Voclosporin with immunosuppressive therapies for treating lupus nephritis [ID3962]:

A Single Technology Appraisal

Produced by	Peninsula Technology Assessment Group (PenTAG) University of Exeter Medical School South Cloisters St Luke's Campus Heavitree Road Exeter EX1 2LU
Authors	**Simone Critchlow**[1] **Ash Bullement**[1] **Hollie Wheat**[1] **Sophie Robinson**[2] **Dominic Taylor**[3] **Preetham Boddana**[4] **Anthony Hatswell**[1] **G.J. Melendez-Torres**[2] **Caroline Farmer**[2] [1] Delta Hat Ltd, Nottingham UK [2] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter [3] North Bristol NHS Trust [4] Gloucestershire Hospitals NHS Foundation Trust
Correspondence to	Caroline Farmer 3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; C.Farmer@exeter.ac.uk
Date completed	21/07/2022
Source of funding	This report was commissioned by the NIHR Evidence Synthesis Programme as project number 135594.

Acknowledgments	The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG).
Declared competing interests of the authors	None
Rider on responsibility for document	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	Critchlow, Bullement, Wheat, Robinson, Taylor, Boddana, Hatswell, Melendez-Torres, Farmer. Voclosporin with immunosuppressive therapies for treating lupus nephritis [ID3962]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2022.

Author Contributions:

Simone Critchlow	Critical appraisal of the economic evidence and analysis submitted by the company, conducted additional economic analyses and drafted sections of the report
Ash Bullement	Critical appraisal of the economic evidence and analysis submitted by the company, conducted additional economic analyses and drafted sections of the report
Hollie Wheat	Critical appraisal of the economic evidence and analysis submitted by the company, conducted additional economic analyses and drafted sections of the report
Sophie Robinson	Critical appraisal of the literature search strategies
Dominic Taylor	Clinical advice and review of draft report
Preetham Boddana	Clinical advice and review of draft report
Anthony Hatswell	Critical appraisal of the economic evidence submitted by the company and drafted sections of the report
G.J. Melendez-Torres	Critical appraisal of the company submission, writing and editorial input
Caroline Farmer	Project lead, critical appraisal of the company submission, writing and editorial input

Table of Contents

List of key issues

List of tables

List of Figures

Abbreviations

ACR	American College of Rheumatology
AD	active disease
AE	adverse event
AZA	azathioprine
BID	twice daily
BT	background therapy
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CKD	chronic kidney disease
CNI	calcineurin inhibitor
CRD	Centre for Reviews and Dissemination
CRR	complete renal response
CQ	clarification question
CSR	clinical study report
CS	company submission
CYC	cyclophosphamide
DIC	deviance information criterion
DP	decision problem
DSA	deterministic sensitivity analysis
EAG	Evidence assessment group
eGFR	estimated glomerular filtration rate
EQ-5D	EuroQol five dimension
EMA	European medicines agency
ERA-EDTA	European Renal Association – European Dialysis and Transplantation Association
ESRD	end-stage renal disease
EULAR	The European Alliance of Associations for Rheumatology
GFR	glomerular filtration rate
HR	hazard ratio
HRQoL	health-related quality of life
HTA	health technology assessment
H-CYC	high dose cyclophosphamide
ICER	incremental cost-effectiveness ratio
ICER	Institute for Clinical and Economic Review

ACR	American College of Rheumatology
ICH	International conference on harmonisation
IQR	interquartile range
ITC	indirect treatment comparison
ITT	intention-to-treat
IV	intravenous
LN	lupus nephritis
LOCF	last observation carried forward
L-CYC	low dose cyclophosphamide
LY	life-year
MCAR	missing completely at random
MD	mean difference
mg	milligram
MHRA	Medicines and Healthcare products Regulatory Agency
MMF	mycophenolate mofetil
MPA	mycophenolic acid
NA	not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	network meta-analysis
NMB	net-monetary benefit
NR	not reported
OR	odds ratio
OWSA	one-way sensitivity analysis
PAS	patient access scheme
PBO	placebo
PRR	partial renal response
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	quality adjusted life year
RCT	randomised controlled trial
RDI	relative dose intensity
RR	relative risk
RTX	rituximab
SAE	serious adverse event

ACR	American College of Rheumatology
SD	standard deviation
SF-36	Short Form (36) health survey
SLE	systemic lupus erythematosus
SLR	systematic literature review
SmPC	Summary of product characteristics
TA	technology appraisal
TAC	tacrolimus
TEAE	treatment-emergent adverse event
TSD	Technical support document
TTD	time to treatment discontinuation
UK	United Kingdom
UPCR	Urine Protein Creatinine Ratio
USA	United States of America
VCS	voclosporin
Vs	versus
WTP	willingness to pay

1. EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the Evidence 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, technology and evidence and information on non-key issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of NICE.

1.1. Overview of the EAG's key issues

A brief overview of the key issues identified by the EAG in their appraisal of the company submission (CS) is provided in Table 1. Further detail of the issues is provided in Sections 1.3, 1.4, 1.5, and 1.6.

Broadly speaking, the key issues related to the company's cost effectiveness model, including limitations with the model structure, estimates of health-related quality of life, and assumptions related to long-term treatment effects. In addition, the EAG highlighted uncertainty in the way voclosporin would be used in practice, leading to uncertainty about the generalisability of the company's model and of clinical effectiveness estimates.

Table 1: Summary of key issues

ID	Summary of issues	Report sections
Key Issue 1	Network meta-analysis estimates may not be reliable and should better account for heterogeneity	3.3, 3.4
Key Issue 2	Model structure	4.2.2, 4.2.6
Key Issue 3	Long-term treatment effect	4.2.2, 4.2.6
Key Issue 4	Utility values	4.2.7
Key Issue 5	Estimation of treatment costs	4.2.6, 4.2.8
Key Issue 6	Transparency of reported model inputs	5.3

ID	Summary of issues	Report sections
Key Issue 7	Uncertainty in how voclosporin will be used in practice	2.4, 3.2.2, 3.2.3.2, 3.3.3, 4.2.3, 4.2.6.7

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are outlined in Table 2.

Table 2: Key differences between the company's preferred assumptions and EAG's preferred assumptions

	Company's preferred assumption	EAG preferred assumption	Report Sections
Utilities	Various see CS B.3.4 for specific details	Amendments to health state utility values for CKD stage 1-3a AD,PR,CR, CKD stage 5 dialysis and transplant	4.2.7
Costs	Various see CS B.3.5 for specific details	Updates to the treatment costs incorporated RDI and amending the MMF dose to 2g daily. Updates of cost inputs to align with sources. Updates to wastage assumptions	4.2.8
Transition probabilities from CKD 1-3a to CKD 3b-4	No movement from CKD 1-3a to CKD 3b-4 in the first 36 months	Movement from CKD 1-3a to CKD 3b-4 in the first 36 months	4.2.6.7
Long-term transition probabilities (36months+)	Application of 'treatment waning' using average of VCS+MMF transitions with MMF transitions applied to VCS+MMF arm	Application of average VCS+MMF and MMF transitions applied to both arms after 36 months	4.2.6.3
Risk on LN deaths in CKD stage 1-3a	Deaths observed within the AURORA 1 and 2 trial are assumed to inform LN related death in CKD stages 1-3a	Removal of 'LN related' deaths from the model transition probabilities from earlier CKD stages 1-3a assuming that death at this stage is non-disease specific and captured by general population mortality	4.2.6.6

Abbreviations: CKD, chronic kidney disease; CS, company submission; LN, lupus nephritis; MMF, mycophenolate mofetil; RDI, relative dosing intensity, VCS, voclosporin

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.

Overall, the technology is modelled to affect QALYs by:

- Increasing the rate of CRR
- Increasing the rate of PRR
- Reducing the risk of CKD progression

Overall, the technology is modelled to affect costs by:

- Drug acquisition costs for voclosporin
- Avoiding/delaying time to more expensive health states related to CKD (such as kidney transplant and dialysis associated with CKD stage 5)

The modelling assumptions that have the greatest effect on the ICER are:

- The application of LN related mortality within the company's model, which may overestimate the number of patients with LN who die as a result of disease (with subsequent impacts on the total costs and QALYs obtained)
- The long-term treatment effect assumptions applied to voclosporin+MMF and MMF. These are primarily:
  - The premise that voclosporin+MMF maintains some level of treatment effect relative to MMF for the entire duration of the modelled time horizon (72 years)
  - The assumption that transition probabilities from the within trial period will be maintained once all patients are removed from treatment at 36 months for the remainder of the model duration

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

The EAG did not identify a key issue solely related to the decision problem; however, in Key Issue 7 (Uncertainty in how voclosporin will be used in practice) the EAG highlights uncertainty

about how voclosporin will be used in practice, including where it will be used in the treatment pathway. This affects the most appropriate comparators for voclosporin.

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

Key Issue 1: Network meta-analysis estimates may not be reliable and should better account for heterogeneity

Report sections	Sections 3.3 and 3.4
Description of issue and why the EAG has identified it as important	Network meta-analyses (NMAs) drew on a heterogeneous evidence base including diverse outcome definitions, follow-up times and populations. However, the company chose to present fixed effects NMAs on the basis that random effects NMAs were judged as not converging. The EAG did not regard that the company had substantiated this claim.
What alternative approach has the EAG suggested?	The EAG suggested exploring informative priors for between-study variance parameters that are appropriate to this context in order to appropriately capture the heterogeneity in the evidence.
What is the expected effect on the cost-effectiveness estimates?	Expected cost-effectiveness estimates are not expected to change substantially, but uncertainty is more likely to be appropriately captured in probabilistic analyses.
What additional evidence or analyses might help to resolve this key issue?	NMAs that use appropriate informative priors, or otherwise clear evidence that no plausible random effects model would lead to convergent estimates in the base case.

Abbreviations: EAG, Evidence Assessment Group; NMA, network meta-analysis

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

Key Issue 2: The company's model structure is subject to a number of structural limitations

Report sections	Sections 4.2.2 and 4.2.6
Description of issue and why the EAG has identified it as important	The company's model is associated with a number of restrictive settings and assumptions which preclude in-depth investigation of the impacts these aspects of the model have on cost-effectiveness results. These features include: • CKD progression is only possible from an 'active disease' sub-state (and so patients with renal response are not subjected to a risk of CKD progression) • No CKD progression events in AURORA 1 or AURORA 2, and so CKD progression is disabled in the company's base-case analysis for the first 3 years, but this is not expected to align with clinical practice • Transitions in the first 3 years are based on the 'count method', which is subject to limitations mostly due to sample size • Very few within-trial deaths, and cause of death is not explicitly captured but is modelled to incur differential costs

Report sections	Sections 4.2.2 and 4.2.6
What alternative approach has the EAG suggested?	The EAG has explored a range of sensitivity analyses where possible within the confines of the company's model structure to investigate these aspects of the model further. These scenarios include permitting CKD progression from 0 years and removing within-trial deaths. However, some scenarios are not possible within the model structure (such as allowing CKD progression for patients with renal response, and re-analysing transition probabilities using a different approach other than the 'count method').
What is the expected effect on the cost-effectiveness estimates?	The scenarios that were possible to explore generally led to an increase in the ICER (further details presented in Section 6.2 of this report). When combined, these scenarios have the potential to lead to a much larger ICER compared with the company's base-case analysis. However, the impact of changing the model structure beyond edits possible for the EAG to make remains unclear.
What additional evidence or analyses might help to resolve this key issue?	Additional structural uncertainty analysis, considering sensitivity analysis allowing different transitions to occur and/or re-analysing the AURORA 1 and AURORA 2 trial data to obtain different transition probabilities may help resolve uncertainty associated with the model structure.

Abbreviations: CKD, chronic kidney disease; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio.

Key Issue 3: The long-term treatment effect of voclosporin + MMF and its comparators is unknown

Report sections	Sections 4.2.2 and 4.2.6
Description of issue and why the EAG has identified it as important	There is uncertainty in the long-term effect of VCS+MMF and how this compares to the long-term effect of MMF alone, as well as other comparators. The company's model requires extrapolation of transition matrices over a lifetime horizon (equivalent to 69 years beyond the initial 3 years of follow-up data available from the AURORA 1 and AURORA 2 studies). The company's application of independent transition matrices from the trial data makes two important assumptions: (1) that short-term data are sufficient to generalise to the longer term, and (2) that the short-term data while patients are on treatment are reflective of longer-term outcomes when patients are no longer receiving the same treatment up until 3 years. The company has assumed a 'waning' effect which takes the average effects across both arms and applied this to the VCS+MMF arm indefinitely. The EAG considered this approach to be inappropriate and unjustified in the absence of long-term data and clear justification within the CS.
What alternative approach has the EAG suggested?	The EAG has explored a range of alternative treatment waning effects, and ultimately prefers to assume the same conditional probabilities for renal response across both arms after 3 years.
What is the expected effect on the cost-effectiveness estimates?	The EAG's preferred approach causes the ICER to increase (further details presented in Section 6.2 of this report).

Report sections	Sections 4.2.2 and 4.2.6
What additional evidence or analyses might help to resolve this key issue?	The EAG feels there is no such evidence that would likely resolve the uncertainty associated with long-term treatment effects, other than longer-term follow-up data or clinical expert opinion.

Abbreviations: CKD, chronic kidney disease; CS, company submission; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; MMF, mycophenolate mofetil; VCS, voclosporin.

Key Issue 4: The utility estimates used in the company's model are inappropriate

Report sections	Section 4.2.7
Description of issue and why the EAG has identified it as important	The EAG has a number of reservations about the appropriateness of the utility values used to populate the model. These include a lack of appropriate analysis methods to derive utility values from the AURORA 1 and AURORA 2 studies, omission of a large quantity of data from AURORA 1 and AURORA 2 from the estimation of utility values, and use of literature-based utility values for later states that reflect a different group of patients.
What alternative approach has the EAG suggested?	Where possible, the EAG undertook sensitivity analyses using alternative utility values attempting to address some limitations of the company's analysis (e.g., using all values from AURORA 1 and AURORA 2, and not just values collected around the end of follow-up in AURORA 2).
What is the expected effect on the cost-effectiveness estimates?	The EAG's preferred utility values cause the ICER to increase slightly (further details presented in Section 6.2 of this report).
What additional evidence or analyses might help to resolve this key issue?	The EAG would prefer the company to re-analyse its utility data collected in AURORA 1 and AURORA 2 in line with standard convention, most likely adopting a regression analysis to explicitly incorporate multiple observations at the patient level.

Abbreviations: EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio.

Key Issue 5: The company has not appropriately calculated the costs of treatment in the model

Report sections	Sections 4.2.6 and 4.2.8
Description of issue and why the EAG has identified it as important	The company's model includes a number of assumptions made with respect to costing VCS, MMF, and other comparators included via the indirect comparison. The EAG considered there to have been a fundamental misinterpretation by the company with respect to the difference between RDI and TTD, which means that while premature discontinuation is captured within the model (through TTD), any dose adjustments are not reflected (through RDI, or an equivalent measure). RDI is not clearly reported in the CS, nor is it contained within the AURORA 1 or AURORA 2 clinical study reports provided within the CS reference pack. For MMF, the company costed this assuming a dose of 2.5 g/day, whereas in

Report sections	Sections 4.2.6 and 4.2.8
	AURORA 1 and AURORA 2 this was dosed at 2 g/day. Moreover, in AURORA 2, MMF dose reductions were permitted per protocol, and this is not reflected within the company's model. For other comparators, TTD is assumed to be 100% which the company justified based on a lack of data to quantify premature treatment discontinuation. The EAG considered this to be inappropriate given that some patients are expected to discontinue treatment due to lack of efficacy or occurrence of AEs.
What alternative approach has the EAG suggested?	The EAG has incorporated a number of edits to address some of the costing issues, and has explored a variety of scenarios to address areas of outstanding uncertainty. These are described throughout Section 6 of this report.
What is the expected effect on the cost-effectiveness estimates?	Incorporating RDI adjustments (assuming 95% for all treatments) causes the ICER to decrease, whereas all other edits to costs generally caused the ICER to increase slightly. However, combining all changes causes the ICER to increase, with details provided in Section 6.2 of this report.
What additional evidence or analyses might help to resolve this key issue?	The EAG notes there is no such evidence that would likely resolve the uncertainty associated with the incorporation of costs within the model unless the company had relative dosing information available, but expects the various changes and sensitivity analyses warrant further discussion at technical engagement and/or by the committee to determine the most suitable basis to inform decision making.

Abbreviations: CS, company submission; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; MMF, mycophenolate mofetil; RDI, relative dose intensity; TTD, time-to-treatment-discontinuation; VCS, voclosporin.

Key Issue 6: There is a lack of transparency around the inputs used in the company's model

Report sections	Section 4.2
Description of issue and why the EAG has identified it as important	The EAG identified a number of issues with respect to transparency of reporting in both the CS and the company's model, which impacted its ability to verify a variety of aspects of the CS. Issues included hardcoded values which did not match source material (due to inflation and/or converting outputs for use within the model), misalignment in source costs with those used in the model, inconsistencies in apparent inflation indices used to adjust costs, and non-systematic identification of drug costs leading to some costs that were higher than other available sources (e.g., prednisolone sourced from BNF and not eMIT).
What alternative approach has the EAG suggested?	The EAG has included edits to model inputs where it could clearly identify discrepancies between source data and intended values for the model. However, it was not possible for the EAG to reconcile all apparent discrepancies with information provided to the EAG, and the timeframe available for the EAG to conduct its review.
What is the expected effect on the cost-effectiveness estimates?	The impact on cost-effectiveness estimates for making these edits is small, if the EAG is correct in its interpretation of the intended use of costs and other inputs, and if any outstanding issues are clarified by the company. However, the EAG considered it important to raise this issue with transparency since the EAG has

Report sections	Section 4.2
	highlighted numerous instances of input parameters which are not clearly referenced and therefore could contain errors but that the EAG could not verify.
What additional evidence or analyses might help to resolve this key issue?	The EAG encourages the company to verify the model input parameters referred to throughout this report to provide reassurance to the committee that the values used are accurate and appropriate to inform decision making.

Abbreviations: BNF, British National Formulary; CS, company submission; EAG, Evidence Assessment Group; eMIT, electronic market information tool.

1.6. Other key issues: summary of the EAG's views

Key Issue 7: Uncertainty in how voclosporin will be used in practice

Report sections	**2.4, 3.2.2, 3.2.3.2, 3.3.3, 4.2.3, 4.2.6.7**
Description of issue and why the EAG has identified it as important	The treatment pathway and the way in which treatments are administered to people with LN is highly variable across the population. The choice of treatment is tailored to patients' needs, and there is a lack of clear evidence about the optimal duration of treatment with immunosuppression. The evidence presented by the company represents one way in which voclosporin may be used: administered at either first or second line (after MMF monotherapy), and with a target duration of 3 years (with a small number of trial participants permitted to withdraw due to response after 2 years of treatment). Clinical effects are based on a combined population of people receiving voclosporin at different lines of treatment, and mostly receiving treatment for close to 3 years. Clinical advice to the EAG is that this may not be how voclosporin is used in practice, as clinicians may seek to continue existing flexibility with treatment choice and duration. Moreover, using voclosporin routinely at first line would be a change in practice, since other CNIs are usually administered later such as when people do not respond to MMF alone. Where variations in practice existed within the trials of voclosporin (such as prior treatment with MMF or treatment discontinuation < 3 years), a lack of statistical power meant that the company was unable to evaluate how these variations influenced the treatment effect. The EAG considered it uncertain but plausible that the effect of voclosporin may vary according to the way it is used. Subgroup analyses from AURORA 1 and AURA-LV suggested that line of treatment may have a significant impact on the magnitude of treatment effect, but the findings between studies were conflicting, and neither the company nor the EAG were able to resolve the reason behind this. The EAG is also aware that variation in the duration of immunosuppression treatment can affect the risk of relapse, but this evidence does not provide a clear steer on the length of time people should receive immunosuppressive treatment. Due to uncertainty in the way treatment for LN is administered, it is likely that further evidence may arise that guides the duration and withdrawal of voclosporin and other treatments. Together, the EAG was unable to rule out that the effect of voclosporin may vary according to how it is used, which has implications for the clinical and cost effectiveness of voclosporin in a way that cannot be fully understood at present.

Report sections	2.4, 3.2.2, 3.2.3.2, 3.3.3, 4.2.3, 4.2.6.7
What alternative approach has the EAG suggested?	While acknowledging the lack of statistical power in the included trials, and quality issues with AURA-LV that affect the feasibility of a pooled meta-analysis, at clarification, the EAG requested further sensitivity analyses from the company to explore the effect of voclosporin according to line of treatment [CQ A15]. The company restricted their response to analyses already presented in the CS, and did not present additional data e.g. for other outcomes or using data from AURORA 2. It is possible that further analyses may have been informative for this matter (e.g. a consistent pattern in effects across outcomes may had increased confidence in the presence of an effect), however, multiple post-hoc analyses that are also under-powered would not have generated estimates with sufficient confidence for decision-making. Furthermore, if differences between the design of AURA-LV and AURORA 1 contributed to the conflict in findings, further analyses would perpetuate these differences without providing insight into the reasons for conflict. Overall, the EAG considered that the company's trial evidence did not sufficiently explore how variation in the use of voclosporin would affect its effect for people with LN, and this is challenging to resolve at this stage.
What is the expected effect on the cost-effectiveness estimates?	Transitions in the company's model are derived from transitions observed within AURORA 1 and AURORA 2, and therefore represent the way in which voclosporin was used within those trials. Separate data were not presented according to whether participants were or were not using MMF at baseline, or according to a different approach to treatment duration. The EAG have no reliable estimates for how the effect of voclosporin may vary across populations and variations in its use, and within the model structure, the EAG was unable to explore how altering the magnitude of treatment effect for voclosporin would affect cost effectiveness. Overall, the EAG considered that the company model likely does not represent solely the way in which voclosporin would be used in practice, but is unable to determine how this has affected cost effectiveness estimates for voclosporin without further analyses from the company.
What additional evidence or analyses might help to resolve this key issue?	The EAG considered that, due to limitations in the trial evidence, this issue cannot be resolved without further evidence generation. However, the company may be able to provide further evidence to inform the committee in its decision-making. For example, the company may be able to provide further analyses that explore the effect of changes to the treatment pathway; such as the position of voclosporin in the treatment pathway, and variation in the duration of treatment. The company may also be able to provide data for the model separated according to MMF use at baseline, which may give an indication for how cost effectiveness may vary according to its use.

Abbreviations: CQ, clarification question; CS, company submission; EAG, Evidence Assessment Group; LN, lupus nephritis; MMF, mycophenolate mofetil

1.7. Summary of EAG's preferred assumptions and resulting ICER

A summary of the ERG's preferred assumptions and resulting ICER is provided in Table 3.

Table 3: Summary of EAG's preferred assumptions and ICER

Preferred assumption	Incremental cost	Incremental QALYs	Cumulative ICER £/QALY
Company base-case			*£19,876*
Company base-case with fix applied	[redacted]	[redacted]	***£19,897***
Align resource use, AE, EOL and drug costs	[redacted]	[redacted]	*£20,114 (£217)*
Add in ½ pack wastage for voclosporin	[redacted]	[redacted]	*£20,413 (+£516)*
Update trial utilities to weighted average from AURORA 1 and AURORA 2 observations	[redacted]	[redacted]	*£21,401(+£1,504)*
Update literature-based utilities for transplant from Li et al.2017	[redacted]	[redacted]	*£20,152(+£255)*
Update literature-based utilities for dialysis from meta-analysis of Cooper et al. 2020	[redacted]	[redacted]	*£19,984(+£87)*
Apply 95% RDI to all treatments	[redacted]	[redacted]	*£18,699 (-£1,198)*
Removal of LN death in CKD stage 1-3a	[redacted]	[redacted]	*£23,497 (+£3,600)*
Allow transitions CKD stage 3b-4 in first 36 months	[redacted]	[redacted]	*£14,811 (-£5,086)*
Use average long-term transition probabilities from VCS+MMF and MMF applied to both arms	[redacted]	[redacted]	*£45,446(+£25,549)*
EAG base case	[redacted]	[redacted]	*£40,029 (+£20,132)*

Abbreviations: CKD, chronic kidney disease; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; LN, lupus nephritis; MMF, mycophenolate mofetil; QALY, quality adjusted life year; RDI, relative dosing intensity, VCS, voclosporin

Modelling errors identified and corrected by the EAG are described in Section 6.1. For further details of the exploratory and sensitivity analyses done by the EAG, see Section 6.2.

2. INTRODUCTION AND BACKGROUND

2.1. Introduction

In this report, the Evidence Assessment Group (EAG) provides a review of the evidence submitted by Otsuka Pharmaceuticals in support of voclosporin (Lupkynis) in combination with immunosuppression therapy for the treatment of lupus nephritis.

Lupus nephritis (LN) is a common complication of systemic lupus erythematosus (SLE), which is an autoimmune condition affecting an estimated 60,000 people in England and Wales. [1] LN is experienced by between 33% - 60% of people with SLE, with higher incidence in people with high-risk disease or those with prior kidney injury. LN occurs when chronic inflammation within the glomerular affects the ability of the kidney to filter waste and excess substances including proteins from the blood. This leads to kidney damage, which can lead to end stage renal disease (ESRD), and serious health outcomes such as heart attacks and strokes. People with LN have a higher standardised risk of mortality compared to the general population (6 – 6.8 vs. 2.4), and have a shorter life expectancy compared to people with SLE who do not have LN. [2]

LN typically develops within 5-years of diagnosis of SLE, though 25% - 50% of people show signs of LN at the time of SLE diagnosis, [3] which may be due to general under-diagnosis of SLE. Data describing the prevalence and incidence of LN in the UK are currently limited. Among publicly available data, the most recent UK-specific study was a 2001 retrospective analysis conducted in England, which reported overall LN prevalence and incidence rates of 4.4 and 0.4 per 100,000 of the population, respectively. [4] While SLE is generally more common amongst females, in general studies report that males with SLE are at higher risk of developing LN. Additional risk factors include people within certain Black, Asian and Hispanic ethnic groups, juvenile onset of SLE, and the presence of high risk genetic markers. In general, 5-yr risk of ESRD in people in LN is 11% (95% CI 10–12%), 10-yr is 17% (95% CI 16–18%), and 15-yr is 22% (95% CI 20–23%). The risk is higher in developing nations, particularly for 15-yr risk. Also higher risk for those with higher class LN, with highest risk in class IV: 5-year, 10-year, and 15-year risks of 19% (95% CI 12–29%), 33% (95% CI 22–44%), and 44% (95% CI 32–56%).[5]

Treatment for LN is similar to the approach used for SLE, and includes high-dose corticosteroids to rapidly control inflammation, followed by immunotherapy (including mycophenolate mofetil [MMF] and cyclosphosphamide). Sometimes additional treatment with a calcineurin inhibitor (CNI; such as tacrolimus), an anti-malarial (hydroxychloroquine), or with

rituximab is indicated. Controlling the inflammation may limit damage to the kidney and reduce the risk of ESRD; however, a third of patients who experience a complete response to treatment nevertheless relapse. Treatments for LN also carry their own risks, and drug-induced toxicity and the increased risk of infections are associated with early mortality and morbidity.

Voclosporin is a novel CNI which, like other CNIs used to treat LN, blocks T-cell activation instrumental in causing inflammation, and independently decreases proteinuria by reinforcing the integrity of podocytes in the glomeruli. Voclosporin does not currently have a licence for use in the UK; in November 2021 the European Medicines Agency (EMA) requested further information from the company, to which it is still preparing its response (as of January 2022).[6] If the company receive a positive decision for voclosporin from the EMA, [redacted].

2.2. Critique of the company's description of the underlying health problem

The EAG considered that the company's description of LN was representative of the condition, and included consideration of relevant available evidence.

2.3. Critique of the company's overview of current service provision

The company accurately summarised treatment recommendations for LN published by EULAR/ERA-EDTA.[7] Clinical advice to the EAG was consistent with statements from the company that people with LN typically receive hydrochloroquine, and that tacrolimus is the CNI treatment most used. However, clinical advisors noted that cyclophosphamide (CYC) is now rarely used within the NHS, due to toxicity. As shown by the EULAR/ERA-EDTA recommendations, initial immunosuppressive treatment for LN is MMF or MPA. The company noted that other treatments, including CNIs, may be used at first-line in certain circumstances, for example if standard doses of MMF or MPA are contra-indicated, or for those with nephrotic-range proteinuria. Clinical advisors to the EAG also noted that an alternative to MMF may be used in case of planned pregnancy. However, advice to the EAG was that alternatives to MMF and MPA in the first-line are rarely used. Advisors also noted that consideration for using a CNI would depend on a person's kidney function, since CNIs are associated with a risk of kidney damage.

Clinical advisors to the EAG noted that, while treatment of LN is evidence-driven, the evidence does not support a one-size-fit-all approach to management. As shown in the EULAR/ERA-

EDTA recommendations reported by the company, there are multiple options available at each stage, and clinicians choose a strategy according to patient preferences, their disease severity and response to previous treatments, and their vulnerability to the safety profile of specific products. EULAR-ERA-EDTA also note that there is yet insufficient evidence to determine the optimum duration of treatment, which should balance the protective effects of treatment for controlling progression of kidney damage with safety risks. Clinicians can vary in their approach to management: one clinical advisor to the EAG reported that they would consider discontinuing treatment after 15- to 18-months, while another of their team typically discontinued after 1-year. Another clinical advisor to the EAG noted that treatment administered longer than 3-years would be consistent with EULAR/ERA-EDTA guidance.

2.4. Critique of company's definition of decision problem

The company submission (CS) was aligned with the decision problem (see Table 4). At clarification, the company noted that the expected licence for voclosporin would be in combination with MMF, which is consistent with the evidence presented.

The EAG were uncertain where in the treatment pathway voclosporin would typically be used. As described in Section 2.3, treatment with a CNI would typically be administered after patients had not responded to treatment with MMF/MPA alone, or if first-line treatment with MMF/MPA was contraindicated. In this case, the main comparators for voclosporin would be azathioprine, rituximab, or tacrolimus. The company do not present a direct comparison between voclosporin and these technologies, and therefore comparative efficacy is demonstrated through the company's network meta-analysis (NMA; Section 3.4). The company proposed that in this position, voclosporin would be used as an alternative to tacrolimus, as both are CNIs are therefore offer a similar mechanism for treating the disease, and potentially carry a similar safety profile (though the company suggested that the safety profile of voclosporin is improved compared to tacrolimus). The company also proposed that voclosporin be considered as an alternative to MMF/MPA in the first-line position. The EAG are unclear if the company intend for voclosporin to be used in the first-line in the same way other CNI therapies are used (i.e. if MMF/MPA is contra-indicated), or whether they intend for voclosporin to be used as an alternative to MMF/MPA in a larger group of patients with LN. Half of all participants included in the AURORA 1 and AURA-LV trials were not receiving MMF at screening for the trial, and it is unclear whether or why these patients were therefore receiving voclosporin as first-line treatment, or if they had previously received and discontinued MMF/MPA. It is therefore

plausible that the company wish the committee to consider a broader use of voclosporin than for other CNIs, though the EAG did not consider the company had substantiated this.

