CPAG Summary Report for Clinical Panel – URN 1673: Infliximab for Progressive Pulmonary Sarcoidosis

The Benefits of the Proposition – Infliximab vs. placebo for sarcoidosis			
No	*Metric*	*Grade of evidence*	*Summary from evidence review*
1.	Survival	Not measured	
2.	Progression free survival	Not measured	
3.	Mobility	Not measured	
4.	Self-care	Not measured	
5.	Usual activities	Not measured	
6.	Pain	Not measured	
7.	Anxiety / Depression	Not measured	
8.	Replacement of more toxic treatment	Not measured	
9.	Dependency on care giver / supporting independence	Not measured	
10.	Safety	Adverse events identified [C]	The percentage of patients reporting adverse events and serious adverse events were reported in the Rossman et al (2006) RCT. Four patients reported serious events (31%) in Group 1 (n=13, infliximab) and one (17%) in Group 2 (n=6, placebo). Adverse events were reported by 92% of Group 1 and 100% of Group 2. Infections were reported by 69% of Group 1 and 50% of Group 2. No significance tests were reported. Most of the participants in the study reported adverse events, and serious adverse events were reported in approximately one third of the patients receiving infliximab.

			This was a randomised, double-blind, placebo controlled phase II trial. However, the study was closed early due to poor enrolment and was therefore underpowered to detect any differences between the groups. The results should be treated with caution.
11.	Delivery of intervention	Not measured	

Other health metrics determined by the evidence review: Infliximab vs. placebo for sarcoidosis			
No	Metric	Grade of evidence	Summary from evidence review
1.	Pulmonary function parameters	Grade C	Pulmonary function was assessed using observed and percent expected vital capacity (VC). As no minimal important difference for change in VC has been defined (van Rijswijk et al 2013) % predicted and proportion of patients with a specified increase e.g. 5%, 10% or 15%, are also reported. In the 6-week randomised phase of the RCT (Rossman et al 2006) there was no significant difference in the mean ± SD relative % change in expected VC between Group 1 (infliximab) (15.2% ± 9.9%) and Group 2 (placebo) (8.4% ± 3.3%), (p=0.65). The observed mean vital capacity was 2.5 litres at baseline and 2.7 litres at week 6 in Group 1 (infliximab) and 2.4 litres at both baseline and week 6 in Group 2 (placebo) (no significance test reported). Three patients (23%) in Group 1 had an improvement of ≥10% predicted VC compared to two patients (33%) in Group 2. Two patients (15%) in Group 1 had an improvement of ≥15% predicted VC compared to no patients (0%) in Group 2. No significance tests were reported for these comparisons. In the phase 2 open label follow up

			from week 12 to week 38, the VC % expected ranged from 65.5 to 67.4 in Group 1 (baseline 59.6) and 70.7 to 72.5 in Group 2 (baseline 65.5). When the results for the first 6 weeks of taking infliximab were combined (week 0 to 6 in Group 1 and week 6 to 12 in Group 2) a significant improvement in mean VC was found ($p<0.02$) (mean figures not reported). An improvement in mean vital capacity % expected was seen in both groups. 23% of the patients in Group 1 (infliximab) and 33% of patients in Group 2 (placebo) achieved an improvement of ≥10% in the randomised phase of the study. No significant differences between infliximab and placebo were reported. This was a randomised, double-blind, placebo controlled phase II trial. However, the study was closed early due to poor enrolment and the study was therefore underpowered to detect any differences between the groups. The results should be treated with caution.
2.	Radiologic improvement on chest x-ray	Grade C	Chest x-rays were assessed as 'markedly worse', 'slightly worse', 'unchanged', 'slightly improved' or 'markedly improved'. In the 6-week randomised phase of the RCT (Rossman et al 2006), 23.0% of Group 1 (infliximab) had radiologic improvement compared to 0% of Group 2 (placebo). No significance test was reported. In the phase 2 open label the percentage of patients showing radiologic improvement appear similar in both groups (Group 1 30.8% vs. Group 2 33.3%). No significance test was reported.

			More patients in Group 1 showed radiologic improvement in the randomised phase of the study however it is not clear if this difference was statistically or clinically significant. It is not clearly stated if the reported improvement included both patients who had 'slightly improved' and 'markedly improved'. This was a randomised, double-blind, placebo controlled phase II trial. However, the study was closed early due to poor enrolment and the study was therefore underpowered to detect any differences between the groups. The results should be treated with caution.
3.	Dyspnea score	Grade C	The Baseline Dyspnea Index (BDI) and Transitional Dyspnea Index (TDI) are used to assess breathlessness and the impact of intervention. The BDI and TDI have three domains: functional impairment, which determines the impact of breathless on the ability to carry out activities; magnitude of task, which determines the type of task that causes breathlessness; and magnitude of effort, which establishes the level of effort that results in breathlessness. The BDI is scored from 0 (very severe impairment) to 4 (no impairment) for each domain and summed to create a focal score (0-12). The TDI is scored from -3 (major deterioration) to +3 (major improvement) for each domain and summed to create a focal score (-9 to +9). An improvement of ≥1.0 has been suggested to be clinically important (Witek & Mahler 2003). In the RCT by Rossman et al 2006, the baselines scores assessed on the BDI were 2.17 for Group 1 and 2.08 for Group 2. Improvements on

