

패널자료 품질개선 연구(XI)

www.kli.re.kr

장인성 · 이지은 · 김기민 · 김우영
이상혁 · 이지희 · 채민희 · 최효미

목 차

요 약	i
제1장 서 론	(장인성) 1
제2장 노동패널 가족력 자료 생성 (최효미)	6
제1절 서 론	6
제2절 가족력 자료 구축	9
1. 가족력 자료 생성 과정 개요	9
2. STEP 1: 가구원 정보 데이터 파일 생성	11
3. STEP 2: 가족관계 PID 파일 생성	13
4. STEP 3: Family Data 생성	16
5. STEP 4: 가구원 소속 변화 데이터 파일	20
6. STEP 5: 혼인력 데이터 생성	21
7. STEP 6: EHC 데이터 파일 생성	22
제3절 자료 특성	24
1. 가족 관계 및 가구원 구성	25
2. 출산 관련 현황	30
제4절 결 론	41
제3장 노동패널 고용지표의 신뢰성 검토 : 경제활동인구조사와 비교	(김우영) 43
제1절 서 론	43

제2절 1998년 표본의 경제활동참가율, 고용률, 실업률 비교	45
1. 전체 인구 대상	45
2. 남성과 여성	50
3. 연령별 비교	53
4. 종사상 지위	60
5. 소 결	62
제3절 2009년 표본의 경제활동참가율, 고용률, 실업률 비교	63
1. 전체 인구 대상	64
2. 남성과 여성	67
3. 연령별 비교	70
4. 종사상 지위	74
5. 소 결	76
제4절 2018년 표본의 경제활동참가율, 고용률, 실업률 비교	77
1. 전체 인구 대상	77
2. 남성과 여성	79
3. 연령별 비교	81
4. 종사상 지위	84
5. 소 결	86
제5절 요약 및 개선방안	87
1. 요약	87
2. 개선 방안	89
제4장 행정데이터와 서베이데이터의 연계 (이상혁)	93
제1절 서 론	93
제2절 선행연구	96
제3절 행정데이터와 서베이데이터 연계의 해외사례	98
1. 덴마크	98
2. 스웨덴, 노르웨이, 핀란드 등	99

3. 미 국	102
4. 호 주	109
제4절 데이터 연계의 잠재적 가치	112
1. 측정오류(Measurement Error)	113
2. 세대간 계층이동성(Intergenerational Mobility)	114
3. 보건의료 연구(Health and Medical Research)	115
제5절 소 결	116

제5장 조사방법에 따른 조사 결과 검증 연구 : 한국노동패널 조사 면접조사, 전화조사, 웹조사의 비교

..... (이지은 · 채민희)	118
제1절 서 론	118
제2절 노동패널 조사방법 개요	120
1. 웹조사 시행과정	120
2. 웹조사 응답자 특성	121
3. 조사도구별 가구 특성(전화조사와 면접조사)	124
4. 조사도구별 개인 응답자 특성	127
제3절 방법론	129
1. 방법론 논의	129
2. 분석 모형	132
제4절 통계적 검증	133
1. 분석 개요	133
2. 분석 결과 : 대면조사 및 전화조사 비교	135
3. 분석 결과 : 대면조사 및 웹조사 비교	143
4. 결과 요약	147
제5절 소 결	148

제6장 Web 조사방식이 응답률과 자료의 질에 미치는 영향 :	
사업체패널조사를 중심으로	(김기민 · 이지희) 153
제1절 서 론	153
제2절 CAWI 활용 현황과 응답 특성	155
1. CAWI 조사의 활용 현황	156
2. CAWI 조사의 응답 특성	159
제3절 CAWI 효과 분석	168
1. 8차 자료를 활용한 회귀분석	168
2. 패널자료를 활용한 이중차분법 분석	175
제4절 요약 및 결론	179
참고문헌	181
부 록	191

표 목 차

<표 1- 1> 패널자료 품질개선 연구(I~X)의 세부 연구주제	2
<표 2- 1> 가구원 데이터 생성 시 활용된 변수	11
<표 2- 2> 가구원 데이터 관측치	12
<표 2- 3> 가족관계 매칭표	14
<표 2- 4> 가족관계 PID 파일 예시	16
<표 2- 5> 가족관계를 나타내는 변수명	19
<표 2- 6> 가구원 정보 변수 지정	20
<표 2- 7> EHC data 파일 변수명	24
<표 2- 8> 차수별 본인 및 가족관계 응답자 수	25
<표 2- 9> 차수별 자녀 응답자 수	26
<표 2-10> 차수별 한부모 가구 비중	28
<표 2-11> 기혼 유배우자 중 부모와 동일 가구 비중	30
<표 2-12> 평균 초산 연령	32
<표 2-13> 평균 자녀 수	32
<표 2-14> 다자녀 출산	33
<표 2-15> 첫째 자녀 성별과 후속 출산	35
<표 2-16> 첫째 및 둘째 자녀 성별과 후속 출산	36
<표 2-17> 분가 연령별 분가 사유	37
<표 2-18> 사망 연령별 사망 사유	38
<표 2-19> 코호트별 평균 초혼 연령	40
<표 2-20> 초혼의 평균 결혼 지속 기간(이혼한 경우만)	41
<표 3- 1> KLIPS 98년 표본과 경제활동인구조사(시 지역)의 고용지표 비교	47

<표 3- 2> KLIPS 98년 표본의 가중치 적용 여부에 따른 경찰조사 (시 지역)와의 고용지표 비교	49
<표 3- 3> KLIPS 98년 남성 표본의 가중치 적용 여부에 따른 경찰조사(시 지역)와의 고용지표 비교	51
<표 3- 4> KLIPS 98년 여성 표본의 가중치 적용 여부에 따른 경찰조사(시 지역)와의 고용지표 비교	52
<표 3- 5> Kolmogorov - Smirnov 테스트 결과	57
<표 3- 6> KLIPS 98년 표본과 경찰조사와의 연령별 고용률 비교	58
<표 3- 7> 2020년 KLIPS 98년 표본과 경찰조사와의 고용률 차이 분해	60
<표 3- 8> KLIPS 98년 표본과 경찰조사(시 지역)와의 종사상 지위 분포 비교	61
<표 3- 9> KLIPS 2009년 표본과 경제활동인구조사의 고용지표 비교	65
<표 3-10> KLIPS 09년 표본과 경찰조사와의 비교	66
<표 3-11> KLIPS 09년 남성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교	68
<표 3-12> KLIPS 09년 여성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교	69
<표 3-13> KLIPS 09년 표본과 경찰조사와의 연령별 고용률 비교	72
<표 3-14> 2020년 KLIPS 09년 표본과 경찰조사와의 고용률 차이 분해	73
<표 3-15> 2020년 KLIPS 09년 표본과 경찰조사와의 15~34세 고용률 차이 분해	74
<표 3-16> KLIPS 09년 표본과 경찰조사와의 종사상 지위 분포 비교	75
<표 3-17> KLIPS 18년 표본과 경제활동인구조사의 고용지표 비교	78

<표 3-18> KLIPS 18년 표본과 경찰조사 고용지표 비교(가중치 사용 안함)	79
<표 3-19> KLIPS 2009년 남성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교	80
<표 3-20> KLIPS 2009년 여성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교	80
<표 3-21> KLIPS 18년 표본과 경찰조사와의 연령별 고용률 비교	83
<표 3-22> KLIPS 18년 표본과 경찰조사와의 성별 종사상 지위에 따른 취업자 수 비교	85
<표 3-23> KLIPS와 경찰조사 자료와의 차이(2020년 현재 기준) ..	87
<표 3-24> KLIPS 98년 표본과 경찰자료의 연령별 가중치 분포 (2020년 기준)	90
<표 4- 1> PVS를 이용한 연계 결과	104
<표 4- 2> PSID 2011 주요 인터뷰 내용	106
<표 4- 3> 미국 인구조사국의 PVS프로젝트 매치 성공률	108
<표 5- 1> 22차년도와 23차년도의 조사방법 분포	122
<표 5- 2> 22차와 23차 모두 웹조사 응답자(6명)	123
<표 5- 3> 22차, 23차 웹조사 응답자의 인구학적 특성	123
<표 5- 4> 가구원 수에 따른 개별 가구원 응답 조사방법 차이	124
<표 5- 5> 동일한 조사방법으로 응답한 가구의 조사방법	125
<표 5- 6> 조사방법에 따른 가구원 수, 자녀 수 특성(면접조사와 전화조사)	126
<표 5- 7> 조사방법에 따른 응답가구 특성(면접조사와 전화조사) ..	126
<표 5- 8> 조사방법에 따른 가구주 특성(면접조사와 전화조사)	127
<표 5- 9> 조사방법에 따른 개인응답자 특성	128
<표 5-10> 주요 변수 설명	134
<표 5-11> 전화조사 및 대면조사 비교-개인 일자리 및 소득 관련	

변수	136
<표 5-12> 전화조사 및 대면조사 비교-건강 및 생활습관 관련 응답	138
<표 5-13> 전화조사 및 대면조사 비교-생활 및 일자리 만족도	139
<표 5-14> 조사방식 영향 추정: 중간값 및 이상적 답변 선택 경향	143
<표 5-15> 웹조사 및 대면조사 비교-개인 소득 및 일자리 관련 변수	144
<표 5-16> 웹조사 및 대면조사 비교-건강 및 생활습관 관련 변수	145
<표 5-17> 웹조사 및 대면조사 비교-생활 및 일자리 만족도	146
<표 5-18> 분석 결과 요약- 개인 소득 및 일자리 관련 변수	150
<표 5-19> 분석 결과 요약- 개인 건강 및 생활습관 관련 변수	150
<표 5-20> 분석 결과 요약- 생활만족도 및 직장만족도 관련 변수	151
<표 6- 1> 8차 조사(WPS2019)의 조사방식별 사업체 현황	156
<표 6- 2> 8차 조사(WPS2019)의 Web 조사 응답 사유	157
<표 6- 3> 8차 조사(WPS2019) 조사방식별 첫 문항과 마지막 문항의 응답시기	158
<표 6- 4> 8차 조사(WPS2019)의 조사방식별 조사성공률	160
<표 6- 5> 사업체패널의 조사구조	161
<표 6- 6> 8차 조사(WPS2019)의 조사기간 현황	162
<표 6- 7> 8차 조사(WPS2019)의 조사방식별 응답자 특성	163
<표 6- 8> 8차 조사(WPS2019)의 조사방식별 사업체 특성	164
<표 6- 9> 8차 조사(WPS2019)의 조사방식별 자료수정률	166
<표 6-10> 8차 조사(WPS2019)의 조사방식별 항목무응답률	167
<표 6-11> 조사방식별 응답률에 대한 기술통계량	169
<표 6-12> 조사방식, 응답자 및 사업체 특성에 대한 기술통계량 (n=2,443)	170

<표 6-13> 조사방식과 응답기간 간의 관계 분석	171
<표 6-14> 조사방식별 자료수정 문항 수와 항목무응답 문항 수에 대한 기술통계량	173
<표 6-15> 조사방식과 응답의 질과의 관계 분석	174
<표 6-16> 분석 데이터셋 구축 현황(패널자료 구축)	175
<표 6-17> 패널자료의 기초통계량	177
<표 6-18> 패널 이중차분법을 활용한 분석결과(n=1,125, t=2)	178

그림목차

[그림 2-1] 가족력 생성 과정 개요	10
[그림 3-1] 1998년 KLIPS 98년 표본과 경찰조사의 연령 분포	54
[그림 3-2] 2009년 KLIPS 98년 표본과 경찰조사의 연령 분포	54
[그림 3-3] 2020년 KLIPS 98년 표본과 경찰조사의 연령 분포	55
[그림 3-4] 2009년 KLIPS 09년 표본과 경찰조사의 연령 분포	70
[그림 3-5] 2015년 KLIPS 09년 표본과 경찰조사의 연령 분포	71
[그림 3-6] 2020년 KLIPS 09년 표본과 경찰조사의 연령 분포	71
[그림 3-7] 2020년 KLIPS 18년 표본과 경찰조사의 남성 연령 분포	81
[그림 3-8] 2020년 KLIPS 18년 표본과 경찰조사의 여성 연령 분포	82
[그림 5-1] 노동패널 웹조사 시행 과정	121
[그림 5-2] 설문조사 데이터 품질 측정 지표	131
[그림 5-3] 조사그룹별 연도별 응답분포:(문항) 현재 건강상태	141
[그림 5-4] 조사그룹별 연도별 응답분포:(문항) 직장만족도 1: 지금 직장은 다닐 만하다	141

요 약

1. 서 론

패널 자료는 동일한 표본을 장기간 조사해야 하는 특성상 시간이 지남에 따라 자료의 품질에 다양한 문제가 발생하게 된다. 한국노동패널 및 사업체패널조사의 연혁이 쌓임에 따라 이러한 문제를 해결하기 위해 지금까지 총 10차례의 연구를 진행해 왔다. 본 연구에서는 노동패널 가족력 자료 생성, 경제활동인구조사와의 비교를 통한 노동패널 고용지표의 신뢰성 검토, 행정 데이터와 서베이 데이터의 연계 문제, 한국노동패널조사 시 면접조사, 전화조사, 그리고 웹조사 등 조사방법에 따른 결과의 차이 분석, 그리고 사업체패널조사에 있어서 CAWI가 응답률 및 자료의 질에 미치는 영향 등 총 5개 주제를 다룬다.

제2장은 KLIPS의 Sub-Data Set 중 하나인 가족력 데이터를 구성하는 방법과 이렇게 해서 만들어진 가족력 데이터의 구조, 기초 통계량 등에 관한 기술적인 내용을 담고 있다. 이를 통해 향후 개인의 혼인상태 변화, 가구원의 분가, 사망, 가족 관계의 식별 및 변화를 보다 쉽게 분석할 수 있으며, 동태적인 분석 시 데이터 정확도 및 신뢰도를 높일 수 있을 것으로 사료된다.

제3장은 노동패널이 경찰조사와 어느 정도 일관성을 가지는지를 검토하고 괴리가 발생하면 어느 항목, 어느 인구집단에서 발생하는지, 그 원인이 무엇인지를 밝히고자 하였다. 본 연구에서 얻은 성과는 향후 한국노동패널조사의 문제점을 개선하고 자료의 품질을 제고하는 데 유용하게 활용될 것으로 기대된다.

제4장은 행정데이터와 서베이데이터를 연계·활용하는 해외사례

를 조사함으로써 국내에서의 행정데이터와 서베이데이터 연계를 위한 시사점을 제공하고자 하였다. 또한 행정데이터와 서베이데이터를 연계·분석한 연구사례를 조사함으로써 행정데이터와 서베이데이터의 연계가 가져올 잠재적 가치를 파악하였다.

제5장에서는 가구조사의 방식에 따른 조사 결과의 차이를 검증해 보았다. 면접조사와 전화조사, 웹조사를 응답하는 집단별 특징의 차이를 분석하는 한편 22차에 면접조사로 응답하였던 개인들이 23차에 전화조사로 응답하였을 때, 혹은 웹조사로 응답하였을 때, 조사도구별로 응답값에 차이가 있는지, 조사결과의 신뢰 정도를 주요 변수별로 분석해 보았다.

제6장에서는 사업체패널 8차 조사에 나타난 CAPI와 CAWI의 응답률, 응답의 질(quality) 혹은 응답 성향(propensity) 등을 비교하여 Web 조사방식이 응답자의 사업체패널조사 참여를 저해하는지, 조사 방식에 따라 응답의 내용에 체계적 차이가 존재하는지 등에 대해 분석해본 후 사업체패널조사에서 CAWI 방식의 향후 활용 수준을 가늠해 보았다.

2. 노동패널 가족력 자료 생성

KLIPS는 조사 연혁이 쌓이면서 장기간의 추적 조사와 표본 추가 등으로 인해 자료 안에서의 가구 동학을 파악하는 것이 점점 어렵고 복잡해지게 되었다. KLIPS의 표본추적 원칙은 원표본 가구뿐 아니라 원가구가 속한 분가가구까지 모두 추적 조사하는 형태이기 때문에 어려움이 가중된다. 표본이 복잡한 구조를 가지다 보니 KLIPS 자료를 횡단면으로 활용하는 경우에는 비교적 쉽게 가구 정보를 활용할 수 있지만, 장기 시계열 혹은 가구의 동학이 반영되어야 하는 종단 분석의 경우에는 자료의 구조 파악이 어려울 뿐 아니라 복잡한 가족 관계가 얽혀 있기 때문에 분석 자료의 구성을 위해 적잖은 시간과 노력을 필요로 하게 된다. 이에 제2장에서는 KLIPS 가족력 데

이터를 작성하여 생애사적 관점에서 KLIPS의 응답자 개인의 가족 관계에 변화를 가져오는 중요한 사건(Event History)을 보다 쉽게 추적할 수 있도록 하였다.

작성된 KLIPS 가족력 자료는 크게 2가지 유형으로 구분된다. 첫 번째 유형은 가족관계를 기준으로 구성된 자료로 각 차수 응답자를 기준으로 형성된 파일(이하 Family Data)이다. 두 번째 유형은 개인의 생애사적 관점에서 가족관계 변화 시점에 초점을 맞추어 작성된 일종의 이력파일(Event History Calendar, 이하 EHC)이다. 이러한 KLIPS 가족력 데이터만을 활용하여 혼인 및 출산과 관련된 특징, 한 부모 가구의 특징, 부모를 부양하는 기혼유배우자 가구의 특징, 그리고 분가 연령 및 사유 등 가족과 관련된 주요 통계량을 시범적으로 측정하여 제시하는 한편 활용법과 주의할 점 등을 밝혔다.

향후 가족력 자료의 활용을 통해 개인의 혼인상태 변화, 가구원의 분가, 사망, 가족 관계의 식별 및 변화를 보다 쉽게 분석할 수 있으며, 무엇보다 장기에 걸친 동태적인 분석 시 자료 작성 시간을 크게 단축시키고, 데이터 정확도 및 신뢰도를 높일 수 있을 것으로 사료된다. 또한, 데이터에 키(key)변수를 다수 포함하여, 기존의 가구용 데이터, 개인용 데이터, 직업력 데이터 등과 쉽게 연계하여 활용이 가능할 것이다.

3. 노동패널 고용지표의 신뢰성 검토 : 경제활동인구조사와 비교

제3장에서는 한국노동패널(KLIPS)에서 제공하는 기초적인 노동 시장 지표, 즉, 경제활동참가율, 고용률, 실업률, 종사상 지위에 따른 취업자 수 등을 성별, 연령별로 살펴보고 통계청의 경제활동인구조사에서 나타나는 지표와 비교하여 괴리의 발생 여부와 원인을 분석하고 개선 방안을 모색하였다. 두 자료 간의 비교는 KLIPS의 대표성을 판단하는 데 있어 필요한 자료가 될 뿐 아니라 우리나라 노동시

장의 특징을 이해하는 데 있어서도 유용한 정보를 제공할 수 있다.

KLIPS 자료와 경활조사를 비교할 때 가장 큰 차이를 보이는 코호트는 KLIPS 98년 표본이다. KLIPS 98년 표본은 2020년에 이미 22년이 경과하였기 때문에 표본 손실과 고령화가 많이 진행되어 가중치를 적용함으로써 대표성을 확보하기에는 한계가 있는 것으로 보인다. 연령을 제한하지 않았을 경우 KLIPS 98년 표본의 경제활동참가율과 고용률은 경활조사에 비하여 각각 8%와 6% 작은 것으로 나타난 반면, 연령을 15~64세로 제한할 경우 경제활동참가율과 고용률의 차이는 4%와 3%로 줄어들었다.

한편, KLIPS 09년 표본과 KLIPS 18년 표본은 대표성에서 큰 차이는 보이지 않고 있다. 성별 차이, 연령 분포, 종사상 지위, 가중치 적용의 효과 등에서도 그 결과가 아주 유사하게 나타나고 있다. 단, 이들 표본은 경활조사와 비교하여 여전히 경제활동참가율, 고용률이 4~6% 정도 낮게 추정되고 있으며, 그 이유는 취업자 수 자체가 과소하게 포함되어 있기 때문이다. 특히, 다른 연령층에 비하여 15~34세의 고용률이 현저하게 낮게 나타나는 현상을 보이고 있다.

KLIPS 표본에 대한 개선점은 첫째, 경활조사의 연령 분포에 맞게 KLIPS 98년 표본의 가중치를 조정하는 것이다. 두 자료 사이에 가장 큰 차이를 보이는 것은 65세 이상이다. 그리고 남성의 경우 35~54세 인구의 비중이 과소하게 나타나며 여성의 경우는 남성과 달리 15~24세, 35~44세의 비중이 경활조사보다 낮게 나타나고 있다. 결국, KLIPS 98년 표본이 경활조사의 지표와 근접하려면 고령자의 가중치를 낮추고, 젊은 층의 가중치를 높여야 한다는 것을 보여준다.

둘째, KLIPS 09년 표본과 18년 표본의 가장 큰 문제점은 취업자 수 자체가 작게 포착되고 있다는 것이다. 특히, 남성보다는 여성의 고용률이 낮게 나타나고, 15~34세의 고용률이 낮게 나타나며, 임시일용직의 수가 작게 포착되고 있어 설문조사 과정에서 이들 계층의 취업자 수를 좀 더 많이 확보하려는 노력이 필요할 것이다.

4. 행정데이터와 서베이데이터의 연계

제4장에서는 주요 선진국의 행정데이터와 서베이데이터의 연계 사례를 살펴보고 그러한 연계가 가져올 잠재적 가치를 검토하였다. 북유럽국가의 경우는 개인의 고유식별자를 이용하여 행정데이터를 체계적으로 수집하고 있으며, 서베이데이터도 이러한 개인의 고유식별자를 포함, 서베이데이터와 행정데이터의 연계가 직관적이고 용이하다. 대표적으로 덴마크의 경우 개인 데이터 가공에 대한 법률(Action Processing of Personal Data)에 의해 모든 데이터는 덴마크 통계청 내에 저장되어 있고 특정 권한이 부여된 환경에서 연구자는 온라인 접근을 통해 데이터를 연구하고 분석할 수 있다. 덴마크의 행정데이터와 서베이데이터 연계활용 방식은 직관적이고 이해하기 쉬우며, 데이터에 포함된 정보가 다양하다. 이는 덴마크 통계청이 개인의 CPR번호와 사업체의 고유번호를 이용하여 행정데이터와 서베이데이터를 수집·관리하기 때문이다. 스웨덴, 노르웨이, 핀란드 등의 데이터 연계활용 역시 덴마크와 기본적으로 유사하며, 덴마크와 마찬가지로 데이터의 연계활용이 활발히 이루어지고 있다.

반면, 호주와 미국은 고유식별자가 없거나 데이터 연계에 활용하기 어려운 실정이라 주로 이름과 주소 등의 개인정보를 이용한 확률적 기록연계 방법에 의해서 서베이데이터와 행정데이터를 연계·활용하고 있다. 확률적 기록연계는 대개 약 80~90% 수준의 매치 성공률을 보이며 확률적 기록연계에 의한 행정데이터와 서베이데이터 연계가 실제로 실현 가능함을 보여준다. 미국의 경우 사회보장번호(Social Security Number)가 미국 인구조사국(US Census Bureau)의 데이터 연계에 있어서 중요한 역할을 해 왔으나 응답자들이 신원도용 위험을 인식하기 시작하면서, 사회보장번호 제공 의향이 감소해 왔고 미국 인구조사국은 확률적 기록연계(Probabilistic Record Linkage) 방법으로 데이터 연계방식을 전환하였다. 미국 인구조사국의 개인식별 검증체계(Person Identification Verification System, 이

하 PVS)를 이용하여 이루어지는 미국의 여러 데이터의 연계는 확률적 기록연계를 적용, 레퍼런스 파일에 매치된 개인에 보호된 식별키(protected identification key, 이하 PIK)를 부여함으로써 다양한 소스의 데이터와의 연계를 가능하게 하고 있다.

행정데이터와 서베이데이터 연계의 잠재적 가치는 다양한 행정데이터를 활용할 수 있다는 점과 서베이데이터의 다양한 변수를 활용할 수 있다는 점이다. 행정데이터는 출생·사망, 노동시장 및 프로그램 참여, 의료 및 교육 서비스 이용, 소득·납세·소비 등 다양한 정보를 비교적 오랜 기간에 대하여 제공할 수 있는 반면, 서베이데이터는, 비록 관측기간의 한계는 존재하지만, 시간 사용, 자기보고 건강 상태, 식습관 등 행정데이터에서 수집하지 않는 보다 다양한 정보를 연구에 활용할 수 있다. 단, 연구자는 이렇게 연계한 데이터가 모집단을 대표하지 않을 수 있다는 점을 유념해야 한다. 특히, 연계성공률이 주요 변수와 연관된 경우, 연계데이터를 이용한 추정결과에 편의가 존재할 수 있음을 고려해야 한다.

5. 조사방법에 따른 조사결과 검증 연구 : 한국노동패널조사 면접조사, 전화조사, 웹조사 비교

노동패널조사는 1998년 조사 시작 때부터 대면 면접조사를 원칙으로 하여 진행해 왔다. 그러나 2020년 코로나19의 영향으로 대면 조사가 사실상 불가능해지는 경우가 발생하면서, 대면 조사 원칙은 유지 하면서, 과거 부득이한 경우에만 사용한 전화조사를 오히려 적극 권유하면서, 노동패널 조사방법에 변화가 발생하였다. 그 결과, 매년 3% 이내에 이루어졌던 전화조사는 18%까지 상승하였고, 매년 95% 이상을 차지하던 면접조사는 80%까지 급감하였다. 또한 2019년에 시범 도입한 웹조사는 73건이 조사되어 작지만 그 비중이 점점 증가하고 있는 추세이다. 이에 면접조사, 전화조사, 웹조사의 조사방법의 차이가 조사결과의 차이를 발생하는지 검증이 필요하였다.

신뢰할 수 있는 패널자료는 한 개인이 주변 여건이 변화되지 않았다면, 연도별로 안정적으로 동일하게 응답되어야 한다. 22차에 면접조사로 응답하였던 개인들이, 23차에 전화조사로 응답하였을 때, 혹은 웹조사로 응답하였을 때, 조사도구별로 응답값에 차이가 있는지, 조사결과의 신뢰 정도를 주요 변수별로 분석해 보았다. 검증결과, 다음과 같은 결론에 도달하였다. 첫째, 면접조사와 전화조사, 웹조사를 응답하는 각각의 집단은 다른 특성을 가지고 있으므로, 조사도구별 개별 분석은 지양하여야 한다. 둘째, 과거 웹조사를 응답했던 응답자가 다음해 조사에서 웹조사만을 고집하는 경향이 크지 않으므로, 웹조사를 면접조사의 보조적 수단으로 계속 사용하는 것에 무리가 없어 보인다. 셋째, 조사도구의 다양화에 대한 조사 결과는 일부 변수에서 통계적으로 유의한 차이를 보인다. 객관적인 판단이 가능한 문항에서는 뚜렷한 차이가 발생하지 않았지만, 주관적인 판단 문항 중 현재 건강상태, 보통사람 대비 건강상태, 음주 여부, 흡연 여부 등은 조사도구에 따라 통계적으로 유의한 차이가 발생하였다. 따라서 이들 문항의 사용에 각별한 주의가 필요하다.

6. Web 조사방식이 응답률 및 자료의 질에 미치는 영향 : 사업체 패널조사를 중심으로

제6장은 사업체패널 8차 조사에 나타난 CAPI와 CAWI의 응답률, 응답의 질 등을 비교하여 Web 조사방식이 응답자의 사업체패널조사 참여를 저해하는지, 조사방식에 따라 응답의 내용에 체계적인 차이가 존재하는지 등에 대해 검증하였다.

본 연구는 조사방식과 응답률 및 자료의 질에 대한 관계를 t-검정과 회귀분석을 통해 살펴보았을 뿐 아니라 인과적 효과를 추정하고자 7차 조사에서 CAPI로 응답한 사업체만을 대상으로 패널자료를 구축하여 이중차분법을 통해 분석하였다. 응답률은 CAWI 조사가 CAPI 조사에 비해 낮은 것으로 나타났다. 자발 CAWI 조사는 CAPI

조사에 비해 조사성공률이 11.5%p 낮고 전체 설문에 대한 응답기간은 4.6일 긴 것으로 확인되었다. 또한 8차 조사 자료를 회귀분석한 결과와 7,8차 패널자료를 이중차분법으로 분석한 결과 모두 CAWI가 CAPI에 비해 응답기간이 각각 4.6일, 6.8일 통계적으로 유의하게 길게 나타나, 다시 한 번 이를 확인하였다.

다음으로, CAWI 조사로 생산된 데이터의 품질은 CAPI와 비교한 결과 유의미한 차이가 없었다. 주요 변수에 대한 자료수정률과 항목무응답률을 자발 CAWI와 CAPI에 대해 비교한 결과, 자료수정률은 12개 중 2개의 변수에서 통계적으로 유의한 차이가 있었으나 항목무응답률은 10개 변수 모두 유의미한 차이가 없어 전반적으로 자료의 질에서 차이가 난다고 하기는 어려웠다. 횡단면 자료를 활용하여 조사방식과 자료수정 및 항목무응답 문항 수와의 관계를 분석한 결과, 통계적으로 유의한 결과를 보이지 않았다. 패널자료를 이중차분법으로 분석한 결과 역시 CAWI는 CAPI에 비해 자료수정 문항 수에서 유의한 차이를 보이지 않았다.

사업체패널조사에서 Web을 전면적으로 활용한다면 응답률이 낮아질 것으로 예상되는바, 패널이탈이 심해지는 경우 패널조사의 의미를 상실할 수 있다. 따라서 사업체패널조사는 현재의 방문조사 원칙을 유지하면서 상황에 따라 Web을 보완적으로 활용하는 방식이 가장 적합한 것으로 보인다. 이때 조사방식을 혼합하더라도 사업체패널자료의 품질 측면에서는 큰 차이가 없을 것으로 예상된다.

제1장 서론

본 연구는 「패널자료 품질개선 연구 시리즈」의 제 11차 연구이다. 패널 자료는 동일한 표본을 장기간 조사해야 하는 특성상 시간이 지남에 따라 자료의 품질에 여러 가지 문제가 발생하게 된다. 표본이탈이나 무응답이 가장 대표적이고, 특정 구간에 응답이 집중되는 히핑(heaping), 모집단 분포의 변화나 표본추가에 따른 가중치의 조정, 측정오차의 보정과 자료의 신뢰성 검증, 임putation(imputation)을 통한 결측치의 추정 등 다양한 부문에서 지속적인 개선이 필요하다. 한국노동패널 및 사업체패널조사의 연혁이 쌓임에 따라 이러한 문제를 해결하기 위해 2012년 첫 번째 패널자료 품질개선 연구를 시작하였으며 지금까지 총 10차례의 연구가 진행되었다. 그동안 진행된 연구의 주제는 <표 1-1>에 제시하였다.

2021년 「패널자료 품질개선 연구(XI)」는 노동패널 가족력 자료 생성, 경제활동인구조사와의 비교를 통한 노동패널 고용지표의 신뢰성 검토, 행정 데이터와 서베이 데이터의 연계 문제, 한국노동패널조사 시 면접조사, 전화조사, 그리고 웹조사 등 조사방법에 따른 결과의 차이 분석, 그리고 사업체패널조사에 있어서 CAWI가 응답률 및 자료의 질에 미치는 영향 등 총 5개 주제를 다루고 있다.

제2장은 KLIPS의 Sub-Data Set 중 하나인 가족력 데이터를 구성하는 방법과 이렇게 해서 만들어진 가족력 데이터의 구조, 기초 통계량 등에 관한 기술적인 내용을 담고 있다. 이는 생애사적 관점에서 KLIPS의 응답자 개인의 가족 관계에 변화를 가져오는 중요한 사건(Event History)을

2 패널자료 품질개선 연구(XI)

〈표 1-1〉 패널자료 품질개선 연구(I~ X)의 세부 연구주제

패널자료 품질개선 연구(I)

- 사업체패널자료 가중치조정 연구

패널자료 품질개선 연구(II)

- 개인조사자료와 사업체조사자료 간 임금분포의 히핑현상 비교
- 조사방법과 히핑의 관계
- 노동패널 표본이탈 연구
- 사업체패널조사에 나타난 표본이탈 현황 및 대체표본에 대한 분석
- 고령화연구패널조사 건강영역의 품질개선 방안

패널자료 품질개선 연구(III)

- 소득 히핑의 보정
- 측정오차에 대한 통계적 보정방법론 연구
- 표본이탈 교정을 위한 계량경제학 방법
- Nonignorable panel attrition에 대한 보정 연구
- Paradata를 활용한 패널 무응답 보정을 위한 가중치 부여 방안

패널자료 품질개선 연구(IV)

- 동태적 패널모형에서 표본이탈 교정
- Paradata를 이용한 무응답 자료 회귀분석
- 사업체패널조사의 표본설계 관련 연구
- 한국에서의 자료검증 연구의 필요성에 대하여

패널자료 품질개선 연구(V)

- 사업체패널조사의 시점 및 표본 대표성
- 사업체패널조사의 표본이탈과 대표성
- 표본 이탈률의 크기와 편향의 관계

패널자료 품질개선 연구(VI)

- 사업체패널조사의 시점 및 표본대표성
- 사업체패널조사의 표본이탈과 대표성
- Propensity 점수를 이용한 사업체조사의 대표성 연구
- 표본 이탈률의 크기와 편향의 관계
- 2016 사업체패널조사의 추가표본설계 및 표본추출
- 노동패널 항목 무응답 처리방법 보완
- 사업체패널조사의 신뢰도와 타당도 검토
- 노동패널 자료 소득세액 추정

〈표 1-1〉의 계속

패널자료 품질개선 연구(VII)

- 한국노동패널조사 표본 추가 연구
- KLIPS 보육 관련 설문조사의 문제와 데이터 품질제고 방안
- 소득세 추정 모형의 필요성과 활용 방안
- 한국노동패널조사와 경제활동인구조사에 나타난 노동시장 활동 비교
- 한국노동패널조사 웹서베이 도입 가능성 검토를 위한 연구
- 웹을 활용한 조사의 효과에 대한 실증연구: 노동패널 CAWI 실험 데이터의 분석

패널자료 품질개선 연구(VIII)

- 제6차 사업체패널조사 가중치 산출 연구
- 웹을 활용한 조사 효과에 대한 실증연구 II
- 출산력변수 형성방법 및 문제점 진단
- 사업체패널조사의 경영의사결정 참여 설문문항들의 효과적 활용 탐색

패널자료 품질개선 연구(IX)

- 한국노동패널조사 신규 패널 특성 분석
- 한국노동패널조사를 이용한 지표 개발
- 사업체패널조사의 가중치 활용법에 대한 고찰
- 내생적 표본이탈을 고려한 비균형패널 분석법
- 보안데이터(restricted-use data) 사용 및 관리에 관한 해외 사례

패널자료 품질개선 연구(X)

- 사업체-근로자 연계자료 구축 연구
- 한국노동패널조사 추가표본에 대한 가중치 연구
- 한국노동패널조사 소득자료 비교: 표본별 및 가계동향조사를 중심으로
- 웹을 활용한 조사 자료 검증: 노동패널자료를 중심으로
- 노동패널조사 교육력 자료 구축 방법 및 문제점 진단

보다 쉽게 추적할 수 있도록 돕기 위한 일종의 기준 파일과 같은 것이다. 그동안 장기간의 추적 조사와 표본 추가 등으로 표본 규모가 크게 확장되면서, KLIPS 자료 안에서의 가구 동학은 더욱 파악하기 어렵고 난해한 상황에 놓였다. 가족력 데이터 구축을 통해 개인의 혼인상태 변화, 가구원의 분가, 사망, 가족 관계의 식별 및 변화를 보다 쉽게 분석할 수 있으며, 무엇보다 장기에 걸친 동태적인 분석 시 자료 작성 시간을 크게 단축

시키고, 데이터 정확도 및 신뢰도를 높일 수 있을 것으로 사료된다.

제3장은 노동패널이 경황조사와 어느 정도 일관성을 가지는지를 검토하고 괴리가 발생하면 어느 항목, 어느 인구집단에서 발생하는지, 그 원인이 무엇인지를 밝히고자 하였다. 노동패널은 1998년부터 시작하여 현재 23차까지 이어오면서 초기 표본의 이탈이 발생하였고, 이를 보정하기 위하여 2009년(12차)과 2018년(21차)에 새로운 표본을 추가하였다. 이 과정에서 경황조사 표본과 괴리를 보일 가능성이 높아졌으며, 그로 인해 노동패널의 대표성에 대한 우려가 발생할 소지가 있다. 본장에서는 두 자료 사이에 가중치의 차이를 반영한 기초적인 노동시장 지표, 즉, 경제활동참가율, 고용률, 실업률, 종사상 지위에 따른 취업자 수 등을 성별, 연령별로 비교해보고 차이가 발생하는 원인 및 해결 방안을 제시하였다.

제4장은 행정데이터와 서베이데이터를 연계·활용하는 해외사례를 조사함으로써 국내에서의 행정데이터와 서베이데이터의 연계 가능성을 모색하고자 하였다. 사회과학 연구뿐만 아니라 다양한 분야에서 행정데이터를 활용한 연구가 꾸준히 증가하고 있다. 또한, 정부에서 정책 수립 시 의견기반정책 중심에서 증거기반정책 중심으로 변화하면서 다양한 정책의 근거를 마련하기 위해서 적극적으로 대처하고 있으며, 이에 행정데이터와 서베이데이터의 연계는 이러한 변화와 대처에 대한 대안 중 하나라고 할 수 있다. 그러나 행정데이터와 서베이데이터의 연계에 대한 국내 연구는 상대적으로 부족하다. 본장에서는 연계 가능성을 고려할 때, 연계 가능한 데이터의 존재와 매치 성공률을 고려하였다. 동시에 행정데이터와 서베이데이터를 연계·분석한 연구사례를 조사함으로써 행정데이터와 서베이데이터의 연계가 가져올 잠재적 가치를 파악하였다.

제5장에서는 가구조사 방식에 따른 조사 결과의 차이를 검증해 보았다.

2020년은 코로나19가 시작된 해로, 대면조사를 기본으로 하는 노동패널 조사환경에 많은 어려움이 발생하였다. 기존 패널 가구들이 면접원이 가정에 방문하는 것을 극도로 꺼려 대면조사를 거절하거나, 조사 방문 시기를 연기하는 경우가 많았다. 이에 그동안 부득이한 경우에만 제한적으로 허용하던 전화 조사나 웹조사 등의 비중이 증가하게 되었다. 대면조사 방법을 기본으로 조사되었던 노동패널조사에 전화조사방법과 웹조사방

법이 혼재되면서, 조사 도구에 따른 조사결과와 신뢰성 검토 필요성이 발생하였다. 본 장에서는 면접조사와 전화조사, 웹조사를 응답하는 집단별 특징의 차이를 분석하는 한편 22차에 면접조사로 응답하였던 개인들이 23차에 전화조사로 응답하였을 때, 혹은 웹조사로 응답하였을 때, 조사도구별로 응답값에 차이가 있는지, 조사결과와 신뢰 정도를 주요 변수별로 분석해 보았다.

마지막으로 제6장에서는 Web을 활용한 조사(CAWI)가 사업체패널조사의 응답률 및 자료의 질에 미치는 영향을 분석해 보았다. 종단 성격을 갖는 대규모의 사업체 단위 조사인 사업체패널조사에 Web 방식을 도입하게 되면, 컨택부터 설문 시작까지의 시간 단축, 비용 절감, 면접원에 의해 발생하는 오차의 감소 등과 같은 장점이 있으나 무응답률 혹은 항목무응답률의 증가, 기존의 컴퓨터를 활용한 조사인 CAPI와 CAWI를 동시에 활용함에 따라 모드(mode)에 의한 오차 발생 등과 같은 단점도 존재한다. 본 연구에서는 사업체패널 8차 조사에 나타난 CAPI와 CAWI의 응답률, 응답의 질(quality) 혹은 응답 성향(propensity) 등을 비교하여 Web 조사 방식이 응답자의 사업체패널조사 참여를 저해하는지, 조사방식에 따라 응답의 내용에 체계적 차이가 존재하는지 등에 대해 분석해본 후 사업체패널조사에서 CAWI 방식의 향후 활용 수준을 가늠해 보았다.

본 연구에서 얻은 성과는 향후 한국노동패널조사 및 사업체패널조사의 문제점을 개선하고 자료의 품질을 제고하는 데 유용하게 활용될 것으로 기대된다.

제 2 장 노동패널 가족력 자료 생성

제1절 서론

KLIPS는 한국의 대표적인 가구패널조사 자료로, 2019년 기준 22차에 해당하는 조사가 누적되어 있다. 즉 1998년 KLIPS 1차 조사 당시 태어난 아이는 2019년에는 22세의 어엿한 성인이 되었으며, 1998년 개인용 설문 의 응답 최소 연령인 만 15세였던 청소년은 이제 만 37세가 되어, 어찌면 결혼과 출산을 이미 경험했을 수도 있다. 10년이면 강산이 변한다는데, KLIPS는 강산이 두 번 변하는 시간 동안 꾸준히 조사를 이어오며, 방대한 양의 자료를 축적하였다. 또, KLIPS는 22년을 지속해오는 과정에서 표본 마모를 보완하고 표본 대표성을 확보하기 위해 12차(2009년)와 21차(2018년)에 추가표본이 더해지면서 표본 규모가 확장되었다.¹⁾ 이를 통해 2018년 통합표본 기준 KLIPS의 표본은 12,134가구에 달하는 명실상부한 대규모 가구패널조사 자료가 되었다.

장기간의 추적 조사와 표본 추가 등으로 표본 규모가 크게 확장되면서, KLIPS 자료 안에서의 가구 동학은 더욱 파악하기 어렵고 난해한 상황에 놓였다. KLIPS의 표본추적 원칙은 원표본 가구뿐 아니라 원표본 가구에

1) KLIPS 홈페이지, 표본추출, <https://www.kli.re.kr/klips/contents.do?key=142> (인출일: 2021. 7. 20).

속했던 원가구원을 추적 조사하는 형태이기 때문에 혼란은 더욱 가중된다. KIPLS의 추적조사 원칙을 간략히 설명하자면, 1차년도 KLIPS에 조사된 98원표본 가구(5,000가구)에 속한 모든 가구원은 이후 추적조사의 대상이 되며, 해당 가구원이 분가하여 다른 가구로 나가더라도 분가가구로 추적조사의 대상이 된다.²⁾ 반면, 이후 분가한 원가구원(A)과 결혼하여 10차년도(2007년)에 분가가구에서 처음 조사된 가구원(B)은 원가구원이 아니기 때문에, 원가구원과 함께 사는 동안은 지속적인 조사대상이 되지만, 원가구원과 분리되어 다른 가구로 다시 분가가 될 경우 더 이상 추적의 대상은 되지 않는다.³⁾ 즉 표본설계 당시의 원표본 가구에 속했던 원가구원만이 지속적인 추적 대상이 된다.

그런데, 이때 KLIPS의 경우 원표본 가구는 98원표본, 09통합표본, 18통합표본으로, 총 3가지 형태가 있다. 앞서 예시에서 A는 98원표본의 원가구원으로 이후 지속적인 추적조사 대상이지만, B는 98원표본이 아니라서 추적조사 대상은 아니었다. 문제는 A가 속한 분가가구가 12차년도(2009년) 조사에 성공하여 해당 분가가구가 09통합표본이 되었다면, B의 경우도 09통합표본의 원가구원이 되어 이후로는 추적조사의 대상이 된다는 점이다.⁴⁾

이처럼 표본이 복잡한 구조를 가지다 보니 KLIPS 자료를 횡단면으로 활용하는 경우에는 비교적 쉽게 가구 정보를 활용할 수 있지만, 장기 시계열 혹은 중단 연구를 위해 자료를 구성하는 과정은 녹록치 않은 작업을 요한다. 특히 가구의 동학이 반영되어야 하는 중단 분석의 경우에는 자료의 구조 파악이 어려울 뿐 아니라 복잡한 가족 관계가 얽혀 있기 때문에, 분석 자료의 구성을 위해 적잖은 시간과 노력을 필요로 하게 된다.

그럼에도 불구하고, 앞서 언급한 바와 같이 22년이라는 장기간에 걸친 자료가 축적됨에 따라 KLIPS는 가구 동학 및 부와 계층의 세대 간 이진 등에 관한 면밀한 고찰이 가능한 연구 자료이다. 또한 생애주기상에서의 개인과 가구의 중요한 사건과 변화에 관한 연구도 얼마든지 가능하다. 예

2) 김유빈 외(2020), 「한국노동패널 1~22차년도 조사자료 User's Guide」, pp.4~5.

3) 김유빈 외(2020), 「한국노동패널 1~22차년도 조사자료 User's Guide」, pp.4~5.

4) 김유빈 외(2020), 「한국노동패널 1~22차년도 조사자료 User's Guide」, pp.11~15.

를 들어, C라는 원가구원이 학령기 동안 어떠한 과정을 거쳐 결혼, 출산에 이르게 되는지, 이를 통해 노동시장 성과가 어떻게 변화하는지 등을 파악할 수 있으며, 부모 세대의 부가 자녀 세대에 어떠한 방식으로 이전되며, 어떤 영향을 미치는지 등 다양하고도 심도 깊은 연구를 수행하기에 KLIPS는 매우 풍부한 정보를 담고 있다.

다만, 이러한 연구를 위해서는 KLIPS의 구조에 대한 깊은 이해와 데이터 작업을 위한 상당한 시간과 노력이 요구된다. 이에 KLIPS에 익숙하지 않은 유저들이 보다 쉽게 KLIPS의 가족력 데이터를 활용할 수 있도록 돕기 위해서, KLIPS 가족력 데이터(일명 Family Data)를 만들어 제공하고자 한다.

KLIPS Data Set 중 개인의 이력에 초점을 두고 작성된 데이터는 직업력 데이터가 대표적으로, KLIPS 직업력 데이터는 일자리 이력을 중심으로 거대한 하나의 파일로 구성되어 있다. 한편, 2020년에는 교육력(educational history) 데이터를 구축한 바 있다.⁵⁾ 본 연구는 교육력 데이터와 유사하게 KLIPS Sub-Data Set으로 가족력 데이터를 구축하여 제공하는 데 목적이 있다.

KLIPS의 가족력 데이터는 생애사적 관점에서 KLIPS의 응답자 개인의 가족 관계에 변화를 가져오는 중요한 사건(Event History)을 보다 쉽게 추적할 수 있도록 돕기 위한 일종의 기준 파일과 같은 것이다.⁶⁾ 본고는 KLIPS의 Sub-Data Set 중 하나인 가족력 데이터를 구성하는 방법과 이렇게 해서 만들어진 가족력 데이터의 구조, 기초 통계량 등에 관한 기술적인 내용을 담고 있다.

KLIPS 가족력 데이터 자료는 KLIPS 1차년도(1998년)부터 22차(2019년)까지의 가구용 설문 중 가구원 정보가 주된 자료이며, 혼인 관련 변수는 개인용 설문조사 자료를 사용하였다.

5) 김유빈 외(2020), 「패널자료 품질개선 연구(X)」, pp.109~131.

6) 즉, KLIPS의 가족력 데이터는 아주 간단한 가구 동학 연구의 경우에는 해당 자료만으로 분석이 가능하나, 대체로 KLIPS의 기존의 Data Set(가구용, 개인용, 직업력) 등의 정보(예컨대, 가구소득, 노동시장 참여 현황 등)를 추가적으로 연결하여 분석하여야 한다.

제2절 가족력 자료 구축

1. 가족력 자료 생성 과정 개요

KLIPS 가족력 자료는 크게 2가지 유형으로 구분된다. 첫 번째 유형은 가족관계를 기준으로 구성된 자료로 각 차수 응답자를 기준으로 형성된 파일(이하 Family Data)이다. 두 번째 유형은 개인의 생애사적 관점에서 가족관계 변화 시점에 초점을 맞추어 작성된 일종의 이력파일(Event History Calendar, 이하 EHC)이다.

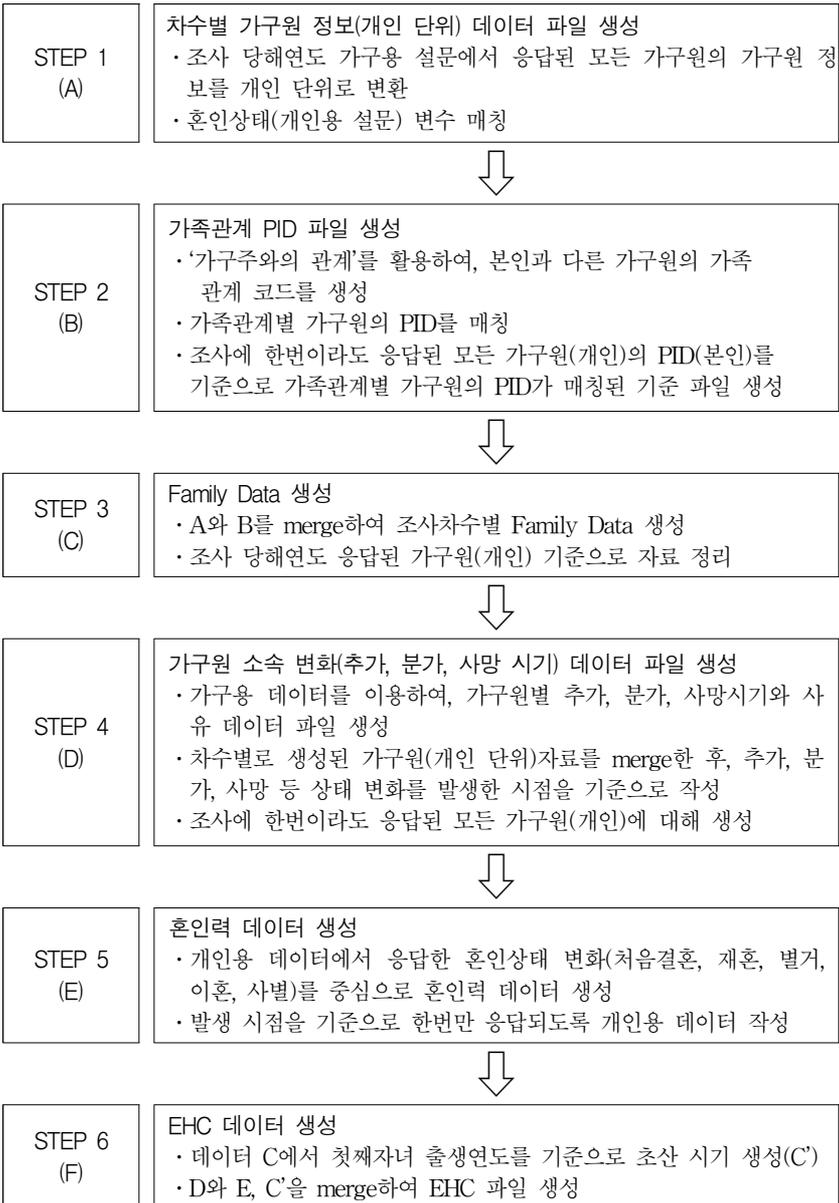
먼저, KLIPS의 Family Data를 생성하는 과정은 크게 6단계로 구분하여 살펴볼 수 있다. 1단계는 기존의 가구용 데이터를 활용하여 가구원 정보 데이터 파일(파일A)을 만드는 과정으로, 조사 당해연도 가구용 설문에서 응답된 모든 가구원에 대한 정보를 차수별로 저장한다. 2단계는 가구용 설문의 ‘가구주와의 관계’를 활용하여, 가구원별로 본인을 기준으로 가족 관계를 다시 재정리하는 과정(파일B)이다. 마지막으로 3단계에서는 파일A와 파일B를 머지(merge)하여, 조사차수별로 Family Data를 생성(파일C)하는 단계로, 파일C가 최종적인 Family Data이다. 이하 제2항부터 제4항까지는 각 단계별 Family Data 생성 과정에 대한 보다 자세한 설명을 제시하였다.

다음으로, 개인의 생애사적 관점에서 가족 구성 변화 시점을 중심으로 작성된 EHC파일도 총 3단계에 걸쳐 작성된다. 먼저, 가족력 데이터 작업의 4단계는 EHC파일 작성의 1단계로, Family Data와 유사하게 가구용 데이터의 가구원 정보를 이용하여 본인의 추가, 분가, 사망 등의 상태를 생성하는 과정이다. 차수별로 형성된 자료를 옆으로 머지하여, 한번이라도 응답한 모든 가구원(개인)에 대해서 한 번의 사건은 한 번만 기록되도록 자료(파일D)를 생성한다. 다음으로 5단계에서는 개인용 데이터에서 조사된 개인의 혼인상태 변화 변수를 활용하여, 개인별 혼인 이력 파일을 작성한다(파일E). 마지막으로 Family Data에서 이미 만들어진 첫째자녀

10 패널자료 품질개선 연구(XI)

의 출생연도를 활용하여 초산 연도를 생성(파일C')하고, 파일들을 연계하여, EHC 데이터를 작성한다.

[그림 2-1] 가족력 생성 과정 개요



2. STEP 1 : 가구원 정보 데이터 파일 생성

KLIPS 가족력 데이터의 1단계는 기존의 가구용 데이터를 활용하여 가구원 정보 데이터 파일을 만드는 과정으로, 혼인상태를 제외한 대부분의 정보는 가구용 데이터의 가구원 정보로부터 추출된다. 가구원 데이터 파일은 가구용 데이터에 가로로(옆으로) 붙어 있는 가구원 정보를 가구원별로 세로로 변환한 파일이다. 이같은 가구원 정보 파일은 개인용 데이터에 붙어 있는 가구원 정보와 일치하는 정보로, 자료 생성 과정 또한 이와 일치한다.

다만, 개인용 데이터는 개인용 설문 응답자를 기준으로 생성된 파일로, KLIPS의 개인용 설문 응답 대상자는 만 15세 이상의 개인이므로 만 15세 미만의 가구원과 개인용 설문에 응답하지 않은 가구원의 정보는 없다. 반면 가구용 자료에는 가구 내 모든 가구원이 응답되어 있으며, KLIPS 가족력 데이터는 모든 가구원 정보를 활용한 개인 단위 데이터(이하 가구원 데이터)를 기준으로 하여 생성된다.

〈표 2-1〉 가구원 데이터 생성 시 활용된 변수

변수 설명	KLIPS_H의 변수명	비고	변수 설명	KLIPS_H의 변수명	비고
가구원 PID	h_0241~h_0255		98원가구원여부	h_0201~h_0215	1~2차 값 없음
성별	h_0241~h_0255		09통합 원가구원	h_4201~h_4215	1~11차 값 없음
양/음	h_0281~h_0295		18통합 원가구원	h_5201~h_5215	1~20차 값 없음
생년	h_0301~h_0315		당해연도 가구번호	hhid01~hhid22	
생월	h_0321~h_0335		지역	h_0141	가구정보
생일	h_0341~h_0355		가구원 수	h_0150	가구정보
만나이	h_0361~h_0375		가구주와의관계	h_0261~h_0275	
학교	h_0661~h_0675		동거여부	h_0381~h_0395	
이수여부	h_0681~h_0695		비동거이유	h_0401~h_0415	
졸업년도	h_0741~h_0755	1~4년도 값 없음	혼인상태	p_0501	개인용데이터
졸업월	h_0761~h_0775				

12 패널자료 품질개선 연구(XI)

가구원 데이터에는 가장 주요한 변수인 가구원 PID를 비롯하여, 가구원별 성별, 생년월일, 학력, 원가구원 여부, 거주지역, 가구원 수, 가구주와의 관계, 동거여부 등의 가구원 정보가 활용되었다. 이때, 혼인상태는 가구용 설문이 아닌 개인용 설문에서 조사되는 변수로, 개인용 자료를 매칭하여 자료를 생성하였으며, 이로 인해 가구원 데이터에서 알 수 있는 혼인상태는 개인용 응답자에 한한다는 점에 유의해야 한다. 가구원 데이터 생성에 활용된 변수는 <표 2-1>과 같다.

한편, 가구원 데이터는 조사 당해연도의 해당 가구 내 가구원을 기준으로 생성되므로, 사망 혹은 분가하여 해당 가구용 데이터에서 PID 이외의 가구원 정보가 조사되지 않은 가구원은 제외된다.⁷⁾ 각 차수별로 조사된 가구원 수와 혼인상태를 알 수 있는 가구원 수는 <표 2-2>와 같다. 22차(2019년) 기준 응답된 가구원은 총 30,051명이며, 혼인상태를 알 수 있는 개인용 설문 응답자는 23,225명이었다.

<표 2-2> 가구원 데이터 관측치

(단위: 명)

차수	모든 가구원	개인용 설문 응답자	차수	모든 가구원	개인용 설문 응답자
1차	17,504	13,314	12차	19,047	14,489
2차	15,753	12,037	13차	18,771	14,118
3차	14,662	11,187	14차	18,587	13,899
4차	14,411	11,051	15차	18,661	13,996
5차	14,308	10,965	16차	18,650	13,887
6차	14,955	11,540	17차	18,651	13,168
7차	15,250	11,660	18차	18,756	14,011
8차	15,210	11,580	19차	18,731	14,202
9차	15,416	11,756	20차	18,824	14,475
10차	15,413	11,855	21차	31,272	23,972
11차	15,213	11,734	22차	30,051	23,225

7) 분가가구에서 응답된 경우에는 분가가구의 가구원으로 가구원 데이터에 포함.

3. STEP 2 : 가족관계 PID 파일 생성

2단계에서는 개별 가구원(이하 본인)을 기준으로 본인과 다른 응답자들과의 가족관계를 매칭한 가족관계 데이터 파일을 생성한다. 현재 KLIPS의 가족관계는 가구를 기준으로 각 가구원이 어떠한 관계인지를 나타내는 형태로 구성되어 있다. 따라서, KLIPS 가족력 데이터의 생성을 위해서는 가구원별로 본인을 기준으로 가족관계를 새롭게 정리하는 과정이 필요했다.

예를 들어, 가구번호가 1번인 가구에 개인고유번호(PID)가 101인 가구가 있고, 그의 배우자(102)가 있다고 생각해보자. 현재 KLIPS 가구용 데이터의 가구원 정보에는 가구원 101의 가구주와의 관계가 10(가구주), 가구원 102의 가구주와의 관계는 20(가구의 배우자)으로 표시된다. KLIPS 가족력 데이터는 가구원별로 본인을 기준으로 하는 가족관계를 생성해야 하므로, 가구용 데이터상의 가구주와의 관계 조합을 이용하여, 본인을 기준으로 각 가족관계에 해당하는 가구원 PID가 입력되도록 변환하였다. 즉 배우자 PID라는 새로운 변수에 해당하는 가구원의 개인고유번호(PID)가 입력되도록 하였다. 제시된 예의 경우 가구원 101의 배우자 PID에는 102가 입력되고, 가구원 102의 배우자 PID에는 101이 입력되는 방식이다.

이때, KLIPS의 가구주 정의는 응답자가 가구주라고 지정한 1인이기 때문에, 특별한 규칙성을 갖지 않는다. 또, 가족의 형태가 매우 다양하여 가구주와의 관계를 기준으로 가족관계를 매칭하는 것이 쉬운 일은 아니었다. 해당하는 가족관계별 조합은 별첨의 자료를 참고하기 바란다. 별첨 자료에 제시된 가족관계 조합은 KLIPS에 응답된 적이 있는 가족관계만을 기준으로 산출한 것으로, 본인-가구주와의 관계만 총 72개의 상이한 가족관계가 발견되었으므로, 수백개의 가족관계 조합을 보였다.

한편, 가족력 데이터 생성을 위해 본인을 기준으로 새롭게 정리되는 가족관계는 부모와 자녀, 배우자에 국한하여 정리되었다. 즉 조사 당시 동일 가구 내에 본인의 조부모 혹은 형제, 자매들이 있었다 할지라도, 가족력 데이터에서는 이들의 정보는 확인이 어렵다. 횡단면상에서의 가구 단

위의 가족관계를 보여주는 가구용 데이터와 달리 가족력 데이터는 개인 단위로 구성되기 때문에, 조부모 및 형제자매의 정보까지를 포괄하기에는 한계가 있었다. 가족력 데이터를 활용하는 대부분의 연구는 부모-본인-자녀의 직계가족에 대한 정보를 주로 활용할 것이며, 실제 이 외의 가족관계 조합에 응답되는 관측치가 많지 않다는 점 등을 감안할 때 이와 같은 구성이 적합하다고 보인다. 연구주제에 따라 조부모의 정보를 꼭 활용해야 하는 경우에는 가족력 데이터의 부모 PID를 이용하여, 부모의 부모(조부모) PID와 가구원 정보를 끌어와서 활용하면 된다.⁸⁾

가족관계 조합의 일부만 예시로 제시하면 <표 2-3>과 같다. 즉 본인이 가구주인 경우 배우자는 20, 부모1은 5(아버지), 부모2는 6(어머니), 자녀1은 11(첫째자녀) 등이다. 이때 부모의 경우 본인이 가구주이거나 가구의 배우자인 경우에는 가구주와의 관계가 아버지, 어머니 등과 같이 성별을 포함하는 관계이지만, 본인이 가구의 첫째자녀(가구주와의 관계 11)인 경우에는 본인의 부모는 10(가구주)으로 부인지 모인지를 알 수가 없다. 따라서, 부모1(가구주-10)과 부모2(가구의 배우자-20)로 코딩한 이후 해당 가구원의 성별을 기준으로 부 PID, 모 PID로 재분류하였다.

<표 2-3> 가족관계 매칭표

본인	배우자	부모1	부모2	자녀1	자녀2	..
10	20	5	6	11	12	..
20	10	7	8	11	12	..
11	21	10	20	111	112	..
:	:	:	:	:	:	..
21	11	-	-	111	112	..
22	12	-	-	121	122	..
23	13	-	-	131	132	..
:	:	:	:	:	:	..
111	211	11	21	1111	1112	..
112	212	11	21	-	-	..

8) 단, 본인 기준 형제자매의 정보는 기존의 가구용 데이터를 활용하여 자료를 새롭게 만들어서 활용해야 한다.

가족관계 PID 파일은 본인의 가족관계별로 해당 가족관계인 개인의 PID가 기록된 파일이다. 배우자를 제외하고 부모와 자녀 관계가 차수가 변한다고 해서 변화하는 것은 아니므로, 모든 차수의 자료를 통합하여, 해당 가족관계에 첫 번째로 응답된 PID를 사용하였다. 즉, 둘 사이의 관계가 부자 관계라고 응답된 최초 응답값이 활용되었으며, 결과적으로 KLIPS 가구용 데이터에서 한 번이라도 응답된 적이 있는 가구의 가족 관계가 매칭된 형태의 단 하나의 파일이 생성된다. 이때 가구주와의 관계가 모든 차수에서 10번 이하인 경우 등 가족관계를 특정할 수 없는 가구의 경우에는 제외되었으므로 주의를 요한다. 가족관계 PID 데이터에 포함된 가구는 총 42,738명이었다.

자료의 구성을 보다 쉽게 이해할 수 있도록 하나의 예를 들어보자. 5차년도에 5100번 가구로 분가한 원가구원인 101인 남성 가구주와 가구주의 배우자(5102), 첫째자녀(5103), 둘째자녀(5104)가 있다고 하자. 가족관계 PID 데이터 파일은 가구원 단위로 구성되므로, 우선 본인 PID는 <표 2-4>와 같이 4개가 생성된다.⁹⁾ 다음으로 101의 배우자는 5102이므로, 배우자 PID는 5102, 자녀1(첫째자녀) PID는 5103, 자녀2(둘째자녀) PID는 5104가 된다. 한편, 본인이 5013이라면 아버지 PID는 101, 어머니 PID는 5102가 될 것이다.

