

4-27-2020

The Inflammatory Bowel Disease Questionnaire in Randomized Controlled Trials of Treatment for Ulcerative Colitis: Systematic Review and Meta-Analysis

Aaron Yaras
Stephen Maher
Martha Bayliss
Andrew Lovley
Joseph C. Cappelleri
Andrew G. Bushmakin
Marco D. DiBonaventura

Follow this and additional works at: <https://aurora.org/jpcrr>



Part of the [Digestive System Diseases Commons](#), [Gastroenterology Commons](#), and the [Health Services Research Commons](#)

Recommended Citation

Yaras A, Maher S, Bayliss M, Lovley A, Cappelleri JC, Bushmakin AG, DiBonaventura MD. The Inflammatory Bowel Disease Questionnaire in randomized controlled trials of treatment for ulcerative colitis: systematic review and meta-analysis. *J Patient Cent Res Rev.* 2020;7:189-205. doi: [10.17294/2330-0698.1722](https://doi.org/10.17294/2330-0698.1722)

Journal of Patient-Centered Research and Reviews (JPCRR) is a peer-reviewed scientific journal whose mission is to communicate clinical and bench research findings, with the goal of improving the quality of human health, the care of the individual patient, and the care of populations.



The Inflammatory Bowel Disease Questionnaire in Randomized Controlled Trials of Treatment for Ulcerative Colitis: Systematic Review and Meta-Analysis

Aaron Yaras, PhD,¹ Stephen Maher, PhD,¹ Martha Bayliss, MA,¹ Andrew Lovley, MA,¹ Joseph C. Cappelleri, PhD,² Andrew G. Bushmakina, MS,² Marco D. DiBonaventura, PhD²

¹Optum Patient Insights, Optum, Inc., Johnston, RI; ²Pfizer Inc., New York, NY

Purpose	The 32-item Inflammatory Bowel Disease Questionnaire (IBDQ-32) is the most frequently used instrument to capture disease-specific quality of life in randomized clinical trials for ulcerative colitis. This review and meta-analysis provides the first synthesis of evidence regarding the sensitivity of IBDQ-32 total and domain scores to treatment efficacy.
Methods	A systematic literature search and risk-of-bias assessment yielded 14 articles that were included in the primary analysis. Treatments were categorized as efficacious if they met the primary efficacy endpoint (which was not the IBDQ-32); otherwise they were categorized as non-efficacious. A continuous measure of treatment efficacy was calculated for each primary efficacy endpoint. Meta-analysis using random-effects models compared standardized mean differences in IBDQ-32 total and domain change scores between target dose and control arms. Meta-regression compared the association between treatment efficacy and these outcomes.
Results	Studies with efficacious treatments showed larger mean improvements relative to controls in IBDQ-32 total scores and all 4 domains (Hedges' <i>g</i> range: 0.49 to 0.67; <i>P</i> <0.001 for all). At the same time, patients in studies with non-efficacious treatments showed small and nonsignificant improvements in these outcomes relative to controls (Hedges' <i>g</i> range: 0.05 to 0.23; <i>P</i> >0.09 for all). Meta-regression models showed that the magnitude of treatment efficacy was a positive predictor of these same IBDQ-32 outcomes.
Conclusions	These analyses found that IBDQ-32 scores are sensitive to treatment. The results provided here support the use of the IBDQ-32 to capture treatment benefits on quality of life for patients with ulcerative colitis. (<i>J Patient Cent Res Rev.</i> 2020;7:189-205.)
Keywords	ulcerative colitis; quality of life; meta-analysis; clinical trials; patient questionnaire

Ulcerative colitis (UC) — one of the two major subtypes of inflammatory bowel disease (IBD), the other being Crohn's disease — is characterized by chronic inflammation and ulceration of tissue within the colon. UC is a relapse-remittent disease, such that patients with UC experience intermittent episodes (flares) that are accompanied

by clinical symptoms, including abdominal pain or cramping, fatigue, diarrhea, rectal bleeding, and frequent and unpredictable urges to defecate. In the United States, the prevalence of UC has been estimated at 28.63 with an annual incidence of 1.22 (both per 10,000).¹ The presence and severity of active UC is associated with impaired health-related quality of life (HRQoL).²⁻⁵

In clinical trials of treatments for UC, primary endpoints are typically disease activity indices, such as the Mayo score,⁶ which are based on ratings of frequency and severity of clinical symptoms and endoscopic activity.

Correspondence: Aaron Yaras, PhD,
Optum, 1301 Atwood Ave., Suite 311N, Johnston, RI 02919
(ayaras@qualitymetric.com)

However, disease activity indices fail to capture the broader humanistic impact of UC on patients' physical, emotional, and social functioning or the humanistic benefits of treatment. To complement the disease activity indices, UC trials often include endpoints capturing change in patients' HRQoL, with the most frequently used instrument being the 32-item Inflammatory Bowel Disease Questionnaire (IBDQ-32).⁷⁻⁹

The IBDQ-32 captures the patient's experience of IBD on 4 domains of functioning and well-being: bowel and systemic symptoms; and emotional and social function.¹⁰ Reviews of the measurement properties of the IBDQ-32 have found evidence supporting its reliability, content validity, construct validity, and responsiveness.^{7,8,11,12} Further, reviews have concluded that the IBDQ-32 has the strongest measurement profile among instruments used to assess IBD-specific HRQoL.^{7,8,11} Other reviewers have recommended that the IBDQ-32 be included as an endpoint in all UC clinical trials in which HRQoL of patients is a relevant outcome.^{9,13}

Despite the evidence supporting the reliability, validity, and responsiveness of the IBDQ-32 when used in observational or noncomparative treatment studies of patients with IBD, to our knowledge there are no comprehensive reviews of the IBDQ-32 when used in randomized controlled trials of patients with UC. Thus, there is a lack of evidence speaking to the degree to which the IBDQ-32 demonstrates sensitivity to treatment.

The objective of this systematic literature review and meta-analysis was to address this evidence gap by examining the magnitude of change in mean IBDQ-32 scores as a function of treatment efficacy. For this purpose, treatment efficacy was defined in two ways: 1) dichotomously (efficacious or non-efficacious), based on whether or not the study's prespecified primary efficacy endpoint (based on a clinical measure of disease activity, not the IBDQ-32) was met; and 2) continuously, based on the effect size (ES) computed for the difference in change on the study's primary efficacy endpoint between target treatment and control arms.

Results from these analyses will add to the evidence of the IBDQ-32's sensitivity to treatment and its utility as an endpoint for assessing the impact of

UC treatments on functioning and well-being in randomized controlled trials. More generally, these analyses address the important question of how well patient-centered outcomes of HRQoL used in clinical research and practice correspond to changes in clinical health as a function of treatment interventions.

METHODS

Inflammatory Bowel Disease Questionnaire (IBDQ-32)

The IBDQ-32 was developed in the late 1980s at McMaster University (Hamilton, Canada).^{10,14} Item selection was based on concept elicitation interviews with patients with UC or Crohn's disease as well as clinical experts, followed by cognitive debriefing among IBD health professionals and patients, resulting in 32 items.¹⁵ All items use 7-point Likert-type scales for capturing symptom-related experiences over the previous 2 weeks, with 1 indicating the highest symptom frequency/severity and 7 indicating the lowest symptom frequency/severity.

Content analysis led to formation of 4 domains: 1) bowel symptoms, 2) systemic symptoms, 3) emotional function, and 4) social function (domain scoring and characteristics are presented in online-only Supplemental Table S1). A total score can also be calculated as the sum of all 32 items (score range: 0–224). Higher domain and total scores indicate better HRQoL.

Literature Search

A systematic search of the published literature, which followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines,¹⁶ identified articles reporting data from randomized controlled trials in which the IBDQ-32 was administered to adults with active UC. (The search protocol will be made available upon request to the authors.) Searches of PubMed, Embase (OvidSP), and the Cochrane Register of Controlled Trials databases were first conducted on September 8, 2017. For the purpose of updating the search results, the search was repeated on September 25, 2019, with the "date of publication" terms updated to restrict the search to newer records that had been published after the first search had been conducted. Search terms included "inflammatory bowel disease questionnaire," "IBDQ," "ulcerative colitis," and "inflammatory bowel disease."

Where appropriate, MeSH terms were used for disease terms (specific terms only, not “exploded”). Articles were filtered for English language only.

In the original search (2017), the publication date of articles was restricted from 2003 to present, since the IBDQ-32 development paper was published in 2003. In the updated search (2019), the publication date of articles was restricted from 2017 to present to capture articles published after the original search was conducted. Articles retrieved from the updated search that were published in 2017 were manually checked to remove duplicates of articles that had been retrieved in the original search. Specific search terms and strings used within each database are provided in online-only Supplemental Figure S1. During full-text review, potentially relevant articles cited by papers were also identified for abstract screening.

