THERAPEUTIC ADVANCES in
Gastroenterology

Original Research

Composite outcomes in observational studies of Crohn's disease: a systematic review and meta-analysis

Fernando Magro*, Catarina Sottomayor*, Catarina Alves, Mafalda Santiago, Paula Ministro, Paula Lago, Luís Correia, Raquel Gonçalves, Diana Carvalho, Francisco Portela, Cláudia Camila Dias, Axel Dignass, Silvio Danese, Laurent Peyrin-Biroulet, Maria Manuela Estevinho and Paula Leão Moreira; on behalf GEDII (Portuguese IBD Group)

Ther Adv Gastroenterol

2022, Vol. 15: 1–19

DOI: 10.1177/17562848221092754

Correspondence to:
Fernando Magro
Department of Biomedicine, Unit of Pharmacology and Therapeutics, Faculty of Medicine, University of Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal

Department of Gastroenterology, São João University Hospital Center (CHUSJ), Porto, Portugal
fm@med.up.pt

Department of Clinical Pharmacology, São João University Hospital Center (CHUSJ), Porto, Portugal

Catarina Sottomayor
Catarina Alves
Faculty of Medicine, University of Porto, Porto, Portugal

Mafalda Santiago
Center for Health Technology and Services Research (CINTESIS), Porto, Portugal

Paula Ministro
Department of Gastroenterology, Tondela-Viseu Hospital Centre, Viseu, Portugal

Paula Lago
Department of Gastroenterology, Santo António University Hospital Center (CHUPorto), Porto, Portugal

Luís Correia
Department of Gastroenterology, Lisbon North Hospital Centre, Santa Maria Hospital, Lisbon, Portugal

Raquel Gonçalves
Gastroenterology Department, Hospital de Braga, Braga, Portugal

Diana Carvalho
Department of Gastroenterology, Santo António dos Capuchos Hospital at Centro Hospitalar Lisboa Central, Lisboa, Portugal

Abstract

Background: This systematic review and meta-analysis aims to assess composite and aggregate outcomes of observational studies in Crohn's disease and to evaluate whether the number and type of variables included affect the frequency of the outcome.

Methods: MEDLINE [via PubMed], Scopus and Web of Science were searched to identify observational studies that enrolled patients with Crohn's disease and evaluated a composite or aggregate outcome. The proportion of patients achieving the outcome was determined and a random-effects meta-analysis was performed to evaluate how the frequency of each outcome varies according to the reporting of predefined variables.

Results: From 10,257 identified records, 46 were included in the qualitative analysis and 38 in the meta-analysis. The frequency for composite and aggregate outcomes was 0.445 [95% confidence interval (CI): 0.389–0.501] and 0.140 (95% CI: 0.000–0.211), respectively. When comparing composite outcomes by number of included variables, the frequency was 0.271 (95% CI: 0.000–0.405) and 0.698 (95% CI: 0.651–0.746), for one and six variables, respectively. The frequency of the composite outcome varied according to the identity of the variables being reported. Specific pairs of predefined variables had a significant effect in the frequency of composite outcomes.

Conclusion: Composite outcomes with increasing number of predefined variables show an increase in frequency. Outcomes including variables such as 'Surgery' and 'Steroids' had higher frequencies when compared with the ones that did not include these variables. These results show that the frequency of composite outcomes is dependent on the number and type of variables being reported.

Keywords: composite outcomes, Crohn's disease, observational studies

Received: 24 November 2021; revised manuscript accepted: 18 March 2022.

Introduction

Observational studies can provide clinicians and policymakers in the health sector with valuable information about the most effective approach in the management of patients diagnosed with a chronic disease.[1] Because of the elevated costs and complex logistics required to monitor patients in the course of these studies, they often exhibit broad differences in their basic design, namely size and duration of the study as well as number and type of endpoint variables being considered for reporting. This heterogeneity in study design becomes a confounding factor at the time of drawing conclusions that could have relevance in a clinical setting.

Francisco Portela
Department of Gastroenterology, University Hospital Centre of Coimbra, Coimbra, Portugal

Cláudia Camila Dias
Center for Health Technology and Services Research (CINTESIS), Porto, Portugal

Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

Axel Dignass
Department of Medicine I, Agaplesion Markus Hospital, Frankfurt, Germany

Silvio Danese
Department of Biomedical Sciences, Humanitas University, Milan, Italy

Inflammatory Bowel Disease (IBD) Center, Department of Gastroenterology, Humanitas Clinical and Research Center (IRCCS), Milan, Italy

Laurent Peyrin-Biroulet
Department of Gastroenterology and Inserm NGERE U1256, University Hospital of Nancy, University of Lorraine, Vandoeuvre-lès-Nancy, France

Maria Manuela Estevinho
Department of Biomedicine, Unit of Pharmacology and Therapeutics, Faculty of Medicine, University of Porto, Porto, Portugal

Department of Gastroenterology, Centro Hospitalar Vila Nova de Gaia/Espinho, Vila Nova de Gaia, Portugal

Paula Leão Moreira
Department of Clinical Pharmacology, São João University Hospital Center (CHUSJ), Porto, Portugal

*Fernando Magro and Catarina Sottomayor shared co-first authorship

Composite and aggregate outcomes are a common strategy employed in the design of observational studies. This strategy, in which outcome is classified either by the presence of any one or by the combination of every individual variable under assessment, is especially useful to maximize the statistical power of a study and overcome limitations related to size of the patient population.[2] Composite and aggregate outcomes have a long tradition in studies related to cardiovascular disease, and the heterogeneity of these outcomes can lead to conflicting conclusions.[3]

Crohn's disease (CD) is a chronic inflammatory gastrointestinal condition that displays remarkable heterogeneity in terms of symptoms, age of onset and disease location. Along with ulcerative colitis (UC), it constitutes the main component of inflammatory bowel disease (IBD), and both its incidence and prevalence have been steadily rising worldwide, although the actual causes for this scenario remain unclear.[4] Consequently, CD has been the focus of numerous observational studies over the years. To tackle the inherent heterogeneity that this body of literature represents, the present meta-analysis was performed with the specific aims of characterizing the frequency of composite and aggregate outcomes included in observational studies on CD and to determine how the number and type of variables reported in the individual studies affect these parameters.

Materials and methods

Search strategy

The bibliographic search was conducted following the Cochrane Collaboration Guidelines for Systematic Reviews[5] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines.[6] Published studies were retrieved using three electronic databases: MEDLINE (via PubMed), Scopus, and Web of Science. The literature search was carried out from inception to 14 July 2020, using the following keywords or medical subject heading (MeSH) terms: [((('aggressive disease') OR ('disabling disease') OR ('disabling outcome') OR ('disabling outcomes') OR ('composite outcome') OR ('composite outcomes') OR ('composite event') OR ('composite events') OR ('composite endpoint') OR ('composite endpoints') OR ('composite') OR (composit*) OR ('progressive disease')) AND (('Colitis, Ulcerative' (MeSH Terms)) OR (ulcerative colitis) OR ('Crohn Disease' (MeSH Terms)) OR (Crohn's disease) OR ('Inflammatory Bowel Diseases' (MeSH Terms)))]. This query was used for PubMed search and adjusted for the other databases. To ensure that all pertinent articles were included, the reference lists of the systematic reviews selected from the databases were manually reviewed.

Eligibility criteria

Any study enrolling both adults and children previously diagnosed with CD using clinical, endoscopic and/or pathological features was considered eligible for inclusion in this systematic review. The inclusion criteria were: (1) cohort, case–control and cross-sectional studies with CD patients; (2) studies evaluating composite or aggregate outcomes; and (3) outcomes representing CD progression. No restrictions in terms of publication dates were applied. The exclusion criteria were: (1) randomized controlled trials and *post hoc* analyses, systematic reviews and meta-analyses, review articles, descriptive and diagnostic studies, animal and *in vitro* studies, study protocols, guidelines, editorials and only abstracts available; (2) studies selecting patients with diseases other than CD; studies evaluating only UC patients; (3) studies that did not define a composite or aggregate outcome of interest; (4) studies reporting an improvement outcome; and (5) studies that did not differentiate between CD and UC in the results.

Study selection and data collection

The studies retrieved from the electronic databases were independently screened by two reviewers. Any study whose title and abstract clearly indicated that it failed to meet the previously described selection criteria was immediately excluded from further analysis. For all the other studies, the full text was considered to determine its inclusion or exclusion. The following information was collected from the selected studies: authors; country of origin and study design; publication year; observation period; number of patients selected [CD and UC]; CD location; cohort's exposure and comparison; outcome definitions and included variables; proportion of patients achieving the defined outcome. The proportion of patients achieving each variable of the outcome was not assessed. The observation

period refers to the mean or median time of follow-up, duration of follow up or the time of occurrence of the outcome, when available. The variables considered in the analysis were selected as being the most clinically relevant parameters in IBD assessment but did not necessarily include every variable reported in the individual studies. Strict definitions for each variable were established from the beginning and used to determine if any given variable was included or not in each study. Each variable was composed of a single or multiple parameters.

