Health Science Monitor
2022; 1(2): 116-124
Published online (http://hsm.umsu.ac.ir)

Original Article

Application of artificial neural network model in studying the mechanism of disease relapse event in patients with tuberculosis

Mehdi Kazempour Dizaji[1,2*], Afshin Moniri, Rahim Roozbahani[3], Mohammad Varahram[1], Payam Tabarsi[3], Ali Zare[2], Parvaneh Baghaei Shiva[3], Atefe Abedini[4], Majid Marjani[3], Mohammadreza Madani[4], Arda Kiani[5], Mohammad Ali Emamhadi[6], Niloufar Alizedeh Kolahdozi[2]

[1] *Mycobacteriology Research Center (MRC), National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2] *Department of Biostatistics, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[3] *Clinical Tuberculosis and Epidemiology Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[4] *Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[5] *Tracheal Diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[6] *Department Forensic Medicine, School of Medicine Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[7] *Virology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

***Corresponding author:** Mehdi Kazempour Dizaji, **Address**: Mycobacteriology Research Center (MRC) Shahid Beheshti University of Medical Sciences, Tehran, Iran, **Email**: mkazempour@sbmu.ac.ir, **Tel**: +982127122080

Abstract

***Background & Aims*:** Today, due to progressing technology and improving the standard of living of humans, the study of diseases has become more complex. This complexity has led to using new methods, such as the model of artificial neural networks (ANNs), to study many chronic diseases, especially tuberculosis (TB). The present study aimed to investigate the mechanism of disease relapse events by applying a multilayer perceptron artificial neural network (MLP-ANN) model among TB patients.

***Materials & Methods*:** This retrospective cohort study examined information of 4,564 TB patients treated in Masih Daneshvari Hospital, Tehran, Iran, from 2005 to 2015. TB disease relapse was considered as a study event, and the relapse mechanism was investigated using an MLP-ANN model consisting of three layers.

***Results*:** Based on an MLP-ANN model comprising three layers, the power to accurately predict disease relapse in TB patients was 96%. Also, variables of family size, adverse effects of exposure to cigarette smoke, patient age, and education as very effective factors, and marital status, history of drug use, imprisonment, pulmonary TB, diabetes, and AIDS as effective factors were identified in predicting the mechanism of TB disease relapse.

***Conclusion*:** Using an ANN model in the study of TB relapse due to its flexibility and high predictive accuracy can clarify any ambiguous aspects of this disease.

***Keywords*:** Artificial neural networks, Perceptron, Relapse, Tuberculosis

Received 20 August 2022; accepted for publication 19 October 2022

Introduction

Tuberculosis (TB) is one of the oldest human diseases considered a cause of mortality and disability in humans for many years. Today, despite the discovery of the disease agent, vaccines, and effective drugs to treat TB, this disease is still a major health problem in the world. Even with the proper treatment techniques and promising reports of TB treatment, the problem of TB relapse remains a challenge to TB control. Disease relapse is the reactivation of the disease in a person who is considered to have improved and strongly depends on the quality of treatment. Disease relapse is very rare in patients who have taken medication regularly and completed the course of treatment. Therefore, patients do not need follow-up after stopping treatment. Disease relapse often occurs in the first year after treatment and will be accompanied by clinical symptoms.

Overall, TB patients with disease relapse can be classified into three categories: patients who have been treated and improved in the past but have presented with positive sputum (relapse), those who have positive sputum after five months of TB medication (treatment failure), and patients who have discontinued their medication for two months or more and return with positive sputum (absence of treatment). TB relapse are mostly caused by incomplete treatment, misdiagnosis, voluntary discharge, nosocomial infection, temporary discharge, and complications of the disease. However, in addition to these reasons, there are many individuals and also demographic and clinical characteristics of patients that can be effective in the disease relapse and the failure of treatment strategies. Despite extensive studies on TB and its relapse, these factors have received little attention. Generally, the success of treatment and reduction in the incidence of disease relapse require detailed studies on the relapse mechanism of this disease and also consideration of various individual and demographic data and clinical factors affecting the disease relapse.

