



## A Study on Deduction Method of Significant Symptoms Combination based on Null Hypotheses

Yong-Taek Oh<sup>1</sup>, An-Na Kim<sup>2</sup>

<sup>1</sup>*Inuri Oriental medical clinic*

<sup>2</sup>*Mibyeong Research Center, Korea Institute of Oriental Medicine*

### ABSTRACT

This study aims to propose an objective analysis method for significant symptoms combination applicable to vast amount of data including big data or clinical data instantly generated in real time, as a substitute for methods based on TKM experts' analysis. A mathematical method based on null hypotheses was used for analyzing relative multi frequency symptom-treatment combinations. These results were used for comparison with TKM experts' analysis on significant symptom-treatment combinations. Based on the correspondence of comparison of the two, for certain purposes relative multi frequency combinations may be considered and used as significant combinations. Therefore, the use of the mathematical method based on null hypotheses for analysing relative multi frequency combinations on text based combinations was appropriate. This method was applied to treatment related formula from Donguibogam and resulted in 39 combinations of significant symptom-symptom, symptom→condition, condition→symptom combinations. This study has some limitations using data only from pain treatment formula. Nevertheless, this limitation may be solved and lead to extended research using various sources of data. Combinations that were not applicable to the hypothesis "Combinations with frequent appearance in certain study subjects are more significant than combinations that are not" were not found. Also, combinations with relatively low frequency in appearance but was significant in result were not found. The limitations of this study requires further research to verify the hypothesis and analysis for combinations that cannot be covered by the hypothesis.

© 2016 KKITS All rights reserved

**KEYWORDS:** Traditional korean medicine, Frequent symptoms combination, Significant symptoms combination, Null hypotheses, Deduction method

**ARTICLE INFO:** Received 10 June 2016, Revised 19 August 2016, Accepted 19 August 2016.

\*Corresponding author is with Mibyeong Research Center, Korea Institute of Oriental Medicine, 1672

Yuseong-daero, Yuseong-gu, Daejeon, 34054, KOREA.  
E-mail address: [ankim2012@kiom.re.kr](mailto:ankim2012@kiom.re.kr)

## 1. 서론

정보통신기술이 발전함에 따라 많은 지식을 빅데이터로 축적하여 활용하는 시대가 되었다. 초기에 데이터가 축적되던 시기에는 빅데이터 자체를 만드는 것이 중요하였으나, 많은 데이터가 축적된 지금은 빅데이터를 분석하여 유용한 정보를 찾아 활용하는 것이 매우 중요해졌다. 이는 의료에서도 마찬가지인데, 과거에는 의료지식을 서적에 기록하는 것만으로도 저장, 활용하는 데에 부족함이 없었으나 학문이 발전하여 축적된 의료지식의 양이 증가하고, 정보통신기술의 발달로 많은 의료데이터가 발생함에 따라 이를 잘 저장, 분석하고 활용하는 연구가 필요해졌다. 이에 대한 연구는 한의학 분야에서도 진행되고 있는데, 한의학 지식을 데이터베이스나 온톨로지 등으로 구축하는 노력이 있었고 [1], 최근에는 구축된 데이터를 통해 새로운 지식을 탐구하거나, 구축된 데이터를 의료 현장에서 손쉽게 활용할 수 있도록 분석, 가공하는 연구가 필요하다.

한의학 빅데이터에는 병증, 치법, 효능, 본초, 처방, 침구, 경혈 등의 여러 가지 콘텐츠가 있는데, 한의학 데이터를 분석하는 데에 있어서는 각 요소들 간의 관계를 정의하고, 요소들 간의 조합을 분석하는 일이 중요하다. 예를 들어 처방과 본초는 구성 대상과 구성 요소의 관계로 정의되고, 구성 요소인 본초들은 구성 대상인 처방 안에서 조합을 이루게 된다. 구성 요소인 본초들의 조합을 이해하는 것은 그 구성 대상인 처방을 이해하고, 활용하는데 중요한 역할을 할 수 있어, 과거 한의학 이론에서는 이를 주로 군신좌사나 약대 등과 같은 이론으로 설명해왔으며, 최근까지 한의학 데이터 내의 조합 분석 연구도 주로 처방 내 본초 조합 분석 연구를 중심으로 발전해 왔다[2-8]. 한의학 구성 요소들 중 중요하지 않은 것은 없겠지만, 한의학은

서양의학과 달리 특정 영상 및 임상물리검사 결과를 진단에 활용하기보다는 전체 증상 조합의 경향성이나 증상들의 공통된 속성을 분석하여 병을 진단하는 학문적 특성이 있으므로 많은 증상 조합들 중에서 학문적, 임상적으로 유의한 조합을 찾는 것은 매우 중요하다. 따라서 본 연구에서는 한의학 빅데이터 내의 많은 증상 조합들 중 유의한 조합을 찾을 수 있는 방법을 연구하고자 하였다.

한의학 콘텐츠의 여러 조합들 중에서 유의한 조합을 찾는 방법은 여러 가지가 있을 수 있겠지만 그동안은 등장 횟수의 절대 값이나 한의학 전문가에 의한 판단에 주로 의존해왔다[2][4-6][9]. 등장 횟수의 절대 값에 의한 분석은 그 데이터의 양이 중요도 판단에 영향을 줄 수 있고, 한의학 전문가에 의한 분석은 임상 데이터와 같이 실시간으로 새로운 데이터가 발생하는 데이터에는 적용하기 힘들다는 한계점이 있다. 따라서 본 연구에서는 타 분야에서 활용되고 있는 수학적 분석 방법을 한의학 콘텐츠에 적용하고, 그 결과와 한의학 전문가에 의한 분석 결과를 비교하여, 한의학 콘텐츠에서 유의 조합을 찾는 데 이용가능한 객관적인 방법을 제시하고자 하였다. 이를 위해 먼저 다수의 증상-치법 조합들을 복수의 전문가가 분석하여 찾은 유의 조합들과 수학적 방법의 하나인 귀무가설을 이용하여 분석하여 찾은 유의 조합들을 비교하여 그 결과의 일치율을 통해 귀무가설의 적용 가능성을 검증하고, 이를 증상-증상, 조건→증상, 증상→조건 조합에 적용하여 유의한 조합을 찾아보고자 하였다.

## 2. 본론

### 2.1 연구대상

증상-치법 조합 분석에 사용한 데이터는 최근

한의학 온톨로지 연구[1]에서도 사용된 전국 한의학 공통교재인 5계 내과학[10-14]에 수록된 수반증상과 처방의 다대다 조합에서 추출 가능한 모든 조합을 대상으로 하였다. 증상 조합 분석에는 동의보감[15] 내 통증과 관련된 처방의 주치를 활용하였는데, 이는 모든 처방의 주치 데이터를 대상으로 하기에는 증상 개념의 수가 매우 많아 본 연구 결과의 적절성을 판단하는데 어려움이 예상되어 일부 처방으로 한정할 필요가 있었고, 통증은 환자의 중요한 자각 증상 중 하나이며 많은 질병에 수반되어 다른 많은 증상과 조합을 갖기 때문이다. 통증과 관련된 처방은 주치에 疼이나 痛이 들어간 654개의 처방으로 선정하였고, 이에서 추출한 증상-증상, 증상→조건과 조건→증상 조합을 대상으로 하였다.

