Calculating Prevalence of Comorbidity and Comorbidity Combinations with Diabetes in Hospital Care in Sweden Using a Health Care Record Database

Hideyuki Tanushi[1], Hercules Dalianis[1], and Gunnar H Nilsson[1,2]

[1] Department of Computer and Systems Sciences (DSV), Stockholm University, Forum 100, SE-164 40 Kista, Sweden
[2] Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, SE-141 83 Huddinge, Sweden

Abstract. We have studied the prevalence of comorbidity in the Stockholm EPR corpus containing almost 600,000 patients from 900 clinics using the ICD- 10 codes assigned to each patient record. The proportion of patients with a valid ICD-10 code was 83.0%, and 41.5% of these had at least one comorbidity. The most frequent comorbidity combination with type 2 diabetes was essential hypertension (43.1%). Our approach seems feasible for large scale analysis of diagnostic codes in EPR databases.

Keywords: comorbidity, chronic disease, ICD-10, medical records systems, computerized medical record, Sweden

1 Introduction

Today a large amount of electronic patient records (EPR) are produced, which are rarely reused. The EPR systems are also more and more centralized covering both several hospitals and clinics. In the Scandinavian countries we have a unique social security number for all citizens following us from birth to death, as well as from clinic to clinic. Diagnoses are often coded by ICD [1]. Among clinical researchers there is a need to study the comorbidity of diseases among patients [2]. An issue that arises is if we can observe any correlation between the comorbidity in our immense database and clinical researchers' findings. Coding and classification of diseases in medical records of individuals have received little attention in the research area. Nowadays health care is more and more required to deal with the management of individuals with multiple coexisting diseases, namely comorbidity [2].

Comorbidity is the presence of no less than two distinct conditions in an individual [2]. One of the attempts for establishing classification of comorbidity is a scheme of taxonomy for classifying diabetic comorbid ailments and prognostic value of the classification [3]. Another attempt is made by Charlson et al. [4] who have developed a prognostic taxonomy for comorbid conditions able to predict the risk of short term mortality for patients enrolled in longitudinal studies.

Starfield et al. [5] describe a categorization of morbidity in order to present a system to measure and compare the burden of illness of patients over time in different ambulatory care facilities, and to show how the system can predict utilization and charges, both concurrently and prospectively.

As the population of elderly people is increasing in many countries the prevelence of chronic diseases are expected to rise. Schellevis et al. [6] and van Weel [7] conducted comorbidity analysis for chronic diseases by calculating combination of these chronic diseases in general practice in the Netherlands. In van Weel's study [7] comorbidity for the ten most common chronic diseases is calculated from approximately 12,000 patients. Davila and Hlaing [8] examined patients admitted with primary diagnosis of essential hypertension and analyzed the frequency of their secondary diagnosis.

Type 2 diabetes is defined as a chronic disease characterized by reduced insulin sensitivity in target tissues [9]. The number of patients with diabetes worldwide is estimated more than 220 million and type 2 diabetes accounts for 90% of them [10]. Several comorbid conditions to type 2 diabetes have been identified in previous studies, such as obesity, which reduces insulin sensitivity, and dyslipidemia and hypertension, which alter vascular and cardiac structure [9].

Detecting comorbidity in a large population is of clinical interest due to the fact that it may reveal new information useful for cause of diseases as well as for new treatment strategies. The aim of this study is to analyze comorbidity in clinical hospital setting in Sweden using an EPR database, and to investigate comorbidity combinations with diabetes.

2 Material and Methods

The tools used in this work[1] are SQL query language and Java programming language. Fig. 1 shows a simple illustration for procedure of comorbidity analysis.

