

# 2005년 인구주택총조사의 범위오차평가와 지역별 차이 연구

2006 12

: 년 월

연구자 : ~~이희용, 이현~~ 지연

## 主 要 內 容

- 본 연구의 목적은 2005년 총조사의 범위오차를 인구분석방법, 센서스 생존비 일관성검사, 사후조사 분석을 통해 종합적으로 평가하고, 향후 오차의 측정과 개선을 위한 시사점을 도출하는 것임
- 2005년 총조사의 오차율은 2000년에 비해 전국적인 수준에서 순 누락은 크게 감소했지만, 남자의 누락율이 여자보다 배 이상 높아지면서 총조사 전체인구의 성비역전 현상이 발생
- 연령별로는 65세 이상 고령자층의 범위오차가 현저히 향상되었으나, 지리적 이동성이 높은 젊은 연령층과 4세 이하의 인구에서는 만성적인 누락이 발생
- ~~측정방식에 따라 범위오차율이 상이하게 나타나고 있는 20-30대 인구의 실제규모는 향후 보다 종합적이고 면밀한 측정이 요구됨~~

## I. 서 론

센서스의 목표는 총인구 수준에서는 순누락<sup>1)</sup>을 줄이는 한편, 조사에서 만 성적으로 누락되기 쉬운 집단(예, 지리적 이동성이 큰 20대와 남자)과 다른 집단과의 계층간 누락을 차이를 줄이고 지역별 격차를 감소시키는 일이다. 2000년 총조사 당시의 인구를 인구균형방정식으로 추정된 결과 총조사의 범 위오차율은 -2.1%로 약 99만 명이 과소집계된 것으로 나타났다. 남녀간의 차 이는 0.58%로 그 차이가 적었지만 여자보다는 남자의 누락이, 노인층보다는 젊은층의 누락이 상대적으로 높았다(통계청, 2002b). 반면에 2005년 총조사의 범 위오차율은 -1.87%로 크게 감소했지만, 남자와 여자의 누락을 차이는 1.1% 로 남자가 여자보다 배 이상 높았다(통계청, 2006d).

그러나, 2005년 사후조사 결과를 보면 총조사의 순누락율은 0.9%로 2000 년 1.56%에 비해 감소한 것과 여자보다 남자의 누락이 많이 발생한 점은 일 치하지만 남녀 간에 차이는 그리 크지 않았다(통계청, 2006c). 이와 같은 결 과는 측정방법의 차이에서 오는 것으로, 한 가지 평가방식만으로 센서스 오 차의 규모와 수준을 완전히 파악하기란 어려운 일이다.

센서스 범위를 다각도로 평가해 보아야 할 필요성을 보여주는 사례는 미 국의 2000년 센서스에서 찾아 볼 수 있다. 2000년 센서스는 우편, 방문, 컴퓨 터, 전화조사 등 다양한 조사방법을 적극적으로 동원하여 30년 만에 처음으로 응답률이 향상되면서, 미국의 센서스 역사상 성공적인 조사로 평가되고 있다(Dimitri and Treat, 2000). 그러나, 센서스 범위를 평가하기 위해 사용하 는 주된 자료원인 사후조사와 인구분석방법에 의한 결과가 서로 일치하지 않는 문제점이 발생했다. 2001년 3월에 발표된 인구분석방법 결과에 따르면

1) 범위오차(coverage error)는 조사대상(거처, 가구, 가구원)이 과소집계(undercount)되거나 과대집계(overcount)될 때 발생한다. 과소집계는 주로 조사되었어야 할 대상이 조사되지 않은 경우(누락)에 발생하며, 과대집계는 두 번 이상 조사(중복)되었거나, 잘못된 장소나 기간에 착오로 조사된 경우에 발생한다. 예를 들어 센서스 기간 이전 사망자나 혹은 기간 이후 출생자가 조사에 포함된 경우이다(UN, 2001). 순오차는 누락과 조사오차(중복+착오에 의한 조사)간의 차이를 의미한다. 일반적으로 센서스에서는 조사오차보다는 누락이 많기 때문에 범위오차가 음의 부호를 갖는 경우가 많은 데 이를 순누락이라고 부른다.

2000년 센서스에서 0.65%(약 180만명)의 과대집계가 발생한 것으로 추정했다. 그러나, 두 달 후에 발표된 사후조사(I 차)에 따르면 2000년 센서스에서 되려 -1.18% 정도(약330만명)의 누락이 발생했다는 결과가 나왔다(U.S. Bureau of the Census, 2003).

미 센서스국은 이 두 가지 측정방식 간에 오차의 총규모뿐만 아니라 오차의 방향도 서로 상이한 결과가 나타나자, 두 결과간의 차이를 발생시킨 원인을 면밀히 검토하기 시작했다. 이후 인구분석방법은 불법체류자의 비중을 조정한 2차 결과를, 다시 불법체류자를 불법이주, 이민, 임시 이주자등으로 구성요인을 분해해서 수정한(recalibration) 3차 결과를 내놓았다. 최종적으로 인구분석방법은 2000년 센서스에서 -0.12%의 과소집계가 발생한 것으로 결론지었다. 마찬가지로 사후조사도 센서스에서 중복으로 조사된 부분을 결과에 반영시켰다. 또한 센서스에서 잘 조사되기 어려운 사람은 사후조사에서도 조사되기 힘든 correlation bias를 보정해서 2차 수정안을 2003년에 발표했다. 그 결과 최종적으로 센서스에서 0.49%의 과대집계가 발생한 것으로 추정했다(Robinson and Adlaka, 2002). 이렇게 다양한 자료원을 이용하여 센서스 범위를 평가하게 되면 오차의 원인에 대한 보다 정확한 이해와 함께 각 평가방법들이 가진 문제점들이 도출되고, 이를 통해 평가방식들이 보다 더 정교해지는 부가적인 효과도 발생한다.

자료	일자	총인구 (백만명)	차이		과 비교
			인구(백만명)	%	
센서스 집계	2000.12	281			
인구분석방법 I	2001.3	279	1.8	0.65	
인구분석방법 II	2001.3	282	0.9	0.32	
인구분석방법 III	2001.10	281	-0.3	-0.12	
사후조사(ACE)	2001.3	284	-3.3	-1.18	
사후조사(ACE) II	2003.3	280	1.3	0.49	

출처 : Robinson and Adlaka (2002), "Comparison of A.C.E. Revision II Results with Demographic Analysis".DSSD A.C.E. Revision II Memorandum Series #PP- 41. U.S. Bureau of Census (2003),<sup>6</sup> "Decision on Intercensal Population Estimates." <http://www.census.gov/dmd/www/dipe.html>

이 연구의 목적은 2005년 총조사의 범위오차를 다각도로 평가하여 오차의 규모와 원인을 파악하고, 향후 총조사의 조사범위 개선 및 평가의 정확성을 향상시키기 위한 시사점을 도출하는 데 있다.

UN의 『인구센서스 권고안』(1998)에 따르면, 센서스의 범위오차를 측정하는 방법은 크게 세 가지로 분류되는데, 인구분석방법, 내적일관성 검사, 사후조사 방법이 있다. 이 연구는 먼저 2005년 총조사 인구와 인구분석방법의 일환인 인구균형방정식, 코호트조성법으로 작성된 장래인구추계, Hamilton-Perry 기법을 이용한 2005년 총조사 기대인구(이하 H-P 기대인구)와의 비교를 통해 2005년 총조사의 범위를 전국수준 및 성과 연령별로 평가해 볼 것이다. 이어서 센서스 생존비를 통한 내적일관성 검사와 2005년 총조사의 사후조사 결과가 심층적으로 분석될 것이다. 지금까지 총조사 범위에 대한 평가는 전국수준에서의 오차 규모에 그 논의가 집중되어왔다. 그러나, 전국수준에서의 오차는 예전과 동일한 수준일 지라도 지역별로 범위가 크게 차이가 날 경우 총조사의 결과를 활용하는데 심각한 영향을 미칠 수 있다. 지역별 범위오차의 차이가 왜 중요한지는 영국의 1991년과 2001년 센서스 사례를 통해간략히 살펴보고자 한다. 마지막으로 장래인구추계와 H-P 기대인구를 기준으로 2005년 총조사의 16개 시도별 범위오차를 측정해 볼 것이다.

## II. 총조사 범위오차 평가

### 1. 인구분석방법에 의한 평가

인구분석방법은 범위오차 추정에 매우 다양한 형태로 활용되는데 다음의 네 가지로 구분해 볼 수 있다: i) 인구균형방정식- 직전 센서스결과를 바탕으로 센서스간의 출생, 사망, 인구이동의 동태자료를 고려하여 작성된 기대인구와 현 센서스 결과를 비교하는 방법, ii) 센서스간 추정방법- 직전 센서스 결과와 출산력, 사망력, 인구이동<sup>7</sup> 결과를 코호트 조성법을 이용해 작성한 인

구추계치와 현 센서스 결과치를 비교하는 방법, iii) 센서스간 코호트 생산률  
 법- 센서스간 코호트 생산률에 기초해서 두 센서스간의 연령분포를 비교하는  
 방법, iv) 코호트 회귀생산계수 비교 방법 등이다(UN, 1998).

이 장에서는 인구분석방법으로써 인구균형방정식, 장래인구추계, H-P 기  
 대인구 비교를, 내적일관성 검사방법으로써 센서스 생존비 일관성 검사를,  
 그리고 사후조사 결과를 이용해서 2005년 총조사의 범위오차 규모와 유형을  
 전국수준에서 평가해 보고자 한다.

### 가. 인구균형방정식

대표적인 인구분석방법인 인구균형방정식은 1940년 미 센서스국에서 센  
 서스 범위오차 평가를 위해 처음 사용되기 시작했다. 이 방법은 출생과 사망  
 에 관한 인구동태자료와 국제이동에 관한 행정자료를 이용하여 특정시점 인  
 구에 관한 전국치 뿐만 아니라 성과 연령, 또는 인종별로도 추정치를 제공할  
 수 있기 때문에 센서스 범위오차 평가의 벤치마크로서 오랫동안 활용되어  
 왔다(Citro, Cork, and Norwood, 2004).

2005.11.1. 추정인구	2000.7.1. 기준인구	인구변동요인			1명, %)
		출생	사망	국제이동	
48,181	47,008	2,723	1,315	-235	

2005년 총조사인구 (외국인포함)	총조사 - 추정인구	범위오차율		
		남녀합	남자	여자
47,279	-902	-1.87	-2.42	-1.32

\* 범위오차율= (총조사 - 추정인구)/추정인구 \* 100

자료 : 통계청(2006d), “장래인구추계 기준인구작성(안)” 내부자료; 통계청(2006a), 「2005  
 인구주택총조사 보고서」.

인구균형방정식으로 2005년 11월 1일자 인구를 추정하는 방법은 2000년  
 7월 1일자 기준인구(47,008천명)에 5년간의 인구증가분을 더한 후 7월에서 11

월까지 4개월간의 증가분을 합산하여 산출된다. 지난 2000년 7월 1일 이후 출생은 272만 명, 사망 131만 명, 국제이동(입국자-출국자)이 -23만 명으로 총 117만 명이 증가한 것으로 나타났다.

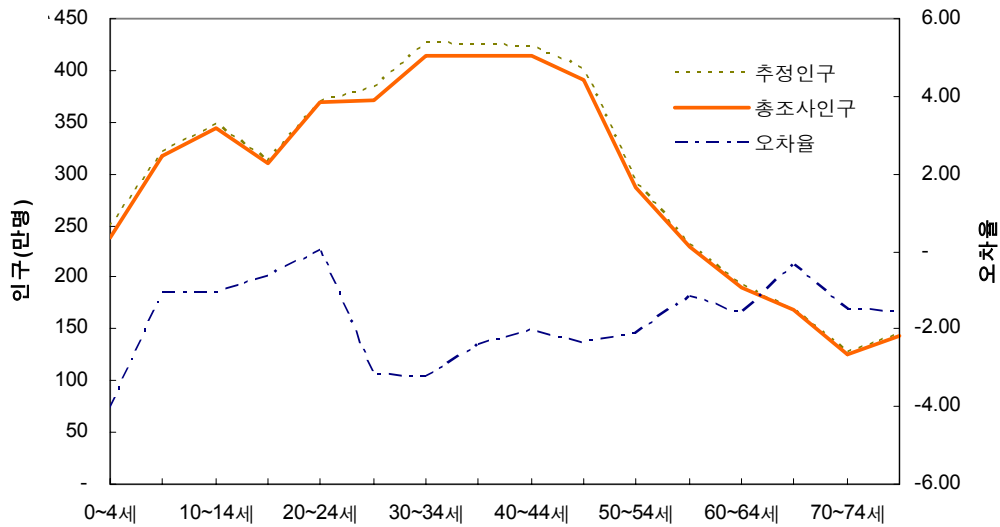
<표 2>를 보면 2005년 11월 1일자 추정인구는 총 4,818만 명으로 2005년 총조사에서 최종 집계된 인구 4,727만 명(외국인 포함) 보다 약 90만 명(1.87%) 정도 더 많은 셈이다. 2000년 추정인구 기준 총조사인구 차이가 약 99만 명(2.1%)이었던 점을 감안하면 2005년 총조사의 순누락율은 크게 향상된 것으로 보인다(통계청, 2002b; 2006d).

