# Discovering Probabilistic Structures of Care

Arjen Hommersom and Sicco Verwer and Peter J. F. Lucas

Institute for Computing and Information Sciences
Radboud University Nijmegen, The Netherlands
{arjenh,s.verwer,peterl}@cs.ru.nl

**Abstract.** Medical protocols and guidelines can be looked upon as concurrent programs, where the patients dynamically change over time. Methods based on verification and model-checking developed in the past have been shown to offer insight into their correctness by adopting a logical point of view. However, there is uncertainty involved both in the management of the disease and the way the disease will develop, and, therefore, a probabilistic view on medical protocols seems more appropriate. On the other hand, representations using Bayesian networks usually involve a single patient group and do not capture the dynamic nature of care. In this paper, we propose a new method inspired by automata learning to represent and identify patient groups for obtaining insight into the care that patients have received. We evaluate this approach using data obtained from general practitioners and identify significant differences in patients who were diagnosed with a transient ischemic attack (TIA). Finally, we discuss the implications of such a computational method for the analysis of medical protocols.

**Topics covered:** Clinical guidelines ; temporal knowledge representations; knowledge extraction from healthcare databases

**Submission category:** Regular paper

## 1 Introduction

Much of the existing clinical knowledge that is concerned with quality of care is summarised in medical protocols and guidelines that describe standards of healthcare. From a computational point of view they can be looked upon as concurrent programs. Methods to investigate properties of protocols and guidelines, based on semi-automatic verification and model-checking, have been developed in the past (e.g. [1–3]). These methods take a logical point of view on protocols and guidelines and only offer insight into their formal correctness. A complementary view on healthcare is to look at the care that is actually given. This will reveal correspondences, usually called *compliance* [4], as well as differences with a given guideline, allowing one to obtain insight into where caregivers deliberately or accidentally departed from a guideline, and where they simply followed the guideline. Probabilistic models, such as Bayesian networks, allow one to capture, in principle, the necessary structural information from recorded data in

such way that the structures can be related to the logical structure of a guideline. Probabilistic approaches are in particular suitable for revealing the probabilistic nature of care processes, clarifying in essence how frequent particular care paths are taken. However, so far most of the research around care processes ignored probabilistic relational information. As a consequence, it is not completely clear which methodology for probabilistic methods can be used for this purpose, and what information they can actually reveal. In this paper we propose novel methods that can be used as a basis for such a methodology.

With the widespread introduction of information systems in healthcare during the last decade, there are now very big healthcare datasets available that enable developing such views on the structure of the given care. Examples of such datasets are those from NIVEL[1], a Dutch institutes with which we collaborate; they collect data of all patients of a large number of representative general practices in the Netherlands. For various diseases, the patients in this dataset have been treated according to guidelines. For example, for patients who were diagnosed with a transient ischemic attack (TIA), Dutch general practitioners generally follow a guideline developed by the NHG (Dutch General Practitioners society)[2]. However, guidelines are mostly concerned with single diseases, despite the fact that the majority of patients have multiple diseases (e.g. two-third of patients older than 65 years have two or more diseases at the same time). To get insight into the relationship between the guidelines and the actual care, which is described in healthcare data, computational learning methods can be of help.

In this paper, we take the first steps toward developing a technique for discovering probabilistic structures in healthcare data. The main idea behind the methodology is to combine ideas from Bayesian network learning with methods from learning automata. In particular, we will focus in this paper on one of the key ingredients of learning automata, which is the identification of *states*. The contribution of this technique is that the probabilistic representation that is learned provides insight into the different subgroups of patients. For example, we may identify patient groups with a different risk profile or patients groups that are treated significantly different from other patient groups. The underlying hypothesis is that these differences will be relevant in the care of the patient, and therefore should have a connection to the guideline.

This paper is organised as follows. In the next section, we discuss the background of computer-based protocols, Bayesian networks and automata learning. Then, in Section 3, the general idea of the paper is discussed and we introduce a new method for learning subgroups of patients by the identification of states. In Section 4, this new learning method is applied to a dataset consisting of patients diagnosed with a TIA and we discuss some possible implications for clinical guidelines. In Section 5, we discuss some related work and in Section 6, we conclude.

---

[1] http://www.nivel.nl

[2] http://nhg.artsennet.nl/kenniscentrum/k_richtlijnen/k_nhgstandaarden/ Samenvattingskaartje-NHGStandaard/M45_svk.htm (in Dutch)

## 2 Background: Protocols and Learning

In this section, we will discuss computer-based protocols and guidelines. After this, we introduce Bayesian networks and briefly introduce background on learning automata, which inspired the work presented in this paper.

### 2.1 Computer-based protocols and guidelines

Medical guidelines and protocols, medical protocols for short, are the main, prescriptive instrument of healthcare to promote quality of care [5]. Many countries have a special institute – e.g., the National Institute of Clinical Excellence (NICE) for the UK – that in collaboration with healthcare professionals, often clinical specialists and epidemiologists, and relevant patient organisations work on the production of such protocols.

