

Disabling and reoperation in patients with Crohn's disease subject to early surgery or immunosuppression: a Bayesian network prognostic model

Cláudia Camila Dias

CIDES & CINTESIS

Faculty of Medicine, University of Porto
Porto, Portugal
camila@med.up.pt

Fernando Magro

MedInUP

Institute of Pharmacology & Therapeutics
Faculty of Medicine, University of Porto
Porto, Portugal
fm@med.up.pt

Pedro Pereira Rodrigues

CINTESIS & LIAAD – INESC TEC

Faculty of Medicine, University of Porto
Porto, Portugal
pprodrigues@med.up.pt

Abstract— Crohn's disease is one type of inflammatory bowel disease whose incidence is currently increasing, subject to relapse and disabling, with unknown etiology, and usually diagnosed between the second and third decade of life. The aim of this work is to develop a Bayesian network tool to predict disabling and reoperation in patients with Crohn's disease subject to early surgery or immunosuppressors intake. Multi-centric study data from patients with surgery or immunosuppression in the first six months after diagnosis was used, focusing on the prognosis and the analysis of factors' interaction. Patients were grouped by the index episode: immunosuppressors intake, and surgery (stratified considering the use or not of immunosuppressors 6 months after surgery). Patient group was associated with disease behavior, upper gastrointestinal tract location (L4) and age at diagnosis, while disease extent was associated to perianal disease. For disabling, association between perianal disease and gender and location was also found. Association between gender and L4 was also found for reoperation. The cross-validated discriminative power of the models were high for both disabling (above 70%) and reoperation (above 80%). The generated models presented interesting insights on factor interaction and predictive ability for the prognosis, supporting their use in future clinical decision support systems.

Keywords: *prognosis; Crohn's disease; clinic decision support; Bayesian network models*

I. INTRODUCTION

Crohn's disease is a chronic illness with unknown etiology usually diagnosed during the second and third decade of life. The interaction between genetic and environment factors had a important role in the etiology of the disease [1].

The disease could had along periods of remission but the more aggressive cases requires more aggressive interventions like surgery [2] but, being a chronic disease, neither treatment nor surgery actually heals the patients, yielding frequent medical visits and hospitalizations, which in turn creates uncertainty about the professional and social future of patients and their families [3].

The treatment of Crohn's disease has changed in the last decades. Strategies like step-up or top down treatments have been more frequently chosen. The top-down strategy is based on the very early use of intensive therapy (immunosuppressive and/or biologics) to maintain a good quality of life from the first flare-up of the disease and prevent any irreversible consequences [4].

In our day the principal focus of treatment has been the improvement of quality of life, reduction of surgeries or hospitalization, besides the sole control of symptoms. Since the treatment plan clearly affects the disease course, focus has been given to identifying good prognostic models based on clinical/demographic factors since they are more easily included in daily clinical practice [5], [6].

However, prognostic studies show heterogeneous results, likely as a result of using different methodologies and/or applying different criteria for selection and evaluation. In addition, predicting the prognosis is a considerably uncertain task, so the development of predictive models also requires research. Also, traditional biostatistics is no longer enough to cope with the real-world biomedical data and it is necessary to look to other techniques [7]. Bayesian networks are a good choice since they deal well with prior knowledge, collected in published evidence included in quality literature or in primary or secondary data sources, and transform the data analyses in a process of updating the prior knowledge with available evidence at inference time [9], dealing better than other models with the uncertainty in the data. Bayesian networks are represented by a qualitative model (which describes the relations among variables) and a quantitative model (which gives the joint probability of all variables) which allows the inspection of marginal probabilities for each variable's state, and their use in the computation of a posteriori probabilities for single patients at inference time [10].

A preliminary study on the problem constituted a proof-of-concept for the use of Bayesian network classifiers as prognostic predictor model for disabling of prognostic prediction of disabling [15]. However, a thorough follow-up clinical appraisal of those results revealed that the analysis of immunosuppressive therapy prior to six months after surgery was misleading, so a new definition of groups should be

considered. Also, an additional important outcome should be modelled, as clinicians usually consider the risk for reoperation as a decisive factor for defining the best intervention for those patients.

The aim of this work is to develop Bayesian networks for the predictive prognosis of patients with Crohn's disease subject to early surgery or immunosuppression, namely targeting disabling disease and reoperation.

