BAYESIAN NETWORKS FOR THE ANALYSIS OF INPATIENT ADMISSIONS

Caterina Conigliani[1]
Dipartimento di Economia, Università Roma Tre, Rome, Italy

Tommasangelo Petitti
Università Campus Bio-Medico, School of medicine, Hygiene, Public health and Statistics research unit, Rome, Italy

Vincenzina Vitale
Dipartimento di Scienze Sociali ed Economiche, Università di Roma La Sapienza, Rome, Italy

Abstract *A care pathway is defined as a complex intervention for the organisation of care processes for a specific group of patients during a specific period. Although the analysis of care pathways has been shown its benefits in clinical practices, little attention has been devoted to study how it can contribute to the optimization of the use of resources. In particular here we focus on the analysis of the history of a large number of patients' admissions, i.e. of data that belong to the routine flow of information that all hospital provide to the Local Healthcare Agency. One goal is the identification of the most likely sequence of wards/clinics for a patient; in fact, knowing which wards/clinics are more interrelated can be useful for a better hospital organization. Moreover we suggest the use of Bayesian Networks to predict the care pathway that each patient will undertake, given his/her history.*

Keywords: *Bayesian networks, care pathways, inpatient admission.*

1. INTRODUCTION

Health Operations Management can be described as the analysis, design, planning and control of all the steps that are required to provide a service to a patient (Vissers and Beech, 2005). In particular, Vissers and Beech distinguish between five levels of Health Operations Management: (*i*) a care plan for each patient (planning and protocol at the level of a patient), (*ii*) the planning of care in care pathways (planning and control for a group of patients), (*iii*) the planning of the capacity of professionals, equipment and space (planning and control at the level of the resource), (*iv*) the planning of the number of patients to be treated and care

[1] Corresponding author: Caterina Conigliani, Dipartimento di Economia, Università Roma Tre, via Silvio D'Amico 77, 00145 Roma, caterina.conigliani@uniroma3.it

activities to be implemented (planning and control at the level of the patient volume), (*v*) the long-term policy of the institution (strategic planning). As noticed in Schrijvers et al. (2012), these levels are strongly connected, and are influenced by the introduction of *care pathways* in health services.

Recall that a care pathway is defined as *a complex intervention for the mutual decision-making and organisation of care processes for a well-defined group of patients during a well-defined period* (Vanhaecht et al., 2007). As pointed out for instance in Aringhieri and Duma (2015), a care pathway can be seen as an algorithm based on a flow chart that specifies all the decisions, the treatments, the events related to a patient with a specific pathology. Notice that a care pathway can be analysed at a single level of care (e.g. a hospital) or globally, considering every possible level of health care (from education and prevention to diagnosis, treatment and recovery). And although it has been shown its benefits in clinical practices, little attention has been devoted to study how it can contribute to the optimization of the use of resources (Aringhieri and Duma, 2015).

In fact, the analysis of care pathways is a powerful bottom-up tool that allows to identify the patient's profiles starting from their socio-demographic features and their pathologies. The most common procedure in this sense is the systematic analysis of the literature in order to determine the diagnostic and clinical best practices, that can be used to build decision trees and top-down guidelines for the clinicians (Twaddle and Qureshi, 2005). An alternative to this approach is that of building the patient clinical pathways starting from the analysis of the history of a large number of patients and of admissions. This procedure has the advantage of taking into account real flows of patients, that can be compared with the clinical best practices described in the literature (Lodewijckx et al., 2012). It is interesting to notice that another possible use of this procedure is that of studying the typical activity of a hospital. This allows to determine the best programming strategies in terms of (*i*) understanding the evolution of cases in short-term treatment, (*ii*) checking the residual production capacity of the facility, (*iii*) estimating the Life Time Values of patients that need to enter the system in order to establish the medium-term sustainability of the facility.

Here we take this second approach, and concentrate on pathways within a clinic (that do not include outpatient department's activities, discharge from the hospital and after-care). In particular, although our approach can be clearly generalised for taking into account all the events that constitute the path that a patient suffering from a disease walks in the National Health System, we focus on patients' admissions. And in order to build clinical pathways from real flows of

patients and admissions, we explore the use of Bayesian networks. These are presented in details in Section 2. Section 3 shows the Bayesian networks and the typical patient profiles estimated using the admissions of 19300 patients of Campus Bio-Medico, a University Hospital in southern Rome (Italy), between 1998 and 2014. The analysis of the admissions is conducted both by ward and by clinic, and in both cases we show how the use of Bayesian networks allows to insert and propagate evidence in the form of the history of a patient up to a certain stage. Section 4 considers the problem of testing the predictive ability of the Bayesian network by applying fivefold cross-validation. A few concluding remarks are presented in the final section.

2. BAYESIAN NETWORKS

Bayesian networks (Cowell et al. (1999); Lauritzen (1996); Pearl (1988)) belong to the wider class of Probabilistic Graphical Models that efficiently encode the joint probability distribution for a large set of variables. In particular, a Bayesian Network (BN) represents a multivariate probability distribution over a set of random variables $\mathbf{X} = \{X_1, \ldots, X_k\}$ by means of *i)* a network structure D expressing the conditional independence statements about the variables, and *ii)* a set P of local probability distributions associated with each variable, and it can handle complex probabilistic models by decomposing them into smaller components. Moreover, the availability of easy and computationally efficient algorithms for evidence propagation makes BNs a very useful statistical tool to carry out what-if analyses. For all these reasons, BNs have been widely applied in several different fields, addressing tasks such as diagnosis, prediction, decision making, classification and data mining.

