Comparability of European League Against Rheumatology-Recommended Pharmacological Treatments of Oral Ulcers Associated with Behçet’s Disease: A Systematic Literature Review of Randomized Controlled Trials

Tara Nazareth1
Erin M Hart2
Sarah M Ronnebaum3
Sandhya Mehta4
Dipen A Patel3
Ina Kötter5

1Amgen Inc., Thousand Oaks, CA, USA; 2Pharmerit - an OPEN Health Company, New York, NY, USA; 3Pharmerit - an OPEN Health Company, Bethesda, MD, USA; 4Celgene Corporation, Summit, NJ, USA; 5Division of Rheumatology and Systemic Inflammatory Diseases, University Hospital Hamburg Eppendorf and Clinic for Rheumatology and Immunology, Bad Bramstedt, Germany

**Objective:** Oral ulcers are the cardinal manifestation in Behçet’s disease (BD). The 2018 European League Against Rheumatism (EULAR) recommendations describe treatments for BD-associated oral ulcers with mucocutaneous involvement; however, little comparative effectiveness information for these agents is available. In the absence of head-to-head trials, an indirect treatment comparison (ITC) could provide useful evidence regarding comparative effectiveness of BD treatments. The purpose of this study was to conduct a comparative systematic literature review (SLR) and similarity assessment of randomized controlled trials (RCTs) investigating the oral ulcer-related efficacy outcomes of EULAR-recommended treatments for BD-associated oral ulcers to determine the feasibility of an ITC.

**Methods:** An SLR was performed to identify relevant RCTs indexed in MEDLINE/Embase before May 29, 2019. RCT similarities for the ITC were assessed based on a step-wise process recommended by the International Society for Pharmacoeconomics and Outcomes Research.

**Results:** In total, 317 articles were identified, of which 14 RCTs, reflecting 11 EULAR-recommended treatments, were evaluated in a similarity assessment. Number of oral ulcers, resolution of oral ulcers, and healing time for oral ulcers were identified as the possible oral ulcer-related outcomes. After completing the similarity assessment of these outcomes, it was determined that a robust ITC was infeasible for the three oral ulcer-related outcomes due to heterogeneity in outcomes reporting, study design, and/or patient characteristics. More broadly, the results underscore the need for and consistent use of standardized measures for oral ulcer outcomes to facilitate comparative research.

**Conclusion:** In the absence of head-to-head RCTs and infeasibility of quantitative ITC, comparative assessments for BD-associated oral ulcers are limited, including comparative effectiveness and cost-effectiveness evaluations. Healthcare decision-makers must continue to base treatment decisions on the extent and strength of available evidence (eg, robust RCTs), clinical guidelines, real-world experience, and patient considerations.

**Keywords:** Behçet syndrome, oral ulcer, comparative effectiveness research

---

**Plain Language Summary**

- Little information is available on the comparative effectiveness of treatments recommended by the 2018 European League Against Rheumatism (EULAR) for oral ulcers.
in patients with Behçet’s disease (BD), given the lack of comparative randomized controlled trials (RCTs).

- An indirect treatment comparison (ITC) may be evaluated when RCTs directly comparing treatments are unavailable.
- A comprehensive review of the scientific literature and assessment of the similarity of RCTs investigating the oral ulcer-related efficacy outcomes of EULAR-recommended treatments in patients with BD was first conducted. Three oral ulcer-related outcomes were relevant and appropriate for assessing the feasibility of an ITC: number of oral ulcers, resolution of oral ulcers, and healing time of oral ulcers.

- Results of our similarity assessment of RCTs identified differences in (a) the ways outcomes are reported, (b) the ways studies are designed, and/or (c) the characteristics of the patients included in the studies (eg, average age, severity of disease).

- A robust ITC was determined to be infeasible for the oral ulcer-related outcomes identified in BD RCTs.

- Results of the similarity assessment also highlight the need for standardized measures for oral ulcer outcomes and comparative research for oral ulcer treatments in BD.

- In the absence of an ITC, the results of this study suggest that treatment decisions for patients with BD and oral ulcers must continue to be based on the extent and strength of available scientific evidence (eg, high-quality RCTs), clinical guidelines, real-world experience, and patient considerations.

Introduction

Behçet’s disease (BD) is a rare, systemic variable vessel vasculitis that involves multiple organ systems. Manifestations of BD may involve the skin and mucous membranes (ie, mucocutaneous involvement), eyes, vascular system (both veins and arteries), gastrointestinal tract, nervous system, and joints. Oral ulcers are the first manifestation experienced by a large majority of patients, appear in the oral cavity, and can be experienced throughout a patient’s lifetime. Oral ulcers are also the most common manifestation of BD. Results from a systematic literature review (SLR) showed that the median prevalence of oral ulcers among 27 observational studies was 100% (interquartile range: 84–100%), suggesting that oral ulcers are experienced by essentially all patients with BD. Furthermore, recurrent oral ulcers are required for diagnosis under the most commonly used set of criteria (the International Study Group criteria). Evidence suggests patients with BD may experience on average approximately 13 oral ulcers per year, and oral ulcers can also occur concurrently with other symptoms, such as fatigue- and sleep-related issues, pain, headache, and other skin lesions (eg, pseudofolliculitis, papulopustular lesions). As BD progresses, patients may experience additional manifestations, including ocular, vascular, and neurological involvement.

The high prevalence and recurrence of oral ulcers and their negative impact on the clinical and humanistic burden of disease represent an important opportunity for treatment. Oral ulcers can lead to poor dental hygiene, limit essential daily activities like eating and talking, and impair quality of life (QoL) overall and across several domains, such as physical and social functioning, sleep and fatigue, and mental health. Furthermore, patients with BD who have oral ulcers have reduced QoL compared with healthy controls and patients with other chronic conditions.