II. MATERIAL AND METHODS

Multi-centric study data of Crohn's patients with surgery or immunosuppression in the first six months after diagnosis with more than 18 years old and at least 3 years of follow were included. A total of 489 patients were included (out of 668 patients collected). Hundred and seventy-nine patients were excluded because of missing values.

A. Studied variables and outcomes

This study was a retrospective study and the collected variables included: characterization of patients (gender), disease (data of diagnosis and intervention (surgery and/or immunosuppressors, location with or not upper gastrointestinal tract, behavior and perianal disease), and follow up data (namely number of surgery and hospitalizations, treatments and adverse events). Patients were grouped by the index episode: immunosuppressors intake, and surgery (stratified considering the use or not of immunosuppressors 6 months after surgery).

The two main outcomes were disabling (defined by the presence of at least one of the following criteria: more than one surgery, excluding the first; more than two hospitalizations, excluding the first; two steroids course requirements per year, steroids dependency, steroids refractory; need to switch immunosuppressors or Anti-TNF therapy; adverse events) and reoperation (defined by the deed of second surgery).

B. Model building and evaluation

Following the preliminary study presented previously [15], Tree Augmented Naïve Bayes (TAN) models were derived using all studied variables to predict disabling and reoperation. To assess the general structure and accuracy of learned models, stratified 10-fold cross-validation was repeated 10 times, estimating accuracy and the area under the ROC curve.

WEKA software [12] was used to learn the Bayesian network structure. gRain [13] and pROC [14] R packages were also used to estimate de parameters of the networks and ROC curves, respectively.

III. RESULTS

The main characteristics of patients are shown in Tab. I. Forty-eight percent of patients took immunosuppressors in the first 6 months after diagnosis and 36% had surgery with immunosuppressors 6 months after surgery. Forty-six percent were male and 21% had more than 40 years old at diagnosis.

Concerning location 47% had ileal disease and 32% had penetrating disease. We observed that 64% of patients had disabling during the course of the disease and 18% needed a second surgery.

TABLE I. MAIN CHARACTERISTICS OF PATIENTS WITH CROHN'S DISEASE INCLUDED IN THE STUDY (N=489).

	n	(%)
Event		
S ₀ surgery	80	(16%)
S ₁ surgery and immuno > 6 mo after surgery	175	(36%)
I immunosuppression	234	(48%)
Gender		
Male	225	(46%)
Age at diagnosis		
A1 & A2 (<40 years)	388	(79%)
A3 (>=40 years)	101	(21%)
Location		
L1 Ileal	232	(47%)
L2 Colonic	45	(9%)
L3 Ileocolic	212	(43%)
L4 upper		
Yes	55	(11%)
Behavior		
B1 Non structuring/non penetrating	158	(32%)
B2 Structuring	176	(36%)
B3 Penetrating	155	(32%)
Perianal disease		
Yes	125	(26%)
Disabling disease	314	(64%)
Reoperation	89	(18%)

A. Bayesian network model qualitative analysis

The qualitative models were show in Fig 1 and 2. Patient group was associated with disease behavior, upper gastrointestinal tract location (L4) and age at diagnosis, while disease extent was associated to perianal disease.

For disabling, association between perianal disease and gender and location was also found. Association between gender and L4 was also found for reoperation.

B. Bayesian network model validation

Fig. 3 presents the in sample, leave-one-out and 10 times 10-fold cross validation ROC curves for disabling and reoperation models, respectively. Resulting validation AUC were between 72% and 73%. (for disabling) and between 79% and 80% (for reoperation).

IV. DISCUSSION

The main contribution of this work was the development of prognostic models for disabling disease and reoperation. These models were developed after a preliminary study [15] done for the adequacy of these techniques into this problem, where comparison with other methods, namely logistic regression, has been assessed. Other clinical problems were also already addressed in the past with good performance, easy interpretation and friendly representations [7], [8], [11].

From models presented in Fig. 1 and 2 it was possible to infer some known clinical associations. In both outcomes, patient group was associated with disease behavior (whether it is penetrating or presenting stenosis or neither), upper gastrointestinal tract location (L4) and age at diagnosis, while disease extent was associated to perianal disease. These global dependencies show a common thread for knowledge representation in Crohn's disease management, specifically tune for each outcome considering the added associations between perianal disease and gender and location (found for disabling) and the association between gender and L4 (found for reoperation).

