



Please download and read the instructions before proceeding to the peer review

Efficacy of Tumor Necrosis Factor-alpha therapy in the management of paediatric perianal Crohn's disease: a systematic review

Journal:	<i>Expert Opinion On Biological Therapy</i>
Manuscript ID	EOBT-2019-ST-0094
Manuscript Type:	Review
Keywords:	biological drugs, Crohn's disease, drug safety, paediatrics

SCHOLARONE™
Manuscripts

Efficacy of Tumor Necrosis Factor-alpha therapy in the management of paediatric perianal Crohn's disease: a systematic review

Keywords: biologic drugs, Crohn’s disease, drug safety, paediatrics

ABSTRACT

INTRODUCTION Anti-Tumor Necrosis Factor-alpha (TNF- α) therapy, primarily infliximab and adalimumab, are now increasingly used to induce and maintain disease remission in the paediatric perianal Crohn’s disease (CD) population, however, their optimal use has not yet been defined in the paediatric setting.

AREAS COVERED In accordance with a published protocol (PROSPERO no. CRD42019118838), we systematically and critically evaluated all published evidence on the efficacy and safety of anti-TNF- α in children with perianal CD, in the PubMed, MEDLINE, Embase, Cochrane and clinical trials.gov databases until October, 18th, 2018. We included in our systematic review 29 articles yielding a total of 565 perianal CD patients aged between 9 months to 18 years.

EXPERT OPINION

According to low-quality evidence from small, uncontrolled and heterogeneous descriptive studies, and very few RCTs, nearly three-fifths children with perianal CD achieved remission with anti-TNF- α treatment and in approximately 40% remission was maintained after 12 months, with practically low discontinuation rate due to serious adverse events. More than half of the patients achieved complete fistula closure. There is still a need for more robust evidence adequately assessing the efficacy and safety of anti-TNF- α therapy in paediatric perianal CD, as well as in comparison with other therapies.

1. INTRODUCTION

The development of perianal abscesses and fistulas are recognised complications of Crohn’s disease (CD). This complication is inflammation at or near the anus, including tags, fissures, fistulae, abscesses, or stenosis, and is accompanied by pain, itching, bleeding, purulent discharge and incontinence of stool. It can lead to significant morbidity and reduced quality of life [1]. Furthermore, a fistulising perianal CD is a risk factor for poor outcomes, with about 70% of patients requiring surgical treatment [2-4].

Recent studies report an increased incidence of perianal CD in paediatric patients, up to 50.7% [5-10]. Especially in children and young adults, perianal CD may be refractory to medical treatment [7,11] with consequent significant impact on the quality of life [12]. These children are at high risk of psychological distress and disability as a result of their perianal CD [13-16].

Therapy with biologic drugs inhibiting Tumor Necrosis Factor- α (TNF- α), primarily infliximab and adalimumab, are now increasingly used to induce and maintain disease remission. In the paediatric CD population and overall good effectiveness on disease course has been described [8,17,18], however, the efficacy on complications is still debated.

While some studies have indeed reported higher rates of fistula closure [19], it appears that a significant number of children fail to achieve disease remission or recurrence of their fistulae despite initial success with an anti-TNF- α therapy. The management of paediatric perianal CD remains thus challenging and frequently relying on surgical resection of the diseased bowel [20,21].

The aim of this study was thus to provide a clearer picture of the actual information on perianal CD by evaluating systematically and critically all published evidence on the efficacy and safety of anti-TNF- α (infliximab, adalimumab and certolizumab) for the induction and maintenance treatment in children with perianal CD. The results we report appear important to guide treatment decision in the clinical practice.

2. METHODS

The protocol for this systematic review was submitted to and approved by PROSPERO (PROSPERO registration no. CRD42019118838) [22]. Reporting is in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) and Meta-analyses Of Observational Studies in Epidemiology (MOOSE) guidelines (Supplementary Table 1 and Table 2, respectively) [23,24]. A multidisciplinary panel of experts was formed to determine the review protocol and carry out all aspects of the review.

2.1 Data Sources and Searches

All studies evaluating the efficacy of anti-TNF- α therapy in the management of paediatric perianal Crohn's disease were searched in the PubMed, MEDLINE, Embase, Cochrane and Clinicaltrials.gov databases from any time until October, 18th, 2018 with no language and study type restriction. The search strategy is available in Supplementary Figure 1. The search strategy was adapted as needed for multiple databases. A hand-search of articles from relevant reviews was conducted to identify studies for potential inclusion. Other studies from cross-references were included (see the PRISMA diagram in Figure 1).

2.2 Study selection

All titles and abstracts were assessed independently and in duplicate to identify potentially relevant articles and exclude the ones that were obviously irrelevant. Studies fulfilling the following criteria were included: articles in English, French and Spanish languages, peer-reviewed studies carried out in children and adolescents with paediatric onset [≤ 18 years] and reporting sufficient data to enable assessment of treatment outcomes of induction or maintenance of response or remission, and/or fistula outcomes in a perianal CD following anti-TNF- α therapy, without study type restriction. We excluded studies that did not report treatment outcomes of perianal CD or the outcomes of perianal CD were not differentiable from the total cohort. Selected full-texts were reviewed in duplicate independently. Reasons for exclusion of full-texts were recorded. Disagreements between reviewers were resolved by consensus and consultation with the expert group.

2.3 Outcome Measures

The outcome measures were the proportion of patients with fistula closure, proportion of patients with response or remission [as classified by Authors of the primary study], proportion of patients categorised as primary non-responders or loss of response, relapse rate at the any of the study time points and the number of patients who underwent surgical resection during anti-TNF- α therapy. The other secondary outcome measures were safety measures defined as the proportion of patients with: any adverse event, serious adverse event [as classified by Authors of the primary study], and withdrawals due to the adverse events.

2.4 Data extraction and quality assessment

Data were independently extracted by two researchers using pre-specified forms. For the included studies, information was extracted on study type (clinical trial; observational study; case study), study design (randomised double-blind; open-label; cross-sectional; prospective; retrospective), duration of follow-up, patient characteristics (age; baseline severity of disease; number of patients with perianal CD; type of perianal manifestations; treatment refractoriness; previous surgeries; percentage of anti-TNF- α -naïve patients), anti-TNF- α therapy (anti-TNF- α drug; dosage), clinical efficacy with other relevant outcome, and adverse effects. Discrepancies between the extracted data were resolved via consensus. Record management was performed using Endnote.

The quality of the individual studies was assessed at the study level by one reviewer and independently checked by a second; disagreements were resolved by consensus. Observational studies were assessed with the Newcastle-Ottawa Scale (NOS) [25]. NOS is a checklist with eight items that outline three quality components: selection, comparability, outcome (Supplementary Table 3). Each item is scored one or two

and summed for a total indicating high (0-4), moderate (5-6), and low (7-9) risk of bias. Any discrepancy was resolved by consultation. Supplementary Table 3 reports the NOS used for this study. For RCTs, we used the Downs and Black Checklist [26]. It consists of 27 questions on reporting quality (10 questions), external validity (3), internal validity (bias and confounding, (13), and statistical power (1). Most scores range from 0 to 1, except one item on the reporting confounders subscale ranging from 0 to 2. The scale has a maximum score of 28; each paper was graded “excellent” (26-28 points), “good” (20-25 points), “fair” (15-19 points) or “poor” (<14 points). The methodological quality of case series and case reports were assessed using a 9-item validated quality appraisal tool for case series [27]. Quality appraisal judgments for each item are determinations of various features of the study including study objectives, population, interventions and co-interventions, outcome measures, statistical analysis, and results.

2.5 Statistical Analysis

Pooled results on the clinical efficacy outcomes [remission, response, fistula Closure] were reported as the percentage of patients with the outcome, calculated as the rate between the total absolute number of patients with the event, on the total absolute number of patients treated at different time points and overall. For the safety outcomes, we used cumulative frequencies of different AE as reported in the primary studies. Parametric data are presented as mean and standard deviation (SD), nonparametric data are presented as median followed by range or interquartile range unless otherwise stated. The pooled proportions and corresponding 95% confidence interval (95% CI) of anti-TNF- α treatment response and remission rates were calculated. As data could not be pooled in forest plots, findings were reported in tables and text.

3. RESULTS

The initial search identified 726 unique articles. After screening the titles and abstracts, a total of 62 articles underwent full-text review, with 29 articles ultimately meeting the inclusion criteria (yielding a total of 565 PCD patients aged between 9 months to 18 years). The study selection process is reported through a PRISMA diagram in Figure 1.

Only two RCTs [28,29] including 38 perianal CD paediatric patients were identified. The remaining 28 studies were either post-hoc analysis of RCTs [30-31], retrospective or prospective observational [8,19-21,32-38], case series or case reports [39-52]. Table 1 presents the characteristics of the included studies. Supplementary Table 1 summarises the risk of bias scores for each included study. The score of quality for randomised controlled trials included in our analysis was fair or poor, and all observational studies scored 6 to 8 on the NOS.

The included studies differed regarding the number of perianal CD patients treated with anti- TNF- α [n=1 to 101], baseline disease severity, concomitant treatment, treatment refractoriness, remission and response criteria (table 2). Only five studies [21,30,32,42,45] evaluated adalimumab in perianal CD patients and no study evaluated certolizumab in perianal CD patients. In observational studies, there were differences between individual patients regarding anti-TNF- α dosage. Moreover, in some studies, characteristics were not differentiable for perianal CD patient from the total patient cohort.

3.1 Efficacy outcome

Pooled remission rates, response rates, fistula closure rate, primary non-responder rates, secondary loss of response rates, and rates of surgical treatment with 95% CI estimates are summarised in table 3. Out of 565 patients included in the analysis, 326 (60%) achieved clinical remission and 310 (55%) had a clinical response. Early remission rate at the 2nd week of therapy ranged from 10% to 37% [31,34,52] and at 6th week of treatment ranged from 35% to 49% [19,20,31,35,36,39,40,47]. Remission rates for maintenance therapy were as follows: ranged from 41% to 76% at 8 weeks [33,35,41]; ranged from 51% to 76% at 10 weeks [28,29,31]; ranged from 10% to 23% at 3 months [8,20,37,43]; ranged from 13% to 36% at 4 to 6 months [20,30,39,44]; and ranged from 25% to 57% at 8 to 9 months [31,45,46]. Delayed/late remission rates were as follows: ranged from 34% to 49% at 12 months [19,32,33,36,43]; ranged from 45% to 70% at 13 to 18 months [30,31,46]; ranged from 7% to 21% at 24 months [8,45] and 14% at 34 months [46].

Four studies [38,41,47,48] reporting short-term response as early as at 1 week ranged from 35% to 92%. Response rates at 2 to 4 weeks ranged from 19% to 39% [20,31,37,42,47]; and at 6 weeks ranged from 65% to 80% [19,31,35,39,40,49]. The late response rates were as follows: at 8th week ranged from 17% to 52% [33,35]; at 10th week ranged from 55% to 77% [29,31,50]; at 7 to 9 months ranged from 3% to 24 % [20,31,45]; at 12 months ranged from 41% to 56% [19,33,36] and 58% at 13 months [30]. Three studies [31,45,46] reported response rates beyond 13 months ranged from 4% to 27%. Twelve studies [8,20,28-30,33,34,36,37,39,40,47] reported the efficacy of anti-TNF alpha treatment on perianal fistula outcome. Overall, a complete closure of the perianal fistula was reported in 127/ 228 (56%) of patients.

Overall, 10/29 studies [8,19,20,29,31,33,35,36,48,51] including 37 patients reported primary non-response rates ranged from 8% to 15%. Loss of response or secondary non-response rates ranged from 14% to 25% [19,20,33-36,39,48]. The pooled incidence rate of surgical interventions required during treatment ranged from 28% to 39% [including 276 patients] [19-21,31,32,36,38,39,50,51].

3.2 Safety outcome

In total, adverse events were reported in 10/29 studies [19,20,30,31,35,36,41,42,48,49] including 105 of 565 perianal CD patients [19%; 95% CI: 16%–22%]. Anaemia and allergic reactions were the most frequent adverse effects (Table 4). Among serious adverse event, leucopenia, varicella infection and anaphylactic shock were the commonest. Three studies [19,35,45] recorded discontinuation rate due to adverse effects which occurred in 9% of patients [95% CI: 5–16 %], with serious infections being the commonest reason cited for treatment withdrawal.

4. EXPERT OPINION

This is the first review that systematically and critically examines all relevant human studies to assess the effectiveness and safety profile of anti-TNF- α agents in the induction and maintenance treatment of perianal fistulising CD in the paediatric population. Our findings show that the majority of paediatric data on the efficacy and safety of anti-TNF agents for perianal fistulising CD treatment are of a descriptive nature; 25 of 29 identified studies are either observation, case series or case reports [8,19-21,32-52].