Endpoints under analysis

A composite outcome was defined as the presence of one or more parameters. Under this definition, to achieve the outcome, patients needed to present at least one parameter, but these may have been included within a single variable. An aggregate outcome was defined as the simultaneous presence of at least two of the parameters considered.[7] The outcomes represented disease progression/disabling disease/therapy failure and included the following 10 variables: Clinical evaluation, Events, Surgery, Hospitalization, Steroids, Biologics, Immunomodulators, Therapy modification, Biomarkers and Endoscopic assessment. Clinical evaluation was defined as reported clinical symptoms or manifestations of CD, extraintestinal manifestations or other clinical aspects, disease activity evaluation with any imaging modality or increase/no change in CD clinical scores. Events was defined as reported CD-related events such as stenosis, fistula or abscess, or change in behaviour according to the Montreal Classification (B2 or B3). Surgery was defined as at least one reported surgical intervention for any cause. Hospitalization was defined as at least one reported inpatient stay for any cause. Steroids was defined as reported *de novo* use of corticosteroids, dose increase, change in corticosteroid drug, or dependency or refractoriness to corticosteroids. Biologics was defined as reported *de novo* use, switch, dose or treatment frequency alteration, or cessation of biological therapy. Immunomodulators was defined as reported *de novo* use, switch or dose increase of immunomodulators, or unspecified immunosuppressive therapy. Therapy modification was defined as reported non-specified medication adjustments for any reason including increase or *de novo* CD-related symptoms or manifestations, or increase in CD activity. Biomarkers was defined as reported evaluation and increase or no change in CD-related biomarkers (C-reactive protein and faecal calprotectin). Endoscopic assessment was defined as reported endoscopic scores or any endoscopic activity change.

Quality assessment

The methodological quality for each study was assessed using the validated Critical Appraisal Skills Programme (CASP) for cohort studies.[8] This validated tool allows assessing and interpreting evidence by systematically assessing its validity, results and relevance. This tool includes 12 categories, each evaluated using a colour scheme: (1) green, if the study met all the parameters included in each item; (2) yellow, if the study met the parameters partially or if it did not have enough information; (3) red, if the study did not meet the parameters included in each item.

Statistical analysis

The main data analysed in this meta-analysis were the proportions of patients achieving composite or aggregate outcomes. The proportion of patients achieving either outcome was compared between study subgroups reporting or not reporting the predefined variables (see the section 'Study selection and data collection'). The following comparisons between subgroups were performed: (1) composite *versus* aggregate outcome; (2) composite outcome by number of variables; (3) composite outcome by presence of each predefined variable; and (4) composite outcome by combination of two or three variables.

To perform the meta-analysis, the 'metaprop' function from the 'meta' package of the R statistical programming language was used.[9] For the pooling of studies, the 'PRAW' summary measure was implemented. Due to the differences observed across studies, a random-effects model was applied. Statistical heterogeneity was assessed using both Cochran's Q test and the I^2 statistic, which estimate the presence of heterogeneity among studies.[10] In addition, Egger's test was used to detect potential publication biases[11] and a sensitivity analysis was performed to assess the influence of any individual study on the overall results.

A Venn diagram and Upset plot were generated using the 'UpsetR' and 'nVennR' packages

included within the R software, to graphically illustrate the distribution of the predefined variables among the individual studies included in the meta-analysis.

All analyses and charts were executed using R software version 4.1.0. A p-value lower than 0.05 was considered statistically significant.

Results

Literature search and study selection

The electronic database search yielded 10,250 records (1885 in MEDLINE, 4323 in Scopus and 4042 in Web of Science); the manual search identified seven additional studies. Following the removal of duplicates ($n = 4444$), 5813 records remained, of which 5582 were excluded. The remaining 231 records were evaluated for eligibility. Following full-text assessment, 185 articles were excluded, 46 articles were selected for inclusion[12–57] in the qualitative analysis and 38 in the meta-analysis[12,13,14,16,17–20,22–34,37,38,40,42,44–52,54–57] (Figure 1).

Quality assessment

The evaluation obtained with CASP Checklists for cohort studies showed that all included studies clearly stated the issue evaluated (Supplementary Table 1). However, due to their observational character and non-randomized selection of patients, all studies showed relevant issues in how the cohort was recruited, potentially introducing selection biases. Most studies had complete and long enough follow-up times, reducing the probability of selection bias due to loss of follow-up.[12,14,16,18,20,26,30,31,34,35,36,38,40,42,44,46,47,51,54–56] In addition, most studies also addressed the most important confounding factors.[14,16,17,22,26,28–30,33,38,39,41,48,49,51,52,56,57] The results of a few studies did not fit well with other available evidence.[14,17,18,19,28,30,32,38,40,41,52,56]

Characteristics of included studies

Study characteristics are summarized in Supplementary Table 2. Thirty-six studies[12,13,14,15,16,18–26,28–32,34,37–42,44–47,48,49,51–53,55] considered patients only affected by CD, while 10 also included patients with UC.[17,27,33,35,36,43,50,54,56,57] Three studies evaluated only paediatric IBD patients[20,39,53] whereas 20 studies assessed adult patients exclusively.[12,16,17,21,23,24,29,31,34,35,37,41,44,46–50,55,57] Some scientific articles[12,20,23–25,34,54] included more than one outcome. In those cases, each outcome was considered independently for the purpose of this analysis. The number of patients included in each study and the observation period varied widely, ranging from 51[23] to 10,367[17] and from 30 days[21,54] to 16 years,[44] respectively. Forty-nine composite outcomes were registered from a total of 43 studies[12,13,14,15,16,17–21,22–27,29,31–33,34–43,44–52,54,55–57] and four aggregate outcomes from four studies.[20,28,30,53] The composite and aggregate outcomes included in the meta-analysis were heterogeneous regarding the predefined variables (Supplementary Table 3): nine outcomes had clinical evaluation,[20,23,24,27,34,49,52] 16 outcomes[16,19,20,24–26,31,34,37,38,44,46,47,56,57] had events, 3 outcomes included endoscopic assessment,[24,28,30] 32 outcomes [13,16,17,18–20,22–25,26,27,29,31,33,34,38,40,42,44,46–48,50–52,54,55,57] had surgery, 19 outcomes [13,20,22,23,27,28,31,34,38,44,46–48,51,52,54,55,57] had hospitalization, 18 outcomes [13,20,22,23,27,28,31,34,38,44,46–48,51,52,54,55,57] included steroids, 17 outcomes [13,14,20,22,23,31–34,40,44,46–48,50,52,54] had immunomodulators, 23 outcomes[12,14,17,22,23,25,29,31–33,40,42,44–48,50,51,54,55,57] had biologics, 5 outcomes[23,28,30,48,49] had therapy modification and four outcomes[23,24,28,30] had biomarkers (Supplementary Figure 1). The number of included variables was also highly variable between outcomes included in the meta-analysis: seven of them reported a single variable[12,18,25,37,45,56] while 17 outcomes[13,20,22–24,27,28,31,34,38,44,46–48,51,52,57] reported four or more variables (Supplementary Figure 2).

Composite and aggregate outcomes

The frequency for composite outcomes was 0.445 [95% confidence interval (CI): 0.389–0.501, $I^2 = 99\%$]. This value was predictably lower in the case of aggregate outcomes (0.140, 95% CI: 0.000–0.211, $I^2 = 84\%$), reflecting the more stringent conditions to achieve this outcome (Figure 2).

The results of Egger's test on the frequency of composite outcomes were not significant ($p = 0.103$), indicating that the dataset was unbiased. This could also be visually appreciated by the symmetry of the corresponding funnel plot in which the standard error was plotted against the outcome frequency for each study (Supplementary Figure 3). The sensitivity analysis for the same dataset failed to reveal the existence of any outlier

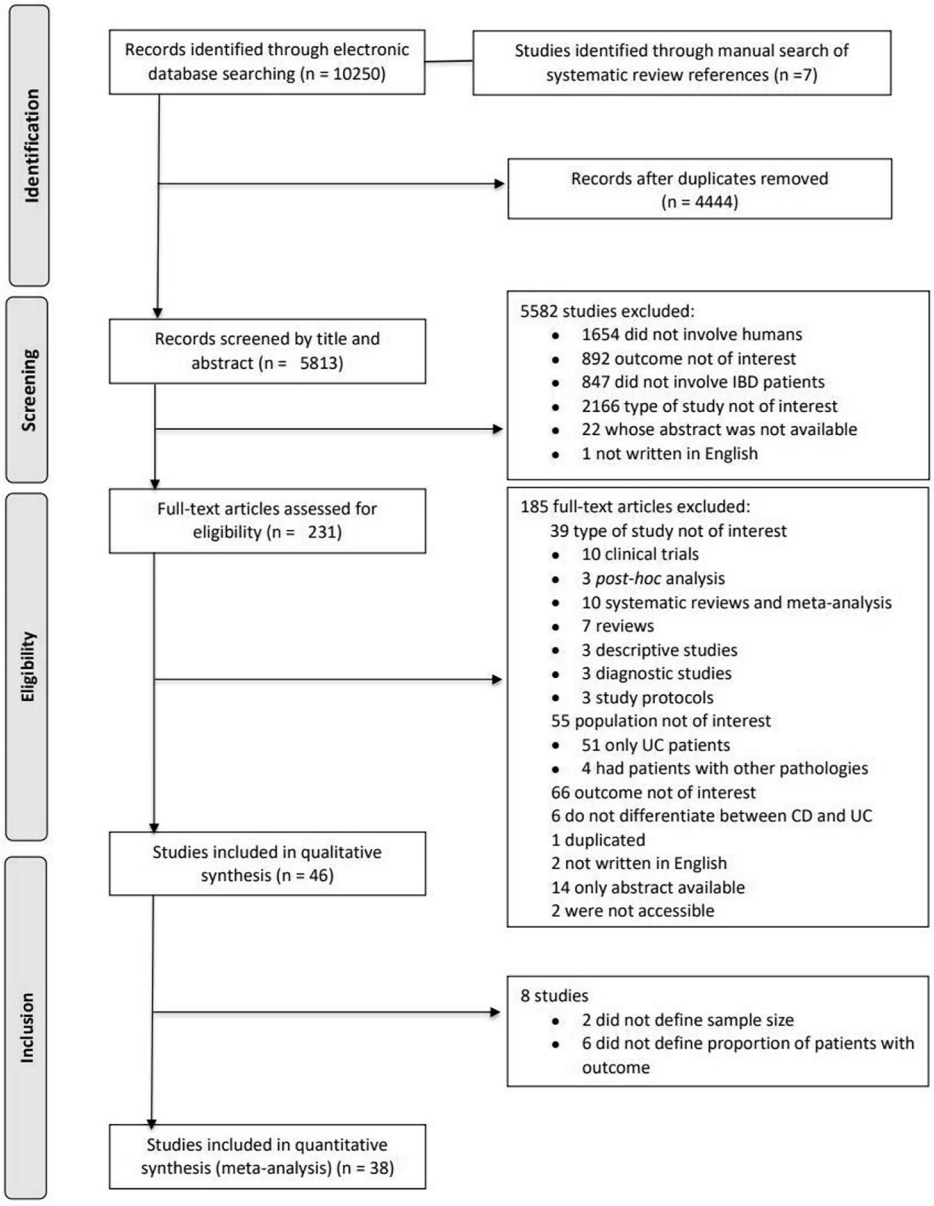

Figure 1. Flow diagram of study selection and data collection process.
CD, Crohn's disease; IBD, inflammatory bowel disease; UC, ulcerative colitis.

among the studies included in the analysis, with the frequency remaining unaltered after the sequential exclusion of each individual study (Supplementary Figure 4), confirming the robustness of the result.

