

# KLIPS에 나타난 혼인상태 측정의 문제점과 개선방안

최 효 미\*

## I. 들어가며

성, 연령, 학력, 혼인상태 등은 개인에 대한 가장 기본적인 정보이다. 따라서 거의 모든 조사(Survey)들은 이들 변수에 대해 조사한다. KLIPS의 경우에도 성, 연령, 학력, 혼인상태, 경제활동상태 같은 변수들은 핵심(key)변수로 구분하여 관리하고 있다. 특히 매년 동일한 가구와 개인을 추적조사하는 KLIPS와 같은 패널조사의 경우에는 key 변수가 얼마나 정확하게 조사되고 있는가 하는 문제가 곧 조사의 신뢰도와 직결되므로 더욱 중요하다.

지금까지 KLIPS 연구진은 모델로 삼을 만한 패널조사가 없는 우리나라의 환경에서 많은 시행착오를 거치며 핵심(key)변수를 정확하게 조사하고 관리하기 위해 노력해 왔다. 그러한 노력의 결과, 핵심(key)변수 중에서 성이나 연령과 같이 한번 결정되면 좀처럼 바뀌지 않는 변수들은 조사가 거듭될수록 오류가 줄어들고 있다.<sup>1)</sup> 이와는 달리 학력이나 혼인상태 등은 시간이 지남에 따라 변화하는 경우가 많다. 게다가 학력이나 혼인상태는 거짓 응답의 유인<sup>2)</sup>이 큰 변수라는 점에서 실사에서부터 클리닝 과정까지 보다 세심한 주의가 요구된다. 특히 개인의 학력은 대부분 일정시점 이후에는 변화하지 않는

\* 한국노동연구원 노동패널팀 연구원(chm@kli.re.kr).

- 1) 조사 초기 성별이나 생년월일과 같은 key 변수가 차수별로 다르게 나타나는 문제가 있었으나, 차수가 거듭되면서 이러한 문제가 거의 사라지고 있다. 7차년도 최초 자료의 경우 이전 조사와 성이 다른 경우가 7개 발견되었으며, 추가적인 확인을 통해 성별과 관련된 오류는 모두 수정할 수 있었다.
- 2) 예를 들어, 50대 주부가 모대학의 사회교육원에서 받은 교육을 정규교육으로 응답하거나 전문대학을 다니는 아들의 학력을 4년제 대학이라고 응답하는 경우와 같이 응답자들은 자신이나 가족의 학력을 높게 응답하려는 경향을 갖고 있다.

데 반해 혼인상태는 언제든지 바뀔 수 있다. 그러므로 혼인과 관련된 오류는 발견하는 것 자체도 어려울 뿐 아니라 수정하기에도 애매한 경우가 많다.

혼인상태는 대부분의 연구에 사용되는 주요변수일 뿐 아니라 혼인상태의 변화 자체가 인구동태 변화와 관련된 연구에서 핵심적인 지표이다. 그러므로 혼인상태와 관련해서는 횡단적인 변수의 일치성뿐만 아니라 종단적인 논리적 합리성이 확보되어야 할 필요가 있다. 이에 본고는 간단하게 핵심(key)변수 클리닝 과정을 소개하고, 현재 KLIPS에 나타나 있는 혼인관련 변수들의 문제점을 살펴본 뒤, 그 해결방안을 모색해 보고자 한다.

## II. 핵심(Key)변수의 클리닝 과정

성, 연령, 학력, 종사상 지위, 혼인상태와 같은 핵심변수들은 실사가 시작될 때부터 데이터가 공개될 때까지 여러 단계의 클리닝 과정을 거치게 된다.

가장 먼저 실사 시작과 함께 면접원들은 각 가구에 대한 info-sheet를 받게 된다. info-sheet는 핵심(key)변수에 관해 해당가구가 이전 조사에서 응답한 내용을 수록하고 있다. 따라서 면접원들은 응답자가 info-sheet 내용과 다른 응답을 할 경우, 현장에서 재 질문을 통해 정확한 응답을 받을 수 있게 된다. 만약 응답자가 이전 조사의 내용이 틀렸다고 응답하는 경우에 면접원은 가구설문의 특기사항 혹은 면접원 기입지에 달라진 응답내용과 그 이유를 명기하도록 하고 있다.

둘째, 핵심변수를 포함한 모든 변수들은 실사기간내내 작성된 설문지의 점검과정을 통해 클리닝된다. 또한 자료를 입력하는 과정에서 더블펀치<sup>3)</sup>를 함으로써 펀치에러를 줄이고 있다.