More specifically, the network structure D of a BN is a directed acyclic graph (DAG) that is defined by a set V of nodes (or vertices), each of which represents a random variable in $\mathbf{X}$, and a set E of directed links (or edges) between pairs of nodes, each of which represents a directed relation of one variable to the other, arranged without producing cycles. Notice that if an arrow points from X_i to X_j, then X_i is named a parent of X_j ($pa(Xj)$) and X_j is named a child of X_i. Moreover, the set P of local probability distributions is linked to $D = (V;E)$ through the Markov condition whereby each variable is conditionally independent of its non-descendants given its parents, so that the joint distribution of $\mathbf{X}$ can be factorised as:

$$p(X_1, \ldots, X_k) = \prod_{j=1}^{k} p(X_j | pa(X_j)) \tag{1}$$

with an evident computational gain.

Notice that once the structure of the network and the probabilities in Equation 1 have been estimated, the BN becomes a very useful tool to perform probabilistic inference. In particular, Jensen et al. (1990) and Lauritzen and Spiegelhalter (1988) have developed efficient algorithms that allow to update the marginal probability distributions of some variables when information on other variables is inserted in the network.

2.1. STRUCTURAL LEARNING

Bayesian networks' structure and parameters (i.e. the probabilities in (1)) can be elicited from a domain expert knowledge, which is often a difficult procedure, or retrivied from the data. It is also possible to induce a network structure and its conditional probabilities on the basis of both data and some domain expert background knowledge; this can speed up the learning process severalfold and improve the accuracy of the elicitation.

More specifically, two different methods can be used for learning the structure of Bayesian networks from the data: *constraint-based* algorithms and *search and score-based* algorithms (see Cooper and Herskovits (1992) and Neapolitan (2004)).

The *score-based* method is an optimization-based search approach, based on a scoring function and a search procedure; it explores the space of all possible candidate networks assigning a score to each one, reflecting its goodness of fit, and selecting that with the highest score. This, however, is often infeasible in practice because the number of possible DAGs grows super-exponentially with the number of nodes. To overcome this limitation, a greedy search strategy such as the *hill climbing* can be applied: starting from a network structure (usually the empty one), in each step the algorithm selects as the new candidate the neighbouring structure of the current network with the highest score. More specifically, it iteratively performs small changes to the current structure (adding, deleting or reversing one edge at a time), and it stops when the score can no longer be improved. Obviously, the main problem of this approach is the possibility to find a local maxima rather than a global one.

Instead, *constraint-based* algorithms test the occurrence of a relation between pairs of variables by means of appropriate hypothesis tests of independence, and then build a graph which satisfies the corresponding independence statements. The most known algorithm beloging to this group, that we are applying in the present work, is the *PC algorithm* due to Spirtes et al. (2000). The version im-

plemented by the software Hugin performs the following steps: *i)* for all pairs of nodes, a statistical test verifies the assumption of conditional independence (unless a structural constraint has been forced by the researcher); *ii)* starting from an empty graph, it creates an undirected graph called the *skeleton*, that includes undirected edges between pairs of nodes for which no conditional independence relations were found; *iii)* the orientation step starts by looking for sets of three variables X_i, X_j, X_l such that X_i and X_l are independent, but dependent conditionally on X_j, that can be uniquely represented by the v-structure $X_i \longrightarrow X_j \longleftarrow X_l$; *iv)* the rest of the edges are then oriented so as to avoiding cycles and new v-structures. Notice that the PC algorithm works under the assumption that the distribution of the observed variables is faithful to a DAG (i.e. that data have been simulated from a probability distribution that factorizes according to a DAG), under the assumption of infinite data sets and no hidden variables.

3. A REAL CASE STUDY

We have considered inpatient admissions in different wards of Campus Bio-Medico between 1998 and 2014 of patients with at least two admissions and no more than seven, and we have estimated a BN with seven nodes, each of which represents the ward of the corresponding admission. The wards included in our analysis are: 1) Surgery, 2) Cardio-Neuro-Vascular, 3) Onco-hematology, 4) Internal Medicine, 5) Diagnostics. Furthermore, we have considered the history of a patient concluded after 18 months of the last admission, so that any further admission of the same person counts as that of a new patient, and we have excluded from the analysis patients whose history is not concluded (i.e. patients whose last admission is less that 18 months old), for a total of 19300 patients.

3.1. THE ANALYSIS OF ADMISSIONS BY WARDS

The estimated BN structure for the 19300 patients in the different wards is shown in Figure 1, and looks quite reasonable: the first inpatient admission directly influences the second, the third, the fourth and the fifth ones; the second inpatient admission directly influences the third and the fourth ones; the third inpatient admission directly influences the fourth and fifth ones; from the fourth inpatient admission, each admission directly influences only the following one. Notice that the BN of Figure 1 was obtained by implementing the PC algorithm with a temporal ordering constraint for nodes.

Note that once the BN of Figure 1 has been estimated, it can be employed in different ways. First, by means of the chain rule decomposition in (1), it can

be used to compute the joint probabilities of the different patient profiles, which provide useful informations for a better hospital organization from the point of view of which wards are more interrelated; Table 1 shows these probabilities for the most likely patient profiles. Second, it can be used to propagate evidence (in our case the history of a patient up to a certain stage) through the graph, and to investigate how such evidence changes the marginal distributions of the remaining nodes.

Assume, for instance, that the first admission of a patient was in the Surgery ward. The BN in Figure 2 shows that this evidence modifies the marginal probabilities of the different wards as far as the second admission is concerned: with respect to Figure 1, the probabilities that the second admission is in Cardio-Neuro-Vascular, in Onco-hematology, in Medicine decrease, while the probability that the second admission is again in Surgery increases from 0.48 (in Figure 1) to 0.86 (in Figure 2). Obviously also the probabilities for the following admissions change. For instance, by comparing Figure 1 with Figure 2 we see that the probability that the third admission is in Surgery increases, even if the most likely category for the third admission is *none*, and the same happens for the fourth admission.

