

## 기생충 감염실태조사를 위한 표본설계

류제복<sup>1)</sup> 이승주<sup>2)</sup> 전성해<sup>3)</sup>

### 요약

『전국 장내 기생충 감염실태조사』를 위해 새로운 표본설계를 하였다. 2000년 인구주택총조사의 10% 표본조사자료를 조사모집단으로 사용하였고, 조사의 특성상 각종 기생충의 감염율이 아주 낮은 관계로 통상적인 분석방법 대신에 상대위험도와 오즈비를 사용하였다. 표본배정은 네이만 배정의 절충형을 사용하였다. 또한 전국 단위와 특성별 추정이 가능하도록 하였고 추정의 정확성을 측정하기 위해 추정량의 분산 식을 유도하였다.

주요용어: 표본설계, 상대위험도, 오즈비, 네이만배정의 절충형

### 1. 서론

우리나라 국민들의 기생충 감염실태를 파악하기 위한 『전국 장내 기생충 감염실태조사』는 1971년부터 5~6년 주기로 실시되었다. 이 조사 결과는 층별, 연령별, 지역별 등으로 감염실태를 예측하여 국가 기생충관리사업의 방향설정과 계획수립을 위한 기초 자료로 사용된다.

1997년에 실시된 조사는 1990년 인구주택총조사의 10% 표본조사자료를 조사모집단으로 하였다. 그러나 기존의 모집단은 낙후되고 모집단의 변동이 생겨서 이들 모집단을 2000년에 실시된 인구주택총조사의 10% 표본조사자료로 교체한다.

기생충 감염과 보균 비율이 점차적으로 줄고 있으며 더욱이 예산의 축소로 표본 수가 감소하게 되어 2004년도는 추정의 정도에 큰 영향이 있을 것으로 판단된다. 과거에는 표본조사구내의 전 가구를 대상으로 조사하였으나 이번에는 조사 가능 범위 내에서 최대한 표본조사구 수를 늘리고 조사구 내에서 일부의 가구를 표본가구로 사용토록 하였다. 지금까지의 조사결과 분석은 단순히 표본자료의 집계에만 그쳐 필요한 추정치와 추정치의 정도에 대한 평가가 불가능 하였다. 그러나 새로운 표본설계에서는 전국단위와 특성별 추정이 가능하도록 추정량과 추정량의 분산 식을 유도하였다.

1) (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 교수  
E-mail: jbyu@cju.ac.kr

2) (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 부교수  
E-mail: access@cju.ac.kr

3) (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 전임강사  
E-mail: shjun@cju.ac.kr

## 2. 현행 조사에 대한 분석

### 2.1. 현행 조사의 개요

1997년도의 조사모집단은 1990년도 인구주택총조사의 10% 표본조사구 중에서 섬지역, 해외공관원, 그리고 특수시설 조사구를 제외한 13,328개의 보통조사구에 거주하는 사람들을 조사모집단으로 하였다. 13,328개의 모집단조사구를 24개의 집단으로 구분(대도시 6, 동부 9, 그리고 읍·면부 9)하여 모집단 가구 수의 크기에 비례하도록 203개의 표본 단위조사구를 층화확률비례추출하였다.

1997년 4월 1일을 기준으로 표본조사구내 15,237가구의 45,832가구원이 조사되었다(조사대상 가구원은 49,977명이었으므로 수검률은 91.7%이었다).

표 2.1: 행정구역별 표본현황

시·도	합계		도시(동부)		농촌(읍·면부)		피검자수
	조사구수	가구수	조사구수	가구수	조사구수	가구수	
전국	203	15,237	154	11,275	49	3,962	45,832
서울	49	3,474	49	3,474	-	-	10,664
부산	17	1,164	17	1,164	-	-	3,878
대구	10	794	10	794	-	-	2,481
인천	10	719	10	719	-	-	2,363
광주	6	443	6	443	-	-	1,240
대전	5	384	5	384	-	-	1,204
울산	3	245	3	245	-	-	781
경기	35	2,664	22	1,663	13	1,001	7,882
강원	7	510	4	293	3	217	1,404
충북	6	550	3	250	3	300	1,584
충남	8	592	2	138	6	454	1,806
전북	8	629	5	389	3	240	1,967
전남	10	850	3	228	7	622	2,082
경북	12	953	5	368	7	585	2,683
경남	14	1,089	8	618	6	471	3,268
제주	3	177	2	105	1	72	545

\* 울산시를 경상남도에서 분할

### 2.2. 지역별, 기생충별 분석

1997년도에 실시한 『제6차 한국 장내 기생충 감염실태조사』에서 피검자 개인별 원 자료를 구할 수 없었기 때문에 조사구별, 성별, 지역별, 연령대별 등으로 집계된 조사보고서

의 범주형자료를 근간으로 통계분석을 수행하였다. 따라서 개인별 원자료를 사용한 경우보다 데이터의 수가 현격하게 작아 이로부터 계산된 각종 수치들은 정확도가 떨어질 것으로 판단된다. 특히, 1997년도 조사에서 동모충, 폐흡충, 축소조충은 전혀 검출되지 않아서 분석에서 제외하였다.

