

아동청소년패널조사의 항목무응답과 단위무응답에 영향을 미치는 요인*

김세원** · 정익중***

초 록

패널조사가 사회과학분야에서 크게 유행하고 있지만 무응답에 의한 결측이 체계적으로 발생할 경우 패널조사 자료에서 편익된 추정이 발생할 수 있다. 본 연구에서는 서울아동패널 자료를 사용하여 항목무응답과 단위무응답에 영향을 주는 요인과, 항목무응답과 단위무응답 간의 관계에 대해 분석하였다. 음이항 회귀분석(negative binomial regression)을 사용하여 분석한 결과, 아동의 성별, 유능감, 학업성취수준, 낮은 자기통제는 1차년도 조사문항 가운데 응답하지 않은 문항이 있는 항목무응답에 유의한 영향을 미치는 것으로 나타났다. 그러나 6회기에 걸친 추적조사에 한 번이라도 참여하지 않은 것을 의미하는 단위무응답의 경우 학업성취수준만이 유의한 영향을 미쳤다. 또한 항목무응답은 단위무응답을 예측하는 것으로 나타났는데, 1차년도 조사에서 한 가지 이상의 설문 문항에 응답을 하지 않았을 경우 이후 추적조사에 참여하지 않을 가능성이 더 높았다. 조사문항의 성격에 따라 항목무응답이 단위무응답에 미치는 영향은 다소 차이를 보였는데 다른 조사문항과 달리 1차년도 조사에서 인적사항에 대한 질문에 응답하지 않은 항목무응답만이 단위무응답을 예측하였다. 이러한 결과는 자료의 질을 향상시키기 위해서는 단위무응답뿐만 아니라 항목무응답을 감소시키기 위한 노력이 동시에 있어야 함을 보여주는 것이다.

주제어 : 패널조사, 항목무응답, 단위무응답, 자기통제

* 본 논문은 2011년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2011-413-B00011).

** 서울대학교 사회과학연구원 선임연구원, 제1저자

*** 이화여자대학교 사회복지전문대학원 부교수, 교신저자, ichung@ewha.ac.kr

I. 서 론

우리나라 사회과학연구에서 동일한 대상 집단을 일정한 시차를 두고 반복 추적하는 자기보고식 패널조사(panel survey)가 크게 유행하고 있다. 1993년 대우경제연구소에서 시작한 한국가구패널조사(KHPS: Korea Household Panel Study), 일명 ‘대우패널’을 통해 우리나라에 패널조사가 최초로 도입되었다. 민간 영역에서 먼저 도입된 대우패널이 IMF 경제위기를 맞아 중단되자, 이를 이어 한국노동연구원에서 1998년 한국노동패널을 운영하면서 공적 영역에서도 패널조사가 도입되었다(이희길, 2008). 한국노동패널이 성공적으로 안착되면서 2000년 이후 국책 연구기관, 공단, 대학을 중심으로 패널조사가 급속히 확산되었다. 2011년 기준 정부예산을 사용하여 중앙 정부 부처, 국책 연구기관 및 공단 등에서 운영 중인 패널조사만 27개에 달한다(이경상, 2007). 민간에서 수행하고 있는 것을 포함하면 그 규모는 더 클 것으로 추측되고 있으며, 지금도 여러 정부부처나 국책 연구기관 등에서 신규 패널조사를 계획하거나 시도하고 있어 향후 패널조사는 계속 증가할 것으로 예상된다.

실증적 통계분석에 기초해서 정책을 수립하거나 그 효과를 측정하고자 하는 정책입안자나 연구자들은 패널조사를 선호하는 경향이 있다(이희길, 2008). 횡단자료를 이용해서 특정 서비스의 효과를 검증하는 경우, 처치를 받는 사람과 그렇지 않는 사람이 원래부터 상이하다는 사실로 인해 연구에 어려움을 겪는다. 반면 패널조사는 동일한 표본을 계속적으로 조사하므로 변화요인이 되는 독립변수와의 노출 전후 상태를 관찰할 수 있어서 독립변수와 종속변수의 시간순서를 확립할 수 있고 전반적인 변화정도를 알 수 있는 장점이 있다. 또한 특정 기간효과(period specific effect), 특정 연령효과(age specific effect), 그리고 동년배효과(cohort effect) 등을 파악할 수 있게 한다(금재호, 1998). 패널자료는 시점 간 동태 정보와 개별 대상에 대한 정태 정보를 동시에 지니고 있기 때문에 결과에 영향을 미치는 측정되지 않은 변수를 통제할 상태에서 처치 효과만을 분리할 수 있다. 유사하게 패널자료는 미관측 이질성을 통제함으로써 시간의 경과에 따른 동태적 변화 및 상태 간 변이과정을 추적할 수 있을 뿐 아니라 관찰 불가능한 변수들을 효과적으로 통제함으로써 보다 엄격한 자연과학의 실험과 유사한 상황에서 특정 변수의 효과를 측정할 수 있다(이상호, 2006; 이희길, 2008).

패널조사가 지닌 이러한 장점에도 불구하고, 과거에 패널조사가 많이 진행되지 못

한 것은 무엇보다도 비용과 시간 측면에서의 현실적인 제약 때문이다. 우선 표본선정 후 한 번의 조사만을 실시하는 횡단조사와 달리, 패널조사는 동일대상을 반복해서 조사해야 하는 까닭에 패널유지와 관리를 위해 엄청난 비용이 소요된다. 시간적으로도 패널조사는 자료 축적에 상당시간을 요하며, 때로 관심 있는 현상에 대한 의미 있는 분석을 위해서는 장시간을 기다려야 하는 문제가 있다.

또한 동일 대상에 대해 반복적 조사를 하게 되면 패널조건화(panel conditioning)가 일어날 수 있다(Frankfort-Nachmias & Nachmias, 2000). 이전 조사를 기억하고 의식적으로 일관된 응답을 하려하기도 하고 반대로 의식적으로 변화를 보이는 방향으로 응답할 수 있어 실질적 변화를 명확하게 알 수 없게 될 가능성도 있다. 게다가 조사 때마다 매번 문제행동에 대한 질문을 받으면서 대상자들이 문제행동을 하지 말아야겠다는 각성이 일어날 수도 있다.

