

데이터 마이닝을 이용한 서울시교직원의 피로요인 탐색연구

이 회 우* · 신 선 미**

* 서울특별시 학교보건진흥원

** 연세대학교 간호정책연구소

An Exploratory Study of Fatigue Related Factors among School Personnel in Seoul by Data mining

Hee Woo Lee*, Sun Mi Shin**

* *Seoul School Health Promotion Center*

** *Nursing Policy Institute, Yonsei Univ. College Of Nursing*

ABSTRACT

Purpose : To identify general characteristics of school personnel with recent fatigue which was the most frequent symptom among subjective symptoms and to explore fatigue-related factors by evaluating physical and perceived health status, life style, and symptoms through data mining techniques.

Methods : We collected a data of the 1,147(male 545, female 602) who were elementary, middle, or high school personnel, answered a questionnaire, and received physical examination in Seoul School Health Center from September to November in 2000. And we investigated the differences between fatigue group and non-fatigue group for demographic characteristics, physical health status, perceived health status, symptoms, and laboratory values by frequency, chi-square test, t-test, or simple logistic regression analysis by SAS package 8.1, and then selected significant variables as input variables of a decision tree analysis of CART model by SAS E-miner.

Results : In general characteristics, the fatigue consisted of 41.1%(male 35.2%, female 46.4%) among 1,147 school personnel. In classical statistics, factors related with fatigue were female, lower means of systolic and diastolic pressure, young age, personnel in middle school, irregular eating habit, no exercise a week or less than 30minutes exercise a day, perception of unhealthy status, and subjective symptoms including short of breath at exercise. In simple logistic regression to examine the relationship between selected independent variables and fatigue as a dependent variable, the odds ratio of gender (female vs male) was 1.58 times, and young age (20s vs 60s) 20.67 times, and middle vs high school personnel 1.86 times.

However, we mined combined several characteristics by SAS-E miner. In CART model, if health perception was healthy, and age was ≥ 37.5 years, the proportion of the fatigue was only 19.3%, but if health perception was not healthy and symptom was severe 'short of breath' during exercise and age was < 53.5 years, and BMI was ≥ 22.69 , the proportion of the fatigue was up to 84.8%.

Conclusions : The fatigue consisted of 41.1%(male 35.2%, female 46.4%). In classical statistics, fatigue-related factors among school personnel were young age, female gender, perceived unhealthy status, subjective physical symptoms, poor life-style, and lower blood pressure rather than only physical health status. However, in data mining, if health perception was healthy and age was ≥ 37.5 years, the proportion of the fatigue was only 19.3%, but if health perception was not healthy and symptom was severe 'short of breath' during exercise and age was < 53.5 years, and BMI was ≥ 22.69 , the proportion of the fatigue was up to 84.8%.

Key words : fatigue, data mining, school personnel

(2006년 4월 29일 접수, 2006년 6월 19일 채택)

교신저자 : 신선미, 연세대학교 간호대학 간호정책 연구소, 120-749 서울시 서대문구 신촌동 134

E-mail : healthteam@hanmail.net

1. 서론

1. 연구 배경

초·중·고에 근무하는 교직원의 40.4%가 이유 없는 '피로'를 경험하고 있다. 2000년도 건강보험 관리공단이 실시한 공무원 및 사립학교 교직원 신체검사 결과를 분석해보니, 불편을 느끼는 신체증상 및 증세에서 교직원의 40.4%가 이유 없이 '피로'하다고 응답하였다(이희우 외, 2001). 그러므로 교직원이 경험하고 있는 '피로'가 신체적 건강문제로 인한 것인지, 직업 구조적 요인인지 또는 주관적 건강인지와 관련된 것인지에 대한 탐색이 필요할 것이다.

임상현장에서 환자들로부터 빈번하게 듣게 되는 '피로'는 미국내 1차 진료에서 7번째로 흔한 증상으로(Rhoten, 1982) 전체 미국 인구의 14%-20%가 '피로'를 호소하고(Chen, 1986) 있으며, 일하는 네덜란드의 인구에서는 22%가 '피로'를 호소하였다(Bultmann et al., 2002). 우리나라에서도 건강검진으로 내원한 사람 중 43%가 만성적인 '피로'를 호소한다 하였고, 또 뚜렷한 이유 없이 심한 '피로'가 6개월 이상 지속되는 만성피로는 아직 역학적으로 많은 부분 설명되지 않아 논란이 많은 증후군이다(천병철, 2001).

'피로'는 정신이나 육체의 지나친 활동으로 야기되는 지친 상태를 의미하는 것(Gilber, 1971)으로 신체적 질병이나 치료와 연관되어 발생하기도 하며, 간혹 몇가지 비정상 상태에 대한 첫 번째 적응증상으로 나타날 수 있다(박기흠, 1996). Grandjean et al.,(1970)은 '피로'는 운동이나 정신적 기술의 감소, 증가된 업무의 회피와 관련된 감각에 대한 주관적 평가이며, '피로'에 대한 경험은 주관적인 피곤함(tiredness)에서 완전히 기진맥진(Exhaustion)한 상태로 가는 연속상의 한 단계로 보았다.

흔히 임상에서는 심근경색, 골연화, 우울, 바이러스 감염, 당뇨병, 만성폐쇄성질환, 만성 신부전, 빈혈, 수술 및 암등의 다양한 질병에서 '피로'가 나타나고 있으며(Hart, 1978), Potmpa et al., (1986)은 '피로'를 신체와 정신적 요인의 복잡한 상호작용에 의해 발생하는 지각으로 정의하였으며, O'conner et al.,(1989)는 '피로'에 관련된 3가지 영역을 객관적인 일의 감소, 신체느낌에 대한 주관적인 평가, 신체기전의 생리적 변화로 정의하였다.