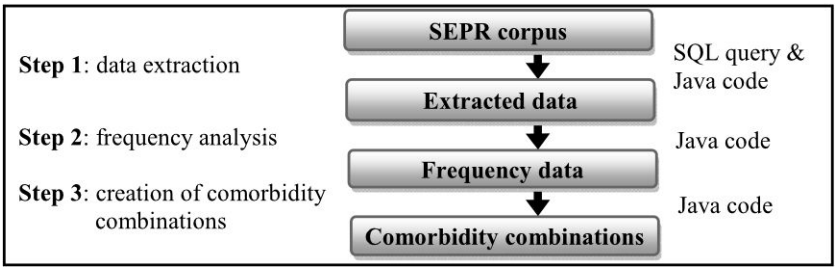

Fig. 1. Illustration for procedure of comorbidity analysis

[1] This research has been carried out after approval from the Regional Ethical Review Board in Stockholm (*Etikprövningsnämnden i Stockholm*), permission number 2009/1742-31/5.

The Stockholm EPR (SEPR) corpus that we have access to contains almost 900 clinics with almost 600,000 patients, that are registered with their social security number, gender, age, admission and discharge date of the patient as well as the ICD code of the diagnosis encompassing the years 2006, 2007 and the first half of 2008 [11]. The unstructured information consists of free text under different headings. Diseases in the SEPR corpus are mainly coded by ICD-10. According to Dalianis et al. [11], 34% of the patients did not have any ICD-10 code.

The main issue in this work is to explore what comorbidity combinations each individual patient has in their medical history. Hence, the data of patient's id, birth year, and diagnosis code are extracted from the SEPR database with a SQL query. In the extracted data there are 2,756,082 diagnosis records from 584,600 patients (=N).

The diagnosis codes in the extracted data are, however, not written in a unified way. For instance, the alphabet is written either in upper-case (e.g. A00) or in lower-case (e.g. a00), and different marks are used (e.g. A00-0, A00.0, A00,0, etc). Hence, the extracted codes need to be normalized. A code expression in upper-case without any marks (e.g. A000) is used in the analyses. Also, there are missing, or syntactically wrong and not to correct diagnosis codes (i.e. invalid codes) in the extracted diagnosis data. Examples of invalid codes are: *AF063*, *KV9*, *KVÅ*, *NCU49*, *NFJ09*, etc. There are 1,772,013 valid ICD-10 codes (13,450 different full level codes and 1,956 different level 3 codes) and 805,568 invalid ICD-10 codes in the extracted data.

Some patients have duplicate ICD-10 codes in their medical records (178,501 duplicate cases). This is presumably because these patients have received medical care in more than one institution. ICD-10 codes are wrong or missing in many cases. In our case the percentage of invalid ICD-10 codes in the extracted data is approximately 31%.

Only valid ICD-10 codes are used and converted to level 3 codes (i.e. truncating the first three characters), and other codes are not converted or interpreted by free text diagnoses. When converting the valid full level ICD-10 codes into level 3 codes, a patient can have the same ICD-10 code no less than twice (e.g. before converting A00, A001 and A009 $\Rightarrow$ after converting A00, A00 and A00). Multiple ICD codes for a patient are therefore counted only once. Also, ICD-10 codes in chapters 19–21 are excluded from the frequency analysis because these codes include causes of diseases and other factors related to health care, and therefore not considered as diseases.

Creating comorbidity combinations for type 2 diabetes makes it possible to focus the analysis on this diagnosis. Also, there are similar comorbidity analyses in previous research which can be a methodological reference to this work as mentioned in the background [6–8]. On the ground of the above comorbidity combinations are used as a method for the comorbidity analysis in this work. For the comorbidity analysis in the following chapter, the frequency of comorbidity combinations for type 2 diabetes is calculated. The frequency is calculated by

counting how many type 2 diabetes patients have additional coded diseases, e.g. comorbid diagnoses.

3 Results

Out of the 584,600 (=N) patients the number of patients with a valid ICD-10 code was 485,271 (83.0%), and 242,435 (41.5%) had at least one comorbidity (see Table 1).

