i} \tag{7}$$

$$\begin{cases} ASFR_{s,i} : s\text{시점 연령}i\text{ 출산율}, \quad w_t = \dfrac{t-s}{l-s} \\ s : \text{단기}(short)\text{추계 마지막 연도} \\ l : \text{장기}(long)\text{추계 시작 연도 2037년} \end{cases}$$

이들 과정에 따른 2023~2072년 합계출산율, 1970~2072년 합계출산율 추이, 그리고 2000~2072년 연령별 출산율은 〈표 3-1〉, [그림 3-2], [그림 3-3]에 제시되어 있다.

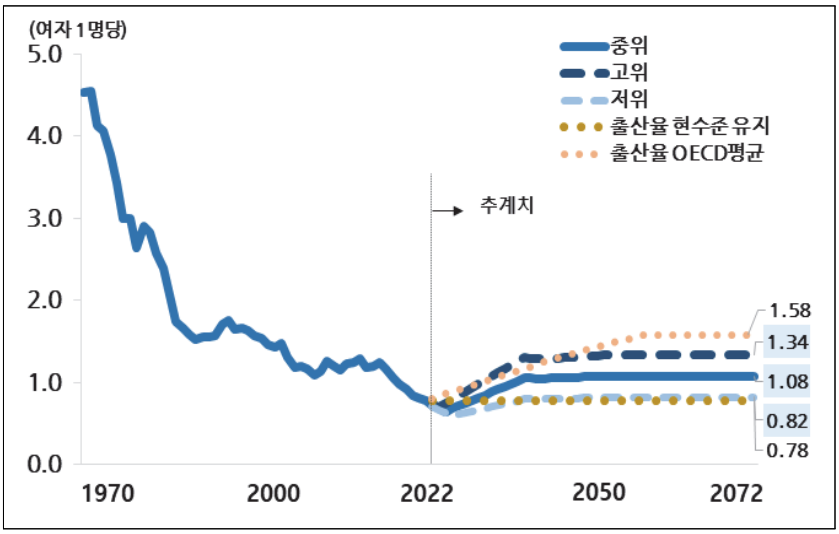
〈표 3-1〉 2023~2072년 합계출산율

(단위: 가임여자 1명당)

가정	2024	2025	2026	2027	2028	2029	2030	2035	2040	2050	2072
중위	0.68	0.65	0.68	0.71	0.75	0.78	0.82	0.99	1.05	1.08	1.08
고위	0.70	0.75	0.80	0.84	0.89	0.93	0.98	1.21	1.30	1.33	1.34
저위	0.67	0.63	0.59	0.61	0.63	0.65	0.67	0.76	0.80	0.82	0.82
출산율 현수준	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
출산율 OECD 평균	0.83	0.86	0.88	0.91	0.94	0.96	0.99	1.12	1.25	1.47	1.58

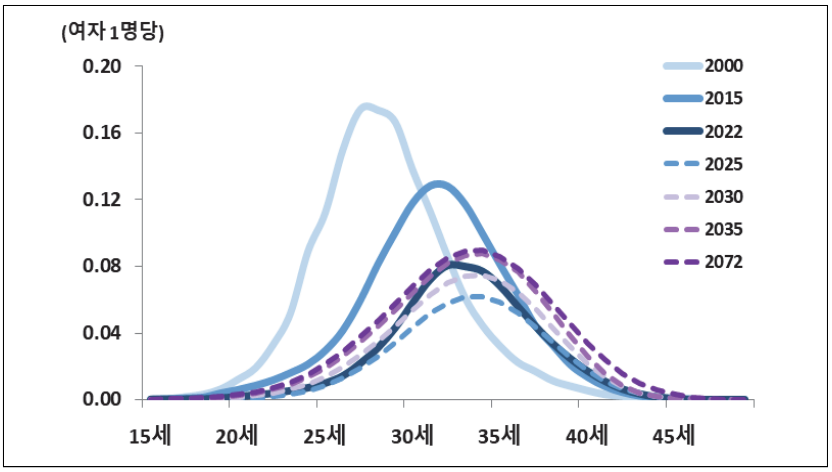
자료: 통계청(2024), 장래인구추계 2022~2072.

〔그림 3-2〕 1970~2072년 합계출산율 추이



자료: 통계청(2024), 장래인구추계 2022~2072.

〔그림 3-3〕 2000~2072년 연령별 출산율



자료: 통계청(2024), 장래인구추계 2022~2072.

2. 출산률 추계 선행연구 검토

가. 연령별 출산율 예측 모형

합계출산율은 가임연령(15~49세) 여성 인구의 연령별 출산율을 모두 합산한 값이다. 이 지표는 연령별 출산율을 연속적인 개념으로 보면 적분 기호로, 출산을 개별적인 사건으로 보면 합계(시그마, Σ) 기호로 표현할 수 있다. 연령별 출산율(age-specific fertility rate, ASFR)은 $f_i = B_i / P_i^f,\ i = 15, 16, \cdots, 49$세이고 이들을 모두 합하면 합계출산율이다. 따라서 이산, 연속적 관점을 적용하여 합계출산율을 표시하면,

$$TFR = \sum_{i=15}^{49} f_i = \sum_{i=15}^{49} \frac{B_i}{P_i^f} \tag{8}$$

$$TFR = \int_{15}^{49} f(x)dx = \int_{15}^{49} \frac{B(x)}{P^f}(x)dx \tag{9}$$

여기서 f_i와 $f(x)$는 1세별(x세) 출산율, B_i와 $B(x)$는 1세별(x세) 여자인구의 출생건수, P^f나 P_i^f는 1세별(x세) 여자인구 수이다. 합계출산율(TFR)은 한 여성이 가임연령(15~49세) 동안 평균적으로 출산하는 자녀 수로 소개된다. 그러나 실제로 한 여성이 미래에 몇 명의 자녀를 출산할지는 미래의 불확실성을 포함한다. 이런 경우에는 확률적인 계산으로 접근한다. 예를 들어, 2023년 TFR 0.721명은 2023년의 15세 여성 출산율, 16세 여성 출산율, $\cdots$, 49세 여성 출산율을 모두 더한 값이다. 이는 2023년 현재 각 연령의 여성이 어떤 출산 행태를 보이고 있는지를 나타낸다.

따라서 연령별 출산율은 일정 기간 동안 관찰된 출산 확률의 이산적 연령분포로 볼 수 있다. TFR은 한 여성이 평생 동안 2023년 현재의 각 연령

여성들과 동일한 출산 행태를 보인다고 가정했을 때의 평균 출생 건수로 해석된다. 그러므로 TFR은 단순히 1명, 2명이 아니라 평균적인 개념으로, 확률적 관점에서 계산된 결과로 0.721명이라는 수치가 나온다.

연령별 출산율을 계산할 때, 장래인구추계에서는 코호트출산율(Cohort Fertility Rate, CFR) 방법이 널리 사용된다. CFR 방법은 여성의 연령별 출산율을 코호트 단위로 관찰하여, 코호트 완결 출생아수와 연령별 출생률을 추계하는 방식이다. 코호트 방법은 〈표 3-2〉의 음영 부분을 의미한다. 코호트로 관찰하는 것으로 15세 t년, 16세 t+1년, 17세 t+2년을 살펴보는 방법이다. '↓'는 t년도 기간별 연령별 출산율(period ASFR)이다.

〈표 3-2〉 모의 연령별 출산율과 코호트 출산율 비교

가임연령	t년	t+1년	t+2년	t+3년	…	t+34년
15	↓ f(15)	f(15)	f(15)	f(15)	…	f(15)
16	↓ f(16)	f(16)	f(16)	f(16)	…	f(16)
17	↓ f(17)	f(17)	f(17)	f(17)	…	f(17)
18	↓ f(18)	f(18)	f(18)	f(18)	…	f(18)
⋮	↓ ⋮	⋮	⋮	⋮	…	⋮
49	↓ f(49)	f(49)	f(49)	f(49)	…	f(49)

주: '↓'는 기간연령별출산율, 음영부분은 코호트출산율을 뜻함
자료: 오진호(2018), pp. 113. 〈표 1〉.

본 연구는 기간 연령별 출산력을 예측하기 위한 다양한 방법을 제시한다. 첫째, 과거 연령별 출산율 자료의 추세가 지속된다는 가정을 바탕으로 통계적 모형을 활용해 미래를 예측하는 수리적 모형 방법이다. 이 방법은 모수적 통계모형(parametric statistical model) 접근 방식을 따른다. 둘째, 출산율 분포에 특정한 모수적 가정을 두지 않고, 자료 자체가 나타내는 분포를 함수적 자료분석(functional data analysis) 패러다임으로 분석하는 방법이다. 이 방법에서는 함수적 자료 모형(functional

data model, FDM)이 자주 활용되며, 비모수적(non-parametric) 모형으로 분류되기도 한다. 셋째, 선진국의 ASFR 분포나 목표치(target point), 또는 출산율 전이 단계를 참고하여 특정 시점의 ASFR을 설정한 후 선형적으로 연장하거나 베이지안 방법을 활용해 예측하는 방법이다. 넷째 ASFR과 사회경제적 외생 변수 간의 관계를 분석하여 다중 회귀분석(multiple regression model) 또는 시계열 모형(time series model)으로 미래 ASFR을 예측하는 방식도 있다. 이러한 방법들은 TFR를 먼저 예측한 뒤 이를 ASFR에 배분하는 방식과, ASFR을 먼저 예측한 뒤 TFR에 맞춰 조정하는 방식으로 구분된다.

ASFR 패턴을 설명하기 위해 다양한 모수적 모형(parametric model)이 연구되어 왔다. 주요 모형들은 〈표 3-3〉에 정리되어 있으며, 대부분 지수족(exponential family) 형태를 따른다. 이는 ASFR의 형태가 평균 출산 연령을 중심으로 대칭적인 패턴을 보이기 때문이다. 대표적인 사례로 Hoem 등(1981)이 제안한 베타 함수 모형이 있지만, 대다수의 모형은 지수족 분포를 채택한다. 모수적 모형은 추정하는 모수의 개수가 많을수록 설명력이 높아지지만, 모수가 적을수록 모형의 단순성과 모수 절약성 측면에서 유리하다. 따라서 설명력과 모수 절약성 간의 균형(trade-off)을 고려해 약 3~4개의 모수를 사용하는 것이 적절하다고 본다. 모형에 대한 세부적인 내용은 박유성 외(2013)와 오진호(2024)의 연구를 참고하면 된다. 〈표 3-3〉에 제시된 모수적 모형 중, 통계청은 ASFR 예측 모형으로 일반화 로그 감마(GLG) 모형(Generalized Log Gamma Model)을 채택하고 있다(Kaneko, 2003).

〈표 3-3〉 출산율 예측모형

모형	모형식	모수
감마(Gamma) 함수	$f(x)=R\frac{1}{\Gamma(b)}c^b(x-14)^{b-1}\exp(-\frac{x-14}{c}),\ x \geqq 15$	R,b,c
Hadwiger 함수 (Hadwiger, 1940)	$f(x)=\frac{ab}{c}(\frac{c}{x})^{3/2}\exp\left(-b^2(\frac{c}{x}+\frac{x}{c}-2)\right)$	a,b,c
베타(Beta) 함수 (Hoem 외, 1981)	$f(x)=R\frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)}(\beta-\alpha)^{(b+c-1)}(x-\alpha)^{b-1}(\beta-x)^{(c-1)}$	R,b,c
혼합 Hadwiger 함수 (Cahandola 외, 1999)	$f(x)=m(\frac{b_1}{c_1})(\frac{c_1}{x})^{3/2}\exp\left(-b_1^2(\frac{c_1}{x}+\frac{x}{c_1}-2)\right)$ $+(1-m)(\frac{b_2}{c_2})(\frac{c_2}{x})^{3/2}\exp\left(-b_2^2(\frac{c_2}{x}+\frac{x}{c_2}-2)\right)$	$m,b_1,b_2,$ c_1,c_2
일반화로그감마함수 (Kaneko, 2003)	$f(x)=\frac{C\lambda\|}{b\Gamma(\lambda^{-2})}(\lambda^{-2})^{\lambda^{-2}}\exp\left[\frac{1}{\lambda}\left(\frac{x-\mu}{b}\right)-\frac{1}{\lambda^2}\exp\left\{\lambda\left(\frac{x-\mu}{b}\right)\right\}\right]$	$C,\mu,$ b,λ
PK1함수 (Peristera and Kostaki, 2007)	$f(x)=c\exp\left(-(\frac{x-\mu}{\sigma(x)})^2\right),\ \begin{cases}\sigma(x)=\sigma_1, & \text{if } x \leq \mu \\ \sigma(x)=\sigma_2, & \text{if } x > \mu\end{cases},$ $c=\sum_{i=15}^{49}ASFR_i$	$c,\mu,$ σ_1,σ_2
PK2함수 (Peristera and Kostaki, 2007)	$f(x)=c_1\exp\left(-(\frac{x-\mu_1}{\sigma_1})^2\right)+c_2\exp\left(-(\frac{x-\mu_2}{\sigma_2})^2\right),$ $c_1=\sum_{i=15}^{\mu}ASFR_i, c_2=\sum_{\mu}^{49}ASFR_i$	$c_1,c_2,\mu_1,$ μ_2,σ_1,σ_2
혼합정규함수 (Mixture of normal functions, (박유성 외, 2013)	$f(x)=c_1\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2\right)I[x \leqq \mu_1]+$ $c_2\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2\right)I[x \geqq \mu_1]$	$c_1,c_2,\mu_1,$ μ_2,σ_1,σ_2

자료: 박유성 외 (2013); 오진호 (2024).

GLG 모형은 출산 순위별 출산율에 확률분포 개념을 적용하며, 확률분포 모형을 활용함으로써 모수 추정에 이론적 기반이 있는 통계적 방법론

을 사용할 수 있다는 장점이 있다. GLG 모형은 4개의 모수(C_i, μ, b, λ)와 오차항을 포함하며, 구체적인 형태는 식 (10)에 제시되어 있다. GLG 모형은 출산율 분포의 특성을 설명하는 데 적합하며, 통계적 안정성을 기반으로 신뢰할 수 있는 결과를 제공한다.

$$f_i(x) = \frac{C_i \mid \lambda \mid}{b\Gamma(1/\lambda^2)}\left(\frac{1}{\lambda^2}\right)^{\lambda^{-2}} \exp\left[\frac{1}{\lambda}\left(\frac{x-\mu}{b}\right) - \frac{1}{\lambda^2}\exp\lambda\left(\frac{x-\mu}{b}\right)\right] + e_i \tag{10}$$

$f(x)$: 연령별 출산율, x(연령)
C: 특정연도에 출생한 코호트가 i번째 자녀를 출산할 확률(i : 출산순위)
μ : 평균출산연령
b : 평균출산연령의 표준편차
γ : 감마함수
λ : 분포형태(고정상수), 유럽은 양의 값을 일본, 한국은 -0.3근방

식 (10)의 $f_i(x)$는 연령 x세의 출산율, C_i는 특정의 출생코호트가 가임연령 동안 출산순위 i번째 자녀의 출산을 경험할 확률이다. u와 b는 출산연령의 평균과 표준편차, λ는 분포형태를 나타내는 모수이다. e_i는 오차이다. 그리고 식 (10)은 식 (3)의 확률적 모형이다.

국내 연구는 주로 모수적 방법에 초점을 맞추고 있다. 대표적인 연구로 로그 감마 모형을 이용해 출산율 추이를 예측한 연구(전광희, 2006)와 확률적 출산율 모형과의 비교를 다룬 연구(박유성 등, 2013)가 있다. 박유성 등(2013)은 출산율 예측에서 결정적(deterministic) 방법과 확률적(probabilistic) 방법을 소개하며 각각의 장단점을 제시하였다.

모수적 접근에서는 〈표 3-3〉에 제시된 다양한 함수, 예를 들어 감마 함수(Hoem 등, 1981), Hadwiger 함수(Hadwiger, 1940), 베타 함수(Hoem 등, 1981), 혼합 Hadwiger 함수(Chandola 등, 1999), 그리고 PK 1, 2 함수(Peristera와 Kostaki, 2007) 등을 활용하여 모수를 추정

하였다. 추정된 모수는 시계열 모형에 대입되어 미래의 출산율을 예측하는 데 사용된다. 이러한 접근법은 과거 출산율 패턴이 미래에도 지속된다는 가정에 기반을 두고 있다.

그러나 우리나라처럼 최근 출산율 변동이 큰 경우, 이러한 가정은 적합하지 않을 수 있다. 이에 대해 박유성 등(2013)은 평균 출산 연령을 중심으로 서로 다른 두 개의 분포가 합쳐진 형태인 혼합 정규 함수(mixture of normal functions)를 제안하였다. 이 접근은 출산율 변화의 복잡성을 보다 효과적으로 설명할 수 있는 방법으로 평가된다.

또한, 출산에 영향을 미치는 혼인 관계를 반영한 연구도 진행되었다. 엄지민과 김기환(2013)은 혼인을 전제로 한 연령별 출산 예측 모형을 제안하였으며, 김기환과 전새봄(2015)은 출산율 예측 이전에 초혼, 재혼, 이혼 등의 혼인 구조를 고려한 혼인 모형을 제안하였다. 이들은 우리나라에서 출산이 혼인과 밀접하게 연관되어 있음을 강조하며, 혼인 변동이 출산율 예측에 미치는 영향을 분석하였다. 이와 같은 연구들은 출산율 변화의 복잡한 요인을 반영하기 위한 다양한 접근법을 제시하며, 특히 사회적·경제적 맥락을 반영한 모형 설계의 중요성을 강조하고 있다.

모수적 방법 외에도 Ramsay와 Silverman(2005), Hyndman과 Ullah(2007)은 함수적 자료분석 패러다임을 활용하여 출산율을 모델링하고 예측하기 위한 비모수적 방법인 함수적 데이터 모형(functional data model; FDM)을 제안하였다. 이들은 관측치에 내재된 측정오차(measurement error)와 전쟁, 질병 등으로 인해 인구 동태 자료에서 나타나는 불규칙적인 패턴을 교정하기 위해 함수적 자료 분석을 적용하였다. 특히, 측정오차와 불규칙 패턴을 보정하기 위해 비모수 평활 기법(non-parametric smoothing)을 사용하여 출산율 모형을 구축하였다(김순영과 오진호, 2017; 김순영 등, 2018). 이러한 FDM 모형은 출산

율 데이터의 변동성을 효과적으로 조정하고, 미래 출산율을 예측하는 데 적합하다. FDM 모형의 구체적인 수학적 표현은 식(11)과 (12)로 제시되며, 이를 통해 향후 출산율의 변화 추이를 예측할 수 있다.

$$f_t(x) = \begin{cases} \frac{1}{\lambda}(f^*{}_t(x)^\lambda - 1) \ , \ 0 < \lambda < 1 \\ \log_e(f^*{}_t(x)) \qquad , \qquad \lambda = 0 \end{cases} \tag{11}$$

식 (11)의 $f^*{}_t(x)$는 시간 t와 연령 x에서 출산율을 의미한다. $f^*{}_t(x)$의 Box-Cox(1964) 변형은 $f^*{}_t(x)$의 값에 따라 증가하는 변동을 줄여주거나 정규화과정, λ는 Box-Cox 변형에서 강도이다.

$$\begin{aligned} f_t(x) &= s_t(x) + \sigma_t(x)\epsilon_{t,x}, \\ s_t(x) &= \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x), \ \ j = 1, \cdots, J \end{aligned} \tag{12}$$

식 (12)의 $f_t(x)$는 시간 t의 연령 x에 대한 관찰된 로그출산율 $\ln f_t(x)$이고, $s_t(x)$는 평활함수(smooth function), $\epsilon_{t,x}$는 독립이고 동일 분포($i.i.d$)인 표준정규 확률변수이다. $\sigma_t(x)$는 시간 t의 연령 x에 따라 변하는 잡음 양이다. 따라서 $\sigma_t(x)\epsilon_{t,x}$는 관측된 로그출산율과 평활한 곡선의 차이인 관측오류이다. $\mu(x)$는 $\sum_{t=1}^{n} s_t(x)/n$에 의해 추정된 평균함수로 평활한 연령에 따른 로그출산율평균, $\beta_{t,j}$, $\phi_j(x)$는 함수적 주성분분석을 사용하여 추정하고, J는 사용된 주성분 수이다. $\Phi = \{\phi_1(x), ..., \phi_J(x)\}$는 J 개의 함수적 주성분의 집합으로 직교 기저함수(orthogonal basis function)이고 $B = \{\beta_1(x), ..., \beta_J(x)\}$는 비상관 주성분 점수(uncorrelated principal component scores)들의 집합으로 시계열 계수를 의미한다. 식 (12)의 두 번째 식은 시간에 따라 변화하는 $s_t(x)$의

변화를 설명하는 부분으로 하나 이상의 주성분을 사용하고 함수적 주성분분석(Functional Principal Components Analysis, FPCA)을 사용하여 평활한 곡선 $s_t(x)$을 직교함수 주성분과 비상관 주성분 점수로 분해한 것이다.

FDM은 첫 번째 주성분에 직교하는 고차원 주성분에 대해서는 다른 시계열 모형들의 주성분 점수가 도출된다. 모든 성분에 FDM 방법은 최적 시계열 모형을 AIC 등과 같은 모형 판별 기준에 의거하여 선택한다. 비모수모형에 대한 자세한 설명은 Hyndman 등(2007, 2013), 김순영 등(2017, 2018)을 참조하면 된다.

FDM 예측은 $Z=\{f_1(x),\ldots,f_n(x)\}$와 $\Phi=\{\phi_1(x),\ldots,\phi_J(x)\}$의 조건부로 $f_{n+h}(x)$의 h단계 예측치를 구할 경우 다음과 같은 식(13)에 의해서 도출 가능하다.

$$\hat{f}_{n+h|n}(x)=E[f_{n+h}(x)|Z,\Phi]=\hat{\mu}(x)+\sum_{j=1}^{J}\hat{\beta}_{n+h|n,j}\phi_j(x) \tag{13}$$

여기서 $\hat{\beta}_{n+h|n,j}$은 Hyndman and Booth(2008)에 의한 지수 평활 또는 ARIMA 모형과 같은 시계열모형을 활용하여 도출한 $\beta_{n+h|j}$의 h 단계 예측을 의미한다. 그리고 모든 주성분들은 직교성을 보이므로 예측분산을 쉽게 식 (14)와 같이 도출할 수 있다.

$$\begin{aligned}\hat{v}_{n+h\mid n}(x)&=Var[f_{n+h}(x)\mid Z,\Phi] \\ &=\sigma_\mu^2(x)+\sum_{j=1}^{J}u_{n+h\mid n,h}\phi_j^2(x)+v(x)+\sigma_t^2(x)\end{aligned} \tag{14}$$

식 (14)의 $\sigma_\mu^2(x)$는 $\hat{\mu}(x)$의 분산, $u_{n+h|n,h}$은 $\beta_{n+h,j}|\beta_{1,j},\ldots,\beta_{n,j}$의 분산이다. 그리고 $v(x),\sigma_t^2(x)$는 각각 $e_t(x)$와 $\sigma_t(x)$의 분산이다.

모수, 비모수 모형 외에[14] 출산율 자료 부족으로 완비성이 낮고 품질까지 좋지 못한 경우 선진국의 출산율 3단계 전이현상을 참고하여 이들의 사전분포를 활용하는 베이지안(Bayesian) 계층모형(Alkema 등, 2011; Raftery 등, 2012, 2014; Sevcikova 등, 2011)이 있다. UN DESA 산하의 유엔인구부(UN Population Division, UNPD)는 출산율 추정과 예측을 위해 출산율 3단계 전이단계[15] 중 2단계에 대해 5년 전후 출산율의 갭(gap, $f_t - f_{t+5}$)을 관찰하여 감소함수(Decline Function)로 정의하고, 식 (15)의 더블로지스틱(Double Logistic) 모형으로 적합한다(Alkema 등, 2011; Raftery 등, 2012, 2014; Sevcikova 등, 2011).

$$
\begin{aligned}
& f_{c,t+5} = f_{c,t} - r(f_{c,t}|\delta^c) + a_{c,t},\ \ \tau_c \le t < \lambda_c \\
& \tau_c = \begin{cases} \max\{t : (M_c - L_{c,t}) < 0.5\}, \text{if}\ \ L_{c,t} > 5.5 \\ < 1950 \sim 1955 \qquad\qquad ,\text{그외} \end{cases} \\
& \lambda_c = \min\{t : f_{c,t} > f_{c,t-1}, f_{c,t+1} > f_{c,t}\ ;\ f_{c,p} < 2, p = t-1, t, t+1\}
\end{aligned}
\tag{15}
$$

$$
r(f_{c,t}|\delta^c) = \frac{-d^c}{1 + \exp[-2\ln(9)(f_{c,t} - \sum_{i=2}^{4} \nabla_i^c + 0.5\nabla_1^c)/\nabla_1^c]} + \frac{d^c}{1 + \exp[-2\ln(9)(f_{c,t} - \nabla_4^c - 0.5\nabla_3^c)/\nabla_3^c]}
$$

식 (15)의 M_c는 국가별 TFR 최댓값, $L_{c,t}$는 부분 최대치(local maxima)이다. 그리고 국가별 전이 3단계의 시작점을 λ_c라고 했을 때, 이는 TFR 2명 이하 부분에서 인접 두 시점에서 증가로 판정된다면 이는 전이 3단계로 접어들었다고 판정한다. 대체적으로 λ_c는 2005~2010년 이후

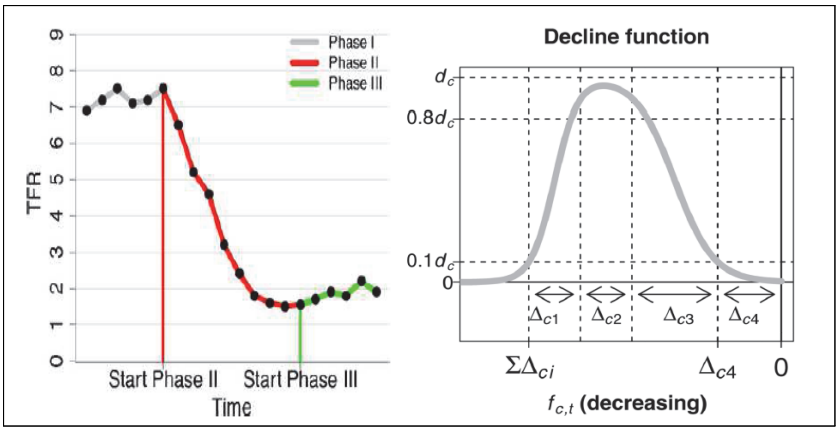
14) 오진호(2018) 논문을 일부 발췌하여 정리한 것임.

15) TFR 진화 추이를 3단계로 나누어서 정의하는 것으로 1단계는 출산율 전이 이전의 높은 출산율 수준, 2단계는 4~5명 이상의 높은 출산율 수준에서 인구대체선 수준이나 미만으로 감소하는 전환단계, 3단계는 낮은 출산율에서 서서히 증가해 인구대체선 2.1명까지 도달하는 단계를 의미.

에 발생한다. 그리고 $\delta^c = (\nabla_1^c, \nabla_2^c, \nabla_3^c, \nabla_4^c, d^c)$는 국가별 모수 벡터, $a_{c,t} \overset{ind}{\sim} N(0, \sigma(t, f_{c,t})^2)$는 출산수준과 시간변동에 따른 표준편차를 의미한다. 일반적으로 TFR이 감소하다가 최소점을 지나 점진적인 증가추세를 보이므로 5년차의 gap을 그려보면 음의 값으로 증가하다가 최고점을 찍고 다시 하락하는 추이를 도출할 수 있다. 이들 값들을 0을 중심으로 부호(sign)를 반대로 부여해 그리면 [그림 3-4]의 우측 그림이 된다. 이때 $\nabla_1^c, \nabla_2^c, \nabla_3^c, \nabla_4^c, d^c$은 [그림 3-4]의 우측 곡선 궤적 단계를 표현하는 값으로 이해하면 된다. 출산율이 점진적으로 감소하는 구간은 ∇_1^c, ∇_2^c에 해당하고 최고점은 d^c, 그리고 최저점에서 점진적으로 증가하는 구간은 ∇_3^c, ∇_4^c에 속한다. 특히 ∇_4^c는 예측구간에 해당되는 경우가 많은데 UN(2017)에서는 상한을 인구대체율인 2.1명으로 하한은 안전선인 1.5명으로 정하고 있다.

그리고 국가별 모수 δ^c는 널리 퍼진 사전분포(diffuse prior distribution)를 가지는 세계 출산 분포로부터 도출되는 것을 가정한다. 이는 마코프 체인 몬테카를로(Markov Chain Monte Carlo, MCMC) 방법을 사용하여 출산 전이 모형에 대한 각 모수들의 사후분포 표본(posterior distribution sample)을 도출한다. 이들이 제시한 MCMC 알고리즘은 깁스샘플링(Gibbs sampling), 메트로폴리스-해스팅(Metropolis-Hastings) 그리고 슬라이스 표본 스텝(Slice Sampling Steps)의 조합이다.

[그림 3-4] TFR 전이 3단계와 감소함수 궤적



자료: Alkema 등 (2011); Raftery 등 (2014).

다음으로 3단계는 국가별(country-specific) 식 (16)과 같은 자기상관회귀(국가별 출산율 평균(μ_c)에 근사하는 1차 자기회귀 시계열 모형 AR(1)) 시계열 모형으로 적합하여 출산율을 추정하고 예측한다.

Alkema 등(2011)이 출산율 전이 3단계를 도출할 때, 21개국 선진국(19개 유럽 국가, 미국, 싱가포르) 출산율을 참고하여 전이 3단계의 출산율이 2.1명(μ)으로 회귀한다는 강한(strong) 가정하에 AR(1)모형으로 예측하였다. 자세한 모형은 식 (16)과 같다.

$$f_{c,t+1} - \mu = \rho(f_{c,t} - \mu) + b_{c,t}, \quad b_{c,t} \sim N(0, \sigma_b^2) \tag{16}$$

그리고 모수 ρ와 σ_b는 출산율 전이 3단계에 접어든 21개국의 1950~2014년 출산율로부터 최대우도추정량으로 추정($\hat{\rho}=0.89$, $\hat{\sigma_b}$=0.10)하였다.

그러나 이 결과와 가정에 대해서 Basten 등(2012)이 동아시아(한국, 일본, 홍콩, 싱가포르, 대만) 출산율 환경과는 괴리감이 있다고 지적하였다. 즉, 이 방법으로 도출되는 TFR은 너무 높기 때문에 Raftery 등(2014)은

국가별(μ_c) 로 AR(1)모형을 적용하는 것으로 수정을 하게 되었다. 또한 Raftery 등(2014)은 21개국 사전분포로 사후분포를 활용하는 방식으로 μ와 ρ를 국가별로 다음 식 (17)과 같이 변경해서 추정하였다.

$$f_{c,t+5}(x) - \mu_c = \rho_c(f_{c,t} - \mu_c) + \epsilon_{c,t}, \quad \epsilon_{c,t} \overset{i.i.d}{\sim} N(0, \sigma_\epsilon^2) \tag{17}$$

$$\mu_c \sim TN_{[0,\infty)}(\overline{\mu}, \sigma_\mu^2), \ \rho_c \sim TN_{[0,1]}(\overline{\rho}, \sigma_\rho^2)$$

여기서, $TN_{[a,b]}(\mu, \sigma^2)$은 a와 b사이로 절단(truncated)되어진 평균이 μ이고 표준편차가 σ인 절단 정규분포(truncated normal distribution)를 의미한다. 그리고 Raftery 등(2014)은 이들 모수에 대한 사전분포를 식 (18)과 같이 제시했다.

$$\overline{\mu} \sim U[0,2.1]; \sigma_\mu \sim U[0,0.318]; \rho_c \sim U[0,1]; \sigma_\rho \sim U[0,0.289]; \sigma_\epsilon \sim U[0,0.5] \tag{18}$$

그리고 Sevcikova 등(2011)은 지금까지 설명한 베이지안 출산율 모형을 R 프로그램 오픈 자료인 bayesTFR 패키지로 구현해 세계 230개국의 출산율을 예측할 수 있게 하였다. Sevcikova 등(2011)은 TFR 추계를 다음 3단계로 제안한다. 첫째 TFR 예측모형 적합, 둘째 미래 TFR 궤적 생성, 셋째 진단 등으로 결과를 분석 보정한다.

정리하면, 출산율 전이 2단계의 시작 시기와 국가별 3단계의 시점을 계산한다. 다음으로 최대우도추정량을 사용하여 전이 3단계에 접어든 국가별 AR(1)모형의 모수를 추정한다. 하지만 이전에도 언급했듯이 Raftery 등(2014)의 방식으로 개선하기 위해 베이지안 방법(Sevcikova 등, 2018)으로 MCMC 알고리즘을 사용하여 전이 2단계 모수의 사후분

포 표본을 도출한다. 그리고 출산율 예측방식은 크게 두 부분으로 구성되는데 현재 출산율 전이 2단계, 3단계에 해당되는 국가에 각각 다르게 접근한다.

먼저 출산율 2단계는 식 (15)의 감소함수와 사후분포 표본 도출 로직에 의거해서 TFR을 예측한다. 그리고 전이 3단계의 예측 시작점은 두 조건 ($\min_t \{f_{c,t}^{(i)}\} \leq \triangle_{c_4}^{(i)}$과 $f_{c,t}^{(i)} > f_{c,t-1}^{(i)}$)을 만족하는 가장 최근의 시점 t로 정한다. 그리고 현재 출산율 3단계에 있는 국가는 식 (16)에 의해 AR(1) 모형을 사용하여 예측한다. 3단계에 접어든 선진국이나 개발도상국 일부 국가들의 TFR의 패턴은 약간의 상승이나 일정한 수준을 유지하는 상수적인 패턴을 보인다. 따라서 AR(1)모형을 적용해도 무리가 없을 것으로 판단된다.

최근 UN에서는 이들이 제안한 모형으로 230여 개국의 출산율, 사망률을 산출한다UN, 2017). 베이지안 방법의 가장 큰 장점은 출산율 자료가 충분치 않은 경우 선진국 출산율의 사전분포를 통해 MCMC 방법으로 관심대상 국가의 사후분포를 생성할 수 있다는 것이다. 따라서 자료 품질이 낮거나 완비성이 떨어지는 개발도상국이나 후진국인 경우 이 방법을 적용하면 출산율 예측이 가능하다. 보다 자세한 설명은 Alkema 등(2011)과 Raftery 등(2012, 2014), Sevcikova 등(2011, 2018), 오진호(2018)의 논문을 참고하면 된다.

나. 출산진도비 접근

동일한 합계출산율이라도 출산이 일부 집단에만 집중된다면 다인가구는 소수화되고 1인가구는 일반화될 수 있다. 이처럼 출산진도비(PPR) 기반 분석이 필요한 이유가 여기에 있다. 단지 평균 자녀수가 아니라, 누가

몇 명을 낳는가의 분포가 가구원수 예측을 좌우한다.

기존의 연령별 출산율 중심의 접근은 출산 시기의 변동을 포착하는 데는 유효하나, 출산의 양(quantum) 또는 출산 순위별 진행경로와 같은 구조적 특성을 반영하는 데는 한계가 있다. 반면, 생존모형(discrete-time survival model) 기반의 PPR 접근(우해봉, 장인수, 2018)은 출산의 누적성과 반복성을 구조화함으로써 가구 내 자녀 수 변화, 결과적으로는 전체 가구원수 변화와의 연계를 보다 정밀하게 분석할 수 있는 장점을 갖는다.

PPR 접근은 출산 간격(tempo)이 아닌 출산의 양(quantum)에 주목하며, 생애 전체에 걸친 출산 행동의 누적 경로를 포착할 수 있다는 점에서 최근 국내·외에서 각광받고 있다(이홍탁, 1994; 우해봉, 장인수, 2018, Ryder, 1986; Feeney와 Yu, 1987; Retherfold 외 2010, Schoumaker, 2013; Van Hook과 Altman, 2013).

이들 연구를 통해 본 절에서는 생존모형 기반의 출산력 분석 결과를 가구추계 변수로 통합할 수 있는 방법론적 가능성을 제시해 본다. 이는 출산 순위별 이행률이 가구원수 분포에 미치는 구조적 영향을 분석하고 나아가 출산력. 혼인력, 가구유형 간 인과적 연결성을 반영한 통합 추계모형의 개념적 틀을 제안하는 것이다.

분석모형으로 complementary log-log link를 가정한 이산형 생존모형을 기초로 출산진도비를 '순위별 PPR'과 'PPR 누적'의 곱으로 TFR을 도출한다. 이때 가구 유형분류는 무자녀+미혼은 1인가구, 유배우+무자녀는 부부가구, PPR 기반 다자녀는 부부+자녀가구, 이혼(사별)+자녀는 한부모가구로 가정한다.

또한 인구-가구 연계 모형은 다음의 단계를 따른다.

1단계는 연령별, 성별, 혼인상태별 순위별 인구를 분류한다. 2단계는 출산진도비 기반 가구 내 자녀수를 결정한다. 3단계는 자녀 수와 혼인상

태에 따라 가구유형을 결정한다. 마지막 4단계는 연령별 사망률을 반영하여 가구 해체나 1인가구 전이를 추정한다.

이런 접근은 기존의 출산정책이 단순 출산 장려에 그쳤으나, 향후에는 가구구조 변화의 흐름을 반영한 정책 설계가 병행되어야 함을 보여주고, 출산력 예측이 곧 가구정책의 핵심 정보 자원으로 작용해야 하며, 특히 1인가구·무자녀 가구 정책에서 중요한 기반이 됨을 알 수 있다. 더불어 통계청 등 인구 관련 연구기관은 인구추계와 가구추계를 연계한 통합 추계 시스템 개발에 착수할 필요가 있음을 시사한다.

정리하면, 생존모형을 활용한 출산력 분석이 단지 인구학적 지표의 정밀화를 넘어서, 가구구조 예측에 필수적인 정보 체계로 활용될 수 있음을 실증적으로 보여준다. 특히 출산진도비는 단순한 평균 출산자녀 수를 넘어, 가구 내 자녀 수 구성의 분포, 더 나아가 가구유형의 다양화를 구조적으로 예측하는 데 주요한 역할을 한다. 향후 연구에서는 출산력뿐 아니라 혼인·이혼·사망·자녀의 분가율 등을 포함한 종합적 생애과정 기반 인구-가구 연계 시뮬레이션 모형 구축이 요구된다.

제2절 출산력 추계 방법론 개선 방안

1. 기존 출산력 추계 방법론의 주요 쟁점

가. 통계청 방법론의 적용상 쟁점

통계청이 채택하고 있는 일반화 로그감마(GLG) 모형은 4개의 모수를 통해 출산율 분포의 특성을 설명하는 통계적으로 안정된 방법론이다. 그

러나 우리나라와 같은 초저출산 상황에서 몇 가지 적용상 쟁점이 제기되고 있다.

첫째, 모수 추정의 안정성 쟁점이다. 합계출산율이 0.72명(2023년)까지 하락한 상황에서 GLG 모형의 모수들이 통계적으로 유의미한 수준에서 안정적으로 추정되는가의 문제이다. 출산율이 극도로 낮아지면서 각 연령별 출산율의 변동성이 커지고, 이는 모형의 적합도와 예측력에 영향을 미칠 수 있다.

둘째, 출산 패턴의 급격한 변화 반영 쟁점이다. GLG 모형은 과거의 안정적인 출산 패턴을 전제로 하는데, 최근 출산 연령의 급격한 상승, 출산 집중도의 변화, 출산 순위별 패턴의 분화 등을 충분히 반영하기 어려울 수 있다. 특히 코로나19 이후 나타난 출산 행태의 구조적 변화는 기존 모형의 적용 범위를 벗어날 가능성이 있다.

셋째, 정책 변수 반영의 제약 쟁점이다. GLG 모형은 본질적으로 인구학적 변수에 기반한 모형으로, 출산지원정책이나 사회경제적 여건 변화가 출산율에 미치는 영향을 직접적으로 모형에 반영하기 어렵다. 이는 정책 효과 분석이나 시나리오별 전망에서 제약으로 작용한다.

나. 연령별 출산율 접근의 구조적 이슈

기존의 연령별 출산율(ASFR) 중심 접근법은 출산력 분석의 표준적 방법이지만, 현재의 복잡한 출산 환경에서는 몇 가지 구조적 쟁점을 안고 있다.

가장 중요한 쟁점은 출산 시기(tempo)와 출산량(quantum)의 분리 문제이다. 연령별 출산율은 특정 연령에서의 출산 집중도를 보여주지만, 이것이 출산 지연인지 출산 포기인지를 명확히 구분하기 어렵다. 특히 우리나라처럼 출산 연령이 급격히 상승하는 상황에서는 단순한 ASFR 추이만으로는 미래 출산력을 정확히 전망하기 어렵다.

또한 출산의 이질성 포착의 한계도 중요한 쟁점이다. 동일한 합계출산율이라도 출산이 소수 집단에 집중되는지, 전 계층에 고르게 분포하는지에 따라 사회적 의미가 전혀 다르다. 연령별 출산율은 평균적 패턴은 보여주지만, 출산 행동의 분화와 양극화 현상을 충분히 반영하지 못한다.

다. 정책 민감도 분석에서의 한계

현행 출산력 추계 방법론은 정책 효과를 분석하고 정책 시나리오를 구성하는 데 있어 몇 가지 중요한 이슈가 있다.

특히 정책 효과의 차별적 반영 어려움이 주요 쟁점이다. 첫째 아이 출산 지원 정책과 둘째 아이 장려 정책은 서로 다른 집단에게 다른 방식으로 영향을 미치는데, 전체적인 합계출산율 변화로는 이러한 차별적 효과를 구분하기 어렵다. 이는 정책의 효과성 평가와 정책 조합의 최적화에 제약을 가한다.

또한 정책 대상 집단의 정확한 식별이 어렵다는 쟁점도 있다. 연령별 출산율 기반으로는 정책이 실제로 영향을 미치는 집단(출산 순위별, 혼인 상태별, 가구 특성별)을 명확히 구분하기 어려워, 정책 설계 시 타겟팅의 정밀도가 떨어질 수 있다.