만약 101이 4차년도까지는 1번 가구에서 조사되었고, 당시 101이 본인의 어머니 102와 함께 살고 있었다면, 101의 모 PID는 102가 된다. 이때, 4차년도 이전 조사에서 1번 가구의 가구주가 102였다면(101은 첫째자녀라고 가정), 본인 PID가 102인 관측치의 자녀1 PID는 101이 된다. 그러나, 4차년도 이전 조사에서 1번 가구의 가구주가 101이었다면, 102의 가구주와의 관계는 6(가구주의 어머니)일 것이다. 이 경우 101의 모가 102인 것은 알 수 있지만, 101이 102의 몇 번째 자녀인지는 알 수가 없는 상황이다. 이처럼 가구주와의 관계 응답값이 10 미만인 경우 본인과 다른 가족과의 관계가 명확하게 식별되지 않은 경우로 KLIPS 가족력 데이터에서 해당 값은 제외된다.

9) 해당 가구가 여러 개 차수에서 응답되었더라도 하나의 PID를 중심으로 구성됨.

〈표 2-4〉 가족관계 PID 파일 예시

본인 PID	배우자1 PID	부 PID	모 PID	자녀1 PID	자녀2 PID	..
101	5102	-	102	5103	5104	..
5102	101	-	-	5103	5104	..
5103	-	101	5102	-	-	..
5104	-	101	5102	-	-	..
:	:	:	:	:	:	..
102	-	-	-	-	-	..
:	:	:	:	:	:	..

한편, 배우자는 부모, 자녀 관계와 달리 다른 사람으로 바뀌는 경우가 발생한다. KLIPS 가족관계 PID 데이터 파일에 포함된 가구원은 42,738명이며, 이 중에 차수가 지나감에 따라 서로 다른 배우자가 발견된 경우(재혼 이상)는 206명이 있었으며, 총 3명의 서로 다른 배우자까지 발견된 경우(3번의 결혼)가 4명 있었다. 이에 따라 배우자 PID는 총 3개까지 생성하였다.

4. STEP 3 : Family Data 생성

3단계는 앞서 생성된 ‘가구원 정보 데이터 파일’과 ‘가족관계 PID 파일’을 머지하여 가족력 데이터의 기본틀을 갖추는 단계이다. KLIPS 가족력 데이터는 가구원(개인) 단위로 자료가 생성되며, 차수별로 생성하였다. 이는 논문의 작성을 위해서는 가족력 데이터만으로 분석이 어렵고, KLIPS의 나머지 데이터의 정보와 결합하여 활용해야 하는 경우가 다수이기 때문에, 이를 보다 쉽게 수행할 수 있도록 하기 위한 것이다. 예컨대, 가족력 데이터와 가구소득을 결합하여 분석을 하려는 경우 차수별로 자료가 구축되어 있는 것이 편의성이 높다고 보았기 때문이다. 또, 가족력 데이터에는 가족관계에 해당하는 가구원의 정보(예: 성별, 연령, 학력 등)를 함께 수록하여 유저 편의를 제고하고자 하였는데, 이러한 정보가 해마다 변동되는 점을 반영하기 위해 차수별로 작성되었다.

다만, 주의할 점은 가족력 데이터에 붙어 있는 각 가족관계에 해당하는 가구원의 PID 정보는 차수와 무관하게 생성된 자료로, 언젠가 한번은 해당 가족관계의 가구원에 대한 정보를 알 수 있음을 표시하는 기준 변수라는 점이다. 즉, 앞서 예시에서 101과 102는 6차년도에 다른 가구에 속한 가구원으로, 6차년도에 102번 가구원 혹은 1번 가구의 조사 성공 여부와 무관하게 101의 모 PID에는 102가 입력되어 있다. 이때, 6차년도에 102가 조사에 성공했는지 여부는 모의 성별 변수의 결측 여부를 통해 확인을 할 수 있는데, 6차년도에 102가 꼭 해당 가구(5100번 가구)가 아니더라도 어딘가에서 응답되었다면 102의 성별 등 가구원 정보는 결측치가 아닌 값을 갖는다.¹⁰⁾

한편, 유저 편의를 위해 해당 가구원이 꼭 동일 가구가 아니더라도 당해연도 조사에 성공한 경우에도 가구원 정보를 붙여 두었으나, 경우에 따라 동일 가구에 속한 경우만을 분석의 대상으로 하는 경우를 고려하여, 본인과 동일 가구 여부 변수를 새롭게 만들어 추가하였다. 이 변수는 각 가족관계에 해당하는 가구원이 현재 본인과 동일 가구에 속한 가구원인지를 표시하는 변수로, 당해연도 가구번호를 통해 확인이 가능한 사항이다. 이때 주의할 점은 본인과 동일 가구에 속한다는 것이 반드시 동거한다는 것이 아닌, 가구경제 등을 공유하는 가족관계라는 점이다.¹¹⁾ 즉 주말부부 등으로 비동거 상태이더라도 동일 가구번호를 가진 가구원은 동일 가구로 구분됨을 유의하여야 한다.

이에 더하여, 유저 편의를 제고하기 위해 차수별로 새롭게 생성하여 추가된 변수로는 한부모 가구 여부 변수이다. 한부모 가구는 본인의 연령이 18세 미만인 가구원만을 대상으로 생성된 변수로, 만 18세 미만의 본인의 가족관계에서 부 혹은 모¹²⁾가 한 명만 발견되는 경우¹³⁾이다. 한부모 가

10) 당해연도 조사된 가구원의 정보는 모두 붙여져 있으므로, 성별 이외의 변수를 활용해도 무방하긴 하나, 문항에 응답하지 않은 경우가 있으므로, 조사에 성공했다면 결측치가 전혀 없는 성별 변수를 사용할 것을 권장함.

11) 이는 KLIPS의 가구 정의에 따른 것임.

12) 성별 기준(가족관계 PID 기준 아님에 주의).

13) 네이버 지식백과, 심리학용어사전 한부모 가정, <https://terms.naver.com/entry.naver?docId=2094195&cid=41991&categoryId=41991> (인출일: 2021. 7. 22).

구여부는 1. 양부모 가구, 2. 편부, 3. 편모로 구성되어 있으며, KLIPS 가족력 데이터의 생성 과정에서의 한계로 인해 부모 둘 다 없는 가구는 없다는 점에 주의를 요한다. 이는 본인을 기준으로 부모와 자녀 관계에 해당하는 직계 가족만을 기준으로 가족력 데이터가 작성된 것에 기인하며, 조손가구 혹은 기타 친인척과 동거하는 경우 등을 알고 싶은 경우에는 부득이 기존의 가구용 데이터를 활용하여 해당 가구를 식별해내는 작업을 해야한다. 또, 앞서 언급한 바와 같이 KLIPS 가족력 데이터는 가구원(개인) 단위로 자료가 구축되어 있으므로, 한부모 가구 내에 두 명 이상의 자녀가 있었던 경우 누적적으로 집계될 수 있다는 점 또한 주지해야 한다. 이는 다자녀 가구, 세대 구성 등의 변수 생성 시에도 동일하게 발생하는 현상으로, 가구 단위의 분석을 원하는 경우 KLIPS 가족력 데이터 가구 단위(가구번호를 기준으로 가로로 된 데이터)로 변환하여 활용하거나, 기존의 가구용 데이터를 활용하여 새롭게 생성해야 한다.

KLIPS 가족력 데이터의 변수 조합은 <표 2-5>와 같다. 먼저 본인을 비롯한 가족관계에 해당하는 가구원의 고유번호(PID) 변수는 본인이 PID, 첫 번째 배우자 SPID1, 두 번째 배우자 SPID2, 세 번째 배우자 SPID3을 순으로 지정하였다. 부는 FPID, 모 MPID로 지정하였으며, 자녀는 출생순위에 따라 CPID1부터 CPID9까지를 지정하였다. 가족관계 PID는 당해연도 조사 성공 여부와 무관하게 모든 차수에 동일하게 수록된 정보라는 점을 재차 강조하는 바이다.

KLIPS 가족력 데이터는 처음 F로 시작하도록 하였으며, F 다음에 이어지는 한자리(백단위) 혹은 두자리(천단위) 숫자는 본인과의 가족관계를 나타낸다. 그리고, 십의 자리와 일의 자리 숫자는 각 가구원(개인)의 정보에 해당하는 내용으로 구성된다.



천단위와 백단위의 숫자는 다음과 같은 순서로 배정하였다. 먼저, 본인의 정보는 백단위 숫자가 0을 갖는다. 뒤에 다시 언급하겠지만 성별은 02

〈표 2-5〉 가족관계를 나타내는 변수명

변수 설명	PID 변수명	F천백단위 변수명	변수 설명	PID 변수명	F천백단위 변수명
본인	PID	F0_ _~F0_ _	자녀1	CPID1	F11_ _~F11_ _
배우자1	SPID1	F1_ _~F1_ _	자녀2	CPID2	F12_ _~F12_ _
배우자2	SPUD2	F2_ _~F2_ _	자녀3	CPID3	F13_ _~F13_ _
배우자3	SPID3	F3_ _~F3_ _	자녀4	CPID4	F14_ _~F14_ _
부	FPID	F6_ _~F6_ _	자녀5	CPID5	F15_ _~F15_ _
모	MPID	F7_ _~F7_ _	자녀6	CPID6	F16_ _~F16_ _
			자녀7	CPID7	F17_ _~F17_ _
			자녀8	CPID8	F18_ _~F18_ _
			자녀9	CPID9	F19_ _~F19_ _

의 값을 가지므로, 본인의 성별은 F002가 된다. 배우자와 부모는 백단위 변수명을 갖도록 배치하였으며, 백단위 변수가 1인 경우가 배우자1, 배우자2가 2, 배우자3이 3, 부는 6, 모는 7을 갖는다. 한편 자녀는 9명인 경우 까지 있으므로, 천단위를 갖도록 배치하여 11이 첫째자녀, 12가 둘째자녀의 가구원 정보를 의미한다.

십의 자리와 일의 자리 숫자는 각각 다음과 같은 가구원 정보를 의미한다. 성별은 02, 생년은 04, 학교는 10, 동거여부 22, 혼인상태 21, 한부모 가구여부 63 등이다. 그러므로, 본인의 성별은 F002이며, 모의 성별은 F702, 첫째자녀의 성별은 F1102 등으로 코딩되었다. 가구원 정보는 본인과 동일 가구인지 여부와 무관하게 당해연도에 응답한 가구원이면, 해당 가족관계의 가구원 정보가 식별될 수 있도록 붙여두었다.

단, 가구원 정보 중에서 가구원 수와 가구주와의 관계는 다른 가구원의 경우에는 정보가 없으며, 본인의 경우만 확인되는 변수이다. 가족력 데이터는 본인과 다른 가구원이 어떤 관계인지를 변수를 통해 식별하도록 만든 자료이므로, 다른 가구원의 가구주와의 관계는 큰 의미가 없기 때문이다. 그러나, 본인이 가족 내에서 어떠한 지위에 있는지를 가늠할 수 있도록 하기 위해 본인의 가구주와의 관계 코드는 남겨두었다. 반면, 본인과

〈표 2-6〉 가구원 정보 변수 지정

변수 설명	가족력 데이터의 지정번호	변수 설명	가족력 데이터의 지정번호	비고
성별	02	09통합	32	
양/음	03	18통합	33	
생년	04	당해연도 가구	41	
생월	05	지역	42	
생일	06	가구원 수	43	본인만
만나이	07	가구주와의관계	51	본인만
학교	10	동거여부	22	본인만
이수여부	11	비동거이유	23	본인만
졸업년도	12	혼인상태	21	
졸업월	13	본인과 동일가구여부	61	본인제외
원가구여부	31	한부모 가구여부	63	만 18세 미만 자녀, 본인만

동일가구 여부는 본인의 경우에는 없는 변수이다. 정리하자면, F061, F1143, F751 등의 변수는 KLIPS 가족력 데이터에는 없는 변수이다.

5. STEP 4 : 가구원 소속 변화 데이터 파일

4~6단계의 작업은 EHC 데이터 파일 생성 과정이다. 먼저 4단계는 앞서 1단계 작업과 유사한 작업으로 가구용 데이터의 가구원 정보를 개인 별 자료로 변환하는 작업이다. 단, 1단계 작업이 각 가구원의 인적 정보를 중심으로 자료를 변환한 것이라면, 4단계 작업은 해당 가구원의 가구원 상태(가구원 소속)의 변화를 중심으로 작성된다. 즉, 해당 가구에 해당 가구원이 추가된 시점과 사유, 분가한 시점과 사유, 사망한 시점과 사유 등이 주요 변수이다. 현재 KLIPS의 가구용 데이터는 가구를 기준으로 가구원 구성의 변화를 확인하기 위해서 가구원별 추가, 분가, 사망과 관련된 정보를 수록하고 있는데, 이를 활용하여 각 가구원이 기존의 가구에 새롭게 추가되거나 반대로 분가하는 경우, 사망 등으로 더 이상 조사되지 않

는 경우 등을 개인별 시점에서 본인이 속한 가구 소속의 변화로 변환한 파일이 파일D이다.

6. STEP 5 : 혼인력 데이터 생성

한편, 5단계에서는 개인용 데이터를 활용하여 개인별 혼인상태의 변화를 중심으로 혼인력 데이터를 생성하였다. KLIPS의 개인별 혼인상태 변화는 개인용 설문과 신규용 설문에서 조사되고 있으며, 조사에 처음 응답한 경우(신규용 설문)에는 혼인상태와 관련된 이력을 전부 조사하나, 기존 응답자(개인용 설문)는 기존 설문 응답 이후의 변화가 있는 경우에만 응답하도록 되어 있다. 단, 1~2차년도 조사의 경우에는 신규용 설문과 마찬가지로 모든 개인용 설문 응답자에게 모든 혼인상태 변화를 응답하도록 되어 있었다. 즉, 2차년도 이전 데이터와 3차년도 이후 데이터는 설문 문항과 구조 자체가 달라, 현재 개인용 데이터에서는 통합이 불가하다.

EHC 파일은 개인의 생애사적 관점에서의 가족관계 변화에 초점이 맞추어진 데이터이므로, 개인용 데이터의 혼인상태 관련 변수들을 개인 이력 파일 형태로 다시 재정렬하고 정리하여 혼인력 데이터를 생성하였다. 우선 3차년도 이후 데이터의 경우 혼인상태 변화 여부와 내용, 변화 시점 등을 조사하고 있으므로, 변화 내용이 초혼인지, 재혼, 별거, 이혼, 사별인지 등을 구분하여 각 혼인상태에 해당하는 변화가 발생한 시점을 기준으로 혼인력 데이터를 형성하였다. 예컨대, 개인용 설문의 혼인상태 변화에서 ‘초혼’이라고 응답하고, 변화 시점이 응답된 경우, 혼인력 데이터의 초혼 시점에 해당 변화 시점이 들어가도록 하였으며, 변화 내용이 ‘이혼’인 경우 이혼 시점에 해당 시점이 입력되도록 하였다.

한편, 1~2차년도의 경우에는 설문 자체가 초혼 시점, 가장 최근 결혼 시점 등으로 되어 있어, 해당하는 혼인상태에 해당하는 변수, 즉, 초혼 시점 등에 해당 결과가 들어가도록 하였다. 따라서, 1차년도와 2차년도에 중복된 응답값이 존재하게 되어(예를 들어, 초혼 시점이 1차년도와 2차년도 모두에 응답되기 때문), 응답값을 머지하여 해당하는 혼인상태에 해당하는 최초의 응답값만 남기는 작업을 거쳤다. 이때, KLIPS의 1~2차년도

혼인상태 문항은 재혼 시기 등의 경우에는 가장 최근 재혼 시기만을 묻고 있기 때문에, 해당하는 재혼이 두 번째 결혼인지 세 번째 결혼인지는 구분할 수 없는 한계가 있다. 따라서, EHC에서는 1~2차년도 of 가장 최근의 재혼이 두 번째 결혼(재혼)으로 분류되었음을 밝혀둔다.

반면, KLIPS의 1~2차년도 자료는 별거/이혼/사별의 경우에도 사유별로 구분되지 않는 한계가 있다. 즉, 별거/이혼/사별 시기를 조사하기는 했으나 해당 시점이 별거 시점인지, 이혼 시점인지, 사별 시기인지는 알 수 없는 한계가 있다. 이에 혼인력 데이터 파일의 별거 시기, 이혼 시기, 사별 시기 등은 3차년도 이후의 응답값을 기준으로 생성된 변수이며, 1~2차년도의 경우에는 별도 변수(eh101~eh106)로 생성되어 있으므로 주의할 것을 요한다.

혼인력 데이터는 각 차수별로 혼인상태 발생 시점을 기준으로 재정렬된 데이터를 머지하여, 하나의 사건당 한 번만 입력되어 있는 이력 파일 형태로 작성한 파일이다. 이때, 초혼의 경우에는 모든 개인에게 한 번만 발생하는 사건이지만, 재혼 등의 사건은 여러 번 발생할 수 있는 사건이다. 각기 다른 시점으로 재혼 시기 등이 응답된 경우는 모두 다른 재혼으로 분류하였으며, 개인에 따라 총 3회까지 재혼상태가 발견된 경우가 있었다. 각기 다른 별거 시점이 관측된 경우 2회, 이혼은 3회, 사별 3회까지 발견되었으며, 따라서 별거 시점_1(첫번째 별거 시기), 별거 시점_2(두번째 별거 시기) 등으로 3차년도 이후 혼인상태가 변화된 경우는 모두 누락이 없도록 작성하였다.¹⁴⁾

7. STEP 6 : EHC 데이터 파일 생성

EHC 데이터 생성의 마지막 단계는 앞서 작성된 Family Data에서 초산 연령을 개인별 이력 파일로 작성하여, 파일D(가구원 소속 변화)와 파일E(혼인력 데이터)와 머지하는 과정이다. 개인별로 초산은 생애사에서 한 번만 발생하는 사건으로, Family Data의 첫째자녀의 출생연도와 일치

14) 단, 해당 변화와 관련된 배우자의 정보는 앞서 작성된 Family Data의 정보를 머지하거나, 개인용 데이터를 활용하여 식별하여야 한다.

한다. 단, Family Data는 각 차수별로 자료가 형성되어 있으므로, 사건 이력 형태의 EHC 데이터 파일의 변수로 넣기 위해서, Family Data를 옆으로 머지하여 처음으로 발견되는 첫째 자녀의 출생연도만을 끌어와서, 초산 시점을 생성하였다(파일C’).

이렇게 생애사적으로 중요한 변화와 관련된 시점을 중심으로 형성된 이력 파일들을 머지하여, 각 가구원(개인)별 사건사 이력파일인 EHC 데이터를 작성하였다. EHC 파일은 응답된 모든 차수에서 해당하는 사건사가 발생하는 시점을 중심으로 작성된 단 하나의 파일로, 각 차수별로 형성되는 Family Data, 기존의 가구용 데이터 혹은 개인용 데이터 등과 연계하여 활용 가능하며, 연계의 편의성을 제고하기 위해 각 차수별 가구번호를 붙여두었다.

EHC 파일의 변수명은 eh로 시작하며, eh 이하에 붙는 변수명 일련번호는 본인의 성별이 1, 출생연도가 2이다. 한편, 본인의 원가구원 여부, 즉 추적대상 표본인지를 표시하는 변수가 11~13번까지 배치되었다. 이어지는 20번대의 변수는 가구원으로 추가된 경우와 관련된 변수, 30번대는 사망 관련 변수, 40번대는 분가와 관련된 변수이다. 추가와 사망은 가구원 본인 기준 한 번만 발생하는 상황이므로, eh11(추가년도), eh22(추가월), eh23(추가 사유) 3개의 변수만을 갖지만, 분가의 경우에는 여러 번 분가가 발생할 수 있어, 첫 번째 분가 시기와 사유, 두 번째 분가 시기와 사유 순으로 작성되었다. 1~22차까지 자료 기준 동일한 개인이 총 3번까지 분가한 경우가 발견되어 변수는 eh41~eh49번까지 생성되었다.

한편, 60번대 변수는 결혼과 관련된 변수로, 초혼 연도 eh61, 초혼 월이 eh62이며, 재혼 연도 eh63, 재혼 월 eh64 등으로 지정하였다. 동일한 개인이 서로 다른 재혼 연도가 관측된 경우는 최대 3회까지로, 재혼 연도는 각기 eh63, eh65, eh67 총 3개가 생성되었다. 별거는 70번대 변수를 갖는데 별거 횟수가 2회까지 발견되는 개인이 있으므로, eh71~eh74까지 생성되었으며, 유사한 방식으로 이혼은 eh81~eh86, 사별은 eh91~eh96에 배치하였다. 이때 앞서 언급한 바와 같이, KLIPS 1~2차 자료의 혼인력은 별거/이혼/사별을 사유별로 구분하지 않았기 때문에, 이를 구분하여 변수 형성이 불가능하였으며, 이에 별도로 분리된 변수 eh101~eh106에

해당 시점이 수록되어 있으므로 주의를 요한다. 마지막으로 첫째 자녀의 출생연도, 즉 초산 시기는 eh201(연도), eh202(월)로 지정하였다.

〈표 2-7〉 EHC data 파일 변수명

변수 설명	지정번호	변수 설명		지정번호	비고
성별	1	분가_첫번째	년도, 월, 사유 순	41~43	
생년	2	분가_두번째	년도, 월, 사유 순	44~46	
98원가구	11	분가_세번째	년도, 월, 사유 순	47~49	
08원가구	12	초혼	년도, 월 순	61~62	
18원가구	13	재혼	년도, 월 순	63~68	재혼 3회
추가년도	21	별거	년도, 월 순	71~74	별거 2회
추가월	22	이혼	년도, 월 순	81~86	이혼 3회
추가사유	23	사별	년도, 월 순	91~96	사별 3회
사망년도	31	별거/이혼/사별	년도, 월 순	101~106	1~2차년도는 사유별 구분 안됨
사망월	21	초산	년, 월 순	201~202	
사망사유	33				

제3절 자료 특성

최종적인 KLIPS 가족력 데이터는 직업력과는 달리 각 차수별로 생성 되는데, 이는 가족력 데이터에 포함되지 않은 당해연도 조사 자료를 보다 손쉽게 붙여서 활용할 수 있도록 하기 위한 것이다. 이하에서는 KLIPS 가족력 데이터만을 활용하여 산출할 수 있는 몇가지 통계량과 활용법을 제시함으로써, KLIPS의 가족력 관련 현황을 간략히 살펴보고자 한다.

1. 가족 관계 및 가구원 구성

가. 가구원 정보 응답자

차수별로 본인을 기준으로 각 가족관계에 해당하는 가구원이 응답된 경우는 <표 2-8>과 같다. 본인과 동일한 가구에 속한 가구원이 아니더라도

<표 2-8> 차수별 본인 및 가족관계 응답자 수

(단위 : 명)

차수	본인	배우자1	배우자2	배우자2	부 PID	모 PID
1차	17,504	8,245	9		7,190	7,871
2차	15,753	7,492	13		6,497	7,155
3차	14,662	7,002	14		6,035	6,655
4차	14,411	6,947	18		5,997	6,629
5차	14,308	6,942	27		5,934	6,567
6차	14,955	7,294	35		6,206	6,900
7차	15,250	7,482	37		6,380	7,087
8차	15,210	7,489	45		6,333	7,066
9차	15,416	7,623	55		6,430	7,188
10차	15,413	7,625	68		6,443	7,207
11차	15,213	7,596	80	1	6,330	7,089
12차	19,047	9,572	93	2	7,548	8,412
13차	18,771	9,486	108	2	7,471	8,374
14차	18,587	9,422	109	2	7,427	8,346
15차	18,661	9,533	112	1	7,477	8,403
16차	18,650	9,554	116	1	7,493	8,407
17차	18,651	9,582	122	1	7,511	8,440
18차	18,756	9,642	128	1	7,603	8,544
19차	18,731	9,676	140	1	7,574	8,543
20차	18,824	9,777	149		7,658	8,631
21차	31,272	16,486	153	1	11,018	12,460
22차	30,051	15,886	150	2	10,655	12,080

도, 당해연도 조사에서 해당 가구원의 정보를 찾아볼 수 있는 경우로, 22차년도를 기준으로 총 30,051명 가운데 첫 번째 배우자가 조사된 관측치는 15,886명으로 절반을 약간 넘는다. 부가 조사된 경우는 10,655명, 모가 조사된 경우는 12,080명이었다.

한편, 출생순위별로 자녀가 조사된 경우는 22차 기준 첫째 자녀가 11,058명으로, 본인 30,051명 중 36.8%였다. 매 차수별로 첫째 자녀의 비중은 조금씩 변화하긴 하나, 30% 후반에서 40% 정도를 보였다. 둘째 자

〈표 2-9〉 차수별 자녀 응답자 수

(단위: 명)

차수	자녀1	자녀2	자녀3	자녀4	자녀5	자녀6	자녀7	자녀8	자녀9
1차	6,589	5,430	1,676	517	193	88	23	12	2
2차	5,941	4,943	1,524	459	167	74	18	11	2
3차	5,490	4,634	1,428	417	151	71	14	9	2
4차	5,494	4,631	1,404	414	157	68	13	8	2
5차	5,479	4,569	1,376	405	143	74	11	8	2
6차	5,806	4,812	1,400	398	158	62	12	8	2
7차	6,038	4,981	1,386	406	151	61	10	8	2
8차	6,037	4,948	1,365	403	144	59	12	7	2
9차	6,178	5,056	1,380	391	125	61	7	7	2
10차	6,210	5,067	1,392	388	115	59	8	6	2
11차	6,131	5,016	1,367	352	110	51	8	6	2
12차	7,365	5,984	1,590	389	116	57	8	4	0
13차	7,302	5,970	1,569	381	111	47	8	3	0
14차	7,324	5,945	1,541	369	116	44	9	2	0
15차	7,373	5,990	1,567	383	111	36	9	1	0
16차	7,414	6,048	1,533	364	105	35	7	1	0
17차	7,477	6,047	1,534	372	102	33	7	1	0
18차	7,611	6,121	1,538	367	102	32	7	1	0
19차	7,582	6,153	1,538	363	97	32	8	1	0
20차	7,703	6,236	1,542	346	93	32	6	1	0
21차	11,414	8,891	2,097	402	108	28	5	1	0
22차	11,058	8,607	2,037	391	103	28	5	1	

녀가 응답된 비중은 1차년도 기준 31.0%(17,504명 중 5,430명)이었으며, 22차 기준 28.6%(30,051명 중 8,607명)으로 감소하였다. 셋째 자녀가 응답된 경우는 크게 감소하여 1차년도 9.6%(1,676명)이었는데, 22차년도에는 6.8%(2,037명)에 그쳤다. 셋째 자녀 이상이 있는 다자녀 가구가 많지 않을 뿐 아니라, 최근으로 올수록 이 비중이 점점 더 줄어드는 경향을 발견할 수 있다.

나. 한부모 가구

만 18세 미만 아동 중 한부모 가구에 속한 아동의 비중은 1차년도 편부 가정 2.2%, 편모 가정 4.2%로 6.4%였다. 22차년도 조사에서는 편부 가정이 1.7%, 편모 가정이 4.5%로 한부모 가구의 비중은 6.2%였다. 양부모 가정의 비중은 전반적으로 90~95% 사이로 특별한 경향성을 보이는 것은 아니었다.¹⁵⁾ 다만, 한부모 가정 중 편부 가정에 비해 편모 가정의 비중이 약 2배가량 많은 수준을 보였다.

다만, 이러한 결과를 가족 해체 과정에서 부모보다 모가 자녀양육을 책임지는 비중이 높다고 해석하기에는 다소 무리가 있으며, 추가적인 분석이 필요하다. 왜냐하면, 부모의 이혼뿐 아니라 사별 등의 이유로 가정이 해체되는 경우 부모 중 누가 양육책임을 갖느냐는 부모의 성별과 무관한 일이기 때문이다. 또, 이혼 등으로 인한 가족 해체 상황에서도 원가구원에 대한 추적조사 과정에서 남성보다 여성의 조사 응답률이 높다면, 표본 이탈에 따른 효과가 반영되었을 가능성을 배제하기 어렵다. 따라서, 연구 목적에 따라 해당 시기에 양부모 가구와 한부모 가구의 차이 등을 비교하기 위한 연구라면, 아래와 같이 차수별 한부모 가구 여부를 활용하여 연구를 수행하면 되지만, 가정 해체 과정에서의 부모 양육책임 분담 등에 관한 사항을 살펴보고자 한다면 가족력 데이터를 균형패널 형태의 자료로 변환하여 분석하여야 함을 주지하는 바이다.

15) KLIPS 가족력 데이터로 산출한 한부모 가구 비중은 어디까지나 당해연도 응답자를 기준으로 하고 있으므로, 일종의 균형 패널로 구성하였을 때 한부모 가정의 비중이 어떻게 변화하는지에 대해서는 추가적인 분석을 요한다.

〈표 2-10〉 차수별 한부모 가구 비중

(단위: 명, %)

차수	전체	양부모+자녀	편부+자녀	편모+자녀
1차	4,593 (100.0)	4,301 (93.6)	100 (2.2)	192 (4.2)
2차	3,933 (100.0)	3,685 (93.7)	81 (2.1)	167 (4.3)
3차	3,502 (100.0)	3,296 (94.1)	66 (1.9)	140 (4.0)
4차	3,359 (100.0)	3,179 (94.6)	62 (1.9)	118 (3.5)
5차	3,238 (100.0)	3,046 (94.1)	73 (2.3)	119 (3.7)
6차	3,321 (100.0)	3,087 (93.0)	85 (2.6)	149 (4.5)
7차	3,386 (100.0)	3,145 (92.9)	83 (2.5)	158 (4.7)
8차	3,335 (100.0)	3,094 (92.8)	88 (2.6)	153 (4.6)
9차	3,326 (100.0)	3,082 (92.7)	90 (2.7)	154 (4.6)
10차	3,301 (100.0)	3,043 (92.2)	102 (3.1)	156 (4.7)
11차	3,259 (100.0)	3,027 (92.9)	81 (2.5)	151 (4.6)
12차	3,631 (100.0)	3,631 (91.7)	109 (2.8)	221 (5.6)
13차	3,521 (100.0)	3,521 (91.2)	122 (3.2)	217 (5.6)
14차	3,473 (100.0)	3,473 (92.0)	103 (2.7)	201 (5.3)
15차	3,457 (100.0)	3,457 (92.7)	87 (2.3)	185 (5.0)
16차	3,418 (100.0)	3,418 (93.1)	87 (2.4)	167 (4.6)
17차	3,336 (100.0)	3,336 (93.5)	83 (2.3)	148 (4.2)
18차	3,367 (100.0)	3,367 (94.1)	77 (2.2)	133 (3.8)
19차	3,296 (100.0)	3,296 (94.2)	66 (1.9)	137 (3.9)
20차	3,288 (100.0)	3,288 (94.4)	67 (1.9)	139 (3.7)
21차	5,032 (100.0)	5,032 (93.3)	90 (1.7)	270 (5.0)
22차	4,726 (100.0)	4,726 (93.8)	86 (1.7)	228 (4.5)

다. 기혼 유배우자의 부모와 동일 가구 비중

기혼 유배우자가 자신의 부모와 함께 사는 경우는 KLIPS 가족력 데이터를 활용하여 쉽게 산출이 가능하다. 가족력 데이터에서는 본인의 혼인 상태와 본인의 부모가 같은 가구번호를 가진 경우(F661, F761)로 쉽게 식별되기 때문이다. 단, 해당 변수는 동거 여부가 아니라 가계경제를 공유하는 동일 가구 내 소속이라는 점에 유의해야 하며, 따라서 동일 공간을

공유하는 동거 개념보다는 경제적인 가계 공유, 즉 부양의 의미가 좀 더 강하다.¹⁶⁾ 여기에서는 기혼 유배우자만을 대상으로 본인의 성별에 따라 자신의 부모와 동일 가구를 형성하는 비중을 살펴보았다. 기혼 유배우 남성의 경우 1차년도 부를 부양하는 비중이 4.5%, 모를 부양하는 비중이 9.9%에 달하였으나, 22차년도에는 부 부양 0.9%, 모 부양 3.4%로 크게 감소하였다. 반면, 기혼 유배우 여성의 경우에는 1차년도에 부를 부양하는 비중이 0.5%, 모 부양은 1.4%로, 22차년도에 부 부양 0.2%, 모 부양 0.9%로 감소하긴 하였으나, 처음부터 기혼 유배우 여성이 자신의 부모를 부양하는 경우가 많지 않았음을 알 수 있다.

즉, 자녀가 결혼한 이후 기혼 남성의 부모, 즉 시부모를 부양하는 문화가 기혼 여성이 자신의 부모를 부양하는 경우에 비해 여전히 좀 더 많은 비중을 차지하고 있음을 알 수 있다. 그러나, 이러한 문화는 20여 년이 지나는 동안 크게 감소하여, 기혼 남성의 부모를 부양하는 비중 자체도 크게 감소하는 경향을 보인다. 한편, 기혼 유배우 남녀 모두 부모 중에서는 부를 부양하는 경우보다 모를 부양하는 비중이 좀 더 높게 나타났다. 아래와 같은 간단한 분석만으로는 이러한 현상의 원인을 정확하게 진단하기 어려우나, 추측컨대 손자녀 돌봄 등으로 인해 모가 함께 세대를 공유하는 경우 때문일 것으로 추측된다.¹⁷⁾

하지만, 가족력 데이터를 활용하여 대가족/핵가족 여부를 정확히 파악하기는 어렵다. 왜냐하면, 가족력 데이터는 본인을 기준으로 본인의 직계 가족(부모와 자녀)만을 포함하고 있기 때문이다. 배우자의 부모가 같은 가구에 동거하는 경우나 애초에 가구주와의 관계가 (응답코드 기준) 10 미만이었던 경우가 가족력 데이터에는 포함되지 않기 때문이다. 해당 변수는 기존의 가구용 데이터를 활용하여 만드는 것이 관측치의 손실이 없이 가장 쉽게 데이터를 구성하는 방법이다.

16) 물론 자녀가 부모를 부양하는 것인지, 부모가 자녀를 부양하는 것인지는 논란의 여지가 있을 수 있다.

17) 가구 내 자녀 연령 등을 추가적으로 고려하여 분석이 가능하며, 이 경우 부모가 기혼 유배우 자녀와 함께 사는 것은 자녀의 부모 부양이라기보다는 오히려 부모가 자녀를 도와주는 형태일 가능성이 큼.

〈표 2-11〉 기혼 유배우자 중 부모와 동일 가구 비중

(단위: 명, %)

차수	기혼 유배우 남성			기혼 유배우 여성		
	전체	부	모	전체	부	모
1차	4,153	187 (4.5)	412 (9.9)	4,154	22 (0.5)	56 (1.4)
2차	3,739	164 (4.4)	386 (10.3)	3,759	25 (0.7)	61 (1.6)
3차	3,470	139 (4.0)	341 (9.8)	3,487	20 (0.6)	49 (1.4)
4차	3,447	122 (3.5)	330 (9.6)	3,451	12 (0.4)	37 (1.1)
5차	3,442	116 (3.4)	321 (9.3)	3,454	10 (0.3)	35 (1.0)
6차	3,593	107 (3.0)	313 (8.7)	3,606	13 (0.4)	36 (1.0)
7차	3,634	93 (2.6)	293 (8.1)	3,664	9 (0.3)	36 (1.0)
8차	3,600	89 (2.5)	266 (7.4)	3,657	9 (0.3)	35 (1.0)
9차	3,654	74 (2.3)	268 (7.3)	3,699	10 (0.3)	42 (1.1)
10차	3,676	77 (2.1)	268 (7.3)	3,715	8 (0.2)	43 (1.2)
11차	3,662	67 (1.8)	244 (6.7)	3,696	6 (0.2)	33 (0.9)
12차	4,507	83 (1.8)	287 (6.4)	4,587	6 (0.2)	35 (0.8)
13차	4,436	64 (1.4)	268 (6.0)	4,513	4 (0.1)	36 (0.8)
14차	4,407	65 (1.5)	254 (5.8)	4,477	7 (0.2)	35 (0.8)
15차	4,485	60 (1.3)	238 (5.3)	4,530	6 (0.1)	37 (0.8)
16차	4,458	56 (1.3)	229 (5.1)	4,504	8 (0.2)	35 (0.8)
17차	4,357	46 (1.1)	217 (5.0)	4,467	8 (0.2)	33 (0.7)
18차	4,479	52 (1.2)	211 (4.7)	4,551	10 (0.2)	31 (0.7)
19차	4,559	46 (1.0)	197 (4.3)	4,584	11 (0.2)	30 (0.7)
20차	4,614	47 (1.0)	185 (4.0)	4,635	11 (0.2)	35 (0.8)
21차	7,614	71 (0.9)	280 (3.7)	7,900	24 (0.3)	64 (0.8)
22차	7,365	68 (0.9)	251 (3.4)	7,616	18 (0.2)	68 (0.9)

2. 출산 관련 현황

이 항에서는 KLIPS 가족력 데이터를 활용하여 출산 관련 현황을 살펴 보았다. 출산 관련 현황 자료 또한 가족력 데이터 구조를 살펴 차수별로 산출이 가능하지만, 이하의 내용은 자료의 활용 가능성을 제시하기 위해 본인의 출생코호트별 분석 결과를 제시하였다. 즉, 자녀의 출산과 관련된 사항은 생애주기상에서 각 자녀에 대해 한 번만 발생하는 사건으로, 관측

되는 마지막 시점까지 해당 사건이 발생했는지 여부를 기준으로 한 번의 정보만을 필요로 한다. 따라서, 차수별로 구분되어 있는 가족력 데이터를 생애주기 형태가 되도록 옆으로 머지하여 활용하면 된다.

가. 초산 연령¹⁸⁾

먼저, 초산 연령은 Family Data를 생애주기형으로 변환하거나 EHC 파일을 활용하여 첫째 자녀의 출생년도에서 본인의 출생년도를 뺀으로써 쉽게 산출된다. 각 코호트별로 여성의 평균 초산 연령은 20대 후반 수준을 보이는데, 1940년대생은 평균 26.2세, 1950년대생은 25.3세, 1960년대 26.5세, 1970년대생 28.4세, 1980년대 28.9세로 초산 연령이 점차 늦어지는 현상이 발견되었다. 남성의 경우 1950년대생의 초산연령은 29.0세, 1960년대생 30.1세, 1970년대생 31.4세로 여성에 비해 2~4세 높았으며, 여성과 마찬가지로 최근 코호트로 올수록 초산 연령이 늦어지는 특징이 관측된다.

이때 1990년대생은 마지막 관측 시점(조사시점)인 22차(2019)년에 20대로 아직까지 초산을 경험하지 못했거나 진행되지 못한 경우가 다수이다. 마찬가지로 1980년대생 또한 30대이기 때문에 표본의 상당 부분은 출산을 하지 않았을 가능성이 높다. 이처럼 초산 연령은 해당 시점까지 출산을 한 경우에만 관측되는 변수로, 이들 코호트의 평균 초산 연령은 향후 더 높아질 가능성이 있다. 반대로, 1949년 이전에 출생한 코호트의 경우에도 KLIPS에는 관측치가 많지 않다. 이는 1998년 1차년도 조사 당시 표본에서 해당 연령이 50대 이상 연령으로 본인이 해당 가구에서 가구주의 부모 등의 위치에 있는 표본들이 많을 것이기 때문이다. 따라서, 출산과 관련된 코호트 분석에서 1950년대생부터 1970년대생 정도가 2019년 기준 출산이 거의 완료되고, 장기간의 가족력 변화를 고찰하기에 적당한 코호트라고 사료된다.

18) 단, 데이터의 클리닝을 요하는(예컨대, 10세 미만의 초산 사례 등) 관측치가 소수 발견되었으며, 본고는 이를 제외하고 산출하였다. 추후 노동패널팀에서 해당 변수에 대한 클리닝을 진행하는 과정에서 결과값이 다소 달라질 수 있음을 밝혀둔다.

〈표 2-12〉 평균 초산 연령

(단위: 명, 세)

코호트	남 성			여 성		
	전체 표본	초산 관측표본	평균 초산연령	전체 표본	초산 관측표본	평균 초산연령
1930년대 이전	385	34	32.7	748	30	27.8
1930~1939년생	1,039	192	31.4	1,391	146	27.3
1940~1949년생	1,878	761	29.6	2,110	527	26.2
1950~1959년생	2,499	1,659	29.0	2,712	1,550	25.3
1960~1969년생	3,058	2,295	30.1	2,961	2,282	26.5
1970~1979년생	3,502	2,190	31.4	3,425	2,497	28.4
1980~1989년생	2,897	951	30.6	2,839	1,366	28.9
1990~1999년생	2,437	42	23.8	2,147	99	23.6

나. 평균 자녀 수

평균 자녀 수는 남성의 경우 1950년대생 1.9명, 1960년대생 1.9명, 1970년대 1.8명이었으며, 여성의 경우 1950년대생 2.0명, 1960년대생 1.9명, 1970년대 1.9명이었다.

〈표 2-13〉 평균 자녀 수

(단위: 명)

코호트	남 성			여 성		
	전체 표본	유자녀 표본	평균 자녀 수	전체 표본	유자녀 표본	평균 초산연령
1930년대 이전	385	88	1.4	748	79	1.1
1930~1939년생	1,039	440	1.6	1,391	360	1.4
1940~1949년생	1,878	1,030	2.0	2,110	911	1.8
1950~1959년생	2,499	1,797	1.9	2,712	1,747	2.0
1960~1969년생	3,058	2,377	1.9	2,961	2,426	1.9
1970~1979년생	3,502	2,194	1.8	3,425	2,511	1.9
1980~1989년생	2,897	953	1.6	2,839	1,366	1.7
1990~1999년생	2,437	43	1.4	2,147	99	1.3

다. 다자녀 출산

가족력 데이터의 자녀 출생순위를 활용하여 다자녀 출산 변수를 생성할 수 있다. 출생순위를 활용한 다자녀 가구의 구성 방식은 해당 가구에 몇 명의 자녀가 있느냐가 아닌 한 명의 자녀가 발견되더라도 그 자녀가 몇 번째 자녀인가를 확인함으로써, 다자녀 출산 이력을 확인하는 방식으로, 다자녀 가구 비중과 약간의 차이를 보인다.

초산(첫째 자녀 출산)이 아닌 둘째아 이상 자녀의 출산 경험이 관측된 비중은 <표 2-14>와 같다. 1950년대생 중 자녀가 관찰되지 않은 비중은 32.0%로 이는 실제 무자녀였다기보다는, 1950년생이 1차년도 조사 당시만 48세로 해당 가구에서 자녀가 조사되지 않았을 가능성이 크다고 보인다. 한편, 1960년대생의 경우 2명(둘째 이상) 출산이 확인된 비중이 51.1%, 3명 이상(셋째 이상 자녀)의 출산이 발견된 비중이 10.3%로 매우 높은 비중을 보였다. 1970년대생은 2명의 자녀 출산이 확인된 비중이 40.5%, 3명 이상 자녀의 출산이 발견된 비중이 8.4%였다. 1960년대생들이 2명의 자녀를 출산한 비중이 다른 코호트에 비해 높은 특징을 보인다.

<표 2-14> 다자녀 출산

(단위: 명, %)

코호트	전체 표본	자녀없음	자녀 수		
			1명	2명	3명이상
1930~1939년생	2,430 (100.0)	1,630 (67.1)	499 (20.5)	216 (8.9)	85 (3.5)
1940~1949년생	3,988 (100.0)	2,047 (51.3)	735 (18.4)	732 (18.4)	474 (11.9)
1950~1959년생	5,211 (100.0)	1,667 (32.0)	929 (17.8)	2,009 (38.6)	606 (11.6)
1960~1969년생	6,019 (100.0)	1,216 (20.2)	1,111 (18.5)	3,075 (51.1)	617 (10.3)
1970~1979년생	6,927 (100.0)	2,222 (32.1)	1,320 (19.1)	2,806 (40.5)	579 (8.4)
1980~1989년생	5,736 (100.0)	3,417 (59.6)	1,046 (18.2)	1,080 (18.8)	193 (3.4)
1990~1999년생	4,584 (100.0)	4,442 (96.9)	98 (2.1)	40 (0.9)	4 (0.1)

라. 남아 선호

첫째아 성별에 따라 둘째아 이상을 출산하는 비중이 어떻게 다른지를

살펴보았다. 전체적으로 첫째아가 남성인 경우 첫째만 낳고 후속 출산이 발견되지 않은 경우가 29.6%, 첫째아가 여성인 경우에는 25.1%로, 후속출산 중단 비중이 다소 높게 나타났다. 첫째아가 남아인 경우 둘째를 출산한 비중은 59.4%였으며, 셋째 이상을 출산한 비중이 11.1%인 데 반해, 첫째아가 여아인 경우에는 둘째아를 출산한 비중이 56.0%, 셋째 이상을 출산한 비중이 18.9%로 특히 셋째 이상 출산 비중이 높은 특징을 보였다.

코호트별로는 첫째아가 남아인 경우 둘째아 이후 후속 출산을 중단한 비중을 기준으로, 1950년대생은 25.9%, 1960년대생은 22.6%, 1970년대생은 28.0%였다. 반면 첫째아가 여아인 경우는 1950년대생은 14.5%, 1960년대생은 17.7%, 1970년대생은 27.9%로, 코호트에 따라 큰 차이를 보였다. 다시 말해, 첫째아가 남아인 경우 후속 출산이 중단되는 비중이 여아인 경우에 비해 1950년대생의 경우 11.5%포인트나 높았으며, 1960년대생은 4.9%포인트 높았고, 1970년대생은 0.1%포인트로 전혀 차이를 보이지 않았다. 셋째아 이상을 출산한 비중 또한 첫째아가 여아인 경우가 남아인 경우에 비해 1950년대생의 경우 16%포인트 높은 데 반해, 1960년대생은 10.2%포인트, 1970년대생은 3.2%포인트 차로 줄어드는 현상이 뚜렷히 관측된다. 이를 통해 최근으로 올수록 남아에 대한 선호가 감소하고 있음을 짐작할 수 있었다.

남아 선호 현상의 변화를 좀 더 면밀히 관찰해보기 위해, 첫째아뿐 아니라 둘째아의 성별까지를 고려한 상태의 후속 출산에 대해서 살펴보면 다음과 같다. 둘째아 이상을 출산한 표본은 11,891명으로, 첫째아가 남아이고 둘째아도 남아인 경우가 3,281명, 첫째는 여아이고 둘째가 남아인 경우가 3,015명, 첫째가 남아이고 둘째는 여아인 경우 2,971명, 첫째가 여아이고 둘째도 여아인 경우가 2,624명이었다.

이때 첫째와 둘째의 성별이 남-남, 남-여, 여-남인 경우 셋째 이상을 출산한 비중이 15% 전후인 데 반해, 첫째가 여아이고 둘째도 여아인 경우에는 셋째 이상을 출산한 비중이 38.3%로 남아선호가 확연하게 관측되었다. 이를 본인의 출생 코호트를 기준으로 구분하여 살펴보면, 첫째-둘째의 성별이 여-여인 경우 셋째아를 출산한 비중이 1950년대생의 경우 54.8%에 달한 반면, 1960년대생은 40.8%, 1070년대생은 25.5%까지 감소

〈표 2-15〉 첫째 자녀 성별과 후속 출산

(단위: 명, %)

코호트	첫째 성별	전체 표본	자녀 수		
			1명	2명	3명 이상
전 체	남	9,023 (100.0)	2,668 (29.6)	5,355 (59.4)	1,000 (11.1)
	여	7,595 (100.0)	1,907 (25.1)	4,252 (56.0)	1,436 (18.9)
1930~1939년생	남	256 (100.0)	126 (49.2)	98 (38.3)	32 (12.5)
	여	82 (100.0)	43 (52.4)	21 (25.6)	18 (22.0)
1940~1949년생	남	783 (100.0)	216 (27.6)	375 (47.9)	192 (24.5)
	여	505 (3 9.2)	101 (20.0)	191 (37.8)	213 (42.2)
1950~1959년생	남	1,743 (100.0)	452 (25.9)	1,098 (63.0)	193 (11.1)
	여	1,466 (100.0)	212 (14.5)	857 (58.5)	397 (27.1)
1960~1969년생	남	2,438 (100.0)	550 (22.6)	1,679 (68.9)	209 (8.6)
	여	2,139 (100.0)	378 (17.7)	1,459 (63.5)	402 (18.8)
1970~1979년생	남	2,426 (100.0)	678 (28.0)	1,485 (61.2)	263 (10.8)
	여	2,261 (100.0)	630 (27.9)	1,415 (58.2)	316 (14.0)
1980~1989년생	남	1,234 (100.0)	537 (43.5)	593 (48.1)	104 (8.4)
	여	1,079 (100.0)	503 (46.6)	487 (45.1)	89 (8.3)
1990~1999년생	남	88 (100.0)	64 (72.7)	21 (23.9)	3 (3.4)
	여	53 (100.0)	33 (62.3)	19 (35.9)	1 (1.9)

하는 것을 볼 수 있다. 즉 남아를 낳기 위해 후속 출산을 하는 비중은 최근으로 올수록 현저히 감소하고 있다고 볼 수 있다.

다만, 1970년대생의 경우에도 첫째아-둘째아의 성별이 남-남, 남-여, 여-남인 경우에는 셋째아 출산 비중이 15% 전후에 그친 반면, 첫째아-둘째아의 성별이 여-여인 경우 10%포인트가량 높은 셋째아 출산 비중을 보여, 자녀 중 남아가 포함되기를 희망하는 남아선호가 여전히 남아 있음을 반증하고 있다. 다만, 앞으로도 출산 가능성이 높은 코호트인 1980년대생 이후 세대에 있어서는 아직까지는 이와 같은 남아선호 현상은 관찰되지 않았다. 다만, 부모의 연령을 고려할 때 1980년대생의 출산, 특히 셋째아 이상의 후속 출산으로의 이행이 아직 완료되지 않았을 가능성이 높으므로, 좀 더 시간을 두고 지켜봐야 할 것으로 사료된다.

〈표 2-16〉 첫째 및 둘째 자녀 성별과 후속 출산

(단위 : 명, %)

코호트	첫째 성별	둘째 성별	전체 표본	자녀 수	
				2명	3명 이상
전 체	남	남	3,281 (100.0)	2,763 (84.2)	518 (15.8)
	여	남	3,015 (100.0)	2,600 (86.2)	415 (13.8)
	남	여	2,971 (100.0)	2,515 (84.7)	456 (15.4)
	여	여	2,624 (100.0)	1,618 (61.7)	1,006 (38.3)
1950~1959년생	남	남	687 (100.0)	600 (87.3)	87 (12.7)
	여	남	677 (100.0)	590 (87.2)	87 (12.9)
	남	여	594 (100.0)	492 (82.8)	102 (17.2)
	여	여	564 (100.0)	255 (45.2)	309 (54.8)
1960~1969년생	남	남	975 (100.0)	856 (87.8)	119 (12.2)
	여	남	965 (100.0)	887 (91.9)	78 (8.1)
	남	여	904 (100.0)	816 (90.3)	88 (9.7)
	여	여	786 (100.0)	465 (59.2)	321 (40.8)
1970~1979년생	남	남	907 (100.0)	760 (83.8)	147 (16.2)
	여	남	871 (100.0)	749 (86.0)	122 (14.1)
	남	여	839 (100.0)	723 (86.2)	116 (13.8)
	여	여	760 (100.0)	566 (74.5)	194 (25.5)
1980~1989년생	남	남	337 (100.0)	279 (82.8)	58 (17.2)
	여	남	271 (100.0)	228 (84.1)	43 (15.9)
	남	여	360 (100.0)	314 (87.2)	46 (12.8)
	여	여	303 (100.0)	257 (84.8)	46 (15.2)

마. 분가 연령 및 분가 사유¹⁹⁾

한편, EHC 데이터 파일을 활용하여 분가 연령과 분가 사유를 간략히 살펴보면 다음과 같다. 먼저, KLIPS의 가구원(개인) 중 분가 경력이 발견

19) 5차년도 이전 가구용 데이터의 분가 관련 변수 체계가 달라 전반적인 클리닝이 진행 중이며, 클리닝 결과에 따라 결과값이 달라질 수 있다. 또, 분가 시기가 출생 연도보다 앞서는 경우 등은 데이터 오류에 해당하는 경우로, 본 연구에서는 해당하는 관측치를 제외하고 결과를 제시하였으며, 향후 클리닝이 필요한 사항이다.

된 경우는 총 5,511명으로, 평균 분가 연령은 31.3세였다. 분가 사유는 장성하여 경제적으로 독립한 경우가 2,204명(40.0%)으로 가장 많았으며, 혼인 1,975명(35.8%), 기타 701명(12.7%), 이혼 543명(9.9%) 순이었다.

분가 연령별로는 20대에 분가한 경우가 가장 많아 2,433명(44.1%)이었으며, 30대에 분가한 경우가 1,657명으로 다음을 이었다. 분가 빈도가 가장 많은 20대의 경우에는 장성하여 경제적으로 독립했다는 응답이 1,167명(48.0%)로 가장 많고, 다음으로 혼인이 1,065명(43.8%)였는 데 반해, 30대의 경우에는 혼인이 838명(50.6%)로 가장 높은 비중을 보이는 분가 사유였으며, 다음으로 장성하여 경제적으로 독립했다는 응답이 586명(35.4%)로 약간의 차이를 보였다.

한편, 10대에 분가가 발생한 경우는 532명으로, 기타 사유가 260명(48.9%)로 가장 많았는데, 이는 부모의 이혼 등으로 인해 자녀가 원가구에서 분리된 경우나 혹은 학업 등으로 인한 분리 등으로 사료된다.

〈표 2-17〉 분가 연령별 분가 사유

(단위 : 명, %)

분가연령	분가 사유					계
	혼인	이혼	경제적독립 (장성)	기타	모름/ 무응답	
전체	1,975 (35.8)	543 (9.9)	2,204 (40.0)	701 (12.7)	88 (1.6)	5,511 (100.0)
0~19세	12 (2.3)	81 (15.2)	164 (30.8)	260 (48.9)	15 (2.8)	532 (100.0)
20~29세	1,065 (43.8)	56 (2.3)	1,167 (48.0)	116 (4.8)	298 (1.2)	2,433 (100.0)
30~39세	838 (50.6)	142 (8.6)	586 (35.4)	70 (4.2)	21 (1.3)	1,657 (100.0)
40~49세	52 (11.9)	167 (38.2)	165 (37.8)	47 (10.8)	6 (1.4)	437 (100.0)
50~59세	7 (3.8)	70 (38.3)	49 (26.8)	53 (29.0)	4 (2.2)	183 (100.0)
60세 이상	1 (0.4)	27 (10.0)	73 (27.1)	155 (57.6)	13 (4.8)	269 (100.0)

바. 사망 연령 및 사망 사유

EHC 데이터를 통해 또한 쉽게 확인이 가능한 또 다른 변수인 사망 관련 현황을 살펴보면, 전체 응답자 42,738명 중 총 1,433명(3.4%)이 사망한 것으로 집계되었으며, 평균 사망 연령은 72.8세였다. 사망 사유별로는 질병이 669명(46.7%)로 가장 많고, 노환이 623명(43.5%), 사고사 106명(7.4%) 순이었다.

연령별로는 70대가 가장 많아 400명(27.9%)이었으며, 80대 397명(27.7%), 60대 233명(16.3%) 순이었다. 사망 원인별로는 질병으로 인한 사망이 30대부터 70대까지 가장 높은 비중을 보였는데, 특히 60대(187명, 80.3%)와

〈표 2-18〉 사망 연령별 사망 사유

(단위: 명, %)

사망연령	사망 사유					계
	노환	사고	질병	기타	모름/ 무응답	
전체	623 (43.5)	106 (7.4)	669 (46.7)	25 (1.7)	10 (0.7)	1,433 (100.0)
0~29세	0 (0.0)	12 (46.2)	11 (42.3)	3 (11.5)	0 (0.0)	26 (100.0)
30~39세	0 (0.0)	11 (37.9)	14 (48.3)	4 (13.8)	0 (0.0)	29 (100.0)
40~49세	0 (0.0)	21 (30.4)	47 (68.1)	1 (1.5)	0 (0.0)	69 (100.0)
50~59세	1 (0.8)	20 (16.1)	94 (75.8)	8 (6.5)	1 (0.8)	124 (100.0)
60~69세	19 (8.2)	20 (8.6)	187 (80.3)	4 (1.7)	3 (1.3)	233 (100.0)
70~79세	153 (38.3)	15 (3.8)	224 (56.0)	4 (1.0)	4 (1.0)	400 (100.0)
80~89세	306 (77.1)	6 (1.5)	82 (20.7)	1 (0.3)	2 (0.5)	397 (100.0)
90세 이상	144 (92.9)	1 (0.7)	10 (6.5)	0 (0.0)	0 (0.0)	155 (100.0)

50대(94명, 75.8%)는 질병으로 인한 사망 비중이 매우 높은 집단이었다. 70대 이후로는 노환으로 인한 사망 비중이 증가하며, 80대는 노환으로 인한 사망이 306명(77.1%), 90대는 144명(92.9%)로 높은 비중을 보였다. 반면, 20대 이하의 경우 사망자가 26명으로 많지는 않으나, 사고사가 12명(46.2%), 질병 11명(42.3%) 순이었다.

사. 초혼 연령 및 혼인상태 변화²⁰⁾

기존 개인용 데이터를 활용하여 복잡하게 산출해야 했던 혼인력 관련 분석은 가족력 데이터의 EHC 파일을 통해 쉽게 산출이 가능하다. EHC 파일은 혼인상태를 기준으로 변수가 생성되어 있기 때문이다. 먼저, 가장 간단하게 산출 가능한 초혼 관련 현황을 살펴보면, 평균 초혼 연령은 26.5세였다. 각 출생코호트별로 평균 초혼 연령의 변화를 살펴보면, 최근으로 올수록 초혼 연령이 점차 증가하는 현상이 뚜렷하게 발견된다. 1950년대생의 경우 평균 초혼연령이 25.7세였으나, 1970년대생은 28.7세로 20년 동안에 초혼 연령이 약 3세가량 늦어졌음을 알 수 있다.

한편, 1980년대생 이후는 아직 혼인으로 이행하지 않은 관측치가 많은 상태로, 평균 초혼 연령은 현재 상태에서 혼인으로 이행한 관측치만을 대상으로 하므로, 다소 낮게 나타나는 것으로 향후 5년 이상 자료가 추가적으로 축적된 이후에서야 초혼 연령의 변화를 좀 더 정확하게 관측할 수 있을 것으로 사료된다.

한편 성별 초혼 연령은 남성의 경우 평균 28.4세, 여성은 24.9세로, 남성이 여성에 비해 3.5세가량 많았다. 남성과 여성의 초혼 연령이 최근 코호트로 올수록 증가하는 양상은 동일한데, 1950년대생의 성별 평균 초혼 연령의 차이가 3.8세인 데 반해, 1970년대생의 성별 평균 초혼 연령 차이는

20) 혼인 시기가 출생 시기보다 먼저인 경우 등 데이터 클리닝을 요하는 관측치는 제외하고 분석하였다. 한편, 초혼 연령이 10세 미만으로 매우 낮은 경우도 데이터 재확인 필요하다고 보이나, 주로 1930년대 이전 출생한 경우로 초혼 등의 관습 등을 고려할 때 반드시 오류라고 보기 어려워 이런 경우는 분석에 포함하였다. 추후 노동패널팀의 자료 재확인 및 클리닝 결과에 따라, 본고의 결과값이 달라질 수 있음을 밝혀둔다.

〈표 2-19〉 코호트별 평균 초혼 연령

(단위: 명, 세)

코호트	표본	평균 초혼연령		
		전체	남성	여성
전체 평균	26,747	26.5	28.4	24.9
1930~1939년생	2,341	23.6	26.2	21.7
1940~1949년생	3,875	25.1	27.1	23.4
1950~1959년생	4,966	25.7	27.7	23.9
1960~1969년생	5,437	27.1	28.9	25.3
1970~1979년생	5,558	28.7	30.3	27.3
1980~1989년생	3,266	28.7	29.8	27.9
1990~1999년생	251	24.4	24.9	24.2

3세로 다소 좁혀지는 양상이 발견되었다.

초혼 이후 이혼한 경우는 1,262명으로, 첫번째 결혼(초혼) 생활을 지속한 기간은 평균 15.8년이였다. 결혼 지속 기간별로 5~9년 사이에 이혼한 경우가 19.3%로 가장 많았고, 10~14년 17.7%, 15~19년 15.7%, 20~24년 14.4% 순이었으며, 결혼 후 4년 이내에 이혼한 경우가 13.7%로 다음을 이었다. 이혼의 발생은 매우 간헐적이며, 초혼 이후 상당 기간이 지난 후에 발견되는 사건이므로, 이를 코호트별로 분석하는 데에는 한계가 있었으나, 향후 KLIPS 자료가 보다 장기에 걸쳐 축적된다면, 결혼 지속 혹은 이혼과 관련된 세대 간 변화에 대한 연구도 가능할 것으로 생각된다.

이때, KLIPS 혼인력 데이터의 한계이기도 한 지점은 어떤 응답자가 원가구원(추적대상 표본)이 아닌 경우 이혼 등으로 인해 분가한 경우 더 이상 KLIPS의 조사 대상으로 포함되지 않고 자료의 누락이 발생한다는 점이다.²¹⁾ 또한, 원가구원이라 할지라도 이혼 등의 사건이 발생한 이후 더 이상 KLIPS 조사에 응하지 않은 경우에도 해당 가구원의 혼인상태 변화를 관측할 수 없기 때문에, KLIPS의 혼인력 데이터에서 별거/이혼/사별 등의 상태는 실제 발생 건수에 비해 적게 조사되고 있을 가능성이 높다는 점에 유의할 필요가 있다.

21) 이 경우 혼인상태 변화 여부 및 시기가 관측되지 않음을 의미한다.

〈표 2-20〉 초혼의 평균 결혼 지속 기간(이혼한 경우만)

(단위: 명, %)

지속기간	N	비중	지속기간	N	비중
0~4년	173	13.7	30~34년	63	5.0
5~9년	244	19.3	35~39년	40	3.2
10~14년	112	17.7	40~44년	16	1.3
15~19년	199	15.7	45~49년	15	1.2
20~24년	182	14.4	50~55년	9	0.7
25~29년	98	7.8	전체	1,262	100.0

제4절 결 론

지금까지 KLIPS의 Sub Data Set에 해당하는 가족력 데이터(각 차수별 Family Data 및 EHC) 생성 과정과 특징을 살펴보고, 해당 자료를 활용한 간단한 분석 예를 살펴보았다. KLIPS 가족력 데이터는 기존의 KLIPS 데이터에 포함되어 있었으나 활용이 어려웠던 가족관계 및 가족사적인 변화를 보다 직관적으로 이해할 수 있도록 데이터 구조를 재정리함으로써, 장기간에 걸쳐 축적된 기존의 KLIPS 데이터가 가진 한계, 즉 복잡한 구조와 장기 시계열로 인해 데이터 작성 과정에 상당한 시간과 노력을 요하는 문제를 획기적으로 경감시킬 수 있을 것으로 기대된다.

이를 통해 개인의 혼인상태 변화, 가구원의 분가, 사망, 가족 관계의 식별 및 변화를 보다 쉽게 분석할 수 있으며, 무엇보다 장기에 걸친 동태적인 분석 시 자료 작성 시간을 크게 단축시키고, 데이터 정확도 및 신뢰도를 높일 수 있을 것으로 사료된다. 또한, 데이터에 키(key)변수를 다수 포함하여, 기존의 가구용 데이터, 개인용 데이터, 직업력 데이터 등과 쉽게 연계하여 활용이 가능하다.

한편, 데이터 작성의 궁극적 목적은 아니었으나, 데이터 작성 과정에서 일부 오류들이 발견되었으며, 이는 이후 노동패널팀의 자료 정제를 위한 노력을 통해 자료 신뢰성을 제고할 수 있는 단초를 제공하고 있다. 이러

한 오류들은 Family Data보다는 EHC 파일 작성 과정에서 발견되었는데, 이는 Family Data 작성을 위해 주로 활용하는 가구용 데이터의 경우 조사 단계에서 인포시트(info-sheet) 등을 통해 오류가 발생하지 않도록 관리되는 데 반해, EHC 파일의 경우 오류 관명 기준이 애매할 뿐 아니라 조사 대상 제외 혹은 조사 거부 등으로 응답 자체를 받기 어려운 한계 등으로 인해 면밀한 클리닝이 어려웠던 것에 기인한 결과로 보여진다. 이러한 오류들은 주로 KLIPS의 데이터 체계 및 클리닝 원칙이 작성되기 이전인 초기 조사에서 발견되는 오류가 다수이므로, 향후에도 이를 완벽하게 복원하거나 클리닝을 완료하기는 어려울 것으로 사료된다. 따라서, 이러한 오류를 어떻게 처리할 것인가는 오롯이 KLIPS 연구진의 판단에 맡겨두고자 한다.

