Dostarlimab for previously treated advanced or recurrent endometrial cancer with high microsatellite instability or mismatch repair deficiency [ID3802]

Produced by	*Warwick Evidence*
Authors	*Toyin Lamina, Independent Scientific Expert, Warwick Evidence, University of Warwick* *Ewen Cummins, Partner/Senior Researcher, McMDC Ltd* *Martin Connock, Honorary Senior Research Fellow, Warwick Evidence, University of Warwick* *Amin Mehrabian, Independent Scientific Expert, Warwick Evidence, University of Warwick* *Rachel Court, Information Specialist, Warwick Evidence, University of Warwick* *Paul Coleman, Honorary Research Fellow, Warwick Medical School, University of Warwick* *Rhona Johnston, Computer modeller, McMDC Ltd* *Emma Crosbie, Professor, Division of Cancer Sciences, University of Manchester* *Yen-Fu Chen, Associate Professor, Warwick Evidence, University of Warwick*
Correspondence to	*Dr Yen-Fu Chen, Warwick Medical School, University of Warwick* *Y-F.Chen@warwick.ac.uk*
Date completed	*05/10/2021 – ICERs updated for revised dostarlimab PAS and company technical engagement response*

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number 13/37/21.

Declared competing interests of the authors
Prof Emma Crosbie is Trustee of Peaches Womb Cancer Trust, which educates, raises awareness and supports survivors of endometrial cancer through social media, learning events and research. She served on a virtual endometrial cancer advisory board (not specifically convened for dostarlimab and the current appraisal) organised by GSK in June 2020 and was paid for her time. All other authors declare that they have no competing interests.

Acknowledgements
We are grateful to Dr Melanie Powell, Consultant Clinical Oncologist, St Bartholomew's Hospital for her expert advice on clinical issues, and to Daniel Gallacher, Research Fellow, Mubarak Patel, Research Assistant, and Angela Noufaily, Research Fellow, Warwick Medical School for their expert advice on statistical issues. The authors would like to acknowledge Dr Dan Todkill who independently quality assessed this report.

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

This report should be referenced as follows:
Lamina T, Cummins E, Connock M, Mehrabian A, Court R, Coleman P, Noufaily A, Johnston R, Crosbie E, Chen Y-F. Dostarlimab for previously treated advanced or recurrent endometrial cancer with high microsatellite instability or mismatch repair deficiency: A Single Technology Appraisal. Warwick Evidence. 2021

Contributions of authors
Toyin Lamina (Independent Researcher) led and Amin Mehrabian (Independent Researcher) supported the critique of clinical effectiveness evidence; Ewen Cummins (Partner/Senior Researcher) and Rhona Johnston (Computer Modeller), supported by Martin Connock (Honorary Senior Research Fellow), reviewed and critiqued the cost-effectiveness evidence and the company model and undertook additional analyses; Paul Coleman (Public Health Specialty Registrar) provided clinical summary and critique of clinical effectiveness evidence and end of life criteria; Rachel Court (Information Specialist) critiqued the company searches and undertook additional searches Yen-Fu Chen (Associate Professor) critiqued the decision problem and indirect comparisons, co-ordinated the project and the report; Emma Crosbie (Professor and Honorary Consultant) provided expert clinical advice and critical comments. All authors contributed towards and commented on drafts of this report and approved the final report.

Please note that: Sections highlighted in [redacted] are '[redacted]. Sections highlighted in [redacted]. Figures that are CIC have been bordered with blue.

Table of Contents

Table of Tables

Table of Issues

Table of Figures

Abbreviations

AE	Adverse event
AUC	Arena under the curve
BGCS	British Gynaecological Cancer Society
BICR	Blinded independent central review
BNF	British National Formulary
BOR	Best overall response
BSA	Body surface area
BSC	Best standard of care
CASP	Critical Appraisal Skills Programme
CDF	Cancer Drugs Fund
CHMP	Committee for Medicinal Products for Human Use

CI	Confidence interval
CR	Complete response
CS	Company submission. Unless otherwise specified, CS section numbers, tables and figures cited in this ERG report refer to those of Document B of the company submission.
CTCAE	Common Terminology Criteria for Adverse Events
CxDx	Cycle X Day X
DCO	Data cut off
DCR	Disease control rate
DG	Diagnostics guidance
dMMR	DNA mismatch repair deficient
DOR	Duration of response
DOST	Dostarlimab
DSA	Deterministic sensitivity analysis
DSU	Decision support unit
EC	Endometrial cancer
ECOG PS	Eastern Cooperative Oncology Group performance score
eGFR	Estimated glomerular filtration rate
EMA	European Medicines Agency
EORTC QLQ-C3-	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
EOT	End-of-treatment
EPAR	European Public Assessment Report
EQ-5D	EuroQoL 5-dimensions 5-levels
ESGO	European Society of Gynecological Oncology
ESMO	European Society for Medical Oncology Annual Meeting

ESP	European Society of Pathology
ESS	Effective sample size
ESTRO	European Society for Radiotherapy and Oncology
FIGO	International Federation of Gynaecology and Obstetrics
HR	Hazard ratio
HR+	Hormone receptor positive
HRQoL	Health-related quality of life
IA	Interim analysis
ICER	Incremental cost-effectiveness ratio
IHC	Immunohistochemistry
I-O	Immuno-oncology
IPD	Individual patient data
IPTW	Inverse probability treatment weighting
IQR	Interquartile range
ITC	Indirect treatment comparison
ITT	Intention-to-treat
KM	Kaplan-Meier
LYG	Life-years gained
MAIC	Matching-adjusted indirect comparison
MHRA	Medicines & Healthcare products Regulatory Agency
MMR	DNA mismatch repair
MSI-H	Microsatellite instability-high
MSS	Microsatellite stable
NCRAS	National Cancer Registry Analysis System
NHS(E)	National Health Service (England)
NNT	Numbers needed to test

NR	Not reported
ORR	Objective response rate
OS	Overall survival
PAS	Patient access scheme
PD	Progressive disease
PD-L1/2	Programmed death-ligand 1/2
PFS	Progression-free survival
PH	Proportional hazards
PLD	Pegylated liposomal doxorubicin
pMMR	DNA mismatch repair proficient
PPS	Post-progression survival
PR	Partial response
PRO	Patient reported outcome
PSA	Probabilistic sensitivity analysis
PSM	Partitioned survival model
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QXW	Once every X weeks
QALY	Quality-adjusted life-years
QOL	Quality of life
RCT	Randomised controlled trial
RECIST	Response evaluation criteria in solid tumours
RWE	Real-world evidence
RWEQ	Real World Evidence Equivalent dataset. This abbreviation is adopted in this report as a more concise way to refer to the main

	comparator patient population/dataset chosen by the company, the GARNET-like Real World Evidence dataset (n=[redacted])
SACT	Systemic Anti-Cancer Therapy (SACT) dataset
SAE	Serious adverse event
SD	Stable disease
SLR	Systematic literature review
STC	Simulated treatment comparisons
STD	Standard deviation
TEAE	Treatment emergent adverse event
ToT	Time on treatment
TTD	Time to discontinuation
TTNT	Time to next treatment
VAS	Visual analogue scale

Executive Summary

1 Executive summary

This summary provides a brief overview of the key issues identified by the evidence review group (ERG) as being potentially important for decision making. It also includes the ERG'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 ERG report.

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

1.1 *Overview of the ERG's key issues*

Table 1: Summary of key issues

Issues	Summary of issue	Report sections
Issue 1	The patient population specified in marketing authorisation and addressed in the company submission (CS) is narrower that what is specified in the final scope	2.3
Issue 2	Patients with advanced disease and with recurrent disease are potentially two distinct populations, but they were identified in different ways between the GARNET trial for dostarlimab and the GARNET-like Real World EQuivalent (RWEQ) cohort	2.3 & 3.3.1.1
Issue 3	Overall the GARNET trial data were fairly immature and may not be sufficient to provide reliable effectiveness and cost-effectiveness estimates	3.2
Issue 4	There are uncertainties over the magnitude of the benefit of dostarlimab relative to comparators due to the single-arm design of the GARNET trial and lack of suitable data for comparator treatments	3.3
Issue 5	GARNET trial population and RWEQ cohort may have fundamental differences that cannot be easily adjusted statistically	3.3.1

Issues	Summary of issue	Report sections
Issue 6	Does the model contain a number of errors, in particular with regards to the waning of the dostarlimab treatment effect after cessation of treatment?	4.3.1
Issue 7	Is the company elicitation exercise for dostarlimab overall survival (OS) mainly relevant to the curves unadjusted for treatment waning, and what does this imply for the choice of the adjusted OS curve?	4.2.6.5 4.3.3.2
Issue 8	Does the company elicitation exercise for current treatment OS suggest that the RWEQ OS data and curves are too pessimistic?	4.2.6.3
Issue 9	Is the company elicitation exercise for dostarlimab treatment discontinuation and waning of treatment effect biased, and if so what does this imply for the values that should be applied?	4.3.3.3
Issue 10	Is the company choice of dostarlimab time to treatment discontinuation (TTD) curve appropriate or would the better fitting Gompertz have been the more natural choice? Is the ERG estimated intention to treat (ITT) TTD generalized gamma a better choice?	4.3.3.8
Issue 11	GARNET had a lot of censoring, quite a lot of which was early censoring. The RWEQ data has much less censoring. Might poorly performing patients have dropped out of GARNET early and if they did how might this have affected results?	4.3.3.5
Issue 12	For the ICERs for dostarlimab compared to individual treatments, does the difference in effect when using RWEQ data compared to when using values within the literature raise questions about the reliability of using the RWEQ data?	4.3.3.6 5.2 6.2

A key difference between the company's preferred assumptions and the ERG's preferred assumptions is whether there are errors in the model implementation. The ERG reports the company base case ICER of £37,311[1] per quality adjusted life year (QALY), but for most of its commentary of Chapter 4 it references the ERG corrected company base case ICER of £49,190 per QALY.

The other main differences between the company and the ERG are:

[1] Updated for the revised PAS of [redacted] presented during technical engagement rather than the [redacted] of the original company submission.

- Should overall survival for dostarlimab be modelled using the generalised gamma or the Weibull?
- Should the dostarlimab time to treatment discontinuation be modelled using the company log-logistic, the company Gompertz or the ERG ITT generalised gamma?
- Is it most reasonable that all but [redacted] of dostarlimab patients will cease or have treatment withdrawn at the [redacted] point, or is [redacted] more reasonable? Would a treatment withdrawal cliff edge be applied in practice?
- When patients have dostarlimab treatment withdrawn, is it reasonable to assume that the full benefits of treatment will be retained for [redacted] or is it more reasonable to assume that there will be some loss of effect, albeit small, from when treatment is withdrawn?

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 QALY. An ICER is the ratio of the extra cost for every QALY gained.

The company estimates the effects of current treatment and dostarlimab as per Table 3, with an [redacted] corresponding central estimate from the probabilistic modelling is £35,492 per QALY.

Table 2: Summary of the company base case

	Current treatment	Dostarlimab	Net
Life years QALYs Costs	[redacted]	[redacted]	[redacted]
ICER			£37,311

Note that the life years reported above are undiscounted, while QALYs and costs are discounted at 3.5%. The ERG applies this convention throughout this report.

Further note that during clarification the company supplied an updated base case due to a slight expansion of the quality of life data set that very slightly worsens its base case, together with a corresponding set of scenario analyses. The ERG has not incorporated these in its report due to time constraints. The ERG thinks the revision

is sufficiently minor to be unlikely to affect Committee deliberations. The ERG revised base case incorporates the change to the quality of life data set.

The company univariate sensitivity analyses find that the ICER is most sensitive to: the baseline quality of life, the quality of life for the main health states of the model and the cost per cycle of dostarlimab.

The company performs a number of comparisons with individual treatments using hazard ratios derived from the company's MAICs based on comparator effectiveness data from the literature.

[redacted]

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

Issue 1: The patient population specified in marketing authorisation and addressed in the company submission (CS) is narrower that what is specified in the final scope

Report section	2.3
Description of issue and why the ERG has identified it as important	The population specified in marketing authorisation and addressed in the CS is required to "have progressed on or following prior treatment with a platinum-containing regimen" rather than simply "previously treated" endometrial cancer (EC) as described in the final scope.
What alternative approach has the ERG suggested?	No alternative approach is required. CS highlighted this difference and ERG critiqued and interpreted the submitted evidence accordingly.
What is the expected effect on the cost-effectiveness estimates?	This issue impacts on applicability (generalisability) of clinical effectiveness and cost-effectiveness estimates.
What additional evidence or analyses might help to resolve this key issue?	No additional evidence or analyses are required. This issue is flagged up to highlight the specific patient population to which the evidence submitted by the company and critiqued by the ERG can be applied.

Issue 2: Patients with advanced disease and with recurrent disease are potentially two distinct populations, but they were identified in different ways between the GARNET trial for dostarlimab and the GARNET-like Real World Evidence equivalent (RWEQ) cohort

Report section	2.3 & 3.3.1.1
Description of issue and why the ERG has identified it as important	The nature of tumour, and hence prognosis, may differ between patients with advanced disease and those with recurrent disease. These two groups of patients also have different treatment histories, which may impact on the response to treatment. While recurrent disease in the GARNET trial was confirmed by radiographic evidence, 'probable' recurrent disease could only be retrospectively identified based on treatment history without supporting radiographic evidence in the GARNET-like real-world evidence equivalent (RWEQ) cohort (which is the main comparator chosen by the company) due to limitations in registry data. This difference may impact on the comparability of patients between GARNET and RWEQ, and may confound the comparison between treatments.
What alternative approach has the ERG suggested?	ERG requested data stratified by advanced vs recurrent disease for both the GARNET and the RWEQ cohorts in ERG's clarification questions. However the company did not provide data for either cohort. The company explained that recurrent and advanced diseases were mentioned in the same inclusion criterion and were not recorded separately in the GARNET trial, and stated that "further subgroup analyses of the licensed dostarlimab indication were not included in the NICE final scope and should not be considered relevant to this appraisal" (company response to ERG clarification question A2).
What is the expected effect on the cost-effectiveness estimates?	The direction and magnitude of the expected effect is not clear, but potential differences in the characteristics and composition between the two cohorts with respect to advanced vs recurrent diseases may confound clinical effectiveness estimates and directly impact on cost-effectiveness estimates.
What additional evidence or analyses might help to resolve this key issue?	Although recurrent and advanced diseases were not separately recorded in the GARNET trial, it should be possible to adopt the same definition of advanced disease being FIGO stage III & IV at diagnosis (or at treatment initiation) and then classify remaining patient groups as recurrent. Comparison between GARNET and RWEQ can then be carried out between the better defined 'advanced disease' groups and the less well-characterised 'recurrent disease' groups (which were defined and identified in different ways) to verify the sources of heterogeneity in the patient characteristics observed between the two cohorts and to explore whether these might have a bearing on observed outcomes.

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

Issue 3: Overall the GARNET trial data were fairly immature and may not be sufficient to provide reliable effectiveness and cost-effectiveness estimates

Report section	3.2
Description of issue and why the ERG has identified it as important	With a medium duration of follow up of [redacted] months and median overall survival time not yet reached, the key effectiveness data for dostarlimab were immature and longer-term effectiveness unknown.
What alternative approach has the ERG suggested?	No alternative seems to be possible within the current appraisal.
What is the expected effect on the cost-effectiveness estimates?	The substantial uncertainties in longer-term effectiveness directly contribute to substantial uncertainties in cos-effectiveness estimates
What additional evidence or analyses might help to resolve this key issue?	Data from longer follow-up might resolve this issue. The committee might consider the option for use within the Cancer Drug Fund while longer-term data are accrued to reduce uncertainties.

Issue 4: There are uncertainties over the magnitude of the benefit of dostarlimab relative to comparators due to the single-arm design of the GARNET trial and lack of suitable data for comparator treatments

Report section	3.1, 3.2 and 3.4
Description of issue and why the ERG has identified it as important	GARNET was a single arm, phase I trial with no comparator. Relative effectiveness needs to be estimated through unanchored indirect comparison. The company made substantial effort in identifying different sources of comparator evidence and undertook a series of matching adjusted indirect comparisons (MAICs), but all of the MAICs were susceptible to bias due to limitations in available data and methods of MAICs.
What alternative approach has the ERG suggested?	Improving some of the MAICs may reduce potential bias (see further Key issues below) but may not eliminate residual confounding, the direction and magnitude of which is difficult to estimate.
What is the expected effect on the cost-effectiveness estimates?	The expected impact on the cost-effectiveness estimates varies depending on individual comparators and MAICs, but may be difficult to estimate because of confounding by indication for different comparators.
What additional evidence or analyses might help to resolve this key issue?	Data on real world use of dostarlimab, possibly collected from a randomized controlled trial (RCT) of dostarlimab versus current clinical management may be needed.

Issue 5: GARNET trial population and RWEQ may have fundamental differences that cannot be easily adjusted statistically

Report section	3.3.1
Description of issue and why the ERG has identified it as important	There are uncertainties around the process used to derive the GARNET-like Real World Evidence equivalent (RWEQ) cohort from the patients with EC diagnosis in the registry and the representation of the UK population. There are major differences in setting, patient characteristics and case definitions between the GARNET trial population and the RWEQ cohort, which was chosen by the company as the main comparator for the base case. There is uncertainty regarding the approaches the company undertook to align the data. The MAIC conducted by the company for GARNET vs RWEQ did not take into account some important prognostic factors and had many methodological issues. There are reservations regarding the validity of the findings from the MAICs.
What alternative approach has the ERG suggested?	In order to characterise the differences between GARNET and RWEQ and to identify potentially more comparable patients between the cohorts, ERG requested data stratified by advanced vs recurrent diseases, and by endometrioid vs other diseases for both cohorts in ERG's clarification questions. However, no data were provided.
What is the expected effect on the cost-effectiveness estimates?	The differences between GARNET and RWEQ are likely to result in effectiveness and cost-effectiveness estimates that are biased in favour of dostarlimab.
What additional evidence or analyses might help to resolve this key issue?	The following analyses may reduce the magnitude of potential bias: • Examination of RWEQ patients with known mismatch repair deficiency (dMMR) status at the time of diagnosis to compare similar tumour biology. • Analyses focusing on more homogeneous groups of patients, e.g. endometrioid disease or advanced disease (which could be operationally defined as International Federation of Gynaecology and Obstetrics [FIGO] stage III and IV). • Comparison of GARNET with subset(s) of patients receiving combination regimens in the RWEQ who might represent fitter patients similar to those recruited in trial settings. • Consider the use alternative sources of more comparable data (such as ZoptEC trial) as primary analyses for base case.

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

Issue 6: Model errors

Report section	4.3.1
Description of issue and why the ERG has identified it as important	There appear to be modelling errors, particularly the implementation of the waning of treatment effect.
What alternative approach has the ERG suggested?	The company applies hazard ratios. But the RWEQ curves of the base case are based upon the RWEQ parameterised curves. Equalising the risk of events between the arms requires that the risk of events in the RWEQ arm be used. There are a number of other more minor modelling errors.
What is the expected effect on the cost-effectiveness estimates?	The company base case ICER of £37,311 per QALY worsens to £49,190 per QALY. Given the importance of this, the ERG thinks that the £37,311 ICER is no longer relevant. For the ERG critique of Chapter 4 the ERG presents the effects that its other changes have upon the ERG corrected company base case ICER of £49,190 per QALY.
What additional evidence or analyses might help to resolve this key issue?	None.

Issue 7: Dostarlimab overall survival (OS) elicitation exercise and choice of OS curve

Report section	4.2.6.5 & 4.3.3.2
Description of issue and why the ERG has identified it as important	The elicitation exercise concentrated upon the unadjusted curves. The ERG thinks that this means it provides values for the unadjusted curves but it is not a good basis for selecting the adjusted curves.
What alternative approach has the ERG suggested?	The ERG prefers the company OS Weibull over the company OS log-logistic.
What is the expected effect on the cost-effectiveness estimates?	This worsens the ERG corrected company base case ICER from £41,190 per QALY to £65,262 per QALY.
What additional evidence or analyses might help to resolve this key issue?	None.

Issue 8: RWEQ OS elicitation exercise and choice of OS curve

Report section	4.2.6.3
Description of issue and why the ERG has identified it as important	The company sponsored experts were also asked about OS at 5, 10 and 15 years under current therapy. Their responses suggest that the curves fitted to the RWEQ data extrapolate too low an OS at 5, 10 and 15 years. The RWEQ data may be poorly aligned with the GARNET population. The OS for the individual treatments within the RWEQ are also hugely different from one another, those receiving combination therapies performing much better than those receiving monotherapies. Aggregating the RWEQ patients into a single treatment groups may not be sensible.
What alternative approach has the ERG suggested?	None.
What is the expected effect on the cost-effectiveness estimates?	If the RWEQ OS underestimates what OS is with current therapy the ICER is biased in favour of dostarlimab.
What additional evidence or analyses might help to resolve this key issue?	GARNET and RWEQ baseline and KM data split by endometrioid disease, Eastern Cooperative Oncology Group (ECOG) performance status and if possible recurrent disease status, possibly drilling down into individual treatments as well.

Issue 9: Dostarlimab time to treatment discontinuation (TTD) elicitation exercise and treatment discontinuations

Report section	4.3.3.3
Description of issue and why the ERG has identified it as important	The elicitation exercise appears to have presented the TTD numbers remaining at risk rather than the TTD KM survival curve. If so, this would seriously bias the presentation. The company sponsored experts were not asked open ended questions but in the main were asked to confirm the company preferences. The company does not present evidence that there will be no loss of effect for any patients for [redacted] after treatment cessation. The ERG thinks that the TTD elicitation exercise is probably biased and at best yields a floor for the treatment cessation percentage. The ERG also thinks that it is more reasonable to assume that some, albeit small, treatment waning will start from treatment cessation.
What alternative approach has the ERG suggested?	The ERG assumes that at [redacted] the proportion remaining on treatment will fall to [redacted], treatment waning will occur over the next [redacted] and all will cease treatment at [redacted].
What is the expected effect on the cost-effectiveness estimates?	This worsens the ERG corrected company base case ICER from £49,190 per QALY to £60,362 per QALY.
What additional evidence or analyses might help to resolve this key issue?	None.

Issue 10: Dostarlimab choice of TTD curve

Report section	4.3.3.8
Description of issue and why the ERG has identified it as important	The company selected the company log-logistic but the company Gompertz has superior information criteria. Due to the treatment cessation assumptions, the dostarlimab TTD curve does not require much extrapolation. The choice of curve can be based upon the internal goodness of fit.
What alternative approach has the ERG suggested?	The ERG prefers the company Gompertz over the company log-logistic due to its better information criteria. The ERG prefers the ERG ITT generalized gamma.
What is the expected effect on the cost-effectiveness estimates?	The company Gompertz worsens the ERG corrected company base case ICER from £49,190 per QALY to £51,804 per QALY. The ERG ITT generalized gamma worsens the ERG corrected company base case from £49,190 per QALY to £52,548 per QALY.
What additional evidence or analyses might help to resolve this key issue?	None.

Issue 11: Censoring and the possibility of informative censoring

Report section	4.3.3.5
Description of issue and why the ERG has identified it as important	There is much higher censoring in GARNET than in the RWEQ data. If this pattern of censoring was observed in a two-arm trial it would be a major concern. Quite a lot of patients in GARNET were censored early in the trial.
What alternative approach has the ERG suggested?	None.
What is the expected effect on the cost-effectiveness estimates?	If censoring in GARNET for reasons other than data cut - off did not occur at random but was in part associated with other factors such as patient baseline characteristics, patient response or patient disease type, the analysis would be biased in favour of dostarlimab.
What additional evidence or analyses might help to resolve this key issue?	GARNET patient baseline characteristics and KM data, with censoring events divided into those due to data cut-off and those for other reasons, split by best response.

1.6 *Other key issues: summary of the ERG's view*

Issue 12: Reliability of comparing GARNET with the RWEQ

Report section	4.3.3.6, 5.2 & 6.2
Description of issue and why the ERG has identified it as important	The ERG has performed exploratory cost effectiveness analyses by fitting curves to some of the RWEQ individual treatment KM data. This suggests that dostarlimab has a somewhat worse ICER when compared to the combination therapies. It also suggests that dostarlimab has a somewhat better ICER when compared to pegylated liposomal doxorubicin (PLD) monotherapy. But the company cost effectiveness estimate for dostarlimab against doxorubicin that used the ZoptEC trial has an ICER that is worse than the company base case ICER. This raises questions about the reliability of the comparison with the RWEQ data and whether it biases the analysis in favour of dostarlimab.
What alternative approach has the ERG suggested?	None.
What is the expected effect on the cost-effectiveness estimates?	If GARNET has tended to recruit fitter patients or patients whose disease has a better prognosis than the RWEQ patients or there is a trial or placebo effect within GARNET the analyses will be biased in favour of dostarlimab.
What additional evidence or analyses might help to resolve this key issue?	GARNET and RWEQ KM data split by endometrioid status and ECOG performance status. RCT data.

1.7 *Summary of ERG's preferred assumptions and resulting ICER*

The ERG corrected company base case is as per Table 3, with an ICER of £49,190 per QALY. The corresponding central ICER of the probabilistic modelling is £48,764 per QALY.

Table 3: Summary of the ERG corrected company base case

	Current treatment	Dostarlimab	Net
Life years QALYs Costs	[redacted]	[redacted]	[redacted]
ICER			£49,190

The ERG preferred assumptions are outlined in Table 4.

Table 4: ERG preferred assumptions and model inputs

Preferred assumption	Section	ICER
Company base-case	5.1	£37,311
ERG corrected company base-case	4.3.1	£49,341
ERG01: Dostarlimab OS Weibull	4.2.6.5 4.3.3.2	£65,454
ERG02: Dostarlimab ERG ITT TTD GGAM	4.3.3.8	£52,709
ERG03: [redacted] dostarlimab continue	4.3.2.1 4.3.3.3	£49,341
ERG04: Waning from treatment cessation	4.3.2.1 4.3.3.3	£55,523
ERG05: Quality of life – no time to death coefficient	4.3.4.1	£49,513
ERG06: Ongoing resource use	4.3.4.6	£48,885
Cumulative effect: ERG02-ERG06	..	£64,006
Cumulative effect: ERG01-ERG06	..	£79,714

ERG: evidence review group; GGAM: generalised gamma; ICER: incremental cost-effectiveness ratio; ITT: intention to treat; OS: overall survival; TTD: time to treatment discontinuation

The ERG presents a range of scenario analysis.

- SA01: Assuming dostarlimab treatment cessation from [redacted] and from [redacted], retaining the assumption that all cease treatment at [redacted].
- SA02: Assuming proportions remaining on dostarlimab at [redacted] of [redacted] and [redacted].
- SA03: Assuming treatment waning starts [redacted] and [redacted] after the treatment cessation at [redacted].
- SA04 Applying the company Gompertz and company log-logistic dostarlimab TTD curves.
- SA05: Applying the dostarlimab TTD KM curve for the first 8 months of the model.
- SA06: Applying the quality of life values of the German study: PFS 0.701 and PPS 0.676.

- SA07: Applying a correction factor to the RWEQ treatment costs to align the modelled treatment duration with the mean stated by the company.
- SA08: Reducing the frequency of visits to the specialist nurse when in PFS off treatment to 12 weekly.
- SA09: Time horizons of 10, 20 and 30 years.

Given the importance of the choice of dostarlimab OS curve, the ERG scenario analyses are presented for the ERG preferred Weibull OS curve and for the company preferred generalised gamma OS curve.