표 2.2: 지역별, 기생충별 감염자수 및 감염비율(%)

	피검자수	회충	구충	편충	간흡충	요코가와 흡충	유무구 조충	왜소 조충	요충	기타
전국	45,832	28 0.06	3 0.007	17 0.04	626 1.4	142 0.3	11 0.02	8 0.02	275 0.6	3 0.007
서울	10,664	7 0.07	0 0.0	0 0.0	13 0.1	6 0.06	0 0.0	0 0.0	46 0.4	0 0.0
부산	3,878	2 0.05	0 0.0	3 0.08	229 5.9	15 0.4	1 0.03	1 0.03	29 0.8	0 0.0
대구	2,481	0 0.0	0 0.0	0 0.0	33 1.3	3 0.1	2 0.08	0 0.0	5 0.2	0 0.0
인천	2,363	1 0.04	0 0.0	0 0.0	4 0.2	2 0.09	1 0.04	1 0.04	12 0.5	0 0.0
광주	1,240	1 0.08	0 0.0	0 0.0	39 3.1	18 1.5	0 0.0	0 0.0	7 0.6	0 0.0
대전	1,204	0 0.0	0 0.0	0 0.0	0 0.0	2 0.2	0 0.0	0 0.0	5 0.4	0 0.0
울산	781	0 0.0	0 0.0	0 0.0	8 1.0	0 0.0	0 0.0	0 0.0	5 0.6	0 0.0
경기	7,882	15 0.2	0 0.0	4 0.05	21 0.3	20 0.2	2 0.03	0 0.0	63 0.8	1 0.01
강원	1,404	0 0.0	0 0.0	0 0.0	1 0.07	4 0.3	0 0.0	3 0.2	6 0.4	2 0.1
충북	1,584	1 0.06	0 0.0	0 0.0	38 2.4	1 0.06	0 0.0	0 0.0	4 0.3	0 0.0
충남	1,806	0 0.0	0 0.0	0 0.0	5 0.3	10 0.6	0 0.0	0 0.0	12 0.7	0 0.0
전북	1,967	0 0.0	0 0.0	2 0.1	37 1.9	14 0.7	0 0.0	0 0.0	30 1.5	0 0.0
전남	2,082	0 0.0	3 0.1	7 0.3	51 2.4	31 1.5	2 0.1	3 0.1	22 1.1	0 0.0
경북	2,683	0 0.0	0 0.0	0 0.0	30 1.1	15 0.6	1 0.03	0 0.0	23 0.9	0 0.0
경남	3,268	1 0.03	0 0.0	1 0.02	117 3.58	0 0.0	0 0.0	0 0.0	3 0.0	0 0.0
제주	545	0 0.0	0 0.0	0 0.0	0 0.0	1 0.2	2 0.4	0 0.0	3 0.6	0 0.0

표 2.2는 지역에 따른 기생충별 감염자수와 감염비율을 나타내고 있다. 표 2.2에 의하면, 전국적으로 간흡충, 요코가와흡충, 요충을 제외하면, 감염비율이 0.1%도 안된다. 즉 1000명당 감염자가 1명 미만이다. 간흡충은 부산지역이 가장 높고, 다음으로 경남, 광주 순이다. 요코가와흡충은 광주와 전남지역이 그리고 요충은 전북과 전남이 다른 지역에 비해 상대적으로 높다.

한편 조사구단위별 전국 CV(coefficient of variation)는 간흡충이 10.7%, 요코가와흡충이 14.08% 그리고 요충이 8.57%수준이다. 그 밖의 기생충들은 CV가 상대적으로 크다. 그리고 층별 피검자를 단순임의추출한 것으로 가정하고 기생충별 전국 CV를 계산하면, 간흡충이 3.92%, 요코가와흡충이 8.36% 그리고 요충이 6.01%로 조사구를 기준으로 한 CV보다 작다.

### 2.3. 연관성분석

본 절에서는 분류변수와 층화변수의 선정을 위해 지역(도시와 농촌)×기생충, 성별×기생충, 주거형태×기생충 등의 분할표를 작성하고 통계적 연관성분석을 실시하였다. 분석 자료로는 조사구단위로 집계된 자료를 이용하였다.

본 조사에서 각종 기생충의 감염율이 대부분 0에 가까울 정도로 낮아서 통상적인 검정 방법을 사용할 수 없어서 상대위험도와 오즈비를 사용하였다.

