

# 사상체질 진단검사 타당성 분석지표의 일반화 연구

황상문 · 박소정 · 강기림 · 권영규 · 채 한\*

부산대학교 한의학전문대학원 양생기능의학부

## Generalized Analysis on Validity of Sasang Type Diagnosis

Sang Moon Hwang, So-Jung Park, Ki-Rim Gang, Young-Kyu Kwon, Han Chae\*

Division of Longevity and Bio-functional Medicine, School of Korean Medicine, Pusan National University

The purpose of this study was to develop generalized validity evaluation indexes for the *Sasang* type diagnostic test. As a generalized evaluation index for *Sasang* type diagnostic test, Generalized Squared Correlation (GC2), Mutual Information (I) and Mutual Information Coefficient (IC) as overall validity index, and Correlation Coefficient (C) and Mutual Information Contribution (ID) as type-specific validity index were suggested and the practicability of these indexes was assessed along with Percentage Correctly Predicted (PCP), adjusted PCP, type-specific sensitivity and type-specific specificity which was proposed previously. The usefulness of the nine validity indexes was examined using previously reported data and three mock *Sasang* type diagnosis results. We demonstrated here that the PCP is not adequate for the analysis of validity. We also showed that our validity index is recommendable for the comparison and improvement of *Sasang* type diagnosis. The generalized validity analysis of *Sasang* type diagnosis can provide a guideline for the development of new *Sasang* type diagnosis or pattern identification. The working sheet for calculating validity indexes can be acquired at <http://www.sasangtypology.org/validity>.

Key words : validity index, sensitivity and specificity, *Sasang* type diagnosis, correlation coefficient, mutual information

### 서론

이제마의 사상의학은 애노희락(哀怒喜樂)의 성정(性情) 또는 기질(temperament)적 차이에서 기인한 장부의 대소에 따라 인간을 태양인, 소양인, 태음인, 소음인의 네 가지로 분류하여, 각 유형에 따른 생리적 병리적 특성에 따라 질병을 예방, 진단, 치료하는 한국 고유의 전통의학이다. 사상의학에서의 체질의 분류는 각 체질에 따른 성품, 정서(emotion)와 기질(temperament), 생리와 병리, 진단과 치료, 예방 뿐 만 아니라 인간의 도덕적 성숙까지 동시에 다루고 있기 때문에, 의학에 있어서 체계적이면서도 독창적인 이론으로서의 고유한 위치를 지니고 있는 생리심리사회학(bio-psycho-sociology)이라 할 수 있다<sup>1-4</sup>.

이에 사상의학에서의 체질 진단은 이론적인 측면과 함께 임상적 활용에 있어서 가장 중요한 의미를 지니게 된다<sup>5</sup>. 체질의 진단에 있어서 동의수세보원에 근간한 객관화 연구들은 용모사기, 체형기상, 향심심육과 성질재간, 병증약리, 소증(素證)

등의 다양한 측면에서 진행되어 왔다<sup>6,7</sup>. 그러나 이러한 각각의 연구들을 통해 획득되었던 사상체질 진단법들의 장단점을 분석하고 평가하는 방법론적 측면에서의 연구는 부족하였던 것이 현실이다<sup>7</sup>.

이에 기존에 제시되었던 진단법들의 장단점과 타당성을 체계적으로 비교하기 위한 수학적 평가모델 및 지표를 제시하는 연구가 최근 들어 시도되었으며<sup>7</sup>, 이를 통해 민감도와 특이도 등에 대한 명확한 정의(definition)가 제시됨으로써 사상의학 진단법에 있어서의 체계적 고찰(systematic review)을 위한 이론적 기초가 마련되었다. 이에 본 연구에서는 이러한 연구를 보다 일반화하여 그 범주를 확장함으로써 보다 효율적인 진단검사법의 개발 지점과 함께 기존 진단검사법의 개선 방안을 제시하고자 한다.

한의학 주변 학문에 있어서는 진단 또는 예측의 정확도를 평가하기 위한 타당성 평가지표들이 다양한 분야에서 활용되고 있다. 통계학(statistics) 측면에서는 예측과 실제 간의 상관성을 평가함에 있어서 두 변수 사이의 선형적 관계의 크기를 표현하는 상관계수(correlation coefficient)가 다양하게 응용되고 있으며<sup>8,9</sup>, 생물 정보학(bioinformatics) 측면에서는 RNA에 의한 단백질

\* 교신저자 : 채 한, 양산시 물금읍 범어리 부산대학교 한의학전문대학원

· E-mail : han@chaelab.org · Tel : 051-510-8470

· 접수 : 2009/08/13 · 수정 : 2009/09/23 · 채택 : 2009/09/28

2차 구조 예측의 정확도 평가에 있어서 정보 이론(information theory)에서 두 변수간의 상호 의존성을 표시하는 상호정보(mutual information) 등이 활용되고 있다<sup>10-12)</sup>.

이에 기존 연구들에서 제시되었던 정확예측율, 민감도와 특이도 이외에 상관계수를 응용한 체질별 상관계수, 일반화 제곱상관과 함께 상호정보를 응용한 상호정보, 상호정보 계수, 체질별 상호정보 기여도 등의 타당성 지표를 설정하였다. 아울러 이들 지표의 의미와 함께 활용상 장단점을 분석하기 위하여 기존 논문에서 보고되었던 연구 결과들과 정보를 지니지 않는 가상의 결과를 함께 비교하였다.

본 연구를 통해 제시되는 일반화된 타당성 평가지표들은 기존의 사상체질 진단검사의 활용 결과에 대한 수학적 이해를 다면적으로 심화시킬 것이며, 진단검사의 평가 요소와 평가 방법을 일반화 및 표준화함으로써 새로운 진단법의 개발이나 기존 진단법의 개선에 활용될 수 있을 것이다. 아울러 본 연구의 저자들은 논문에서 개발한 타당성 평가지표 계산용 워크시트를 <http://www.sasangtypology.org/validity> 에서 자유롭게 활용할 수 있도록 함으로써 활용의 편의성을 높였다.

## 재료 및 방법

### 1. 진단검사의 타당성 평가지표

분류나 예측 또는 진단검사의 타당성을 표시함에 있어서는 통상 2X2 테이블(Table 1)을 사용하며 TP, FN, FP, TN의 4가지 지표로 설명되고 있다<sup>13)</sup>.

Table 1. Typical 2 x 2 contingency table for general diagnostic test.