Table 4: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	Adults with active lupus nephritis	As per scope	N/A	The evidence submitted by the company was appropriate to the NICE scope
Intervention	Voclosporin with immunosuppressive therapies	As per scope	N/A	The evidence submitted by the company was appropriate to the NICE scope. The evidence presented evaluated the effectiveness of voclosporin in combination with MMF and immunosuppressive therapies. At clarification, the company confirmed that this is consistent with the expected licence for voclosporin.
Comparator(s)	Standard therapy for lupus nephritis without voclosporin including the following induction treatments, followed by maintenance treatment with mycophenolate plus corticosteroids or azathioprine plus corticosteroids:	As per scope	N/A	The evidence submitted by the company was appropriate to the NICE scope. As stated in Key Issue 7, the EAG were uncertain which comparators would be most appropriate for voclosporin, as it was unclear where in the treatment line voclosporin would be used.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	• mycophenolate plus corticosteroids • cyclophosphamide plus corticosteroids • azathioprine plus corticosteroids • rituximab • a calcineurin inhibitor plus mycophenolate and corticosteroids.			
Outcomes	The outcome measures to be considered include:	As per scope	N/A	The evidence submitted by the company was appropriate to the NICE scope
Economic analysis	The NICE reference case stipulates that: • the cost-effectiveness should be expressed as cost per quality-adjusted life year in a cost-utility analysis framework with fully incremental analysis where required • the model time horizon should be sufficiently long to fully capture all differences in costs and outcomes being compared between the technologies • Costs should be considered from an NHS	The EAG considered that the economic analysis largely matched the analysis outlined within the scope: • Cost effectiveness was expressed as a cost per quality adjusted life year • A lifelong time horizon was considered • Costs were considered from an NS and Personal Social Services perspective • Health effects were mapped to the EQ-5D • Costs and health effects were discounted	N/A	Mostly in line with the NICE scope, with concerns relating to model structure and the utility values obtained (see Section 4.2). Incremental analyses were not presented but have been provided by the EAG in Section 5.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	and Personal Social Services perspective • Health effects should be expressed in QALYs with the EQ-5D being the preferred measure in adults with sources of data being a representative sample of UK patients • Costs and health effects should be discounted at 3.5%			
Subgroups	None specified	N/A	N/A	N/A
Special considerations including issues related to equity or equality	None specified	N/A	N/A	N/A

Abbreviations EAG, Evidence Assessment Group; NICE, National Institute for Health and Care Excellence; N/A, not applicable

3. CLINICAL EFFECTIVENESS

3.1. Critique of the methods of review(s)

The company undertook a systematic literature review (SLR) to identify evidence from randomised controlled trials (RCTs) for voclosporin and its comparators for the treatment of active class III-IV LN. Overall, the EAG considered the review methods used by the company to be acceptable, though raised some concerns about the company's literature search strategy and its methods of quality appraisal. The EAG did not consider that issues with the company's search strategy would have a major impact on the findings of the review, as it considered it likely that all relevant evidence for voclosporin had been identified. This evidence includes a direct comparison with MMF, and a search of recent literature reviews by the EAG suggested that the company's review also included all relevant trials of tacrolimus + MMF, which the EAG considered the other principal comparator of interest. However, the EAG did have concerns about aspects of the quality assessment conducted by the company, which it considered underestimated risk of bias of the included trials.

Table 5: Summary of EAG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix D (D1.1.1)	The company literature searches were carried out in Proquest, which the EAG were unable to access and so searches were not tested. The company searched several databases together in one strategy, which is not best practice as, for example, terms can vary between databases. Moreover, the RCT filter that was used by the company is not the recognised, validated filter from the Cochrane Handbook; in clarification the company stated that they used a mixture of different filters, though that is not how they are designed to be used[8] and this makes the effectiveness of the search uncertain. Overall, the EAG considered it likely that the company's search strategy missed relevant papers. Clinical trials registers were not searched so relevant (unpublished, ongoing) trials may have been missed. The company stated that targeted PubMed searches were carried out for adverse events but the strategies were not provided in clarification, therefore it was not possible to assess the effectiveness of these. It is possible that exclusion of cohort, case-control, cross-sectional and case series as publication types in the literature searches (due to the use of an RCT filter) meant that papers reporting adverse events have been missed.
Inclusion criteria	Appendix D (D1.1.2)	The inclusion criteria were appropriate to the aims of the review and consistent with NICE methods. The criteria were limited to RCT evidence; while RCTs are the gold standard for determining relative efficacy, they often lack external validity, and in some topic areas, restriction to RCT evidence can result in a limited evidence base. New NICE guidance[9] allows for inclusion of non-randomised studies to supplement a limited evidence base, provide a counterpoint to RCT evidence, and provide insight into any concerns about the generalisability of trial evidence. Given the small evidence base for treatments for LN, the EAG considered it may have been valuable for the company to have broadened their SLR to include non-randomised evidence. However, the EAG were unable to identify non-randomised studies either of voclosporin, or including a comparison of tacrolimus+MMF (a comparator of interest for which there is no direct RCT evidence). The company also confirmed at clarification [A19] that it had been unable to find a non-randomised comparison of tacrolimus plus MMF. Ultimately therefore, the EAG considered it unlikely that the inclusion of non-randomised evidence would have contributed significantly to the evidence base.

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Screening	Appendix D (D1.1.2)	Screening methods were described in full, and were conducted according to gold standard practice.
Data extraction	NR	Methods for extraction of clinical effectiveness data were not reported
Tool for quality assessment of included study or studies	CS B.2.5	Critical appraisal of the trials of voclosporin and those included in the company's NMA was conducted using an appropriate checklist (NICE quality appraisal tool). The company did not nuance their appraisal according to outcome, which is a limitation in their approach. Overall the EAG judged the company's quality appraisal to be acceptable, though disagreed with several of their assessments, judging that these underestimated the risk of bias of the included trials. Moreover, the EAG noted that a response to an item in their appraisal of AURORA 2 was incongruent with the risk of bias under assessment. The company's quality appraisal of trials included in the NMA highlighted several issues with the included studies, though these were not discussed by the company.
Evidence synthesis	Paired meta-analysis: CS B.8. Network meta-analysis: CS B.2.9	The company did not conduct a substantial narrative synthesis of treatment effects across the included trials of voclosporin. The company did conduct a paired meta-analysis of data from comparable treatment arms in AURORA 1 and AURA-LV in an effort to capitalise on a larger sample size. The outcomes considered by the analysis were limited in scope, which limited the utility of the analysis in the appraisal. The utility of the analysis was also limited by concerns about the potential imbalance between treatment arms in AURA-LV. The company NMAs to evaluate the comparative effectiveness of voclosporin versus other treatments for LN. The analyses were restricted to two outcomes only (complete renal response and partial renal response), which despite being non-independent were analysed separately. The EAG considered that a multivariate analysis to include both outcomes would have been preferable. The EAG also considered that the findings of random effects models should have been prioritised in the base case, and that alternative priors should have been explored.

Abbreviations: CS, Company submission; EAG, Evidence Assessment Group; LN, lupus nephritis; MMF, mycophenolate mofetil; NMA, network meta-analysis; NR, not reported; RCT, randomised controlled trial; SLR, systematic literature review

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

3.2.1. Studies included in the clinical effectiveness review

The CS described three trials of voclosporin that were identified by the company's SLR (Table 6). These comprise a phase III, double-blind, placebo-controlled RCT (AURORA 1) and its extension (AURORA 2), and a Phase IIb, double-blind, placebo-controlled, three-armed RCT (AURA-LV). The company also conducted a pooled meta-analysis using data from AURORA 1 and AURA-LV (using those participants from AURA-LV who were randomised to the low dose arm). An overview of the methods of the included trials is provided in the following sections.

Table 6: Clinical trials included in the CS

Study name and acronym	Study design	Population	Intervention	Comparator	Location
AURORA 1	Phase III, double-blind, parallel-group, two-arm, multicentre RCT Follow-up: 52 weeks	Adult patients with SLE and LN class III – V as determined by a kidney biopsy, and who were considered to require high-dose corticosteroid and immunosuppressive treatment N=357	Voclosporin (23.7 mg BID) with MMF (2g) and low-dose corticosteroids	Placebo with MMF (2g) and low-dose corticosteroids	International (Europe 40 sites; USA 29 sites; Latin America 32 sites, South Africa 3 sites, Asia 38 sites). Sites in the UK: 0
AURORA 2	Phase III extension to AURORA 1. Double-blind, parallel-group, placebo-controlled, multicentre extension to a RCT Follow-up: 2 years	Patients recruited for AURORA 1 who completed 52 weeks of treatment in either arm N=216	Voclosporin (23.7 mg BID up to 12 months, then patients with controlled UPCR become eligible for a dose reduction to 15.8mg BID for the final 12 months; otherwise dosage remains the same) with MMF (2g) and low-dose corticosteroids	Placebo with MMF (2g) and low-dose corticosteroids	International (Europe 30 sites; USA 24 sites; Latin America 23 sites; South Africa 3 sites; Asia 25 sites) Sites in the UK: 0
AURA-LV	Phase IIb double-blind placebo-controlled, three-arm, multicentre study Follow-up: 48 weeks	Adult patients with SLE and LN class III – V N=265	Voclosporin 23.7mg BID or Voclosporin 39.5 mg BID, with MMF (2g) and low-dose corticosteroids	Placebo with MMF (2g) and low-dose corticosteroids	International (Europe, Americas, Asia) Sites in the UK: 0

Abbreviations: BID, twice daily; LN, lupus nephritis; MMF, mycophenolate mofetil; RCT, randomised controlled trial; SLE, Systemaic Lupus Erythematosus; UPCR, Urine Protein Creatinine Ratio

3.2.2. Description and critique of the design of the studies

3.2.2.1. Design of the studies

None of the trial sites were based in the UK, which at clarification the company stated was due to an understanding during the feasibility assessment of the trial that interest in the clinical trial of voclosporin in the UK would be less than elsewhere in Europe. Clinical advisors to the EAG were unable to explain why this might be the case but did not consider management of LN to vary greatly between countries. However, they noted that the incidence of prognostic markers in the LN population may vary between locations (for example, variation in the proportion of the population from certain ethnic minority groups).

AURORA 1 was an international multicentre placebo-controlled RCT with follow-up of one year. The EAG considered that the trial was of high quality (see Section 3.2.2.5), however had concerns about the length of follow-up and the lack of statistical power.

The EAG considered that the follow-up for the trial was short given that voclosporin (and its comparators) may be expected to be administered over several years. The company implemented a target for 3-years of treatment with voclosporin, and during AURORA 1 participants were permitted to withdraw or reduce their dose only for safety concerns (withdrawal after 2-years of treatment due to response was possible for those participants who continued from AURORA 1 into AURORA 2). The use of voclosporin over multiple years is consistent with current use of other immunosuppression treatments; this is done to ensure a complete renal response (CRR) and to protect against renal flares. The EMA advise that rates of renal response may be detected within 1-year of treatment, [10] though a minority of people may experience a response after more than 1-year of treatment. [11] For this reason the EAG considered it reasonable that a difference in renal response would be detected during the follow-up of AURORA 1, but that it was plausible that some but not all renal responses would be identified. The EAG were more concerned that AURORA 1 would be unable to detect incidence of renal flares, which would require follow-up of longer than 1 year. [10] Clinical advice to the EAG was also that the follow-up of the included trials may be limited for detecting differences how response to treatment may be sustained over time, and the effect of treatment on CKD progression. The company noted that only a minority of participants would be expected to transition from CKD stages 3b - 4 to CKD stage 5 within 1 year (CS, p.116). For adverse events (AE), the EAG considered that a 1 year follow-up would capture initial tolerance to the

treatments, but would not capture longer term toxicity effects associated with immunosuppressant therapy, such as infections and malignancies.

AURORA 1 was the largest of the included trials, but this trial only had sufficient statistical power for detecting change in its primary outcome, and was not powered for subgroup analyses. This seriously limits the scope of the evidence base for exploring variation in treatment effect across groups of interest, such as according to line of treatment, geographical location, and disease staging at baseline.

Participants in ***AURORA 2*** were those that completed the treatment regime in AURORA 1, chose to participate in the follow-on study, and met the trial inclusion criteria (see Section 3.2.2.2). Group allocation was maintained as in AURORA 1, and participants continued to be blinded. Follow-up was 2 years, thus completing follow-up for the target 3-year treatment period of voclosporin. There was a substantial loss of participants between AURORA 1 and AURORA 2: a total of 39.5% of participants did not participate (35.2% of the voclosporin arm and 43.8% of the placebo arm). The reasons for participants not continuing with AURORA 2 are summarised in Table 7; the major reasons were due to AEs, lack of efficacy, and a withdrawal of physician or participant consent. High levels of attrition, particularly where these are related to treatment, increase the risk of bias associated with trial data (see Section 3.2.2.5). This is attenuated slightly as the rate of discontinuation was comparable between arms, as were the reasons for discontinuation, though the EAG noted that the rate of withdrawal due to a lack of efficacy was greater in the placebo arm. Overall, the EAG concluded that absolute rates of events for all outcomes from AURORA 2 were subject to a high risk of bias, as they do not include consideration of participants who chose to discontinue treatment prior to AURORA 2. Relative risk estimates from AURORA 2 may be more reliable, provided that treatment effects are stable across LN populations; this is typically the case, though the EAG did not have clear evidence for this within LN. Finally, the EAG noted that AURORA 2 was underpowered to detect statistical significance in any clinical outcome, including primary trial outcomes, and no subgroup analyses were conducted. This further limits the utility of the AURORA 2 trial.

Table 7: Reasons that participants from AURORA 1 did not enrol in AURORA 2

	AURORA 1	
	VCS (n=63)	PbO (n=78)
Permanent treatment discontinuation	[redacted]	[redacted]

	AURORA 1	
	VCS (n=63)	PbO (n=78)
AE	[redacted]	[redacted]
Protocol non-compliance	[redacted]	[redacted]
Pregnancy	[redacted]	[redacted]
Physician decision	[redacted]	[redacted]
Prohibited medication required	[redacted]	[redacted]
Lack of efficacy	[redacted]	[redacted]
Other	[redacted]	[redacted]
Withdrew from AURORA 1 prematurely	[redacted]	[redacted]
Intolerable AE	2 (1.1)	0 (0.0)
Death	1 (0.6)	5 (2.8)
Lost to follow-up	1 (0.6)	3 (1.7)
Physician decision	2 (1.1)	3 (1.7)
Prohibited medication required	1 (0.6)	0 (0.0)
Pregnancy	1 (0.6)	0 (0.0)
Protocol non-compliance	1 (0.6)	1 (0.6)
Withdrawal of consent	7 (3.9)	14 (7.9)
Lack of efficacy	0 (0.0)	1 (0.6)
'Other'	0 (0.0)	4 (2.2)
Administrative reasons	[redacted]	
Did not give consent due to life circumstances	[redacted]	
Not recorded	[redacted]	

Abbreviations: AE, adverse event; PbO, placebo; VCS, voclosporin

^rates for each arm not reported; * note that sub-categories total more than 15. EAG is unclear whether this is because a participant gave more than one reason for discontinuing; #the EAG cannot account for 10 participants missing from AURORA 2 in the breakdown of reasons provided by the company

Source: Table B.2-5 of the CS, p.37; company clarification response A10

AURA-LV was an international multicentre phase IIb dose-finding trial, comparing two doses of voclosporin with each other and with a matching placebo. The trial appeared well-conducted, however an anomalous high mortality rate in the low-dose arm of voclosporin led to the company concluding that a chance imbalance in randomisation had undermined the internal validity of the trial. At clarification [A26] the company provided a report summarising the deliberations of an internal board that reviewed the mortality data in AURA-LV, which concluded that the deaths were unrelated to treatment, and may have resulted from an imbalance in

disease severity and treating centre. [12] The EAG accepted the conclusions of the report, noting that chance imbalances in baseline characteristics can occur no matter how rigorous the methods used, particularly for smaller trials. However, the EAG considered that the findings of the AURA-LV trial are therefore at a higher risk of bias, as where one imbalance is noted, more may be present and undetected.

3.2.2.2. Population

Trial inclusion criteria

Population inclusion and exclusion criteria for the included trials are provided in Table 8. The EAG considered that these criteria were reasonable and aligned with the target patient population for voclosporin. While the criteria excluded people with significant comorbid health conditions and a medical history with severe infections or cardiovascular conditions, clinical advice to the EAG was that these criteria would not exclude a high proportion of people with LN in clinical practice. This is because many people with LN are younger and are less likely to have these serious conditions.

People with CKD stage 3b and above at screening were also excluded, as were those who were expected to need a transplant during the trial duration. The EAG considered that this was also consistent with the intended use of voclosporin.

Table 8: Key inclusion/exclusion criteria for the included trials

	AURORA 1	AURORA 2	AURA-LV
Inclusion	Adults aged 18 – 75 years Diagnosis of SLE (per ACR criteria) LN, as defined as class III-V, including mixed class Active LN according to a kidney biopsy* Requires high-dose corticosteroids and immunosuppression therapy	Completed 52 weeks of treatment with study drug in the AURORA 1 study, including anyone who had discontinued and re-started treatment. Continued immunosuppressive therapy was required	Adults aged 18 – 75 years Diagnosis of SLE (per ACR criteria) LN, as defined as class III-V, including mixed class Active LN according to laboratory findings# Requires high-dose corticosteroids and immunosuppression therapy
Exclusion	eGFR ≤45 ml/min/1.73 m2 at screening Requires renal dialysis at screening or during the trial period	Requires renal dialysis at screening or during the trial period Planned kidney transplant	eGFR ≤45 ml/min/1.73 m2 at screening Requires renal dialysis at screening or during the trial period

	AURORA 1	AURORA 2	AURA-LV
	Previous or planned kidney transplant Current or medical history of malignancy^ or severe viral infection. Current severe active conditions, including infections requiring antibiotics, severe cardiovascular disease, liver disease	A medical condition with increased risk to the patient or may interfere with assessments	Previous or planned kidney transplant Current or medical history of malignancy^ or severe viral infection. Current severe active conditions, including infections requiring antibiotics, severe cardiovascular disease, liver disease

Abbreviations: ACR, American College of Rheumatology; eGFR, estimated glomerular filtration rate; LN, lupus nephritis; SLE, systemic lupus erythematosus

*within 2 years or 6 months prior to baseline, depending on UPCR rate (see Table B.2-3, p. 31 CS); # Further details in table B.2-13, p.49 CS)

Baseline characteristics

Key baseline characteristics for the included trials are summarised in Table 9.

The EAG considered that the trial populations appeared comparable with the target LN population for voclosporin: participants were mostly female and in early to mid-age, were in biopsy class III-IV and IV, and had active LN at the time of screening. Baseline measurement of eGFR and UPCR was consistent with active LN and concurrent kidney damage. Trial participants had been diagnosed with LN approximately 3 to 5 years prior to the trials.

The EAG considered that the company had reported a reasonable scope of baseline characteristics, though noted the omission of some characteristics that have prognostic value (e.g. incidence of those with juvenile-onset, high risk biomarkers), and that there was a lack of information about the previous treatment received by those in the trials. As treatment efficacy may vary according to the aggressiveness of a person's disease and their previous treatment, the EAG considered it could not rule out differences between trials and trial arms that may have affected trial outcomes. This concern was bolstered given that disease characteristics for those in the low-dose voclosporin arm of AURA-LV appeared comparable to those in the other arms and trials using the characteristics reported, but they subsequently had a higher risk of mortality, which may in part have been due to higher disease severity. [12]

Participants were randomised to treatment arms on a 1:1 ratio, stratified by biopsy class (class V or other), MMF use at baseline (yes/no), and region (North America vs Latin America vs Europe and South Africa vs Asia-Pacific). Within AURORA 1, trial arms were reasonably well-

balanced (noting the concern above). Fewer characteristics were reported for participants entering AURORA 2, though in the characteristics reported there was also reasonable balance. The EAG noted that those in the placebo arm of AURORA 2 were more likely to be in biopsy class III, and those in the voclosporin arm were more likely to be in biopsy class IV; such differences would be unsurprising given attrition between AURORA 1 and 2 effectively breaking randomisation, and the overall trial sample size. Several minor differences between trial arms were noted within AURA-LV: median age in the placebo arm was lower, and more participants in the low dose voclosporin arm were treated within Asian settings, and were White or Asian.

Overall, the EAG considered that the trial arms appeared well-balanced across most characteristics, including disease severity, but could not conclude that participants were entirely comparable due to missing details for some characteristics (e.g. previous treatment), and because of the lack of stable prognostic measures within LN.

Table 9: Demographic characteristics of included trial samples

		AURORA 1		AURORA 2		AURA-LV	
	VCS	Placebo	VCS	Placebo	VCS (low dose)	VCS high dose	Placebo
Age, median (range), years	31 (18–62)	32 (18–72)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Female, n (%)	161 (90)	152 (85)	[redacted]	[redacted]	76 (85.4)	81 (92.0)	73 (83.0)
Region (%)	Asia Pacific 29% Europe and South Africa 29% Latin America 27% North America 15%	Asia Pacific 29% Europe and South Africa 29% Latin America 27% North America 15%	NR	NR	Asia: 58.4% Europe: 28.1% Americas 13.5%	Asia: 48.9% Europe: 28.4% Americas: 22.7%	Asia: 39.8% Europe: 38.6% Americas: 21.6%
Race	White 38% Black 15% Asian 30% Other 18%	White 34% Black 11% Asian 31% Other 24%	[redacted]	[redacted]	White: 33.7% Black: 3.4% Asian Indian subcontinent: 24.7% Asia other: 33.7% Other: 4.5%	White: 40.9% Black: 6.8% Asian Indian subcontinent: 22.7% Asia other: 27.3% Other: 2.3%	White: 47.7% Black: 5.7% Asian Indian subcontinent: 20.5% Asia other: 20.5% Other: 5.7%
Ethnicity	Hispanic or Latino 32% Other 68%	Hispanic or Latino 33% Other 66% Unknown 1%	[redacted]	[redacted]	Hispanic or Latino 10.1% Other 89.9%	Hispanic or Latino 14.8% Other 85.2%	Hispanic or Latino 14.8% Other 85.2%
Time since initial LN diagnosis, mean (SD), years	4.6 (5.1)	4.7 (4.9)	[redacted]	[redacted]	4.2 (5.1)	3.2 (4.4)	3.5 (4.0)
Time since SLE diagnosis, mean (SD), years	6.6 (6.4)	6.9 (6.1)	NR	NR	[redacted]	[redacted]	[redacted]

		AURORA 1		AURORA 2		AURA-LV	
Biopsy class, n (%)	Pure class III: 11% Pure class IV: 51% Pure class V: 14% Class II and V only: 0% Class III and V only: 13% Class IV and V only: 11%	Pure class III: 16% Pure class IV: 43% Pure class V:14% Class II and V only: <1% Class III and V only: 11% Class IV and V only: 15%	[redacted]	[redacted]	Pure class V: 13.5% Class III/IV: 62.9% Class III+V or IV+V: 23.6%	Pure class V: 15.9% Class III/IV: 71.6% Class III+V or IV+V: 12.5%	Pure class V: 14.8% Class III/IV: 67% Class III+V or IV+V: 18.2%
Baseline eGFR Mean (SD), mL/min/1.73 m²	92.1 (30.6)	90.4 (29.0)	[redacted]	[redacted]	95.3 (28.4)	104.0 (27.3)	100.2 (27.1)
Mean (SD) baseline UPCR, mg/mg	4.14 (2.71)	3.87 (2.36)	[redacted]	[redacted]	5.16 (4.2)	4.48 (3.0)	4.43 (3.6)
SELENA-SLEDAI, mean (SD); n	13.2 (6.5); n=177	11.8 (6.1); n=177	NR	NR	NR	NR	NR
MMF use at screening, n (%)	100 (56)	96 (54)	NR	NR	31 (34.8)	29 (33.0)	32 (36.4)

Abbreviations: eGFR, estimated glomerular filtration rate; LN, lupus nephritis; MMF, mycophenolate mofetil; SD, standard deviation; UPCR, urine protein/creatinine ratio; VCS, voclosporin

Source: CS; AURORA 2 CSR; AURA-LV CSR

3.2.2.3. Intervention and comparator

Treatment characteristics and dose modifications for the included trials are summarised in Table 10. Additional details about the tapering of MMF for those not receiving this at baseline are provided in the CS, along with a list of permitted concomitant therapies (AURORA Table B.2-4, p. 34; ADD other refs).

Intervention characteristics for AURORA 1 were the same as those used in the low dose (23.7mg) arm of AURA-LV. In AURORA 2, intervention characteristics were similar but different rules about dose modification were used to account for participants having received treatment for 1-year prior to the trial. All three trials permitted dose modification due to safety events, but in AURORA 2, participants receiving voclosporin with controlled UPCR could also receive a reduction in dose to 15.8mg (2 capsules, twice daily). [redacted] in each arm were receiving a reduced dose by the end of AURORA 2 ([redacted] in the voclosporin arm and [redacted] in the placebo arm), though [redacted] participants received a reduced dose due to UPCR being well controlled ([redacted] after two years of treatment). Compliance to voclosporin across trials was high (>99%).

Exposure to MMF and corticosteroids were provided by the company at clarification (A18). Exposure to MMF was [redacted]. Median exposure to oral prednisone [redacted].

The comparator to voclosporin for all three trials was a matching number of capsules containing a placebo. All other treatment details were the same as the intervention arm. Dose reductions in placebo were managed by altering the number of capsules administered.

Table 10: Intervention characteristics of the included trials

	AURORA 1		AURORA 2		AURA-LV		
Voclosporin and comparator	23.7 mg voclosporin (administered as three 7.9 mg capsules) BID	Matching placebo	23.7 mg voclosporin (administered as three 7.9 mg capsules) BID	Matching placebo	23.7 mg voclosporin (administered as three 7.9 mg capsules) BID	39.5 mg voclosporin (administered as five 7.9 mg capsules) BID	Matching placebo (three or five capsules BID)

	AURORA 1	AURORA 2	AURA-LV
Treatments administered to both arms	2g MMF daily Days 1&2: IV methylprednisolone once daily (0.25 – 0.5g according to weight) Day 3: Oral prednisone (20 - 25mg/day according to weight). Tapering to begin on subsequent days Week 16: Oral prednisone 2.5mg/day	2g MMF daily Days 1&2: IV methylprednisolone once daily (0.25 – 0.5g according to weight) Day 3: Oral prednisone (20 - 25mg/day according to weight). Tapering to begin on subsequent days Week 16: Oral prednisone 2.5mg/day	2g MMF daily Days 1&2: IV methylprednisolone once daily (0.25 – 0.5g according to weight) Day 3: Oral prednisone (20 - 25mg/day according to weight). Tapering to begin on subsequent days Week 16: Oral prednisone 2.5mg/day
Dose modification	Modification was permitted due to a decrease in renal function, increased blood pressure, or an abnormal heart rhythm.	After 1 year in AURORA 2 (i.e. 2 years of treatment), participants were permitted to reduce the dose of voclosporin to 15.8mg (2 capsules) provided UPCR was controlled. Dose modification was also permitted due to adverse events, included but not limited to those specified for AURORA 1.	Modification was permitted due to a decrease in renal function, increased blood pressure, or an abnormal heart rhythm.

Abbreviations: BID, twice daily; MMF, mycophenolate mofetil; mg, milligram; UPCR, urine protein/creatinine ratio

3.2.2.4. Outcomes

The outcomes reported in the trials are summarised in Table 11. Outcomes measured consistently across trials were CRR, PRR, change in serum creatinine, urine protein, UPCR and eGFR, immunology parameters, and SELENA-SLEDAI (SLE disease activity). All trials also captured safety outcomes. AURA-LV measured a broader range of outcomes related to CRR

and PRR, such as time to event outcomes and the rate of sustained response. In AURORA 1 and 2, these outcomes were replaced by measures specific to UPCR. HRQoL was measured in AURORA 1 and 2; both trials measured generic HRQoL using the SF-36, while AURORA 1 also reported disease-specific HRQoL using the Lupus Pro measure.

The EAG concluded that the definitions of CRR and PRR used within the trials were clinically relevant. Data for each of the outcomes making up the composite CRR outcome were provided by the company for AURORA 1, and were provided for AURORA 2 and AURA-LV at clarification. EULAR/ERA-EDTA (2019) guidelines note that proteinuria and serum creatinine in particular are strongly associated with long-term kidney outcomes, and that treatment should aim for ≥25% reduction in proteinuria at 3 months, ≥50% at 6 months and complete renal response (<500–700 mg/day) at 12 months. Thresholds for change in UPCR used by the company were therefore considered to be predictive of longer-term outcomes. On the whole, advice to the EAG was that smaller changes in renal response7 outcomes are generally considered to be unreliable, due to natural fluctuation in measurements over time.

The EAG noted that the company varied the threshold at which safety events were reported across trials, and that this variation was not justified by the company, pre-specified in trial protocols, or tied to the sample size:

- AURORA 1: TEAEs at ≥4%, serious TEAEs at ≥2 patients; TEAEs leading to discontinuation or dose modification at ≥2%; no threshold for all others.
- AURORA 2: TEAEs at ≥3%, serious TEAEs at ≥2%; no other thresholds
- AURA-LV: TEAEs at ≥5%; serious TEAEs at ≥2 patients, TEAEs leading to discontinuation at ≥2%.

A different threshold for AEs was also used in the company model (grade 3 or 4 AEs were included where these were reported by ≥1% of participants). Variation in reporting thresholds across outcomes and trials is an indication of reporting bias (see Section 3.2.2.5), as it may occlude events and patterns in events across trials. In this case, the EAG were concerned that variation in threshold was occluding AE events that were high severity but low incidence; however, the EAG did not identify evidence of this from the trial CSRs.