The European League Against Rheumatism (EULAR) is the primary source of treatment recommendations for BD-associated oral ulcers. In 2018, the EULAR recommendations suggested that topical treatments (eg, corticosteroids) should be considered first for the treatment of oral ulcers. Colchicine was recommended to be tried first for the prevention of recurrent mucocutaneous lesions, including oral ulcers, due to its good safety and tolerability profile; however, the EULAR recommendations also note conflicting efficacy data for colchicine in oral ulcers. Immunomodulatory and immunosuppressive drugs, such as azathioprine, thalidomide, interferon-α, tumor necrosis factor (TNF)-α inhibitors, and apremilast, are recommended by EULAR based on evidence from at least one randomized controlled trial (RCT) for select cases, such as patients whose oral ulcers continue to recur despite the use of colchicine. In 2019, Phase 3 RCT (RELIEF) data became available for apremilast, leading to a formal indication for BD-associated oral ulcers and subsequent approvals in the United States and Japan (2019) and the European Union (2020). Although real-world evidence indicates oral corticosteroids may be used to treat BD, they are not recommended by EULAR to treat BD-associated oral ulcers; furthermore, a variety of published research challenges their use as chronic therapy due to substantial associated morbidity.

Evidence-based healthcare decision-making requires a comparison of relevant competing interventions, and systematic reviews of RCTs are a standard method of analyzing information in the healthcare setting. Comparative evidence is also typically included in health technology assessments (HTAs), which is a multidisciplinary research process designed to inform policy decisions and to provide decision-
makers with a summary of information related to the use of a health technology. The SLR underlying the EULAR recommendations for mucocutaneous manifestations found heterogeneously reported outcomes for BD-associated oral ulcers (eg, number, duration, frequency, severity), discrepancies in statistical significance of efficacy outcomes, a dearth of adverse events reporting, and lack of head-to-head comparisons. The authors noted that the limitations observed in the literature hindered a comparison of the efficacy of treatments and that a formal comparative effectiveness analysis would be valuable.

Although the SLR underlying the EULAR recommendations provides a valuable assessment of efficacy and safety outcomes for individual treatments for mucocutaneous manifestations (including oral ulcers), the objective of the SLR was not to perform an in-depth comparison of trial similarity (eg, trial design, patient and treatment characteristics, outcome reporting methods) for oral ulcer treatments. Thus, we conducted a comparative SLR and similarity assessment of RCTs investigating the oral ulcer-related efficacy outcomes of EULAR-recommended treatments for oral ulcers associated with BD to determine the feasibility of a comparative effectiveness study via an indirect treatment comparison (ITC). Given that no direct comparative effectiveness evidence is available, the comparative SLR and RCT similarity assessment would inform the potential to proceed with an ITC and would provide useful evidence regarding comparative effectiveness of treatments for consideration by healthcare decision-makers, including HTA bodies.

Materials and Methods
Systematic Literature Review: Search Strategy and Article Selection
An SLR was performed in accordance with standard practice using a predefined search strategy (Supplementary Table 1) to identify RCTs (articles and their supplements or letters) indexed in MEDLINE/Embase, written in English, and published any time up to May 29, 2019. The references identified in the literature search were reviewed for potential inclusion.

Articles were included if they i) were RCTs, ii) enrolled adult (≥18 years of age) patients with BD and oral ulcers, iii) assessed EULAR-recommended systemic or topical pharmacologic treatment for mucocutaneous manifestations (ie, topical steroids, colchicine, azathioprine, thalidomide, interferon-α, TNF-α inhibitors, and apremilast) or placebo, and iv) described efficacy or patient-reported outcomes related to improvement in oral ulcer-related outcomes. Excluded were non-randomized trials, single-arm or dose-escalation trials lacking placebo, observational studies, or results published only as conference abstracts. Full inclusion and exclusion criteria can be found in Table 1.

Table 1 Eligibility Criteria for Randomized Controlled Trials (PICOS Model)

| Inclusion                                                                 | Exclusion                                                                 |
|--------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Population                                                               | Exclusion                                                                 |
| Patients with Behçet’s disease aged ≥18 years                             | Lifestyle/behavioral interventions                                        |
| Intervention                                                             | Other EULAR-recommended therapies or placebo                             |
| EULAR-recommended therapies*                                             | Non-EULAR-recommended treatments or none                                   |
| Comparator                                                               | Outcomes other than those relating to BD-associated oral ulcers           |
| Other                                                                    |                                                                            |
| Randomized, parallel-arm, placebo-controlled, withdrawal trials; randomized, crossover trials reporting results at first cross* English language Articles or letters indexed in MEDLINE/Embase Any publication date | Non-randomized trials Single-arm trials Dose-escalation trials lacking placebo Observational studies Conference abstracts |

Notes: *Includes topical and systemic measures such as steroids, colchicine, azathioprine, thalidomide, interferon-α, tumor necrosis factor-α inhibitors, and apremilast, per the 2018 EULAR recommendations for the management of mucocutaneous involvement (including oral ulcers) in Behçet’s disease. †Only parallel-arm trials or crossover trials reporting results at first cross can be considered for the indirect treatment comparison.

Abbreviations: BD, Behçet’s disease; EULAR, European League Against Rheumatism; PICOS, participant, intervention, control, outcome, study design.
The literature search results were downloaded into a single Excel file, and one reviewer (EH) independently screened all titles and abstracts resulting from the search methodology to identify articles for full-text review. Citations selected for full-text evaluation were reviewed by two independent reviewers for potential inclusion in the feasibility assessment (EH and SR).

Data Extraction

One reviewer extracted all relevant data into an extraction sheet developed in Excel (EH), and a second reviewer performed quality assurance (SR). Extracted data included key trial characteristics, oral ulcer-related outcomes (eg, frequency, number, size, duration, pain, response, severity), and key baseline patient demographics and clinical characteristics. Such key baseline clinical characteristics included oral ulcer-related outcomes (eg, duration, number, healing time, pain), prior treatments, concomitant treatments, disease duration, and presence of other BD manifestations.

