한국청소년연구
Studies on Korean Youth
2017. Vol. 28. No. 2 pp. 157~195.
ISSN 1225-6366(Print) / 2288-8799(Online)
http://dx.doi.org/10.14816/sky.2017.28.2.157

한국청소년패널조사에서의 종단적 최초표본이탈에 영향을 미치는 표본특성요인

이경상* · 이순래**

초 록

이 연구에서는 한국청소년패널조사 중2패널 1~6차년도, 초4패널 1~5차년도 데이터에서의 종단적 최초표본이탈에 영향을 미치는 표본특성요인에 대해서 경험적으로 분석해보고자 하였다. 이를 위해 선행연구분석을 통해 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모의 독립변수가 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈의 종속변수에 미치는 영향을 연구모형으로 설정한 후, 생존분석기법(survival analysis) 중 콕스의 비례위험 회귀분석 모형(cox proportional hazard model)을 적용해 경험적으로 분석하였다. 분석결과, 중2패널에서는 여성에 비해 남성일수록, 학업성적이 낮을수록, 모(여자보호자)학력이 높을수록, 월평균가구소득(log)이 높을수록 최초표본이탈위험이 높은 것으로 나타났고, 초4패널에서는 모(여자보호자)학력이 높을수록, 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록, 거주지가 중소도시에 비해 대도시일수록 최초표본이탈위험이 높은 것으로 나타났다. 중2패널과 초4패널의 최초표본이탈위험에 공통적으로 영향을 미치는 것으로 나타난 변수는 모(여자보호자)학력 변수뿐이었고, 중2패널에는 남성더미(여성기준), 학업성적, 월평균가구소득(log) 변수가, 초4패널에는 한부모가구더미(양부모가구기준), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준) 변수가 각각 따로 최초표본이탈위험에 영향을 미치는 것으로 나타났다. 이러한 분석결과는 향후 아동 및 청소년 대상의 패널데이터를 구축할 때 표본이탈을 최소화하기 위한 표본관리방법 강화에 도움을 줄 수 있을 것이다.

주제어: 한국청소년패널조사, 표본이탈, 표본특성

* 한국청소년정책연구원 선임연구위원, 제1저자
** 원광대학교 경찰행정학과 교수, 교신저자, soonlee@wonkwang.ac.kr

Ⅰ. 서 론

한국사회에서는 2000년대부터 동일표본 반복추적조사 방법에 의한 패널데이터 구축이 활발하게 진행되어오고 있다(김영석, 1999; 이경상, 2007, 2017; Babbie, 2001). 1975년 한국행동과학연구소에서 한국 아동의 종단적 연구를 최초로 실시한 이래 한동안 패널데이터 구축이 이뤄지지 않고 있다가, 1990년대 들어 1993년 대우경제연구소에서 한국가구패널조사를, 1998년 한국노동연구원에서 한국노동패널조사를 선구적으로 실시하였고, 이후 2000년대부터 여러 연구기관 및 대학 등에서 많은 패널데이터 구축을 진행해, 2017년 현재 40개가 넘는 패널데이터가 구축되었거나 구축 중에 있는 것으로 파악되고 있다(대우경제연구소, 1994; 이경상, 2007, 2017; 이규용, 이지은, 신선옥, 이혜정, 김기홍, 2015; 이성진, 2005).

현재까지 진행된 패널조사 중 아동 및 청소년을 주된 대상으로 한 패널조사를 조사시작년도 순으로 살펴보면, 2003년 한국청소년정책연구원의 한국청소년패널조사, 2004년 한국직업능력개발원의 한국교육고용패널조사, 2004년 서울대 서울아동패널연구단의 서울아동패널조사, 2005년 한국교육개발원의 한국교육종단연구, 2010년 서울특별시교육연구정보원의 서울교육종단연구, 2010년 한국청소년정책연구원의 한국아동청소년패널조사, 2011년 한국청소년정책연구원의 다문화청소년패널조사, 2011년 한국청소년정책연구원의 취약계층아동청소년종단조사, 2011년 한국교육개발원의 탈북청소년교육종단연구, 2011년 세종대 박현선 교수 연구진의 경기교육복지패널조사, 2011년 강원도교육연구원의 강원학생교육성취도종단연구, 2012년 경기도교육연구원의 경기교육종단연구, 2014년 가톨릭대 이상균 교수 연구진의 서울교육건강복지패널조사 등이 있다(강원도교육연구원, 2013; 경기도교육연구원, 2014; 김세원, 정익중, 2012; 서울특별시교육연구정보원, 2011; 신부용, 박현선, 2016; 이상균, 그레이스정, 유조안, 2015; 한국교육개발원, 2010, 2012; 한국직업능력개발원, 2015; 한국청소년정책연구원, 2009a, 2009b, 2010a, 2010b, 2011).

이러한 아동 및 청소년 대상의 패널데이터는, 횡단데이터에 비해, 부모의 이혼 등 인생사적 사건이 아동 및 청소년의 정신건강에 미치는 영향과 같이 변수간의 시간적 선후관계를 명확히 설정할 수 있고, 청소년 음주경험 횟수의 종단적 변화실태와 같이 변수들의 시간적 변화추세를 파악할 수 있는 등의 장점을 갖고 있다(민주홍, 2005;

이경상, 2007; 이지연, 김진, 2007; 한국청소년개발원, 2001; Diggle, Heagerty, Liang & Zeger, 2002). 이러한 장점때문에 많은 연구자들이 이들 패널데이터들을 활용하여 기존 횡단데이터에서는 가능하지 않거나 제한적으로만 가능했던 주제형태에 대한 경험적 분석을 실시하여 아동 및 청소년 분야의 학문적, 정책적 발전에 많은 기여를 해왔다.

반면 아동 및 청소년 대상의 패널데이터는, 횡단데이터에 비해, 상대적으로 연구설계가 까다롭고 많은 비용이 소요된다는 일반적인 단점 이외에, 패널조사가 진행됨에 따라 강력거절, 해외이주 등으로 인해 표본이 이탈한다는 단점도 아울러 갖고 있다(이경상, 2007; 이지연, 김진, 2007). 표본이탈은 패널데이터의 횡종단적 대표성에 문제를 야기하며, 특히 표본이탈이 무작위로 발생하지 않고 표본대상의 고유한 특성에 의해 발생하면 패널자료를 활용한 연구결과가 왜곡될 수도 있다(이상협, 박찬용, 정성석, 최혜미, 2011; 이화정, 강석복, 2014; Allison, 2005; Taris, 2000). 이러한 단점은 사후 가중치 부여를 통해 통계적으로 보완될 수 있지만, 패널데이터의 대표성 제고를 위해서는, 가급적 사전 조사과정에서의 표본이탈을 최소화한 후에 통계적 보완을 실시하는 것이 더 좋을 것이다(남기성, 이기성, 박근화, 홍기학, 손창균, 2014; Taris, 2000). 그래야만 더 나은 패널데이터를 통해 더 나은 논문들이 양산될 수 있을 것이다.

향후 아동 및 청소년 대상의 패널데이터의 표본이탈을 최소화하기 위해서는 기 진행된 아동 및 청소년 대상의 패널데이터들의 표본이탈에 영향을 미치는 표본특성요인들에 대한 분석들을 많이 축적하여 향후의 표본관리에 참고하는 것이 필요할 것이다. 그럼에도 불구하고 이들 패널데이터들의 표본이탈에 영향을 미치는 표본특성요인들에 대한 연구는 그리 많이 수행되지 않은 실정이다. 따라서 이 연구에서는 향후 아동 및 청소년 대상의 패널데이터의 표본이탈 최소화에 도움이 되고자, 한국청소년정책연구원의 한국청소년패널조사에서의 종단적 최초표본이탈에 영향을 미치는 표본특성요인에는 어떠한 것들이 있는지를 분석해보고자 한다.

Ⅱ. 선행연구의 검토

이 연구의 목적인 한국청소년패널조사에서의 종단적 최초표본이탈에 영향을 미치는 표본특성요인을 분석하기 위해서는 기존 관련 선행연구들에서 주로 어떤 독립변수들이 영향을 미치고 있는 변수들로 나타났는가를 살펴 참조하는 것이 필요하고, 아울러 한국청소년패널조사 데이터에서 실제로 독립변수로 변수화할 수 있는 것에는 어떠한 것들이 있는가를 고려하는 것이 필요하다.

패널조사의 표본이탈에 관한 대표적 선행연구로는 Fitzgerald, Gottschalk & Moffitt의 연구(1998)를 들 수 있다. Fitzgerald, Gottschalk & Moffitt(1998)는 1968년부터 1986년까지 19년간 PSID(Panel Study of Income Dynamics)의 표본이탈을 분석했다. 그 결과 기혼자보다는 미혼자가, 고령층일수록, 백인보다는 유색인종이, 고소득 가구일수록 표본이탈율이 높다는 것을 알 수 있었다. 또한 근로소득의 변동폭이 컸던 집단에서도 표본이탈률이 높게 나타났다(Fitzgerald, Gottschalk & Moffitt, 1998).

Lillard & Panis(1998)는 1968년부터 1988년까지 21년간의 PSID 자료를 분석했는데, 표본유지율이 높은 집단은 백인남성과 오랜 혼인기간의 기혼부부들이었다. 반면 표본이탈율이 높은 집단은 배우자와의 이혼이나 별거, 배우자 사망과 같은 부정적 가정변화를 겪은 사람들이었다(Lillard & Panis, 1998).

Zabel(1998)과 Hill & Willis(2001)는 인구통계학적 요인과 함께 조사기법 관련 요인에 주목했다. Zabell(1998)은 1988년까지의 PSID와 1984년과 1990년의 SIPP(Survey of Income and Program Participation)를 대상으로 표본이탈 모형을 검증했다. 그 결과 인구통계학적 요인의 영향은 그다지 크지 않았고, 조사기법 관련 요인들의 영향이 유효한 것으로 나타났다. 구체적으로 조사횟수가 거듭될수록 표본이탈이 증가했고, 면접시간이 짧을수록 그리고 동일 면접원이 반복해 조사할수록 표본이탈이 감소하는 경향이었다(Zabel, 1998). Hill & Willis(2001)는 HRS(Health and Retirement Study)를 분석했다. 그 결과는 Zabel의 연구와 유사했다. 동일한 면접원이 반복조사했을 때에 표본유지율이 증가하는 경향이었다(Hill & Willis, 2001).

MaCurdy, Mroz & Gritz(1998)은 1979년부터 1991년까지 13년간의 NLSY79 (National Longitudinal Survey of Youth 79)자료를 분석했다. NLSY79는 출생코호트 조사로 1979년부터 자료수집을 시작했다. 코호트는 1957-1964년 출생아들이며 1979년

당시에 표본대상은 14-22세이었다. 이 연구에서는 남성의 경우에 백인이며 대졸 학력의 미취업자 집단이 취업자들에 비해 표본이탈율이 높았다. 여성은 다소 다른 양상이었는데, 고졸과 대졸 집단에서 소득이 많을수록 표본이탈률이 증가하는 경향이었다(MaCurdy, Mroz & Gritz, 1998).

Watkins & Melde(2007)은 중학교 청소년들 대상의 갱단예방 교육훈련프로그램에 대한 종단적 평가자료인 GREAT(the longitudinal evaluation of the Gang Resistance Education and Training program)자료를 분석했다. 이 연구에서는 연령, 성, 인종, 미국출생여부, 양부모가족여부, 학교환경, 학교전념, 재산범죄, 폭력범죄, 피해, 자아통제를 표본이탈 관련요인으로 고려했는데, 분석결과 12세에 비해 13세이상이, 백인에 비해 다른 인종들이, 양부모가족에 비해 한부모가족이 표본이탈율이 높게 나타났고, 학교에 전념할수록 표본이탈율은 낮게 나타났다. 특히 이 연구에서는 Gottfredson & Hirschi(1990)의 자아통제론을 바탕으로 자아통제와 이와 연관된 비행(범죄)을 표본이탈의 중요한 관련요인으로 고려했는데, 자아통제와 비행(범죄)은 표본이탈에 유의한 영향을 미치지 않는 것으로 나타났다(Watkins & Melde, 2007).