Inclusion/Exclusion Criteria: To be included in the review, identified papers needed to describe a double-blinded randomized controlled trial in which the IBDQ-32 was administered to adult patients with active UC before and after the patients had received either an active treatment or a true control (eg, placebo for drug studies, conventional care for psychotherapeutic studies). In addition, articles needed to report IBDQ-32 data that afforded calculation of ES estimates for mean differences in total and/or domain scores. Finally, the success or failure of treatment to meet the study’s primary (clinical) efficacy endpoint needed to be reported in the article or reported for that trial in a different article for which there was a citation.

Article Screening: Screenings of abstracts and full-text articles were performed by at least 2 independent reviewers (from among authors A.Y., S.M., and A.L.) for each article. Any discrepancy among reviewers in a selection decision was resolved by discussion among all 3 reviewers until a consensus decision was reached.

Extraction of Data: Relevant data were extracted from each selected article by 1 researcher (from among authors A.Y., S.M., and A.L.) and independently reviewed for accuracy by at least 1 other researcher. Any discrepancy in extraction was resolved by discussion among at least 2 researchers until a consensus decision was reached.

Values for IBDQ-32 scores reported numerically in an article were extracted directly and added to a database. If the IBDQ-32 scores were reported only graphically in an article (ie, displayed in a figure but not as numeric values), values were estimated using the software WebPlotDigitizer-Desktop, Version 2.8 (<https://automeris.io/WebPlotDigitizer>, Ankit Rohatgi, San Francisco, CA), which converts the spatial distance of points on a graph into numeric values. WebPlotDigitizer has been demonstrated to have high levels of intercoder reliability and validity when used for this purpose.¹⁷ For all studies, IBDQ-32 scores were extracted only for the assessment visits at which the primary efficacy endpoint was evaluated.

Derivation of ES estimates were calculated using Comprehensive Meta-Analysis (CMA) software, Version 3.3 (Biostat, Inc., Englewood, NJ).

Risk-of-Bias Assessment

Risk of bias was assessed for each included study to identify threats to internal validity due to systematic errors in the design, procedures, or reporting of the study. The risk of bias tool provided by the Cochrane Collaboration¹⁸ was used to evaluate the severity of risk (low, high, or unknown) for each study across 6 domains: sequence generation, allocation concealment, blinding, incomplete outcome data, selective reporting, and other sources of bias. For studies that did not report information relevant for assessing risk of bias, we conducted searches of selected online clinical trial registries (eg, clinicaltrials.gov, chictr.org.cn) and performed internet searches for study protocols or other available documentation.

Any study for which a high risk was not identified for any of the 6 domains was considered a low-risk-of-bias study; a study with 1 or 2 high-risk domains was considered a moderate-risk-of-bias study; and a study with 3 or more high-risk domains was considered to be a high-risk-of-bias study. To enhance the reliability of our findings, it was determined that any studies rated as having high risk of bias would be excluded from the primary analysis. If the assessment identified 1 or more studies as having a high risk of bias, a separate sensitivity analysis, with high-risk-of-bias studies included in meta-analysis models of mean differences in IBDQ-32 total and domain change scores, would be conducted.

Analysis Methods

All models focused on establishing the relationship between treatment arm (active treatment vs control) and changes in IBDQ-32 scores. Changes in IBDQ-32 scores were measured using standardized ES (ie, Hedges' *g* statistic) to allow for comparison across domains and total scores, which had different scaling.

Two moderator variables also were included in the primary analyses to establish whether the relationship between treatment arm and IBDQ-32 score changes may vary systematically. A dichotomous measure of treatment efficacy — coded as “efficacious” if the study’s primary efficacy endpoint was successfully achieved or as “non-efficacious” if the study’s primary efficacy endpoint was not achieved — was included as a categorical moderator in all meta-analysis models. A continuous measure of treatment efficacy, defined as the Hedges' *g* ES for the comparison of change in the primary efficacy endpoint between the treatment and control arms, was included as a continuous moderator (predictor) in all meta-regression models. (Hedges' *g* was calculated directly or estimated using the logit method,¹⁹ as implemented in CMA Version 3.3.²⁰) For studies using multiple treatment arms (eg, treatment at different doses), the primary analysis only calculated treatment efficacy for the treatment arm using the dosage recommended or approved for use in this population or, if this information was not available for a treatment type, the highest dosage administered during the trial (referred to hereafter as the “target dose”).

Additional meta-analyses were conducted to examine effects of other categorical moderating variables on the magnitude of treatment differences in standardized mean change (Hedges' *g*) for IBDQ-32 scores.²⁰ These additional moderators included treatment type, treatment duration, and baseline disease severity.

A sensitivity analysis was conducted in which meta-analysis models were tested across all active treatment arms within each study rather than only the target dose as in the primary analysis. Studies with a single active treatment arm were included as in the primary analysis. However, for studies with multiple active treatment arms, treatment effects for all combined treatment arms were simultaneously compared to the control arm. Treatment effects based on changes in mean scores

were calculated for combined treatment arms based on averaging of means and standard deviations when weighting for sample size. Treatment effects based on proportions were calculated for combined treatment arms by summing the number of subjects who met the primary endpoint criterion (eg, clinical response) and summing the number of total subjects.

For cases in which some treatment arms were relatively efficacious (vs control) while others were non-efficacious within the same study, rules for categorization were used to classify the overall efficacy of treatment as follows. First, if more than half of treatment arms were efficacious or non-efficacious, the category followed the majority of treatment arms. Second, if there were the same number of efficacious and non-efficacious treatment arms, a statistical test similar to those used in the original study (eg, independent-samples *t*-test, *z*-test for difference in proportions) was conducted, and the statistical significance of the results of that test (ie, $P < 0.05$) was used to determine the category of treatment efficacy.

All meta-analyses were conducted using random-effects models to calculate pooled ES estimates within each subgroup of the treatment efficacy categorical moderator (ie, efficacious or non-efficacious treatment) and to compare ES estimates between subgroups.²⁰ Random-effects models were chosen because inclusion of different treatment methods across studies led to the assumption that there could be different “true” ES for each study.²⁰ Individual treatment comparisons were weighted within treatment efficacy subgroups using inverse variance derived from the random-effects models. Meta-analyses across studies were based on Hedges' *g* ES for standardized differences between mean total or domain scores. Interpretation of ES magnitude, both within and across studies, followed Cohen's conventions: 0.20 indicated a small effect, 0.50 indicated a medium effect, and 0.80 indicated a large effect.²¹ Heterogeneity within subgroups of studies (ie, efficacious studies, non-efficacious studies) was assessed using the I^2 statistic, which estimates the percentage of variability due to heterogeneity among studies rather than sampling error. I^2 values of 25%, 50%, and 75% can be interpreted as indicating low, moderate, and high heterogeneity, respectively.²² Given the small numbers of studies being compared across

moderator subgroups, and thus the low statistical power to detect group differences in ES, statistical significance for between-groups heterogeneity (ie, Cochran's Q test) was tested using α of 0.10, as has been recommended elsewhere.²³⁻²⁵

Meta-regression models for IBDQ-32 endpoints were conducted using random-effects models with maximum likelihood estimation for deriving the coefficient, with treatment efficacy as the sole continuous predictor (Hedges' g ES estimates for efficacy of primary endpoint) of the treatment difference for mean IBDQ-32 domain or total scores (also Hedges' g ES estimates).²⁰

Publication bias was examined using a funnel plot and Egger's test (with α of 0.05) for the distribution of Hedges' g for mean differences in IBDQ-32 total scores by the standard error observed within each study reporting this outcome.²⁰

All meta-analyses, meta-regressions, and publication bias analyses reported here were conducted using CMA Version 3.3,²⁰ which also was used to generate the corresponding forest plots and scatterplots.

RESULTS

Literature Search

The PRISMA diagram, which combines results from the searches conducted in September 2017 and September 2019, outlines the sources included in the search, the number of articles retrieved from each source, and the number of articles excluded at each stage of the screening process (Figure 1). IBDQ-32 data were extracted from 15 articles that met all criteria for inclusion in this review.²⁶⁻⁴⁰ Reasons for exclusion of articles during abstract screening, full-text screening, and data extraction are presented in online-only Supplemental Tables S2–S4, respectively.