IFX therapy was assessed in two RCTs, which included 38 paediatric perianal CD patients [28,29]. These studies were not placebo-controlled due to ethical concerns. The results of our review reflect the practical situation in paediatric settings where treatment is often based on expert opinion, and whose assumed benefits are extrapolated from studies conducted in adults and from descriptive studies rather than prospective controlled trials.

Moreover, from the perspective of ethical considerations, research in paediatric settings should not only be ethically feasible but also essential to good clinical practice. This is particularly challenging in children with perianal CD because the fistulising disease is associated with high morbidity and significant impairment in patients' quality of life and often requires continued surgical intervention.

Moreover, since the approval of both IFX and ADA in paediatric CD, no prospective trial has been published that compares the effectiveness of anti-TNF- α antibodies with alternative treatments. An ongoing trial compares the efficacy of remission induction with IFX, prednisolone or exclusive enteral nutrition (EEN) in newly diagnosed paediatric CD; unfortunately, patients with the active perianal disease were excluded [53].

Overall, the pooled remission rates following the different follow-up time points between 2 weeks and 34 months of therapy occur in approximately 60% of patients, whereas in 55% a clinical response was observed. When the effect of anti-TNF- α therapy was evaluated over time, the percentage of patients in clinical remission increased at the 6th, 8th and 10th week, compared to the 12th week (Table 3).

Whereas remission was maintained in 41% after 12 months [19,32,33,36,43] and in 13% of patients after 2 years of treatment [8,45]. infants with severe fistulising perianal CD seems to achieve remission with complete fistula closure during the induction therapy with anti-TNF- α agents [48,49,51]. This evidence, however, has to be considered of low-quality as it emerged only from sporadic clinical case reports. Although this treatment showed promising results in the short-term, long-term studies are thus needed to confirm these initial findings (even after stopping the corticosteroid or immunosuppressant treatment).

Overall, pooled perianal fistula closure rate after anti-TNF- α treatment was 56% [8,20,28-30,33,34,36,37,39,40,47]. The variation in these remission rates largely depends on patient or treatment factors. Some studies on adult patients reported radiographic evidence of persistent fistula tracts in patients who had a clinical response to anti-TNF- α therapy, suggesting that the therapy suppresses the inflammatory response and fistula drainage but may not eradicate the epithelialized tracts [54]. Any recurrent fistula or perianal abscesses may be explained by the closure of the external orifice.

Eleven percent of patients were not responsive at all to treatment [8,19,20,29,31,33,35,36,48,51], while 19% of patients had a secondary loss of response [19,20,33-36,39,48]. The causes of the significant reduction in remission rate and clinical response after one or two years of follow-up are not clearly established. Patients' characteristics can have a high impact on drug effectiveness. A recent meta-analysis revealed that the presence of a perianal lesion and younger age at CD onset were relative risk factors for loss of response to anti-TNF- α therapy in adults [55]. Moreover, the response that an individual patient will have to a specific anti-TNF- α and dose is difficult to predict. The reason for primary non-response or secondary loss of response is multifactorial and probably related to the metabolism of the drug and to the development of anti-drug antibody levels [56]. Antibody levels are found to be higher in primary or secondary non-responders when compared to responders, whereas concomitant therapy with immunomodulators seems to reduce the risk of anti-drug antibody development [57].

Descriptive reports are a useful source of evidence on safety as they have the distinct advantage of examining the long-term effects and assessing for rare adverse events [58]. Safety issues of long-term use of anti-TNF therapies represent a significant concern. In the present review, of the 29 studies included, only 10 assessed adverse events as an outcome, for a total of 105 patients analysed, with acute infusion reactions being the most prevalent [19,20,30,31,35,36,41,42,48,49]. A limited number of studies reported safety data related to perianal CD patients. Although serious adverse events leading to drug suspension seem relatively low (serious infection was the most frequent cause of drug withdrawal), the overall rate of serious adverse events seems considerably higher: it appears in six out of ten children.

Most of the serious adverse events are infectious. This may be explained by the fact that a large number of patients included in our study had moderate to severe disease activity and were on a combination of immunosuppressive medications which is a widely recognised risk factor for opportunistic infections in CD patients [59-61]. Furthermore, potential selection bias may overestimate this incidence rate. There is also some evidence that anti-TNF- α agents are associated with a potential risk of malignancies; however, in the present review, no case of malignancy was identified. Still, a systematic review by Dulai *et al* [62] reported 2 cases of lymphoma among paediatric patients with IBD who received IFX with thiopurine. However, new evidence suggests that increased risk of malignancies is not linked to anti-TNF- α treatment in itself [63].

The impact of an anti-TNF- α therapy on the natural history of fistulising CD is not well established, but the combined use of the different medical strategies available at the moment (anti-TNF- α and immunomodulators) together with surgical evaluation and eventual placement of a non-cutting seton before initiating biologic therapy seem to improve clinical results and quality of life in these patients, reducing the use of corticosteroids and the adverse events related to their use. The benefits of corticosteroids seem higher when given earlier in the course of the disease. However, it remains difficult to predict responsiveness to these therapeutic options, so further studies are needed to determine biomarkers that can predict response to therapies; this should help to establish the more appropriate and personalised management of fistulising disease in paediatric patients [56].

Limitations and Strengths

This review has several limitations: (1) most of the studies available on this subgroup of CD patients are of lowest level of evidence (retrospective or prospective observational studies, case series or case report) that are prone to several biases; (2) questionable methodological quality of most of the studies; (3) most of the studies include patients with both systemic and perianal disease, from which the outcomes of interest have been extrapolated by means of additional analyses; (4) we included only those studies that report at-least one outcome for perianal CD patients in our analysis, leading to potential selection bias; (5) in primary studies, there was considerable heterogeneity about treatment protocols and the definition of the outcomes examined (response and remission criteria); some authors used PCDAI for outcome measure, that globally assesses the severity of CD in paediatric patients and not only perianal disease whereas other authors instead consider fistula closure or reduction/cessation of fistula drainage as outcome criteria; these facts additionally weaken the applicability of this systematic review; (6) finally, endoscopic endpoints or biochemical outcomes, were neither assessed nor discussed.

Despite these limitations, this is the first systematic review examining all available evidence on anti-TNF- α treatment in paediatric perianal CD. We used a comprehensive search strategy to identify relevant studies,

complete reporting and analysis of the data. We attempted to identify all relevant studies regardless of language or publication status. One of the greatest strengths of this review is the important addition to existing literature regarding paediatric perianal CD because no review has been addressed the present issue on this subgroup so far.

5. CONCLUSION

With all the mentioned limitations the findings of the present review imply that nearly three-fifths children with perianal CD achieved remission with anti-TNF- α treatment and in approximately 40% remission was maintained after 12 months, with a low discontinuation rate due to serious adverse events. Complete fistula closure was reported in more than half of the patients. There is still a need for more robust evidence adequately assessing the efficacy and safety of anti-TNF- α therapy in paediatric perianal CD, as well as in comparison with other therapies.

FUNDING

None

DISCLOSURE OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article

REFERENCES

1. Schwartz DA, Loftus EV, Jr., Tremaine WJ, et al. The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. *Gastroenterology*. 2002 Apr;122(4):875-80.
2. Beaugerie L, Seksik P, Nion-Larmurier I, et al. Predictors of Crohn's disease. *Gastroenterology*. 2006 Mar;130(3):650-6.
3. Loly C, Belaiche J, Louis E. Predictors of severe Crohn's disease. *Scand J Gastroenterol*. 2008 Aug;43(8):948-54.
4. Beaugerie L, Sokol H. Clinical, serological and genetic predictors of inflammatory bowel disease course. *World J Gastroenterol*. 2012 Aug 7;18(29):3806-13.
5. Lionetti HJ, Oh SH, Kim DY, et al. Clinical Characteristics and Long-Term Outcomes of Paediatric Crohn's Disease: A Single-Centre Experience. *J Crohns Colitis*. 2017 Feb;11(2):157-164.
6. Assa A, Amitai M, Greer ML, et al. Perianal Pediatric Crohn Disease Is Associated With a Distinct Phenotype and Greater Inflammatory Burden. *J Pediatr Gastroenterol Nutr*. 2017 Sep;65(3):293-298.
7. Panes J, Rimola J. Perianal fistulizing Crohn's disease: pathogenesis, diagnosis and therapy. *Nat Rev Gastroenterol Hepatol*. 2017 Nov;14(11):652-664.
8. Keljo DJ, Markowitz J, Langton C, et al. Course and treatment of perianal disease in children newly diagnosed with Crohn's disease. *Inflamm Bowel Dis*. 2009 Mar;15(3):383-7.
9. Eglinton TW, Roberts R, Pearson J, et al. Clinical and genetic risk factors for perianal Crohn's disease in a population-based cohort. *Am J Gastroenterol*. 2012 Apr;107(4):589-96.
10. de Zoeten EF, Pasternak BA, Mattei P, et al. Diagnosis and treatment of perianal Crohn disease: NASPGHAN clinical report and consensus statement. *J Pediatr Gastroenterol Nutr*. 2013 Sep;57(3):401-12.
11. Zwintscher NP, Shah PM, Argawal A, et al. The impact of perianal disease in young patients with inflammatory bowel disease. *Int J Colorectal Dis*. 2015 Sep;30(9):1275-9.
12. Mahadev S, Young JM, Selby W, et al. Self-reported depressive symptoms and suicidal feelings in perianal Crohn's disease. *Colorectal Dis*. 2012 Mar;14(3):331-5.
13. Greenley RN, Hommel KA, Nebel J, et al. A meta-analytic review of the psychosocial adjustment of youth with inflammatory bowel disease. *J Pediatr Psychol*. 2010 Sep;35(8):857-69.
14. Mackner LM, Crandall WV. Psychological factors affecting pediatric inflammatory bowel disease. *Curr Opin Pediatr*. 2007 Oct;19(5):548-52.

15. Mackner LM, Crandall WV. Long-term psychosocial outcomes reported by children and adolescents with inflammatory bowel disease. *Am J Gastroenterol*. 2005 Jun;100(6):1386-92.

16. Mackner LM, Crandall WV, Szigethy EM. Psychosocial functioning in pediatric inflammatory bowel disease. *Inflamm Bowel Dis*. 2006 Mar;12(3):239-44.

17. Hyams J, Crandall W, Kugathasan S, et al. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. *Gastroenterology*. 2007 Mar;132(3):863-73; quiz 1165-6.

18. Hyams JS, Griffiths A, Markowitz J, et al. Safety and efficacy of adalimumab for moderate to severe Crohn's disease in children. *Gastroenterology*. 2012 Aug;143(2):365-74.e2.

19. Dupont-Lucas C, Dabadie A, Alberti C, et al. Predictors of response to infliximab in paediatric perianal Crohn's disease. *Aliment Pharmacol Ther*. 2014 Oct;40(8):917-29.

20. de Ridder L, Escher JC, Bouquet J, et al. Infliximab therapy in 30 patients with refractory pediatric crohn disease with and without fistulas in The Netherlands. *J Pediatr Gastroenterol Nutr*. 2004 Jul;39(1):46-52.

21. Seemann, N. M., King, S. K., Elkadri, A., et al. The operative management of children with complex perianal Crohn's disease. *Journal of pediatric surgery* 2016; 51(12):1993-1997.

22. Prospero International Prospective Register of Systematic Reviews. Available from: <https://www.crd.york.ac.uk/PROSPERO/> Accessed 2 October 2018.

23. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009 Aug 18;151(4):264-9, w64.

24. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000 Apr 19;283(15):2008-12.

25. GA Wells, B Shea, D O'Connell, J Peterson, V Welch, M Losos et al. (Accessed 2 Feb 2019) The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

26. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998 Jun;52(6):377-84.

27. NIH Study Quality Assessment Tools. Available from <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>

28. Ruemmele FM, Lachaux A, Cezard JP, et al. Efficacy of infliximab in pediatric Crohn's disease: a randomized multicenter open-label trial comparing scheduled to on demand maintenance therapy. *Inflamm Bowel Dis*. 2009 Mar;15(3):388-94.

