NHS England

Evidence review: Abatacept for refractory idiopathic inflammatory myopathies

NHS England URN: 1925

NHS England

Evidence review: Abatacept for refractory idiopathic inflammatory myopathies

Drafted: June 2020

Updated: Month Year

Prepared by: Solutions for Public Health (SPH) on behalf of NHS England Specialised Commissioning

Contents

1. Introduction

This review examines the clinical effectiveness, safety and cost effectiveness of abatacept compared to intravenous immunoglobulin (IVIg) or intravenous cyclophosphamide in patients with refractory idiopathic inflammatory myopathies (IIMs) (dermatomyositis, polymyositis and juvenile dermatomyositis and excluding inclusion body myositis).

2. Executive summary of the review

One paper was included in this review (Tjärnlund et al 2018). This was a multi-centre randomised controlled trial (RCT) comparing immediate treatment with abatacept (n=11) with delayed treatment with abatacept (n=9). The trial was conducted in the UK, Sweden, and the Czech Republic. All patients received abatacept for six months, with the delayed treatment group starting treatment three months after the immediate treatment group. Treatment received by the delayed treatment group in the three months prior to their receipt of abatacept is not described by the study authors but this period is referred to as standard treatment in this review. Results after six months of abatacept were pooled for all patients and compared to baseline scores.

In children and adults with refractory IIMs (including intolerance to or an inadequate response to rituximab in patients with myositis-specific and myositis-associated antibodies), what is the clinical effectiveness of abatacept compared with standard treatment?

Critical Outcomes. The critical outcomes for decision making are total improvement score, muscle strength, disability/ function, physician global activity score, patient global activity score and muscle enzymes. Certainty in the quality of the evidence for these critical outcomes was very low using modified GRADE.

Total improvement score[1]

For people with refractory IIMs, a statistically significant benefit of abatacept compared with standard treatment in terms of total improvement score was reported at three months in one RCT (n=20). The median (interquartile range (IQR)) was 28.8 (IQR 15 to 37.5) for abatacept and 5.0 (IQR 0 to 12.5) for standard treatment (p=0.03). The percentage of patients achieving a minimal total improvement (≥20 points) at three months was 60% for patients receiving abatacept and 20% for patients receiving standard treatment (statistical comparison between groups not reported). The total improvement score after six months of abatacept was only presented graphically. However, the proportion of patients achieving the minimum total improvement after receiving six months of abatacept was 90% in the immediate treatment group and 40% in the delayed treatment group (statistical comparison between groups not reported). The proportion of patients achieving a moderate improvement (≥40 points) was 40% in the immediate treatment group and 10% in the delayed treatment group (time period not reported). No patients achieved a major improvement (≥60 points). The certainty of the evidence was very low.

[1] Scored from 0 to 100 with higher scores indicating more improvement

Muscle strength[2]
For people with refractory IIMs, a statistically significant benefit of abatacept compared with standard treatment in terms of muscle strength was reported at three months in one RCT (n=20). The mean difference (standard deviation (SD)) between baseline and three months was 2.5 (SD 4.7) for patients receiving abatacept and -4.9 (SD 9.1) for standard treatment (p=0.038). A statistically significant benefit in muscle strength after six months treatment with abatacept (median 74, IQR 68.5 to 78) compared to baseline (median 70, IQR 64 to 73) (p=0.047) was also reported by the same RCT. The certainty of the evidence was very low.

Disability/ function[3]
For people with refractory IIMs, there was no statistically significant benefit of abatacept compared with standard treatment in terms of disability/ function at three months in one RCT (n=20). The mean difference between baseline and three months was -0.2 (SD 0.4) for patients receiving abatacept and -0.0006 (SD 0.2) for standard treatment (p=0.296). There was also no statistically significant benefit in disability/ function after six months treatment with abatacept (median 1.00, IQR 0.38 to 1.44) compared to baseline (median 1.00, IQR 0.63 to 1.81) (p=0.427) in the same RCT. The certainty of the evidence was very low.

Physician global activity score[4]
For people with refractory IIMs, there was no statistically significant benefit of abatacept compared with standard treatment in terms of physician global activity score at three months in one RCT (n=20). The mean difference between baseline and three months was -10.8 (SD 13.7) for patients receiving abatacept and 0.3 (SD 13.8) for standard treatment (p=0.096). There was also no statistically significant benefit in physician global activity score after six months treatment with abatacept (median 20.0, IQR 10.0 to 40.5) compared to baseline (median 30.0, IQR 22.5 to 46.0) (p=0.063) in the same RCT. The certainty of the evidence was very low.

Patient global activity score[4]
For people with refractory IIMs, there was no statistically significant benefit of abatacept compared with standard treatment in terms of patient global activity score at three months in one RCT (n=20). The mean difference between baseline and three months was -1.1 (SD 15.8) for patients receiving abatacept and 2.1 (SD 18.5) for standard treatment (p=0.434). There was also no statistically significant benefit in patient global activity score after six months treatment with abatacept (median 29.0, IQR 13.5 to 69.5) compared to baseline (median 42.0, IQR 24.5 to 74.0) (p=0.458) in the same RCT. The certainty of the evidence was very low.

Muscle enzymes[5]
For people with refractory IIMs, there was no statistically significant benefit of abatacept compared with standard treatment in terms of muscle enzymes (creatine kinase or lactate

[2] Assessed by the Manual Muscle Test (0-80) with higher scores indicating greater muscle strength
[3] Assessed by the Health Assessment Questionnaire (0-3) with higher scores indicating greater disability
[4] Scored on a Visual Analogue Scale (0-100) with higher scores indicating a higher level of disease activity
[5] Measured in µcat/L with higher scores indicating greater injury/ tissue damage

dehydrogenase) at three months in one RCT (n=20). For creatine kinase mean difference between baseline and three months was -3.2 (SD 10.9) for patients receiving abatacept and 13.5 (SD 18.7) for standard treatment (p=0.094). For lactate dehydrogenase mean difference between baseline and three months was -0.3 (SD 1.3) for patients receiving abatacept and 1.9 (SD 3.3) for standard treatment (p=0.065). There was also no statistically significant benefit after six months treatment with abatacept compared to baseline for creatine kinase levels (median 2.8, IQR 1.5 to 7.1 vs median 3.0, IQR 2.0 to 30.4, p=0.438) or for lactate dehydrogenase levels (median 4.0, IQR 3.1 to 4.6 vs median 4.5, IQR 3.8 to 7.1, p=0.299) in the same RCT. The certainty of the evidence was very low.

Important Outcomes. Outcomes important to decision making are disease activity, quality of life and number of relapses. Certainty in the quality of the evidence for disease activity and quality of life was very low using modified GRADE. No evidence was identified for number of relapses.

Disease activity[6]
For people with refractory IIMs, a statistically significant benefit of abatacept compared with standard treatment in terms of disease activity was identified at three months in one RCT (n=20). The mean difference between baseline and three months was -12.7 (SD 14.5) for patients receiving abatacept and 1.4 (SD 12.2) for standard treatment (p=0.0353). However, there was no statistically significant benefit in disease activity after six months treatment with abatacept (median 23.0, IQR 9.0 to 36.0) compared to baseline (median 30.0, IQR 15.5 to 43.5) (p=0.1958) in the same RCT. The certainty of the evidence was very low.