There are many methods to study the incidence of disease relapse in TB patients, but due to their limitations and low flexibility, the possibility of studying the exact mechanism of the disease relapse deprives based on these methods. Today, owing to the development of treatment methods for TB and improving patients' quality of life, the use of new techniques that can accurately estimate the status of this disease has been considered. One of the methods designed to study the mechanism of disease-related events such as disease relapse in TB patients is the artificial neural network (ANN) model, the use of which has increased in recent decades. Currently, the ANN models are widely used in complex processes and various fields of medical sciences, such as diagnosis and treatment, because of their high predictive accuracy and flexibility. The term ANN refers to a family of models inspired by studies of the human brain. It is like a processor that has a natural desire to store experimental knowledge and makes it usable, and its function is very similar to the human brain. Perceptron, the smallest processing unit in an ANN, is a mathematical model of a biological neuron in different layers of the ANN, such as the input, hidden, and output layers, which are connected by synaptic weights. One of the most widely used ANNs is the multilayer perceptron (MLP-ANNs). The most common type of the MLP-ANN model is the three-layer perceptron (Figure 1)

ANNs examine and evaluate information in two phases: training and testing. In the network training phase, receiving a random sample of data, usually, more than 50% of the information, learns and evaluates the relationship between the data and the studied variables. After completing the training phase, the network performance and predictive power will be tested with the remaining observations. In the model of ANNs, there is no limit to the type of data studied; therefore, these models can present replies logically and plausible with missing data, data with a heavy censoring rate, and data with bias.

Due to the high flexibility and accuracy of prediction, the ANNs models can be one of the best tools to investigate health data. Therefore, using an ANN model to study the mechanism of TB relapse can clarify many obscure aspects of this disease. Consequently, this study was designed to investigate the relapse mechanism of TB disease by applying a three-layer MLP-ANN model in TB patients.

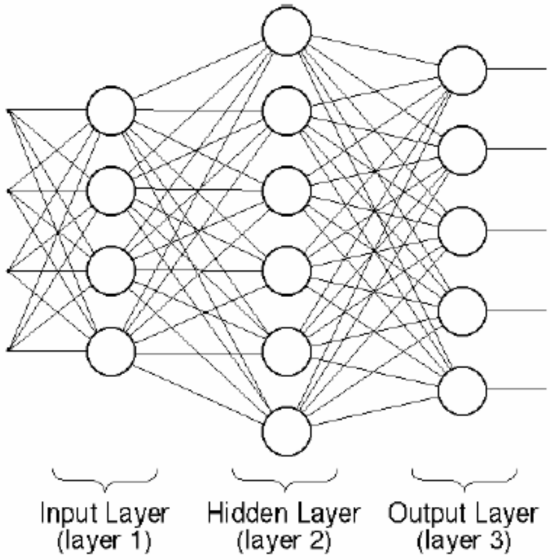

Fig. 1. Schematic of a three-layer MLP-ANN model.

Materials & Methods

A retrospective cohort study was performed at Dr. Masih Daneshvari Hospital, Tehran, Iran, from 2005 to 2015. The data of this study were obtained from the medical records of 4,564 TB patients who referred to the aforesaid hospital and treated by the National Institute of TB and lung disease research of this hospital. The time to occurrence of disease relapse was determined after treatment and recovery of these patients, and disease relapse (based on sputum culture) was considered the study event. To follow the patient's condition until the disease relapse, information was collected through their clinical records, and the last follow-up for patients who were discharged without a history of relapse was carried out through contact with them or their relatives. Patients whose status was without disease relapse until the end of the study, and after that, no information was available about them. Also, patients whose information was complete until the end of the survey and indicated no disease relapse were also considered censored.

In the present study, the mechanism of TB disease relapse, the effect of individual, demographic and clinical variables such as gender, patient's age, marital status, education level, residency area, nationality, family size, medicines adverse effects, smoking, exposure to secondhand smoke (passives smoker), history of drug use, contact with a TB patient, imprisoned, pulmonary TB, extra-pulmonary TB, diabetic mellitus, HIV positive, and comorbidities were examined on the time to relapse event and disease relapse status.