- 1
2
3 29. Kierkus J, Iwanczak B, Wegner A, et al. Monotherapy with infliximab versus combination therapy
4 in the maintenance of clinical remission in children with moderate to severe Crohn disease. *J*
5 *Pediatr Gastroenterol Nutr.* 2015 May;60(5):580-5.
6
7 30. Ruemmele FM, Rosh J, Faubion WA, et al. Efficacy of Adalimumab for Treatment of Perianal
8 Fistula in Children with Moderately to Severely Active Crohn's Disease: Results from IMaGNE 1
9 and IMaGNE 2. *J Crohns Colitis.* 2018 Nov 9;12(10):1249-1254.
10
11 31. Crandall W, Hyams J, Kugathasan S, et al. Infliximab therapy in children with concurrent
12 perianal Crohn disease: observations from REACH. *J Pediatr Gastroenterol Nutr.* 2009
13 Aug;49(2):183-90.
14
15 32. Viola F, Civitelli F, Di Nardo G, et al. Efficacy of adalimumab in moderate-to-severe pediatric
16 Crohn's disease. *Am J Gastroenterol.* 2009 Oct;104(10):2566-71.
17
18 33. Kim MJ, Lee JS, Lee JH et al. Infliximab therapy in children with Crohn's disease: a one-year
19 evaluation of efficacy comparing 'top-down' and 'step-up' strategies. *Acta Paediatrica* 2011, 100:
20 451-455.
21
22 34. Lionetti P, Bronzini F, Salvestrini C, et al. Response to infliximab is related to disease duration in
23 paediatric Crohn's disease. *Aliment Pharmacol Ther.* 2003 Aug 15;18(4):425-31.
24
25 35. Hukkinen M, Pakarinen MP, Piekala M, et al. Treatment of complex perianal fistulas with seton
26 and infliximab in adolescents with Crohn's disease. *J Crohns Colitis.* 2014 Aug;8(8):756-62.
27
28 36. Iwanczak BM, Ryzko J, Jankowski P, et al. Induction and Maintenance Infliximab Therapy for
29 the Treatment of Crohn's Disease with Perianal Fistulas in Children: Retrospective, Multicenter
30 Study. *Adv Clin Exp Med.* 2016 May-Jun;25(3):523-30.
31
32 37. Cezard JP, Nouaili N, Talbotec C, et al. A prospective study of the efficacy and tolerance of a
33 chimeric antibody to tumor necrosis factors (remicade) in severe pediatric crohn disease. *J Pediatr*
34 *Gastroenterol Nutr.* 2003 May;36(5):632-6.
35
36 38. Hyams JS, Markowitz J, Wyllie R. Use of infliximab in the treatment of Crohn's disease in
37 children and adolescents. *J Pediatr.* 2000 Aug;137(2):192-6.
38
39 39. Sinitsky DM, Lemberg DA, Leach ST, et al. Infliximab improves inflammation and
40 anthropometric measures in pediatric Crohn's disease. *J Gastroenterol Hepatol.* 2010
41 Apr;25(4):810-6.
42
43 40. Sykora J, Huml M, Siala K, et al. Fistulizing anorectal Crohn disease in a child with severe HLA-
44 B27-associated enthesitis-related arthritis. *J Pediatr Gastroenterol Nutr.* 2009 Apr;48(4):491-4.
45
46 41. Akman SA, Arikan C, Sozen G, et al. Infliximab treatment of pediatric refractory Crohn's
47 disease: a case report. *Turk J Gastroenterol.* 2006 Jun;17(2):133-6.
48
49
50
51
52
53
54
55
56
57
58
59
60

42. Russell RK, Wilson ML, Loganathan S, et al. A British Society of Paediatric Gastroenterology, Hepatology and Nutrition survey of the effectiveness and safety of adalimumab in children with inflammatory bowel disease. *Aliment Pharmacol Ther.* 2011 Apr;33(8):946-53.

43. Munoz Urribarri A, Sylvester F, Delgado Godos A, et al. [Management of colonic Crohn disease with no response to corticoids and sulfazalazine: first case of pediatric use of infliximab in Peru]. *Rev Gastroenterol Peru.* 2013 Apr-Jun;33(2):162-6.

44. Barabino A, Castellano E, Gandullia P, et al. A girl with severe fistulizing Crohn's disease. *Dig Liver Dis.* 2000 Dec;32(9):792-4.

45. Rosenbach Y, Hartman C, Shapiro R, et al. Adalimumab treatment in children with refractory Crohn's disease. *Dig Dis Sci.* 2010 Mar;55(3):747-53.

46. Singer AA, Gadepalli SK, Eder SJ, et al. Fistulizing Crohn's Disease Presenting After Surgery on a Perianal Lesion. *Pediatrics.* 2016 Mar;137(3): e20152878.

47. Serrano MS, Schmidt-Sommerfeld E, Kilbaugh TJ, et al. Use of infliximab in pediatric patients with inflammatory bowel disease. *Ann Pharmacother.* 2001 Jul-Aug;35(7-8):823-8.

48. Uslu N, Usta Y, Saltik-Temizel IN, et al. Ineffectiveness of infliximab therapy in severe infantile Crohn's disease. *Journal of Crohn's and Colitis.* 2010;4(1):106-109.

49. Saadah OI. Anti-TNFalpha antibody infliximab treatment for an infant with fistulising Crohn's disease. *BMJ Case Rep.* 2010;2010.

50. Gouldthorpe O, Catto-Smith AG, Alex G, et al. Loss of response to long-term infliximab therapy in children with Crohn's disease. *Pharmaceuticals (Basel).* 2013 Oct 16;6(10):1322-34.

51. Eradi B, Sandhu BK, Spray C, et al. Highly destructive perianal crohn disease. *J Pediatr Gastroenterol Nutr.* 2005 Nov;41(5):667-9.

52. Bicette R, Tenjarla G, Kugathan S, et al. A 14-Year-Old Girl with Recurrent Vulvar Abscess. *Journal of Pediatric and Adolescent Gynecology.* 2014 2014/08/01;27(4): e83-e86.

53. Cozijnsen MA, van Pieterse M, Samsom JN, et al. Top-down Infliximab Study in Kids with Crohn's disease (TISKids): an international multicentre randomised controlled trial. *BMJ Open Gastroenterol.* 2016;3(1):e000123.

54. Sands BE, Blank MA, Diamond RH, et al. Maintenance infliximab does not result in increased abscess development in fistulizing Crohn's disease: results from the ACCENT II study. *Aliment Pharmacol Ther.* 2006 Apr 15;23(8):1127-36.

55. Zhang QW, Shen J, Zheng Q, et al. Loss of response to scheduled infliximab therapy for Crohn's disease in adults: A systematic review and meta-analysis. *J Dig Dis.* 2019 Feb;20(2):65-72.

56. Adegbola SO, Sahnun K, Warusavitarne J, et al. Anti-TNF Therapy in Crohn's Disease. *International journal of molecular sciences.* 2018;19(8):2244.

- 1
2
3 57. Mahadevan U, Cucchiara S, Hyams JS, et al. The London Position Statement of the World
4 Congress of Gastroenterology on Biological Therapy for IBD with the European Crohn's and
5 Colitis Organisation: pregnancy and pediatrics. *Am J Gastroenterol*. 2011 Feb;106(2):214-23;
6 quiz 224.
7
8
9 58. Chou R, Helfand M. Challenges in Systematic Reviews That Assess Treatment Harms. *Annals of*
10 *Internal Medicine*. 2005;142(12_Part_2):1090-1099.
11
12 59. Veereman-Wauters G, de Ridder L, Veres G, et al. Risk of infection and prevention in pediatric
13 patients with IBD: ESPGHAN IBD Porto Group commentary. *J Pediatr Gastroenterol Nutr*. 2012
14 Jun;54(6):830-7.
15
16 60. De Ridder L, Turner D, Wilson DC, et al. Malignancy and mortality in pediatric patients with
17 inflammatory bowel disease: a multinational study from the porto pediatric IBD group. *Inflamm*
18 *Bowel Dis*. 2014 Feb;20(2):291-300.
19
20 61. Navas-Lopez VM, Blasco-Alonso J, Giron-Fernandez-Crehuet F, et al. Efficacy and safety of
21 adalimumab in the treatment of Crohn s disease in children. *Rev Esp Enferm Dig*. 2013 Nov-
22 Dec;105(10):579-84.
23
24 62. Dulai PS, Thompson KD, Blunt HB, et al. Risks of serious infection or lymphoma with anti-
25 tumor necrosis factor therapy for pediatric inflammatory bowel disease: a systematic review. *Clin*
26 *Gastroenterol Hepatol*. 2014 Sep;12(9):1443-51; quiz e88-9.
27
28 63. Cozijnsen MA, Samsom JN, de Ridder L. Anti-Tumour Necrosis Factor Therapy for Paediatric
29 Crohn's Disease: Improved Benefits Through Treatment Optimisation, Deeper Understanding of
30 Its Risks, and Reduced Costs due to Biosimilar Availability. *Paediatr Drugs*. 2018 Feb;20(1):19-
31 28.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Identification

Records identified through database Searching
(Pubmed n=479; Embase n=57; Cochrane n=185; Clinical
trials.gov n=5)
n=726

Duplicates(n=18)

Records after deduplicates removed (n=708)

Screening

Records screened (n=708)

Records excluded after title and abstract screening (n=646)

Eligibility

Full-text articles assessed for eligibility (n=62)

Excluded after full text examination (n=33)
Reason studies that did not report outcomes of interest (n=29)
Reason language (n=4)

Included

Studies included in qualitative synthesis
N=Studies included in metanalysis
N=29

Table 1. Description of the studies included in the qualitative analysis.

STUDY CHARACTERISTICS		ANTI-TNF THERAPY			OUTCOME	
First Author, year [Ref]	Design	Anti-TNF drug	Induction	Maintenance	Response criteria	Remission criteria
Akman, S. A., et al. (2006) [41]	Case report	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	-	-
Barabino, A., et al. (2000) [44]	Case report	IFX	5 mg/kg/d IV at 0 W (single dose)	-	-	PCDAI \leq 10
Bicette, R., et al. (2014) [52]	case report	IFX	NR	NR	NR	NR
Cezard et al. (2003) [42]	Prospective Observational	IFX	5 mg/kg/d IV at 0, 2, 6 W (one pt single dose)	NR	NR	HBI \leq 4 and the decrease or tapering of steroids
Crandall, W., et al. (2009) [31]	Post hoc analysis of RCT	IFX	5 mg/kg/d IV at 0, 2, 6 W	patients in clinical response were randomized to receive 5 mg/kg either q8w beginning at week 14, or q12w beginning at week 18	partial response: PCDAI < 5	Complete response: PCDAI = 0
De Ridder, L., et al. (2004) [20]	Retrospective cross-sectional	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	PCDAI \leq 10 or a decline of 20 points, fistula closure or cessation of drainage maintained for >4 weeks by physical examination.	modified PCDAI=10
Dupont-Lucas, C., et al. (2014) [19]	retrospective cohort study	IFX	5 mg/kg/d IV at 0, 2, 6 W	NR	Partial response: -50% or more from baseline in the number of draining fistulas and/or partial healing of ulcers	HBI < 5 ; Fistula drainage assessment; complete closure of all fistulas and healing of all ulcers.

Eradi, B., et al. (2005) [51]	case report	IFX	NR	NR	NR	NR
Gouldthorpe, O., et al. (2013) [50]	retrospective case series	IFX	NR	5 mg/kg q8w	PGA	modified PCDAI ≤ 10
Hukkinen, M., et al. (2014) [35]	retrospective cross-sectional	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	Partial response: Reduction in the size, number, or drainage of the fistulas	Complete response: cessation of drainage for at least 4 weeks recurrence
Hyams, J. S., et al. (2000) [38]	Cross-sectional	IFX	5 mg/kg/d IV at 0 W [n = 4]; 5 mg/kg/d IV at 0, 4 W[n=9]; 5 mg/kg/d IV at 0, 4, 8 W[n=4]; 5 mg/kg/d IV at 0, 2, 6 W[n=2]	-	PGA; PCDAI (cutt-off NR)	NR
Iwanczak, B. M., et al. (2016) [36]	retrospective observational	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	PCDAI (cutt-off NR)	NR
Keljo, D. J., et al. (2009) [8]	prospective, multicenter observational	IFX	NR	NR	NR	PCDAI=0 within 12 Months
Kierkuś, J., et al. (2015) [29]	RCT	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	PCDAI ≤ 30 or decrease of 15 points	PCDAI ≤ 10 points after 1 year
Kim, M. J., et al. (2011) [33]	retrospective observational	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w (patients in clinical response)	NR	PCDAI ≤ 10 points after 1 year
Lionetti, P., et al. (2003) [34]	retro cohort study	IFX	5 mg/kg/d IV at 0 W. (Decisions for re-infusion were made on the physician's judgement)	NR	minimal response: reduction $< 50\%$ or more in fistula diametre, discharge, pain and restriction of activity ; Partial response: reduction $> 50\%$ or more in fistula diametrer, discharge, pain and restriction of activity ;	PCDAI ≤ 15
Munoz Urribarri, A., et al. (2013) [43]	case report	IFX	5 mg/kg/d IV	5 mg/kg q8w. After 8 mo: 5 mg/kg q4w	NR	NR