Subgroup analysis outcomes

When the frequency of composite outcomes was discriminated according to the total number of variables reported in the study, significant differences emerged between the subgroups (Figure 3). Specifically, the subgroup of outcomes that reported a single variable exhibited the lowest frequency of composite outcomes (0.271; 95% CI: 0.000–0.405, $I^2 = 97\%$), while those that included five and six variables had a significantly higher frequency (0.722; 95% CI: 0.603–1.000; $I^2 = 98\%$ and 0.698; 95% CI: 0.651–0.746; $I^2 = 86\%$, respectively) in comparison to the rest of the subgroups (Figure 3, Supplementary Figure 5). When the frequency of

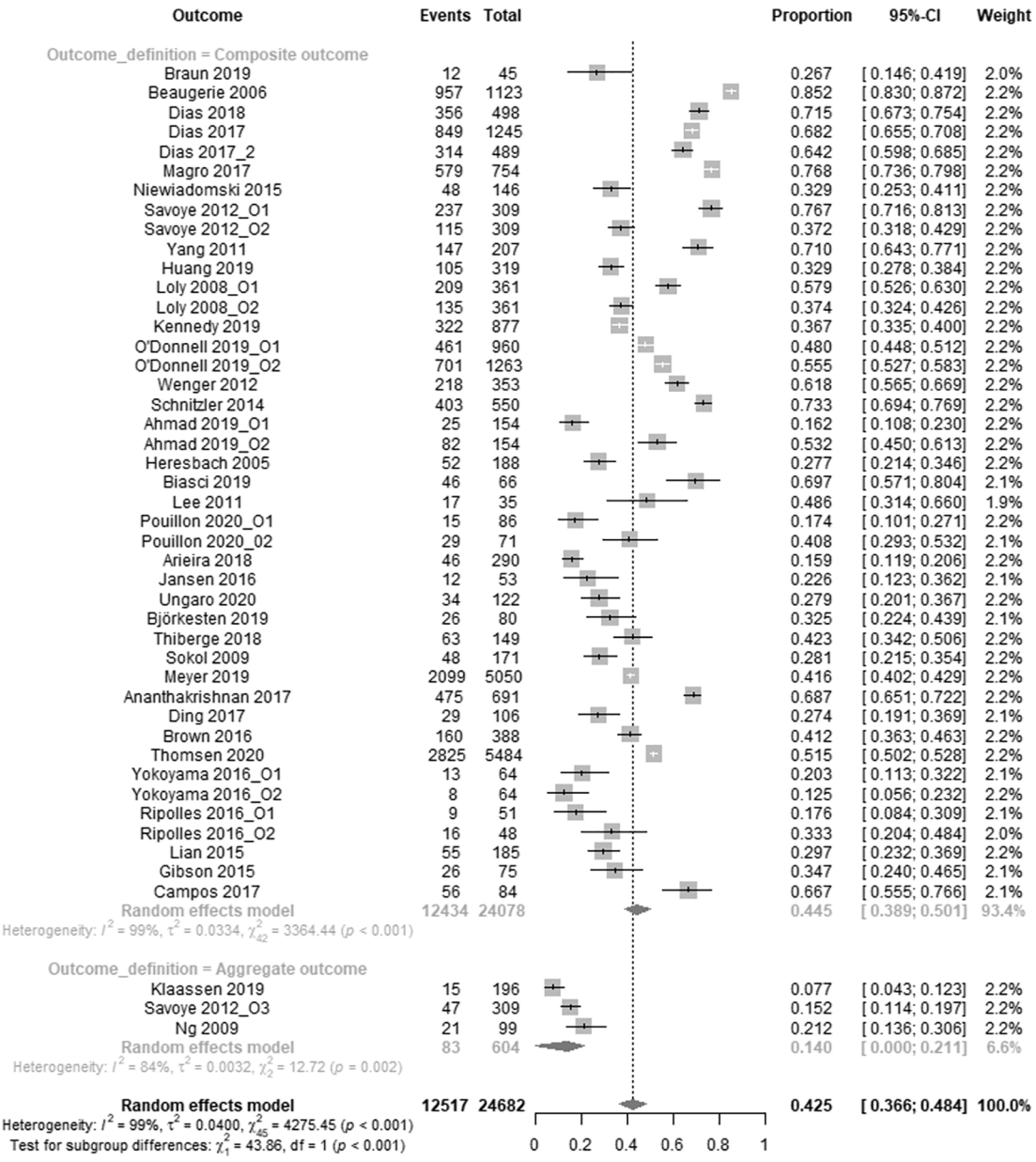

Figure 2. Frequency of composite and aggregate outcomes and corresponding 95% confidence interval ($n = 46$).

composite outcomes for each individual subgroup was compared with that corresponding to all the other subgroups considered together, the average number for the subgroups reporting five and six variables was again significantly higher ($p < 0.001$ in both cases), whereas no significant differences were identified in the case of the other four subgroups (Supplementary Figure 5).

Influence of specific variables on composite outcome frequencies

The reporting of specific variables had an effect on the frequency of composite outcomes. The subgroup of outcomes including the variable 'Surgery' exhibited a significantly higher frequency of composite outcomes *versus* the subgroup that did not include this variable (Yes:

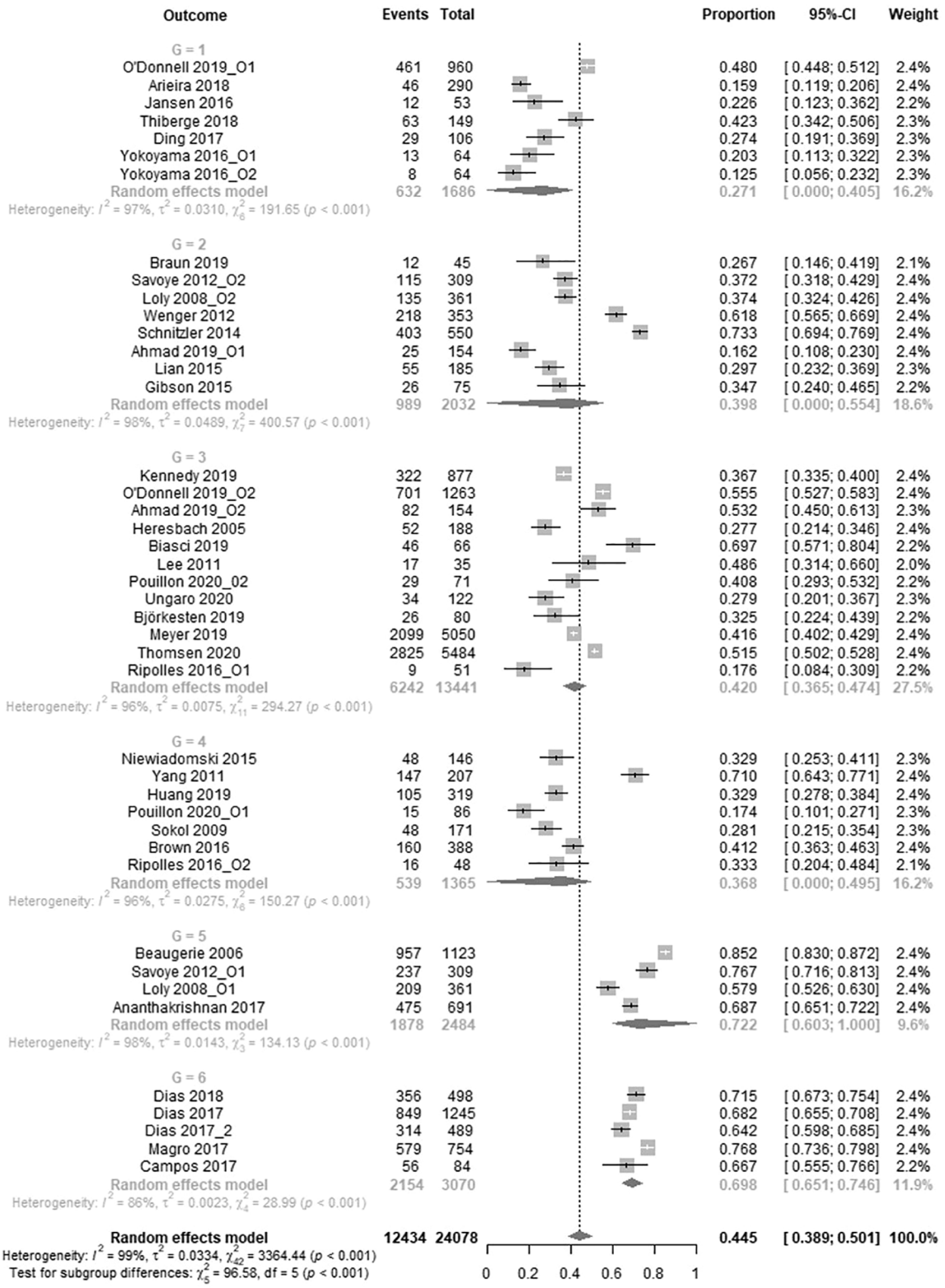

Figure 3. Subgroup analysis: frequency of composite outcomes and 95% confidence interval (CI) according to the number of variables (G = number of variables) reported in the outcome (n = 43).