2. 출산진도비 접근의 방법론적 강점

가. 생존모형 기반 접근의 분석적 강점

출산진도비 접근법이 생존모형을 기반으로 하는 것은 출산 행동의 본질적 특성을 반영한 것으로, 기존 방법론 대비 여러 분석적 강점을 제공한다.

첫째, 출산의 누적적 특성을 자연스럽게 반영할 수 있다. 출산은 본질적으로 순차적이고 누적적인 과정인데, 생존모형은 이러한 시간적 순서와 누적 확률을 체계적으로 다룰 수 있다. 이는 개별 여성의 생애 출산 경로를 보다 현실적으로 모델링할 수 있게 해준다.

둘째, 관측되지 않은 이질성(unobserved heterogeneity)을 효과적으로 처리할 수 있다. 같은 연령, 같은 사회경제적 조건에서도 개인의 출산 성향은 다를 수 있는데, 생존모형은 이러한 개인차를 모형 내에서 체계적으로 고려할 수 있다.

셋째, 시간 변화하는 변수들의 영향을 동적으로 분석할 수 있다. 소득 변화, 고용 상태 변화, 주거 여건 변화 등이 각 출산 단계에서 어떤 영향을 미치는지를 시점별로 추적할 수 있어, 보다 정교한 인과관계 분석이 가능하다.

나. 연령별 출산율 접근의 구조적 쟁점

출산진도비 접근법의 가장 큰 강점 중 하나는 출산 순위별로 차별화된 분석이 가능하다는 점이다.

각 출산 단계의 고유한 결정 요인을 분석할 수 있다. 첫째 아이 출산 결정은 주로 경제적 안정성과 관련이 높은 반면, 둘째 아이 출산 결정은 육아 부담과 일-가정 양립 가능성이 더 중요한 요인이 될 수 있다. 출산진도비 접근은 이러한 단계별 차이를 명시적으로 모델링할 수 있다.

또한 정책 효과의 순위별 차별성을 정확히 측정할 수 있다. 보육료 지원 정책이 첫째 아이 출산에는 큰 영향을 미치지 않지만 둘째 아이 출산에는 상당한 효과를 보인다면, 이러한 차별적 효과를 출산진도비 분석을 통해 정량적으로 파악할 수 있다.

다. 혼인-출산 연계 모델링의 강점

우리나라의 출산은 혼인과 밀접하게 연계되어 있어, 혼인 패턴의 변화가 출산력에 미치는 영향을 정확히 파악하는 것이 중요하다. 출산진도비 접근법은 이러한 연계를 자연스럽게 모델링할 수 있는 강점을 갖는다.

혼인상태별 출산진도비의 차별적 적용을 통해 만혼이 출산력에 미치는 복합적 영향을 분석할 수 있다. 단순히 출산 연령이 늦어지는 효과뿐만 아니라, 늦은 결혼이 둘째, 셋째 아이 출산 확률에 미치는 영향까지 체계적으로 파악할 수 있다.

또한 혼인 지속기간과 출산 확률의 관계를 동적으로 모델링할 수 있다. 신혼 초기, 결혼 3-5년차, 그 이후 등 혼인 지속기간별로 각 출산 순위의 확률이 어떻게 변화하는지를 추적하여, 보다 현실적인 출산력 전망을 제공할 수 있다.

3. 출산력 기반 정책 시뮬레이션 체계에서의 강점

가. 출산 순위별 정책 효과 차별화의 강점

출산진도비 기반 접근법은 정책 효과를 출산 순위별로 세분화하여 분석할 수 있어, 보다 정밀한 정책 설계가 가능하다.

첫째 아이 출산 지원 정책(신혼부부 주택 지원, 난임 치료비 지원 등)의 효과는 주로 PPR_0(무자녀→첫째)에 반영되고, 둘째 아이 장려 정책(육아휴직 확대, 보육료 지원 등)의 효과는 PPR_1 (첫째→둘째)에 반영되는 방식으로 정책별 타겟 효과를 명확히 구분할 수 있다.

이는 정책 조합의 최적화에도 기여한다. 예를 들어, 전체 출산율 제고

라는 목표하에서 첫째 아이 출산 진입에 집중할지, 둘째 아이 출산 확대에 집중할지를 출산진도비별 현황과 정책 효과를 바탕으로 과학적으로 결정할 수 있다.

나. 가구 특성별 정책 설정에서의 강점

출산진도비 접근법은 가구 특성과 출산 결정을 직접 연결하여, 보다 정확한 정책 타겟팅을 가능하게 한다.

가구주 연령, 혼인상태, 기존 자녀 수, 소득 수준 등 가구의 다양한 특성이 각 출산 단계에서 어떤 영향을 미치는지를 체계적으로 분석할 수 있다. 이를 통해 동일한 정책이라도 어떤 가구 유형에게 더 효과적인지를 사전에 예측할 수 있다.

또한 정책 수혜 집단의 명확한 식별이 가능하다. 예를 들어, 다자녀 가구 지원 정책의 실제 수혜자가 누구인지(2자녀→3자녀 가구인지, 3자녀→4자녀 가구인지)를 정확히 파악하여, 정책 설계의 정밀도를 높일 수 있다.

다. 시나리오 기반 정책 효과 분석의 강점

출산진도비 접근법은 다양한 정책 시나리오에 대한 정량적 분석을 가능하게 하여, 정책 결정 과정에서 중요한 정보를 제공한다.

정책 강도별 효과 분석이 가능하다. 예를 들어, 육아휴직급여를 현재 수준에서 10%, 20%, 30% 인상했을 때 각각 둘째 아이 출산확률(PPR_1)이 어떻게 변화하는지를 정량적으로 분석하여, 정책의 한계효과와 최적 수준을 도출할 수 있다.

또한 정책 조합의 시너지 효과도 분석할 수 있다. 주택 지원과 보육료

지원을 동시에 시행했을 때의 효과가 각각을 별도로 시행했을 때의 효과의 단순합보다 큰지를 검증하여, 정책 패키지의 효율성을 평가할 수 있다.

이러한 강점들을 통해 출산진도비 기반 접근법은 기존의 연령별 출산율 중심 방법론을 보완하여, 보다 정밀하고 정책 친화적인 출산력 추계 체계를 구축할 수 있는 기반을 제공한다. 특히 현재와 같은 초저출산 상황에서는 평균적 지표보다는 구조적 분석이 더욱 중요하므로, 출산진도비 접근법의 도입이 시급한 과제라 할 수 있다.

제3절 출산력 추계모형 구축 및 결과 검토

1. 출산력 추계모형 설계 및 구축

가. 모형의 기본 구조

출산진도비 기반 출산력 추계는 혼인상태별 인구 구성과 PPR을 결합하여 출생아수를 산출하는 방식으로 설계된다. 전체 추계 과정은 〈표 3-4〉와 같이 크게 세 개의 모듈이 유기적으로 연계되어 작동한다.

〈표 3-4〉 출산력 추계 방법론 구조

출산력 추계 방법론 구조
연도 t 시작
⇓
1. 혼인상태별 인구(MaritalTransitionCalculator) - t년도 혼인상태별 인구 (미혼/유배우/사별/이혼) - 성별, 연령별 (15~100세)
⇓
2. PPR 적용 (FertilityCalculator → PopulationCalculator) - 유배우 여성 × PPR → 출생아수 산출 - 출생아 / 여성인구 → ASFR, TFR 역산
⇓
3. 코호트요인법 (PopulationCalculator) - 출생, 사망, 이동 반영 - t+1년도 총인구 (성별 × 연령별) 산출
⇓
4. 혼인상태 조정 (MaritalTransitionCalculator) - t+1년도 총인구와 혼인상태별 인구 합계 매칭 - 전이율 역산 및 평활화
⇓
연도 t+1로 이동

자료: 저자 작성.

첫 번째는 혼인상태별 전이율 추정 모듈(MaritalTransitionCalculator)이다. 이 모듈은 매년 혼인상태별 인구(미혼, 유배우, 사별, 이혼)의 전이를 추정하고, 인구추계 결과와의 정합성을 확보하기 위해 전이율을 동적으로 조정한다. 2020년 인구총조사를 기준으로 초기화되며, 혼인·사망·이혼 실적 데이터를 바탕으로 전이확률을 추정한다.

두 번째는 출산율 추정 모듈(FertilityCalculator)이다. 이 모듈은 연령별 출산진도비(PPR)를 추정하고, 이를 유배우 여성 인구에 적용하여 출생아수를 산출하는 데 필요한 계수를 제공한다. PPR 추정 방식은 분리형

(무자녀→첫째, 유자녀→추가 출산)과 통합형(전체 유배우 여성 대상)으로 구분되며, 본 연구에서는 정합성 확보를 위해 통합형 방식을 채택하였다.

세 번째는 인구추계 모듈(PopulationCalculator)이다. 이 모듈은 혼인상태별 유배우 여성 인구에 PPR을 적용하여 출생아수를 집계하고, 이를 바탕으로 연령별 출산율(ASFR)과 합계출산율(TFR)을 역산한다. 산출된 출생아는 코호트요인법의 0세 인구로 투입되며, 사망·이동 요인과 함께 다음 연도 인구를 추계한다.

이러한 세 모듈은 매년 순환적으로 작동한다. t년도의 혼인상태별 인구 → PPR 적용 → 출생아 산출 → 코호트요인법 → t+1년도 총인구 → 혼인상태 조정 → t+1년도 혼인상태별 인구의 과정이 반복되며, 2023년부터 2055년까지는 PPR 기반 추계, 2056년 이후는 2055년 ASFR을 고정하여 추계한다.

나. 혼인상태별 전이율 추정 모듈

(1) 전이 구조 및 전이 행렬 추정

인구 · 가구 통합 추계를 위해 혼인상태별 인구의 전이를 모델링하는 혼인상태별 전이율 계산(MaritalTransitionCalculator) 클래스를 개발하였다. 이 클래스에서는 4가지 혼인상태(미혼, 유배우, 사별, 이혼) 간의 전이확률을 추정하고, 각 상태에서의 사망확률을 반영하여 매년 혼인상태별 인구를 업데이트한다.

혼인상태별 전이율 계산에서는 5차원 전이확률을 행렬($T_{t,g,a,m,m'}$) 구조로 상정한다. 초기 전이율의 추정은 2020년 인구동향 실적 데이터를 기반으로 수행되며, 장래 가구추계 모형구축(송창길 외, 2024)의 방법을 준용한다.

전이율의 전망에서는 기존 장래가구추계의 혼인상태별 추계 방법론과는 달리 인구추계와의 정합성을 고려해야한다. 이를 위해, 혼인상태별 인구의 합계가 인구추계 결과와 일치하도록 전이율을 역산하여 조정하는 방식을 크게 7단계로 구현하였다.

첫 번째 단계로, 초기 전이율을 활용하여 다음 연도 혼인상태별 인구를 계산한다. 시점 t의 혼인상태별 인구와 전이율을 사용하여 시점 t+1의 혼인상태별 인구를 계산한다.

$$P_{t+1,g,a,m} = \Sigma_{a'=0}^{100} \Sigma_{m'=0}^{3} P_{t,g,a'-1,m'} \times T_{t,g,a'-1,m',m} \tag{19}$$

여기서 사망($m' = 4$)으로 전이한 인구는 다음 연도 인구에 포함되지 않는다.

두 번째 단계에서는 인구추계로부터 얻은 시점 t+1의 총인구($P_{t+1,g,a}^{Real}$)와 첫 번째 단계에서 계산한 혼인상태별 인구의 합계($\Sigma_m P_{t+1,g,a,m}$) 간의 차이를 계산한다.

$$\Delta_{t+1,g,a} = P_{t+1,g,a}^{Real} - \Sigma_{m=0}^{3} P_{t+1,g,a,m} \tag{20}$$

세 번째 단계에서는 $\Delta_{t+1,g,a}$를 각 혼인상태에 비례하여 배분한다.

$$P_{t+1,g,a,m} += \Delta_{t+1,g,a} \times (P_{t+1,g,a,m} / \Sigma_{m=0}^{3} P_{t+1,g,a,m'}) \tag{21}$$

네 번째 단계는 전이율을 역산하는 작업이다. 조정된 혼인상태별 인구 P_{t+1}와 이전 연도 인구 P_t로부터 전이율 T_{t+1}을 역산한다. 각 출발 상태 m에서 도착 상태 m'로의 전이 인구는 다음 식과 같다.

$$N_{t,g,a,m,m'} = P_{t,g,a,m} \times T_{t,g,a,m,m'} \tag{22}$$

이렇게 계산된 전이 인구는 다음 연령으로 이동하므로, 연령 $a+1$에서의 혼인상태 m' 인구에 기여한다. 역산된 전이율은 다음과 같다.

$$T^{adjust}_{t,g,a,m,m'} = N_{t,g,a,m,m'} / P_{t,g,a,m} \tag{23}$$

다섯 번째로, 급격한 전이율 변화를 방지하기 위해 이전 전이율과 역산 전이율을 가중평균하기 위한 평활화 파라미터(damping factor) λ를 적용한다.

$$T_{t+1,g,a,m,m'} = \lambda \times T_{t,g,a,m,m'} + (1-\lambda) \times T^{adjust}_{t,g,a,m,m'} \tag{24}$$

여기서 $\lambda = 1$이면, 기존 전이율이 유지되는 것이고, $\lambda = 0$이면 역산된 전이율을 완전히 반영하는 의미가 된다. 본 연구에서는 전이율 추계의 안전성을 담보하기 위해 평활화 파라미터를 적용하여 검토하였으며, 연구 결과 산출 시에는 $\lambda = 0$으로 설정하여 추계 인구 데이터를 최대한 반영하였다.

여섯 번째 단계에서는 각 출발 상태에서 모든 도착 상태(생존 상태만)로의 전이확률 합이 1이 되도록 정규화한다.

$$T_{t+1,g,a,m,m'} = T_{t+1,g,a,m,m'} / \Sigma^{3}_{m''=0} T_{t+1,g,a,m,m''} \tag{25}$$

여기서 사망 전이($m' = 4$)는 정규화에서 제외한다.

마지막 단계에서는 시나리오 적용을 다룬다. 본 연구에서는 전이율 조정에 따라 출산력 시나리오가 달라지도록 조정하였다. 초혼율의 변화 시나리오에 따라 출산율이 변화하는 시나리오이므로 출산 가능 연령(15~49세)에만 적용하며, 다음과 같이 계산한다.

$$T_{t,g,a,0,1} = T_{t,g,a,0,1} \times (1+\tau_{growth}) \tag{26}$$

여기서 τ_{growth}는 시나리오별 초혼율 증가율이며, 조정 후에는 반드시 정규화를 수행하여 확률 조건을 유지하도록 설정하였다.

(2) 구현 구조

혼인상태별 전이율 추정 모듈(MaritalTransitionCalculator)은 다음과 같은 python 구조로 설계되었다(〈표 3-5〉 참조).

〈표 3-5〉 MaritalTransitionCalculator pseudo 코드

혼인상태별 전이율 추계 MaritalTransitionCalculator pseudo 코드

```
# 혼인상태별 전이율 추계
class MaritalTransitionCalculator:
    # 초기설정
    def __init__(self, base_year, end_year, reestimate, scenario):
        """
        Args:
            base_year: 기준연도 (2020 고정)
            end_year: 전망 종료 연도
            reestimate: 재추정 여부 (False면 기존 파일 로드)
            scenario: 'low', 'mid', 'high'
        """
        # 시나리오별 초혼율 증가율 설정
        self.growth_rate = {'low': 0.000, 'mid': 0.015, 'high': 0.030}[scenario]

        if not reestimate and all_files_exist():
            # 기존 데이터 로드
            self.pop2020_marital = np.load(pop2020_file)
            self.initial_tp = np.load(initial_tp_file)
        else:
            # 새로 추정
            self.load_actual_data()  # 혼인·사망·이혼 실적 로드
            self.calculate_spouse_distribution()  # 배우자 연령 분포 계산
            self.initial_tp = self.estimate_initial_transitions()  # 2020년 전이율 추정

    # 2020~2030년 확장
    def _extend_marital_transitions_to_2023(self):
        # DB에서 2020-2022년 실제 인구 로드
        pop2020_total, pop2021_total, pop2022_total = db.load([2020, 2021, 2022])

        # 2020→2021 전이
        pop2021_marital, tp_2021 = self.adjust_marital_population(pop2020_marital,
        tp_2020, pop2021_total, damp=0.0)

        # 2021→2022 전이
        pop2022_marital, tp_2022 = self.adjust_marital_population(pop2021_marital,
        tp_2021, pop2022_total, damp=0.0)
```

혼인상태별 전이율 추계 MaritalTransitionCalculator pseudo 코드
2022→2023 전이 (2022년 인구로 한 번 더 추정) _, tp_2023 = self.adjust_marital_population(pop2022_marital, tp_2022, pop2022_total, damp=0.0) # 2023년 전이율 저장 self.initial_tp = tp_2023 self.pop2022_marital = pop2022_marital

자료: 저자 작성.

다. 출산율 모듈

본 연구에서 구현한 출산율 모듈(FertilityCalculator)은 연령별 출산율(ASFR)이 가정값으로 주어졌을 때, 이를 인구추계 모듈에 전달해주는 부분과 PPR 방식을 선택했을 때, 출산진도비가 인구추계에 전달되도록 계산하는 부분으로 나뉜다.

PPR 방식의 적용에 있어서는 분리형 방식과 통합형 방식으로 나누어 구현된다. 분리형 방식은 자녀가 없는 여성이 출산하는 비율과 자녀가 있는 여성이 추가 출산을 하는 비율이 적용되도록 출산진도비가 산출되며, 통합형 방식에서는 유배우자가 있는 여성의 출산율이 직접 산출되도록 로직을 구현하고 있다.

※ 분리형 방식 (27)

$$B_t(a) = P^f_{t,a,1,i=0} \times PPR_{0\to1}(a) + P^f_{t,a,1,i\geq1} \times PPR_{1+\to2+}(a)$$

※ 통합형 방식

$$B_t(a) = P^f_{t,a,1} \times PPR_{unified}(a)$$

여기서 $B_t(a)$는 t시점 a연령 여성이 출산한 아이 수이다. 현재 모형에서는 유배우 여성 인구를 무자녀/유자녀로 구분하지 않고, 전체 유배우 여성에게 통합 PPR을 적용한다. 분리형 방식을 적용하기 위해서는 배우

자가 있지만 자녀가 없는 여성과 배우자와 자녀가 모두 있는 여성, 즉, 가구 구성 정보를 가정해야 하나, 출산으로 인한 가구원수 변화 가정 간 정합성을 고려하기 위해 본 연구에서는 가구별 자녀수 정보 없이도 추계가 가능하도록 단순화한 통합형 방식을 활용하여 추계 결과를 산출하고 있다.

세 번째 단계는 인구추계 모듈에서 출생아수를 집계하고, 연령별 출산율(ASFR)과 합계출산율(TFR)을 계산하는 모듈이다. 이는 인구추계모듈(PopulationCalculator)에서 구현된다. 인구추계 모듈에서는 앞서 산출된 출생아수에 출생성비(남아 105 : 여아 100)를 적용하여 성별 출생아수를 계산한다.

$$B_t \tag{28}$$

이어서 출생아수를 여성 인구로 나누어 연령별 출산율(ASFR)을 계산하고, 이를 합산하여 합계출산율(TFR)을 도출한다.

$$ASFR_t(a) = B_t(a)/P_t^{female}(a), \quad TFR_t = \Sigma_{a=15}^{49} ASFR_t(a) \tag{29}$$

여기서, $P_t^{female}(a)$는 t시점 a연령의 여성 인구 수 이다. 이러한 3단계 과정은 매년 반복되며, 2023년부터 2055년까지는 PPR 기반 출생아 산출 → ASFR/TFR 역산 과정을 거치고, 2056년 이후는 2055년의 ASFR을 고정하여 기존 코호트요인법 추계 방법과 동일한 방식으로 추계가 진행된다.

FertilityCalculator 클래스에서는 초기 입력 변수로 시나리오를 구분할 수 있도록 시나리오 생성기관(agency: kostat or kihasa)과, 추계 실적 연도(base_year), 추계 종료연도(end_year), 출산력 추계 방법론('GLG', 'PPR', 'PPR_UNIFIED')를 입력값으로 활용한다. 여기서 추계 시작 연도는 base_year+1로 결정되며, agency와 model을 각각 kostat과 GLG를 선택할 경우, 통계청의 가정이 직접 적용될 수 있도록 구현하였다.

〈표 3-6〉 FertilityCalculator pseudo 코드

출생아수 추계 FertilityCalculator pseudo 코드

```
# 출산진도비 기반 출산력 추계

class FertilityCalculator:
    def __init__(self, agency, base_year, end_year, model='PPR_UNFIED'):
        """
        Args:
            agency: 기관 ('kostat', 'kihasa')
            base_year: 기준연도 (실적 데이터의 마지막 연도)
            end_year: 전망 종료 연도
            model: 모델 종류 ('GLG', 'PPR', 'PPR_UNIFED')
        """

    def estimate_ppr(self):
        """
        PPR 추정 메인 함수
        - 2015년, 2020년 PPR 추정
        - 요약 결과 엑셀 저장
        - GLG 형식 결과 반환
        """
        # 1. 2015년, 2020년 각각 추정
        for birth_year in [2015, 2020]:
            # 1-1. 인구 데이터 로드 (인구총조사 2% 표본)
            df_pop = pd.read_csv(f'{birth_year}_2%_표본_인구.csv')
            # 여성, 유배우, 15-49세 필터링

            # 1-2. 출생 데이터 로드
            df_birth = pd.read_csv(f'{birth_year}_출생_연간자료.csv')
            # 모 연령, 총출생아수 정보 추출

            # 1-3. PPR 계산
            # 15-24세: 그룹 평균
            # 25-49세: 각 연령별 개별 계산
            ppr_0_to_1[age] = births_첫째 / pop_무자녀
            ppr_1_to_2[age] = births_둘째이상 / (pop_1자녀이상 - births_첫째)
            ppr_unified[age] = births_전체 / pop_유배우전체

        # 2. 요약 저장
        # ppr_estimation_summary_2022.xlsx

        # 3. ASFR 형식으로 변환
        # ASFR배열 생성
        # 2020년까지: 실적 데이터
        # 2021년 이후: PPR 고정
        return results  # {'low': {afr, cfr}, 'mid': {afr, cfr}, 'high': {afr, cfr}}
```

자료: 저자 작성.

agency와 model을 각각 kihasa와 PPR을 선택할 경우 PPR 방식의 출산진도비를 계산하여 PopulationCalculator 클래스로 전달하게 되며, PopulationCalculator 클래스에서 인구추계가 수행되며 출생아수와 각 연령별 출산율(ASFR) 및 합계출산율(TFR)을 계산하여 출력하게 된다. PopulationCalculator 클래스는 이후 5장 통합 추계모형 구축에서 더 자세히 다루도록 한다.

2. 출산진도비(PPR) 추정

가. 주요 활용 데이터

출산진도비 추정을 위해서는 두 가지 핵심 데이터가 필요하다. 첫째는 특정 시점의 여성 인구 구성 정보이고, 둘째는 해당 기간의 출생 이력 데이터이다.

인구 구성 데이터는 통계청 「인구주택총조사」의 2% 표본조사 자료를 활용한다. 이 데이터에는 여성의 연령(만 15-49세), 혼인상태(미혼, 유배우, 사별, 이혼), 총출산아수(여아수 + 남아수)가 포함되어 있다. 본 연구에서는 2015년과 2020년 인구총조사 자료를 사용하여 각 시점의 출산진도비를 추정한다.

특히 유배우 여성 인구가 분모의 기준이 되므로, 혼인상태 코드가 2인 여성만을 추출한다. 또한 15-24세 구간은 표본 수가 적어 추정의 불안정성이 크므로 10개 연령을 하나의 그룹으로 묶어 처리한다. 25세 이상은 각 단일연령별로 구분하여 보다 정밀한 추정을 수행한다.

출생 데이터는 통계청 「출생통계」의 연간 마이크로데이터를 활용한다. 출생신고서에 기재된 모의 연령, 모의 총출생아수(현재 출생아 포함) 정

보를 통해 각 출산 순위별 출산 건수를 파악한다. 2015년 추정을 위해서는 2014-2015년 2개년 평균을, 2020년 추정을 위해서는 2019-2020년 2개년 평균을 사용하여 단년도 변동성을 완화한다.

코로나19의 영향을 고려하여 2020년 데이터의 가중치를 조정하는 방안도 검토하였으나, 본 연구에서는 2019-2020년 평균을 그대로 사용하는 것으로 결정하였다. 이는 2020년 이후의 출산율 하락이 일시적 충격이 아닌 구조적 변화일 가능성을 반영한 것이다.

나. 출산진도비 추정 방법

출산진도비는 특정 출산 순위에서 다음 순위로 진행하는 비율로 정의된다. 본 연구에서는 세 가지 유형의 출산진도비를 추정한다.

분리형 출산진도비는 출산 단계를 세분화하여 각 단계별 진행률을 개별적으로 산출하는 방식이다. 먼저, 'PPR 0→1'(무자녀 → 첫째아)는 무자녀 유배우 여성 중 첫째아를 출산하는 비율로 다음과 같이 산출 가능하다.

$$PPR_{0\to1}(a) = \frac{\text{첫째아 출산 건수}(a)}{\text{무자녀 유배우 여성 인구}(a)} \tag{30}$$

'PPR 1+→2+'(1자녀 이상 → 추가 출산)은 1자녀 이상 유배우 여성 중 추가 자녀를 출산하는 비율로 다음과 같이 산출 가능하다.

$$PPR_{1+\to2+}(a) = \frac{\text{둘째아 이상 출산 건수}(a)}{\text{1자녀 이상 유배우 여성 인구}(a) - \text{첫째아 출산 건수}(a)} \tag{31}$$

여기서 분모에서 첫째아 출산 건수를 제외하는 이유는, 해당 연도에 첫째아를 출산한 여성은 더 이상 1자녀 이상 상태에서 추가 출산의 위험에

노출되지 않기 때문이다.

통합형 출산진도비는 출산 단계를 구분하지 않고 전체 유배우 여성의 출산율을 산출하는 단순화된 방식으로 산출하며 유배우 여성의 출산율과 같다.

$$PPR_{unified}(a) = \frac{\text{전체 출산 건수}(a)}{\text{전체 유배우 여성 인구}(a)} \tag{32}$$

분리형 및 통합형 출산진도비 추정에서 15-24세 그룹 처리는 표본 수가 적은 저연령층의 추정 불안정성을 해소하기 위한 것이다. 15-24세 전체를 하나의 그룹으로 묶어 총 출산 건수와 총 인구를 합산한 후 비율을 계산하고, 이 그룹 평균값을 각 단일 연령(15세, 16세, ..., 24세)에 동일하게 적용한다. 아래 수식은 통합형 PPR의 15~24세 산식이다.

$$PPR_{15-24} = \frac{\Sigma_{a=15}^{24}\text{출산 건수}(a)}{\Sigma_{a=15}^{24}\text{유배우 여성 인구}(a)} \tag{33}$$

이러한 그룹화 처리는 25세 이상에서는 적용하지 않으며, 각 단일 연령별로 개별 추정을 수행한다.

3. 주요 결과 검토

가. 혼인상태별 이동 추정 결과

본 연구에서 구축한 혼인상태별 전이 모형의 타당성을 검증하기 위해, 2020년 인구총조사 기반 초기 전이율을 2022년까지 적용하여 추정한 혼인상태별 인구 구성비를 통계청의 2022년 실적과 비교하였다.

검증 절차는 다음과 같다. 첫째, 2020년 인구총조사의 혼인상태별 인구 비중을 2020년 확정인구에 적용하여 초기값으로 설정한다. 둘째, 2020년 실적 데이터로 추정한 전이율을 적용하여 2021년, 2022년 혼인상태별 인구를 순차적으로 계산한다. 이 과정에서 인구추계로부터 산출된 2021년, 2022년 총인구와의 정합성을 확보하기 위해 앞서 제시한 조정 절차를 적용한다. 셋째, 2022년 추정 결과를 통계청의 2022년 혼인상태별 실적과 비교한다.

〈표 3-7〉은 2022년 기준 본 연구의 추정 결과와 통계청 실적을 연령대별·성별로 비교한 것이다. 전반적으로 본 연구의 추정치는 통계청 실적과 매우 유사한 패턴을 보이며, 대부분의 연령대에서 ±1%p 이내의 오차를 나타냈다.

남성의 경우, 15~19세는 미혼이 100.0%로 실측치와 동일하다. 2024세는 미혼 99.4%(통계청 99.2%), 유배우 0.5%(통계청 0.6%)로 0.2%p 이내의 차이를 보인다. 25~29세에서는 미혼 93.5%로 통계청과 정확히 일치하며, 유배우는 5.8%(통계청 6.0%)로 0.2%p 차이가 나타난다.

30대 이후 혼인 구성이 본격적으로 변화하는 연령대에서도 추정의 정확도가 유지된다. 30~34세는 미혼 69.7%(통계청 70.7%), 유배우 29.0%(통계청 28.2%)로 미혼에서 1.0%p, 유배우에서 0.8%p의 차이를 보인다. 3539세는 미혼 41.1%(통계청 42.1%), 유배우 56.1%(통계청 55.5%)로 각각 1.0%p, 0.6%p 차이이다. 40대에서는 미혼·유배우 모두 0.5%p 이내로 거의 일치하는 것으로 나타났다.

이혼과 사별 비율은 상대적으로 작지만 연령 증가에 따라 상승하는 패턴을 정확히 포착하였다. 40~44세 이혼 비율은 4.4%(통계청 4.2%), 45~49세는 7.2%(통계청 6.7%)로 각각 0.2~0.5%p 차이를 보인다. 사별 비율은 모든 연령대에서 0.5% 미만으로 낮으며, 추정치와 실측치가 0.1%p 이내에서 일치하는 것으로 나타났다.

〈표 3-7〉 혼인상태별 이동 추정 결과 비교

(단위: %)

연령대	본 연구				통계청			
	2022년				2022년			
	미혼	유배우	사별	이혼	미혼	유배우	사별	이혼
남성								
15-19세	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
20-24세	99.4	0.5	0.0	0.1	99.2	0.6	0.0	0.1
25-29세	93.5	5.8	0.2	0.5	93.5	6.0	0.1	0.4
30-34세	69.7	29.0	0.2	1.1	70.7	28.2	0.2	0.9
35-39세	41.1	56.1	0.2	2.6	42.1	55.5	0.2	2.2
40-44세	28.7	66.5	0.3	4.4	29.1	66.4	0.3	4.2
45-49세	22.4	69.9	0.5	7.2	22.1	70.6	0.5	6.7
여성								
15-19세	99.9	0.1	0.0	0.0	99.9	0.1	0.0	0.0
20-24세	98.2	1.6	0.0	0.2	97.6	2.2	0.0	0.1
25-29세	85.4	13.7	0.0	0.8	84.9	14.4	0.0	0.6
30-34세	50.3	47.5	0.2	2.1	51.8	46.2	0.1	1.9
35-39세	26.0	69.6	0.3	4.1	26.4	69.5	0.3	3.9
40-44세	16.7	76.1	0.7	6.5	16.7	76.4	0.7	6.2
45-49세	11.3	76.9	1.5	10.3	11.3	77.4	1.5	9.8

주: 통계청 결과는 2022-2052 가구추계의 실적.
자료: 본 연구의 결과는 저자 산출.

여성의 경우에도 유사한 정확도를 보인다. 15~19세와 20~24세는 미혼이 각각 99.9%, 98.2%로 통계청(99.9%, 97.6%)과 거의 일치한다. 25~29세는 미혼 85.4%(통계청 84.9%), 유배우 13.7%(통계청 14.4%)로 0.5~0.7%p 차이이다.

여성은 남성보다 결혼 연령이 빠르므로 30대에서 유배우 비율이 높게 나타나는데, 이러한 패턴도 정확히 반영되었다. 30~34세는 유배우 47.5%(통계청 46.2%)로 1.3%p 차이, 35~39세는 69.6%(통계청 69.5%)로 0.1%p 차이를 보인다. 40대 이후에도 유배우 비율이 76% 수준을 유지하며, 추정

치는 통계청 대비 ±0.5%p 이내로 거의 일치하는 것으로 나타났다.

이혼 비율은 연령 증가에 따라 상승하여 45~49세에서 10.3%(통계청 9.8%)에 달하며, 0.5%p의 차이를 보인다. 사별 비율도 연령 증가에 따라 상승하지만 40대에서는 1% 미만 수준으로 낮으며, 추정치와 실측치가 0.1%p 이내로 정확히 일치하는 것으로 나타났다.

본 연구의 혼인상태별 전이 모형은 2020년 전이율을 2년간 적용하여 2022년 혼인상태 구성을 매우 정확하게 추정하였다. 대부분의 연령대·성별·혼인상태에서 통계청 실측치와 ±1%p 이내의 오차를 보이며, 특히 주요 출산 연령대인 25~39세 유배우 여성 비율의 차이는 0.1~1.3%p의 높은 정확도를 나타냈다.

이러한 결과는 다음 두 가지를 시사한다. 첫째, 인구동향 실적 데이터(혼인·사망·이혼)를 기반으로 한 전이율 추정 방법이 타당함을 보여준다. 2019~2020년 평균 실적으로 추정한 전이율이 2021~2022년에도 안정적으로 적용되었다는 것은 단기적 추계에서 실적 기반 전이율의 신뢰성이 높음을 의미한다.

둘째, 인구추계 결과와의 정합성을 확보하기 위한 7단계 역산 조정 절차가 효과적이다. 전이율 적용 후 총인구와의 차이를 비례 배분하고 전이율을 역산하는 방식은 혼인상태별 인구의 합계를 인구추계와 일치시키면서도, 각 혼인상태의 구성비를 실측치에 가깝게 유지하였다.

향후 장기 추계(2023~2055년)에서는 이러한 단기 검증 결과를 바탕으로, 초혼율 증가 시나리오(저위 0%, 중위 1.5%, 고위 3.0%)를 적용하여 혼인상태 구성의 변화를 전망한다.

나. PPR 추정 결과

2015년과 2020년의 출산진도비 추정 결과 다음과 같다.

〈표 3-8〉 출산진도비 추정 결과

(단위: %)

연령	2015년			2020년		
	PPRunifed	PPR0→1	PPR1+→\2+	PPRunifed	PPR0→1	PPR1+→\2+
15-24세	34.5	62.1	26.5	27.3	42.2	22.3
25세	25.9	46.2	18.4	23.5	33.9	20.3
30세	23.4	42.0	17.3	16.8	25.6	12.5
35세	10.6	31.0	8.0	10.1	24.4	7.2
40세	1.7	8.9	1.2	2.0	8.2	1.3
45세	0.1	0.4	0.0	0.1	0.5	0.0
49세	0.0	0.0	0.0	0.0	0.0	0.0

자료: 인구총조사 2%샘플 및 인구동향조사 출산 자료 활용 저자 계산.

출산진도비 추정 결과 ‘PPR 0→1’은 모든 연령대에서 2015년 대비 2020년 PPR이 하락하였다. 반면, 'PPR 1+→2+'의 2015년 대비 2020년 추이를 살펴보면, 30대 중반까지는 'PPR 0→1'과

마찬가지로 하락세를 보이나, 30대 중반(37세) 이후부터는 소폭 증가하였다. 이는 첫째아 출산이 지연되면서 늦은 나이에도 둘째아 이상 출산을 시도하는 여성이 증가했음을 시사한다. 다만 절대적인 수준은 여전히 낮아(40세 1.3%), 고령 출산의 어려움이 반영된 것으로 해석된다. 이렇게 추정된 출산진도비는 이후 인구추계 모형 PopulationCalculator 클래스에서 직접 활용되며, 추계 기간(2023-2055년)에 걸쳐 2020년 수준이 유지된다고 가정한다. 2055년 이후는 PPR을 활용하여 산출한 ASFR과 TFR이 지속 유지되는 가정을 활용한다.

다. 출산력 추계 결과

본 연구의 출산력 추계 결과는 〈표 3-9〉와 같다. 본 연구의 첫 번째 시나리오는 혼인상태 전이율(MaritalTransitionCalculator) 모듈에서 '미혼→유배우' 전이율(이하 초혼율)을 유지(+0.0%)하는 시나리오이며, 두 번째 시나리오는 초혼율이 '1.5%/연' 증가하는 시나리오다. 세 번째 시나리오는 초혼율이 '3.0%/연' 증가하는 것을 가정한 시나리오이다.

이러한 가정의 이유는 PPR 방식이 유배우 여성의 출산율을 기반으로 하므로, 혼인율 변화가 전체 출산율에 결정적 영향을 미치기 때문이며, 초혼율 증가 수준의 가정은 통계청의 TFR 가정과 직접 비교 가능하도록 설계하였다. 또한, 추후 혼인상태의 변화 시나리오를 상정하거나 혼인상태별 출산율의 변화 시나리오를 상정하는 경우 등 다양한 시나리오 적용이 가능하다.

현재 설정되어 있는 초혼율 '+1.5%/연' 시나리오(2055년 TFR 0.84)는 통계청의 저위 시나리오(0.82)와 유사한 수준이며, 초혼율 '+3.0/연' 시나리오(2055년 TFR 1.08)는 통계청의 중위(1.08) 수준이다. 이는 PPR이 2020년 수준으로 고정되어 있고, 혼인율 개선만으로는 통계청이 가정한 출산율 회복 시나리오에도 도달하기 어렵다는 것을 의미한다.

특히 본 연구에서 가정한 초혼율 '+3.0/연' 시나리오는 상당히 낙관적인 가정임에도 불구하고 2050년 TFR이 1.00에 그치고 있다. 이는, 초혼율 개선만으로는 출산율 제고에 한계가 있으며, 유배우 여성의 출산 의향 자체를 높이는 정책(PPR 상승)이 병행되어야 함을 시사한다.

〈표 3-9〉 합계출산율 비교 결과

(단위: 명)

연도	통계청			본 연구		
	저위	중위	고위	초혼율 유지	초혼율 +1.5%/연	초혼율 +3.0%/연
2023년	0.72	0.72	0.72	0.74	0.74	0.74
2025년	0.63	0.65	0.75	0.69	0.70	0.71
2030년	0.67	0.82	0.98	0.63	0.67	0.71
2035년	0.76	0.99	1.21	0.62	0.69	0.77
2040년	0.80	1.05	1.30	0.62	0.73	0.85
2045년	0.81	1.06	1.31	0.62	0.77	0.93
2050년	0.82	1.08	1.33	0.62	0.80	1.00
2055년	0.82	1.08	1.34	0.62	0.84	1.08

자료: 저자 산출.

라. 구축 모형의 한계 및 개선 방향

현재 구현된 PPR 기반 출산력 추계모형의 한계를 정리하면 크게 세 가지로 정리해볼 수 있다.

먼저, PPR이 고정되는 가정이다. 사회경제적 여건, 정책 변화에 따라 변화할 가능성이 큰 상황에서 2020년 수준의 PPR이 2055년까지 유지된다고 가정하는 것은 강한 가정으로 볼 수 있다. 다만, 본 연구에서는 이를 시뮬레이션 할 수 있는 구조로 설계한 것이며, 추후 PPR의 시나리오에 따라 여러 상황을 점검해 볼 수 있을 것으로 기대한다.

두 번째는 출산-가구 구성 간 순환 구조의 문제다. 출생아가 즉시 가구원수 증가로 반영되는 부분 구현의 한계로 인해 기존 가구 유형(부부+자녀) 추계를 직접 적용하지 않았다. 5장에서 상세하게 논의하겠지만, 인구와 가구 추계의 통합 부분에서 반드시 개선되어야 할 부분이다. 본 연구에서는 출산순위별 PPR을 산출하였으나, 이러한 한계로 통합 PPR만 사

용하여 인구추계 결과를 산출하였다. 추후 정책 효과의 순위별 차별성 분석 등을 위해 이러한 한계는 추후 과제로 다뤄져야 할 것이다.

세 번째는 혼인상태 전이의 불확실성이다. 혼인상태별 인구 구성이 PPR 방식의 핵심 요소로, 혼인상태 전이율의 예측 불확실성이 출산력 추계에 직접 영향을 미치기 때문에 혼인 추계의 합리성 역시 제고될 필요가 있다.

다만 이러한 한계에도 불구하고, PPR 기반 접근법은 출산 행동의 미시적 메커니즘을 명시적으로 모델링하고, 혼인-출산 연계를 체계적으로 반영할 수 있다는 점에서 의의가 있다. 향후 단계적 개선을 통해 보다 정교한 추계 체계로 발전시킬 수 있을 것이다.

제4장

사망력 추계 방법론 개선

제1절 기존 사망력 추계 방법론 검토
제2절 사망력 추계 개선 방안
제3절 사망력 추계 개선 구현 및 결과 검토

제4장 사망력 추계 방법론 개선

제1절 기존 사망력 추계 방법론 검토

이번 절은 인구변동요인 가운데 사망 관련 사망력(mortality)을 전망하는 방법론을 소개한다. 선행적으로 통계청의 방법론을 소개하고, 사망력 파트의 개선사항을 도출한다.

1. 통계청 방법론

가. 사망 추이 분석

통계청(2024)은 사망추계 방법을 세 부분으로 나누어 살펴본다. 먼저 사망자수와 사망률 추이를 살펴보고, 이들 자료를 생명표 작성방법에 대입하여 기대수명 추이를 알아본다.

우리나라 사망률은 지속적으로 개선되고 있으나 고연령층 인구의 증가 속도가 빨라지면서 사망자수가 예전에 비해 높아지고 있다. 기대수명은 1970~2021년 기간 동안 지속적으로 증가되어 왔으나 최근 추세는 둔화됨을 보인다. 남녀 간 기대수명 차이는 1970년 7.1년에서 1985년 8.6년까지 증가한 후 감소하여 2022년 5.8세로 나타났다. 다음으로 사망 추계는 Li-Lee-Gerland(2013)[16] 확장모형을 적용하여 성·연령별 사망률을

16) Li-Lee-Gerland(2013)에서 제안한 LC-ER모형을 다중인구로 확장해서 적용함. Li, N., Lee, R., and Gerland, P.(2013), "E of mortality decline for long-term projections", Demography, 50(6), 2037-2051xtending the Lee-Carter method to model the

예측한다. 1970~2021년 성과 연령별 사망률 자료를 활용하여 2072년까지 성, 연령별로 사망률을 예측한다. 이들 모형에 대한 변천과 내용은 다음 절 사망률 전망모형에서 상세히 다룬다. 끝으로 사망 추계가정이다. 장래 사망의 변화를 감안하여 4개의 기대수명 수준을 가정한다. 중위가정 기대수명은 2072년 남자 89.5년, 여자 92.7년, 고위가정은 2072년 남자 90.3년, 여자 94.0년, 저위가정[17]은 2072년 남자 88.3년, 여자 91.1년이다. 그리고 사망력 코로나19 장기 영향 가정 기대수명은 2072년 남자 88.8년, 여자 92.3년이다. 〈표 4-1〉은 2022~2072년 통계청(2024) 기대수명 추계결과이다.

