

A Comparison of the Charlson Comorbidities Derived from Medical Language Processing and Administrative Data

Jen-Hsiang Chuang, M.D., M.S.^{1,3}, Carol Friedman, Ph.D.^{1,2}, George Hripcsak, M.D., M.S.¹

¹Department of Medical Informatics, Columbia University, New York, NY

²Department of Computer Science, Queens College CUNY, New York, NY

³Institute of Health Informatics & Decision Making, National Yang-Ming University, Taiwan

The objective of this study was to develop a medical language processing (MLP) system, which consisted of MedLEE and a set of inference rules, to identify 19 Charlson comorbidities from discharge summaries and chest x-ray reports. We used 233 cases to learn the patterns that were indicative of comorbidities for developing the inference rules. We then used an independent data set of 3,662 pneumonia patients to identify comorbidities by MLP compared with administrative data (ICD-9 codes). A stratified random sample of 190 records from disagreement cases was manually reviewed. The sensitivity, specificity, and accuracy for the MLP system/ICD-9 codes in this testing set were 0.84/0.16, 0.70/0.30, and 0.77/0.23 respectively. Thirteen of the 19 comorbidities studied were underreported in the administrative data. The kappa values ranged from 0.19 for peptic ulcer to 0.70 for lymphoma. We conclude that comorbidities derived from natural language processing of medical records can improve ICD-9-based approaches.

INTRODUCTION

A comorbidity is a clinical condition that exists before a patient's admission to the hospital, is not responsible for causing the hospitalization, and is likely to be a significant factor influencing mortality and resource use in the hospital¹. Identification of patient comorbidities is crucial for enrolling eligible patients into a clinical trial, implementing a clinical decision support system², and developing risk adjustment models for evaluating medical services. Charlson³ developed a list of 19 comorbidities and produced a weighted index, which took into account both the number and the seriousness of the comorbidities. She used medical record data from 604 consecutively admitted patients to empirically construct the Charlson index based on their one-year survival status after admission. Due to relative inexpensiveness and availability of administrative databases, researchers usually used different compositions of ICD-9 (International Classification of Disease) codes to construct their adapted Charlson indices for risk-adjusted outcomes research.

However, administrative databases tend to underestimate both the comorbid conditions and the Charlson index compared to the clinical data derived from chart review⁴⁻⁶.

Although computerized medical records provide a rich and convenient source of clinical data, most of them are stored in unstructured free text, which is difficult to use for decision support or data analysis. At Columbia Presbyterian Medical Center (CPMC), a medical language processing (MLP) system, called MedLEE (Medical Language Extraction and Encoding System)⁷, has been developed and used for actual patient care. MedLEE uses a grammar to determine the structure of the text and uses a lexicon to classify words and phrases and to specify their target forms. MedLEE, which was originally developed for the domain of chest radiograph reports⁸, has also been extended to discharge summaries, and used to automate a severity score guideline for community-acquired pneumonia². MedLEE performed very well in identifying comorbidities and symptoms from discharge summaries (average sensitivity and specificity were 92% and 93%)². However, inference rules for only five comorbid diseases were developed in that study.

The goals of this study were to develop an MLP system to automatically generate 19 Charlson comorbidities from discharge summaries and chest x-ray (CXR) reports, to evaluate whether individual comorbidity identified by MLP correlated with hospital deaths, to compare the prevalence rates of Charlson comorbidities derived from MLP with the ones derived from ICD-9 codes, and to determine how well each predicted inpatient mortality.

METHODS

Study population

This retrospective study focused on all adult pneumonia patients (20 years and older), admitted to CPMC after Jan. 1, 1995 and discharged before Jan. 1, 2001. Pneumonia cases were identified by ICD-9-CM principal diagnosis codes (480-483, 485-487.0).

Exclusion criteria included cases where there were no computerized discharge summaries, the patient's discharge status was "left against medical advice", "transferred to acute facilities" or "transferred to hospice facilities", or cases contained in the training set used for developing the inference rules. For patients with multiple admissions for pneumonia, we chose only the last hospitalization with an available discharge summary to be included into the study.

