

코호트 기반 조사 공변수 자료의 신뢰도 평가 연구: 원전주변지역주민 역학조사연구

배상혁¹, 박보영¹, 이충민², 안윤옥^{1,2}

¹서울대학교 의과대학 예방의학교실; ²서울대학교 의학연구원 원자력영향역학연구소

Reliability of Covariates in Baseline Survey of a Cohort Study: Epidemiological Investigation on Cancer Risk Among Residents Who Reside Near the Nuclear Power Plants in Korea

Sanghyuk Bae¹, Bo-Young Park¹, Zhong Min Li², Yoon-Ok Ahn^{1,2}

¹Department of Preventive Medicine, College of Medicine, Seoul National University;

²Institute of Radiation Effect & Epidemiology, Seoul National University Medical Research Centre

Objectives: We evaluated the reliability of the possible covariates of the baseline survey data collected for the Epidemiological Investigation on Cancer Risk Among Residents Who Reside Near the Nuclear Power Plants in Korea.

Methods: Follow-up surveys were conducted for 477 participants of the cohort at less than 1 year after the initial survey. The mean interval between the initial and follow-up surveys was 282.5 days. Possible covariates were identified by analyzing the correlations with the exposure variable and associations with the outcome variables for all the variables. Logistic regression analysis with stepwise selection was further conducted among the possible covariates to select variables that have covariance with other variables. We considered that these variables can be representing other variables. Seven variables for the males and 3 variables for the females, which had covariance with other possible covariates, were selected as representative variables. The Kappa index of each variable was calculated.

Results: For the males, the Kappa indexes were as follow; family history of cancer was 0.64, family history of liver diseases in parents and siblings was 0.56, family history of hypertension in parents and siblings was 0.51, family history of liver diseases was 0.50, family history of hypertension was 0.44, a history of chronic liver diseases was 0.53 and history of pulmonary tuberculosis was 0.36. For females, the Kappa indexes were as follow; family history of cancer was 0.58, family history of hypertension in parents and siblings was 0.56 and family history of hypertension was 0.47.

Conclusions: Most of the possible covariates showed good to moderate agreement.

Key words: Baseline survey, Cohort study, Confounding factors, Covariates, Effect modifiers, Reliability
J Prev Med Public Health 2010;43(2):159-165

서론

설문조사는 비교적 작은 노력으로 대규모 자료를 수집하는데 용이하여, 역학 연구 시 자료를 수집하기 위해서 널리 쓰이는 방법이다. 그러나 응답자의 기억력에는 한계가 있어 올바르게 않은 응답을 할 수도 있고, 질문의 방식이나 질문에 대한 응답자의 이해 정도에 따라 응답이 달라질 수 있다는 단점이 있어 설문지를 개발할 때 설문항목을 표준화하는 등의 노력이 필요하다 [1]. 이런 단점을 보완하기 위해

세심한 주의를 기울여 설문조사를 시행하였다 하더라도 응답자가 사실과 다른 응답을 하거나, 설문 문항에 대하여 연구자가 의도한 바와 다르게 응답자가 이해할 가능성은 여전히 존재하는 것이므로 [2], 자료 수집이 완료된 후 수집된 자료의 타당성 또는 신뢰성을 확인하는 것이 필요하다.

원전주변지역 역학조사연구는 원전주변지역 거주주민들과 대조지역 거주주민들을 추적 관찰하여 원전주변지역 거주주민들에게서 방사선 관련 암의 발생이 늘어나는지를 확인하기 위한 연구로, 1992년부터 2005년까지 기반코호트가

Table 1. Basic characteristics of subjects according to survey intervals

	Survey intervals						All cohort
	<1y	>=1y, <2y	>=2y, <3y	>=3y, <4y	>=4y, <5y	>=5y	
N	477	766	401	312	748	1,505	36,606
Age (mean \pm SD)	59.6 \pm 13.4	61.6 \pm 12.3	61.8 \pm 11.9	63.5 \pm 10.7	60.5 \pm 11.6	61.4 \pm 11.7	54.6 \pm 14.5
Sex (%)							
Male	188 (39.4)	301 (39.3)	144 (35.9)	112 (35.9)	256 (34.2)	496 (33.0)	15,223 (41.6)
Female	289 (60.6)	465 (60.7)	257 (64.1)	200 (64.1)	492 (65.8)	1,009 (67.0)	21,383 (58.4)
Group (%)							
Near	258 (56.2)	252 (32.9)	240 (59.8)	180 (57.7)	482 (64.4)	786 (52.2)	11,403 (31.1)
Intermediate	164 (34.4)	501 (65.4)	120 (29.9)	127 (40.7)	258 (34.5)	532 (35.4)	10,496 (28.7)
Far	45 (9.4)	13 (1.7)	41 (10.2)	5 (1.6)	8 (1.1)	187 (12.4)	14,707 (40.2)

