Evaluation of a program for identifying patients with diabetes from electronic health records in the information system

Leandro Biagini[a] , Alejandro Mauro[a], Claudio Torres Casanelli[a], Luisa Legorburu[a], Lisette Lucic[a], Marcela Pezzani[a], Pelayo Navarro[a], Daniel Luna[b] , Fernán González Bernaldo de Quirós[b], Marcelo Maira[a]

[a] *Megasalud, Chile*
[b]*Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*

Abstract

Diabetes is a major cause of morbidity and mortality. Multiple strategies have been tried for managing populations that suffer from this disease; they all identified the need for correct patient identification. The objective was to assess the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of HIBA's Terminology Server (TSHIBA) for detection of diabetic patients in the Megasalud legacy health system.

Keywords:

Diabetes, Problem-oriented health record, Sensitivity and specificity, Clinical terminology.

Methods

The Terminology Server (TSHIBA) is a software composed of a local interface vocabulary (thesaurus) mapped to a reference vocabulary, SNOMED CT. The thesaurus is a list of terms created from almost 2 million free text inputs extracted from the clinical data repository. The terms included in the thesaurus are divided into concepts (real clinical entities) and descriptions (different ways of naming these clinical entities). The TSHIBA facilitates the encoding process by analyzing text entered in an EHR by a user.

For the purpose of this analysis, we used TSHIBA to search for diabetic candidates by reading several fields of the Megasalud legacy EHR trying to identify descriptions from the Interface Vocabulary already coded in TSHIBA. Historically, Megasalud clinicians entered symptom and diagnosis data in an unstructured manner (free text). We used TSHIBA Lexical Concept Lookup software service; a keyword search retrieved exact matches or a set of similar concepts from a given subset.

Our objective was to evaluate TSHIBA's ability to correctly identify patients with diabetes. The software "reviewed" 9438 patient medical records who had visited "La Florida" Megasalud Medical Center between January 1st and June 30th 2008. Concurrently, a manual review of these records was done and used as the "Gold Standard".

Results

The results showed a sensitivity of 86.19% (95% CI = 83-89) and a specificity of 98.37% (95% CI = 98-98,6). Positive Predictive Value (PPV) was 76.11% and Negative Predictive Value (NPV) was 99.16%. The Positive and Negative Likelihood Ratio and its confidence intervals were 53 (95% CI = 45-62,4) and 0.14 (95% CI = 0,11-0,17), respectively (see Table 1).

Table 1 – Evaluation of the TS for sensibility and specificity for diabetes

	Manual Revision (+)	**Manual Revision (-)**	
TSHIBA (+)	TP = 462	FP = 145	T = 607
TSHIBA (-)	FN = 74	TN = 8757	T = 8831
	T = 536	T = 8902	T Pat=9438
Sensitivity	TP/(TP+FN) = 86.19%		
Specificity	TN/(TN+FP) = 98.37%		
PPV	TP/(TP+FP) = 76.11%		
NPV	TN/(TN+FN) = 99.16%		
LR+	Sensitivity/(1-specificity) = 53		
LR-	(1-sensitivity)/specificity = 0.14		

Conclusion

We noted that the results achieved in this first terminiology server evaluation using a repository of foreign clinical data was more than satisfactory.