2×2 분할표에서 행 변수는 성별이고 열 변수는 기생충 감염여부라 하자. 남성과 여성의 기생충 감염율을 각각  $\pi_1$ 과  $\pi_2$ 라 하면, 남성과 여성간의 기생충 감염율의 비  $r = \pi_1/\pi_2$ 를 상대위험도라 하고  $\theta = [\pi_1(1 - \pi_2)]/[\pi_2(1 - \pi_1)]$ 를 오즈비라한다. 한편 상대위험도의 신뢰구간은 상대위험도가 정규분포를 따르지 않기 때문에 우선 상대위험도에 자연로그를 취하여 정규분포를 따르게 한 후 신뢰구간을 구한다. 다음에 다시 지수를 취하여 역치환하여 상대위험도의 신뢰구간을 구한다. 이렇게 구한 상대위험도의 100(1 -  $\alpha$ )% 대표본 신뢰구간은 다음과 같다.

$$(\hat{r} \times \exp(-z_{\alpha/2} \cdot ASE(\log \hat{r})), \hat{r} \times \exp(z_{\alpha/2} \cdot ASE(\log \hat{r})))$$

여기서,  $ASE(\log \hat{r})$ 는  $\log \hat{r}$ 의 점근적 표준오차(asymptotic standard error: ASE)

$$ASE(\log \hat{r}) = \sqrt{\frac{1 - n_{11}/n_1}{n_{11}} + \frac{1 - n_{21}/n_2}{n_{21}}}$$

이며  $\hat{r}$ 는 표본상대위험도이다. 상대위험도의 신뢰구간에 1이 포함되지 않으면 두 그룹간의 유의한 차이가 있다고 할 수 있다. 한편 오즈비의 100(1 -  $\alpha$ )% 신뢰구간은  $\theta$ 대신에  $\theta$ 의 동치인  $\log \theta$ 를 사용하여 신뢰구간을 구한다. 즉,  $\log \theta$ 의 100(1 -  $\alpha$ )% 대표본 신뢰구간은

$$\log \hat{\theta} \pm z_{\alpha/2} \cdot ASE(\log \hat{\theta})$$

이 된다. 여기서,  $ASE(\log \hat{\theta})$ 는  $\log \hat{\theta}$ 의 점근적 표준오차

$$ASE(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

이며  $\hat{\theta}$ 는 표본오즈비이다. 따라서 오즈비  $\theta$ 에 대한  $100(1 - \alpha)\%$  대표본 신뢰구간은

$$\left( \exp(\log \hat{\theta} - z_{\alpha/2} \cdot ASE(\log \hat{\theta})), \exp(\log \hat{\theta} + z_{\alpha/2} \cdot ASE(\log \hat{\theta})) \right)$$

이 된다.

표 2.3에 의하면, 남성이 여성보다 기생충의 감염율이 높게 나타나고 있다. 대부분의 경우 감염율이 낮아서 단순한 비율 차이로 분석하는 것은 적절치 않다. 따라서 앞서 언급한 바와 같이 상대위험도와 오즈비를 사용한다.

표 2.4에서 간흡충의 상대위험도는 2.2119로, 남성의 감염율이 여성의 2.2119배가 됨을 알 수 있다. 또한 상대위험도의 95% 신뢰구간이 1을 포함하고 있지 않기 때문에 성별에 따

표 2.3: 성별에 따른 기생충별 감염자수 및 감염비율(%)

구분	전체	남	여
대상자수(%)	49,977(100.0)	24,952( 49.1)	25,025( 50.1)
피검자수(피검율%)	45,832( 91.7)	22,484( 90.1)	23,348( 93.3)
회충(비율%)	28( 0.06)	16( 0.07)	12( 0.05)
구충(비율%)	3(0.007)	2(0.009)	1(0.004)
편충(비율%)	17( 0.04)	9( 0.04)	8( 0.03)
간흡충(비율%)	626( 1.40)	426( 1.90)	200( 0.90)
요코가와흡충(비율%)	142( 0.30)	93( 0.40)	49( 0.20)
유무구조충(비율%)	11( 0.02)	7( 0.03)	4( 0.02)
왜소조충(비율%)	8( 0.02)	7( 0.03)	1(0.004)
요충(비율%)	275( 0.60)	143( 0.70)	132( 0.60)
기타(비율%)	3(0.007)	3( 0.01)	0( 0.00)

표 2.4: 성별에 대한 기생충별 유의성

기생충	상대 위험도	상대위험도의 95% 신뢰구간	오즈비	오즈비의 95% 신뢰구간
회충	1.3846	(0.6552, 2.9261)	1.3848	(0.6550, 2.9279)
구충	2.0769	(0.1883, 22.9034)	2.0773	(0.1883, 22.9111)
편충	1.1682	(0.4508, 3.0274)	1.1685	(0.4507, 3.0291)
간흡충	2.2119	(1.8717, 2.6139)	2.2123	(1.8683, 2.6196)
요코가와흡충	1.9709	(1.3951, 2.7843)	1.9713	(1.3941, 2.7875)
유무구조충	1.8173	(0.5320, 6.2070)	1.8176	(0.5320, 6.2100)
왜소조충	7.2690	(0.8944, 59.0784)	7.2704	(0.8944, 59.0996)
요충	1.1250	(0.8886, 1.4242)	1.1252	(0.8875, 1.4265)

라 감염 차이가 있는 것으로 볼 수 있다. 요코가와흡충의 상대위험도는 1.9709로, 남성의 감염율이 여성의 1.9709배가 된다. 상대위험도의 95% 신뢰구간도 1을 포함하고 있지 않기 때문에 성별에 따라 감염율에 차이가 있는 것으로 볼 수 있다. 한편 왜소조충은 비록 상대위험도의 95% 신뢰구간에서 1의 값을 포함하고 있지만 신뢰구간의 크기가 지나치게 크다. 이는 표본감염자수가 작고 성별 간 차이가 크기 때문이다. 상대위험도로 판단하게 되면 성별에 따라 기생충별 감염율에 차이가 있음을 알 수 있다. 한편 오즈비의 경우도 상대위험도와 비슷한 수준이다. 오즈비가 1에서 멀리 떨어져 있을수록 연관성은 더욱 강한 것을 의미한다. 간흡충의 경우 오즈비가 2.2123 이므로 이는 남성이 여성에 비해 간흡충에 감염될 가능성이 2.2123배라고 할 수 있다.