구축되었다. 연구자료로 기초조사자료는 거주지역을 비롯하여 대상자의 연령, 생활수준, 결혼 상태 등의 인구학적 요인, 생화학적 및 이학적 검사, 직업력, 과거력, 가족력, 생활습관 요인 등이 수집되었고 추적관찰조사를 통하여서 암의 발생과 암으로 인한 사망 등을 수집하고 있다. 이 중 노출변수인 거주지역, 결과 변수인 암의 발생 및 사망, 그리고 생화학적 및 이학적 검사 자료를 제외한 나머지 자료는 표준화된 설문을 이용한 면접조사 방식으로 수집되었다. 이러한 설문조사로 수집된 자료들은 노출변수와 결과변수의 관계에 공변수로 작용할 가능성이 있는 정보들을 조사한 것으로, 노출변수와 결과변수 사이의 관련성을 평가하기 위해서는 이들 자료의 질적 수준, 즉 타당성과 신뢰성을 확인하여야 할 필요가 있다 [3,4].

타당성을 확인하는 가장 확고한 방법은 응답자 중 표본 집단을 선정하여 설문조사로 수집된 정보와 응답자의 실제 정보를 비교하여 얼마나 정확히 조사된 것인지를 평가하는 것일 것이나, 설문조사의 항목이 313문항으로 그 자료의 양이 방대하고 자세하여 다른 정보원으로부터 정확한 정보를 수집하는 것이 어려울 뿐 아니라 과거 의료 이용 등 일부 항목을 제외한 다른 항목들은 설문조사 이외의 이용 가능한 정보원이 존재하지 않는 것들이기 때문에 타당성을 이러한 방법으로 확인하는 것은 거의 불가능하다 [5]. 타당성을 확인하는 다른 방법으로는 조사된 설문 항목들 사이의 관련성이 얼마나 논리적으로 합당한지 평가하는 방법이나 임상 검사결과와 설문의 항목의 정보를 대조하여 관련된 항목의 타당성을 평가하는 방법 등이 있을 것이나 이 역시 방대한 자료를 평가하기 어렵고 모든 설문 항목이 임상검사와 관련성이 있는 것이 아니기에 평가에 포함되지 않을 수도 있어 충분하지 않다. 또, 타당성은 설문지를 개발하거나 설문조사 방법을 적용하는 단계에서 주로 평가되는 것 [6]이므로 설문조사가 이미 수행된 상태에서 자료의 질을 평가하기에는 적절하지 않다.

한편, 자료가 이미 수집된 상태에서는 신뢰성을 평가함으로써 자료의 질적 수준을 가늠할 수도 있다. 신뢰성을 확

인하는 방법으로는 같은 대상을 반복 조사된 설문조사의 항목 간의 일치도를 평가하거나 동시에 수집된 자료를 반복으로 나누어 그 사이의 일치도를 평가하는 방법 등이 있다.

이 연구에서는 같은 연구 대상자에게 반복 시행된 설문조사 자료를 이용하여 그 자료의 신뢰성을 평가하고자 하였다. 모든 설문조사 항목에 대하여 신뢰성을 확인하는 것이 가능하기는 하지만, 전체 설문조사 자료의 신뢰성을 대표적으로 보여 줄 수 있는 방법이라 하기 어려우며, 실제 연구 성적에 영향을 주는 것은 공변수로 작용할 수 있는 정보들일 것이므로 이 연구에서는 교란변수와 변조변수일 가능성이 있는 항목들을 대상으로 신뢰성을 확인하고자 하였다.

연구방법

1. 분석대상자료

1992년 7월부터 2005년 12월까지 원전주변지역 역학조사 연구 기반코호트에 등록된 사람은 36,606명이었다. 이 중 2회에 걸쳐 반복 조사 자료를 연구대상으로 선정하였다. 2회에 걸쳐 반복 조사된 자료의 수는 4209건으로, 조사 간격에 따른 대상자들의 기본적인 정보는 Table 1에 정리되어 있다. 조사 간격이 길어지는 경우, 시간의 흐름이 신뢰도에 영향을 줄 수 있고, 병력과 같은 변수들은 신뢰도와 무관하게 시간에 따라 변화할 수 있는 것이므로 반복조사가 최초 조사 시기 후 1년 이내인 자료들만 대상으로 연구를 시행하였다. 조사 간격이 1년 미만인 자료는 모두 477건으로, 남자가 188명, 여자가 289명이었다. 노출 변수인 거주지역은 근거리지역이 258명, 근거리대조지역이 164명, 원거리대조지역이 45명이었다. 연구대상자의 평균 나이는 59.6세였다. 2회 반복 조사 사이의 기간은 최저 77일, 평균 282.5일이었다.