Table 11: Outcomes measured by the included trials

	AURORA 1 Final follow-up: 1 year	AURORA 2 Final follow-up: 2 years	AURA-LV Final follow-up: 1 year	Pooled analysis of AURORA 1 and AURA-LV
CRR, defined as all the following: • UPCR of ≤0.5 mg/mg • eGFR of ≥60 ml/min/1.73^2 or no confirmed eGFR decrease of >20% from baseline • no rescue medication • no more than 10 mg prednisone equivalent per day for ≥3 consecutive days or for ≥7 days in total during final 8 weeks	✓	✓	✓	✓
Time to CRR	×	×	✓	×
Duration of CRR	×	×	✓	×
PRR, defined as 50% reduction in UPCR from baseline	✓	✓	✓	✓
Time to PRR	×	×	✓	×
Duration of PRR	×	×	×, though measured	×

	AURORA 1 Final follow-up: 1 year	AURORA 2 Final follow-up: 2 years	AURA-LV Final follow-up: 1 year	Pooled analysis of AURORA 1 and AURA-LV
			‘sustained’ PRR	
Reductions in UPCR	✓	×	×	✓
Time to reductions in UPCR	✓	×	×	✓
Duration of reductions in UPCR	✓	×	×	×
Change in serum creatinine, urine protein, and eGFR from baseline	✓	✓	✓	×
Change from baseline in immunology parameters (complement 3 (C3), C4, and anti-ds DNA) at weeks 24 and 52	✓	✓	✓	×
Renal flares	✓	✓	×	×
Extra-renal flares	×	✓	×	×
Generic HRQoL (SF-36)	✓	✓	×	×
Disease specific HRQoL (LupusPRO)	✓	×	×	×

	AURORA 1 Final follow-up: 1 year	AURORA 2 Final follow-up: 2 years	AURA-LV Final follow-up: 1 year	Pooled analysis of AURORA 1 and AURA-LV
SLE disease activity (SELENA-SLEDAI)	✓	✓	✓	×
Safety	✓	✓	✓	×
Subgroup analyses conducted (including age, gender, race, biopsy class, region, MMF use at baseline)	✓	×	✓	✓

Abbreviations: AE, adverse event; CRR, complete renal response; eGFR, estimated glomerular filtration rate; g, gram; MMF, mycophenolate mofetil; PRR, partial renal response; SAE, serious adverse event; UPCR, urine protein/creatinine ratio

^ provided at clarification at request of the EAG

3.2.2.5. Critical appraisal of the design of the studies

The company provided quality assessment ratings of the included trials using the critical appraisal checklist recommended by NICE, [13]. Although this is an acceptable tool, ratings presented by the company did not include consideration of how risk of bias may vary across outcome. Of relevance for the included trials, risk of bias ratings may vary between objective (e.g. clinical measures) and subjective outcomes (e.g. HRQoL), and risk of bias may be greater for some outcomes due to specific issues with their measurement. The company's ratings were reported in Section B.25 of the CS.

The EAG agreed with most of the ratings provided by the company, but considered there were some items of note:

- All trials were described as double-blind, and the company stated that patients, clinicians and all trial personnel were blinded to treatment allocation throughout the trials. It was unclear to the EAG which of the trial personnel were un-blinded, and therefore preventing the trials from being characterised as triple blind. On the whole, the EAG did not consider any lack of blinding to affect the measurement of most trial outcomes, though (depending on which personnel were not blinded and their role), this could affect subjective outcomes such as the two measures of HRQoL.
- The EAG did not consider that the company appraisal had sufficiently considered the impact of drop-out between AURORA 1 and AURORA 2 on the randomisation process of AURORA 2. As AURORA 2 was conducted as a separate trial to AURORA 1, and participants who started treatment in AURORA 1 but discontinued prior to AURORA 2 were not included in analyses of AURORA 2, this breaks the randomisation process. Few baseline characteristics were reported to determine the comparability of participants remaining in AURORA 2 across trial arms, and while reasons for discontinuation appeared comparable across arms, the EAG nevertheless considered the break in randomisation to be a high risk of bias in AURORA 2. Absolute rates of clinical outcomes were considered to be at particular risk of bias, though the EAG did not have evidence to confirm that relative effects would be stable once participants choosing to discontinue treatment were removed from the analysis. The EAG further noted that the company's response to the item on whether prognostic characteristics for AURORA 2 were balanced across arms was irrelevant and did not address the issue.
- The EAG were unclear why thresholds for reporting safety events varied across trials, when these were not explained, pre-specified in trial protocols, or appeared to be connected to

sample size. Changing thresholds across trials and/or outcomes is a signal of reporting bias, as thresholds may be changed to occlude patterns in the data. However, the EAG inspected the original safety data in the trial CSRs and did not identify any clear pattern of effect of concern.

- Sample sizes for AURORA 1 and AURA-LV were powered for the primary outcome only, which meant that it was not possible for the company to detect a reliable difference in effect on outcomes requiring greater power (e.g. those with low event rates), or to detect variation in effect across subgroups. AURORA 2 included only those participants who chose to continue from AURORA 1, and due to a high level of attrition at this time, AURORA 2 was under-powered for all its analyses.

3.2.3. Description and critique of the results of the studies

3.2.3.1. Clinical effectiveness results

Clinical effectiveness data for key outcomes from the included trials are shown in Table 12.

Renal response outcomes

Participants in both arms of the included trials experienced CRR, though the rate of CRR was higher for those receiving CRR across the trials. The breakdown in the composite outcome for CRR showed that voclosporin was beneficial for all outcomes, but the biggest effect was shown for proteinuria. This is exemplified by data from AURORA 1 showing that more than two thirds of those in the placebo arm met the required CRR criteria for eGFR and the use of rescue medication and prednisone, but only 23% of them also showed the required reduction in UPCR. A larger effect of voclosporin for proteinuria is consistent with voclosporin having an additional independent mechanism for reducing proteinuria in addition to its immunosuppressant mechanism. Clinical advice to the EAG was that both mechanisms – an improvement in kidney functioning as shown across outcomes of the CRR composite, and an independent reduction in proteinuria – would be beneficial for kidney function. Proteinuria is also a validated prognostic marker of longer-term kidney functioning.[7] However, clinical advice also cautioned that a reduction in proteinuria that does not result in disease modification may result in a corresponding level of nephrotoxicity.

There were limited data concerning the time to response, but some data from AURA-LV (time to response) and AURORA 1 (time to UPCR ≤0.5mg/mg) suggested that voclosporin may also lead to an earlier renal response, though this varied from a difference of weeks in AURA-LV to

days in AURORA 1. Clinical advisors to the EAG were uncertain whether this difference would be of clinical benefit to patients, noting that this may be the case for some participants who are experiencing a rapid decline in kidney function prior to treatment. There was a paucity of data concerning the duration of response; though on the whole, the EAG considered that the evidence did not demonstrate that duration of effect would differ between arms. In both arms of AURORA 2, the number of participants in CRR reduced between years 1 and 3, suggesting that participants began to relapse. However, the EAG also noted that the relative effect of voclosporin for CRR fluctuated in magnitude over the follow-up of AURORA 2, which may be consistent with the fluctuating nature of LN. Independent PRR data was not reported for AURORA 2 or AURA-LV, and were not calculable by the EAG on the data provided, but at clarification [A9] the company provided independent PRR data for AURORA 1. These data showed that amongst participants who did not achieve a CRR within 1 year, more participants in the voclosporin arm exhibited a PRR, though these effects were not statistically significant. Overall, the EAG concluded that the primary advantage of voclosporin was that people with LN may be more likely to achieve a renal response than with MMF and immunosuppressive treatment alone.

Renal relapse/flares

Data from AURORA 2 did not show a difference in the risk of renal flares up until end of the trial. The EAG concluded that these data suggested that those additional participants in the voclosporin arm who achieved a CRR were not more likely to relapse within 3 years of starting treatment. However, clinical advice to the EAG was that this follow-up is nevertheless still short for determining the long-term impact of renal response, including the nature and impact of relapse.

Table 12: Trial outcomes for renal response

		AURORA 1 Final follow-up: 1 year		AURORA 2 Final follow-up: 2 years		AURA-LV Final follow-up: 1 year	
	VCS (N=179)	**Placebo (N=178)**	**VCS (N=116)**	**Placebo (N=100)**	**VCS (N=89)**	**VCS high dose (N=88)**	**Placebo (N=88)**
CRR							
CRR	Week 24 32.4% OR 2.23 (1.3, 3.7)*	Week 24 19.7%	18 months [redacted]	18 months [redacted]	Week 24 32.6% OR 2.03 (1.01, 4.05)*	Week 24 27.3% OR NR	Week 24 19.3%
	Week 52 73 (40.8%) OR 2.65 (1.6, 4.3)*	Week 52 40 (22.5%)	24 months [redacted]	24 months [redacted]	Week 48 49.4% OR 3.21 (1.68, 6.13)*	Week 48 39.8% OR 2.10 (1.09, 4.02)*	Week 48 23.9%
			30 months [redacted]	30 months [redacted]			
			36 months [redacted]	36 months [redacted]			
Time to CRR	-	-	-	-	Median time: 19.7 weeks (16.1, 36.1) HR 2.26 (1.45, 3.51)*,≠	Median time: 23.4 weeks (13.7, 33.4) HR 2.25 (1.46, 3.47)*,≠	Median time: NR
Sustained CRR	-	-	-	-	[redacted]	[redacted]	[redacted]
Duration of CRR	-	-	-	-	[redacted]	[redacted]	[redacted]
Composite of CRR							
UPCR ≤ 0.5 mg/mg	52 weeks 81 (45.2%) OR 3.11 (1.9, 5.0)*	52 weeks 41 (23.0%)	[redacted]	[redacted]	[redacted]	[redacted]	NR

		AURORA 1 Final follow-up: 1 year		AURORA 2 Final follow-up: 2 years		AURA-LV Final follow-up: 1 year	
			[redacted]	[redacted]			
			[redacted]	[redacted]			
			[redacted]	[redacted]			
eGFR of ≥60 ml/min/1.73^2 or no confirmed eGFR decrease of >20% from baseline	52 weeks 147 (82.1%) 1.50 (0.9, 2.5)	52 weeks 135 (75.8%)	[redacted]	[redacted]	[redacted]	[redacted]	NR
			[redacted]	[redacted]			
			[redacted]	[redacted]			
			[redacted]	[redacted]			
Received no rescue medication for LN	52 weeks 163 (91.1%) 1.62 (0.8, 3.2)	52 weeks 154 (86.5%)	[redacted]	[redacted]	[redacted]	[redacted]	NR
Did not receive > 10 mg/day prednisone for ≥ 3 consecutive days or for ≥ 7 days in total during Weeks 44 through 52	52 weeks 156 (87.2%) 1.26 (0.7, 2.3)	52 weeks 152 (85.4%)	[redacted]	[redacted]	[redacted]	[redacted]	NR
PRR							

		AURORA 1 Final follow-up: 1 year		AURORA 2 Final follow-up: 2 years		AURA-LV Final follow-up: 1 year	
PRR (all patients who achieved a PRR)	24 weeks 126 (70%) OR 2.43 (1.56, 3.79)*	24 weeks 89 (50%)	18 months [redacted]	18 months [redacted]	24 weeks 69.7% OR 2.33 (1.26, 4.33)*,≠	24 weeks 65.9% OR 2.03 (1.10, 3.76)*	24 weeks 49.4%
	52 weeks 125 (70%) 2.26 (1.45, 3.51)*	52 weeks 92 (52%)	24 months [redacted]	24 months [redacted]	48 weeks NR	48 weeks NR OR 2.68 (1.43, 5.02)*,≠	48 weeks NR
			30 months [redacted]	30 months [redacted]			
			36 months [redacted]	36 months [redacted]			
PRR (patients who only achieved a PRR; i.e. did not achieve a CRR during follow-up)^	[redacted]	[redacted]	NR	NR	NR	NR	NR
Time to PRR	-	-	-	-	Median time: 1.3 weeks (2.6, 5.9) HR 1.63 (1.16, 2.27) *,≠	Median time: 4.4 weeks (4.1, 6.1) HR 1.74 (1.25, 2.43) *,≠	Median time: 6.6 weeks (4.6, 8.6)
Additional outcomes							
Time to UPCR of ≤0.5 mg/mg	Median 169 days HR 2.0 (1.5, 2.7) 64.8% of patients reached this at some point	Median 372 days 43.8% of patients reached this at some point	-	-	-	-	-
Time to 50% reduction in UPCR from baseline	Median 29 days 96.6% HR 2.05 (1.6, 2.6)*	Median 63 days 75.8%	-	-	-	-	-

		AURORA 1 Final follow-up: 1 year		AURORA 2 Final follow-up: 2 years		AURA-LV Final follow-up: 1 year	
Duration of UPCR of ≤0.5 mg/mg	Mean 163.3 days (1, 356)	Mean 158.8 days (1, 358)	-	-	-	-	-
Flares							
Renal flares (after achieving a UPCR of ≤0.7 mg/mg)	[redacted]	[redacted]	[redacted]	[redacted]	-	-	-
			[redacted]	[redacted]			
			[redacted]	[redacted]			
Extra-renal flares	-	-	[redacted]	[redacted]	-	-	-
			[redacted]	[redacted]			
			[redacted]	[redacted]			
			[redacted]	[redacted]			

Abbreviations: AE, adverse event; CRR, complete renal response; eGFR, estimated glomerular filtration rate; g, gram; HR, hazard ratio; MD, mean difference; MMF, mycophenolate mofetil; OR, odds ratio; NR, not reported; PRR, partial renal response; SAE, serious adverse event; UPCR, urine protein/creatinine ratio; VCS, voclosporin

Notes: * statistically significant (i.e. p value <0.05); ^analysis requested by the EAG; ≠ compared with placebo

Source: CS; clarification response [A7]

Health-related quality of life

HRQoL data as assessed using SF-36 were reported in appendices to the CS (Appendix N2), though disease –specific HRQoL data measured by Lupus Pro were not reported. Data from AURORA 1 were provided to the EAG by the company within the trial CSR, though this was not the case for AURORA 2, as while the trial CSR was provided, the accompanying data tables were not. The data for AURORA 1 showed that there was no difference in HRQoL between treatment arms at any timepoint, as measured using SF-36

and LupusPro. Change in HRQoL showed that there was a mean increase in HRQoL in both trial arms, though this was highly variable across the trial sample. The company reported that there was also no difference in HRQoL between treatment arms in AURORA 2. Clinical advice to the EAG was that it is plausible that people can experience a response to treatment that is clinically meaningful to their condition without showing a corresponding benefit in HRQoL. This is because the impacts of active disease and receiving immunosuppressive treatment can be detrimental to HRQoL, and improvements in HRQoL may not be seen until a response is stable and people have been withdrawn from treatment.

Additional clinical outcomes of interest

There was no difference in SLE disease activity between trial arms.

	AURORA 1 **Final follow-up: 1 year**		**AURORA 2** **Final follow-up: 2 years**		**AURA-LV** **Final follow-up: 1 year**		
	VCS (N=179)	Placebo (N=178)	VCS (N=116)	Placebo (N=100)	VCS (N=77)	VCS high dose (N=82)	Placebo (N=79)
SELENA-SLEDAI	Week 24 Mean change: -4.5 (5.4, -3.7) MD: -0.5 (-1.6, 0.6)	Week 24 Mean change: -4.1 (-5.0, -3.2)	[redacted]	[redacted]	Week 24 Mean change (range): -6.3 (5.86; -25, 6)*	Week 24 Mean change (range): -7.1 (7.41, -26.10)*	Week 24 Mean change (range):-4.5 (7.09, -26.12)
	Week 52 Mean change -6.0 (-6.7, -5.2)	Week 52 Mean change -5.5 (-6.3, -4.7)			Week 48 Mean change (range):-7.9 (6.39, -25.8)*	Week 48 Mean change (range): -8.3 (6.93, -26.6)*	Week 48 Mean change (range): -5.3 (6.85, -28.8)

	AURORA 1 **Final follow-up: 1 year**		**AURORA 2** **Final follow-up: 2 years**		**AURA-LV** **Final follow-up: 1 year**		
	MD: -0.5 (-1.4, 0.4)						

Abbreviations: MD, mean difference; VCS, voclosporin

Pairwise meta-analyses

The company presented pooled data for AURORA 1 and AURA-LV for a limited set of outcomes up to 1 year, using participants from AURA-LV who received the low (target) dose of voclosporin. Generally speaking, pooled analyses are preferred data as they draw upon a larger body of evidence, and have greater statistical power for conducting sensitivity analyses. However, few outcomes were considered within the pooled analysis, and therefore these data were considered by the EAG alongside the data from each of the included trials. Subgroup analyses from the pooled data were not provided in the CS, but were included in a confidential document submitted by the company with the CS (Aurinia Pharmaceuticals data on file, 2021).[14]

The results of the pairwise meta-analyses for CRR are shown in Table 13. The results were generally consistent with those reported for the individual trials. Data for change in eGFR and serum creatinine were discussed in a confidential file provided to the EAG by the company, but the tables containing the data were not accessible.

Table 13: Results of the pairwise meta-analyses

	Voclosporin	**Placebo**
CRR	24 weeks: 31.7% OR 2.01 [redacted]^* 52 weeks:43.7% OR 2.76 [redacted]^*	24 weeks: 20.3% 52 weeks: 23.3%
PRR	24 weeks: 70.1% OR 2.42 [redacted]52 weeks: 69.4% OR 2.26 [redacted]	24 weeks: 49.8% 52 weeks: 50.6%
≥50% UPCR reduction	52 weeks: 93.7% Median time to reduction: 29 days HR 1.96 [95% CI [redacted]	52 weeks: 75.2% Median time to reduction: 58 days
UPCR ≤0.5mg/mg	[redacted]	[redacted]
Change in UPCR	[redacted]	[redacted]

Abbreviations: CRR, complete renal response; HR, hazard ratio; NE, not estimable; OR, odds ratio; PRR, partial renal response; UPCR, urine protein/creatinine ratio

^ compared to placebo; *statistically significant at p<.05

Source: CS, clarification response [A17], and additional confidential data provided by the company[14]

3.2.3.2. Subgroup analyses

All subgroup analyses conducted by the company evaluated whether rates of CRR varied across population subgroups. Due to a formatting issue in the CS, at clarification the EAG requested that the company re-submit all subgroup and covariate analyses with their response [A14] to ensure completeness. In addition, the EAG expressed an interest in further subgroup analyses to explore the effect of previous MMF treatment at screening on the treatment effect (for example across additional outcomes, and/or using data from AURORA 2 [A15]). Finally, the EAG requested the company conduct a subgroup analysis limited to centres within Europe [A16]. The company re-submitted the subgroup analyses for CRR from AURORA 1 and AURA-LV, and conducted the requested analysis within European centres. The company did not expand their choice of analyses to explore variation in effect according to MMF use at baseline.

Overall, subgroup analyses showed that participants receiving voclosporin had a greater chance of achieving a CRR than those in the placebo arm across all population subgroups. The EAG noted some variation in the magnitude of effect across groups, though in most cases this was inconclusive, and due to limitations in statistical power the EAG did not draw firm conclusions about variation in effect across these populations. However, the EAG did note that the subgroup analyses appeared to show a smaller effect of voclosporin amongst White participants and those in Europe. There is evidence that people with LN from certain minority ethnic groups have an increased likelihood of having a more aggressive course of LN, which may explain the smaller effect in White trial participants. However, there was no further evidence to consider this further.

In addition, the EAG noted a difference in the magnitude of effect according to whether participants were receiving MMF at baseline in AURORA 1 or AURA-LV. In those receiving MMF at baseline in AURORA 1, [redacted] of those receiving voclosporin achieved a CRR compared to only [redacted] in the placebo arm; however in those not receiving MMF at baseline, rates of response were [redacted] ([redacted] in the voclosporin arm and [redacted] in the placebo arm). However, in AURA-LV, rates of response were greater in the voclosporin arm regardless of MMF use at baseline, and in contrast to the AURORA 1 data, a larger treatment effect was noted amongst those not receiving MMF at baseline. Pooling of the two data points generated a pooled effect consistent with the AURORA 1 findings, but given the unexplained heterogeneity between the two trials, the EAG were concerned about the validity of the pooled estimate. Neither the company nor the EAG were able to explain the conflicting findings. At clarification

[A15], the company suggested the difference in effect between those receiving and not receiving MMF at baseline was due to random variation, and therefore not indicative of a true difference in effect. The EAG accepted that random variation may explain the large difference in effect in both trials, and the conflicting findings between trials, but did not consider that other causes had been satisfactorily explored. For example, as noted in Section 3.2.2.2, the company did not collect data about previous treatments received by participants, and while all participants receiving MMF at baseline were receiving this for the treatment of LN (confirmed by the company to CQ 15), they did not collect information about the length of time they had been receiving it. It was therefore not possible for the EAG to compare whether the trial samples differed in their use of MMF at baseline. Clinical advice to the EAG was that a different magnitude of response might be seen between those who had only recently started MMF, and those who had received MMF for some time and who had not achieved a response or had relapsed. As noted in Section 3.2.2.2, the EAG also considered it plausible that samples differed in characteristics that were unmeasured at baseline, such as those related to disease prognosis. A clinical advisor to the EAG considered it more likely that treatment with voclosporin would have a greater effect at the first-line, as at subsequent lines there may be greater resistance to response in the population. This view may support the findings from AURORA 1, where a greater rate of CRR was seen in those not receiving MMF at baseline who received placebo ([redacted] vs. [redacted] amongst those already receiving MMF), and so explains why the relative benefit of voclosporin was not statistically different. However, the EAG's other advisor did not consider there was yet sufficient evidence to determine why rates of CRR appeared to differ according to MMF use at baseline. Overall, the EAG considered it plausible but uncertain that the magnitude of treatment effect for voclosporin may vary according to the way it is used. This uncertainty is covered by Key Issue 7.

3.2.3.3. Adverse effects

Safety data were presented by the company for each of the included trials within the CS, though rates of serious treatment-related TEAEs were re-submitted by the company at clarification (Section C) due to an error in the CS. The EAG considered that safety data presented for AURORA 1 were the most reliable: safety data from AURORA 2 were considered to be flawed as they do not include participants from AURORA 1 who chose not to continue with the trial; data from AURA-LV were affected by a potential imbalance in disease characteristics and treating centre, which a panel concluded may have contributed to the high mortality rate in the low-dose arm. [12]

The evidence did not show that the addition of voclosporin resulted in an unacceptable rise in safety events: while treatment-related adverse events were reported in the voclosporin arm, there was no difference in the number of serious adverse events. Moreover, while acknowledging the limitations in the AURORA 2 data, treatment–related AEs were comparable between arms by the end of AURORA 2, supporting the company's claim that these events were temporary and/or treatable.

Voclosporin appears to be associated with an increased risk of gastrointestinal and skin disorders, and a higher risk of hypertension, which may be of interest given the increased risk of cardiovascular disorders amongst people with SLE. Notably however, there was no increase in the risk of infections within the trials. As noted in Section 3.2.2.4, the EAG did not consider the follow-up of the trials to be sufficient to conclude whether voclosporin was associated with an increased risk of malignancy. Paradoxically, the EAG noted that voclosporin was associated with an increased risk of a decline in kidney function, including GFR decreases, renal impairment, and proteinuria. This is a known risk associated with prolonged use of CNIs, and clinical advisors to the EAG suggested that people with LN receiving voclosporin should receive similar monitoring for kidney function as those who receive treatment with other CNIs.

Table 14: Key safety data for voclosporin across all included trials

	AURORA 1		AURORA 2		AURA-LV		
	VCS (n=178)	Control (n=178)	VCS (n=116)	Control (n=100)	VCS low dose	VCS high dose+	Control
Any AE	162 (91%)	158 (88.8%)	[redacted]	[redacted]	82 (92.1)	85 (96.6)	75 (85.2)
Any serious AE	37 (20.8)	38 (21.3)	[redacted]	[redacted]	25 (28.1)	22 (25.0)	14 (15.9)
AE leading to discontinuation	20 (11.2)	26 (14.6)	[redacted]	[redacted]	16 (18.0)	14 (15.9)	9 (10.2)
AE leading to dose adjustment	80 (44.9)	47 (26.4)	[redacted]	[redacted]	48 (53.9)	51 (58.0)	28 (31.8)
All cause death	0	3 (1.7)	[redacted]	[redacted]	10 (11.2)	2 (2.3)	1 (1.1)
Treatment-related AE	80 (44.9)	45 (25.3)	[redacted]	[redacted]	45 (50.6)	55 (62.5)	15 (17.0)
Serious treatment-related AE	8 (4.5)	8 (4.5)	[redacted]	[redacted]	4 (4.5)	7 (8.0)	1 (1.1)
Treatment-related AE leading to discontinuation	-	-	-	-	11 (12.4)	8 (9.1)	2 (2.3)
Treatment-related death	0	0	[redacted]	[redacted]	0 (0.0)	0 (0.0)	0 (0.0)
Any infections or infestation	115 (64.6)	101 (56.7)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Any gastrointestinal	83 (46.6)	61 (34.3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GFR decrease	43 (24.2)	15 (8.4)	[redacted]	[redacted]	27 (30.3)	27 (30.7)	12 (13.6)
Renal impairment	13 (7.3)	6 (3.4)	[redacted]	[redacted]	Acute renal failure: 5 (5.6)	Acute renal failure: 8 (9.1)	Acute renal failure: 0 (0.0)
Proteinuria	0 (0.0)	8 (4.5)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

	AURORA 1		AURORA 2		AURA-LV		
Lupus nephritis	2 (1.1)	12 (6.7)	[redacted]	[redacted]	0 (0.0)	1 (1.1)	3 (3.4)
Anaemia	21 (11.8)	10 (5.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Hypertension	36 (20.2)	15 (8.4)	[redacted]	[redacted]	15 (16.9)	16 (18.2)	8 (9.1)
Skin disorders	42 (23.6)	31 (17.4)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Neoplasm	[redacted]	[redacted]	-	-	[redacted]	[redacted]	[redacted]

Abbreviations: AE, adverse event; GFR, glomerular filtration rate; VCS, voclosporin

Source: CS, trial CSRs, and clarification response [Section C]

3.3. Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

The company identified a total of 17 trials to include in their network meta-analyses (NMAs), as well as an additional two trials providing 'non-essential' data on comparators. NMAs focused on CRR and PRR outcomes only, and thus trials not including these outcomes were excluded; moreover, base case NMAs excluded the two trials providing 'non-essential' data, which the EAG judged was appropriate as these comparators were not most relevant to the decision problem. Appraisals of the 19 trials were presented in CS Table B.5-22, in which summary judgments by risk of bias item were tabulated without justification. It is notable that 12 of the 17 key trials did not include blinding of providers, participants or outcome assessors; otherwise, risk of bias domains did not suggest any additional notable threats to validity.

The company undertook an assessment of heterogeneity in included trials. Key features relevant to assessing transitivity in NMAs related to variation in dosages of MMF, which was the reference treatment for all NMAs; six trials with exclusively Asian patients; variable length of follow-up; and outcome definitions for CRR and PRR. The last two points are considered in depth below.

3.3.1. Follow-up times

According to the CS, the longest available follow-up was included in analyses, with a maximum of two years and a modal follow-up time of six months; thus, AURORA-2 was excluded from NMAs (CS document B, p. 84). In the base case, all longest follow-ups were pooled, though it was not clear from the information provided exactly which follow-up points were used in the base case NMA, precluding a clear view as to the inconsistency of follow-up times across networks. This is a potential threat to transitivity if follow-up times are unbalanced over nodes in the evidence networks. A related issue arose from the digitization of curve data from two trials to include in NMAs. The choice of time points for digitization, and how this accounted for censoring where appropriate, created an additional source of ambiguity in the analysis.

3.3.2. Outcome definitions for CRR and PRR

Included trials defined CRR and PRR in a range of ways. As acknowledged in the CS (appendix D, p. 51), though most definitions of CRR included a proteinuria component, the stringency of this component (e.g. proteinuria of <0.5 g/day, or of <0.3 g/day) varied; and more recent trials included eGFR as part of CRR definitions. CRR definitions were tabulated in Table B.5-10. At

clarification, the EAG requested a similar tabulation for PRR definitions; this was presented as clarification Table 20. PRR definitions were considerably heterogeneous, including in the components included; for example, several trials defined PRR as response from baseline (e.g. in UPCR or proteinuria), whereas others defined PRR with respect to specific thresholds (e.g. urinary protein excretion).

While CRR and PRR definitions were broadly consistent within group in considering improvements in renal function, it was not clear that CRR and PRR definitions would be consistent enough to generate measures equivalent between studies in the effectiveness of included comparators. The company asserted in response to CQ A20 that clinical experts were consulted as to the similarity of definitions, and that the company regarded outcome definitions were similar across trials on the basis of inclusion of components such as assessment of proteinuria or UPCR. However, several trials used different combinations of renal function measures to assess PRR, so that even if the component measures included were similar, trials differed in the 'ways' patients could meet effectiveness thresholds.

This is important because it is a threat to transitivity in evidence networks. If a drug would appear more effective under one definition of CRR as compared to another definition but the favourable definition is more prevalent with respect to some nodes in the network as compared to others, then the resultant comparative effectiveness estimates will be biased in favour of the drug meeting an 'easier' threshold for effectiveness. However, the small number of trials relative to the number of nodes precludes any formal or qualitative investigation of this problem.

Relatedly, it is not obvious that CRR and PRR are ordinal outcomes, as might be expected. In response to CQ A23, the company notes that patients achieving CRR are not necessarily subsets of patients achieving PRR. This is a conceptual challenge to interpreting the results of included trials collectively and was reflected in the company's analytic strategy for the NMA.

3.3.3. Similarity of trial populations across the network

A final point relates to the distribution of effect modifiers across the network on the basis of the characteristics of patient populations in the included trials. First, and possibly most importantly, trials in the network include combinations of patients on first, second and third line treatment. This is not explicitly formalised in the table of characteristics for included trials, but it does mean that comparative effectiveness estimates may not be proper to a line of treatment, and if imbalanced over the network, lines of treatment may generate biased estimates of comparative

effectiveness. Moreover, the company acknowledges that a potential source of heterogeneity is the subset of trials enrolling exclusively Asian patients; however, it appears possible, if not likely, that disease characteristics are unequally distributed over the network. Presented in Table B.5-9 (CS Appendix D), the baseline characteristics of patients enrolled in NMA-included trials represent a wide variety of disease characteristics. The range of patients in biopsy class IV ranges from 0% to 100%, with many trials not reporting biopsy results. Demographically, the sex of patient samples ranges from 55% to 100% female. It is unclear how this would influence effectiveness estimates from NMAs.

3.4. Critique of the indirect comparison and/or multiple treatment comparison

3.4.1. Methods used to undertake NMAs

Network meta-analyses (NMAs) were undertaken using standard methods as described in TSD2. [15] CRR and PRR were modelled separately using a logit link, with standard Markov chain Monte Carlo methods implemented using Rstan. The company used generally appropriate and standard statistical methods to estimate both base case and scenario NMAs. Code and data supplied by the company were fully reproducible, and confirmed that the number of iterations used after burn-in was sufficient to achieve convergence for base case NMAs. As noted in Section 3.3.2, CRR and PRR were not regarded by the company to be ordinal outcomes and thus these outcomes were analysed separately. The EAG noted that even if an ordinal model was considered unsuitable, a multivariate NMA might have improved the stability of estimates. The company did not appear to consider this option. Missingness across included trials was also not discussed in sufficient depth to understand how this was addressed.

Fixed effects models and random effects models both used weakly informative priors for treatment effects. Random effects models additionally used an informative prior for between-study standard deviation (half normal with mean 0 and standard deviation 5). At clarification, the EAG questioned the choice of informative prior for between-study standard deviation; in response to CQ A22, the company specified that the source was an example used in TSD2 related to beta blockers, and that further informative priors were not considered. The EAG did not regard this was sufficient justification, especially given the availability of more plausible 'off the shelf' priors (from e.g. Turner (2015)). [16] The company did not present random effects models for base case NMAs, asserting that this was due to lack of convergence. However, this claim was not substantiated with respect to specific model diagnostics, and the EAG could not

trace where and to what degree the company detected evidence of non-convergence. Thus, the EAG presents random effects estimates alongside fixed effects estimates below. This is important as well because the heterogeneity in both NMAs suggests that a random effects model more appropriately reflects the included data.