			the TDI in the 6-week randomised phase were reported for both groups in mean functional impairment (Group 1 0.38 vs. Group 2 0.17) and mean magnitude of effort (Group 1 0.23 vs. Group 2 0.17). No significance tests were reported. No change in magnitude of task was reported (figures not reported). The meaning of the dyspnea scores are not clearly reported and it is not clear if the difference in scores were statistically significant. The scores of between 0 and 1 on the TDI would suggest a small, but not clinically important, improvement. This was a randomised, double-blind, placebo controlled phase II trial. However, the study was closed early due to poor enrolment and the study was therefore underpowered to detect any differences between the groups and the results should be treated with caution.
4.	Quality of life	Grade C	Quality of life was assessed using the SF-36 which is scored from 0-100 with higher scores indicating better functioning. An improvement of 10 points was considered clinically relevant in one study (Vorselaars et al 2015). In the RCT, SF-36 scores were similar in both groups at baseline (Group 1 26.7 vs. Group 2 26.4) and at the end of the 6-week randomised phase (Group 1 27.1 vs. Group 2 26.4). No significance tests were reported (Rossman et al 2006). No significant improvement in quality of life was reported for either group. This was a randomised, double-

			blind, placebo controlled phase II trial. However, the study was closed early due to poor enrolment and the study was therefore underpowered to detect any differences between the groups and the results should be treated with caution.

The Benefits of the Proposition – Infliximab for sarcoidosis (no comparator)			
No	*Metric*	*Grade of evidence*	*Summary from evidence review*
1.	Survival	Not measured	
2.	Progression free survival	Not measured	
3.	Mobility	Not measured	
4.	Self-care	Not measured	
5.	Usual activities	Not measured	
6.	Pain	Not measured	
7.	Anxiety / Depression	Not measured	
8.	Replacement of more toxic treatment	Benefit determined [A]	The dose of corticosteroid at baseline and follow-up is compared to assess the corticosteroid sparing effect of treatment. In the prospective study (Vorselaars et al 2015, n= 56) with 26 weeks follow-up there was a statistically significant reduction in mean corticosteroid dose by 8.8mg for 19 patients who received concomitant corticosteroids (p=0.001). The baseline and follow-up dose per day was not reported. In the largest retrospective study (Jamilloux et al 2017, n=132) with a median of 20.5 months follow-up, there was a statistically significant reduction in the mean dose of corticosteroid from 23 ± 20 mg/day at baseline to 11 ± 11 mg/day in 113 patients who received concomitant corticosteroids. A reduction in corticosteroid use is a positive outcome. In these studies the

			mean reduction in corticosteroids was approximately 9-12 mg/ day. The clinical meaningfulness of a reduction of this magnitude in this population is not clear. The evidence comes from uncontrolled studies with relatively large sample sizes. Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
9.	Dependency on care giver / supporting independence	Benefit determined [A]	The modified Rankin score measures the degree of disability or dependence in the daily activities of people who have suffered a stroke or other causes of neurological disability. It is scored from 0 (no symptoms at all) to 6 (dead). A score of 3 is defined as moderate disability; requiring some help, but able to walk without assistance. A score of 1 is defined as no significant disability despite symptoms; able to carry out all usual duties and activities[1]. A statistically significant improvement in modified Rankin score was reported from 3 at baseline to 1 at a median follow-up of 20 months. The modified Rankin scores reported translate to an improvement from moderate disability to no significant disability despite symptoms. It is likely that this improvement would be meaningful to the patient. This uncontrolled retrospective review (Cohen Aubart et al 2017) included a small number of patients from three centres (n=18). The retrospective design and lack of comparator limit the strength of the conclusions that can be

[1] https://www.mdcalc.com/modified-rankin-scale-neurologic-disability

			drawn.
10.	Safety	Adverse events identified [A]	The percentage of patients reporting side effects and severe side effects were reported in the uncontrolled prospective study by Vorselaars et al (2015). Severe side effects were reported in 7 patients (13%). These included 3 patients with pneumonia requiring hospitalisation, 2 patients with allergic reaction with antibody formation,1 patient with peritonitis and1 patient with severe gastrointestinal complaints. Other side effects were reported for 13 patients (23%) including mild infection of the respiratory tract (n=5), oedema (n=3), headache (n=2), joint pain (n=2) and dizziness (n=1). 61% of the patients did not experience any side effects from infliximab. 13% of the study population experienced severe side effects. This uncontrolled prospective study had a relatively large sample size (n=56). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
11.	Delivery of intervention	Not measured	