		Gold Standard Classification	
		Disease	Healthy
Test Result	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

일반적인 의미에서 불 때 의학 분야에서 질병 진단 등과 관련되어 사용되는 타당성 평가지표인 민감도(Sensitivity), 특이도(Specificity), 정확예측률(Percentage Correctly Predicted)의 의미는 다음과 같다.

정확예측률 (PCP) = (TP + TN) / N (= 정확하게 진단된 비율)

민감도 (Sensitivity) = TP / (TP + FN) (= 질병이 실제로 있는 경우 있다고 진단된 비율)

특이도 (Specificity) = TN / (FP + TN) (= 질병이 실제로 없는 경우 없다고 진단된 비율)

Here,

TP (true positive) = 질병이 있는데 양성으로 나온 수

FN (false negative) = 질병이 있는데 음성으로 나온 수

FP (false positive) = 질병이 없는데 양성으로 나온 수

TN (true negative) = 질병이 없는데 음성으로 나온 수

N = TP + FN + FP + TN

사상체질의 진단 또한 실제 체질구별(Gold Standard Type

Classification)과 진단검사를 통해 예측(진단)한 체질간의 일치도를 평가함으로써 그 타당성이 측정될 수 있다. 그러나 질병의 유무라는 두 가지만으로 분류되는 질병 진단과는 달리, 사상체질은 태양인, 소양인, 태음인, 소음인의 4가지의 분류를 사용하므로 4X4 표(Table 2a)를 활용하여 그 타당성을 평가하게 되거나, 실제적인 통계적 유의성을 고려하여 태양인을 제외한 3가지의 분류를 이용한 3X3 표(Table 2b)를 이용하게 된다<sup>7)</sup>.

Table 2. Contingency table for Sasang type diagnostic test.

(a) 4X4 table

		Gold Standard type classification				Total
		Tae-Yang	So-Yang	Tae-Eum	So-Eum	
Test Result	Tae-Yang	a	b	c	d	M1
	So-Yang	e	f	g	h	M2
	Tae-Eum	i	j	k	l	M3
	So-Eum	m	n	o	p	M4
Total		D1	D2	D3	D4	N

(b) 3X3 table excluding Tae-Yang type

		Gold Standard type classification			Total
		So-Yang	Tae-Eum	So-Eum	
Test Result	So-Yang	a	b	c	M1
	Tae-Eum	d	e	f	M2
	So-Eum	g	h	i	M3
Total		D1	D2	D3	N

이와 같은 상황에서는 타당성의 평가를 위해 요구되는 지표들이 2X2 표에서의 TP, FN, FP, TN의 4개가 아닌, a~p의 16개 (또는 a~i의 9개)가 되며, 이에 따라 정확예측률, 민감도, 특이도 또한 개념의 확장을 통해 일반화된 정의가 요구된다<sup>7)</sup>. a~p의 16개의 수를 이용할 경우에 있어서 정확예측률(PCP)은 (a+f+k+l)/N로 정의되며, 민감도와 특이도는 체질에 따라 각각 일반화 유형 민감도(Q<sup>D</sup>, 이하 민감도), 일반화 유형 특이도(Q<sup>M</sup>, 이하 특이도)로 각각 정의된다<sup>7)</sup>.

민감도와 특이도, 정확예측율은 비교적 간단하고 질병의 진단과 분류 검사의 타당성 평가에 가장 널리 사용되는 지표이지만 지나치게 단순화됨으로 인하여 중요하고 상세한 타당성 정보를 포괄하지 못한다는 단점을 지니고 있을 수 있다<sup>14)</sup>. 예를 들어 정확예측률은 진단검사의 유용성을 하나의 숫자로 간단히 표현할 수 있지만, 2X2 표의 경우에 있어서는 FN과 FP의 정보가 전혀 반영되지 않으며, 4X4표에 있어서는 b, c, d, e, g, h, i, j, l, m, n, o의 정보가 반영되지 못한다.

이에 보다 종합적이고 균형 있는 타당성 분석을 위해서는 상관계수나 상호정보와 같은 지표들의 필요성이 제기된다. 상관계수(Correlation Coefficient, C), 일반화 제곱상관(Generalized Squared Correlation, GC<sup>2</sup>), 상호정보(Mutual Information, I), 상호정보 계수(Mutual Information Coefficient, IC), 상호정보 기여도(Mutual Information Contribution) 등은 진단검사의 타당성을 계산함에 있어서 관련된 a~p의 정보가 모두 이용되며, 전체적으로 또는 체질 유형별로 한 개의 숫자로 나타내어지는 타당성 지표로서 진단이나 예측의 정확도를 평가하는데 널리 쓰이고 있으므로<sup>14-16)</sup>, 사상체질 진단검사의 타당성 분석에도 적용될 수 있을 것이다.

이에 본 연구에서는 기존의 연구<sup>7)</sup>에서 이원분류 또는 다원 분류적 접근으로 고찰하였던 것과는 달리, 진단 검사법 자체의 타당성을 평가함과 동시에 다원분류적 접근을 위주로 한 체질별 타당성을 평가하는 접근법을 사용하였다. 이러한 전략은 진단검사 자체의 타당성을 평가함에 있어서 정확예측율(PCP)의 한계성을 막연하게 실감하고 있었던 것<sup>7)</sup>에서 진일보하여, 이를 구체적으로 대체할 수 있는 일반화된 타당성 평가지표를 제시할 수 있을 것이다.

본 연구에서는 진단검사 자체의 타당성을 한 개의 지표로 표현한 것을 포괄지표(overall validity index)로, 각 체질별 타당성을 한 개의 지표로 표현한 것을 체질별 지표(type-specific validity index)로 나누었다. 이에 일반화 제곱상관(GC<sup>2</sup>)과 상호정보(I), 상호정보계수(IC), 정확예측율(PCP), 수정 정확예측율(Adjusted PCP)은 포괄지표로 분류될 수 있으며, 체질별 상관계수(C)와 체질별 상호정보 기여도(I<sup>D</sup>)는 체질별 민감도(type-specific sensitivity), 체질별 특이도(type-specific specificity)와 함께 체질별 지표로 각각 분류된다.

