



OPEN ACCESS

# Construction of an evaluation index system for determining the academic impact of military medical scholars

Xuan Li,<sup>1</sup> J Y Hao<sup>2</sup>

<sup>1</sup>Military Medical Research Institute, Academy of Military Science, Chinese People's Liberation Army, Beijing, China  
<sup>2</sup>Military Science Information Research Center, Academy of Military Science, Chinese People's Liberation Army, Beijing 100142, China

## Correspondence to

J Y Hao, Military Science Information Research Center, Academy of Military Science, Chinese People's Liberation Army, Beijing 100142, China; tanhjy@163.com

Received 26 September 2017  
Revised 29 November 2017  
Accepted 30 November 2017  
Published Online First  
13 January 2018

## ABSTRACT

**Introduction** Academic papers are an essential manner for describing new ideas and consolidating existing concepts in the field of military medicine. The academic impact of military medical publications reflects the extent and depth of recognition, acceptance and utilisation of the concepts transmitted in these publications. The aim of this research was to construct an evaluation index system suitable for evaluating the academic influence of scholars in the field of military medicine.

**Methods** Using the Delphi consensus methodology, 30 experts from the field of military medicine, military medical information and library and information science were asked during three rounds of questioning to score the feasibility and importance of indicators that could be used to determine academic impact. An analytic hierarchy process method was used to calculate the relative weighting of each indicator in determining the final level of academic impact.

**Results** Eight evaluation indicators were agreed on to potentially determine academic impact. These comprised: 'Web of Science documents', 'Citation impact', 'h-index', 'Percentage of international collaborations', 'Percentage of the top 10% of the cited frequency', 'Category normalised citation impact', 'Percentage of documents cited' and 'The number of F1000 Recommended papers'.

**Conclusions** The evaluation index system determined from this study combines the advantages of both qualitative and quantitative recognised evaluation indicators, which are subsequently weighted according to their importance in the field of military medicine. It is hoped that this framework will provide a manner in the future for comparing the potential academic impact of military medical scholars.

## INTRODUCTION

Academic papers are an essential manner for describing new ideas and consolidating existing concepts in the field of military medicine. By limiting their size and by conforming to a relatively set structure, information can be transmitted more effectively. Most importantly though the use of academic papers ensures peer review and often enables the general public to understand advances that are happening in the field. The academic impact of military medical publications reflects the extent and depth of recognition, acceptance and utilisation of the concepts transmitted in these publications.<sup>1 2</sup> Appraising their academic impact may in some cases enable scholars to adjust their research direction towards

## Key messages

- ▶ In this paper, we used Delphi method and analytic hierarchy process method to construct an evaluation index system.
- ▶ The constructed comprehensive evaluation index system combines the advantages of qualitative indicators and quantitative indicators.
- ▶ The evaluation index system will be used to evaluate military medical scholars' academic impact according to their published articles.

that perceived to be more influential. It can also promote more extensive academic communication and cooperation.

Current methods of evaluating scholars' academic impact include peer review, bibliometric methods,<sup>3</sup> social network analysis<sup>4</sup> and comprehensive evaluation.<sup>5</sup> Peer review is a qualitative method for evaluating scientific research activities and related elements (such as researchers, research projects or achievements) by one or more people in a similar research field with proper competence. The bibliometric method statistically analyses certain numerical values to reveal the characteristics and value of the literature. An example of this is the InCites database, published by the Thomson Reuters Group in 2011. It combines nearly 30 years' information from Web of Science to deliver reliable metrics and indicators. Due to its inherent wealth of bibliometric indicators, standardised national benchmark data for each subject, and intuitive visualisation of the results, InCites has become a powerful research tool. The social network analysis method evaluates the cooperation of scholars and the effect on a citation network. This includes the node degree, shortest path, weight of nodes and edge data. Finally when faced with complex issues, the comprehensive evaluation method uses multiple indicators to determine the academic impact. The aim of this research was to construct an evaluation index system suitable for evaluating the academic influence of scholars in the field of military medicine.