If we then insert a second evidence in the second node of the network, for instance an admission in Onco-hematology (Figure 3), the situation changes again. With respect to Figure 2, the probability that there is not a third admission decreases from 0.62 (in Figure 2) to 0.43 (in Figure 3), the probability that the third admission is in Surgery decreases from 0.29 (in Figure 2) to 0.16 (in Figure 3), while it increases significantly the probability that the third admission is in Onco-hematology, from 0.04 (in Figure 2) to 0.36 (in Figure 3). Similarly, the probability that there is not a fourth admission decreases from 0.82 (in Figure 2) to 0.69 (in Figure 3), the probability that the fourth admission is in Surgery decreases from 0.13 (in Figure 2) to 0.08 (in Figure 3), while the probability that the fourth admission is in Onco-hematology increases from 0.02 (in Figure 2) to 0.20 (in Figure 3).

If we then insert a third evidence in the third node of the network, for instance an admission in Surgery (Figure 4), the situation changes again. With respect to Figure 3, the probability that there is not a fourth admission decreases from 0.69 (in Figure 3) to 0.59 (in Figure 4), the probability that the fourth admission is in Onco-hematology decreases from 0.20 (in Figure 3) to 0.10 (in Figure 4), while it increases significantly the probability that the fourth admission is in Surgery, that is now the ward with the highest probability, from 0.08 (in Figure 3) to 0.28

(in Figure 4). And something similar happens for the following admissions: it is unlikely that there will be a fifth, sixth or a seventh admission, but if there is one, then it is probably in Surgery.

3.2. THE ANALYSIS OF ADMISSIONS BY CLINICS

Notice that a similar analysis of that presented in the previous section can be performed also with respect to the inpatient admissions in the different clinics of Campus Biomedico, which are shown in Table 2 and whose number is rather large compared of that of wards. In this case the analysis with the BN is made quite difficult by the large number of possible clinics at each node of the network, i.e. by the enormous number of possible patient profiles, most of which are in fact very unlikely. For this reason, here we are going to limit the number of possible patient profiles by considering two different scenarios. In the first one we consider inpatient admission in different clinics of Campus Biomedico of patients whose first two admissions were in the Surgery ward, for a total of 7878 patients; from Table 1 we can see that there are at least seven patient profiles with this feature which receive a high probability. In the second scenario we consider inpatient admission in different clinics of Campus Biomedico of patients whose first admission was in the Cardio-Neuro-Vascular ward, for a total of 4788 patients; again, from Table 1 we can see that this corresponds to different high probability profiles.

Consider initially the first scenario. Figure 5 shows the estimated BN structure for the patients whose first two admissions were in the Surgery ward with respect to their admissions in the different clinics: we can see that in this case each admission directly influences only the following one, so that for instance given the second admission, the clinic of the third admission is independent of that of the first admission. As in the previous section, this BN can be used to compute the joint probabilities of the different patient profiles by means of the chain rule decomposition in (1), which are shown in Table 3. Moreover, one could insert and propagate different evidence in the BN, that would change the marginal distributions of the various nodes.

Consider now the second scenario. Figure 6 shows the estimated BN structure for the patients whose first admission was in the Cardio-Neuro-Vascular ward with respect to their admissions in the different clinics; also in this case each admission directly influences only the following one, and this BN can be used to compute the joint probabilities of the different patient profiles, which are shown in Table 4.