Table 1. Frequency of patients by the number of valid ICD-10 codes (except codes in chapters 19-21)

No valid ICD-10 code	
No. of valid ICD-10 codes	*No. of patients*
0	99,329

At least one valid ICD-10 code	
No. of valid ICD-10 codes	*No. of patients*
1	242,836
2+	242,435
Total	485,271

More than one valid ICD-10 code	
No. of valid ICD-10 codes	*No. of patients*
2	104,420
3	52,405
4	29,589
5	17,904
6	11,396
7	7,627
8	5,238
9	3,748
10+	10,108
Total	242,435

The total number of patients with type 2 diabetes were 14,162 (=n), out of which 13,487 (95.2%) had at least one comorbidity. The most frequent comorbidity combinations with type 2 diabetes were essential (primary) hypertension (43.1%), and heart failure (17.0%) (see Table 2).

4 Discussion

In our study the proportion of patients with missing valid ICD-10 codes (17.0%) was somewhat lower than expected. A study [11] on the same material showed about 34% missing ICD-10 codes, however another study [12] showed 1.2% missing main diagnosis on other material. Another issue is the quality of the set ICD codes. A study [13] found about 20% wrongly set ICD codes, and Roque et al. [14] found 15.9% wrongly set ICD-10 codes.

The comorbidity figures in our study (41.5%) were higher than for example in Westert et al. [15] showing about one fifth of patients with more than one chronic condition. However, these figures include all health care settings and are limited to chronic diagnoses.

The most frequent comorbidity combinations with type 2 diabetes in our study were essential hypertension (43.1%) and heart failure (17.0%). Our figures

Table 2. Top 20 comorbidity combinations for patients with type 2 diabetes (n=14,162)

	Comorbid diagnosi(e)s	*Description*	*No. of patients*	*%*
1	[I10]	Essential (primary) hypertension, #1	6,099	43.1
2	[I50]	Heart failure, #2	2,410	17.0
3	[I25]	Chronic ischaemic heart disease, #3	2,119	15.0
4	[E10]	Insulin-dependent diabetes mellitus, #4	2,069	14.6
5	[I20]	Angina pectoris, #5	2,042	14.4
6	[I48]	Atrial fibrillation and flutter, #6	1,941	13.7
7	[E78]	Disorders of lipoprotein metabolism and other lipidaemias, #7	1,856	13.1
8	[E78, I10]	cf. #1, #7	1,378	9.7
9	[I10, I20]	cf. #1, #5	1,256	8.9
10	[N39]	Other disorders of urinary system	1,245	8.8
11	[I10, I25]	cf. #1, #3	1,200	8.5
12	[N18]	Chronic renal failure	1,174	8.3
13	[I10, I50]	cf. #1, #2	1,156	8.2
14	[I21]	Acute myocardial infarction	1,109	7.8
15	[I10, I48]	cf. #1, #6	978	6.9
16	[I48, I50]	cf. #2, #6	967	6.8
17	[E10, I10]	cf. #1, #4	960	6.8
18	[I20, I25]	cf. #3, #5	940	6.6
19	[H36]	Retinal disorders in diseases classified elsewhere	905	6.4
20	[I70]	Atherosclerosis	866	6.1

are in line with Caughey et al. [16] reporting about 51–53% comorbidity with hypertension, and 12–42% with cardiovascular disease. However, their study includes only eight chronic diagnoses. Similar comorbidity figures are also found in Finland in a population study by Reunanen et al. [17].

Our method to extract and present information needs to become more streamlined since today there are some manual steps that have to be carried out. The strength of our method is that we use continuously growing everyday clinical data enabling more detailed analysis. The main weakness is that there is the large proportion of patients without a valid ICD-10 code in EPR databases. One method to solve this problem is to match the free text in the patient record with the ICD-10 code's textual description to populate the record with ICD-10 codes (see [14]).

We believe that in the future hospital management and clinical research will monitor and analyze the ICD-10 codes and also SNOMED-CT [18] to assess health care and also to predict future needs.

5 Conclusion

In Swedish hospital care the proportion of patients with a valid ICD-10 codes seemed to be fairly high (83.0%), and the comorbidity about 41.5%. The most

frequent comorbidity combinations with type 2 diabetes were essential hypertension and heart failure. Our study raises questions on the quality and analysis of diagnosis coding in hospital settings. However, our approach seems feasible for large scale analysis of diagnostic codes in EPR databases. Our results on high rates for diabetes comorbidity may have implication for both health care planning and delivery.