제 3 장

노동패널 고용지표의 신뢰성 검토 : 경제활동인구조사와 비교

제1절 서론

본 연구는 한국노동패널(KLIPS)에서 제공하는 기초적인 노동시장 지표, 즉 경제활동참가율, 고용률, 실업률, 종사상 지위에 따른 취업자 수 등을 성별, 연령별로 살펴보고 통계청의 경제활동인구조사(이하 경활조사라고 함)에서 나타나는 지표와 비교하는 것을 목적으로 한다. 노동패널은 1998년부터 시작하여 현재 23차까지 이어오면서 초기 표본의 이탈이 발생하였고, 이를 보정하기 위하여 2009년(12차)과 2018년(21차)에 새로운 표본을 추가하였다. 이 과정에서 경활조사 표본과 괴리를 보일 가능성이 높아졌으며 이에 따라 노동패널의 대표성에 대한 우려가 발생할 소지가 있다. 본 연구에서는 노동패널이 경활조사와 어느 정도 일관성을 가지는지를 검토하고 괴리가 발생하면 어느 항목, 어느 인구집단에서 발생하는지, 그 원인이 무엇인지를 밝히고자 한다. 두 자료 모두 표본조사이기 때문에 두 자료 사이에 차이가 생긴다면 결국 가중치와 관계가 있다고 할 수 있다. 따라서 두 자료 사이의 가중치 차이도 함께 분석에 고려할 것이다.

경활조사는 매월 조사·발표되며 우리나라 노동시장에 대한 정보 제공을 목적으로 한다. 노동시장의 대표적인 지표로는 경제활동참가율, 고용률, 실업률 등이 있는데 경활조사를 통하여 이들 노동공급관련 지표를 얻

을 수 있다. 한편, 노동패널을 통해서도 이들 지표를 구할 수 있는데 노동패널은 1년 중 특정 주를 대상으로 경제활동을 조사하는 것이 아니라 조사가 가능한 시점에서 조사되기 때문에 월간 지표는 비교적 정확하지 않다. 하지만 노동패널에서 계산된 경제활동참가율과 고용률의 연간 지표는 경찰조사와 큰 차이를 보이고 있지 않다. 다만, 노동패널에서 계산되는 실업률은 1년 동안 사람에 따라 기준 시기(reference week)가 다르고, 실업과 비경제활동 사이의 구분이 모호하기 때문 경찰조사의 실업률과는 상당한 차이를 보이고 있다(김우영, 2017).

노동패널과 경찰조사의 비교를 통하여 노동패널의 신뢰성을 높이려는 연구는 홍민기 외(2015)에서도 있었다. 특히 이상호(2015)는 노동패널에서 산출되는 고용지표와 경찰조사에서 제공하는 고용지표를 비교하였는데 노동패널의 고용률이 경찰조사보다 전반적으로 낮게 나타나며, 특히 청년층의 경우에 괴리가 크다는 것을 발견하였다. 이상호(2015)는 2009년(12차) 통합표본을 이용하여 2014년(17차)까지 분석하였기 때문에 1998년 표본이나 2014년 이후에는 경찰조사와 어떤 차이를 보이는지 알 수 없다. 본 연구에서는 1998년 표본뿐 아니라 2020년까지 분석기간을 확대하여 노동패널의 횡단면적 대표성이 확보되고 있는지 확인하고자 한다.

통계청의 경찰조사는 약 35,000가구를 표본으로 하는 가구조사인데 15세 이상의 모든 가구원을 대상으로 개인의 특성과 노동시장 활동에 대한 정보를 묻고 있다. 이 조사는 매월 이루어지며 응답자의 응답부담 완화와 표본의 노후화 현상 및 신규 계열 괴리 현상 감소를 위해 약 970가구를 매월 교체하여 주기적으로 표본이 재설정되고 있다.²²⁾ 한편, KLIPS는 인구주택 총조사를 모집단으로 하여 1998년 1차 조사(5,000개 도시 가구)를 시작으로 지금 현재까지 23차에 이르고 있다.²³⁾ 또한, 표본이 점점 고령화됨에 따라 2009년에 전국을 대상으로 1,415가구를 새롭게 표본에 추가하였고, 2018년(21차)에 또 다시 5,044가구를 추가하여 총 12,134가구에 이르고 있다. 2018년 현재 KLIPS는 1998년 원표본, 2009년 통합표본,

22) 경제활동인구조사 통계 설명: <https://meta.narastat.kr/metascv/index.do?confmNo=101004>.

23) 1998년 표본은 시 지역에 한정하며 제주지역을 제외하고 있음.

2018년 통합표본이 합쳐져서 아주 복잡한 구조를 갖고 있다. 이와 함께 사용하는 자료에 따라 그 표본에 맞는 가중치(weight)를 사용해야 하는 번거로움도 존재한다.

한편, KLIPS는 일정 기간 동안 개인과 가구의 동태적 변화를 살펴보기 위하여 구축된 패널자료(panel data)이기 때문에 한 시점의 단면(cross-section)을 보는 경제활동인구조사와는 목적이 다르다고 할 수 있다. 하지만 많은 연구자들이 KLIPS를 이용하여 횡단면 분석을 할 뿐 아니라 반복적 횡단분석(repeated cross-sectional analysis)을 하고 있기 때문에 횡단면 자료로서의 대표성을 확보하는 것도 매우 중요하다.

본 연구에서는 1998년(1차)~2020(23차)년까지 KLIPS에서 측정될 수 있는 각종 노동시장 지표를 통계청의 경황조사에서 제공되는 지표들과 비교하고자 한다. 구체적으로 본 연구에서는 15세 이상 인구, 경제활동참가인구(율), 고용률, 실업률, 취업자 수 등을 성별, 연령별, 종사상 지위별로 구분하여 비교할 것이며 이들 간 차이 정도를 확인하고, 그 원인과 개선방안에 대해서 제언하고자 한다.

제2절 1998년 표본의 경제활동참가율, 고용률, 실업률 비교

1. 전체 인구 대상

경제활동참가율, 고용률, 실업률을 알기 위해서는 15세 이상 인구, 취업자 수, 실업자 수를 알아야 한다. 경제활동참가율은 15세 이상 인구 중 경제활동에 참가하고 있는 사람(취업자+실업자)들의 비중, 고용률은 15세 이상 인구 중 취업자의 비중, 실업률은 경제활동인구 중 실업자의 비중으로 각각 계산된다. 한 가지 주의하여야 할 것은 실업자의 정의인데 통계청의 경황조사에서는 2000년 이전에는 ILO 기준(지난 1주일간 구직활동 여부)에 따라 실업자를 판단하고, 2000년부터는 OECD 기준(지난 4주간

의 구직활동 여부)에 따라 실업자 여부를 결정하고 있다는 것이다. 따라서 노동패널로부터 실업자를 정의하고 실업률과 경제활동참가율을 계산할 때도 1998년과 1999년은 ILO 기준을 따르고, 2000년부터는 OECD 기준을 따르기로 한다.

<표 3-1>은 KLIPS 1998년 표본을 대상으로 1998~2020년까지 횡단면 가중치를 적용하여 15세 이상 인구, 취업자, 실업자 및 경제활동참가율, 고용률, 실업률을 계산하고 통계청 경황조사에서 시(동) 지역에 거주하는 가구를 대상으로 해당 지표들을 계산한 뒤 비교한 것이다. 참고로 경황조사에서 비농가 가구와 제주도를 제외한 전체 15세 이상 전체 인구를 대상으로 비교한 결과는 <부표 3-1>과 <부표 3-2>에 제시된다.

우선 15세 이상 인구의 절대적인 수를 보면 1998년 초기에는 KLIPS와 경황조사가 상당히 유사하다가 그 이후 다소 격차가 벌어져 2020년에는 KLIPS의 15세 이상 인구가 경황조사보다 약 9% 적게 나타나고 있다.²⁴⁾ 취업자 수는 두 자료 사이에 괴리가 좀 더 심하게 나타나는데 2020년 현재 KLIPS의 취업자 수는 경황조사 취업자 수의 약 85%에 그치고 있다. 한편, 실업자 수는 KLIPS와 경황조사 사이에 차이가 가장 심하게 나타나는데 KLIPS의 실업자 수는 초기에는 경황조사 실업자 수보다 많은 경우도 있었으나 최근에 들어서서는 50%대로 낮아지는 현상도 보이고 있다. 참고로 <표 3-1>과 <부표 3-1>, <부표 3-2>와 비교하면 KLIPS와 경황조사 사이에 15세 이상 인구, 취업자 수, 실업자 수의 차이가 시(동) 지역 거주자를 대상으로 한 <표 3-1>에서 가장 적다는 것을 알 수 있다. 따라서 KLIPS 98년 표본은 우리나라 시 지역 거주자의 노동시장 상황을 보여주는 데 적합한 자료라고 할 수 있다.

한편, KLIPS 98년 표본에서 계산된 경제활동참가율과 고용률 역시 통계청의 경제활동인구조사에서 제공하는 수치보다는 전반적으로 낮게 나

24) KLIPS에서 15세 이상 인구, 취업자 수, 실업자 수를 계산할 때 빈도를 사용하는 대신 해당 노동시장 상태의 가중치를 더해서 구해짐. 빈도를 사용할 경우 가중치가 정수이어야 하는 조건이 필요하기 때문에 반올림으로 생기는 오차가 발생할 수 있다. 여기서는 이러한 오차를 줄이기 위하여 가중치를 더하는 방식을 사용하였으며 따라서 15세 이상 인구, 취업자 수, 실업자 수에 대한 좀 더 정확한 수치라고 말할 수 있다.

〈표 3-1〉 KLIPS 98년 표본과 경제활동인구조사(시 지역)의 고용지표 비교

(단위: 천 명, %)

	KLIPS						경제활동인구조사						KLIPS/경제조사					
	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률
1998	30,040	14,497	2,291	55.9	48.3	13.6	28,168	15,487	1,306	59.6	55.0	7.8	1.07	0.94	1.75	0.94	0.88	1.75
1999	30,254	15,727	1,123	55.7	52.0	6.7	29,879	16,615	1,245	59.8	55.6	7.0	1.01	0.95	0.90	0.93	0.94	0.96
2000	30,508	15,778	851	54.5	51.7	5.1	34,345	20,005	950	61.0	58.2	4.5	0.89	0.79	0.90	0.89	0.89	1.13
2001	30,743	16,119	777	55.0	52.4	4.6	29,190	16,906	786	60.6	57.9	4.4	1.05	0.95	0.99	0.91	0.90	1.04
2002	30,921	16,674	656	56.0	53.9	3.8	29,576	17,465	661	61.3	59.1	3.6	1.05	0.95	0.99	0.91	0.91	1.04
2003	31,082	16,596	803	56.0	53.4	4.6	30,023	17,504	721	60.7	58.3	4.0	1.04	0.95	1.11	0.92	0.92	1.16
2004	31,205	16,641	872	56.1	53.3	5.0	30,423	17,951	763	61.5	59.0	4.1	1.03	0.93	1.14	0.91	0.90	1.23
2005	31,272	16,475	950	55.7	52.7	5.4	30,839	18,177	794	61.5	58.9	4.2	1.01	0.91	1.20	0.91	0.89	1.29
2006	31,436	16,596	893	55.6	52.8	5.1	31,399	18,558	743	61.5	59.1	3.8	1.00	0.89	1.20	0.90	0.89	1.33
2007	31,596	16,677	748	55.1	52.8	4.3	31,958	18,944	707	61.5	59.3	3.6	0.99	0.88	1.06	0.90	0.89	1.19
2008	31,836	16,891	547	54.8	53.1	3.1	32,521	19,177	695	61.1	59.0	3.5	0.98	0.88	0.79	0.90	0.90	0.89
2009	32,001	16,984	758	55.4	53.1	4.3	32,986	19,080	793	60.2	57.8	4.0	0.97	0.89	0.96	0.92	0.92	1.08
2010	32,160	17,305	501	55.4	53.8	2.8	33,398	19,391	815	60.5	58.1	4.0	0.96	0.89	0.61	0.92	0.93	0.69
2011	32,409	17,755	484	56.3	54.8	2.7	33,894	19,849	765	60.8	58.6	3.7	0.96	0.89	0.63	0.93	0.94	0.73
2012	32,580	17,770	558	56.3	54.5	3.0	34,362	20,237	736	61.0	58.9	3.5	0.95	0.88	0.76	0.92	0.93	0.85
2013	32,728	17,937	510	56.4	54.8	2.8	34,848	20,575	710	61.1	59.0	3.3	0.94	0.87	0.72	0.92	0.93	0.84
2014	32,934	18,455	531	57.6	56.0	2.8	35,312	21,127	828	62.2	59.8	3.8	0.93	0.87	0.64	0.93	0.94	0.74
2015	33,109	18,037	455	55.9	54.5	2.5	35,667	21,396	851	62.4	60.0	3.8	0.93	0.84	0.53	0.90	0.91	0.65
2016	33,240	18,442	470	56.9	55.5	2.5	35,952	21,562	883	62.4	60.0	3.9	0.92	0.86	0.53	0.91	0.93	0.64
2017	33,334	18,587	527	57.3	55.8	2.8	36,185	21,837	895	62.8	60.3	3.9	0.92	0.85	0.59	0.91	0.92	0.71
2018	33,493	18,927	536	58.1	56.5	2.8	36,341	21,836	940	62.7	60.1	4.1	0.92	0.87	0.57	0.93	0.94	0.68
2019	33,559	18,498	797	57.5	55.1	4.1	36,559	22,039	915	62.8	60.3	4.0	0.92	0.84	0.87	0.92	0.91	1.03
2020	33,626	18,695	551	57.2	55.6	2.9	36,850	21,875	955	62.0	59.4	4.2	0.91	0.85	0.58	0.92	0.94	0.69

주: 경활자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임. 98-99년 실업률은 ILO 기준, 2000-2020년은 OECD 기준. 모든 수치는 가중치 부여.

타나고 있다.²⁵⁾ 하지만 고용률의 경우 취업자 수에서 보였던 격차보다는 훨씬 작게 나타나고 있으며 꾸준히 10% 이내의 차이를 보이고 있다. 다만, KLIPS의 실업률은 몇 개 연도를 제외하면 다른 지표에 비하여 경찰조사와 큰 차이를 보이고 있으며 이러한 경향을 최근에 더 심하게 나타나고 있다.

결국, KLIPS 98년 표본은 경찰조사와 비교할 때 경제활동참가율, 고용률을 과소추정하고 있는데 이러한 결과는 KLIPS가 비경제활동인구를 과다하게 포함하고 있기 때문으로 볼 수 있다. 즉, 시간이 지남에 따라 표본 이탈과 함께 표본에 포함된 사람들이 고령화되면서 비경제활동인구 비중이 높아지고 있고 이를 가중치로 충분히 보정하지 못해서 나타나는 현상으로 보여진다.

KLIPS 98년 표본과 통계청의 경찰조사와의 차이가 표본 자체의 문제인지 아니면 가중치의 문제인지를 확인하기 위하여 가중치를 사용하지 않은 KLIPS 98년 표본으로 경제활동참가율, 고용률, 실업률을 재계산하였고 이를 경제활동인구조사(시 지역 거주)와 비교하여 보았다. 그 결과는 <표 3-2>와 <부표 3-3>에 제시된다. 또한, 가중치를 적용한 경우와 그렇지 않은 경우 어느 쪽이 더 경찰조사와 근접한지를 알아보기 위하여 조정된 평균제곱근오차(Root Mean Squared Error)를 식 (1)과 같이 계산한다.

$$\text{조정된 } RMSE = \frac{\sqrt{\sum_{t=1998}^{2020} (x_t - \mu_t)^2 / n}}{\mu} \quad (1)$$

식 (1)에서 x_t 는 t 년도의 KLIPS 지표, μ_t 는 t 년도의 경찰조사 지표, μ 는 해당 지표의 경찰조사 평균, n 은 기간 수, 즉 23을 나타낸다. 즉, 식 (1)은 해당 지표에 대한 KLIPS와 경찰조사의 차이의 제곱을 해당 지표의 평균으로 나눈 것이며 이는 마치 변동계수(coefficient of variation)와 유사하다.

25) 이러한 현상은 2009년 표본을 사용한 이상호(2015), 김우영(2017)에서도 발견되었다.

〈표 3-2〉 KLIPS 98년 표본의 가중치 적용 여부에 따른 경활조사(시 지역)와의 고용지표 비교

(단위 : %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경활(가중치 적용)			(A-C)			(B-C)		
	경활참가율	고용률	실업률	경활참가율	고용률	실업률	경활참가율	고용률	실업률	경활참가율	고용률	실업률	경활참가율	고용률	실업률
1998	55.9	48.3	13.6	55.8	48.2	13.6	59.6	55.0	7.8	-3.7	-6.7	5.8	-3.8	-6.8	5.8
1999	55.7	52.0	6.7	55.6	51.9	6.7	59.8	55.6	7.0	-4.1	-3.6	-0.3	-4.2	-3.7	-0.3
2000	54.5	51.7	5.1	54.6	51.9	5.0	61.0	58.2	4.5	-6.5	-6.5	0.6	-6.4	-6.3	0.5
2001	55.0	52.4	4.6	55.2	52.7	4.4	60.6	57.9	4.4	-5.6	-5.5	0.2	-5.4	-5.2	0.0
2002	56.0	53.9	3.8	56.4	54.3	3.7	61.3	59.1	3.6	-5.3	-5.2	0.2	-4.9	-4.8	0.1
2003	56.0	53.4	4.6	56.5	53.8	4.7	60.7	58.3	4.0	-4.7	-4.9	0.6	-4.2	-4.5	0.7
2004	56.1	53.3	5.0	56.7	53.9	4.9	61.5	59.0	4.1	-5.4	-5.7	0.9	-4.8	-5.1	0.8
2005	55.7	52.7	5.4	56.6	53.6	5.2	61.5	58.9	4.2	-5.8	-6.2	1.2	-4.9	-5.3	1.0
2006	55.6	52.8	5.1	56.8	54.0	4.9	61.5	59.1	3.8	-5.9	-6.3	1.3	-4.7	-5.1	1.1
2007	55.1	52.8	4.3	56.2	53.9	4.2	61.5	59.3	3.6	-6.4	-6.5	0.7	-5.3	-5.4	0.6
2008	54.8	53.1	3.1	56.1	54.3	3.2	61.1	59.0	3.5	-6.3	-5.9	-0.4	-5.0	-4.7	-0.3
2009	55.4	53.1	4.3	56.6	54.2	4.2	60.2	57.8	4.0	-4.8	-4.7	0.3	-3.6	-3.6	0.2
2010	55.4	53.8	2.8	57.2	55.6	2.7	60.5	58.1	4.0	-5.1	-4.3	-1.2	-3.3	-2.5	-1.3
2011	56.3	54.8	2.7	57.9	56.4	2.6	60.8	58.6	3.7	-4.5	-3.8	-1.0	-2.9	-2.2	-1.1
2012	56.3	54.5	3.0	58.8	57.1	2.9	61.0	58.9	3.5	-4.7	-4.4	-0.5	-2.2	-1.8	-0.6
2013	56.4	54.8	2.8	58.6	57.3	2.2	61.1	59.0	3.3	-4.7	-4.2	-0.5	-2.5	-1.7	-1.1
2014	57.6	56.0	2.8	59.9	58.5	2.4	62.2	59.8	3.8	-4.6	-3.8	-1.0	-2.3	-1.3	-1.4
2015	55.9	54.5	2.5	58.4	57.1	2.2	62.4	60.0	3.8	-6.5	-5.5	-1.3	-4.0	-2.9	-1.6
2016	56.9	55.5	2.5	58.9	57.7	2.1	62.4	60.0	3.9	-5.5	-4.5	-1.4	-3.5	-2.3	-1.8
2017	57.3	55.8	2.8	58.8	57.6	2.0	62.8	60.3	3.9	-5.5	-4.5	-1.1	-4.0	-2.7	-1.9
2018	58.1	56.5	2.8	59.2	57.9	2.2	62.7	60.1	4.1	-4.6	-3.6	-1.3	-3.5	-2.2	-1.9
2019	57.5	55.1	4.1	58.9	57.4	2.5	62.8	60.3	4.0	-5.3	-5.2	0.1	-3.9	-2.9	-1.5
2020	57.2	55.6	2.9	58.1	56.7	2.4	62.0	59.4	4.2	-4.8	-3.8	-1.3	-3.9	-2.7	-1.8
조정된 RMSE										0.09	0.09	0.36	0.07	0.07	0.39

<표 3-2>를 보면 가중치를 사용하였을 때보다 오히려 가중치를 사용하지 않았을 때 KLIPS 98년 표본의 경제활동참가율과 고용률이 경황조사의 해당 지표에 더 근접하게 나타남을 알 수 있다. 이러한 내용은 <부표 3-3>에서도 확인된다. 다만, 실업률의 경우는 가중치를 사용하였을 때가 경황조사와 더 근접한 것으로 나타나지만 경제활동참가율과 고용률에 비하여 조정된 평균제곱근오차가 워낙 크기 때문에 그리 신뢰할 만한 수준은 아니다.

KLIPS 98년 표본에서 얻어진 경제활동참가율과 고용률이 경황조사에서 구해진 값보다 낮게 나타나는 것은 KLIPS 98년 표본의 가중치를 재산정할 필요성이 있다는 것을 의미한다. 현재로서는 KLIPS 98년 표본에 가중치를 적용하지 않고 사용하는 것이 오히려 경제활동참가율이나 고용률을 더 정확하게 추정할 수 있으며, Probit이나 Logit을 이용하여 경제활동참여 결정이나 취업결정 모형을 추정할 때도 가중치를 사용하지 않는 것이 더 정확하게 계수를 추정할 수 있을 것으로 사료된다. 하지만 KLIPS 98년 표본에 가중치를 사용하지 않는 경우에도 경제활동참가율이나 고용률이 경황조사보다 일관되게 낮게 나타나기 때문에 궁극적으로는 KLIPS 1998년 표본에 대한 가중치를 재부여하는 작업이 필요할 것이다.

2. 남성과 여성

KLIPS 98년 표본을 남성과 여성으로 구분하여 경제활동참가율, 고용률, 실업률을 계산하고 이를 경황조사 자료의 해당 지표와 비교한 결과는 <표 3-3>과 <표 3-4>에 각각 제시한다. 앞서와 마찬가지로 KLIPS 98년 표본에 가중치를 적용한 경우와 그렇지 않은 경우를 구분하며 경황조사는 비교를 위하여 시(동) 지역 거주자만을 대상으로 경제활동참가율, 고용률, 실업률을 계산하였다. 우선 남성을 대상으로 한 <표 3-3>을 보면 가중치의 적용 여부와 상관없이 KLIPS의 경제활동참가율과 고용률은 전반적으로 경황조사의 해당 지표보다 낮게 나타남을 알 수 있다. 또한, 실업률의 경우에도 최근에 들어서서는 경황조사보다 낮게 나타나는 경향을 보인다. 이는 결과적으로 KLIPS에서 비경제활동인구가 표본에 과도하게

〈표 3-3〉 KLIPS 98년 남성 표본의 가중치 적용 여부에 따른 경찰조사(시 지역)와의 고용지표 비교

(단위: %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률
1998	69.8	60.9	12.7	69.7	60.9	12.6	74.6	68.2	8.5	-4.8	-7.4	4.2	-4.9	-7.3	4.1
1999	69.2	64.4	7.0	69.1	64.2	7.1	74.0	68.3	7.8	-4.8	-3.9	-0.8	-4.9	-4.1	-0.7
2000	68.1	64.8	4.9	68.3	65.0	4.9	74.4	70.6	5.1	-6.3	-5.8	-0.2	-6.0	-5.6	-0.2
2001	67.8	64.8	4.5	68.4	65.4	4.3	74.2	70.5	4.9	-6.4	-5.7	-0.5	-5.8	-5.1	-0.6
2002	68.4	66.2	3.2	69.4	67.2	3.2	74.9	71.9	4.0	-6.4	-5.6	-0.8	-5.5	-4.7	-0.9
2003	68.9	65.7	4.6	70.1	66.8	4.7	74.4	71.4	4.1	-5.6	-5.7	0.5	-4.3	-4.5	0.6
2004	68.7	65.4	4.8	70.3	67.1	4.7	74.7	71.6	4.3	-6.0	-6.1	0.5	-4.4	-4.5	0.4
2005	68.8	65.0	5.5	70.8	67.2	5.1	74.4	71.1	4.4	-5.6	-6.1	1.0	-3.5	-3.9	0.7
2006	68.1	64.9	4.7	70.6	67.5	4.5	74.0	70.9	4.2	-5.9	-6.0	0.5	-3.4	-3.4	0.3
2007	67.2	64.3	4.4	70.1	67.2	4.2	73.9	70.9	4.0	-6.7	-6.7	0.4	-3.8	-3.8	0.1
2008	66.6	64.5	3.0	70.5	68.3	3.2	73.4	70.5	3.9	-6.8	-5.9	-0.9	-2.9	-2.2	-0.8
2009	66.8	64.0	4.2	70.4	67.6	4.0	72.6	69.4	4.4	-5.8	-5.4	-0.2	-2.2	-1.8	-0.4
2010	66.1	64.2	2.8	70.8	68.9	2.7	72.7	69.6	4.3	-6.6	-5.3	-1.5	-1.9	-0.6	-1.7
2011	67.5	65.6	2.8	72.0	70.0	2.7	73.0	70.1	3.9	-5.5	-4.5	-1.1	-1.1	-0.1	-1.2
2012	66.5	64.4	3.2	72.4	70.3	2.9	73.1	70.4	3.7	-6.6	-6.0	-0.5	-0.7	-0.2	-0.7
2013	66.5	64.3	3.3	72.4	70.6	2.5	73.0	70.4	3.5	-6.5	-6.1	-0.2	-0.6	0.2	-1.1
2014	68.8	66.9	2.8	74.9	73.3	2.2	73.9	71.1	3.8	-5.1	-4.2	-0.9	1.0	2.2	-1.6
2015	65.3	64.1	1.9	72.2	70.9	1.8	73.7	70.9	3.9	-8.4	-6.8	-1.9	-1.6	0.0	-2.0
2016	66.3	64.6	2.5	72.5	71.1	2.0	73.7	70.7	4.0	-7.4	-6.1	-1.5	-1.1	0.4	-2.0
2017	66.1	63.9	3.3	71.6	70.1	2.1	73.8	70.8	4.1	-7.7	-6.9	-0.8	-2.2	-0.7	-2.0
2018	65.5	63.6	2.9	71.6	70.1	2.2	73.3	70.1	4.2	-7.7	-6.5	-1.4	-1.6	-0.1	-2.1
2019	65.8	62.6	4.9	71.5	69.7	2.5	72.9	69.9	4.1	-7.1	-7.3	0.8	-1.5	-0.3	-1.6
2020	65.2	63.1	3.2	70.4	68.7	2.5	72.0	69.1	4.1	-6.8	-6	-0.9	-1.6	-0.4	-1.6
조정된 RMSE										0.09	0.08	0.28	0.05	0.05	0.33

〈표 3-4〉 KLIPS 98년 여성 표본의 가중치 적용 여부에 따른 경찰조사(시 지역)와의 고용지표 비교

(단위: %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치비 적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률	경찰참가율	고용률	실업률
1998	42.8	36.3	15.0	42.7	36.3	15.0	45.5	42.5	6.6	-2.7	-6.1	8.5	-2.7	-6.2	8.5
1999	43.1	40.5	6.2	43.1	40.4	6.1	46.4	43.7	5.8	-3.2	-3.2	0.4	-3.3	-3.2	0.3
2000	41.8	39.5	5.4	41.9	39.7	5.2	48.5	46.6	3.8	-6.6	-7.1	1.7	-6.6	-6.9	1.5
2001	43.0	40.9	4.8	42.9	41.0	4.6	47.8	46.1	3.7	-4.9	-5.2	1.1	-4.9	-5.1	0.9
2002	44.4	42.3	4.6	44.3	42.3	4.5	48.5	47.0	3.1	-4.1	-4.7	1.5	-4.3	-4.7	1.3
2003	43.9	41.9	4.6	43.8	41.7	4.7	47.8	46.0	3.7	-3.9	-4.1	0.9	-4.0	-4.3	1.0
2004	44.2	41.8	5.3	43.9	41.6	5.2	49.0	47.1	3.8	-4.8	-5.3	1.5	-5.1	-5.6	1.4
2005	43.5	41.1	5.4	43.4	41.0	5.4	49.3	47.4	3.9	-5.9	-6.3	1.6	-6.0	-6.4	1.5
2006	43.8	41.3	5.6	43.9	41.5	5.4	49.6	47.9	3.3	-5.8	-6.6	2.3	-5.7	-6.5	2.1
2007	43.5	41.7	4.1	43.2	41.4	4.1	49.7	48.2	3.0	-6.2	-6.5	1.2	-6.5	-6.8	1.2
2008	43.5	42.0	3.3	42.6	41.3	3.2	49.4	48.0	2.9	-6.0	-6.0	0.4	-6.8	-6.7	0.3
2009	44.5	42.6	4.3	43.6	41.6	4.5	48.5	46.9	3.3	-3.9	-4.2	1.0	-4.9	-5.2	1.1
2010	45.0	43.8	2.8	44.5	43.2	2.9	48.9	47.1	3.6	-3.8	-3.3	-0.8	-4.4	-3.9	-0.8
2011	45.6	44.4	2.4	44.9	43.8	2.5	49.2	47.5	3.4	-3.6	-3.1	-0.9	-4.3	-3.7	-0.9
2012	46.4	45.1	2.8	46.0	44.7	2.9	49.5	47.9	3.3	-3.1	-2.8	-0.5	-3.5	-3.2	-0.4
2013	46.6	45.6	2.0	45.7	44.9	1.9	49.7	48.2	3.1	-3.1	-2.6	-1.0	-4.0	-3.3	-1.2
2014	47.2	45.9	2.7	46.3	45.0	2.7	51.0	49.1	3.7	-3.8	-3.2	-1.0	-4.7	-4.1	-1.1
2015	46.7	45.2	3.2	45.6	44.3	2.7	51.5	49.6	3.8	-4.8	-4.4	-0.6	-5.9	-5.2	-1.1
2016	47.7	46.6	2.4	46.2	45.2	2.2	51.7	49.7	3.9	-4.0	-3.1	-1.4	-5.5	-4.5	-1.7
2017	48.8	47.8	2.0	46.7	45.9	1.7	52.3	50.4	3.7	-3.6	-2.6	-1.7	-5.6	-4.4	-2.0
2018	50.9	49.5	2.6	47.6	46.5	2.3	52.6	50.5	4.0	-1.7	-0.9	-1.4	-4.9	-3.9	-1.7
2019	49.5	47.9	3.2	47.3	46.1	2.5	53.1	51.1	3.8	-3.6	-3.1	-0.7	-5.8	-4.9	-1.3
2020	49.5	48.3	2.4	46.7	45.6	2.3	52.3	50.1	4.3	-2.8	-1.8	-1.9	-5.6	-4.5	-2.0
조정된 RMSE										0.09	0.09	0.57	0.10	0.11	0.58

포함되고 있음을 의미한다. 한편, 조정된 RMSE를 보면 KLIPS에 가중치를 적용하지 않을 경우가 적용할 때보다 오히려 경제활동참가율과 고용률에 있어서 경황조사와의 격차가 더 적어지는 것을 알 수 있다. 따라서 남성을 대상으로 하는 분석에 있어서는 가중치를 적용하지 않는 것이 더 대표성이 높다고 할 수 있을 것이다.

한편, 여성을 대상으로 한 <표 3-4>에서도 남성과 마찬가지로 KLIPS의 경제활동참가율과 고용률이 경황조사의 해당 지표보다 낮게 나타남을 알 수 있다. 실업률 역시 2010년 이후부터는 경황조사의 실업률보다 지속적으로 낮게 나타나고 있다. 하지만 가중치 적용 여부에 따른 결과는 남성과는 다른 양상을 보이는데 여성의 경우에는 가중치를 적용할 때가 그렇지 않을 때보다 경황조사에 좀 더 근접하는 것으로 나타난다. 따라서 여성을 대상으로 한 분석에서는 가중치를 사용하는 것이 바람직할 것이다. 앞서 전체를 대상으로 한 분석에서 가중치의 조정이 필요하다고 언급하였는데 이는 여성보다는 남성에 해당된다고 할 수 있을 것이다.

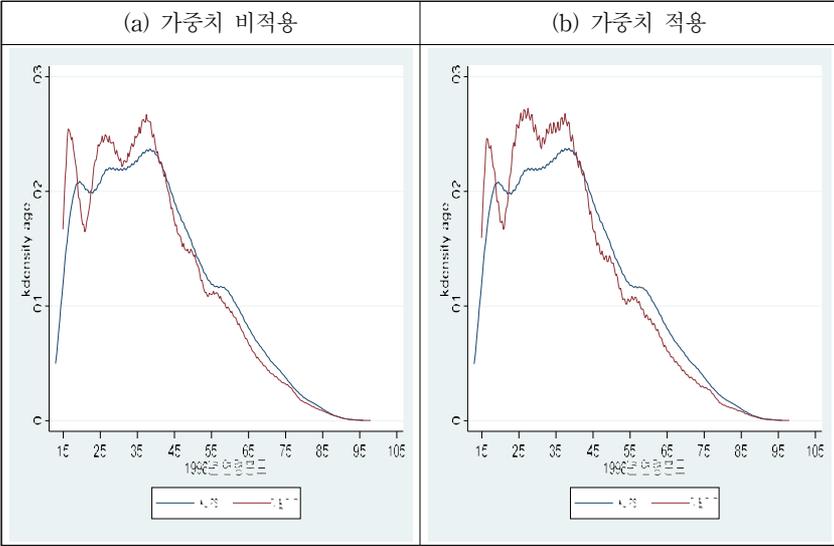
3. 연령별 비교

지금까지 KLIPS 98년 표본으로부터 계산되는 경제활동참가율과 고용률은 경황조사에서 구해진 해당 지표보다 낮게 나타나는데 이는 KLIPS 98년 표본의 노후화로 고령자의 비중이 많아지고, 따라서 비경황인구가 과다하게 포함되어 있기 때문으로 추측되었다. 여기서는 이러한 가능성을 확인하고자 한다. 하지만 표본의 노후화의 경우에도 가중치를 적절히 부여한다면 경황조사와 근접할 수 있기 때문에 가중치를 사용하는 경우와 그렇지 않은 경우 모두 고려하여야 할 것이다.

KLIPS 98년 표본과 경황조사(가중치 적용)의 연령 분포를 비교하기 위하여 1998년, 2009년, 2020년 3개년도의 연령 분포를 Kernel method를 이용하여 계산하였으며 KLIPS의 경우 가중치를 사용한 것과 가중치를 사용하지 않은 경우를 구분하여 연령 분포를 추정하였다.

우선 1998년 KLIPS를 보면 가중치를 사용한 경우와 그렇지 않은 경우 사이에 연령 분포에서 큰 차이를 보이고 있지 않다. 한편 경황조사는 가

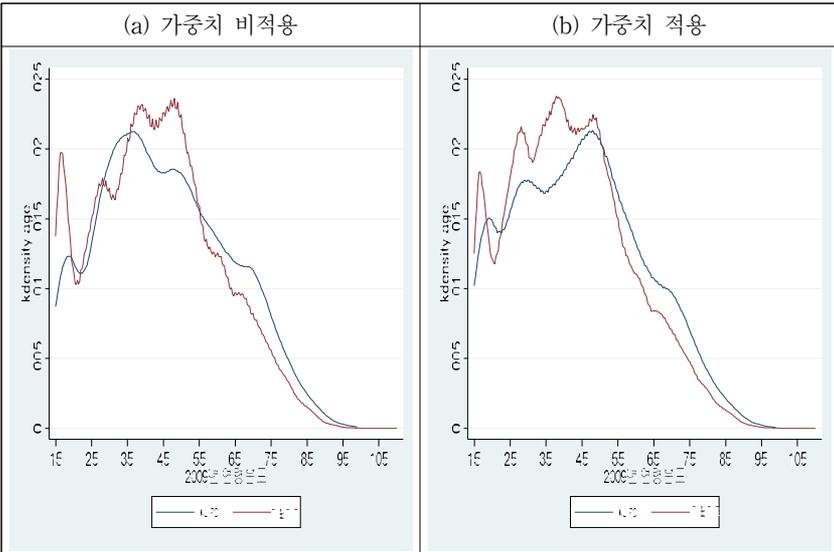
(그림 3-1) 1998년 KLIPS 98년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 39.7, 표준편차: 15.8
경찰조사: 평균: 38.3, 표준편차: 16.2

주: KLIPS: 평균: 39.7, 표준편차: 16.2
경찰조사: 평균: 37.7, 표준편차: 15.4

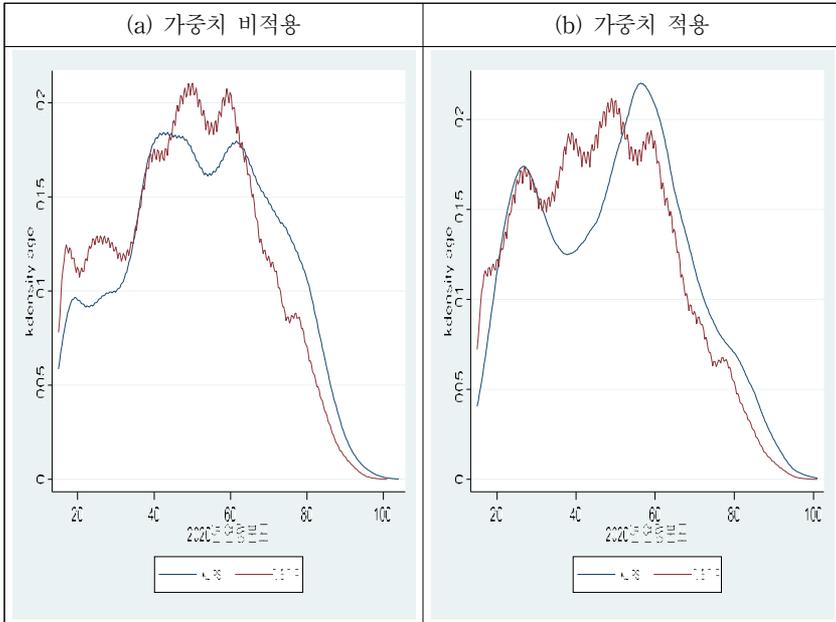
(그림 3-2) 2009년 KLIPS 98년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 45.5, 표준편차: 17.8
경찰조사: 평균: 43.4, 표준편차: 16.8

주: KLIPS: 평균: 44.4, 표준편차: 17.7
경찰조사: 평균: 42.1, 표준편차: 16.3

[그림 3-3] 2020년 KLIPS 98년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 51.9, 표준편차: 18.9
경찰조사: 평균: 48.8, 표준편차: 18.1

주: KLIPS: 평균: 49.4, 표준편차: 18.6
경찰조사: 평균: 46.4, 표준편차: 17.6

중치를 사용할 경우 평균 연령과 분산이 약간 줄어드는데 KLIPS의 평균 연령이 39.7세, 경찰조사(가중치 적용)의 평균 연령은 37.7세로 약 2살 정도 KLIPS의 평균 연령이 높은 것으로 나타나고 있다. 또한, [그림 3-1]을 보면 KLIPS의 45세 이상 비중이 경찰조사의 45세 이상 비중을 초과하고 있어 1998년 표본이 시작하는 해부터 KLIPS 표본이 중·고령층을 과다하게 포함하고 있다는 것을 알 수 있다. 이러한 현상은 가중치를 적용한 경우에도 유사하게 나타나고 있어 적절하게 가중치 설정되었는지 의문이다.

[그림 3-2]에 나타난 2009년 연령 분포를 보면 역시 KLIPS의 평균 연령이 경찰조사(가중치 적용)의 평균 연령보다 약 2살 정도 많은 것으로 나타나며 KLIPS의 55세 이상 비중이 경찰조사의 55세 이상 비중을 초과하는 것을 보이고 있다. 한편, [그림 3-3]은 2020년 KLIPS의 연령 분포와 경찰조사의 연령 분포 사이에 가장 큰 차이를 보여준다. KLIPS의 55세

이상 비중이 경찰조사의 55세 이상 비중을 초과할 뿐 아니라 35~55세의 비중이 KLIPS에서는 35세 미만에 비하여 더 작게 나타나는 반면 경찰조사에서는 그 반대를 보이고 있다. 즉, 2020년 KLIPS 표본은 경찰조사에 비하여 35~55세가 과소하게, 55세 이상은 과다하게 포함되고 있다. 따라서 이전 연도와 달리 두 자료 사이에 평균 연령도 3살의 차이를 나타내고 있다.

결국, [그림 3-1]~[그림 3-3]은 KLIPS 98년 표본이 젊은 층을 과소하게, 중·고령층을 과다하게 포함하고 있고, 특히 2019년에는 35~55세도 과소하게 반영되어 있다는 것을 보여주고 있다. 또한, KLIPS 98년 표본에 가중치를 적용하였을 경우에도 경찰조사와의 격차를 줄이고는 있지만 여전히 연령 분포를 경찰조사 수준으로 맞출 정도의 역할은 하고 있지 못하다.

[그림 3-1]~[그림 3-3]에서 보이는 KLIPS와 경찰조사의 차이가 통계적으로 유의한지를 검증하기 위해서는 좀 더 분석이 필요한데 여기서는 Kolmogorov-Smirnov test를 사용하여 통계적 검증을 하고자 한다 (Stata, 2021). K-S 테스트는 두 분포 사이의 누적분포의 차이를 계산하고 그중 가장 큰 수치를 채택하여 그 차이가 통계적으로 유의한지를 검증하는 것이다. 즉 우리의 연령 분포의 경우 K-S 통계량은 다음과 같이 계산된다.

$$\begin{aligned}
 K-S \text{ 통계량} &= \max [|D_{\max}|, |D_{\min}|], & (2) \\
 D_{\max} &= \max [F(x) - G(x)], \\
 D_{\min} &= \min [F(x) - G(x)]
 \end{aligned}$$

위에서 x 는 연령, F 와 G 는 각각 KLIPS와 경찰조사의 연령누적분포를 의미한다. K-S 통계량은 누적분포의 차이가 가장 큰 수치를 의미하여 Kolmogorov(1933)과 Smirnov(1933)은 이 통계량에 대한 p값을 제시하였다. <표 3-5>는 Kolmogorov - Smirnov test 결과를 보여준다.

<표 3-5>를 보면 1998년부터 KLIPS와 경찰조사의 연령 분포는 통계적으로 유의한 차이가 있었던 것으로 나타나고 있다. 그리고 그 차이는 해가 거듭되면서 더 벌어지고 있다는 것을 K-S 통계량으로부터 확인할

〈표 3-5〉 Kolmogorov - Smirnov 테스트 결과

	D_{\max}	D_{\min}	K-S 통계량	p-value
1998	0.042	-0.001	0.042	0.000
2009	0.064	-0.000	0.064	0.000
2020	0.081	0.001	0.081	0.000

수 있다. 결국, KLIPS 98년 표본이 중·고령층을 상대적으로 더 많이 포함하고 있어 경찰조사의 연령 분포와 다르게 나타난다고 볼 수 있을 것이다.

KLIPS 98년 표본이 중·고령층을 과다하게 포함하고 있다는 것은 비경찰인구를 상대적으로 더 많이 포함하고 있다는 것을 의미하며 이는 KLIPS의 경제활동참가율, 고용률이 경찰조사에 비하여 낮게 나타나는 이유가 될 수 있다. 이를 구체적으로 확인하기 위하여 KLIPS 98년 표본과 경찰조사의 연령별 경제활동참가율, 고용률, 실업률을 비교하기로 한다. 앞서 연령 분포의 그림에서 대략 35세, 55세를 분기점으로 KLIPS와 경찰조사가 달라지는 것으로 나타났다. 따라서 15~34세, 35~54세, 55세 이상으로 연령층을 구분하고 각각에 대하여 경제활동참가율, 고용률, 실업률을 계산하기로 한다. 또한, KLIPS의 연령 분포는 가중치를 적용하였을 때 경찰조사와 더 가깝게 나타났기 때문에 모든 지표는 가중치를 적용하여 계산한다.

〈표 3-6〉은 KLIPS 98년 표본과 경찰조사의 연령별 고용률과 둘 사이의 차이를 보여준다. 경제활동참가율과 실업률에 대한 내용은 각각 〈부표 3-4〉와 〈부표 3-5〉에 제시하는데 경제활동참가율의 차이는 고용률과 매우 유사하기 때문에 고용률만을 대상으로 설명하기로 한다.

연령을 35세 미만, 35~54세, 55세 이상으로 구분하여 고용률을 계산한 결과, KLIPS 98년 표본의 고용률이 경찰조사의 고용률보다 모든 연령계층에서 낮게 나타나는데 그중에서도 35세 미만 집단에서 가장 큰 차이를 보이고 있다. 따라서 KLIPS 98년 표본의 고용률이 경찰조사보다 낮게 나타나는 이유는 단지 표본에 중·고령층이 과다하게 포함되어 있기 때문만은 아니며 취업자 자체가 적게 포함되어 있기 때문이라고 할 수 있다.

〈표 3-6〉 KLIPS 98년 표본과 경찰조사와의 연령별 고용률 비교

(단위: %, %p)

	35세 미만			35~54세			55세 이상		
	KLIPS98	경찰	차이	KLIPS98	경찰	차이	KLIPS98	경찰	차이
1998	38.0	47.8	-9.8	66.6	71.6	-5.0	34.2	37.0	-2.8
1999	41.9	48.1	-6.2	71.6	72.2	-0.5	35.9	38.8	-2.9
2000	41.5	50.6	-9.0	71.5	74.7	-3.2	34.5	43.2	-8.7
2001	42.2	51.2	-9.0	72.3	74.0	-1.7	34.9	38.1	-3.2
2002	43.7	52.4	-8.7	74.3	74.9	-0.6	35.8	39.5	-3.7
2003	44.1	52.1	-8.0	72.7	74.2	-1.6	35.9	37.2	-1.3
2004	43.9	52.7	-8.8	72.5	74.6	-2.2	35.6	38.7	-3.1
2005	43.4	52.4	-9.1	72.0	74.6	-2.6	35.1	38.7	-3.5
2006	43.4	51.7	-8.3	71.9	75.3	-3.4	35.2	39.2	-4.0
2007	41.6	51.1	-9.5	73.5	75.5	-2.0	36.1	40.6	-4.6
2008	41.7	50.2	-8.5	74.4	75.7	-1.3	36.5	40.4	-3.9
2009	42.3	48.5	-6.2	73.3	74.6	-1.3	38.0	40.3	-2.3
2010	42.0	48.7	-6.7	75.7	75.0	0.7	38.1	40.9	-2.8
2011	41.9	49.2	-7.2	77.6	75.5	2.1	39.3	42.1	-2.9
2012	41.0	49.6	-8.6	76.3	75.6	0.6	41.5	43.4	-1.9
2013	40.4	49.4	-9.0	77.0	75.9	1.1	42.7	44.6	-2.0
2014	41.9	50.3	-8.3	77.3	76.6	0.7	43.7	46.1	-2.4
2015	40.6	51.0	-10.4	75.7	76.7	-1.1	44.4	46.3	-1.9
2016	41.6	50.9	-9.4	76.6	77.1	-0.5	46.1	46.6	-0.5
2017	42.3	51.1	-8.9	76.9	77.7	-0.8	47.7	47.3	0.4
2018	44.9	51.5	-6.6	77.2	77.4	-0.2	48.2	46.9	1.3
2019	44.6	52.4	-7.7	74.8	77.0	-2.2	47.4	47.7	-0.3
2020	45.8	51.5	-5.7	75.9	75.8	0.1	47.4	47.6	-0.2

주: 1) 경제활동인구자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임.

2) 모든 수치는 가중치를 사용하여 구한 것임.

특히, 35세 미만에서는 KLIPS 98년 표본의 고용률이 모든 연도에서 상당히 낮게 나타나는데 이 계층에 대한 취업자 표본을 확충하거나 가중치를 조정할 필요가 있어 보인다.

이상호(2015) 역시 경활조사와 비교하여 KLIPS 1998년과 2009년 표본에서 경제활동참가율과 고용률이 낮게 나타나고 있음을 발견하였다. 또한, KLIPS에서 특히 청년층 고용률이 과소하게 나타나는 이유를 분석한 결과, ① 시간이 지남에 따라 표본 손실(sample attrition)이 발생하고, ② 청년층의 응답률이 낮으며, ③ 취업자로 분류될 경우 설문부담이 크기 때문에 취업자로 응답하기를 꺼리고, ④ 재학생의 아르바이트 등 취업자 비중이 낮게 나타나고 있는 점들을 지적하였다. 따라서 여기서도 동일한 이유가 가능하다고 볼 수 있다. 결국, 이 문제를 해결하기 위해서는 가중치를 조정하는 것뿐 아니라 표본 손실과 응답률을 높일 수 있도록 설문참가자에 대한 인센티브를 제공하고, 취업자에 대한 설문부담이 크지 않도록 설문지를 개편하는 등의 노력이 필요하다고 본다.

KLIPS 98표본과 경활조사의 전체 고용률 차이의 원인을 좀 더 살펴보기 위해서 다음과 같은 식 (3)을 생각해 보자.

$$\mu = \phi_{1534}\mu_{1534} + \phi_{3554}\mu_{3554} + \phi_{55ov}\mu_{55ov} \quad (3)$$

위에서 μ 는 전체 고용률, ϕ_{1534} , ϕ_{3554} , ϕ_{55ov} 는 각각 15~34세 비중, 35~54세 비중, 55세 이상 비중을 말하며, μ_{1534} , μ_{3554} , μ_{55ov} 는 각 연령대의 고용률을 나타낸다. 2020년을 대상으로 전체 고용율을 식 (3)으로 분해한 결과는 <표 3-7>에 제시되었다.

<표 3-7>은 2020년 KLIPS 98년 표본의 고용률이 경활조사보다 3.8%포인트 낮은 이유를 잘 보여주고 있다. 우선, 비중으로 볼 때 KLIPS 98년 표본은 경활조사에 비하여 35~44세 비중, 15~34세 비중이 낮고, 대신 55세 이상 비중은 높은 것으로 나타나고 있다. KLIPS 98년 표본이 경활조사의 연령 비중과 같을 경우의 전체 고용률을 계산하면 57.7%로 경활조사와의 격차는 -1.7%(=57.7-59.4)포인트로 낮아져 연령 비중의 차이가 전체의 58%를 차지하는 것을 보여준다. 나머지 차이는 고용률 자체의 차이인데 가장 큰 차이는 15~34세로 약 -5.7%포인트의 차이로 나타나고 있다. 따라서 KLIPS 98년도의 고용률을 높이기 위해서는 55세 이상의 고령자 비중을 낮추는 대신 35~54세의 표본을 늘리거나 응답률을 높이고, 15~34세의 경우 취업자 여부를 좀 더 명확히 판단하여 그 비중을 높이는

〈표 3-7〉 2020년 KLIPS 98년 표본과 경찰조사와의 고용률 차이 분해

(단위: %, %p)

	15~34세		35~54세		55세 이상		전체
	비중	고용률	비중	고용률	비중	고용률	고용률
KLIPS 98표본	0.271	45.8	0.303	75.9	0.425	47.4	55.6
경찰자료	0.286	51.5	0.377	75.8	0.337	47.6	59.4
차이	-0.015	-5.7	-0.074	0.1	0.088	-0.3	-3.8

주: 1) 비중의 합은 1이고 고용률의 단위는 %

2) 경제활동인구자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임.

3) 모든 수치는 가중치를 사용하여 구한 것임.

것이 주요 과제가 될 것이다.

4. 종사상 지위

〈표 3-8〉은 취업자를 대상으로 종사상 지위의 비중을 제시한다. KLIPS 98년 표본에서는 상용직, 임시직, 일용직, 고용주/자영업자, 무급가족종사자로 구분하고 있다.²⁶⁾ 여기서는 임시직과 일용직을 묶어 임시일용직으로 정하고 분석하기로 한다.

KLIPS 98년 표본과 경찰조사의 차이는 확연하게 나타난다. 즉, KLIPS 98년 표본의 상용직 비중은 경찰조사보다 상당히 높게 나타나고, 반대로 임시일용직 비중은 낮게 나타나고 있다. 최근에 들어서 상용직 비중의 격차가 줄어들고는 있지만, 임시일용직 비중은 여전히 KLIPS 98년 표본이 경찰조사보다 약 5%포인트 정도 낮게 나타나고 있다. 이는 왜 KLIPS에서 고용률이 낮게 나타나는지를 설명하는 단초가 될 수 있을 것이다. 다만, 2020년의 경우 그 격차는 상당히 줄고 있다.

KLIPS 98년 표본에서 임시일용직 비중이 낮은 이유는 아르바이트 일자리 등의 일시적인 직업을 가진 사람들이 상대적으로 많이 제외되었기 때문인 것으로 추측된다. 이러한 가능성에 대해서는 이상호(2015)에서도

26) 사업장에 근로자가 있는지 여부로 고용주와 자영업자를 구분할 수는 있으나 그렇게 할 경우 노동연구원에서 제시하는 고용주/자영업자 수와 차이를 보이기 때문에 이 둘을 묶어 분석하기로 한다.

〈표 3-8〉 KLIPS 98년 표본과 경찰조사(시 지역)와의 종사상 지위 분포 비교

(단위: %)

	상용직			입시일용직			임금근로자			자영업주			무급가족종사자			비임금근로자		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
1998	50.4	36.0	14.5	12.1	32.2	-20.1	62.5	68.2	-5.7	29.6	25.1	4.5	7.9	6.8	1.1	37.5	31.8	5.7
1999	47.2	32.5	14.7	16.9	34.7	-17.8	64.1	67.2	-3.1	27.5	25.7	1.7	8.5	7.1	1.4	35.9	32.8	3.1
2000	49.9	30.7	19.3	14.8	33.6	-18.7	64.8	64.2	0.6	26.8	27.1	-0.3	8.4	8.6	-0.2	35.2	35.8	-0.6
2001	51.8	33.3	18.6	14.6	35.2	-20.6	66.4	68.5	-2.0	25.4	25.3	0.1	8.2	6.3	1.9	33.6	31.5	2.0
2002	52.4	32.9	19.5	14.8	36.0	-21.2	67.2	68.9	-1.7	25.4	25.2	0.2	7.4	5.9	1.4	32.8	31.1	1.7
2003	53.7	35.2	18.5	14.4	34.8	-20.4	68.1	70.0	-2.0	24.4	24.4	-0.1	7.6	5.5	2.0	31.9	30.0	2.0
2004	53.0	36.0	16.9	15.1	34.6	-19.5	68.1	70.6	-2.6	24.7	24.4	0.3	7.3	5.0	2.3	31.9	29.4	2.6
2005	51.5	37.1	14.5	14.6	34.2	-19.6	66.2	71.3	-5.1	27.2	24.0	3.3	6.6	4.7	1.9	33.8	28.7	5.1
2006	53.0	38.0	15.0	14.9	34.0	-19.1	67.9	72.0	-4.1	25.8	23.4	2.4	6.3	4.5	1.7	32.1	28.0	4.1
2007	53.3	39.4	13.8	14.9	33.5	-18.6	68.2	72.9	-4.7	25.9	22.9	3.0	5.9	4.2	1.7	31.8	27.1	4.7
2008	53.4	40.9	12.5	15.2	32.4	-17.2	68.6	73.3	-4.7	25.4	22.4	2.9	6.0	4.2	1.8	31.4	26.7	4.7
2009	52.4	42.7	9.8	16.7	31.6	-15.0	69.1	74.3	-5.2	24.9	21.6	3.3	6.0	4.1	1.9	30.9	25.7	5.2
2010	52.5	44.9	7.6	17.9	30.3	-12.4	70.4	75.2	-4.8	24.0	21.0	3.0	5.6	3.9	1.7	29.6	24.8	4.8
2011	52.8	46.5	6.2	18.9	29.0	-10.1	71.6	75.5	-3.9	23.2	20.7	2.5	5.2	3.8	1.4	28.4	24.5	3.9
2012	52.4	47.6	4.8	19.2	27.9	-8.7	71.6	75.6	-4.0	23.3	20.8	2.5	5.1	3.6	1.4	28.4	24.4	4.0
2013	53.0	49.4	3.6	19.4	26.8	-7.5	72.4	76.3	-3.9	22.9	20.3	2.6	4.8	3.4	1.3	27.6	23.7	3.9
2014	53.4	50.0	3.4	18.8	26.7	-8.0	72.1	76.7	-4.6	22.9	19.9	3.0	5.0	3.4	1.6	27.9	23.3	4.6
2015	54.3	50.9	3.4	19.4	26.4	-7.1	73.6	77.3	-3.7	22.0	19.5	2.5	4.4	3.2	1.2	26.4	22.7	3.7
2016	54.5	51.6	2.9	18.8	25.9	-7.2	73.3	77.6	-4.3	22.0	19.3	2.8	4.7	3.2	1.5	26.7	22.4	4.3
2017	54.4	52.5	1.8	19.7	25.1	-5.5	74.0	77.7	-3.6	21.2	19.2	2.0	4.7	3.1	1.6	26.0	22.3	3.6
2018	55.8	53.9	1.8	19.1	24.2	-5.2	74.8	78.2	-3.3	20.7	18.9	1.8	4.4	2.9	1.5	25.2	21.8	3.3
2019	55.3	55.1	0.1	20.0	23.6	-3.6	75.3	78.8	-3.5	20.4	18.5	1.8	4.4	2.7	1.7	24.7	21.2	3.5
2020	55.9	57.0	-1.1	20.6	22.2	-1.6	76.6	79.1	-2.5	19.4	18.3	1.1	4.1	2.5	1.6	23.5	20.9	2.6

주: 자영업주는 고용주와 자영업자를 합한 것임. 모든 수치는 가중치를 사용하여 구한 것임. 차이는 KLIPS-경찰조사.

제시되었다. 하지만 상용직과 임시일용직을 합친 임금근로자 비중은 KLIPS 98년 표본과 경활조사 사이에 차이가 5%포인트 이내로 나타나고 있기 때문에 KLIPS 98년 표본에서 상용직과 임시일용직을 구분하기보다는 임금근로자 전체로 분석할 경우 좀 더 대표성을 갖는다고 볼 수 있다.

한편, 임금근로자와는 달리 비임금근로자의 비중은 KLIPS 98년 표본이 경활조사보다 더 크게 나타나고 있다. KLIPS의 자영업주(고용주+자영업자) 비중은 경활조사보다 약 2~3%포인트 더 높게 나타나고, 무급가족종사자는 약 1~2%포인트 정도 높게 나타나고 있어 비임금근로자 전체로 보면 KLIPS에서의 비중이 경활조사보다 약 3~4%포인트 높게 나타나고 있다. 결국 KLIPS 98년 표본을 통한 임금근로자 비중은 과소, 비임금근로자 비중은 과다하게 추정된다고 할 수 있을 것이다. <부표 3-6>은 KLIPS 98년 표본에 가중치를 적용하지 않았을 경우를 제시하고 있는데 주요 결과는 크게 달라지지 않고 있다.

5. 소 결

지금까지 KLIPS 98년 표본을 이용하여 경제활동참가율, 고용률, 실업률을 구하였고 이를 경활조사와 비교한 결과는 다음과 같이 정리될 수 있다. 첫째, KLIPS 98년 표본에서 가중치를 사용하여 구해진 취업자 수는 지속적으로 경활조사보다 작게 나타나고 실업자 수 역시 최근 들어서는 경활조사보다 작게 나타나고 있기 때문에 KLIPS 98년 표본을 이용하여 취업자, 실업자, 경제활동인구의 규모를 파악하는 것은 이들 지표를 과소 평가하게 될 위험이 있다.

둘째, KLIPS 98년 표본과 경활조사 사이에 경제활동참가율과 고용률의 격차는 가중치를 사용하지 않았을 때보다 오히려 가중치를 사용하였을 때 더 커지는 것으로 나타났다. 따라서 KLIPS의 가중치를 조정하는 것이 필요하다.

셋째, KLIPS 98년 표본에서 구해진 경제활동참가율과 고용률이 경활조사보다 낮게 나타나는 현상은 남성과 여성 모두에서 동일하게 나타나고 있다. 따라서 KLIPS 98년 표본이 남녀 모두에 있어서 비경활인구를

과도하게 포함하고 있다고 할 수 있다.

셋째, KLIPS 98년 표본과 경활조사 사이에 가장 큰 차이는 연령 분포에서 나타나고 있다. KLIPS 98년 표본은 경활조사에 비하여 55세 이상 비중을 과도하게 포함하고 있으며, K-S 검증결과 이 두 연령 분포는 서로 다른 것으로 나타났다. 2020년 KLIPS 98년 표본과 경활조사의 전체 고용률 차이를 연령 비중과 각 집단의 고용률 차이로 분해한 결과, KLIPS 98년 표본은 고령층을 과도하게 포함하고 있고, 15~34세 연령층의 고용률은 과소하게 포착하는 것으로 나타났다. 따라서 KLIPS 98년 표본의 고용률을 높이기 위해서는 55세 이상의 고령자 비중을 낮추는 것과 함께 35~54세의 표본을 늘리고, 15~34세의 경우 취업자 여부를 좀 더 명확히 판단하여 그 비중을 높이는 것이 필요하다.

마지막으로 취업자를 종사상 지위로 나누어 보면, KLIPS 98년 표본에서 계산된 임금근로자 비중은 경활조사에 비하여 과소하게, 비임금근로자 비중은 과도하게 나타난다. 특히, KLIPS 98년 표본에서는 임시일용직 비중이 낮게 나타나고 있는데 이는 경활조사에 비하여 임시일용직 표본이 더 적게 포함되어 있기 때문일 수도 있고, 아니면 면접 대상자가 자신을 임시일용직으로 보고하기보다는 상용직으로 보고하기 때문일 가능성도 존재한다. 한편, 비임금근로자의 비중이 높은 것은 이들이 임금근로자에 비하여 상대적으로 높은 응답률을 보이는 것과 관계가 있는 것으로 판단된다.²⁷⁾

제3절 2009년 표본의 경제활동참가율, 고용률, 실업률 비교

KLIPS 1998년 표본은 시간이 지남에 따라 표본 마모가 심해질 뿐 아니라 도시지역만을 모집단으로 하고 있다는 한계가 있다. 따라서 2009년

27) 1998~1999년, 2009~2010년, 2018~2019년을 대상으로 패널을 만들어 표본이탈을 계산한 결과, 비임금근로자의 표본유지율이 임금근로자보다 약 3~4%포인트 높게 나타나고 있다.

에는 이러한 한계를 극복하기 위하여 제주도를 포함한 우리나라 전 지역을 모집단으로 하여 1,415가구를 기존 가구에 추가하였고, 이에 따라 가중치도 새롭게 구축되었다. 이하에서는 KLIPS 2009년 표본을 대상으로 계산된 경제활동지표와 통계청의 경제활동인구조사와 비교하기로 한다.

1. 전체 인구 대상

<표 3-9>는 KLIPS 2009년 표본(이하에서는 KLIPS 09년 표본으로 표시)을 대상으로 2009~2020년까지 횡단면 가중치를 적용하여 15세 이상 인구, 취업자, 실업자 및 경제활동참가율, 고용률, 실업률을 계산하고 이를 통계청 경황조사에서 계산된 해당 지표들과 비교한 것이다.

우선 15세 이상 인구의 절대적인 수를 보면 2009~2020년 기간 동안 KLIPS와 경황조사 사이에 큰 차이를 보이지는 않으나 KLIPS 09년 표본이 경황조사보다 15세 이상 인구를 일관되게 약 2% 정도 과다하게 추정하고 있다. 이러한 차이는 KLIPS 98년 표본에서 나타난 9% 차이보다는 훨씬 작은 것이다.