'피로'가 축적됨으로서 초래되는 문제점은 일상생활 능력의 약화와 기분(mood)의 저하가 있으며, 심한 경우는 치료를 연기하거나 중단하게 되어 삶의 질을 저하시키게 된다(Fieler, 1991). 그럼에도 불구하고 '피로' 현상은 대단히 복잡할 뿐 아니라 주관적 현상으로 측정되어야 할 변수의 질과 양이 너무 광범위하여 그 종합적 측정에 어려움이 있으며(김돈균, 1976), 우리나라에서는 노동 작업능률과 관계된 산업피로와 근육피로에 초점을 둔 연구가 주로 행하여 지거나, 또는 일부 임상환자를 대상으로 피로의 요인을 알기 위해 시도되고 있다. 이처럼 '피로'측정의 광범위성을 보완하기 위해서는 여러 특성이 결합된 분류규칙을 알 수 있는 데이터 마이닝 기법이 '피로'요인 탐색에 좋은 방법이 될 것이라고 생각한다.

본 연구에서 '피로'요인의 규명하기 위한 데이터 마이닝은 대상자의 다양한 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하여 모험화 함으로서 유용한 지식을 추출하는 일련의 과정이라고 정의할 수 있다. 또한 통계적 관점에서 간략하게 표현한다면 데이터 베이스 속에서 의미 있는 정보를 발견하기 위한 탐색적 분석이라고 할 수도 있으며, 의사결정을 위한 응용프로그램으로 데이터 내의 패턴 및 상관관계를 자동적으로 발견해 내는 과정이라 할 수 있다. 그러나 데이터 마이닝은 다소 새로운 분야이고 지금도 발전되어 가는 과정에 있기 때문에 그 정의를 한마디로 요약하기는 어렵다(강현철과 최종후, 1999).

그러므로 본 연구에서는 교직원이 지각하는 건강문제 중 많은 범주를 차지하고 있는 '피로'의 현황 및 일반적 특성을 일반통계를 통하여 파악한 후, 지식추출기법인 데이터 마이닝 기법을 활용하여 일반통계에서 알 수 없었지만 잠재적으로 유용성이 있는 '피로' 대상자의 특성 및 관계를 탐색하고자 한다.

2. 연구의 목적

교직원이 경험하고 있는 신체 증상 중 가장 많은 범주를 차지하고 있는 '피로'의 현황 및 일반적 특성을 파악한 후, 새로운 지식추출 기법인 데이터 마이닝 기법을 활용하여 '피로'와 건강상태와의 관계를 파악하고자 한다. 구체적인 연구목적은 다음과 같다.

첫째, '피로'를 느끼는 교직원과 느끼지 않는 교직원의 신체건강상태를 비교하여 파악한다.

둘째, 로지스틱 회귀분석을 통해 일반적 특성, 주관적 건강상태 인지, 건강행위에서의 '피로'변인을 파악한다.

셋째, '피로'를 경험하는 교직원의 특성파악과 예측을 위한 최적의 의사결정나무 모형을 구축하기 위해 로지스틱 회귀모형, C4.5, CHAID, CART의 모형의 안정성과 예측력을 평가한다.

넷째, 모형평가 결과 가장 정확도가 높은 의사결정나무 모형을 이용하여 '피로'경험 교직원의 특성을 파악한다.

II. 연구 방법

1. 연구 설계

'피로'를 경험하는 교직원의 현황을 파악하고, '피로'경험 교직원과 경험하지 않는 교직원의 일반적 특성을 비교한 후, 데이터 마이닝 기법을 이용하여 '피로'경험교직원의 분류특성을 파악하기 위한 탐색연구이다.

2. 연구 대상

본 연구의 모집단은 초·중·고등학교에 근무하는 남·여 교직원들이고, 근접 모집단으로서 20세 이상 65세 이하의 서울시 남·여 교직원으로 선정하였다. 이들 중 표집대상은 서울시 학교보건원에서 2000년도 9월부터 11월 사이에 교직원 건강진단을 받은 초·중·고등학교 교직원 총 1,147명을 연구대상으로 하였다.

3. 연구 자료

연구대상자 1,147명을 대상으로 조사된 신체검사 결과와 자가 응답 설문자료를 이용하여 데이터 베이스를 구축한 후, 본 연구의 목적에 맞게 자료 변환을 거쳐 분석항목을 선정하였다.

4. 분석 항목

분석항목은 인구학적 요인(성별, 학교종류, 연령 등 3문항), 신체건강상태(신체체질량지수, 혈압, 혈색소, 혈당, 총콜레스테롤, 뇨당, 뇨단백, 뇨잠혈 등

8문항) 현재건강상태인지(1문항), 지각된 신체증상 및 증세(18 문항), 흡연(1문항), 음주(1문항), 식생활(1문항), 운동습관(1문항) 등 총 34문항으로 구성되었다.

분석항목 중 종속변수는 '피로'로 '최근 특별한 이유 없이 푹 쉬고 싶을 정도로 '피로'한 적이 있다'라는 질문에 '예'라고 응답한 교직원을 '피로'군으로 분류하였으며, '아니오'라고 응답한 교직원을 '비피로'군으로 정의하였다.

5. 분석방법

'피로'를 인지하는 교직원의 현황과 관련 요인의 파악을 위해 SAS package와 SAS E-miner를 이용하여 다음과 같이 분석하였다.

1) '피로'군 교직원과 '비피로'군 교직원의 분포와 신체건강상태의 비교는 빈도와 백분율, Chi-square test, t-test를 통해 알아보았다.