지난 5월에 발표된 2005년 인구주택총조사 전수집계 결과에 따르면 2005년 11월 1일 현재 남자는 2,362만 명, 여자는 2,365만 명으로 여자가 남자보다 3.1만 명 정도 더 많이 집계되어 총조사 역사상 처음으로 전체 성비가 역전된 것으로 나타났다(통계청, 2006b). 그러나 인구균형방정식에 의해 추정된 남녀간의 범위오차 차이를 보정하게 되면 2005년도 추정인구는 남자가 2,421만 명, 여자가 2,397만 명이 되기 때문에 실제 여자보다 남자가 여전히 24만 명 정도 더 많을 것으로 추정된다.

<그림 1>은 추정인구를 기준으로 총조사 범위오차를 5세 연령계층별로 살펴본 것이다. 그림의 왼쪽 축은 인구를 오른쪽 축은 총조사의 범위오차율을 나타낸다. 오차율은 4세 이하 연령에서 높았다가 5세 이후에는 -1% 전후로 낮아지기 시작해서 20-24세에는 0.01%로 전 연령에서 가장 오차율이 낮았다. 이후부터 오차율이 급격히 증가하여 25-34세까지 -3%대로 가장 높았고, 그 다음 연령부터는 점차 오차율이 향상되고 있는 것으로 보인다.

20-24세 연령의 오차율이 전 연령층에서 가장 낮다는 사실은 일반적으로 센서스에서 만성적인 누락집단이 지리적 이동성이 높은 20대 젊은 연령층이라는 사실과는 일견 상충되는 것처럼 보인다. 이러한 결과가 발생한 원인은 20-24세 인구에서 남녀간 범위오차의 방향이 서로 다르기 때문이다. 20-24세 남자의 오차율은 인구분석방법에 따르면 1.24%로 총조사에서 중복이 가장 많이 발생한 집단인데 비해, 여자의 오차율은 -1.32%로 누락이 발생하고 있기 때문이다. 따라서 남녀합은 서류의 영향이 상쇄되면서 가장 오차율이 낮아진 것이다.

<그림 1> 추정인구와 총조사인구 및 연령별 범위오차율



자료 : 통계청(2006d), “장래인구추계 기준인구작성(안)” 내부자료; (2006a), 「2005 인구주택총조사 보고서」.

#### 나. 장래인구추계 비교

총조사의 범위를 측정해 볼 수 있는 또 다른 인구분석방법은 2000년에 작성된 2005년도 장래인구추계 결과와 비교해 보는 방법이다. 장래인구추계는 코호트조성법을 사용하는데 작성과정은 다음과 같다. 먼저 인구주택총조사 결과를 기초로 인구분석방법에 의해 성과 연령별 기준인구를 작성한다. 인구변동을 발생시키는 세 가지 요인인 출생, 사망, 국제이동이 향후에 어떤 방향으로 변화될 것이라는 가정들을 세운다. 각 가정들을 조합해서 중위, 고위, 저위의 세 가지 수준별 시나리오를 작성한 후 2000년 기준인구를 출발점으로 성과 연령별로 장래인구를 추계한다.

추계인구는 1960년부터 5년 주기로 작성되었는데 2000년 총조사 결과를 기초로 지난 2001년에 장래인구추계가 발표된 바 있다. 그러나, 최근 출산율이 급격히 떨어지자 2001년에서 2004년까지의 출생, 사망, 국제이동의 실적치를 반영해서 2005년에는 특별장래인구추계(전국편 1월, 시도편 4월)를 다시

공표했다. 본 연구에서는 최근의 인구동태 실적자료가 반영된 2005년 특별장래인구추계 결과를 이용하였다.

<표 3> 특별장래인구추계 기준 2005년 인구와 총조사 성별 범위를

2005년 특별추계인구	총조사인구 - 추계인구	범위오차율 (1명, %)		
		남녀합	남자	여자
48,294	-1,015	-2.1	-2.9	-1.3

\* 장래인구추계는 7.1일자 인구이며, 외국인이 포함된 인구임  
 자료: 통계청(2005), 「특별장래인구추계」; (2006a), 「2005 인구주택총조사 보고서」

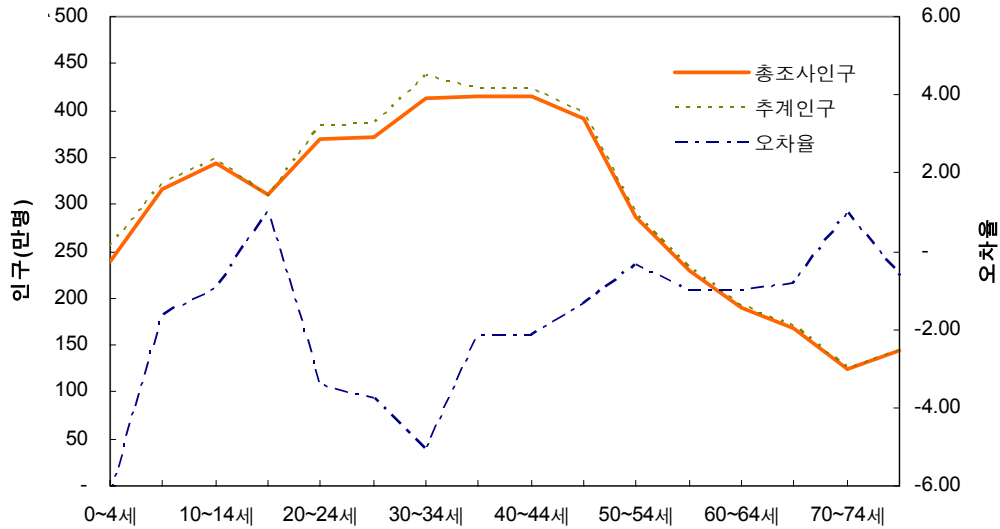
장래인구추계에서 추정한 2005년도 인구를 기준으로 놓고 보면 총조사인구는 101만명 정도의 누락이 발생한 셈이다<sup>2)</sup>. 1995년에 추정한 2000년도 추계인구와 2000년 총조사 인구의 차이가 -2.73%였던 것에 반해 2005년에는 -2.1%로 나타나 2005년 총조사의 누락율이 감소했음을 알 수 있다. 그러나, 총조사에서 여자의 오차율(-1.3%)은 인구균형방정식에 의한 추정인구나 장래인구추계 모두 유사한 수준이었지만 남자의 오차율(-2.9%)은 장래인구추계가 약간 더 높았다.

<그림 2>는 장래인구추계 기준으로 총조사의 오차율을 연령별로 추정해 본 것인데, 이 그림을 보면 지난 총조사에서 누락이 가장 많이 발생한 집단은 4세 이하 연령층(6.28%)과 30-34세(5.08%) 연령에서 가장 높았다. 반면에 총조사인구가 장래인구추계 보다 더 많았던 계층은 15-19세 인구로 총조사인구가 약 3만 명 정도 과대집계된 것으로 나타났다. 35세 부터는 누락율이 현저히 줄어들다가 70세 이상의 고령층에서는 총조사에서 과대집계된 것으로 나타난다.

2) 장래인구추계는 7월1일자 연앙인구이기 때문에 총조사 인구와 정확하게 비교하기 위해서는 장래인구추계의 결과를 각 연령별로 11월 1일자 인구로 환산해 주어야 한다. 그러나, 이 연구에서는 분석의 편의를 위하여 7월 1일자 장래인구추계 결과를 그대로 사용했음을 밝혀둔다.



<그림 2> 추계인구와 총조사인구 및 연령별 범위오차율



자료 : 통계청(2005), 「특별장래인구추계」 ; (2006a), 「2005 인구주택총조사 보고서」

추계인구를 기준으로 총조사인구를 비교할 때 저연령층의 과소집계와 노인층의 과대집계 경향은 일정부분은 추계의 특성에 그 원인이 있다. 위의 총조사 범위율 계산식에서 총조사 대신 추계인구를 기준으로 놓게 되면 인구 추계가 총조사에 비해 얼마만큼 정확한지를 알 수 있다. Bongaarts와 Bulatao(2000)가 이와 같은 방식으로 세계 여러 나라에서 행해지고 있는 추계의 정확성을 종합적으로 분석한 결과 공통적으로 나타나는 현상이 추계인구가 저연령층의 규모를 너무 많이, 고연령층은 너무 적게 추계하는 경향이 있다는 점을 발견했다. 환언하자면, 추계인구를 기준으로 센서스의 오차를 측정하면, 정도의 차이는 있겠지만 센서스에서 저연령층은 상대적으로 과소 집계 된 것으로, 고연령층은 과대집계된 것으로 나타난다(이지연, 2004).

#### 다. Hamilton-Perry 기대인구 비교

Hamilton과 Perry는 1962년에 코호트조성법처럼 코호트 생산률을 사용하지만 조성법의 복잡성을 보완하여<sup>12</sup> 별도의 인구동태자료(출생, 사망) 없이도

센서스 결과만으로 인구를 추계하는 방법을 제안했다. H-P방법은 최근 두 센서스간의 코호트 변화비(cohort-change ratio)가 다음번 센서스에서 지속된다는 가정 하에서 장래인구를 추계한다. 두 시점 간에 코호트 변화비는 이동과 사망이 발생한 결과로 보기 때문에, 이동과 사망에 대한 별도의 동태자료를 요구하지 않는다. 따라서 정확한 인구이동 자료를 구하기 힘든 소지역의 경우 이 방법을 활용하면 인구추계가 가능하다는 장점이 있다 (Smith and Tayman, 2003).

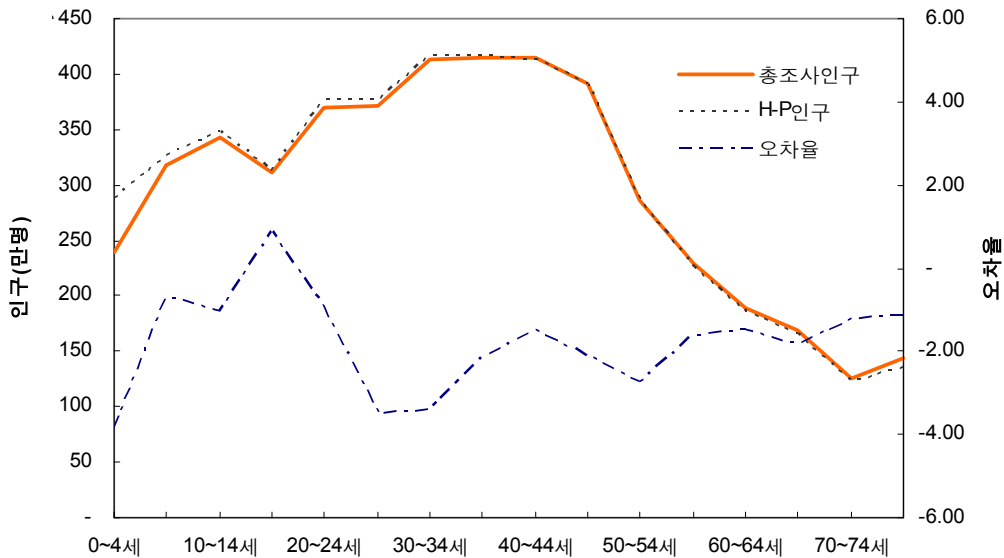
또한, 코호트조성법은 인구동태 자료인 연령별 출산율만을 사용해야 하는 반면, H-P방법은 출산율 자료도 사용할 수도 있지만 센서스 자료상에서 직접 구할 수 있는 모아비(child-women ratio)를 사용해서도 추계가 가능하다. 센서스라는 단일 자료원으로부터 추계에 필요한 모든 정보를 얻기 때문에, 속성이 다른 자료들을 연결시킬 때 발생하는 인구의 정의나 기준일자 조정 등의 문제로부터 자유로울 수 있다는 점은 H-P 방법이 가진 장점중 하나이다(Smith, Tayman, and Swanson, 2001).

장래인구추계가 전국 및 시도별로만 작성되기 때문에 이지연(2004)은 H-P 방법을 이용해서 234개 시군구별로 1995-2000년까지 총조사의 범위오차 수준이 그대로 유지된다면 2005년 총조사에서 집계될 것으로 기대되는 인구를 추정했었다. 이 장에서는 2005년 H-P 기대인구를 기준으로 총조사 범위를 평가해 보았다. H-P 기대인구는 외국인 포함되지 않은 시군구별 인구였기 때문에 여기서는 2005년 총조사의 내국인인구 자료를 사용하였다.

2005년 H-P 기대인구	총조사 - 추정인구	범위오차율			오차율 (인명, %)
		남녀합	남자	여자	
47,822	-781	-1.63	-2.01	-1.01	

자료 : 이지연(2004) "센서스 커버리지 측정을 위한 기대인구 추정기법," 통계청(2006a) 「2005 인구주택총조사 보고서」.

<그림 3> 2005년 Hamilton-Perry 기대인구와 총조사인구 및 연령별 범위오차율



자료 : 이지연(2004) "센서스 커버리지 측정을 위한 기대인구 추정기법", 통계청(2006a) 「2005 인구주택총조사 보고서」.

H-P 기대인구에 따르면 2005년도 총조사에서 집계될 인구를 총 4,782만 명으로 추정했다. 이 수치는 인구분석방법을 이용한 결과중 실제 총조사 인구와 가장 가까운 값으로, 총조사의 누락수준을 78만명, 오차율은 -1.63%로 추정했다. 남자의 누락율은 2.01%이고 여자는 그 절반에 해당하는 1.01%로 나타났다. <그림 3>의 연령별 누락율 차이를 살펴보면 장래인구추계를 기준으로 비교했을 때와 마찬가지로 총조사의 저연령층의 과소집계, 고연령층의 과대집계 패턴이 나타난다.