Study Characteristics

Sample and design characteristics of studies in the 15 selected articles are presented in Table 1. Several of these articles included multiple comparisons of an endpoint due to either reporting of findings from multiple independent studies, or because arms for more than one treatment dose were compared to the control arm. In total, there were 33 treatment arm comparisons across endpoints, of which 18 included the target dose. Seven articles^{30-33,37,38,40} reported findings from biologic

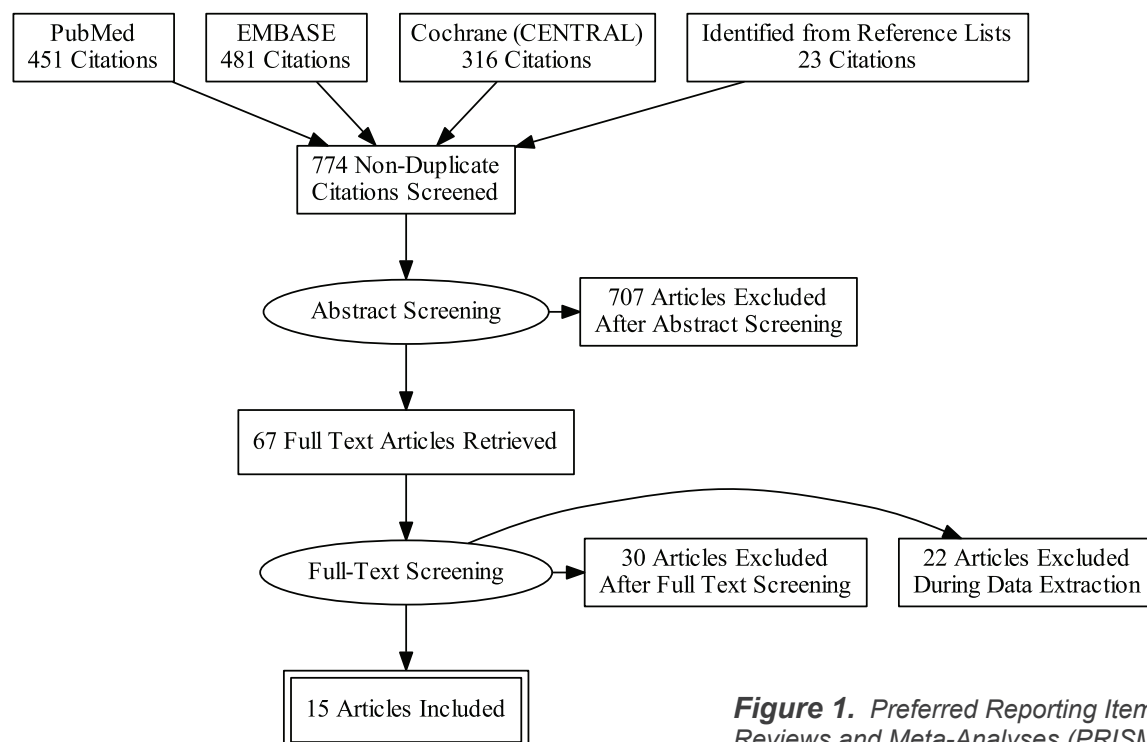


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

Table 1. Sample and Study Characteristics of Reviewed Articles

Author	Risk of Bias ^a	Sample	Target Treatment	Control	Treatment Duration	Primary Efficacy Endpoint	Efficacy on Primary Endpoint? (Hedges' g)	IBDQ Outcomes
Bloom (2004) ²⁶	Low	106 pts with mild-to-moderate active UC at hospital outpatient clinics in Europe & Canada	Tinzaparin 175 anti-Xa IU/kg/day for 14 days followed by tinzaparin 4500 anti-Xa IU/day for 28 days (n=48)	Placebo (n=52)	6 weeks	Ad hoc CAI (bowel frequency, rectal bleeding, histology, sigmoidoscopy)	No (-0.04)	• Domains
Boye (2011) ²⁷	Moderate	58 pts with mostly mild-to-moderate active UC at university clinics in Norway & Germany	Stress management EW (n=35)	TAU (n=23)	18 months	RCAI (modified)	Yes (0.59)	• Total
Cross (2012) ²⁸	Low	47 pts with mild-to-moderate UC at a university hospital or veterans clinic in the U.S.	Home telemanagement (n=25)	BAC (n=22)	12 months	Seo index	Yes (1.69)	• Total
Eisenbruch (2005) ²⁹	High	30 pts with mild active UC recruited via public advertisements in Germany	Mind-body therapy EW (n=15)	Wait-list control (n=15)	10 weeks	RCAI	No (-0.12)	• Total • Domains
Feagan (2005) ³⁰	Low	181 pts with mild-to-moderate active UC at university medical centers globally	VDZ 2 mg/kg (n=60)	Placebo (n=63)	6 weeks	TMS	Yes (0.70)	• Total
Feagan (2007) ³¹	Low	ACT 1: 364 pts with moderate-to-severe active UC at medical sites globally	IFX 5 mg/kg (n=121)	Placebo (n=121)	8 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	Yes (0.74)	• Total • Domains
Feagan (2013) ³²	Low	ACT 2: 364 pts with moderate-to-severe active UC at medical sites globally	IFX 5 mg/kg (n=121)	Placebo (n=123)	8 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	Yes (0.81)	• Total • Domains
Leiper (2011) ³³	Low	895 pts with moderate-to-severe active UC at medical centers in 34 countries	VDZ 300 mg (n=225)	Placebo (n=149)	6 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	Yes (0.53)	• Total
Panaccione (2014) ³⁴	Low	24 pts with moderate-to-severe active UC at a university hospital in U.K.	RTX 1 g (n=16)	Placebo (n=8)	12 weeks	Remission: (TMS < 3)	No (0.26)	• Total
	Low	181 pts with moderate-to-severe active UC at medical centers in Belgium	AZA+IFX (n=63)	AZA-only (n=53) + IFX-only (n=65)	16 weeks	Remission: (TMS < 3) & (all MS items < 2) & (No CS use)	Yes (0.43)	• Total

Table continued on next page

Table 1 (continued). Sample and Study Characteristics of Reviewed Articles

Author	Risk of Bias ^a	Sample	Target Treatment	Control	Treatment Duration	Primary Efficacy Endpoint	Efficacy on Primary Endpoint? (Hedges' g)	IBDQ Outcomes
Panes (2015) ³⁵	Low	194 pts with moderate-to-severe active UC at medical centers globally	Tofacitinib 10 mg BID (n=33)	Placebo (n=48)	8 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	No (0.42)	• Total • Domains
Panes (2018) ^{36*}	Low	OCTAVE Induction 1: 614 pts with moderate-to-severe active UC at medical sites globally	Tofacitinib 10 mg BID (n=476)	Placebo (n=122)	8 weeks	Remission: (TMS < 3) & (all MS items < 2) & (RB = 0)	Yes (0.51)	• Total • Domains
Probert (2003) ³⁷	Low	OCTAVE Induction 2: 547 pts with moderate-to-severe active UC at medical sites globally	Tofacitinib 10 mg BID (n=429)	Placebo (n=112)	8 weeks	Remission: (TMS < 3) & (all MS items < 2) & (RB = 0)	Yes (0.93)	• Total • Domains
Rutgeerts (2015) ³⁸	Low	43 pts with moderately active, glucocorticoid-resistant UC at medical centers in U.K. & Germany	IFX 5 mg/kg (n=23)	Placebo (n=20)	6 weeks	Remission: (TMS < 3) or (Baron score = 0)	No (0.22)	• Total
Sandborn (2003) ³⁹	Low	291 pts with moderate-to-severe active UC at medical centers globally	GLM 4 mg/kg (n=77)	Placebo (n=77)	6 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	No (0.27)	• Total
Sandborn (2014) ⁴⁰	Low	88 pts with mild-to-moderate UC at medical centers in U.S.	Repifermin 50 μ g/kg (n=14)	Placebo (n=28)	4 weeks	Remission: (TMS EA = 0) & (RB = 0) & (SF < 2) & (PGA < 2)	No (-0.75)	• Total
	Low	Phase II: 164 pts with moderate-to-severe active UC at medical centers globally	GLM 200/100 mg (n=41)	Placebo (n=40)	6 weeks	TMS	No (0.28)	• Total
	Low	Phase III: 761 pts with moderate-to-severe active UC at medical centers globally	GLM 200/100 mg (n=252)	Placebo (n=252)	6 weeks	Response: (TMS decrease \geq 3 pts/30%) & (RB decrease \geq 1 pt or RB < 2)	Yes (0.48)	• Total

^aBased on number of domains with high risk of bias: 0 = low risk; 1-2 = moderate risk; 3 or more = high risk. High risk of bias studies were excluded from the primary analysis.

*Data in the meta-analysis were based on those reported in a published erratum to this article.⁵³

5-ASA, 5-aminosalicylic acid; AZA, azathioprine; BAC, best available care; BID, twice daily; CAI, Colitis Activity Index; CS, corticosteroids; DAI, disease activity index; EA, endoscopic appearance; EW, every week; GLM, golimumab; IBDQ, Inflammatory Bowel Disease Questionnaire; IFN, interferon- β -1a; IFX, infliximab; IU, international unit; kg, kilogram; MS, Mayo score; PGA, Physician's Global Assessment; pts, patients; RB, rectal bleeding; RCAI: Rachmilewitz clinical activity index; RTX, rituximab; SF, stool frequency; TAU, treatment as usual; TMS, total Mayo score; UC, ulcerative colitis; VDZ, vedolizumab.

treatments (golimumab, infliximab, rituximab, or vedolizumab), 4 articles^{26,35,36,39} reported small-molecule treatments (tinzaparin, tofacitinib, or repifermin), 2 articles^{27,29} reported findings from psychological therapy (stress management or mind-body therapy), 1 article reported treatments of management training,²⁸ and 1 article reported combined small-molecule and biologic treatment (azathioprine and infliximab).³⁴

Treatment duration, defined as treatment onset through assessment of the primary efficacy endpoint (and thus the assessment of IBDQ-32 scores analyzed here), ranged from 4 to 16 weeks for biologic and small-molecule treatments, with 6- and 8-week assessments most commonly used. Treatment duration for studies using psychological therapy varied from 10 weeks to 18 months, while the efficacy of treatment management training was assessed after 12 months.