마지막으로 패널조사에서 가장 크게 부각되는 문제점 중 하나는 패널 유실(panel attrition) 문제이다. 회기(wave)가 반복될수록 이사, 응답거부, 행방불명, 사망 등의 사유로 1차년도 원표본에 대해 계속 조사하기가 어렵다. 특히, 표본 유실이 체계적으로 발생한 경우 남아 있는 패널의 대표성을 유지하기 어렵다. 예를 들어, 문제성향이 높은 개인은 패널조사에서 빨리 탈락하기 때문에 모집단에서 이러한 성향을 가진 집단이 과소 대표되는 문제가 발생할 수 있다. 그리고 이러한 성향과 결측이 체계적으로 관계가 있을 경우 편의된 추정이 발생할 수 있다. 이렇게 자기보고식 조사의 가장 부정적인 반응은 결측이고, 이는 항목무응답(item nonresponse)과 단위무응답(unit nonresponse)의 두 가지 형태로 나타난다. 항목무응답은 조사에 참여했으나 설문지의 모든 문항에 답변하지 않은 경우를 의미하고, 단위무응답은 표본에 속했던 대상이 거부, 이사 등 여러 가지 사유로 다음 회기 조사 자체에 참여하지 않는 것을 의미한다.

현재 국내외에서 패널조사 자료를 활용하여 수많은 원인론 연구들이 진행되고 있다. 자료의 이용가능성이 증가하면서 패널조사 자료를 분석한 연구 또한 기하급수적으로 증가하고 있다. 예를 들어 한국청소년패널조사 자료를 활용한 논문이 제1회 한국청소년패널학술대회에서 15편에 불과하던 것이 현재는 그 규모를 파악할 수 없을 만큼 그 숫자가 증가하였다. 하지만 패널조사 자체에 대한 연구는 절대적으로 부족하다. 따라서 본 연구는 패널조사 자체의 무응답에 대한 연구를 진행하려고 한다. 동일한 예측요인이 항목무응답과 단위무응답에 영향을 미치는지, 두 가지 유형의 무응답

은 서로 관련이 있는지에 대해 체계적으로 살펴보려고 한다. 본 연구는 무응답이 발생하는 원인을 파악하여 패널조사 자료의 결측치 처리방법과, 항목무응답과 단위무응답을 감소시킬 수 있는 방안에 대한 유용한 기초자료를 제공할 수 있을 것이다.

II. 이론적 배경

패널조사를 통해 수집된 자료는 개인의 행동이나 정서를 보다 심도 있게 이해할 수 있게 해주었다. 그러나 조사대상자들이 무작위로 패널조사에서 탈락하는 것이 아니라 특정 변수와 관련되어 있거나 의도적으로 특정 문항에 응답하지 않는다면, 패널조사의 자료는 편의된 추정의 문제점을 갖게 될 것이다. 현재 패널자료를 활용한 많은 연구가 진행되고 있음에도 불구하고, 아직까지 국내뿐만 아니라 외국에서도 패널자료 내에 발생한 무응답과 관련된 연구는 상당히 부족한 상황이다. 일부 연구에서는 비행이나 자기통제력, 범죄행동 등과 같은 개인의 특성이 무응답 발생과 관련이 있는 것으로 나타났다(Hirschi & Gottfredson, 1993; Piquero, MacIntosh & Hickman, 2000; Watkins & Melde, 2007).

비행성향이 높은 개인이 참여하면서 그들이 제공하는 대답은 덜 신뢰성이 있고 덜 타당한 응답을 하기 쉽다. 비행성향이 높은 개인은 그렇지 않은 개인에 비해 조사에 참여하지 않을 가능성이 높고 특히 반복적이고 지속적인 참여가 필요한 패널연구에서 이러한 가능성이 크다(Watkins & Melde, 2007). 일부 연구에서는 범죄행위가 사회조사 참여와 부적(-) 관계가 있는 것으로 나타났다(Hindelang, Hirschi & Weis 1981; Hirschi & Gottfredson, 1993). 반면 자기유능감이 높거나 학업성취가 높은 긍정적인 성향을 가진 개인은 결측없이 응답할 확률도 높고 장기조사에 거부없이 참여할 가능성이 높다. 예를 들어, Watkins과 Melde(2007)의 연구에서 학업성적이 중요하다고 인식하는지 여부 등과 같은 학교에 대한 헌신은 항목무응답과 단위무응답 모두와 관련이 있는 것으로 나타났다. 한편, 이들 연구에서는 개인의 심리적 및 행동적 특성 외에 한부모 가족 유형도 항목무응답과 단위무응답에 유의한 영향을 미치는 것으로 나타났다. 따라서 패널조사에서의 무응답을 정확하게 설명하기 위해서는 이러한 인구사회학적 요인들을 함께 고려할 필요가 있을 것이다.

비행행동은 개인 욕구의 즉각적 충족이 가능하기 때문에 지속되는 경향이 있다. 하지만 대부분 사람들은 일정 수준의 자기통제력이 있기 때문에 비행행동에 연루되지 않는다. 왜냐하면 개인적 억지력을 충분히 발휘하여 순간의 만족에 굴복하지 않을 수 있기 때문이다. 반대로 자기통제력이 부족한 개인은 충동적이고 모험을 좋아한다. 이는 비행행동에 연루될 가능성을 높인다. 이를 조사 맥락으로 가져온다면 자기통제력이 부족한 사람이 답변하는 자기보고식 응답 자료의 신뢰성과 타당성에 의문을 제기할 수 있다.

조사응답과 자기통제간의 직접적 관계가 있을 수 있기 때문에(Piquero et al., 2000; Watkins & Melde, 2007), 일반적으로 자기통제력이 낮은 개인은 자기보고식 설문에서 본인 행동에 대한 솔직한 응답을 하지 않거나 응답 자체를 하지 않을 수 있다. 설문조사 참여에 따른 인센티브가 낮음에도 불구하고 설문조사에 응하는 것은 자기통제를 통해 본인의 시간과 노력을 희생할 수 있기 때문이다. 따라서 자기통제력이 낮은 개인은 자기통제력이 높은 개인보다 설문조사에 참여할 가능성도 낮고 참여하더라도 항목마다 성실하게 답변하지 않을 가능성이 높다. 즉, 설문조사에 참여하기 위해서는 잠재적 응답자의 즉각적인 욕구를 희생해야 하므로, 자기통제력이 낮은 개인은 설문조사에 성실하게 참여하려는 결정을 하는 확률이 낮아진다. 자기통제력이 낮은 개인은 설문조사 문항에도 차등적으로 응답하고 설문 자체에도 차등적으로 참여할 것이라고 예상할 수 있다.