Nowadays, there are also computer-based representations of medical protocols and this research has made considerable progress in the last decade [6]. A popular way to describe protocol modelling is through the paradigm of 'task-network models'. A task consists of a number of steps, each step having a specific function or goal [7, 8, 6]. Examples of languages that support task models, and which have been evolving since the 1990s, include PRO*forma* [9], Asbru [10], EON [11], and GLIF3 [8]. The network-of-task approach allows modelling the plan-like execution of protocols, which can also be modelled using logical methods (e.g. [12]). Computer-based versions of medical protocols allow adding support for their maintenance and updating without going through the entire text again as is still standard practice in protocol development.

### 2.2 Bayesian networks

Bayesian networks are powerful graphical representations that represent conditional independence assumptions [13], i.e., information about whether or not sets of random variables influence other sets of variables under the assumption that other variables have been observed for a problem at hand. There is a considerable body of work (e.g. [14, 15]), indicating that Bayesian networks offer a natural and intuitive formalism for constructing clinically relevant models.

Formally, a *Bayesian network* is a tuple $B = (G, X, P)$, with $G = (V, E)$ a directed acyclic graph (DAG), $X = \{X_v \mid v \in V\}$ a set of random variables indexed by $V$, and $P$ a joint probability distribution of the random variables in $X$. $P$ is represented as a Bayesian network with respect to the graph $G$ if $P$ can be written as a product of the probability of each random variable, conditional on their parents:

$$P(X_1, \ldots, X_n) = \prod_{v \in V} P(X_v \mid X_{\mathrm{pa}(v)})$$

where $\mathrm{pa}(v)$ is the set of parents of $v$. In the following, we will assume that each variable is binary with values *true* and *false*. We will write $x_i$ for $X_i = true$ and $\neg x_i$ for $X_i = false$.
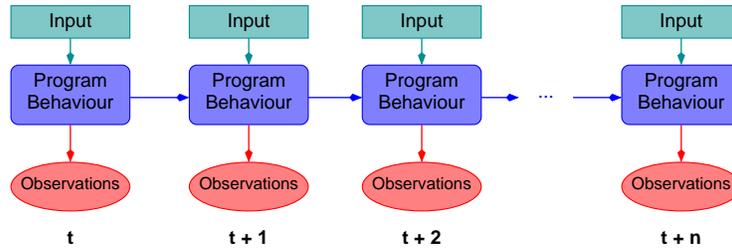
**Fig. 1.** Abstract temporal process, with the form of a Markov Process, driven by the execution of a program. Note that probabilistic graphical models support decomposing the program state by means of a graphical independence structure.

Bayesian networks allow modelling evolution of stochastic processes as a function of time; various types of so-called temporal Bayesian networks, also called dynamic Bayesian networks, have been proposed for this purpose [16]. A simple example of a temporal Bayesian network describing the state change of a program based on its execution is shown in Fig. 1. Use of techniques from temporal Bayesian networks offer interesting possibilities for studying program behaviour in detail. In particular, these allow exploring run-time behaviour of a given protocol by showing the interactions at different points in time.

Bayesian networks can also be learnt from data, which encompasses both learning the graph structure of the model and its associated parameters [17]. A major problem is that the search space of network structures (directed acyclic graphs) is extremely large [18] even if one takes into account that many different networks represent the same conditional independence information [19], i.e., are *Markov equivalent*. There are different ways to learn a Bayesian network from data using search-based, dependency-analytic and hybrid approaches, and the results obtained by these methods are generally good. Finally, there are methods available to learn temporal Bayesian networks from data [20].

### 2.3 Identification of automata

Another research area that is of immediate relevance to this paper is known as *automaton identification*, which concerns itself with constructing (learning) state machine models automatically from execution traces [21]. Since state machines are key models for the design and analysis of computer systems [22], the problem of learning finite state machines from data enjoys a lot of interest from the software engineering and formal methods communities. They use learned automaton models for providing insight into complex software systems and test their properties using model checking and testing techniques. In the literature, this approach has been used for learning and analysing models for different types of complex software systems such as web-services [23], the biometric passport [24], and java programs [25].

