Journal of Crohn's and Colitis, 2017, 445–453
doi:10.1093/ecco-jcc/jjw171
Advance Access publication September 28, 2016
Original Article

Original Article

Development and Validation of Risk Matrices for Crohn's Disease Outcomes in Patients Who Underwent Early Therapeutic Interventions

Cláudia Camila Dias,[a,b] Pedro Pereira Rodrigues,[a,b] Rosa Coelho,[c] Paula Moura Santos,[d] Samuel Fernandes,[d] Paula Lago,[e] Cidalina Caetano,[e] Ângela Rodrigues,[e] Francisco Portela,[f] Ana Oliveira,[f] Paula Ministro,[g] Eugénia Cancela,[g] Ana Isabel Vieira,[h] Rita Barosa,[h] José Cotter,[i] Pedro Carvalho,[i] Isabelle Cremers,[j] Daniel Trabulo,[j] Paulo Caldeira,[k,l] Artur Antunes,[l] Isadora Rosa,[m] Joana Moleiro,[m] Paula Peixe,[n] Rita Herculano,[n] Raquel Gonçalves,[o] Bruno Gonçalves,[o] Helena Tavares Sousa,[k,p] Luís Contente,[p] Henrique Morna,[q] Susana Lopes,[c] Fernando Magro[c,r,s]; on behalf GEDII

[a]Health Information and Decision Sciences Department, Faculty of Medicine of the University of Porto, Porto, Portugal [b]CINTESIS – Center for Health Technology and Services Research, Porto, Portugal [c]Gastroenterology Department, Hospital São João, Porto, Portugal [d]Gastroenterology Department, Faculty of Medicine, Centro Hospitalar Lisboa Norte, Hospital de Santa Maria, Lisboa, Portugal [e]Gastroenterology Department, Centro Hospitalar do Porto, Porto, Portugal [f]Gastroenterology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal [g]Gastroenterology Department, Centro Hospitalar Tondela e Viseu, Tondela e Viseu, Portugal [h]Gastroenterology Department, Hospital Garcia da Orta, Lisboa, Portugal [i]Gastroenterology Department, Centro Hospitalar do Alto Ave, Guimarães, Portugal [j]Gastroenterology Department, Centro Hospitalar de Setúbal, Hospital São Bernardo, Setúbal, Portugal [k]Department of Biomedical Sciences and Medicine, University of Algarve, Faro, Portugal [l]Gastroenterology Department, Hospital de Faro, Faro, Portugal [m]Instituto Português de Oncologia Francisco Gentil, Lisboa, Portugal [n]Gastroenterology Department, Centro Hospitalar Lisboa Oriental Portugal, Lisboa, Portugal [o]Gastroenterology Department, Hospital de Braga, Braga, Portugal [p]Gastroenterology Department, Portimão Unit, Centro Hospitalar do Algarve, Portimão, Portugal [q]Gastroenterology Department, Hospital Nélio Mendonça, Funchal, Portugal [r]Institute of Pharmacology and Therapeutics Faculty of Medicine of the University of Porto, Porto, Portugal [s]MedInUP – Center for Drug Discovery and Innovative Medicines, University of Porto, Porto, Portugal

Corresponding Author: Fernando Magro, MD, PhD, Institute of Pharmacology and Therapeutics, Faculdade de Medicina da Universidade do Porto, Alameda Prof Hernâni Monteiro, 4200–319 Porto, PORTUGAL. Tel: +351-22-5513642; Fax: +351-22-5513643; Email: fm@med.up.pt

Abstract

Introduction: The establishment of prognostic models for Crohn's disease [CD] is highly desirable, as they have the potential to guide physicians in the decision-making process concerning therapeutic choices, thus improving patients' health and quality of life. Our aim was to derive models for disabling CD and reoperation based solely on clinical/demographic data.

Methods: A multicentric and retrospectively enrolled cohort of CD patients, subject to early surgery or immunosuppression, was analysed in order to build Bayesian network models and

risk matrices. The final results were validated internally and with a multicentric and prospectively enrolled cohort.

Results: The derivation cohort included a total of 489 CD patients [64% with disabling disease and 18% who needed reoperation], while the validation cohort included 129 CD patients with similar outcome proportions. The Bayesian models achieved an area under the curve of 78% for disabling disease and 86% for reoperation. Age at diagnosis, perianal disease, disease aggressiveness and early therapeutic decisions were found to be significant factors, and were used to construct user-friendly matrices depicting the probability of each outcome in patients with various combinations of these factors. The matrices exhibit good performance for the most important criteria: disabling disease positive post-test odds = 8.00 [2.72–23.44] and reoperation negative post-test odds = 0.02 [0.00–0.11].

Conclusions: Clinical and demographical risk factors for disabling CD and reoperation were determined and their impact was quantified by means of risk matrices, which are applicable as bedside clinical tools that can help physicians during therapeutic decisions in early disease management.

Key Words: Disabling disease; reoperation; Crohn's disease; risk matrices

1. Introduction

Crohn's disease [CD] is a chronic and progressive disease of unknown etiology, prone to relapses and disabling events. The clinical course is usually characterized by intermittent relapses, although half of the patients express a mild disease with low propensity to recurrent episodes. On the other hand, the more aggressive cases may require surgery.[1] As a chronic disease, neither treatment nor surgery actually heal the patients, which are usually the subject of frequent medical visits and hospitalizations, creating an aura of uncertainty surrounding their professional and social future that also affects their families.[2,3]

The new concepts on CD treatment are leaving behind the classical approach of controlling the disease symptoms—instead, studies focusing on improvement of quality of life and on the reduction of hospitalizations and surgeries are now emerging. Since the treatment schedule clearly affects the disease course, identifying good prognostic models based on genetic/serological and clinical/demographic factors have been a focus of research. The latter option is more appealing, as it can be supported by data collected during the daily clinical practice.[4,5] The identification of clinical criteria that can predict CD outcomes at an early phase of the disease is therefore crucial, as it can guide the decision-making process on the therapeutic options.

