Automated Integration of a Diagnostic Decision Support System with a Large Scale Clinical Information System - Aspects of Feasibility

Gai Elhanan, M.D., James J. Cimino, M.D.
Department of Medical Informatics, Columbia-Presbyterian Medical Center, New York

Background. One role for generic Medical Decision Support Systems (MDDSS) is to serve as sentinels against diagnostic errors of omission. For such a task, MDDSS must be fully integrated into institutional Clinical Information Systems (CIS) and allow for automated transfer of data between the two. The abundance of available data in today's CIS and the loss of contextual relationship in the process make the utilization of these large scale data sets more difficult. We identified two focal points for the submission of large scale data sets to MDDSS: *Organization* and *Interpretation*. Organizational issues stem from the occurrence of multiple instances of identical tests (i.e. *Plasma Sodium Test, Plasma Sodium Ion Test.*…) and that for an MDDSS related tests may represent identical concepts (i.e. *Serum Sodium Test, Plasma Sodium Test*…). Interpretation issues result from the need to re-apply context to the data set to identify the "most representative" type of result among multiple instances of the same test. These processes are executed intuitively by physicians but are poorly understood, and are not incorporated into MDDSS. To cope with the first we utilized the hierarchic and semantic network of the controlled medical vocabulary in use at CPMC[1] while for the second we applied two dedicated submission algorithms. We describe the differential effect of the two algorithms on the diagnoses lists produced by DXplain in a non-clinical evaluation.

Methods.

Data submission algorithms

1 - *Simple summation*. All instances of results for a specific measurable substance were categorized and counted. Based on pre-defined rules a selection was made for the more dominant result.

2 - *Time-scored*. Based on the assumption that results around admission time better represent the disease state. Tests results were scored in an exponentially declining manner based on the elapsed time between admission and the individual test. Scored results were summed-up and selected similar to the simple summation algorithm.

Data sets & Evaluation

Data were collected from randomly selected adult patients who were hospitalized for at least 7 days in non-specialized wards. Each data set was submitted to DXplain using each algorithm. The resulting diagnoses lists were compared.

	Test	Decreased	Normal	Increased	Irrelevant	Include ?	All
22	Intravascular Calcium Test	27	**14**	0	0	Yes No	43
23	Intravascular Uric Acid Tets	3	20	**10**	0	Yes No	33
22	Intravascular Calcium Test	**30**	33	0	0	Yes No	63
23	Intravascular Uric Acid Tets	3	**20**	21	0	Yes No	44

Summ.

Time

Figure 1: At the top the Summation algorithm (Summ.) selected decreased calcium and normal uric acid while the Time-score algorithm (Time) switched the selected findings to normal calcium and increased uric acid. Values for the Time-scored algorithm represent the sum of scored results and not the number of instances, as shown in the Summation algorithm rows.

Results. 231 data sets were submitted, averaging 945 laboratory tests per set (33-14,027), of which 42.5% were used for submission to the MDDSS. The resulting diagnoses lists contained one or more diagnoses in 180 (77.9%) sets. The overall rate of shared diagnoses between lists based on different algorithms was 54.7%. Identical lists were produced for 52 sets only. Switched findings (Fig.1) occurred in 208 data sets (90%, 3.9 per set), resulting in 128 diagnoses lists that differed at varying levels.

Discussion. This study was designed to evaluate the effect of differing approaches toward interpretation of large data sets. It should be expected that different algorithms will result in different diagnoses lists, but the algorithms used represent two reasonable lay-person approaches toward the problem and therefore, the degree for which the diagnoses list varied is significant. Our experience suggests that large scale automated integration between MDDSS and CIS is feasible, but dedicated, standardized algorithms are required to deal with the progression of data over time and to make the results comparable across systems.

1. Elhanan G, Socratous SA, Cimino JJ. Integrating DXplain into a clinical information system using the world wide web. *Proc Amia Annu Fall Symp* 1996;:348-352.