도시(동부)와 농촌(읍·면부)간에도 기생충별 감염상태는 차이가 있다. 회충의 경우 상대위험도가 1.6048로 농촌이 도시보다 1.6048배 정도 감염율이 높고, 간흡충은 농촌이 도시보다 2.3829배, 유무구조충은 1.8173배, 그리고 요충은 2.6702배 감염율이 높다고 할 수 있다. 그리고 간흡충과 요충의 상대위험도에 대한 95% 신뢰구간이 1을 포함하고 있지 않기 때문에 도시와 농촌간의 감염율에 차이가 있는 것으로 볼 수 있다. 오즈비의 결과도 비슷한 수준을 나타내고 있다. 주거형태에 따른 기생충별 감염자율에 대한 카이제곱 검정결과 간흡충과 요코가와흡충에서 주거형태별 차이를 보이고 있다(한국조사연구학회(2004)).

### 3. 새로운 표본설계

#### 3.1. 모집단 분석

새로운 표본설계에서 사용할 모집단조사구는 2000년도에 실시한 인구주택총조사의 10% 표본조사구에서 통계청에서 실시하고 있는 각종 조사(경제활동인구조사, 도시가계조사 등)에 사용되고 있는 조사구를 제외한 23,536개로 구성되었다. 이는 통계청에서 사용하고 있는 표본조사구가 중복될 경우 조사 수행의 어려움을 감안한 것이다.

표 3.1: 조사구 종류별 분포

표 본	합 계	아파트	일 반	섬	기숙사	사회 복지시설	농 촌
조사구수	23,536	8,178	15,186	126	28	14	4
(%)	(100)	(34.747)	(64.522)	(0.535)	(0.119)	(0.060)	(0.017)
가구수	1,332,046	475,614	848,063	7,129	773	404	63
(%)	(100)	(35.706)	(63.666)	(0.535)	(0.058)	(0.030)	(0.005)
인구수	4,121,507	1,608,338	2,492,433	17,222	2,165	1,268	81
(%)	(100)	(39.023)	(60.474)	(0.418)	(0.053)	(0.030)	(0.002)

아파트와 일반 조사구를 제외한 조사구에서는 현실적으로 조사가 어려우므로 본 표본설계에서는 23,364개의 아파트조사구와 일반조사구만을 모집단조사구로 사용한다. 표 3.1에

표 3.2: 지역별 조사구 분포(아파트, 일반)

지역	합계			동부			읍·면부		
	계	아파트	일반	계	아파트	일반	계	아파트	일반
총계 (%)	23,364 (100.0)	8,178 (100.0)	15,186 (100.0)	18,263 (78.2)	7,300 (89.3)	10,963 (72.2)	5,101 (21.8)	878 (10.7)	4,223 (27.8)
서울	5,297	1,576	3,721	5,297	1,576	3,721	0	0	0
부산	1,837	669	1,168	1,791	658	1,133	46	11	35
대구	1,242	476	766	1,167	444	723	75	32	43
인천	1,175	524	651	1,137	522	615	38	2	36
광주	626	326	300	626	326	300	0	0	0
대전	634	287	347	634	287	347	0	0	0
울산	462	207	255	383	173	210	79	34	45
경기	4,381	1,840	2,541	3,468	1,588	1,880	913	252	661
강원	804	254	550	438	199	239	366	55	311
충북	725	255	470	399	203	196	326	52	274
충남	951	261	690	280	142	138	671	119	552
전북	1,017	333	684	611	313	298	406	20	386
전남	1,002	239	763	371	183	188	631	56	575
경북	1,457	395	1,062	629	272	357	828	123	705
경남	1,529	505	1,024	880	384	496	649	121	528
제주	225	31	194	152	30	122	73	1	72

\* 총 조사구 23,364개는 아파트와 일반 조사구만을 합한 것이다.

의하면, 아파트 조사구와 일반 조사구는 전체의 99.3%를 차지하고 있으며, 가구수와 인구수도 각각 전체의 99.4%와 99.5%로 거의 대부분을 차지하고 있다.

새로운 표본설계에서는 23,364개의 아파트조사구와 일반조사구만을 모집단조사구로 사용한다. 표 3.2에서 동부의 조사구는 18,263개로 전체 조사구의 78.2%로 읍·면부(21.8%)의 3.6배가 된다. 동부는 주로 도시지역으로 전체 아파트조사구의 89.3%를 차지하고 있다. 반면에 농촌지역인 읍·면부는 상대적으로 일반조사구의 비중이 높다. 서울, 광주, 그리고 대전에는 읍·면부가 없고 모두 동부로 구성되어 있으며, 나머지 광역시에는 약간의 읍·면부가 있다. 충남, 전남과 경북을 제외하고는 동부의 조사구가 읍·면부의 조사구보다 많다.