The notion of disabling disease was introduced in 2006 by Beaugerie *et al.*[6] Two years later, Lolly *et al.*[7] presented another study on this subject. Later, in 2011, Yang *et al.*[8] used a similar definition in their study. Currently, and due to the emergence of the above-mentioned new strategies for disease control, there is no consensus regarding the concept of 'disabling' in CD.

The studies seeking to identify factors that can improve the CD outcomes have conflicting conclusions, as the different cohorts yield heterogeneous results due to different methodologies and/or different criteria for patient selection and evaluation. Although the computation of prognostic models can be considered a crucial step towards CD control and management, these are, unfortunately, seldom-addressed topics in the current literature. Moreover, the intricate nature of real-world biomedical data requires the utilization of analyses in which complexity goes beyond traditional biostatistics,[9] without losing the necessary formality.[10,11]

2. Material and Methods

2.1. Derivation and validation cohorts

Data from CD patients being followed by 14 secondary and tertiary care centres was collected in a retrospective fashion [between April and December 2013]. Inclusion criteria were defined as follows: patients aged more than 18 years who had undergone surgery or immunosuppressive therapy in the initial 6 months after diagnosis, and who had at least 3 years of follow-up. All data was collected through a web database, and all missing data or discrepancies were reviewed by the investigators. The study was monitored by the national coordinator of the Portuguese IBD group [GEDII].

The independent validation cohort included patients [from five hospitals] who were enrolled in a prospective fashion. Their data were registered in a national clinical database [gediibasedados.med.up.pt] of IBD patients. Inclusion criteria were similar to those used in the derivation cohort.

Patients from both cohorts were stratified according to the early therapeutic strategies followed by their physicians: Group I patients started immunosuppression during the initial 6 months after diagnosis [index event] and prior to any surgical procedure; Group S_0 patients underwent a surgical intervention [index event] in the initial 6 months after diagnosis and had no immunosuppression therapy during the known follow-up period; and Group S_1 patients underwent a surgical intervention [index event] in the initial 6 months after diagnosis and started immunosuppression within the 6 months after surgery.

2.2. Clinical and demographic variables

Besides the basic demographic data, clinical information was collected for each patient, including the Montreal classification[12] and the follow-up data [total number of surgeries and hospitalizations; treatment, namely corticoids, immunosuppression or anti-TNF, and disease adverse events—stenosis, abscess, perforation and anal disease]. Regarding Montreal classification, patients were classified according to disease extent, behaviour and age. Steroid resistance was defined as the presence of active disease despite a prednisolone dose of up to 0.75 mg/kg per day over a period of 4 weeks.[13,14]

2.3. Outcomes analysed

The primary outcome of this study was the occurrence of disabling disease, whereas reoperation was studied as a secondary outcome. We have defined disabling disease as the occurrence of at least one of the following events: one or more surgeries in the first 5 years after diagnosis [excluding the index surgery, if applicable]; more than one surgery during follow-up [also excluding the index surgery, if applicable]; more than two hospitalizations [excluding the index episode and hospitalization for infliximab infusion]; at least two steroid course requirements per year, steroid dependency and steroid refractoriness; need to switch immunosuppression [AZA or MTX], and anti-TNF drugs [infliximab or adalimumab]; new events such as stenosis, penetrating disease or anal disease. This definition of disabling was determined taking into account the following aspects: [1] the introduction of immunosuppression and anti-TNF per se were not interpreted as disabling disease based on the development of new therapeutic strategies; [2] due to low efficacy of 5-ASA in CD patients, many newly diagnosed patients were treated with immunosuppression and anti-TNF; [3] switching treatments was seen as disabling and [4] events after the index event were interpreted as disabling due to signs of disease progression.

2.4. Statistical analyses

Prognostic models were defined by means of Bayesian networks [BNs] built over the set of available variables.[15] Bayesian networks can be seen as an alternative to logistic regression, where statistical dependence and independence are not hidden in approximating weights, but rather explicitly represented by links in a network of variables.[11] Generally, a Bayesian network represents a joint distribution of one set of variables, specifying the assumption of independence between them, with the interdependence between variables being represented by a directed acyclic graph. Each variable is represented by a node in the graph, and is dependent on the set of variables represented by its ascendant nodes.[16] This dependence is represented by a conditional probability table that describes the probability distribution of each variable, given their ascendant variables. The Tree Augmented Naïve Bayes [TAN] classifier model, used in this study, includes two assumptions: [1] all explanatory variables are conditioned by the outcome, i.e. all will directly influence the outcome during inference; and [2] an optional additional dependence is allowed for each variable, i.e. each variable's effect might be adjusted by one additional covariable.[17] TAN classifiers were built from the derivation cohort. Model parameters were validated by comparing the AUC in the derivation cohort with those calculated from a leave-one-out and a 10 times 2-fold cross-validation [for variability assessment with independent training and testing], and by an independent comparable validation cohort.

The application of the prognostic models generated in this work can be visualized by means of [a] an online tool for direct BN inference [in beta testing phase], and [b] appropriately defined risk matrices. In order to choose which variables should be included in the risk matrices, we applied a logistic regression with all independent variables using the enter method. Variables with statistical significance [or with clinical relevance for the course of the disease, e.g. age at diagnosis[4]] were chosen as factors for the matrices. Each cell of the matrices represents the marginal posterior outcome probability estimate for that subgroup of patients. The precision of such estimates is given by a 95% credible interval, computed from a Monte Carlo simulation of one million samples from the derived joint probability model [i.e. the BN].[18] The risk values in each cell of the matrix represent the expected risk for a patient in that subgroup, while the credible interval encloses 95% of risk estimates for patients in that subgroup [i.e. only 5% of patients in that subgroup have a risk estimate outside the credible interval]. We believe that this approach is more interesting from the clinical point of view than the usual one, in which a confidence interval [CI] of the expected risk of all patients in each subgroup is computed and presented. Since patients with an early surgery [Group S] may have either engaged in immunosuppression therapy or not [S_0 *vs.* S_I], and in order to have an accurate prognosis for these patients, the posterior probability P[D|S] and the corresponding credible intervals were computed for Groups S_0 and S_I, taking into account the probability of needing immunosuppressive therapy after surgery [Table 1]. The cell risk matrices for Group S patients were calculated as follows:

$$P(D|S) = P(D|S_I) \times P(I|S) + P(D|S_0) \times P(\sim I|S).$$

To assess the discriminative ability of the risk matrices for each outcome, specific cut-off values were chosen after performing a ROC analysis of the derivation cohort. For disabling disease, and taking into account its expected prevalence and the impact of a positive prediction, a rule-in approach was applied aiming at a high positive predictive value [~80%]. For reoperation, and taking into account its expected prevalence and the impact of a negative prediction, a rule-out approach was applied aiming at a high negative predictive value [also ~80%]. The thus derived decision rules were then evaluated on both cohorts, estimating sensitivity, specificity, accuracy, the predictive values, likelihood ratios, and the post-test odds.