1) 5계 내과학 데이터의 정제

5계 내과학 교재에는 총 834개의 질병명이 존재하는데 각 질병에 제시된 수반증상과 처방을 띄어쓰거나 쉼표를 기준으로 정제하여 각각 4,173개, 834개의 데이터를 얻었다[16]. 처방 용어는 정제 및 개념 단위로의 분해와 관련된 사전 연구[17]가 있어 이를 따랐으며 그 예는 다음과 같다.

표 1. 처방 데이터 정제, 분해 예  
Table 1. Examples of treatment method data refinement and decomposition

치료대상	처방 원문	처방 정제 및 분해
氣虛陽微噎隔	溫補脾腎	溫補脾
		溫補腎
	益氣回陽	益氣
		回陽
氣虛中風	益氣活血	益氣
		活血
	補益健脾	補益健脾

2) 동의보감 주치 데이터의 정제

(1) 동의보감 주치 데이터의 특징

동의보감 처방의 주치 데이터에는 증상과 병의 개념이 혼재되어 있는데, 그 구분이 명확한 것도 있으나 咳嗽(해수, 기침), 眩暈(현훈, 어지러움)과 같이 둘 중 한 개념으로만 분류하기에는 모호한 것들도 있다. 이에 관한 논의는 본 연구 범위에서 벗어난다고 판단하여 본 연구에서는 증상과 병의 구분 없이 주치를 모두 증상으로 처리하여 연구를 진행하기로 하였다.

주치 데이터에는 매우 많은 증상 개념이 있고, 각 개념 간 계층 구조가 복잡하며, 유사한 개념에 다양한 표현 용어가 존재할 뿐 아니라 증상 개념 외에도 시간, 병인, 병기 등과 같은 다양한 증상 발현 조건 개념들이 혼재되어 있다. 기 연구의 주치 데이터의 정제, 분해 방법[18]과 조합 추출 방법[19]에 따르면 동의보감 처방의 주치 종류는 약 4,400여개이고, 약 14,500여개의 조합이 존재하는데, 5계 내과학에 수록된 처방의 종류가 약 400여개, 증상-처방 조합 수는 약 10,000여개인 것에 비하면 주치 조합의 수에 비해 주치 용어의 종류가 많다는 것을 알 수 있다. 따라서 본 연구에서는 다양한 유사 용어를 포괄하고 복잡한 계층 구조를 간략화 할 수 있도록 대표 용어를 선정하였다. 대표 용어의 선정은 주치를 증상 데이터와 조건 데이터(비 증상 데이터)로 나누어 진행하였는데, 증상 데이터는 통증과 관련된 증상들의 조합을 중심으로 살펴보기 위해 疼(동)이나 痛(통)이 들어간 증상 용어들은 모두 痛(통)을 대표 용어로 선정하였고, 그 외의 증상 용어들은 복잡한 표현의 용어는 간략한 표현의 용어로, 하위 개념 용어는 상위 개념 용어로, 유의 개념의 다양한 표현 용어는 그 중 한 가지 용어를 대표 용어로 선정하였다. 조건 데이터는 모든 개념을 다 연구 대상으로 할 수 없기에 병인 개념만 추출하여 대표 병인 용어로 선정하였다.

(2) 증상 데이터에서 대표 용어 선정

증상 데이터에서 대표 용어의 선정은 다음과 같다.

표 2. 증상 데이터에서 선정한 대표 용어의 예  
Table 2. Examples of representative terms in symptom data

	증상 데이터 원문	선정한 대표 용어
상위 개념 용어를 대표 용어로 선정한 경우	嘔逆	嘔吐
	嘔	嘔吐
	痒	瘙癢
	太陽瘡	瘡
	痔	痔瘻
	外痔	痔瘻
	膏淋	淋
	沙石淋	淋
	五淋	淋
	赤帶下	帶下
	熱痞	痞
	偏正頭風	頭風
	癰疽陽證	癰疽
	風痰	痰
간략한 용어를 대표 용어로 선정한 경우	婦人帶下	帶下
	婦人血結胸	血結胸
	室女月經不通	月經不通
	上脰胞腫如桃	上脰胞腫
	腸澀下血作派	腸澀下血
	大惡風寒	惡風寒
	微惡寒飲	惡寒飲
	痰盛	痰
	痰多	痰
	白帶久不止	白帶
	下血如豚肝狀	下血
	小便淋瀝	淋
	淋瀝	淋
	淋瀝不止	淋
	生瘡	瘡
	噫宿腐氣	噫氣
	煩悶欲死	煩悶
泄瀉不止	泄瀉	
다양한 표현 용어 중 하나를 대표 용어로 선정한 경우	有汗	汗
	有熱	發熱
	體熱	發熱
	身熱	發熱
	大便自利	泄瀉
	瀉泄	泄瀉
	大便滑泄	泄瀉
	手不敢近	手不可近
	不可屈伸	不能屈伸
	久無嗣息	無子
	頭目昏暈	眩暈
	目暈	眩暈
	頭暈	眩暈
脈數	脈數	

	증상 데이터 원문	선정한 대표 용어
	脈洪	脈洪
	爬搔	瘙癢
	昏花	眼花
	眼見黑花	眼花
	眼目昏花	眼花
	乳汁壅塞不行	乳汁不行
	渴	口渴
	月經瘀閉	月經不通
	月經不行	月經不通
	大便結燥	大便燥結
	胸煩	心煩

(3) 대표 조건 용어 선정

조건 데이터에서 대표 병인 용어의 선정은 다음과 같이 하였다.

표 3. 조건 데이터에서 선정한 대표 병인 용어의 예  
Table 3. Examples of representative etiology terms in condition data

조건 용어 데이터 원문	선정한 대표 병인 용어
沈寒	寒
寒邪客搏	寒
傷寒陰證	寒
孕婦傷寒	寒
傷寒	寒
風濕相搏	風濕
風濕所侵	風濕
風熱毒氣上攻	風熱
陽明風熱攻注	風熱
風熱毒氣上攻	風熱
傷風見寒脈	風
產後傷風	風
風毒	風
熱邪傳裏	熱
邪熱流入於腎經	熱
熱傷肺氣	熱
浮熱	熱
熱結	熱
氣分之火	火
思慾動火	火
瘀血在太陽經	瘀血
惡血留於脇下	惡血
水濕停蓄	水濕
邪水流注	水
痼冷	冷
傷酒	酒

## 2.2 귀무가설을 이용한 연구 방법

증상 조합의 유의성 판단에 적용 가능한 방법을 찾기 위해 먼저 유의한 증상-치법 조합 도출 연구를 시행하였다. 치법은 증상과 마찬가지로 한의학을 구성하는 중요한 요소 중 하나이며 효능과 연계하여 다양한 활용이 가능하데[20], 치법 용어와 관련하여 최소 의미 단위 분해[17]나 증상-치법 분류체계 연구[16] 등과 같은 기 연구가 있어 방법론의 탐색 및 검증에 적합하다고 판단하였다.