ICD-9 comorbidities

The ICD-9 comorbidities were constructed using ICD-9-CM codes obtained from the Clinical Data Warehouse at CPMC, which was funded as an infrastructure project starting in 1994. We linked each index admission to the patient's date of birth, gender, and race, and we retrieved linked ICD-9-CM codes for all previous visits for one year prior to the index admission. We used the algorithm developed by Deyo⁹ for ICD-9-CM adaptations of the Charlson comorbidities. Because some of the diagnoses included in the Charlson comorbidities, such as acute myocardial infarction, new stroke, congestive heart failure, etc., could be complications during hospitalizations, Deyo included such diagnoses into ICD-9 comorbidities only if they occurred during the year prior to the index admission. However, other diagnoses, such as old myocardial infarct, diabetes, connective tissue disease, etc, were included in the ICD-9 comorbidities whether they occurred during the index admission or a prior visit. Finally, we adopted Elixhauser's algorithm to avoid double-counting closely related comorbidities¹.

MLP comorbidities

This MLP system consisted of MedLEE and a set of inference rules processing the structured XML output generated by MedLEE¹⁰. The contributions of this study were to develop the inference rules and to expand the MedLEE lexicon for automated abstraction of comorbidities.

To abstract various comorbid diagnoses, we had to define our own comorbidities (Table 1), which were derived from the definitions of the Charlson comorbidities³. We studied the relevant terms in ICD-9-CM and ICD-O-3¹¹, and investigated the natural histories of the comorbid diseases introduced in the medical textbooks or other online resources. We used each relevant term or corresponding key words to search all discharge summaries in a given year to analyze how physicians used the term in this setting. We also used the same approach by searching the MedLEE lexicon to analyze whether it had been contained in the lexicon. If a term was not contained in the lexicon and it was possible that physicians

would use it in the discharge summaries, we either wrote a specific inference rule for correctly identifying this concept from the MedLEE output if it was possible or added a new term to the lexicon.

Table 1 Brief definitions of comorbidities used for developing medical language processing system

Comorbidity	MLP criteria
AIDS	Definite AIDS or AIDS related complex
Cancer	History of any cancer ¹¹ except basal- or squamous-cell cancer of the skin
CHF	History of CHF or LVEF \leq 35% or cardiac diseases with pulmonary edema by CXR reports at admission
Ch. renal dis.	History of chronic renal failure, dialysis, or post renal transplantation
COPD	History of asthma, emphysema, chronic bronchitis
CTD	History of SLE, systemic sclerosis, rheumatoid arthritis, etc.
CVA	History of transient ischemic attack or stroke with minor or no sequelae
Dementia	Documented notion of dementia
Diabetes	History of diabetes without DM C'x
DM C'x	History of diabetes with retinopathy, nephropathy, or neuropathy
Hemiplegia	History of hemiplegia, paraplegia, or quadriplegia
Leukemia	History of acute and chronic myelogenous or lymphocytic leukemia, and polycythemia vera
Lymphoma	History of Hodgkins, myeloma, lymphosarcoma, etc.
Metastasis	Documented metastatic cancer
MI	History of myocardial infarction
MLD	History of chronic hepatitis or cirrhosis without portal hypertension
MSLD	History of cirrhosis with portal hypertension +/- variceal bleeding
Peptic ulcer	History of peptic ulcers
PVD	History of claudication or prior limb arterial revascularization, etc.

*Abbreviation: CHF: congestive heart failure; Ch. renal dis.: chronic renal disease; COPD: chronic pulmonary disease; CTD: connective tissue disease; CVA: cerebrovascular disease; CXR: chest x-ray; DM C'x: diabetic complication; LVEF: left ventricular ejection fraction; MI: myocardial infarction; MLD: mild liver disease; MSLD: moderate or severe liver disease; PVD: peripheral vascular disease; SLE: systemic lupus erythematosus.