Other health metrics determined by the evidence review: Infliximab for sarcoidosis (no comparator)			
No	Metric	Grade of evidence	Summary from evidence review
1.	Pulmonary function parameters	Grade A	Pulmonary function was assessed using forced vital capacity (FVC), forced expiratory volume in 1 second (FEV_1) and diffusing capacity of the lung for carbon monoxides, corrected for haemoglobin (DLCOc). As no minimal important difference for

			change in FVC has been defined (Vorselaars et al 2015) studies also report % predicted and percentage of patients with a specified increase e.g. 5%, 10% or 15%. An improvement in forced expiratory volume of ≥10% was considered to be a clinically relevant change for an individual patient (van Rijswijk et al 2013). Function was also assessed by the 6-minute walking test which assesses how far someone can walk in 6 minutes. In patients with a pulmonary treatment indication there was a significant improvement from baseline for FVC % predicted (+6.6 from a baseline of 73.6), FEV_1 % predicted (+5.8 from a baseline of 55.8) and DLCOc (+4.1 from a baseline of 56.6) at 26 week follow-up. 71% of patients had an improvement of ≥5% FVC predicted and 46% of patients had an improvement of ≥10% FVC predicted. For FEV_1 % predicted, 64% had an improvement of ≥5% and 46% an improvement of ≥10%. The 6-minute walking distance % predicted improved by +4.2 from a baseline of 61. No unit of measurement or significance test was reported for this measure. In patients with an extrapulmonary treatment indication there was a significant improvement from baseline for FVC % predicted (+3.9) and FEV_1 % predicted (+3.5) (baseline figures not reported for these patients). 37% of patients had an improvement of ≥5% FVC predicted and FEV_1 % predicted. A statistically significant improvement was seen for all measures where significance testing was reported. An improvement of ≥10% was considered clinically relevant. This was achieved by 46% patients with a pulmonary treatment indication. This uncontrolled prospective study by

			Vorselaars et al (2015) had a relatively large sample size (n=56). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
2.	Clinical response (physician evaluation)	Grade A	Clinical response was assessed by physicians in five uncontrolled retrospective studies. The response definitions used varied for each study. In the largest of these studies (Jamilloux et al 2017) the following categories were used: • Complete response (disappearance of clinical signs (excluding sequelae) upon use of corticosteroids <10mg) • Partial response (an improvement of clinical and para-clinical parameters upon >50% reduction of the initial corticosteroids dose) • Stable (non-responders with no change) • Progressive disease (non-responders with either new organ involvement, worsening of an organ involvement or need for increased corticosteroid dosage). After a median follow-up of 20.5 months, a complete or partial response was reported in 18% and 46% of patients respectively. Non-responders included 25% of patients with no change and 11% of patients with progressive disease. In multivariate analysis comparing responders with non-responders, pulmonary involvement was associated with a lower clinical response (OR =0.38, 95%CI 0.14 to 0.92). A complete or partial response was seen in 64% patients, with most of these showing a partial response. The clinical meaningfulness of a partial

			response is not clear although the definition includes improvement with a reduction in corticosteroid use. This uncontrolled retrospective review included patients from 25 centres and had a relatively large sample size (n= 132). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement and a small number of patients included in the sample (5%) did not receive infliximab as the anti-TNF treatment. The retrospective design and lack of comparator limit the strength of the conclusions that can be drawn.
3.	Organ assessment	Grade B	The Extrapulmonary Physician Organ Severity Tool (ePOST) examines 17 extrapulmonary organs and assigns each a score from 0 (not affected) to 6 (very severely affected) to each organ. Intermediate scores are slight (1), mild (2), moderate (3), moderate to severe (4) and severe (5). ePOST scores were provided for a range of organs. Pulmonary and neurological outcomes are of particular interest in this review. Statistically significant improvements between baseline and a median follow-up of 20.5 months were reported for the upper respiratory tract (2.3 to 1.8), central nervous system (CNS) (3.8 to 2.6) and peripheral nervous system (PNS) (1.1 to 0.24). There was no significant improvement for lungs between baseline (2.1) and follow-up (1.9). Statistically significant improvements were reported for the upper respiratory tract, CNS and PNS. The difference in scores between baseline and follow-up was greater for the nervous system scores. For the CNS this equates to an improvement from ‘moderate to severe’ to ‘mild to