Formally, automaton identification can be seen as a grammatical inference [26] problem in which the traces are modelled as the words of a language, and the
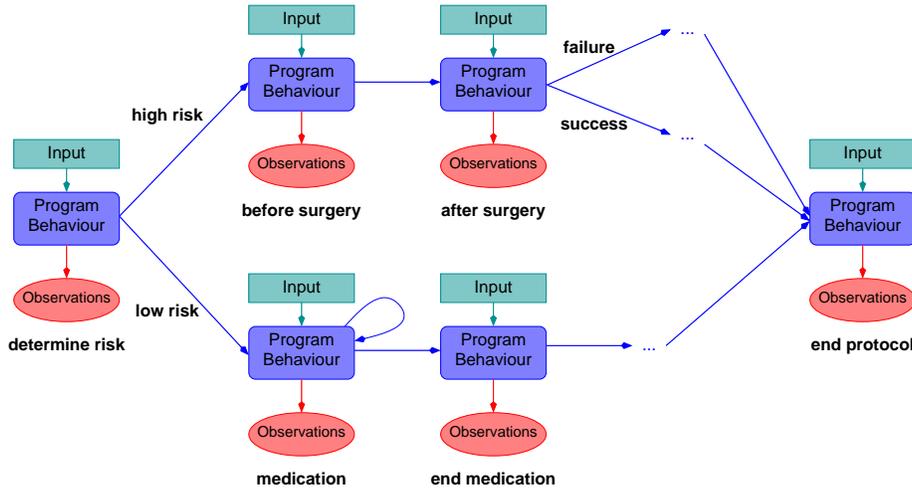
**Fig. 2.** An automaton representation of a medical protocol. The model contains guards that makes the future execution dependent on the value of a certain variable. In addition there is recursion that can loop back to already visited states. Every state contains a Bayesian network model for the properties of patients that visit those states, at the time(s) they visit them.

goal is to find a model for this language. The most commonly used language model is the deterministic finite state automaton (DFA) [27]. Hence, its learning (identification) problem is one of the best studied problems in grammatical inference, and many algorithms have been developed for this purpose.

## 3  Methods

### 3.1  General idea

It is surprising that learning Bayesian networks from program state data has never been tried since it is well known that Hidden Markov Models (HMMs, a type of temporal Bayesian network) and probabilistic automata are equivalent in terms of the distributions they can represent [28]. The standard method of adding behavioural abstractions to HMMs is to generalise the relations within a state to be an arbitrary Bayesian network [29], ending up with a model such as Fig. 1. Technically, there is no reason why this standard generalisation cannot be applied to an automaton model instead of an HMM, ending up with a model such as Fig. 2.

In this paper, we investigate how to identify the *states*, which is the key ingredient to learn automata. For example, the question is if we can identify that certain patient groups should be treated differently, e.g., low or high risk patient groups. If this is the case, then this (*i*) provides information about compliance in case the patients are treated according to some protocol, or (*ii*) indicates
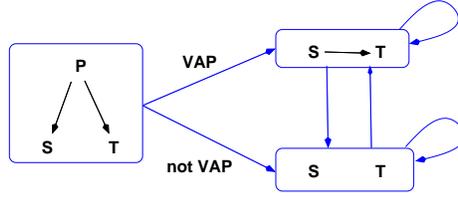
**Fig. 3.** A simplified example of contextual independence as it appeared in [31]. Here $P$ is pneumonia, $T$ is elevated body temperature, and $S$ is increased sputum production.

that this difference should be taken into account into the development of a new protocol.

### 3.2 Representation

The state can be seen a particular configuration of characteristics that hold for a patient group. In this first paper on learning automata of Bayesian networks, we take the most simple case: we condition on a single characteristic of the patient. This is essentially a multinet representation [30]. Suppose, for example, we have a joint distribution $P(X_V)$ over all variables $X_V$ represented by a Bayesian network $B = (G, X, P)$. By the chain rule, taking into account the independences represented by the graph, we can pick a single $v \in V$ and write:

$$P(X_V) = P(X_{V \setminus \{v\}} \mid X_v) P(X_v)$$

Let $X^* = X_{V \setminus \{v\}}$, then, assuming all variables are binary, consider the two conditional distributions:

$$P^{x_v}(X^*) = P(X^* \mid x_v)$$
$$P^{\neg x_v}(X^*) = P(X^* \mid \neg x_v)$$

and the distribution $P(X_v)$. Clearly, the distribution $P(X_v)$ is easy to represent by a single number $P(x_v)$ as $P(\neg x_v) = 1 - P(x_v)$. The distributions $P^{x_v}$ and $P^{\neg x_v}$, on the other hand, can be represented by Bayesian networks $B^{x_v}$ and $B^{\neg x_v}$. Obviously, the triple $\langle B^{x_v}, B^{\neg x_v}, P(x_v) \rangle$, which we call a *split Bayesian network* can represent exactly the same distribution as the original Bayesian network $B$. However, this more extended representation may indicate different relationships between variables in the populations where $x_v$ or $\neg x_v$ hold, and thus, can provide more insight than $B$ alone. It is not difficult to see that this can be done recursively by further conditioning on other characteristics.

Consider for example Fig. 3, showing a simplified example of relationships in patients with ventilator-associated pneumonia (VAP). When patients arrive at the ICU, they may already have pneumonia ($P$), which (indirectly) connects sputum production ($S$) with an elevated body temperature ($T$). Pneumonia diagnosed after they arrive at the ICU is classified as VAP. Only for the patients

with VAP, there is a relationship between $S$ and $T$. While the model on the left-hand-side would be an appropriate model for all patients at the ICU, the two models allow for a richer representation of the relevant knowledge.

### 3.3  Learning models

Typically, a split model is, while often more insightful, also a more complex model. In this paper, we propose two statistically motivated ways to determine whether these more complex models should be chosen over a Bayesian network representation.