### 3.2. 층화

2004년 표본설계에서 사용할 모집단조사구는 2000년 인구주택총조사의 10% 표본조사자료이다. 본 조사를 통해서 전국단위와 지역별, 그리고 특성별 통계를 얻고자 한다. 적절

한 층화변수를 선정해서 층을 나누고 표본크기를 정하며, 표본배정에 사용한다.

1997년도의 조사 자료를 바탕으로 한 2.3절의 연관성분석에 의하면, 기생충 감염에서의 유의적인 차이가 발생하는 변수로는 성별, 도시(동부)와 농촌(읍·면부), 그리고 주거형태가 있다. 이들을 층화변수로 사용하기 위해서는 모집단조사구의 정보가 이들을 어느 정도 반영할 수 있는지를 가늠해야 한다. 또한 여러 변수를 사용하여 층화를 하게 되면 층의 수가 많게 되어 표본수가 작을 경우에는 상당수의 층에 표본이 배정되지 않는 문제가 발생한다.

2004년 표본설계에서는 전국단위의 통계와 함께 광역시와 광역자치단체별 통계작성을 염두에 두었다. 따라서 전국의 행정구역을 층화 기준으로 하여 7대 광역시와 9개 광역자치단체로 구분하여 16개 층을 형성하였다. 그리고 각 층 내에서는 도시와 농촌을 구분하였다.

성별과 주거형태 변수는 조사 후 사후층화방법을 적용해서 추정치의 보정에 반영하는 것이 바람직하다.

### 3.3. 표본크기의 결정 및 배분

새로운 표본설계에서는 표본규모가 4만 5천여 명에서 2만 3천여 명으로 감소하였으나 대신에 표본조사구수를 300개로 늘려서 표본의 대표성과 추정의 정도를 높이도록 하였다.

본 조사의 목적은 기생충별 감염자수(또는 감염율)을 파악하고 이에 대한 적절한 대책을 세우는 것이다. 감염율이 높은 기생충에 대해서는 특별히 관리할 필요가 있다. 1997년도 조사결과에 의하면, 3개 기생충(간흡충, 요코가와흡충, 요충)을 제외하고는 감염자수가 너무 작고 CV가 크다. 따라서 통상적으로 CV가 큰 것을 사용할 경우 표본규모가 너무 크게 되어 주어진 조사환경(비용과 시간 및 조사관리 등)에 대처할 수가 없다. 또한 이들 3개의 기생충을 제외하고는 대부분의 기생충들이 여러 층에서 발생하지 않으므로 CV가 계산되지 않아 네이만배정을 적용하여 층별 표본배정을 할 수 없다. 그러므로 표본크기의 결정에 감염율이 높은 간흡충과 요코가와흡충, 그리고 전 영역에서 기생충이 분포되고 있는 요충을 표본크기를 결정하는 변수로 사용하였다. 표본규모를 결정하기 위해서 비례배정과 네이만배정을 고려하여 표본크기를 식(3.1)과 식(3.2)를 사용하여 계산하였다.

$$\text{비례배정 : } n = \frac{N \sum N_h P_h Q_h}{N^2 V' + \sum N_h P_h Q_h} \quad (3.1)$$

$$\text{네이만배정 : } n = \frac{(\sum N_h \sqrt{P_h Q_h})^2}{N^2 V' + \sum N_h P_h Q_h} \quad (3.2)$$

여기서  $V'$ 는 허용오차로  $V' = (C' \sum W_h P_h)^2$ 이고  $C'$ 는 전국단위의 통계를 생산하는 데 사용할 변동계수로 일정한 값을 갖으며,  $P_h$  대신에 1997년도 자료를 이용한 추정치를 사용한다. 그리고  $N$ 과  $N_h$ 는 2000년도 인구주택총조사의 10% 표본자료를 사용한다.

표 3.3은 CV의 크기에 따라 3개 기생충별로 비례배정과 네이만배정을 사용한 경우의 표본크기이다. 2000년도 인구주택총조사의 10% 표본자료에 의하면, 조사구당 평균가구 수는 56.6이고 인구수는 가구당 평균 3.1명이다. 조사여건을 고려해서 질병관리본부와 실제 조사를 수행할 한국건강관리협회와의 협의를 거쳐서 300조사구를 표본조사구로, 조사구당