## METHOD

### Evaluation indicators for academic papers

The following 10 evaluation indicators were selected, of which the first nine were determined from the InCites database:



To cite: Li X, Hao JY. *J R Army Med Corps* 2018;**164**:164–169.



### Web of Science documents

The total number of documents published by the scholar in this database within a certain period was used as a measure of a scholar's academic productivity; the more documents published by an author, the more likely the author is to be a scholar with a high academic impact.

### Citation impact

The value of the citation impact is obtained by dividing the total cited frequency of the published documents by the total number of documents of a scholar. Although the author's total cited frequency also reflects the extent of his or her influence, this study selects citation impact because it shows the average number of citations of a scholar, which can eliminate the influence of the total number of papers on the total cited frequency. Additionally, citation impact has stronger stability.

### H-index

The h-index<sup>6</sup> was suggested in 2005 by Jorge E Hirsch. A scholar with an index of 'h' will have published 'h' papers, each of which will have been cited in other papers at least 'h' times. Thus, the h-index reflects both the number of publications and the number of citations per publication. The h-index is considered to be a measure of a scholar's personal achievement in scholarly evaluations.

### Percentage of international collaborations

This is the percentage of the number of published papers with international coauthors out of the total number of papers published by a scholar. Some researchers have suggested that international cooperation can help enhance the scientific research impact. This indicator reflects the scholar's ability to attract international cooperation. The cooperation impact and citation impact complement each other and reflect the scholar's academic impact together.<sup>7</sup>

### Percentage of domestic collaborations

This is the percentage of the number of published papers with domestic coauthors out of the total number of papers published by a scholar.

### Percentage of the top 10% of the cited frequency

This is the ratio of the number of papers in the top 10% of the cited frequency to the total number of papers published by the scholar in a subject area, in a particular year and in a specific document type. The indicator reflects a high level of scientific research.

**Table 1** Basic information on the experts

Basic information	Number of experts (%)
<b>Years of work</b>	
Less than 10 years	2 (6.67)
10–20 years	16 (53.34)
More than 20 years	12 (40)
<b>Positional titles</b>	
Intermediate	2 (6.67)
Associate professor	12 (40)
Professor	16 (53.34)
<b>Affiliation</b>	
Military medical institutions	9 (30)
Local research institutes	9 (30)
Colleges and universities	12 (40)
<b>Professional field</b>	
Military medicine	2 (6.67)
Military medical information	7 (23.33)
Library and information	21 (70)
<b>Gender</b>	
Male	13 (43.33)
Female	17 (56.67)

### Category normalised citation impact

Each of the papers belongs to one or several disciplines and the citations may be a discrepancy between different disciplines. This indicator means the actual cited frequency divided by the expected cited frequency—which is in the same subject area—with the same type of document and published in the same year. The scholar's category normalised citation impact is the average of the category normalised citation impact of all the published papers by the scholar. This indicator can eliminate the influence of the subject area, publication year and type of literature, and so it can be used to evaluate the literature collection of different disciplines.

### Category normalised citation impact: country adjusted

The category normalised citation impact is divided by the square root of the category normalised citation impact in its country. This indicator can eliminate the influence of different political backgrounds, different languages and different scales of scientific research networks on the publication and citation of the paper due to different regions.

### Percentage of documents cited

The percentage of the number of papers being cited at least once, out of the total number of papers published by the scholar. This

**Table 2** The results of the second round of the expert questionnaire

List of indicators	Unimportant (1 point)	Generally important (3 points)	Comparatively important (5 points)	Important (7 points)	Very important (9 points)	The average score
The number of F1000 Recommended papers	0	0	5	13	10	7.36
Percentage of the top 10% of the cited frequency	0	2	4	12	10	7.14
Citation impact	0	0	5	16	7	7.14
Web of Science documents	0	1	5	16	6	6.86
H-index	0	4	6	11	7	6.5
Category normalised citation impact	1	2	10	10	5	6.14
Percentage of documents cited	2	1	12	9	4	5.79
Percentage of international collaborations	1	6	6	12	3	5.68