A search-and-score-based method for learning Bayesian network uses a scoring function to measure the goodness of fit of a structure to the data. This score typically approximates the probability of the structure given the data and represents a trade-off between how well the network fits the data and how complex the network is. There are several ways to search for the optimal networks, e.g., a tabu search is often used. There are also various scoring functions for Bayesian networks. For example, in our experience, the Akaike information criterion (AIC) score works well for learning models from epidemiological datasets.

The AIC score of a split network can be derived as follows. Suppose have a dataset $D$ and $B$ a candidate Bayesian network with distribution $P$, let $L = \Pr(D \mid B) = \prod_{r \in D} P(r \mid B)$ be called the *likelihood* of the Bayesian network, where $r \in D$ is a record in dataset $D$, and the probability distribution of the Bayesian network $B$, $P$, is used to compute Pr using the common assumption that the records are independent and identically distributed. Furthermore, let $k$ be the number of parameters in the network, where $k = \sum_{v \in V} 2^{|pa(v)|}$, if the network contains only binary variables. Then the AIC score is defined as:

$$\text{AIC} = 2k - 2 \log L$$

Note that models with the lowest AIC are selected, i.e., with the highest likelihood and lowest number of parameters. Furthermore, suppose we split on $X_v$, let $D^{x_v}$ be the datarecords $D^{x_v} \subseteq D$ where $x_v$ holds, and $D^{\neg x_v} = D \setminus D^{x_v}$. Let $L^{x_v} = P^{x_v}(D^{x_v} \setminus \{x_v\}) P(x_v)$ and $L^{\neg x_v} = P^{\neg x_v}(D^{\neg x_v} \setminus \{\neg x_v\}) P(\neg x_v)$. Given a split model $M = \langle B^{x_v}, B^{\neg x_v}, P(x_v) \rangle$, it follows that the likelihood is the product of $L^{x_v}$ and $L^{\neg x_v}$. The number of of parameters is the number of parameters used to represent the Bayesian networks, plus one to represent $P(x_v)$. This yields:

$$\text{AIC} = 2(k^{x_v} + k^{\neg x_v} + 1) - 2 \log(L^{x_v} \times L^{\neg x_v}) = \text{AIC}^{x_v} + \text{AIC}^{\neg x_v} + 2$$

Of course, several other methods can be used to determine splits, such as the BIC or BDE score which are often used in Bayesian network learning.

Besides the score-based methods, automata learning uses hypothesis testing to determine whether a split should occur, in particular we will use a likelihood-ratio test. In that case, we consider a test statistic, which looks similar to the AIC:

$$D = -2 \log \frac{L}{L^{x_v} \times L^{\neg x_v}} = 2(\log L^{x_v} + \log L^{\neg x_v} - \log L)$$

$D$ is distributed according to a chi-squared distribution with $k^{x_v} + k^{\neg x_v} + 1 - k$ degrees of freedom. A significance test can thus be used to decide whether to split on a certain variable.

## 4  Experiments

Below we discuss our first experimental results, indicating that the proposed methodology is promising.

### 4.1  Data

The data used for analysis were obtained from the Netherlands Information Network of General Practice (LINH). All Dutch inhabitants are obligatory registered with a general practice, and the LINH registry contains information of routinely recorded data from about all patients of approximately 90 general practices. Longitudinal data of approximately one and half million patient years, covering the decade 2002-2011, were considered. From this data, we selected patients who were diagnosed with a transient ischemic attack (TIA) during this time-frame.

From this data, we selected a number of variables. This included the gender of the patient, a number of cardiovascular diseases (atherosclerosis, angina pectoris, stroke, cerebral infarction, hypertension, and heart failure), relevant classes of drugs that may be prescribed (antihypertensives, antilipemics, antithrombics, and antidiabetics), and a number of possible consequences of cardiovascular diseases (atrial fibrillation, orthostatic hypotension, and ankle edema).

### 4.2  Learning of networks

For the patients diagnosed with TIA we first learnt a Bayesian network from the available data, consisting of 600 patients who suffered a TIA. The resulting graph is shown in Fig. 4. The thickness of the arcs indicate the significance of the relationships, which was obtained by bootstrapping. The graph shows the statistical (in)dependences amongst variables. While this graph provides insight into the dependences between variables in the whole group of TIA patients, it is difficult to identify sub-groups where the relationships between variables are significantly different – which would suggest that this group may need to be managed differently from other patients.