2) 건강행위, 주관적 건강상태와 '피로'와의 관계를 알고, '피로' 모형투입변수를 선정하기 위해 단순 로지스틱 회귀분석을 이용하였다.

3) 데이터 마이닝 기법을 이용한 '피로' 분류 모형 구축에 앞서, Cross-validation 평가방법을 이용하여 다중 Logistic 회귀모형과 데이터 마이닝 모형 중 CHAID, C4.5, CART모형의 안정성과 예측력을 평가하였다.

4) 그 중 가장 정확률이 높은 CART 모형을 선택하여 '피로'를 호소하는 교직원의 분류모형을 구축 후 '피로'를 경험하는 교직원의 분류특성을 알아보았다.

III. 연구 결과

1. 연구대상자의 일반적 특성에 따른 '피로'의 현황

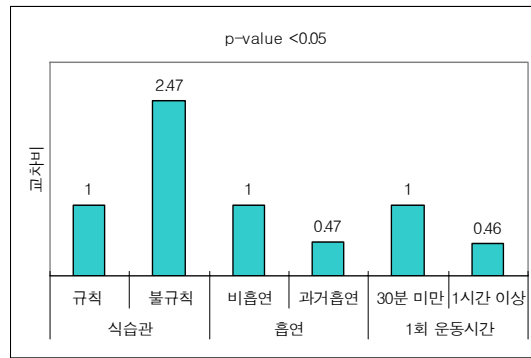
전체대상자 1,147명 중 '피로'의 증상을 가진 교직원은 남자35.2%, 여자는 46.4%로 여자교직원이 '피로'를 더욱 많이 경험하고 있었고($p < 0.001$), 나이별로 볼 때 20대 50.9%, 30대 53.6%, 40대 40.3%, 50대 31.3%, 60대 4.8%로 젊은 교직원에서 '피로'의 증상이 많았다($p < 0.001$). 근무학교별로 볼 때 초등학교 교직원 42.0%, 중학교 50.5%, 고등학교 교직원 35.4%가 '피로'의 증상을 가지고 있어 중학교 교직원이 가장 많이 '피로'를 자각하고 있었다($p < 0.001$) (Table 1).

〈Table 1〉 대상자의 일반적 특성

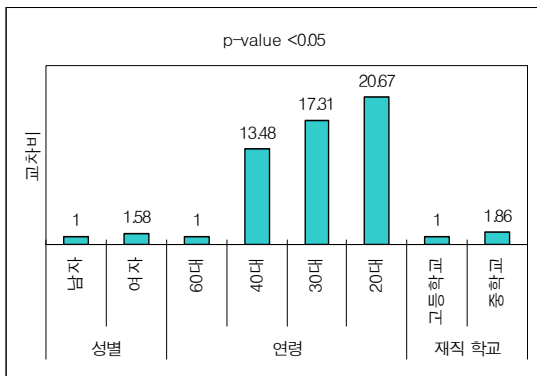
변수	구분	'비피로'군(%)	'피로'군(%)	χ^2	p-value
성별	남자	353 (64.8)	192 (35.2)	14.6	<0.001
	여자	323 (53.7)	279 (46.4)		
	계	676 (58.9)	471 (41.1)		
나이별	20대	57 (49.1)	59 (50.9)	29.8	<0.001
	30대	211 (53.6)	183 (46.5)		
	40대	234 (59.7)	158 (40.3)		
	50대	154 (68.8)	70 (31.3)		
	60대	20 (95.2)	1 (4.8)		
	계	676 (58.9)	471 (41.1)		
초·중·고별	초등학교	196 (58.0)	142 (42.0)	17.5	<0.001
	중학교	139 (49.5)	142 (50.5)		
	고등학교	341 (64.6)	187 (35.4)		
	계	676 (58.9)	471 (41.1)		

2. 신체건강상태와 '피로'의 분포

신체건강상태에 따른 '피로'의 분포를 알기위해 교직원의 신체검사 결과들 즉 신체체질량 지수, 혈압, 혈색소, 혈당, 총콜레스테롤, 뇨당, 뇨단백, 뇨잠혈 등의 결과와 '피로'와의 관련성을 살펴보았다. 그 결과, p-value 0.05수준에서 유의한 관련성은 수축기 혈압만이 '피로'군에서 '비피로'군보다 낮은 양상을 보였고 (남 p=0.01, 여 p=0.05) 기타 다른 신체검사의 결과는 '피로'군과 '비피로'군과의 통계학적 차이가 없었다. 이처럼 교직원이 지각하는 '피로'는 신체건강상태와 관련성은 크지 않음을 알 수 있다.



