Data Mining as a Tool of Quality in Obstetrics

Diana Lungeanu[a], Stefan Holban[b], Daniela Zaharie[c], Dan B. Navolan[d], Raul F. Horhat[a], Calin Muntean[a], Anca Kigyosi[a, d]

[a] *Department of Medical Informatics, University of Medicine and Pharmacy of Timisoara, Romania*
[b] *Department of Computer Science, Polytechnic University of Timisoara, Romania*
[c] *School of Mathematics and Informatics, West University of Timisoara, Romania*
[d] *University Hospital of Obstetrics-Gynaecology „Dr. Dumitru Popescu" from Timisoara, Romania*

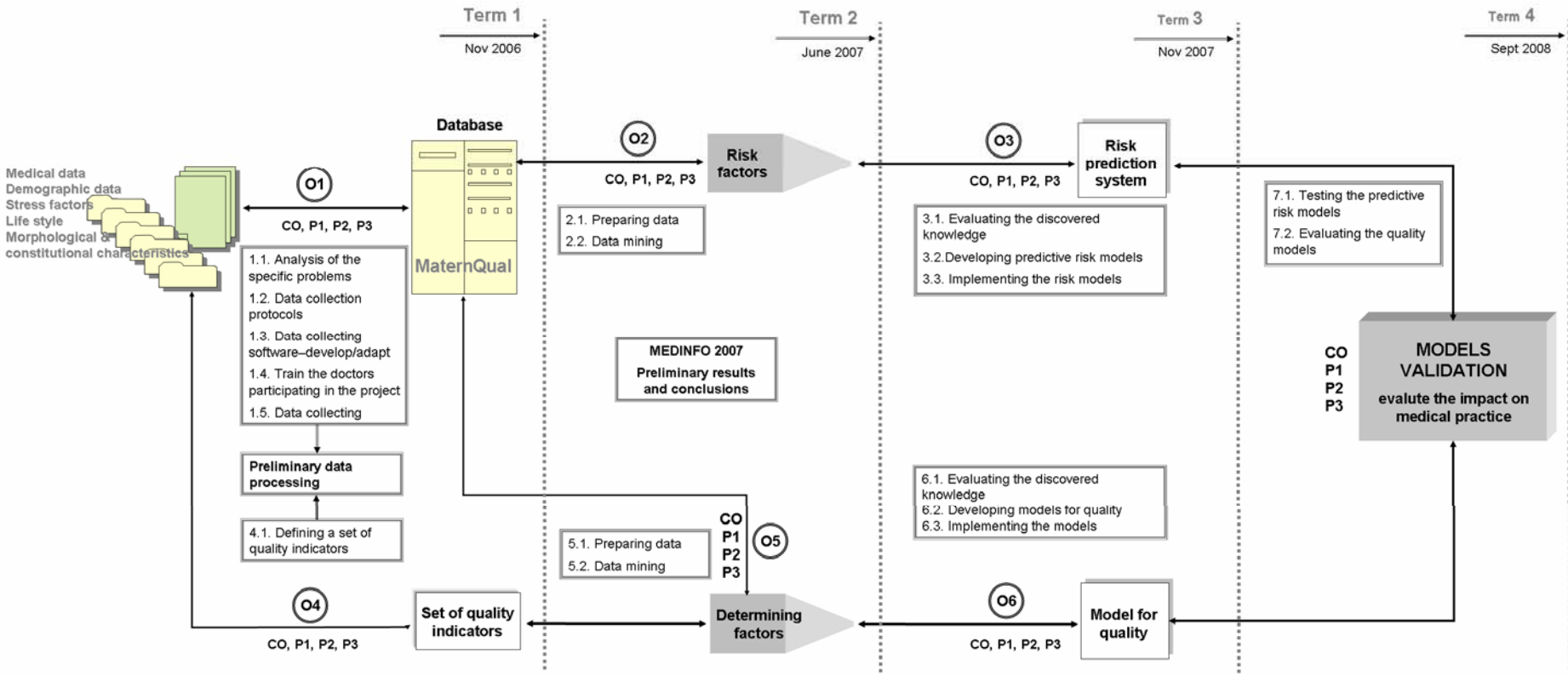
Medinfo 2007
Brisbane

Project MaternQual

- **identifying the factors** associated with
  - obstetrical outcomes
  - pre-term risk factors
- **developing**
  - suitable models
  - prediction systems