Table 5: ERG scenario analyses

	ICER	
	Weibull	**Gen.Gamm.**
Base case	£79,714	£64,006
SA01a: Cessation at [redacted]	£81,853	£63,583
SA01b: Cessation at [redacted]	£83,990	£63,140
SA02a: [redacted] dostarlimab [redacted] continuing treatment	£73,411	£59,041
SA02b: [redacted] dostarlimab [redacted] continuing treatment	£83,336	£66,859
SA03a: Waning starts [redacted] after [redacted] cessation	£77,378	£60,153
SA03b: Waning starts [redacted] after [redacted] cessation	£75,813	£57,082
SA04a: Dostarlimab Gompertz TTD curve	£80,921	£64,733
SA04b: Dostarlimab log-logistic TTD curve	£75,198	£60,225
SA05: Dostarlimab KM TTD for 8 months	£75,457	£60,429
SA06: German QoL values	£79,263	£63,465
SA07: RWEQ treatment cycles adjustment	£80,083	£64,296
SA08: Reduced specialist nurse frequency	£79,290	£64,170
SA09a: 10 year time horizon	£90,563	£74,322
SA09b: 20 year time horizon	£81,822	£65,962
SA09c: 30 year time horizon	£79,911	£64,186

KM: Kaplan Meier; ICER: incremental cost-effectiveness ratio; TTD: time to treatment discontinuation

The ERG also analyses the RWEQ individual treatment data, which results in the cost effectiveness estimates of Table 6. There is a modelling issue as to whether dostarlimab treatment waning should be based upon the pooled RWEQ curves or upon the individual comparator curves. The ERG thinks that it should be based upon the individual comparator curves.

Table 6: ERG scenario analyses: Individual treatment comparisons

Waning	RWEQ curves used			Comparator curves used		
Comparator	Δ QALY	Δ Cost	ICER	Δ QALY	Δ Cost	ICER
Carb+Pac	[redacted]	[redacted]	£104k	[redacted]	[redacted]	£108k
Carb+PLD	[redacted]	[redacted]	£88,929	[redacted]	[redacted]	£102k
PLD mono	[redacted]	[redacted]	£53,080	[redacted]	[redacted]	£58,120
Carb+Pac: Carboplatin + paclitaxel, Carb+PLD: Carbplatin + PLD, PLD mono: PLD monotherapy						

Evidence Review Group Report

2 INTRODUCTION AND BACKGROUND

2.1 *Introduction*

This Evidence Review Group (ERG) report provides a detailed critique of the company submission (CS) presented to the National Institute for Health and Care Excellence (NICE) for the single technology appraisal (STA) on dostarlimab for previously treated advanced or recurrent endometrial cancer with high microsatellite instability or mismatch repair deficiency. Related NICE STAs include the currently suspended pembrolizumab for previously treated endometrial cancer NICE technology appraisal [ID1205][1] and the ongoing lenvatinib with pembrolizumab for previously treated advanced endometrial cancer NICE technology appraisal [ID3811].[2]

Dostarlimab has been approved by the European Medicines Agency (EMA) in April 2021.

The ERG noted that the positive opinion from the Committee for Medicinal Products for Human Use (CHMP) of the EMA was for a conditional marketing authorisation, which is granted for a medicine that *"fulfils an unmet medical need when the benefit to public health of immediate availability outweighs the risk inherent in the fact that additional data are still required. The marketing authorisation holder is expected to provide comprehensive clinical data at a later stage*."[3]

The primary evidence that supported the conditional marketing authorisation and that forms the key part of clinical effectiveness evidence for dostarlimab in the CS for this STA is data from a single arm phase I trial, GARNET (Clinical Study Report dated July 2019 provided with the CS, and Oaknin et al. 2020[4]). Comparative effectiveness and cost-effectiveness between dostarlimab and current practice in the National Health Services (NHS) therefore need to be derived from indirect comparisons between data from GARNET trial and those sourced from elsewhere. Consequently, ERG's critique focuses on the limited volume of clinical evidence, the validity of indirect comparisons based on the evidence and associated

uncertainties related to the findings, and the derivation of cost-effectiveness estimates based on these.

2.2 *Background*

2.2.1 Endometrial cancer

The company provided detailed overview of endometrial cancer in CS Section B.1.3.1. Endometrial cancer forms the vast majority (94%) of uterine cancer, which is the 4th most common cancer and accounts for 5% of all new cancer cases in female in the UK.[5] The incidence of endometrial cancer peaks among 75-79 age group, with a total of 9,494 incident cases diagnosed in 2017 and an estimated prevalence of 70,200 women who had been diagnosed between 1991 and 2010 being still alive at the end of 2010.[5] Around 80% of uterine cancer are diagnosed at an early stage (I/II), with older age associated with late stage diagnosis.[5]

2.2.2 Marketing authorisation for dostarlimab

Dostarlimab is approved by the EMA "*as monotherapy for the treatment of adult patients with mismatch repair deficient (dMMR)/microsatellite instability-high (MSI-H) recurrent or advanced endometrial cancer (EC) that has progressed on or following prior treatment with a platinum-containing regimen*".[6]

The targeting of patients with dMMR/MSI-H reflects the proposed mechanism of action for dostarlimab, which is an inhibitor of programmed cell death protein (PD-1) that is implicated in preventing immune cells to kill cancer cells.

2.2.3 Classification

EC has traditionally been classified into two types (Type I and Type II) proposed by Bokhman based on endocrine and metabolic features.[7] Classifications based on histology and molecular features of the tumours have subsequently been incorporated into the classification as described below. While it is acknowledged that the dualistic classification does not fully reflect the heterogeneity of EC which has become apparent with more recent knowledge particularly in genetic epidemiology,[8] the classification is a commonly used prognostic factor, the information of which is relatively easy to obtain.

Type I EC is moderately or well differentiated tumours associated with oestrogen excess, obesity, hormone receptor positivity and endometrial hyperplasia. Type I EC typically includes grade 1 and grade 2 endometrioid EC and constitutes around 60-70% of cases, who tend to have better prognosis.[9]

Type II EC is poorly differentiated tumours with low or absence of hormone receptor expression and is associated endometrial atrophy. Type II EC typically includes serous and clear-cell carcinoma and constitutes around 30-40% of cases, who tend to have worse prognosis.

2.2.4 Staging

Currently the most widely used staging for endometrial cancer is the classification proposed by the International Federation of Gynecology and Obstetrics (FIGO) in 2009[10] as shown in Table 7 below.

Table 7: FIGO staging for endometrial cancer

Stage	Description
I	**Tumour confined to the corpus uteri (i.e. the body of the uterus, or the womb)**
IA	No or less than half myometrial invasion
IB	Invasion equal to or more than half of the myometrium (middle layer of the uterine wall consisted mainly of muscle cells)
II	**Tumour invades cervical stroma (dense, fibromuscular tissue through which vascular, lymphatic, and nerve supplies to the cervix pass), but does not extend beyond the uterus**
III	**Local and/or regional spread of the tumour**
IIIA	Tumour invades the serosa of the corpus uteri (outer layer of uterus) and/or adnexae (the region adjoining the uterus that contains the ovary and fallopian tube)
IIIB	Vaginal and/or parametrial involvement (connective tissue that surrounds the uterus and connect the uterus to other tissues in the pelvis)
IIIC	Metastases to pelvic and/or para-aortic lymph nodes
IV	**Tumour invades bladder and/or bowel mucosa, and/or distant metastases**
IVA	Tumour invasion of bladder and/or bowel mucosa
IVB	Distant metastases, including intra-abdominal metastases and/or inguinal lymph nodes

Adapted from Pecorelli 2009.[10]

2.2.5 Prognostic factors

The company conducted a targeted literature review in which published literature reviews of prognostic factors associated with survival in EC were identified by Google searches (CS Appendix M). Adjustment for all potential prognostic factors and effect modifiers is required to minimise bias for unanchored indirection comparison that relies on data from individual arms from different studies as in this STA. Details regarding this are discussed in Section 3.4 of this report.

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

ERG's critique of company's definition of decision problem is shown in Table 8.

The population addressed in the CS is narrower than what was specified in the final scope in that patients are required to "have progressed on or following prior treatment with a platinum-containing" rather than simply "previously treated" EC. This stricter requirement seems to be in line with the marketing authorisation for dostarlimab received from EMA. However, ERG notes that relevant inclusion criteria specified in the GARNET trial clinical study report (CSR) are even more restrictive in that patients were required to:

- Have progressed on or after platinum *doublet therapy*.
- Have received *no more than 2 lines of anticancer therapy for recurrent or advanced (≥Stage IIIB) disease.*

These may have implications related to generalisability of GARNET trial evidence as well as selection of comparators for undertaking indirect comparison to generate estimates for relative effectiveness.

For comparator, the company used a basket of treatments found to be most commonly used in current clinical management according to real world evidence (RWE) obtained from UK registry in its base case. Treatment regimens included in the base case are broadly in line with comparators listed in the final scope, with the following exceptions:

- Pegylated liposomal doxorubicin (PLD) monotherapy was included instead of doxorubicin monotherapy.

- While hormone therapy was included in company's base case, no empirical data for its effectiveness was used as hormone therapy was not adequately captured in the registry and the literature review did not identify any studies that provided relevant evidence. Instead, an assumption that its effectiveness would be as good as the basket of chemotherapies was made.

- Carboplatin plus PLD was included in the basket of treatments in company's base case but was not listed in the final scope.

- Best supportive care was excluded from the company's decision problem.

Overall, the ERG considered the company's approach to using a basket of most commonly used treatments based on UK registry reasonable given the large number of diverse regimens used in clinical practice. However, this approach raises challenges in finding a patient population and retrieving data that are directly comparable with the well-defined GARNET trial population and data. Pertinent issues related to these are highlighted in Section 3 of this report. The ERG agrees that best supportive care is not particularly relevant in the targeted place in the treatment pathway.

While no patient subgroups were specified in the final scope of this STA and in the company submission (CS), the target population includes patients with advanced disease or patients with recurrent disease. As the company acknowledge (CS Section B1.3.2), these two groups of patients may different treatment history, and potentially different response to treatment and prognosis. Potential analyses stratified by these subgroups may reduce heterogeneity within the patient population included in this appraisal. Nevertheless, no such analyses were conducted and presented in the CS.

Table 8: Summary of decision problem addressed in the company submission and ERG's critique

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Population	People with previously treated advanced or recurrent endometrial cancer with high microsatellite instability or mismatch repair deficiency.	Patients with recurrent or advanced dMMR/MSI-H EC that has progressed on or following prior treatment with a platinum-containing regimen.	The patient population is aligned with the NICE final scope, though it is important to note that patients eligible for dostarlimab must *have progressed on or following prior treatment with a platinum-containing regimen*. This is in line with the marketing authorisation for dostarlimab in this indication and the patient population included in the pivotal GARNET trial (see CS Section B.2.3.1).	As the company highlighted, the additional eligibility criterion regarding prior treatment with a platinum-containing regimen conforms to the marketing authorisation granted by the EMA and reflects the inclusion criteria of the GARNET trial. The company suggests that platinum-containing regimen is a standard of care in the UK for first-line treatment for recurrent or advanced EC (CS Section B.1.3.4.2). ERG agrees with this.
Intervention	Dostarlimab	Dostarlimab	NA – aligned with the NICE final scope.	No concern.
Comparator(s)	**Chemotherapy**, including: • Carboplatin and paclitaxel	**Base case cost-effectiveness analysis:** A basket of treatments representing current clinical management, comprising:	**Current clinical management** • In the absence of a definitive standard of care or clear treatment guidelines for this indication, the base case cost-effectiveness analysis	ERG recognises that there is no definitive guideline for the choice of treatment in this setting, and various combination and monotherapy including

	• Paclitaxel monotherapy • Doxorubicin monotherapy • Carboplatin monotherapy **Hormone therapy** (such as medroxyprogesterone acetate and megestrol) **Best supportive care**	• Carboplatin plus paclitaxel • Paclitaxel monotherapy • Carboplatin plus pegylated liposomal doxorubicin (PLD) • PLD monotherapy • Carboplatin monotherapy • Hormone therapy (50:50 ratio of medroxyprogesterone and letrozole) **Scenario analyses:** Individual comparisons versus: • Carboplatin plus paclitaxel • Paclitaxel monotherapy • Doxorubicin monotherapy • Carboplatin monotherapy • Hormone therapy (50:50 ratio of medroxyprogesterone and letrozole)	compares dostarlimab to current clinical management in the UK as a basket of comparator therapies. This consists of aggregate data for patients receiving a range of the most commonly prescribed chemotherapy regimens in patients with recurrent or advanced EC who have progressed on or after a platinum-containing regimen in clinical practice, based on a GSK-initiated real-world evidence (RWE) study using data from the National Cancer Registry Analysis System (NCRAS) in England (hereafter referred to as the UK RWE study). • The treatments included in this aggregate data include the individual chemotherapy regimens listed in the final scope, as well as carboplatin plus PLD. As the UK RWE study could not capture hormone therapy, the costs of hormone therapy (a weighted average of medroxyprogesterone acetate and letrozole based on UK clinical expert feedback) have instead been	platinum-based chemotherapy (e.g. carboplatin and cisplatin), anthracyclines (e.g. doxorubicin) and taxanes (e.g. paclitaxel, docetaxel) have been used in clinical practice depending on characteristics of the tumour, individual patient's treatment history, fitness and other factors. Ideally, comparison of dostarlimab with individual comparators would allow more precise evaluation of relative effectiveness and cost-effectiveness which take into account potential association between patient characteristics and treatment choice. Nevertheless, the large number of possible regimens and the paucity of effectiveness data related to individual regimens mean comparison with individual comparator may only be feasible for most commonly used regimens such as carboplatin + paclitaxel and doxorubicin monotherapy. Given the

			incorporated within the basket. • An SLR was conducted to identify relevant clinical evidence for the individual therapies listed in the NICE final scope however these data were extremely limited; most studies in the relevant patient population were observational studies, where patient characteristics and Kaplan-Meier (KM) survival data were poorly reported. Where possible, scenario analyses have been conducted versus the comparators for which data were identified in the literature in the post-platinum chemotherapy setting. • No data were identified for either carboplatin monotherapy or hormone therapy. Despite efforts made to identify alternative sources of data for these comparators, feedback from UK clinical experts strongly indicated that any data for patients not in the post-platinum chemotherapy setting would not be suitable to use as a proxy for these comparators. The UK clinical	above consideration, ERG agrees that using a basket of treatments found in current practice may be a reasonable comparator. ERG considers that hormone therapy is a relevant comparator in second-line, recurrent or advanced EC setting. ERG agrees that patients who would be considered for best supportive care alone are likely to be those who are not well enough to be considered for dostarlimab treatment, and therefore best supportive care is not a relevant comparator in the targeted place in the treatment pathway.

			experts also indicated that survival with hormone therapy or carboplatin monotherapy would not be expected to exceed that observed in the UK RWE study. As such, individual comparisons have been explored between dostarlimab and carboplatin monotherapy and hormone therapy in scenario analyses, using efficacy data for doxorubicin monotherapy and current clinical management as a proxy, respectively (See Section B.3.8.3). **Removal of BSC** • BSC was not fully defined in the NICE final scope, and there is a lack of standardised definition in the literature. It is likely to consist of pain and symptom management or relief with treatment such as analgesics and corticosteroids. • BSC is not considered a relevant comparator to dostarlimab in this submission and a comparison versus BSC has not been included, for the following reasons:	

			○ Feedback from UK clinical experts is that, for most patients, BSC would be used as an add-on therapy to chemotherapy and thus is expected to be used as an add-on therapy to dostarlimab.16 Accordingly, UK clinical experts agreed that BSC would not represent a relevant comparator to dostarlimab.16 ○ Whilst a small proportion of patients with recurrent or advanced EC who have progressed on or after a platinum-containing regimen may receive palliative therapy as BSC, these patients reflect a different patient population (of more severely unwell patients) compared to the proposed target population for dostarlimab.	
Outcomes	The outcome measures to be considered include: • progression-free survival • overall survival • response rates • duration of response	• Progression-free survival • Overall survival • Response rates (overall response rates, disease control rate) • Duration of response	NA – aligned with the NICE final scope.	No concern.

	• adverse effects of treatment • health-related quality of life	• Adverse effects of treatment • Health-related quality of life		
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any commercial arrangements for the intervention, comparator and subsequent treatment	• An economic analysis has been conducted with the cost-effectiveness of treatments expressed in terms of incremental cost per quality-adjusted life year. • A lifetime time horizon has been adopted to reflect all differences in costs and outcomes between the technologies being compared. • Costs are considered from an NHS and Personal Social Services perspective. • A confidential commercial discount to the list price of dostarlimab has been adopted within the base case analysis. • Any commercial arrangements for the comparators are not	• Regarding the costs associated with diagnostic testing, NICE diagnostics guidance DG42 recommends that all patients with EC should be tested using immunohistochemistry (IHC) to identify tumours with dMMR/MSI-H.18 DG42 recommends that IHC testing for dMMR is the preferred approach, and clinical expert opinion sought by GSK agreed with this.16 Additionally, discussions with NHSE at a surgery confirmed that testing would not be an issue for access to dostarlimab. • Furthermore, given the availability of nivolumab through the Cancer Drugs Fund (CDF) for patients with dMMR/MSI-H, dMMR testing is already in use in clinical practice to identify eligible patients, and therefore resources for dMMR testing are already being embedded within usual practice. • As such, dMMR testing will soon become standard of care for all patients with EC and no additional diagnostic tests will be	The ERG agrees that test costs should not be included.

	technologies will be taken into account. The economic modelling should include the costs associated with diagnostic testing for microsatellite instability status in people with endometrial cancer who would not otherwise have been tested. A sensitivity analysis should be provided without the cost of the diagnostic test. See section 5.9 of the Guide to the Methods of Technology Appraisals.	known and have therefore not been taken into account. • The inclusion of diagnostic testing for dMMR/MSI-H status has been explored within a scenario analysis, which considers dMMR/MSI-H testing for recurrent patients only (see Section B.3.8.3).	required to facilitate the prescribing of dostarlimab beyond those already conducted for patients with EC in UK NHS clinical practice. These costs have therefore not been included within the base case economic analysis, but a scenario analysis has been conducted to explore the impact of the inclusion of diagnostic testing costs for dMMR status for recurrent patients only.	
Subgroups	None specified	Not discussed in the CS.	No comments provided.	The target population includes two subgroups of patients (i.e. patients with advanced or recurrent disease) with different treatment history and prognosis, and could potentially be evaluated separately.
Special considerations including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific	Not discussed in the CS.	No comments provided.	No concern.

	treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.			

3 CLINICAL EFFECTIVENESS

3.1 *Critique of the methods of review(s)*

The company conducted an original and updated clinical systematic review to identify evidence for the efficacy and safety of dostarlimab and the chemotherapy comparators listed in the NICE final scope for the treatment of recurrent or advanced EC that has progressed on or after platinum-based chemotherapy. A range of study types (both interventional and observational) are included (CS Appendix D.4.1).

A summary of the ERG's quality assessment of the company's systematic review of clinical evidence is presented in ERG report Appendix (Table 57). While the overall risk of bias was judged to be low, the ERG has some concern. The company did not use the NICE-preferred tool for assessing methodological quality, heterogeneity in study results was not addressed in their analysis, and no information on predefined analyses in a referenced protocol or in the submission was provided. Results from the clinical systematic review were analysed with narrative description. Given the nature and differences in the study designs and outcomes across included studies, a quantitative synthesis may not be appropriate. The ERG considers the narrative analyses method appropriate for the SLR.

The company did not initially consider hormone therapy (which was within the NICE final scope) as one of the comparators, and thus was not included in the original or update clinical SLR. However, the company provided a targeted literature review (TLR) for hormone therapy (CS Appendix L). A summary of the ERG's quality assessment of the company's the hormone therapy TLR is presented in the ERG report Appendix. No studies from the hormone therapy TLR were found relevant by the company for this submission; thus, none was included in the cost-effectiveness analysis. The company made an assumption that hormone therapy has the same effectiveness as other therapies in the basket of

treatments was made; thus, conducted a scenario analysis with hormone therapy, using the UK RWE study as a proxy to validate the base-case.

The ERG examined the studies included and excluded in the company's clinical systematic review as well as the hormone therapy TLR. In addition, the ERG conducted searches for recent relevant systematic reviews and examined their bibliographies for studies of comparator treatments listed in the NICE final scope. No additional relevant studies were identified by the ERG.

Quality Assessment

The company states that they assessed study quality using the Appendix C of PMG6 methodology checklist for randomised controlled trials in the old NICE guidelines manual,[11] Critical Appraisal Skills Programme (CASP) check list for Non-RCTs,[12] and ROBINS I assessment tool for the UK RWE study (non-RCT study)[13] (CS section B.2.3.1.4 and Appendix D.7). The latest NICE guidance[14] recommends the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) checklist for RCTs, and the Institute of Health Economics (IHE) Quality Appraisal Checklist for case series (non-RCTs). Therefore, the ERG conducted an independent assessment of the eight studies included in the indirect treatment comparisons (ITCs) (GARNET, UK RWE study, ZoptEC trial, McMeekin *et al*. (2015), Rubinstein *et al*. (2019), Mazgani *et al*. (2008), Julius *et al*. (2013), and Makker *et al*. (2013))[15-21]) using both tools. A comparison of the ERG and company quality assessments using the company's preferred tools are provided in ERG report appendix (Table 58, Table 59, Table 60 respectively). A single ERG reviewer conducted these assessments, with a second reviewer checking all items where the ERG and company disagreed.

ERG points of critique: The ERG has few concerns over the overall low risk of bias of company's clinical SLR. In addition to the observed differences between the ERG and the company's judgements, the choice of checklist for the quality appraisal appears to be important given the differences in ERG overall risk of judgments using the company preferred checklist compared to the NICE preferred checklist, particularly for GARNET, where the ERG reported a low risk of bias rating using NICE preferred checklist and moderate risk rating using the company preferred check list. GARNET is noted to be of higher quality compared to the UK RWE study (the key comparator study), using the NICE preferred checklist.

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

The key study in the CS is the GARNET (NCT02715284), a Phase 1, single-arm, open-label, multicentre, non-randomised study of dostarlimab (see ERG report Appendix for details on the study quality assessment).

GARNET (data cut 1 March 2020) has not previously been published and data are presented in the CS and the CSR provided to the ERG.

3.2.1 GARNET trial

3.2.1.1 GARNET method

GARNET is an ongoing multi-cohort study conducted in 9 countries (including 9 centres in UK) to evaluate the antitumor activity of dostarlimab in participants with recurrent and advanced endometrial cancer with only the relevant Cohort A1 included in the submission. This cohort included patients with recurrent or advanced dMMR/MSI-H EC that has progressed after treatment with a platinum-containing chemotherapy regimen and have histologically or cytologically

proven recurrent or advanced EC with measurable lesion(s) per RECIST v1.1.[22] Patients had to have received no more than 2 lines of anticancer therapy for recurrent or advanced (≥Stage IIIB) disease. All EC histologies were allowed, except endometrial sarcoma. Participants were also required to have an Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0 or 1 and adequate organ function. Key exclusion criteria were prior therapy with an anti-PD-1, anti-PD-L1, or anti-programmed cell death-ligand 2 agent, uncontrolled central nervous system metastases and/or carcinomatous meningitis, and additional malignancy that progressed or required active treatment within the last 2 years. CS Appendix N, Table 96 has detailed patient eligibility criteria. The key patient flow of the study is provided in CS Appendix D.6 Table 64.

Eligible patients received dostarlimab 500 mg via IV infusion every 3 weeks (Day 1 of each 21-day cycle) for the first 4 cycles, followed by dostarlimab 1000 mg via IV infusion every 6 weeks (Day 1 of each 42-day cycle) for all subsequent cycles. The median follow-up in the submission was [redacted] months (see CS section B.2.4 for follow-up for specific outcomes). The ERG considered this a relatively short follow-up duration. 129 patients received any amount of dostarlimab (intention-to-treat (ITT) population/safety analysis set). This population was used in the base case cost-effectiveness evaluation. The company described a number of pre-specified analysis populations, including: the efficacy population analysis set (n = [redacted]) and immune-related efficacy population set (n = [redacted]) (CS Table 8).

Baseline characteristics of the participants in GARNET were reported by the company for the ITT population/safety and efficacy population analysis sets, discussed in more detail in CS section B.2.3.1.2 (CS Table 7). The ERG verified these data using the tables and figures provided by the company in the submission as these were not reported in the company CSR. [redacted] patients received more than 2 prior lines of anticancer therapy for recurrent or advanced disease, which appear to contradict with the specified inclusion criteria of no more than 2 prior lines of anticancer therapy (CS Appendix

N). The company explained in factual accuracy check that the 2 prior lines of anticancer therapy for trial inclusion refer specially to platinum-based therapy. However the ERG could not verify this based on the published trial protocol. The ERG considered the inclusion of these [redacted] patients important as it is unclear if any adjustments were made in the indirect comparisons (CS Appendix D.5).

The clinical advisors consulted for the ERG considered the GARNET participants to be generally representative of UK patients (CS section B.2.3.1.2, CS Table 7).

3.2.1.2 Efficacy outcomes

The company describes the primary and secondary efficacy outcomes in CS Table 6. Key safety measures and health-related quality of life (HRQoL) outcome measures are also described in CS Table 6.
For the key efficacy outcomes, the efficacy evaluable set (n=[redacted]) was used, excluding progression free survival (PFS) and overall survival (OS), where the ITT population/safety set was used in the economic evaluation (n =129). For immune-related efficacy outcomes, a different population was used (n=[redacted]). HRQoL and safety outcomes were also derived from ITT population/safety set. As GARNET is a single arm study, the statistical assessment of outcomes was descriptive. Kaplan-Meier methods were used to estimate PFS and OS.

Table 9: Key efficacy outcomes from GARNET, summarises the key clinical effectiveness outcomes of GARNET. Fuller details are presented in the CS section B.2.4.1, B.2.4.2, B.2.4.3, B.2.4.5 – B.2.4.8.

Table 9: Key efficacy outcomes from GARNET

Efficacy outcomes	Efficacy evaluable set, (n=[redacted]); ITT population (n=129)*
ORR (95% CI)[a]	[redacted]
Complete response	[redacted]
Partial response	[redacted]
DOR, median (95% CI)[b] months Median follow-up [redacted] months	[redacted]
DCR (95% CI)[a]	[redacted]
irORR[c,d]	[redacted]
irDOR[c,e] median (95% CI) months	[redacted]
irDCR[c,d]	[redacted]
*PFS[b], median (95% CI) months Median follow-up [redacted] months	[redacted]
irPFS[c,d], median (95% CI) months	[redacted]
*OS[b], median (95%) CI months Median follow-up [redacted] months	[redacted]

Footnotes: [a]Two-sided 95% exact Clopper–Pearson confidence interval (CI); [b]Two-sided 95% CI from Kaplan–Meier; [c] Immune-related efficacy population; [d]Exact 2-sided 95% CI for the binomial proportion; [e] 95% CIs from Brookmeyer and Crowley (1982) method; *ITT population;**PFS estimate from non-rounded up individual patient PFS estimates; +indicates response is still ongoing
Abbreviations: ORR: Overall response rate; DOR: Duration of Response; DCR: Disease control rate; irORR: immune-related ORR; irDOR: immune-related DOR; irDCR: immune-related DCR; irPFS: immune-related PFS; PFS: Progression Free survival; OS: Overall Survival.

