

# Association Rule Mining to Identify Critical Demographic Variables Influencing the Degree of Burnout in A Regional Teaching Hospital

Yii-Ching Lee<sup>1</sup>, Chih-Hsuan Huang<sup>2</sup>, Yi-Chen Lin<sup>3</sup>, Hsin-Hung Wu<sup>3</sup>

<sup>1</sup> Department of Medical Quality Management, Cheng Ching General Hospital-Chung Kang Branch, Taichung City 407, Taiwan

<sup>1</sup> Department of Health Business Administration, Hung Kuang University, Taichung City 433, Taiwan

<sup>1</sup> The School of Health Policy and Management, Chung Shan Medical University, Taichung City 402, Taiwan

<sup>2</sup> School of Business Administration, Hubei University of Economics, Wuhan City 430205, China

<sup>3</sup> Department of Business Administration, National Changhua University of Education, Changhua 500, Taiwan

**Abstract** –This study uses apriori algorithm of IBM SPSS Modeler 14.1 on nine questions of emotional exhaustion dimension along with ten demographic variables from a regional teaching hospital in Taiwan in 2014 to identify critical demographic variables that influence the degree of burnout. By setting up the support of 25%, confidence of 80%, and lift of 1.5, twenty nine rules are found. To further refine the rules by their similarities, seven major combinations are summarized. The major characteristics are depicted below. Female medical staffs with college/university education who are not in charge of supervisor/manager with very often direct patient contacts feel much stressful to work with people directly and all day. That is, they have relatively higher degree of burnout. In summary, four demographic variables are found to be the major variables that influence emotional exhaustion, including gender, supervisor/manager, education, and direct patient contact.

**Keywords** – Apriori algorithm, Association rule, Emotional exhaustion, Demographic variable, Burnout.

## 1. Introduction

In recent years, the quality of care and patient safety issues in the world have gained attention

---

DOI: 10.18421/TEM63-10

<https://dx.doi.org/10.18421/TEM63-10>

**Corresponding author:** Hsin-Hung Wu,  
National Changhua University of Education, Changhua,  
Taiwan 500.

**Email:** [hhuw@cc.ncue.edu.tw](mailto:hhuw@cc.ncue.edu.tw)

 © 2017 Yii-Ching Lee et al; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

The article is published with Open Access at [www.temjournal.com](http://www.temjournal.com)

gradually, and the cooperation of a medical team is the primary key to shape patient safety culture. During the operational needs, the support from one another within the medical team members is critically important. Many researches have indicated that burnout or fatigue of medical personnel and poor health care outcomes or medical errors are positively correlated. For instance, Shanafelt et al. [1] pointed out that physicians' overwork, stress, and fatigue are contributing factors to medical errors. That is, medical errors are highly related to physicians' degree of burnout and mental quality of life. In addition to physicians, Cimiotti et al. [2] stated that nurse burnout is linked to job dissatisfaction and patient healthcare quality. In fact, an increase in a nurse' workload could result in higher infections. In order to assess the patient safety culture of a particular healthcare organization and medical staffs' degree of burnout, the Chinese version of safety attitudes questionnaire (SAQ) has been used in Taiwan annually [3].

The Chinese version of SAQ has eight dimensions, including emotional exhaustion dimension borrowed from Maslach burnout inventory-human services survey (MBI-HSS) that can be used to assess the medical staffs' degree of fatigue [3-5]. Because burnout is highly related to healthcare quality and the patient safety culture, it is critically important for the hospital management to regularly assess the medical staffs' degree of burnout. Besides, different medical staffs with different demographic variables might perceive different emotional exhaustion in terms of nine questions. It would be of interest to identify the characteristics of the medical staffs with high degree of burnout from the viewpoints of the hospital management.

Data mining, particularly the association rule, plays an essential role to knowledge discovery. The association rule, a well-recognized data mining and knowledge discovery approach, has been widely applied in medical areas [6]. However, few studies

using the association rule on the burnout have been found. Therefore, this study intends to use the association rule, i.e., apriori algorithm, to attempt to find rules and associations that exist between the demographic variables and nine questions of emotional exhaustion particularly on the high degree of burnout.