Consistency checks did not reveal evidence of inconsistency in the PRR NMA; however, the company noted some evidence of inconsistency in the CRR NMA arising from a small trial providing direct evidence of the comparison between MMF and L-CYC. Because of the Bayesian framework used to undertake analyses, consistency was checked by comparing unrestricted mean effects models against the base case estimate. The EAG agreed that the evidence of inconsistency in the CRR NMA was ultimately not consequential enough to invalidate the model, as evidenced by DIC values that were approximately 3 points apart between the fixed effects and unrestricted mean effects models.

The company's critical appraisal of trials included in the NMA identified several issues with the included trials, including: a lack of information about whether appropriate methods for randomisation and allocation concealment were used; imbalance in prognostic factors across trial arms; and analyses not using an ITT approach. These issues are known to affect the reliability of treatment effects.

3.4.2. NMA results

Pairwise odds ratios for each comparator against MMF are presented below, both for the company's fixed effects model and the EAG's random effects model.

Findings from the fixed effects NMA (see Table 15) suggested that voclosporin with MMF is the only treatment statistically superior to MMF in achieving CRR. Pairwise odds ratios suggested that voclosporin with MMF was statistically superior to all comparators with the exception of azathioprine. Unsurprisingly, a random effects model generated substantially wider confidence intervals, though with qualitatively similar point estimates. Voclosporin with MMF was still the only treatment statistically superior to MMF in achieving CRR.

Table 15: Pairwise odds ratios vs MMF for CRR network meta-analysis

	Fixed effects OR (95% CrI)	Random effects OR (95% CrI)
VCS+MMF	[redacted]	[redacted]
AZA	[redacted]	[redacted]
H-CYC	[redacted]	[redacted]

	Fixed effects OR (95% CrI)	Random effects OR (95% CrI)
L-CYC	[redacted]	[redacted]
RTX+MMF	[redacted]	[redacted]
TAC	[redacted]	[redacted]
TAC+MMF	[redacted]	[redacted]
Model fit	Residual deviance 41.8, pD 24.3, DIC 66.1	Residual deviance 39.3, pD 27.7, DIC 67.0

Abbreviations: AZA = azathioprine; CrI = credible Interval; CRR = complete renal response; DIC = deviance information criterion; H-CYC = high-dose cyclophosphamide; L-CYC = low-dose cyclophosphamide; MPR = methylprednisolone; MMF = mycophenolate mofetil; OR = odds ratio; pD = parameters; PR = prednisolone; RTX = rituximab; TAC = tacrolimus; VCS = voclosporin

Source: CS Table B.2-31, EAG calculations

At clarification, the company disclosed that NMAs for PRR were incorrectly estimated due to data extraction errors. The revised estimates, presented in response to CQ A9, are presented below (see Table 16). Only rituximab with MMF was significantly better than MMF at producing PRR outcomes in the fixed effects NMA, with few meaningful differences between the remaining comparators in effectiveness. Unsurprisingly, estimates from the random effects NMA did not suggest any significant differences between any comparators in effectiveness.

Table 16: Pairwise odds ratios vs MMF for PRR network meta-analysis

	Fixed effects OR (95% CrI)	Random effects OR (95% CrI)
VCS+MMF	[redacted]	[redacted]
H-CYC	[redacted]	[redacted]
L-CYC	[redacted]	[redacted]
RTX+MMF	[redacted]	[redacted]
TAC	[redacted]	[redacted]
Model fit	Residual deviance 17.9, pD 15.2, DIC 32.3	Residual deviance 17.9, pD 16.5, DIC 34.4

Abbreviations: AZA = azathioprine; CrI = credible Interval; DIC = deviance information criterion; H-CYC = high-dose cyclophosphamide; L-CYC = low-dose cyclophosphamide; MPR = methylprednisolone; MMF = mycophenolate mofetil; OR = odds ratio; pD = parameters; PR = prednisolone; PRR = partial renal response; RTX = rituximab; TAC = tacrolimus; VCS = voclosporin

Source: Clarification Table 6, EAG calculations

Of note is that for both outcomes, random effects models suggested similar fit as compared to fixed effects models, especially as measured by the deviance information criterion (DIC). One approach would be to state that when two models have similar fit indices, the more parsimonious model should be chosen. However, the EAG regards that based on heterogeneity in outcome definition and follow-up time, there is a strong conceptual basis to prefer a random

effects model; and indeed, TSD3 notes that information criteria alone should not determine choice of model in the face of a conceptual rationale for model choice.

A range of scenario analyses were provided for both CRR and PRR outcomes, including restricting follow-up to six months or 12 months; excluding trials with a significantly different outcome definition; and excluding trials with 100% Asian populations (presented in CS Appendix D.1.1.4.1.9 for CRR, and in clarification responses for PRR). Results were qualitatively similar to base case NMAs.

3.5. Additional work on clinical effectiveness undertaken by the EAG

The EAG reproduced base case NMAs for CRR and PRR outcomes, including scrutiny of model diagnostics and results. The EAG were unable to consider alternative base cases using, for example, informative prior distributions for the between-study variance due to time and resource constraints.

3.6. Conclusions of the clinical effectiveness section

The EAG considered the clinical evidence to demonstrate that treatment with voclosporin + MMF is associated with an increased likelihood of renal response than treatment with MMF alone. There was a lack of reliable data for the effectiveness of tacrolimus + MMF, however evidence from the company's NMA appeared to demonstrate that voclosporin + MMF was more effective for renal response. Evidence from the clinical trials suggested that the addition of voclosporin to MMF did not increase rates of serious adverse events, though prolonged use of voclosporin may carry similar risks to kidney function as other CNIs. Within the trial follow-up, people receiving voclosporin + MMF did not show an improvement in HRQoL compared to those treated with MMF alone. If longer-term evidence demonstrated that voclosporin was associated with a higher rate of sustained response, clinical experts to the EAG considered that improvements in HRQoL may be seen later, following discontinuation from treatment.

There are several limitations with the trial evidence for voclosporin, including a chance but meaningful imbalance in the trial arms of AURA-LV, issues with the selection of participants in AURORA 2, and the lack of statistical power in the trials. While the EAG considered the length of trial follow-up to be acceptable for evaluating renal response, the trials were too short to detect the medium- to long-term implications of treatment, including the impact of treatment on CKD progression, and outcomes following discontinuation from voclosporin. The EAG also highlighted uncertainty about the generalisability of trial evidence to the way voclosporin would

be used in practice (Key Issue 7), and considered that the treatment effect may vary according to variation in the treatment pathway and the duration of treatment. The possibility of effect modification could not be explored within the clinical trials, and this issue was also present in the NMAs.

4. COST-EFFECTIVENESS

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

The company carried out a SLR, using a single search strategy, to identify existing cost-effectiveness evidence, HRQoL evidence, and cost and resource use evidence for voclosporin in LN. A summary of the EAG's critique of the methods implemented by the company to identify relevant evidence is presented in Table 17.

Table 17. Summary of EAG's critique of the methods implemented by the company to identify health economic evidence

Systematic review step	Section of CS in which methods are reported			EAG assessment of robustness of methods
	Cost-effectiveness evidence	HRQoL evidence	Cost and resource use evidence	
Searches	Appendix G	Appendix G	Appendix G	Search strategies by the company were provided in clarification [CQ B1]. The company literature searches were carried out in Proquest which we do not have access to so searches cannot be tested; several databases were searched together in one strategy which is not best practice, it is likely that the strategy may have missed some relevant papers. The cost effectiveness filter that was used does not appear to be a tested filter; [17] this makes the effectiveness of the search uncertain and it is possible that some relevant papers may have been missed. It appears as if the company conducted additional 'targeted' searches for evidence, including data relevant for input into the company model, however the details of these searches were not provided.
Inclusion criteria	Appendix G (Section G.1.1.1.1)	Appendix G (Section G.1.1.1.1)	Appendix G (Section G.1.1.1.1)	Inclusion criteria for the company's SLR were appropriate. Inclusion criteria for any targeted searches conducted by the company were not provided, though the EAG understands this included a search for data on re-transplantations rates (CQ B10) and a search for AE disutility values (though no such data were identified; CQ A5).

Systematic review step	Section of CS in which methods are reported			EAG assessment of robustness of methods
Screening	Appendix G (Section G.1.1.1.2)	Appendix G (Section G.1.1.1.2)	Appendix G (Section G.1.1.1.2)	Screening methods were described in full, and were conducted according to gold standard practice
Data extraction	Appendix G (Section G.1.1.1.3)	Appendix G (Section G.1.1.1.3)	Appendix G (Section G.1.1.1.3)	Data extraction was described in full, and was conducted according to gold standard practice
QA of included studies	Appendix G (Section G.1.1.1.4)	NA	NA	Quality appraisal of economic evaluations reported in full-text publications was conducted using the Drummond checklist, [18] as per best practice. The evidence submitted was consistent with the NICE reference case

Abbreviations: CQ, clarification question; CS, Company Submission; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; NA, not applicable; QA, quality assessment

At clarification stage, the company confirmed that four published cost-effectiveness models and a cumulative cost analysis for LN were identified within its SLR, and that commonalities across these models were used by the company to inform the health states and the decision to build a Markov model to inform this submission (CQ B3). The EAG highlighted that only one of the identified studies considered a comparison of VCS+MMF to MMF, which is discussed further alongside the company's chosen model structure in Section 4.2.2 of this report.

Overall, the EAG was satisfied that the company's health economic SLR was broadly appropriate, and it is unlikely that any cost-effectiveness, cost and resource use, or HRQoL evidence that is directly related to this appraisal was not identified from the searches run. In spite of this, the EAG noted that various sources are used to populate the model that were not identified from the SLR, owing to model's use of data from a non-LN population for various input parameters (e.g., utility values and unit costs). These are discussed in turn in the relevant sub-sections of Section 4.2 of this report.

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

4.2.1. NICE reference case checklist

Table 18: NICE reference case checklist

Attribute	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	✓ No comment
Perspective on costs	NHS and PSS	✓ No comment
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	× The model only presents pairwise analyses not a fully incremental analysis and the EAG has considerable concerns with the chosen model structure
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	✓ No comment
Synthesis of evidence on health effects	Based on systematic review	× Utility obtained from one time point in the AURORA 2 study via mapping, though inappropriate analysis methods used. Dialysis and transplant utilities deemed unsuitable
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.	✓ Health effects expressed as QALYs (although captured from SF-36 mapped to EQ-5D)
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	× The approach taken although informed by patients within the AURORA-2 trial was analysed using methods inappropriate for decision making
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	× Generalisability of data unknown as trial did not have any UK centres
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓ No comment

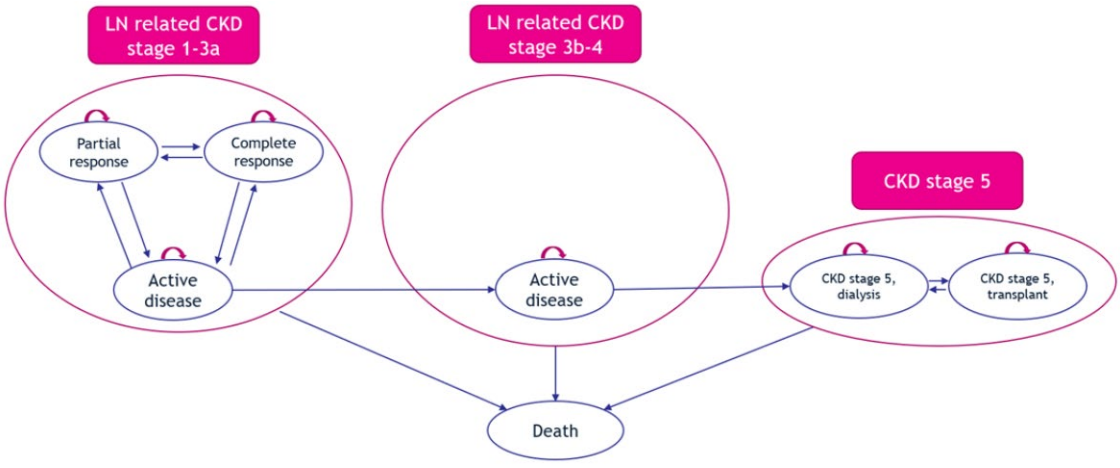
Attribute	Reference case	EAG comment on company's submission
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	✓ No comment
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	✓ No comment

Key: EQ-5D, EuroQol 5 dimension; HRQoL: health-related quality of life; NHS, National Health Service; PSS, Personal Social Services; QALY: quality-adjusted life year; TA: technology appraisal

4.2.2. Model structure

The company developed a *de novo,* cohort-level state-transition Markov model to estimate the cost-effectiveness of voclosporin + MMF (VCS+MMF) versus placebo + MMF (referred to simply as 'MMF' henceforth) in adult patients with LN. A schematic of the submitted model is provided in Figure 1 (replicated based on Figure B.3-1 from the CS with health states removed which are not considered in the model base case).

Figure 1: Cost-effectiveness model structure (company base case)



Source: Adaptation of Figure B.3-1 in the CS, adapted to remove health states not considered in the model base case

Abbreviations: CKD, chronic kidney disease; LN, lupus nephritis

In its submission, the company describes how its cost-effectiveness model structure was informed by previously published models (identified via SLR) due to no previously established NICE guidance concerning people with LN (CS Section B.3.2.2). Although the company states

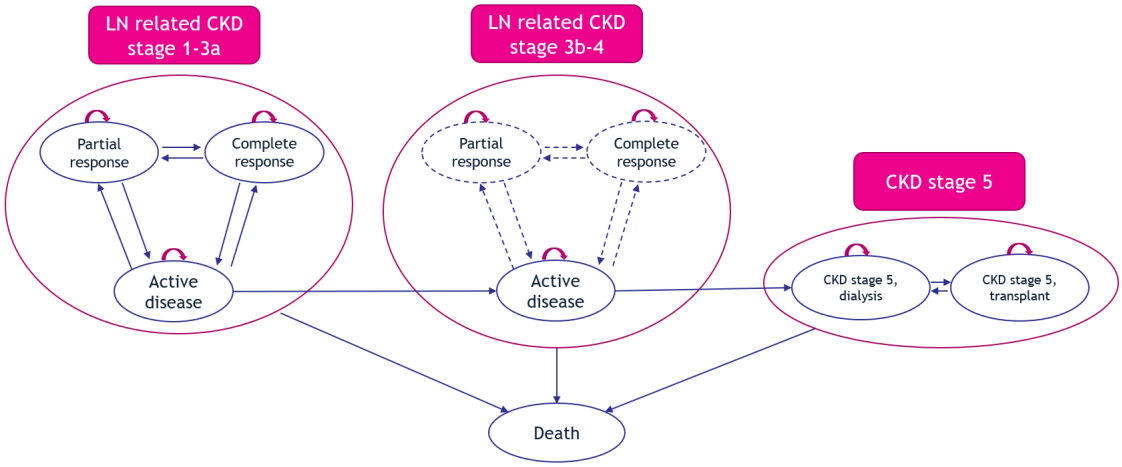
that its model is based on structures identified from studies identified via the SLR, the specific papers are not cited within the CS as to disclose which previously implemented models were used to inform this latest approach.

At clarification, the EAG highlighted the ICER report, to seek justification for the differences in modelling approaches between this paper and the structure used by the company (CQ B3). The company noted that the model in this report did not aptly consider CKD stages, consequently not capturing renal flares. Within a report by the Institute for Clinical and Economic Review (ICER) about its cost-effectiveness analysis of LN treatments, renal flares were included as a parameter and so were explicitly captured within the modelling; although patients receiving belimumab did experience fewer renal flares, the difference between the amount experienced between this arm and the placebo arm was not statistically significant. [19]

Consistent with the expected licence and use for VCS, people are assumed to enter the model in the CKD stage 1-3a active disease (AD) health state. Within CKD stage 1-3a, transitions between partial (renal) response (PR), complete (renal) response (CR) and active disease (AD) health states may occur, with movements between any of these states deemed possible. Importantly, people in either of the response states (i.e., PR or CR) must return to AD before they progress to CKD stage 3b-4 (see Figure 1 for the EAG's edited version of the company's model structure to illustrate non-zero transitions). Here, the EAG highlights the arrow connecting the two AD states, which illustrates that patients must progress through the AD states to move into CKD stage 3b-4.

Although not in the company's base case analysis, it is possible (in terms of model functionality) for patients to move between AD, PR, and CR states within CKD stage 3b-4 (see Figure 2 for the model-permitted transitions, including movements into CR and PR in CKD stage 3b-4 which are set to 0% in the company's base-case analysis and hence enabling this transition has no impact on cost-effectiveness results).

Figure 2: Cost-effectiveness model structure (provided by the company, including scenario analysis health states)



Source: CS Figure B.3-1, Section B.3.2.2.

Abbreviations: CKD, chronic kidney disease; LN, lupus nephritis

Owing to the progressive nature of CKD, the company's base-case analysis does not permit movements from later stages to earlier stages of CKD. CKD stage 5 establishes health states by either use of dialysis or undergoing kidney transplant, in which the company have demonstrated that movements between the two may occur (e.g., patients could undergo transplant but then later require dialysis). People can die while in any model health state.

At clarification stage, the EAG raised concerns with the following features of the company's model structure:

- The recurrent transitions within CKD stage 5 in the model between dialysis and transplant health states given that patients have a 90% probability of receiving a kidney transplant within two years (CS Section B.3.3.2, Table B.3-5) (CQ Question B10)
- The movement (or lack thereof) between various health states within CKD stage 1-3a and stage 3b-4 (CS Section B.3.3.2.2, Table B.3-3) (CQ Question B8)
- The capturing of renal flares within the model (CS Section B.3.2) (CQ Question B4)

The EAG consideration of these features of the model are discussed in the sections that follow.

4.2.2.1. Dialysis and transplant

As described above, the company's model includes the possibility of patients requiring dialysis or undergoing kidney transplantation upon experience of CKD progression to stage 5. Both dialysis and kidney transplant are associated with substantial medical resource use costs, and by extension have important impacts on the overall modelled costs reflected by the company's model.

The company's model includes an estimated probability of undergoing kidney transplant which is equivalent to 90% of patients receiving a kidney transplant within two years from developing stage 5 CKD (estimate obtained from clinical opinion provided to the company). Based on these estimates, the company estimated a per-cycle probability of transplant for patients with CKD stage 5 receiving dialysis of 43.77%. A clinical advisor to the EAG indicated it may be feasible that LN patients could receive transplant more quickly than other patients requiring a transplant as LN patients tend to be relatively younger and fitter, and so would usually be considered more suitable candidates for transplant versus an all-comer population with stage 5 CKD. Despite this, advice from the EAG's clinicians indicated that 90% appeared high, and the EAG were advised that 65% per 2-years may be more reflective of practice. A value of 65% per 2-years translates to a per 6-month cycle rate of 23.08%.

An important feature of the Markov model structure is that it is possible for patients to incur the costs of several kidney transplants, as patients can move between the CKD stage 5 dialysis and transplant states repeatedly. The company noted this within its submission (Table B.3-5) and assumed 2.96% of transplant patients move to dialysis, after which they experience the same probability of transplant (i.e., 43.77% as described above). The EAG considered it unlikely that the transitions between dialysis and transplant in the model are reflective of UK clinical practice, principally owing to the memoryless property of the model as well as the fact that a subsequent transplant is associated with the same probability of occurring versus a first transplant.

The EAG believes that modelling transplants in this way could have been avoided by having a series of sub-models to track (some) event history, which patients could enter upon developing CKD stage 5. This could therefore avoid the 'memoryless' property of the originally imposed Markov model and avoid the possibility that patients may experience multiple transplants. Within the timeframe the EAG had to conduct its review, it was not possible for it to restructure the company's model to explore this further; however, the EAG conducted a sensitivity analysis to

limit patients to only one transplant to ascertain the impact on the ICER of reducing transplant rates in the model (see Section 6.2 for further details).

4.2.2.2. Health states within CKD stages 1-3a and 3b-4

People enter the model within CKD stage 1-3a and may progress from this stage to either death or CKD stage 3b-4. As previously noted, it is crucial to note that movement from PR and CR within the LN related CKD stage 1-3a health state to any sub-state within LN related CKD stage 3b-4 is impossible, i.e., it is only possible to progress to CKD stage 3b-4 if patients have AD due to initial structural decisions made by the company (further discussed in Section 4.2.6). Relatedly, patients cannot achieve a PR or CR from AD CKD stage 3b-4, as these transition probabilities are set to 0% in the company's base-case analysis (given that no patients in AURORA 1 or AURORA 2 developing CKD stage 3b-4 during the period of follow-up).

The EAG received clinical expert advice that it is possible for patients to progress from any health state within CKD stage 1-3a (i.e., AD, PR, or CR) to CKD stage 3b-4, rather than limiting this to only movements from CKD stage 1-3a AD to CKD stage 3b-4 AD. As with the inclusion of non-base case functionality between health states within CKD 3b-4, the EAG believe that it may be useful to include similar capabilities for transitions between all health states, despite the limited data from the AURORA 1 and AURORA 2 trials. After receiving clinical expert advice indicating that patients may be able to progress CKD stage without the presence of AD, the EAG requested justification for the inability to transition between CKD stages (CQ Question B8). The company acknowledged that a person must "*go through a period of disease activity in order for their kidney to accumulate damage*", which is in line with logic regarding how people experience renal flares (CQ Question B8 p.80).

With respect to achieving PR or CR from AD in the CKD stage 3b-4 state, the company chose to use a 'conservative approach' in the model on the basis of feedback from clinical experts that response is rare in patients who reach CKD 3b-4 (CS Section B.3.2.2, p.112). While the EAG acknowledges that there are no data from the AURORA 1 or 2 studies to populate these transitions, the EAG considered it plausible that a PR or CR could theoretically be achieved by patients in either arm, potentially as a result of subsequent therapy use. Therefore, by disabling these transitions, tied with the fact that the PR and CR states have a 'protective' property with respect to CKD progression, it may instead be the case that disabling these transitions introduces a bias in favour of VCS+MMF. However, owing to the paucity of evidence to determine response rates to subsequent therapies in a more advanced CKD population, the

EAG did not explore this feature of the model further, and on balance considered the fact that these transitions are set to 0% in the company's base-case analysis to be reasonable (yet still subject to uncertainty).

4.2.2.3. Capturing renal flares

Renal – and extra-renal – flares are mentioned on several occasions within the CS, included as an outcome specified in the final scope issued by NICE (CS Section B.2.2, Table B.2-1) and equally reported as a key secondary outcome (CS Section B.2.3.2.1, Table B.2-7).

Within the CS, renal flares and extra-renal flares are only reported as an efficacy outcome for the AURORA 2 trial as the follow-up data from AURORA 1 were deemed too short to be considered meaningful (CS Section B.2.6.2). Upon initial inspection of the CS, the EAG could not readily identify precisely how the company's model captures renal flares, therefore the EAG queried how the company captured flares within the model for the avoidance of doubt (CQ Question B4). The company did not clarify whether flares were captured in the model from the AURORA 1 trial; however, justification was provided for how flares were captured.

Based on clinical advice provided to the EAG, renal flares are recognised to be an important aspect of LN, reflecting a key aspect of the natural course of the disease. The CS explains that *"in order to be considered to have experienced a renal flare, patients must first achieve an adequate renal response"*, thus people experiencing renal flares are assumed to be a sub-population of the people with this adequate renal response (CS Section B.2.6.2.3, p.66). A number of patients in AURORA 1 and AURORA 2 were reported to have experienced flares in Table B.2-24 (CS Section B.2.6.2.3). Reporting of flares was limited in both AURORA 1 and AURORA 2, especially given that not all patients from AURORA 1 enrolled in AURORA 2, and so the EAG was unable to fully verify how accurately the company's model captures flares, but considered this an important limitation of the company's model (given the importance of flares in clinical practice).

4.2.3. Population

The population included within the model reflects the population of the AURORA 1 study. Although the company's model classifies patients in terms of CKD stage and renal response, patients must also have been experiencing LN classes III, IV and V or mixed classes of III/V and IV/V to meet the inclusion criteria of AURORA 1. The model does not explicitly capture LN class, but these classes would be expected to be referred to in NHS clinical practice in order for

patients to be deemed suitable candidates for treatment with VCS (in combination with MMF). Owing to the need to capture the downstream costs and effects associated with CKD progression, the EAG considered it appropriate to have not constructed model health state around LN class, but highlights for completeness that LN class is used in clinical practice but is not an explicit feature of the company's model.

Within the CS, the company clarifies that treatment using VCS + MMF should be considered for all active LN patients, *"including patients at initial diagnosis of LN, those with newly flaring disease (previously in remission), and those previously diagnosed but inadequately treated for LN"* (CS Section B.1.3.8, p.26). Patients enrolled in the AURORA 1 study were screened for LN both with and without prior MMF use, and those who experienced successful treatment could progress into the subsequent AURORA 2 follow-on 2-year trial. Approximately 60.5% of patients enrolled in AURORA 2 after completing AURORA 1 (see Section 3.2.2). All patients entered the economic model with CKD stage 1-3a. The model base case was informed using the combined AURORA 1 and AURORA 2 population using data across 36-months. The use of data from both studies is discussed further in Section 4.2.6, and prior use of MMF highlights a key issue for this appraisal concerning the positioning of VCS+MMF in NHS practice (see Section 1.6, Key Issue 7).

4.2.4. Interventions and comparators

The intervention considered, VCS, is described in the CS as being used in combination with background immunosuppressive therapies. At clarification stage, the company confirmed that the licensed indication for VCS will likely restrict background immunosuppressive therapies to MMF specifically, in line with the use of VCS in AURORA 1 and AURORA 2. The cost-effectiveness model considered VCS + MMF as the intervention and as such, the model is therefore aligned with both the AURORA 1 and AURORA 2 trials as well as the anticipated marketing authorisation for VCS.

VCS is administered as 7.9 mg oral tablets (capsules), dispensed in pack sizes of 180. Patients require six capsules daily to achieve a total daily dose of 47.4 mg. Dosing within the cost-effectiveness model is aligned with that of the AURORA 1 and AURORA 2 trials.

In combination with VCS, patients receive MMF (also orally administered). Within the AURORA 1 trial, for patients who had not previously received MMF prior to randomisation, 1 g/day would be administered initially, increasing to 2 g/day starting from day 8. Conversely, for patients who

had been taking MMF prior to the commencement of AURORA-1, a dose of 2g/day was administered. In AURORA 1, 54.9% of patients had experienced prior MMF use at screening. The company's cost-effectiveness model differs from the clinical trial dosing with regard to MMF dosing, as MMF is assumed to be dosed at 2.5g/day irrespective of prior use.

The final scope for the appraisal outlined that several treatments should be considered comparators to VCS:

- MMF
- Cyclophosphamide
- Azathioprine
- Rituximab
- A calcineurin inhibitor + MMF

The CS stated that MMF was regarded as the most commonly used initial therapy, however all comparators listed in the final scope were incorporated into the cost-effectiveness analysis. The company submission compared VCS + MMF with seven comparator regimens:

- MMF
- Low-dose cyclophosphamide
- High-dose cyclophosphamide
- Azathioprine
- Rituximab + MMF
- Tacrolimus + MMF
- Tacrolimus

Clinical advice to the EAG emphasised that MMF was the primary treatment used in current clinical management of LN. Clinicians highlighted that rituximab and tacrolimus are occasionally used if the patient is pregnant or contemplating pregnancy.

To inform the cost-effectiveness analysis, trial data were used to inform several inputs for VCS+MMF and MMF. For other comparators, an NMA was conducted to compare VCS+MMF to other relevant comparators included within the final scope due to a lack of direct evidence for each of these comparators versus VCS+MMF (see Section 3.4).

4.2.5. Perspective, time horizon and discounting

The company's model adopts an NHS and PSS perspective on costs and outcomes, discounted at 3.5% per annum in line with the NICE methods manual. [9] The model output refers to QALYs, LYs and pairwise ICERs for VCS+MMF versus each comparator. Overall, the EAG were satisfied that the perspective adopted, and discounting applied are aligned with the NICE reference case.

The model calculates costs and outcomes over 72 years, which is considered to be a 'lifetime' horizon. The company justify the use of 72 years as based on the extrapolated outcomes, it is the point at which <0.1% of patients are alive. With a mean starting age of patients being 33.2 years (based on the average from the AURORA-1 study), [20] this equates to a maximum age within the model of 105.2 years. The EAG therefore considered a 72-year time horizon to be sufficiently reflective of the lifetime of patients.

The company applied a 6-month cycle length (with a half-cycle correction), justified on the basis of clinical expert advice (CQ Question B8). The company stated at clarification that, in line with clinical expert advice, 6-month cycles were adequate to assess patient response and progression, whilst half-cycle correction accounted for the incidence of events not occurring at the beginning or end of every cycle (CQ Question B8) (CS Section B.3.2.2).

The EAG believe that a 6-month cycle length is suitable for decision making within the company's model but draw attention to two factors that should be considered. Firstly, the duration of treatment effect after the 3-year stopping rule may not be reflected with such long cycle lengths, thus implications of treatment waning may not be correctly gauged (further discussed in Section 4.2.6.3). Secondly, the long length of cycle could potentially mask differences in resource use and treatment costs, which may be inflated in a real-world scenario with a shorter cycle length.

4.2.6. Treatment effectiveness and extrapolation

4.2.6.1. Overview of treatment effectiveness reflected within the model

The company's model captures the impact of treatment through transitions between health states linked with renal response (CR, PR, and AD), as well as CKD stage (1 to 3a, 3b to 4, and 5), details of which are provided in CS Section B.3.3. Of note, the transitions between the renal response health states were derived from data collected in the AURORA 1 and AURORA 2 clinical trials, whereas transitions associated with CKD stage were based on external evidence (i.e., not based on data from AURORA 1 and AURORA 2). No data from the AURA-LV trial were considered in the company model, which was considered appropriate given some of the quality issues associated with this trial (see Section 4.2.6.2).

To facilitate comparisons to other comparators not included in the AURORA 1 and AURORA 2 clinical trials, the company undertook an NMA. A detailed critique of the NMA can be found in Section 3.4 of this report. The company also included within its model assumptions about long-term transitions, both with respect to extrapolation in general and extrapolation of treatment effects. Finally, the company performed time-to-event analyses of treatment discontinuation data to populate its model.

The following sub-sections contain the EAG's critique of these aspects of the company's model.

4.2.6.2. Renal response transitions

All patients enter the model with AD and CKD Stages 1-3a. Then, in terms of renal response, patients can either remain in AD, or achieve either a PR or CR. Transitions up to 36 months were derived from data collected in AURORA 1 (0 to 12 months) and AURORA 2 (12 to 36 months). After 36 months, transitions estimated in the final one or two model cycles were then assumed to be 'carried forward' and applied to later model cycles (discussed further in later parts of this sub-section). Consideration was also given to the possibility of treatment effect waning, described further in Section 4.2.6.3 of this report.

