Validation and update of the Lémann index to measure cumulative structural bowel damage in Crohn's disease

Benjamin Pariente, Joana Torres, Johan Burisch, Naila Arebi, Brigida Barberio, Dana Duricova, Pierre Ellul, Adrian Goldis, Ioannis P. Kaimakliotis, Konstantinos H. Katsanos, et al.

Validation and update of the Lémann index to measure cumulative structural bowel damage in Crohn's disease.

Short Title: Validation and update of the Lémann index

Benjamin Pariente[1], Joana Torres[2], Johan Burisch[3], Naila Arebi[4], Brigida Barberio[5], Dana Duricova[6], Pierre Ellul[7], Adrian Goldis[8], Ioannis Kaimakliotis[9], Konstantinos Katsanos[10], Željko Krznaric[11], Deirdre Mc Namara[12], Natalia Pedersen[13], Shaji Sebastian[14], Mustapha Azahaf[15], Petra Weimers[16], Philip Lung[17], Carmelo Lacognata[18], Martin Horak[19], Dimitrios Christodoulou[10], Viktor Domislovic[11], Ian Murphy[20], Jérôme Lambert[21], Ryan Ungaro[22], Jean-Frédéric Colombel[22], Jean-Yves Mary[21]

[1] Hepato-Gastroenterology Department, Claude Huriez hospital, University of Lille, Lille, France.

[2] Division of Gastroenterology, Surgical Deparment, Hospital Beatriz Ângelo, Loures, Portugal

[3] Department of Gastroenterology, medical division, Hvidovre University Hospital, Hvidovre, Denmark

[4] Department of Inflammatory Bowel Disease, St Mark's Hospital and Imperial College London

[5] Division of Gastroenterology, Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy.

[6] IBD clinical and research center, ISCARE a.s., Prague, Czech Republic; Institute of Pharmacology, First Faculty of Medicine, Charles University, Prague, Czech Republic

[7] Mater Dei Hospital, Department of Medicine, Division of Gastroenterology, Msida MSD 2020, Malta

[8] Gastroenterology, Department of University of Medicine, Timișoara, România

[9] American Gastroenterology Center, Cyprus

[10] Division of Gastroenterology, Department of Internal Medicine, Faculty of Medicine University of Ioannina School of Health Sciences, 45110 Ioannina, Greece.

[11] Department of Gastroenterology, Hepatology and Nutrition, University Hospital Centre Zagreb, Zagreb, Croatia

[12] TAGG Research Centre, School of Medicine, Trinity College Dublin

[13] Department of Gastroenterology, Slagelse Hospital, Denmark

[14] IBD Unit, Hull University Teaching Hospitals NHS Trust, Hull, United Kingdom

[15] Radiology Department, Claude Huriez hospital, University of Lille, Lille, France.

[16] Department of Gastroenterology, North Zealand University Hospital, Capital Region, Denmark

[17] Radiology Department, St Mark's Hospital and Imperial College

[18] Radiology Unit, University of Padova, Padova, Italy.

[19] Department of Radiology, Na Homolce Hospital, Prague, Czech Republic.

[20] Department of Radiology, Tallaght University Hospital, Dublin 24, Ireland

[21] INSERM U1135 CRESS 'Centre de Recherche Epidémiologie et Statistiques', Equipe ECSTRRA, Université de Paris, Hôpital Saint-Louis, Paris, France.

[22] Department of Medicine, Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York.

Correspondance: Pr Benjamin Pariente, Service des maladies de l'appareil digestif, Center hospitalier Claude Huriez, 2 avenue Oscar Lambret, 59000, Lille, France. E-mail: benjamin.pariente@chru-lille.fr; Telephone: +33 320 44 53 43; Fax: +33 320 44 55 64.

Disclosures:

Name	**Disclosures**
B. Pariente	Consulting fees: AbbVie, MSD, Takeda, Janssen, Lilly, Pfizer, Biogaran, Biogen, Mylan, and Sandoz Lecture fees: Abbvie, MSD, Takeda, Janssen, Ferring, Lilly, and Mylan
J. Torres	Consulting fees: Takeda, Janssen Lecture fees: Janssen
J. Burisch	Consulting fees from AbbVie, Janssen-Cilag, Celgene, MSD, Pfizer, Takeda, Tillots Pharma, Samsung Bioepis Grants from: MSD, Takeda, Tillots Pharma, Novo Nordisk and Bristol Myers Squibb
N. Arebi	Consulting fees: Janssen Lecture fees: Pfizer, Janssen
J.Lambert	None
B.Barberio	None
D. Duricova	Lecture fees: Takeda, Janssen, Pfizer
P. Ellul	None
A. Goldis	None
J. Kaimakliotis	None
K. Katsanos	Honoraria from AbbVie, Enorasis, Ferring, Janssen, MSD, Shire,Takeda
Z. Krznaric	Consulting fees: AbbVie, MSD, Takeda, Janssen, Fresenius, Mylan, Oktal Pharma and Sandoz Lecture fees: Abbvie, MSD, Takeda, Janssen, Fresenius, Oktal Pharma and Mylan
D. Mc Namara	None
N. Pedersen	None
S. Sebastian	Holds research grants from Biogen, Takeda, AbbVie, Warner Chilcott, Ferring, MSD, Biohit and Celgene, serves on the advisory boards of Takeda, AbbVie, Merck, Ferring, Pharmacocosmos, Warner Chilcott, Janssen, Falk Pharma, Biohit, TriGenix,Cellgene and Tillots Pharma,

	and has received speaker fees from Abbvie, Janssen, Merck, Warner Chilcott and Falk Pharma
M. Azahaf	None
P. Weimers	Consulting fees from Vifor Pharma Nordiska AB, grants from Ferring lægemidler and Tillotts Pharma AG, as well as non-financial support from Janssen-Cilag A/S, Calpro AS, Pharmacosmos A/S and Vifor Pharma Nordiska AB.
PFC. Lung	None
C. Lacognata	None
M. Horak	None
D.Christodoulou	Honoraria from Abbvie, Janssen, Takeda, MSD
V. Domislovic	None
I. Murphy	None
J.Lambert	None
R. Ungaro	RCU has served as an advisory board member or consultant for Eli Lilly, Janssen, Pfizer, and Takeda; and research support from AbbVie, Boehringer Ingelheim, and Pfizer. RCU is supported by an NIH K23 Career Development Award (K23KD111995-01A1).
JF. Colombel	Dr. Colombel reports receiving research grants from AbbVie, Janssen Pharmaceuticals and Takeda; receiving payment for lectures from AbbVie, Amgen, Allergan, BMS, Inc. Ferring Pharmaceuticals, Shire, and Takeda; receiving consulting fees from AbbVie, Amgen, Arena Pharmaceuticals, Boehringer Ingelheim, BMS, Celgene Corporation, Celltrion, Eli Lilly, Enterome, Ferring Pharmaceuticals, Geneva, Genentech, Gilead, Iterative Scopes, Ipsen, Imedex, Immunic, lmtbio, Inotrem, Janssen Pharmaceuticals, Landos, LimmaTech Biologics AG, Medimmune, Merck, Novartis, O Mass, Ostuka, Pfizer, Shire, Takeda, Tigenix, Viela bio; and hold stock options in Intestinal Biotech Development.
JY. Mary	None