H-P 기대인구를 기준으로 보면 5세 이하의 오차율이 -16.8%로 유난히 높게 나타나는 데, 이로 인해 총조사 인구와 H-P 기대인구간에 발생한 차이의 60%가 이 4세 이하 연령에서 집중적으로 발생하고 있다. 저연령층에서 H-P 기대인구와 총조사인구가 차이가 많이 나는 이유 중 하나는 H-P 기대인구는 4세 이하 연령에 대한 인구를 1995년에서 2000년 까지 총조사의 모아비 변화 추세만을 이용해 2005년 인구를 추정하는데, 최근 5년간의 출산율이 유례

없이 급격하게 저하되었기 때문이다.

4세 이하 연령을 제외할 때 총조사에서 가장 많이 누락된 계층은 20대 연령층으로 특히 20-24세 인구(-2.84%)가 가장 많았고, 그 다음은 -2.55%를 보이고 있는 25-29세 연령으로 나타났다. 이것은 앞의 장래인구추계나 기준 인구를 기준으로 볼 때 최대 누락인구를 30-34세 연령으로 추정된 결과와는 약간의 차이가 있다. 그러나, 다음에 센서스 생존비를 이용한 결과도 역시 최대 누락계층을 20대 연령층으로 추정하고 있다.

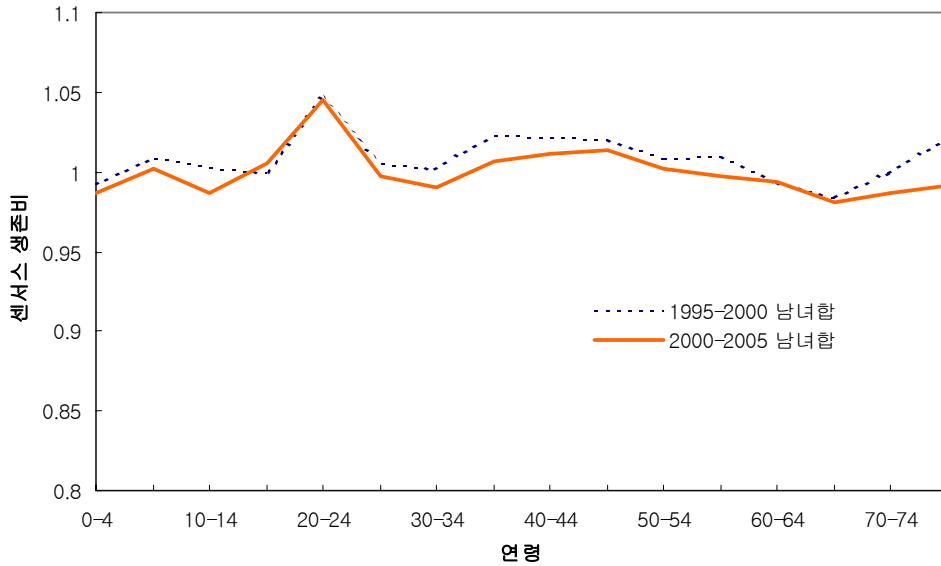
## 2. 센서스 생존비 일관성 검사

지난 센서스에서 조사된 코호트가 이번 센서스에서는 과연 얼마나 조사되었는지를 알아보게 되면 이를 통해서도 범위오차를 측정할 수 있다. 이때 고려해야 할 사항은 해당 코호트의 사망율이다. 센서스 생존비란 실제 생존 비율을 센서스 코호트 변화비로 나눈 것이다. 센서스 코호트 변화비란 이전 센서스에서 조사된 연령별 인구가 차기 센서스에서 조사된 비율이다. 생존비율은 두 센서스의 중간 시점에 발표된 생명표에서 해당 연령집단별 정지인구( ${}_nL_x$ )의 변화비율을 사용했다(Feeney, 1997; Coale and Demeny 1983; Preston and Bennet, 1983).

생존비가 1에 가까울수록 해당 코호트가 센서스에서 완전하게 조사되었음을 의미한다. 1보다 크면 차기 센서스에서 과소집계(또는 전기 센서스의 과대집계)가, 1보다 작으면 차기 센서스에서 과대집계(또는 전기 센서스의 과소집계)가 발생했음을 의미한다.

<그림 4-1>을 보면 1995년에서 2000년까지의 총조사 생존비에 비해 2000년에서 2005년 총조사는 전연령층에서 생존비 수치가 1에 가까워지고 있다. 또한 연령간의 생존비 차이가 줄어들어 안정적인 패턴을 보이는 등 최근에 올수록 총조사의 범위오차가 많이 향상되었음을 알 수 있다. 연령별로 센서스 생존비를 살펴보면 1995년부터 2005년까지의 총조사 모두 젊은층보다는 고연령층이 보다 완전하게 집계되는 계층임을 알 수 있다.

<그림4-1> 센서스 생존비 비교: 1995-2000년, 2000-2005년

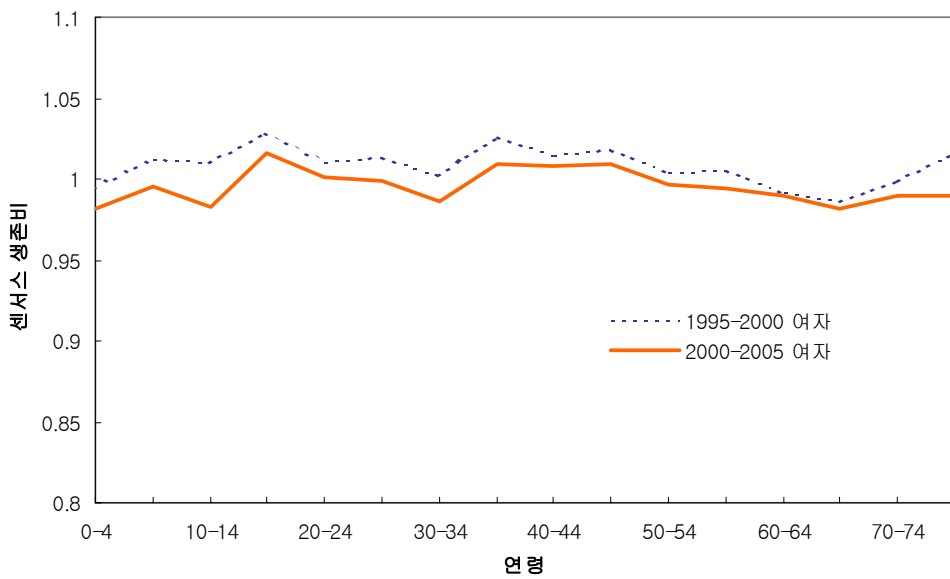
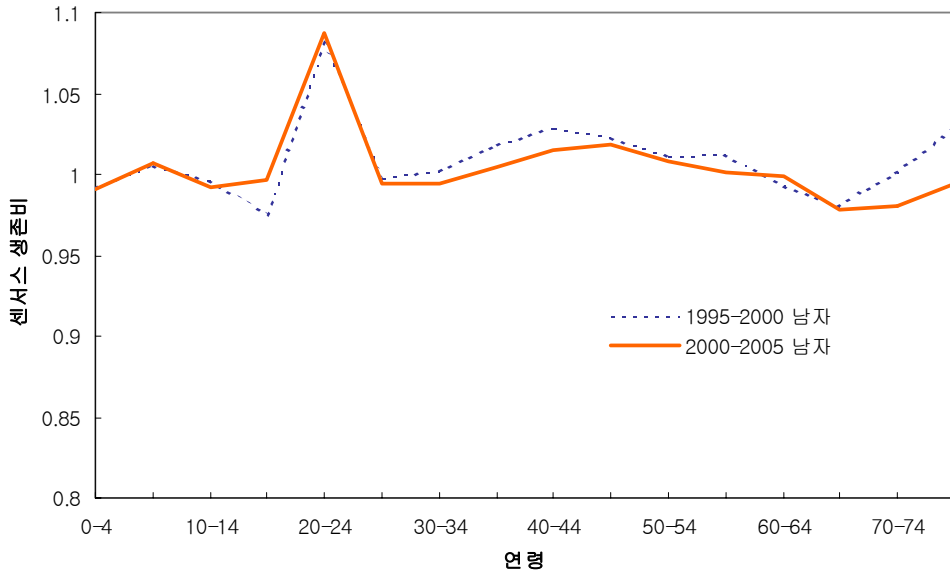


자료 : 통계청(2006a, 2002a, 1997), 해당연도 인구주택총조사 보고서.

2000년 총조사부터는 고령자(65세 이상)집단 내부의 차이가 줄어들면서 생존비가 전반적으로 많이 향상되었다. 그러나 이 연령층의 생존비가 지속적으로 1.0이하로 떨어지는 원인 중 하나는 생명표상에서 고연령층의 생존확률이 상대적으로 과소추정되는 경향 때문이기도 하다. 총조사에서 만성적으로 누락되는 계층인 20-24세(2005년 총조사 기준 25-29세)인구의 생존비는 1995년에서 2005년까지 유사한 수준에 머물러 있다. 최대 누락집단인 이 계층을 각 세별로 살펴보면 2000년에 20-21세 연령층으로 생존비가 약 1.08이었다. 이것은 생존확율을 따를 때 2005년에 108명이 있어야 한다면, 총조사에서는 100명만 조사되었음을 의미한다.

남자의 생존비는 <그림 4-2>를 보면 젊은층과 다른 연령계층간의 차이가 크고, 젊은층 내부에서도 범위오차의 차이가 큰 것을 알 수 있다. 34세의 이하의 연령에서 범위는 실제로 지난1995년에서 2005년까지 별다른 수준 차이가 없었다. 반면에 이후의 연령층에서는 최근에 올수록 총조사의 범위오차가 많이 향상되었다.

<그림4-2> 센서스 생존비 비교(남자): 1995-2000년, 2000-2005년



년

자료 : 통계청(2006a, 2002a, 1997), 해당연도 인구주택총조사 보고서

<그림 4-3>을 보면 남자에 비해 여자는 전 연령층에서 범위오차가 줄어들고 있으며, 연령간의 차이도 크게 줄었다. 특히 남자의 범위는 최근에 올수록 인구규모가 상대적으로 적은 35세 이상에서 주로 향상되었다면 여자는 대부분 인구규모가 큰 34세 이하에서 크게 향상되었다. 이러한 차이가 2005 총조사에서 남자와 여자의 성비 역전을 발생시킨 직접적인 원인으로 보인다.

### 3. 사후조사

사후조사(Post Enumeration Survey)는 센서스 결과를 자체적으로 평가하는 도구이다. 이를 통해 센서스의 프레임과 방법론등 조사과정상에서 발생하는 문제점과 센서스 결과의 범위오차 및 내용오차(content error)를 측정해 볼 수 있는 중요한 조사다(Whitford and Banda, 2001). UN의 『인구주택센서스 권고안』(1998)에 따르면 사후조사의 목적은 이용자에게는 센서스 수치의 정확성에 대한 정보를 제공하고, 센서스 담당자들에게는 향후 센서스 개선사업에 도움을 주기 위한 것이다. 또한, 센서스와는 독립적으로 실시되는 사후조사를 통해 누락된 사례와 착오로 조사된 사례를 모두 조사할 수 있기 때문에 인구균형방정식과는 달리 전국치 뿐만 아니라 지역별로도 순오차율을 측정할 수 있다는 점은 사후조사가 가진 큰 장점이다.

사후조사는 단수체계 추정방식(single-system estimation 또는 종속방식)과 이중체계 추정방식(Dual-system estimation 또는 독립방식)으로 나뉜다. 인구주택총조사의 사후조사는 단수체계 추정방식을 채택하고 있다. 이 방식은 총조사의 조사프레임과 진행방식이 사후조사에서도 그대로 활용한다(김민경, 2000). 이중체계 추정방식은 센서스에서 사용된 프레임이나 참여한 조사원, 관리자들과는 독립적인 방식으로 사후조사를 실시하고, 그 결과를 센서스 결과와 비교해서 범위의 수준을 측정하는 방법이다(UN, 1998).

사후조사에서 이중체계 추정을 사용하고 있는 대표적인 사례는 미국이다. 미센서스국은 센서스 실시 후에 특정 지역을 선정한 후 센서스와는 독립적인 표본들과 별도의 조사요원을 통해서 사후조사를 실시한다. 사후조사 결과

와 센서스 결과를 비교하기 위해 두개의 샘플이 구성된다. P-sample은 사후조사 표본지역에서 사후조사요원의 개별면접을 통해 조사한 모든 사람으로 구성된다(Hogan, 2000). E-sample은 사후조사 지역과 동일한 지역에서 센서스 조사(enumeration)에 응답한 사람들로 구성된다.

