Optimization Model Research on Major Underlying Factors in the Subhealth Condition Evaluation in 1 City and 7 Provinces in China

Hong-Mei Ni*, Yu-Min He†, Xu-Ming Yang‡
*Basic Medical College, Shanghai University of Traditional Chinese Medicine, †College of Acupuncture and Massage, Shanghai University of Traditional Chinese Medicine, Shanghai, China

Abstract

**Background:** The study aimed to analyze major underlying factors of the subhealth condition evaluation and find the optimization model. **Methods:** Selected 524 cases of health state and 453 cases of subhealth state from the research objects. A genetic algorithm was applied to discover the optimization model. The decision tree algorithm was used to find the main performance in the areas of physical, psychological, and social adaptation in the two populations which were the health state and the subhealth state. **Conclusions:** To establish the optimization model, the author would set up a curve-fitting equation between reduction of health self-assessment score (S-G1) and white blood cell (WBC) value in routine blood, so as to establish the relationship between S-G1 and WBC and found the approximate minimum solution of each equation. Besides, the author would analyze the differences between two populations in WBC examination to seize the difference. Revealed the differences of two populations in the areas of physical, psychological, and social adaptation by data mining and got the result that WBC of “the health state population” was higher than that of “the sub-health state population” in the index changes of WBC. The problem in social adaptation area of the subhealth state population was more serious in degree than the health state population; the reason was complex. If WBC of the health state population was near or below 5.5079, the object may be in the state of subhealth. However, if WBC was near or below 4.35, it is possible to enter the “morbid state.”

**Keywords:** Curve fitting, decision tree, genetic algorithm, health state, subhealth state, white cell count value

**INTRODUCTION**

N. Berkman, the former Soviet union, first put forward “the third state” concept.[1] Since then, many domestic and foreign scholars had discussed and put forward a similar name. Chinese scholars named it “sub-health” in the 1980s.[2] In response to this new concept, Chinese scholars published articles to explain the subhealth state.[3,4] Due to the wide range of subhealth state, the evaluation of subhealth state had become the focus of subhealth study. The assessment methods needed multidimensional and multilevel study. At present, the main method of the health condition evaluation was the combination of scale analysis and biological indicators.[5] The subhealth state could be evaluated synthetically by exploring the correlation of important variables in detection indicators and subhealth state measurement.

**METHODS**

During the period from 2006 to 2012, Shanghai University of Traditional Chinese Medicine carried out the research on “sub health category and measurement standards research.” A test was carried out in 1 city and 7 Provinces in China, which...
were located in Shanghai, Shanxi, Jiangsu, Yunnan, Sichuan, Hebei, Henan, and Anhui. Fifteen thousand people, aged from 20 to 60 years, were surveyed. Based on the principles and methods of scale development, a subhealth assessment form was developed, which had good reliability and validity. The questionnaire came from “the Chinese sub-health state measurement scale (CHSHS-2)” which was provided by Shanghai University of Traditional Chinese Medicine. It contained a total of 78 entries, four areas (body performance, psychological performance, social adaptation, and sex life) and 20 aspects (fatigue, digestion, sleep, dysfunction, immune disorders, allergies, aging, pain, constipation, depression, anxiety, learning and memory, pressure, satisfaction, adaptability, security, self-confidence, self-realization, social support, sex). The response option of each entry adopted level 5 Likert method for quantitative determination: 1 not at all, 2 a little, 3 moderately, 4 mostly, 5 completely, which represented the frequency of 0, 25%, 50%, 75%, 100% respectively. The degree of subjective symptom could be divided into 5 levels: “Never” assignment 1; “rarely” assignment 2; “general” assignment 3; “regular” assignment 4; “always” assignment 5. The higher the score was, the more serious the state of health would be. In addition, the questionnaire also included three self-assessment topics: health self-assessment score (S-G1), psychological health self-assessment score (S-G2), and quality of life. Higher scores (1-100 points) indicated better health state. We Use data mining method for knowledge acquisition [Table 1].

**Results**

This study tested 1055 cases who had physical examinations in Shanghai Shuguang Hospital medical center from 2007 to 2008. All cases should include healthy, disease, and subhealthy people, so people who do not meet the inclusion criteria must be excluded. At the same time, routine blood test was conducted to all the objects which were investigated. This study was approved by the university institutional review boards. Written informed consent was obtained from each subject prior to inclusion in the study. After the data acquisition is completed, the decision tree algorithm is used to excavate the main influencing factors of different healthy people, and the genetic algorithm is used to establish the optimization model to compare the differences of white blood cells (WBCs) in the two populations.