**Table 3** The results of the third round of the expert questionnaire

List of indicators	Unimportant (1 point)	Generally important (3 points)	Comparatively important (5 points)	Important (7 points)	Very important (9 points)	The average score
The number of F1000 Recommended papers	0	0	4	14	10	7.43
Citation impact	0	0	3	17	8	7.36
Percentage of the top 10% of the cited frequency	0	0	5	14	9	7.29
Web of Science documents	0	0	5	16	7	7.14
H-index	0	1	7	16	4	6.64
Category normalised citation impact	0	1	11	11	5	6.43
Percentage of documents cited	0	2	14	7	5	6.07
Percentage of international collaborations	1	6	13	7	1	5

indicator reveals the extent of the citation of all the documents published by the scholar.

### The number of F1000 Recommended papers

The F1000 database is a secondary literature database for evaluation and recommendation. Top scientists and outstanding researchers in biology and medicine read, evaluate and recommend the published papers. This indicator is based on the expert's qualitative evaluation. The number of F1000 Recommended papers is converted into quantitative indicators. Although this indicator is not from the InCites database, it can be retrieved through F1000 or PubMed. It combines qualitative and quantitative evaluation so that the construction of the evaluation index system will be more scientific and reliable.<sup>5</sup>

### Constructing the comprehensive evaluation index system using the Delphi method and analytic hierarchy process (AHP)

To evaluate the impact of academics in the field of military medicine, different weights should be given to indicators. In this paper, we chose 30 Chinese experts from the field of military medicine, military medical information and library and information science. To be included in this study, experts had to be researchers who are active in the field of military medicine, and they were representative of all subspecialties within the discipline (Table 1). Web page links of the questionnaire were sent to experts through email, WeChat (a popular social platform in China) or SMS so that they could complete the survey online. They were asked to score the feasibility and importance of the indicators, and feedback was provided on the results of each round of the questionnaire. After three rounds of the questionnaire survey, the relative importance of each indicator was calculated. The AHP was then used to construct the judgement matrix to calculate the weight of each indicator.

### Questionnaire design

The questionnaire comprised three parts. The first part described the purpose of the study, how to complete it and the criteria for

$$B = \begin{bmatrix} 0 & 0.07 & 0.14 & 0.29 & 0.79 & 1 & 1.36 & 2.43 \\ -0.07 & 0 & 0.07 & 0.22 & 0.72 & 0.93 & 1.29 & 2.36 \\ -0.14 & -0.77 & 0 & 0.15 & 0.65 & 0.86 & 1.22 & 2.29 \\ -0.29 & -0.22 & -0.15 & 0 & 0.5 & 0.71 & 1.07 & 2.14 \\ -0.79 & -0.72 & -0.65 & -0.5 & 0 & 0.21 & 0.57 & 1.64 \\ -1 & -0.93 & -0.86 & -0.71 & -0.21 & 0 & 0.36 & 1.43 \\ -1.36 & -1.29 & -1.22 & -1.07 & -0.57 & -0.36 & 0 & 1.07 \\ -2.43 & -2.36 & -2.29 & -2.14 & -1.64 & -1.43 & -1.07 & 0 \end{bmatrix}$$
**Figure 1** The judgement matrix B.

scoring. The second part was the question-and-answer section and listed the preliminary evaluation indicators and the connotations of each indicator. It asked the experts to evaluate and score the indicators according to feasibility and importance while using them to evaluate scholars' academic impact. Moreover, we set up open-ended questions and asked experts to make suggestions on the evaluation index system. Finally, the third part of the questionnaire asked for experts' personal information and the basis for estimate.

### First round of questioning

The first round of questioning produced a response rate of 100%.<sup>8</sup> It asked experts to evaluate the feasibility of the primary evaluation indicators. Several experts proposed deleting the indicator of 'percentage of domestic collaborations', because the indicator 'percentage of international collaboration' could better reflect the scholar's academic impact. Additionally, seven experts believed that the feasibility of the 'category normalized citation impact – country adjusted' was poor. More than 85% believed that the other eight indicators had satisfactory feasibility, and therefore the other two indicators were removed before the second round.