Table 2. Possible covariates in male

Variables(variable name)	Association with cancer*		Correlation with residential area
	HR	95% CI	p-value [†]
History of TB(TBHX)	1.37	1.07-1.77	0.002
History of chronic liver diseases(GGANHX1)	1.73	1.03-2.92	<0.001
History of pneumonia(PNHX)	1.61	1.19-2.17	<0.001
History of acute liver diseases(GGANHX)	1.38	1.13-1.68	<0.001
History of allergic diseases(ALLERGHX)	0.77	0.62-0.97	<0.001
History of vasectomy(VASEXTHX)	0.38	0.21-0.73	<0.001
Family history of hypertension(TCVAOPF)	0.54	0.38-0.76	<0.001
Family history of liver diseases(TGANF)	0.38	0.21-0.70	<0.001
Family history of cardiac diseases(TSIMF)	0.48	0.24-0.98	<0.001
Family history of cancer(TCANF)	0.67	0.53-0.87	<0.001
History of chemotherapy(CHEMOT)	1.10	1.02-1.18	<0.001
Smoking of more than 400 cigarettes(SMOK400)	1.30	1.17-1.44	<0.001
Drinking history(ALCOHL)	1.87	0.80-0.96	<0.001
Family history of hypertension in parents and siblings(CVAOPF1)	0.62	0.42-0.90	<0.001
Family history of liver diseases in parents and siblings(GANF1)	0.41	0.19-0.87	<0.001
Family history of cancer in paternal grandparents, uncles, aunts and cousins(CANF2)	0.40	0.16-0.98	<0.001
Frequency of drinking beer(BEERFRQ)	0.78	0.73-0.84	<0.001
Frequency of drinking liquor(GINFRQ)	0.60	0.45-0.78	<0.001

HR: hazard ratio.
 *Cox proportional hazard model.
[†]p-value: Chi-square test.

Table 3. Possible covariates in female

Variables(variable name)	Association with cancer*		Correlation with residential area
	HR	95% CI	p-value [†]
History of GB stone(GBSTONHX)	1.61	1.13-2.29	<0.001
Family history of hypertension(TCVAOPF)	0.60	0.40-0.91	<0.001
Family history of stroke(TCVAF)	0.63	0.43-0.92	<0.001
Family history of cancer(TCANF)	0.63	0.45-0.89	<0.001
History of chemotherapy(CHEMOT)	1.20	1.06-1.37	<0.001
Smoking of more than 400 cigarettes(SMOK400)	1.36	1.18-1.57	<0.001
Family history of hypertension in parents and siblings(CVAOPF1)	0.62	0.40-0.97	<0.001

HR: hazard ratio.
 *Cox proportional hazard model.
[†]p-value: Chi-square test.

2. 분석 대상 변수의 선정

분석 대상 변수로는 전체 변수 중에서 공변수일 가능성이 있는 변수들을 선정하고 이들 중 대표할 수 있는 변수를 선정하였다.

공변수일 가능성이 있는 변수를 찾기 위하여 먼저, 결과 변수와의 연관성을 가질 수 있는 변수를 파악하였다. 이를 위해 결과변수인 암 발생 또는 암으로 인한 사망 여부에 대하여 전체 코호트 자료 중 1차 조사로 얻어진 자료의 변수들을 대상으로 콕스 비례 위험 모형(Cox proportional hazard model)을 이용한 단변수분석을 시행하여 위험비를 분석하였다. 대상 질병인 암은 남성과 여성에서 그 원인과 양상이 다른 질병으로, 원인과 결과의 연관성에 있어서도 차이가 있을 것이므로 남성과 여성을 나누어서 분석을 수행하였다.