한편, KLIPS 09년 표본의 취업자 수는 경황조사보다 약 2% 작게 나타나고 있는데 이 역시 KLIPS 98년 표본과 경황조사 사이에서 나타난 취업자 수의 차이보다는 훨씬 작은 것이라 할 수 있다. 하지만 KLIPS 09년 실업자 수는 여전히 경황조사와 큰 차이를 보이고 있으며 최근 들어 약 20~30% 낮게 추정되고 있다.

KLIPS 09년 표본의 취업자 수와 실업자 수의 과소추정은 경제활동참가율과 고용률에도 편의를 불러오고 있는데, 경황조사에 비하여 경제활동참가율은 약 5%, 고용률은 약 4% 정도 낮게 나타나고 있다. KLIPS 09년 표본에서 가장 큰 차이를 보이는 지표는 실업률인데 해마다 변동률이 크며, 2020년 현재는 경황조사의 실업률에 비하여 약 27% 낮게 나타나고 있다. KLIPS 09년 표본이 비경제활동인구를 과다하게 포함하고 있다는 점에서 KLIPS 98년 표본과 유사하다고 말할 수 있지만 KLIPS 09년 표본이 KLIPS 98년 표본보다는 경황조사와의 차이가 더 작다는 것은 긍정적이라고 볼 수 있을 것이다.

〈표 3-9〉 KLIPS 2009년 표본과 경제활동인구조사의 고용지표 비교

(단위: 천 명, %)

	KLIPS						경제활동인구조사						KLIPS/경제조사					
	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률
2009	41,079	22,320	1,056	56.9	54.3	4.5	40,301	23,688	894	61.0	58.8	3.6	1.02	0.94	1.18	0.93	0.92	1.25
2010	40,204	22,427	622	57.3	55.8	2.7	40,825	24,033	924	61.1	58.9	3.7	0.98	0.93	0.67	0.94	0.95	0.73
2011	42,165	24,124	739	59.0	57.2	3.0	41,387	24,527	863	61.3	59.3	3.4	1.02	0.98	0.86	0.96	0.96	0.88
2012	42,623	24,385	801	59.1	57.2	3.2	41,857	24,955	826	61.6	59.6	3.2	1.02	0.98	0.97	0.96	0.96	1.00
2013	43,037	24,687	655	58.9	57.4	2.6	42,304	25,299	808	61.7	59.8	3.1	1.02	0.98	0.81	0.95	0.96	0.84
2014	43,533	25,250	695	59.6	58.0	2.7	42,795	25,897	939	62.7	60.5	3.5	1.02	0.98	0.74	0.95	0.96	0.77
2015	44,115	25,362	691	59.1	57.5	2.7	43,239	26,178	976	62.8	60.5	3.6	1.02	0.97	0.71	0.94	0.95	0.75
2016	44,353	25,582	693	59.2	57.7	2.6	43,606	26,409	1,009	62.9	60.6	3.7	1.02	0.97	0.69	0.94	0.95	0.70
2017	44,638	26,093	652	59.9	58.5	2.4	43,931	26,725	1,023	63.2	60.8	3.7	1.02	0.98	0.64	0.95	0.96	0.65
2018	45,017	26,443	671	60.2	58.7	2.5	44,182	26,822	1,073	63.1	60.7	3.8	1.02	0.99	0.63	0.95	0.97	0.66
2019	45,274	26,341	852	60.1	58.2	3.1	44,504	27,123	1,063	63.3	60.9	3.8	1.02	0.97	0.80	0.95	0.96	0.82
2020	45,484	26,292	797	59.6	57.8	2.9	44,785	26,904	1,108	62.5	60.1	4.0	1.02	0.98	0.72	0.95	0.96	0.73

주: 경활자료는 KOSIS를 이용하여 구한 것임. 모든 수치는 가중치 부여.

KLIPS 09년 표본이 경제활동참가율, 고용률, 실업률을 과소하게 추정하고 있는 것이 가중치의 문제인지를 확인하기 위하여 KLIPS 09년 표본에 가중치를 사용하지 않고 이들 고용지표를 계산한 결과를 <표 3-10>에 제시한다.

<표 3-9>와 <표 3-10>을 비교하면 KLIPS 09년 표본에 가중치를 사용하지 않았을 경우 경황조사와의 격차가 더 커짐을 알 수 있다. 따라서 KLIPS 09년 표본으로 고용지표를 계산할 경우에는 가중치를 사용하는 것이 더 정확한 값을 얻을 수 있다고 할 것이다. 참고로 KLIPS 98년 표본에서는 오히려 가중치를 적용할 경우 격차가 더 크게 나타남을 상기할 필요가 있다. 따라서 KLIPS 98년 표본은 가중치를 조정할 필요가 시급한 반면, KLIPS 09년 표본은 가중치 조정이 상대적으로 덜 중요하다고 볼 수 있다. 하지만 KLIPS 09년 표본에서도 여전히 통계청 경황조사에 비해

<표 3-10> KLIPS 09년 표본과 경황조사와의 비교

(단위: %)

	KLIPS 09년 표본			경황조사			KLIPS/경황조사		
	경황 참가율	고용률	실업률	경황 참가율	고용률	실업률	경황 참가율	고용률	실업률
2009	56.1	53.7	4.2	61.0	58.8	3.6	0.92	0.91	1.17
2010	56.8	55.3	2.6	61.1	58.9	3.7	0.93	0.94	0.70
2011	57.5	56.0	2.6	61.3	59.3	3.4	0.94	0.94	0.76
2012	58.3	56.6	2.8	61.6	59.6	3.2	0.95	0.95	0.88
2013	58.1	56.8	2.3	61.7	59.8	3.1	0.94	0.95	0.74
2014	59.5	58.2	2.3	62.7	60.5	3.5	0.95	0.96	0.66
2015	58.1	56.8	2.3	62.8	60.5	3.6	0.93	0.94	0.64
2016	58.5	57.2	2.1	62.9	60.6	3.7	0.93	0.94	0.57
2017	58.7	57.4	2.1	63.2	60.8	3.7	0.93	0.94	0.57
2018	58.9	57.7	2.0	63.1	60.7	3.8	0.93	0.95	0.53
2019	58.7	57.2	2.6	63.3	60.9	3.8	0.93	0.94	0.68
2020	57.9	56.5	2.3	62.5	60.1	4.0	0.93	0.94	0.58

주: 1) 경제활동인구자료는 KOSIS를 통해서 구한 것임.

2) KLIPS 수치는 가중치를 사용하지 않은 것임.

여 경제활동참가율, 고용률, 실업률이 낮게 나타나기 때문에 이 점을 개선하기 위한 방안은 필요할 것이다.

2. 남성과 여성

KLIPS 09년 표본을 남성과 여성으로 구분하여 경제활동참가율, 고용률, 실업률을 계산하고 이를 경황조사 자료의 해당 지표와 비교한 결과는 <표 3-11>과 <표 3-12>에 각각 제시된다. 앞서와 마찬가지로 가중치를 적용한 경우와 그렇지 않은 경우를 구분하였으며, 경황조사는 전국을 대상으로 한 값으로 KOSIS를 통해서 얻어진 값이다. 따라서 <표 3-11>과 <표 3-12>에 나타난 경황조사의 수치는 가중치를 적용한 값이다.

남성의 경제활동참가율과 고용률을 보면 가중치 적용 여부와 상관없이 KLIPS 09년 표본의 수치가 경황조사보다 낮게 나타남을 알 수 있다. 또한, 가중치를 부여했을 때가 부여하지 않았을 때보다 오히려 경황과의 격차가 약간 커지는 것을 알 수 있다. 이러한 현상은 새로운 표본이 시작된 2009년에도 발생하기 때문에 KLIPS 09년 표본의 가중치 설정에 있어서 약간의 문제가 있었던 것으로 보여진다. 한편, KLIPS 09년 실업률도 경황 실업률보다 다소 작게 나타나고 있는데 이는 <표 3-9>에서 보듯이 KLIPS 09년 표본의 실업자 수 자체가 경황보다 적게 포착되기 때문으로 볼 수 있다.

여성을 대상으로 한 <표 3-12>를 보면, 남성과 마찬가지로 KLIPS 09년 표본의 경제활동참가율과 고용률이 경황조사의 해당 지표보다 낮게 나타나고 있으며 KLIPS의 실업률도 2009년을 제외하고는 경황조사보다 지속적으로 낮게 나타나고 있다. 하지만 남성과의 차이점은 가중치 적용 여부인데, 남성의 경우에는 가중치를 적용하지 않을 때 경황 수치와 더 근접한 반면, 여성의 경우에는 가중치를 적용했을 때 경황과의 차이가 줄어드는 것으로 나타나고 있다. 하지만 남성의 경우에 가중치 적용 여부에 따른 차이가 크지 않기 때문에 KLIPS 09년 표본의 가중치 설정에 있어서 큰 문제는 없다고 판단된다.

〈표 3-11〉 KLIPS 09년 남성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교

(단위: %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률
2009	68.9	65.9	4.3	70.0	67.2	4.0	73.2	70.2	4.1	-4.3	-4.3	0.2	-3.2	-3.0	-0.1
2010	69.5	67.7	2.7	70.5	68.6	2.7	73.2	70.3	4.0	-3.7	-2.6	-1.3	-2.7	-1.7	-1.3
2011	71.4	69.1	3.2	71.5	69.5	2.8	73.4	70.7	3.6	-2.0	-1.6	-0.4	-1.9	-1.2	-0.8
2012	71.2	68.9	3.2	71.9	69.9	2.7	73.5	71.1	3.4	-2.3	-2.2	-0.2	-1.6	-1.2	-0.7
2013	70.9	68.8	3.0	71.7	69.9	2.5	73.5	71.1	3.3	-2.6	-2.3	-0.3	-1.8	-1.2	-0.8
2014	71.9	70.0	2.7	74.2	72.7	2.1	74.4	71.7	3.5	-2.5	-1.7	-0.8	-0.2	1.0	-1.4
2015	70.6	69.0	2.2	71.7	70.2	2.1	74.1	71.4	3.6	-3.5	-2.4	-1.4	-2.4	-1.2	-1.5
2016	70.4	68.5	2.7	71.8	70.3	2.1	74.0	71.2	3.8	-3.6	-2.7	-1.1	-2.2	-0.9	-1.7
2017	70.0	68.2	2.5	71.1	69.5	2.2	74.1	71.2	3.8	-4.1	-3.0	-1.3	-3.0	-1.7	-1.6
2018	70.0	68.4	2.4	70.9	69.4	2.1	73.7	70.8	3.9	-3.7	-2.4	-1.5	-2.8	-1.4	-1.8
2019	69.9	67.7	3.2	70.8	69.0	2.6	73.5	70.7	3.9	-3.6	-3.0	-0.7	-2.7	-1.7	-1.3
2020	69.7	67.6	3.0	69.8	68.3	2.2	72.6	69.8	3.9	-2.9	-2.2	-0.9	-2.8	-1.5	-1.7

〈표 3-12〉 KLIPS 09년 여성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교

(단위: %)

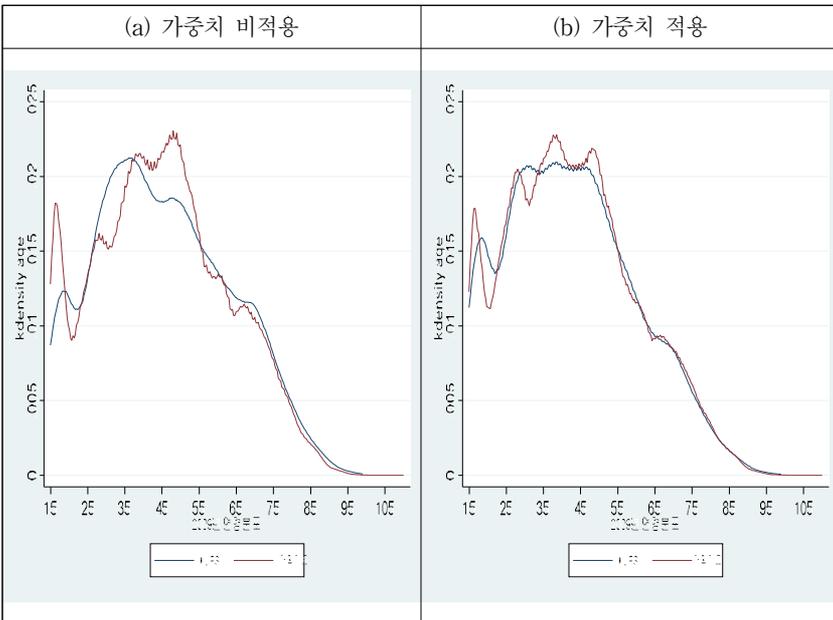
	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률
2009	45.0	42.8	4.8	43.3	41.3	4.5	49.3	47.8	3.0	-4.3	-5.0	1.8	-6.0	-6.5	1.5
2010	45.5	44.3	2.7	44.3	43.2	2.6	49.6	47.9	3.3	-4.1	-3.6	-0.6	-5.3	-4.7	-0.7
2011	46.6	45.4	2.6	44.8	43.7	2.4	49.8	48.3	3.1	-3.2	-2.9	-0.5	-5.0	-4.6	-0.7
2012	47.0	45.5	3.1	45.7	44.4	2.9	50.1	48.6	3.0	-3.1	-3.1	0.1	-4.4	-4.2	-0.1
2013	46.9	46.0	2.0	45.6	44.8	1.9	50.3	48.9	2.8	-3.4	-2.9	-0.8	-4.7	-4.1	-0.9
2014	47.4	46.1	2.7	46.4	45.2	2.5	51.5	49.7	3.5	-4.1	-3.6	-0.8	-5.1	-4.5	-1.0
2015	47.6	46.0	3.3	45.7	44.5	2.5	51.9	50.1	3.5	-4.3	-4.1	-0.2	-6.2	-5.6	-1.0
2016	48.2	46.9	2.5	46.2	45.2	2.2	52.2	50.3	3.6	-4.0	-3.4	-1.1	-6.0	-5.1	-1.4
2017	49.9	48.7	2.4	47.1	46.2	1.9	52.7	50.8	3.5	-2.8	-2.1	-1.1	-5.6	-4.6	-1.6
2018	50.5	49.2	2.6	47.8	46.9	2.0	52.9	50.9	3.7	-2.4	-1.7	-1.1	-5.1	-4.0	-1.7
2019	50.2	48.7	3.1	47.7	46.5	2.5	53.5	51.6	3.6	-3.3	-2.9	-0.5	-5.8	-5.1	-1.1
2020	49.5	48.1	2.8	47.0	45.8	2.5	52.8	50.7	4.0	-3.3	-2.6	-1.2	-5.8	-4.9	-1.5

3. 연령별 비교

KLIPS 09년 표본과 경찰조사의 연령 분포를 비교하기 위하여 앞서와 마찬가지로 2009년, 2015년, 2020년 3개년도의 연령 분포를 Kernel method 를 이용하여 계산하기로 한다. KLIPS 09년 표본은 가중치를 사용한 것과 가중치를 사용하지 않은 경우를 모두 사용하며, 경찰자료는 원자료를 이용하여 가중치를 적용해서 연령 분포를 구하기로 한다.

[그림 3-4]부터 [그림 3-5]는 2009년, 2015년, 2020년의 KLIPS 09년 표본과 경찰조사의 15세 이상 연령 분포를 보여주고 있다. 여기서 한 가지 주의할 사항은 앞의 [그림 3-2]의 경찰 2009년, [그림 3-3]의 경찰 2020년과 아래의 경찰 2009년과 2020년은 다르다는 것이다. 전자는 경찰조사 중 도시 거주자만을 대상으로 한 것이고, 아래 경찰의 연령 분포는 우리나라 전 지역을 대상으로 한 연령 분포이기 때문이다.

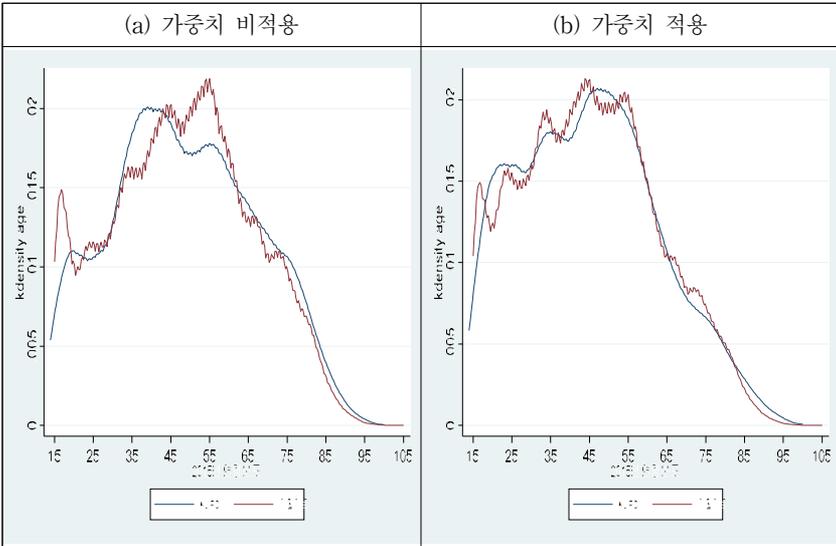
[그림 3-4] 2009년 KLIPS 09년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 45.5, 표준편차: 17.8
 경찰조사: 평균: 45.2, 표준편차: 17.5

주: KLIPS: 평균: 42.7, 표준편차: 17.1
 경찰조사: 평균: 43.2, 표준편차: 16.9

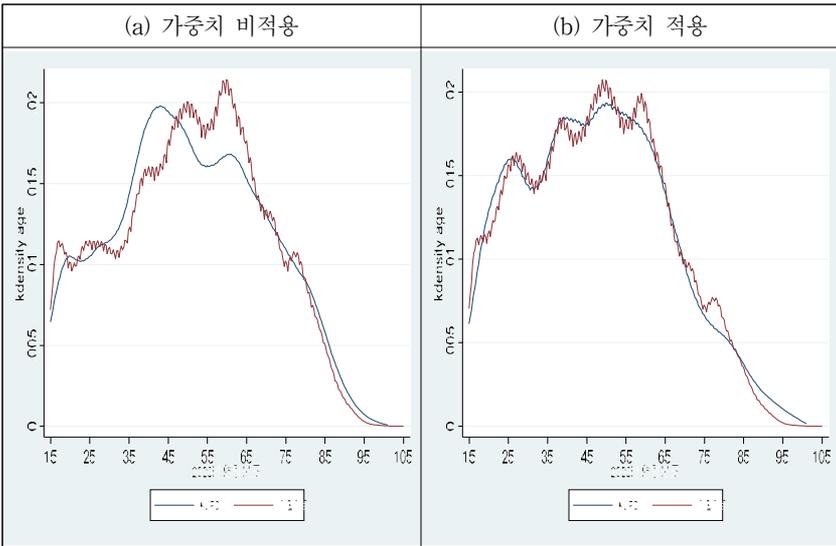
(그림 3-5) 2015년 KLIPS 09년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 48.6, 표준편차: 18.4
경찰조사: 평균: 48.0, 표준편차: 18.1

주: KLIPS: 평균: 45.1, 표준편차: 17.8
경찰조사: 평균: 45.2, 표준편차: 17.5

(그림 3-6) 2020년 KLIPS 09년 표본과 경찰조사의 연령 분포



주: KLIPS: 평균: 50.4, 표준편차: 18.9
경찰조사: 평균: 50.7, 표준편차: 18.5

주: KLIPS: 평균: 47.3, 표준편차: 18.3
경찰조사: 평균: 47.4, 표준편차: 17.9

위의 연령 분포를 보면 전반적으로 KLIPS 09년 표본과 경찰조사 사이에 큰 차이가 없는 것을 알 수 있다. 특히 두 자료 사이에 평균 연령은 아주 근접하게 나타나고 있다. 이는 KLIPS 98년 표본에서 평균 연령이 경찰보다 약 2세 정도 많은 것과는 아주 대조적이다.

<표 3-13>은 KLIPS 09년 표본과 경찰조사의 연령별 고용률과 둘 사이의 차이를 보여준다. 경제활동참가율과 실업률에 대한 내용은 각각 <부표 3-7>와 <부표 3-8>에 제시하는데 경제활동참가율의 차이는 고용률과 매우 유사하기 때문에 고용률만을 대상으로 설명하기로 한다.

KLIPS 09년 표본과 경찰의 고용률에서 가장 큰 차이를 보이는 연령층은 35세 미만으로서 KLIPS 09년 표본이 경찰보다 5%에서 9%까지 낮게 나타나고 있다. 55세 이상도 2009년부터 2015년까지는 2~4%의 차이를 보였으나 그 이후에는 차이가 사라지고 있으며, 35~54세는 KIPS와 경찰조사 사이에 고용률의 차이가 크지 않은 것으로 나타난다. 결국, KLIPS

<표 3-13> KLIPS 09년 표본과 경찰조사와의 연령별 고용률 비교

(단위:%)

	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2009	42.4	48.8	-6.4	73.5	75.5	-2.0	40.3	44.5	-4.2
2010	43.7	49.0	-5.3	76.3	75.8	0.5	40.7	44.4	-3.7
2011	45.0	49.3	-4.3	77.8	76.2	1.6	42.0	45.3	-3.3
2012	43.9	49.8	-5.9	77.1	76.4	0.7	44.4	46.4	-2.0
2013	43.0	49.5	-6.5	77.9	76.8	1.1	45.3	47.4	-2.1
2014	44.2	50.5	-6.3	78.1	77.4	0.7	46.2	48.4	-2.2
2015	43.5	51.1	-7.6	77.2	77.5	-0.3	46.9	48.3	-1.4
2016	41.9	51.0	-9.1	77.4	77.8	-0.4	48.5	48.8	-0.3
2017	42.8	51.0	-8.2	78.2	78.2	0.0	49.6	49.5	0.1
2018	44.0	51.4	-7.4	77.9	77.9	0.0	49.8	49.4	0.4
2019	43.2	52.2	-9.0	77.2	77.7	-0.5	49.9	50.1	-0.2
2020	42.9	51.3	-8.4	76.9	76.3	0.6	49.3	50.2	-0.9

주: 모든 수치는 가중치를 사용하여 구한 것임.

09년 표본의 고용률이 경찰보다 낮게 나타나는 이유는 35세 미만의 고용률이 낮게 추정되기 때문으로 볼 수 있다. 또한, 이러한 현상은 KLIPS 98년 표본에서도 동일하게 나타나고 있음을 상기할 필요가 있다. 따라서 이들 젊은 층에 대하여 취업자 표본을 확충하거나 가중치를 조정할 필요가 있어 보인다.

KLIPS 09표본과 경찰조사의 전체 고용률 차이의 원인을 좀 더 살펴보기 위해서 앞의 식 (3)을 이용하여 2020년을 분해한 결과를 <표 3-14>에 제시한다. 표를 보면 연령 비중에 있어서는 두 자료 사이에 거의 차이가 없는 것으로 나타나고 있다. 따라서 KLIPS 98년 표본에서 젊은 층의 비중이 작은 것과는 달리 KLIPS 09년 표본에서는 연령 비중의 차이보다는 젊은 층에서 취업자라고 응답한 비중이 작게 나타나는 것이 문제라고 할 수 있다.

2020년 KLIPS 09년 표본의 15~34세 사람들의 고용률이 왜 낮은가를 보기 위해서 연령대를 좀 더 세분화하여 경찰조사와 비교한 결과는 <표 3-15>에 제시된다.

<표 3-15>를 보면 15~34세를 5세 단위로 분해하였을 때 이들이 전체 인구에서 차지하는 비중은 KLIPS 09년 표본이나 경찰조사 사이에 큰 차이를 보이지 않음을 알 수 있다. 이 두 자료 사이에 최대 격차는 0.7%포인트, 최소 격차는 -0.7%포인트로 아주 근접하게 나타난다. 따라서 15~34세의 고용률 차이는 세부 연령의 비중 차이는 아닌 것으로 결론지을 수 있다.

한편, 고용률을 보면 각 연령층마다 상당한 차이를 보이는데 특히 20~

<표 3-14> 2020년 KLIPS 09년 표본과 경찰조사와의 고용률 차이 분해

	15~34세		35~54세		55세 이상		전체
	비중	고용률	비중	고용률	비중	고용률	고용률
KLIPS 09표본	0.274	42.9	0.371	76.9	0.355	49.3	57.8
경찰자료	0.272	51.3	0.366	76.3	0.361	50.2	60.1
차이	0.002	-8.4	0.005	0.6	-0.006	-0.9	-2.3

주: 비중의 합은 1이고 고용률의 단위는 %. 모든 수치는 가중치를 사용하여 구한 것임.

〈표 3-15〉 2020년 KLIPS 09년 표본과 경찰조사와의 15~34세 고용률 차이 분해

연령층	KLIPS		경찰		KLIPS-경찰	
	비중	고용률	비중	고용률	비중	고용률
15~19	0.055	1.9	0.055	6.6	0.000	-4.7
20~24	0.072	27.8	0.064	41.0	0.007	-13.2
25~29	0.082	60.1	0.080	67.6	0.002	-7.4
30~34	0.066	72.2	0.073	76.0	-0.007	-3.9

주: 비중의 합은 1이고 고용률의 단위는 %. 모든 수치는 가중치를 사용하여 구한 것임.

24세에서 가장 큰 차이를 보이며, 25~29세에서도 7.4%포인트의 적지 않은 차이를 나타내고 있다. 20~24세는 대학에 재학 중인 연령이라고 볼 수 있는데 KLIPS에서 이 연령층의 고용률이 낮게 추정되는 것은 결국 경찰조사에 비하여 아르바이트 등 임시일용직 일자리가 취업으로 잘 포착되지 못하고 있음을 시사한다. 이제 이를 확인하기 위하여 종사상 지위에 따른 고용률을 살펴보기로 한다.

4. 종사상 지위

지금까지 분석을 통하여 KLIPS 09년 표본은 경찰조사에 비하여 경제활동참가율과 고용률이 낮는데 이러한 현상은 남녀 모두에게 공통적으로 나타나지만, 연령별로는 특히 재학 중에 있는 사람들의 고용률이 작게 나타나고 있음을 발견하였다. 따라서 아르바이트 등의 임시일용직으로의 취업이 KLIPS에서 잘 포착되지 않을 가능성을 제시하였다.

〈표 3-16〉은 종사상 지위로 구분하여 KLIPS 09년 표본과 경찰조사의 취업자 수를 계산한 결과이다. 취업자 중 종사상 지위별 비중을 계산한 결과는 〈부표 3-9〉에 제시한다. 2020년 현재 KLIPS 09년 표본의 취업자 수는 경찰조사보다 약 61만 명 적은 것으로 나타나고 있는데, 이를 종사상 지위로 나누어 보면 상용직은 약 76만 명 과대 추정하고 있고, 임시일용직은 103만 명 과소 추정하고 있다. 한편, 자영업주와 무급가족 종사자는 두 자료 사이에 상대적으로 작은 차이를 보이고 있다. 이러한 현상은 〈부표 3-9〉에서도 재확인된다.

〈표 3-16〉 KLIPS 09년 표본과 경찰조사와의 증사상 지위 분포 비교

(단위: 천 명)

	KLIPS					경찰조사					KLIPS-경찰조사				
	취업자	상용직	임시 일용	자영 업주	무급 가족	취업자	상용직	임시 일용	자영 업주	무급 가족	취업자	상용직	임시 일용	자영 업주	무급 가족
2009	22,320	11,830	3,763	5,430	1,297	23,688	9,479	7,107	5,749	1,352	-1,368	2,351	-3,344	-320	-55
2010	22,427	11,897	3,953	5,328	1,248	24,033	10,178	6,933	5,642	1,279	-1,606	1,719	-2,979	-314	-31
2011	24,126	12,956	4,454	5,478	1,234	24,526	10,785	6,811	5,657	1,273	-401	2,171	-2,357	-179	-39
2012	24,388	13,144	4,573	5,480	1,190	24,955	11,250	6,671	5,768	1,266	-567	1,895	-2,098	-288	-76
2013	24,688	13,462	4,554	5,471	1,200	25,299	11,847	6,518	5,703	1,232	-612	1,615	-1,964	-232	-31
2014	25,251	13,843	4,772	5,415	1,221	25,897	12,319	6,639	5,720	1,219	-646	1,523	-1,867	-305	2
2015	25,364	14,121	4,708	5,424	1,111	26,178	12,716	6,686	5,622	1,153	-814	1,405	-1,978	-199	-42
2016	25,584	14,226	4,770	5,462	1,126	26,409	13,062	6,607	5,614	1,126	-825	1,164	-1,838	-152	0
2017	26,094	14,554	4,986	5,428	1,127	26,725	13,428	6,506	5,681	1,110	-631	1,126	-1,520	-254	17
2018	26,443	15,025	4,888	5,406	1,123	26,822	13,772	6,311	5,638	1,101	-379	1,253	-1,423	-231	22
2019	26,339	15,105	4,961	5,247	1,026	27,123	14,216	6,224	5,606	1,077	-783	889	-1,263	-359	-51
2020	26,292	15,283	4,785	5,241	980	26,904	14,521	5,811	5,531	1,042	-612	762	-1,025	-290	-62

주: 자영업주는 고용주와 자영업자를 합한 것임. 모든 수치는 가중치를 사용하여 구한 것임. 차이는 KLIPS-경찰조사.

결국, 앞서 예상한 바와 같이 KLIPS 09년 표본의 고용률이 경찰조사보다 작게 나타나는 주된 이유는 KLIPS에서 임시일용직 근로자 수가 너무 과소하게 포함되어 있기 때문이라고 볼 수 있다. 이는 30세 미만 연령층에서 고용률이 왜 낮게 나타나는지를 설명하는 단초가 된다.

한편, 최근에 들어오면서 KLIPS 09년 표본의 상용직 근로자 수는 상대적으로 감소하고, 임시일용직 근로자 수는 상대적으로 증가하여 경찰조사와의 격차가 감소하고 있는 것은 고무적이지만 여전히 격차의 크기는 무시할 만한 수준은 아니다. 결국, 이러한 문제를 해결하기 위해서는 KLIPS에서 상용직의 구분을 엄격하게 적용해야 할 뿐 아니라 임시일용직 취업자를 설문과정에서 좀 더 많이 포착하는 노력이 필요할 것이다.

5. 소 결

지금까지 KLIPS 09년 표본을 이용하여 경제활동참가율, 고용률, 실업률을 구하였고 이를 경찰조사와 비교하였다. KLIPS 09년 표본은 KLIPS 98년 표본과 비교할 때 좀 더 경찰조사에 근접하고 있다고 말할 수 있을 것이다. 그럼에도 불구하고 몇 가지 측면에서 여전히 개선의 여지가 존재한다.

첫째, KLIPS 09년 표본에서 계산된 경제활동참가율과 고용률은 KLIPS 98년 표본과 마찬가지로 여전히 경찰조사보다 낮게 나타나고 있다. 그 원인으로서는 KLIPS 09년 표본에서 15세 이상 인구가 과대하게 추정되고, 취업자 수는 과소하게 추정되기 때문이다.

둘째, KLIPS 09년 표본에서 구해진 경제활동참가율과 고용률이 경찰조사보다 낮게 나타나는 현상은 남성과 여성 모두에서 동일하게 나타나고 있다. 따라서 KLIPS 09년 표본이 남녀 모두에 있어서 비취업자를 과도하게 포함하고 있다고 할 수 있다.

셋째, KLIPS 09년 표본과 경찰조사의 고용률을 연령별로 구분한 결과, 연령 비중은 두 자료 사이에는 큰 차이가 발생하고 있지 않지만, 30세 미만 사람들의 고용률에서는 큰 격차를 보이고 있다. 이는 KLIPS가 청년 특히 20~24세 대학 재학 중인 사람들 중 취업자를 파악하는 데 어려움을

겪고 있는 것으로 볼 수 있다.

마지막으로 취업자를 종사상 지위로 나누어 본 결과, KLIPS 09년 표본에서 계산된 상용직 근로자 수는 경찰조사보다 과다하며, 임시일용직 근로자 수는 과소하게 나타나고 있다. 이는 경찰조사에 비하여 KLIPS 09년 표본이 자신을 임시일용직보다는 상용직으로 보고할 가능성과 아예 자신을 미취업자로 보고할 가능성 모두를 시사하고 있다. 후자의 경우는 자신이 취업자로 분류될 경우 상대적으로 많은 설문문항에 답해야 하는 부담 때문에 발생할 수 있을 것이다(이상호, 2015).

제4절 2018년 표본의 경제활동참가율, 고용률, 실업률 비교

KLIPS 2009년 표본은 9년이 경과한 2018년에 약 5.2%의 표본 손실이 발생하였고, 패널 참가자가 또 다시 고령화됨에 따라 횡단면 자료로서의 대표성에서 한계를 가지게 되었다. 따라서 이를 극복하기 위하여 2018년에 5,044가구를 기존 가구에 추가하였고, 가중치도 새롭게 부가하였다.

2018년 현재 KLIPS에 포함된 총가구 수는 12,134가구이며, 개인으로는 23,972명에 이른다. 이는 2009년 6,721가구, 14,489명에 비하여 거의 두 배로 표본이 늘어난 것이다. 이하에서는 앞서와 같이 KLIPS 2018년 표본(이하 KLIPS 18년 표본이라 함)을 대상으로 경제활동지표를 계산하고 이를 통계청의 경제활동인구조사와 비교하기로 한다.

1. 전체 인구 대상

<표 3-17>은 KLIPS 18년 표본을 대상으로 2018~2020년까지 횡단면 가중치를 적용하여 15세 이상 인구, 취업자, 실업자 및 경제활동참가율, 고용률, 실업률을 계산하고 이를 통계청 경찰조사에서 계산된 해당 지표들과 비교한 것이다.

우선 15세 이상 인구를 보면 KLIPS 18년 표본이 경활조사보다 약 2% 과대 추정하고 있는데 이는 KLIPS 09년 표본에서도 공통적으로 나타나고 있는 현상이다. 한편, 취업자 수는 KLIPS 18년 표본에서 약 3~4% 과소 추정되고 있고, 특히 실업자 수는 20~30% 과소 추정되고 있다. 이러한 차이로 인하여 KLIPS 18년 표본은 경활에 비하여 경제활동참가율은 약 6%, 고용률은 5% 작게 나타나며, 실업률의 경우는 상당한 차이를 보이고 있다. 결국, 앞의 다른 KLIPS 코호트와 마찬가지로 KLIPS 18년 표본도 경활조사에 비하여 비경제활동인구가 과대하게 포함되어 있다고 할 수 있을 것이다.

〈표 3-17〉 KLIPS 18년 표본과 경제활동인구조사의 고용지표 비교

(단위: 명, %)

		2018	2019	2020
KLIPS	15세 이상 인구	45,017	45,274	45,484
	취업자 수	25,913	26,154	25,960
	실업자 수	676	819	888
	경활참가율	59.1	59.6	59.0
	고용률	57.6	57.8	57.1
	실업률	2.5	3.0	3.3
경제활동 인구조사	15세 이상 인구	44,182	44,504	44,785
	취업자 수	26,822	27,123	26,904
	실업자 수	1,073	1,063	1,108
	경활참가율	63.1	63.3	62.5
	고용률	60.7	60.9	60.1
	실업률	3.8	3.8	4.0
KLIPS/ 경활조사	15세 이상 인구	1.02	1.02	1.02
	취업자 수	0.97	0.96	0.96
	실업자 수	0.63	0.77	0.80
	경활참가율	0.94	0.94	0.94
	고용률	0.95	0.95	0.95
	실업률	0.66	0.79	0.83

주: 모든 수치는 가중치를 적용하여 얻은 것임.

〈표 3-18〉 KLIPS 18년 표본과 경찰조사 고용지표 비교(가중치 사용 안함)

	2018	2019	2020
경찰참가율	0.90	0.89	0.89
고용률	0.91	0.91	0.90
실업률	0.58	0.68	0.66

주: 모든 수치는 해당지표의 KLIPS/경찰조사.

위의 차이가 가중치와 관련이 있는지를 확인하기 위해서 KLIPS 18년 표본에 가중치를 적용하지 않은 경제활동참가율과 고용률, 실업률을 계산하여 경찰조사와 비교하였는데 그 결과는 <표 3-18>에 제시된다. 표에서 보듯이 오히려 두 자료 사이의 격차는 더 증가하는 것으로 나타나고 있다. 따라서 KLIPS 18년 가중치가 긍정적으로 작용하고 있다고 볼 수 있다.

2. 남성과 여성

KLIPS 18년 표본을 남성과 여성으로 구분하여 경제활동참가율, 고용률, 실업률을 계산하고 이를 경찰조사 자료의 해당 지표와 비교한 결과는 <표 3-19>과 <표 3-20>에 각각 제시된다. 가중치를 적용한 경우와 그렇지 않은 경우로 구분하였으며, 경찰조사는 전국을 대상으로 한 것이기 때문에 가중치를 적용한 값이라고 할 수 있다.

<표 3-19>과 <표 3-20>를 보면 KLIPS 18년 표본에 가중치를 적용하였을 때가 그렇지 않을 때보다 경찰조사에 더 근접하고 있음을 알 수 있다. 이는 앞서 KLIPS 09년 표본에서 남성의 경우에는 가중치를 적용하였을 때가 오히려 경찰조사와의 격차가 더 커진 것과는 대조된다. 결국 KLIPS의 코호트에 따라 가중치 적용 여부의 결과가 다른 양상을 보인다는 것인데 이는 KLIPS의 코호트에 맞는 가중치를 재조정할 필요성을 제기한다.

하지만 KLIPS 18년 표본에 가중치를 부여하였을 경우에도 경제활동참가율, 고용률, 실업률 모두에서 경찰보다 낮은 수치를 보이고 있는데 특히, 여성에서 격차가 더 크게 나타나고 있음을 알 수 있다. 따라서 앞서

〈표 3-19〉 KLIPS 2009년 남성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교

(단위: %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률
2018	70.2	68.4	2.5	69.4	67.9	2.2	73.7	70.8	3.9	-3.5	-2.4	-1.4	-4.3	-2.9	-1.7
2019	70.6	68.6	2.9	69.0	67.3	2.5	73.5	70.7	3.9	-2.9	-2.1	-1.0	-4.5	-3.4	-1.4
2020	70.4	68.0	3.4	68.1	66.4	2.6	72.6	69.8	3.9	-2.2	-1.8	-0.5	-4.5	-3.4	-1.3

〈표 3-20〉 KLIPS 2009년 여성 표본의 가중치 적용 여부에 따른 경찰조사와의 고용지표 비교

(단위: %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰(가중치 적용)			(A-C)			(B-C)		
	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률
2018	48.0	46.7	2.7	45.2	44.2	2.2	52.9	50.9	3.7	-4.9	-4.2	-1.0	-7.7	-6.7	-1.5
2019	48.6	47.0	3.2	45.5	44.3	2.6	53.5	51.6	3.6	-4.9	-4.6	-0.4	-8.0	-7.3	-1.0
2020	47.7	46.2	3.2	44.4	43.2	2.7	52.8	50.7	4.0	-5.1	-4.5	-0.8	-8.4	-7.5	-1.3

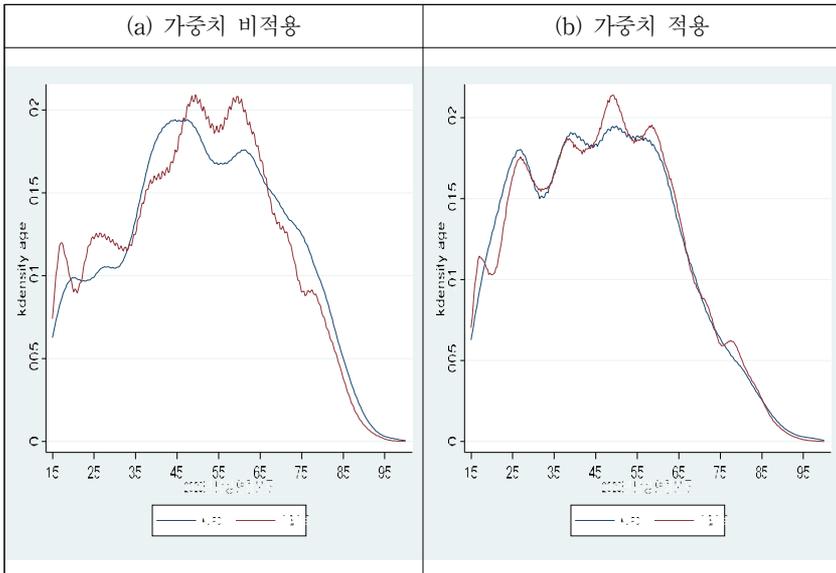
KLIPS 18년 표본 전체에서 경활조사와의 차이는 남성보다는 여성에서 기인한다고 말할 수 있을 것이다.

3. 연령별 비교

앞선 연령별 비교에서는 연도별로 연령 분포를 구하여 KLIPS 표본과 경활조사 표본을 비교하였는데 KLIPS 18년 표본은 경활조사와 비교할 때 남성과 여성에 따라 차이가 다르게 나타나기 때문에 여기서는 연도별 보다는 2020년을 기준으로 남성, 여성의 연령 분포를 구하고 KLIPS 18년 표본과 경활조사를 비교하기로 한다.

[그림 3-7]과 [그림 3-8]은 2020년 남성과 여성의 KLIPS 18년 표본과 경활조사의 15세 이상 연령 분포를 보여주고 있다. 연령 분포의 형태를 보면 두 그래프 모두 가중치를 적용하였을 때가 그렇지 않았을 때보다 KLIPS 18년 표본과 경활조사 사이에 차이가 작아지는 것으로 나타나고

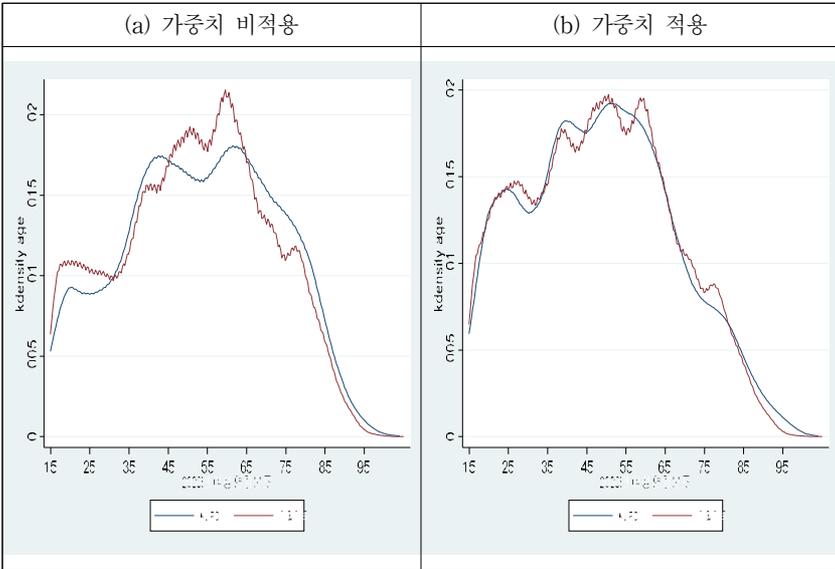
[그림 3-7] 2020년 KLIPS 18년 표본과 경활조사의 남성 연령 분포



주: KLIPS: 평균: 50.8, 표준편차: 18.5
경활조사: 평균: 49.7, 표준편차: 18.1

주: KLIPS: 평균: 50.0, 표준편차: 17.5
경활조사: 평균: 46.5, 표준편차: 17.3

[그림 3-8] 2020년 KLIPS 18년 표본과 경찰조사의 여성 연령 분포



주: KLIPS: 평균: 53.1, 표준편차: 19.3
경찰조사: 평균: 51.6, 표준편차: 18.9

주: KLIPS: 평균: 48.4, 표준편차: 18.7
경찰조사: 평균: 48.3, 표준편차: 18.5

있다. 따라서 가중치를 사용하는 것이 좀 더 대표성을 가진다고 말할 수 있을 것이다.

한편, 여성을 보면 가중치를 적용하였을 때 두 자료의 연령 분포가 유사할 뿐 아니라 평균 연령도 두 자료 모두 48세로 아주 근접하게 나타나고 있다. 따라서 앞서 남성에 비하여 여성의 경우 KLIPS 18년 표본의 고용률이 경찰조사보다 더 낮게 나타나는 것이 여성의 연령 분포가 경찰조사와 상이하기 때문에 발생하는 것으로 보기는 어려울 것이다. 그렇다면 KLIPS 18년 표본과 경찰조사 사이에 연령계층별 고용률의 차이 때문에 여성의 고용률이 경찰조사보다 낮게 나타난다고 볼 수 있는데 이를 확인하기 위하여 전체, 남성, 여성을 각각 연령을 3구간으로 하여 고용률을 구하였고 그 결과는 <표 3-21>에 제시하였다.

<표 3-21>을 보면 두 자료 사이에 고용률의 차이가 가장 크게 나타나는 연령층은 15~34세라는 것을 알 수 있다. 이는 앞선 KLIPS 98년, 09년 표본에서도 공통적으로 나타나는 현상이다. 다른 연령층에서도 KLIPS

18년 표본의 고용률이 경찰조사보다 작게 나타나기는 하지만 그 차이는 우려할 만한 수준은 아니다. 특히, 55세 이상에서는 두 자료의 고용률이 아주 근접하고 있다.

성별로 보면, 남성의 경우 35세 미만에서 KLIPS 18년 표본이 경찰조사보다 고용률이 매우 낮게 추정되는 반면, 다른 연령층에서는 약간 과대하게 추정되어 전체적으로는 이들이 서로 상쇄하는 경향을 보이고 있다. 한편, 여성의 경우에는 35세 미만뿐 아니라 35~54세, 55세 이상 모두에서 KLIPS의 고용률이 경찰조사에 비하여 과소하게 추정되는 것으로 나타나고 있지만 남성에 비하여 35세 미만에서의 차이는 크지 않게 나타나고 있

〈표 3-21〉 KLIPS 18년 표본과 경찰조사와의 연령별 고용률 비교

(단위: %, %p)

	전 체								
	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2018	43.0	51.4	-8.4	75.7	77.9	-2.2	49.4	49.4	0.0
2019	43.6	52.2	-8.6	75.6	77.7	-2.1	49.9	50.1	-0.2
2020	42.7	51.3	-8.6	75.0	76.3	-1.3	49.4	50.2	-0.8
	남 성								
	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2018	44.0	53.6	-9.6	91.8	91.0	0.8	63.3	62.0	1.3
2019	44.2	54.5	-10.3	92.2	90.3	1.9	63.2	61.7	1.5
2020	42.4	53.7	-11.3	91.5	89.1	2.4	63.4	61.6	1.8
	여 성								
	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2018	41.9	49.2	-7.3	59.2	64.4	-5.2	37.4	38.6	-1.2
2019	42.9	49.8	-6.9	58.7	64.7	-6.0	38.2	40.1	-1.9
2020	43.0	48.9	-5.9	57.8	63.1	-5.3	37.3	40.3	-3.0

주: 모든 수치는 가중치를 사용하여 구한 것임.

다. 결국, 앞서 경찰조사와 비교하여 KLIPS 18년 표본에서 남성에 비하여 여성의 취업자 수, 고용률이 더 낮게 나타나는 것은 어떤 특정 연령에서의 문제가 아니라 모든 연령에서 공통적으로 나타나는 현상이라고 볼 수 있다.

경찰조사와 비교하여 KLIPS 09년 표본에서 고용률이 낮게 나타나는 이유가 젊은 층에서의 임시일용직 취업자가 과소하게 추정되어서 나타나는 현상임을 확인하였다. KLIPS 18년 표본에서도 젊은 층뿐 아니라 여성에게서 전반적으로 고용률이 낮게 나타나는데, 이는 여성의 임시일용직 비중이 높기 때문으로 생각해 볼 수 있다. 따라서 왜 KLIPS 18년 표본의 고용률이 경찰조사에 비하여 낮은지를 확인하기 위해서는 성별로 구분하여 종사상 지위를 살펴볼 필요가 있다.

4. 종사상 지위

<표 3-22>는 종사상 지위로 구분하여 KLIPS 18년 표본과 경찰조사의 취업자 수를 성별로 보여주고 있다. 예상한 바와 같이 KLIPS 18년 표본의 여성 취업자 수는 경찰조사보다 상당히 작은 것으로 추정되고 있다. 2020년 현재 KLIPS 18년 표본의 여성 취업자 수는 경찰조사보다 약 100만명 적은 것으로 나타나고 있는데, 이를 종사상 지위로 나누어 보면 상용직에서 약 26만명, 임시일용직에서 약 61만명, 자영업주에서 약 17만명이 과소 추정되고 있으며 무급가족종사자만 약 3만 7천명이 과대 추정되고 있다.

한편, 남성의 경우에는 2020년 현재 KLIPS 18년 표본의 취업자 수가 경찰조사보다 약 5만 7천명 과대 추정되고 있는데 과대 추정의 원인은 상용직 때문으로 나타나고 있다. 상용직을 제외한 임시일용직과 비임금근로자는 과소 추정되고 있으며 특히, 임시일용직은 약 40만명이 적게 추정되고 있다.

결국, KLIPS 18년 표본에서 남성과 여성 고용률이 경찰조사보다 낮게 나타나는 주된 원인은 임시일용직 근로자가 과소하게 포함되어 있기 때문이라고 볼 수 있다. 특히, 이러한 현상은 여성에서 더 두드러지게 나타

〈표 3-22〉 KLIPS 18년 표본과 경찰조사와의 성별 종사상 지위에 따른 취업자 수 비교

(단위: 천 명)

		KLIPS					경찰조사					KLIPS-경찰조사				
		취업자	상용직	임시 일용	자영 업주	무급 가족	취업자	상용직	임시 일용	자영 업주	무급 가족	취업자	상용직	임시 일용	자영 업주	무급 가족
남 성	2018	15,379	8,921	2,324	3,946	188	15,372	8,347	2,879	4,007	139	7	574	-555	-61	49
	2019	15,495	9,134	2,302	3,909	150	15,463	8,536	2,819	3,965	143	32	598	-517	-56	7
	2020	15,438	9,174	2,236	3,869	155	15,381	8,667	2,637	3,920	157	57	507	-401	-51	-2
여 성	2018	10,534	5,448	2,588	1,507	991	11,450	5,425	3,433	1,631	961	-916	23	-845	-124	30
	2019	10,659	5,616	2,634	1,447	961	11,660	5,680	3,405	1,641	935	-1,001	-64	-771	-194	26
	2020	10,523	5,594	2,559	1,441	921	11,523	5,855	3,173	1,611	884	-1,000	-261	-614	-170	37

주: 자영업주는 고용주와 자영업자를 합한 것임. 모든 수치는 가중치를 사용하여 구한 것임. 차이는 KLIPS-경찰조사.

나고 있다. 한편, KLIPS 18년 표본의 남성 취업자는 경황과 큰 차이를 보이지 않지만 여전히 임시일용직의 수는 과소하게 추정되는 것은 주목할 만한 사항이다. 결국, KLIPS 09년 표본과 마찬가지로 18년 표본에서도 임시일용직의 확보가 KLIPS 18년 표본의 종사상 지위의 대표성을 확보하는 데 중요한 과제가 될 것이다.

5. 소 결

지금까지 KLIPS 18년 표본을 이용하여 경제활동참가율, 고용률, 실업률을 구하였고 이를 성별, 연령별, 종사상 지위별로 경황조사와 비교하였다. KLIPS 18년 표본은 KLIPS 09년 표본과 비교할 때 가구와 개인을 표본에 더 많이 추가하였지만 여전히 경황조사와 비교할 때 고용률, 경제활동참가율, 실업률이 과소 추정되고 있으며 특히 이러한 문제는 여성에서 좀 더 심하게 나타나고 있다.

KLIPS 18년 표본의 고용률과 경제활동참가율이 경황조사에 비하여 낮게 추정되는 이유는 KLIPS 18년 표본에서 15세 이상 인구가 과대하게 추정되는 반면, 취업자 수는 과소하게 추정되기 때문이다. 이러한 문제는 이전 KLIPS 코호트에서도 공통적으로 나타나고 있다. 한편, KLIPS 18년 표본을 남성과 여성을 구분하여 경황조사와 비교한 결과, 남성보다는 여성에서 경제활동참가율과 고용률의 차이가 큰 것으로 나타나고 있다. 구체적으로 남성의 경우는 15~34세만 경제활동참가율과 고용률이 과소 추정되고 있으나 여성의 경우는 전 연령층에서 이들 지표가 경황조사와 비교하여 낮게 나타나고 있다. 취업자를 종사상 지위별로 구분하여 KLIPS 18년 표본과 경황조사를 비교한 결과, 남성과 여성 모두에서 임시일용직 근로자가 과소 추정되고 있는 것으로 나타났다. 결국, KLIPS 18년 표본에서 취업자 수가 작게 나타나는 이유는 젊은 연령층에서 취업자 수가 작게 추정되기 때문이며, 임시일용직 취업자가 설문에서 과소하게 포착되기 때문이라고 볼 수 있다.

제5절 요약 및 개선방안

1. 요약

지금까지 KLIPS 98년, 09년, 18년 표본을 이용하여 경제활동참가율, 고용률, 실업률, 종사상 지위별 취업자 수 등을 성별, 연령별로 구하였고 이를 통계청의 경황조사와 비교하였다. 본 연구에서 제시하고 있는 KLIPS와 경황조사의 노동시장 지표, 또 이 두 자료 간의 비교는 KLIPS의 대표성을 판단하는 데 있어 필요한 자료가 될 뿐 아니라 우리나라 노동시장의 특징을 이해하는 데 있어서도 유용한 정보를 제공할 것으로 기대된다.

지금까지 KLIPS 자료와 경황조사를 비교 분석한 결과를 2020년 기준으로 정리한 내용은 <표 3-23>으로 요약될 수 있다.

<표 3-23> KLIPS와 경황조사 자료와의 차이(2020년 현재 기준)

	KLIPS 98년 표본	KLIPS 09년 표본	KLIPS 18년 표본
15세 이상 인구	-9%	+2%	+2%
취업자수	-15%	-2%	-4%
경제활동참가율	-8%	-5%	-6%
고용률	-6%	-4%	-5%
남녀차이 (경제활동참가율, 고용률)	남성이 여성보다 격차가 더 큼	여성이 남성보다 격차가 더 큼	여성이 남성보다 격차가 더 큼
연령 분포	55세 이상 비중이 과다함	연령 분포의 차이가 없음	연령 분포의 차이가 없음
15~34세 고용률	-5.7%	-8.4%	-8.6%
종사상 지위	임시일용직 과소	임시일용직 과소	임시일용직 과소
가중치	적용시 오히려 격차가 커짐	적용시 격차가 적어짐	적용시 격차가 적어짐

<표 3-23>은 지금까지 분석 중 가장 특징적인 점들만을 정리한 것이며 2020년을 기준으로 KLIPS와 경활조사를 비교하고 있다. 우선, 정량적인 지표를 기준으로 보면 경활조사와 가장 큰 차이를 보이는 코호트는 KLIPS 98년 표본인데 이는 어느 정도 예상할 수 있는 결과이다. KLIPS 98년 표본은 2020년에 이미 22년이 경과하였기 때문에 표본 손실과 고령화가 많이 진행되어 가중치를 적용함으로써 대표성을 확보하기에는 한계가 있는 것으로 보인다. 이는 55세 이상 비중이 경활조사와 비교하여 과다하게 포함되어 있고, 가중치를 적용할 경우 오히려 경활조사 지표와의 격차가 더 커지는 것에서도 알 수 있다.

KLIPS 98년 표본이 고령자를 많이 포함하고 있기 때문에 연령을 15~64세로 한정하여 경제활동참가율과 고용률을 구하고 이를 경활조사와 비교하여 보았다. 그 결과는 <부표 3-10>에 제시한다. 연령을 제한하지 않았을 경우 KLIPS 98년 표본의 경제활동참가율과 고용률은 경활조사에 비하여 각각 8%와 6% 작은 것으로 나타난 반면, 연령을 15~64세로 제한할 경우 경제활동참가율과 고용률의 차이는 4%와 3%로 줄어드는 것을 알 수 있다. 하지만 15세 이상 인구와 취업자 수 자체는 오히려 연령을 제한할 경우 경활조사와의 격차가 증가한다. 이는 KLIPS 98년 표본이 고령자를 과다하게 포함할 뿐 아니라, 본문 <표 3-7>에서 보듯이 젊은 층 자체도 과소하게 포함하고 있다는 것을 의미한다. 따라서 KLIPS 98년 표본의 젊은 연령층에 대한 가중치를 높여서 이들의 비중을 높이는 것을 고려할 필요가 있다.

한편, KLIPS 09년 표본과 KLIPS 18년 표본은 대표성에서 큰 차이는 보이지 않고 있다. KLIPS 09년 표본이 KLIPS 18년 표본보다 경제활동참가율, 고용률에서 좀 더 경활조사에 접근하고 있지만 그 차이는 그리 크지 않다. 또한, 성별 차이, 연령 분포, 종사상 지위, 가중치 적용의 효과 등에서도 그 결과가 아주 유사하게 나타나고 있다.

하지만 KLIPS 98년 표본보다는 좀 더 대표성을 확보하고 있지만 KLIPS 09년 표본과 KLIPS 18년 표본의 문제점이 없는 것은 아니다. 이들 표본은 경활조사와 비교하여 여전히 경제활동참가율, 고용률이 4~6% 정도 낮게 추정되고 있으며, 그 이유는 취업자 수 자체가 과소하게 포함

되어 있기 때문으로 보인다. 특히, 다른 연령층에 비하여 15~34세의 고용률이 현저하게 낮게 나타나는 현상을 보이고 있다. 따라서 청년층을 대상으로 하는 연구에서는 이 점을 분석에 감안할 필요가 있을 것이다. 또한, 취업자를 종사상 지위로 구별할 때 임시일용직의 수와 비중이 경활조사와 비교하여 현저히 낮게 나타나는 문제점이 있다. 하지만 KLIPS 09년 표본과 KLIPS 18년 표본의 연령 분포는 경활조사의 연령 분포와 거의 차이가 없고, 가중치를 적용할 경우 경활조사로부터 얻어진 노동시장 지표에 좀 더 근접하게 나타난다는 점은 긍정적이다.

2. 개선 방안

KLIPS 표본에 대한 개선점은 KLIPS 98년 표본에 대한 제안과 KLIPS 09년, 18년 표본에 대한 제안으로 구분할 수 있다. 우선, KLIPS 98년 표본의 경제활동참가율, 고용률, 실업률은 경활조사와 비교할 때 상당히 큰 차이를 보이며, 가중치를 사용하지 않았을 때보다 오히려 가중치를 사용하였을 때 그 격차가 더 커지는 것으로 나타났다. 또한, 55세 이상의 비중은 과다하고, 15~34세, 35~54세의 비중은 과소하게 나타나고 있다. 이미 진행되고 있는 KLIPS 98년 코호트에 젊은 층의 표본을 추가할 수 없는 상황에서 이러한 문제를 해결할 수 있는 방안은 결국 경활조사의 연령 분포에 맞게 KLIPS 98년 표본의 가중치를 조정하는 것이 필요할 것이다.

<표 3-24>는 현재 KLIPS 98년 표본과 경활조사의 2020년 연령별 가중치와 비중을 나타내고 있다. 표를 보면 두 자료 사이에 가장 큰 차이를 보이는 것은 65세 이상임을 알 수 있다. 따라서 KLIPS 98년 표본을 이용하여 전체 인구를 대상으로 분석할 경우에는 65세 이상 인구를 제외하는 것이 좀 더 대표성을 갖는 데 도움을 줄 것으로 사료된다.

하지만 65세 이상을 제외한다고 두 자료 사이에 차이가 해소되는 것은 아니다. 남성의 경우 35~54세 인구의 비중이 과소하게 나타나는데 이들의 고용률이 다른 연령층에 비하여 높다고 할 때 KLIPS 98년 표본에서 35~54세 인구 비중이 과소하게 나타나는 것은 KLIPS 98년 표본 전체의 고용률을 낮추게 하는 것으로 작용할 것이다. 따라서 이들 연령층에 대한

가중치를 높일 필요가 있겠다.

한편, 여성의 경우는 남성과 달리 15~24세, 35~44세의 비중이 경찰조사보다 낮게 나타나고 있다. 여성의 경우 이 연령대의 경제활동이 상대적으로 활발하다고 본다면 이 역시 여성의 고용률을 낮추는 방향으로 작용할 것이다. 또한, 여성의 15~24세 비중이 낮은 것은 KLIPS 표본에서 임시일용직의 비중이 낮게 나타나는 것과도 관련이 높다고 보여진다. 결국, <표 3-24>는 KLIPS 98년 표본이 경찰조사의 지표와 근접하려면 고령자의 가중치를 낮추고, 젊은 층의 가중치를 높여야 한다는 것을 보여주고 있다.

하지만 연령에 대한 가중치 조정만으로 KLIPS 98년 표본의 대표성을 확보하기는 어렵다. 이는 KLIPS 98년 표본에서 15~34세 연령층의 고용률 자체가 과소하게 나타나기 때문이다. 이는 이상호(2015)의 연구에서도 확인되고 있다. 이상호(2015)는 청년층의 고용률이 낮은 이유로 “청년층

<표 3-24> KLIPS 98년 표본과 경찰자료의 연령별 가중치 분포(2020년 기준)

	연령구간	KLIPS98 가중치	비중	경찰조사 가중치	비중	KLIPS/ 경찰
남성	15~24세	1,831,043	0.054	2,157,292	0.059	0.93
	25~34세	2,905,479	0.086	3,131,221	0.085	1.02
	35~44세	2,155,618	0.064	3,350,074	0.091	0.71
	45~54세	2,808,731	0.084	3,618,714	0.098	0.85
	55~64세	3,447,799	0.103	3,155,774	0.086	1.20
	65 이상	3,431,250	0.102	2,604,813	0.071	1.44
여성	15~24세	1,832,545	0.054	2,404,074	0.065	0.84
	25~34세	2,557,833	0.076	2,843,861	0.077	0.99
	35~44세	2,062,421	0.061	3,246,139	0.088	0.70
	45~54세	3,167,861	0.094	3,679,987	0.100	0.94
	55~64세	3,477,105	0.103	3,276,895	0.089	1.16
	65 이상	3,947,996	0.117	3,381,309	0.092	1.28

주: 연령계층별 가중치는 각 연령의 가중치를 한한 것임. 비중의 합은 1.
KLIPS/경찰은 비중을 사용하여 계산한 것임.

비재학자와 비동거자의 낮은 응답률 문제”와 “대학 재학생의 취업자 식별 문제”를 제시하고 있다. 결국, 15~34세 인구에 대해서 취업자를 정확히 파악하는 노력이 필요하다. 이를 위하여 1차 설문조사 후 잠정적으로 고용률을 계산하고 경황조사에 비하여 너무 낮게 나타날 경우에는 미취업자에 대한 후속 설문조사가 이루어지는 등 설문조사 체계를 보완할 필요가 있을 것이다.

다음으로 KLIPS 09년 표본과 18년 표본을 보면, KLIPS 98년 표본보다는 경황조사에 훨씬 근접하고 있지만 여전히 경제활동참가율과 고용률이 다소 낮게 나타나고 있다. 하지만 본문의 연령 분포 분석에서 보듯이 두 표본의 연령 분포는 경황조사와 크게 다르지 않다. 따라서 KLIPS 09년, 18년 표본의 연령계층에 따른 가중치 조정은 필요하지 않다고 판단된다. 또한, 가중치를 적용했을 때가 그렇지 않을 때보다 경황자료의 지표와 더 근접하기 때문에 가중치의 역할도 충분히 하고 있다고 볼 수 있다.

KLIPS 09년 표본과 18년 표본의 가장 큰 문제점은 성별의 분포나 연령의 분포가 왜곡되고 있는 것이 아니라 해당 계층에서 취업자 수 자체가 적게 포착되고 있다는 것이다. 특히, <표 3-23>에서 보듯이, 남성보다는 여성의 고용률이 낮게 나타나고, 15~34세의 고용률이 낮게 나타나며, 임시일용직의 수가 작게 포착되고 있어 설문조사 과정에서 이들 계층의 취업자 수를 좀 더 많이 확보하려는 노력이 필요할 것이다.