김대일, 남재량과 류근관(2000)은 국내에서 처음으로 표본이탈을 체계적으로 연구했다. 이들은 대우경제연구소의 1993년부터 1998년까지 6년간의 한국가구패널(KHPS: Korea Household Panel Study) 자료와 통계청의 경제활동인구조사를 사용하여 구축한 1993년 2월부터 1997년 2월까지 5년간의 경제활동패널을 대상으로 표본이탈에 영향을 미치는 요인들을 살펴보았다. 분석결과 두 패널 모두에서 성, 연령, 학력, 혼인상태, 가구주와의 관계, 경제활동상태, 종사자의 지위에 따라 표본이탈율의 차이가 있었다. 여성에 비해 남성이, 고연령층에 비해 저연령층이, 미혼이거나 이혼한 집단, 전년도에 실업상태에 있던 집단, 그리고 취업상태일 때는 임금근로자인 집단에서 표본이탈율이 높은 경향이었다(김대일, 남재량, 류근관, 2000).

심영상과 허명회(2002)는 한국노동연구원의 1998년부터 1999년까지 2년간의 한국노동패널조사(KLIPS: Korean Labor & Income Panel Study) 자료를 사용해 가구 표본이탈의 영향요인을 로지스틱 회귀모형으로 분석했다. 분석결과, 서울에 비해 대구, 광주, 충북에서, 전월세가구에서, 고소득가구일수록, 가구주가 미혼인 경우, 가구주가 고학력일수록, 면접이 어려워 방문횟수가 많을수록 표본이탈이 높은 경향이었다(심영상, 허명회, 2002).

이상호(2005)는 한국노동연구원의 1998년부터 2003년까지 6년간의 한국노동패널조사(KLIPS: Korean Labor & Income Panel Study) 자료를 사용해 가구 표본이탈의 영향요인과 표본이탈이 가구소득에 미치는 영향의 두 단계로 나누어 분석했다. 첫 번째 가구 표본이탈의 영향요인에 대해 프로빗 기법(probit analysis)을 사용해 분석한 결과, 가구주의 연령이 높을수록, 가구의 거주형태가 자가인 경우 표본이탈이 낮은 경향이었고, 가구주의 혼인상태가 미혼인 경우, 가구주가 실업자인 경우, 가구의 소비수준이 높을수록 표본이탈이 높은 경향이었다. 두 번째, 첫 번째 단계에서 산출된 Inverse Mills Ratio을 포함한 고정효과 모형(fixed effect model)과 포함하지 않은 고정효과 모형을 비교분석한 결과, 시간가변(time variant) 요인들의 효과는 크게 변하지 않았다. 따라서 표본이탈로 인해 외형적으로 대표성에 의문이 제기될 수 있지만, 표본이탈이 연구결과를 크게 왜곡하지는 않는 것으로 나타났다. 표본이탈이 발생해도 관련 요인들에 가중치를 부여하거나 기존 가중치를 보정하면 충분히 신뢰할 수 있는 연구결과를 얻을 수 있다는 것이다(이상호, 2005).

천영민, 윤정혜와 오민홍(2009)은 한국고용정보원의 2006년부터 2007년까지 2년간의 대졸자직업이동경로조사(GOMS: Graduate Occupational Mobility Survey) 자료를 활용해 표본이탈의 영향요인을 로짓 기법으로 분석했다. 분석결과, 기혼남성의 표본유지율이 가장 높았고, 미혼자의 표본이탈율이 높게 나타났다. 연령대에서는 20대 후반이 20대 초반보다 표본이탈율이 높았고, 전문대 졸업자가 4년제 대학이나 교육대학 졸업자보다 표본이탈율이 높았다. 경제활동상태에서는 미취업자가 취업자나 학생보다 이탈율이 높았고, 부모와 동거하지 않거나 가구 월평균 총소득이 낮은 집단에서 표본이탈율이 높게 나타났다. 동일면접원 여부에서는 같은 면접원이 반복조사한 경우에 그렇지 않을 때에 비해 표본유지율이 1.5배 정도 높은 수준이었다(천영민, 윤정혜, 오민홍, 2009).

이상협, 박찬용, 정성석과 최혜미(2011)는 한국노동연구원의 1998년부터 2008년까지 11년간의 한국노동패널조사(KLIPS: Korean Labor & Income Panel Study) 자료 중 가구주 자료를 대상으로 가구주의 표본유지기간에 따른 이탈여부를 콕스 비례위험모형을 이용해 분석했다. 분석결과, 취업여부별로는 미취업자의 이탈위험이 취업자에 비해 1.28배 높은 것으로 나타났다. 연령별로는 60세 이상인 집단이 30세 미만에 비해 이탈위험이 낮아지는 경향이었고, 지역별로는 서울에 비해 경기도를 제외한 다른 지역에서 이탈위험이 감소하는 경향이었다. 학력별로는 초등학교 이하에 비해 전문대

이상인 경우 이탈위험이 1.13배 높았고, 결혼상태별로는 미혼인 집단이 기혼이나 이혼·별거 집단에 비해 이탈위험이 높았다. 조사방법별로는 방문면접에 비해 유치조사한 경우에 이탈위험이 1.13배 높게 나타났다. 입주형태별로는 자가에 비해 전세인 집단에서 이탈위험이 1.1배 높았고, 자가에 비해 월세인 집단에서 이탈위험이 1.237배 높았으며, 주거형태별로는 단독주택에 비해 상가주택인 경우에 0.85배로 이탈위험이 낮은 경향이었다(이상협, 박찬용, 정성석, 최혜미, 2011).

김세원과 정익중(2012)은 서울아동패널연구단의 2004년부터 2010년까지 7년간의 서울아동패널조사(SPSC: Seoul Panel Study of Children) 자료의 표본이탈을 음이항 회귀분석을 이용해 분석하였다. 김세원과 정익중(2012)은 7차년도까지의 표본이탈 횟수를 종속변수로 설정하고, 성별, 한부모가구여부, 학교폭력피해경험정도, 중비행정도, 유능감, 학업성취수준, 자기통제를 독립변수로 설정하였다. 분석결과 다른 변수들은 통계적으로 유의한 영향을 미치지 않는 것으로 나타났고, 학업성취수준만이 통계적으로 유의한 영향을 미치고 있는 것으로 나타났는데, 학업성취수준이 낮을수록 표본이탈 횟수가 높게 나타났다. 이 연구에서도 Watkins & Melde(2007)의 연구에서처럼 자아통제와 중비행 변수를 독립변수로 투입하였는데 표본이탈 횟수에 유의한 영향을 미치지 않는 것으로 나타났다(김세원, 정익중, 2012).

송헌재(2012)는 한국조세재정연구원의 2008년부터 2009년까지 2년간의 재정패널조사(NSTB: National Survey of Tax and Benefit) 자료의 가구 이탈여부를 강력거절이탈, 추적실패이탈, 기타이탈로 나누어 다항로지스틱 회귀분석 모형을 이용해 분석했다. 강력거절이탈의 경우에는, 가구주가 여성인 경우, 가구소득이 높을수록, 가구순자산이 많을수록 강력거절로 표본이탈할 가능성이 높고, 1차년도 설문에 가구주가 응답한 경우, 동일한 면접원을 투입할 경우 강력거절로 표본이탈할 가능성이 낮은 것으로 나타났다. 추적실패이탈의 경우에는, 가구원수가 적을수록, 가구소득이 높을수록, 응답자가 여성인 경우, 응답자 연령이 어릴수록 추적실패로 표본이탈할 가능성이 높고, 동일면접원을 투입할 경우 추적실패로 표본이탈할 가능성이 낮은 것으로 나타났다(송헌재, 2012).

이경희와 민인식(2016)은 한국노동연구원의 1998년부터 2016년까지 17년간의 한국노동패널조사(KLIPS: Korean Labor & Income Panel Study) 자료의 표본가구 이탈여부를 로그로지스틱 생존분석 모형을 이용해 분석했다. 분석결과 고졸가구주인 경우,

가구주가 임금근로자인 경우, 가구주 연령이 높을수록, 가구소득이 높을수록, 가구원수가 적을수록, 비광역시에 거주하는 경우 표본이탈 위험이 낮은 경향이었다(이경희, 민인식, 2016).

이상의 국내외 관련 선행연구의 분석결과를 요약해 살펴보면, 표본(개인 혹은 가구주)의 특성과 관련해서는 대체적으로 백인에 비해 유색인이, 기혼자에 비해 미혼자가, 배우자와의 이혼, 별거, 사망을 겪은 사람이, 학교전념이 낮을수록, 학업성취수준이 낮을수록, 취업자에 비해 미취업자들이, 가구소비수준이 높을수록, 양부모가구에 비해 한부모가구인 경우, 부모와 동거하지 않는 경우, 자가에 비해 전월세가구인 경우 표본이탈이 높은 것으로 나타났다. 성별, 연령, 학력, 가구원수는 연구결과가 다소 상반되게 나타났고, 가구소득의 경우에는 전반적으로 가구소득이 높을수록 표본이탈이 높게 나타난 가운데 일부 상반된 연구결과도 있는 것으로 나타났다. 지역도 서울 등 대도시가 표본이탈이 높게 나타난 가운데 상반된 연구결과도 있는 것으로 나타났다.

일부 연구에서 표본특성과 함께 분석된 조사과정의 특성과 관련해서는 조사횟수가 많아질수록, 면접시간이 길수록, 면접원이 변경될수록, 방문면접조사에 비해 유치조사를 한 경우 일관되게 표본이탈이 높은 것으로 나타났다.

이러한 분석결과를 바탕으로 이 연구에서 투입할 독립변수를 다음과 같이 선정해보았다.

우선 조사과정의 특성과 관련된 변수들인 조사횟수, 면접시간, 면접원 변경여부, 유치조사여부 변수들은 한국청소년패널조사에서 제공하고 있지 않으므로 독립변수로 투입할 수가 없다. 다음으로 표본의 특성과 관련된 변수들 가운데에서 청소년들의 개인변수로 투입가능한 변수들 중, 인종, 학교전념 변수는 한국청소년패널조사에서 제공하고 있지 않고, 연령, 학력, 혼인상태, 취업여부 변수는 초4, 중2 패널인 한국청소년패널조사의 청소년에게는 해당되지 않으므로 독립변수로 투입할 수가 없다. 또한 표본의 특성과 관련된 변수들 가운데에서 청소년들의 사회경제적 배경변수로 투입가능한 변수들 중, 부모의 성, 부모의 연령, 부모의 인종, 가구소비수준 변수도 한국청소년패널조사에서 제공하고 있지 않으므로 독립변수로 투입할 수가 없다.

따라서 이 연구에서는, 이들 변수들을 제외하고, 청소년들의 개인변수로 성별, 학업성적 변수를, 청소년들의 사회경제적 배경변수로 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 부모동거여부, 가구원수, 월평균가구소

득, 자가거주여부, 거주지규모 변수를 독립변수로 설정하였다. 그런데 설정된 독립변수들 중, 부모동거여부 변수는, 중2패널(1차년도)의 경우 부모동거=99.0%(3,385명), 부모동거안함=1.0%(35명)이고 초4패널(1차년도)의 경우 부모동거=99.2%,(2,815명) 부모동거안함=0.8%(24명)로, 거의 대부분의 청소년들이 부모와 동거하고 있고 동거하고 있지 않는 경우가 너무 적은 것으로 나타나 통계적인 측면에서 변수로서의 가치가 적다고 판단되어 독립변수에서 제외하고, 나머지 변수들을 최종적으로 독립변수로 설정하였다.

한편 이 연구에서는 패널데이터의 매 차수마다의 정태적(static) 측면에서의 표본이탈이 아니라 패널데이터의 전 차수에 걸친 동태적(dynamic) 측면에서의 표본이탈을 종속변수로 설정하였다.

패널조사에서 표본이탈은 시간의 흐름을 바탕으로 한 일종의 사건(event)이다(박재빈, 2007; Vermunt, 1997). 이혼(divorce)은 결혼을 한 부부가 일정 기간이 지난 후에 결별하는 사건이며, 전직(job transition)은 어떤 직장에 다니다가 시간이 흐른 뒤에 다른 직장을 찾는 사건이다.

이러한 사건(event)은 시간의 흐름에 따라 그 실질적 의미가 다르다. 결혼 후 1년 뒤의 이혼과 30년 뒤의 이혼을 같다고 할 사람은 없을 것이다. 전직의 경우도 취업하자마자 직장을 바꾸는 전직과 15년 후의 전직은 다른 의미의 사건이다. 마찬가지로 패널조사의 2차년도에 발생한 표본이탈과 3차년도, 4차년도 그 후에 발생한 표본이탈의 경우 시간의 흐름을 고려한 동태적 측면에서는 서로 다른 표본이탈이다.