Logistic regression was applied with IBM SPSS v23.0, BN structures were created with WEKA software,[19] posterior probabilities were inspected using SamIam software[20]; the exact inference procedures for validation and risk matrices definition were available in R package gRain[21] using Lauritzen-Spiegelhalter algorithm,[22] while ROC curves and corresponding AUCs were computed using R package pROC.[23]

3. Results

3.1. Population characteristics and analysed outcomes

The derivation cohort analysed in this study consisted of 489 CD patients, of which 46% were male, and 79% were 40 years old or younger when diagnosed [Table 2]. Most of the patients had either an ileal or an ileocolonic location of the disease, 12% had upper tract involvement, and 26% had perianal disease. Disabling disease was observed in 64% [CI 95%: 60–68%] of the enrolled patients, while 18% [CI 95%: 15–21%] needed reoperation [i.e. more than one surgery].

Table 1. Probability of immunosuppression after surgery: P[I|S].

	Perianal disease					
	No			Yes		
	Behaviour					
	B1	B2	B3	B1	B2	B3
Age at diagnosis						
≤40	35%	80%	75%	71%	71%	83%
>40	43%	47%	67%	68%	68%	68%

I = Immunosuppression; S = Surgery.

Table 2. Characteristics of the derivation cohort enrolled in this study [*n* = 489].

	Total [*n* = 489]		Group S_0 [*n* = 80]		Group S_I [*n* = 175]		Group I [*n* = 234]		*p*-value[1]
	n	[%]	*n*	[%]	*n*	[%]	*n*	[%]	
Gender									0.914
Male	225	[46%]	36	[45%]	79	[45%]	110	[47%]	
Location									<0.001
L1. Ileal	232	[47%]	55	[69%]	95	[54%]	82	[35%]	
L2. Colonic	45	[10%]	1	[1%]	4	[2%]	40	[17%]	
L3. IleoColonic	212	[43%]	24	[30%]	76	[43%]	112	[48%]	
Upper tract involvement [L4]									<0.001
Yes	55	[12%]	2	[2%]	11	[6%]	42	[18%]	
Disease behaviour									<0.001
B1. Non-structuring/non-penetrating	158	[32%]	17	[21%]	14	[8%]	127	[54%]	
B2. Structuring	176	[36%]	35	[44%]	80	[46%]	61	[26%]	
B3. Penetrating	155	[32%]	28	[35%]	81	[46%]	46	[20%]	
Perianal disease									0.001
Yes	125	[26%]	11	[14%]	38	[22%]	76	[32%]	
Age at diagnosis									<0.001
≤40 years	388	[79%]	48	[60%]	132	[75%]	208	[89%]	
>40 years	101	[21%]	32	[40%]	43	[25%]	26	[11%]	
Follow-up time, median [IQR]	9[6–14]		13 [8–19]		13 [8–19]		6 [5–10]		<0.001
Disabling disease	314	[64%]	15	[19%]	137	[78%]	162	[69%]	<0.001
CI 95%	[60–68%]		[10–28%]		[72–84%]		[63–75%]		
Reoperation	89	[18%]	12	[15%]	70	[40%]	7	[3%]	<0.001
CI 95%	[15–21%]		[7–23%]		[33–47%]		[0.8–5%]		
Total of surgeries [including index]									–
None	132	[31%]	–		–		132	[78%]	
1	204	[48%]	68	[85%]	105	[60%]	31	[18%]	
2	52	[12%]	9	[11%]	39	[22%]	4	[2%]	
3	22	[5%]	1	[1%]	18	[10%]	3	[2%]	
4	10	[2%]	1	[1%]	9	[5%]	0	[0%]	
5	3	[1%]	1	[1%]	2	[1%]	0	[0%]	
6	1	[1%]	0	[0%]	1	[1%]	0	[0%]	
9	2	[1%]	0	[0%]	1	[1%]	1	[1%]	
Required anti-TNF	176	[37%]	0	[0%]	74	[42%]	102	[45%]	<0.001
Steroids	318	[67%]	21	[28%]	105	[63%]	192	[83%]	<0.001
1 cycle/years	57	[29%]	6	[46%]	18	[28%]	33	[28%]	0.670
≥2 cycles/year	11	[6%]	0	[0%]	5	[8%]	6	[5%]	
1 cycle for each 3 years	70	[36%]	6	[46%]	24	[37%]	40	[34%]	
Steroid dependent	46	[24%]	1	[8%]	14	[22%]	31	[27%]	
Steroid resistant	11	[6%]	0	[0%]	4	[6%]	7	[6%]	
New events									
Stenosis	74	[15%]	6	[8%]	43	[25%]	25	[11%]	<0.001
Penetrating disease	5	[1%]	2	[3%]	3	[2%]	0	[0%]	0.089
Anal disease	51	[10%]	4	[4%]	25	[14%]	23	[10%]	0.036

[1] Chi-Squared test; IQR: interquartile range; 95% CI: 95% confidence interval.