표 3.3: 표본규모

C.V. (%)	비례배정			네이만배정		
	간흡충	요코가와 흡충	요충	간흡충	요코가와 흡충	요충
1	623,655	1,808,087	1,167,430	400,223	1,294,959	1,085,900
2	175,987	675,351	371,091	112,938	483,689	345,175
3	80,127	330,384	173,660	51,421	236,623	161,532
4	45,460	192,631	99,528	29,174	137,963	92,577
5	29,211	125,405	64,260	18,746	89,816	59,772
6	20,330	87,908	44,840	13,047	62,960	41,708
7	14,956	64,955	33,039	9,598	46,521	30,732
8	11,461	49,917	25,344	7,355	35,751	23,574
9	9,061	39,541	20,051	5,815	28,320	18,651
10	7,342	32,087	16,256	4,712	22,981	15,121
11	6,070	26,555	13,444	3,896	19,019	12,505
12	5,102	22,337	11,303	3,274	15,998	10,514
13	4,348	19,048	9,635	2,790	13,642	8,962
14	3,750	16,434	8,310	2,406	11,771	7,730
15	3,267	14,324	7,241	2,097	10,259	6,736

표 3.4: 예상 CV(%)

	간흡충	요코가와흡충	요충
예상 CV	4.48575	9.94149	8.05577

25가구를 표본가구로 결정하였다. 이에 따라 표본크기는 최종 23,250명으로 정하였다. 표본크기가 23,250명이면 네이만배정에서 3가지 기생충에 대한 예상 CV는 표 3.4와 같다.

한편 300개의 표본조사구를 배정하는 방법으로 비례배정(식 3.3)과 네이만배정(식 3.4)을 고려하였다(참고; 표 3.5).

$$\text{비례배정 : } n_h = n \frac{N_h}{N} \tag{3.3}$$

$$\text{네이만배정 : } n_h = n \frac{N_h \sqrt{P_h Q_h}}{\sum N_h \sqrt{P_h Q_h}} \tag{3.4}$$

비례배정은 층별 조사구수를 기준으로 하였으므로 서울과 경기와 같이 조사구가 많은 곳에 표본조사구가 많이 배정된다. 그러나 이들 지역에는 기생충 발생이 다른 층에 비해 상대적으로 낮다. 특히 3개의 주요 기생충의 발생이 낮으므로 이들을 기준으로 볼 때 표본이

표 3.5: 표본조사구 배정

	비례배정	네이만배정			
		간흡층	요코가와 흡층	요층	절층 (평균)
전국	300	300	300	300	300
서울	72	27	37	64	43
부산	25	65	33	29	43
대구	15	18	10	9	12
인천	15	6	9	14	10
광주	9	17	22	9	16
대전	7	3	6	7	5
울산	4	5	3	5	4
경기	52	29	54	62	49
강원	10	3	10	9	7
충북	9	13	5	6	8
충남	12	6	18	13	12
전북	12	17	21	19	19
전남	15	25	37	20	27
경북	18	20	28	22	23
경남	21	43	3	8	18
제주	4	3	4	4	4

너무 많이 배정된다. 반면에 3개의 주요 기생충에 대해 계산한 네이만배정의 경우는 감염율과 CV를 고려한 관계로 간흡층의 경우는 부산, 경남 지역에 그리고 요코가와흡층은 경기, 전남 지역에 상대적으로 표본조사구가 많이 배정되는 등 기생충별로 층별 표본조사구 배정에 큰 차이를 보이고 있다. 그러나 조사의 목적이 감염율이 높은 기생충에 대해서 특별한 관리와 감독이 요구되고, 발생이 높은 기생충에 대한 보다 정확한 통계생산이 요구되므로 이들 3개 주요 기생충에 대한 네이만배정의 절층형(3개 기생충에 대한 네이만배정의 평균값)을 적용하여 표본조사구를 배정하는 것이 적절하다고 판단하였다.

300개의 표본조사구를 우선 16개 층(광역시와 광역자치단체)에 네이만의 절층형을 사용하여 배정하고 층내에서 동부와 읍면부의 조사구수(참고; 표 3.2)에 비례해서 표본조사구를 배정하였다. 이때 표본조사구수는 층에 최소 3개를 배정하였다. 표 3.2에 의하면, 7개 광역시에는 읍·면부의 조사구가 동부에 비해 상당히 작으므로(울산은 전체 조사구에서 읍·면부가 17.1%를 차지하고 있으나 배정된 표본조사구가 작다) 동부에만 표본조사구를 배정하였다. 최종적으로 층별, 동부와 읍·면부에 배정된 표본조사구수는 표 3.6에 있다.

표 3.6: 동부와 읍·면부의 표본조사구

	합 계	동 부	읍·면부
전국	300	228	72
서울	43	43	0
부산	43	43	0
대구	12	12	0
인천	10	10	0
광주	16	16	0
대전	5	5	0
울산	4	4	0
경기	49	39	10
강원	7	4	3
충북	8	4	4
충남	12	4	8
전북	19	11	8
전남	27	10	17
경북	23	10	13
경남	18	10	8
제주	4	3	1

### 3.4. 표본추출

16개 층(광역시와 광역자치단체)으로부터 표본조사구를 추출할 때 층내를 도시(동부)와 농촌(읍·면부)으로 나누어 계통추출법을 적용한다. 계통추출을 위해서 조사구를 행정구역으로 정렬하고 아파트조사구와 일반조사구 순으로 정돈한 후에 각각 가구 수에 확률비례하는 층화확률비례계통추출법을 사용하였다. 이러한 추출방법은 표본조사구를 지역별로 고르게 뽑을 수 있고 또한 행정구역 내에서 아파트조사구와 일반조사구가 균형있게 추출되어서 표본의 대표성이 커진다. 한편 표본가구는 표본조사구로부터 비복원단순임의추출법을 사용한다.