Next we considered on which variables we could split and compared the AIC-motivated criterion to the likelihood-ratio criterion. There was a high degree of agreement between the two criteria, see Fig. 5. All the possible splits where there was an improvement in the AIC score were highly significant on the likelihood-ratio test ($p < 0.001$) and vice versa. The top five possible splits were *atherosclerosis*, *myocardial infarction*, *cerebral infarction*, *heart failure*, and *orthostatic hypertension*. The first, atherosclerosis, is somewhat of an outlier and this is almost never diagnosed directly. Instead, this may indicate patients which
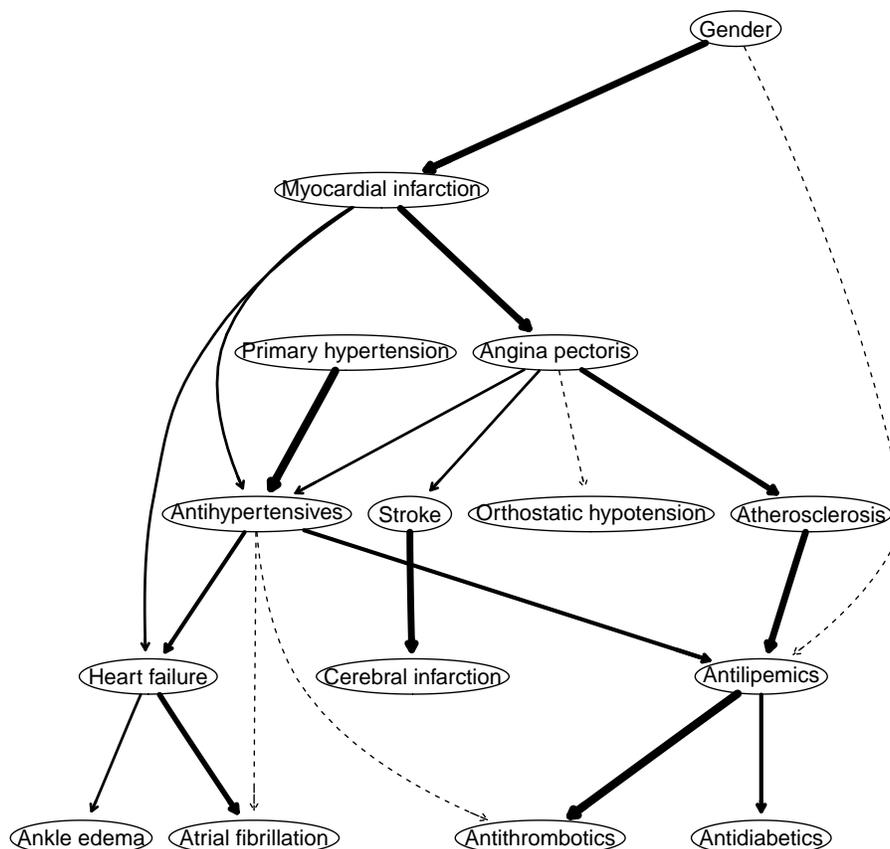
**Fig. 4.** Bayesian network structure learnt from patient data consisting of TIA patients.

have several cardiovascular diseases. The others are clearly patient-groups who were at higher risk than other patients.

To illustrate the results of this analysis, we take the most significant one, which was myocardial infarction. The data were split to find differences in network structure that allow explaining differences in the course of the disease, as shown in Figs. 6 and 7. These differences are related to both the patient characteristics and the way that patients are treated. For example, in the group of myocardial infarction, there is no clear (other) association with heart failure, because heart failure is very common in this group and the remaining patients are at high risk for this in any case. In the other group, heart failure is related to antilipemics (patients treated for high cholesterol) and antihypertensives (patients treated for high blood pressure). Clearly, in this patient group it makes sense to reduce the risk using several drugs. Another interesting difference is in the
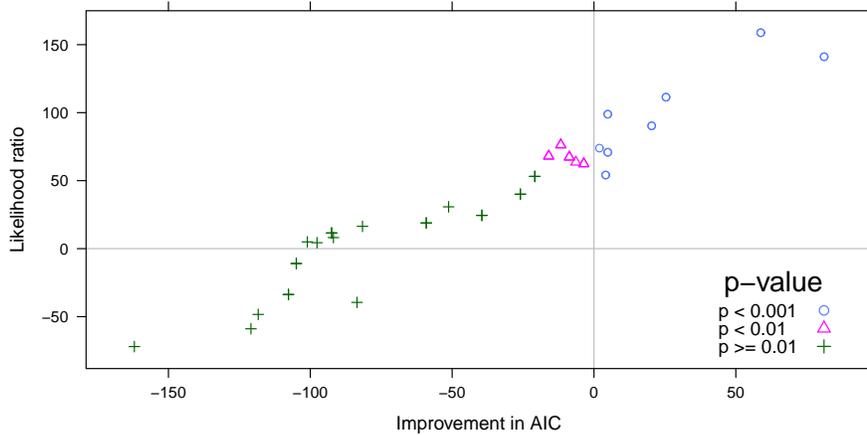
**Fig. 5.** Relationship between the improvement in AIC compared to a likelihood ratio.

treatment: for example, the gender of the patient seems relevant for prescribing antilipemics to patients who had a myocardial infarction, i.e., these drugs are much more likely to be prescribed to males who had heart attacks. For those who did not have myocardial infarction, the gender was less significant, i.e., most likely the cholesterol level is reason alone. While this gender difference is not noted in the guideline for TIA or in the Dutch guideline for cardiovascular risk management (2006), statins have gender-specific differences [32].