The ERG verified the above data using the tables and figures provided by the company in the submission and the CSR. The ERG could only verify PFS, irPFS and OS information for the ITT population using the tables and figures provided by

the company in the submission as these were not reported in the company CSR. The company reported immune end points to provide more specific information for the tumour response to dostarlimab as an immunotherapy. The ERG agrees with this rationale. The ORR majorly consisted of partial response. The median DOR was not reached. The median PFS was associated with very wide confidence intervals (CIs) and was very sensitive to very small changes in individual patient PFS estimates, leading to different PFS estimates for rounded up and unrounded up individual patient PFS estimates. Median PFS estimate of [redacted] months (from non-rounded up individual patient PFS estimates) informed the economic evaluation. Most of the progression occurred in the first 6 months. The median OS was not evaluable. Figure 1 and Figure 2 below show progression free survival (PFS) and overall survival (OS) from GARNET. The blue lines are the survival outcomes for the ITT/safety population. PFS and OS information from the ITT population/safety set (n =129) was used in the economic evaluation. The ERG notes that the flat tail in GARNET OS curve is predictive of long term effectiveness; however, it may be due to insufficient follow-up duration/immature data, and small sample size.

Figure 1: PFS from GARNET (efficacy population and ITT population) (BICR)

Figure 2: OS from GARNET (efficacy population and ITT population)

The CS presents pre-specified subgroup analyses for ORR in the efficacy population in CS Figure 20. There appears to be overlapping 95% CIs within the subgroups as well as with the overall population ORR. However, this was not observed for the subgroup analysis by ECOG performance status. The ERG agrees with the company that ORR was ≥20% (null hypothesis; expected ORR for conventional therapy) for all of the subgroup estimates, suggesting a treatment benefit of dostarlimab for all subgroups. However, the ERG notes that numbers for many of these subgroups are small with wide confidence intervals suggesting uncertainty. The ERG could only verify these data using the figure provided by the company as there was no information in the CSR.

ERG points of critique: Patients with more than 2 lines of prior anti-cancer treatment were included in the GARNET study, which was not consistent with the pre-specified eligibility criteria. Clinical effectiveness outcomes were reported

over a relatively short time frame and have the potential for positive response to treatment with dostarlimab in most participants. Some outcomes do not have enough data to be fully reported (such as DOR and OS). The median PFS is unstable and varies with the decimal place of individual PFS estimates. With no comparator group it is unclear what magnitude of benefit dostarlimab offers over established clinical management. This is discussed in more detail in Section 3.4 of the ERG report (Critique of the indirect comparison).

3.2.1.3 HRQoL

The EORTC-QLQ-C30 and the EQ-5D-5L were assessed following a protocol amendment, and therefore not all participants were assessed for the effects of dostarlimab on HRQoL. The HRQoL data was from participants in the ITT/safety population set. No HRQoL outcomes are reported in the CSR provided by the company. The ERG could only verify these data using the tables and figures provided in CS section B.2.4.8. For EORTC-QLQ-C30, [redacted] participants had evaluable data and the mean scores generally showed improvement in HRQoL from baseline to week 24, except for deterioration in some domains in the initial month (CS section B.2.4.8). The ERG notes that not all domains and items of the EORTC-QLQ-C30 were reported. Over the period of follow-up, the minimally important difference appears to be achieved by patient-reported pain, fatigue symptoms, physical functioning, and symptomatic adverse events (AEs). For the EQ-5D-5L index score, [redacted] participants had evaluable data. The change from baseline was submitted by the company in response to clarification question A12b, where the initial 18 weeks showed improvements, followed by fluctuation to week 54, and thereafter an improvement to week 78 with a decline to week 96. These EQ-5D-5L scores were used in the economic evaluation, see CS section B.3.4.1 for further description. For the EQ-VAS, [redacted] participants had evaluable data. The change from baseline is seen in CS Figure 16, where mean scores showed fluctuation throughout the study. The most notable improvement in the scores were seen after end of treatment. The ERG notes that a small number of participants were evaluated from week 18 onwards.

ERG points of critique: The effects of dostarlimab on HRQoL is unclear. Not all participants were assessed for HRQoL, the CS does not report the mean change from baseline for all domains of the EORTC-QLQ-C30 and no discussion of the minimally important differences of these outcomes were reported.

3.2.1.4 Safety

The safety data reported were from the dostarlimab ITT/safety population set, n=129 as a secondary outcome in the CSR and CS section B.2.8. Most participants receiving dostarlimab had at least one treatment-emergent adverse event (TEAE) ([redacted]) and [redacted] experienced at least one Grade 3 or higher TEAE. Serious adverse events occurred in [redacted] of patients. The most frequently reported Grade ≥3 TEAEs were anaemia and abdominal pain, see CS Table 34. Grade ≥3 TEAEs with an incidence of ≥5% were included in the economic model, see CS section B.3.3.8. Death occurred in participants ([redacted]) while in the study, with disease progression as the most common reason ([redacted]129, [redacted]). Adverse event was the cause of death in [redacted] patients.

ERG points of critique: The ERG clinical advisors considered the toxicity of dostarlimab to be at an acceptable rate for an immunotherapy (I-O therapy).

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

Evidence for the comparator is taken from a real-word evidence (UK RWE) study which was funded by GSK (the company). Data for this study are provided in the CS and a report provided to the ERG in response to clarification question C2. In addition, six studies (Rubinstein *et al*. (2019); Mazgani *et al*. (2008); McMeekin *et al*. (2015); Julius *et al*.

(2013); Makker *et al*. (2013); and ZoptEC study)[15-21] were included in the indirect comparison and used by the company in scenarios for the economic model (see CS section B.3.8.3).

3.3.1 UK RWE study

The UK RWE study was used as the main comparative evidence (for current clinical management) in the indirect treatment comparison and economic model. It has not previously been published and data are presented in the CS and a report provided to the ERG in response to clarification question C2.

The UK RWE study was a UK national retrospective observational study (see ERG report Appendix for details on the study quality assessment), conducted by GSK to fill evidence gaps relating to the current clinical management for patients diagnosed with recurrent or advanced EC in the UK due to the lack of data identified for comparator therapies in the company's clinical SLR (CS section B.2.3.2 and B.2.4.5). UK RWE study used routine, linked patient-level UK health data available through the National Cancer Registration and Analysis Service (NCRAS), which combines linked data from several health and population databases. UK RWE study collected data for patients diagnosed between 1st January 2013 and 31st December 2018, with data extraction up until 30th September 2020.

To identify patients with EC, the UK RWE study used an initial inclusion criteria and exclusion criteria (CS Table 10). In addition, more inclusion and exclusion criteria were applied to identify patients with recurrent or advanced EC (CS Table 11). Further inclusion and exclusion criteria were applied to align the patient population more closely to GARNET (CS Table 12 and Appendix O.1). The key inclusion criteria for the UK RWE study to align the patient population more closely to GARNET included: a diagnosis of recurrent or advanced EC, and patients who received exactly one prior platinum doublet therapy for recurrent or advanced disease. The study identified a large population of patients (n=[redacted]), further known as UK RWE GARNET-like cohort or abbreviated as RWEQ (Real World EQuivalent) cohort for brevity in the ERG

report. The ERG notes that there are uncertainties around the impact of the possible selection bias associated with the complex process of deriving the UK RWEQ cohort from the patients with EC diagnosis in the registry. Figure 3 shows a flow of patients included in the RWEQ.

Figure 3: Patients included in the UK Real World EQuivalent (RWEQ) cohort (reproduced from CS Figure 8)

Abbreviations: 2L: second-line; EC: endometrial cancer; RWE: real-world evidence.

3.3.1.1 UK RWE study methods

The inclusion criteria of the GARNET trial and UK RWE study have been considered by the ERG. The key eligibility criteria for GARNET were presented in CS Table 6 (and Appendix N.1, Table 96) and the UK RWE in CS Table 10 - 12 and Appendix O.1.

The ERG notes that these criteria appear to be similar on many key factors but that there are differences; those with potential relevance are:

- In GARNET, participants were required to have received no more than two lines of systemic anticancer therapy. In the UK RWE study, the requirement was for exactly one prior platinum doublet therapy. Based on the company's response to ERG clarification question A15, the ERG considers the definitions of lines of prior therapy dissimilar

between GARNET and the UK RWE study. Adjuvant and neoadjuvant chemotherapy were excluded from the data shown in CS Table 7 for GARNET, but this could not be verified in the UK RWE study. Also, it was not required to have platinum-based doublet therapy as the last line of therapy prior to dostarlimab in GARNET. In the UK RWE study, it was required to have received only one line of platinum-based doublet chemotherapy, progressed, and received further second line treatment. Given the differences in the demographic and clinical baseline characteristics (see Table 10 below) between the two studies, it is unclear how the differences in the prior lines of therapy may impact the benefit of dostarlimab over established clinical management.

- In GARNET, participants were required to have histologically or cytologically proven recurrent solid tumour with measurable lesion(s) per RECIST v1.1. In the UK RWE study, the requirement for recurrence was probable recurrence, defined as patients who were FIGO Stage I/II and received surgery, systemic anti-cancer therapy or radiation therapy and then had a treatment gap greater than 90 days, followed by treatment with any treatment. The company notes that this definition was supported by their clinical advisors/UK clinical expert opinion; however, the ERG's clinical advisor noted that some sort of radiographic evidence is required to confirm recurrence.

  Based on company's response to ERG clarification question A16, the ERG has been able to consider the validity the definition of recurrence in the UK RWE study with the number of patients identified in the UK RWE study compared to the estimated incidence of patients with recurrent EC based on published epidemiological estimates for the UK which was submitted. The ERG notes that these estimates appear to be similar, but some uncertainty remains in the robustness of the definition of recurrence in the UK RWE study as a difference of about 2% in recurrence rate was observed when recurrence was defined as >180 days in the post-hoc sensitivity analyses conducted by the company in response to ERG clarification question A16. It is unclear how the difference in definition of recurrence between GARNET and UK RWE study might impact baseline prognosis.

- In GARNET, patients were required to have dMMR/MSI-H EC, this was not stated in the eligibility criteria for the UK RWE study. The CS states that "*MSI-H or dMMR EC represents a subgroup where PD-1/PD-L1 inhibition with I-O therapy (such as dostarlimab) is most effective*". Also, the company referred to a systematic literature review (SLR) conducted by GSK,[23] stating "there is no evidence MSI-H or dMMR biomarker status has any prognostic or predictive value for efficacy and survival outcomes (including recurrence, relapse-free survival, PFS and OS) among patients with advanced or recurrent EC receiving non-anti-PD-(L)1 therapy". The ERG notes that the full report for the SLR was not provided by the company. While ERG is not aware of evidence which contradicts this claim, ERG's clinical advisor pointed out that the inclusion of exclusively patients with dMMR/MSI-H in the GARNET may have resulted in the selection of a higher proportion of patients with better prognosis compared with RWEQ cohort, which was not selected based on MMR/MSI status. This is because dMMR/MSI-H is predominantly found within in Type I endometrioid tumours (28-40%), which tends to have better prognosis (as described earlier in Section 2.2.3) and is rarely found within other histological subtypes (serous, clear cell and other types, 0-2%) which tends to be more aggressive.[8] The was reflected in the much higher proportion of patients with endometrioid EC in the GARNET compared with RWEQ (see Table 10 below).
- In GARNET, participants were required to have adequate organ function; this was not stated in the eligibility criteria for the UK RWE study. This could also have led to the selection of fitter patients with better prognosis into the GARNET trial.

Demographic and clinical characteristics of the participants in the RWEQ cohort were reported by the company (CS section B.2.4). The ERG verified these data using the tables and figures provided by the company in the submission as there was no published study report for the RWEQ cohort. In the RWEQ, patients were required to have an ECOG PS of ≤1. However, patients with an ECOG PS of 'not recorded (NR)' (n=[redacted]) were not excluded by the company from the UK

RWEQ cohort for the purpose of a larger sample size of patients, longer follow-up, and prevention of potential unknown bias associated with non-recording. The company highlighted that information on the classification of patients as ECOG PS of 'not recorded' is not provided in the NCRAS dataset and the chances that patients with an ECOG PS of 'not recorded' had an ECOG PS >1 was negligible as patients with an ECOG PS >1 comprised a small percentage (N=[redacted]) of the overall UK RWE study patients with recurrent or advanced EC. The ERG has not been able to verify this estimate. The company provided a sensitivity analysis of patients with a known ECOG PS of 0 or 1 in the RWEQ cohort subsequently referred to as 'RWEQ ECOG PS ≤1' cohort (CS Appendix O.2 and reproduced in ERG report Table 6). The ERG agrees that the overall patient characteristics and efficacy outcomes of the RWEQ ECOG PS ≤1 cohort appear to be similar to the RWEQ cohort and excluding patients with an ECOG PS of 'not recorded' does not seem to have a major impact.

The ERG notes that the most common chemotherapy regimens received by patients also appear to be similar between REWQ ECOG PS ≤1 cohort and the RWEQ cohort (CS Table 14 and Appendix Table 128). The ERG observed that despite that carboplatin plus pegylated liposomal doxorubicin (PLD) was not listed in the NICE final scope as a relevant comparator, it was included by the company. The company noted data completeness as carboplatin plus PLD was received by a substantial proportion of the RWE population as the rationale for inclusion.

The ERG found several differences in the demographic and clinical baseline characteristics between the RWEQ cohort and GARNET ITT population (see Table 10 below) for the following characteristics: age (younger population in GARNET); FIGO stage (RWEQ population had more advanced disease); Grade of disease (highest portion was grade 3 in the RWEQ population, and grade 2 in the GARNET population); ECOG PS (GARNET had higher proportion in ECOG status 0 and 1, and half of the RWEQ population had their ECOG status unknown); histology (GARNET had a higher proportion

of endometroid disease); prior lines of therapy (RWEQ population had exactly one prior platinum doublet therapy while GARNET may have had 1 or more than prior lines of therapy, where one prior therapy must be specific to platinum doublet therapy); and prior surgery (GARNET had higher proportions). It is unclear how exactly these imbalances might affect baseline prognosis at the start of the second-line treatment and therefore subsequent outcomes in advanced/recurrence setting for the two groups, although many of the above differences may suggest more advanced and aggressive disease among the RWEQ cohort.

The ERG found differences in the company's presentation of patients' ECOG PS and FIGO stage. The company provided information on the ECOG PS and FIGO stage at study entry for GARNET whereas the ECOG PS and FIGO stage recorded at "registry diagnosis" was provided for the RWEQ study participants (see Table 10 below). The company explained in their response to ERG clarification question A20 that registry diagnosis is "the date a patient is entered in the NCRAS registry, and not necessarily the date of cancer diagnosis". As both ECOG PS and FIGO stage are well recognised prognostic factors and they may have changed (likely deteriorated) between registry entry and start of second-line therapy, ERG considered the discrepant timing of measuring these variables between GARNET and RWEQ to be a crucial issue that could invalidate any adjustments made in the indirect comparisons using these data (this issue is discussed further in Section 3.4.1).

Differences in the PFS time definition between RWEQ participants and the GARNET participants were also observed by the ERG. Time to next treatment (TTNT) was used as a proxy for PFS for the RWEQ due to lack of progression information within the NCRAS database and following advice from the company's clinical experts. The CS anticipates using TTNT as proxy for PFS may favour current treatment management. The ERG notes that there is uncertainty around the robustness of this proxy measure.

Table 10: Baseline characteristics and efficacy outcomes in the GARNET-like ECOG PS ≤1, UK RWE GARNET-like (RWEQ) cohort, and GARNET ITT population

Characteristic	GARNET-like UK RWE (RWEQ) ECOG PS ≤1 cohort (N=[redacted])	GARNET-like UK RWE (RWEQ) cohort (N=[redacted])	GARNET ITT population (N=129)
Mean age, years (STD)	[redacted]	[redacted]	[redacted]
Median age, years (range)	[redacted]	[redacted]	[redacted]
Age group, n (%)			
<65 years	[redacted]	[redacted]	[redacted]
65 to <75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]
Race, n (%)			
White	[redacted]	[redacted]	[redacted]
Black	[redacted]	[redacted]	[redacted]
Asian	[redacted]	[redacted]	[redacted]
Other	[redacted]	[redacted]	[redacted]
Unknown	[redacted]	[redacted]	[redacted]
Most recent ECOG PS at registry diagnosis (RWEQ) or study entry (GARNET), n (%)			
0	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]
Not recorded	[redacted]	[redacted]	[redacted]
Histology at diagnosis, n (%)			
Endometrioid	[redacted]	[redacted]	[redacted]
Non-endometroid	[redacted]	[redacted]	[redacted]
Serous carcinoma	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]
FIGO stage at the time of registry diagnosis (RWEQ) or Most recent FIGO stage at study entry, (GARNET), n (%)			

I	[redacted]	[redacted]	[redacted]
II	[redacted]	[redacted]	[redacted]
III	[redacted]	[redacted]	[redacted]
IV	[redacted]	[redacted]	[redacted]
Unknown	[redacted]	[redacted]	[redacted]
Grade of disease at diagnosis, n (%)			
Grade 1	[redacted]	[redacted]	[redacted]
Grade 2	[redacted]	[redacted]	[redacted]
Grade 3	[redacted]	[redacted]	[redacted]
Grade 4	[redacted]	[redacted]	[redacted]
Not assessable	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]
Prior anticancer treatment, n (%)			
Any prior anti-cancer treatment	[redacted]	[redacted]	[redacted]
Prior surgery	[redacted]	[redacted]	[redacted]
Number of prior lines of therapy post advanced/recurrent diagnosis, n (%)			
1	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]
≥4	[redacted]	[redacted]	[redacted]
Most common chemotherapy regimens			
	Carboplatin plus paclitaxel	Carboplatin plus paclitaxel	NA
	Carboplatin plus PLD	Paclitaxel monotherapy	NA
	Paclitaxel monotherapy	Carboplatin plus PLD	NA
	PLD monotherapy	PLD monotherapy	NA
	Carboplatin monotherapy	Carboplatin monotherapy	NA

	Cisplatin plus doxorubicin	Cisplatin plus doxorubicin	NA
	Carboplatin plus gemcitabine	Carboplatin plus gemcitabine	NA
	Doxorubicin monotherapy	Carboplatin plus doxorubicin	NA
	Cisplatin monotherapy	Doxorubicin	NA
	Carboplatin plus doxorubicin	Cisplatin	NA
Median PFS (months) (95% CI)	[redacted]	[redacted]	[redacted]
Median OS (months) (95% CI)	[redacted]	[redacted]	[redacted]

Abbreviations: ECOG PS: Eastern Cooperative Oncology Group Performance Status; FIGO: International Federation of Gynaecology and Obstetrics; ITT: intention-to-treat; RWE: real-world evidence; STD: standard deviation; NA: Not applicable

The ERG notes that hormone therapy is not included as part of the current clinical management of recurrent or advanced EC in the UK RWE study, and aware that it was incompletely captured in the NCRAS database. This is further verified through the company's response to clarification question A14, where the company re-iterated that patients receiving hormone therapy were not purposely excluded from the RWEQ cohort, rather hormone therapy was poorly reported in the NCRAS database, as it is dispensed in primary care or community pharmacies.

3.3.1.2 UK RWE study results

The primary efficacy outcome measures of the UK RWE study are PFS and OS. Safety measures were not recorded in the UK RWE study. As the UK RWE study is a single arm, retrospective observational study, the statistical assessment of outcomes was descriptive. Kaplan-Meier methods were used to estimate PFS and OS. Summaries of the PFS and OS outcomes from the final RWEQ cohort are presented in Table 10 above. The ERG verified these data using the tables and figures provided by the company in the submission. PFS and OS information from the RWEQ cohort was used in the cost effectiveness analysis. A naïve comparison of RWEQ cohort versus GARNET trial patients (CS section B.2.4.5.2 and B.2.4.6.2) showed RWEQ cohort had an increased risk of death, and a reduced risk of progression before month 9. The ERG notes that the results of the native comparison should be treated with caution due to the methodological differences in PFS definitions as well as the sensitivity of GARNET's PFS estimates and immaturity of the GARNET trial data.

ERG points of critique: Overall, the ERG notes there is considerable uncertainty as to the similarity between the RWEQ cohort and GARNET ITT population and its representation of the UK population. In addition, there are concerns about the appropriateness of the definition of recurrence and using TTNT as a proxy for PFS. In order to characterise the differences between GARNET and RWEQ cohort and to identify potentially more comparable patients between the cohorts, data stratified by advanced versus recurrent diseases, and by endometrioid versus other diseases for both cohorts may be valuable. The ERG requested these data as part of the clarification questions; however, no data was provided.

3.3.2 Published studies identified from the company's clinical SLR and included in the indirect comparisons

The UK RWE study was the main comparative efficacy evidence submitted by the company. However, indirect treatment comparisons (ITCs) between dostarlimab and comparators listed in the NICE final scope (including: carboplatin plus

paclitaxel, paclitaxel monotherapy and doxorubicin monotherapy) were carried by the company, based on the studies identified in the clinical SLR (including: Rubinstein *et al*. (2019); Mazgani *et al*. (2008); McMeekin *et al*. (2015); Julius *et al*. (2013); Makker *et al*. (2013); and ZoptEC study).[15-21] These comparisons include an inverse probability treatment weighting (IPTW) ITC between GARNET and the ZoptEC trial[15, 16] and a series of matching-adjusted indirect comparisons (MAICs) between GARNET and the remaining 5 studies included from the SLR (see CS B.2.7.2 and Appendix D.5.2 and D.5.3).

3.3.2.1 Methods of published studies included in the indirect comparisons

The study characteristics, clinical and demographic characteristics, and efficacy outcomes measures (see CS Appendix D.4.3 to D.4.6, D.5.2, and D.5.3 and summarised in ERG report Table 11: Baseline characteristics and efficacy outcomes in the studies included in the ITCs, and GARNET ITT population).

The ERG notes that the study characteristics, clinical and demographic characteristics, and efficacy outcomes measures appear to have some differences; those with potential relevance are:

- Study design: 2 RCTs and 4 non-RCTs.
- Sample size: Ranged from 17 to 255. McMeekin *et al*. (2015)[17] and ZoptEC[15, 16] provide a far greater sample size compared to the other studies.
- Clinical and demographic characteristics: Variance was observed in age; ethnicity, ECOG PS; FIGO stage; histology; and lines of therapy.
- Efficacy outcomes: Response rate was the main efficacy outcome for most of the studies except ZoptEC[15, 16], McMeekin *et al*. (2015)[17] and Julius *et al*. (2013),[20] where PFS or OS was their primary outcome.

- Definition of response rates, PFS and OS: The definitions were either not reported or varied between studies. Of relevance is the difference in PFS definition between ZoptEC[15, 16] and GARNET. Due to the differences in PFS between the two studies, a descriptive-only KM analysis was conducted to compare PFS between GARNET and ZoptEC;[15, 16] but an adjusted comparison of OS between ZoptEC[15, 16] and GARNET was conducted by the company.
- Tumour assessments: Studies used tumour assessments per RECIST v1.1 by blinded independent central review (BICR) or investigator[22] and RECIST v1 (for trials performed prior to 2009 when the RECIST v1.1 was published).

Owing to lack of data on patient characteristics and prognostic variables, and limitations in the study design from the published studies (see Table 11), the indirect treatment comparisons (ITCs) cannot account for any prognostic variable imbalances that are not reported, introducing an unknown level of bias. The ITCs were provided for completeness as supportive comparative efficacy evidence only and are used by the company in scenarios for the economic model. The ERG partly agrees with this.

The ERG considered the doxorubicin arm of ZoptEC[15, 16] a potential primary comparative effectiveness evidence alongside UK RWE study. The baseline characteristics of patients (excluding: ethnicity, ECOG PS, and FIGO stage) (see Table 11 and Clarification question A17, Table 14), setting and data collection methods were similar between GARNET and ZoptEC.[15, 16] There were differences in the presentation of information on stage of endometrial cancer (ZoptEC included an additional stage – “metastatic disease”), definition of PFS and timings of re-evaluation for response between GARNET and ZoptEC.[15, 16] Some of the differences in definition and baseline characteristics were accounted for by the choice of ITC method – inverse-probability weighted (IPTW) and excluding patients before the indirect comparison was conducted (see CS Appendix D.5.2, Table 44). In addition, relative to the studies included in the ITCs, individual patient-

level data (IPD) on ZoptEC large patient sample were available, thus, allowing the GARNET population to be matched with ZoptEC[15, 16] populations as closely as possible, minimising the heterogeneity between the two study populations, and resulting in more robust comparisons. The number of lines of prior anti-cancer treatment and tumour grade (key prognostic variables) were missing in the ZoptEC trial,[15, 16] which may impact the robustness of the study.

At the check point meeting, the company highlighted that because doxorubicin monotherapy is captured in the UK RWE study it was not necessary to include the ZoptEC trial[15, 16] as a primary comparator. The ERG notes that a comparative analysis to verify the similarities between the efficacy outcomes of the pegylated liposomal doxorubicin (PLD) or doxorubicin monotherapy in UK RWEGARNET-like cohort versus ZoptEC[15, 16] was not provided by the company.

Information on individual treatment regimens (including from PLD monotherapy) in the UK RWE GARNET-like cohort was provided by the company in response to clarification questions A3 and A9. Data on doxorubicin was not provided, as the company only presented information on treatments prescribed to ≥5% of patients in the UK RWE study GARNET-like population. The ERG found several differences in the demographic and clinical baseline characteristics between the UK RWE GARNET-like (PLD monotherapy) cohort and the doxorubicin arm of ZoptEC[15, 16] for the following characteristics: age (younger population in ZoptEC); ethnicity (ZoptEC had higher proportion of white ethnicity); ECOG PS (ZoptEC had higher proportion of patients in ECOG status 0 and 1, and about half of the GARNET like UK RWE (PLD monotherapy) population had their ECOG status unknown); FIGO stage (GARNET like UK RWE (PLD monotherapy) population had more advanced disease); and histology (Zoptec had grater endometroid disease) (see Table 11 below). It is unclear how these differences might affect baseline prognosis at the start of the second-line treatment for both groups, although many of the above differences may suggest less aggressive disease among the ZoptEC[15, 16] population. Further work on the comparative analysis has been conducted by the ERG (see ERG report section 3.5).

Besides ZoptEC,[15, 16] the ERG considered McMeekin *et al*. (2015)[17] (an RCT which provides evidence for doxorubicin or paclitaxel monotherapy) a reasonably robust study as it also had more information on inclusion/exclusion criteria, patient characteristics and prognostic data with large sample size relative to other studies included in the ITCs. However, the ERG notes that there were differences in the baseline characteristics of patients (including: ethnicity and histology) between the McMeekin *et al*. (2015) study[17] and GARNET trial. Also, the classification of patient's performance status differed between GARNET and McMeekin *et al*. (2015),[17] with the use of widely accepted ECOG status[24] and Karnofsky Performance Status (KPS),[25] respectively. The company matched KPS scale in McMeekin *et al*. (2015)[17] to ECOG status scale to align the performance measure across studies in this submission (see CS Appendix D.4.3, Table 19); however, KPS 90, 80, 70 and 60 were mismatched to their respective ECOG status. The ERG matched the performance scales (see Table 11 below) using the guidance provided by the ECOG-ACRIN Cancer Research Group.[26]

Some key prognostic variables were missing in the McMeekin *et al*. (2015) study,[17] including: FIGO stage, prior surgery, and number of lines of prior anti-cancer treatment, which may impact the robustness of the study. In addition, the IPDs were not available for McMeekin *et al*. (2015),[17] thus matching the study population with GARNET may lead to less robust comparisons compared to ZoptEC,[15, 16] consequently limiting it as a potential primary comparative efficacy evidence.