〈표 4-1〉 통계청 사망력 가정별 기대수명

(단위: 년)

가정	구분	2022*	2023*	2024	2025	2030	2040	2050	2060	2072
중위	전체	82.7	83.5	84.3	84.5	85.5	87.2	88.6	89.8	91.1
	남자(A)	79.9	80.6	81.4	81.6	82.8	84.8	86.5	88.0	89.5
	여자(B)	85.6	86.4	87.1	87.3	88.1	89.5	90.7	91.7	92.7
	A-B	5.8	5.9	5.7	5.7	5.3	4.7	4.2	3.7	3.3
고위	전체	82.7	83.7	84.8	85.1	86.3	88.2	89.8	91.0	92.2
	남자(A)	79.9	80.8	81.9	82.2	83.6	85.8	87.6	88.9	90.3
	여자(B)	85.6	86.6	87.7	88.0	89.0	90.6	91.9	93.0	94.0
	A-B	5.8	5.8	5.8	5.8	5.4	4.8	4.3	4.1	3.7
저위	전체	82.7	83.3	83.7	83.8	84.4	86.0	87.3	88.4	89.7
	남자(A)	79.9	80.3	80.8	81.0	81.9	83.7	85.3	86.7	88.3
	여자(B)	85.6	86.2	86.5	86.5	87.0	88.3	89.3	90.1	91.1
	A-B	5.8	5.9	5.7	5.5	5.1	4.6	4.0	3.4	2.8
사망력 코로나19 장기 영향	전체	82.7	83.5	83.7	83.9	84.9	86.6	88.1	89.3	90.6
	남자(A)	79.9	80.6	80.9	81.1	82.2	84.2	85.9	87.3	88.8
	여자(B)	85.6	86.4	86.6	86.8	87.6	89.0	90.2	91.2	92.3
	A-B	5.8	5.9	5.8	5.7	5.4	4.8	4.3	3.9	3.5

주: * 2022년 생명표와 2023년 인구동향조사(사망통계) 등을 기초로 생명표를 별도 작성.
자료: 통계청(2024), 장래인구추계: 2022~2072년.

rotation of age pattern.

17) 중위, 고위, 그리고 저위는 Li-Lee-Gerland(LLG)(2013) 확장모형의 제50분위수, 제99백분위수, 제1백분위수로 각각 설정함.

나. 사망률 추계 모형

사망률 추계를 위한 대표적인 모형으로는 Lee-Carter(1992) 모형을 비롯하여 이를 개선한 다양한 모델이 존재한다. 대표적인 개선 모형으로는 LM(2001), BMS(2002), Li-Lee(2005), 젊은 층과 노년층의 사망률 개선 양상이 과거와 다르게 변화하는 사망률 개선 교대(declined mortality rate with rotation) 현상을 반영한 LC-ER 모형(Li-Lee-Gerland, 2013)이 있다. 또한, LC-ER 모형에 인구 부집단(sub-population)의 정합성을 확보하기 위해 Li-Lee(2005) 방식을 추가한 LC-ER and LL 조합 모형(Statistics Canada, 2015)이 개발되었다.

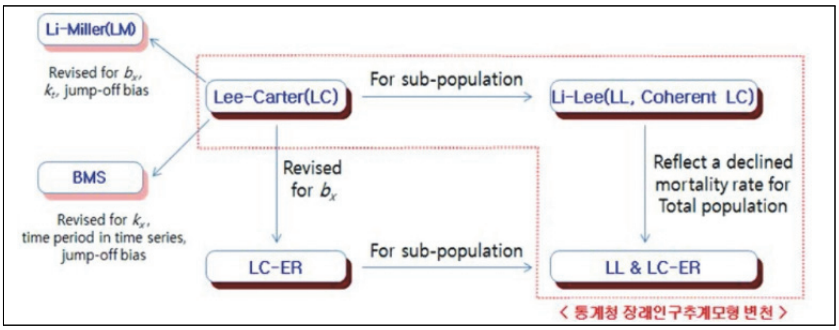
통계청은 [그림 4-1]과 같이 사망률 예측 모형으로 LC, LL, LC-ER and LL 모형(통계청, 2016, 'LLG 확장 모형'으로 명칭)을 순차적으로 개선과 수정을 통해 적용해 왔다. 이들 모형의 특징은 다음과 같다.

LC 모형은 1992년 Lee와 Carter가 제안한 대표적인 사망률 예측 모형이다. 이 모형은 로그 변환된 사망률을 연령 효과(age effect)와 시간 효과(period effect)의 선형 결합으로 표현하며, 로그 연령별 사망률(age-specific mortality) 예측 모델의 수식 구조는 식 (34)와 같다.

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t} \quad , t = 1, 2, \cdots, n \tag{34}$$

식 (34)의 $m_{x,t}$는 연령 x와 시간 t에서의 사망률(death rate), a_x는 연령에 따른 로그사망률의 평균수준, b_x는 각 연령에서 로그사망률의 변화, k_t는 시간 t에 따른 로그사망률 변화 수준을 나타내고, $\epsilon_{x,t}$는 평균 0, 분산 σ_ϵ^2이고, 관측되지 않은 오차를 의미한다.

〔그림 4-1〕 통계청 사망률 모형 변천



자료: 오진호·김순영(2018), figure 2.3.

연령에 따른 평균적인 로그 사망률 수준(a_x)과 각 연령에서 로그 사망률의 변화율(b_x)은 시간에 무관한 상수이며, 전반적인 사망률 개선 정도를 나타내는 사망률 지수(mortality index, k_t)만이 시간에 의존한다. 따라서 k_t의 예측을 통해 미래 사망률을 전망하게 된다. 모수 a_x는 시간에 대한 $\ln(m_{x,t})$의 평균으로 계산된다. 또한, b_x와 k_t는 $[\ln(m_{x,t}) - a_x]$에 대해 비정칙분해(Singular Value Decomposition, SVD) 방법을 적용하여 얻어진 첫 번째 주성분을 사용해 추정된다. 이때, b_x와 k_t의 유일한 추정치를 보장하기 위해 식(35)과 같은 제약조건을 적용한다.

$$\sum_t k_t = 0,\ \sum_x b_x = 1 \tag{35}$$

또한, 추정된 $\hat{k}_t$는 총 사망자수($D_t = \sum_x D_{x,t}$)가 실제 사망자수와 일치하도록 조정한다. 여기서, $D_{x,t}$는 t연도, x연령의 사망자수이며, $N(x,t)$는 t연도, x연령의 인구수이다. 세 개의 모수 중 a_x와 b_x는 자료로부터 직접 추정되고, k_t만이 시간에 의존하는 모수이므로, LC모형에서의 핵심

은 k_t를 얼마나 정확하게 추정하고 예측하는지에 달려 있다. 추정된 $\hat{k}_t$의 예측은 기본적으로 식 (35)과 같은 절편을 포함한 임의보행모형(random walk with drift:RWD)을 사용하며, 필요에 따라 ARIMA모형을 적용하여 적합성을 평가할 수도 있다.

$$k_t = k_{t-1} + d + e_t \tag{36}$$

식 (36)의 d는 임의보행과정의 추세를 나타내는 절편모수로 k_t의 평균 연간 변화를 나타내고, e_t는 평균이 0이고 분산이 σ_ϵ^2인 정규분포를 따르는 확률변수이다.

예측된 $\hat{k}_{n+j}$ $(j = 1, 2, \cdots)$와 추정된 $\hat{a}_x$와 $\hat{b}_x$을 식 (28)에 대입하면 미래의 로그사망률을 예측할 수 있다. 즉, $k_{n+h|n}$의 h-단계 예측(h-step-ahead forecasts)을 $\hat{k}_{n+h|n}$로 표기한다면 $\ln(m_{x,n+h})$의 h-단계 예측은 식 (37)와 같다.

$$\ln(m_{x,n+h}) = \hat{a}_x + \hat{b}_x \hat{k}_{n+h|n} \tag{37}$$

또한 첫 번째 주성분(b_x)과 식 (34)의 오차항 간의 직교성(orthogonality)으로 인해, 전반적인 예측분산은 식 (38)과 같이 도출된다.

$$Var[\ln m_{x,t}] \approx \hat{b}_x^2 u_{n+h|n} + v_x \tag{38}$$

식 (38)의 $\hat{b}_x^2$는 첫 번째 주성분 분산이며, $u_{n+h|n} = Var(k_{n+h}|k_1, \cdots, k_n)$는 시계열 모형으로부터 도출된 예측 오차이다.

또한 모형의 잔차 분산(v_x)은 식 (34)에서 각 연령별 잔차 제곱 $\{\epsilon_{x,1}^2, \cdots, \epsilon_{x,n}^2\}$의 평균을 통해 추정된다. 이러한 예측 분산을 이용하면 미

래 사망률 예측의 불확실성을 반영한 신뢰구간(confidence interval)을 도출할 수 있다.

LC 모형은 인구집단을 특정 특성(예: 성별(Sex) 또는 지역(Region))을 기준으로 나누어 다중 인구(multiple population) 또는 인구 부집단(sub-population)으로 분석할 경우, 사회경제적 조건이 유사한 인구집단 간 사망률 패턴에는 일정한 연관성이 존재한다.

그러나 이러한 다중 인구 그룹에 LC 모형을 적용할 경우, 개별 집단 간 사망률 변화(k_t)값이 서로 다르게 추정되므로 집단 간 연관성을 반영할 수 없다. 이로 인해, 집단 간 사망률 격차가 발생하며, 장기 예측에서 합리적인 결과를 도출하기 어려울 가능성이 높다.

이에 따라, 인구학자들은 인구사회학적으로 밀접하게 연관된 하위 집단 간 사망률 격차가 장기적으로 지속적으로 확대되지 않도록 보장하는 내적 일관성(coherence)을 충족하는 사망률 예측 모형이 필요하다고 주장하였다. 이에 따라, Li와 Lee(2005)는 사회경제적 조건이 유사한 인구집단의 사망률 패턴이 장기적으로 공통 사망률(common mortality)로 수렴할 것이라는 가정을 기반으로, 기존 LC 모형을 다중 인구 분석이 가능하도록 확장하였다. 이를 위해, 기존 LC 모형에 공통(coherent) 사망 경향을 반영한 시간×연령 효과(time-age effect)를 추가하여, 식 (34)를 개선한 식 (39)의 사망률 예측 모형을 제시하였다.

$$\ln(m_{x,t,i}) = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + \epsilon_{x,t,i} \tag{39}$$

LC 모형과 비교할 때, 가장 큰 차이점은 우변의 $B_x K_t$항이다. 이 항은 모든 그룹의 공통 사망 경향을 반영하며, 전체 집단의 연령별 로그 사망률 변화 정도(B_x)와 시간에 따른 로그 사망률 수준 변화(K_t)를 나타낸다. 또한, 하위 집단의 사망률이 장기적으로 발산하지 않도록 조정하는 역할

을 수행한다.

한편, $b_{x,i}\,k_{t,i}$는 i번째 그룹의 개별 사망 경향을 의미하며, 개별 집단의 연령별 로그사망률 변화 정도($b_{x,i}$)와 시간에 따른 로그사망률 수준 변화($k_{t,i}$)를 반영한다. 이는 하위 집단들이 공통 추세에서 벗어나는 단기적 변동을 설명하는 역할을 한다. 또한 $a_{x,i}$는 그룹 i의 연령에 따른 평균적인 로그 사망률 수준을 나타낸다. 여기서, 그룹은 인접 국가(지역) 또는 동일 국가 내 성별 등의 구분 기준이 될 수 있다. 이러한 특징을 고려할 때, LL 모형은 LC 모형의 확장 형태로 볼 수 있으며, 공통 사망률 패턴을 반영한 사망률 추계에 활용된다.

LL모형은 LC모형의 모수추정 과정을 두 번 적용한다. 먼저 그룹을 구분하지 않고 전체 집단을 대상으로 식 (34)와 같이 LC 모형을 적용하여 비정칙분해(SVD) 방법에 의하여 공통 사망 경향 $B_x K_t$을 추출한다. 여기서, A_x는 공통 사망 경향에 대한 평균사망률을 의미한다.

LL 모형은 LC 모형의 모수 추정 과정을 두 번 적용하는 방식으로 구성된다. 먼저, 그룹을 구분하지 않고 전체 집단을 대상으로 식 (34)에 따라 LC 모형을 적용하여 비정칙 분해(Singular Value Decomposition, SVD) 방법을 이용해 공통 사망 경향($B_x K_t$)을 식 (40)와 같이 추정한다.

$$\ln(m_{x,t}) = A_x + B_x K_t + e_{x,t} \tag{40}$$

여기서, A_x는 공통 사망 경향에 대한 평균 사망률을 의미한다.

다음 단계에서는 앞서 추출한 공통 사망 경향의 추정치($\widehat{B_x}\widehat{K_t}$)를 기반으로, 각각의 인구 부집단(sub-population)에서 도출된 잔차(residuals)($\left[\ln(m_{x,t,i}) - a_{x,i} - \widehat{B_x}\widehat{K_t}\right]$)에 비정칙 분해(Singular Value Decomposition, SVD) 방법을 다시 적용하여 개별 그룹의 사망 경향

($b_{x,i}\,k_{t,i}$)을 추정한다. 또한, 추정된 $\widehat{k_t}$와 $\widehat{K_t}$의 예측은 LC 모형과 마찬가지로 기본적으로 임의 보행 모형(random walk model)을 사용하지만, 필요에 따라 ARIMA 모형을 적용하여 적합성을 평가할 수 있다.

최근 사망률이 낮은 선진국과 우리나라는 과거에 비해 유소년(영아와 아동, infant and child) 사망률 개선 속도는 점진적으로 감소하는 반면, 고령층의 사망률 개선 속도는 점차 빨라지는 현상을 보이고 있다. 이러한 사망률 개선 교대(declined mortality rate with rotation) 현상은 여러 선행연구에서 보고되었다(Horiuchi와 Wilmoth, 1995; Li와 Gerland, 2011; Li 등, 2013; 김순영과 오진호, 2017). 이는 유소년층과 노년층의 사망률 감소 패턴이 교대(rotation)되는 현상으로, 기대수명이 증가함에 따라 연령별 사망률 변화 양상이 달라지고 있음을 의미한다. 즉, LC(1992) 모형에서 연령별 사망률 변화를 나타내는 b_x값이 시간의 경과에 따라 변하고 있음을 시사한다. Li와 Gerland(2011)는 이러한 현상을 사망률 개선 교대(mortality decline with rotation)라고 명명하고, 시간의 변화에 따라 연령별 사망률 개선 패턴이 달라지는 현상을 반영하기 위해 LC 모형의 b_x에 강건한 순환(robust rotation)을 도입한 LC-RR(Lee-Carter method with robust rotation) 모형을 제안하였다. 이후 Li 등(2013)은 LC-RR 모형을 확장하여 회전(rotation) 모델이 보다 유연하게 변할 수 있도록 개선된 LC-ER 모형을 제안하였다. LC-ER 모형(Lee-Carter model with extended rotation)은 LC 모형에서 시간의 변화와 관계없이 일정했던 b_x값을 시간 변수(time-varying parameter)를 고려한 $B_{x,t}$형태로 변환한 방식으로, 그 구조는 식 (41)과 같다.

$$\ln(m_{x,t}) = a_x + B_{x,t}K_t + \epsilon_{x,t}, \tag{41}$$

식 (41)의 $B_{x,t}$는 교대 이전의 연령별 사망률인 LC 모형의 b_x와 사망률 개선의 최고점 사망률 b_x^u의 선형 가중평균으로 식 (42)과 같이 정의된다. 식 (42)의 e_0^u는 사망률 개선 교대 현상이 종료되는 시점의 기대수명을 의미한다. 이는 LC 모형 계열에서 연령별 사망률 개선 패턴이 매년 동일하게 유지된다는(b_x) 단점을 보완한 개념이다. LC-ER 모형은 선진 20개국의 사망률 분석을 기반으로 몇 가지 실험적 결과값과 가정을 적용한 모델이다. 식 (42)에 따르면, 연령별 사망률 개선 패턴은 연령 80세에 도달한 시점에서 교대 현상이 시작되며, 100세까지 지속된 후 일정하게 유지되는 것으로 가정한다. 이는 65세 미만 연령대에서는 사망률 개선 패턴이 동일하지만, 65세 이후에는 연령이 증가할수록 사망률 개선 속도가 점차 감소함을 의미한다.

$$B_{x,t} = \begin{cases} b_x, & e_0^t < 80, \\ (1 - w_s(t))b_x + w_s(t)b_x^u, & 80 \le e_0^t < e_0^u \\ b_x^u & e_0^u < e_0^t \end{cases} \tag{42}$$

$$b_x^u = \begin{cases} \bar{b}_{15-64}, & 0 \le x \le 64 \\ b_x \times \dfrac{b_{u,60-64}}{b_{65-70}}, & 65 \le x \end{cases}$$

$$w(t) = \frac{e_0^t - 80}{e_0^u - 80};\ w_s(t) = \left[0.5\{1 + \sin(\frac{\pi}{2}(2w(t) - 1))\}\right]^p$$

또한, $B_{x,t}$의 임의 선형 변환을 통해, K_t는 교대 없이 도출된 기존 LC 모형의 기대수명(k_t)와 비교하여 두 값의 차이가 가장 작아지는 K_t를 찾기 위해 반복 계산을 수행한다. 이러한 과정을 통해 b_x^u, e_0^u 그리고 $B_{x,t}$를 도출한 후, 성별과 K_t값을 기반으로 e_0^u을 예측할 수 있다.

캐나다 통계청(2015)과 통계청(2016)은 사회·경제적 조건이 유사한 인구 부집단(예: 시도별)의 사망률 패턴을 반영하기 위해 LL 모형과 LC-ER

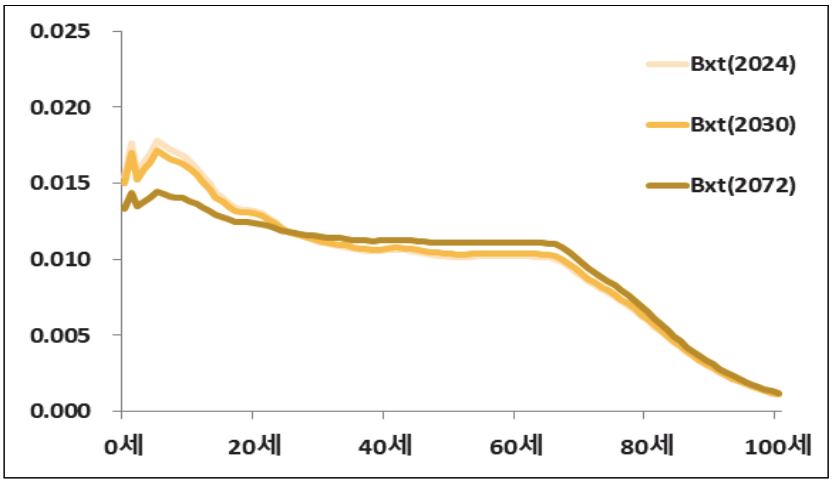
모형을 결합한 LL and LC-ER모형을 식 (43)과 같이 제안하였다.

$$\log m_{x,t,i} = a_{x,i} + B_{x,t}K_t + b_{x,i}k_{t,i} + \epsilon_{x,t,i} \tag{43}$$

여기서 $b_{x,i}k_{t,i}$는 개별집단(예: 남녀)의 연령별 사망률 변화 패턴과 시간 지수를 나타낸다.

정리하면, 통계청은 기존 LC 모형을 기반으로 인구부집단 간 차이, 젊은 층과 장년층의 사망률 개선 속도 차이를 보완하면서 LL모형, 그리고 LL and LC-ER (2016, 'LLG 확장 모형')까지 적용해 오고 있다. 최근 확인 결과 통계청(2024) 사망률 모형은 LLG확장 모형을 적용하고 있다. 〔그림 4-2〕는 LLG 확장모형에서 $B_{x,t}$의 추이를 보여준다. 시간에 흐름에 따라 젊은 층 사망률 개선은 낮아지고, 65세 이후는 높아지는 경향을 확인할 수 있다.

〔그림 4-2〕 LLG 확장모형의 $B_{x,t}$ 변화



자료: 통계청(2024), 장래인구추계: 2022~2072년.

2. 사망력 추계 선행연구 검토

가. 사망률 추이와 추계 모형의 발전

사망력 추계 모형은 시대별 사망률 개선 양상을 반영하기 위해 다양한 방식으로 발전해 왔다. 대표적으로 Lee-Carter(LC) 모형을 기반으로 한 계열 모형들이 있으며, 사망률 교차 현상(mortality crossover)을 보정하거나 고령층 사망률 개선 효과를 반영하는 LL, LC-ER, 그리고 LL과 LC-ER의 결합 모형 등이 연구되어 왔다.

이러한 모형 변천은 공중보건 향상, 보건의료 기술 발전 등으로 인한 사망률 감소, 특히 영유아 사망률의 급감과 고령층 생존율 증가에 대한 대응의 일환이다. 통계청(2024)은 국내 사망률 추이를 보다 정확히 반영하기 위해 LC, LL, LC-ER 및 이들의 조합 모형을 혼합하여 적용하고 있다.

전 세계적으로 가장 널리 사용되는 모형은 LC 모형이지만, 고령화가 급속히 진행되었거나 이미 초고령사회에 진입한 국가들—예: 미국, 일본, 프랑스, 독일, 한국, 중국, 캐나다—에서는 LC-ER 또는 LL과 LC-ER 조합모형이 보다 적합한 대안으로 활용되고 있다.

나. 사망률 자료 접근 방식과 모형 선택

사망률 모형은 자료의 질과 정밀도에 따라 크게 달라진다. 선진국의 경우, 성별과 연령별 사망자수가 0세부터 100세 이상까지 매년 집계되며, 이를 기반으로 생명표와 기대여명 자료가 생산된다. 이러한 상세한 데이터를 토대로 모수적(parametric) 접근법을 적용하여 장기 사망률을 예측할 수 있다.

반면, 개발도상국이나 저소득국가에서는 사망률 자료의 정비 수준이 낮

고, 성·연령별 정보가 5세 또는 10세 단위로 제공되며, 통계 기간도 1년이 아닌 5년 또는 10년 단위인 경우가 많다. 이로 인해 완비된 자료에 기반을 둔 모수적 접근은 어려우며, 일반적으로 비모수적(non-parametric), 베이지안(Bayesian) 접근법을 활용하게 된다.

유엔(UN)은 「세계인구전망(WPPs 2024)」에서 약 200개국의 사망률을 추정하고 있으나, 국가별 통계체계의 차이를 고려하여 선진국의 사망률 전이 패턴을 개발도상국이 모방한다고 가정하는 베이지안 추정기법을 사용하고 있다.

이와 더불어 사망률 개선의 지속성과 인간 수명의 한계 논의 역시 지속되고 있다. 전반적으로 사망률은 꾸준하게 개선되어 왔으나, 이러한 개선이 어디까지 지속될 수 있는지에 대해서는 의견이 분분하다. 일부 학자들은 인간 수명에 자연적 한계가 존재한다고 주장하며, 최고령 인구 증가 추세가 점차 정체된다는 점에 주목한다.

예를 들어, Jan Vijg(2010)와 Dong 등(2016)은 인간 수명의 상한선이 존재하며, 생존곡선의 평탄화(flattening of the longevity curve)가 그 증거라고 지적한다. 이들은 115세에서 125세 사이를 인간 수명의 한계로 제시하고, 특히 100세 이상의 초고령 인구의 증가율이 일정 수준에서 정체되고 있다는 점을 강조하였다. 이러한 현상은 인구통계학적 개념인 수명의 엔트로피(entropy of life table)와 연결되며, 수명에는 '천정(entropy)'가 있다는 견해를 뒷받침한다.

반면, 인간 수명의 한계를 부정하고 150세 수명을 주장하는 학자들도 있다. Oeppen과 Vaupel(2002), Vaupel(2010), Austad(2015), Hughes와 Hekimi(2017) 등은 의료 혁신과 유전체 연구의 발전을 바탕으로 인간 수명은 도약(jump)의 가능성을 지니고 있으며, 수명의 상한은 존재하지 않을 수 있다고 본다.

요컨대, 사망률 개선의 향후 지속성과 인간 수명의 한계에 대한 논의는 단순히 평균수명이 아닌 최고령 인구의 증가 추이와 밀접한 관련이 있으며, 이는 향후 사망률 전망 모형과 정책 대응 방향 설정에 중요한 시사점을 제공한다.

다. 연령대별 사망률 보정 방법론

1) 생명표

x세의 기대여명(Expected years of life remaining 또는 age-specific life expectancy)은 생명표(life table)에서 x세 이상의 각 연령에서 생존한 사람들의 총합을 x세의 생존자 수로 나눈 값이다. 예를 들어, 2020년 출생한 신생아의 기대수명(life expectancy at birth)은 남아 80.8년, 여아 86.5년, 전체 평균 83.5년이다. 또한, 2020년 기준 50세의 기대여명은 남자 32.2년, 여자 37.7년, 전체 34.9년이다.

남자의 경우, 2020년 신생아 기대수명이 80.8년이고, 50세의 기대여명이 32.2년이라면, 50+32.2=82.2년으로 신생아 기대수명(80.8년)보다 높은 수치를 보인다. 이는 시간이 지남에 따라 사망률이 점진적으로 낮아지는 경향이 반영되었기 때문이다. 따라서 일반적으로 기대수명보다 x세의 기대여명이 더 높은 값을 나타낸다.

생명표는 현재의 사망 상태를 기준으로 하여 장래 생존 기간을 추정하는 표이다. 이것은 가상의 10만 명이 출생했다는 가정(출발점이 되는 10만 명은 정의상, 동일한 코호트를 구성, 기수(radix))에서 출발하며, 이 10만 명이 시간의 흐름에 따라 다양한 위험 요인에 의해 사망하게 되는 과정을 반영한다. 예를 들어, 시간의 경과를 연령의 증가로 가정하면 다

음과 같이 사망 패턴을 설명할 수 있다. 특정 연령까지 누적적으로 2만 명이 사망하면, 생존자는 10만 명-2만 명 = 8만 명이 된다. 이후 연령이 증가하면서 절반 수준(5만 명)이 생존하고, 연령이 더욱 증가하면 10만 명의 1/10 수준(1만 명)이 남는 식으로 생존자 수를 추정할 수 있다. 이러한 생존자 수 변화를 곡선으로 연결하여 도식화한 것이 생명표 곡선이다. 또한, 생명표는 확률적 방식으로도 표현할 수 있으며, 이를 사망확률(mortality rate)이라고 한다.

생명표를 작성하기 위해서는 다양한 생명표 함수(life table functions)가 요구된다. 생존자수(Number left alive at age x)는 특정 연령 x세까지 생존한 사람의 수이다. 사망확률(probability of dying between ages x and $x+1$)은 연령 x에서 $x+1$사이에 사망할 확률이다. 생존확률(probability of surviving from ages x and $x+1$)은 연령 x에서 $x+1$까지 생존할 확률이다. 정지인구(person-years lived between ages x and $x+1$)는 연령 x에서 $x+1$까지 생존한 인구가 살아간 총 연수를 의미한다. 그리고 기대여명과 기대수명이 여기에 해당된다. 생명표는 기수(radix)라고 하는 출발점 10만 명($l_0 = 100,000$)에서 시작한다. 생명표를 구성하는 출생 코호트 중에서 특정 연령 x세까지 생존하는 사람의 수를 x세의 생존자수(l_x)라고 한다. 마지막으로, 해당 출생 코호트에서 최후의 한 명이 사망하는 연령을 w라고 한다면 $l_w = 0$이 된다.

생명표에서 l_x가 $x+1$세에 도달하기 전 사망할 확률을 x세의 사망확률이라고 하며, 이를 q_x라고 표기한다. 실제 사망률과 생명표상의 사망확률은 서로 변환할 수 있는 수식이 존재한다. x세의 실제 연령별 사망률(age-specific mortality rate, ASMR)을 M_x라고 정의하면, 연앙 사망률을 의미하는 조사망률(crude mortality rate)의 정의에 따라 $M_x = D_x/N_x$가 된다. 여기서 D_x는 x세 인구의 1년간 사망자수, N_x는

연앙인구(mid-year population)이다. 또한, 기수(radix), 즉 연초 인구를 K_x라고 하면, 연말 인구는 K_{x+1}이 된다. 만약 사망이 연중 균증하게 발생한다고 가정하면, 연앙인구는 $K_x - 0.5D_x = N_x$가 된다. 이를 이용하여 생명표에서 사망확률q_x를 계산하면 $q_x = D_x/K_x = D_x/(N_x + 0.5D_x)$이므로, 결국 q_x는 식 (44)과 같다.

$$q_x = (D_x/N_x)/(1 + 0.5D_x/N_x) \tag{44}$$

식 (44)에서 실제 x세의 연령사망률을 M_x을 q_x와 연결 지으면 식 (45)과 같다.

$$q_x = (D_x/N_x)/(1 + 0.5D_x/N_x) = \frac{M_x}{1 + \frac{1}{2}M_x} \tag{45}$$

식 (45)는 생명표의 사망확률이 실제 사망률만으로 표현될 수 있음을 보여준다. 이러한 관계를 바탕으로 사망률과 사망확률을 혼용하여 사용하는 경우도 있다. 따라서 실제 성별·연령별 사망자수 및 성별·연령별 인구가 인구동태통계(Vital Statistics)와 인구주택총조사(Census)에서 집계된다면, 생명표에서 해당 연령(x)의 성별·연령별 사망확률을 계산할 수 있다. 생존확률은 l_x가 $x+1$세에 도달할 확률을 의미하며, x세의 생존확률이라 하며, p_x로 표기한다. 일반적으로 p_x는 $p_x = l_{x+1}/l_x = 1 - q_x$이다. 정지인구는 l_x가 $x+1$세에 이르기까지의 살아간 총인년수(person-years)를 의미하고, L_x로 표기한다. 이는 식 (46)와 같이 계산된다.

$$L_x = \int_x^{x+1} l_t dt \tag{46}$$

연령 x에서 $x+1$세까지의 사망은 다양한 형태로 발생할 수 있으나, 계산의 편의를 위해 사망이 균등하게 발생한다고 가정한다. 이런 경우, 7월 1일 기준으로 절반은 생존하고, 나머지 절반은 사망한다고 볼 수 있다. 따라서 사망자수와 생존자수는 $0.5 \times (l_x - l_{x+1}) = 0.5 d_x$가 된다. 여기서 d_x는 생명표에서 x세의 사망자수이다. 따라서 L_x는 식 (47)과 같이 계산될 수 있다.

$$L_x = 0.5 \times d_x + l_{x+1} = 0.5 \times (l_x + l_{x+1}) \tag{47}$$

기대여명은 x세에서 $x+1$세에 이르기까지의 생존 총인년수 총합을 평균화한 값이다. 따라서 기대여명은 $T_x = \sum_{i=x}^{\infty} L_i$이다. 여기서 T_x는 x세 이후의 총 생존인년수(person-years lived above age x)이다. 따라서 T_x를 l_x로 나누면 기대여명($e_x = T_x / l_x$)이 된다. 출생 시 기대여명(기대수명 e_0^o)는 $x=0$일 때의 평균 여명을 의미하며, 이를 기대수명(life expectancy at birth)이라 한다. 이상의 내용을 정리하면, 생명표는 식 (48)과 같은 순서로 계산된다.

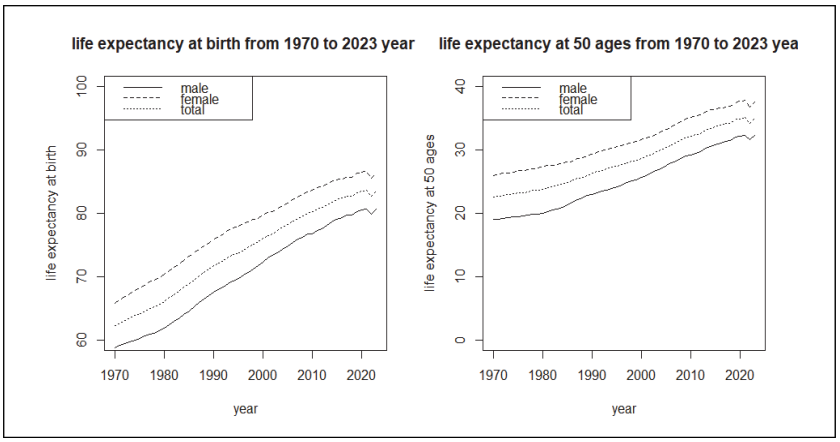
$$l_0 \rightarrow d_0 \rightarrow l_1 \rightarrow d_1 \rightarrow l_2 \rightarrow L_2 \rightarrow \ldots \rightarrow L_w \rightarrow T_x \rightarrow e_x^o \tag{48}$$

즉, 생존자수(l_x)로부터 시작하여, 사망자수(d_x), 사망확률(q_x), 생존확률(p_x), 정지인구(L_x), 총 생존인년수(T_x), 최종적으로 기대여명(e_x)을 계산하는 과정으로 생명표가 완성된다.

〔그림 4-3〕은 1970년부터 2023년까지의 사망률을 생명표에 적용하여

전체 및 성별에 따른 0세 기대여명과 50세 기대여명을 산출한 결과이다.

[그림 4-3] 성별 기대수명과 50세 기대여명 추이



자료: 통계청, KOSIS, 생명표.

기대여명은 여성 〉 전체 평균 〉 남성 순으로 나타났으며, 이는 전쟁이나 극심한 기아와 같은 예외적인 상황을 제외하면 대부분의 국가에서 공통적으로 관찰되는 경향이다. 이러한 차이는 여성의 사망률이 남성보다 낮아 기대수명이 상대적으로 높게 나타나기 때문이다. 1970년대 남성의 출생 시 기대수명은 60세 미만이었으나, 2023년에는 약 83세에 도달하여 50여 년간 약 23세가 증가하였다. 여성의 경우, 1970년대 67세에서 2023년 약 86세로 증가하여 약 21세 상승하였다. 또한, 50세 기준 기대여명을 살펴보면, 남성은 19세에서 32.3세로 증가하였고, 여성은 26세에서 37.6세로 증가하였다. 이처럼 단기간 내 기대수명이 급격히 증가한 사례는 매우 드물다. 이러한 결과는 경제의 급속한 성장, 공중보건의 개선, 보편적 의료보험 혜택 확대 등으로 인해 모든 연령대에서 사망률이 획기적으로 감소한 영향으로 해석할 수 있다.

2) 저연령층 사망확률 보정

사망률과 사망확률의 관계식은 식 (49)과 같다.

$$ {}_nq_x = \frac{n\,{}_nm_x}{1+(n-{}_na_x)\,{}_nm_x} \tag{49} $$

여기서 ${}_nq_x$와 ${}_nm_x$는 각각 n세 간격 x연령의 사망확률과 사망률이다. 만약 1세 간격의 사망률 자료라면, $n=1$이 된다. 그리고 ${}_na_x$는 해당 기간 동안 사망한 사람들이 해당 기간 동안 생존한 평균 인년수(average number of person-years)이다. 만약 사망이 시간 경과에 따라 직선적으로 발생한다고 가정하거나, 그 기간에 사망한 사람이 평균적으로 그 기간의 중간 지점에서 사망한다면 ${}_na_x = 1/2 = 0.5$가 된다.

생명표 작성에서 사망률을 사망확률로 전환하여 보정한다. 통계청(2016)[18]은 신, 구 생명표 비교과 개선사항을 소개한다. 주요 내용을 소개하면, 사망률은 $q'_x = m_x/(1+0.5m_x)$을 적용하여 사망확률로 전환한다. 이때 q'_x는 Greville 9차항 계수를 이용하여 1~99세까지 연령별 사망확률을 보정한다. 그리고 최종상한연령은 115세로 정하고, 85~115세까지 고령층 사망확률은 Coale-Kisker모형을 적용하여 보정한다. 그리고 '01~'14년 0세 사망자수를 사용하여 0세 사망확률를 재작성하여 보정한다(〈표 4-2〉 참조). 그리고 고령층 사망확률 연장 방법에 대해서는 다음 절에 상세히 소개한다.

18) 통계청(2016), 1970~2014년 생명표.

〈표 4-2〉 신, 구 생명표 개선사항

	구 생명표	신 생명표
사망자수	◦ '70-'04년 : 3년 평균 사망자 수 ◦ '05-'14년 : 단년 사망자 수(16개월 방식)	◦ '70-'00년 : 3년 평균 사망자수 ◦ '01-'14년 : 단년 사망자수(16개월 방식)
연앙인구	◦ '70-'92년 : 추계인구, ◦ '93-'14년 : 주민등록인구	
기초자료 보정	◦ '70-'04년 : 연령왜곡신고 보정 ◦ '05-'14년 : 주민등록연령 보정	◦ '70-'00년 : 연령왜곡신고 보정 ◦ 주민등록보정계수 적용절차 삭제
사망률→ 사망확률 전환	◦ '70-'00년 : $_nq_x = \frac{_nm_x}{[\frac{1}{n} + {_nm_x}[\frac{1}{2} + \frac{n}{12}({_nm_x} - 0.095)]]}$ ◦ '01-'14년 : $q_x' = \frac{m_x}{1+\frac{1}{2}m_x}$	◦ '70-'14년 : 전환방법 통일 $q_x' = \frac{m_x}{1+\frac{1}{2}m_x}$
사망확률 보정	◦ '01-'14년 : 집계된 0세 사망자 수를 사용하여 0세 사망확률 재작성 ◦ '01-'14년 : Greville 9차항 계수를 이용, 연령별 사망확률 보정(1-99세)	
고령층 사망확률 보정	◦ '99, '01-'14년 : Coale-Kisker모형(85-∞세)	◦ '70-'00년 : Brass-logit 모형(75-85세) ◦ '70-'00년 : Coale-Kiske모형(85-110세) ◦ '01-'14년 : Coale-Kiske모형(85-115세)
최종상한 연령 정지인구	◦ '70-'92년 : $_\infty L_{90} = 3.725 \times (l_{90}) + 0.0000625 \times (l_{90})^2$ ◦ '93-'98년 : $_\infty L_{95} = l_{95} \times \log(l_{95})$ ◦ '99-'00년 : $_\infty L_{95} = \sum L_x \ (x = 95, \cdots, \infty)$ ◦ '01-'14년 : $_\infty L_{100} = \sum L_x \ (x = 100, \cdots, \infty)$	◦ '70-'99년 : 한계연령 110세로 고정 $_{110}L_{100} = \sum L_x \ (x = 100, \cdots, 110)$ ◦ '01-'14년 : 한계연령 115세로 고정 $_{115}L_{100} = \sum L_x \ (x = 100, \cdots, 115)$

* '00년 생명표는 '99년 사망확률과 '01년 사망확률을 기초로 기하평균보간법으로 작성

자료: 통계청(2016), 2015년 생명표 작성결과 및 1970~2014년 생명표(新) 작성결과.

기대수명은 연령별 사망률 추정 절차 중 특히 저연령층과 고연령층, 즉 사망률이 상대적으로 높은 연령대의 처리 방식에 가장 민감하게 반응한다. 이로 인해 기대수명 추정치는 이러한 연령대에서 적용되는 방법론이나 보정 절차에 따라 큰 변동을 보일 수 있다.

식 (43)에서는 사망률을 사망확률로 변환할 때, 변환 과정에 사용되는 보정 계수인 $_na_x$의 값이 사망률 수준에 따라 달라지는 경험적 특성이 있다. 일반적으로 사망률이 낮은 경우에는 영아 사망이 생후 초기 단계에 집중되는 경향이 있으며, 이때 출생 이전, 즉 임신기 환경의 영향력이 출생 후 환경보다 더 크게 작용하는 것으로 나타난다. 따라서 0세에서 5세 미만 연령대의 사망위험이 연령에 따라 어떻게 변화하는지를 분석하는 것은 보건정책 수립과 평가에 매우 중요한 기초 자료가 된다.

Coale와 Demeny(1983)는 다양한 인구 집단의 자료를 바탕으로 이와 같은 관계를 검토하고, 5세 미만 연령층에 적용할 수 있는 $_na_x$값의 대표적인 추정치를 제안하였다. 〈표 4-3〉은 이들이 제시한 5세 미만 연령

층에 대한 ${}_{n}a_{x}$값을 요약한 것이다.

〈표 4-3〉 5세 이하의 ${}_{n}a_{x}$ 값

구분	남아	여아
Value of ${}_{1}a_{0}$		
만일 ${}_{1}m_{0} \geqq 0.107$	0.330	0.350
만일 ${}_{1}m_{0} < 0.107$	$0.045 + 2.684 \times {}_{1}m_{0}$	$0.053 + 2.800 \times {}_{1}m_{0}$
Value of ${}_{4}a_{0}$		
만일 ${}_{1}m_{0} \geqq 0.107$	1.352	1.361
만일 ${}_{1}m_{0} < 0.107$	$1.651 - 2.816 \times {}_{1}m_{0}$	$1.522 - 1.518 \times {}_{1}m_{0}$

자료: Coale과 Demeny(1983) West model life tables.

최근 UN-IGME(Inter-agency Group for Child Mortality Estimation)는 「Levels & Trends in Child Mortality Report 2023」을 통해 5세 미만 저연령층의 사망률 보정 방법을 제시하였다. 〔그림 4-4〕는 해당 보고서에서 제시한 한국의 5세 미만 사망률 추이를 보여준다. 1990년에는 1,000명당 남아 17명, 여아 14명이 영아 사망으로 기록되었으나, 2022년에는 남녀 모두 1,000명당 3명 이하로 감소하여 0.3% 수준에 도달하였다.

