一般口演

一般口演7
医療データ分析３（レセプトデータ・治験）

2018年11月23日(金) 14:20 〜 15:50 H会場 (福岡サンパレスHパレスホール)

[2-H-2-2] 肺炎入院患者の重症度をレセプトデータから事後的に生成できるか？

○小原 仁[1], 水谷 駿介[2], 小森園 康二[3] （1.久留米大学データセンター, 2.南風病院医事課, 3.南風病院肝臓内科）

【目的】各医療機関において日常的に生成・蓄積されているレセプトデータをもとに、肺炎入院患者の重症度を事後的に判別する判別モデルの開発と開発したモデルの判別能を評価する。【方法】肺炎入院患者の重症度は、調査施設で記録された A-DROPによる肺炎重症度情報を用いた。解析用のデータセットは、重症度を付与された肺炎患者のレセプトデータをもとに、入院時に提供されたすべての診療行為の実績情報を有するデータセットを生成した。生成したデータセットは統計モデル構築用の訓練用データセットと構築したモデルの評価に用いる評価用データセットに二分割を行った。統計モデルの構築と評価は以下の手順とした。（1）診療行為実績の該当を示す各変数の単変量解析を行い説明変数候補のスクリーニングを行った。（2）スクリーニングされた変数からグラフィカルモデリング法を用いて、変数間の関連が弱い組み合わせとなる変数を選定した。（3）４つの重症度区分をもつ肺炎重症度を順序ロジスティック回帰モデルの目的変数とし、統計的有意水準0.2未満を基準とするステップワイズ法をもとに説明変数の変数選択を行った。（4）構築したモデルの判別能は、事前に付与された A-DROPの重症度と評価用データに構築したモデルを適用させて判別された重症度の一致割合（95%信頼区間）をもとに評価した。【結果】レセプトデータから8つの説明変数を有する判別モデルが構築された。説明変数は年齢、性別のほか、酸素吸入やフローボリュームなどの診療行為実績を示す変数が選択された。未知となる評価用データに構築したモデルを適用させた判別能は、 A-DROPによる肺炎重症度との一致割合0.779（95%信頼区間:0.713-0.836）であった。【結論】保険診療を行う医療施設に蓄積されているレセプトデータをもとに、肺炎入院患者の重症度を事後的に生成することは可能である。

Predictive capability of the receipt data for pneumonia inpatients severity

Hitoshi Obara[*1], Syunsuke Mizutani[*2], Yasuji Komorizono[*2]
*1 Kurume University Data Center, *2 Nanpuh Hospital

[Background] Severity of disease should be adjusted for when the average length of hospital stay, medical expenses and others are compared between hospitals. In this study, a discriminant function that discriminates mild and sever pneumonia patients is developed based on health insurance claim data that are available from any hospitals in Japan. The function may be used for adjusting for the severity of pneumonia patients when comparison of medical care services is conducted between hospitals in Japan. [Method] The severity data of pneumonia patients that are available from Nanpuh Hospital in Kagoshima and also the health insurance claim data from the same hospital are employed to establish the discriminant function. After dividing the study subjects into two groups that are called the training dataset and test dataset, respectively, the discriminant function is constructed based on the training dataset in the following way: (1) items in health insurance claim data are screened by the simple regression, (2) the screened items are classified into several groups by graphical modeling technique so that items in the same group have week correlations, (3) applying variable selection technique a tentative best model is constructed in each group, (4) Akaike's Information Criterion(AIC) of the tentative best model is computed and the model with the smallest AIC throughout groups is selected as the best model. (5) the best cut-off point that has the largest sum of sensitivity and specificity rates is selected, and finally the behavior of the established discriminant function is evaluated by using the test dataset. [Results and Discussions] The sensitivity and specificity of the established discriminant function are shown to be 0.844 and 0.743. Many hospitals in Japan have no severity information of inpatient, but any hospitals have the health insurance claim data. The function is developed based on the health insurance claim data and may be applied to adjusting for severities of patients when conducting comparison of medical care services between medical institutions in Japan.

Keywords: discriminant function, Receipt data, pneumonia, severity, comparative study

---

1. Introduction

According to medical statistics by the Ministry of Health, Labor and Welfare, Japanese Government, the number of inpatients contracted with pneumonia is the largest among many diseases.[1)] It is expected to further increase in future with the progress of aged society. [2)] Because of the limited medical resources, it is important to develop effective medical care service to patients. Comparative studies among hospitals would be useful for this purpose. To conduct such studies it would be essential to adjust for the severity of inpatients at the beginning of hospitalization in addition to type and size of hospitals for comparability between hospitals. Although the data on the severities of patients at the first examination are given by hospitals covered by DPC (Diagnosis Procedure Combination) in Japan, they are not given in many hospitals that are not covered by DPC. [3)-4)]

The purpose of this paper is to develop a formula to discriminate between severe cases and mild cases at the first examination of pneumonia patients based on the data of health insurance claims, which are called the Receipt data and available from all hospitals in Japan. The formula may be used for adjusting for the severities of disease when comparison of medical care services is undertaken between hospitals in Japan

2. Materials and Method

2.1. Source of data

Receipt data from Nanpuh hospital in Kagoshima are used in this study. The Nanpuh hospital has 338 beds and is one of the DPC hospitals where severity data of each pneumonia patient is available.