그 뒤, 노출변수인 거주지역과의 상관성을 가지는 변수를 찾기 위해 단변수분석에서 유의한 결과가 나온 변수들을 대상으로 노출변수와의 관련성을 보기 위하여 카이제곱검정을 시행하였다. 결과변수와 연관성을 가지는 동시에 노출변수와의 상관성을 가지는 변수를 공변수로 작용할 가능성이 있는 변수로 생각하였다. 선정된 변수 중 남성에서는 3개, 여성에서는 1개의 변수가 결측치가 50% 이상으로, 분석에 적합하지 않은 것으로 생각되어 분석 대상에서 제외되었다. 공변수로 작용할 가능성이 있는 변수로 최종 선정된 변수는 남성과 여성에서 각각 Table 2,3과 같다.

이렇게 하여 찾아진 공변수로 작용할 가능성이 있는 변수들 중에서 다른 변수들과 공변성이 가장 큰 변수들을 대표 변수로 삼고자 하였다. 이 대표적인 변수들을 찾기 위해서, 전체 코호트 자료 중 1차 조사로 얻어진 자료를 이용하

Table 4. The selection of variables in logistic regression model with stepwise selection method in male

Tested as dependent variable	Tested as independent variables																		
	PNH X	TBH X	GGA NHX	GGA NHX 1	ALLE RGH X	TCV AOP F	VAS ECT HX	TGA NF	TSIM F	TCA NF	CHE MOT	SMO K400	ALC OHL	CVA OPF 1	GAN F1	CAN F2	BEE RFR Q	GINFRQ	
GINFRQ	X	X	X	O	X	X	X	X	X	X	X	X	O	X	X	X	O		
BEERFRQ	X	O	X	O	X	X	X	X	X	X	X	O	O	X	X	X		O	
CANF2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	
GANF1*	X	X	X	O	X	X	X	X	X	X	X	X	X			O	X	X	
CVAOPF1*	X	X	X	O	X	X	X	X	O	X	X	X	X		X	X	O	X	
ALCOHL	X	X	X	O	X	X	X	X	X	X	O			X	X	X	O	O	
SMOK400	X	O	X	O	X	X	O	X	X	X	X		O	X	X	X	X	X	
CHEMOT	X	X	O	O	O	X	O	X	X	X		O	O	X	X	X	O	X	
TCANF*	X	X	O	O	X	X	O	X	X		X	X	O	X	X	X	O	X	
TSIMF	O	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	
TGANF*	X	X	X	X	O	X	X		O	X	X	X	X	X	X	X	O	X	
VASECTHX	X	X	O	O	O	X		X	X	X	O	O	O	X	X	X	O	O	
TCVAOPF*	X	X	X	X	X		X	X	O	X	X	X	O	X	X	X	X	X	
ALLERGHX	X	X	O	O		X	X	X	X	X	O	X	X	X	X	X	X	X	
GGANHX1*	X	O	O		O	X	O	X	X	X	O	O	O	X	X	X	O	O	
GGANHX	O	O		O	O	X	O	X	X	X	X	X	X	X	X	X	O	X	
TBHX*	O		O	O	O	X	X	X	X	X	O	O	O	X	X	X	X	X	
PNHX		O	O	O	X	X	O	X	X	X	X	X	X	X	X	X	X	X	

O: selected in the model. X: not selected in the model, GINFRQ: frequency of drinking liquor, BEERFRQ: frequency of drinking beer, CANF2: family history of cancer in paternal grandparents, uncles, aunts and cousins, GANF1: family history of liver diseases in parents and siblings, CVAOPF1: family history of hypertension in parents and siblings, ALCOHL: drinking history, SMOK400: smoking of more than 400 cigarettes, CHEMOT: history of chemotherapy, TCANF: family history of cancer, TSIMF: family history of cardiac diseases, TGANF: family history of liver diseases, VASECTHX: history of vasectomy, TCVAOPF: family history of hypertension, ALLERGHX: history of allergic diseases, GGANHX1: history of chronic liver diseases, GGANHX: history of acute liver diseases, TBHX: history of TB, PNHX: history of pneumonia.
* selected for reliability test.

Table 5. The selection of variables in logistic regression model with stepwise selection method in female

Tested as dependent variable	Tested as independent						
	GBS TON HX	TCV AOP F	TCV AF	TCA NF	CHE MOT	SMO K400	CVA OPF 1
CVAOPF1*	X	X	O	X	O	O	
SMOK400	O	X	O	O	X		X
CHEMOT	O	X	O	O		O	X
TCANF*	O	X	O		O	O	X
TCVAF	O	X		O	X	O	X
TCVAOPF*	O		O	O	X	O	X
GBSTONHX		X	O	O	O	O	X

O: selected in the model. X: not selected in the model, CVAOPF1: family history of hypertension in parents and siblings, SMOK400: smoking of more than 400 cigarettes, CHEMOT: history of chemotherapy, TCANF: family history of cancer, TCVAF: family history of stroke, TCVAOPF: family history of hypertension, GBSTONHX: history of GB stone.
* selected for reliability test.