〔그림 4-4〕 UN-IGME의 5세 미만 사망력

STATISTICAL TABLE (CONTINUED)

Country, regional and global estimates of mortality among children under age 5

Country	Under-five mortality rate (USMR) with 90 per cent uncertainty interval (deaths per 1,000 live births)			Annual rate of reduction (ARR) (per cent)	Number of under-five deaths with 90 per cent uncertainty interval (thousands)*			Sex-specific under-five mortality rate (deaths per 1,000 live births)			
								1990		2022	
	1990	2000	2022	1990–2022	1990	2000	2022	Male	Female	Male	Female
Republic of Korea	16 (15 - 17)	8 (7 - 8)	3 (3 - 3)	5.4 (5.1 - 5.7)	11 (10 - 11)	5 (4 - 5)	1 (1 - 1)	17 (16 - 18)	14 (13 - 15)	3 (3 - 3)	3 (2 - 3)

Country	Infant mortality rate (deaths per 1,000 live births)		Number of infant deaths (thousands)*		Neonatal mortality rate (deaths per 1,000 live births)			Annual rate of reduction (ARR) (per cent)	Number of neonatal deaths (thousands)*		
	1990	2022	1990	2022	1990	2000	2022	1990–2022	1990	2000	2022
Republic of Korea	13 (13 - 14)	2 (2 - 3)	9 (9 - 10)	1 (1 - 1)	7 (6 - 8)	3 (3 - 4)	1 (1 - 1)	5.3 (4.7 - 6)	5 (4 - 6)	2 (2 - 2)	0 (0 - 0)

자료: UN-IGME(2023).

KOSIS의 2022년 자료에 따르면 0세 남아의 사망률은 0.00251, 1~4세는 0.00071이며, 여아의 경우 각각 0.00198, 0.00064로 나타나고 있다. 5세 미만 사망률 패턴을 추정하기 위해 전통적으로 Coale과 Demeny의 모델 생명표와 유엔 모델 생명표가 활용되어 왔으나, 이들 모델은 다음과 같은 몇 가지 한계점을 갖는다.

첫째, 이들 모델은 5세 미만 연령을 0세와 1~4세의 두 연령구간으로만 구분하고 있어, 신생아 사망률 추정이나 보다 세부적인 연령대 분석에는 적합하지 않다. 특히 생후 12개월에 초점을 두고 있어 다양한 국가나 맥락에 따라 실제 자료와 부합하지 않는다는 비판이 제기되어 왔다(Galley와 Woods, 1998; Knodel과 Kintner, 1977; Lantoine와 Pressat, 1984; Lynch 등, 1998; Manfredini, 2004).

둘째, Coale과 Demeny(1983) 및 유엔 모델 생명표(1992)는 1980년대 초반의 자료에 기반하고 있어, 현재의 사망률 추세를 반영하기에는 시대적 한계가 있다.

셋째, 이들 모델은 지역별 그룹화(regional groupings)를 기반으로 연령 패턴을 '가족(family)' 단위로 요약하는 방식으로 구성되어 있어, 사망률 연령 패턴을 연속적인 형태로 설명하기보다는 불연속적 특징을 가진다.

이러한 한계를 보완하기 위해 최근에는 보다 정교한 수학적 모델이 제안되고 있다. 예를 들어, Murray 등(2003)은 수정된 로짓 시스템(modified logit system)을, Wilmoth 등(2012)은 로그-이차(log-quadratic) 모델을, Clark(2019)은 특이값 분해(SVD)를 활용한 구성요소 모델(component model)을 제안하였다. 이들 모델은 연속형 매개변수를 사용하고, 최신 데이터를 기반으로 계수를 추정함으로써 기존 모델 생명표의 제한점을 보완한다.

그러나 Murray 및 Wilmoth의 모델은 여전히 0세와 1~4세의 두 구간에 의존하고 있으며, Clark의 SVD 모델 또한 단일 연도 연령 구간 이하의 세부 사망률을 제시하지 않는다는 제한이 있다.

이에 따라 Guillot 외(2022)는 보다 세분화된 연령 정보를 반영하고, 5세 미만 사망률(Under-5 Mortality Database, U5MD)의 연령별 패턴을 보다 정밀하게 설명하기 위한 새로운 logQuadratic 모형을 제안하였다. 이 모델은 Under-5 Mortality Database(U5MD)에 수록된 세부 연령별 자료를 바탕으로 개발되었으며, 특히 생후 첫해(0세)의 연령 구간을 세분화하여 기존 모델보다 유연하고 정교한 접근이 가능하다.

logQuadratic 모형은 5세 미만 전체 사망률 수준을 나타내는 두 개의 매개변수와, 연령별 사망률 패턴의 형태를 설명하는 두 개의 매개변수를 이용해, 0~5세 미만 연령대의 세부 사망률 일정을 예측할 수 있다. 이로써 해당 모형은 기존 모델 생명표의 구조적 한계를 효과적으로 보완하면서도, 실제 정책 수립에 활용 가능한 고해상도의 사망률 예측을 가능하게 한다.

logQuadratic 모형(식 (50) 참조)은 Wilmoth 등(2012)의 log-quadratic model(식 (51) 참조)에서 변형되었다고 볼 수 있다. 이는 5세 미만 범위 내의 각 세부 연령 x에 대해서 출생부터 연령 x, $q(x)$까지 누적 사망 확률과 5세 미만 사망률 $q(5y)$ 사이의 log-quadratic 관계 관찰을 기반으로 한다.

$$\log[q(x)] = a_x + b_x \ln(q(5y)) + c_x \ln[q(5y)]^2 + v_x k, \;\; k=0, \;\; k \neq 0 \qquad (50)$$

$$\log(m_x) = a_x + b_x h + c_x h^2 + v_x k, \quad h = {}_5q_0, \;\; k = (-2, 2) \qquad (51)$$

식 (50)에서 $k=0$일 때, 모델은 최종 U5MD(under-5 Mortality Database)에 포함된 국가별 연도 집합의 평균 사망률 스케줄인 일반적

인 패턴을 예측한다. $k \neq 0$일 때, 모델은 $q(x)$가 연령에 따른 비감소 함수(a nondecreasing function of age)이며, 주어진 $q(5y)$수준에서 $q(x)$의 연령 패턴의 특이성에 따라 사망할 확률을 조정한다. 주어진 $q(5y)$ 수준에서 k의 값에 따라 사망률의 연령 패턴은 신생아(neonatal)와 유아(infant) 사망률이 상대적으로 높은 '초기' 또는 상대적으로 낮은 '후기'가 된다.

식 (50)은 2단계 추정법에 의해 도출된다. 먼저 $q(x)$와 $q(5y)$에 대한 회귀방정식의 최소제곱법에 의해서 a_x, b_x, c_x을 추정한다. 다음으로 v_x를 추정하기 위해 SVD 방법을 적용하여 추정한다. 이들 방법에 대한 자세한 설명은 Guillot, Prieto와 Gerland(2022, p.8)를 참고하면 되며, 이들 추정 과정을 R코드로 소개하고 있다. 이들은 식 (51)을 식 (50)으로 개선할 경우 네 가지 이점이 있다고 밝히고 있다.

첫째 예측된 $q(x)$집합과 그에 상응하는 ${}_nM_x$ 값은 항상 식 (51)의 우측에서 예측 변수로 선택된 $q(5y)$수준과 일치한다. 둘째 이 모델은 사망률을 사용할 때 21개의 계수 세트와 22개의 계수 세트로 더 간결하다. 셋째 모델은 사망 시 잘못 보고된 연령으로 인해 발생할 수 있는 사망 일정의 변동에 덜 민감하다. 넷째 초기 신생아, 신생아 및 유아 사망률과 같은 고전적인 사망률 지표를 직접 예측할 수 있으며, 이는 실제로 사망할 확률이 각각 $q(7d)$, $q(28d)$, $q(12m)$이다.

그러나 이 모델에서 누적 사망 확률을 사용하는 데 한 가지 단점이 있다.

신생아 사망에 대한 과소 보고와 같은 초기 연령의 데이터 오류는 전체 $q(x)$곡선을 통해 반영된다. 따라서 이는 최종 U5MD에 국가별 연도를 포함하는 것과 관련하여 식 (44)은 다소 보수적인 접근 방식이라고 볼 수 있다.

이들이 제안한 모델은 연령별 사망률이 아닌 사망 누적 확률을 예측하

지만, 해당 사망률은 22개의 작은 연령 간격 내에서 일정한 사망률의 가정을 사용하여 예측된 $q(x)$값에서 식 (52)를 이용해 계산할 수 있다.

$$ {}_nM_x = -\frac{\ln\left[\frac{1-q(x+n)}{1-q(x)}\right]}{n} \tag{52} $$

앞서 소개한 선행연구들과 마찬가지로, 5세 미만 사망률을 산정하고 보정하는 데에는 다양한 통계적 접근이 활용되어 왔다. 대표적인 방법으로는 수정된 로짓 시스템(modified logit system), 로그-이차 모형(log-quadratic model), 특이값 분해(SVD)를 활용한 구성요소 모델(component model), 그리고 확장된 log-quadratic 모형 등이 있다. 이들 방법은 각각의 이론적 기반과 데이터 특성에 따라 사망률의 연령별 패턴을 보다 정확히 설명하기 위해 개발되었다.

또한 일부 연구에서는 개별 연령의 사망률을 직접 사용하는 대신, 누적 사망률(cumulative mortality) 또는 이의 변환 값을 활용하여 모형의 안정성과 예측력을 향상시키고자 하였다.

이와 같은 선행연구들을 통해 다음과 같은 주요 시사점을 도출할 수 있다.

첫째, 0에 가까운 극미한 수준의 사망률이나 출산율의 변화를 분석할 때는 누적값 또는 로그 변환(log transformation)이 자주 활용된다. 이러한 방식은 지표의 미세한 변동성을 포착하고, 분석 결과의 해석 가능성을 높이기 위한 통계적 기법으로 널리 사용된다.

둘째, 5세 미만 사망률 추정에는 logit, log-quadratic 등 다양한 연결함수(link function)가 활용된다. 이는 사망확률이나 사망사건을 선형모형의 틀 안에서 표현하기 위한 것으로, 비선형적인 특성을 갖는 사망률 데이터를 보다 효과적으로 설명할 수 있는 방법이다.

셋째, 5세 미만 사망률은 생애 초기 단계에서 급격한 감소 양상을 보이

는 특성이 있어 단순 선형모형으로는 설명하기 어렵다. 이에 따라 이차 곡선(quadratic curve) 형태의 모형이 자주 적용되며, 보다 정확한 예측과 해석을 위해 다양한 형태의 함수 변환과 선형화 기법이 필수적으로 동반된다.

이러한 시사점은 향후 5세 미만 사망률 추정의 정밀도 향상뿐 아니라, 연령별 보건정책 수립의 기초자료로서의 활용 가능성을 제고하는 데 기여할 수 있다.

3) 고연령층 사망률 연장

일반적으로 사망률은 5세 이하 연령대와 85세 이상 고령층에서 변동성이 크다. 사망률은 '해당 연령대 사망자수/해당 연령대 주민등록 인구'로 계산되지만, 저연령층과 고연령층은 분모(해당 연령대 인구)가 다른 연령대에 비해 작아 상대적으로 큰 변동성을 보일 수밖에 없다. 여러 선행 연구에 따르면 80세 또는 85세를 초과하면 사망률 개선 추세가 둔화하거나 감소하는 경향을 보인다(Coale과 Guo, 1989; Horiuchi와 Coale, 1982; Coale과 Kisker, 1990; UN, 2022). 또한, 성별과 관계없이 고령층 사망률을 연령대별로 확장하면 초고령 연령대에서 비현실적인 교차(crossover)가 발생할 수 있다(UN, 2022).

이러한 문제점을 지적한 선행 연구들은 기존 사망률 모형을 그대로 적용할 경우 고령층의 사망률 패턴을 제대로 반영하지 못할 가능성이 있음을 지적하며, 고령층 사망률 개선에 적합한 새로운 모형을 적용할 필요성을 제기하였다. 이에 따라 고령층 사망률의 추세를 연장하는 방법으로 외삽방법(old ages mortality extrapolation methods)이 제시되었으며, Gompertz-Makeham 모형 (GPZ), Weibull 모형 (WBL), Helligman-

Pollard 모형 (HP), Coale-Guo 및 Coale-Kisker 방법 (CK), Himes-Preston-Condran 방법 (HPC), Perks 모형 (PRK; 1932), Thatcher 모형 (THT; 1999), Kannisto 모형 (KST; 1992), Denuit와 Goderniaux 방법 (DG, 2007) 등 다양한 모형이 활용되어 왔다. 이들 모형을 설명하기에 앞서 식 (53)과 같은 개념정의가 필요하다.

x: 사망연령; $f(x)$: x의 확률밀도함수; $F(x)$: x의 누적분포함수
$S(x) = 1 - F(x)$: $F(x)$의 생존함수 (53)
$h(x) = f(x)/[1 - F(x)]$: 사망력 또는 위험함수(hazard function)
E_x: 연령 $(x, x+1)$사이의 이벤트 노출 인구수(exposed population)

Gompertz-Makeham 모형은 예전부터 고전적 사망률 모델로 사망률을 노년층으로 확장하는 데 사용되었다. Gompetrz(1825)는 사망력(the force of mortality)이 연령에 따른 지수 함수로 표현될 수 ($h(x) = B\exp(Cx)$) 있음을 발견했다. 여기서 B(scale parameter)는 사망력 수준이고, C (positive shape parameter)는 연령별 사망력 증가율이다. Makeham(1860)은 이모형을 개선하여, Gompertz-Makeham 모형(GPZ)인 $h(x) = A + B\exp(Cx)$을 제안하였다. 여기서 A는 교통사고(사고), 질병과 같은 사망원인으로부터의 사망력을 의미한다. Gompertz 모형은 $A = 0$일 때, Makeham 모형의 특수한 경우이지만, Makeham모형은 Gompertz모형에 비해 거의 더 나은 적합도를 가지고 있다. 두 모형 모두 사망률을 80세 이상으로 확장하는 데 사용되었으며, Makeham 유형의 함수는 유엔 모델 생명표를 80세 이상으로 확장하는 데 사용되었다(UN, 1981). 사용된 함수는 $q_x/(1-q_x) = A + B^x$이다.

Weibull(1951)은 Weibull's 모형(WBL)에서 $h(x) = Bx^C$를 제안하였다.

Logistic 모형의 일반적인 형태는 $h(x)=\theta+\frac{\lambda B\exp(Cx)}{1+B\exp(Cx)}$이다. Thatcher 등(1998)이 13개 선진국의 고령 사망률을 측정하는 데 사용된 이 모형을 사용하려고 했을 때, λ가 1에 매우 가깝다는 것을 발견했다(Thatcher, 1999). Thatcher 등(1998)은 이 모형이 3개의 독립 모수로 간결하게 표현할 수 있음을 증명하고 식 (54)를 제안하였다.

$$h(x)=\theta+\frac{B\exp(Cx)}{1+B\exp(Cx)} \tag{54}$$

그리고 생명표에서 $h(x)$값은 Perks(1932)가 제안한 로지스틱 곡선으로 적합할 수 있으며, Perks의 모형에서 위험 함수는 $h(x)=A+\frac{B\exp(Cx)}{1+D\exp(Cx)}$이다. Beard(1971)과 Kannisto(1992) 모형과 같은 다른 고연령 사망률 모형은 로지스틱 모형의 특수한 경우이며, $h(x)$에서 $A=0$를 설정하면 3개 매개변수 Beard 모형인 $h(x)=\frac{B\exp(Cx)}{1+D\exp(Cx)}$이 된다. Kannisto(1992)는 고연령대에서 $h(x)$에 대한 현대 데이터에 따른 $Logit(h(x))$이 x의 선형 함수인 로지스틱 모형의 가장 간단한 형태이면서 적합력이 높다는 것을 찾았으며, 이런 Kannisto 모형의 $h(x)$은 $h(x)=\frac{B\exp(Cx)}{1+B\exp(Cx)}$, Logit 형태는 $Logit(h(x))=\ln(B)+Cx$이다.

Heligman-Pollard(1980)은 Heligman-Pollard 모형에서 $q_x=d_x/l_x$를 식 (55)과 같이 모델링하였다.

$$\frac{q_x}{1-q_x}=A^{(x+B)^C}+D\exp[-E(\ln x-\ln F)^2]+GH^x \tag{55}$$

여기서 A, B, C, D, E, F, G, H는 상수, $A^{(x+B)^C}$, $D\exp[-E(\ln x - \ln F)^2]$, GH^x는 각각 유아, 사고, 그리고 노령 사망률이다. 특히 GH^x는 Gompertz 사망률의 법칙을 포착하는 용어로 볼 수도 있다.

Heligman-Pollard(1980)는 1946~1948, 1960~1962, 1970~1972년의 이들 세 기간 동안 85세까지의 호주 사망률에 대한 모형 적합도를 도출하였다.

Coale-Kisker 방법(Coale과 Kisker, 1990, CK)은 선진국의 사망 확률을 모델링하는 것으로 잘 알려져 있다(Boleslawski과 Tabeau, 2001). 이 방법은 식 (56)과 같은 재귀적(recursive) 형태를 보인다.

$$k(x) = k(x-1) - R, \quad x \geqq x_0 \tag{56}$$

식 (56)의 $k(x) = \ln(m_x / m_{x+1})$이고, R은 결정 상수, 그리고 외삽은 정수 연령 x_0에서 시작한다. m_x는 x세의 사망률이고, $x = x_1$까지 공식을 적용하면, R은 $\dfrac{(x_1 - x_0)k(x_0) + \ln m_{x_0} - \ln m_{x_1}}{1 + 2 + \cdots + (x_1 - x_0)}$이다. 그런데 이 방법은 생명표의 상한 연령 x_1과 상한연령의 m_x값을 가정해야 한다. 또한 외삽을 시작할 연령도 주관적으로 결정해야한다. Coale과 Kisker(1990)는 $x_0 = 84$, $x_1 = 110$, $m_{x_1} = 1$을 가정한다. 참고로 CK 모형은 통계청(2024)이 85세 이상의 연령 자료를 연장하거나 보정할 때 사용한 모형이다.

Coherent Kannisto 모형은 남성과 여성의 사망률을 동반(coherent) 처리하는 Kannisto 방법(1992)의 수정모형이다. Kannisto 모형은 $m_x = \dfrac{ce^{d_x}}{1 + ce^{d_x}} e^{\epsilon_x}$ 또는 $logit(m_x) = \log c + dx + \epsilon_x$이다. 여기서 ϵ_x는 평균이 0인 랜덤 오차이다. 이 모형은 각 성별에 대해 독립적으로 추정되며, 최대우도법(Thatcher 등, 1998)을 사용하고, 연령과 ϵ_x의 정상성에 대한

독립성을 가정한다. 이 결과 성별 d, c모수($d_{M,}d_{F,}c_M, c_F$, $M(male)$, $F(Female)$)가 추정된다. Sevicikova 등(2016)이 제안한 Coherent Kannisto 모형은 성별 매개변수 d_M과 d_F를 동일하게 간주하고($d_M = d_F = d$), 매개변수 c_M과 c_F는 성별 간에 서로 다르게 적용하는 방식이다. 결과적으로 식 (57)와 같은 식이 도출되고,

$$logit(m_x^g) = \log c_g + d_x + \epsilon_x^g, \;\; g = M, F \tag{57}$$

가변수와 변수명을 통일하여 식 (58)과 같은 모델로 표현 가능하다.

$$\begin{aligned} & logit(m_x^g) = \beta_0 + \beta_1 I_{(g=M)} + \beta_2 x + \epsilon_x^g \\ & \begin{cases} I_{(g=M)} = 1, \;\; g = M \\ I_{(g=M)} = 0, \;\; o.w \end{cases} \end{aligned} \tag{58}$$

Sevicikova 등(2016)은 β 모수를 추정하기 위해 80~99세의 관찰된 m_x에 모형을 적합하고, ϵ_x^g의 독립성과 정규성을 가정했을 때, 최대 우도에 해당하는 일반최소제곱회귀(GLS)를 통해 모형에 적용했다고 밝혔다. 그런 다음, $\hat{c}_F = e^{\hat{\beta}_0}, \hat{c}_M = e^{\hat{\beta}_0 + \hat{\beta}_1}, \hat{d} = \hat{\beta}_2$을 추정한다. UN(2019, 2022, 2024)의 WPP 2019, 2022, 2024에서 고연령 상한연령 연장은 이 방법을 적용하여 130세까지 사망확률을 도출한다.

Himes, Preston과 Condran(1994)은 45-99세 연령대에서 단일 연령대 사망률 관찰을 기반으로 외삽법을 제안했다. 이들이 제안한 아이디어는 관찰된 사망률 곡선을 식 (59)의 2단계 중 첫 번째 단계인 3차 이동평균으로 매끄럽게 만드는 것이다.

$$g_x = \frac{h(x-1) + h(x) + h(x+1)}{3} \tag{59}$$

그런 다음, 평활 사망률 g_x는 식 (60)를 사용하여 2단계의 평활화과정을 거친다.

$$\mu_x = \frac{g_{x-1} + g_x + g_{x+1}}{3} \tag{60}$$

이는 평활화 과정을 거친 사망률은 다음 식 (61)으로 이어진다.

$$\mu_x = \frac{h(x-2) + 2h(x-1) + 3h(x) + 2h(x+1) + h(x+2)}{3} \tag{61}$$

고려된 연령대에서 사망률은 지수 함수를 따르며, 이는 Gompetz모형 또는 Helligman과 Pollard 모형의 세 번째 항으로 근사될 수 있다. 그리고 Himes, Preston과 Cardan(1994)은 식 (62)을 사용하여 사망률을 115세까지 연장할 것을 제안했다.

$$Logit(h(x)) = \ln\left(\frac{h(x)}{1-h(x)}\right) = \theta + \Pi x \tag{62}$$

여기서 모수 θ와 Π는 45~79세에서 추정한 값이다.

Denuit et Goderniaux 모형(2005)은 식 (63)의 로그 사망률의 3차 다항식 공식에 의존한다.

$$\ln(q_x) = a + bx + cx^2 \tag{63}$$

논문에서 저자들은 궁극적인 생존 연령을 130세로 가정하였으며, 이러한 제약 조건을 적용하기 위해 $q_{130} = 1$과 $p_{130} = 0$을 부과했다고 밝히고 있다.

제2절 사망력 추계 개선 방안

1. 기존 사망력 추계 방법론의 주요 쟁점

가. 사망력 모형 관리와 고령화 대응

앞서 살펴본 바와 같이, 현재 통계청을 비롯한 주요국 통계기관들은 사망률 추이를 정확히 반영하기 위해 LC, LL, LC-ER, LLG 등 다양한 모형을 구축하여 활용하고 있으며, 본 연구원 역시 해당 모형을 구축하고 관리 중에 있다. 이러한 다중 모형 체계는 각각의 장점을 상황에 맞게 활용할 수 있다는 이점이 있지만, 동시에 모형 관리와 운영 측면에서 중요한 쟁점을 제기한다.

첫째, 모형별 특성에 따른 차별적 관리 필요성의 쟁점이다. 우리나라의 급속한 고령화 진행 속도를 고려할 때, 고령층 사망률 모형의 정확성이 전체 추계 결과에 미치는 영향이 점차 커지고 있다. 65세 이상 인구 비율이 2025년 20.6%에서 2070년 46.4%로 증가할 것으로 전망되는 상황에서, 고령층 사망률의 작은 변화도 전체 기대수명과 생존 인구에 큰 영향을 미친다.

이는 기존의 LC나 LL 모형보다는 고령층 사망률 개선을 보다 정교하게 반영하는 LC-ER이나 LLG 모형의 구현과 유지보수에 더 많은 관심과 자원을 투입해야 함을 의미한다. 특히 85세 이상 초고령층의 경우 사망률 패턴이 중년층과는 전혀 다른 양상을 보이므로, 이들 연령대에 특화된 모형의 정밀도 관리가 핵심적 과제가 된다.

둘째, 모형 교체와 결과 산출의 일관성 확보 쟁점이다. 인구구조 변화나 사망률 패턴의 구조적 변화에 따라 적절한 모형을 선택하고 교체해야

하는 상황이 발생할 수 있는데, 이 과정에서 추계 결과의 연속성과 일관성을 어떻게 유지할 것인가의 문제이다. 특히 모형 전환 시점에서 발생할 수 있는 불연속성을 최소화하고, 결과 해석의 일관성을 확보하는 것이 중요한 기술적 과제이다.

셋째, 급속한 인구변동에 대한 모형 적응성 관리의 쟁점이다. 코로나19와 같은 외부 충격이나 의료기술의 급속한 발전은 기존 모형들이 전제하고 있는 사망률 개선 패턴을 근본적으로 변화시킬 수 있다. 이러한 변화를 감지하고 적절한 모형 조정이나 교체를 수행할 수 있는 모니터링 체계의 구축이 필요하다.

나. 고연령층 데이터 품질과 추계 변동성

사망력 추계의 정확성은 기초 데이터의 품질에 직접적으로 의존하는데, 특히 고연령층 데이터의 품질 문제는 전체 추계 결과에 상당한 영향을 미치는 중요한 쟁점이다.

고연령층 사망 데이터의 신뢰성 확보가 가장 핵심적인 쟁점이다. 85세 이상, 특히 100세 이상 초고령층의 경우 연령 오보고(age misreporting) 문제가 심각하며, 이는 해당 연령대의 실제 사망률 추정에 상당한 오차를 발생시킨다. 또한 고연령층 인구 자체가 절대적으로 적어 통계적 변동성이 크므로, 연도별 사망률의 불규칙한 변동이 추계 모형의 안정성을 해칠 수 있다.

이러한 문제는 우리나라의 급속한 고령화와 맞물려 더욱 심각해진다. 85세 이상 인구가 2025년 약 110만 명에서 2070년 약 530만 명으로 급증할 것으로 예상되는 상황에서, 이들 연령대의 사망률 패턴을 정확히 파악하지 못하면 전체 인구추계의 오차가 기하급수적으로 증가할 수 있다.

또한 고연령층의 사망원인 구조와 건강 상태의 복잡성도 중요한 쟁점이다. 85세 이상에서는 단일 사망원인보다는 복합적 요인에 의한 사망이 일반적이며, 이는 전통적인 사망률 모형으로는 충분히 설명하기 어려운 특성이다. 특히 치매, 복합 만성질환 등 고령층 특유의 건강 문제들이 사망률에 미치는 영향을 어떻게 모형화할 것인가는 중요한 기술적 과제이다.

국제 비교 데이터의 활용 가능성도 제약적이다. 각국의 고령층 사망률 패턴은 의료체계, 사회보장제도, 문화적 요인 등에 따라 상당한 차이를 보이므로, 선진국의 경험을 그대로 적용하기 어렵다. 이는 특히 85세 이상 초고령층에서 더욱 두드러지는 문제로, 우리나라만의 고유한 고령층 사망률 패턴을 파악하고 이를 반영한 추계 방법론 개발의 필요성을 제기한다.

2. 한계연령 상향 조정의 의의

가. 추계 합리성 제고

현재 대부분의 국가에서 채택하고 있는 100세 한계연령 설정은 과거 인구구조와 기대수명을 기반으로 한 것으로, 급속한 수명 연장과 의료기술 발전을 충분히 반영하지 못하는 한계가 있다. 한계연령을 130세 수준으로 연장하는 접근법은 이러한 현실 변화를 보다 적절히 반영할 수 있는 강점을 제공한다. 우리나라의 기대수명은 지속적으로 증가하여 2023년 기준 남성 80.6세, 여성 86.4세에 도달했으며, 통계청 전망에 따르면 2072년에는 남성 89.5세, 여성 92.7세까지 증가할 것으로 예상된다. 이러한 추세에서 100세 이상 생존자는 급격히 증가할 것이며, 실제로 100세 이상 인구는 2025년 약 8천 명에서 2072년 약 21만 명으로 증가할

것으로 전망된다. 한계연령 상향 조정의 가장 큰 강점은 이러한 초고령 인구의 급증을 체계적으로 반영할 수 있다는 점이다. 현재 100세 한계연령하에서는 100세 이상 인구가 모두 동일한 사망률을 갖는 것으로 가정되어, 실제 100세, 110세, 120세 간의 사망률 차이를 반영하지 못한다. 반면 130세 한계연령하에서는 연령별 사망률의 자연스러운 증가 패턴을 보다 합리적으로 모형화할 수 있다. 또한 모형의 안정성 측면에서도 개선 효과를 기대할 수 있다. 현행 100세 한계연령하에서는 95세 이상에서 급작스럽게 사망률이 1.0에 수렴하는 인위적 패턴이 나타난다. 이는 실제 사망률의 자연스러운 증가 패턴과 다른 것으로, 모형의 적합도를 저하시키고 예측 오차를 증가시킬 수 있다. 한계연령을 130세로 연장하면 이러한 경계 효과(boundary effect)를 완화하여 보다 자연스러운 사망률 곡선을 구현할 수 있다. 향후 기대수명이 더욱 연장될 경우에 대비한 확장성도 확보할 수 있다. 현재의 의료기술 발전 속도를 고려할 때, 향후 50년 내에 평균 기대수명이 90세를 넘어설 가능성이 있으며, 이 경우 100세 한계연령은 더욱 부적절해진다. 130세 한계연령은 이러한 미래 변화에 대비한 여유를 제공한다. LC-ER, LLG 등 고령층 특화 모형들도 충분한 연령 범위가 확보되어야 모수 추정이 안정적으로 이루어지므로, 한계연령 연장은 이러한 모형들의 통계적 성능을 향상시키고 추계 결과의 신뢰구간을 보다 합리적으로 산출할 수 있게 해준다.

나. 정책분석 신뢰성 향상

한계연령 상향 조정은 주요 정책분석 영역에서 추계 결과의 신뢰성을 제고할 수 있다. 특히 연금 재정추계, 의료비 지출 전망, 장기요양 수요 예측 등 고령층 인구 규모에 민감한 정책 영역에서 그 효과가 두드러진다.

연금 재정추계에서의 신뢰성 제고가 대표적이다. 국민연금의 경우 가입자의 기대수명이 1년 증가할 때 연금 지급액이 물가상승률의 기대 수준인 약 2%씩 증가한다. 100세 한계연령하에서는 100세 이상 연금 수급자를 체계적으로 반영할 수 없어 재정추계에서 과소추정 위험이 발생할 수 있다. 특히 우리나라의 급속한 고령화를 고려할 때, 2050년 이후에는 이러한 추정 오차가 크게 증가할 가능성이 있다.

의료비 지출 전망의 신뢰성도 개선된다. 85세 이상 초고령층의 1인당 의료비는 전 연령 평균의 3-4배에 달하며, 연령이 증가할수록 의료비 증가율도 가파르게 상승한다. 한계연령 연장을 통해 90세, 95세, 100세 이상 각 연령대별 인구 규모를 합리적으로 전망할 수 있다면, 의료비 지출 전망의 신뢰성을 제고할 수 있다.

장기요양 수요 예측에서도 개선 효과를 기대할 수 있다. 장기요양 인정률은 85세 이상에서 급격히 증가하며, 특히 95세 이상에서는 70% 이상의 높은 인정률을 보인다. 100세 한계연령하에서는 이러한 초고령층의 장기요양 수요를 체계적으로 예측하기 어렵지만, 한계연령 연장을 통해서는 보다 합리적인 수요 전망이 가능하다.

다. 국제적 정합성 확보

한계연령 상향 조정은 국제 비교 연구와 과학적 근거 측면에서도 타당성을 갖는다.

유엔 세계인구전망(UN World Population Prospects)은 2019년부터 Coherent Kannisto 모형을 적용하여 130세까지 사망확률을 도출하고 있다. 이 방법은 80~99세의 관찰된 사망률 데이터를 바탕으로 최대우도법을 통해 모수를 추정하고, 130세까지 사망률을 외삽하는 방식이다.

UN WPP 2019, 2022, 2024 모두 이 방법론을 적용하고 있다.

최근 장수명 연구도 한계연령 상향 조정의 타당성을 뒷받침한다. Pearce and Raftery(2021)는 베이지안 인구추계를 활용하여 2100년까지의 최대 수명 도달 확률을 정량화했다. 연구 결과에 따르면, 현재 최고령 기록인 122세를 넘어설 확률은 거의 100%에 가까우며, 124세 도달 확률은 99%, 127세 도달 확률은 68%, 130세 도달 확률은 13%로 추정되었다(Pearce & Raftery, 2021). 이는 130세 한계연령 설정이 과도하지 않으며, 이 세기 말까지 실현 가능한 수명 범위 내에 있음을 보여준다.

유럽의 주요 인구연구기관들도 극고령층 사망률 연구를 활발히 수행하고 있다. 독일의 막스플랑크 인구연구소(Max Planck Institute for Demographic Research)는 International Database on Longevity (IDL)의 관리 기관으로 참여하여 110세 이상 초장수자에 대한 데이터를 수집·분석하고 있다(Jdanov et al., 2021).

이러한 국제적 동향을 고려할 때, 한계연령 연장은 다음과 같은 이점을 제공한다. 첫째, OECD 국가 간 기대수명이나 고령화 지표 비교 시 한계연령 차이로 인한 왜곡을 줄일 수 있다. 둘째, 일본과 유럽 등 선진국의 100세 이상 인구 사망률 데이터를 참조하여 추계의 합리성을 높일 수 있다. 셋째, 국제 인구연구에서 우리나라의 데이터를 활용할 수 있는 기반을 마련할 수 있다.

제3절 사망력 추계 개선 구현 및 결과 검토

1. 한계연령 연장의 기본 구조

본 연구의 한계연령 연장은 생명표 모듈(LifeTableCalculator 클래스)을 통해 구현하였으며, 크게 두 단계로 이루어진다.

첫 번째 단계는 기본 생명표 계산(0~100세)이다. 전통적인 생명표 계산 방식을 따라 0~100세의 생명표를 계산하며, 방법론은 송창길 외(2023)를 준용한다. 이 단계에서는 사망률(m_x)로부터 연령 구간 내 평균 생존 기간(a_x), 사망확률(q_x), 생존자수(l_x), 사망자수(d_x), 정지인구(L_x), 총 생존 연수(T_x), 기대여명(e_x) 등 생명표의 핵심 지표를 산출한다.

두 번째 단계는 고령 연장(100~130세)이다. 100세 이후의 사망률과 생명표 지표를 추정하기 위해 앞서 검토한 Coale-Kisker 방법과 Coherent Kannisto 방법을 모두 활용할 수 있도록 구현하였다. 사용자는 extension_method 파라미터를 통해 연장 방법을 선택할 수 있으며, 선택된 방법에 따라 80~99세 사망률 패턴을 기반으로 100~130세의 사망률을 외삽한다.

전체적인 계산 흐름은 〈표 4-4〉의 pseudo 코드에 제시된 바와 같이, 방법론 선택에 따라 성별별 독립 계산 또는 남녀 통합 회귀를 수행한 후, 남성과 여성의 생명표로부터 전체(성별=0) 생명표를 재계산하는 구조로 이루어진다. 이러한 모듈화된 설계를 통해 향후 다른 연장 방법론을 추가하거나, 시나리오별로 다른 방법을 적용하는 확장성을 확보할 수 있게 하였다.

〈표 4-4〉 LifeTableCalculator pseudo 코드

생명표 모듈 LifeTableCalculator pseudo 코드

```
# 한계연령 연장 생명표 계산

class LifeTableCalculator:
    def __init__(self, extension_method):
        """
        Args:
            extension_method: 'coale_kisker' or 'coherent_kannisto'
        """
        self.extension_method = extension_method

    def calculate(self, mx):
        """
        생명표 계산 (0-130세)

        Args:
            mx: (n_years, 3, 131) 사망률 (성별: 0=전체, 1=남, 2=여)

        Returns:
            dict: 'ax', 'qx', 'lx', 'dx', 'Lx', 'Tx', 'ex'
                  각각 (n_years, 3, 131) 형태
        """

    ## Coherent Kannisto 방법
    def _calculate__coale_kisker(self, ax, qx, lx, dx, Lx, Tx, mx):

    ## Coherent Kannisto 방법
    def _calculate_coherent_kannisto(self, mx, gender):
        """
    ## 기본 생명표 계산 (0~100세): 송창길 외(2023) 방법론 준용
    def _calculate_basic(self, mx, gender):
        """
        기본 생명표 계산 (0~100세)
        Returns:
            tuple: (ax, qx, lx, dx, Lx, Tx, ex) 각 (n_years, 101)
        """
    ## 단일 성별 생명표 계산(0~130세)
    def _calculate_single_gender(self, mx_gender, gender, method):

    ## 남녀 생명표로부터 전체(성별=0) 생명표 계산(0~130세)
    def _calculate_total_from_sex(self, results):
```

자료: 저자 작성.

2. Coale-Kisker 방법

Coale-Kisker 방법은 80~99세의 사망률 패턴을 선형 회귀로 외삽하여 100130세 사망률을 추정한다. 구현의 핵심은 다음과 같다.

첫째, 사망률의 logit 변환이다. 사망률 m_x는 0과 1 사이로 제한되므로 선형 회귀 적용이 어렵다. 이를 해결하기 위해 logit 변환($logit(m_x) = \ln(m_x/(1-m_x))$)을 사용하여($-\infty$,+$\infty$) 범위로 확장한다.

둘째, 연령별 회귀 추정이다. 80~99세의 logit 변환된 사망률을 연령에 대해 회귀하여 $logit(m_x) = \alpha + \beta \times x$ 관계식을 추정한다. 이를 100~130세로 외삽한 후 역변환하여 사망률을 구한다.

셋째, 성별별 독립 계산이다. 남성과 여성의 사망률을 각각 독립적으로 추정하며, 각 성별에 대해 별도의 회귀식을 사용한다. 이후 남녀의 생명표를 합산하여 전체(성별=0) 생명표를 산출한다.

넷째, 생명표 지표의 순차 계산이다. 추정된 사망률로부터 a_x, q_x를 계산하고, l_x를 순차적으로 산출한 후, d_x, L_x, T_x, e_x를 도출한다. 130세에서는 $q_{130} = 1.0$으로 설정한다.

구현 구조는 〈표 4-5〉와 같다. _calculate_coale_kisker() 함수는 연도별로 80~99세 데이터를 사용하여 회귀식을 추정하고, 이를 100~130세로 외삽하여 확장된 생명표를 반환한다.

〈표 4-5〉 Coale-Kisker 기반 한계연령 상향 조정 방법 pseudo 코드

Coale-Kisker 기반 한계연령 상향 조정 방법 pseudo 코드

```
# Coale-Kisker 기반 한계연령 상향 조정 사망력 추계

def _calculate_coale_kisker(self, ax, qx, lx, dx, Lx, Tx, mx):
    """
    Args:
        ax, qx, lx, dx, Lx, Tx: (n_years, 101) 기본 생명표
        mx: (n_years, 131) 사망률
    Returns:
        tuple: 확장된 생명표 (n_years, 131)
    """

    # 1. 확장 배열 초기화 및 0~100세 복사

    # 2. 연도별 반복
    for year_idx in range(n_years):

        # 3. 80~99세 mx 추출 및 logit 변환
        mx_fit = mx[year_idx, 80:100]
        logit_mx = log(mx_fit / (1 - mx_fit))

        # 4. 선형 회귀: logit(mx) = α + β*age
        beta, alpha = polyfit(ages_80_99, logit_mx, degree=1)

        # 5. 100~130세 예측 및 역변환
        logit_pred = alpha + beta * ages_100_130
        mx_pred = exp(logit_pred) / (1 + exp(logit_pred))

        # 6. ax, qx 계산 (130세는 qx=1.0)

        # 7. lx 순차 계산

    # 8. dx, Lx, Tx, ex 재계산

    return ax_ext, qx_ext, lx_ext, dx_ext, Lx_ext, Tx_ext, ex_ext
```

자료: 저자 작성.

3. Coherent Kannisto 방법

Coherent Kannisto 방법은 유엔(UN)이 World Population Prospects 2019부터 채택한 방법으로, 남녀 통합 회귀를 통해 성별 간 사망률 개선의 일관성(coherence)을 확보한다.

첫째, 남녀 통합 데이터셋 구성이다. 80~99세의 남성과 여성 사망률 데이터를 하나의 데이터셋으로 결합한다. 각 관측값에는 연령(x)과 성별 지시변수(I_{male})가 포함되어 총 40개의 관측값(남성 20개 + 여성 20개)을 생성한다.

둘째, 다중 선형 회귀 추정이다. logit 변환된 사망률을 연령과 성별에 대해 회귀하여 $logit(m_x) = \beta_0 + \beta_1 \times I_{male} + \beta_2 \times x$ 관계식을 추정한다. 여기서 β_0 는 여성의 기준 수준, β_1은 남성의 추가 사망률(성별 효과), β_2는 연령 증가에 따른 사망률 증가를 나타낸다. 이 구조는 남녀가 동일한 연령 기울기(β_2)를 공유하되, 수준의 차이(β_1)를 허용하여 일관성을 확보한다.

셋째, 성별별 예측이다. 추정된 회귀식을 사용하여 100~130세의 남성과 여성 사망률을 각각 예측한다. 남성 예측 시에는 $I_{male} = 1$, 여성 예측 시에는 $I_{male} = 0$을 사용하여 성별 효과를 반영한다.

넷째, 생명표 지표 계산이다. Coale-Kisker와 동일하게 추정된 사망률로부터 a_x, q_x, l_x, d_x, L_x, T_x, e_x를 순차적으로 계산한다.

구현 구조는 〈표 4-6〉과 같다. _calculate_coherent_kannisto() 함수는 전체 사망률 배열(m_x)을 입력받아 특정 성별의 생명표를 계산하되, 회귀 추정 단계에서는 남녀 데이터를 통합하여 사용한다.