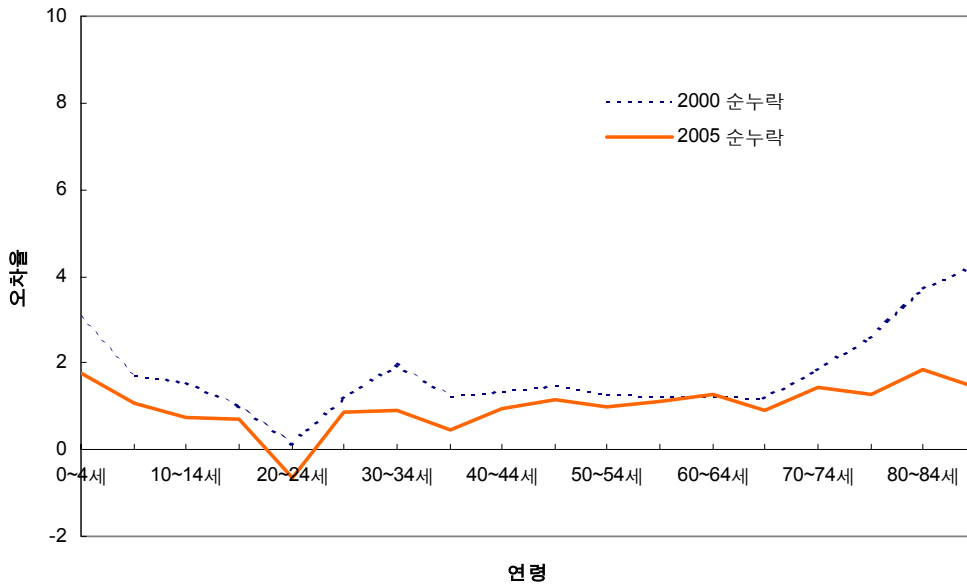
사후조사 이후 두 표본에 있는 사람들을 매칭시키는 작업을 통해서 센서스의 범위가 측정된다. 한 응답자에 관한 정보가 P-sample에는 있고 E-sample에는 없다면 이 응답자는 센서스에서 누락된 사람이다. 그러나 모든 경우가 다 센서스 누락을 의미하는 것은 아니다. 두 표본간의 불일치는 두개의 레코드를 연결시키기에 충분한 정보가 없거나, 전국비교가 아닌 P-sample 지역과 E-sample 지역이라는 한정된 장소만을 비교하기 때문일 수도 있고, 매칭하는 과정 자체에 오류가 발생해도 생길 수 있기 때문이다. 반대로 P-sample에는 없고, E-sample에만 있다면 이것은 센서스에서 착오로 조사되었거나 사후조사의 누락일 수 있다(White and Rust, 1997).

2005년 총조사 사후조사는 총조사 기준일로 부터 한 달 후인 12월 1일부터 8일간 가구원(13개), 가구(2개), 및 응답상황(3개)와 관련된 총 18개 항목을 조사했다. 단수추정방식으로 사후조사가 진행되기 때문에 통계청 지방청 조사 직원들이 총조사 기간에 조사원에 의해서 작성된 조사구 지도와 가구명부를 바탕으로 해당가구를 방문해서 총조사에서 조사된 내용을 확인하고, 누락과 중복(실제 착오로 조사된 경우를 말함), 전입 및 전출가구 등을 조사했다.

표본은 730개 조사구 43,242가구를 지역층화방법으로 추출하였다. 이번 조사에서는 일반가구 이외에도 30개 기숙시설 및 특수사회시설조사구에서도 처음으로 조사가 실시되었다. 유의할 사항은 조사구 추출확률이 평균적으로 0.27% 이지만, 지역당 최소 30개 이상의 조사구를 확보하기 위해 제주지역의 표본조사구는 1.02%(30개)로 추출확률이 큰 반면, 인천은 0.24%(33개)로 불균등추출되었다는 사실이다.



<그림 5> 사후조사 결과 총조사 순누락율: 2000년, 2005년



자료 : 통계청(2006c, 2001), 해당연도 인구주택총조사 사후조사 결과분석(내부자료)

사후조사 결과를 조사구 특성별로 살펴보면, 연립과 다세대 및 공동주택을 포함하는 아파트조사구는 가구의 오차율이 1.26%인 반면, 가구원의 오차율은 2.4%로 높았다. 이는 아파트 조사구의 경우 거처 자체가 누락될 확률이 낮고, 여러 가구가 함께 사는 비율이 낮기 때문에 가구가 누락될 확률은 적기 때문이다. 이에 반해 일반적인 단독주택 조사구의 가구와 가구원의 오차율이 모두 아파트 조사구보다 높았으나, 가구와 가구원이 각각 4.26%, 4.9%로 누락되는 정도는 유사했다.

이제 개인차원에서 사후조사 결과를 살펴보자. 사후조사에서는 총조사의 오차는 누락율, 조사착오율을 더한 값인 총오차율과 누락률에서 조사착오율을 뺀 값인 순누락율로 표현한다. 2005년 사후조사 결과 누락은 2.39%, 착오에 의한 조사가 1.49% 였다. 2000년 사후조사에서 총조사의 총오차율을 5.04%(누락 3.30% + 중복 1.74%)로 추정했는데, 2005년의 총오차율은 3.88%로 크게 감소했으며, 특히 누락율이 1% 가량 현저히 줄어들었다. 반면에 2005년 총조사의 순누락율은 0.90%로, 2000년의 순누락율 1.56%에 비해 크게

줄었다(통계청, 2006c; 2001).

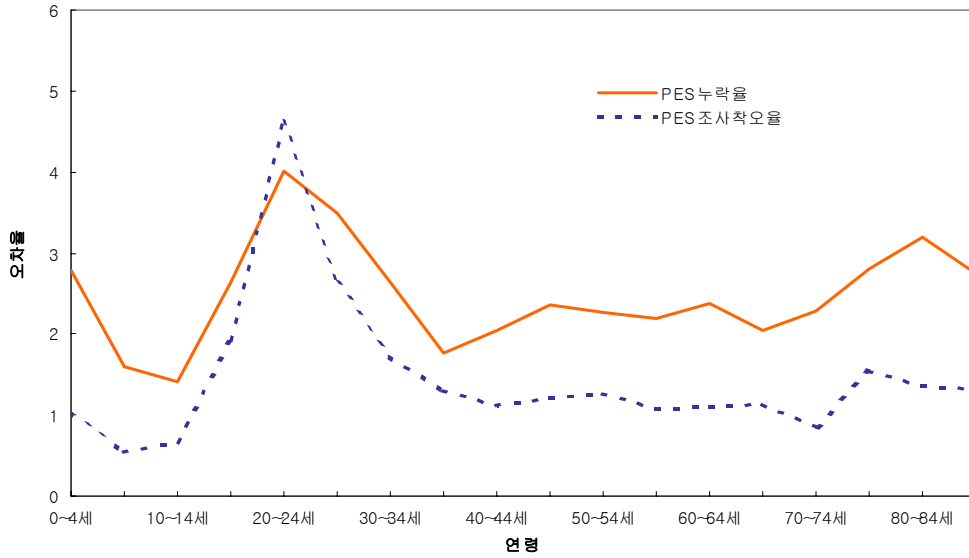
연도별 총조사의 순누락율을 연령별로 살펴보면 전 연령층에서 오차율이 줄어든 것을 알 수 있다. 2000년에 비해 2005년은 특히 65세 이상 고령자 인구, 15세 이하 유소년 인구, 30대 연령층에서 보다 완전하게 조사된 것을 알 수 있다(<그림 5> 참조).

순누락율은 총조사의 범위오차를 간략하게 표현하는 효율적인 지표이지만 이용에 주의해야할 사항이 있다. 만약, 과대집계 및 과소집계 경향이 둘 다 동일하게 높을 경우, 서로의 영향이 상쇄되어 순오차율이 0에 가깝게 나타난다는 문제점이 있다. 예를 들어 20-24세 연령은 누락(4.01%)과 조사착오(4.64%)가 가장 많이 발생한 집단이다. 총오차율은 8.65%로 가장 높았지만, 순누락율이 -0.63%로 다른 연령에 비해 가장 낮을 뿐만 아니라 오히려 총조사에서 과대집계가 발생한 것으로 나타났다(통계청, 2001).

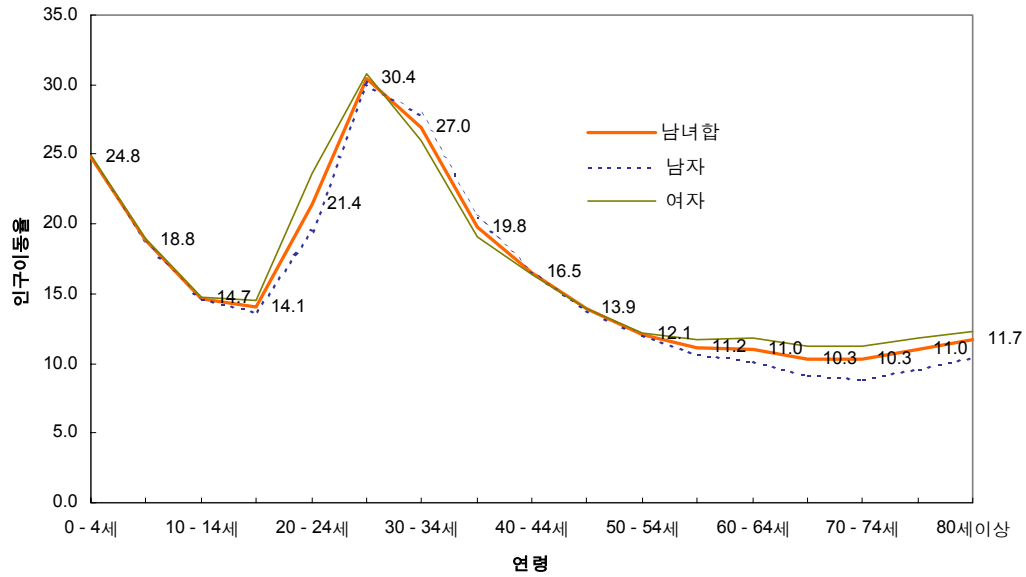
이러한 현상은 사후조사 연도별로도 나타난다. 1966년 총조사 이후부터 사후조사가 실시된 이래로 총조사의 순누락율이 가장 낮았던 해는 -0.14%였던 1990년이다. 그러나, 1990년이 착오에 의한 조사(3.46%)와 누락(3.43%) 모두 가장 많아 총오차율은 6.98%로 총조사 사상 가장 높았지만, 순누락율은 가장 낮게 나타난다(통계청, 2006c).

<그림 6>은 사후조사 결과에서 나타난 2005년 총조사의 누락율과 조사착오율을 연령별로 살펴본 것이다. 전체적으로 누락율이 조사착오율보다 높은 편이지만 두 오차의 패턴이 연령에 따라 매우 유사한 점을 알 수 있다. 누락과 조사착오 모두 4세 이하 어린이에게서 높았다가 5세-14세 까지 취학연령에서는 가장 낮았고, 20-24세에는 오차가 정점에 이른다. 이후로 오차가 낮아져서 35세 이후 부터는 안정된 수준을 유지하다가 65세를 기점으로 오차가 점차 높아지고 있다.

<그림 6> 2005년 사후조사 결과 총조사의 연령별 누락율과 조사착오율



자료 : 통계청(2006c), 인구주택총조사 사후조사 결과분석(내부자료)



자료 : 통계청 (2006e), KOSIS. 주민등록인구 <http://kosis.nso.go.kr/>

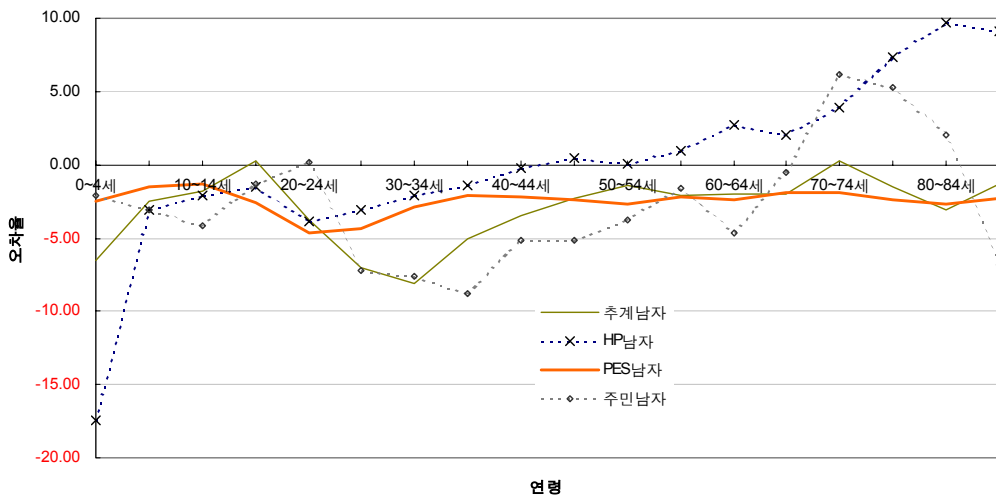
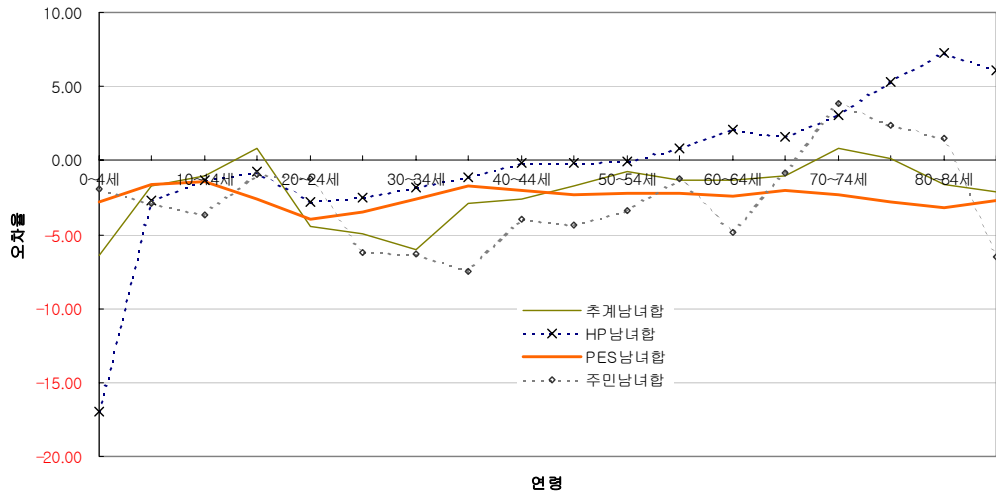
UN(2001)에 의하면 센서스에서 과소집계나 과대집계가 발생하는 근본적인 원인은 동일한데 다음의 다섯 가지 경우 중 하나다: i) 부정확하거나 불완전한 조사지도나 조사명부, 지리적으로 접근이 어려운 경우, ii) 이동 중이거나 거처가 불명확해 조사되기 어려운 경우, iii) 상주인구 개념과 같은 센서스 인구의 주요 요건이 공중에게 잘못 전파된 경우, iv) 조사 요원들이 조사의 정의와 절차를 잘못 이해한 경우, v) 관리 감독의 부재로 인해 조사활동에 대한 품질확신이 부족한 경우이다. 다시 말해서 총조사에서 누락이나 조사착오는 조사원이 응답자를 쉽게 만날 수 없거나 응답자의 상주지를 쉽게 구별할 수 없을 때이다.