### Second round of questioning

The second questionnaire was sent to 30 people with a response rate of 93%. It asked experts to evaluate and score the indicators according to importance while using them to evaluate the scholars' academic impact.<sup>9</sup> For scoring, the importance of each indicator was divided into five levels: not important: 1 point; generally important: 3 points; comparatively important: 5 points; important: 7 points; and very important: 9 points (Table 2).

### Third round of questioning

Twenty-eight questionnaires were sent out in the third round with a 100% response rate.<sup>10</sup> Indicators were sorted by the average score of importance of the two rounds of survey results. The order of the indicators was basically the same in the second

$$A = \begin{bmatrix} 1.0000 & 1.0725 & 1.1503 & 1.3364 & 2.2034 & 2.7183 & 3.8962 & 11.3589 \\ 0.9324 & 1.0000 & 1.0725 & 1.2461 & 2.0544 & 2.5345 & 3.6328 & 10.5910 \\ 0.8694 & 0.9324 & 1.0000 & 1.1618 & 1.9155 & 2.3632 & 3.3872 & 9.8749 \\ 0.7483 & 0.8025 & 0.8607 & 1.0000 & 1.6487 & 2.0340 & 2.9154 & 8.4994 \\ 0.4538 & 0.4868 & 0.5220 & 0.6065 & 1.0000 & 1.2337 & 1.7683 & 5.1552 \\ 0.3679 & 0.3946 & 0.4232 & 0.4916 & 0.8106 & 1.0000 & 1.4333 & 4.1787 \\ 0.2567 & 0.2753 & 0.2952 & 0.3430 & 0.5655 & 0.6977 & 1.0000 & 2.9154 \\ 0.0880 & 0.0944 & 0.1013 & 0.1177 & 0.1940 & 0.2393 & 0.3430 & 1.0000 \end{bmatrix}$$
**Figure 2** The judgement matrix A.