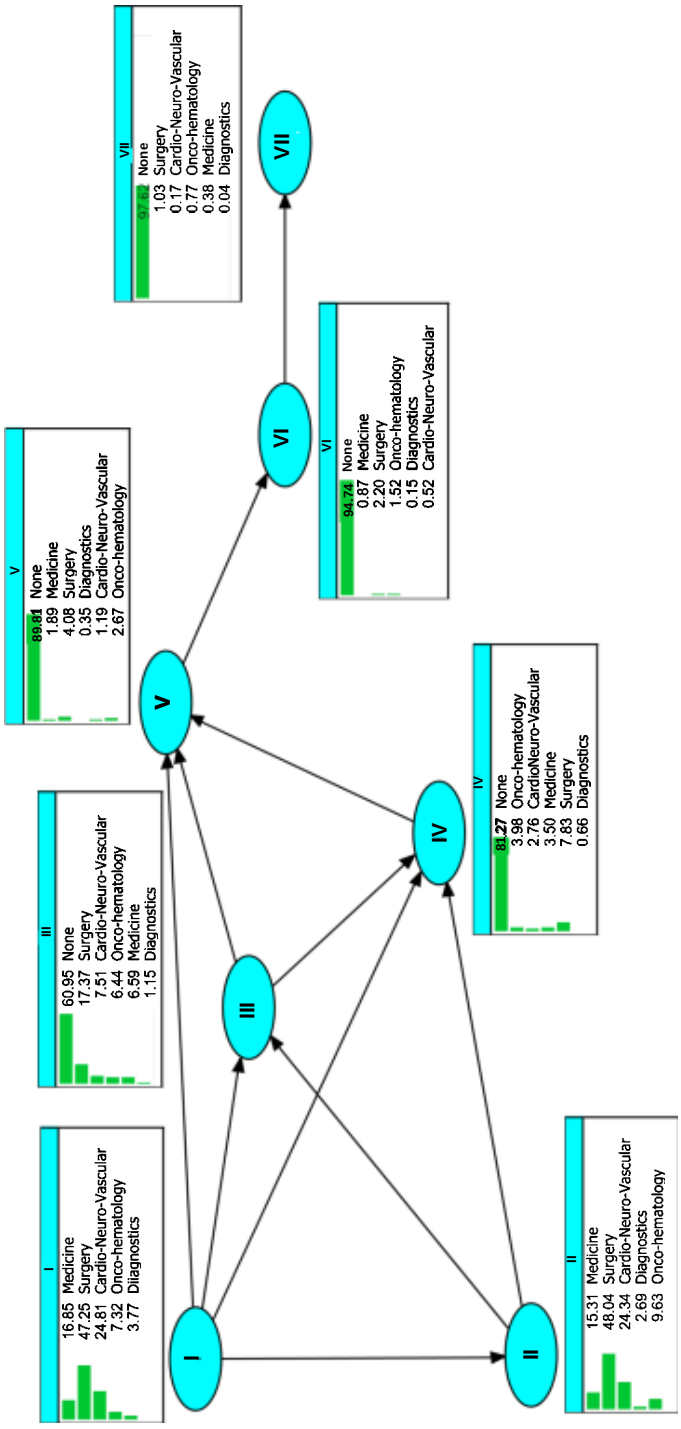

Fig. 1: The BN for the analysis of admission by ward

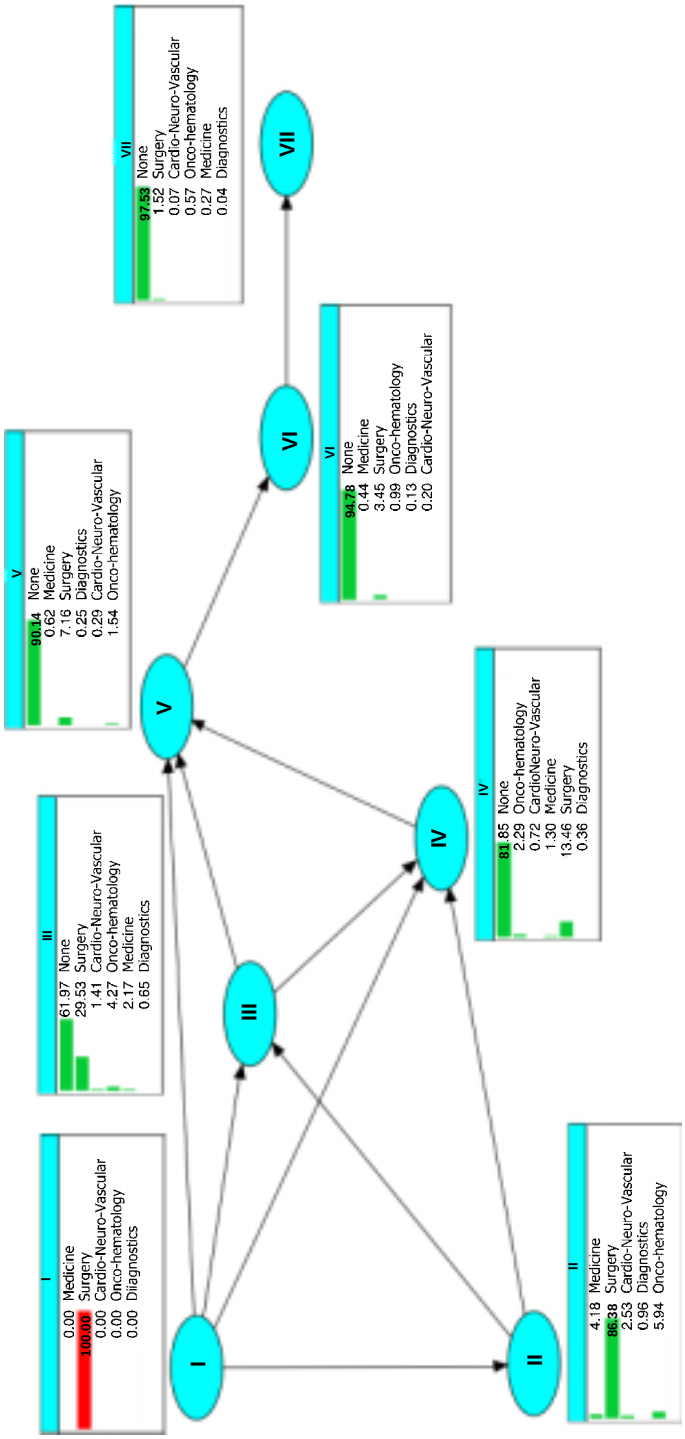

Fig. 2: The BN with evidence of a first admission in Surgery

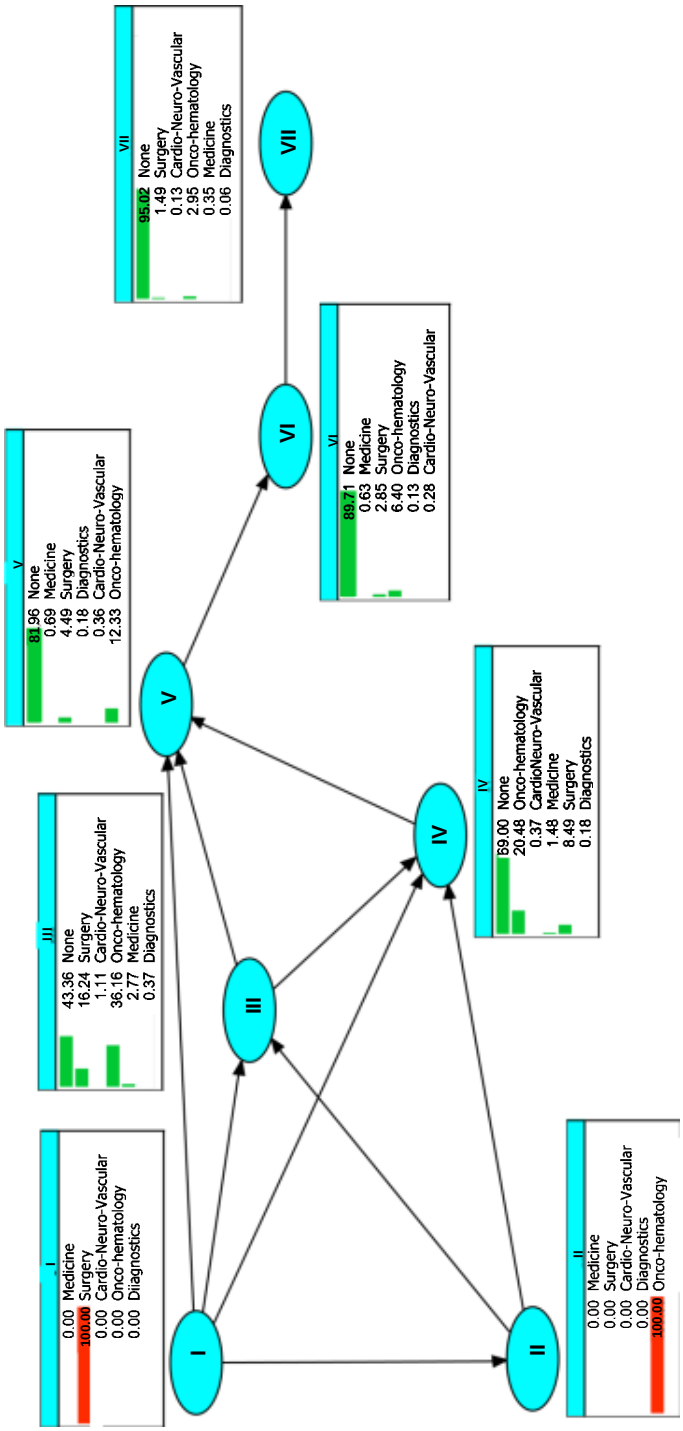

Fig. 3: The BN with evidence of a first admission in Surgery and a second admission in Onco-hematology

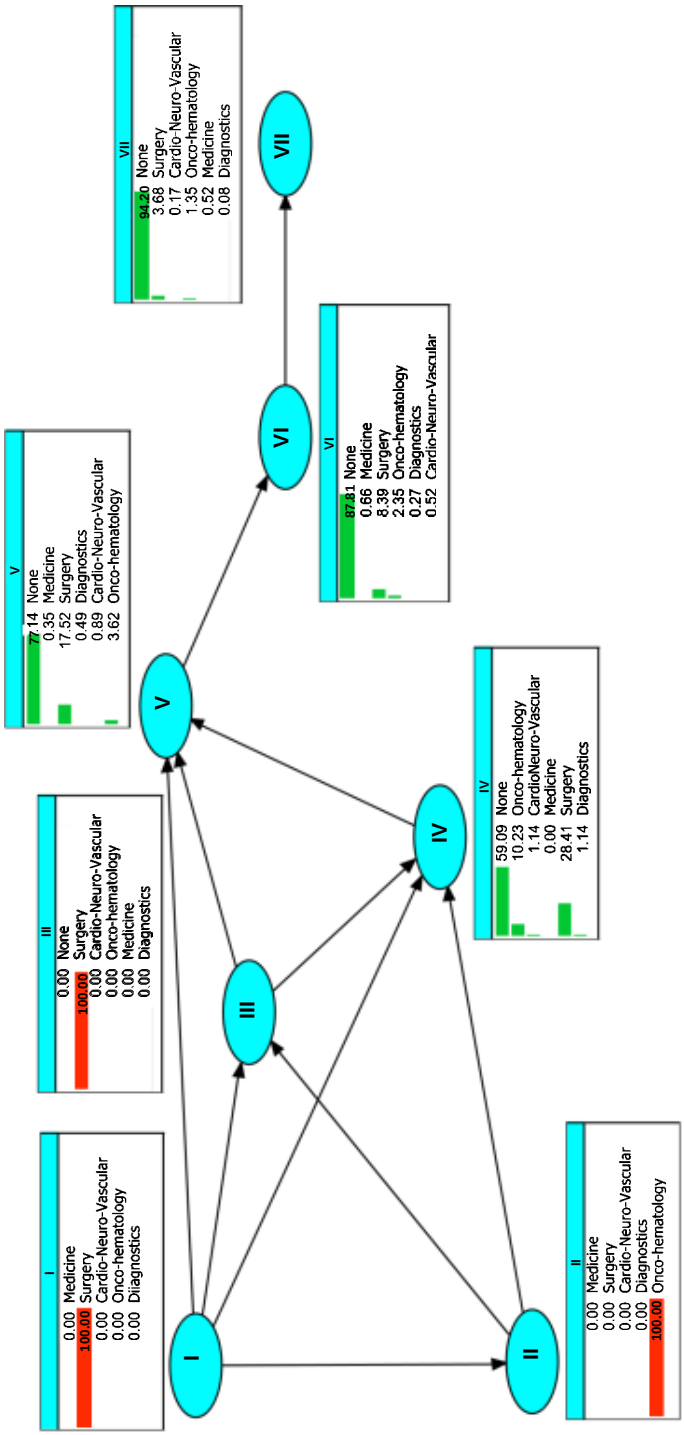

Fig. 4: The BN with evidence of a first admission in Surgery, a second admission in Onco-hematology, a third admission in Surgery

Tab. 1: The patient ward profiles with highest probabilities

Probability	I admission	II admission	III admission	IV admission	V admission	VI admission	VII admission
0.260	Surgery	Surgery					
0.149	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular					
0.068	Surgery	Surgery	Surgery				
0.067	Medicine	Medicine					
0.037	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular				
0.024	Surgery	Surgery	Surgery	Surgery			
0.021	Onco-hematology	Onco-hematology					