All studies defined their primary efficacy endpoint as treatment differences in change in a disease activity index based on clinical and/or endoscopic activity. All studies used well-established disease activity indices for use in clinical trials, such as Rachmilewitz's clinical activity index,⁴¹ Mayo score,⁶ Seo index,⁴² and Sutherland's UC disease activity index,⁴³ with the exception of one study²⁶ that used an ad hoc disease activity index capturing clinical, endoscopic, and histologic activity. Six studies used differences in mean change on the disease activity index as their primary endpoint, 6 used differences in proportion of patients showing clinical response, and 6 used differences in proportion of patients achieving clinical and/or endoscopic remission (with the latter two approaches each using the Mayo score to define response or remission).

For 11 articles, ES estimates for treatment differences in mean change scores were calculated directly from reported numeric data. For the remaining 4 articles,^{29,31,32,34} ES estimates were calculated from data extracted from figures using WebPlotDigitizer.

Risk of Bias

The risk-of-bias analysis by individual study (online-only Supplemental Figure S2) and across all studies (online-only Supplemental Figure S3) showed that the risk of bias was low across most risk-of-bias domains

and studies. Only 2 of 15 articles described studies that had 1 or more domains that were judged to have a high risk of bias: Boye et al²⁷ had 1 high-risk-of-bias domain and was considered a moderate-risk-of-bias study; while Elsenbruch et al²⁹ had 3 high-risk-of-bias domains and thus was considered a high-risk-of-bias study. Following the prespecified criteria, this latter study was excluded, leaving 14 articles for the primary analysis. However, the sensitivity analysis conducted included the high-risk-of-bias study (and, therefore, reflected all 15 articles).

Primary Analysis

Of the 14 articles included in the primary analysis, 13 reported differences in mean IBDQ-32 total score for a total of 16 comparisons that included the target dose (3 articles^{31,36,40} each reported results from 2 independent studies in which target dose was used), and 4 reported differences in mean IBDQ-32 domain scores for a total of 5 comparisons that included target dose. Of the 16 comparisons for mean IBDQ-32 total scores, 10 were categorized as having efficacious treatments while 6 were categorized as having non-efficacious treatments. Of the 5 comparisons for mean IBDQ-32 domain scores, 3 were categorized as having efficacious treatments while 2 were categorized as having non-efficacious treatments.

Meta-Analyses: Pooled estimates of Hedges' *g* for mean differences in IBDQ-32 total scores (Figure 2, Table 2) were medium-sized (ES: 0.67; 95% confidence interval [CI]: 0.51, 0.83) and statistically larger than zero ($P < 0.001$) when summarized across the 10 efficacious treatments, but were small (ES: 0.23; 95% CI: -0.04, 0.50) and not statistically different from zero ($P = 0.09$) across the 6 non-efficacious treatments. The pooled ES estimate was statistically larger for the efficacious studies than for non-efficacious treatment ($Q(1): 7.5; P = 0.006$).

High heterogeneity was observed among the 10 efficacious studies ($I^2 = 74.1\%$), whereas little heterogeneity was observed among the 6 non-efficacious studies ($I^2 = 0\%$). Further examination indicated that the high heterogeneity among efficacious studies was not driven by a single outlier; when removing the study by Cross et al²⁸ — which reported a much larger treatment effect (ES: 2.02), more than

Total IBDQ-32 score - mean difference

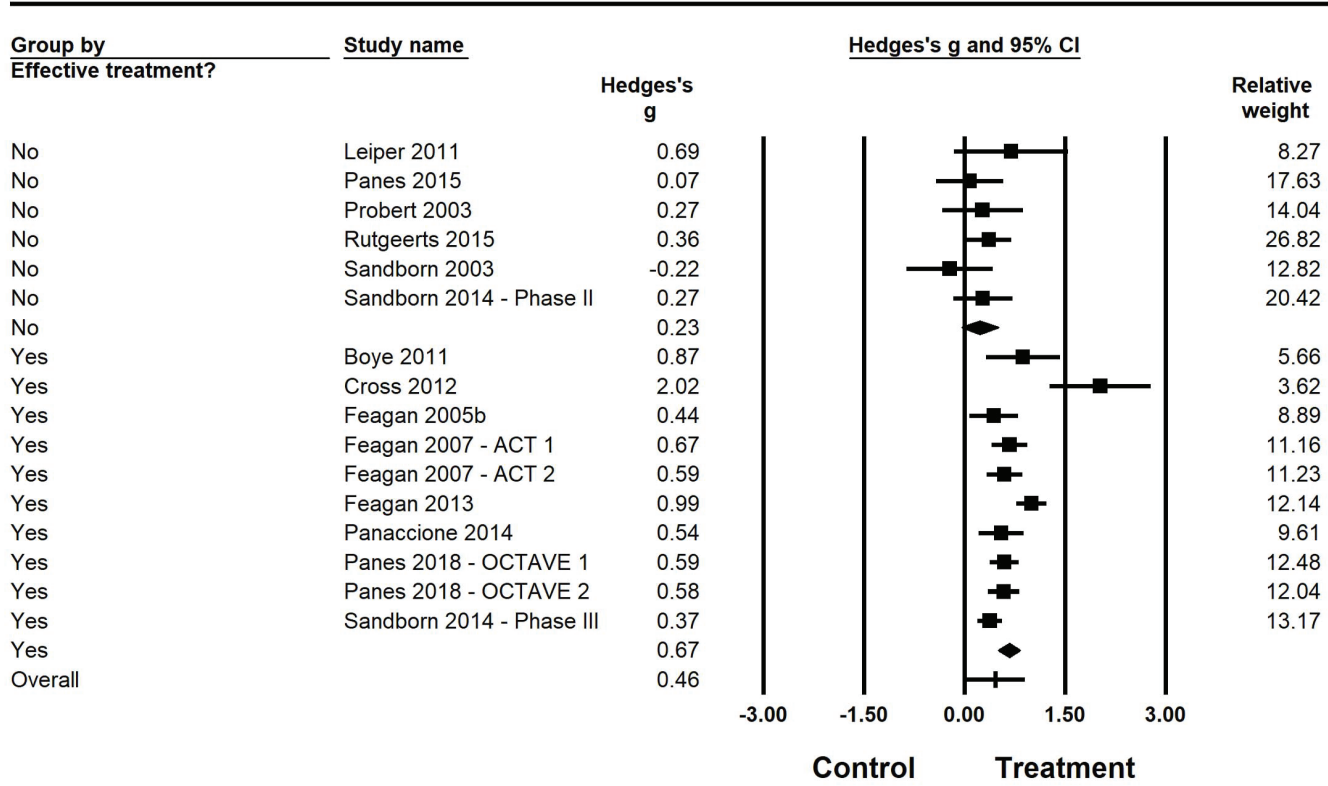


Figure 2. Forest plot for treatment comparisons of mean differences in Inflammatory Bowel Disease Questionnaire (IBDQ-32) total scores for efficacious and non-efficacious treatments. Pooled effect estimates were calculated based on random-effects models, with relative weights for individual comparisons based on inverse variance. The diamond represents a pooled effect estimate for each treatment efficacy subgroup. Data are Hedges' g with 95% confidence intervals.

Table 2. Results From Primary Meta-Analysis

Endpoint	Efficacy Subgroup*	No. of Studies	Within Subgroups			Between Subgroups (Heterogeneity)	
			Pooled Hedges' g (95% CI)	P	I ²	Q (df)	P
Total score	Non-efficacious	6	0.23 (-0.04, 0.50)	0.092	0%	7.45 (1)	0.006
	Efficacious	10	0.67 (0.51, 0.83)	<0.001			
Bowel symptoms	Non-efficacious	2	0.16 (-0.22, 0.53)	0.404	0%	4.51 (1)	0.034
	Efficacious	3	0.58 (0.47, 0.70)	<0.001			
Systemic symptoms	Non-efficacious	2	0.07 (-0.30, 0.45)	0.701	0%	4.50 (1)	0.034
	Efficacious	3	0.50 (0.38, 0.61)	<0.001			
Emotional function	Non-efficacious	2	0.10 (-0.27, 0.47)	0.600	0%	3.82 (1)	0.051
	Efficacious	3	0.49 (0.37, 0.60)	<0.001			
Social function	Non-efficacious	2	0.05 (-0.32, 0.42)	0.793	0%	6.41 (1)	0.011
	Efficacious	3	0.55 (0.44, 0.67)	<0.001			

*Efficacious treatments are those which achieved their primary endpoint; non-efficacious treatments are those for which the primary endpoint was not met.

twice as large as the other studies in this category — there remained substantial heterogeneity among the 9 remaining studies ($I^2=61.9\%$). (Even with the removal of this study from the model, the pooled ES for the 9 remaining efficacious studies [ES: 0.62, 95% CI: 0.48, 0.75] was still statistically significantly greater than for the non-efficacious studies [ES: 0.24; 95% CI: 0.00, 0.47]; $Q(1): 7.5; P=0.006$.)