### 1) 증상-치법 조합의 추출

각 교과서[16-20]에는 병증명과 수반증상, 치법, 처방 등의 정보가 존재하는데 각 병증명에 제시된 수반증상군과 치법군에서 단일 개념의 증상-치법의 조합을 추출하였다.

표 4. 風寒感冒에서 추출 가능한 모든 증상-치법 조합  
Table 4. Possible combinations of symptom-treatment method extracted from common cold caused by wind-cold

		치법			
		疏風	散寒	解表	宣肺
증상	惡寒	惡寒-疏風	惡寒-散寒	惡寒-解表	惡寒-宣肺
	微熱	微熱-疏風	微熱-散寒	微熱-解表	微熱-宣肺
	嘔噦	嘔噦-疏風	嘔噦-散寒	嘔噦-解表	嘔噦-宣肺
	流清涕	流清涕-疏風	流清涕-散寒	流清涕-解表	流清涕-宣肺
	咽痒	咽痒-疏風	咽痒-散寒	咽痒-解表	咽痒-宣肺
	咳嗽	咳嗽-疏風	咳嗽-散寒	咳嗽-解表	咳嗽-宣肺
	無汗	無汗-疏風	無汗-散寒	無汗-解表	無汗-宣肺
	頭痛	頭痛-疏風	頭痛-散寒	頭痛-解表	頭痛-宣肺
	小便清長	小便清長-疏風	小便清長-散寒	小便清長-解表	小便清長-宣肺

예를 들어 폐계내과학에는 風寒感冒(풍한감모, 풍한사에 의한 감기)의 수반증상으로 惡寒(오한), 微熱(미열), 嘔噦(분체, 재채기), 流清涕(류청체, 맑은 콧물이 흐름), 咽痒(인양, 목이 가려움), 咳嗽(해수, 기침), 無汗(무한, 땀이 나지 않음), 頭痛(두통), 小便清長(소변청장, 소변색이 맑고 양이 많음)이, 치법으로 疏風(소풍), 散寒(산한), 解表(해표), 宣肺(선폐)가 제시되어 있어 모두 36개의 증상-치법 조합이 추출 가능하다.

### 2) 조합 유의성 분석 방법 연구

#### (1) 귀무가설을 이용한 분석의 필요성

風寒感冒에서 추출 가능한 36개의 증상-치법 조합에는 惡寒-散寒, 無汗-解表와 같이 유의하다고 생각되는 조합과 咽痒-散寒, 頭痛-宣肺와 같이 유의하지 않다고 생각되는 조합이 섞여있다. 만약 증상-치법 조합을 진료지원을 위해 증상에 적절한 처방을 추천하는 것[21]과 같이 활용하고자 한다면 조합 가능한 모든 증상-치법 조합에서 유의한 조합을 구분하는 과정이 필요하다. 이런 과정은 주로 전문가의 판단에 의존해왔기 때문에 모든 데이터에 대해 전문가의 손을 거쳐야만 활용이 가능하다는 문제점이 있었다. 본 연구의 목적은 전문가의 판단을 거치지 않더라도 유의성을 판단할 수 있는 객관적이고 보편적인 방법을 찾는 것인데, 증상-치법 조합이나 증상-증상, 증상-조건, 조건-증상 조합은 문자로 구성된 데이터이므로, 일반적인 통계 분석법을 적용하기 어렵다. 따라서 본 연구에서는 ‘특정 문헌에서 등장 횟수가 많은 조합은 그렇지 않은 조합보다 유의할 가능성이 높을 것이다’ 라는 생각아래 등장 횟수가 많은 조합을 추출하고 이를 한의학 전문가 분석 결과와 비교해 확인해보고자 하였다. 여기서 중요한 것은 많은 것을 정의하는 것인데, 단순히 등장 횟수의 절대 값이 큰 것을 많다고 정의하면 그 유의성과 상관없이

많이 언급된 증상이나 치법이 포함된 증상-치법 조합만이 유의하다고 판단되는 문제가 생기므로 각 증상 별로 많다는 기준을 달리 적용할 필요가 있었다. 따라서 본 연구에서는 “특정 증상에 대하여 상대적으로 등장 횟수가 많은 치법 조합은 그렇지 않은 치법 조합에 비해 유의성이 높다” 라고 가정하고 상대적으로 등장 횟수가 높은 치법 조합을 찾아보았다.

가정에 따라 각 증상별로 등장 횟수가 상대적으로 많은 치법을 찾기 위하여 귀무가설과 대립가설을 이용하였다.

(2) 귀무가설, 대립가설의 설정

귀무가설이란, 통계적 가설검정에서 쓰는 수리통계학 용어 중 하나로, 검정하고자 하는 가설을 대립가설( $H_A$ )로, 그와 반대되는 가설을 귀무가설( $H_0$ )로 설정한 후 귀무가설이 맞지 않음을 통계학적 증거를 통해 증명하여 결과적으로 대립가설이 맞음을 증명하는 것인데, 유의수준을 설정한 후 검정 결과 P-value가 유의수준 미만이면 귀무가설을 기각하고 대립가설을 받아들이고 P-value가 유의수준 이상이면 귀무가설을 받아들인다[22]. 여기서 유의수준이란 귀무가설을 기각할 확률의 크기를 의미하며 흔히 0.05 또는 0.01로 정하여 사용하는데 검정 대상의 특성에 따라 그 값을 달리하며, 적절한 값은 관련 전문가들이 결정한다. 하지만 한의학 콘텐츠의 증상 조합이나 증상-치법 조합의 유의성을 판단하는데 적절한 유의수준에 관한 연구는 아직 시행된 바가 없으므로 본 연구에서는 일반적으로 사용되는 유의수준 값인 0.05로 우선 설정하였다.

본 연구에서의 귀무가설과 대립가설의 설정은 다음과 같다.

특정 증상 P에 대한 치법 조합의 등장 횟수를 각각  $P_n$ 이라하고,

( $P_a$  : 증상P-치법a의 등장 횟수,  $P_b$  : 증상P-치법b의 등장 횟수)

귀무가설( $H_0$ )과 대립가설( $H_A$ )을 다음과 같이 설정하였다.

$H_0$  : 모든  $P_n$ 은 수학적으로 차이가 없다.  
( $P_a=P_b=P_c=...=P_n$ )

$H_A$  : 모든  $P_n$ 이 같지는 않다 [수학적으로 차이가 있는  $P_n$ 이 존재한다]

즉,  $H_0$ 는 특정 증상과 조합되는 모든 치법의 등장 횟수에 수학적으로

크고 작음의 차이가 없다는 것이고,  $H_A$ 는 수학적으로 크거나 작은 차이가 존재한다는 것이다.

(3) 데이터의 구조 및 분석 과정

증상-치법 조합은 먼저 등장 횟수가 상대적으로 많은 증상-치법 조합이 존재하지 않는 경우와 존재하는 경우로 나눌 수 있다. 존재하지 않는 경우는 모든 조합의 등장 횟수가 비슷하게 적거나 많아서 치법들의 등장 횟수에 상대적인 차이가 없는 경우이다.