**2. Literature review**

**2.1 Emotional exhaustion**

MBI-HSS has three dimensions including emotional exhaustion, personal accomplishment, and depersonalization with the respective nine, eight, and five questions [7]. In the Chinese version of safety attitudes questionnaire, only emotional exhaustion dimension with nine questions is included to assess the medical staffs’ perceptions of burnout in healthcare organizations in Taiwan as shown in Table 1. [3-5]. Each medical staff is required to answer all of the questions in emotional exhaustion dimension based on a five-point Likert scale ranging from strongly agree to strongly disagree. Further, these nine questions are reversed questions so each respondent’s answer needs to be adjusted. For instance, if the original answer of Q2 (I feel burned out from my work.) is strongly agree, which indicates the particular medical staff has the poor perception; the adjustment is to use the numerical value of one. By the same token, the answers from the other questions are also adjusted.

*Table 1. Nine questions of Emotional Exhaustion Dimension from the Chinese Version of SAQ Survey*

No.	Description of the question
Q1	I feel like I’m at the end of my rope.
Q2	I feel burned out from my work.
Q3	I feel frustrated by my job.
Q4	I feel I’m working too hard on my job.
Q5	I feel emotionally drained from my work.
Q6	I feel used up at the end of the workday.
Q7	I feel fatigued when I get up in the morning and have to face another day on the job.
Q8	Working with people all day is really a strain for me.
Q9	Working with people directly puts too much stress on me.

**2.2 Apriori algorithm**

Data mining technique is an alternative approach to analyze social science areas such as psychology, cognitive science, and human behaviors and can be a driving force to explore new applications in the future [8]. The association rule, which is a well-recognized data mining and knowledge discovery approach, intends to discover interesting relations between variables in large databases [9]. The

advantage of association rule algorithms that attempt to find many rules and associations can exist between any of the attributes, and each of which may have a different conclusion [10].

Apriori algorithm which is one of the association rule approaches was first proposed by Agrawal and Srikant and is a standard approach that starts with a dataset containing transactions and aims to construct frequent item sets by setting up a user specified threshold, namely support, confidence, and lift [4,10,11,12]. The support for a particular association rule  $A \Rightarrow B$  is the proportion of transactions in the database containing both A and B and is formulated as below [12]:

$$\text{Support} = P(A \cap B) = \frac{\text{number of transactions containing both A and B}}{\text{total number of transactions}}$$

The confidence of the association rule  $A \Rightarrow B$  is a measure of the accuracy of the rule, which is determined by the percentage of transactions in the database containing A that also contains B and is defined below [12]:

$$\text{Confidence} = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\text{number of transactions containing both A and B}}{\text{number of transactions containing A}}$$

Lift is defined as a simple correlation that measures if A and B are independent or dependent and correlated events and is expressed by the following equation [11]:

$$\text{Lift (A,B)} = \frac{P(A \cup B)}{P(A)P(B)}$$

If a particular rule has a lift of one, it indicates that the probabilities of A and B are independent. When two events are independent, there is no rule drawn involving these two events. In contrast, if a particular rule has a lift greater than one, it indicates A and B are dependent and positively correlated [11]. Larose [12] pointed out that analysts might prefer rules that have either high support or high confidence, and usually both. A strong rule is the rule that meets or surpasses certain minimum support and confidence criteria.

Apriori algorithm that uses prior knowledge of frequent itemsets employs an iterative approach namely a level-wise search, where k-itemsets are used to explore (k+1)-itemsets [11]. The set of frequent 1-itemsets is first found by scanning the database to accumulate the count for each item and

collecting those items that satisfy the minimum support. The resulting set is denoted as  $L_1$ . Next,  $L_1$  is used to find  $L_2$  (the set of frequent 2-itemsets), which is used to find  $L_3$ , and so on, until no more frequent k-itemsets are found. In fact, the finding for each  $L_k$  requires one full scan of the database [11]. For more information about apriori algorithm, please refer to Larose [12] and Han and Kamber [11].

### 3. A case study

This study focuses on the medical staff including physicians, nurses, technicians, pharmacists, medical administrators, and respiratory therapists from a regional teaching hospital in central Taiwan. The internal data set of the Chinese version of safety attitudes questionnaire in 2014 from this case hospital is used with eight dimensions along with forty six questions, including nine questions of emotional exhaustion dimension. The demographic variables summarized in Table 2. include gender, age, supervisor/manager, respondents reporting events in the past 12 months, job position, job status, experience in organization, experience in position, education, and direct patient contact. There are 793 questions received but the number of effective questionnaire in emotional exhaustion is 680, representing an 85.8% effective response rate.