Pooled estimates of Hedges' g for mean differences in each of the IBDQ-32 domain scores (Figures 3–6, Table 2) were medium-sized (ES range: 0.49 to 0.58) and statistically larger than zero ($P<0.001$ for all) when summarized across the 3 efficacious treatments, but were very small (ES range: 0.05 to 0.16) and not statistically different from zero ($P>0.10$ for all) across the 2 non-efficacious treatments. The pooled ES estimate was statistically larger for the efficacious than non-efficacious studies for all domains ($P<0.10$ for all). Little heterogeneity was observed within studies by efficacy category for domain scores ($I^2=0\%$ for all).

Meta-Regression: Regression coefficients of continuous treatment efficacy (Hedges' g ES) as a predictor of Hedges' g for mean differences for each IBDQ-32 outcome are presented in Table 3. Results from all models indicate a positive association, such that increases in the magnitude of treatment efficacy on the primary efficacy endpoint predict larger mean differences in IBDQ-32 total and domain scores. The regression coefficient of 0.71 found for the ES estimate of treatment efficacy on the primary efficacy endpoint for IBDQ-32 total scores indicates that a 1-unit increase in the ES estimate (ie, an increase of 1 standard deviation in IBDQ total score) for treatment differences in mean change of IBDQ-32 total scores corresponds to a 0.71 increase in the ES estimate for the primary efficacy endpoint (ie, an increase of approximately seven-tenths of a standard deviation in the primary efficacy endpoint). The associations were statistically significant ($P<0.001$) for mean differences in total score, marginally significant for bowel and systemic symptoms domains ($P=0.074$ and $P=0.101$, respectively), and not significant for emotional or social function domains ($P=0.20$ for both).

Meta-Analysis of Other Categorical Moderator Variables: Additional meta-analyses were conducted for treatment comparisons of mean change in IBDQ-32

total scores across other categorical moderator variables for which there were at least 2 included treatment comparisons per category. These analyses found no differences in pooled ES estimates as a function of treatment type (biologic [$n=9$; ES: 0.54, 95% CI: 0.35, 0.72] vs small-molecule [$n=4$; ES: 0.38, 95% CI: 0.10, 0.66]; $P=0.370$), treatment duration (6 weeks [$n=6$; ES: 0.49, 95% CI: 0.28, 0.71] vs 8 weeks [$n=5$; ES: 0.54, 95% CI: 0.33, 0.76]; $P=0.743$), or baseline disease severity (mild-to-moderate [$n=4$; ES: 0.68, 95% CI: 0.31, 1.05] vs moderate-to-severe [$n=11$; ES: 0.54, 95% CI: 0.36, 0.72]; $P=0.508$). Heterogeneity was moderate to high within most subgroups for each of these moderator variable analyses (I^2 values ranged from 63.3% to 86.6% for all subgroups except for those for studies with an 8-week treatment duration [$I^2=11.4\%$]).

Sensitivity Analysis

All Treatment Arms Included: Sensitivity meta-analyses were conducted when combining IBDQ-32 outcomes across all reported treatment arms (not just the treatment arms of the target dose) from studies described in the 14 articles included in the primary analysis.

Pooled ES estimates and 95% CI for treatment comparisons of IBDQ-32 outcomes in this sensitivity analysis are presented in online-only Supplemental Table S5. Relative magnitudes of pooled Hedges' g estimates were generally similar to those observed in the primary analyses. Differences in pooled ES for IBDQ-32 total scores were statistically significant between efficacious studies, for which there was a medium-sized effect (ES: 0.64, 95% CI: 0.48, 0.80), and non-efficacious studies, for which there was a small effect (ES: 0.27, 95% CI: 0.02, 0.51); $Q(1): 6.3; P=0.012$. As in the primary analysis, high heterogeneity was observed among efficacious studies ($I^2=76.6\%$) while there was very little heterogeneity among non-efficacious studies ($I^2=0.6\%$).

Further, statistically larger ES for efficacious studies than non-efficacious studies were found for differences in change scores for 3 of the 4 IBDQ-32 domains: systemic symptoms, emotional function, and social function ($P<0.10$ for all).

High-Risk-of-Bias Studies Included: Sensitivity meta-analyses were conducted when including the high-risk-of-bias study²⁹ ($n=15$ articles) in meta-analytic models based on target treatment arms.

Bowel symptoms domain - mean difference

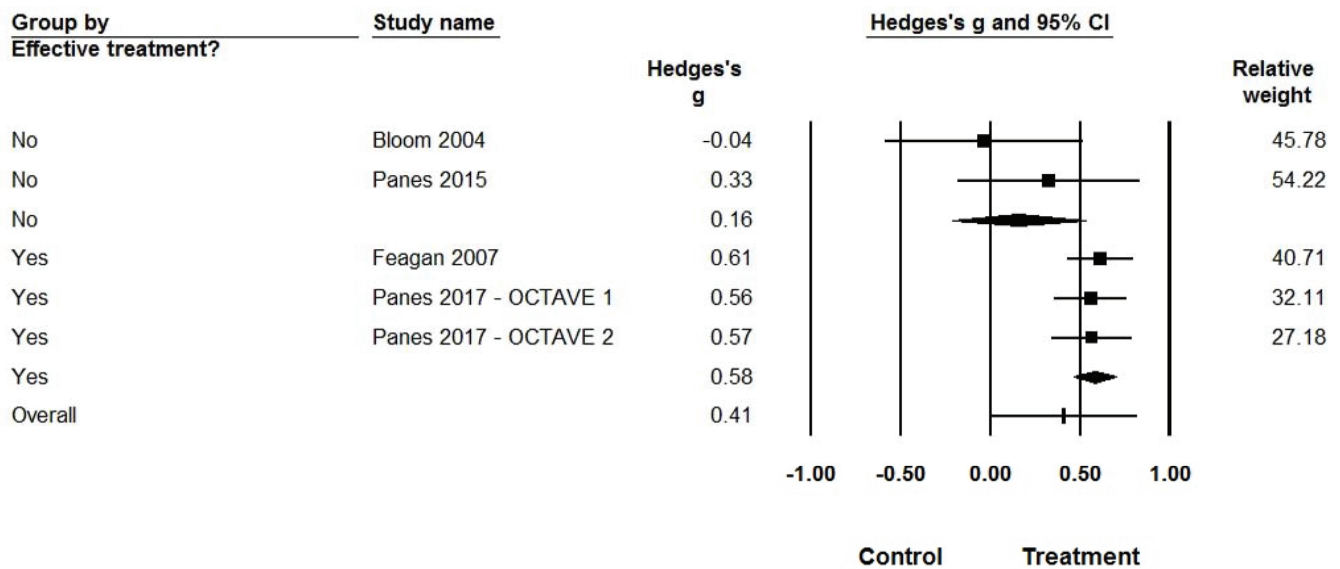


Figure 3. Forest plot for treatment comparisons of mean differences in Inflammatory Bowel Disease Questionnaire (IBDQ-32) bowel symptom domain scores for efficacious and non-efficacious treatments. Pooled effect estimates were calculated based on random-effects models, with relative weights for individual comparisons based on inverse variance. The diamond represents a pooled effect estimate for each treatment efficacy subgroup. Data are Hedges' g with 95% confidence intervals.

Systemic symptoms domain - mean difference

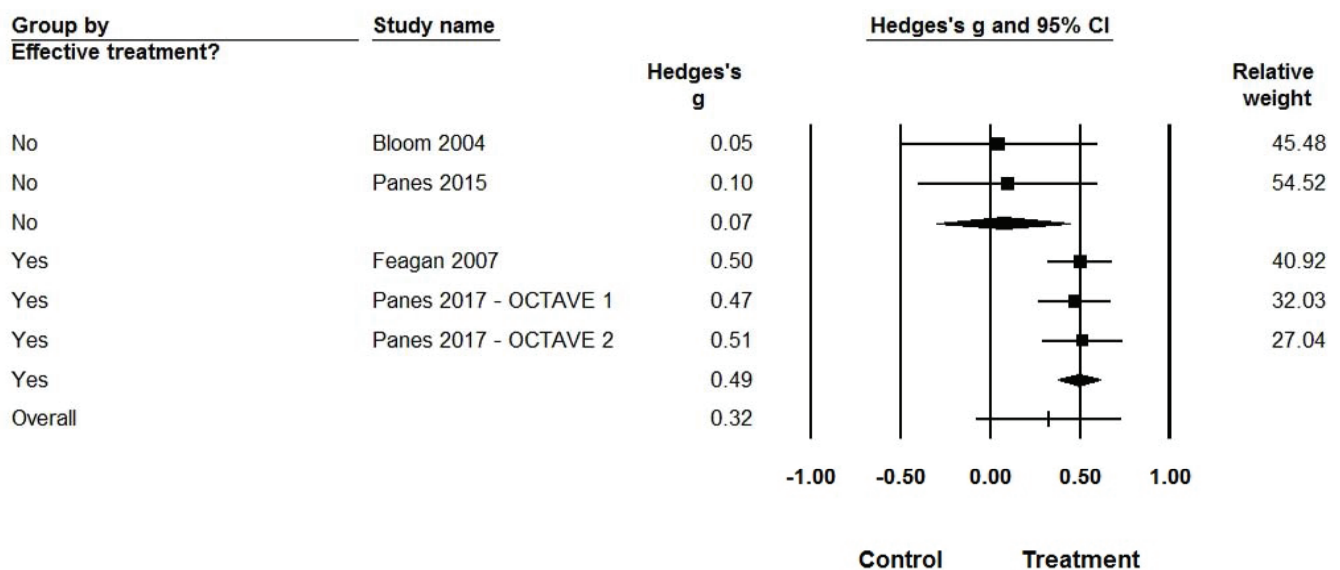


Figure 4. Forest plot for treatment comparisons of mean differences in Inflammatory Bowel Disease Questionnaire (IBDQ-32) systemic symptom domain scores for efficacious and non-efficacious treatments. Pooled effect estimates were calculated based on random-effects models, with relative weights for individual comparisons based on inverse variance. The diamond represents a pooled effect estimate for each treatment efficacy subgroup. Data are Hedges' g with 95% confidence intervals.

