CONFIDENTIAL UNTIL PUBLISHED

External Assessment Group Report commissioned by the NIHR Evidence Synthesis Programme on behalf of NICE

Spesolimab for treating generalised pustular psoriasis flares

---

Produced by	**Southampton Health Technology Assessments Centre (SHTAC)**
Authors	Inês Souto Ribeiro, Senior Research Assistant, Health Economics Lois Woods, Senior Research Assistant, Evidence Synthesis and Information Specialist Asyl Hawa, Research Fellow, Health Economics Joanna Picot, Senior Research Fellow, Evidence Synthesis
Correspondence to	Dr Jo Picot Southampton Health Technology Assessments Centre (SHTAC) School of Healthcare Enterprise and Innovation University of Southampton Alpha House, Enterprise Road, University of Southampton Science Park, Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	18 October 2024
Source of Funding	This report was commissioned by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as project number NIHR135887.

Acknowledgements

We thank the following for providing a quality assurance review of the draft report:

- Dr Jonathan Shepherd, Principal Research Fellow, SHTAC
- Dr Marcia Takahashi, Research Fellow SHTAC
- Dr Emma Maund, Research Fellow, SHTAC

We also thank Professor David Scott, Professorial Fellow – Research, for contributing to the clarification questions and providing statistical advice.

Declared competing interests of the authors and advisors

The report authors declare none.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Souto Ribeiro, I; Woods, L. Hawa, A.; Picot, J. Spesolimab for treating generalised pustular psoriasis flares. A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2024.

Contributions of authors

Ines Ribeiro critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Lois Woods critically appraised the clinical effectiveness systematic review and drafted the report; Asyl Hawa critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Joanna Picot critically appraised the clinical effectiveness systematic review, drafted the report and is the project co-ordinator and guarantor.

Commercial in confidence (CIC) information in blue

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

BAC	Best available care
BMI	Body mass index
BNF	British National Formulary
BSA	Body surface area
CEE	Central Eastern European
CI	Confidence interval
CIC	Commercial in confidence
CPRD	Clinical Practice Research Datalink
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
DLQI	Dermatology Life Quality Index
DRESS	Drug reaction with eosinophilia and systemic symptoms
DSA	Deterministic sensitivity analyses
EAG	External Assessment Group
eMIT	Electronic Market Information Tool
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
EQ-VAS	EuroQol Visual Analogue Scale
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue scale
GPP	Generalised pustular psoriasis
GPPASI	Generalised Pustular Psoriasis Area Severity Index
GPPGA	Generalised Pustular Psoriasis Physician Global Assessment
HCRU	Health care resource use
HES	Hospital Episode Statistics
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
ICU	Intensive care unit
IL36	Interleukin-36
IL36Ra	Interleukin-36 receptor antagonist

IL36RN	Interleukin-36 receptor antagonist protein
ITT	Intent to treat
IV	intravenous
JDA	Japanese Dermatological Association
MCID	Minimum clinically important difference
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PAS	Patient Access Scheme
PGA	Physician Global Assessment
PRO	Patient reported outcome
PSA	Probabilistic sensitivity analysis
PSS	Psoriasis Symptom Scale
PSSRU	Personal Social Services Research Unit
PV	Psoriasis vulgaris
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RWE	Real-world evidence
SD	Standard deviation
SEE	Structured expert elicitation
SLR	Systematic literature review
SmPC	Summary of product characteristics
SNDS	Système Nationale des Donneés de Santé = French National Health Data System
TA	Technology appraisal
TEAE	Treatment-emergent adverse event
TNF	Tumour necrosis factor
UK	United Kingdom
US	United States
VAS	Visual analogue scale

1 EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the external assessment group (EAG) as being potentially important for decision making. It also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, health technology, evidence and information on the issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1 Overview of the EAG's key issues

Table 1 Summary of Key Issues

ID	Summary of issue	Report sections
1	The trial evidence is from a narrower population than defined in the NICE scope and company's decision problem.	2.3, 3.2.1, 4.2.3
2	Comparator in the Effisayil 1 trial is placebo not established clinical management without spesolimab	2.3, 3.2.1, 3.2.5.1, 3.2.5.2, 4.2.6
3	A second/recurrent GPP flare is not implemented in the model	4.2.2.2.1
4	Treatments included in the comparator arm	4.2.4
5	Short time horizon	4.2.5
6	Efficacy of the comparator arm (BAC): response to treatment	4.2.6.2
7	Proportion of patients treated as inpatients in the spesolimab arm	4.2.8.3

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are the composition of the comparator in the first week of the model and the efficacy of BAC.

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Following their response to the clarification questions, the company updated their economic model. The company revised base case deterministic cost-effectiveness results are shown in

Table 2 with a confidential PAS discount applied for spesolimab. The ICER is -£172,713 per QALY for spesolimab versus BAC (spesolimab dominates BAC), with a QALY gain of [redacted] and a reduced cost of [redacted].

Table 2 Company revised base case results after clarification responses using PAS price for spesolimab

Treatment	Total		Incremental		ICER
	Cost (£)	QALYs	Cost (£)	QALYs	(£/QALY)
BAC	[redacted]	[redacted]			-£172,713 (spesolimab dominates)
Spesolimab	[redacted]	[redacted]	[redacted]	[redacted]	

Source: Clarification response document Table 13, company revised economic model (Deterministic Results!F15)
BAC, best available care; QALY, quality adjusted life year; ICER, incremental cost-effectiveness ratio.

1.3 The decision problem: summary of the EAG's key issues

Although no key issues were identified with respect to the decision problem we note that the comparator in the company's key trial, Effisayil 1, comes from a narrower patient group (see Key Issue 1) and the comparator in the Effisayil 1 trial is placebo not established clinical management without spesolimab (see Key Issue 2 and Key Issue 4).

1.4 The clinical effectiveness evidence: summary of the EAG's key issues

Issue 1 The trial evidence is from a narrower population than defined in the NICE scope and company's decision problem

Report section	2.3, 3.2.1, 4.2.3
Description of issue and why the EAG has identified it as important	The NICE scope and company's decision problem define the population for this appraisal as "*Adult patients with generalised pustular psoriasis presenting with flares*". This aligns with the therapeutic indication for spesolimab which is "*the treatment of flares in adult patients with generalised pustular psoriasis as a monotherapy*". However, the pivotal clinical trial that informs the company's submission and economic model, the Effisayil 1 trial, enrolled adult patients with generalised pustular psoriasis presenting with flares of moderate-to-severe intensity (section 3.2.1.1). Therefore, the trial evidence is drawn from a narrower population group (i.e. only people experiencing flares of moderate-to-severe intensity) than defined in the NICE scope and company decision problem. Additionally, the Structured Expert Elicitation (SEE) exercise for best available care efficacy and safety asked clinical experts to make estimates for patients with a moderate or a severe flare.

What alternative approach has the EAG suggested?	We do not suggest an alternative approach because there is no alternative source of trial evidence that is fully aligned with the NICE scope, company decision problem and economic model population.
What is the expected effect on the cost-effectiveness estimates?	Unknown, but the results from the cost effectiveness model may not be generalisable to a population of adults with generalised pustular psoriasis experiencing flares of any severity.
What additional evidence or analyses might help to resolve this key issue?	Clinical expert opinion about the patients expected to receive spesolimab in clinical practice and any potential for differences in spesolimab efficacy and best available care (BAC) efficacy for mild flares in comparison to moderate-to-severe flares.

Issue 2 Comparator in the Effisayil 1 trial is placebo not established clinical management without spesolimab

Report section	2.3, 3.2.1, 3.2.5.1, 3.2.5.2, 4.2.6
Description of issue and why the EAG has identified it as important	The NICE scope and company's decision problem define the comparator for this appraisal as established clinical management without spesolimab. The company's Figure showing the clinical pathway indicates which therapies are used at first-, second- and third-line or later in the treatment of moderate-to-severe GPP flares in the UK (section 2.2.3 of this report). However, the comparator arm in the company's pivotal clinical trial Effisayil 1 received placebo only (prior to randomisation, patients had to discontinue biologic therapies, systemic non-biological therapies and other treatments such as phototherapy and topical treatments). The economic model assumes that patients in the comparator arm receive no treatment during the first week (link with Key Issue 4). Therefore, clinical effectiveness evidence for the comparator arm in the first week of the trial and the associated period for the economic model does not align with clinical practice.
What alternative approach has the EAG suggested?	There is no alternative source of comparative trial evidence for spesolimab where the comparator is established clinical management without spesolimab.
What is the expected effect on the cost-effectiveness estimates?	The treatment effect for the first week for a comparison of spesolimab versus best available care is unknown. In the EAG's opinion the model may over-estimate the treatment difference between spesolimab and best available care in the first week because it is based on data from spesolimab and placebo treated trial participants. However, we acknowledge the absence of any reliable data to show whether there is a difference in effectiveness between best available care and placebo. Also see Key Issue 4
What additional evidence or analyses might help to resolve this key issue?	Given the nature of the available evidence it may be difficult to resolve this key issue. Clinical expert opinion may be able to provide an indication of the impact on flare symptoms when a patient does not receive any treatment for a week.

1.5 The cost-effectiveness evidence: summary of the EAG's key issues

Issue 3 A second/recurrent GPP flare is not implemented in the model

Report section	4.2.2.2.1
Description of issue and why the EAG has identified it as important	Patients who respond to treatment (i.e., have a GPPGA pustulation subscore of 0 or 1) are assumed to remain responders for the remainder of the modelled time horizon. However, evidence from the Effisayil 1 trial shows that 11.3% of patients received rescue treatment with spesolimab to treat second/recurrent flares, and eight patients received a standard of care escape treatment after day 8 although there is no information on how many of these patients have achieved a clinical response before getting worse. The company did not include second or recurrent flares in the model. We consider that understanding how modelling second/recurrent flares might affect the model conclusions is relevant for the committee's decision.
What alternative approach has the EAG suggested?	The EAG don't have data to model second or recurrent flares. It is unclear whether there is currently sufficient evidence (efficacy, safety, and costs) available to incorporate this in the model.
What is the expected effect on the cost-effectiveness estimates?	Uncertain; we note that the total costs will increase to account for the additional drugs used but we are unclear about the efficacy of the rescue treatments for patients with second/recurrent flares.
What additional evidence or analyses might help to resolve this key issue?	Evidence on the efficacy and safety of the rescue treatments and further clinical expert opinion on whether the company's assumption is reasonable or not.

Issue 4 Treatments included in the comparator arm

Report section	4.2.4
Description of issue and why the EAG has identified it as important	There are no licensed treatments specifically approved to treat GPP flares in the UK aside from spesolimab. In the company's model, it was assumed that patients in the comparator arm (BAC) do not receive any drugs during the first week (to align with the placebo data of the Effisayil 1 trial where patients in the placebo arm did not receive any active drugs during the first week). In the EAG's view, it is unlikely that patients in UK clinical practice would not receive any active drugs to treat GPP flares for a whole week (see Key Issue 2). Beyond that, data from the Effisayil 1 historical cohort study informed the efficacy of BAC (see Key Issue 6 for further details), although the treatments used to treat each flare in this study are not possible to derive. So, the treatments included in the BAC arm of the company's model beyond week 1 were obtained from a group of UK experts participating in a SEE exercise. We believe that modelling the treatments indicated by the UK experts as part of the

	SEE exercise is a reasonable approach in the absence of better UK evidence. However, it is unclear how closely these treatments match the treatments used to treat moderate-to-severe GPP flares in the Effisayil 1 historical cohort study (see Key Issue 6 for further details).
What alternative approach has the EAG suggested?	The EAG included active treatments in the comparator arm in the first week of the model in our base case, obtained from the SEE exercise. This is aligned with the EAG base case for the efficacy of the comparator arm (see Key Issue 6 below).
What is the expected effect on the cost-effectiveness estimates?	Including active treatments in the comparator arm in the first week of the model leads to ICERs of £15,680 (using the company's base case source of efficacy for BAC - Effisayil 1 1 historical cohort) and to £143,574 (using the EAG base case source of efficacy for BAC – SEE exercise).
What additional evidence or analyses might help to resolve this key issue?	Further clinical expert opinion on which treatments are currently considered the best available care in UK clinical practice to treat GPP flares.

Issue 5 Short time horizon

Report section	4.2.5
Description of issue and why the EAG has identified it as important	A time horizon of 12 weeks was implemented in the company's model, which is the follow-up period of the Effisayil 1 trial. The EAG notes that relevant evidence to inform the model is only available for a period of 12 weeks and therefore modelling a longer time horizon would be challenging. The company also stated that evidence on flare frequency shows that patients are unlikely to have more than two flares per year. However, it is possible that treating one flare with spesolimab might affect the efficacy and safety of spesolimab or other treatments in a subsequent flare, even if that only happens after a period of 12 weeks. Therefore, we consider the appropriateness of a 12-week time horizon to be uncertain and questionable because the long-term consequences of the treatment with spesolimab are still unclear.
What alternative approach has the EAG suggested?	The EAG is unable to model a longer time horizon as long-term data does not seem to be currently available.
What is the expected effect on the cost-effectiveness estimates?	Uncertain as we are unclear about the long-term effects of spesolimab.
What additional evidence or analyses might help to resolve this key issue?	Further clinical data on the long-term effects of spesolimab (if possible) and clinical expert opinion on the plausibility of assuming that spesolimab has similar incremental effectiveness/costs versus BAC in the treatment of individual flares, regardless of being the first or subsequent flares.

Issue 6 Efficacy of the comparator arm (BAC): response to treatment

Report section	4.2.6.2
Description of issue and why the EAG has identified it as important	For the first week of the model, treatment response to BAC was obtained from the Effisayil 1 trial and, beyond that, from the Effisayil 1 historical cohort, as crossover occurred for more than 80% of patients in the placebo arm on day 8 in the Effisayil 1 trial. Effisayil 1 trial provides a direct comparison between the intervention and the comparator. However, patients in the placebo arm of the Effisayil 1 trial do not receive any standard of care treatments although they may be admitted to hospital. It is unknown if receiving active treatments would have led to different effectiveness outcomes. Therefore, using trial data to inform the efficacy of BAC does not seem to appropriately reflect UK reality. The Effisayil 1 historical cohort uses the same cohort of patients as in the Effisayil 1 trial, however it is unclear how long in the past the flares have occurred and whether the standard of care at that time is similar to the standard of care currently used as we do not know which treatments were included in standard of care in this study. Therefore, it is uncertain how generalisable the results from Effisayil 1 historical cohort are to the current UK reality. The SEE exercise has limitations, such as being a low quality evidence source when compared to clinical trials and RWE studies; the lack of consistency in the definition of response to treatment; the discrepancy in the estimates of efficacy between clinicians; and the potential biased interpretation of best available care by the experts (see further details in section 4.2.6.2.). However, the SEE exercise provides relevant data to inform the economic model as the estimates elicited for the efficacy of BAC were in relation to the best available care as seen in the experts' practice, which would therefore be relevant to the NHS. Furthermore, we consider that it is more appropriate and relevant to the NHS to use the SEE estimates as they are related to the treatments included in the comparator arm in the model (also elicited by the SEE exercise experts).
What alternative approach has the EAG suggested?	The EAG uses the SEE exercise estimates from day 1 until the end of the time horizon, based on a GPPGA pustulation subscore of 0 or 1 in our base case. We explored the uncertainty around this assumption by conducting alternative scenario analyses, as follows: • Effisayil 1 trial (first week) + SEE exercise (GPPGA pustulation subscore of 0 or 1) • SEE exercise (GPPGA pustulation subscore of 0) • Effisayil 1 trial (first week) + SEE exercise (GPPGA pustulation subscore of 0)

	• Effisayil 1 trial (first week) + Effisayil 1 historical cohort (GPPGA pustulation subscore of 0 or 1, company's base case) • Effisayil 1 historical cohort (GPPGA pustulation subscore of 0 or 1, company's scenario)
What is the expected effect on the cost-effectiveness estimates?	Using the SEE exercise estimates for the efficacy of BAC from day 1 (based on a GPPGA pustulation subscore of 0 or 1) leads to an increase in the ICER of £311,357 per QALY (from -£167,783 to £143,574) for the EAG corrected company's base case.
What additional evidence or analyses might help to resolve this key issue?	Further clinical expert on which are the most appropriate estimates for the efficacy of BAC to be used in the economic model.

Issue 7 Proportion of patients treated as inpatients in the spesolimab arm

Report section	4.2.8.3
Description of issue and why the EAG has identified it as important	In the BAC arm, the proportion of patients treated as inpatients was based on the study by Wolf et al. (77.6%). For the spesolimab arm, the company assumed that only half of the proportion considered for BAC would be treated as inpatients (33.8%). This assumption was based on the relative reduction of 48.4% in the proportion of patients with an active flare (GPPGA pustulation subscore higher than 1) for spesolimab versus placebo in the Effisayil 1 trial. We note that the reduction of patients with an active flare directly leads to a reduction in the number of hospitalisations, as the clinical experts advising the company stated that patients with a resolved flare defined by a GPPGA pustulation subscore of 0 or 1 are discharged from hospital, and therefore we assume that they are not at risk of being hospitalised. Thus, it remains uncertain whether spesolimab has an additional residual benefit in reducing hospitalisation rates for patients not responding to treatment and therefore with an active flare. We note that there is no evidence showing such a benefit. Also, in the case of spesolimab having a benefit in reducing hospitalisation rates, the size of the benefit remains uncertain. Different inpatient rates for spesolimab changes the ICER considerably. So, in the absence of data, we consider the company's assumption is likely to be optimistic.
What alternative approach has the EAG suggested?	We have not changed our base case but we explored the uncertainty around this assumption by conducting alternative scenario analyses, as follows: • no reduction • reduction of 10% • reduction of 20% • reduction of 30%

What is the expected effect on the cost-effectiveness estimates?	Applying a reduction in inpatient rates for spesolimab versus BAC leads to the following results in the EAG corrected company's base case: • no reduction: ICER increases from -£167,783 to -£5,009 per QALY. • reduction of 10%: ICER increases from -£167,783 to -£37,564 per QALY. • reduction of 20%: ICER increases from -£167,783 to -£70,119 per QALY. • reduction of 30%: ICER increases from -£167,783 to -£102,674 per QALY.
What additional evidence or analyses might help to resolve this key issue?	Further clinical expert opinion on the potential residual benefit of spesolimab in reducing inpatient rates for patients with an active flare.

1.6 Other issues: summary of the EAG's view

There are two other issues to draw to the committee's attention. We deem these not sufficiently likely to affect decision making that they should be a 'key' issue, but we believe they should be noted.

- The full extent of the duration of the response to spesolimab is unknown. Duration of response is one of the outcomes listed in the NICE scope and the company's decision problem. The Effisayil 1 trial had a duration of 12 weeks and 60% of participants originally randomised to spesolimab had a GPPGA pustulation subscore of 0 (clear) at Week 12. However, we do not know how much longer this response endured beyond 12 weeks or whether spesolimab treatment has any effect on the length of the interval between GPP flares.
- We were not able to obtain any expert clinical input into our report. Of 20 clinicians or dermatology centres contacted, 12 responded but none were able to be of assistance (four were conflicted as they had contributed to the CS, three had no capacity to take on the task, and five were out of office for an extended period due to illness or work-related reasons). In addition to the areas noted above across our seven key issues where we think clinical expert opinion could be useful, there are other questions we would have asked clinicians to help inform the report. These include:
  - What are the prognostic factors associated with GPP flares. In particular do race or age affect either flare intensity or frequency or response to flare treatment?
  - Are the treatments in Figure 1 of this report a reasonable reflection of the standard of care for moderate-to-severe GPP flares in UK clinical practice. Would topical steroids be used alongside second- and third-line treatments?

- Is the Effisayil 1 trial population generalisable to the population of GPP patients experiencing moderate-to-severe flares in England? Are the characteristics of the economic model population representative of the patients who may receive spesolimab in UK clinical practice?
- If a patient was not treated in the first week of a flare what impact would this have? Would a delay in starting treatment mean the flare would take longer to resolve?
- Questions about the appropriateness of all the model assumptions.

1.7 Summary of EAG's preferred assumptions and resulting ICER

Based on the EAG's critique of the company's model (discussed in section 4.2.2), we have identified the following key aspects of the company base case with which we disagree. Our preferred model assumptions are the following:

- **Composition of the comparator arm**: inclusion of active treatments during the first week of the model (linked to the following change in the efficacy of BAC).
- **Efficacy of BAC**: SEE exercise estimates based on a GPPGA pustulation subscore of 0 or 1 from day 1 until the end of the time horizon (see Table 19).

Table 3 shows the cost-effectiveness results of applying the EAG preferred model assumptions to the company's base case including the PAS discount for spesolimab. Incorporating the EAG assumptions, the ICER for spesolimab versus BAC increases to £143,574 per QALY.

The changes that have the most significant impact on the cost-effectiveness results are changing the assumptions for the efficacy of BAC, the proportion of inpatients on spesolimab, and the proportion of patients in the spesolimab arm with active flare by the end of the time horizon.

Table 3 EAG's preferred base case results using PAS price for spesolimab

Treatment	Total		Incremental		ICER
	Cost (£)	QALYs	Cost (£)	QALYs	(£/QALY)
BAC	[redacted]	[redacted]			£143,574
Spesolimab	[redacted]	[redacted]	[redacted]	[redacted]	

BAC, best available care; QALY, quality adjusted life year; ICER, incremental cost-effectiveness ratio; PAS, patient access scheme.

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Boehringer Ingelheim Ltd on the clinical effectiveness and cost effectiveness of spesolimab (Spevigo) for treating adults with generalised pustular psoriasis presenting with flares. It identifies the strengths and weaknesses of the CS.

Clarification on some aspects of the CS was requested from the company by the EAG via NICE on 3 September 2024. A response from the company via NICE was received by the EAG on 18 September 2024 and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on generalised pustular psoriasis

Generalised pustular psoriasis (GPP) is a chronic autoinflammatory disease affecting the skin with additional systemic symptoms.[1,2] It is phenotypically distinct from the most common type of psoriasis, plaque psoriasis (psoriasis vulgaris (PV)) although some patients may have both types.[3,4] GPP is characterised by intermittent flares with partial or complete remission occurring in between each flare episode.[1] A GPP flare typically develops rapidly and affects large areas of any part of the body. The skin becomes covered with sterile pus-filled blisters which can merge to become 'lakes' and there can be associated redness, itching, pain and scaling. Systemic features of flares are associated with inflammation and include fever, swelling, malaise, joint pain, fatigue, headache, increased white blood cell count, increased neutrophil count (a type of white blood cell) and increased c-reactive protein.[1,2,5] A GPP flare can be life-threatening and requires emergency treatment.[1,2,6]

Although the actual cause of GPP is unknown, it is known that interleukin-36 (IL36) signalling is part of the pathogenic process. In the skin, IL-36 contributes to host defence through inflammatory response, but if it becomes dysregulated it can induce psoriatic-like skin disorder; genetic mutations of antagonist IL-36Ra are associated with GPP.[5,7]

The only published data available for prevalence and incidence of GPP in England are from a recent cohort study (POLARIS) carried out by Boehringer Ingelheim in which data in the Clinical Practice Research Datalink (CPRD) Aurum database were linked with Hospital Episode Statistics (HES) data, as reported in CS section B.1.3.2.1.[8] GPP is rare. In England, between 2008 and 2019, there were 206 incident cases of GPP (in comparison the same

study identified 84,081 cases of plaque psoriasis over the same period): incidence was stable with an age- and sex-adjusted incidence rate of 0.14 to 0.34 cases per 100,000 person-years, and prevalence increased from 1.61 per 100,000 patients in 2008 to 3.00 per 100,000 patients in 2019.[8] During the twelve years analysed, 50.5% of patients experienced one flare (requiring three or more days of hospitalisation), 7.8% of patients experienced two flares, and 4.4% of patients experienced three or more flares.[8] GPP occurs more frequently in women than men: the POLARIS study found that females were 62.5% of the GPP population in England, which is reflective of the proportion of females with GPP in other epidemiological studies and reviews.[9,10]

The disease course for GPP is variable: for each patient GPP flares can be of different levels of intensity or severity and can recur more or less frequently.[1] Therefore, the periods of remission (when skin is clear) or partial remission in between flares are of varying lengths.[1] This appraisal is focused on treating patients who are experiencing a GPP flare.

The causes of GPP flares are not known, however, some, but not all, patients can identify triggers for the flares such as stress, alcohol, infections or other illnesses, menstruation, pregnancy, environmental triggers, or stopping high-dose steroids.[2 5,6]

There are limited data for prognostic factors for GPP disease course or for treatment outcomes for GPP flares, and Clarification Response A25 refers to the global consensus statement (Puig et al. 2023)[1] which shows a lack of predictability for flare duration, severity and frequency. The UK extension of the global consensus study (Barker et al. 2024)[11] noted uncertainty around the prognostic value of age because this could potentially be correlated to complication-related morbidity rather than flare frequency. The EAG has not been able to discuss prognostic factors associated with GPP with any clinical experts to verify this, and we agree that scientific literature investigating prognostic factors is lacking.

2.2.2 Background information on spesolimab

Spesolimab, brand name Spevigo, is a first in class humanised antagonistic monoclonal immunoglobulin 1 antibody. It blocks human interleukin-36 receptor (IL-36R) signalling by binding to IL-36R which prevents activation of the autoinflammatory response.

Spesolimab received a conditional marketing authorisation from the MHRA in July 2023. The authorisation was conditional because the data on treatment of subsequent flares was not considered comprehensive. Another study, Effisayil REP, is in progress to investigate the treatment of repeated flares with spesolimab.

Spesolimab is indicated for treatment of flares in adult patients with GPP as monotherapy (i.e. it should not be used in combination with other GPP treatments such as systemic immunosuppressants).[12] It is administered as an intravenous infusion of 900 mg. If flare symptoms persist, an additional 900 mg dose may be administered after one week. In response to clarification question A1 the company stated that there are no data on the use of a third dose of spesolimab for a specific flare and therapy choice if a flare remains unresolved after two doses of spesolimab would be at the discretion of the treating physician. It is not clear to us if clinicians would view a third spesolimab dose as an option in this circumstance. The company confirmed that they have not recommended a specified minimum period between spesolimab dosage courses for subsequent flares (response to clarification question A1ii) so there does not appear to be a maximum dose of spesolimab that can be received within a specified time period.

