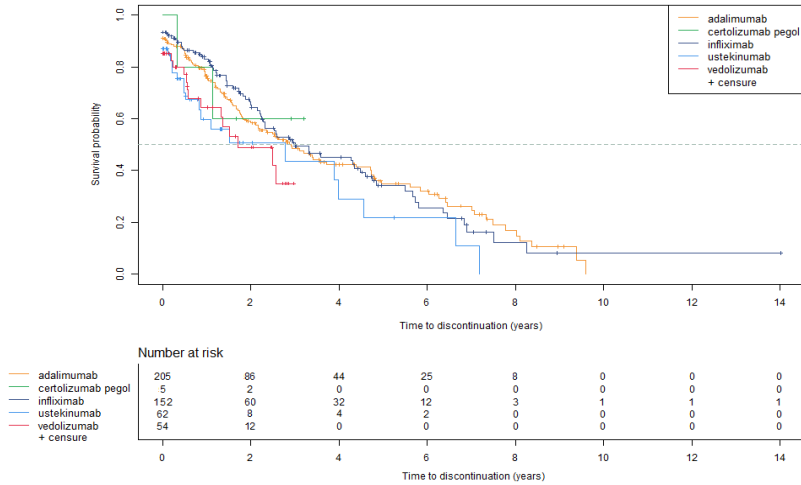


**IDENTIFICATION OF PROGNOSTIC FACTORS IN CROHN'S DISEASE USING  
NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING – A PILOT STUDY**