0.020	Medicine	Medicine	Medicine				
0.019	Medicine	Surgery					
0.013	Surgery	Surgery	Surgery	Surgery	Surgery		
0.012	Surgery	Onco-hematology					
0.011	Surgery	Medicine					
0.011	Onco-hematology	Onco-hematology	Onco-hematology				
0.011	Diagnostics	Diagnostics					
0.011	Cardio-Neuro-Vascular	Surgery					
0.009	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular	Cardio-Neuro-Vascular			
0.009	Medicine	Cardio-Neuro-Vascular					
0.009	Medicine	Medicine	Medicine	Medicine			
0.008	Cardio-Neuro-Vascular	Medicine					
0.007	Diagnostics	Surgery					
0.007	Surgery	Surgery	Surgery	Surgery	Surgery	Surgery	
0.007	Surgery	Cardio-Neuro-Vascular					
0.006	Onco-hematology	Onco-hematology	Onco-hematology	Onco-hematology			
0.006	Surgery	Surgery	Surgery	Surgery	Surgery	Surgery	Surgery
0.005	Onco-hematology	Onco-hematology	Onco-hematology	Onco-hematology	Onco-hematology		
0.005	Medicine	Medicine	Medicine	Medicine	Medicine		
0.005	Onco-hematology	Surgery					
0.004	Surgery	Surgery	Onco-hematology				