for the two related aspects in obstetrical care

2006 - 2008


Term 1
Nov 2006
Term 2
June 2007
Term 3
Nov 2007
Term 4
Sept 2008
Medical data
Demographic data
Stress factors
Life style
Morphological & constitutional characteristics
O1
CO, P1, P2, P3
Database
MaternQual
1.1. Analysis of the specific problems
1.2. Data collection protocols
1.3. Data collecting software–develop/adapt
1.4. Train the doctors participating in the project
1.5. Data collecting
Preliminary data processing
4.1. Defining a set of quality indicators
O4
CO, P1, P2, P3
Set of quality indicators
O2
CO, P1, P2, P3
Risk factors
2.1. Preparing data
2.2. Data mining
MEDINFO 2007
Preliminary results and conclusions
5.1. Preparing data
5.2. Data mining
CO
P1
P2
P3
O5
Determining factors
O3
CO, P1, P2, P3
Risk prediction system
3.1. Evaluating the discovered knowledge
3.2.Developing predictive risk models
3.3. Implementing the risk models
6.1. Evaluating the discovered knowledge
6.2. Developing models for quality
6.3. Implementing the models
O6
CO, P1, P2, P3
Model for quality
7.1. Testing the predictive risk models
7.2. Evaluating the quality models
CO
P1
P2
P3
MODELS VALIDATION
evalute the impact on medical practice

Specific research questions at this stage

- Are there any patterns apparent in obstetrical DRG code utilisation?
- Could we identify the main factors which determine the obstetrical care outcomes?
- Could we identify the factors associated with the pre-term spontaneous delivery?

Statistical analysis of reported outcome data

- two university hospitals in distant geographic locations of each other
- DRG reported data for each hospital
- EHR data set for Hospital T – more data (demographic, blood chemistry, personal obstetrical history, pregnancy evolution)
- ICD-10 codes reported as primary and secondary diagnostic codes

Comparison between university hospitals B and T

Year 2006	Hospital B DRG report	Hospital T DRG report	Hospital T EHR data set
No of new born babies	413	2864	211
No of pre-term babies	47 (11.4%)		14 (6.6%)
No of pregnant women (not necessarily distinct cases)	587	4560	211
ICD-10 O80 and O83 vaginal delivery	250 (42.5% of total ob-gyn cases)	2232 (49% of total ob-gyn cases)	157 (74% of deliveries)
ICD-10 - O82.0 elective caesarian delivery	4 (0.7%)	6 (0.1%)	54 (26%)
ICD-10 - O82.1 emergency caesarian section	74 (12.6%)	903 (19.8%)	0
ICD-10 - P03.4 newborn affected by caesarian delivery	40 (10.9% of full-term babies)	263 (9.3%)	22 (11.2% of full-term babies)

Newborn babies can be seriously affected by a caesarian delivery

Discrepancy between the DRG reported data and the internal EHR data set

Overall use of ICD-10 obstetrical codes

(mothers & newborns, primary and secondary codes)

- Hospital T – EHR data set

  (211 records after pre-processing)

{{Z391, 269}, {Z392, 265}, {Z390, 255}, {Z370, 254}, { O731, 88}, **{O821, 70}**, {O347, 68}, {Z380, 50}, {P599, 39}, {P025, 26}, {P034, 22}, {O990, 20}, {P128, 19}, {O722, 17}, {O342, 17}, {O713, 16}, {O730, 13}, {O60, 11}, {O700, 10}, {O664, 8}, {P081, 7}, {O321, 7}, {P211, 6}, {P073, 6}, {P030, 6}, {D180, 6}, { P910, 5}……}

Reported O82.1 - emergency caesarian section – 70/211 – 33%

Recorded "elective caesarian delivery" in EHR – 54/211 – 26%

- Hospital B

  (1000 records)

{{Z390, 402}, {Z370, 398}, {Z391, 396}, {Z392, 395}, {P599, 362}, {Z380, 361}, {O731, 361}, {Z300, 357}, {Z480, 350}, {O655, 348}, {O628, 337}, {Z298, 284}, {Z703, 269}, {O838, 258}, {P831, 134}, {O366, 109}, {O335, 107}, {O691, 105}, {O990, 104}, {Z540, 99}, {P121, 91}, {O358, 91}, {O420, 89}, {Z321, 82}, {O260, 79}, {O438, 78}, {P154, 77}, {O998, 77}, **{O821, 74}**, {O224, 74}, {P034, 73}, {O664, 73}, {O60, 73}, {O200, 62}, {O361, 57}, {O800, 55}, {O479, 55}, {O713, 46}, {O365, 46}, {O235, 46}, {P550, 45}, {P073, 45}, {P025, 41}, …..}

Reported O82.1 - emergency caesarian section – 74/100 – 7.4%

Risk factors associated with pre-term birth

Code	Attribute	Code	Attribute
A1	Maternal age	A12	Weight gain during pregnancy
A2	Body Mass Index	A13	*Fundus uterus* - height
A3	Smoking	A14	Gestational age
A4	Parity	A15	Type of birth – vaginal/CSection
A5	No. Pregnancies	A16	Child sex
A6	Hemoglobin level	A17	Child head perimeter
A7	Low/high red cell count	A18	Child weight
A8	Glucose level	A19	Child length
A9	Systolic BP	A20	Apgar score
A10	Diastolic BP	A21	Live/still birth
A11	Abdominal perimeter	A22	*Full-term / Pre-term (class label)*