2.2. Study subjects

The study subjects are 590 pneumonia patients who are older than or equal to 15 years old and are completed admission and discharge between April 1, 2012 and March 31, 2014 in Nanpuh hospital. Pneumonia patients are defined according to the definition of DPC, namely, pneumonia, acute bronchitis, acute bronchiolitis, influenza, viral pneumonia. Patients who were hospitalized for treatment other than pneumonia and who developed pneumonia during the hospitalization period are not included in this study subject.

2.3. Constructing the discriminant function

2.3.1. Training dataset and test dataset

590 study subjects are divided into two classes of

sizes 400 and 190 randomly, and called the training dataset and test dataset, respectively. The training dataset is used for constructing the discriminant function and test dataset is used for evaluating the behavior of the constructed function.

2.3.2. Items

In addition to the age at hospitalization, gender and the presence or absence of hospitalization by ambulance, the items taken into account are 606 items that are recorded in the Receipt data at the first and second days of hospitalization. All of those items in the Receipt data are binary data. The age of hospitalization was dichotomized under and above 65 years old.

2.3.3. Objective variable

In DPC hospitals severities of pneumonia patients are given by a measuring system called the A-DROP system. [5] The patient whose severity score is zero at hospitalization given by the A-DROP system is defined to be mild, while the other scores are defined to be severe in this study. The dichotomized severity data are used as objective variable in this study.

2.3.4. Selection of explanatory variables

The explanatory variables of the discriminant function are selected by 3 steps. In the first step, simple logistic regression is applied to each item by using the binary severity data (mild or severe) as objective variable, and items whose p-value are less than 20% are selected from 606 items. In the second step the items selected in the first step are classified into several classes so that items in the same class have weak correlations, using the graphical modeling technique.[6] Using items in a class as explanatory variables and severity as objective variable, items in the class are further selected by using stepwise logistic regression. In the last step Akaike's Information Criterion(AIC) is computed by establishing logistic regression model whose objective variable is the binary severity data (mild or severe) and explanatory variables are variables select in the second step and finally the logistic model that has the smallest AIC is selected as the best model.

2.3.5. The best cut-off point

Prediction probability of being a severe patient is computed for all patients in training dataset using the best model. A patient is decided to be severe if and only if his/her prediction probability is larger than or equal to c. The c is called the cut-off point. Giving the value of c from 0.1 to 0.9 (step 0.1) compute the sensitivity and specificity of the decision and the value of c that has the largest sum of sensitivity and specificity are selected as the best cut-off point. The decision model that is constructed by the best model with the best cut-off point is called the discriminant function in this paper.

2.4. Evaluation of the discriminant function

Applying the constructed discriminant function to each patient in the test dataset, the severity (mild or severe) of the patient is estimated. The results and the binary severity data (mild or severe) obtained from A-DROP system are cross classified into a 2x2 table, and finally the sensitivity and specificity are computed from the table. They are used for evaluating the behavior of the discriminant function. For all statistical analyzes, STATA/IC 14.0 was used.

3. Results

20 items are selected in the first step, excluding items with less than 10 patients. These 20 items are given in Table 1.

Table 1. 20 variables selected in the first step

①Age at hospitalization, ②Gender, ③Presence or absence of hospitalization of ambulance, ④Oxygen inhalation, ⑤Installation of indwelling catheter, ⑥Total bilirubin, ⑦Blood gas analysis, ⑧Biochemical examination 2 decision fee, ⑨Lmmunological examination decision fee, ⑩Microbiological examination decision fee, ⑪Electrocardiography, ⑫Ardiac ultrasoud examination, ⑬D imaer, ⑭Cerebral natriuretic peptide, ⑮Sample inspection management addition 4, ⑯Dietary expenses at hospitalization 1, ⑰Resprene tablets 20mg, ⑱Astomin tablets 10mg, ⑲Rob tablets 60mg, ⑳Carcysteine tablets 500mg

Fig 1 shows the result of the graphical modeling technique in the step 2, where a line connecting two items indicates a significant relationship between items with significance level less than 0.01.

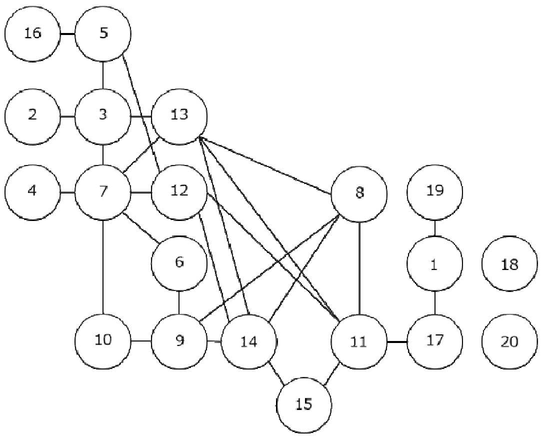

Figure 1. Relationship of items by graphical modelling technique in Step 2

4 different combinations of items were selected in the second step, which are listed in the 2nd column in Table 2. We call those combinations, M1, M2, M3 and M4, respectively. The results of AIC computed from each model are shown in the last column in Table 2. The table shows that the minimum AIC value is attained by M3.