## 5    Related work

As already mentioned, the representation that was discussed here can be seen as a multinet [30]. These multinets were proposed to represent contextual independence, whereas we learn these networks to discover different subpopulations. As a consequence, by recursively applying the learning approach, we obtain a much richer representation than multinets alone. However, it is beyond of this paper to learn a complete automaton, as we focused on the identification of states.

A more related approach to learn Bayesian networks in context of subgroup discovery is by Duivesteijn et al. [33]. The idea of this paper is to compare structural differences between subgroups. This ideas is different from our paper, which aims to actually use subgroup discovery to uncover potentially different care paths. As a consequence, we proposed statistical criteria based on the AIC and the likelihood-ratio test. Furthermore, we aim to extend this approach to learn automata of Bayesian networks, rather than do subgroup discovery alone.

Furthermore, there have been some proposals to induce guidelines from data. For example, one could look upon learning guidelines as learning a decision tree [34], and possibly integrate these models with additional medical background
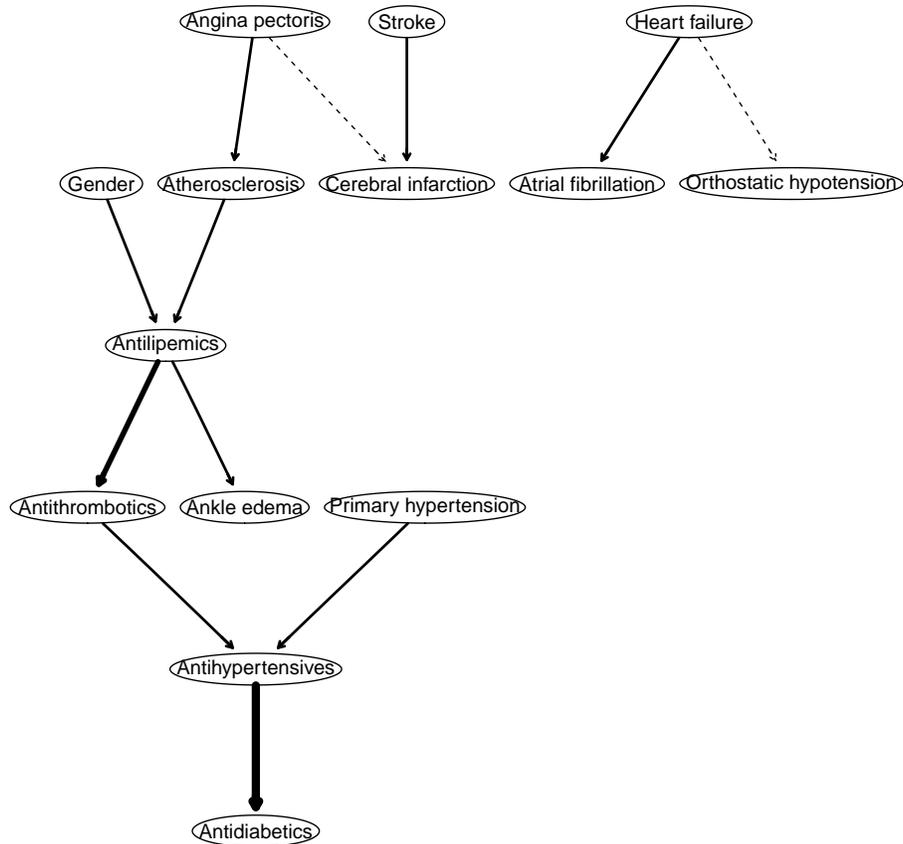
**Fig. 6.** Split of TIA patients also diagnosed with myocardial infarction.