따라서 이 연구에서는, 패널조사에서의 이러한 측면을 고려하여, 한국청소년패널조사에서의 시간의 흐름을 고려한 종단적 최초표본이탈(event)을 종속변수로 설정하였다. 그런데 종속변수인 종단적 최초표본이탈에는 중2패널, 초4패널 각각 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈의 3가지 하위유형이 있어, 이 6가지 표본이탈유형 모두를 최종적으로 종속변수로 설정하였다. 청소년최초표본이탈, 부모(보호자)최초표본이탈 외에 청소년부모(보호자)동시최초표본이탈을 종속변수의 하위유형으로 포함한 이유는 청소년부모(보호자)동시최초표본이탈이 어느 한쪽의 강제, 설득 혹은 상호간의 협의를 통해 일어나는 사건으로 표본이탈에 있어 청소년최초표본이탈, 부모(보호자)최초표본이탈과는 의미가 다르다고 판단했기 때문이다.

그러므로 이 연구의 최종 연구모형은 한국청소년패널조사에서 독립변수인 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모가 종속변수인 종단적 측면에서의 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈에 미치는 영향을 경험적으로 분석해보는 것으로 설정되었다.

Ⅲ. 연구방법

1. 분석모형

앞에서 살펴본 것처럼 이 연구의 연구모형은 한국청소년패널조사에서의 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모가 종단적 측면에서의 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈에 미치는 영향을 살펴보는 것이었다.

이러한 연구모형은 생존분석(survival analysis) 중 콕스의 비례위험 회귀분석 모형(cox proportional hazard model)을 적용해 분석해 볼 수 있다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997). 생존분석(survival analysis) 혹은 사건사분석(event history analysis)은 시간의 흐름을 반영한 종속변수의 변화를 분석하는 방법이다. 콕스의 비례위험 회귀분석 모형은, 준모수적 추정방법(semi-parametric estimation method)으로 독립변수의 효과를 판별하는 방법으로, 종속변수인 비례위험도(위험요인이 없는 경우의 위험도와 위험요인이 있는 경우의 위험도의 비례)에 여러 독립변수들이 미치는 영향을 검증하는 분석모형을 말한다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997).

콕스의 비례위험 회귀분석 모형은 다음과 같은 지수함수의 변형들에 자연로그를 취해 일반화선형모형의 한 형태로 나타내볼 수 있다. 여기에서 $h_0(t)$는 위험요인이 전혀 없을 때의 t시점에서의 사건(event)발생의 위험도를 의미하고 $h_i(t)$는 i개체의 t시점에서의 사건(event)발생의 위험도를 위미하며, $\exp(\beta_k)$는 위험요인이 있을 때

($x_k = 1$)의 비례위험도를 의미한다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997).

$$h_i(t) = h_0(t) \times \exp(\beta_1 x_1) \times \exp(\beta_2 x_2) \times \cdots \times \exp(\beta_k x_k)$$

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta_1 x_1) \times \exp(\beta_2 x_2) \times \cdots \times \exp(\beta_k x_k)$$

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

$$\ln \frac{h_i(t)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

이 연구의 연구모형을 콕스의 비례위험 회귀분석 모형에 적용해보면, 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모와 같은 표본특성들이 독립변수인 위험요인들이며, 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈과 같은 종단적 표본이탈의 가능성이 종속변수인 비례위험도에 해당된다.

콕스의 비례위험 회귀분석 모형은 생존시간의 분포에 대해서는 아무런 가정을 하지 않지만 비례위험도에 대해서는 시간에 대해 일정하다고 가정(비례위험의 가정)하는 준모수적 모형이다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997). 따라서 모형의 타당성을 검증하기 위해서는 비례위험의 가정에 대해 검증하는 절차가 필요한데, 일반적으로 이 연구에서 활용한 SPSS에서는 LML(log minus log survial) plot을 통해 독립변수의 각 하위범주에 해당되는 곡선들이 수직으로 어느 정도 일정한 거리를 유지하고 있다면 비례위험의 가정을 만족하고 있는 것으로 검증하고 있다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997).

2. 분석자료

이 연구에서는 한국청소년패널조사에서의 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모와 같은 표본특성들이 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈과 같은 종단적 표본이탈에 미치는 영향을 콕스의 비례위험 회귀분석 모형을 통해 경험적으로 검증해보려고 하고 있으므로, 당연히 한국청소년패널조사 중2패널 1~6차년도, 초4패널 1~5차년도 데이터를 분석자료로 활용하였다(한국청소년정책연구원, 2009a, 2009b).

한국청소년패널조사(KYPS: Korea Youth Panel Survey)의 중2패널 데이터는 2003년 기준 전국 중학교 2학년 청소년들을 모집단으로 층화다단계 집락표집방법에 의해 표본추출한 3,449명의 중학교 2학년 청소년들과 보호자들을 대상으로 6년간 반복추적조사하여 구축한 자료이고, 초4패널 데이터는 2004년 기준 전국 초등학교 4학년 청소년들을 모집단으로 층화다단계 집락표집방법에 의해 표본추출한 2,844명의 초등학교 4학년 청소년들과 보호자들을 대상으로 5년간 반복추적조사하여 구축한 자료이다(한국청소년정책연구원, 2009a, 2009b).

1차년도 조사대상 표본수 대비 최종년도 청소년 조사성공 사례수의 비율은 중2패널 82.1%, 초4패널 86.1%로 나타났고, 최종년도 청소년·부모 조사성공 사례수의 비율은 중2패널 81.8%, 초4패널 85.9%로 나타났다. 그리고 1차년도 조사대상 표본수 대비 중2패널 청소년 1~6차년도 계속응답 사례수의 비율은 71.3%, 초4패널 청소년 1~5차년도 계속응답 사례수의 비율은 79.7%인 것으로 나타났다(한국청소년정책연구원, 2009a, 2009b).

3. 분석변수

이 연구의 종속변수는 한국청소년패널조사에서의 종단적인 측면에서의 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈이다.

실제 분석시에는, 종속변수 모두, 중2패널 1~6차년도 데이터와 초4패널 1~5차년도 데이터에서 각각 종단적 최초 표본이탈 발생상황을, 콕스의 비례위험 회귀분석의

통계적 논리에 맞게, 최초 표본이탈(event)의 발생무유(0=censored, 1=event) 변수와 최초 표본이탈(event)이 발생하기까지 걸린 년수(time to event) 변수의 두 가지 분석변수로 만든 후 분석에 투입하여 활용하였다. 최초 표본이탈을 사건(event)으로 설정하였으므로, 최초 표본이탈후 복귀하거나 복귀후 다시 이탈하는 경우에도 최초 표본이탈 시점을 기준으로 발생무유와 년수를 산정하였다.

이 연구의 독립변수는 한국청소년패널조사에서의 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모이다.

실제 분석시에는, 독립변수 모두, 중2패널 1차년도 데이터와 초4패널 1차년도 데이터에서 각각, 남성더미(여성기준), 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)미취업더미(부모(보호자)모두취업기준), 한부모가구더미(양부모가구기준), 가구원수, 월평균가구소득(log), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준), 거주지농촌더미(중소도시기준)의 11개 분석변수로 만든 후 분석에 투입하여 활용하였다.

남성더미(여성기준) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 성별 변수인 '(1) 남자', '(2) 여자'로 측정된 것을, 여성=0, 남성=1로 하는 남성더미 변수로 만들어 분석에 활용하였다.

학업성적 변수는 중2패널의 경우에는 1차년도 자료에서 1학기 반성적을 %로 측정한 것을 그대로 분석에 활용하였다. 따라서 점수가 높을수록 학업성적이 낮음을 의미한다. 초4패널의 경우에는 1차년도 자료에서 1학기 반성적을 국어, 영어, 수학, 사회, 과학 5개 과목별로 각각 '(1) 매우 못하는 수준'에서 '(5) 매우 잘하는 수준'의 5점 척도로 측정된 것을, 모두 합산한 후 과목수인 5로 나눈 다음, 다시 다른 변수들과의 설명의 방향성을 맞추기 위해 6에서 빼는 방식으로 역코딩한 것을 분석에 활용하였다. 따라서 점수가 높을수록 학업성적이 낮음을 나타낸다.

부(남자보호자)학력 변수와 모(여자보호자)학력 변수는 중2패널, 초4패널 모두 1차년도 자료에서 '(1) 무학', '(2) 초등학교', '(3) 중학교', '(4) 고등학교', '(5) 전문대', '(6) 대학교', '(7) 대학원 석사', '(8) 대학원박사'로 측정된 것을 실제 분석에서는 각각 무학=0, 초등학교=6, 중학교=9, 고등학교=12, 전문대=14, 대학교=16, 대학원 석사=18, 대학원박사=21의 교육연수로 환산하여 활용하였다. 따라서 모두 점수가 높을수록 학

력이 높음을 의미한다.

부모(보호자)미취업더미(부모(보호자)모두취업기준) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 현재 직업을 측정한 것을, 부모(보호자) 모두가 취업한 경우=0, 부(남자보호자) 혹은 모(여자보호자) 중 1명이상 미취업한 경우=1로 하는 부모(보호자) 미취업더미 변수로 만들어 분석에 활용하였다.

한부모가구더미(양부모가구기준) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 현재 부모와의 가족구성형태를 측정한 것을, 부모가 모두 있는 양부모가구=0, 부 또는 모만 있는 한부모가구=1로 하는 한부모가구더미 변수로 만들어 분석에 활용하였다.

가구원수 변수는 중2패널, 초4패널 모두 1차년도 자료에서 측정한 부모수(동거여부 관계없음), 언니/누나/오빠/형수(동거여부 관계없음. 단, 결혼해서 분가한 형제자매는 제외), 동생수(동거여부 관계없음. 단, 결혼해서 분가한 형제자매는 제외), 할아버지수(동거인 경우만 해당), 할머니수(동거인 경우만 해당), 친척수(동거인 경우만 해당)에다가 응답자인 청소년 본인을 더해서 가구원수 변수로 만들어 분석에 활용하였다.

월평균가구소득(log) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 가구 월평균 소득으로 측정된 것을 log값으로 변환하여 분석에 활용하였다. 따라서 모두 점수가 높을수록 월평균가구소득(log)이 높음을 의미한다.

자가거주아님더미(자가거주기준) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 '(1) 자기집', '(2) 전세', (3) '월세', (4) '무상', (5) '기타'로 측정된 것을, 자가거주=0, 자가거주아님=1로 하는 자가거주아님더미 변수로 만들어 분석에 활용하였다.

거주지대도시더미(중소도시기준) 변수와 거주지농촌더미(중소도시기준) 변수는 중2패널, 초4패널 모두 1차년도 자료에서 거주지를 직접 측정한 후 대도시(특별시 및 광역시), 중소도시, 농촌(군)의 세 가지 하위변수로 리코딩한 후, 중소도시를 기준변수로 하는 거주지대도시더미(중소도시기준) 변수와 거주지농촌더미(중소도시기준) 변수를 만들어 분석에 활용하였다.

4. 분석방법

이 연구에서 분석을 위해 투입된 독립변수는 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소

득, 자가거주여부, 거주지규모이고 종속변수는 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈이었다.

먼저 이러한 변수들의 실태를 살펴보기 위해서 빈도분석, 평균분석, 교차분석을 실시하였다. 그리고 독립변수들이 종속변수에 미치는 영향을 경험적으로 살펴보기 위해서 SPSS 20.0을 사용하여 콕스의 비례위험 회귀분석을 실시하였다(박재빈, 2007; 송경일, 최종수, 2008; Vermunt, 1997).

콕스의 비례위험 회귀분석시 투입된 독립변수들간의 상관관계 분석을 실시한 결과 다중공선성의 우려는 없는 것으로 나타났다(표 1, 2 참조). 그리고 콕스 비례위험 회귀모형의 비례위험의 가정 검증을 위해 LML(log minus log survial) plot분석을 실시한 결과 비례위험의 가정도 충족되는 것으로 나타났다(부록 참조).