Table 2. Demographic information of this case hospital

Demographic Variable		Frequency	Percentage
Gender	1. Male	116	17.0
	2. Female	564	83.0
Age	1. Less than 20 years old	9	1.3
	2. 21-30 years old	210	30.9
	3. 31-40 years old	276	40.6
	4. 41-50 years old	155	22.8
	5. 51-60 years old	29	4.3
	6. 61 years old and above	1	0.1
Supervisor/Manager	1. Yes	44	6.5
	2. No	636	93.5
Reporting Events in the past 12 Months	1. 0 event	362	53.2
	2. 1-5 events	278	40.9
	3. 6-10 events	32	4.7
	4. 11-15 events	5	0.7
	5. 16 events and above	3	0.4
Job Position	1. Physician	49	7.2
	2. Nurse	382	56.2
	3. Technician	69	10.1
	4. Pharmacist	39	5.7
	5. Medical Administrator	126	18.5
	6. Respiratory Therapist	15	2.2
Job Status	1. Full Time	612	90.0
	2. Part Time	40	5.9
	3. Agency	8	1.2
	4. Contract	20	2.9

Experience in Organization	1. Less than 6 months	53	7.8
	2. 6 to 11 months	26	3.8
	3. 1 to 2 years	105	15.4
	4. 3 to 4 years	91	13.3
	5. 5 to 10 years	135	19.9
	6. 11 to 20 years	252	37.1
	7. 21 years and more	18	2.7
Experience in Position	1. Less than 6 months	74	10.9
	2. 6 to 11 months	29	4.3
	3. 1 to 2 years	110	16.2
	4. 3 to 4 years	99	14.6
	5. 5 to 10 years	145	21.3
	6. 11 to 20 years	215	31.6
	7. 21 years and more	8	1.2
Education	1. Junior High School and Below	1	0.2
	2. Senior High School	18	2.7
	3. College/University	608	89.4
	4. Graduate School	53	7.8
Direct Patient Contact	1. No	44	6.5
	2. Rare	81	11.9
	3. Very Often	555	81.6

Each question in emotional exhaustion dimension uses a five-point Likert scale ranging from strongly agree to strongly disagree. As mentioned early, all of the questions are reversed questions. Thus, the first step is to adjust each respondent's answer. Later, in apriori algorithm, the scale, however, should be in the binary format, i.e, true/false or zero/one. At this stage, the scales in strongly agree, agree, and neutral are transformed into value zero, while the scales in disagree and strongly disagree are transformed into value one. The purpose of the transformation is intended to identify those medical staffs with relatively higher degree of burnout. In addition, there are antecedents and consequents in the settings of apriori algorithm. In this study, antecedents consist of ten demographic variables and nine questions of emotional exhaustion dimension, whereas consequents are composed of nine questions of emotional exhaustion dimension. By performing apriori algorithm of IBM SPSS Modeler 14.1, 4,944 rules are generated. In order to simplify the rules, this study uses the following parameters: support of 25%, confidence of 80%, and lift of 1.5. The number of rules has been drastically reduced from 4,944 to 29 as shown in Table 3.

Table 3. Results generated by apriori algorithm

No.	Consequent	Antecedent	Support %	Confidence %	Lift
1	Q9=1	Q8 = 1 Gender = Female Job Status = Full Time	27.79	93.12	2.97

2	Q9=1	Q8 = 1 Gender = Female	32.06	92.66	2.09
3	Q9=1	Q8 = 1 Gender = Female Education = College/University	28.97	92.39	2.08
4	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Gender = Female Education = College/University	25.29	91.86	2.07
5	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Gender = Female	27.06	91.85	2.07
6	Q9=1	Q8 = 1 Job Status = Full Time	33.68	91.70	2.07
7	Q9=1	Q8 = 1 Education = College/University Job Status = Full Time	29.12	91.41	2.06
8	Q9=1	Q8 = 1	38.82	91.29	2.06
9	Q9=1	Q8 = 1 Gender = Female Supervisor/Manager = No Education = College/University	25.15	91.23	2.05
10	Q9=1	Q8 = 1 Gender = Female Supervisor/Manager = No	26.62	91.16	2.05
11	Q9=1	Q8 = 1 Education = College/University	34.12	90.95	2.05
12	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Job Status = Full Time	28.82	90.31	2.03
13	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Education = College/University Job Status = Full Time	25.59	90.23	2.03
14	Q9=1	Q8 = 1 Direct Patient Contact = Very Often	33.09	90.22	2.03
15	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Education = College/University	29.85	90.15	2.03
16	Q9=1	Q8 = 1 Supervisor/Manager = No	29.706	89.60	2.018