KLIPS 09년 표본과 18년 표본에서 여성의 고용률이 낮게 추정되고, 15~34세, 임시일용직의 취업자 수가 작게 나타나는 것은 결국 노동시장 정착도(labor market attachment)가 낮은 사람들을 KLIPS 조사과정에서 제대로 파악하지 못하기 때문으로 볼 수 있다. 여성의 경우는 출산, 육아 등으로 경력단절이 빈번하며, 젊은 층 역시 아르바이트나 임시직으로 일하는 사람들이 많기 때문에 이들 계층의 취업 여부를 파악하는 것은 상대적으로 어려울 수 있을 것이다. 통계청의 경황조사는 응답자의 부담 완화와 표본의 노후화를 막기 위해 일부 조사가구를 매월 교체하여 주기적으로 표본을 재설정하고 있다. 하지만 KLIPS의 경우 패널자료이기 때문에 이러한 표본교체가 쉽지 않다는 한계를 가진다. 하지만 취업자에 대한 설문부담을 덜어 줌으로써 취업자로 분류되는 것을 좀 더 용의하게 할 수는

있을 것이다. 호주의 HILDA는 자산 등 몇몇 변수에 대해서는 rotating base로 물으면서 설문부담을 덜어주고 있다. KLIPS도 일부 설문항목은 격년 등으로 물으면서 설문량을 줄일 수 있을 것이다. 또한, KLIPS 98년 표본에 대한 제언과 마찬가지로 초기 설문조사 후 잠정적으로 고용률을 계산하고 경황조사와 비교함으로써 취업자, 특히 임시일용직 취업자를 올바르게 파악하려는 노력이 필요할 것이다.

제 4 장

행정데이터와 서베이데이터의 연계

제1절 서론

사회과학 연구뿐만 아니라 다양한 분야에서 행정데이터를 활용한 연구가 꾸준히 증가하고 있다. 행정데이터를 활용한 연구는 서베이데이터로 불가능했던 연구를 실질적으로 가능케 한다. 또한, 정부의 정책 수립이 의견기반정책 중심에서 증거기반정책 중심으로 변화하면서 각 분야에서 다양한 정책의 근거를 마련하기 위해서 적극적으로 대처하고 있으며, 행정데이터와 서베이데이터의 연계는 이러한 변화와 대처에 대한 대응들 중 하나라고 할 수 있다. 물론, ICT 발전으로 정보의 축적 및 저장이 용이하여 과거와 비교하면 활용 가능한 행정데이터의 양이 증가하고 있다는 점, 회상적 질문, 응답자의 오류 등으로 인해 서베이데이터 내에 존재하는 측정오차 등도 이 같은 데이터 연계의 원인이라고 할 수 있다.

일반적으로 행정데이터는 한정된 정보를 포함하기 때문에 다른 행정데이터와의 연계가 필요하다. 행정데이터는 연구 목적으로 수집된 자료가 아니라 여러 분야에서 행정상의 관리 및 통제를 위하여 수집된 자료이다. 당연히 연구대상을 상정하고 다양한 연구자들의 니즈를 고려하여 수집된 서베이데이터와 비교할 때, 단독의 행정데이터는 포함된 정보가 한정적이다. 따라서 행정데이터는 다양한 출처의 행정데이터를 연계함으로써

써 포함정보의 범위를 확장할 필요가 있고 그렇게 함으로써 비로소 사회과학, 보건의료 등의 다양한 분야의 연구에 활용할 수 있다. 특히, Røed와 Raaum(2003)는 노르웨이 행정데이터를 연구했던 풍부한 경험을 바탕으로 행정데이터가 활용될 수 있는 연구 분야를 제시하면서 행정데이터를 활용함으로써 설명변수에서 외생적 변화를 찾을 수 있다고 언급하였다.

다양한 출처의 행정데이터의 연계에도 불구하고, 행정데이터는 서베이데이터가 포함하는 정보를 대체할 수 없는 경우가 드물지 않다. 예를 들면, 시간 사용, 영양 섭취, 가치판단 등에 관한 부분 등은 행정데이터로부터 수집하기 어려운 정보들이다. 그리고 정보 수집의 대상, 정보의 수집 방식, 정보 수집의 환경 등의 변화로 행정데이터의 연계가 예상치 못한 오류를 포함할 수도 있다. 따라서 행정데이터와 서베이데이터의 연계는 이러한 행정데이터가 가진 정보의 한계를 보완하고 연구자가 인지하지 못하는 행정데이터의 잠재적 오류를 탐지하는 것을 가능케 한다.

행정데이터와 서베이데이터의 연계에 대한 국내 연구는 상대적으로 부족하다. 일부 국내 연구는 해외 국가의 데이터 연계 시스템을 비교·분석함으로써 한국의 행정데이터 연계의 활성화를 위한 정책적 제언을 하거나(예를 들면, 유종성 외, 2020), 행정데이터의 연계가 활발히 이루어지고 있는 대표적인 국가를 중심으로 추진과정, 현황, 연계방법, 데이터 연계로 인한 연구성과 등을 조사·분석하였다(시승·토머스클먼, 2020; 백상숙, 2020; 로위 닐센, 2020; 한은희, 2020; 오미애, 2020; 신광영 2017; 윤광석·이건, 2017). 또한 일부 연구는 북유럽국가의 데이터 연계방식을 조사하면서 행정데이터와 서베이데이터의 연계 가능성만을 언급하였다. 본 연구와 같이 행정데이터와 서베이데이터의 연계에 초점을 두고 해외 국가의 사례연구를 진행하여 구체적 연계 성과를 제공하는 다른 국내 연구는 찾을 수가 없었다.

본 연구는 행정데이터와 서베이데이터를 연계·활용하는 해외사례를 조사함으로써 해외의 연계 성과를 바탕으로 국내에서의 행정데이터와 서베이데이터의 연계 가능성을 고려하고자 한다. 특히, 연계 가능성을 고려할 때, 연계 가능한 데이터의 존재와 매치 성공률을 고려한다. 동시에 행정데이터와 서베이데이터를 연계·분석한 연구사례를 조사함으로써 행

정데이터와 서베이데이터의 연계가 가져올 잠재적 가치를 파악한다.

본 연구는 연구의 범위를 다음과 같이 제한하고자 한다. 첫째, 해외에서 이루어진 서베이데이터와 행정데이터의 연계사례 연구는 북유럽국가, 호주, 미국 등으로 한정한다. 북유럽국가는 오래된 행정데이터 연계의 역사와 고유식별자를 이용하여 데이터를 관리·연계하는 고유의 특성 때문에 연구범위에 포함한다. 미국은 확률적 기록 연계의 대표적 국가로서 비교적 낮은 시작에도 데이터 연계를 통한 연구가 활발히 이루어지고 있어서 연구범위에 포함한다. 호주는 한국과 같이 데이터의 연계를 저해하는 환경(분산된 정보의 수집·관리 체계, 엄격한 개인정보보호 등)에서 최근 광범위한 데이터를 연계하기 위한 노력을 하고 있어서 벤치마크 대상으로 연구범위에 포함한다. 둘째, 개인 단위의 서베이데이터와 행정데이터의 연계에 연구의 초점을 둔다. 왜냐하면, 일반적으로 고용자(혹은 사업체) 데이터는 고용자의 고유식별자를 포함하고 있어서 행정데이터와 서베이데이터의 연계가 북유럽의 개인수준 데이터의 연계와 흡사하기 때문이다.²⁸⁾

본 장의 나머지는 다음과 같이 구성된다. 제2절에서는 데이터 연계와 관련된 국내 선행연구를 검토한다. 제3절에서는 북유럽국가, 미국, 호주 등을 중심으로 행정데이터의 연계방법을 알아보고 이를 바탕으로 이들 국가에서 행정데이터와 서베이데이터의 연계가 어떻게 이루어지는지를 분석한다. 제4절에서는 행정데이터와 서베이데이터를 연계한 연구사례를 조사함으로써 행정데이터와 서베이데이터의 연계가 가져오는 이점을 살펴본다. 제5절에서 국내의 대표적 서베이데이터인 한국노동패널과 행정데이터의 연계 가능성을 논하면서 본 연구를 마무리한다.

28) 북유럽국가의 개인수준 데이터와 일반적인 고용자 데이터는 고유식별자를 갖고 있어서 데이터 연계 측면에서 유사성을 가진다. 그러나, 데이터의 성질이나 데이터의 구조에서 개인 데이터와 고용자 데이터는 차이가 있을 수 있다. 고용자 데이터를 연계한 스웨덴 사례(Tomaskovic-Devey 외, 2015), 프랑스 사례(Abowd와 Kramarz, 2005), 미국 사례 (Abowd 외, 2018) 등을 참고한다.

제2절 선행연구

국내의 기존 연구에는 해외 각국의 행정데이터 연계를 조사한 다양한 연구가 다수 존재한다. 이들 연구는 국가별 행정데이터 연계를 소개하고 비교하거나(유종성 외, 2020), 특정 국가의 행정데이터 연계에 초점을 두고 행정데이터 연계의 발전과정, 관련 기관, 연계 방법, 연계 시 문제점 및 해결방안 등을 자세하게 조사하였다(시승·토머스 콜먼, 2020; 백상숙, 2020; 로위 닐센, 2020; 한은희, 2020; 오미애, 2020; 신광영 2017; 윤광석·이건 2017).

유종성 외(2020)는 덴마크, 스웨덴, 영국, 미국 등 해외의 구체적인 사례를 통해 사회과학 연구를 위해 행정데이터를 구축·활용할 수 있기까지의 과정, 그 과정에서 제기된 문제와 그에 대한 대처방안 등에 대해서 연구를 진행하였다. 그리고 이를 통해서 한국에서 행정데이터를 구축·활용하는 데 저해 요인을 제시하고 사회과학 연구 및 정책 제안에서 행정데이터를 적극적으로 활용하기 위한 몇 가지 방안을 제안하였다. 유종성 외(2020)는 행정데이터 연계방법보다는 행정데이터 연계의 활성화 방안의 도출을 위하여 관심 국가별 연계 현황과 문제점 및 시사점에 초점을 두었다.

국제사회보장리뷰 2020년 9월 제14호는 다음과 같이 행정 데이터 연계와 관련된 다수의 연구를 제공하였다.

시승·토머스콜먼(2020)은 급변화하는 데이터 환경을 언급하면서 서베이데이터와 행정데이터의 비교를 통해 행정데이터가 갖는 장점을 강조하는 동시에 행정데이터를 연구에 활용할 때 고려해야 하는 부분을 제시하였다. 특히, 미국의 구체적 사례를 통해 행정데이터의 잠재적인 문제점과 한계점 등을 기술하였고, 이후 미국의 행정데이터와 서베이데이터 연계 활용의 현황을 제공하였다.

백상숙(2020)은 호주의 행정데이터 연계활용 추진과정을 소개하고 보건·복지 행정데이터에 초점을 두고 연계 원칙과 연계 기관, 여러 부처

간의 정보 연계, 호주 보건복지연구소의 역할, 행정데이터 연계에 따른 프라이버시 보호, 2020년 처음 공개하는 “나의건강기록(My Health Record)” 등에 대하여 연구를 진행하였다.

로위 닐센(2020)은 노르웨이의 행정데이터 연계를 소개하고 그 방법과 과정을 설명하면서 노르웨이 행정데이터를 이용한 다양한 연구사례를 제공하였다.

한은희(2020)는 영국의 데이터 연계를 추진하는 관련 기관들의 보고서를 중심으로 행정데이터 연계기관의 설립 배경, 행정데이터 연계와 관련된 기관, 행정데이터 연계를 가능하게 하는 법률, 행정데이터의 이용 절차 및 연계방식, 그리고 연계된 행정데이터를 이용한 몇몇 연구사례 등을 조사하였다.

오미애(2020)는 보건·복지·사회보장 등을 중심으로 캐나다 내의 행정 데이터 활용을 조사하였다.

신광영(2017)은 스웨덴의 행정데이터 구축 및 연계 과정과 연계된 행정데이터를 활용한 사례를 연구하였다. 스웨덴의 행정데이터 구축의 역사, 행정데이터의 구조, 행정데이터를 연계·활용한 구체적 사례, 개인정보 보호를 위한 조치 등을 조사하였다.

윤광석·이건(2017)은 영국의 행정데이터 운영방식과 운영시스템을 통해서 한국 정부가 보유한 행정데이터를 활용할 수 있는 방안을 찾고자 하였다. 내외부 조직들 사이의 협력체계와 개인정보보호 등의 이슈에 초점을 두고 영국의 구체적 사례를 통해서 한국 행정데이터의 효율적 활용을 위한 대안을 제시하였다.

그 밖에 행정데이터의 연계에 대한 국내 연구보고서를 소개하면 다음과 같다.

강신욱 외(2018)는 행정데이터의 연계가 활발히 이루어지는 유럽국가들을 중심으로 행정데이터 연계를 연구하였다. 구체적으로 강신욱 외(2018)는 북유럽국가의 데이터 구축 및 연계에 대한 역사적 배경과 덴마크와 스웨덴의 추진과정 사례를 조사하였으며, 연계된 데이터를 이용한 연구성과를 제공하였다. 또한, 행정데이터 연계·활용의 후발주자인 영국에 대하여도 연계 시스템 수립과정, 주요활동 및 운영 프로세스, 시스템을 활

용한 성과 등을 제공하였다.

오미에 외(2016)는 보건복지 데이터포털과 관련하여 데이터베이스 구축과 포털 운영, 포털의 개편 내용 및 고도화 방안을 조사하면서, 보건복지데이터와 다른 출처의 데이터와의 연계 가능성을 언급하였다. 오미에 외(2016)는 2015년 센서스 데이터를 데이터 연계 사례로서 언급하면서 데이터 연계를 위한 이론적 근거를 제공하였지만, 연계된 데이터와 연계 결과에 대한 구체적인 내용은 제공하지 않았다.

종합하자면, 국내의 기존 연구는 대개 해외의 행정데이터 연계 과정, 연계 현황, 연계 시 문제점 및 해소방안 등을 조사함으로써 한국의 행정데이터 연계활용을 활성화하기 위한 근거자료나 활성화 방안을 제시하고자 하였다. 조사 대상국가는 덴마크, 스웨덴 등의 북유럽국가뿐만 아니라 영국, 미국, 호주, 캐나다 등 다양한 국가를 포함하였다. 대부분의 연구는 행정데이터의 연계 활용에 초점을 두고 진행되었다.

제3절 행정데이터와 서베이데이터 연계의 해외사례

본 절에서는 북유럽국가, 미국, 호주 등에서 수행되고 있는 데이터 연계를 조사함으로써 일반적인 데이터 연계 시스템과 그 안에서의 행정데이터 및 서베이데이터의 연계 현황과 연계 가능성에 대하여 알아보고자 한다.

1. 덴마크

덴마크의 주민등록은 1924년에 처음 시작되어 각 가구의 구성원들에 대한 정보가 색인카드에 수기로 작성되어 관리되었다. 이후 1968년에 전산화와 함께 덴마크의 주민등록은 고유의 식별번호인 CPR 번호(CPR-number)가 부여되어 관리되었다(Pedersen, 2011). Thygesen 외(2011)에 따르면, 덴마크 영주권을 가진 모든 개인은 고유의 식별번호인 CPR 번호

(CPR-number)가 부여되며 이를 이용하여 개인수준에서 하나의 레지스터, 복수의 레지스터, 혹은 다른 출처의 데이터를 레지스터 기반 정보와 연계할 수 있다. 각 덴마크 사업체도 고유의 식별번호를 부여받게 되는데, 이를 통해서 고용자와 피고용인 사이의 연결이 가능하게 된다. 개인 데이터 가공에 대한 법률(Act on Processing of Personal Data)로 인해, 모든 데이터는 덴마크 통계청 내에 저장되어 있고 특정 권한이 부여된 환경에서 연구자는 온라인 접근을 통해 데이터를 연구하고 분석할 수 있다. 특히, 덴마크 행정데이터의 사회보장번호는 조사기관 및 통계기관에만 접근이 허용되며, 연구자는 사회보장번호가 아닌 개별 식별번호에 접근이 가능하다(Bingley와 Martinello, 2014).

덴마크의 행정데이터와 서베이데이터 연계활용 방식은 직관적이고 이해하기 쉬우며, 또한 데이터에 포함된 정보가 다양하다. 이는 덴마크는 덴마크 통계청이 개인의 CPR 번호와 사업체의 고유번호를 이용하여 행정데이터와 서베이데이터를 수집·관리하기 때문이다. Petersson 외(2011)는 노동시장 가입(labour market affiliation)에 대한 정보를 포함하는 덴마크 행정데이터와 서베이데이터를 선택하여 보여주고, 이들의 연계로 인해 획득할 수 있는 정보, 유효성과 포함범위 등을 제공하였다. 주민등록 시스템(Danish Civil Registration System)을 갖추기까지의 과정, 주민등록 대상, CPR 번호 생성, 등록시스템에 포함된 정보 등에 관한 상세정보는 Pedersen(2011)을 참고한다.

특히, 덴마크는 다양한 행정정보가 중앙에서 수집·관리되기 때문에 데이터 연계의 범위가 개인이나 사업체 관련 정보를 연계하는 수준을 넘어선다. 예를 들면, 동거인의 경우에는 건물 식별번호를 통해서 동거인 식별번호가 부여되고, 이 동거인 식별번호를 이용하여 동거인의 행정데이터와의 연계가 가능하다(Bingley와 Martinello, 2014).

2. 스웨덴, 노르웨이, 핀란드 등

스웨덴, 노르웨이, 핀란드 등도 덴마크와 같이 개인 고유 식별번호를 부여하여 데이터를 수집·관리하므로 데이터 연계 활용이 덴마크와 유사

하다. 마찬가지로 스웨덴, 노르웨이, 핀란드 등에서도 덴마크와 같이 데이터의 연계 활용이 활발히 이루어지고 있다.

가. 스웨덴

스웨덴의 전 중앙국세청(National Tax Board)은 1947년 이후로 스웨덴에서 영주(residence on a permanent basis)한 개인들에게 개인 식별번호(Personal Identification Number, 이하 PIN)를 부여·유지해 왔고 보건·의료 등을 포함한 광범위한 스웨덴의 행정데이터가 PIN을 기반으로 하여 수집·관리되고 있다. 스웨덴의 PIN은 10자리로 구성되며 한국의 주민등록번호처럼 생년월일 정보를 포함하고 있다. 스웨덴의 인구등록 전산화는 1967년에 이루어졌다.

스웨덴에는 두 개의 인구등록부가 존재하는데, 하나는 스웨덴 국세청에서 관리하는 인구등록(Population Register, Folkbokföringsregistret)이고 다른 하나는 스웨덴 통계청에서 관리하는 총인구등록부(Total Population Register, Registret över totalbefolkningen)이다(Ludvigsson 외, 2016). 각 PIN과 관련된 정보의 변화는 스웨덴 국세청을 중심으로 공유가 이루어지고 이러한 정보는 스웨덴 통계청에 저장되어 관리된다(Ludvigsson 외, 2009).

특히, Ludvigsson 외(2009)는 연구자가 특정한 질병을 경험한 환자의 행정데이터를 요청하는 과정과 특정한 질병에 대한 데이터를 개인적으로 구축한 연구자가 행정데이터 연계를 요청하는 과정 등을 언급하였다. 특히, 전자의 경우 연구자가 만성소화장애가 있는 환자의 정보에 접근하는 과정을 가정하여 다음과 같이 설명하였다. 우선, 연구자는 국제질병분류(International Classification of Diseases, 이하 ICD)의 만성소화장애에 해당하는 코드를 가진 환자를 식별하여 데이터를 제공할 것을 보건복지국가위원회(National Board of Health and Welfare, 이하 NBHW)에 요청한다. NBHW는 해당 질병이 있는 환자의 PIN을 스웨덴 통계청으로 보내고 스웨덴 통계청은 나이와 연령이 매치되는지 확인한 후 각 환자들에게 고유 일련번호를 부여한다. 끝으로 연구자는 PIN을 제거한 고유 일련번호

호를 가진 데이터를 수령하는데, 이 고유 일련번호는 PIN에 대응하지만 PIN의 정보를 포함하고 있지 않기 때문에 고유 일련번호를 이용하여 PIN을 복원할 수는 없다. 반면, 후자의 경우, 연구자 개인이 수집한 정보를 스웨덴의 행정데이터에 연계하는 가능성에 대해서도 언급하였다. 이때, 성별, 나이, 카운티, 시민권 여부, 사회경제적 위치, 이주 및 사망 등의 정보를 이용하여 매치할 수 있다고 했다. 행정데이터 연계 과정과 달리, 구체적인 연계 과정을 제공하지 않았다.

나. 노르웨이

노르웨이의 개인 식별번호 시스템 시작에는 다양한 기관과 단체가 관제되어 있었다. 그중 재계가 하나의 주요한 견인차였고, 노르웨이 국세청(Norwegian Tax Administration), 노르웨이 통계청(Statistics Norway) 등이 중요한 참가자였다(Furseth, 2014). 노르웨이에서는 영주권을 가진 사람들에게는 생년월일을 포함한 11자리의 PIN을 부여하고, 세금납부와 의료보험의 대상이지만 영주권자가 아닌 이들에게는 D번호(D-number)를 부여한다. 이렇게 부여된 번호를 이용하여 출생, 혼인, 출산, 사망, 교육, 세금, 의료 등에 관한 정보가 중앙인구등록부(Central Population Register)에서 계속 갱신된다. Furseth(2014)는 연구 당시 노르웨이 통계청이 27개의 공공단체로부터 100개 이상의 다른 행정 데이터를 제공받고 있다고 하였다.

노르웨이는 행정데이터의 수집·관리가 고유 개인 식별번호에 기반하므로 다른 기관들로부터 정보를 수집·취합하고 기존의 데이터를 업데이트하는 것이 비교적 용이하다. 한번 부여된 개인 식별번호는 사라지지 않으며, 노르웨이에 다시 이주해 오더라도 같은 개인 식별번호를 이용하게 된다(Nielsen 외, 2014).

Furseth(2014)는 노르웨이 통계청에서 면접 기반 서베이를 진행할 때 다른 데이터에서 정보를 가져옴으로써 면접시간을 절약할 수 있다고 하면서, 취득·수료한 교육과정에 대한 정보를 그 사례로 언급하였다.

다. 핀란드

핀란드의 개인 식별코드는 1963년에 시작되었다. 핀란드도 다른 북유럽 국가와 마찬가지로 PIN을 기반으로 행정데이터를 수집·관리한다. 또한, 고유한 부동산 식별자(real estate identifier), 거주지 코드(domicile code), 사업체 식별 코드(business identification code)를 각각 부동산, 거주지, 사업체에 부여함으로써, 동거인 식별, 거주지 이동 확인, 고용, 실업 등의 상태 변화 등을 각 코드를 연계함으로써 확인될 수 있다. 단순한 양적 측면에서 볼 때, 핀란드 통계청 통계의 96%가 행정데이터로부터 나온다(Statistics Finland, 2004).

핀란드의 기본등록시스템(basic register system)은 건물과 거주지 정보를 포함한 인구 정보 시스템, 부동산 정보 시스템, 사업체 정보 시스템 등을 포함한다.²⁹⁾ 이외의 주요 등록시스템은 과세 기록, 고용 기록, 연금 기록, 취업지원자 기록 등이 있다. 또한, 핀란드 통계청은 통계적 목적으로 최종 교육 및 학위 통계(Register of Completed Education and Degree), 사업체 등록부(Business Register), 건물과 거주지 등록부(Register of Buildings and Dwellings), 학생 등록부(Student Register) 등의 데이터를 자체적으로 수집한다.

유효한 PIN이 있는 경우 데이터를 연계하는 것은 직관적이다. 하지만, 핀란드의 경우 유효한 PIN이 없는 경우가 발생할 수 있는데, 이때 거주지 코드나 생년월일을 이용하여 유효한 PIN을 찾아야 한다(Törmälehto, 2008).

3. 미국

미국 사회보장번호(Social Security Number)는 1936년 사회보장국(Social Security Administration)이 사회보장연금 목적으로 소득을 추적하기 위하여 만들었고, 이후 사회보장 관리 이외의 목적으로 널리 사용되

29) 핀란드 통계청(2004)은 핀란드의 기본등록시스템이 세계에서 유일하다고 기술하였다.

고 있다(Puckett, 2009). 실제로 사회보장번호는 미국 인구조사국(US Census Bureau)의 데이터 연계에 있어서 중요한 역할을 해왔다. 하지만, 응답자들이 신원 도용 위험을 인식하기 시작하면서 사회보장번호 제공 의향이 감소해 왔고, 미국 인구조사국은 확률적 기록연계(Probabilistic Record Linkage) 방법으로 데이터 연계방식을 전환하게 된다(McNabb 외, 2009). McNabb 외에 따르면, Survey of Income and Program Participation (이하 SIPP)의 경우, 1996년 응답자의 12%가 사회보장번호 제공을 거부하였으나, 2004년에는 응답자의 35%가 사회보장번호 제공을 거부하였다. 마찬가지로, Current Population Survey(이하 CPS)의 경우 1994년 응답자의 10%가 사회보장번호 제공을 거부하였으나, 2003년에는 응답자의 23%가 사회보장번호 제공을 거부하였다. 이에 인구조사국은 직접적인 질문을 그만두고 행정데이터 연계 거부에 대한 문항을 포함하여 확률적 기록연계 방법으로 데이터 연계를 도모하였다.

Wagner와 Layne(2014)은 미국 인구조사국의 개인 식별 검증체계(Person Identification Verification System, 이하 PVS)를 이용하여 이루어지는 여러 데이터의 연계에 대하여, 이론적 근거를 포함하여 PVS를 이용한 연계의 과정을 구체적으로 설명하였다. Wagner와 Layne에 따르면, PVS는 확률적 기록연계를 적용, 레퍼런스 파일에 매치된 개인에 보호된 식별키(protected identification key, 이하 PIK)를 부여함으로써 다양한 소스의 데이터와 연계를 가능하게 한다(기록연계 방법론은 Herzog 외(2007)와 Fellagi와 Sunter(1969)를 참고한다). Wagner와 Layne은 상거래 데이터, 2011년 연방데이터, 2010년 센서스데이터 등과의 연계 성공률을 보고하였으나, 서베이데이터에 대한 연계 성공률은 보고하지 않았다.

<표 4-1>은 Wagner와 Layne(2014)에서 보여준 PVS를 이용한 2개의 상거래 기록 파일, 2011년 연방데이터 파일, 2010년 센서스데이터를 미국 인구조사국의 마스터 파일에 연계한 결과이다. 패널 1의 기록 수는 각 데이터 파일이 포함하는 모든 기록의 수를 나타낸다. 패널 2는 검증과정(verification process)에 의한 연계 결과를 보여준다. 연방데이터 파일의 경우 검증과정에 의한 연계 시 이미 높은 매치 성공률을 보이는데, 이는 검증과정이 사회보장번호를 통해서 매치를 찾고 매치된 기록들 사이의

〈표 4-1〉 PVS를 이용한 연계 결과

	상거래 파일 2개의 평균	2011년 연방 데이터 파일	2010년 센서스 데이터
패널 1. 기록#	339,295,722	53,181,072	312,471,327
패널 2. Verification 프로세스			
입력 기록#	257,979,068	53,118,553	0
매치된 기록#	195,571,749	53,058,202	0
매치 안된 기록#	62,407,320	60,351	0
매치 성공(%)	75.81	99.89	0
패널 3. Geo Search 프로세스			
입력 기록#	143,701,483	122,870	302,103,352
매치된 기록#	41,847,479	68,525	259,873,717
매치 안된 기록#	101,854,004	54,345	42,229,635
매치 성공(%)	29.12	55.77%	86.02
패널 4. Name Search 프로세스			
입력 기록#	101,854,004	54,345	42,229,635
매치된 기록#	5,328,014	13,158	19,960,452
매치 안된 기록#	96,525,990	41,187	22,269,183
매치 성공(%)	5.23	24.21	47.27
패널 5. DOB Search			
입력 기록#	96,525,990	41,187	22,269,183
매치된 기록#	73,802	587	304,690
매치 안된 기록#	96,452,188	40,600	21,964,493
매치 성공(%)	0.08	1.43	1.37
패널 6. HHComp Search 프로세스			
입력 기록#	0	0	21,964,493
매치된 기록#	0	0	1,975,295
매치 안된 기록#	0	0	19,989,198
매치 성공(%)	0	0	8.99
패널 7. PVS 연계 결과			
PVS가능 기록#	339,273,232	53,181,072	302,103,352
매치된 기록#	242,821,044	53,140,472	282,114,154
매치 안된 기록#	96,452,188	40,600	19,989,198
매치 성공(%)	71.57	99.92	93.38

주: 1) #는 수를 나타냄.

2) 센서스 데이터는 사회보장번호(Social Security Number, SSN)가 없으므로 Verification에 대한 값이 모두 0이며, 센서스 데이터만 HHComp Search를 진행하므로 상거래 및 연방 파일에 대한 값은 모두 0임. Verification 모듈은 SSN, 성, 이름, Geo Search는 우편번호 처음 3자리, Name Search는 성, 이름, 가운데 이름, 세대 접미사, 생년월일, 성별, DOB Search는 성, 이름, 생년월일, HHComp Search는 동일 가구 내 PIK를 이용함.

자료: Wagner와 Layne(2014)의 Table 2. Sample PVS Results를 각색.

이름과 생년월일이 일치하면 매치가 성공하는 방식이기 때문이다. 상거래 파일 2개는 입력 기록의 수도 연방 데이터 파일에 비해 적다. 그리고 검증과정에 의한 매치 성공률은 약 76% 수준으로 높지 않다. 2010년 센서스데이터 파일의 경우 사회보장번호가 없어서 검증과정에 입력할 수 없으므로 모든 셀에서 0의 값을 갖는다.

패널 3, 4, 5는 Geo Search(지리정보 검색) 프로세스, Name Search(이름 검색) 프로세스, DOB Search(출생일 검색) 프로세스에 의한 연계 결과를 보여준다. 검증과정에서 매치되지 않았거나 입력되지 못한 기록들을 Geo Search, Name Search, DOB Search의 순서대로 프로세스에 입력하여 매치되는 기록을 찾는다. 상거래데이터 파일과 연방데이터 파일의 경우 처음 4개의 모듈로 충분한 매치 성공률을 달성하였기 때문에 HHComp Search(Household Composition Search, 가구구성 검색) 프로세스에 입력하지 않는다. 2010년 센서스데이터 파일은 HHComp Search 프로세스에 입력하여 추가적으로 매치되는 기록을 찾는다. 패널 7은 PVS에 의한 매치의 결과를 보여주는데, PVS에 의한 확률적 기록연계가 데이터 종류에 따라 약 72~99%까지의 매치 성공률을 보여준다. 이는 데이터의 종류에 따라 다소 차이는 있지만 이름과 주소와 같은 개인정보를 이용하여 동일한 인물의 데이터를 연계하는 것이 충분히 가능하다는 것을 보여준다.

한편, McGonagle 외(2011)는 세계에서 가장 긴 경시적 데이터인 Panel Study of Income Dynamics(이하 PSID)에 대하여 표본설계와 추적원칙, 응답자 및 표본크기, 주요조사와 추가조사, 데이터의 배포 및 문서화, 윤리문제, 펀딩, 운영, 그리고 인생과정(life course) 연구에 있어서 PSID가 갖는 잠재성 등을 조사하였다. McGonagle 외는 외부 행정데이터 연계가 확대되고 있다고 덧붙이면서, 다음 행정데이터가 PSID에 연계되고 있다고 하였다. 국립보건통계센터 국민사망지수(National Center for Health Statistics National Death Index)로부터의 1979년 이후의 사망일 및 사망원인 데이터, 보건부 산하의 메디케어와 메디케이드 센터(Centers for Medicare and Medicaid Services)로부터 1991~2008년의 의료비 청구 데이터, 주택도시개발부(Department of Housing and Urban Development)

로부터 주거보조비 수령 여부에 대한 데이터, 국립교육통계센터(National Center for Education Statistics)로부터 응답자가 다닌 초중등 국공립학교에 대한 데이터 등이다.

〈표 4-2〉 PSID 2011 주요 인터뷰 내용

	첫수집 연도
◆ 경제 관련 항목	
· 주거 환경 특성, 유틸리티	1968
· 고용	1968
· 소득	1968
· 정부이전프로그램(Government Transfer Program) 참여	1968
· 소비지출	1984
· 부와 저축(Wealth and active savings)	1968
· 연금(Pensions)	1976
· 기부 및 자선(Philanthropic giving and volunteering)	2001
· 가사노동 및 육아(Housework and child care)	1968
◆ 건강 관련 항목	
· 일과 활동의 제한(Work and activity limitations)	1968
· 건강보험	1968
· 흡연과 음주	1968
· 산재보상(Workers' compensation)	1977
· 병원 및 간호(Hospital and nursing care)	1981
· 사회보장 장애 보험(Social security disability insurance)	1984
· 일반 건강 상태(General health status)	1984
· 운동(Exercise)	1986
· 키와 몸무게(Height and weight)	1986
· 하루 중 활동(Activities/ Instrumental activities)	1992
· 건강 상태(Health conditions)	1999
· 정신적 고통(Emotional distress)	2001
· 전반적 삶의 만족도(Overall life satisfaction)	2009
· 건강 지출(Health expenditures)	1999
· 아동기 전반적 건강상태	1999

〈표 4-2〉 계속

	첫수집 연도
· 아동기 건강문제	2007
· 약물치료 사용(Medication use)	2011
◆ 인구사회 관련 항목	
· 가족구성 및 거주지 변화	1968
· 사망사건	1968
· 혼인상태 및 출산 기록	1968
· 가족구성원의 교육수준	1968
· 가구주 및 배우자의 종사 직업 및 산업	1968
◆ 계약 하에서 접근 가능한 항목	
· 지리 정보	1968
· 차량 모델	1999
· 사망원인 및 사망일(국립보건통계센터)	1979
· 메디케어 청구(메디케어와 메디케이드 센터)	1991
· 초중등학교 특성 정보(국립교육통계센터)	1968
· 주거보조비(주택도시개발부)	1968

자료: McGonagle 외(2011)의 표 3. PSID 2011 주요 인터뷰 내용을 각색.

〈표 4-2〉는 2011년 PSID의 메인 인터뷰 데이터가 포함하는 정보의 종류와 정보의 종류별로 처음 수집된 연도를 보여준다. “계약하에서 접근 가능한 항목”에서는 비밀보호를 위하여 계약에 의해서만 제공될 수 있는 정보의 종류를 보여준다. 특히, 해당 항목 중에는 행정데이터로부터 연계된 정보가 포함되어 있다. 사망원인이나 사망일, 메디케어 청구내역, 초중등학교 특성정보, 주거보조비 등 모두 행정데이터의 연계가 없었다면 신뢰할 만한 정보를 구하기 어렵거나 정보수집 비용이 상당할 것으로 예상되는 정보들이다. 이는 행정데이터와 서베이데이터를 연계함으로써 가용한 정보범위와 정보수집 비용을 획기적으로 개선할 수 있다는 것을 보여준다.

Mulrow 외(2011)는 다양한 서베이데이터를 직접 연계함으로써 행정데이터와 서베이데이터 연계 시 미국 인구조사국의 PVS 이용 가능성을 평가하였다. 〈표 4-3〉은 미국 인구조사국의 PVS를 이용하여 행정데이터

〈표 4-3〉 미국 인구조사국의 PVS프로젝트 매치 성공률

인커밍 데이터 (incoming data)	Verification	Geo Search	Name Search	최종 매치
서베이 데이터				
ACS 2001	N/A	86.30	58.12	93.49
ACS 2002	N/A	86.27	57.57	93.12
ACS 2003	N/A	87.05	54.15	92.39
ACS 2004	N/A	88.16	53.63	92.60
ACS 2005	N/A	89.93	44.77	92.90
ACS 2006	N/A	87.87	47.53	92.03
ACS 2007	N/A	89.06	41.76	91.65
ACS 2008	N/A	88.08	46.07	91.71
ACS 2009	N/A	84.02	52.23	90.82
SIPP 2001	93.74	69.57	33.19	93.06
CPS 2001	94.07	82.20	32.28	76.53
센서스 데이터				
201년 센서스	N/A	83.04	57.57	91.14

자료: Mulrow 외(2011)의 Table 1. Match Percentages for Census Bureau PVS Projects 를 각색.

와 서베이데이터의 매치 성공률을 보여준다. 매치 성공률은 각 모듈에서 매치된 수를 모듈에 입력된 기록의 수로 나누어 계산한 값이다. 이를 위하여 2001~2009년 미국지역사회조사(American Community Survey, 이하 ACS), SIPP, CPS를 미국 인구조사국의 마스터 파일에 연계하였다. ACS의 매치 성공률은 조사연도에 따른 차이가 존재하지만 대체로 약 91~93% 수준을 유지한다. 2001년 SIPP 서베이데이터는 약 93%의 매치 성공률을 보여준다. 2001년 CPS 서베이데이터의 매치 성공률은 약 77%로 다른 서베이데이터에 비교해 많이 낮은 수준이다.³⁰⁾ 마지막 행은 2010

30) Wagner와 Layne(2014)에서 보여준 2개 상거래 행정 데이터의 평균 매치 성공률은 약 72%로 Mulrow 외(2011)의 2001년 CPS 서베이 데이터의 매치 성공률보다 낮았다.

년 센서스 데이터 연계 시 매치 성공률로 약 91%의 매치 성공률을 보여 준다.³¹⁾

행정데이터와 서베이데이터의 연계·활용에 그치지 않고, 연계한 행정·서베이데이터를 공개데이터로 만든 경우도 존재한다. Abowd 외(2006)는 미국의 인구조사국, 사회보장국, 국세청, 의회예산국 사이의 협력으로 완성한 SIPP/SSA/IRS PUF (Public Use File, 공공 용도 파일)에 대하여 연구목적과 배경, 성과 등에 대하여 보고하였다. SIPP/SSA/IRS PUF는 서베이데이터인 SIPP에 다양한 출처의 행정데이터를 연계하여 개인정보보호 조치를 취한 공개데이터라는 점이 주목할 만하다.

4. 호주

호주의 기록연계에 대한 역사는 오래되어 1970년에 Hobbs 교수가 기록연계의 잠재적 가치를 거론한 바 있다(Hobbs, 1970). 특히, 서호주의 데이터 연계 계획은 1970년 후반에 이미 시작되었으나 특정 목적에 한정된 수준이었다. 이후 1995년 서호주에서 서호주대학(University of Western Australia)과 서호주보건국(Western Australia Department of Health)이 협력하여 데이터 연계 목적의 최초 데이터통합 인프라스트럭처인 서호주 데이터통합 시스템(Western Australia Data Linkage System, 이하 WADLS)을 출범하였다(Holman 외, 2008). 1995년부터 2003년까지 258개의 프로젝트가 승인되어 170여 개의 학술지 논문이 게재되는 등 상당한 성과를 얻었으며, 특히 일부 프로젝트는 보건 정책과 의료 관행에 실질적 변화를 가져왔다(Brook, Rosman, Holman, 2008). WADLS의 성공 이후, 2006년에는 호주 수도 특별지역(ACT)과 뉴사우스웨일즈의 데이터 연계를 수행하기 위한 건강기록연계센터(Centre for Health Record Linkage, CheHeReL)가 뉴사우스웨일즈(NSW) 주에 설립되었다. 나아가, 같은 해에 연방정부, 주정부, 연구기관 등이 상당한 투자를 통해 인구건강 연구

31) 비슷한 수준의 매치 성공률을 다른 연구에서 발견할 수 있었다. Meyer와 Mittag(2019)에서 행정 데이터들은 98% 이상 PIK에 매치가 되었으며, 반면 CPS 데이터는 가구의 91%가 적어도 한 명의 PIK가 매치된 가구원을 갖고 있었다.

네트워크(Population Health Research Network, 이하 PHRN)라 불리는 데이터 연계 인프라스트럭처를 구축하였다(Boyd 외, 2012).

현재 호주는 주별로 데이터 연계 유닛을 갖추고 각 데이터 연계 유닛이 해당 주의 데이터 연계를 담당한다. PHRN의 경우, 관할구역 간의 데이터 연계는 관련 데이터 연계 유닛과 호주 보건복지연구소(Australian Institute of Health and Welfare, AIHW)가 협업한다. 호주가 이러한 접근법을 취해온 데에는 다양한 요인들이 있다. 그중, 호주의 데이터 연계가 서호주와 뉴사우스웨일즈의 성공적인 데이터 연계에 힘입어 추진되었다는 점, 호주가 6개의 주와 2개의 자치 특별지역으로 구성되어 있다는 점, 그리고 연방정부와 주정부가 모두 건강정보 수집에 공동책임이 있다는 점 등을 생각해 볼 수 있다(Flack과 Smith, 2019).

호주는 고위험의 민감한 정보를 연계하는 연구프로젝트에 대하여 당국의 승인받은 기관이 관리하게 하고 있다. 2012년 호주통계청(Australian Bureau of Statistics, ABS)과 호주 보건복지연구소가 데이터통합당국(Accredited Integrating Authorities)으로 승인받은 이후, 호주 가족연구소(The Australian Institute of Family Studies, AIFS) 등 2019년 기준으로 7개 기관이 데이터통합당국으로 승인받았다.³²⁾

호주는 북유럽국가의 등록번호나 미국의 사회보장번호와 같은 개인 식별번호가 존재하지 않는다. 여권번호나 운전면허번호가 존재하지만, 신청을 통해 발급받지 않는 경우, 이 번호들은 존재하지 않는다. 지역 서비스 프로그램 운영 시 수집되는 정보들은 고유식별자를 이용하여 관리되는 경우를 발견할 수 있으나, 이러한 식별자가 다른 행정데이터의 수집에서 사용되는 경우는 드물다. 대개 데이터 연계는 이름, 출생연월일, 주소 등의 정보를 이용하여 확률적 기록연계 방법으로 이루어지지만, 확률적 연계와 결정적 연계를 같이 사용하기도 한다(Flack과 Smith, 2019).

호주 통계청(Australian Bureau of Statistics, 2013)은 2011년 센서스데이터와 학생등록 행정데이터를 다양한 방법으로 연계, 각 연계데이터의 품질을 평가하였다. 주별로 차이가 있지만, 이름과 주소만을 이용한 확률적 연계로 약 80~90% 수준의 매치 성공률을 보이며 센서스와 행정데이

32) <https://toolkit.data.gov.au/>

터의 연계가 성공적인 것으로 판단하였다. Wagner와 Layne(2014)이 제공한 미국 인구조사국의 매치 성공률과 비교 가능하지만, Wagner와 Layne은 특정 행정데이터가 아니라 미국 인구조사국의 마스터 파일에 연계했다는 점을 유념해야 한다.

행정데이터와 서베이데이터의 연계에 대한 고려도 있었다. Soloff 외(2007)는 다양한 행정데이터와 호주아동 경시적연구(Longitudinal Study of Australian Children, 이하 LSAC)의 연계 가능성을 비용, 접근성, 정보보호, 데이터의 품질, 정보종류 등의 측면에서 검토하였다.³³⁾ Soloff 외에 따르면, LSAC가 현실적으로 연계 가능한 행정데이터는 보건, 교육, 지리정보, 보육 등이다. Genici와 Nguyen(2013)은 Soloff 외(2007)와 유사한 연구를 호주청년 경시적조사(Longitudinal Survey of Australian Youth, LSAY)에 대하여 진행하면서 LSAC를 행정데이터와 서베이데이터 연계의 성공적 모형으로 소개하였다. 예를 들면, Daraganova 외(2013)는 LSAC와 국민평가프로그램의 문해 능숙도와 산술능력(National Assessment Program-Literacy and Numeracy, 이하 NAPLAN) 데이터를 성공적으로 연계, LSAC에서 수집하지 않는 학업성취의 직접적인 측정치를 연계·활용할 수 있음을 보였다.³⁴⁾ 또한, 연계데이터와 행정데이터를 비교, LSAC와 NAPLAN 연계데이터가 모집단을 대표하지 않는다는 점을 발견하였는데, 행정데이터와 서베이데이터 연계의 잠재적 가치를 강조하는 동시에 연계데이터의 대표성 문제 등은 연구자들이 주의해야 할 부분이라고 덧붙였다.

33) LSAC는 약 10,000여 가구의 두 개 코호트(만 0~1세 유아와 만 4~5세 아동)를 대상으로 아동의 발달과 복지를 연구하기 위하여 2004년부터 시작된 조사이다.

34) NAPLAN은 2008년 이후 매년 같은 날에 3학년, 5학년, 7학년, 9학년의 호주 학생들을 대상으로 읽기, 쓰기, 문법, 산술능력을 평가하는 프로그램이다. 반면, LSAC는 NAPLAN과 같은 직접적인 학업성취 측정변수는 수집하지 않으며, 대신 교사와 학부모에 의한 학업성취 정보를 수집한다.

제4절 데이터 연계의 잠재적 가치

행정데이터와 서베이데이터 연계의 잠재적 가치는 다양한 행정데이터와 서베이데이터에 포함된 변수를 분석할 수 있다는 사실로 요약할 수 있다. 구체적으로는 다음과 같은 잠재적 가치를 생각할 수 있다. 첫째, 행정데이터는 서베이데이터와 달리 오보고의 가능성이 낮다. 실제로 여러 기존 연구에서 연계된 행정데이터와 서베이데이터를 비교해서 소득, 세금환급, 소비 등의 변수에서 보고된 값의 차이가 상당하다는 것을 발견하였다. 둘째, 연계 가능한 행정데이터가 사전에 결정된다면, 설문에서 이들 문항을 제외함으로써 조사 비용을 상당히 줄일 수 있으며, 이는 나머지 문항의 품질 향상에도 기여할 수 있다(Penner와 Dodge, 2019). 셋째, 행정데이터와 서베이데이터의 연계는 서베이데이터의 관측기간 한계를 획기적으로 개선할 수 있다. 예를 들어, 노동시장 참여 시의 선택이 향후 퇴직·은퇴에 미치는 영향을 연구하기 위해서는 약 30여 년의 관측 기간이 필요하다. 이때, 행정데이터와 서베이데이터를 연계·활용하면 서베이데이터는 노동시장 참여 시점의 다양한 정보를 제공하고 행정데이터는 향후의 퇴직·은퇴에 대한 의사결정 결과를 제공함으로써 이러한 장기적인 효과 연구가 가능하다. 넷째, 행정데이터는 외생적으로 변하는 설명변수를 제공할 수 있으며, 그 결과 연계데이터는 인과적 추론을 가능케 한다(Røed와 Raaum, 2003).

본 절의 나머지는 행정데이터와 서베이데이터를 연계·활용한 연구사례를 소개함으로써 행정데이터 및 서베이데이터의 연계가 갖는 잠재적 가치를 시사하고자 한다. 여기에서 고용자·피고용자 데이터를 연계하는 연구사례는 제외한다. 고용자·피고용자 데이터 연계에 대한 전반적인 내용은 Abowd와 Kramarz(1999)를 참고한다. 고용자·피고용자 연계데이터를 분석한 최근 연구는 Tomaskovic-Devey 외(2015)와 Lastrapes 외(2020) 등을 참고한다.

1. 측정오류(Measurement Error)

연구자는 행정데이터와 서베이데이터를 연계함으로써 데이터 간의 품질을 비교·평가할 수 있다. 서베이데이터의 데이터 품질을 평가하는 연구는 데이터 연계 관련 초기 연구에서 자주 발견할 수 있다. 이때, 흔히 행정데이터 내의 정보를 참값으로 보고 다양한 서베이데이터가 제공하는 정보의 편의와 불확실성을 평가한다.

측정오류에 관련된 다수의 연구를 발견할 수 있다. 소득(Abowd와 Stinson, 2013), 건강보험 이용정보(Card 외, 2004), 연금 및 세금(Kapteyn와 Ypma, 2007), 교육 및 고용(Bingley와 Martinello, 2014) 등 다양한 정보 내의 측정오류를 고려하고 평가하였다. 국가별로는 미국(Card 외, 2004; Brummet, 2014), 스웨덴(Kapteyn와 Ypma, 2007), 덴마크(Bingley와 Martinello, 2014; Kreiner, Lassen, Leth-Petersen, 2015), 호주(Daraganova 외, 2013; Lujic 외, 2017) 등을 포함한다.

측정오류에 대한 연구 결과 측면에서는 다양한 결론이 혼재한다. Card 외(2004)는 SIPP 데이터와 캘리포니아 주 행정데이터를 연계하여 메디케어 지원을 받지 않으나 지원을 받는다고 오보할 확률이 1.3% 수준임을 보였고, 오보율이 소득과 관련 있음을 밝혔다. 반면, Bingley와 Martinello (2014)는 보건·노화·퇴직 조사(Survey on Health, Aging, and Retirement, 이하 SHARE) 덴마크 부표본과 덴마크 행정데이터를 연계하여 행정데이터와 서베이데이터의 소득정보 차이가 통계적으로 유의하지 않으며 고전적인 형태의 측정오류를 띠는 것을 보였다. 마찬가지로, Brummet (2014)는 서베이, 연방정부, 상거래 등의 주소 데이터를 성공적으로 연계하여 각 출처의 주소 데이터가 양질의 주소정보를 갖고 있다고 밝혔다. 이러한 상반된 결론에 대하여, Kapteyn와 Ypma(2007)는 서베이데이터와 행정데이터의 측정오류를 발견한 기존 연구에 대하여 서베이데이터와 행정데이터가 같은 개념의 변수를 수집하지 않아서 발생한 차이를 측정오류로 오해할 수 있다고 하였다.

한 국가 내에서도 정보의 종류에 따라 상반된 결론을 보이기도 한다. 예를 들면, Kreiner, Lassen, Leth-Petersen(2015)은 덴마크 가구소비조사

(Danish Family Expenditure Survey)의 소비정보와 다른 서베이데이터의 소득정보를 덴마크 행정데이터에 연계함으로써 서베이를 통해 수집된 소비정보와 소득정보의 품질을 평가하였다. 그 결과, 서베이 기반 소비정보는 잡음은 크지만 편의는 없고, 반면 서베이 기반 소득정보는 잡음도 크고 편의도 존재한다는 것을 보였다.

2. 세대간 계층이동성(Intergenerational Mobility)

세대간 계층이동성 연구는 흔히 데이터 관측기간을 넘어가는 정보가 필요한 경우가 흔하다. 때문에, 응답자에 대한 추가적인 부담 없이 하나의 데이터 출처로 세대간 계층이동성에 관하여 신뢰할 수 있는 결론을 도출할 수 있는 경우는 아주 드물다. 데이터 연계는 이러한 한계점을 극복할 수 있게 하며, 이는 최근 행정데이터를 이용한 연구가 늘어난 중요한 이유 중 하나이다. 특히, 행정데이터와 서베이데이터의 연계는 행정데이터 연계를 통해 상당한 기간의 행정정보를 연계·활용할 수 있게 할 뿐만 아니라 연구목적으로 수집된 다양한 정보를 연구에 활용할 수 있게 한다. 같은 맥락에서 O'Hara 외(2017)는 연방정부 서베이데이터와 행정데이터의 연계를 통한 가족 관련 연구의 개선 가능성을 논하였다.

세대간 계층이동에 대한 연구 중 데이터를 연계한 연구는 미국과 북유럽국가에서 다수 발견할 수 있었다. 이들 연구는 덴마크(Boserup 외, 2016), 스웨덴(Elinder 외, 2018; Erixson, 2017), 노르웨이(Bø 외, 2019), 미국(Bricker 외, 2016; Abowd 외, 2018) 등을 포함한다. Boserup 외(2016), Elinder 외(2018), Erixson(2017), Bø 외(2019)는 상속재산이 다음 세대에 미치는 영향을 분석했는데, 이들 연구는 다양한 기관에서 수집되는 행정데이터만을 연계·활용하였다. Bricker 외(2016)는 고소득층의 비율을 추정하였고, Abowd 외(2018)는 고용자가 소득불균형에 미치는 영향을 분석하였다. 특히, Bricker 외(2016)와 Abowd 외(2018)는 행정데이터와 서베이데이터를 연계하여 분석을 진행했는데, 전자는 소비자금융조사(Survey of Consumer Finances)를, 후자는 현재인구조사(Current Population Survey, 이하 CPS)와 ACS 서베이데이터를 행정데이터에 연계하였다.

한편, 사회보장프로그램에 대한 연구도 세대간 계층이동성에 관련된 분야로서 행정데이터와 서베이데이터를 연계·분석한 연구를 발견할 수 있다. Meyer와 Mittag(2019)는 4개의 이전프로그램(transfer program)에 대한 행정데이터를 CPS 데이터와 연계하여 모든 이전프로그램의 빈곤 감소 효과를 추정하였다. Powers와 Neumark(2005)는 SIPP 데이터와 행정데이터를 연계하여 보조적 보장소득(Supplemental Security Income, SSI) 프로그램과 양로보험(Old Age Insurance, OAI)의 상호작용으로 인한 조기퇴직 유인을 연구하였다.

3. 보건의료 연구(Health and Medical Research)

보건의료 연구는 데이터의 연계가 가장 활발한 분야이다. 잘 확립된 보건의료서비스 체계에서는 서비스 이용자에게 고유식별자가 부여되어 대부분의 서비스 이용정보가 수집·관리된다. 수집된 정보는 출생 및 사망 기록, 진료기록, 처방기록, 수술기록, 암기록 등 거의 모든 의료서비스 이용정보를 포함한다. 고유식별자가 존재하기 때문에 시스템 내에서 행정데이터 간의 연계도 비교적 용이하며, 다른 출처의 데이터를 연계하는 경우 이름과 주소정보를 이용해서 연계할 수 있다. 다만, 호주처럼 연방체계하에서 주정부가 정보를 수집하고 그에 대한 책임을 갖는 특수한 상황하에서는 데이터 연계가 제약적일 수 있고(Flack과 Smith, 2019), 일반진료, 의사 상담내용, 자기보고 건강상태 등 체계적으로 수집되지 않는 정보는 이용이 불가능하다(Erlangsen와 Fedyszyn, 2015).

보건의료 연구에서 행정데이터와 서베이데이터를 연계·활용한 연구를 상당수 찾을 수 있었다. Schmidt 외(2014)는 덴마크의 주민등록 시스템을 역학연구에 활용할 수 있는 접근방법에 초점을 두고 이와 유사한 연구를 수행하였다. 특히, 역학연구에서 행정데이터와 서베이데이터를 연계·활용하는 구체적 사례를 조사하여 제공하였다. Brønnum-Hansen 외(2015)는 덴마크인에게 부여되는 고유 식별번호를 이용하여 교육 및 사망 행정 데이터를 SHARE 덴마크 데이터와 연계하여 교육수준별 건강 결과 차이를 연구하였다. 마찬가지로 Foverskov 외(2018)는 덴마크 행정데이

터와 SHARE 덴마크 데이터를 연계하여 교육과 인지능력 노화 간의 관계를 연구하였다. Carrieri와 Wuebker(2016)는 SHARE와 행정데이터를 연계하여 유럽국가들에 대하여 건강정보가 예방적 의사결정에 미치는 영향을 연구하였다.

제5절 소 결

본 연구에서는 행정데이터와 서베이데이터의 연계 가능성을 검토하였다. 우선, 북유럽국가, 미국, 호주를 중심으로 행정데이터와 서베이데이터 연계 현황을 살펴보았다. 북유럽국가는 개인의 고유식별자를 이용하여 행정데이터를 체계적으로 수집하고 있으며, 서베이데이터도 이러한 개인의 고유식별자를 포함하여 서베이데이터와 행정데이터의 연계가 직관적이고 용이하다. 반면, 호주와 미국은 고유식별자가 없거나 데이터 연계에 활용하기 어려운 실정이므로 주로 이름과 주소 등의 개인정보를 이용한 확률적 기록연계 방법으로 서베이데이터와 행정데이터를 연계·활용하고 있다. 확률적 기록연계는 대개 약 80~90% 수준의 매치 성공률을 보이며 확률적 기록연계에 의한 행정데이터와 서베이데이터 연계가 실제로 실현 가능함을 알 수 있었다.

또한, 다양한 행정데이터와 서베이데이터를 연계한 다양한 연구사례를 소개함으로써 행정데이터와 서베이데이터의 연계가 가지는 잠재적 가치를 알아보았다. 행정데이터와 서베이데이터 연계의 잠재적 가치는 다양한 행정데이터를 활용할 수 있다는 점과 서베이데이터의 다양한 변수를 활용할 수 있다는 점이다. 우선, 행정데이터는 출생·사망, 노동시장 및 프로그램 참여, 의료 및 교육 서비스 이용, 소득·납세·소비 등 다양한 정보를 비교적 오랜 기간에 대하여 제공할 수 있다. 반면, 서베이데이터는, 비록 관측 기간의 한계는 존재하지만, 시간사용, 자기보고 건강상태, 식습관 등 행정데이터에서 수집하지 않는 보다 다양한 정보를 연구에 활용할 수 있다. 따라서, 행정데이터와 서베이데이터의 연계는 연구자의 데

이터 이용의 범위를 한층 더 확장할 수 있게 한다.

행정데이터와 서베이데이터를 연계·활용하는 연구에서 주의해야 할 부분이 존재한다. 우선 연구자는 행정데이터와 서베이데이터에서 수집하는 정보의 차이에 대한 이해가 필요하다. 이러한 차이는 개념적 차이, 수집대상이나 수집 시점의 차이, 응답자의 차이 등 여러 측면에서 나타날 수 있다. 또한, 연구자는 이렇게 연계한 데이터가 모집단을 대표하지 않을 수 있다는 점을 염두에 두어야 한다. 특히, 연계 성공률이 주요 변수와 연관된 경우, 연계데이터를 이용한 추정 결과에 편의가 존재할 수 있음을 고려해야 한다.

끝으로, 본 장의 연구를 토대로 한국의 행정데이터와 서베이데이터의 연계 가능성을 짐작해 볼 수 있다. 예를 들면, 한국의 대표적인 패널데이터인 한국노동패널은 주민등록번호를 수집하지 않으며, 이름, 성별, 주소, 연락처, 가구구성원 등의 정보를 수집한다. 본문의 미국과 호주 사례에 비추어 볼 때, 이름과 주소만을 이용하여 확률적 기록연계 방식으로 한국 노동패널 데이터를 (모집단)행정데이터에 연계한다면, 80~90% 수준의 매치 성공률을 예상할 수 있다. 또한, 매치에 실패하면 연락처와 가구구성원 등의 정보를 이용하여 다시 매치를 시도함으로써 이름과 주소만을 이용한 매치 성공률을 개선할 수 있을 것이다. 다만, 이러한 행정데이터와 서베이데이터의 연계를 위해서는 일종의 응답자 동의가 필요한데, 데이터 연계에 대한 동의율은 조사대상, 조사지역, 설문방식 등에 의존하므로 예측하기 어렵다. 응답자가 데이터 연계에 대하여 충분히 동의한다면, 한국에서도 이름, 주소, 연락처, 가구구성원 등의 정보를 이용하여 행정데이터와 서베이데이터의 연계가 성공적으로 이루어질 수 있을 것이다.

제 5 장

조사방법에 따른 조사 결과 검증 연구 :

한국노동패널조사 면접조사, 전화조사, 웹조사의 비교

제1절 서론

한국노동패널조사는 1998년에 시작하여 2021년 현재 24차 조사를 진행 중이다. 2020년은 코로나19가 시작된 해로, 대면조사를 기본으로 하는 노동패널 조사환경에 많은 어려움이 발생하였다. 수십년 동안 조사를 진행하면서 면접원과 라포가 생성된 패널 가구들도 면접원이 가정에 방문하는 것을 극도로 꺼려 조사를 거절하거나, 사회적 거리두기 단계가 낮아지면 방문하기를 요청하면서, 조사 방문 시기를 연기하는 경우가 많았다. 이에 노동패널팀은 대면 면접조사를 기본으로 하고, 부득이한 경우에만 다른 조사방법을 제한적으로 허용하던 기본 원칙을 유연하게 가져가기로 했다. 일차적으로 면접조사를 권유하지만, 조사를 거절하는 경우 전화조사를 적극적으로 권유하도록 하였다.

그 결과, 매년 3% 이내에서 이루어졌던 전화조사 응답비율이 18%까지 급증하였고, 상대적으로 매년 95% 이상을 차지하던 면접조사 응답비율은 80%까지 급감하였다.³⁵⁾

또한 22차(2019년 조사)에서 응답률 제고를 위해 시범 도입한 웹조사

35) 각 연도별 조사방법 분포

도 23차(2020년조사)에서 그 응답 인원이 예년에 비해 증가하였다. 웹조사는 가구 설문에 성공한 가구의 비성공 가구원에 대해서만 진행하였고, 22차에 26건이 성공하였고, 23차는 73건이 조사되었다. 이에 노동패널 기본 조사방법인 면접조사와 코로나19의 상황에 따라 급증한 전화조사, 이 사후적실패 등에 따른 유지율 하락을 방지하기 위한 웹조사 등이 노동패널의 조사방법으로 혼재되면서, 조사방법에 따른 모드 효과 검증이 필요

(단위: 명, %)

	개인 응답자 수	전체 비중	면접	유치	전화	면접+ 전화	유치+ 전화	면접+ 유치	면접+ 유치+ 전화	웹
1차년도	13,321	96.5	64.4	21.8	1.5	2.6	3.7	2.6	0.1	-
2차년도	12,037	100	77.8	7.6	7.4	4.7	2.5	0.1	0.0	-
3차년도	11,205	99.8	83.0	5.1	5.3	3.9	1.5	0.9	0.2	-
4차년도	11,051	100	83.7	3.4	4.6	4.9	1.2	1.8	0.5	-
5차년도	10,965	100	83.8	2.6	3.9	6.4	1.0	1.0	1.3	-
6차년도	11,540	100	86.3	1.7	3.8	5.7	0.9	1.3	0.4	-
7차년도	11,660	100	88.0	1.8	4.1	3.7	0.4	1.6	0.5	-
8차년도	11,580	100	85.4	1.6	2.8	6.1	1.4	1.7	1.0	-
9차년도	11,756	100	89.7	1.5	2.9	2.8	0.7	2.3	0.3	-
10차년도	11,855	100	91.3	1.1	3.2	3.1	0.6	0.6	0.1	-
11차년도	11,734	100	95.7	0.6	1.8	1.3	0.3	0.2	0.1	-
12차년도	14,489	100	93.2	1.5	2.3	2.1	0.5	0.3	0.1	-
13차년도	14,118	100	94.5	0.4	2.5	2.2	0.3	0.1	0.1	-
14차년도	13,899	100	93.6	0.8	2.9	2.0	0.4	0.3	0.1	-
15차년도	13,998	100	94.0	0.4	4.3	1.2	0.1	0.0	0.1	-
16차년도	13,887	100	95.6	0.4	2.8	0.9	0.1	0.1	0.0	-
17차년도	13,168	100	96.8	0.3	2.1	0.6	0.1	0.1	0.0	-
18차년도	14,011	100	95.4	0.5	3.1	0.7	0.2	0.1	0.1	-
19차년도	14,202	100	96.5	0.4	1.9	0.6	0.5	0.1	0.0	-
20차년도	14,475	100	96.9	0.3	2.1	0.3	0.4	0.0	0.0	-
21차년도	23,972	100	95.6	0.3	2.9	0.5	0.3	0.0	0.4	-
22차년도	23,225	100	97.1	0.3	1.8	0.6	0.1	0.0	0.0	0.1
23차년도	22,964	100	80.4	0.4	17.9	0.8	0.1	0.1	0.1	0.3

자료: 한국노동패널조사 1~23차 학술대회용 자료.

한 상태이다.

웹조사와 면접조사의 조사결과 검증연구³⁶⁾는 기수행된 적이 있으며, 본 연구는 이에 대한 후속연구이면서, 동시에 전화조사의 모드 효과도 함께 고려하는 연구로 진행할 예정이다.