1) 체질별 지표 (type-specific validity index)

(1) 체질별 상관계수(Correlation Coefficient, C)

Table 2a를 사용할 경우 수식은 다음과 같다<sup>14)</sup>.

$$C = \frac{TP \times TN - FP \times FN}{\sqrt{(TP \times FP)(TP \times FN)(TN \times FP)(TN \times FN)}}$$

Here,  
 For Tae-Yang Type(TY) : TP=a, TN=f+g+h+j+k+l+n+o+p, FP=b+c+d, FN=e+i+m  
 For So-Yang Type(SY) : TP=f, TN=a+c+d+i+k+l+m+o+p, FP=e+g+h, FN=b+j+n  
 For Tae-Eum Type(TE) : TP=k, TN=a+b+d+e+f+h+m+n+p, FP=i+j+l, FN=c+g+o  
 For So-Eum Type(SE) : TP=p, TN=a+b+c+e+f+g+i+j+k, FP=m+n+o, FN=d+h+l

상관계수(C)는 두 변수 사이의 선형적인 상관성의 정도를 나타내는 양으로서, 사상체질 진단검사에 있어서는 실제 체질과 진단된 체질간의 일치도를 의미한다. 상관계수는 -1에서 1 사이의 값을 가지며, 양의 상관관계가 있을 경우 양의 값, 음의 상관관계가 있을 경우 음의 값, 상관관계가 전혀 없을 경우에는 0의 값을 가진다<sup>17)</sup>.

예를 들어 태음인을 진단하는 경우에 있어서, 실제 태음인인 경우에는 상관계수가 1에 가까워지고 일치하지 않는다면 0에 가깝게 되며, 만약 거꾸로 진단되는 경우가 많다면 -1에 가깝게 된다.

상관계수는 TP, TN, FP, FN(또는 a~p)의 값들을 모두 사용하게 되므로 민감도나 특이도 등과 같은 단순한 백분율 값보다 더 균형 있는 평가를 할 수 있다는 장점을 지니고 있으나, TP가 아주 작은 경우에도 FP나 FN이 작다면 상관계수가 비교적 큰 값으로 나올 수도 있으므로 주의할 필요가 있다<sup>14)</sup>.

(2) 체질별 상호정보 기여도 (Mutual Information Contribution, I<sup>D</sup>)

Table 2a를 사용할 경우 수식은 다음과 같다<sup>14)</sup>.

For Tae-Yang Type : (TY-I<sup>D</sup>)  
 = -(D1/N)xlog(D1/N) + [(a/N)xlog(a/M1) + (e/N)xlog(e/M2) + (i/N)xlog(i/M3) + (m/N)xlog(m/M4)]  
 For So-Yang Type : (SY-I<sup>D</sup>)  
 = -(D2/N)xlog(D2/N) + [(b/N)xlog(b/M1) + (f/N)xlog(f/M2) + (j/N)xlog(j/M3) + (n/N)xlog(n/M4)]  
 For Tae-Eum Type : (TE-I<sup>D</sup>)  
 = -(D3/N)xlog(D3/N) + [(c/N)xlog(c/M1) + (g/N)xlog(g/M2) + (k/N)xlog(k/M3) + (o/N)xlog(o/M4)]  
 For So-Eum Type : (SE-I<sup>D</sup>)  
 = -(D4/N)xlog(D4/N) + [(d/N)xlog(d/M1) + (h/N)xlog(h/M2) + (l/N)xlog(l/M3) + (p/N)xlog(p/M4)]

상호정보(I)는 그 정의상 한 변수에 대한 정보를 획득하는 것이 다른 변수가 지닌 불확실성을 줄여주는 정도를 나타내는 것으로서, 정보이론(information theory) 분야에 있어서 두 변수 사이의 상호 의존성(dependence)을 측정하기 위하여 사용되고 있다<sup>15)</sup>. 예를 들어 X와 Y가 완전히 독립적이라면 I(X, Y) = 0 이며, X와 Y간 관련성이 높을수록 I(X, Y)의 값은 커지게 된다. 사상체질 진단검사에 있어서는 X는 실제 사상체질, Y는 진단검사를 통해 예측된 사상체질이 되며, 상호정보 I(X, Y)는 실제 체질과 진단 체질간의 관련성 즉 진단검사의 정확성을 측정하는 지표가 된다.

상호정보는 체질별 상호정보 기여도 (Mutual Information Contribution, I<sup>D</sup>)들의 단순 합으로서, 체질별 상호정보 기여도 (I<sup>D</sup>)는 상호정보(I)에 각각의 사상체질 유형 즉 태양인, 소양인, 태음인, 소음인 별로 기여도가 얼마나 되는가를 나타내는 척도이다. 일반적으로 해당 체질에 해당하는 샘플 수가 많을수록, 체질별 정확도가 높을수록 체질별 상호정보 기여도는 높아지게 된다.

2) 포괄지표 (overall validity index)

(1) 일반화 제곱 상관(Generalized Squared Correlation, GC<sup>2</sup>)

Table 2a를 사용할 경우 수식은 다음과 같다<sup>14)</sup>.

$$GC^2 = (a-Za)^2 / Za + (b-Zb)^2 / Zb + (c-Zc)^2 / Zc + \dots + (p-Zp)^2 / Zp$$

Here,  
 Za = D1xM1/N, Zb = D2xM1/N, Zc = D3xM1/N, Zd = D4xM1/N,  
 Ze = D1xM2/N, Zf = D2xM2/N, Zg = D3xM2/N, Zh = D4xM2/N,  
 Zi = D1xM3/N, Zj = D2xM3/N, Zk = D3xM3/N, Zl = D4xM3/N,  
 Zm = D1xM4/N, Zn = D2xM4/N, Zo = D3xM4/N, Zp = D4xM4/N

일반화 제곱상관(GC<sup>2</sup>)은 두 변수 사이의 상관성의 정도를 나타내는 상관계수를 둘 이상의 변수에까지 사용할 수 있도록 카이제곱검정(X<sup>2</sup> statistic)과의 유사성을 이용하여 일반화한 값으로서<sup>14)</sup>, 사상체질 진단검사에 있어서 실제 사상체질과 검사를 통해 진단된 사상체질 간의 상관성을 측정하는 지표가 된다. 두 변수 사이에 상관관계가 전혀 없을 경우 0의 값이 나오고, 상관관계가 강할수록 1에 가까운 값을 보이게 된다. 즉 아무 정보 없이

임의로 진단한 경우와 얼마만큼 다른가(차이가 존재하는가)를 나타내는 양으로서, 양의 상관관계가 클 경우와 함께 음의 상관관계가 클 경우에도 큰 값이 나올 수 있으므로 주의해야 한다.

