Identification of fraud in healthcare system `upcoding detection`– documentation.

Authors:

Karol Krzyżak - karol.krzyzak@gmail.com

Łukasz Pińkowski -lpinkowski@gmail.com

Table of contents

Codes and scripts

Solution consists of two elements:

- Database based on csv available for problem. Database is provided as SQL file. It is guaranteed to work only with MySQL database server.

- Application written in Java language. It was tested with Java8 only. Application is really simple it just runs some SQL queries and makes internal aggregates on it.

- As a result of application run result file is created. It has csv format and consist DRG groups which are classified as upcoding along with ICD10 codes used for it.

Methodology

The main problem in the task was the form of data - providing them in an aggregated form definitely hampered the search for potential frauds. It seems to us that it would be much easier to work on data in the form of records for individual patients.

Considering the following factors that may affect the change of data over time:

- Change of medical knowledge in the treatment of specific diseases
- Definitions changes within the DRG groups themselves
- Territorial changes in the hospital network
- Changes in the limits of benefits between individual years
- Demographic changes affecting the age of patients (progressing fast enough that they should be taken into account)

We found that the least-prone factor in time should be recognition, i.e. ICD10 codes. We decided to designate these diagnoses, the number of which increased the most, with the simultaneous increase in the average valuation of the DRG groups to which these diagnoses belonged. We took into account the diagnoses, the number of which increased by at least 10% yoy, with the simultaneous increase in the valuation of DRG groups related to them by at least 15%. The above values have been chosen experimentally. We did not take into account records for which the number of diagnoses in the previous year was very small - we considered this as a data error.

The above procedure allowed to designate 21 groups of DRG with potential upcoding. From the received codes it would be necessary to exclude those for whom the definitions of diagnoses in the ICD10 dictionary have changed.

We tried to apply a similar analysis to diagnoses, but the results were less convincing, which is probably related to possible changes in the treatment regimen.

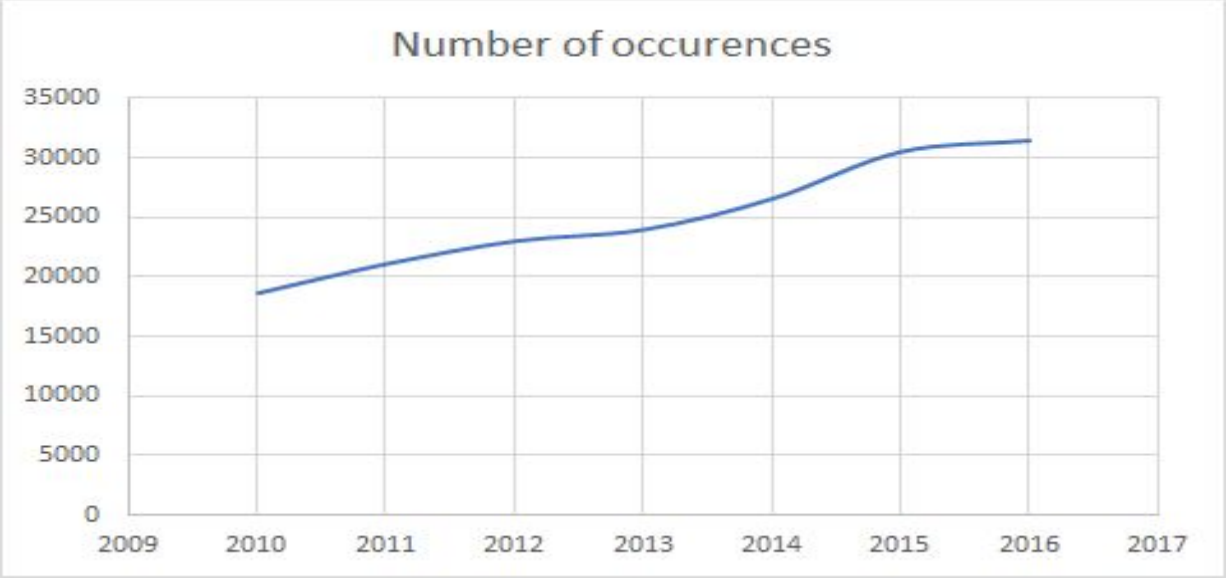
Results

The table below presents the DRG groups received in the results of our program. Together with groups we present related diagnoses, the number of which has increased, and the year in which the increase occurred (in comparison to the previous year)

Code DRG	Code ICD10	Year
03.4656.033.02	I42.9	2016
5.51.01.0003013	J35	2013
5.51.01.0004004	J93.0	2010
	J93.9	2010
	J93.1	2010
	J93.8	2010
	D38.6	2010
	D38.1	2010
	R91	2010
5.51.01.0005004	I21.4	2012
5.51.01.0005062	I49.3	2016
5.51.01.0008096	M05.3	2010
5.51.01.0009007	N62	2012
5.51.01.0009020	T33.8	2010
	N17.9	2010
5.51.01.0010056	C74.9	2011
5.51.01.0011008	D41.0	2012
	D30.7	2012
	D41.4	2012
	C67.2	2012
	C67.8	2012
	C67.9	2012
5.51.01.0011044	D30.7	2010
5.51.01.0011087	N17.9	2010
	N17.9	2012
	N17.0	2010
5.51.01.0012006	I13.9	2010
5.51.01.0014006	B34.9	2013
	A49.9	2013
	R59.9	2013
	I88.9	2013
	L08.9	2013
	B34.8	2013
	B27.9	2013
	A38	2013
	B08.2	2013
	L04.0	2013

5.51.01.0014026	C71.6	2016
5.51.01.0014027	C47.9	2016
5.51.01.0014027	C71.6	2016
5.51.01.0014028	C74.9	2011
5.52.01.0000243	J96.0	2010
5.52.01.0001363	C75.9	2011
	L40.5	2010
	C75.9	2010
	E80.2	2011
	K43.9	2013
5.52.01.0001424	Z51.0	2013
	C73	2015
5.52.01.0001470	C73	2015
	C61	2015
5.52.01.0001475	I35.0	2011

Below you can find an in-depth analysis for one pair of DRG codes, ICD10 - specifically for 5.52.01.0001475, I35.0.



The analysis showed an increase in 2011 related to the increase in the value of valuation of one of the DRG groups designated for this diagnosis. It is possible that further increases are also related to upcoding.