This cohort was stratified according to the early therapeutic strategies followed for each patient: Group I consisted of patients who started immunosuppression in the initial six months after diagnosis and prior to any surgical procedure, and Group S consisted of patients who underwent surgery in the initial 6 months following diagnosis. Group S was further divided into patients who started immunosuppression 6 months after the initial surgery [S_I], and patients who did not follow any immunosuppression therapy during the follow-up time considered in this study [S_0]. Eighty [16%] patients were in Group S_0, 175 [36%] in Group S_I and 234 [48%] in Group I, and significant differences were observed between these groups for all variables accounted for in this study, with the exception of gender [Table 2]. Concerning outcome, disabling disease occurred most frequently among S_I patients [78% of S_I, 69% of I and 19% of S_0, $p < 0.001$], as did the need for reoperation [40% of S_I, 3% of I and 15% of S_0, $p < 0.001$].

3.2. Bayesian prognostic models and relevant risk factors

In order to unveil the interdependent relationships between the analysed CD outcomes and the variables considered, BN-based models were built for the presence of disabling disease and the need for reoperation [Figure 1]. In both cases, patient group was associated [arc between variables] with disease behaviour, upper gastrointestinal tract location [L4], and age at diagnosis, while location was associated with perianal disease. For disabling disease, an association between perianal disease and gender was also found, whereas an association between gender and upper tract involvement was found for reoperation.

To determine which of the factors listed above were significant and should be included in risk matrices, a logistic regression was carried out using all the independent variables considered. Those that were statistically significant and those that had been shown by previous

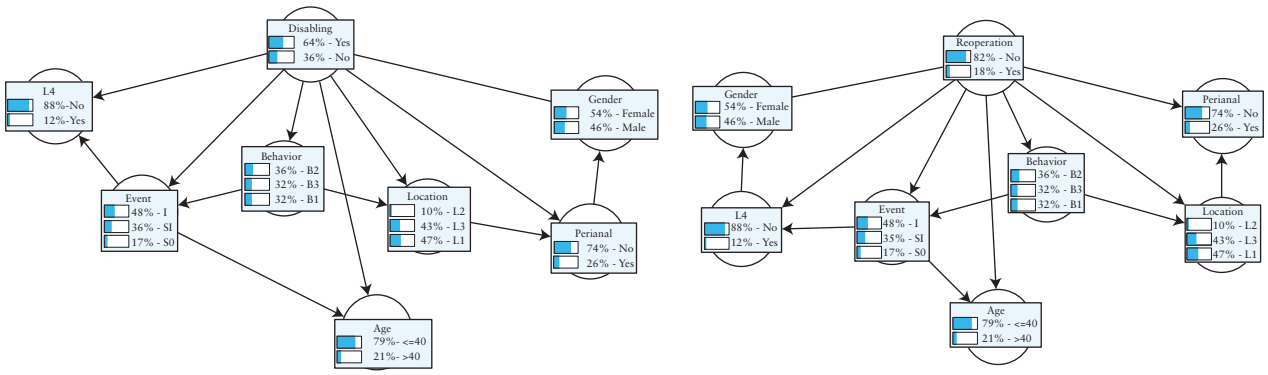

Figure 1. Bayesian network representing the relationships between each outcome [disabling disease and reoperation] and each demographic and clinical variables, and relationships between predictive factors. The bars within each variable represent the prior marginal probabilities for each variable's category. Arrows represent association between variables, but do not convey any causal relationship, the association between the outcomes and each of the remaining variables being imposed on the model.

studies to be important for the outcomes analysed[4] were selected and included in the final matrices as risk factors: for disabling disease—age at diagnosis [knowledge from literature], behaviour phenotype (B2: OR = 2.359 [1.318–4.22], B3: OR = 2.926 [1.567–5.463]), perianal disease (OR:4.340 [2.399–7.852]), and the therapeutic-defined patient group (S_I: OR=15.134 [7.483–30.607] and I: OR = 12.797 [6.128–26.723])—and for reoperation age at diagnosis (A3: OR = 0.364 [0.184–0.723]), behaviour phenotype (B2: OR = 6.443 [2.054–20.206], B3: OR = 4.815 [1.525–15.201]), perianal disease (OR: 2.189 [1.144–4.187]) and therapeutic-defined patient group (S_I: OR = 3.080 [1.500–6.325] and I: OR = 0.204 [0.073–0.571]).

3.3. Risk matrices

The risk matrices convey the risk of each outcome [disabling disease and reoperation] stratified by the relevant factors. In the particular case of this study, one of the factors considered to be relevant is the patient's group according to the initial therapeutic approaches. In this regard, it should be noticed that two of the groups are intimately linked: in fact, 52% of the patients [n = 255] were submitted to a surgical intervention during the initial six months after diagnosis, but only a fraction of those [31%] did not enter into immunosuppressive therapy afterwards, yielding Groups S_I and S_0. Therefore, and in order to accurately depict the risk in patients who undertook surgery but have not yet started immunosuppression, one has to take into account the probability of a patient starting immunosuppression after the initial surgery, i.e. p [I|S]. Such probabilities are shown in Table 1, stratified by the other three risk factors considered to be relevant [age at diagnosis, presence of perianal disease, and disease phenotype]. The highest value [83%] was observed for patients who were 40 years old or younger at diagnosis, had perianal disease, and a penetrating disease behaviour [B3]. The lowest value [35%] was observed for patients who were 40 or younger at diagnosis, had no perianal disease, and presented the least aggressive phenotype [B1]. These results were accounted for in the construction of the final risk matrices.

3.3.1. Disabling disease

The risk for CD patients of facing disabling disease, taking into consideration the relevant factors determined previously, is stated in Table 3. Patients undergoing early surgery [Group S] had a lower probability of facing disabling disease than patients in the other two groups. For the other three risk factors considered, one could detect an increased risk when patients were 40 years old or younger at diagnosis, had perianal disease, and a penetrating disease behaviour [B3]. The highest risk was observed for patients included in Group S_I, with perianal disease, penetrating disease behaviour [B3], and older than 40 years at diagnosis: 94% [88–98%]. The lowest risk was observed for Group S patients, who had no perianal disease, non-structuring/non-penetrating disease behaviour [B1], and who were 40 or younger at diagnosis: 27% [21–36%].