0.494, 95% CI: 0.431–0.557, $I^2 = 99\%$; No: 0.316, 95% CI: 0.000–0.421, $I^2 = 97\%$; $p = 0.004$). The differences were also significant for the following variables: 'Hospitalization' (Yes: 0.530, 95% CI: 0.448–0.613, $I^2 = 99\%$; No: 0.375, 95% CI: 0.000–0.449, $I^2 = 97\%$; $p = 0.006$); 'Steroids' (Yes: 0.569, 95% CI: 0.488–0.650, $I^2 = 98\%$; No: 0.364, 95% CI: 0.311–0.417, $I^2 = 98\%$; $p < 0.001$) and 'Immunomodulators' (Yes: 0.586, 95% CI: 0.506–0.666, $I^2 = 98\%$; No: 0.355, 95% CI: 0.302–0.408, $I^2 = 98\%$; $p < 0.001$; Figure 4). No significant differences between the two subgroups were identified for the predefined variables 'Clinical Evaluation', 'Events', 'Endoscopic Assessment', 'Biologics', 'Therapy Modification' and 'Biomarkers' (Supplementary Figure 6).

When the study subgroups were created based on the reporting of paired variables rather than individual variables, the presence of the following variable pairs had a significant effect in increasing the frequency of composite outcomes: 'Biologics' and 'Biomarkers', 'Biologics' and 'Therapy modification', 'Clinical Evaluation' and 'Immunomodulators', 'Clinical Evaluation' and 'Steroids', 'Events' and 'Biologics', 'Events' and 'Immunomodulators', 'Events' and 'Steroids', 'Hospitalizations' and 'Biologics', 'Hospitalization' and 'Immunomodulators', 'Hospitalization' and 'Steroids', 'Hospitalization' and 'Therapy Modification', 'Immunomodulators' and 'Therapy Modification', 'Surgery' and 'Biologics', 'Surgery' and 'Hospitalization', 'Surgery' and 'Immunomodulators', 'Surgery' and 'Steroids', 'Surgery' and 'Therapy Modification', 'Steroids' and 'Biologics', 'Steroids' and 'Immunomodulators', 'Steroids' and 'Therapy Modification'. On the contrary, the following paired variables significantly decreased the frequency of composite outcomes: 'Clinical Evaluation' and 'Biomarkers', 'Clinical Evaluation' and 'Events', 'Clinical Evaluation' and 'Therapy Modification', 'Therapy Modification' and 'Biomarkers'. The entire set of comparisons performed using paired variables and their corresponding statistical significance are summarized in Table 1.

The analysis of the data based on the simultaneous reporting of three different variables showed a significant effect on the frequency of composite outcomes by several combinations. When 'Clinical Evaluation' was combined at the same time with 'Events' and 'Hospitalization', with 'Events' and 'Surgery' and with 'Therapy Modification' and 'Biomarkers' the frequency of composite outcomes decreased. On the contrary, many three-variable combinations had the opposite effect, significantly increasing the frequency of composite outcomes compared with their respective control subgroups. These values and their statistical significance are summarized in Table 2.

Discussion

This study assesses the relative impact of composite and aggregate outcomes in patients with CD via a comprehensive meta-analysis of the observational studies available in the literature. An important aim was to determine how the specific set of variables being reported by a particular study affects the reporting of composite and aggregate outcomes, which in turn has important implications in the development of clinical guidelines for the management of the disease.

One of the first conclusions made evident by the collected metadata is the remarkable heterogeneity presented by these observational studies, both in terms of total number of patients considered and on the total duration of the included studies. Perhaps even more importantly, this heterogeneity also extended to the total number and the identity of the variables employed to assess disease outcome. Around 16% of the outcomes included in the meta-analysis after passing the eligibility criteria (7 out of 43) reported a single variable. It is to be expected that the frequency of these outcomes would steadily increase as the number of variables being reported goes up, and that is indeed what was observed in the present meta-analysis. The frequency of these outcomes was the highest for those studies that included five and six variables, but no evident increase was observed in the transition between five and six (Supplementary Figure 7). This suggests that the frequency may reach a plateau when five different variables are considered, beyond which further improvement would be only marginal. Therefore, based on the present set of results, five variables would be the most appropriate number to maximize the frequency of composite outcomes.

In addition to the total number of variables, the choice between specific variables that were reported had consequences on the frequency of

(a)

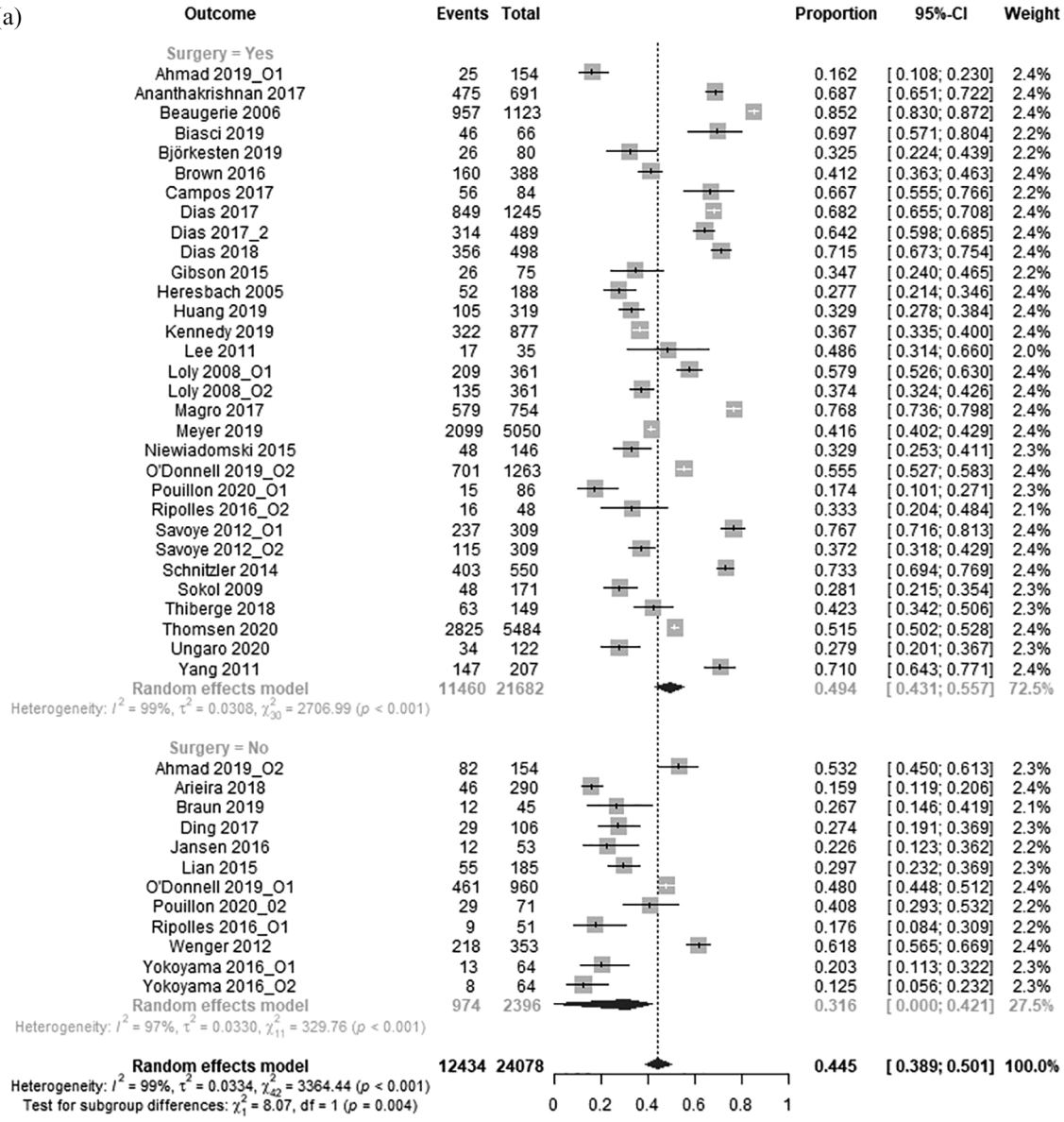

Figure 4. *(Continued)*

(b)