표 1

중2패널 투입 독립변수들간의 상관관계 분석결과

	A	B	C	D	E	F	G	H	I	J	K
A	1	.052**	.014	.009	.005	-.010	-.083**	.015	-.023	-.014	-.055**
B	.052**	1	-.303**	-.260**	-.052**	.152**	-.036*	-.256**	.132**	-.009	.011
C	.014	-.303**	1	.696**	.047**	-.131**	-.063**	.464**	-.098**	.119**	-.175**
D	.009	-.260**	.696**	1	-.001	-.137**	-.073**	.454**	-.099**	.124**	-.195**
E	.005	-.052**	.047**	-.001	1	-.018	-.010	-.052**	-.036*	.054**	-.025
F	-.010	.152**	-.131**	-.137**	-.018	1	-.250**	-.337**	.146**	-.013	.017
G	-.083**	-.036*	-.063**	-.073**	-.010	-.250**	1	.042*	-.131**	-.066**	.165**
H	.015	-.256**	.464**	.454**	-.052**	-.337**	.042*	1	-.243**	.047**	-.148**
I	-.023	.132**	-.098**	-.099**	-.036*	.146**	-.131**	-.243**	1	.048**	-.036*
J	-.014	-.009	.119**	.124**	.054**	-.013	-.066**	.047**	.048**	1	-.254**
K	-.055**	.011	-.175**	-.195**	-.025	.017	.165**	-.148**	-.036*	-.254**	1

* $p<.05$, ** $p<.01$, *** $p<.001$

A 남성더미, B 학업성적, C 부학력, D 모학력, E 부모미취업더미, F 한부모가구더미, G 가구원수, H 월평균가구소득log, I 자가거주아님더미, J 거주지대도시더미, K 거주지농촌더미

표 2

초4패널 투입 독립변수들간의 상관관계 분석결과

	A	B	C	D	E	F	G	H	I	J	K
A	1	-.016	.004	-.005	-.021	.014	-.040*	.024	.002	.011	.019
B	-.016	1	-.274**	-.259**	-.061**	.105**	.030	-.260**	.151**	-.001	.058**
C	.004	-.274**	1	.677**	.051**	-.121**	-.058**	.438**	-.149**	.038*	-.125**
D	-.005	-.259**	.677**	1	.001	-.108**	-.062**	.450**	-.134**	.064**	-.093**
E	-.021	-.061**	.051**	.001	1	.007	-.042*	-.174**	-.041*	.052**	-.049**
F	.014	.105**	-.121**	-.108**	.007	1	-.145**	-.300**	.099**	.020	.024
G	-.040*	.030	-.058**	-.062**	-.042*	-.145**	1	.067**	-.070**	-.059**	.089**
H	.024	-.260**	.438**	.450**	-.174**	-.300**	.067**	1	-.245**	.059**	-.064**
I	.002	.151**	-.149**	-.134**	-.041*	.099**	-.070**	-.245**	1	.024	.011
J	.011	-.001	.038*	.064**	.052**	.020	-.059**	.059**	.024	1	-.286**
K	.019	.058**	-.125**	-.093**	-.049**	.024	.089**	-.064**	.011	-.286**	1

* p<.05, ** p<.01, *** p<.001

A 남성더미, B 학업성적, C 부학력, D 모학력, E 부모미취업더미, F 한부모가구더미, G 가구원수, H 월평균가구소득log, I 자가거주아님더미, J 거주지대도시더미, K 거주지농촌더미

Ⅳ. 분석결과

1. 투입변수들의 기본적 특성

여기에서는 연구모형에 투입된 변수들의 기본적인 특성에 대해 살펴보았다. 표3은 중2패널, 초4패널의 투입변수들에 대해 빈도분석, 평균분석을 실시한 결과를 표로 나타낸 것이고, 표4, 표5는 중2패널, 초4패널의 최초표본이탈여부별 표본이탈시점의 분포에 대해 교차분석을 실시한 결과를 표로 나타낸 것이다.

중2패널의 경우부터 살펴보면, 독립변수로 투입된 변수 중, 성별은 여성이 50.0%, 남성이 50.0%의 분포를 보이고 있고, 학업성적은 반성적을 %로 측정한 것인데 평균 44.99%,인 것으로 나타났다. 부(남자보호자)학력은 평균 13.32년, 모(여자보호자)학력은 평균 12.24년이었다. 부모(보호자)취업여부는 부모(보호자)모두취업인 경우가 48.0%, 부모(보호자)중1명이상미취업인 경우가 52.0%인 것으로 나타났고, 한부모가구여부는 양부모가구가 94.5%, 한부모가구가 5.5%인 것으로 나타났다. 가구원수는 평균 4.30명인 것으로 나타났고, 월평균가구소득은 299.73만원인 것으로 나타났다. 자가거주여부는 자가거주가 69.4%, 자가거주아님이 30.6%로 나타났고, 거주지규모는 대도시 46.1%, 중소도시 46.9%, 농촌지역 7.0%인 것으로 나타났다.

종속변수로 투입된 변수 중, 청소년최초표본이탈여부는 패널 종료시점까지 표본이탈이 없어 중도절단(censored)으로 분류된 경우가 71.3%, 표본이탈(event)로 분류된 경우가 28.7%인 것으로 나타났고, 부모(보호자)최초표본이탈여부는 중도절단(censored)으로 분류된 경우가 68.6%, 표본이탈(event)로 분류된 경우가 31.4%인 것으로 나타났으며, 청소년부모(보호자)최초동시표본이탈은 중도절단(censored)으로 분류된 경우가 71.7%, 표본이탈(event)로 분류된 경우가 28.3%인 것으로 나타났다.

표본이탈여부별 표본이탈시점의 분포를 살펴보면, 청소년최초표본이탈의 경우에는 표본이탈 전체 100% 중에서 2차년도 26.4%, 3차년도 17.3%, 4차년도 10.8%, 5차년도 19.1%, 6차년도 26.5%의 분포로 표본이탈이 발생했고, 부모(보호자)최초표본이탈의 경우에는 2차년도 31.6%, 3차년도 17.2%, 4차년도 10.8%, 5차년도 17.7%, 6차년도 22.7%의 분포로 표본이탈이 발생했으며, 청소년부모(보호자)최초동시표본이탈의 경우에는 2차년도 26.7%, 3차년도 17.5%, 4차년도 10.2%, 5차년도 19.1%, 6차년도 26.4%의 분포로 표본이탈이 발생했다.

표 3

투입변수들에 대한 기술통계 분석결과(빈도(%), 평균(표준편차))

		중2패널(n=3,449)	초4패널(n=2,844)
성별	여성	1724(50.0%)	1320(46.4%)
	남성	1725(50.0%)	1524(53.6%)
학업성적		44.99(27.72)	2.62(0.64)
부학력		13.32(2.98)	13.80(2.59)
모학력		12.24(2.69)	12.91(2.24)
부모취업여부	부모모두취업	1591(48.0%)	1353(48.9%)
	부모1명이상미취업	1725(52.0%)	1413(51.1%)
한부모가구여부	양부모가구	3229(94.5%)	2725(96.2%)
	한부모가구	188(5.5%)	107(3.8%)
가구원수		4.30(0.88)	4.32(0.84)
월평균가구소득		299.73(216.89)	302.14(176.51)
자가거주여부	자가거주	2356(69.4%)	1768(62.4%)
	자가거주아님	1039(30.6%)	1064(37.6%)
거주지 규모	대도시	1591(46.1%)	1316(46.3%)
	중소도시	1616(46.9%)	1281(45.0%)
	농촌지역	242(7.0%)	247(8.7%)
청소년최초 표본이탈여부	중도절단(censored)	2459(71.3%)	2267(79.7%)
	표본이탈(event)	990(28.7%)	577(20.3%)
부모최초 표본이탈여부	중도절단(censored)	2365(68.6%)	2213(77.8%)
	표본이탈(event)	1084(31.4%)	631(22.2%)
청소년부모최초동시 표본이탈여부	중도절단(censored)	2473(71.7%)	2280(80.2%)
	표본이탈(event)	976(28.3%)	564(19.8%)

표 4

중2패널-최초표본이탈여부별 표본이탈시점의 분포(빈도, %)

		표본이탈시점(year)					전체
		2	3	4	5	6	
청소년 최초 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2459 (100.0%)	2459 (100.0%)
	표본이탈 (event)	261 (26.4%)	171 (17.3%)	107 (10.8%)	189 (19.1%)	262 (26.5%)	990 (100.0%)
부모 최초 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2365 (100.0%)	2365 (100.0%)
	표본이탈 (event)	343 (31.6%)	186 (17.2%)	117 (10.8%)	192 (17.7%)	246 (22.7%)	1084 (100.0%)
청소년 부모 최초동시 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2473 (100.0%)	2473 (100.0%)
	표본이탈 (event)	261 (26.7%)	171 (17.5%)	100 (10.2%)	186 (19.1%)	258 (26.4%)	976 (100.0%)

표 5

초4패널-최초표본이탈여부별 표본이탈시점의 분포(빈도, %)

		표본이탈시점(year)				전체
		2	3	4	5	
청소년 최초 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2267 (100.0%)	2267 (100.0%)
	표본이탈 (event)	137 (23.7%)	100 (17.3%)	194 (33.6%)	146 (25.3%)	577 (100.0%)
부모 최초 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2213 (100.0%)	2213 (100.0%)
	표본이탈 (event)	174 (27.6%)	125 (19.8%)	193 (30.6%)	139 (22.0%)	631 (100.0%)
청소년 부모 최초동시 표본이탈 여부	중도절단 (censored)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2280 (100.0%)	2280 (100.0%)
	표본이탈 (event)	137 (24.3%)	96 (17.0%)	183 (32.4%)	148 (26.2%)	564 (100.0%)

초4패널의 경우를 살펴보면, 독립변수로 투입된 변수 중, 성별은 여성이 46.4%, 남성이 53.6%의 분포를 보이고 있고, 학업성적은 반성적을 1~5점으로 측정한 것인데 평균 2.62점으로 나타났다. 부(남자보호자)학력은 평균 13.80년, 모(여자보호자)학력은 평균 12.91년이었다. 부모(보호자)취업여부는 부모(보호자)모두취업인 경우가 48.9%, 부모(보호자)중1명이상미취업인 경우가 51.1%인 것으로 나타났고, 한부모가구여부는 양부모가구가 96.2%, 한부모가구가 3.8%인 것으로 나타났다. 가구원수는 평균 4.32명인 것으로 나타났고, 월평균가구소득은 302.14만원인 것으로 나타났다. 자가거주여부는 자가거주가 62.4%, 자가거주아님이 37.6%로 나타났고, 거주지규모는 대도시 46.3%, 중소도시 45.0%, 농촌지역 8.7%인 것으로 나타났다.

종속변수로 투입된 변수 중, 청소년최초표본이탈여부는 패널 종료시점까지 표본이탈이 없어 중도절단(censored)으로 분류된 경우가 79.7%, 표본이탈(event)로 분류된 경우가 20.3%인 것으로 나타났고, 부모(보호자)최초표본이탈여부는 중도절단(censored)으로 분류된 경우가 77.8%, 표본이탈(event)로 분류된 경우가 22.2%인 것으로 나타났으며, 청소년부모(보호자)최초동시표본이탈은 중도절단(censored)으로 분류된 경우가 80.2%, 표본이탈(event)로 분류된 경우가 19.8%인 것으로 나타났다.

표본이탈여부별 표본이탈시점의 분포를 살펴보면, 청소년최초표본이탈의 경우에는 표본이탈 전체 100% 중에서 2차년도 23.7%, 3차년도 17.3%, 4차년도 33.6%, 5차년도 25.3%의 분포로 표본이탈이 발생했고, 부모(보호자)최초표본이탈의 경우에는 2차년도 27.6%, 3차년도 19.8%, 4차년도 30.6%, 5차년도 22.0%의 분포로 표본이탈이 발생했으며, 청소년부모(보호자)최초동시표본이탈의 경우에는 2차년도 24.3%, 3차년도 17.0%, 4차년도 32.4%, 5차년도 26.2%의 분포로 표본이탈이 발생했다.

2. 종단적 최초표본이탈의 영향요인

이어서 이 연구의 연구모형인 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모의 독립변수가 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈의 종속변수에 미치는 영향을 살펴보았다. 실제 분석을

위해, 앞의 연구방법 부분에서 언급했던대로, 독립변수들을 남성더미(여성기준), 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)미취업더미(부모(보호자)모두취업기준), 한부모가구더미(양부모가구기준), 가구원수, 월평균가구소득(log), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준), 거주지농촌더미(중소도시기준)으로 전환한 후 투입하였다. 표 6, 표 7은 중2패널, 초4패널의 최초표본이탈에 대한 콕스비례위험회귀분석 결과를 표로 나타낸 것이다.