2.2.3 The position of spesolimab in the treatment pathway

There are no specific guidelines for treating GPP and GPP flares. The NICE Clinical Guideline for the assessment and management of psoriasis (CG153) refers specifically to GPP only in the context of referral to dermatology specialist advice and treatment which should be immediate and same-day, and in consideration of the use of acitretin for people with pustular forms of psoriasis.[13] Apart from spesolimab, there are no licensed treatments in the UK specifically for GPP so treatment typically draws on the licensed therapies used for treating other forms of psoriasis. The only information on the current treatment pathway for patients experiencing a GPP flare is from a global consensus study (Puig et al. 2023)[1], for which there is a UK consensus panel published (Barker et al. 2024),[11] and from the company's structured expert elicitation (SEE).[14]

The company's outline of the treatment pathway for GPP flares is shown in CS Figure 4, reproduced below in Figure 1. It shows a range of therapies are used at first-, second- and third-line or later including topical corticosteroids, systemic non-biological therapies such as ciclosporin, and biological therapies such as TNF-alpha inhibitors (infliximab), an IL-17 inhibitor (secukinumab), and IL-23 inhibitors (guselkumab, ustekinumab).

Figure 1 Clinical pathway of care for moderate-to-severe GPP flares in the UK

Source: Reproduced from CS Figure 4 with title amended by the EAG.
Abbreviations: 1L, first line; 2L, second line; 3L+, third- or later-line; GPP, generalised pustular psoriasis.

The company acknowledge that this diagram is simplified and represents an undefined clinical pathway. We agree, and this is due to a lack of clinical guidelines or licensed treatments as noted above. The pathway in Figure 1 does not appear to take into account whether the severity of a flare determines choice of treatment. Clarification Response A2i confirms that the pathway is based on evidence from the company's SEE (critiqued in section 3.3.2) and represents UK and Ireland clinical expert opinion of best available care for treatment of moderate-to-severe flares. The EAG has been unable to obtain further clinical expert opinion for verification, however, this illustration of the treatment pathway for flares does not contradict the global consensus study nor the UK additional panel to the global consensus study because they do not state what treatments should be used and instead agree on treatment goals and that treatments should have a rapid onset of action.[1,11]

The pathway in Figure 1 does not include the proposed position of spesolimab. The SmPC does not specify a line of treatment, but CS section B.3.2.3.1 states that spesolimab is expected to be used in first-line treatment of GPP flares, confirmed in Clarification Response A2ii, which further elaborates that it is expected to displace the current first-line and subsequent-line treatment options, providing an alternative option to the multiple lines of treatment approach.

The SmPC indicates use of spesolimab as a monotherapy.[12] It does not recommend using spesolimab in combination with immunosuppressants, including biologics, because combination treatment has not been evaluated systematically. Currently, topical steroids can be used alongside the other therapies ([redacted] at first line in Figure 1), but it is unclear whether they can be used alongside spesolimab. This is because the pivotal trial (see section 3.2.1 below) did not permit the use of therapies in addition to spesolimab: topical treatments could not be initiated in the week prior to visit 2 (treatment initiation) and washout periods ranging from seven days to 5.5 months were required for other medications (CS Appendix N). Therefore, it is unclear whether patients receiving preventive topical or systemic therapies for GPP will be able to receive emergency treatment for flares with spesolimab (when there is no opportunity for a washout period) and to what extent this affects the size of the eligible population. The EAG did not have access to clinical expertise to clarify the proportion of the UK GPP population that are on maintenance topical steroids or systemic therapies.

EAG conclusion on GPP and spesolimab in the treatment pathway

GPP is a rare systemic inflammatory disease and GPP flares require emergency treatment for which the company propose spesolimab as first-line therapy. The company accurately reports disease, pathogenesis and symptoms, and provides epidemiologic data relevant to England from their analysis of national CPRD Aurum and HES data. Due to the limitations of clinical guidelines which focus on PV, the company conducted a structured expert elicitation[14] which has informed the treatment pathway. There is limited information around some aspects of the dosing of spesolimab. For example, what is the minimum interval between treating subsequent flares with spesolimab and is there a maximum dose of spesolimab that can be received within a specified time period? There is also a lack of evidence on concomitant use of spesolimab with topical steroids and systemic treatments because these were not permitted in the clinical trial.

2.3 Critique of the company's definition of the decision problem

Table 4 summarises the decision problem addressed by the company in the CS in relation to the final scope issued by NICE and the EAG's comments on this.

Table 4 Summary of the decision problem

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Population	Adult patients with generalised pustular psoriasis presenting with flares	Adult patients with generalised pustular psoriasis presenting with flares	Not applicable	CS Table 2 states the therapeutic indication is for the treatment of flares in adult patients with generalised pustular psoriasis as a monotherapy. However, in the company's pivotal trial, Effisayil 1,[a] only adult patients with GPP experiencing moderate-to-severe intensity flares were randomised to treatment.[b] The clinical evidence is therefore drawn from a narrower population than defined in the NICE scope and the company's decision problem. The patient cohort that start in the economic model are in the moderate-to-severe flare health state.
Intervention	Spesolimab	Spesolimab	Not applicable	The intervention matches the NICE scope and in the pivotal Effisayil 1 trial, the licensed dose of spesolimab was used.
Comparators	Established clinical management without spesolimab which may include: Systemic non-biological therapies such as ciclosporin Biological therapies (such as TNF-alpha	Established clinical management without spesolimab	Not applicable	The comparator in the company's Effisayil 1 trial was placebo. Prior to randomisation patients had to discontinue biologic therapies, systemic non-biological therapies and other treatments (e.g. phototherapy and topical treatments). Therefore, there is no direct comparative effectiveness evidence from the company's trial for spesolimab versus established clinical management without spesolimab. The economic model assumes that patients in the

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
	inhibitors, IL-17 and IL-23 family inhibitors)			comparator arm receive no treatment during the first week and thereafter receive treatment from a basket of options (TNF inhibitors, IL-23 inhibitors, IL-17 inhibitors and ustekinumab in the base case) identified by a group of UK clinical experts who participated in the company's structured expert elicitation (SEE) exercise. Therefore, the EAG note that the comparator in the economic model is not fully aligned with the NICE scope and company definition of the decision problem
Outcomes	The outcome measures to be considered include: • Symptoms specific to GPP including pain • Severity of flares • Mortality • Response rate • Duration of response • Relapse rate • Adverse effects of treatment • Health-related quality of life	The outcome measures to be considered include: • Symptoms specific to GPP including pain • Severity of flares • Mortality • Response rate • Duration of response • Relapse rate • Adverse effects of treatment • Health-related quality of life	Not applicable	The outcomes listed by the company match those in the NICE scope and the majority are clearly reported in the CS. An exception is duration of response for which there is limited data due to the length of the Effisayil 1 trial (response was still ongoing for a proportion of patients at the end of this 12-week study).
Economic analysis	Not included in CS Table 1			The company's cost-utility analysis adheres to the NICE reference case. Whether the time horizon used is sufficiently long to reflect all important differences in

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
				costs or outcomes between spesolimab and BAC is unclear (Key Issue 5).
Subgroups	Not included in CS Table 1. No subgroups are noted in the NICE scope.			No comments.
Special Considerations including issues related to equity or equality	Not included in CS Table 1. No special considerations are noted in the NICE scope.			No comments.

Source: CS Table 1 with the addition of EAG comments in the final column.
BAC, best available care; CS, company submission; EAG, External Assessment Group; GPP, generalised pustular psoriasis; IL-17, interleukin-17; IL-23, interleukin-23; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; SEE, structured expert elicitation; TNF, tumour necrosis factor.
[a] Throughout the CS the Effisayil trial names are followed by the trademark symbol (TM in superscript) but because the NICE style guide does not permit the use of this symbol we have not included this in our report and refer solely to the Effisayil 1, Effisayil 2,Effisayil ON and Effisayil REP trials.
[b] A moderate-to-severe intensity flare was defined as a Generalised Pustular Psoriasis Physician Global Assessment (GPPGA) score of at least 3 (moderate) and presence of fresh pustules (new appearance or worsening of existing pustules), and a GPPGA pustulation subscore of at least 2 (mild), and at least 5% of body surface area covered with erythema (abnormal redness of the skin or mucous membranes) and the presence of pustules (CS section B.2.3.1).

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review

The company conducted a systematic literature review (SLR) to identify all relevant evidence of clinical efficacy and safety associated with spesolimab and current available treatments for GPP (CS Appendix D). Bibliographic searches of the main healthcare databases and clinical trials registers were run in December 2022, with update searches run in May 2024. Several specific dermatology and psoriasis conferences were hand searched from 2022 to May 2024. Published search filters for randomised controlled trials (RCTs) and observational studies were used, and the searches were generally comprehensive. We do not believe that any relevant studies would be missed by the searches.

Screening was carried out according to eligibility criteria that were broader than the NICE scope and company decision problem because they included a population of all GPP patients (CS Appendix D Table 10) not just GPP patients presenting with flares (CS Table 1). A total of 79 unique studies were identified (CS Appendix D.1.2) or 81 unique trials according to CS section B.2.2; we are unable to explain the reason for this inconsistency. A further screening stage was carried out to identify the studies "*that reported data specific to GPP flare*s" to align with the company decision problem (CS Appendix D.1.2), and Clarification Response A4 confirms that the eligibility criterion was a population including patients experiencing a GPP flare. Therefore, the studies identified by this additional screening stage were identified based on the population only, not because of any specific outcomes related to GPP flare treatment. The company assessed fifteen studies as including data specific to a population experiencing a GPP flare (CS Appendix Table 11 and CS section B.2.2). A list of the excluded studies with reasons for exclusion was provided in Clarification Response A3.

Of the 15 studies that "*reported data specific to GPP flares and were thus aligned to the decision problem*" (CS Appendix D.12), the company identified four studies that evaluated the efficacy and safety of spesolimab. Of these, the Effisayil 1 study assessing spesolimab IV infusion for treatment of GPP flares, was considered by the company as most relevant to this submission (section 3.2 below). The other three studies evaluating spesolimab were the company's Phase I, Effisayil 2 and Effisayil ON studies. The Phase 1 and Effisayil ON open label extension study safety results are considered in section 3.2.5.5. Three of the remaining 11 studies, were considered further as a source of possible comparator data for an indirect treatment comparison (ITC) (see section 3.3 below).

In Clarification Response A5 the company confirm that the quality assessment, relevance of the evidence base, and heterogeneity assessment were performed for all 79 studies included in the broader scope of the SLR. This makes it more difficult to have clarity about the methods and criteria that were used to identify the evidence specific to the company's decision problem for this appraisal.

The EAG's summary appraisal of the company's review methods is provided in Appendix 1.

3.2 Critique of studies of the technology of interest, the company's analysis and interpretation (and any standard meta-analyses of these)

3.2.1 Included studies

The company's systematic literature review identified the four pivotal trials of spesolimab for the treatment of adults with GPP but only one of these, the Effisayil 1 trial, provides evidence fully relevant to this technology appraisal and informs treatment effectiveness in the economic model. The outcomes from Effisayil 1 that inform that company model are:

- Comparative effectiveness data for Week 1 in terms of the response to treatment (GPPGA pustulation subscore 0 or 1 representing flare resolution) for both the spesolimab and BAC treated model cohorts. The Effisayil 1 trial also informs efficacy of the spesolimab treated cohort beyond Week 1 to the end of the model time horizon.
- Rates of serious infection and liver injury (adverse events) in the spesolimab model cohort
- Utility values for the active flare state for both the spesolimab and BAC treated model cohorts. We note that the Effisayil 2 trial informed the utilities for the resolved flare health state (further details in section 4.2.7.2)

The key features and roles of the four spesolimab studies in this appraisal are summarised in Table 5.

Table 5 Summary of the company trials of spesolimab and their role in the appraisal

	NCT02978690	Effisayil 1 (NCT03782792)	Effisayil 2 (NCT04399837)	Effisayil ON (NCT03886246)
Role in this appraisal	Evidence provided in Appendix F for information.	Key source of clinical effectiveness evidence and informs the economic model.	Supportive clinical effectiveness evidence. Effisayil 2 informs the utilities for the resolved flare health state in the economic model. Effisayil ON does not inform economic model.	
Study type	Phase I, PoC study.	Phase II double-blind[a] RCT.	Phase IIb double-blind RCT	Open-label extension study
Patient group	Adult patients with GPP presenting with flare involving 10% or more of their BSA with erythema and the presence of pustules and a GPPGA score of 3 or higher.	Adult patients with GPP presenting with flares of moderate-to-severe intensity (GPPGA score of at least 3 and presence of fresh pustules, and a GPPGA pustulation subscore of at least 2, and at least 5% of BSA covered with erythema and the presence of pustules).	Patients aged ≥ 12–75 years at screening (weight ≥ 40 kg) with a history of at least two prior presentations of GPP flares with fresh pustulation. At screening and randomisation a GPPGA score of 0 or 1.[15]	Patients who completed treatment in either Effisayil 1 or Effisayil 2 without premature discontinuation (patients with evidence of moderate-to-severe flare at screening are excluded).
Intervention	Spesolimab IV infusion (10mg/kg, single dose) for treatment of GPP flares.	Spesolimab IV infusion, 900mg single initial dose for treatment of GPP flares. A second open-label dose at Day 8 if prespecified criteria were met[b] and one rescue dose for a recurrence of GPP flare.[c] were also permitted.	Spesolimab SC with randomisation to one of 3 arms for prevention of GPP flares: 600mg LD then 300mg q4w 600mg LD then 300mg q12w 300mg LD then 150mg q12w	Spesolimab SC (300mg) as a maintenance treatment: q4w q6w q12w
Comparator	None	Placebo. An open-label dose of spesolimab at Day 8 was permitted if prespecified criteria were met[b] and one rescue dose for a recurrence of GPP flare.[c] was also permitted.	Placebo SC q4w	None
Duration	20 weeks	12 weeks	48 weeks	252 weeks

Source: EAG generated table based on information presented in CS section B.2.2, CS Appendix F.1.1 and cited references.
BSA, body surface area; GPP, generalised pustular psoriasis; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment; IV, intravenous; LD, loading dose; PoC, proof-of-concept; q4w, every 4 weeks; q6w, every 6 weeks; q12w, every 12 weeks; RCT, randomised controlled trial; SC, subcutaneous.

[a] Note the study was double-blind for the first week only and placebo-controlled for the first week only because most placebo arm participants received open-label spesolimab at Day 8.
[b] The criteria were a GPPGA total score and a GPPGA pustulation sub score of at least 2, and no receipt of escape treatment during Week 1.
[c] The criteria for a rescue dose of spesolimab after Day 8 and through to Week 12 were at least a 2-point increase in GPPGA score and GPPGA pustular sub score after achieving a clinical response to initial treatment (randomised treatment, escape treatment or open-label spesolimab)

3.2.1.1 Study characteristics

Effisayil 1 is the company-sponsored, multi-centre, randomised, double-blind phase II study of spesolimab versus placebo for the treatment of adult patients with GPP flares of moderate-to-severe intensity. The design of Effisayil 1 is described in CS section B.2.3.1 and shown in CS Figure 5. Patient flow is described in CS section B.2.3.2 and shown in CS Figure 6.

Eligible patients aged 18-75 years were first enrolled into the study if they met at least one of the three eligibility criteria summarised in CS section B.2.3.1 (the full trial eligibility criteria for enrolment are provided in CS Appendix N). Patients could be enrolled into the study if they satisfied the eligibility criteria regardless of their current flare status. However, to be randomised to either spesolimab or placebo, patients had to be currently experiencing a GPP flare of moderate-to-severe intensity. This was defined as:

- A GPPGA score of at least 3 (moderate), and
- Presence of fresh pustules (new appearance or worsening of existing pustules), and
- A GPPGA pustulation subscore of at least 2 (mild), and
- At least 5% of body surface area covered with erythema (abnormal redness of the skin or mucous membranes) and the presence of pustules

Prior to randomisation, participants had to stop receipt of their current therapies as described in detail in CS Appendix N Table 62 and were not permitted to use these therapies during the trial. In brief for the main types of therapy:

- Biological therapies had to stop at least 7 days (range 7 days to 5.5 months depending on the biological therapy) prior to randomisation.
- Other systemic immunomodulating treatments and other systemic psoriasis treatments were not permitted 30 days prior to randomisation.
- Methotrexate, cyclosporine and retinoids could not be started or have the dose escalated in the 2 weeks prior to randomisation and these therapies had to be discontinued prior to receipt of randomised treatment.
- Topical treatments and phototherapy for psoriasis or any other skin condition could not be started in the week prior to randomisation.
- Investigational products for psoriasis had to stop 3 months prior to randomisation and other investigational products or devices were not permitted 30 days prior to randomisation.
- IL-36R inhibitors were not permitted before or during trial participation.

Fifty-three trial participants were randomised 2:1 to either spesolimab (n=35) or placebo (n=18). In response to clarification question A12 the company provided a breakdown of the numbers of participants by country (both for enrolled and for randomised participants). The proportions of participants randomised from the various countries were: Malaysia (23%), France (19%), Tunisia (13%), China (11%), Germany (9%), Taiwan (9%), United States (6%), Japan (4%), Singapore, Switzerland and Thailand each 2% (percentages calculated by the EAG).

The key features of treatment and trial events after randomisation are shown in CS Figure 5 and summarised briefly below:

- Day 1: participants randomised to spesolimab received a single 900mg IV dose. This is the dose recommended in the Summary of Product Characteristics (SmPC). Participants randomised to placebo received a single IV dose.
- Days 2-7: Escape treatment (physician's choice) was permitted in the case of disease worsening.
- Day 8:
  - Primary and key secondary endpoint analyses were conducted. In these analyses participants who had received escape treatment during days 2-7 were assumed not to have responded.
  - Participants from both trial arms with a GPPGA score of at least 2 and a GPPGA pustulation subscore of at least 2 (flare graded at least as mild, confirmed in response to clarification question A11), and who had not received escape treatment during days 2-7 were eligible to receive an open-label spesolimab 900mg IV dose (i.e. participants from the placebo arm meeting the criteria could receive open label spesolimab and participants from the spesolimab arm meeting the criteria could receive a second dose of spesolimab).
- Day 8 – week 12: Safety endpoints were assessed.
- Day 9 – week 12: Participants with a recurrence of GPP flare (defined as "*at least a 2-point in GPPGA score and GPPGA pustular subscore after achieving a clinical response to initial treatment*") could receive a rescue dose of spesolimab 900mg IV. In response to clarification question A9, the company confirmed that clinical response to initial treatment was defined as a GPPGA total score of 0 or 1, and we note that the criteria for a flare mean that it is graded at least as mild. All participants (regardless of which randomised treatment they had received and whether they had received escape treatment or open-label spesolimab on Day 8) were able to receive

a rescue dose of spesolimab if they met the criteria. Only one rescue dose of spesolimab was permitted per patient.

- Beyond trial end at week 12: If participants with clinical improvement completed the trial up to week 12 and were without flare symptoms, they were eligible to enter the open-label extension study Effisayil ON.

As CS Figure 5 and the summary above indicate, it was possible for participants to receive three spesolimab doses during the trial – i) in the arm randomised to spesolimab, ii) open-label spesolimab on Day 8 providing criteria met, iii) rescue medication due to a recurrence of GPP flare. In response to clarification question A10 the company state that two patients received three spesolimab doses for these reasons.

Participant flow through the Effisayil 1 trial is described in CS section B.2.3.2 and shown in CS Figure 6. This shows that of the 53 participants who were randomised, 52 completed the first week of the trial. At day 8, 12 participants (34%) in the spesolimab group received open-label spesolimab (i.e. a second dose) and 15 of the 18 participants (83%) in the placebo group also received open-label spesolimab (their first dose).

3.2.1.2 Patients' baseline characteristics

Patients' baseline data are reported for demographic characteristics, disease severity characteristics associated with skin condition and quality of life, and IL36RN mutation in CS Table 9. Trial publication Table S1 additionally reports geographic region and systemic disease characteristics (baseline temperature (fever/no fever), white blood cell count, and C-reactive protein level).[16] We have combined them, and added [redacted] from the CSR, in Table 6 below:

Table 6 Patients' baseline characteristics in Effisayil 1

Characteristic	Spesolimab (n = 35)	Placebo (n = 18)
Age, mean years (SD)	43.2 (12.1)	42.6 (8.4)
Weight, kg (SD)	73.7 (24.0)	68.8 (26.6)
Female, n (%)	21 (60)	15 (83)
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Characteristic	Spesolimab (n = 35)	Placebo (n = 18)
Race, n (%)[a]		
Asian	16 (46)	13 (72)
White	19 (54)	5 (28)
Region, n (%)		
US, n (%)	2 (5.7)	1 (5.6)
Asia, n (%)	14 (40.0)	13 (72.2)
Europe, n (%)	14 (40.0)	2 (11.1)
Africa, n (%)	5 (14.3)	2 (11.1)
GPPGA total score, n (%)[b]		
3 (moderate)	28 (80)	15 (83)
4 (severe)	7 (20)	3 (17)
GPPGA pustulation subscore — n (%)[c]		
2 (mild)	6 (17)	5 (28)
3 (moderate)	16 (46)	7 (39)
4 (severe)	13 (37)	6 (33)
Median GPPASI total score (IQR) [d]	**27.4 (15.5–36.8)**	**20.9 (12.0–32.0)**
Median DLQI score (IQR)[e]	**19.5 (16–25)**	**19.5 (14–24)**
Median pain VAS score (IQR)[f]	**79.8 (70.5–87.8)**	**70.0 (50.0–89.4)**
Median PSS score (IQR)[g]	**11.0 (9–12)**	**10.5 (9–11)**
Median FACIT-Fatigue score (IQR)[h]	**14.0 (7–28)**	**18.0 (6–33)**
***IL36RN* mutation, n (%)[i]**	**8 / 29 (28)**	**6 / 17 (35)**
Baseline temperature[j]		
< 38.5 C, n (%)	29 (85.3)	18 (100.0)
≥ 38.5 C, n (%)	5 (14.3)	0 (0.0)
Baseline leucocytes[j] [k]		
< 10,000 µL, n (%)	14 (40.0)	6 (33.3)
≥ 10,000 µL, n (%)	18 (51.4)	10 (55.5)
Baseline C-reactive protein[j] [l]		
≥ 3mg/L-70 mg/L, n (%)	20 (58.8)	12 (66.7)
≥ 70 mg/L, n (%)	11 (31.4)	4 (22.2)

Characteristic	Spesolimab (n = 35)	Placebo (n = 18)
Sources: CS Table 9; Bachelez 2021 Table S1;[16] Effisayil 1 CSR.[17] Footnotes relabelled and reordered by EAG. Abbreviations: [redacted]DLQI, Dermatology Life Quality Index; FACIT, Functional Assessment of Chronic Illness Therapy; GPPASI, Psoriasis Area and Severity Index for Generalised Pustular Psoriasis; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment; IQR, interquartile range; n, number; PSS, Psoriasis Symptom Scale; SD, standard deviation; VAS, Visual Analogue Scale. [a] Race was reported by the patient; [b] Scores on the GPPGA range from 0 (clear skin) to 4 (severe disease); [c] GPPGA pustulation subscores range from 0 (no visible pustules) to 4 (severe pustulation); [d] Scores on the GPPASI range from 0 (least severe) to 72 (most severe); [e] Scores on the DLQI range from 0 (no effect) to 30 (extremely large effect) – spesolimab n = 34; [f] Scores on the pain VAS range from 0 (no pain) to 100 (very severe pain); [g] Scores on the PSS range from 0-16, with higher scores indicating more severe symptoms; [h] Scores on the FACIT-Fatigue range from 0-52, with lower scores indicating greater impact of fatigue on daily activities; [i] Samples from seven patients were missing; [j] Data for 52 participants (34 spesolimab, 18 placebo); [k] Four patients had missing values at baseline; [l] Five patients had missing values at baseline.		

The Effisayil 1 trial population may not be representative of UK patients because 55% of patients in Effisayil 1 were Asian, 51% were at sites in Asia, compared to only 30% from Europe (France, Germany and Switzerland). There were no UK sites or patients. The company's clinical experts validated the evidence from Effisayil 1 as relevant (CS section B.2.12.3), however, the EAG was unable to obtain expert opinion to independently verify whether the baseline characteristics of the Effisayil 1 trial population are similar to GPP patients experiencing flares in England.

Characteristics were not balanced between trial arms. There were proportionally more Asian people in the placebo arm (72%) compared to the spesolimab arm (46%). The EAG was unable to obtain expert opinion on whether race could be a prognostic factor for treatment outcomes in GPP flares. Disease severity scores according to the GPPASI score and GPPGA pustulation subscore were slightly worse in the spesolimab arm, proportionally more patients had fever (Suppl. Table 1)[16] in the spesolimab arm (14% compared to 0%), and therefore there could be a bias in favour of the placebo arm. It may be that the characteristics aren't entirely balanced because of the small sample size and the especially small placebo group.

It is not clear whether all possible prognostic factors for flare severity and duration were assessed at baseline because there are very limited data about which prognostic characteristics predict flare severity and duration (as confirmed in company response to clarification question A25, see also section 2.2.1 above). Therefore, we cannot be certain if there was any further bias or consistent bias in favour of the placebo arm. Baseline

characteristics that informed participant selection for the pre-specified subgroup analyses but which are not reported in the CS are: BMI, presence of plaque psoriasis, background GPP treatment prior to randomisation, and the Japanese Dermatological Association (JDA) GPP severity score (though this characteristic is less relevant to UK clinical practice). These characteristics are reported in CSR Tables 10.4.1:1, 10.4.3:3, 10.4.5:1 and 10.4.2:1 respectively.[17] There are [redacted].