특히 항목무응답은 자기통제력이 떨어지는 대상에게서 더 빈번하게 발생할 수 있다. 적어도 1-2시간을 시험을 보듯 꾸준히 앉아서 응답할 수 있는 능력을 가져야 설문지에 결측 없이 응답할 수 있다. 반면 단위무응답은 자기통제보다는 권위에 대해 거부하는 비행성향과 관련될 수 있다. 하지만 주요 비행이론에서는 자기통제력이 낮은 개인이 비행행동을 할 가능성이 높은 것으로 설명하고 있기 때문에 단위무응답 또한 자기통제와 관련될 수 있다. 이렇게 자기통제는 부정확한 응답으로 차등적 타당도를 가져올 수 있고 설문을 거부하면서 차등적 응답률을 가져올 수 있다. 하지만 아직까지 이러한 항목무응답과 단위무응답에 대한 경험적인 연구가 절대적으로 부족한 실정이다. 따라서 본 연구는 자기통제와, 단위무응답과 항목무응답으로 조작화된 차등적 응답 간의 관련성을 살펴보고자 한다.

또한 조사항목별로 무응답을 많이 했던 개인은 단위무응답을 할 확률이 높을 수

있다(Loosveldt, Pickery & Billiet, 2002). 항목무응답은 다양한 이유에서 발생하는데, 대답하기 너무 어렵거나 조사에 무관심하거나 또는 너무 창피하고 위협적이기 때문에 대답하지 않을 수 있다. 따라서 항목무응답을 이해하기 위해서는 어떤 항목에 대한 무응답인지를 구분할 필요가 있다고 판단된다. 특히 다른 항목에 대한 무응답과 달리, 개인의 인구통계학적 특성과 같은 응답하기 쉬운 항목에 대한 무응답은 질문 자체가 어렵다기 보다 단지 자기통제가 낮은 개인 성향과 관련될 가능성이 높다.

이와 같이 자기통제력이 차등적인 설문응답 및 수집된 자료의 질적 차이와 관련된다는 것이 사실이라면, 낮은 자기통제력을 가진 개인을 체계적으로 배제하지 않도록 하는 조치가 조사계획과 자료수집과정에서 반드시 고려되어야 한다. 하지만 현재 결측치 처리방법(목록별 제거, 대응별 제거) 등은 낮은 자기통제력을 가진 개인을 더 쉽게 배제하여 설명될 수 있는 변량을 제한할 수 있다. 이러한 문제는 여러 회기를 조사하는 패널조사에서 더 악화될 수 있다. 대부분의 자기통제력이 낮은 개인이 회기가 거듭될수록 점점 표본에서 탈락하면서 설명 변량이 더 감소하게 된다. 따라서 이러한 내용을 반영할 수 있는 결측치 처리방법을 개발하기 위해서는 이러한 관계에 대한 확인 작업을 통해 기초자료를 만들 필요가 있다.

III. 연구방법

1. 자료 및 표본

본 연구는 서울아동패널(Seoul Panel Study of Children)의 자료를 이용하여 항목무응답 및 단위무응답과 관련 있는 변수를 분석하고자 하였다. 서울아동패널은 2004년 서울지역 내 11개 초등학교 4학년에 재학 중이었던 1,785명의 아동 및 부모를 매년 1회 추적 조사를 하고 있으며, 사회정서적 발달, 학업성취, 건강, 부모 및 가족과 관련된 조사 자료를 구축하고 있다(정익중, 김세원, 2011). 현재 분석에 이용 가능한 자료는 7차년도 조사까지 구축되어 있다. 7차년도 조사당시 응답자는 고등학교 1학년에 재학 중이었으며, 총 1,373명이 조사에 참여하였다. 모든 자료는 자기보고식 설문조사를 통하여 수집되었다.

서울아동패널과 같은 종단자료는 조사의 특성 상 표본유실과 결측 자료가 발생하게 된다. 만약 결측 자료가 연구에서 다루는 현상과 직접적으로 관련이 있다면, 결측이 무작위로 발생했음을 가정한 상태에서 결측치를 처리하는 방법은 문제가 될 수 있다 (Watkins & Melde, 2007). 예를 들어, 자기통제수준이 낮은 개인이 개별 문항에 응답하지 않을 가능성이 높거나 추후 조사에 참여하지 않을 가능성이 높다고 간주한다면, 자기통제력이 낮은 개인의 자료가 분석에서 체계적으로 제외되지 않았다고 확신하기 어려울 것이다.

이러한 문제를 다루는데 있어 학교를 기반으로 하여 수집된 아동 및 청소년 자료를 연구에서 활용하는 것은 장점이 있다. 이 연령대의 응답자를 대상으로 수집된 자료는 성인을 대상으로 하는 표본에 비해 타당도가 높고, 이 시기 청소년은 사회적 바람직성 편향에 의해 영향을 덜 받기 때문이다(Junger-Tas & Marshall, 1999; Watkins & Melde, 2007에서 재인용). 따라서 아동기 및 청소년기에 수집된 자료는 항목무응답과 단위무응답을 설명하는데 있어 좋은 정보를 제공할 수 있을 것이다.

본 연구에서는 서울아동패널 1차년도 조사에 참여한 1,785명 아동의 자료를 활용하여 항목무응답을 설명하였다. 그리고 단위무응답 분석에서는 1차년도 조사 참여자의 2차-7차년도 추적조사에 대한 참여 여부를 활용하였다.

2. 변수

1) 종속변수

첫째, 항목무응답 분석에서 종속변수는 서울아동패널 1차년도 응답자가 전체 조사 문항 가운데 응답하지 않은 문항의 수로 계산하였다. 척도의 성격에 따라 하나의 질문은 여러 개의 하위척도로 구성된 경우도 있었으며, 이런 경우 하위척도를 구성하는 각각의 문항을 개별적인 조사 문항으로 간주하였다. 1차년도 총 조사문항 수는 157개였으며, 이 가운데 응답하지 않은 문항의 수가 최종분석에 사용되었다. 1차년도 157개 조사문항별로 응답한 경우는 '0', 응답하지 않은 경우는 '1'로 코딩하였으며, 이들 문항을 모두 합산하여 1차년도 항목무응답 수를 계산하였다. 단, 질문 문항에 '해당 없음'인 경우는 무응답으로 간주하지 않았다. 예를 들어, 지난 1년간 전학횟수를 묻는

질문은 전학한 경험이 있는 응답자에게만 해당된다. 따라서 전학경험이 없어 전학횟수에 응답하지 않은 경우는 항목무응답 횟수에 포함하지 않았다. 이와 같이 계산된 항목무응답의 최소값은 0, 최대값은 69였다.

둘째, 단위무응답 분석에서 종속변수는 서울아동패널 1차년도 조사자의 2차-7차년도 추적조사에 대한 무응답 횟수이다. 항목무응답과 달리 단위무응답은 해당년도의 조사 자체에 참여하지 않은 경우를 의미한다. 구체적인 계산방법은 다음과 같다. 1차 조사 이후 6개 회기별로 조사에 참여한 경우는 '0', 조사에 참여하지 않은 경우는 '1'로 코딩하였으며, 6개 회기별 무응답여부를 모두 합하여 단위무응답 횟수를 계산하였다. 한편, 6번의 추후조사에 모두 참여한 응답자는 1,165명(65.3%), 추후조사 가운데 1번 참여하지 않은 응답자가 218명(12.2%), 2번-5번 참여하지 않은 응답자는 332명(18.6%)이었다. 그리고 1차년도 조사 이후 추후 조사에 한 번도 참여하지 않은 응답자가 70명(3.9%)이었다.