여 공변수로 작용할 가능성이 있는 변수 각각을 종속변수로, 나머지 변수를 독립변수로 사용하여 단계적 선택법(stepwise selection method)을 사용하는 로짓회귀분석을 시행하였다. 로짓회귀분석의 단계적 선택법에 의해 선택된 독립변수는 종속변수의 변화에 대하여 유의한 영향을 미치는 것이므로 [7], 많은 수의 종속변수에 유의한 영향을 미치는 독립변수가 이 종속변수들을 대표하는 것으로 생

각하였다. 각 독립변수가 각각의 로짓 회귀분석 중 몇 개에서 선택되었는지를 비교하였다 (Table 4,5). 이 로짓 회귀분석 결과를 토대로 많은 분석에서 선택된 변수와(남자: GGANHX1, TCANF, 여자: TCANF), 종속변수로 사용되었을 때 독립변수가 하나도 선택되지 않아 다른 변수로 대표될 수 없는 변수들(남자: GANF1, CVAOPF1, TCVAOPF; 여자: TCVAOPF, CVAOPF1), 그리고 이들 변수에 의해서 대표될 수 없는 나머지 변수들을 반영할 수 있는 변수(남자: TBHX, TGANF)들을 일치도 분석 대상으로 최종 선정하였다. 남성에서는 총 18개의 공변수로 작용할 가능성이 있는 변수 중 7개, 여성에서는 7개의 공변수로 작용할 가능성이 있는 변수 중 3개의 변수가 대표 변수로 선택되었다. 선택된 변수는 모두 비연속 변수로, 연속 변수는 선택된 것이 없었다.

3. 신뢰도 분석

선정된 변수들에 대해서 반복 측정된 자료 간의 일치도를 평가하기 위해서 각 변수의 카파값을 구하였다. 신뢰성은 Altman [8]이 제시한 바에 따라 카파값이 1.0-0.8이면 매우 좋은 일치(very good), 0.8-0.6이면 좋은 일치(good), 0.6-0.4이면 중등도의 일치(moderate), 0.4-0.2이면 적당한 일치(fair), 0.2이하의 일치도가 낮은 것(poor)으로 판단하였다.

연구결과

남성에서 선택된 변수들의 일치도 평가는 Table 6에 제시되어 있다. '암의 가족력'은 카파값이 0.64로 가장 높았고, '결핵의 과거력'이 0.36으로 가장 낮았다. Altman이 제시한 바에 따르면, '암의 가족력'이 좋은 일치도를 보였고, '부모나 형제에서의 간질환의 가족력', '부모와 형제에서의 뇌혈관질환의 가족력', '간질환의 가족력', '고혈압의 가족력', '만성간질환의 과거력'이 중등도의 일치, '결핵의 과거력'이 적당한 일치도를 보였다.

여성에서 선택된 변수들의 일치도 평가는 Table 7에 제시되어 있다. '암의 가족력'은 카파값이 0.58로 가장 높았고, '고혈압의 가족력'이 카파값이 0.47로 가장 낮았다. Altman의 기준에 따르면, '부모나 형제에서의 고혈압의 가족력', '암의 가족력', '고혈압의 가족력' 모두 중등도의 일치도를 보였다.

고찰

코호트연구에서 연구성적을 분석할 때 결과변수에 영향을 주는 공변수들의 영향을 보정하기 위해 다변량분석법을 적용하는 것이 일반적이다. 따라서, 수집된 자료 중 노출변수와 결과변수뿐만 아니라 분석에 같이 포함되는 공변수들의 자료의 질을 평가하는 것이 필요하다. 그러나 이 전 연구들에서 노출자료를 수집하는 설문조사의 타당성과 신뢰성을 확인하는 연구 [9,10]는 여럿 있었으나 공변수들의 자료의 질을 평가한 경우는 거의 없었다.

본 연구는 원전주변지역 역학조사에서 대상자들에 대한 자료를 수집하기 위해 시행된 설문조사의 신뢰성을 확인하기 위해서 수행되었다. 이를 위해서 반복 측정된 자료 중 선택된 공변수들의 일치도를 평가하였다.