EAG conclusion on included studies

The Effisayil 1 trial provides evidence fully relevant to this appraisal and informs the economic model. It was one of four pivotal trials of spesolimab identified by the company's systematic review. Supportive clinical evidence is reported in the CS from two other studies, Effisayil 2 and Effisayil ON, and results from the fourth study, NCT02978690, are provided for information. The Effisayil 2 trial informed the utilities for the resolved flare health state (see section 4.2.7.2). Effisayil 1 compared spesolimab versus placebo for the treatment of adult patients with GPP flares of moderate-to-severe intensity. Therefore, the trial evidence is drawn from a narrower population than defined in the NICE scope and company's decision problem (Key Issue 1). The comparator in the Effisayil trial is placebo, not established clinical management without spesolimab which is the comparator defined in the NICE scope and company decision problem (Key Issue 2). No UK participants were recruited to the trial and a high proportion of the participants were Asian which could impact the generalisability of the evidence to patients experiencing moderate-to-severe GPP flares in England. Knowledge about prognostic factors for flare severity and duration is limited, and because we were unable to obtain clinical expert opinion to help inform this report there is uncertainty about whether the high proportion of Asian participants in the trial and the imbalance in the proportions of Asian people between trial arms (spesolimab arm 46% vs placebo arm 72%) could have affected treatment outcomes.

3.2.2 Risk of bias assessment

The company critically appraised Effisayil 1 using the RCT checklist from the Centre for Reviews and Dissemination. The level of risk of bias is reported in CS Appendix Table 12, and supporting statements are provided in CS Table 11. The company judged the risk of bias in Effisayil 1 as low in all domains except for detection bias where the risk of bias was 'unclear' but without any justification made, and this judgement is not reported in CS Table 11. The EAG agree the Effisayil 1 trial is mainly at low risk of bias overall, although because

prognostic factors for GPP flares are not defined it is unclear which characteristics are most important to be balanced between groups (and there was slight imbalance for some characteristics). Other characteristics were also not balanced between trial arms (GPPASI, GPPGA pustulation subscore and proportion with fever were slightly worse in the spesolimab arm) so there may be bias in favour of the placebo arm. A summary of our assessments for each domain are presented alongside the company assessments in Appendix 2 of this report.

CS Table 11 adds an assessment question asking whether the study reflects routine clinical practice in England which refers to generalisability rather than risk of bias and we discuss this aspect of the study in section 3.2.1.2.

3.2.3 Outcomes assessment

The company lists the reported outcomes that are specified in the decision problem in CS Table 8 with the outcomes that inform the economic model indicated by bold type: GPPGA pustulation subscore, the occurrence of treatment emergent adverse effects (TEAEs), EQ-5D. A full list of outcomes for the Effisayil 1 trial is provided in CS Appendix N Table 62 and a description (including summary of scoring) of the GPPGA, GPPASI, PSS, Pain VAS, DLQI, FACIT-Fatigue and EQ-5D-5L is provided in CS Appendix N Table 63. We therefore focus here on the efficacy outcomes, the EQ-5D and safety outcomes as these inform the economic model, and a description of the Pain VAS, FACIT-Fatigue, DLQI, and PSS outcomes is provided in Appendix 3.

3.2.3.1 Efficacy outcomes

Efficacy in the Effisayil 1 trial was primarily measured using the GPPGA pustulation subscore and the GPPGA total score. A GPPGA pustulation subscore of 0 at Week 1 was the primary outcome of the trial and a GPPGA total score of 0 or 1 at Week 1 was the key secondary endpoint (CS section B.2.3.1). For the economic model, response to treatment was modelled in terms of flare resolution, defined as a GPPGA pustulation subscore 0 or 1 (CS section B.3.3.1).

The GPPGA is a relatively recent modification of the Physician Global Assessment (PGA) which has been commonly used in clinical trials.[18] It was first described in use in the 2019 publication[19] for the NCT02978690 proof-of-concept study of spesolimab. The PGA has been criticised for being variable between trials, in terms of range of the scales used to evaluate the three aspects of psoriasis being assessed (erythema, induration and scale) and the definitions used for each point on the scale. In the GPPGA, which appears to have been developed by the company, the induration component of the PGA has been replaced by

pustulation. Since the Effisayil 1 trial was completed, the GPPGA has been assessed and validated in a company sponsored study. In this study GPP-experienced dermatologists participating in the Effisayil 2 trial were trained on the GPPGA using a guide that includes descriptions and photographs to illustrate how each component of the GPPGA is scored. Then they were asked to score 16 representative photographs twice (10-14 days apart) taken from patients with GPP who participated in the Effisayil 1 trial.[20] The inter-rater reliability between dermatologists (26 completing the first assessment and 20 completing both assessments), assessed by the intraclass correlation coefficient, was described as 'generally excellent'.[20]

The way in which erythema, pustules and scaling are scored and contribute to the GPPGA total score is shown in Table 7. To achieve a GPPGA total score of 0 or 1 at Week 1 (key secondary endpoint) the mean composite score has to be less than 1.5. This can be achieved by 35 different combinations of the component parts of the score as shown in Appendix 4.

Because the GPPGA is a relatively new measure it is not used in clinical practice.[14] When the company conducted their SEE exercise for efficacy[14] the participating clinicians were provided with pictures and definitions of the GPPGA components.

Table 7 Scoring for the Generalised Pustular Psoriasis Physician Global Assessment (GPPGA).

Score	Erythema	Pustules	Scaling
0 (clear)	Normal or post-inflammatory hyperpigmentation	No visible pustules	No scaling or crusting
1 (almost clear)	Faint, diffuse pink, or slight red	Low-density occasional small discrete pustules (noncoalescent)	Superficial focal scaling or crusting restricted to periphery of lesions
2 (mild)	Light red	Moderate-density groups discrete small pustules (noncoalescent)	Predominantly fine scaling or crusting
3 (moderate)	Bright red	High-density pustules with some coalescence	Moderate scaling or crusting covering most or all lesions
4 (severe)	Deep fiery red	Very-high-density pustules with pustular lakes	Severe scaling or crusting covering most or all lesions

Each component is graded separately and then the average composite mean score is used to produce a total GPPGA score		
Mean composite score	**Total GPPGA score**	**Description**
0	0	Clear
> 0 to < 1.5	1	Almost clear
≥ 1.5 to 2.5	2	Mild
≥ 2.5 to 3.5	3	Moderate
≥ 3.5	4	Severe

Source: EAG table drawing on information published in Burden et al. 2023[20] and CS Appendix N Table 63.
GPPGA, Generalised Pustular Psoriasis Physician Global Assessment

The GPPGA pustulation subscore of 0 at week 1 (primary outcome) is the score given based on the physician's assessment of pustules only. As Appendix 4 shows, a patient with a GPPGA pustulation subscore of 0 could still be assessed as having moderate GPP, for example, if their erythema and scaling scores were both 4 (severe).

3.2.3.2 EQ-5D

Effisayil 1 collected EQ-5D-5L assessments daily for the first week, then weekly until weeks 4, 8 and 12. No results are reported for the group originally randomised to placebo who received open label spesolimab at Day 8. MCID for the EQ-5D health index was estimated at 0.07 (CS Figure 15 footnote). The data were mapped to EQ-5D-3L using the NICE preferred mapping function (CS section B.3.4.2). The company suggest this outcome measure is not particularly sensitive in acute conditions, nonetheless the data inform the model as required for utilities but not for disutilities due to adverse events (CS Table 16). The EAG believe it is unlikely that that any single PRO for quality of life fully covers all aspects of a GPP flare in order to be an alternative to EQ-5D, and no alternative to EQ-5D was suggested by the company.

3.2.3.3 Safety outcomes

Effisayil 1 assessed mortality and adverse effects of treatment; occurrence of treatment emergent adverse events was used in the economic model (CS Table 8). Severity of adverse events was measured appropriately using the Rheumatology Common Toxicity Criteria (CS Table 13). Adverse events of special interest were pre-specified as hepatic injury, systemic hypersensitivity reactions (e.g. infusion reactions and anaphylactic reactions), severe infections, and opportunistic and mycobacterium tuberculosis infections (CSR 9.5.3.2.1),[17] which are relevant to IV infusion administration and a treatment that acts on the immune system.

The CS additionally reports adverse events from the Phase I spesolimab study in CS Appendix F.1.2, and interim safety results from the Effisayil ON open-label extension study for patients receiving maintenance spesolimab in CS section B.2.10.3 (also published in the Navarini et al. 2023 conference abstract).[21]

EAG comment on outcomes assessment

There is no standard definition of GPP flare resolution but the company's use of the GPPGA pustulation subscore seems acceptable. A GPPGA pustulation subscore of 0 (clear, no visible pustules) at Week 1 was the primary outcome of the trial and a GPPGA total score of 0 (clear) or 1 (almost clear) at Week 1 was the key secondary endpoint. A GPPGA pustulation subscore of 0 or 1 is the outcome that informs the economic model. The GPPGA (total score and subscores) do not appear to be used in clinical practice but clinical expert opinion on whether a GPPGA pustulation subscore of 0 or 1 appropriately reflects GPP flare resolution would be welcome.

The patient reported outcomes used in the Effisayil 1 trial are not specific to GPP but two, the Psoriasis Symptom Scale and the Dermatological Quality of Life Index, are specific for psoriasis and skin conditions respectively. Three generic measures were also used (pain VAS, Functional Assessment of Chronic Illness Therapy – Fatigue scale, and EQ-5D) with the EQ-5D informing the economic model.

3.2.4 Statistical methods of the included studies

The company's statistical methods for Effisayil 1 are reported in CS section B.2.4, CS Table 10 and section 9.7 of the clinical study report (CSR)[17]. We summarise the methods in Table 8 below.

Table 8 Summary and EAG critique of the statistical methods used in Effisayil 1

Analysis populations
Analyses were conducted on the randomised set for the primary outcome (GPPGA pustulation subscore, Week 1) and the key secondary outcome (GPPGA total score, Week 1) which compared spesolimab against placebo (CS section B.2.4). This is an appropriate ITT analysis. [redacted].
After Week 1, on Day 8 15/18 participants randomised to the placebo group received open label spesolimab and were therefore imputed with non-response or worst possible outcome. The company argue that an ITT analysis of outcomes after Week 1 would be

noninformative. Therefore, analyses are reported in the CS for four post-hoc groups where the original randomised ITT set is split (CS B.2.4):

- Patients randomised to spesolimab who received,
    - either 1 dose (Day1) or 2 doses (Day 1 and Day 8), n=35
    - 1 dose (Day 1), n=23
    - 2 doses (Day 1 and Day 8), n=12
- Patients randomised to placebo who received,
    - spesolimab (Day 8), n=15

The secondary outcomes assessed at Week 4, GPPASI score, pain VAS, PSS and FACIT-Fatigue, were reported descriptively according to these four post-hoc groups and the planned hierarchical statistical testing of these outcomes was abandoned. They are also reported descriptively for the ITT analysis in CSR section 11.1.2.2.[17]

Analyses of other outcomes beyond Week 1:

- GPPGA pustulation subscore and total score through Week 12 were analysed descriptively according to the four post-hoc subgroups (CS section B.2.6.3.1.1).
- PROs through Week 12 (pain VAS, FACIT-Fatigue, DLQI and PSS) were analysed descriptively for the randomised ITT populations (CS section B.2.6.3.1.2)
- Systemic features (neutrophil and CRP levels) are reported descriptively for the spesolimab group only (CS section B.2.6.3.1.3).

The safety set included all randomised patients who received at least one dose of study medication (CS section B.2.10) [redacted].

EAG comment: The ITT analysis used for the primary outcome and the secondary outcome is appropriate. Only the results for the primary outcome and key secondary outcome at Week 1 can be considered comparatively meaningful. All other results can only be considered illustrative because they are reported according to assignment to post-hoc groups, not per-protocol, which leads to increased risk of bias and the company have not explained the relevance of each group nor how the results should be interpreted. [redacted].

Sample size calculations

It was calculated that 51 patients were needed to provide ≥90% power (CS section B.2.4 and CS Table 10), and this was achieved. This only provides power for the primary outcome and the key secondary outcome.

EAG comment: The target sample size to achieve the required power for the primary outcome and key secondary outcome was achieved.

Methods to account for multiplicity

The primary outcome and key secondary outcome were tested hierarchically, i.e. the key secondary outcome was only tested if the primary outcome results were statistically meaningful. Hierarchical testing of the secondary outcomes was planned but abandoned due to 15/18 participants in the placebo arm receiving open label spesolimab on Day 8, whereby the analysis changed to assessing results descriptively according to four post-hoc population groups instead. The trial was not powered for statistical comparisons of the secondary outcomes, nor were the secondary outcomes statistically tested as planned, so the results of the secondary outcomes are uncertain.

EAG comment: due to cross-over at Day 8 multiplicity is only relevant to the primary outcome and key secondary outcome, and this has been accounted for. All other outcomes are not statistically analysed and are illustrative only.

Analysis of outcomes

The GPPGA pustulation subscore, GPPGA total score, and GPPASI score were analysed using the Suissa-Shuster Z-pooled method, which is an exact unconditional test suitable for these binary outcomes. Confidence intervals were determined using the Chan and Zhang method which is suitable for small sample sizes.

All outcomes, except for the primary outcome and key secondary outcome, were analysed descriptively. Clarification Response A17 confirms that all comparisons according to randomised treatment as originally planned are non-informative after Day 8.

EAG comment: the company have used standard statistical analyses appropriate to the type of outcome and suitable for small sample sizes.

Handling of missing data

Missing data for the binary outcomes (includes primary outcome and key secondary outcome) was imputed as non-response, [redacted].

The last observation carried forward (LOCF) method [redacted].was planned to handle missing data for the continuous outcomes (pain VAS, PSS, FACIT-Fatigue, DLQI). However, the high crossover of participants into spesolimab treatment after Day 8 meant that this was replaced with descriptive analysis based on observed cases, including values after rescue therapy or treatment discontinuation (Clarification Response A13). Therefore, results using the LOCF method were not reported in the CS, but are reported

for pain-VAS and DLQI in Clarification Response A13 for participants initially randomised to spesolimab.
EAG comment: the amount of missing data is not clear. However, the methods appear appropriate, the primary outcome and key secondary outcome are treated conservatively in the primary analysis, [redacted].
Sensitivity analyses
Sensitivity analyses were performed for the primary outcome and the key secondary outcome: - post-hoc use of linear regression and to adjust for the imbalance of covariates at baseline, incl. sex, race, and GPPASI score (CS Table 10). [redacted]
EAG comment: The sensitivity analyses are relevant, e.g. adjusting for imbalanced covariates at baseline, [redacted] due to small numbers could affect results more than if there was a larger sample size. We have not identified any other sensitivity analyses that would have been useful.

Source: CS Table 10, Clarification Response A13, CSR section 9.7.[17]
Abbreviations: CRP, C-reactive protein; CSR, clinical study report; DLQI, Dermatology Life Quality Index; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy – Fatigue scale; GPP, Generalised Pustular Psoriasis; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment; GPPASI, Generalised Pustular Psoriasis Physician Area Severity Index; ITT, intention to treat; LOCF, last observation carried forward; PROs, patient reported outcomes; PSS, Psoriasis Symptom Scale; VAS, visual analogue scale.

EAG conclusion on study statistical methods

The EAG have not identified any methodological issues with the statistical methods of Effisayil 1. Only the results for the primary outcome and the key secondary outcome (both at Week 1) can be interpreted as statistically meaningful, because high crossover to spesolimab treatment after Day 8 rendered the planned statistical analysis of the other outcomes noninformative.

3.2.5 Efficacy results of the intervention studies

In this section we focus on the primary outcome (GPPGA pustulation subscore 0 at Week 1) and the key secondary outcome (GPPGA total score of 0 or 1 at week 1). We also report on the GPPGA pustulation subscore over time because within these data we can observe something of how the proportions with a GPPGA pustulation subscore 0 or 1 vary over time (response to treatment in the economic model is modelled in terms of flare resolution,

defined as a GPPGA pustulation subscore of 0 or 1). We do not report on the GPPASI 75 outcome here but descriptive post-hoc analyses for four participant groups, based on treatment received, are presented in CS Appendix E for this outcome.

3.2.5.1 GPPGA pustulation subscore 0 at Week 1 (Primary outcome)

At Week 1, 19 of the 35 participants in the spesolimab arm (54.3%) achieved a GPPGA pustulation subscore of 0 (i.e. they had no visible pustules). In the placebo arm only one participant (5.6%) achieved this (Table 9 below and CS Figure 7). The risk difference between the spesolimab and placebo arms of the trial was 48.7 percentage points (95% CI 22 to 67, p-value <0.001). In this analysis, any participants who withdrew from the study (one participant in the spesolimab arm) or who received escape medication during Week 1 (two participants in the spesolimab arm, one in the placebo arm) were assumed not to have responded.

A post-hoc sensitivity analysis was conducted with linear regression that adjusted for sex, race, and baseline GPPASI value because these were imbalanced covariates at baseline. As can be observed from Table 9, similar results were obtained.

Table 9 GPPGA pustulation subscore 0 at Week 1 and results of post-hoc sensitivity analysis

Primary outcome	**Spesolimab (N=35)**		**Placebo (N=18)**	
GPPGA pustulation subscore of 0 at week 1, n/N (%)	19/35 (54.3%)		1/18 (5.6%)	
Risk difference percentage points (95% CI), p-value	48.7 (21.5 to 67.2), p<0.001			
Post-hoc sensitivity analyses of primary outcome[a]				
	Spesolimab	**Placebo**	**Spesolimab**	**Placebo**
Sex	Female		Male	
Responders/total patients	11/21	1/15	8/14	0/3
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	48.2 (26.9 to 69.5), p<0.001			
Race	Asian		White	
Responders/total patients	10/16	1/13	9/19	0/5
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	52.2 (31.4 to 72.9), p<0.001			

Baseline GPPASI value	
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	46.9 (25.4, 68.3), p<0.001[b]

Source: Based on CS Appendix E Table 14 with data from CS B.2.6.1 added.
CI, confidence interval; GPPASI, Generalised Pustular Psoriasis Area and Severity Index; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment
[a] The post-hoc sensitivity analyses are based on a linear regression model using the primary endpoint as the dependent variable, and treatment assignment plus the respective covariate (sex, race or baseline GPPASI value) as the independent variable. The 95% CI and p-value are based on robust standard errors.
[b] The EAG believes incorrect values are reported in CS Appendix E Table 14 and so we report the values from Bachelez et al. (2021)[16] Table S5.

3.2.5.1.1 Time to complete pustular clearance

This analysis, reported in CS Figure 8, was conducted without censoring for participants receiving escape medication. In the spesolimab arm, complete pustular clearance was achieved at Day 2 in four participants, Day 3 in 11 participants and day 8 in 21 participants. All the participants in the spesolimab arm who achieved a GPPGA pustulation subscore of 0 with a single dose of spesolimab (n=19/35) did this by Week 1 (Day 8). The single participant who achieved complete pustular clearance in the placebo arm did this at Day 8.

CS Figure 9 shows pictures of one participant's skin before (Day 1, baseline) and after treatment (Day 3 and Week 1) with a single dose of spesolimab and a second participant's skin before (Day 1, baseline) and after treatment with a single dose of spesolimab (Day 8, Week 4, Week 12).

3.2.5.1.2 GPPGA pustulation subscores over time by randomised treatment at Day 1 and open-label spesolimab treatment at Day 8

CS Figure 12 panel A shows the GPPGA pustulation subscores over time for the 35 participants randomised to spesolimab. Data from this panel informs the efficacy data that is used in the cost-effectiveness model (see section 4.2.6.1 and Table 14 of this report for further details). At Day 8, 57% (n=20/35) of participants had a GPPGA pustulation score of 0 or 1, rising to 80% (n=28/35) had a GPPGA pustulation subscore of 0 or 1 at week 2, this proportion was maintained at week 4 and then fell slightly such that at Week 12 71% (n=25/35) had a GPPGA pustulation score of 0 or 1. We note that an outcome in the NICE scope and company decision problem is duration of response. Because the trial ended at Week 12 we do not know how much longer the response endured for the 60% of participants who had a GPPGA pustulation subscore of 0 at Week 12 (Key issue 4).

CS Figure 12 panel B shows the GPPGA pustulation subscores over time for the 23 participants who were randomised to spesolimab and received only a single spesolimab dose during the trial. CS Figure 12 panel C shows the data for the remaining participants randomised to spesolimab (n=12) who received their initial randomised dose and then a second open-label spesolimab dose at Day 8 because they still had a GPPGA pustulation subscore of at least 2. CS Figure 12 panel D shows the GPP pustulation subscores over time for the 15 participants randomised to placebo who then received open-label spesolimab at Day 8.

These data provide some indication of the duration of response after receipt of spesolimab for the treatment of a moderate-to-severe flare of GPP but, as patients were only followed-up to Week 12 in the Effisayil 1 trial, the data on duration of response are limited.

3.2.5.2 GPPGA total score of 0 or 1 at week 1 (Key secondary outcome)

The proportion of participants with a GPPGA total score of 0 or 1 at Week 1 was 42.9% (15/35 participants) in the spesolimab arm compared to 11.1% (2/18 participants) in the placebo arm. The risk difference between the trial arms was 31.7 percentage points, statistically significantly in favour of spesolimab. The analysis was conducted by assuming that any participants who withdrew from the study or who received escape medication during Week 1 had not responded.

The EAG infers from CS Figure 11 that all the participants meeting the criteria for this outcome did so with a GPPGA total score of 1 (i.e. no participants had a GPPGA total score of 0 at week 1).

A linear regression that adjusted for sex, race, and baseline GPPASI value because these were imbalanced covariates at baseline was also conducted for this key secondary outcome as a post-hoc sensitivity analysis. Table 10 shows that similar results were obtained.

Table 10 GPPGA total score 0 or 1 at Week 1 and results of post-hoc sensitivity analysis

Key secondary outcome	Spesolimab (N=35)	Placebo (N=18)
GPPGA total score of 0 or 1, n/N (%) at week 1	15/35 (42.9%)	2/18 (11.1%)
Risk difference percentage points (95% CI), p-value	31.7 (2.2 to 52.7), p<0.02	
Post-hoc sensitivity analyses of key secondary outcome[a]		

Key secondary outcome	Spesolimab (N=35)		Placebo (N=18)	
	Spesolimab	Placebo	Spesolimab	Placebo
Sex	Female		Male	
Responders/total patients	10/21	2/15	5/14	0/3
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	34.6 (11.3 to 57.9), p=0.005			
Race	Asian		White	
Responders/total patients	8/16	2/13	7/19	0/5
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	35.4 (12.5 to 58.3), p=0.004			
Baseline GPPASI value				
Adjusted risk difference percentage points (95% CI), p-value for treatment difference	29.7 (5.8, 53.5), p=0.018			

Source: Based on CS Appendix E Table 15 with data from CS B.2.6.2 added.
CI, confidence interval; GPPASI, Generalised Pustular Psoriasis Area and Severity Index; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment
[a] The post-hoc sensitivity analyses are based on a linear regression model using the primary endpoint as the dependent variable, and treatment assignment plus the respective covariate (sex, race or baseline GPPASI value) as the independent variable. The 95% CI and p-value are based on robust standard errors.

3.2.5.2.1 Time to first achievement of a GPPGA total score of 0 or 1

This analysis was conducted without censoring for participants receiving escape medication. The earliest achievement of a GPPGA total score of 1 (almost clear) in the spesolimab arm was day 3 (in 2 participants). By Day 8, which is the next time point shown in CS Figure 11, 17 participants in the spesolimab arm had a GPPGA total score of 1. In contrast, the earliest achievement of a GPPGA total score of 1 in the placebo arm was at Day 8 (in 3 participants). As CS Figure 11 shows, no participants achieved a GPPGA total score of 0 or 1 during Week 1.

3.2.5.2.2 GPPGA total score over time by randomised treatment at Day 1 and open-label spesolimab treatment at Day 8 (exploratory outcome)

The proportion of participants randomised to spesolimab (CS Figure 13 panel A, n=35) with GPPGA total scores of 0 or 1 increases from 43% at day 8 to 66% at week 4 (which is the

first time point that scores of 0 are observed) before levelling off at 60% at week 8 and week 12.

CS Figure 13 panel B shows that among the 23 participants randomised to spesolimab who received a single dose of treatment 15 (65%) had a GPPGA total score of 1 at day 8. From Day 8 to Week 12 the proportion of participants with a GPPGA total score of either 0 or 1 reached a maximum of 74% at week 4 (3/23 participants with a GPPGA total score of 0 and 14/23 participants with a GPPGA total score of 1) dropping back down to 61% at week 12 (4/23 participants with a GPPGA total score of 0 and 10/23 participants with a GPPGA total score of 1). The pattern of response among participants who were randomised to placebo and who received open-label spesolimab on Day 8 was a little more variable (potentially due to the smaller numbers of participants) but supports these observations (CS Figure 13 panel D).

Among the 12 participants who received two doses of spesolimab (CS Figure 13 panel C), none achieved a GPPGA total score of 0 during the 12-week trial but 58% (n=7/12) had a GPPGA total score of 1 at week 12 (the remaining 42% received escape or rescue treatment).

3.2.5.3 HRQoL outcomes

3.2.5.3.1 Pain VAS, FACIT-Fatigue, DLQI, and PSS

CS section B.2.6.3.1.2 reports mean score change from baseline over time (through to Week 12, end of study) descriptively for the following PROs: pain VAS, FACIT-Fatigue, PSS and DLQI. This includes the first week of comparator placebo evidence. CS Figure 14 shows that in the first week by Day 8, the placebo-controlled period, the spesolimab group (n=35) mean score achieved the MCID, shown by the dashed orange line, in all PRO scores, though we note that the 95% confidence intervals crossed the MCID boundary for the Pain VAS and DLQI. Although the mean score for FACIT-Fatigue and PSS scores of the placebo group (n=18) also achieved the MCID 95% confidence intervals crossed the MCID boundary for both and the spesolimab group achieved a greater improvement in both mean scores. Median score change from baseline at Week 1 is reported in the CSR in addition and compiled in Table 11 below.