2) 독립변수

본 연구에서 항목무응답과 단위무응답 정도를 설명하기 위해 투입된 독립변수의 정의는 다음과 같다. 모든 독립변수는 1차년도 조사에 응답한 값을 기준으로 계산하였다.

첫째, 아동의 성별은 더미변수로 남자=1, 여자=0으로 구분하였다.

둘째, 한부모가구 여부는 더미변수로, 아동의 부 또는 모만 있는 가구=1, 부모가 모두 있는 가구=0으로 구분하였다.

셋째, 학교폭력피해경험은 지난 1년 동안 학교에서 '다른 아이들이 듣기 싫은 별명을 부르며 나를 놀리거나 조롱한 적이 있다', '다른 아이들이 나를 고의로 어떤 일에 끼워주지 않거나 따돌리고 무시한 적이 있다', '다른 아이들이 나를 손이나 주먹으로 때리거나 발로 걷어찬 적이 있다' 등의 6개 문항에 대한 경험정도로 측정하였다. 각 문항은 4점 리커트 척도로 측정되었으며, '전혀 없다'에서부터 '4번 이상'까지 구분되어 있다. 각 문항의 점수를 모두 합한 총점이 높을수록 피해경험이 많은 것을 의미한다.

넷째, 아동의 비행정도는 지난 한 달 동안 '남의 돈이나 물건을 슬쩍 훔친 적이 있다', '다른 사람을 심하게 때린 적이 있다', '남의 돈이나 물건을 뺏은(뺨 뜯은) 적이 있다'로 측정하였다(있다=1, 없다=0). 3개 문항의 총점이 높을수록 비행 수준은 높은

것을 의미한다.

다섯째, 아동의 유능감은 ‘나는 재주가 많다고 생각한다’, ‘나는 다른 사람들보다 의지가 강하다’, ‘나는 처음에 못할지라도 잘 할 때까지 열심히 한다’의 3개 문항으로 측정하였다. 각 문항의 응답범주는 ‘전혀 그렇지 않다’에서 ‘매우 그렇다’로 구분된 4점 리커트 척도이며, 총점이 높을수록 유능감이 높은 것을 의미한다.

여섯째, 아동의 학업성취수준은 전문가가 출제한 국어와 수학시험문제로 측정하였다. 시험은 약 30분간 진행되었으며, 과목별 만점은 각각 15점이었다. 최종분석에서는 두 과목의 점수를 모두 합하여 사용하였다.

일곱째, 아동의 낮은 자기통제는 지난 6개월 동안 ‘집중력이 없고 어떤 일에 오래 주의를 기울이지 못한다’, ‘가만히 앉아 있지 못하고 안절부절못하며 지나치게 많이 움직인다’, ‘충동적이고 생각없이 행동한다’ 등의 문항으로 측정하였다. 각 문항의 응답범주는 3점 리커트 척도로 ‘전혀 아니다’, ‘그런 편이다’, ‘자주 그렇다’로 구분되어 있으며, 총점이 높을수록 자기통제수준이 낮은 것을 의미한다.

마지막으로 항목무응답을 단위무응답에 대한 예측변수로 투입한 분석모형에서는 조사문항을 표 1과 같이 구분하여 각 영역별 항목무응답 여부(무응답 문항이 1개 이상 있다=1, 없다=0)를 사용하였다. 각 영역별 조사문항의 내용 및 수는 다음과 같다.

표 1

설문문항의 각 영역별 내용 및 문항 수

영역	관련 문항 내용	문항 수
학교 및 학업	학교의 환경 및 유대감, 학교폭력, 학업성적	34
아동청소년 행동평가척도	우울불안, 주의집중, 비행, 공격성	62
심리적 특성	자존감, 또래관계	21
부모 및 가족	부모와의 유대감, 성인역 아동부담, 아동학대	26
개인적 특성	성별, 생년월일, 가족관계, 건강상태, 전학 등	14
	합 계	157

3. 분석방법

본 연구에서는 항목무응답과 단위무응답을 설명하는 변수들의 영향력을 살펴보고, 항목무응답과 단위무응답 간의 관계를 설명하고자 하였다. 여기에서 무응답 횟수와 같은 가산자료(count data)는 비음정수(non-negative integer)이기 때문에, 흔히 적용되는 선형회귀분석을 이용하여 모형화를 해서는 안 되며 특별히 고안된 가산자료모형을 이용해야 한다(Lee, Stevenson, Wang & Yau, 2002; Lord, Washington & Ivan, 2005; 유승훈, 양창영, 2005, 재인용).

일반적으로 가산자료모형에는 포아송 모형(Poisson model)을 많이 사용한다. 가산 변수 u_i 는 수식 (1)과 같은 포아송 분포를 갖는다(Muthen & Muthen, 2009). 이 때 λ 는 사건이 발생할 확률로 사건의 평균과 분산을 의미하며, $\log \lambda$ 에 대한 회귀식은 수식 (2)와 같다.

$$P(u_i = r) = \frac{\lambda_i^r e^{-\lambda_i}}{r!}, \quad \text{where } u_i = 0, 1, 2, \dots \quad (1)$$

$$\log \lambda_i = \ln \lambda_i = \beta_0 + \beta_1 x_i \quad (2)$$

포아송 모형은 다음과 같이 자료의 평균과 분산이 같다고 가정한다.

$$E(u_i | x_i) = \lambda_i \quad (3)$$

$$V(u_i | x_i) = \lambda_i \quad (4)$$

본 연구에서 항목무응답의 횟수의 범위는 0-69, 평균은 0.68, 표준편차는 2.75, 단위무응답의 값의 범위는 0-6, 평균과 표준편차는 각각 1.01과 1.74였다. 이와 같이 본 연구에서 항목무응답과 단위무응답 횟수 모두 평균보다 표준편차가 더 큰 것으로 나타났다. 과대산포(overdispersion)는 응답치가 '0'인 경우가 많거나 자료 내의 관찰되지 않은 이질성에서 비롯되는데(Sheu, Hu, Keeler, Ong & Sung, 2004), 이와 같이 자료에 과대산포 문제가 있을 경우 포아송 모형은 올바른 추정치를 제공할 수 없다.