Quality of life[7]
For people with refractory IIMs, there was a statistically significant benefit in quality of life after six months treatment with abatacept (median 37, IQR 24 to 45) compared to baseline (median 31, IQR 24 to 35) (p=0.005) in one RCT (n=20). The certainty of the evidence was very low.

Number of relapses
No evidence was available for number of relapses.

In children and adults with refractory IIMs, what is the safety of abatacept compared with current standard treatment?

Safety outcomes important to decision making are adverse events. Certainty in the quality of the evidence for safety outcomes was very low using modified GRADE.

Adverse events
For people with refractory IMMs, eight adverse events related to treatment with abatacept were reported after six months treatment with abatacept. All of these adverse events were

[6] Assessed by extra-muscular disease activity, scored on a Visual Analogue Scale (0-100) with higher scores indicating a higher level of disease activity
[7] Assessed by the SF-36 physical health component (0-100) with higher scores indicating better quality of life

moderate (n=4) or mild (n=4). Further details for the abatacept related adverse events were not reported.

In children and adults with refractory IIMs, what is the cost effectiveness of abatacept compared with current standard treatment?

No evidence was identified on the cost effectiveness of abatacept compared with current standard treatment.

From the evidence selected is there any data to suggest that there are subgroups of patients with refractory IIMs that would benefit from treatment with abatacept more than others?

No evidence was identified suggesting that any subgroups of patients would benefit from treatment more.

Limitations. The evidence base was limited to a single small, underpowered study with a lack of patient blinding potentially affecting the reliability of self-reported outcomes.

Conclusions. Additional benefit was found for abatacept compared to standard treatment for two critical (total improvement score and muscle strength) and one important outcome (disease activity). Evidence for this comparison was only available over a three month timeframe and there was a lack of information on what standard treatment involved. These improved outcomes did not translate to improvement in other critical and important outcomes including disability/ function, physician and patient global assessment scores or muscle enzymes. Treatment with abatacept was not associated with any serious or severe adverse events.

3. Methodology

Review questions

The review questions for this evidence review are:

1. In children and adults with refractory IIMs (including intolerance to or an inadequate response to rituximab in patients with myositis-specific and myositis-associated antibodies), what is the clinical effectiveness of abatacept compared with standard treatment?
2. In children and adults with refractory IIMs, what is the safety of abatacept compared with current standard treatment?
3. In children and adults with refractory IIMs, what is the cost effectiveness of abatacept compared with current standard treatment?
4. From the evidence selected is there any data to suggest that there are subgroups of patients with refractory IIMs that would benefit from treatment with abatacept more than others?

See Appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2019).

The searches for evidence were informed by the PICO document and were conducted on 30th January 2020 and updated on 11th May 2020.

See Appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO framework. Full text references of potentially relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See Appendix C for evidence selection details and Appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were quality appraised using a checklist appropriate to the study design.

See Appendices E and F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE.

See Appendix G for GRADE Profiles.

4. Summary of included studies

One paper was identified for inclusion (Tjärnlund et al 2018). Table 1 provides a summary of this included study and full details are given in Appendix E.

This RCT compared immediate treatment with abatacept (n=11) with delayed treatment with abatacept (n=9). All patients received abatacept for six months, with the delayed treatment group starting treatment three months after the immediate treatment group.

Table 1 Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Tjärnlund et al 2018 RCT Sweden, UK, Czech Republic	Patients with refractory[a] dermatomyositis or polymyositis n=20 Immediate treatment (n=11) Delayed treatment (n=9) No subgroups reported	**Intervention** Immediate treatment with abatacept **Comparison** Delayed treatment with abatacept i.e. standard treatment for 3 months, before starting abatacept In both groups, patients received abatacept intravenously for 6 months (7 infusions). Dosage based on body weight was 500mg for patients <60kg, 750mg for patients 60-100kg and 1,000mg for patients >100kg Concomitant methotrexate was allowed, with stable doses ≥ 1 months prior to inclusion in the study	**Critical Outcomes** • Total improvement score at month 3 and month 9[b] • Muscle strength (assessed by MMT) at month 3 and after 6 months of abatacept[c] • Disability/ function (assessed by HAQ) at month 3 and after 6 months of abatacept • Physician global disease activity score at month 3 and after 6 months of abatacept • Patient global disease activity score at month 3 and after 6 months of abatacept • Muscle enzymes (creatine, lactate dehydrogenase) at month 3 and after 6 months of abatacept **Important outcomes** • Disease activity (extra-muscular global assessment) at month 3 and after 6 months of abatacept • Health-related quality of life after 6 months of abatacept **Safety** Adverse events after 6 months of abatacept

Abbreviations: HAQ **-** health assessment questionnaire; kg – kilogram; mg – milligram; MMT **-** manual muscle test; RCT – randomised controlled trial; UK – United Kingdom

[a] Patients had active disease after treatment with glucocorticoids (≥0.5 mg/kg/day for ≥1 month), in combination with at least one other immunosuppressive drug, methotrexate (minimum dose 15 mg/week) or azathioprine (minimum dose 100 mg/day) for at least 3 months

[b] At month 3, the immediate treatment group had received abatacept for 3 months and the delayed treatment group were receiving standard treatment and had not yet received abatacept. By month 9 both groups had completed 6 months of abatacept and the immediate treatment group had received an additional 3 months follow-up after completing their abatacept treatment

[c] Results after 6 months of abatacept were pooled for all patients and compared to baseline scores

5. Results

In children and adults with refractory IIMs (including intolerance to or an inadequate response to rituximab in patients with myositis-specific and myositis-associated antibodies), what is the clinical effectiveness and safety of abatacept compared with standard treatment?

Outcome	Evidence statement
Clinical Effectiveness	
Critical outcomes	
Total improvement score **Certainty of evidence:** Very low	Total improvement score is a composite measure[a] and is relevant to patients because it provides an overview of their improvement across 6 core measures that can relate to functionality and quality of life. 1 study (RCT) provided evidence relating to total improvement score measured at 3 months. At this timepoint, the comparison was abatacept (immediate treatment group) compared to standard treatment (delayed treatment group). This study also provided evidence relating to minimum total improvement score after six months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed a *statistically significantly higher* median (IQR) total improvement score at 3 months favouring abatacept (28.8, IQR 15 to 37.5) compared to standard treatment (5.0, IQR 0 to 12.5) (p=0.03). At 3 months the percentage of patients achieving a minimal total improvement (≥20 points) was 60% for patients receiving abatacept and 20% for patients receiving standard treatment. Total improvement score after 6 months of abatacept was only presented graphically. However, 90% of patients in the immediate treatment group achieved a minimum total improvement after 6 months of abatacept. This was 40% in the delayed treatment group (statistical comparison between groups not reported). The study reported moderate improvement (≥40 points) for 40% of the immediate treatment group and 10% of the delayed treatment group but did not report a time period for this assessment. No patients achieved a major improvement (≥60 points). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does improve total improvement score at 3 months.**
Muscle strength **Certainty of evidence:** Very low	Muscle strength is relevant to patients because it can relate to mobility and independence and can impact quality of life. 1 study (RCT) provided evidence relating to muscle strength (assessed by MMT-8[b]) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to muscle strength measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed a *statistically significantly higher* mean difference (SD) between baseline and month 3 for muscle