Similarity Assessment for an ITC

Evaluation of RCTs in the similarity assessment for an ITC was conducted using the step-wise process (Figure 1) recommended by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).32 First, RCTs reporting oral ulcer-related outcomes that were unique to one treatment were excluded from further review, as these would not be comparable against other treatments. Second, oral ulcer-related outcomes commonly reported for at least two treatments were identified. Third, for each commonly reported outcome, the following steps were followed to determine whether RCTs reporting that outcome should be excluded in an ITC: i) exclude RCTs if the outcome reporting method was different from the other RCTs, such as different outcome definitions (eg, new oral ulcers vs any oral ulcers) or different units of measure (eg, mean number vs probability of an event, where the endpoint could not be easily transformed into another commonly reported outcome); ii) exclude RCTs if the outcome was reported at different time points (ie, by at least 2 weeks; the 2-week time point was chosen because individual oral ulcers heal in approximately 2 weeks) compared with other RCTs; and iii) exclude RCTs if the patient demographics and clinical characteristics were not similar to other RCTs. Expert opinion was applied to qualitatively determine whether differences were clinically meaningful. After completion of the step-wise process described above, an ITC would be feasible for an oral ulcer-related outcome if at least two treatments were included.

Results

Study Selection

Overall, 317 articles were identified, including 310 identified through a pre-specified search of MEDLINE/Embase and seven articles identified in published systematic reviews (Figure 2). Of these 317 articles, 302 were excluded during the title/abstract screening. Thus, 15 articles were screened with full text, and one article assessing topical cyclosporine-A was excluded because of its publication type (ie, conference abstract).33 The remaining 14 articles (representing 14 RCTs) were evaluated in the similarity assessment for an ITC and included the following treatments: apremilast,34,35 colchicine,36–40 colchicine + benzathine penicillin,39 interferon-α-2a,41 triamcinolone acetonide ointment,42 interferon-α lozenges,43 cyclosporine,48 interferon-α-2c hydrogel,44 thalidomide,45 etanercept,46 and azathioprine.47

Efficacy Outcomes

In the first step of the similarity assessment for an ITC (Figure 3), six RCTs only reporting oral ulcer-related outcomes unique to one treatment were excluded. The excluded RCTs were i) an RCT evaluating colchicine, which rated aphthous stomatitis on a scale ranging from 0 to 3+(numerical values used in this scoring system were not clearly defined)35; ii) an RCT evaluating colchicine, which reported Iranian BD dynamic activity measure scores for oral ulcers37; iii) an RCT evaluating colchicine and cyclosporine, which reported the percentage of patients who experienced “alleviated oral ulcers” (“alleviated” was not defined)38; iv) an RCT evaluating interferon-α lozenges, which reported the mean change in the total surface area of all oral ulcers present43; v) an RCT evaluating triamcinolone acetonide ointment, which reported the percentage of patients experiencing a “positive response” (“positive” was not defined)42; and vi) an RCT evaluating interferon-α-2a, which reported the number of patients who responded to treatment (authors considered all signs and symptoms of BD when determining response and results were not specific to oral ulcers).41 In the last RCT listed here, additional oral ulcer-specific endpoints were assessed; however, these outcomes were not further assessed for comparability.
because the authors only reported $P$ values and no estimates were provided.\textsuperscript{41}

In the second step of the similarity assessment (Figure 3), eight RCTs reporting on seven active treatments were evaluated.\textsuperscript{34,35,39,40,44–47} The following BD-associated oral ulcer-related outcomes were reported for multiple treatments: i) number of oral ulcers; ii) resolution of oral ulcers (ie, patients who were oral ulcer-free); and iii) healing time of oral ulcers. Not all eight RCTs reported on each of the three outcomes.

In the third step of the similarity assessment, we separately assessed whether an ITC would be feasible for each of the commonly reported oral ulcer-related efficacy outcomes.
Figure 2 PRISMA diagram.
Notes: *The 14 included articles represent 14 RCTs.
Abbreviations: BD, Behçet’s disease; RCT, randomized controlled trial.

Figure 3 Overview of the results of the similarity assessment to determine the feasibility of performing an ITC for oral ulcer-related outcomes in RCTs.
Abbreviations: ITC, indirect treatment comparison; RCT, randomized controlled trial.
## Table 2 Comparison of Outcome Reporting Methods for Oral Ulcers: Number, Resolution, and Healing Time

| Treatment Assessed (vs Placebo) | Citation | Number of Oral Ulcers | Comparability |
|--------------------------------|----------|-----------------------|---------------|
| **Outcome Reporting Method**   |          |                       |               |
| • Colchicine                   | Calguneri 1996<sup>39</sup> | Number and frequency of oral ulcers before/after treatment; definitions of number, frequency, and before/after were not defined | X             |
| • Colchicine + benzathine penicillin<sup>a</sup> | Yurdakul 2001<sup>40</sup> | Mean number of oral ulcers during the 24-month study period | X             |
| • Thalidomide 100 mg/day       | Hamuryuden 1998<sup>42</sup> | Number of oral ulcers reported as combined data obtained at study visits and from reports by patients on occurrence of lesions between study visits; only new oral ulcers were reported | X             |
| • Thalidomide 300 mg/day       | Hamuryuden 1998<sup>42</sup> | Mean sum of oral ulcers observed over all visits during the 24-week study period | X             |
| • Interferon-α-2c hydrogel     | Hamuryuden 1991<sup>44</sup> | Mean number of oral ulcers at visit | ✓             |
| • Apremilast                   | Hatemi 2015<sup>34</sup> | Least-squares mean number of oral ulcers at visit | ✓             |
| • Apremilast                   | Hatemi 2019<sup>35</sup> | Mean number of oral ulcers at visit | ✓             |
| • Etanercept                   | Melikoğlu 2005<sup>46</sup> | Mean number of oral ulcers at visit | ✓             |

| Treatment Assessed (vs Placebo) | Citation | Resolution of Oral Ulcers | Comparability |
|--------------------------------|----------|---------------------------|---------------|
| **Outcome Reporting Method**   |          |                          |               |
| • Thalidomide 100 mg/d         | Hamuryuden 1998<sup>42</sup> | Complete response defined as absence over the 24-week study period of oral or genital ulcers of any size at and between visits | X             |
| • Thalidomide 300 mg/d         | Hamuryuden 1998<sup>42</sup> | Complete response defined as the percentage of patients with no oral ulcers | ✓             |
| • Apremilast                   | Hatemi 2015<sup>34</sup> | Complete response defined as the percentage of patients with no oral ulcers | ✓             |
| • Apremilast                   | Hatemi 2019<sup>35</sup> | Complete response defined as the percentage of patients with no oral ulcers | ✓             |
| • Etanercept                   | Melikoğlu 2005<sup>46</sup> | Percentage of patients free of oral ulcers | ✓             |
| • Azathioprine                 | Yazici 1990<sup>47</sup> | Percentage of patients with oral ulcers | ✓             |

| Treatment Assessed (vs Placebo) | Citation | Healing Time for Oral Ulcers | Comparability |
|--------------------------------|----------|-------------------------------|---------------|
| **Outcome Reporting Method**   |          |                               |               |
| • Colchicine                   | Hatemi 2019<sup>35</sup> | Time to resolution of oral ulcers (ie, no oral ulcers) in weeks | X             |
| • Colchicine + benzathine penicillin<sup>a</sup> | Calguneri 1996<sup>39</sup> | Healing time in days of oral ulcers determined for each episode | X             |
| • Colchicine                   | Yurdakul 2001<sup>40</sup> | Distribution of time to first occurrence of sustained absence of oral ulceration | X             |