셋째, 입력이 끝난 원자료는 다시 최종차수<sup>4)</sup> 응답내용과 당해연도 응답내용을 비교하여 오류가 있는지를 다시 확인한다. 이 과정에서 발견된 오류는 설문지 확인과 재조사 등을 통해 다시 수정하게 된다. 이때 핵심(Key)변수는 자료의 일관성을 위해 이전차수 응답 내용까지 모두 소급하여 수정하는 것을 원칙으로 하고 있다. 따라서 조사차수가 늘어날

3) 동일한 설문을 각기 다른 사람으로 하여금 입력하도록 하는 과정으로, 두번째 입력하는 사람이 첫 번째 입력내용과 다른 내용을 입력할 경우 화면에 오류메시지가 뜨며 이후 올바른 값을 선택하여 입력하도록 하는 시스템이다.

4) KLIPS의 설문 문항은 “지난조사 이후”의 변화내용을 묻는 경우가 많다. 따라서 조사 당해연도를 제외하고 응답자가 가장 최근에 응답한 내용이 최종차수 응답내용이 되며, info-sheet에 수록되는 내용도 최종차수 응답이다.

수록 수정해야 하는 개체수는 점점 늘어나게 될 것이다. 즉 1차년도에 여성이라고 응답한 개인이 2차년도 확인결과 남성인 것이 확인되었다면, 1차년도의 성만 수정하면 된다. 그러나 만약 7차년도에 와서야 남성임이 확인되었다면, 1~6차까지 모든 성별을 수정해야 할 것이다. 따라서 조사차수가 더해지기 전에 오류가 있는 가구와 개인의 정보를 가능한 빨리 발견해내는 것이 클리닝 시간과 노력을 절약하는 관건이 될 것이다.

### Ⅲ 혼인상태 측정상의 문제점

#### 1. 사실혼 포함에 따른 문제

여기에서는 KLIPS에서 혼인상태가 어떻게 측정되고 있는지 알아보도록 하자. 현재 KLIPS는 실사가이드를 통해 기혼자의 범주에 사실혼을 포함하고 있다. 반면, 통계청의 인구동태조사는 시·구청 또는 읍·면·동사무소에 제출된 혼인(혹은 이혼)신고서를 기준으로 하고 있으므로 사실상 법적 혼인상태만을 조사하고 있다고 볼 수 있다.<sup>5)</sup> 그러나 노동시장 참여에 대한 의사결정을 비롯하여 가구단위에서 이루어지는 많은 의사결정들은 법적인 상태가 아니라 실제 배우자 존재여부에 따라 영향을 받을 것이다. 따라서 사실혼을 기혼으로 간주하는 현행 KLIPS의 기준은 가구단위에서 이루어지는 다양한 의사결정에 대해 분석하려는 조사의 목적에 부합한다.

그러나 이러한 이점에도 불구하고 사실혼을 기혼으로 받아들임으로 인해 KLIPS는 여러 가지 복잡한 문제에 봉착하게 된다. 첫째로 사실혼 관계에 있었던 응답자들은 사실혼이 종료되었을 때, 기존의 사실혼 자체를 없었던 일로 하려는 경향을 가지고 있다.

둘째로 응답자가 ‘현재는 재결합 의사가 없고 완전히 혼인관계가 종료되었다’고 생각하여 이혼한 것으로 응답했다고 하자. 그러나 시간이 지남에 따라 응답자가 동일한 배우자와 다시 사실혼 관계로 회귀할 경우 재혼으로 보아야 할 것인지 재결합으로 보아야 할 것인지 모호해지게 된다. 이는 사실혼의 종료라는 것이 법적 이혼이 아닌 주관적 판단에 따른 기준이기 때문이다. 따라서 사실혼 종료의 기준을 어떻게 설정할 것인가는 매우 중요한 문제이다.

현재 KLIPS는 사실혼의 종료와 관련하여 “오랜기간에 걸쳐 숙고한 끝에 혼인상태를

5) 혼인 혹은 이혼신고서 작성시 사실상 혼인관계(이혼관계)가 성립된 시기를 기록하도록 하고 있으나, 동거의 이유로 사실혼 관계에 있으면서 신고하지 않은 경우에는 조사에서 누락되게 된다.