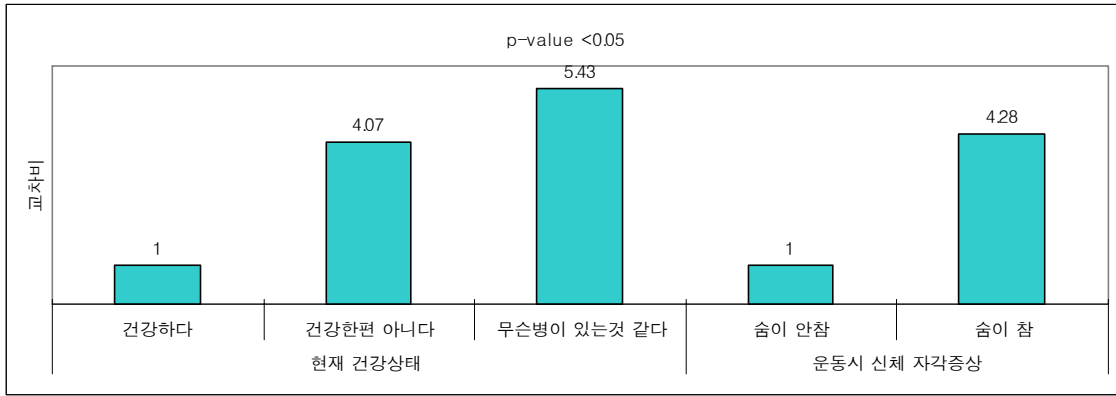
[Figure 2] 건강행위에서의 '피로' 관련 요인



[Figure 1] 일반적 특성에서의 '피로' 관련 요인

3. 일반적 특성, 건강행위, 현재건강상태 인지에서의 '피로'관련 변인

신체건강상태 외 교직원의 '피로'와 관련 있는 다른 변수들을 알기위해 단순 로지스틱 회귀분석에 의해 일반적 특성과 '피로', 건강행위와 '피로', 현재건강상태인지와 '피로'의 교차비를 구하였다. 그 결과, 남자에 비해 여자는 1.6배(p<0.001)이었고, 60대에 비해 20대는 20.7배(p< 0.001), 30대는 17.3배(p=0.01), 40대는 13.5배(p=0.04)로 '피로'를 자각하고 있었다. 또 고등학교 교직원에 비해 중학교 교직원은 1.9배(p<0.001), 현재 '건강하다'고 생각하는 교직원에 비해 '건강한 편은 아니다'라고 생각하는 교직원은 4.1배



[Figure 3] 건강상태인지에서의 '피로' 관련 요인

($p=0.01$), '무슨 병이 있는 것 같다'라고 응답한 교직원은 5.4배($p<0.01$)로 '피로'를 많이 호소하고 있었다 (Figure 1, 3).

식사습관이 규칙적인 교직원에 비해 시간이 불규칙한 경우 2.5배($p=0.02$)로 '피로'를 느끼고 있었고, 담배를 '(거의)피우지 않는다'는 교직원에게 비해 '과거에 피웠으나 지금은 끊었다'는 교직원은 '피로'를 0.5배($p=0.01$)로 '피로'를 덜 느끼고 있었다. 또 1회 운동시간이 '30분 미만'이라는 교직원에게 비해 '1시간~2시간미만'은 0.5배($p=0.05$) '피로'를 느끼고 있었고, 운동시 숨이 찬다는 교직원은 그렇지 않은 교직원보다 4.3배로 '피로'를 자각하고 있었다(Figure 2, 3).

4. 데이터 마이닝을 이용한 '피로' 모형의 투입 변수

모형평가에서 종속변수는 '피로'이고 독립변수는 신체건강 변수 와 단순 로지스틱 회귀분석에서 유의한 변수 등 이었다.

5. 모형 평가

실제로 모형 분석을 수행함에 있어서 여러가지 기준과 규칙을 어떻게 지정하느냐에 따라서 서로 다른 의사결정나무를 얻게 된다(강현철 등, 1998).

그러므로 본 연구에서는 모형의 평가에 있어 각각의 모형 모두 사후확률¹⁾이 0.5²⁾로 동일하다고 가정하였다. 로지스틱 회귀모형에서는 입력변수를 다중으

로 투입하였고, 의사결정나무분석에서는 각 모형의 가지분리기준을 각 모형 모두 Default로 지정하여 동일한 정지기준을 설정하였다. 그 이유는 대부분의 경우에 디폴트 옵션들은 훌륭하게 작동하기 때문이며(강현철 등, 1998), 동일한 조건에서 모형을 비교하고 검토하여 가장 적절한 의사결정나무를 얻을 수 있어야 하기 때문이다. 즉 끝마디에 포함될 관측개체의 최소개수는 1, 지정된 분리기준에 이해 부모마디가 자식마디로 분리되기 위해 요구되는 관측개체의 수³⁾는 11로 이수보다 부족하면 더 이상 분리가 일어나지 않는다. 자식마디가 형성될 때 최대의 분리개수는 이지 분리를 기본으로 하였다. 마지막으로 뿌리마디로부터 끝마디까지의 깊이는 7로 이수보다 큰 경우엔 더 이상 분리를 수행하지 않았다.

지나치게 많은 마디를 갖는 의사결정나무는 새로운 자료에 적용시 예측오차가 매우 클 가능성이 있기 때문에 적절한 부나무(Subtree)를 선택하기 위해 나무형성시 최적의 평가 타당도를 갖는 부나무를 형성하였다.

또 이런 모형의 평가는 안정성과 예측력을 위해 연구대상자 1,147명중 75%를 분석용(Train), 25%를 평가용(Validation)으로 무작위 샘플링에 의해 분할한 후 평가하였고, 이런 과정을 5회 반복(Five fold cross-validation)하여 각 자료의 정분류율⁴⁾, 민감도⁵⁾, 특이도⁶⁾는 오분류 행렬을 기초로 하여 계산한

1) 각 교직원이 '피로'군이 될 확률

2) 일반적으로 분류의 기준으로 삼는 사후확률의 경계는 '1/목표범주의 범주개수'이다.

3) 데이터의 전체관측개체의 수 ÷100

4) 정분류율 : (관찰=1∩예측=1)의 빈도+(관찰=2∩예측=2)의 빈도/전체빈도

〈Table 2〉 5-Fold cross validations의 평균

구분	CHAID		C4.5		CART		Logistic regression	
	분석용	평가용	분석용	평가용	분석용	평가용	분석용	평가용
정분류율	71.56	67.67	73.86	68.22	75.33	68.43	69.60	68.50
민감도	67.56	61.74	63.89	57.21	65.37	56.18	55.00	52.66
특이도	74.39	71.88	80.68	76.37	82.24	77.11	79.70	79.96

후 5회의 평균을 구하였다. 그 결과 정분류율과 민감도가 가장 좋은 모형은 CART모형이었다(Table 2).