본 연구에서와 같이 분산이 평균보다 큰 자료의 경우 포아송 모형보다는 음이항

모형(negative binomial model)이 더 적합하다. 음이항 모형은 가산자료모형에서 분산이 평균을 초과하는 것을 허용하며, 과대산포와 관련된 문제를 다룰 수 있다(Long, 1997). 음이항 회귀모형은 수식 (2) 포아송 모형에 관찰되지 않은 이질성인 ε_i 를 포함하며, 수식 (5)와 같이 나타낼 수 있다(Muthen & Muthen, 2009).

$$\ln \lambda_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{where } \exp(\varepsilon) \sim \Gamma \quad (5)$$

음이항 회귀모형의 평균과 분산은 다음과 같이 가정된다. $\alpha > 0$ 인 경우, 분산이 평균보다 커지게 되어 과대산포 문제를 다룰 수 있다.

$$E(u_i | x_i) = \lambda_i \quad (6)$$

$$V(u_i | x_i) = \lambda_i(1 + \lambda_i \alpha) \quad (7)$$

과대산포 여부는 포아송 모형과 음이항 모형의 우도비 검정통계량(likelihood ratio test = $-2\{\text{LL(Poisson)} - \text{LL(negative binomial)}\}$, 영가설 기각: $\text{LR} > \chi^2(1-2\alpha, 1 \text{ df})$)을 이용하여 판단할 수 있다. 이 검정은 분산이 평균보다 큰 것을 허용하는 대안적 방법과 대조적으로 포아송 분포에서 요구되는 평균과 분산의 동일성을 비교·검증한다(Computer Center, 2012). 본 연구에서는 LR 검증 결과 항목무응답 횟수와 단위무응답 횟수 모두 분산이 평균보다 유의미하게 큰 것으로 나타났다. 최종 분석에서는 항목무응답과 단위무응답을 설명하기 위해 Mplus 5.21을 이용하여 음이항 회귀분석을 실시하였다.

IV. 연구결과

1. 주요변수의 기술통계

본 연구에서 사용된 주요변수에 대한 기술통계 분석결과는 표 2와 같다. 항목무응답의 평균 횟수는 0.68, 단위무응답의 평균 횟수는 1.01이었으며, 분석대상자 가운데

남자와 여자의 비율은 각각 절반 정도였으며, 한부모가구인 경우가 전체의 7.3%였다. 학교폭력피해경험의 평균은 9.08, 비행행동의 평균은 0.39, 유능감의 평균은 8.94, 학업성취의 평균은 19.17, 낮은 자기통제의 평균은 4.58로 나타났다. 1차 조사 문항 영역별로 한 개 문항 이상에 응답을 하지 않았는지 여부를 살펴본 결과, 여러 영역 가운데 아동청소년 행동평가척도에 대한 무응답 비율이 10.2%로 가장 높았다.

표 2

주요변수의 기술통계(N=1,785)

변 수	평균	표준편차	최소값	최대값
항목무응답	0.68	2.75	0	69
단위무응답	1.01	1.74	0	6
성별(%)				
남자	52.0			
여자	48.0			
한부모(%)				
예	7.3			
아니오	92.0			
학교폭력피해경험	9.08	3.30	1.00	24.00
비행행동	0.39	0.81	0.00	3.00
유능감	8.94	2.00	1.00	12.00
학업성취	19.17	4.48	3.00	30.00
낮은 자기통제	4.58	1.45	1.00	9.00
영역별 항목무응답 여부(%)				
학교 및 학업	9.8			
아동청소년 행동평가척도	10.2			
심리적 특성	4.9			
부모 및 가족	4.8			
개인적 특성	7.8			

한편, 추적기간 동안 발생한 단위무응답의 총 횟수와 각 조사연도별 단위무응답의 비율은 표 3과 같다. 먼저, 추적기간 동안 모든 조사에 응답한 사례의 비율은 65.3%였으며, 6번의 추적조사 기간 동안 조사에 1번 응답하지 않은 비율이 12.2%로 가장

많았으며, 2번에서 5번까지는 약 4% 정도로 모두 비슷하게 나타났다. 1차 조사 이후 한 번도 추적조사에 참여하지 않은 경우는 3.9%였다.

연도별로 조사에 참여하지 않은 단위무응답의 비율은 2차년도 조사에서 6.2%였으며, 이후 연도별 조사에서 단위무응답 발생 비율은 점차 증가하여 7차년도 조사에서는 전체 사례의 23.1%가 조사에 응답하지 않은 것으로 나타났다.

표 3
단위무응답 발생 빈도(N=1,785)

	구 분	사례 수	백분율(%)
추적 기간 내 총 횟수	0번	1,165	65.3
	1번	218	12.2
	2번	84	4.7
	3번	86	4.8
	4번	77	4.3
	5번	85	4.8
	6번	70	3.9
연도별 단위 무응답 빈도	1차년도	0	0.0
	2차년도	111	6.2
	3차년도	239	13.4
	4차년도	318	17.8
	5차년도	345	19.3
	6차년도	372	20.8
	7차년도	412	23.1

표 3과 같이 추적기간 내 발생한 총 단위무응답 발생 횟수의 분포와 연도별 단위무응답 발생 빈도의 차이는 무응답의 발생 패턴에서 비롯된다. 서울아동패널에서 대부분의 단위무응답 사례는 그림 2와 같이 단조형태(monotone pattern)가 아니라 일반 형태(general pattern)로 발생했기 때문이다.

t1	t2	t3	t4	t5	t6	t7

그림 1. 단조형태 무응답

t1	t2	t3	t4	t5	t6	t7

그림 2. 일반형태 무응답

2. 항목무응답 예측요인

본 연구에서는 항목무응답의 수는 서울아동패널 1차년도 조사 문항 가운데 응답하지 않은 문항의 수로 계산하였다. 항목무응답에 대한 독립변수의 영향력을 살펴보기 위하여 음이항 회귀모형을 사용하였으며, 분석결과는 다음과 같다.

먼저, 카이제곱검증결과 모형이 통계적으로 유의하였으며, 이는 상수항만 있는 모형보다 독립변수를 투입한 모형의 설명력이 유의하게 높은 것을 의미한다. 독립변수 가운데 아동의 성별이 남자인 경우 자기보고식 설문조사에서 항목무응답의 횟수가 더 많은 것으로 나타났다. 그리고 아동의 유능감과 학업성취수준이 높을수록 항목무응답 횟수는 유의하게 낮았으며, 자기통제 수준이 낮을수록 항목무응답 횟수는 유의하게 높았다. 즉, 자신을 유능감이 있는 존재로 인식하고, 학업수준이 높은 아동은 전체 설문문항에 모두 응답할 가능성이 높은 것을 알 수 있었다. 그러나 자기통제 수준이 낮은 응답자는 그렇지 않은 응답자에 비해 전체 설문문항에 대한 응답을 완료하는데 상대적으로 시간과 노력을 덜 할애하는 것으로 볼 수 있으며, 이는 선행연구 결과와 일관되는 것이다(Wakins & Melde, 2007).

