

A. Abu-Hanna<sup>1</sup>, P. J. F. Lucas<sup>2</sup>

<sup>1</sup>Dept. of Medical Informatics,  
Academic Medical Center-University  
of Amsterdam, The Netherlands,  
<sup>2</sup>Dept. of Computing Science,  
University of Aberdeen, Scotland, UK

# Editorial

## Prognostic Models in Medicine

### AI and Statistical Approaches

#### 1. Introduction

Prognosis (pro: before; gnoscerere: to know) literary means to know beforehand or, as a noun, foreknowledge. The key concept behind prognosis is the *prediction* of an event *before* its possible occurrence. One could argue that the prediction of anything without *knowing* about its possible occurrence is a form of prognosis. An example of the latter is used in the evaluation of a clinician's performance on predicting the disease outcome of a patient without knowledge of the actual outcome. Either way, the essence of prognosis is that the predicted event occurs in the *future* relative to the information available at the time of prediction. Time is thus inherent to the concept of prognosis and distinguishes it, for example, from that of diagnosis, where the future plays a less important role. Medical prognosis is defined here as:

*the prediction of the future course and outcome of disease processes, which may either concern their natural course or their outcome after treatment.*

The course and outcome of a disease are usually expressed in terms of the patient, rather than in terms of the disease or treatment. Examples of prognostic information are prediction statements about survival chance and time ("the patient has 75% chance to survive for 5 years after being diagnosed with the disease"), predictions about events such as complications of a chosen therapy

and treatment response ("the patient will probably respond positively to the therapy"), and of disease evolution ("the patient is expected to progress from a high risk group to a lower risk group within three days of intensive care").

Prognosis is interwoven with the other tasks of clinical management of the patient, namely those of diagnosis, and therapy selection and planning. Diagnostic information and information about medical actions, mainly therapy, strongly influence prognosis [14]. This is apparent from the action or therapy-based association that practicing clinicians have with the notion of prognosis: the clinician views prognosis as a *decision tool* that is conditioned on diagnostic and treatment information [28]. When viewed in this way, the value of diagnostic information that does not contribute to prognosis may be questioned; the same is true for the value of prognostic information that does not contribute to the choice of therapy. This does not, however, exclude the fact that knowing for the sake of knowing can also be valuable in itself [5]. In any case, from the increasing awareness in clinical practice of evidence-based medicine and the central role that prognosis plays in the management of patients, it is clear that interest in prognostic models will only increase in the future. Prognostic models are already useful aides to clinical management [20]. It therefore seems to be the proper time to pay attention to the subject of prognosis in medical informatics. The development of prognostic models for the delivery

of improved medical care is the theme of the papers included in this special issue of *Methods of Information in Medicine*.

The special issue has culminated from a series of successful events starting with the invited session on "Intelligent Prognostic Methods in Medical Diagnosis and Treatment Planning" in 1998 [1] during the conference "Computational Engineering in Systems Applications 1998" (CESA '98) in Hammamet, Tunisia. This invited session has resulted in a special issue on prognosis in medicine of the journal *Artificial Intelligence in Medicine* in February 1999 [17]. Subsequently, a new initiative has resulted in the workshop Prognostic Models in Medicine during the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM '99) in Aalborg, Denmark [2]. The current special issue includes a selection of the best papers of the workshop held in Aalborg, supplemented with selected papers directly submitted for review for this special issue. The remainder of this editorial is meant to place these papers in the perspective of on-going research on prognostic models.

#### 2. Approaches to Prognostic Models

There are many different ways in which prognostic models can be used and developed. Prognostic models can be constructed by choosing from a

plethora of different techniques. In this section, we will shed some light on issues related to their use, construction, and evaluation.

### 2.1 Uses of Prognostic Models

Prognostic models have many uses, such as: (a) guiding health-care policy by generating global predictive scenarios, (b) assisting comparative audit among hospitals by (case-mix adjusted) mortality predictions, (c) determining study eligibility of patients for new treatments, (d) defining inclusion criteria for clinical trials to control for variation in prognosis, (e) cost reimbursement programs, and (f) selecting appropriate tests and therapies in individual patient management including supporting decisions on withholding or withdrawing therapy.