Emotional function domain - mean difference

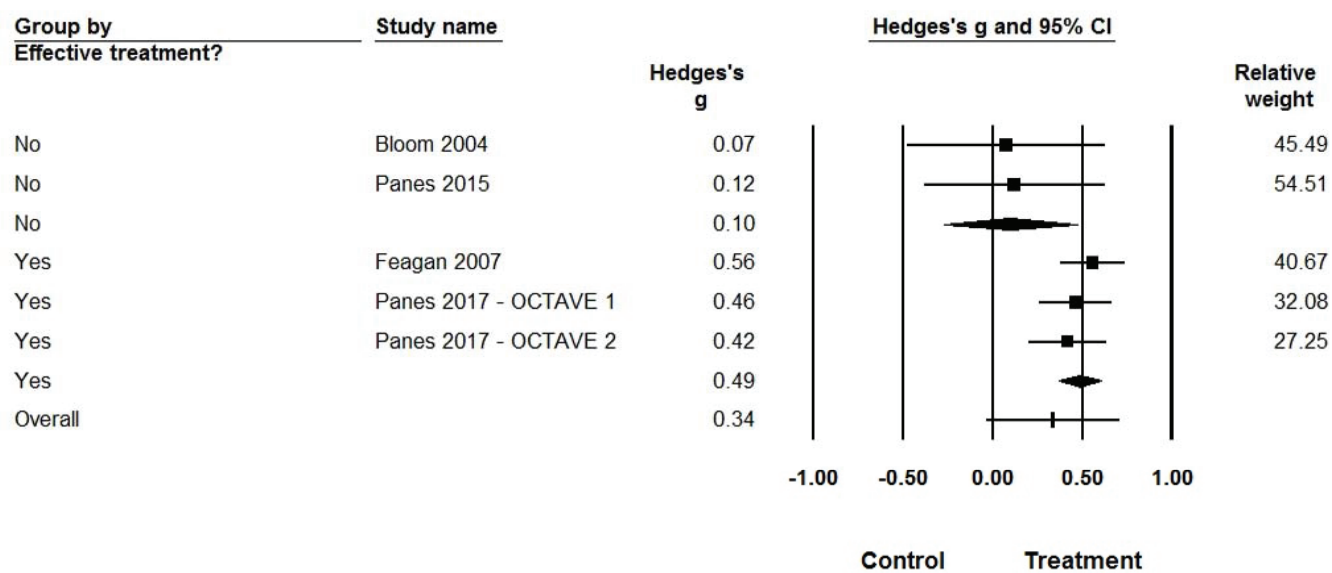


Figure 5. Forest plot for treatment comparisons of mean differences in Inflammatory Bowel Disease Questionnaire (IBDQ-32) emotional function domain scores for efficacious and non-efficacious treatments. Pooled effect estimates were calculated based on random-effects models, with relative weights for individual comparisons based on inverse variance. The diamond represents a pooled effect estimate for each treatment efficacy subgroup. Data are Hedges' g with 95% confidence intervals.

Social function domain - mean difference

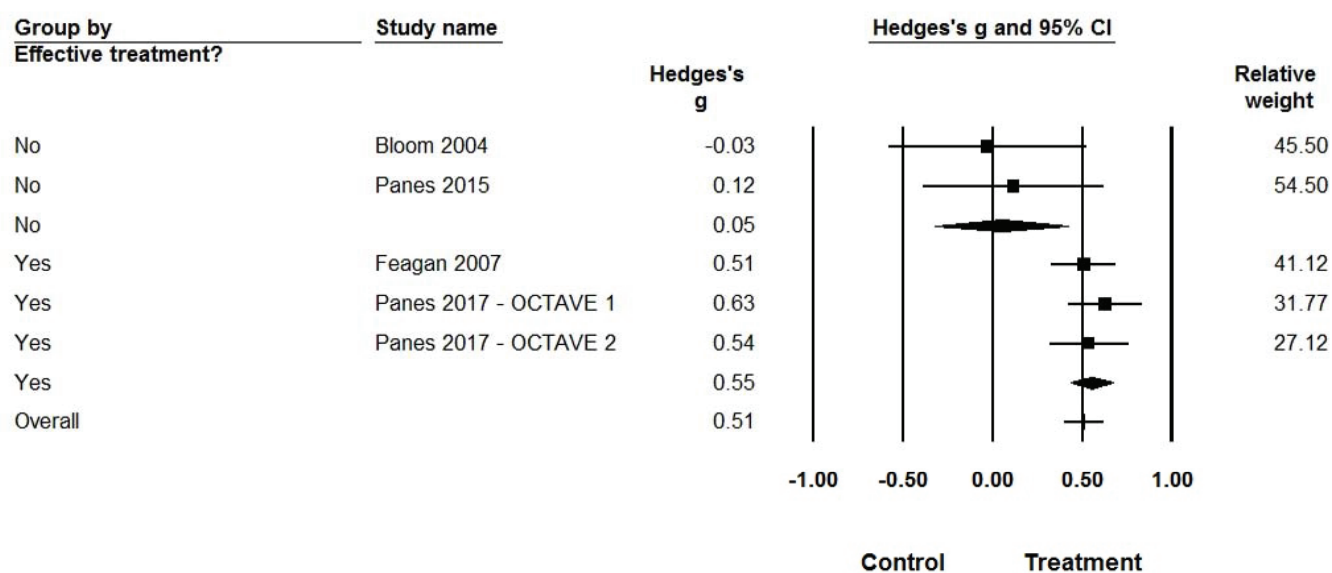


Figure 6. Forest plot for treatment comparisons of mean differences in Inflammatory Bowel Disease Questionnaire (IBDQ-32) social function domain scores for efficacious and non-efficacious treatments. Pooled effect estimates were calculated based on random-effects models, with relative weights for individual comparisons based on inverse variance. The diamond represents a pooled effect estimate for each treatment efficacy subgroup. Data are Hedges' g with 95% confidence intervals.

Table 3. Results From Primary Meta-Regression Analysis

Outcome Model	Number of Comparisons	Coefficient for Treatment Efficacy*	Standard Error	Z-Score	P
Total score	16	0.71	0.18	3.96	<0.001
Bowel symptoms	5	0.45	0.25	1.79	0.074
Systemic symptoms	5	0.41	0.25	1.64	0.101
Emotional function	5	0.32	0.25	1.29	0.196
Social function	5	0.32	0.25	1.30	0.195

*Based on Hedges' g effect size for treatment effects on primary efficacy variable.

Pooled ES estimates and 95% CI for treatment comparisons of IBDQ-32 outcomes in this sensitivity analysis are presented in online-only Supplemental Table S6. Relative magnitudes of pooled Hedges' g estimates for differences in IBDQ-32 total scores were similar to those observed in the primary analyses, with a statistically significantly larger effect for studies with efficacious treatment (ES: 0.67, 95% CI: 0.51, 0.83) than for studies with non-efficacious treatments (ES: 0.26, 95% CI: 0.01, 0.52; Q(1): 6.99; P<0.001). Pooled estimates of Hedges' g for mean differences for bowel symptoms, systemic symptoms, and social function domain scores for efficacious studies were medium-sized and statistically larger than zero (ES range: 0.50 to 0.58; P<0.001 for all) but for non-efficacious studies were small and not statistically different from zero (ES range: 0.08 to 0.25, P>0.10 for all), with the pooled ES estimate statistically significantly larger for the efficacious than non-efficacious studies for the 3 domains (P<0.10 for all).

The pooled estimate of Hedges' g for mean differences in emotional function domain scores for efficacious studies was medium-sized and statistically larger than zero (ES: 0.49, 95% CI: 0.37, 0.60; P<0.001). For non-efficacious studies it was small and not significantly different from zero (ES: 0.24, 95% CI: -0.09, 0.58; P=0.148), although the pooled ES estimate did not statistically differ between the efficacious and non-efficacious studies for this domain (Q(1): 1.86; P=0.172).