To establish a training set for developing the inference rules, we obtained a set of 300 discharge

summaries, which were randomly sampled from the pool of discharge summaries from the year 1995. We excluded non-adult patients, maternal cases, and patients admitted for mental diseases. There remained 233 cases. We developed an abstraction manual, which provides comorbid definitions, data locations in the discharge summaries, lists of abbreviations and synonyms for comorbidities, and explicit rules to clarify missing or ambiguous data. One of the authors used the developed manual to manually abstract the Charlson comorbidities and other risk variables from these 233 training cases.

To develop the inference rules, we studied the patterns in 233 discharge summaries that were indicative of a comorbidity or complication (a complication is defined as an adverse event, occurring during hospitalization, which is independent of the patient's underlying disease¹²) based on the natural history of a given disease and the locations of the findings in different sections of discharge summaries and CXR reports. Among these, congestive heart failure (CHF) is an ill-defined diagnosis. We used a previously developed inference rule⁸ to interpret a series of the patient's parsed CXR reports which were performed during hospitalization to distinguish whether any pulmonary edema occurred before or after admission. If pulmonary edema was induced immediately after surgery or the patient did not have any cardiac diseases (determined by parsed discharge summaries), such cases were excluded from the cases of CHF complication (see Table 1 for definition of CHF comorbidity). Determinations of other comorbidities or complications were only based on the findings from the parsed discharge summaries.

We used MedLEE to parse all medical records, imported the generated XML files to SAS version 8.2 (SAS Institute Inc, Cary, NC), and transformed XML data to SAS datasets. All of the inference rules were implemented using the SAS language. In addition, we post-processed the XML output to incorporate the dates when the CXRs were performed to facilitate temporal reasoning. We tested the developed inference rules on the 233 training cases. This test identified problems leading to further refinement in the MedLEE system, the inference rules, and the abstraction manual.

Evaluation on the pneumonia cases

We applied Deyo's algorithm to the pneumonia cases for determining the ICD-9 comorbidities compared with the comorbidities determined by the MLP. To evaluate and improve the performance of the MLP system to correctly classify comorbidities in a very

large scale independent data set, we used stratified random sampling to select 190 disagreement cases between MLP and ICD-9 systems by comorbidity. Although we had expected that applying the MLP system to this testing set would perform worse, our purpose was to use the limited resources to try to improve the system. One of the authors applied the definitions of comorbidities in the abstraction manual to the chosen discharge summaries and CXR reports to determine the true comorbid status for each case. Overall sensitivity, specificity, and accuracy for MLP and ICD-9 systems among these 190 testing cases were calculated.

Statistical analysis

We used chi-square test to evaluate the association of individual MLP comorbidity with hospital mortality. A valid comorbidity was expected, at least, to have an odds ratio (OR) for hospital mortality greater than 1. Stepwise logistic regression was used to assess the validity of comorbidities determined by different methods to predict the inpatient mortality after controlling for patient age (age was transformed to three dummy variables to represent <60, 60-69, 70-79, and ≥ 80 years). To assess the performance of each system on the pneumonia cases, area under the receiver operating characteristic (ROC) curve and Hosmer-Lemeshow goodness-of-fit test¹³ were used to measure discrimination and calibration respectively. A high p value for goodness-of-fit test would suggest good fit, while a low p value ($p < 0.05$) would indicate lack of fit¹⁴. A non-parametric approach was used to compare two correlated ROC curves¹⁵. We also computed kappa values (range from -1 to 1; higher kappa is better) to measure the agreement of individual comorbidity determined by two different methods. All analyses were performed using SAS.

RESULTS

We wrote 13 main SAS programs to determine 19 comorbidities. CANCER program determines comorbidities of cancer, lymphoma, leukemia, and metastasis. Each of CVA, DM, and LIVER programs determines two related comorbidities. The sizes of the programs ranged from 2 KB (including COPD, DEMENTIA, MI, CTD, and ULCER) to 17KB (including CANCER, AIDS, and CHF).