더 이상 유지할 수 없다고 판단하여 배우자와 함께 살고 있지 않은 경우”라는 가이드를 가지고 있다. 그러나 이러한 기준은 자의적일 수 있고 별거와의 구분 또한 모호하다는 단점이 있다. 다만 한번 “이혼이라고 응답한 후에는 같은 사람과 사실혼 관계가 다시 성립한다 하더라도 재혼”으로 파악한다는 가이드를 가지고 있다. 이에 따라 사실혼이 종료되어 이혼이라고 응답한 개인이 몇 년 후 ‘혼인상태는 변화가 없었다’고 응답하고 기혼으로 회귀해 버리는 오류를 없앨 수 있다.

## 2. 설문 항목의 부재

한편, 현재 설문 문항은 별거중이던 응답자가 재결합을 할 경우 오류 발생이 불가피한 구조를 지니고 있다. 즉, 별거중이던 응답자가 이전의 배우자와 재결합할 경우 혼인상태는 별거에서 기혼으로 바뀌지만, ‘혼인상태의 변화’를 묻는 질문에는 변화하지 않은 것으로 응답할 수밖에 없어 양자가 일치하지 않는 문제를 유발시키고 있다.<sup>6)</sup>

## 3. 응답 오류

이외에 응답자가 의도하지는 않았으나 응답자에 의해 발생된 오류도 상당수 존재한다. 즉 실제 혼인상태가 변화하지 않았으나 응답자가 혼인상태를 매해 다르게 응답하는 경우를 말한다. 예를 들면, 배우자의 가출로 인해 오랫동안 배우자와 비동거중인 응답자가 혼인상태에 변화가 없음에도 불구하고 별거-이혼-사별 등을 왔다갔다 하는 경우를 들 수 있다. 또 몇 년 전에 이혼한 배우자가 사망했을 경우, 응답자는 자신의 혼인상태를 이혼이 아닌 사별로 응답함으로써 오류를 발생시키기도 한다.

이와 관련해서는 배우자의 오랜 가출 등으로 배우자의 생사를 모른다거나 이혼한 것과 마찬가지로 상태라 하더라도, 배우자의 사망이 확인되거나 재결합 가능성이 전혀 없는 경우가 아니라면 ‘별거’로 간주한다는 기준을 7차년도 조사부터 적용하고 있다. 그리고 이혼 후에 배우자가 사망했다면 응답자의 혼인상태는 사별이 아닌 ‘이혼’으로 보고 클리닝하고 있다.

## 4. 거짓 응답

거짓 응답이란 응답자가 일부러 사실과 다른 응답을 하는 것을 의미한다. 앞서 지적

6) 현재 별거후 재결합의 경우에 변화내용 항목 중 응답할 만한 항목이 없다.

한 것처럼 사실혼이 종료되었을 경우 기존의 사실혼 자체를 없었던 일로 하려는 의도에서 본인의 혼인상태를 미혼이라고 응답하는 경우가 이에 포함된다. 이외에도 이혼을 경험한 응답자가 의도적으로 미혼 혹은 사별이라고 응답하는 경우도 상당히 존재한다.

그러나 이러한 문제를 실사과정에서 엄격하게 통제하기는 어려운데, 이는 정확한 응답을 강요할 경우 차기년도에 해당 응답자가 패널에서 이탈하는 현상이 발생할 수 있기 때문이다.

#### IV. KLIPS의 혼인상태 오류 및 클리닝 방안

##### 1. KLIPS의 혼인상태 오류현황

KLIPS의 혼인상태는 차수별로 항목값이 다르기 때문에 주의가 요구된다. 이에 필자는 해당차수별로 혼인상태를 ‘1.미혼, 2.기혼, 3.별거, 4.이혼, 5.사별’로 항목 값을 통일시킨 후 어느 정도의 오류가 있는지를 관찰하였다.

<표 1>은 혼인상태 변수에 존재하는 오류의 종류와 빈도를 잘 보여준다.<sup>7)</sup> 그런데 주의할 점은 KLIPS의 혼인상태가 크게 1~2차와 3~4차, 5차 이후로 구조가 변경되었다는 점이다. 그러므로 각 시점별로 조금 다른 방식으로 오류를 확인할 필요가 있었다. 따라서 <표 1>에 제시된 “—” 부분은 해당 사유를 구성할 수 없는 경우를 의미하며, 그러한 유형의 오류가 없음을 의미하는 것이 아니다.

그 결과, 1~2차년도의 경우에는 오류발생률이 상대적으로 낮은 것으로 나타났다. 이는 조사 초기라는 점에 기인하는 것으로, 데이터상에서 그것을 오류로 볼 만한 비교집단이 존재하지 않음을 의미한다. 다만, 1차년도와 2차년도 데이터가 비교적 동일한 항목에 대해 질문하고 있으므로, 2개 차수를 연결할 경우 총 76개의 데이터 불일치를 발견할 수 있었다.