Author Contributions

- Study concept and design: Benjamin PARIENTE, Jean-Frederic COLOMBEL, and Jean-Yves MARY
- Acquisition of data: all authors except Jérôme LAMBERT and Jean Yves MARY
- Interpretation of data: Benjamin PARIENTE, Jérôme LAMBERT, and Jean-Yves MARY
- Drafting of the manuscript: Benjamin PARIENTE and Jean-Yves MARY
- Critical revision of the manuscript for important intellectual content: all authors
- Statistical analysis: Jean-Yves MARY
- Study supervision: Benjamin PARIENTE, Jean-Frederic COLOMBEL, and Jean-Yves MARY
- Approval of the final version: all authors

Acknowledgments

The authors would like to acknowledge:

The European Crohn's and Colitis Organisation

Vicent Hernandez (Gastroenterology Department, Complexo Hospitalario Universitario de Vigo, Vigo, Spain)

Alexandros SKAMNELOS (Department of Radiology, Faculty of Medicine, University of Ioannina School of Health Sciences, 45110 Ioannina, Greece)

Christina NAKA (Department of Radiology, Faculty of Medicine, University of Ioannina School of Health Sciences, 45110 Ioannina, Greece)

This study was partially funded by an Abbvie Investigator Initated STudy: "CROCO - Crohn´s disease Cohort Study" and by GEDII (Portuguese Group for the Study of IBD)

ABSTRACT

Background: The Lémann index is a tool measuring cumulative structural bowel damage in Crohn's disease (CD). We here report its validation and updating.

Methods: This was an international, multicenter, prospective, cross-sectional observational study. At each center, 10 inclusions, stratified by CD duration and location, were planned. For each patient, the digestive tract was divided into 4 organs, upper tract, small bowel, colon/rectum, anus, and subsequently into segments, explored systematically by magnetic resonance imaging (MRI), and by endoscopies in relation to disease location. For each segment, investigators retrieved information on previous surgical procedures, identified predefined strictures and penetrating lesions of maximal severity (grades 1-3) at each organ investigational method (gastroenterologist and radiologist for MRI), provided segmental damage evaluation ranging from 0.0 to 10.0 (complete resection). Organ resection-free cumulative damage evaluation was then calculated from the sum of segmental damages. Then, investigators provided a 0-10 global damage evaluation from the 4 organ standardized cumulative damage evaluations. Simple linear regressions of investigator damage evaluations on their corresponding Lémann index were studied, as well as calibration plots. Finally, updated Lémann index was derived through multiple linear mixed models applied to combined development and validation samples.

Results: In 15 centers, 134 patients were included. Correlation coefficients between investigator damage evaluations and Lémann indexes were above 0.80. When analyzing data in 272 patients from both samples and 27 centers, the unbiased correlation estimates were 0.89, 0,97, 0,94, 0.81, 0.91 for the 4 organs and globally, and stable when applied to one sample or the other.

Conclusions: The updated Lémann index is a well-established index to assess cumulative bowel damage in Crohn's disease that can be used in epidemiological studies and disease modification trials.

Key words: Crohn's' disease, bowel damage, Lémann index, validation

INTRODUCTION

Crohn's disease (CD) is a chronic inflammatory bowel disorder, characterized by periods of clinical remission alternating with periods of relapse. Persistent and under-treated inflammation over time is believed to lead to cumulative bowel damage including the development of fibrostenotic strictures, abscesses or fistulae [1]. These complications frequently lead to altered intestinal function and represent the main indications for surgical resection, a major contributor to disability [2]. The goals of therapy in CD have evolved from merely controlling symptoms and improving quality of life to blocking disease progression and improving long-term disease outcomes by reducing structural damage, disability, and irreversible disease complications [3–5]. However, current available clinical and endoscopic indexes only measure CD activity at a specific time point and do not measure the cumulative impact of the disease [6].

To address this need, the first instrument measuring cumulative structural bowel damage in CD, the Lémann index, was recently developed by an international panel of experts including gastroenterologists and radiologists [7]. This study also included the development of a specific index for each of the four following organs: upper digestive tract, small bowel, colon and anus and of a global index. An internal cross-validation was performed using bootstrap method. According to this validation, the unbiased correlation coefficients between predicted Lémann indexes and investigator damage evaluations were 0.85, 0.98, 0.90, 0.82 for upper tract, small bowel, colon/rectum, anus, respectively, and 0.84 globally. In addition, an inter-observer variation study of organ and global Lémann indexes using magnetic resonance imaging (MRI) data only was performed in a sub-sample of patients and investigators. However, this original work did not include external validation, which is essential for clinical applications.

Therefore, the goals of the present study were first to validate the Lémann index in an

independent cohort of CD patients with a different set of investigators not originally involved in its development, and second to update the index by combining data from the development and validation study.

METHODS

Investigators

Investigators were recruited from European tertiary centers in inflammatory bowel disease belonging to the Epidemiological Committee of the European Crohn's and Colitis Organisation (ECCO), and from the Mount Sinai Hospital in New York City, USA. All investigators (except BP) had not been involved in the initial development of the index. They were trained by the main investigator (BP) during dedicated meetings, to ensure consistency in understanding the concept of bowel damage, identification of damage components, organ damage evaluation, and the global digestive tract damage evaluation.

Study design

This was a multicenter, cross-sectional, prospective, observational study. Institutional review boards and independent ethics committees of the participating institutions approved the protocol as required by local regulations. Informed consent was obtained from all the participating patients according to national guidelines. At each center, 10 patients were planned to be included with a stratified recruitment on known or suspected CD location (upper gastrointestinal tract, small bowel, colon/rectum, and anus) and disease duration (< 2 years, ≥ 2 years and < 10 years, and ≥ 10 years), as described below. Similar to the Lémann index development phase, patients' recruitment per center was stratified as follows: one patient with upper gastrointestinal tract involvement, three patients with small bowel involvement, three patients with colon and/or rectum involvement, three patients with anal involvement, with one patient having each disease duration category for the latter three disease location groups. Disease location was not mutually exclusive [7]. For each patient, the following examinations were performed as originally described: clinical examination and abdominal MRI in all cases, and upper endoscopy, colonoscopy, pelvic MRI according to known or suspected CD location; all examinations had to be performed in a maximum time

window < 120 days [7]. Upper tract, small bowel, colon/rectum, and anus were divided in 3, 20, 6 and 1 segments, respectively [7].