Outcome	Evidence statement
	strength favouring abatacept (2.5, SD 4.7) compared to standard treatment (-4.9, SD 9.1) (p=0.038). There was also a *statistically significant improvement* in median (IQR) muscle strength favouring 6 months treatment with abatacept (74, IQR 68.5 to 78) compared to baseline (70, IQR 64 to 73) (p=0.047). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does improve muscle strength at 3 months and that compared to baseline, abatacept does improve muscle strength after 6 months.**
Disability/ function **Certainty of evidence:** Very low	Disability/ function is relevant to patients because it can relate to independence and quality of life and identify unknown and unquantified benefits and risks of the intervention. 1 study (RCT) provided evidence relating to disability/ function (assessed by HAQ[c]) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to disability/ function measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed *no statistically significant difference* in mean difference (SD) between baseline and month 3 for disability/ function for abatacept (-0.2, SD 0.4) compared to standard treatment (-0.0006, SD 0.2) (p=0.296). There was also *no statistically significant difference* in median (IQR) disability/ function after 6 months treatment with abatacept (1.00, IQR 0.38 to 1.44) compared to baseline (1.00, IQR 0.63 to 1.81) (p=0.427). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does not improve disability/ function at 3 months and that compared to baseline, abatacept does not improve disability/ function after 6 months.**
Physician global activity score **Certainty of evidence:** Very low	Physician global activity score is relevant to patients because it is an assessment of disease activity and can relate to quality of life. 1 study (RCT) provided evidence relating to physician global activity score (assessed by VAS[d]) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to physician global activity score measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed *no statistically significant difference* in mean difference (SD) between baseline and month 3 for physician global activity score for abatacept (-10.8, SD 13.7) compared to standard treatment (0.3, SD 13.8) (p=0.096). There was also *no statistically significant difference* in median (IQR) physician global activity score after 6 months treatment with abatacept (20.0, IQR 10.0 to 40.5) compared to baseline (30.0, IQR 22.5 to 46.0) (p=0.063). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does not improve physician global activity score at 3 months and that compared to baseline,**

Outcome	Evidence statement
	abatacept does not improve physician global activity score after 6 months.
Patient global activity score **Certainty of evidence:** Very low	Patient global activity score is relevant to patients because it is an assessment of disease activity and can relate to quality of life. 1 study (RCT) provided evidence relating to patient global activity score (assessed by VAS[d]) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to patient global activity score measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed *no statistically significant difference* in mean difference (SD) between baseline and month 3 for patient global activity score for abatacept (-1.1, SD 15.8) compared to standard treatment (2.1, SD 18.5) (p=0.434). There was also *no statistically significant difference* in median (IQR) patient global activity score after 6 months treatment with abatacept (29.0, IQR 13.5 to 69.5) compared to baseline (42.0, IQR 24.5 to 74.0) (p=0.458). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does not improve patient global activity score at 3 months and that compared to baseline, abatacept does not improve patient global activity score after 6 months.**
Muscle enzymes **Certainty of evidence:** Very low	Muscle enzymes are relevant to patients because they are an indicator of muscle injury or disease. Higher creatine kinase levels indicate muscle injury. Higher lactate dehydrogenase levels indicate tissue damage. 1 study (RCT) provided evidence relating to muscle enzymes (µcat/L) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to muscle enzymes measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed *no statistically significant difference* in mean difference (SD) between baseline and month 3 for creatine kinase levels for abatacept (-3.2, SD 10.9) compared to standard treatment (13.5, SD 18.7) (p=0.094). There was also *no statistically significant difference* in mean difference (SD) between baseline and month 3 for lactate dehydrogenase levels for abatacept (-0.3, SD 1.3) compared to standard treatment (1.9, SD 3.3) (p=0.065). There was *no statistically significant difference* in median (IQR) creatine kinase levels after 6 months treatment with abatacept (2.8, IQR 1.5 to 7.1) compared to baseline (3.0, IQR 2.0 to 30.4) (p=0.438). There was also *no statistically significant difference* in median (IQR) lactate dehydrogenase levels after 6 months treatment with abatacept (4.0, IQR 3.1 to 4.6) compared to baseline (4.5, IQR 3.8 to 7.1) (p=0.299). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does not improve muscle enzymes**

Outcome	Evidence statement
	at 3 months and that compared to baseline, abatacept does not improve muscle enzymes after 6 months.
Important outcomes	
Disease activity **Certainty of evidence:** Very low	Disease activity is relevant to patients because it can relate to quality of life. 1 study (RCT) provided evidence relating to disease activity (assessed by extra-muscular global assessment, VAS[d]) measured at 3 months. At this timepoint, the comparison was abatacept compared to standard treatment. This study also provided evidence relating to disease activity measured at baseline and after 6 months treatment with abatacept. 1 RCT (Tjärnlund et al 2018) (n=20) showed a *statistically significantly higher* mean difference (SD) between baseline and month 3 for disease activity favouring abatacept (-12.7, SD 14.5) compared to standard treatment (1.4, SD 12.2) (p=0.0353). There was *no statistically significant difference* in median (IQR) disease activity after 6 months treatment with abatacept (23.0, IQR 9.0 to 36.0) compared to baseline (30.0, IQR 15.5 to 43.5) (p=0.1958). **(VERY LOW)** **This study provides very low certainty evidence that compared to standard treatment, abatacept does improve disease activity at 3 months. However, compared to baseline, abatacept does not improve disease activity after 6 months.**
Quality of life **Certainty of evidence:** Very low	Quality of life is relevant to patients because it provides an indication of an individual's general health and ability to participate in and enjoy life events. Analysis from 1 study (RCT) provided evidence relating to quality of life (assessed by SF-36 physical health component[e]) measured at baseline and after 6 months treatment with abatacept. 1 study (Tjärnlund et al 2018) (n=20) showed a *statistically significant improvement* in median (IQR) quality of life favouring 6 months treatment with abatacept (37, IQR 24 to 45) compared to baseline (31, IQR 24 to 35) (p=0.005). **(VERY LOW)** **This study provides very low certainty evidence that compared to baseline, abatacept does improve quality of life after 6 months.**
Number of relapses **Certainty of evidence:** Not applicable	Number of relapses is relevant to patients because it relates to the return of the condition and can negatively impact quality of life. **No evidence was identified for this outcome.**
Safety	
Adverse events	Adverse events are relevant to patients because they can result in death or be life threatening and can result in persistent or significant

Outcome	Evidence statement
Certainty of evidence: Very low	disability or incapacity. They can also require hospitalisation, prolong existing hospitalisation or require additional treatment. Analysis from 1 study (RCT) provided evidence relating to adverse events after 6 months treatment with abatacept. No comparative evidence was provided for this outcome. 1 study (Tjärnlund et al 2018) (n=20) reported 8 adverse events that were considered related to treatment with abatacept. None of these adverse events were described as serious or severe. Four of the adverse events reported were moderate and four were mild. Further details for the abatacept related adverse events were not reported. **This study provides very low certainty evidence that 6 months treatment with abatacept is associated with a small number of moderate or mild adverse events.**

[a] Total improvement score (0-100) is a consensus-based response score (the EULAR response criteria) that includes 6 core set measures (physician, patient and extra-muscular global activity, muscle strength, Health Assessment Questionnaire and muscle enzyme levels). Higher scores indicate more improvement. There are agreed thresholds for minimal, moderate and major improvement (Aggarwal et al 2017)
[b] The Manual Muscle Test is scored from 0 to 80 with higher scores indicating greater muscle strength
[c] The Health Assessment Questionnaire is scored from 0 to 3 with higher scores indicating greater disability
[d] Visual Analogue Scales are 0 to 100mm with higher scores indicating a higher level of disease activity
[e] The SF-36 is scored from 0 to 100 with higher scores indicating better quality of life. Only the physical health component score was reported

Abbreviations: HAQ – health assessment questionnaire; IQR – interquartile range; L – litre; MMT – manual muscle test; RCT – randomised controlled trial; SD – standard deviation; VAS – visual analogue scale; µcat - microcat

From the evidence selected is there any data to suggest that there are subgroups of patients with refractory IIMs that would benefit from treatment with abatacept more than others?