조사된 자료의 질을 평가하는 경우 전체 설문지의 신뢰성을 평가하거나 [11] 혹은 설문지로 조사된 변수 중 특정 변수를 임의로 선정하여 그 일치도를 평가하는 것이 일반적이다 [12-14]. 전자의 경우는 설문지 전체가 진단 또는 평가의 도구인 경우로 전체의 신뢰성을 평가하는 것이 타당할 것이나, 본 연구의 대상이 되는 설문조사 자료의 경우 노출변수와 결과변수의 연관성 분석에는 공변수로 작용할 가능성이 있는 자료만 사용될 것이므로 모든 변수의 신뢰성을 평가하는 것은 불필요하다. 후자의 경우 임의로 선정된 변수는 자료에 포함되어 있는 다른 변수의 신뢰성에 대해 아무런 정보를 제공하지 않는다는 단점이 있다.

따라서 본 연구에서는 노출변수와 결과변수 사이에 실제

Table 6. Reliability test of selected variables in male

Variable name	1st survey	2nd survey			Kappa
		No	Yes	Don't know	
GANF1	No	124	1	0	0.56
	Yes	2	2	0	
	Don't know	0	0	0	
CVAOPF1	No	105	6	0	0.51
	Yes	9	10	0	
	Don't know	0	0	0	
TCANF	No	71	4	19	0.64
	Yes	4	21	9	
	Don't know	3	4	56	
TGANF	No	136	2	20	0.50
	Yes	3	2	1	
	Don't know	3	3	18	
TCVAOPF	No	110	10	19	0.44
	Yes	12	11	3	
	Don't know	3	2	18	

		No	Yes, cured	Yes, not cured	Don't know	
GGANHX1	No	46	0	1	0	0.53
	Yes, cured	1	1	0	0	
	Yes, not cured	0	0	1	0	
	Don't know	42	1	0	95	
TBHX	No	148	0	0	22	0.36
	Yes, cured	4	6	0	3	
	Yes, not cured	0	0	0	1	
	Don't know	1	0	0	3	

GANF1: family history of liver diseases in parents and siblings, CVAOPF1: family history of hypertension in parents and siblings, TCANF: family history of cancer, TGANF: family history of liver diseases, TCVAOPF: family history of hypertension, GGANHX1: history of chronic liver diseases, TBHX: history of TB.

Table 7. Reliability test of selected variables in female

Variable name	1st survey	2nd survey			Kappa
		No	Yes	Don't know	
CVAOPF1	No	139	6	0	0.56
	Yes	8	11	0	
	Don't know	0	0	0	
TCANF	No	108	11	24	0.58
	Yes	7	29	17	
	Don't know	10	7	76	
TCVAOPF	No	150	10	32	0.47
	Yes	11	11	7	
	Don't know	16	2	50	

CVAOPF1: family history of hypertension in parents and siblings, TCANF: family history of cancer, TCVAOPF: family history of hypertension.

로 영향을 줄 가능성이 있으면서 다른 변수들의 신뢰도를 대표할 수 있는 변수를 찾아 일치도를 확인하였다. 공변수로 작용할 수 있는 몇 개의 변수 중 일치도를 평가할 최종 분석 대상 변수를 선정하기 위해서 본 연구에서는 이 변수들 사이에서 서로 간의 공변성이 가장 많은 변수를 골라 대

표적인 변수로 삼았다. 이 변수를 고르기 위해 종속 변수의 변화에 유의하게 영향을 주는 독립적인 변수를 선정하는 로짓 회귀분석의 단계적 선택법 [7]을 사용하여 각 로짓 회귀 모형에 선택되는 수가 많은 변수들을 선정하였다. 만약 분석 대상 변수들이 연속변수였다면 이 변수들간의 상관계수를 비교하여 대표적인 변수를 선택할 수 있었을 것이나, 분석 대상이 된 변수들은 모두 비연속변수로 이러한 방법을 사용할 수 없었다. 차선책으로 각각의 변수들의 연관성을 지표로 하여 변수를 선택하였다. 상관계수를 사용할 수 있는 연속변수의 경우와 비교할 때, 이 방법은 연관성의 방향이 변수들 간에 상호적이지 않고, 변수들의 관계가 정량적이지 아니라 정성적으로만 고려된다는 단점이 있다.

원전주변지역 역학조사연구 기반코호트의 결과변수인 암은 남녀의 생리적, 유전적 차이로 인해 여성에게서 주로 생기는 자궁암이나 유방암뿐만 아니라 다른 암종에서도 그 발생양상과 위험요인이 성별에 따라 다르다 [15,16]. 또한 설문조사의 일치도 역시 성별에 의해 영향을 받는다는 것이 보고된 바 있다 [2]. 따라서 본 연구에서는 남성과 여성을 나누어 따로 분석을 수행하였다. 실제로 공변수로 작용할 수 있는 변수들을 선정하기 위해 단변수분석을 수행한 결과 성별에 따라 공변수로 작용할 수 있는 변수들이 서로 다른 것을 확인할 수 있었다.