Data Collection

History of surgical procedures of maximal severity (grade 1 to 3), including the percentage of resected segment in case of partial resection, was collected per segment (Supplementary Table 1). For each segment, one gastroenterologist and one radiologist per center identified the presence of predefined stricturing and penetrating lesions of maximal severity (grade 1 to 3) at each examination (Supplementary Table 1). Investigators provided a resection-free damage evaluation for each non-resected segment ranging from 0 to 10 (10 corresponding to the damage of a completely resected segment), according to the history of surgical procedures of maximal severity grade on this segment, except resection, and the presence of stricturing and penetrating lesions of maximal severity in this segment at each examination MRI (and CT if available), anus clinical exam, and upper endoscopy or colonoscopy or pelvic MRI) as originally described and as illustrated in the Supplementary Table 2 [7]. For each patient, the investigator organ resection-free cumulative damage evaluation was calculated as the sum of resection-free segmental damage evaluations using weights of 2.0 for duodenum, 1.5 for distal ileum segment, 2.0 for rectum, and 1.0 for all other segments, as determined in the development study [7]. In parallel, investigator organ resection-free cumulative damage evaluations were also calculated using a weight of 1.0 for all segments. Finally, for each organ, a standardized cumulative damage evaluation was calculated as the weighted sum of the segmental damage evaluations provided by the investigators plus the damage attributed to the resected segments in case of previous resection, standardized to a scale of 0.0 to 10.0 through division by the total number of segments of the organ. The calculation of the contribution of a segment with partial resection to standardized cumulative damage is described in the Supplementary Methods, data

collection. Investigators provided a global damage evaluation from 0 to 10 for each patient according to the 4 investigator organ standardized cumulative damage evaluations, taking into account the relative importance they gave to each organ in terms of damage as illustrated in the Supplementary Table 3.

In case of an investigator evaluation detected as "aberrant" compared to the evaluations provided by other investigators at segmental, organ, or global level, the investigator was contacted about this discrepancy (over or under estimation), and informed that he/she could modify his/her evaluation. For each "aberrant" evaluation, only one contact was permitted, even if the second evaluation was still detected as "aberrant". Methods used to detect possibly "aberrant" values of investigator damage evaluations at the segmental, organ or global level, during data collection are described in the Supplementary Materials, data collection.

In each center, investigators prospectively collected patients' data at enrollment using an anonymous electronic case report form (e-CRF). The enrollment form of the e-CRF included stratification factors, birth date, sex, date of CD diagnosis, date of enrollment, and known or suspected CD locations at the time of enrollment. No information on previous or on-going treatment was collected. Previous surgical procedures of each grade and observed stricturing and penetrating lesions of each grade were recorded per segment on the e-CRF. Newly recorded information was automatically transferred to the organ damage evaluation form allowing investigators to record their estimates of segmental damage evaluations. Then, the 4 organ standardized cumulative damage evaluations were calculated as described above and automatically transferred to the global damage form allowing investigators to record their global damage evaluation estimate.

Statistical analyses

Qualitative and quantitative characteristics of included patients were described through n and %, n, median and inter-quartile range (IQR), respectively. Recorded surgical procedures and observed lesions across the different examinations were summarized per segment of each organ as patient numbers per most severe grade.

Each organ Lémann index was calculated for each patient from the numbers of segments involved by stricturing and penetrating lesions of maximal grade across examinations, as described in the development paper [7]. Validation of each organ Lémann index was undertaken by estimating the correlation and the characteristics of the linear relationship, slope and intercept, between the investigator organ resection-free cumulative damage evaluation and the organ Lémann index in the sub-sample of patients with organ damage known or suspected at inclusion and all patients with organ damage found during at least one investigational method. This relationship was described by the scatterplot of investigator organ resection-free cumulative damage evaluation as a function of organ Lémann index. In addition, correlation was estimated between investigator organ resection-free cumulative damage evaluation using a weight of 1.0 for each segment and the organ Lémann index.

Global Lémann index was calculated for each patient from the 4 organ standardized cumulative damage evaluations, as published [7]. Validation of the global Lémann index was undertaken by estimating the correlation and the characteristics of the linear relationship, slope and intercept, between the investigator global damage evaluation and the global Lémann index in the whole sample. This relationship was described by the scatterplot of investigator global damage evaluation as a function of global Lémann index.

For organ and global Lémann indexes, a perfect validation should correspond to a rather narrow scatterplot, a correlation estimate near 1.00, a slope and an intercept estimate of

the linear relationship near 1.00 and 0.00, respectively. For each organ and globally, the proportion of the variance of the investigator damage evaluation explained by the Lémann index was estimated through adjusted square correlation value and test of slope to 1 was performed through t-test. In addition, calibration-in-the-large and slope calibration plots were drawn [8].

Following the careful examination of differences between development and validation sample as described in the Supplementary Methods, statistical analyses, and of calibration plots, we analyzed together data of the development (n=138) and validation (n=134) samples [9]. The updated organ Lémann indexes were then constructed through a multiple linear mixed model [10] applied to the combined samples: in this model, the dependent variable was the calculated organ resection-free cumulative damage evaluation as the sum of the segmental damage evaluations ignoring surgical resections. Following the results observed in the first part of the validation study, all segments of all organs were given the same weight of 1.0 in these calculations. The independent variables of the model were the numbers of segments with surgical interventions of each maximal severity grade, except resections, the numbers of segments with stricturing lesions of each maximal severity grade, and the number of segments with penetrating lesions of each maximal severity grade. Study and center were used as random factors. The multiple linear mixed model allowed the estimation of the coefficients of the linear combination of independent variables showing the best fit with investigator organ resection-free cumulative damage evaluation through maximal restricted likelihood [10]. The independent variables were selected manually using backward selection and likelihood ratio test. The same method was used to construct the updated global Lémann index, using investigator global damage evaluation as the dependent variable, the 4 calculated organ standardized cumulative damage evaluations as independent variables, study and centers as random factors.

The quality of the updated organ and global Lémann indexes was assessed through the percentage of the variance of the dependent variable explained by the multiple linear mixed model [11], the scatterplot of the updated organ Lémann indexes or global Lémann index as a function of the investigator organ resection-free cumulative damage evaluations or global damage evaluation. In addition, the unbiased correlation values between investigator organ resection-free cumulative damage evaluations and updated organ Lémann indexes, between investigator global damage evaluation and updated global Lémann index, were estimated ignoring the random factors as the indexes will be used in practice [12].

For details on random factors, percentage of the variance explained by the model, bootstrap method used, see Supplementary Methods, statistical analyses.