(2) 상호정보 (Mutual Information, I)

수식은 다음과 같다<sup>15)</sup>.

$$I = (TY-I^D) + (SY-I^D) + (TE-I^D) + (SE-I^D)$$

상호정보(I)는 체질별 상호정보 기여도(I<sup>D</sup>)의 단순 합이다. 앞의 상호정보 기여도(I)에서 설명한 바와 같이 상호정보는 실제 사상체질과 진단된 사상체질 사이의 관련성, 즉 진단검사의 정확성을 측정하는 지표가 된다. 상호정보는 일반화 제곱상관과 마찬가지로 상관관계가 전혀 없을 경우 0의 값이 나오게 되므로, 정보 없이 임의로 진단한 경우와 얼마나 다른지를 나타내는 양으로 사용할 수 있다. 또한 Table 2의 일부 값만 사용하는 것이 아니라 전체 값을 사용하여 계산하는 값이므로 좀 더 균형 있는 평가를 할 수 있다는 장점을 지니고 있다.

(3) 상호정보계수(Mutual Information Coefficient, IC)

Table 2a를 사용한 수식은 다음과 같다<sup>15)</sup>.

$$IC = I / H(X)$$

Here, H(X) is an entropy and can be expressed as :

$$H(X) = - (D1/N) \times (\log D1/N) - (D2/N) \times (\log D2/N) - (D3/N) \times (\log D3/N) - (D4/N) \times (\log D4/N)$$

상호정보계수(IC)는 상호정보(I)를 엔트로피 H(X)로 나누어 최소값이 0, 최대값이 1이 되도록 정규화한 값이다. 사상체질 진단검사에 있어서 실제 사상체질과 검사를 통해 진단된 사상체질 사이에 전혀 관련성이 없는 경우에는 0, 실제 사상체질과 검사를 통해 진단된 사상체질이 완전히 일치할 경우에는 1이 된다. 상호정보가 갖고 있는 장점을 똑같이 갖고 있으면서, 최대값이 1이 되도록 정규화한 값<sup>14)</sup>이므로 서로 다른 진단평가 사이의 값을 비교하기에도 적절하다.

(4) 수정 정확예측율 (Adjusted Percentage Correctly Predicted, Adjusted PCP)

Table 2를 사용할 경우 수식은 다음과 같다.

$$Adjusted\ PCP = (PCP-K)/(1-K)$$

$$= ((a+e+i)/N-K)/(1-K)$$

K는 아무 정보도 지니지 않은 임의의 진단검사에서의 나타나는 정확예측율

이는 정확예측율(PCP)에서 아무 정보도 지니지 않은 임의의 진단검사에서의 나타나는 정확예측율(K)을 제하여 0~1 사이의 값을 지닐 수 있도록 수정한 값이다.

2. 진단검사의 타당성 분석

이상에서 기술된 사상체질 진단검사의 타당성을 체계적으로 비교하기 위하여 제시된 진단검사 타당성 지표들을 기존에 발표되었던 수편의 논문<sup>5,18,19)</sup>에 기재되어 있는 데이터를 사용하여 계산하고, 이를 기존에 발표되었던 논문의 데이터 분석 결과와<sup>5,7,18,19)</sup> 비교 고찰하여, 각 진단검사의 장단점 분석 및 특성 파악에 활용될 수 있는지 살펴보았다. 다만 분석에 활용하기 위한 데이터가 기재된 논문이 많지 않았으며, 본 연구의 목적이 타당성 분석지표의 일반화 여부 및 활용 가능성에 있었으므로, 서로 다른 측정방법들을 활용하였던 수 편의 연구를 우선적으로 사용하여 분석의 대상으로 삼았다<sup>7)</sup>.

Laryngograph와 EEG를 사용하였던 김성형 등의 연구<sup>18)</sup>(진단검사 A), TS\_QSCD를 사용하였던 김영원 등의 연구<sup>19)</sup>(진단검사 B), 개정 사상체질분류검사지 II를 사용하였던 최경주 등의 연구<sup>5)</sup>(진단검사 C)가 세 종류의 가상의 진단검사<sup>14)</sup>(진단검사 D, E, F)와 함께 분석에 사용되었다. 이러한 세 종류의 “정보 없는” 가상의 임의의 진단검사는 검사대상에 대한 정보가 전혀 없이 사전에 계획된 전략에 따라서 임의로 체질을 분류하는 방법이다. 진단검사 D (Random1/3)는 소양인, 태음인, 소음인을 각각 1/3의 같은 비율로 임의로 분류하는 방법이며, 진단검사 E (RandomBG)는 기존에 알려져 있는 비율<sup>20)</sup>에 따라 소양인/태음인/소음인을 30/50/20의 비율로 임의의 분류하는 방법이고, 진단검사 F (UniformTE)는 가장 비율이 높은 것으로 알려진 태음인<sup>20)</sup>으로만 분류하는 방법이다.

본 연구에서는 태양인의 경우 분석에서 제외하고 소양인, 태음인, 소음인의 세 가지 체질만을 분석대상으로 삼았으며, 진단검사의 타당성 분석에 있어서는 위에서 제시한 정확예측율(PCP), 상호정보(I), 상호정보계수(IC), 일반화 제곱상관(GC<sup>2</sup>), 체질별 민감도(Q<sup>D</sup>), 체질별 특이도(Q<sup>M</sup>), 체질별 상호정보 기여도(I<sup>D</sup>), 체질별 상관계수(C), 수정 정확예측율(adjusted PCP)의 아홉 가지 지표를 사용하였다.

결 과

본 연구에 사용된 진단검사(A~F) 별 타당성 평가지표 분석 결과는 Table 3과 같다. 전체적으로 살펴볼 때 기존 연구 결과를 통해 얻었던 진단검사법 A, B, C는 모두 “정보 없는” 진단검사 D, E, F 보다 모든 항목에 있어서 높은 수치를 보이고 있음을 확인할 수 있었으며, 진단검사 F의 태음인 민감도만 100%로 높게 나타남을 알 수 있었다. 또한 각 진단별 타당성 지표에 있어서, 정확예측율(PCP), 수정된 정확예측율(adjusted PCP), 상호정보(I), 상호정보계수(IC), 일반화 제곱상관(GC<sup>2</sup>), 체질별 민감도(Q<sup>D</sup>), 체질별 특이도(Q<sup>M</sup>), 상호정보 기여도(I<sup>D</sup>), 상관계수(C) 등은 서로 비슷하게 높거나 낮은 수치를 보임을 알 수 있었다.