한편, 표 4에서와 같이 본 연구에서는 dispersion 값이 99% 신뢰수준에서 통계적으로 유의하여 음이항 모형이 자료를 잘 설명하는 것으로 볼 수 있었다. 만약 dispersion 값이 0과 유의미하게 다르지 않다면, 음이항 모형 대신 포아송 모형이 자료에 더 적합하다는 것을 의미한다.

표 4

항목무응답 음이항 회귀분석결과

변 수	Estimate	S.E.
아동성별(남=1)	0.366	0.111**
한부모가구	0.152	0.184
학교폭력피해경험	0.039	0.028
중비행	0.031	0.062
유능감	-0.097	0.029**
학업성취수준	-0.089	0.012**
낮은 자기통제	0.051	0.022*
상수항	1.071	0.540*
Dispersion parameter(a)	2.043	0.267***
Log Likelihood	-1709.983	
LR χ^2	367.422***	

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

3. 단위무응답 예측요인

본 연구에서 단위무응답 횡수는 서울아동패널 1차년도 조사에 참여한 아동이 총 6 번의 추적조사에 참여하지 않은 횡수를 의미한다. 항목무응답과 마찬가지로 단위무응답 횡수는 가산자료이며, 분산이 평균보다 크기 때문에 음이항 회귀분석방법을 사용하였고, 분석결과는 표 5와 같다.

여러 변수 가운데 학업성취수준만이 단위무응답횡수에 유의한 영향을 미치는 것으로 나타났다. 즉, 1차 조사에서 측정한 학업성취수준이 낮은 아동은 이후 추적조사에 응답하지 않을 가능성이 더 높았다. 이러한 분석결과는 학업성취수준이 높은 아동일수록 패널에 좀 더 오래 남아 있을 수 있음을 보여주는 것이다.

그러나 항목무응답의 분석결과와는 달리 유능감이나 자기통제수준은 단위무응답에 영향을 미치지 않았다. 선행연구(Wakins & Melde, 2007)에서도 다른 독립변수를 투

입하지 않았을 때는 낮은 자기통제수준이 단위무응답에 유의한 영향을 미쳤으나, 독립변수를 모두 통제한 상태에서는 자기통제수준이 더 이상 유의한 영향을 미치지 않는 것으로 나타났다.

표 5

단위무응답 음이항 회귀분석결과

변 수	Estimate	S.E.
아동성별(남=1)	-0.128	0.085
한부모	0.140	0.150
학교폭력피해경험	-0.005	0.013
중비행	0.023	0.053
유능감	0.012	0.021
학업성취수준	-0.018	0.009*
자기통제	-0.006	0.025
상수항	0.344	0.315
Dispersion parameter(a)	3.303	0.189***
Log Likelihood	-2284.264	
LR χ^2	99.336***	

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

4. 항목무응답과 단위무응답 간의 관계

표 6은 항목무응답과 단위무응답 간의 관계를 파악하기 위하여 단위무응답을 종속 변수로, 항목무응답 여부를 독립변수로 투입하여 음이항 회귀분석을 한 결과이다. 1차년도 조사에서의 항목무응답 여부는 이후 추적조사에서의 단위무응답에 유의한 영향을 미쳤다. 즉, 1차년도 조사에서 한 가지 이상의 설문 문항에 응답을 하지 않았을 경우, 이후 추적조사에서 설문조사에 응답하지 않을 가능성이 더 높았다.

표 6

항목무응답과 단위무응답 간의 관계

변 수	Estimate	S.E.
1차년도 항목무응답 여부	0.047	0.021*
상수항	-0.032	0.044
Dispersion parameter(a)	3.276	0.185***
Log Likelihood	-2331.032	
LR χ^2	5.800*	

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

표 7은 1차년도 조사문항을 영역별로 구분하여, 각 영역별 항목무응답과 단위무응답 간의 관계를 음이항 회귀분석으로 설명한 결과이다. 1차 조사 설문지의 여러 영역 가운데 아동의 생년월일, 가족관계, 건강상태, 전학 등과 같은 개인적 특성에 대한 문항에 응답을 하지 않은 아동은 이후 추적조사에 참여하지 않을 가능성이 유의하게 높았다. 다른 질문영역과 달리 개인적 특성 영역은 응답할 의향만 있다면 쉽게 응답할 수 있으며, 폐쇄형 질문보다 개방형 질문을 주로 사용하였다는 차이점이 있다. 이와 같이 자신의 인적사항과 관련된 문항에 성실하게 응답하지 않은 아동은 이후 패널조사에 참여하지 않을 가능성이 높다는 것을 알 수 있었다.

표 7

영역별 항목무응답과 단위무응답 간의 관계

변 수	Estimate	S.E.
학교 및 학업 영역 무응답 여부	0.200	0.130
아동청소년 행동평가척도 영역 무응답 여부	0.142	0.134
심리적 특성 영역 무응답 여부	0.052	0.188
부모 및 가족 영역 무응답 여부	0.124	0.179
개인적 특성 영역 무응답 여부	0.382	0.124**
상수항	-0.078	0.048
Dispersion parameter(a)	3.251	0.185***
Log Likelihood	-2328.984	
LR χ^2	9.896†	

†p<.10, **p<.01, ***p<.001

V. 결 론

자기보고를 통한 패널조사는 사회과학에서 가장 신뢰성 높은 자료수집방법으로 사용되고 있다. 그러나 패널조사는 항목무응답 뿐만 아니라 단위무응답으로 인해 자료에 문제가 발생할 수 있다. 자기통제력이 낮은 개인은 자기보고식 조사를 완료하는데 필요한 집중력이 부족하고, 이로 인해 만족스럽지 않은 경험인 설문응답을 가능한 빨리 종료하기 위해 모든 문항에 일일이 응답하지 않을 수 있다. 또한 자기통제력이 낮은 개인은 조사 참여 자체를 거부할 수도 있다.

본 연구는 서울아동패널 자료를 사용하여 항목무응답과 단위무응답에 영향을 주는 요인과, 항목무응답과 단위무응답 간의 관계에 대하여 설명하였다. 분석결과, 아동의 성별, 유능감, 학업성취수준, 낮은 자기통제 정도는 항목무응답에 유의한 영향을 미치는 것으로 나타났다. 그동안 자기통제수준이 낮은 응답자는 자기통제수준이 높은 응답자에 비해 응답을 부정확하게 하는 경향이 있으며(Hindelang et al., 1981) 자기통제는 응답에서의 차이와 관련이 있을 것이라는 주장이 있었다(Keane, Maxim & Teevan, 1993). 그러나 자기통제와 응답 간의 관계를 직접적으로 검증한 연구는 상대적으로 부족한 상황이었다. 본 연구의 결과는 낮은 자기통제수준과 무응답 간의 유의한 관계에 대한 실증적 근거를 제공한 것으로 볼 수 있다.