〈표 4-6〉 Coherent Kannisto 기반 한계연령 상향 조정 방법 pseudo 코드

Coherent Kannisto 기반 한계연령 상향 조정 방법 pseudo 코드

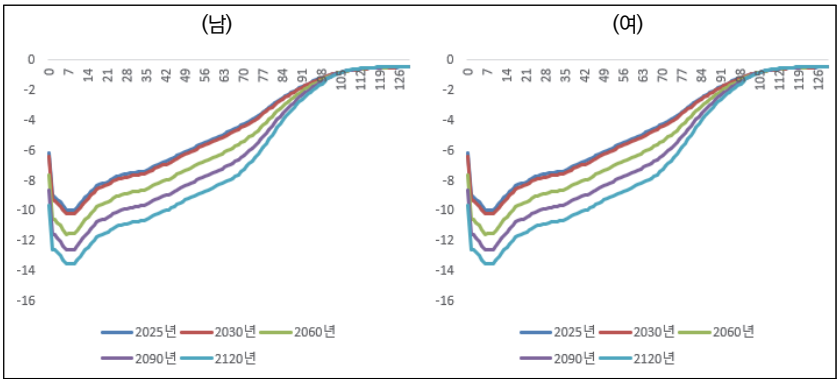
```
# Coherent Kannisto 방법 (남녀 통합 회귀)

def _calculate_coherent_kannisto(self, mx, gender):
    """
    Args:
        mx: (n_years, 3, 131) 전체 사망률
        gender: 1(남) or 2(여)
    Returns:
        dict: 해당 성별 생명표 (n_years, 131)
    """
    # 1. 기본 생명표 계산 (0~100세)
    ax, qx, lx, dx, Lx, Tx, ex = _calculate_basic(mx[:, gender, :], gender)

    # 2. 확장 배열 초기화 및 0~100세 복사

    # 3. 연도별 반복
    for year_idx in range(n_years):

        # 4. 80~99세 남녀 데이터 추출
        mx_male = mx[year_idx, 1, 80:100]
        mx_female = mx[year_idx, 2, 80:100]

        # 5. 통합 데이터셋 구성
        ages_combined = [80, ..., 99, 80, ..., 99]        # 40개
        mx_combined = [mx_male, mx_female]                # 40개
        is_male = [1, ..., 1, 0, ..., 0]                  # 40개

        # 6. Logit 변환
        logit_mx = log(mx_combined / (1 - mx_combined))

        # 7. 다중 선형 회귀: logit(mx) = β0 + β1*I(male) + β2*age
        X = [ones, is_male, ages_combined]
        model.fit(X, logit_mx)

        # 8. 100~130세 예측 (해당 성별)
        is_male_pred = 1.0 if gender == 1 else 0.0
        X_pred = [ones, is_male_pred, ages_100_130]
        logit_pred = model.predict(X_pred)

        # 9. 역변환
        mx_pred = exp(logit_pred) / (1 + exp(logit_pred))

        # 10. ax, qx 계산 (130세는 qx=1.0)

        # 11. lx 순차 계산

    # 12. dx, Lx, Tx, ex 재계산

    return {ax_ext, qx_ext, lx_ext, dx_ext, Lx_ext, Tx_ext, ex_ext}
```

자료: 저자 작성.

4. 주요 결과 검토

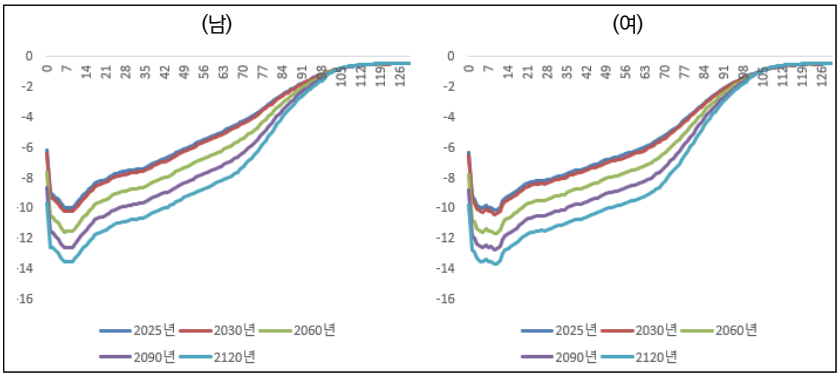
〔그림 4-5〕, 〔그림 4-6〕은 Coale-Kisker 방법과 Coherent Kannisto 방법으로 추정한 연령별 사망확률 추이를 보여준다.

〔그림 4-5〕 한계연령 상향 성별 사망확률(Coale-Kisker)



자료: 저자 산출.

〔그림 4-6〕 한계연령 상향 성별 사망확률(Coherent Kannisto)



자료: 저자 산출.

두 방법 간 사망확률 패턴은 거의 동일하게 나타났으며, 특히 100세 이상 초고령층에서도 두 방법의 결과가 매우 유사함을 확인할 수 있다. 이는 두 방법 모두 80~99세의 사망률 패턴을 선형적으로 외삽한다는 공통된 기반을 가지고 있기 때문이다.

〈표 4-7〉은 두 방법으로 추정한 생명표 주요 지표를 2025년, 2060년, 2120년에 대해 비교한 것이다. 100세 미만 연령(0, 20, 60, 80세)에서는 두 방법의 사망확률과 기대여명이 완전히 일치한다. 예를 들어 2025년 80세 남성의 사망확률은 두 방법 모두 0.05065, 기대여명은 8.6년으로 동일하다.그러나 100세 이상 초고령층에서는 두 방법 간 차이가 나타나기 시작한다. 2025년 100세 남성의 사망확률은 Coale-Kisker 방법에서 0.34924인 반면 Coherent Kannisto 방법에서는 0.36499로 약 1.6%p 높다. 이러한 차이는 연령이 증가할수록 확대되어, 105세에서는 약 2%p, 110세에서는 약 2.7%p의 차이를 보인다. Coherent Kannisto 방법이 상대적으로 높은 사망확률을 추정하는 것은 남녀 통합 회귀에서 성별 간 일관성을 유지하면서 보수적으로 추정하기 때문으로 판단된다. 기대여명 측면에서는 100세에서 Coale-Kisker 방법이 남성 2.1년, 여성 2.4년인 반면 Coherent Kannisto 방법은 남성 2.0년, 여성 2.5년으로 남성에서 약간 낮고 여성에서 약간 높게 나타난다. 그러나 이러한 차이는 0.1년 수준으로 실질적으로 매우 미미하다.시간이 경과할수록 두 방법 간 차이는 다소 확대되는 경향을 보인다. 2120년 100세 사망확률의 경우 남성에서 Coale-Kisker 0.28486, Coherent Kannisto 0.30201로 약 1.7%p, 여성에서는 0.24885와 0.23263으로 오히려 Coale-Kisker가 더 높다. 이는 장기 추계에서 회귀 방식의 차이가 누적되어 나타나는 효과로 해석된다.

〈표 4-7〉 상한연령 상향 설정에 따른 생명표(Coale-Kisker vs Coherent Kannisto)

(단위: 년)

연도	Coale-Kisker					연도	Coherent Kannisto				
	성별	남		여			성별	남		여	
	연령	사망 확률	기대 여명	사망 확률	기대 여명		연령	사망 확률	기대 여명	사망 확률	기대 여명
2025	0	0.00231	81.2	0.00196	87.0	2025	0	0.00231	81.2	0.00196	87.0
	20	0.00031	61.5	0.00024	67.3		20	0.00031	61.5	0.00024	67.3
	60	0.00616	23.8	0.00220	28.6		60	0.00616	23.8	0.00220	28.6
	80	0.05065	8.6	0.02619	11.1		80	0.05065	8.6	0.02619	11.1
	100	0.34924	2.1	0.30284	2.4		100	0.36499	2.0	0.28692	2.5
	105	0.44590	1.6	0.41887	1.7		105	0.46492	1.5	0.39723	1.8
	110	0.51998	1.3	0.51035	1.4		110	0.53702	1.3	0.48976	1.4
	115	0.56861	1.2	0.56833	1.2		115	0.58124	1.1	0.55289	1.2
	120	0.59745	1.1	0.60029	1.1		120	0.60582	1.1	0.59027	1.1
	125	0.61357	1.0	0.61663	1.0		125	0.61877	1.0	0.61063	1.1
	130	1.00000	0.4	1.00000	0.4		130	1.00000	0.4	1.00000	0.4
2060	0	0.00051	88.1	0.00044	91.8	2060	0	0.00051	88.1	0.00044	91.8
	20	0.00008	68.2	0.00006	71.9		20	0.00008	68.2	0.00006	71.9
	60	0.00180	28.9	0.00065	32.3		60	0.00180	28.9	0.00065	32.3
	80	0.02403	11.1	0.01238	13.3		80	0.02403	11.1	0.01238	13.3
	100	0.32245	2.2	0.28052	2.5		100	0.33883	2.1	0.26444	2.7
	105	0.44257	1.6	0.41961	1.7		105	0.46229	1.5	0.39744	1.8
	110	0.53058	1.3	0.52463	1.3		110	0.54663	1.2	0.50537	1.4
	115	0.58211	1.1	0.58348	1.1		115	0.59246	1.1	0.57101	1.2
	120	0.60855	1.1	0.61128	1.1		120	0.61444	1.0	0.60438	1.1
	125	0.62125	1.0	0.62340	1.0		125	0.62438	1.0	0.61988	1.0
	130	1.00000	0.4	1.00000	0.4		130	1.00000	0.4	1.00000	0.4
2120	0	0.00007	93.9	0.00006	96.1	2120	0	0.00007	93.9	0.00006	96.1
	20	0.00001	73.9	0.00001	76.1		20	0.00001	73.9	0.00001	76.1
	60	0.00025	34.0	0.00009	36.1		60	0.00025	34.0	0.00009	36.1
	80	0.00686	14.5	0.00353	16.4		80	0.00686	14.5	0.00353	16.4
	100	0.28486	2.4	0.24885	2.7		100	0.30201	2.3	0.23263	2.8
	105	0.44324	1.6	0.42553	1.6		105	0.46387	1.5	0.40250	1.7
	110	0.54994	1.2	0.54805	1.2		110	0.56392	1.2	0.53136	1.3
	115	0.60041	1.1	0.60264	1.1		115	0.60742	1.1	0.59435	1.1
	120	0.62047	1.0	0.62235	1.0		120	0.62355	1.0	0.61885	1.0
	125	0.62792	1.0	0.62894	1.0		125	0.62919	1.0	0.62756	1.0
	130	1.00000	0.4	1.00000	0.4		130	1.00000	0.4	1.00000	0.4

자료: 저자 산출.

〈표 4-8〉은 본 연구의 사망력 추계 결과를 통계청의 「장래인구추계: 2022~2072년」 중위 시나리오와 비교한 것이다. 본 연구의 두 방법(Coale-Kisker, Coherent Kannisto)은 0세 기대수명에서 동일한 결과를 나타냈으며, 이는 100세 이상 초고령층의 사망확률 차이가 전체 기대수명에 미치는 영향이 극히 제한적임을 보여준다.

〈표 4-8〉 사망력 가정별 기대수명

(단위: 년)

가정	구분	2023*	2024	2025	2030	2040	2050	2060	2072
통계청 -중위	전체	83.5	84.3	84.5	85.5	87.2	88.6	89.8	91.1
	남자(A)	80.6	81.4	81.6	82.8	84.8	86.5	88.0	89.5
	여자(B)	86.4	87.1	87.3	88.1	89.5	90.7	91.7	92.7
	A-B	5.9	5.7	5.7	5.3	4.7	4.2	3.7	3.3
[본 연구] Coale-Kisker	전체	83.9	84.1	84.3	85.4	87.2	88.8	90.1	91.4
	남자(A)	80.9	81.2	81.4	82.7	84.9	86.7	88.2	89.8
	여자(B)	86.8	87.0	87.2	88.1	89.6	90.8	91.9	93.0
	A-B	5.9	5.8	5.8	5.4	4.7	4.1	3.7	3.2
[본 연구] Coherent Kannisto	전체	83.9	84.1	84.3	85.4	87.2	88.8	90.1	91.4
	남자(A)	80.9	81.2	81.4	82.7	84.9	86.7	88.2	89.8
	여자(B)	86.8	87.0	87.2	88.1	89.6	90.8	91.9	93.0
	A-B	5.9	5.8	5.8	5.4	4.7	4.1	3.7	3.2

주: * 2022년 생명표와 2023년 인구동향조사(사망통계) 등을 기초로 생명표를 별도 작성.
자료: 통계청(2023) '장래인구추계: 2022~2072년', 본 연구는 저자 산출.

본 연구의 기대수명은 통계청 추계보다 전반적으로 약간 높게 나타났다. 2023년의 경우 본 연구는 전체 83.9세(남성 80.9세, 여성 86.8세)인 반면 통계청은 83.5세(남성 80.6세, 여성 86.4세)로 약 0.3~0.4세의 차이를 보인다. 이러한 차이는 추계 기간 동안 일관되게 유지되어, 2072년에는 본 연구 91.4세(남성 89.8세, 여성 93.0세), 통계청 91.1세(남성 89.5세, 여성 92.7세)로 약 0.3세의 차이가 나타난다.

이러한 차이는 기준연도 생명표 작성 방식, 모형 파라미터 설정 등의 차이에 기인하며, 추계의 전반적인 경향에는 영향을 미치지 않는다. 본 연구의 기대수명은 통계청 추계와 유사한 수준으로, 두 방법론 모두 장기적인 사망률 개선 추세에 대해 일관된 전망을 제시하고 있음을 확인할 수 있다.

성별 간 기대수명 격차(A-B)는 두 추계 모두 유사한 패턴을 보인다. 2023년 5.9년에서 시작하여 2072년 3.2~3.3년으로 감소하는 추세는 남성의 사망률 개선 속도가 여성보다 빠르다는 가정을 공통적으로 반영한 결과이다. 본 연구와 통계청의 성별 격차 추이가 거의 동일한 것은 두 추계 모두 LLG 확장 모형을 기반으로 하여 남녀 사망률 개선의 상대적 속도를 유사하게 가정하고 있기 때문이다.

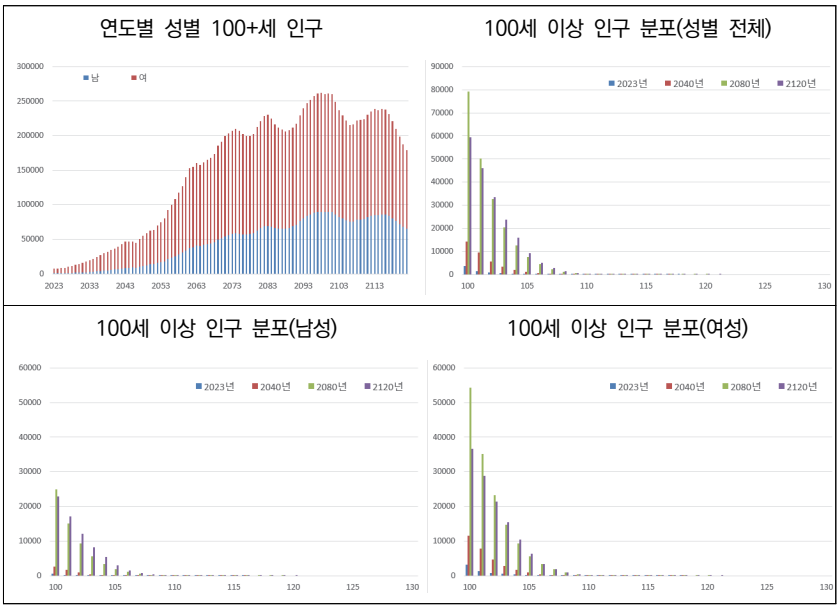
한계연령을 100세에서 130세로 연장한 효과는 연령대에 따라 차별적으로 나타난다. 0세 기대수명에 미치는 영향은 0.3~0.4년 수준으로 제한적이지만, 100세 기대여명은 Coale-Kisker 방법 기준 2025년 남성 2.1년, 여성 2.4년으로 100세 이후에도 평균 2년 이상 생존하는 것으로 추정된다. 이는 100세 한계 설정하에서는 반영할 수 없는 초장수 가능성을 포착한 것이다.

또한 한계연령 연장은 100세 이상 인구의 연령별 분포를 세밀하게 파악할 수 있게 한다. 표에서 확인할 수 있듯이 105세, 110세, 115세 등 5세 간격으로 사망확률과 기대여명을 추정함으로써, 초고령층 내부의 이질성을 반영할 수 있다. 예를 들어 2025년 남성의 경우 100세 사망확률 0.349에서 110세 0.520, 120세 0.597로 점진적으로 증가하는 패턴을 보이며, 이는 극고령층 정책 설계에 유용한 정보를 제공한다.

Coale-Kisker와 Coherent Kannisto 두 방법의 선택이 실질적 정책 분석에 미치는 영향은 제한적이다. 0세 기대수명이 완전히 동일하고,

100세 이상에서만 차이가 나타나므로, 일반적인 인구추계나 정책 분석에서는 두 방법 중 어느 것을 선택하더라도 큰 문제가 없다. 다만 100세 이상 초장수자 연구나 극고령층 특화 정책(예: 100세 이상 의료비, 장기요양 수요 등)을 분석하는 경우에는 방법 선택이 의미를 가질 수 있다. 이러한 경우 국제 표준과의 정합성 측면에서 Coherent Kannisto 방법을 우선적으로 고려할 수 있다.

〔그림 4-7〕 100세 이상 인구 수(Coherent Kannisto)



자료: 저자 산출.

위 〔그림 4-7〕은 Coherent Kannisto 방법으로 추계한 100세 이상 인구의 규모와 연령별 분포를 보여준다(Coale-Kisker 방법과 큰 차이 없음). 좌상단 그래프는 2023년부터 2122년까지 100세 이상 인구 추이를 보여주는데, 2023년 약 7만 명(남성 1.5만 명, 여성 5.5만 명)에서

2080년대 약 20만 명(남성 5만 명, 여성 15만 명 수준)까지 급증한 후 유지하다가 감소하는 패턴을 나타낸다. 여성이 남성보다 2배 이상 많은 것은 여성의 높은 기대수명과 성비 불균형을 반영한 것이다.

우상단, 좌하단, 우하단 그래프는 각각 전체, 남성, 여성의 100세 이상 연령별 인구 분포를 보여준다. 모든 연도에서 100세 인구가 가장 많고 연령이 증가할수록 급격히 감소하는 지수함수적 감소 패턴이 나타난다. 2023년에는 105세 이상 1천 명 미만으로 인구가 거의 없지만, 2080년과 2120년에는 105세 이상 약 1~2만 명, 110세 이상은 1천 명 미만 수준으로 관찰된다. 특히 100~105세 구간에 상당한 인구가 분포하고 있고, 초고령층 정책 수요가 단일 연령대에 집중되지 않고 여러 연령대에 걸쳐 나타날 것임을 시사한다.

제5장

통합 추계 체계 구축 및 활용 방안

제1절 인구·가구 통합 추계 모형 구조 및 설계
제2절 통합 추계 모형 결과 검토
제3절 KIHASA-SIM 연계 시뮬레이션 체계 설계

제 5 장 통합 추계 체계 구축 및 활용 방안

제1절 인구·가구 통합 추계 모형 구조 및 설계

1. 전체 시스템 아키텍처

가. 데이터 구조 설계

본 연구에서 구축한 인구추계 시스템은 데이터 흐름의 명확성과 재현 가능성을 확보하기 위해 3계층 데이터 구조를 채택하였다. 전체 데이터 디렉토리는 /data/ 하위에 actuals/, assumptions/, projections/의 세 폴더로 구성되며, 각각 입력 데이터, 가정 데이터, 산출 결과의 역할을 명확히 구분한다.

1) 실적 데이터 구조 (actuals/)

실적 데이터 계층에는 1960년부터 2022년까지의 인구 관련 실적 통계가 저장된다. 주요 데이터는 다음과 같다.

- 확정인구 (1960~2022년): 성별, 연령별 총인구
- 생명표 (1970~2022년): 연도별, 성별, 연령별 사망률 및 생명표 지표
- 출산 실적 (2000~2022년): 모의 연령별 출생아수
- 인구동향 실적 (2014~2022년): 혼인, 사망, 이혼 신고 자료

- 인구주택총조사 샘플 (2015, 2020년): 2% 표본 자료

이들 실적 데이터는 가정 추정의 기초 자료로 활용되며, 데이터베이스(PostgreSQL)로 관리되어 DBManager 클래스를 통해 접근한다. 실적 데이터의 최신성 유지를 위해 db_init_run.sh (초기 로드)와 db_update_run.sh (업데이트) 스크립트가 제공된다.

2) 가정 데이터 구조 (assumptions/)

가정 데이터 계층은 입력과 출력의 이중 역할을 수행한다. 사망력, 이동력 가정은 각 전문 모듈에서 생성되어 이 계층에 저장되며(출력), 인구추계 모듈에서 이를 읽어들여 사용한다(입력). 가정 데이터는 5개 하위 폴더로 구성된다.

- fertility/: 연령별 출산율(ASFR) 가정 또는 출산진도비(PPR) 계수
- mortality/: 연령별 사망률(m_x) 가정 및 생명표
- mobility/: 국내이동률 및 국제이동률 가정
- household/: 가구주율, 가구구성비율 가정 (PPR 방식 사용 시)
- marital/: 혼인상태별 전이율 및 초기 인구 (PPR 방식 사용 시)

특히 PPR 방식의 출산력 추계에서는 추계 실행 후 역산된 ASFR이 assumptions/fertility/ 폴더에 저장된다. 예를 들어 kihasa_fertility_ppr_2022_2122.xlsx 파일에는 2023~2055년의 연령별 ASFR이 시나리오별로 기록되며, 이는 추계 결과의 재현성을 확보하고 동일 시나리오 재실행 시 일관성을 유지하는 역할을 한다.

3) 추계 결과 저장 구조 (projections/)

모형의 산출 결과는 시나리오별 인구추계 결과가 저장된다. 각 시나리오 조합마다 별도의 폴더가 생성되며, 폴더명은 {출산기관}-{출산시나리오}_{사망기관}-{사망시나리오}_{이동기관}-{이동시나리오} 형식을 따른다. 예를 들어 kihasa-mid_kihasa-mid_kostat-mid/는 출산과 사망은 한국보건사회연구원 중위 시나리오, 이동은 통계청 중위 시나리오를 의미한다. 각 시나리오 폴더 내에는 결과 관리를 위한 두 개의 하위 폴더가 생성된다.

- latest/: 가장 최신 추계 결과 (타임스탬프 제거된 파일명)
- archive/: 과거 추계 결과 보관 (타임스탬프 포함 원본 파일명)

추계 실행 시 population_2023_2122_20250115_1430.xlsx와 같이 타임스탬프가 포함된 파일명으로 저장되며, 추계 완료 후 자동 아카이빙 프로세스를 통해 최신 파일은 latest/population_2023_2122.xlsx로 복사되고, 모든 버전은 archive/로 이동된다. 이러한 구조는 결과의 버전 관리를 가능하게 하면서도, 사용자가 최신 결과에 쉽게 접근할 수 있도록 한다.

나. 모듈 구조 설계

인구추계 시스템은 기능별로 분리된 6개의 핵심 모듈로 구성되며, 각 모듈은 독립적으로 실행 가능하면서도 상호 연계된다. 모듈 구조는 다음과 같다.

1) 공통 모듈 (common/)

공통 모듈은 전체 시스템의 설정과 공용 함수를 제공한다. config.py에는 데이터 경로, 연령 범위(N_AGES=131), 출생성비 등 시스템 전반의 상수가 정의되어 있다.

2) 데이터 관리 모듈 (dbloader/)

데이터베이스 관리를 담당하는 모듈로, DBManager 클래스를 통해 PostgreSQL 데이터베이스의 실적 데이터에 접근한다. init_loader.py는 초기 데이터 적재를, update_loader.py는 최신 데이터 업데이트를 담당한다.

3) 사망력 모듈 (mortality/)

사망력 전망을 담당하는 모듈로, MortalityCalculator 클래스가 LC, LL, LC-ER, LLG-E 등 다양한 사망률 모형을 구현한다. 추정된 사망률 가정은 assumptions/mortality/에 저장된다.

4) 출산력 모듈 (fertility/)

출산력 가정 생성을 담당하는 모듈로, FertilityCalculator 클래스가 GLG 방식의 ASFR 가정 로드 또는 PPR 방식의 출산진도비 계산을 수행한다. PPR 방식 선택 시 초기 PPR 계수를 추정하며, 추계 실행 후 역산된 ASFR은 assumptions/fertility/에 저장된다.

5) 이동력 모듈 (mobility/)

국내이동과 국제이동 가정을 생성하는 모듈로, MobilityCalculator 클래스가 과거 추세 기반 이동률을 추정한다. 결과는 assumptions/mobility/에 저장된다. 현재 이동력 추계 방법론은 미구현 상태이며, 통계청의 시나리오별 가정 파일들이 저장되어있다.

6) 인구추계 모듈 (population/)

핵심 추계 모듈로, 다음의 하위 클래스들이 통합되어 작동한다.

- AssumptionLoader: 가정 데이터 로드
- LifeTableCalculator: 생명표 계산
  (Base/Coale-Kisker/Coherent Kannisto 확장방법)
- MaritalTransitionCalculator: 혼인상태별 전이율 추정 및 조정
- HouseholdCalculator: 가구주율, 가구구성비율 추계
- PopulationCalculator: 코호트요인법 기반 통합 인구추계

PopulationCalculator는 --fertility-method 옵션을 통해 GLG 또는 PPR 방식을 선택할 수 있으며, PPR 선택 시 MaritalTransitionCalculator와 HouseholdCalculator가 자동으로 초기화되어 통합 추계를 수행한다.

다. 통합 추계 파이프라인 구조

전체 추계 프로세스는 ProjectionPipeline 클래스를 통해 통합 관리된다. 파이프라인은 5단계로 구성된다.

1단계의 Fertility 전망에서는 fertility_run.py를 호출하여 출산력 가정을 생성한다. GLG 방식이면 통계청 또는 연구원의 ASFR 가정을 그대로 사용하고, PPR 방식이면 초기 PPR 계수를 추정한다. 결과는 assumptions/fertility/에 저장된다.

2단계의 Mortality 전망에서는 mortality_run.py를 호출하여 사망률 가정을 생성한다. 선택된 모형(LC/LL/LC-ER/LLG-E)과 시뮬레이션 횟수에 따라 사망률을 추정하고, 결과를 assumptions/mortality/에 저장한다.

3단계의 Mobility 전망에서는 mobility_run.py를 호출하여 국내이동과 국제이동 가정을 생성한다. 결과는 assumptions/mobility/에 저장된다.

4단계의 Population 추계에서는 population_run.py를 시나리오 조합별로 반복 호출하여 인구추계를 수행한다. 예를 들어 출산·사망·이동 각각 3개 시나리오(low/mid/high)가 있으면 총 3×3×3=27개 조합을 순차 실행한다. 각 추계 결과는 projections/시나리오명/에 저장되며, PPR 방식의 경우 역산된 ASFR이 assumptions/fertility/에도 저장된다.

5단계의 결과 아카이빙은 추계 완료 후 archive_projection_results() 함수가 자동 호출되어, 타임스탬프가 포함된 모든 결과 파일을 정리한다. 각 시나리오 폴더에서 동일 유형의 파일 중 최신 버전을 latest/에 복사하고, 모든 버전을 archive/로 이동한다.

또한 추계의 파이프라인 실행은 다음 세 가지 방식을 지원한다. 먼저, '전체 실행'으로, 1~5단계 모두 수행하는 옵션이다. 두 번째는 부분 건너뛰기는 옵션이다. --skip-assumptions 옵션으로 특정 가정 전망 생략(예: 이동 가정은 기존 파일 사용) 가능하다. 세 번째 옵션은 인구추계만 실행하는 옵션이다. --only-population 옵션으로 4~5단계만 수행 (모

든 가정 파일 재사용) 가능하다. 이러한 유연성은 가정 변경 없이 추계만 반복 실행하거나, 특정 요인의 가정만 업데이트하는 등의 효율적인 작업을 가능하게 한다.

2. 주요 모듈 설계

가. DBManager: 실적 데이터 관리

DBManager 클래스는 PostgreSQL 데이터베이스에 저장된 실적 데이터에 대한 인터페이스를 제공한다. 주요 기능은 다음과 같다. 먼저, db_to_numpy()는 특정 연도의 인구, 생명표, 출산, 사망, 이동 데이터를 numpy 배열로 로드한다. 두 번째로 get_latest_year()는 데이터베이스의 최신 데이터 연도를 조회하며, 세 번째로, validate_data()는 데이터 무결성 검증(연령 합계, 성비 등)한다.

나. AssumptionLoader: 가정 데이터 로드

AssumptionLoader 클래스는 assumptions/ 폴더에 저장된 다양한 형식의 가정 데이터를 표준화된 numpy 배열로 변환한다. 주요 메서드는 다음과 같다. 먼저, load_all()은 출산, 사망, 이동 가정을 일괄 로드하며, load_fertility()은 ASFR 가정(Excel 파일)을 파이썬 numpy (n_years, 3, 131) 배열로 전환하는 기능을 한다. load_mortality()은 사망률 가정(Excel/npy)을 파이썬 numpy (n_years, 3, 131) 배열로 전환한다. load_mobility() 함수는 이동률 가정(Excel 파일)을 numpy (n_years, 3, 131) 배열로 전환한다. 각 가정 파일은 기관(agency), 시나리오(scenario),

연도 범위에 따라 명명되며, AssumptionLoader는 파일명 규칙을 해석하여 올바른 파일을 자동 선택한다.

다. PopulationCalculator: 통합 인구추계

PopulationCalculator는 인구추계의 핵심 클래스로, 코호트요인법을 구현하며 GLG와 PPR 두 가지 출산력 방식을 지원한다. 초기화 시 〈〉와 같은 파라미터를 입력받는다.

GLG 방식에서는 ASFR 가정을 직접 사용하여 출생아를 계산하며, PPR 방식에서는 혼인상태별 유배우 여성 인구에 출산진도비를 적용한다. PPR 방식 선택 시 `MaritalTransitionCalculator`와 `HouseholdCalculator`(PPR_SIMPLE은 제외)가 자동으로 초기화되며, 매년 다음 순서로 계산이 진행된다.

1. t년도 혼인상태별 인구 확인
2. 유배우 여성 인구 × PPR → 출생아 산출
3. 코호트요인법 (출생, 사망, 이동) → t+1년도 총인구 산출
4. 혼인상태 조정 (7단계 역산 절차) → t+1년도 혼인상태별 인구

추계 완료 후 출생아수를 여성 인구로 나누어 ASFR을 역산하고, 이를 'assumptions/fertility/' 폴더에 저장한다. 예를 들어 'kihasa_fertility_ppr_2022_2122.xlsx' 파일에는 2023~2055년의 연령별 ASFR이 시나리오별 시트로 기록된다. 2056년 이후는 2055년의 ASFR을 고정하여 GLG 방식과 동일하게 추계한다.

〈표 5-1〉 PopulationCalculator 주요 파라미터

PopulationCalculator 주요 파라미터
PopulationCalculator(
base_year=2022, # 기준연도
end_year=2122, # 추계 종료연도
fertility_file="...", # 출산 가정 파일
mortality_file="...", # 사망 가정 파일
domestic_mobility_file="...", # 국내이동 가정 파일
foreign_mobility_file="...", # 국제이동 가정 파일
fertility_scenario='mid', # 출산 시나리오
mortality_scenario='mid', # 사망 시나리오
domestic_mobility_scenario='mid', # 이동 시나리오
foreign_mobility_scenario='mid',
extension_method='coale_kisker', # 생명표 연장 방법
fertility_method='PPR', # 출산력 방식 (GLG/PPR/PPR_SIMPLE)
marital_reestimate=False # 혼인상태 재추정 여부)

자료: 저자 작성.

라. 연계 모듈

그 밖에 PopualationCalculator에서 활용하기 위한 연계 모듈이 구현되어 있다.

먼저, LifeTableCalculator에서는 사망률(m_x)로부터 생명표를 계산한다. 0~100세는 전통적 방식, 100~130세는 선택된 연장 방법(Coale-Kisker 또는 Coherent Kannisto)을 적용한다. 결과는 a_x, q_x, l_x, d_x, L_x, T_x, e_x 의 7개 지표로 구성된다. 이는 송창길 외(2023)의 연구에서 구현한 생명표 추계 방식에 본 연구에서의 한계연령 확장 방식을 추가 구현한 모듈이다.

두 번째로 MaritalTransitionCalculator는 혼인상태별 전이율을 추정하고, 매년 인구추계 결과와 정합성을 확보하도록 전이율을 동적으로 조정한다. 초기 전이율은 2020년 인구동향 실적에서 추정하며, 시나리오에 따라 초혼율을 연간 0% (저위), 1.5% (중위), 3.0% (고위) 증가시킨다.

추계 과정에서 역산 조정과 평활화를 통해 혼인상태별 인구의 합계가 총인구와 일치하도록 관리한다.

세 번째로, HouseholdCalculator는 PPR 방식에서 가구주율과 가구구성비율을 추계한다. 과거 인구총조사 자료를 기반으로 연령별, 성별, 혼인상태별 가구주율을 추정하고, 1세대 부부 가구와 2세대 부부+자녀 가구의 비율을 전망한다. 이들 가정은 유배우 인구를 가구 단위로 변환하는 데 사용된다. 대부분의 로직은 송창길 외(2024)에서 구현한 가구추계 방식이다.

마지막으로, FertilityCalculator이다. 해당 모듈의 GLG 방식에서는 ASFR 가정을 로드하는 역할만 수행한다. PPR 방식에서는 과거 출생 신고 자료와 가구 자료를 결합하여 초기 PPR 계수(무자녀→첫째, 유자녀→둘째)를 추정한다. 본 연구에서는 통합형 PPR을 사용하므로, 전체 유배우 여성에 대한 단일 PPR 계수를 산출한다.

3. 시나리오 관리 체계

가. 시나리오 조합 방식

본 연구의 인구추계는 출산, 사망, 이동의 세 요인에 대해 독립적인 시나리오를 설정하고, 이를 조합하여 다양한 미래 경로를 전망한다. 각 요인은 일반적으로 저위(low), 중위(mid), 고위(high)의 3개 시나리오를 갖지만, 필요에 따라 더 많은 시나리오를 추가할 수 있다.

각 시나리오들의 조합은 'kihasa-mid_kihasa-mid_kostat-mid'와 같은 방식으로 식별할 수 있다. 이는 출산과 사망은 한국보건사회연구원 중위 시나리오, 이동은 통계청 중위 시나리오를 의미한다. 기관을 요인별

로 다르게 설정할 수 있어, 예를 들어 출산은 연구원 가정, 사망과 이동은 통계청 가정을 조합하는 등 유연한 시나리오 구성이 가능하다.

나. 결과 저장 구조

각 시나리오 조합마다 projections/ 하위에 별도 폴더가 생성되며, 추계 결과는 다음 파일들로 저장된다.

- population_2023_2122_YYYYMMDD_HHMM.xlsx: 인구
- births_2023_2122_YYYYMMDD_HHMM.xlsx: 출생아수
- deaths_2023_2122_YYYYMMDD_HHMM.xlsx: 사망자수
- migrations_2023_2122_YYYYMMDD_HHMM.xlsx: 순이동자 수
- summary_1960_2122_YYYYMMDD_HHMM.xlsx: 종합 요약

여기서 YYYYMMDD_HHMM은 추계 실행 시각을 나타내는 타임스탬프이다. 이러한 명명 규칙은 동일 시나리오를 반복 실행하더라도 결과가 덮어씌워지지 않고 모두 보존되도록 한다. 추계 완료 후 아카이빙 프로세스가 자동 실행되어, 각 시나리오 폴더에 'latest/'와 'archive/' 하위 폴더를 생성하며, 'latest/' 폴더에는 각 유형별 최신 파일을 타임스탬프 없는 파일명으로 복사(예: population_2023_2122.xlsx)한다. 'archive/'에는 모든 과거 버전을 타임스탬프 포함 원본 파일명으로 보관하고 있다. 이러한 구조는 사용자가 최신 결과에 빠르게 접근하면서도, 필요시 과거 버전을 참조할 수 있는 이력 관리 기능을 제공한다.

다. 실행 명령어 예시

전체 인구추계 파이프라인은 〈표 5-2〉와 같은 명령어로 실행된다. 특정 가정을 재사용하기 위한 '--skip-assumptions' 옵션을 사용하며, 〈표 5-2〉의 '예시-1'은 사망력 가정만 새로 추정하고, 출산과 이동은 기존 파일을 사용하는 것을 의미한다. 가정 변경 없이 인구추계만 재실행하려면 '--only-population' 옵션을 사용한다. 〈표 5-2〉의 '예시-2'는 'assumptions/' 폴더의 모든 파일을 재사용하고 4~5단계(Population 추계 + 아카이빙)만 수행하는 사례이다.

〈표 5-2〉 인구추계 모형 실행 방법

인구추계 모형 실행 방법

``## 전체 실행 코드
python -m projection_pipeline \
  --fertility-agency kihasa \
  --mortality-agency kihasa \
  --mobility-agency kostat \
  --base-year 2022 \
  --end-year 2122 \
  --fertility-scenarios mid \
  --mortality-scenarios mid \
  --mobility-scenarios mid \
  --extension-method coale_kisker \
  --fertility-method PPR_SIMPLE \
  --mortality-method LLGE \
  --mortality-sims 1000
 	
## 예시-1 	
python -m projection_pipeline \
  --skip-assumptions fertility mobility \
  ... (기타 파라미터)
 	
## 예시-2 	
python -m projection_pipeline \
  --only-population \
  ... (기타 파라미터)```

자료: 저자 작성.

제2절 통합 추계 모형 결과 검토

1. 추계 시나리오 설정

가. 시나리오 조합 구성

통합 추계의 검토를 위해 통합 추계 시스템을 활용하여 출산력 시나리오를 중심으로 3개의 인구추계를 수행하였다. 출산력은 한국보건사회연구원(KIHASA)의 저위, 중위, 고위 시나리오를 각각 적용하였으며, 사망력과 이동력은 단일 시나리오를 사용하였다.

출산력 시나리오를 3개로 설정한 이유는 다음과 같다. 첫째, 출산력은 인구 변동의 핵심 요인으로, 장기 추계에서 인구 규모와 구조에 가장 큰 영향을 미친다. 둘째, 저출산 대응 정책의 효과는 출산력 변화로 나타나므로, 정책 시나리오 분석을 위해서는 출산력의 다양한 경로를 전망할 필요가 있다. 셋째, 본 연구의 PPR 방식은 초혼율 증가를 핵심 정책 변수로 설정하므로, 초혼율 수준에 따른 민감도를 평가하기 위해 3개 시나리오를 구성하였다.

반면 사망력은 중위 시나리오만 사용하였다. 제4장 3절에서 확인한 바와 같이, Coale-Kisker와 Coherent Kannisto 두 방법 간 기대수명 차이가 0.1세 이내로 미미하며, 저위·중위·고위 시나리오 간 차이는 지난 연구에서 검토하였기 때문에 중위 시나리오만으로도 합리적인 추계 결과를 도출할 수 있다고 판단하였다. 이동력은 아직 구현 전이므로 역시 단일 시나리오(통계청 중위)를 사용하였다.

나. 주요 추계 파라미터

본 절에서 비교할 3개 시나리오에 공통으로 적용된 추계 파라미터는 다음과 같다. 기준연도는 2022년, 추계 시작연도는 2023년, 추계 종료연도는 2122년(100년 추계) 이다. 기준 인구는 DB에 저장된 2022년 성별·연령별 확정인구를 활용한다. 출산력 추계 방식은 PPR_SIMPLE (단순화된 출산진도비 방식)을 활용하며, 2023~2055년까지 전망한다. 이후 2056~2122년은 2055년에 산출된 ASFR이 고정되는 시나리오다.

출산력 추계에서 PPR_SIMPLE 방식을 선택한 이유는 다음과 같다. 본 연구에서 제안한 PPR 방식은 원칙적으로 가구 구성(무자녀 부부 / 유자녀 부부)을 구분하여 첫째 출산 PPR과 둘째 이상 출산 PPR을 별도로 적용하는 분리형 구조이다. 그러나 분리형 방식은 매년 출산으로 인한 가구원 수 변화를 추적해야 하므로, 가구 추계와의 정합성 유지가 중요하다. 그러나 기존 연구에서 구현한 가구추계는 수정지수모형을 활용하여 단순 추세 연장을 활용하기 때문에 출산 → 자녀 수 증가 → 가구 구성 변화 → 다음 연도 PPR 적용의 순환 과정에서 불일치가 발생할 위험이 있다. 이에 본 연구에서는 가구 구성 정보 없이도 추계가 가능한 통합형 PPR을 적용하는 PPR_SIMPLE 방식을 채택하였다. 이 방식은 전체 유배우 여성 인구에 단일 PPR을 적용하므로, 출산 순위별 의사결정의 차이를 명시적으로 반영하지 못하는 한계가 있다. 그러나 추계의 안정성과 재현 가능성을 확보하고, 가구 추계 없이도 출산진도비 개념을 인구추계에 적용할 수 있다는 장점이 있다. 향후 가구 추계와의 완전한 통합이 구현된다면, 분리형 방식으로 전환할 수 있다.

사망력 추계 방식에서는 사망률 모형을 LLG-E(Lee-Li-Gerland Extended)모형을 활용하며, 시뮬레이션을 1,000회 시행하여 중위 (50th

percentile) 수준을 적용한다. 생명표는 0~100세는 전통적 방식(Base 모듈)으로 계산하고, 100~130세는 Coale-Kisker 방법으로 연장하였다. 제4장에서 확인한 바와 같이, Coale-Kisker와 Coherent Kannisto 두 방법 간 차이가 미미하므로, 구현이 상대적으로 단순한 Coale-Kisker를 선택하였다. 한계연령 130세 설정은 향후 기대수명 연장에 대비하고, 100세 이상 초고령층 인구를 연령별로 세분화하여 정책 분석의 정밀도를 높이기 위함이다.

이동력은 통계청의 중위 가정을 그대로 사용하였다. 통계청은 과거 출입국 추세와 정책 방향을 고려하여 국제이동 가정을 설정하고 있으며, 본 연구에서 이를 독자적으로 재추정할 명확한 근거가 없다고 판단하였다.

다. 시나리오별 가정 차이

3개 시나리오의 핵심 차이는 초혼율 증가 속도이다. PPR 방식에서 출산은 주로 유배우 여성에서 발생하므로, 초혼율이 유배우 여성 인구 규모를 결정하는 핵심 변수이다. 본 연구에서는 초혼율의 연간 증가율을 저위·중위·고위로 차별화하여 시나리오를 구성하였다.