1. 포괄 지표

진단검사 자체의 타당성을 하나의 숫자로 나타내는 포괄지표로서는 정확예측율(PCP)과 수정된 정확예측율(adjusted PCP), 상호정보(I), 상호정보 계수(IC), 일반화 제곱상관(GC<sup>2</sup>)을

계산하였으며, 예상했던 것과 같이 정확예측률이 높을수록 다른 포괄지표들도 높게 나온다는 것, 즉 지표 간에 양의 상관관계가 있다는 것을 확인할 수 있었다(Table 3, Fig. 1). 그러나 정확예측율(PCP)과 다른 네 지표간의 뚜렷한 차이점 또한 확인할 수 있었는데, 정확예측율(PCP)는 정보를 지니고 있거나(진단검사 A~C) 또는 정보를 지니지 않았거나(가상의 진단검사 D~F) 모두 0.3 이상의 값을 보이고 있었지만, 나머지 네 지표들은 정보를 지니지 않았을 경우(가상의 진단검사 D~F)에는 0의 값을 보임을 알 수 있었다.

진단검사 A, B, C에 있어서 정확예측율(PCP)은 각각 약 92%, 76%, 58% 로 나타나는데 비하여 일반화 제곱상관(GC<sup>2</sup>)은 각각 0.75, 0.42, 0.17를, 상호정보(I)는 0.81, 0.40, 0.16를, 상호정보계수(IC)는 0.75, 0.38, 0.15를 나타내었다.

각 타당성 평가지표에 있어서 가장 높은 값을 보인 진단검사 A와 가장 낮은 값을 보인 진단검사 C 사이의 차이를 비교해 보면, 정확예측율(PCP)은 그 차이가 0.34에 그친 데 반하여 수정된 정확예측율(adjusted PCP), 일반화제곱상관(GC<sup>2</sup>), 상호정보(I), 상호정보계수(IC)는 각각 0.67, 0.58, 0.65, 0.60으로, 세 타당성 지표 모두 각 진단검사 간의 비교에 있어서 정확예측율(PCP)보다 훨씬 큰 차이를 보여 주고 있는 것을 알 수 있다.

가상의 임의 진단검사 D, E, F에 있어서는 피험자의 비율에 따라 결과를 계산해낼 수 있었다. 피험자들을 각각 1/3의 비율로 소양인, 태음인, 소음인으로 진단한 진단검사 D에 있어서는 정확예측율(PCP)은 약 33.3%가 된다.

**Table 3. Validity indexes of three previously reported Sasang type diagnostic test A, B, C and three "non-informative" diagnostic test D, E, F.**

Test	PCP	Adjusted PCP (K=0.5)	I	IC	GC <sup>2</sup>	Q <sup>D</sup>	Q <sup>M</sup>	I <sup>D</sup>	C	Sasang Type
A (n=24)	0.917	0.83	0.807	0.749	0.749	0.667	1.000	0.155	0.775	SY
						1.000	0.889	0.327	0.913	TE
						1.000	0.909	0.325	0.919	SE
B (n=151)	0.762	0.52	0.402	0.375	0.420	0.744	0.711	0.132	0.616	SY
						0.818	0.692	0.156	0.677	TE
						0.818	0.692	0.156	0.677	SE
C (n=587)	0.584	0.17	0.158	0.151	0.165	0.413	0.685	0.029	0.319	SY
						0.675	0.664	0.058	0.443	TE
						0.769	0.415	0.071	0.429	SE
D (Random 1/3)	0.33	-	0	0	0	0.33	0.30	0	0	SY
						0.33	0.20	0	0	TE
						0.33	0.20	0	0	SE
E (Random mBG)	0.375	-	0	0	0	0.30	0.30	0	0	SY
						0.50	0.50	0	0	TE
						0.20	0.20	0	0	SE
F (Uniform TE)	0.50	-	0	0	0	0	-	0	0	SY
						0	-	0	0	TE
						0	-	0	0	SE

Test D(Random 1/3) is a random classification giving each category of So-Yang, Tae-Eum and So-Eum type the same probability of 1/3. Test E(RandomBG) is a random classification reflecting the generally known probability of 30 : 50 : 20<sup>20)</sup>, and Test F(UniformTE) is a uniform classification into Tae-Eum category, which is known as the most abundant of the three. The value K=0.5 (PCP of the test F) was used in the calculation of the adjusted PCP. \* PCP : Percentage Correctly Predicted, I : Mutual Information, IC : Mutual Information Coefficient, GC<sup>2</sup> : Generalized Squared Correlation, Q<sup>D</sup> : Generalized Type Sensitivity, Q<sup>M</sup> : Generalized Type Specificity, I<sup>D</sup> : Mutual Information Contribution, C : Correlation Coefficient

소양인/태음인/소음인을 30/50/20의 비율로 임의 진단하는 진단검사 E의 경우 정확예측율(PCP)은 37.5%이며, 가장 비율이 높은(50%) 것으로 알려진 태음인으로만 진단하는 진단검사 F의 경우에는 예상대로 정확예측율(PCP)이 태음인의 비율과 같은 50%에 달하는 것을 알 수 있었다. 이처럼 정확예측율(PCP)은 아무런 정보 없이도 최대 50%의 값을 지닐 수 있으나, 일반화 제곱상관(GC<sup>2</sup>), 상호정보(I), 상호정보계수(IC)는 모두 0의 값을 보이는 것을 알 수 있다.

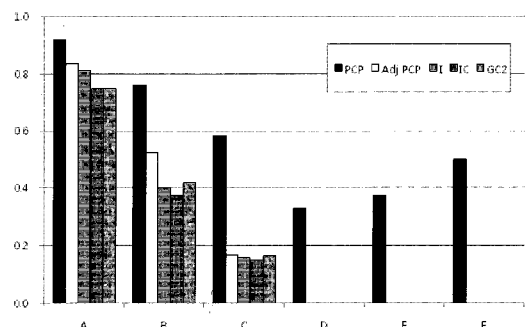
2. 체질별 지표

체질별 지표인 체질별 민감도(Q<sup>D</sup>), 체질별 특이도(Q<sup>M</sup>), 상호정보 기여도(I<sup>D</sup>), 상관계수(C)에 있어서, 민감도와 특이도가 높은 경우에 상호정보 기여도(I<sup>D</sup>)와 상관계수(C)가 대체적으로 높게 나오는 것을 확인할 수 있었다. 다만 이들 값의 크기가 체질별로 상반되는 경향성을 보일 수도 있으므로, 해석에 있어서는 주의할 필요가 있음도 확인할 수 있었다. 또한 상호정보 기여도(I<sup>D</sup>), 상관계수(C)에 있어서는 정보를 지니지 않는 가상의 진단검사 D~F는 0의 값을 보이고 있었다.