존재하는 경우는 증상의 특성에 따라 단 1개의 증상-치법 조합만이 다른 증상-치법 조합보다 조합 횟수가 높은 경우부터, 1개를 제외한 모든 조합이 조합 횟수가 높은 경우까지 다양하게 존재할 수 있다. 만약 첫 분석에서 p-value가 0.05 이상이면 귀무가설이 받아들여져서 특정 증상에 대해 치법별 등장 횟수가 수학적으로 차이가 없다는 의미이므로 등장 횟수가 많은 조합이 존재하지 않는 경우에 해당하고, 0.05 미만이면 특정 증상에 대해 치법별 등장 횟수가 수학적으로 차이가 있다는 의미이므로 등장 횟수가 적은 조합과 많은 조합이 함께 있는 경우에 해당한다. 등장 횟수가 많은 조합이 존재하지 않는 경우는 본 연구 대상에 해당하지 않으므로 배제하고, 존재하는 경우에 대해 등장 횟수가 적은 조합이 1개라면 처음 p-value 검사 후 가장 적은 등장 횟수의 조합을 제거하고 다시 p-value 검사를 한다면 그 값은 0.05 이상이 되고, 2개라면 가장 적은 등장 횟수의 조합 2

개를 제거한 후 p-value 검사를 하면 그 값이 0.05 이상이 된다. 이를 이용하여 만약 등장 횟수가 작은 순서대로 차례로 제거해가면서 p-value가 0.05 이상이 될 때까지 p-value 검사를 반복하면 등장 횟수가 상대적으로 큰 조합만 추출할 수 있다. 예를 들어 증상 가와 조합된 치법 a, b, c, d, e의 등장 횟수가 각각 30, 29, 3, 2, 1이고 a와 b의 등장 횟수가 c, d, e에 비해 상대적으로 크다면 p-value는 a, b, c, d, e를 모두 대상으로 했을 때부터 e, d를 제거한 후까지는 0.05 미만이 될 것이며, c를 제거한 후 a, b만을 대상으로 했을 때 비로소 0.05 이상이 될 것이므로 상대적으로 등장 횟수가 많은 치법 a, b를 도출할 수 있다. 이를 순서대로 표현하면 <그림1>과 같다.

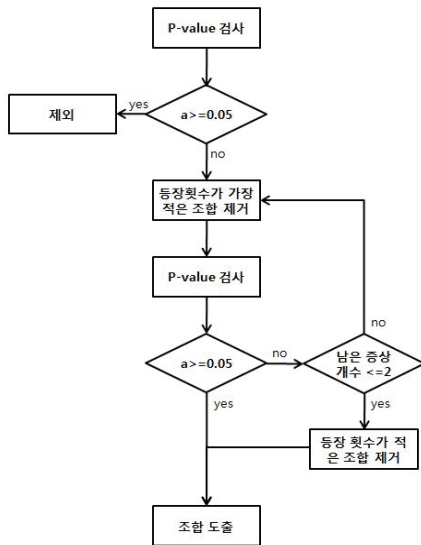


그림 1. 다양한 경우를 고려한 상대적 다빈도 조합 도출 순서도

Figure 1. Flow chart of deducting relative multi frequency combinations on multiple occasions

위의 방법에 따라 통계 분석 프로그램 중 하나인 SAS 9.4를 이용하여 추출 가능한 모든 증상-치법 조합에 대해 분석을 실시하였다.

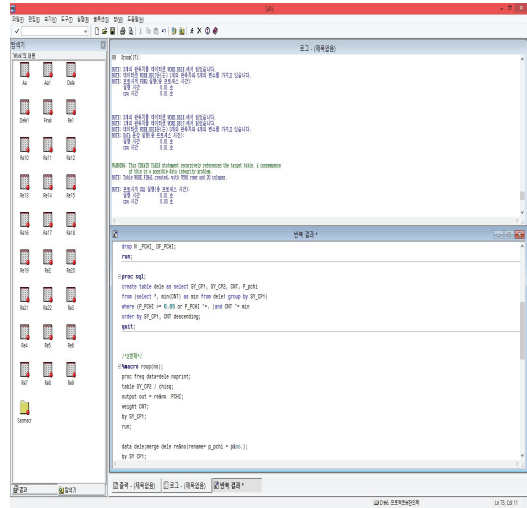


그림 2. SAS coding 화면  
Figure 2. Screen of SAS coding

### 2.3 귀무가설을 이용한 분석의 적절성 검증

귀무가설을 이용한 증상-치법 조합 유의성 분석의 적절성을 검증하기 위하여 복수의 한의학 전문가에 의한 증상-치법 조합의 유의성 분석 결과를 활용하였다. 기본 자료는 온톨로지 연구 과정에서 사용한 증상-치법의 데이터를 활용하였으며, 기존에 입력된 용어를 개념 단위 용어로 정제하여 증상-치법 조합을 재생성하여 한의학 전문가에게 제시하고 조합의 유의성 여부에 대하여 판정하도록 하여 기준으로 삼았다. 예를 들면, 폐계내과학에는 風寒感冒의 수반증상으로 惡寒, 微熱, 噴嚏, 流鼻涕, 咽痒, 咳嗽, 無汗, 頭痛, 小便清長이, 치법으로 疏風, 散寒, 解表, 宣肺이 제시되어 있어 증상 9개와 치법 4개가 상호 조합되어 36개의 증상-치법 조합이 생성될 수 있다. 이와 같은 조합을 대상으로 전문가들이 판정한 결과와 귀무가설을 이용한 분석 결과를 비교하여 귀무가설을 이용한 조합 유의성 분석 방법의 적절성을 검증하도록 하였다.

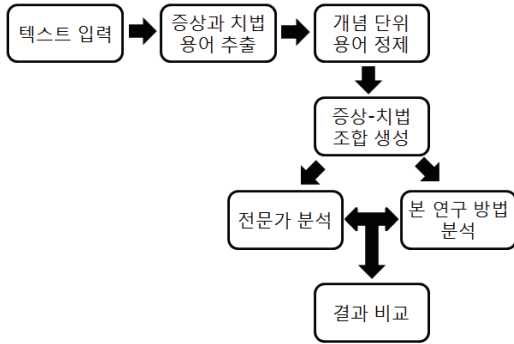


그림 3. 귀무가설을 이용하여 도출한 증상-치법 조합의 적절성 검증 과정

Figure 3. Appropriacy verification process of symptom-treatment method combinations using null hypotheses

귀무가설을 이용한 수학적 방법으로 얻어진 상대적 다빈도 증상-치법 조합은 1,589개인데, 이를 다수의 한의학전문가에 의해 시행된 유의한 증상-치법 조합 분석 결과[16]와 비교해 보니 그 일치율이 조합의 등장 횟수가 10회 이상일 때 43%이었고 등장 횟수가 많아질수록 점차 상승하여 30회 이상일 때 60%, 60회 이상일 때 88%, 66회 이상일 때 100%가 되었다. 이 중 등장 횟수가 60회 이상인 조합들을 대표로 살펴보면 <표 5>와 같다.