6. CART알고리즘에 의한 의사결정 나무 모형

Cross-validation결과 모형의 정분류율이 가장 높은 모형은 CART모형이었다. 이에 본 연구에서는 CART모형의 의사결정나무 모형을 제시하였다.

제시된 모형의 구축은 모두 모형 평가과정과 동일하며, 모형구축의 효율성을 위해 모형 평가 때와 역시 마찬가지로 분석용(Train set)으로 860명(75%), 평가용 데이터(Validation set)로 287명(25%)을 분할하였다. 그러므로 마디내 왼쪽 값은 분석용 데이터에 대한 결과이며, 오른쪽 값은 평가용 데이터에 대한 결과이고, 마디내 2는 '피로'군, 1은 '비피로'군을 의미한다.

분석용 860명중 '피로'군은 348명으로 40.5%였다. 이중 자신의 건강인지상태를 묻는 질문에서 '건강하다'라고 응답한 교직원의 '피로'율은 26.9%로서 '건강하지 않다'라고 응답한 교직원의 '피로'율 56.1%보다 낮았다(Figure 4).

'건강하다'라고 응답한 교직원은 다시 '나이 37.5세'를 기준으로 가치가 분류되었는데 37.5세보다 작은 경우의 교직원의 '피로'율은 41.9%이었고, 나이가 37.5세 이상일 때 '피로'율은 19.3%로서 '피로'율이 37.5세를 기준으로 급격히 낮아짐을 알 수 있었다.

건강하다 생각하면서 나이가 37.5세 미만의 마디에서는 하루 운동시간을 기준으로 하여 새로운 가치가 생성되었는데 '하루 30분 미만 운동시' '피로'군은 57.7% 인 반면, '운동을 30분 이상 할 때 34.0%가 '피로'군임을 알 수 있었다. '하루 30분 이상 운동'하는

마디에서는 콜레스테롤 156.5mg%를 기준으로 그 156.5mg% 이상은 26.5%가 '피로'를 인지하고 있었고, 156.5mg% 미만일 때 65%가 '피로'를 인지함으로써 콜레스테롤이 낮을 때 더 '피로'를 자각함을 알 수 있었다. 또 콜레스테롤이 156.5mg% 이상인 경우 음식을 짜게 먹는 편이라고 응답한 교직원의 '피로'율은 53.3%인 반면, 싱겁게 먹거나 보통이라고 응답한 교직원은 20.6%가 '피로'를 인지하고 있었다.

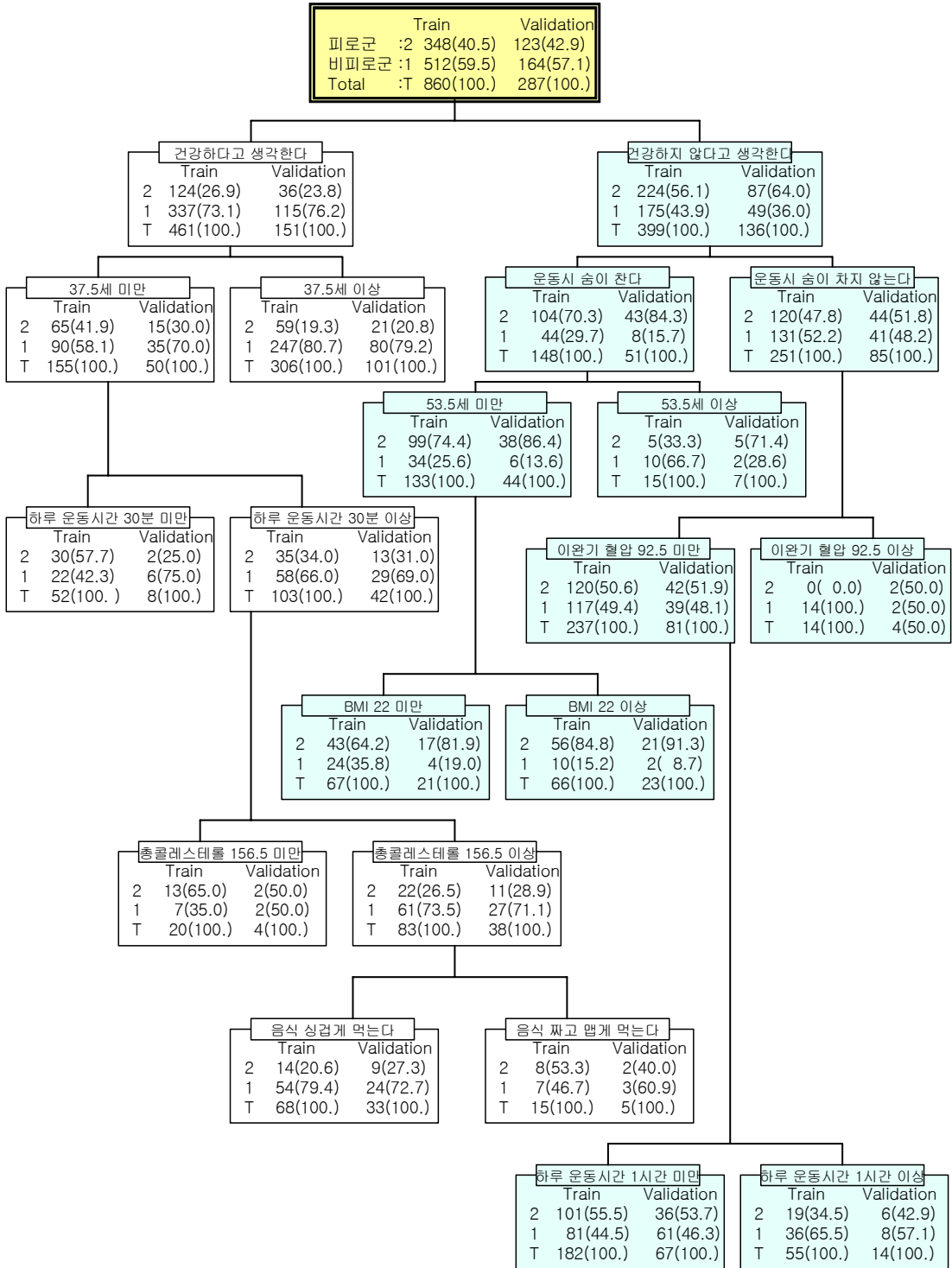
가장 첫번째 분류 노드에서 건강하다고 생각하지 않는다고 응답한 교직원의 분류가지를 따라 내려가 보면, 운동시 숨이 찬다고 응답한 교직원의 '피로'율은 70.3%인 반면, 아니오라고 응답한 사람은 47.8%만이 '피로'를 인지하고 있었다.