제2절 노동패널 조사방법 개요

1. 웹조사 시행과정

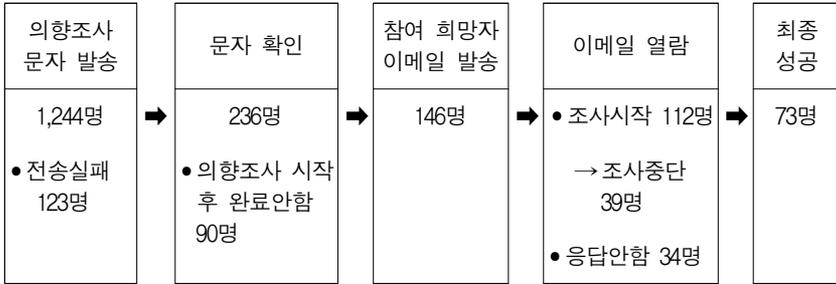
노동패널 웹조사 대상자는 기본적으로 23차 가구 설문에 응답한 가구 중, 15세 이상 가구원으로 개인 설문에 응답해야 하지만, 응답하지 않은 비성공 가구원이다.

23차 비성공 가구원에게 웹조사에 참여할 의향이 있는지 확인하는 의향조사 문자를 먼저 발송한다. 의향조사는 이름, 연락처, 이메일 등 본인 확인 문항과 웹조사 참여 의향을 묻는 문항으로 구성되었다. 의향조사 문자 발송 대상자는 23차 비성공 가구원³⁷⁾ 중, 지난 5개년 동안 연속으로 응답하지 않아 설문 응답 확률이 희박한 가구원 및 개인연락처가 없는 가구원 등을 제외한 1,244명이다. 1,244명 중 의향조사 문자를 확인한 가구

36) CAPI와 CAWI 조사 효과 실험 연구에 대한 자세한 내용은 김유빈 외(2017), 『패널자료 품질개선연구(VII)』 제7장 「웹을 활용한 조사의 효과에 대한 실증연구: 노동패널 CAWI 실험 데이터의 분석」(이지은·신선옥·이혜정·김기홍 저술)과 김유빈 외(2018), 『패널자료 품질개선연구(VIII)』 제3장 「웹을 활용한 조사의 효과에 대한 실증연구 II: 노동패널 CAWI 1차, 2차 실험 연구 결과 분석」(이지은·이혜정·신선옥·정현상 저술) 및 김유빈 외(2020), 『패널자료 품질개선연구(X)』 제5장 「웹을 활용한 조사 자료 검증: 노동패널자료를 중심으로」(이지은·채민희 저술) 참조.

37) 원칙적으로는 노동패널 본조사가 모두 완료된 이후인 12월에 웹조사를 시행하여야 하지만, 2020년은 코로나19의 영향으로 조사가 늦게 시작되어, 12월이 되어도 본조사가 완료되지 못했다. 따라서 웹조사 대상자 1,244명은 12월 1일 기준 조사 성공 가구의 비성공 가구원임을 밝힌다.

〔그림 5-1〕 노동패널 웹조사 시행 과정



원은 236명으로 19.0%이다. 이 중에서 의향조사를 시작하였지만, 완료하지 않은 인원은 90명이다. 문자를 확인하고, 웹조사를 실시하겠다고 응답한 146명에게 웹조사 링크가 있는 이메일을 발송하였다. 이 중에서 이메일을 열람한 사람은 총 112명이고, 참여희망 의사를 보였지만, 열람조사하지 않은 사람이 34명이다. 조사를 시작했지만, 39명은 조사를 중단하였고, 최종 성공한 인원은 73명이다.

본 연구에서는 웹조사에 최종적으로 성공한 73명에 대하여 분석을 시도하였다.

2. 웹조사 응답자 특성

노동패널 웹조사는 앞서 언급한 바와 같이 2019년도(22차조사)에 시범 도입되어, 2020년도(23차 조사)에서 본격적으로 시작되었다고 할 수 있다.

<표 5-1>은 22차년도와 23차년도의 개인 조사의 응답 조사방법의 분포를 보여준다. 노동패널 조사방법은 (1) 면접, (2) 유치, (3) 전화, (4) 면접 후 전화 보완, (5) 유치 후 전화 보완, (6) 면접 후 유치, (7) 면접·유치·전화가 모두 사용, (8) 웹조사 등 8가지 방법이 있다. 매해 95% 이상이 면접조사로 이루어졌으며, 전화조사방법은 3% 내외로 사용되었다.

22차에 웹조사로 응답한 26명의 23차 조사 결과를 살펴보면, 응답안함 7명, 면접조사 8명, 전화조사 5명, 그리고 웹조사로 다시 응답한 사람이 6명이었다. 각 경우에 해당하는 사례 수가 너무 적어, 특성을 파악하기는 어렵다. 다만 추후 충분한 사례 수가 확보된다면, 대상자별 선호 조사방

〈표 5-1〉 22차년도와 23차년도의 조사방법 분포

(단위: 명)

		23차									Total
		응답 안함	(1) 면접	(2) 유치	(3) 전화	(4) 면접 + 전화	(5) 유치 + 전화	(6) 면접 + 유치	(7) 면접+ 유치+ 전화	(8) 웹조사	
22 차	응답 안함	0	855	10	309	16	5	2	0	30	1,227
	(1)면접	1,389	17,357	60	3,535	123	17	23	12	34	22,560
	(2)유치	5	22	29	6	0	0	0	0	0	62
	(3)전화	62	146	0	211	3	3	0	0	3	428
	(4)면접 + 전화	19	55	0	21	32	1	0	0	0	128
	(5)유치 + 전화	3	6	0	10	2	0	1	0	0	22
	(6)면접 + 유치	1	1	0	1	0	0	0	0	0	3
	(7)면접 + 유치 + 전화	2	1	0	1	1	1	0	0	0	6
	(8) 웹조사	7	8	0	5	0	0	0	0	6	26
	합계	1,488	18,451	99	4,099	177	27	26	12	73	24,452

자료: 한국노동연구원, 「한국노동패널」 제22~23차 조사(2019~2020년도).

법을 파악하여 패널 유지율을 높일 수 있을 것으로 기대된다. 또 한 가지 주목할 점은 22차에 응답하지 않았던 30명이 23차에서 웹조사로 응답한 것인데, 이들의 특성도 추후 분석된다면, 탈락위기에 있는 패널을 다시 응답하도록 하는 방법을 제시할 수 있을 것이라 생각된다.

1차와 2차 모두 웹으로 응답한 6명의 응답자 특성을 간략하게 살펴보고자 한다. 6명의 응답자로 모든 패널에 대한 일반화는 어렵지만, <표 5-2>에 보여지듯이, 모두 4년제 대학 졸업 또는 휴학의 학력을 가지고 있다. 모두 취업자이며, 6명 중 5명이 임금근로자였음에 주목할 필요가 있다.

〈표 5-2〉 22차와 23차 모두 웹조사 응답자(6명)

pid	가구주와의 관계	성별	연령	학력	혼인 상태	고용 형태	월소득	사업장 규모	평균 근로시간
25201	가구주	남성	53	대졸 이상	기혼	임금	750	1,000명 이상	40
29601	가구주	남성	64	대졸 이상	기혼	임금	500	30~50명	40
47103	첫째 자녀	남성	23	대학휴학	미혼	임금	49	1~4인	35
68303	가구주 둘째 형제자매	여성	52	대졸 이상	미혼	비임금	200	-	40
362605	둘째 자녀	여성	28	대졸 이상	미혼	임금	250	-	50
482103	첫째 자녀	여성	28	대졸 이상	미혼	임금	180	1~4인	30

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 원자료.

〈표 5-3〉 22차, 23차 웹조사 응답자의 인구학적 특성

(단위: %)

		22차 (N=26)	23차 (N=73)			22차 (N=26)	23차 (N=73)	
성별	여성	38.5	49.3	고용 형태	임금	61.5	53.4	
	남성	61.5	50.7		비임금	7.7	11.0	
연령	20대	46.2	56.2	미취업자	미취업자	30.8	35.6	
	30대, 40대	38.5	27.4		월평균 소득 ¹⁾	100만 원 미만	11.8	13.6
	50대 이상	15.4	16.4			100만~200만 원	23.5	27.3
혼인 상태	미혼	69.2	72.6	200만~300만 원	35.3	34.1		
	기혼(유배우자)	30.8	23.3	300만~400만 원	5.9	6.8		
	이혼, 별거, 사별	-	4.1	400만 원 이상	23.5	18.2		
학력	중졸 이하	-	9.6					
	고졸	3.9	17.8					
	전문대졸	26.9	34.3					
	대졸 이상	69.2	38.4					

주: 월평균 소득은 미취업자를 제외한 취업자만으로 구한 것임.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

〈표 5-3〉은 22차(26명)와 23차(73명) 웹조사 응답자의 인구학적 특성의 차이를 살펴본 것이다. 22차에 비해 23차 조사에서 여성 및 20대, 미혼의 비율이 높아졌다. 학력은 22차에는 4년제 대졸 이상이 70%로 매우 높

왔던 것과 달리, 23차에는 전문대졸과 4년제 대졸 이상이 비슷한 비율로 조사되었다. 고용형태는 22차, 23차 모두 임금근로자의 비율이 절반 이상이고, 월평균 소득은 200만 원 이상 300만 원 미만의 비율이 가장 높았다.

3. 조사도구별 가구 특성(전화조사와 면접조사)

웹조사는 비성공 개인을 대상으로 제한적으로 수행한 반면, 전화조사는 코로나19 상황에 따라 비대면 조사를 독려하면서 급증한 방법으로, 개인조사뿐만 아니라 가구조사도 이루어졌다. 이에 기존 조사방법인 면접조사 응답가구와 전화조사 응답가구의 특성 차이를 분석하는 것은 의미가 있다.

각 가구에 속한 가구원 개인은 자신의 선호 조사방법에 따라 조사에 응할 수 있다. 한 가구의 모든 가구원이 동일한 조사방법을 사용한 경우와 한 가구에 속한 여러 가구원이 다양한 조사방법을 선호에 따라 선택한 경우로 구분하여 살펴보았다.

가구원 수에 따른 조사방법 차이를 살펴보면, 23차 응답가구 11,669가구 중 95.2%가 모든 가구원이 동일한 조사방법으로 응답하였다. 가구원 수가

〈표 5-4〉 가구원 수에 따른 개별 가구원 응답 조사방법 차이

(단위: 가구, %)

가구원 수	동일한 조사 방법		다양한 조사 방법		합계
	빈도	비율	빈도	비율	
1	3,897	100.0	-	-	3,897
2	5,083	95.3	250	4.7	5,333
3	1,342	89.6	155	10.4	1,497
4	694	84.5	127	15.5	821
5	83	81.4	19	18.6	102
6	11	64.7	6	35.3	17
7	1	50.0	1	50.0	2
합계	11,111	95.2	558	4.8	11,669

자료: 한국노동연구원, 「한국노동패널」23차 조사(2020년도) 원자료.

많아질수록 다양한 조사방법을 사용하는 경우가 많았다(표 5-4 참조).

대부분의 가구에서 모든 구성원이 동일한 조사방법으로 조사를 한다는 것은, 면접원이 가구에 연락하여 조사를 의뢰할 때, 해당 가구에서 선호하는 조사방법을 추천하는 것이 가능하다는 것을 의미한다. 따라서 단일 방법으로 조사한 가구의 조사방법 및 가구의 특성을 살펴보는 것은 의미가 있다. <표 5-5>에 의하면, 면접조사가 82.7%, 전화조사가 16.4%이다.

면접조사와 전화조사 응답가구의 가구 특성을 비교해보면, 가구원 수는 전화조사 응답가구가 평균 2.8명으로 면접조사 2.4명보다 많았다. 코로나19의 영향으로 비대면조사인 전화조사의 응답이 급격하게 증가하였는데, 그 이유는 가구에 아이가 있을 경우에 비대면 조사를 더욱 선호했을 것이라고 생각할 수 있다. 실제로 0~6세의 어린 가구원이 있는 경우, 7~12세 가구원이 있는 경우, 13~15세 가구원이 있는 경우에 모두 면접조사보다는 전화조사로 응답한 비율이 더 높았다(표 5-6 참조).

조사방법에 따른 응답가구 특성을 살펴보면, 가구소득, 거주지역, 주거형태 모두 가구 특성에서 차이가 있는 것으로 나타났다. 전화조사로 응답한 가구의 가구소득이 연평균 5,391.4만 원으로 면접조사보다 높았고, 수

<표 5-5> 동일한 조사방법으로 응답한 가구의 조사방법

(단위: 가구, %)

	빈도	비율
(1) 면접	9,192	82.7
(2) 유치	18	0.2
(3) 전화	1,818	16.4
(4) 면접+전화	67	0.6
(5) 유치+전화	6	0.1
(6) 면접+유치	8	0.1
(7) 면접+유치+전화	2	0.0
(8) 웹조사		
합 계	11,111	100.0

자료: 한국노동연구원, 「한국노동패널」23차 조사(2020년도) 원자료.

〈표 5-6〉 조사방법에 따른 가구원 수, 자녀 수 특성(면접조사와 전화조사)

(단위: 가구, %)

		면접조사 (N=9,192)	전화조사 (N=1,818)	전 체 (N=11,010)
가구원 수	1	27.5	21.7	26.5
	2	31.1	20.9	29.4
	3	17.7	21.3	18.3
	4	19.1	28.8	20.7
	5명 이상	4.7	7.3	5.2
	(평균)	2.4명	2.8명	2.5명
0~6세	0	91.4	85.0	90.3
	1	6.2	10.8	7.0
	2	2.3	3.7	2.5
	3명 이상	0.1	0.6	0.2
7~12세	0	89.2	84.3	88.4
	1	7.4	11.9	8.1
	2	3.2	3.5	3.3
	3명 이상	0.2	0.3	0.2
13~15세	0	94.3	90.2	93.6
	1	5.2	8.8	5.8
	2	0.5	1.1	0.6

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 원자료.

〈표 5-7〉 조사방법에 따른 응답가구 특성(면접조사와 전화조사)

(단위: 가구, %)

		면접조사 (N=9,192)	전화조사 (N=1,818)	T-value
가구소득	(평균)	4,402.6	5,391.4	-4.87***
주거지역	수도권	46.0	63.6	13.85***
	비수도권	54.0	36.4	
주거형태	자가	63.8	57.8	-3.83***
	전세	15.2	18.9	
	월세	16.5	18.0	
	기타	4.5	5.3	

주: * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

〈표 5-8〉 조사방법에 따른 가구주 특성(면접조사와 전화조사)

(단위: 가구, %)

		면접조사 (N=9,192)	전화조사 (N=1,818)	T-value
성별	남성	72.3	80.1	6.92***
	여성	27.7	19.9	
연령	10대, 20대	2.3	4.5	18.33***
	30대, 40대	27.7	42.7	
	50대 이상	70.0	52.8	
	(평균)	59.8세	52.4세	
학력	중졸 이하	35.9	16.9	-17.06***
	고졸	29.3	29.4	
	전문대졸	12.1	18.7	
	대졸 이상	22.6	35.0	
혼인상태	미혼	8.9	13.0	9.67***
	기혼(유배우자)	62.2	68.7	
	이혼, 별거, 사별	28.9	18.3	

주: * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

도권에 거주 비율도 63.6%로 높았다. 주거형태는 자가의 경우는 면접조사가, 전세 월세의 경우는 전화조사가 높았다(표 5-7 참조).

〈표 5-8〉은 면접조사 응답가구와 전화조사 응답가구의 가구주 특성을 살펴본 것이다. 전화조사의 평균 연령이 52.4세로 상대적으로 더 젊고, 고학력자가 더 많았다.

4. 조사도구별 개인 응답자 특성

조사방법에 따른 전체 응답자(22,964명) 대비 조사방법별 비중은 면접조사 80.3%(18,451명), 전화조사 17.8%(4,099명), 웹조사 0.3%(73명), 그 외 혼합방식이 1.5%(341명)이다. 성별, 연령, 결혼상태, 학력, 경제활동상태, 월평균 소득 등 모든 변수에서 조사방법에 따른 응답자 특성은 통계

적으로 유의한 차이가 있는 것으로 나타났다. 성별은 면접조사가 남성이 많았고, 전화조사와 웹조사는 여성이 더 많았다. 평균 연령은 면접조사는 53.4세, 전화조사는 46.2세, 웹조사는 32.7세로 큰 차이를 보였다. 결혼상태는 면접조사와 전화조사 모두 기혼(유배우자)이 60% 이상으로 가장 많았던 반면, 웹조사는 72.6%가 미혼이었다. 학력은 면접조사는 중졸, 고졸이 많았고, 전화조사는 고졸, 대졸 이상이 많고, 웹조사는 대졸 이상, 전문대졸 이상 순으로 많았다. 경제활동상태는 면접조사와 전화조사는 미취업자, 임금근로자 순으로 많고, 웹조사는 임금근로자 미취업자 순으로 많았다. 월평균 임금은 전화조사 308.9만 원, 면접조사 269.7만 원, 웹조사 248.2만 원 순으로 많았다(표 5-9 참조).

<표 5-9> 조사방법에 따른 개인응답자 특성

(단위 : %)

		면접 (N=18,451)	전화 (N=4,099)	웹 (N=73)	F-value
성별	남성	53.9	48.9	49.3	33.74***
	여성	46.1	51.1	50.7	
평균 연령(세)		53.4	46.2	32.7	562.25***
결혼상태	미혼	19.9	26.1	72.6	107.06***
	기혼(유배우자)	64.1	64.6	23.3	
	별거, 이혼, 사별	16.0	9.3	4.1	
학력	중졸	34.0	19.5	9.6	139.52***
	고졸	29.7	31.0	17.8	
	전문대졸	15.4	20.4	34.3	
	대졸 이상	20.9	29.2	38.4	
경제활동상태	임금	37.9	44.1	53.4	
	비임금	15.5	13.3	11.0	
	미취업	46.7	42.6	35.6	
월평균소득(만 원)		269.7	308.9	248.2	28.13***

주: * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

제3절 방법론

앞서 언급한 바와 같이 한국노동패널 23차에 응답한 22,623명의 응답자들 중 18.12%(4,099명)가 전화조사로 응답하였고 0.32%(73명)이 웹조사로 응답을 하였다. 본 절에서는 이러한 조사방식 차이가 설문 응답 내용에 미치는 영향을 분석하기에 앞서 관련 방법론들을 논의하고자 한다.

1. 방법론 논의

조사방식 영향 분석에 있어 핵심적인 이슈는 크게 두 가지로 나눌 수 있다. 첫째, 잠재적 내생성(endogeneity) 문제이고 둘째, 응답 품질 측정의 문제이다.

먼저 내생성 문제를 고려해야 하는 이유는 설문조사 방식을 대개 응답자가 선택하거나 응답자의 연령이나 직업 등 개인 혹은 가구적 특성과 관련이 높기 때문이다. 예를 들어 웹조사 표본과 대면조사 표본의 생활만족도 문항 내용을 비교한다고 가정해보자. 응답자가 설문조사 방식을 선택할 경우 상대적으로 디지털 기기 사용에 익숙한 젊은 층이 노년층에 비해 웹조사를 선택할 확률이 높기 마련이다. 이런 경우 웹조사 표본과 대면조사 표본의 평균 생활만족도 지수를 비교한다면 만족도 차이가 연령의 차이에서 기인하는지 혹은 실제 만족도 차이인지 판단하기 어렵다. 즉 개인적 특성 차이를 충분히 통제하지 않으면 응답 내용 차이가 조사방식 영향인지 단순히 관찰되지 않는(unobservable) 개인 특성 차이를 반영하는 것인지 구분하기 어렵다.

내생성 문제를 해결하기 위해 기존 문헌들은 1) 동일한 개인을 상이한 조사방식으로 설문하거나 2) 다양한 개인 특성 변수를 통제하거나 혹은 3) 조사방식 그룹에 무작위 배정(random assignment) 한 후 응답 내용을 비교하는 방식을 주로 택했다. Heiervang(2009)는 동일한 응답자들을 면접조사로 설문한 뒤 일정 기간 이후 웹조사로 다시 설문하여 두 응답값

평균을 비교하였다. Zhang(2017)은 면접조사, 전화조사 및 웹조사 응답결과들을 비교 분석하였는데, 조사 그룹 간 응답자 특성 차이를 고려하기 위해 성별, 나이 그리고 교육수준을 통제한 후 비교하였다. Norman(2010)의 경우 응답자들을 각각 조사방식 그룹으로 무작위 배정한 후 그룹 간 응답 내용을 비교하여 대면조사와 전화조사 대비 웹조사로 응답한 경우 응답 성실도가 낮다고 밝혔다.

무작위 배정방식은 앞서 언급한 내생성 문제를 상당 부분 극복할 수 있다는 장점이 있다. 그러나 노동패널과 같은 기존 설문조사들은 조사방법론 검증이 아닌 데이터 수집에 그 주목적이 있기 때문에 조사 품질 면에서나 비용 면에서나 무작위 배정을 수행하기 어렵다. 일부 응답자들을 대상으로 엄밀히 통제된 환경에서 소규모 실험을 진행할 수도 있겠지만 이 경우 분석 결과의 대표성에 대한 의문도 제기될 수 있다.

이와 같은 이슈들을 고려하여 김유빈 외(2020)는 두 가지 방법론을 개발했다. 첫째로는 응답자의 과거 대면조사 응답 내용을 이후 웹조사 응답 내용과 비교하는 방식이다. 이 방법은 동일한 개인이 상이한 조사방식으로 응답한 내용을 비교하여 관찰되지 않는 개인 특성들을 통제할 수 있다는 장점이 있으나 조사방식별로 응답한 시점이 다르다는 한계가 존재한다. 즉 응답자 A가 2015년에 대면조사로 응답한 내용을 2020년 전화조사 응답 내용과 비교할 경우 5년의 시간차가 존재하기 때문에 응답 내용 차이가 조사방식 영향인지 시간차를 반영하는지 구분하기 어렵다. 김유빈 외(2020)가 활용한 두 번째 방법론은 이러한 시간차 문제를 일부 해결한다. 이는 과거에 대면조사하에서 유사한 응답을 한 개인들을 그룹화하고 같은 그룹 내에서 이후에 다른 조사방식으로 응답한 내용들을 비교하는 방법이다. 즉 2015년에 유사한 응답한 개인들을 모은 후 그룹 내에서 일부가 2020년에 웹조사로, 나머지가 대면조사로 응답한 경우 그들간의 응답 차이를 분석하는 것이다. 본 분석에서는 이들이 개발한 두 번째 방법론의 기본적인 아이디어를 바탕으로 회귀모형을 추정하고자 한다. 자세한 내용은 다음 2. 분석 모형을 참고하기 바란다.

앞서 언급한 바와 같이 방법론 논의에 있어 두번째로 중요한 사항은 응답 품질의 측정이다. 다양한 조사방식들을 비교하기 위해서는 어떤 방

식하에서의 응답 품질이 나은지를 판단해야 하기 때문이다. 이러한 품질을 측정하기 위해서는 실제값과 응답 내용을 비교하는 것이 가장 이상적인 방법이나 설문조사의 경우 실제값을 알기 어렵다는 한계가 존재한다. 즉, 품질을 측정하기 위해 다른 방법들이 필요하다. 또한 주관적인 의견이나 만족도와 같은 설문의 경우 실제값이 측정 불가능한 경우도 있다. 따라서 다양한 조사방식으로 동일한 개인을 동일한 시점에서 설문하더라도 이 경우 어떤 응답이 실제에 가까운지 판단할 수 없다.

기존 문헌들은 응답 품질 측정을 위해서 설문조사 응답 신뢰도를 측정하는 지표들을 개발해왔다. [그림 5-2]는 de leeuw and van der zouwen (1988)에서 제시한 응답 품질 측정들로 1) 실제값 대비 응답 내용의 정확성, 2) 편향성 존재 여부, 3) 문항별 응답률, 그리고 4) 조사방식 간 응답 내용의 일관성 등을 포함한다. Bowling(2005)은 이 다섯 가지 지표 외에도 1) 응답 시 ‘예’를 선택하는 행태(Yes-saying bias), 2) 응답자들의 기억력의 정확성(Recall bias), 3) 조사자 특성에 응답자가 영향을 받는 경향성(Interviewer bias) 등을 품질 측정 지표로 제시하였다.

Bowling(2005)은 여러 가지 조사방식 영향을 분석한 기존 논문들에서 도출한 결과들을 살펴본 후 de leeuw and van der zouwen(1988)에서 제시한 지표를 적용하여 다음과 같이 요약했다. 첫째, 대면조사는 전화조사나 웹조사에 비해 설문 참여율이 높고 각 문항 응답률도 높다. 둘째, 그러나 면접조사에서는 응답자들이 사회적으로 이상적인 답변들을 고르는 경향(social desirability bias)이나 ‘예’를 선택할 확률(Yes-saying bias)이 더 높다. 이러한 경향은 전화조사에서도 관찰되는데 웹조사에서 상대적으로 최소화된다. 다만 웹조사의 경우 응답자들이 부정확한 기억력으로

[그림 5-2] 설문조사 데이터 품질 측정 지표

De Leeuw and Van Der Zouwen(1988) 설문조사 데이터 품질 측정 기준

- 1) 응답 내용이 실제치(True Value) 대비 얼마나 정확한가
- 2) 사회적 바람직성 편향(Social Desirability Bias)이 존재하지는 않는가
- 3) 문항별로 응답률이 높은가
- 4) 자유응답의 경우 응답 분량이 충분히 많은가
- 5) 여러 가지 조사방식으로 동일한 설문에 응답했을 때 그 내용이 유사한가

응답하는 경향(recall bias)이 상대적으로 강하게 나타났다.

본 분석에서도 이들의 방식을 참고하여 각각 조사방식이 응답 신뢰도 즉, 품질에 영향을 미치는지 간략하게 살펴보기로 한다. 노동패널 자료 특성을 고려하여 사회적으로 이상적인 답변을 택하는 경향이 있는지와 중간값을 선택하는 경향이 있는지를 분석하고자 한다. 자세한 내용은 다음을 참고하기 바란다.

2. 분석 모형

앞서 논의한 바와 같이 김유빈 외(2020)는 다른 모든 조건이 동일하다면 과거 응답값이 같은 경우 미래 응답값도 같다는 가정하에 분석을 수행하였다. 즉 두 사람이 $t-1$ 기에 동일한 응답을 했다면 t 기에도 응답값이 동일해야 한다. 만약 t 기에 두 사람의 응답값이 다르고 각자 다른 조사방식하에서 설문에 응했다면 그 응답값 차이는 조사방식의 영향으로 간주된다. 본 분석에서도 이 방법론의 기본적인 아이디어를 바탕으로 하되 일부 한계점을 보완한 회귀모형을 활용하고자 한다.

분석 모형은 아래 식 (1)과 식 (2)이다. 식 (1)은 전화조사와 대면조사 그룹을 비교하고 식 (2)는 웹조사와 대면조사 그룹을 비교한다. 실질적으로 두 식은 동일하나 방법론은 상이하다. 웹조사 영향 분석의 경우 웹조사 표본 수가 대면조사 표본 수에 비해 현저히 적기 때문에 모형 추정 이전에 매칭 단계를 추가적으로 수행하였다.

$$y_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 phone_{i,t} + \beta_3 phone_{i,t} \times y_{i,t-1} + X_i + u_{i,t} \quad (1)$$

$$y_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 web_{i,t} + \beta_3 web_{i,t} \times y_{i,t-1} + X_i + u_{i,t} \quad (2)$$

$y_{i,t}$ 는 개인 i 의 t 기(23차) 응답값으로 소득이나 건강, 만족도 등이다. $y_{i,t-1}$ 는 동일한 문항 y 에 대한 i 의 과거 응답값이며 22차 응답값을 활용했다. 따라서 β_1 은 22차 y 와 23차 y 의 상관관계를 보여준다. $phone_{i,t}$ ($web_{i,t}$)은 더미변수로 i 가 23차에 전화조사(웹조사)로 응답한 경우 1, 대면조사로 응답한 경우 0의 값을 가진다. 앞서 언급한 바와 같이 김유빈 외(2020)는 추정 모형에 조사방식 더미변수를 포함하여 조사방식 영향을

추정하였다. 즉 이는 조사방식 효과가 개인 특성에 관계없이 동일하다는 가정이 수반된다. 본 분석에서는 개인별로 그 효과가 다를 수 있다는 점을 고려하여 $phone_{i,t} \times y_{i,t-1}$ 는 22차 응답값과 23차 $phone_{i,t}$ 간의 교차항을 추가하였다. 예를 들어 저소득자의 경우 소득 관련 문항을 답변할 때 조사원을 직접 대면하지 않는 설문 방식하에서 실제값을 응답할 가능성이 평균수준 소득자에 비해 높을 수 있다. 따라서 β_3 은 조사방식의 영향이 기존 값 $y_{i,t-1}$ 에 따라 상이한지를 테스트한다. 따라서 $\beta_3(\neq 0)$ 이 통계적으로 유의하다면 조사방식의 영향이 개인적 특성에 따라 다르다고 결론내릴 수 있다. X_i 는 그 외 개인 특성 변수로 교육수준, 성별과 연령 더미변수 및 거주지역 더미변수를 포함한다.

웹조사 영향 분석에는 식 (2)를 추정하기에 앞서 프로빗 모형을 활용하여 간단한 표본 매칭 단계를 수행하였다. 잠재적 종속변수는 웹조사에 참여할 확률이고 관측되는 종속변수는 $web_{i,t}$ 이다. 따라서 $web_{i,t}$ 를 종속변수로 하고 독립변수로 웹조사에 영향을 미칠 수 있는 변수들을 포함하였다. 이는 개인 교육연수, 성별더미, 거주지역 더미, 연령, 연령 제곱항, 연령 제곱항, 현재 건강상태 및 가구주와의 관계 더미변수이다. 이를 통해 적합값(Fitted value)을 구한 후 웹조사 그룹과 유사한 대면조사 그룹 응답자들만을 표본에 포함하였다.³⁸⁾ 이후 분석 절차는 전화조사 효과 분석과 동일하다.

제4절 통계적 검증

1. 분석 개요

이 절에서는 앞서 소개한 방법론을 적용해 대면조사, 전화조사 및 웹조

38) 추정된 적합값들은 0.00000422에서 0.374839 범위 내에 있다. 매칭을 위해 먼저 적합값들을 소수점 다섯째자리에서 반올림한 후 '이 값이 동일한 웹조사 응답자가 존재하는 대면조사 응답자'만을 표본에 포함하였다.

사 방식이 응답 행태에 미치는 영향을 분석한다. 분석 문항은 크게 세 가지로 분류된다. 1) 일자리 및 소득, 2) 건강 및 생활습관 관련 문항 그리고 3) 직장 및 생활 만족도 관련 문항들이며, <표 5-10>에 적시하였다. 분석 자료는 노동패널 22차와 23차 개인 조사 원자료를 활용한다.

<표 5-10> 주요 변수 설명

변수	설명
(1) 개인 일자리 및 소득	
월평균 임금액수	로그값
월평균 소득액수(비임금)	로그값
정규/비정규 여부	(1) 정규직 (0) 비정규직
주당 근무시간(비임금)	단위: 시간
근로계약기간 유무	(1) 근로계약기간이 있다 (0) 없다
근로시간 규칙성	(1) 규칙적이다 (0) 규칙적이지 않다
(2) 개인 건강 및 생활습관	
현재 ___님의 건강상태는 어떻다고 생각하십니까?	(1) 아주 건강하다 (2) 건강한 편이다 (3) 보통이다 (4) 건강하지 않은 편이다 (5) 건강이 아주 안 좋다
그렇다면 ___님 나이의 보통사람들과 비교해볼 때 ___님의 건강상태는 어떻다고 생각하십니까?	
(1) 시각, 청각, 언어장애 등 감각기관의 장애가 있다	(1) 예 (0) 아니오
(2) 걷기, 계단 오르기, 물건을 들거나 운반하기 등의 육체적 활동에 상당한 제약이 있다	
활동 제약 여부	
1. 배우기, 기억하기, 집중하기에 어려움을 느낀다 2. 옷입기, 목욕하기 등 집안 돌아다니는 데에 어려움을 느낀다 3. 쇼핑하기, 병원가기 등 집밖 돌아다니는 데에 어려움을 느낀다 4. 직업활동을 하는 데 어려움을 느낀다	
(3) 개인 직장 및 생활 만족도	
직장만족도	1. 지금 근무하고 있는 직장은 다닐만한 좋은 직장이다 (1) 전혀 그렇지 않다 (2) 그렇지 않은 편이다

〈표 5-10〉의 계속

변수		설명
직장만족도	2. 나는 이 직장에 들어온 것을 기쁘게 생각한다 3. 직장을 찾고 있는 친구가 있으면 나는 이 직장을 추천하고 싶다 4. 나는 내가 다니고 있는 직장을 다른 사람들에게 자랑할 수 있다	(3) 보통이다 (4) 그런 편이다 (5) 아주 그렇다
생활만족도	1. 가족의 수입 2. 여가활동 3. 주거환경 4. 가족관계 5. 친인척 관계 6. 사회적 친분 관계	(1) 매우 만족 (2) 만족 (3) 보통 (4) 불만족 (5) 매우 불만족

2. 분석 결과 : 대면조사 및 전화조사 비교

가. 개인 일자리 및 소득 변수

〈표 5-11〉은 조사방식이 개인 일자리 및 소득 응답 내용에 미치는 영향을 추정한 결과이다. 종속변수로는 총 6가지 문항 - 월평균 임금액수(로그값), 월평균 소득액수(로그값), 정규직 더미, 주당 근무시간, 근로계약기간 존재 더미, 근로시간 규칙성 여부 더미변수이다. 식 (1)에서 $y_{i,t-1}$ 에 해당하는 과거 응답 내용은 22차 자료를 활용하였다.

먼저 첫 번째 행은 추정된 β_1 으로 22차 응답 내용과 23차 응답 내용의 상관도를 보여준다. 예를 들어 종속변수가 월평균 임금액 로그값인 경우에 ‘과거 응답’ 변수는 22차 월평균 임금액 로그값이다. 추정된 계수값들은 모두 0.5 이상이고 1% 유의 수준하에서 통계적으로 유의하여 과거 변수와 현재 변수의 상관도가 높음을 알 수 있다.

다음으로 3행은 ‘전화조사’ 변수의 계수 β_2 추정값을 보여준다. 주목할 점은 ‘근로계약기간 있음’ 및 ‘근로시간 규칙성’을 제외하고는 모두 통계

적으로 유의하지 않다는 점이다. 다시 말해서 월평균 임금 및 소득, 정규직 여부, 주당근무시간 관련 응답 시 응답 행태에 조사방식이 유의미한 영향을 준다고 판단하기 어렵다. 다만 한 가지 주목할 만한 점은 월평균 임금, 월평균 소득, 정규직 여부, 근로계약기간 관련 문항은 계수 추정값이 음수라는 점이다. 즉, 전화조사에 참여한 응답자일수록 월평균 임금 및 소득과 정규직 확률 등을 낮게 보고하는 경향이 있다.

다음으로는 교차항(과거응답*전화조사) 계수 추정값을 살펴보겠다. 전화조사 참여 효과와 마찬가지로 교차항 또한 대부분의 경우 통계적으로 유의하지 않다. 이는 응답자들의 소득 및 일자리 관련 응답 행태가 전화조사나 대면조사 방식에 좌우되지 않으며 이러한 영향은 개인적 특성과도 크게 관련이 없다는 해석이 가능하다. 다만 근로시간 규칙성의 경우 교차항 및 전화조사 영향이 1% 유의수준하에서 통계적으로 유의하게 나타났다. 해당 변수는 근로시간이 규칙적일 경우 1을, 그렇지 않을 경우 0이기 때문에 과거 응답값이 1이고 전화조사에 참여(=0.096-0.097)한 경우

<표 5-11> 전화조사 및 대면조사 비교 - 개인 일자리 및 소득 관련 변수

	월평균 임금 (1)	월평균 소득 (비임금) (2)	정규직 여부 (3)	주당 근무시간 (비임금) (4)	근로계약 기간 있음 (5)	근로시간 규칙성 (6)
과거응답	0.849*** [0.015]	0.803*** [0.020]	0.924*** [0.008]	0.652*** [0.016]	0.612*** [0.014]	0.677*** [0.012]
과거응답* 전화조사	0.013 [0.019]	0.003 [0.047]	0.002 [0.011]	-0.013 [0.050]	-0.051 [0.039]	-0.097*** [0.028]
전화조사	-0.074 [0.103]	-0.091 [0.265]	-0.009 [0.010]	0.010 [2.346]	-0.013* [0.006]	0.096*** [0.030]
관측치 수	7522	2417	7554	3085	7560	10877
R-squared	0.869	0.742	0.904	0.555	0.483	0.538

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

2) [] 안은 표준편차 값임.

3) * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」23차 조사(2020년도) 학술대회용 자료.

에는 두 계수값의 합이 0에 가까우므로 전화조사의 영향이 있다고 보기 어렵다. 그러나 과거 응답값이 0이고 23차에 전화조사에 참여(=0.096)한 경우에는 전화조사 계수값이 0.096으로 과거에 근로시간이 규칙적이지 아니라고 응답한 이들이 23차에 전화조사로 응답한 경우에 (대면조사 대비) 근로시간이 규칙적이라고 말할 확률이 더 높다는 해석이 가능하다.

나. 건강 및 생활습관 관련 변수

다음으로는 건강 및 생활습관 관련 변수들을 살펴보겠다. 이는 주관적인 건강상태를 묻는 문항인 ‘현재 건강상태는 어떠한가’ 및 ‘보통사람 대비 건강상태는 어떠한가’와 감각기관이나 육체적 제약 및 활동제약 여부를 묻는 문항들을 포함한다. 또한 민감문항으로 분류될 수 있는 생활습관 관련 변수인 흡연 여부와 음주 여부를 살펴본다.

<표 5-12>는 추정 결과를 보여준다. 주목할 만한 점은 현재 건강상태와 보통사람 대비 건강상태를 묻는 응답의 경우 전화조사 영향이 통계적으로 유의하게 나타났다는 점이다. 반면 감각기관 장애나 활동 제약 등을 묻는 질문에는 조사방식의 유의한 영향이 관찰되지 않는다. 민감문항인 흡연 여부와 음주 여부도 마찬가지이다. 이는 스스로 건강상태를 진단하고 응답해야 하는 경우 주관적인 판단이 요구되고 응답시간이 상대적으로 길 수 있기 때문에 조사방식의 영향이 상대적으로 클 수 있다. 즉 조사방식이 응답자가 민감하게 느끼는 문항보다는 상대적으로 응답시간이나 더 많은 사고를 요구하는 문항 응답행태에 더 영향을 끼치는 것으로 관찰된다.

<표 5-12>의 첫 번째와 두번째 열을 보면 전화조사로 응답한 그룹이 대면조사 그룹에 비해 건강상태가 더 안 좋다고 응답하였다(0.170:0.264). 교차항 계수는 모두 음수인 -0.073 및 -0.088이므로 전화조사 그룹의 경우 과거응답 추정계수가 0.459(=0.532-0.073) 및 0.441(=0.529-0.088)로 대면조사 그룹보다 작다. 즉 22차에 대면조사로 응답하고 23차에 다시 대면조사로 응답한 그룹에 비해서 23차에 전화조사로 응답한 그룹의 경우 23차 응답값이 과거 대비 더 많이 바뀌었다. 이는 다시 말해 2년간 조사방

<표 5-12> 전화조사 및 대면조사 비교 - 건강 및 생활습관 관련 응답

	현재 건강상태	보통사람 대비 건강상태	감각기관 장애	육체적 제약	흡연 여부	음주 여부
과거응답	0.532*** [0.008]	0.529*** [0.008]	0.447*** [0.027]	0.495*** [0.014]	0.950*** [0.005]	0.935*** [0.004]
과거응답* 전화조사	-0.073*** [0.015]	-0.088*** [0.017]	0.001 [0.093]	-0.054 [0.036]	0.010 [0.007]	0.009 [0.005]
전화조사	0.170 [0.034]	0.264*** [0.046]	0.000 [0.001]	0.004 [0.003]	-0.001 [0.002]	-0.002 [0.003]
관측치 수	21377	21378	21379	21379	21,382	21,382
R-squared	0.511	0.439	0.321	0.425	0.935	0.902
	활동제약1 (배우기, 집중)	활동제약2 (웃입기 등)	활동제약3 (쇼핑, 병원가기)	활동제약4 (직업활동 등)		
과거 응답	0.512*** [0.022]	0.405*** [0.028]	0.426*** [0.018]	0.539*** [0.014]		
과거응답* 전화조사	-0.015 [0.039]	0.017 [0.066]	-0.046 [0.046]	-0.002 [0.024]		
전화조사	0.002 [0.002]	0.001 [0.002]	0.003 [0.003]	-0.004 [0.003]		
관측치 수	21380	21379	21379	21379		
R-squared	0.407	0.285	0.356	0.488		

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

현재 건강상태: 매우 좋음(1), 좋음(2), 보통(3), 안 좋음(4), 매우 안 좋음(5); 보통 사람 대비 건강상태: 매우 좋음(1), 좋음(2), 보통(3), 안 좋음(4), 매우 안 좋음(5); 감각기관 장애, 육체적 제약, 활동제약1(배우기, 집중 등), 활동제약2(웃입기 등), 활동제약3(쇼핑, 병원가기 등), 활동제약4(직업활동 등): 있다(1) 없다(0); 흡연 여부, 음주 여부: 흡연/음주를 한다(1) 안한다(0).

2) [] 안은 표준편차 값임.

3) *10% 유의수준, **5% 유의수준, ***1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

식의 차이가 응답값에 영향을 끼쳤을 수 있다는 해석이 가능하다.

그 외 문항은 감각기관 장애 여부, 배우기나 집중에 어려움을 느끼는지를 묻는 활동제약 1, 웃입기 등 실내 활동에 어려움이 있는지를 묻는 활

동제약 2, 쇼핑을 하거나 병원에 가는 등 실외 활동에 어려움이 있는지를 묻는 활동제약 3 등과 앞서 언급한 흡연 여부와 음주 여부 문항들이다. 대부분의 계수가 0에 가깝고 통계적으로 유의하지 않으며 계수 부호 또한 일관적인 패턴을 보이지 않아 조사방식이 이들 응답행태에 영향을 미친다고 보기 어렵다.

다. 생활 및 직장 만족도

<표 5-13>은 만족도 문항 분석 결과이다. 먼저 생활만족도 문항 분석 결과를 보면 ‘가족관계 만족도’와 ‘주거환경 만족도’ 문항을 제외하고는 모든 경우에 전화조사 영향이 10% 유의수준하에서 유의하게 나타났다. 전화조사 그룹이 대면조사 그룹에 비해 가족의 수입이나 여가활동 및 주거환경에 더 만족한다고 응답하였다. 반면 친인척관계나 사회적 친분관계는 대면조사 그룹이 더 만족하는 경향(0.080:0.078)을 보였다.

반면 직장만족도는 전화조사 영향이 통계적으로 유의하지 않으며 계수 값이 거의 0에 가까운 것으로 보아 조사방식이 이들 문항 응답행태에 영향을 미친다고 보기 어렵다. 앞서 건강 관련 문항 응답 행태와 마찬가지로 조사방식이 전반적인 응답행태에 영향을 미친다기보다는 문항 종류에

<표 5-13> 전화조사 및 대면조사 비교 - 생활 및 일자리 만족도

	생활만족도(1: 매우 만족~5: 매우 불만족)					
	가족의 수입	여가활동	주거환경	가족관계	친인척관계	사회적 친분관계
과거응답	0.471 *** [0.009]	0.398 *** [0.008]	0.433 *** [0.007]	0.445 *** [0.008]	0.438 *** [0.007]	0.414 *** [0.007]
과거응답* 전화조사	0.028 [0.019]	0.020 [0.016]	0.006 [0.019]	-0.028 [0.020]	-0.029* [0.016]	-0.035** [0.016]
전화조사	-0.133** [0.055]	-0.076* [0.044]	-0.047 [0.044]	0.045 [0.048]	0.080* [0.041]	0.078** [0.039]
관측치 수	21,316	21,318	21,317	21,317	21,318	21,318
R-squared	0.304	0.240	0.231	0.243	0.250	0.237

〈표 5-13〉의 계속

	직장 만족도(1: 전혀 그렇지 않다~5: 아주 그렇다)			
	직장만족도 1 좋은 직장이다	직장만족도 2 기쁘게 다닌다	직장만족도 3 추천하고 싶다	직장만족도 4 자랑할 수 있다
과거 응답	0.512*** [0.022]	0.405*** [0.028]	0.426*** [0.018]	0.539*** [0.014]
과거응답* 전화조사	-0.015 [0.039]	0.017 [0.066]	-0.046 [0.046]	-0.002 [0.024]
전화조사	0.002 [0.002]	0.001 [0.002]	0.003 [0.003]	-0.004 [0.003]
관측치 수	21380	21379	21379	21379
R-squared	0.407	0.285	0.356	0.488

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

생활만족도 문항: 매우만족(1), 만족(2), 보통(3), 불만족(4), 매우불만족(5); 직장 만족도 문항: 전혀 그렇지 않다(1), 그렇지 않은 편이다(2), 보통이다(3), 그런 편이다(4), 아주 그렇다(5).

2) [] 안은 표준편차 값임.

3) *10% 유의수준, **5% 유의수준, ***1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

따라 다르다는 결론을 내릴 수 있다.

라. 조사방식이 특정값 선택 행태에 미치는 영향

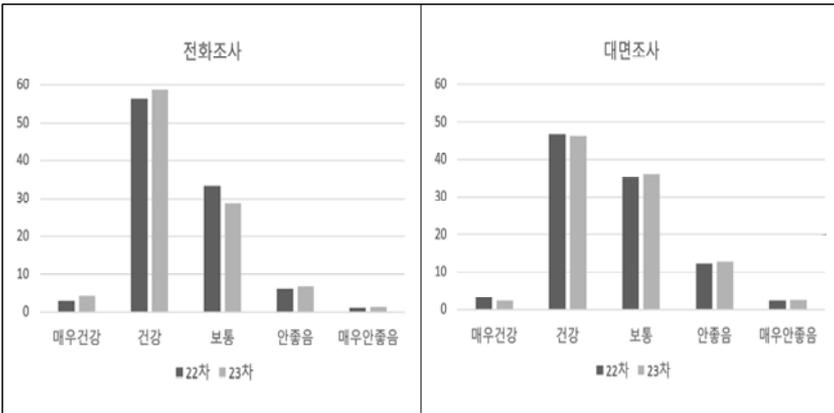
앞서 논의한 바와 같이 기존 문헌에서는 조사방식이 Yes-saying이나 중간값 혹은 사회적으로 바람직해 보이는 특정값을 집중적으로 선택하는 행태에 영향을 미칠 수 있다고 밝혔다. 따라서 건강 및 만족도 관련 문항들 중에 앞서 통계적으로 유의한 영향이 관찰된 경우에 한해서, 이러한 영향이 특정값을 선택하는 경향과 연관이 있는지를 간단히 살펴보고자 한다.

[그림 5-3]은 현재 건강상태에 대한 응답 분포로 왼쪽은 전화조사 그룹 응답 분포, 오른쪽은 대면조사 그룹 응답 분포이다. 먼저 전반적으로 조사방식과 관계 없이 22차와 23차 응답분포가 유사하지만 전화조사 그룹

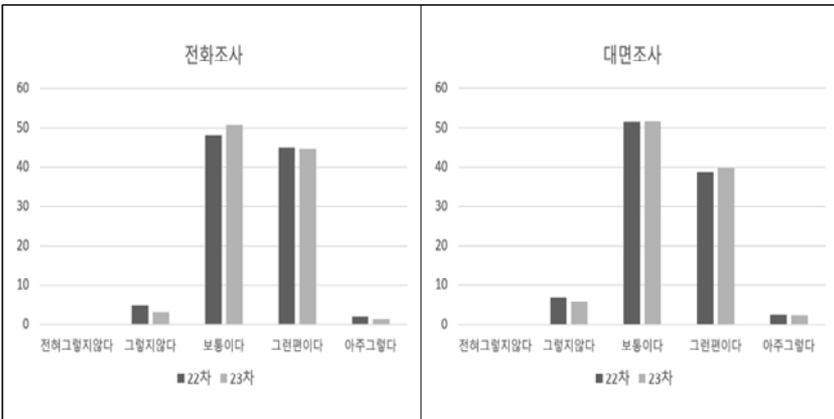
이 더 큰 차이를 보인다. 전화조사 그룹에서 22차에 건강상태를 ‘보통’ 이라고 응답한 비율은 약 33%이며 23차에는 30% 미만으로 감소했다.

[그림 5-4]는 직장만족도 ‘지금 직장은 다닐만 하다’에 대한 응답 분포이다. 전화조사 그룹의 경우 22차 대비 23차에 ‘보통이다’는 비율이 증가하고 ‘그렇지 않다’는 비율이 소폭 감소했는데 차이 폭이 크지 않고 대면 조사의 경우에도 소폭 변화가 관찰되어 조사방식의 영향으로 보기는 무리가 있다.

(그림 5-3) 조사그룹별 연도별 응답분포 : (문항) 현재 건강상태



(그림 5-4) 조사그룹별 연도별 응답분포 : (문항) 직장만족도 1 : 지금 직장은 다닐 만하다



단순 응답 분포에서는 조사방식의 영향으로 볼 만한 상당한 차이는 관찰되지 않았다. 그러나 앞서 언급한 바와 같이 조사방식 그룹 간에 개인 특성 차이가 상당 부분 존재하기 때문에 단순 분포 비교만으로 조사방식의 영향이 미미하다고 결론 내리기 어렵다. 따라서 아래 추정 모형으로 추가적인 분석을 수행하고자 한다.

$$D_{i,t} = \beta_0 + \beta_2 phone_{i,t} + W_i + u_{i,t}$$

먼저 $D_{i,t}$ 는 23차에 중간값과 바람직한 값을 선택했는지 여부를 나타내는 더미변수이다. 예를 들어 현재 건강상태 중간값 더미변수는 ‘보통’으로 응답한 경우 1, 그렇지 않은 경우 0의 값을 갖는다. 바람직한 값 더미변수는 ‘매우 좋음’으로 응답한 경우 1, 그렇지 않은 경우 0이다. $phone_{i,t}$ 은 23차에 전화조사 응답 여부를 나타내는 더미변수이다. W_i 는 연령 더미변수, 성별, 교육연수와 지역더미변수를 포함하고 이에 22차 응답값 더미변수를 추가로 통제하였다. 예를 들어 $D_{i,t}$ 가 현재 건강상태 중간값 더미변수인 경우 W_i 는 22차 현재 건강상태 응답 더미변수 4개(매우 좋음, 좋음, 보통, 좋지않음 여부 더미변수들)를 통제한다. 이는 과거에 동일한 응답을 한 조건부하에 전화조사 방식이 중간값을 선택하는 경향에 영향을 미치는지 보기 위함이다.

<표 5-14>는 추정결과를 보여준다. 먼저 전체적으로 계수가 0에 가깝고 통계적으로 유의하게 추정된 계수가 많지 않으며 계수 부호도 변수 간에 일정하지 않다. 전화조사 그룹이 대면조사 그룹에 비해 현재 건강상태를 ‘보통’으로 응답할 확률이 약 2.2%포인트 높고 가족수입만족도의 경우 ‘보통’으로 응답할 확률도 더 낮다. 전화조사 그룹이 직장만족도를 ‘매우 만족’으로 응답할 확률이 대면조사 그룹에 비해 더 낮은데, 이는 유의수준 5%하에서 유의하나 계수값이 1%포인트 이하이다. 즉 조사방식이 중간값이나 이상적 답변을 선택하는 경향에 영향을 미친다는 가설을 뒷받침할 뚜렷한 패턴이 보이지 않는다.

〈표 5-14〉 조사방식 영향 추정 : 중간값 및 이상적 답변 선택 경향

	현재 건강상태 '보통' 응답	직장만족도 1 '보통' 응답	가족수입만족도 '보통' 응답	사회적 친분관계 만족도 '보통' 응답
전화조사 그룹	0.022** [0.009]	0.002 [0.013]	-0.030*** [0.010]	-0.007 [0.010]
관측치 수	21,377	7,550	21,316	21,318
R-squared	0.228	0.221	0.153	0.200
	현재 건강상태 '매우 좋음' 응답	직장만족도 1 '매우 만족' 응답	가족수입만족도 '매우 만족' 응답	사회적 친분관계 만족도 '매우만족' 응답
전화조사 그룹	0.001 [0.002]	-0.009** [0.004]	0.000 [0.001]	-0.002 [0.001]
관측치 수	21,377	7,550	21,316	21,318
R-squared	0.176	0.12	0.042	0.115

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

2) [] 안은 표준편차 값임.

3) *10% 유의수준, **5% 유의수준, ***1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

3. 분석 결과 : 대면조사 및 웹조사 비교

웹조사의 경우 이중차분방식을 적용하므로 표본 수에 한계가 존재한다. 앞서 언급한 대로 23차에 총 73명만이 웹조사에 참여하였고 이와 더불어 문항에 따라 무응답이 존재하기 때문에 실질적으로 활용할 수 있는 관측치 수는 많지 않다. 예를 들어 월평균 임금액수의 경우 23차와 22차에 모두 응답한 웹조사 그룹 응답자 수는 8명, 23차와 21차에 모두 응답한 응답자 수는 12명에 불과하다. 따라서 결과 해석에 있어 주의가 필요하다.

<표 5-15>, <표 5-16> 및 <표 5-17>은 분석 결과를 보여준다. 앞서 논의한 바와 같이 간단한 표본 매칭 단계를 거친 후 전화조사 경우와 동일한 방식으로 모형을 추정하였다.

먼저 <표 5-15>는 월평균 소득(비임금)과 정규직 여부, 주당근무시간

〈표 5-15〉 웹조사 및 대면조사 비교 - 개인 소득 및 일자리 관련 변수

	월평균 임금 (1)	월평균 소득 (비임금) (2)	정규직 여부 (3)	주당 근무시간 (비임금) (4)	근로계약기 간 있음 (5)	근로시간 규칙성 (6)
과거응답	0.861*** [0.024]	0.792*** [0.042]	0.899*** [0.018]	0.614*** [0.030]	0.610*** [0.037]	0.682*** [0.021]
과거응답* 웹조사	-0.125 [0.134]	-0.651*** [0.067]	0.213*** [0.060]	-1.337*** [0.376]	-0.424 [0.378]	-0.010 [0.245]
웹조사	0.815 [0.806]	3.688*** [0.253]	-0.145*** [0.040]	43.177*** [11.521]	0.214 [0.184]	-0.061 [0.221]
관측치 수	1,383	461	1,386	569	1,389	1,998
R-squared	0.888	0.792	0.886	0.628	0.457	0.569

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

2) [] 안은 표준편차 값임.

3) * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

(비임금) 응답 항목에 있어 통계적으로 유의한 결과를 보여준다. 대면조사 그룹 대비 웹조사 그룹은 월평균 소득을 더 높게(3.688), 정규직 확률을 더 낮게(-0.145), 그리고 주당 근무시간(43.177)을 더 높게 응답하는 경향이 있다고 나타났다. 그러나 추정계수 크기가 지나치게 크다는 점과 결측치 값들로 인해 표본에 포함된 웹조사 응답자들이 10명 미만이라는 점 등을 고려했을 때 결과 해석에 주의가 필요하다.

<표 5-16>은 건강 및 생활습관 관련 응답 분석을 보여준다. 이 분석은 웹조사 응답자 수가 총 43명으로 앞서 분석 대비 신빙성이 높다고 가정하고자 한다. 먼저 ‘현재 건강상태’와 ‘보통사람 대비 건강상태’ 응답 분석 결과, 주요 추정계수들이 5% 유의 수준하에서 통계적으로 유의하게 나타나지 않았다. 그러나 전화조사 영향 추정 계수가 0.835 및 0.866으로 종속변수가 1에서 5의 값을 가진다는 것을 감안했을 때 그 크기가 작지 않다. 또한 <표 5-12>에서 살펴본 전화조사 영향 추정계수와 부호가 일치한다는 점에서 주목할 만하다. 즉 웹조사로 응답한 경우 대면조사 대비 현재 건강을 더 안 좋다고 응답하는 경향이 있다. 그 외 장애나 제약, 흡연 여부 및

음주 여부의 경우에도 웹조사 영향이 통계적으로 유의하거나 계수 크기가 의미가 있는 것으로 나타났다. 예를 들어 웹조사 그룹이 대면조사 그룹에 비해 흡연을 한다고 응답할 확률이 더 높고 음주 여부도 마찬가지이다.

〈표 5-16〉 웹조사 및 대면조사 비교 - 건강 및 생활습관 관련 변수

	현재 건강상태	보통사람 대비 건강상태	감각기관 장애	육체적 제약	흡연 여부	음주 여부
과거응답	0.507*** [0.016]	0.536*** [0.015]	0.419*** [0.049]	0.462*** [0.026]	0.944*** [0.010]	0.934*** [0.007]
과거응답* 웹조사	-0.210 [0.166]	-0.226 [0.179]	0.000 [0.000]	-0.508*** [0.058]	-0.054 [0.056]	-0.593*** [0.136]
웹조사	0.835* [0.434]	0.866 [0.571]	0.047 [0.032]	0.021 [0.023]	0.098* [0.056]	0.357*** [0.088]
관측치 수	3,862	3,862	3,862	3,862	3,862	3,862
R-squared	0.519	0.469	0.278	0.417	0.923	0.893
	활동제약 1 (배우기, 집중)	활동제약 2 (옷입기 등)	활동제약 3 (쇼핑, 병원가기)	활동제약 4 (직업활동 등)		
과거 응답	0.465*** [0.050]	0.398*** [0.049]	0.427*** [0.040]	0.558*** [0.026]		
과거응답* 웹조사	0.000 [0.000]	-0.411*** [0.055]	0.000 [0.000]	-0.775*** [0.080]		
웹조사	0.126** [0.051]	0.025 [0.023]	0.024 [0.025]	0.145* [0.078]		
관측치 수	3,862	3,862	3,862	3,862		
R-squared	0.379	0.313	0.381	0.482		

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

현재건강상태: 매우좋음(1), 좋음(2), 보통(3), 안좋음(4), 매우안좋음(5);

보통사람 대비 건강상태: 매우좋음(1), 좋음(2), 보통(3), 안좋음(4), 매우안좋음(5); 감각기관 장애, 육체적 제약, 활동제약1(배우기, 집중 등), 활동제약2(옷입기 등), 활동제약3(쇼핑, 병원가기 등), 활동제약4(직업활동 등): 있다(1) 없다(0); 흡연 여부, 음주 여부: 흡연/음주를 한다(1), 안한다(0).

2) [] 안은 표준편차 값임.

3) * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

<표 5-17> 웹조사 및 대면조사 비교 - 생활 및 일자리 만족도

	생활만족도(1: 매우 만족~5: 매우 불만족)					
	가족의 수입	여가활동	주거환경	가족관계	친인척관계	사회적 친분관계
과거응답	0.490*** [0.016]	0.405*** [0.018]	0.456*** [0.018]	0.453*** [0.018]	0.437*** [0.018]	0.407*** [0.018]
과거응답* 웹조사	0.146 [0.206]	0.109 [0.156]	-0.020 [0.142]	0.010 [0.126]	-0.409 [0.265]	-0.378 [0.248]
웹조사	-0.316 [0.553]	-0.138 [0.453]	0.313 [0.332]	-0.020 [0.310]	1.256* [0.718]	1.091 [0.667]
관측치 수	3,859	3,859	3,859	3,859	3,859	3,859
R-squared	0.316	0.268	0.266	0.263	0.269	0.263
	직장 만족도(1: 전혀 그렇지 않다~5: 아주 그렇다)					
	직장만족도 1 좋은 직장이다	직장만족도 2 기쁘게 다닌다	직장만족도 3 추천하고 싶다	직장만족도 4 자랑 할 수 있다		
과거 응답	0.440*** [0.028]	0.453*** [0.023]	0.479*** [0.025]	0.504*** [0.029]		
과거응답* 웹조사	0.076 [0.508]	-0.062 [0.348]	0.025 [0.420]	0.250* [0.132]		
웹조사	-0.111 [1.583]	0.323 [1.136]	0.147 [1.421]	-0.663 [0.541]		
관측치 수	1,390	1,390	1,390	1,390		
R-squared	0.338	0.322	0.357	0.358		

주: 1) 상세한 설문 내용은 <표 5-10> 참고.

생활만족도 문항: 매우만족(1), 만족(2), 보통(3), 불만족(4), 매우불만족(5);

직장만족도 문항: 전혀 그렇지 않다(1), 그렇지 않은 편이다(2), 보통이다(3), 그런 편이다(4), 아주 그렇다(5).

2) [] 안은 표준편차 값임.

3) * 10% 유의수준, ** 5% 유의수준, *** 1% 유의수준.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

마지막으로 <표 5-17>은 생활만족도 및 일자리 만족도 응답 분석 결과를 보여준다. 대부분의 경우 주요 추정계수가 통계적으로 유의한 패턴을 보이지 않는다. 그러나 웹조사 더미변수의 계수 부호가 앞서 <표

5-13>에서 추정된 전화조사 영향 추정계수와 부호가 일치한다는 점에서 전화조사와 웹조사가 대면조사 대비 응답행태에 유사한 영향을 줄 수 있다는 해석을 가능하게 한다.

4. 결과 요약

분석 결과는 다음과 같다. 첫째, 주관적인 판단이나 기준을 요하는 설문 문항 응답 경향이 상대적으로 조사방식에 민감하게 영향을 받는 것으로 보인다. 건강상태 자가 진단이나 생활 만족도 문항의 경우 조사방식에 따라 응답 경향이 달라지는 반면, 상대적으로 객관적인 판단이 가능한 근로 활동 제약 여부 등에 관한 문항들은 조사방식에 따라 응답 경향이 뚜렷한 차이를 보이지 않는다. 이는 주관적인 내용에 관한 설문 문항은 주로 응답 시간이 더 소요되거나 사고 및 판단이 요구되기 때문에 응답 품질이 조사방식에 민감하게 영향을 받을 가능성이 있다. 둘째, 전화조사와 웹조사가 응답행태에 미치는 영향이 유사한 것으로 보인다. 이러한 패턴은 주관적인 생활만족도나 건강 의견을 묻는 문항들에서 주로 관찰된다. 웹조사의 경우 자료의 한계로 그 패턴이 뚜렷하지는 않으나 전반적으로 추정계수 부호가 전화조사 분석과 일치하는 등 이를 뒷받침하는 결과들을 얻었다. 셋째, 전화조사 방식이 중간값이나 특정값 혹은 이상적인 답변을 선택하는 경향을 유도한다는 가설을 뒷받침할 만한 패턴은 관찰되지 않는다. 마지막으로 본 분석에서는 민감문항으로 흡연 여부와 음주 여부를 살펴보았는데 해당 문항 응답 행태에 전화조사 영향은 관찰되지 않은 반면 웹조사 영향은 통계적으로 유의하고 추정계수도 상대적으로 크게 나타난바, 조사원과 직접 소통하지 않는 웹조사 방식이 민감문항 설문에 유용할 가능성이 있다. 그러나 웹조사 분석의 한계를 고려하여 결과 해석에 주의를 요한다.

제5절 소 결

한국노동패널 조사는 1998년부터 2021년 현재까지 24년 동안 면접조사 방법을 주요 조사도구로 이용하여 패널 조사를 실시하였다. 그러나 2020년(23차) 코로나19의 영향으로, 대면조사를 기피하는 응답자들이 발생하면서 조사 거절 사례가 많아져, 패널조사에서 가장 중요한 요소인 조사성공 유지율이 급격하게 떨어질 위기에 처하게 되었다. 이에 노동패널팀에서는 조사성공률을 높이기 위해, 면접조사가 불가능한 경우에 한해, 보조적으로 사용되던 전화조사방법을 대면조사를 기피하는 패널들에게 적극적으로 권유하였다. 전화조사는 패널가구를 관리하는 면접원이 직접 패널에게 전화하여 문항을 설명하고, 응답자가 대답하면, 면접원이 CAPI 시스템에 기입하는 방식으로 이루어지기 때문에, 대면조사와 거의 유사한 조사방법이라고 판단하였다. 조사 결과, 매년 3% 이내에서 이루어졌던 전화조사 응답비율이 18%까지 급증하였고, 매년 95% 이상을 차지하던 면접조사 비율은 80%까지 급감하였다.