The statistical analysis was carried out on 19.1 SPSS software (SPSS, Chicago, USA).

RESULTS

Patients

From June 2017 to December 2018, a total of 134 patients were enrolled in 15 centers (7 to 10 patients per center). The main baseline characteristics of the 134 included patients are described in Table 1.

All 134 patients had a clinical examination and an abdominal MRI (including 8 patients with CT-scan); 22 patients an upper endoscopy; 107 patients a colonoscopy; and 39 patients a pelvic MRI. The median delay between all performed examinations was 0.92 month, with a delay less than 1.9 months for 75% of the patients. The numbers of patients with 1, 2, 3 or more segments involved by surgical procedures, stricturing lesions, and penetrating lesions of maximal grade are summarized in Table 2. One patient had a by-pass in the duodenum, 24 and 23 patients had a history of small bowel and colon/rectum surgical resection, respectively. Two patients had a total colectomy and two had a proctectomy.

Validation of the organ and global Lémann indexes

Mean ± SD (range) of the investigator organ resection-free cumulative damage evaluations and of the calculated organ Lémann indexes were 2.2 ± 3.2 (0.0 – 9.0) and 1.3 ± 2.3 (0.0 – 7.0) for upper tract (n=22), 5.9 ± 6.2 (0.0 – 29.8) and 4.3 ± 4.8 (0.0 – 24.5) for small bowel (n=102), 6.5 ± 8.0 (0.0 – 42.9) and 5.4 ± 6.4 (0.0 – 35.5) for colon/rectum (n=116), and 3.1 ± 1.6 (0.0 – 7.0) and 1.9 ± 1.6 (0.5 – 5.0) for anus (n=45). Correlation coefficients between investigator organ resection-free cumulative damage evaluation and each organ Lémann index were 0.91, 0.96, 0.95, and 0.81, for upper tract, small bowel, colon/rectum, and anus, respectively (Figure 1A-D). When ignoring the particular weights of duodenum, distal ileum and rectum, the correlation values were 0.89, 0.96, and 0.96 for upper tract, small bowel, and colon/rectum. Proportions of the investigator organ resection-free cumulative damage evaluation variance explained by each organ Lémann index were 82%,

91%, 89%, and 65%, for upper tract, small bowel, colon/rectum, and anus, respectively. Slope and intercept estimates and standard errors of the linear regression of the investigator organ resection-free cumulative damage evaluation on the organ Lémann index are shown in Table 3.

Mean ± SD (range) of the investigator global damage evaluation and of the calculated global Lémann index were 0.8 ± 0.8 (0.0 – 5.0) and 1.2 ± 1.3 (0.0 – 7.9), respectively. Correlation coefficient between investigator global damage evaluation and the global Lémann Index was 0.98 (Figure 1E). Correlation coefficient estimate varied across centers, with a median (IQR) of 0.989 (0.986 – 0.994). Proportion of the investigator global damage evaluation variance explained by the global Lémann index was 96%. Slope and intercept estimates of the linear regression of the investigator global damage evaluation on the global Lémann index are shown in Table 3. All slope estimates were significantly different from 1.0 for all organs and globally, except for upper tract, for which the test was borderline but not significant. Calibration-in-the-large and slope calibration plots are provided in the supplementary figure 1,

We observed some differences between the development and validation samples: surgical procedures of grade 2 (bypass diversion or stricturoplasty) were never observed in the upper tract in the development sample and were rarely observed in the small bowel and in the colon/rectum in the validation sample; stricturing lesions of grade 3 were never observed in the anus in the validation sample; penetrating lesions of grade 2 were rarely (1 case with 1 segment) observed in the upper tract, and those of grade 3 were never observed in the validation sample. A large difference was observed in the investigator global damage evaluation between the development and the validation samples (median of 1.50 and 0.50, $p<0.0001$) with a parallel variation in the investigator organ standardized cumulative damage evaluations, which was significant for the small bowel (median of 0.49 and 0.22, $p=0.001$).

Only a few borderline significant differences were observed in the numbers of segments with interventional procedures, stricturing or penetrating lesions of each grade when comparing the development and the validation samples.

Updated organ and global Lémann indexes obtained through the analysis of the development and validation samples together

Table 4 provides the coefficient estimates for each of the 4 organs with its standard error and p-value, and the rounded coefficient to be applied to the numbers of segments with stricturing and penetrating lesions of each grade of severity for calculation of the organ Lémann index. The percentage of the variance of the investigator calculated organ damage evaluation explained by the model was 81%, 96%, 91%, and 74% for the upper digestive tract, the small bowel, the colon and rectum, and the anus, respectively. Figures 2A-D show the scatterplot of the predicted organ Lémann index versus the investigator organ damage evaluation. The scatterplot of the residual as a function of the predicted Lémann index is presented per organ in Supplementary Figures 2A-D. The raw correlations between the updated organ Lémann indexes and investigator organ resection-free cumulative damage evaluations ignoring random factors were globally (in the development sample and in the validation sample) 0.904 (0.851 and 0.943), 0.976 (0.980 and 0.943), 0.947 (0.923 and 0.963), 0.811 (0.821 and 0.823), for the upper tract, the small bowel, the colon/rectum, and the anus, respectively. The optimisms of these correlation coefficient estimates were 0.012, 0.001, 0.003, and 0.006, for the upper tract, small bowel, colon/rectum, and anus, respectively. This led to unbiased correlation coefficients 0.891, 0.975, 0.944, 0.805, for upper tract, small bowel, colon/rectum, anus, respectively, as shown in Figures 2A-D.

Table 4 gives the coefficient estimates with standard error and p-value and the rounded coefficients to be applied to the cumulative standardized damage evaluation per organ in order to calculate the global Lémann index. The percentage of the variance of the investigator

global damage evaluation explained by the model was 81%. Figure 2E shows the scatterplot of the global Lémann index *versus* the investigator global damage evaluation. The scatterplot of the residual as a function of the predicted global Lémann index is presented in Supplementary Figure 2E. The raw correlation between updated global Lémann index and investigator global damage evaluation ignoring random factors were globally (in the development sample and in the validation sample) 0.920 (0.904 and 0.979). The optimism of this global correlation coefficient estimate was 0.005, leading to an unbiased correlation coefficient 0.915, for upper tract, small bowel, colon/rectum, anus, respectively (Figures 2E). The estimates of the variance of the random factors were 0.297 for the between studies factor, 0.150 for the between investigators factor nested within the study factor and 0.249 for the residuals.