3.3.2. Reoperation

The risk for CD patients of undergoing more than one surgery [reoperation] during the course of the disease, taking into consideration the relevant factors determined previously, is stated in Table 4. Patients who had an early start of immunosuppression therapy had a lower probability of reoperation than those who had an early surgery. Moreover, younger age at diagnosis and perianal disease increased the risk of reoperation, whereas a non-structuring/non-penetrating [B1] disease behaviour decreased it. The lowest probability of reoperation was observed for patients in Group I who were older than 40 years at diagnosis, had no perianal disease, and a non-structuring/non-penetrating [B1] disease behaviour: 0.4% [0.2–0.9%]. The highest probability of reoperation was observed for patients who had undergone an early surgery and later entered into an immunosuppression therapy [Group S_I], were diagnosed before or at 40, and had perianal disease and a structuring [B2] disease behaviour: 54% [26–73%].

3.4. Model validation

The Bayesian prognostic models and the resulting risk matrices were validated following two different approaches: an internal one, which consisted of two different tests [leave-one-out and ten times 2-fold cross-validation]; and an external one, following the analysis of a prospectively recruited validation cohort. ROC analyses were performed independently for the derivation cohort and for each of the validation sets of data, and the respective AUCs, along with their 95% CIs, are illustrated in Figure 2. As was desirable, the AUC values of the validation cohort nearly overlapped those of the derivation cohort, and the AUC values of the generated sets of data for the internal validation were rather similar to the later ones. Furthermore, the overall discrimination power was high for both outcomes [78% for disabling disease and 86% for reoperation, using the derivation cohort]. Based on this, the following cut-offs were

Table 3. Risk matrix showing the probability [%] of having disabling disease during the course of the disease.

		Perianal Disease						
		No			Yes			
		B1	B2	B3	B1	B2	B3	
Age at diagnosis	≤40	27% [21–36%]	60% [57–66%]	64% [62–65%]	72% [59–78%]	77% [67–87%]	85% [74–94%]	Group S
	>40	31% [17–39%]	43% [36–50%]	60% [49–62%]	66% [50–73%]	73% [58–83%]	76% [59–87%]	
	≤40	55% [43–69%]	71% [67–79%]	77% [75–80%]	82% [75–90%]	90% [83–96%]	92% [84–97%]	Group I
	>40	45% [34–60%]	62% [58–72%]	69% [67–72%]	75% [67–86%]	86% [77–94%]	88% [78–96%]	
	≤40	53% [42–66%]	71% [67–79%]	76% [74–77%]	81% [74–89%]	90% [83–96%]	91% [83–97%]	Group S_I
	>40	63% [52–75%]	78% [75–84%]	82% [81–84%]	86% [81–92%]	93% [88–98%]	94% [88–98%]	

Patient's therapeutic group: S = surgery; S_I = surgery and immunosuppression 6 months after surgery; I = immunosuppression.

Reading example: A patient 40 years old or younger at diagnosis, without perianal disease, and disease phenotype B1 had a probability of disabling disease ranging between 27% (if he or she had a surgery [Group S]) and 55% [if in the first 6 months after diagnosis he or she had immunosuppression]. But if the patient had perianal disease, the probability of disabling disease increased: 72% with a surgery in the first 6 months after diagnosis [Group S] and 82% if immunosuppression [Group I] was conducted.

Colour scheme: White = ≤10%; green = 11–19%; yellow = 20–49%; orange = 50%-74%; red = ≥75%.

Table 4. Risk matrix showing probability [%] of reoperation during the course of the disease.

		Perianal Disease						
		No			Yes			
		B1	B2	B3	B1	B2	B3	
Age at diagnosis	≤40	9% [6–13%]	42% [36–51%]	37% [29–51%]	22% [10–32%]	44% [23–50%]	48% [23–36%]	Group S
	>40	6% [3–8%]	19% [15–26%]	20% [15–31%]	11% [4–17%]	26% [9–44%]	28% [13–36%]	
	≤40	1% [0.5–2%]	9% [5–14%]	3% [2–6%]	1% [0.5–3%]	11% [3–24%]	4% [2–7%]	Group I
	>40	0.4% [0.2–0.9%]	5% [3–8%]	2% [1–3%]	0.7% [0.2–1%]	6% [2–14%]	2% [1–4%]	
	≤40	21% [14–32%]	49% [42–58%]	41% [32–56%]	30% [12–45%]	54% [26–73%]	51% [28–59%]	Group S_I
	>40	9% [6–15%]	26% [21–34%]	20% [15–32%]	14% [5–23%]	31% [11–50%]	28% [13–35%]	

Patient's therapeutic group: S = surgery; S_I = surgery and immunosuppression 6 months after surgery; I = immunosuppression.

Reading example: A patient 40 years old or younger at diagnosis, without perianal disease, and B1 phenotype had the probability of reoperation ranging between 1% (if he or she had immunosuppression in the first 6 months after diagnosis [Group I]) and 21% (if he or she had surgery in the first 6 months after diagnosis and immunosuppression thereafter [Group S_I]). But if the patient had perianal disease, the probability of reoperation increased to 30% if a surgery occurred during the first 6 months after diagnosis and the patitent had immunosuppression 6 months after surgery [Group S_I].

Colour scheme: White = ≤10%; green = 11–19%; yellow = 20–49%; orange = 50%-74%; red = ≥75%.

determined: values above 75% [for disabling disease] and above 19% [for reoperation] were considered to be positive test results, i.e. to predict the occurrence of the respective outcome.

Table 5 presents the performance of the chosen cut-offs for each outcome in the derivation and validation cohorts. The CIs for each of the performance measures computed overlapped between the two cohorts [with only two exceptions], further validating our model. Overall, the application of the cut-offs to the validation cohort resulted in 94 % [83–98%] specificity and 89% [70–97%] PPV for disabling disease, and 96% [78–100%] sensitivity and 98% [90–99%] NPV for reoperation.