일치도를 분석한 결과, 남성과 여성의 경우 모두 암의 가족력이 가장 높은 일치도를 보였다. 이탈리아에서 수행된 연구에서는 면접에 의해 조사된 암의 가족력은 암의 발병 부위에 따라 카파값이 0.5~1.0이었고, 전체 암의 경우 카파값은 0.7로 본 연구의 결과와 비슷한 결과를 나타낸 바 있다 [17].

다른 질환의 가족력의 경우 대체로 중등도의 일치도를 보였다. 일본에서 행해진 한 연구에 의하면 고혈압의 가족력의 경우 카파값이 0.7, 뇌혈관질환의 경우 0.5로 보고된 바 있다 [18]. 국내의 경우 설문지로 조사된 가족력의 신뢰도가 비교적 높은 것으로 보고된 바 있다 [19]. 본 연구의 결과는 이러한 과거의 다른 연구 결과와 비슷한 경향을 나타낸 것으로 생각된다.

남성에서 결핵의 경우 과거력의 카파값이 0.36으로 비교적 낮은 값을 나타내었다. 일치도가 낮은 이유로 여러 가지 원인이 있을 수 있을 것이나 가장 의심되는 것은 두 번의 조사 사이에 대상자가 질문의 내용을 이해하는 데 차이가 있지 않았나 하는 것이다. 그 이유는 첫 번째 조사에서 '아니오' 또는 '예' 라고 대답한 사람이 두 번째 조사에서 '모름' 이라고 대답한 경우가 가장 많았기 때문이다. 조사 대상자가 질문의 내용을 잘못 이해하는 원인으로서는 질문의 구성이 애매하거나 면접을 수행하는 조사원의 태도가 다른 것

등을 생각할 수 있다 [20].

또 다른 가능성으로 첫 번째 조사 시기에는 결핵에 걸려 있지 않았으나 두 번째 조사가 수행되기 전에 병에 걸린 경우 이러한 실제 변화가 반영된 경우를 생각할 수 있다. 그러나 본 연구의 결과에서는 대상자의 응답이 '아니오' 또는 '모름' 에서 '예' 로 바뀐 대상자는 한 명도 없었으므로 두 조사 간의 기간 동안 결핵의 과거력에 생긴 실제 변화로 인해 일치도가 낮게 나타날 가능성은 없다. 두 번의 반복 조사 간의 시간이 대상자의 기억을 흐리게 하여 일치도가 낮게 나타났을 가능성은 있으나 [21] 두 조사 간의 시간이 1년 미만으로 기억의 부정확성이 일치도에 크게 영향을 주었을 것으로 생각되지 않는다.

본 연구가 가지는 제한점으로 분석대상 선정 과정에서 위음성이 생길 가능성을 들 수 있다. 즉, 잠재적인 공변수를 골라내는 과정에서 실제로는 결과변수와 연관성이 있으나 그 변수에 대한 자료의 신뢰도가 낮아 연관성이 없는 것으로 나타나 공변수가 아닌 것으로 잘못 분류되어 분석 대상에 포함되지 않는 변수들이 있을 수 있다는 점이 있다. 그러나 실제로 이런 이유로 탈락된 변수들이 있는지 확인하는 것은 전체 변수들의 신뢰도를 평가하기 전에는 불가능한 것으로 생각된다.

또 다른 한계점으로, Table 1에 제시된 바와 같이 반복조사 간의 기간이 1년 미만인 자료를 제공한 설문 대상자의 특성이 전체 코호트와는 다르다는 것이 있다. 그러나 반복조사 간의 시간이 길어질수록 과거 정보에 대한 기억이 흐려지는 정도가 더욱 커져서 자연스럽게 일치도가 낮아지는 경향이 있으므로 [21], 본 연구에서는 이러한 문제점을 극복하기 위해 이러한 한계점에도 불구하고 조사 대상을 반복조사 간의 기간이 1년 미만인 자료로 한정하였다.

본 연구에서 분석된 변수들은 원전주변지역 역학조사연구 기반코호트의 설문조사 자료 중 공변수로 작용할 수 있는 변수들로 기반코호트의 자료를 사용하여 노출변수와 결과변수의 관련성을 분석할 때 이들 변수가 보정되어야 할 것이다. 본 연구에서는 이들 변수 대부분이 중등도에서 좋은 일치도를 가지는 것을 확인하였다.