Tab. 2: Clinics corresponding to the different wards

Cardio-Neuro-Vascular	**Surgery**	**Medicine**	**Onco-hematology**	**Diagnostics**
Cardiac surgery	Short stay surgery	Endocrinology	Hematology	Endoscopy
Cardiology	General surgery	Hepatology	Oncology	Interventional radiology
Neurology	Geriatric surgery	Gastroenterology		
Cardiac intensive care unit (CICU)	Reconstructive surgery	Geriatrics		
	Thoracic surgery	Internal medicine		
	Gynecology	Nephrology		
	Onco-gynecology			
	Oculistics			
	Orthopedics			
	Otorinolaringoiatry			
	Rehabilitation			
	Breast surgery			
	Urology			

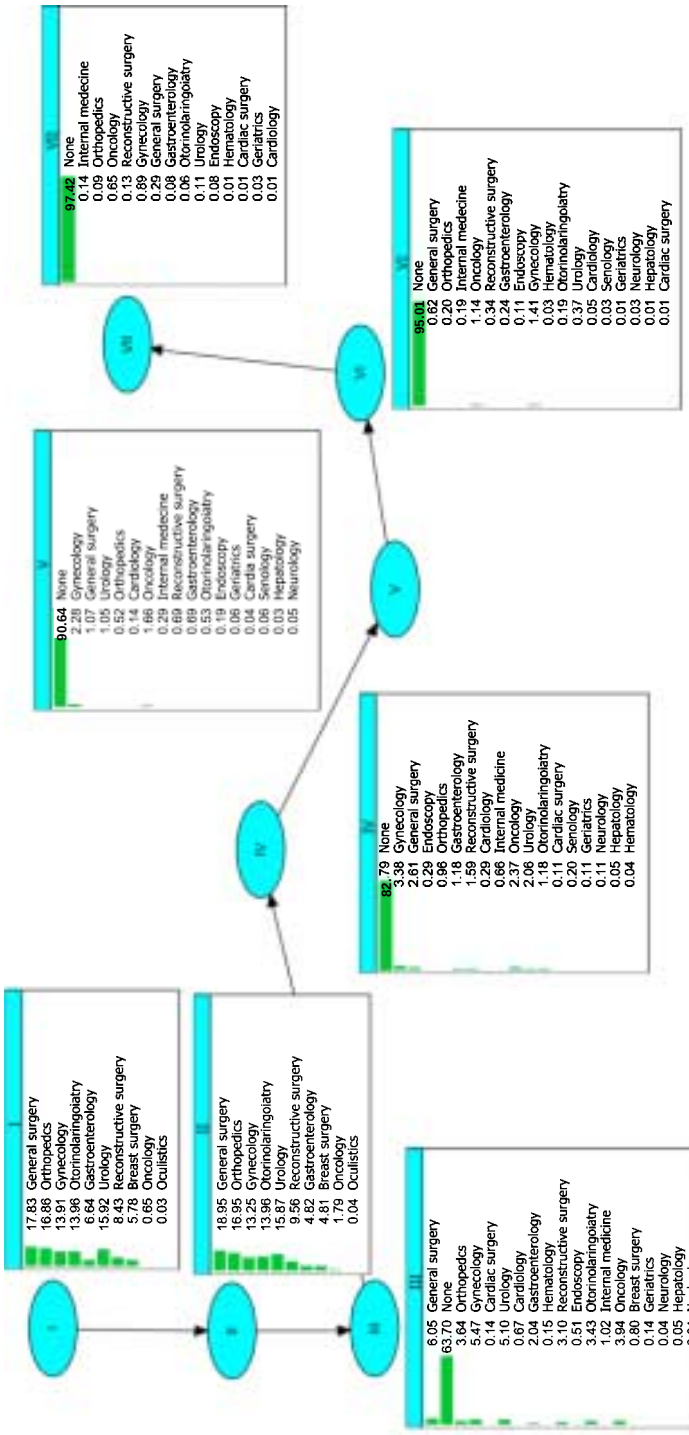

Fig. 5: The BN for the analysis of admissions by clinic (conditional on the first two admissions being in Surgery)

Tab. 3: The patient clinic profiles with highest probabilities (conditional on the first two admissions being in Surgery)

Probability	I admission	II admission	III admission	IV admission	V admission	VI admission	VII admission
0.116	Orthopedics	Orthopedics	None	None	None	None	None
0.091	Otorinolaringoiatry	Otorinolaringoiatry	None	None	None	None	None
0.091	Urology	Urology	None	None	None	None	None
0.088	General surgery	General surgery	None	None	None	None	None
0.063	Gynecology	Gynecology	None	None	None	None	None
0.051	Reconstructive surgery	Reconstructive surgery	None	None	None	None	None
0.024	Urology	Urology	Urology	None	None	None	None
0.022	Breast surgery	Breast surgery	None	None	None	None	None
0.021	General surgery	General surgery	General surgery	None	None	None	None
0.021	Orthopedics	Orthopedics	Orthopedics	None	None	None	None
0.019	Gastroenterology	Gastroenterology	None	None	None	None	None
0.018	Otorinolaringoiatry	Otorinolaringoiatry	Otorinolaringoiatry	None	None	None	None
0.017	Gynecology	Gynecology	Gynecology	None	None	None	None
0.013	Gastroenterology	General surgery	None	None	None	None	None
0.010	Reconstructive surgery	Reconstructive surgery	Reconstructive surgery	None	None	None	None
0.007	Urology	Urology	Urology	Urology	None	None	None
0.007	Gynecology	Gynecology	Gynecology	Gynecology	None	None	None
0.006	General surgery	General surgery	General surgery	General surgery	None	None	None
0.006	Breast surgery	Reconstructive surgery	None	None	None	None	None
0.005	General surgery	Orthopedics	None	None	None	None	None
0.005	Gynecology	Gynecology	Gynecology	Gynecology	Gynecology	None	None
0.005	Gastroenterology	Gastroenterology	Gastroenterology	None	None	None	None
0.004	Otorinolaringoiatry	Otorinolaringoiatry	Otorinolaringoiatry	Otorinolaringoiatry	None	None	None
0.004	Gynecology	Gynecology	Gynecology	Gynecology	Gynecology	Gynecology	Gynecology
0.004	General surgery	Urology	None	None	None	None	None