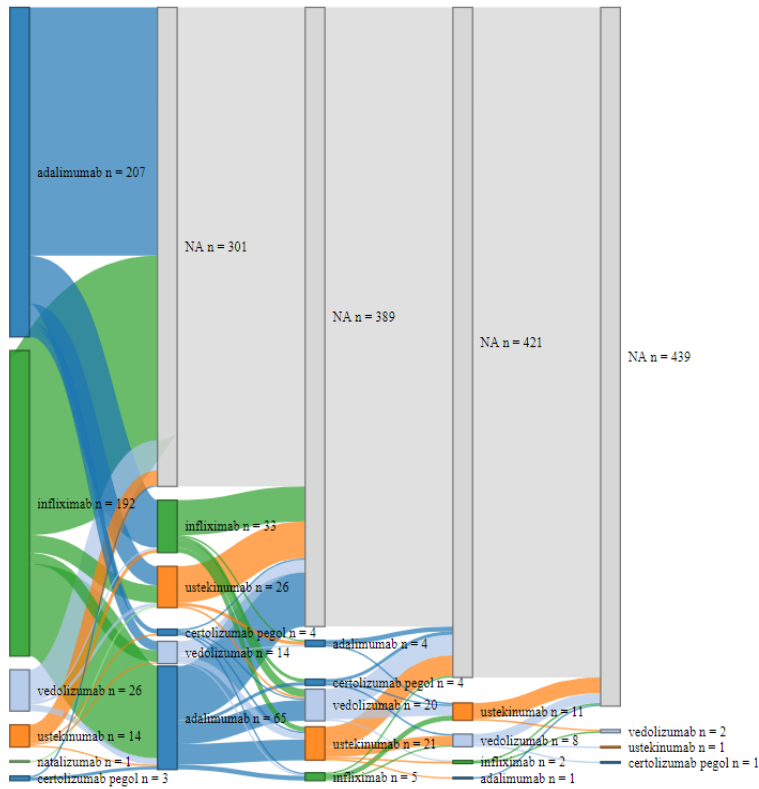
**SUPPLEMENTAL FIGURES AND TABLES**

## Supplemental Figures

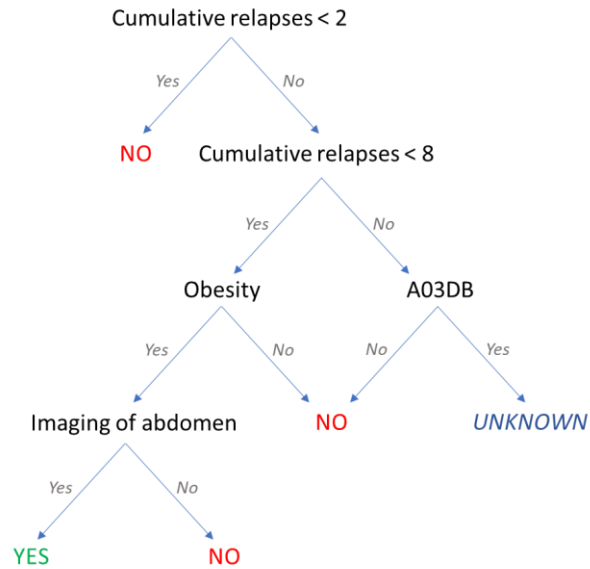


**Figure S1. Time to treatment discontinuation for each biologic agent.** Survival plot (Kaplan-Meier estimator) indicating the survival probability for each biologic across time in years (top) and corresponding data (bottom). Time to discontinuation was evaluated among those subjects that were treated with biologic agents during the last year of the study period. Because the Index Date (time of diagnosis) could be prior to the onset of the study period, time to discontinuation can exceed the length of the study window.

Formatted: Centered



**Figure S2. Patients assigned to different treatment lines with biologics.** Sankey plot depicting the flow of patients by treatment line during the last year of the study period. Each horizontal bar represents a transition across treatment lines (L1-5, from left to right). The link between nodes represents the number of patients that switched from one treatment to another and is represented with arcs that have a width proportional to the patient switch fraction. The grey areas represent the number of patients from whom we do not have information regarding the subsequent treatment lines. NA = not available, representing either patients who a) continued treatment with the same biologic until end of the study period, b) discontinued the treatment and did not receive any other biologic, or c) information regarding biologic treatment was not available in the EHRs.



**Figure S3. Decision tree of relevant clinical variables to predict CD relapse at 3 months.** The combination of number of cumulative relapses, obesity, imaging of abdomen and ATC code A03DB (belladonna and derivatives in combination with analgesics) yields a pathway that predicts whether a patient could have a CD relapse (yes, green or blue, unknown) or not (no, red). Patients with less than 2 cumulative relapses, or those with less than 8 but that are not obese, are predicted not to incur in a further relapse. On the other hand, patients with more than 8 cumulative relapses who are obese and have had an imaging test of abdomen were predicted to suffer from a further relapse at the 3-month timepoint. Regarding the use of belladonna and derivatives, patients with more than 8 cumulative relapses who had taken these substances appeared in the decision tree predictive variables, though with an unknown outcome of relapse occurrence.

## Supplemental Tables

**TABLE S1. Number of variables per variable category included in the predictive model for CD-related flare-ups.**

Category	Number of variables
Medical History	11515
Surgical procedures	3885
Signs and Symptoms findings	3580
Evaluation Procedures	2570
Concomitant Medication	1745
Evaluation Findings	485
Diagnostic procedures	390
Biopsy Findings	375
Family History	321
Hospital Exposure medication (Study treatments) & Exposure medication (Study treatments)	105
Disease Characteristics	105
Disposition Events (comorbidities, surgery/ procedures, signs, and symptoms)	75
Disposition Events (flare-up, EIMs, complications)	33
Substance use (habits)	15
Laboratory	15
Procedures History	10
Demographics	3

**TABLE S2. Relative importance of the most relevant variables included in the predictive model for CD-related flare-ups ([full list](#)).**