진단검사 A<sup>18)</sup>의 타당성 평가지표(Fig. 2a)에 있어서, 민감도는 소양인의 경우 가장 낮지만 특이도는 이와 반대로 소양인이 가장 높게 나오는 것을 알 수 있었다. 소양인/태음인/소음인에 있어서 상호정보 기여도(I<sup>D</sup>)는 0.1551/0.3269/0.3251의 수치를 보였으며, 상관계수(C)에 있어서는 0.7746/0.9129/0.9188로 소양인이 태음인이나 소음인보다 낮은 수치를 보였다.

진단검사 B<sup>19)</sup>에 있어서는 체질별로 각 타당성 평가지표의 크기 순서가 달라지는 것을 확인할 수 있었다(Fig. 2b). 민감도에 있어서는 소음인이 가장 높았지만 특이도에서는 태음인이 가장 큰 것을 알 수 있었다. 소양인/태음인/소음인에 있어서 상호정보 기여도(I<sup>D</sup>)는 0.1320/0.1063/0.1560로 나타났으며, 상관계수(C)는 0.6155/0.6318/0.6765로 나타났는데, 이는 소음인에 있어서 다른 체질들에서보다 높은 수치를 보이는 것이다.

진단검사 C<sup>19)</sup>에서 민감도는 소양인(41.3%)의 경우 가장 낮았지만 특이도는 소음인(41.5%)에 있어서 가장 낮게 나타났다. 상호정보 기여도(I<sup>D</sup>)와 상관계수(C)는 소양인/태음인/소음인에 있어서 각각 0.029/0.058/0.071, 0.319/0.443/0.429로 나타났(Fig. 2c).



**Fig. 1. PCP, adjusted PCP, I, IC and GC<sup>2</sup> in the diagnostic test A, B, C and 'non-informative' mock test D, E, F. \* PCP : Percentage Correctly Predicted, I : Mutual Information, IC : Mutual Information Coefficient, GC<sup>2</sup> : Generalized Squared Correlation**

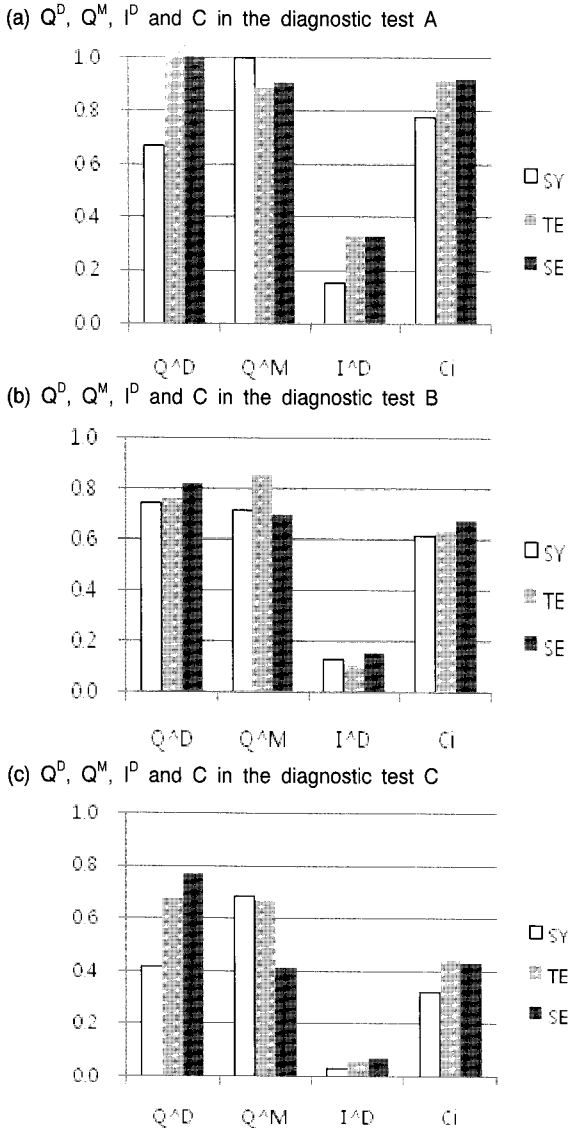


Fig. 2.  $Q^D$ ,  $Q^M$ ,  $I^D$  and  $C$  in the diagnostic test A, B, C. \*  $Q^D$  : Generalized Type Sensitivity,  $Q^M$  : Generalized Type Specificity,  $I^D$  : Mutual Information Contribution,  $C$  : Correlation Coefficient

### 고찰 및 결론

본 연구에서는 사상체질 진단검사의 타당성 분석을 위하여 일반화된 타당성 지표로서 상호정보(I), 상호정보 계수(IC), 일반화 제곱상관( $GC^2$ )과 같은 포괄지표(overall validity index)와 상호정보 기여도( $I^D$ ), 상관계수(C)와 같은 체질별 지표(type-specific validity index)를 제시하고, 이의 효용성을 기존의 정확예측율(PCP), 수정된 정확예측율(adjusted PCP), 민감도와 특이도와 함께 분석하였다.

사상체질 진단검사법 자체의 타당성을 하나의 수치로 표현한 포괄지표(overall validity index)에 있어서는 기존의 정확예측율(PCP)의 문제점을 명확히 분석해 낼 수 있었으며, 상호정보(I), 상호정보계수(IC), 일반화제곱상관( $GC^2$ )의 유용성을 확인할 수 있었다. 정확예측율(PCP)은 임의의 정보 없는 진단검사에 대해서도 33~50%라는 높은 값을 보이고 있었으나, 본 연구에서 제시

한 상호정보(I), 상호정보계수(IC), 일반화제곱상관( $GC^2$ )은 정보를 지니지 않은 다양한 임의의 진단검사에 있어서 모두 동일하게 0의 값을 보이고 있었다(Table 3).