There were 5,546 hospitalizations for pneumonia during the study period. After applying the exclusion criteria, we found that 3,662 unique patients met the study criteria. The mean age was 65.5 years old, 54.4% were women, and inpatient mortality rate was 9.1%. The study sample consisted of 35.4% whites, 28.6% blacks, 22.5% Hispanics, and 13.5% others.

The sensitivity, specificity, and accuracy for the MLP system/ICD-9 codes in the testing set of 190 sampled disagreement cases were 0.84/0.16, 0.70/0.30, and 0.77/0.23 respectively.

Table 2 compares the prevalence rates of the Charlson comorbidities in 3,662 patients detected by ICD-9 with the ones detected by MLP. Thirteen of the 19 comorbidities studied were underreported in the administrative data. The kappa values ranged from 0.19 for peptic ulcer to 0.70 for lymphoma.

We found ORs of Hemiplegia, COPD, CTD, MLD, DM C’x, and AIDS determined by MLP for hospital mortality were all less than 1. MLP system also identified 8.9% patients with CHF complications, 0.4% with CVA complications and 0.3% with acute MI during hospitalization.

Table 2 Prevalence rates of comorbidities by data source and calculated kappa and agreement by comorbidity in 3662 patients

Comorbidity	ICD-9	MLP	Kappa (Agreement)
AIDS	2.9%	5.1%	0.61 (0.97)
Cancer	5.8%	12.0%	0.31 (0.89)
CHF	12.7%	22.5%	0.40 (0.82)
Ch. renal dis.	4.9%	6.1%	0.41 (0.94)
COPD	29.5%	23.6%	0.66 (0.87)
CTD	2.2%	2.8%	0.68 (0.98)
CVA	5.2%	13.5%	0.34 (0.89)
Dementia	6.0%	9.8%	0.43 (0.92)
Diabetes	16.1%	14.0%	0.67 (0.92)
DM C’x	3.4%	6.3%	0.48 (0.95)
Hemiplegia	1.8%	0.7%	0.28 (0.98)
Leukemia	0.9%	1.3%	0.49 (0.99)
Lymphoma	2.2%	2.4%	0.70 (0.99)
Metastasis	4.8%	4.7%	0.46 (0.95)
MI	3.6%	10.7%	0.27 (0.90)
MLD	1.3%	1.0%	0.51 (0.99)
MSLD	0.9%	0.7%	0.53 (0.99)
Peptic ulcer	3.0%	4.8%	0.19 (0.94)
PVD	2.7%	5.3%	0.27 (0.94)

After finding the incorrectly classified cases by MLP in the testing set, we refined the inference rules or definitions again if changing the inference rules or definitions could solve the problems we identified. We implemented stepwise logistic regression to develop (i) original MLP, (ii) refined MLP, and (iii) ICD-9 risk-adjustment model for predicting hospital mortality using pneumonia cases. The following comorbidities were entered into the refined MLP model: cancer (OR=1.6), metastasis (OR=3.7), lymphoma (OR=2.3), CHF (OR=1.5), PVD

(OR=1.6), and MSLD (OR=5.6). In the original MLP model, the entered comorbid variables were the same except for slightly different ORs. However, cancer (OR=2.1), metastasis (OR=3.1), dementia (OR=1.7), CHF (OR=1.4), MSLD (OR=4.7), and MLD (OR=2.9) were entered into the ICD-9 model. The area under the ROC curve was 0.695 in the original MLP system and 0.705 in the refined MLP system ($p = 0.022$). There was no significant difference in the area under the ROC curves between the refined MLP and ICD-9 systems (ROC=0.704, $p = 0.904$). Goodness-of-fit tests indicated that the model performed well in the refined MLP system ($p = 0.296$) but not well in the original MLP system ($p = 0.091$) or in the ICD-9 system ($p = 0.041$).