Supplementary Table 4 provides for each of the 4 organs the rounded coefficients to be applied to the number of segments with stricturing and penetrating lesions of each severity grade in order to estimate the updated predicted organ Lémann index, resections excluded, coefficients derived from the analyses performed on the development sample and on both samples together. Supplementary Table 4 provides also the rounded coefficients to be applied to the 4 organ cumulative standardized damage evaluations in order to estimate the updated predicted global Lémann index, coefficients derived from the analyses performed on the development sample and on both samples together. Of note, rounded coefficients were: 2.0 for upper tract, 4.0 for small bowel, 3.0 for colon/rectum, and 2.5 for anus. Comparison between Supplementary Figure 3 and Figure 2E illustrates the significant improvement in the relationship between the predicted global Lémann index and the investigator global damage evaluation: when using the original Lémann index, the scatterplot is splitting into two clouds as compared to one cloud when using the updated Lémann index.

To facilitate the calculation of the updated organ Lémann indexes and of the updated global Lémann index, an excel sheet (Figures 3A) has been developed and is available online and a user's guide describing the 4 steps necessary to calculate the Lémann index has been added in Figure 3B. An example of a Lémann index calculation is presented in Supplementary Figure 4A-D. Of note, the Lémann index can range between 0 (no damage) and 115 (maximal damage). The Lémann index cannot be calculated if (a) a mandatory examination is not performed or not dated, and/or (b) maximal delay between performed examinations is 120 days or more. In that case, the user is informed through comments 1 or 2.

DISCUSSION

The Lémann index is a unique tool that measures cumulative bowel damage in CD by accounting for surgical procedures and the extent and severity of lesions in the digestive tract. We here report its validation in a cohort of 134 CD patients from 15 international centers. We found a high correlation between investigator organ damage evaluation and the organ Lémann index for all 4 organs (all coefficients over than 0.80). We also observed a high correlation coefficient (0.98) between investigator global damage evaluation and the global Lémann Index, but slope and intercept of their linear relationships deviate largely from expected values. In addition, differences in investigator damage evaluations were observed between the development and validation samples. Thus, we performed analyses of pooled data of 272 patients (138 from development and 134 from validation studies), and derived updated organ and global Lémann indexes showing very similar performances to original non updated indexes [7]. Recently, the Selecting EndPoInts foR Disease ModIfication Trials (SPIRIT) initiative involving experts from the International Organization for the Study of Inflammatory Bowel Diseases (IOIBD), proposed goals to achieve in disease-modification trials for preventing disease progression in patients with inflammatory bowel disease [13]. The Lémann index was voted by the panel as the reference tool to assess bowel damage in CD. It was also agreed that the calculation of the Lémann index should be centralized in future clinical trials, and should be assessed every 12 to 24 months[13].

In the validation study, even though correlation values between investigator organ resection-free cumulative damage evaluation and each organ Lémann index were high, the estimate of the slope of the relationship of the investigator organ resection-free cumulative damage evaluation and the organ Lémann Index was significantly greater than 1.00 for upper tract, small bowel and colon/rectum, and significantly lower than 1.00 for anus, whereas intercept estimate was greater than 0 for upper tract, small bowel and anus. It means that for a

similar level of damage severity as assessed by the organ Lémann Index, the organ resection-free cumulative damage evaluations obtained by the investigators of the validation study were higher on average than those reported in the development sample for upper tract, small bowel and colon/rectum, and lower on average for anus. These differences increased with higher organ Lémann indexes values for upper tract, small bowel and anus. Along the same line, even if the correlation value between investigator global damage evaluation and the global Lémann Index was 0.98, the estimate of the slope of the relationship of the investigator global damage evaluation on the global Lémann Index was significantly lower than 1.00, whereas intercept estimate was greater than 0. This means that for the same level of damage severity as assessed by the global Lémann Index, the global damage evaluations reported by the investigators of the validation study were lower on average than the evaluations performed by the investigators of the development sample. This difference increased with higher global Lémann Index values. These sub-optimal results were confirmed by the calibration plots. During the development phase, some results were imposed without statistical grounds: the number of penetrating lesions of grade 3 in the upper tract was forced into the model despite not significant due to the rarity of the lesion, the surgical procedure of grade 2 in the anus was given a priori a coefficient of 7, sum of the coefficients associated to stricturing and penetrating lesions of grade 3. Some lesions were only observed in the development or in the validation study. Moreover, we observed a large variation between the development and validation samples in the median of global damage evaluations. Altogether, this led us to update the Lémann indexes using a linear mixed model as in the development study, adding a random factor, the study, illustrating the variation in the intercept of the linear regression from one study to another and re-estimating the coefficients to take into account the different situations observed in the two samples.

As shown in Supplementary Table 4, coefficients of the Lémann indexes were different when estimated in the development sample and in both development and validation samples. Still, it is reassuring that the correlations between predicted damage indexes and investigator damage evaluations were very similar when applied to the development sample and to the validation samples in spite of the differences cited above, underlying the robustness of the updated Lémann indexes.

When putting into perspective with the other disease severity indexes validated in inflammatory bowel disease, the results of the validation of the Lémann index are satisfactory. The Crohn's disease endoscopic index of severity (CDEIS), the simplified endoscopic severity index in Crohn's disease (SES-CD) and the ulcerative colitis endoscopic index of severity (UCEIS) have been developed and then validated [14–17]. The CDEIS was validated in a sample of 103 endoscopies, with a Pearson's correlation between the calculated CDEIS and the global evaluation of lesion severity assessed by the endoscopists of 0.81[14]. Similarly, the SES-CD showed a correlation with the CDEIS of 0.89 on the validation sample of 121 endoscopies [15]. The median correlation between UCEIS and overall assessment of severity was 0.93 in the first validation sample, and was 0.90 and 0.93 among investigators blinded and unblinded of patients' clinical status in a second validation sample [16,17]. When using magnetic resonance enterography to evaluate CD disease activity, the Magnetic Resonance Index of Activity (MaRIA) was developed at the segmental level, in 5 colonic segments and terminal ileum, from the corresponding segmental adaptation of the CDEIS (n=213 segments and 50 patients) [18]. Correlation between MaRIA and CDEIS was high as expected in the development study, but also in the validation study, 0.80 and 0.84, at the segmental (n=248) and patient (n=48) level, respectively [19]. In these validation studies, the correlation values, as well as the explained proportion of variability of the investigators' overall severity assessment, are in the same order of magnitude as those here observed.

Another aim of this study was to assess the respective weight of each organ in the damage index. Interestingly, the weight of the damage caused to the entire small bowel was considered as dominating over other organs by investigators. This highlights the need to consider the entire small bowel when assessing the progression of CD and efficacy of medical therapies in clinical practice and clinical trials for patients with small bowel disease. Of note, lesions in the anus had also a significant impact on global Lémann Index. This is mainly related to the fact that there was only one segment in the anus (compared to 6 segments for the colon/rectum for example).