A three-layer perceptron ANN was applied to study the mechanism of disease relapse in the patients studied. In this model, the first layer was considered for individual, demographic and clinical variables of patients, the second layer as a hidden layer, and the third layer for the time to relapse event and disease relapse status. According to the studied variables, 39 neurons were considered for the input layer and three for the output layer. Also, according to the best structure for the ANN, 21 neurons were determined as the appropriate number of hidden or middle-layer

neurons. In this model, the effect of each factor on predicting the disease relapse was evaluated by the importance index and normalized importance. All analyses were performed using SPSS 19 software.

Results

In this study, the information of 4,564 TB patients was examined based on the three-layer perceptron ANN model. For 166 (3.64%) patients, the disease relapse occurred as the desired event after treatment, and 4,398 (96.36%) patients were considered censored. To determine the structure and number of neurons in each model layer, the patients' individual, demographic, and clinical data as input layer variables and time to relapse event and disease relapse status as output layer variables of the ANN model were studied (Table 1).

Table 1. Variables of TB patients and number of neurons for input and output layers in perceptron ANN model

Layers	Variables	Category	Number of neurons
Output	Time to relapse event	-	1
Output	Disease relapse status	Relapse Censored	2
Input	Gender	Female Male	2
Input	Age	-	1
Input	Marital Status	Single Married Widow Divorced	4
Input	Education	Illiterate Primary Secondary High School Higher education	5
Input	Residency area	Rural Urban	2
Input	Nationality	Iranian Non-Iranian	2
Input	Family size	-	1
Input	Adverse effect	No Yes	2
Input	Smoker	No Yes	2
Input	Passives smoker	No Yes	2
Input	Drug user	No Yes	2
Input	TB contact	No Yes	2
Input	Imprisoned	No Yes	2
Input	Pulmonary TB	No Yes	2
Input	Extra-pulmonary	No Yes	2
Input	Diabetic mellitus	No Yes	2
Input	HIV positive	No Yes	2
Input	Comorbidities	No Yes	2

To study the disease relapse event based on the three-layer perceptron ANN model, 70% of patients were randomly placed in the training and 30% in the testing phases. To determine the best model structure and specify the number of neurons suitable for the hidden layer, a set of three-layer perceptron ANNs with

different numbers of neurons was examined for the hidden layer based on the training phase. These models were compared using indices such as the sum of square error, average overall relative error, and area below the ROC curve. A three-layer perceptron ANN model with 39 neurons for the input layer, 21 neurons for the hidden layer, and three neurons for the output layer with the lowest total sum of square error, average overall relative error, and also the highest area under the ROC curve was selected as the best structure among other ANN models (Table 2).

Table 2. The best structure for the perceptron ANN model based on varying number of neurons in the hidden layer

ANNs structure (Input/hidden/output)	Sum of squares error	ROC curve area	Average overall relative error
39/6/3	124.88	0.60	0.80
39/7/3	125.72	0.57	0.81
39/8/3	125.32	0.56	0.80
39/9/3	125.22	0.53	0.78
39/10/3	125.29	0.58	0.82
39/11/3	124.70	0.60	0.79
39/12/3	124.02	0.61	0.78
39/13/3	125.32	0.59	0.80
39/14/3	124.46	0.59	0.79
39/15/3	125.33	0.59	0.80
39/16/3	124.19	0.61	0.78
39/17/3	125.27	0.58	0.80
39/18/3	125.72	0.55	0.81
39/19/3	125.52	0.55	0.82
39/20/3	125.28	0.58	0.80
39/21/3*	123.52	0.61	0.76
39/22/3	125.31	0.61	0.82
39/23/3	125.52	0.48	0.79
39/24/3	127.52	0.46	0.80
39/25/3	125.82	0.60	0.82

Results are based on the training phase.

Based on the results of this model, the correct predictive power of disease relapse events in TB patients for both training and testing phases were 96.3% and 96.5%, respectively. This model has only 3.7% and 3.5% errors in predicting the disease relapse event in TB patients, respectively. In other words, this model could predict the correct mechanism of disease relapse events for approximately 96 out of every 100 patients with TB. To evaluate the importance of individual, demographic, and clinical factors in predicting the disease relapse event in TB patients, a scale was used based on the normalized importance index. This study recognized variables with normalized importance index above 60% as very effective factors, variables with normalized importance index between 20 to 60% as effective factors, and variables with normalized importance less than 20% as ineffective factors.