또한 노동패널조사는 2019년도(22차)부터 이사추적 실패 등의 이유로 방문할 수 없는 패널들을 추적 조사하기 위해, 휴대전화번호와 이메일을 이용한 웹조사를 시범 도입하였고, 23차에도 연속적으로 진행하였다. 웹조사방법은 가구 조사에 성공한 가구에 속한 응답하지 않은 가구원을 대상으로 실시하였으며, 그 결과 2019년(22차) 26명, 2020년(23차) 73명의 개인이 응답하였다.

본 연구는 대면조사방법을 기본으로 조사되었던 노동패널조사에 23차에는 전화조사방법과 웹조사방법이 혼재되면서, 조사 도구에 따른 조사 결과의 신뢰성 검토 필요성 차원에서 수행되었다.

우선 웹조사 개인 응답자 특성은, 22차 웹조사 응답자 26명 중, 23차에도 웹조사로 응답한 인원은 6명(23.1%)인 반면, 면접조사로 응답한 인원은 8명(30.8%)이었다. 이는 웹조사로 응답한 이후에는 웹조사에 대한 선호도가 높아져 다시 면접조사로 회귀하기 어려울 것이라는 우려와는 다

른 결과였다. 따라서 추후에도 개인 조사에 응답하지 않은 가구원에게 대안적인 조사방법으로 웹조사를 권유하는 것은, 면접조사를 기본으로 하는 노동패널의 기본방침에 큰 혼란을 초래하지 않을 것이라 조심스럽게 추측해 볼 수 있다.

조사방법별 응답 가구의 특성을 살펴보면, 코로나19의 영향으로 자녀가 있는 경우, 대면조사보다 전화조사를 선호할 것이라는 예측대로 전화조사 응답가구에서 0~15세 자녀가 더 많이 발견되었다. 가구소득, 거주지역, 주거형태 모두 조사방법에 따라 차이가 있는 것으로 나타났다. 전화조사 응답가구의 가구소득이 더 높고, 거주지역이 수도권이 더 많고, 주거형태는 면접조사보다 자가의 비율이 낮았다.

조사방법별 개인 응답자 특성을 살펴보았는데, 성별, 연령, 결혼상태, 경제활동상태, 월평균 소득 모든 항목에서 면접조사, 전화조사, 웹조사의 응답자 특성이 통계적으로 차이가 있는 것으로 나타났다. 평균연령은 면접조사가 53.4세, 전화조사 46.2세, 웹조사 32.7세로 큰 차이를 보였고, 학력은 면접조사는 중졸, 고졸 순으로 많은 반면, 전화조사는 고졸, 대졸 이상 순으로, 웹조사는 대졸 이상, 전문대졸 순으로 많았다. 경제활동상태는 면접조사는 미취업, 임금근로자, 비임금 근로자 순으로 많고, 전화조사와 웹조사는 임금근로자, 미취업자, 비임금 근로자 순으로 많았다. 월평균 소득은 전화조사 308.9만 원, 면접조사 269.7만 원, 웹조사 248.2만 원이었다. 이상의 분석들로, 조사도구별로 응답 집단의 특성이 상이한 것으로 추측되므로, 조사도구별로 개별 분석하는 것은 지양해야 한다.

신뢰할 수 있는 패널자료는 한 개인이 주변 여건이 변화되지 않았다면, 연도별로 안정적으로 동일하게 응답되어야 한다. 22차에 면접조사로 응답하였던 개인들이, 23차에 전화조사로 응답하였을 때, 혹은 웹조사로 응답하였을 때, 조사도구별로 응답값에 차이가 있는지, 조사결과의 신뢰 정도를 주요 변수별로 분석해 보았다. 다만 웹조사는 응답 사례 수가 적어 결과 해석에 주의해야 한다.

분석변수는 개인의 소득 및 일자리 관련 변수(월평균 임금액수, 월평균 소득액수(비임금), 정규직 여부, 주당 근무시간, 근로계약기간 존재 여부, 근로시간 규칙성 여부)와 건강 및 생활 습관 관련 변수(현재 건강상태, 보

통 사람 대비 건강상태, 감각기관이나 육체적 제약 및 활동 제약 여부, 흡연 여부, 음주 여부), 생활 만족도, 직장 만족도 등이다.

〈표 5-18〉 분석 결과 요약 - 개인 소득 및 일자리 관련 변수

	전화조사 및 대면조사			웹조사 및 대면조사		
	과거응답	교차항	전화조사	과거응답	교차항	웹조사
월평균 임금	○(+)	×	×	○(+)	×	×
월평균 소득(비임금)	○(+)	×	×	○(+)	○(-)	○(+)
정규직 여부	○(+)	×	×	○(+)	○(+)	○(-)
주당 근무시간(비임금)	○(+)	×	×	○(+)	○(-)	○(+)
근로계약 기간 유무	○(+)	×	□(-)	○(+)	×	×
근로시간 규칙성	○(+)	○(-)	○(+)	○(+)	×	×

주: 1) 통계적 유의성 ○:1%, △:5%, □:10%, ×: 유효하지 않음.

2) (+), (-)는 방법론에서 사용된 계수의 부호로, 과거응답은 β_1 , 교차항은 β_2 , 전화조사(웹조사)는 β_3 의 부호임.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

〈표 5-19〉 분석 결과 요약 - 개인 건강 및 생활습관 관련 변수

	전화조사 및 대면조사			웹조사 및 대면조사			
	과거응답	교차항	전화조사	과거응답	교차항	웹조사	
현재 건강상태	○(+)	○(-)	○(+)	○(+)	×	×	
보통사람 대비 건강상태	○(+)	○(-)	○(+)	○(+)	×	×	
감각기관 장애	○(+)	×	×	○(+)	×	×	
육체적 제약	○(+)	×	×	○(+)	○(-)	×	
흡연 여부	○(+)	×	×	○(+)	×	□(+)	
음주 여부	○(+)	×	×	○(+)	○(-)	○(+)	
활동 제약	1. 배우기, 집중	○(+)	×	×	○(+)	×	×
	2. 옷입기 등	○(+)	×	×	○(+)	○(-)	×
	3. 쇼핑, 병원가기	○(+)	×	×	○(+)	×	×
	4. 직업활동 등	○(+)	×	×	○(+)	○(-)	□(+)

주: 1) 통계적 유의성 ○:1%, △:5%, □:10%, ×: 유효하지 않음.

2) (+), (-)는 방법론에서 사용된 계수의 부호로, 과거응답은 β_1 , 교차항은 β_2 , 전화조사(웹조사)는 β_3 의 부호임.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

개인 소득 및 일자리 관련 변수 분석 결과, 전화조사와 웹조사 모두, 모든 분석 변수에서 과거 22차 응답과 현재 23차 응답이 통계적으로 유의한 상관관계가 있는 것으로 나타났다. 교차항은 전화조사와 대면조사 분석에서, 근로시간 규칙성만 유의하게 추정되었고, 웹조사와 대면조사 분석에서는 월평균 소득(비임금), 정규직 여부, 주당 근무시간(비임금)이 유의하게 추정되었다(표 5-18 참조).

<표 5-19>는 개인 건강 및 생활습관 관련 변수 분석 결과이다. 전화조사 및 대면조사 분석에서는 현재 건강상태, 보통사람 대비 건강상태가 유의하게 나타났다. 웹조사 및 대면조사 분석에서는 육체적 제약, 음주 여부, 흡연 여부, 활동제약 4(직업활동 등)가 유의미하게 추정되었다.

<표 5-20>은 생활만족도와 직장만족도 변수 분석 결과를 요약한 것으로, 전화조사 및 대면조사 분석에서는 생활만족도 중 친인척관계, 사회적 친분관계가 유의미한 결과로 나타났고, 웹조사 및 대면조사 분석에서는 친인척 관계에서만 유의미하게 추정되었다.

<표 5-20> 분석 결과 요약 - 생활만족도 및 직장만족도 관련 변수

		전화조사 및 대면조사			웹조사 및 대면조사		
		과거응답	교차항	전화조사	과거응답	교차항	웹조사
생활 만족 도	1. 가족의 수입	O(+)	×	×	O(+)	×	×
	2. 여가활동	O(+)	×	×	O(+)	×	×
	3. 주거환경	O(+)	×	×	O(+)	×	×
	4. 가족관계	O(+)	×	×	O(+)	×	×
	5. 친인척관계	O(+)	□(-)	□(+)	O(+)	×	□(+)
	6. 사회적 친분관계	O(+)	△(-)	△(+)	O(+)	×	×
직장 만족 도	1. 좋은 직장이다	O(+)	×	×	O(+)	×	×
	2. 기쁘게 다닌다	O(+)	×	×	O(+)	×	×
	3. 추천하고 싶다	O(+)	×	×	O(+)	×	×
	4. 자랑할 수 있다	O(+)	×	×	O(+)	□(+)	×

주: 1) 통계적 유의성 O:1%, △:5%, □:10%, ×: 유효하지 않음.

2) (+), (-)는 방법론에서 사용된 계수의 부호로, 과거응답은 β_1 , 교차항은 β_3 , 전화조사(웹조사)는 β_2 의 부호임.

자료: 한국노동연구원, 「한국노동패널」 23차 조사(2020년도) 학술대회용 자료.

종합적으로 볼 때, 첫째, 면접조사와 전화조사, 웹조사를 응답하는 집단의 특징은 차이가 있으므로, 조사도구별 개별 분석은 지양해야 한다. 둘째, 과거 웹조사로 응답한 경험이 있는 응답자가 다음해에 웹조사만을 고집하는 경향이 크지 않으므로, 추후에도 응답거절자에 대하여 보조적 수단으로 웹조사를 활용하는 것은 큰 무리가 없어 보인다. 셋째, 패널 조사의 신뢰성 여부, 즉 개인의 외부 환경이 변하지 않았다면, 조사도구의 변화에도 매해 응답값의 차이가 없어야 하는 측면에서는 일부 변수에서 통계적으로 유의한 차이가 발생하였다. 전화조사와 웹조사 모두 상대적으로 객관적인 판단이 가능한 문항에서는 조사 도구에 따라 응답 경향에 뚜렷한 차이를 보이지 않았다. 다만 주관적 판단 문항 중 현재 건강상태, 보통사람 대비 건강상태, 음주 여부, 흡연 여부 등은 조사도구에 따른 유의한 차이를 보였다. 특히 이들 변수는 노동패널 웹조사 실험설계³⁹⁾와 22차 웹조사 분석⁴⁰⁾에서도 차이를 보였던 문항으로 사용에 각별한 주의가 필요하다.

2021년도에도 코로나19의 영향이 계속되어, 전화조사 응답비율이 높을 것으로 예상된다. 이에 조사방법에 따른 차이 분석 연구가 계속되어야 할 것이며, 그 차이를 극복할 수 있는 후속 연구가 필요하다.

39) 2017년도 실험조사의 결과는 김유빈 외(2017), 『패널자료 품질개선 연구(VII)』 제7장 「웹을 활용한 조사의 효과에 대한 실증연구: 노동패널 CAWI 실험데이터의 분석」(이지은·신선옥·이혜정·김기홍 저술)을 참고하기 바란다.

2018년도 실험조사의 결과는 김유빈 외(2018), 『패널자료 품질개선 연구(VIII)』 제3장 「웹을 활용한 조사의 효과에 대한 실증연구Ⅱ: 노동패널 CAWI 1차, 2차 실험 연구 결과 분석」(이지은·신선옥·이혜정·정현상 저술)을 참고하기 바란다.

40) 2020년도 웹조사 시범도입의 결과는 김유빈 외(2020), 『패널자료 품질개선 연구(X)』 제5장 「웹을 활용한 조사 자료 검증: 노동패널자료를 중심으로」(이지은·채민희 저술)을 참고하기 바란다.

제 6 장

Web 조사방식이 응답률과 자료의 질에 미치는 영향 : 사업체패널조사를 중심으로

제1절 서론

사업체패널조사는 2005년에 컴퓨터를 활용한 조사(Computer Assisted Personal Interviewing : CAPI)를 도입한 후, 1차 조사(WPS2005)부터 6차 조사(WPS2015)까지 CAPI 방식으로 조사를 수행하였다. 6차 조사에서는 신규표본의 추가⁴¹⁾로 인해 원패널뿐 아니라 신규패널에 대한 조사를 동시에 진행하면서 과거에 비해 조사 완료까지 오랜 기간이 소요되었다.⁴²⁾ 조사기간이 길어지면 설문지의 응답 기준시점⁴³⁾과 실제 응답시점 간의 차이가 커져 측정 오차를 발생시킬 가능성이 높기 때문에⁴⁴⁾ 7차 조사

41) 사업체패널조사가 시행된 지 10년이 지나면서 모집단 변동과 누적된 패널 이탈 등의 문제를 해결하고자 6차 조사부터 신규표본 약 2,000여 개를 추가하였다.

42) 1~5차 조사기간은 일반적으로 적게는 5개월, 많게는 7개월 정도 소요되었다. 6차 조사에서는 늘어난 표본만큼 컨택원과 면접원을 충분히 투입하고 과거와 유사한 시기에 조사를 시작했음에도 조사 기간이 9개월이 걸려, 최대 길었던 조사기간보다도 약 2개월 정도 더 소요되었다.

43) 사업체패널조사는 당해에 전년도 12월 말 혹은 지난 2년간을 기준으로 조사한다.

44) 1~5차 조사와 6차 조사의 응답결과는 다소 차이가 있는데, 주된 원인은 모집단 변화로 인한 것으로 판단되나 인사담당자의 응답 부담 증가와 조사기간이 길어짐에 따른 설문지의 기준시점과 실제 응답시점 간의 차이 증가 또한 무시할 수 없는 이유로 보인다.

(WPS2017)부터는 조사 기간을 단축시키고자 분량이 짧은 무노조 사업장의 근로자대표 설문(35문항)과 CAPI 조사로 전체 설문을 완료하지 못한 경우, 그리고 방문조사에 심한 거부감을 드러내는 일부 사업체에 대해 Web 조사방식을 도입했다. 중단 성격을 갖는 대규모의 사업체 단위 조사인 사업체패널조사에 Web 방식⁴⁵⁾을 도입하게 되면, 컨택부터 설문 시작까지의 시간 단축⁴⁶⁾, 비용 절감⁴⁷⁾, 면접원에 의해 발생하는 오차의 감소⁴⁸⁾ 등과 같은 장점이 있으나 무응답률 혹은 항목무응답률의 증가⁴⁹⁾, CAPI와 CAWI를 동시에 활용함에 따른 모드(mode)에 의한 오차 발생⁵⁰⁾ 등과 같은 단점도 존재한다.

만약 CAWI 조사를 실시해도 어느 정도 유의미한 응답률을 보이고 CAPI와 CAWI의 응답결과 또한 체계적인 차이를 보이지 않으면, CAWI 방식의 활용은 설문의 양이 많으며 현장방문이 쉽지 않을 수 있고 예산

45) Web 조사는 일반적으로 비용 절감, 시간 단축, 설문 구성의 유연성, 면접원에 의한 오차 감소 등의 장점과 적용대상(coverage)의 편향, 낮은 응답률 등의 단점이 있다고 알려져 있다(Symth and Pearson, 2011). 그러나 이러한 특징은 조사의 구조, 목적 및 대상 등에 따라 차이가 날 수 있다.

46) 사업체패널조사는 이미 Web을 기반으로 한 CAPI 시스템을 활용하고 있어 skip 문항이나 오류 검증 등은 자동으로 처리되고 조사 후 곧장 서버에 자료가 축적되기 때문에 이로 인한 시간 단축은 발생하지 않는다. 다만, Web조사는 컨택 후 e-mail을 보내 곧장 응답을 유도하기 때문에 컨택 후 면접원의 사업체 방문까지의 기간은 줄일 수 있다. 실제로 컨택 후 방문까지의 기간이 상당히 길거나 방문 약속의 변경으로 인해 이 기간이 길어지는 경우는 적지 않게 발생한다.

47) 사업체패널조사는 근로자 현황을 종이설문으로 진행하여 별도의 자료 처리 과정을 거치는데, CAWI로 조사하는 사업체는 이 부분 또한 자동처리되어 종이설문 인쇄비와 자료 입력 비용 등이 절감된다.

48) 면접원이 응답자의 속도를 좌우하거나 응답자에게 불명확하게 용어를 설명하는 경우 응답자의 응답 편의를 일으킬 가능성이 있다.

49) Web 조사는 다른 방식의 조사(우편, 전화, 방문 등)에 비해 응답률은 낮고, 항목 무응답의 비율이 높다고 알려져 있다(이지은 외, 2017; Fricker, Galesic, Tourangeau and Yan, 2005; de Leeus, 2005; Lozar et al., 2008; Shih and Fan, 2008; Kirchner and Felderer, 2017; Daikeler et al., 2020).

50) 같은 사람에게 동일한 설문을 하는 경우에도 모드(mode)의 차이로 인해 때때로 다른 결과가 발생되기도 한다(de Leeuw, 1992; Dillman et al., 1996; Dillman, 2000; Michael and Ali, 2006). 사업체패널조사는 CAPI와 CAWI를 동일한 디자인의 Web 기반 설문을 사용하지만, 응답자의 컴퓨터 사용 환경이나 선택한 웹 브라우저의 차이로 인해 시각적 효과나 속도 등에서 모드 효과가 나타날 수 있다.

제약이 심한 사업체패널조사에 상당한 도움이 될 것이다. 즉 대면조사를 하기 어렵거나 한 번에 설문을 마치지 못한 경우 등의 상황에서 Web은 효과적일 수 있다. 최근 진행된 8차 조사(WPS2019)는 2020년에 진행되었는데 상당수의 응답자들이 Covid-19로 인해 비대면 방식의 조사를 요구하여 CAWI 방식을 유용하게 활용할 수 있었고, 이로 인해 CAWI 조사의 효과를 실증분석할 수 있는 기회가 주어졌다. 즉, 7차 조사에서는 심한 거절로 방문을 거부하는 응답자만 전체 설문을 CAWI로 조사하였으나⁵¹⁾ 8차 조사에서는 CAPI 조사와 동일한 컨택(혹은 응답) 상태에서 CAWI로 조사한 사업체가 있어, 유사한 조건에서 CAPI와 CAWI 조사의 효과를 비교할 수 있게 되었다. 본 연구는 8차 조사에 나타난 CAPI와 CAWI의 응답률, 응답의 질(quality) 혹은 응답 성향(propensity) 등을 비교하여 Web 조사방식이 응답자의 사업체패널조사 참여를 저해하는지, 조사방식에 따라 응답의 내용에 체계적인 차이가 존재하는지 등에 대해 검증하고자 한다. 이를 통해 사업체패널조사에서 CAWI 방식의 향후 활용 수준에 대해 가늠해 볼 수 있을 것으로 기대한다.

본 장은 다음과 같이 구성된다. 먼저 제2절에서는 사업체패널의 CAWI 조사 현황 및 응답 특성에 대해 탐색적으로 살펴본다. 제3절은 조사방식이 조사성공률 및 자료의 질에 미치는 영향에 대해 통계 모형을 활용하여 분석한다. 마지막으로 제4절에서는 지금까지 분석한 결과를 요약하고 이상의 분석 결과가 향후 조사방식 결정에 어떠한 함의를 제공하는지에 대해 논의한다.

제2절 CAWI 활용 현황과 응답 특성

본 절은 사업체패널 8차 조사의 CAWI 조사 활용 현황과 응답 특성을

51) 거절이 심한 경우 불성실한 응답을 할 가능성이 높기 때문에, CAPI와 CAWI의 7차 조사 응답 결과를 단순비교하게 되면 CAWI가 CAPI에 비해 응답률은 매우 낮고 응답의 질 저하를 유발한다는 편향적인 결론이 나올 수 있다.

CAPI와 비교하여 살펴보겠다. 먼저, 8차 조사의 CAWI 조사 활용 현황을 살펴보고, 그 사유를 분석하겠다. 다음으로 CAWI 조사의 응답률(조사성공률, 응답기간 등), 응답 성향(응답자 특성, 사업체 특성 등), 응답의 질(자료수정률, 항목무응답률 등)을 CAPI 조사와 비교·분석하고자 한다.

1. CAWI 조사의 활용 현황

사업체패널 8차 조사에서 활용된 조사방식을 살펴보겠다. 8차 조사에서 조사에 성공한 사업체를 조사방식별로 보면, CAPI가 1,446개(51.7%), CAPI와 CAWI를 혼합한 방식(CAPI+CAWI)이 553개(19.8%), CAWI가 796개(28.5%)였다(표 6-1 참조). 즉 Web 방식으로 전체 설문을 조사한 사업체 수는 796개로, 전체 조사된 사업체 중 28.5%를 차지한다고 나타나 그 비중이 적지 않음이 확인되었다.⁵²⁾

지금부터는 8차 조사에서 Web으로 조사한 사유에 대해 자세히 살펴보겠다(표 6-2 참조). 먼저, CAPI와 CAWI를 혼합한 방식으로 조사한 사업체의 Web 활용 사유를 보면, ‘참고자료를 미리 확인하지 못해 즉각적인 응답이 어려워서’가 21.7%로 가장 많았고, 그다음으로 ‘응답시간이 길어져서’(18.5%), ‘응답 중간에 급한 일이 생겨서(0.7%)’ 순으로 나타났다. 대부분 방문조사 후 남은 부분을 Web으로 응답한 경우로, 이는 7차 조사

〈표 6-1〉 8차 조사(WPS2019)의 조사방식별 사업체 현황

(단위: 개, %)

	사업체 수	비율
CAPI	1,446	51.7
CAPI+CAWI (혼합조사)	553	19.8
CAWI	796	28.5
전 체	2,795	100.0

52) 사업체패널 7차 조사에서도 Web 방식을 활용하였는데, 그 비율은 48.8% 정도로 나타났다. 그러나 전체 설문을 Web으로 응답한 사례 수는 적고 대다수 CAPI 조사 후 남은 일부 설문에 한해서 실시되었다.

때와 동일하였다.

다음으로, 설문 전체를 CAWI 조사로 응답한 사유를 보면, ‘Covid-19로 인해 사업체에서 비대면 조사를 요청해서’가 34.2%로 가장 많았고, 그다음으로 ‘회사 내에서 방문조사하는 것이 눈치 보여서(10.2%)’, ‘조사에 대한 심한 거절로 방문이 불가하여 Web 외의 다른 방식을 활용할 수 없어서’(7.3%), ‘회사 내 외부인 출입 금지 규정으로 인해서’(7.2%) 순으로 나타났다. Web을 활용한 사유를 요약해보면, 응답자의 의지에 따라 자발적으로 CAWI 조사로 응답한 경우와 그렇지 않은 경우로 나뉘질 수 있다. 동일한 CAWI 조사를 하였더라도 응답자의 의지에 따라 수집된 자료는 다른 응답 특성을 가질 수 있다. 거절이 심해 Web으로 응답한 경우 불성실한 응답을 했을 가능성이 높는데, 이 경우까지 CAWI로 포함하여 CAPI와 비교한다면 CAWI가 응답률과 응답의 질이 모두 낮다는 편향적인 결과가 나타날 수 있어 향후 CAWI 조사의 모드 효과를 분석할 때 이를 구별하여 분석하겠다. 이때, Web 조사를 자발적으로 요청한 경우는 ‘자발 CAWI’, 그렇지 않은 경우는 ‘비자발 CAWI’라 일컫겠다.

<표 6-3>은 조사방식에 따른 인사담당자 설문 의 첫 문항과 마지막 문항의 응답월별 사업체 수와 비율을 보여준다. CAPI, CAPI+CAWI, 자발 CAWI 조사는 월별 첫 문항과 마지막 문항의 응답 사업체 비율이 큰 차이를 보이지 않았으나(즉, 매월 꾸준히 조사가 시작되고 완료되고 있

<표 6-2> 8차 조사(WPS2019)의 Web 조사 응답 사유

(단위: 개, %)

사유		사업체 수	비율	
CAPI+CAWI	참고자료를 미리 확인하지 못해서	293	21.7	
	응답시간이 길어져서	250	18.5	
	급한 일이 생겨서	10	0.7	
CAWI	자발	Covid-19로 인해	462	34.2
		회사 내 방문조사를 진행하는 것이 눈치 보여	138	10.2
		회사의 외부인 출입 금지 규정으로 인해	97	7.2
	비자발	조사 거절로 인해	99	7.3
전체		1,349	100.0	

음), 비자발 CAWI는 2020년 7월부터 12월까지 매월 10개 내외 사업체가 첫 문항을 응답하였으나 마지막 문항 응답은 모두 2021년 1월에 이루어졌다. 이를 통해 비자발적으로 Web을 활용하여 조사된 사업체는 조사기간이 종료되는 시기에 간신히 응답이 완료된 것으로 보인다. 이와 같은 상황에서는 응답자가 불성실하게 답변했을 가능성이 있기 때문에 비자발 CAWI 조사는 다른 조사방식과 다른 특성을 가질 수 있음이 짐작된다.

〈표 6-3〉 8차 조사(WPS2019) 조사방식별 첫 문항과 마지막 문항의 응답시기
(단위: 개, %)

		2020						2021	전체
		7월	8월	9월	10월	11월	12월	1월	
첫 문항 기준	CAPI	24 (1.7)	146 (10.1)	180 (12.4)	234 (16.2)	228 (15.8)	443 (30.6)	191 (13.2)	1,446 (100.0)
	CAPI+CAWI	2 (0.4)	85 (15.4)	63 (11.4)	49 (8.9)	54 (9.8)	142 (25.7)	158 (28.6)	553 (100.0)
	자발 CAWI	15 (2.2)	128 (18.4)	85 (12.2)	97 (13.9)	84 (12.1)	222 (31.9)	66 (9.5)	697 (100.0)
	비자발 CAWI	3 (3.0)	9 (9.1)	5 (5.1)	10 (10.1)	6 (6.1)	8 (8.1)	58 (58.6)	99 (100.0)
	전 체	44 (1.6)	368 (13.2)	333 (11.9)	390 (14.0)	372 (13.3)	815 (29.2)	473 (16.9)	2,795 (100.0)
마지막 문항 기준	CAPI	9 (0.6)	62 (4.3)	192 (13.3)	260 (18.0)	245 (16.9)	325 (22.5)	353 (24.4)	1,446 (100.0)
	CAPI+CAWI	2 (0.4)	47 (8.5)	40 (7.2)	54 (9.8)	61 (11.0)	117 (21.2)	232 (42.0)	553 (100.0)
	자발 CAWI	15 (2.2)	72 (10.3)	63 (9.0)	96 (13.8)	113 (16.2)	219 (31.4)	119 (17.1)	697 (100.0)
	비자발 CAWI	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	99 (100)	99 (100.0)
	전 체	26 (0.9)	181 (6.5)	295 (10.6)	410 (14.7)	419 (15.0)	661 (23.6)	803 (28.7)	2,795 (100.0)

주: 1) 응답시기 비율 = 해당 월 응답 사업체 수/(해당 방식의 전체 사업체 수)×100
2) () 안은 비율임.

지금까지 사업체패널 8차 조사의 조사방식별 응답 현황과 Web을 활용한 경우 그 사유에 대해 살펴보았다. 8차 조사는 과거 차수의 조사와 달리 Covid-19 등으로 방문조사가 어려워지면서 응답자들이 자발적으로 Web 조사를 요청하였고, 결과적으로 4개 사업체 중 1개는 자발 CAWI 조사로 전체 설문을 완료하였다. 이 자료는 CAPI와 CAWI의 모드 효과를 검증하는 데 유용할 것이다. 즉, 조사 상황의 차이를 고려하지 않고 CAPI와 CAWI 조사를 비교할 수 있으므로 모드 효과를 살펴보는 데 효과적일 것으로 기대된다.

2. CAWI 조사의 응답 특성

이제 8차 조사에서 CAPI와 CAWI의 모드 효과(mode effect)에 대해 살펴보겠다. CAWI와 CAPI 조사방식을 비교·분석할 때 비자발 CAWI를 포함하게 되면 해당 조사방식의 특이점에 의해 잘못된 결과가 도출될 수 있으므로, 본 분석에서는 CAPI와 자발 CAWI로 조사가 완료된 사업체만을 활용하겠다. 모드 효과(mode effect)는 조사방식(면접, 전화, 우편, Web 조사 등)의 변화로 인해 발생하는 표본이탈·무응답 오차·측정오차 등을 종합적으로 말하는데(이상호, 2012: 27), 여기서는 무응답 오차와 측정 오차를 중심으로 CAPI와 CAWI의 응답 특성을 분석하겠다.

가. 응답률

저조한 응답률이 반드시 높은 무응답 오차⁵³⁾로 이어지는 것은 아니지만 (Keeter, Miller, Kohut, Groves and Presser, 2000; Groves and Peytcheva, 2008), 일반적으로 응답률이 높으면 수집된 데이터의 대표성이 보장되며 무응답 편향에 대한 우려를 줄일 수 있다고 알려져 있다(Rogelberg and Stanton, 2007).

보통의 경우 조사방식별 응답률을 산출하여 비교할 때 무작위로 조사

53) 무응답 오차란 무응답자가 응답을 했을 때 지금 조사된 결과와 다른 답변을 할 경우 발생한다.

방식을 배정하는 등의 실험설계가 이루어지지만 사업체패널 8차 조사는 이렇게 설계되지는 않았다. 따라서 사후적으로 조사된 결과에 따라 조사 방식별 응답률을 계산하기 때문에 다소 한계가 존재한다. 즉 응답자 혹은 사업체의 특성에 따라 응답률의 차이가 나타날 수 있는데 이 특성이 조사 방식의 선택에도 영향을 미칠 수 있음을 유념해야 한다. 응답자 특성에 따라 응답률 간의 차이가 있음을 발견한 연구들이 있다(Anseel et al., 2010; Baruch, 1999).

본 연구에서는 응답률을 두 변수를 통해 살펴보겠다. 하나는 조사성공률이고 다른 하나는 조사기간이다. 조사성공률은 말 그대로 해당 조사방식으로 조사를 시작한 사업체 중 조사가 완료된 사업체의 비율이다. 조사기간은 컨택부터 인사담당자 설문⁵⁴⁾의 마지막 문항 완료까지의 기간인데, 조사기간이 길어질수록 조사완료 확률이 감소하기 때문에 이 또한 간접적인 응답률을 보여줄 것으로 판단하여 활용하였다.

먼저, 8차 조사의 조사방식별 조사성공률을 살펴보았다(표 6-4 참조). CAPI로 조사한 사업체의 조사성공률은 96.4%, 자발 CAWI는 84.9%로 나타났다. 자발 CAWI로 조사한 사업체의 조사성공률은 CAPI에 비해 11.5%포인트 낮았다.⁵⁵⁾ 이는 CAWI 조사의 조사성공률이 CAPI에 비해 낮은 것일 수 있지만, CAWI 조사를 선택한 응답자나 사업체의 특성에 의한 영향일 수도 있다.

〈표 6-4〉 8차 조사(WPS2019)의 조사방식별 조사성공률

(단위: 개, %)

	대상 사업체	완료 사업체	중단 사업체	조사성공률
CAPI	1,500	1,446	54	96.4
자발 CAWI	821	697	124	84.9

54) 본 연구에서는 인사담당자 설문에 대해서만 분석한다. 사업체패널조사에서 인사담당자가 응답하는 설문 분량은 노동조합 유무에 따라 다소 차이가 존재하나 전체 설문 분량 중 적게는 86.7%, 많게는 92.3%를 차지한다.

55) Web과 다른 조사방식 간의 응답률 차이에 대한 가장 대표적인 연구인 Lozar et al.(2008)은 Web 설문의 응답률이 다른 조사방식보다 평균 약 11%포인트 낮다고 보고하였다.

조사기간에 대한 분석에 앞서, 향후 분석에 대한 이해를 돕고자 사업체 패널의 조사구조와 조사진행 과정을 간략히 설명하겠다.⁵⁶⁾ 사업체패널의 조사구조는 크게 사전조사와 본조사로 나뉜다. 사전조사는 사업체의 근로자 현황에 대한 내용⁵⁷⁾인데, 이는 본조사의 설문과 연계되어 있어 반드시 본조사 전에 완료되어야 한다. 본조사는 인사담당자가 응답하는 설문과 근로자대표가 응답하는 설문으로 구성되며, 설문 내용은 <표 6-5>과 같다. 이와 같이 구조화된 사업체패널조사는 사전컨택, 근로자 현황조사표 작성, 본조사, 자료 검수 및 보완조사 등의 단계를 거쳐 진행된다.

이제 조사방식별 조사기간의 차이를 비교·분석하겠다(표 6-5 참조). 본 분석에서 총조사기간(Dur)은 첫 컨택부터 마지막 문항의 응답까지로 정의⁵⁸⁾하였는데, 이 기간은 다음과 같이 4가지로 분류할 수 있다. 첫 번째 기간(Dur1)은 사업체를 컨택하여 근로자 현황 조사 시작까지, 두 번째 기간(Dur2)은 근로자 현황 조사 완료까지, 세 번째 기간(Dur3)은 인사담당자 설문의 첫 문항 응답까지, 네 번째 기간(Dur4)은 인사담당자 설문의

<표 6-5> 사업체패널의 조사구조

사전 조사		EPQ. 근로자 현황
본조사	인사담당자 응답용 설문	AQ. 사업장 특성 BQ. 고용현황 및 고용관리 CQ. 보상 및 평가 DQ. 인적자원관리 및 작업조직 EQ. 인적자원개발 FQ. 기업복지와 안전보건 MQ. 유노조사업장의 노사관계 NQ. 무노조사업장의 노사관계
	근로자대표 응답용 설문	PQ. 유노조사업장의 노사관계 (노동조합 대표) RQ. 무노조사업장의 노사관계 (노사협의회 대표)

56) 자세한 사항은 사업체패널 홈페이지에 업로드된 유저가이드를 참고하라.

57) 사업체패널 1~6차 조사까지는 'FPQ. 재무현황'도 함께 조사하였으나, 통계청의 정기간질진단의 권고에 따라 7차 조사부터 나이스신용평가(주)의 정보로 대체하였다.

58) Re-check 조사 기간은 포함되지 않는다.

〈표 6-6〉 8차 조사(WPS2019)의 조사기간 현황

(단위: 일)

	CAPI	자발 CAWI	diff.	t-value
Duration: 총조사기간 (Dur1 → Dur4)	119.33	115.54	3.79	1.70**
• Dur1: 첫 컨택 → 근로자 현황 조사 시작	1.13	1.02	0.11	5.10***
• Dur2: 근로자 현황 조사 시작 → 근로자 현황 조사 완료	5.24	10.17	-4.93	-24.82***
• Dur3: 근로자 현황 조사 완료 → 설문 첫 문항 응답	103.68	90.50	13.18	5.66***
• Dur4: 설문 첫 문항 응답 → 마지막 응답 (응답기간)	9.28	13.85	-4.57	-5.61***
전체 사업체 수 (개)	1,446	697		

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

마지막 문항 응답까지로 분류하였다. 네 번째 기간(Dur4)은 총조사기간 중 실제 설문을 응답하는 기간이므로, 사업체패널처럼 많은 분량의 설문을 응답하는 경우 특히 이 기간(Dur4)이 길어질수록 이탈 확률이 높을 것이다.

<표 6-6>에 나타난 총조사기간(Duration)을 분석한 결과를 보면, CAPI는 119.3일이, 자발 CAWI는 115.5일이 소요되어, 자발 CAWI가 3.8일 정도 짧은 것으로 나타났다. 근로자 현황 조사 완료 후 인사담당자 첫 문항 응답까지의 기간(Dur3)은 자발 CAWI(90.5일)가 CAPI(103.7일)에 비해 13.2일 정도 짧은 것으로 나타나는데, 이 기간이 자발 CAWI의 총 조사기간을 상당히 감소시키는 데 큰 역할을 하는 것으로 보인다. Dur3은 응답자에게 설문을 즉각적으로 보내는 Web 조사의 장점이 잘 드러나는 구간이다. 응답률과 가장 상관성을 보일 것으로 생각되는 설문의 응답기간(Dur4)을 분석하면, CAPI는 9.3일, 자발 CAWI는 13.9일이 소요되어, 자발 CAWI가 약 4.6일 긴 것으로 나타났다. 이를 통해 자발 CAWI가 CAPI에 비해 응답기간(Dur4)이 긴 것을 확인하였다.

사업체패널조사는 면접원이 방문하여 CAPI 조사를 하는 경우 방문일에 조사를 완료하는 비율은 60%인 데 반해, 자발 CAWI 조사는 응답자가 설문의 첫 접속일에 조사를 완료하는 비율이 38%으로 나타났다. 또한, CAPI 조사의 경우 평균 2.8회 정도 방문을 하고 CAWI 조사는 자발적으

로 선택한 경우라도 평균 7회 정도의 독려 전화를 한다고 나타났다. 이러한 실사 과정의 기록을 통해서도 CAWI의 응답기간은 CAPI에 비해 길 것으로 유추된다.

지금까지 CAPI와 자발 CAWI의 응답률을 조사성공률과 응답기간(Dur4)을 통해 확인하였다. CAWI는 CAPI에 비해 조사성공률은 낮고 전체 설문 응답기간(Dur4)이 길게 나타나, CAWI가 CAPI 조사 방식에 비해 응답률이 더 낮을 것으로 추정된다.

나. 응답 성향

앞의 조사방식별 응답률에 대한 분석을 통해 자발 CAWI가 CAPI에 비해 응답률이 더 낮을 것으로 추정하였다. 그러나 사업체패널 8차 조사는 실험설계 방식으로 조사되지 않았기 때문에 응답률의 차이가 응답자 혹은 사업체의 특성 차이에 의한 것일 수 있다. 여기서는 자발 CAWI와 CAPI 조사의 응답 성향을 비교하였다. 사업체패널조사는 격년마다 동일한 사업체의 인사담당자에게 설문을 조사하는 방식으로 진행된다. 따라

〈표 6-7〉 8차 조사(WPS2019)의 조사방식별 응답자 특성

	CAPI	자발 CAWI	diff	t-value	
연령(세)	40.40	40.07	0.330	0.80	
남성 비율	0.590	0.528	0.062	2.72***	
직 급	최고경영자·임원	0.047	0.053	-0.006	-0.61
	부장·차장	0.200	0.316	-0.116	-5.95***
	과장	0.243	0.232	0.011	0.56
	대리 이하	0.510	0.399	0.111	4.84***
근속년수	9.32	9.67	-0.344	-1.14	
7차 조사와 동일응답자 ²⁾	0.682	0.687	-0.005	-0.21	
전체 사업체 수(개)	1,446	697			

주: 1) ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

2) 7차 조사 정보를 활용한 분석은 CAPI는 1,329개, 자발 CAWI는 579개 사업체를 대상으로 함.

서 사업체의 특성뿐 아니라 응답자(인사담당자)의 특성에 따라 선호하는 조사방식이 다를 수 있을 것이다. 특히, 사업체패널조사는 패널조사임에도 인사담당자가 바뀌면 응답자 특성이 변경된다는 독특한 특징을 갖고 있다.

먼저, 응답자 특성에 따라 선호하는 조사방식에 차이가 있는지 살펴보았다. <표 6-7>은 조사방식별 응답자 특성을 분석한 결과이다. 이에 따르면, 평균 연령, 최고경영자·임원·과장 비중, 평균 근속, 7차 조사와 동일 응답자의 비중은 조사방식에 따라 통계적으로 유의미한 차이를 보이지 않았다. 자발 CAWI는 CAPI에 비해 부장·차장급의 비중은 높고

<표 6-8> 8차 조사(WPS2019)의 조사방식별 사업체 특성

(단위: %)

		CAPI	자발 CAWI	diff	t-value	
규모	30~99인	0.511	0.515	-0.004	-0.17	
	100~299인	0.272	0.281	-0.009	-0.42	
	300~499인	0.099	0.096	0.003	0.20	
	500인 이상	0.118	0.108	0.010	0.68	
산업	제조업	경공업	0.107	0.110	-0.004	-0.28
		화학공업	0.098	0.079	0.019	1.45
		금속·자동차·운송	0.151	0.148	0.004	0.22
		전기·전자·정밀	0.075	0.090	-0.015	-1.20
	비제조업	건설업	0.013	0.014	-0.001	-0.23
		전기·가스·수도업	0.073	0.082	-0.008	-0.69
		개인서비스업	0.127	0.119	0.007	0.49
		유통서비스업	0.093	0.083	0.010	0.77
		사업서비스업	0.169	0.191	-0.021	-1.22
		사회서비스업	0.093	0.083	0.009	0.72
유노조 사업체		0.292	0.251	0.041	1.97**	
공공부문		0.024	0.040	-0.017	-2.16**	
서울·경기·인천		0.380	0.575	-0.195	-8.65***	
전체 사업체 수(개)		1,446	697			

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

남성, 대리 이하의 비중은 낮았는데, 이는 통계적으로 유의하였다.

다음은 조사방식별 사업체 특성 차이가 있는지 분석하였고, 그 결과는 <표 6-8>과 같다. 사업체 규모와 산업은 통계적으로 유의하지 않았다. 자발 CAWI로 응답한 사업체는 CAPI로 응답한 사업체에 비해 유노조 사업체의 비중은 낮고 공공부문과 수도권(서울, 경기, 인천)에 위치한 사업체의 비중은 높았으며, 이는 통계적으로 유의하였다.

지금까지 조사방식별 응답자와 사업체 특성에 대해 살펴보았는데, 응답자 특성 중 성별, 일부 직급(부장·차장, 대리 이하)과 사업체 특성 중 노동조합 유무, 공공부문 여부, 사업체의 수도권 위치 여부에 따라 CAWI와 CAPI에 대한 성향 차이가 있는 것으로 나타났다. 앞에서 CAWI는 CAPI에 비해 응답률이 더 낮은 것으로 추정하였는데, 응답 특성 또한 응답률에 영향을 미친 것일 수도 있다. 향후 통계모형을 활용한 분석을 통해 이에 대해 살펴보겠다.

다. 자료의 질

여기서는 조사방식별 자료의 질 차이가 있는지에 대해 자료수정률과 항목무응답률을 통해 알아보겠다. 사업체패널조사는 CAWI든 CAPI든 동일한 에디팅 정보가 탑재된 설문 시스템으로 조사를 하지만 모드 효과에 의해 오류 발생 정도가 차이날 수 있는데, 이를 위의 두 가지 정보(자료수정률, 항목무응답률)를 통해 살펴보겠다. 사업체패널조사의 인사담당자 데이터에 포함된 변수는 약 1,400여 개로 매우 많기 때문에 모든 변수에 대한 분석은 불가능하다. 또한 사업체패널의 자료수정률이나 항목무응답률은 전반적으로 매우 낮아 대부분의 변수가 분석에 적합하지 않다. 따라서 본 연구에 적합한 주요 변수를 선정하여 분석하였다.

먼저, 자료수정률에 대해 살펴보았다. 자료수정률은 에디팅(editing)이나 클리닝(cleaning) 과정을 전혀 거치지 않은 응답자가 설문을 완료하여 서버(server)에 처음 쌓인 자료(원자료)와 최종 배포 자료(제11회 사업체패널 학술대회 버전 데이터) 간에 자료 수정이 어느 정도 있었는지를 측정할 수 있다. 이때, 원자료에는 항목무응답으로 처리되었으나 Re-check

〈표 6-9〉 8차 조사(WPS2019)의 조사방식별 자료수정률

(단위 : %)

	CAPI		자발 CAWI		diff	t-value
	평균	SD	평균	SD		
단독/다수 사업장	0.022	0.147	0.024	0.154	-0.002	-0.33
조직유형	0.024	0.152	0.019	0.135	0.005	0.72
산업	0.004	0.064	0.013	0.113	-0.009	-2.28**
대졸 신입사원 초임	0.346	0.476	0.275	0.447	0.071	3.30***
신입부터 과장까지 소요 기간	0.027	0.162	0.032	0.175	-0.005	-0.60
교육훈련 전담부서 유무	0.003	0.059	0.007	0.084	-0.004	-1.18
안전보건관리 전담 부서 유무	0.008	0.091	0.013	0.113	-0.005	-1.02
사업장의 노조 존재 유무	0.013	0.114	0.014	0.119	-0.001	-0.23
단체교섭과 관련 과업 여부	0.008	0.091	0.007	0.084	0.001	0.27
본사/다른 사업장의 노조 존재 유무	0.016	0.125	0.023	0.150	-0.007	-1.14
노사협의회 존재 유무	0.002	0.046	0.003	0.054	-0.001	-0.36
노사협의회 개최 횟수	0.006	0.074	0.006	0.076	0.0002	-0.06
전체 사업체 수(개)	1,446		697			

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

조사를 통해 값이 입력된 경우는 자료수정으로 처리하였다. 자료수정률 분석 시 활용한 변수는 총 12개인데, 이는 오류가 발생하여 Re-check 조사를 실시하였던 변수를 중심으로 선정하였다.

12개 변수에 대해 원자료와 최종 배포 자료를 비교하여 조사방식별 평균 자료수정률을 계산하였고, 그 결과는 <표 6-9>와 같다. 자발 CAWI와 CAPI의 자료수정률을 보면, 8개 변수는 자발 CAWI의 자료수정률이, 4개 변수는 CAPI의 자료수정률이 높다고 나타났다. 이 중 2개 변수, 산업, 대졸 신입사원 초임만 CAPI와 자발 CAWI 간에 통계적으로 유의한 차이가 있다고 나타났다. 산업은 자발 CAWI가, 대졸 신입사원 초임은 CAPI가 자료수정률이 높았다. 그 외의 변수는 자발 CAWI와 CAPI 간에 자료수정률에서 유의한 차이를 보이지 않았다.

다음은 항목무응답률에 대해 분석하였다. 항목무응답은 설문 설계상 반드시 응답해야 하는 문항임에도 값이 결측된 경우를 말한다. 항목무응

<표 6-10> 8차 조사(WPS2019)의 조사방식별 항목무응답률

(단위 : %)

	CAPI		자발 CAWI		diff	t-value
	평균	SD	평균	SD		
하청업체 평균임금	0.008	0.091	0.009	0.092	-0.0003	-0.07
고용형태별/직종별 빈 일자리	0.016	0.125	0.016	0.125	0.0001	0.02
사내하도급 하청업체 정보	0.003	0.059	0.007	0.085	-0.004	-1.18
정규직 대비 비정규직 임금 수준	0.012	0.111	0.017	0.130	-0.005	-0.88
무기계약직 전환 근로자 수	0.008	0.087	0.010	0.100	-0.002	-0.58
직급별 기본급 밴드의 폭	0.093	0.290	0.089	0.285	0.004	0.28
근로시간	0.003	0.053	0.006	0.076	-0.003	-1.06
예산에 배정된 교육훈련비	0.010	0.098	0.014	0.119	-0.005	-0.96
교육훈련에 지출된 총비용	0.003	0.059	0.004	0.066	-0.001	-0.30
육아휴직 근로자 수	0.004	0.064	0.001	0.038	0.003	1.03
전체 사업체 수(개)	1,446		697			

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

답은 응답거절, 모름을 포함한다. 항목무응답률 분석에 활용한 변수는 총 10개로, 최종 배포 자료 중 항목무응답이 존재하는 변수를 중심으로 선정하였다. 분석 결과는 <표 6-10>과 같다.

자발 CAWI와 CAPI의 항목무응답률을 비교하면, 7개 변수는 자발 CAWI가, 3개 변수는 CAPI가 항목무응답률이 높았지만 이 중 통계적으로 유의한 변수는 없었다.

지금까지 조사방식별 자료수정률과 항목무응답률의 차이가 있는지 살펴해보았다. 자료수정률 분석을 위해 선정한 변수 12개 중 2개 변수 외에 통계적으로 유의미한 차이를 보이지 않았다. 산업은 CAWI가, 대졸 신입 사원 초임은 CAPI가 자료수정률이 통계적으로 유의하게 높았다. 항목무응답률 분석을 위해 선정한 변수 10개 모두 CAPI와 CAWI 간에 통계적으로 유의한 차이가 없었다. 이를 통해 CAPI와 CAWI 방식 간에 자료의 질은 큰 차이를 보이지 않는 것으로 판단된다.

제3절 CAWI 효과 분석

앞서 사업체패널 8차 조사를 활용하여 CAPI와 자발 CAWI로 응답한 사업체를 대상으로 응답률과 자료의 질을 탐색적으로 비교하였다. 본 절에서는 통계모형을 활용하여 CAWI가 응답률과 자료의 질에 미친 영향을 살펴보겠다. 먼저, 8차 조사를 활용하여 회귀분석을 통해 조사방식과 응답률 및 자료의 질 간의 관계를, 다음으로, 7차 조사에서 CAPI로 응답한 사업체 중 8차 조사에서 CAPI와 자발 CAWI로 응답한 사업체에 대한 패널자료를 구축하여 패널 이중차분법(difference-in-differences)을 통해 조사방식이 응답률 및 자료의 질에 미친 영향을 분석하겠다.

1. 8차 자료를 활용한 회귀분석

이제 8차 자료를 활용하여 식 (1)에 대한 회귀분석을 실시하겠다.

$$Y_i = \beta_0 + \beta_d D_i + \beta'_x X_i + U_i \quad (1)$$

여기서 Y_i 는 개별 사업체(i)의 응답률과 자료의 질을 나타내는 값, D_i 는 조사방식에 대한 더미변수, X_i 는 응답자 및 사업체 특성, U_i 는 오차항이다. 이때, 응답률은 응답기간을, 자료의 질은 자료수정 및 항목무응답 문항 수를, 조사방식은 CAPI, CAPI+CAWI(혼합조사), 자발 CAWI, 비자발 CAWI에 대한 더미변수를 활용하였다.

가. 응답률

여기서는 응답자와 사업체 특성을 통제하여 조사방식과 응답률 간의 관계에 대해 분석하였다. 응답률의 대리변수(proxy)로 응답기간(Dur4)을 활용하였다. 조사성공률은 조사가 중단된 사업체의 응답자 정보와 일부 사업체 정보 등이 없어 활용하지 못하였다. 설문을 응답하는 기간이 지나

치게 길어지는 경우 패널 탈락 확률이 높아질 수 있기 때문에 이를 활용하였다. 제2절에서 설명하였듯이 응답기간(Dur4)은 인사담당자 설문 의 첫 문항부터 마지막 문항 응답까지로 정의한다(표 6-6 참조).

분석에 앞서 활용된 자료의 기술통계량을 살펴보겠다. 먼저, 종속변수를 보면(표 6-11 참조), 전체 사업체에 대한 응답기간은 13.5일로 나타났다. 조사방식별 응답기간을 보면, CAPI가 9.8일로 가장 짧고 비자발 CAWI가 46.2일로 가장 길었다. 조사참여에 대해 부정적인 성향을 보일 것으로 판단되는 비자발 CAWI로 조사하였을 때의 응답기간이 가장 길어, 응답기간이 길어질수록 패널 이탈로 이어질 확률이 높음을 확인하였다.⁵⁹⁾

다음으로 독립변수에 대한 기술통계를 살펴보겠다(표 6-12 참조). 전체 분석 자료 중 조사방식의 비중은 CAPI가 54.2%, 자발 CAWI가 23.7%, CAPI+CAWI(혼합조사)가 19.2%, 비자발 CAWI가 3.0%로 나타났다. 응답자 특성을 보면, 7차 조사와 동일응답자의 비중은 67.0%, 남성 응답자의 비중은 56.9%, 평균 연령은 40.6세, 평균 근속년수는 9.7년이고, 직급별 비중은 최고경영자 및 임원이 5.0%, 부장 및 차장이 25.6%, 과장이 25.3%, 대리 이하가 44.1%로 나타났다. 사업체 특성을 보면, 규모별 비중은 30~99인 이하가 49.4%, 100~299인이 28.5%, 300~499인이 10.2%, 500인 이상이 11.8%,이고, 수도권(서울, 경기, 인천)에 위치한 사업체는 44.2%, 노동조합이 있다고 응답한 사업체는 29.4%, 공공부문은 3.3%로

〈표 6-11〉 조사방식별 응답률에 대한 기술통계량

(단위: 일, 개)

	전체		CAPI		CAPI +CAWI		자발 CAWI		비자발 CAWI	
	평균	SD	평균	SD	평균	SD	평균	SD	평균	SD
응답기간	13.5	23.2	9.8	17.7	17.0	27.6	15.0	19.5	46.2	56.0
전체 사업체 수	2,443		1,324		468		578		73	

59) 비자발 CAWI의 조사성공률은 73.3%로, CAPI(96.4%)와 자발 CAWI(84.9%)의 조사성공률에 비해 상당히 낮았다.

〈표 6-12〉 조사방식, 응답자 및 사업체 특성에 대한 기술통계량(n=2,443)

		평균	표준편차		
조사방식	CAPI	0.542	0.498		
	혼합조사	0.192	0.394		
	자발 CAWI	0.237	0.425		
	비자발 CAWI	0.030	0.170		
응답자 특성	7차 조사와 동일응답자=1	0.670	0.470		
	남성=1	0.569	0.495		
	연령(세)	40.64	8.95		
	근속년수	9.73	6.61		
	직급	최고경영자, 임원급	0.050	0.218	
		부장, 차장급	0.256	0.437	
		과장급	0.253	0.435	
		대리급 이하	0.441	0.497	
	사업체 특성	규모	99인 이하	0.494	0.500
			100~299인	0.285	0.452
300~499인			0.102	0.303	
500인 이상			0.118	0.323	
수도권 입지=1		0.442	0.497		
유노조 사업체=1		0.294	0.456		
공공부문=1		0.033	0.179		
산업		경공업	0.110	0.313	
		화학공업	0.090	0.286	
		금속·자동차·운송	0.150	0.357	
	전기·전자·정밀	0.083	0.276		
	전기·가스·수도업	0.013	0.114		
	건설업	0.075	0.264		
	개인서비스업	0.119	0.324		
	유통서비스업	0.091	0.288		
	사업서비스업	0.173	0.378		
사회서비스업	0.096	0.294			

〈표 6-13〉 조사방식과 응답기간 간의 관계 분석

		응답기간(일)	
		Coef.	Std. Err.
상수항		18.92	(8.70)**
조사방식	[기준: CAPI]		
	CAPI+CAWI	6.46	(1.22)***
	자발 CAWI	4.55	(1.13)***
	비자발 CAWI	35.52	(2.69)***
응답자	2017년과 동일응답자=1	-1.97	(1.06)*
	남성=1	0.07	(0.98)
	연령	-0.09	(0.41)
	연령 ² /100	0.03	(0.47)
	근속년수	0.16	(0.24)
	근속년수 ² /100	-0.28	(0.84)
	[기준: 최고경영자, 임원]		
특성	부장, 차장	0.52	(2.35)
	과장	-0.36	(2.46)
	대리 이하	-2.38	(2.54)
사업체	[기준: 99인 이하]		
	100~299인	-0.74	(1.10)
	300~499인	-2.74	(1.60)*
	500인 이상	-0.19	(1.63)
	유노조=1	-1.93	(1.24)
	공공부문=1	11.18	(2.67)***
	수도권 위치=1	0.68	(0.96)
	[기준: 경공업]		
	화학공업	-4.85	(2.02)**
	금속·자동차·운송	-0.40	(1.79)
	전기·전자·정밀	1.01	(2.07)
	전기·가스·수도업	-3.37	(4.17)
	건설업	-5.52	(2.15)***
	개인서비스업	-8.00	(1.89)***
유통서비스업	-7.81	(2.10)***	
사업서비스업	-7.31	(1.78)***	
사회서비스업	-3.96	(2.00)**	
R ²		0.110	
전체 사업체 수(개)		2,443	

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

나타났다. 산업별 비중은 경공업이 11.0%, 화학공업이 9.0%, 금속·자동차·운송이 15.0%, 전기·전자·정밀이 8.3%, 전기·가스·수도업이 1.3%, 건설업이 7.5%, 개인서비스업이 11.9%, 유통서비스업이 9.1%, 사업서비스업이 17.3%, 사회서비스업이 9.6%로 나타났다. 전체 사업체 중 제조업의 비중은 43.3%이고 서비스업의 비중은 47.9%이다.

이제 조사방식과 응답기간과의 관계에 대해 분석한 결과를 살펴보겠다(표 6-13 참조). 응답기간은 CAPI 조사에 비해 CAPI+CAWI(혼합조사), 자발 CAWI, 비자발 CAWI 각각 6.5일, 4.6일, 35.5일 길게 나타났고 이는 모두 통계적으로 유의하였다. CAPI는 다른 조사방식에 비해 응답기간이 짧다고 나타나, 방문면접 방식은 다른 조사방식에 비해 응답률을 높이는 데 긍정적인 역할을 하는 것으로 판단된다. 이를 통해 CAWI가 CAPI보다 높은 응답률을 기대하기는 어려울 것으로 보인다.

응답자 특성별로 보면, 7차 조사와 동일응답자가 설문에 응한 경우 그렇지 않은 경우에 비해 응답기간은 짧다고 나타났으나 이는 유의하지 않았다. 그 외 성별, 연령, 근속, 직급 모두 통계적으로 유의하지 않았다.

사업체 특성별로 보면, 응답기간은 99인 미만 사업체에 비해 300~499인 사업체가, 공공부문에 비해 민간부문이 짧다고 통계적으로 유의하게 나타났다. 경공업 사업체에 비해 화학공업, 건설업, 서비스업 사업체의 응답기간이 통계적으로 유의하게 짧은 것으로 나타났다.

나. 자료의 질

이제 조사방식과 자료의 질과의 관계를 살펴보겠다. 자료수정 문항 수는 50개 변수 중 자료가 수정된 문항의 합계로, 항목무응답 문항 수는 68개의 변수 중 항목무응답으로 처리한 문항의 합계로 측정하였다.

<표 6-14>에 나타난 전체 분석자료의 평균적인 자료수정 문항 수는 0.30개, 항목무응답 문항 수는 0.26개이다. 조사방식별로 보면 CAPI와 자발 CAWI는 평균과 비슷하고, CAPI+CAWI(혼합조사)는 항목무응답 문항 수가 0.29개로 다소 많았다. 비자발 CAWI는 자료수정 문항 수가 0.34개로 전체 평균 비해 많았고 항목무응답 문항 수는 0.19개로 적었다. 그

〈표 6-14〉 조사방식별 자료수정 문항 수와 항목무응답 문항 수에 대한 기술통계량
(단위: 개)

	전체		CAPI		CAPI +CAWI		자발 CAWI		비자발 CAWI	
	평균	SD	평균	SD	평균	SD	평균	SD	평균	SD
자료수정 문항 수	0.30	0.55	0.31	0.55	0.28	0.55	0.29	0.54	0.34	0.56
항목무응답 문항 수	0.26	0.58	0.24	0.56	0.29	0.65	0.26	0.58	0.19	0.43
전체 사업체 수(개)	2,443		1,324		468		578		73	

외 응답자와 사업체의 특성은 앞의 <표 6-12>를 참고한다.

<표 6-15>는 조사방식과 자료의 질 간의 관계를 분석한 결과이다. 자료수정 문항 수는 CAPI 조사에 비해 CAPI+CAWI(혼합조사)와 자발 CAWI는 적고 비자발 CAWI는 많다고 나타났으나 이는 통계적으로 유의하지 않았다. 항목무응답 문항 수는 CAPI 조사에 비해 CAPI+CAWI(혼합조사)와 자발 CAWI는 많고 비자발 CAWI는 적다고 나타났으나 CAPI+CAWI(혼합조사)만 통계적으로 유의하였다. 조사방식과 자료의 질 간에 유의미한 관계는 보이지 않는 것으로 판단된다.

응답자 특성별로 보면, 7차 조사와 동일 응답자인 경우 그렇지 않은 경우에 비해 자료수정 문항 수와 항목무응답 문항 수 모두 적다고 나타났는데, 이때, 자료수정 문항 수만 통계적으로 유의하였다. 이를 통해 과거 사업체패널조사를 경험한 응답자가 재응답하는 경우 자료의 질에 긍정적인 영향을 미칠 수 있음을 확인하였다. 사업체패널조사는 사업체를 대상으로 하는 패널조사이기 때문에 인사담당자가 변경되면 응답자가 바뀌는 경우가 존재한다. 그러나 자료의 질을 위해 응답자를 최대한 패널화할 수 있도록 노력을 기울여야 할 것이다. 그리고 여성에 비해 남성이 자료수정 문항 수와 항목무응답 문항 수가 많다고 나타났고 이때 자료수정 문항 수에 대해서만 유의하였다. 그 외 성별, 연령, 근속년수, 직급은 자료의 질에 유의미한 영향을 미치지 않는 것으로 나타났다.

사업체 특성별로 보면, 99인 이하 소규모 사업체에 비해 300-499인과 500인 이상 사업체가 자료수정 문항 수와 항목무응답 문항 수가 더 많다고 나타났고 이는 통계적으로 유의하였다. 유노조 사업체는 무노조에 비

〈표 6-15〉 조사방식과 응답의 질과의 관계 분석

		자료수정 문항 수		무응답 문항 수	
		Coef.	Std. Err.	Coef.	Std. Err.
상수항		0.236	(0.213)	0.238	(0.227)
조사방식	[기준 : CAPI]				
	CAPI+CAWI	-0.004	(0.030)	0.057	(0.032)*
	자발 CAWI	-0.003	(0.028)	0.029	(0.030)
	비자발 CAWI	0.010	(0.066)	-0.069	(0.070)
응답자 특성	2017년과 동일응답자=1	-0.053	(0.026)**	-0.027	(0.028)
	남성=1	0.061	(0.024)**	0.014	(0.026)
	연령	0.005	(0.010)	-0.003	(0.011)
	연령 ² /100	-0.008	(0.012)	0.001	(0.012)
	근속년수	0.004	(0.006)	0.001	(0.006)
	근속년수 ² /100	0.002	(0.021)	0.008	(0.022)
	[기준 : 최고경영자, 임원]				
부장, 차장	0.050	(0.057)	0.069	(0.061)	
과장	0.019	(0.060)	0.100	(0.064)	
대리 이하	0.020	(0.062)	0.078	(0.066)	
사업체 특성	[기준 : 99인 이하]				
	100~299인	0.041	(0.027)	0.032	(0.029)
	300~499인	0.112	(0.039)***	0.101	(0.042)**
	500인 이상	0.146	(0.040)***	0.084	(0.043)**
	유노조=1	0.032	(0.030)	0.083	(0.032)***
	공공부문=1	-0.187	(0.065)***	-0.076	(0.070)
	수도권 입지=1	-0.067	(0.024)***	-0.045	(0.025)*
	[기준 : 경공업]				
	화학공업	-0.045	(0.049)	-0.051	(0.053)
	금속·자동차·운송	-0.016	(0.044)	-0.016	(0.047)
전기·전자·정밀	-0.090	(0.051)*	-0.040	(0.054)	
전기·가스·수도업	0.093	(0.102)	-0.085	(0.109)	
건설업	0.006	(0.053)	-0.024	(0.056)	
개인서비스업	-0.029	(0.046)	-0.012	(0.049)	
유통서비스업	-0.172	(0.051)***	0.027	(0.055)	
사업서비스업	-0.126	(0.044)***	-0.048	(0.046)	
사회서비스업	-0.143	(0.049)***	-0.004	(0.052)	
R ²		0.041		0.022	
전체 사업체 수(개)		2,443			

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

해 자료수정 문항 수와 항목무응답 문항 수 모두 많다고 나타났는데, 항목무응답 문항 수만 통계적으로 유의하였다. 공공부문은 민간부문에 비해 자료수정 문항 수와 항목무응답 문항 수가 모두 적었는데, 자료수정 문항 수만 통계적으로 유의하였다. 수도권에 위치한 사업체는 비수도권에 위치한 사업체에 비해 자료수정 문항 수와 항목무응답 문항 수가 모두 적고 통계적으로 유의하였다. 산업별로 보면, 자료수정 문항 수는 경공업 사업체에 비해 전기·전자·정밀, 유통·사업·사회 서비스업 사업체가 통계적으로 유의하게 적었고, 항목무응답 문항 수는 통계적으로 유의하지 않았다.