표 6

중2패널-최초표본이탈에 대한 콕스비례위험회귀분석 결과

	청소년 최초표본이탈(event)		부모 최초표본이탈(event)		청소년부모동시 최초표본이탈(event)	
	B	Exp(B)	B	Exp(B)	B	Exp(B)
남성더미(여성기준)	.196**	1.217	.188**	1.206	.198**	1.219
학업성적	.008***	1.008	.008***	1.008	.008***	1.008
부학력	.023	1.024	.016	1.016	.021	1.022
모학력	.055**	1.057	.055**	1.057	.058**	1.060
부모미취업더미 (부모취업기준)	.021	1.021	.052	1.053	.028	1.028
한부모가구더미 (양부모가구기준)	.296	1.345	.178	1.195	.309	1.362
가구원수	-.044	.957	-.071	.931	-.047	.954
월평균가구소득(log)	.503**	1.653	.371*	1.450	.493**	1.637
자가거주아님더미 (자가거주기준)	.099	1.104	.129	1.138	.107	1.113
거주지대도시더미 (중소도시기준)	-.109	.897	-.079	.924	-.096	.909
거주지농촌더미 (중소도시기준)	-.015	.985	-.057	.944	.016	1.016

* $p<.05$, ** $p<.01$, *** $p<.001$

먼저 표 6에 나타난 중2패널의 최초표본이탈에 대한 콕스비례위험회귀분석 결과부

터 살펴보면, 청소년최초표본이탈(event), 부모(보호자)최초표본이탈(event), 청소년부모(보호자)동시최초표본이탈(event) 모두 동일하게, 남성더미(여성기준), 학업성적, 모(여자보호자)학력, 월평균가구소득(log) 변수가 통계적으로 유의한 것으로 나타났고, 나머지 부(남자보호자)학력, 부모(보호자)미취업더미(부모(보호자)모두취업기준), 한부모가구더미(양부모가구기준), 가구원수, 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준), 거주지농촌더미(중소도시기준) 변수는 통계적으로 유의하지 않은 것으로 나타났다.

즉, 중2패널의 경우에는 여성에 비해 남성일수록, 학업성적이 낮을수록, 모(여자보호자)학력이 높을수록, 월평균가구소득(log)이 높을수록 최초표본이탈위험이 높은 것으로 나타났다. 이 중에서 월평균가구소득(log) 변수가 최초표본이탈위험에 미치는 영향력이 제일 큰 것으로 나타났고, 남성더미(여성기준) 변수의 영향력이 그 다음 큰 것으로 나타났으며, 이어서 모(여자보호자)학력 변수, 학업성적 변수의 순으로 영향력이 큰 것으로 나타났다.

월평균가구소득 변수는 기존 선행연구들에서 일부 상반된 연구결과도 있긴 했지만 전반적으로 가구소득이 높을수록 표본이탈이 높게 나왔던 변수였는데, 기존 선행연구 중의 Fitzgerald, Gottschalk & Moffitt(1998), MaCurdy, Mroz & Gritz(1998), 심영상과 허명회(2002), 송헌재(2012)의 연구결과와 유사하게 이 연구에서도 월평균가구소득이 높을수록 표본이탈위험이 높게 나타났다. 이는 가구소득이 높은 가구일수록 패널조사에 대한 협조정도가 낮음을 의미한다.

성별 변수는 기존의 몇몇 선행연구들에서 다소 상반된 결과가 나왔던 변수였는데, 이 연구에서는 기존 선행연구 중의 김대일, 남재량과 류근관(2000)의 연구결과와 비슷하게 여성에 비해 남성일수록 표본이탈위험이 높게 나타났다. 이는 남성의 패널조사에 대한 협조정도가 상대적으로 낮음을 의미한다.

학력 변수도 기존 몇몇 선행연구들에서 다소 상반된 결과가 나왔던 변수였는데, 이 연구에서는 기존 선행연구 중의 심영상과 허명회(2002), 이상협, 박찬용, 정성석과 최혜미(2011)의 연구과 비슷하게 모(여자보호자)학력이 높을수록 표본이탈위험이 높게 나타났다. 부학력은 통계적으로 유의하지 않고 모학력이 통계적으로 유의하게 나타난 것은 패널조사 협조에 부보다는 모의 영향이 더 크게 작용하고 있음을 나타내는 것으로 볼 수 있는데, 특히 모학력이 높을수록 표본이탈위험이 높게 나타난 것은 모학력

이 높을수록 패널조사에 대한 협조정도가 낮음을 의미한다.

학업성적 변수는 기존의 선행연구 중에서 유일하게 서울아동패널조사를 대상으로 분석했던 김세원과 정익중(2012)의 연구에 투입된 변수였는데, 김세원과 정익중(2012)의 연구결과와 유사하게 이 연구에서도 학업성적이 낮을수록 표본이탈위험이 높게 나타났다. 이는 학업성적이 낮을수록 패널조사에 대한 협조정도가 낮음을 의미한다.

표 7

초4패널-최초표본이탈에 대한 콕스비례위험회귀분석 결과

	청소년 최초표본이탈(event)		부모 최초표본이탈(event)		청소년부모동시 최초표본이탈(event)	
	B	Exp(B)	B	Exp(B)	B	Exp(B)
남성더미(여성기준)	.174	1.190	.089	1.093	.141	1.151
학업성적	.053	1.054	.102	1.108	.061	1.063
부학력	.020	1.020	.008	1.008	.017	1.017
모학력	.088**	1.092	.082**	1.086	.093**	1.097
부모미취업더미 (부모취업기준)	.120	1.128	.071	1.073	.094	1.099
한부모가구더미 (양부모가구기준)	.703*	2.019	.634*	1.884	.706*	2.025
가구원수	-.043	.958	-.086	.918	-.048	.953
월평균가구소득(log)	.518	1.679	.347	1.415	.450	1.569
자가거주아님더미 (자가거주기준)	.262**	1.299	.212*	1.236	.243*	1.275
거주지대도시더미 (중소도시기준)	.190	1.209	.207*	1.230	.202*	1.223
거주지농촌더미 (중소도시기준)	-.400	.670	-.332	.718	-.416	.660

* $p<.05$, ** $p<.01$, *** $p<.001$

다음으로 표 7에 나타난 초4패널의 최초표본이탈에 대한 콕스비례위험회귀분석 결과를 살펴보면, 청소년최초표본이탈(event)의 경우에는 모(여자보호자)학력, 한부모가

구더미(양부모가구기준), 자가거주아님더미(자가거주기준) 변수가 통계적으로 유의한 것으로 나타났고, 나머지 남성더미(여성기준), 학업성적, 부(남자보호자)학력, 부모(보호자)미취업더미(부모(보호자)모두취업기준), 가구원수, 월평균가구소득(log), 거주지대도시더미(중소도시기준), 거주지농촌더미(중소도시기준)는 통계적으로 유의하지 않은 것으로 나타났다.

즉, 모(여자보호자)학력이 높을수록, 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록 최초표본이탈위험이 높은 것으로 나타났다. 이 중에서 한부모가구더미(양부모가구기준) 변수가 최초표본이탈위험에 미치는 영향력이 제일 큰 것으로 나타났고, 그 다음으로 자가거주아님더미(자가거주기준) 변수의 영향력이 큰 것으로 나타났으며, 다음으로 모(여자보호자)학력 변수의 영향력이 큰 것으로 나타났다.

부모(보호자)최초표본이탈(event), 청소년부모(보호자)동시최초표본이탈(event)은 모두 동일하게, 모(여자보호자)학력, 한부모가구더미(양부모가구기준), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준) 변수가 통계적으로 유의한 것으로 나타났고, 나머지 남성더미(여성기준), 학업성적, 부(남자보호자)학력, 부모(보호자)미취업더미(부모(보호자)모두취업기준), 가구원수, 월평균가구소득(log), 거주지농촌더미(중소도시기준)는 통계적으로 유의하지 않은 것으로 나타났다.

즉, 모(여자보호자)학력이 높을수록, 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록, 거주지가 중소도시에 비해 대도시일수록 최초표본이탈위험이 높은 것으로 나타났다. 이 중에서 한부모가구더미(양부모가구기준) 변수가 최초표본이탈위험에 미치는 영향력이 제일 큰 것으로 나타났고, 자가거주아님더미(자가거주기준) 변수의 영향력이 그 다음으로 큰 것으로 나타났으며, 그 다음으로 거주지대도시더미(중소도시기준) 변수의 영향력이 큰 것으로 나타났고, 그 다음으로 모(여자보호자)학력 변수의 영향력이 큰 것으로 나타났다.

그런데 청소년최초표본이탈(event)의 경우에도 거주지대도시더미(중소도시기준) 변수의 통계적 유의도가 0.050으로 95% 유의수준에 거의 가깝게 나왔다는 점을 감안하면, 초4패널도 중2패널과 마찬가지로 청소년최초표본이탈(event), 부모(보호자)최초표본이탈(event), 청소년부모(보호자)동시최초표본이탈(event) 모두 비슷한 결과가 나왔다고도 할 수 있다.

한부모가구여부 변수는 기존 몇몇 선행연구들에서 한부모가구일수록 표본이탈이 높

게 나왔었는데, 기존 선행연구 중의 Lillard & Panis(1998), Watkins & Melde(2007), 김대일, 남재량과 류근관(2000)의 연구결과와 유사하게 이 연구에서도 양부모가구에 비해 한부모가구일수록 표본이탈위험이 높게 나타났다. 이는 한부모가구일수록 패널조사에 대한 협조정도가 낮음을 의미한다.

자가거주여부 변수도 기존 몇몇 선행연구들에서 자가거주가 아닐수록 표본이탈이 높게 나왔었는데, 기존 선행연구 중의 심영상과 허명회(2002), 이상호(2005), 이상협, 박찬용, 정성석과 최혜미(2011)의 연구결과와 유사하게 이 연구에서도 자가거주에 비해 자가거주가 아닐수록 표본이탈위험이 높게 나타났다. 이는 자가거주가 아닐수록 패널조사에 대한 협조정도가 낮음을 의미한다.

거주지대도시더미(중소도시기준) 변수는 기존 몇몇 선행연구들에서 일부 상반된 연구결과도 있긴 했지만 전반적으로 대도시에 거주할수록 표본이탈이 높게 나왔던 변수였는데, 기존 선행연구 중의 이상협, 박찬용, 정성석과 최혜미(2011), 이경희와 민인식(2016)의 연구결과와 유사하게 이 연구에서도 거주지가 중소도시에 비해 대도시일수록 표본이탈위험이 높게 나타났다. 이는 거주지가 대도시일수록 패널조사에 대한 협조정도가 상대적으로 낮음을 의미한다.

학력 변수는 기존 몇몇 선행연구들에서 다소 상반된 결과가 나왔던 변수였는데, 기존 선행연구 중의 심영상과 허명회(2002), 이상협, 박찬용, 정성석과 최혜미(2011)의 연구결과 비슷하게 이 연구에서도 모(여자보호자)학력이 높을수록 표본이탈위험이 높게 나타났다. 앞의 중2패널 분석결과에서 살펴본 바와 같이, 부학력은 유의하지 않은 반면 모학력이 유의하게 나타난 것은 패널조사 협조에 부보다는 모의 영향이 상대적으로 더 크게 작용하고 있음을 나타내는 것으로 볼 수 있다. 그리고 모학력이 높을수록 표본이탈위험이 높게 나타난 것은 모학력이 높을수록 패널조사에 대한 협조정도가 상대적으로 낮음을 의미한다.

중2패널 분석결과와 초4패널 분석결과를 비교해보면 종단적 최초표본이탈의 영향요인에 대한 분석결과의 양상이 상당히 다르게 나타났음을 알 수 있다.

앞에서 살펴보았듯이 중2패널 분석결과는 여성에 비해 남성일수록, 학업성적이 낮을수록, 모(여자보호자)학력이 높을수록, 월평균가구소득(log)이 높을수록 최초표본이탈위험이 높은 것으로 나타난 반면, 초4패널 분석결과는 모(여자보호자)학력이 높을수록, 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록, 거

주지가 중소도시에 비해 대도시일수록 최초표본이탈위험이 높은 것으로 나타났다.

모(여자보호자)학력 변수만이 공통적으로 최초표본이탈에 영향을 미치는 것으로 나타났고, 중2패널에는 남성더미(여성기준), 학업성적, 월평균가구소득(log) 변수가, 초4패널에는 한부모가구더미(양부모가구기준), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준) 변수가 각각 고유하게 최초표본이탈에 영향을 미치는 것으로 나타났다.

이러한 분석결과는 중2패널이든 초4패널이든 모두 패널조사 협조정도에 모가 영향을 미치고 있는 가운데, 중2패널의 경우에는 남성더미(여성기준), 학업성적과 같은 개인적인 특성이 패널조사 협조정도에 영향을 미치고 있는 것으로 나타난 반면, 초4패널의 경우에는 한부모가구더미(양부모가구기준), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준)와 같은 배경적인 특성이 패널조사 협조정도에 영향을 미치고 있음을 의미한다.