또한 3차년도의 오류발생률이 낮은 이유는 3차년도 조사에서 이전 조사 이후 변화여부만을 질문하고 혼인상태를 질문하지 않았기 때문에 변화가 없다고 응답한 경우는 모두 오류에서 제외되기 때문이다.

7) 오류발생률은 1차 클리닝이 완료된 파일을 기준으로 최종차수 응답과 당해년도 응답을 비교한 값이다. 즉, 6차년도 오류발생률은 5차년도까지의 응답 중 가장 최근 응답을 기준으로 6차년도 자료의 오류 비중을 본 것이다.

반면, 4차년도와 5차년도 자료에서 나타나는 높은 오류발생률은 설문구조 변경으로 인해 조사당시 발견이 불가능했던 3차년도 자료의 오류가 이후 조사까지 누적적으로 이어졌기 때문인 것으로 보인다.

그러나 6차년도 조사에서 오류발생률이 급격하게 감소하였고, 7차년도에서는 더욱 낮아졌다. 이는 6차 이후 혼인상태와 관련된 오류를 줄이려는 노력이 집중적으로 이루어졌기 때문이다. 그러나 <표 1>에 제시된 오류발생률은 최종차수 응답과 당해연도 응답을 비교한 값이므로, 5차년도 이전의 자료가 가지고 있는 오류를 감안하면 6차년도와 7차년도에 실제로 존재하는 오류는 이보다는 큰 수치일 것이다.<sup>8)</sup>

<표 1> 혼인상태에서 발견되는 오류 종류 및 빈도

(단위: 명, %)

		1차	2차	3차	4차	5차	6차	7차
횡단면	설문지 선택 오류	-	6	17	49	10	1	0
오류	혼인상태와 변화내용 불일치	2	38	-	12	0	12	0
종단면	혼인상태가 변화하지 않았다고 응답했으나 전년과 다름	-	-	-	215	317	39	47
	혼인상태가 변화했다고 응답했으나 전년과 동일	-	-	9	12	7	24	1
	1차-2차년도 자료간 불일치	-	76	-	-	-	-	-
합 계		27	95	26	288	334	76	48
전체 개인수		13,321	12,037	11,205	11,051	10,966	11,541	11,661
오류발생률		0.2	0.8	0.2	2.6	3.0	0.6	0.4

주: 1) 6차년도 release version. 단, 7차년도는 7차년도 학술대회 version임.

2) - 표시는 설문구조상 해당 사유구성이 불가능한 경우를 의미함.

## 2. 데이터 클리닝 기준 설정 및 적용

혼인상태와 관련된 오류를 수정하는데 가장 어려운 점은 오류 유형이 매우 다양하다는 점이다. 가장 안정적이라고 볼 수 있는 7차년도 자료에서조차도 혼인상태와 관련한 오류의 유형은 14가지나 있다(표 2 참조).

통상 데이터 클리닝은 응답자에게 동일한 질문을 다시 하여 정확한 응답을 얻어내는 것을 원칙으로 하고 있다. 그러나 혼인상태의 경우 최근 조사에서 거짓 응답을 한 응답자는 재질문에 대해서도 거짓 응답을 할 가능성이 매우 높으므로 재질문 방식이 반드시 좋은 클리닝 기준이라고 하기도 어렵다.<sup>9)</sup> 게다가 혼인상태에 대한 정확한 응답을 꺼리

8) 이후에 제시된 기준에 따라 1차년도부터 순차적으로 자료를 클리닝할 경우 6차년도에는 총 92명, 7차년도에는 67명의 개인의 혼인상태가 수정된다.

는 응답자에게 집요하게 정확한 응답을 강요할 경우 표본 탈락의 위험이 있다.

<표 2> 7차년도 조사당시 발견된 오류 유형

(단위: 명, %)

최종차수 응답내용	7차 조사 응답내용	
	변화여부	혼인상태
미 혼	변화없음	기혼
		이혼
		사별
기 혼		별거
		미혼
		기혼
별 거		이혼
		미혼
		기혼
이 혼		별거
	미혼	
	기혼	
사 별	미혼	
	기혼	
	이혼	
기혼/별거/이혼/사별	변화했음	최종차수 응답과 동일