총조사 오차의 연령별 차이는 연령에 따른 지리적 이동율과 밀접히 연관된다. <그림 7>은 주민등록자료를 이용해서 2005년 성과 연령별 인구이동율을 계산한 것이다. 여기서 인구이동율은 2005년 한 해 동안에 읍면동별로 신고된 전출자인구를 당해 년도 주민등록인구로 나눈 값으로 산출했다. 2005년 인구이동율은 약 18%로, 한국인 100명중 18명이 읍면동의 경계를 넘어서 다른 지역으로 이사했음을 의미한다. 이 그림을 보면 2005년 한 해 동안 25-29세 인구의 약 1/3이 이동한 것으로 나타나 전연령층에서 이동율이 가장 높았던 것을 알 수 있다. 이동율은 20대 후반과 30대 초반에서 정점에 이르다가 이후 나이가 들수록 감소한 후 65세 이후 부터 다시 상승한다. 젊은층에서 이동율이 높은 이유는 주로 이 연령이 생애사건(예, 취업, 결혼)이 가장 많이 나타나는 연령들이기 때문이다.

#### 4. 성과 연령별 오차율

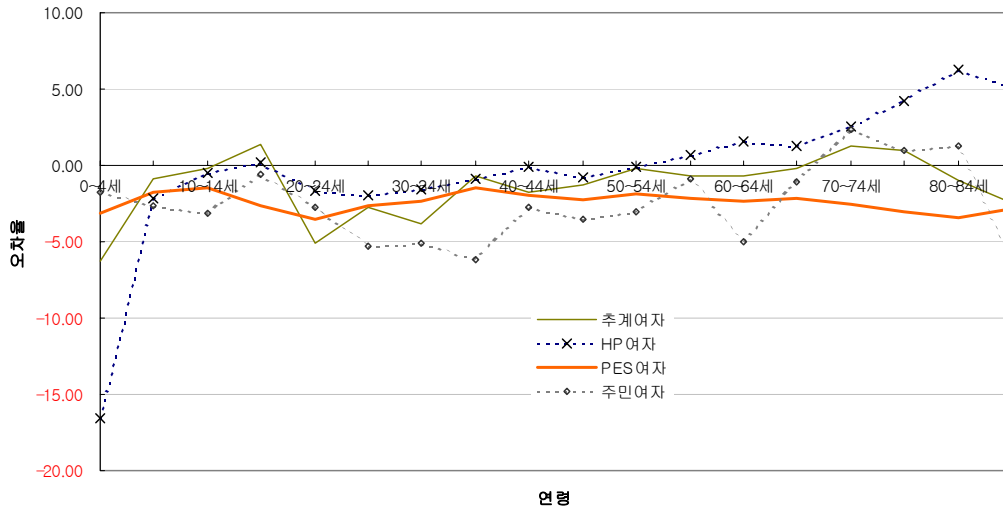
과연 2005년 총조사에서 누락이 가장 많이 발생한 집단과 가장 완전하게 조사된 계층은 누구일까? 이 장에서는 총조사에서 남자와 여자의 연령별 오차율을 장래인구추계, H-P 기대인구, 사후조사, 주민등록인구를 기준으로 살펴보았다. 사후조사는 누락율을 사용하였으며, 다른 추정결과들과 동일한 기준으로 비교하기 위해 마이너스 값을 부여했다(그림 8-1, 8-2, 8-3 참조).

<그림 8-1> 2005년 총조사 연령별 오차율(남녀합) : 장래인구추계, H-P 기대 인구 사후조사 주민등록인구 비교



자료 : 통계청(2006a, 2006c, 2006e, 2005); 이지연(2004).

<그림 8-3> 2005년 총조사 연령별 오차율(여자)



자료 : 통계청(2006a, 2006c, 2006e, 2005); 이지연(2004).

여기서 한 가지 알 수 있는 사실은 장래인구, H-P 기대인구의 오차율은 추계의 특성 때문에 오차율의 전반적인 패턴(저연령층 과소집계, 고연령층 과대집계 경향)이 유사하다는 점이다. 사후조사 결과치와 가장 근접한 오차율은 장래인구추계기준 결과이다. 이에 반해 주민등록인구는 다른 측정방법에 비해 결과 차이가 많은 편이다. 주민등록인구의 경우 5세 이하 연령의 오차율이 다른 측정치에 비해 낮은 반면 다른 결과들과는 달리 상대적으로 안정적인 것으로 추정된 35-39세, 60-64세 인구의 오차율을 높게 추정하고 있었다.

총조사에서 누락된 계층과 완전하게 조사된 계층에 대해 장래인구추계, H-P 기대인구, 사후조사, 주민등록인구를 기준으로 평가해 보면 누락보다는 완전하게 조사된 계층에 대해 더 많은 일치점을 찾을 수 있다. 추계인구, H-P 기대인구, 주민등록인구 모두 15-19세 연령층이 가장 완전하게 조사된 것으로 나타났다. 그러나, 최대 누락집단에 대해서는 H-P 기대인구와 사후조사 결과 기준으로는 20-24세 연령이, 장래인구추계는 30-34세, 주민등록인구는 35-39세 연령으로 나타나고 있다<sup>25</sup>. 각 자료간에 불일치를 보이고 있는 20

대와 30대 인구의 실제 규모에 대해서는 향후 보다 세밀한 논의와 분석이 필요할 것으로 보인다.

<그림 8-2, 8-3>은 총조사의 연령별 오차율을 남자와 여자로 구분해서 살펴본 것이다. 남자는 기준이 되는 자료가 무엇인가에 따라 오차율 차이가 크게 달라지고 있다. 그러나 한 가지 유사한 점은 어떤 추정기준을 사용하더라도 남자의 오차율이 여자보다는 높았고, 연령간의 등락폭이 크다는 점이다. 특히 젊은층과 노인층의 오차율이 여자에 비해 상대적으로 큰 편이다.

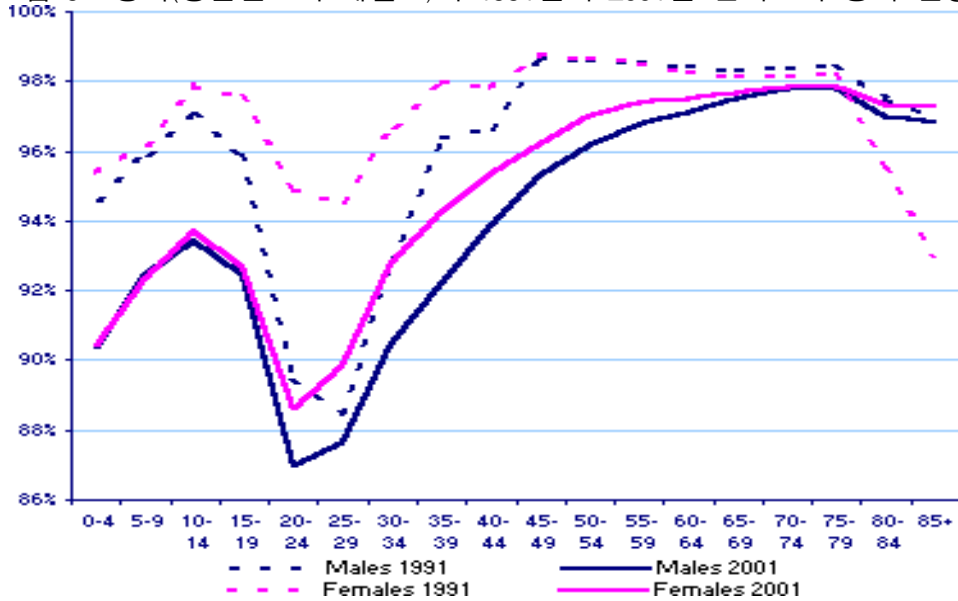
### Ⅲ. 총조사 지역별 범위오차 평가

#### 1. 영국 1991년 센서스 범위오차 사례

지금까지 총조사의 범위에 대한 논의는 주로 전국수준에서 성과 연령계층간 차이에 집중되었다. 그러나 총조사 범위의 차이는 지역에 따라서도 다를 수 있다. 예를 들어 특정지역에서는 심각한 누락이 발생했어도 다른 지역에서는 예전에 비해 완전하게 조사된다면 총조사의 범위오차는 전국수준에서는 향상될 수 있다. 그러나 이렇게 지역간 범위오차 차이가 커지게 되면 센서스 결과를 사용하는데 있어서 심각한 문제가 발생할 수 있다. 지역별 센서스 범위의 차이가 왜 중요한지는 최근 영국의 사례를 통해서 알 수 있다. 영국은 센서스 범위오차의 지역간 격차 때문에 1991년 이후 10년간 1991년 센서스 결과 대신 1981년 센서스에 출생과 사망, 인구이동 규모를 고려한 인구학적 결과를 기준인구로 사용했다.

영국의 센서스는 조사원이 방문하여 조사표를 전달하고, 우편으로 조사표를 회신하는 방법을 사용한다. 복수체계추정 방식으로 진행되는 Census Validation Survey(CVS)는 CVS 응답자와 센서스 결과를 상호 비교해서 중복되는 비율로 조사범위의 완전성을 평가한다. CVS는 범위오차 평가 뿐만 아니라 CVS 표본에 만 있는 응답자의 특성을 센서스 자료의 해당 주소지역에 외삽(extrapolate) 시키는 역할을 수행한다.

<그림 9> 영국(잉글랜드와 웨일즈)의 1991년과 2001년 센서스의 성과 연령별 응답율



출처 : ONS (2002), "National and Local Response rates (Annex B)",  
<http://www.statistics.gov.uk/census2001/annexb.asp>

1991년 CVS 결과 센서스 조사범위의 완전성을 총 98%로 추정했는데, 이는 약 1.2백만 명의 누락이 있었음을 의미한다. 실제 회수된 조사표는 96%(-2.1백만명)이지만 조사원의 추정에 의해서 보고된 인구가 약 2%(90만명 추가)가량 된다. 조사원이 추정한 90만 명의 특성은 최근 센서스 자료를 이용해서 대체(imputation)되었다. 누락된 1.2백만 명은 top-down 방식으로 전국적으로 골고루 배분되었다.

영국은 인구센서스 결과를 바탕으로 지역별 재정자원이 분배되기 때문에 전국뿐만 아니라 지역별 인구가 중요하다. 그러나, 문제는 런던 및 36개 도시에서만 20-34세(특히 남자)의 약 20%의 누락이 집중적으로 발생했다는 사실이다. 스코틀랜드는 지역 보건서비스 및 등록자료를 바탕으로, 잉글랜드와 웨일즈는 과거의 성비자료와 비교한 결과 이러한 사실이 재확인되었다. 이로 인해서 영국통계청은 1991년 센서스 결과 대신 1981년 센서스에 출생과 사망, 인구이동자 규모가 고려된 인구학적 결과를 지역별 기준인구로 사용했다 (Simpson, 1994).



<표 5> 영국의 1991년과 2001년 센서스 응답율 성과 지역별 비교

	남자(%)		여자(%)	
	1991	2001	1991	2001
런던시내	86	77	90	79
런던외곽	95	89	96	90
전체	95	93	97	94

출처 : ONS (2002), "National and Local Response rates (Annex B)",  
<http://www.statistics.gov.uk/census2001/annexb.asp>

영국에서 전체적인 응답율 하락과 지역별 누락율 격차라는 문제는 2001년 센서스에서는 더 심각해졌다. <그림 9>는 영국의 1991년과 2001년 센서스 남자와 여자의 응답율 그래프이다. 영국도 한국과 마찬가지로, 여자보다는 남자의 누락이 많고, 최대 누락계층은 20대 그 다음은 4세 이하, 그리고 연령이 높아지면서 누락이 줄어들지만, 70세 이상의 고연령층에서는 누락이 다시 증가하는 추세이다. 1991년과의 차이는 남자와 여자의 응답율이 모두 다 떨어졌지만, 젊은층에서의 남녀간 격차는 줄어들었다는 점이다.

<표 5>는 1991년과 2001년 센서스를 전국 및 런던 시내와 런던 외곽지역으로 구분한 도표이다. 2001년 센서스에서는 런던 시내에 거주하는 남자와 여자 100중 20명은 센서스에 응답하지 않았다. 런던 시내와 외곽간의 응답율 차이는 더 커졌고, 남자뿐만 아니라 여자의 응답율 하락도 두드러졌다(ONS, 2002). 이러한 이유들로 인해서, 영국은 1991년 이후로 총조사 응답율의 하락 및 지역·계층별 오차율 차이를 보정하는 One number census의 필요성을 역설해 왔다 (ONS, 2001).