[redacted]Table 11: Baseline characteristics and efficacy outcomes in the published studies included in the ITCs, RWEQ and GARNET ITT population

Trial	GARNET ITT population (N=129)	GARNET-like UK RWE (RWEQ) cohort (N[redacted]	GARNET-like UK RWE (RWEQ) - PLD monother	Rubinstein *et al*. (2019) (N=20) [18]	Mazgani *et al*. (2008) (N=31) [19]	McMeekin *et al*. (2015) (N=248)* [17]	ZoptEC (N=255) [15, 16]	Julius *et al*. (2013) (N= 60) [20]**	Makker *et al*. (2013) (N= 17) [21]

			apy cohort (N=[redacted])						
Study design	Phase I open-label, single-arm (only Part 2B, Cohort A1 of interest)	Retrospective observational study	Retrospective observational study	Retrospective review of medical records of patients	Retrospective review of medical records of patients	Phase III open-label RCT	Phase III open-label RCT	Retrospective review of medical records of patients	Retrospective review of medical records of patients
Intervention	Dostarlimab	Basket of chemotherapy	PLD	Carboplatin + paclitaxel	Carboplatin + paclitaxel	Doxorubicin or paclitaxel monotherapy	Doxorubicin	PLD	Doxorubicin
Mean age, years (STD)	[redacted]	[redacted]	[redacted]	NR	NR	NR	63.8 (8.81)	66.8	NR
Median age, years (range)	[redacted]	[redacted]	[redacted]	67 (40 – 83)	NR	64 (33 – 88)	64 (28 – 87)	67 (34 – 87)	56 (36 – 78)
Age group n (%)									
<65 years	[redacted]	[redacted]	[redacted]	NR	NR	NR	136 (53.3)	NR	NR
65 to <75 years	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
≥ 65 years	[redacted]	[redacted]	[redacted]	NR	NR	NR	119 (46.7)	NR	NR
≥75 years	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Race n (%)									
White	[redacted]	[redacted]	[redacted]	NR	NR	213 (86)	240 (94.1)	44 (73.3)	16 (94.1)
Black	[redacted]	[redacted]	[redacted]	NR	NR	18 (7)	7 (2.7)	10 (16.7)	1 (5.9)

Asian	[redacted]	[redacted]	[redacted]	NR	NR	5 (2)	5 (2.0)	NR	NR
Other[a]	[redacted]	[redacted]	[redacted]	NR	NR	12 (5)	3 (1.2)	NR	NR
Unknown[b]	[redacted]	[redacted]	[redacted]	NR	NR	NR	0 (0.0	NR	NR
Performance status, n (%)	Study entry	Registry diagnosis	Registry diagnosis						
ECOG 0 (KPS 90-100)	[redacted]	[redacted]	[redacted]	NR	NR	165 (66.5)[c]	125 (49.0)	NR	NR
ECOG 1 (KPS 70-80)	[redacted]	[redacted]	[redacted]	NR	NR	80 (32.3)[c]	118 (46.3)	NR	NR
ECOG 2 (KPS (50-60)	[redacted]	[redacted]	[redacted]	NR	NR	2 (0.8)[c]	11 (4.3)	NR	NR
Not recorded	[redacted]	[redacted]	[redacted]	NR	NR	1 (0.4)[c]	1 (0.4)	NR	NR
Histology at diagnosis, n (%)									
Endometrioid	[redacted]	[redacted]	[redacted]	3 (15)	19 (61)	138 (55.6)	164 (64.3)	NR	5 (29.4)
Non-Endometrioid	[redacted]	[redacted]	[redacted]	17 (85)	12 (39)	109 (44.0)	91 (35.7)	NR	12 (70.6)
Missing	[redacted]	[redacted]	[redacted]	NR	NR	1 (0.4)	NR	NR	NR
FIGO stage[d], n (%)									
I	[redacted]	[redacted]	[redacted]	5 (25.0)	NR	NR	NR	NR	NR
II	[redacted]	[redacted]	[redacted]	3 (15.0)	NR	NR	NR	NR	NR
III	[redacted]	[redacted]	[redacted]	7 (35.0)	NR	NR	NR	NR	3 (17.6)
IV	[redacted]	[redacted]	[redacted]	5 (25.0)	NR	NR	NR	NR	14 (82.4)
Unknown	[redacted]	[redacted]	[redacted]	0 (0)	NR	NR	NR	NR	NR
Advanced (FIGO III or IV)	[redacted]	[redacted]	[redacted]	NR	NR	NR	94 (36.9)	NR	NR

Metastatic	[redacted]	[redacted]	[redacted]	NR	NR	NR	90 (35.3)	NR	NR
Recurrent	[redacted]	[redacted]	[redacted]	NR	NR	NR	71 (27.8)	NR	NR
Grade of disease at diagnosis, n (%)									
Grade 1	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Grade 2	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Grade 3	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Grade 4	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Not assessable	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Missing	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Prior anticancer treatment, n (%)									
Any prior anti-cancer treatment	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Surgery	[redacted]	[redacted]	[redacted]	NR	NR	NR	222 (89.2)	NR	NR
Radiotherapy	[redacted]	[redacted]	[redacted]	NR	NR	NR	138 (55.4)	NR	NR
Prior adjuvant chemotherapy	[redacted]	[redacted]	[redacted]	NR	NR	140 (57.0)	92 (36.9)	NR	NR
Number of prior lines of therapy post advanced/recurrent diagnosis, n (%)									
1	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
2	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
3	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
≥4	[redacted]	[redacted]	[redacted]	NR	NR	NR	NR	NR	NR
Median PFS (months) [e] (95% CI)	[redacted]	[redacted]	[redacted]	10.0 (2.0, 47.0)	Endometroid: 8.0 (5.02, 12.72)	4.0 (2.7, 4.3)	[redacted]	7.0 (NR)	2.1 (0.97, 2.7)

					Serous: 9.0 (3.59, 35.4)				
Median OS (months) (95% CI)	[redacted]	[redacted]	[redacted]	27.0 (6.0, 117.0)	Endometroid: 15.0 (9.13, 30.36) Serous: 26.0 (9.72, 71.4)	12.3 (10.7, 15.4)	10.8 (9.8, 12.6)	7.0 (NR)	5.8 (1.0, 15.0)

Footnotes: [a] Includes American Indian or Alaska Native. [b] Includes 'Not reported'. [c] McMeekin *et al.* (2015) reported Karnofsky performance status scale (100, 90, 80, 70, 60, NR), rather than ECOG PS. [d] FIGO: For the RWE study this is at registry diagnosis and for Rubinstein et al. (2019) this is at diagnosis. [e] PFS was estimated using time to next therapy (TTNT) as a proxy for RWEQ and RWEQ PLD monotherapy cohorts. ZoptEC baseline estimates N= 255 were provided in response to clarification question A17. *For McMeekin *et al*. 2015, the 248 sample relates to the comparator arm of interest (Paclitaxel or doxorubicin monotherapy). For McMeekin *et al*. 2015, PFS is calculated from efficacy set (N = 223). ** Only the 40mg/m^2 dose (standard clinical) of PLD has been used from Julius *et al.* (2013) in the Matched adjusted indirect comparison (MAICs): other doses have insufficient bases.
Abbreviations: ECOG PS: Eastern Cooperative Oncology Group Performance Status; KPS: Karnofsky performance status; FIGO: International Federation of Gynaecology and Obstetrics; STD: standard deviation; PFS: Progression Free survival; OS: Overall Survival.

3.3.2.2 Results of published studies included in the indirect comparisons

The median PFS for the published studies included in the indirect comparisons ranged from 2.1 (95% CI 0.97, 2.7) months in the Makker *et al.* 2013 study (doxorubicin)[21] to 10.0 (95% CI 2.0, 47.0) months in the Rubinstein *et al.* 2019 trial (carboplatin plus paclitaxel).[18] The median OS for the studies included in the indirect comparisons ranged from 5.8 (95% CI 1.0, 15.0) months in the Makker *et al.* 2013 study (doxorubicin)[21] to 27.0 (95% CI 6.0, 117.0) months in the Rubinstein *et al.* 2019 trial carboplatin plus paclitaxel[18] (see Table 11). Only the studies which included carboplatin plus paclitaxel therapy[18, 19] reported longer PFS and OS than GARNET; however, the ERG highlights that the wide confidence intervals reported and small sample sizes in the studies lead to uncertainties regarding these results.

ERG points of critique: Overall, the ERG notes the limited information available and associated uncertainties from most of the published studies makes it difficult to draw any meaningful conclusions. In addition to the UK RWEGARNET-like cohort, the doxorubicin arm of the ZoptEC trial[15, 16] may offer a valuable comparator population as the setting, data collection methods and patient characteristics were relatively aligned to the GARNET trial.

Safety

From the relevant published studies identified in the clinical SLR and included in the ITCs, only the ZoptEC trial[15, 16] (doxorubicin monotherapy) and McMeekin *et al*. (2015)[17] (paclitaxel or doxorubicin monotherapy) study had recorded safety information. Adverse events (AEs) were graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE).[27] The same NCI CTCAE version (version 4.03) was used in the GARNET and ZoptEC trial,[15, 16] while version 3.0 was used by McMeekin *et al*. (2015).[17] Table 12 below (reproduced from CS Table 41) shows a naïve comparison of the treatment-related TEAEs in GARNET, ZoptEC[15, 16]

and McMeekin *et al*. (2015).[17] Overall, [redacted] of patients in GARNET experienced any treatment-related TEAEs in comparison to 90% of patients in the McMeekin *et al.* (2015) study,[17] and 96.4% (nearly all) in the ZoptEC trial.[15, 16] Notable differences were also observed in the frequently of the type of individual treatment-related TEAEs (occurring in ≥5% of patients) between GARNET and ZoptEC[15, 16] (see CS Table 42). McMeekin *et al*. (2015)[17] reported treatment-related TEAEs occurring in ≥20% of patients. The ERG notes that only the ZoptEC trial[15, 16] reported raw AE data. Grade ≥3 TEAEs from the ZoptEC trial[15, 16] were included in the individual scenario analyses in the cost effectiveness evaluation.

ERG points of critique: Overall, the ERG notes due to the differences in trial protocols, the comparisons of safety information between studies should be approached with caution. Chemotherapy interventions appear to exhibit higher toxicity relative to dostarlimab; however, the lack of data from most of the published studies is associated with some uncertainties with regards to toxicity.

Table 12: Treatment-related TEAEs in GARNET, ZoptEC and McMeekin et al. (2015) (reproduced from CS Table 41)

Trial	GARNET ITT population (N=129)	ZoptEC (N=249)[15, 16]	McMeekin *et al.* (2015) (N=239)[17]
Intervention	Dostarlimab	Doxorubicin monotherapy	Paclitaxel *or* doxorubicin monotherapy
Any treatment-related TEAEs, n (%)	[redacted]	240 (96.4)	215 (90.0)
Any Grade ≥3 treatment related TEAEs, n (%)	[redacted]	NR	NR
Any treatment-related SAE, n (%)	[redacted]	NR	29 (12.0)

Abbreviations: ITT: intention-to-treat; NR: not reported; SAE: serious adverse event; SLR: systematic literature review; TEAE: treatment-emergent adverse event.

3.4 *Critique of the indirect comparison*

3.4.1 Company's approaches and general caveats for unanchored indirect comparison

The GARNET trial is a single-arm trial and did not include any comparators. It is necessary to derive estimates of relative effectiveness between dostarlimab and other treatments through unanchored indirect comparison. This means there is no shared common comparator (e.g. placebo) through which comparisons between dostarlimab and other comparators of interest can be 'calibrated' in some way using data from RCTs that preserve random allocation of treatments and balance known and unknown confounders between treatment arms within individual studies. Consequently, unanchored indirect comparison heavily relies on comprehensive identification and adjustment of all prognostic factors and effect modifiers. Even if this can be achieved, there is still risk of residual confounding caused by unknown confounders. Failure to account for major imbalance in prognostic factors and effect modifiers between treatment arms being compared will result in biased estimates, the accuracy of which is unknown. Where there is insufficient evidence that the degree of bias arising from imbalance in confounders remaining unaccounted for is acceptable, NICE Technical Support Document (TSD) 18 recommended that the findings "*should be heavily caveated by noting: the amount of bias (systematic error) in these estimates is unknown, is likely to be substantial, and could even exceed the magnitude of treatment effects which are being estimated*".[28]

In unanchored indirect comparisons, attempts are often made to generate 'adjusted' results using the individual patient data (IPD) available from an index study, which is usually the study for the technology of interest, or GARNET trial in this STA. The adjustments aim to predict what results might have been observed in the GARNET trial population if its distribution of prognostic factors and effect modifiers were similar to the patient population in the comparator study. Ideally the latter would include a representative 'target population' for whom the new technology is indicated, as the findings from the indirect comparison would reflect the expected clinical effectiveness in

the target population. Findings from unanchored indirect comparisons therefore need to be interpreted with the nature of the comparator population in mind.

As described in Section 3.3, the company identified various sources of data from their SLR in order to inform unanchored indirect comparisons between dostarlimab and relevant comparators. It is unclear if the inclusion and exclusion criteria applied in the feasibility assessment for indirect comparisons were specified post hoc. The ERG reviewed the company's stated reasons for excluding or including individual studies for the indirect comparisons and considered them reasonable.

Individual studies/datasets used as comparators and corresponding indirect comparisons are summarised in Table 13. The company has chosen the matching-adjusted indirect comparison (MAIC) method for its primary indirect comparison with miscellaneous treatments used in clinical management in England using an REWQ dataset obtained from registry (described earlier in Section 3.3.1). Separate MAICs were also conducted for supportive indirect comparisons with other individual comparators using data from published trials in the literature.

The company justified the choice of MAIC over an alternative method of simulated treatment comparison (STC), described alongside MAIC in NICE Technical Support Document (TSD) 18, by suggesting that MAIC produces a marginal (population-level) treatment effect while STC produces only conditional (patient-level) treatment effects and citing a commentary[29] that mainly focused on anchored rather than unanchored indirect comparison (see CS Section B.2.7.1). The ERG is not entirely convinced by this, as the availability of IPD from GARNET means the predicted outcomes for individual patients can be used to construct population-level treatment effect.

Table 13: Comparator datasets and corresponding indirect comparisons included in the CS

Comparator dataset	Nature	Comparator(s) included in the dataset	Methods of indirect comparison used	Company's designation of the analysis	Findings informed economic model (CS Section 3.8.3)
GARNET-like RWE (RWEQ)	IPD from registry	Wide range of treatment regimens used in clinical practice in England	MAIC	Primary	Scenarios 6 & 7
ZoptEC	IPD from RCT	Doxorubicin monotherapy	IPTW	Supportive	Scenario 35
Makker et al. (2013)	Aggregated data from literature	Doxorubicin monotherapy	MAIC	Supportive	Scenarios 35, 36, 37, 38,39
McMeekin et al. (2015)	Aggregated data from literature	Doxorubicin monotherapy & paclitaxel monotherapy	MAIC	Supportive	Scenarios 36 & 39
Julius et al. (2013)	Aggregated data from literature	PLD	MAIC	Supportive	Scenario 38
Rubinstein et al. (2019)	Aggregated data from literature	Carboplatin plus paclitaxel	MAIC	Supportive	Scenario 40
Mazgani et al. (2008)	Aggregated data from literature	carboplatin plus paclitaxel	MAIC	Supportive	Scenario 41

Abbreviations: IPD: individual patient data; IPTW: inverse probability treatment weighting; MAIC: matching-adjusted indirect comparison; PLD: pegylated liposomal doxorubicin; RCT: randomised controlled trial

For the supportive indirect comparison between dostarlimab and doxorubicin monotherapy using individual patient data (IPD) obtained from the ZoptEC trial, the inverse probability treatment weighting (IPTW) method was used. The company justified the choice of IPTW in preference over propensity score matching (CS Appendix D, Section D.5.2) given the relatively small number of patients from each of the trial arms and that many patients may be eliminated in the matching process, which would impact

on interpretation of findings and reduce statistical power. ERG agrees with this. However, the rationale for choosing IPTW over STC method was not clearly stated. Given the challenges in clearly specifying the correct model for unanchored indirect comparison, there may be scope for using both methods (or adopting a doubly robust estimation methods described in TSD 18) to verify the validity of the analyses and robustness of the findings.

More detailed critique of individual unanchored indirect comparisons is provided below. The ERG focussed on GARNET versus (vs) clinical management using RWEQ and GARNET versus doxorubicin using ZoptEC as IPD for these two comparators were available to the company.

3.4.2 GARNET vs RWEQ (dostarlimab vs current clinical management)

3.4.2.1 Comparability of patient characteristics and datasets

In view of the scarcity of alternative data, the company sourced data from the NCRAS to create RWEQ cohort (see Section 3.3.1), which could potentially provide a suitable comparator dataset that represents current UK practice. Nevertheless, the difference in nature between GARNET (with data collected following a strict protocol in a trial setting) and RWEQ (with data retrospectively retrieved from registry collected during routine practice) poses substantial challenges in harmonising the two datasets and allowing a fair comparison to be made. Having examined the methods and findings of this unanchored indirect comparison, the ERG has strong reservation concerning its validity and the suitableness of the findings to support the base case.

As described in Section 3.3.1, there are major differences between GARNET and RWEQ, both in terms of the characteristics of patients included and in terms of the methods by which and settings in which the data were collected. Imbalance in patient characteristics (that are likely to be prognostic factors and effect modifiers) could be adjusted to some extent using appropriate statistical techniques. However, more concerning are systematic differences between the two cohorts of patients and related

data arising from methodological issues associated with data collection, case definition and selection (in particular, the necessary and yet complicated processes of reducing from 45,494 patients with EC diagnosis in the registry to the [redacted] patients included in the final RWEQ cohort). These systematic biases may not be easily recognised and cannot be ‘adjusted away’ by statistical means.

The major differences in patient characteristics between GARNET and RWEQ as described in Section 3.3.1.1 and Table 10 (e.g. a much higher proportion of patients with endometrioid disease in GARNET, [redacted] vs RWEQ, [redacted]) suggested a systematic difference in how patients were selected into the two cohorts, which raise some concerns regarding the comparability of the two datasets even after statistical adjustment.

In addition to the clear difference in baseline characteristics between GARNET and RWEQ, findings from company’s analysis to verify prognostic factors also provide strong evidence that the two cohorts may have some fundamental differences. For example, the effect of tumour grade on OS was shown to be in opposite directions in separate Cox regression models for the two cohorts: HR (grade 3/4 vs 1/2) [redacted] (95% CI [redacted] to [redacted]) for the GARNET cohort compared to HR [redacted] (95% CI [redacted] to [redacted]) for the RWEQ cohort.

The marked differences between GARNET and RWEQ populations also raised the issue of whether the findings of the MAICs reflect what would be observed in the target population as defined in final scope. Data obtained from a registry are often considered more representative of patients encountered in clinical practice than patients recruited into clinical trials, and therefore using RWEQ as the comparator could be an advantage in the context of unanchored indirect comparison because the process of statistical adjustment aims to predict what outcome would look like if the GARNET trial population had a similar distribution of prognostic factors as seen in the comparator population. Nevertheless, representativeness of RWEQ here may be compromised by the many selection criteria retrospectively applied to the original RWE dataset and the imprecise methods for identifying recurrent cases to reach the highly selective RWEQ cohort. The resultant unanchored indirect comparison may therefore reflect findings that are

applicable only to a patient population that is difficult to define and not necessarily reflecting what would be expected in the target population.

In addition to the very limited prognostic factors taken into account in the MAICs, ERG noted several other issues in the process of selecting matching variables:

- Using ECOG PS at treatment initiation for GARNET ITT but using ECOG PS at registry (initial) diagnosis for RWEQ
- Modelling a very small number of patients with unknown histology and cancer grade as a separate category rather than treating them as missing data
- Lumping FIGO stage 3 and stage 4, which could be associated with quite different prognosis together in the analysis.

3.4.2.2 Methods of MAIC

In CS B.2.7.1, the company stated that "*The primary endpoint analysis considered in the UK RWE study MAIC utilised a Cox proportional hazards model, using weights obtained using the MAIC method.*" The company started with a list of potential prognostic factors identified from a 'targeted literature review' and subsequently selected by an expert panel (CS Appendix M); and then narrowed down the final matching variables by fitting two separate Cox proportional hazard models (one for GARNET and one for RWEQ) and retaining any variables that attained the level of significance p≤0.1 in at least one of the two datasets.

ERG considers the list generated by the expert panel (see Table 14) to be reasonably comprehensive but makes the following observations:

(1) Based on another systematic review conducted by the company (only a conference abstract was cited),[23] there is no evidence that MMR/MSI status has prognostic value among patients with recurrent or advanced EC receiving non-anti-PD-(L)1 therapy (CS Section B.2.3.2, page 50). However, as noted earlier in ERG report Section 3.3.1, the prevalence of dMMR/MSI-H differs between type I and type II EC, which in turn are

associated with various prognostic factors; therefore the differences in the distribution of MMR/MSI status between GARNET and RWEQ cohorts could still result in confounding and cause bias in the indirect comparison.

(2) The following potential prognostic factors were identified in the literature but were not selected by the expert panel:

- For good prognosis: absence of other systemic disease, smaller tumour size, resectability, longer disease-free interval, positive oestrogen and progesterone receptor, PTEN mutations.
- For poor prognosis: advanced EC (relative to recurrent EC), increased number of positive lymph nodes, substantial lymphovascular space invasion, desmoplasia in lymph nodes, extension of carcinoma into perinodal adipose tissue, distant recurrence, P53 gene mutation.

The rationale for excluding these potential prognostic factors was not described. ERG considers some of these factors such as disease-free interval and advanced vs recurrent EC to be potentially important.[30]

Table 14: Comparison of possible prognostic factors between those identified in the literature, selected company's expert panel and included in company's MAIC for GARNET vs RWEQ

Potential prognostic factors	Identified from company's targeted literature review	Selected by company's expert panel	Included in MAIC scenario 1	Included in MAIC scenario 2
Absence of other systemic disease	Yes	No	No	No
Race (Non-Hispanic White)	Yes	Yes	No	Yes
Increased Age	Yes	Yes	No	No
Smaller tumour size	Yes	No	No	No
Resectability / Prior surgery for study indication	Yes	Yes	No	Yes
Longer disease-free interval	Yes	No	No	No
Good Performance status	Yes	Yes	No	No
Advanced EC vs recurrent EC	Yes	No	No	No
FIGO	Yes*	Yes	No	Yes
Grade of disease at diagnosis	No	Yes	No	No
Number of prior platinum-based therapies	No	Yes	Yes	No
Histology: Serous & clear cell cancer	Yes	Yes	Yes	Yes
Increased number of positive lymph nodes	Yes	No	No	No
Substantial lymphovascular space invasion	Yes	No	No	No
Desmoplasia in lymph nodes	Yes	No	No	No
Extension of carcinoma into perinodal adipose tissue	Yes	No	No	No
Distant recurrence	Yes	No	No	No
Positive oestrogen and progesterone receptor	Yes	No	No	No
PTEN mutations	Yes	No	No	No
P53 gene mutation	Yes	No	No	No
MMR/MSI status	No	Yes	No	No

Footnote: *Described as: "Histology: FIGO grade 3"

The company constructed two scenarios (two final models): scenario 1 was based on prognostic factors identified by the expert panel; scenario 2 was based on variables identified from the above Cox proportional hazard model selection process. ERG is highly concerned with regard to whether the very limited matching variables included in these two scenarios enabled sufficient adjustment of imbalance in key prognostic factors between GARNET and RWEQ (see Table 14 below). No information on goodness of fit for the models or assessment of the magnitude of potential residual bias were presented.

As the company had access to IPD for both GARNET and RWEQ, it could have been possible for the company to carry out the MAIC by matching RWEQ to GARNET and created an adjusted RWEQ to be compared with unadjusted GARNET ITT as a sensitivity analysis and validity check.

Given the issues highlighted above related to both the datasets and the methods, ERG has strong reservations regarding the validity of the findings from these MAICs.

3.4.3 GARNET vs ZoptEC (dostarlimab vs doxorubicin)

3.4.3.1 Comparability of patient populations and datasets

As described in Section 3.3.2, the company sourced IPD from a ZoptEC trial identified in their SLR, which allow an unanchored indirect comparison to be carried out between dostarlimab and doxorubicin. Table 11 in Section 3.3.2 of this report and CS Appendix Table 40 shows that the baseline characteristics of patients were broadly similar between GARNET and ZoptEC, except for ethnicity, ECOG PS, and possible FIGO stage. Some of the differences were removed by excluding patients before indirect comparison was performed (see CS Appendix D.5.2, Table 44). Primarily, [redacted] patients with ECOG PS score 2 from ZoptEC trial were excluded as GARNET trial only included patients with ECOG PS score 0 or 1), and [redacted] patients in GARNET who had more than one prior line of platinum therapy were excluded because patients in ZoptEC only had one prior line of platinum therapy. These exclusions seem reasonable, but reduced the

sample sizes and thus statistical power for the indirect comparison. The company also excluded [redacted] patients with follow-up of longer than 36 months for doxorubicin group of the ZoptEC trial (CS Appendix D.5.2, Table 44). This exclusion might have introduced bias as the excluded patients would have had longer survival.

3.4.3.2 Methods for MAIC

The MAIC was carried out using a stabilised inverse probability of treatment weighting (IPTW) approach. This method was chosen in preference over propensity score matching (PSM) because of the relatively small sample sizes of the trials, as more patients may be eliminated during the PSM process. ERG agrees with this rationale, although it is not clear whether an alternative method of simulated treatment comparison was considered.

Overall, the methods for the MAIC using IPTW were described in good detail and were justified. The company stated that grade of tumour could not be included in matching due to violation of positivity assumption (CS Appendix D.5.2, page 117). This suggested patients with certain tumour grade rarely or never received either dostarlimab or doxorubicin, which would cause technical problems during the matching process, but further details were not provided. Analysis of potential impact of unmeasured confounding was provided and showed the findings of the MAIC were reasonably robust. The company did not perform IPTW for PFS, citing the differences in the definitions of PFS and the timepoints of tumour assessments between GARNET and ZoptEC (CS Section B.2.7.2). ERG believes such analysis could have been undertaken as a sensitivity analysis.

3.4.4 GARNET vs other comparators

3.4.4.1 Dostarlimab vs carboplatin + paclitaxel

Combination therapy of carboplatin plus paclitaxel is the most commonly used treatment regimen in the NHS for the target patient population, as reflected in RWEQ (used by [redacted]% of patients, see CS Table 14). The company identified two studies (Rubinstein *et al*.2019 and Mazgani *et al*. 2008) [18, 19] providing potentially relevant data for this comparator (see ERG report Section 3.3.2 and Table 11). ERG noted that the median PFS reported in these studies was [redacted] than that was reported for dostarlimab in GARNET before any adjustments were made. Both were retrospective studies of small sample sizes (n=20 and 31 respectively) and reported very limited information concerning prognostic factors and effect modifiers that would allow adjustments be made through MAICs (see Table 15 below). Because of these limitations, the findings from the MAICs were highly uncertain.

3.4.4.2 Dostarlimab vs paclitaxel monotherapy, doxorubicin monotherapy or pegylated liposomal doxorubicin (PLD) monotherapy

The company identified three additional studies in which relevant data for patients receiving paclitaxel or doxorubicin monotherapy (McMeekin et al. 2015), doxorubicin monotherapy (Makker et al. 2013) and PLD monotherapy (Julius et al. 2013) were available. Of these, only McMeekin et al. 2015 was a prospective trial with a relatively large sample size, but it also reported very limited information on prognostic factors and effect modifier to allow comprehensive adjustment (see Table 15 below). MAICs undertaken using the other two studies suffered from very small sample sizes (the effective sample sizes for GARNET also became much smaller during the matching process) and very limited adjustment and so the findings were also highly uncertain.