그러므로 이전차수 혼인상태 응답내용과 가구정보 등을 조합하여 각각의 경우를 클리닝하는 것이 현실적인 대안이라 볼 수 있다. 본고가 혼인상태 클리닝과 관련하여 선택한 기준은 ① 정보의 양이 많은 것<sup>10)</sup>을 지속적 응답보다 우선시하고, ② 정보의 양이 동일할 경우에는 지속적 응답을 바른 응답으로 간주하였다. 단, ③ 이혼 후 사별한 경우는 사별이 아닌 이혼으로 혼인상태를 클리닝하였으며, ④ 별거 후 동일한 배우자와 재결합한 경우<sup>11)</sup>에는 변화내용에 6번을 삽입하여 ‘별거 후 재결합’으로 구분하는 등의 기준을 적용하였다. 특히 ①번의 기준과 관련하여 결혼이나 이혼, 사별 등의 년월을 응답한 응답자가 이듬해부터 지속적으로 미혼이라고 응답하는 경우, 미혼이라고 거짓말 할 유인이 결혼이나 이혼 등으로 응답할 유인보다는 크다는 판단에 기인한다.

<표 3>과 <표 4>를 통해 알 수 있듯이 클리닝 전에 비해 클리닝 후에는 미혼 혹은 기혼의 비중이 다소 감소한 반면, 이혼이나 별거의 비중이 증가하였음을 알 수 있다. 7차

9) 다만, 패널조사의 특성상 다음 연도에도 응답자는 동일한 질문을 받게 될 것이므로, 이 과정에서 당시의 정보가 틀린 것이었다는 면접원의 확인기록이 있을 때에는 이를 최우선으로 클리닝하고 있다.

10) 변화의 내용, 변화한 년월 등이 기록되어 있는 경우를 의미한다.

11) 이전차수에 별거였다가 다음차수에 기혼으로 변한 경우는 크게 재혼한 경우와 재결합한 경우로 나뉜다. 따라서, 별거 이후 기혼으로 변화했다고 하여 일률적인 기준을 적용해서는 안된다. 그러므로 이 경우 개인이 속한 가구 데이터의 내용을 조사하여 동거중인 배우자가 누구인지를 판별한 다음, 이혼(혹은 사별) 후 재혼이나 별거 후 재결합으로 수정하였다.

년도의 경우 클리닝 후 미혼은 7명, 기혼은 18명, 사별은 7명이 감소하였으며, 별거와 이혼은 각각 5명, 13명 증가하였다. 아울러 별거 후 재결합한 응답자가 12명, 혼인상태 구분이 불분명한 응답자가 2명 있었다.

〈표 3〉 1차 클리닝 완료 자료의 1~7차년도 혼인상태 빈도

(단위: 명, %)

	1차	2차	3차	4차	5차	6차	7차
미 혼	3,854(29.0)	3,426(28.5)	3,164(28.2)	3,069(27.8)	2,970(27.1)	3,156(27.4)	3,130(26.8)
기 혼	8,286(62.4)	7,619(63.3)	7,061(63.0)	6,956(62.9)	6,897(62.9)	7,201(62.4)	7,301(62.6)
별 거	134( 1.0)	75( 0.6)	71( 0.6)	53( 0.5)	50( 0.5)	62( 0.5)	71( 0.6)
이 혼	148( 1.1)	129( 1.1)	130( 1.2)	127( 1.2)	176( 1.6)	221( 1.9)	244( 2.1)
사 별	854( 6.4)	788( 6.6)	776( 6.9)	754( 6.7)	873( 8.0)	901( 7.8)	915( 7.9)
구분모호	12( 0.1)	-	2( 0.0)	101( 0.9)	-	-	-
전 체	13,298	12,037	11,204	11,051	10,966	11,541	11,661

- 주: 1) 6차년도 release version. 단, 7차년도는 7차년도 학술대회 version임.  
 2) '1차 클리닝 자료'란 최종차수 응답과 당해연도 응답을 비교할 때 발견된 오류를 수정한 것으로, 여기에서는 제II장에서 언급된 클리닝 과정이 모두 끝난 상태의 파일을 의미함.  
 3) 구분모호란 '변화내용을 모름'이라고 응답한 경우 혹은 1차년도 혼인상태 응답거부로 인해 별거/이혼/사별 중 어떤 것인지 구분할 수 없는 경우로 모두 기혼무배우자에 해당함.