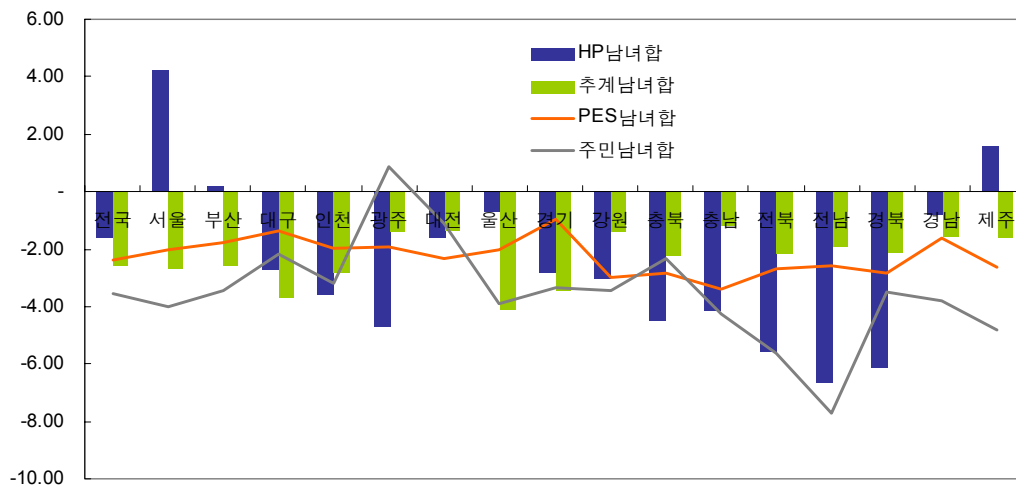
## 2. 2005년 총조사 지역별 범위오차 평가

이제 2005 총조사에서 누락이 많이 발생한 지역은 어디인지, 또 지역별 누락율에 영향을 미치는 특정한 요인들이 있는지 살펴보자. 먼저 평가방법에 따라 지역별 오차율 추정결과를 비교해 보자. 다음의 <그림 10>은 총조사에서 연령별 오차율을 장래인구추계,<sup>28</sup>H-P 기대인구, 사후조사, 주민등록인구를

기준으로 살펴본 것이다. 전국수준에서 총조사인구와 차이가 가장 적은 방식은 H-P 기대인구, 사후조사, 추계인구 순으로 총조사와 가장 차이가 많이 나는 것은 주민등록인구이다.

각 평가방식에 따라서 지역별 총조사 오차를 추정하는데 차이가 있다. H-P 인구는 도지역으로 갈수록 총조사의 오차율을 높게 추정하는 반면, 장래인구추계는 오차율이 시지역에서 높았던 것으로 추정하고 있다. 주민인구는 서울, 부산, 인천, 대전, 울산등의 시지역에서의 총조사의 범위오차율에 대해 추계인구와 유사한 추정을 보여주고 있는 반면 강원, 충남, 전북, 전남, 경북등의 도지역에서는 H-P 인구기준 오차율과 유사했다.

<그림 10> 총조사 지역별 오차율 비교: 장래인구추계 H-P 기대인구 사후조사,



자료 : 통계청(2006a, 2006c, 2006e, 2005); 이지연(2004).

이러한 차이는 각 자료의 특성과도 연관된다. 이지연(2004)이 2000년 총조사 사후조사를 분석한 결과에 따르면 조사착오를 일으킬 확률은 연령과 같은 개인차원의 변수들과 관련된다면, 누락은 주로 가구유형이나 거처의 종류와 같이 가구차원의 변수들에 영향을 받는다. 연령은 앞서 살펴보았듯이 지리적 이동성과 밀접히 관련되고, 공동주택에 비해서 일반주택이 그리고, 주

택내 많은 가구들이 거주할 수록 누락될 확률이 높다. 그렇다면 이렇게 총조사에 오차를 가져오는 요인들과 각 평가방식들에 의해서 산출된 지역별 총조사 범위오차율과는 어떤 관계가 있는지 알아볼 필요가 있다.

다음의 <표 6>는 평가방식에 따른 총조사의 지역별 오차율 추정치와 2005년도 총조사 결과 산출된 인구이동변수, 주택내 가구수와의 상관관계를 살펴본 것이다. 총조사의 인구이동 항목은 지난 5년간 비이동자와 시군구내 이동, 시군구외 이동, 다른 시도로의 이동 4가지 범주가 있다. 이중 유의미한 결과들은 주로 비이동자와 시군구내이동 범주에서 나타났기 때문에 두 범주의 결과만 수록하였다. 주택내 가구수 항목은 1가구 부터 6가구 이상까지 범주가 있으나 1가구와 4가구 이상으로 분류했을 때 유의미한 결과들이 많았다.

총조사의 지역별 오차를 설명하는데 있어 H-P 기대인구와 장래인구추계가 반응하는 요인이 서로 상이하다는 점을 알 수 있다. H-P 기대인구의 경우 시군구내 인구이동과 음의 상관관계를 보였다. 다시 말해서 시군구내 인구이동이 많은 지역에서는 H-P 기대인구 기준 총조사 오차율이 낮아진다. 반면에 장래인구추계를 기준으로 할때는 시군구내 인구이동과 양의 상관관계를 보인다. 주택내 가구수가 단독가구인 비중이 높은 지역은 주로 농촌지역이 많은데, 이런 곳에서는 H-P 기대인구 기준 총조사 오차율이 높아진다. 반면에 한 주택내 4가구 이상 다가구 비중이 높은 지역은 주로 시지역인데, 이런 곳에서는 추계인구를 기준으로 한 총조사 오차율이 높아진다. 주민인구는 주택내 가구수 보다는 인구이동율과 높은 상관관계를 보이고 있다.

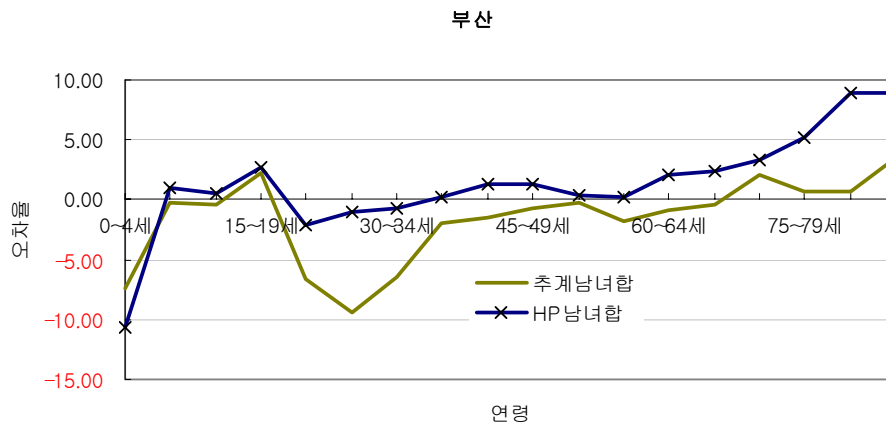
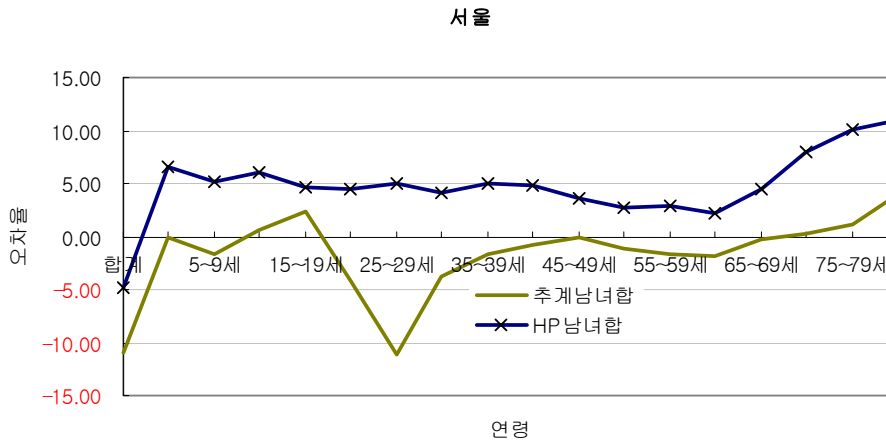
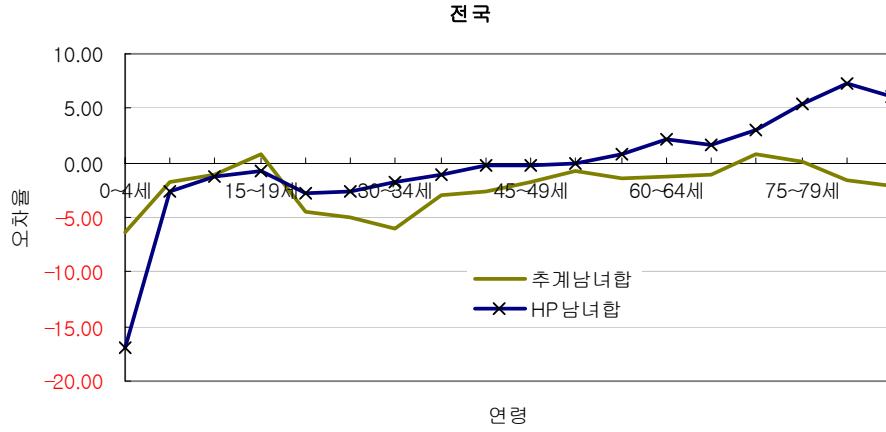
		HP오차율	추계오차율	주민오차율	PES오차율	
인구 이동	비이동	.454	-.433	.566*	.684**	계
	시군구내	-.570*	.528*	-.537*	-.657**	
주택의 가구수	단독가구	.588*	-.347	.393	.522*	
	4가구이상	-.355	.538*	-.331	-.559*	

\* p<.05, \*\*p<.01

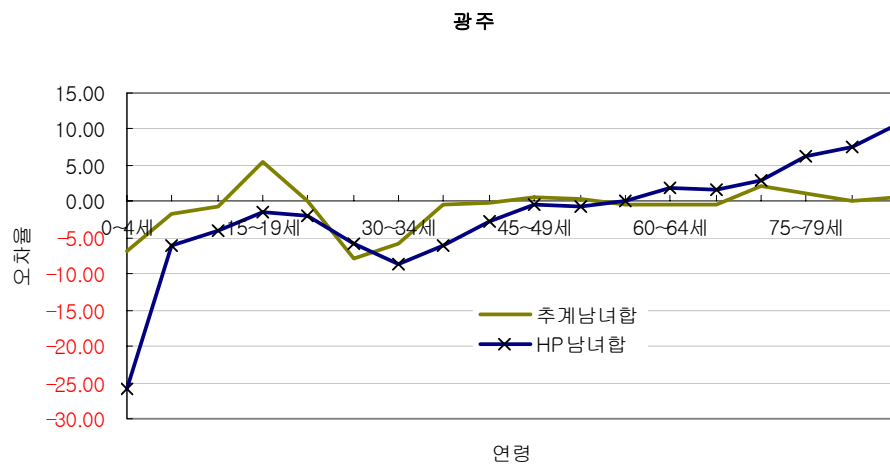
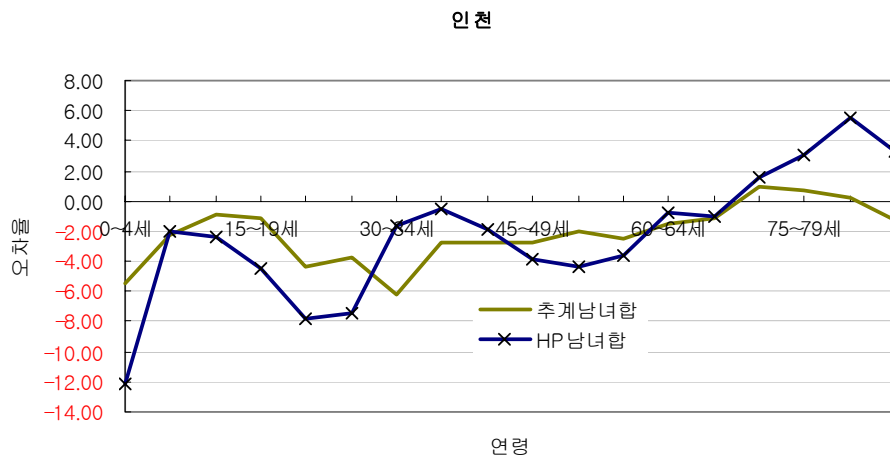
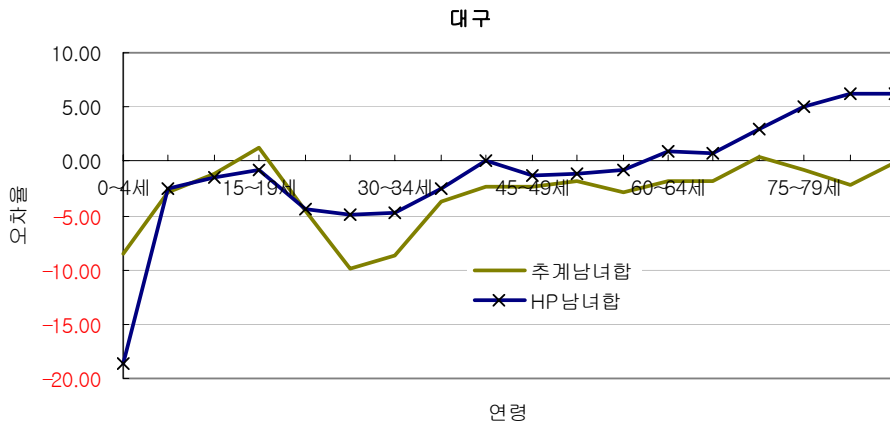
이상의 논의를 통해 장래인구추계와 H-P 기대인구에 의한 총조사 지역별 오차율이 서로 다른 경향을 가지고 있다는 사실이 확인되었기 때문에 이 두 자료를 이용하면 총조사의 지역별 오차를 보다 다각도로 측정해볼 수 있을 것이다. <그림 11>을 보면 전국수준에서는 장래추계와 H-P 모두 유사한 패턴을 보이는데, 추계기준으로는 30-34세 연령이 H-P 기준으로는 20-24세의 누락이 많았고, 그 이후부터는 총조사의 누락이 점차 줄어들면서 나이가 들수록 총조사에서 과대집계가 발생하고 있는 것으로 나타났다.