Comparative analysis

M1: a filter type ranking of attributes based on the informational gain (IG) [9];

M2: a filter type ranking of attributes based on the symmetrical uncertainty (SU) [9];

M3: a ranking based on the coefficients obtained by logistic regression (LR); in this case, the higher was the absolute value of a regression coefficient, the higher was the corresponding attribute's relevance;

M4: a ranking method based on the evolutionary multi-objective approach (EMOA) proposed in [10]

Data processing: Weka 3.4.8; DataFit 7.1.44

Classification results

Rank	M1 (IG)	M2 (US)	M3 (LR)	M4 (EMOA)	
				Attribute	Average rank (stdev)
1	A13	A13	A13	A13	4.1 (2.7)
2	A16	A16	A12	A12	5.8 (2.9)
3	A3	A3	A10	A9	6.4 (3.6)
4	A6	A6	A4	A11	6.4 (4.0)
5	A7	A7	A2	A4	6.6 (4.4)
6	A5	A5	A6	A10	6.9 (3.9)
7	A4	A4	A7	A5	7.5 (4.1)
8	A2	A2	A9	A2	7.97 (4)
9	A11	A11	A16	A7	7.98 (3.1)
10	A1	A1	A8	A6	8.1 (4)
11	A12	A12	A3	A3	8.4 (3.5)
12	A8	A8	A11	A1	8.6 (4.1)
13	A10	A10	A1	A8	9.5 (3.8)
14	A9	A9	A5	A16	10.2 (2.7)

Ranks of selected attributes	Correct classification ratio (%)		
	M1(IG), M2(US)	M3 (LR)	M4 (EMOA)
1-14	77.96	77.96	77.96
1-13	78.53	78.53	79.09
1-12	76.27	76.83	77.40
1-11	76.27	77.4	78.53
1-10	78.53	79.09	80.79
1-9	78.53	78.53	80.79
1-8	79.09	80.79	80.22
1-7	79.66	80.22	80.79
1-6	77.96	79.66	81.92
1-5	76.83	81.35	81.92
1-4	78.53	*82.48*	*83.05*
1-3	76.83	81.35	82.48
1-2	*80.79*	81.92	81.92
1	79.09	79.09	79.09

Conclusions

- major problem – sparse data
- we should reconsider the strategy for data collection

References

1. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press, 2001.
2. Marshall M, Leatherman S, Mattke S. OECD working paper no. 16: Selecting indicators for the quality of health promotion, prevention and primary care at the health systems level in OECD countries. 2004: *http://www.oecd.org/dataoecd/27/52/33865865.pdf* (Last access 28 May 2007).
3. Cios KJ, Moore GW. Uniqueness of medical data mining. Artificial Intelligence in Medicine 2002; 26: 1-24.
4. Spangler WE, May JH, Strum DP, Vargas LG. A data mining approach to characterizing medical code usage pattern. Journal of Medical Systems 2002; 26(3): 255-275.
5. Neumann A, Holstein J, Le Gall JR, Lepage E. Measuring performance in health care: case-mix adjustment by boosted decision trees. Artificial Intelligence in Medicine 2004; 32: 97-113.
6. Goodwin L, Maher S. Data mining for preterm birth prediction. Proceedings of the 2000 ACM Symposium on Applied Computing (SAC'00): 46-50.
7. Lovell DR, Rosario B, Niranjan M, Prager RW, Dalton KJ, et al. Design, construction and evaluation of systems to predict risk in obstetrics. International Journal of Medical Informatics 1997; 46: 159-173.
8. Resnik R. Issues in the management of preterm labor. J. Obstet. Gynaecol. Res. 2005; 31(5): 354-358.
9. Witten IH, Frank E. Data mining. Practical machine learning tools. Berlin: Elsevier, 2005.
10. Zaharie D, Holban S, Lungeanu D, Navolan D. A computational intelligence approach for ranking risk factors in preterm birth. 4th International Symposium on Applied Computational Intelligence and Informatics SACI 2007; in press.

Acknowledgements

This work is supported by grant 99-II CEEX 03 - INFOSOC 4091/31.07.2006 from the Romanian Ministry of Education and Research.

We kindly thank **Dr. Maria Bari** from the University of Medicine and Pharmacy of Bucharest for her valuable contribution to this project.

Address for correspondence

Diana Lungeanu

Department of Medical Informatics, University of Medicine and Pharmacy

P-ta E. Murgu 2, 300041 Timisoara, Romania. Tel/fax +40-256-490288;

dlungeanu@gmail.com