Rosenbach, Y., et al. (2010) [45]	case series	ADA	first dose: 160mg/1.73 m ³ ; 2nd dose: 80mg /1.73 m ³	40/1.73 m ³ EOW	Partial response: HBI <8	HBI <4
Ruemmele, F. M., et al. (2009) [28]	RCT	IFX	5 mg/kg/d IV at 0, 2, 6 W	scheduled: 5 mg/kg q8w; on demand: repeated if relapse	NAAP	HBI<5
Ruemmele, F. M., et al. (2018) [30]	post-hoc analysis RCT	ADA	> 40 kg: 160/80mg; < 40 kg: 80/40mg	High-dose group: >40 kg = 40mg, < 40 kg = 20mg ; Low-dose group: > 40 kg = 20mg, < 40 kg = 10mg	PCDAI decrease > 15 points from baseline	PCDAI ≤ 10 ; fistula closure or a decrease in number by ≥50%, respectively, for at least two consecutive visits.
Russell, R. K., et al. (2011) [42]	case series	ADA	160/80 mg [n = 3]; 80/40 mg [n = 41]; 24mg/m ² [n = 16]; Other doses [n = 10];	40mg [n = 61] 80mg [n = 1] 24mg/m ² [n = 3] Other doses [n = 3]	PCDAI decrease > 12.5 points and/or PGA	PCDAI ≤ 10 or PGA
Saadah, O. I. (2010) [49]	case report	IFX	5 mg/kg/d IV at 0, 2, 6 W	-	NR	NR
Seemann, N. M., et al. (2016) [21]	retrospective chart review	IFX/ADA	NR	NR	NR	NR
Serrano, M. S., et al. (2001) [47]	case series	IFX	5 mg/kg/d IV at 0 w[n=4]; 5 mg/kg/d IV at 0,2 W [n=1]; 5 mg/kg/d IV at 0,2 6 W[n=1]	NR	PGA, patient reported outcomes	NR
Singer, A. A. M., et al. (2016) [46]	case series	IFX	NR	NR	PGA	PGA
Sinitsky, D. M., et al. (2010) [39]	case series	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	PCDAI decrease ≥ 15 points and/or PGA	PCDAI ≤ 15

Sýkora, J., et al. (2009) [40]	case report	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w	NR	NR
Uslu, N., et al. (2010) [48]	Case report	IFX	5 mg/kg/d IV at 0, 2, 6 W	5 mg/kg q8w (1 pt)	NR	NR
Viola, F., et al. (2009) [32]	Prospective study	ADA	160/80mg [n = 13] 120/80mg [n = 2] 80/40mg [n = 8]	80 [n = 15] 40 [n = 8]	PCDAI decrease \geq 50% from baseline value	PCDAI \leq 10

ADA = adalimumab; HBI = Harvey Bradshaw Index; IFX = infliximab; IV = intravenously; NR = not reported; PCDAI= paediatric Crohn’s disease activity index; PGA= physician global assessment; pt = patient; q8w = every eight weeks; RCT = randomised control trial; W = weeks;

Table 2. Description of the cohorts of patients with PCD exposed to anti-TNF.

Study [Ref]	Total number of patients	N° of patients with PCD treated with anti-TNF (%)	Age (years[range]; Mean±SD)	Gender/Male gender (%)	Baseline severity (PCDAI mean±SD; HBI)	Proportion of PCD patients naïve to biological therapy	Concomitant medication	Refractory to drug therapy for PCD
Akman, S. A., et al. (2006) [41]	1	1	14	M	NR	1/1	AZA, CS	Yes
Barabino, A., et al. (2000) [44]	1	1	10	F	PCDAI=60	1/1	CS	Yes
Bicette, R., et al. (2014) [52]	1	1	14	F	NR	1/1	-	No
Cezard et al. (2003) [42]	21	16/21 (76)	*15±2	*52%	*HBI=8	16/16	CS, AZA, MTX= 10	No
Crandall, W., et al. (2009) [31]	112	23/112(20)	13.1±2.6	82.6%	PCDAI=44.9±10.1	23/23	6-MP, AZA = 21/23; MTX =3/23; 5-ASA= 12/23; CS = 10/23	No
De Ridder, L., et al. (2004) [20]	30	12/30(40)	14.9±2	50%	modified PCDAI=29.2 [15-45]	12/12	AZA=11/12; CS=8/12; MTX=2/12 (temporary switch)	Yes
Dupont-Lucas, C., et al. (2014) [19]	101	101/101(100)	12.8±3.3	62.4%	NR	101/101	5-ASA=18/101; AB=45/101; CS=43/101; AZA/6-MP=66/101; MTX=9/101;	No

							EN=21/101; PN=5/101	
Eradi, B., et al. (2005) [51]	3	1/3 (33)	3	M	NR	1/1	NR	Yes
Gouldthorpe, O., et al. (2013) [50]	71	28/71 (39)	*14.4 [3.95-20.1]	*67%	PCDAI>30	28/28	DND	No
Hukkinen M., et al. (2014) [35]	13	13/13(100)	14.4 [14.0-16.8]	84%	PCDAI=31.3±10.9	13/13	5-ASA=7/13; AB=13/13; CS=6/13 AZA=3/13	2/13
Hyams, J. S., et al. (2000) [38]	19	1/19 (5)	*14.4±3.0	*52%	PCDAI>30	1/1	CS=1/1; for other concomitant medication DND	DND
Iwanczak, B. M., et al. (2016) [36]	221	50/221 (23)	<10Y: 2/50 10-18Y:48/50	*32%	PCDAI=47±17	50/50	NR	Yes
Keljo, D. J., et al. (2009) [8]	276	97/276 (35);	11.8±2.6	59%	PCDAI>0	NR	NR	NR
Kierkuś, J., et al. (2015) [29]	99	25/99 (25)	*14.54±2.61	*62%	PCDAI>30	25/25	*AZA or MTX=84/84 (only for maintenance)	Yes
Kim, M. J., et al. (2011) [33]	29	18/29 (62)	*13.9±1.6	*82%	*PCDAI=40.7±11.7	18/18	step up strategy: CS=11/29, 5-ASA=17/29, AZA=29/29; top down: No concomitant tx	6/18
Lionetti, P., et al. (2003) [34]	22	15/22 (68)	*13±3.2	*45%	PCDAI=41.2±21	15/15	* AZA/6-MP= 12/22 CS=11/22 5-ASA= 19/22	15/22

							EN= 2/22 TPN= 1/22 MTX= 2/22 AB=13/22		
Munoz Urribarri, A., et al. (2013) [43]	1	1/1 (100)	9	F	NR	1/1	AZA	Yes	
Rosenbach, Y., et al. (2010) [45]	14	6/14 (43)	*13.9 [1.9-19.1]	*64%	HBI=11.5±1.64	2/6 [ADR ON IFX N=2; LOR ON IFX N=2]	EN=1/6; MTX=1/6	conventional therapy (n=6/6), IFX (n=2/6)	
Ruemmele, F. M., et al. (2009) [28]	40	13/40 (32)	*13.9±2.2	*55%	*HBI=7.6±3.5	13/13	NR	Yes	
Ruemmele, F. M., et al. (2018) [30]	LOW DOSE	21	21/21 (100)	14.3±2.1	81%	PCDAI=42.5 [30.0- 60.0]	14/21	CS=10/21, AZA/6- MP=10/21, MTX=3/21, AB=6/21	21/21 [resistant or intolerant to IFX (n=7)]
	HIGH DOSE	15	15/15 (100)	14.5±2.3	46.7%	PCDAI=45.0 [32.5- 62.5]	9/15	CS=3/15, AZA/6- MP=9/15, MTX=3/15, AB=0/15,	15/15 [resistant or intolerant to IFX (n=6)]
Russell, R. K., et al. (2011) [42]	72	29/72 (41) [‡]	*14.8 [6.1-17.8]	NR	PCDAI=37.5 [7.5- 65] [N=48]	DND	*Steroids:26/70 AZA/6MP: 27/70 MTX: 19/70	DND	
Saadah, O. I. (2010) [49]	1	1/1 (100)	9 mo	F	NR	1/1	NR	Yes	
Seemann, N. M., et al. (2016) [21]	57	48/57 (84)	*Simple PCD: 12.2 (1–17) ; *Complex PCD: 10.7 (5–14)	*Simple PCD: 69.8% *Complex PCD:	NR	NR	*Simple PCD: CS=14/43, AZA=30/43,	NR	

				42.9%			MTX=15/43, AB=1/43, 5-ASA=28/43 ; Complex PCD: CS=9/14, AZA=4/14, MTX=9/14, AB=14/14, 5-ASA=5/14	
Serrano, M. S., et al. (2001) [47]	16	6/16 (41)	16.2±3.7	87%	NR	NR	NR	NR
Singer, A. A. M., et al. (2016) [46]	8	7/8 (88)	11.8±2.9	63%	NR	7/7	CS=1/7; MTX=1/7; AB=6/7	No
Sinitsky, D. M., et al. (2010) [39]	16	8/16 (50)	*13.0±4.2	*81%	*PCDAI=31.3 [17.5-57.5]	8/8	*CS=7/16; AZA=14/16; MTX=2/16; 6-MP=1/16; Tacrolimus=1/16	Yes
Sýkora, J., et al. (2009) [40]	1	1/1 (100)	15	M	PCDAI=32.5	1/1	5-ASA, AB, MTX	No
Uslu, N., et al. (2010) [48]	12	2/2 (100)	case 1: 9 mo case2: 6 mo	case1: F case2: M	case1: PCDAI>30 case2: PCDAI=55	2/2	case 1: AZA,CS; case 2:AZA, CS	Yes
Viola, F., et al. (2009) [32]	23	4/23 (17)	*16.1 [9-20]	*52.17	*PCDAI=36±5.7	DND	*CS: 18/23 MTX: 2/23 AZA/6 MP: 11/23	Yes

AB= antibiotics; AZA= Azathioprine; CS=corticosteroids; DND= data not differentiable; EN=enteral nutrition; F=female; HBI= Harvey Bradshaw Index; IFX= infliximab; M= male; mo= months; MTX=methotrexate; NR=not reported; PCDAI=Pediatric Crohn Disease Activity index; PN=parenteral nutrition; 5-ASA =5-aminosalicylic acid; 6-MP= 6-mercaptopurine.

£ : information about perianal disease available only for one patient

*: data related to the total cohort of patients

For Peer Review Only

Table 3. Overview of the studies according to remission and response rates to anti-TNF in patients with PCD.

Remission	N° of studies	Overall [%; 95% CI]	RCT [%; 95% CI]	Observational studies [%; 95% CI]	Case series [%; 95% CI]	Case reports [%; 95% CI]
2 weeks	3	8/39 [21;10-37]	5/23[22;8-44]	2/15[13;2-42]	NA	1/1[100;6-100]
6 weeks	8	89/214[42;35-49]	13/23[57;35-76]	72/176[41;34-49]	3/14[21;6-51]	1/1[100;6-100]
8 weeks	3	19/32[59; 41-76]	NA	18/31[58;39-75]	NA	1/1[100;6-100]
10 weeks	3	39/61[64;51-76]	39/61[64;51-76]	NA	NA	NA
3 months	4	19/126[15;10-23]	NA	18/125[14;9-22]	NA	1/1[100;6-100]
4 to 6 months	4	13/57[23;13-36]	9/36[25;13-43]	2/12[17;3-49]	1/8[13;1-53]	1/1[100;6-100]
8 to 9 months	4	16/40[40;25-57]	13/23[57;35-76]	NA	3/13[23;6-54]	NA
12 months	5	71/174[41;34-49]	NA	70/173[41;33-48]	NA	1/1[100;6-100]
13 to 18 months	3	38/66[58;45-70]	35/59[59;46-72]	NA	3/7[43;12-80]	NA
24 months	2	13/103[13;7-21]	NA	11/97[11;6-20]	2/6[33;6-76]	NA
34 months	1	1/7[14;1-58]	NA	NA	1/7[14;1-58]	NA
Response	N° of studies	Overall [%; 95% CI]	RCT [%; 95% CI]	Observational studies [%; 95% CI]	Case series [%; 95% CI]	Case reports [%; 95% CI]
1 week	4	7/10[70; 35-92]	NA	1/1[100;6-100]	4/6[67;24-94]	2/3[67;13-98]
2 to 4 weeks	5	24/86[28;19-39]	4/23[17;6-40]	17/28[61;41-78]	3/35[9;2-24]	NA
6 weeks	6	107/147[73;65-80]	3/23[13;3-35]	98/114[86;78-92]	4/8[50;18-83]	2/2[100;20-100]
8 weeks	2	10/31[32;17-52]	NA	10/31[32;17-52]	NA	NA
10 weeks	3	51/76[67;55-77]	23/48[48;34-63]	NA	28/28[100; 85-100]	NA
7 to 9 months	3	4/41[10;3-24]	2/23[9;2-30]	1/12[8;0-40]	1/6[17;1-64]	NA
12 months	3	82/169[49;41-56]	NA	82/169[49;41-56]	NA	NA
13 months	1	21/36[58;41-74]	21/36[58;41-74]	NA	NA	NA
14-18 months	3	4/36[11;4-27]	1/23[4;0-24]	NA	3/13[23;6-54]	NA
Primary non-response	10	37/342[11;8-15]	5/48[10;4-24]	30/291[10;7-15]	NA	2/3[67;13-98]
Secondary loss of response	8	42/219[19;14-25]	NA	40/209[19;14-25]	1/8[13;1-53]	1/2[50;3-97]
Underwent surgery	10	91/276[33;28-39]	1/23[4;0-24]	82/244[34;28-40]	1/8[13;1-53]	1/1[100;6-100]

during treatment						
------------------	--	--	--	--	--	--

For Peer Review Only

Table 4. Adverse events during anti-TNFα therapy.

