

A young woman with brown hair tied back, wearing a white lab coat, is smiling broadly. She is in a clinical or hospital setting, with medical equipment visible in the background. The image is partially overlaid by a blue semi-transparent box on the left and a red and white graphic on the right.

Rare Diseases

Geert Smits
Universitair Ziekenhuis Antwerpen

Health Information Technology
October 13th, 2010

Rare diseases

Kennis / Ervaring / Zorg

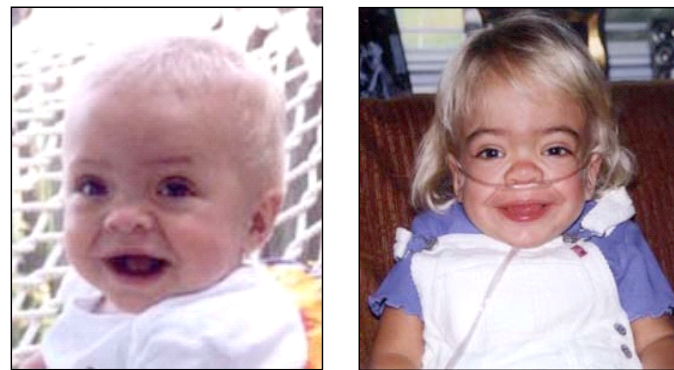


Impact of rare diseases

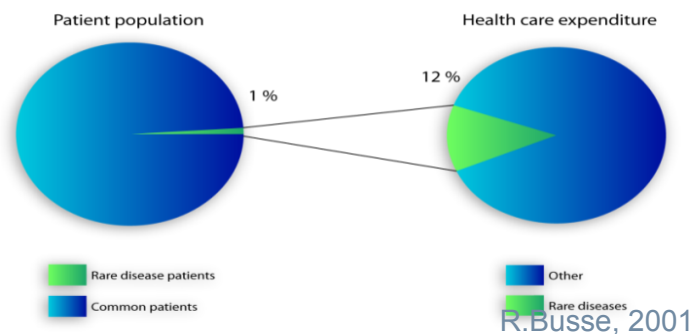
29
million
affected people
in EU – 27 member states