Outcome	Evidence statement
Subgroups	**No evidence was identified suggesting that any subgroups of patients would benefit from treatment more**

In children and adults with refractory IIMs, what is the cost effectiveness of abatacept compared with current standard treatment?

Outcome	Evidence statement
Cost effectiveness	**No evidence was identified for cost effectiveness**

6. Discussion

This review considers the evidence of clinical effectiveness and safety of abatacept in patients with refractory IIMs, compared to standard treatment. The critical outcomes of interest are total improvement score, muscle strength, disability/ function, physician global activity score, patient global activity score and muscle enzymes. Important outcomes are disease activity, quality of life and number of relapses.

Evidence was available from one RCT. This study was at high risk of bias. Certainty in the comparative evidence for critical and important outcomes was very low using modified GRADE.

This small pilot RCT included 20 adult patients with this rare condition. The use of delayed treatment as the comparator was reasonable for a pilot RCT with a population of patients with refractory disease. As a result of the design, the period of time where one group (immediate treatment) was receiving abatacept and the other group (delayed treatment) was receiving standard treatment was three months. This provides some comparative data on the effectiveness of abatacept. However, no information was provided on treatments received during the standard treatment period for the delayed treatment group. The study authors reported that approximately one third of patients had received prior intravenous immunoglobulin (IVIG) at baseline, but it is not clear if any patients received IVIG during the study period. The sample size was based on feasibility rather than a power calculation. It is not clear that the study was sufficiently powered to demonstrate efficacy.

As all patients received six months treatment with abatacept at some point, the study authors also pooled the six month data for all patients and compared this to baseline scores. Although the data was generated as part of a controlled trial, this was not a comparison of patients randomised to different intervention groups.

Ten patients were initially randomised to each group. However, one patient was changed from delayed to immediate treatment due to aggressive progress of weakness. This introduces a risk of bias as patients were not analysed according to the original randomisation. Three patients discontinued during the trial with reasons for discontinuation provided. This included one patient from the immediate treatment group and two patients from the delayed treatment group. One of these three patients was not included in the analysis as they withdrew, due to disease progression, before any efficacy assessments were performed. The comparison of groups at baseline was based on the 11 patients in the immediate treatment group and nine patients in the delayed treatment group. Patients in the immediate treatment group were older, but the groups were similar on treatment and health measures. The confidence intervals were wide around these measures for both groups, reflecting the small sample size.

This was an open label trial, however outcome assessors were blinded to group. The objective nature of some of the outcomes, e.g. muscle enzymes, and the blinding of outcome assessors will have minimised the risk of bias for some outcomes. However, other outcomes, such as quality of life were self-reported and may have been affected by patient's knowledge of their treatment group.

Outcomes at three months, when the immediate treatment group had received abatacept for three months and the delayed treatment group were receiving standard treatment, were

reported as mean difference from baseline. Although statistically significant differences were reported for some outcomes (total improvement score, muscle strength and disease activity), the clinical importance of the results is unclear as the improved outcomes did not translate to improvement in other outcomes including disability/ function, physician and patient global assessment scores or muscle enzymes. Median scores were reported for baseline and after six months treatment with abatacept for some outcomes. However, this was pooled data for all patients rather than a comparison between treatment groups.

No evidence on number of relapses was identified. Quality of life and safety outcomes were only reported as pooled data following six months treatment with abatacept. The SF-36 was used to assess quality of life. However, only the physical health component was reported representing a partial assessment of quality of life. A full quality of life assessment using the SF-36 would also include mental health component and total scores.

Adverse events were reported after six months of abatacept. The number and severity of those thought to be associated with abatacept was reported but the type of adverse events associated with abatacept was not specified.

The trial was conducted at centres in three countries, one of which was the UK. It is not clear how many patients were treated in each centre. No statement was provided regarding the training or co-ordination of assessment across centres. However, outcomes were assessed using published assessment measures which may have limited the risk of bias.

The trial received some funding from industry, however the authors stated that the study was carried out independently from the pharmaceutical company.

7. Conclusion

Clinical and safety outcomes comparing immediate treatment to delayed treatment with abatacept was available from one small pilot RCT. This reported outcomes at three months, at which stage the immediate treatment group had received abatacept for three months and the delayed treatment group were receiving standard treatment. Outcomes following six months of abatacept were also reported in a pooled analysis of all patients compared to baseline scores. Evidence certainty for all outcomes was very low.

Very low certainty evidence suggests that three month's treatment with abatacept is associated with a statistically significant improvement in the critical outcomes of total improvement score and muscle strength and the important outcome of disease activity (assessed by extra-muscular global assessment) compared to standard treatment. Very low certainty evidence also suggests that six months treatment with abatacept is associated with a statistically significant improvement in the critical outcome of muscle strength and the important outcome of quality of life compared with baseline scores. No statistically significant differences were observed for the critical outcomes of disability/ function, physician global activity score, patient global activity score, or muscle enzymes. No evidence was identified on number of relapses.

Treatment with abatacept was not associated with any serious or severe adverse events. Eight moderate or mild adverse events were reported.

No evidence on cost effectiveness was identified.

There was no evidence to identify subgroups of patients who might benefit more from treatment with abatacept.

The limited evidence available from one small, low quality study is insufficient to draw reliable conclusions about the efficacy of abatacept in patients with refractory IIMs.

Appendix A PICO Document

The review questions for this evidence review are:

1. In children and adults with refractory IIMs (including intolerance to or an inadequate response to rituximab in patients with myositis-specific and myositis-associated antibodies), what is the clinical effectiveness of abatacept compared with standard treatment?
2. In children and adults with refractory IIMs, what is the safety of abatacept compared with current standard treatment?
3. In children and adults with refractory IIMs, what is the cost effectiveness of abatacept compared with current standard treatment?
4. From the evidence selected is there any data to suggest that there are subgroups of patients with refractory IIMs that would benefit from treatment with abatacept more than others?