〈표 5-3〉 출산율 시나리오별 주요 가정

시나리오	가정	대응 시나리오
초혼율 유지	현 수준 유지(초혼율 정체)	합계출산율 0.62 수준 - 현 출산율 유지
초혼율 '+1.5%/연'	완만한 증가(정책 효과)	합계출산율 0.84 수준 - 통계청 저위
초혼율 '+3.0%/연'	빠른 증가(적극적 정책 효과)	합계출산율 1.08 수준 - 통계청 중위

주: 통계청 시나리오는 장기 수렴 기준을 적용(장래인구추계: 2022~2072년).
자료: 저자 작성.

1) 초혼율 유지 시나리오

앞서 3장에서 언급하였듯이 초혼율 유지 시나리오는 시나리오 이름 그대로 초혼율이 현재 수준을 유지하는 가정이다. 이는 저출산 대응 정책이 혼인 증가로 이어지지 못하거나, 사회·경제적 여건 악화로 혼인이 더욱 지연·회피되는 경우를 상정한다. 이 시나리오에서는 유배우 여성 비율이 현재 수준에서 정체되거나 완만히 감소하므로, 출산율 회복이 제한적이다. 통계청의 합계출산율 현 수준 유지 시나리오는 2023년부터 0.78 수준을 유지하는 것을 가정하지만 본 연구의 초혼율 유지 시나리오는 상대적으로 더 비관적인 시나리오로 설계되었다.

2) 초혼율 '+1.5%/연' 시나리오

초혼율 '+1.5%/연' 시나리오는 초혼율이 매년 1.5%씩 증가하는 가정이다. 이는 저출산 대응 정책(예: 주거 지원, 육아 지원, 일·가정 양립 확대)이 일정 수준 효과를 발휘하여 혼인이 점진적으로 증가하는 경우를 상정한다. 1.5% 증가율 시나리오와 대응되는 통계청의 저위 시나리오(2025년 0.63에서 장기적으로 0.82 수준까지 상승)와 유사한 경로를 가정한다. 본 연구의 중위 시나리오는 초혼율 증가를 통해 유배우 여성 비율을 높이고, PPR 적용을 통해 유사한 출산율 회복 경로를 달성하도록 설계되었다.

3) 초혼율 '+3.0%/연' 시나리오

초혼율 '+3.0%/연' 시나리오는 초혼율이 매년 3.0%씩 증가하는 가정

이다. 이는 적극적인 저출산 대응 정책과 사회 인식 개선이 결합하여 혼인이 빠르게 증가하는 낙관적 경우를 상정한다. 3.0% 증가율은 정책 효과의 상한선을 고려한 것으로, 실현 가능성은 정책 강도와 사회 변화 속도에 달려 있다. 통계청의 중위 시나리오는 합계출산율이 2025년 0.73에서 장기적으로 1.08 수준까지 상승하는 경로를 가정한다. 본 연구의 고위 시나리오는 초혼율의 빠른 증가를 통해 이와 유사한 출산율 상승을 달성하도록 설계되었지만, 회복하는 시기가 통계청 가정에 비해 더디게 가정되었다고 할 수 있다.

4) 시나리오 간 비교 가능성

본 연구에서 상정한 3개 시나리오는 초혼율 증가 속도만 다를 뿐, 사망력과 이동력은 동일하므로, 시나리오 간 차이는 순수하게 출산력 변화의 효과로 해석할 수 있다. 이러한 비교는 저출산 대응 정책이 혼인율에 미치는 영향을 정량화하고, 정책 강도에 따른 인구 변동의 민감도를 평가하는 데 적절할 수 있다. 다만 본 연구의 PPR 방식은 통계청의 GLG 방식과 구조적으로 다르므로, 동일한 합계출산율 수준을 가정하더라도 연령별 출산율 패턴이나 출산 시기는 다를 수 있다. 이러한 차이는 제2절 3항 "통계청 추계와의 비교"에서 상세히 분석한다.

라. 추계 실행 환경

추계는 Linux 서버 환경(Ubuntu 22.04)에서 Python 3.11 기반으로 실행되었다. 주요 패키지는 NumPy 1.26, Pandas 2.1, scikit-learn 1.3이며, 데이터베이스는 PostgreSQL 15를 사용하였다. 3개 시나리오

는 순차적으로 실행되었다.

추계 결과는 projections/ 폴더에 시나리오별로 저장되었으며, 자동 아카이빙을 통해 latest/ 폴더에는 최신 결과가, archive/ 폴더에는 전체 실행 이력이 보관되었다. 또한 PPR 방식에서 역산된 ASFR은 assumptions/fertility/kihasa_fertility_ppr_2022_2122.xlsx에 시나리오별 시트로 저장되어 재현 가능성을 확보하였다.

2. 주요 추계 결과

가. 총인구 및 인구 규모 변화

1) 총인구 추이

3개 시나리오의 총인구는 2023년 5,169만 명에서 출발하여 2050년 4,507~4,625만 명, 2070년 3,364~3,616만 명, 2120년 1,238~1,803만 명으로 감소한다. 시나리오 간 차이는 2050년 118만 명(2.6%), 2070년 252만 명(7.5%), 2120년 565만 명(45.7%)으로 시간이 경과할수록 확대된다. 인구성장률은 초혼율 유지 시나리오 기준 2040년 -0.59%, 2060년 -1.47%, 2090년 -2.12%로 악화되었다가 2120년 -1.58%로 소폭 완화된다. 시나리오 간 인구성장률 차이는 2070년 0.31%p(초혼율 유지 -1.70%, 초혼율 '+3.0%/연' -1.39%), 2120년 0.73%p(초혼율 유지 -1.58%, 초혼율 '+3.0%/연' -0.85%)이다.

〈표 5-4〉 총인구 추계 결과

(단위: 천 명)

연도	초혼율 유지			초혼율 '+1.5%/연'			초혼율 '+3.0%/연'		
	전체	남	여	전체	남	여	전체	남	여
2023	51,692	25,821	25,871	51,693	25,821	25,871	51,693	25,821	25,872
2030	51,067	25,478	25,589	51,113	25,502	25,611	51,162	25,527	25,635
2040	48,912	24,349	24,563	49,162	24,477	24,685	49,426	24,612	24,814
2050	45,070	22,386	22,685	45,642	22,679	22,963	46,253	22,992	23,261
2060	39,564	19,699	19,864	40,455	20,156	20,298	41,415	20,649	20,766
2070	33,642	16,852	16,790	34,832	17,463	17,369	36,157	18,143	18,014
2080	28,121	14,154	13,966	29,622	14,926	14,696	31,355	15,816	15,539
2090	22,992	11,573	11,419	24,785	12,495	12,290	26,909	13,586	13,323
2100	18,454	9,313	9,140	20,499	10,365	10,134	22,982	11,642	11,340
2110	14,892	7,555	7,337	17,175	8,729	8,446	20,021	10,193	9,828
2120	12,378	6,344	6,034	14,861	7,619	7,242	18,032	9,247	8,785

자료: 저자 산출.

2) 성비 및 성별 인구 구성

전체 성비(여성 100명당 남성 수)는 2023년 99.8에서 2050년 98.798.8로 소폭 하락한 후 지속 상승하여 2070년 100.4~100.7, 2120년 105.1~105.3으로 초기 가정(출생 성비 105) 수준에 수렴한다. 시나리오 간 성비 차이는 0.20.4로 미미하다.

성별 인구 구성을 보면, 2023년 남성 2,582만 명(49.9%), 여성 2,587만 명(50.1%)에서 2120년 초혼율 유지시나리오는 남성 634만 명(51.2%), 여성 603만 명(48.8%), 초혼율 '+3.0%/연' 시나리오는 남성 925만 명(51.3%), 여성 879만 명(48.7%)로 남성 비율이 증가한다. 시나리오 간 성별 구성비 차이는 0.1%p 이내로 거의 없다.

〈표 5-5〉 인구성장률 및 성비 추계 결과

(단위: %)

연도	초혼율 유지		초혼율 '+1.5%/연'		초혼율 '+3.0%/연'	
	인구성장률	성비	인구성장률	성비	인구성장률	성비
2023	0.12	99.8	0.13	99.8	0.13	99.8
2030	-0.25	99.6	-0.23	99.6	-0.21	99.6
2040	-0.59	99.1	-0.53	99.2	-0.47	99.2
2050	-1.03	98.7	-0.94	98.8	-0.85	98.8
2060	-1.47	99.2	-1.37	99.3	-1.26	99.4
2070	-1.70	100.4	-1.55	100.5	-1.39	100.7
2080	-1.85	101.3	-1.65	101.6	-1.44	101.8
2090	-2.12	101.3	-1.87	101.7	-1.59	102.0
2100	-2.16	101.9	-1.83	102.3	-1.49	102.7
2110	-2.08	103.0	-1.68	103.4	-1.27	103.7
2120	-1.58	105.1	-1.21	105.2	-0.85	105.3

자료: 저자 산출.

나. 연령 구조 변화

유소년 인구(0-14세) 비율은 2023년 11.0%에서 2040년 5.7~6.7%, 2080년 3.4~6.5%로 감소한다. 2120년에는 초혼율 유지 4.1%, 초혼율 '+1.5%/연' 5.7%, 초혼율 '+3.0%/연' 7.7%로 시나리오 간 격차가 3.6%p에 달한다.

생산가능인구(15-64세) 비율은 2023년 70.6%에서 2040년 58.5~59.1%, 2080년 39.242.1%로 급감한다. 2120년에는 초혼율 유지 43.7%, 초혼율 '+1.5%/연' 46.0%, 초혼율 '+3.0%/연' 48.3%로 회복된다. 시나리오 간 차이는 2040년 0.6%p, 2080년 2.9%p, 2120년 4.6%p이다.

고령 인구(65세 이상) 비율은 2023년 18.4%에서 2040년 34.8~35.2%, 2080년 51.4~57.4%로 상승한다. 2120년에는 초혼율 유지 52.2%, 초

혼율 '+1.5%/연' 48.3%, 초혼율 '+3.0%/연' 44.0%로 시나리오 간 격차가 8.2%p로 확대된다.

노인부양비는 2023년 26.0에서 2040년 59.6, 2080년 122.2~146.1까지 상승하며, 2120년에는 초혼율 유지 119.3, 초혼율 '+1.5%/연' 104.9, 초혼율 '+3.0%/연' 91.1로 나타난다. 중위연령은 2023년 45.0세에서 2080년 65.0~69.0세로 상승한 후 2120년 59.0~66.0세로 변화한다.

〈표 5-6〉 연령대별 인구 구성 추계 결과

(단위: %, 세)

시나리오	연도	0-14세	15-64세	65세 이상	노인부양비	중위연령
초혼율 유지	2023	11.0	70.6	18.4	26.0	45.0
	2040	5.7	59.1	35.2	59.6	55.0
	2080	3.4	39.2	57.4	146.1	69.0
	2120	4.1	43.7	52.2	119.3	66.0
초혼율 '+1.5%/연'	2023	11.0	70.6	18.4	26.0	45.0
	2040	6.2	58.8	35.0	59.6	55.0
	2080	4.8	40.7	54.5	133.6	67.0
	2120	5.7	46.0	48.3	104.9	63.0
초혼율 '+3.0%/연'	2023	11.0	70.6	18.4	26.0	45.0
	2040	6.7	58.5	34.8	59.6	55.0
	2080	6.5	42.1	51.4	122.2	65.0
	2120	7.7	48.3	44.0	91.1	59.0

주: 노인부양비는 15-64세 인구 대비 65세 이상 인구 비율을 의미함.
자료: 저자 산출.

다. 혼인상태별 인구 구성

가임기 여성(15~49세)의 혼인상태별 구성은 초혼율 증가 시나리오에 따라 뚜렷한 차이를 보인다. 특히 30대에서 시나리오 간 격차가 두드러진다.

30~34세 여성의 유배우 비율은 초혼율 유지 시나리오에서 2023년 12.7%에서 2050년 12.0%로 거의 변화가 없으나, 초혼율 '+1.5%/연' 시나리오는 12.9%에서 17.2%로, 초혼율 '+3.0%/연' 시나리오는 13.1%에서 24.1%로 상승한다. 2050년 기준 시나리오 간 격차는 초혼율 유지 12.0%, 초혼율 '+1.5%/연' 17.2%, 초혼율 '+3.0%/연' 24.1%로 12.1%p에 달한다. 미혼 비율은 반대 방향으로 변화하여, 2050년 초혼율 유지 87.3%, 초혼율 '+1.5%/연' 81.7%, 초혼율 '+3.0%/연' 74.3%이다.

35-39세 여성의 유배우 비율은 초혼율 유지 시나리오에서 2023년 45.4%에서 2050년 40.3%로 감소하는 반면, 초혼율 '+1.5%/연'는 45.7%에서 52.5%로, 초혼율 '+3.0%/연'는 46.0%에서 64.9%로 상승한다. 2050년 시나리오 간 격차는 24.6%p(초혼율 유지 40.3%, 초혼율 '+3.0%/연' 64.9%)이다. 미혼 비율은 2050년 초혼율 유지 57.8%, 초혼율 '+1.5%/연' 45.0%, 고위 31.8%로 나타난다.

40-44세 여성의 유배우 비율은 2050년 초혼율 유지 56.7%, 초혼율 '+1.5%/연' 68.5%, 초혼율 '+3.0%/연' 78.6%로 시나리오 간 격차가 21.9%p이다. 45-49세 여성은 2050년 초혼율 유지 60.8%, 초혼율 '+1.5%/연' 70.3%, 초혼율 '+3.0%/연' 78.4%로 격차가 17.6%p이다.

25-29세 이하 연령에서는 시나리오 간 차이가 미미하다. 25-29세 유배우 비율이 2050년 초혼율 유지 1.6%, 초혼율 '+1.5%/연' 2.5%, 초혼율 '+3.0%/연' 3.7%로 모두 5% 미만이며, 15-24세는 전 시나리오에서 0.1~0.2% 수준이다.

사별과 이혼 비율은 가임기에서 낮은 수준(각 0~7%)을 유지하며, 시나리오 간 차이는 거의 없다.

〈표 5-7〉 가임기 여성의 혼인상태별 구성비 추계 결과 - 초혼율 유지

(단위: %)

연령	연도	미혼	유배우	사별	이혼	계
15-24세	2023	99.9	0.1	0.0	0.0	100.0
	2030	99.9	0.1	0.0	0.0	100.0
	2040	99.9	0.1	0.0	0.0	100.0
	2050	99.9	0.1	0.0	0.0	100.0
25-29세	2023	98.2	1.6	0.0	0.2	100.0
	2030	98.1	1.6	0.0	0.2	100.0
	2040	98.0	1.7	0.0	0.2	100.0
	2050	98.1	1.6	0.0	0.2	100.0
30-34세	2023	86.5	12.7	0.1	0.7	100.0
	2030	87.1	12.1	0.1	0.7	100.0
	2040	87.5	11.8	0.1	0.7	100.0
	2050	87.3	12.0	0.1	0.7	100.0
35-39세	2023	52.4	45.4	0.2	2.1	100.0
	2030	58.5	39.6	0.1	1.7	100.0
	2040	58.8	39.4	0.1	1.7	100.0
	2050	57.8	40.3	0.2	1.7	100.0
40-44세	2023	27.5	68.0	0.4	4.1	100.0
	2030	37.2	58.9	0.4	3.6	100.0
	2040	39.5	56.8	0.4	3.4	100.0
	2050	39.6	56.7	0.4	3.3	100.0
45-49세	2023	17.7	75.1	0.7	6.5	100.0
	2030	25.8	67.3	0.7	6.2	100.0
	2040	33.2	60.8	0.6	5.4	100.0
	2050	33.3	60.8	0.6	5.3	100.0

자료: 저자 산출.

〈표 5-8〉 가임기 여성의 혼인상태별 구성비 추계 결과 - 초혼율 '+1.5%/연'

(단위: %)

연령	연도	미혼	유배우	사별	이혼	계
15-24세	2023	99.9	0.1	0.0	0.0	100.0
	2030	99.9	0.1	0.0	0.0	100.0
	2040	99.9	0.1	0.0	0.0	100.0
	2050	99.9	0.1	0.0	0.0	100.0
25-29세	2023	98.1	1.6	0.0	0.2	100.0
	2030	97.9	1.8	0.0	0.2	100.0
	2040	97.4	2.3	0.0	0.3	100.0
	2050	97.1	2.5	0.0	0.3	100.0
30-34세	2023	86.3	12.9	0.1	0.7	100.0
	2030	85.6	13.5	0.1	0.7	100.0
	2040	84.1	15.0	0.1	0.9	100.0
	2050	81.7	17.2	0.1	1.0	100.0
35-39세	2023	52.1	45.7	0.2	2.1	100.0
	2030	55.4	42.7	0.2	1.8	100.0
	2040	51.0	46.9	0.2	2.0	100.0
	2050	45.0	52.5	0.2	2.3	100.0
40-44세	2023	27.4	68.1	0.4	4.1	100.0
	2030	34.9	61.0	0.4	3.7	100.0
	2040	31.9	63.9	0.4	3.8	100.0
	2050	27.0	68.5	0.4	4.1	100.0
45-49세	2023	17.7	75.1	0.7	6.5	100.0
	2030	24.9	68.2	0.7	6.2	100.0
	2040	27.6	65.9	0.7	5.8	100.0
	2050	22.7	70.3	0.7	6.2	100.0

자료: 저자 산출.

〈표 5-9〉 가임기 여성의 혼인상태별 구성비 추계 결과 - 초혼율 '+3.0%/연'

(단위: %)

연령	연도	미혼	유배우	사별	이혼	계
15-24세	2023	99.9	0.1	0.0	0.0	100.0
	2030	99.9	0.1	0.0	0.0	100.0
	2040	99.8	0.2	0.0	0.0	100.0
	2050	99.8	0.2	0.0	0.0	100.0
25-29세	2023	98.1	1.7	0.0	0.2	100.0
	2030	97.6	2.1	0.0	0.3	100.0
	2040	96.6	3.0	0.1	0.4	100.0
	2050	95.7	3.7	0.1	0.5	100.0
30-34세	2023	86.1	13.1	0.1	0.7	100.0
	2030	84.1	15.0	0.1	0.8	100.0
	2040	80.0	18.8	0.1	1.1	100.0
	2050	74.3	24.1	0.1	1.4	100.0
35-39세	2023	51.8	46.0	0.2	2.1	100.0
	2030	52.1	45.8	0.2	1.9	100.0
	2040	42.7	54.7	0.2	2.4	100.0
	2050	31.8	64.9	0.3	3.0	100.0
40-44세	2023	27.3	68.2	0.4	4.1	100.0
	2030	32.6	63.3	0.4	3.8	100.0
	2040	24.7	70.6	0.4	4.2	100.0
	2050	15.9	78.6	0.5	4.9	100.0
45-49세	2023	17.6	75.2	0.7	6.5	100.0
	2030	24.0	69.0	0.7	6.3	100.0
	2040	22.2	70.9	0.7	6.2	100.0
	2050	13.7	78.4	0.8	7.1	100.0

자료: 저자 산출.

라. 출생아수 및 합계출산율

출생아수는 2023년 24.3~24.4만 명에서 2030년 20.4~22.9만 명, 2050년 12.8~20.2만 명으로 감소한다. 2120년에는 초혼율 유지 2.9만

명, 초혼율 '+1.5%/연' 5.1만 명, 초혼율 '+3.0%/연' 8.5만 명으로 시나리오 간 격차가 5.6만 명에 달한다.

합계출산율(TFR)은 초혼율 유지 시나리오에서 2030년 0.63까지 하락한 후 0.62 수준에서 고정된다. 초혼율 '+1.5%/연' 시나리오는 2040년 0.73에서 점진적으로 상승하여 2060년 이후 0.84를 유지한다. 초혼율 '+3.0%/연' 시나리오는 2040년 0.85, 2050년 1.00을 거쳐 2060년 이후 1.08로 수렴한다.

시나리오 간 TFR 격차는 2030년 0.08(초혼율 유지 0.63, 초혼율 '+3.0%/연' 0.71), 2040년 0.23(초혼율 유지 0.62, 초혼율 '+3.0%/연' 0.85), 2050년 이후 0.46(초혼율 유지 0.62, 초혼율 '+3.0%/연' 1.08)로 확대된다. 초혼율 '+3.0%/연' 시나리오에서는 2050년 TFR 1.00로 상승하지만, 통계청 중위 수준(1.08)에도 미치지 못하는 수준이다.

〈표 5-10〉 출생아수 및 합계출산율 추계 결과

(단위: 천 명)

연도	초혼율 유지		초혼율 '+1.5%/연'		초혼율 '+3.0%/연'	
	출생아수	TFR	출생아수	TFR	출생아수	TFR
2023	243	0.74	244	0.74	244	0.74
2030	204	0.63	216	0.67	229	0.71
2040	154	0.62	181	0.73	210	0.85
2050	128	0.62	163	0.80	202	1.00
2060	80	0.62	111	0.84	147	1.08
2070	65	0.62	97	0.84	137	1.08
2080	55	0.62	88	0.84	130	1.08
2090	43	0.62	71	0.84	108	1.08
2100	36	0.62	61	0.84	98	1.08
2110	32	0.62	56	0.84	94	1.08
2120	29	0.62	51	0.84	85	1.08

자료: 저자 산출.

마. 사망자수 및 기대수명

사망자수는 2023년 33.3만 명(남 18.0만 명, 여 15.3만 명)에서 2060년 73.9만 명까지 증가한 후 점차 감소하여 2120년 28.5~29.7만 명 수준이 된다. 시나리오 간 사망자수 차이는 인구 규모 차이에서 발생하며, 2120년 초혼율 유지 28.5만 명, 초혼율 '+1.5%/연' 29.1만 명, 초혼율 '+3.0%/연' 29.7만 명으로 1.2만 명 차이를 보인다.

성별 사망자수는 2023~2060년에는 남성이 더 많으나(2060년 남 36.3만 명, 여 37.6만 명), 2070년 이후에는 비슷한 수준을 유지한다. 2120년에는 초혼율 유지 시나리오 기준 남 14.1만 명, 여 14.4만 명으로 거의 균형을 이룬다.

〈표 5-11〉 연간 사망자수 추계 결과

(단위: 천 명)

연도	초혼율 유지			초혼율 '+1.5%/연'			초혼율 '+3.0%/연'		
	전체	남	여	전체	남	여	전체	남	여
2023	333	180	153	333	180	153	333	180	153
2030	400	211	189	400	211	189	400	211	189
2040	512	269	243	512	269	244	512	269	244
2050	664	337	326	664	337	327	664	337	327
2060	739	363	376	739	363	376	739	363	376
2070	710	348	362	710	348	362	710	348	362
2080	647	325	322	647	325	322	648	325	323
2090	602	306	296	602	306	296	602	306	296
2100	502	251	251	502	252	251	503	252	251
2110	407	202	205	408	203	205	409	203	205
2120	285	141	144	291	145	146	297	149	148

자료: 저자 산출.

〈표 5-12〉 기대수명 추계 결과

(단위: 세)

연도	전체	남	여	성별 격차
2023	83.9	80.9	86.8	5.9
2030	85.4	82.7	88.1	5.4
2040	87.2	84.9	89.6	4.7
2050	88.8	86.7	90.8	4.1
2060	90.1	88.2	91.9	3.7
2070	91.2	89.5	92.8	3.3
2080	92.2	90.6	93.6	3.0
2090	93.0	91.6	94.3	2.7
2100	93.8	92.5	95.0	2.5
2110	94.4	93.2	95.6	2.3
2120	95.1	93.9	96.1	2.2

자료: 저자 산출.

0세 기대수명은 2023년 83.9세(남 80.9세, 여 86.8세)에서 지속적으로 상승하여 2070년 91.2세(남 89.5세, 여 92.8세), 2090년 93.0세(남 91.6세, 여 94.3세)에 달한다. 성별 격차는 2023년 5.9세에서 2090년 2.7세로 축소된다. 사망력 가정이 3개 시나리오 공통이므로 기대수명은 동일하다.

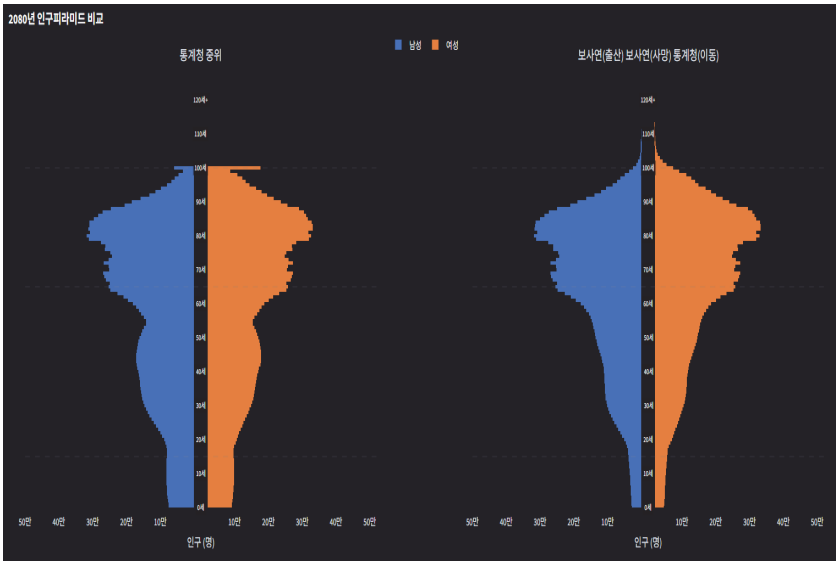
바. 100세 이상 인구

한계연령을 130세로 연장한 결과, 100세 이상 초고령층 인구의 연령별 분포를 세밀하게 추적할 수 있게 되었다. [그림 5-1]은 2023년과 2080년의 인구 피라미드를 비교한 것으로, 좌측은 한계연령 100세 설정 시, 우측은 130세 연장 시의 결과를 보여준다. 한계연령 100세 설정하에서는 100세 이상 인구가 단일 연령대에 집계되어 인구 피라미드 상단이 'T자형'으로 급격히 돌출되는 형태로 나타난다. 이는 100세 이상 생존자를 모두 100세로 간주하여 실제 연령 분포를 반영하지 못하는 한계를 시

각적으로 보여준다.

반면 한계연령을 130세로 연장하면 100세 이상 인구가 100세, 101세, 102세 등으로 세분화되어 피라미드 상단이 부드럽게 감소하는 형태로 나타난다. 2023년에는 100세 이상 인구 규모가 작아 두 방법 간 차이가 크지 않지만, 2080년에는 100세 이상 인구가 약 50만 명에 달하면서 한계연령 연장의 효과가 뚜렷하게 나타난다. 100세부터 105세까지 구간에만 수만 명 단위의 인구가 분포하고, 110세 이상에도 수천 명 규모의 인구가 존재하여, 초고령층 내부의 연령 구조를 현실적으로 반영할 수 있게 된다. 이러한 세밀한 연령 분포 정보는 100세 이상 대상 의료·요양·돌봄 정책을 연령대별로 차별화하여 설계하는 데 필수적인 기초 자료로 활용될 수 있다.

〔그림 5-1〕 인구피라미드 비교((좌) 통계청, (우) 본 연구: 2080년)



자료: 저자 작성.

3. 통계청 추계와의 비교

가. 비교 범위 및 방법

본 연구의 결과 비교를 위해 본 연구의 초혼율 '+1.5%/연' 시나리오와 통계청의 「장래인구추계: 2022~2072년」 중위 시나리오를 비교하였다. 비교 기간은 통계청 추계가 종료되는 2112년까지이며, 주요 지표는 총인구, 연령 구조, 출생아수, 합계출산율, 기대수명이다. 두 추계의 주요 차이점은 다음과 같다.

〈표 5-13〉 인구 추계 주요 가정 차이

구분	본 연구	통계청 중위
출산력	유배우 여성 출산율 가정	연령별 출산율 가정
사망력	LLG 확장모형	LLG 확장모형
한계연령	130세(Coal-Kisker)	110세(Coal-Kisker), 100+세 제공
혼인상태	적용	적용X

자료: 통계청(2023) '장래인구추계: 2022~2072년', 본 연구는 저자 산출.

나. 총인구 비교

본 연구의 초혼율 '+1.5%/연' 시나리오의 총인구는 통계청 중위 시나리오 대비 전 기간 낮게 전망되었다. 2023년 차이는 2만 명(-0.0%)으로 미미하나, 2040년 90만 명(-1.8%), 2070년 235만 명(-6.3%), 2120년 480만 명(-24.4%)으로 격차가 확대된다. 2050년까지는 차이가 147만 명(-3.1%) 수준이지만, 2060년 이후 격차 확대 속도가 빨라진다. 이는 PPR 방식의 출산율 회복 속도가 통계청 GLG 방식보다 느리고, 가임기 여성 인구 감소가 누적되면서 장기적으로 출생아수 차이가 확대되기 때문으로 해석된다.

〈표 5-14〉 총인구 추계 결과 비교

(단위: 천 명)

연도	본 연구(혼인율 '+1.5%/연')	통계청(중위)	차이(본 연구-통계청)
2023	51,693	51,713	-20
2030	51,113	51,306	-193
2040	49,162	50,059	-897
2050	45,642	47,107	-1,465
2060	40,455	42,302	-1,847
2070	34,832	37,182	-2,350
2080	29,622	32,567	-2,945
2090	24,785	28,208	-3,423
2100	20,499	24,399	-3,900
2110	17,175	21,565	-4,390
2120	14,861	19,657	-4,796

자료: 통계청(2023) '장래인구추계: 2022~2072년'; 본 연구는 저자 산출.

다. 연령 구조 비교

유소년 인구(0-14세) 비율은 본 연구가 통계청보다 일관되게 낮다. 2040년 본 연구 6.2%, 통계청 7.7%로 1.5%p 차이가 나며, 2072년에는 2.0%p 차이(본 연구 4.6%, 통계청 6.6%)로 확대된다. 이는 본 연구의 출산율이 통계청 대비 낮은 수준에서 회복되는 경로를 반영한다.

생산가능인구(15-64세) 비율은 2040년 본 연구 58.8%, 통계청 58.0%로 0.8%p 높으나, 2072년에는 2.0%p 낮아진다(본 연구 43.8%, 통계청 45.8%). 이는 유소년 인구 차이가 시차를 두고 생산가능인구에 반영되는 효과이다.

고령 인구(65세 이상) 비율은 2040년 본 연구 35.0%, 통계청 34.3%로 유사하나, 2072년에는 본 연구 51.5%, 통계청 47.7%로 3.8%p 높다. 노인부양비(생산가능인구 100명당 고령 인구)도 2072년 본 연구 117.6, 통계청 104.2로 13.4p 차이가 나며, 이는 본 연구의 생산가능인구 비율

이 더 낮기 때문이다.

중위연령은 2072년 본 연구 66.0세, 통계청 63.4세로 2.6세 차이를 보인다.

〈표 5-15〉 연령대별 인구 구성 추계 결과 비교

(단위: %, %p, 세)

시나리오	연도	0-14세	15-64세	65 세이상	노인부양비	중위연령
본 연구(혼인율 '+1.5%/연')	2023	11.0	70.6	18.4	26.0	45.0
	2040	6.2	58.8	35.0	59.6	55.0
	2072	4.6	43.8	51.5	117.6	66.0
통계청(중위)	2023	11.0	70.7	18.2	25.8	45.5
	2040	7.7	58.0	34.3	59.1	54.6
	2072	6.6	45.8	47.7	104.2	63.4
차이 (본 연구-통계청)	2023	0.0	-0.1	0.2	0.2	-0.5
	2040	-1.5	0.8	0.7	0.5	0.4
	2172	-2.0	-2.0	3.8	13.4	2.6

주: 노인부양비는 15-64세 인구 대비 65세 이상 인구 비율을 의미함.
자료: 통계청(2023) '장래인구추계: 2022~2072년'; 본 연구는 저자 산출.

라. 출산력 비교

출생아수는 2023년 본 연구 24.4만 명, 통계청 23.0만 명으로 본 연구가 1.4만 명 많으나, 2030년 이후 역전되어 KIHASA가 지속적으로 적다. 2030년 차이는 4.7만 명(-17.9%), 2040년 7.8만 명(-30.1%), 2070년 6.5만 명(-40.1%)이며, 2120년에는 4.0만 명(-44.0%) 차이로 다소 축소된다.

합계출산율(TFR)은 2023년 본 연구 0.74, 통계청 0.72로 유사하나, 2030년 이후 본 연구가 낮다. 2030년 본 연구 0.67, 통계청 0.82로 0.15 차이가 나며, 2040년에는 0.32 차이(본 연구 0.73, 통계청 1.05)로 확대된다. 2050년 이후 격차는 0.24~0.28 수준으로 유지된다.

〈표 5-16〉 출생아수 추계 결과 비교

(단위: 천 명)

연도	본 연구(혼인율 '+1.5%/연')	통계청(중위)	차이(본 연구-통계청)
2023	244	230	14
2030	216	263	-47
2040	181	259	-78
2050	163	212	-49
2060	111	156	-45
2070	97	162	-65
2080	88	144	-56
2090	71	116	-45
2100	61	108	-47
2110	56	103	-47
2120	51	91	-40

자료: 통계청(2023) '장래인구추계: 2022~2072년', 본 연구는 저자 산출.

〈표 5-17〉 합계출산율 추계 결과 비교

(단위: 세)

연도	본 연구(혼인율 '+1.5%/연')	통계청(중위)	차이(본 연구-통계청)
2023	0.74	0.72	0.20
2030	0.67	0.82	-0.15
2040	0.73	1.05	-0.32
2050	0.80	1.08	-0.28
2060	0.84	1.08	-0.24

자료: 통계청(2023) '장래인구추계: 2022~2072년', 본 연구는 저자 산출.

이는 PPR 방식의 초혼율 연 1.5% 증가 가정이 통계청의 ASFR 직접 가정보다 보수적인 출산율 회복 경로를 제시하기 때문이다. 특히 2030~2050년 기간의 차이가 크며, 이는 두 추계의 출산력 가정 설정 방식과 수준의 차이를 반영한다.

마. 사망력 비교

0세 기대수명은 본 연구와 통계청 모두 LLG-E 모형을 사용하여 매우 유사한 추이를 보인다. 2030~2040년에는 차이가 0.1세 이내로 거의 일치하나, 2050년 이후 본 연구가 0.20.5세 높게 나타난다.

2023년 기대수명은 본 연구 전체 83.9세(남 80.9세, 여 86.8세), 통계청 전체 83.5세(남 80.6세, 여 86.4세)로 KIHASA가 0.4세 높다. 2070년에는 본 연구 91.2세(남 89.5세, 여 92.8세), 통계청 90.9세(남 89.2세, 여 92.5세)로 0.3세 차이를 보이며, 2120년에는 0.5세(남 0.4세, 여 0.4세) 차이로 소폭 확대된다.

이러한 차이는 한계연령 연장(본 연구 130세, 통계청 100세)과 생명표 연장 방법(Coale-Kisker)의 효과로 해석된다. 그러나 0세 기대수명 기준으로는 영향이 0.3~0.5세 수준으로 제한적이다.

성별 격차는 양 추계 모두 2023년 약 5.85.9세에서 2120년 약 2.22.4세로 축소되는 유사한 패턴을 보인다

〈표 5-18〉 기대수명 추계 결과 비교

(단위: 세)

연도	본 연구(혼인율 '+1.5%/연')			통계청(중위)			차이(본 연구-통계청)		
	전체	남	여	전체	남	여	전체	남	여
2030	85.4	82.7	88.1	85.5	82.8	88.1	-0.1	-0.1	0.0
2040	87.2	84.9	89.6	87.2	84.8	89.5	0.0	0.1	0.1
2050	88.8	86.7	90.8	88.6	86.5	90.7	0.2	0.2	0.1
2060	90.1	88.2	91.9	89.8	88.0	91.7	0.3	0.2	0.2
2070	91.2	89.5	92.8	90.9	89.2	92..5	0.3	0.3	0.3
2080	92.2	90.6	93.6	91.8	90.3	93.3	0.4	0.3	0.3
2090	93.0	91.6	94.3	92.6	91.3	94.0	0.4	0.3	0.3
2100	93.8	92.5	95.0	93.4	92.1	94.7	0.4	0.4	0.3
2110	94.4	93.2	95.6	94.0	92.9	95.2	0.4	0.3	0.4
2120	95.1	93.9	96.1	94.6	93.5	95.7	0.5	0.4	0.4

자료: 통계청(2023) '장래인구추계: 2022~2072년', 본 연구는 저자 산출.

바. 종합 평가

본 연구의 초혼율 '+1.5%/연' 시나리오와 통계청 중위 시나리오의 비교 결과, 총인구는 본 연구가 2040년 1.8%, 2070년 6.3%, 2120년 24.4% 낮게 전망되었다. 주요 차이는 출산력 추계에서 발생하며, 본 연구의 TFR이 2030~2060년 기간 0.15~0.32 낮은 수준을 보인다.

연령 구조에서는 본 연구의 유소년 인구 비율이 2040년 1.5%p, 2072년 2.0%p 낮으며, 이는 고령 인구 비율과 노인부양비가 상대적으로 높게 나타나는 요인이다. 2072년 고령 인구 비율은 본 연구 51.5%, 통계청 47.7%로 3.8%p 차이를 보인다.

사망력은 양 추계 모두 LLG-E 모형 기반으로 기대수명 차이가 0.3~0.5세로 미미하다. 한계연령 연장 효과는 0세 기대수명에는 제한적이나, 초고령층 기대여명에서는 보다 명확히 나타날 것으로 예상된다.

본 연구의 PPR 방식은 초혼율 증가를 명시적 정책 변수로 설정하여 정책 효과 평가가 가능한 구조를 제공하나, 통계청 GLG 방식 대비 보수적인 출산율 회복 경로를 제시한다. 두 추계 방식은 출산력 모델링 접근에서 차이가 있으며, 이는 장기 인구 전망의 불확실성을 반영한다.

제3절 KIHASA-SIM 연계 시뮬레이션 체계 설계

1. KIHASA-SIM 연계 방안

2016년 한국보건사회연구원 주관으로 만든 마이크로시뮬레이션 DOSA[19]는 최근 공적연금의 개혁 효과 분석을 위한 동태적 미시 모의실

험 모형 개발로 확대 적용 중에 있다. 본 연구는 코호트요인법 기반의 집계 인구·가구추계 모형과 동태적 마이크로시뮬레이션 모형인 KIHASA SIM 간의 연계 방안을 검토한다. 여기서 '연계'란, 집계 모형이 생산하는 연령·성별 단위의 인구 및 가구 경로를 마이크로시뮬레이션 모형에 외생변수로 투입하여, 동일한 인구·가구 시나리오하에서 노동, 소득, 연금 등의 동태적 결과를 시뮬레이션하도록 만드는 일련의 통합 과정을 의미한다.

코호트요인법 기반의 인구추계 모형은 출생, 사망, 국제이동, 혼인, 이혼 등의 인구학적 사건을 연령·성별 코호트 흐름으로 구성하여 장래 인구 및 가구 구조를 예측한다. 반면, KIHASA SIM은 출생, 혼인, 교육, 노동, 소득, 연금, 사망 등 개인 및 가구가 경험하는 생애 사건을 연도별로 시뮬레이션함으로써, 정책 변화에 따른 미시적 파급 효과를 분석하는 데 중점을 둔다.

이처럼 목적과 기능이 상이한 두 모형을 연계할 경우, 인구 시나리오의 일관성과 정책 효과 분석의 정밀도를 동시에 제고할 수 있다는 장점이 있다. 또한 출산율, 사망률, 혼인율, 건강상태 등 인구학적 가정을 변화시킬 때, 노동시장, 복지재정, 연금재정에 미치는 민감도를 동시에 파악할 수 있다는 점 역시 주요한 이점이다.

한편, 모형 간의 연계는 두 가지 방향성을 가질 수 있다. 양쪽 모형이 서로의 산출물을 상호 투입하거나 참조하는 양방향 형태일 수도 있고, 한 모형의 산출물을 다른 모형에 일방적으로 적용하는 단방향 형태일 수도 있다. 본 연구에서 시도하는 연계는 시뮬레이션 결과의 총량 정합성을 확

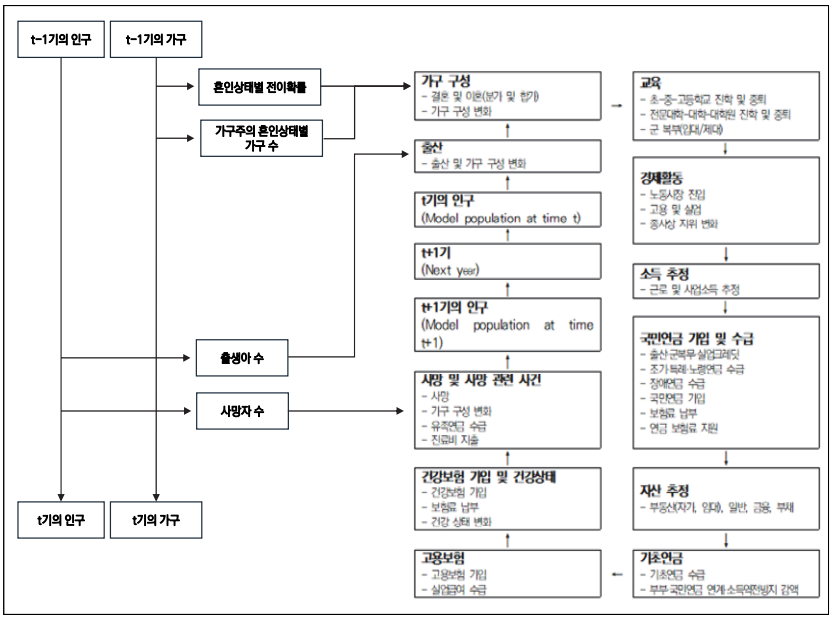
---

19) Dynamic micro-simulation Outlook model for Social policy Analysis (DOSA) DOSA는 개인의 출생에서부터 결혼, 이혼, 자녀출산, 사망과 같은 인구·가구 동태와 교육과정 이수, 경제활동을 통한 소득의 창출과 조세부담, 자산 형성, 사회급여(국민연금, 기초연금 등)의 수급과정까지 국내 관련 모형들 중 가장 포괄적인 범위에서 개별 경제주체들은 물론 제도들 간의 상호작용을 통한 사회경제적 상태변화를 확률적으로 시뮬레이션하는 동태적 모형이다.(한국보건사회연구원, 2016).

보하고, 두 모형이 동일한 인구·가구 시나리오를 기반으로 작동하는지를 검증하는 데 목적을 두고 있다. 따라서 인구 및 가구추계 모형의 산출물을 KIHASA SIM에 외생변수로 투입하는 단방향 연계 방식을 적용한다.