표 5. 등장 횟수가 60회 이상인 상대적 다빈도 조합의 전문가 유의성 분석 결과와의 비교  
Table 5. Comparison of relative multi frequency combinations above 60 times and expert analysis results

상대적 다빈도 증상-치법 조합	등장 횟수	전문가 유의성 분석 결과와의 일치 여부
腹大堅滿-化瘀	60	일치함
腹痛-清熱	65	일치하지 않음
偏身麻木-活血	73	일치함
高熱-清熱	74	일치함
高熱-解毒	74	일치함
大便色黑-活血	116	일치함
大便色黑-化瘀	118	일치함

이를 통해 상대적 다빈도 조합이 전문가 분석을 통한 유의 조합과 상당부분 일치하며 등장 횟수가 늘어날수록 그 일치율이 증가하는 것을 알 수 있어 분석 대상의 규모가 커질수록 “특정 증상 대하여 등장 횟수가 상대적으로 많은 치법 조합은 그렇지 않은 치법 조합에 비해 유의성이 높다” 라는 가정이 어느 정도 맞음을 확인할 수 있었다. 따라서 귀무가설을 이용한 수학적 분석 방법이 한의학 콘텐츠에 적용 가능한 방법이라 판단하였다.

## 2.4 귀무가설을 이용한 유의 증상 조합 탐색 연구

### 1) 증상 조합 추출

麻黃湯(마황탕)의 경우 동의보감에 그 주치는 太陽傷寒 頭痛 身疼 百節痛 無汗 惡風寒(태양상한 두통 신동 백절통 무한 오통한, 태양상한으로 머리와 몸과 관절이 모두 아프고 땀이 나지 않으며 오한을 느끼고 바람이 싫은 것)이라 되어있는데, 이를 분해, 정제하여 대표 용어로 변환하면 寒(한), 痛(통), 無汗(무한), 惡風寒(오통한)의 용어가 추출 가능하여 12개의 증상 조합이 추출 가능하다.

표 6. 麻黃湯에서 추출 가능한 모든 증상 조합  
Table 6. Possible combinations of symptoms extracted from Mahwangtang

	寒	痛	無汗	惡風寒
寒	X	寒-痛	寒-無汗	寒-惡風寒
痛	痛-寒	X	痛-無汗	痛-惡風寒
無汗	無汗-寒	無汗-痛	X	無汗-惡風寒
惡風寒	惡風寒-寒	惡風寒-痛	惡風寒-無汗	X

이와 같은 방법으로 총 3,818개의 증상 조합을 추출하였으며 그 예는 <표7>과 같다.

표 7. 통증 관련 처방에서 추출 가능한 증상 조합의 예  
Table 7. Possible combinations of symptoms extracted from formula related with pain

口乾-心煩, 口乾-盜汗, 口乾-胸滿, 口乾-脈數, 吞酸-嘈雜, 吞酸-嘔吐, 吞酸-惡心, 吞酸-泄瀉, 吞酸-痛, 吞酸-痰, 咳嗽-咽痒, 咳嗽-多臥少起, 咳嗽-寒, 咳嗽-熱, 咳嗽-痛, 咳嗽-痰, 咳嗽-發熱, 咳嗽-盜汗, 咳嗽-聲重, 咳嗽-胸滿, 咳嗽-項強, 咳嗽-風, 咳嗽-風寒, 咳嗽-鼻塞, 嘔吐-噎痞, 嘔吐-噤口痢, 嘔吐-四肢冷, 嘔吐-多臥少起, 嘔吐-好食酸鹹, 嘔吐-心悸, 嘔吐-心煩, 嘔吐-怠惰嗜臥, 嘔吐-惡心, 嘔吐-手足熱, 嘔吐-暑風, 嘔吐-泄瀉, 嘔吐-濕, 嘔吐-熱, 四肢緩縱-痛, 四肢緩縱-皮膚不仁, 四肢緩縱-肌肉重着, 四肢冷-泄瀉, 四肢冷-痛, 四肢冷-身重, 大便不利-小便不利, 大便不利-痛, 大便不利-癥瘕, 心煩-發熱, 心煩-目赤, 心煩-眩暈, 心煩-眼花, 心煩-癩瘦, 心煩-風, 心煩-鼻衄, 心煩-惡聞食氣, 怔忡-痛, 怔忡-發熱, 怔忡-盜汗, 怔忡-耳鳴, 怔忡-遺精, 怔忡-驚悸, 怔忡-氣鬱, 怔忡-涎, 怔忡-痛, 怔忡-眩暈, 惡寒-手足寒, 惡寒-拘急, 惡寒-時氣瘟疫, 惡寒-汗, 惡寒-沈困無力, 惡寒-無汗, 惡寒-痛, 惡寒-發熱, 惡寒-脈沈, 惡寒-腸澀下血, 惡寒-自汗, 惡寒-身涼, 惡寒-風, 惡寒-風寒, 惡寒-風濕寒, 惡寒-麻木不仁, 惡寒-惡風, 惡心-泄瀉, 惡心-痛, 惡心-痰, 惡心-肌瘦, 惡心-霍亂, 惡心-面黃 ………
---

이 증상 조합을 대상으로 증상-치법 조합 분석과 마찬가지로 유의수준 0.05에서 본 연구 방법을 적용하여 유의할 것으로 기대되는 조합을 얻었다.

2) 결과

귀무가설을 이용한 수학적 분석을 통해 유의성 있는 증상-증상 조합 22개, 증상→조건 조합 6개, 조건→증상 조합 11개로 총 39개의 증상 조합을 얻었다.

표 8. 본 연구 방법론을 통해 유의할 것으로 기대되는 증상 조합

Table 8. Symptom combinations expected to be significant based on this study's research method

증상-증상	痛-泄瀉, 泄瀉-痛, 痛-發熱, 嘔吐-痛, 痛-嘔吐, 腎膜-痛, 痛-眩暈, 癰疽-痛, 瘡-痛, 無汗-痛, 眩暈-痛, 無汗-惡寒, 腳氣-痛, 無汗-發熱, 痢-痛, 無汗-脈浮, 淋-痛, 無汗-脈緊, 發熱-痛, 無汗-項強, 疝氣-痛, 疝氣-奔豚
증상→조건	痛→寒, 痛→風, 腳氣→濕熱, 無汗→風寒, 無汗→寒, 無汗→風
조건→증상	痰→痛, 熱→痛, 風→痛, 風濕→痛, 風熱→痛, 風寒→痛, 寒→痛, 濕熱→痛, 食積→痛, 氣滯→痛, 濕熱→腳氣