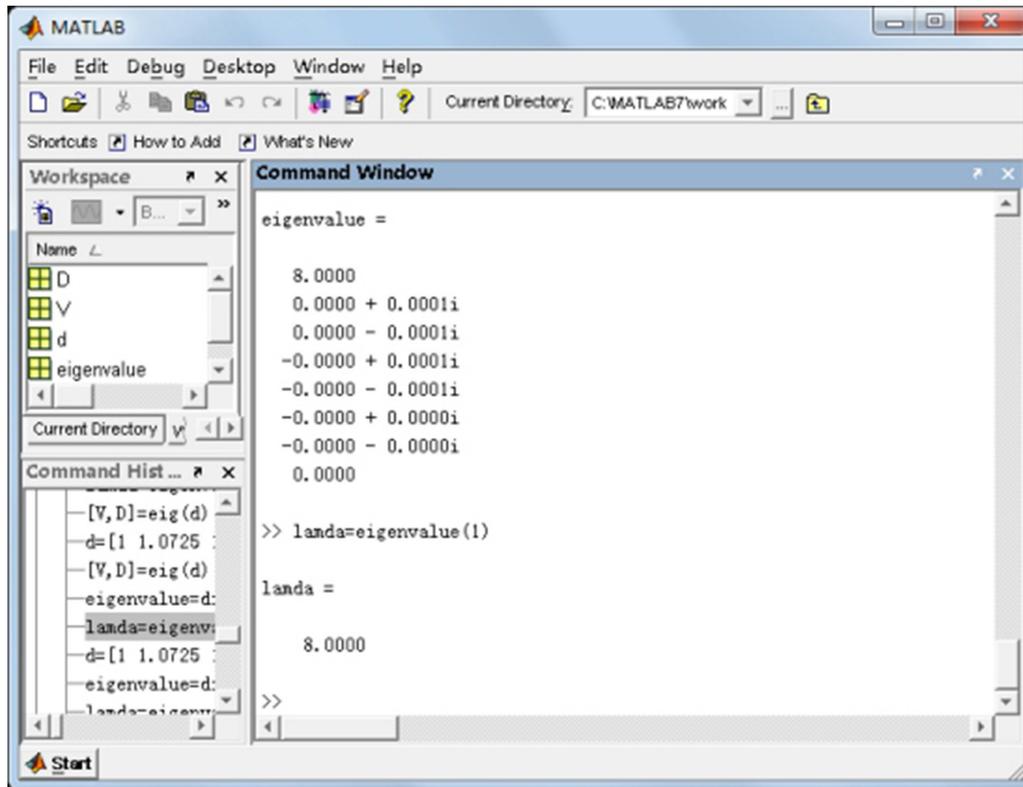


Figure 3 Calculating the eigenvector of the judgement matrix and the weight of each indicator by MatLab.

and third rounds of the expert questionnaire; almost every indicator's average score variability was less than 5% (Table 3).

### Calculating the weight of each evaluation indicator through the AHP

The AHP is a commonly used method to calculate the relative importance of each indicator. In this study, the average score of each indicator minus the average score of other indicators was used to construct the judgement matrix.<sup>11</sup>

$$B = (b_{ij})_{n \times n} \quad (b_{ij} = b_i - b_j, \quad i, j = 1, 2, \dots, n) \quad (1)$$

According to formula 1, we got the judgement matrix B (Figure 1).

Then, according to the reciprocity of the comparison matrix, the form of matrix B was converted by the exponential function  $e^x$  to construct judgement matrix A (Figure 2).

The eigenvector of judgement matrix A (Figure 3) was calculated, and then the weight of each indicator was obtained.

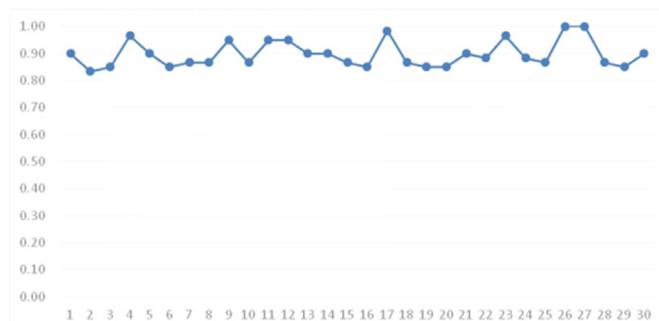


Figure 4 Experts' authority coefficient.

### Determining the authority of experts

The experts' authority has considerably impacted the reliability of questionnaire results. In this study, we considered the factors as 'the experts' own academic achievements', 'the main basis for the estimate' and 'the familiarity of evaluation',<sup>12</sup> to calculate the expert's authority coefficient (Figure 4). The lowest expert's authority coefficient was 0.83 in the case of one expert. Hence, the chosen experts had a relatively high degree of authority.

### Determining the reliability of the questionnaire

Reliability is an estimate of the consistency and stability of the results of the questionnaire. In this study, we used the Cronbach's  $\alpha$  method<sup>13</sup> to detect the credibility of the questionnaire results. The larger the Cronbach's  $\alpha$  coefficient is, the higher is the reliability of the questionnaire results. Between 0 and 1, if the  $\alpha$  coefficient is more than 0.8, this indicates that the reliability is very good; if the  $\alpha$  coefficient ranges from 0.7 to 0.8, then it has considerable reliability; if the  $\alpha$  coefficient is less than 0.7, there are indicators

Table 4 The average score and weight value of each evaluation indicator

List of indicators	Average score of importance	Weights
The number of F1000 Recommended papers	7.43	0.212
Citation impact	7.36	0.198
Percentage of the top 10% of the cited frequency	7.29	0.184
Web of Science documents	7.14	0.159
H-index	6.64	0.096
Category normalised citation impact	6.43	0.078
Percentage of documents cited	6.07	0.054
Percentage of international collaborations	5	0.019

**Table 5** The value of random index (RI) of the reciprocal matrix of 1–10 orders

Matrix order	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

in the evaluation index system that need to be modified or deleted; and if the  $\alpha$  coefficient is less than 0.35, then the credibility is low. In this study, SPSS software was used to test the reliability of the questionnaire results.