Eurordis2

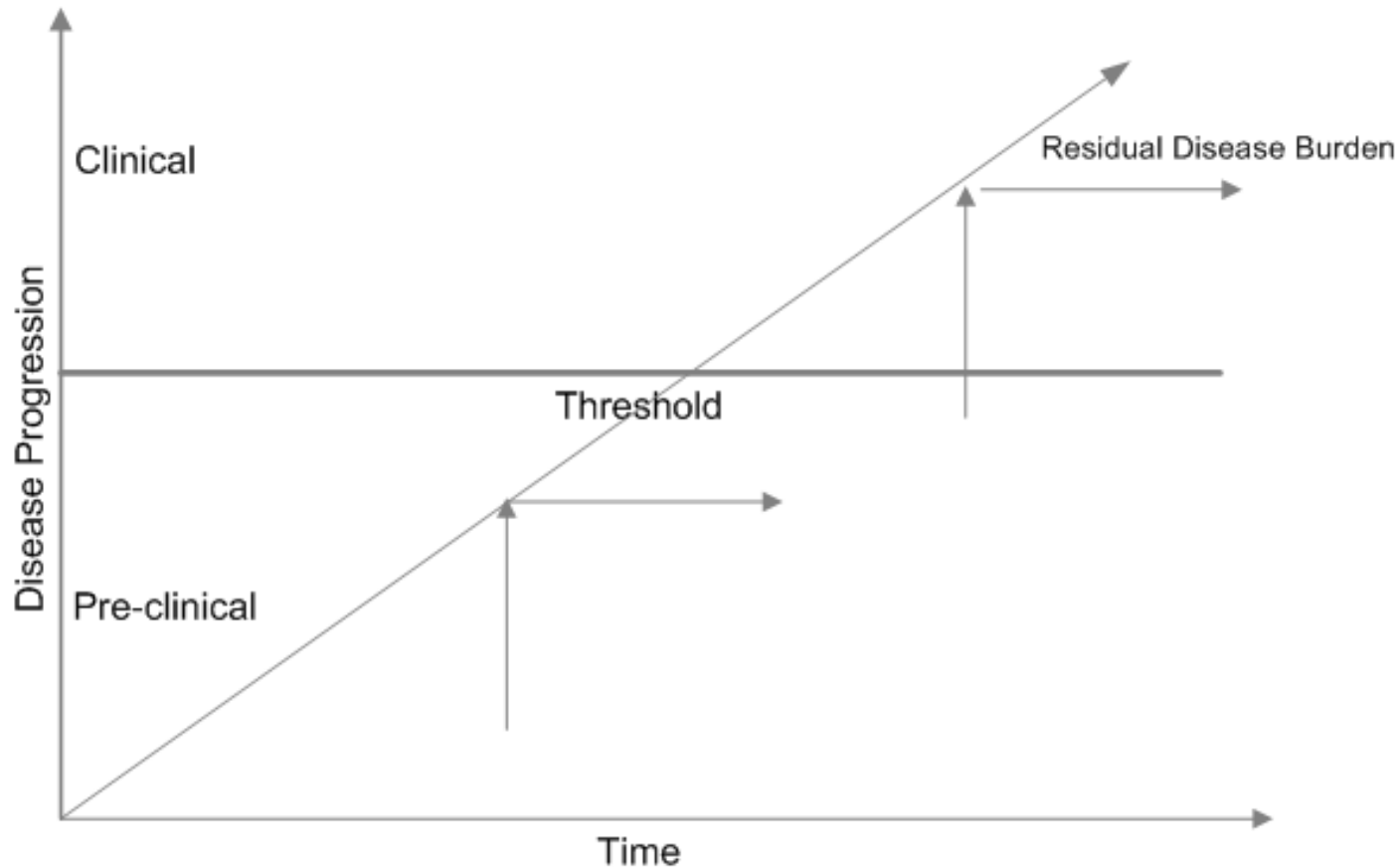
Huge ethical impact



Big economic impact



The need for faster and more accurate diagnosis....



Disease Progression: Severe MPS I



10 months



12 months



22 months

Kennis / Ervaring / Zorg



34 months



39 months

Data mining in medicine



Kennis / Ervaring / Zorg

UZA

Data mining in medicine

RESEARCH ARTICLE

Open Access

Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes

Wei Yu^{*}, Tiebin Liu, Rodolfo Valdez, Marta Gwinn, Muin J Khoury

Research article

Open Access

Using data mining techniques to explore physicians' therapeutic decisions when clinical guidelines do not provide recommendations: methods and example for type 2 diabetes

Massoud Toussi^{*1}, Jean-Baptiste Lamy¹, Philippe Le Toumelin² and Alain Venot^{1,2}

Address: ¹Laboratoire d'Informatique Médicale et Bioinformatique (LIM&BIO EA 3969), UFR SM&H, Université Paris 13, Bobigny, France and ²Département Inter-hospitalier de Santé publique, Hôpital Avicenne, Assistance publique-Hôpitaux de Paris, Université Paris 13, Bobigny, France

Email: Massoud Toussi^{*} - massoudtoussi@gmail.com; Jean-Baptiste Lamy - jblamy@free.fr; Philippe Le Toumelin - philippe.letoumelin@ap-hop-paris.fr; Alain Venot - alain.venot@univ-paris13.fr

^{*} Corresponding author

Data mining and clinical data repositories: Insights from a 667,000 patient data set

Irene M. Mullins^a, Mir S. Siadaty^a, Jason Lyman^a, Ken Scully^a, Carleton T. Garrett^b, W. Greg Miller^b, Rudy Muller^b, Barry Robson^c, Chid Apte^c, Sholom Weiss^c, Isidore Rigoutsos^c, Daniel Platt^c, Simona Cohen^d, William A. Knaus^{a,*1}

^aDepartment of Public Health Sciences, University of Virginia Health System, Charlottesville, VA, USA

^bDepartment of Pathology, Virginia Commonwealth University, Richmond, VA, USA

^cIBM T.J. Watson Research Center, IBM Life Sciences, Yorktown Heights, New York, USA

^dIBM Research, Haifa, Israel

Received 28 January 2005; accepted 22 August 2005

A data mining framework for time series estimation

Xiao Hu^{*}, Peng Xu, Shaozhi Wu, Shadnaz Asgari, Marvin Bergsneider

Neural Systems and Dynamics Lab, Department of Neurosurgery, Geffen School of Medicine at University of California, Los Angeles, CA 90095, USA