건강하지 않다고 생각하면서 운동시 숨이 찬다고 응답한 대상자의 '피로'율은 70.3%으로 그렇지 않은 교직원의 '피로'율 47.8%보다 '피로'인 지율이 높음을 알 수 있다. 또 운동시 숨이 찬다고 응답한 마디에서는 나이가 53.5세를 기준으로 새로운 마디가 형성되었는데 53.5세 미만일 때 '피로'율은 74.4%인 반면, 나이가 53.5세 이상일 때 '피로'율은 33.3%이었다. 또 건강하지 않으면서 운동시 숨이 차면서, 나이가 53.5세 미만인 노드에서는 BMI 22를 기준으로 새로운 노드가 형성되었는데 BMI가 22 미만일 때 64.2%가 '피로'군인 반면, BMI가 22이상일 때 '피로'군은 84.8% 있었다.

건강하지 않다고 생각하면서, 운동시 숨이 찬다라는 질문에서 '아니오'라고 응답한 교직원 노드의 분류가지를 따라가 보면 이완기혈압 92.5mmHg를 기준으로 그 미만인 경우 50.6%가 '피로'군임을 알 수 있었다. 이완기혈압이 92.5mmHg 미만의 마디에서는 하루운동을 1시간 이상 한다는 교직원의 '피로'율이 34.5%인 반면, 하루 운동시간이 1시간 미만인 경우의 '피로'율은 55.5%이었다.

5) 민감도 : (관찰=1∩예측=1)의 빈도/(관찰=1)의 빈도

6) 특이도 : (관찰=2∩예측=2)의 빈도/(관찰=2)의 빈도



[Figure 4] '피로'의 의사결정나무

IV. 고찰

본 연구에서 '피로'를 호소하는 교직원은 41.1%로서 검진결과 자료에서 43%가 만성적인 '피로'를 호소한다는 결과와 유사함을 알 수 있었다(2001, 천병철). 본 연구에서 남자 교직원의 '피로'율은 35.2%, 여자교직원의 '피로'율은 46.4%이었고, 남자에 대한 여자의 교차비는 1.6배 높음을 알 수 있었다. 이런 결과는 '피로'에 대한 선행연구 중 미국 여성이 남성보다 더 '피로'를 호소하였고(Torres-Harding et al., 2002), 장세진(2001)의 근로자를 대상으로 한 연구에서 남자보다 여자가 '피로'율이 높다는 것과 같다.

본 연구에서 일반통계를 통해 알 수 있었던 '피로' 관련 변인을 살펴보면 성별, 나이, 초·중·고등학교별 근무지, 현재 건강상태 인지, 식습관, 흡연, 운동, 낮은 수축기 혈압 등 있었는데, 이런 결과는 선행연구 결과와 다소 유사한 것도 있고 다소 차이가 있는 것도 있다.

나이와 '피로'의 관계를 살펴보면 정은리(1983)의 연구에서 연령이 낮을수록 '피로' 호소의 빈도가 높다는 것, 김윤주(1984)의 연구에서 교직경력에 따라 높아질수록 '피로'율이 점차로 낮아지는 경향이 있다는 것, 장세진(2001)의 연구에서 고연령보다는 20대와 30대에서 '피로'수준이 높다고 하였는데, 본 연구에서도 60대에 비해 20대의 교차비가 20.7배로 매우 높음을 알 수 있었다.

초·중·고등학교별 '피로'를 살펴보면, 본 연구결과 남자교직원은 중학교 교직원이 가장 '피로'율이 높음을 알 수 있었고($p=0.03$), 여자교직원에서도 중학교 교직원이 '피로'율이 높은 경향이 있었다($p=0.06$). 이런 결과는 김윤주(1984)의 연구에서 중학교 교사가 가장 높은 '피로' 호소를 하고 있었던 것과 같은 결과임을 알 수 있었다.

신체건강상태와 '피로'의 관계를 살펴보면, 남·녀 모두 '피로'군의 수축기 혈압은 '비피로'군보다 낮았고 통계학적으로 유의함을 알 수 있었다.