서울지역은 H-P기대인구로 추정된 오차율(4.20%)과 추계오차율(-2.70%)간의 차이가 가장 큰 지역이다. 1995년에서 2000년까지 서울이 이입보다는 이출인구가 많아 H-P기대인구에서는 2005년 인구를 감소세로 추정했다. 그러나 2001년 이후 실제로는 이입인구가 다시 많아지면서 인구가 증가했는데 이러한 경향이 2005년에 추계된 특별장래인구추계에서는 반영되었기 때문이다. 장래인구추계를 기준으로 보면 15-19세 인구는 과대집계된 반면 최대 누락연령은 25-29세 연령층이었다. 반면에 H-P기대인구와 추계인구의 오차율이 유사한 지역은 대전과 충북이었다.

<그림 11> 총조사 지역별 오차율

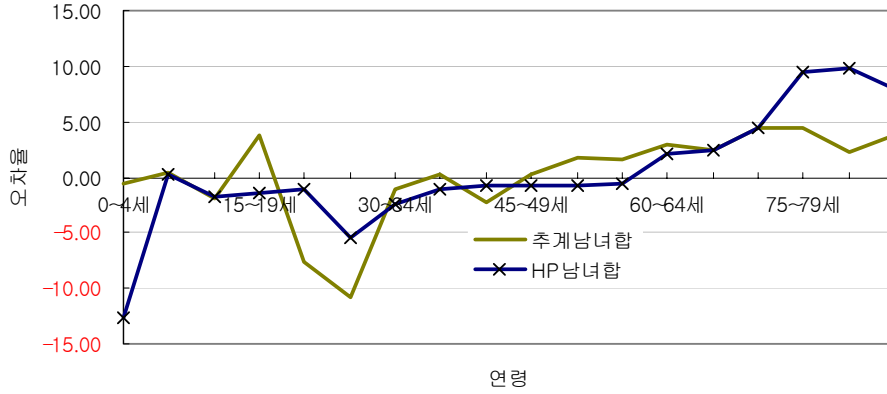


<그림 11> 총조사 지역별 오차율(계속)

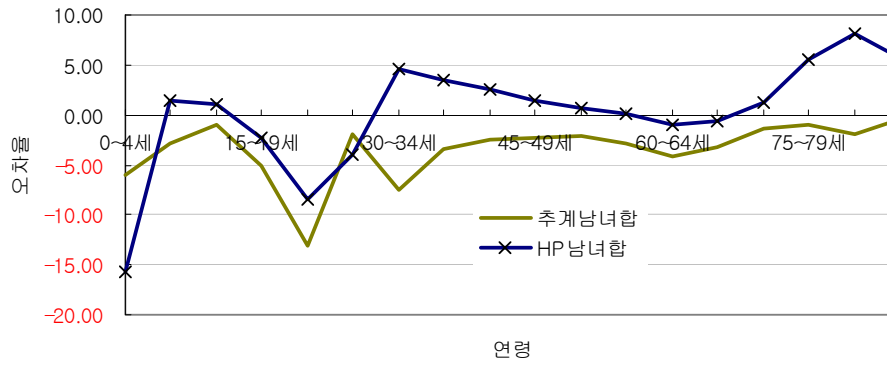


<그림 11> 총조사 지역별 오차율(계속)

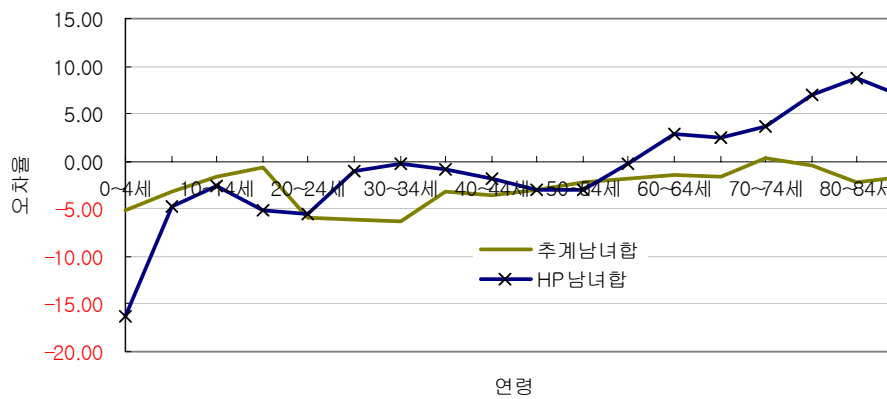
대전



울산

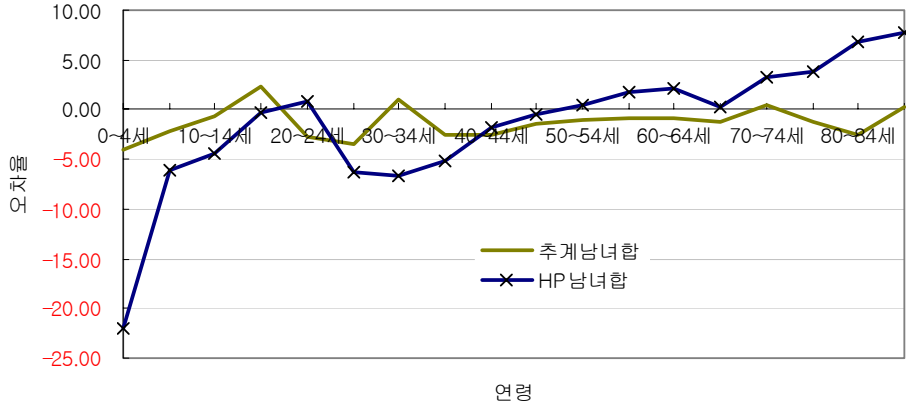


경기

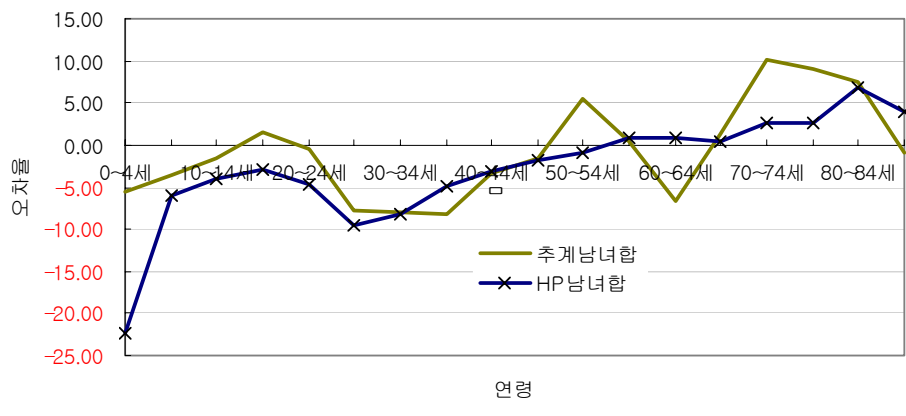


<그림 11> 총조사 지역별 오차율(계속)

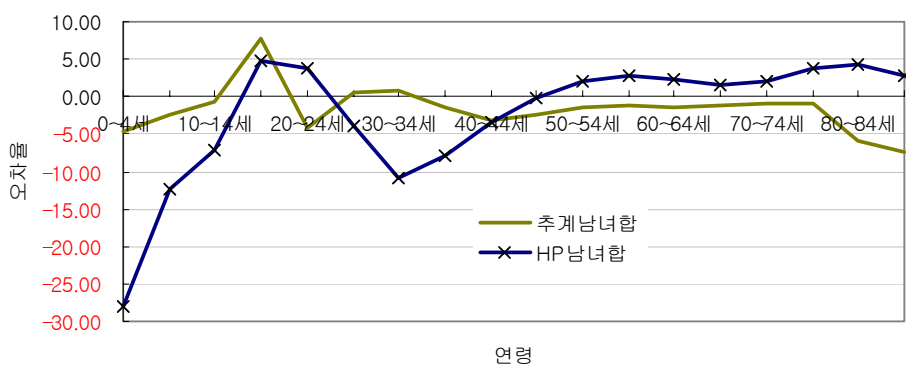
강원



충북



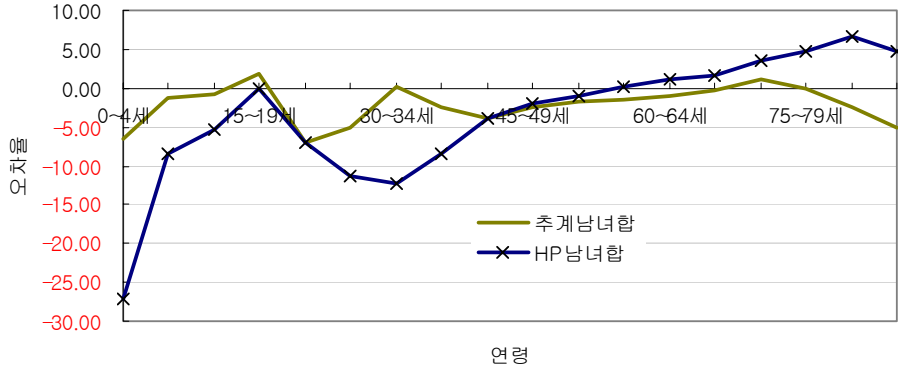
충남



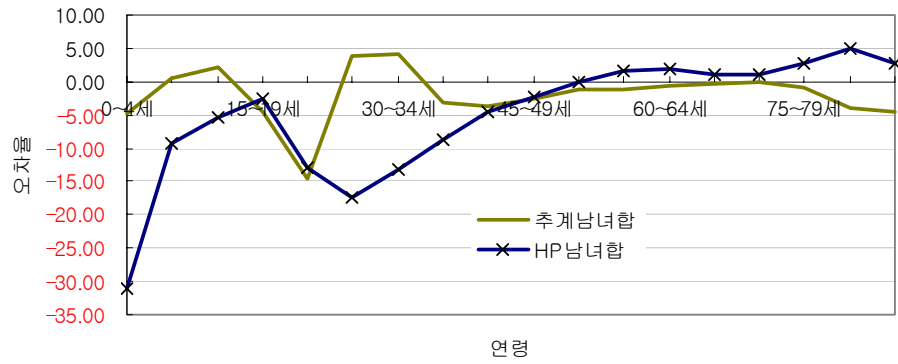


<그림 11> 총조사 지역별 오차율(계속)

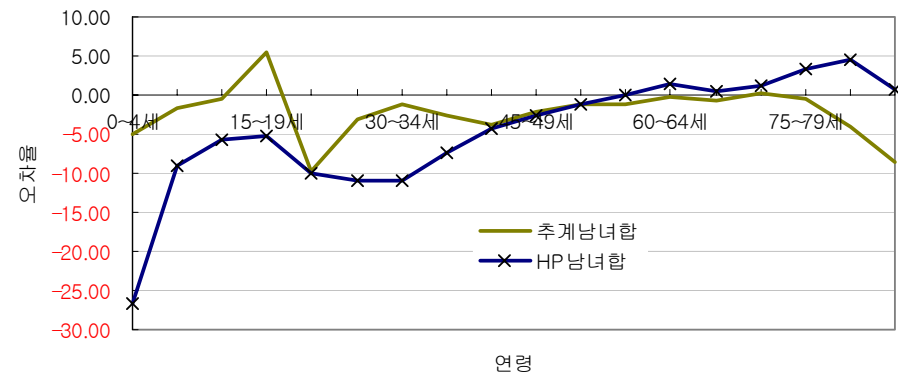
전북



전남

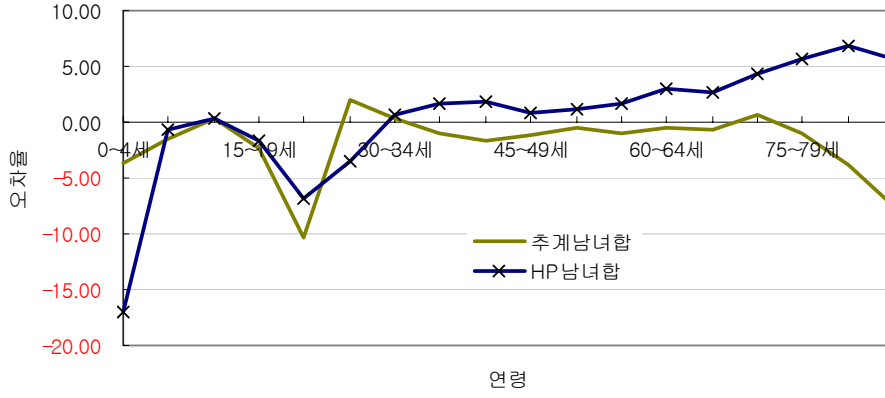


경북

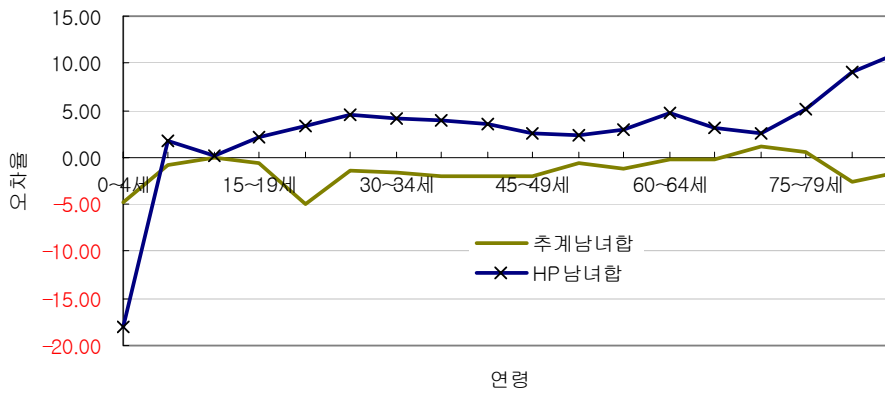


<그림 11> 총조사 지역별 오차율(계속)

경남



제주



자료 : 통계청(2006a, 2005); 이지연(2004).