Table 15: Methodological features of MAICs presented in the CS

Source of comparator	Design	Therapy	Analysis	Matching variables	Validity assessment
RWEQ (n=[redacted])	Retrospective, UK registry	Clinical management	MAIC, scenario 1, vs GARNET (ESS=[redacted])	Histology Number of prior platinum-based therapies	Limited matching; possible violation of PH assumption; no assessment of residual bias
RWEQ (n=[redacted])	Retrospective, UK registry	Clinical management	MAIC, scenario 2, vs GARNET (ESS=[redacted])	Race/ethnicity Stage at diagnosis ECOG PS Histology Prior surgery	Limited matching; no assessment of residual bias
RWEQ ECOG PS ≤1 (n=[redacted])	Retrospective, UK registry	Clinical management	MAIC, scenario 1 (sensitivity analysis) vs GARNET (ESS=[redacted])	Histology Number of prior platinum-based therapies	Limited matching; possible violation of proportional hazard assumption
RWEQ ECOG PS ≤1 (n=[redacted])	Retrospective, UK registry	Clinical management	MAIC, scenario 2 (sensitivity analysis) vs GARNET (ESS=[redacted])	Race/ethnicity Stage at diagnosis ECOG PS Histology Prior surgery	No assessment of residual bias
ZoptEC (n=[redacted])	Trial	Doxorubicin	IPTW, main analysis, vs GARNET (n=[redacted]) OS only	Age Race ECOG PS Histology FIGO stage at baseline (Stage I/II versus Stage III/IV) Prior surgery	Tumour grade could not be adjusted due to violation of the positivity assumption; did not adjust for prior lines of therapy..

ZoptEC (n=[redacted]) [15, 16]	Trial	Doxorubicin	IPTW, sensitivity analysis, vs GARNET (n=129) OS only	Age Race ECOG PS Histology FIGO stage at baseline (Stage I/II versus Stage III/IV) Prior surgery	Tumour grade could not be adjusted due to violation of the positivity assumption; did not adjust for prior lines of therapy..
Rubinstein *et al*.2019 (n=20) [18]	Retrospective, single centre, USA	Carboplatin + paclitaxel	MAIC vs GARNET (ESS [redacted])	Histology	Very limited matching; violation of proportional hazard assumption for both PFS & OS
Mazgani *et al*. 2008 (n=31) [19]	Retrospective, single agency, Canada	Carboplatin + paclitaxel	MAIC vs GARNET (ESS [redacted])	Histology	Very limited matching; possible violation of proportional hazard assumption for PFS
McMeekin *et al*. 2015 (n=248) [17]	Trial	Paclitaxel (n=68) or doxorubicin (n=171)	MAIC, vs GARNET (ESS [redacted]) OS only	Race ECOG PS Histology	Very limited matching
Makker *et a*l. 2013 (n=17) [21]	Retrospective, single centre, USA	Doxorubicin	MAIC, vs GARNET (ESS=[redacted])	Race ECOG PS Histology	Very limited matching
Julius *et al*. 2013 (n=60) [20]	Retrospective, single centre, USA	PLD (n=41 for 40 mg/m^2)	MAIC, vs GARNET (ESS=[redacted]) OS only	Race	Very limited matching

Abbreviation: ESS: effective sample size

3.4.4.3 Dostarlimab vs hormone therapy

The company conducted a targeted literature review (CS Appendix L), but did not identify any studies that provide suitable data for the population of interest to enable an

indirect comparison. ERG checked the reasons stated by the company for study exclusion and considered them to be reasonable. ERG also undertook a separate search and did not identify any additional studies (see Section 3.1). Therefore, ERG agrees that there is currently a lack of data to allow reliable comparison be made between dostarlimab and hormone therapy in the population of interest.

3.4.5 Summary of critique of the indirect comparisons

As GARNET is a single arm trial without including a comparator, relative effectiveness between dostarlimab and comparator treatments has to be estimated through unanchored indirect comparisons, which are very susceptible to biases arising from differences in clinical and methodological features between different studies/data sources. The company identified two datasets with IPD and several other published studies with aggregate data, and undertook a suite of unanchored indirect comparisons using MAICs. However, ERG considered findings from all these MAICs to be highly uncertain due to a combination of the nature of the IPD datasets, limited information presented in published literature and issues related to MAIC methodology. The findings expressed as hazard ratios are summarised in Table 16, which should be interpreted with caveats highlighted below:

- The RWEQ cohort has very different characteristics compared with the GARNET population and the differences suggest RWEQ cohort was likely have more aggressive and advanced diseases and to be less fit compared with the GARNET trial population. Many issues related to the nature of the datasets and methods indicate that the MAICs comparing GARNET with RWEQ, which produced estimates more favourable for dostarlimab, are unlikely to be valid. ERG therefore prefers the unadjusted comparison over any of the MAICs for GARNET vs RWEQ, acknowledging that the estimates are likely to be biased in favour of dostarlimab.

- RWEQ included a basket of different treatments used in the UK clinical practice. A significant proportion of patients in the cohort were offered single agent

regimens that mean they were not fit for combination regimens, likely reflecting disease burden in stage 4 disease. They were more likely to be advanced stage at diagnosis than recurrent after successful initial management and therefore their overall outlook was likely worse from the start compared with GARNET trial population. It would be extremely difficult (if not impossible) to fully address the imbalance in known and unknown prognostic factors between the cohorts by statistical adjustment. An RCT of dostarlimab vs standard care might be the only way to obtain unbiased estimates.

- The IPTW unanchored indirect comparison between dostarlimab and doxorubicin using IPD from ZoptEC trial overcame some of the inherent limitations in registry data (i.e. RWEQ) that may be intractable. However, some important factors such as tumour grade and prior lines of therapy could not be matched.

- Most of the remaining MAICs based on published literature were limited by small sample sizes and very limited matching and therefore the level of uncertainty associated with the validity and representativeness of these findings is very high. ERG noted that (given similar comparator treatments, e.g. doxorubicin or PLD monotherapy), the estimated benefits for dostarlimab tend to be larger when the comparator data were sourced from retrospective studies than from prospective trials.

Table 16: Findings from company's MAICs, expressed as hazard ratios (HRs) for PFS and OS

Study/data set & design	Comparator	Analysis	ESS for GARNET	HR dostarlimab vs comparator	
				PFS	OS
RWEQ Retrospective (n=[redacted])	Clinical management	Unadjusted	129	Not estimated	[redacted]
RWEQ Retrospective (n=[redacted])	Clinical management	MAIC, scenario 1	[redacted]	Not estimated	[redacted]
RWEQ Retrospective (n=[redacted])	Clinical management	MAIC, scenario 2	[redacted]	Not estimated	[redacted]
ZoptEC Trial (n=[redacted])	Doxorubicin	IPTW, main analysis	[redacted]	Not estimated	[redacted]
ZoptEC Trial [15, 16] (n=[redacted])	Doxorubicin	IPTW, sensitivity analysis	129	Not estimated	[redacted]
Rubinstein *et al*.2019 [18] Retrospective (n=20)	Carboplatin + paclitaxel	MAIC	[redacted]	[redacted]	[redacted]
Mazgani *et al*. 2008 [19] Retrospective (n=31)	Carboplatin + paclitaxel	MAIC	[redacted]	[redacted]	[redacted]
McMeekin *et al*. 2015 Trial [17] (n=239)	Paclitaxel (n=68) or doxorubicin (n=171)	MAIC	[redacted]	No data	[redacted]
Makker *et al*. 2013 [21] Retrospective (n=17)	Doxorubicin	MAIC	[redacted]	[redacted]	[redacted]
Julius *et al*. 2013 [20] Retrospective (n=41)	PLD	MAIC	[redacted]	No data	[redacted]

3.5 *Additional work on clinical effectiveness undertaken by the ERG*

This section describes two pieces of additional work undertaken by the ERG to facilitate interpretation of clinical effectiveness evidence. The first work involves an unadjusted comparison of PFS and OS survival curves between the GARNET trial and other trials of PD-1 or PD-L1 inhibitors for recurrent or advanced EC to verify the company's claim that extended (flat) tails are a 'hallmark of I-O therapy' (CS page 146 and 199). The second work explored the possibility that data from trial settings tend to over-estimate treatment effectiveness compared with data obtained from real-world setting by making an unadjusted comparison of PFS and OS outcomes between ZoptEC trial (doxorubicin monotherapy) and the subset of RWEQ data for patients treated with PLD (pegylated liposomal doxorubicin) monotherapy provided by the company in response to ERG's clarification questions.

3.5.1 GARNET versus other trials of PD-1 or PD-L1 inhibitors

In the absence of longer-term data from GARNET, the ERG considered evidence from trials for other PD-1 or PD-L1 inhibitors with longer follow-up periods and reported survival curves in post platinum, second line treatment of recurrent or advanced EC, and conducted a rapid analysis to assess if the shape of the survival curves from GARNET are truly unique or characteristic of I-O therapy. The ERG is aware that the shape of survival curves and the extent and positioning of flat tails is dependent on many factors, not only class of intervention (e.g. PD-1 or PD-L1 inhibitors), but including maturity of data (proportion of participants experiencing the event) which in turn is influenced by the length of follow up, the severity of the disease and the effect on event rate of interventions, and heterogeneity of the included population. Table 17 summarises the study characteristics and survival outcomes for other PD-1 or PD-L1 targeted interventions.

Makker *et al*. 2019 and 2020[31, 32] is a single arm phase 2 study of pembrolizumab plus lenvatinib (from an interim analysis and more mature analysis, respectively), with longer study follow-up than GARNET, and patient characteristics similar to GARNET. Figure 4 shows PFS and OS KM plots for Makker *et al*. (2019 and 2020) study[31, 32] versus GARNET. More mature data from Makker reduces the flat tail and introduces events that move the PFS KM plot more toward baseline. It seems possible that more mature data for GARNET might have the same PFS and OS pattern as the Makker et al. (2020) study.[32]

Ott *et al*. (2017)[33] is a single arm phase 1 study of pembrolizumab, with longer study follow-up than GARNET, smaller sample size and less comparable patient characteristics (such as age) to GARNET. Figure 5 shows PFS and OS KM plots for Ott *et al*. (2017)[33] study versus GARNET. The shapes of the plots are similar; however, the faster rate of events in Ott *et al*. (2017)[33] means the flat tail gets closer to zero survival and becomes less influential.

Overall, the rapid analyses conducted by the ERG showed that the extended tail in I-O therapies is likely subdued when follow up is sufficiently extended. This is supported by further exploratory analyses of survival data from trials of check point drugs in non-small cell lung cancer (NLSCLC) shown in ERG Appendix 9.2.

Table 17: Study characteristics and survival outcomes for other PD or PD-L targeted interventions

Author	• **Study design** • **Follow-up** • **Prior platinum therapy** • **Sample size** • **Age (mean), years** • **FIGO stage** • **ECOG PS**	**Intervention**	• **Definition of PFS** • **PFS (months)**	**OS (months)**
Makker *et al*. (2020)[32]	• Ongoing phase 2 study • Median follow-up of 18.7 months	Oral lenvatinib 20 mg once daily plus 200 mg intravenous	• Median PFS: 7.4	Median OS: 16.7

	• Yes • 108 patients • 65.1 • FIGO stage: 1 (n =12), 2 (n =19), 3 (n =24), not reported (n =53) • ECOG PS: 0 (n =53), 1 (n =55)	pembrolizumab once every 3 weeks, in 3-week cycles.		
Makker *et al*. (2019)[31]	• Ongoing phase 2 study • Median study follow-up was 13·3 months • Yes • 53 patients • 64 • FIGO stage: 1 (n =5), 2 (n =11), 3 (n =6), not reported (n = 31) • ECOG PS: 0 (n =20), 1 (n =33)	Oral lenvatinib 20 mg daily plus 200 mg intravenous pembrolizumab once every 3 weeks.	• Defined as the time from first study dose to date of first documented disease progression or death, whichever occurred first • With a median follow-up for progression free survival of 7·7 months • 27 (51%) patients had disease progression or had died, median progression-free survival was 7·4 months (95% CI 5·0 to not estimable).	NR
Ott *et al*. (2017)[33]	• Multicohort phase Ib KEYNOTE-028 trial • Median follow-up duration was 76.2 weeks • Yes (mostly n =25) • 24 patients • 67 • FIGO stage: NR • ECOG PS: 0 (n =7), 1 (n =16), not reported (n=1)	Pembrolizumab ,10 mg/kg every 2 weeks for up to 24 months or until progression or unacceptable toxicity.	• PFS defined as time from allocation to the first documented disease progression according to RECIST (version 1.1) or death resulting from any cause. • Median PFS: 1.8 (95% CI, 1.6 -2.7)	Median OS: 4.3 to not reached. 6-months OS rates: 67% 12-months OS rates: 51%

			• 6-months PFS rates: 19% • 12-months PFS rates: 14.3%	

Figure 4: PFS and OS KM plots for Makker et al. (2019 and 2020) study versus GARNET

Figure 5: PFS and OS KM plots for Ott et al. (2017) study versus GARNET

3.5.2 [redacted] (doxorubicin arm)

In addition to the differences in the patient characteristics between RWEQ PLD and ZoptEC (doxorubicin arm)[15, 16] described by the ERG in report section 3.3.2.1), the ERG

conducted analyses to assess the potential difference in effectiveness outcomes between RWEQ PLD and ZoptEC (doxorubicin arm)[15, 16] (Figure 6 and Figure 7 below). *Note: ZoptEC n=[redacted] populations represent the derived main analysis set used for the PFS and OS ITCs*. Given the broad equivalence between doxorubicin and PLD, better outcomes observed for doxorubicin monotherapy in ZoptEC [15, 16] compared with PLD monotherapy in RWEQ would suggest potential under-estimation of treatment effects of chemotherapy in real-world setting compared with those obtained in a trial setting. This in turn would lend support to the possibility that the use of RWEQ might have resulted in an under-estimation (of a similar magnitude) of the effects of the basket of therapies used in real-life clinical practice, compared with if they had been evaluated in a trial setting that is more comparable to the GARNET.

Figure 6: PFS for RWEQ PLD monotherapy versus ZoptEC (doxorubicin arm)

Figure 7: OS for RWEQ PLD monotherapy versus ZoptEC (doxorubicin arm)

3.6 *Conclusions of the clinical effectiveness section*

The CS presents evidence from GARNET, a Phase 1, single-arm, open-label study of dostarlimab conducted in 9 countries (including 9 centres in UK).

A total of 129 patients received any amount of dostarlimab, and this population was used in the base case cost-effectiveness analyses. Clinical outcomes suggested a potential for positive response to treatment with dostarlimab; however, the pivotal trial of dostarlimab has a short follow-up time frame and some outcomes do not have enough data to be fully informed. In the absence of a comparator group, it is unclear whether there is a meaningful improvement over established clinical management.

Evidence for the comparator (basket of chemotherapies) was taken from the RWEQ cohort of the UK RWE study funded by GSK. The RWEQ cohort included [redacted] patients. Supportive indirect comparisons with other individual comparators were also conducted using data from published studies in the literature.

Overall, the ERG's key concerns in the clinical effectiveness are:
The magnitude of the benefit of dostarlimab over treatment with chemotherapy and hormone therapy is uncertain. The main source of evidence was a phase I trial, with immature data and no comparator arm, and comparison with chemotherapy was from unanchored indirect treatment comparisons.

There are uncertainties with regard to whether the procedures for retrospectively selecting patients into the final RWEQ cohort in the UK RWE study produced a patient cohort that is representative of the target patient population in the UK. There are major differences in setting, patient characteristics and case definitions between the GARNET trial population and the RWEQ cohort. The major differences between the GARNET trial population and the RWEQ cohort remained after the matching process in the primary MAICs. Limited prognostic factors could be adjusted for in the supportive MAICs using other sources of comparator evidence. Estimates of relative effectiveness between dostarlimab and comparator treatments obtained from both unadjusted comparisons and the MAICs presented in the CS are highly uncertain and are likely to be biased in favour of dostarlimab.

4 COST EFFECTIVENESS

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

4.2 *The company presents an extensive systematic literature review of economic evaluations, quality of life values and resource use. This appears to have been competently conducted, is well summarised but is of limited use given the disease area and in particular the lack of relevant quality of life studies.* *Summary of the company's submitted economic evaluation*

4.2.1 NICE reference case checklist

Table 18: NICE reference case checklist

Element of health technology assessment	Reference case	ERG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes.
Perspective on costs	NHS and PSS	Yes.
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes, cost utility analysis. The company base case makes a naïve comparison between dostarlimab and a pooled real world data comparison. Scenarios that compare dostarlimab with individual treatments are also presented. A fully incremental analysis is not presented. The ERG thinks this is reasonable given the base case and that the individual treatments will be used for different groups of patients based upon their fitness.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes. 40 years.

Element of health technology assessment	Reference case	ERG comment on company's submission
Synthesis of evidence on health effects	Based on systematic review	The base case compares dostarlimab with a real world basket of treatments. The scenarios around individual treatments are rooted in a systematic review.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes. EQ-5D-5L cross walked to EQ-5D-3L and valued using the standard UK social tariff.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes. The standard UK social tariff.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes. The standard UK social tariff.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes.
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes.
PSS, personal social services; QALYs, quality-adjusted life years; EQ-5D, standardised instrument for use as a measure of health outcome.		

4.2.2 Model structure

The company presents a partitioned survival analysis with the usual three main health states of progression free survival (PFS), post progression survival (PPS) and dead. This uses a Markov model with a 3 week cycle to match the dostarlimab infusion

frequency. The distribution of patients between the three main health states is determined by the overall survival (OS) curve and the PFS curve.

The OS and PFS curves for dostarlimab are derived by fitting parameterized curves to the GARNET Kaplan Meier (KM) OS and PFS data. For the comparator arm the OS and PFS curves are estimated by fitting parameterized curves to the RWEQ KM OS and Time to Next Treatment (TTNT) data, TTNT being used as a proxy for PFS due to progression data not being available for the RWEQ.

The time on treatment curves are estimated by fitting parameterized curves to the KM Time to Treatment Discontinuation (TTD) data of GARNET and the RWEQ.

Unusually, and in part justified by the approach of TA571, the company imposes stopping rules for dostarlimab, assuming that at [redacted] all but [redacted] of patients stop treatment and at [redacted] all patients stop treatment.

Due to the dostarlimab treatment stopping rules the company applies a waning treatment effect to the dostarlimab OS and PFS curves. The company assumes that the treatment effect is retained for [redacted] after stopping dostarlimab, so the extrapolated dostarlimab OS and PFS curves are unaffected by treatment cessation. After this it takes another [redacted] for all the treatment effect to be lost, with the dostarlimab OS and PFS efficacy being equalized with the contemporaneous RWEQ OS and PFS efficacy.

As the ERG found some of the company submission difficult to follow and lacking some detail, the detail of the company modelling is presented in sections:

4.2.6 from page 100100: treatment effects and extrapolation;

4.2.7 from page 115115: health related quality of life; and

4.2.8 from page 116116: resource use and costs.

While many readers will prefer to skip forward to section 4.3 on page 120120 which presents the main ERG critique of the company economic modelling, the graphical

presentation of the company curves and expert responses of section 4.2.6 may be more easily digestible than those of the company submission.

4.2.3 Population

The population reflects the scope but is subject to the concerns raised about the naïve comparison in the clinical review section.

- For dostarlimab the efficacy estimates are drawn from the GARNET population.
- For the comparator arm the efficacy estimates are drawn from the RWEQ population, pooled across the various chemotherapy regimens in the RWEQ data set.

4.2.4 Interventions and comparators

For the company base case the company compares dostarlimab with the basket of chemotherapy treatments of the RWEQ data set, though for costing only includes treatments which comprised more than 5% of the RWEQ data set. For costing it is also assumed that some comparator arm patients will receive hormone therapy.

4.2.5 Perspective, time horizon and discounting

The perspective and discounting is as per the NICE reference case. The time horizon is 40 years, which is sufficient to capture the extrapolated OS curves.

4.2.6 Treatment effectiveness and extrapolation

Unusually, given the assumptions about dostarlimab stopping rules and treatment waning, the OS and PFS modelling is best understood by reviewing the TTD curves first, followed by the comparator RWEQ OS and PFS curves. The dostarlimab OS and PFS curves estimated from GARNET can then be presented, followed by a presentation of how the treatment stopping rules and waning to RWEQ effectiveness affects these curves.

4.2.6.1 TTD Curve: dostarlimab

The company states that it fits a range of parameterized curves to the ITT (N=129) GARNET TTD KM data.

Table 19: Company GARNET TTD parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM	SPL1	SPL2
AIC									
BIC									
Sum									

The information criteria minima are highlighted in bold, with the company choice highlighted by a bold border. The company selects the log-logistic curve, stating that "*the Gompertz and the log-logistic models were considered to provide the best statistical fit*". The ERG notes that the Gompertz has better AIC and BIC than the log-logistic, with their combined total being somewhat below that of the log-logistic.

In GARNET the KM proportion remaining on treatment at [redacted] Despite this, and partly justified by the approach of TA571, the company assumes that at [redacted] all but [redacted] of patients will discontinue dostarlimab and that at [redacted] all patients will discontinue dostarlimab.

[redacted]9

If the spline models are discounted as unnecessary due to long term extrapolations being unnecessary the Gompertz has the best information criteria. The reasoning behind the choice of the log-logistic is unclear. Within the company model the average time spent on treatment is [redacted] months if the log-logistic is applied and [redacted] months if the Gompertz is applied: a difference of [redacted].

The dostarlimab TTD curve of [redacted] is critical to the modelling. Most obviously, it determines the costs of dostarlimab within the model. But perhaps even more importantly it determines the OS and PFS curves in the dostarlimab arm. Given the discontinuations at year [redacted] the company assumes that the treatment effect of dostarlimab is retained in full for [redacted] year after this, but then wanes during years [redacted]

and [redacted] so that "*at* [redacted] *years … the efficacy associated with dostarlimab was assumed to be equal to the efficacy associated with current clinical management*".

4.2.6.2 TTD Curve: comparator RWEQ

While the comparator RWEQ TTD curve does not affect anything in the dostarlimab arm, it's derivation is presented here so as to sit alongside that of dostarlimab.

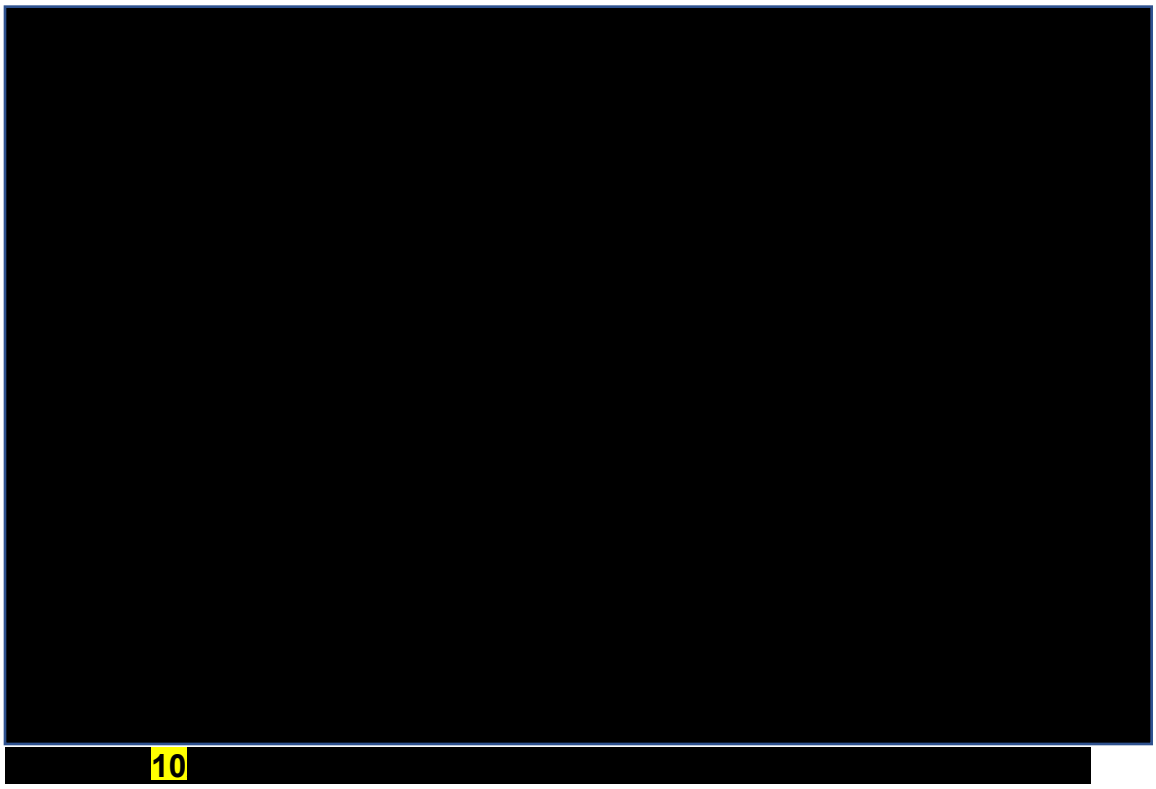


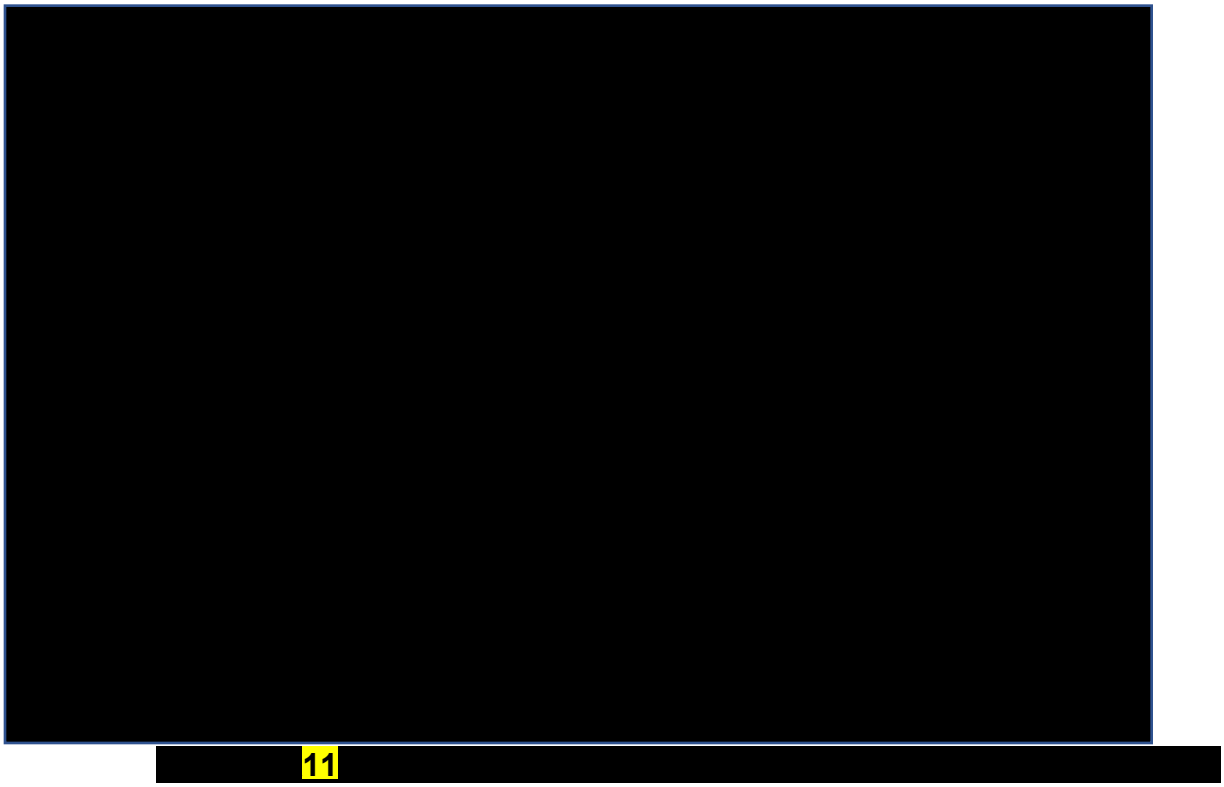
Table 20: Company RWEQ TTD parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM
AIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sum	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The company states "*the generalised gamma and gamma model provided the best fit to the observed ToT data from the UK RWE study. The generalised gamma model was therefore included in the base case*". Despite the gamma having a lower sum of

information criteria and perhaps being the more natural choice, it can be noted that the modelled discounted time on treatment is virtually identical for the two curves.