knowledge [35]. Another approach uses process mining techniques [36–38], where the idea is to extract process models from event logs [36, 37]. The main difference is that in process mining there is no abstraction of the events into (probabilistic) states. We think this is an important step, especially for learning clinical models, as it is important for clinicians have understandable models that describe a certain healthcare process. Since there are both probabilistic aspects in guidelines as well as the patients that are being treated by a guideline, a probabilistic model of the process seems more appropriate. Nonetheless, some of the ideas from the process mining field could be combined with abstraction into states, as presented in this paper.
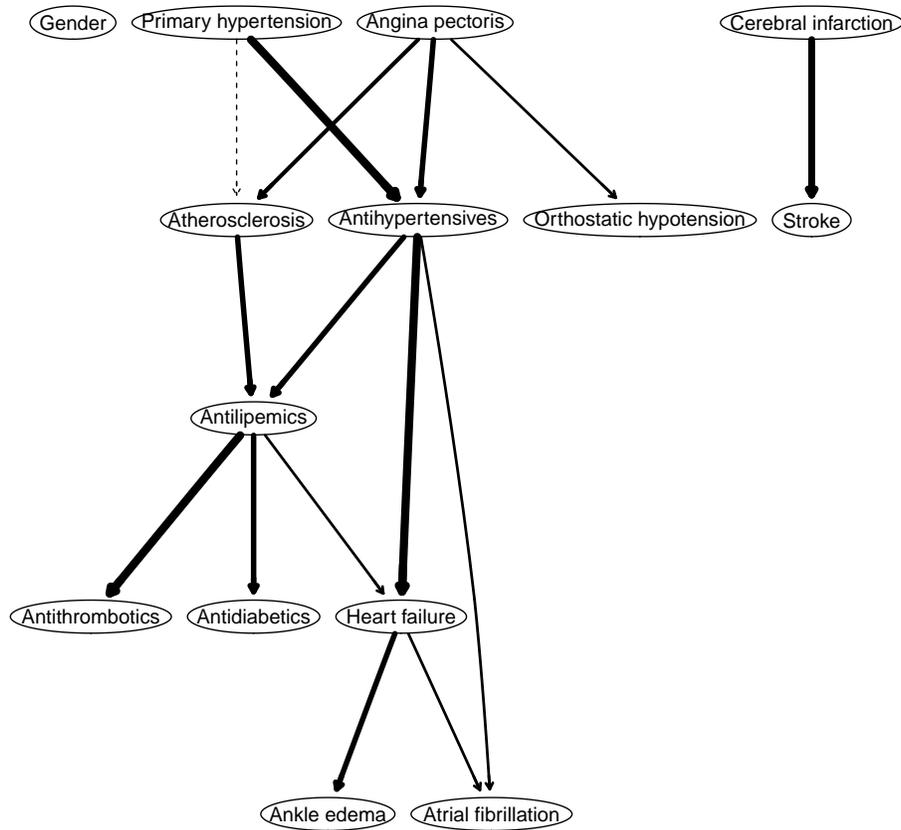
**Fig. 7.** Split of TIA patients who have not been diagnosed with myocardial infarction.

## 6 Conclusions

In this paper, we introduced a new method for discovery of structure in epidemiological datasets. We view this as the essential step to learn automata that describe care processes. In this paper, we introduced the necessary learning methods and applied these ideas to a dataset consisting of patients diagnosed with a TIA. We argued that the technique identifies subpopulations that can be seen as groups that are different from the others.

The evaluation we presented in this paper is still fairly static: we included all TIA patients, regardless of temporal relationships between events. In future work, we will include time, so that complete automata can be learned from the data. We believe that this can be used to learn representations of the *actual*

care, which can then be compared more formally to the care recommended by guidelines.

## References

1. ten Teije, A., Marcos, M., Balser, M., van Croonenborgd, J., Duellic, C., van Harmelena, F., Lucas, P., Miksch, S., Reif, W., Rosenbrand, K., Seyfang, A.: Improving medical protocols by formal methods. Artificial Intelligence in Medicine **63**(3) (2006) 193–209
2. Hommersom, A., Groot, P., Lucas, P., Balser, M., Schmitt, J.: Verification of medical guidelines using background knowledge in task networks. IEEE Transactions on Knowledge and Data Engineering **19**(6) (2007) 832–846
3. Bottrighi, A., Giordano, L., Molino, G., Montani, S., Terenziani, P., Torchio, M.: Adopting model checking techniques for clinical guidelines verification. Artificial Intelligence in Medicine **48**(1) (2010) 1 – 19
4. Quaglini, S.: Compliance with clinical practice guidelines. In Teije, A.t., Miksch, S., Lucas, P., eds.: Computer-based Medical Guidelines and Protocols: A Primer and Current Trends. Volume 139 of Studies in Health Technology and Informatics. IOS Press (2008) 160–179
5. Field, M., Lohr, K., eds.: Clinical Practice Guidelines: Directions for a New Program. National Academy Press, Institute of Medicine, Washington D.C. (1990)
6. ten Teije, A., Miksch, S., Lucas, P., eds.: Computer-based Clinical Guidelines and Protocols: a Primer and Current Trends. IOS Press, Amsterdam (2008)
7. Fox, J., Das, S.: Safe and Sound: Artificial Intelligence in Hazardous Applications. AAAI Press (2000)
8. Peleg, M., Boxwala, A., Ogunyemi, O., Zeng, P., Tu, S., Lacson, R., Begnstam, E., Ash, N.: GLIF3: The evolution of a guideline representation format. In: Proc. AMIA Annual Symposium. (2000) 645–649
9. Fox, J., Johns, N., Rahmanzadeh, A., Thomson, R.: PRO*forma*: a general technology for clinical decision support systems. Computer Methods and Programs in Biomedicine **54** (1997) 59–67
10. Shahar, Y., Miksch, S., Johnson, P.: The Asgaard project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. Artificial Intelligence in Medicine **14** (1998) 29–51
11. Tu, S., Musen, M.: From guideline modeling to guideline execution: Defining guideline based decision-support services. In: Proceedings of American Medical Informatics Association Symposium, Los Angeles, CA (1999) 863–867
12. Hommersom, A., Groot, P., Lucas, P., Balser, M., Schmitt, J.: Verification of medical guidelines using background knowledge in task networks. IEEE Transactions on Knowledge and Data Engineering **19**(6) (2007) 832–846
13. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)
14. Andreassen, S.: Planning of therapy and tests in causal probabilistic networks. Artificial Intelligence in Medicine **4**(3) (1992) 227–241
15. Lucas, P., van der Gaag, L., Abu-Hanna, A.: Bayesian networks in biomedicine and health-care. Artificial Intelligence in Medicine **30** (2004) 201–214
16. Dagum, P., Galper, A., Horvitz, E.: Dynamic network models for forecasting. In: Proceedings of UAI92. (1992) 41–48
17. Neapolitan, R.: Learning Bayesian Networks. Pearson (2004)