**Note:** <sup>a</sup>Not a placebo-controlled trial.
Number of Oral Ulcers

Seven RCTs evaluating six active treatments reported the number of oral ulcers (Table 2).43,35,39,40,44-46 Four RCTs were not comparable to the other three RCTs due to differences in outcome reporting methods. Differences in outcome reporting methods included reporting any vs only new oral ulcers and reporting the sum of oral ulcers over the entire study period vs at specified time points.

Of the four RCTs, one trial evaluating colchicine and colchicine + benzathine penicillin reported the number of oral ulcers and the frequency of oral ulcers before and after treatment.39 The definitions of the number and frequency of oral ulcers, the definitions of before and after treatment, and the measure (eg, mean vs median) for the number of oral ulcers were unclear. Another RCT evaluating colchicine reported the mean number of oral ulcers during the entire 24-month study period; however, results at specific interim time points were not reported.40 A third RCT evaluating thalidomide reported the number of new oral ulcers as combined data obtained at study visits and from reports by patients on the occurrence of lesions between study visits.45 The fourth RCT evaluating interferon-α-2c hydrogel was excluded because it reported the mean sum of oral ulcers observed over all visits during the 24-week study period.44 Results at interim time points were not reported.

Three of the seven RCTs (evaluating two active treatments) reporting the number of oral ulcers had comparable outcome reporting methods with regard to the outcome metric and reporting interval (Table 2). Two multicountry RCTs evaluating apremilast43,35 and one RCT in Turkey evaluating etanercept46 reported the mean number of oral ulcers at specified time points, and all three RCTs reported the mean number at Week 4. However, the two apremilast RCTs43,35 had dissimilar baseline patient demographics and disease characteristics compared with the etanercept RCT46 (Table 3). First, both of the multicountry apremilast RCTs required the presence of at least two oral ulcers as an inclusion criterion, whereas the Turkish etanercept RCT required mucocutaneous disease and/or arthritis and a positive pathergy test.43,35,46 As such, it is unclear how many patients in the etanercept RCT had oral ulcers, while all patients in the apremilast RCTs had oral ulcers; the mean number of oral ulcers per patient at baseline was lower in the etanercept RCT (placebo group=1.8 and etanercept group=2.0)46 than in the apremilast RCTs (Phase 2 RCT: placebo group=3.1 and apremilast group=3.2; Phase 3 RCT “RELIEF”: placebo group=3.9 and apremilast group=4.2).43,35 Second, patients in the etanercept RCT had shorter disease duration compared with the apremilast RCTs (~3 years vs ~6 years). Third, the etanercept RCT allowed patients to use topical/oral corticosteroids and azathioprine for severe disease (eg, presence of deep vein thrombosis or painful arthritis), whereas the apremilast RCTs did not permit the use of corticosteroids during the double-blind phase of the RCTs. Lastly, the etanercept RCT46 included only male patients (N=40; n=20 in the active arm), whereas the apremilast RCTs43,35 included male and female patients and did not report data by sex (ie, apremilast Phase 2 RCT comprised 31% male [34/111] and 69% female [77/111], and the apremilast Phase 3 RCT comprised 39% male [80/207] and 61% female [127/207]).43,35 After following the step-wise approach for the similarity assessment, based on these differences, it was determined that an ITC would not be feasible for the number of oral ulcers.

oral Ulcer Resolution

Five RCTs evaluating four active treatments reported the resolution of oral ulcers (Table 2).43,34,35,45-47 One RCT evaluating thalidomide was not comparable to the other four RCTs due to differences in outcome reporting method; the complete response was defined as the absence of oral or genital ulcers of any size at and between visits over a 24-week period, and results were not reported by ulcer type.45 The remaining four RCTs assessed the absence of oral ulcers only and had comparable outcome reporting methods to discern the percentage of patients who were oral ulcer-free. Two RCTs evaluating apremilast43,35 defined complete response as the percentage of patients who were oral ulcer-free; one RCT evaluating etanercept46 reported the percentage of patients who were oral ulcer-free, and one RCT evaluating azathioprine47 reported the percentage of patients who had oral ulcers. Therefore, the percentage of patients who were oral ulcer-free could be calculated. Although these four RCTs had similar outcome reporting methods to discern the percentage of patients who were oral ulcer-free, they were not comparable due to differences in time reporting. The two RCTs evaluating apremilast43,35 reported results at week 12, the RCT evaluating etanercept46 reported results at week 4, and the RCT evaluating azathioprine47 reported results at month 24. After following the step-wise approach for the similarity
Table 3 Comparison of Trial Characteristics, Baseline Patient Demographics, and Baseline Patient Clinical Characteristics Among Trials Assessing the Mean Number of Oral Ulcers

| Trial Characteristics | Melikoglu 2005\textsuperscript{16} | Hatemi 2015\textsuperscript{34} | Hatemi 2019\textsuperscript{15} |
|-----------------------|------------------------------------|---------------------------------|---------------------------------|
| Country               | Turkey                             | Turkey, United States           | Multi-Country                    |
| Manifestations required for inclusion | Mucocutaneous disease and/or arthritis + positive pathology test | ≥2 oral ulcers | ≥2 oral ulcers |
| Major organ involvement | None permitted                     | None permitted                   | None permitted                   |
| Have joint involvement/arthritis | NR but allowed                     | NR but allowed                   | NR but allowed                   |
| Concomitant treatment use during trial | Corticosteroids and azathioprine were allowed for patients with relatively severe disease during the entire study period | None related to BD were allowed in placebo-controlled period | Colchicine, glucocorticoids, immunosuppressants (including azathioprine), and biologics were not permitted in placebo-controlled period |