DISCUSSION

In this study, we developed an MLP system to identify comorbidities from discharge summaries and CXR reports and demonstrated that the MLP system detected more comorbidities in cases and was more accurate than ICD9 codes. In addition, the MLP system could be tuned to distinguish comorbidities from complications, while 30% of medical complications identified by ICD-9-CM codes lacked any documented evidence¹⁶.

We believed that some of the differences in prevalence rates of comorbidities between MLP and ICD-9 systems might be due to the different definitions or different classification algorithms in each system. For example, cases with both diabetes and chronic renal disease were classified by the MLP to the diabetic complication group only. On the other hand, the reasons for underreporting comorbidities in the ICD-9 codes may be because complications take precedence in coding over comorbidities during hospitalization⁴ and patient’s previous illnesses are treated in the other hospitals.

This study found low to moderate kappa values between the MLP and ICD-9 comorbidities; however, our results were very similar to other studies comparing the comorbidities derived from chart review of medical records with the ones derived from ICD-9 codes^{5,6}.

ORs of six comorbidities detected by MLP for hospital mortality were less than 1. It is possible that physicians only record a short death note and forget to mention these diseases in the discharge summaries once the patient dies in the hospital. In addition, AIDS determined by MLP and ICD-9 both had very low mortality rates in this sample. It may be due to improvements in AIDS care in recent years. However, in the Charlson index the same weighting

score of six was assigned to AIDS and metastasis comorbidities 14 years ago. When applying the original index for predicting hospital mortality, researchers should be cautious.

The refined MLP system achieved the same discrimination (ROC=0.705) for predicting hospital mortality as the ICD-9 system. Ghali¹⁷ used the Charlson index to predict hospital mortality for patients receiving coronary artery bypass graft surgery and the reported ROC was 0.704. We plan to use MLP to study more comorbid diseases, e.g., hypertension, and develop a new comorbidity index for risk adjustment modeling.

CONCLUSIONS

This study demonstrated that natural language processing of medical records detected more comorbidities than an ICD-9 code-based approach. The accuracy of the MLP system for detecting comorbidities was also better than the ICD-9 codes. We plan to continuously improve the performance of the MLP system and evaluate it further.

Acknowledgments

The authors thank Drs. Stephen Johnson and Alan Moskowitz for their valuable comments on the planning proposal of this project; Lyudmila Shagina for her help in adding new terms to the MedLEE lexicon; Alla Balbina for her help in accessing data on Clinical Data Warehouse at CPMC. This work was supported in part by National Library of Medicine grants R01 LM06910 and LM06274 and by Agency for Healthcare Research and Quality 1 R18 HS11806. Dr. Chuang was supported in part by the Medical Foundations in Memory of Dr. Chi-Shuen Tsou and Dr. Albery Ly-Young Shen, Taiwan, ROC.

References

1. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8-27.
2. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. *Proc AMIA Symp* 1999:256-60.
3. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-83.
4. Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease. an Australian perspective. *J Clin Epidemiol* 2001;54:687-93.

5. Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol* 1999;52:137-42.
6. Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. *J Clin Epidemiol* 1994;47:1027-32.
7. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1994;1:161-74.
8. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 1995;122:681-8.
9. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613-9.
10. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. *J Am Med Inform Assoc* 1999;6:76-87.
11. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al., editors. International classification of diseases for oncology. Third ed. Geneva: World Health Organization; 2000.
12. Giraud T, Dhainaut JF, Vaxelaire JF, Joseph T, Journois D, Bleichner G, et al. Iatrogenic complications in adult intensive care units: a prospective two-center study. *Crit Care Med* 1993;21:40-51.
13. Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92-106.
14. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957-63.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
16. McCarthy EP, Iezzoni LI, Davis RB, Palmer RH, Cahalane M, Hamel MB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care* 2000;38:868-76.
17. Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA. Searching for an improved clinical comorbidity index for use with ICD-9- CM administrative data. *J Clin Epidemiol* 1996;49:273-8.