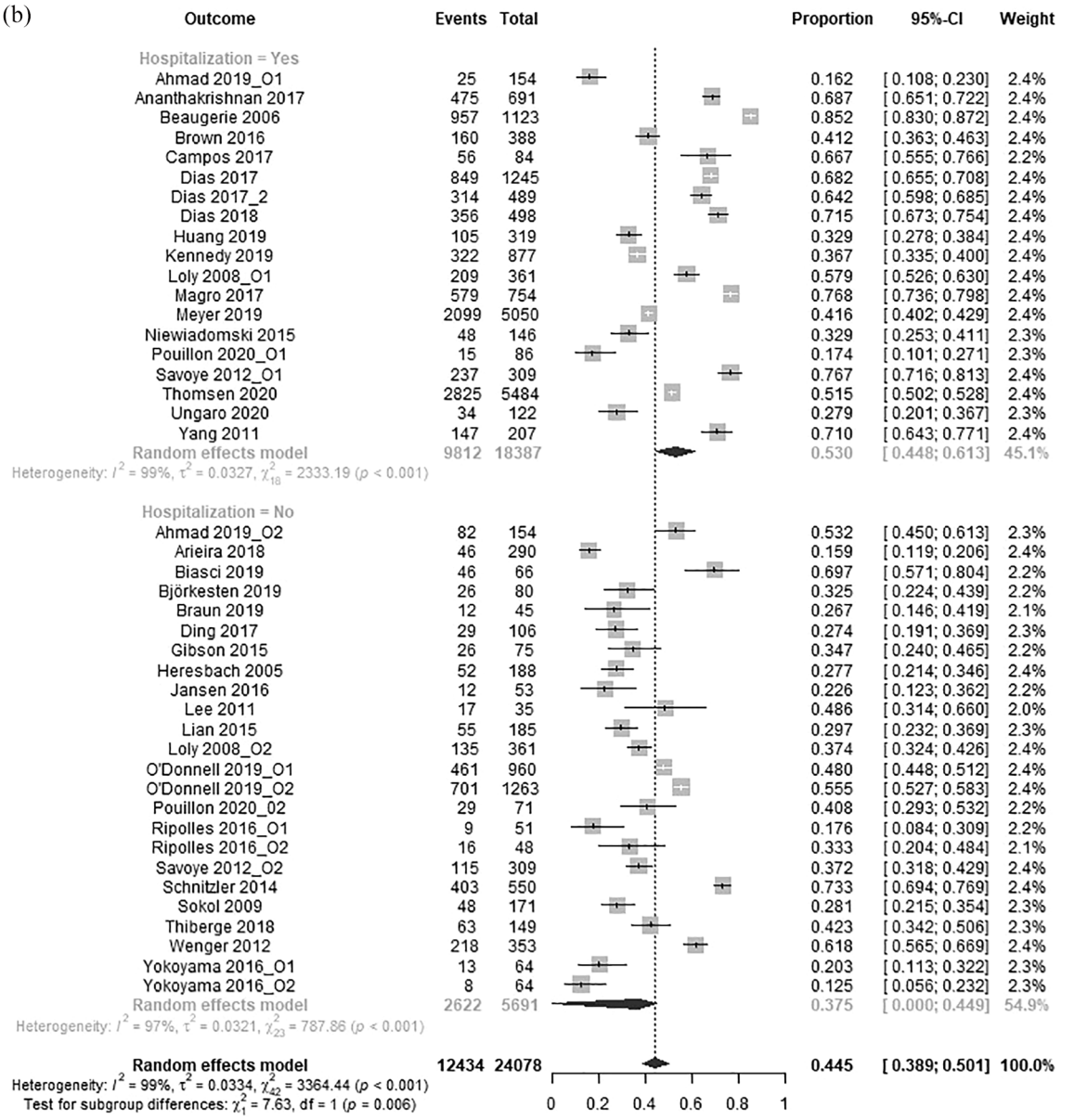

Figure 4. *(Continued)*

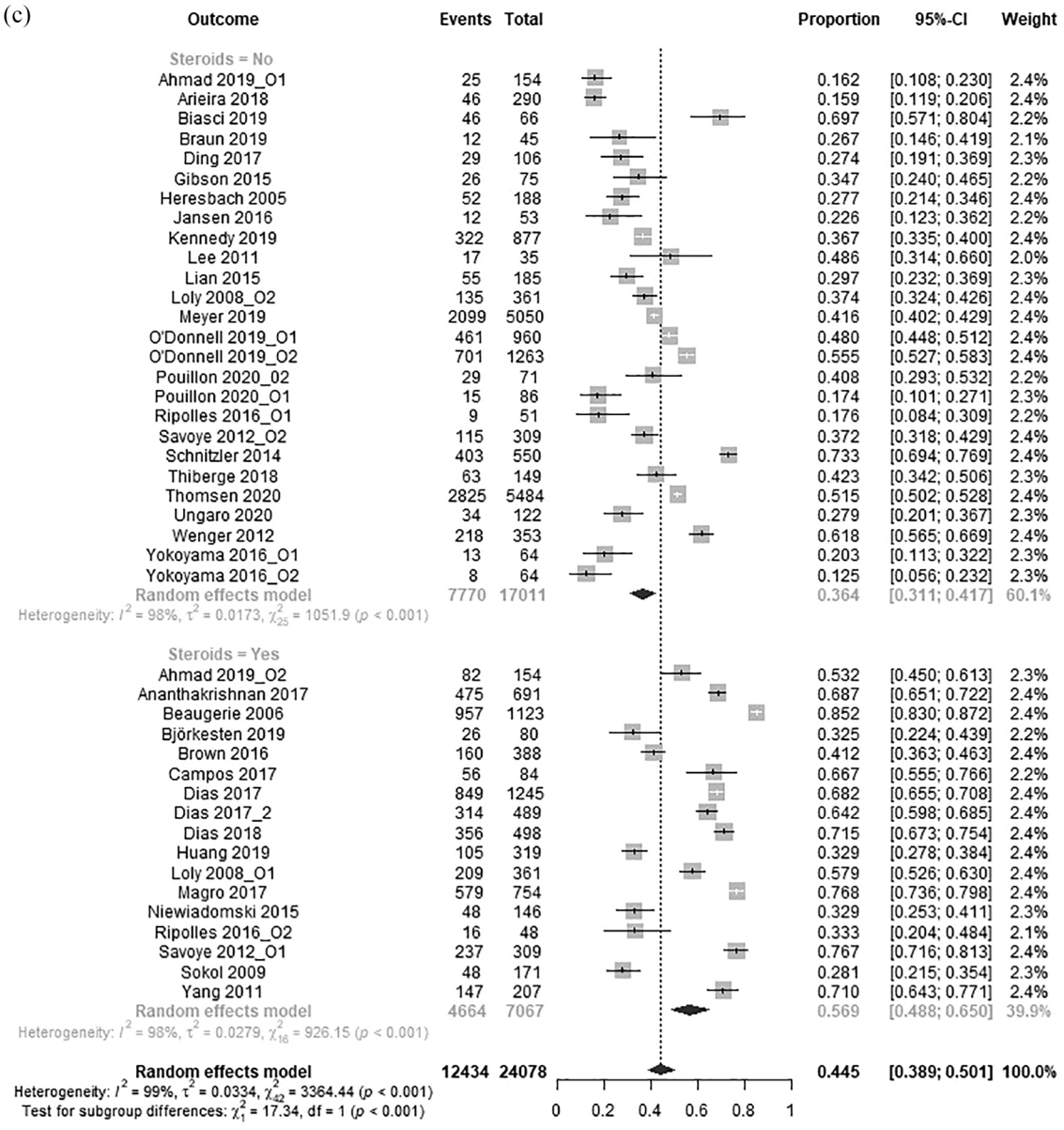

Figure 4. *(Continued)*

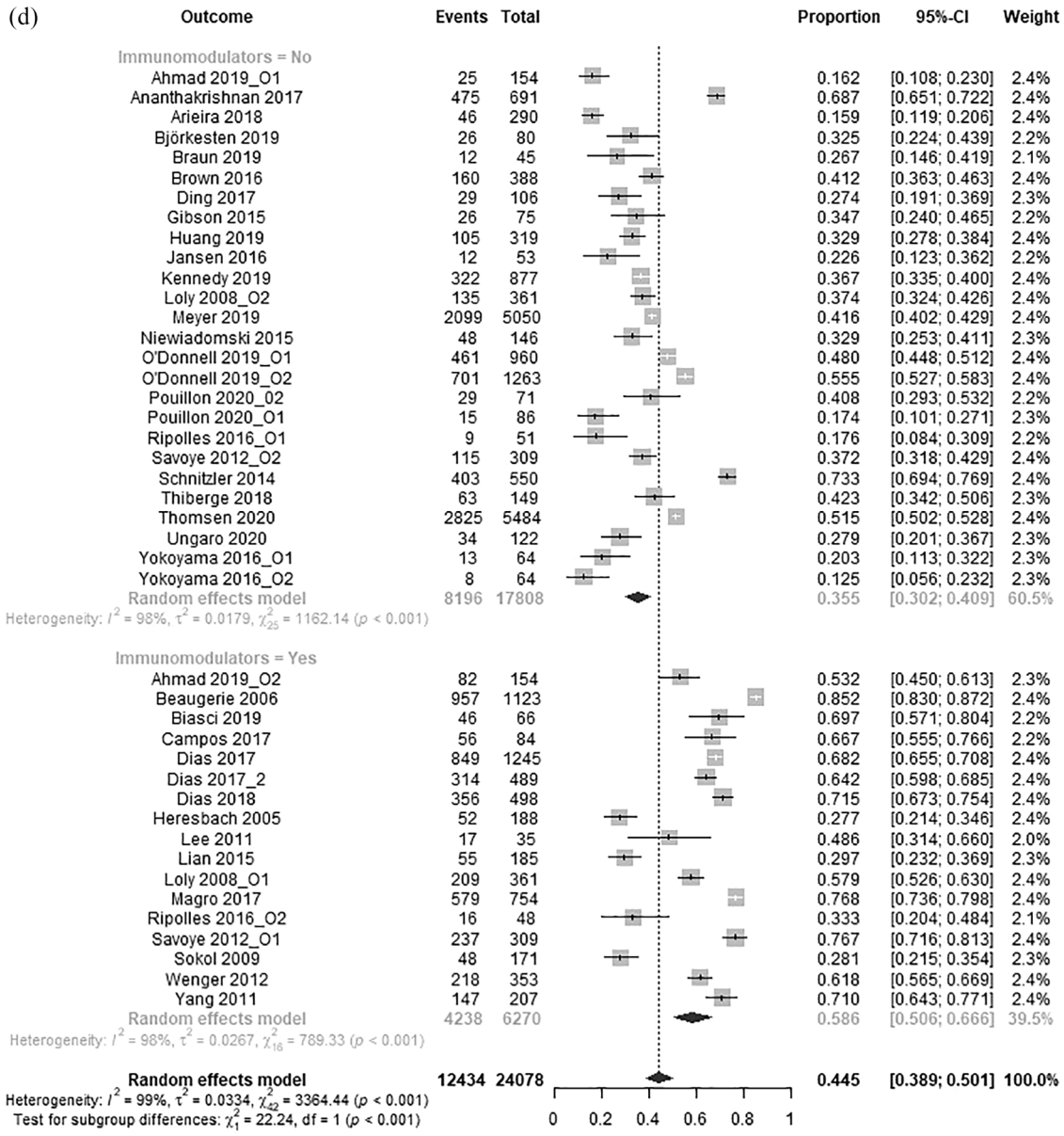

Figure 4. Subgroup analysis: frequency of composite outcomes and 95% confidence intervals (CIs) according to the presence of individual predefined variables reported in the outcome ($n = 43$). (a) Subgroups determined by the presence or absence of the variable 'Surgery'; (b) subgroups determined by the presence or absence of the variable 'Hospitalization'; (c) subgroups determined by the presence or absence of the variable 'Steroids'; and (d) subgroups determined by the presence or absence of the variable 'Immunomodulators'.