| Baseline Patient Characteristics | Melikoglu 2005\textsuperscript{16} | Hatemi 2015\textsuperscript{34} | Hatemi 2019\textsuperscript{15} |
|---------------------------------|------------------------------------|---------------------------------|---------------------------------|
| Treatment | Etanercept | Placebo | Apremilast | Placebo | Apremilast | Placebo |
| N      | 20         | 20      | 55         | 56      | 104        | 103      |
| Oral ulcers at baseline, %     | NR        | NR      | NR         | NR      | NR         | NR       |
| Age, mean (SD), years          | 28.5 (5.3) | 30.8 (6.2) | 34.3 (NR) | 34.7 (NR) | 38.5 (NR) | 40.6 (NR) |
| Male, %                         | 100       | 100     | 29         | 32      | 38.5       | 38.8     |
| Disease duration, mean (SD), years | 2.8 (2.6) | 4.3 (3.2) | 4.92 (NR) | 5.72 (NR) | 6.7 (NR)  | 6.9 (NR) |
| Baseline number of oral ulcers, mean (SD) | 2.0 (1.3) | 1.8 (1.6) | 3.2 (2.0) | 3.1 (1.3) | 4.2 (NR)  | 3.9 (NR) |
| Colchicine use prior to trial, % | 55        | 35      | NR         | 50      | 54         |          |
| Concomitant treatment use during trial, % | 0         | 0       | NR         | 0       | 0          | 0        |
| Colchicine                      | 0         | 0       | NR         | 7.7     | 5.8        |          |
| Azathioprine                    | 5         | 5       | 0          | 0       | 0          | 0        |
| Topical corticosteroids         | 0         | 0       | 0          | NR      | NR         | NR       |
| Oral corticosteroids (prednisolone) | 10       | 10      | 20.5       | 16.5    | 8.7        | 8.7      |
| Indomethacin                    | 5         | 0       | 0          | NR      | NR         | NR       |
| Naproxen                        | 0         | 5       | 8.7        | 8.7     | 8.7        |          |
| Ibuprofen                       | NR        | NR      | NR         | NR      | NR         | NR       |

Notes: Patients who had relatively severe disease include two patients with recent exacerbation of deep vein thrombosis (one patient each in etanercept and placebo arm) and one patient in the placebo arm with painful arthritis. Abbreviations: BD, Behçet’s disease; N, number; NR, not reported; SD, standard deviation.

Assessment, it was determined that an ITC is not feasible for the resolution of oral ulcers.

Healing Time of Oral Ulcers

Three RCTs evaluating three active treatments reported the healing time for oral ulcers (Table 2). However, these RCTs were not comparable due to differences and lack of clarity in outcome reporting methods, specifically whether healing time referred to the time needed to be oral ulcer-free or the duration of an oral ulcer. An RCT evaluating apremilast reported the time to resolution of oral ulcers (ie, oral ulcer-free) in weeks. An RCT evaluating colchicine and colchicine + benzathine penicillin reported the healing time in days for oral ulcers for each attack, but did not clarify whether the healing time was the duration of individual oral ulcers or all oral ulcers; thus, this RCT could not be compared with the apremilast RCT. Lastly, an RCT evaluating colchicine reported the distribution of time to the first occurrence of the sustained absence of oral ulceration. The probability of complete resolution of oral ulcers was reported, but not the time it took for patients to become oral ulcer-free; thus, this RCT could not be compared with the
apremilast RCT. After following the step-wise approach for the similarity assessment, it was determined that an ITC is not feasible for the healing time of oral ulcers.

**Summary of Results**

Among the 14 RCTs identified in the SLR, three oral ulcer-related efficacy outcomes (ie, number of oral ulcers, resolution of oral ulcers, and healing time for oral ulcers) were identified as possible oral ulcer-related outcomes to assess in an ITC. After completion of the similarity assessment, it was determined that a robust ITC was infeasible for the three oral ulcer-related outcomes due to heterogeneity in outcomes reporting, study design, and/or patient characteristics. More broadly, the results demonstrate that standardized measures for oral ulcer outcomes and comparative research for oral ulcer treatments are lacking.

**Discussion**

Comparative effectiveness is an important component of regulatory and reimbursement decision-making. ITCs in particular are commonly used in HTAs to provide comparative evidence, especially in the absence of direct evidence from head-to-head trials; ITCs are also important in providing robust estimates that combine direct and indirect evidence when head-to-head data are available.\(^{29,30}\)

However, comparative effectiveness research is often challenging to conduct and therefore unavailable, particularly in rare diseases such as BD, and in the clinical trial setting.\(^{48}\) We performed a comparative SLR and similarity assessment to determine the feasibility of an ITC of the EULAR-recommended treatments for BD-associated oral ulcers. Results demonstrated that, among the 14 RCTs identified in our similarity assessment, a robust ITC was not feasible based on heterogeneous outcomes reporting, study design, and baseline patient demographics and clinical characteristics. Therefore, our results provide an additional comparative analysis to consider along with findings from the SLR informing the 2018 EULAR recommendations for mucocutaneous involvement in BD.\(^{21}\)

Although a robust indirect comparison with apremilast and treatments used off-label for BD-associated oral ulcers would provide additional insight for healthcare decision-makers, our study results demonstrate that this is not possible without additional data from RCTs specifically designed to support such comparisons (ie, data which are currently unavailable). The inability to perform such an assessment dramatically limits the comparative-effectiveness and cost-effectiveness evaluations of various treatments for BD-associated oral ulcers. Furthermore, the dissimilarities found in this study highlight the lack of standardized measures in BD overall and for BD-associated oral ulcers specifically (eg, disease activity, severity). The Outcome Measures in Rheumatology (OMERACT) Behçet’s Syndrome Working Group also identified the variability and lack of standardized definitions of outcomes across trials evaluating treatments in BD as an area of unmet need.\(^{49,50}\)