In the present study, the results of the three-layer perceptron ANN model for selecting important variables in predicting the mechanism of disease relapse events showed that the variables of family size, adverse effects, exposure to cigarette smoke, patient age, and education level had the greatest effect and were considered as one of the most effective factors on the disease relapse event. Also, the variables of marital status, history of drug use, imprisoned, having pulmonary TB, diabetes, and AIDS were among the effective factors, while variables of patient gender, patient nationality, place of residence, smoker, history of contact with a TB person, extra-pulmonary TB and comorbidities were among the least effective factors in

predicting the disease relapse event in TB patients (Table 3).

Table 3. Importance of the variables on the mechanism of disease relapse event based on perceptron ANN model

Variables	Importance	Normalized importance (%)
Gender	0.030	19.0
Age	0.098	62.2
Marital status	0.073	46.2
Education	0.100	63.4
Residency area	0.024	15.2
Nationality	0.031	19.9
Family size	0.158	100
Adverse effect	0.109	68.9
Smoker	0.016	10.2
Passives smoker	0.102	64.8
Drug user	0.040	25.5
TB contact	0.022	13.8
Imprisoned	0.034	21.2
Pulmonary TB	0.045	28.6
Extra-PulmonaryTB	0.025	15.9
Diabetic mellitus	0.032	20.6
HIV positive	0.034	21.2
Co morbidities	0.026	16.2

Also, to better understand the importance of the factors affecting the prediction of disease relapse events in TB patients, information on the normalized importance index and importance index graphically (from the most to the least impact) has been marked (Figure 2).

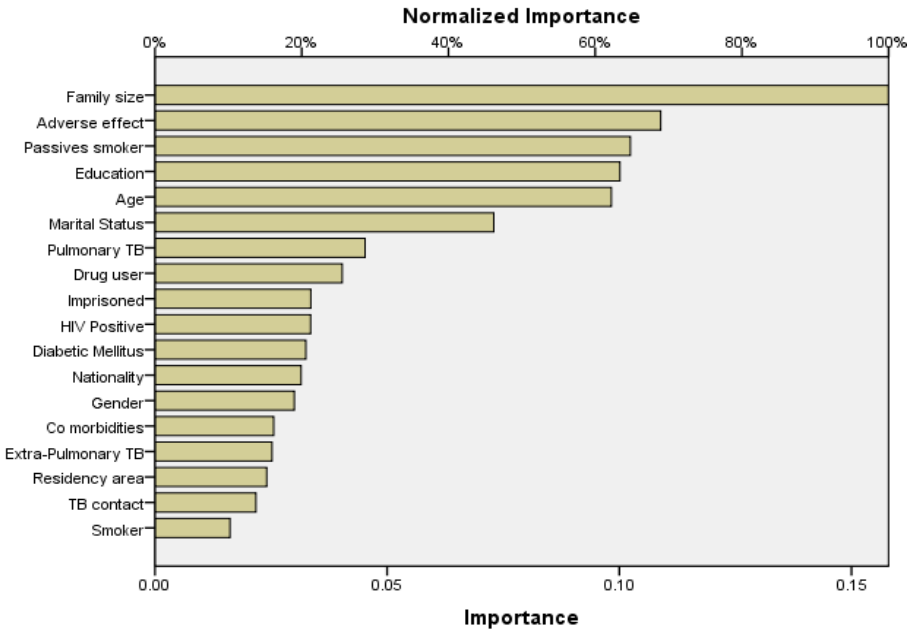

Fig. 2. The order of importance of the variables on the mechanism of disease relapse event in perceptron ANNs model based on normalized importance and importance index.