PICO Table

P –Population and Indication	Patients with refractory IIMs (dermatomyositis, polymyositis and juvenile dermatomyositis and excluding inclusion body myositis). [Refractory idiopathic inflammatory myopathy is defined as: the intolerance to or an inadequate response to glucocorticoids and at least two other conventional immunosuppressive or immunomodulatory agents (1st line treatment), and rituximab for patients with myositis-specific or myositis-associated antibodies. Conventional immunosuppressive or immunomodulatory agents may include, methotrexate, azathioprine, cyclosporine, mycophenolate mofetil, leflunomide, tacrolimus.] Subgroups of interest include patients with severe skin ulceration and patients with skin or muscle calcinosis. Other subgroups of interest are age of onset (juvenile onset versus adult onset) and presence of autoantibodies relevant to myositis.
I – Intervention	Intravenous or subcutaneous abatacept given either as monotherapy or combination therapy.
C – Comparator(s)	• Intravenous immunoglobulin (IVIg) • Intravenous cyclophosphamide [IVIg is currently used as third line treatment after rituximab treatment. Intravenous cyclophosphamide can be fourth line treatment or used earlier in the pathway in patients with severe life-threatening disease.]
O – Outcomes	**Critical to decision-making:** • Total improvement score according to EULAR response criteria[8] at 6 months (0-100 points). • Muscle strength e.g. manual muscle test (MMT)

[8] For EULAR response criteria for adult dermatomyositis thresholds for minimal (≥20 points), moderate (≥40 points), major improvement (≥60 points) see reference Aggarwal et al 2017. For juvenile dermatomyositis score thresholds for minimal (≥30 points), moderate (≥45), and major improvement (≥70) see reference Rider et al 2017

	• Disability/function measured using validated health questionnaires such as Health Assessment Questionnaire (HAQ) or Childhood Health Assessment Questionnaire (CHAQ)
	• Physician/parent global activity score
	• Muscle enzymes (creatinine kinase and lactate dehydrogenase)
	
	Important to decision-making:
	• Disease activity
	  • Skin involvement assessed by Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI) in adults
	  • Disease Activity Score (DAS) (DASmuscle, DASskin DASmajor organ) in juvenile dermatomyositis
	  • Global extra-muscle disease severity score
	  • Lung function measured as Forced Vital Capacity (FVC) for patients with interstitial lung disease
	• Quality of life measured using a validated questionnaire such as Dermatology Life Quality Index (DLQI) or Hospital Anxiety and Depression Scale (HADS)
	• Number of relapses
	
	Safety
	• Adverse events
	
	Cost effectiveness

Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher-level quality evidence is found, case series can be considered.
Language	English only
Patients	Human studies only
Age	Adults and children
Date limits	2010 to 2020
Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials and guidelines
Study design	Case reports and resource utilisation studies

Appendix B Search strategy

Medline, Embase and Cochrane Library were searched limiting the search to papers published in English Language in the last 10 years. Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials and guidelines, case reports and resource utilisation studies were excluded.

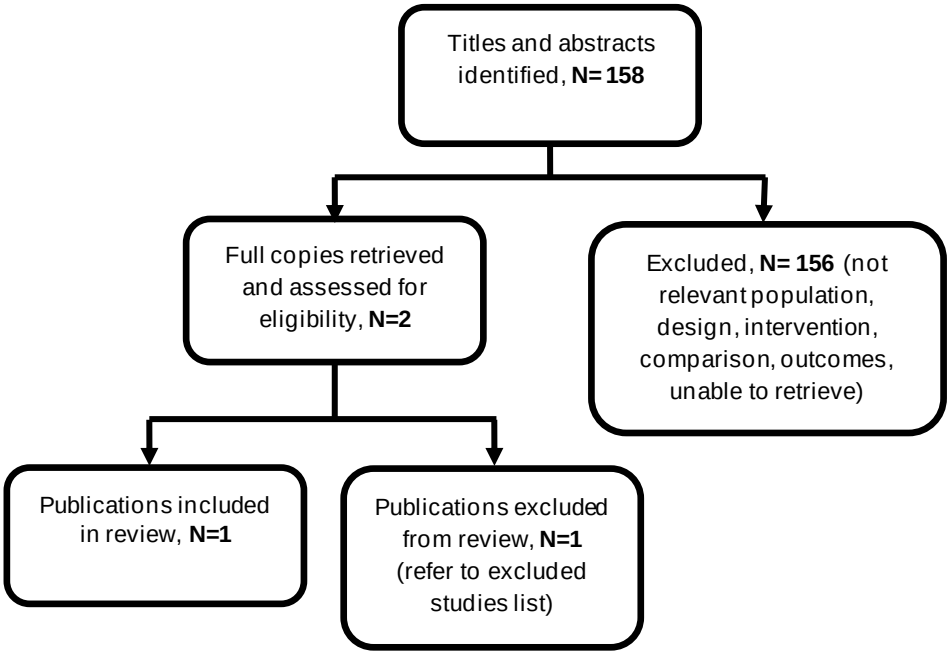
Search dates: 1st January 2010 to 30th January 2020. Search updated 11th May 2020

Embase search

1	myositis/ or exp dermatomyositis/ or polymyositis/
2	(((refractory or recalcitrant) adj5 myopath*) or riim).ti,ab,kw.
3	(myositis or polymyositis or poly-myositis or dermatomyositis or dermato-myositis).ti,ab,kw.
4	1 or 2 or 3
5	Abatacept/
6	(abatacept or orencia or ohrencia or CTLA4-Ig).ti,ab,kw.
7	5 or 6
8	4 and 7
9	(exp animals/ or nonhuman/) not human/
10	8 not 9
11	conference*.pt.
12	10 not 11
13	limit 12 to (english language and yr="2010 -Current")

Appendix C Evidence selection

Figure 1 – Study selection flow diagram



References submitted with Preliminary Policy Proposal

Reference	Paper selection decision and rationale if excluded
Tjärnlund A. Tang Q. Wick C. Dastmalchi M. Mann H. Tomasová Studýnková J. Chura R. Gullick N.J. Salerno R. Rönnelid J. Alexanderson H. Lindroos E. Aggarwal R. Gordon P. Vencovsky J. Lundberg I.E. 2018. Abatacept in the treatment of adult dermatomyositis and polymyositis: a randomised, phase IIb treatment delayed-start trial. *Ann Rheum Dis*, 77, 55-62	Included
Tang Q. Ramsköld D. Krystufkova O. Mann H.F. Wick C. Dastmalchi M. Lakshmikanth T. Chen Y. Mikes J. Alexanderson H. Achour A. Brodin P. Vencovsky J. Lundberg I.E. Malmström V. 2019. Effect of CTLA4-Ig (abatacept) treatment on T cells and B cells in peripheral blood of patients with polymyositis and dermatomyositis. Scand J Immunol, 89, e12732	Excluded. This paper describes impact of treatment on T and B cells in a subset of patients from the Tjärnlund RCT. Clinical and safety outcomes for these patients are reported by Tjärnlund et al. This paper does not include results for the outcomes listed in the PICO
Kerola AM. Kauppi M.J. 2015. Abatacept as a successful therapy for myositis-a case-based review. Clin Rheumatol, 34, 609-12	Excluded. Case report

Appendix D Excluded studies table

Study reference	Reason for exclusion
Tang Q. Ramskòld D. Krystufkova O. Mann H.F. Wick C. Dastmalchi M. Lakshmikanth T. Chen Y. Mikes J. Alexanderson H. Achour A. Brodin P. Vencovsky J. Lundberg I.E. Malmstròm V. 2019. Effect of CTLA4-Ig (abatacept) treatment on T cells and B cells in peripheral blood of patients with polymyositis and dermatomyositis. Scand J Immunol, 89, e12732	This paper describes impact of treatment on T and B cells in a subset of patients from the Tjärnlund RCT. Clinical and safety outcomes for these patients are reported by Tjärnlund et al. This paper does not include outcomes listed in the PICO