Ⅴ. 결 론

이 연구에서는, 2000년대 들어와서 패널데이터 구축이 활발하게 진행되어오고 있는 상황에서, 향후 아동 및 청소년 대상 패널데이터 구축시 표본이탈을 최소화해 좋은 데이터를 만드는데 도움을 드리고자, 기 진행된 한국청소년패널조사 데이터의 종단적 최초표본이탈에 영향을 미치는 표본특성요인에는 어떠한 것들이 있는지를 경험적으로 분석해보고자 하였다.

이를 위해 기존 국내외 관련 선행연구들에서 주로 어떤 변수들이 패널데이터 표본이탈의 영향요인으로 분석되었는지를 검토하고, 한국청소년패널조사 데이터에서 실제로 독립변수로 추출할 수 있는 변수에는 어떠한 것들이 있는가를 검토하여, 성별, 학업성적, 부(남자보호자)학력, 모(여자보호자)학력, 부모(보호자)취업여부, 한부모가구여부, 가구원수, 월평균가구소득, 자가거주여부, 거주지규모의 독립변수가 청소년최초표본이탈, 부모(보호자)최초표본이탈, 청소년부모(보호자)동시최초표본이탈의 종속변수에 미치는 영향을 연구모형으로 설정한 후, 한국청소년패널조사 중2패널 1~6차년도, 초4패널 1~5차년도 데이터를 대상으로 생존분석(survival analysis) 중 콕스의 비례위험

회귀분석 모형(cox proportional hazard model)을 적용해 경험적으로 분석하였다.

분석결과, 중2패널에서는 여성에 비해 남성일수록, 학업성적이 낮을수록, 모(여자보호자)학력이 높을수록, 월평균가구소득(log)이 높을수록 최초표본이탈위험이 높게 나타났고, 초4패널에서는 모학력(여자보호자)이 높을수록, 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록, 거주지가 중소도시에 비해 대도시일수록 최초표본이탈위험이 높게 나타났다.

중2패널과 초4패널의 최초표본이탈위험에 공통적으로 영향을 미치는 것으로 나타난 변수는 모(여자보호자)학력 변수뿐이었고, 중2패널에는 남성더미(여성기준), 학업성적, 월평균가구소득(log) 변수가, 초4패널에는 한부모가구더미(양부모가구기준), 자가거주아님더미(자가거주기준), 거주지대도시더미(중소도시기준) 변수가 각각 따로 최초표본이탈위험에 영향을 미치는 것으로 나타났다.

이러한 분석결과는 향후 아동 및 청소년 대상의 패널데이터를 구축할 때 표본이탈을 최소화하기 위한 표본관리방법 강화에 어느 정도 도움을 줄 수 있다.

현재도 패널데이터 구축시 표본이탈을 최소화하기 위해 홈페이지 구축, 정기적인 연락, 생일축하카드, 패널진행경과에 대한 팜플렛 발송, 동의하에 주변지인들 연락처 확보, 경품지급 등 일반적인 표본관리방법과 더불어 표본들 중 이탈위험이 큰 표본들을 선정하여 더 집중적으로 관리하는 이탈위험 표본관리방법을 사용하고 있다. 현재의 이탈위험 표본관리는 조사 및 표본관리과정에서 타 표본에 비해 소극적이고 비협조적인 태도를 보인다거나 타 표본에 비해 연락이 잘 안된다거나 하는 표본을 조사면접원 및 표본관리담당자의 견해를 바탕으로 선정해서 더 많은 예산을 책정하여 더 많은 연락을 취하여 표본이탈을 최소화하는 방법으로 진행하고 있다. 이러한 이탈위험 표본관리방법은 기 구축된 패널들의 표본이탈을 최소화하는데 중요한 기여를 해 왔다.

그런데 향후 아동 및 청소년대상의 패널데이터 구축시 표본이탈을 최소화하기 위해서는 현재의 이탈위험 표본관리방법에 더해서 이탈위험 표본특성을 바탕으로 한 모니터링을 실시해서 이탈위험 표본관리방법을 한층 더 강화하는 것이 좋을 것이라고 생각된다. 기존 아동 및 청소년대상의 표본이탈 영향요인에 대한 연구결과들을 바탕으로, 1차년도 패널조사 표본구축 후에 데이터를 분석해서 이탈위험이 높은 표본특성군을 선정해서 종단적 조사가 끝나는 시점까지 지속적으로 모니터링하면서 이탈위험 표본관리를 시행하면 더 좋을 것이다. 가령, 기존 연구결과에서 표본이탈에 영향을 미

치는 것으로 나타난 표본특성요인을 영향력의 크기별로 상, 중, 하로 구분해놓고, 1차년도 패널조사 구축후 표본의 특성들을 분석해서, 어떤 특성들을 지닌 표본은 이탈위험이 상상인 표본으로, 어떤 특성들을 지닌 표본은 이탈위험이 중하인 표본으로 구분한 후, 그에 맞게 표본관리 예산을 차등적으로 배분하고, 표본관리 빈도 및 시간을 다르게 시행하는 것이 이탈위험 표본관리 모니터링의 구체적인 예가 될 수 있을 것이다.

이러한 모니터링 실시에 이 연구의 연구결과도 하나의 참조가 될수 있을 것이다. 중등패널이든 초등패널이든 모(여자보호자)학력이 높을수록 최초표본이탈위험이 높게 나타났으므로 아동 및 청소년 전체를 대상으로 패널을 구축할 때의 모니터링에 참조하면 좋을 것이다. 또한 중등패널의 경우에는 여성에 비해 남성일수록, 학업성적이 낮을수록, 월평균가구소득(log)이 높을수록 최초표본이탈위험이 높게 나타났으므로 중등패널을 구축할 때의 모니터링에 참조하면 좋을 것이다. 그리고 초등패널의 경우에는 양부모가구에 비해 한부모가구일수록, 자가거주에 비해 자가거주가 아닐수록, 거주지가 중소도시에 비해 대도시일수록 최초표본이탈위험이 높게 나타났으므로 초등패널을 구축할 때의 모니터링에 참조하면 좋을 것이다. 물론 이 연구결과이외에 기 구축된 아동 및 청소년 대상의 다른 패널데이터들의 표본이탈의 영향요인에 대한 연구들도 축적되어 함께 참조하면 좋을 것이다.

참고문헌

강원도교육연구원 (2013). **강원 학생 교육성취도 종단연구 결과보고서.** 강원: 강원도교육연구원.

경기도교육연구원 (2014). **경기교육종단연구 2차년도 기초분석보고서.** 경기: 경기도교육연구원.

김대일, 남재량, 류근관 (2000). 한국노동패널표본의 대표성과 패널조사 표본이탈자의 특성 연구. **노동경제논집, 23,** 1-33.

김세원, 정익중 (2012). 아동청소년패널조사의 항목무응답과 단위무응답에 영향을 미치는 요인. **한국청소년연구, 23**(3), 27-50.

김영석 (1999). **사회조사방법론.** 서울: 나남출판.

남기성, 이기성, 박근화, 홍기학, 손창균 (2014). 2012년 고령화패널조사 가중값 부여 방법 연구. **Journal of the Korean Data Analysis Society, 13**(3), 1291-1301.

대우경제연구소 (1994). **한국가구경제활동연구.** 서울: 대우경제연구소.

민주홍 (2005). 패널조사의 필요성과 연구동향. **직업과 인력개발, 8**(4), 108-117.

박재빈 (2007). **생존분석-이론과 실제.** 서울: 신광출판사.

서울특별시교육연구정보원 (2011). **서울교육종단연구2010 사용자매뉴얼.** 서울: 서울특별시교육연구정보원.

송경일, 최종수 (2008). **SPSS15를 이용한 생존자료의 분석.** 서울: 한나래출판사.

송헌재 (2012). 재정패널 2차년도 이탈가구와 대체가구의 특성 분석. **재정학연구, 5**(1), 97-133.

신부용, 박현선 (2016). 다문화가정 자녀의 부모-자녀관계와 성인역할모델이 학교적응에 미치는 영향: 자기효능감의 매개과정을 중심으로. **학교사회복지, 33,** 1-24.

심영상, 허명회 (2002). 한국노동패널조사에서 표본이탈에 대한 통계적 모형화 및 활용. **응용통계, 17,** 31-46.

이경상 (2007). 국내 패널조사의 현황 및 시사점. **통계로 본 인적자원 동향 제2호.** 서울: 교육인적자원부.

이경상 (2017). 국내 패널조사 진행현황. 미간행 출판물.

이경희, 민인식 (2016). 패널조사 응답지속성에 관한 연구: 한국노동패널조사를 중심

으로. **조사연구, 17**(3), 1-24.

이규용, 이지은, 신선옥, 이혜정, 김기홍 (2015). **제16차년도 한국 가구와 개인의 경제활동-한국노동패널 기초분석보고서.** 서울: 한국노동연구원.

이상균, 그레이스정, 유조안 (2015). 잠재계층분석을 통한 아동의 건강증진행위 유형화와 영향요인 분석. **보건사회연구, 35**(2), 477-510.

이상협, 박찬용, 정성석, 최혜미 (2011). 한국노동패널 탈락 분석. **한국데이터정보과학회지, 22**(1), 1-8.

이상호 (2005). 한국노동패널(KLIPS)의 표본이탈분석: 가구소득을 중심으로. **노동리뷰, 11,** 66-80.

이성진 (2005). **한국인의 성장 · 발달-30년 종단적 연구.** 서울: 교육과학사.

이지연, 김진 (2007). 횡단조사자료 종단화의 가치와 한계: 경제활동인구조사와 도시가계조사. **여성가족 패널브리프 2007,** 가을 2호, 32-59.

이화정, 강석복 (2014). 한국아동청소년패널조사 자료에서 단위무응답의 실태 및 가중치 적용. **한국데이터정보과학회지, 25**(6), 1397-1405.

천영민, 윤정혜, 오민홍 (2009). 대졸자 직업이동 경로조사에서 패널탈락분석. **응용통계연구, 22**(5), 981-993.

한국교육개발원 (2010). **탈북청소년의 교육 종단연구** Ⅰ. 서울: 한국교육개발원.

한국교육개발원 (2012). **한국교육종단연구2005(VIII).** 서울: 한국교육개발원.

한국직업능력개발원 (2015). **한국교육고용패널 기초분석보고서(2014).** 세종: 한국직업능력개발원.

한국청소년개발원 (2001). **청소년패널조사 기초연구.** 서울: 한국청소년개발원.

한국청소년정책연구원 (2009a). **한국청소년패널조사 중2패널 1~6차년도 User's Guide.**

한국청소년정책연구원 (2009b). **한국청소년패널조사 초4패널 1~5차년도 User's Guide.**

한국청소년정책연구원 (2010a). **한국아동청소년패널조사2010 I.** 서울: 한국청소년정책연구원.

한국청소년정책연구원 (2010b). **다문화가족 아동 · 청소년의 발달과정 추적을 위한 종단연구** Ⅰ. 서울: 한국청소년정책연구원.

한국청소년정책연구원 (2011). **취약계층 아동 · 청소년 종단조사** Ⅱ. 서울: 한국청소년정책연구원.

Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. North Carolina: SAS Institute INC.

Babbie, Earl (2001). *The Practice of Social Research.* Wadsworth Group, 고성호 외 역 (2002). **사회조사방법론.** 서울: 도서출판 그린.

Diggle, P. J., Heagerty, P. J., Liang, K. Y., & Zeger, S. T. (2002). *Analysis of longitudinal data*. London: Oxford University Press.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *The Journal of Human Resources, 33*(2), 251-299.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Hill, D., & Willis, R. J. (2001). Reducing panel attrition: A search for effective policy instruments. *The Journal of Human Resources, 36*(3), 416-438.

Lillard, L. A., & Panis, C. W. A. (1998). Panel attrition from the panel study of income dynamics: Household income, marital status, and mortality. *The Journal of Human Resources, 33*(2), 437-457.

MaCurdy, T., Mroz, T., & Gritz, R. M. (1998). An evaluation of the national longitudinal survey of youth. *The Journal of Human Resources, 33*(2), 345-436. University of Wisconsin Press.

Taris, T. (2000). *A primer in longitudinal data analysis.* London: Sage Publications.

Vermunt, J. K. (1997). *Log-linear models for event histories*. London: Sage Publications.

Watkins, A. M., & Melde, C. (2007). The effect of self-control on unit and item nonresponse in an adolescent sample. *Journal of Research in Crime and Delinquency, 44*(3), 267-294.

Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *The Journal of Human Resources, 33*(2), 479-506.