3. 고찰

본 연구에서는 한의학 진단에서 중요한 의미를 갖는 유의한 증상 조합을 얻을 수 있는 객관적이고 보편적인 방법을 찾고자 하였다. 하지만 의미있는 조합을 찾는데 보편적으로 사용되고 있는 검증 방법들은 주로 숫자 데이터를 대상으로 하고 있어 문자로 구성된 증상 조합 데이터에는 적용하기 어려웠다. 문자로 구성된 데이터에서 유의한 조합을 찾는 연구에 관해서는 귀무가설을 통해 유의한 증상-치법 조합을 찾는 사전 연구가 있었다[19]. 본 연구에서는 기존 연구의 프로세스를 다소 개선하고, 이를 증상 데이터에도 적용하여 유의한 증상 조합을 찾는 연구를 수행하고자 하였다. 본 연구에서는 특정 문헌에서 등장 횟수가 많은 조합은 그렇지 않은 조합보다 유의할 가능성이 높다는 기존 연구의 생각아래 등장 횟수가 많은 조합을 추출하고자 하였는데, 많다는 것을 각 증상의 특성에 상관없이 단순히 절대 값이 큰 것으로 정의하면 유의성을 떠나 단지 많이 언급된 증상이 포함된 조합이 유의하다고 판단되는 한계점이 있을 수 있다. 예를 들어 본 연구에서 가장 많은 등장 횟수를 기록한 寒-痛 조합은 그 등장 횟수가 36회로 절대 값이 큰 만큼 중요한 조합이라 할 수 있지만, 등장 횟수가 7회인 氣滯-痛(기체-통, 기가 통하지 않음-통증) 조합도 비록 그 절대 값은 작으나 氣滯와 조합된 병증들 중에는 등장 횟수가 가장 크므로 이 역시 중요한 조합이라 할 수 있어 조합의 유의성을 판단할 때는 절대적 다빈도 만큼이나 상대적 다빈도 역시 중요하다는 것을 알 수 있다. 따라서 본 연구에서는 이러한 점을 반영하기 위해 각 증상별로 많다는 정의를 상대적으로 내릴 수 있는 방법을 고려하였다. <그림 4>와 <그림 5>는 각각 發熱(발열), 腳氣(각기, 다리가 나무처럼 뻣뻣해지는 병증)와 함께 등장한 조합 중 상대적으로 등장횟수

가 큰 조합을 도출하는 과정을 보여주는데, 發熱-寒(발열-한) 조합과 脚氣-濕熱(각기-습열) 조합은 조합 횟수의 절대 값이 각각 7회와 4회이지만 發熱-寒 조합은 조합 횟수의 상대 값이 작아 유의하지 않다고 도출된 경우이고 脚氣-濕熱 조합은 조합 횟수의 상대 값이 커서 유의하다(T로 표시)고 도출된 경우이다. 이는 연구 대상에서 發熱의 등장 횟수 절대 값이 脚氣보다 크기 때문이다. 즉, 發熱은 많이 언급된 증상이므로 發熱과 유의한 조합을 이루려면 脚氣와 유의한 조합을 이루기위해서 보다 많은 조합수가 필요한 것이다.

4805	發熱	痛	26	T	2.47E-44	1.17E-15	2.49E-10	1.18E-08	3.97E-07	0.000941
4806	發熱	寒	7		2.47E-44	1.17E-15	2.49E-10	1.18E-08	3.97E-07	0.000941
4807	發熱	惡寒	5		2.47E-44	1.17E-15	2.49E-10	1.18E-08	3.97E-07	
4808	發熱	惡寒	5		2.47E-44	1.17E-15	2.49E-10	1.18E-08	3.97E-07	
4809	發熱	無汗	5		2.47E-44	1.17E-15	2.49E-10	1.18E-08	3.97E-07	
4810	發熱	風寒	4		2.47E-44	1.17E-15	2.49E-10	1.18E-08		
4811	發熱	風	4		2.47E-44	1.17E-15	2.49E-10	1.18E-08		
4812	發熱	脉浮	3		2.47E-44	1.17E-15	2.49E-10			
4813	發熱	煩渴	3		2.47E-44	1.17E-15	2.49E-10			
4814	發熱	脉緊	2		2.47E-44	1.17E-15				
4815	發熱	痲	2		2.47E-44	1.17E-15				
4816	發熱	口渴	2		2.47E-44	1.17E-15				
4817	發熱	自汗	2		2.47E-44	1.17E-15				
4818	發熱	咳嗽	2		2.47E-44	1.17E-15				
4819	發熱	汗	2		2.47E-44	1.17E-15				
4820	發熱	頭痛	1		2.47E-44					
4821	發熱	耳鳴	1		2.47E-44					
4822	發熱	飲食無味	1		2.47E-44					
4823	發熱	遺精	1		2.47E-44					
4824	發熱	外感內傷	1		2.47E-44					
4825	發熱	嘔逆	1		2.47E-44					
4826	發熱	泄瀉	1		2.47E-44					
4827	發熱	惡風	1		2.47E-44					
4828	發熱	暈明中風	1		2.47E-44					
4829	發熱	眼花	1		2.47E-44					
4830	發熱	心腎虛	1		2.47E-44					
4831	發熱	心煩	1		2.47E-44					
4832	發熱	身無寒	1		2.47E-44					
4833	發熱	身涼	1		2.47E-44					
4834	發熱	時疫難	1		2.47E-44					
4835	發熱	癩泄	1		2.47E-44					
4836	發熱	少陰中風	1		2.47E-44					
4837	發熱	小便不瀉	1		2.47E-44					
4838	發熱	泄瀉	1		2.47E-44					

그림 4. 등장 횟수의 절대 값이 크더라도 상대 값은 작은 조합의 예(發熱-寒)

Figure 4. Examples of combinations with high absolute value of appearance frequency but low relative value (fever-cold)

본 연구에서는 귀무가설을 이용하여 상대적으로 등장 횟수가 큰 조합을 찾았는데, 유의수준을 타

3007	脚氣	痛	10	T	9.52E-05	0.038774	0.108809			
3008	脚氣	濕熱	4	T	9.52E-05	0.038774	0.108809			
3009	脚氣	筋脉拘攣	2		9.52E-05	0.038774				
3010	脚氣	痿躄	1		9.52E-05					
3011	脚氣	步履艱苦	1		9.52E-05					
3012	脚氣	便尿	1		9.52E-05					
3013	脚氣	白虎歷節	1		9.52E-05					
3014	脚氣	筋攣	1		9.52E-05					
3015	脚氣	肝腎虛	1		9.52E-05					
3016	脚氣	行步不隨	1		9.52E-05					
3017	脚氣	風濕	1		9.52E-05					
3018	脚氣	癱瘓麻木	1		9.52E-05					
3019	脚氣	足弱病	1		9.52E-05					
3020	脚氣	阻隔	1		9.52E-05					
3021	脚氣	赤腫	1		9.52E-05					

그림 5. 등장 횟수의 절대 값이 작더라도 상대 값은 큰 조합의 예(脚氣-濕熱)

Figure 5. Examples of combinations with low absolute value of appearance frequency but high relative value (limb flaccidity-dampness heat)

분야에서 일반적으로 사용되고 있는 0.05로 설정하였다. 유의수준은 분석 대상의 특성에 따라 관련 전문가들이 적절한 값을 정해야 하는 것인데 한의학 콘텐츠의 조합 분석에 있어 적절한 유의수준 값은 아직 알려진 바가 없다. <그림 6>은 眩暈(현훈, 어지러움)과 상대적으로 등장 횟수가 많은 조합을 도출하는 과정을 보여주는데, 유의수준이 0.05인 본 연구에서는 痛만이 많이 조합되었다고 판단되었지만, 유의수준을 0.01로 설정한다면 痛뿐만 아니라 嘔吐(구토)와 心煩(심번, 가슴이 두근거림) 역시 상대적으로 등장 횟수가 많은 조합으로 판단될 것이다.