Table 11 PROs: median score change from baseline at Week 1 in Effisayil 1 (pain VAS, FACIT-Fatigue, DLQI, PSS)

Outcome	Spesolimab N=35	Placebo N=18
Pain VAS. CFB, median (Q1, Q3)	[redacted]	[redacted]
FACIT-Fatigue. CFB, median (Q1, Q3)	[redacted]	[redacted]
PSS. CFB, median (Q1, Q3)	[redacted]	[redacted]
DLQI. CFB, median (Q1, Q3)	[redacted]	[redacted]

Source: CSR Tables 11.1.3.1.4:1 (pain VAS), 11.1.3.1.4:2 (PSS), 11.1.3.1.4:3 (FACIT-Fatigue), 15.2.4.10.3:2 (DLQI). Use of escape medication represents non-response. Last observation carried forward imputation for any missing data.
Abbreviations: CFB, change from baseline; DLQI, Dermatology Life Quality Index; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapies – Fatigue scale; IQR, interquartile range; NA, not applicable; PSS; Psoriasis Symptom Scale; VAS, visual analogue scale.

Using median score change from baseline instead of mean score change from baseline shows that the spesolimab group [redacted] MCID for the pain VAS score (a [redacted] points, MCID is a decrease of =>30 points), and that the placebo group [redacted] MCID for the PSS score. Thus, the median score change from baseline is more conservative than the mean score change.

After Day 8, CS Figure 14 compares the placebo group who received open label spesolimab (15/18) with the group initially randomised to spesolimab (35/35, 12 of whom received a second (open label) dose of spesolimab). The group initially randomised to placebo met the MCID for the scores where they had not met it already, pain VAS and DLQI, within a week (by Week 2). Generally, both groups show continued improvement in all PRO scores until the end of the study with confidence intervals for the two groups overlapping from the Day 8 time point onwards. We note that all participants had received one or two doses of spesolimab for the results reported after Day 8 and so these results do not support efficacy of spesolimab against placebo nor against standard of care.

The per-protocol secondary outcomes for Week 4 are reported within CS Appendix E Table 16, but descriptively according to the post-hoc subgroups noted in CS section B.2.4 instead of the original randomised groups.

3.2.5.3.2 EQ-5D

CS Figure 15 shows that the group initially randomised to spesolimab achieved an increase in EQ-5D median score change from baseline [redacted] by Week 1,

[redacted]. After Week 1 the results for the group initially randomised to spesolimab are inconsistent, and there are no further results reported for the group originally randomised to placebo who received open label spesolimab at Day 8. EQ VAS score change from baseline, by visit, was not reported in the CS.

3.2.5.4 Subgroup analyses

No subgroups were specified in the NICE scope or the company decision problem. Pre-specified exploratory subgroup analyses for the Effisayil 1 trial are reported for the primary outcome and the key secondary outcome in CS section B.2.7 and CS Appendix E.1 Figures 5 and 6. These results have been published in the trial publication Burden et al. 2023.[22]

Results are reported for the pre-specified subgroups of sex, race (Asian/White), BMI, plaque psoriasis at baseline, IL36RN mutation status, baseline GPPGA total score, baseline GPPGA pustulation subscore, baseline GPPASI total score, baseline Japanese Dermatological Association (JDA) GPP Severity Index, and medication for GPP prior to randomisation. The trial publication explains that results for the pre-specified age subgroups were not reported because there were only two patients in the ≥65 years subgroup. [22]

For both the primary outcome (the proportion of participants achieving a GPPGA pustulation subscore of 0 at Week 1) and the key secondary outcome (the proportion of patients achieving a GPPGA total score of 0 or 1 at Week 1) the majority of the subgroups had similar treatment effect estimates to the overall trial population that are favourable for spesolimab.

Some of the subgroups were very small, particularly for the placebo group which only numbered 18 in total. Consequently, the subgroup confidence intervals tend to be wider than that for the overall trial population and often cross the null line making the results less certain. The subgroup analyses do not raise any significant concerns but should be viewed as illustrative.

3.2.5.5 Safety results

Safety results for Effisayil 1 are reported in CS sections B.2.10.1-2 and CS Table 13 for adverse events at Week 1 (to align with the timing of the primary outcome) and at Week 12 for all randomised patients who received at least one dose of the study drug.

At Week 1 the proportions of participants experiencing adverse events, severe adverse events, or drug-related adverse events were similar between the spesolimab and placebo groups. There were only slightly more adverse events in the spesolimab group (23/35, 66%) than in the placebo group (10/18, 56%). The placebo group fared much worse for pyrexia with 22.2% of participants experiencing pyrexia in the placebo group compared to 5.7% in the spesolimab group. The pyrexia events are discussed in the trial publication as occurring in the context of the underlying GPP flare but the authors were unable to rule them out as drug-related events (Bachelez et al. 2021).[16] Infections were the adverse events that occurred among a greater proportion of the spesolimab group (6/35, 17.1%) than the placebo group (1/18, 5.6%) in week 1. There were four serious adverse events among two participants in the spesolimab group and none in the placebo group. Three of the serious adverse events [redacted] being drug reaction with eosinophilia and systemic symptoms (DRESS), serious urinary tract infection, and drug-induced hepatic injury (the fourth was arthritis).

At Week 12 there were 51 participants in total, according to CS Table 13, who had received up to three doses of spesolimab during the course of the study. At Week 12 82% of participants had experienced an adverse event and 55% experienced drug-related adverse events. All incidence rates for adverse events had decreased since Week 1. There were eight serious adverse events experienced by six participants. Two of the serious adverse events were reported as DRESS but the CS states these were re-classified by independent external expert assessment as ‘no DRESS’ and as ‘possible DRESS’ which both resolved without drug treatment. The serious adverse events do not raise any further concerns.

The trial publication for Effisayil 1 reports on antidrug antibodies, the occurrence of which may lead to loss of efficacy of the study drug and potential risk for increased toxicity. Antidrug antibodies were detected in 46% of participants treated with spesolimab, at a median of 2.3 weeks after spesolimab administration.[16] This does not appear to have affected safety results at Week 12 whereby the incidence rate of adverse events decreased.

Throughout the Effisayil 1 trial there were no deaths and no adverse events that led to discontinuation of the trial drug.

Results of the Phase I spesolimab study show that all seven participants experienced an adverse event, all graded as mild or moderate, and there were no severe or serious adverse events (CS Appendix F.1.2). Interim results for Effisayil ON, the open label extension study, show that nine out of ten participants experienced an adverse event, and there was only one

severe adverse event and one serious adverse event both of which were pustular psoriasis in the same patient (CS section B.2.10.3). These studies do not raise any safety concerns.

EAG conclusion on the safety results

Spesolimab appears to be well tolerated and the safety results to date raise no concerns. However, we note that the data are limited.

3.2.6 Pairwise meta-analysis of intervention studies

Pairwise meta-analysis was not conducted because only one trial provides evidence fully relevant to this appraisal.

3.3 Critique of studies included in the indirect comparison and/or multiple treatment comparison

In CS section B.2.9 the company state that there was an absence of robust comparator data for formal indirect treatment comparison but did not elaborate further nor signpost to other parts of the company submission for this information. Our summary of the location and sources of evidence the company identified and included in their feasibility assessment for indirect treatment comparison is provided in Appendix 5.

The EAG has reviewed the information on the nine retrospective studies identified by the company and presented in CS Appendix D Table 11, the three real-world evidence sources and the two SEE sources (from CS Appendix M) but none could provide comparator data for formal indirect treatment comparison (Appendix 6). In addition, we asked the company to explain their rationale for not conducting unanchored population matching using trial or real-world data (clarification question A27). The company tabulated the trial and population characteristics for studies we listed in clarification question A27 in their response (Clarification response Table 8) and explained that there were no other studies (RCT or real-world) that reported efficacy or safety findings for a population of GPP patients experiencing flares as defined in the Effisayil 1 trial and receiving the comparator treatment for this appraisal, established clinical management (also described as best available care in the CS). We independently checked a proportion of the wider set of studies identified by the company's systematic literature review (clinical SLR tables embedded in CS Appendix D.1.2) and did not find any comparator data that would be suitable for use in a formal indirect treatment comparison.

3.3.1 The Effisayil 1 historical cohort.

In the absence of an indirect treatment comparison to inform the economic model the company used data from the Effisayil 1 historical cohort.[23] The way these data are used in

the model are described in more detail in section 4.2.6.2.1 of this report. We have some concerns about how well the flares and treatments received for flares represent current best available care/established clinical management in England.

3.3.1.1 Effisayil 1 historical cohort study aims and methods

The Effisayil 1 historical cohort,[23] (n=53) provides data on the characteristics and clinical course of past GPP flares. In this retrospective study, investigators used a standard study questionnaire to log data which included the clinical and laboratory characteristics of past flares that were identified as being the typical, most severe, and longest flares experienced by each patient. There was no standard definition for typical, most severe and longest flares so this was based on investigator interpretation. For these three categories of flares, the treatment used, duration of hospitalisation and time to clearance of pustules, erythema and scaling was also logged.

3.3.1.2 Effisayil 1 historical cohort study generalisability and patient characteristics

We asked the company to provide evidence to show how representative the Effisayil 1 historical cohort,[23] is of the generalised pustular psoriasis patient population experiencing flares in England (clarification question A24). The company provided patient demographic and flare characteristics data (clarification response Table 6) for the POLARIS[8] and SCRIPTOR[24] real-world evidence sources whilst also stating that caution should be applied when making comparisons across data sources because of differences in definitions and collection methods. We note that patients in the POLARIS study in England with incident GPP (n=206), who experienced a flare (n=129) had a higher mean age than those in the Effisayil 1 trial at baseline (mean 57.3 years, SD 19.0 in POLARIS versus 43.0, SD 10.9 in Effisayil 1). The patients in the SCRIPTOR study (n=27 from the UK) were also older than those in the Effisayil 1 trial at baseline (mean 53.1, SD 19.7 at diagnosis). Due to the locations where the Effisayil 1 trial took place (see section 3.2.1.1) over half of the participants (54.7%) were of Asian ethnicity with the remainder being White (45.3%). A similar proportion (46.1%) were White in the POLARIS study but the proportion of Asian patients was lower (8.6%) (the remaining patients' ethnicities were Black/Caribbean 2.1%, mixed 41.3% or other/unknown 1.9%).[8] In the SCRIPTOR study a much higher proportion were documented as White, Caucasian and/or of European decent (77.8%) with 14.8% as Central, South and/or East Asian descent[24] (we believe there is an error in Clarification response Table 6 which gives a value of 1.3%). The remaining two participants were either of mixed ethnicity or of unknown/not reported ethnicity). For the other characteristics it is

difficult to compare due to missing data or data for a particular characteristic not being reported in different sources or data being reported in a different format.

3.3.1.3 Effisayil 1 historical cohort study flare treatment

Most patients in the Effisayil 1 historical cohort,[23] (n=46/53, 86.6%) received treatment for past GPP flares with at least one medication and 24.5% of the cohort (n=13/53) received at least one biologic therapy. We note that the use of prior treatments for past GPP flares in the historical cohort is not specific for moderate-to-severe flares as data were not reported in this way in the publication. In the EAG's opinion the use of biologic treatments in the Effisayil 1 historical cohort (24.5%),[23] seems lower than might be expected with current treatment in England [CS Figure 4 shows [redacted]% at first-line (Week 1) and [redacted]% at second-line (weeks 2-4) and [redacted]% at third and later lines of therapy (Weeks 5+)] but the EAG acknowledges that three UK clinical experts consulted by the company confirmed the prior treatments for past GPP flares in the historical cohort was similar to the basket of treatments defined in the SEE. The time period during which the historical flares occurred is not known (clarification question A22) so it is possible that some flares could have occurred many years ago when biological treatments were not available or as widely used. We are concerned that time to pustular clearance in the Effisayil 1 historical cohort (CS Table 18) may have been longer than it would be for a similar patient group treated in the NHS because of these treatment differences. If this were the case it would disadvantage the comparator group, increasing the difference in time to flare resolution between the comparator and intervention groups.

Although the data on treatments received for past GPP flares were captured for the Effisayil 1 historical cohort,[23] the study did not report on which medications were received for the three categories of flare or the treatment composition for individual flares. Therefore, in the economic model, the evidence on treatments received during a GPP flare were based on the company's SEE exercise (CS Table 24). Consequently, the evidence on comparator efficacy and the composition of the comparator treatments received for a flare come from different sources. This is discussed further in section 4.2.6.2 of our report.

3.3.1.4 Effisayil 1 historical cohort outcomes

Descriptive results are presented which provide information on the length of time flares took to resolve. Data was not collected using the GPPGA, instead the proportions of patients with pustular, erythema and scaling resolution within different time bands (less than 1 week, 1-2 weeks, 3-4 weeks, 5-8 weeks 9-12 weeks and more than 12 weeks) were reported for each of the three types of flare category. In the economic model the company use time to

pustular clearance as a proxy for the GPPGA pustulation subscore of 0 or 1 which is used in the model as the measure of response to treatment (see section 4.2.6). It is unclear how well pustular clearance corresponds to a GPPGA pustulation subscore of 0 or 1. Duration of hospital treatment was similarly reported for each of the three types of flare category giving the proportions of patients with different durations of hospitalisation (less than 1 week, 1-2 weeks and 3-4 weeks).

3.3.1.5 Effisayil 1 historical cohort study risk of bias

The company had not conducted a risk of bias assessment for the Effisayil 1 historical cohort study[23] so we requested this (clarification question A23). The company judged four of seven domains as having a moderate risk of bias which, following the guidance on the use of the ROBINS-I[25] should have led to an overall risk of bias of moderate. However, the company gave an overall risk of bias judgement of 'Low' which we believe is inappropriate. After our independent risk of bias assessment (presented with that of the company in Appendix 7) we judged the Effisayil 1 historical cohort study[23] to be at a serious risk of bias.

3.3.2 Alternative comparator data sources

We reviewed the studies identified by the company (Appendix 6) to see if any provided data that could be used as an alternative to the Effisayil 1 historical cohort.[23] In particular we considered the CEE GPP Expert Network retrospective case-series by Wolf et al.[26] because information on this study is presented in CS section B.2.9.2. However, the focus of the Wolf et al. study is to describe GPP flares and treatments received rather than the outcomes of flare treatment. This study did not provide the level of detail required to inform effectiveness in the cost effectiveness model. Data from this study were used to help determine which flare category from the Effisayil 1 historical cohort.[23] was the most appropriate to use in the economic model and this is described in more detail in section 4.2.6.2.1 of our report.

The only other data source that does provide the level of detail needed to inform comparator effectiveness in the economic model is the company's SEE - BAC and efficacy,[14] which as noted above, provides the evidence for treatments received for a GPP flare in current UK practice. The company describe this SEE in CS section B.2.9.3 (with the summary of results in CS Appendix M.1.1). The company states "*predictions for BAC efficacy were generally aligned with the RWE outcomes observed in the Effisayil 1 historical cohort and CEE GPP Expert Network data*" however they ultimately decided that "*the data are not sufficient for decision making and are not used to inform the economic modelling*". Whilst we agree that the efficacy evidence elicited from the UK experts may be subject to a risk of bias and associated with uncertainty, we also recognise that it may better reflect the experience of

patients treated for moderate-to-severe GPP flares in the UK and be better aligned to the composition of comparator treatments. We therefore believe this data source should be considered and discuss this further in section 4.2.6.2.2 and 4.2.6.2.3. The SEE - BAC and efficacy report[14] states that [redacted].We note that an important potential source of bias is [redacted].

3.4 Conclusions on the clinical effectiveness evidence

The company's decision problem matches the NICE scope, however, the key clinical effectiveness evidence from the Effisayil 1 trial (spesolimab n=35, placebo n=18) differs from the NICE scope and company decision problem in two important areas:

- The population is adult patients with GPP flares of moderate-to-severe intensity which is a narrower population than defined in the NICE scope and company's decision problem (Key Issue 1)
- The Effisayil trial comparator is placebo, not established clinical management without spesolimab as defined in the NICE scope and company's decision problem (Key Issue 2)

The participants enrolled in the trial included a high proportion (55%) who were Asian which was partly a consequence of the countries from which participants were recruited. No participants were recruited from England. The company assessed the Effisayil 1 population as applicable to routine clinical practice [in the NHS] despite the high proportion of Asian people in the cohort (CS Table 11; CS section B.2.12.3), however, the EAG was unable to independently recruit clinical experts to verify this.

We agreed with the company's judgement that the risk of bias in the Effisayil 1 RCT was low. Although there were some differences in baseline characteristics between the groups, there is little published information about the prognostic factors for flare resolution. Some of the differences may bias the results in favour of the placebo arm.

There is no standard definition of flare resolution. In the Effisayil 1 trial the primary outcome was a GPPGA pustulation subscore of 0 (clear, no visible pustules) at Week 1. However,

clinical effectiveness (flare resolution) in the economic model is defined as a GPPGA pustulation subscore of 0 or 1 which is a less stringent response. Independent expert clinical opinion to verify the company's validation of the appropriateness of this definition of flare resolution would be welcome.

The Effisayil 1 RCT found that at Week 1 there was a statistically significant difference in favour of spesolimab in the proportion of participants who achieved a GPPGA pustulation subscore of 0 (primary outcome) and a GPPGA total score of 0 or 1 (Key secondary outcome). Similar results were obtained from post-hoc sensitivity analyses of these outcomes that adjusted for covariates that were not balanced at baseline (sex, race, baseline GPPASI value).

Participants in the placebo arm meeting pre-specified criteria were eligible to receive open label spesolimab on Day 8. The majority of placebo arm participants met the criteria and chose to receive open label spesolimab. Therefore, there is no comparative evidence for the efficacy of spesolimab in comparison to placebo beyond Week 1 of the trial. The planned testing of secondary outcomes, including for three patient reported outcome measures (Pain VAS, PSS and FACIT-Fatigue score), at Week 4 was abandoned and instead, outcomes after Day 8 were analysed descriptively. The EQ-5D-5L (health index and VAS) was an exploratory outcome and contributes data to the economic model, but in common with other trial outcomes, there is no comparative evidence beyond Week 1. The EAG view the PRO results from Effisayil 1 as illustrative. The total trial period was 12 weeks which means that we do not know the extent of the duration of the response to spesolimab.

The reported adverse events raised no concerns, although data both in terms of the number of people who had received spesolimab (n=51 participants in total) and the duration of the trial (12 weeks) are limited. There were no deaths during the trial and no adverse events that led to discontinuation of the trial drug (section 3.2.5.5).

In the absence of comparative data from the Effisayil 1 trial beyond Week 1 the company sought to identify studies that could be included in an indirect treatment comparison. No studies were identified that provided comparator data suitable for use in a formal indirect treatment comparison. The company therefore uses data from the Effisayil 1 historical cohort to inform cost-effectiveness modelling of the comparator arm beyond Week 1. We assessed this study as being at a serious risk of bias. The Effisayil 1 historical cohort provides data on the characteristics and clinical course of past GPP flares experienced by the participants who took part in the Effisayil 1 RCT. The past GPP flares were categorised by investigators as being either typical, the most severe or the longest flares experienced by

patients but there was no standard definition for these categories. It is difficult to know how representative the historical data on flare are to the GPP patient population experiencing flares in England. Additionally, the time period when the historical flares occurred is not known, and this may have impacted the range of treatments that people had received for their historic flares. In particular, the use of biologic treatments in the Effisayil 1 historical cohort seems lower than might be expected with current treatment in England. We have concerns that treatment differences could mean that pustular clearance in the historical cohort took longer than would be expected for a similar patient group receiving treatment in the NHS now. If this was the case this would disadvantage the comparator group in the economic model.

The company ruled out the use of their SEE – BAC and efficacy to inform comparator effectiveness in the economic model. Although we acknowledge that these data are also subject to a risk of bias as well as being associated with some uncertainty the results are relevant to the treatment of GPP flares in the NHS. Additionally, the composition of comparator treatments received that is used in the economic model also comes from the SEE – BAC and efficacy. We therefore use data from this source to inform comparator effectiveness in the EAG base case.

4 COST EFFECTIVENESS

4.1 EAG comment on company's review of cost-effectiveness evidence

The company reports their economic search strategy in CS section B.3.1 and CS Appendix G. They conducted simultaneous searches to identify cost-effectiveness studies, costs, and health care resource use studies for the treatment of patients with GPP. A single search strategy was carried out and consisted of an original search on 27 December 2022 and an updated search on 6 May 2024. CS Appendix G Table 31 presents the inclusion and exclusion criteria. The searches were well-constructed; the company searched relevant sources and used published search filters which they adapted by adding further search terms. Thus, we consider the searches to be sensitive, up-to-date, and not likely to miss any studies.

The systematic literature review did not identify any cost-effectiveness studies for patients with GPP, and there are no previous NICE appraisals assessing GPP. We are not aware of any cost-effectiveness studies that have been missed by the company.

4.2 Summary and critique of the company's submitted economic evaluation by the EAG

The company developed a de novo economic model to assess the cost-effectiveness of spesolimab compared with BAC in the treatment of adult patients with GPP presenting with flares.

4.2.1 NICE reference case checklist

The company's economic model fulfils the requirements of NICE's reference case (Table 12), except for the time horizon. It is unclear whether the time horizon used in the company's submission (12 weeks) is long enough to reflect all important differences in costs or outcomes between spesolimab and BAC. For further details, please see section 4.2.5.

Table 12 NICE reference case checklist

Element of health technology assessment	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and Personal Social Services	Yes

Element of health technology assessment	Reference case	EAG comment on company's submission
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Unclear
Synthesis of evidence on health effects	Based on systematic review	Yes
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
Evidence on resource use and costs	Costs should relate to NHS and Personal Social Services resources and should be valued using the prices relevant to the NHS and Personal Social Services	Yes

Element of health technology assessment	Reference case	EAG comment on company's submission
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	No discounting applied due to the short time horizon

Source: EAG assessment based on the company submission.
EAG, External Assessment Group; NHS, National Health Service; NICE, The National Institute for Health and Care Excellence; QALY, quality-adjusted life-years; UK, United Kingdom.

4.2.2 Model structure

4.2.2.1 Overview of the model structure

The company developed a de novo cost-effectiveness model, which is described in CS section B.3.2.2. The model parameters are presented in CS sections B.3.3 to B.3.5, the base case inputs in CS section B.3.9.1, and the model assumptions in CS section B.3.9.2. The company developed a Markov state transition model with three mutually exclusive health states: GPP flare, resolved flare, and death. A daily cycle length was adopted, and the company did not apply a half-cycle correction. All patients start the model in the GPP flare health state, which is defined as a GPPGA score ≥ 3, new or worsening pustules, GPPGA pustulation subscore ≥ 2 and ≥ 5% of body surface area with erythema and the presence of pustules. At each cycle in the model, patients can remain in the GPP flare health state, move to the resolved flare health state (defined as a GPPGA pustulation subscore of 0 or 1) or die. We note that death was not initially modelled by the company for patients in the resolved flare health state. In response to clarification question B4, the company updated their model and included general population mortality to reflect the risk of death for patients with a resolved flare. The updated model structure is presented in Figure 2 below.

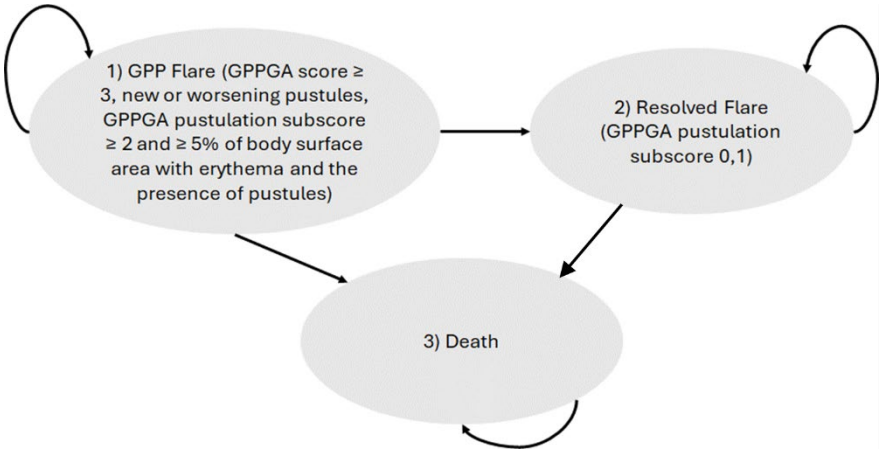

Figure 2 Updated model structure

Source: Reproduced from Figure 3 of the clarification response document.
GPP, generalised pustular psoriasis; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment.

In the company submission, response to treatment was defined as a GPPGA pustulation subscore of 0 or 1. The primary outcome of the Effisayil 1 trial is a GPPGA pustulation subscore of 0.[16] However, UK clinicians consulted by the company as part of the model developed for this submission validated the use of a pustulation subscore of 0 or 1 to represent a resolved flare as it better reflects the impact of GPP in UK clinical practice.[28] Also, the UK clinicians advising the company are of the opinion that any patients who had been in hospital would likely be discharged at this stage.[28] We were unable to confirm this assumption with clinical experts, but we consider that further clinical opinion would be relevant.

4.2.2.2 EAG critique of model assumptions

4.2.2.2.1 A second/recurrent GPP flare is not implemented in the model

Patients who respond to treatment (i.e., have a GPPGA pustulation subscore of 0 or 1) are assumed to remain responders for the remainder of the modelled time horizon. The company argues that evidence on flare frequency shows that patients are unlikely to have more than two flares per year. But the EAG notes that it does not mean that patients cannot have two flares in a 12-week period.

Based on information from the company submission and clarification response to question A1i, the EAG notes that within the 12-week Effisayil 1 trial period:

- Two patients in the spesolimab arm received a single dose of spesolimab (at baseline) for their first flare and then a rescue dose for a second/recurrent flare.

- Two patients in the spesolimab arm received two doses of spesolimab (at baseline and at day 8) for the first flare and then a rescue dose for a second/recurrent flare.
- Two patients in the placebo arm received a single dose of spesolimab at day 8 and then a rescue dose for a second/recurrent flare.
- Four patients in the spesolimab arm and another four patients in the placebo arm received standard of care escape treatment after day 8 – this was for patients with disease worsening after day 8 who either had not achieved a clinical response (GPPGA pustulation subscore of 0 or 1) and had disease worsening, or who had achieved a clinical response but subsequently experienced disease worsening that was not severe enough to meet criteria for flare recurrence and receive a rescue dose of spesolimab.