4.2.6.3 OS curve: comparator RWEQ



At 65 months, the parameterised curves have broadly grouped into those suggesting around 5% survival, the log-normal, log-logistic, Gompertz and generalised gamma, and those that suggest minimal survival, the exponential, Weibull and gamma.

Table 21: Company RWEQ OS parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM
AIC							
BIC							
Sum							

The company noted that both the log-logistic and the log-normal has good information criteria, but that both tended to underestimate overall survival when compared to the

company clinical expert responses. The company selected the log-logistic due to its information criteria and it predicting marginally higher survival than the log-normal. The mean survival estimates of the company experts were more than double that of the log-logistic curve at 5 years, and roughly treble those of the log-logistic curve at 10, 15 and 20 years, the individual responses being the small back dots and their average the larger diamonds. It is unclear why no expert responses were elicited for 3 years for RWEQ, particularly given its shorter anticipated OS and PFS compared to dostarlimab.

12

4.2.6.4 PFS curve: comparator RWEQ

Due to the RWEQ data not recording progression the company uses time to next treatment (TTNT) as a proxy. The company notes that this may bias the analysis against dostarlimab because it is likely that progression will occur before TTNT.

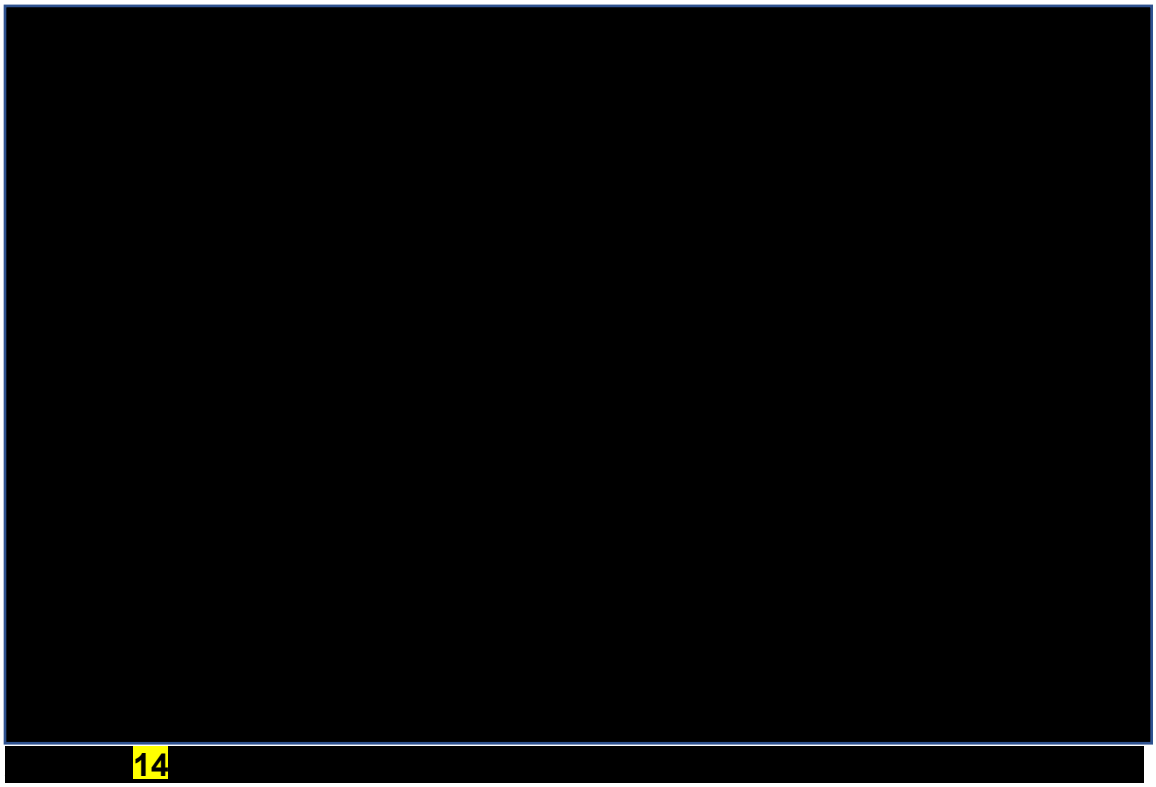
It may be questionable whether any of the TTNT parameterized curves fits the RWEQ KM data particularly well. The curves all tend to lie above the KM S(t) curve from month 9 to 24 and then tend to fall below it.

Table 22: Company RWEQ TTNT parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM
AIC							
BIC							
Sum							

The company notes that the RWEQ TTNT extrapolation is less sensitive to the choice of curve. Based upon the information criteria the company selected the log-logistic, this also estimating slightly higher percentages than the other curves. But similar to the RWEQ OS curve, the company noted that the mean survival estimates of the company

experts were roughly treble that of the log-logistic curve at 5 years, and more than treble those of the log-logistic curve at 10, 15 and 20 years.



14

4.2.6.5 OS curve: dostarlimab

The company fits the same set of parameterized curves to the GARNET OS KM data as it does the TTD data.

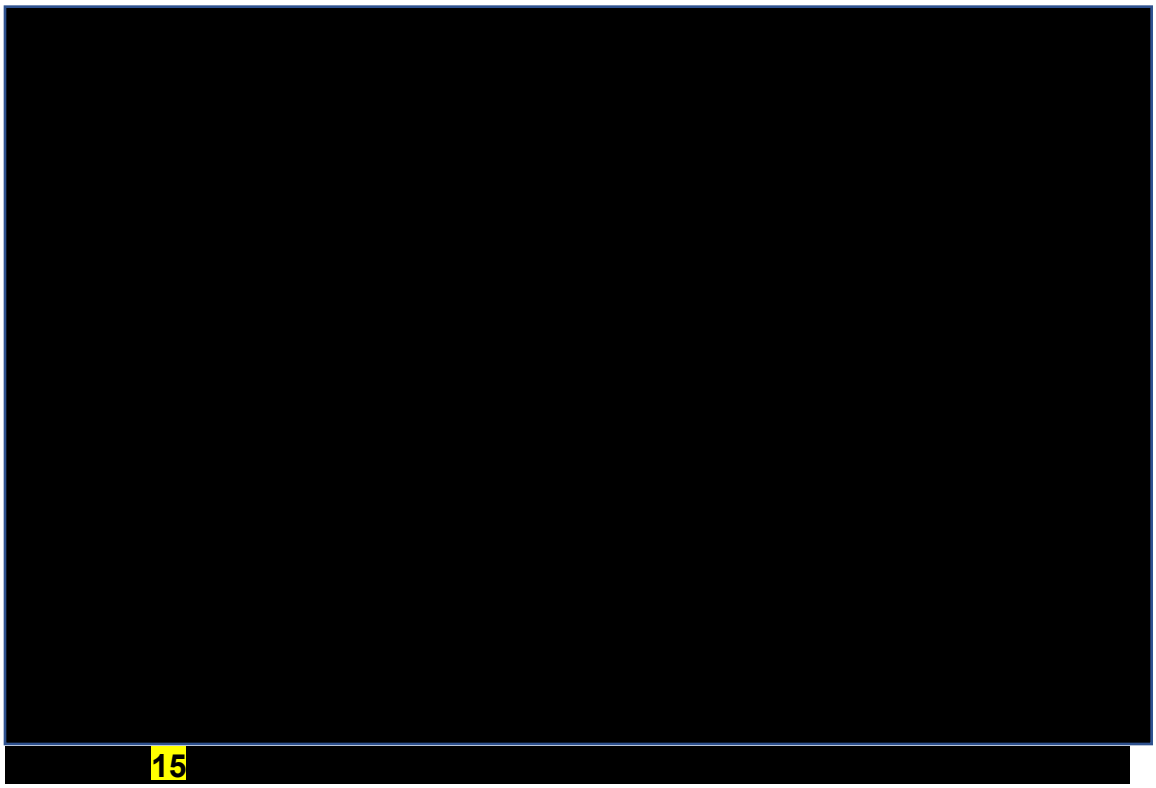


Table 23: Company GARNET OS parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM	SPL1	SPL2
AIC									
BIC									
Sum									

16

The unadjusted curves diverge markedly after two years. The company experts' estimates of the probable survival at 3, 5, 10, 15 and 20 years also show a large spread. The company base case adjusted the OS curves for treatment waning between year [redacted] and year [redacted], due to the treatment cessation assumption at year [redacted].

17

Despite the log-normal having a similar AIC and a superior BIC to the generalised gamma, the company selected the generalised gamma due to its waned curve conforming more closely to the means prediction of the company experts. The log-normal curve was deemed to provide too low an estimate of overall survival for dostarlimab.

4.2.6.6 PFS curve: dostarlimab

The parameterized curves fitted to the GARNET PFS KM data is shown below.

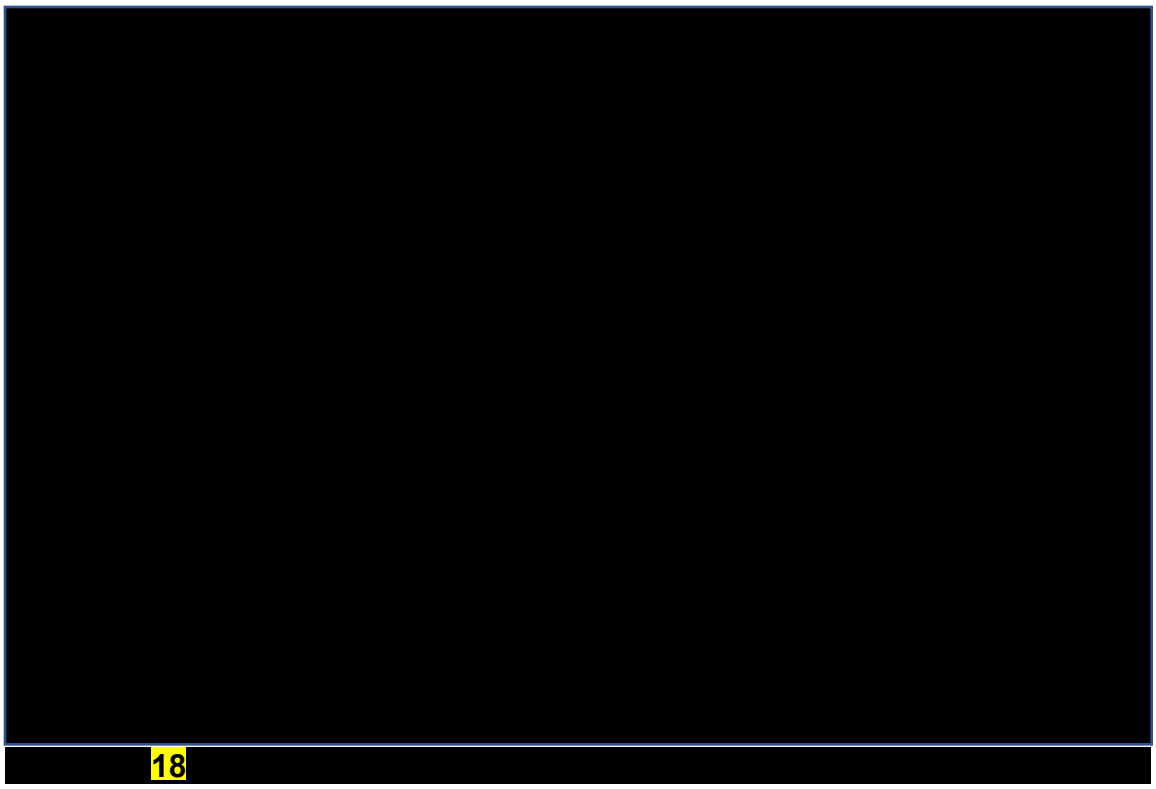
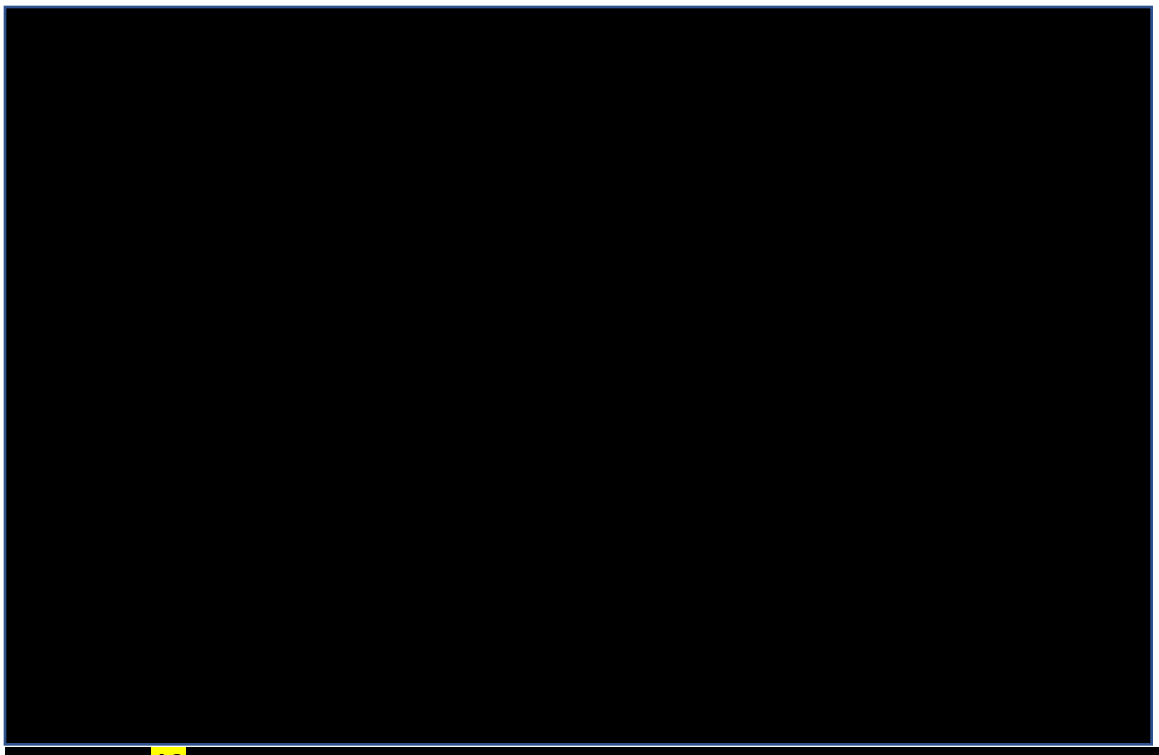


Table 24: Company GARNET PFS parameterised curves information criteria

	EXPO	WEIB	GOMP	LOGN	LOGL	GAMM	GGAM	SPL1	SPL2
AIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sum	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]



[redacted]19

Adjusting the PFS curves for treatment waning between years [redacted] and [redacted] due to treatment cessation at year [redacted] has less effect upon the dostarlimab PFS curves.

The company identified the generalised gamma and the Gompertz as having the best fit to the Kaplan Meier data, based partly on expert opinion. But given the dostarlimab OS curve the company selected the log-normal PFS curve as a more conservative and better aligned PFS curve.

4.2.6.7 Modelled curves

The final set of curves that are applied in the company base case are presented in Figure 21 and Figure 22.

21

Even given the major adjustment to the dostarlimab OS curve there is still a considerable divergence between the OS curve and the PFS curve. The modelling consequently estimates that in the dostarlimab arm a considerable amount of overall survival is spent in the PPS health state after progression has occurred. A similar picture emerges in terms of the PFS curve and the TTD curve, the modelling estimating that much of the time spent in PFS occurs after cessation of treatment.

For the RWEQ comparator the OS and PFS curves are much more closely aligned and relatively little of overall survival is spent in the PPS health state after progression has occurred.

4.2.7 Health related quality of life

The company analyses the GARNET EQ-5D-5L data of the [redacted] patients reporting their baseline EQ-5D and at least one subsequent EQ-5D. The small number of patients reporting EQ-5D appears was due to EQ-5D data only being collected from study protocol 3. The EQ-5D-5L data was cross walked to EQ-5D-3L using the standard algorithm and evaluated using the UK social tariff. A range of models were explored. Generalised estimating equation (GEE) was used with patient identifiers to identify repeat sampling from individuals. The final company model included baseline quality of life, post progression survival (PPS) and being within 15 weeks of death, equivalent to 5 three week model cycles. An alternative model, Model 2, excluding the time to death variable was included as a scenario analysis.

Table 25: Company quality of life models

	Model 1	Model 2
Constant	[redacted]	[redacted]
Baseline	[redacted]	[redacted]
PPS	[redacted]	[redacted]
< 15 weeks to death	[redacted]	[redacted]

Given the mean baseline quality of life of [redacted] this resulted in the following quality of life values.

Table 26: Company quality of life values

	Model 1	Model 2
PFS	[redacted]	[redacted]
PFS and <15 weeks to death	[redacted]	[redacted]
PPS	[redacted]	[redacted]
PPS and < 15 weeks to death	[redacted]	[redacted]

A multiplicative age adjustment to these quality of life values was applied using the standard reference.

Note that the submission values were originally based upon [redacted] patients. The company has since updated this to the [redacted] patients.

4.2.8 Resources and costs

4.2.8.1 Dostarlimab drug and administration costs

Dostarlimab is initially administered every 3 weeks, but from the 5th administration the dose and treatment interval are both doubled. The original company submission included a [redacted] PAS. This has been increased to [redacted] at technical engagement. All costs and ICERs within this document reflect the increased [redacted] PAS. This results in the following costs by model cycle. The simple IV 1st infusion reference cost of £241 is applied to the first cycle, with subsequent administrations being costed using the £332 reference cost for subsequent administrations. This results in the following drug and administration costs by 3 week model cycle.

Table 27: Dostarlimab drug and administration costs per model cycle

	Cost	PAS	PAS inc.	Size mg	
Dostarlimab	£5,887	[redacted]	[redacted]	500	
	Dose mg	Days	Per cycle	Admin	Total
Dose Cycle 1	500	21	500	£241	[redacted]
Dose Cycles 2-4	500	21	500	£332	
Dose Cycles 5+	1000	42	500	£166	

4.2.8.2 RWEQ drug and administration costs

The company costs the individual treatments that comprise more than 5% of the RWEQ basket using the CMU EMIT database, and where this lacks entries the BNF. Combination therapies incur the Complex IV 1st administration £307 NHS reference cost. The company further assumes that 20% of patients will receive hormone therapy. A weighted average of the resulting costs is applied to the RWEQ TTD curve of the model. This results in the following costs per 3 week model cycle.

Table 28: RWEQ drug and administration costs per model cycle

	CARP	CPLD	PLDM	PACM	CARM	HORM	Average
Weight	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Drug costs	£37	£1,084	£1,069	£37	£15	£21	[redacted]
1st admin	£307	£181	£181	£723	£181	£0	
Subs admin	£496	£249	£249	£996	£249	£0	
CARP: carboplatin + paclitaxel, CPLD: carboplatin + PLD, PLDM: PLD monotherapy, PACM, paclitaxel monotherapy, CARM: carboplatin monotherapy, HORM: hormone therapy							

4.2.8.3 Ongoing monitoring costs

Ongoing monitoring costs are based upon expert opinion and costed using the usual NHS reference costs and PSSRU costs. Resource use, unit costs and total costs by health state are as below.

Table 29: Ongoing monitoring resource use and costs

	PFS On Tx	PFS Off Tx	PPS	Cost
OP Consultant Follow-Up visit	1.0	0.3	0.3	£176
Blood test	1.0	0.3	0.3	£3
CT scan	0.3	0.3	0.3	£97
Specialist Nurse	1.0	1.0	1.0	£50
GP visit	1.0	1.0	1.0	£39
GP Nurse visit	0.3	0.3	0.3	£48
Cost per 3 week cycle	£312	£186	£186	

4.2.8.4 Subsequent treatment costs

GARNET suggests that of those who have ceased dostarlimab treatment [redacted] received a subsequent treatment. The distribution of these between chemotherapy treatments is assumed to be as per the RWEQ arm; i.e. the RWEQ 2nd line treatment. The company further assumes that 10% will be radiotherapy and 5% hormone therapy, reducing the proportions of chemotherapy treatments to proportionately so that the treatment distribution sums to 100%; i.e. those who receive a subsequent treatment receive 1 subsequent treatment.

The RWEQ data suggests that after their 2nd line treatment [redacted] of patients received a subsequent treatment. But it appears that the RWEQ may not have collected radiotherapy data or hormone therapy data. The company adds an absolute 10% radiotherapy and 5% hormone therapy, resulting in a proportion receiving 3rd line treatment in the RWEQ arm of [redacted]. The distribution between the chemotherapy treatments is that of the 3rd line RWEQ data.

No administration costs are applied.

The duration of subsequent treatments is largely taken from the RWEQ data set, being [redacted] model cycles for 2nd line and [redacted] model cycles for 3rd line. The durations of radiotherapy and hormone therapy are taken from the literature.

The total cost is applied to the proportion falling out of PFS each cycle.

Table 30: Subsequent treatment costs

				Model cycles	
	DOST	RWEQ	Drug	2nd line	3rd line
Paclitaxel monotherapy	[redacted]	[redacted]	£37	[redacted]	[redacted]
Carboplatin monotherapy	[redacted]	[redacted]	£15	[redacted]	[redacted]
PLD monotherapy	[redacted]	[redacted]	£1,069	[redacted]	[redacted]
Carboplatin + PLD	[redacted]	[redacted]	£1,084	[redacted]	[redacted]
Carboplatin + paclitaxel	[redacted]	[redacted]	£37	[redacted]	[redacted]
Carboplatin + gemcitabine	[redacted]	[redacted]	£66	[redacted]	[redacted]
Radiotherapy	[redacted]	[redacted]	£2,723	8.7	8.7
Hormone therapy	[redacted]	[redacted]	£21	4.6	4.6
Total Cost	£3,011	£2,883			

4.2.9 Adverse events

While clinically important, adverse events have relatively little effect upon the model outcomes and so the ERG does not present the detail of their cost and QALY calculations. In brief, for dostarlimab adverse event rates are taken from GARNET. For the comparator arm the rates of adverse events for the individual treatments are taken from papers in the literature. These are then combined into a weighted average for RWEQ. Each adverse event is typically associated with a relevant inpatient NHS reference cost while the QALY impacts are typically taken from a range of previous NICE assessments.

Table 31: Adverse events: Costs and QALYs

	DOST	RWEQ	Cost	QALY
Abdominal pain	[redacted]		£375.46	-0.069
Allergic reactions		3%	£404.26	-0.116
Fatigue		4%	£0.00	-0.073
Anaemia	[redacted]	4%	£485.28	-0.119
Neutropenia		25%	£431.19	-0.090

Thrombocytopenia		5%	£655.62	-0.090
Nausea		1%	£447.58	-0.045
Vomiting		1%	£447.58	-0.103
Leukopenia		1%	£431.19	-0.090
Sensory neuropathy		2%	£351.03	-0.116
Hand and foot syndrome		3%	£404.26	-0.116
Mucosal inflammation		1%	£391.93	-0.151
Stomatitis		1%	£391.93	-0.151
Dostarlimab total			[redacted]	-0.021
RWEQ total			£214.93	-0.049

4.2.10 Other comparators

Given the extent of the submission and the focus on the company base case, the ERG has had only limited time to review the company modelling for the comparisons with the individual treatments. For each comparator it appears that this applies:

- The relevant OS hazard ratio to the unadjusted dostarlimab OS curve.
- The relevant PFS hazard ratio to the unadjusted dostarlimab PFS curve.
- The relevant PFS hazard ratio to the unadjusted dostarlimab TTD curve, but caps treatment at a maximum of 6 model cycles.
- The relevant direct drug costs and administration cost.

The company also performs similar scenario analyses using the company hazard ratios that it derives the RWEQ compared to dostarlimab.

4.3 *ERG critique of the company economics*

4.3.1 Model validation

The ERG has rebuilt the company model using the company assumptions and gets good agreement with the company model.

Table 32: Company model vs ERG model rebuild

	Company model			ERG model rebuild		
	RWEQ	DOST	net	RWEQ	DOST	net
QALYs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Costs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ICER			£37,311			£37,075

The ERG rebuild has identified one major error and a number of more minor errors in the company model structure.

- The major error is the calculation of treatment waning and the equalizing of dostarlimab effectiveness with the comparator RWEQ effectiveness as reviewed in greater detail in section 4.3.1.1 below. Correcting this error worsens the company base case ICER from £37,311 per QALY to £46,314 per QALY.

- There is an error in the calculation of the dostarlimab [redacted] cessation percentage. Correcting this error worsens the company base case ICER from £37,311 per QALY to £38,126 per QALY.

- The model assumes 3 weekly dosing of dostarlimab when from the 5th administration it is 6 weekly. Correcting this error worsens the company base case ICER from £37,311 per QALY to £38,098 per QALY.

- The company model assumes that dostarlimab patients who receive a subsequent treatment receive only 1 subsequent treatment while the GARNET trial data suggests more than 1 subsequent treatment. Correcting this error worsens the company base case ICER from £37,311 per QALY to £37,821 per QALY.

- For the scenario that includes a screening cost there is an error in the number needed to screen. This does not affect the company base case.

- While not a modelling error the company excludes doxorubicin + cisplatin from the RWEQ costing on the basis of it comprising less than 5% of those treated, but at [redacted] ([redacted]) this is peculiar and the ERG thinks it an error of judgement, in particular because it means that the company has not presented the effectiveness estimates for doxorubicin + ciplatin. But including doxorubicin + cisplatin has minimal effect upon the company base case, worsening the company base case ICER from £37,311 per QALY to £37,411 per QALY.

- While not a modelling error, at clarification the company noted that the submission quality of life values had been based on an [redacted] subset of the GARNET trial and not the [redacted] subset of the GARNET trial. Correcting this has little effect, worsening the company base case ICER from £37,311 per QALY to £37,428 per QALY.

The corrections worsen the company base case ICER from £37,428 per QALY to £49,190 per QALY. Sections 5.1 and 5.2, from page 146, reports the detail of the results for the company submission base case of £37,311 per QALY. But the intervening sections work with the £49,190 per QALY ICER, which the ERG will refer to as the ERG corrected company base case. The ERG thinks that the ERG corrected company base case is the more relevant figure to work with.

4.3.1.1 Treatment waning and equalisation of hazards with RWEQ

The company submission states that "*Treatment waning was assumed to end at* [redacted] *years, at which point, the efficacy associated with dostarlimab was assumed to be equal to the efficacy associated with current clinical management*". The company model applies the MAIC adjusted RWEQ hazard ratios to the dostarlimab curve hazards. It does not apply the RWEQ hazards from [redacted]. The OS hazards of the company base case are shown below.

23

The MAIC OS HR of [redacted] is not applied in full at the start of [redacted]. Rather it is increased linearly from 0.000 at the start of [redacted] to 2.857 at the end of [redacted], and thereafter remains at [redacted]. The effect of this is seen in the upward tick in the OS hazard for dostarlimab during [redacted]. But the OS hazard for dostarlimab remains below that of RWEQ. The company model does not equalize the hazards between the arms from [redacted] onwards. Hazards are only equalized between the arms from around 20 years due to general population mortality rising above the company modelled hazards. The company base case assumes that a treatment effect from dostarlimab will be retained for around [redacted] after the vast majority have ceased treatment and [redacted] after all have ceased treatment.