RESEARCH ARTICLE

Open Access

Regression tree construction by bootstrap: Model search for DRG-systems applied to Austrian health-data

Thomas Grubinger^{*}, Conrad Kobel and Karl-Peter Pfeiffer

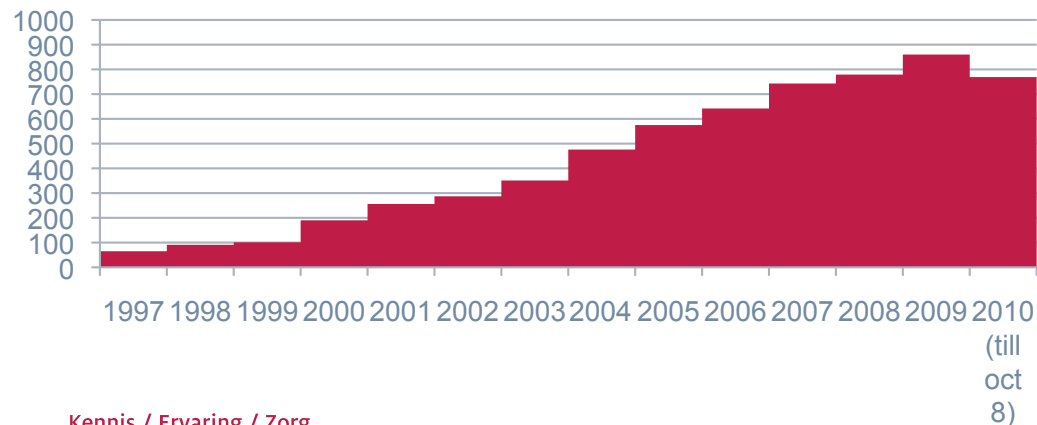
RESEARCH ARTICLE

Open Access

A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay

Andrew A Kramer^{*1} and Jack E Zimmerman²

Data Mining Publications in PubMed



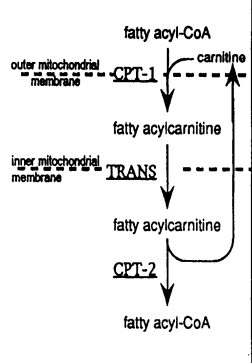
Kennis / Ervaring / Zorg

UZA'

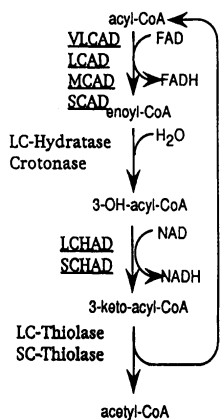
- Medium chain acyl CoA dehydrogenase (MCAD) deficiency
- Autosomal recessive disorder of beta-oxidation of fatty acids
- Causes hepatomegaly, cardiomyopathy, ...
- Can be controlled by diet restrictions if diagnosed early