**The decision tree algorithm digging out the main factors in Chinese subhealth state measurement scale**

Classification algorithm was a data analysis method, which was one kind of prediction of data mining. Its aim, based on the important sample data set, was to find out the model which could accurately describe and distinguish data classes or concepts. According to the attribute’s value of entity and other constraints, the model could be divided into the data classes. The current techniques and methods were mainly the algorithm of decision tree classification, Bias classification and Bias network, neural network, genetic algorithm, rough sets, and case-based reasoning. Decision tree learning method was inductive learning method which was based on the given data sample. In the decision tree, from the root to the leaf nodes of a path corresponds to a conjunctive rules, the decision tree corresponds to a set of disjunctive expression rules.

The subhealth state involved in many aspects of the body area: fatigue, digestion, sleep, vegetative nerve disorders (dysfunction), immunity, allergies, aging, and pain. In these aspects, fatigue is the most important performance. There were five entries of fatigue in the CHSHS-2: Did you have fatigue, weakness, and tired in the latest month (ID: F101)? Whether the fatigue could be eased after the break (ID: F102)? Did you feel tired and you just wanted to rest and did not want to move in weekends and holidays (ID: F103)? Did you feel particularly tired when you had pressure (ID: F104)? Did you also feel tired when you had a good mood (ID: F105)?

The psychological area of subhealth state involved in some aspects: depression, anxiety, learning, and memory. There are six entries of depression in the CHSHS-2: Did you have a sense of loneliness (ID: F1001)? Did you often sigh (ID: F1002)? Did you have the desire to talk about depression with others (ID: F1003)? Did you blame yourself (ID: F1004)? Did you feel life meaningless and lose interest in life (ID: F1005)? Did you have a depressed mood in about a month (ID: F1006)? There were five entries of anxiety: Did you have a psychological sense of security (ID: F1101)? Did you feel nervous when something happens (ID: F1102)? Did you feel worried or afraid without reason (ID: F1103)? Did you feel calm in about a month (ID: F1104)? Did you often feel upset in about a month (ID: F1105)? There were three entries of learning and memory: Did you experience the situation that you couldn’t remember what happened recently (ID: F1201)? Were you distracted involuntarily when you did something (ID: F1202)? Were you efficient recently (ID: F1203)? In addition, there was a survey of giving marks on your own mental state of mind: How many grades would you hit (the best state for 100 points) (ID: S-G2)?

The subhealth state involved many aspects in the social adaptation areas: pressure, satisfaction, adaptability, sense of security, self-confidence, self-realization, and social support. Pressure and satisfaction are the representatives among these aspects. There were five entries of pressure: Did you have the feeling what you had done would develop as you would (ID: F1301)? Did you feel that you had many difficulties which could not overcome (ID: F1302)? Did you feel the pressure of survival (ID: F1303)? Did you feel that the society developed so fast that you could not adapt it (ID: F1304)? Were you worried about your future work and life (ID: F1305)? There were two entries of satisfaction: Were you satisfied with your current occupations (ID: F1401)? Were you satisfied with your competence (ID: F1402)?

**The main performance of subhealth state population in the psychological area for decision tree analysis**

Among 1055 cases of physical examination population, a subhealth crowd of 453 cases was obtained by removing the
Ni, et al. Optimization model research on major underlying factors
disease crowd and the health population whose S-G2 were higher than 85. On that basis, a crowd of 313 cases in subhealth state which was obtained by deleting the health population whose psychological score were higher than 85 were treated as the study object. 15 properties, related to mental subhealth state “depression,” “anxiety,” “learning and memory,” were treated as input by decision tree analysis. The output variable was S-G2 < 70 and 70 ≤ S-G2 < 85, which were set to 0 and 1. The dependency network was treated as result by decision tree analysis.

The dependency network intensity from strong to weak was followed by F1105->S-G2; F1006->S-G2; F1001->S-G2.

The entry “Did you often feel upset in about a month? (ID: F1105)” had the most powerful influence on the scores of psychological area, followed by “Did you have a depressed mood in about a month (ID: F1006)?”, and third, “Did you have a sense of loneliness (ID: F1001)” Figure 1 was the weakest dependency network.