단위무응답의 경우 학업성취수준이 낮을수록 추적조사에 응답하지 않는 횟수가 유의하게 증가하였다. 그러나 학업성취수준 이외의 변수는 단위무응답에 유의한 영향을 미치지 않았으며, Watkins와 Melde(2007)의 연구에서도 낮은 자기통제수준은 항목무응답에는 영향을 미쳤지만 단위무응답에는 유의한 영향을 미치지 않는 것으로 나타났다. 한편, 횡단연구나 종단연구에서 발생한 결측치를 목록별 제거(listwise deletion) 방법으로 처리했을 경우, 상대적으로 긍정적인 발달지표를 보이는 아동만이 분석대상에 포함될 가능성이 높다. 결측치를 포함한 사례를 제외하는 것은 단순히 표본의 크기를 감소시키는 것이 아니라 표본의 대표성과 관련된 문제인 것이다. 따라서 연구자는 결측치를 어떻게 처리할 것인가에 대해 충분히 고민할 필요가 있다.

또한 항목무응답은 단위무응답을 예측하는 것으로 나타났는데, 조사문항의 성격에 따라 항목무응답이 단위무응답에 미치는 영향은 다소 차이를 보였다. 다른 조사문항과 달리 1차년도 조사에서 응답자의 생년월일이나 전학, 가족관계 등과 같이 인적사

항에 대한 질문에 응답하지 않은 경우만이 이후 추적조사 자체에 참여하지 않는 경우가 더 많았다. 항목무응답을 감소시키기 위해서는 특정 주제의 포함여부, 질문유형, 조사방식 등에 좀 더 민감할 필요가 있을 것이다.

지금까지 항목무응답과 단위무응답 간의 관계를 설명한 연구는 거의 없었으며, 단위무응답이 항목무응답에 비해 조사연구의 질에 더 부정적인 영향을 미치는 것으로 간주되었고 두 가지 무응답 형태는 자료의 질과 관련되어 있기는 하나 각기 다른 원인과 해결책을 가진 서로 구분되는 별개의 문제로 생각되었다(Yan & Curtin, 2010). 본 연구의 결과에서는 항목무응답과 단위무응답이 유의한 관계가 있음을 실증적으로 설명했다는 점에서 그 의미를 찾을 수 있을 것이며, 자료의 질을 향상시키기 위해서는 단위무응답뿐만 아니라 항목무응답을 감소시키기 위한 노력이 동시에 있어야 할 것이다.

이번 연구는 크게 두 가지 함의를 가지고 있다. 첫째, 자기보고식 패널조사를 활용한 연구를 진행할 때 결측치를 처리함에 있어 항목무응답의 체계적인 오류를 점검할 것을 요구한다. 목록별 제거를 통해 결측치를 처리하는 방법은 자기통제력이 낮은 개인을 훨씬 더 많이 배제하게 되면서 설명변량이 작아지고 유의한 결과를 확인하지 못할 가능성이 높다. 특히 항목무응답이 단위무응답을 예측할 수 있다는 것은 이후 패널조사에서 표본유실을 미리 예방할 수 있는 준비를 할 수 있다는 것을 보여준다. 특히 인구통계학적 특성에 대한 항목무응답이 있을 경우, 다음 회기조사에서 패널 유지와 관리에 좀 더 주의하여 조사를 진행해야 할 것이다. 둘째, 자기통제력이 낮은 개인을 참여시키기 위한 노력을 강화해야 한다. 특히 부모나 교사를 활용하여 조사에 대한 지속적인 참여를 유도하고, 이들을 대상으로 하여 자료를 수집하는 다원측정(triangulation)과 같은 방법은 자기보고 조사의 한계를 일부 극복할 수 있도록 도와줄 것이다.

패널조사의 계획과 관리가 어려운 이유는 항상 장기적으로 설정된 목표와 단기간에 동원 가능한 자원 사이에 격차가 존재하기 때문이다. 따라서 조사 초기부터 장기적인 목표와 자원을 충분히 고려한 상태에서 진행하지 않으면 패널조사는 지속되기 어렵다(이지연, 2005). 패널조사는 장기 반복조사의 특성으로 인해 반드시 조사의 질 확보를 위한 충분한 연구예산 및 패널조사에 전문성을 가진 연구인력의 안정적 운영이 요구된다(이희길, 2008). 국가차원에서 충분한 예산이나 인력이 지원되지 않는다면 표집대상 지역을 제한하는 것이 표본의 추적과 유지에 효과적이다(Thornberry & Krohn, 2002). 미국의 대표적인 패널연구들도 대부분 특정 지역중심인 경우가 많다. 따라서 서울아

동패널이 표집대상 지역을 확대하기보다 한정짓는 방법을 사용한 것은 바람직한 전략이었다고 판단된다.

다년도 추적조사에서는 조사대상의 표본유실을 최소화하는 것이 연구의 타당도를 유지하기 위해 매우 중요하다(Ribisl et al., 1996). 물론 결측치의 문제를 최소화하여 처리하는 다중대체(multiple imputation)방법이 있지만 가능한 항목무응답과 단위무응답 모두를 감소시키려는 노력을 해야 한다. 거의 대부분 단위무응답을 감소시키기 위한 방법만을 고려하는 경향이 있는데, 본 연구에서 살펴본 것처럼 항목무응답이 단위무응답에 영향을 미치기 때문에 항목무응답 패턴을 파악하고 이를 통해 밝혀진 단위무응답의 확률이 높은 조사대상의 경우 특별 관리하도록 해야 할 것이다. 단위무응답을 줄이기 위해서는 일반적인 표집방식인 가구단위 표집을 이용한 추적조사보다는 센터나 학교를 통한 군집표집이 추적에 용이하며, 조사대상의 유실률도 훨씬 낮은 것으로 보고되고 있다(Hawkins, Catalano, Kosterman, Abbott & Hill, 1999). 기존 추적조사연구에서 효과적인 것으로 간주되는 다양한 표본추적 및 유지기법을 사용하는 것도 단위무응답을 감소하기 위해 바람직하다(Ribisl et al., 1996). 서울아동패널도 학교를 통한 표집을 사용했으며 교육청과 같은 공공기관의 협조, 입학·졸업 축하편지 및 선물, 뉴스레터 활용, 홈페이지나 자원봉사 확인증 등과 같이 대상자와 지속적으로 연락을 유지하여 소속감 부여하고 적절한 인센티브를 제공하는 등 다양한 방법으로 패널유지와 관리에 노력하고 있다. 또한 조사원과 조사대상자간의 좋은 신뢰관계는 응답률을 높이는데 매우 중요하므로 조사원을 장기간 안정적으로 활용하는 것이 중요하다. 다양한 방법을 통해 조사의 질을 높이는 것이 연구의 질을 높이는 가장 쉽고 바람직한 방법이라는 것을 다시 한 번 명심해야 할 것이다.