3.5. Derivation vs. validation cohorts

The validation cohort consisted of 129 patients who were prospectively enrolled in this study, and whose demographic and clinical characteristics are depicted in Table 6. The derivation and validation cohorts were shown to be similar for all variables analysed, with the exception of the presence of perianal disease [26% in the derivation cohort vs. 13% in the validation cohort, $p = 0.003$] and stratification in the three different patients' groups considered [$p < 0.001$]. To exclude the hypothesis that these differences could significantly impact the validation analysis, a new set of comparisons was carried out: the performance of the determined cut-offs was computed for the validation cohort stratified according to the presence of perianal disease [see Supplementary Table S1] and the patient's group [see Supplementary Table S2]. The performance measures were similar [i.e. the 95% CIs overlap], with a few noted exceptions that occurred in performance measures that were less relevant for the corresponding outcome.

4. Discussion

Because CD is a disabling disease that has a significant negative impact on the patient's quality of life, the construction and validation of predictive models that can anticipate negative outcomes is a cornerstone for preventive therapeutics, allowing physicians to adjust the medication in a prophylactic fashion, instead of doing so as a reaction to a flare. A few attempts to analyse risk factors from a prognosis perspective have been done in the past, but they usually involved results from genetic or serologic tests,[24,25] which are both expensive and time-consuming. Our study is, to our best knowledge, the first one to build and validate risk models for CD outcomes—disabling disease and reoperation—based solely on clinical and demographic variables, which have the key advantage of

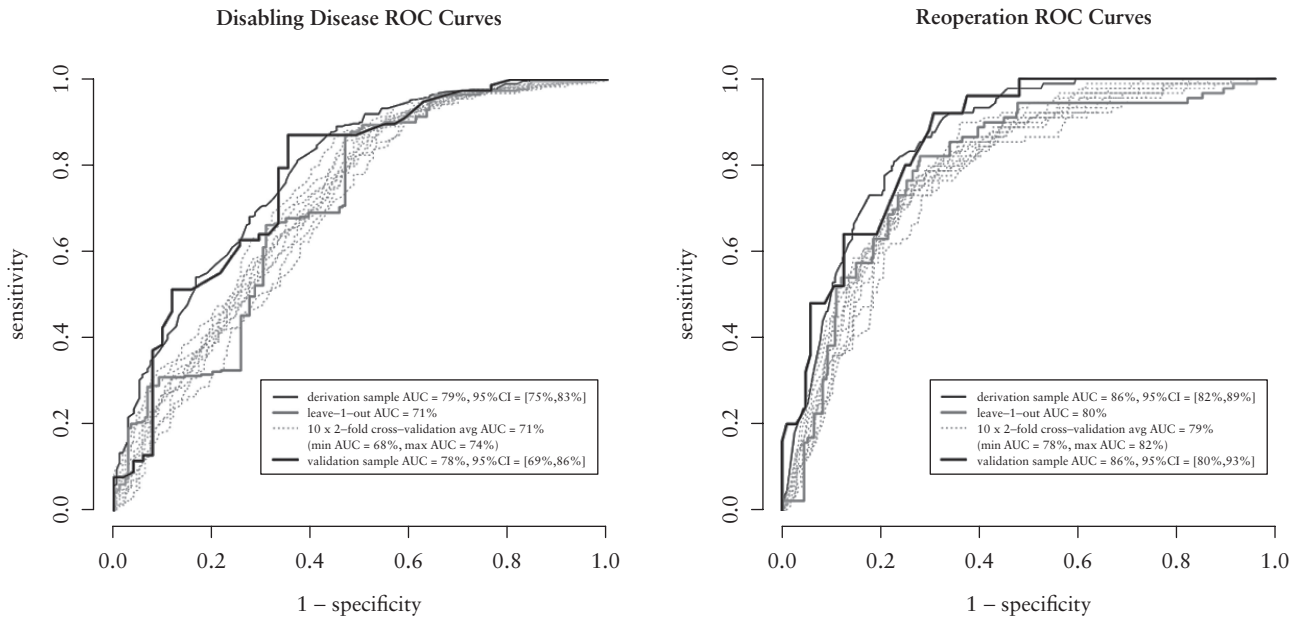

Figure 2. Receiver operating characteristic analyses and area under the curve values for the outcomes of disabling disease and reoperation in the derivation and validation cohorts, as well as for the internal validation procedures.

Table 5. Performance of risk matrix in derivation and validation cohort for disabling disease and reoperation (% [CI 95%]).

	Disabling [>75%]		Reoperation [>19%]	
	Derivation	Validation	Derivation	Validation
Sens	53% [47–58%]	31% [21–42%]	91% [85–97%]	96% [78–100%]
Spec	74% [68–80%]	94% [83–98%]	66% [61–71%]	61% [50–70%]
PPV	78% [73–83%]	89% [70–97%]	37% [31–43%]	37% [26–49%]
NPV	47% [41–53%]	47% [37–57%]	97% [95–99%]	98% [90–99%]
Accuracy	60% [55–64%]	55% [46–64%]	70% [66–74%]	67% [59–75%]
LR+	2.04 [1.55–2.68]	5.23 [1.66–16.47]	2.67 [2.29–3.09]	2.43 [1.89–3.13]
LR–	0.63 [0.57–0.72]	0.73 [0.63–0.85]	0.14 [0.07–0.27]	0.07 [0.01–0.46]
Odds post test+	3.67 [2.80–4.80]	8.00 [2.72–23.44]	0.59 [0.48–0.72]	0.58 [0.41–0.84]
Odds post test–	1.14 [1.02–1.29]	1.13 [0.91–1.38]	0.03 [0.01–0.06]	0.02 [0.00–0.11]

Sens: sensibility; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR: negative likelihood ratio.

being easily and quickly acquired. The final results were arranged into color-coded and user-friendly matrices that constitute a preliminary but useful tool that can be used by physicians in the therapeutic decision-making process.