### Determining the consistency of the judgement matrix

Inconsistencies are inevitable when experts compare different indicators. Therefore, it is necessary to test the consistency of the judgement matrix. Consistency ratio (CR), proposed by Satty, is generally used as the criterion for measuring the consistency of a judgement matrix. When CR is less than 0.1, it can be considered that the judgement matrix has satisfactory consistency.<sup>14</sup>

$$CR = CI/RI \quad (2)$$

CI represents the consistency index, and RI represents random index.

$$CI = (\lambda_{\max}n)/(n-1) \quad (3)$$

$\lambda_{\max}$  means the largest eigenvalue of the judgement matrix, and  $n$  means the order of the matrix.

## RESULTS

### Indicators and weight value of each indicator

Following three rounds of questioning, eight evaluation indicators were agreed on (Table 4).

### Determining the reliability of the questionnaire

The Cronbach's  $\alpha$  coefficient of the expert questionnaire results was 0.777, indicating that the results of the study had considerable intrinsic reliability.

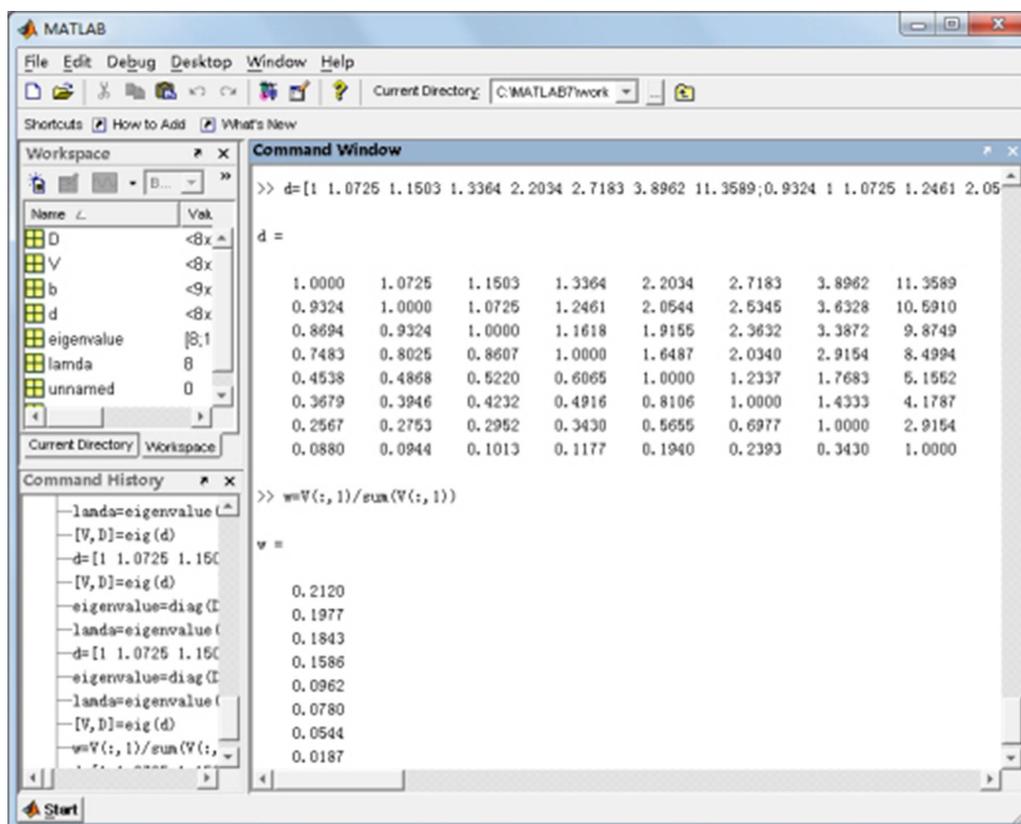
### Determining the consistency of the judgement matrix

The value of the RI of the reciprocal matrix of 1–10 orders is given in Table 5 below.