composite outcomes. In this study, some of the previously defined variables increased the frequency of composite outcomes when included on a given observational study. These included 'Surgery', 'Hospitalization', 'Steroids' and 'Immunomodulators'. In contrast, the reporting of the remaining variables ('Clinical Evaluation', 'Events', 'Endoscopic Assessment', 'Biologics', 'Therapy Modification' and 'Biomarkers') appear to have little effect by themselves. It should be noticed, however, that the inclusion of certain combinations of variables that do not have an effect by themselves nevertheless increases the frequency of composite outcomes. Two main conclusions can be derived from these results: (1) the frequency of composite outcomes appears to

Table 1. Subgroup analysis.

First variable	Second variable	Number of outcomes	Frequency of composite outcome in subgroup		Significance
			Both variables reported	At least one not reported	
Biologics	Biomarkers	1	0.715 [0.675; 0.755]*	0.438 [0.382; 0.495]	61.74, $p < 0.001$
	Immunomodulators	12	0.525 [0.425; 0.626]	0.414 [0.349; 0.480]	3.27, $p = 0.070$
	Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
Clinical evaluation	Biologics	1	0.408 [0.294; 0.523]	0.446 [0.389; 0.502]	0.33, $p = 0.570$
	Biomarkers	1	0.176 [0.072; 0.281]*	0.451 [0.395; 0.507]	20.55, $p < 0.001$
	Endoscopic assessment	1	0.408 [0.294; 0.523]	0.446 [0.389; 0.502]	0.33, $p = 0.570$
	Events	1	0.174 [0.094; 0.255]*	0.451 [0.395; 0.507]	30.80, $p < 0.001$
	Hospitalization	5	0.543 [0.000; 1.000]	0.432 [0.381; 0.483]	0.89, $p = 0.350$
	Immunomodulators	3	0.734 [0.000; 1.000]*	0.423 [0.373; 0.474]	13.73, $p < 0.001$
	Surgery	5	0.543 [0.000; 1.000]	0.432 [0.381; 0.483]	0.89, $p = 0.350$
	Steroids	4	0.634 [0.000; 1.000]*	0.426 [0.375; 0.477]	4.11, $p = 0.040$
	Therapy modification	2	0.213 [0.126; 0.300]*	0.455 [0.398; 0.512]	20.89, $p < 0.001$
Endoscopic assessment	Biomarkers	1	0.408 [0.294; 0.523]	0.446 [0.389; 0.502]	0.33, $p = 0.570$
Events	Biologics	6	0.675 [0.611; 0.738]*	0.406 [0.344; 0.468]	35.43, $p < 0.001$
	Hospitalization	9	0.519 [0.398; 0.640]	0.425 [0.362; 0.488]	1.82, $p = 0.180$
	Immunomodulators	4	0.703 [0.651; 0.754]*	0.418 [0.359; 0.476]	51.54, $p < 0.001$
	Surgery	13	0.516 [0.424; 0.609]	0.413 [0.343; 0.483]	3.03, $p = 0.080$
	Steroids	6	0.639 [0.544; 0.734]*	0.413 [0.353; 0.472]	15.55, $p < 0.001$
Hospitalization	Biologics	9	0.611 [0.521; 0.701]*	0.399 [0.000; 0.479]	11.86, $p < 0.001$
	Immunomodulators	9	0.711 [0.649; 0.773]*	0.374 [0.327; 0.421]	72.17, $p < 0.001$
	Steroids	13	0.628 [0.545; 0.710]*	0.364 [0.315; 0.414]	28.87, $p < 0.001$
	Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
Immunomodulators	Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
Surgery	Biologics	17	0.522 [0.456; 0.588]*	0.394 [0.000; 0.493]	4.43, $p = 0.040$
	Hospitalization	24	0.530 [0.448; 0.613]*	0.375 [0.00; 0.449]	7.63, $p = 0.006$
	Immunomodulators	14	0.609 [0.527; 0.692]*	0.369 [0.318; 0.419]	23.78, $p < 0.001$
	Steroids	16	0.571 [0.488; 0.655]*	0.370 [0.318; 0.422]	16.12, $p < 0.001$
	Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
Steroids	Biologics	11	0.556 [0.472; 0.640]*	0.407 [0.342; 0.473]	7.53, $p = 0.006$
	Immunomodulators	11	0.614 [0.543; 0.685]*	0.388 [0.323; 0.452]	21.61, $p < 0.001$
	Therapy modification	2	0.766 [0.585; 0.948]*	0.430 [0.380; 0.480]	12.27, $p < 0.001$
Therapy modification	Biomarkers	1	0.176 [0.072; 0.281]*	0.451 [0.395; 0.507]	20.55, $p < 0.001$

Frequency of composite outcomes according to the presence of pairs of predefined variables reported in the study.
*Statistically significant from the mean of the subgroup that does not include both variables, $p < 0.05$.

Table 2. Subgroup analysis.

First variable	Second variable	Third variable	Number of outcomes	Frequency of composite outcome in subgroup		Significance
				All variables reported	At least one not reported	
Clinical evaluation	Endoscopic assessment	Biomarkers	1	0.408 [0.294; 0.523]	0.446 [0.389; 0.502]	0.33, $p = 0.570$
	Events	Hospitalization	1	0.174 [0.094; 0.255]*	0.451 [0.395; 0.507]	30.80, $p < 0.001$
		Surgery	1	0.174 [0.094; 0.255]*	0.451 [0.395; 0.507]	30.80, $p < 0.001$
	Hospitalization	Immunomodulators	3	0.734 [0.000; 1.000]*	0.423 [0.373; 0.474]	13.73, $p < 0.001$
		Steroids	5	0.543 [0.000; 1.000]	0.432 [0.381; 0.483]	0.89, $p = 0.350$
		Surgery	4	0.634 [0.000; 1.000]*	0.426 [0.375; 0.477]	4.11, $p = 0.040$
	Surgery	Immunomodulators	3	0.734 [0.000; 1.000]*	0.423 [0.373; 0.474]	13.73, $p < 0.001$
		Steroids	4	0.634 [0.000; 1.000]*	0.426 [0.375; 0.477]	4.11, $p = 0.040$
	Steroids	Immunomodulators	3	0.734 [0.000; 1.000]*	0.423 [0.373; 0.474]	13.73, $p < 0.001$
	Therapy modification	Biomarkers	1	0.176 [0.072; 0.281]*	0.451 [0.395; 0.507]	20.55, $p < 0.001$
Events	Biologics	Immunomodulators	4	0.703 [0.651; 0.754]*	0.418 [0.359; 0.476]	51.54, $p < 0.001$
		Hospitalization	5	0.700 [0.659; 0.741]*	0.410 [0.351; 0.469]	62.65, $p < 0.001$
	Hospitalization	Immunomodulators	4	0.703 [0.651; 0.754]*	0.418 [0.359; 0.476]	51.54, $p < 0.001$
		Steroids	6	0.639 [0.544; 0.734]*	0.413 [0.353; 0.472]	15.55, $p < 0.001$
	Surgery	Biologics	6	0.675 [0.611; 0.738]*	0.406 [0.344; 0.468]	35.43, $p < 0.001$
		Hospitalization	9	0.519 [0.398; 0.640]	0.425 [0.362; 0.488]	1.82, $p = 0.180$
		Immunomodulators	4	0.703 [0.651; 0.754]*	0.418 [0.359; 0.476]	51.54, $p < 0.001$
		Steroids	6	0.639 [0.544; 0.734]*	0.413 [0.353; 0.472]	15.55, $p < 0.001$
	Steroids	Biologics	5	0.700 [0.659; 0.741]*	0.410 [0.351; 0.469]	62.65, $p < 0.001$
		Immunomodulators	4	0.703 [0.651; 0.754]*	0.418 [0.359; 0.476]	51.54, $p < 0.001$
Hospitalization	Steroids	Biologics	8	0.661 [0.594; 0.728]*	0.394 [0.332; 0.456]	32.90, $p < 0.001$
		Immunomodulators	8	0.711 [0.644; 0.778]*	0.384 [0.336; 0.431]	61.33, $p < 0.001$
Surgery	Biologics	Immunomodulators	10	0.559 [0.458; 0.660]*	0.411 [0.349; 0.474]	5.91, $p = 0.020$
		Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
	Hospitalization	Biologics	10	0.621 [0.535; 0.706]*	0.389 [0.000; 0.471]	14.59, $p < 0.001$
		Immunomodulators	8	0.711 [0.644; 0.778]*	0.384 [0.336; 0.431]	61.33, $p < 0.001$
		Steroids	13	0.628 [0.545; 0.710]*	0.364 [0.315; 0.414]	28.87, $p < 0.001$
		Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
	Immunomodulators	Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
		Steroids	10	0.636 [0.551; 0.721]*	0.388 [0.340; 0.437]	24.77, $p < 0.001$
	Steroids	Biologics	11	0.572 [0.490; 0.655]*	0.402 [0.337; 0.466]	10.25 $p = 0.001$
		Therapy modification	1	0.667 [0.566; 0.767]*	0.440 [0.383; 0.496]	14.81, $p < 0.001$
Steroids	Biologics	Immunomodulators	8	0.585 [0.494; 0.676]*	0.414 [0.352; 0.475]	9.35, $p = 0.002$

Frequency of composite outcomes according to the presence of triads of predefines variables reported in the study.
*Statistically significant from the mean of the subgroup that does not include all three variables, $p < 0.05$.

be particularly sensitive to the inclusion of certain variables and, therefore, the exclusion of these variables from certain observational studies suggest that the real frequency of these outcomes might have been underestimated and (2) some variables whose inclusion does not have an effect by itself may still increase the frequency when reported in combination. This was the case, for instance, for the variables 'Biologics' and 'Biomarkers'. Those studies included in this analysis that included both variables reported significantly higher frequencies of composite outcomes compared with those that did not include this combination, but in those that included either one or the other, this effect was not observed.