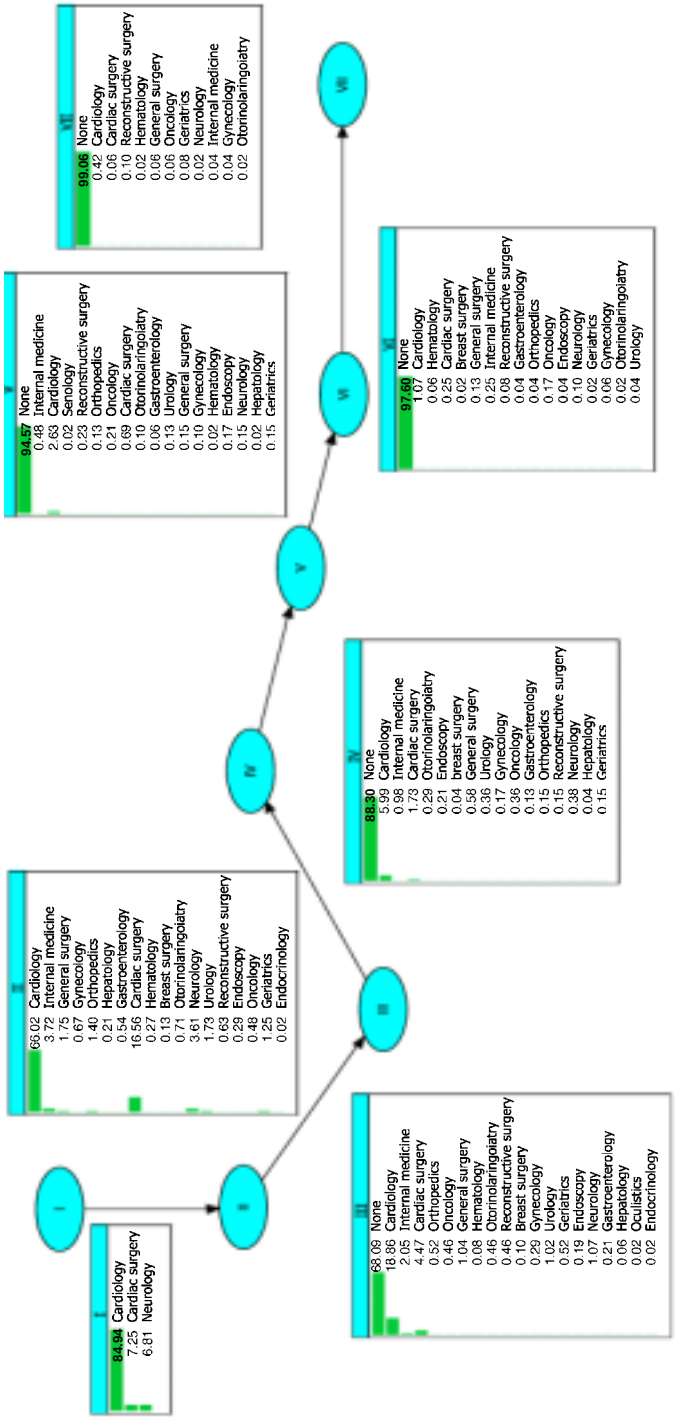

Fig. 6: The BN for the analysis of admissions by clinic (conditional on the first admission being in Cardio-Neuro-Vascular)

Tab. 4: The patient clinic profiles with highest probabilities (conditional on the first admissions being in Cardio-Neuro-Vascular)

Probability	I admission	II admission	III admission	IV admission	V admission	VI admission	VII admission
0.433	Cardiology	Cardiology	None	None	None	None	None
0.107	Cardiology	Cardiology	Cardiology	None	None	None	None
0.096	Cardiology	Cardiac surgery	None	None	None	None	None
0.029	Cardiac surgery	Cardiac surgery	None	None	None	None	None
0.024	Neurology	Neurology	None	None	None	None	None
0.023	Cardiology	Cardiology	Cardiology	Cardiology	None	None	None
0.015	Cardiology	Internal medicine	None	None	None	None	None
0.012	Cardiology	Cardiac surgery	Cardiac surgery	None	None	None	None
0.012	Cardiology	Cardiology	Cardiac surgery	None	None	None	None
0.010	Cardiac surgery	Cardiology	None	None	None	None	None
0.010	Cardiology	Cardiology	Cardiology	Cardiology	Cardiology	None	None
0.008	Cardiology	General surgery	None	None	None	None	None
0.008	Cardiology	Urology	None	None	None	None	None
0.006	Neurology	Cardiology	None	None	None	None	None
0.005	Cardiology	Orthopedics	None	None	None	None	None
0.005	Neurology	Neurology	Neurology	None	None	None	None
0.004	Neurology	Geriatrics	None	None	None	None	None
0.004	Cardiology	Geriatrics	None	None	None	None	None
0.004	Cardiology	Cardiology	Cardiology	Cardiac surgery	None	None	None
0.004	Cardiac surgery	Cardiac surgery	Cardiac surgery	None	None	None	None
0.003	Cardiology	Cardiology	Cardiology	Cardiology	Cardiology	Cardiology	None
0.003	Cardiac surgery	Internal medicine	None	None	None	None	None
0.003	Cardiology	Cardiac surgery	Cardiology	None	None	None	None
0.003	Cardiology	Internal medicine	Internal medicine	None	None	None	None
0.003	Cardiology	Cardiology	Internal medicine	None	None	None	None
0.003	Cardiology	Otorinolaringoiatry	None	None	None	None	None
0.003	Neurology	Internal medicine	None	None	None	None	None

4. PREDICTION

We now consider the problem of testing the predictive ability of the BN. It is important to stress that, given the history of a patient, the task of predicting if there is going to be a new admission (within 18 months from the last one) is indeed rather difficult. However, if we condition on the fact that there will be a new admission, it is interesting to check the performance of the Bayesian network in predicting the ward (or the clinic) that this will involve. For simplicity this is be done here by considering patients with exactly five admissions, for a total of 951 patients, and with reference to wards.