식생활과 '피로'의 관계를 살펴보면, 평소 식습관이 남녀 모두 '규칙적이다'에 비해 '거의 매일 횡수 시간 불규칙하다'라고 응답한 대상자의 교차비는 2.5배로 높음을 알 수 있었다. 이 결과는 정원훈과 신미경(1998)의 연구에서 결식집단이 정상 식이집단보다 높은 '피로'증가를 보인다는 것과 유사한 결과임을 알

수 있었다. 흡연력과 '피로'의 관계에서는 '담배를 피우지 않는 교직원'에 비해 '과거에 피웠으나 현재 끊은' 교직원의 교차비는 0.5배로 낮음을 알 수 있었다. 운동 습관과 '피로'의 관계를 살펴보면, 1회 운동시간이 '30분 미만'이라는 교직원에 비해 '1시간~2시간 미만'은 0.5배($p=0.042$) '피로'를 덜 자각함을 알 수 있었다.

이처럼 일반통계를 통해 알 수 있었던 '피로'의 관련변인은 신체건강상태 중에서는 단지 낮은 수축기혈압과 관련이 있고, 대부분은 일반적 특성이나 건강습관, 기타 주관적 증상 등이었다. 그러므로 '피로'는 신체건강상태를 나타내는 지표는 아님을 알 수 있었고 이는 Potmpa et al.,(1986)과 O'conner et al.,(1989) 등이 밝힌 신체건강상태와의 관련성과는 다소 다른 결과이었는데 이는 본 연구에서는 임상의 환자가 아니라 비교적 건강한 인구집단을 연구대상으로 하였기에 선행연구와 직접비교하는 것은 타당하지 않을 수 있을 것이다.

한편, 데이터 마이닝 분석을 통해 특정 특성을 가진 대상자의 '피로' 관련요인을 좀더 자세히 알 수 있었다. 즉 건강하다고 인지하면서, 나이가 37.5세 이상인 경우는 피로가 단지 19.3%이었으나, 건강하지 않다고 생각하면서, 운동시 숨이 심하게 차고, 나이는 53.5세 미만이며, BMI가 22.7이상인 경우 피로가 84.8% 까지 높아졌다. 또 건강인지군인면서, 나이 37.5세 미만이며, 하루 운동을 30분이상을 하고, 콜레스테롤이 156.5mg/%미만인 특성을 가진 대상자군에서는 는 피로가 65%로 같은 조건에서 단지 콜레스테롤이 156.5mg/%이상인 대상자의 피로율인 28.9%보다 높음을 알 수 있다. 또 동질한 특성의 대상자군에서 콜레스테롤이 156.5mg/%이상인 경우는 음식을 짜게 먹을때 피로가 53.3%이나 싱겁게 먹을때 20.6%에 불과하였다.

이처럼 데이터 마이닝의 의사결정나무분석을 통해 콜레스테롤, 낮은 이완기 혈압, BMI 등은 동질한 특성을 가진 어떤 집단에서 피로와 관련이 있음을 알 수 있었다. 이는 데이터 마이닝은 일반통계에서 통계학적으로 유의하지 않았던 정보를 버리지 않고 특정 특성을 가진 대상자의 정보를 가급적 많이 사용하여 분류를 보여준다는 장점이 있기 때문이다(Buja et al., 1999; Friedman & Fesher, 1997; 신선미, 2002). 특

히 의사결정나무 분석은 시각적 효과로 해석이 용이하여 연구결과를 이해하고 이를 이용하려는 독자의 의사결정을 도울 수 있다는 장점이 있으며, 동일한 분류기준으로 분석하였을 때 전통적 통계에서 많이 사용하고 있는 로지스틱 회귀모형보다 '피로'분류에 있어 정확률이 높았다. 그 이유는 로지스틱 회귀모형은 변수들간의 관계가 복잡한 비선형성(nonlinearity)을 가지는 경우에는 예측의 유용성 측면에서는 한계가 있고, 각 입력변수의 영향이 다른 입력변수에 종속되어 있지 않다고 가정하고 있어 유용한 교호작용(interaction)을 탐색하는 것이 실제적으로 어렵기 때문인 반면(강현철 등, 2000; 신선미 등, 2002), 의사결정나무분석은 여러 특성이 결합된 규칙을 도출할 수 있다는 장점을 가지고 있기 때문이다.

그러나 데이터 마이닝의 의사결정나무분석은 분리 기준, 정지규칙, 평가기준 등을 어떻게 지정하느냐에 따라서 서로 다른 나무가 형성되어 실제로 단 한번만에 분석이 끝나지 않아 연구자는 필요한 과정을 반복적으로 수행하여 실제 적용시의 문제점을 비교 검토하여야 하는 통계적 통찰력이 요구되었다.

또 데이터 마이닝 분석에 있어 생각해야 할 조건은 다른 어떤 통계방식보다 사전계획과 양질의 데이터 준비가 중요하다는 것이다. 오류가 있는 데이터에는 아무리 훌륭한 방법을 적용한다고 해도 얻어진 결과가 만족스럽지 않을 수 있으며, 데이터 정제하는 시간과 경비가 전통적 통계방식보다 더욱 많이 들기 때문이다(김지은, 2001). 이런 면에서 볼 때 본 연구자료의 제한점은 '피로'를 규명하기 위한 데이터 마이닝을 염두에 두고 계획된 데이터가 아니어서 '피로'와 관련성이 있을 법한 우울, 사회적 지지, 근무처 환경 및 조직문화, 직무만족도등 등과 '피로'와의 관계를 분석할 수 없었다는 점을 들 수 있을 것이다. 그러나 다른 인구집단에 비해 교직원은 성실한 응답을 하는 집단특성을 가지고 있으며, 결측치도 거의 없어 데이터의 질은 높다고 생각할 수 있다. 또 본연구의 대상자는 몇몇 학교의 교직원만을 대상으로 한 것이 아니라 서울시 학교보건원에서 수검 받은 대상자 전수를 대상으로 하여 서울시 교직원을 대표할 수 있다 할 수 있으나, 그러나 전국적 규모의 자료가 아니기에 우리나라 전체 교직원에게 일반화하는데 다소 어려움이 있을 것이다.