To the best of our knowledge, this report represents the first systematic review and meta-analysis of outcomes in observational studies of CD. The data presented here complement previous reports that also focused on CD outcomes but were restricted to randomized clinical trials (RCTs),[58,59] which were specifically excluded from our analysis. In fact, those reports explicitly state the need to consider nonrandomized controlled trials and observational studies in addition to RCTs to get a fuller picture of outcome reporting in CD.[59] A comparable heterogeneity in the reporting of outcomes as reported for studies identified in RCT analysis is also evident in observational studies of CD. This highlights the lack of consensus on the clinical outcomes normally reported in studies involving CD patients, and the current need for the development of a core outcome set (COS) to bring the necessary standardization in the reporting of results.[60] In addition, our study identifies certain key variables, and combinations of variables, that appear to have the most pronounced effect for increasing the frequency of composite outcomes, an important piece of information in the development of a COS. A particular advantage of taking observational studies into consideration is that this results in the inclusion of long-term studies (as long as 16 years of continuous monitoring in this report), which is not realistic for RCTs. This may provide additional insight on how specific outcome variables behave in the long term, something particularly relevant given the chronic characteristic of CD.

A limitation of this study is that it did not include patient-reported outcome measures (PROMs) among the outcome variables considered. Although the use of PROMs has become more widespread recently, their validation against more traditional endpoints used in CD is still pending.[61] The reliability of the outcome variables considered in the analysis has also not been assessed here. In addition, the results present high heterogeneity ($I^2 > 50\%$), even after performing subgroup analysis. However, we presume it is mainly due to the baseline characteristics of the included studies, which are highly different from each other.

In summary, the present meta-analysis illustrates the importance of considering the number of variables to get an accurate estimate of the frequency of composite outcomes. Moreover, it identifies a group of variables that appear to be particularly important for the determination of composite indices for CD, and whose absence from the study report may lead to underestimation of such outcomes. This effect is not limited to individual variables but also applies to certain combinations of variables that appear to be linked to each other. We hope this will become a valuable resource in the development of tools for the standardization of outcome reporting, a yet unmet need in the field, and in the design of future cohort studies in CD.

Author contribution(s)

Fernando Magro: Conceptualization; Methodology; Supervision; Validation; Writing – original draft.

Catarina Sottomayor: Data curation; Formal analysis; Investigation; Writing – original draft.

Catarina Alves: Data curation; Formal analysis; Investigation; Writing – original draft.

Mafalda Santiago: Data curation; Formal analysis; Methodology; Writing – original draft.

Paula Ministro: Data curation; Writing – review & editing.

Paula Lago: Data curation; Writing – review & editing.

Luís Correia: Data curation; Writing – review & editing.

Raquel Gonçalves: Data curation; Writing – review & editing.

Diana Carvalho: Data curation; Writing – review & editing.

Francisco Portela: Data curation; Writing – review & editing.

Cláudia Camila Dias: Data curation; Writing – review & editing.

Axel Dignass: Data curation; Writing – review & editing.

Silvio Danese: Data curation; Writing – review & editing.

Laurent Peyrin-Biroulet: Data curation; Writing – review & editing.

Maria Manuela Estevinho: Data curation; Formal analysis; Investigation; Writing – original draft.

Paula Leão Moreira: Data curation; Writing – review & editing.

ORCID iDs
Fernando Magro https://orcid.org/0000-0003-2634-9668

Cláudia Camila Dias https://orcid.org/0000-0001-9356-3272

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Portuguese Group of Studies in Inflammatory Bowel Disease (GEDII).

Conflict of interest statement
The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: FM received a fee for presenting from AbbVie, Ferring, Falk, Hospira, PharmaKern, MSD, Schering, Laboratórios Vitoria, Vifor Pharma and OM Pharma. AD reports fees for participation in clinical trials, review activities, such as data monitoring boards, statistical analysis, end point committees from Falk, Abbvie, Janssen, Gilead and Pfizer; consultancy fees from Abbvie, MSD, Ferring, Roche/Genentech, Takeda, Vifor, Pharmacosmos, Boehringer-Ingelheim, Falk, Janssen, Pfizer, Sandoz/Hexal, BMS/ Celgene, Tillotts, Galapagos, Amgen and Fresenius Kabi; payment from lectures including service on speakers bureaus from Falk Foundation, Ferring, MSD, Abbvie, Vifor, Janssen, Pfizer, Tillotts, Takeda, Gilead/Galapagos; payment for article preparation from Falk Foundation, Thieme, Takeda and UniMed Verlag. SD has served as a speaker, consultant and advisory board member for Schering-Plough, AbbVie, MSD, UCB Pharma, Ferring, Cellerix, Millenium Takeda, Nycomed, Pharmacosmos, Actelion, Alphawasserman, Genentech, Grunenthal, Pfizer, Astra Zeneca, Novo Nordisk, Cosmo Pharmaceuticals, Vifor and Johnson & Johnson, Nikkiso Europe GMBH and Theravance. LP-B has served as a speaker, consultant and advisory board member for Merck, Abbvie, Janssen, Genentech, Mitsubishi, Ferring, Norgine, Tillots, Vifor, Hospira/Pfizer, Celltrion, Takeda, Biogaran, Boerhinger-Ingelheim, Lilly, HAC Pharma, Index Pharmaceuticals, Amgen, Sandoz, Forward Pharma GmbH, Celgene, Biogen, Lycera, Samsung Bioepis and Theravance.

Writing assistance
Writing assistance was provided by Ana Rolo and funded by the Portuguese Group of Studies in Inflammatory Bowel Disease (GEDII).

Data availability statement
The data that support the findings of this study are available from the corresponding author on reasonable request.

Supplemental material
Supplemental material for this article is available online.

References
1. Goldberg R, Gore JM, Barton B, *et al*. Individual and composite study endpoints: separating the wheat from the chaff. *Am J Med* 2014; 127: 379–384.
2. Ferreira-González I, Permanyer-Miralda G, Busse JW, *et al*. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol* 2007; 60: 651–657.
3. Kip KE, Hollabaugh K, Marroquin OC, *et al*. The problem with composite end points in cardiovascular studies. The story of major adverse cardiac events and percutaneous coronary intervention. *J Am Coll Cardiol* 2008; 51: 701–707.
4. Baumgart DC. *Crohn's disease and ulcerative colitis: from epidemiology and immunobiology to a rational diagnostic and therapeutic approach*. 2nd ed. Berlin: Springer, 2017, pp. 1–695.