		Education = College/University			
17	Q9=1	Q8 = 1 Supervisor/Manager = No Job Status = Full Time	26.62	89.50	2.02
18	Q9=1	Q8 = 1 Supervisor/Manager = No	31.77	89.35	2.01
19	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Supervisor/Manager = No Education = College/University	25.88	88.64	2.00
20	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Supervisor/Manager = No	27.65	88.30	1.99
21	Q8=1	Q3 = 1	26.62	85.08	1.92
22	Q8=1	Q3 = 1	26.62	83.43	2.15
23	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Education = College/University	33.38	80.62	2.08
24	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Gender = Female Education = College/University	28.82	80.61	2.08
25	Q8=1	Q9 = 1 Gender = Female Education = College/University	33.24	80.53	2.07
26	Q8=1	Q9 = 1 Gender = Female	36.91	80.48	2.07
27	Q8=1	Q9 = 1 Education = College/University	38.68	80.23	2.07
28	Q8=1	Q9 = 1 Gender = Female Supervisor/Manager = No	30.29	80.10	2.06
29	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Gender = Female	31.03	80.10	2.06

The explanation of rules taking Rule 1 in Table 3. as an example is that a female with full time job answering working with people all day is really a strain is associated with working with people directly put too much stress. That is, they feel stressful to work with people directly and all day. This rule has the support of 27.794 (189/680) and confidence of

93.122 (176/189). The rule support can be calculated based upon support  $\times$  confidence, i.e.,  $27.794\% \times 93.122\% = 25.882\%$ . The lift of 2.097 shows that the antecedents and consequents are correlated positively.

Due to the similarities of rules, this study further refines the rules from 29 to 7 as shown in Table 4. For instance, a medical staff with college/university degree and very often direct patient contacts feel stressful to work with people directly and all day. Therefore, Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the medical staff is college/university educated with very often direct patient contacts.

Table 4. Refined similar rules

No.	Consequent	Antecedent	Support %	Confidence %	Lift
1	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Education = College/University	29.85	90.15	2.03
	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Education = College/University	33.38	80.62	2.08
2	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Gender = Female Education = College/University	25.29	91.86	2.07
	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Education = College/University	28.82	80.61	2.08
3	Q9=1	Q8 = 1 Gender = Female Education = College/University	28.97	92.39	2.08
	Q8=1	Q9 = 1 Gender = Female Education = College/University	33.24	80.53	2.07
4	Q9=1	Q8 = 1 Gender = Female	32.06	92.661	2.09
	Q8=1	Q9 = 1 Gender = Female	36.912	80.478	2.07
5	Q9=1	Q8 = 1 Education = College/University	34.118	90.948	2.05
	Q8=1	Q9 = 1 Education = College/University	38.676	80.228	2.07

6	Q9=1	Q8 = 1 Gender = Female Supervisor/Manager = No	26.618	91.16	2.05
	Q8=1	Q9 = 1 Gender = Female Supervisor/Manager = No	30.294	80.097	2.06
7	Q9=1	Q8 = 1 Direct Patient Contact = Very Often Gender = Female	27.06	91.85	2.07
	Q8=1	Q9 = 1 Direct Patient Contact = Very Often Gender = Female	31.03	80.10	2.06

Rule 2 shows that a female medical staff with college/university education and very often direct patient contact answering strongly disagree or disagree on Q8 (adjusted reversed question) would answer strongly disagree or disagree on Q9 (adjusted reversed question). Also, a female medical staff with college/university education and very often direct patient contact answering strongly disagree or disagree on Q9 (adjusted reversed question) would answer strongly disagree or disagree on Q8 (adjusted reversed question). Therefore, Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the female medical staff is college/university educated with very often direct patient contact.