Therefore, 11.3% of patients in the Effisayil 1 trial received rescue treatment with spesolimab to treat second/recurrent flares, and eight patients received a standard of care escape treatment, [redacted]. The company did not include second or recurrent flares in the model. Although we acknowledge the low number of patients in this situation, we believe the company should have explored this in scenario analyses, as it would be valuable to see how modelling second/recurrent flares may affect the model conclusions. We also note that we were unable to confirm with clinical experts whether the company's assumption may be reasonable. We consider this to be a key issue (Key Issue 3) as we believe this uncertainty can be resolved with further clinical data and further clinical input.

4.2.2.2.2 Severity of a GPP flare (based on GPPGA pustulation subscore) is not implemented in the model

The company argued that this was a simplifying assumption designed to increase the transparency and interpretability of the model. The EAG agrees that modelling the severity of GPP flares would be challenging as the available data to inform the economic model does not seem to be categorised in that way.

Nevertheless, we note that most of the evidence informing the model is drawn from patients with moderate-to-severe GPP flares (Effisayil 1 trial and the SEE exercise). Therefore, it is unclear whether the model results are generalisable to the patients with mild flares. This is further discussed in section 4.2.3.

4.2.2.2.3 Maximum flare duration is 12 weeks

In the company's model, all patients are assumed to respond to treatment (spesolimab or best available care) by Week 12. The company considers this to be a conservative assumption as the study informing the efficacy of the comparator arm, the Effisayil 1 historical cohort, shows that around 12% of patients have not responded in 12 weeks.[23] We therefore tested this in a scenario analysis (see section 6).

Evidence from the Effisayil 1 trial shows that 25 of the participants at Week 12 had a GPPGA pustulation subscore of either 0 (n=21) or 1 (n=4) and the remaining 10 participants had received escape/rescue therapy; therefore there is no information on their Week 12 GPPGA pustulation subscore. In the best-case scenario, all patients achieved a GPPGA pustulation subscore of 0 or 1 by Week 12 (company's base case); in the worst-case scenario, none of them resolved their flare by Week 12. The EAG believes that it is likely that some of these patients have resolved their flares by Week 12 while others have not. Thus, we tested the following alternatives in scenario analyses (see section 6):

- The worst-case scenario (7 out of 35 patients have not responded to treatment by Week 12): 20.0%. According to the text above, the worst-case scenario is 10/35= 28.6%, however due to model limitations and the model assumption that patients cannot get a second/recurrent flare within the time horizon, it is only possible to model a maximum of 7 patients not responding to treatment by Week 12.
- The same proportion of patients as for the comparator arm have not responded to treatment by Week 12: 12.37%.
- The proportion of patients who received escape/rescue therapy at the latest time (between Week 8 and Week 12) have not responded to treatment by Week 12: 5.7%. This assumes that these patients did not have enough time to recover after receiving the escape/rescue therapy.

4.2.2.2.4 No half-cycle correction and no discounting of outcomes

The EAG agrees with these assumptions as the duration of both the cycle length (one day) and the time horizon (12 weeks) are short. Therefore, the impact of half-cycle correction and discounting would be negligible.

EAG comment on model structure

The EAG considers the model structure to be appropriate for this condition and to reasonably reflect its pathway. The company assumed that a GPP flare that appeared to initially improve cannot get worse again within a 12-week period, and that all

patients have responded to treatment (both to spesolimab and to BAC) by week 12. We are unclear whether these assumptions are appropriately representing UK reality, and therefore we consider that further clinical expert would be valuable.

4.2.3 Population

The population considered in the company model is described in CS section B.3.2.1 and consists of adult patients with GPP presenting with flares. This is aligned with the population defined in the NICE scope and the licensed population for spesolimab. However, we note that most of the evidence informing the model is drawn from the moderate-to-severe flare population (Effisayil 1 trial and the SEE exercise). It is unclear whether the model results are generalisable to the patients with mild flares and further clinical expert opinion would be helpful to clarify this. In the absence of that, we consider that the model conclusions are only relevant for the moderate-to-severe flare population.

The baseline characteristics of the model population are presented in CS Table 14. These were taken from the Effisayil 1 trial.[16] Table 13 shows the model inputs for baseline age, sex and weight. As we were not able to identify any clinical experts to advise the EAG, it is unclear whether the modelled baseline characteristics are representative of the patients who may receive spesolimab treatment in UK clinical practice. We note that using other values for the baseline characteristics has a limited impact on the model results.

Table 13 Baseline characteristics of the population

Characteristic	Baseline inputs
Average age (years)	43.00
Proportion female (%)	67.90
Average weight (kg)	72.03

Source: Reproduced from CS Table 14.

EAG comment on model population

The patient population included in the company's cost-effectiveness analysis aligns with the NICE scope and the licensed population. However, it is unclear whether the model results are generalisable to patients experiencing mild flares as the population from the studies informing the economic model is mainly a moderate-to-severe flare population. This is discussed in Key Issue 1. It is unclear whether the baseline characteristics based on the Effisayil 1 trial are reflective of UK clinical practice.

4.2.4 Interventions and comparators

CS section B.3.2.3 describe the intervention and comparators. The economic model compares spesolimab versus best available care (BAC).

Spesolimab is administered by intravenous infusion at the recommended dose of 900 mg on day 1. For patients with persistent symptoms, another dose of 900 mg can be offered on day 8. In the company's model, it was assumed that 80% of patients who received spesolimab on day 1 and have persistent symptoms on day 8 (GPPGA pustulation subscore ≥2) receive a second infusion of spesolimab. This was based on data from the Effisayil 1 trial, in which 12 out of 15 patients who did not achieve a GPPGA pustulation subscore of 0 or 1 received a second dose of spesolimab.[16] The remaining patients who have persistent symptoms on day 8 (20%) were assumed to receive treatment with ciclosporin.

There are no licensed treatments specifically approved to treat GPP flares in the UK aside from spesolimab. During the first week, the company used data from the placebo arm of the Effisayil 1 trial to inform the efficacy of the comparator arm (BAC). Beyond that, the company used data from the Effisayil 1 historical cohort study (further details are provided in section 4.2.6).[23] In the Effisayil 1 trial, patients in the placebo arm did not receive any active drugs during the first week although they could be admitted to hospital for supportive care.[16] For the Effisayil 1 historical cohort study, it was not possible to derive the treatment composition for an individual flare.

Therefore, in the company's model, it was assumed that patients in the comparator arm do not receive any drugs during the first week (to align with the placebo data of the Effisayil 1 trial) and, beyond the first week, evidence on the treatments received for a GPP flare in the UK were obtained from a group of UK experts participating in a SEE exercise (see Figure 1 above). The experts were asked to respond to the following question: "Please list up to 30 treatments that you believe would comprise currently best available care for first-line active treatment during a GPP flare, either moderate or severe (i.e., treatments used in practice)". Further clinical expert opinion on whether the treatments in Figure 1 reasonably reflect the standard of care for moderate-to-severe GPP flares in UK clinical practice would be useful as we were unable to confirm that with other clinical experts.

In the EAG's view, it is unlikely that patients in UK clinical practice would not receive any active drugs to treat GPP flares for a whole week. In the absence of better UK evidence for the composition of the BAC arm, we believe that modelling the treatments indicated by the UK experts as part of the SEE exercise is a reasonable approach. However, we are concerned that these treatments do not match the treatments used to treat GPP flares in the

Effisayil 1 historical cohort study, which is the study used to inform the efficacy of BAC after week 2 in the company's base case. It is therefore uncertain how generalisable the efficacy of the comparator (obtained from the Effisayil 1 historical cohort) is to the UK population (for further details, see section 4.2.6.2).

EAG comment on intervention and comparators

Although the intervention and comparators in the economic model appear to be broadly consistent with the NICE scope, the EAG is concerned that the composition of the comparator arm (BAC) is not aligned with UK clinical practice and/or the studies informing its efficacy. We therefore consider this to be a Key Issue (Key Issue 4).

4.2.5 Perspective, time horizon and discounting

The perspective of the analysis is the National Health Service (NHS) and Personal Social Services in England, in line with the NICE reference case. There is no discounting for costs and outcomes in the model as the time horizon is too short (12 weeks) (see section 4.2.2.2.4).

A time horizon of 12 weeks was implemented in the company's model. This is the follow-up period of the Effisayil 1 trial,[16] and the company considers this to be the period of most relevance in the emergency medical setting. Spesolimab is indicated to treat GPP flares. The company stated that evidence on flare frequency shows that patients are unlikely to have more than two flares per year.

The EAG notes that relevant evidence to inform the model is only available for a period of 12 weeks and therefore modelling a longer time horizon would be challenging. However, it is possible that treating one flare with spesolimab might affect the efficacy and safety of spesolimab or other treatments in a subsequent flare. Evidence on this does not seem to be currently available, and we are unclear whether any data are being collected to inform this (further information can be seen in the company's response to clarification question B1). We consider the appropriateness of a 12-week time horizon to be uncertain and questionable because the long-term consequences of the treatment with spesolimab are still unclear.

EAG comment on perspective, time horizon and discounting

The company uses the recommended perspective, which is in line with NICE guidelines. No discounting was applied in the company's model due to a short time horizon (<1 year), and we agree with this assumption. We are uncertain whether a time horizon of 12 weeks is long enough to capture all the costs and consequences of

the treatment with spesolimab. Therefore, we consider this to be a Key Issue (Key Issue 5).

4.2.6 Treatment effectiveness and extrapolation

4.2.6.1 Response to treatment: spesolimab

The approach to modelling the response to spesolimab is discussed in CS section B.3.3.1. The company used data from the Effisayil 1 trial to model the effectiveness of spesolimab. Evidence on flare response was available at day 2, day 3, day 8, week 2, week 3, week 4 and week 12.

As no standard definition of GPP flare resolution is available, the company defined response to treatment as a GPPGA pustulation subscore of 0 or 1, as previously discussed in section 4.2.2. However, we were not able to confirm whether the use of a GPPGA pustulation subscore of 0 or 1 appropriately reflects the resolution of a flare and further expert opinion would be beneficial. We tested the impact of this assumption by running an alternative scenario in which the response to treatment is defined as a GPPGA pustulation subscore of 0 (primary endpoint of the Effisayil 1 trial).

The company adjusted the trial data to the model daily cycle by assuming that no patients would respond on Day 1 and then applying a linear cumulative response in between weeks (see Table 14 below for the model inputs).

Table 14 Efficacy of spesolimab based on Effisayil 1 trial

Timepoint	GPPGA subscore 0 or 1 (company's base case)	GPPGA subscore 0 (EAG scenario analysis)
Day 2	13/35 = 37.1% (13 responders)	4/35 = 11.4%
Day 3	6/35 = 17.1% (6 new responders)	7/35 = 20.0%
Day 3-8	1/35 = 2.9% (1 new responder)	8/35 = 22.9%
Week 2	8/35 = 22.9% (8 new responders)	4/35 = 11.4%
Week 3	0/35 = 0.0% (no new responders)	0/35 = 0.0%
Week 4	0/35 = 0.0% (no new responders)	0/35 = 0.0%
Week 12	7/35 = 20.0% (remaining)	12/35 = 34.3%

Source: Partially reproduced from CS Table 17.
GPPGA, Generalised Pustular Psoriasis Physician Global Assessment.

4.2.6.2 Response to treatment: BAC

The company's approach to modelling the response to BAC is discussed in CS section B.3.3.2. For the first week of the model, response to BAC was obtained from the Effisayil 1 trial using a similar approach as the one used for spesolimab (see section 4.2.6.1).

After the first week, response to BAC was not obtained from the Effisayil 1 trial because crossover occurred for more than 80% of patients in the placebo arm, who received spesolimab on Day 8. In their base case, the company used data from the Effisayil 1 historical cohort to model the efficacy of BAC after the first week.[23]

4.2.6.2.1 Effisayil 1 historical cohort (company's base case)

In Effisayil 1 historical cohort, investigators collected historical data from patients included in the Effisayil 1 trial to describe the characteristics and outcomes of GPP flares prior to patients' enrolment in the trial.[23] The Effisayil 1 historical cohort did not collect data on GPPGA pustulation subscores of 0 or 1 and the company used time to pustular clearance as a proxy to the Effisayil 1 trial GPPGA pustulation subscore outcome. It is not clear to us whether this is appropriate since we were not able to confirm it with clinical experts. Time to pustular clearance was collected for patients' typical flare, most severe flare and longest past flare and was categorised as <1 week, 1-2 weeks, 3-4 weeks, 5-8 weeks, 9-12 weeks and >12 weeks.[23] However, it is unclear how the investigator-assigned flare categories ("typical flare", "most severe flare" and "longest past flare") relate to the type of flares included in the Effisayil 1 trial.

The study by Wolf et al.,[26] details of which are provided in CS section B.2.9.2, reported GPPGA pustulation subscores by flare type ("most severe flare" and "other flare"). The company compared the baseline GPPGA pustulation subscores observed in the Effisayil 1 trial to the GPPGA pustulation subscores reported by flare type in Wolf et al.[26] CS Figure 23 shows this comparison and indicates that the trial cohort appears to be more aligned with the "most severe flare" cohort from Wolf et al.

However, the company were not certain whether the Effisayil 1 historical cohort investigator-assigned flare categories and the Wolf et al. flare categories were aligned. Therefore, the company's approach for their base case was to use a weighted average of the "most severe flare" and "typical flare" of the Effisayil 1 historical cohort based on the proportion of most recent flares that were categorised as "most severe" and "other flares" from the Wolf study (55% and 45%, respectively). Table 15 below presents the efficacy of BAC based on Effisayil 1 historical cohort data.[23]

Table 15 Efficacy of BAC based on Effisayil 1 historical cohort

Timepoint	Time to pustular clearance, n (%)			
	Typical past flare (N=28)	Most severe past flare (N=28)	Longest past flare (N=11)	Pooled (weights 0.45 for typical, 0.55 for most severe past flare)
< 1 week	5 (16.7%)	3 (10.7%)	0 (0.0%)	13.40%
1-2 weeks	11 (40.0%)	3 (10.7%)	1 (9.1%)	23.89%
3-4 weeks	6 (23.3%)	13 (46.4%)	3 (27.3%)	36.01%
5-8 weeks	3 (10.0%)	4 (14.3%)	4 (36.4%)	12.37%
9-12 weeks	0 (0.0%)	1 (3.6%)	2 (18.2%)	1.98%
>12 weeks	3 (10.0%)	4 (14.3%)	1 (9.1%)	12.37%

Source: Partly reproduced from CS Table 18 and EAG Table 1 of clarification response document (question B6)
BAC, best available care.

To adjust the Effisayil 1 historical cohort data from Table 15 to the timepoints of the model, the company made the following assumptions: responses for weeks 1 and 2 are informed by "1-2 weeks" by assuming that half of the proportion responds in week 1 and the other half in week 2; and responses for weeks 3 and 4 are informed by "3-4 weeks" by assuming the same as before. It was also assumed that all flares would be resolved by week 12 with a linear cumulative response occurring between week 4 and week 12.

The company's model inputs for the response to BAC are presented in Table 16.

Table 16 Efficacy of BAC based on Effisayil 1 trial and Effisayil 1 historical cohort (company's base case)

Timepoint	Assumption	Efficacy outcome
Day 2	No response	0.0%
Day 3	Effisayil 1 trial: 1/18 (1 new responder)	5.6%
Day 3-8	Effisayil 1 trial: 1/18 (1 new responder)	5.6%
Week 2	Effisayil 1 historical cohort 1-2 weeks/2	11.9%
Week 3	Effisayil 1 historical cohort 3-4 weeks/2	18.0%

Timepoint	Assumption	Efficacy outcome
Week 4	Effisayil 1 historical cohort 3-4 weeks/2	18.0%
Week 12	Remaining	40.9%

Source: Partly reproduced from company's model cells 'Data Library'!U36:U42

4.2.6.2.2 Structured Expert Elicitation (SEE) exercise

The SEE exercise is described in CS section B.2.9.3 and Appendix M. It was carried out by the company with the aim of identifying the treatments used in the UK to treat GPP flares and the efficacy and safety profiles of the current treatments.[14] It comprised two rounds of elicitation (one individual round and one group round), concluding in an expert consensus response to each question. [redacted] experts from the UK and Ireland participated in the elicitation exercise, [redacted] of whom worked in the NHS. [redacted] experts attended the consensus workshop. Analysis of results utilised the SHELF software, through R, to fit probabilistic distributions on absolute and multinomial probabilities.[14]

Table 17 shows the efficacy of BAC elicited by the group of experts participating in the SEE exercise for a GPPGA pustulation subscore of 0 and a GPPGA pustulation subscore of 0 or 1.[14] The experts elicited estimates for moderate and severe flares separately. We have applied the same weight as the company for moderate (45%) and severe flares (55%) and calculated the pooled efficacy of BAC for overall moderate-to-severe GPP flares.

Table 17 Efficacy of BAC for GPPGA pustulation subscore of 0 and 0 or 1, based on SEE exercise (consensus output)

Timepoints	GPPGA pustulation subscore 0			GPPGA pustulation subscore 0, 1		
	LL	Median	UL	LL	Median	UL
Moderate						
Day 1-2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 2-3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 3-8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 2-4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Severe						
Day 1-2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 2-3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 3-8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Timepoints	GPPGA pustulation subscore 0			GPPGA pustulation subscore 0, 1		
	LL	Median	UL	LL	Median	UL
Week 2-4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pooled (45% moderate, 55% severe)						
Day 1-2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 2-3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 3-8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 2-4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Boehringer Ingelheim. Structured Expert Elicitation, 2022[14]
BAC, best available care; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment; LL, lower level; UL, upper level.

To adjust the SEE estimates from Table 17 to the timepoints of the Effisayil 1 trial, we made the following assumptions aligned with the previous company's approach: no patient responds on day 1; responses for day 2 are informed by "day 1-2"; responses for day 3 are informed by "day 2-3"; responses for day 8 are informed by "day 3-8"; responses for weeks 2, 3 and 4 are informed by "week 2-4" by assuming that a third of the proportion responds in week 2, a third responds in week 3 and the remaining third responds in week 4. It was assumed that all flares would be resolved by Week 12 with a linear cumulative response occurring between week 4 and week 12.

The model inputs for the response to BAC based on the SEE exercise are presented in Table 18.

Table 18 Efficacy of BAC based on SEE exercise to use in the model

Timepoint	GPPGA subscore 0	GPPGA subscore 0 or 1
Day 2	[redacted]	[redacted]
Day 3	[redacted]	[redacted]
Day 8	[redacted]	[redacted]
Week 2	[redacted]	[redacted]
Week 3	[redacted]	[redacted]
Week 4	[redacted]	[redacted]
Week 12	[redacted]	[redacted]

Source: EAG calculations
BAC, best available care; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment.

4.2.6.2.3 EAG view

Using data from Effisayil 1 trial to inform the efficacy of BAC for the first week of the model is appropriate, in the sense that it provides a direct comparison between the intervention and the comparator. However, it seems unrealistic that patients in the placebo arm of the Effisayil 1 trial do not receive any standard of care treatments. The EAG is uncertain on how the relative efficacy of spesolimab may change if patients receive any standard of care treatments. Therefore, using trial data to inform the efficacy of BAC does not seem to appropriately reflect UK reality. Consequently, we do not use the Effisayil 1 trial for BAC efficacy in the EAG base case, however we explore this assumption in alternative scenarios (see Table 19 below).

The Effisayil 1 historical cohort[23] uses the same cohort of patients as in the Effisayil 1 trial, however it is unclear how long in the past the flares have occurred and whether the standard of care at that time is similar to the standard of care currently used. It is also uncertain which treatments were included in standard of care for individual flares and how they align with the standard of care for the UK, and therefore how generalisable the results from Effisayil 1 historical cohort are to the current UK reality.

The SEE exercise[14] provides relevant data to inform the economic model for the efficacy of BAC, as the estimates elicited were in relation to the best available care as seen in the experts' practice, which would therefore be relevant to the NHS. Currently, the company's base case matches historical cohort efficacy with SEE best available care information. The EAG considers to be more appropriate and relevant to the NHS to match SEE estimates of BAC efficacy with SEE estimates of best available care treatments.

The limitations of the SEE exercise in relation to the estimation of the efficacy of BAC are:

- Lower quality source of evidence compared to clinical trials and RWE studies.
- Potential for lack of consistency in the definition of response to treatment – although clinicians had pictures and definitions of the GPPGA components to aid them, there could have been variation in how they interpreted treatment response. The EAG considers that other studies might have a similar problem in the interpretation of treatment response between different clinicians.
- Discrepancy in the estimates of efficacy between clinicians in the first round of elicitation – however, when the results were pooled, all the experts agreed with the outcomes obtained, which may provide some confidence on the results.

- Expert interpretation of best available care may have included more efficacious treatments than commonly used in UK or Ireland practice (for instance, other clinicians might use less aggressive treatment to reduce the possibility of adverse events).

As mentioned in section 4.2.6.1 for spesolimab, whether the use of a GPPGA pustulation subscore of 0 or 1 appropriately reflects the resolution of a flare is uncertain. We tested alternative scenarios assuming that treatment response is defined as a GPPGA pustulation subscore of 0 for the BAC arm (see Table 19 below).

Based on the discussion above, we use the SEE exercise estimates for a GPPGA pustulation subscore of 0 or 1 in the EAG base case from Day 1 until the end of the time horizon (Table 19). Nevertheless, we acknowledge the high uncertainty of this assumption and the limitations of each source; therefore, we tested all the alternatives presented in Table 19 in scenario analyses. We note that for each scenario not using the Effisayil 1 trial to inform the efficacy of BAC during the first week (including the EAG base case), we also changed the composition of the comparator arm to include active treatments during this week. The active treatments included are the ones from the SEE exercise (see Figure 1 above).

Table 19 Alternative scenarios for the efficacy of BAC

	GPPGA subscore of 0 or 1				GPPGA subscore of 0	
	Effisayil 1 trial (week 1) + Effisayil 1 historical cohort (company's base case)	**Effisayil 1 historical cohort (company's scenario)**	**SEE exercise (EAG base case)**	**Effisayil 1 trial (week 1) + SEE exercise**	**SEE exercise**	**Effisayil 1 trial (week 1) + SEE exercise**
Day 2	0.0%	0.0%	[redacted]	[redacted]	[redacted]	[redacted]
Day 3	5.6%	5.74%	[redacted]	[redacted]	[redacted]	[redacted]
Day 8	5.6%	19.60%	[redacted]	[redacted]	[redacted]	[redacted]
Week 2	11.9%	11.9%	[redacted]	[redacted]	[redacted]	[redacted]
Week 3	18.0%	18.0%	[redacted]	[redacted]	[redacted]	[redacted]
Week 4	18.0%	18.0%	[redacted]	[redacted]	[redacted]	[redacted]

EAG, External Assessment Group; GPPGA, Generalised Pustular Psoriasis Physician Global Assessment; SEE, structured expert elicitation.

4.2.6.3 Mortality

The mortality risk associated with a GPP flare is described in CS section B.3.3.4. The company applied a daily rate of death of 0.096% for patients in the ICU unit. This rate was derived from a French SNDS study in which 2.6% of patients (15 out of 569) died within four weeks after their last flare.[29]

The company stated that the French study was aligned with other European studies – Augey et al.[30] (reporting a rate of approximately 2%) and Kromer et al.[31] (reporting a rate of approximately 3%).

Experts participating in the SEE exercise were asked to predict the number of patients with a GPP flare lasting beyond four weeks who would die due to any reason. The final consensus output shows [redacted].[14] This results in a weighted average mortality of [redacted] for overall moderate-to-severe flares, based on the weights of 45% for moderate flares and 55% for severe flares, as above. We note that using this mortality rate has a low impact on the model results.

In response to clarification question B4, the company updated their model to add non-disease-specific general population mortality for patients resolved from a GPP flare in both the intervention and comparator arms. This were obtained from the UK Office of National Statistics National Life Tables for 2017-2019 in England (pre-COVID). The yearly rates have been converted to daily rates to fit the model's daily cycle. The EAG is unclear on why the company have not used the most recent version of the National Life Tables for 2020-2022, but we note that adding general population mortality to the model has a low impact on the model results.

4.2.6.4 Adverse event incidence

Adverse events are described in CS section B.3.3.3. The following adverse events were included in the company's base case: serious infection, tuberculosis reactivation and liver injury. Serious infection and tuberculosis reactivation were identified as important adverse events in a Cochrane review of biological side effects and a review of related NICE technology appraisals for psoriasis, psoriatic arthritis, rheumatoid arthritis, ankylosing spondylitis, ulcerative colitis and Crohn's disease.[32-40] Liver injury was included based on clinical advice. The company applied a weekly probability of experiencing an adverse event while receiving treatment in the economic model.

The incidences of serious infection and tuberculosis for the BAC arm were obtained from previous NICE appraisals in other disease areas, TA375 (rheumatoid arthritis) and TA383 (ankylosing spondylitis and non-radiographic axial spondyloarthritis), for which similar types of treatment to the BAC arm are used.[37,38] The incidence of liver injury was based on the SEE exercise as no other source was identified by the company.[14] The company did not model any adverse events for patients receiving BAC during the first week of treatment since these patients were assumed not to receive any active treatments during this period.

The incidences of serious infection and liver injury for the spesolimab arm were obtained from the Effisayil 1 trial and it was assumed to be equal to the probability of experiencing an event over the first week of treatment (after the first dose of spesolimab).[16] Due to lack of evidence, rates of tuberculosis were assumed to be equal to the rates for BAC. CS Table 20 shows the incidence rates for each adverse event.

In the SEE exercise, the incidence of adverse events elicited by the experts were [redacted] for serious infection and [redacted] for tuberculosis.[14] We note that changing these values has a negligible impact on the model results.

EAG comment on treatment effectiveness and extrapolation

In the company's base case, the efficacy of spesolimab is based on Effisayil 1 trial data while the efficacy of BAC is based on Effisayil 1 trial data for the first week and, beyond the first week, it is based on data from the Effisayil 1 historical cohort.