18. Robinson, R.: Counting unlabeled acyclic graphs. In: LNM 622, Springer, NY (1977) 220–227
19. Gillespie, S.B., Perlman, M.D.: Enumerating Markov Equivalence Classes of Acyclic Digraph Models. In: UAI-2001. (2001)
20. Ghahramani, Z.: Learning dynamic bayesian networks. In: Adaptive Processing of Sequences and Data Structures, Springer-Verlag, Heidelberg (1998) 168–197
21. Cook, J.E., Wolf, A.L.: Discovering models of software processes from event-based data. ACM Trans. Softw. Eng. Methodol. **7** (July 1998) 215–249
22. Lee, D., Yannakakis, M.: Principles and methods of testing finite state machines - a survey. Proceedings of the IEEE **84** (1996) 1090–1123
23. Bertolino, A., Inverardi, P., Pelliccione, P., Tivoli, M.: Automatic synthesis of behavior protocols for composable web-services. In: Proceedings of the joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering, ACM (2009) 141–150
24. Aarts, F., Schmaltz, J., Vaandrager, F.: Inference and abstraction of the biometric passport. In: Proceedings of the 4th international conference on Leveraging applications of formal methods, verification, and validation - Volume Part I, Springer-Verlag (2010) 673–686
25. Walkinshaw, N., Bogdanov, K., Holcombe, M., Salahuddin, S.: Reverse engineering state machines by interactive grammar inference. In: Proceedings of the 14th Working Conference on Reverse Engineering, IEEE (2007) 209–218
26. de la Higuera, C.: Grammatical Inference: Learning Automata and Grammars. Cambridge University Press, New York, NY, USA (2010)
27. Sudkamp, T.A.: Languages and Machines: an introduction to the theory of computer science. Third edn. Addison-Wesley (2006)
28. Dupont, P., Denis, F., Esposito, Y.: Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms. Pattern Recognition **38** (2005) 1349–1371
29. Boutilier, C., Dearden, R., Goldszmidt, M.: Exploiting structure in policy construction. In: IJCAI, AAAI (1995)
30. Geiger, D., Heckerman, D.: Knowledge representation and inference in similarity networks and Bayesian multinets. Artificial Intelligence **82** (1996) 45–74
31. Visscher, S., Lucas, P.J.F., Flesch, I., Schurink, K.: Using temporal context-specific independence information in the exploratory analysis of disease processes. In: AIME. (2007) 87–96
32. Gutierrez, J., Ramirez, G., Rundek, T., Sacco, R.L.: Statin therapy in the prevention of recurrent cardiovascular events: a sex-based meta-analysis. Arch. Intern. Med. **172**(12) (Jun 2012) 909–919
33. Duivesteijn, W., Knobbe, A., Feelders, A., van Leeuwen, M.: Subgroup discovery meets bayesian networks – an exceptional model mining approach. In: Proceedings of the 2010 IEEE International Conference on Data Mining. ICDM '10, Washington, DC, USA, IEEE Computer Society (2010) 158–167
34. Bohada, J.A., no, D.R., López-Vallverdú, J.A.: Automatic generation of clinical algorithms within the state-decision-action model. Expert Systems with Applications **39**(12) (2012) 10709 – 10721
35. López-Vallverdú, J.A., no, D.R., Bohada, J.A.: Improving medical decision trees by combining relevant health-care criteria. Expert Systems with Applications **39**(14) (2012) 11782 – 11791
36. Van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: Discovering process models from event logs. Knowledge and Data Engineering, IEEE Transactions on **16**(9) (2004) 1128–1142

37. Mans, R., Aalst, W.v.d., Vanwersch, R., Moleman, A.: Process mining in healthcare data challenges when answering frequently posed questions. In Lenz, R., Miksch, S., Peleg, M., Reichert, M., Riao, D., ten Teije, A., eds.: Process Support and Knowledge Representation in Health Care (BPM 2012 Joint Workshop, ProHealth 2012/KR4HC 2012. Volume 7738 of LNCS., Springer (2012) 140–153

38. Kaymak, U., Mans, R., van de Steeg, T., Dierks, M.: On process mining in health care. In: SMC. (2012) 1859–1864