This knowledge representation allows an easy visualization of the model, supporting the experts' decision beyond the use of risk factors and discriminative cut points, into a support on the interdependences of studied variables and their induced causality.

From the quantitative analysis, and following the discussion from previous preliminary study - which showed that no significant decrease in predictive accuracy rises (beyond the naïve Bayes approach) with the enhancement of graphical analysis of variables' dependences - the Tree Augmented Naïve Bayes proved to be an accurate prognostic model, usable in clinical settings, even though it might (as expected) slightly over fit the training cohort (i.e. higher in-sample AUC), losing some generalization ability (i.e. lower quality on cross-validation). Further analysis shall confirm such suspicions using independent cohorts.

V. CONCLUSIONS

The cross-validated evaluation of the Bayesian network classifiers derived in this study resulted in high accuracy and discriminative power for both disabling (above 70%) and reoperation (above 80%) outcomes in Crohn's disease patients subject to early surgery or immunosuppression. The generated models presented interesting insights on factor interaction and predictive ability for the prognosis, supporting their use in future clinical decision support systems.

Current path of research includes the definition of clinically usable decision support tools, based on the hereby derived Bayesian network classifiers and temporal Bayesian networks for the modeling of temporal interdependences useful for mid- and long-term prognosis of Crohn's disease course, taking into account the specific modeling of immunosuppressors intake.

ACKNOWLEDGMENT

The authors thank the investigators of all hospitals who included data for this study, to GEDII - Grupo de Estudo da Doença Intestinal Inflamatória - for all the support, and project NanoStima "NORTE-01-0145-FEDER-000016" which is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

REFERENCES

- [1] G. Latella and C. Papi, "Crucial steps in the natural history of inflammatory bowel disease," *World J. Gastroenterol.*, vol. 18, no. 29, pp. 3790–3799, 2012.
- [2] J. Cosnes, C. Gower-Rousseau, P. Seksik, and A. Cortot, "Epidemiology and natural history of inflammatory bowel diseases," *Gastroenterology*, vol. 140, no. 6, pp. 1785–1794, May 2011.
- [3] D. C. Baumgart and W. J. Sandborn, "Inflammatory bowel disease: clinical aspects and established and evolving therapies," *Lancet*, vol. 369, no. 9573, pp. 1641–1657, 2007.
- [4] S. M. Devlin and R. Panaccione, "Evolving Inflammatory Bowel Disease Treatment Paradigms: Top-Down Versus Step-Up," *Medical Clinics of North America*, vol. 94, no. 1, pp. 1–18, 2010.
- [5] C. C. Dias, P. P. Rodrigues, A. da Costa-Pereira, and F. Magro, "Clinical prognostic factors for disabling Crohn's disease: A systematic review and meta-analysis," *World J. Gastroenterol.*, vol. 19, no. 24, pp. 3866–3871, 2013.
- [6] C. C. Dias, P. P. P. Rodrigues, A. da Costa-Pereira, and F. Magro, "Clinical predictors of colectomy in patients with ulcerative colitis: systematic review and meta-analysis of cohort studies," *J. Crohns. Colitis*, vol. 9, no. 2, pp. 156–163, 2015.
- [7] P. J. F. Lucas, L. C. Van Der Gaag, and A. Abu-Hanna, "Bayesian networks in biomedicine and health-care," *Artif. Intell. Med.*, vol. 30, no. 3, pp. 201–214, 2004.
- [8] C. A. M. Schurink, P. J. F. Lucas, I. M. Hoepelman, and M. J. M. Bonten, "Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units," *Lancet Infect. Dis.*, vol. 5, no. 5, pp. 305–312, 2005.
- [9] P. Lucas, "Bayesian analysis, pattern analysis, and data mining in health care," *Curr. Opin. Crit. Care*, vol. 10, no. 5, pp. 399–403, 2004.
- [10] A. Darwiche, "Bayesian Networks," *Commun. ACM*, vol. 53, no. 12, pp. 80–90, 2010.
- [11] C. C. Dias, C. Granja, A. Costa-Pereira, J. Gama, and P. P. Rodrigues, "Using Probabilistic Graphical Models to Enhance the Prognosis of Health-Related Quality of Life in Adult Survivors of Critical Illness," *2014 IEEE 27th Int. Symp. Comput. Med. Syst.*, pp. 56–61, 2014.
- [12] I. Witten, F. Eibe, and M. A. Hall, *Data Mining- Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, 2011.
- [13] S. Højsgaard, "Graphical independence networks with the gRain package for R," *J. Stat. Softw.*, vol. 46, no. 10, 2012.
- [14] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC*