The OMERACT Behçet’s Syndrome Working Group proposed a main core set of outcomes that evaluate all organ systems and should be used in all trials of BD, as well as organ-specific subdomains that can be used in trials focusing on a particular organ system (eg, for assessing oral ulcer-related efficacy).\(^{49,50}\) In 2016, the OMERACT Behçet’s Syndrome Working Group suggested that the most important measures related to oral ulcers included oral ulcer number, duration, and associated pain.\(^{49}\) Standardized definitions would ideally be adopted for frequently used outcomes to evaluate oral ulcer-related efficacy, such as response, relapse, and remission.\(^{49}\) Furthermore, patients expressed a need to assess the effect of oral ulcers on difficulty with household tasks, stress, and issues in relationships with partners.\(^{49}\)

A reliable and validated patient-reported outcome measure that incorporates domains endorsed by patients and physicians would also be valuable.\(^{49}\) Most of the EULAR-recommended therapies for oral ulcers were not robustly nor recently studied among patients with BD-associated oral ulcers. Similarly, no validated patient-reported outcome instrument as suggested by OMERACT currently exists.

The RCTs identified in our literature search had several limitations that ultimately contributed to the infeasibility of an ITC, including lack of clarity in study designs, particularly in ambiguous inclusion criteria; baseline characteristics that were not fully reported; and outcomes and time points that were often not clearly defined. In addition, only the apremilast RCTs included patients from more than one country,\(^{34,35}\) and several RCTs had relatively short study durations (eg, 4 weeks) that may not sufficiently predict patient experience on treatment intended to be used as chronic therapy (Supplementary Table 2). Finally, some studies had very small sample sizes (eg, the etanercept RCT from Turkey with 20 patients in the active, non-placebo arm); if undertaken, an ITC performed on small sample sizes would lead to imprecise effect estimates with wide confidence intervals and inconclusive results.\(^{51}\)

There are significant unmet needs in patients with BD. Real-world evidence further suggests that approximately 10% of patients with BD do not receive treatment,\(^{8}\)
possibly due to concerns about treatment (ie, actual or potential ineffectiveness of therapy, cost of treatment, and/or side effects). Oral ulcers are the most common manifestation among all patients with BD; yet evidence suggests the treatments commonly prescribed for patients with BD to address oral ulcers (eg, colchicine, corticosteroids) are used off-label, indicated for short-term use, have non-specific actions, and/or temporarily relieve symptoms. Many of the treatments for BD are also associated with serious safety concerns, for example, long-term use of topical or oral corticosteroids with several serious side effects, short-term use of oral corticosteroids with sepsis, venous thromboembolism, and fracture, and immunosuppressants with serious adverse events. Since the publication of the EULAR guidance in 2018, the only new RCT evaluating BD-associated oral ulcers is the multi-country Phase 3 RCT RELIEF study assessing apremilast for the treatment of active oral ulcers in patients with BD. Results of the Phase 3 study are consistent with the Phase 2 RCT in which patients receiving apremilast experienced a significant and sustained reduction in the number of oral ulcers, pain from oral ulcers, and disease activity, as well as a significant and sustained improvement in QoL. The safety profile of apremilast in RELIEF was consistent with other currently approved indications (ie, chronic plaque psoriasis and psoriatic arthritis), with approximately 509,000 patients treated as of March 31, 2020, since its first approval in 2014. Apremilast has subsequently gained regulatory approval for the treatment of oral ulcers in BD in several countries, including the United States, Japan, Switzerland, and the European Union.

Acknowledging the practical challenges with conducting assessments focused on this rare disease, future comparative studies and real-world studies assessing the efficacy, safety, and humanistic value of treatments used for BD-associated oral ulcers across longer time frames would be valuable in informing the comparative effectiveness of available therapies. The consistent use of a recommended set of relevant endpoints and reporting intervals to consider for implementation for each manifestation of BD, including oral ulcers, given the near-universal prevalence (such as the Core Domain Set proposed by the OMERACT Behçet’s Syndrome Working Group), would help to facilitate consistency in research, foster comparability of treatments, and ultimately help to guide healthcare decision-making.

This study has some limitations. First, the research question is primarily focused on EULAR-recommended treatments as opposed to any treatment for BD that might be used off label for BD-associated oral ulcers; however, we believe EULAR to be an important foundation for the treatment set evaluated, and no evidence from the SLR performed for this study suggested that non-recommended agents would be sufficiently similar and appropriate for an ITC. Second, the similarity assessment relies on qualitative comparison using expert feedback, given that there is no recommended quantitative method for determining similarity; however, our approach utilized a rigorous step-wise process recommended by ISPOR, and criteria for RCT exclusion were objective and defined a priori. Third, the evidence was limited to RCTs and no observational data were included; however, ITC methodology typically relies on RCT evidence to ensure a robust analysis.

Conclusions

Results of this evaluation show the clinical evidence for EULAR-recommended treatments for oral ulcers associated with BD cannot be robustly (ie, quantitatively) compared; therefore, comparative effectiveness insight via an ITC is currently not possible, thereby also limiting cost-effectiveness insights as a result. Where formal assessments are not possible, healthcare decision-makers, including HTA bodies, payers, and clinicians, must continue to base treatment decisions on the extent and strength of available evidence (eg, robust RCTs), clinical guidelines, real-world experience, and patient considerations. Consistent use of standardized measures for oral ulcer outcomes (such as those proposed by the OMERACT Behçet’s Syndrome Working Group) across clinical trials is needed to facilitate comparative research of BD-associated oral ulcer treatments.

Abbreviations

BD, Behçet’s disease; EULAR, European League Against Rheumatism; HTA, health technology assessment; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; ITC, indirect treatment comparison; OMERACT, Outcome Measures in Rheumatology Clinical Trials; QoL, quality of life; RCT, randomized controlled trial; SLR, systematic literature review; TNF, tumor necrosis factor.

Data Sharing Statement

Data are available from the corresponding author (DP) upon request.
Ethics Approval
None.

Patient Consent for Publication
Not required.

Acknowledgments
Writing support was funded by Celgene and Amgen Inc. and provided by Pharmed - an OPEN Health Company. Editorial support was funded by Celgene and Amgen Inc. and provided by Larry Radican, PhD, MPH, of Peloton Advantage, LLC, an OPEN Health company.