			moderate'. For the PNS this equates to an improvement from 'slight' to 'not affected to slight'. This uncontrolled retrospective review (Jamilloux et al 2017) included patients from 25 centres and had a relatively large sample size (n= 132). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement and a small number of patients included in the sample (5%) did not receive infliximab as the anti-TNF treatment. The retrospective design and lack of comparator limit the strength of the conclusions that can be drawn.
4.	Inflammatory response	Grade A	Measures of inflammatory response included the biomarkers soluble interleukin-2-receptor (sIL-2R), angiotensin-converting enzyme (ACE) and F-fluorodeoxyglucose (FDG) by positron emission tomography (PET) maximum standardised uptake value (SUV_{max}). A statistically significant improvement from baseline to 26 week follow-up was reported for each of the inflammatory response measures in Vorselaars et al (2015). For F-FDG PET SUV_{max} the mean improvement was -4.0 from a baseline of 6.6 for lung parenchyma, -3.0 from a baseline of 5.7 for mediastinum and -5.8 from a baseline of 9.0 for index localisation. For ACE the improvement after infliximab was -28.2 from a baseline of 89.7 and for sIL-2R the improvement was -4269 from a baseline of 8,824. No information about the clinical relevance of the results reported was provided by the study authors. This uncontrolled prospective study by Vorselaars et al (2015) had a

			relatively large sample size (n=56). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
5.	Quality of life	Grade A	Quality of life was assessed using a Patient Global Assessment (PGA) with scores on a visual analogue scale ranging from 0 (best imaginable health status) to 100 (worst imaginable health status); and the SF-36 to assess physical functioning. The SF-36 is scored from 0-100 with higher scores indicating better functioning. An improvement of 10 points was considered clinically relevant (Vorselaars et al 2015). A statistically significant improvement was seen in both quality of life measures used from baseline to follow up at 26 weeks. For the PGA the mean score improved by -14.6 from a baseline of 61.0. For the SF-36 the score improved by 8.2 from a baseline of 40.6. An improvement of 10 points was considered clinically relevant. An improvement of >10 points was reported for the PGA but not for the SF-36. This uncontrolled prospective study by Vorselaars et al (2015) had a relatively large sample size (n=56). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
6.	Clinical response (patient reported)	Grade C	For patient-reported symptomatic response Russell et al (2013) used the following categories: • 'Resolved' = complete resolution of clinical disease activity

			• 'Improved' = organs with reduced sarcoid burden or reduced frequency in disease activity but still with evidence of disease • 'Unchanged' = disease activity clinically no different than prior to infliximab • 'Progressed' = clinical features of progressive disease despite infliximab Separate scores were provided for all organs and individual organs. Pulmonary and neurological outcomes are of particular interest in this review. When all organs were included, an improvement was reported by 73% of patients, consisting of 20% who were 'resolved' and 53% who were 'improved'. Of those that had not seen an improvement, 23% were 'unchanged' and 5% had 'progressed'. For lungs, an improvement was reported by 60% of patients, all of whom were 'improved'. For the central nervous system, an improvement was reported by 75% patients consisting of 25% 'resolved' and 50% 'improved'. The majority of patients reported an improvement in their symptoms with 20% reporting a complete resolution of disease activity. This uncontrolled retrospective review by Russell et al (2013) included only a small number of patients from a single centre (n=26). The retrospective design and lack of comparator limit the strength of the conclusions that can be drawn.
7.	Composite overall response	Grade A	The composite overall response included organ function, inflammation and quality of life (Vorselaars et al 2015). This was an author-designed non-validated tool. Improvement in a category was scored only when one of the parameters improved significantly without deterioration of the others. A

			good or excellent response was a clinically relevant improvement in 2 or 3 categories, a partial response was a clinically relevant improvement in one category and no improvement in any category was a nonresponse. After 26 weeks follow-up a response was reported in 96% of patients. This included 40% showing an excellent response, 39% a good response and 17% a partial response. The definitions used for this outcome measure include the clinical relevance of the improvement observed. 40% of patients showed an 'excellent' response which equates to a clinically relevant improvement in all three categories. The composite tool used has not been validated and the study authors advised that it should be interpreted with care. This uncontrolled prospective study by Vorselaars et al (2015) had a relatively large sample size (n=56). Patients had a range of clinical manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.
8.	Infliximab trough levels	Grade A	The trough level is the lowest concentration reached by a drug before the next dose is administered. It can be used to evaluate appropriate dosage levels. The mean trough level was 18.0 $\mu g/mL^{-1}$. There was no significant correlation between trough level and response. This uncontrolled prospective study by Vorselaars et al (2015) had a relatively large sample size (n=56). Patients had a range of clinical

			manifestations for sarcoidosis including pulmonary and/or CNS involvement. The lack of comparator in these studies limits the strength of the conclusions that can be drawn.