2. 패널자료를 활용한 이중차분법 분석

이제 패널 이중차분법을 활용하여 CAWI 방식에 대한 모드 효과를 분석하겠다. 분석 데이터셋은 7차 조사에서 CAPI로 조사한 사업체 중 8차 조사에서 CAPI와 자발 CAWI로 응답한 사업체를 대상으로 하여 구축하였다(표 6-16 참조). 이중차분법은 정책 혹은 제도의 시행 전후 관심대상 변수에 대한 처치집단(treatment group)과 통제집단(control group) 간의 차이를 통해 정책이나 제도의 인과적 효과를 추정할 수 있다고 알려진 방법이다.⁶⁰⁾

〈표 6-16〉 분석 데이터셋 구축 현황(패널자료 구축)

	2017	2019
처치집단 (Treatment Group) (Q=1, n=53)	CAPI	자발 CAWI
통제집단 (Control Group) (Q=0, n=1,072)	CAPI	CAPI

60) 이중차분법은 정책이나 제도 시행을 하지 않았더라면 두 집단의 결과변수가 비슷한 추세를 보인다는 가정하에서 인과 관계를 추정한다. 보통 정책이나 제도 시행 이전의 두 집단 결과변수 값에 대한 유사 궤적을 보여줌으로써 이를 검증한다. 그런데 사업체패널 6차 조사는 7, 8차 조사와 다른 조사업체, 다른 CAPI 시스템, 다른 면접원 등에 의해 조사되어 조사 상황에 영향을 미치는 요인들이 많아 이에 대한 별도의 분석은 하지 않았다.

분석의 기본이 되는 모형은 식 (2)와 같이 설정하였다.

$$Y_{it} = \beta_t + \beta_q Q_i + \beta_d D_{it} + \beta'_w W_{it} + \delta_i + U_{it} \quad (2)$$

여기서 Y_{it} 는 개별 사업체(i)의 기간(t)에 대한 응답기간과 자료수정 문항 수, Q_i 는 처치집단으로, 7차 조사에서는 CAPI로 조사하고 8차 조사에서는 자발 CAWI로 조사한 집단에 대한 더미변수이고, D_{it} 는 Q_i 와 2019년을 의미하는 더미변수의 교호항($D_{it} = Q_i \times 1[2019 = 1]$)이다. 그리고 X_i 는 응답자와 사업체 특성, δ_i 는 개별 사업체(i)에 대한 미관측 이질성, U_{it} 는 통상적인 오차항이다.

$$\Delta Y_{i,2019} = \Delta \beta_{2019} + \beta_d Q_i + \beta'_x \Delta x_{i,2019} + \Delta U_{i,2019} \quad (3)$$

식 (2)를 그대로 추정해도 되나, 미관측 이질성이 존재하는 경우 내생성이 발생할 수 있으므로 차분한 식 (3)을 OLS로 추정하였다.

가. 기술통계량

분석에 앞서, 기초통계량을 살펴보겠다(표 6-17 참조). 7차 조사에서 전체 설문을 CAPI로 조사한 사업체는 1,125개인데, 이 중 8차 조사에서 CAPI로 조사한 사업체는 1,072개, 자발 CAWI로 조사한 사업체는 53개이다. 여기서 처치집단은 8차에서 자발 CAWI로 조사한 사업체이다.

먼저, 응답기간에 대해 살펴보겠다. 처치집단(Q=1)인 8차 조사에서 자발 CAWI로 조사한 사업체의 응답기간은 2017년에 11.4일에서 2019년에 18.6일로 7.2일만큼, 통제집단(Q=0)인 항상 CAPI로 조사한 사업체의 응답기간은 2017년에 10.5일에서 2019년에 11.0일로 0.4일만큼 증가하였다. 처치집단은 통제집단에 비해 응답기간이 6.8일 더 길게 나타났다. 이때, 표준오차는 3.88이므로 유의수준 10%하에서 통계적으로 유의하였다.

다음으로, 자료수정 문항 수를 살펴보겠다. 처치집단(Q=1)은 2017년에 1.85개에서 2019년에 0.38개로 1.47개만큼, 통제집단(Q=0)은 2017년에 2.06개에서 2019년에 0.31개로 1.75개만큼 감소하였다. 처치집단은 통제집단

〈표 6-17〉 패널자료의 기초통계량

		2017				2019				
		통제집단 (Q=0)		처치집단 (Q=1)		통제집단 (Q=0)		처치집단 (Q=1)		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	
종속 변수	응답기간(일)	10.54	20.75	11.38	23.69	10.95	18.43	18.55	23.01	
	자료수정 문항 수(개)	2.063	1.231	1.849	1.099	0.310	0.559	0.377	0.562	
통제 변수	남성=1	0.629	0.483	0.491	0.505	0.593	0.491	0.472	0.504	
	연령(세)	40.35	8.89	40.13	8.58	40.68	9.10	39.64	8.43	
	근속년수	9.17	6.39	9.21	6.86	9.60	6.54	9.36	7.01	
	직급	최고경영자·임원	0.069	0.254	0.075	0.267	0.055	0.228	0.057	0.233
		부장·차장	0.281	0.450	0.283	0.455	0.228	0.419	0.302	0.463
		과장	0.249	0.433	0.208	0.409	0.266	0.442	0.170	0.379
		대리 이하	0.342	0.475	0.358	0.484	0.451	0.498	0.472	0.504
상시근로자 수		241.5	551.8	261.8	495.5	249.9	570.0	266.8	515.1	
전체 사업체 수(개)		1,072		53		1,072		53		

에 비해 자료수정 문항 수가 0.28개 더 많다고 나타났다. 이때, 표준오차는 0.19로 통계적으로 유의하지 않았다.

통제변수를 보면, 처치집단은 통제집단에 비해 여성의 비율이 높고 평균 상시근로자 수가 많은 것으로 나타났다.

나. 실증분석

이제 패널 이중차분법으로 CAWI가 응답기간과 자료수정 문항 수에 미치는 효과에 대해 분석한 결과를 살펴보겠다(표 6-18 참고).

자발 CAWI로 조사한 경우, 응답기간은 6.8일 길고 자료수정 문항 수는 0.27개 정도 증가한다고 나타났다. 이때, 응답기간에 대해서만 통계적으로 유의하였다. 이를 통해 CAWI 방식은 CAPI에 비해 응답률이 낮은 것으로 추정되나 자료의 질에 특별한 영향을 미치지 않는 것으로 보인다. 응답기간은 2017년에 비해 2019년에 약 0.9일 정도 길고 자료수정 문항

〈표 6-18〉 패널 이중차분법을 활용한 분석결과(n=1,125, t=2)

	응답기간(Dur4)		자료수정 문항 수	
	Coeff.	Std. Err	Coeff.	Std. Err
[2019=1]	0.884	(0.898)	-1.742	(0.044) ***
[2019=1]×[자발 CAWI=1]	6.798	(3.891) *	0.267	(0.191)
남성=1	1.788	(2.715)	-0.037	(0.134)
연령	-0.642	(0.825)	-0.015	(0.041)
연령 ² /100	0.723	(0.965)	0.007	(0.047)
근속년수	0.299	(0.441)	-0.022	(0.022)
근속년수 ² /100	-1.531	(1.597)	0.105	(0.079)
[기준: 최고경영자, 임원]				
부장·차장급	-2.387	(2.673)	-0.061	(0.132)
과장급	-2.513	(2.805)	-0.156	(0.138)
대리급 이하	-4.351	(2.807)	-0.114	(0.138)
log(상시근로자)	-1.175	(2.487)	0.227	(0.122) *
R ²	0.008		0.010	

주: ***는 1% 수준, **는 5% 수준, *는 10% 수준에서 통계적으로 유의함.

수는 2017년에 비해 2019년에 1.7개 정도 작았는데, 자료수정 문항 수에 대해서만 통계적으로 유의하였다. 이 효과는 조사업체가 두 번째 조사를 수행하면서 첫 번째 조사를 통해 축적된 경험과 숙련도의 향상에 힘입은 것으로 보인다. 그 외, 성별, 연령, 근속, 직급에 대해서는 모두 유의하지 않았다. 자료수정 문항 수는 규모가 클수록 통계적으로 유의수준 10%하에서 정(+)의 효과를 보였다.

지금까지 횡단면 자료와 패널자료를 활용하여 통계모형을 통해 분석하였다. 응답기간은 CAPI 조사에 비해 자발 CAWI 조사가 길다고 나타났고, 이는 횡단면 분석과 패널 분석 모두 유의하였다. 이는 기존의 방문조사방식이나 전화조사방식은 Web 조사방식에 비해 응답률이 더 높다는 선행연구들과 동일한 결과다(이지은 외, 2017; Fricker, Galesic, Tourangeau and Yan, 2005; De Leeuw, 2005; Lozar et al., 2008; Shih and Fan, 2008; Kirchner and Felderer, 2017; Daikeler et al., 2020). 다만,

응답률의 대리변수로 응답기간을 활용하였기 때문에 조사방식별 응답률의 차이를 추정하지 못하였다. 자료수정 문항 수는 CAPI와 자발 CAWI 간에 횡단면 분석과 패널 분석 모두에서 유의미한 차이가 나타나지 않아, 조사방식에 따라 자료의 질은 체계적인 차이를 보이지 않는 것으로 판단된다.

제4절 요약 및 결론

이 장에서는 CAWI와 CAPI로 응답한 사업체를 비교하여 Web 조사방식이 응답률과 자료의 질에 어떠한 영향을 미치는지를 살펴보았다.

우선, 응답률은 CAWI 조사가 CAPI 조사에 비해 낮은 것으로 나타났다. 자발 CAWI 조사는 CAPI 조사에 비해 조사성공률이 11.5%포인트 낮고 전체 설문에 대한 응답기간은 4.6일 긴 것으로 확인되었다. 또한 8차 조사 자료를 회귀분석한 결과와 7, 8차 패널자료를 이중차분법으로 분석한 결과 모두 CAWI가 CAPI에 비해 응답기간이 각각 4.6일, 6.8일 통계적으로 유의하게 길게 나타나, 다시 한 번 이를 확인하였다.

다음으로, CAWI 조사로 생산된 데이터의 품질은 CAPI와 비교해 유의미한 차이가 없었다. 주요 변수에 대한 자료수정률과 항목무응답률을 자발 CAWI와 CAPI에 대해 비교한 결과, 자료수정률은 12개 중 2개의 변수에서 통계적으로 유의한 차이가 있었으나 항목무응답률은 10개 변수 모두 유의미한 차이가 없어 전반적으로 자료의 질에서 차이가 난다고 하기는 어려웠다. 횡단면 자료를 활용하여 조사방식과 자료수정 및 항목무응답 문항 수와의 관계를 분석한 결과, 통계적으로 유의한 결과를 보이지 않았다. 특히 불성실한 응답 가능성이 높은 비자발 CAWI도 자료의 질은 다른 조사방식과 별반 차이를 보이지 않았다. 패널자료를 이중차분법으로 분석한 결과 역시 CAWI는 CAPI에 비해 자료수정 문항 수에서 유의미한 차이를 보이지 않았다.

지금까지의 분석 결과를 통해 볼 때, 만약 사업체패널조사에 전면적으로

로 Web을 활용하게 된다면 응답률은 지금보다 낮아질 것이다. 패널조사에서 패널 이탈(attrition)이 심한 경우 조사의 의미는 상실된다. 따라서 사업체패널조사는 향후에도 지금과 같이 방문조사를 원칙으로 하되, 응답자가 원하거나 방문이 거절된 경우 혹은 방문조사에서 완료하지 못한 경우 등에 한해서만 보완적으로 Web을 활용하는 것이 가장 적합할 것으로 보인다. 이때 조사방식을 혼합하여 활용하더라도 사업체패널 자료의 품질 측면에서는 큰 차이가 없을 것으로 예상된다.

최근 국내외 패널조사에서 부분적으로 Web을 활용하여 응답자에 대한 접근성을 높이고 있다. 이는 Web을 통해 과거에는 접근하지 못했던(거리의 한계 등) 응답자나 상황에서 조사를 성공시킴으로써, 조사 전체의 응답률을 상승시키기 위한 것이다. 본 연구는 CAWI와 CAPI의 응답률과 자료의 질 비교에만 초점을 맞췄지만, 향후 사업체패널조사에서 지금과 같은 혼합된 Web 활용이 전체 조사성공률에 미치는 영향⁶¹⁾ 등과 같은 후속 연구가 필요할 것이다. 이는 추후의 연구과제로 남겨 두겠다.

61) 사업체패널 8차 조사의 이전 차수 대비 조사성공률(85.5%)은 7차 조사(80.5%)에 비해 5.0%포인트 증가되었다.

참고문헌

- 강신욱·정해식·오욱찬 외(2018), 「유럽 국가들의 행정자료 통합 및 활용」, 연구보고서, 『소득불평등 심화의 원인과 정책적 대응 효과 연구』, 한국보건사회연구원.
- 김우영(2017), 「한국노동패널조사와 경제활동인구조사에 나타난 노동시장활동 비교」, 『패널자료 품질개선연구 VII』, 한국노동연구원, pp. 244~271.
- 김유빈·박민규·변종석·최효미·권혁진·김우영·성재민·이지은·신선옥·이혜정·김기홍(2017), 『패널자료 품질개선 연구 VII』, 한국노동연구원.
- 김유빈·고영우·김기민·박민규·이지은·채민희(2020), 『패널자료 품질개선 연구 X』, 한국노동연구원.
- 김유빈·이지은·신선옥·정현상·권익성(2020), 「한국노동패널 1~22차년도 조사자료 User's Guide」, 한국노동연구원.
- 네이버 지식백과, 「심리학용어사전 한부모 가정」, <https://terms.naver.com/entry.naver?docId=2094195&cid=41991&categoryId=41991>(인출일: 2021. 7. 22).
- 로위 닐센(2020), 「사회 정책 계획 또는 평가를 위한 노르웨이 행정 데이터 활용」, 『국제사회보장리뷰』 14(1), pp.30~37.
- 백상숙(2020), 「호주의 보건 복지 공공 데이터 융합에 대한 고찰」, 『국제사회보장리뷰』 14(1), pp.16~29.
- 시숭·토머스 콜먼(2020), 「미국의 사회과학 및 정책 연구를 위한 행정 빅데이터 활용」, 『국제사회보장리뷰』 14(1), pp.5~15.
- 신광영(2017), 「스웨덴의 행정 데이터 통합과 활용에 관한 연구」, 『스칸디나비아 연구』 20(1), pp.83~109.
- 오미애(2020), 「캐나다의 보건, 복지, 사회보장 행정 데이터 및 공공 데이

- 터 연계」, 『국제사회보장리뷰』 14(1), pp.53~64.
- 오미애·최현수·박민규·권성훈·진재현·천미경·김솔휘(2016), 『2016년 보건복지통계정보 통합 관리 및 운영』, 한국보건사회연구원.
- 유종성·전병유·신광영·이도훈·최성수(2020), 「증거기반 정책연구를 위한 행정 데이터의 활용」, 『한국사회정책』 27(1), pp.5~37.
- 윤광석·이건(2017), 「공공데이터 활용 행정 촉진을 위한 거버넌스 모색: 영국의 ADRN 사례를 중심으로」, 『국가정책연구』 31(1), pp.1~27.
- 이상호(2012), 「패널조사의 비표본오차와 CAPI에 관한 이론적 검토」, 『패널자료의 품질 개선을 위한 조사기법 연구』, 한국노동연구원
- _____ (2015), 「노동패널 고용지표의 신뢰성 검토」, 『노동패널자료 연구 I』, 한국노동연구원.
- 이지은·신선옥·이혜정·김기홍(2017), 「웹을 활용한 조사의 효과에 대한 실증연구: 노동패널 CAWI 실험 데이터의 분석」, 『패널자료 품질개선 연구(Ⅶ)』, 한국노동연구원.
- 통계청, 「경제활동인구조사 통계 설명 자료」, <https://meta.narastat.kr/metasvc/index.do?confmNo=101004>
- 한국노동패널조사 홈페이지, 표본추출, <https://www.kli.re.kr/klips/contents.do?key=142>(인출일: 2021. 7. 20).
- 한은희(2020), 「영국의 연구를 목적으로 한 행정 데이터 연계 활용」, 『국제사회보장리뷰』 14(1), pp.38~52.
- 홍민기·강신옥·성재민·이상호·장지연·김재광·이지은(2015), 『노동패널자료 연구 I: 패널자료 비교연구를 중심으로』, 한국노동연구원.

Abowd, J. M., and F. Kramarz(2005), “Human Capital and Worker Productivity: Direct Evidence from Linked Employer–employee Data,” *Annales d’Economie et de Statistique*, pp.323~338.

Abowd, J. M., K. L. McKinney, and N. L. Zhao(2018), “Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer–employee Data,” *Journal of Labor Economics* 36(S1),

S183-S300.

- Abowd, J. M. and M. H. Stinson(2013), “Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data,” *Review of Economics and Statistics* 95(5), pp.1451~1467.
- Abowd, J., M. Stinson, and G. Benedetto(2006), “Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project,” pp.1~159.
- Anseel, F., F. Lievens, E. Schollaert, and B. Choragwicka,(2010), “Response rates in organizational science, 1995~2008: A meta-analytic review and guidelines for survey researchers,” *Journal of Business and Psychology* 25(3), pp.335~349.
- Alexander, M., K. Polimis, and E. Zagheni(2020), “Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States,” *Population Research and Policy Review*, pp.1~28.
- Australian Government(2018), “Framework to Guide the Secondary Use of My Health Record System Data, Department of Health,” Viewed 24 July 2012. [https://www1.health.gov.au/internet/main/publishing.nsf/content/F98C37D22E65A79BCA2582820006F1CF/\\$File/MHR_2nd_Use_Framework_2018_ACC_AW3.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/content/F98C37D22E65A79BCA2582820006F1CF/$File/MHR_2nd_Use_Framework_2018_ACC_AW3.pdf).
- Baruch, Y.(1999), “Response rate in academic studies - A comparative analysis,” *Human Relations* 52(4), pp.421~438.
- Bingley, P. and A. Martinello(2014), “Measurement Error in the Survey of Health, Ageing and Retirement in Europe: A Validation Study with Administrative Data for Education Level, Income and Employment,” *Work Pap Ser*, pp.1~33.
- Bø, E. E., E. Halvorsen, and T. O. Thoresen(2019), “Heterogeneity of the Carnegie effect,” *Journal of Human Resources* 54(3), pp.726~759.

- Bohensky, M. A., D. Jolley, V. Sundararajan, S. Evans, J. Ibrahim, and C. Brand(2011), “Development and Validation of Reporting Guidelines for Studies Involving Data linkage,” *Australian and New Zealand Journal of Public Health* 35(5), pp.486~489.
- Boserup, S. H., W. Kopczuk, and C. T. Kreiner(2016), “The Role of Bequests in Shaping Wealth Inequality: Evidence from Danish Wealth Records,” *American Economic Review* 106(5), pp.656~661.
- Bowling, Ann(2005), “Mode of Questionnaire Administration Can Have Serious Effects on Data Quality,” *Journal of Public Health* 27(3), pp.281~291.
- Brønnum-Hansen, H., M. Baadsgaard, M. L. Eriksen, K. Andersen-Ranberg, and B. Jeune(2015), “Educational Inequalities in Health Expectancy During the Financial Crisis in Denmark,” *International Journal of Public Health* 60(8), pp.927~935.
- Brummet, Q.(2014), “Comparison of Survey, Rederal, and Commercial Address Data Quality,” *Center for Administrative Records Research and Applications Working Paper* 14(6), pp.1~28.
- Brzeziński, M., K. Sałach, and M. Wroński(2020), “Wealth Inequality in Central and Eastern Europe: Evidence from Household Survey and Rich Lists’ Data Combined,” *Economics of Transition and Institutional Change* 28(4), pp.637~660.
- Card, D., A. K. Hildreth, and L. D. Shore-Sheppard(2004), “The Measurement of Medicaid Coverage in the SIPP: Evidence from a Comparison of Matched Records,” *Journal of Business and Economic Statistics* 22(4), pp.410~420.
- Carrieri, V. and A. Wuebker(2016), “Quasi Experimental Evidence on the Effects of Health Information on Preventive Behaviour in Europe,” *Oxford Bulletin of Economics and Statistics* 78(6), pp.765~791.

- Daikeler, J., M. Bošnjak, and L. K. Manfreda(2020), “Web versus other survey modes: an updated and extended meta-analysis comparing response rates,” *Journal of Survey Statistics and Methodology* 8, pp.513~539
- De Leeuw, E.(1992), “Data Quality in Mail, Telephone and Face To Face Surveys,” Amsterdam: TT Publications.
- _____(2005), “To mix or not to mix. Data collection modes in surveys,” *Journal of Official Statistics* 21(2), pp.233~255.
- De Leeuw Ed, Van Der Zouwen J.(1988), “Data Quality in Telephone and Face-to-face Surveys: A Comparative Meta-analysis,” In: Groves RM, Biemer PP, Lyberg LE et al. Eds. *Telephone Survey Methodology*, New york: John Wiley and Sons, 1988.
- Dillman, D. A.(2000), *Mail and Internet Surveys: The Tailored Design Method*, NewYork: John Wiley & Sons, Inc.
- Dillman, D. A., R. L. Sangster, J. Tanari, and T. Rockwood(1996), “Understanding Differences in People’s Answers to Telephone and Mail Surveys,” In New Directions for Evaluation Series, 70, *Advances in Survey Research*, M.T. Braverman and J.K. Slater (eds). San Francisco: Jossey-Bass.
- Elinder, M., Erixson, O., and D. Waldenström(2018), “Inheritance and Wealth Inequality: Evidence from Population Registers,” *Journal of Public Economics* 165, pp.17~30.
- Erlangsen, A., and I. Fedyszyn(2015), “Danish Nationwide Registers for Public Health and Health-related Research,” *Scandinavian Journal of Public Health* 43(4), pp.333~339.
- Erixson, O.(2017), “Health Responses to a Wealth Shock: Evidence from a Swedish Tax Reform,” *Journal of Population Economics* 30(4), pp.1281~1336.
- Fellegi, I. P., and A. B. Sunter(1969), “A Theory for Record Linkage,” *Journal of the American Statistical Association* 64(328), pp.1183

~1210.

- Foverskov, E., M. M. Glymour, E. L. Mortensen, A. Holm, T. Lange, and R. Lund(2018), "Education and Cognitive Aging: Accounting for Selection and Confounding in Linkage of Data from the Danish Registry and Survey of Health," *Ageing and Retirement in Europe, American Journal of Epidemiology* 187(11), pp.2423~2430.
- Frikler S., M. Galesic, R. Tourangeau, and T. Yan(2005), "An experimental comparison of web and telephone surveys," *Public Opinion Quarterly* 69(3), pp.370~392
- Furseth, J.(2014), "Personal Identification Numbers: 50 Years Old and Ripe for Upgrading," MIMEO. https://www.ssb.no/en/omssb/samarbeid/internasjonalt-utviklingssamarbeid/_attachment/216789?_ts=14b3aaed7c0
- Groves, R., M., and E. Peytcheva(2008), "The impact of nonresponse rates on nonresponse bias: a meta-analysis," *Public Opinion Quarterly* 72(2), pp.167~189
- Heerwegh, Dirk(2009), "Mode Differences Between Face-to-face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects," *International Journal of Public Opinion Research* 21(1), pp.111~121.
- Herzog, T. N., F. J. Scheuren, and W. E. Winkler(2007), "Data Quality and Record Linkage Techniques," Springer Science and Business Media.
- Kapteyn, A. and J. Y. Ypma(2007), "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics* 25(3), pp.513~551.
- Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser(2000), "Consequences of reducing nonresponse in a national telephone survey," *Public Opinion Quarterly* 64(2), pp.125~148

- Kirchner, A. and B. Felderer(2017), “The effect of nonresponse and measurement error on wage regression across survey modes: A validation study.” In *Total Survey Error in Practice*, edited by Paul P. Biemer, Edith De Leeuw, and Stephanie Eckman, Hoboken, NJ: Wiley, pp.531~556.
- Kreiner, C. T., D. D. Lassen, and S. Leth-Petersen(2015), “10 Measuring the Accuracy of Survey Responses Using Administrative Register Data,” *University of Chicago Press*, pp.289~307.
- Lastrapes, W. D., I. M. Schmutte, and T. Watson(2020), “A Comparison of Firm Age in the Survey of Business Owners and the Longitudinal Business Database,” Center for Economic Studies, US Census Bureau 20(8).
- Lozar Manfreda, K., M. Bosnjak, J. Berzelak, I. Haas, and V. Vehovar(2008), “Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates,” *Journal of the Market Research Society* 50(1), pp.79~104.
- Ludvigsson, J. F., P. Otterblad-Olausson, B. U. Pettersson, and A. Ekblom(2009), “The Swedish Personal Identity Number: Possibilities and Pitfalls in Healthcare and Medical Research,” *European journal of Epidemiology* 24(11), pp.659~667.
- McGonagle, K. A., R. F. Schoeni, N. Sastry, and V. A. Freedman(2012), “The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research,” *Longitudinal and Life Course Studies* 3(2).
- McNabb, J., D. Timmons, J. Song, and C. Puckett(2009), “Uses of Administrative Data at the Social Security Administration,” *Social Security Bulletin* 69(1).
- Meyer, B. D., and N. Mittag(2019), “Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net,”

American Economic Journal: Applied Economics 11(2), pp.176~204.

Michael W. Link and Ali Mokdad(2006), “Can Web and Mail Survey Modes Improve Participation in an RDD-based National Health Surveillance?” *Journal of Official Statistics* 22(2), pp.293~312

Mulrow, E., A. Mushtaq, S. Pramanik, and A. Fontes(2011), “Assessment of the US Census Bureau’s Person Identification Validation System,” NORC at the University of Chicago.

Nielsen, V. O., H. Brunborg, V. Aalandslid, D. Roll-Hansen, and C. Hendriks(2014), “Status Analysis on Civil Registration and Vital Statistics(CRVS)”.

Norman R., M. T. King, D. Clarke, R. Viney, P. Cronin, D. Street(2010), “Does Mode of Administration Matter? Comparison of Online and Face-to-face Administration of a Time Trade-off Task,” *Qual Life Res* 19(4), pp.499~508.

O’Hara, A., R. M. Shattuck, and R. M. Goerge(2017), “Linking Federal Surveys with Administrative Data to Improve Research on Families,” *The ANNALS of the American Academy of Political and Social Science* 669(1), pp.63~74.

Padmanabhan, S., L. Carty, E. Cameron, R. E. Ghosh, R. Williams, and H. Strongman(2019), “Approach to Record Linkage of Primary Care Data from Clinical Practice Research Datalink to Other Health-related Patient Data: Overview and Implications,” *European Journal of Epidemiology* 34(1), pp.91~99.

Pedersen, C. B.(2011), “The Danish Civil Registration System,” *Scandinavian Journal of Public Health* 39(7), pp.22~25.

Penner, A. M., and K. A. Dodge(2019), “Using Administrative Data for Social Science and Policy,” *The Russell Sage Foundation Journal of the Social Sciences* 5(2), pp.1~18.

Petersson, F., Baadsgaard, M., and L. C. Thygesen(2011), “Danish

- Registers on Personal Labour Market Affiliation,” *Scandinavian Journal of Public Health* 39(7), pp.95~98.
- Powers, E. T., and D. Neumark(2005), “The Supplemental Security Income Program and Incentives to Claim Social Security Retirement Early,” *National Tax Journal* 58(1), pp.5~26.
- Røed, K. and O. Raaum(2003), “Administrative Registers: Unexplored Reservoirs of Scientific Knowledge,” *The Economic Journal* 113(488), F258-F281.
- Rogelberg, S. G. and J. M. Stanton(2007), “Understanding and dealing with organizational survey nonresponse,” *Organizational Research Methods* 10(2), pp.195~209.
- Schmidt, M., L. Pedersen, and H. T. Sørensen(2014), “The Danish Civil Registration System As a Tool in Epidemiology,” *European Journal of Epidemiology* 29(8), pp.541~549.
- Shih, T. H. and X. Fan(2008), “Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis,” *Field Methods* 20(3), pp.249~271.
- Smyth, J. D. and J. E. Pearson(2011), “Internet survey methods: A review of strengths, weaknesses, and innovations,” In M. Das, P. Ester, & L. Kaczmirek(Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies*, Routledge/Taylor & Francis Group, pp.11~44.
- Stata(2021), Stata Manual, <https://www.stata.com/manuals/rksmirmov.pdf>.
- Statistics Finland(2004), “Use of Registers and Administrative Data Sources for Statistical Purposes,” Best Practices of Statistics Finland, Helsinki.
- Thygesen, L. C., C. Daasnes, I. Thaulow, and H. Brønnum-Hansen(2011), “Introduction to Danish(Nationwide) Registers on Health and Social Issues: Structure, Access, Legislation, and Archiving,”

Scandinavian Journal of Public Health, 2011; 39(Suppl 7), pp.12~16.

Tomaskovic-Devey, D., M. Hällsten, and D. Avent-Holt(2015), “Where Do Immigrants Fare Worse?: Modeling Workplace Wage Gap Variation with Longitudinal Employer-employee Data,” *American Journal of Sociology* 120(4), pp.1095~1143.

Törmälehto, V. M.(2008), “Social Statistics - Integrated Use of Survey and Administrative Data at Statistics Finland,” International Association for Official Statistics Conference on Reshaping Official Statistics.

Wagner, D. and M. Lane,(2014), “The Person Identification Validation System(PVS): Applying the Center for Administrative Records Research and Applications’(CARRA) Record Linkage Software,” Center for Economic Studies, US Census Bureau.

Warren, J. R., F. T. Pfeffer, J. Helgertz, and D. Xu(2020), “Linking 1940 US Census Data to the Panel Study of Income Dynamics: Technical Documentation,” *University of Michigan PSID Technical Series Paper* 20(2), pp.1~9.

Zhang, X., L. Kuchinke, M. L. Woud, J. Velten, and J. Margraf(2017), “Survey Method Matters: Online/Offline Questionnaires and Face-to-face or Telephone Interviews Differ,” *Computers in Human Behavior* 71, pp.172~180.

[부 록]

<부표 3-1> KLIPS 98년 표본과 경제활동인구조사(비농가 가구)의 고용지표 비교

(단위:천 명, %)

	KLIPS						경제활동인구조사						KLIPS/경제조사					
	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률
1998	30,040	14,497	2,291	55.9	48.3	13.6	31,702	17,330	1,441	59.2	54.7	7.7	0.95	0.84	1.59	0.94	0.88	1.77
1999	30,254	15,727	1,123	55.7	52.0	6.7	32,210	17,765	1,332	59.3	55.2	7.0	0.94	0.89	0.84	0.94	0.94	0.96
2000	30,508	15,778	851	54.5	51.7	5.1	32,675	18,626	945	59.9	57.0	4.8	0.93	0.85	0.90	0.91	0.91	1.06
2001	30,743	16,119	777	55.0	52.4	4.6	33,098	19,079	865	60.3	57.6	4.3	0.93	0.84	0.90	0.91	0.91	1.06
2002	30,921	16,674	656	56.0	53.9	3.8	33,630	19,749	725	60.9	58.7	3.5	0.92	0.84	0.90	0.92	0.92	1.07
2003	31,082	16,596	803	56.0	53.4	4.6	34,247	19,885	797	60.4	58.1	3.9	0.91	0.83	1.01	0.93	0.92	1.19
2004	31,205	16,641	872	56.1	53.3	5.0	34,735	20,429	896	61.2	58.8	3.9	0.90	0.81	1.04	0.92	0.91	1.27
2005	31,272	16,475	950	55.7	52.7	5.4	35,198	20,679	864	61.2	58.8	4.0	0.89	0.80	1.10	0.91	0.90	1.35
2006	31,436	16,596	893	55.6	52.8	5.1	35,767	21,087	808	61.2	59.0	3.7	0.88	0.79	1.11	0.91	0.90	1.38
2007	31,596	16,677	748	55.1	52.8	4.3	36,440	21,536	773	61.2	59.1	3.5	0.87	0.77	0.97	0.90	0.89	1.24
2008	31,896	16,891	547	54.8	53.1	3.1	37,083	21,796	761	60.8	58.8	3.4	0.86	0.77	0.72	0.90	0.90	0.92
2009	32,001	16,984	758	55.4	53.1	4.3	37,572	21,705	877	60.1	57.8	3.9	0.85	0.78	0.86	0.92	0.92	1.11
2010	32,160	17,305	501	55.4	53.8	2.8	38,066	22,072	899	60.3	58.0	3.9	0.84	0.78	0.56	0.92	0.93	0.72
2011	32,409	17,755	484	56.3	54.8	2.7	38,626	22,552	838	60.6	58.4	3.6	0.84	0.79	0.58	0.93	0.94	0.75
2012	32,580	17,770	558	56.3	54.5	3.0	39,103	22,953	803	60.8	58.7	3.4	0.83	0.77	0.69	0.93	0.93	0.89
2013	32,728	17,937	510	56.4	54.8	2.8	39,579	23,320	787	60.9	58.9	3.3	0.83	0.77	0.65	0.93	0.93	0.86
2014	32,934	18,455	531	57.6	56.0	2.8	40,139	23,953	915	62.0	59.7	3.7	0.82	0.77	0.58	0.93	0.94	0.76
2015	33,109	18,037	455	55.9	54.5	2.5	40,826	24,411	948	62.1	59.8	3.7	0.81	0.74	0.48	0.90	0.91	0.67
2016	33,240	18,442	470	56.9	55.5	2.5	41,489	24,830	992	62.2	59.8	3.8	0.80	0.74	0.47	0.91	0.93	0.65
2017	33,334	18,587	527	57.3	55.8	2.8	41,808	25,153	1,002	62.6	60.2	3.8	0.80	0.74	0.53	0.92	0.93	0.73
2018	33,493	18,927	536	58.1	56.5	2.8	41,998	25,177	1,053	62.5	59.9	4.0	0.80	0.75	0.51	0.93	0.94	0.70
2019	33,559	18,498	797	57.5	55.1	4.1	42,278	25,425	1,044	62.6	60.1	3.9	0.79	0.73	0.76	0.92	0.92	1.04
2020	33,626	18,665	551	57.2	55.6	2.9	42,493	25,171	1,083	61.8	59.2	4.1	0.79	0.74	0.51	0.93	0.94	0.71

주: 경제활동인구조사는 비농가를 대상으로 한 것임. 1998~1999년의 실업률은 ILO 기준, 2000~2019년은 OECD 기준. 모든 수치는 가중치를 부여하여 얻은 것임.

〈부표 3-2〉 KLIPS 98년 표본과 경제활동인구조사(제주도 제외)의 고용지표 비교(전체 인구 대상)

(단위: 천 명, %)

	KLIPS						경제활동인구조사						KLIPS/경찰조사					
	15세 이상 인구	취업자 수	실업자 수	경찰 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경찰 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경찰 참가율	고용률	실업률
1998	30,040	14,497	2,291	55.9	48.3	13.6	34,958	19,693	1,481	60.6	56.3	7.0	0.86	0.74	1.55	0.92	0.86	1.94
1999	30,254	15,727	1,123	55.7	52.0	6.7	35,365	20,048	1,365	60.5	56.7	6.4	0.86	0.78	0.82	0.92	0.92	1.05
2000	30,508	15,778	851	54.5	51.7	5.1	35,798	20,912	971	61.1	58.4	4.4	0.85	0.75	0.88	0.89	0.89	1.16
2001	30,743	16,119	777	55.0	52.4	4.6	36,210	21,349	891	61.4	59.0	4.0	0.85	0.76	0.87	0.90	0.89	1.15
2002	30,921	16,674	656	56.0	53.9	3.8	36,612	21,959	745	62.0	60.0	3.3	0.84	0.76	0.88	0.90	0.90	1.15
2003	31,082	16,596	803	56.0	53.4	4.6	37,004	21,940	816	61.5	59.3	3.6	0.84	0.76	0.98	0.91	0.90	1.28
2004	31,205	16,641	872	56.1	53.3	5.0	37,363	22,401	855	62.2	60.0	3.7	0.84	0.74	1.02	0.90	0.89	1.35
2005	31,272	16,475	950	55.7	52.7	5.4	37,707	22,547	880	62.1	59.8	3.8	0.83	0.73	1.08	0.90	0.88	1.42
2006	31,436	16,596	893	55.6	52.8	5.1	38,214	22,899	830	62.1	59.9	3.5	0.82	0.72	1.08	0.90	0.88	1.46
2007	31,596	16,677	748	55.1	52.8	4.3	38,759	23,274	784	62.1	60.0	3.3	0.82	0.72	0.95	0.89	0.88	1.30
2008	31,836	16,891	547	54.8	53.1	3.1	39,350	23,486	771	61.6	59.7	3.2	0.81	0.72	0.71	0.89	0.89	0.97
2009	32,001	16,984	758	55.4	53.1	4.3	39,871	23,399	889	60.9	58.7	3.7	0.80	0.73	0.85	0.91	0.90	1.16
2010	32,160	17,305	501	55.4	53.8	2.8	40,389	23,749	919	61.1	58.8	3.7	0.80	0.73	0.55	0.91	0.91	0.76
2011	32,409	17,755	484	56.3	54.8	2.7	40,942	24,233	858	61.3	59.2	3.4	0.79	0.73	0.56	0.92	0.93	0.79
2012	32,580	17,770	558	56.3	54.5	3.0	41,403	24,652	821	61.5	59.5	3.2	0.79	0.72	0.68	0.92	0.92	0.94
2013	32,728	17,937	510	56.4	54.8	2.8	41,839	24,990	802	61.6	59.7	3.1	0.78	0.72	0.64	0.92	0.92	0.90
2014	32,934	18,455	531	57.6	56.0	2.8	42,316	25,575	932	62.6	60.4	3.5	0.78	0.72	0.57	0.92	0.93	0.80
2015	33,109	18,037	455	55.9	54.5	2.5	42,744	25,839	969	62.7	60.5	3.6	0.77	0.70	0.47	0.89	0.90	0.69
2016	33,240	18,442	470	56.9	55.5	2.5	43,094	26,054	1,001	62.8	60.5	3.7	0.77	0.71	0.47	0.91	0.92	0.68
2017	33,334	18,587	527	57.3	55.8	2.8	43,403	26,351	1,016	63.1	60.7	3.7	0.77	0.71	0.52	0.91	0.92	0.76
2018	33,493	18,927	536	58.1	56.5	2.8	43,638	26,450	1,066	63.1	60.6	3.9	0.77	0.72	0.50	0.92	0.93	0.72
2019	33,559	18,498	797	57.5	55.1	4.1	43,946	26,741	1,055	63.3	60.8	3.8	0.76	0.69	0.76	0.91	0.91	1.08
2020	33,626	18,635	551	57.2	55.6	2.9	44,222	26,526	1,098	62.5	60.0	4.0	0.76	0.70	0.50	0.92	0.93	0.73

주: 경찰자료는 KOSIS를 통해서 구했으며 제주도는 제외된 것임. 1998~1999년의 실업률은 ILO 기준, 2000~2019년은 OECD 기준. 모든 수치는 가중치를 부여.

〈부표 3-3〉 KLIPS 98년 표본의 가중치 적용 여부에 따른 경찰조사(시 지역 거주자)와의 고용지표 비교

(단위: 천 명, %)

	A: KLIPS(가중치 적용)			B: KLIPS(가중치 비적용)			C: 경찰조사(가중치 적용)			A/C			B/C		
	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률	경찰 참가율	고용률	실업률
1998	55.9	48.3	13.6	55.8	48.2	13.6	59.6	55.0	7.8	0.94	0.88	1.75	0.94	0.88	1.74
1999	55.7	52.0	6.7	55.6	51.9	6.7	59.8	55.6	7.0	0.93	0.94	0.96	0.93	0.93	0.95
2000	54.5	51.7	5.1	54.6	51.9	5.0	61.0	58.2	4.5	0.89	0.89	1.13	0.90	0.89	1.12
2001	55.0	52.4	4.6	55.2	52.7	4.4	60.6	57.9	4.4	0.91	0.90	1.04	0.91	0.91	1.01
2002	56.0	53.9	3.8	56.4	54.3	3.7	61.3	59.1	3.6	0.91	0.91	1.04	0.92	0.92	1.02
2003	56.0	53.4	4.6	56.5	53.8	4.7	60.7	58.3	4.0	0.92	0.92	1.16	0.93	0.92	1.18
2004	56.1	53.3	5.0	56.7	53.9	4.9	61.5	59.0	4.1	0.91	0.90	1.23	0.92	0.91	1.19
2005	55.7	52.7	5.4	56.6	53.6	5.2	61.5	58.9	4.2	0.91	0.89	1.29	0.92	0.91	1.25
2006	55.6	52.8	5.1	56.8	54.0	4.9	61.5	59.1	3.8	0.90	0.89	1.33	0.92	0.91	1.28
2007	55.1	52.8	4.3	56.2	53.9	4.2	61.5	59.3	3.6	0.90	0.89	1.19	0.91	0.91	1.15
2008	54.8	53.1	3.1	56.1	54.3	3.2	61.1	59.0	3.5	0.90	0.90	0.89	0.92	0.92	0.91
2009	55.4	53.1	4.3	56.6	54.2	4.2	60.2	57.8	4.0	0.92	0.92	1.08	0.94	0.94	1.05
2010	55.4	53.8	2.8	57.2	55.6	2.7	60.5	58.1	4.0	0.92	0.93	0.69	0.95	0.96	0.68
2011	56.3	54.8	2.7	57.9	56.4	2.6	60.8	58.6	3.7	0.93	0.94	0.73	0.95	0.96	0.70
2012	56.3	54.5	3.0	58.8	57.1	2.9	61.0	58.9	3.5	0.92	0.93	0.85	0.96	0.97	0.83
2013	56.4	54.8	2.8	58.6	57.3	2.2	61.1	59.0	3.3	0.92	0.93	0.84	0.96	0.97	0.68
2014	57.6	56.0	2.8	59.9	58.5	2.4	62.2	59.8	3.8	0.93	0.94	0.74	0.96	0.98	0.63
2015	55.9	54.5	2.5	58.4	57.1	2.2	62.4	60.0	3.8	0.90	0.91	0.65	0.94	0.95	0.58
2016	56.9	55.5	2.5	58.9	57.7	2.1	62.4	60.0	3.9	0.91	0.93	0.64	0.94	0.96	0.53
2017	57.3	55.8	2.8	58.8	57.6	2.0	62.8	60.3	3.9	0.91	0.92	0.71	0.94	0.96	0.50
2018	58.1	56.5	2.8	59.2	57.9	2.2	62.7	60.1	4.1	0.93	0.94	0.68	0.94	0.96	0.54
2019	57.5	55.1	4.1	58.9	57.4	2.5	62.8	60.3	4.0	0.92	0.91	1.03	0.94	0.95	0.63
2020	57.2	55.6	2.9	58.1	56.7	2.4	62.0	59.4	4.2	0.92	0.94	0.69	0.94	0.95	0.57

주: 경제활동인구자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임. 1998~1999년의 실업률은 ILO 기준, 2000~2019년은 OECD 기준.

〈부표 3-4〉 KLIPS 98년 표본과 경찰조사와의 연령별 경제활동참가율 비교

(단위: %)

	35세 미만			35~54세			55세 이상		
	KLIPS98	경찰	차이	KLIPS98	경찰	차이	KLIPS98	경찰	차이
1998	46.7	53.2	-6.5	74.5	76.3	-1.8	39.0	39.2	-0.2
1999	46.5	53.0	-6.5	75.0	76.4	-1.3	38.4	40.8	-2.5
2000	45.4	54.2	-8.8	74.1	77.4	-3.3	35.9	44.3	-8.4
2001	45.9	54.7	-8.9	74.3	76.4	-2.1	36.3	39.2	-2.9
2002	47.0	55.6	-8.6	76.0	76.6	-0.6	36.6	40.3	-3.7
2003	48.0	55.7	-7.6	75.1	76.1	-1.0	36.6	38.1	-1.5
2004	48.5	56.4	-7.9	74.7	76.6	-1.9	36.6	39.7	-3.1
2005	48.0	56.1	-8.1	74.8	76.9	-2.1	36.4	39.8	-3.4
2006	47.6	55.3	-7.7	74.8	77.2	-2.4	36.2	40.2	-4.0
2007	45.4	54.4	-9.1	75.5	77.4	-1.8	37.3	41.6	-4.3
2008	44.3	53.4	-9.1	76.1	77.6	-1.4	37.2	41.3	-4.1
2009	45.8	52.0	-6.2	75.6	76.7	-1.1	39.1	41.4	-2.3
2010	44.5	52.2	-7.7	77.1	77.1	0.0	38.9	42.4	-3.5
2011	44.4	52.5	-8.1	78.8	77.4	1.4	40.1	43.4	-3.3
2012	43.4	52.7	-9.3	78.1	77.5	0.6	42.4	44.7	-2.3
2013	43.1	52.7	-9.6	78.3	77.5	0.8	43.4	45.7	-2.3
2014	45.4	54.0	-8.6	78.5	78.5	0.0	44.4	47.4	-3.0
2015	43.3	54.8	-11.5	76.7	78.7	-2.0	45.1	47.8	-2.6
2016	44.4	55.2	-10.8	77.9	78.9	-1.0	46.6	48.0	-1.4
2017	45.1	55.4	-10.3	78.6	79.5	-0.9	48.2	48.8	-0.6
2018	48.0	55.9	-7.9	78.4	79.5	-1.1	49.1	48.5	0.6
2019	49.0	56.4	-7.4	76.8	79.1	-2.3	48.8	49.4	-0.6
2020	49.5	55.5	-6.0	77.0	77.9	-0.9	48.1	49.5	-1.5

주: 경제활동인구자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임.
모든 수치는 가중치를 사용하여 구한 것임.

〈부표 3-5〉 KLIPS 98년 표본과 경찰조사와의 연령별 실업률 비교

(단위 : %)

	35세 미만			35~54세			55세 이상		
	KLIPS98	경찰	차이	KLIPS98	경찰	차이	KLIPS98	경찰	차이
1998	18.5	10.1	8.4	10.6	6.2	4.5	12.3	5.5	6.8
1999	9.9	9.3	0.6	4.5	5.5	-0.9	6.5	5.1	1.5
2000	8.5	6.7	1.8	3.4	3.4	0.0	3.8	2.4	1.4
2001	8.0	6.5	1.5	2.7	3.2	-0.5	3.8	2.7	1.1
2002	7.1	5.8	1.3	2.3	2.3	-0.1	2.1	2.0	0.0
2003	8.3	6.4	1.8	3.2	2.4	0.8	1.7	2.3	-0.6
2004	9.4	6.6	2.8	3.0	2.6	0.4	2.7	2.5	0.2
2005	9.6	6.5	3.1	3.7	2.9	0.8	3.4	2.8	0.6
2006	8.8	6.5	2.3	3.8	2.4	1.4	2.9	2.6	0.3
2007	8.2	6.1	2.1	2.6	2.4	0.3	3.3	2.3	1.0
2008	5.8	6.0	-0.2	2.3	2.4	-0.1	1.8	2.1	-0.3
2009	7.7	6.7	0.9	3.1	2.8	0.3	2.8	2.6	0.2
2010	5.6	6.7	-1.1	1.8	2.7	-0.9	2.1	3.5	-1.4
2011	5.6	6.3	-0.7	1.5	2.5	-1.0	2.2	3.0	-0.8
2012	5.7	6.0	-0.3	2.3	2.3	0.0	2.2	3.0	-0.8
2013	6.3	6.3	0.0	1.7	2.1	-0.3	1.8	2.3	-0.5
2014	7.6	6.9	0.7	1.5	2.4	-0.9	1.6	2.8	-1.2
2015	6.2	6.9	-0.7	1.3	2.4	-1.2	1.6	3.1	-1.5
2016	6.3	7.7	-1.4	1.6	2.3	-0.6	1.1	3.0	-2.0
2017	6.3	7.8	-1.5	2.2	2.3	-0.1	1.0	3.0	-2.0
2018	6.3	7.8	-1.5	1.5	2.6	-1.1	1.9	3.3	-1.4
2019	8.9	7.1	1.8	2.5	2.6	0.0	2.8	3.5	-0.6
2020	7.4	7.2	0.2	1.5	2.7	-1.3	1.5	3.8	-2.4

주: 경제활동인구자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임.
모든 수치는 가중치를 사용하여 구한 것임.

〈부표 3-6〉 KLIPS 98년 표본과 경찰조사(시 지역)와의 종사상 지위 분포 비교(기중치 비적용)

(단위: %)

	상용직			입시일용직			입금근로자			자영업주			무급가족종사자			비입금근로자		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
1998	50.5	36.0	14.5	12.0	32.2	-20.2	62.5	68.2	-5.7	29.6	25.1	4.6	7.9	6.8	1.1	37.5	31.8	5.7
1999	47.1	32.5	14.6	16.8	34.7	-17.9	64.0	67.2	-3.2	27.5	25.7	1.8	8.5	7.1	1.5	36.0	32.8	3.2
2000	49.5	30.7	18.8	14.9	33.6	-18.7	64.4	64.2	0.1	26.9	27.1	-0.2	8.7	8.6	0.1	35.6	35.8	-0.1
2001	51.4	33.3	18.1	14.6	35.2	-20.6	65.9	68.5	-2.6	25.5	25.3	0.2	8.6	6.3	2.3	34.1	31.5	2.6
2002	51.9	32.9	19.0	14.8	36.0	-21.2	66.8	68.9	-2.2	25.6	25.2	0.5	7.6	5.9	1.7	33.2	31.1	2.2
2003	53.5	35.2	18.2	14.3	34.8	-20.5	67.8	70.0	-2.2	24.5	24.4	0.0	7.7	5.5	2.2	32.2	30.0	2.2
2004	52.7	36.0	16.6	15.1	34.6	-19.5	67.7	70.6	-2.9	24.8	24.4	0.4	7.5	5.0	2.5	32.3	29.4	2.9
2005	51.0	37.1	14.0	14.8	34.2	-19.5	65.8	71.3	-5.5	27.4	24.0	3.4	6.8	4.7	2.1	34.2	28.7	5.5
2006	52.7	38.0	14.7	15.0	34.0	-19.0	67.7	72.0	-4.3	25.8	23.4	2.3	6.5	4.5	2.0	32.3	28.0	4.3
2007	53.2	39.4	13.7	14.8	33.5	-18.6	68.0	72.9	-4.9	25.8	22.9	3.0	6.1	4.2	1.9	32.0	27.1	4.9
2008	53.2	40.9	12.3	14.9	32.4	-17.5	68.1	73.3	-5.2	25.7	22.4	3.3	6.2	4.2	2.0	31.9	26.7	5.2
2009	52.2	42.7	9.5	16.4	31.6	-15.2	68.5	74.3	-5.7	25.4	21.6	3.8	6.0	4.1	2.0	31.5	25.7	5.7
2010	52.0	44.9	7.1	17.6	30.3	-12.7	69.6	75.2	-5.6	24.6	21.0	3.6	5.8	3.9	2.0	30.4	24.8	5.6
2011	52.6	46.5	6.1	18.1	29.0	-10.9	70.7	75.5	-4.8	23.9	20.7	3.2	5.4	3.8	1.6	29.3	24.5	4.8
2012	52.2	47.6	4.6	18.2	27.9	-9.7	70.5	75.6	-5.1	24.2	20.8	3.4	5.3	3.6	1.7	29.5	24.4	5.1
2013	52.6	49.4	3.2	18.4	26.8	-8.4	71.0	76.3	-5.3	23.8	20.3	3.5	5.2	3.4	1.8	29.0	23.7	5.3
2014	52.4	50.0	2.4	18.6	26.7	-8.2	70.9	76.7	-5.8	23.8	19.9	3.9	5.3	3.4	1.9	29.1	23.3	5.8
2015	53.8	50.9	2.9	18.1	26.4	-8.3	71.9	77.3	-5.4	23.2	19.5	3.7	4.9	3.2	1.7	28.1	22.7	5.4
2016	53.8	51.6	2.2	18.4	25.9	-7.5	72.2	77.6	-5.3	23.1	19.3	3.8	4.7	3.2	1.5	27.8	22.4	5.3
2017	54.0	52.5	1.5	19.1	25.1	-6.0	73.1	77.7	-4.5	22.3	19.2	3.0	4.6	3.1	1.5	26.9	22.3	4.5
2018	55.0	53.9	1.0	18.7	24.2	-5.5	73.7	78.2	-4.5	21.8	18.9	2.9	4.5	2.9	1.6	26.3	21.8	4.5
2019	55.4	55.1	0.2	19.1	23.6	-4.5	74.5	78.8	-4.3	21.4	18.5	2.8	4.2	2.7	1.5	25.5	21.2	4.3
2020	56.0	57.0	-1.0	18.6	22.2	-3.6	74.6	79.1	-4.5	21.4	18.3	3.1	4.0	2.5	1.5	25.4	20.9	4.5

주: 자영업주는 고용주와 자영업자를 합한 것임. KLIPS는 기중치를 사용하지 않은 것임. 차이는 KLIPS-경찰조사.

〈부표 3-7〉 KLIPS 09년 표본과 경찰조사와의 연령별 경제활동참가율 비교

(단위:%)

	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2009	46.3	52.1	-5.8	75.8	77.5	-1.7	41.5	45.4	-3.9
2010	46.2	52.4	-6.1	77.5	77.8	-0.3	41.5	45.6	-4.1
2011	47.8	52.5	-4.7	79.0	78.0	1.0	43.2	46.5	-3.3
2012	46.6	52.8	-6.2	78.8	78.1	0.7	45.4	47.5	-2.1
2013	45.5	52.8	-7.3	79.1	78.3	0.8	46.2	48.3	-2.1
2014	47.2	54.1	-6.9	79.3	79.2	0.0	46.8	49.6	-2.8
2015	46.4	54.8	-8.3	78.4	79.4	-1.0	47.5	49.7	-2.1
2016	44.8	55.1	-10.3	78.6	79.5	-0.9	49.2	50.1	-0.9
2017	45.6	55.2	-9.6	79.3	80.0	-0.7	50.1	50.7	-0.6
2018	46.6	55.6	-9.0	79.0	79.9	-0.8	50.7	50.9	-0.2
2019	47.0	56.2	-9.2	78.5	79.6	-1.1	50.8	51.7	-0.9
2020	46.6	55.2	-8.6	78.3	78.4	-0.2	50.0	52.0	-2.0

주: 모든 수치는 가중치를 사용하여 구한 것임.

〈부표 3-8〉 KLIPS 09년 표본과 경찰조사와의 연령별 실업률 비교

(단위:%)

	35세 미만			35~54세			55세 이상		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2009	8.4	6.5	1.9	3.0	2.7	0.4	2.7	1.9	0.8
2010	5.5	6.4	-1.0	1.5	2.6	-1.1	2.0	2.7	-0.7
2011	5.9	6.1	-0.2	1.6	2.3	-0.8	2.6	2.4	0.3
2012	5.9	5.7	0.1	2.2	2.2	0.0	2.2	2.4	-0.2
2013	5.4	6.1	-0.7	1.5	2.0	-0.5	1.9	1.9	-0.1
2014	6.4	6.7	-0.3	1.5	2.3	-0.8	1.3	2.4	-1.1
2015	6.3	6.7	-0.4	1.5	2.4	-0.8	1.3	2.7	-1.3
2016	6.5	7.4	-1.0	1.5	2.2	-0.7	1.5	2.6	-1.2
2017	6.2	7.6	-1.4	1.5	2.2	-0.8	1.1	2.5	-1.4
2018	5.7	7.6	-1.9	1.4	2.5	-1.0	1.8	2.9	-1.1
2019	8.1	7.1	1.0	1.6	2.4	-0.8	1.9	3.1	-1.2
2020	7.9	7.1	0.8	1.8	2.7	-0.9	1.3	3.4	-2.1

주: 모든 수치는 가중치를 사용하여 구한 것임.

〈부표 3-9〉 KLIPS 09년 표본과 경찰조사와의 종사상 지위 분포 비교

(단위: %)

	상용직			임시일용직			임금근로자			자영업주			무급가족종사자			비임금근로자		
	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이	KLIPS	경찰	차이
2009	53.0	40.0	13.0	16.9	30.0	-13.1	69.9	70.0	-0.1	24.3	24.3	0.0	5.8	5.7	0.1	30.1	30.0	0.1
2010	53.0	42.4	10.6	17.6	28.8	-11.2	70.7	71.2	-0.5	23.8	23.5	0.3	5.6	5.3	0.3	29.3	28.8	0.5
2011	53.7	44.0	9.7	18.5	27.8	-9.3	72.2	71.7	0.5	22.7	23.1	-0.4	5.1	5.2	-0.1	27.8	28.3	-0.5
2012	53.9	45.1	8.8	18.8	26.7	-7.9	72.6	71.8	0.8	22.5	23.1	-0.6	4.9	5.1	-0.2	27.4	28.2	-0.8
2013	54.5	46.8	7.7	18.4	25.8	-7.4	73.0	72.6	0.4	22.2	22.5	-0.3	4.9	4.9	0.0	27.0	27.4	-0.4
2014	54.8	47.6	7.2	18.9	25.6	-6.7	73.7	73.2	0.5	21.4	22.1	-0.7	4.8	4.7	0.1	26.3	26.8	-0.5
2015	55.7	48.6	7.1	18.6	25.5	-6.9	74.2	74.1	0.1	21.4	21.5	-0.1	4.4	4.4	0.0	25.8	25.9	-0.1
2016	55.6	49.5	6.1	18.6	25.0	-6.4	74.2	74.5	-0.3	21.3	21.3	0.0	4.4	4.3	0.1	25.8	25.5	0.3
2017	55.8	50.2	5.6	19.1	24.3	-5.2	74.9	74.6	0.3	20.8	21.3	-0.5	4.3	4.2	0.1	25.1	25.4	-0.3
2018	56.8	51.3	5.5	18.5	23.5	-5.0	75.3	74.9	0.4	20.4	21.0	-0.6	4.2	4.1	0.1	24.7	25.1	-0.4
2019	57.3	52.4	4.9	18.8	22.9	-4.1	76.2	75.4	0.8	19.9	20.7	-0.8	3.9	4.0	-0.1	23.8	24.6	-0.8
2020	58.1	54.0	4.1	18.2	21.6	-3.4	76.3	75.6	0.7	19.9	20.6	-0.7	3.7	3.9	-0.2	23.7	24.4	-0.7

주: 자영업주는 고용주와 자영업자를 합한 것임. 모든 수치는 가중치를 사용하여 구한 것임. 차이는 KLIPS-경찰조사.

〈부표 3-10〉 KLIPS 98년 표본과 경제활동인구조사(시 지역)의 고용지표 비교(15~64세)

(단위: 천 명, %)

	KLIPS						경제활동인구조사						KLIPS/경제조사					
	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률	15세 이상 인구	취업자 수	실업자 수	경활 참가율	고용률	실업률
1998	27,439	14,022	2,229	59.2	51.1	13.7	26,325	15,155	1,294	62.5	57.6	7.9	1.04	0.93	1.72	0.95	0.89	1.73
1999	27,423	15,163	1,087	59.3	55.3	6.7	27,712	16,157	1,237	62.8	58.3	7.1	0.99	0.94	0.88	0.94	0.95	0.94
2000	27,446	15,200	838	58.4	55.4	5.2	31,280	19,154	944	64.3	61.2	4.7	0.88	0.79	0.89	0.91	0.91	1.11
2001	27,515	15,487	752	59.0	56.3	4.6	26,970	16,456	781	63.9	61.0	4.5	1.02	0.94	0.96	0.92	0.92	1.02
2002	27,439	16,010	654	60.7	58.3	3.9	27,227	16,966	655	64.7	62.3	3.7	1.01	0.94	1.00	0.94	0.94	1.05
2003	27,475	15,882	794	60.7	57.8	4.8	27,536	17,031	717	64.5	61.8	4.0	1.00	0.93	1.11	0.94	0.94	1.20
2004	27,416	15,900	849	61.1	58.0	5.1	27,793	17,404	755	65.3	62.6	4.2	0.99	0.91	1.12	0.94	0.93	1.21
2005	27,184	15,651	927	61.0	57.6	5.6	28,054	17,594	785	65.5	62.7	4.3	0.97	0.89	1.18	0.93	0.92	1.30
2006	27,211	15,797	878	61.3	58.1	5.3	28,444	17,926	734	65.6	63.0	3.9	0.96	0.88	1.20	0.93	0.92	1.36
2007	27,154	15,792	728	60.8	58.2	4.4	28,816	18,231	698	65.7	63.3	3.7	0.94	0.87	1.04	0.93	0.92	1.19
2008	27,153	15,933	539	60.7	58.7	3.3	29,197	18,442	685	65.5	63.2	3.6	0.93	0.86	0.79	0.93	0.93	0.92
2009	27,199	15,917	729	61.2	58.5	4.4	29,502	18,308	777	64.7	62.1	4.1	0.92	0.87	0.94	0.95	0.94	1.07
2010	27,197	16,250	472	61.5	59.7	2.8	29,761	18,594	784	65.1	62.5	4.0	0.91	0.87	0.60	0.94	0.96	0.70
2011	27,147	16,620	447	62.9	61.2	2.6	30,106	18,997	735	65.5	63.1	3.7	0.90	0.87	0.61	0.96	0.97	0.70
2012	27,144	16,586	516	63.0	61.1	3.0	30,382	19,283	706	65.8	63.5	3.5	0.89	0.86	0.73	0.96	0.96	0.86
2013	27,099	16,609	480	63.1	61.3	2.8	30,618	19,513	688	66.0	63.7	3.4	0.89	0.85	0.70	0.96	0.96	0.82
2014	26,900	16,948	503	64.9	63.0	2.9	30,857	19,962	796	67.3	64.7	3.8	0.87	0.85	0.63	0.96	0.97	0.76
2015	26,975	16,420	441	62.5	60.9	2.6	30,992	20,189	812	67.8	65.1	3.9	0.87	0.81	0.54	0.92	0.94	0.67
2016	27,024	16,743	463	63.7	62.0	2.7	31,078	20,289	844	68.0	65.3	4.0	0.87	0.83	0.55	0.94	0.95	0.68
2017	26,825	16,774	514	64.4	62.5	3.0	31,065	20,506	843	68.7	66.0	3.9	0.86	0.82	0.61	0.94	0.95	0.77
2018	26,799	17,120	515	65.8	63.9	2.9	30,974	20,426	883	68.8	65.9	4.1	0.87	0.84	0.58	0.96	0.97	0.71
2019	26,539	16,590	776	65.4	62.5	4.5	30,937	20,456	847	68.9	66.1	4.0	0.86	0.81	0.92	0.95	0.95	1.13
2020	26,246	16,567	520	65.1	63.1	3.0	30,864	20,119	872	68.0	65.2	4.2	0.85	0.82	0.60	0.96	0.97	0.71

주: 경활자료는 원자료를 이용하여 시(동) 거주자만을 대상으로 구한 것임. 1998~1999년 실업률은 ILO 기준, 2000~2020년은 OECD 기준. 모든 수치는 가중치 부여.

◆ 執筆陣

- 장인성(한국노동연구원 선임연구위원)
- 이지은(한국노동연구원 전문위원)
- 김기민(한국노동연구원 전문위원)
- 김우영(공주대학교 교수)
- 이상혁(고려대학교 연구교수)
- 이지희(닐슨코리아 과장)
- 채민희(난카이대학교 교수)
- 최효미(육아정책연구원 연구위원)

패널자료 품질개선 연구(XI)

- | | |
|-----------|--|
| ▪ 발행연월일 | 2021년 12월 24일 인쇄
2021년 12월 30일 발행 |
| ▪ 발 행 인 | 황 덕 순 |
| ▪ 발 행 처 | 한국노동연구원
☎ 30147 세종특별자치시 시청대로 370
세종국책연구단지 경제정책동
☎ 대표 (044) 287-6080 Fax (044) 287-6089 |
| ▪ 조판·인쇄 | 거목정보산업(주) (044) 863-6566 |
| ▪ 등 록 일 자 | 1988년 9월 13일 |
| ▪ 등 록 번 호 | 제2015-000013호 |

© 한국노동연구원 2021 정가 7,000원

ISBN 979-11-260-0498-0