The major underlying factors of health state population in the psychological areas for decision tree analysis
Among 1055 cases of being investigated, there were 524 cases of healthy controls, excluding the disease crowd and the subhealth population whose S-G1s were lower than 85. Then, 466 cases of psychological health were obtained as study objects. The decision tree analysis was adopted. 15 properties, related to mental subhealth state “depression,” “anxiety,” “learning and memory,” were treated as input. Then, the objects of 85 ≤ S-G2 < 95, 95 ≤ S-G2 < 100 were reset as 0 and 1. They were treated as output by decision tree analysis. Figure 2 was the dependent network in psychological health population.

We could get only one dependency from the dependency network: F1101->S-G2. Namely, only this entry “Did you have a psychological sense of security? (ID: F1101)” had impact on psychology. Thus, it could be seen that the influence of the subhealth state population, and the health state population in the psychological area was different. So it needed to be treated differently.

Further analysis showed that the performance of subhealth population in psychological area was wider in involvement aspects and deeper in degree. The emotion of “a psychological sense of security,” “upset,” “depressed,” “loneliness” were more common. The health state people just showed a slight anxiety about “a psychological sense of security,” while the people of subhealth state showed more serious symptoms of “upset,” “depressed” even “loneliness with depression characteristics.”

The relevant factors of subhealth state population in the fatigue aspect for decision tree analysis
Fatigue was one of the most common manifestations of body area. Therefore, in the subhealth state population, S-G1 represented S-G1 of the total state of each of the respondents. S-G1 < 70, 70 ≤ S-G1 < 80, 80 ≤ S-G1 < 85 in overall grade were set as 0, 1, 2. Five properties of “the fatigue” in the body area were treated as input, S-G1 were treated as output, analyzed by decision tree. We could get the dependence network shown in Figure 3.

As shown in Figure 3, the entry F101 “Did you have fatigue, weakness, and tiredness in about a month (ID: F101)” had the closest relationship with fatigue in the somatic manifestations of subhealth state population, which was associated with the S-G1 in overall grade.

The relevant factors of health state population in the fatigue aspect for decision tree analysis
In the health state population, S-G1 < 90, 90 ≤ S-G1 < 95, 95 ≤ S-G1 ≤ 100 in overall grade were set as 0, 1, 2. Five properties of “the fatigue” in the body area were treated as input, S-G1 were treated as output, analyzed by decision tree. We could get the dependence network shown in Figure 4.

As shown in Figure 4, the entry F103 “During the weekends or holidays, when you felt tired, did you just want to rest and did not want to move? (ID: F103)” had the closest
relationship with fatigue in the somatic manifestations of health state population, which was related to the S-G1 in overall grade. Thus, it could be seen that the influence of the subhealth state population and the health state population in the psychological area was different. So it needed to be treated differently.

Further analysis showed that the fatigue in subhealth state population was more severe than that of the health state population. What’s more, the duration of fatigue in subhealth state population was longer, while the fatigue of health state population would disappear after a rest.

The relevant factors of subhealth state population in the social adaptation area for decision tree analysis

In the subhealth state population, $S\text{-}G1 < 70, 70 \leq S\text{-}G1 < 80, 80 \leq S\text{-}G1 \leq 85$ in overall grade were set as 0, 1, 2. Seven properties of “the pressure” and “satisfaction” in the social adaptation area were treated as input, S-G1 were treated as output, analyzed by decision tree. We could get the dependence network shown in Figure 5. The strong dependency relationship was $F1305 \rightarrow S\text{-}G1$, and the weak dependency relationship was $F1401 \rightarrow S\text{-}G1$.

As shown in Figure 5, the social adaptation area of subhealth state people mainly involved the entry $F1305$ “Are you worried about your future work and life? (ID: F1305)” and $F1401$ “Were you satisfied with your current occupations? (ID: F1401),” which were related to the S-G1 in overall grade.

The relevant factors of health state population in the social adaptation area for decision tree analysis

In the health state population, $S\text{-}G1 < 90, 90 \leq S\text{-}G1 < 95, 95 \leq S\text{-}G1 \leq 100$ in overall grade were set as 0, 1, 2. Seven properties of “the pressure” and “satisfaction” in the social adaptation area are treated as input, S-G1 were treated as output, analyzed by decision tree. We could get the dependence network shown in Figure 6. The strong dependency relationship was $F1301 \rightarrow S\text{-}G1$; and the weak dependency relationship was $F1302 \rightarrow S\text{-}G1$.