구체적인 연계 절차는 다음과 같다([그림 5-2] 참조). 먼저 인구추계 모형에서 산출된 연령·성별 기준의 출생아와 사망자수 집계값을 KIHASA SIM의 인구모듈과 사망모듈에 각각 외생적으로 주입하여 거시적 집계치를 조정하는 데 활용한다. 또한, 가구추계에 활용된 혼인상태별 전이확률과, 가구추계 결과로 산출된 가구주의 혼인상태별 가구수를 KIHASA SIM의 가구구성 모듈에 외생 입력값으로 적용한다.

[그림 5-2] 통합 인구추계 모형과 KIHASA SIM의 연계(안)



자료: 저자 작성.

2. 주요 연계 변수 및 환경

가. 주요 변수 연계구조 설계

통합 인구추계 모형과 KIHASA-SIM 간의 연계를 위해서는 두 모형 간 공통으로 활용되는 핵심 변수들의 정의와 구조를 일치시키는 것이 필수적이다. 첨부 그림에서 제시된 바와 같이, 연계 변수는 크게 인구학적 기본 변수, 가구 구성 변수, 그리고 정책 시뮬레이션을 위한 확장 변수로 구분된다.

인구학적 기본 변수는 연령, 성별, 혼인상태를 중심으로 구성된다. 통합 인구추계 모형에서 산출되는 '가구 구성'과 '출산' 관련 변수들은 KIHASA-SIM의 인구 모듈에 직접 연계된다. 특히 혼인상태별 전이확률은 양 모형에서 동일하게 적용되어야 하며, 이는 가구 형성과 해체 과정의 일관성을 보장하는 핵심 요소이다.

가구 구성 변수는 가구주의 성별, 연령, 혼인상태와 함께 가구원수, 자녀수 등을 포함한다. 통합 인구추계의 '가구주의 혼인상태별 가구수'는 KIHASA-SIM의 가구구성 모듈과 직접 연계되며, 이를 통해 미시 수준의 가구 생성과 거시 수준의 가구 총량이 정합성을 유지하게 된다.

정책 시뮬레이션을 위한 확장 변수로는 교육 수준, 경제활동 상태, 소득 수준 등이 있다. 이러한 변수들은 통합 인구추계 모형에서는 직접 다루지 않지만, KIHASA-SIM에서 인구학적 변수와 결합하여 노동시장 참여, 연금 가입, 복지 수급 등의 정책 효과를 분석하는 데 활용된다.

나. 모형 간 연계 프로세스

모형 간 연계는 〔그림 5-1〕에서 제시된 바와 같이 세 가지 주요 경로를 통해 이루어진다. 첫째, '혼인상태별 전이확률'은 통합 인구추계 모형에서 KIHASA-SIM으로 직접 전달되어 개인의 혼인상태 변화를 시뮬레이션하는 데 활용된다. 둘째, '가구주의 혼인상태별 가구수'는 KIHASA-SIM의 가구 생성 및 해체 과정을 제어하는 총량 제약으로 작동한다. 셋째, '출생아수'와 '사망자수'는 각각 KIHASA-SIM의 출산 모듈과 사망 모듈에 외생적으로 입력되어 인구 재생산 과정의 거시적 일관성을 확보한다. 이러한 연계 프로세스는 연도별로 반복되며, 각 시점에서 통합 인구추계의 결과가 KIHASA-SIM의 제약 조건으로 작동한다. 이를 통해 미시 시뮬레이션의 확률적 변동성을 유지하면서도 거시적 인구 경로와의 정합성을 확보할 수 있다.

연계 환경 구축을 위해서는 데이터 형식의 표준화가 중요하다. 두 모형이 서로 다른 데이터 구조와 코딩 체계를 사용할 수 있으므로, 중간 변환 계층(interface layer)을 구축하여 데이터 교환의 효율성과 정확성을 확보한다. 또한 연령 구간, 혼인상태 분류, 가구 유형 정의 등 주요 분류 체계를 통일하여 연계 과정에서 발생할 수 있는 불일치를 최소화한다.

한편, 정책 시뮬레이션의 정밀도를 높이기 위해서는 인구학적 변수 외에 건강상태와 노동시장 참여 관련 변수의 추가 연계가 필요하다. 특히 고령화 사회에서 건강수명과 노동수명은 연금재정, 의료비 지출, 노동력 공급 등 주요 정책 이슈와 직결되는 핵심 변수이다. 따라서 통합 인구추계 체계에서 이러한 변수들을 추가로 생성하여 KIHASA-SIM과 연계함으로써, 보다 현실적이고 정책 활용도가 높은 시뮬레이션 결과를 도출할 수 있을 것이다.

3. 건강수명 및 노동수명 연계 방안

가. 건강수명 추계 방법론

2) 주요 방법론

사망률 개선의 결과로 나타나는 건강수명(Healthy Life Expectancy, HALE)은 특정 연령에서 건강한 상태로 기대할 수 있는 평균 생존연수를 의미하고, 질병과 장애로 인한 건강 손실을 반영하여 계산한다. 이것은 연령대별 장애 조정계수(Disability Weight, DW)를 도출하여 건강조정 기대수명을 산정한다. 특히 노인층의 건강수명은 예전과 다른 노인 지위의 질적, 양적 성장을 의미한다. WHO(2024)에서 제시하는 HALE은 식(64)와 같다. DW는 0과 1 사이이며, 0은 완전 건강, 1은 장애로 인해 삶의 질이 완전히 손실된 상태를 의미한다.

$$HALE = LE - \sum_{i=x}^{w}(LE \times DW), \tag{64}$$

i = 장애 시작연령, w = 사망연령(상한연령)

WHO(2024)[20]는 2000~2021년까지 0세와 60세의 건강수명을 제시한다. 2010년 0세의 건강수명은 남자 67.9년, 여자 72.2년, 2015년 남자 69.5년, 여자 73.2년, 2021년 남자 70.7년, 74.1년이다. 통계청은 2017년 생명표부터 2016년 사회조사에 기반을 둔 '유병기간 제외 기대여명(Life expectancy excluding period of illness)과 '주관적 건강평가 기대여명(Subjective health assessment life expectancy)'를 2022년

20) https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-ghe-hale-healthy-life-expectancy-at-age-60.

까지 식 (65)을 적용하여 공표해 오고 있다.

$$HALE = \frac{1}{l_x}\sum_{i=x}^{w}(L_i \times (1-\pi_i)),\ \pi_i : i\text{연령대 장애유병률} \tag{65}$$
$$= \frac{T_x}{l_x} - \frac{\sum_{i=x}^{w}(L_i\pi_i)}{l_x} = LE - \frac{\sum_{i=x}^{w}(L_i\pi_i)}{l_x}$$

식 (64)과 (65)를 비교하면 거의 유사한 구조를 보이며, 이는 0세의 기대수명과 연령별 기대여명에서 장애유병률을 차감하여 건강수명(HALE)을 산출하는 방식이다. 따라서 건강수명을 계산하는 핵심은 연령별 장애유병률을 정확히 추정하는 데 있다. 통계청은 이를 위해 매년 사회조사에서 수집한 유병기간과 건강평가 자료를 활용하고 있다. KOSIS 생명표는 2012년부터 2022년까지 2년 간격으로 5세 단위(0, 1, 5, … , 80, 85세 이상)의 건강수명을 공표하고 있다. 이들 자료를 살펴보면 85세 이상의 건강수명을 요약(abridged)하여 제시하고 있다. 또한 건강수명의 장애유병률 정보가 없다.

따라서 본 연구는 먼저 건강수명과 연령별 기대여명과의 연도별 차이와 식(65)을 활용하여 장애유병률(π_i)을 추정해 본다. 추정방법은 생명표로부터 식 (65)의 LE와 l_x, L_x자료를 참고하여 85세 이하는 84세, 83세, 82세, …, 5세, 1세, 0세의 π_x를 산정한다. 그리고 85세 이상은 85세부터 100세까지의 건강수명이 요약(abridged)된 값이다. 이들 값을 분해하기 위해 100세 건강수명은 0이라는 가정과 85세 이상의 요약 건강수명을 1세별로 전환하는 spline(FMM, periodic, cubic, natural, monoHFC, Hyman, GAM 등)으로 분해해서 적용하였다. 이 방법은 5세별 사망률을 spline 보간법으로 1세 간격으로 확장하는 통계청(2024) 방법과 동일하다. 〈표 5-19〉는 3차(cubic) 선형 결과를 보여준다. 그리

고 5세 간격의 건강수명은 〈표 5-20〉에 제시한다. 〈표 5-19〉와 〈표 5-20〉의 'a'와 'b' 표기는 유병기간 제외 기대여명과 주관적 건강평가 기대여명을 의미한다. 〈표 5-19〉는 연령별 장애 유병률을 제시한 자료로, 100세의 건강수명을 0으로 가정함에 따라 해당 연령의 유병률은 1로 설정되었다. 이러한 가정과 85세 유병률 값을 기준으로 spline 보간법을 활용하여 0세부터 100세까지의 연령별 유병률을 추정하였다. 전체적으로 남성이 여성보다 높은 유병률을 보이나, 해당 수치는 사회조사에 기반을 둔 것이므로 연령 증가나 시간 흐름에 따른 명확한 변화 패턴은 관찰되지 않는다. 이와 같이 산출된 연령별 유병률 값을 생명표 작성 방식과 동일한 로직에 적용하여 산정한 건강수명 결과는 〈표 5-20〉에 제시되어 있다.

〈표 5-19〉 2012~2022년 연령별 장애유병률

연도 / 구분 / 성별 / 연령	2012				2014			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	0.16	0.21	0.14	0.22	0.18	0.23	0.14	0.21
1	0.16	0.21	0.14	0.22	0.18	0.23	0.14	0.21
5	0.16	0.22	0.15	0.23	0.18	0.23	0.15	0.22
10	0.17	0.23	0.16	0.25	0.19	0.24	0.16	0.23
15	0.18	0.24	0.17	0.26	0.20	0.26	0.17	0.25
20	0.20	0.26	0.19	0.28	0.21	0.28	0.18	0.27
25	0.21	0.27	0.20	0.30	0.23	0.30	0.20	0.28
30	0.23	0.29	0.22	0.32	0.25	0.32	0.21	0.30
35	0.25	0.32	0.23	0.35	0.28	0.35	0.23	0.33
40	0.27	0.35	0.26	0.38	0.30	0.38	0.25	0.36
45	0.30	0.38	0.28	0.42	0.34	0.41	0.28	0.39
50	0.34	0.42	0.31	0.46	0.37	0.46	0.30	0.43
55	0.37	0.46	0.34	0.50	0.42	0.50	0.34	0.48
60	0.42	0.50	0.38	0.55	0.47	0.56	0.38	0.52
65	0.45	0.53	0.43	0.58	0.51	0.60	0.42	0.57
70	0.48	0.55	0.48	0.62	0.56	0.63	0.48	0.63
75	0.50	0.57	0.52	0.66	0.59	0.65	0.53	0.65
80	0.52	0.54	0.54	0.66	0.59	0.65	0.57	0.68
85	0.59	0.56	0.60	0.70	0.63	0.68	0.64	0.73

연도 / 구분 / 성별 / 연령	2012				2014			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
90	0.70	0.66	0.70	0.78	0.72	0.76	0.75	0.81
95	0.84	0.82	0.84	0.88	0.85	0.87	0.87	0.90
100+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
연도 / 구분 / 성별 / 연령	2016				2018			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	0.18	0.24	0.13	0.20	0.20	0.24	0.13	0.19
1	0.18	0.24	0.13	0.20	0.20	0.24	0.13	0.20
5	0.19	0.24	0.14	0.21	0.20	0.25	0.14	0.21
10	0.20	0.26	0.15	0.23	0.21	0.27	0.15	0.22
15	0.21	0.27	0.16	0.24	0.23	0.28	0.16	0.24
20	0.23	0.29	0.17	0.26	0.24	0.30	0.17	0.25
25	0.25	0.31	0.19	0.27	0.26	0.32	0.18	0.27
30	0.26	0.33	0.20	0.29	0.28	0.35	0.20	0.29
35	0.29	0.36	0.21	0.31	0.31	0.37	0.22	0.31
40	0.32	0.39	0.23	0.34	0.34	0.40	0.24	0.33
45	0.36	0.43	0.25	0.37	0.37	0.44	0.26	0.36
50	0.39	0.48	0.28	0.41	0.41	0.49	0.28	0.40
55	0.44	0.53	0.31	0.45	0.46	0.53	0.32	0.44
60	0.49	0.58	0.35	0.49	0.51	0.59	0.36	0.48
65	0.54	0.62	0.39	0.54	0.56	0.63	0.40	0.53
70	0.57	0.66	0.43	0.59	0.60	0.67	0.46	0.59
75	0.62	0.68	0.48	0.63	0.65	0.70	0.53	0.63
80	0.64	0.68	0.52	0.65	0.69	0.71	0.58	0.65
85	0.70	0.71	0.60	0.70	0.75	0.75	0.66	0.70
90	0.78	0.79	0.71	0.78	0.83	0.82	0.76	0.78
95	0.88	0.89	0.85	0.88	0.91	0.90	0.87	0.89
100+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

주: a (유병기간 제외 기대여명), b(주관적 건강평가 기대여명).

〈표 5-20〉의 주요 결과에 따르면, 사회조사 자료를 활용해 산출한 유병기간 제외 기대여명은 주관적 건강평가에 기반을 둔 기대여명보다 전반적으로 낮은 수준을 보였다. 이는 주관적 건강평가가 질병에 대한 심리적 낙관성을 반영하는 반면, 유병기간 제외 기대여명은 실제 질병 경험에 기반을 둔 객관적 지표이기 때문이다. 유병기간 제외 기준으로 본 0세의

건강수명은 지난 10여 년간 남성 약 65년, 여성 65~67년 수준에서 큰 변동 없이 유지되고 있다.

〈표 5-19〉 2012~2022년 연령별 장애유병률 〈계속〉

연도 / 구분 / 성별 / 연령	2020				2022			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	0.19	0.22	0.12	0.18	0.19	0.22	0.10	0.15
1	0.19	0.23	0.12	0.18	0.19	0.23	0.10	0.15
5	0.20	0.23	0.13	0.19	0.19	0.23	0.11	0.16
10	0.21	0.25	0.14	0.20	0.20	0.25	0.12	0.17
15	0.22	0.26	0.15	0.21	0.22	0.26	0.13	0.18
20	0.24	0.28	0.16	0.22	0.23	0.28	0.14	0.19
25	0.26	0.30	0.17	0.24	0.25	0.30	0.14	0.20
30	0.28	0.32	0.18	0.26	0.27	0.32	0.15	0.22
35	0.30	0.35	0.19	0.28	0.29	0.35	0.17	0.23
40	0.33	0.38	0.21	0.30	0.32	0.38	0.18	0.26
45	0.36	0.41	0.23	0.33	0.35	0.42	0.20	0.28
50	0.40	0.46	0.26	0.36	0.40	0.45	0.22	0.30
55	0.44	0.50	0.28	0.39	0.44	0.50	0.24	0.34
60	0.49	0.55	0.32	0.44	0.49	0.55	0.27	0.38
65	0.54	0.60	0.36	0.49	0.54	0.61	0.31	0.43
70	0.58	0.64	0.41	0.55	0.60	0.65	0.35	0.48
75	0.60	0.67	0.47	0.60	0.62	0.68	0.40	0.54
80	0.63	0.69	0.54	0.65	0.66	0.69	0.48	0.61
85	0.68	0.74	0.64	0.71	0.72	0.73	0.60	0.70
90	0.77	0.82	0.75	0.80	0.81	0.81	0.72	0.79
95	0.88	0.90	0.87	0.90	0.90	0.90	0.86	0.89
100+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

주: a (유병기간 제외 기대여명), b(주관적 건강평가 기대여명).
자료: 통계청 사회조사 활용 저자 작성.

그리고 주관적 건강평가 기반 기대여명은 남성 66.6년에서 71.6년, 여성 65.8년에서 73.0년으로 점진적 증가세를 나타냈다. 이는 유병기간 제외 기대여명의 개선 속도가 주관적 건강평가 기대여명보다 완만하며, 장애유병률의 개선이 생명표 기대여명이나 사망률 개선 속도와 일치하지 않음을 시사한다. 또한, 85세 이상 인구의 건강수명을 100세까지 추계한 결과, 이 연령대의 건강수명은 약 12년 수준으로 전망된다.

〈표 5-20〉 2012~2022년 연령별 건강수명

연도 / 구분 / 성별 / 연령	2012				2014			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	65.0	66.5	66.6	65.8	64.7	65.7	67.7	67.5
1	64.4	65.8	65.8	65.0	64.1	65.0	66.9	66.7
5	61.0	62.4	61.8	61.1	60.7	61.6	62.9	62.8
10	56.3	57.6	56.9	56.1	56.0	56.9	57.9	57.8
15	51.5	52.8	52.0	51.2	51.1	52.1	53.1	52.8
20	46.7	48.1	47.2	46.4	46.4	47.3	48.3	48.0
25	42.1	43.4	42.5	41.8	41.6	42.5	43.6	43.3
30	37.4	38.7	37.9	37.1	36.8	37.9	38.8	38.7
35	32.7	34.1	33.4	32.4	32.2	33.2	34.3	34.0
40	28.2	29.4	28.9	27.8	27.6	28.6	29.8	29.4
45	23.8	25.0	24.5	23.5	23.3	24.1	25.4	24.9
50	19.7	20.7	20.4	19.2	19.1	19.7	21.3	20.6
55	15.8	16.8	16.6	15.4	15.2	15.7	17.4	16.6
60	12.4	13.1	13.1	11.8	11.7	11.9	13.7	12.8
65	9.4	10.1	9.8	9.0	8.7	9.0	10.4	9.5
70	6.9	7.6	7.0	6.4	6.2	6.6	7.3	6.7
75	5.0	5.6	4.8	4.5	4.4	4.8	5.0	4.8
80	3.5	4.4	3.4	3.2	3.2	3.6	3.3	3.3
85	1.3	1.7	1.1	1.6	1.2	1.5	1.1	1.5
90	0.7	1.1	0.7	0.9	0.7	1.0	0.7	0.9
95	0.3	0.6	0.3	0.4	0.3	0.5	0.3	0.4
100+	0	0	0	0	0	0	0	0

〈표 5-20〉 2012~2022년 연령별 건강수명 〈계속〉

연도 / 구분 / 성별 / 연령	2016				2018			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	64.7	65.2	68.8	68.4	64.0	64.9	69.1	69.0
1	64.0	64.5	68.0	67.5	63.3	64.1	68.4	68.2
5	60.5	61.0	64.0	63.6	59.8	60.6	64.4	64.3
10	55.8	56.3	59.1	58.6	55.1	55.8	59.4	59.3
15	51.0	51.5	54.2	53.7	50.3	51.0	54.5	54.4
20	46.2	46.7	49.4	48.9	45.5	46.2	49.7	49.6
25	41.4	41.9	44.7	44.3	40.8	41.5	45.1	44.9
30	36.8	37.2	40.1	39.7	36.1	36.8	40.3	40.2
35	32.1	32.6	35.6	35.1	31.4	32.2	35.7	35.5
40	27.5	28.0	31.1	30.4	26.9	27.7	31.2	31.0
45	23.0	23.5	26.6	25.9	22.6	23.2	26.8	26.5
50	18.8	19.1	22.4	21.7	18.5	18.9	22.6	22.2
55	14.9	15.0	18.4	17.6	14.6	15.0	18.5	18.1
60	11.5	11.5	14.7	13.9	11.1	11.4	14.7	14.2
65	8.5	8.5	11.3	10.5	8.2	8.5	11.2	10.7
70	6.2	6.1	8.2	7.5	5.9	6.1	8.0	7.5
75	4.2	4.5	5.7	5.2	3.9	4.3	5.2	5.2
80	2.9	3.3	3.9	3.6	2.5	3.0	3.4	3.6
85	1.1	1.3	1.3	1.4	1.2	1.2	1.1	1.5
90	0.6	0.9	0.8	0.7	0.5	0.8	0.6	0.7
95	0.2	0.4	0.4	0.3	0.1	0.4	0.3	0.2
100+	0	0	0	0	0	0	0	0

〈표 5-20〉 2012~2022년 연령별 건강수명 〈계속〉

연도 / 구분 / 성별 / 연령	2020				2022			
	a		b		a		b	
	남자	여자	남자	여자	남자	여자	남자	여자
0	65.6	67.2	70.7	71.3	65.1	66.6	71.6	73.0
1	64.8	66.4	69.9	70.5	64.3	65.7	70.8	72.1
5	60.9	62.6	65.9	66.5	60.7	62.0	66.8	68.1
10	56.1	57.7	60.9	61.5	55.9	57.1	61.9	63.2
15	51.2	52.8	56.1	56.6	51.1	52.3	57.0	58.2
20	46.4	48.0	51.3	51.8	46.3	47.4	52.1	53.3
25	41.6	43.2	46.6	47.1	41.6	42.8	47.4	48.6
30	36.9	38.5	41.9	42.4	36.9	38.1	42.7	43.9
35	32.3	33.9	37.3	37.7	32.3	33.5	38.1	39.3
40	27.8	29.3	32.7	33.2	27.7	28.8	33.6	34.5
45	23.5	24.8	28.2	28.6	23.4	24.3	29.1	30.0
50	19.3	20.4	23.9	24.1	19.1	20.1	24.8	25.6
55	15.4	16.3	19.8	19.9	15.2	16.1	20.6	21.2
60	12.0	12.6	15.9	15.9	11.6	12.3	16.6	17.0
65	8.8	9.4	12.3	12.1	8.5	9.0	12.9	13.1
70	6.5	6.9	9.0	8.7	5.9	6.4	9.5	9.4
75	4.6	4.9	6.1	5.9	4.1	4.5	6.6	6.4
80	3.1	3.3	3.8	3.8	2.7	3.1	4.1	3.9
85	1.2	1.2	1.1	1.4	1.0	1.1	1.3	1.5
90	0.6	0.8	.07	0.7	0.5	0.7	0.7	0.7
95	0.2	0.4	0.3	0.3	0.2	0.4	0.3	0.2
100+	0	0	0	0	0	0	0	0

자료: 통계청 사회조사 활용 저자 작성.

2) 방법론 설계 및 적용

건강수명 추계를 통합 인구추계 모형과 KIHASA-SIM에 연계하기 위해서는 앞서 도출한 연령별 장애유병률을 활용한 체계적인 방법론 설계가 필요하다. 본 연구에서는 다음과 같은 단계별 접근을 통해 건강수명

추계를 구현하고 정책 시뮬레이션에 활용 가능한 형태로 변환한다.

첫째, 장애유병률의 시계열 모형화이다. 〈표 5-19〉에서 추정된 연령별 장애유병률을 기반으로, 향후 장애유병률의 변화를 예측하는 모형을 구축한다. 이를 위해 Lee-Carter 모형의 변형을 적용하여 연령별 장애유병률의 시간적 변화 패턴을 포착한다. 구체적으로는 다음과 같은 모형을 적용한다.

$$\log(\text{장애유병률}[x,t]) = \alpha[x] + \beta[x] \times \kappa[t] + \varepsilon[x,t]$$

여기서 x는 연령, t는 연도, α[x]는 연령별 평균 로그 장애유병률, β[x]는 연령별 변화 민감도, κ[t]는 시간 추세를 나타낸다. 이 모형은 사망률 추계에 사용되는 Lee-Carter 모형과 유사한 구조를 가지므로, 사망률과 장애유병률 간의 일관성 있는 추계가 가능하다.

둘째, 혼인상태별 건강수명 차별화이다. 통합 인구추계 모형의 핵심인 혼인상태별 접근을 건강수명에도 확대 적용한다. 기존 연구들은 혼인상태가 건강 상태와 밀접한 관련이 있음을 보여주고 있으므로, 다음과 같은 상대위험도를 적용한다.

$$\text{장애유병률}[\text{혼인상태}] = \text{장애유병률}[\text{전체}] \times \text{RR}[\text{혼인상태}]$$

여기서 RR은 혼인상태별 상대위험도로, 기혼자를 기준(1.0)으로 할 때 미혼, 이혼, 사별 상태의 상대적 장애 위험을 나타낸다. 이는 4장에서 적용한 사망률의 혼인상태별 보정과 동일한 논리 구조를 따른다.

셋째, 건강수명의 동적 추계 체계 구축이다. 매년 갱신되는 생명표와 장애유병률을 결합하여 건강수명을 동적으로 산출하는 체계를 구축한다. 이를 pseudo-code로 표현하면 다음과 같다.

〈표 5-21〉 건강수명 추계 pseudo-code

건강수명 추계 pseudo 코드

```
def calculate_health_expectancy(year, age, gender, marital_status):
    # 기본 생명표 정보
    life_expectancy = get_life_expectancy(year, age, gender)

    # 장애유병률 추계
    base_disability = project_disability_rate(year, age, gender)

    # 혼인상태별 보정
    marital_adjustment = get_marital_RR(marital_status, age, gender)
    adjusted_disability = base_disability * marital_adjustment

    # 건강수명 계산
    health_expectancy = life_expectancy * (1 - adjusted_disability)

    return health_expectancy
```

자료: 저자 작성.

넷째, KIHASA-SIM 연계를 위한 데이터 구조화이다. 산출된 건강수명은 연령, 성별, 혼인상태별로 구조화하여 KIHASA-SIM의 건강 모듈에 입력 가능한 형태로 변환한다. 이때 건강수명뿐만 아니라 장애유병률 자체도 함께 전달하여, 개인별 건강 상태 전이 확률을 시뮬레이션하는 데 활용할 수 있도록 한다.

다섯째, 정책 시나리오별 건강수명 변화 반영이다. 건강증진 정책, 의료기술 발전, 생활습관 개선 등 다양한 정책 시나리오에 따른 장애유병률 변화를 모형에 반영할 수 있는 구조를 마련한다. 예를 들어, 만성질환 관리 프로그램의 효과를 장애유병률 감소율로 변환하여 적용하거나, 의료기술 발전에 따른 건강수명 연장 효과를 시나리오로 구성할 수 있다.

이러한 방법론을 통해 구축된 건강수명 추계는 단순히 평균적인 건강 상태를 예측하는 데 그치지 않고, 인구 구조 변화와 정책 개입이 국민 건강에 미치는 영향을 종합적으로 평가할 수 있는 기반을 제공한다. 특히 혼인상태별 차별화를 통해 가구 구조 변화가 건강 결과에 미치는 영향을

명시적으로 고려할 수 있으며, 이는 고령화 사회의 건강 정책 수립에 중요한 시사점을 제공할 것이다.

나. 노동수명 추계 방법론 검토

1) 노동수명 추계의 배경과 필요성

최근 건강노동기대여명(Healthy Working Life Expectancy) 또는 건강조정노동생명표(Health-adjusted Working Life Table)에 대한 관심이 높아지고 있다. 이는 고령화가 심화되면서 연령(age)을 기준으로 한 정책 설계에서 건강 역량(health capacity)을 중심으로 한 설계로 정책 패러다임이 전환되고 있기 때문이다(Parker et al., 2020, p. 395). 동일한 연령대에서도 최근 출생코호트의 건강 수준이 과거보다 높아짐에 따라, 연령을 노동능력의 대리변수로 사용하는 데 한계가 커지고 있다. 고령사회에서는 사람들이 얼마나 오랫동안 건강하게 일할 수 있는지를 파악하는 것이 무엇보다 중요하다(Milligan & Wise, 2012, p. 6). 또한 고령화에 따른 재정 압박 속에서 정책 효과를 평가하려는 수요 증가 역시 건강노동기대여명에 대한 관심을 높이는 주요 요인 중 하나다. 사람들이 법정 은퇴연령에 따라 은퇴한다는 가정의 현실성이 낮아지면서, 건강을 고려한 실질적인 은퇴 시점 변화에 대한 관심도 함께 증가하고 있다.

전통적으로 노동수명은 사망이나 은퇴로 인한 이탈을 고려하여 산출되어 왔으며(Wolfbein & Wool, 1950), 퇴출과 재진입을 반영하는 방식(Pollard et al., 1974; Schoen, 1980), 또는 다상태(multi-state) 과정을 고려하는 방식(Willekens, 1980; Schoen & Woodrow, 1980; Schoen, 1988)으로 발전해왔다. 건강수명 측정 역시 Sullivan(1971)이

인구집단의 특정 시점에서의 건강 상태를 반영하는 방법론을 제시한 이후, 시간에 따른 상태 변화를 고려한 다상태 접근 방식으로 발전해왔으며, 이는 동태적 분석과 생애경로 연구에 적합한 틀로 자리잡았다.

그러나 이러한 두 영역의 방법론이 각각 발전해 온 데 반해, 건강과 노동을 통합적으로 고려하려는 시도는 오랜 기간 이루어지지 않았다. 비교적 최근에 이르러서야 Lièvre et al.(2007)이 다상태 틀에 건강지표를 포함함으로써 기대노동수명과 건강조정 노동수명을 동시에 추계할 수 있는 방법론을 제시하였다. 이후 Laaksonen et al.(2022), Epping et al.(2024), Parker et al.(2020), Boissonneault & Rios(2021) 등은 이 방법론을 적용한 연구를 활발히 수행하고 있다.

한편, 우리나라에서는 건강노동기대여명과 관련된 연구가 매우 드물다. 노동기대여명을 산출한 기존 연구들은 대부분 생명표에 경제활동참가율 또는 취업률을 가중하는 방식으로 노동생명표를 작성하였으며, 건강 상태는 고려하지 않았다(황수경, 2005; 박경애·최기홍, 2006; 신경진·최보승, 2013; 양원창 외, 2024).[21] 건강노동기대여명을 직접 산출한 국내 연구는 우해봉(2014)이 유일하다.

이처럼 우리나라에서 건강기대여명에 대한 연구가 부족하지만, 건강기대여명에 대한 필요는 매우 높다. 우리나라처럼 고령화가 급속히 진행되는 상황에서 건강상태를 반영한 노동수명 추계가 체계적으로 이루어지지 않는다면, 고령자의 실제 노동역량을 파악하고 이에 기초한 정책 대응 가능성을 정량적으로 평가하는 데 한계가 발생할 수 있다. 특히, 생산가능인구의 감소에 대응하는 정책 설계에는 건강과 노동이 결합된 정교한 지

21) 김도형(2019)은 Milligan and Wise(2012)의 방법론을 적용하여 고령자의 건강역량(health capacity)과 추가 노동역량(additional work capacity)을 추정하였다. 그러나 이때의 추가 노동역량은 _"얼마나 더 오래 건강하게 일할 수 있는가"를 의미하기보다는, 동일한 건강 수준에도 불구하고 실제로는 경제활동에 참여하지 않는 인구의 규모, 즉 잠재적인 건강역량에 비해 활용되지 않은 노동력 규모를 지칭하는 개념에 더 가깝다.

표 개발이 필수적이다. 본 절에서는 이러한 문제의식에 기반하여, 건강노동기대여명의 추계를 시도한다.

2) 노동수명 추계 방법론

노동생명표 산출 시 가장 널리 사용되는 방식은 사망과 은퇴를 고려한 Wolfbein & Wool(1950)의 방식이다. 이는 사망과 은퇴를 함께 반영하는 이중감소(double decrement) 모형으로, 생명표상의 정지인구(stationary population)에 연령별 경제활동참가율을 곱해 정지 노동인구를 계산하고, 이를 이용해 기대노동수명을 추정한다(Wolfbein & Wool, 1950). 보다 구체적으로는, 연령별 생존자 수 l_x와 연령별 노동시장 참여율 p_x를 곱하여 정지인구 기준의 기대노동연수를 계산하는 방식이다(식 (66) 참조). 이 모형은 주된 일자리 퇴직 후 재취업과 같은 노동시장 내 이동성, 여성의 출산·육아, 군복무, 교육 연장 등의 일시적 이탈과 복귀를 반영하지 못한다는 한계를 지닌다(박경애, 최기홍, 2006).

$$e_x^w = \frac{T_x^w}{l_x} = \frac{\sum_{a=x}^{w} l_a \, p_a}{l_x} \tag{66}$$

여기서, e_x^w는 x세에서 기대되는 향후 노동시장 참여 연수

T_x^w는 x세 이후 총 노동시장 내 인원 연수

이에 따라 노동시장에서의 퇴출과 재진입, 즉 사망에 따른 이탈과 취업에 따른 진입을 동시에 고려하는 증감(increment/decrement) 모형이 등장하였다(Pollard et al., 1974; Schoen & Woodrow, 1980 등). 이 방식은 행동률에 기반한 명확한 전이 규칙을 반영하므로 요약 기대치

(summary measures) 계산 시 발생하는 편향을 최소화할 수 있고, 여성 노동시장 행태를 현실적으로 반영할 수 있으며, 노동력 흐름(flows between statuses)에 대한 상세한 측정이 가능하다는 장점을 지닌다(Schoen & Woodrow, 1980). 대표적으로 Schoen & Woodrow(1980)는 미국의 노동시장 상태를 노동시장 참여, 비참여, 사망의 세 가지 상태로 구분하고, 이들 간 전이를 참여→비참여, 비참여→참여, 참여 또는 비참여→사망으로 설정하였다. 이들은 x세에서의 기대노동기간 e_x^L을 x세 생존자를 기준으로 이후 각 연령 $a\ (x \leq a \leq w)$에서 노동시장 참여 상태에 있을 조건부 확률의 총합으로 정의하였다(식 (67) 참조). 이 모형은 재취업 경로를 부분적으로 반영할 수 있다는 장점이 있으나, 상태를 단지 두 가지(참여/비참여)로 한정함으로써 실업, 가사, 교육 등 다양한 비경제활동 상태를 구분하지 못하고, 전이확률을 상수로 가정한다는 한계를 지닌다.

$$e_x^L = \sum_{a=x}^{w} P(a\text{세의 생존 및 노동시장 참여상태} \mid x\text{세 생존}) \qquad (67)$$

가장 최근에 등장한 방법론은 다상태(multi-state) 모형 또는 마르코프 기반 모형이다. 이는 한 개인이 시간의 흐름에 따라 여러 상태로 전이하는 과정을 고려하여 기대여명을 추정하는 동태적(dynamically structured) 접근 방식이다(Schoen, 1988; Preston et al, 2001). 이러한 방식의 모형은 취업, 실업, 비경제활동, 사망, 건강/비건강, 장애 상태 등 복수의 상태를 설정하고, 이들 간 전이확률을 이용하여 기대여명 또는 상태별 체류기간(expected duration in state)을 산출한다. 다상태 모형의 개념을 정립한 대표적 연구인 Willekens(1980)은 취업(E), 실업(U), 비경제활동(N), 사망(D)의 네 상태를 연속 마르코프 과정으로 정식화하고,

연령별 전이확률 행렬을 이용해 상태별 기대 체류기간(expected years spent in state)을 계산한 후, 이를 바탕으로 기대노동수명을 추정하였다. 예를 들어, x세에서의 노동기대여명, 즉 취업상태에서 향후 기대되는 잔류기간은 다음의 식(68)과 같이 표현될 수 있다.

$$e_x^E = \sum_{t=x}^{w} P(E_t \mid x) \tag{68}$$

여기서, e_x^E는 x세 기준 노동상태에 대한 기대체류 기간

$P(E \mid x)$는 x세에서 시작하여 t세에 취업상태일 조건부 확률

이러한 방식은 연령 또는 기간별 전이확률 추정을 통해 퇴출과 재취업의 동태를 보다 정밀하게 반영할 수 있다는 장점이 있어, 최근까지 가장 활발히 이용되고 있다. 그러나 건강 상태나 기능 제한이 외생 변수로 처리되어 개인의 노동역량(capacity)을 내포하지 못한다는 한계가 있으며, 전이확률 산출을 위해 장기간의 패널 자료가 요구된다.

한편, 한국에서는 Wolfbein-Wool의 이중감소 모형이 여전히 기대노동수명 산출의 대표적 방식으로 활용되고 있다. 방하남(2002)은 인구주택총조사의 생명표와 경제활동인구조사 자료를 결합하여 1990년대 후반 한국인의 기대노동수명을 산출하였다. 황수경(2005)은 동일한 자료원을 활용하되, 연도별 고용률 변화를 반영하여 남녀·연령별 노동기대여명 추세를 제시하였다. 이들 연구는 모두 사망만을 탈락 요인으로 고려하였으며, 재취업 흐름이나 건강 상태의 변화는 반영하지 않았다. 비교적 최근의 연구인 박시내(2012)와 양원창 외(2024) 역시 Wolfbein-Wool의 방법론을 각기 다른 분석 목적에 적용하여 노동생명표를 구성하였으나, 건강 상태는 고려하지 않았다. 한편, 다상태 모형을 적용한 연구로는 신경

진·최보승(2013)이 있으며, 이들은 KLIPS 자료를 활용하여 4상태 취업 전이 행렬을 제시하였으나 표본의 제약으로 인해 건강 상태나 직무 특성은 분석에 포함되지 않았다.

건강노동기대여명(Healthy Working Life Expectancy)은 한 개인이 건강한 상태로 경제활동에 참여할 것으로 기대되는 남은 생애 기간을 의미한다(Lièvre et al., 2007). 즉, 건강 수준과 노동시장 참여율이 앞으로도 지속된다고 가정할 때, 특정 연령의 개인이 얼마나 오랫동안 건강하게 일할 수 있을지를 추정한 지표이다. 이는 건강기대여명(healthy life expectancy)과 노동기대여명(working life expectancy)의 개념이 결합된 것으로(우해봉, 2014, p. 82), 전통적인 기대여명 또는 건강기대여명의 확장 개념에 해당한다.

건강노동기대여명은 건강기대여명 산출법을 기반으로, 연령별 건강 상태와 고용 상태의 분포를 생명표에 결합함으로써 추정할 수 있다. 가장 직관적이고 간단한 방법은, 주어진 기간의 생명표와 동일 기간의 유병률을 결합하여 건강기대여명을 구하는 Sullivan(1970)의 방법을 응용하는 것이다. 이 방법에서는 연령별 건강한 삶의 비율을 생명표상의 생존연수에 곱하여 기대여명을 건강 상태로 분할하며, 이를 확장하면 연령별 건강 상태와 고용 상태를 모두 고려하여 기대여명을 이중 분할할 수 있다. 즉, 각 연령에서 건강하고 취업 상태에 있는 인구의 비율을 사용하여 생명표를 가중함으로써, 건강하게 경제활동을 할 것으로 기대되는 기간을 산출하는 방식이다.

이를 조금 더 구체적으로 살펴보면, 우선 생명표(기간 생명표)를 사용하여 연령별 생존율, 생존인구, 생존연수(L_x)를 구한다. 여기서 L_x는 생명표에서 x연령부터 $x+1$연령까지 생존한 인년(person-years)을 의미한다. 건강상태를 나타내는 지표를 정의하고, 해당 지표로 연령별 '건강

한 인구'의 비율(h_x)을 구한다. 예를 들어 주관적 건강상태가 '양호'인 인구 비율, 또는 장애·만성질환이 없는 인구 비율 등이 해당 지표로 활용될 수 있다. 같은 방식으로, 연령별 취업자 비율(e_x)도 산출한다. 고용률은 경제활동인구조사에서 연령별 취업자 비율로 계산하며, 필요시 경제활동 참가율을 활용할 수도 있다. 이후 각 연령에서 건강하고 취업 상태'에 있는 인구 비율($h_x \times e_x$)을 구하고, 생명표상의 생존연수(L_x)에 이 값을 곱하여, 해당 연령 구간에서 기대되는 건강하게 경제활동을 한 인년을 계산한다. 선택한 기준연령 a에서의 건강기대여명은 다음의 식(69)와 같이 계산된다.

$$HWLE(a) = \frac{\sum_{x=a}^{w} L_x \, h_x \, e_x}{l_a} \tag{69}$$

여기서, a는 기준연령, l_a은 a세 생존인구.

위의 식에서 예를 들어 $a = 0$(출생 시)인 경우, 식 (69)의 값은 출생 코호트가 평생 동안 보낼 건강+취업 상태 인년의 총합이며, 이를 생존인구로 나누어 출생 시 건강노동기대여명을 산출할 수 있다. 만일 $a = 50$이라면, 50세부터의 건강한 경제활동 가능기간을 얻을 수 있다.

이상의 산출 과정은 Sullivan 방법론을 따른 것으로, 현재의 연령별 건강 수준과 고용 수준이 불변한다고 가정한 단일 시점의 기대값이다. 이 방법의 장점은 단면자료(prevalence data)만으로 건강 상태별 기대여명을 산출할 수 있어 실용적이며, 사망률 자료와 건강조사 자료를 별도로 수집하여 결합할 수 있다는 장점도 있다. 또한 패널 탈락(attrition) 문제가 없어, 서로 다른 시점 간 비교에도 유리하다.

그러나 이 방법론은 현재 관찰된 연령별 건강상태 유병률이 향후에도

동일하게 유지된다는 전제를 두므로, 시간의 흐름에 따른 건강 상태의 변화를 반영하지 못한다는 한계가 있다. 또한, 생명표나 유병률 자료가 충분하지 않은 소규모 집단에 대해서는 건강기대여명을 산출하기 어렵고, 개인별 건강경로의 이질성을 반영하는 데에도 한계가 존재한다.

한편, 전술한 Sullivan 방법론의 한계를 극복하기 위한 대안으로 다중상태 접근이 제안된다(Lièvre et al., 2007). 이 방법에서는 일정 기간 동안 개인이 경험할 수 있는 다양한 상태 간 전이를 고려한다. 예컨대, 건강에서 불건강 상태로의 전이, 불건강 상태에서 건강을 회복하는 전이, 그리고 초기 상태에 따라 달라지는 사망 확률 등을 추정한다. 이렇게 추정된 상태 전이확률을 토대로 다중상태 생명표를 구성하고, 연령의 흐름에 따라 사람들이 건강을 잃거나 회복하거나 사망에 이르는 이행 경로를 시뮬레이션하여 각 건강 상태에서의 기대 연수를 산출한다.