그러나 본 연구는 교직원의 '피로' 변인을 신체적 건강상태를 포함하여 구체적으로 알아보았고, 학교보건연구에서 아직은 자주 사용되지 않는 데이터 마이닝 활용을 시도하여 여러 특성이 결합된 규칙을 도출하였으며, 전통적 통계방법과 데이터 마이닝을 비교하여 향후 학교보건연구에서 데이터 마이닝의 유용성을 평가하였다는 점에서 그 의의를 찾고자 한다.

V. 결론 및 제언

본 연구는 Chi-square와 logistic regression 등의 일반적 통계방식과 데이터 마이닝의 의사결정나무분석을 통해 교직원이 경험하고 있는 '피로'의 현황을 알아보고 그 관련요인을 탐색하는 것을 목적으로 하고 있다.

그 결과 남자 교직원은 35.2%, 여자는 46.4%가 '피로'를 경험하고 있었다. 전통적 통계방식에서 알게된 '피로'의 관련 변인은 낮은 수축기 혈압과 남자 보다는 여자, 젊은 연령, 중학교 교직원, 불규칙적인 식습관, 운동을 안하거나, 1회 운동시간이 30분 미만, 건강하지 않다고 인지하는 군, 그리고 운동시 숨이 차는 것과 같은 자각증상들과 관련이 있음을 알 수 있었다.

한편, 데이터 마이닝 분석을 통해 여러 특성이 결합된 규칙을 도출할 수 있었다. 즉 건강하다고 인지하면서, 나이가 37.5세 이상인 경우는 피로가 단지 19.3%이었으나, 건강하지 않다고 생각하면서, 운동시 숨이 심하게 차고, 나이는 53.5세 미만이며, BMI가 22.7이상인 경우 피로가 84.8% 까지 높아졌다.

그러나 '피로'와 관련이 있을 가능성이 있는 교직원의 직무만족도, 근무환경 및 조직문화등의 특성은 자료수집의 제한으로 인해 본 연구에서 확인하지 못하였다. 그러므로 후행연구를 통해 향후 확인할 것을 제안할 수 있다.

참고 문헌

- 강현철, 서두성, 최종후(1998). Enterprise Miner의 의사결정나무분석 알고리즘. SUSI-K 발표논문집, 9, 169-185.
- 강현철, 최종후(1999). 데이터마이닝-방법론 및 활용.

- 자유아카데미.
- 김돈균(1976). 산업'피로'의 자각증상에 관한 연구. 한국산업의학, 15(4), 1-14.
- 김윤주(1984). 일부 초,중,고교 교사들의 '피로' 자각증상에 관한 비교분석. 연세대학교 석사학위 논문.
- 박기흠(1996). 만성'피로'(연수강좌). 가정의학회지, 17(3), 19-23.
- 신선미(2002). 저체중 청소년의 건강증진을 위한 데이터 마이닝 응용-서울시 고등학생을 중심으로-. 연세대학원 석사학위 논문.
- 이희우, 장철환, 신선미(2001). 교직원 건강상태와 생활습관에 관한 연구. 서울상록과학학술재단지.
- 장세진(2001). 사업장 근로자들의 '피로'와 의료이용, 사고재해 및 결근간의 관련성. 제 53차 대한예방의학회 추계학술대회 연제집, 348-349
- 정원훈, 신미경(1998). 결식에 따른 '피로'도와 노성분 변화에 관한 연구. 대한보건협회지, 24(1), 138-152.
- 정은리(1983). 전문직에 종사하는 일부 여성들이 자각하는 '피로'에 관한 연구. 이화여자대학교 석사학위 논문
- 천병철(2001). 만성'피로'와 관련증상의 요인분석. 제 53차 대한예방의학회 추계학술대회 연제집, 420-421
- Buja A., Lee YS.(1999). Data mining criteria for tree-based regression and classification, <http://www.research.att.com/~andreas/papers/trees.ps.gz>.
- Bultmann U., Kand I., Kasal SV., Beurskens AJ., van den Brandt PA(2002). Fatigue and psychological distress in the working population. Psychometrics, prevalence, and correlates. J Psychosom Res, 52(6), 445-452.
- Chen MK.(1986). The epidemiology of self perceived fatigue among adults. Preventive medicine, 15:74-81.
- Fieler, V.K.(1991). Side effects in the patients receiving radiation therapy for prostate cancer, Oncology Nursing Form, 18(2), 353.
- Friedman JH., Fesher NI(1997). Bump hunting in high-dimensional data. Tech. report. Dept of Statistics, Stanford University.
- Gilber FR(1971).. Highlights form recent seminar on fatigue. Family Practice, 10(7), 309-310
- Grandjean EP., Wotzka G., Schhadd A., Gilgen A.(1970). Fatigue and stress in air traffic controllers. Ergonomic, 14(1), 159-165.
- Hart L.K.(1978). Fatigue in the patient with multiple sclerosis. Research in nursing & Health, 1(4), 147-157.
- O'conner G.T., Buring J.E, Yusuf(1989). An overview of randomized trial of rehabilitation with exercise after myocardial infarction. Circulation, 80, 234-244.
- Potmpa K., Lopez M., Reid C., Lawson L.(1986). Chronic fatigue. Image J Nurs Sch, 18(4), 165-169.
- Rhoten D.(1982). Fatigue: Its physiological and psychological significance. Ergonomics, 11, 427-436.
- Torres-Harding s., Jason L., Taylor R.(2002). Fatigue severity, attributions, Medical utilization, and symptoms in persons with chronic fatigue. J Behav Med, 5(2), 99-113.