MCADD Disease

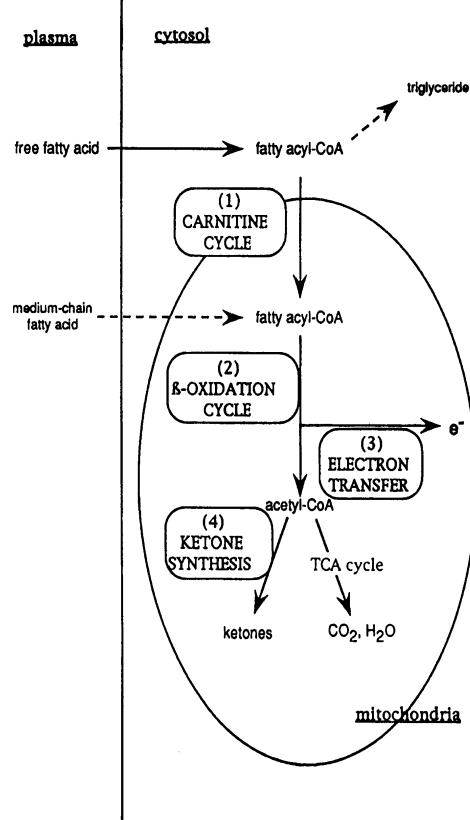
(1) CARNITINE CYCLE



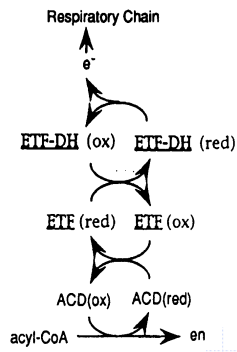
(2) β-OXIDATION CYCLE



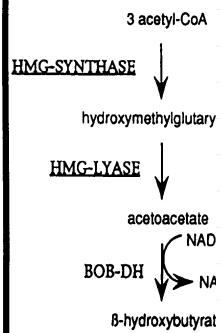
FATTY ACID OXIDATION PATHWAY



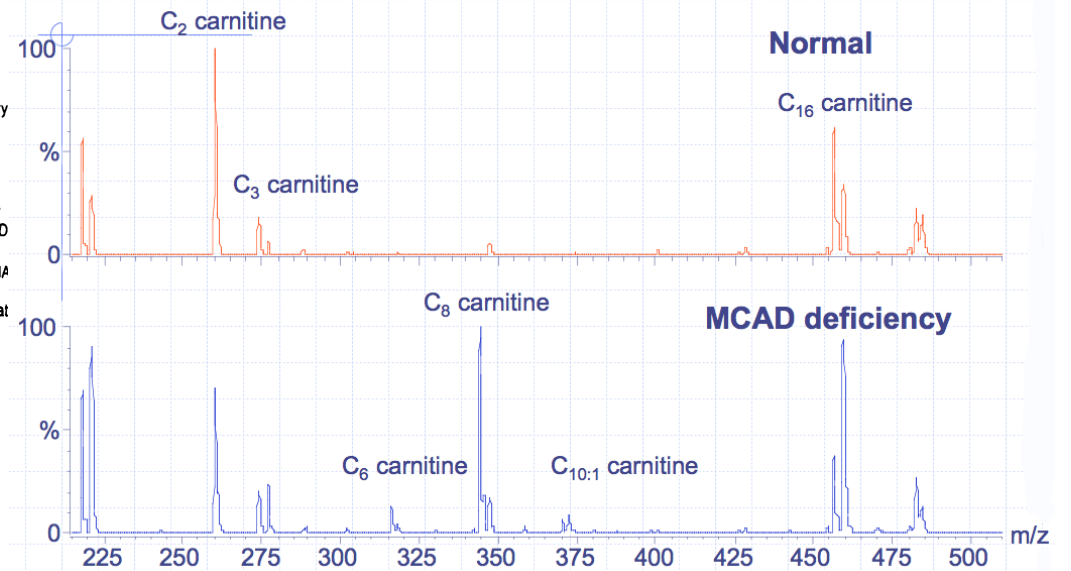
(3) ELECTRON TRANSFER



(4) KETONE SYNTHESIS



Normal vs MCAD deficiency



We started from an anonymised dataset, listing about 32.000 blood sample analyses for the acylcarnitines such as C6, C8, C10, C10:1 in blood which were obtained through a systematic screening of newborns during the 1st semester 2009 in the province of Antwerp. This dataset was further enriched with blood samples of MCAD deficient patients. In total, 9 MCAD cases were present.

- Source: PCMA – Provinciaal Centrum voor Metabole Aandoeningen

Methods

- Data set from the 'provinciaal centrum voor metabole aandoeningen'
- Clinical expert practise
- Data mining methods
 - Decision trees
 - Logistic regression
 - Naive Bayes
- Compared the different methods
- Determined the important parameters

Methodology:

- 9-fold (stratified) cross-validation (10 randomizations each)
- Optimization of the included variables
- Optimization of the model parameters
- Optimization of the classification threshold

Results


model	Included model parameters	TN	FN	FP	TP
Decision trees (C4.5)	C6, C8, C10, C6/C8, C6/C10, C8/C10, C8/C2	32085,1	0,1	8,9	8,9
Logistic regression	C8	32082,7	0	11,3	9
Logistic regression	C8, C8/C2, C6/C8, C8/C10, C6/C10	32087	0	7	9
Logistic ridge regression	C8, C8/C2, C6/C8, C8/C10, C6/C10	32089,5	0	4,5	9