### 4. 추정

본 조사의 표본추출방법인 층화이단집락추출법을 통하여 얻어진 표본조사 자료를 이용하여 모집단에 대한 추정치(전국, 층별, 특성별)를 산출하는 추정방법을 제시하였다. 새로운 표본설계에서 사용된 조사모집단은 2000년 인구주택 총조사의 10% 표본자료를 사용한 관계로 전국 또는 지역별(층별) 추정에 있어서는 표 4.1의 가중치를 사용해야 한다. 그리고 특성별(성별, 도시, 농촌 등) 전국 추정치 계산을 위해서는 통계청의 자료(<http://www.nso.go.kr>)

를 이용하여 표 4.1과 같은 방법으로 계산한 가중치를 사용한다. 상세한 추정과정과 특성별 추정치는 한국조사연구학회(2004)에 있다.

실제 조사를 마치고 얻은 자료들에는 무응답, 부재, 응답 저절 등의 이유로 결측치가 생기며, 또한 조사 결과의 특성차별(예를 들면, 성별, 도시·농촌별 등) 비율이 모집단의 비율을 따르지 않는다. 이와 같은 경우 가중치를 조정하여 추정에 반영한다.

표 4.1: 가중치(2000년 인구주택총조사 기준)

지역	전수조사자료		조사모집단 (10% 표본조사자료)		가중치( $W_h$ )	
	가구수	인구수	가구수	인구수	가구수	인구수
전국	14,391,374	46,136,101	1,323,677	4,100,771	10.8723	11.2506
서울	3,109,809	9,895,217	29,1046	905,906	10.6849	10.9230
부산	1,124,001	3,662,884	105,398	332,654	10.6644	11.0111
대구	762,194	2,480,578	69,793	220,604	10.9208	11.2445
인천	751,130	2,475,139	66,965	216,405	11.2168	11.4375
광주	409,571	1,352,797	36,155	115,920	11.3282	11.6701
대전	415,134	1,368,207	35,980	113,665	11.5379	12.0372
울산	307,819	1,014,428	26,442	84,085	11.6413	12.0643
경기	2,691,510	8,984,134	250,865	814,746	10.7289	11.0269
강원	489,226	1,487,011	44,823	130,745	10.9146	11.3734
충북	463,524	1,466,567	41,774	124,789	11.0960	11.7524
충남	591,971	1,845,321	53,470	157,545	11.0711	11.7130
전북	604,082	1,890,669	57,581	171,467	10.4910	11.0264
전남	666,389	1,996,456	57,517	164,324	11.5860	12.1495
경북	891,774	2,724,931	83,774	240,631	10.6450	11.3241
경남	955,081	2,978,502	88,840	265,803	10.7506	11.2057
제주	158,159	513,260	13,254	41,482	11.9329	12.3731

\* 조사모집단은 통계청에서 사용하고 있는 표본가구를 제외한 아파트와 일반 가구.

#### 4.1. 층별(지역별) 추정

$h$ 층(광역시 또는 광역자치단체)에서 특정 기생충의 감염자 총수와 감염율에 대한 추정을 위해서 가중치를 사용한다. 10% 표본조사자료를 조사모집단으로 사용한 관계로 가중값을 가구기준 가중값과 인구기준 가중값을 생각할 수 있는데, 여기서는 최종단위가 인구가므로 인구기준 가중값을 사용한다.

$h$ 층에서 특정 기생충에 감염된 총 감염자수에 대한 추정량은 다음 식과 같다.

$$\hat{Y}_h = W_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} = W_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi} = W_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} \hat{Y}_{hi}. \quad (4.1)$$

여기서,

$W_h$ : 표 4.1에 있는  $h$ 층에 대한 가중값(모집단 확대 인자)

$N_h$ :  $h$ 층의 모집단 조사구수

$n_h$ :  $h$ 층의 표본조사구수

$M_{hi}$ :  $h$ 층  $i$ 번째 표본조사구내의 총 가구수

$m_{hi}$ :  $h$ 층  $i$ 번째 표본조사구에서 추출된 표본가구수

$y_{hij}$ :  $h$ 층  $i$ 번째 표본조사구의  $j$ 번째 표본가구내에서 특정 기생충에 감염된 사람 수

이다. 그리고  $h$ 층에서 특정 기생충에 감염된 총 감염자수에 대한 추정량의 분산추정량은 식(4.2)와 같다.

$$\text{var}(\hat{Y}_h) = W_h^2 \left( N_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 \frac{M_{hi} - m_{hi}}{M_{hi}} \frac{s_{hi}^2}{m_{hi}} \right). \quad (4.2)$$

여기서,

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (\hat{Y}_{hi} - \hat{Y}_h)^2,$$

$$s_{hi}^2 = \frac{1}{m_{hi} - 1} \sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2$$

이다. 그리고  $\hat{Y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{Y}_{hi}$  이고  $\bar{y}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$  이다.