As decision-support tools, it is useful to distinguish between two kinds of prognostic models: models at the patient *population* level, and at the *individual* patient level. These two levels imply different requirements on the prognostic model in terms of the methods used for building and evaluating the model.

Generally speaking, the population versus individual level aspect roughly corresponds to two levels of Shortliffe's typology of decision-support functions [24]. Shortliffe distinguishes between the following three types of decision support functions: information management, focusing attention, and patient-specific consultation. The population level maps roughly onto the second function type and the individual level onto the third type. For example, a prognostic model at the population level used in a quality assurance program can lead to the identification of a discrepancy between the expected and actual rates of complications resulting from surgery for a group of patients, and thus help to identify the reasons for this discrepancy. On the other hand, a prognostic model meant for use as a basis for providing treatment advice [9, 7] for a specific patient is of the patient-specific consultation function type. The concept of "group" is, of course, coarse as it covers situations where the whole patient population is taken as one group to situations where

patients are stratified in many different sub-groups, e.g., risk groups.

### 2.2 Model Building: Statistics, Machine learning and AI

Various methods have been suggested for the representation of prognostic models ranging from quantitative and probabilistic approaches to symbolic and qualitative ones. In some situations these models are integrated in the medical management of patients (diagnosis, treatment, follow-up), where assessment of the expected prognosis constitutes an integral part.

A huge body of experience in developing prognostic models exists in the field of medical statistics. Commonly used techniques are simple decision rules based on the categorization of a prognostic score, Bayes' rule, and logistic regression. The parameters of these models are almost without exception assessed from data especially collected for that purpose.

A traditional prognostic instrument in medicine is the use of simple decision rules based on a *prognostic score* to classify a patient into a future risk category [19]. A prognostic scoring system provides a means for the quantification of the severity of disease or health status of a patient, where usually a higher score corresponds to greater severity. Different elements contribute to this total score, including, for example, physiological variables such as heart rate, demographic variables such as age and sex, history, laboratory findings, and so on, where each element contributes a number of "penalty points". An example of a prognostic scoring system is the Pre-Arrest Morbidity or PAM score [11] that is used to predict the outcome of in-hospital cardiopulmonary resuscitation based on comparing the total score with some threshold. This kind of *representation*, i.e. rules, is quite popular in artificial intelligence (AI), although they would usually be logical rather than numerical in nature.

In a prognostic context, Bayes' rule and logistic regression are mostly used to estimate the probability of a dichotomous outcome of interest, such as mortality. A prognostic score can be used as an independent variable in the analysis. For example, to estimate the probabil-

ity of in-hospital death of intensive care patients, the SAPS-II logistic regression model takes the SAPS-II score as a covariate [15]. One could also use the output itself of Bayes' rule or the logistic regression formula in decision rules as above, where the probability of the outcome in the conditions may be viewed as a score.

Survival analysis comprises another popular collection of statistical methods used when data is time-dependent. Special techniques have been developed in medical statistics for survival analysis. Here one is usually interested in the prediction of the time that would elapse before an event will occur. This has obvious appeal for predicting survival times of patients, but in spite of its name, survival analysis is applied to a wide range of time prediction problems, such as expected time to relapse after treatment or even expected time to appearance of a symptom after being exposed to particular conditions. An important aspect in survival analysis is the use made of *censored* data. A censored subject or case is one in which the event of interest has not been observed for the period of time that the subject has been followed. Cox's regression is capable of dealing with censored data; it is the technique most frequently used in developing this type of prognostic model.

Prognostic models are often used in the context of decision analysis: by augmenting a prognostic model with utility assessments of potential outcomes, and indicating particular variables as decision variables, optimal decisions for groups of patients or individual patients can be determined [21, 25]. The actual result that is conveyed to the clinician is often formulated in terms of actions to be taken in a particular clinical situation, without explicitly referring to the prognostic model used [26].