Appendix E Evidence Table

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Tjärnlund A. Tang Q. Wick C. Dastmalchi M. Mann H. Tomasová Studýnková J. Chura R. Gullick N.J. Salemo R. Rönnelid J. Alexanderson H. Lindroos E. Aggarwal R. Gordon P. Vencovsky J. Lundberg I.E. 2018. Abatacept in the treatment of adult dermatomyositis and polymyositis: a randomised, phase IIb treatment delayed-start trial. *Ann Rheum Dis*, 77, 55-62 **Study location** Sweden, UK, Czech Republic **Study type** RCT	**Inclusion criteria** Patients aged 18 to 80 years with confirmed diagnosis of dermatomyositis or polymyositis and active disease[9] after treatment with glucocorticoids (≥0.5 mg/kg/day for ≥1 month), in combination with at least one other immunosuppressive drug, methotrexate (minimum dose 15 mg/week) or azathioprine (minimum dose 100 mg/day) for at least 3 months. Concomitant methotrexate was allowed, with stable doses ≥ 1 month prior to inclusion in the study **Exclusion criteria** Patients with other types of inflammatory	**Intervention** Immediate treatment with abatacept **Comparator** Delayed treatment with abatacept i.e. standard treatment for 3 months, before starting abatacept. Treatment received by patients for the 3 months before starting abatacept not described In both groups, patients received abatacept intravenously for 6 months (7 infusions). Dosage based on body weight was 500mg for patients <60kg, 750mg for patients 60-100kg	**Critical outcomes** ***Total improvement score*** *(0-100 with higher scores indicating more improvement)* Significantly higher for immediate treatment (median 28.8, IQR 15 to 37.5) vs delayed treatment (median 5.0, IQR 0 to 12.5) at month 3[10] (p=0.03) Median total improvement score after 6 months of abatacept only presented graphically *Percentage achieving minimal improvement (≥20 points)* Immediate treatment vs delayed treatment[11]: • Month 3: 60% vs 20% • Month 9: 90% vs 40% No statistical test between groups reported *Percentage achieving moderate improvement (≥40 points)* • Immediate treatment group: 40% • Delayed treatment group: 10%	This study was appraised using the JBI checklist for RCTs 1. Yes 2. Yes 3. Yes 4. No 5. No 6. Yes 7. Not applicable 8. Yes 9. No 10. Yes 11. Unclear 12. Unclear 13. Unclear This study was judged to be at high risk of bias. There is some risk of bias from the lack of blinding for patients, impacting self-reported outcomes. However, outcome assessors were blinded to treatment group. No statement was made regarding the training of outcome assessors across the 3 study centres resulting in an

[9] Active disease was defined as persisting or worsening muscle weakness (MMT -8 bilaterally <150) or low endurance measured by Functional Index for myositis <20% of upper value, together with at least one other sign of active disease: elevated (above upper limit of normal) serum levels of muscle enzymes, inflammation in a recent muscle biopsy (<1 month) or on MRI findings consistent with inflammation, or active extra-muscular disease.
[10] At month 3, the immediate treatment group had received abatacept for 3 months and the delayed treatment group had yet to receive abatacept
[11] By month 9 both groups had completed 6 months of abatacept and the immediate treatment group had received an additional 3 months follow-up after completing their abatacept treatment

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study aim Study investigating the efficacy and safety of abatacept in patients with dermatomyositis or polymyositis refractory to conventional treatment using a randomised trial design with delayed start in 1 arm **Study dates** 2011 to 2013	myopathies including drug induced myositis, inclusion body myositis, malignancy associated myositis **Sample size** n=20 Immediate treatment (n=11) Delayed treatment (n=9) **Baseline characteristics** Baseline characteristics were similar between the 2 groups for demographics, previous treatment, concomitant medication and baseline disease activity and health assessment scores. Median age was higher for patients in the immediate treatment group than patients in the delayed treatment group (58.0, IQR 46.0 to 61.0 vs 47.0, IQR 40.5 to 54.0) (p=0.0375) At baseline, 25% of patients had received prior rituximab and 35% had received previous IVIG treatment	and 1,000mg for patients >100kg	Time period not reported No statistical test between groups reported No patients achieved major improvement (≥60 points) ***Muscle strength assessed by MMT-8*** *(0-80 with higher scores indicating greater muscle strength)* Significantly greater mean difference (SD) between month 0 (baseline) and month 3 for immediate treatment (2.5, SD 4.7) vs delayed treatment (-4.9, SD 9.1) (p=0.038) Significant improvement after 6 months of abatacept compared to baseline. 6 months median 74 (IQR 68.5 to 78) vs baseline median 70 (IQR 64 to 73) (p=0.047) ***Disability/ function assessed by HAQ*** *(0-3 with higher scores indicating greater disability)* No significant difference in mean difference (SD) between month 0 and month 3 for immediate (-0.2, SD 0.4) vs delayed (-0.0006, SD 0.2) treatment (p=0.296) No significant difference after 6 months of abatacept compared to baseline. 6 months median 1.00 (IQR 0.38 to 1.44) vs baseline median 1.00 (IQR 0.63 to 1.81) (p=0.427) ***Physician global activity score*** *(VAS, 0-100mm with higher scores indicating a higher level of disease activity)*	unclear risk of bias. There is high risk of bias for outcomes comparing the randomised groups as the sample size was small and based on feasibility rather than a power calculation. The study may not have been sufficiently powered to demonstrate efficacy Although 10 patients were randomised to each group, 1 patient was changed from delayed to immediate treatment due to aggressive progress of weakness. Baseline characteristics were reported after the transfer of this patient 1 patient was not included in the analysis due to early withdrawal from the trial (disease progression) It is not clear if intention-to-treat or per protocol analysis was used for the outcomes of interest As all patients received six months treatment with abatacept at some point, the study authors also pooled the six month data for all patients and compared this to baseline scores. Although the data was generated as part of a controlled trial, this was not a comparison of patients randomised to different intervention groups. **Other comments** Outcomes comparing 6 months of abatacept with 3 months of abatacept are not reported as this

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			No significant difference in mean difference (SD) between month 0 and month 3 for immediate (-10.8, SD 13.7) vs delayed (0.3, SD 13.8) treatment (p=0.096) No significant difference after 6 months of abatacept compared to baseline. 6 months median 20.0 (IQR 10.0 to 40.5) vs baseline median 30.0 (IQR 22.5 to 46.0) (p=0.063) ***Patient global activity score*** *(VAS, 0-100mm with higher scores indicating a higher level of disease activity)* No significant difference in mean difference (SD) between month 0 and month 3 for immediate (-1.1, SD 15.8) vs delayed (2.1, SD 18.5) treatment (p=0.434) No significant difference after 6 months of abatacept compared to baseline. 6 months median 29.0 (IQR 13.5 to 69.5) vs baseline median 42.0 (IQR 24.5 to 74.0) (p=0.458) ***Muscle enzymes*** *Creatine kinase (µcat/L) (higher scores indicate muscle injury)* No significant difference in mean difference (SD) between month 0 and month 3 for immediate (-3.2, SD 10.9) vs delayed (13.5, SD 18.7) treatment (p=0.094) No significant difference after 6 months of abatacept compared to baseline. 6 months median 2.8 (IQR 1.5 to 7.1) vs baseline median 3.0 (IQR 2.0 to 30.4) (p=0.438)	is not a comparison of interest listed in the PICO **Source of funding** The study was funded by grants from the pharmaceutical company and Swedish government and charitable organisations. The authors state that the work was carried out independently from the pharmaceutical company