Variable	Relative importance
Accumulative past flare	0.358901
AGE	0.029441
Difference between event value and basal value Leukocytes	0.023361
Difference between event value and basal value Hemoglobin	0.010256
Increment respect to maximum normal value Fibrinogen	0.009158
Disposition events- past admissions	0.008597
Montreal Scale	0.008082
Proton pump inhibitors -A02BC	0.007368
Substance use Findings (habits) -TOBACCO USE	0.006733
Acetic acid derivatives and related substances -M01AB	0.006705
Belladonna and derivatives in combination with analgesics -A03DB	0.006624
Evaluation Procedures -CT of chest and abdomen	0.005912
methylprednisolone -H02AB04	0.005684
ciprofloxacin -J01MA02	0.005551
prednisone -H02AB07	0.005274
Evaluation Procedures -Source specific culture	0.004973
Diagnostic procedures -Laboratory procedure	0.004925
Medical History -Infection due to Enterobacteriaceae	0.004657
Difference between event value and basal value CRP -48	0.004641
SEX	0.004507
mesalazine -A07EC02	0.004463
Increment respect to maximum normal value Leukocytes -9	0.004463
Evaluation Procedures -Imaging of abdomen	0.004428
Increment respect to maximum normal value CRP -48	0.004324
Evaluation Procedures -Blood gas measurement	0.004311
Signs and Symptoms findings -Gastrointestinal tract finding	0.004250
metronidazole -J01XD01	0.004230
Bisphosphonates -M05BA	0.004085
Medical History -Obesity	0.003902

Combinations and complexes of aluminum, calcium and magnesium compounds -A02AD	0.003743
infliximab -L04AB02	0.003694
Medical History -Lymphadenopathy	0.003680
Signs and Symptoms findings -Fatigue	0.003607
Diagnostic procedures -Bacterial disease screening	0.003504
Evaluation Procedures -Inspection	0.003432
Difference between event value and basal value Fibrinogen -71	0.003414
Descent respect to minimum normal value Hemoglobin -1	0.003410
Benzodiazepine derivatives -N05BA	0.003272
Anilides -N02BE	0.003251
Disease characteristics -Crohn's disease of ileum	0.003230
azathioprine -L04AX01	0.003028
Signs and Symptoms findings -Altered bowel function	0.003003
Disease characteristics -L2	0.002998
Diphenylmethane derivatives -N05BB	0.002955
Signs and Symptoms findings -Mucosal finding	0.002950
Evaluation Findings -Assessment scales	0.002948
Biopsy Findings -Biopsy	0.002945
Medical History -Ileitis	0.002888
Descent respect to minimum normal value Leukocytes -9	0.002874
Medical History -Disorder of coronary artery	0.002867
Evaluation Procedures -Glucose measurement, body fluid	0.002794
Signs and Symptoms findings -Erythema	0.002772
Immunoglobulins, normal human -J06BA	0.002755
Signs and Symptoms findings -Pain in limb	0.002752
Glucocorticoids -H02AB	0.002722
Signs and Symptoms findings -Platelet production finding	0.002722
Surgical procedures -Partial excision of large intestine	0.002718
Bile acid sequestrants -C10AC	0.002696
Evaluation Procedures -Body fluid analysis	0.002688
Diagnostic procedures -Exploration of digestive system	0.002670
Combinations of penicillins, incl. beta-lactamase inhibitors -J01CR	0.002663
Electrolyte solutions -B05XA	0.002618

Procedures History -Other	0.002576
Opioids in combination with non-opioid analgesics -N02AJ	0.002550
Signs and Symptoms findings -Finding of quantity of appetite	0.002541
Medical History -Structural disorder of heart	0.002501
Medical History -Colitis	0.002489
Medical History -Disorder of spleen	0.002472
Surgical procedures -Operation on prepuce	0.002453
Evaluation Procedures -Red blood cell indices determination	0.002447
Procedures History -Therapeutic procedure	0.002444
Medical History -Disorder of abdomen	0.002437
Medical History -Disorder of colon	0.002428
Propionic acid derivatives -M01AE	0.002425
Medical History -Disorder of pancreas	0.002417
Signs and Symptoms findings -Chest wall pain	0.002402
Progestogens and estrogens, fixed combinations -G03AA	0.002400
Evaluation Procedures -Complete blood count	0.002343
Surgical procedures -Partial colectomy with anastomosis	0.002340
Beta blocking agents, selective, and other diuretics -C07CB	0.002326
Evaluation Procedures -In-vitro immunologic test	0.002314
Evaluation Procedures -Endoscopy of large intestine	0.002305
Disease characteristics -Stricture Crohn's disease	0.002261
ACE inhibitors, plain -C09AA	0.002259
budesonide -A07EA06	0.002235
Medical History -Inflammation of large intestine	0.002235
Medical History -Respiratory tract infection	0.002235
Other drugs for functional gastrointestinal disorders -A03AX	0.002217
Detoxifying agents for antineoplastic treatment -V03AF	0.002198
Substance use Findings (habits) -ALCOHOL USE	0.002184
Substituted alkylamines -R06AB	0.002169
H2-receptor antagonists -A02BA	0.002165
Evaluation Procedures -Antibody measurement	0.002158
Substance use Findings (habits) -OTHER SUBSTANCES	0.002127
Esters of aminobenzoic acid -N01BA	0.002125