Any adverse event [number of events (n) in 105/ 565 patients]	Serious adverse event [number of events (n) in 105/565 patients]	Withdrawals due to adverse events [number of patients 11/120]
Overall events (91)	Overall (65)	Overall (11)
Anaemia (9)	Leucopaenia (7)	Serious infections (8)
Allergic Reaction (7)	Varicella-Zona Virus Infection (7)	Acute infusion reaction (1)
Mild Allergic Skin Rash (3)	Death (2)	Fever (1)
Opportunistic Inf (3)	Anaphylactic Shock (2)	Shivering (1)
Transient Arthritis/Arthralgia (3)	Psoriasis (2)	
Flushing (2)	Serious Infection Not Specified (2)	
Hepatic Event (2)	Septicaemia (2)	
Liver Enzyme Elevation (2)	Acute Infusion Reaction [1]	
Neutropenia (2)	Conjunctivitis (1)	
Sinusitis (2)	Hepatitis B (1)	
Alopecia (1)	Herpes (1)	
Atopic Dermatitis (1)	Molluscum Contagiosum (1)	
Chest Pain (1)	Otitis (1)	
Delayed-Type Hypersensitivity (1)	Pneumonia (1)	
Eczema and Cheilitis (1)	Psoriasiformic Lesions (1)	
Fever (1)	Repeated Viral Diarrhoeas (1)	
Headaches (1)	Sacral Osteomyelitis (1)	
Nausea (1)	Sepsis (1)	
Perianal Abscess (1)	Septic Shock (1)	
Perianal Dermatitis (1)	Serious AE Not Specified (29)	
Tachycardia (1)		
Tachypnea (1)		
Other ADE not specified (28)		
Other infection not specified (16)		

* Reasons of withdrawals due to adverse events were specified in three studies: Rosenbach, Y., et al. (2010) [45]; Dupont-Lucas, C., et al. (2014) [19]; Hukkinen, M., et al. (2014) [35]

Case Series (NIH)

First author [ref.] Question (Yes/No)	Gouldthorpe et al. [50]	Rosenbach et al. [45]	Russell et al. [42]	Serrano et al. [47]	Singer et al. [46]	Simitsky et al. [39]
Was the study question or objective clearly stated?	Yes	Yes	Yes	Yes	Yes	Yes
Was the study population clearly and fully described, including a case definition?	Yes	Yes	Yes	Yes	Yes	Yes
Were the cases consecutive?	No	Yes	No	Yes	Yes	Yes
Were the subjects comparable?	Yes	Yes	Yes	Yes	Yes	Yes
Was the intervention clearly described?	Yes	Yes	Yes	Yes	No	Yes
Were the outcome measures clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	No	No	Yes	Yes
Was the length of follow-up adequate?	Yes	Yes	No	Yes	Yes	Yes
Were the statistical methods well-described?	Yes	Yes	Yes	Yes	No	Yes
Were the results well-described?	Yes	Yes	Yes	Yes	Yes	Yes
Rate	Good	Good	Fair	Good	Fair	Good

Case report (CARE checklist)

First author [Ref.]		Akman et al. [41]	Barabino et al. [44]	Bicette et al. [52]	Eradi et al. [51]	Munoz Urribarri et al. [43]	Saadah et al. [49]	Sýkora et al. [40]	Uslu et al. [48]
Title	The area of focus and “case report” should appear in the title	Yes	No	No	Yes	Yes	No	No	No
Key words	Two to five key words that identify topics in this case report	Yes	Yes	Yes	No	Yes	No	No	Yes
Abstract	Introduction – What is unique and why is it important?	Yes	No	Yes	Yes	Yes	No	No	Yes
	The patient’s main concerns and important clinical findings.	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
	The main diagnoses, interventions, and outcomes.	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
	Conclusion—What are one or more “take-away” lessons?	Yes	No	Yes	No	No	No	No	Yes
Introduction	Briefly summarize why this case is unique with medical literature references.	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Patient information	De-identified demographic and other patient information.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Main concerns and symptoms of the patient.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Medical, family, and psychosocial history including genetic information.	Yes	No	Yes	No	No	Yes	No	Yes
	Relevant past interventions and their outcomes.	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Clinical findings	Relevant physical examination (PE) and other clinical findings.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Timeline	Relevant data from this episode of care organized as a timeline (figure or table)	Yes	No	No	No	Yes	Yes	No	No
Diagnostic assessment	Diagnostic methods (PE, laboratory testing, imaging, surveys).	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Diagnostic challenges. (e.g financial, language/cultural)	No	No	Yes	No	No	No	Yes	No
	Diagnostic reasoning including differential diagnosis.	Yes	No	Yes	No	No	Yes	Yes	Yes
	Prognostic characteristics (e.g. staging) when applicable.	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Therapeutic intervention	Types of intervention (pharmacologic, surgical, preventive).	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Administration of intervention (dosage, strength, duration).	Yes	Yes	No	No	Yes	Yes	Yes	Yes
	Changes in the interventions with explanations.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Follow-ups and outcomes	Clinician and patient-assessed outcomes when appropriate.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Important follow-up diagnostic and other test results.	Yes	Yes	Yes	No	Yes	Yes	No	Yes
	Intervention adherence and tolerability (how was this assessed)?	Yes	Yes	No	No	Yes	No	No	No
	Adverse and unanticipated events.	Yes	Yes	No	No	Yes	Yes	No	Yes
Discussion	Strengths and limitations in the approach to this case.	No	Yes	Yes	No	Yes	Yes	Yes	Yes
	Discussion of the relevant medical literature.	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
	The rationale for your conclusions.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	The primary “take-away” lessons from this case report.	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Patient perspective	The patient can share their perspective on their case.	No	No	No	No	No	No	No	No
Informed consent	The patient should give informed consent.	No	No	No	No	No	Yes	No	No
Total score		25/30	21/30	23/30	14/30	23/30	23/30	17/30	23/30

Observational studies (NOS)

First author [Ref.]	Selection				Comparability		Outcome			Rate
	Representativeness of the exposed cohort (A,B stars)	Selection of the non-exposed cohort (A star)	Ascertainment of exposure (A,B star)	Demonstration that outcome of interest was not present at start of study (A star)	The study controls for age, sex and marital status (star)	Study controls for other factors (star)	Assessment of outcome (A,B star)	Was follow-up long enough for outcomes to occur (A star)	Adequacy of follow-up of cohorts (A,B star)	
Cezard et al.[37]	B star	C	A star	A star	star	star	B star	A star	A star	Good
De Ridder et al. [20]	B star	C	A star	A star	star	star	B star	A star	A star	Good
Dupont-Lucas et al [19]	A star	C	A star	A star	star	star	B star	A star	A star	Good
Hukkinen et al. [35]	A star	C	A star	A star	star	star	B star	A star	A star	Good
Hyams et al. [38]	B star	C	A star	A star	star	star	B star	B	B star	Good
Iwanczak et al. [36]	A star	C	A star	A star	star	-	B star	A star	A star	Good
Keljo et al. [8]	A star	C	A star	A star	star	-	B star	A star	D	Fair
Kim et al. [33]	B star	C	A star	A star	star	star	B star	A star	A star	Good
Lionetti et al. [34]	B star	C	A star	A star	star	star	B star	A star	A star	Good
Seemann et al. [21]	A star	C	A star	A star	star	star	B star	A star	C	Fair
Viola et al. [32]	B star	C	A star	A star	star	star	B star	A star	A star	Good

Downs and Black quality assessment: Crandall et al. [31]

Criteria	Scoring criteria	Score	Notes/justifications
Is the hypothesis/aim/objective of the study clearly described?	A point was given if the hypothesis aim or objective of the study was implicitly or explicitly indicated anywhere in the article. (No=0; Yes=1)	1	post-hoc analysis evaluated the effect of infliximab upon concurrent perianal Crohn disease(CD)in a subpopulation of 31 patients from REACH, a randomized trial of 112 children with moderately to severe active luminal CD.
Are the main outcomes to be measured clearly described in the "Introduction" or "Methods" section?	A point was given if the main outcomes to be measured were clearly described in the "Introduction" or "Methods" section. (No=0; Yes=1)	1	signs and symptoms of perianal disease were assessed at each visit using the PCDAI perirectal subscore. Patients with no symptoms or asymptomatic tags received a score of 0;those with "1-2 indolent fistula, scant drainage, no tenderness" received a score of 5; and those with [...]
Are the characteristics of the patients included in the study clearly described?	A point was given if the inclusion or exclusion criteria, or both, were indicated. (no=0; Yes=1)	1	eligible patients were 6 to 17 years of age, had CD of at least 3 months duration, and had a baseline PCDAI score greater than 30. Concomitant therapies included azathioprine, 6-mercaptopurine, methotrexate, aminosalicylates and oral corticosteroids. patients received previous treatment with infliximab or other agents targeting TNF-alpha were ineligible. patients with active luminal disease and concomitant perianal disease were included but those with isolated perianal disease were excluded.
Are the interventions of interest clearly described?	A point was given if the criteria for guideline adherence were described in detail. (No=0; Yes=1)	1	eligible patients received an induction regimen of infliximab 5mg/kg at week 0,2, and 6. At week 10 patients in clinical response were randomized to receive 5mg/kg either every 8 weeks (q8w) beginning at week 1 or every 12 weeks (q12w) beginning at week 18.
Are the distributions of principal confounders for each group of participants to be compared clearly described?	Two points were awarded if a study reported any possible confounders (e.g., sex ratios, age, comorbidities, and severity of injury) that might account for	1	See table 1

	differences between groups clearly in table format. One point was awarded if the study indicated that groups were matched for any such demographical variables or if potential confounders were mentioned in the text of the article but not clearly listed in table format. No points were awarded if the study did not report any confounders. (no=0; partially=1; Yes=2)		
Are the main findings of the study clearly described?	A point was awarded if quantitative data were reported for all of the main outcome measures indicated in the "Introduction" or "Methods" section. (0 = No; 1 = Yes)	1	See figure 2
Does the study provide estimates of the random variability in the data for the main outcomes?	A point was awarded if the interquartile range (for non-normally distributed data), standard error, standard deviation, or confidence intervals (for normally distributed data) were reported. If the distribution of the data was not described, we assumed that the estimates used were appropriate, and we answered "Yes" (1 point). (0 = No; 1 = Yes)	0	
Have all of the important adverse events that may be a consequence of the intervention been reported?	A point was awarded if any adverse events, unwanted side effects, or lack thereof were explicitly indicated from either adherence or failure to adhere to recommended guidelines. A point was not awarded if the study made no mention of the presence or absence of adverse events. (0 = No; 1 = Yes)	1	See "safety"
Have the characteristics of patients lost to follow-up been describe?	The authors of this tool indicated that this question should be answered "Yes" when clear reasons for loss to follow-up were described. For the purposes of this review, a point was awarded if a study explicitly reported the number and reason for patients	1	See Figure 1 and 2

	lost to follow-up. (0 = No; 1 = Yes)		
Have actual probability values been reported (e.g. 0.035 rather than < 0.05) for the main outcomes except where the probability value is less than 0.001?	A point was awarded if the exact P value was provided for both statistically significant and non-significant results for at least the main outcome measures. A point was not awarded if a study simply indicated that the results for the main outcome measures were not significant without providing the exact P value. (No=0; Yes=1)	0	
Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	A point was awarded if the study identified the source population for patients and described how the patients were selected. Patients were determined to be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample (only feasible where a list of all members of the relevant population exists). Where a study did not report the proportion of the source population from which the patients are derived, the question was answered as unable to determine. (no=0; Yes=1)	0	Post hoc analyses evaluated the effect of infliximab upon concurrent perianal Crohn disease (CD) in a subpopulation of 31 patients from REACH, a randomized trial of 112 children with moderately to severely active luminal CD.
Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	The proportion of patients included in the study were representative of the population. Those asked who agreed to participate or responded should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population. No point was awarded if the proportion of those asked who agreed to participate or responded was not stated. (No=0; Yes=1)	0	Post hoc analyses evaluated the effect of infliximab upon concurrent perianal Crohn disease (CD) in a subpopulation of 31 patients from REACH, a randomized trial of 112 children with moderately to severely active luminal CD.
Were the staff, places and facilities where	A point was awarded unless the study	1	