ABSTRACT

The sample characteristics affecting longitudinal initial sample attrition in the Korea youth panel survey

Lee, Kyeong-Sang* · Lee, Soon-Rae**

In this study, an attempt was made to empirically analyze the sample characteristics that affect the longitudinal initial sample attrition in data taken from the Korea Youth Panel Survey. The data was taken from the 1st through to the 6th annual datasets of the second grade junior high school youth panel data and the 1st through the 5th annual datasets of the fourth grade elementary school youth panel data. The independent variables consisted of gender, academic performance, father(male guardian)'s level of education, mother(female guardian)'s level of education, parental(guardian's) employment status, single parent households, number of household members, monthly average household income, whether the primary property is owner occupied or rented, and size of the primary property. The dependent variables consisted of the adolescent's longitudinal initial sample attrition, the parent(guardian)'s longitudinal initial sample attrition, and the adolescent- parent(guardian)'s concurrent longitudinal initial sample attrition. The cox proportional hazard model was applied for the purposes of analysis. The results of this study were as follows: The higher the proportion of males compared to females, the lower the level of academic performance. Similarly, the higher the level of the mother(female guardian)'s level of education. Finally it was found that the higher the monthly average household income, the higher the risk of

* First Author, National Youth Policy Institute, Senior Researcher
** Corresponding Author, Wonkwang University, College of Police Administration, Professor, soonlee@wonkwang.ac.kr

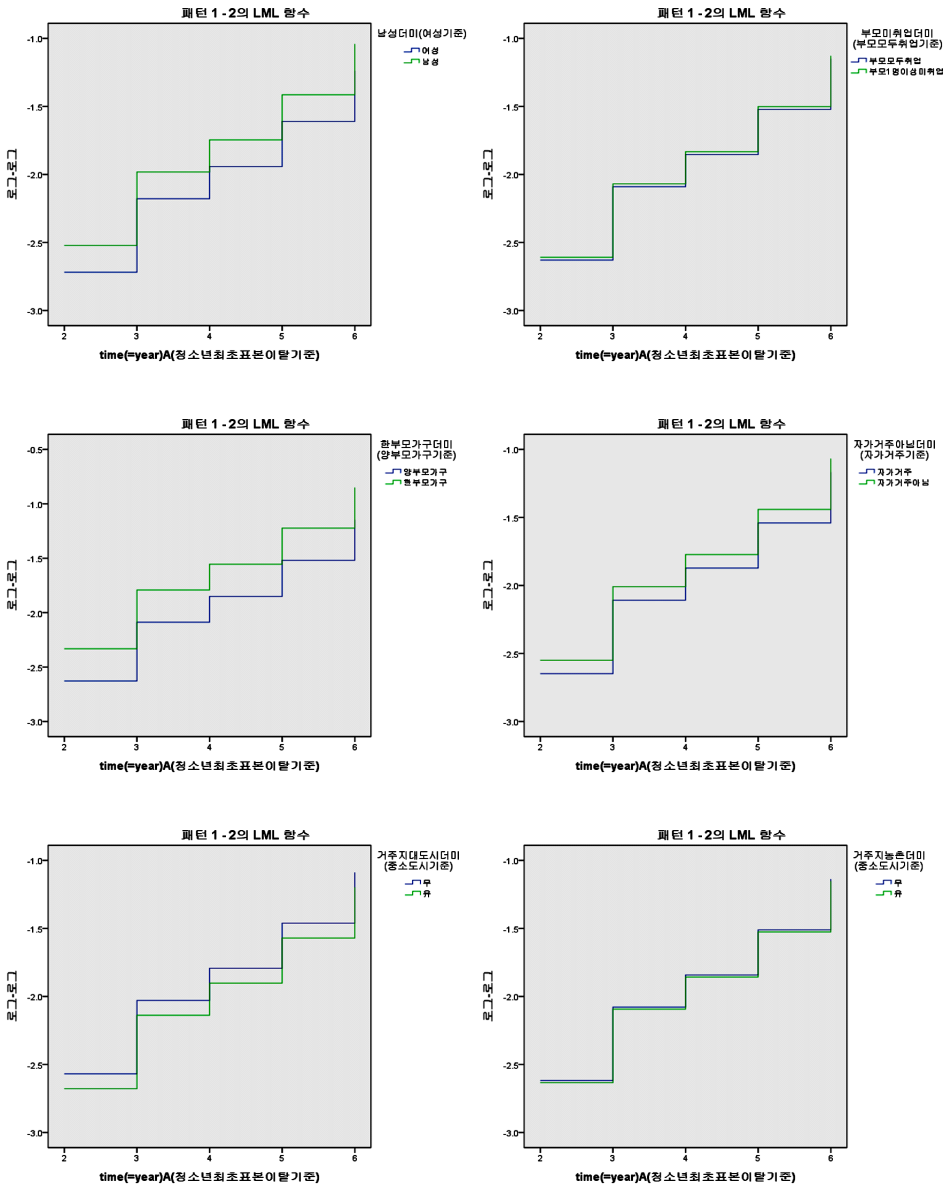
longitudinal initial sample attrition was in the analysis of the 1st through to the 6th annual datasets of the second grade junior high school youth panel data. The following relationships led to a higher risk of longitudinal initial sample attrition in the analysis of 1st through the 5th annual datasets of the fourth grade elementary school youth panel. The higher the mother(female guardian)'s level of education, the higher the single parent households compared to the two parent households, the higher the proportion of primary properties being rented rather than owner occupied, and the higher proportion of metropolitan city residence compared to small-medium city residence –all led to higher risks of attrition. The only variable that had a common effect on the risk of longitudinal initial sample attrition of the second grade junior high school youth panel and fourth grade elementary school youth panel was the variable of the mother(female guardian)'s level of education. The variables such as the proportion of males compared to females, the level of academic performance, and the monthly average household income affected only the risk of longitudinal initial sample attrition of the second grade junior high school youth panel. The variables such as single parent households compared to more conventional two parent households, ownership status of the primary property and the proportion of metropolitan city residence compared to small-medium city residence affected only the risk of longitudinal initial sample attrition in the fourth grade elementary school youth. panel. The results of this analysis should help to strengthen sample management methods to minimize sample attrition when creating panel data for children and adolescents in the future.

Key Words: Korea youth panel survey, sample attrition, sample characteristics

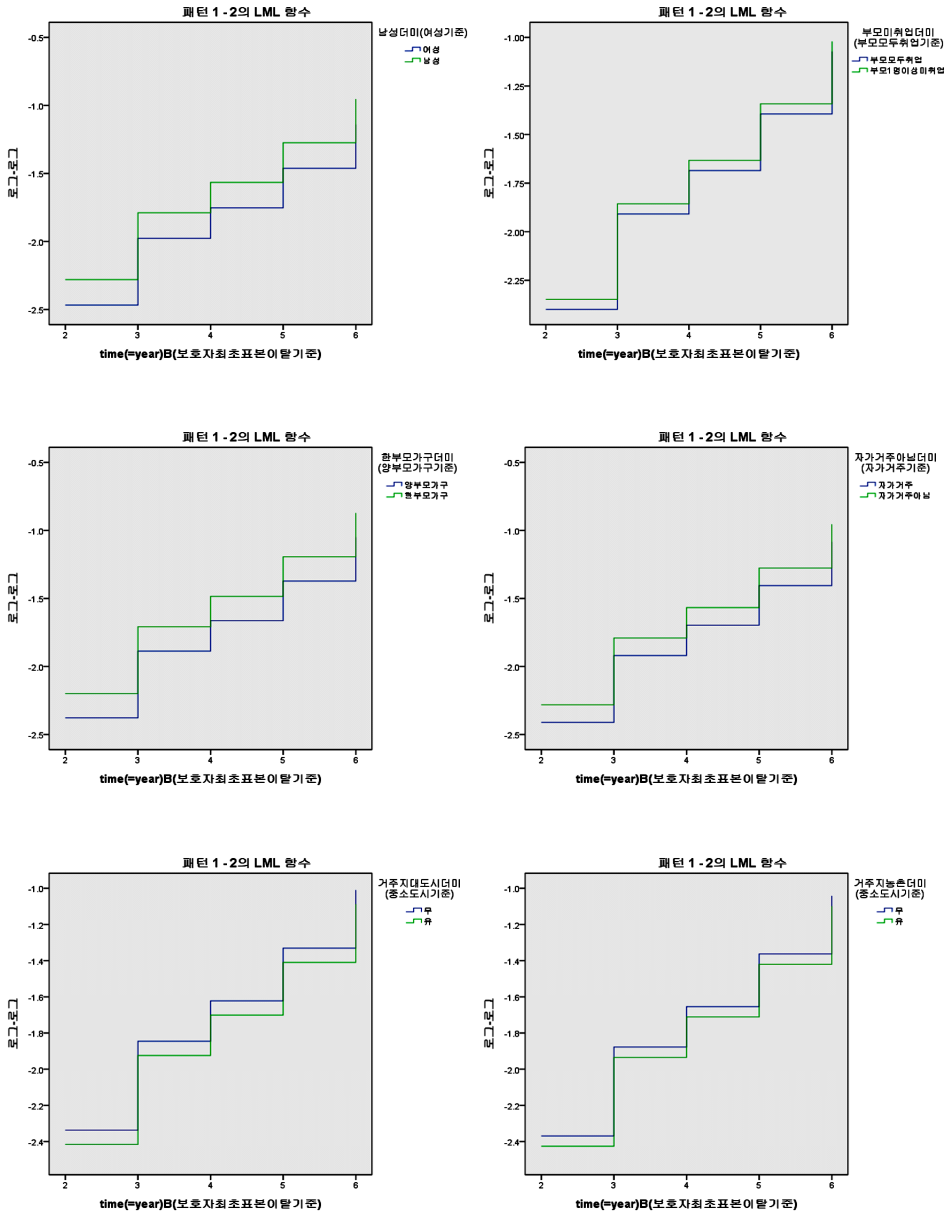
투고일: 2017. 3. 13, 심사일: 2017. 5. 2, 심사완료일: 2017. 5. 12

[부 록]