## IV. 결 론

센서스와 표본조사의 가장 큰 차이는 전자가 일정한 지역이나 영토내의 모든 대상을 조사한다면, 후자는 표본으로 추출된 대상만을 조사한다는 점이다. 센서스의 목적은 한 시점에서 전체 인구의 규모와 구조에 대해 완전한 스냅샷을 제공하는 일이다. 범위오차 평가의 목적은 센서스 과정에서 발생했던 조사범위의 완전성을 저해하는 요인들을 선별해내고, 이를 통해 앞으로 실시될 센서스의 조사범위의 개선과 완성도를 높이는데 도움을 주기 위한 것이다.

2005년 총조사의 범위는 측정하는 방식에 따라 유의미한 차이가 있었다: 인구균형방정식 결과 -1.86%, 장래인구추계 기준 -2.1%, H-P 기대인구 기준으로는 1.63%, 센서스 생존비 검사 결과 -1.56%, 사후조사 결과 0.9%. 그러나, 2005년 총조사가 2000년 총조사에 비해 총인구 수준에서 순누락이 감소했다는 사실은 어떤 평가방식을 사용하든 일치된 결과를 보이고 있다. 특히 이번 총조사에서는 센서스에서 누락되기 쉬운 계층인 65세 이상 고령자층에서 범위가 크게 향상된 것으로 보인다.

2005년 총조사는 총인구수준에서 누락은 감소했지만 2000년 총조사에 비해 남녀간의 범위오차 차이가 현저히 커졌고, 지리적 이동성이 높은 20대 젊은이와 4세 이하의 연령층에서의 만성적인 누락은 여전히 발생하고 있었다. 남자에 비해 여자는 전 연령층에서 범위오차가 줄어들었고, 일반적으로 센서스에서 다른계층에 비해 누락될 확률이 높은 20대 이하와 70세 이상의 고연령층의 범위가 크게 향상되면서 연령간의 차이도 줄었다. 반면에 남자의 범위는 인구규모가 큰 30대 이하 연령에서는 예전과 별다른 차이가 없었고, 그 이후의 연령층에서는 전반적으로 향상되었다. 이러한 남녀간의 범위오차 차이가 2005년 총조사에서 남자와 여자의 성비 역전을 발생시킨 원인으로 보인다. 향후에 2005년 총조사의 조사과정에서 어떤 요인들로 인해 남자와 여자의 범위오차가 크게 벌어졌는지에 대해서는 보다 세밀한 논의와 분석이 필요할 것으로 보인다.

2005년 총조사의 범위를 측정하는 방식에 따라 결과 차이가 크게 나는 부분은 전연령층에서 누락이 가장 크게 발생한 계층에 대한 문제였다. H-P 기대인구, 사후조사, 센서스 생존비 검사결과는 20대 인구가 가장 많은 누락이 발생한 것으로 추정했다. 이에 비해 인구균형방정식에 의해 작성된 보정인구와 장래인구추계는 30-34세 인구가, 주민등록인구를 기준으로 보면 35-39세 인구가 가장 많이 누락된 것으로 추정된다. 가장 완전하게 조사된 계층은 추계인구, H-P기대인구, 주민등록인구 모두 남녀 모두 15-19세, 70-74세 인구로 동일하게 추정하고 있다. 각 자료간에 불일치를 보이고 있는 20대와 30대 인구의 실제 규모에 대해서는 향후 다양한 자료원을 동원해서 보다 종합적인 규명이 필요할 것으로 보인다. 예를 들어 주민등록증 발급현황자료, 장애인 복지인증카드 발급자료, 의료보험자료, 4대 연금자료, 운전면허등록자료, 외국인 등록현황자료 등을 연계하여 문제가 되고 있는 20대와 30대의 실제 규모를 추정해 볼 수도 있을 것이다.

한 가지 예로 2002년 시도별 주민등록증 발급현황(6.30. 현재)을 보면 발급대상자 3,692만명중 약 152만명(약 4.1%)이 주민등록증을 발급받지 않았다고 한다. 1980년 주민등록법 개정이후 17세 이상인 사람에 대한 주민등록증 소지가 의무화되었다. 성인의 경우 각종 의료, 행정, 복지서비스를 받기 위해서는 주민등록증 소지가 필수적이다. 주민등록증을 발급받기 위해서는 본인의 지문날인 절차가 필요한 만큼, 주민등록 발급자는 실제 현존했거나 현재 국내에 거주하고 있는 인구로 추정할 수 있다. 이러한 자료들을 분석해 보면 사망, 행방불명, 해외이주 등으로 실제 국내에 거주하지 않지만 행정자료상에는 남아있는 인구의 규모를 추정하는데 보조적인 정보로 활용될 수 있을 것이다.

또한 이번 총조사에서 가장 완전하게 조사된 15-19세 인구가 다음 총조사에서는 만성적인 누락집단이 되는 20-24세가 된다. 이 계층의 누락을 변화를 계속해서 주의깊게 지켜본다면 향후 측정방법에 따른 범위오차의 불일치 문제를 해결하는데 중요한 자료가 될 수 있을 것이다.

총조사 범위의 지역별 차이도 <sup>39</sup>측정방법에 따라 차이가 있다. H-P 인구를 기준으로 보면 도지역으로 갈수록 총조사의 오차율을 높게 추정하는 반면,

장래인구추계는 오차율이 시지역에서 높았던 것으로 추정하고 있다. 장래인구기준으로는 울산의 누락이 높았고, H-P기대인구를 기준으로 볼 때 전남과 경북의 오차율이 가장 높았다. 두 측정방법 모두 오차율이 높게 나타나는 지역은 20-24세 연령의 누락율이 특히 높다는 공통점이 있다.

영국이 2001년에 대대적으로 추진했던 One-number census는 낮아지는 센서스 응답율과 커져가는 계층과 지역에 따른 범위오차의 격차를 해결하기 위한 방안이었다. 지역과 계층별 오차율 차이가 보정된 센서스 결과, 즉 실측치 보다는 추정치의 성격을 가진 센서스 결과를 제공하겠다는 의미이다. One-number census를 실시하려면 센서스의 범위오차 차이를 보다 종합적으로 측정하고, 측정한 결과에 따라 지역별 인구의 규모와 구조를 사후에 통계적인 근거를 가지고 보정할 수 있어야 한다. 영국에서 One-number census가 가능했던 이유는 정교하게 설계된 사후조사가 뒷받침 되었기 때문이다.

UN의 인구주택센서스 권고안(1998)에 따르면 사후조사의 목적은 이용자에게는 센서스 수치의 정확성에 대한 정보를 제공하고, 센서스 담당자들에게는 향후 센서스 개선사업에 도움을 주기 위한 것이다. 그러나, 현재 총조사의 사후조사 결과는 총조사 자체평가 자료로서 정확성에 대한 정보만 제공할 뿐 이후 총조사 개선 및 인구추계 자료로서의 활용도가 매우 낮은 실정이다.

사후조사의 장점은 다른 평가방식들이 범위오차 만을 측정할 수 있는 반면 내용오차를 평가할 수 있다는 장점이 있다. 사후조사 결과는 전국 및 시군구 인구추계의 기준인구 자료로 활용될 수 있는데 이를 위해서는 정확한 조사오차와 누락율 측정이 선행되어야 할 것이다. 향후 사후조사의 정확성 향상을 위해서는 복수추정방식으로의 전환이 필요하다. 현재의 단수추정방식으로는 총조사에서 조사되기 어려운 사람은 사후조사에서도 조사되기 힘든 correlation bias가 복수추정방식에 비해 현저히 크다. 또한 시군구 인구추계를 위해 소지역까지 오차의 규모와 수준이 정확하게 평가 가능하도록 사후조사 표본이 재설계되어야 할 것으로 보이며, 앞으로 이에 관한 세밀한 논의와 연구가 필요할 것으로 보인다. 40

## 참 고 문 헌

- [1] 김민경. 2000. 『인구센서스의 이해』, 글로벌.
- [2] 이지연. 2004. “센서스 커버리지 측정을 위한 기대인구 추정기법.” 통계연구 보고서 04-03. 통계청.
- [3] 통계청. 2006a. 『2005 인구주택총조사 보고서』 .
- [4] ----- . 2006b. “2005 인구주택총조사 전수 (인구부문) 집계 결과.” 통계청 보도자료(5.26.).
- [5] ----- . 2006c. 『인구주택총조사 사후조사 결과분석』 . 내부자료.
- [6] ----- . 2006d. "장래인구추계 기준인구작성(안)." 내부자료.
- [7] ----- . 2006e. KOSIS. <http://kosis.nso.go.kr/>
- [8] ----- . 2005. 『특별장래인구추계』 .
- [9] ----- . 2002a. 『2005 인구주택총조사 보고서』 .
- [10] ----- . 2002b. 『인구추계작성방법』 . 내부자료.
- [11] ----- . 2001. 『인구주택총조사 사후조사 결과분석』 . 내부자료.
- [12] ----- . 1997. 『1995 인구주택총조사 보고서』 .
- [13] 행정자치부. 2002. “시도별 주민등록증 발급현황.” 정보공개청구 답변자료.
- [14] Bongaarts, John and Rodolfo A. Bulatao (eds.). 2000. *Beyond Six Billion*. Washington, DC: National Academy Press.
- [15] Citro, Constance F., Daniel L. Cork, and Janet L. Norwood (eds.). 2004. *Panel to Review the 2000 Census.; The 2000 Census, Counting Under Adversity*. Washington, D.C.: National Academies Press.

- [16] Coale, Ansley J. and Paul Demeny. 1983. *Regional Model Life Tables and Stable Populations*, Princeton, N.J.: Princeton University Press.
- [17] Dimitri, Robert C. and James B. Treat. 2000. "Examination of Census 2000 Initial Response Rates and the '90 PLUS FIVE PROJECT." *American Statistical Association Proceedings 2000*. Alexandria, VA: American Statistical Association.
- [18] Feeney, Griffith. 1997. "Research Notes: 1997 Census Survival Ration Consistency Check."  
<http://www.gfeeney.com/rnotes/1997-census-survival-ratio-check/csrcc.html>
- [19] Hogan, Howard. 2000. "Accuracy and Coverage Evaluation: Theory and Application." Dual System Estimation Workshop of the National Academy of Sciences Panel to Review the 2000 Census. U. S. Bureau of the Census.
- [20] Office for National Statistics. 2002. "National and Local Response rates (Annex B)", <http://www.statistics.gov.uk/census2001/annexb.asp>
- [21] ----- . 2001. "A Guide to the One Number Census",  
<http://www.statistics.gov.uk/census2001/pdfs/oncguide.pdf>.
- [22] Preston, Samuel H. and Neil D. Bennett. 1983. "A Census-based Method of Estimating Adult Mortality", *Popualtion Studies* 37:91-104.
- [23] Robinson, Geregory J. and Adlaka, Arjun. 2002. "Comparison of A.C.E. Revision II Results with Demographic Analysis". DSSD A.C.E. Revision II Memorandum Series #PP- 41. U. S. Bureau of the Census.
- [24] Simpson, Stephen N. 1994. "Editorial : Coverage of the Great Britain Census of Population and Housing". *Journal of the Royal Statistical Society. Series A (Statistics in <sup>42</sup>Society)*, 157: 313-316.

- [25] Smith, K. Stanley and Jeff Tayman. 2003. "An Evaluation of Population Projections by Age." *Demography* 40: 741-757.
- [26] Smith, K. Stanley, Jeff Tayman, and D. A. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic.
- [27] UN, 2001, *Handbook on Census Management for Population and Housing Censuses*, United Nations Department of Economic and Social Affairs Statistics Division: New York.
- [28] -----, 1998. *Principles and Recommendations for Population and Housing Censuses (Revision 1)*, United Nations Department of Economic and Social Affairs Statistics Division: New York.
- [29] U. S. Bureau of the Census. 2003. "Decision on Intercensal Population Estimates." Washington, DC: U.S. Census Bureau.
- [30] <http://www.census.gov/dmd/www/dipe.html>
- [31] White, Andrew A. and Keith F. Rust. 1997. *Preparing For the 2000 Census: Interim Report II*. Report of the Panel to Evaluate Alternative Census Methodologies. Washington, D.C.: National Academy Press.
- Whitford, David C., and Jeremiah P. Banda. 2001. "Post-enumeration Surveys (PES): Are They Worth It?" Presented at the Symposium on Global Review of the 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects. Statistics Division. Department of Economic and Social Affairs. United Nations Secretariat. New York