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Lactate dehydrogenase (µcat/L) (higher scores indicate tissue damage) No significant difference in mean difference (SD) between month 0 and month 3 for immediate (-0.3, SD 1.3) vs delayed (1.9, SD 3.3) treatment (p=0.065) No significant difference after 6 months of abatacept compared to baseline. 6 months median 4.0 (IQR 3.1 to 4.6) vs baseline median 4.5 (IQR 3.8 to 7.1) (p=0.299) **Important outcomes** ***Disease activity assessed by extra-muscular global assessment*** *(VAS, 0-100mm with higher scores indicating a higher level of disease activity)* Significantly greater mean difference (SD) between month 0 (baseline) and month 3 for immediate treatment (-12.7, SD 14.5) vs delayed treatment (1.4, SD 12.2) (p=0.0353) No significant difference after 6 months of abatacept compared to baseline. 6 months median 23.0 (IQR 9.0 to 36.0) vs baseline median 30.0 (IQR 15.5 to 43.50) (p=0.1958) ***Quality of life assessed by SF-36 physical health component*** *(0-100 with higher scores indicating a better quality of life)* Significant improvement after 6 months of abatacept compared to baseline. 6 months median 37 (IQR 24 to 45) vs	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			baseline median 31 (IQR 24 to 35) (p=0.005) No results were reported for the SF-36 mental component score or SF-36 total score No between group assessment reported **Safety** ***Adverse events*** *(mild, moderate, severe or serious)* After 6 months of abatacept: 8 (out of 36) adverse events were considered related to abatacept: • Serious 0 (0%) • Severe: 0 (0%) • Moderate: 4 (50%) • Mild: 4 (50%) 28 (out of 36) adverse events were not considered to be related to abatacept: • Serious: 4 (14.3%) • Severe: 2 (7.1%) • Moderate: 4 (14.3%) • Mild: 17 (60.7%) • Not classified: 1 (3.6%) Adverse events included: • Infections (n=14) • Cardiovascular events (n=4) • Tumours (n=3) • Skin manifestations (n=3) • Musculoskeletal system effects (n=3) • Gastrointestinal effects/ nausea (n=2) • Urinary tract effects (n=1) • Neuropathological effects (n=1) • Other (n=5)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Type of adverse events considered related to abatacept not specified The study authors stated that there was no antibody positivity that was associated with clinical response or no response to treatment The study authors also stated that they could not demonstrate a particular baseline phenotype that was associated with response to abatacept	

HAQ – health assessment questionnaire; IQR – interquartile range; IVIG – intravenous immunoglobulin; JBI – Joanna Briggs Institute; kg – kilogram; L – litre; mg – milligram; MMT – manual muscle test; RCT – randomised controlled trial; SD – standard deviation; VAS – visual analogue scale; µcat - microcat

Appendix F Quality appraisal checklists

JBI Critical Appraisal Checklist for RCTs

1. Was true randomisation used for assignment of participants to treatment groups?
2. Was allocation to treatment groups concealed?
3. Were treatment groups similar at the baseline?
4. Were participants blinded to treatment assignment?
5. Were those delivering treatment blind to treatment assignment?
6. Were outcomes assessors blind to treatment assignment?
7. Were treatment groups treated identically other than the intervention of interest?
8. Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analysed?
9. Were participants analysed in the groups to which they were randomised?
10. Were outcomes measured in the same way for treatment groups?
11. Were outcomes measured in a reliable way?
12. Was appropriate statistical analysis used?
13. Was the trial design appropriate, and any deviations from the standard RCT design (individual randomisations, parallel groups) accounted for in the conduct and analysis of the trial?

Appendix G GRADE Profiles

Table 1: In children and adults with refractory IIMs (including intolerance to or an inadequate response to rituximab in patients with myositis-specific and myositis-associated antibodies), what is the clinical effectiveness and safety of abatacept compared with standard treatment?

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Immediate treatment with abatacept	Delayed treatment with abatacept i.e. standard treatment for 3 months, before starting abatacept	Result		
Total improvement score (1 RCT)									
Median total improvement score at month 3 (0-100 with higher scores indicating more improvement)									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10[3]	n=9	Abatacept median 28.8 (IQR 15 to 37.5) vs standard treatment median 5.0 (IQR 5.0 to 12.5) (p=0.03)	Critical	Very low
Percentage achieving minimal total improvement (≥20 points) at month 3									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept 60% vs standard treatment 20%. No statistical comparison reported	Critical	Very low

Percentage achieving minimal total improvement (≥20 points) after 6 months of abatacept									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Immediate treatment with abatacept 90% vs delayed treatment with abatacept 40%. No statistical comparison reported	Critical	Very low
Percentage achieving moderate total improvement (≥40 points). Time period not reported									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Immediate treatment with abatacept 40% vs delayed treatment with abatacept 10%. No statistical comparison reported	Critical	Very low
Muscle strength (1 RCT)									
Mean difference in muscle strength from baseline to month 3 assessed by MMT-8									
1 RCT Tjärnlund et al 2018	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference 2.5 (SD 4.7) vs standard treatment mean difference -4.9 (SD 9.1) (p=0.038)	Critical	Very low
Median muscle strength at baseline and after 6 months of abatacept assessed by MMT-8 (0 to 80 with higher scores indicating greater muscle strength)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[5]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 70 (IQR 64 to 73) vs 6 months abatacept median 74 (IQR 68.5 to 78) (p=0.047)	Critical	Very low

Disability/ function (1 RCT)									
Mean difference in disability/ function from baseline to month 3 assessed by HAQ									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -0.2 (SD 0.4) vs standard treatment mean difference -0.0006 (SD 0.2) (p=0.296)	Critical	Very low
Median disability/ function at baseline and after 6 months of abatacept assessed by HAQ (0-3 with higher scores indicating greater disability)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[6]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 1.00 (IQR 0.63 to 1.81) vs 6 months abatacept median 1.00 (IQR 0.38 to 1.44) (p=0.427)	Critical	Very low
Physician global activity score (1 RCT)									
Mean difference in physician global activity score from baseline to month 3 assessed by VAS									
1 RCT Tjärnlund et al 2018	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -10.8 (SD 13.7) vs standard treatment mean difference 0.3 (SD 13.8) (p=0.096)	Critical	Very low
Median physician global activity score at baseline and after 6 months of abatacept assessed by VAS (0-100mm with higher scores indicating a higher level of disease activity)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[5]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 30.0 (IQR 22.5 to 46.0) vs 6 months abatacept median 20.0 (IQR 10.0 to 40.5) (p=0.063)	Critical	Very low