The overall disabling rate was 64%, ranging from 19% in Group S_0 patients to 78% for Group S_I patients, values that are lower than those previously published.[6,8] However, our definition of disabling disease was stricter than that used in those studies, as variables such as 'need for immunosuppression' or 'need for anti-TNF' were not interpreted as disabling. In fact, the notion of disabling disease published by Beaugerie *et al.* has suffered a few modifications with the advent of new therapeutic strategies, namely the acceleration of step-up and the implementation of top-down treatments. We have taken these factors into account, and adjusted the concept of disabling disease to the present context. Moreover, one should keep in mind that the concept of disabling is different from that of disability. In fact, the disability index published by Peyrin-Biroulet *et al.*[26] is a multidimensional approach to disabling events, encompassing body function and structure, activities, participation, and environmental factors. As our main aim was to predict the disabling events from the clinical and demographic context of the patient, we have chosen to use the concept of disabling instead of the disability index mentioned above.

The need for more than one surgery [reoperation] occurred in 18% of all patients, a value that ranged from 40% in Group S_I to 3% in Group I. In our study, a cumulative risk of reoperation for Group S_0 ranged between 1% and 15% at 5 and 30 years of follow-up, respectively. For patients in Group S_I, the probability ranged from 0% to 41% for 5 and 30 years of follow-up, while for Group I the cumulative probability ranged between 0% and 3% for 5 and 30 years of follow-up, respectively. These values were lower than those found by Frolkis *et al.*,[27] but our cohort had a subgroup of patients who had a more aggressive treatment from diagnosis, namely immunosuppression within 6 months after diagnosis. The ability to anticipate flares and to establish preventive therapeutic strategies are invaluable steps for better disease management and a rational utilization of the available resources, especially in the case of a highly disabling and chronic disease such as CD. In this context, the Pocer study was a seminal work, establishing that retreatment according to the clinical risk of recurrence, with an early colonoscopy and treatment step-up if needed, was better than conventional drug therapy alone for the prevention of postoperative CD

Table 6. Comparison between derivation and validation cohort.

	Derivation [*n* = 489]		Validation [*n* = 129]		*p*-value
	n	[%]	*n*	[%]	
Gender					0.831
Male	225	[46%]	58	[45%]	
Location					0.817
L1. Ileal	232	[47%]	64	[50%]	
L2. Colonic	45	[9%]	13	[10%]	
L3. IleoColonic	212	[43%]	52	[40%]	
Upper tract involvement [L4]					0.714
Yes	55	[11%]	16	[12%]	
Disease behaviour					0.266
B1. Non-structuring /non-penetrating	158	[32%]	49	[38%]	
B2. Structuring	176	[36%]	37	[29%]	
B3. Penetrating	155	[32%]	43	[33%]	
Perianal disease					**0.003**
Yes	125	[26%]	17	[13%]	
Age at diagnosis					0.945
≤40 years	388	[79%]	1028	[79%]	
>40 years	101	[21%]	27	[21%]	
Disabling disease	314	[64%]	78	[61%]	0.432
Reoperation	89	[18%]	25	[19%]	0.759
Patients group					**<0.001**
S_0	80	[17%]	53	[41%]	
S_I	175	[35%]	30	[23%]	
I	234	[47%]	46	[36%]	

recurrence.[28] The rationale for our study was different, and so were the variables and the outcomes analysed, namely by including the impact of early strategies [immunosuppression and/or surgery] in the CD development. However, it is our belief that combining the results of an early endoscopic examination—after the Pocer study—with our risk matrices could greatly improve the predictive ability and increase the chances of effectively adjusting the medication in a preventive fashion.

The variables considered in the risk matrices were included after a significant result in logistic regression analysis [behaviour, patient group, and perianal disease] or because they have been previously described as predictive factors [age at diagnosis].[4,6,8] The disabling disease risk matrix showed that patients who were 40 years or less at diagnosis, have perianal disease, an aggressive disease phenotype, and who are medicated with immunosuppressors [either before any surgery or after an initial one] have a higher risk of undergoing the disabling events considered. The need for a reoperation, on the other hand, is more likely to occur in patients diagnosed at 40 or less, who have perianal disease, an aggressive disease phenotype, and who underwent an early surgery upon diagnosis. These results are in line with past research, supporting the accuracy of our matrices.[4] The embedded models' discriminative power ranges between 78% for disabling disease and 86% for reoperation. These values are higher than those presented by Siegel *et al.*,[23] a study that involved more variables, some of which of a genetic nature; therefore, our study opens new windows of opportunity for simpler yet usable models in clinical practice. The cut-offs established in our study for detecting disabling disease and reoperation [75% and 19%, respectively] were shown to have quite good performance, presenting a positive post-test odds of 8.00 [2.72–23.44] for disabling disease, and a negative post-test odds of 0.02 [0.00–0.11] for reoperation.

The BMs and risk matrices built in this study were validated using internal leave-one-out and crossed validations, as well as an independent and prospectively recruited validation cohort. The ROC curves for disabling disease and for need for reoperation were similar for the derivation and the validation cohorts. The differences between the derivation and the validation cohorts—present in the occurrence of perianal disease and formation of groups based on early therapeutic decisions—were thoroughly investigated. Whereas differences regarding the occurrence of perianal disease did not seem to impact the results, those concerning unbalanced groups of patients may convey a slight but undefined bias in the overall predictive quality of the models. In fact, in the derivation cohort, Group S_I had a lower discriminative ability for reoperation, whereas Group I had a higher discriminative ability for the same outcome, the undefined nature of bias coming from both the groups having had a lower frequency in the validation cohort when compared with the derivation cohort.

This study had a few limitations that we acknowledge as follows: the retrospective nature of the derivation cohort, the inability to provide a close set of treatment strategies for all patients [due to lack of data], the smaller follow-up time for Group I, and the absence of data regarding patients' smoking habits. Nevertheless, we believe we have limited these drawbacks by: [i] monitoring the inclusion with a data entry monitor; [ii] precisely defining the inclusion and exclusion criteria before study start; and [iii] building a web platform for study purposes that automatically sends missing data reports. Regarding smoking status, the lack of this variable in the case report of many patients prevented its use. However, we have tested the models, including the smoking status whenever possible, and the results were similar to those we have obtained without this variable [data not shown].