유의수준은 본 방법을 통해 얻어진 유의한 증상 조합을 어떤 목적으로 사용하느냐에 따라 달라질 수 있다. 예를 들어 한의사의 진료 과정에 도움을 주는 진료 지원 시스템 등에서 활용한다면 보다 엄격한 기준을 적용해야 할 것이며, 일반인을 대상으로 한의학 콘텐츠의 검색 성능을 높이고자 시맨틱 검색 엔진 개발 등에 활용한다면 다소 완화된 기준을 적용할 수 있을 것이므로 향후 목적에 맞는 적절한 유의수준을 찾는 연구가 필요하다.

2624	眩暈	痛	18 T	3.02E-20	2.43E-07	0.001558	0.027807
2625	眩暈	嘔吐	7	3.02E-20	2.43E-07	0.001558	0.027807
2626	眩暈	心煩	3	3.02E-20	2.43E-07	0.001558	
2627	眩暈	惡心	2	3.02E-20	2.43E-07		
2628	眩暈	神思不清	2	3.02E-20	2.43E-07		
2629	眩暈	頭風	2	3.02E-20	2.43E-07		
2630	眩暈	痰	2	3.02E-20	2.43E-07		
2631	眩暈	耳鳴	1	3.02E-20			
2632	眩暈	元氣虛冷	1	3.02E-20			
2633	眩暈	惡阻痛	1	3.02E-20			
2634	眩暈	涎	1	3.02E-20			
2635	眩暈	惡露不盡	1	3.02E-20			
2636	眩暈	惡聞食氣	1	3.02E-20			
2637	眩暈	心悸	1	3.02E-20			
2638	眩暈	神昏	1	3.02E-20			
2639	眩暈	身重	1	3.02E-20			
2640	眩暈	神思昏憒	1	3.02E-20			
2641	眩暈	產後去血暈	1	3.02E-20			
2642	眩暈	產後	1	3.02E-20			
2643	眩暈	泄瀉	1	3.02E-20			
2644	眩暈	四肢厥冷	1	3.02E-20			
2645	眩暈	脾胃虛	1	3.02E-20			
2646	眩暈	鼻淵	1	3.02E-20			
2647	眩暈	婦人血風	1	3.02E-20			
2648	眩暈	發寒熱	1	3.02E-20			
2649	眩暈	發熱	1	3.02E-20			
2650	眩暈	目眩	1	3.02E-20			
2651	眩暈	目不聞	1	3.02E-20			
2652	眩暈	面鳴	1	3.02E-20			
2653	眩暈	羸瘦	1	3.02E-20			
2654	眩暈	眼花	1	3.02E-20			
2655	眩暈	動氣躁發汗	1	3.02E-20			
2656	眩暈	痰欬	1	3.02E-20			
2657	眩暈	多眠少起	1	3.02E-20			

그림 6. 眩暈과 상대적으로 등장 횟수가 큰 조합을 도출하는 과정

Figure 6. Process of deduction on dizziness and relatively frequently appeared combinations

본 연구는 개별 한의학전문가에 의해 분석되던 한의학 콘텐츠 조합의 유의성을 이미 타 분야에서 사용되고 있는 수학적 방법을 통해 분석함으로써 객관성과 보편성을 확보하고, 새로 축적되는 데이터를 시스템적으로 분석할 수 있게 함으로써 규모가 커져 몇몇 전문가의 노력만으로는 분석이 불가능한 빅데이터나 실시간으로 데이터가 축적되는 임상 빅데이터 등에 적용이 가능하고자 하였다. 다만 본 연구방법은 “특정 연구 대상에서 등장 횟수가 상대적으로 많은 조합이 그렇지 않은 조합보다 유의할 가능성이 높다”는 가정에서 벗어난 조합은 찾을 수 없으며, 등장 횟수가 상대적으로 많진 않으나 유의한 조합 역시 찾을 수 없다는 한계가 있어 가정을 검증하는 연구나 가정 외의 조합

도 찾을 수 있는 연구가 요구된다. 기존에 한정되어 있는 인터페이스 레이아웃을 자신이 원하는 인터페이스 레이아웃으로 적용하여 사용할 수 있는 시스템을 개발진행 하였다. 언제, 어디서든지 네트워크를 통해 사용자가 원하는 인터페이스 레이아웃을 적용할 수 있기 때문에, 스마트 모바일을 사용하는데 있어 최적의 환경을 만들어 준다. 또한, 장애인, 노인, 어린이와 같이 특정계층에게 맞는 인터페이스 레이아웃을 제공할 수 있기 때문에, 보다 스마트 모바일은 사용하는데 있어 불편함을 많이 줄여줄 것이다. 인터페이스 레이아웃의 버튼은 사용자가 원하는 만큼 크기를 늘리고 줄일 수 있다.

#### 4. 결론

본 연구에서는 한의학 전문가들의 분석 과정을 대신하여 빅데이터나 실시간으로 데이터가 축적되는 임상 빅데이터 등에 적용이 가능한 유의성 있는 증상 조합을 분석할 수 있는 방법을 찾고자 하였다. 이에 다음과 같은 결론을 얻었다.

1. 상대적 다빈도 조합은 전문가의 조합 유의성 분석 결과와 어느 정도 일치하여, 목적에 따라 상대적 다빈도 조합을 유의한 조합으로 판단하고 활용할 수 있음을 알 수 있었다.
2. 문자로 구성된 조합의 상대적 다빈도 조합을 구하는데 있어 귀무가설을 이용한 본 연구 방법은 적절하였다.
3. 통증의 치료와 관련된 처방에서 찾아낸 유의성 있는 증상-증상, 증상→조건, 조건→증상 조합은 39개가 있었다. 본 연구가 『東醫寶鑑』에서 통증 관련 처방의 정보에 국한하여 진행된 것이라는 한계가 있으나, 이는 연구 대상 데이터의 영역을 확대하는 것으로 극복할 수 있다.
4. 다만 “특정 연구 대상에서 등장 횟수가 상대적으로 많은 조합이 그렇지 않은 조합보다 유의할

가능성이 높다” 는 가정에서 벗어난 조합은 찾을 수 없으며, 등장 횟수가 상대적으로 많진 않으나 유의한 조합 역시 찾을 수 없다는 한계가 있어 가정을 검증하는 연구나 가정 외의 조합도 찾을 수 있는 연구가 지속되어야 한다.