Table 1: selection of the most predictive models

- **Decision trees** have a very good classification accuracy but occasionally fail to identify a MCAD case in the different randomizations.
- **Logistic regression**
 - C8 only is an excellent MCAD predictor (FP: 11,3)
 - Number of false positives is further reduced by including more variables (11,3 → 7)
 - Ridge regression reduces the number of false positives even further by performing variable selection on this subset (7 → 4,5)
- **Naïve Bayes** resulted overall in poor classification (~4600 false positives)

Conclusion

- Data mining techniques can be used to automate the support for early diagnosis of MCAD.
- C8 is the most important parameter in the calculated models ...
- But based on (only) 9 MCADD cases. More complex models could be identified as more data is available.
- The models are straightforward and intuitive and can be used to support the diagnosis (medical decision support).

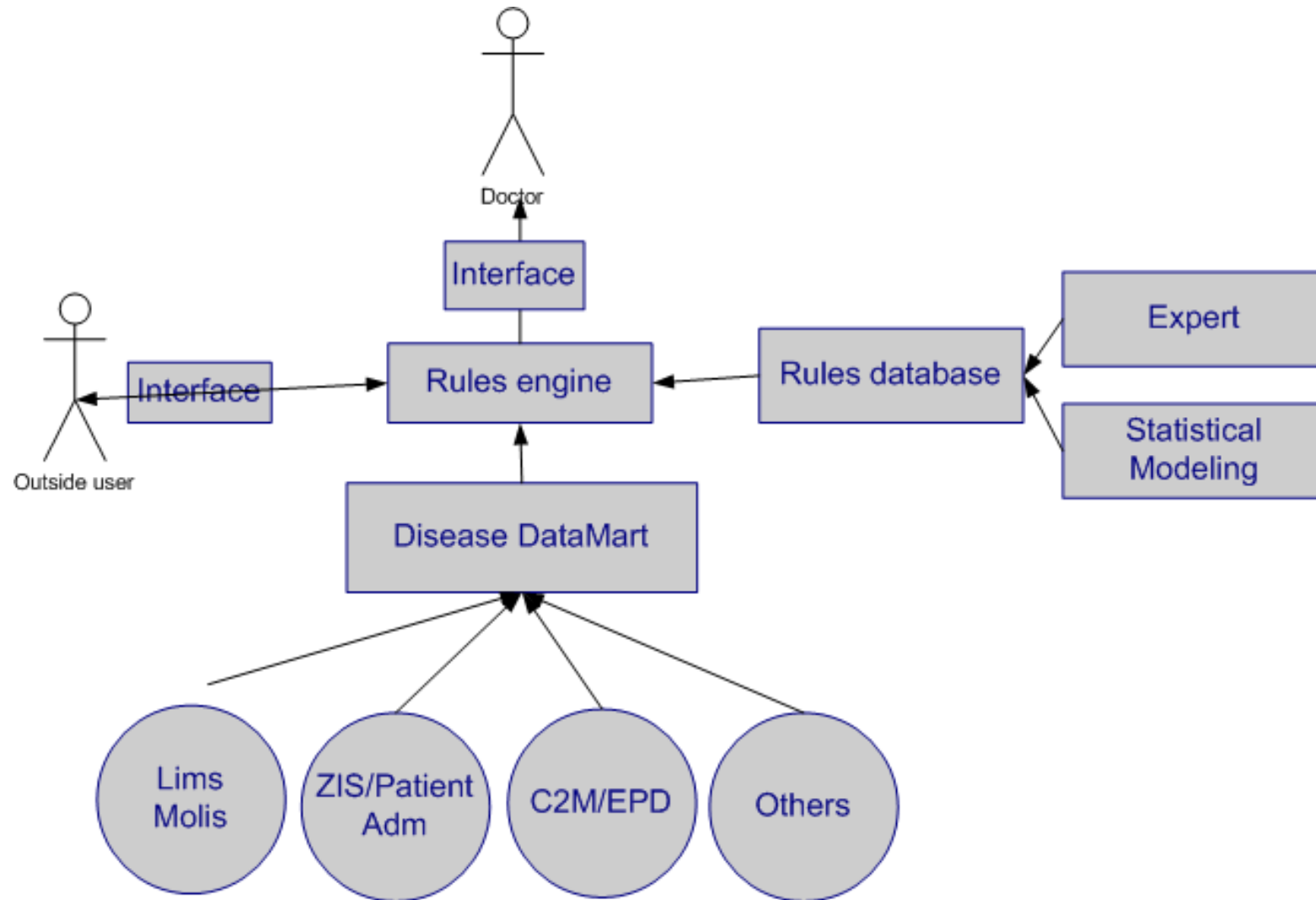


Improved diagnostics in
hospital settings en
beyond...