참고문헌

1. Ahn YO, Yoo KY, Park BJ, Kim DH, Bae JM, Kim DH, et al. *Epidemiology: The Principles and Applications*. Seoul: Seoul National University Press; 2005, p.169-170. (Korean)
2. Stein AD, Courval JM, Lederman RI, Shea S. Reproducibility of responses to telephone interviews: Demographic predictors of discordance in risk factor status. *Am J*

- Epidemiol* 1995; 141(11): 1097-1105.
3. Littman AJ, White E. Reliability and validity of self-reported male balding patterns for use in epidemiologic studies. *Ann Epidemiol* 2005; 15(10): 771-772.
 4. Post A, Gilljam H, Rosendahl I, Meurling L, Bremberg S, Galanti MR. Validity of self reports in a cohort of swedish adolescent smokers and smokeless tobacco (snus) users. *Tob Ccontrol* 2005; 14(2): 114-117.
 5. White E, Hunt JR, Casso D. Exposure measurement in cohort studies: The challenges of prospective data collection. *Epidemiologic Rev* 1998; 20(1): 43-56.
 6. Chaiky S. *Methods of Research in Social Science*. 2nd ed. Seoul: Hak Hyun Sa; 1994. (Korean)
 7. Wang D, Zhang W, Bakhai A. Comparison of bayesian model averaging and stepwise methods for model selection in logistic regression. *Stat Med* 2004; 23(22): 3451-3467.
 8. Altman DG. *Practical Statistics for Medical Research*. 1st ed. London: Chapman and Hall; 1991.
 9. Ahn Y, Lee JE, Cho NH, Shin C, Park C, Oh BS, et al. Validation and calibration of semi-quantative food frequency questionnaire. *Korean J Community Nutr* 2004; 9(2): 173-182. (Korean)
 10. Choi SW, Ju YS, Kim DS, Kim JY, Kwon HJ, Kang DH, et al. Reliability and validity of the korean version of isaac questionnaire. *Korean J Prev Med* 1998; 31(3): 361-371. (Korean)
 11. Brazier JE, Harper R, Jones NM, O' Cathain A, Thomas KJ, Usherwood T, et al. Validating the sf-36 health survey questionnaire: New outcome measure for primary care. *BMJ* 1992; 305(6846): 160-164.
 12. Kapp JM, Jackson-Thompson J, Petroski GF, Schootman M. Reliability of health-related quality-of-life indicators in cancer survivors from a population-based sample, 2005, BRFSS. *Public Health* 2009; 123(4): 321-325.
 13. Jee SH, Yoon JE, Won S, Kim S, Suh I. Test-retest reliability of a self-administered smoking questionnaire for middle and high school students. *Korean J Epidemiol* 2003; 25(1): 16-23. (Korean)
 14. Evenson KR, Eyster AA, Wilcox S, Thompson JL, Burke JE. Test-retest reliability of a questionnaire on physical activity and its correlates among women from diverse racial and ethnic groups. *Am J Prev Med* 2003; 25(3 Suppl 1): 15-22.
 15. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: A systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; 371(9612): 569-578.
 16. Belani CP, Marts S, Schiller J, Socinski MA. Women and lung cancer: Epidemiology, tumor biology, and emerging trends in clinical research. *Lung Cancer* 2007; 55(1): 15-23.
 17. Bravi F, Bosetti C, Negri E, Lagiou P, La Vecchia C. Family history of cancer provided by hospital controls was satisfactorily reliable. *J Clin Epidemiol* 2007; 60(2): 171-175.
 18. Saito T, Nanri S, Saito I. Reliability of family history of lifestyle-related diseases on questionnaire. *Pediatr Int* 2009; 51(4): 514-519.
 19. Kim CY, Huh BY. Reliability of a self-administered health questionnaire-a survey in a male college students group-. *J Korean Public Health Assoc* 1989; 15(2): 45-54. (Korean)
 20. Beckett M, Weinstein M, Goldman N, Yu-Hsuan L. Do health interview surveys yield reliable data on chronic illness among older respondents? *Am J Epidemiol* 2000; 151(3): 315-323.
 21. Zhu S, Toyoshima H, Kondo T, Tamakoshi K, Yatsuya H, Hori Y, et al. Short- and long-term reliability of information on previous illness and family history as compared with that on smoking and drinking habits in questionnaire surveys. *J Epidemiol* 2002; 12(2): 120-125.