〈표 4〉 논리적 오류 클리닝 후 1~7차년도 혼인상태 빈도

(단위: 명, %)

	1차	2차	3차	4차	5차	6차	7차
미 혼	3,845(28.9)	3,415(28.4)	3,169(28.3)	3,082(27.9)	2,970(27.1)	3,147(27.3)	3,123(26.8)
기 혼	8,305(62.4)	7,498(62.3)	6,953(62.2)	6,894(62.4)	6,879(62.7)	7,180(62.2)	7,283(62.5)
별 거	134( 1.0)	91( 0.8)	73( 0.7)	66( 0.6)	49( 0.5)	63( 0.6)	76( 0.7)
이 혼	146( 1.1)	148( 1.2)	142( 1.3)	159( 1.4)	194( 1.8)	233( 2.1)	257( 2.2)
사 별	856( 6.4)	884( 7.3)	847( 7.6)	844( 7.6)	858( 7.8)	900( 7.8)	908( 7.8)
재 결 합	-	-	-	3( 0.0)	14( 0.1)	16( 0.1)	12( 0.1)
구분모호	28( 0.2)	1( 0.0)	3( 0.0)	3( 0.0)	2( 0.0)	2( 0.0)	2( 0.0)
전 체	13,314	12,037	11,204	11,051	10,966	11,541	11,661

- 주: 1) 6차년도 release version. 단, 7차년도는 7차년도 학술대회 version임.  
 2) '논리적 오류 클리닝'이란 제IV장 2항에서 제시된 내용을 기준으로 클리닝을 완료한 상태를 의미함.  
 3) 1차년도 개인설문에 응답하였으나, 혼인상태와 관련하여 응답을 하지 않은 개인이 23명 존재하였는데, 클리닝으로 결혼시기 등을 알 수 있었던 경우가 16건 있었음. 1차년도 개인의 수가 일치하지 않는 이유는 이 때문임.  
 4) 구분모호란 '변화내용을 모름'이라고 응답한 경우 혹은 1차년도 혼인상태 응답거부로 인해 별거/이혼/사별 중 어떤 것인지 구분할 수 없는 경우로 모두 기혼무배우자에 해당함.

한편, 7차년도 기준으로 클리닝 후 기혼 유배우자는 전체 개인의 62.5%인 7,283명이었으며, 미혼/별거/이혼/사별이 각각 3,123명, 76명, 257명, 908명인 것으로 조사되었다.



## V. KLIPS의 혼인상태와 인구동태조사와의 비교

이 장에서는 클리닝 후 혼인력 파일을 가지고 KLIPS에 나타난 혼인상태의 특징에 대해서 좀 더 자세하게 살펴보도록 하자. <표 5>는 우리나라 혼인에 관한 공식통계라 할 수 있는 통계청의 인구동태조사와 KLIPS의 혼인율<sup>12)</sup>을 비교한 표이다.

이 표에서 보는 것처럼 KLIPS의 경우 2001년의 혼인율은 9.8%이었으며, 2002년에 다소 감소하였으나, 2004년에는 다시 9.8%를 나타냈다.

또 인구동태조사에 비해 KLIPS의 혼인율이 1.5배 가량 높다. 이는 혼인율 산출방식의 차이 때문인 것으로 보인다. 인구동태조사의 혼인율은 혼인건수를 15세 이상 인구로 나눈 후 100을 곱하여 산출된다. 다시 말해 분모인 15세 이상 인구에는 각각의 개인이 모두 합산되지만 분자인 혼인건수는 2명당 1건만 합산되기 때문에 혼인율이 과소추정된다. 이에 더하여 KLIPS가 사실혼 관계를 혼인으로 보고 있기 때문에 인구동태조사에 비해 높은 혼인율을 보이게 된다.<sup>13)</sup>

그런데 KLIPS의 혼인율이 인구동태조사의 2배가 되지 않는 이유는 크게 두 가지를 생각할 수 있다. 첫째는 KLIPS의 조사 원칙상 개인용에 15세 이상 모든 가구원이 응답

<표 5> KLIPS와 인구동태조사의 혼인율 비교

(단위: 명, 건, %)

	KLIPS		인구동태조사	
	처음결혼/재혼	혼인율	혼인건수	혼인율
2001	307,745	9.8	320,063	6.7
2002	255,220	8.0	306,573	6.4
2003	314,106	9.7	304,932	6.3
2004	342,172	9.8	310,944	6.4

- 주: 1) 6차년도 release version 단, 7차년도는 7차년도 학술대회 version임.  
 2) 인구동태조사 혼인율은 15세 이상 인구 1,000명당 혼인건수의 비율임.  
 3) KLIPS의 혼인건수는 횡단면 가중치를 적용한 값임.