As shown in Figure 6, the social adaptation area of health state people mainly involved the entry $F1301$ and $F1302$, “Did you have the feeling what you had done would develop as your will? (ID: F1301),” “Did you feel that you had many difficulties which could not overcome? (ID: F1302),” which were related to the S-G1 in overall grade. Thus, it could be seen that the influence of the subhealth state population and the health state population in the social adaptation area was different. So it needed to be treated differently.

Further analysis showed that the performance of subhealth population in social adaptation area was worse in involvement aspects than that of health population. The social adaptation of subhealth people was related to “pressure” and “satisfactory,” while the health state people was just related to “pressure.” To some extent, the pressure of health population came from the short term, however, the pressure of subhealth population originated from the long term, such as “future life and work.”

Moreover, the state of subhealth people also involved the aspect of “satisfaction.”
Establishing optimal model of the relevant factors in health and subhealth population by Genetic algorithm

Fatigue was one of the main manifestations of subhealth state. Studies had shown that WBC were related to fatigue. Therefore, all the study objects in this study had already finished self-evaluation and had done the routine blood test.

To analyze the possible differences between S-G1 and WBC in health and subhealth populations, first of all, we did some reduction processing to S-G1 and WBC value. Then, the polynomial curve fitting method was adopted to establish all kinds of equations of “the independent variable is S-G1 (after reduction)” and “the dependent variable was WBC value (after reduction).” Because the relation equation was a nonlinear equation, and this complex equation had no conventional precise solution method, the evolutionary algorithms such as genetic algorithm, immune algorithm, and particle swarm algorithm must be applied in relation equations to obtain the optimal solutions.

Among 1055 cases of being investigated, there were 524 cases of health people, excluding the diseased people and those people whose S-G1 were <85. Then according to the value of S-G1 and WBC value (after reduction) could be obtained after the normalization processing of WBC. In order to determine the correlation between S-G1 and WBC value in health population, the polynomial curve fitting method was adopted to establish the equations of “the independent variable is S-G1 (after reduction)” and “the dependent variable is WBC value (after reduction).” According to the existing data, a quartic equation was obtained as shown: $Y = 0.001027x^4 - 0.378290x^3 + 52.198309x^2 - 3198.025556x + 73398.81207$, which is fitting curve for health S-G1 and WBC value (after reduction) in health population. This equation means the relationship between S-G1 and WBC value in health population. Therefore, when making out the global optimal solution of the relation equation, genetic algorithm was used and worked out as follows: $X = 86.74; Y = 5.51$. Namely, when S-G1 was 86.74, the approximate minimum value of WBC was 5.51.

Similarly, methods based on the above health population, the normalized data of S-G1 and WBC value in sub-health population were shown in Table 3, the fitting curve equation is $y = 0.0000115x^4 - 0.0028810x^3 + 0.6067352x^2 -10.0219783x+143.1585412$, genetic algorithm was used and solved out as follows: $X = 41.10; y = 4.35$. When S-G1 was 41.10, the approximate minimum value of WBC was 4.35. Based on the analysis of the above, it could be discovered that there existed differences between health population and sub-health population. Moreover, in any physical, pathology, and treatment process of biomedicine, a lot of optimization problem, without exception, could be classified as an optimization problem.
as the topic of system state trend, owing to the participation of thermodynamic energy oscillation and system state. Thus, it could be seen, the state of population in subhealth was in the trough, and dropped constantly. But if it was clear that one was in the subhealth state, appropriate traditional Chinese and western medicine treatment and prevention measures could be carried out to promote the improvement of their state. That was an effective treatment and prevention which could drive the system state toward “positive valley.” When S-G1 was 86.74, the approximate minimum value was 5.51 in health people. When S-G1 people was 41.10, the approximate minimum value was 4.35 in subhealth. Obviously, when the WBC value of health state population was close to or below 5.51, it was more likely to show the symptomatic “sub-health state,” which should be paid more attention. In addition, compared to “the health state population,” the approximation minimum of WBC value in “sub-health population” was lower.