Transitions for the first 36 months were estimated using the 'count method', using data for patients residing in a given health state at the end of each model cycle to then determine movements from the previous cycle. As an example, at baseline all patients in the VCS+MMF arm were in the AD state (n=179 patients).[21] At 6 months, based on information contained within the company's model, there were n=[redacted] patients still evaluable for renal response, of which n=[redacted] achieved and maintained a CR, n=[redacted] achieved and maintained a PR, and n=[redacted]

remained in AD (either following a temporary renal response, or no change in terms of their renal response). Using this information, the transition probability from baseline to 6 months for the movement AD to CR was estimated as [redacted].

Related to the above, the EAG highlights the following excerpt from the CS: *"A transition probability was then generated for each transition within the CKD stages 1-3a by dividing the number of transitions from health state A to health state B by the total number of patients starting in health state A at the beginning of the six-month period."* (CS Section B.3.3.2.1). Here, it is implied that transitions are calculated based on patients being in a given health state at the start of a model cycle. However, instead of this, the model calculates transitions on the basis of patients being in a given health state at the end of a model cycle, which is of particular relevance for the first transition matrix since all patients enter the model in the AD CKD stage 1-3a health state. This is an important distinction to make since some patients can be lost to follow-up part-way through a model cycle.

At clarification stage, the EAG asked the company to provide further information about the approach taken to censoring patients with missing data to inform the 'count' method. In response, the company confirmed that censored observations were essentially removed from the analysis, by subtracting the relevant number of patients with missing data from both the numerator and denominator (company's response to CQ B5). This means that patients are assumed to be missing completely at random (MCAR) and can therefore be effectively removed from the analysis with no adjustment to the resultant transitions other than to re-scale the probabilities so that they sum to one.

The EAG asked the company to provide two alternative analyses to explore the impact of missing data on the transition probabilities, and in particular attempting to account for the potential reasons for the data being missing. These scenarios were:

- To allocate patients with missing data to the health state they last occupied (i.e., a last observation carried forward [LOCF] -type approach)
- To allocate patients with missing data to the AD state (i.e., a 'worst-case scenario' approach)

As the company notes in its response, both of these analyses should be interpreted with caution, since they involve imputing missing data while making explicit assumptions about what the missing data are mostly likely to have been if they were not missing. Furthermore, while the

company provided results for all comparators, the EAG's commentary is limited here to only the comparison of VCS+MMF to MMF since it is not possible to produce 'fair' comparisons to the other treatments given that individual patient-level data are not available to the company nor the EAG for other treatments.

The company notes that censoring affects the MMF arm mostly in the AD state, whereas censoring affects the VCS+MMF arm mostly in the CR and PR states (company's response to CQ B5, Table 22). Therefore, the company explains that the LOCF-type approach is expected to benefit VCS+MMF (i.e., 'carries forward' patients in broadly better response states), whereas the 'worst-case scenario' approach is expected to disadvantage VCS+MMF (i.e., 'forces' more VCS+MMF patients into the worse AD state, relative to the MMF arm). While the EAG is broadly in agreement with the company's view of these exploratory analyses, these interpretations should be viewed as being relative to the company's base-case approach (i.e., an alternative censoring approach may appear to advantage or disadvantage VCS+MMF versus the company's base-case approach, but all three approaches are estimates and are not 'true' data).

A further complication with the 'count method' in addition to determining how to account for censoring is the need to 'switch' from using AURORA 1 data (up to 12 months) to AURORA 2 data (after 12 months). This is challenging since not all patients that were followed up until the end of AURORA 1 continued in/ transferred to the AURORA 2 study. More specifically, taking the VCS+MMF arm as an example, n=162 patients completed AURORA 1 (CS Section B.2.3.1.5.1), and only n=116 entered AURORA 2 (CS Section B.2.3.2.5.1), meaning that n=46 VCS+MMF patients completed 52 weeks of study follow-up in AURORA 1 but did not enrol in AURORA 2. The CS contains information about different reasons that some patients did not enrol in AURORA 2 (CS Section B.2.3.2.5.1), discussed further in Section 3.2.2.1 of this report.

As AURORA 2 does not provide long-term follow-up data for all patients enrolled in AURORA 1, it was necessary for the company to impose an assumption about the impact on transitions for 'removing' the patients effectively lost to follow-up (by virtue of recruiting less than 100% of patients into AURORA 2; or in other words, that 116 ≠ 162). In the model, it is assumed that patients that did not enrol in AURORA 2 could be taken as uninformative censored observations. By extension, this means that data about these patients' long-term renal response outcomes are assumed to be MCAR (i.e., the same rationale of missing data was assumed to apply here as per the assumption made for patients lost to follow-up in general).

Given the above commentary related to the sample size and designs of AURORA 1 and AURORA 2, the EAG has prepared a simple schematic to illustrate the number of patients 'at risk' for specific transitions from a given health state at each model cycle from 0 to 36 months, shown in Figure 3. As can be inferred from this diagram, there is a large proportion of patients considered 'missing' from AURORA 2 (either due to censoring, death, or non-enrolment from AURORA 1 to AURORA 2) when making the switch from AURORA 1 to AURORA 2 data in the company's model. It may also be speculated that the proportion of patients in AD at the end of AURORA 1 was greater than the proportion of patients in AD that entered AURORA 2 [redacted].

Figure 3: Patients by renal response category over time

Abbreviations: AD, active disease; CR, complete (renal) response; MMF, mycophenolate mofetil; PR, partial (renal) response; VCS, voclosporin.

Note: This figure presents data for all AURORA 1 patients until 12 months, and then all AURORA 2 patients from 12 to 36 months. 'Missing' refers to a patient no longer being considered in either CR, PR, or AD for any reason (including death, loss to follow-up, not enrolling in AURORA 2, etc.).

It is the EAG's view that the approach taken to censoring patients from AURORA 1 who did not enrol in AURORA 2 has the potential to have led to overly optimistic estimates of transition probabilities between 12 and 36 months (and therefore, by consequence, for the remainder of the modelled time horizon). This is because these latter transitions are based only on AURORA 2 patients, and patients who did not enrol in AURORA 2 may be more likely to either continue with AD (if they were in AD at the end of AURORA 1) or 'lose' their renal response (by virtue of having VCS+MMF or MMF either at 12 months [when completing the study], or prior to 12 months [if they discontinued treatment prematurely]). The company did not provide alternative analyses to account for this aspect of censoring as part of the EAG's request for clarification about the overall approach taken to account for missing data (CQ B5).

An alternative statistical analysis approach was noted in the CS, with further details provided by the company at clarification stage. In summary, the company explained that a multinomial logit model was considered to derive transition probabilities, but found that model fit was poor, both in terms of reflecting the available trial data and the direction of transitions over time (company's response to CQ B6). In response, the company explained that there were issues with model fit in this analysis, and in particular that *"…the health state distribution of patients for VCS + MMF did not reflect the trial data whatsoever…"* (company's response to CQ B6). The EAG agreed with the company's general summary of the model fit being poor and therefore not useful to inform the model, though the EAG would have ideally preferred the company to elaborate further as to the reason(s) why the fit was so poor. The EAG speculated that the poor fit was likely caused by a small number of patients at risk for each transition over time.

For completeness, the EAG observes that the company provided a scenario analysis in which transitions were estimated using data from AURORA 2 only. While this analysis avoids the issue relating to 'switching' from AURORA 1 to AURORA 2 data, there is a clear issue with this approach in that randomisation is not only broken, but the comparability of the two groups is determined on the basis of a measure taken post-baseline (more specifically, at 12 months). Patients only entered AURORA 2 if they completed 12 months of treatment as part of AURORA 1), and so it would therefore be expected that transitions from 0 to 12 months based on AURORA 2 data only would appear more favourable (in terms of achieving PR or CR, for either treatment arm) versus including data for all AURORA 1 patients. Consequently, the EAG does not consider this scenario analysis further.

Based on the structure of the company's model, achieving either a PR or CR is associated with a 'protective' property in terms of CKD progression – in other words, patients can only progress from CKD stage 1-3a to CKD stage 3b-4 if they have AD, whereas patients with PR or CR must first 'lose' their renal response before being eligible to transition from CKD stage 1-3a to CKD stage 3b-4. This feature of the company's model is especially noteworthy given the specification of a relatively long model cycle length of 6 months. In theory, a hypothetical patient could 'lose' their renal response at any time within a 6-month period but would only be subject to the risk of CKD progression in the next cycle (i.e., it is not possible within the company's base case model for a patient with renal response to experience CKD deterioration to stage 3b-4 without first moving to AD, and so this takes place over a 12-month period [PR/CR CKD stage 1-3a $\xrightarrow[6\ months]{}$ AD CKD stage 1-3a $\xrightarrow[6\ months]{}$ CKD stage 3b-4]).

At clarification stage, the company explained that this feature of the model was designed to reflect cumulative kidney damage associated with LN, and that during the natural course of disease, patients with LN transition to AD after experiencing a relapse (i.e., renal flare), and it can take some time for the flare to manifest in irreparable kidney damage i.e., progression of CKD (company's response to CQ B8). The company clarified that analyses in which CKD progression are permitted from the CR and PR health states are *"not appropriate"* and that *"clinical experts have verified the assumption that requires patients to first enter and spend some time in AD CKD* [stage] *1-3a before transitioning to AD in CKD* [stage] *3b-4"* (company's response to CQ B8). Therefore, such analyses were not provided as part of the company's response to this request.

Contrary to the view expressed by the company in its response to the aforementioned CQ, based on clinical opinion provided to the EAG it is expected that some patients could experience CKD progression outside of experiencing renal flare (i.e., it is entirely possible for patients with CR or PR to experience CKD progression outside of a renal flare). The EAG acknowledges that all models represent a simplification of reality, and so the decision to only allow CKD progression to occur from the AD state in CKD stage 1-3a may be reasonable, yet there is no other evidence provided in the CS to further substantiate this structural feature of the model. As the requested sensitivity analyses to explore this further within the company's model were not provided (per the EAG's request in CQ B8), and it is beyond the remit of the EAG to re-structure the company's model to permit such transitions, the EAG highlights this as a limitation of the company's model structure, and the impact of this restriction of the model structure on results is unclear.

As described previously, from 36 months, transitions that were estimated for the previous one or two model cycles were assumed to be 'carried forward' and applied to later model cycles. In the base-case analysis, the 'average' transitions from 24 to 30 months and 30 to 36 months were assumed to serve as the renal response transitions for the remainder of the modelled time horizon. However, limited explanation concerning the calculation of these long-term transition probabilities is provided in the CS, though calculations were clearly presented in the company's model in order to understand how they were estimated.

Let us consider the transition AD CKD stage 1-3a to CR CKD stage 1-3a. For the VCS+MMF arm, the transition probability applied in the base-case analysis for 24 to 30 months is [redacted] (call this x_{24}) and for 30 to 36 months is [redacted] (call this x_{30}). The transitions estimated for 24 to 30

months were based on a sample of n=[redacted] patients residing in the AD CKD stage 1-3a health state (call this n_{24}), which decreased to n=[redacted] patients for 30 to 36 months (call this n_{30}). Therefore, in the base-case analysis, the transition probability estimated to apply from 36 months (call this x_{36+}) is calculated as follows:

$$x_{36+} = \frac{(x_{24} \times n_{24}) + (x_{30} \times n_{30})}{(n_{24} + n_{30})}$$

Or

[redacted]

In this worked example, x_{36+} is estimated to be [redacted]. While the EAG raises no issues with the calculation approach to obtain these 'average' transitions, the approach in general is heavily reliant on small numbers of patients still considered to be 'at risk' for a given set of transitions. Taking the example above, for 24 to 30 months n=[redacted] of n=[redacted] patients moved from AD CKD stage 1-3a to CR CKD stage 1-3a*, but for 30 to 36 months n=[redacted] of n=[redacted] patients experienced the same transition. The EAG noted that in response to CQ B14, the company explained that patients *"tend to respond quickly with* [VCS] *treatment"*, [redacted]. Therefore, the EAG highlights that the long-term transitions included within the model are subject to substantial uncertainty and appear to lack a degree of face validity in terms of how renal response is likely to be achieved in the long term after cessation of treatment with VCS+MMF or MMF.

The EAG has undertaken further exploratory analyses concerning the duration of treatment effect, and how this impacts transitions within the model. Further details of these analyses are provided within Section 6 of this report, and additional discussion pertaining to the expected duration of treatment effect is contained within Section 4.2.6.3 of this report.

* Note: the 'final' transition probability of [redacted] is not equal to [redacted] as the model also accounts for the risk of death. Mortality model inputs and calculations are discussed further in Section 4.2.6.7 of this report.

4.2.6.3. Treatment efficacy waning

The company describes within its submission how *"uncertainty related to any sustained efficacy following treatment discontinuation…"* was *"… accounted for by applying a long-term treatment waning effect to* [VCS + MMF] *and all comparators"* (CS Section B.3.3.2.1, p.116). Further detail concerning the application of treatment efficacy waning is provided in Table B.3-2 in the CS. In summary, the model assumes that when all patients permanently discontinue VCS + MMF (assumed to be 36 months in the base-case analysis), transition probabilities 'wane' to reflect an average of the estimated probabilities for the last two model cycles across both treatment arms from AURORA 2 (i.e., VCS + MMF versus MMF). The EAG noted that this application is based on transition probabilities from VCS+MMF and MMF whilst patients remain on treatment and does not capture what happens to patients who discontinue treatment on either arm. These transition probabilities based on patients receiving treatment are applied for the remainder of the time horizon (i.e., from 36 months to 72 years).

Second to this, the EAG also notes that while this aspect of the model transitions reflects *some* loss of treatment effect from 36 months, it should not be mistaken as an assumption of loss of *all* treatment effect (since some residual treatment effect is maintained from 36 months). In the context of the model, here 'loss of treatment effect' refers to the difference between treatment arms for transitions between PR, CR, and AD after cessation of treatment (at 36 months in the company's base-case analysis). The company's base-case approach to capturing long-term treatment effect means that patients that received VCS+MMF are associated with 'better' transition probabilities for the remainder of their lifetime – for example, a lower risk of losing their renal response. At clarification stage, the EAG asked the company to provide further information concerning the application of treatment effect, and to provide scenarios such that the model can reflect partial, total, or no treatment effect waning (clarification question B7).

In response, the company explained that any loss of treatment effect is *"unlikely to occur instantaneously following treatment discontinuation"*, as well as adding that it was unaware of any data or studies concerning treatment waning effects in an LN population (company's response to CQ B7). The company did not provide any of the requested sensitivity analyses concerning differential approaches to capturing potential treatment waning effects within its model.

The EAG highlights a study by Jourde-Chiche *et al.*, (2022)[11] which reports findings from the WIN-Lupus trial: a multicentre RCT investigating weaning of maintenance immunosuppressive

therapy in LN. While the EAG acknowledges that this study was published after the company made its submission, the study provides some evidence related to the waning of treatment effects over time for immunosuppressive therapies in an LN population. Acknowledging a number of limitations of this study and its direct relevance to this appraisal (different treatments, non-inferiority study, limited sample size of n=88 patients, amongst others), the EAG highlights the following conclusion reached by the authors of this study: *"*[Immunosuppressive therapy] *discontinuation was associated with a higher risk of severe* [systemic lupus erythematosus] *flares (renal or extra-renal) requiring induction* [immunosuppressive therapy]*"* (Jourde-Chiche *et al.*, [2022], p.4).[11] The EAG therefore considered it entirely possible that the effect of treatment (with either VCS+MMF or MMF alone) could indeed wane over time, and that there is no guarantee that it would persist over a lifetime.

The EAG contends that there are different ways one could hypothesise about the long-term effect of VCS+MMF treatment, relative to MMF alone, after discontinuation. Of note, the EAG highlights the importance of separating two distinct concepts:

- loss of effect in terms of assuming an immediate loss of renal response
- loss of effect in terms of assuming no further difference in gaining or losing renal response over time

The first concept above is not reflected by the model, which the EAG agrees is appropriate and does not advocate any immediate reversal of renal response upon discontinuation of treatment effect. However, the second concept is *partially* reflected by the model, but the EAG maintains its view expressed at clarification stage that scenarios reflecting 'no' waning or 'full' waning may be suitable scenarios to consider within decision making.

A related issue with respect to long-term treatment effect is that by carrying forward transition probabilities after patients have discontinued treatment, patients can continue to achieve a renal response. The EAG concedes that this may be possible in reality due to the use of subsequent therapies (costs for which are captured within the model). However, the effects of subsequent therapies are not explicitly captured within the model, and there is no guarantee that subsequent therapy would yield 'similar' response rates to those implied by the latter transition matrices estimated from the AURORA 2 trial data. This is especially important to consider in light of the fact that subsequent therapy use is not directly linked to renal response health state within the company's model.

Overall, the EAG considered the duration of treatment effect, and the method most appropriate to reflect this, as a key area of uncertainty inherent within the company's model and has therefore conducted additional exploratory analyses to investigate this further. Details of the scenarios undertaken, and associated results are provided in Section 6.2 of this report.

4.2.6.4. Indirect treatment comparison

A full critique of the company's network meta-analysis (NMA) is provided in Section 3.4 of this report, but here the EAG focuses on the application of the NMA of odds ratios (ORs) comparing the probability of transitioning from AD CKD stage 1-3a to either PR or CR (separate ORs for each transition). As discussed in Section 3.4 of this report, the EAG would have preferred an analysis that used random effects, likely with a better choice of informative prior to improve the credibility of estimates. Though the EAG present estimates from a random effects NMA, the fixed effects NMA is used pending resolution of questions about optimal estimation of the NMA.

Overall, the EAG has no major concerns with the application of the NMA outputs within the company's model but highlights two relatively minor points for completeness below.

The first of these pertains to the comparison of VCS+MMF versus MMF. In the company's model, while ORs are presented to compare MMF with VCS+MMF, these do not inform any model calculations. Instead, transitions for the VCS+MMF arm are based solely on the patient-level data from the AURORA 1 and AURORA 2 studies. While the EAG considered this approach to be sensible in light of the availability of individual patient-level data for both arms, it is evident that transitions for VCS+MMF would be different if the ORs were used to derives transitions instead of estimating transitions from the AURORA 1 and AURORA 2 trial data, if only because the OR provides a summary measure assuming a time-invariant difference in transitions over the course of the modelled time horizon.

The second point relates to a specific transition probability for the MMF arm, which serves as the baseline from which the ORs are applied. Over the time period 18 to 24 months, n= patients in the MMF arm transitioned from AD CKD stage 1-3a to CR CKD stage 1-3a. [redacted]. Similar to aspects of the model highlighted earlier in this report, this is another example of where the model calculations are adversely affected by the number of patients at risk of a given transition at a given time point across the 36 month-period over which data are available from AURORA 1 and AURORA 2. However, the EAG noted that this specific issue affects only one transition at one model cycle.

4.2.6.5. Time to treatment discontinuation and stopping rule

To inform treatment discontinuation rates within the model, the company undertook parametric survival analyses of time to treatment discontinuation (TTD) data for VCS+MMF and MMF. Additional information was provided by the company at clarification stage, at the request of the EAG, concerning the overall approach taken, provision of supporting plots, statistical goodness-of-fit scores, and justification of the base-case model selected (CQs B11 and B12).

Initially, the EAG was unsure why a parametric model was fitted to these data, given that all patients are subjected to a stopping rule at 36 months, and that trial follow-up was sufficient to allow estimation of drug costs without needing to fit a parametric model. In response to the EAG's request for further information, the company explained the benefit of parametric survival models providing a 'smooth' curve, versus a stepped Kaplan-Meier estimate. The EAG accepts that a parametric model has this advantage, but notes that a sensitivity analysis using the Kaplan-Meier estimate may have also been helpful to consider for completeness.

The company selected a log-logistic model, based on visual fit, statistical goodness-of-fit scores, and consideration of proportional hazards. The EAG considered the choice of a log-logistic model to be acceptable, though asked the company to provide alternative analyses at clarification stage for completeness. The company provided additional results with four alternative parameterisations (exponential, generalised gamma, lognormal, and Weibull), which had very little impact on the company's base-case ICER. As such, the company's base-case approach was deemed acceptable and is maintained in the EAG's preferred analysis.

For all comparators within the model except for MMF, a TTD curve was not considered, and all patients were assumed to receive treatment until they stopped treatment (i.e., there was no treatment discontinuation for any proportion of patients for reasons such as adverse events or lack of efficacy). The EAG considered this to be a very limited analysis which will likely overestimate costs associated with all other comparators (except MMF). As such an exploratory scenario analysis is considered by the EAG (outlined in Section 6.2) which assumes that all comparators follow the same TTD curve as the observed MMF data.

In the company's base-case analysis, patients are assumed to receive treatment with either VCS+MMF or MMF until 36 months. The stopping rule of 36 months was chosen by the company on the basis of the AURORA 1 and AURORA 2 trials providing follow-up until this point in time and based on clinical expert opinion provided to the company. Independent clinical expert feedback provided to the EAG suggested that 36 months was likely a suitable stopping

rule in most cases. However, the EAG was also advised about the heterogeneous nature of LN and how different patients respond to treatment, meaning that some patients may discontinue earlier than 36 months, or potentially (if permitted according to both the marketing authorisation for VCS and the recommendation reached by NICE) could be treated beyond 36 months. The draft SmPC included within the CS states with respect to treatment duration: "[redacted]" (CS, Appendix C, p.1). Consequently, the EAG highlights that there is currently no restriction made within the SmPC that would limit the maximum duration of treatment with VCS+MMF to 36 months. At clarification stage, the EAG asked the company to confirm the rationale behind the 36-month stopping rule, and in response it reaffirmed that this was in keeping with the available data from AURORA 1 and AURORA 2, as well as clinical expert opinion it received.

Taking into consideration the position of the company (informed by both trial data and advice from its clinical experts), as well as the views expressed by the EAG's clinical advisers, the EAG tentatively adopts a 36-month stopping rule to inform its preferred analysis. The EAG considered treatment duration as a key issue with regard to positioning for voclosporin+MMF and the long-term efficacy assumptions related to a 36-month stopping rule (see Key Issue 3 and Key Issue 7), in so far as an imposed stopping rule constitutes a restriction of the use of VCS+MMF in practice relative to its licensed indication. However, this restriction may also mean that some patients would need to discontinue treatment at 36-months who may have otherwise continued in the absence of a stopping rule. Further to this, for MMF and the analysis of the data from AURORA 1 and 2, and inclusion within the model, the imposed stopping rule by the company does not take into consideration the duration of therapy for patients who were receiving MMF prior to entering the trial, again bringing into question the appropriate positioning of MMF for cost-effectiveness estimates.

As a scenario analysis, the company included within its model the ability to impose an early stopping rule at 18 months. This earlier stopping rule was explored as a result of findings from a US-based survey of 96 clinicians, which was contained within feedback provided as part of ICER's independent assessments of VCS and belimumab for the treatment of LN. For complete context, the exact quote included in the feedback is as follows: *"Underpinning this, a survey of 96 treating U.S. physicians suggests that the majority would keep patients on treatment for no more than 1.5 years after achieving a complete renal response"* (Aurinia Comments on the Institute for Clinical and Economic Review's Draft Evidence Report, p.1-2).[22] The EAG highlights

that the time point of 18 months refers to treatment after achieving a CR, not from initiation of treatment prior to response.

The EAG highlights that the 18-month stopping rule scenario comprises (i) a simple 'cap' on the treatment duration curve used within the company's model to affect costs, and (ii) use of transitions up until 18 months (with the 'carrying forward' approach of the base-case analysis applied relatively earlier). While this scenario happens to yield [redacted] QALYs compared with the company's base-case analysis, the EAG investigated results accounting for all possible ranges of treatment duration from 12 months to 36 months, in 6-month intervals, to investigate the relationship between outcomes and duration of treatment in the company's model.

The results of this analysis are presented in Figure 4. As can be inferred from this figure, there is a non-linear relationship between the imposed stopping rule and the total QALYs estimated by the model. In reality, it would be expected that a longer duration of treatment should yield increasing QALYs, and so this analysis sheds further light of the overall uncertainty in the company's model with respect to transitions based on limited trial data, and how these impact on lifetime estimates of QALYs (including the approach taken to account for treatment waning, as discussed in Section 4.2.6.3 of this report) over the modelled time horizon.

Figure 4: Total QALYs by treatment arm based on duration of treatment

Abbreviations: MMF, mycophenolate mofetil; QALY, quality-adjusted life year; VCS, voclosporin.

4.2.6.6. Mortality

In the company's model, background mortality was applied for all health states before any other transitions between health states occurred, which was independent of health state occupancy. However, additional mortality risk was included for specific health states in the model at all model cycles for the following health states:

- **AD CKD stage 1-3a:** Based on patient-level data for the MMF arm from AURORA 1 and AURORA 2, a probability of [redacted] per 6-month model cycle was estimated on the basis of [redacted] deaths being recorded over a total of [redacted] 'at risk' periods of 6 months
- **AD CKD stage 3b-4:** Based on a study by Sugrue *et al.*, (2019), [23] a probability of 3.92% per 6-month model cycle was estimated. The population included in the study by Sugrue *et al.* reflected a broader CKD population, and so this estimated mortality risk may be higher than the 'true' value expected for a relatively younger LN population
- **CKD stage 5 (dialysis):** Based on Sugrue *et al.*, (2019), [23] 7.47% per model cycle
- **CKD stage 5 (transplant):** Based on Sugrue *et al.*, (2019), [23] 2.62% per model cycle

No excess mortality risk was applied for patients residing in either the PR or CR health states for every model cycle. However, for specific model cycles, some specific transition matrices included non-zero probabilities for death for the PR and CR health states, based on the count method. Owing to the low number of deaths that occurred during follow-up in the AURORA 1 and AURORA 2 studies, these mortality risks are small.

The EAG considered the company's overall approach to incorporating mortality within the model to be appropriate but given the small number of observed deaths in the AURORA 1 and AURORA 2 studies, it is important to acknowledge that the incorporation of these deaths within the first 36 months of follow-up in the model can have a marked impact on results due to sample size of AURORA 1 and AURORA 2. For example, in the MMF arm [redacted] with PR CKD stage 1-3a died between 24- and 30-months, and because [redacted] were still at risk at this time, this ultimately translated to a [redacted] probability of death in this cycle specifically, versus [redacted] for the VCS+MMF arm [redacted]. Moreover, the EAG considered it counter-intuitive that patients can die of their disease from the PR or CR states (i.e., there is at

least one non-zero transition from either the PR or CR CKD stage 1-3a state to Dead, on at least one treatment arm), but cannot experience CKD progression from this same state.

A further issue with the application of mortality data from the AURORA 1 and AURORA 2 studies is that deaths from CR and PR are factored into the model using a different approach versus deaths from AD (all in CKD stages 1-3a). Deaths from CR or PR are both time-varying and arm-specific (i.e., could be different values for each model cycle, and can be different for VCS+MMF versus MMF), whereas deaths from AD are applied based on count data from AURORA 1 and 2 varying over time (and arm-specific) with a further additional constant over time independent of treatment arm (i.e., one transition probability is applied to both arms, across all model cycles).

The EAG considers the description of how mortality is captured within the company's model via the CS to be somewhat misleading, as the application of mortality risk for AD CKD stages 1-3a is described very briefly, and within a sub-section titled: "Transitions between AD CKD 1-3a and AD CKD 3b–4". The relevant text in the CS states: *"In addition, the transition probability from AD CKD 1-3a to death could be informed by mortality data collected in the MMF arm in AURORA 1 and AURORA 2 ([redacted]% per 6-month cycle")"* (CS, Section B.3.3.2.2, p.116-117). Here, the CS acknowledges that the same probability is applied by treatment arm but is based only on data collected for the MMF arm. No explanation is provided for why this specific transition probability was estimated only using data for the MMF arm, nor is it clear why this particular probability was necessary to consider fixed over time.

Given the small number of deaths that occurred during the follow-up period of AURORA 1 and AURORA 2, the EAG prefers the approach taken to capture deaths from AD over the time-varying/ arm-specific approach taken for PR and CR deaths, but both approaches are subject to both misinterpretation given both the description provided within the model and uncertainty given the small numbers of events within the trial. As such, the impact of LN deaths is explored further as part of the EAG's exploratory analysis.

4.2.6.7. CKD progression transitions

In its base-case analysis, the company assumes that progression from CKD stage 1-3a to CKD stage 3b-4 is not possible in the first 36 months of the model (i.e., 3 years). This is justified by the company in its submission on the basis of no patients in AURORA 2 progressing to CKD stage 3b-4 over the course of three years of follow-up. The EAG noted that while it is correct that no patients in AURORA 2 experienced progression to CKD stage 3b-4, it is clinically

plausible that patients *could* progress to CKD stage 3b-4 in the first 3 years of the model, and arguably the most likely patients to progress within the first 3 years of treatment would be those patients that discontinued treatment before 12 months, or those that completed AURORA 1 but did not enrol in AURORA 2 (including, 13 patients (of n=55) who discontinued due to a lack of efficacy – see responses to CQ A10). In addition, the EAG noted that 54.9% of the AURORA 1 population were already receiving MMF at screening, and so may be considered to have been 'at risk' for CKD progression prior to enrolment (though if they had already progressed to CKD stage 3b-4, would not have met the inclusion criteria of the study). This restriction within the company model structure emphasises why the positioning of voclosporin within the treatment pathway is a key consideration and noteworthy to the EAG (see Key Issue 7).

In light of the considerations above, while the EAG acknowledges the rationale behind disabling this transition for the first 3 years of the model, it expects that in reality this transition should be considered possible, as there is no biological basis from which to assume CKD progression cannot occur until after 3 years of treatment (with any regimen included within the model). This transition is therefore permitted within the EAG's preferred analysis, presented in Section 6 of this report.

After 3 years, patients are permitted to experience progression from CKD stage 1-3a to CKD stage 3b-4. As discussed previously, the company's model includes a 'protective' property linked with renal response, such that patients cannot progress to CKD stage 3b-4 unless they currently reside in AD. This means that by extension, VCS+MMF is associated with an indirect benefit in terms of CKD progression through keeping patients in either PR or CR for longer versus MMF. Given the irreversible nature of CKD progression, this indirect treatment effect constitutes an important assumption within the company's model. Advice from clinical experts to the EAG indicated that it may be possible for patients to progress CKD stage whilst still maintaining renal response.

For patients with AD, the risk of progressing to CKD stage 3b-4 is fixed at 3.05% per 6-monthly model cycle. This value was estimated on the basis of clinical expert opinion that approximately 6% of patients in CKD stage 1-3a will progress to CKD stage 3b-4 per year (CS Table B.3 3). The EAG noted that this probability was not estimated using empirical evidence, but rather was derived from clinical expert opinion (with further details about elicitation of this opinion not clear from the CS), and so it is subject to uncertainty. Nevertheless, the EAG understands that there is a paucity of evidence available concerning the long-term disease progression for patients with

LN (confirmed also by the fact that the EAG was also unable to identify relevant evidence to inform CKD progression rates within the company's model for an LN population), and so recognises the need to rely on experts to populate these aspects of the model. However, the EAG's principal concern relating to this transition is about the 'protective' property of renal response in the model with respect to CKD progression.