the patients were treated, representative of the treatment the majority of patients receive?	specifically stated that patients were treated by a therapist who received specialized training relative to guideline recommendations. (1 = Yes; 0 = No; 0 = Unable to determine)		
Was an attempt made to blind study subjects to the intervention they have received?	A point was awarded if the patients were not aware of, or would have no way of knowing (as in the case of retrospective studies), which intervention they received. The study was not awarded a point if it was prospective and failed to mention whether the patients had knowledge of whether they were assigned to the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Open-label
Was an attempt made to blind those measuring the main outcomes of the intervention?	A point was awarded if the study specifically stated that those assessing the outcome measures were unaware of (or would have no way of knowing) whether the patients were in the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Open-label
If any of the results of the study were based on "data dredging", was this made clear?	A point was awarded if no retrospective unplanned (at the outset of the study) subgroup analyses were reported. (1 = Yes; 0 = No; 0 = Unable to determine)	0	This study is limited by the observational nature of results that were not prespecified in the protocol,
In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	A point was awarded if all patients had same follow-up. (1 = Yes; 0 = No; 0 = Unable to determine)	1	See figure 2
Were the statistical tests used to assess the main outcomes appropriate?	If the distribution of the data (normal or not) was not described, it was assumed that the estimates used were appropriate, and a point was awarded. No point was awarded for studies that reported qualitative or	0	

	quantitative data without any form of statistical comparisons or if the statistical tests reported were not appropriate. (1 = Yes; 0 = No; 0 = Unable to determine)		
Was compliance with the intervention/s reliable?	If the authors in prospective studies reported non-adherence to physical therapy intervention or adherence could not be determined, the study was not awarded a point. In retrospective studies, data were collected only for those patients who completed their episode of care (adherence to physical therapy assumed), and a point was awarded. For studies where the effect of any non-adherence was likely to bias any association to the null, the study was not awarded point. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the main outcomes measures used accurate (valid and reliable)?	A point was awarded if the primary outcome measures were thought to be valid and reliable (e.g., number of physical therapy visits per chart report), regardless of whether reliability or validity was reported. A point was not awarded if at least one of the primary outcome measures in the study was not valid or reliable or if this information was not reported or could not be determined (i.e., a questionnaire without reported validity or reliability). (1 = Yes; 0 = No; 0 = Unable to determine)	1	Signs and symptoms of perianal disease were assessed at each visit using the PCDAI perirectal subscore.
Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	A point was awarded when participants from both adherence and non-adherence groups were recruited from the same population. Otherwise, a point was not awarded (e.g., a point was not awarded when all participants from the adherence	1	31 patients from REACH, a randomized trial of 112 children with moderately to severely active luminal CD.

	group received care at clinic A and all participants in the non-adherence group received care at clinic B, because they could have represented 2 distinct populations). (1 = Yes; 0 = No; 0 = Unable to determine)		
Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	A point was awarded when the study provided a specific time line for patient recruitment (prospective studies) or when data were collected between reported dates of patient care (retrospective studies). (1 = Yes; 0 = No; 0 = Unable to determine)	1	Patients were enrolled from February 24, 2003, to March 31, 2004 (from the original paper)
Were study subjects randomized to intervention groups?	A point for random allocation was awarded if random allocation of patients was stated in the “Method” section of the article. The precise method of randomization need not be specified. Quasi-randomization allocation procedures, such as allocation by bed availability, did not satisfy this criterion. For crossover study designs, a point was awarded when participants were randomly allocated in the order in which treatments were received. (1 = Yes; 0 = No; 0 = Unable to determine)	1	patients in clinical response were randomized to receive[.]
Was the randomized intervention assignment concealed from both patients and healthcare staff until recruitment was complete and irrevocable?	The study did not receive a point unless the participants were randomly allocated and the methods for ensuring random allocation were specified. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	A point was awarded unless the effect of the main confounders was not investigated or confounding was demonstrated, but no adjustment was made in the final analyses. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were losses of patients to follow-up taken into account?	point was awarded as long as the number of dropouts lost to follow-up accounted for	1	

	less than 10% of the initial number of total participants or a maximum of 5% from each group. The question was answered with “unable to determine” if the number of patients lost to follow-up were not reported, could not be deduced from the outcome data (i.e., initial and final sample sizes not indicated) or the study methodology would not infer such information. (1 = Yes; 0 = No; 0 = Unable to determine)		
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	A point is awarded if power was stated. (No=0; Yes=1)	0	
Total score		15/28	

Score	Rate
28-26	Excellent
25-20	Good
19-15	Fair
≤14	Poor

Downs and Black quality assessment: Kierkus et al. [29]

Criteria	Scoring criteria	Score	Notes/justifications
Is the hypothesis/aim/objective of the study clearly described?	A point was given if the hypothesis aim or objective of the study was implicitly or explicitly indicated anywhere in the article. (No=0; Yes=1)	1	The aim of the present study was to compare the efficacy and safety of 2 protocols of maintenance therapy with infliximab (IFX) and an immunomodulatory agent in pediatric patients with Crohn disease (CD): withdrawal of immunomodulators versus continuation of immunosuppressants.
Are the main outcomes to be measured clearly described in the "Introduction" or "Methods" section?	A point was given if the main outcomes to be measured were clearly described in the "Introduction" or "Methods" section. (No=0; Yes=1)	1	Patients were assessed at baseline and at weeks 10 and 54. Each visit included a physical examination and laboratory tests measuring the levels of inflammatory indicators. Also, body mass index (BMI) was calculated during each visit. Additionally, the clinical activity of the disease was assessed using PCDAI, and the endoscopic activity was scored using SES-CD at weeks 0, 10 and 54.
Are the characteristics of the patients included in the study clearly described?	A point was given if the inclusion or exclusion criteria, or both, were indicated. (no=0; Yes=1)	1	The Concomitant Immunomodulator for Maintenance Infliximab. Therapy study included 7- to 17-year-old patients in whom the diagnosis of CD was confirmed by endoscopy and biopsy. The differential diagnoses included infectious (eg, C difficile) and allergic colitis, which have been ruled out in our patients. All of the participants had a history of moderate to severely active CD, as defined by Pediatric Crohn's Disease Activity Index (PCDAI) values>30 points (22), and lacked or lost the response to previously given pharmacotherapy other than biological therapy. The endoscopic activity of CD was scored with the Simplified Endoscopic Activity Score for Crohn's Disease (SES-CD) at weeks 0, 10, and 54. SES-CD is a simple, reproducible, and easy-to-use endoscopic scoring system for patients with CD based on the following 4 variables, each scored between 0 and 3 points: ulcer size, ulcerated surface, affected surface, and stenosis determined in 5 ileocolonic segments (23). The presence of active fistula was not the exclusion criterion from the trial. Permitted concomitant

			treatments included all of the therapies approved by the doctor in charge of the patient taking part in the trial. Patients were ineligible if they had received any biological agent within 8 weeks before or during the trial.
Are the interventions of interest clearly described?	A point was given if the criteria for guideline adherence were described in detail. (No=0; Yes=1)	1	During the induction phase (weeks 0–10), the remission of CD was induced with 3 doses of IFX (5 mg/kg body weight [b.w.]) given at weeks 0, 2, and 6. Simultaneously, steroid therapy was withdrawn in patients who had received such a treatment previously. In accordance with the protocol, the dose of steroids was decreased gradually (5 mg/week), up to complete withdrawal in week 2. Reimplementation of steroids, however, was considered to be an option of rescue therapy. On week 10, the therapeutic responses were evaluated on the basis of PCDAI score and endoscopic examination. Patients with the clinical response (ie, decrease in PCDAI 15 points and 30 points) were centrally randomized into 1 of the following maintenance protocols: IFX with an immunomodulatory agent (azathioprine [AZA] 1.5–3 mg/kg b.w. per day or methotrexate 10–25 mg/week; group I) or IFX with the immunomodulatory agent discontinued after 26 weeks of treatment (group II). Randomization list was prepared for 100 patients in groups of 10. A total of 100 randomization envelopes were prepared with subsequent numbers. Central off-site randomization service by envelope system provided treatment allocation. Because both AZA and methotrexate have been introduced before the beginning of the study, their doses were stable and maintained during the trial. During the maintenance phase (weeks 10–54), the infusions of IFX (5 mg/kg b.w.) were given every 8 weeks, and the follow-up visits were scheduled at weeks 14, 22, 30, 38, and 46 for both groups. At week 26, the immunomodulatory agent was discontinued in patients from group II. The duration of concomitant therapy has been

			tempered by data from a study that showed that there may be no additional benefit with continuation of immunosuppressants beyond 6 months of combination therapy (21). The therapeutic response was evaluated at week 54 based on PCDAI score and endoscopic examination, and the patients were assessed for the presence of any potential serious AEs (SAE). The flowchart illustrating the course of the study is presented in Figure 1.
Are the distributions of principal confounders for each group of participants to be compared clearly described?	Two points were awarded if a study reported any possible confounders (e.g., sex ratios, age, comorbidities, and severity of injury) that might account for differences between groups clearly in table format. One point was awarded if the study indicated that groups were matched for any such demographical variables or if potential confounders were mentioned in the text of the article but not clearly listed in table format. No points were awarded if the study did not report any confounders. (no=0; partially=1; Yes=2)	1	The Mann-Whitney U test was used for the analysis of intergroup differences in the level of quantitative variables (BMI, hemoglobin concentration, platelet count, C-reactive protein, SES-CD score, and PCDAI score), whereas the intragroup differences in the levels of these variables were analyzed with the Wilcoxon signed-rank test.
Are the main findings of the study clearly described?	A point was awarded if quantitative data were reported for all of the main outcome measures indicated in the "Introduction" or "Methods" section. (0 = No; 1 = Yes)	1	See figure 2 e table 2
Does the study provide estimates of the random variability in the data for the main outcomes?	A point was awarded if the interquartile range (for non-normally distributed data), standard error, standard deviation, or confidence intervals (for normally distributed data) were reported. If the distribution of the data was not described, we assumed that the estimates used were appropriate, and we answered "Yes" (1 point). (0 = No; 1 = Yes)	1	See table 2
Have all of the important adverse events	A point was awarded if any adverse events,	1	See table 3

that may be a consequence of the intervention been reported?	unwanted side effects, or lack thereof were explicitly indicated from either adherence or failure to adhere to recommended guidelines. A point was not awarded if the study made no mention of the presence or absence of adverse events. (0 = No; 1 = Yes)		
Have the characteristics of patients lost to follow-up been describe?	The authors of this tool indicated that this question should be answered “Yes” when clear reasons for loss to follow-up were described. For the purposes of this review, a point was awarded if a study explicitly reported the number and reason for patients lost to follow-up. (0 = No; 1 = Yes)	0	
Have actual probability values been reported (e.g. 0.035 rather than < 0.05) for the main outcomes except where the probability value is less than 0.001?	A point was awarded if the exact P value was provided for both statistically significant and non-significant results for at least the main outcome measures. A point was not awarded if a study simply indicated that the results for the main outcome measures were not significant without providing the exact P value. (No=0; Yes=1)	1	See table 1 and 2. Mean PCDAI score at baseline was 48.10 points (median 51, interquartile range [IQR] 35–55) compared with 12.50 (median 10, IQR 5–17.5) after completing the induction phase at week 10; this difference proved statistically significant ($P<0.05$; Fig. 2A). Moreover, there was a significant decrease in SES-CD score; the mean scores at baseline and after the induction phase were 16.62 points (median 15, IQR 9–24) and 5.22 points (median 3, IQR 0–8), respectively ($P<0.05$; Fig. 2B).
Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	A point was awarded if the study identified the source population for patients and described how the patients were selected. Patients were determined to be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample (only feasible where a list of all members of the relevant population exists). Where a study did not report the proportion of the source population from which the patients are derived, the question was answered as	0	During screening phase, the results of laboratory tests (days 14–0) and endoscopic examination (up to 3 months before day 0) were obtained to enroll the eligible patients.