또한 정확예측율(PCP)은 각 진단검사의 타당성을 민감하게 반영하지도 못하고 있었는데, 이는 본 연구에서 제시한 상호정보(I), 상호정보계수(IC), 일반화 제곱상관( $GC^2$ )이 진단검사의 타당성을 보다 예민하게 반영한다는 것을 의미한다. 이는 정확예측율이 피험자에 대한 아무런 진단 정보 없이도 최대 50%(진단검사 F)를 얻을 수 있다는 점에 기인하는 것으로, 이러한 상황을 고려하여 정확예측율(PCP)을 보정한 수정 정확예측율(adjusted PCP)이 본 연구에서 제시된 상호정보(I), 상호정보계수(IC), 일반화 제곱상관( $GC^2$ )과 유사한 경향성을 보이게 되는 점에서도 확인할 수 있다(Fig. 1).

이에 정확예측율(PCP)이라는 개념은 진단검사 타당성을 나타내는 지표로서의 치명적인 문제점을 지니고 있다 할 수 있다. 본 연구에서 고찰한 바와 같이 임의의 진단의 전략에 따라서 기본적으로 33.3~50%의 점수를 더 얻게 되므로 실제 피험자의 체질 구성이나 진단검사의 내부 전략에 따라 타당성의 기본 점수가 달라지게 되어(Table 3), 각 진단 검사에 대한 기본적인 정보들이 충분히 제시되고 고려되지 않는 한 타당성의 상호 비교는 불가능하게 된다.

예를 들면, 92%의 정확예측율(PCP)을 보인 진단검사 A가 76%의 정확예측율(PCP)을 보인 진단검사 B보다 더 타당한 검사법이라고 말할 수 있는지도 분명하지 않을 것이며, 본 연구에서의 진단검사 C의 58%라는 정확예측율이 과연 아무런 정보도 없이 임의로 진단한 진단검사 D, E, F보다 얼마나 더 타당한 지를 계산할 수도 없을 것이다. 이에 진단검사 타당성을 분석하는 지표로서의 정확예측율(PCP)은 논리적 정당성을 지니지 못하는 것이라 사료된다.

본 연구에서 제시된 일반화 제곱상관( $GC^2$ ), 상호정보(I), 상호정보계수(IC)는 기존의 정확예측율(PCP)과는 달리 정보 없는 가상의 임의 분류법인 진단검사 D, E, F에서 모두 0의 값을 얻을 수 있었으며, 정확예측율의 보정 없이도 각 진단검사(진단검사 A, B, C)의 결과 수치만으로도 그 타당도를 바로 평가할 수 있었으며, 진단검사 간의 차이 또한 정확예측율(PCP) 보다 민감하게 반응하고 있음을 알 수 있다. 아울러 Table 2의 a~p에 이르는 모든 값들이 반영되어 있는 보다 체계적이고 종합적인 수치이므로, 진단검사의 타당성 평가 지표로서는 매우 바람직하다 할 수 있다.

다만 본 연구에 사용된 세 진단 검사(A, B, C)의 경우 사용된 샘플의 크기가 각각 24, 151, 587로 매우 큰 차이를 보이고 있으므로, 이러한 값이 고려되지 않는 상황에서는 타당성 지표들이 모두 높게 나온 진단검사 A가 통계적으로 가장 우수한 진단검사법이라고 단정짓기에는 어려움이 있을 것이라 사료된다. 아울러 통계적 우수성을 논의함에 있어서는 진단검사에 사용되었던 측정 방법 및 측정치들의 안정성이나 정확도와 함께 기준으로 삼은 검사(gold standard)의 특성 등에 따라서도 전혀 다른 평가가 가능하다는 점도 충분히 고려되어야 할 것이다.

사상체질 진단검사법에 있어서 각각의 체질별 검사 타당성

을 측정하는 체질별 지표(type-specific validity index)에 있어서, 기존 연구 결과에 대한 평가에 있어서 그 의미에 대한 논의가 다시금 이루어질 필요성이 제기되었다. 진단검사 A를 사용한 기존의 연구<sup>18)</sup>에서는 “체질별 정관별력”이 소양인의 경우에서 가장 낮기 때문에 좋지 못하다고 하였는데, 이에 대해 이수진 등<sup>7)</sup>은 민감도를 기준으로 한 평가이므로 특이도를 평가 기준으로 한 경우에는 소양인이 가장 낮을 것이라는 전혀 다른 결론을 내렸을 것이라는 견해를 밝혔다. 이러한 평가는 진단검사 타당성 지표의 표준화가 필요하다는 근거로서는 매우 적절하나, 민감도나 특이도보다 더 종합적으로 타당도를 반영할 수 있는<sup>14)</sup> 상호정보 기여도( $I^D$ )와 상관계수(C) 등이 함께 고려되었다면, 소양인에 있어서의 유용성이 다른 체질에서 보다 낮다는 결론은 그대로 유지할 수 있었을 것이라 사료된다. 다만, 이 경우에 있어서 체질별 샘플의 수가 작으므로 통계적 유의성에 대해서는 별도의 평가가 필요할 것으로 보인다(Fig. 2a).

진단검사 B를 사용하였던 기존의 연구<sup>19)</sup>에서, 저자들은 체질별 민감도를 의미하는 “일치율”을 사용하여 소음인의 일치도가 가장 높다고 결론을 내렸으며, 이에 대해 이수진 등<sup>7)</sup>은 체질별 특이도를 사용하였다면 태음인의 일치도가 가장 높을 것이라는 분석을 시도하였다. 그러나 상호정보 기여도( $I^D$ )와 상관계수(C)에 있어서 소음인의 경우 다른 체질에 비해 다소 높은 것을 확인할 수 있으므로 기존의 저자들<sup>19)</sup>이 내렸던 결론이 조금 더 타당할 것으로 보이므로, 다양한 지표들을 함께 고려하여 종합적으로 평가하는 것이 필요하다고 사료된다. 다만 이들 타당성 지표들의 체질별 차이가 그리 크지 않으므로 이에 대한 결론을 내리는 데에 있어서는 신중해야 할 것으로 보인다(Fig. 2b).

진단검사 C를 사용하였던 기존 연구<sup>3)</sup>에서는 소양인의 민감도(41.3%)와 소음인의 특이도(41.5%) 값이 다른 지표에 비해서 눈에 띄게 낮은 것을 볼 수 있는데, Table 3에서 확인할 수 있듯이 이는 정보 없는 진단검사법에 의한 값, 예를 들어 검사 D의 소양인 일반화 유형 민감도 33.3%, 소음인 일반화 유형 특이도 25%에 비해서 그리 높다고는 할 수 없다. 이러한 점은 포괄 지표인 정확예측률(PCP)이 58.4% 정도로 비교적 낮게 나오는 것<sup>7)</sup>과 함께 체질별 상호정보 기여도( $I^D$ )와 상관계수(C)가 다른 진단검사에 비해 낮게 나오게 되는 원인이 된다. 각 체질별 상관계수(C)의 비교에 있어서는 다른 체질에 비해 민감도와 특이도가 비교적 고르게 나타난 태음인의 상관계수가 가장 높은 것을 확인할 수 있었다(Fig. 2c).