Whereas medical statistics has been somewhat restrictive in its use of techniques for the development of prognostic models, the field of machine learning offers a very wide range of representations and techniques. We have provided a brief description of the main machine learning approaches in medical prognosis in [17]. As machine learning is actually a subfield of artificial intelligence, a number of issues typical for AI are

naturally borrowed, such as: knowledge representation, automated reasoning, problem solving and search, and knowledge engineering. Symbolic representations, for example logic representations, are used in logical classification rules; graph representations are used in decision trees [4], Bayesian networks [27] and neural networks [6]. In general *declarative* representations, which are representations that can be used in different ways, are favored in AI and machine learning.

The differences in views among statisticians, AI and machine-learning researchers, even when dealing with the same formalism, can be illustrated by the formalism of Bayesian networks. Basically, a Bayesian network is just a multivariate probability distribution with independence assumptions, which is a view with which statisticians would feel comfortable. However, a Bayesian network can also be viewed as a knowledge-representation formalism, with an explicit structure, an associated semantics and reasoning methods. The strong relationship between qualitative causal models and the semantics of Bayesian networks makes it possible to construct such models with the assistance of domain experts, in a way similar to knowledge engineering. It is well known that the structure of the graphical part of a Bayesian network facilitates the assessment of probabilities, even to the extent that reliable probabilistic information can be obtained from experts [16]. Hence, Bayesian networks do allow AI-like interpretations. Finally, both the structure of a Bayesian network, albeit with difficulty, and its underlying probability distribution can be learnt from data. Hence, there is also a strong relationship to machine learning.

Bayesian networks are unique in the sense that they share the characteristics of the fields of AI, machine learning and statistics. Other techniques do not usually provide such flexibility in their interpretation, due to their more restricted nature. When model construction is hard, less explicit domain models may be used, such as case-based and neural network representations. However, it might still be possible to incorporate domain knowledge in such restricted models. The integration of domain knowledge elicited from the

experts into the learning approach usually results in better performing models.

Construction of a prognostic model is conceptually a *search process*: the algorithms used for their construction are searching a model space for the model that is most appropriate or fits best in some predefined sense. The more expressive the model is, the more complex the model space. This explains, on the one hand interest in straightforward models like the independent form of Bayes' rule [18] or in logistic regression, as their construction will require limited search, and, on the other hand, in methods for searching the model space efficiently. Modern techniques, like genetic algorithms, offer a method inspired by the evolution theory to perform parallel search in the model space. This approach can mitigate the problem of finding a model which is only locally optimal.

### 2.3 Evaluation of Prognostic Models

The evaluation of prognostic models has become an increasingly important topic in the last few years. The commentary in [29] discusses the clinical usefulness of prognostic models. Before one uses a suggested model it is important to have an indication that the model would work well on a population different from that which was used to develop the model. There are many studies reporting on model validation, certainly the lack thereof, e.g. [10]. One may distinguish between two types of evaluation: laboratory evaluation and clinical evaluation [3, 30]. A *laboratory evaluation* usually focuses on the performance of the model, expressed in statistical terms. Relevant questions in laboratory evaluation are whether the model passes the appropriate statistical tests, usually on a new data set, and whether it is the best model given the available factors. In [12] a good overview is given about different aspects relevant to statistical model performance: inaccuracy, imprecision, inseparability, and resemblance. In a *clinical evaluation* one is interested in the question whether the model is satisfactory for its *clinical purpose*, and whether the model's performance is good enough to assist in dealing with a clinical prediction problem. For example, to see

whether a model is sensitive to the quality of, e.g., intensive care one could simulate different care qualities and see whether the model can detect relevant differences [31]. We have already commented on the difference between models that operate at the individual level and models that operate at the population level. This is important when considering the purpose of the model. For example a model's purpose could be the reliable probabilistic prediction of death over a wide range of risk groups, rather than the classification of vital status of each individual patient, and this should be considered in the clinical evaluation. Note that it is possible to have a statistically but yet not clinically valid model and vice versa.