As a global conclusion, we may say that our study added important knowledge to the state of the art regarding CD development—namely unveiling important risk predictive factors, including the impact of early therapeutic strategies in the disease development. Moreover, that knowledge is delivered in the form of an intuitive and user-friendly bedside predictive tool, which can be used by any

physician to quantify the likelihood of disabling events or the need for reoperation. The fact that no genetic or serologic tests results are included allows an immediate reading of this tool, allowing the early adjustment of medication and contributing to a prophylactic approach concerning CD's negative outcomes.

Funding

This work was funded by GEDII – Grupo de Estudo da Doença Inflamatória Intestinal.

Conflict of Interest

Fernando Magro received a fee for presenting from: AbbVie, Ferring, Falk, Hospira, PharmaKern, MSD, Schering, Lab. Vitoria, Vifor, and OmPharma.

Acknowledgments

The authors thank all investigators at the hospitals who provided data for the EASY study, GEDII – Grupo de Estudo da Doença Inflamatória Intestinal – for all their support, and Sandra Dias for all her assistance during the data collection. This work was partially developed under the scope of 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]. The authors also acknowledge the help of Raphael Oliveira in the development of the online inference tool, and Catarina L. Santos for medical writing assistance.

Author Contributions

Dias CC was involved in the conception and design of the study, acquisition, analysis and interpretation of data, and was responsible for drafting the manuscript. Rodrigues PP was involved in the design, analysis and interpretation of data and drafting the manuscript. Coelho R was involved in interpretation of data and drafting the manuscript. Magro F was involved in the conception and design of the study, interpretation of data, and drafting and revising the manuscript. All other authors were responsible for data collection. All authors read and approved the final version of the manuscript.

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

1. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* 2011;**140**(6):1785–94.
2. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007;**369**(9573):1641–57.
3. Devlin SM, Panaccione R. Evolving inflammatory bowel disease treatment paradigms: top-down versus step-up. *Med Clin North Am* 2010;**94**(1):1–18.
4. Dias CC, Rodrigues PP, da Costa-Pereira A, Magro F. Clinical prognostic factors for disabling Crohn's disease: a systematic review and meta-analysis. *World J Gastroenterol* 2013;**19**(24):3866–71.
5. Dias CC, Rodrigues PPP, Costa-Pereira A da, Magro F. Clinical predictors of colectomy in patients with ulcerative colitis: systematic review and meta-analysis of cohort studies. *J Crohns Colitis* 2015;**9**(2):156–63.
6. Beaugerie L, Seksik P, Nion-Larmurier I, Gendre J-P, Cosnes J. Predictors of Crohn's disease. *Gastroenterology* 2006;**130**(3):650–6.
7. Loly C, Belaiche J, Louis E. Predictors of severe Crohn's disease. *Scand J Gastroenterol* 2008;**43**(8):948–54.
8. Yang CH, Ding J, Gao Y, Chen X, Yang Z Bin, Xiao SD. Risk factors that predict the requirement of aggressive therapy among Chinese patients with Crohn's disease. *J Dig Dis* 2011;**12**(2):99–104.
9. Lucas PJF, Van Der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. *Artif Intell Med* 2004;**30**(3):201–14.
10. Schurink CAM, Lucas PJF, Hoepelman IM, Bonten MJM. Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units. *Lancet Infect Dis* 2005;**5**(5):305–12.
11. Lucas P. Bayesian analysis, pattern analysis, and data mining in health care. *Curr Opin Crit Care* 2004;**10**(5):399–403.
12. Satsangi J, Silverberg MS, Vermeire S, Colombel J-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;**55**(6):749–53.
13. Van Assche G, Dignass A, Panes J, *et al*. The second European evidence-based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *J Crohns Colitis* 2010;**4**(1):7–27.
14. Dignass A, Van Assche G, Lindsay JO, *et al*. The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: current management. *J Crohns Colitis* 2010;**4**(1):28–62.
15. Darwiche A. Bayesian Networks. *Commun ACM* 2010;**53**(12):80–90.
16. Mitchell T. *Machine Learning*. New York, NY: McGraw-Hill; 1997.
17. Huang K, King I, Lyu M. *Constructing a large node Chow-liu tree based on frequent itemsets*. In: Proceedings of the 9th International Conference on Neural Information Processing; November 18–22, Singapore, 2002; 498–502.
18. Lecoute B, Poitevineau J. *The Significance Test Controversy Revisited—the Fiducial Bayesian Alternative*. New York, NY: Springer; 2014.
19. Witten I, Frank E, Hall MA. *Data Mining—Practical Machine Learning Tools and Technique*. 3rd edn. Burlington, MA: Morgan Kaufmann; 2011.
20. Darwiche A. *Modeling and Reasoning with Bayesian Networks*. Cambridge, UK: Cambridge University Press; 2009.
21. Højsgaard S. Graphical independence networks with the gRain package for R. *J Stat Softw* 2012;**46**(10):1–26.
22. Lauritzen, S. L. and Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. *J R Stat Soc Ser B* 1988;**50**(2):157–224.
23. Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77.
24. Beaugerie L, Sokol H. Clinical, serological and genetic predictors of inflammatory bowel disease course. *World J Gastroenterol* 2012;**18**(29):3806–13.
25. Louis E, Belaiche J, Reenaers C. Do clinical factors help to predict disease course in inflammatory bowel disease? *World J Gastroenterol* 2010;**16**(21):2600–3.
26. Peyrin-Biroulet L, Cieza A, Sandborn WJ, *et al*. Development of the first disability index for inflammatory bowel disease based on the international classification of functioning, disability and health. *Gut* 2012;**61**(2):241–7.
27. Frolkis AD, Lipton DS, Fiest KM, *et al*. Cumulative incidence of second intestinal resection in Crohn's disease: a systematic review and meta-analysis of population-based studies. *Am J Gastroenterol* 2014;**109**(11):1739–48.
28. De Cruz P, Kamm MA, Hamilton AL, *et al*. Crohn's disease management after intestinal resection: a randomised trial. *Lancet* 2015;**385**(9976):1406–17.