12) KLIPS는 현재의 혼인상태와 지난조사 이후 혼인상태 변화내용을 묻고 있다. 따라서 분고는 지난해에 응답한 개인의 경우에는 당해연도 조사에서 응답한 경우를 모두 당해연도 혼인건수로 계산하였다. 그러나 지난해 조사에 응답하지 않은 개인의 경우에는 지난 조사 종료시점인 9월을 기준으로 지난해 10월부터 올해 9월까지 혼인한 건수만을 당해연도 혼인건수로 계산하였다.

13) 인구동태조사의 경우, 신고건수를 기준으로 혼인건수를 산출하므로 사실혼이 제외되어 있을 가능성이 매우 높다.

하는 것이 맞겠으나, 부득이 부부 중 한쪽만이 응답한 경우가 존재하기 때문이다.<sup>14)</sup> 둘째, 원표본 개인이 혼인으로 인해 분가할 경우 이후 조사에서 이탈되어 혼인건수에서 누락되는 표본들이 존재하기 때문인 것으로 보인다.<sup>15)</sup>

## VI 맺음말

지금까지 KLIPS의 혼인상태 관련변수에 나타난 오류와 혼인상태의 특징에 대해 살펴 보았다. KLIPS는 사실혼 관계를 혼인상태에 포함시키고 있으며 이로 인한 몇 가지 문제가 발생하고 있었다. 즉, 응답자가 사실혼 종료 이후 사실혼 경험을 부정하는 경향을 보인다거나 시간이 지남에 따라 동일한 배우자와 사실혼 관계로 회귀하는 등의 이유로 혼인력 데이터에 오류가 발생하고 있었다. 또한 설문구조상의 문제로 별거 후 재결합과 같은 경우 응답할 만한 항목 값이 존재하지 않아 오류를 일으키고 있었다. 이외에도 응답자가 의도적으로 거짓 응답을 하는 경우가 있어 혼인력 데이터의 클리닝을 더욱 어렵게 만들고 있었다.

그러나 다행스러운 것은 앞서 소개한 클리닝 노력의 결과 조사차수가 거듭될수록 혼인상태와 관련된 변수들의 오류발생률이 낮아지고 있으며, 향후 지속적인 노력이 이뤄질 경우 현저하게 오류를 줄일 수 있을 것이라는 사실이다.

이를 위해서는 KLIPS는 첫째, 연구자의 편의와 연구의 신뢰도를 높이기 위해 통일된 형태의 혼인력 파일을 제공할 필요가 있다. 그러나 클리닝으로 인해 오히려 데이터가 잘못될 가능성을 부정할 수 없으므로, 원자료의 혼인력은 그대로 놔둔 채 추가적인 정보를 제공하는 방식<sup>16)</sup>으로 클리닝 데이터를 활용하는 것이 보다 바람직할 것으로 보인다. 추가적인 정보제공 방식으로는 개인용 자료에 클리닝된 혼인력 관련 변수를 삽입한다거나 별도의 패치파일<sup>17)</sup>을 구성하여 제공하는 방법 등이 있다.

둘째, 여러 차례 응답오류가 발생했던 개인이나 거짓 응답할 가능성이 높은 개인들은

14) 7차년도에 조사된 기혼 유배우자 중 가구주만 응답한 경우가 48건(1.4%), 가구주의 배우자만 응답한 경우는 38건(1.1%)이다.

15) KLIPS의 7차년도 분가가구조사 성공률은 58.7%이다.

16) 반면, 현재 성별, 생년월일, 학력과 같은 변수들은 별도의 확인과정을 거친 후 클리닝된 변수값만을 제공하고 있다.

17) 현재 KLIPS 홈페이지에는 Key File(1차~4차), 학과코드, 직업 및 산업코드 전환프로그램 등의 패치 파일이 수록되어 있다.

별도의 파일로 관리한다면 실사과정에서의 오류를 줄이고 클리닝에 소요되는 시간을 단축시킬 수 있을 것이다.

셋째, 잦은 설문구조 변경으로 인해 혼란이 야기된 점을 감안하여 설문구조의 변경은 신중하게 고려되어야 할 것이다. 그럼에도 불구하고 변화내용에 ‘별거 후 재결합’을 포함시킨다거나 같은 차수 안에서는 항목 값을 일치시키는 등의 설문지 수정은 불가피할 것으로 보인다. **표 11**

## [별 첨] KLIPS의 혼인관련 설문 의 구조

KLIPS의 혼인관련 문항들은 5차년도 이후 비교적 유사한 틀을 유지하며 지속되어 오고 있으나, 1~4차년도 조사에서는 문항이 계속 바뀌고 있어 연구자의 주의가 요구된다. 게다가 문항의 통일이 이뤄지지 않아 연구자들이 혼인력 데이터를 구성할 때 다소의 어려움이나 수고스러움을 겪고 있다.