Funding
This study was funded by Celgene Corporation.

Disclosure
Ms Nazareth was employed by Celgene Corporation at the time of the study. Ms Hart, Dr Ronnebaum, and Dr Patel are employees of Pharmed - an OPEN Health Company, which was the recipient of consulting fees from Celgene and Amgen Inc. Dr Mehta was employed by Celgene Corporation at the time of the study. Prof Dr med Kötter has served as a speaker for AbbVie, Boehringer Ingelheim, Chugai, Eli Lilly, Medac, Merck Sharp & Dohme, Novartis, Pfizer, Roche, and Sanofi and reports participation in a clinical trial of apremilast vs placebo (RELIEF) for Celgene. The authors report no other potential conflicts of interest for this work.

References
1. Hatemi G, Christensen R, Bang D, et al. 2018 update of the EULAR recommendations for the management of Behçet’s syndrome. Ann Rheum Dis. 2018;77(6):808–818.
2. Cansu D, Kaşifoglu T, Kortmaz C. Is there any relationship between season/weather and oral ulcer in Behçet’s disease? Eur J Rheumatol. 2014;1(3):89–91. doi:10.5152/eurjrheumatol.2014.031
3. Alpsoy E, Donmez L, Bacani A, Apaydin C, Butun B. Review of the chronological of clinical manifestations in 60 patients with Behçet’s disease. Dermatolog. 2003;207(4):354–356. doi:10.1159/000074113
4. Davatchi F, Shahram F, Chams-Davatchi C, et al. Behçet’s disease in Iran: analysis of 6500 cases. Int J Rheum Dis. 2010;13(4):367–373. doi:10.1111/j.1756-185X.2010.01549.x
5. Gur A, Sarac AJ, Burkan YK, Nas K, Cevik R. Arthropathy, quality of life, depression, and anxiety in Behçet’s disease: relationship between arthritis and these factors. Clin Rheumatol. 2006;25(4):524–531. doi:10.1007/s10067-005-0100-6
6. Krause L, Köhler A-K, Altenburg A, et al. Ocular involvement in Adhamatiades-Behçet’s disease in Berlin, Germany. Graefes Arch Clin Exp Ophthalmol. 2009;247(5):661–666. doi:10.1007/s00417-008-0983-4
7. Kural-Sevahri E, Fresko I, Seyahi N, et al. The long-term mortality and morbidity of Behçet’s syndrome: a 2-decade outcome survey of 387 patients followed at a dedicated center. Medicine. 2003;82(1):60–76. doi:10.1097/00005792-200301000-00006
8. Solan-Morales O, Mehta S, Ronnebaum S, et al. A systematic literature review of healthcare burden in patients with Behçet’s disease. Poster presented at: the European Congress of Rheumatology EULAR 2019; 12–15 June 2019; Madrid, Spain. 2019.
9. International Study Group for Behçet’s Disease. Criteria for diagnosis of Behçet’s disease. Lancet (London, England). 1990;335(8697):1078–1080.
10. Akkoç N. Update on the epidemiology, risk factors and disease outcomes of Behçet’s disease. Best Pract Res Clin Rheumatol. 2018;32(2):261–270. doi:10.1016/j.berh.2018.08.010
11. Naito M, Suzukamo Y, Wakai K, et al. One-year period prevalence of oral aphthous ulcers and healthcare-related quality of life in patients with Behçet’s disease. Genet Res Int. 2014;2014:930348.
12. Ambrose NL, Haskard DO. Differential diagnosis and management of Behçet syndrome. Nat Rev Rheumatol. 2013;9(2):79–89.
13. Greco A, De Virgilio A, Ralli M, et al. Behçet’s disease: new insights into pathophysiology, clinical features and treatment options. Autoimmun Rev. 2018;17(6):567–575. doi:10.1016/j.autrev.2017.12.006
14. Mumcu G, Ergun T, Inane N, et al. Oral health is impaired in Behçet’s disease and is associated with disease severity. Rheumatol. 2004;43(8):1028–1033. doi:10.1093/rheumatology/keh236
15. Mumcu G, Inanc N, Ergun T, et al. Oral health related quality of life is affected by disease activity in Behçet’s disease. Oral Dis. 2006;12(2):145–151. doi:10.1111/j.1601-0825.2005.01173.x
16. Sensu A, Higgins S, Fortune F. The influence of oral health and psychosocial well-being on clinical outcomes in Behçet’s disease. Rheumatol Int. 2018;38(10):1873–1883. doi:10.1007/s00296-018-4117-y
17. Canpolat O, Yurtseven S. The quality of life in patients with Behçet’s disease. Asian Nurs Res. 2011;5(4):229–235. doi:10.1016/j.anr.2011.12.003
18. Güler T, Çiçli Aslan N, Gariş Y, Dörtbaş F, Aşlıhan A. Disease activity in Turkish patients with Behçet’s disease: association with fatigue, psychological status and quality of life. Türkiye Klinikleri J Med Sci. 2017;37(3):130–135. doi:10.5336/medsci.2017-56200
19. Melikoglu M, Melikoglu MA. What affects the quality of life in patients with Behçet’s disease? Acta Reumatol Port. 2014;39(1):46–53.
20. Bernabé E, Mercenes W, Mather J, Phillips C, Fortune F. Impact of Behçet’s syndrome on health-related quality of life: influence of the type and number of symptoms. Rheumatol. 2010;49(11):2165–2171. doi:10.1093/rheumatology/keq251
21. Lecesse P, Ozguler Y, Christensen R, et al. Management of skin, mucosa and joint involvement of Behcet’s syndrome: a systematic review for update of the EULAR recommendations for the management of Behçet’s syndrome. Semin Arthritis Rheum. 2019;48(4):752–762. doi:10.1016/j.semarthritis.2018.05.008
22. Otezla® tablets [updated]; 2019. Available from: https://otezla-japan.jp/dr/. Accessed November 11, 2020.
23. Otezla® (apremilast) [package insert]. Summit, NJ: Amgen Inc.; June 2020.
24. Otezla (apremilast) [SmPC]. Breda, Netherlands: Amgen Europe B. V.; 2020.