본 연구는 항목무응답과 단위무응답에 영향을 미치는 요인을 파악하였으나 차후 연구에서는 이러한 항목무응답과 단위무응답이 결과에 미치는 편의(bias)정도나 패널조건화 정도와 그것이 결과에 영향을 미치는 편의정도에 대한 추정 등 다양한 형태의 패널조사 자체에 대한 연구들이 다각적으로 진행되어야 할 것이다.

참고문헌

- 금재호 (1998). **노동관련 패널조사의 의의와 필요성**. 서울: 한국노동연구원.
- 유승훈, 양창영 (2005). 가산자료모형을 이용한 해양오염사고 발생횟수의 분석. **해양정책연구**, 20(2), 33-56.
- 이경상 (2007). **국내 패널조사의 현황 및 시사점**. 서울: 교육인적자원부 인적자원통계협력망.
- 이상호 (2006). 패널연구: 서구 주요 패널의 조사전략 변화와 시사점. **노동리뷰**, 16, 63-78.
- 이지연 (2005). **가구조사자료의 종단화 방안**(통계연구보고서 05-03). 대전: 통계청.
- 이희길 (2008). 국내 패널조사의 현황 분석. 이희길 외 (편), **통계조사의 정확성과 효율성 제고방안: 현장조사 및 표본추출을 중심으로**. 대전: 통계개발원.
- 정익중, 김세원 (2011). 아동·청소년발달의 위험요인 및 보호요인 잠재적 유형 간 관계. **한국청소년연구**, 22(3), 197-230.
- 정진성, 박현호 (2010). 지역사회의 구조적 특성이 살인범죄에 미치는 영향: 전국 시군구를 대상으로 한 음이향 회귀분석. **형사정책연구**, 81, 91-119.
- Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences* (6th ed.). New York: Worth Publishers.
- Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatric and Adolescent Medicine*, 153, 226-234.
- Hindelang, M. J., Hirschi, T., & Weis, J. G. (1981). *Measuring delinquency*. Beverly Hills, CA: Sage.
- Hirschi, T., & Gottfredson, M. R. (1993). Commentary: Testing the general theory of crime. *Journal of Research in Crime and Delinquency*, 30, 47-54.
- Junger-Tas, J., & Marshall, I. H. (1999). The self-report methodology in crime research. In M. Tonry (Ed.), *Crime and justice: A review of research* (pp.

- 291-367). Chicago: University of Chicago Press.
- Keane, C., Maxim, P. S., & Teevan, J. J. (1993). Drinking and driving, self-control, and gender: Testing a general theory of crime. *Journal of Research in Crime and Delinquency*, 30, 30-46.
- Lee, A. H., Stevenson, M. R., Wang, K., & Yau, K. K. W. (2002). Modeling young driver motor vehicle crashes: Data with extra zeros. *Accident Analysis and Prevention*, 34, 515-521.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item nonresponse as a predictor of unit nonresponse in a panel survey. *Journal of Official Statistics*, 18(4), 545-557.
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis and Prevention*, 37, 35-46.
- Muthen, L. K., & Muthen, B. (2009). Regression analysis, exploratory factor analysis, confirmatory factor analysis, and structural equation modeling for categorical, censored, and count outcomes. *Mplus short courses topic 2*. Retrieved July 30, 2012, from <http://www.statmodel.com/download/Topic%202-v14.pdf>.
- Piquero, A. R., MacIntosh, R., & Hickman, M. (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. *Criminology*, 38, 897-929.
- Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson II, W. S., & Bootsmiller, B. J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning*, 19, 1-25.
- Rodríguez, G. (2007). Lecture notes on generalized linear models. Available at <http://data.princeton.edu/wws509/notes/>

- Sheu, M. L., Hu, T. W., Keeler, T. E., Ong, M., & Sung, H. Y. (2004). The effect of a major cigarette price change on smoking behavior in California: A zero-inflated negative binomial model. *Health Economics*, 13(8), 781-791.
- Thornberry, T. P., & Krohn, M. D. (2002). *Taking stock of delinquency: An overview of findings from contemporary longitudinal studies*. New York: Kluwer Academic/Plenum Publishers.
- UKIT. (2012). Negative binomial: Testing for overdispersion in Poisson regression. Retrieved May 20, 2012, from http://www.uky.edu/ComputingCenter/SSTARS/www/documentation/P_NB_3.htm.
- 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.
- Yan, T., & Curtin, R. (2010). The relation between unit nonresponse and item nonresponse: A response continuum perspective. *International Journal of Public Opinion Research*, 22(4), 535-551.

ABSTRACT

Factors influencing item nonresponse and unit nonresponse in a youth panel survey

Kim, Se-Won* · Chung, Ick-Joong**

In recent years, panel surveys have become a popular tool in the social sciences. However, the problem of nonresponse gives rise to a number of issues. Nonresponse can create bias in the survey findings if appropriate adjustments are not made. The purpose of this study is to identify the factors influencing item nonresponse and unit nonresponse, and to examine the relationship between item nonresponse and unit nonresponse. This study makes use of data from the Seoul Panel Study of Children(SPSC). The results of negative binomial regression analyses indicated that gender, competence, academic achievement, low self-control were significant predictors of item nonresponse in the 1st year survey. However, only academic achievement was found to be a significant predictor of unit nonresponse during six-year follow-ups. In addition, item nonresponse helped predict unit nonresponse. Moreover, the effects on unit nonresponse were different, depending on the nature of item nonresponse. Item nonresponse that failed to answer questions regarding personal information was found to be related to the degree of unit nonresponse of subsequent follow-up surveys. In order to improve the quality of data obtained through Panel surveys, greater efforts need to be made to simultaneously reduce item nonresponse as well as unit nonresponse.

Key Words : panel survey, item nonresponse, unit nonresponse, self-control

투고일 : 6월 11일, 심사일 : 7월 26일, 심사완료일 : 8월 10일

* Senior Researcher, The Center for Social Sciences, Seoul National University
(sewonkim1@naver.com)

** Associate Professor, Graduate School of Social Welfare, Ewha Womans University
(ichung@ewha.ac.kr)