**Conclusions**

“The health state population” and “the sub-health state population,” as the objects of study, were selected from the 1055 cases of physical examination population. The decision tree algorithm was adopted to excavate the main related entries in physical area, psychological area, and social adaptation area. The main entries of subhealth state population in the psychological area were: “Did you often feel upset in about a month?” “Did you have a depressed mood in about a month?” “Did you have a sense of loneliness?” However, there was only one entry of health state population in the psychological field: “Did you have a psychological sense of security?” The main related entry of subhealth state population in the body area was: “Did you have fatigue, weakness, and tiredness in about a month?” “During the weekends or holidays, when you felt tired, did you just want to rest and did not want to move?” The main related entries of sub-health population in the social adaptation area: “Were you worried about your future work and life?” and “Were you satisfied with your current occupation?” while the main related entries of health state population were: “Did you have the feeling what you had done would develop as your will?” and “Did you feel that you had many difficulties which could not overcome?” When furthering study the main related factors which were excavated from these areas in different populations, special attention and verification should be given to grasp the differences of different populations.

At the same time, depending on the progression level of such illnesses, divided the process into three stages, i.e., a normal state, a predisesease state (or a critical state), and a disease state. According to this idea, in a healthy and sub-healthy population, the situation of WBC value was compared.

Because each population studied the blood routine examination, to reveal possible differences between S-G1 and WBC value of the routine blood test in healthy and subhealthy populations, the author, first, did some reduction processing to S-G1 and WBC value. Then, the polynomial curve fitting method was adopted to establish all kinds of equations of “the independent variable was S-G1 (after reduction)” and “the dependent variable was WBC value (after reduction)” That is, the relationship between S-G1 and WBC was established.

The modern intelligence optimization algorithm – genetic algorithm was chosen to apply in relation equation to obtain the optimal solutions. According to the result, when WBC value of health state population was close to or below 5.51, it was more likely to show the symptom of “sub-health state,” which should be paid more attention. But if WBC value of subhealth state population was close to or below the minimum 4.35, then this person might be “sick.”

In summary, in order to describe the underlying dynamical mechanism of complex human body health changes, their evolutions are often modeled as time-dependent nonlinear dynamical system. We evaluated the subhealth state by mining the different main performance of health population and the subhealth population, and established an evaluation model between S-G1 and WBC. It is crucial to detect the difference between subhealthy state and healthy state, so as to prevent qualitative deterioration by taking appropriate intervention actions, such as treatment of Traditional Chinese Medicine. This study contributed to extract the potential rules and knowledge which hided in the research evidence. It also provided a new way of thinking for the assessment of healthy and subhealthy status.

**Acknowledgments**

This work was supported by Chinese “Disease” Sub-health Medicine Research and Intervention of the Eleventh Five-Year Science and Technology Support Project of China (No. 2006BAI13B01); Financial Support Case Studies of Traditional Chinese Medicine Treatment of Disease and Health

**Table 2: S-G1 and white blood cell value (after reduction) in health population (units:×10^9/L)**

<table>
<thead>
<tr>
<th>WBC value</th>
<th>6.21</th>
<th>4.80</th>
<th>6.30</th>
<th>6.03</th>
<th>5.05</th>
<th>6.09</th>
<th>6.85</th>
<th>6.37</th>
<th>5.00</th>
<th>5.47</th>
<th>6.59</th>
<th>5.89</th>
<th>6.87</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-G1</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td>90</td>
<td>92</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>100</td>
</tr>
</tbody>
</table>

WBC: White blood cell, S-G1: Health self-assessment score

**Table 3: S-G1 and White blood cell value (after reduction) in subhealth population**

<table>
<thead>
<tr>
<th>WBC value</th>
<th>8.60</th>
<th>4.50</th>
<th>4.70</th>
<th>5.00</th>
<th>7.90</th>
<th>5.96</th>
<th>7.15</th>
<th>7.00</th>
<th>6.23</th>
<th>6.23</th>
<th>6.10</th>
<th>3.80</th>
<th>6.10</th>
<th>5.97</th>
<th>5.20</th>
<th>4.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-G1</td>
<td>30</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>59</td>
<td>60</td>
<td>65</td>
<td>68</td>
<td>70</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>80</td>
<td>82</td>
<td>83</td>
</tr>
</tbody>
</table>

WBC: White blood cell, S-G1: Health self-assessment score
Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Lisitsyn YP. Pre-Disease—“Third State”, Social Hygiene Organization. Kazan NGOs, Medical Service; 1999.
7. Xu Li. Research on Statistical Modeling of Chinese Sub-Health Scale Xi’an: Fourth Military Medical University; 2008.