	unable to determine. (no=0; Yes=1)		
Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	The proportion of patients included in the study were representative of the population. Those asked who agreed to participate or responded should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population. No point was awarded if the proportion of those asked who agreed to participate or responded was not stated. (No=0; Yes=1)	1	See figure 1
Were the staff, places and facilities where the patients were treated, representative of the treatment the majority of patients receive?	A point was awarded unless the study specifically stated that patients were treated by a therapist who received specialized training relative to guideline recommendations. (1 = Yes; 0 = No; 0 = Unable to determine)	1	It involved 15 Polish pediatric gastroenterology centers (all of the academic tertiary institutions), and thus the hereby reported findings are likely generalizable for the other patient populations.
Was an attempt made to blind study subjects to the intervention they have received?	A point was awarded if the patients were not aware of, or would have no way of knowing (as in the case of retrospective studies), which intervention they received. The study was not awarded a point if it was prospective and failed to mention whether the patients had knowledge of whether they were assigned to the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	The study was designed as a multicenter randomized open label trial
Was an attempt made to blind those measuring the main outcomes of the intervention?	A point was awarded if the study specifically stated that those assessing the outcome measures were unaware of (or would have no way of knowing) whether the patients were in the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	The study was designed as a multicenter randomized open label trial

1 2 3 4 5 6 7 8 9 10 11	If any of the results of the study were based on "data dredging", was this made clear?	A point was awarded if no retrospective unplanned (at the outset of the study) subgroup analyses were reported. (1 = Yes; 0 = No; 0 = Unable to determine)	1	
12 13 14 15 16 17 18 19 20 21 22	In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	A point was awarded if all patients had same follow-up. (1 = Yes; 0 = No; 0 = Unable to determine)	1	See figure 1
23 24 25 26 27 28 29 30 31 32 33 34 35 36 37	Were the statistical tests used to assess the main outcomes appropriate?	If the distribution of the data (normal or not) was not described, it was assumed that the estimates used were appropriate, and a point was awarded. No point was awarded for studies that reported qualitative or quantitative data without any form of statistical comparisons or if the statistical tests reported were not appropriate. (1 = Yes; 0 = No; 0 = Unable to determine)	1	
38 39 40 41 42 43 44 45 46	Was compliance with the intervention/s reliable?	If the authors in prospective studies reported non-adherence to physical therapy intervention or adherence could not be determined, the study was not awarded a point. In retrospective studies, data were collected only for those patients who completed their episode of care (adherence to physical therapy assumed), and a point was awarded. For studies where the effect of any non-adherence was likely to bias any association to the null, the study was not awarded point. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
	Were the main outcomes measures used accurate (valid and reliable)?	A point was awarded if the primary outcome measures were thought to be valid and reliable (e.g., number of physical therapy visits per chart report), regardless	1	

	of whether reliability or validity was reported. A point was not awarded if at least one of the primary outcome measures in the study was not valid or reliable or if this information was not reported or could not be determined (i.e., a questionnaire without reported validity or reliability). (1 = Yes; 0 = No; 0 = Unable to determine)		
Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	A point was awarded when participants from both adherence and non-adherence groups were recruited from the same population. Otherwise, a point was not awarded (e.g., a point was not awarded when all participants from the adherence group received care at clinic A and all participants in the non-adherence group received care at clinic B, because they could have represented 2 distinct populations). (1 = Yes; 0 = No; 0 = Unable to determine)	1	It involved 15 Polish pediatric gastroenterology centers (all of the academic tertiary institutions), and thus the hereby reported findings are likely generalizable for the other patient populations
Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	A point was awarded when the study provided a specific time line for patient recruitment (prospective studies) or when data were collected between reported dates of patient care (retrospective studies). (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were study subjects randomized to intervention groups?	A point for random allocation was awarded if random allocation of patients was stated in the “Method” section of the article. The precise method of randomization need not be specified. Quasi-randomization allocation procedures, such as allocation by bed availability, did not satisfy this criterion. For crossover study designs, a point was awarded when participants were randomly allocated in the order in which	1	The study was designed as a multicenter randomized open label trial

	treatments were received. (1 = Yes; 0 = No; 0 = Unable to determine)		
Was the randomized intervention assignment concealed from both patients and healthcare staff until recruitment was complete and irrevocable?	The study did not receive a point unless the participants were randomly allocated and the methods for ensuring random allocation were specified. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	A point was awarded unless the effect of the main confounders was not investigated or confounding was demonstrated, but no adjustment was made in the final analyses. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were losses of patients to follow-up taken into account?	A point was awarded as long as the number of dropouts lost to follow-up accounted for less than 10% of the initial number of total participants or a maximum of 5% from each group. The question was answered with “unable to determine” if the number of patients lost to follow-up were not reported, could not be deduced from the outcome data (i.e., initial and final sample sizes not indicated) or the study methodology would not infer such information. (1 = Yes; 0 = No; 0 = Unable to determine)	1	
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	A point is awarded if power was stated. (No=0; Yes=1)	1	Assuming 0.8 statistical power
Total Score		19/28	

Score	Rate
28-26	Excellent
25-20	Good
19-15	Fair
≤14	Poor

For Peer Review Only

Downs and Black quality assessment: Ruemmele et al. (2009) [28]

Criteria	Scoring criteria	Score	Notes/justifications
Is the hypothesis/aim/objective of the study clearly described?	A point was given if the hypothesis aim or objective of the study was implicitly or explicitly indicated anywhere in the article. (No=0; Yes=1)	1	"study aimed to evaluate in a prospective manner the efficacy and safety of IFX as maintenance therapy of severe pediatric CD comparing scheduled and "on demand" treatment strategies."
Are the main outcomes to be measured clearly described in the "Introduction" or "Methods" section?	A point was given if the main outcomes to be measured were clearly described in the "Introduction" or "Methods" section. (No=0; Yes=1)	1	The endpoints of this study were for phase 1: rate of remission at W10, and for phase 2: rate of remission at W60, with the secondary criteria: number of relapses in group A versus group B and number of IFX infusions in group A versus group B.
Are the characteristics of the patients included in the study clearly described?	A point was given if the inclusion or exclusion criteria, or both, were indicated. (no=0; Yes=1)	1	The endpoints of this study were for phase 1: rate of remission at W10, and for phase 2: rate of remission at W60, with the secondary criteria: number of relapses in group A versus group B and number of IFX infusions in group A versus group B.
Are the interventions of interest clearly described?	A point was given if the criteria for guideline adherence were described in detail. (No=0; Yes=1)	1	The study was separated into 2 parts: phase 1 was considered the induction period based on 3 consecutive IFX infusions (5 mg/kg) at week (W)0, W2, and W6 allowing determination of IFX responders included in phase 2, the maintenance therapy. Only patients who responded to IFX with complete
Are the distributions of principal confounders for each group of participants to be compared clearly described?	Two points were awarded if a study reported any possible confounders (e.g., sex ratios, age, comorbidities, and severity of injury) that might account for differences between groups clearly in table format. One point was awarded if the study indicated that groups were matched for any such demographical variables or if potential confounders were mentioned in the text of the article but not clearly listed in table format. No points were awarded if the study did not report any confounders. (no=0; partially=1; Yes=2)	0	

Are the main findings of the study clearly described?	A point was awarded if quantitative data were reported for all of the main outcome measures indicated in the "Introduction" or "Methods" section. (0 = No; 1 = Yes)	1	See figure 1 and figure 2
Does the study provide estimates of the random variability in the data for the main outcomes?	A point was awarded if the interquartile range (for non-normally distributed data), standard error, standard deviation, or confidence intervals (for normally distributed data) were reported. If the distribution of the data was not described, we assumed that the estimates used were appropriate, and we answered "Yes" (0 = No; 1 = Yes)	0	
Have all of the important adverse events that may be a consequence of the intervention been reported?	A point was awarded if any adverse events, unwanted side effects, or lack thereof were explicitly indicated from either adherence or failure to adhere to recommended guidelines. A point was not awarded if the study made no mention of the presence or absence of adverse events. (0 = No; 1 = Yes)	1	See section on Safety Data and Adverse Events, infectious diseases
Have the characteristics of patients lost to follow-up been describe?	The authors of this tool indicated that this question should be answered "Yes" when clear reasons for loss to follow-up were described. For the purposes of this review, a point was awarded if a study explicitly reported the number and reason for patients lost to follow-up. (0 = No; 1 = Yes)	1	Nine of initially 40 patients (27.3%) did not enter into phase 2 of the study since they failed to enter into clinical remission (n = 6) and/or had ESR 20 mm/h (n=9).
Have actual probability values been reported (e.g. 0.035 rather than < 0.05) for the main outcomes except where the probability value is less than 0.001?	A point was awarded if the exact P value was provided for both statistically significant and non-significant results for at least the main outcome measures. A point was not awarded if a study simply indicated that the results for the main outcome measures were not significant without providing the exact P value. (0 = No; 1 = Yes)	0	
Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	A point was awarded if the study identified the source population for patients and described how the patients were selected. Patients were determined to be representative if they comprised	1	After screening 47 patients a total of 40 pediatric CD patients were included in this study. Seven patients were not included, since in 4 children disease activity dropped between the screening and inclusion visits and in 2

	the entire source population, an unselected sample of consecutive patients, or a random sample (only feasible where a list of all members of the relevant population exists). Where a study did not report the proportion of the source population from which the patients are derived, the question was answered as unable to determine. (1 = Yes; 0 = No; 0 = Unable to determine)		patients no clear written consent was obtained, whereas 1 eligible patient was not enrolled for serious doubts about compliance.
Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	The proportion of patients included in the study were representative of the population. Those asked who agreed to participate or responded should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population. No point was awarded if the proportion of those asked who agreed to participate or responded was not stated. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the staff, places and facilities where the patients were treated, representative of the treatment the majority of patients receive?	A point was awarded unless the study specifically stated that patients were treated by a therapist who received specialized training relative to guideline recommendations. (1 = Yes; 0 = No; 0 = Unable to determine)	1	
Was an attempt made to blind study subjects to the intervention they have received?	A point was awarded if the patients were not aware of, or would have no way of knowing (as in the case of retrospective studies), which intervention they received. The study was not awarded a point if it was prospective and failed to mention whether the patients had knowledge of whether they were assigned to the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Was an attempt made to blind those measuring the main outcomes of the intervention?	A point was awarded if the study specifically stated that those assessing the outcome measures were unaware of (or would have no way of	0	

	knowing) whether the patients were in the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)		
If any of the results of the study were based on "data dredging", was this made clear?	A point was awarded if no retrospective unplanned (at the outset of the study) subgroup analyses were reported. (1 = Yes; 0 = No; 0 = Unable to determine)	1	
In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	A point was awarded if all patients had same follow-up. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the statistical tests used to assess the main outcomes appropriate?	If the distribution of the data (normal or not) was not described, it was assumed that the estimates used were appropriate, and a point was awarded. No point was awarded for studies that reported qualitative or quantitative data without any form of statistical comparisons or if the statistical tests reported were not appropriate. (1 = Yes; 0 = No; 0 = Unable to determine)	1	Statistical analyses was preformed using the chi ² test and where applicable a paired t-test. Statistical significance was set at P <0.05.
Was compliance with the intervention/s reliable?	If the authors in prospective studies reported non-adherence to physical therapy intervention or adherence could not be determined, the study was not awarded a point. In retrospective studies, data were collected only for those patients who completed their episode of care (adherence to physical therapy assumed), and a point was awarded. For studies where the effect of any non-adherence was likely to bias any association to the null, the study was not awarded point. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the main outcomes measures used accurate (valid and reliable)?	A point was awarded if the primary outcome measures were thought to be valid and reliable (e.g., number of physical therapy visits per chart	1	See "Efficacy".

	report), regardless of whether reliability or validity was reported. A point was not awarded if at least one of the primary outcome measures in the study was not valid or reliable or if this information was not reported or could not be determined (i.e., a questionnaire without reported validity or reliability). (1 = Yes; 0 = No; 0 = Unable to determine)		
Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	A point was awarded when participants from both adherence and non-adherence groups were recruited from the same population. Otherwise, a point was not awarded (e.g., a point was not awarded when all participants from the adherence group received care at clinic A and all participants in the non-adherence group received care at clinic B, because they could have represented 2 distinct populations). (1 = Yes; 0 = No; 0 = Unable to determine)	0	40 pediatric patients with CD enrolled from 11 different French Speaking Group of Pediatric Gastroenterologists in France and Belgium. These patients were randomized and allocated to interventions.
Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	A point was awarded when the study provided a specific time line for patient recruitment (prospective studies) or when data were collected between reported dates of patient care (retrospective studies). (1 = Yes; 0 = No; 0 = Unable to determine)	1	Patients were enrolled during a 3-year inclusion period from May 2002 to April 2005
Were study subjects randomized to intervention groups?	A point for random allocation was awarded if random allocation of patients was stated in the "Method" section of the article. The precise method of randomization need not be specified. Quasi-randomization allocation procedures, such as allocation by bed availability, did not satisfy this criterion. For crossover study designs, a point was awarded when participants were randomly allocated in the order in which treatments were received. (1 = Yes; 0 = No; 0 = Unable to determine)	1	For randomization, a computer-generated adaptive randomization scheme was used, which used as stratification factors age and sex of patients, type, location and activity of disease, and past and present treatments, allowing to obtain comparable groups (no significant differences for sex, age, disease location, disease activity, and treatment).