5. Higgins G. *Cochrane handbook for systematic reviews of interventions*. London: Wiley 2011, pp. 649.
6. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*; 2009; 6: e1000097.
7. Cordoba G, Schwartz L, Woloshin S, *et al*. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review *BMJ* 341: c3920.
8. Oxford Centre for Triple Value Healthcare. CASP Checklists, https://casp-uk.net/casp-tools-checklists/
9. Balduzzi S, Rücker G and Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* 2019; 22: 153–160.
10. Higgins JPT, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560.
11. Egger M, Smith GD, Schneider M, *et al*. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997; 315: 629–634.
12. Yokoyama K, Yamazaki K, Katafuchi M, *et al*. A retrospective claims database study on drug utilization in Japanese patients with Crohn's disease treated with adalimumab or infliximab. *Adv Ther* 2016; 33: 1947–1963.
13. Yang CH, Ding J, Gao Y, *et al*. Risk factors that predict the requirement of aggressive therapy among Chinese patients with Crohn's disease. *J Dig Dis* 2011; 12: 99–104.
14. Wenger S, Nikolaus S, Howaldt S, *et al*. Predictors for subsequent need for immunosuppressive therapy in early Crohn's disease. *J Crohn's Colitis* 2012; 6: 21–28.
15. Ungaro RC, Limketkai BN, Jensen CB, *et al*. Stopping Mesalamine therapy in patients with Crohn's disease starting biologic therapy does not increase risk of adverse outcomes. *Clin Gastroenterol Hepatol* 2020; 18: 1152.e1–1160.e1.
16. Ungaro RC, Yzet C, Bossuyt P, *et al*. Deep remission at 1 year prevents progression of early Crohn's disease. *Gastroenterology* 2020; 159: 139–147.
17. Thomsen SB, Allin KH, Burisch J, *et al*. Outcome of concomitant treatment with thiopurines and allopurinol in patients with inflammatory bowel disease: a nationwide Danish cohort study. *United Eur Gastroenterol J* 2020; 8: 68–76.
18. Thiberge C, Charpentier C, Gillibert A, *et al*. Lower subcutaneous or visceral adiposity assessed by abdominal computed tomography could predict adverse outcome in patients with Crohn's disease. *J Crohn's Colitis* 2018; 12: 1429–1437.
19. Schnitzler F, Friedrich M, Wolf C, *et al*. The NOD2 p.Leu1007fsX1008 mutation (rs2066847) is a stronger predictor of the clinical course of Crohn's disease than the FOXO3A Intron variant rs12212067. *PLoS ONE* 2014; 9: e108503.
20. Savoye G, Salleron J, Gower-Rousseau C, *et al*. Clinical predictors at diagnosis of disabling pediatric Crohn's disease. *Inflamm Bowel Dis* 2012; 18: 2072–2078.
21. Samimi R, Flasar MH, Kavic S, *et al*. Outcome of medical treatment of stricturing and penetrating Crohn's disease: a retrospective study. *Inflamm Bowel Dis* 2010; 16: 1187–1194.
22. Sokol H, Polin V, Lavergne-Slove A, *et al*. Plexitis as a predictive factor of early postoperative clinical recurrence in Crohn's disease. *Gut* 2009; 58: 1218–1225.
23. Ripollés T, Paredes JM, Martínez-Pérez MJ, *et al*. Ultrasonographic changes at 12 weeks of anti-TNF drugs predict 1-year sonographic response and clinical outcome in Crohn's disease: a multicenter study. *Inflamm Bowel Dis* 2016; 22: 2465–2473.
24. Pouillon L, Remen T, Amicone C, *et al*. Risk of late postoperative recurrence of Crohn's disease in patients in endoscopic remission after ileocecal resection, over 10 years at multiple centers. *Clin Gastroenterol Hepatol* 2021; 19: 1218.e4–1225.e4.
25. O'Donnell S, Borowski K, Espin-Garcia O, *et al*. The unsolved link of genetic markers and Crohn's disease progression: a North American cohort experience. *Inflamm Bowel Dis* 2019; 25: 1541–1549.
26. Kennedy NA, Jones GR, Plevris N, *et al*. Association between level of fecal calprotectin and progression of Crohn's disease. *Clin Gastroenterol Hepatol* 2019; 17: 2269–2276.
27. Niewiadomski O, Studd C, Hair C, *et al*. Prospective population-based cohort of inflammatory bowel disease in the biologics era : disease course and predictors of severity. *J Gastroenterol Hepatol* 2015; 30: 1346–1353.
28. Ng SC, Lied GA, Kamm MA, *et al*. Predictive value and clinical significance of myenteric plexitis in Crohn's disease. *Inflamm Bowel Dis* 2009; 15: 1499–1507.
29. Meyer A, Rudant J, Drouin J, *et al*. Effectiveness and safety of reference infliximab and biosimilar

in Crohn disease: a French equivalence study. *Ann Intern Med* 2019; 170: 99–107.

30. Klaassen MAY, Imhann F, Collij V, *et al*. Anti-inflammatory gut microbial pathways are decreased during Crohn's disease exacerbations. *J Crohn's Colitis* 2019; 13: 1439–1449.
31. Magro F, Dias CC, Coelho R, *et al*. Impact of early surgery and immunosuppression on Crohn's disease disabling outcomes. *Inflamm Bowel Dis* 2017; 23: 289–297.
32. Lian L, Stocchi L, Shen B, *et al*. Prediction of need for surgery after endoscopic balloon dilation of ileocolic anastomotic stricture in patients with Crohn's disease. *Dis Colon Rectum* 2015; 58: 423–430.
33. Lee JC, Lyons PA, McKinney EF, *et al*. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* 2011; 121: 4170–4179.
34. Loly C, Belaiche J and Louis E. Predictors of severe Crohn's disease. *Scand J Gastroenterol* 2008; 43: 948–954.
35. Kochar B, Barnes EL, Long MD, *et al*. Depression is associated with more aggressive inflammatory bowel disease. *Am J Gastroenterol* 2018; 113: 80–85.
36. Koutroubakis IE, Ramos-Rivers C, Regueiro M, *et al*. Persistent or recurrent anemia is associated with severe and disabling inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2015; 13: 1760–1766.
37. Jansen I, Prager M, Valentini L, *et al*. Inflammation-driven malnutrition: a new screening tool predicts outcome in Crohn's disease. *Br J Nutr* 2016; 116: 1061–1067.
38. Huang S, Li L, Ben-Horin S, *et al*. Mucosal healing is associated with the reduced disabling disease in Crohn's disease. *Clin Transl Gastroenterol* 2019; 10: e00015.
39. Hradsky O, Copova I, Zarubova K, *et al*. Time to relapse in children with Crohn's disease treated with azathioprine and nutritional therapy or corticosteroids. *Dig Dis Sci* 2016; 61: 2041–2050.
40. Heresbach D, Alexandre JL, Branger B, *et al*. Frequency and significance of granulomas in a cohort of incident cases of Crohn's disease. *Gut* 2005; 54: 215–222.
41. Galazzo G, Tedjo DI, Wintjens DSJ, *et al*. Faecal microbiota dynamics and their relation to disease course in Crohn's disease. *J Crohn's Colitis* 2019; 13: 1273–1282.
42. Gibson DJ, Murphy DJ, Smyth AE, *et al*. Magnetic resonance enterography findings as predictors of clinical outcome following antitumor necrosis factor treatment in small bowel Crohn's disease. *Eur J Gastroenterol Hepatol* 2015; 27: 956–962.
43. Gonsky R, Deem RL, Landers CJ, *et al*. IFNG rs1861494 polymorphism is associated with IBD disease severity and functional changes in both IFNG methylation and protein secretion. *Inflamm Bowel Dis* 2014; 20: 1794–1801.
44. Dias CC, Portela F, Fernandes S, *et al*. The timing of early therapeutic strategies has a significant impact on Crohn's disease prognosis. *Dig Liver Dis* 2018; 50: 462–468.
45. Ding NS, Malietzis G, Lung PFC, *et al*. The body composition profile is associated with response to anti-TNF therapy in Crohn's disease and may offer an alternative dosing paradigm. *Aliment Pharmacol Ther* 2017; 46: 883–891.
46. Dias CC, Rodrigues PP, Coelho R, *et al*. Development and validation of risk matrices for Crohn's disease outcomes in patients who underwent early therapeutic interventions. *J Crohn's Colitis* 2017; 11: 445–453.
47. Dias CC, Rodrigues PP, Fernandes S, *et al*. The risk of disabling, surgery and reoperation in Crohn's disease – a decision tree-based approach to prognosis. *PLoS ONE* 2017; 12: e0172165.
48. Campos C, Perrey A, Lambert C, *et al*. Medical therapies for stricturing Crohn's disease: efficacy and cross-sectional imaging predictors of therapeutic failure. *Dig Dis Sci* 2017; 62: 1628–1636.
49. Braun T, Di Segni A, Benshoshan M, *et al*. Individualized dynamics in the gut microbiota precede Crohn's disease flares. *Am J Gastroenterol* 2019; 114: 1142–1151.
50. Biasci D, Lee JC, Noor NM, *et al*. A blood-based prognostic biomarker in IBD. *Gut* 2019; 68: 1386–1395.
51. Brown P, Clark T, Dowson G, *et al*. Relationship of body mass index to clinical outcomes after infliximab therapy in patients with Crohn's disease. *J Crohn's Colitis* 2016; 10: 1144–1150.
52. Beaugerie L, Seksik P, Nion-Larmurier I, *et al*. Predictors of Crohn's disease. *Gastroenterology* 2006; 130: 650–656.
53. Baldassano RN, Han PD, Jeshion WC, *et al*. Pediatric Crohn's disease: risk factors for postoperative recurrence. *Am J Gastroenterol* 2001; 96: 2169–2176.

54. Ahmad W, Nguyen NH, Boland BS, *et al*. Comparison of multiplex gastrointestinal pathogen panel and conventional stool testing for evaluation of diarrhea in patients with inflammatory bowel diseases. *Dig Dis Sci* 2019; 64: 382–390.

55. Af Björkesten CG, Jussila A, Kemppainen H, *et al*. Relationship of faecal calprotectin and long-term outcomes in Finnish patients with Crohn's disease: retrospective multi-centre chart review study. *Scand J Gastroenterol* 2019; 54: 1226–1232.

56. Arieira C, Cúrdia Gonçaves T, Dias de Castro F, *et al*. Clinical course in Crohn's disease: factors associated with behaviour change and surgery. *Scand J Gastroenterol* 2018; 53: 1222–1227.

57. Ananthakrishnan AN, Sakuraba A, Barnes EL, *et al*. The benefit of combination therapy depends on disease phenotype and duration in Crohn's disease. *Aliment Pharmacol Ther* 2017; 46: 162–168.

58. Ma C, Hussein IM, Al-Abbar YJ, *et al*. Heterogeneity in definitions of efficacy and safety endpoints for clinical trials of Crohn's disease: a systematic review. *Clin Gastroenterol Hepatol* 2018; 16: 1407.e22–1419.e22.

59. Catt H, Hughes D, Kirkham JJ, *et al*. Systematic review: outcomes and adverse events from randomised trials in Crohn's disease. *Aliment Pharmacol Ther* 2019; 49: 978–996.

60. Ma C, Panaccione R, Fedorak RN, *et al*. Development of a core outcome set for clinical trials in inflammatory bowel disease: study protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. *BMJ Open* 2017; 7: e016146.

61. de Jong MJ, Huibregtse R, Masclee AAM, *et al*. Patient-reported outcome measures for use in clinical trials and clinical practice in inflammatory bowel diseases: a systematic review. *Clin Gastroenterol Hepatol* 2018; 16: 648–663.