체질별 민감도와 특이도가 각각 “실제로 해당 체질인 경우 진단이 정확할 확률”, “해당 체질로 진단한 것이 실제로 정확할 확률”을 의미하는 것으로서 중요한 정보의 양쪽 측면을 반영하는 것임에 비해서, 상호정보 기여도( $I^D$ )와 상관계수(C)는 보다 종합적인 지표<sup>14)</sup>로서 해당 정보를 보다 균형 있게 반영하므로 이들에게 전반적인 경향성을 파악하는 지표로서의 역할을 부여하는 것이 바람직하다고 보인다. 이에 상호정보 기여도( $I^D$ )와 상관계수(C)를 기준으로 타당도는 진단검사 A에서는 소양인이 가장 낮고, 진단검사 B에서는 소음인이 가장 높고, 진단검사 C에서는 소양인이 가장 낮다고 보는 것이 타당하다고 사료된다.

과거 사상체질 진단검사의 타당성 지표 및 상호 비교의 근거로 사용되었던 정확예측률(PCP)은, 본 연구에서 살펴본 것과 같이 각 진단검사의 타당성들을 비교함에 있어 치명적인 결함을 지니고 있음을 확인할 수 있었다. 또한 기존 연구<sup>7)</sup>에서 제시된 민감도( $Q^D$ )와 특이도( $Q^M$ )는 진단검사의 장단점을 분석하는 타당성 분석지표로서 가장 간단하고도 강력한 도구가 될 것이나, 지나친 단순화로 인해 관련 정보를 포괄하는 데에는 한계를 지닐 수 있다. 이에 진단검사의 전반적인 성능 평가에 있어서는 상호정보( $I$ ), 상호정보 계수(IC), 일반화 제곱상관( $GC^2$ ), 상호정보 기여도( $I^D$ ), 상관계수(C)의 활용이 보다 종합적인 의미를 지닐 것이라 보인다.

아울러 본 연구에서 제시된 타당성 평가지표들은 기존 진단검사의 개선이나 새로운 진단검사의 개발에 중요하게 활용될 수 있을 것으로 기대된다. 진단검사법의 개발 과정에 있어서 본 연구에서 제시된 각종 타당성 평가 지표들은 항목별 가중치나 문턱값의 조정, ROC 곡선의 조정<sup>17)</sup> 등에 활용될 수 있으며, 이를 통해 최적화된 진단검사의 개발이 가능해질 것이다.

## 감사의 글

이 논문은 2009년도 한국한의학연구원의 지원을 받아 기관 고유사업의 일환으로 수행된 연구임(K09011).

## 참고문헌

1. 전국한의과대학사상의학교실. 개정증보 사상의학. 서울, 집문당, 2004.
2. 채한, 엄일규, 김병철, 김영인, 김병주, 권영규. 신경망을 사용한 사상체질 진단검사 개발 연구. 동의생리병리학회지 23(4):765-771, 2009.
3. Park, S.H., Kim, M.G., Lee, S.J., Kim, J.Y., Chae, H. Temperament and Character Profiles of Sasang Typology in an Adult Clinical Sample. Evid Based Complement Alternat Med, 2009.
4. Chae, H., Park, S.H., Lee, S.J., Kim, M.G., Wedding, D., Kwon, Y.K. Psychological Profile of Sasang Typology: A Systematic Review. In Press.
5. 최경주, 최양식, 차재훈, 황민우, 이수경, 고병희, 송일병. 개정된 사상체질분류검사지Ⅱ의 신뢰도와 타당도에 대한 연구. 사상체질의학회지 18: 62-74, 2006.
6. 채한, 홍무창, 배현수, 신민규. 사상인의 생리적 특성 연구. 동의생리병리학회지 19: 304-314, 2005.
7. 이수진, 김명근, 채한. 사상체질진단검사 타당성 분석에 대한 연구. 대한한의학회지 29: 7-14, 2008.
8. Davies, R.J., Ali, N.J., Stradling, J.R. Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. Thorax 47: 101-105, 1992.

9. Foa, E.B., Riggs, D.S., Dancu, C.V., Rothbaum, B.O. Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress* 6: 459-473, 1993.
10. A.J. B., I.S, K. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomp*, 2000.
11. Atchley, W.R., Wollenberg, K.R., Fitch, W.M., Terhalle, W., Dress, A.W. Correlations Among Amino Acid Sites in bHLH Protein Domains: An Information Theoretic Analysis. *Molecular Biology and Evolution* 17: 164-178, 2000.
12. Huynen, M., Snel, B., III, W.L., Bork, P. Predicting Protein Function by Genomic Context: Quantitative Evaluation and Qualitative Inferences. *Genome Research* 10: 1204-1210, 2000.
13. Dawson, B., Trapp, R.G., Trapp, R. *Basic & Clinical Biostatistics* (4th Ed.). Mc Graw Hill Medical, 2005.
14. Baldi, P., Brunak, S., Chauvin, Y., Andersen C.A.F., Nielsen, H. Assessing the accuracy of prediction algorithms for classification : an overview. *Bioinformatics* 16: 412-424, 2000.
15. Guiasu, S. *Information Theory with Applications*. New York: Mc Graw Hill, 1977.
16. Cohen, J., Cohen, P., West, S.G., Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New Jersey : Lawrence Erlbaum Associates, 2003.
17. Petrie, A., Sabin, C. *Medical statistics at a glance* (2nd Ed.). Massachusetts: Blackwell Publishing, Ltd. 2005.
18. 김선형, 신미란, 김달래, 권기록. Laryngograph와 EEG를 이용한 음향특성과 사상체질간의 상관성 연구. *사상체질의학회지* 12: 144-156, 2000.
19. 김영원, 신동윤, 김정호, 최대성, 임미경, 이경로, 송정모. 사상체질 진단을 위한 2단계 설문지(TS-QSCD)의 개발 연구. *사상체질의학회지* 18: 75-90, 2006.
20. 채한, 박수잔, 이수진, 고평찬. 사상 유형학의 성격심리학적 교찰. *대한한의학회지* 25: 151-164, 2004.