KLIPS의 혼인관련 변수의 구성은 1차년도에는 혼인경험을 기준으로 질문하였고, 4차년도에는 “미혼, 기혼유배우, 기혼무배우”로 구분하여 질문하였다(<표 6>과 <표 7> 참조).

반면, 2차와 5~7차년도 설문의 혼인상태는 “미혼, 기혼, 별거, 이혼, 사별”의 5개 항목으로 구분하여 응답하도록 되어 있으나, “별거”와 “이혼”의 번호가 차수별로 조금씩 다르다. 즉, 2차년도에는 “1.미혼, 2.기혼, 3.별거, 4.이혼, 5.사별”의 순으로 질문하였으나, 5~7차년도 조사에서는 “1.미혼, 2.기혼, 3.이혼, 4.별거, 5.사별” 순으로 질문하였다.

게다가 5~7차년도 설문의 경우에는 개인용의 혼인상태는 ‘3.이혼, 4.별거’를 의미하지만, 혼인상태 변화내용에서는 ‘3.별거, 4.이혼’으로 질문하고 있어 같은 차수 안에서도 항목 값이 일치하지 않은 상태이다. 더욱이 신규용 설문의 혼인상태 문항은 ‘3.별거, 4.이혼’으로 되어 있으므로 혼동을 일으킬 위험이 크다.<sup>18)</sup>

한편, 3차년도의 경우에는 혼인상태의 변화와 관련된 질문만 있고 현재의 혼인상태를 질문하지 않고 있어 3차년도 혼인상태를 알기 위해서는 1차년도와 2차년도 자료를 이용하여 새로운 변수를 생성해야만 한다.

더욱이 4차년도의 경우에는 개인용이 아닌 가구용 설문에서 가구원의 혼인상태를 물었으며, 4차년도 개인용 설문에서는 3차년도와 마찬가지로 변화와 관련된 내용만을 질문하였다.

따라서 연구자들은 이렇듯 복잡한 설문의 구조를 잘 이해한 후 데이터를 사용하여야 데이터 핸들링 과정에서 발생하는 어려움을 줄이고 연구의 신뢰성을 확보할 수 있을 것이다.

18) 개인용 설문과 신규용 설문의 항목 값이 다른 문제는 7차년도 학술대회 버전부터 개인용에 신규용 자료를 통합하여 제공하고 있으므로, 이에 따른 연구자들의 불편은 해소시킬 수 있을 것으로 보인다.

〈표 6〉 1~7차년도 혼인상태 변수

	1차	2차	3차	4차	5차	6차	7차
변수명	P01521	P02185 N02432	-	P04013 N04013 A04013	P05331 N05893	P06335 N06818	P07310
항목값	1.미혼 2.결혼하고 있거나 결혼 한적 있음	1.미혼 2.기혼유배우 3.별거 4.이혼 5.사별	-	1.미혼 2.기혼유배우 3.기혼무배우	1.미혼 3.이혼	<개인용 > 2.기혼유배우 4.별거 5.사별 <신규용> 1.미혼 2.기혼유배우 3.별거 4.이혼 5.사별	

주 : 6차년도 release version 기준. 단, 7차년도는 7차년도 학술대회 version 기준변수임.

〈표 7〉 1~7차년도 혼인관련 질문내용

	1차	2차	3차	4차	5차	6차	7차
혼인 시기 질문	초혼시기(년/월/만나이) 초혼배우자 동거여부 초혼종료사유 초혼종료시기(년) 재혼여부 재혼시기(년)	초혼시기(년/월/만나이) 초혼배우자 동거여부 초혼종료사유 초혼종료시기(년/월) 재혼여부 재혼시기(년/월)	-	-	-	-	-
		총결혼횟수 최근결혼시기(년/월) 최근배우자동거여부 최근배우자비동거사유 최근배우자비동거시기 (년/월)					
혼인 종료 사유		1. 별 거 2. 사 별 3. 이 혼					
변화 여부 질문		-	-	변화여부 변화횟수 변화내용 변화시기(년/월)			
	변화 내용	-	-	1. 처음결혼 2. 기 혼 3. 별 거 4. 이 혼 5. 사 별			

주 : 1) 6차년도 release version 기준. 단, 7차년도는 7차년도 학술대회 version 기준변수임.

2) 변화내용, 변화시기(년), 변화시기(월)은 최대 5번까지 응답가능하도록 하였음.