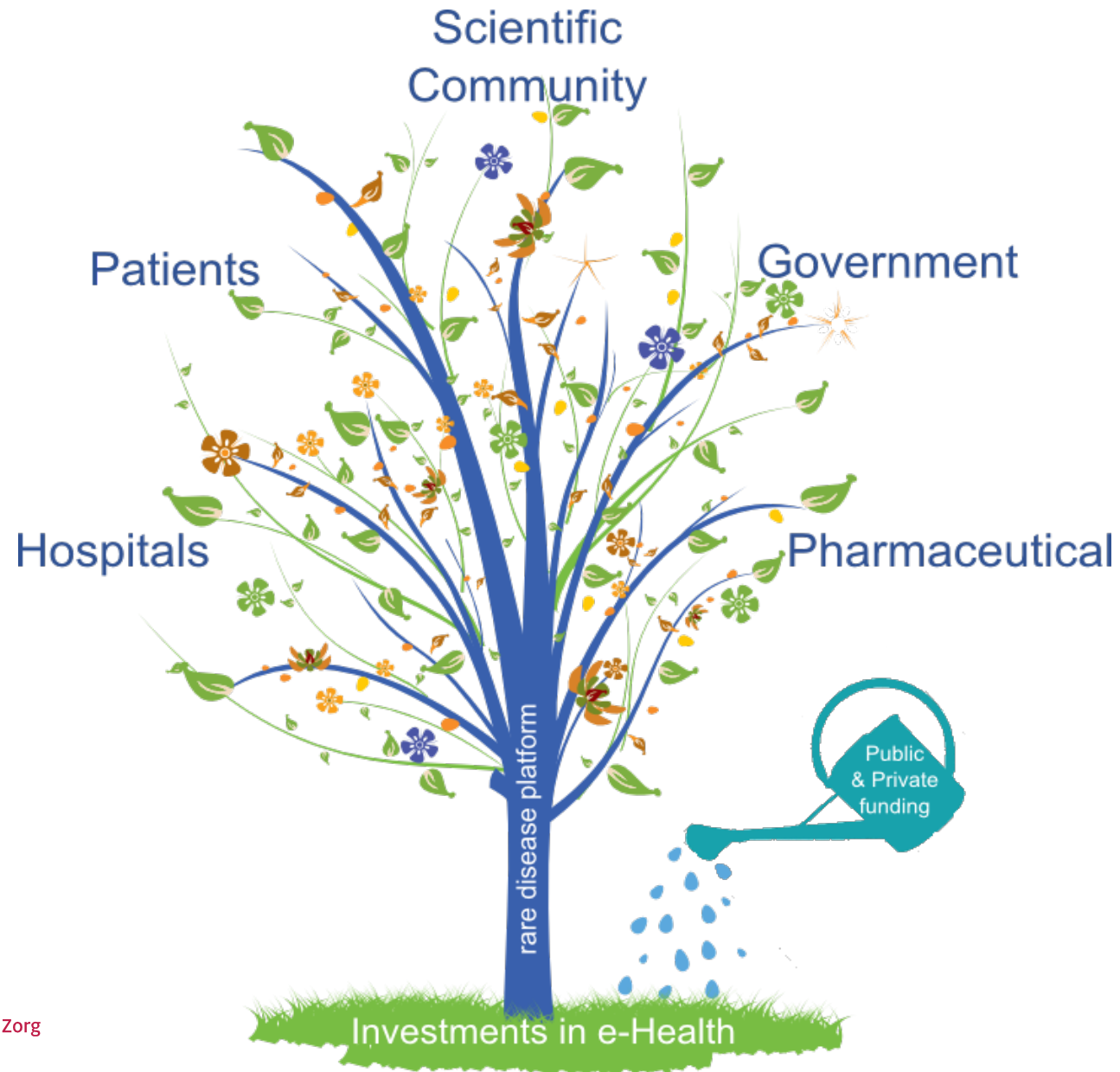
Kennis / Ervaring / Zorg

UZA'