Patient global activity score (1 RCT)									
Mean difference in patient global activity score from baseline to month 3 assessed by VAS									
1 RCT Tjärnlund et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -1.1 (SD 15.8) vs standard treatment mean difference 2.1 (SD 18.5) (p=0.434)	Critical	Very low
Median patient global activity score at baseline and after 6 months of abatacept assessed by VAS (0-100mm with higher scores indicating a higher level of disease activity)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[6]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 42.0 (IQR 24.5 to 74.0) vs 6 months abatacept median 29.0 (IQR 13.5 to 69.5) (p=0.458)	Critical	Very low
Muscle enzymes (1 RCT)									
Mean difference in creatine kinase from baseline to month 3 (µcat/L)									
1 RCT Tjärnlund et al 2018	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -3.2 (SD 10.9) vs standard treatment mean difference 13.5 (SD 18.7) (p=0.094)	Critical	Very low
Mean difference in lactate dehydrogenase from baseline to month 3 (µcat/L)									
1 RCT Tjärnlund et al 2018	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -0.3 (SD 1.3) vs standard treatment mean difference 1.9 (SD 3.3) (p=0.065)	Critical	Very low

Median creatine kinase at baseline and after 6 months of abatacept (µcat/L) (higher scores indicate muscle injury)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[5]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 3.0 (IQR 2.0 to 30.4) vs 6 months abatacept median 2.8 (IQR 1.5 to 7.1) (p=0.438)	Critical	Very low
Median lactate dehydrogenase at baseline and after 6 months of abatacept (µcat/L) (higher scores indicate tissue damage)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[5]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 4.5 (IQR 3.8 to 7.1) vs 6 months abatacept median 4.0 (IQR 3.1 to 4.6) (p=0.299)	Critical	Very low
Disease activity (1 RCT)									
Mean difference in extra-muscular global assessment score from baseline to month 3 assessed by VAS									
1 RCT Tjärnlund et al 2018	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable[2]	n=10	n=9	Abatacept mean difference -12.7 (SD 14.5) vs standard treatment mean difference 1.4 (SD 12.2) (p=0.0353)	Important	Very low
Median extra-muscular global assessment score at baseline and after 6 months of abatacept assessed by VAS (0-100mm with higher scores indicating a higher level of disease activity)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[5]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 30.0 (IQR 15.5 to 43.5) vs 6 months abatacept median 23.0 (IQR 9.0 to 36.0) (p=0.1958)	Important	Very low

Quality of life (1 RCT)									
Median quality of life at baseline and after 6 months of abatacept assessed by SF-36 physical health component (0-100 with higher scores indicating a better quality of life)									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[6]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Baseline median 31 (IQR 24 to 35) vs 6 months abatacept median 37 (IQR 24 to 45) (p=0.005)	Important	Very low
Adverse events									
Number of mild, moderate, severe or serious adverse events considered related to abatacept									
Analysis from 1 RCT Tjärnlund et al 2018	Very serious limitations[7]	No serious indirectness	Not applicable	Not calculable[2]	n=19	---	Serious: 0 Severe: 0 Moderate: 4 Mild: 4	Important	Very low

HAQ – health assessment questionnaire; IQR – interquartile range; kg – kilogram; L – litre; mg – milligram; MMT – manual muscle test; RCT – randomised controlled trial; SD – standard deviation; VAS – visual analogue scale; µcat - microcat

1 Risk of bias due to lack of patient blinding as this outcome included subjective reporting; study underpowered for this comparison between randomised groups with uncertainty about the study's ability to demonstrate effectiveness
2 Imprecision assessment not possible as no confidence intervals reported and insufficient information to assess effect size
3 N is 10 rather than 11 because 1 patient was not included in the analysis due to early withdrawal from the trial (disease progression)
4 Risk of bias due to the study being underpowered for this comparison between randomised groups with uncertainty about the study's ability to demonstrate effectiveness
5 Risk of bias for this outcome due to comparison using data pooled across groups with differences in when patients received abatacept
6 Risk of bias for this outcome due to comparison using data pooled across groups with differences in when patients received abatacept; lack of patient blinding as this outcome included subjective reporting
7 Risk of bias for this outcome due to descriptive data pooled across groups with differences in when patients received abatacept

Glossary (adapted from the NICE Glossary)

- Baseline. The set of measurements assessed at the beginning of a study with which subsequent results are compared.

- Bias. Systematic (as opposed to random) deviation of the results of a study from the 'true' results, which is caused by the way the study is designed or conducted.

- Blinding. A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results.

- Clinical importance. A benefit from treatment that relates to an important outcome such as length of life and is large enough to be important to patients and health professionals.

- Confidence interval (CI). A way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment, often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate. The confidence interval is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value.

- Intention-to-treat analysis (ITT). An assessment of the people taking part in a trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully adhered to the treatment or switched to an alternative treatment. ITT analyses are often used to assess clinical effectiveness because they mirror actual practice, when not everyone adheres to the treatment, and the treatment people have may be changed according to how their condition responds to it. Studies of drug treatments often use a modified ITT analysis, which includes only the people who have taken at least one dose of a study drug.

- P value. The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing two treatments found that one seems to be more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance), it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 0.1% probability that the results occurred by chance), the result is seen as highly significant. However, a statistically significant difference is not necessarily clinically significant.

- Per-protocol analysis. A comparison of treatment groups in a trial that includes only those patients who completed the treatment they were originally allocated to. If done alone, this analysis leads to bias.

- Randomised controlled trial. A study in which a number of similar people are randomly assigned to two (or more) groups to test a specific drug, treatment or other intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, a dummy intervention (placebo) or no intervention at all. The groups are followed up to see how effective the experimental intervention was. Outcomes are measured at specific

times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.

- Standard deviation. A measure of the spread, scatter or variability of a set of measurements. Usually used with the mean (average) to describe numerical data.

- Statistical significance. A statistically significant result is one that is assessed as being due to a true effect rather than random chance.

References

Included studies

- Tjärnlund A, Tang Q, Wick C, Dastmalchi M, Mann H, Tomasová Studýnková J, Chura R, Gullick NJ, Salerno R, Rönnelid J, Alexanderson H, Lindroos E, Aggarwal R, Gordon P, Vencovsky J, Lundberg IE. Abatacept in the treatment of adult dermatomyositis and polymyositis: a randomised, phase IIb treatment delayed-start trial. Ann Rheum Dis. 2018;77: 55-62.

Other references

- Aggarwal R, Rider LG, Ruperto N, Bayat N, Erman B, Feldman BM, Oddis CV, Amato AA, Chinoy H, Cooper RG, Dastmalchi M, Fiorentino D, Isenberg D, Katz JD, Mammen A, de Visser M, Ytterberg SR, Lundberg IE, Chung L, Danko K, Garciá-De la Torre I, Song YW, Villa L, Rinaldi M, Rockette H, Lachenbruch PA, Miller FW, Vencovsky J, International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. Ann Rheum Dis. 2017;76(5): 792–801.
- Rider LG, Aggarwal R, Pistorio A, Bayat N, Erman B, Feldman BM, Huber AM, Cimaz R, Cuttica RJ, Knupp de Oliveira S, Lindsley CB, Pilkington CA, Punaro M, Ravelli A, Reed AN, Rouster-Stevens K, van Royen-Kerkhof, Dressler F, Saad Mgalhaes C, Constantin T, Davidson JE, Magnusson B, Russo R, Villa L, Rinaldi M, Rockette H, Lachenbruch PA, Miller FW, Vencovsky J, Ruperto N, International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism Criteria for Minimal, Moderate, and Major Clinical Response in Juvenile Dermatomyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. Arthritis Rheumatol. 2018;69(5): 911–923.