Publication Bias

Evaluation of publication bias was conducted for treatment comparisons of mean differences in IBDQ-32 total scores for the primary analyses, which

included 13 of the 14 articles (93%) and 16 of the 19 comparisons (84%). Visual examination of the funnel plot (online-only Supplemental Figure S4) and results from Egger's regression test (2-tailed P=0.888) indicated that publication bias was unlikely an issue of concern for this meta-analysis.

DISCUSSION

To the best of our knowledge, this paper provides the first systematic literature review and synthesis of evidence regarding the sensitivity of the IBDQ-32 to treatment for UC. Specifically, the objective of this study was to examine the impact of efficacious treatment on mean change in IBDQ-32 outcomes. We defined treatment efficacy both as a dichotomous variable, based on success of meeting the primary study endpoint, and also as a continuous variable, by calculating an ES estimate for the treatment efficacy of that endpoint. For both of these approaches, results were pooled across studies reported in 14 published articles to investigate whether treatment efficaciousness was a predictor of change in IBDQ-32 total and domain scores.

For the dichotomous efficacy approach, the primary meta-analysis found that patients in studies with efficacious treatments showed larger mean improvements relative to controls in IBDQ-32 total scores and all 4 domains. At the same time, patients in studies with non-efficacious treatments showed small and nonsignificant differences in these outcomes relative to controls. For the continuous efficacy approach, the primary meta-regression models showed that the magnitude of treatment efficacy was a positive predictor of these same IBDQ-32 outcomes. Results from sensitivity meta-analyses (of 15 articles) were generally supportive of findings from the primary

analysis. Thus, this disease-specific patient-centered measure of HRQoL showed strong correspondence to changes in clinical health as a function of treatment interventions.

Meta-analyses were conducted to examine evidence that effects for treatment comparisons of mean changes in IBDQ-32 total score were associated with potential moderator variables, including treatment type, treatment duration, and severity of patients' baseline disease activity. No differences in pooled ES estimate across subgroups were found for any of these factors. High heterogeneity was found across ES estimates within most of these subgroups, which may be due to the fact that each subgroup contained a mix of efficacious and non-efficacious studies, across which there were established differences in ES magnitude. In the primary meta-analysis, high heterogeneity also was observed across ES estimates for mean changes in IBDQ-32 total scores within the subgroup of efficacious studies. Even when removing a study with an outlying ES relative to the rest within the subgroup, a moderate amount of heterogeneity was still observed. Within this subgroup, there were a number of key differences among studies, including factors such as types of treatments administered, type of disease activity index measure used to capture clinical efficacy, treatment duration, baseline severity of UC, and potentially demographic factors. While moderator analyses found that some of these factors did not explain statistically significant differences in ES estimates across studies, other factors varied too greatly among studies to combine into meaningful subgroups. The combined and possibly interactive impacts of each of these factors on heterogeneity in ES estimates cannot be easily assessed given the limited number of studies.

The current analysis focused on the inclusion of the IBDQ-32 in studies of patients with UC. However, the IBDQ-32 is often included as an HRQoL endpoint in studies of patients with Crohn's disease.⁹ While UC and Crohn's disease are often studied together, they represent distinct pathophysiological entities.^{44,45} Because our a priori research interest was in UC, it would have been beyond the scope of our research question to explore Crohn's disease in this study. We predict that if the same type of meta-analysis was

conducted for studies using the IBDQ-32 with patients with Crohn's disease, the results would be similar to those reported here for UC. In the absence of such data, however, we defer to future research to answer whether our results also apply to Crohn's disease.

The primary reason for including the IBDQ-32 as a key secondary endpoint in clinical trials is that it enables a broader interpretation of treatment benefit. When asked to describe their experiences with the disease, patients with UC often mention clinical symptoms that are typically assessed using disease activity indices. Nevertheless, these patients also express concerns about their disease and its impact on their everyday lives that go far beyond what is captured by those indices. They report anxieties stemming from a lack of control over their bodily functions, fear of disease progression, hospitalization or surgery, and fear of not having immediate access to a toilet.⁴⁶⁻⁴⁸ These concerns impact their employment opportunities and work productivity and limit their ability to engage in social and recreational activities, which can impair their ability to develop and maintain relationships with others and subsequently lead to difficulties achieving intimacy and to feelings of isolation and depression.^{46,47,49-51} Thus, comprehensively evaluating benefits of a treatment for patients with UC entails more than merely assessing changes in clinical and endoscopic activity — it also requires measuring changes in patient-reported HRQoL. The results provided here support the use of the IBDQ-32 to capture treatment benefits on functioning and well-being in patients with UC.

Limitations

A limitation of this study is that the literature search was restricted to papers published in peer review journals, with findings in the "gray" literature (eg, conference presentations) not represented. Given that our study objective was to evaluate what is typically used as a secondary outcome measure, and not to provide evidence supporting efficacy of any particular treatment regimen, we believe that publication bias would be less impactful for our review and that the consequences of possible bias would be relatively low. As the ratio of efficacious treatments to non-efficacious treatments was fairly even (10:9), evidence from our examination of publication bias shows that it is unlikely that an exhaustive search of unpublished studies meeting our

selection criteria would result in substantially different findings than those observed here.

Another limitation of our review stems from the decision to restrict our data sources to only randomized controlled trials, such that our review and meta-analysis did not include data on changes in IBDQ-32 scores from interventional studies of patients with UC that used other study designs, such as single-arm or nonrandomized comparative designs. This decision was made because the design of randomized controlled trials reduces bias by minimizing confounds between patient groups (due to randomization) and by controlling for placebo effects on HRQoL, which have been demonstrated to be substantial in studies of patients with UC.⁵² However, by restricting our evidence base to only randomized controlled trials, we limit the generalizability of our findings and ignore data that could be relevant to the question of the degree to which IBDQ-32 scores are sensitive to the clinically efficacious UC treatments. Future research incorporating findings from studies using nonrandomized design could potentially provide additional information with respect to the research questions addressed in this review.

Another potential limitation of this analysis was a lack of uniformity in determining whether or not a treatment was efficacious. We relied on each study's authors for the determination about whether the primary efficacy endpoint was met, according to the a priori definition for that study. The fact that many studies used similar definitions of efficacy (ie, significant mean difference between target and placebo groups on a disease activity measure, difference in proportion achieving response or remission as defined by changes on disease activity indices) reduced the potential variability in these definitions across studies. The lack of overlap in ES estimates for non-efficacious and efficacious target treatments (Hedges' *g* ranging from -0.75 to 0.42 and from 0.43 to 1.66, respectively) supports that the dichotomous classification fairly represented actual treatment efficacy.

CONCLUSIONS

This review provides new evidence that IBDQ-32 total and domain scores are sensitive and specific to efficacious treatment, such that treatments that effectively produced reduction of clinical and

endoscopic UC symptoms also produced improvements in IBDQ-32 scores, while treatments that did not reduce UC disease activity did not produce substantial changes in IBDQ-32 outcomes. Integrating these findings with those from reviews providing evidence for the reliability, validity, and responsiveness of the IBDQ-32 for measuring IBD-specific HRQoL in patients with UC^{7,8,11,12} supports the appropriateness of including this instrument in clinical trials.

Patient-Friendly Recap

- Research studies on ulcerative colitis often use a tool called the Inflammatory Bowel Disease Questionnaire (IBDQ-32) to measure changes in a patient's health-related quality of life following treatment. However, this tool has not been validated as useful in randomized clinical trials.
- The authors reviewed published results from 14 randomized controlled trials that used the IBDQ-32 in this patient population to determine whether the tool performed to expectations.
- They found that IBDQ-32 scores do indeed reflect changes from treatment, supporting the use of this tool to capture quality-of-life outcomes in patients participating in ulcerative colitis clinical trials.

Author Contributions

Study design: Yarlas, Maher, Bayliss, Cappelleri, Bushmakin, DiBonaventura. Data acquisition or analysis: Yarlas, Maher, Lovley. Manuscript drafting: Yarlas. Critical revision: Maher, Bayliss, Lovley, Cappelleri, Bushmakin, DiBonaventura.

Conflicts of Interest / Funding Sources

This research was funded by Pfizer Inc. (New York, NY). Authors Yarlas, Maher, Bayliss, and Lovley are employees of Optum, Inc. (Johnston, RI), which received funding from Pfizer to conduct the literature review and develop this manuscript. Authors Cappelleri, Bushmakin, and DiBonaventura are employees and shareholders of Pfizer.