This validation and updating have several strengths. When combining the development and validation study, we included a large number of patients (n=272) and worldwide inflammatory bowel disease centers (n=27) with one gastroenterologist and one radiologist in each center. In the development study, investigators were familiar with the Lémann index, as they belonged to the group that developed its original concept and construction. In the validation study, investigators were specialized in inflammatory bowel disease but without any experience regarding the Lémann index, thus more reflecting gastroenterologists that will use the index in the future. Patients' recruitment was stratified according to CD location and CD duration, ensuring a large range of damage severities within each organ. Stricturing lesions, penetrating lesions, surgical procedures per grade of severity were well defined and described. Moreover, by pooling the development and the validation samples, all lesions of each severity grade were observed in at least one segment of each organ. The heterogeneity between the two samples was also taken into account in the analysis, an important point when considering using the Lémann indexes in future studies. Finally, the observed phenotypic differences between the development sample and the validation sample are an asset. Indeed, if an index does perform as well when applied to two different cohorts, it could mean that this index could be used in a forthcoming one. It also implies that this index

could be used in clinical practice (most likely in referral centers) as part of the longitudinal monitoring of patients which is now recommended in international guidelines [13].

We also acknowledge several limitations. First, we did not perform central reading of MRIs and endoscopies. However, absence of central reading increases the general applicability of the Lémann Index in a more real-world setting, but a training of gastroenterologists and radiologists to identify stricturing and penetrating lesions included in the index is still necessary. Secondly, no data on previous or on-going treatment were collected. As a matter of fact, the damage severity index is a transversal measure at a given time point Previous and/or current treatments may have an impact on surgical procedures and/or on stricturing or penetrating lesions at the time of damage evaluation but should not influence its calculation. Thirdly, the absence of a reproducibility study on lesion detection at the different examinations is another limitation even though a reproducibility study of the organ and global Lémann indexes derived from MRI data was performed during the development study. Finally no sensitivity-to-change evaluation of the index was performed and this should be addressed in further studies.

In conclusion, the Lémann index to assess cumulative bowel damage in CD is now validated and updated. It can be easily calculated by inputting the identified lesions and the history of surgical procedures into an excel calculation sheet, and is ready to be used in clinical research and clinical practice.

REFERENCES

1. Pariente B, Cosnes J, Danese S, et al. Development of the Crohn's disease digestive damage score, the Lémann score. Inflamm Bowel Dis 2011;17:1415–1422.

2. Peyrin-Biroulet L, Loftus EV, Colombel J-F, et al. The Natural History of Adult Crohn's Disease in Population-Based Cohorts. Am J Gastroenterol 2010;105:289–297.

3. Allen PB, Peyrin-Biroulet L. Moving towards disease modification in inflammatory bowel disease therapy: Curr Opin Gastroenterol 2013;29:397–404.

4. Panaccione R, Colombel J-F, Louis E, et al. Evolving definitions of remission in Crohn's disease. Inflamm Bowel Dis 2013;19:1645–1653.

5. Colombel J-F, Mahadevan U. Inflammatory Bowel Disease 2017: Innovations and Changing Paradigms. Gastroenterology 2017;152:309–312.

6. Sandborn WJ, Feagan BG, Hanauer SB, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. Gastroenterology 2002;122:512–530.

7. Pariente B, Mary J-Y, Danese S, et al. Development of the Lémann index to assess digestive tract damage in patients with Crohn's disease. Gastroenterology 2015;148:52-63.

8. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York, NY: Springer; 2009.

9. Janssen KJM, Moons KGM, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76–86.

10. Pinheiro JC, Bates DM. *Mixed-effects models in S and S-PLUS*. New York, NY: Springer; 2000.

11. Xu R. Measuring explained variation in linear mixed effects models. Stat Med 2003;22:3527–3541.

12. Harrell FE, Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY, Springer series in statistics, 2010.

13. Le Berre C, Peyrin-Biroulet L, SPIRIT-IOIBD study group. Selecting Endpoints for Disease-Modification Trials in Inflammatory Bowel Disease: the SPIRIT consensus from the IOIBD. Gastroenterology 2021.

14. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983–989.

15. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505–512.

16. Travis SPL, Schnell D, Feagan BG, et al. The Impact of Clinical Information on the Assessment of Endoscopic Activity: Characteristics of the Ulcerative Colitis Endoscopic Index Of Severity [UCEIS]. J Crohns Colitis 2015;9:607–616.

17. Travis SPL, Schnell D, Krzeski P, et al. Reliability and Initial Validation of the Ulcerative Colitis Endoscopic Index of Severity. Gastroenterology 2013;145:987–995.

18. Rimola J, Rodriguez S, García-Bosch O, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. Gut 2009;58:1113–1120.

19. Rimola J, Ordás I, Rodriguez S, et al. Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. Inflamm Bowel Dis 2011;17:1759–1768.

FIGURES LEGENDS

Figure 1: Scatterplot of the investigator organ resection-free cumulative damage evaluation as a function of the organ Lémann Index in the validation sample, upper tract (A), small bowel (B), colon/rectum (C), anus (D), and of the investigator global damage evaluation as a function of the global Lémann Index in the validation sample (E).

The area of each circle is related to the number of coincident data plotted there.

Figure 2: Scatterplot of the updated organ predicted index *versus* the organ resection-free cumulative damage evaluation given by the investigator for (A) upper tract, (B) small bowel, (C) colon and rectum, (D) anus, and of the updated global predicted index *versus* the global damage evaluation given by the investigator (E).

Results obtained on the development sample in blue and validation sample in red when analyzed together.

For small bowel, one patient with a high level of damage (investigator damage evaluation of 77 was not shown in the Figure 2B).

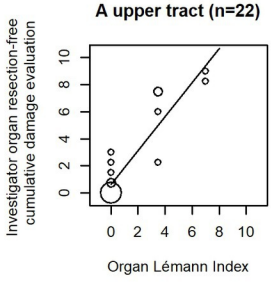
Predicted indexes were estimated from the multiple linear mixed model.

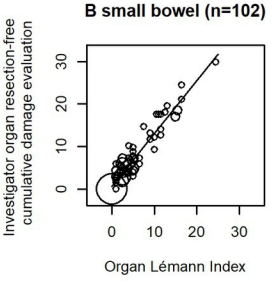
The area of each circle is proportional to the number of coincident data plotted there.

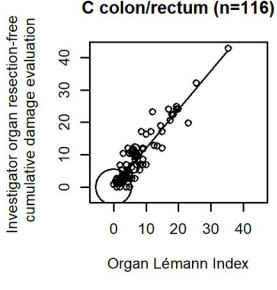
"r" is the unbiased estimate of the correlation coefficient between predicted organ (and global) damage indexes and investigator organ (and global) damage evaluations, derived from the bootstrap method.