In the future we plan to populate our data with more ICD-10 codes extracted from diagnosis expressions in the free text, this is similar to the approach described by Roque et al. [14].

Acknowledgments. We would like to thank Maria Skeppstedt for her invaluable help in posing smart SQL queries to extract data from the Stockholm EPR corpus.

References

1. International Classification of Diseases (ICD), `http://www.who.int/classifications/icd/en`
2. Valderas, J.M., Starfield, B., Sibbald, B., Salisbury, C., Roland, M.: Defining Comorbidity: Implications for Understanding Health and Health Services. Ann Fam Med. 7(4), 357–363 (2009)
3. Kaplan, M.H., Feinstein, A.R.: The importance of classifying initial co-morbidity in evaluating the outcome of diabetes mellitus. J Chronic Dis. 27(7-8), 387–404 (1974)
4. Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, R.C.: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 40(5), 373–83 (1987)
5. Starfield, B., Weiner, J., Mumford, L., Steinwachs, D.: Ambulatory care groups: a categorization of diagnoses for research and management. Health Serv Res. 26(1), 53–74 (1991)
6. Schellevis, F.G., van der Velden, J., van de Lisdonk, E., van Eijk, J.T.M., van Weel, C.: Comorbidity of chronic diseases in general practice. J Clin Epidemiol. 46(5), 469–73 (1993)
7. van Weel, C.: Chronic diseases in general practice: the longitudinal dimension. European Journal of General Practice. 2, 17–21 (1996)
8. Davila, E.P., Hlaing, W.M.: Comorbidities of Patients with Hypertension Admitted to Emergency Departments in Florida Hospitals. Florida Public Health Review. 5, 84–92 (2008)
9. Hollander, P.A., Kushner, P.: Type2 Diabetes Comorbidities and Treatment Challenges: Rationale for DPP-4 Inhibitors. Postgraduate Medicine. 122(3), 71–80 (2010)
10. Fact sheet on diabetes, `http://www.who.int/mediacentre/factsheets/fs312/en/index.html`
11. Dalianis, H., Hassel, M., Velupillai, S.: The Stockholm EPR Corpus Characteristics and Some Initial Findings. In: Proceedings of ISHIMR 2009, 14th International Symposium for Health Information Management Research, Kalmar (2009)
12. *Socialstyrelsen* (The National Board of Health and Welfare, In Swedish): *Kodningskvalitet i patientregistret - Slutenvård 2007.* Artikelnummer: 2009-125-1 (2009)

13. *Socialstyrelsen* (The National Board of Health and Welfare, In Swedish): *Diagnosgranskningar utförda i Sverige 1997-2005 samt råd inför granskning.* Artikelnummer: 2006-131-30 (2006)
14. Roque, F.S., Jensen, P.B., Schmock, H., Andreatta, M., Hansen, T., Søeby, K., Brædkjaer, S., Juhl, A., Werge, T., Jensen, L.J., Brunak, S.: Using electronic patient records to discover disease correlations and stratify patient cohorts, submitted manuscript (2011)
15. Westert, G.P., Satariano, W.A., Schellevis, F.G., van den Bos, G.A.M.: Patterns of comorbidity and the use of health services in the Dutch population. Eur J Public Health. 11(4), 365–72 (2001)
16. Caughey, G.E., Vitry, A.I., Gilbert, A.L., Roughead, E.E.: Prevalence of comorbidity of chronic diseases in Australia. BMC Public Health. 221(8) (2008)
17. Reunanen, A., Kangas, T., Martikainen, J., Klaukka, T.: Nationwide survey of comorbidity, use, and costs of all medications in Finnish diabetic individuals. Diabetes Care. 23(9), 1265–71 (2000).
18. SNOMED Clinical Terms User Guide, `http://www.ihtsdo.org`