Discussion

Advances in science and improvement of living standards have affected the treatment and control of various diseases. In general, not only lifestyle and health-related factors have changed but also the study of disease processes has become more complex. In such situations, conventional methods cannot provide accurate disease status to researchers due to their patient data and information limitations. This limitation has led to applying new methods more than conventional statistical methods in studying many chronic diseases, such as TB. As a result of recent advances in medical science and the implementation of appropriate treatment strategies, new methods for studying TB are further considered. Despite appropriate treatment regimens for TB, the problem of relapse, control, and therapy of this disease is globally still a threat. Due to the occurrence of death events and the low incidence of TB, the disease relapse event has received less attention. Indeed, most studies on TB have focused on the diagnosis and treatment, and the factors influencing the mechanism of disease relapse events have not been recognized.

Some studies have been conducted to evaluate the factors affecting the survival of TB patients during treatment. Most of these studies have been designed to investigate the factors affecting events such as death and to estimate the effectiveness of treatment methods and improve the quality of life of TB patients. The lack of studies on the incidence of relapse in TB patients has made disease relapse a health problem in controlling this disease. In the pre-chemotherapy period, the risk of relapse was high among TB patients, but after chemotherapy, this figure decreased. In Iran, the TB relapse rate is estimated at about 7% over ten years. The disease relapse strongly depends on the quality of treatment strategies that clinical specialists consider for the patient. Therefore, studying the mechanism of disease relapse in TB patients can evaluate the success of treatment regimens. However, this requires a careful study of the mechanism of disease relapse in TB patients.

So far, many studies on various diseases have been carried out using new methods such as ANN models, but in the case of TB and especially on the mechanism of TB relapse, this study is one of the few studies that have been conducted. Assessing the mechanism of disease relapse using conventional statistical methods will not be accurate due to patient information and data limitations. This restrictions can lead to incorrect decisions about the quality of treatment and improper prediction of the disease relapse. Therefore, it is necessary to have a method relative to the data structure with high accuracy and flexibility in predicting. Accordingly, this study was performed to design a modern method for analyzing TB patients' data to achieve the mechanism of disease relapse. To reach this goal, a three-layer perceptron ANN model was applied with 39 neurons in the input layer, 21 in the latent layer, and three in the output layer.

Based on the results of this model, the accuracy of predicting the disease relapse mechanism in TB patients was estimated to be 96%, which is a relatively high prediction accuracy. To better understand the predictive power of this model, it should be noted that this model can accurately predict the disease relapse in 96 out of every 100 TB patients. Various studies have also shown that neural networks have higher predictive power than conventional statistical methods. In this study, in addition to determining the effect of different variables on the disease relapse mechanism, the order of their importance was also assessed graphically based on the indicators of importance and normalized importance. One of the advantages of using ANNs is to present the order of importance of affective factors in the prediction of the disease relapse mechanism in TB patients. This issue is especially important in health policy and prevention at the community level.

In this study, the variables of family size, adverse effects, exposure to cigarette smoke, patient age, and level of education with a normalized importance index above 60% showed the greatest effect in predicting the disease relapse mechanism in TB patients. Variables of marital status, history of drug use, imprisoned, pulmonary TB, diabetes, and AIDS with a normalized

importance index of between 20% to 60% were introduced as effective factors in predicting the disease relapse in TB patients. As mentioned earlier, studies on pulmonary disease and TB and also identification of factors affecting the disease relapse have paid less attention to patients' individual, as well as demographic and clinical characteristics. According to the results of this study, based on the characteristics of patients and even before starting treatment, it can be predicted which group of patients is more likely to have disease relapse. For instance, according to the results of the present study, having a crowded family, adverse drug effects, exposure to secondhand smoke, the patient's age, and low education significantly increase the likelihood of TB relapse. As a result, there should be more sensitivity in the choice of treatment strategies for these patients than for other patients.

Conclusion

The model of ANNs has high predictive power and great flexibility due to not being sensitive to patient data structure and information. Therefore, using ANN models to study chronic diseases such as TB can clarify many aspects of the disease.

Acknowledgments

Thanks to the staff of Masih Daneshvari Hospital who helped us in this research

Conflict of interest

The authors have no conflict of interest in this study.

Funding/support

The Shahid Beheshti University of Medical Sciences financially supported this study.

Data availability

The raw data supporting the conclusions of this article are available from the authors upon reasonable request.