- [15] C. C. Dias, F. Magro, and P. P. Rodrigues, “Preliminary Study for a Bayesian Network Prognostic Model for Crohn’s Disease,” in

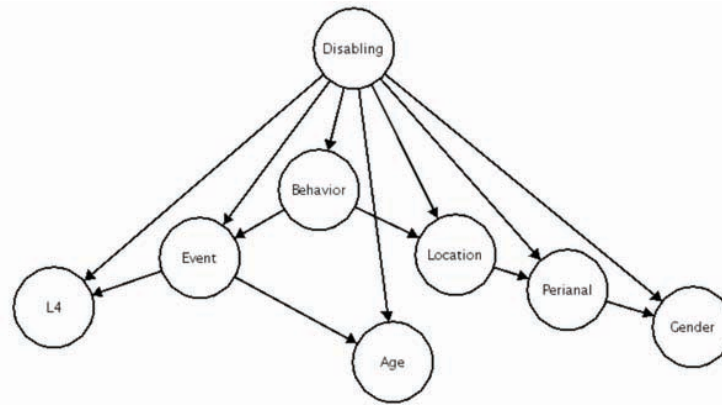


Figure 1. Modeling disabling in Crohn’s disease patients subject to early surgery or immunosuppression – Tree Augmented Naïve Bayes

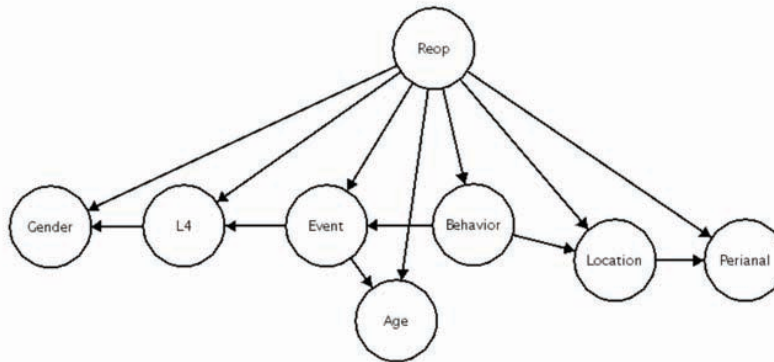


Figure 2. Modeling reoperation in Crohn’s disease patients subject to early surgery or immunosuppression – Tree Augmented Naïve Bayes

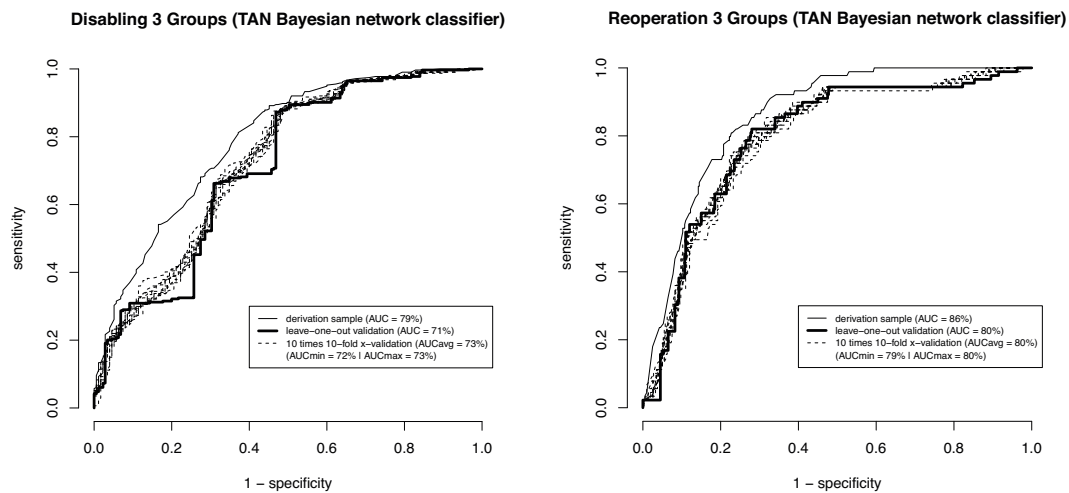


Figure 3. In sample , leave-one-out and 10 cross validation ROC curves of TAN models for disabling and reoperation outcomes.