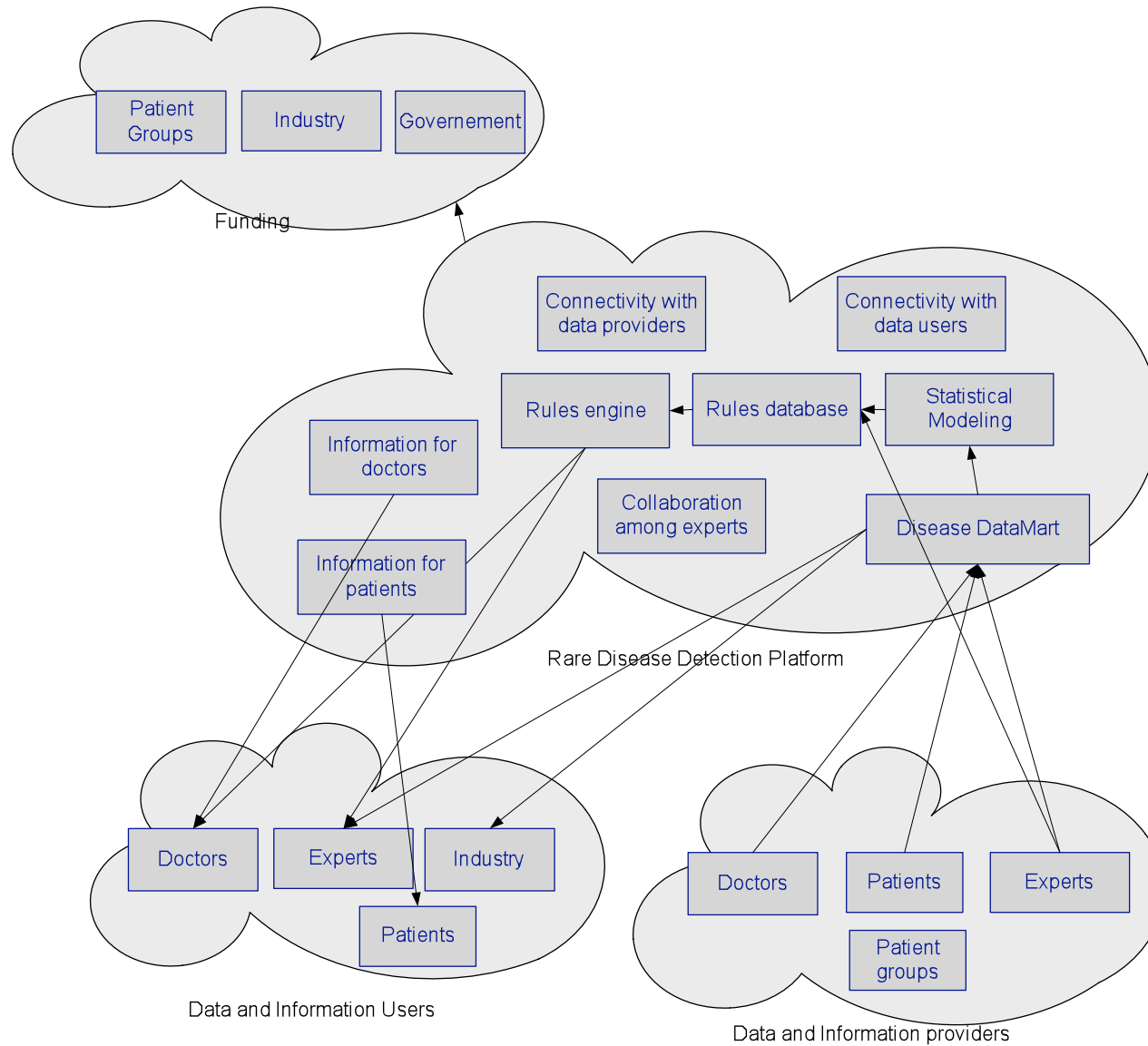
Within the hospital...



Stakeholders...



In the world...



Conclusion

Kennis / Ervaring / Zorg