Evaluation of models should be handled with care. In the machine-learning community new models and methods are constantly developed and compared with other, usually more established, models. The question often arises whether one is using the right statistical tests e.g., to show that one model is better than another. There always is a danger that very search-intensive algorithms come up with patterns that are not really there. Hence, the tests should be scrutinized and adapted accordingly. For a description of phenomena that might invalidate the experimental comparisons between classifiers see [22]. In [8] nine statistical questions in machine learning have been formulated. The question of how to choose between two learning algorithms when the data set is small is described in detail in this latter paper and observations are made about the validity of independence assumptions. Finally, note that the use of very expressive models such as neural networks could easily lead to overfitting the data. Concerns from medical statisticians about the misuses of neural networks in prognostic models are expressed in [23]. These include, among others, fitting implausible (biological) functions, tuning and testing on the same data set (ideally three data sets are required), and insufficient comparison with statistical methods.

In short, in model evaluation one should not disregard the extensive work done in statistics. On the other hand, as argued in [13] the statistical community

---

should indulge in machine learning and datamining research and not miss the boat this time.

### 3. Papers in this Special Issue

The papers in this special issue of *Methods of Information in Medicine* on prognostic models in medicine cover a wide variety of different approaches to the construction, evaluation and clinical use of such models. We shall briefly summarize the content of these papers.

The paper by Dréau et al. presents an evaluation study of various statistical predictive models of cardiovascular risk taken from the literature. The study investigates the usability of the models and whether the risk estimations fit the medical decisions taken by physicians following an international practice guideline for treating hypertensive patients. The study shows that between-model agreement is better for the lower and upper risk quintiles than the other three quintiles and that the models vary considerably in their classification of patients into the high risk category when a threshold is used on the probability. This means that when considering the high risk category, the same patient will be often classified differently by the different models. When considering the decision to treat taken by the physicians, there was a poor agreement between the risk models and the physicians' decisions. The study emphasizes the importance of validating models before their diffusion and reflects on the factors that might lead to invalidation.

In the paper of Vaillant et al. a non-parametric method for predictor selection and its application to survival analysis in breast cancer is described. The method is based on the information theoretic concept of entropy, in this case the mutual information between a survival variable and its respective co-variables, and can be used without making assumptions about the distributions of the survival and predictor variables. The predictors are iteratively selected based on their ability to reduce the entropy in the survival variable. A decomposition of the entropy of the survival variable shows the parallels with the square sum decomposition as in the tra-

ditional variance analysis. Entropy is also used to aggregate the survival curves by lumping together subsets of the predictor's values without loss of information. The method is applied in survival analysis to breast cancer and yields comparable results to a logistic and Cox's regression models.

The paper by Anand et al. introduces a novel method for handling censored observations in the k-Nearest Neighbors algorithm. The new algorithm called Ck-NN, is applied to the prediction of survival times of patients with colorectal cancer. The algorithm provides two major enhancements to the traditional k-NN algorithm. The first is that the distance between two values of a categorical variable is based on a distance metric between the Kaplan-Meier survival curves that are induced by the two values. In this way censored data are incorporated in a natural way into the distance metrics of the k-NN framework. The second enhancement consists of the use of elements of evidence theory for the incorporation of censored observations of the nearest neighbors in the final prediction of the survival time. The prediction maintains a balance between informativeness and correctness. The Ck-NN algorithm is employed with various distance measures and comparisons are provided with k-NN measures.

The paper by Zupan et al. describes a way to improve the classification performance of a prognostic model by augmenting the given features, i.e. independent variables, with domain knowledge elicited from the expert. The model predicts the patient's long-term clinical status after hip arthroplasty. The given features are organized by a domain expert in a hierarchy where they constitute the leaves. Intermediate features are then introduced as aggregates of the leaves by the expert who also provides a mapping between combinations of the (discrete) values of the given features to values of the aggregate features. It is shown that using the higher level attributes significantly improves the classification accuracy of a naive Bayes classifier. A comparison with the performance of a decision tree classifier is also provided. In both cases the added value of the expert's higher level features is illustrated.