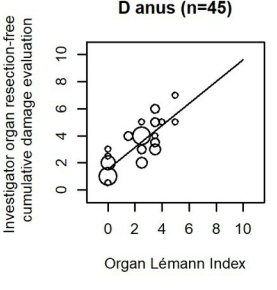
Figure 3: (A) User's guide describing the 4 steps necessary to calculate the Lémann index. (B) Excel file allowing calculation of the Lémann index in one patient.

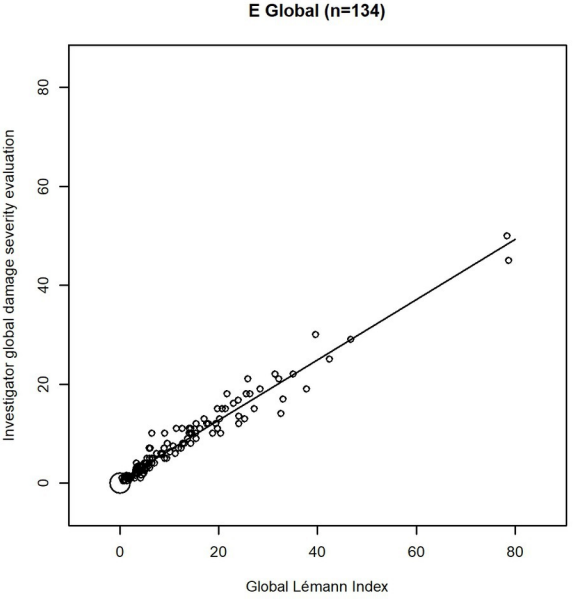
In case of previous surgical resection the investigator should type 100 if complete resection of the segment and the percentage of resected segment in case of partial resection. After identifying strituring and penetrating lesions and their grade of severity at each mandatory examination according to Crohn's disease location, the investigator should type the grade of maximal severity (1, 2, or 3) of each lesion present in a segment at each examination. At the end the most severe lesion in each segment will be automatically selected. The Lémann index can range between 0 (no damage) and 115 (maximal damage).

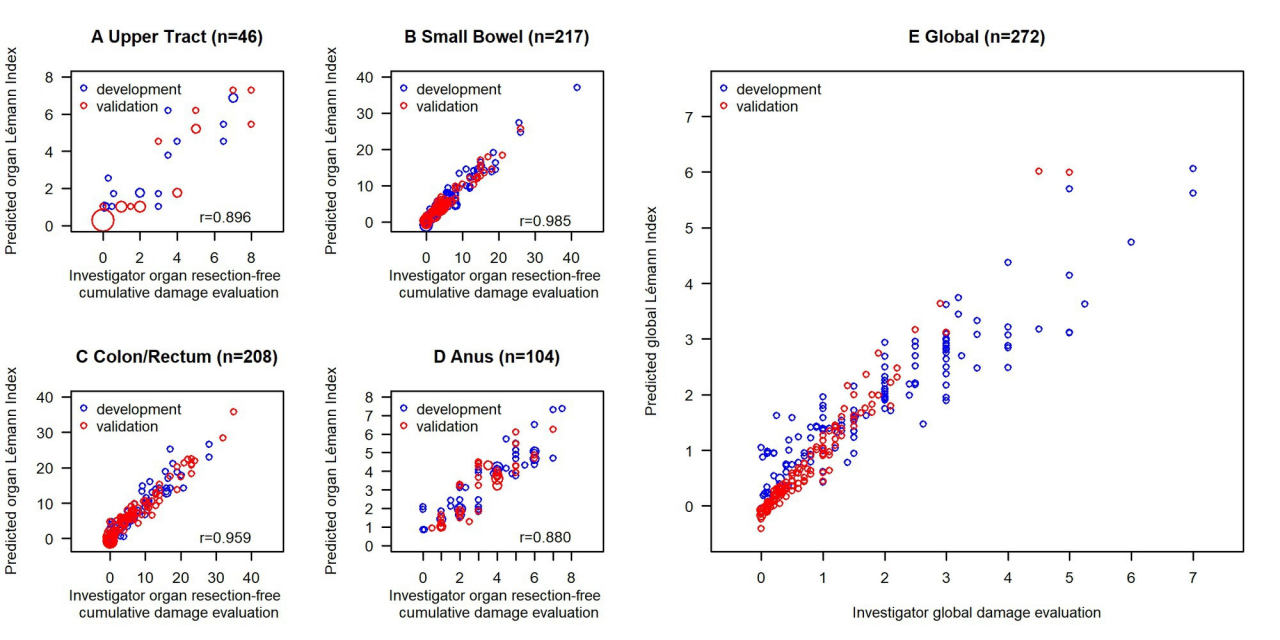

A upper tract (n=22)
Investigator organ resection-free cumulative damage evaluation
0 2 4 6 8 10
0 2 4 6 8 10
Organ Lémann Index


B small bowel (n=102)
Investigator organ resection-free cumulative damage evaluation
0 10 20 30
0 10 20 30
Organ Lémann Index


C colon/rectum (n=116)
Investigator organ resection-free cumulative damage evaluation
0 10 20 30 40
0 10 20 30 40
Organ Lémann Index


D anus (n=45)
Investigator organ resection-free cumulative damage evaluation
0 2 4 6 8 10
0 2 4 6 8 10
Organ Lémann Index


E Global (n=134)
Investigator global damage severity evaluation
0 20 40 60 80
0 20 40 60 80
Global Lémann Index


A Upper Tract (n=46)
development
validation
r=0.896
Predicted organ Lémann Index
Investigator organ resection-free cumulative damage evaluation
B Small Bowel (n=217)
development
validation
r=0.985
Predicted organ Lémann Index
Investigator organ resection-free cumulative damage evaluation
C Colon/Rectum (n=208)
development
validation
r=0.959
Predicted organ Lémann Index
Investigator organ resection-free cumulative damage evaluation
D Anus (n=104)
development
validation
r=0.880
Predicted organ Lémann Index
Investigator organ resection-free cumulative damage evaluation
E Global (n=272)
development
validation
Predicted global Lémann Index
Investigator global damage evaluation

A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1					Examination											
2					Abdominal MRI and CT if available		Clinical Exam		Upper endoscopy		Colonoscopy		Pelvic MRI and CT if available			
3				To be done (Y)												
4				Performed (Y)												
5				Date												
	Organ	Known or suspected CD location (Y)	Segment per organ	Percentage of resection	Maximal grade of severity among segmental lesions		Maximal grade of severity among segmental lesions		Maximal grade of severity among segmental lesions		Maximal grade of severity among segmental lesions		Maximal grade of severity among segmental lesions		Lémann Index	
					Stricturing	Penetrating	Stricturing	Penetrating	Stricturing	Penetrating	Stricturing	Penetrating	Stricturing	Penetrating		
					1, 2 or 3	2 or 3	1, 2 or 3	1, 2 or 3	2 or 3	1, 2 or 3	2 or 3	1, 2 or 3		1, 2 or 3		
	Upper tract		Esophagus													
			Stomach													
			Duodenum													
																Upper tract
	Small bowel		Segment 1													
			Segment 2													
			Segment 3													
			Segment 4													
			.....													
			Segment 19													
			Segment 20													
																Small bowel
	Colon/ rectum		Cecum													
			Ascending colon													
			Transverse colon													
			Descending colon													
			Sigmoid colon													
			Rectum													
																Colon/Rectum
	Anus		Anus													
																Anus
		Comment 1														
		Comment 2														Global