Once patients have progressed to CKD stage 3b-4, a risk of progressing to CKD stage 5 is included within the model. CKD stage 5 is separated by 'Dialysis' and 'Transplant', with patients initially moving to 'Dialysis' from CKD Stage 3b-4. The probability of moving into 'Dialysis' from CKD stage 3b-4 is fixed at 13.91% per 6-month model cycle, applied across both arms equally, based on clinical opinion provided to the company (CS Table B.3-4). Similar to the EAG's commentary concerning movements from AD CKD stage 1-3a to AD CKD stage 3b-4, there is a paucity of evidence to inform this latter aspect of the model, yet it is clear that this transition probability is subject to substantial uncertainty.

Of greater concern, however, are transitions between 'Dialysis' and 'Transplant'. The probabilities applied in the model for these transitions are as follows:

- From 'Dialysis' to 'Transplant': 43.77% (based on clinical opinion)
- From 'Transplant' to 'Dialysis': 2.96% (based on Palmer *et al.*, [2004]) [24]
- From 'Dialysis' to 'Dead': 7.47% (based on Sugrue *et al.*, [2019]) [23]
- From 'Transplant' to 'Dead': 2.62% (based on Sugrue *et al.*, [2019]) [23]

Taking these probabilities together, it is possible to track over a given time horizon how many transplants would be modelled for a hypothetical cohort of patients starting in the 'Dialysis' health state. Taking a 10-year time horizon as an example, the average patient starting in 'Dialysis' would be modelled as receiving 1.12 transplants over 10 years, and slightly more than half (50.5%) of the starting cohort would be modelled to have died by 10 years. While these estimates are hypothetical, the EAG considered it important to acknowledge that due to the memoryless property of the company's Markovian model, many surviving patients will undergo at least one transplant, and a notable proportion will have two or more transplants over their lifetime. Overall, the EAG has concerns with the face validity of the estimated number of transplants that occur within the company's model, owing mostly to the specification of a time-

invariant probability of transplant occurring from the dialysis health state, and a lack of consideration of event history when considering eligibility for re-transplantation.

Ideally, the EAG would have preferred the company's model structure to introduce an element of memory to better account for the probability of additional transplants, and potentially adjust the subsequent chance of transplant success or failure. In lieu of a model structure that explicitly captures differences in outcomes based on re-transplantation rates, the EAG has conducted an exploratory analysis to ascertain the impact on results if re-transplantation is disabled within the company's model (see Section 6.2 of this report).

4.2.7. Health-related quality of life

4.2.7.1. Overview of HRQoL within the model and EAG critique

The AURORA 1 and AURORA 2 trials included both the SF-36 (v2) and the LupusPRO (v.1.7) disease-specific measure. In order to generate health state utilities, the company used a mapping from the SF-36 to generate EQ-5D utilities (Rowen et al., 2009).[25] Given the EQ-5D was not directly measured this does provide reference case utilities, albeit with uncertainty inherent through the use of a mapping algorithm. In response to CQ B20, the company confirmed that SF-6D utilities[26] had not been generated, and thus there remains uncertainty regarding the validity of the mapping in this patient population as the mapping was derived in different disease areas, and may not reflect the specific issues faced by people with LN. The company however did provide plots comparing the LupusPRO and mapped EQ-5D utility (CQ B21), which appear to support the mapped EQ-5D values reflecting the patient experience according to the disease specific measure.

Although the company used data collected from the AURORA trial programme to populate the model, the approach used to estimate health state utilities for use in the economic model is methodologically wrong, certainly biased, with the resulting values unreliable for decision making. The EAG has used the values provided by the company in some instances due to the lack of other values in the company submission, however the EAG has substantial reservations regarding the conduct of the utility analysis, and consequently the robustness of the utility values provided.

To populate the model, a mix of trial data and data from the literature was used. For CR/PR/AD in CKD stages 1-3a, the approach taken by the company was to use only the utility values from AURORA 2 observed in Month 36. These were split by patients in CR, PR and AD, and the

mean utility in each of the groups used (taking values of 0.83, 0.80, and 0.71, respectively). For values not available from the AURORA 1 and AURORA 2 trials (i.e., CKD stage 3b-4, CKD stage 5 [dialysis and transplant]), literature values were used. For CR/PR/AD in CKD stages 3b-4, a study by Jesky *et al.,* (2016) [27] was used. This is a study of a broader population of patients with CKD (the most common cause being diabetes), where a decrement of 0.055 was assumed to apply relative to the values derived from the AURORA trials. This appears to have been derived from Table 3 of the Jesky *et al.*, (2016) study, though the exact methodology is not clear from the CS. In response to clarification (CQ B.22), the company confirm the EAG's understanding that the approach taken was to average the EQ-5D Index scores between CKD stages 1/2 and 3a and deduct the average score from stage 3b and 4.

Further utilities are then used for patients receiving dialysis and post-transplant, taken from a publication by Lee *et al.,* (2005). [28] This study used the EQ-5D in transplant recipients and compared results between groups, finding that transplant recipients had a higher utility (0.71) than patients receiving haemodialysis (0.44) or peritoneal dialysis (0.53). The company then assumed an equal 50:50 split between the two forms of dialysis, giving a mean utility of 0.485 for dialysis.

The company considered the application of disutilities associated with AEs for VCS+MMF and MMF, which were applied as a one-off disutility at the start of the model. Disutilities were estimated based on incidence of AEs observed within AURORA 1 and reported as Grade 3/4 TEAEs with an incidence of ≥1% with impact on HRQoL and assumed duration of each AE sourced from the literature. For comparators outside of the trial, an assumption was made by the company that regimens which contain MMF would have the same disutility as MMF, with all other comparator disutilities associated with AEs set to zero. The company considered this a conservative approach (with respect to comparisons against VCS+MMF).

The EAG has major concerns with all of the approaches/sources used for utility data, which are addressed in turn throughout this section. The EAG presents alternative approaches to informing health-state utility values within the model with a description outlined in Section 6.2.

4.2.7.2. Issues relating to the analysis of trial data

By taking the mean values of month 36 data to derive health state utility values by CR/PR/AD, the company's approach omits all other trial HRQoL data from consideration. The uncertainty associated with these values is therefore likely higher than implied by the stated SDs/SEs, any patients who did not provide a value at month 36 are not represented in the analysis, and if

patients have provided multiple observations, the correlation between these is not used. To underline how much data are omitted, it is the EAG's understanding that not a single value from AURORA 1 informs the estimates used in modelling (CQ B15 and CS Table B.3-10).

The company justifies its approach by stating that utilities increase then decrease in the period between months 0 and 36 (CQ B15). However not including this 'area under the curve' ignores any differences seen within the study period and is highly inappropriate, and unsuitable for use in calculating either the within trial period (given values are not stable), or for use extrapolating the likely outcomes seen in patients over time. There would appear to be two obvious appropriate methods for analysing the trial utility data (neither of which have been provided): either to analyse in a regression model by timepoint, or to estimate health state utility values from all data. The equations for such regressions are shown below for clarity:

$$MappedUtility \sim as.factor(response) + as.factor(time\ period)$$

$$MappedUtility \sim as.factor(response)$$

The first of these methods would specify a regression model incorporating time periods for which the relevant utility values would be estimated (e.g., by model cycle and response status), with values then used in the relevant model periods. The second method would specify a regression with only response status, and all values able to inform the estimate, to generate an overall health state utility value for each response category (i.e., CR, PR, and AD). The huge amount of omitted data (every observation apart from month 36) and lack of appropriate analysis method means that the EAG does not consider the utility values estimated to provide a reliable basis from which to inform decision making. Given the non-linear nature of the model, and unknown effects of proper analysis, it is not possible to speculate whether the result is biased, and in which direction any bias would impact the analysis.

Even given the company approach, there are further issues with the values used. The values presented in CS Table B.3-10 appear to be a tabulation of the mean (and SDs) of all observations which exceed the number of patients at risk in each time period. This implies multiple observations per patient were available and used in calculations of values – however the patient level values will be correlated, again meaning that SDs (and indeed means) are also unreliable. Thus, even the simplistic analysis performed by the company is inappropriate in mean values, with incorrect SDs, and is unsuitable for use in decision making.

4.2.7.3. Issues relating to the use of a decrement for CKD stages 3b-4

Although the issues relating to this assumption are of less concern to the EAG than those regarding the analysis of trial data, the approach used by the company is also limited.

The first limitation regarding this decrement is that the population in the Jesky *et al.* [27] study is a much older population (age at baseline 64 years, versus 31 years in the AURORA 1 VCS+MMF arm), predominantly with diabetes, and as such it is a strong assumption that the same decrement would apply (confirmed in CQ B22). Notwithstanding this limitation, the approach taken to uncertainty by the company is to assume the SE of the newly calculated decrement is "SE assumed to be 20% of utility value due to no SE reported in publication", however this relates to the decrement, and not the overall value.

This uncertainty is exacerbated further as utilities are then age adjusted using the often cited Ara & Brazier (2011) study. [29] Although age-adjustment applies to all health states, as CKD stage 3b-4 patients have already had a decrement applied (from an older age group), they may be impacted to a larger degree.

4.2.7.4. Issues relating to the utilities used for dialysis and transplant

In addition to the above issues, the EAG has further concerns about the approach taken to populating the model with dialysis and transplant utilities. As noted above, the company makes use of values from a study by Lee *et al.,* (2005). [28] The date of publication of this study (2006) should be acknowledged, as the underlying data informing the analysis by Lee *et al.* are now (at the time of writing) approximately 20 years old, and as with the Jesky *et al.* study, [27] the data are not specific to a population with LN, which constitutes a further limitation, but not the only concern relating to the approach.

Firstly, the company assume a 50:50 split between haemodialysis and peritoneal dialysis. Based on data reported by UK Renal Registry in its 23rd Annual Report (published in 2019), [30] 87.6% of all UK dialysis patients receive haemodialysis dialysis. This would impact the weighted utility and costs.

The second concern is the data source used. This compares the utility values cross-sectionally, which, when used directly in the model, implicitly assumes patients are similar between groups. This is unlikely to be the case in practice, where receiving a transplant is informative, and patients are generally younger and healthier i.e., they would be expected to have higher utility than non-recipients, regardless of the receipt of a transplant. This can be seen in the Lee *et al.*

study where the transplant recipients were around 10 years younger than the dialysis patients (53 versus 60-67 years [depending on sex and timing of dialysis]). [28]

These issues arose in the recent NICE appraisal of imlifidase for enabling transplant [ID1672], where the EAG identified a number of relevant references which warrant consideration in this appraisal. This includes a systematic review of utility weights through different stages of CKD by Cooper *et al.*, (2020), [31] and a regression analysis of utility values in patients waiting for transplants by Li *et al.*, (2017). [32] This latter paper by Li *et al.* presents seven regression models with various characteristics which would appear relevant to this appraisal (predominantly female, nondiabetic, younger patients), and the impact of transplant on the same patients (i.e., not comparing cross-sectionally with data taken from the UK).

4.2.7.5. Issues relating to the disutilities associated with AEs

As noted previously, the company's model included disutilities for AEs based on incidence of AEs observed within AURORA 1 and reported as Grade 3/4 TEAEs with an incidence of ≥1% with impact on HRQoL and assumed duration of each AE sourced from the literature. For treatments other than VCS+MMF or MMF, the company assumed that disutilities for MMF apply to all comparators containing MMF.

The EAG noted that a coding error was found in the company's model (in the versions provided at company submission and the revised model provided at clarification stage) which incorrectly adjusts the disutilities applied within the model based on the 6-month cycle length. The approach to estimate disutilities associated with AEs already captures the assumed duration from the literature, and therefore duration is already captured in the one-off value applied. The subsequent adjustment in the calculation for cycle length is inaccurate and halves the disutilities associated with AEs. The EAG has provided an updated analysis as part of Section 6.1.

4.2.8. Resources and costs

4.2.8.1. Overview of costs reflected within the model

The company's model includes costs relating to treatment and comparators, medical resource use, the resolution of AEs, background therapy and death (death from background mortality or death as a result of underlying LN). The costs captured by the model are discussed in turn below.

4.2.8.2. Treatment costs

All costs were presented within the model in terms of either the number of packs or vials (dependent on whether each drug was to be orally or intravenously administered).

Voclosporin

As stated within the CS, the indicative NHS list price of VCS is [redacted] per pack. At the time of writing, a proposed commercially-sensitive simple patient access scheme (PAS) is applied to the cost of VCS in the company's model. The discount is equivalent to a [redacted] discount on the list price of VCS equating to a final price of [redacted] per pack. Functionality to apply this discount to VCS is included in the model. However, the EAG noted that the PAS discount has been applied to the cost per mg rather than the cost per pack. Given the discount is based on the price per pack, the application within the model should be aligned. Within the scope of the model, acknowledging that there is only one pack size of VCS included and the dose is fixed over time, this application of the PAS discount has no impact on any cost-effectiveness analyses and is therefore not discussed further within this report.

The EAG noted that the cost of VCS is applied in the company's model based on a fixed supply of treatment with the assumption that there is no wastage associated with treatment discontinuation. In reality, it is expected that some product wastage would arise for patients that discontinue treatment part-way through a pack of treatment, though this is not explicitly reflected with the company model. For simplicity, the EAG has explored a sensitivity analysis which adds on half a pack cost of voclosporin treatment to reflect wastage of the treatment (please refer to Section 6.2 for further details).

MMF

MMF is costed within the model differently dependent on the treatment arm considered. For MMF and VCS+MMF, the dosing assumed within the model is 2.5 g/day, despite dosing schedules in AURORA 1 and AURORA 2 considering MMF at a dose of 2 g/day. Explanation for the assumption of a 2.5 g/day dose was provided at clarification (CQ B26), where reference was made to the EULAR/ERA-EDTA guidelines where the recommended dose for MMF was between 2-3 g/day. [7] The company took the average of the upper and lower bounds to inform its base-case analysis.

Within both AURORA 1 and AURORA 2 trials, a dose of 2 g/day for MMF was predominant. The EAG believes that efficacy data based on this 2 g/day dose should have informed the model rather than the 2.5 g/day dose the company implemented within its model. This is further explored in Section 6.2 where a scenario analysis using a 2 g/day dose of MMF is presented.

Comparator treatment costs

The EAG cross-checked the company's calculations of the cost per mg for each treatment within the model. Comparator costs were aligned with those referenced.

Application of relative dosing intensity (RDI)

The company applied an RDI of 100% for all treatments except for tacrolimus + MMF, which instead had an RDI estimate of 95%. In response to CQ B27, the company emphasised that treatment with tacrolimus + MMF was adjusted for TTD by instead setting RDI to 95%. Justification for the decision to substitute TTD compliance for RDI was purely cost-based, since the company stated that: *"treatment acquisition and administration costs are reduced by 5%"*, thus treatment efficacy or informative dropouts for tacrolimus + MMF adherence will not be accounted for (Company's response to CQ B27).

In the absence of mean RDI reported in the CSRs of AURORA 1 and AURORA 2, the EAG was unable to adjust the company's model to account for dose adjustments for VCS + MMF or MMF. Within the CS, the company addressed treatment discontinuation in Section B.2.3.2.3.3, highlighting that patients may have their dose reduced after 12 months within the AURORA 2 trial (i.e., after 2-years of treatment) on consultation with the Medical Monitor at the Investigator's discretion: in these instances, patients taking 23.7 mg BID of VCS could have their dose reduced to 15.8 mg BID (from three down to two capsules).

The EAG believed that there is a fundamental misinterpretation between the use of the TTD curve and the application of RDI in the model since the two are not interchangeable. Therefore, the EAG considered the company's approach to capture treatment costs to be inappropriate. The length of time that patients received treatment is not comparable to how much treatment a patient obtained relative to the anticipated (or 'target') dose, and as such the analysis presented is limited. In the absence of alternative information, the EAG has undertaken a simplified scenario analysis outlined in Section 6.2.7.

4.2.8.3. Administration costs

Administration costs were incorporated within the model if patients were administered treatment as an intravenous infusion (IV). IV costs were split into two separate costs for IV *first attendance* and IV *subsequent cycles*. Oral administration was assumed to have a cost of £0. Within the CS, it is noted that "*costs have been adjusted for inflation using the NHS cost inflation index*" (CS, Table B.3-16, pp.134). Both IV costs were sourced from the NHS National Schedule 2019/20 (version 1).

The two administration costs associated with IV attendance (first and subsequent using costs SB14Z and SB15Z respectively) could not be validated alongside the original source. For IV first attendance, the company used a cost of £404.89 in their model but referenced the SB14Z currency code on the "Total HRGs" sheet which was priced at £406.04, while the company used a cost of £339.75 within the model for the subsequent administration cycle cost, however the original source indicated that this cost would be £341.30. [33,34] Given the company states that costs were adjusted for inflation, and the NHS cost inflation index are positive, the EAG are unsure why the costs included in its model are lower than those reported in the source documentation. Further to this, as the costs are hardcoded inputs within the model (rather than inputted using the original source and inflated within the model for transparency), it is unclear to the EAG how the respective IV administration costs have been obtained given the reported references. However, the EAG accepts that these differences are relatively minor (in the region of £1-£2) and so are unlikely to have a marked impact on model results.

4.2.8.4. Background therapy (BT) costs

BT costs are applied to each comparator based on receiving tapered glucocorticoids (with dosing options from either the AURORA 1 and AURORA 2 trials, or the literature) and hydroxychloroquine. Glucocorticoids referred to methylprednisolone and prednisolone. The EAG's main concern regarding BT is the difference between tapered glucocorticoids from either the AURORA trials or the literature. These were dosed differently within the model, with a higher dose of up to 2,500 mg used outside of the AURORA 1 and AURORA 2 trials. The AURORA trial protocols outlined rapid glucocorticoid tapering to 2.5mg/day at week 16. No justification was provided in the CS as to why glucocorticoid tapering would not be considered for the alternative comparators.

Costs associated with BT were aligned except for a few instances of prednisone, where the company referenced the British National Formulary (BNF) as their cost source of this drug;

however, on inspection, only predniso*lone* was available. If the company did use predniso*lone* costs instead of predniso*ne*), the EAG identified lower price alternatives for this via eMIT. For simplicity, the EAG considered predniso*ne* to be interchangeable with predniso*lone* for costing purposes. A comparison of the company's costs of prednisone/prednisolone versus the costs sourced on eMIT are reported in Table 19, however due to the low cost of the treatment, the EAG did not anticipate this to be a driver of the cost-effectiveness results.

Table 19: Alternative prednisolone costs sourced from eMIT

Company reported costs from the BNF			Costs sourced from eMIT		
Dose	Packsize	Price	Dose	Packsize	Price
Prednisolone 1mg	28	£0.88	Prednisolone 1mg	28	£0.16
Prednisolone 2.5mg	30	£1.42	Prednisolone 2.5mg	28	£0.71
Prednisolone 5mg	30	£0.95	Prednisolone 5mg	28	£0.41
Prednisolone 10mg	30	£1.90	Prednisolone 10mg	N/A	N/A
Prednisolone 20mg	30	£3.80	Prednisolone 20mg	28	£3.30
Prednisolone 25mg	56	£40.00	Prednisolone 25mg	56	£17.72
Prednisolone 30mg	28	£29.12	Prednisolone 30mg	N/A	N/A

Abbreviations: BNF, British National Formulary; eMIT, electronic market information tool; mg, milligram; N/A not applicable.

4.2.8.5. Resource use and monitoring

4.2.8.5.1. CKD-based health states

The model considers a cost per cycle related to the occupancy of each CKD-based health state and LN stage:

- CKD stage 1-3a
    - AD
    - PR
    - CR

- CKD stage 3b-4
    - AD
    - PR (only included in scenario analysis – see Section 4.2.2.2)
    - CR (only included in scenario analysis – see Section 4.2.2.2)
- CKD stage 5
    - Dialysis
    - Transplant

Two types of costs are included per health state – a cost of health state entry (referred to as cycle 1 in the company model) and a cost applied within the health state thereafter (referred to as cycle 2+ in the company model). This distinction in costs by cycle of entry (cycle 1) versus later cycles (cycle 2+) allows for the specification of additional costs that are applied upon a particular movement, typically reflecting initial additional monitoring/ investigations. Resource use categories and frequency estimates were applied per cycle and based on clinical guidelines and KOL expert feedback to the company, with key assumptions listed:

- Given there was a paucity of evidence to inform resource use for the PR health state, the resource use frequency is an average of CR and AD (which reflect patients with an absence of flare or AD and patients in an AD health state).
- Urinalysis, complete blood count and anti-dsDNA, C3 and C4 levels monitoring occur every visit
- Serum immunoglobulin measurement, antibody tests, chronic infection screening and cholesterol and lipid monitoring occur every visit in AD, and every second visit in CR

In addition to these assumptions, the CS also states that *"resource use is identical between response states across different CKD stages, except for CKD-specific categories"* (CS, Section B.3.5.2, p.136).

A list of the resource use frequency per health state was provided in Table B.3-18 of the CS. As outlined above, resource use differed by health state, and differential resource use was applied on entry to the model health state. Entrance to the health state was determined by 'Costing

transitions' presented within the company model which used the transition probabilities to derive entrants to new health states. Overall, the EAG was satisfied with the approach taken to estimating medical resource use costs by CKD stage.

4.2.8.5.2. Additional monitoring for CNI-based treatments

In addition to CKD-based health state resource use, a further monitoring cost is applied to tacrolimus. The company did not consider it relevant to apply to the VCS+MMF arm, despite VCS being a CNI, due to an improved immunosuppressive potency, tolerable safety profile and broader therapeutic index which the company explained eliminates the need for regular therapeutic drug monitoring (CS Section B.3.5.2). This additional cost was assumed to apply at every nurse and specialist visit, with frequency dependent on CKD-stage. The EAG has explored a scenario analysis where this cost is also applied to the VCS+MMF treatment arm (given that VCS is also a CNI treatment), which is described further in Section 6.2.

As is the case for a variety of costs included within the company's model, the CS states that the costs for additional monitoring for CNI-based treatments have been adjusted for inflation indices from the NHS cost inflation index from the PSSRU 2021, however the company include no description of the exact indices used and the value incorporated within the model is a hardcoded input within the model. Without transparent explanation, the EAG is unable to cross-check the application.

4.2.8.5.3. Resource use costs incorporated within the model

Resource use costs were calculated predominantly using three sources; the PSSRU 2021, [34] NHS National Schedule 2019/2020[33] and an NHS report by Kerr (2012) [35] on costs for CKD in England.

Costs incorporated within the model were reported in Table 3.4-17 of the CS. PSSRU costs were included to account for costs associated with primary care (e.g., nurse visits). The PSSRU was also used to inflate costs to 2021 costs where appropriate. The NHS National Schedule 2019/2020 was predominantly used to inform non-Kidney specific secondary care costs and testing (e.g., ultrasound scans). The report by Kerr (2012) relating to CKD in England was used to inform CKD-specific costs (predominantly those related to transplant).

On cross-checking of the model inputs, the EAG found that costs from the NHS National Schedule 2019/2020 could not be matched with their original source. Although these costs were inputted as hard coded values, on further inspection, the EAG were able to back calculate that

the costs included were taken from the NHS National Schedule 2019/2020 version 1 and uplifted by an inflationary factor of 1.002 to reflect current prices. The 1.002 inflationary factor can be obtained from the PSSRU 2020/2021 Hospital and Community Health Services (HCHS) prices inflation index (with a NHSCII Pay & Price index of 2.21 for the year) 2019/2020. Therefore, the EAG have assumed that this was the process undertaken for informing NHS National Schedule costs within the model.[34] Despite querying some anomalous costs, and clarification provided by the company (CQ B25), there still remained a few instances where the costs included within the model could not be matched using the same methodology; these are provided in Table 21 for transparency.

The EAG was also unable to consolidate costs used from Kerr (2012).[35] An inflation rate of 1.2636 seemed consistent amongst most costs taken from this document by the company (for urinalysis, initial assessment for kidney transplant, waiting list clinic attendance, post-kidney transplantation year 2+ and anti-hypertensive medication). This cost was calculated within the model by dividing the cost included in the model for each resource unit cost by the corresponding price within the '*CKD in England'* reference. Although this rate is consistent within the model, the EAG could not re-calculate this rate of inflation (by taking the product of inflation rates provided within the relevant PSSRU resources). Calculations are provided in Table 20 below.

Table 20: Inflation rates as calculated by the EAG using PSSRU 2021 inflation indices

Sector	Years of inflation included in product	Overall inflation rate from year specified to 2021
HCHS prices	Inflation rate 2013-2021 inclusive	1.1549
	Inflation rate 2011-2021 inclusive	1.2359
HCHS pay & prices	Inflation rate 2013-2021 inclusive	1.1743
	Inflation rate 2011-2021 inclusive	1.2350

Abbreviations: EAG, evidence assessment group; HCHS, hospital and community health services; PSSRU, Personal Social Services Research Unit

Inflation rates in Table 20 were derived from taking the product of 2011-2021 inflation rates inclusive (since several costs were taken from 2010) and 2012-2021 inflation rates inclusive (publication year of this guidance) using both HCHS prices and HCHS pay & prices.

4.2.8.6. AE costs

Adverse events (AE) costs were again predominantly calculated using the NHS National Schedule 2019/20. The EAG followed calculations provided by the company, however some costs could not be matched. The EAG present a table of the costs of AEs set by the company (Table 21), highlighting differing costs upon re-calculation. Some discrepancies are thought to be rounding errors, and the EAG anticipated the impact on the model results would be minimal.

Table 21: Costs for treatment-emergent Grade 3/4 adverse events shown in company model and re-calculated by the EAG

Adverse event	Value in model	EAG re-calculated value	EAG comments
Pneumonia	£2,701.93	£2,701.93	N/A
Gastroenteritis	£2,490.47	£2,490.30	Potential rounding error
Urinary tract infection	£2,418.10	£2,423.42	Cost was not inflated
Hypertension/hypertensive crisis	£640.41	£640.22	Potential rounding error
Anaemia	£872.29	£1,352.15	Calculated using same weighted average method as for epilepsy (weighting non-elective long stay, non-elective short stay and day-case costs, then inflated by a factor of 1.0022)
Neutropenia	£619.36	£673.88	Cost could not be matched
Bronchitis	£2,299.17	£2,304.23	Cost was not inflated
Herpes zoster/ Varicella zoster virus	£8,868.09		Could not find within reference
Upper respiratory tract infection	£1,458.20	£1,458.21	Potential rounding error
Epilepsy	£1,472.93	£1,472.93	N/A
Septicaemia / Sepsis	£2,422.00	£2,422.00	N/A

Abbreviations: EAG, external assessment group; N/A; not applicable.

4.2.8.7. Second-line therapy costs

Although not explicitly outlined in the CS, subsequent therapy costs were incorporated within the model structure based on a proportion of patients receiving either MMF, azathioprine, rituximab+MMF or tacrolimus+MMF. Proportions were informed based on data from Otsuka

Pharmaceutical market estimates considerations for VCS+MMF and MMF. The assumption was made that all other model comparators would have the same subsequent treatments as MMF. A further assumption was made that besides MMF, patients would not be able to receive the same subsequent therapy as they received in the prior line (for example, patients receiving tacrolimus on the comparator arm would not receive tacrolimus as a second-line treatment). A summary of the proportions are provided in Table 22 alongside the assumed treatment duration.

Whilst the EAG do not have any major concerns with the approach taken by the company, the EAG do note two minor details of the approach taken which lack justification. Firstly the assumption that no patients can receive the same second-line therapy as they had first line – this seems justified, however the approach is not taken for MMF and it's assumed that [redacted] of patients on the MMF arm have receive subsequent MMF. Second to this, the company patients cannot receive the same second-line therapy as their first-line therapy (and this proportion is removed from the model except in the case of MMF). The EAG considered that this may be implausible and instead alternative regimens may have been administered. A different approach could have been taken by the company to redistribute the removed patients to the alternative second-line treatment options.

Table 22: Second-line therapies applied within the CEM

Comparator	**Subsequent treatment**			
	MMF	Azathioprine	Rituximab+MMF	Tacrolimus + MMF
Assumed treatment duration	8 weeks	8 weeks	6 weeks	2 weeks
Voclosporin+MMF	[redacted]	[redacted]	[redacted]	[redacted]
MMF	[redacted]	[redacted]	[redacted]	[redacted]
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]
Aza	[redacted]	[redacted]	[redacted]	[redacted]
Rituximab+MMF	[redacted]	[redacted]	[redacted]	[redacted]
Tacrolimus+MMF	[redacted]	[redacted]	[redacted]	[redacted]
Tacrolimus	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: H-CYC, high dose cyclophosphamide; L-CYC, low dose cyclophosphamide; MMF, mycophenolate mofetil;

4.2.8.8. Mortality costs

Within the CS, end-of-life (EOL) care costs was costed differently depending on whether deaths were LN-related mortality or assumed to be background mortality.

4.2.8.8.1. LN-related mortality

If a death was classed as LN-related, the company costed these events at £12,636. In PSSRU 2021, this was the average cost in the final year of life for a patient diagnosed with renal failure. LN-related deaths were defined as those that either occurred during the period of follow-up in AURORA 1 or AURORA 2, or based on a mortality risk explicitly linked to a CKD-based health state. Deaths captured from background mortality rates were considered separately.

KDIGO guidelines define CKD stage 5 as synonymous with kidney failure. [36] This implies that, within the model, this cost should only relate to people within CKD Stage 5 (i.e., at a point of needing a transplant), and as a result LN-related mortality costs may be overestimated. Some deaths were recorded within AURORA 1 and AURORA 2 (described further in Section 4.2.6), and as mentioned are defined as LN-related within the context of the model. These deaths incur the 'renal failure' EOL cost of £12,636. The EAG noted that deaths that occurred in AURORA 1 or AURORA 2 could be linked to any cause, and could in theory be partially linked to LN (e.g., a cardiovascular event associated with CKD, since CKD is associated with increased risk of cardiovascular events), [37] but could plausibly be any other cause not associated with LN.

The EAG highlights that since no patient in AURORA 1 or AURORA 2 was recorded as having progressed to CKD stage 3b-4 (and by consequence, no patient progressed to CKD stage 5 either), this EOL cost is likely inappropriate because not all deaths within the trial are LN-related and none appear to fulfil the traditional definition of renal failure. The EAG considered a scenario where the LN-related mortality cost within the model is the same as the background mortality cost, £9,590 taken sourced from the PSSRU 2021 and defined as 'any diagnosis'.

4.2.8.8.2. Background related mortality

The company applied a cost of £9,590 to people in their final year needing hospital care for a non-LN-related death. Again, to reiterate the point above, it is unclear how patients should incur this cost in comparison to the renal failure cost since cause of death is not explicitly modelled. Ultimately, the EAG highlights that all patients in the model reside within a CKD-related health state for the duration of the model time horizon, and so to an extent, it could be argued that a

large proportion of deaths are likely to be linked to either LN or CKD, yet it is less clear if this should result in a large difference in EOL costs across arms.