In the paper of Ohno-Machado and Vinterbo, the effects of case removal on the performance of prognostic models are studied and tested on acute trauma, and myocardial infarction data sets. Influential cases are sought which are defined here as those whose removal from the database would result in better performance of the model, or at least would not result in its deterioration. The paper compares various univariate and multivariate approaches to the identification and removal of cases that might affect the model's performance. Two multivariate approaches are considered, one based on a sequential backward selection of cases and the other on a non-sequential genetic algorithm which can handle subsets of cases. It is shown that the genetic algorithm, which is less greedy than the sequential approach in its search for candidate cases, provides the best results. This is because the genetic algorithm can identify cases which are influential *en bloc* which, alone, may not be influential.

The paper by Ramoni, Sebastiani and Dybowski applies a new classification method, called the robust Bayes classifier to mortality prediction in intensive care. This classifier, which is based on the naive Bayes classifier, is robust in the sense that it does not assume any missing data mechanism that had led to the missing data in the database. Instead it reasons about probability intervals. Two modes of the robust Bayes classifier are applied and compared to a logistic regression model which relies on the imputation of the missing data. Higher accuracy of the robust classifier is reported, this is measured on the cases which could be classified. This increase in accuracy comes at the cost of coverage as there are situations in which the classifier is unable to assign a class to a case. Adopting weaker criteria to increase the coverage reduces accuracy to a level comparable to that achieved by the logistic regression model. Cost criteria are provided to facilitate the choice of the appropriate classifier.

The paper by Armengol, Palaudàries and Plaza describes the application of a case base reasoning method called Lazy Induction of Descriptions (LID). LID is applied in a prognostic model to predict long-term risks for individual diabetic

patients. A case in the case base has a structure that includes characteristics of the patient, (qualitative) findings, disease, therapy, and outcome (macro- and microcomplications). The method works by building an increasingly homogeneous set, in terms of the outcome, of cases that match the case at hand. This is achieved by using a heuristic for iteratively selecting the best features that induce partitions that most resemble a correct partition that is based on the actual possible outcome values (risk patterns in this case). A quantitative estimate of the system's performance in terms of correct classification when compared to a gold standard based on an expert's opinion is provided.

The paper by Chan and Naghdy describes a decision-support system that is able to assist anaesthetists in prescribing fluid replacement therapy to surgical patients. The system uses fuzzy control rules to predict the body fluid level of a patient; the output of the system is a suggested infusion rate for the patient concerned. Interesting in this paper is the combined use of a statistical technique – cluster analysis – with fuzzy logic in the development of the rule base of the system. The paper also includes a preliminary evaluation of the system using a small data set (18 patients), suggesting that the system might indeed offer valuable advice.

These papers illustrate a typical sample of the wide spectrum of research in medical prognosis. They illustrate ideas from AI and statistics in the context of solving the medical problem of prognosis. The importance of prognosis in medicine and the arising mutual awareness between the AI and statistical communities are promising conditions for the methodological progress and practical applications of medical prognostic models.

#### Acknowledgments

We are grateful to our colleagues who served on the program committee of the invited session on *Prognostic Models in Medicine, Artificial Intelligence and Decision Analytic Approaches* during the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99) in Aalborg, Denmark, and the colleagues who have reviewed the additional papers submitted to this special issue. These colleagues are: K. P. Adlassing, AU, S. S. Anand, UK, S. Andreassen, DK,

P. M. M. Bossuyt, NL, J. Fox, UK, L. C. van der Gaag, NL, J. D. F. Habbema, NL, P. Haddawy, USA, P. Hammond, UK, E. Keravnou, CY, N. Lavrac, SL, J. van der Lei, NL, L. Ohno-Machado, USA, N. Peek, NL, M. Ramoni, UK, M. Stefanelli, IT, Th. Wetter, DE, J. Wyatt, UK.