#### 4.2. 전국 추정

특정 기생충에 대한 전국 총 감염자수에 대한 추정량은 식(4.1)로부터 얻는다. 즉,

$$\hat{Y} = \sum_{h=1}^{16} \hat{Y}_h = \sum_{h=1}^{16} W_h \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}. \quad (4.3)$$

그리고 특정 기생충에 대한 전국 총 감염자수에 대한 분산 추정량은 식(4.2)로부터 바로 얻을 수 있다.

$$\begin{aligned} \text{var}(\hat{Y}) &= \sum_{h=1}^{16} \text{var}(\hat{Y}_h) \\ &= \sum_{h=1}^{16} W_h^2 \left( N_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 \frac{M_{hi} - m_{hi}}{M_{hi}} \frac{s_{hi}^2}{m_{hi}} \right). \end{aligned} \quad (4.4)$$

## 5. 결론

2004년도 『제7차 전국 장내 기생충 감염실태조사』를 위한 새로운 표본설계는 2000년에 실시된 인구주택총조사의 10% 표본조사자료를 조사모집단으로 사용하였다. 2004년도 조사에는 예산의 감소로 부득이 표본규모를 축소할 수밖에 없었다. 표본의 축소로 인한 추정 정도의 정도를 고려하여 표본조사구수를 300개로 확대하고 대신에 표본조사구내에서 일부의 표본가구를 사용하였다. 조사의 특성상 각종 기생충의 감염율이 매우 낮은 관계로 통상적인 분석방법 대신에 상대위험도와 오즈비를 사용하였다. 또한 표본배정에 있어서도 특정 기생충들의 변동계수가 큰 차이가 나고, 특정 지역에서는 기생충 감염이 발생하지 않는 등 비례배정이나 특정 기생충만을 기준으로 한 네이만 배정은 적절치 않아 네이만 배정의 절충형을 사용하였다. 그리고 과거에는 조사결과를 단순히 집계하는 수준이었으나 새로운 표본설계에서는 전국 단위와 특성별 추정이 가능하도록 하였고 추정의 정확성을 측정하기 위해 추정량의 분산 식을 유도하였다.

## 참고문헌

- 김영원·류제복·박진우·홍기학 공역 (1998). <표본조사의 이해와 활용>. 자유아카데미.  
 박홍래 (2000). <통계조사론>. 영지문화사.  
 보건복지부·한국건강관리협회 (1992). <제5차 한국장내 기생충감염현황>.  
 보건복지부·한국건강관리협회 (1997). <제6차 한국장내 기생충감염현황>.  
 한국조사연구학회 (2000). <2001년도 국민건강·영양조사 표본설계 및 표본조사구 추출 연구용역 최종보고서>.  
 한국조사연구학회 (2004). <제 7차 전국 장내 기생충실태조사 표본설계 연구용역 최종보고서>.  
 한국통계학회 (2001). <시민보건지표조사 및 건강증진 프로그램개발 표본설계 및 표본조사구 추출 연구용역 최종보고서>.  
 Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc.  
 Cochran, W. G. (1977). *Sampling Techniques*, John Wiley & Sons, Inc.  
 U.S. Department of Health and Human Services. (1999). *Vital and Health Statistics - National Health Interview Survey: Research for the 1995-2004 Redesign*, Series 2, No. 126.  
 U.S. Department of Health and Human Services. (2000). *Vital and Health Statistics-Design and Estimation for the National Health Interview Survey, 1995-2004*, Series 2, No. 130.

[ 2004년 7월 접수, 2004년 8월 채택 ]

## A Sample Design for Intestinal Parasitic Infection Survey

Jea-Bok Ryu <sup>1)</sup> Seung-Joo Lee<sup>2)</sup> Sung-Hae Jun<sup>3)</sup>

### ABSTRACT

We made a new sample design for intestinal parasitic infection survey in 2004. We used the 10% sample survey data of 2000 population and housing census as a survey population. Since the infection rates of intestinal parasitics are very low, we applied the relative risk and odds ratio instead of ordinary method such as t-test to study the characteristics from the 1997 survey data. In order to allocate samples to stratum, we used the compromise of Neyman allocation which is the average of three Neyman allocations. And also, we derive estimators and variance estimators of the estimators.

*Keywords:* Sample design, Relative risk, Odds ratio, Compromise of Neyman allocation

- 
- 1) Professor, Division of Life Science-Genetic Engineering-Statistics, Cheongju University, 36 Naedok-Dong, Sangdang-Gu, Cheongju, Chungbuk, 360-764, Republic of Korea.  
E-mail : jbryu@cju.ac.kr
  - 2) Associate Professor, Division of Life Science-Genetic Engineering-Statistics, Cheongju University, 36 Naedok-Dong, Sangdang-Gu, Cheongju, Chungbuk, 360-764, Republic of Korea.  
E-mail : access@cju.ac.kr
  - 3) Full Time Lecturer, Division of Life Science-Genetic Engineering-Statistics, Cheongju University, 36 Naedok-Dong, Sangdang-Gu, Cheongju, Chungbuk, 360-764, Republic of Korea.  
E-mail : shjun@cju.ac.kr