Notice that the fundamental idea of predictive validation is to split the data into two subsets, the *training data*, that are used to build a model (a BN in our case), and the *validation data*, that are used to test the performance of the model for prediction. With respect to resubstitution validation, this approach has the advantage of avoiding over-fitting, but the results can be highly dependent on the choice for the training/validation split. One popular way to deal with this is to consider cross-validation: in k-fold cross-validation the data is first partitioned into k equally sized folds; subsequently k iterations of training and validation are performed, in such a way that within each iteration a different fold of the data is held-out for validation while the remaining k-1 folds are used for learning. Notice that this procedure allows for overlapping training sets, while keeping the validation sets independent, and that the performance of the model on each fold can be tracked using some performance measure; upon completion, k samples of the performance measure will be available, and different methodologies such as averaging can be used to obtain an aggregate measure.

Table 5 shows the results of fivefold cross-validation for predicting the wards of the fifth admission of patients, given their first four admissions; the proportion of correct predictions reaches nearly the 82%.

5. CONCLUSIONS AND FUTURE STEPS

The aim of care pathways is to improve the quality of health care from the point of view of the patient outcomes, safety and satisfaction, and from the point of view of optimizing the use of resources. One of the key ingredients is the integration of evidence-based knowledge (Lodewijckx et al., 2012). Care pathways have been suggested as a way to translate national guidelines into local protocols, and from a clinical point of view there is enough evidence in their favour to justify further evaluation of their impact (Campbell et al., 1998). However little attention has

Tab. 5: Results of fivefold cross-validation for predicting the wards of the fifth admission

	Predicted				
Observed	CNV	S	M	OH	D
Cardio-Neuro-Vascular (CNV)	105	12	8	6	0
Surgery (S)	13	309	26	17	6
Medicine (M)	14	18	161	8	3
Onco-hematology (OH)	3	18	3	180	3
Diagnostics (D)	4	9	3	0	22

been devoted to studying how care pathways can contribute to the management of a facility critical resources, such as beds and operating rooms.

The analysis presented in this paper shows how the use of Bayesian networks allows to obtain important informations about a patient profile that can easily converge into a care pathway. In particular, we have only been looking at the patients admissions, i.e. at their administrative history, without introducing purely clinical variables. One of the main advantages of this approach is that it uses only data belonging to the routine flow of information that all hospitals provide to the Local Healthcare Agency and to the Ministry of Health.

Notice that the fact that we have not included in the analysis any clinical variable implies that the quality of the estimates depends significantly on the number of events that make up the history of the patient. In this sense it is clearly interesting to explore to what extent the estimates of patient clinical paths can be improved by including in the model further informations both about each admission (i.e. diagnosis, duration of the hospitalization, time elapsed since previous admission, medical treatments provided) and about each patient (i.e. age, sex, potential chronic deseases). The possibility to use these informations in order to reliably predict both the probability of a future admission and its time frame will be the object of future research.

REFERENCES

Aringhieri, R. and Duma, D. (2015). The optimization of a surgical clinical pathway. In M. Obaidat, T. ÃÚren, J. Kacprzyk, and J. Filipe, eds., *Simulation and Modeling Methodologies, Technologies and Applications. Advances in Intelligent Systems and Computing, vol 402*. Springer, Cham.

Campbell, H., Hotchkiss, R., Bradshaw, N. and Porteous, M. (1998). Integrated care pathways. In *BMJ*. 316: 133-137.

Cooper, G.F. and Herskovits, E.A. (1992). A bayesian method for the induction of probabilistic networks from data. In *Machine Learning*. 9: 309-347.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York.

Jensen, F.V., Lauritzen, S.L. and Olesen, K.G. (1990). Bayesian updating in causal probabilistic networks by local computations. In *Computational Statistics Quarterly*. 4: 269-282.

Lauritzen, S.L. (1996). *Graphical Models*. Clarendon Press, Oxford.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. In *Journal of the Royal Statistical Society B*. 50: 157-224.

Lodewijckx, C., Decramer, M., Sermeus, W., Panella, M., Deneckere, S. and Vanhaecht, K. (2012). Eight-step method to build the clinical content of an evidence-based care pathway: the case for copd exacerbation. In *Trials*. 13: 229.

Neapolitan, R.E. (2004). *Learning Bayesian Networks*. Pearson Prentice Hall, NewHaven, Connecticut.

Pearl, J. (1988). *Probabilistic Inference in Intelligent Systems*. Morgan Kaufmann Publishers, San Mateo, CA.

Schrijvers, G., van Hoorn, A. and Huiskes, N. (2012). The care pathway: concepts and theories: an introduction. In *International Journal of Integrated Care*. 12(Special Edition Integrated Care Pathways): e192. doi: 10.5334/ijic.812

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Adaptive Computation and Machine Learning, second edition.

Twaddle, S. and Qureshi, S. (2005). Scottish intercollegiate guilines network. In *Evidence-Based Healthcare & Public Health*. 9: 405-409.

Vanhaecht, K., De Witte, K. and Sermeus, W. (2007). The care process organisation triangle: a framework to better understand how clinical pathways work. In *Int J Care Pathw*. 11: 1-8.

Vissers, J. and Beech, R. (2005). *Health operations management: patient flow statistics in health care*. Routledge, Taylor & Francis Group Publishers, New York.