The standard of care treatments used to treat flares in the Effisayil 1 historical cohort are unknown for individual types of flares and therefore it is unclear how similar or different they might be from the best available care treatments currently used in UK practice. The EAG considers the SEE exercise estimates to reflect UK reality more closely and to be aligned with the modelled comparator treatments (which were elicited by the same experts). We use the SEE exercise estimates based on a GPPGA pustulation subscore of 0 or 1 from day 1 until the end of the time horizon in the EAG base case. We acknowledge, however, the uncertainty of this assumption and the limitations of each source and therefore we consider this to be a key issue (Key Issue 6). We tested several alternatives for the efficacy of BAC in scenario analyses.

The mortality and adverse event inputs seem reasonable, and we note that changing these input values has a low impact on the model results.

4.2.7 Health related quality of life

4.2.7.1 Systematic literature review for utilities

The company conducted a systematic literature review of HRQoL studies in patients with GPP. The methodology is described in CS Appendix I. The search date was 27 December 2022, with an update on 6 May 2024. CS Appendix I Table 46 presents the inclusion and exclusion criteria. We consider that the company searched an adequate range of sources, and the searches are adequately up to date. The searches identified 29 studies describing HRQoL in patients with GPP, but none of these reported health state utility or disutility values to be used in the model.

4.2.7.2 Study-based health-related quality of life

The health-related quality of life data used in the model is described in CS section B.3.4.5. EQ-5D-5L data were collected from the Effisayil 1 trial daily during the first week, then every week until week 4, and then in week 8 and 12.[16] The company mapped the EQ-5D-5L individual domain scores into a single utility index score using tariffs for the UK and mapped EQ-5D-5L to EQ-5D-3L using the approach described in Hernandez-Alava et al.[41]

The baseline utility in the Effisayil 1 trial was used as the utility of patients in the GPP flare health state in the economic model, regardless of treatment. The utility values for patients who recovered from a GPP flare (resolved flare health state) were based on data from the Effisayil 2 trial using a repeated-measures mixed effects model, which accounts for multiple measures per patient (clarification response B8). It is unclear why the company did not use data from the Effisayil 1 trial for patients with a GPPGA pustulation subscore of 0 or 1 to reflect a resolved flare. We note that the objective of Effisayil 2 trial was to assess the efficacy and safety of subcutaneous spesolimab for GPP flare prevention. The reason why the company used data from this trial for the utility of the resolved flare health state was because patients in Effisayil 2 trial spent most of their time in the trial without experiencing a GPP flare. In addition, the utility value for general population with a similar age and sex distribution is 0.85,[42] which is similar to the estimated utility for a resolved flare. Patients in the ICU had a utility of zero, as assumed in previous appraisals.[43] Table 20 presents the utility values used in the company's base case.

It seems reasonable to the EAG to assume that the quality of life of patients that have had resolved a flare for a while would be similar to the quality of life of patients not experiencing a flare. However, we would expect that the quality of life of patients immediately after resolving a flare is still slightly lower than if they have not had a flare for several months (or if their last flare was a mild one). This is mainly important for patients with moderate-to-severe

GPP flares that have required hospital stay and/or for patients that experienced some treatment adverse events. Although we were unable to check this assumption with clinical experts, we note that increasing or decreasing the utilities does not have a significant impact on the model results.

Table 20 Utility values used in the company's base case

State	Utility value, mean (standard error)	95% confidence interval
Active flare	[redacted]	[redacted]
Resolved flare	[redacted]	[redacted]
Hospitalisation ICU	0 (0)	0-0

Source: Partly reproduced from CS Table 21.
ICU, intensive care unit.

4.2.7.3 Adverse event disutilities

The adverse event disutilities were applied in the company's model and are described in CS section B.3.3.4.

The duration of each disutility was assumed to be 28 days. The disutility of serious infection was obtained from NICE TA375 (0.156),[44] and the company assumed a similar disutility for tuberculosis in the absence of any other evidence. For liver injury, a disutility of 0.0956 was obtained from a UK study by Sullivan et al.[45]

EAG comment on HRQoL

In the company's base case, utility values were informed by EQ-5D-5L data (mapped to EQ-5D-3L) derived from the Effisayil 1 and 2 trials. We agree with the company's approach to model health state utilities and adverse event disutilities. The EAG notes that no other sources reported utility values for GPP flares and also that changing the utility values do not have a significant impact on the model results.

4.2.8 Resources and costs

4.2.8.1 Literature review of cost and resource studies

The company's systematic literature review used to identify healthcare resource use studies is described in section 4.1 above. The results of the review are shown in CS Appendix H. The systematic literature review did not identify any UK-specific studies for patients with GPP with resource use data differentiated by whether patients experienced a flare or not.

The company assumed that European studies might also be applicable to the UK setting and therefore considered two European studies to be relevant: Viguier et al. 2024[46] and Wolf et al. 2024.[26] Viguier et al. is a retrospective study using the SNDS to identify healthcare resource use for patients who were hospitalised with GPP in France.[46] Wolf et al. is a retrospective study reporting rates of hospitalisation along with length of stay for patients with a GPP flare in Central and Eastern European countries.[26] CS Table 7 shows the healthcare resource use evidence from these studies.

In the company's base case, the study by Wolf et al. was used to inform the proportion of patients treated as inpatients. It was also used, as well as the Viguier et al. study, to inform company's sensitivity and scenario analyses (see Table 23).

4.2.8.2 Drug acquisition and administration costs

CS section B.3.5.1 presents the drug acquisition and administration costs. Acquisition costs were obtained from the British National Formulary (BNF)[47] and Drugs and Pharmaceutical Electronic Market Information Tool (eMIT).[48]

CS Table 22 shows the route of administration, pack cost, pack size and strength for each treatment used in the economic model. For all treatments, the company applied the minimum cost per mg. We note that the price for acitretin in CS Table 22 is incorrect and the company clarified that a price of £17.10, based on eMIT, should be considered for the economic model (clarification question B6). Similarly, the price for infliximab reported in CS Table 22 is incorrect and the company clarified that the price of £377.66 for a strength of 100mg (already used in the economic model) is correct (clarification question B7).

CS Table 23 shows the treatment dosing schedules for each treatment. Only spesolimab has a recommended treatment dosing regimen for GPP. The proportion of patients receiving a second dose of spesolimab was obtained from Effisayil 1 trial. For the patients eligible for a second dose of spesolimab who have not received it, the company assumed that this was because of previous adverse events and that these patients have ciclosporin on Day 8 instead (see section 4.2.4 above). For the BAC comparators, the dosing schedule was sourced from BNF assuming the same dosing as for severe psoriasis or plaque psoriasis, where available. If applicable, the company used the higher dosing regimens for each of the drugs to represent GPP treatment. In response to clarification question B7, the company explained that clobetasol propionate (a topical steroid) is only used in Week 1 of the economic model in a scenario where active treatment in the first week is modelled.

CS Table 24 shows the composition of the comparator arm in the company's base case. The weekly acquisition costs for spesolimab and the comparators are shown in Table 21.

Table 21 Weekly acquisition costs for spesolimab (PAS price) and the comparators (list price)

Treatment	Weekly cost	Source
Spesolimab	[redacted]	Company data on file
Immunosuppressant (ciclosporin)	£34.95	BNF[47]
TNF inhibitors (infliximab)	£1,360.14	eMIT[48]
Methotrexate	£14.55	
Retinoid (acitretin)	£5.98	
IL-23 inhibitors (guselkumab)	£2,250.00	
IL-17 inhibitors (secukinumab)	£1,218.78	
Ustekinumab	£2,147.00	
Topical steroids (clobetasol propionate)	£3.76	

Source: Partly reproduced from CS Table 25

In response to clarification question B13, the company provided an updated table with the correct administration costs to be considered in the model (Table 9 of the clarification response document). We reproduce these costs in Table 22 below.

Table 22 Administration costs used in the company revised base case model

Administration route	Cost in model	Source and justification
Oral	£0	Assumption: TA442, TA475, TA511, TA907
Subcutaneous (first dose)	£141.00	PSSRU 2023, unit costs of health and social care 2022/23, Nurse (Community-based band 5 without qualifications), wage cost per hour (3 hours); Cost incurred only once on the first use of subcutaneous therapy, as patients self-administer thereafter. As per TA935.[36,49]

Administration route	Cost in model	Source and justification
Subcutaneous (subsequent doses)	£0	See above.
Intravenous	£174.89	Assumed to be the total unit cost of a dermatology outpatient appointment (as per TA907) applied for each intravenous administration. National Schedule of NHS Costs 2021–2022. WF01A: Non-Admitted Face-to-Face Attendance, Follow-up. Dermatology (Service Code 330). Outpatient procedure[50]

Source: Reproduced from Table 9 of clarification response document
PSSRU, Personal Social Services Research Unit

4.2.8.3 Health state costs

The resource use and costs applied to each health state are described in CS section B.3.5.3, including outpatient and inpatient hospital visits, and ICU stay.

Resource use data were obtained from the two relevant European sources identified through the systematic literature review (see section 4.2.8.1) and the SEE exercise.[51] The company considered these sources more appropriate than the Effisayil 1 historical cohort because the location of treatment centres was mostly outside of Europe in the historical cohort study. CS Table 7 shows the healthcare resource use evidence for the two European studies while CS Table 27 shows the resource use elicited by the experts participating in the SEE exercise.

Table 23 presents the resource use modelled in the company's base case and scenario analyses.

Table 23 Modelled resource use implemented in the company revised base case model

Resource use during a flare	Value	Source
Proportion of patients treated as inpatients (spesolimab)	38.8%	Assumed 50% reduction in inpatient rates, based on the relative reduction of 48.4% in the proportion of patients with an active flare for spesolimab versus placebo in the Effisayil 1 trial.

Resource use during a flare	Value	Source
Proportion of patients treated as inpatients (BAC)	77.6%	Overall inpatient rate for a patient's last flare by Wolf et al.[26] Company scenarios: 93% (inpatient rate for a patient's most severe flare by Wolf et al.)[26] [redacted] (inpatient rate elicited for moderate/severe flares from the SEE exercise)[51]
Proportion of patients treated as outpatients (spesolimab)	61.2%	Assumption that all patients that did not have a hospitalisation would be treated as outpatients
Proportion of patients treated as outpatients (BAC)	22.4%	As above
Average weekly number of outpatient appointments: active flare	[redacted]	HCRU SEE[51]
Average annual number of outpatient appointments: resolved flare	[redacted]	HCRU SEE[51]
Proportion of inpatients treated in ICU (spesolimab)	0%	Assumption based on the fact that rapid resolution of flares as observed with spesolimab would alleviate the risk of the need for ICU care Company scenario: [redacted] based on HCRU SEE[51]
Proportion of inpatients treated in ICU (BAC)	[redacted]	HCRU SEE[51] Other sources: 25% (142/569) from Viguier et al.[46] 11.5% (3/26) from Wolf et al.[26] This uncertainty was explored via deterministic sensitivity analyses.

Resource use during a flare	Value	Source
Proportion of patients in ICU requiring mechanical ventilation	[redacted]	HCRU SEE[51]
Length of stay in ICU without mechanical ventilation, days	[redacted]	Maximum length of stay derived from the mean length of stay elicited in the SEE exercise and the response rates over time for BAC. SEE exercise: [redacted] days (mean length of stay) – not a company scenario.[51]
Length of stay in ICU with mechanical ventilation, days	[redacted]	Maximum length of stay derived from the mean length of stay elicited in the SEE exercise and the response rates over time for BAC. Company scenario: SEE exercise: [redacted] days (mean length of stay) [51]
Proportion of hospitalised patients treated in day care	[redacted]	Hospital Episodes Statistics data for GPP Non elective events, April 2016 – February 2022. Of [redacted] inpatient admissions where the primary patient diagnosis was GPP (ICD-10 code L401), [redacted] ([redacted]) were day cases (Clarification Response B9).

Source: Partly reproduced from CS Table 29 and model sheet "Data Library", cells U62:U95.
BAC, best available care; GPP, generalised pustular psoriasis; HCRU, health care resource use; ICU, intensive care unit; SEE, structured expert elicitation.

The company assumed that inpatients will be discharged once their pustules have resolved, i.e., once they reached a GPPGA pustulation subscore of 0 or 1 in the model (see section 4.2.2.1 above).

For the BAC arm, the proportion of patients treated as inpatients (77.6%) was based on the overall inpatient rate for a patient's last flare reported in Wolf et al.[26] We consider that the company have conducted relevant scenarios to check this assumption (see Table 23). Spesolimab dominates BAC in both scenarios. The proportion of inpatients was modelled by assuming that all patients (77.6%) were hospitalised on Day 1 (day of treatment) and, after that, patients could only be discharged from hospital. We consider that this assumption lacks

face validity as patients are likely to be hospitalised beyond Day 1. However, there is no evidence on which proportion of patients were hospitalised at each day and this was modelled similarly for both arms, BAC and spesolimab. We consider therefore the company's approach to be reasonable.

The company assumed that the proportion of the spesolimab arm patients treated as inpatients was half that of the BAC arm (33.8%) (see Table 23). This assumption was based on the relative reduction of 48.4% in the proportion of patients with a GPPGA pustulation subscore ≥ 2 (active flare) for spesolimab versus placebo in the Effisayil 1 trial. As such, spesolimab increases the number of patients with resolved flares (by reducing the number of patients with active flares, see section 4.2.6.1) which, in practice, directly leads to a reduction in the number of patients that will be admitted to hospital. However, the EAG considers it to be very uncertain whether spesolimab has an additional residual benefit in reducing hospitalisation rates for patients not responding to treatment (patients with active flares) as there is no evidence showing that benefit or the size of that benefit. We consider the company's assumption of a 50% reduction is likely to be optimistic in the absence of data. The EAG considers this uncertainty to be a key issue (Key Issue 7) and we believe that further clinical expert opinion may be valuable. Although the company mentioned in the CS that alternative inpatient rates for spesolimab were explored in scenario analyses, these are not reported. We therefore tested alternative reductions in scenario analyses (30%, 20%, 10% and 0%) and further clinical expert opinion would help in reducing this uncertainty.

The proportion of patients hospitalised treated in ICU for the BAC arm was based on the SEE exercise ([redacted]). Viguier et al.[46] reported that 25% of hospitalised patients were treated in ICU while a 11.5% proportion was reported by Wolf et al.[26] The uncertainty around this assumption was tested in deterministic sensitivity analysis, by changing the base case value by ± 20%. We note that the results have not changed significantly and spesolimab dominates in both situations.

The company assumed that patients treated with spesolimab would not require treatment in the ICU. The company argued that spesolimab has demonstrated rapid and sustained control of disease, with skin clearance as early as day 3 and pustulation clearance as early as day 2. Spesolimab also demonstrated clinically meaningful improvements in patient reported outcomes and markers of inflammation. Therefore, the company assumed that spesolimab preventing patients from requiring admission to an ICU is not unreasonable (clarification question B12). In the company's scenario where the same proportion of patients in ICU for the BAC arm was used for spesolimab ([redacted]), spesolimab dominates the

comparator. Again, as above, we are uncertain whether spesolimab reduces the risk of being admitted to ICU for those patients with active flares in addition to the already modelled reduction of active flares. We would benefit from further clinical expert opinion, and we tested alternative proportions in scenario analyses: 5%, 10%, 15% and [redacted].

The company applied the maximum length of stay in ICU in the model and that was derived from the mean length of stay elicited by clinical experts in the SEE exercise[51] and the response rates over time for BAC. As time in hospital was modelled as the time needed for patients reaching a GPPGA pustulation subscore of 0 or 1, the purpose of this adjustment was to ensure that the modelled output mean length of stay for BAC was similar to the elicited values. We note that removing this adjustment has a very low impact on the model results.

Costs were mainly obtained from published literature, 2021/2022 NHS reference costs[50] and Unit Costs of Health and Social Care 2023;[49] these are summarised in CS Table 28.

4.2.8.4 Adverse event costs

CS section B.3.5.4 and CS Table 30 presents the unit costs applied to adverse events included in the model: serious infection, tuberculosis and liver injury. The costs were obtained from the NHS reference costs and were updated to 2021/2022 in response to clarification question B11. The updated costs are shown in Table 24 below.

Table 24 Unit costs for the treatment of adverse events (updated in response to clarification question B11)

Adverse event	Unit cost	Source
Serious infection	£2,718.58	National schedule of NHS costs 2021/2022. Total HRGs, weighted average of DZ11K-DZ11V and WJ06A-WJ06J (sepsis and pneumonia)
Tuberculosis	£4,253.13	National schedule of NHS costs 2021/2022. Total HRGs, weighted average of DZ14F-DZ14J (pulmonary, pleural or other tuberculosis)
Liver injury	£2,262.30	National schedule of NHS costs 2021/2022. Total HRGs, weighted average of GC01E-GC01F (liver failure disorders without interventions)

Source: Reproduced from Clarification Response B11 and partly reproduced from CS Table 30.
HRG, healthcare resource group; NHS, National Health Service.

4.2.8.5 Other costs

CS section B.3.5.5 reports an end-of-life cost of £5,877.88 (one-off cost) implemented by the company in the economic model. This cost was inflated from the original cost of £4,580 obtained from Georghiou et al.[52]

EAG comment on resources and costs

The costs for the administration and acquisition costs, adverse events and end-of-life are reasonable and based on relevant sources.

The health state costs were based on the SEE exercise estimates, two European prospective studies and some assumptions. We are unclear about two of the company's assumptions: the proportion of inpatients and proportion of patients admitted to ICU in the spesolimab arm. These assumptions are associated with a lot of uncertainty and therefore several scenarios were conducted by the EAG to test their impact on the model results (Key Issue 7).

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company's original base case results are shown in CS Table 34, with an incremental cost-effectiveness ratio (ICER) of -£194,860 per QALY for spesolimab versus best available care (spesolimab dominates). This and all other cost-effectiveness results presented in this report were conducted with a confidential Patient Access Scheme (PAS) price discount for spesolimab of [redacted] In response to the EAG's clarification questions, the company corrected the following input parameters in a company revised base case:

- Cost of acitretin: £17.10 instead of £17.92 (clarification question B6)
- Cost of ciclosporin: £41.59 instead of £68.28 (clarification question B7)
- Treatment dosing schedule for ustekinumab (clarification question B7)
- Weekly acquisition cost for clobetasol propionate: £3.76 instead of £0.04 (clarification question B7)
- Adverse event costs updated using the National Schedule of NHS costs 2021/22 (clarification question B11)
- Subcutaneous administration cost: £141 instead of £57 (clarification question B13)
- Inclusion of general population mortality (clarification question B4)

The company revised base case results using the PAS price for spesolimab are shown in Table 25. The ICER for spesolimab versus best available care is -£172,713 per QALY (spesolimab dominates).

Table 25 Company revised base case results after clarification responses using PAS price for spesolimab

Treatment	Total		Incremental		ICER (£/QALY)
	Cost (£)	QALYs	Cost (£)	QALYs	
BAC	[redacted]	[redacted]			-£172,713 (spesolimab dominates)
Spesolimab	[redacted]	[redacted]	[redacted]	[redacted]	

Source: Clarification response document Table 13, company revised economic model (Deterministic Results!F15)
BAC, best available care; QALY, quality adjusted life year; ICER, incremental cost-effectiveness ratio.

5.2 Company's sensitivity analyses

5.2.1 Deterministic sensitivity analyses

The company submission reports the original deterministic sensitivity analyses (DSA) in section B.3.11.2. The results of the updated DSA conducted on the company revised model are provided in Figure 9 of the clarification response document. Parameters are varied by +/- 20% of the deterministic point estimate for each parameter. The top 10 most influential variables are presented in a tornado diagram (clarification response document Figure 9), which shows the DSA results including the PAS price for spesolimab. The three main drivers of the model are the percentage of patients treated as inpatients, the acquisition cost of spesolimab, and the daily cost of inpatient care. However, the ICER remained below -£50,000 per QALY in all cases (spesolimab still dominated the comparator).

5.2.2 Probabilistic sensitivity analyses

The company conducted a probabilistic sensitivity analysis (PSA) with input parameter uncertainty distributions as presented in CS Appendix O. The PSA was run for 1,000 iterations, and mean results for the PSA conducted on the company revised model were reported in Table 14 of the clarification response document using the PAS price for spesolimab. The cost-effectiveness plane and cost-effectiveness acceptability curve are presented in Figure 7 and Figure 8 of the clarification response, respectively. All required variables were included with appropriate distributions. The probabilistic results were in line with the deterministic results when run by the EAG.

5.2.3 Scenario analyses

The company reports the original results of nine scenario analyses in CS Table 38 (PAS price for spesolimab included), while the updated results of the same scenario analyses conducted on the company revised base case model are reported in Table 15 of the clarification response document (PAS price for spesolimab included). The scenario analyses explored the efficacy of BAC, inpatient proportions and costs, and length of stay in an ICU unit. Only one scenario, using the Effisayil 1 historical cohort data from day 0 to inform the efficacy of BAC, resulted in a positive ICER of £16,523 per QALY. For all other scenarios, spesolimab dominated best available care. The company reports in CS section B.3.5.3 that they would run scenario analyses on the proportion of patients receiving spesolimab that were hospitalised, however the EAG were unable to find these scenarios in the company submission or the clarification response document. The EAG will perform these as exploratory analyses in section 6.1 of this report.

5.3 Model validation and face validity check

We conducted a range of checks on the company model using an EAG checklist:

- Input checks: comparison of all parameter values in the model against the values stated in the company submission and cited sources.
- Output checks: replication of results reported in the company submission using the company model.
- 'White box' checks: manual checking of formulae working from the cohort-level Markov model, which includes reviewing the calculations across each cycle and working backwards to trace links to input parameters and forwards to the results.
- 'Black box' checks: working through a list of tests to assess whether changes to key model inputs or assumptions have the expected effects on the model results.

The model is well-implemented, and no coding errors were identified.

5.3.1 EAG corrections to the company model

The EAG corrected the application of the utility value for the ICU in the model. As described in section 4.2.7, the utility score for patients in the ICU is assumed to be zero. However, in the model the company have only applied this utility score to patients in the ICU on mechanical ventilation. The EAG have corrected the company model so that the utility score applies to all patients in the ICU, both on and off mechanical ventilation. The corrected model results are shown in Table 26. The corrected company revised model gives an ICER of -£167,783 per QALY for spesolimab versus best available care (i.e., spesolimab dominates best available care).

Table 26 EAG corrections to the company revised base case model using the PAS price for spesolimab

Treatment	Total		Incremental		ICER
	Cost (£)	QALYs	Cost (£)	QALYs	(£/QALY)
BAC	[redacted]	[redacted]			-£167,783 (spesolimab dominates)
Spesolimab	[redacted]	[redacted]	[redacted]	[redacted]	

BAC, best available care; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio.

5.3.2 EAG summary of key issues and additional analyses

A full summary of EAG observations on key aspects of the company's economic model is presented in Table 27.

Table 27 EAG observations of the key aspects of the company's economic model

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
Key model features			
Model structure	Section 4.2.2 Markov model	We agree	No change
Model assumptions (Relapse of a GPP flare is not considered in the model)	Section 4.2.2.2.1	11.3% of patients in the Effisayil 1 trial received rescue treatment with spesolimab to treat second/recurrent flares, and eight patients received a standard of care escape treatment, [redacted]. We consider this to be a key issue (Key Issue 3) as we believe this uncertainty can be resolved with further clinical data and further clinical input.	No change

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
Model assumptions (Maximum flare duration is 12 weeks)	Section 4.2.2.2.3	Evidence from the studies informing the efficacy of spesolimab and BAC shows that a proportion of patients have not responded to treatment or that it is unknown whether they have responded to treatment by week 12.	EAG base case: no change EAG scenarios: • 12.37% of patients in the BAC arm with active flares by the end of the time horizon. • 20.0% of patients in the spesolimab arm with active flares by the end of the time horizon. • 12.37% of patients in the spesolimab arm with active flares by the end of the time horizon. • 5.7% of patients in the spesolimab arm with active flares by the end of the time horizon.
Model assumptions (others)	Section 4.2.2.2	We agree	No change
Population	Section 4.2.3	It is unclear whether the model results can be generalisable to patients experiencing mild flares as the population from the studies informing the economic	No change

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
		model is mainly a moderate-to-severe flare population (Key Issue 1).	
Comparators	Section 4.2.4	The EAG is concerned that the composition of the comparator arm is not aligned with the UK clinical practice and/or the studies informing its efficacy (Key Issue 4).	EAG base case: active treatments from the SEE exercise applied from day 1 until the end of the time horizon (linked to change in efficacy of BAC)
Perspective	NHS and Personal Social Services	We agree	No change
Time horizon	85 days	It is unclear whether a 12-week time horizon is long enough to capture all the consequences of the treatment with spesolimab (Key Issue 5)	No change
Discounting	0%	We agree	No change
Model inputs			
Baseline characteristics	Section Population4.2.3	We agree	No change
Efficacy of spesolimab	Section 4.2.6.1	We are unclear whether the use of a GPPGA pustulation	EAG base case: no change

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
	Effisayil 1 trial, based on a GPPGA pustulation subscore of 0 or 1	subscore of 0 or 1 appropriately reflects the resolution of a flare in UK practice.	EAG scenarios: • Effisayil 1 trial with treatment response defined as a GPPGA pustulation subscore of 0
Efficacy of BAC	Section 4.2.6.2 Effisayil 1 trial (week 1) plus Effisayil 1 historical cohort, based on a GPPGA pustulation subscore of 0 or 1	The standard of care treatments used to treat flares in the Effisayil 1 historical cohort are unknown for the individual types of flares and therefore it is unclear how similar or different they might be from the best available care treatments currently used in UK practice. The EAG considers the SEE exercise estimates to reflect UK reality more closely and to be aligned with the modelled comparator treatments (which were elicited by the same experts). We acknowledge however the uncertainty of this	EAG base case: SEE exercise estimates with treatment response defined as a GPPGA pustulation subscore of 0 or 1 from day 1 until the end of the time horizon. EAG scenarios: • Effisayil 1 trial (first week) + Effisayil 1 historical cohort (GPPGA pustulation subscore of 0 or 1, company's base case) • Effisayil 1 historical cohort (GPPGA pustulation subscore of 0 or 1, company's scenario) • Effisayil 1 trial (first week) + SEE exercise (GPPGA pustulation subscore of 0 or 1) • SEE exercise (GPPGA pustulation subscore of 0) • Effisayil 1 trial (first week) + SEE exercise (GPPGA pustulation subscore of 0)

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
		assumption and the limitations of each source (Key Issue 6).	
Mortality	Section 4.2.6.3 Daily rate of death 0.096%	We agree	No change
Adverse event incidence	Section 4.2.6.4	We agree	No change
Utilities			
Health state utilities	Section 4.2.7.2	We agree	No change
Adverse event disutilities	Section 4.2.7.3	We agree	No change
Severity modifier	Not applied	We agree	No change
Resource use and costs			
Acquisition costs	Section 4.2.8.2	We agree	No change
Administration costs	Section 4.2.8.2	We agree	No change
Proportion of patients treated as inpatients (spesolimab)	Section 4.2.8.3 38.8% (reduction of 50%)	As there is no evidence, it is very uncertain whether patients not responding to treatment with spesolimab, and therefore with an active flare, will have any reduction in hospitalisation rates compared with patients in a	EAG base case: no change EAG scenarios: • 77.6% (no reduction) • 69.84% (reduction of 10%) • 62.08% (reduction of 20%) • 54.32% (reduction of 30%)

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
		similar situation in the BAC arm. We consider the company's assumption is likely to be optimistic in the absence of data (Key Issue 7).	
Proportion of patients treated as inpatients (BAC)	Section 4.2.8.3 77.6%	We agree	No change
Proportion of inpatients treated in ICU (spesolimab)	Section 4.2.8.3 0%	We are uncertain whether spesolimab reduces the risk of being admitted to ICU for those patients with active flares in addition to the already modelled reduction of active flares.	EAG base case: no change EAG scenarios: • 5% • 10% • 15% • [redacted] (company's scenario)
Proportion of inpatients treated in ICU (BAC)	Section 4.2.8.3 [redacted]	We agree	No change
Proportion of patients in ICU requiring mechanical ventilation	Section 4.2.8.3 [redacted]	We agree	No change

Parameter	Company base case	EAG comment	EAG base case/ EAG scenarios
Other health state costs	Section 4.2.8.3	We agree	No change
Adverse event costs	Section 4.2.8.4	We agree	No change
End of life costs	Section 4.2.8.5	We agree	No change
NHS, National Health Service; BAC, best available care.			