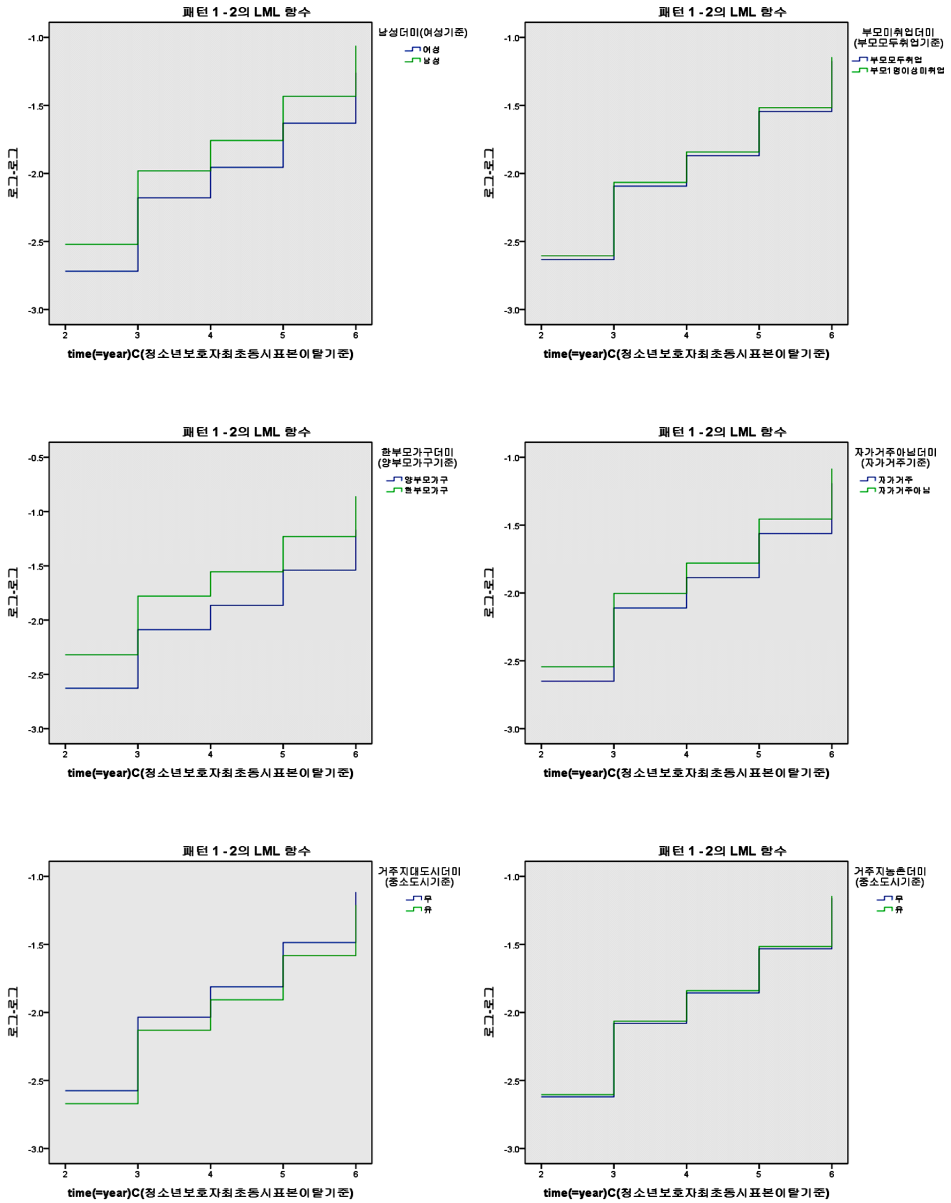
중2패널-청소년최초표본이탈에 대한 콕스비례위험회귀분석 LML PLOT



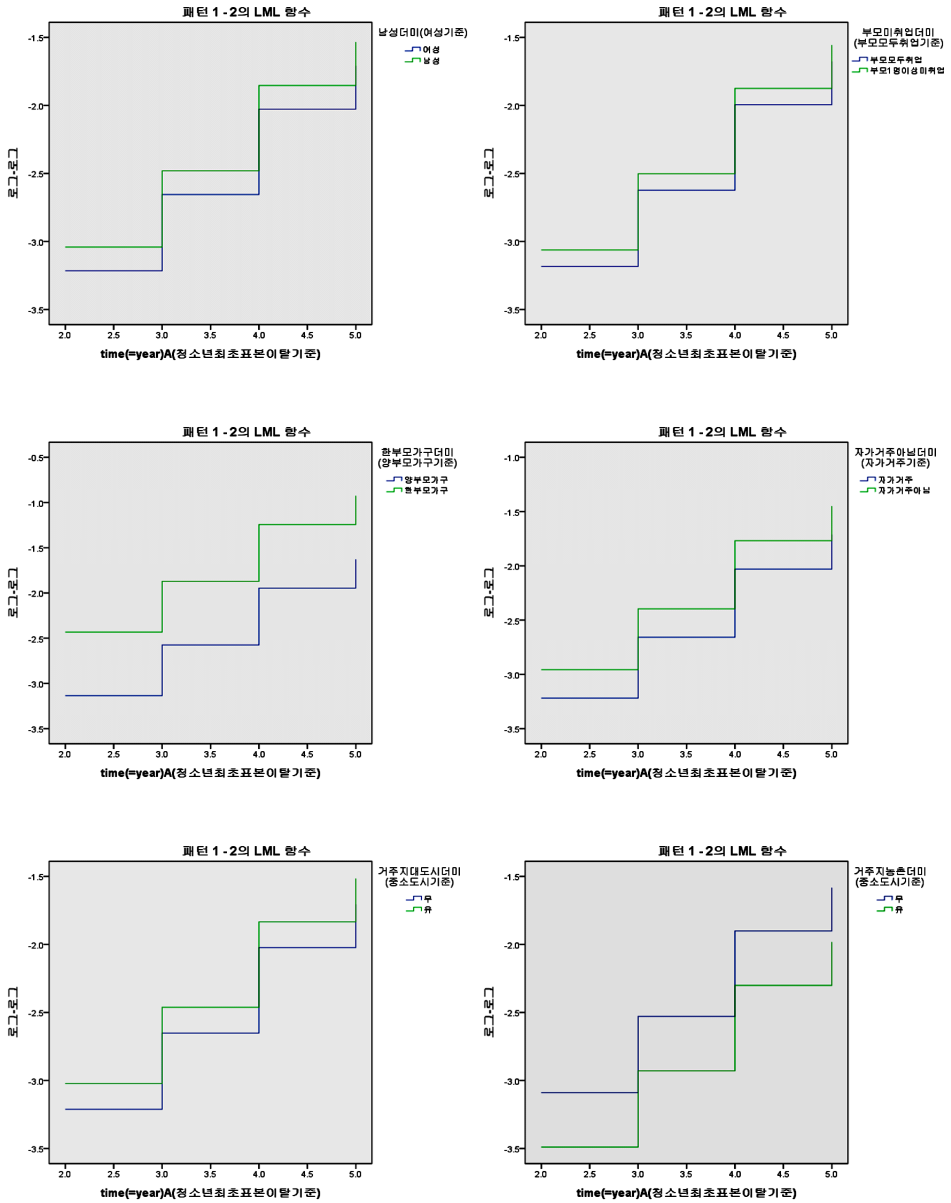
중2패널-보호자최초표본이탈에 대한 콕스비례위험회귀분석 LML PLOT



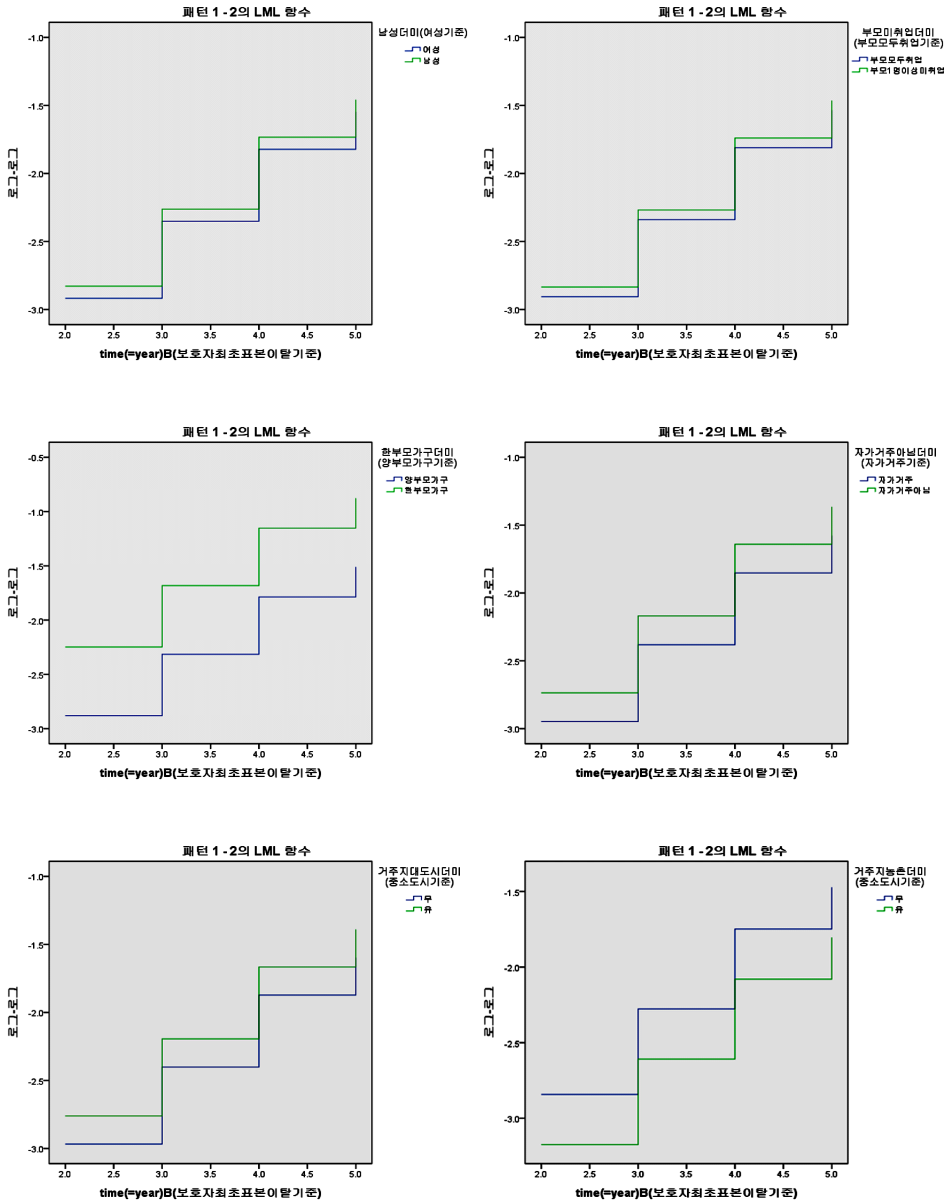
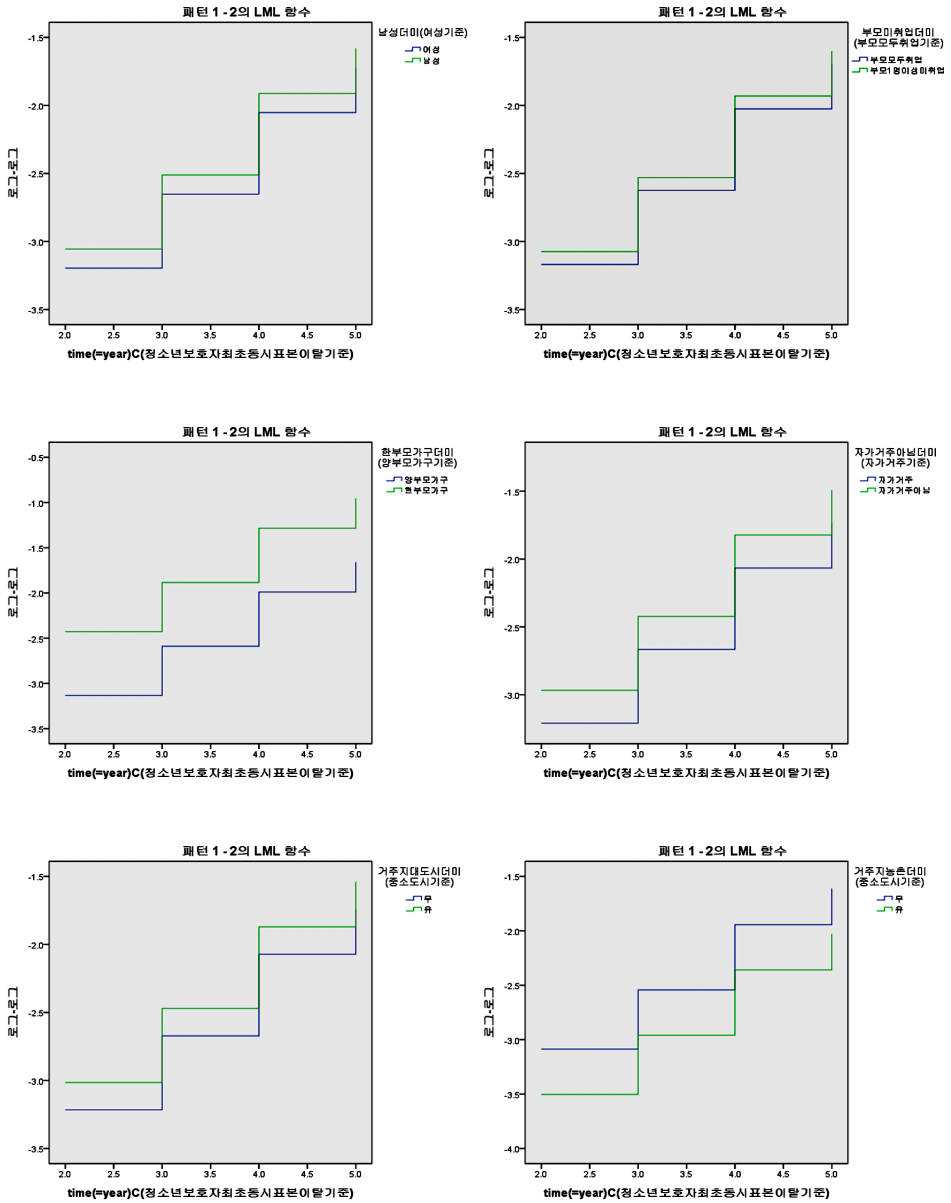
중2패널-청소년보호자최초동시표본이탈에 대한 콕스비례위험회귀분석 LML PLOT



초4패널-청소년최초표본이탈에 대한 콕스비례위험회귀분석 LML PLOT



초4패널-보호자최초표본이탈에 대한 콕스비례위험회귀분석 LML PLOT



초4패널-청소년보호자최초동시표본이탈에 대한 콕스비례위험회귀분석 LML PLOT