4.2.8.8.3. Issues related with company's approach to mortality costing

Taking into consideration the points raised above, the EAG believed it is unjustified to differentiate between EOL mortality costs since all patients must either be within a CKD stage thus experiencing a "renal failure" death, or a death unrelated to LN, in which case could experience an "any diagnosis" death (acknowledging that it is hard to capture EOL LN costs from this source). As part of EAG exploratory analyses, LN-related deaths are removed from the earlier CKD stages (1-3a). This analysis mitigates (to an extent) this issue of EOL costs, although these will still be applied for CKD states 3b-4 and 5.

5. COST-EFFECTIVENESS RESULTS

5.1. Company's cost-effectiveness results

5.1.1. Base case results

An updated model was provided by the company at clarification stage with several edits provided to the cost-effectiveness calculations. These are described by the company as:

- Updates to AEs
- Updates to the medical resource use costs
- Updates to NMA results for PR
- Connecting RDI for TAC+MMF
- Fixing the error on the Outcomes sheet described as 'some numbers in column A which are used as indices for arrays of results

The results presented within the model did not consistently align with the results presented alongside the CQs (see Table 40 – clarification response). The EAG has assumed that results within the model file are correct, and any discrepancies in results presented in the company's clarification response were minor typographical or copy/paste errors. The results within the company model are shown in Table 23. The deterministic ICER for VCS+MMF versus MMF alone was £19,876. Updated probabilistic results were not provided by the company and therefore have been run and presented by the EAG using the company's updated model provided at clarification (also as part of Table 23).

All results versus the listed comparators were presented by the company as pairwise analyses, not incremental analyses, therefore the EAG has provided full incremental analysis of the comparators listed (shown in Table 24).

Table 23: Company base case results

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Pairwise cost per QALY gained
Company deterministic base case (results taken from EAG from updated company CEM provided at clarification stage)					

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Pairwise cost per QALY gained
VCS + MMF	[redacted]	[redacted]		[redacted]	
MMF	[redacted]	[redacted]	[redacted]	[redacted]	£19,876
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£11,411
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£10,914
AZA	[redacted]	[redacted]	[redacted]	[redacted]	£15,855
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,848
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,169
TAC	[redacted]	[redacted]	[redacted]	[redacted]	£17,833
Company probabilistic base case (analysis run by EAG from updated company CEM provided at clarification stage)					
VCS + MMF	[redacted]	[redacted]	[redacted]	[redacted]	-
MMF	[redacted]	[redacted]	[redacted]	[redacted]	£21,086
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£11,962
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£11,458
AZA	[redacted]	[redacted]	[redacted]	[redacted]	£17,041
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£20,683
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,364
TAC	[redacted]	[redacted]	[redacted]	[redacted]	£18,331

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALY, quality adjusted life year; RTX, rituximab; TAC, tacrolimus; VCS

Table 24: Full incremental analysis of voclosporin+MMF versus comparators – company base case

Treatment	Total discounted costs	Total discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	ICERs (following re-baseline)
Company incremental base case (results taken by EAG from updated company CEM provided at clarification stage)					
MMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AZA	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated

Treatment	Total discounted costs	Total discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	ICERs (following re-baseline)
TAC	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
VCS + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£19,897

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALY, quality adjusted life year; RTX, rituximab; TAC, tacrolimus; VCS

5.2. Company's sensitivity analyses

5.2.1. Deterministic sensitivity analysis

The company presented the results of a one-way sensitivity analysis to explore the sensitivity of the base case results by varying key parameters within plausible 95% confidence intervals. The included parameters are respective ranges presented as an Appendix to the company submission document (CS Appendix O). The EAG noted that as part of the original submission the company did not present DSA results against any comparison besides VCS+MMF vs. MMF. Further to this, in response to the CQs, the company did not provide an updated deterministic sensitivity analysis, following revisions to the model. The EAG have therefore re-ran the analysis presented within the model for VCS+MMF vs MMF and results are presented in Figure 5, Figure 6 and Figure 7 for the impact on the incremental costs, incremental QALYs and the ICER respectively.

Figure 5: DSA: Incremental costs from company model (analysis ran by EAG on the updated company CEM provided at clarification stage)

Abbreviations: AD, active disease; CKD, chronic kidney disease; LN, lupus nephritis; MMF, mycophenolate mofetil

Figure 6: DSA: Incremental QALYs from company model (analysis ran by EAG on the updated company CEM provided at clarification stage)

Abbreviations: AD, active disease; CKD, chronic kidney disease; partial response; MMF, mycophenolate mofetil; QALY, quality-adjusted life year

Figure 7: DSA: ICER from company model (analysis ran by EAG on the updated company CEM provided at clarification stage)

Abbreviations: AD, active disease; CKD, chronic kidney disease; ICER, incremental cost-effectiveness ratio; MMF, mycophenolate mofetil

The EAG has a fundamental issue with the company's DSA. Firstly, the inclusion of interlinked parameters within the DSA (for example transition probabilities from AD CKD 1-3a to death). Whilst important to test parameter uncertainty associated with such transitions, this parameter is linked with several other transitions within the model to ensure that the transition probabilities sum to 100%. As the model contains adjustments to account for differences in transition probabilities, varying this probability to death has a knock-on implication for other transition probabilities from the AD CKD 1-3a health states (for patients remaining in AD CKD 1-3a). This is illustrated in Table 25 which shows the transition probabilities when varying this parameter at its lower and upper bound, and the impact on the VCS+MMF 6-month transition probabilities. Given the parameters are interlinked, the description provided by the company of a deterministic one-way sensitivity analysis is inaccurate as all parameters were not varied one at a time. This is problematic for two of the top ten results in the DSA ('AD CKD 1-3a -> Death' and 'AD CKD 1-3a -> AD CKD 3b-4'). It is the opinion of the EAG that interlinked parameters should not be included in a DSA framework and instead should be explored through PSA and scenario analysis to avoid misinterpretation of results.

Table 25: CKD 1-3a AD transition probability at 6-months for VCS + MMF

To:		From CKD 1-3a AD		
		Deterministic	**Upper bound**	**Lower bound**
CKD 1-3a	**CR**	[redacted]	[redacted]	[redacted]
	PR	[redacted]	[redacted]	[redacted]
	AD	[redacted]	[redacted]	[redacted]
CKD 3b-4	**CR**	[redacted]	[redacted]	[redacted]
	PR	[redacted]	[redacted]	[redacted]
	AD	[redacted]	[redacted]	[redacted]
CKD 5	**Dialysis**	[redacted]	[redacted]	[redacted]
	Transplant	[redacted]	[redacted]	[redacted]
Death		[redacted]	[redacted]	[redacted]
Sum		[redacted]	[redacted]	[redacted]

Abbreviations: AD, active disease; CKD, chronic kidney disease; CR, complete response; PR, partial response

Second to this issue, the EAG believed that in other instances parameters lacked face validity when tested at their lower and upper bound and may substantially over-estimate the volume of uncertainty each parameter is associated with. For example, the utility value for CR CKD Stage 1-3a is varied between bounds of 0.433 and 0.997 (with a deterministic input of 0.83). Given the utilities were derived from the SF-36 in AURORA 2 (at Month 36), it is probably that a realistic lower bound of the CR CKD Stage 1-3a utility would also translate to a similarly lower utility for patients in the PR and AD health states (which remain constant in the DSA framework at 0.8 and 0.71 respectively), implying that a patient has a substantially lower HRQoL in the best health state feasible within the model (CR in CKD stage 1-3a); this lacks face validity.

Finally, the EAG considered that the company should have considered presenting DSA in the context of a net-monetary benefit (NMB) as opposed to the ICERs, given results produce negative values. In the context of negative ICERs, it is not possible without further investigation to understand where on a cost-effectiveness plane the results are positioned (i.e., is the intervention less costly and more effective and therefore dominant, or conversely less effective and more costly and therefore dominated by the comparator).

In summary, the EAG does not consider the specific outputs of the DSA to be relevant for decision making except to illustrate that parameters included (isolated and linked) impact model results, and edits should be made to exclude inappropriate parameters before results can be interpreted in a meaningful way.

5.2.2. Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) to explore parameter uncertainty with 1,000 iterations conducted. The company did not provide an updated PSA as part of the response to clarification and therefore the EAG have re-ran the analysis on the updated model provided by the company at clarification stage. In line with the format presented by the company in the submission, the EAG provide Figure 8, Figure 9 and Figure 10 which illustrate the PSA results in a PSA scatterplot for total discounted costs and QALYs, the PSA scatterplot for incremental discounted costs and QALYs and the cost effectiveness acceptability curve (CEAC) respectively.

Figure 8: Cost-effectiveness plane – total discounted costs and QALYs (analysis re-ran by the EAG in the updated company's CEM provided at clarification stage)

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALYs, quality adjusted life years; RTX, rituximab; TAC, tacrolimus

Figure 9: Cost-effectiveness plane – incremental discounted costs and QALYs (re-ran by the EAG in the updated company's CEM provided at clarification stage)

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALYs, quality adjusted life years; RTX, rituximab; TAC, tacrolimus

Figure 10: CEACs (re-run by the EAG in the updated company's CEM provided at clarification stage)

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; RTX, rituximab; TAC, tacrolimus

Given the number of comparators included within the graphs, the EAG provide a further diagram (Figure 11) which illustrates the parameter uncertainty within the PSA for VCS+MMF vs. MMF. In addition to this, for ease of interpretation, the EAG have also added in the deterministic result and the average result from the 1,000 iterations to the graph taken from the updated company model provided at clarification stage. As illustrated within the revised diagram, the incremental

costs associated with VCS+MMF vs. MMF alone are always positive (indicating that VCS+MMF costed more in each of the 1,000 iterations) and vary between £2,872 and £32,411. Incremental QALYs associated with VCS+MMF vs MMF alone varied between -1.17 and 3.591 cross the 1,000 iterations. These results indicate that there is a wide range of parameter uncertainty within the company model and selected model base case. Overall, the deterministic and probabilistic mean values were similar with similar incremental costs and slightly lower probabilistic incremental QALYs.

The individual diagrams for VCS+MMF vs the other comparators included within the model are provided as an Appendix (Appendix A).

Figure 11: Cost-effectiveness plane for voclosporin+MMF vs MMF

Abbreviations: MMF, mycophenolate mofetil; QALYs, quality adjusted life years

5.2.3. Scenario analyses

The company undertook a range of scenario analyses to consider alternative data sources and assumptions within the economic model. Full details of this are provided in CS Section B.3.11.3. The company provide scenario analysis of voclosporin+MMF versus MMF related to:

- Time horizon
- Discount rates
- Stopping rules
- Utilities

- TTD
- Wastage

The EAG considered the range of scenarios presented by the company to be limited in range, and hence have limited ability to wholly explore structural uncertainty within the model and decision problem.

5.3. Model validation and face validity check

An overview of the company's approach taken to validate the submitted cost-effectiveness analysis is provided in Section B.3.14.1 of the CS. The company notes that a technical validation of the model was undertaken internally to ensure that calculations of the model were correct prior to submission. The company also stated that an external health economist reviewed the CS with feedback incorporated prior to submission. Details of the technical validity were not provided by the company, nor were details of the type of review undertaken by the external health economist, and so the EAG does not discuss this further. However, further details of the EAG's corrections and adjustments to the company's model are provided in Section 6.1 of this report.

In addition to the technical validation, the company also sought to compare data from AURORA 1 to the outputs of the model as an internal validation exercise (CS Section B.3.14.1). The company presented estimates of the proportion of patients with PR or CR at 12 months in the model, versus the 'true' results of AURORA 1 (CS Table B.3-28). At clarification stage (CQ B32), the EAG queried an apparent discrepancy between the 'published' PR value for VCS + MMF of 70% (125/179, from Rovin *et al.*, 2021) and the implied CR+PR 'count' value of 74.86% (134/179, which can be inferred from CS Table B.3-28). In response, the company explained that PR is not mutually exclusive from CR using the definition of response in Rovin *et al.*, (2021), and that most but not all patients who achieved CR also achieved PR. The EAG therefore does not consider this discrepancy to be an error, but instead highlights the difficulties associated with comparing PR and CR rates, given that most but not all CRs can also be considered PRs.

With respect to the internal validation exercise, the EAG again noted that because of the designs of AURORA 1 and AURORA 2, the company could not present the results of an equivalent internal validation exercise for a time horizon longer than 1 year. Consequently, the

EAG considered the internal validation to have limited merit beyond confirming that the 'count method' yields transition probabilities that largely reflect the data collected in AURORA 1.

Outside of the remit of model validation, the EAG highlighted that the company's model included a number of apparent input parameters which have no influence on model calculations. The EAG considered the inclusion of these parameters to be problematic in terms of transparency; however, since they do not compromise the model calculations, these 'unused' parameters are not considered further as part of the EAG's critique.

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

Due to the size and complexity of the model, paired with limited description within the CS, a thorough cell-by-cell inspection of the model was not feasible within the timeframe available. However, the EAG conducted black box (i.e., face validity) tests on the model in Excel alongside a crosscheck of inputs included within the model. The structure of the company's model was somewhat rigid in terms of how it captured health and cost outcomes associated with LN. Given the rigid structure, the EAG's ability to incorporate additional flexibilities to adequately understand uncertainty associated with the decision problem was limited.

This section is organised as follows: Section 6.1 details the impact of errors identified in the EAG's validation of the executable model (focused mostly on 'black box' tests and crosschecking input parameters). Section 6.2 details a series of scenario analyses exploring the areas of concern identified by the EAG (as discussed throughout Section 4 of this report). A summary of the scenarios explored by the EAG are provided at the end of Section 6.2. Following identification of corrections and investigation of the scenarios undertaken by the EAG, combined with alternative functionality included by the company in its submitted model, the EAG presents its preferred base-case analysis in Section 6.3. Finally, Section 6.4 presents the EAG's conclusions of the cost-effectiveness section of the CS.

6.1. EAG corrections and adjustments to the company's base case model

Below is a short list of errors that the company identified after submission and resolved in a revised model provided at clarification stage:

- RDI for tacrolimus+MMF was taking a value of 100% rather than 95% as intended
- Error in inflationary costs
- Error in results sheet where resource use was referring to incorrect cell ranges
- Error in NMA application for PRR

EAG also noted an error found in application of disutilities, as these values were mistakenly halved. The disutilities associated with AEs affect the 'QALYs' sheet, on rows 5:6 columns Q, AB, CP, DA. While this errors only affects MMF containing regimens, since VCS+MMF contains MMF the company's base-case results are affected as a result of resolving this error. Table 26 provides a summary of the EAG-corrected company base-case results, and Table 27 provides a

breakdown of revised incremental analysis. All scenarios provided by the EAG are with the correction for disutilities applied. While the EAG consider MMF to be the main comparator for consideration and hence provide a breakdown of results for voclosporin+MMF versus MMF, advice to the EAG indicated that tacrolimus may also be a comparator of interest. As such, for the EAG corrected base case and the summary of EAG preferred base case, full incremental analysis is presented which shows a comparison of voclosporin+MMF versus all comparators within the model (including the key comparators of interest, MMF and tacrolimus).

Table 26: EAG-corrected company base case results

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained
EAG corrected company deterministic base case					
VCS + MMF	[redacted]	[redacted]	-	-	-
MMF	[redacted]	[redacted]	[redacted]	[redacted]	£19,897
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£11,468
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£10,966
AZA	[redacted]	[redacted]	[redacted]	[redacted]	£15,947
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,882
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,189
TAC	[redacted]	[redacted]	[redacted]	[redacted]	£17,969
EAG corrected company probabilistic base case					
VCS + MMF	[redacted]	[redacted]	-	-	-
MMF	[redacted]	[redacted]	[redacted]	[redacted]	£21,508
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£12,191
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£11,754
AZA	[redacted]	[redacted]	[redacted]	[redacted]	£17,422
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£21,854
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£18,782
TAC	[redacted]	[redacted]	[redacted]	[redacted]	£19,186

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALYs, quality adjusted life years; RTX, rituximab; TAC, tacrolimus; VCS

Table 27: EAG corrected: Full incremental analysis of voclosporin+MMF versus comparators – company base case –

Treatment	Total discounted costs	Total discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	ICERs (following re-baseline)
Company incremental base case (results taken by EAG from updated company CEM provided at clarification stage with fix applied)					
MMF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AZA	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
TAC	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Strictly Dominated
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
VCS + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£19,897

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALY, quality adjusted life year; RTX, rituximab; TAC, tacrolimus; VCS

6.2. Exploratory and sensitivity analyses undertaken by the EAG

The EAG have undertaken a range of alternative exploratory analyses within the company's model. Whilst some exploratory analysis links to functionality already within the model provided by the company, further model edits have also been undertaken to try to explore structural uncertainty, where possible. Each model scenario is discussed in turn throughout this section.

6.2.1. Scenario 1: Amending the approach to applying trial-based utility values to CKD states 1-3a (AD, PR and CR)

In the company's model, the heath state utility values were based on 36-month data from AURORA 2. The EAG considered this approach to be inappropriate as it ignores all data from month 0 to month 36 (see Section 4.2.7.2). Based on this, the EAG considered an analysis which produces a weighted average utility value per health state (CKD stage 1-3a for AD, PR and CR) based on the information provided within the company submission (CS Table B.3-10)

and applied the calculated values as health state utilities within the model. Table presents a summary of the utility values obtained and applied within the EAG analysis.

Table 28: EAG scenario - weighted average of mapped utility values

Health state	Health state utility values applied in company base case	Health state utility values applied in EAG analysis
CKD Stage 1-3a CR	0.830	0.814
CKD Stage 1-3a PR	0.800	0.800
CKD Stage 1-3a AD: Non-response	0.710	0.749

Abbreviations: AD, active disease; CR, complete response; PR, partial response

As shown in the approach the company have taken, the utilities for the CR health state are higher (+0.162) than when applying a weighted average, and the AD value is lower (-0.385). The EAG considered that this approach may favour the VCS+MMF arm within the model where CR rates are higher. Whilst the differences may appear small between the two methods, the company's base case ICER for VCS+MMF versus MMF increases from £19,897 to £21,401 (+£1,504) when applying the weighted values. In the absence of a regression model, the EAG considered this scenario to represent a more reasonable approach to modelling utility values based on the data available.

6.2.2. Scenario 2: Amending health state utility values for CKD Stage 5 (transplant and dialysis)

As outlined in Section 4.2.7.4, the EAG had concerns regarding the approach taken to populating the model with dialysis and transplant utilities. The company makes use of values from a study by Lee *et al.,* (2005) [28], which, as the EAG outlined previously, has limitations in comparability with the LN patient population relevant to this appraisal (e.g., 53 versus 60-67 years [depending on sex and timing of dialysis]).

The EAG identified Cooper *et al*., 2020[31] which was a systematic review of utility weights through different stages of CKD, and a study by Li *et al.,* (2017), [32] which presents regression models with various characteristics relevant for consideration in an LN setting (e.g., predominantly female, nondiabetic, younger patients).

The EAG explored three alternative approaches to applying health state utility values for the CKD Stage 5 transplant and dialysis utility values.

6.2.2.1. Scenario 2A: the EAG applied a transplant utility value taken from the Li *et al.* [32] regression analysis

Within the Li *et al.,* 2017 publication[32] eight regression analyses are presented which include predictive variables on health state utility values for waiting list patients and transplant recipients. The EAG considers 'model 7' to be the most relevant for consideration with transplant values versus waiting list, age, gender and diabetes status. Using the regression model with the average age (33.2†) and proportion female (87.7%) from the company model, a revised estimate for transplant patients was estimated using the following formula (assuming no patients were diabetic):

$$Transplant\ utility = baseline\ (0.830) + age + diabetic + female + transplant$$

$$Transplant\ utility = 0.830 + (-0.036) + 0 + \ (-0.033 * 0.877) \ + \ (+0.053)$$

$$Transplant\ utility = 0.818$$

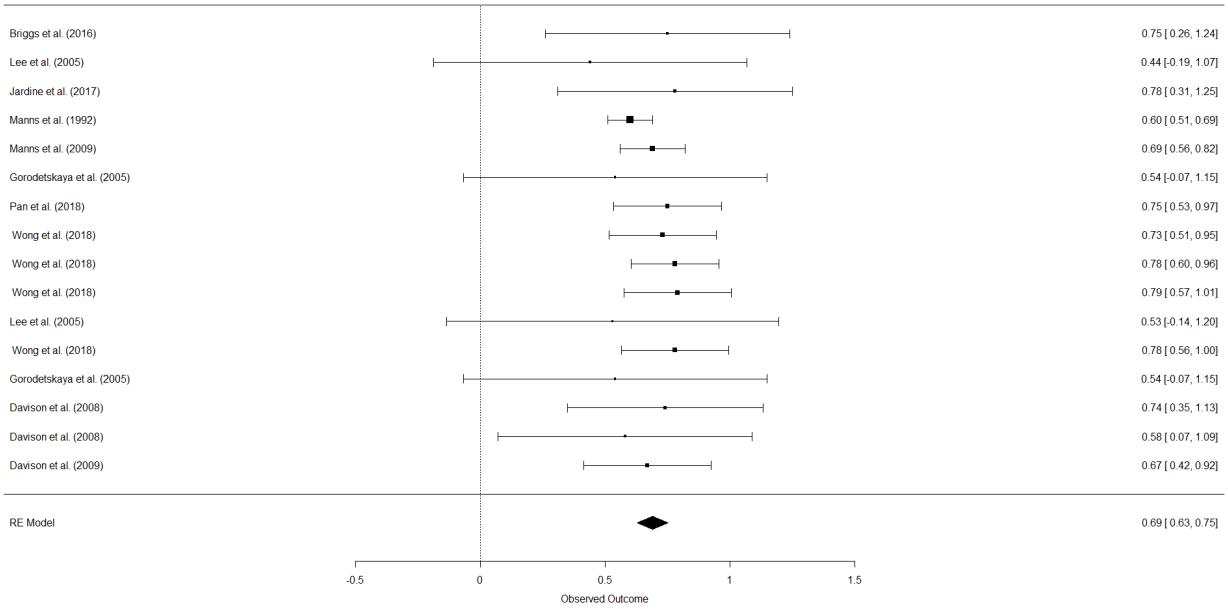
The company's base case ICER for VCS+MMF versus MMF increases from £19,897 to £20,152 (+£255) when applying the alternative transplant value.

6.2.2.2. Scenario 2B: the EAG applied a dialysis utility value taken from a meta-analysis of values presented within Cooper et al. [31]

In the systematic review by Cooper *et al.,* [31] utility weights through different stages of CKD are presented. Within the paper, Table 4 presents a summary of all CKD Stage 5 utilities, split by dialysis and transplantation. The EAG meta-analysed the dialysis values presented in Cooper *et al.* to obtain a mean estimate of 0.69. This scenario explored the impact of applying the meta-analysed value to the CKD stage 5 dialysis health state. Using this value increases the company base case ICER by £87 (from £19,897 to £19,983 for VCS+MMF versus MMF).

† Age is a categorical variable in 'model seven' from Li *et al.*, (2017)

Figure 12: Meta-analysis of dialysis utilities outlined in Cooper et al. 2020[31]



6.2.2.3. Scenario 2C: the EAG apply a dialysis utility value taken from the largest source of EQ-5D data for dialysis patients with an applied utility increment for transplant patients (taken from Li *et al*. [32])

In this final scenario, the EAG amended both the CKD stage 5 (dialysis and transplantation) values simultaneously. This scenario uses the 0.75 dialysis value from Briggs *et al.,* 2016 as presented in Cooper *et al.,* (2020) [31] This was selected as the largest source of EQ-5D-3L data. The transplantation utility was calculated by using the Briggs et al 0.75 value and applying the transplant increment reported in Li et al. (+0.053 as outlined in 'model 7'). [32] The resulting change in the company ICER when this scenario is applied is a slight increase of £334 (£19,897 to £20,230 for VCS+MMF versus MMF).

6.2.3. Scenario 3: Wastage applied to voclosporin

Voclosporin is expected to be dispensed in packs providing a 30-day supply (180 tablets of 7.9mg dose). However, in the company's model, patients are modelled to incur the cost of treatment based on the half-cycle corrected LYs within a model cycle and based on a time-to-treatment discontinuation curve. Hence, patients are costed to receive the precise number of tablets within a model cycle that are needed, with no rounding to account for the number of tablets dispensed. In reality, it is expected that some product wastage for voclosporin would arise for patients that discontinue due to any cause part-way through a pack. As this is not

explicitly modelled within the company base case, the EAG has explored a simple analysis which adds on half of an additional pack of voclosporin to the overall incremental costs projected by the model to ascertain the potential impact of including wastage within the model results. This analysis causes the company's base case ICER to increase from £19,897 to £20,413 (+£516).

6.2.4. Scenario 4: 2g dose of MMF

The company base case applied a 2.5g dose of MMF daily based on referenced guidelines from EULAR/ERA-EDTA, which suggested a recommended dose between 2-3g. To align with the AURORA 1 and AURORA 2 trial, the EAG considered a scenario applying a 2g dose. The impact of this scenario is minimal on the base case results producing a revised ICER that is £13 less than the deterministic result (£19,897 versus £19,884).

6.2.5. Scenario 5: Additional monitoring for CNI treatment applied to voclosporin

As outlined within Section 4.2.8.5.2, the company included additional monitoring for CNI treatments (tacrolimus). The company did not consider it relevant to apply this additional cost to the VCS+MMF arm, despite VCS being a CNI, as they state it has an improved immunosuppressive potency, a tolerable safety profile and broader therapeutic index. Expert advice to the EAG suggested that, based on the current available evidence, voclosporin would be considered comparable to other CNIs with regard to monitoring. Therefore, for completeness the EAG has explored a scenario where the cost is applied to all CNI treatments within the model (i.e., the VCS+MMF and tacrolimus arms).

This scenario increases the company base case ICER from £19,897 to £20,862 (+£965).

6.2.6. Scenario 6: Amendments to cost inputs to align with referenced sources

As outlined throughout Section 4.2.8, there were several instances where the company's description of a given cost did not align with the original sources. As such, the EAG conducted a scenario that aligned the costs to the original sources, applied cheaper drug cost prices where available, and inflated costs to current prices where relevant. In addition to this, the EAG also adjusted the LN-related mortality cost to be aligned with 'any diagnosis' end of life cost as reported within the PSSRU 2021. The rationale for this was two-fold: firstly, the description of renal failure within the PSSRU may relate to the later CKD stages (i.e., CKD stage 5), and

therefore death from earlier states may be overestimating costs applied; second to this, the costs included within the PSSRU account for costs within the final year of life, and resource use within the model already varies by health state, and as such these differences may already be captured within the model resource use calculations. A description of the amendments made are shown in Table 29. Whilst the amendments to the costs are small, the resulting impact is a +£217 on the company base case ICER (from £19,897 to £20,114).

Table 29: EAG amended costs

Cost type	Cost description	Company model cost	EAG scenario cost
AE cost	Urinary tract infection	£2,418.10	£2,423.42
	Anaemia	£872.29	£1,352.15
	Neutropenia	£619.36	£673.88
	Bronchitis	£2299.17	£2,304.23
Resource use costs	Initial assessment for kidney transplant	£3,205.72	£3,135.49
	Waiting list clinic attendance (pre-transplant)	£3,754.12	£3,617.87
	Post-kidney transplantation year 2+	£9,246.94	£9,044.35
	Anti-hypertensive medication	£166.79	£163.14
Prednisone/prednisolone costs	1mg	£0.88 (28 pack)	£0.16 (28 pack)
	2.5mg	1.42 (30 pack)	£0.71 (28 pack)
	5mg	£0.95 (30 pack)	£0.41 (28 pack)
	20mg	£3.80 (30 pack)	£3.30 (28 pack)
	25mg	£40 (56 pack)	£17.72 (56 pack)
EOL cost	LN related death	£12,636	£9,590

Abbreviations: AE, adverse event; LN, lupus nephritis

6.2.7. Scenario 7: Amendments to estimating treatment costs for the intervention and comparators

The company applied an RDI of 100% for all treatments except for tacrolimus + MMF, which instead had an RDI estimate of 95%. Further to this, TTD curves were applied to the VCS+MMF and MMF arms but all other comparators were assumed to have no treatment discontinuation. Based on responses to clarification questions (and outlined in Section 4.2.8.2), the EAG believe

that there is a fundamental misinterpretation between the use of the TTD curve and the application of RDI in the model with regard to estimating treatment costs. A TTD curve provides information about the duration of time that patients spend on treatment before permanent discontinuation, whereas RDI provides an estimate of the proportion of treatment that was administered relative to the planned dose (for those patients still receiving treatment). As such the EAG has conducted two additional scenarios in relation to estimating treatment costs within the economic model. These are discussed in turn.

6.2.7.1. Scenario 7A: the EAG assuming an RDI of 95% for all comparators

In this scenario, the EAG apply an RDI value of 95% for all treatment options included in the model (i.e., all comparators and VCS+MMF). Whilst a simplified scenario using an arbitrary number (though the estimate of 95% was applied to tacrolimus within the company base case), in the absence of alternative data, either from the literature or from the AURORA studies, this scenario considered that not all patients will receive 100% of the planned dose.

Though the EAG acknowledges the limitations of using essentially arbitrary values to inform RDI, in the absence of an alternative approach which exhibits face validity, the EAG deems the use appropriate for exploration. The application of the 95% RDI reduces the ICER from £19,897 to £18,699 (-£1,198) within the comparison of VCS+MMF versus MMF.

6.2.7.2. Scenario 7B: the EAG assuming TTD equivalent to MMF for all other model comparators

In this scenario, the model assumes that for all comparators (but not the VCS+MMF arm), that TTD is equivalent to the curve informing the MMF arm. Similar to the scenario above (7A), this scenario serves as an exploratory analysis to illustrate that not all patients are likely to remain on treatment throughout the duration of the model and may discontinue for a plethora or reasons (including but not limited to lack of efficacy and occurrence of adverse events).

Table 30 reports the impact of this scenario in comparison to the company base case. The largest impact is on the VCS+MMF versus rituximab+MMF scenario, where the ICER increases by £4,922. All other comparisons have a relatively small impact on the ICER (varying from £22 difference for VCS+MMF versus AZA, to £681 for VCS+MMF versus tacrolimus).

Table 30: EAG exploratory analysis: comparison of ICERs when applying a treatment costing scenario assuming TTD for non-MMF comparators is equivalent to MMF

Comparisons	Company base case ICER	EAG treatment costing scenario: assuming TTD for comparators is equivalent to MMF
VCS + MMF vs MMF	£19,897	£19,897
VCS + MMF vs L-CYC	£11,468	£11,833
VCS + MMF vs H-CYC	£10,966	£11,316
VCS + MMF vs AZA	£15,947	£15,968
VCS + MMF vs RTX + MMF	£18,882	£23,804
VCS + MMF vs TAC + MMF	£18,189	£18,663
VCS + MMF vs TAC	£17,969	£18,649

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; RTX, rituximab; TAC, tacrolimus; TTD, time-to-treatment discontinuation; VCS, voclosporin.