6 EAG'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

The EAG performed scenario analyses based upon the uncertainties discussed in section 4 using the EAG corrected company base case. The results of these scenario analyses are presented in Table 28. The EAG have made the following additional changes to the model when altering the efficacy of BAC:

- For scenarios using a GPPGA pustulation subscore of 0, the corresponding efficacy of spesolimab from the Effisayil 1 trial also uses a GPPGA pustulation subscore of 0.
- For the scenarios using the Effisayil 1 historical cohort from Day 1 or the SEE exercise from Day 1, comparator treatments are also included from Day 1.

Spesolimab dominates the comparator (i.e., is more effective and less expensive) in all but two scenarios: using the Effisayil 1 historical cohort from Day 1 with a GPPGA pustulation subscore of 0 or 1, where spesolimab remains cost-effective; and using the SEE exercise from Day 1 with a GPPGA pustulation subscore of 0 or 1, where spesolimab is no longer cost-effective at a willingness-to-pay (WTP) threshold of £30,000.

Table 28 EAG scenario analyses on the EAG corrected company base case using the PAS price for spesolimab

Scenario	Incremental costs	Incremental QALYs	ICER (£/QALY)
EAG corrected company base case	[redacted]	[redacted]	-£167,783 (spesolimab dominates)
Model assumptions			
12.37% of patients in BAC arm with active flare by end of time horizon	[redacted]	[redacted]	-£214,995 (spesolimab dominates)
5.7% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	-£160,012 (spesolimab dominates)
12.37% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	-£149,083 (spesolimab dominates)

Scenario	Incremental costs	Incremental QALYs	ICER (£/QALY)
20% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	-£133,250 (spesolimab dominates)
Comparator costs			
Cost of ciclosporin: £48.50 (NICE requested scenario)	[redacted]	[redacted]	-£167,783 (spesolimab dominates)
Efficacy of BAC: GPPGA pustulation subscore of 0 or 1			
Effisayil 1 historical cohort	[redacted]	[redacted]	£15,680
SEE exercise	[redacted]	[redacted]	£143,574
Effisayil 1 trial (first week) + SEE exercise	[redacted]	[redacted]	-£143,513 (spesolimab dominates)
Efficacy of BAC: GPPGA pustulation subscore of 0			
SEE exercise	[redacted]	[redacted]	-£118,683 (spesolimab dominates)
Effisayil 1 trial (first week) + SEE exercise	[redacted]	[redacted]	-£167,597 (spesolimab dominates)
Proportion of inpatients on spesolimab			
77.6% (0% reduction)	[redacted]	[redacted]	-£5,009 (spesolimab dominates)
69.84% (10% reduction)	[redacted]	[redacted]	-£37,564 (spesolimab dominates)
62.08% (20% reduction)	[redacted]	[redacted]	-£70,119 (spesolimab dominates)
54.32% (30% reduction)	[redacted]	[redacted]	-£102,674 (spesolimab dominates)

Scenario	Incremental costs	Incremental QALYs	ICER (£/QALY)
Proportion of inpatients treated in the ICU on spesolimab			
5%	[redacted]	[redacted]	-£166,247 (spesolimab dominates)
10%	[redacted]	[redacted]	-£164,703 (spesolimab dominates)
15%	[redacted]	[redacted]	-£163,151 (spesolimab dominates)
[redacted]	[redacted]	[redacted]	-£161,151 (spesolimab dominates)

Source: EAG corrected company base case model.
QALY, quality adjusted life-year; ICER, incremental cost-effectiveness ratio.

6.2 EAG's preferred assumptions

Based on the EAG critique of the company's model discussed in section 4, we have identified one significant aspect of the company base case with which we disagree. Our preferred model assumption is to use the SEE exercise with a GPPGA pustulation subscore of 0 or 1 from Day 1 for the efficacy of BAC rather than the Effisayil 1 trial (first week) plus the Effisayil 1 historical cohort. As noted above, this also includes using comparator treatments from Day 1. Although the EAG has concerns with other assumptions made by the company (see Key Issues in section 1.5), we do not have the required data to make an informed decision on what the correct assumptions should be. The EAG base case results are reported in Table 29 below. When making this sole change, the ICER increases to £143,574 per QALY for spesolimab versus best available care.

Table 29 EAG preferred base case results using PAS price for spesolimab

Treatment	Total		Incremental		ICER (£/QALY)
	Cost (£)	QALYs	Cost (£)	QALYs	
BAC	[redacted]	[redacted]			£143,574
Spesolimab	[redacted]	[redacted]	[redacted]	[redacted]	

6.2.1 EAG scenario analyses

For consistency, we performed the same scenarios on our EAG base case model as previously run on the EAG corrected company base case model. Table 30 summarises these results. The ICERs varied between -£167,783 (the EAG corrected company base case model using the Effisayil 1 trial and the Effisayil 1 historical cohort for the efficacy of BAC) and £581,059 (20% of patients in the spesolimab arm with active flare by the end of the 12-week time horizon).

Table 30 EAG scenario analyses on the EAG preferred base case using the PAS price for spesolimab

Scenario	Incremental costs	Incremental QALYs	ICER (£/QALY)
EAG base case	[redacted]	[redacted]	£143,574
Model assumptions			
12.37% of patients in BAC arm with active flare by end of time horizon	[redacted]	[redacted]	-£12,812 (spesolimab dominates)
5.7% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	£213,829
12.37% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	£335,032
20% of patients in spesolimab arm with active flare by end of time horizon	[redacted]	[redacted]	£581,059
Comparator costs			
Cost of ciclosporin: £48.50 (NICE requested scenario)	[redacted]	[redacted]	£143,351
Efficacy of BAC: GPPGA pustulation subscore of 0 or 1			
Effisayil 1 trial (first week) + Effisayil 1 historical cohort (company base case)	[redacted]	[redacted]	-£167,783 (spesolimab dominates)
Effisayil 1 historical cohort	[redacted]	[redacted]	£15,680
Effisayil 1 trial (first week) + SEE exercise	[redacted]	[redacted]	-£143,513 (spesolimab dominates)

Scenario	Incremental costs	Incremental QALYs	ICER (£/QALY)
Efficacy of BAC: GPPGA pustulation subscore of 0			
SEE exercise	[redacted]	[redacted]	-£118,683 (spesolimab dominates)
Effisayil 1 trial (first week) + SEE exercise	[redacted]	[redacted]	-£167,597 (spesolimab dominates)
Proportion of inpatients on spesolimab			
77.6% (0% reduction)	[redacted]	[redacted]	£455,571
69.84% (10% reduction)	[redacted]	[redacted]	£393,172
62.08% (20% reduction)	[redacted]	[redacted]	£330,772
54.32% (30% reduction)	[redacted]	[redacted]	£268,373
Proportion of inpatients treated in the ICU on spesolimab			
5%	[redacted]	[redacted]	£148,852
10%	[redacted]	[redacted]	£154,190
15%	[redacted]	[redacted]	£159,590
[redacted]	[redacted]	[redacted]	£166,593

6.3 Conclusions on the cost effectiveness evidence

The company developed a three-state Markov model to estimate the cost-effectiveness of spesolimab compared to best available care for patients with generalised pustular psoriasis flares. Response to treatment was obtained from the Effisayil 1 trial for the spesolimab arm and from the Effisayil 1 trial (Week 1) combined with the Effisayil 1 historical cohort (beyond Week 1) for the best available care arm. The EAG considers the overall structure of the model to be appropriate, however there are concerns with the length of the time horizon (12 weeks), the lack of implementation of second/recurrent GPP flares and the assumption that all patients have a resolved flare by Week 12 (see section 4.2.2). The company made some minor changes to the model in response to clarification questions. The company's revised base case produced an ICER of -£172,713 per QALY, using a confidential PAS discount of [redacted] for spesolimab (spesolimab dominates best available care). The EAG identified an inconsistency regarding the application of utility scores for patients in the ICU in the model; this was corrected and resulted in an ICER of -£167,783 per QALY (see section 5.3.1).

The EAG's preferred base case comprises a singular change to the EAG corrected base case: the efficacy of BAC using the SEE exercise from Day 1 (based upon a GPPGA pustulation subscore of 0 or 1). This also includes comparator treatments from Day 1. Although the EAG has concerns with other assumptions made by the company (see Key Issues in section 1.5), no data was available to make informed decisions about alternative assumptions for the EAG base case. The EAG preferred base case increases the ICER to £143,574 per QALY gained for spesolimab compared to best available care. Both the company and EAG base cases are most sensitive to changes in the source of efficacy for the BAC arm, the proportion of patients in the spesolimab arm with active flares by the end of the 12-week time horizon, and the proportion of patients in the spesolimab arm admitted into hospital.

7 SEVERITY

Severity is described in CS section B.3.6. The company used the QALY shortfall calculator[53] to calculate the expected total QALYs for the general population, using the baseline characteristics of the Effisayil 1 trial (age: 43, 68% female). This results in a remaining discounted QALYs of 17.91.

UK evidence suggests that patients with GPP experience between 0.9 to 1.79 flares per year. However, the company concluded that a patient would have to experience more than 1.79 flares per year over 100 years to be eligible for a severity modifier. Therefore, a severity modifier was not applied for this appraisal. The EAG agrees with the company's conclusion.

8 REFERENCES

1. Puig L, Choon SE, Gottlieb AB, et al. Generalized pustular psoriasis: A global Delphi consensus on clinical course, diagnosis, treatment goals and disease management. *Journal of the European Academy of Dermatology and Venereology* 2023;37(4):737-52.
2. Rivera-Díaz R, Daudén E, Carrascosa JM, et al. Generalized Pustular Psoriasis: A Review on Clinical Characteristics, Diagnosis, and Treatment. *Dermatol Ther (Heidelb)* 2023;13(3):673-88.
3. Navarini AA, Burden AD, Capon F, et al. European consensus statement on phenotypes of pustular psoriasis. *Journal of the European Academy of Dermatology and Venereology* 2017;31(11):1792-99.
4. Bachelez H, Barker J, Burden AD, et al. Generalized pustular psoriasis is a disease distinct from psoriasis vulgaris: evidence and expert opinion. *Expert Rev Clin Immunol* 2022;18(10):1033-47.
5. Hoegler KM, John AM, Handler MZ, Schwartz RA. Generalized pustular psoriasis: a review and update on treatment. *Journal of the European Academy of Dermatology and Venereology* 2018;32(10):1645-51.
6. Psoriasis Association. *Pustular psoriasis*. https://www.psoriasis-association.org.uk/pustular-psoriasis#gpp (accessed 08/10/2024).
7. Zhou L, Todorovic V. Interleukin-36: Structure, Signaling and Function. In: Atassi MZ (ed.) *Protein Reviews : Volume 21*. Cham: Springer International Publishing; 2021 p191-210.
8. Frysz M, Patel S, Li MOY, et al. Prevalence, incidence, mortality, and healthcare resource use for generalised pustular psoriasis, palmoplantar pustulosis, and plaque psoriasis in England: a population-based cohort study. *British Journal of Dermatology* 2024.
9. Kharawala S, Golembesky AK, Bohn RL, Esser D. The clinical, humanistic, and economic burden of generalized pustular psoriasis: a structured review. *Expert Rev Clin Immunol* 2020;16(3):239-52.
10. Lebwohl M, Medeiros RA, Mackey RH, et al. The Disease Burden of Generalized Pustular Psoriasis: Real-World Evidence From CorEvitas' Psoriasis Registry. *Journal of Psoriasis and Psoriatic Arthritis* 2022;7(2):71-78.
11. Barker JN, Becher G, Burden DA, et al. Clinical course, treatment and management of generalised pustular psoriasis from a United Kingdom extension of a global Delphi panel. *Skin Health and Disease* 2024;4(3):e359.
12. Boehringer Ingelheim International GmbH. Spevigo 450 mg concentrate for solution for infusion (SPC). *Summary of Product Characteristics*, 2023.
13. National Institute for Health and Care Excellence (NICE). Psoriasis: assessment and management. Clinical Guideline [CG153], 2012; last updated September 2017.
14. Boehringer Ingelheim. Structured expert elicitation in general pustular psoriasis report, 2022.
15. Morita A, Choon SE, Bachelez H, et al. Design of Effisayil™ 2: A Randomized, Double-Blind, Placebo-Controlled Study of Spesolimab in Preventing Flares in Patients with Generalized Pustular Psoriasis. *Dermatol Ther (Heidelb)* 2023;13(1):347-59.
16. Bachelez H, Choon SE, Marrakchi S, et al. Trial of Spesolimab for Generalized Pustular Psoriasis. *New England Journal of Medicine* 2021;385(26):2431-40.
17. Boehringer Ingelheim (ed.) Effisayil™ 1: Multi-center, double-blind, randomized, placebo controlled, Phase II study to evaluate efficacy, safety and tolerability of a single intravenous dose of BI 655130 in patients with Generalized Pustular Psoriasis (GPP) presenting with an acute flare of moderate to severe intensity; 2021.
18. Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): Why do both? A systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. *Journal of the American Academy of Dermatology* 2012;66(3):369-75.

19. Bachelez H, Choon S-E, Marrakchi S, et al. Inhibition of the interleukin-36 pathway for the treatment of generalized pustular psoriasis. *New England Journal of Medicine* 2019;380(10):981-83.
20. Burden AD, Bachelez H, Choon SE, et al. The Generalized Pustular Psoriasis Physician Global Assessment (GPPGA) score: online assessment and validation study of a specific measure of GPP disease activity. *British Journal of Dermatology* 2023;189(1):138-40.
21. Navarini AA, Bachelez H, Choon SE, et al. 43008 Effisayil ON, an open-label, long-term extension study of spesolimab treatment in patients with generalized pustular psoriasis: interim results for flare treatment. *Journal of the American Academy of Dermatology* 2023;89(3):AB44.
22. Burden AD, Okubo Y, Zheng M, et al. Efficacy of spesolimab for the treatment of generalized pustular psoriasis flares across pre-specified patient subgroups in the Effisayil 1 study. *Experimental Dermatology* 2023;32(8):1279-83.
23. Choon SE, Lebwohl MG, Turki H, et al. Clinical Characteristics and Outcomes of Generalized Pustular Psoriasis Flares. *Dermatology* 2023;239(3):345-54.
24. Boehringer Ingelheim (ed.) Observational and Non-Interventional Study (ONIS) Report. Data on File; 2022.
25. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
26. Wolf P, Ceovic R, Conrad C, et al. Characteristics and management of generalized pustular psoriasis (GPP): Experience from the Central and Eastern Europe (CEE) GPP Expert Network. *Journal of the European Academy of Dermatology and Venereology* 2024.
27. Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. 2021;25:37.
28. Boehringer Ingelheim. CEM validation meeting of SPEVIGO in GPP flares 2024.
29. Bachelez H, Massol J, Pouvourville G, et al. Characterization of flares in patients with generalized pustular psoriasis – a population-based study from the French National Health Data System database (SNDS) *American Academy of Dermatology Virtual Meeting Experience*; 2021.
30. Augey F, Renaudier P, Nicolas J-F. Generalized pustular psoriasis (Zumbusch): a French epidemiological survey. *European Journal of Dermatology* 2006;16(6):669-73.
31. Kromer C, Loewe E, Schaarschmidt ML, et al. Drug survival in the treatment of generalized pustular psoriasis: a retrospective multicenter study. *Dermatologic Therapy* 2021;34(2):e14814.
32. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev* 2011;2011(2):CD008794.
33. National Institute for Health and Care Excellence. Infliximab for the treatment of adults with psoriasis [TA134], 2008.
34. National Institute for Health and Care Excellence. Infliximab and adalimumab for the treatment of Crohn's disease [TA187], 2010.
35. National Institute for Health and Care Excellence. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy [TA329], 2015.
36. National Institute for Health and Care Excellence. Secukinumab for treating moderate to severe plaque psoriasis [TA350], 2015.
37. National Institute for Health and Care Excellence. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed [TA375], 2016.
38. National Institute for Health and Care Excellence. TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis [TA383], 2016.
39. National Institute for Health and Care Excellence. Ixekizumab for treating moderate to severe plaque psoriasis [TA442], 2017.

40. National Institute for Health and Care Excellence. Ixekizumab for treating active psoriatic arthritis after inadequate response to DMARDs [TA537], 2018.
41. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study. *Pharmacoeconomics* 2023;41(2):199-207.
42. McNamara S, Schneider PP, Love-Koh J, et al. Quality-adjusted life expectancy norms for the English population. *Value in Health* 2023;26(2):163-69.
43. National Institute for Health and Care Excellence. Casirivimab plus imdevimab, nirmatrelvir plus ritonavir, sotrovimab and tocilizumab for treating COVID [TA878], 2023.
44. Stevenson M, Archer R, Tosh J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technology Assessment* 2016;20(35):1-610.
45. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Medical Decision Making* 2011;31(6):800-04.
46. Viguier M, Bentayeb M, Azzi J, et al. Generalized pustular psoriasis: A nationwide population-based study using the National Health Data System in France. *Journal of the European Academy of Dermatology and Venereology* 2024;38(6):1131-39.
47. British National Formulary (BNF). https://bnf.nice.org.uk/ (accessed June 2024).
48. Department of Health and Social Care. *Drugs and pharmaceutical electronic market information tool (eMIT)*. https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit (accessed 13 July 2024).
49. Jones KWH, Birch, S., Castelli, A., Chalkley, M., Dargan, A., Findlay, D., Forder, J., Gao, M., Hinde, S., Markham, S. Premji, S. Teo, H.,. Unit Costs of Health and Social Care 2023 Manual.
Technical report. In: Personal Social Services Research Unit, editor. Kent, UK: (University of Kent), 2024.
50. National Health Service England. *NHS trust and NHS foundation trusts. National Cost Collection for the NHS: 2020/21 data* https://www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/ (accessed March 2024).
51. Boehringer Ingelheim. Structured expert elicitation in general pustular psoriasis report: Hospitalisation and length of hospital stay under best available care, 2024.
52. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. *London: Nuffield Trust Research Report* 2014.
53. Shchneider P, McNamara S, Love-Koh J, et al. *QALY Shortfall Calculator*. https://shiny.york.ac.uk/shortfall.
54. Choon SE, Lebwohl MG, Marrakchi S, et al. Study protocol of the global Effisayil 1 Phase II, multicentre, randomised, double-blind, placebo-controlled trial of spesolimab in patients with generalized pustular psoriasis presenting with an acute flare. *BMJ Open* 2021;11(3):e043666.
55. Burden AD, Choon SE, Gottlieb AB, et al. Clinical Disease Measures in Generalized Pustular Psoriasis. *Am J Clin Dermatol* 2022;23(1):39-50.
56. Kersten P, White PJ, Tennant A. Is the Pain Visual Analogue Scale Linear and Responsive to Change? An Exploration Using Rasch Analysis. *PLOS ONE* 2014;9(6):e99485.
57. Rentz AM, Skalicky AM, Burslem K, et al. The content validity of the PSS in patients with plaque psoriasis. *Journal of Patient-Reported Outcomes* 2017;1(1):4.
58. Madrigal-Cadavid J, Estrada-Acevedo J, Maria Jaramillo A, et al. Rasch analysis of the dermatology life quality index (DLQI) in patients with mild to moderate-severe psoriasis. *Indian Journal of Dermatology, Venereology and Leprology*;90.
59. Yang Z, Li S, Wang X, Chen G. Health state utility values derived from EQ-5D in psoriatic patients: a systematic review and meta-analysis. *Journal of Dermatological Treatment* 2022;33(2):1029-36.

60. Swinburn P, Lloyd A, Boye KS, et al. Development of a Disease-Specific Version of the EQ-5D-5L for Use in Patients Suffering from Psoriasis: Lessons Learned from a Feasibility Study in the UK. *Value in Health* 2013;16(8):1156-62.
61. Morita A, Strober B, Burden AD, et al. Efficacy and safety of subcutaneous spesolimab for the prevention of generalised pustular psoriasis flares (Effisayil 2): an international, multicentre, randomised, placebo-controlled trial. *Lancet* 2023;402(10412):1541-51.
62. Warren RB, Reich A, Kaszuba A, et al. Imsidolimab, an anti-interleukin-36 receptor monoclonal antibody, for the treatment of generalized pustular psoriasis: results from the phase II GALLOP trial. *British Journal of Dermatology* 2023;189(2):161-69.
63. Yang BCJ, Zhang Z, Schwabe C, et al. HB0034, a novel anti-IL-36R inhibitor, showed a promising efficacy in controlling acute GPP flare 2023 [03 June, 2024] *32nd EADV Congress 2023, Berlin*: European Academy of Dermatology and Venereology; 2024.
64. Tseng CY, Lu CW, Chung WH, Lin FJ. EE412 Unraveling the Disease Burden of Generalized Pustular Psoriasis: A Comparative Analysis of Flares with Psoriasis Vulgaris in Taiwan. *Value in Health* 2023;26(12):S130.
65. Hu K, Liu Y, Liu Y, et al. Rapid and sustained resolution in generalized pustular psoriasis with IL-17A inhibitors required high adherence: a 96-week analysis in a real-life setting. *International Journal of Dermatology* 2024.
66. Zema CL, Valdecantos WC, Weiss J, et al. Understanding Flares in Patients With Generalized Pustular Psoriasis Documented in US Electronic Health Records. *JAMA Dermatol* 2022;158(10):1142-48.
67. Reisner DV, Johnsson FD, Kotowsky N, et al. Impact of Generalized Pustular Psoriasis from the Perspective of People Living with the Condition: Results of an Online Survey. *Am J Clin Dermatol* 2022;23(Suppl 1):65-71.
68. Milan R, Read S, Couture-Lapointe C, et al. EPH32 Canadian Pustular Psoriasis Study (CAPPS): Examining Disease Burden, Treatments, and Healthcare Resource Utilization for Generalized Pustular Psoriasis Flares. *Value in Health* 2023;26(12):S208.
69. Bellinato F, Gisondi P, Marzano AV, et al. Characteristics of patients experiencing a flare of generalized pustular psoriasis: a multicenter observational study. *Vaccines* 2023;11(4):740.
70. Boehringer Ingelheim. Patient pathway analysis in Generalised Pustular Psoriasis (GPP), 2022.