References

1. Organization, WH, Global tuberculosis report. 2015. p. 204.
2. Diacon AH, Pym A, Grobusch MP, de los Rios JM, Gotuzzo E, Vasilyeva I, Leimane V, Andries K, Bakare N, De Marez T, Haxaire-Theeuwes M. Multidrug-resistant tuberculosis and culture conversion with bedaquiline. New England Journal of Medicine. 2014 Aug 21;371(8):723-32.
3. Quy HT, Lan NT, Borgdorff MW, Grosset J, Linh PD, Tung LB, van Soolingen D, Raviglione M, Co NV, Broekmans J. Drug resistance among failure and relapse cases of tuberculosis: is the standard re-treatment regimen adequate?. The International Journal of Tuberculosis and Lung Disease. 2003 Jul 1;7(7):631-6.
4. Sevim T, Atac G, Güngör G, Törün T, Aksoy E, Gemci I, Tahaoglu K. Treatment outcome of relapse and defaulter pulmonary tuberculosis patients. The International Journal of Tuberculosis and Lung Disease. 2002 Apr 1;6(4):320-5.
5. Farley JE, Ram M, Pan W, Waldman S, Cassell GH, Chaisson RE, Weyer K, Lancaster J, Van der Walt M. Outcomes of multi-drug resistant tuberculosis (MDR-TB) among a cohort of South African patients with high HIV prevalence. PloS one. 2011 Jul 22;6(7):e20436.
6. Pietersen E, Ignatius E, Streicher EM, Mastrapa B, Padanilam X, Pooran A, Badri M, Lesosky M, van Helden P, Sirgel FA, Warren R. Long-term outcomes of patients with extensively drug-resistant tuberculosis in South Africa: a cohort study. The Lancet. 2014 Apr 5;383(9924):1230-9.
7. Salaniponi FM, Nyirenda TE, Kemp JR, Squire SB, Godfrey-Faussett P, Harries AD. Characteristics, management and outcome of patients with recurrent tuberculosis under routine programme conditions in Malawi. The international journal of tuberculosis and lung disease. 2003 Oct 1;7(10):948-52.
8. Organization, WH, *Global tuberculosis control: WHO report 2010*. 2010: World Health Organization.
9. Organization, W.H., *Anti-tuberculosis drug resistance in the world: third global report.* 2004.
10. Akl A, Ismail AM, Ghoneim M. Prediction of graft survival of living-donor kidney transplantation:

nomograms or artificial neural networks?. Transplantation. 2008 Nov 27;86(10):1401-6.

11. Chi CL, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. InAMIA annual symposium proceedings 2007 (Vol. 2007, p. 130). American Medical Informatics Association.
12. Yegnanarayana B. Artificial neural networks. 2009: PHI Learning Pvt.
13. Mirsaeidi SM, Tabarsi P, Khoshnood K, Pooramiri MV, Rowhani-Rahbar A, Mansoori SD, Masjedi H, Zahirifard S, Mohammadi F, Farnia P, Masjedi MR. Treatment of multiple drug-resistant tuberculosis (MDR-TB) in Iran. International journal of infectious diseases. 2005 Nov 1;9(6):317-22.
14. Wright A, Zignol M. Anti-tuberculosis drug resistance in the world: fourth global report: the world health organization/international union against tuberculosis and lung disease (who/union) global project on anti-tuberculosis drug resistance surveillance, 2002-2007. World Health Organization; 2008.
15. Mirsaeidi MS, Tabarsi P, Farnia P, Ebrahimi G, Morris MW, Masjedi MR, Velayati AA, Mansouri D. Trends of drug resistant Mycobacterium tuberculosis in a tertiary tuberculosis center in Iran. Saudi medical journal. 2007 Apr 1;28(4):544.
16. Mansoori, S., et al., *Comparative study of initial and acquired drug resistance in pulmonary tuberculosis.* Revue internationale des services de santé des forces armées, 2003. 76(1): p. 45-49.
17. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2001 Apr 15;91(S8):1636-42.
18. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. Molecular cancer. 2005 Dec;4(1):1-2.