대표적으로 Lièvre et al.(2007)은 이러한 다중상태 기법을 건강노동기대여명(Healthy Working Life Expectancy, HWLE) 산출에 적용하였다. 이들은 건강노동기대여명을 “50세에서 70세 사이에 건강하고 취업상태로 보내는 기대여명”으로 정의하였으며(Lièvre et al., 2007, p. 509), 건강 상태와 고용 상태를 동시에 반영하기 위해 개인의 상태를 건강+취업, 건강+비취업, 불건강+취업, 불건강+비취업의 네 가지 상태로 구분하였다. 이들은 유럽 공동체 가구패널(ECHP) 자료를 활용해 이들 네 상태 간 전이확률을 추정하고, 50~70세 구간에서 각 상태에 머무는 평균 지속기간(mean duration)을 산출하였다. 보다 구체적으로, 마르코프 체인에 기반한 다항 로지스틱 회귀모형을 통해 초기 상태에서 다음 조사 시점의 건강-고용 상태로의 전이확률을 추정하고, 이를 연령별로 보간·적용하여 다중상태 건강조정노동생명표를 구축하였다(Lièvre et al., 2007).

이 방법은 건강 상태 간 가역적 이동(reversibility)을 명시적으로 모델링할 수 있어, 현재의 이행확률에 기반한 기간 지표임에도 장애나 건강 상태의 회복 가능성을 반영할 수 있다는 장점이 있다. 즉, 시간의 흐름에 따른 건강 변화 과정을 포함함으로써, Sullivan 지표보다 현실에 가까운 건강노동기대여명 추계가 가능하다. 그러나 이 방법은 자료에 대한 의존도가 크다는 한계도 지닌다. 동일한 개인을 반복 관찰한 종단자료가 필수적이며, 상태별 전이확률 추정을 위한 통계모형이 필요하므로 분석 과정이 복잡하고 통계적 불확실성이 더 크게 개입한다. 특히, 자료의 제약으로 인해 전이확률 추정에 편향이 생길 경우 전체 기대여명 추계에 영향을 줄 수 있다.

이상의 논의를 정리하면, Sullivan 방법은 간편성과 시계열 비교의 용이성을 장점으로 하나 정태적이며, 다중상태 접근은 정교하고 포괄적이지만 자료 집약적이다. 따라서 연구 목적과 데이터 환경에 따라 두 방법을 적절히 선택하거나 병행하는 방식이 바람직하다.

우리나라의 경우, 건강노동기대여명 산출을 시도한 연구로는 우해봉(2014)이 대표적이다. 이 연구는 Lièvre et al.(2007)의 방법론을 적용하여 건강 상태와 노동 상태를 동시에 고려하였다. 보다 구체적으로는, 건강과 근로활동 두 변수를 조합하여 건강-취업, 건강-비취업, 불건강-취업, 불건강-비취업, 사망의 다섯 가지 상태를 정의하고, 이들 간 전이확률을 추정하여 건강노동생명표를 산출하였다.

3) 방법론 설계

우리나라는 장기 시계열을 갖춘 다양한 패널자료가 구축되어 있으므로, Sullivan 방식뿐 아니라 다중상태 모형도 적용 가능하다. 다만 본 연

구는 인구 및 가구추계 모형을 개선하고, 건강수명 및 노동수명과 같은 지표를 산출하여 인구추계 결과의 활용도를 제고하는 데 목적이 있으므로, 계산의 직관성과 적용의 용이성을 고려해 Sullivan 방식을 적용한다.

Sullivan 방식은 주어진 시점의 생명표에 연령별 건강률과 고용률을 결합하여 기대여명을 상태별로 분할하는 방식이다. 먼저, 연령별로 '건강하고(healthy)' + '취업한(working)' 상태에 있는 인구의 비율을 추정하고, 이를 생명표상의 생존연수에 가중하여 건강노동기대여명을 산출한다. 이를 위한 구성 요소는 아래와 같다.

생존연수(L_x): 연령 x에서 $x+1$세까지 생존한 인년(person-years)

l_a: 기준연령 a세의 생존자 수

h_x: 연령 x에서의 건강상태 비율

e_x: 연령 x에서의 고용률

$h_x \times e_x$: 연령 x에서 건강하고 취업 상태'에 있는 인구 비율

이때, 기준연령에서의 건강노동기대여명은 앞서 제시한 식 (69)와 동일한 방식으로 산출된다. 이를 구현하기 위해 필요한 자료는 크게 세 가지로 구분된다.

첫째, 연령별 사망률 또는 생존율 정보를 포함한 생명표, 둘째, 건강 상태에 대한 연령별 분포, 셋째, 연령별 고용상태 분포 자료의 활용이 중요하다. Sullivan 방식을 적용하기 위해서는 건강수명 추계 결과와 연령별 고용상태 분포를 결합해야 한다. 다만 단순히 전체 고용률을 건강기대여명에 곱하는 방식은 건강과 고용이 독립이라는 가정을 전제하는데, 이는 특히 고령층에서 성립하기 어렵다. 건강 악화는 고용 이탈과 밀접하게 연결되어 있으므로, 건강과 고용 간 상관성을 고려할 수 있는 자료 활용이 바람직하다. 한국노동패널조사(KLIPS), 한국의료패널(KHP) 등은 동일

개인의 건강과 고용 정보를 동시에 수집하여 보다 현실적인 건강노동기대여명 산출을 가능하게 한다.

넷째, 건강지표와 고용지표의 정의가 핵심적이다. 건강의 기준(주관적 건강, 활동 제약, 만성질환 유무 등)과 고용의 범위(고용률, 경제활동참가율, 비공식 노동 포함 여부 등)에 따라 결과가 크게 달라질 수 있다. 우리나라 특성에 부합하는 변수를 사용하되, 국제 비교를 위해 EU-SILC 등 해외 지표도 함께 고려할 필요가 있다.

다섯째, 연령 기준은 지표 해석에 직접적인 영향을 미친다. 0세 기준 건강노동기대여명은 평생 기준이나 비경제활동 기간이 포함되어 실제 노동연령대 활동을 반영하기 어렵다. 국제적으로는 15세, 20세, 50세, 55세 기준이 자주 사용되며, 우리나라는 노동시장 진입이 늦고 고령까지 노동시장에 잔류하는 비율이 높다는 점에서 보다 높은 연령기준도 고려할 수 있다.

여섯째, 성별 구분이 필요하다. 남성은 경제활동참가율이 높고 취업기간이 길어 건강노동기대여명이 높게 나타나는 반면, 여성은 출산·육아 등으로 취업률이 낮고 기대여명은 길지만 건강-비취업 상태의 기대여명이 상대적으로 길다. 다만 최근 고령 여성의 노동시장 잔류 또는 재진입이 증가하고 있다는 점에 주목할 필요가 있다.

일곱째, 지표 해석 시 단일 수치만이 아니라 상태별 기대여명(건강-취업, 건강-비취업 등)을 함께 제시하고, 전체 노동수명 대비 건강노동수명 비율을 병행 분석하는 것이 바람직하다. 교육 수준, 소득 수준, 직종 등 집단 간 격차에 따른 건강노동기대여명 비교는 조기 은퇴의 원인과 구조적 취약성을 진단하는 데 유용한 정보를 제공한다.

다. 주요 이슈 및 향후 개선 방향

1) 구현 과정의 기술적 이슈

PPR_SIMPLE 방식은 가구 구성 정보 없이도 통합형 PPR을 적용하여 추계의 안정성을 확보하였으나, 출산 순위별(첫째/둘째 이상) 의사결정 차이를 명시적으로 반영하지 못하는 한계가 있다. 향후 가구 추계와의 완전한 통합을 통해 분리형 PPR 방식으로 전환할 필요가 있다.

혼인상태 전이율 조정 과정에서 평활화 파라미터를 0으로 설정하여 역산 전이율을 100% 반영하였으나, 일부 연령대에서 연도 간 변동이 크게 나타났다. 향후 0.3~0.5 수준의 평활화 또는 3개년 이동평균 등의 안정화 기법을 검토할 필요가 있다.

PPR 방식은 매년 혼인상태 전이와 가구 계산을 반복하여 GLG 방식 대비 계산 시간이 길다. 대규모 시나리오 분석을 위해서는 코드 최적화(numpy 벡터화, numba JIT 컴파일) 또는 병렬 처리가 필요하다.

건강수명 및 노동수명 추계를 위해서는 건강과 고용 정보를 동시에 포함하는 패널 자료 확보가 필수적이다. 한국노동패널조사, 한국의료패널 등을 활용할 수 있으나, 표본 크기와 연령별 대표성 확보, 그리고 건강지표와 고용지표의 정의 표준화가 향후 과제로 남아 있다.

2) 데이터 및 방법론의 한계

현재 혼인상태 전이 모형은 과거 실적 기반 전이율을 선형적으로 조정하는 방식으로, 성비 불균형, 지역 간 격차, 경제적 요인 등 혼인 시장의 구조적 변화를 반영하지 못한다. 향후 혼인 매칭 모형을 도입하여 연령

별·지역별 성비, 교육 수준, 소득 수준 등을 고려한 혼인 확률 추정이 필요하다.

현재 모형은 모든 혼인상태에 동일한 사망률을 적용하나, 실제 유배우자의 사망률은 미혼·사별·이혼자보다 20~40% 낮다. 혼인상태별 상대위험도(Relative Risk)를 추정하고 전체 사망률을 인구 구성비에 따라 분해하는 방법을 검토할 필요가 있다.

이동력 가정은 통계청의 과거 추세 기반 전망을 준용하였으나, 향후 이민 정책 변화(외국인 근로자 확대, 결혼이민 증가)가 인구 구조와 출산력에 미치는 영향을 명시적으로 모델링할 필요가 있다.

건강수명과 노동수명 연계를 위해서는 건강지표와 고용지표의 정의를 표준화하고, 국제 비교 가능성을 확보하는 것이 중요하다. 또한 성별, 교육 수준, 소득 수준별 격차 분석을 통해 정책 타깃 설정에 활용할 수 있는 세분화된 지표 생산이 필요하다.

3) 모형 확장 방향

현재 모형은 전국 단위 추계이나, 지역별 인구 이동, 혼인 시장, 출산력 차이를 반영한 시·도별 추계 모형으로 확장이 필요하다. 이를 위해서는 지역 간 이동 매트릭스, 지역별 혼인 실적, 지역별 가구 특성 데이터와 지역 간 정합성 확보 메커니즘이 요구된다.

출산력과 이동력에도 불확실성을 반영한 확률적 추계(stochastic projection)로 확장하면 추계 결과의 신뢰구간을 제시할 수 있으며, 연금이나 의료비 재정 추계에서 리스크 평가에 유용하다.

출생 코호트의 교육 수준, 소득 수준, 건강 상태 등을 함께 추적하면 정책 시뮬레이션의 정교도를 높일 수 있다. 특히 건강수명과 노동수명을 연

령별, 성별, 교육 수준별로 세분화하여 산출하고, 이를 KIHASA-SIM의 노동 모듈, 연금 모듈, 의료비 모듈과 연계함으로써 고령화 정책의 통합적 효과 분석이 가능해진다.

KIHASA-SIM과의 연계를 강화하기 위해서는 인구추계 모형에서 산출된 혼인상태별 전이확률, 가구주의 혼인상태별 가구수, 출생아수, 사망자수를 표준화된 형식으로 제공하는 인터페이스 계층을 구축해야 한다. 또한 향후 건강수명과 노동수명 추계 기능을 구축하여 연도별 결과를 제공함으로써, 미시 시뮬레이션의 확률적 변동성을 유지하면서도 거시적 인구 경로와의 정합성을 확보할 수 있다.

제6장

결론

제1절 주요 연구 결과
제2절 결론 및 향후 과제

제6장 결론

제1절 주요 연구 결과

본 연구는 저출산·고령화 시대에 부응하는 인구추계 방법론 개선과 통합 추계 체계 구축을 목적으로 수행되었다. 주요 연구 결과는 다음과 같다.

1. 인구추계 방법론 현황 및 개선 방향

제2장에서는 국내외 인구추계 방법론을 검토하고, 인구·가구 추계의 통합적 접근과 정책시뮬레이션 연계 체계의 필요성을 제시하였다.

기존 GLG(Generalized Log Gamma) 분포 등 모수 분포를 이용한 출산율 전망 방식은 연령별 출산율(ASFR) 패턴을 함수로 적합하여 추계하므로 정책 경로가 명시적이지 않은 한계가 있다. 이에 대한 대안으로 PPR(Parity Progression Ratio, 출산진도비) 방식을 제안하였다. PPR 방식은 혼인상태별 인구 구성과 출산 의사결정을 명시적으로 모델링하여, 초혼율 증가와 같은 정책 변수가 출산으로 연결되는 경로를 추적할 수 있다는 장점이 있다.

또한 인구추계와 가구추계를 별도로 수행하는 기존 방식의 문제점을 지적하고, 혼인상태별 인구 전이를 기반으로 두 추계를 통합하는 방안을 제시하였다. 정책시뮬레이션 연계 측면에서는 KIHASA-SIM과 같은 마이크로시뮬레이션 모형과의 연계를 통해 거시적 인구 경로의 일관성을 유지하면서도 미시적 정책 효과 분석이 가능한 체계를 설계하였다.

2. 출산력 전망 방법론 개선

제3장에서는 출산력 전망 방법론 개선을 위해 PPR 방식의 이론적 배경과 구현 방안을 제시하였다.

기존 출산력 추계 방법론의 한계를 검토한 결과, 모수 분포 기반 ASFR 추정 방식은 혼인·가구 구성과의 연계가 약하고, 정책 시나리오 설정이 자의적일 수 있다. 이에 대한 개선 방안으로 혼인상태별·가구 구성별 출산진도비를 활용하는 PPR 방식을 제안하였다.

PPR 추계모형은 분리형(무자녀 부부/유자녀 부부 구분)과 통합형(전체 유배우 여성 대상)으로 구현 가능하며, 본 연구에서는 가구 정합성 확보를 위해 통합형 PPR을 적용하는 PPR_SIMPLE 방식을 채택하였다. 초혼율을 핵심 정책 변수로 설정하여 유지, '+1.5%/연', '+3.0%/연' 시나리오를 구성하고, 2023~2055년 기간 동안 PPR 기반 추계를 수행한 후 역산된 ASFR을 저장하여 재현 가능성을 확보하였다.

3. 사망력 전망 방법론 및 수명 지표 확장

제4장에서는 사망력 전망 방법론 개선과 생명표 한계연령 연장 방안을 제시하였다.

기존 사망력 추계 방법론으로 Lee-Carter(LC), Lee-Li(LL), Lee-Carter with Extended Rotation(LC-ER), Lee-Li-Gerland Extended(LLG-E) 모형을 검토하고, 본 연구에서는 성별 간 사망률 개선의 일관성을 확보하고 연령대별 패턴 변화를 반영할 수 있는 LLG-E 모형을 채택하였다. 1,000회 시뮬레이션을 통해 불확실성을 정량화하고 중위(50th percentile) 시나리오를 추계에 활용하였다.

생명표 한계연령을 기존 100세에서 130세로 연장하기 위해 Coale-Kisker 방법과 Coherent Kannisto 방법을 비교 검토하였다. 두 방법 간 0세 기대수명 차이는 0.1세 이내로 미미하였으나, 85세 이상 고령층 기대여명에서는 의미 있는 차이가 나타났다. 구현의 단순성을 고려하여 Coale-Kisker 방법을 채택하였으며, 한계연령 연장으로 인한 0세 기대수명 증가 효과는 0.3~0.4세 수준이었다.

4. 통합 추계 체계 구축 및 결과 검토

제5장에서는 출산력, 사망력, 이동력, 혼인상태, 가구를 통합하는 추계 시스템을 구축하고, 3개 시나리오(저위/중위/고위)의 추계 결과를 검토하였다.

통합 추계 모형은 3계층 데이터 구조(actuals/assumptions/projections)와 6개 핵심 모듈(공통/dbloader/mortality/fertility/mobility/population)로 설계되었다. ProjectionPipeline 클래스를 통해 5단계 추계 프로세스(Fertility 전망 → Mortality 전망 → Mobility 전망 → Population 추계 → 결과 아카이빙)를 통합 관리하며, 시나리오별 결과는 latest/archive 폴더로 자동 정리되어 버전 관리와 재현 가능성을 확보하였다.

2022년 기준 2122년까지 100년 추계 결과, 총인구는 저위 시나리오에서 5,169만 명에서 1,238만 명(76.1% 감소), 중위 1,486만 명, 고위 1,803만 명으로 전망되었다. 시나리오 간 격차는 2050년 118만 명(2.6%)에서 2120년 565만 명(45.7%)으로 확대되어, 초혼율 증가 효과가 장기적으로 누적됨을 확인하였다.

연령 구조는 급속한 고령화 추세를 보여, 65세 이상 인구 비율이 2023년

18.4%에서 2080년 51.457.4%까지 상승하였다. 노인부양비는 2023년 26에서 2080년 122~146으로 급등하며, 중위연령은 45세에서 67세로 상승하였다.

30~34세 여성의 유배우 비율은 저위 시나리오에서 2023년 12.7%에서 2050년 12.0%로 정체된 반면, 중위는 17.2%, 고위는 24.1%로 상승하여 초혼율 증가 가정의 효과가 뚜렷하게 나타났다.

합계출산율(TFR)은 저위 시나리오에서 0.62 수준에 고정된 반면, 중위는 2060년 이후 0.84, 고위는 1.08로 수렴하였다. 출생아수는 2023년 24.4만 명에서 2120년 저위 2.9만 명, 중위 5.1만 명, 고위 8.5만 명으로 감소하였다.

기대수명은 2023년 83.9세(남 80.9세, 여 86.8세)에서 2090년 93.0세(남 91.6세, 여 94.3세)로 상승하며, 성별 격차는 5.9세에서 2.7세로 축소되었다.

통계청 중위 시나리오와의 비교 결과, 본 연구의 초혼율 '+1.5%/연' 시나리오의 총인구는 2040년 1.8%, 2070년 6.3%, 2120년 24.4% 낮게 전망되었다. 주요 차이는 출산력에서 발생하였으며, 본 연구의 TFR이 2030~2060년 기간 0.15~0.32 낮았다. 사망력은 LLG-E 모형 기반으로 기대수명 차이가 0.3~0.5세로 미미하였다.

5. KIHASA-SIM 연계 시뮬레이션 체계 설계

제5장 3절에서는 코호트요인법 기반 집계 인구추계 모형과 동태적 마이크로시뮬레이션 모형인 KIHASA-SIM 간의 연계 방안을 제시하였다.

연계 체계는 단방향 방식으로 설계되어, 인구추계 모형에서 산출된 혼인상태별 전이확률, 가구주의 혼인상태별 가구수, 출생아수, 사망자수를

KIHASA-SIM에 외생변수로 투입한다. 이를 통해 동일한 인구·가구 시나리오하에서 노동, 소득, 연금 등의 정책 효과를 일관되게 분석할 수 있다.

건강수명 및 노동수명 연계 방안을 검토하였으며, Sullivan 방식을 활용한 건강수명 추계 방법론과 건강-고용 상태를 동시에 고려한 건강노동기대여명 산출 방법론을 제시하였다. 향후 건강수명과 노동수명 추계 기능을 구축하여 KIHASA-SIM과 연계하면 고령화 정책의 통합적 효과 분석이 가능해질 것으로 기대된다.

제2절 결론 및 향후 과제

1. 연구의 의의

본 연구는 저출산·고령화 대응을 위한 인구추계 방법론 개선과 통합 추계 체계 구축이라는 측면에서 다음과 같은 의의를 갖는다.

첫째, PPR 방식 도입으로 출산력 추계의 정책 연계성을 강화하였다. 초혼율 증가를 명시적 정책 변수로 설정함으로써, 저출산 대응 정책의 효과를 정량적으로 평가할 수 있는 분석 틀을 제공하였다. 이는 기존 ASFR 직접 가정 방식의 한계를 보완하고, 혼인·가구 정책과 출산력 전망을 연계할 수 있는 토대를 마련하였다.

둘째, 생명표 한계연령 130세 연장으로 초고령층 인구 분석의 정밀도를 높였다. 100세 이상 인구를 연령별로 세분화하여 추적함으로써, 향후 기대수명 연장에 대비하고 초고령층 대상 정책(요양, 의료, 돌봄 등) 수립에 필요한 정보를 제공할 수 있게 되었다.

셋째, 인구·가구·정책시뮬레이션을 통합하는 추계 체계를 구축하였다.

혼인상태별 인구 전이를 중심으로 인구추계와 가구추계를 통합하고, KIHASA-SIM과의 연계 방안을 설계함으로써 거시적 인구 경로의 일관성과 미시적 정책 효과 분석을 동시에 확보할 수 있는 기반을 마련하였다.

넷째, 재현 가능하고 확장 가능한 추계 시스템을 구현하였다. 모듈화된 구조, 표준화된 데이터 형식, 명시적인 가정 저장, 자동 아카이빙 등을 통해 추계 과정의 투명성과 재현 가능성을 확보하였으며, 향후 지역별 추계, 확률적 추계 등으로의 확장 가능성을 열어두었다.

2. 정책적 시사점

본 연구의 결과는 저출산·고령화 대응 정책 수립에 다음과 같은 시사점을 제공한다.

첫째, 초혼율 증가 정책의 장기적 효과가 확인되었다. 초혼율 연 1.5% 증가 시나리오 대비 초혼율 연 3.0% 증가 시나리오는 2120년 총인구가 348만 명(23.4%) 많으며, TFR도 0.24 높다. 이는 혼인 지원 정책(주거, 경제적 안정, 일·가정 양립 등)이 장기적으로 인구 감소 완화에 기여할 수 있음을 시사한다.

둘째, 출산력 정책의 효과가 나타나기까지 상당한 시간이 소요된다. 초혼율 증가 효과가 유배우 인구 증가로 이어지고, 이것이 다시 출산으로 연결되기까지 10~20년의 시차가 존재한다. 2040년까지는 시나리오 간 TFR 격차가 0.23에 불과하지만, 2050년 이후 0.46으로 확대된다. 따라서 저출산 대응 정책은 즉각적 효과를 기대하기보다 장기적 관점에서 지속적으로 추진해야 한다.

셋째, 인구 감소 자체를 막기는 어렵지만 감소 속도는 조절 가능하다. 초혼율 연 3.0% 증가 시나리오도 2120년 1,803만 명으로 현재의 35%

수준이며, TFR 1.08은 인구 대체 수준(2.1)에 크게 미달한다. 그러나 초혼율 유지 시나리오 대비 565만 명(45.7%) 많아, 정책 노력에 따라 인구 감소 속도와 인구 구조를 일정 수준 관리할 수 있다.

넷째, 고령화 속도는 출산력 정책만으로 완화하기 어렵다. 2080년 고령 인구 비율은 초혼율 유지시나리오 57.4%, 초혼율 '+3.0%/연' 시나리오 51.4%로 시나리오 간 차이가 6.0%p에 불과하다. 이미 고령층으로 진입한 베이비붐 세대의 영향이 크기 때문이다. 따라서 고령 인구 비율 자체를 낮추려는 시도보다는, 건강수명 연장, 노동수명 확대, 노인부양 체계 개선 등 고령사회 적응 정책에 집중하는 것이 현실적이다.

3. 향후 과제

본 연구의 한계와 향후 개선 방향은 다음과 같다.

방법론 측면에서는 첫째, PPR 방식을 분리형(무자녀/유자녀 구분)으로 확장하여 출산 순위별 의사결정을 명시적으로 반영할 필요가 있다. 이를 위해서는 가구 추계와의 완전한 통합이 선행되어야 한다. 둘째, 혼인상태 전이 모형에 혼인 매칭 이론을 도입하여 성비, 교육 수준, 소득 수준, 지역 등을 고려한 구조적 혼인 확률 추정이 필요하다. 셋째, 혼인상태별 사망률 차이를 반영하여 유배우자의 낮은 사망률을 명시적으로 모델링해야 한다.

데이터 측면에서는 첫째, 건강수명 및 노동수명 추계를 위해 건강과 고용 정보를 동시에 포함하는 패널 자료 구축이 필요하다. 한국노동패널조사, 한국의료패널 등을 활용하되, 표본 크기와 연령별 대표성을 강화해야 한다. 둘째, 건강지표와 고용지표의 정의를 표준화하고 국제 비교 가능성을 확보해야 한다. 셋째, 지역별 추계를 위한 시·도별 혼인 실적, 가구 특

성, 이동 매트릭스 등의 데이터 확보가 필요하다.

시스템 측면에서는 첫째, 전국 단위 추계를 시·도별 추계로 확장하여 지역별 인구 이동, 혼인 시장, 출산력 차이를 반영하고 지역 간 정합성을 확보해야 한다. 둘째, 확률적 추계(stochastic projection)로 확장하여 추계 결과의 신뢰구간을 제시하고, 연금·의료비 재정 추계에서 리스크 평가에 활용할 수 있도록 해야 한다. 셋째, KIHASA-SIM과의 연계를 강화하여 건강수명과 노동수명 추계 결과를 제공하고, 인터페이스 계층을 통해 데이터 교환의 효율성과 정확성을 높여야 한다.

정책 활용 측면에서는 첫째, 다양한 정책 시나리오(예: 이민 정책 변화, 육아휴직 확대, 주거 지원 강화 등)를 설정하고 효과를 시뮬레이션함으로써 정책 의사결정을 지원해야 한다. 둘째, 교육 수준, 소득 수준, 직종별 건강수명과 노동수명 격차를 분석하여 취약 집단에 대한 타깃 정책 수립에 활용해야 한다. 셋째, 정책 효과의 불확실성을 정량화하여 제시함으로써 정책 리스크 관리에 기여해야 한다.

본 연구에서 구축한 통합 추계 체계는 저출산·고령화 대응 정책의 장기적 효과를 평가하고, 인구 변동이 사회경제에 미치는 영향을 종합적으로 분석하는 데 활용될 수 있을 것이다. 향후 지속적인 방법론 개선과 시스템 확장을 통해 정책 수립의 과학적 기반을 강화하고, 인구 변화에 능동적으로 대응하는 사회 시스템 구축에 기여할 수 있기를 기대한다.

참고문헌

김기환, 전새봄 (2015), 출산율 예측모형을 이용한 한국의 출산력 시나리오 분석, 응용통계분석, 28(4), pp. 685-701.

김순영, 오진호 (2017), 모수와 비모수 모형을 활용한 사망률 예측 비교 연구, 응용통계연구, 31(5), pp. 701-717.

김순영, 오진호, 김기환 (2018), 시계열 이용기간에 따른 사망률 예측 비교, 응용통계연구, 31(1), pp. 41-65.

김승현 (2020), 통계적 노출 제어 방법 및 활용사례 연구: 재현데이터 방법론을 중심으로, 주택금융리서치, 9호, pp. 18~35.

김현태, 장가영 (2023), 데이터 가명·익명처리 기법의 현황과 대안: 재현데이터를 중심으로, 보험연구원.

박경애, 최기홍 (2006), 증감 노동생명표에 의한 노동기대여명의 측정과 전망, 한국인구학, 29(3), pp. 51-72.

박시내 (2012), 노동생명표 작성을 통한 중·고령 세대의 은퇴 연령 분석, 통계개발원 연구보고서.

박유성, 김미리, 김성용 (2013), 확률적 출산율 모형과 한국의 미래인구 구조, 조사연구, 14(3), pp. 49-78.

방하남 (2002), Work-life profiles of Korean workers: life-table analysis, 한국노동패널 연구, 2002-05.

송창길, 류한별, 나원희, 백주하, 한신실, 고경표 (2023), 인구추계모형 구축과 활용 방안, 한국보건사회연구원, 연구보고서, 2023-28.

송창길, 오진호, 고경표 (2024), 장래가구추계 모형 구축, 한국보건사회연구원, 연구보고서, 2024-26.

신경진, 최보승 (2013), 생명표 활용에 대한 고찰 및 분석 방법 소개, 한국데이터정보과학회지, 24(6), pp. 1159-1175.

양원창, 김수범, 임달오, 송수진 (2024), 노동생명표를 이용한 노동기간 변화와

특성 분석, 보건정보통계학회지, 49(4), pp. 383-391.

엄지민, 김기환 (2013), 여성 초혼율을 사용한 합계출산율 예측에 관한 연구, 한국자료분석학회지, 15(3), pp. 1261-1271.

오진호 (2018), 미완결 코호트출산율 예측에 관한 연구, 한국인구학, 41(1), pp. 109-134.

오진호 (2018), 모수, 비모수, 베이지안 출산율 모형을 활용한 합계출산율 예측과 비교, 응용통계연구, 31(6), pp. 677-692.

오진호, 김순영 (2018), The Korean journal of applied statistics, 응용통계연구, 31(5), pp. 637-653.

오진호 (2024), 출산순위별 출산율 확률적 예측 연구, 한국인구학, 47(1), pp. 19-40.

우해봉 (2014), 건강노동기대여명 지표를 활용한 중고령층의 건강과 근로활동 분석, 사회복지정책, 41(4), pp. 77-99.

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

전광희 (2006), 인구학적 관점에서 본 적정인구의 추계, 한국인구학, 29(1), pp. 209-239.

통계청 (2024), 장래인구추계: 2022~2072년.

황수경 (2005), 노동생명표 작성법을 이용한 은퇴연령의 추정, 노동리뷰, 한국노동연구원, pp. 45-55

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., and Heilig, G. K. (2011), probabilistic projections of the total fertility rate for all countries, Demography, 48, pp. 815-839.

Austad, S. N. (2015), In Molecular and Cellular Biology of Aging (eds Vijg, J., Campisi, J. and Lithgow, G.) Ch. 2 (The Gerontological Society of America, 2015).

Basten, S. A., Coleman, D. A., and Gu, B. (2012), Re-examining the

fertility assumptions in the UN's 2010 World Population Prospects: Intentions and fertility recovery in East Asia? Presented at the Annual Meeting of the Population Association of America; San Francisco.

Boleslawski, L. and Tabeau, E. (2001), Comparing theoretical age patterns of mortality beyond the age of 80, In E. Tabeau, A. Van Den Berg Jeths, and C. Heathcote (Eds.), Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective, Kluwer Academic Publishers, Dordrecht.

Clark, SJ (2019), A general age-specific mortality model with an example indexed by child mortality or both child and adult mortality, Demography, 56, pp. 1131-1159.

Coale, A. J. and Demeny, P. (1983), Regional Model Life Tables and Stable Populations, Academic Press, New York.

Coale, A., & Guo, G. (1989), Revised regional model life tables at very low levels of mortality, Population Index, 55(4), pp. 613-643.

Coale, A. J. and Kisker, E. E. (1990), Defects in data on old-age mortality in the United States: New procedures for calculating schedules and life tables at the highest ages, Asian and Pacific Population Forum, 4, pp. 1-31.

Denuit, M. et Goderniaux, A. (2005), Closing and projecting life tables using log-linear models, Bulletin de l'Association Suisse des Actuaries, 1, pp. 29-49.

Dong X, Milholland B, and Vijg J. (2016), Evidence for a limit to human lifespan, Nature, 538, pp. 257-259.

ECB (2011), European Central Bank Household Finance and Consumption Network Core Output Variables, March.

European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank (2008), System of National Accounts 2008.

Federal Interagency Forum Aging Related Statistics (FIFARS) (2010), Older Americans 2004: Key indicators of wellbeing, Federal Interagency Forum on Aging-Related Statistics, Washington, DC: U.S. Government Printing Office, July 2010.

Feeney, G. and Yu, J. (1987), Period parity progression measures of fertility in China, Population Studies, 41, pp. 77-102.

Galley C, and Woods R (1998), Reflections on the distribution of deaths in the first year of life, Population, 53, pp. 921-946.

Gompertz, B. (1825), On the nature of the function expressive of the law of mortality, Philosophical Transactions, 27, pp. 513-585.

Guillot, M., Prieto, J. R., Verhulst, A., and Gerland, P. (2022), Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data, Demography, 59(1), pp. 321-347.

Hadwiger, H. (1940), Eine Analytische Reprodutions-Funktion fur Biologische Gesamtheiten, Skandinavisk Aktuarietidskrift, 23, pp. 101-113.

Heligman, L. and Pollard, J. H. (1980), The age pattern of mortality, Journal of the Institute of Actuaries, 107, pp. 49-80.

Himes, C. L., Preston, S. H., and Condran, G. A. (1994), A relational model of mortality at older ages in low mortality countries, Population Studies, 48(2), pp. 269-291.

Hoem, J. M., Madsen, D., Nielsen, J. L., Ohlsen, E., Hansen, H. O., and Rennermalm, B. (1981), Experiments in modelling recent Danish fertility curves, Demography, 18, pp. 231-244.

Horiuchi, S. and Wilmoth, J. (1995), Annual meeting of the population association of America, San Francisco, CA: Population Association of America.

Horiuchi, S. and Coale, A. J. (1982), A simple equation for estimating the expectation of life at old ages, Population Studies, 36(2), pp. 317-326.

Hughes, B. G. and Hekimi, S. (2017), Many possible maximum lifespan trajectories, Nature, 546, pp. E8-E9.

Hyndman, R. J. and Booth, H. (2008), Stochastic population forecasts using functional data models for mortality, fertility and migration, International Journal of Forecasting, 24, pp. 323-342.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013), Coherent mortality forecasting: the product-ration method with functional time series models, Demography, 50, pp. 261-283.

Hyndman, R. J. and Ullah, S. (2007), Robust forecasting of mortality and fertility rates: a functional data approach, Computational Statistics & Data Analysis, 51, pp. 4942-4956.

ILO (2003), Household Income and Expenditure Statistics, Report II of the 17th International Conference of Labour Statisticians in 2003, published by the International Labour Office.

Kannisto, V. (1992), Development of oldest-old mortality, 1950-1990, Evidence from 28 developed countries, Monographs on Population Aging, Vol. 1, Odense: Odense University Press, pp. 1-108.

Knodel J, and Kintner H (1977), Impact of breast feeding patterns on biometric analysis of infant mortality, Demography, 14, pp. 391-409.

Lantoine C., and Pressat R. (1984), New aspects of infant-mortality, Population, 39, pp. 253-264.

Lee, R. D. and Carter, L. R. (1992), Modeling and forecasting U.S. mortality, Journal of the American Statistical Association, 87, pp. 659-671.

Li, N and Lee, R (2005), Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method, Demography, 42(3), pp. 575-594.

Lee, R. D. and Miller, T. (2001), Evaluating the performance of the Lee-Carter method for forecasting mortality, Demography, 38, pp. 537-549.

Li, N. and Lee R. (2005), Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method, Demography, 42, pp. 575-594.

Li, N. and Gerland, P. (2011), Modifying the Lee-Carter Method to Project Mortality Changes up to 2100, the Population Association of America 2011 Annual meeting-Washington, DC, Session 125, formal Demography I: Mathematical Models and Methods.

Li, N., Lee, R., and Gerland, P. (2013), Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections, Demography, 50, pp. 2037-2051.

Lynch KA, Greenhouse JB, and Brandstrom A (1998), Biometric modeling in the study of infant mortality: Evidence from nineteenth-century Sweden, Historical Methods, 31, pp. 53-64.

Makeham, W. (1860), On the law of mortality and the construction of annuity tables, Journal of the Institute of Actuaries, 13, pp. 325-358.

Manfredini M (2004), The Bourgeois-Pichat's biometric method and the influence of climate: New evidences from late 19th-century Italy, Social Biology, 51, pp. 24-36.

Murray CJL, Ferguson BD, Lopez AD, Guillot M, Salomon JA, and Ahmad O. (2003), Modified logit life table system: Principles, empirical validation, and application, Population Studies, 57, pp. 165-182.

Oeppen, J. & Vaupel, J. W. (2002), Demography, Broken limits to life expectancy, Science 296, pp. 1029-1031.

Pearce, M., & Raftery, A. E. (2021), Probabilistic forecasting of maximum human lifespan by 2100 using Bayesian population projections, Demographic Research, 44, pp. 1271-1294. https://doi.org/10.4054/DemRes.2021.44.52.

Peristera, P. and Kostaki, A. (2007), Modeling fertility in modern populations, Demographic Research, 16, pp. 141-194.

Perks, W. (1932), On some experiments in the graduation of mortality statistics, Journal of the Institute of Actuaries, 63, pp. 12-27.

Preston, S. H., Heuveline, P. and Guillot, M. (2000), Demography: Measuring and Modeling Population Processes, Blackwell Publishing Ltd, Oxford.

Ramsay, J. O. and Silverman, B. W. (2005), Functional Data Analysis (2nd ed), Springer-Verlag, New York.

Raftery, A. E., Li, N., Sevcıkova, H., Gerland, P., and Heilig, G. K. (2012), Bayesian probabilistic population projection for all countries, In Proceedings of the National Academy of Sciences of United States of America, 109, pp. 13915-13921.

Raftery, A. E., Alkema, L., and Gerland, P. (2014), Bayesian population projections for the United Nations, Statistical Science, 29, pp. 58-68.

Ryder, N. B. (1986), Observations on the history of cohort fertility in the United states, Population and Development Review, 12(4), pp. 617-643.

Schoumaker, B. (2013), A Stata module for computing fertility rates and TFRs from birth histories: TFRs, Demography Research, 28(38), pp. 1093-1144.

Sevc ı kova, H., Alkema, L., and Raftery, A. E. (2011), bayesTFR: an R package for probabilistic projections of the total fertility rate, Journal of Statistical Software, 43, pp. 1-29.

Sevcikova, H., Alkema L., Raftery, A. E., Fosdick, B., and Gerland, P. (2018), Package 'bayesTFR', R demography.

Statistics Canada (2015), Population Projections for Canada(2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions.

Statistics Canada (2025), Demosim microsimulation model at Statistics Canada: A tool for policy planning and evaluation.

Thatcher, A. R., Kannisto, V., Vaupel, J. W. (1998), The force of mortality at ages 80 to 120, Monographs on Population Aging, Vol. 5, Odense: Odense University Press.

Thatcher, A. R. (1999), The Long-Term Pattern of Adult Mortality and the Highest Attained Age, Journal of the Royal Statistical Society, Series A (Statistics in Society), 162(1), pp. 5-43.

UN (2019), World Population Prospects 2019.

UN (2022), World Population Prospects 2022.

UN (2024), World Population Prospects 2024.

UNECE (2006), Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing, published by the United Nations Economic Commission for Europe (UNECE) in 2006.

UN-IGME (Inter-agency Group for Child Mortality Estimation) (2023), Levels & Trends in Child Mortality Report 2023.

Van Hook, J. and Altman, C.E. (2013), Using discrete-time event history fertility models to simulate total fertility rates and other fertility measures, Population and Policy Review, 32(4), pp. 585-610.

Vijg, J. (2010), Aging of the Genome: The dual role of DNA in life and death, Oxford University Press, 2010 Apr.

Vaupel, J. W. (2010), Biodemography of human ageing, Nature 464, pp. 536–542.

Weibull, W. (1951), A Statistical Distribution Function of Wide Applicability, Journal of Applied Mechanics, 18, pp. 293-297.

WHO (2024), Healthy life expectancy (HALE) at age 60 (years), https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-ghe-hale-healthy-life-expectancy-at-age-60.

Wilmoth, J., (1995), Are mortality rates falling at extremely high ages: An investigation based on a model proposed by Coale and Kisker, Popul. Stud., 49(2), pp. 281-295.

Wilmoth J, Zureick S, Canudas-Romo V, Inoue M, and Sawyer C (2012), A flexible two-dimensional mortality model for use in indirect estimation, Population Studies, 66, pp. 1-28.

[인터넷 사이트]

https://www.abs.gov.au/statistics/people/population/househld-and-family-projections-australia.

https://www.jchs.harvard.edu/sites/default/files/Harvard_JCHS_McCue_Household_Projections_Rev010319.pdf.

https://assets.cmhc-schl.gc.ca/sites/cmhc/professional/housing-markets-data-and-research/market-reports/housing-market-insight/2021/housing-market-insight-69701-m01-en.pdf?rev=05ff122f-8c1b-4fec-8f3f-db5e935a59cf.

https://www.nrscotland.gov.uk/files/statistics/household-projections/18/household-proj-18-report.pdf.
https://www.nrscotland.gov.uk/files/statistics/household-projections/18/household-proj-18-report.pdf.
https://www.gov.wales/sites/default/files/statistics-and-research/2018-11/population-household-statistics-quality-report.pdf.
https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/HP16-methodology.pdf.
https://www.pbl.nl/uploads/default/downloads/pbl-cbs-2022-regionale-bevolkingsprognose-2022-2050-4977.pdf.
https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf.
https://www.nhis.or.kr/announce/wbhaec11403m01.do.
https://www.healthdata.org/research-analysis/gbd.
https://www.un.org/development/desa/pd/data/household-size-and-composition?utm_source=chatgpt.com.
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Household_composition_statistics.
https://www150.statcan.gc.ca/n1/pub/17-20-0001/172000012025001-eng.htm.

Abstract

Improving Population Projection Methodology
: Focusing on Fertility and Mortality

Project Head: Song, Changgil

South Korea is experiencing unprecedented demographic transitions with a total fertility rate of 0.72 (2023) and entry into a super-aged society. Existing parametric distribution-based fertility projection methods lack explicit policy pathways, while limited upper age bounds in life tables and disconnected population and household projection frameworks hinder long-term policy analysis. This study aims to establish an independent population projection system by introducing the Parity Progression Ratio (PPR) method, extending life tables to age 130, and constructing an integrated population-household projection framework.

We introduced the PPR method to incorporate first marriage rates as an explicit policy variable and adopted the Lee-Li-Gerland Extended (LLG-E) model for mortality projection with life table extension to age 130. An integrated projection model based on marital status transitions was developed to conduct 100-year projections under low, medium, and high first marriage rate scenarios, confirming the long-term cumulative effects of marriage policies. Results were compared with

Co-Researchers: Ryu, Jaerin · Oh, Jinho · Son, Hyunsub · Oh, Daeun

official projections by Statistics Korea, and a framework for integration with KIHASA-SIM microsimulation and methodologies for projecting healthy and working life expectancy were presented.

The PPR method strengthened policy linkages in fertility projections, and life table extension to advanced ages enhanced precision in analyzing the oldest-old population. First marriage rate policies demonstrate long-term effects but require 10-20 years to materialize. Population aging cannot be mitigated by fertility policies alone, emphasizing the importance of adaptation policies for aged societies. Future research directions include parity-specific PPR models, marriage market matching models, subnational and stochastic projection frameworks, and enhanced KIHASA-SIM integration.

Key words: Population Projection, Fertility, Parity Progression Ratio(PPR), Mortality, Life Table Extension