There is much better correspondence between the two arms during the waning of PFS when the HR of [redacted] is applied. The odd behaviour of the hazards towards the end of the time horizon is due to the PFS ≤ OS constraint, hence the OS hazards being applied. Very few patients remain alive at this point.

[24]

Given the company intention to equalize hazards between the arms from year [redacted] onwards, when equalizing hazards the ERG will equalize the dostarlimab hazard with the RWEQ hazard. During any period of waning, within the dostarlimab arm the ERG will take a weighted average of the dostarlimab hazard and the RWEQ hazard. If the number of cycles during the adjustment period is N the weight for the RWEQ hazard for the n^{th} cycle of this adjustment period will be n/N.

4.3.1.2 GARNET subsequent treatments

When costing subsequent treatments the company notes that [redacted] of those who ceased dostarlimab received a subsequent treatment. Among these [redacted] patients the average number of subsequent treatments, including [redacted] radiotherapy treatment and [redacted] letrozole treatments, was [redacted]. The company assumes that [redacted] will receive radiotherapy, [redacted] hormone therapy and the remainder the balance of the RWEQ 2L chemotherapies. But this yields an average number of subsequent treatments of [redacted] rather than the [redacted] of GARNET.

For the RWEQ data the proportion receiving a subsequent chemotherapy was [redacted]. It appears that radiotherapy and hormone therapy subsequent treatment data was not available. The company adds 10% radiotherapy and 5% hormone therapy to suggest a retreatment rate of [redacted]. This may be reasonable if the radiotherapy and hormone therapy data was not available within the RWEQ data.

The ERG thinks that it is more reasonable to apply the average number of subsequent treatments for dostarlimab, because this is what generated the clinical effectiveness estimates.

Note that subsequent treatment costs only include the direct drug costs. There are no drug administration costs. Including administration costs in the ERG model rebuild raises costs in both arms, but net costs and the ICER are barely affected by this omission. The ERG does not explore this further.

4.3.1.3 Quality of life values

The company submission notes that in GARNET only N=106 patients in the ITT population had EQ-5D data available due to EQ-5D only being collected following protocol amendment 3. The quality of life values are based upon the subset who have both a baseline and at least 1 post baseline value. The ERG assumes this is the reason for the reduction in the sample size from N=106 to N=[redacted]. The mean baseline quality of life value relates to the N=[redacted] and not the N=[redacted]. The ERG thinks that the company should supply the mean quality of life value for the N=[redacted] as well, as there may be an issue around which is the most appropriate to use for the calculation of the quality of life values within the model. This issue can be resolved at technical engagement by a presentation of both values and their standard errors.

4.3.1.4 Number needed to test

The calculation of the number needed to test (NNT) for testing costs suggests that of the 42% of recurrent patients, all 42% need tested. Given the £210 test cost this results in an average testing cost of £88. But the calculation incorrectly applies the assumed 23% dMMR prevalence, in effect not applying it. Applying this results in an NNT of 186% and an average testing cost of £390. The ERG is also unclear why only the

recurrent need to be tested. If all patients need to be tested the NNT rises to 443% and the average testing cost to £929.

The £210 cost per test is taken from NICE DG42, IHC screening for Lynch syndrome in people with endometrial cancer. Note that DG42 also includes a genetic counselling cost of £563. If this is included and all need to be tested, costs in the dostarlimab arm would increase by £1,268.

But ERG expert opinion notes that if NICE guidance is followed testing will be routine in all centres within the next 12-18 months. The ERG thinks that despite the NICE scope the company is correct not to include the costs of the tests.

4.3.2 Correspondence between model inputs and cited sources

4.3.2.1 TA571 treatment discontinuations

The company states that similar discontinuation assumptions were made during the STA of Avelumab for treating metastatic Merkel cell carcinoma (TA517). This is a slightly partial account.

- The company submission for TA517 assumed that 1/3 of patients projected to remain on treatment at 2 years by the log-logistic curve would continue treatment beyond it. All patients would stop treatment at 5 years.
- The ERG preferred to apply the Weibull with no treatment cessation rules as it seemed unethical to cease treatment for those continuing to benefit from it.
- The NHS England submission noted that "*several other PD-L1 drugs have 2 year maximum treatment durations in use, particularly in lung cancer. In those diseases in which PD-L1 drugs have been used for the longest, there is an increasing perception amongst clinicians that very long treatment durations may not be necessary and may cause harm in view of the uncommon but potentially very serious immune-related toxicities that are being encountered with prolonged treatment durations.*"
- The FAD concluded that "*The committee agreed that the company's assumptions appeared to reflect clinical practice with regard to stopping treatment. However, it*

concluded that it would consider both the company's and the ERG's assumptions in its decision-making."

4.3.3 ERG critique: Main Issues

4.3.3.1 Uncertainty around long term clinical effect

The limited duration of follow-up during GARNET and the structural uncertainties around treatment cessation and duration of benefit mean there is considerable uncertainty about the reliability of the long-term modelling. This is reflected in the ERG corrected company base case ICER sensitivity to the time horizon that is applied. Quite a long extrapolation is required for the ICER to approach the NICE upper End of Life willingness to pay (WTP) threshold of £50k/QALY.

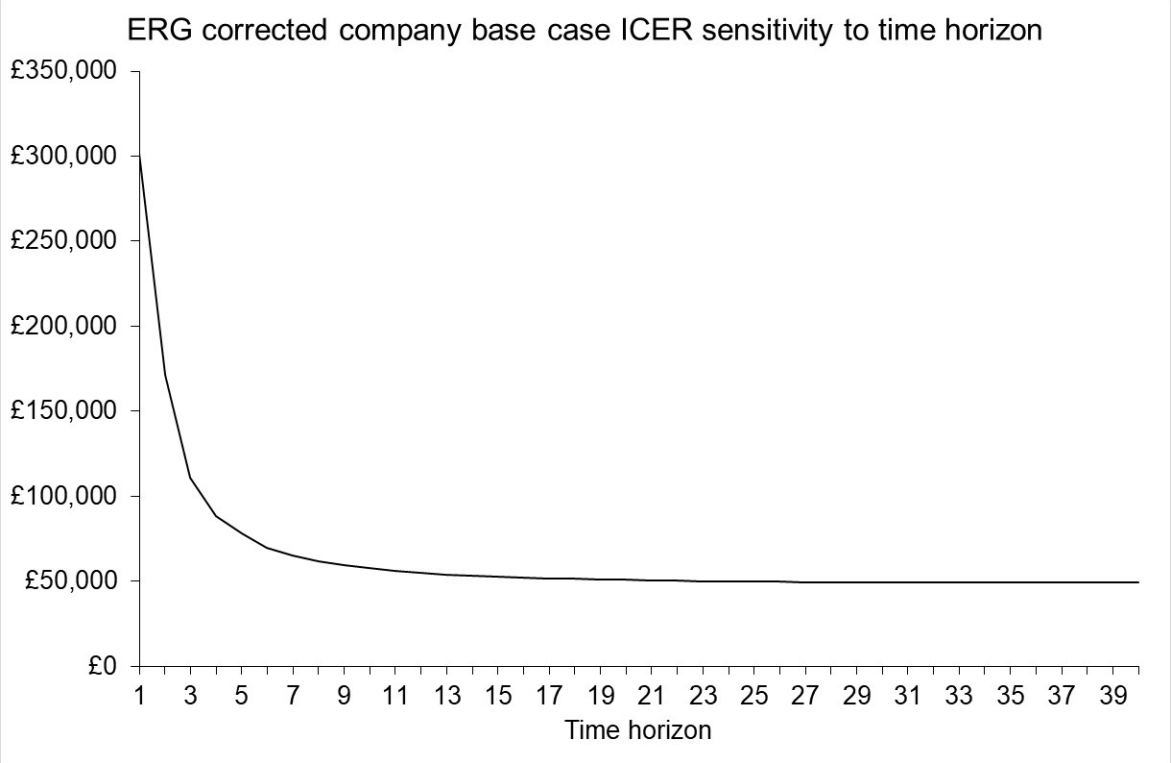


Figure 25: ERG corrected company base case: ICER sensitivity to time horizon

4.3.3.2 OS and PFS extrapolation: Elicitation

The company's seven experts were shown the GARNET Kaplan Meier S(t) curves and the 6 monthly numbers remaining at risk for the ITT and the evaluable efficacy populations, and the equivalent of this for the RWEQ data set. They were asked to complete the following table.

Table 33: Company expert elicitation: OS projections

	6mth	12mth	18mth	24mth	3yr	5yr	10yr	15yr	20yr
DOST									
RWEQ									

The experts were then shown the GARNET Kaplan Meier OS S(t) curve with the unadjusted OS parameterised curves fitted to it and extrapolated to 20 years. They were asked to state which of the unadjusted parameterised OS curves best represented the proportion who would remain alive. A similar exercise was then performed for the RWEQ Kaplan Meier OS S(t) curve and OS parameterised curves.

A parallel exercise was then undertaken for PFS, with the experts being asked to complete the following table and then decide on which of the unadjusted parameterised curves was the most reasonable extrapolation.

Table 34: Company expert elicitation: PFS projections

	6mth	12mth	18mth	24mth	3yr	5yr	10yr	15yr	20yr
DOST									
RWEQ									

The key point is that the experts were never asked about the parameterised curves adjusted for dostarlimab treatment stopping rules. The OS and PFS elicitation exercises were conducted prior to the discussions around treatment stopping rules and treatment waning. It can be argued that the experts might have this in the back of their mind in any case, but the presentation of the unadjusted curves during the elicitation exercise suggests the opposite was anticipated by the company.

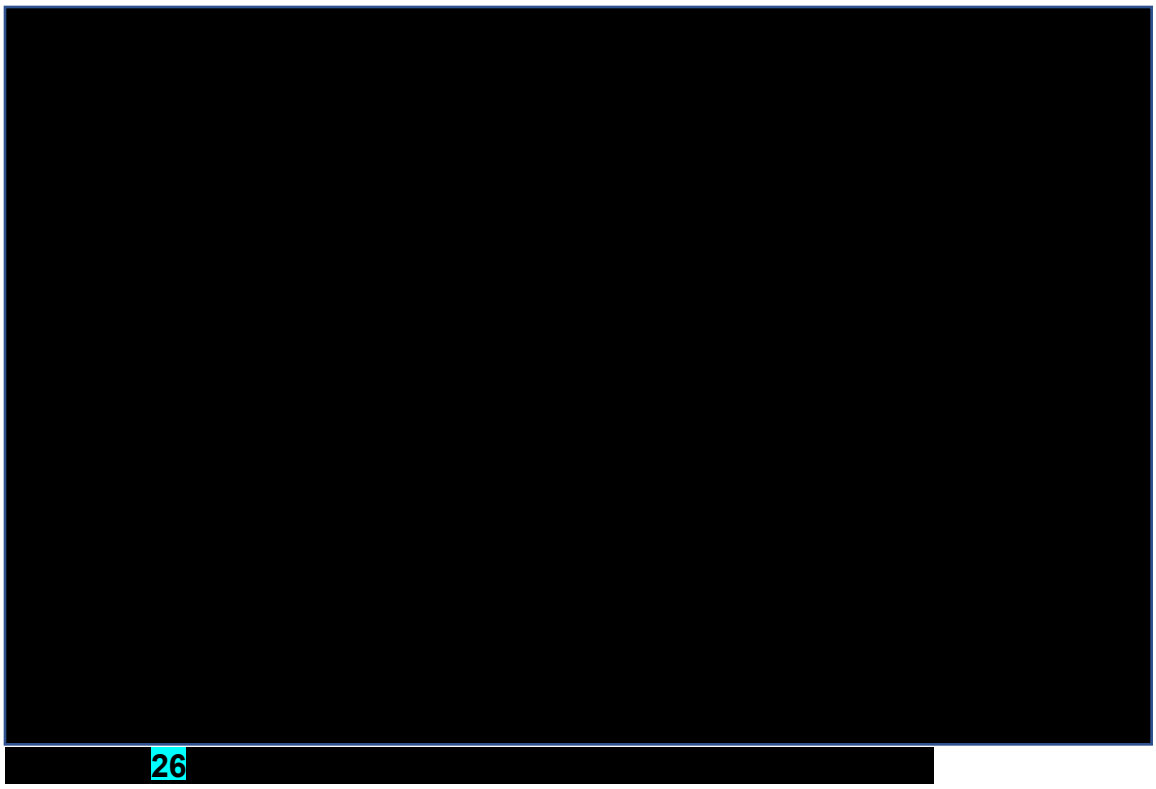
The ERG thinks that the most reasonable interpretation of the company expert estimates for OS and PFS relate to the GARNET data and to the unadjusted curves.

The ERG thinks that it is unreasonable for the company to have presented these results within its submission results overlaid on the dostarlimab adjusted curves. The ERG thinks that the company expert responses will be biased and too high for an assessment of the reasonableness of the adjusted curves.

Given the issues highlighted above, the ERG undertook further in-depth critique of the company's approaches to modelling OS and PFS extrapolation, and selected Weibull as the preferred parametric model for OS (see Appendices 9.3 and 9.4).

4.3.3.3 Treatment discontinuation and waning: Elicitation

Subsequent to the OS and PFS elicitation, the company experts were shown a graph similar to Figure 26 below. The ERG has superimposed the TTD Kaplan Meier S(t) curve and the Kaplan Meier % N at risk curve out to two years, though note that as presented in Figure 8 on page 101 above the GARNET TTD KM data extends beyond this. The ERG Kaplan Meier % N at risk are typically higher than those of the company. It is difficult to know quite what data points the company presentation relates to as some span periods up to 6 weeks, but even given this the ERG cannot align its N at risk with that of the company presentation. This could be due to ERG error, company error or the company presentation may be using an earlier data cut which would result in earlier censoring due to data cut off and hence lower numbers remaining at risk than applies in the IA2 data cut used by the ERG. The reason for the increase from [redacted] for 31w-36wpatients to [redacted] patients for 37w-42w in the company presentation is apparently due to the inclusion of a patient who received a delayed dose.



26

The experts were also shown the following tabulated values.

Table 35: Company TTD elicitation table

	0mth	3mth	6mth	9mth	12mth	>12mth	2yr
%							
N							

In effect, the experts were shown the blue bars of Figure 26 with labels showing the proportion and number of patients remaining on treatment, and the values of Table 35.

Note that the experts were not asked to complete the final table entry as in the OS and PFS elicitation exercises, but were rather presented with it prefilled at . The ERG thinks that it would have been better to have asked the experts to complete this themselves much as with the OS and PFS elicitation.

The key point is that the company seems to have presented the number remaining at risk and not the Kaplan Meier TTD S(t) curve. If so the company presentation assumes

that censoring due to data cut off is a discontinuation event. This would be incorrect and would seriously bias the presentation.

It is also notable that the company only presents the KM numbers remaining at risk to "[redacted]" weeks and "[redacted]" with the value for this "timepoint" being [redacted]. The [redacted] remaining at risk applies to weeks [redacted], somewhat closer to week 54 than the uninformed observer might be expected given the company presentation. But noting this might have resulted in an infeasibly low proportion being estimated to remain on treatment at the [redacted] point, a reflection of the number at risk falling off due to data cut off despite the Kaplan Meier S(t) curve being maintained.

The Kaplan Meier TTD data extends some time beyond this but the longer presentation would have shown a further decline in the KM numbers at risk due to the data cut-off, as per the ERG superimposed curves of Figure 26.

The ERG thinks that to elicit the desired result the company has presented an invalid data set that appears to show a smooth steady fall in the number of patients remaining on dostarlimab, and hence the reasonableness of assuming that this smooth steady discontinuation rate will broadly continue to yield around [redacted]. The experts are not presented with the resulting modelled curve which applies the [redacted] assumptions to the curve fitted to the Kaplan Meier S(t) data, resulting in the cliff edge discontinuation at [redacted]. The modelled TTD curve bears no resemblance to the numbers remaining on treatment that were presented to the experts.

The ERG thinks that the company should have presented the Kaplan Meier S(t) curve estimates for TTD. The ERG thinks that presenting the Kaplan Meier number remaining at risk and so in effect treating data cut off as a discontinuation event renders the company TTD elicitation exercise largely meaningless. At best it would seem to put a lower floor on what proportion might remain on treatment but the values cannot be used as central estimates.

Of the questions:

- [redacted] As this was a Yes/No question it appears there was no way for the experts to dissent by suggesting a different percentage.
- [redacted] As this was a Yes/No question it appears there was no way for the experts to dissent by suggesting another timepoint. The ERG also notes that there is no stopping rule in the SmPC. It is unclear whether the company is suggesting that if NICE approves dostarlimab that a stopping rule at [redacted] should be a part of the recommendation and funding. ERG expert opinion suggests that there will only be a cliff edge if funding is withdrawn at this point.
- [redacted] This presumes that a stopping rule at [redacted] will be introduced. The restriction of the responses to be no more than [redacted] is also a concern, particularly in the light of two respondents choosing this value and possibly being constrained by it causing the [redacted] to be biased and too low.

The presentation also suggests that the experts were asked:

[redacted] The ERG has not been able to find any responses to these questions. The second question is slightly loosely worded in that it does not specify that this should be among those who have discontinued treatment. It also has surprisingly

long durations as options, with there being no means for the experts to be any more explicit about short durations such as “[redacted]” other than by stating “[redacted]”.

The company did not ask the seven experts about complete cessation of dostarlimab. It appears that this was only asked of two of the seven experts during follow-up one-to one interviews. It is unclear whether either of these experts were either of the two of seven experts who tended to disagree with the pre-specified responses of the main elicitation exercise. The two experts who were consulted apparently noted that [redacted]The SmPC states “Treatment can be continued as long as Jemperli continues to work. The doctor may interrupt Jemperli treatment or stop it altogether if certain side effects occur”. This appears to put the emphasis on reacting to the occurrence of side effect, rather than pre-emptively withdrawing treatment.

ERG expert opinion thinks that the cliff edge discontinuation of the company base case is only likely to apply if funding is withdrawn after [redacted] of treatment. Both experts note the possibility of a range of adverse events. Patients remaining progression free and doing well while receiving dostarlimab may not want to have it withdrawn from them. One ERG expert notes that patients find repeated ongoing treatment a burden which could be a contributory factor to treatment cessation in addition to the side effects mentioned in the SmPC.

There is some disagreement between the ERG experts as to when patients remaining progression free while on dostarlimab might start to have treatment withdrawn. One suggests that toxicity could see some withdrawing from treatment as early as [redacted], though the ERG thinks that withdrawals while progression free that are related to toxicity might already be reflected in the GARNET TTD data. This expert suggests that most patients would have withdrawn from treatment at [redacted] but that some would continue beyond this point, while the other expert suggests that rather more patients

remaining progression free could continue dostarlimab treatment beyond [redacted]. Similarly, one ERG expert thinks a [redacted] total cessation point is reasonable, while the other queries why patients who are progression free would cease treatment even at the [redacted] point.

The ERG questions why the TTD and stopping rules elicitation exercise was conducted after the OS and PFS elicitation exercise. Unbiased OS and PFS estimates adjusted for the TTD and treatment stopping rules obviously require prior consideration of the TTD and stopping rules.

The ERG thinks that the company TTD and stopping rules elicitation exercise was poorly constructed, that its results are likely to be biased and that at best it provides a floor to the cessation percentage. Given this the ERG will apply a percentage of [redacted] and will explore [redacted] and [redacted]. In the light of the SmPC the ERG will constrain the proportion on treatment after the first cessation point to be the lesser of the TTD curve and the PFS curve.

The company does not appear to have presented data to support its assumption that those ceasing dostarlimab would continue to receive the full benefits of treatment for [redacted] after stopping treatment. There may be retention of benefits but the ERG thinks it unlikely that no patient would have any loss of effect for [redacted] after treatment cessation.

The ERG thinks that the more natural assumption is that for some patients some loss of effect, albeit small, would start from treatment cessation. As a consequence, the ERG base case will assume that treatment waning occurs from the point of treatment cessation. This does not assume that patients revert to the RWEQ risks immediately upon treatment cessation, only that they move towards these risks from treatment cessation. A treatment benefit is still assumed for [redacted] after treatment cessation, which in itself may be optimistic.

4.3.3.4 Treatment discontinuation and waning: Scenarios

Given the questionable reliability of the company TTD elicitation exercise the ERG presents various scenarios to illustrate the effect that altering these assumptions has upon the ERG corrected company base case. For all scenario all dostarlimab patients are assumed to cease treatment at the start of year 5

Table 36: Corrected Company ICER: Sensitivity to discontinuation assumptions

First discontinuation rules		Waning of effect period		
Year	% remaining	Start Year	End Year	ICER
[redacted]	[redacted]	[redacted]	[redacted]	£49,190
[redacted]	[redacted]	[redacted]	[redacted]	£55,354
[redacted]	[redacted]	[redacted]	[redacted]	£51,900
[redacted]	[redacted]	[redacted]	[redacted]	£47,223
[redacted]	[redacted]	[redacted]	[redacted]	£53,590
[redacted]	[redacted]	[redacted]	[redacted]	£60,362
[redacted]	[redacted]	[redacted]	[redacted]	£56,568
[redacted]	[redacted]	[redacted]	[redacted]	£51,429
[redacted]	[redacted]	[redacted]	[redacted]	£57,990
[redacted]	[redacted]	[redacted]	[redacted]	£65,369
[redacted]	[redacted]	[redacted]	[redacted]	£61,235
[redacted]	[redacted]	[redacted]	[redacted]	£55,635
[redacted]	[redacted]	[redacted]	[redacted]	£56,315
[redacted]	[redacted]	[redacted]	[redacted]	£54,034
[redacted]	[redacted]	[redacted]	[redacted]	£51,894
[redacted]	[redacted]	[redacted]	[redacted]	£59,563
[redacted]	[redacted]	[redacted]	[redacted]	£57,139
[redacted]	[redacted]	[redacted]	[redacted]	£54,864
[redacted]	[redacted]	[redacted]		

Assuming that waning starts immediately upon treatment cessation or [redacted] after treatment cessation worsens the ICER by a reasonable amount. Note that these scenarios still retain a waning dostarlimab treatment effect out to [redacted], [redacted] after treatment has ceased for most patients.

Results are particularly sensitive to moving the timepoint of the first main discontinuation from [redacted] years to [redacted] years,

There needs to be detailed consideration of the cliff edge that is assumed for discontinuations.

4.3.3.5 Censoring by arm and informative censoring

The OS KM S(t) and N at risk as a proportion of baseline N can be presented for GARNET ITT and the RWEQ population.

27

There is much higher censoring in the GARNET data than the RWEQ data. If the above pattern of censoring was observed in a two arm trial it would raise major concerns. There is also a large amount of early censoring in the GARNET data, which is a concern.

It is possible that those who performed badly during GARNET were more likely to drop out of the trial and be censored while those with a better performance were more likely to continue with treatment and remain in the trial.

At clarification the ERG asked for GARNET KM data restricted to those with a CR or PR response. The company declined to supply this on the grounds that as this would only apply to [redacted] patients the reduced sample size means that it would not be appropriate to draw any conclusions from this data.

Fully exploring this would need to take into account censoring due to data cut off and censoring due to other reasons. The ERG clarification KM data request for the ITT, evaluable efficacy and those with a CR or PR response populations would need to be augmented by splitting the censoring column into censoring due to data cut off and censoring due to other reasons. Patient baseline characteristics split by best response and reason for censoring, data cut or other, would also be required. This is an issue that can be addressed at technical engagement.

4.3.3.6 RWEQ individual treatment effects

The pooled RWEQ GARNET like OS KM curve can be compared with the RWEQ 5 most common treatments' individual KM curves. Note that the company declined to supply the KM curve for cisplatin + doxorubicin mainly due to it falling marginally below the arbitrary 5% of RWEQ patients threshold that the company uses for costing purposes ([redacted]) and not being within the NICE scope. With regards the latter is can be noted that the cisplatin + doxorubicin clinical effectiveness remains within the RWEQ data, that the NICE scope specifies chemotherapy "including" a number of named treatments and that the NICE scope also does not specifically name carboplatin+PLD. Treatments comprising more than 5% of the RWEQ GARNET like ([redacted]) data set accounted for [redacted] of patients.

28

Figure 28 shows the marked differences in overall survival by treatment within the RWEQ data set. The combination therapies had better survival and the monotherapies worse survival.

Table 37: RWEQ baseline characteristics by treatment

	Carb+Pac	Carb+PLD	PLD mono	Pac mono	Carb mono
N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age					
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
< 65 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
65 - 75 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥ 75 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG at registry diagnosis					
Unknown	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Known	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
of which 0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
of which 1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Histology at diagnosis					
Clear cell carc.	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Endometrioid	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mixed carc.	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Non-spec. carc.	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Serous	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Other	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
FIGO at registry diagnosis					
I	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
II	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
III	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
IV	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Grade at diagnosis					
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Not assessable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

There are few marked differences in the patient baseline characteristics presented by the company that could account for these large differences, though the following might be noted:

- Fewer younger patients for PLD monotherapy and carboplatin monotherapy

- Higher unknown ECOG for PLD monotherapy and carboplatin monotherapy

The lower proportion of younger patients under 65 for PLD monotherapy and carboplatin monotherapy mirrors the lower proportion of younger patients in the RWEQ population, [redacted], compared to the GARNET population [redacted]. The GARNET forest plot of Figure 20 (Document B, page 74) showed no difference in ORR between those under 65 and those over 65, but this does not necessarily imply that there was no difference in overall survival.

The high proportion with unknown ECOG status may be of concern, given that GARNET found it to be a statistically significant determinant of the likelihood of response.

ERG expert opinion is that there are likely to be possibly quite large imbalances between the GARNET and RWEQ populations, and that the best means of exploring this might be to consider the endometrioid subgroups of GARNET and RWEQ populations.

At clarification the ERG requested GARNET and RWEQ KM data split by ECOG status and by endometrioid status. The company declined to supply this, though noted that it was exploring the possibility of supplying data according to endometrioid status. This can be resolved during technical engagement.

4.3.3.7 Dostarlimab PFS vs TTD

As outlined in section 4.2.6.7, the company base case PFS and TTD parameterized curves almost coincide for the first two years of the model, the areas under the curves (AUC) being [redacted] months and [redacted] months respectively: a ratio of 1.04.

29

During GARNET a reasonably higher proportion of patients remained on treatment compared to remaining in PFS between months 2 and 8. The ERG will present a scenario that applies the GARNET TTD KM curve for the first 8 months of the model.

4.3.3.8 Choice of TTD curve

The company choice of the log-logistic TTD curve does not appear to be justified on statistical grounds. As outlined in greater detail in section 4.2.6.1 on page 101, the Gompertz TTD curve has lower information criteria. Since the dostarlimab TTD curve is mainly being applied prior to the first cessation point and so during the period for which Kaplan Meier data is available, there is less need to assess the reasonableness of extrapolation. The information criteria can be used to assess the internal goodness of fit.

Following detailed critique of the company's approaches and alternative options (see ERG report Appendix 9.5), the ERG prefers the company Gompertz over the company log-logistic dostarlimab TTD curve. This worsens the ERG corrected company base

case from £49,190 per QALY to £51,804 per QALY. But for its base case the ERG prefers the ERG ITT generalized gamma which worsens the ERG corrected company base case from £49,190 per QALY to £52,548 per QALY.

4.3.4 ERG critique: Other issues

4.3.4.1 Quality of life model

The company supplies a range of additional quality of life regressions that explore varying the number of 3 week cycles from death. The quality of life values for the various health states are the exponential of the sum of the relevant coefficients, the baseline QoL coefficient being qualified by the GARNET baseline quality of life of [redacted].