Table 2. 4 different sets of explanatory variables selected in Step 2

Model	Explanatory variable	AIC
M1	Age at hospitalization, Gender, Blood gas analysis, Astomin tablets, Biochemical examination 2 decision fee, Sample inspection management addition 4	257.6
M2	Age at hospitalization, Gender, Blood gas analysis, Astomin tablets, Ardiac ultrasoud examination,	258.0
M3	Age at hospitalization, Gender, Blood gas analysis, Astomin tablets, Ardiac ultrasoud examination, Cerebral natriuretic peptide	256.9
M4	Age at hospitalization, Gender, Blood gas analysis, Astomin tablets, Sample inspection management addition 4	261.5

The selected model, namely M3, is given as follows.

$$\mathrm{Log}\{P(sever)/1\text{-}P(sever)\}=$$
$$-2.052+3.033X1+1.038X2+0.859X3+1.444X4+$$
$$0.834X5+0.722X6.$$

Where X1 = { 1: Age≧65, 0: Age<65 }, X2 = {1: Male, 0: Female }, X3 = {1: Blood gas analysis, 0: No Blood gas analysis }, X4 = {1: Astomin tablets 10mg, 0: No Astomin tablets 10mg }, X5 = {1: Electrocardiography, 0: No Electrocardiography }, X6 = {1: Cerebral natriuretic peptide, 0: No Cerebral natriuretic peptide }.

The cut-off point was selected as c=0.8, thus the final decision rule is given as follows.

P(severe) > 0.8 ⇒ Judge as severe,
P(severe)≦0.8 ⇒ Judge as mild.

Finally, the established discriminant function is applied to the test dataset. The results are summarized in Table 3. The table shows that the sensitivity and specificity of the discriminant function are 0.88 and 0.74, respectively.

Table 3. Results of the established discriminant function when applied to the test data

		A-DROP*		
		mild	sever	total
established discriminant function	mild	26	18	44
	sever	9	137	146
	total	35	155	190
sensitvity	0.884			
specificity	0.743			

*A-DROP score : 0 points = mild、1 or more = severe

4. Discussion

Among those six explanatory variables adopted in our discriminant function, four variables, age at a hospitalization, gender, blood gas analysis, and cerebral natriuretic peptide, are also taken into account in the severity judgment system by A-DROP system. Astomin tablets and electrocardiography are included in our discriminant function, whereas consciousness disturbance and blood pressure are considered in the A-DROP system. The difference is due to the fact that blood pressure and consciousness disturbance are items not included in the Receipt data, and that Astomin and electrocardiography are items served after hospitalization and do not exist when A-DROP system is applied. Since Astomin tablets are prescribed to many more mild patients than severe patients as cough medicine, and also electrocardiography examination is applied to many more severe patients than mild patients, we think it reasonable that these two variables are included in our discriminant function.

The A-DROP system defines score 1 or 2 to be moderate, but we defined it severe in this paper. It is because if we apply the definition of the A-DROP system to patients in non DPC hospitals it is expected that there would be only few patients who are decided to be severe. The difference of definition would be no problem so long as the developed function is used for adjusting for severities of patients between hospitals. Severity of disease is not given in many medical institutions except for DPC hospitals, but Receipt data are available from any hospitals in Japan. The established discriminant function is based on Receipt data, and would be able to apply various problems in flexible other than comparative studies between hospitals.

5. Conclusion

A function is established in this paper for discriminating between severe and mild pneumonia patients based on Receipt data that are routinely available in all medical institutions in Japan. The sensitivity and specificity of the discriminant function are shown to be 0.88 and 0.74, respectively.

Conflict of interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Ministry of Health, Labor and Welfare. Estimated number of patients (per day) by classification of diseases: Patient Survey 2014. [http://www.mhlw.go.jp/english/database/db-hss/dl/sps_2014_01.pdf (cited 2017-Dec-21)].

[2] Miki M, Watanabe A. Does the epidemiology of pneumonia tell the truth? : The mortality of pneumonia will indicate the future of the chest physicians. Annals of The Japanese Respiratory Society 2013; 2(6): 663-671. (in Japanese)

[3] Matsuda S, Ishikawa KB, Kuwabara K, Fujimori K, Fushimi K, Hashimoto H. Development and use of the Japanese case-mix system. Eurohealth 2008; 14: 25-30.

[4] Ishii M. DRG/PPS and DPC/PDPS as prospective payment systems. Japan Med Assoc J 2012; 55: 279-291.

[5] Miyashita N, Matsushima T, Oka M. Japanese Respiratory Society. The JRS Guidelines for the management of community acquired pneumonia in adults: an update and new recommendations. Intern Med 2006; 45: 419-428.

[6] David Edwards. Introduction to graphical modeling, Springer, 2nd Edition, 2003.