Medical History -Disorder of lower gastrointestinal tract	0.002081
Third-generation cephalosporins -J01DD	0.002077
Insulins and analogues for injection, long-acting -A10AE	0.002053
Bulk-forming laxatives -A06AC	0.002052
Medical History -Disorder due to infection	0.002038
Selective beta-2-adrenoreceptor agonists -R03CC	0.002012
Medical History -Injury of pancreas	0.002000
Iron bivalent, oral preparations -B03AA	0.001994
Signs and Symptoms findings -Pain / sensation finding	0.001959
Evaluation Procedures -Imaging	0.001946
Evaluation Procedures -Ultrasonography of abdomen	0.001941
Signs and Symptoms findings -Abdominal pain	0.001920
Surgical procedures -Scraping	0.001870
Belladonna alkaloids, tertiary amines -A03BA	0.001862
Evaluation Procedures -Evaluation procedure	0.001858
Evaluation Procedures -Protein electrophoresis	0.001818
Biopsy Findings -Biopsy of large intestine	0.001800
Evaluation Procedures -Soft tissue X-ray of abdomen	0.001796
Other immunosuppressants -L04AX	0.001790
Combinations of sulfonamides and trimethoprim, incl. derivatives -J01EE	0.001743
Evaluation Procedures -Delayed hypersensitivity skin test for t...	0.001730
Medical History -Intestinal infectious disease	0.001677
Other antiinflammatory and antirheumatic agents, non-steroids -M01AX	0.001674
Biopsy Findings -Removal	0.001670
Softeners, emollients -A06AA	0.001667
Evaluation Procedures -Electrophoresis measurement	0.001658
Evaluation Procedures -Urinary endoscopy	0.001646
Beta blocking agents, non-selective -C07AA	0.001643
Fluoroquinolones -S01AE	0.001624
Selective estrogen receptor modulators -G03XC	0.001616
Evaluation Procedures -Radiographic procedure of chest	0.001614
Evaluation Procedures -CT of bone	0.001613
Evaluation Procedures -Imaging of head	0.001606

Beta blocking agents, selective -C07AB	0.001587
Signs and Symptoms findings -Pain in male genitalia	0.001585
Disease characteristics -B1	0.001566
adalimumab -L04AB04	0.001564
Evaluation Procedures -Endoscopy of intestine	0.001548
Medical History -Diverticular disease of large intestine	0.001546
Heparin group -B01AB	0.001537
Signs and Symptoms findings -Cramp	0.001533
Signs and Symptoms findings -Joint pain	0.001532
sulfasalazine -A07EC01	0.001525
Other antibacterials -J01XX	0.001520
Signs and Symptoms findings -Incontinence	0.001471
Evaluation Procedures -Diagnostic radiography, posteroanterior ...	0.001451
Medical History -Gastroenteritis	0.001441
Surgical procedures -Ileorectostomy	0.001422
Family History -Crohn's disease of small intestine	0.001406
Evaluation Procedures -Laboratory test	0.001383
Signs and Symptoms findings -Nausea and vomiting	0.001367
Evaluation Procedures -Heart procedure	0.001361
Corticosteroids, potent, combinations with antibiotics -D07CC	0.001357
Medical History -Myositis	0.001351
Imidazole and triazole derivatives -D01AC	0.001346

---