9 APPENDICES

Appendix 1 EAG summary appraisal of the company's systematic review methods

Table 31 EAG appraisal of systematic review methods

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
Was the review question clearly defined using the PICOD framework or an alternative?	Yes	The SLR sought to identify all relevant evidence of clinical efficacy and safety associated with spesolimab and current available treatment options for GPP (CS Appendix D.1.1) and the eligibility criteria are defined according to a PICOS framework in CS Appendix Table 10.
Were appropriate sources of literature searched?	Yes	Core healthcare databases, MEDLINE, Embase, CDSR and CENTRAL were searched, alongside hand searches of several dermatology and psoriasis conferences and ISPOR, three clinical trials platforms and reference lists of relevant systematic reviews and meta-analyses.
What time period did the searches span and was this appropriate?	Yes	The time period was database inception to 14th December 2022, with an update search on 6th May 2024. Conference abstracts were restricted to the preceding two years, i.e. since 2022. The searches are recent and the EAG has not carried out any updates.
Were appropriate search terms used and combined correctly?	Yes	Search strings included both relevant subject headings and appropriate free text terms. Relevant published search filters were used.
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes, (clarity issue resolved in clarification response A4)	A broad set of eligibility criteria is reported in CS Appendix Table 10, however, eligibility criteria to align with the decision problem were not clear (whether data specific to GPP flares was relevant to population, or outcomes, or both – resolved in Clarification Response A4) and appeared to have been conflated with a

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
		feasibility assessment for the indirect treatment comparison that included evidence not eligible for this appraisal. Eligibility criteria for the evidence included in this appraisal are not clearly distinguished from the eligibility criteria and feasibility assessment aligned with the broader scope of the SLR (see Clarification Response A5).
Were study selection criteria applied by two or more reviewers independently?	Yes	Two independent researchers performed title and abstract screening and full-text screening; a third independent researcher was invited to resolve any discrepancies (CS Appendix D.1.1.4).
Was data extraction performed by two or more reviewers independently?	No, but methods are adequate	One researcher performed data extraction with a second researcher checking to ensure accuracy, with any discrepancies resolved by mutual discussion (CS Appendix D.1.1.5).
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes	RCTs were assessed using the quality assessment checklist adapted from CRD, University of York guidance (CS Appendix D. 1.3.1). Non-randomised and observational studies were assessed using the Downs and Black checklist (CS Appendix D.1.3.2).
Was risk of bias assessment (or other study quality assessment) conducted by two or more reviewers independently?	Yes	Each reviewer performed the quality assessment independently (CS Appendix D.1.3.1).
Is sufficient detail on the individual studies presented?	Yes	Effisayil 1 – yes, the entire CSR and all trial publications Effisayil 1 historical cohort – study publication (no company study report)

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
		CEE cohort study – study publication and extracted data alongside other studies in Appendix M. SEE study – yes, technical report with methods and results supplied with CS (copies of training and preparation material not included), summary in CS Appendix M.1.1.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Not applicable.	Statistical analysis of treatment comparisons was not carried out. The company used historical chart review of participants in their pivotal RCT, supplemented with data from a case series study of patients with GPP in CEE countries and a company SEE study to provide a descriptive comparison. This is discussed in section 3.3.

Abbreviations: CDSR, Cochrane Database of Systematic Reviews; CEE, Central Eastern European; CENTRAL, Cochrane Central Register of Controlled Trials; CRD, Centre for Reviews and Dissemination; GPP, generalised pustular psoriasis; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; ITC, indirect treatment comparison; MEDLINE, Medical Literature Analysis and Retrieval System Online; NMA, network meta-analysis; PICOD/PICOS, Population Intervention Comparator Outcome Study Design framework; RCT, randomised controlled trial; SEE, structured expert elicitation.

Appendix 2 Company and EAG risk of bias assessment for Effisayil 1

Table 32 Risk of bias assessment for Effisayil 1

Question	Company conclusion	Company comments	EAG assessment
Was randomisation carried out appropriately?	Yes	Patients were randomised to receive spesolimab or placebo in a ratio of 2:1 with stratification for Japanese vs non-Japanese ethnic group	Low risk of bias. An interactive response technology system was used to implement randomisation. Participants were stratified according to ethnicity (Japanese/not Japanese).
Was concealment of treatment allocation adequate?	Yes	Interactive response technology was used to perform treatment allocation and manage ordering of drug supplies. Clinicians were provided with a solution for infusion which could be spesolimab or a placebo solution.	Agree. Low risk of bias. The study drug was allocated using computerised interactive response technology (Study protocol, Choon et al. 2021). [54]
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	There were some differences observed in some demographic and clinical characteristics at baseline, namely gender, race, GPPASI score, and GPPGA pustulation subscore. Clinical characteristics representative of more severe disease were observed in the spesolimab arm so any resulting bias would be in	Not enough information. Unclear risk of bias. Without clinical expertise we have been unable to determine which characteristics are prognostic factors for flares (sections 2.2.1 and 3.2.1.2). We agree that there were some differences in demographic and clinical characteristics at baseline, for example, disease severity according to GPPASI and GPPGA was slightly greater in the spesolimab arm,

Question	Company conclusion	Company comments	EAG assessment
		favour of the placebo arm.	systemic symptoms, e.g. fever (Suppl. Table 1) was higher in the spesolimab arm, but having the IL36RN mutation was similar between arms. We observe a slight bias in favour of the placebo arm, however, disease severity is not a confirmed prognostic factor.
Were the care providers, participants and outcome assessors blinded to treatment allocation?	Yes	Patients and clinicians were blinded to randomised treatment allocation until after the database lock for final analysis.	Agree. Low risk of bias. [redacted]
Were outcome assessors blind to treatment allocation?	Unclear (CS Appendix Table 12)	[CS Appendix Table 12 does not include supporting statements]	Agree. Unclear risk of bias. It is not clear to the EAG whether the outcome assessors were the same people as the care providers, if so then they probably were blinded, but we cannot be sure if they were the same people.
Were there any unexpected imbalances in dropouts between groups?	No	91% of patients to the spesolimab arm and 94% of patients randomised to the placebo arm completed the 12-week trial period.	Agree. Low risk of bias. The patient flow diagram in CS Figure 6 shows that early discontinuations were few and roughly proportional to the 2:1 randomisation. During the follow-up period

Question	Company conclusion	Company comments	EAG assessment
			proportionally more participants in the placebo group received open label spesolimab or rescue treatment with spesolimab, but this would be expected.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	[No company justification made]	Unclear risk of bias. The published study protocol, Choon et al. 2021,[54] and CS Appendix N, Table 62 have GPPASI-related secondary outcomes but the results for the GPPASI outcomes are not reported in the CS, and limited results in Supplementary Figure 6 of the trial publication (Bachelez et al. 2021)[16]. All other planned outcomes are reported.
Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Primary endpoint and key secondary endpoint analysis at Week 1 was based on an ITT analysis with patients withdrawing or receiving escape medication assumed not to have responded. Due to high levels of open-label spesolimab use in the placebo arm after Week 1, changes at Week 4 for secondary endpoints and further	Agree. Low risk of bias. An ITT analysis was carried out. Methods to account for missing data were to impute non-response for the binary outcomes and to use the last observation carried forward for the continuous outcomes. These methods appear appropriate, but the amount of missing data is unclear. See also section 3.2.4 for EAG discussion of statistical

Question	Company conclusion	Company comments	EAG assessment
		endpoints were reported descriptively. The ITT principle was still adopted with appropriate methods used to account for missing data.	methods around missing data.
Does the study reflect routine clinical practice in England?	Yes	The study population, intervention, comparator and outcomes are applicable to routine clinical practice as discussed in Section B.2.12.3, despite not being directly reflective due to a lack of active treatment for the control arm and a high Asian cohort in the study population.	This question, not in the original CRD checklist, is about generalisability rather than risk of bias. No EAG risk of bias statement made; generalisability of the trial to the UK/England is discussed in sections 3.2.1.1 and 3.2.1.2.

Source: Reproduced from CS Table 11 and CS Appendix Table 12; with EAG additional comments.
Abbreviations: CRD, Centre for Reviews and Dissemination; CSR, clinical study report; GPP, generalised pustular psoriasis; GPPASI, Psoriasis Area and Severity Index for Generalized Pustular Psoriasis; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; IL36RN, interleukin-36 receptor antagonist protein; ITT, intention to treat.

Appendix 3 Outcomes assessment: Pain VAS, FACIT-Fatigue, DLQI, and PSS

The following patient reported outcome measures (PROs) used in Effisayil 1 are not specific to GPP. Two are specific to psoriasis (Psoriasis Symptom Scale) or skin conditions (Dermatological Quality of Life Index) and the others are generic PROs (pain VAS, Functional Assessment of Chronic Illness Therapy – Fatigue scale, and EQ-5D). The EAG has found no scientific literature reporting validation of these outcome measures or consensus on minimum clinically important difference (MCID) in GPP. A recent paper states the need for validation and international consensus for the consistent use of these measures (and efficacy measures) in GPP.[55] The EAG agree this would be useful as we have been unable to verify the use of these measures in clinical practice and what might constitute a valid response due to lack of clinical expertise available to us. All PROs, except EQ-5D, were secondary outcomes, measuring change from baseline at Week 4, that were included in the planned hierarchy for statistical testing. Measurements at other timepoints, and for EQ-5D, were exploratory outcomes. EQ-5D is the only PRO that informs the economic model.

Table 33 Summary of PRO outcomes assessed in Effisayil 1

Outcome	Description / Validation / MCID
Pain VAS	• Generic scale to record current pain intensity by marking a point on a single line 100mm in length representing 0 (no pain) to 100 (very severe pain) • Score can be influenced by other comorbid conditions,[55] and another study warns it does not behave linearly with raw change scores either under- or over-estimating true change. [56] • Wide range of MCIDs reported (9 to 30 mm) but none for GPP[56] • Effisayil 1 used an MCID of a 30-point decrease from baseline (CS Appendix N Table 63) which is at the more conservative end of that range requiring a larger decrease in pain to show meaningful response.
FACIT-Fatigue	• 13-item questionnaire to assess fatigue and its impact on daily activities and functioning over a 1-week recall period. • Each question is scored 0 to 4 total score 0 to 52, higher scores indicate a lower impact on fatigue/function. • Validated in several chronic diseases, but not for GPP.[55] • Effisayil 1 used an MCID of a 4-point increase from baseline (CS Appendix N Table 63)

Outcome	Description / Validation / MCID
PSS	• 4-item scale (symptoms: pain, redness, itching, and burning) to assess skin condition over a 1-day recall period. • Each symptom scored 0 to 4, total score 0 to 16, higher scores indicate increased severity. • Validated in plaque psoriasis patients,[57] but not for GPP.[55] • Effisayil 1 used an MCID of a 2-point decrease in score from baseline (CS Appendix N Table 63)
DLQI	• 10-item questionnaire to assess symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment, over a 1-week recall period. • Each question is scored 0 to 3, total score 0 to 30, higher scores indicate worse quality of life. • Validated for use in psoriasis generally,[58] but not for GPP specifically.[55] It does not assess any of the systemic symptoms associated with GPP. • Effisayil 1 used an MCID of a 4-point decrease from baseline (CS Appendix N Table 63)
EQ-5D-5L	• A 5-item questionnaire to assess current mobility, self-care, usual activities, pain/discomfort, and anxiety/depression; and a visual analogue scale (EQ-VAS, a vertical scale for patients to provide a global assessment of their health that takes values between 100 (best imaginable health) and 0 (worst imaginable health)). • Each question is scored 1 to 5, a summary score is calculated from a regional value set, higher scores indicate worse quality of life. • Evaluated for psoriasis generally;[59] bolt-ons for skin irritability and self-confidence contribute to improved content validity for patients with psoriasis;[60] but it still does not assess pustular or systemic symptoms specific to GPP and the bolt-ons were not used in Effisayil 1. • Mapped to EQ-5D-3L using the NICE preferred mapping function (CS section B.3.4.2)

Sources: CS Table 8, CS Appendix N Table 63,
Abbreviations: DLQI, Dermatology Life Quality Index; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy – Fatigue scale; GPP, generalised pustular psoriasis; MCID, minimum clinically important difference; PSS, Psoriasis Symptoms Scale; VAS, visual analogue scale.

Appendix 4 Mean composite GPPGA score in relation to the total GPPGA score

Table 34 Components of the mean composite GPPGA score (erythema, pustules and scaling) and how this relates to total GPPGA score

Erythema Score component	Pustules score component					Scaling Score component
	0	1	2	3	4	
0	0.00	0.33	0.67	1.00	1.33	0
0	0.33	0.67	1.00	1.33	1.67	1
0	0.67	1.00	1.33	1.67	2.00	2
0	1.00	1.33	1.67	2.00	2.33	3
0	1.33	1.67	2.00	2.33	2.67	4
1	0.33	0.67	1.00	1.33	1.67	0
1	0.67	1.00	1.33	1.67	2.00	1
1	1.00	1.33	1.67	2.00	2.33	2
1	1.33	1.67	2.00	2.33	2.67	3
1	1.67	2.00	2.33	2.67	3.00	4
2	0.67	1.00	1.33	1.67	2.00	0
2	1.00	1.33	1.67	2.00	2.33	1
2	1.33	1.67	2.00	2.33	2.67	2
2	1.67	2.00	2.33	2.67	3.00	3
2	2.00	2.33	2.67	3.00	3.33	4
3	1.00	1.33	1.67	2.00	2.33	0
3	1.33	1.67	2.00	2.33	2.67	1
3	1.67	2.00	2.33	2.67	3.00	2
3	2.00	2.33	2.67	3.00	3.33	3
3	2.33	2.67	3.00	3.33	3.67	4
4	1.33	1.67	2.00	2.33	2.67	0
4	1.67	2.00	2.33	2.67	3.00	1
4	2.00	2.33	2.67	3.00	3.33	2
4	2.33	2.67	3.00	3.33	3.67	3
4	2.67[a]	3.00	3.33	3.67	4.00	4
	Cells above show mean composite scores					

Source: EAG table

KEY:

Mean composite score	0	> 0 to < 1.5	≥ 1.5 to 2.5	≥ 2.5 to 3.5	≥ 3.5
Total GPPGA score	0	1	2	3	4
Description	Clear	Almost clear	Mild	Moderate	Severe

GPPGA, Generalised Pustular Psoriasis Physician Global Assessment

[a] A patient with a GPPGA pustulation subscore of 0 could still be assessed as having moderate GPP, if their erythema and scaling scores were both 4 because the mean composite score would be 2.67.

Appendix 5 Location and sources of evidence identified by the company for their indirect treatment comparison feasibility assessment

Appendix D.1.5 (Relevance of evidence base and heterogeneity assessment findings) provides some information on the company's systematic review which would have been expected to identify comparator data. The company's systematic review identified 79 unique studies from which the company identified a sub-set of 15 studies that they stated were relevant to their decision problem because they reported data specific to GPP flares (these 15 studies are listed in CS Appendix D Table 11).

It was not clear to us whether the feasibility assessment for indirect treatment comparison was performed on the broader set of 79 studies identified by the company's systematic literature review described in CS Appendix D or the sub-set of 15 studies the company stated were relevant to the decision problem. In response to clarification question A5 the company confirmed all studies identified in the clinical systematic literature review were assessed.

The evidence identified by the company's systematic review that the company stated was relevant to the decision problem (i.e. the 15 studies listed in CS Appendix D Table 11), comprised:

- Four studies related to the Effisayil trial program[16,19,21,61]
- Two small (10 patients or fewer) prospective studies of currently unlicensed products for the treatment of GPP flare (imsidolimab[62] and HB0034[63])
- Nine retrospective studies:
  - One was conducted in Taiwan[64]
  - One was conducted in China[65]
  - Three were conducted in North America (two in the USA[66,67] and one in Canada[68]
  - Three were conducted in Europe (France[46], Italy[69] and Central and Eastern Europe[26]
  - One was a retrospective analysis of the Effisayil 1 trial cohort. This is described as the Effisayil 1 historical cohort[23]

Although not explicitly stated, it seems that the company have assessed the 11 studies that were not part of the Effisayil trial program for their suitability to provide comparator data in an indirect treatment comparison. From the information provided in CS Appendix D.1.5 it appears that the two prospective studies on unlicensed products were excluded as they

were not part of the clinical pathway of care in the UK, the studies conducted in China, Taiwan and North America were excluded on the grounds that they lacked generalisability to the UK and the remaining four studies (the three conducted in Europe[26,46,69]) and the Effisayil 1 historical cohort[23] were reviewed to discern whether they could inform the efficacy and safety of BAC in formal evidence synthesis. The company presents the data from these three European studies in CS Appendix M (the Effisayil 1 historical cohort is not included in Appendix M). In addition to the three European studies identified by the company's systematic review, CS Appendix M Table 57 provides data for another three real-world evidence sources and two Structured Expert Elicitation sources. The three real-world evidence sources are:

- HES (Health Episode Statistics data[70]);
- POLARIS[8] and
- SCRIPTOR[24]

It is not clear to us how the POLARIS study was identified and included as it does not appear within the clinical SLR tables (CS Appendix D.1.2).

The two Structured Expert Elicitation (SEE) sources are:

- SEE – Best Available Care (BAC) efficacy and safety[14] and
- SEE – Health Care Resource Use (HCRU)[51].

Appendix 6 EAG review of studies identified in the CS to ascertain if any could contribute data in an indirect treatment comparison

Table 35 EAG review of studies identified in the CS

Study ID & design	GPP Population	Intervention	EAG comments
Tseng et al. 2023[64] Retrospective cohort	243 patients in Taiwan with data on 1,151 flares between January 2017 and September 2020.	Systemic treatment (not specified)	No treatment or efficacy data reported.
Hu et al. 2024[65] Retrospecti	18 patients in China with flares between June 2019 and October 2023.	Secukinumab (n=13) or ixekizumab (n=5)	Reports response at 4 weeks over time to 96 weeks.

Study ID & design	GPP Population	Intervention	EAG comments
ve case series			
Zema et al. 2022[66] Retrospective cohort	271 patients in the US with data on 513 flares between July 2015 and June 2020.	Topical corticosteroids (35% of flares), opioids (21%), other oral treatments[a] (13%), oral corticosteroids (11%)	No efficacy data reported.
SNDS Viguier et al.	569 hospitalized patients in France with flares between 2012 and 2015.	Not reported.	Reports duration of hospital stay.

Study ID & design	GPP Population	Intervention	EAG comments
2024[46] Retrospective cohort			
CAPPS Milan et al. 2023[68] Retrospective cohort	15 patients in Canada 66.7% of whom had a least one flare between January 2011 and December 2020.	Topical corticosteroids (73.3%), oral corticosteroids (46.7%), biologics (13.3%)	No detailed efficacy data reported.

Study ID & design	GPP Population	Intervention	EAG comments
Effisayil™ 1 historical cohort Choon et al. 2023[23] Retrospective cohort	53 patients, 29 patients with data for flares per year (time period of study not defined).	For 46 patients with at least one historical medication for past flare. At least one biologic (24.5%)[b], acitretin (45.3%), methotrexate (43.4%), cyclosporine (30.2%), Other systemic and topical therapies[c].	Reports typical time to pustular clearance for each flare category but flare resolution not linked to a set of particular treatments for a category of flare.

Study ID & design	GPP Population	Intervention	EAG comments
Bellinato et al. 2023[69] Retrospective cohort	66 patients in Italy with a history of flares between 2018 and 2022.	Reported as proportion of patients: corticosteroids alone (69.7%) or in combination with systemic retinoids (36.4%), biologics[d], other treatments[e]	No efficacy data reported.
Reisner et al. 2022[67] Online surve	66 patients in the US, all had experienced at least one flare in the past 12 months (survey took place 4-14th August 2020).	Treatments added during a flare (n=21 patients): topicals (38%); oral corticosteroids (38%), biologics (IL-12/23 inhibitors 14%, IL-17 inhibitors 5%), other (14%).	No efficacy data reported.

Study ID & design	GPP Population	Intervention	EAG comments
y of people living with GPP			
CEE GPP Expert Network Wolf et al. 2024[26] Retrospecti	58 patients in Central and Eastern Europe. Captured characteristics of most recent flare and most severe flare within the past 10 years up to December 2022.	For historical most severe flare (n=26); retinoids (53.8%), methotrexate (23.1%), systemic steroids (23.1%), PUVA (19.2%), biologics (15.4%), anti-TNF-α (7.7%), Anti-IL-12/23 (3.8%), Anti-IL-36R (7.7%), UVB (15.4%).[f]	Reports time from initiation of treatment to flare resolution for subgroups by flare type.

Study ID & design	GPP Population	Intervention	EAG comments
ve case series			
HES[70] Retrospective cohort	[redacted] patients in England with a GPP diagnosis ([redacted])	Not reported	Relevant population but no data on treatment received for flares or its efficacy.
POLARIS[8] Retrospective	373 patients in England with a GPP diagnosis (registered in CPRD Aurum database between 1 January 2008 and 31 December 2019)	Not reported but likely to have been best available care in England	Relevant population and a subgroup analysis of patients with GPP flares but no data on treatment efficacy.

Study ID & design	GPP Population	Intervention	EAG comments
cohort			
SCRIPTOR[24] Retrospective chart review	27 patients [redacted] in the UK with a confirmed diagnosis of GPP [redacted]. A total of [redacted] flare episodes amongst [redacted] patients.	Reported for patients with at least one drug-based treatment for flare [redacted]	Relevant population and includes [redacted] flare episodes [redacted] does not provide what would be needed for the economic model.
SEE – BAC efficacy and	[redacted] clinical experts ([redacted])	Best available care [redacted]	Expert predictions for BAC efficacy and safety based on their experience of treating GPP patients at specialist centres.

Study ID & design	GPP Population	Intervention	EAG comments
safety[14] Expert opinion (note it is not clear if these experts were the same exper			

Study ID & design	GPP Population	Intervention	EAG comments
ts who contributed to the SEE – HCRU)			
SEE – HCRU[51] Expert opinion (note	[redacted]	[redacted]	Expert predictions for healthcare resource use.

Study ID & design	GPP Population	Intervention	EAG comments
it is not clear if these experts were the same experts who contributed to the SEE –			

Study ID & design	GPP Population	Intervention	EAG comments
BAC efficacy and safety).			

Source: EAG created table with information drawn from cited sources and CS Table 11
BAC, best available care; CAPPS, Canadian Pustular Psoriasis Study; CEE GPP, Central and Eastern Europe generalised pustular psoriasis; CPRD, Clinical Practice Research Datalink; GPP, generalised pustular psoriasis; GPPGA, generalised pustular psoriasis global assessment; HCRU, health care resource utilisation; HES, Hospital Episode Statistics; IL, interleukin; PUVA, psoralen plus ultraviolet A; SEE, structured expert elicitation; SNDS, Système National des Données de Santé; TNF, tumour necrosis factor; [redacted] US, United States; UVB, ultraviolet B

[a] Other oral treatments e.g. methotrexate, cyclosporine, tacrolimus
[b] Ustekinumab (13.2%), Adalimumab (9.4%), Infliximab (7.5%), Ixekizumab (5.7%) or Secukinumab (5.7%)
[c] **Systemic:** Acitretin (45.3%), Methotrexate (43.4%), Cyclosporine (30.2%), Cefuroxime axetil (11.3%), Prednisolone (11.3%), Etanercept (9.4%), Cetirizine hydrochloride (7.5%), Hydroxyzine hydrochloride (7.5%), Amoxicillin; clavulanic acid (5.7%), Ampicillin sodium; sulbactam sodium (5.7%), Azithromycin (5.7%), Cefuroxime (5.7%), Cetirizine (5.7%), Dexchlorpheniramine maleate (5.7%), Etretinate (5.7%), Isotretinoin (5.7%), Loratadine (5.7%). **Topical:** Clobetasol propionate (20.8%), Potassium permanganate (17.0%), Betamethasone (15.1%), Betamethasone; calcipotriol (15.1%), Betamethasone valerate (11.3%), Clobetasone butyrate (11.3%), Mometasone furoate (11.3%), Calcipotriol (9.4%), Camphor; coal tar; salicylic acid; sulfur (9.4%), Coal tar; Pinus spp. tar; salicylic acid (9.4%), Antipsoriatics for topical use (7.5%), Betamethasone valerate; fusidic acid (7.5%), Clobetasol (7.5%), Emulsifying wax; liquid paraffin; white soft paraffin (7.5%), Hydrocortisone (7.5%), Betamethasone; salicylic acid (5.7%), Coal tar; salicylic acid (5.7%)
[d] ixekizumab (22.7%), risankizumab (16.6%), secukinumab (19.7%), adalimumab (9.1%), guselkumab (3%), tildrakizumab (3%), and brodalumab (3%)
[e] non-biologic treatments (cyclosporine and methotrexate), NB-UVB and PUVA phototherapy.
[f] Also reported summary of treatments used for 'Last flare: other' and 'Last flare = most severe'.

Appendix 7 Company and EAG risk of bias assessment for the Effisayil 1 historical cohort study

Table 36 Summary of company and EAG risk of bias assessment of the Effisayil 1 historical cohort study[23] using the ROBINS-I tool

	Effisayil 1 historical cohort, Choon et al. 2023[23]	
	Company	**EAG**
Bias due to confounding	Moderate	Serious
	The study acknowledges various confounding factors such as stress, infections, and treatment withdrawal that can trigger GPP flares. However, no adjustment has been made; retrospective nature of data collection and lack of control for these confounders may introduce bias.	Confounding expected (pre-intervention prognostic factors may have influenced receipt of particular interventions for GPP flare) but prognostic factors for flare resolution not discussed. No control for potential confounders described.
Bias in selection of participants into the study	Low	Low
	53 participants were selected based on specific inclusion criteria related to their history of GPP and systemic inflammation, which were clearly defined. This reduces the risk of selection bias.	All the patients eligible for the Effisayil1 RCT were included in the Effisayil1 historical cohort.
Bias in classification of interventions	Low	Low
	The classification of interventions was based on historical medical data and standardized questionnaires, reducing the risk of misclassification.	Interventions received were identified from retrospective chart review. Presume charts would have been completed at the time of the intervention.
Bias due to deviations from the intended intervention	Moderate	Moderate
	The study relies on retrospective data, and deviations from intended interventions were not systematically recorded or controlled.	There was potential for deviations from the intended interventions.

	Effisayil 1 historical cohort, Choon et al. 2023[23]	
	Company	**EAG**
Bias due to missing data	Moderate	Moderate
	There were instances of missing data for several parameters. No analysis was done to assess the effect of missing data, which could affect the study's conclusions.	Patients with no historical flare could not be included. Patients with constant flare could not contribute flare frequency data. Not all patients contributed treatment data.
Bias in measurement of outcomes	Moderate	Serious
	Outcome measurements were based on patient recall, chart review and investigator interpretation, which could introduce measurement bias.	Outcomes were assessed for three categories of flare, but there was no standard definition of these three flare categories. Flare clinical course obtained from patients (potential for recall bias) and from chart review.
Bias in selection of the reported result	Low	Low
	The study appears to report relevant outcomes related to GPP flares comprehensively.	For the outcome of interest (duration of past flares) reporting appears comprehensive.
Overall risk of bias	Low	Serious
	The study provides valuable insights into GPP flares, but the retrospective design introduces some bias. The overall risk of bias is assessed as Low.	Study judged to be at a serious risk of bias in two domains, a moderate risk of bias in two domains, a low risk of bias in three domains.

Source: Part reproduction on company response to clarification question A23 Table 5 with EAG decisions and comments added.
GPP, generalised pustular psoriasis; RCT, randomised controlled trial.