From Rule 3, we can summarize that Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the female medical staff is college/university educated. From Rule 4, the conclusion is that Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the medical staff is female. From Rule 5, Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the medical staff is college/university educated. From Rule 6, Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the female medical staff is not in charge of supervisor/manager. Finally, from Rule 7, Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated when the female medical staff has very often direct patient contact.

#### 4. Conclusions

By applying apriori algorithm on nine questions of emotional exhaustion dimension of the Chinese version of safety attitudes questionnaire from a regional teaching hospital in Taiwan in 2014, twenty nine rules are found when the support, confidence,

and lift are set to 25%, 80%, and 1.5, respectively. In addition, by further summarizing the similarities, seven combinations are found. The major characteristics are that Q8 (adjusted reversed question) and Q9 (adjusted reversed question) are associated with each other when the female medical staff with college/university education and very often direct patient contact is not in charge of supervisor/manager. That is, disagree or strongly disagree in working with people all day is not really a strain (Q8 adjusted reversed question) is associated with disagree or strongly disagree in working with people directly does not put too much stress. Therefore, the female medical staff who is not in charge of supervisor/manager with college/university education and very often direct patient contact have relatively higher degree of burnout. In summary, four demographic variables including gender, supervisor/manager, education, and direct patient contact are the major variables to influence the degree of emotional exhaustion. From the practical viewpoints, the hospital management needs to pay much attention to the medical staff with these demographic characteristics and then try to take actions to reduce the degree of burnout in order to increase the patient safety in this regional teaching hospital.

#### **Institutional Review Board Approval**

The clinical trial approval certificate (ethic statement) was approved by Cheng Ching General Hospital in Taichung City, Taiwan with protocol number of HP150029.

#### **Acknowledgements**

*This study was partially supported by Ministry of Science and Technology in Taiwan with the grant number of MOST 105-2221-E-018-015.*

#### **References**

- [1]. Shanafelt, T.D., Balch, C.M., Bechamps, G., Russell, T., Dyrbye, L., Satele, D., Collicott, P., Novotny, P.J., Sloan, J., & Freischlag, J. (2010). Burnout and medical errors among American surgeons. *Annals of Surgery*, 251(6), 995-1000.
- [2]. Cimiotti, J.P., Aiken, L.H., Sloane, D.M., & Wu, E.S. (2012). Nurse staffing, burnout, and health care-associated infection. *American Journal of Infection Control*, 40, 486-490.
- [3]. Lee, Y.C., Shieh, J.I, Huang, C.H., Wang, C.Y., & Wu, H.H. (2017). Analyzing patient safety culture from viewpoints of physicians and nurses - A case of a regional teaching hospital in Taiwan. *Journal for Healthcare Quality*, doi: 10.1097/JHQ.000000000000065.
- [4]. Chi, C.Y., Wu, H.H., Huang, C.H., & Lee, Y.C. (2017). A linear regression approach to identify critical demographic variables affecting the patient safety culture from viewpoints of physicians and nurses. *Hospital Practices and Research*, 2(2), 47-53.
- [5]. Cheng, C.Y., Lee, Y.C., Huang, C.H., & Wu, H.H. (2017). Identifying critical dimensions of the patient safety culture by linear regression approach. *Proceedings of 2nd International Multidisciplinary Research Conference (IMRC 2017)*, 27-36.
- [6]. Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 40, 1086-1093.
- [7]. Lee, H. F., Chien, T. W., & Yen, M. (2013). Examining factor structure of Maslach burnout inventory among nurses in Taiwan. *Journal of Nursing Management*, 21, 648-656.
- [8]. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, 39, 11303-11311.
- [9]. Kamsu-Foguem, B., Rigal, F., & Mauget, F. (2013). Mining association rules for the quality improvement of the production process. *Expert Systems with Applications*, 40, 1034-1045.
- [10]. IBM Corporation (2012), *IBM SPSS modeler 15 modeling nodes*. IBM Corporation.
- [11]. Han, J. & Kamber, M. (2006). *Data mining: Concepts and techniques*, 2nd ed. New York: Morgan Kaufmann Publishers.
- [12]. Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data mining*. Hoboken, New Jersey: John Wiley and Sons.