References

1. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV Jr. Incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota from 1970 through 2010. *Clin Gastroenterol Hepatol*. 2017;15:857-63. [CrossRef](#)

2. Gibson PR, Vaizey C, Black CM, et al. Relationship between disease severity and quality of life and assessment of health care utilization and cost for ulcerative colitis in Australia: a cross-sectional, observational study. *J Crohns Colitis*. 2014;8:598-606. [CrossRef](#)
3. Han SW, McColl E, Barton JR, James P, Steen IN, Welfare MR. Predictors of quality of life in ulcerative colitis: the importance of symptoms and illness representations. *Inflamm Bowel Dis*. 2005;11:24-34. [CrossRef](#)
4. Yarlás A, Bayliss M, Cappelleri JC, et al. A systematic review of the SF-36® Health Survey for measuring health-related quality of life in patients with ulcerative colitis. (abstr.) *Value Health*. 2016;19(7):A514. [CrossRef](#)
5. Zahn A, Hinz U, Karner M, Ehehalt R, Stremmel W. Health-related quality of life correlates with clinical and endoscopic activity indexes but not with demographic features in patients with ulcerative colitis. *Inflamm Bowel Dis*. 2006;12:1058-67. [CrossRef](#)
6. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med*. 1987;317:1625-9. [CrossRef](#)
7. Alrubaiy L, Rikaby I, Dodds P, Hutchings HA, Williams JG. Systematic review of health-related quality of life measures for inflammatory bowel disease. *J Crohns Colitis*. 2015;9:284-92. [CrossRef](#)
8. Chen XL, Zhong LH, Wen Y, et al. Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties. *Health Qual Life Outcomes*. 2017;15(1):177. [CrossRef](#)
9. Williet N, Sandborn WJ, Peyrin-Biroulet L. Patient-reported outcomes as primary end points in clinical trials of inflammatory bowel disease. *Clin Gastroenterol Hepatol*. 2014;12:1246-56.e6. [CrossRef](#)
10. Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology*. 1989;96:804-10. [CrossRef](#)
11. Achleitner U, Coenen M, Colombel JF, Peyrin-Biroulet L, Sahakyan N, Cieza A. Identification of areas of functioning and disability addressed in inflammatory bowel disease-specific patient reported outcome measures. *J Crohns Colitis*. 2012;6:507-17. [CrossRef](#)
12. Pallis AG, Mouzas IA, Vlachonikolis IG. The inflammatory bowel disease questionnaire: a review of its national validation studies. *Inflamm Bowel Dis*. 2004;10:261-9. [CrossRef](#)
13. D'Haens G, Sandborn WJ, Feagan BG, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology*. 2007;132:763-86. [CrossRef](#)
14. Mitchell A, Guyatt G, Singer J, et al. Quality of life in patients with inflammatory bowel disease. *J Clin Gastroenterol*. 1988;10:306-10. [CrossRef](#)
15. Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 1999;28:S23-7. [CrossRef](#)
16. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med*. 2009;3:e123-30.
17. Drevon D, Fursa SR, Malcolm AL. Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behav Modif*. 2017;41:323-39. [CrossRef](#)
18. Higgins PD, Schwartz M, Mapili J, Zimmermann EM. Is endoscopy necessary for the measurement of disease activity in ulcerative colitis? *Am J Gastroenterol*. 2005;100:355-61.
19. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med*. 2000;19:3127-31. [CrossRef](#)
20. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to Meta-Analysis*. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2009.
21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences, Second Edition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
22. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60. [CrossRef](#)
23. Haidich AB. Meta-analysis in medical research. *Hippokratia*. 2010;14(Suppl 1):29-37.
24. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med*. 1998;17:841-56. [CrossRef](#)
25. Lee YH. An overview of meta-analysis for clinicians. *Korean J Intern Med*. 2018;33:277-83. [CrossRef](#)
26. Bloom S, Kiilerich S, Lassen MR, et al. Low molecular weight heparin (tinzaparin) vs. placebo in the treatment of mild to moderately active ulcerative colitis. *Aliment Pharmacol Ther*. 2004;19:871-8. [CrossRef](#)
27. Boye B, Lundin KE, Jantschek G, et al. INSPIRE study: does stress management improve the course of inflammatory bowel disease and disease-specific quality of life in distressed patients with ulcerative colitis or Crohn's disease? A randomized controlled trial. *Inflamm Bowel Dis*. 2011;17:1863-73. [CrossRef](#)
28. Cross RK, Cheevers N, Rustgi A, Langenberg P, Finkelstein J. Randomized, controlled trial of home telemanagement in patients with ulcerative colitis (UC HAT). *Inflamm Bowel Dis*. 2012;18:1018-25. [CrossRef](#)
29. Elsenbruch S, Langhorst J, Popkirowa K, et al. Effects of mind-body therapy on quality of life and neuroendocrine and cellular immune functions in patients with ulcerative colitis. *Psychother Psychosom*. 2005;74:277-87. [CrossRef](#)
30. Feagan BG, Greenberg GR., Wild G, et al. Treatment of ulcerative colitis with a humanized antibody to the alpha4beta7 integrin. *N Engl J Med*. 2005;352:2499-507. [CrossRef](#)
31. Feagan BG, Reinisch W, Rutgeerts P, et al. The effects of infliximab therapy on health-related quality of life in ulcerative colitis patients. *Am J Gastroenterol*. 2007;102:794-802.
32. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med*. 2013;369:699-710. [CrossRef](#)
33. Leiper K, Martin K, Ellis A, et al. Randomised placebo-controlled trial of rituximab (anti-CD20) in active ulcerative colitis. *Gut*. 2011;60:1520-6. [CrossRef](#)
34. Panaccione R, Ghosh S, Middleton S, et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. *Gastroenterology*. 2014;146:392-400.e3. [CrossRef](#)
35. Panés J, Su C, Bushmakina AG, Cappelleri JC, Mamolo C, Healey P. Randomized trial of tofacitinib in active ulcerative colitis: analysis of efficacy based on patient-reported outcomes. *BMC Gastroenterol*. 2015;15:14. [CrossRef](#)
36. Panés J, Vermeire S, Lindsay JO, et al. Tofacitinib in patients with ulcerative colitis: health-related quality of life in phase 3 randomised controlled induction and maintenance studies. *J Crohns Colitis*. 2018;12:145-56. [CrossRef](#)

37. Probert CS, Hearing SD, Schreiber S, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. *Gut*. 2003;52:998-1002. [CrossRef](#)
38. Rutgeerts P, Feagan BG, Marano CW, et al. Randomised clinical trial: a placebo-controlled study of intravenous golimumab induction therapy for ulcerative colitis. *Aliment Pharmacol Ther*. 2015;42:504-14. [CrossRef](#)
39. Sandborn WJ, Sands BE, Wolf DC, et al. Repifermin (keratinocyte growth factor-2) for the treatment of active ulcerative colitis: a randomized, double-blind, placebo-controlled, dose-escalation trial. *Aliment Pharmacol Ther*. 2003;17:1355-64. [CrossRef](#)
40. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2014;146:85-95; quiz e14-5. [CrossRef](#)
41. Rachmilewitz D. Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. *BMJ*. 1989;298:82-6. [CrossRef](#)
42. Seo M, Okada M, Yao T, Ueki M, Arima S, Okumura M. An index of disease activity in patients with ulcerative colitis. *Am J Gastroenterol*. 1992;87:971-6.
43. Sutherland LR, Martin F, Greer S, et al. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. *Gastroenterology*. 1987;92:1894-8. [CrossRef](#)
44. de Souza HS, Fiocchi C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol*. 2016;13:13-27. [CrossRef](#)
45. Feakins RM. Ulcerative colitis or Crohn's disease? Pitfalls and problems. *Histopathology*. 2014;64:317-35. [CrossRef](#)
46. Devlen J, Beusterien K, Yen L, Ahmed A, Cheifetz AS, Moss AC. The burden of inflammatory bowel disease: a patient-reported qualitative analysis and development of a conceptual model. *Inflamm Bowel Dis*. 2014;20:545-52. [CrossRef](#)
47. Sammut J, Scerri J, Xuereb RB. The lived experience of adults with ulcerative colitis. *J Clin Nurs*. 2015;24:2659-67. [CrossRef](#)
48. Waljee AK, Joyce JC, Wren PA, Khan TM, Higgins PD. Patient reported symptoms during an ulcerative colitis flare: a qualitative focus group study. *Eur J Gastroenterol Hepatol*. 2009;21:558-64. [CrossRef](#)
49. McCormick JB, Hammer RR, Farrell RM, et al. Experiences of patients with chronic gastrointestinal conditions: in their own words. *Health Qual Life Outcomes*. 2012;10:25. [CrossRef](#)
50. Wolfe BJ, Sirois FM. Beyond standard quality of life measures: the subjective experiences of living with inflammatory bowel disease. *Qual Life Res*. 2008;17:877-86. [CrossRef](#)
51. Jansen F, van Uden-Kraan CF, Braakman JA, van Keizerswaard PM, Witte BI, Verdonck-de Leeuw IM. A mixed-method study on the generic and ostomy-specific quality of life of cancer and non-cancer ostomy patients. *Support Care Cancer*. 2015;23:1689-97. [CrossRef](#)
52. Estevinho MM, Afonso J, Rosa I, et al. Placebo effect on the health-related quality of life of inflammatory bowel disease patients: a systematic review with meta-analysis. *J Crohns Colitis*. 2018;12:1232-44. [CrossRef](#)
53. Panés J, Vermeire S, Lindsay JO, et al. Tofacitinib in patients with ulcerative colitis: health-related quality of life in phase 3 randomised controlled induction and maintenance studies. *J Crohns Colitis*. 2019;13:139-40. [CrossRef](#)

© 2020 Aurora Health Care, Inc.