Using MatLab to calculate the maximum eigenvalue of the matrix, as shown in (Figure 5), the  $\lambda_{\max}$  was found to be 8, and the order of the matrix  $n$  was also 8.  $CI=(8-8)/(8-1)=0$ , so  $CR=0$ , indicating that the study of the structure of the judgement matrix had strong consistency.

## DISCUSSION

The aim of this research was to use a Delphi methodology and AHP, in conjunction with a qualitative evaluation and quantitative evaluation to construct an evaluation index system for evaluating the academic impact of military medical scholars. Thirty experts from the field of military medicine, military medical information and library and information science were asked three rounds of questioning to score the feasibility and importance of the indicators. This resulted in a list of evaluation indicators comprising: 'Web of Science documents', 'Citation impact', 'h-index', 'Percentage of international collaborations', 'Percentage of the top 10% of the cited frequency', 'Category normalised citation impact', 'Percentage of documents cited' and 'The number of F1000 Recommended papers'. An AHP was subsequently used to construct a judgement matrix to calculate the weight of each indicator. After analysing the authority of the experts, the questionnaire results and consistency of the matrix were found to be reliable.



**Figure 5** Calculating the maximum eigenvalue of a matrix using MatLab.

No evaluation method can ever fully describe the potential academic impact of military medical publications as there will always be a degree of individual opinion and thereby bias when interpreting scientific research. However, the evaluation index system determined from this study combines the advantages of both qualitative and quantitative recognised evaluation indicators, which are subsequently weighted according to their importance in the field of military medicine. It is hoped that this framework will provide a manner in the future for comparing the potential academic impact of military medical scholars.

**Contributors** JYH provided the overall idea and methodological guidance of the article, besides, she wrote part of the article and completed the verifying and revision of the article. XL acquired and analysed all the data of the article and wrote most of the paper.

**Funding** This work was supported by General Logistics Department of the Chinese People's Liberation Army. Grant Number (CWS14R075).

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** All data are included in the article.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

- Pan RK, Fortunato S. Author impact factor: tracking the dynamics of individual scientific impact. *Sci Rep* 2014;4:7–8.
- Gao Z, Zhang Z. Q. A review of quantitative evaluation methods for individual academic impact. 2016;39:133–8.
- Wildgaard L, Schneider JW, Larsen B. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics* 2014;101:125–58.
- Hoffmann CP, Lutz C, Meckel M. Impact factor 2.0: applying social network analysis to scientific impact assessment [J]. *Hawaii Interna Conference on System Sci* 2014;47:1576–85.
- Ma L, Liu YT, W. Y. G, et al. Bibliometrics-based evaluation index system of research outputs in clinical medicine of Chinese university and empirical research. *Chin J Med Libr Inf Sci* 2015;24:37–40.
- Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* 2005;102:16569–72.
- Didegah F, Thelwall M. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *J Informetr* 2013;7:861–73.
- Li X. The first round of the expert questionnaire. 2017 <https://sojump.com/jq/8235116.aspx> (accessed 04 May 2017).
- Li X. The second round of the expert questionnaire. 2017 <https://sojump.com/jq/8826717.aspx> (accessed 04 May 2017).
- Li X. The third round of the expert questionnaire. 2017 <https://sojump.com/jq/9253235.aspx> (accessed 04 May 2017).
- Luo Z. Q. A new method of constructing the judgment matrix in analytic hierarchy process [J]. *J Univer Electro Scie Tech China* 1999;05:557–61.
- Qi DG. *The construction of the indicator system for normally evaluating clinical teaching quality of military medical universities [D]*. Chongqing: Third Military Medical University, 2011.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- Chen JQ. *Evaluation of scientific research: methodology and application approach [D]*. Wuhan: Wuhan University, 2004.