B

STEP	ACTION
1	Collect the percentage of previous resection for each segment from patient's history: → percentage stored in column D of the line of each segment.
2	Identify suspected or known Crohn's disease location: →Y stored in column B of each organ line.
3	Program the mandatory examinations to be performed according to Crohn's disease location; abdominal MRI, and CT if available, and perianal clinical exam for all patients, and if mandatory upper endoscopy, colonoscopy, pelvic MRI: →Y stored in line 4 and date stored in line 5 of the column of each examination, when performed.
4	Collect the most severe stricturing and the most severe penetrating lesions within each segment of each organ identified on each examination: →maximal grade for each type of lesion stored in the line of the segment of the columns E and F for MRI and CT if available, G and H for clinical exam, I and J for upper endoscopy, K and L for colonoscopy, M and N for pelvic MRI.

Table 1. Baseline characteristics of the 134 CD patients included in the validation sample

Characteristics	N = 134
Females, n (%)	57 (42)
Age (years), median (IQR)	33 (25-43)
Time since diagnosis (years), median (IQR)	5 (2-11)
Number of inclusions according to stratification factors, n	
- Upper digestive tract	13
- Small bowel (<2, 2-10, >10 years)	14/15/14
- Colon/rectum(<2, 2-10, >10 years)	14/14/14
- Perianal (<2, 2-10, >10 years)	11/11/14
Known or suspected CD location, n (%)	
- Upper digestive tract	13
- Small bowel	93
- Colon/rectum	87
- Perianal	39
Duration of the disease, n (%)	
- < 2 years	40
- ≥ 2 and < 10 years	47
- ≥ 10 years	47

CD: Crohn's disease; IQR: interquartile range; n: number of patients;

Table 2. Number of patients with different numbers of segments involved by surgical procedures, stricturing lesions, and penetrating lesions, of maximal grade (N = 134)

Organ	Total number	Without procedures and lesions	Number of segments	Maximal grade of surgical procedure				With no lesion	Maximal grade stricturing lesions*				Maximal grade penetrating lesions*			
				none	1	2	3°		none	1	2	3	none	1	2	3
Upper tract	22	10	1	21		1	0	10	16	0	2	2	13	7	1	0
			2			0	0			0	2	0		1	0	0
Small Bowel	102	16	1	78		0	12	23	25	12	39	15	79		11	9
			2			2	8			2	7	7			1	3
			3			0	3			0	3	2			0	1
			>3			0	1			0	0	0			0	0
Colon/ rectum	118	23	1	94		1	12[a]	31	57	6	25	4	37	21	21	13
			2			1	4[b]			3	8	3		8	9	1
			3			0	2[a]			3	4	0		7	3	1
			>3			0	5[c]			4	7	0		13	6	0
Anus	47	0	1	27	15	3	2	4	36	6	5	0	6	14	15	12

°segmental surgical resection [a] including two partial resections; [b] including three partial resections; [c] including one partial resection.
* Lesions of maximal grade of severity identified by anal clinical examination or MRI and CT if available or endoscopy or pelvic MRI.

Table 3: Slope and intercept of the linear relationship between the investigator organ resection-free cumulative damage evaluation and the organ Lémann index, and between the investigator global damage evaluation and the global Lémann index.

	Slope		Intercept
	estimate ± standard	p-value[1]	estimate ± standard error
Upper tract (n=22)	1.26±0.13	0.057	0.62±0.33
Small bowel (n=102)	1.24±0.04	<0.0001	0.61±0.24
Colon/Rectum (n=116)	1.19±0.04	<0.0001	0.08±0.32
Anus (n=45)	0.81±0.09	0.037	1.55±0.22
Global (n=134)	0.61±0.01	<0.0001	0.51±0.20

[1] test of slope at 1.00.
Slope of the linear relationship represents the variation in the mean of the investigator damage evaluations for a variation of 1 of the Lémann index. Thus, a slope equal to 1.00 means that a variation of Δ in the Lémann index is associated with the same Δ variation in the mean of the investigator damage evaluations. Intercept is the estimate of the mean of the investigator evaluations when Lémann index is 0.

Table 4: Estimated and rounded coefficients to be applied (a) to the numbers of segments with stricturing and penetrating lesions of each maximal grade of severity in order to calculate the organ Lémann indexes and (b) to the calculated investigator organ standardized cumulative damage evaluations in order to calculate the global Lémann index (n=272).

Organ	Type of lesion	Grade of severity	Coefficient		p-value	Rounded coefficient
			Estimate	SE		
(a) Organ indexes						
Upper tract (n=46)	Stricturing	1	-	-	-	-
		2	3.51	0.35	< 0.001	3.5
		3	5.16	0.53	< 0.001	5.0
	Penetrating	1	0.75	0.25	0.004	1.0
		2-3	1.43	0.54	0.011	1.5
Small bowel (n=217)	Stricturing	1	1.20	0.21	<0.001	1.0
		2	2.92	0.15	< 0.001	3.0
		3	4.97	0.11	< 0.001	5.0
	Penetrating	2	1.56	0.19	< 0.001	1.5
		3	3.88	0.22	< 0.001	4.0
Colon and rectum (n=208)	Stricturing	1	0.54	0.18	0.003	0.5
		2	2.18	0.15	< 0.001	2.0
		3	5.18	0.44	< 0.001	5.0
	Penetrating	1	1.09	0.11	< 0.001	1.0
		2	2.58	0.15	< 0.001	2.5
		3	4.56	0.35	< 0.001	4.5
Anus (n=104)	Stricturing	1	-	-	-	-
		2	1.80	0.29	< 0.001	2.0
		3	3.47	0.71	< 0.001	3.5
	Penetrating	1	-	-	-	-
		2	2.26	0.22	< 0.001	2.5
		3	3.05	0.27	< 0.001	3.0
(b) Global index						
Calculated investigator organ standardized cumulative damage evaluation	Upper digestive tract		0.182	0.051	<0.001	2.0
	Small bowel		0.0380	0.041	<0.001	4.0
	Colon/Rectum		0.290	0.017	< 0.001	3.0
	Anus		0.271	0.016	< 0.001	2.5

SE: standard error