25. Nava F, Ghilotti F, Maggi L, et al. Biologics, colchicine, corticosteroids, immunosuppressants and interferon-alpha for neuro-Behçet’s syndrome. Cochrane Database Syst Rev. 2014;12:CD010729.
26. Waljee AK, Rogers MA, Lin P, et al. Short term use of oral corticosteroids and related harms among adults in the United States: population based cohort study. BMJ. 2017;357:j1415. doi:10.1136/bmj.j1415
27. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on indirect treatment comparisons good research practices; part 1. Value Health. 2011;14(4):417–428.
28. Facey KM. As health technology assessment evolves so must its approach to patient involvement. J Comp Eff Res. 2019;8(8):549–554. doi:10.2217/cer-2019-0039
29. Es-Skalii IJ, Spoors J. Analysis of indirect treatment comparisons in national health technology assessments and requirements for industry submissions. J Comp Eff Res. 2018;7(4):397–409. doi:10.2217/cer-2017-0092
30. Laws A, Kendall R, Hawkins N. A comparison of national network for meta-analysis. Value Health. 2014;17(5):642–654. doi:10.1016/j.jval.2014.06.001
31. Higgins JP, Green S Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from: www.handbook.cochrane.org.
32. Cope S, Zhang J, Saletan S, Smiechowski B, Jansen JP, SchmID P. A process for assessing the feasibility of a network meta-analysis: a case study of everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. BMC Med. 2014;12:93. doi:10.1186/1741-7015-12-93
33. Ergun T, Gurbuz O, Yurdakul S, Hamuryudan V, Bekiroglu N, Yazici H. Topical cyclosporine-A for treatment of oral ulcers of Behçet's syndrome. Int J Dermatol. 1997;36(9):720. doi:10.1111/j.1365-4326.1997.tb03133.x
34. Hatemi G, Melikoglu M, Tunc R, et al. Apremilast for Behçet's syndrome - a phase 2, placebo-controlled study. N Engl J Med. 2015;372(16):1510–1518. doi:10.1056/NEJMoA1408684
35. Hatemi G, Mahr A, Ishigatsu Y, et al. Trial of apremilast for oral ulcers in Behçet's syndrome. N Engl J Med. 2019;381(20):1918–1928. doi:10.1056/NEJMoA1816594
36. Aktulga E, Altac M, Mufuoglu A, et al. A double blind study of colchicine in Behçet's disease. Haematologica. 1989;63(5):399–402.
37. Davatchi F, Sadeghi Abdollahi B, Tehrani Banihashemi A, et al. Colchicine versus placebo in Behçet's disease: randomized, double-blind, controlled crossover trial. Mod Rheumatol. 2009;19(5):542–549. doi:10.3109/10494041.2009.381143
38. Masuda K, Nakajima A, Urayama A, Nakae K, Kogure M, Inaba G. Double-masked trial of cyclosporin versus colchicine and long-term open study of cyclosporin in Behçet's disease. Lancet (London, England). 1989;333(8647):1093–1096. doi:10.1016/S0140-6736(89)92381-7
39. Calguneri M, Ertendi I, Kiraz S, Erman M, Celik I. Effect of prophylactic benzathine penicillin on mucocutaneous symptoms of Behçet's disease. Dermatology. 1996;192(2):125–128. doi:10.1159/000146336
40. Yurdakul S, Mat C, Tuzun Y, et al. A double-blind trial of colchicine in Behçet's syndrome. Arthritis Rheum. 2001;44(11):2686–2692. doi:10.1002/1529-0131(200111)44:11<2686::AID-ART448>3.0.CO;2-H
41. Alpsoy E, Durusoy C, Yilmaz E, et al. Interferon alfa-2a in the treatment of Behçet disease: a randomized placebo-controlled and double-blind study. Arch Dermatol. 2002;138(4):467–471. doi:10.1001/archderm.138.4.467
42. Fani MM, Ebrahimii H, Pourshahidi S, Afkazi E, Shafiee Sarvestani S. Comparing the effect of phenytoin syrup and trimethadione acetone ointment on aphthous ulcers in patients with Behçet's syndrome. Iran Red Crescent Med J. 2012;14(2):73–78.
43. Kilic H, Zeytin HE, Korkmaz C, et al. Low-dose natural human interferon-alpha lozenges in the treatment of Behçet's syndrome. Rheumatology. 2009;48(1):1388–1391. doi:10.1093/rheumatology/kep237
44. Hamuryudan V, Yurdakul S, Rosenkaimer F, Yazici H. Inefficacy of topical alpha interferon in the treatment of oral ulcers of Behçet's syndrome: a randomized, double blind trial. Br J Rheumatol. 1991;30(5):395–396. doi:10.1093/rheumatology/30.5.395-a
45. Hamuryudan V, Mat C, Saip S, et al. Thalidomide in the treatment of the mucocutaneous lesions of the Behçet syndrome. A randomized, double-blind, placebo-controlled trial. Ann Intern Med. 1998;128(6):443–450. doi:10.7326/0003-4819-128-6-199803150-00004
46. Melikoglu M, Fresko I, Mat C, et al. Short-term trial of etanercept in Behçet's disease: a double blind, placebo controlled study. J Rheumatol. 2005;32(1):98–105.
47. Yazici H, Pazarli H, Barnes CG, et al. A controlled trial of azathioprine in Behçet’s syndrome. N Engl J Med. 1990;322(5):281–285. doi:10.1056/NEJM199002013220501
48. Rath A, Salamon V, Peixoto S, et al. A systematic literature review of evidence-based clinical practice for rare diseases: what are the perceived and real barriers for improving the evidence and how can they be overcome? Trials. 2017;18(1):556. doi:10.1186/s13063-017-2287-7
49. Hatemi G, Meara A, Oguzler Y, et al. Developing a core set of outcome measures for Behçet disease: report from OMERACT 2016. J Rheumatol. 2017;44(11):1750–1753. doi:10.3899/jrheum.161352
50. Hatemi G, Meara A, Oguzler Y, et al. THU0303 The OMERACT core domain set for clinical trials in Behçet’s syndrome. Ann Rheum Dis. 2019;78(Suppl 2):431.
51. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. Res Synth Methods. 2017;8(1):79–91. doi:10.1002/1756-185X.13109