6.2.8. Scenario 8: Exploratory impact of restricting movement from CKD 5 transplant back to CKD 5 dialysis

As outlined within Section 4.2.6.7, the memoryless nature of the Markovian model and the ability for patients to move backwards and forwards between the CKD stage 5 health states (dialysis and transplant), means that it is possible for patients to undergo multiple transplants (with the same probability per model cycle) over the modelled lifetime horizon. The EAG considered this to lack face validity, and as such have explored a scenario (already existing within the company model) which disables movement from re-transplantation. This analysis in isolation had a low impact on the modelled ICER with a difference of +£460 from the company base case (£20,357 and £19,897 respectively).

6.2.9. Scenario 9: Reduction in transplantation rates (CKD 5 dialysis to CKD 5 transplant)

Advice to the EAG indicated that transplantation rates included within the company's base case model may be too high (90% of patients receiving a kidney transplant within two years from developing stage 5 CKD – translating to a per-cycle rate of 43.77%). Expert advice suggested that 65% (within two years) may serve as a better reflection of current clinical practice. As such, the EAG have considered a scenario which reduces the transplantation rate to be 65% over two

years (translating to a per cycle rate of 23.08%). This exploratory scenario has a relatively small impact on the base case ICER for VCS+MMF versus MMF (£19,897 in the base-case compared to £19,526).

6.2.10. Scenario 10: Removal of LN related deaths for CKD stages 1-3a

The EAG previously highlighted limitations of including LN-related mortality within the model for the early CKD stages (1-3a). The EAG has therefore conducted a scenario which removes LN death from the model in the first 36 months from CR and PR CKD stages 1-3a (Scenario 10A), and another scenario removing LN death from the model in the first 36 months from CR, PR, *and* AD CKD stages 1-3a (Scenario 10B). The rationale for undertaking these scenarios is two-fold. Firstly, the LN-related deaths incorporated within the company's cost-effectiveness model were based on a small number of observed deaths in AURORA 1 and AURORA 2, and the methodology used to estimate transition probabilities within the model means that the deaths can have a marked impact on results, which may not be a true reflection of reality and instead an artefact of a within trial analysis and small sample size. This issue is exacerbated further by the fact that the approach taken to capture LN-related deaths differs according to health state (i.e., CR and PR deaths are estimated as time-varying and arm-specific, whereas AD deaths are constant over time, and equal across arms). Secondly, the EAG considered it counter-intuitive that it was infeasible within the model structure for patients to progress CKD stage within the 36-month window however they could experience LN-related death. Based on this, the EAG believe it is possible that LN-related death could be overestimated within the model, and with the application of mortality specific costing (see Section 4.2.8.8), this could in turn overestimate total costs within the model and underestimate total QALYs gained across treatment arms.

6.2.10.1. Scenario 10A: Remove LN deaths for CR and PR, CKD stages 1-3a using count method

In this scenario, the EAG removed the impact of the 'count method' deaths that apply to some model cycles for the CR, PR and AD health states (in CKD stages 1-3a) based on data from only AURORA 1 and 2. The removal of the LN-related deaths for CR, PR and AD, CKD stages 1-3a, has a substantial impact on the company's modelled ICER, increasing the base-case ICER from £18,897 to £23,497 (+£3,600). The removal of these deaths adjusts the ICER, however LN death still occurs due to an additional model parameter (explored in Scenario 10B).

6.2.10.2. Scenario 10B: Remove LN deaths for CR, PR, and AD, CKD stages 1-3a using count method and additional model scenario

As an alternative to Scenario 10A, in Scenario 10B the EAG assumed that the risk of death in each of the CKD stage 1-3a states (for first 36-months of the model) would be captured by background mortality (which is also accounted for within the company base case model), however two methods are used to remove early-stage CKD deaths from the model. This scenario involves the adjustments to the 'count method' outlined in Scenario 10A, with a further adjustment to a switch within the company's model labelled "Transitions shared between all treatments, AD CKD 1-3a -> Death". The company base case inputs the count data method and a further proportion of 1.729% also referenced as being count method data. Therefore, without further description the EAG considered there could be potential risk of double counting of deaths within the model for CKD stages 1-3a. The removal of the LN related deaths for CR, PR and AD, CKD stages 1-3a, in the 36-month transition probabilities as well as amending the additional parameter to 0% has an strikingly large impact on the company's modelled ICER, increasing the base-case ICER from £18,897 to £38,125 (+£18,228).

6.2.11. Scenario 11: Inclusion of transitions into CKD 3b-4 and 5 in the first 36 months

Within the company base case, the model framework does not allow patients to experience CKD progression within the first 36 months of the time horizon. Whilst CKD progression was not observed within the AURORA 1 and AURORA 2 trial follow-up, the EAG considered it feasible that some patients may experience CKD disease progression, and this transition may be of particular relevance for those patients who do not respond to treatment (and hence remain in an AD health state). This 'protective' assumption by the company may be particularly problematic when considering patients who have received prior treatment with MMF (54.9% of the AURORA 1 population), and who still do not achieve response (e.g., within the current model framework and based on the anticipated patient population, it is feasible that a patient could have been receiving MMF for several months with no response to treatment, enters the model, receives VCS+MMF, still does not achieve response, and yet their CKD is still contains a protective property which means their CKD cannot progress for 36 months).

The EAG therefore explored a scenario analysis which already exists within the company's economic model allowing patients to transition from CKD stages 1-3a to 3b-4 within the first 36 months. The transition in this scenario is only considered for movements from AD and patients

in a PR and CR health state are still 'protected' from CKD progression unless they lose response (i.e., move to AD). The movement from CKD stage 1-3a AD to CKD stage 3b-4 AD is 3.05%, which the company derived from KOL expert feedback which indicated that the probability of patients progressing CKD stage was 6% per year.

The inclusion of this scenario has a large impact on the company ICER and reduces the base case ICER for VCS+MMF versus MMF from £19,897 to £14,811 (-£5,086) highlighting the extent of structural uncertainty within the model.

6.2.12. Scenario 12: Long-term transition probabilities for VCS+MMF and MMF and the implementation–

The company describe how uncertainty related to sustained efficacy within the model was captured by applying a long-term waning effect for VCS+MMF which assumed that when patients stopped treatment at 36-months within the model, transition probabilities were averaged between the treatment arms from AURORA 2 (i.e., VCS+MMF and MMF). The EAG considered two main limitations with this application:

1. This application of a treatment waning effect is still based on patients receiving treatment in the AURORA 2 trial (and therefore the implicit assumption is made that the treatment effect for both VCS+MMF and MMF alone would be maintained after stopping treatment at 36-months for the remainder of the 72 year time horizon within the model).

2. The assumption made by the company is not that the treatment effect of VCS wanes for all patients, but rather is that an average between the two arms is taken (inherently assuming that some treatment effect is maintained for VCS+MMF versus MMF).

With a lack of longer-term data, the EAG are unable to explore uncertainty with regard to how VCS+MMF would compare to MMF once patients have stopped treatment. Despite this, findings from the literature (Jourde-Chiche 2022[11] – as outlined in Section 4.2.6.3) found evidence related to the waning of treatment effectiveness over time in an LN-specific population. As such, the EAG consider it reasonable to assume that the effect of VCS+MMF or MMF alone could wane over time and there is no guarantee that the transition probabilities observed within the AURORA 2 trial would be maintained over the remainder of the model.

The EAG explored two scenarios related to the long-term transition probabilities within the model. These scenarios make the implicit assumption that differences beyond 36 months are

driven by the patient health state occupancy at 36-months rather than the treatment arm i.e., a higher proportion of patients achieving CR on the VCS+MMF arm would still translate to a more favourable long-term outcome as the transition probabilities for progression of CR are more favourable than those patients with AD. The approach is slightly different between the two scenarios; scenario 12A assumes that VCS+MMF has the same long-term transition probabilities as MMF, which are derived from the MMF arm of the AURORA 2 data, while scenario 12B assumes that VCS+MMF has the same long-term transition probabilities as MMF, which are derived from averaging the VCS+MMF and MMF transition probabilities from AURORA 2 data.

6.2.12.1. Scenario 12A: the EAG assumed that long-term transition probabilities for voclosporin + MMF are the same as the long-term transition probabilities for MMF

The first scenario assumes that the point where patients are removed from voclosporin treatment within the model (36 months), thus transition probabilities thereafter are based on the MMF arm alone. This scenario could be considered conservative in the sense that it assumes there is no long-term treatment effect associated with voclosporin specifically in terms of the risk of achieving or losing response. However, the counter to this argument is that this scenario *does* in fact assume that there is a long-term effect of MMF which is applied beyond 36-months (despite the assumption that patients are no longer on treatment), as health state occupancy differs between the two arms at 36 months, and transition probabilities are a function of the current health state.

This scenario has a dramatic increase on the company's base case more than doubling the ICER (£18,897 to £46,412). This analysis indicates how sensitive the model results are to key structural uncertainties relating to the long-term transition probabilities within the model and the assumption that VCS+MMF not only maintains a level of treatment effect over time, but that this is maintained when patients are no longer receiving treatment.

6.2.12.2. Scenario 12B: the EAG assumed that the long-term transition probabilities for voclosporin+MMF and MMF are the same and the average is taken from AURORA-2

The second scenario considered by the EAG applied the average transition probabilities from the AURORA 2 study to both arms within the model (VCS+MMF and MMF). The EAG's understanding based on expert advice is that achieving and maintaining response is what is

important for patients, and response is what primarily drives progression through the model. As such pooling of the transition probabilities allows utilisation of the trial data in this way inherently assumes that the transition probabilities applied at 36-months are driven by health state occupancy rather than the individual treatment arms. Similar to scenario 12A, the impact of this scenario has a marked increase on the company base case ICER (increasing by £25,549 from £19,897 to £45,446), indicating just how sensitive the cost-effectiveness estimates are to the assumption that there is a long-term difference in the expected transitions for VCS+MMF versus MMF alone (which is not driven by the proportion of patients that achieved response).

6.2.13. Impact on the ICER of additional clinical and economic analyses undertaken by the EAG

The EAG made the changes described in Sections 6.2.1 to 6.2.12. Each change was made individually. The results of the EAG's exploratory analyses are provided in Table 31 for voclosporin+MMF versus MMF.

Table 31: EAG's exploratory analyses of voclosporin+MMF versus MMF

EAG assumption	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- company base case
EAG corrected company base-case	*6.1*	[redacted]	[redacted]	[redacted]	
Scenario 1: Utility values - use weighted average of Table B.3.10 (observed in AURORA 1 and 2) - EQ-5D by visit and status	*6.2.1*	[redacted]	[redacted]	[redacted]	*+£1,504*
Scenario 2A: Transplant utility - taken from Li et al. 2017	*6.2.2.1*	[redacted]	[redacted]	[redacted]	*+£255*
Scenario 2B: Dialysis utility - taken from meta-analysed dialysis values presented in Cooper 2020	*6.2.2.2*	[redacted]	[redacted]	[redacted]	*+£87*
Scenario 2C: Dialysis utility - Briggs et al. 2016 (presented in Cooper 2020) with the transplant increment from Li et al. 2017	*6.2.2.3*	[redacted]	[redacted]	[redacted]	*+£334*
Scenario 3: 1/2 additional pack of VCS for wastage	*6.2.3*	[redacted]	[redacted]	[redacted]	*+£516*
Scenario 4:2g dose of MMF applied to VCS+MMF and	*6.2.4*	[redacted]	[redacted]	[redacted]	*-£13*

EAG assumption	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- company base case
MMF (MMF for other regimens and subsequent treatments remain the same)					
Scenario 5 Additional monitoring for all CNI treatments	*6.2.5*	[redacted]	[redacted]	[redacted]	*+£965*
Scenario 6: Amend treatment, resource use and EOL costs within the model to match original source	*6.2.6*	[redacted]	[redacted]	[redacted]	*+£217*
Scenario 7A: Application of 95% RDI to treatments	*6.2.7.1*	[redacted]	[redacted]	[redacted]	*-£1,198*
Scenario 7B: Application of MMF TTD to other comparator treatments	*6.2.7.2*	[redacted]	[redacted]	[redacted]	*N/A**
Scenario 8: Restricted movement from transplant to dialysis: set to 0%	*6.2.8*	[redacted]	[redacted]	[redacted]	*+£460*
Scenario 9: Percentage reduction in transplantation rates from current value (43.77% per 6 months) – reduction to 23.08%	*6.2.9*	[redacted]	[redacted]	[redacted]	*-£371*
Scenario 10A: Removal of LN related death in CKD stage 1-3a from count method	*6.2.10*	[redacted]	[redacted]	[redacted]	*+£3,600*
Scenario 10B: Removal of LN related death in CKD stage 1-3a (CR, PR and AD removal from count method and additional model input capturing AD -> death in CKD stage 1-3a)	*6.2.6*	[redacted]	[redacted]	[redacted]	+£18,228
Scenario 11: Company setting: Model transitions: allow transitions to CKD 3b-5 in the first 36 months	*6.2.11*	[redacted]	[redacted]	[redacted]	*-£5,086*
Scenario 12A: Removal of long-term treatment effect for VCS+MMF (set transitions from 36 months equal to placebo)	*6.2.12.1*	[redacted]	[redacted]	[redacted]	*+£26,515*
Scenario 12B: Application of average transition probabilities from 36-months applied to both arms	*6.2.12.2*	[redacted]	[redacted]	[redacted]	*+£25,549*

Abbreviations: CKD, chronic kidney disease; CR, complete response; EAG, Evidence Assessment Group; EOL, end-of-life; ICER, incremental cost-effectiveness ratio; LN, lupus nephritis; MMF, mycophenolate mofetil; PR, partial response; QALY, quality adjusted life year; TTD, time-to-treatment discontinuation; VCS, voclosporin

Notes: * this does not affect the main comparison of voclosporin+MMF versus MMF but results have been presented within section 6.2.7 to understand the impact on the results of voclosporin+MMF versus other model comparators (and results are presented as part of the EAG preferred assumptions in a fully incremental format within section 6.3)

6.3. EAG's preferred assumptions

The EAG did not consider it possible to provide a preferred ICER that was able to address all of the described limitations/uncertainties inherent within the company's submitted model. This is largely because limitations pertinent to the model structure and uncertainty in the long-term transition probabilities could not be resolved. Despite this, the EAG has identified several alternative assumptions that are considered to represent a more suitable basis from which to understand the likely cost-effectiveness of voclosporin+MMF.

The tentative preferred base case ICER is £40,029 as shown in Table 32 below for voclosporin+MMF versus MMF. This table shows the cumulative change on the ICER for each change made within the model. The increase in the ICER is mostly driven by the removal of any long-term treatment differences associated with voclosporin+MMF and MMF alone.

Pairwise results of voclosporin+MMF versus all comparators when applying the EAG base case are presented in Table 33 with a full incremental provided in Table 34.

Table 32: EAG's preferred model assumptions

Preferred assumption	Section in EAG report	Cumulative ICER £/QALY
Company base-case		*£19,876*
Company base-case with fix applied		*£19,897*
Align resource use, AE, EOL and drug costs	*4.2.8*	*£20,114*
Add in ½ pack wastage for voclosporin	*4.2.8*	*£20,631*
Update trial utilities to weighted average from AURORA 1 and AURORA 2 observations	*4.2.7*	*£22,190*
Update literature-based utilities for transplant from Li et al.2017	*4.2.7*	*£22,496*
Update literature-based utilities for dialysis from meta-analysis of Cooper et al. 2020	*4.2.7*	*£22,603*
Apply 95% RDI to all treatments	*4.2.8*	*£21,291*
Stop LN death in CKD stage 1-3a	*4.2.6*	*£25,605*

Preferred assumption	Section in EAG report	Cumulative ICER £/QALY
Allow transitions CKD stage 3b-4 in first 36 months	*4.2.6*	*£18,488*
Use average long-term transition probabilities from VCS+MMF and MMF applied to both arms	*4.2.6*	*£40,029*

Abbreviations: EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Note: *The EAG prefers the incorporation of Scenario 12B (average across arms) over Scenario 12A (same as MMF); however, due to the need for additional functionality in the company's model to allow this scenario to be included, the EAG was unable to implement equivalent functionality to apply to the indirect comparators that feature in the fully-incremental analysis within the timeframe for preparing the EAG's report.

Table 33: EAG preferred analysis: pairwise comparison

	Total discounted costs	Total discounted QALYs	Incremental discounted costs versus VCS + MMF	Incremental discounted QALYs versus VCS + MMF	ICER versus VCS + MMF
EAG base case pairwise incremental results					
VCS + MMF	[redacted]	[redacted]	-	-	-
MMF	[redacted]	[redacted]	[redacted]	[redacted]	£40,029
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£8,743
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	£8,038
AZA	[redacted]	[redacted]	[redacted]	[redacted]	£14,555
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£29,958
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£16,550
TAC	[redacted]	[redacted]	[redacted]	[redacted]	£17,895

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALYs, quality adjusted life years; RTX, rituximab; TAC, tacrolimus; VCS, voclosporin

Note: *The EAG prefers the incorporation of Scenario 12B (average across arms) over Scenario 12A (same as MMF); however, due to the need for additional functionality in the company's model to allow this scenario to be included, the EAG was unable to implement equivalent functionality to apply to the indirect comparators that feature in the fully-incremental analysis within the timeframe for preparing the EAG's report.

Table 34: Full incremental analysis of voclosporin+MMF versus comparators: EAG preferred assumptions

Treatment	Total discounted costs	Total discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	ICERs (following re-baseline)
Company incremental base case (results taken by EAG from updated company CEM provided at clarification stage)					

Treatment	Total discounted costs	Total discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	ICERs (following re-baseline)
MMF	[redacted]	[redacted]	[redacted]	[redacted]	
AZA	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
TAC + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
TAC	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
L-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
H-CYC	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
RTX + MMF	[redacted]	[redacted]	[redacted]	[redacted]	Extendedly dominated
VCS + MMF	[redacted]	[redacted]	[redacted]	[redacted]	£40,029.31

Abbreviations: AZA = azathioprine; CEM, cost-effectiveness model; CYC, cyclophosphamide; H-CYC, high-dose CYC; L-CYC, low-dose CYC; MMF, mycophenolate mofetil; QALY, quality adjusted life year; RTX, rituximab; TAC, tacrolimus; VCS, voclosporin

Note: *The EAG prefers the incorporation of Scenario 12B (average across arms) over Scenario 12A (same as MMF); however, due to the need for additional functionality in the company's model to allow this scenario to be included, the EAG was unable to implement equivalent functionality to apply to the indirect comparators that feature in this analysis within the timeframe for preparing the EAG's report.

6.4. Conclusions of the cost-effectiveness section

6.4.1. The company's choice of model structure and approach to informing transition probabilities is subject to substantial uncertainty

Whilst the company's model broadly reflects the progression nature of CKD in an LN population, it is subject to several important structural limitations which restrict the ability to fully understand and interpret the uncertainty associated with the cost-effectiveness of voclosporin+MMF as a treatment for LN. These issues include the derivation of transition probabilities, the rigid model structure which forces patients to follow a certain trajectory (examples here include no CKD progression within 3 years, inability to achieve response in CKD stages 3b-4, inability to progress CKD stage for patients in CR and PR with earlier CKD stage 1-3a), the application of health state utility values and the long-term treatment effect assumptions associated. The EAG was only able to partially address some of the limitations in the company's model framework based on the information available

6.4.2. Several of the company's model inputs lacked transparency

As highlighted throughout the report, the EAG raised a number of concerns with respect to the transparency of model inputs, notable the cost inputs incorporated within the model. Owing to the fact that costs (whether uplifted or not), were included as inputs with limited description of their original source, the EAG has had to make assumptions when crosschecking the company's model with the referenced inputs.

6.4.3. The company's approach to analysing trial utilities was inappropriate and not fit for decision making

Importantly, the EAG considered that the company's approach to analysing trial utilities was wholly inappropriate and should not be used to inform decision making. Although the company used data collected from the AURORA trial programme to populate the utility values within the model, the approach used was considered to be methodologically wrong, with an assumption made which negated several months of informative HRQoL data. The EAG has substantial reservations in relation to the conduct of the utility analysis and recommends that a regression model should have been used to derive health state utility values.

6.4.4. The company's sensitivity analyses were subject to a number of limitations

Though the company provided scenario analysis associated with cost-effectiveness results, the EAG considered the analyses presented (CS Table B.3-25) to be uninformative and surface level, without inclusion of the larger more important structural issues within the model and hence preventing a clearer picture of true uncertainty associated with the decision problem under consideration. To illustrate this, only ten scenario analyses were presented, of which four related to adjusting the time horizons and varying the discount rates. No scenarios were presented which explored the impact of structural assumptions on the model such as allowing specific movements between health states, or alterations to the approaches taken to estimate transition probabilities. While utilities values were tested, only two scenarios were presented, a literature based analysis, and the exclusion of age-adjustment.

6.4.5. The EAGs tentative preferred base-case analysis yields an ICER in excess of £20,000 per QALY gained and is subject to substantial

structural uncertainty owing to limitations of the company's economic model that were not possible for the EAG to address

The EAG's preferred base-case analysis included several changes to the company's analysis in attempt to address limitations highlighted throughout the report. It should be emphasised that the EAG was not able to illustrate all uncertainty and limitations associated with the company's analysis and this was a result of the company's selected model structure alongside data availability. When considering the EAG's preferred settings, the changes resulting in slightly smaller total costs and fewer projected incremental QALYs gains. This resulted in an increase in the ICER by over 100% (from £19,876 estimated by the company to an EAG preferred base case of £40,029).

7. DISEASE SEVERITY

The company considered that the condition does not meet the criteria associated with a severity modifier and therefore did not present the calculation of the QALY shortfall in line with the new methods and processes. [9]

For completeness the EAG have assessed the appropriateness of a severity modifier by calculating the QALY shortfall using the Schneider et al. (2021) estimator tool. [38] This tool uses data from the Office of National Statistics (ONS) for England[39] to generate general population survival with various sources of data to inform utility estimates. The two are combined to estimate anticipated QALYs based on user inputted age of the patient population (assumed to be 33 from the company model) and percentage female in the patient population (assumed 87% rounded to the nearest integer from the company model). Using the company's modelled deterministic QALYs on the MMF arm (13.08) the QALY shortfall was estimated and is presented in Table 35. For further description of the methods used to estimate the QALY shortfall, the EAG refer to the NICE new methods manual[1] and the description of the references provided in Schneider et al. 2021. [2] The EAG are aligned with the company that the population does not meet the criteria associated with a severity modifier.

Table 35: Assessment of severity modifier by EAG

Alternative HRQoL norms provided in the Schneider et al. estimator tool	Absolute shortfall	Proportional shortfall	Corresponding QALY weight
Reference case: Hernandez Alava et al., EQ-5D-5L to 3L mapping + HSE 2017-2018	6.32	32.59%	x 1
Alternative A: van Hout et al., EQ-5D-5L to 4L mapping + HSE 2017-2018	6.45	33.02%	x 1
Alternative B: MVH, EQ-5D-3L value set + health state profiles	7.20	35.52%	x 1
Alternative C: MVH, EQ-5D-3L value set + HSE 2012+14	7.12	32.25%	x 1

Abbreviations: HRQoL, health-related quality of life; QALY, quality adjusted life-year.

References

1. Rees F, Doherty M, Grainge M, Davenport G, Lanyon P, Zhang W. The incidence and prevalence of systemic lupus erythematosus in the UK, 1999-2012. Ann Rheum Dis. 2016;75(1):136-41.

2. Almaani S, Meara A, Rovin BH. Update on Lupus Nephritis. Clin J Am Soc Nephrol. 2017;12(5):825-35.

3. Anders H-J, Saxena R, Zhao M-h, Parodis I, Salmon JE, Mohan C. Lupus nephritis. Nature reviews Disease primers. 2020;6(1):1-25.

4. Patel M, Clarke AM, Bruce IN, Symmons DP. The prevalence and incidence of biopsy-proven lupus nephritis in the UK: Evidence of an ethnic gradient. Arthritis Rheum. 2006;54(9):2963-9.

5. Tektonidou MG, Dasgupta A, Ward MM. Risk of End-Stage Renal Disease in Patients With Lupus Nephritis, 1971-2015: A Systematic Review and Bayesian Meta-Analysis. Arthritis Rheumatol. 2016;68(6):1432-41.

6. European Medicines Agency (EMA). Committee for medicinal products for human use (CHMP): Minutes for the meeting on 24-27 January 2022. 2022.

7. Fanouriakis A, Kostopoulou M, Cheema K, Anders HJ, Aringer M, Bajema I, et al. 2019 Update of the Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of lupus nephritis. Ann Rheum Dis. 2020;79(6):713-23.

8. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, et al. Technical Supplement to Chapter 4: Searching for and selecting studies. In: Higgins J, Thomas J, Chandler L, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions Version 63 (updated February 2022). 6.3: Cochrane; 2022.

9. Excellence NIfH. NICE health technology evaluations: the manual 2022.

10. European Medicines Agency (EMA). Guideline on clinical investigation of medicinal products for the treatment of systemic lupus erythematosus and lupus nephritis. Committee for Medicinal Products for Human use (CHMP); 2015.

11. Jourde-Chiche N, Costedoat-Chalumeau N, Baumstarck K, Loundou A, Bouillet L, Burtey S, et al. Weaning of maintenance immunosuppressive therapy in lupus nephritis (WIN-Lupus): results of a multicentre randomised controlled trial. Annals of the Rheumatic Diseases. 2022:annrheumdis-2022-222435.

12. Aurinia Pharmaceuticals I. Data and safety monitoring board charter for AURA-LV [confidential report]. 2015.
13. Excellence NIfH. Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template. 2022.
14. Aurinia Pharmaceuticals I. Data on file (CONFIDENTIAL). 2.7.3 Summary of Clinical Efficacy. 2021.
15. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011 (updated 2014).
16. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med. 2015;34(6):984-98.
17. Glanville J, Fleetwood K, Yellowlees A, Kaunelis D, Mensinkai S. Development and testing of search filters to identify economic evaluations in MEDLINE and EMBASE. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2009.
18. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ. 1996;313(7052):275-83.
19. Institute for C, Economic R. Lupus Nephritis An assessment of voclosporin and belimumab; Belimumab and Voclosporin for Lupus Nephritis: Effectiveness and Value. Final Report. 2021 2021/04/16/.
20. Aurora CSR. Final clinical study report – AUR-VCS-2016-01 (AURORA 1) a randomized, controlled double-blind study comparing the efficacy and safety of Orelvo (voclosporin) (23.7 mg twice daily) with placebo in achieving renal response in subjects with active lupus nephritis. Otsuka Pharmaceutical Europe Ltd. 2020.
21. Rovin BH, Teng YKO, Ginzler EM, Arriens C, Caster DJ, Romero-Diaz J, et al. Efficacy and safety of voclosporin versus placebo for lupus nephritis (AURORA 1): a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet. 2021;397(10289):2070-80.
22. consultation P. Comments on ICER's Draft Evidence Report - Aurinia Pharmaceuticals: Institute for Clinical and Economic Review; 2021. Available from: https://icer.org/wp-content/uploads/2020/11/ICER_Lupus-Nephritis_Public-Comments-Folio_031221.pdf.
23. Sugrue DM, Ward T, Rai S, McEwan P, van Haalen HGM. Economic Modelling of Chronic Kidney Disease: A Systematic Literature Review to Inform Conceptual Model Design. Pharmacoeconomics. 2019;37(12):1451-68.

24. Palmer AJ, Annemans L, Roze S, Lamotte M, Rodby RA, Bilous RW. An economic evaluation of the Irbesartan in Diabetic Nephropathy Trial (IDNT) in a UK setting. J Hum Hypertens. 2004;18(10):733-8.
25. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health and Quality of Life Outcomes. 2009;7(1):27.
26. University of Sheffield. SF-6D & SF-6Dv2 - Calculating QALYs from the SF-36, SF-12 and the standalone SF-6Dv2 2021. Available from: https://licensing.sheffield.ac.uk/product/SF-6D.
27. Jesky MD, Dutton M, Dasgupta I, Yadav P, Ng KP, Fenton A, et al. Health-Related Quality of Life Impacts Mortality but Not Progression to End-Stage Renal Disease in Pre-Dialysis Chronic Kidney Disease: A Prospective Observational Study. PLoS One. 2016;11(11):e0165675.
28. Lee AJ, Morgan CL, Conway P, Currie CJ. Characterisation and comparison of health-related quality of life for patients with renal failure. Current Medical Research and Opinion. 2005;21(11):1777-83.
29. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(4):539-45.
30. UK Renal Registry. 23rd Annual Report: Data to 31/12/2019. 2021.
31. Cooper JT, Lloyd A, Sanchez JJG, Sörstadius E, Briggs A, McFarlane P. Health related quality of life utility weights for economic evaluation through different stages of chronic kidney disease: a systematic literature review. Health and Quality of Life Outcomes. 2020;18(1):310.
32. Li B, Cairns JA, Draper H, Dudley C, Forsythe JL, Johnson RJ, et al. Estimating Health-State Utility Values in Kidney Transplant Recipients and Waiting-List Patients Using the EQ-5D-5L. Value Health. 2017;20(7):976-84.
33. National Health S. National Schedule of NHS Costs 2019-2020. 2021.
34. Jones K, Burns A. Unit Costs of Health and Social Care. Canterbury: Personal Social Services Research Unit (PSSRU), University of Kent; 2021.
35. Kerr M. Chronic Kidney Disease in England: The Human and Financial Cost 2012. Available from: https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Chronic-Kidney-Disease-in-England-The-Human-and-Financial-Cost.pdf.

36. Kidney Disease: Improving Global O. KDIGO 2021: Clinical Practice Guideline for the Management of Glomerular Diseases. Kidney Int. 2021;100(4S):S1–S276.

37. Jankowski J, Floege J, Fliser D, Böhm M, Marx N. Cardiovascular Disease in Chronic Kidney Disease: Pathophysiological Insights and Therapeutic Options. Circulation. 2021;143(11):1157-72.

38. Schneider P, McNamara S, Love-Koh J, Doran T, Gutacker N. QALY shortfall calculator. https://r4scharr.shinyapps.io/shortfall/2021.

39. Office for National S. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland (mid-2020 edition of this dataset 2021 local authority boundaries) 2021. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.

Appendix A: PSA output: cost-effectiveness planes voclosporin+MMF versus individual comparators

Figure 13: Cost-effectiveness plane for voclosporin+MMF vs. low dose CYC

Abbreviations: CYC, cyclophosphamide; MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.

Figure 14: Cost-effectiveness plane for voclosporin+MMF vs. high dose CYC

Abbreviations: CYC, cyclophosphamide; MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.

Figure 15: Cost-effectiveness plane for voclosporin+MMF vs. azathioprine

Abbreviations: MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.

Figure 16: Cost-effectiveness plane for voclosporin+MMF vs. rituximab+MMF

Abbreviations: MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.

Figure 17: Cost-effectiveness plane for voclosporin+MMF vs. tacrolimus+MMF

Abbreviations: MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.

Figure 18: Cost-effectiveness plane for voclosporin+MMF vs. tacrolimus

Abbreviations: MMF, mycophenolate mofetil, QALY, quality-adjusted life-year; vs, versus.