## References

- [1] H-C. Jang, *Ontology-based traditional Korean medicine knowledge framework*, Korea Institute of Oriental Medicine, 2013.
- [2] W-S. Cha, T-H. Lee, B-W. Lee, *Analysis of herbal combination frequency on clinical herbal formulation*, The Journal of Herbal Formula Science. Vol. 19, No. 2. pp. 1-10, 2011.
- [3] B-S. Park, E-H. Kim, S-A. Lee, B-W. Lee, *The comparative study between the effect of herbal formula in schoolbook and the effect deduced from compositional herbal effects*, Journal of Oriental Medical Classics. Vol. 21, No. 1. pp. 88-89, 2008.
- [4] W-Y. Park, J-U. Baek, and B-W. Lee, *Analysis of Prescriptions from somunsumyungronbang and nansilbijang*, Journal of Korean Society of Medical History. Vol. 24, No. 2. p. 117, 2011.
- [5] J-U. Baek, and B-W. Lee, *A study on the frequencies of medicinal herb combinations in the prescriptions of Bangyakhappyeon*, Journal of Oriental Medical Classics. Vol. 24, No. 4. pp. 55-67, 2011.
- [6] S-W. Shin, J-U. Baek, and B-W. Lee, *Study on the frequencies of medicinal herb combinations in the prescriptions of wenreijingwei*, Journal of Physiology & Pathology in Korean Medicine. Vol. 25, No. 5. pp. 759-764, 2011.
- [7] J-M. Park, B-C. Shin, G-H. Heo, and B-W. Lee, *The comparison and consideration of indications of herbal medicine through analysis about insured herbal extracts and clinical prescriptions - Focusing on Bojungikgi-tang -*, Korean journal of oriental preventive medical society. Vol. 17, No. 2. pp. 129-138, 2013.
- [8] D-S. Park, B-K. Lee, and B-W. Lee, *A study on the comparative method of prescription using herb weight ratio*, The Journal of Herbal Formula Science. Vol. 21, No. 2. pp. 121-132, 2013.
- [9] B-W. Lee, B-I. Seo, D-M. Um, and S-C. Kim, *A study on formula analysis using single concepted herbal effect database*, Journal of Oriental Medical Classics. Vol. 20, No. 4. pp. 319-324, 2007.
- [10] Joint Text Compilation Committee for Korean Internal Medicine(Liver), *Korean internal medicine(Liver)*, Institute of Oriental Medicine, 2001.
- [11] Joint Text Compilation Committee for Korean Internal Medicine(Heart), *Korean internal medicine(Heart)*, Gunja Publishing Ltd., 2006.
- [12] Joint Text Compilation Committee for Korean Internal Medicine(Spleen), *Korean internal medicine(Spleen)*, Gunja Publishing Ltd., 2009.
- [13] Joint Text Compilation Committee for Korean Internal Medicine(Lung), *Korean internal medicine(Lung)*, Nado Publishing Ltd., 2007.
- [14] H-G. Park, *Korean internal medicine(Kidney)*, Sungbo Publishing Ltd., 2006.
- [15] J. Heo, *Donguibogam*, BubIn Publishers

- Company, 2009.
- [16] Y-T. Oh, A-N. Kim, S-K. Kim, J-S. Seo, and H-C. Jang, *Study on systematizing the combination of method of treatment and symptoms using the basic traditional medicine theory*, Journal of Physiology & Pathology in Korean Medicine. Vol. 27, No. 4. pp. 383-390, 2013.
- [17] Y-T. Oh, B-U. Jeon, H-C. Jang, C-S. Kim, J-H. Kim, S-K. Kim, and M-Y. Song, *Study of connections between efficacy terms and remedy terms by conversion the raw terms into the representative terms*, Journal of Oriental Medical Classics. Vol. 24, No. 5. pp. 99-105, 2011.
- [18] A-N. Kim, Y-T. Oh, S-K. Kim, S-H. Kim, and H-C. Jang, *A study of disassembling major indication terms into minimum meaning units and linking to diseases*, The Journal of Herbal Formula Science. Vol. 21, NO 2. pp. 74-75, 2013.
- [19] Y-T. Oh, B-R. Nam, and A-N. Kim, *Study on analysis of significance of symptom-treatment method combination*, Journal of Applied Mathematics & Informatics. Vol. 32, No. 5-6. pp. 737-746, 2014.
- [20] Y-T. Oh, A-N. Kim, S-K. Kim, and H-C. Jang, *Study on the ontology linking by acupuncture terms*, The Journal of the Korea Contents Association Vol. 14, NO. 9. pp. 565, 2014.
- [21] Y-T. Oh, A-N. Kim, S-K. Kim, and H-C. Jang, *A study on diagnosis support using knowledge of diseases from literature*, Journal of Daejeon university Traditional Korean Medicine Institute. Vol. 23, No. 1. pp. 13-20, 2014.

- [22] T-M. Song, and J-Y Song, *Big data analysis methodology for science research*, Hannarae publishing Company, 2013.

---

## 귀무가설을 이용한 유의한 증상 조합 도출 방법 연구

오용택<sup>1</sup>, 김안나<sup>2</sup>

<sup>1</sup>아이누리한의의원

<sup>2</sup>한국한의학연구원 미병연구단

---

### 요 약

본 연구에서는 한의학 전문가들의 분석 과정을 대신하여 빅데이터나 실시간으로 데이터가 축적되는 임상 빅데이터 등에 적용이 가능한 유의성 있는 증상 조합을 분석할 수 있는 방법을 찾고자 하였다.

귀무가설을 이용한 수학적 방법으로 상대적 다빈도 증상-치법 조합을 찾은 후 이를 전문가의 증상-치법 조합 유의성 분석 결과와 비교해보니, 상대적 다빈도 조합은 전문가의 조합 유의성 분석 결과와 어느정도 일치하여, 목적에 따라 상대적 다빈도 조합을 유의한 조합으로 판단하고 활용할 수 있음을 알 수 있었다. 따라서 문자로 구성된 조합의 상대적 다빈도 조합을 구하는데 있어 귀무가설을 이용한 수학적 방법은 적절하다고 판단하였다. 본 연구 방법을 『東醫寶鑑』에서 통증의 치료와 관련된 처방에 적용한 결과, 유의성 있는 증상-증상, 증상→조건, 조건→증상 조합 39개를 얻을 수 있었다. 본 연구가 통증 관련 처방의 정보에 국한하여 진행된 것이라는 한계가 있으나, 이는 연구 대상 데이터의 영역을 확대하는 것으로 극복할 수 있다. 다만 “특정 연구 대상에서 등장 횟수가 상대적으로 많은 조합이 그렇지 않은 조합보다 유의할 가능성이 높다”는 가정에서 벗어난 조합은 찾을 수 없으며, 등장 횟수가 상대적으로 많진 않으나 유의한 조합 역시 찾을 수 없다는 한계가 있어 가정을 검증하는 연구나 가정 외의 조합도 찾을 수 있는 연구가 지속되어야 한다.

---



**Yong Taek Oh** received a Doctor's degree of Traditional Korean Medicine from Daegu Haany University. From 2010 to 2014, he was a researcher at

Korea Institute of Oriental Medicine. He has been at Inuri Oriental medical clinic since 2015. His current research interests include Korean Traditional Medicine, Ontology and Medical treatment support system.

*E-mail address:* ydydxor@gmail.com



**An Na Kim** received a Master degree of Traditional Korean Medicine from Woo Suk University. She has been a researcher in the Mibyeong Research

Center at Korea Institute of Oriental Medicine since 2009. Her current research interests include Korean Traditional Medicine, Ontology and Medical treatment support system.

*E-mail address:* ankim2012@kiom.re.kr