Was the randomized intervention assignment concealed from both patients and healthcare staff until recruitment was complete and irrevocable?	The study did not receive a point unless the participants were randomly allocated and the methods for ensuring random allocation were specified. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Open-label
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	A point was awarded unless the effect of the main confounders was not investigated or confounding was demonstrated, but no adjustment was made in the final analyses. 1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were losses of patients to follow-up taken into account?	A point was awarded as long as the number of dropouts lost to follow-up accounted for less than 10% of the initial number of total participants or a maximum of 5% from each group. The question was answered with “unable to determine” if the number of patients lost to follow-up were not reported, could not be deduced from the outcome data (i.e., initial and final sample sizes not indicated) or the study methodology would not infer such information. (1 = Yes; 0 = No; 0 = Unable to determine)	1	Nine of initially 40 patients (27.3%) did not enter into phase 2 of the study since they failed to enter into clinical remission (n= 6) and/or had ESR =20 mm/h (n=9). 3 patients were excluded during phase 2 (2 patients in group A and 1 in group B) due to a secondary loss of response to IFX (between W30 and W38), failing to come into remission on 2 consecutive visits.
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	A point is awarded if power was stated. (1=Yes; 0=No)	0	
Total score		15/28	

Score	Rate
28-26	Excellent

25-20	Good
19-15	Fair
≤14	Poor

For Peer Review Only

Downs and Black quality assessment: Ruemmele et al. (2018) [30]

Criteria	Scoring criteria	Score	Notes/justifications
Is the hypothesis/aim/objective of the study clearly described?	A point was given if the hypothesis aim or objective of the study was implicitly or explicitly indicated anywhere in the article. (No=0; Yes=1)	1	In the current analysis of IMaGInE 1/2, we examined the efficacy of adalimumab for fistula healing over time, covering a total period of 292 weeks of adalimumab exposure.
Are the main outcomes to be measured clearly described in the "Introduction" or "Methods" section?	A point was given if the main outcomes to be measured were clearly described in the "Introduction" or "Methods" section. (No=0; Yes=1)	1	Fistula closure and fistula improvement were defined as closure of all IMaGInE 1 baseline fistulae or a decrease in number by $\geq 50\%$. [...]In addition, maintenance of fistula closure was analysed to Week 240 of IMaGInE 2 in patients with fistula closure at Week 0 of IMaGInE 2.
Are the characteristics of the patients included in the study clearly described?	A point was given if the inclusion or exclusion criteria, or both, were indicated. (no=0; Yes=1)	1	IMaGInE 1 [NCT00409682] ¹⁰ and IMaGInE 2 [NCT00686374] ¹¹ methods have been described previously. Patients from the overall intention-to-treat population with fistulae at screening and baseline of IMaGInE 1 were assessed in the current analysis. Fistulae were defined by draining cutaneous fistulae upon gentle compression during physical examination, without use of pelvic magnetic resonance imaging. All fistulae were perianal. Further details are provided in the Supplementary Methods.
Are the interventions of interest clearly described?	A point was given if the criteria for guideline adherence were described in detail. (No=0; Yes=1)	0	
Are the distributions of principal confounders for each group of participants to be compared clearly described?	Two points were awarded if a study reported any possible confounders (e.g., sex ratios, age, comorbidities, and severity of injury) that might account for differences between groups clearly	2	See Table 1

	in table format. One point was awarded if the study indicated that groups were matched for any such demographical variables or if potential confounders were mentioned in the text of the article but not clearly listed in table format. No points were awarded if the study did not report any confounders. (no=0; partially=1; Yes=2)		
Are the main findings of the study clearly described?	A point was awarded if quantitative data were reported for all of the main outcome measures indicated in the "Introduction" or "Methods" section. (0 = No; 1 = Yes)	1	See Figure 1 and 2
Does the study provide estimates of the random variability in the data for the main outcomes?	A point was awarded if the interquartile range (for non-normally distributed data), standard error, standard deviation, or confidence intervals (for normally distributed data) were reported. If the distribution of the data was not described, we assumed that the estimates used were appropriate, and we answered "Yes" (0 = No; 1 = Yes)	1	See Table 1 and Supplementary tables
Have all of the important adverse events that may be a consequence of the intervention been reported?	A point was awarded if any adverse events, unwanted side effects, or lack thereof were explicitly indicated from either adherence or failure to adhere to recommended guidelines. A point was not awarded if the study made no mention of the presence or absence of adverse events. (0 = No; 1 = Yes)	1	See Supplementary Tables
Have the characteristics of patients lost to follow-up been describe?	The authors of this tool indicated that this question should be answered "Yes" when clear reasons for loss to follow-up were described. For the purposes of this review, a point was awarded if a study explicitly reported the number and reason for patients lost to follow-up. (0 = No; 1 = Yes)	0	
Have actual probability values been reported (e.g. 0.035 rather than < 0.05)	A point was awarded if the exact P value was provided for both statistically significant and	0	

for the main outcomes except where the probability value is less than 0.001?	non-significant results for at least the main outcome measures. A point was not awarded if a study simply indicated that the results for the main outcome measures were not significant without providing the exact P value. (0 = No; 1 = Yes)		
Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	A point was awarded if the study identified the source population for patients and described how the patients were selected. Patients were determined to be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample (only feasible where a list of all members of the relevant population exists). Where a study did not report the proportion of the source population from which the patients are derived, the question was answered as unable to determine. (1 = Yes; 0 = No; 0 = Unable to determine)	0	A total of 36 children/adolescents had fistulae at baseline of IMaGINE 1 and were included in the analysis.
Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	The proportion of patients included in the study were representative of the population. Those asked who agreed to participate or responded should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population. No point was awarded if the proportion of those asked who agreed to participate or responded was not stated. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the staff, places and facilities where the patients were treated, representative of the treatment the majority of patients receive?	A point was awarded unless the study specifically stated that patients were treated by a therapist who received specialized training relative to guideline recommendations. (1 = Yes; 0 = No; 0 = Unable to determine)	1	

Was an attempt made to blind study subjects to the intervention they have received?	A point was awarded if the patients were not aware of, or would have no way of knowing (as in the case of retrospective studies), which intervention they received. The study was not awarded a point if it was prospective and failed to mention whether the patients had knowledge of whether they were assigned to the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Open -label
Was an attempt made to blind those measuring the main outcomes of the intervention?	A point was awarded if the study specifically stated that those assessing the outcome measures were unaware of (or would have no way of knowing) whether the patients were in the guideline adherence group. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Open-label
If any of the results of the study were based on "data dredging", was this made clear?	A point was awarded if no retrospective unplanned (at the outset of the study) subgroup analyses were reported. (1 = Yes; 0 = No; 0 = Unable to determine)	0	Post-hoc analysis
In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	A point was awarded if all patients had same follow-up. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were the statistical tests used to assess the main outcomes appropriate?	If the distribution of the data (normal or not) was not described, it was assumed that the estimates used were appropriate, and a point was awarded. No point was awarded for studies that reported qualitative or quantitative data without any form of statistical comparisons or if the statistical tests reported were not appropriate. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Was compliance with the intervention/s	If the authors in prospective studies reported	0	

reliable?	non-adherence to physical therapy intervention or adherence could not be determined, the study was not awarded a point. In retrospective studies, data were collected only for those patients who completed their episode of care (adherence to physical therapy assumed), and a point was awarded. For studies where the effect of any non-adherence was likely to bias any association to the null, the study was not awarded point. (1 = Yes; 0 = No; 0 = Unable to determine)		
Were the main outcomes measures used accurate (valid and reliable)?	A point was awarded if the primary outcome measures were thought to be valid and reliable (e.g., number of physical therapy visits per chart report), regardless of whether reliability or validity was reported. A point was not awarded if at least one of the primary outcome measures in the study was not valid or reliable or if this information was not reported or could not be determined (i.e., a questionnaire without reported validity or reliability). (1 = Yes; 0 = No; 0 = Unable to determine)	1	Fistulae were defined by draining cutaneous fistulae upon gentle compression during physical examination, without use of pelvic magnetic resonance imaging.
Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	A point was awarded when participants from both adherence and non-adherence groups were recruited from the same population. Otherwise, a point was not awarded (e.g., a point was not awarded when all participants from the adherence group received care at clinic A and all participants in the non-adherence group received care at clinic B, because they could have represented 2 distinct populations). (1 = Yes; 0 = No; 0 = Unable to determine)	1	
Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls	A point was awarded when the study provided a specific time line for patient recruitment (prospective studies) or when data were	1	

(case-control studies) recruited over the same period of time?	collected between reported dates of patient care (retrospective studies). (1 = Yes; 0 = No; 0 = Unable to determine)		
Were study subjects randomized to intervention groups?	A point for random allocation was awarded if random allocation of patients was stated in the "Method" section of the article. The precise method of randomization need not be specified. Quasi-randomization allocation procedures, such as allocation by bed availability, did not satisfy this criterion. For crossover study designs, a point was awarded when participants were randomly allocated in the order in which treatments were received. (1 = Yes; 0 = No; 0 = Unable to determine)	1	Post-hoc of a RCT
Was the randomized intervention assignment concealed from both patients and healthcare staff until recruitment was complete and irrevocable?	The study did not receive a point unless the participants were randomly allocated and the methods for ensuring random allocation were specified. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	A point was awarded unless the effect of the main confounders was not investigated or confounding was demonstrated, but no adjustment was made in the final analyses. (1 = Yes; 0 = No; 0 = Unable to determine)	0	
Were losses of patients to follow-up taken into account?	A point was awarded as long as the number of dropouts lost to follow-up accounted for less than 10% of the initial number of total participants or a maximum of 5% from each group. The question was answered with "unable to determine" if the number of patients lost to follow-up were not reported, could not be deduced from the outcome data (i.e., initial and final sample sizes not indicated) or the study methodology would not infer such information. (1 = Yes; 0 = No; 0 = Unable to determine)	0	

Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	A point is awarded if power was stated. (1=Yes; 0=No)	0	
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	A point is awarded if power was stated. (1=Yes; 0=No)	0	
Total score		13/28	

Score	Rate
28-26	Excellent
25-20	Good
19-15	Fair
≤14	Poor

Article highlights box

- Therapy with biologic drugs inhibiting Tumor Necrosis Factor-alpha (TNF- α), primarily infliximab and adalimumab, are now increasingly used to induce and maintain disease remission in Crohn's disease (CD); however, the efficacy on fistulizing perianal CD is still debated, especially in children and young adults. A significant number of children fail to achieve disease remission or recurrence of their fistulae despite initial success with an anti-TNF- α therapy. The management of paediatric perianal CD remains thus challenging and frequently relying on surgical resection of the diseased bowel.
- A systematic review was conducted of 29 studies (yielding a total of 565 perianal CD patients aged between 9 months to 18 years), showing that the majority of paediatric data on the efficacy and safety of anti-TNF agents for perianal fistulising CD treatment are of a descriptive nature: only two RCTs were identified; the remaining 28 studies were either post-hoc analysis of RCTs, retrospective or prospective observational, case series or case report, reflecting the practical situation in paediatric settings where treatment is often based on expert opinion, and whose assumed benefits are extrapolated from studies conducted in adults and from descriptive studies rather than prospective controlled trials.
- Overall, the pooled remission rates following the different follow-up time points between 2 weeks and 34 months of therapy occur in approximately 60% of patients, whereas in 55% a clinical response was observed. Infants with severe fistulising perianal CD seems to achieve remission with complete fistula closure during the induction therapy with anti-TNF- α agents. This evidence, however, has to be considered of low-quality as it emerged only from sporadic clinical case reports. Although this treatment showed promising results in the short-term, long-term studies are thus needed to confirm these initial findings.
- Safety issues of long-term use of anti-TNF therapies represent a significant concern. Acute infusion reactions were the most prevalent side effects. Serious adverse events leading to drug suspension seem relatively low (serious infection was the most frequent cause of drug withdrawal).
- Although the combined use of the different medical strategies available at the moment (anti-TNF- α and immunomodulators) together with surgical evaluation seem to improve clinical results and quality of life in these patients, there is a continued need for prospective real-world effectiveness and safety studies to characterise better the safety and efficacy profile of anti-TNF therapies in paediatric setting.