#### REFERENCES

1. Abu-Hanna A, Lucas PJF. Intelligent Prognostic Methods in Medical Diagnosis and Treatment Planning. In: Borne P, Iksouri, el Kamel A, eds. Proceedings of Computational Engineering in Systems Applications (IMACS-IEEE, UCIS, Lille). 1998; 312-7.
2. Abu-Hanna A, Lucas PJF, eds. Prognostic Models in Medicine: AI and decision analysis approaches. Working Notes of the workshop on Prognostic Models in Medicine. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99) Aalborg, Denmark.
3. Altman DG, Royston, P. What do we mean by validating a prognostic model? *Stat Med* 2000; 19: 453-73.
4. Aitchison TC, Sirel JM, Watt DC, MacKie RM. Prognostic trees to aid prognosis in patients with cutaneous malignant melanoma. *BMJ* 1995; 311: 1536-9.
5. Asch D, Patton J, Hershey J. Knowing for the sake of knowing: the value of prognostic information. *Med Decis Making* 1990; 10: 47-57.
6. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press 1995.
7. Chang RWS. Individual outcome prediction models for the intensive care units. *Lancet* 1989; 1: 143-6.
8. Dietterich TG. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput* 1998; 10 (7): 1895-924.
9. De Dombal FT, Clamp SE, Softley A, Unwin BJ, Staniland JR. Prediction of Individual Patient Prognosis: Value of Computer Aided Systems. *Med Decis Making* 1986; 6: 18-22.
10. Ebell MH, Kruse JA, Smith M, Novak J, Drader-Wilcox J. Failure of three decision rules to predict the outcome of in-hospital cardiopulmonary resuscitation. *Med Decis Making* 1997; 17: 171-7.
11. George AL, Folk BP, Crecelius PL, Campbell WB. Pre-arrest morbidity and other correlates of survival after in-hospital cardiopulmonary arrest. *Am J Med* 1989; 87: 28-34.
12. Hand DJ. *Construction and assessment of classification rules*. Chichester: John Wiley and Sons 1997.
13. Hand DJ. *Datamining: Statistics and more?* *Am Stat* 1998; 52 (2): 112-8.
14. Hilden J, Habbema JDF. Prognosis in medicine: an analysis of its meaning and role. *Theor Med* 1987; 8: 249-365.
15. Le Gall JR, Lemeshow SL, Sauliner F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270 (24): 2957-63.
16. Lucas PJF, Boot H, Taal BG. Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma. *Meth Inform Med* 1998; 37: 206-19.
17. Lucas PJF, Abu-Hanna A. Prognostic methods in medicine. *Artif Intell Med* 1999; 15 (2): 105-19.
18. Lucas PJF, Van der Gaag LC. *Principles of Expert Systems*. Wokingham: Addison-Wesley 1991.
19. Mazumdar M and J. Glassman, Tutorial in biostatistics, categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat Med* 2000; 19: 113-32.
20. Neuberger J. Prognostic models – what is their future? *Eur J Gastroenterol Hepatol* 1997; 9: 1145-7.
21. Raiffa H. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley 1968.
22. Salzberg SL. On comparing classifiers: a critique of current research and methods. *Data Mining Knowl Discov* 1999; 1: 1-12.
23. Schwarzer G, Vach W, Schumacher M. On misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* 2000; 19: 541-61.
24. Shortliffe EH. Computer programs to support clinical decision making. *JAMA* 1987; 258: 61-6.
25. Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Boston: Butterworths 1988.
26. Steyerberg WW, Keizer HJ, Fossa SD, et al., Prediction of residual retroperitoneal mass histology following chemotherapy for metastatic nonseminomatous germ cell tumor: multivariate analysis of individual patient data from 6 study groups. *J Clin Oncol* 1995; 13: 1177-87.
27. Van der Gaag LC. Bayesian belief networks: odds and ends. *Comput J* 1996; 39: 79-113.
28. Windeler J. Prognosis - what does the clinician associate with this notion? *Stat Med* 2000; 19: 425-30.
29. Wyatt JC, Altman DGF. Commentary – prognostic models: clinically useful or quickly forgotten. *BMJ* 1995; 311: 1539-41.
30. Wyatt JC, Spiegelhalter DJ. Evaluating medical expert systems: what to test for and how? *Med Inform* 1990; 15 (3): 205-17.
31. Zhu BP, Lemeshow S, Hosmer DH, Klaar J, Avrunin J, Teres D. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996.

Address of the authors:

A. Abu-Hanna, PhD,  
Department of Medical Informatics,  
AMC-University of Amsterdam,  
Meibergdreef 15,  
1105 AZ Amsterdam  
The Netherlands  
E-mail: A.Abu-Hanna@amc.uva.nl