Cycles to death	0	1	2	3	4
Constant Baseline QoL Progressed Cycles to death	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cycles to death	5	6	7	8	9
Constant Baseline QoL Progressed Cycles to death	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
**significant at 1%, *significant at 5%					

In general, the inclusion of a time to death variable makes the coefficient for progressed disease not statistically significant. The exception to this is the model that examines 4 cycles from death. But it seems likely that there is a high degree of multicollinearity between the two variables. The QIC criteria also do not obviously favour the choice of 5 cycles.

Given the centrality of the PFS and PPS health states to the model the ERG thinks it is peculiar to introduce the time to death variable if this renders the PPS coefficient not statistically significant. The QIC criteria also tend to favour either not including time to

death, or including a lengthy time to death which in effect makes consideration of progression redundant. The company is also concerned about the number of observations retained in each analysis, the ERG noting that the analysis with no time to death variable has largest number of observations. But the differences in the main quality of life values are not large.

[redacted]30[redacted]

The ERG also notes that the application of the PFS EoL QoL value and the PD EoL QoL value requires that deaths occurring in the next 5 cycles be modelled as occurring either from PFS or from PD. The model assumption is that these deaths will be proportionate to the number of patients in PFS compared to the number of patients in PD which may not be realistic. The ERG thinks that this is a further argument against including the time to death variable.

The ERG thinks that the natural approach for the base case is not to include the time to death variable. This revision has minimal effect, only slightly worsening the ERG corrected company base case from £49,190 per QALY to £49,513 per QALY.

4.3.4.2 Quality of life values in the literature

The NICE scope does not list any relevant previous STAs. The company SLR identifies 3 studies from the literature, dismissing them due to either small sample size or being based upon expert opinion. The ERG broadly agrees with this but notes that the German study (n=20) reports EQ-5D values of 0.701 for primary disease (N=9) and 0.676 for advanced disease (N=11), though these are valued using a German TTO tariff. These are broadly similar to the company model with no time to death variable estimates of [redacted] for PFS and [redacted] for PD.

4.3.4.3 RWEQ mean number of model treatment cycles

Within its costings of RWEQ as subsequent treatment to dostarlimab the company uses the RWEQ 2nd line data of the RWEQ arm. The company also notes that the median number of model cycles that patients remained on treatment within this was [redacted]. The company base case simulates a mean number of model cycles of [redacted]. This may suggest reducing the modelled RWEQ drug and administration costs to [redacted] cycles; i.e. multiplying by a factor of 84%. This worsens the ERG corrected company base case ICER from £49,190 per QALY to £49,443 per QALY.

4.3.4.4 RWEQ costing

The RWEQ costing calculates an average cost per model cycle based upon the baseline balance of treatments, then applies this to the pooled RWEQ TTD curve. The individual treatment TTD KM curves vary wildly, much as per the individual treatment OS KM curves. It is not obvious whether this is likely to result in much bias, but it can be noted that PLD is one of the more expensive treatments but has very poor KM curves.

4.3.4.5 RWEQ PLD costing

The PLD cost is based upon an average BSA of [redacted] resulting in a dose of [redacted]. This is marginally above the 70mg dose that could be supplied with a 20mg and a 50mg vial at a drug cost of £1,073 rather than £1,425. It might have been more reasonable to

assume a 50:50 split between dosing under and dosing over 70mg, which would slightly lessen the direct drug cost per 4 week treatment cycle from £1,425 to £1,248.

RWE data apparent suggests a [redacted] balance between doxorubicin and PLD, but model costing assumes all PLD. Doxorubicin is somewhat cheaper than PLD.

The above considerations would worsen the ICER. Time constraints mean that the ERG has not explored this. The ERG thinks that the effect would be relatively minor.

4.3.4.6 Ongoing costs

The PFS on treatment, PFS off treatment and PPS health states incur reasonable costs due to ongoing monitoring.

ERG expert opinion suggests that the company estimates for PFS on treatment may be too resource intensive. Given the hospital based monitoring GP and community nurse visits may be less likely.

ERG expert opinion suggests that the company estimates for PFS off treatment may be too resource intensive. Consultant OP visits might be only every 12 weeks. And OP specialist nurse visits might also be less frequent. CT scans might initially be 3 monthly but this would probably extend to 6 months.

In the light of this the ERG will for:

- PFS on treatment, exclude GP and community nurse visits
- PFS off treatment, extend consultant OP visits to 12 weeks and CT scans to 6 monthly

This improves the ERG corrected company base case ICER from £49,190 per QALY to £48,735 per QALY.

The ERG will also present a scenario the extends the PFS off treatment OP specialist nurse visit frequency to 12 weekly.

4.3.4.7 GARNET trial population and test sensitivity and specificity

The company reports that the test is also associated with a sensitivity of 96.2% and a specificity of 88.4%. If these values carry across to the current setting the relatively low specificity may be a concern. Given the assumed prevalence of 23%, it suggests that

among those with a positive test 71% would be true positives and 29% would be false positives. Almost a third of those testing positive and so being treated with dostarlimab may not be dMMR. How this tallies with the GARNET population and what impact this might have upon the real world effectiveness of dostarlimab compared to that in GARNET is difficult for the ERG to speculate about.

5 COST EFFECTIVENESS RESULTS

5.1 *Company's cost effectiveness results*

The original company submission included a dostarlimab PAS of [redacted], and a base case ICER of £50,221 per QALY. During technical engagement the company has increased its PAS to [redacted]. The results of this section revise the company estimates of its original submission by applying the [redacted] PAS rather than the original [redacted] PAS.

The undiscounted life years and discounted QALYs are presented in Table 38.

Table 38: Company base case: Survival and QALYs

	Undiscounted LY			Discounted QALYs		
	RWEQ	DOST	Net	RWEQ	DOST	Net
PFS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PPS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The company base case anticipated that around two thirds of survival in the dostarlimab arm will occur after progression, with around three quarters of the net QALY gain also occurring after progression.

The disaggregate discounted costs are presented in Table 39.

Table 39: Company base case: Disaggregate costs

	RWEQ	DOST	Net
Diagnostic	[redacted]	[redacted]	[redacted]
Drug	[redacted]	[redacted]	[redacted]
AEs	[redacted]	[redacted]	[redacted]
PFS ongoing	[redacted]	[redacted]	[redacted]
Subsequent Treatment	[redacted]	[redacted]	[redacted]
PPS ongoing	[redacted]	[redacted]	[redacted]
End of Life	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]

These results in the cost effectiveness estimate of Table 40.

Table 40: Company base case: Summary

	RWEQ	DOST	Net
LY	[redacted]	[redacted]	[redacted]
QALY	[redacted]	[redacted]	[redacted]
Cost	[redacted]	[redacted]	[redacted]
ICER			£37,311

As noted in the executive summary, at clarification the company supplied slightly revised quality of life values due to basing this on [redacted] rather than [redacted], which very slightly worsens its base case, together with a set of scenario analyses. Due to time constraints the ERG has not updated the company model results for this. This is unlikely to affect Committee deliberations. The ERG revised base case and scenario analyses do reflect the revised quality of life model.

The probabilistic model has a slightly better central estimate of £35,492 per QALY with the associated CEAC being presented in Figure 31.

31

The probabilities of dostarlimab being cost effective at the various NICE willingness to pay thresholds are presented in Table 41.

Table 41: Company base case probabilities of cost effectiveness

	Standard		End of Life	
	Threshold	Probability	Threshold	Probability
Lower threshold Upper threshold				
* Applying the NICE methods 1.7 QALY multiplier				

5.2 *Company's sensitivity analyses*

The company presents a range of univariate sensitivity analyses, the tornado diagram for the 10 most influential variables being presented in Figure 61 on page 197 of Document B of the company submission. The company base case is most sensitive to: baseline utility, the health state utilities and the cost per cycle of dostarlimab. Given the company base case ICER these are all prone to pushing the ICER further above and below £50k per QALY.

The company also presents an extensive range of scenario analyses in Table 85 on page 206 of Document B of the company submission. The ERG does not replicate these in full but highlights a subset of them in Table 42. Unfortunately, time constraints mean that the ERG has not been able to update these company scenario analyses for the revised [redacted] PAS. As a consequence, they relate to the original company PAS of [redacted] and associated base case of [redacted] per QALY.

Table 42: Selection of company scenario analyses

	Δ Cost	Δ QALY	ICER
Company base case (Previous PAS of [redacted])	[redacted]	[redacted]	£50,221
RWEQ based upon EGOG01 ([redacted])	[redacted]	[redacted]	£49,155
RWEQ based upon MAIC OS HR [redacted]	[redacted]	[redacted]	£54,249
RWEQ based upon MAIC OS HR [redacted]	[redacted]	[redacted]	£52,917
DOST OS log-normal, no waning	[redacted]	[redacted]	£50,997
DOST OS generalised gamma, no waning	[redacted]	[redacted]	£33,677
[redacted]	[redacted]	[redacted]	£55,804
[redacted]	[redacted]	[redacted]	£45,439
[redacted]	[redacted]	[redacted]	£53,633
Inclusion of screening test	[redacted]	[redacted]	£50,261
Individual treatment comparator: Doxorubicin monotherapy			
PFS HR [redacted] Makker, OS [redacted] HR Zoptec	[redacted]	[redacted]	£63,144
PFS HR [redacted] Makker, OS [redacted] HR McMeekin	[redacted]	[redacted]	£55,284
PFS HR [redacted] Makker, OS [redacted] HR Makker	[redacted]	[redacted]	£41,337
PFS HR [redacted] Makker, OS [redacted] HR Julius	[redacted]	[redacted]	£40,439
Individual treatment comparator: Paclitaxel monotherapy			
PFS HR [redacted] Makker, OS [redacted] HR McMeekin	[redacted]	[redacted]	£56,911
Individual treatment comparator: Carboplatin + paclitaxel			
PFS HR [redacted] and OS [redacted] Rubenstein	[redacted]	[redacted]	Dom'ted
PFS HR [redacted] and OS [redacted] Mazgani	[redacted]	[redacted]	£106k
Dom'ted: Carboplatin + paclitaxel dominates dostarlimab			

The ERG also highlights that the company restricts it exploration of the alternative functional forms of dostarlimab OS to the log logistic, log normal and generalised gamma.

The ERG has not had time to check whether the no waning scenarios also assume no treatment cessation with patients following the base case TTD curve. The ERG has not had time to check whether the treatment waning that applies to dostarlimab when the

individual comparators are being considered applies the hazards of the individual comparators or retains the hazards of the pooled RWEQ curves. This can be addressed during technical engagement.

5.3 *ERG corrected company base case*

For completeness the ERG presents the ERG corrected company base case results. The undiscounted life years and discounted QALYs are presented in Table 43.

Table 43: ERG corrected company base case: Survival and QALYs

	Undiscounted LY			Discounted QALYs		
	RWEQ	DOST	Net	RWEQ	DOST	Net
PFS	[redacted]					
PPS						
AEs						
Total						

The ERG corrected company base case still anticipates that around two thirds of survival in the dostarlimab arm will occur after progression, with around two thirds of the net QALY gain also occurring after progression.

The disaggregate discounted costs are presented in Table 44.

Table 44: ERG corrected company base case: Disaggregate costs

	RWEQ	DOST	Net
Diagnostic	[redacted]		
Drug			
AEs			
PFS ongoing			
Subsequent Treatment			
PPS ongoing			
End of Life			
Total			

These results in the cost effectiveness estimate of Table 45.

Table 45: ERG corrected company base case: Summary

	RWEQ	DOST	Net
LY	[redacted]		

QALY Cost	[redacted]	[redacted]	[redacted]
ICER			£49,190

The probabilistic model has a slightly better central estimate of £48,764 per QALY with the associated CEAC being presented in

Figure [redacted]32.

[redacted]32

The probabilities of dostarlimab being cost effective at the various NICE willingness to pay thresholds are presented in Table 46.

Table 46: ERG corrected company base case probabilities of cost effectiveness

[redacted]

	Standard		End of Life	
	Threshold	Probability	Threshold	Probability

Lower threshold Upper threshold	[redacted]	[redacted]	[redacted]	[redacted]
* Applying the NICE methods 1.7 QALY multiplier				

5.4 *Company base case: Technical engagement*

As noted in the ERG review of the company TE submission the model submitted by the company at TE was submitted late and did not obviously implement the revised company waning. As a consequence, the ERG cannot replicate the company TE base case. The company TE base case with the ERG waning method applied is stated as resulting in a similar ICER of £49,608 per QALY. This is very similar to the ERG corrected company base case of £49,190 per QALY of Table 45 above, though it should be borne in mind that the ERG corrected company base case applies the original dostarlimab ToT data and not the company TE updated dostarlimab ToT data.

6 EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1 *Exploratory and sensitivity analyses undertaken by the ERG*

6.1.1 ERG preferred assumptions

The ERG prefers the Weibull for dostarlimab OS whereas the company prefers the generalised gamma. The choice of dostarlimab is a major driver of results. Due to the company error the company submission presentation of the OS curves adjusted for waning is also incorrect. The ERG presents these for the Weibull and the generalised gamma over the 40 year time horizon of the model, alongside the GARNET OS KM curve. The waning assumptions are the ERG preferred [redacted] at [redacted] with waning over the next [redacted], after which all cease dostarlimab. The ERG also tabulates the modelled OS percentages for the range of curves available.

The proportions of patients modelled as surviving are presented in Table 47.

Table 47: ERG adjusted dostarlimab modelled OS by curve

Year	GGAM	WEIB	GAMM	EXPO	LOGL	LOGN	GOMP	KM
0.5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
15	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
20	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Given the centrality of the choice of dostarlimab OS curve, the ERG presents a full set of analyses for its preferred base case using the Weibull, and also for the scenario of using the company preferred generalised gamma.

Table 48: ERG preferred model assumptions

Preferred assumption	Section	ICER
Company base-case	5.1	£37,311
ERG corrected company base-case	4.3.1	£49,341
ERG01: Dostarlimab OS Weibull	4.2.6.5 4.3.3.2	£65,454
ERG02: Dostarlimab ERG ITT TTD GGAM	4.3.3.8	£52,709
ERG03: [redacted] dostarlimab continue	4.3.2.1 4.3.3.3	£49,341
ERG04: Waning from treatment cessation	4.3.2.1 4.3.3.3	£55,523
ERG05: Quality of life – no time to death coefficient	4.3.4.1	£49,513
ERG06: Ongoing resource use	4.3.4.6	£48,885
Cumulative effect: ERG02-ERG06	..	£64,006
Cumulative effect: ERG01-ERG06	..	£79,714

6.1.2 ERG preferred base case

The undiscounted life years and discounted QALYs are presented in Table 49.

Table 49: ERG base case: Survival and QALYs

	Undiscounted LY			Discounted QALYs		
	RWEQ	DOST	Net	RWEQ	DOST	Net
PFS PPS AEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The ERG base case anticipates that survival in the dostarlimab arm is split more equally between PFS and PPS, though the majority of the QALY gain is still modelled as occurring after progression has occurred. This arises because the modelled OS curve lies some what above the modelled PFS curve.

The disaggregate discounted costs are presented in Table 50.

Table 50: ERG base case: Disaggregate costs

	RWEQ	DOST	Net
Diagnostic	£0	£0	£0
Drug + admin AEs PFS ongoing Subsequent Treatment PPS ongoing End of Life	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]

These results in the cost effectiveness estimate of Table 51.

Table 51: ERG base case: Summary

	RWEQ	DOST	Net
LY QALY Cost	[redacted]	[redacted]	[redacted]
ICER			£79,714

The probabilistic model has an ICER of £80,640 per QALY with the associated CEAC being presented in Figure 34.

34

The probabilities of dostarlimab being cost effective at the various NICE willingness to pay thresholds is presented in Table 52.

Table 52: ERG base case probabilities of cost effectiveness

	Standard		End of Life	
	Threshold	Probability	Threshold	Probability
Lower threshold Upper threshold				
* Applying the NICE methods 1.7 QALY multiplier				

As already noted the Weibull OS curve has a major impact upon results. If the company preferred generalised gamma is applied the deterministic ICER is £64,006 per QALY. The probabilistic ICER is £63,366 per QALY, the CEAC being presented in Figure 35.

35

If the company OS generalised gamma is retained the probabilities of dostarlimab being cost effective at the various NICE willingness to pay thresholds is presented in Table 53.

Table 53: ERG base case probabilities of cost effectiveness but retaining company OS generalised gamma.

	Standard		End of Life	
	Threshold	Probability	Threshold	Probability
Lower threshold Upper threshold				
* Applying the NICE methods 1.7 QALY multiplier				

6.1.3 ERG scenario analyses

The ERG presents the following scenario analyses:

- SA01: Assuming dostarlimab treatment cessation from [redacted] and from [redacted], retaining the assumption that all cease treatment at [redacted].

- SA02: Assuming proportions remaining on dostarlimab at [redacted] of [redacted] and [redacted].
- SA03: Assuming treatment waning starts [redacted] and [redacted] after the treatment cessation at [redacted].
- SA04 Applying the company Gompertz and company log-logistic dostarlimab TTD curves, based upon the original TTD KM data of the original company submission, and the log-normal dostarlimab TTD curve based upon the updated TTD KM data, as used within the company TE submission. Note that the ERG base case generalised gamma TTD curve has always been based upon the updated TTD KM data.
- SA05: Applying the dostarlimab TTD KM curve for the first 8 months of the model.
- SA06: Applying the quality of life values of the German study: PFS 0.701 and PPS 0.676.
- SA07: Applying a correction factor to the RWEQ treatment costs to align the modelled treatment duration with the mean stated by the company.
- SA08: Reducing the frequency of visits to the specialist nurse when in PFS off treatment to 12 weekly.
- SA09: Time horizons of 10, 20 and 30 years.
- SA10: Applying the upper and lower confidence intervals of the dostarlimab OS curve
- SA11: Applying the upper and lower confidence intervals of the dostarlimab OS, PFS and TTD curve

The deterministic cost effectiveness estimates for these scenarios are presented in Table 54.

Table 54: Scenarios around the ERG base case that applies the dostarlimab Weibull OS curve and ERG base case retaining the company preferred dostarlimab generalised gamma OS curve

	ICER	
	Weibull	**Gen.Gamm.**
Base case	£79,714	£64,006
SA01a: Cessation at [redacted]	£81,853	£63,583
SA01b: Cessation at [redacted]	£83,990	£63,140
SA02a: [redacted] dostarlimab [redacted] continuing treatment	£73,411	£59,041
SA02b: [redacted] dostarlimab [redacted] continuing treatment	£83,336	£66,859
SA03a: Waning starts [redacted] after [redacted] cessation	£77,378	£60,153
SA03b: Waning starts [redacted] after [redacted] cessation	£75,813	£57,082
SA04a: Dostarlimab Gompertz TTD curve (old data)	£80,921	£64,733
SA04b: Dostarlimab log-logistic TTD curve (old data)	£75,198	£60,225
SA04c : Dostarlimab log-normal TTD curve	£76,679	£61,392
SA05: Dostarlimab KM TTD for 8 months	£75,457	£60,429
SA06: German QoL values	£79,263	£63,465
SA07: RWEQ treatment cycles adjustment	£80,083	£64,296
SA08: Reduced specialist nurse frequency	£79,290	£64,170
SA09a: 10 year time horizon	£90,563	£74,322
SA09b: 20 year time horizon	£81,822	£65,962
SA09c: 30 year time horizon	£79,911	£64,186
SA10a : Lower CI OS curve	£139k	£95,290
SA10b : Upper CI OS curve	£57,484	£50,177
SA11a : Lower CI all curves	£123k	£84,118
SA11b : Upper CI all curves	£64,477	£56,329
SA02a + SA03a	£71,029	£55,327
SA02a + SA03b	£69,448	£52,411
SA02b + SA03a	£81,338	£63,164
SA02b + SA03b	£80,040	£60,184
SA02a + SA03a + SA04c	£68,811	£53,432
SA02a + SA03b + SA04c	£67,282	£50,626
SA02b + SA03a + SA04c	£79,558	£61,603
SA02b + SA03b + SA04c	£77,777	£58,328

It can be noted that in the above the scenario of SA02a + SA03b + SA04c and its ICER of £50,626 per QALY is the closest scenario the ERG presents to the company TE base case without company waning with its associated ICER of £49,608 per QALY. Removing the GARNET number of subsequent treatments multiplier and retaining the company preferred quality of life model further revises the SA02a + SA03b + SA04c scenario ICER to £49,795 per QALY and near complete alignment with the company TE estimate.

6.2 *ERG exploratory analyses against single RWEQ comparators*

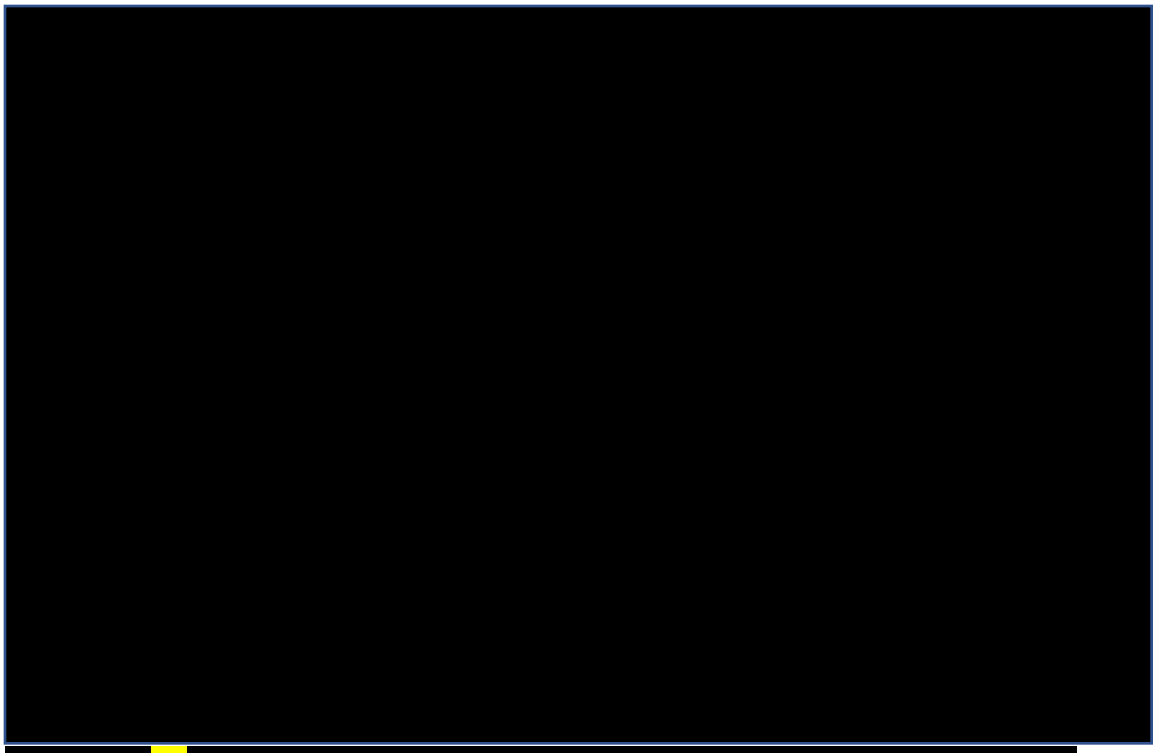
The ERG has fitted curves to the RWEQ individual treatment KM data for carboplatin + paclitaxel, carboplatin + PLD and PLD monotherapy. The ERG prefers the log-logistic parameterisation for these, with the exception of preferring the Weibull for the carboplatin + PLD TTD curve. This latter choice has little effect upon model outputs due to the company model limiting treatment to a maximum of [redacted] model cycles.

These can be used to estimate the cost effectiveness of dostarlimab against the individual treatments. A modelling issue arises as to whether the waning of effect for dostarlimab should be based upon the pooled RWEQ curves or upon the curves of the individual RWEQ treatment that is under consideration. The ERG will present the results of both approaches.

The ERG parameterised curves for the individual treatments are presented below in Figure 36, Figure 37 and Figure 38.

36

37



38

The cost effectiveness estimates for these are presented in Table 55.

Table 55: ERG RWEQ single treatment scenarios

Waning	RWEQ curves used			Comparator curves used		
Comparator	Δ QALY	Δ Cost	ICER	Δ QALY	Δ Cost	ICER
Carb+PAC	[redacted]	[redacted]	£104k	[redacted]	[redacted]	£108k
Carb+PLD	[redacted]	[redacted]	£88,929	[redacted]	[redacted]	£102k
PLD mono	[redacted]	[redacted]	£53,080	[redacted]	[redacted]	£58,120

What is perhaps most noteworthy is how much better the cost effectiveness estimate is for PLD monotherapy compared to the ERG base case of £79,714 per QALY: an improvement of 27-34%.

The company scenario analysis that compares dostarlimab with doxorubicin using the Zoptec trial suggests an ICER that is 25% worse than the company base case.

The ERG views the Zoptec trial as the most reliable of the company comparisons with doxorubicin. This may raise questions about patient recruitment during GARNET and Zoptec, how this compares with the RWEQ patient group and whether the comparison of GARNET with the RWEQ patient group may be biased.

The other company scenario analyses that compare dostarlimab with doxorubicin suggest ICERs that are 10% above, 18% below and 19% below the company base case ICER. None approach being below the base case ICER by the 27-34% of the ERG exploratory analyses.

7 END OF LIFE

Based on survival benefit estimated by the company (see CS section B.2.4.6), dostarlimab appears to meet the NICE efficacy criteria of extending life (more than 3 months survival than the current clinical management, and current clinical management survival of less than 24 months) for patients with recurrent or advanced EC that has progressed on or after platinum-based chemotherapy. However, there is uncertainty around the survival estimates as GARNET's data is immature and there are many issues surrounding data for comparators and longer-term outcomes beyond two years.

The company model estimates the following undiscounted life years. There may be some concerns around whether the values for the comparator arm are underestimates given the company experts' opinions as summarised in section ERG report section 4.2.6.3.

Table 56: Modelled undiscounted mean survival

Preferred assumption	RWEQ	DOST	Net
Company base case	[redacted]	[redacted]	[redacted]
ERG corrected company base case	[redacted]	[redacted]	[redacted]
ERG base case	[redacted]	[redacted]	[redacted]

8 References

1. National Institute for Health and Care Excellence (NICE). *Pembrolizumab for previously treated endometrial cancer [ID1205]: Project information*. 2021. URL: https://www.nice.org.uk/guidance/indevelopment/gid-ta10243 (Accessed 16 June 2021).

2. National Institute for Health and Care Excellence (NICE). *Lenvatinib with pembrolizumab for previously treated advanced endometrial cancer [ID3811]: Project information*. 2021. URL: https://www.nice.org.uk/guidance/indevelopment/gid-ta10692 (Accessed 16 June 2021).

3. European Medicines Agency (EMA). *Summary of opinion (initial authorisation): Jemperli*. 2021. URL: https://www.ema.europa.eu/en/documents/smop-initial/chmp-summary-positive-opinion-jemperli_en.pdf (Accessed 17 June 2021).

4. Oaknin A, Tinker A, Gilbert L, Samouëlian V, Mathews C, Brown J*, et al.* Clinical Activity and Safety of the Anti–Programmed Death 1 Monoclonal Antibody Dostarlimab for Patients With Recurrent or Advanced Mismatch Repair–Deficient Endometrial Cancer. A Nonrandomized Phase 1 Clinical Trial. *JAMA Oncology* 2020;6(11):1766-72.

5. Cancer Research UK. *Uterine cancer statistics*. 2020. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer (Accessed 17 June 2021).

6. European Medicines Agency (EMA). *Dostarlimab (Jemperli). Summary of Product Characteristics*. 2021. URL: https://www.ema.europa.eu/en/documents/product-information/jemperli-epar-product-information_en.pdf (Accessed 17 June 2021).

7. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol* 1983;15(1):10-7. http://dx.doi.org/10.1016/0090-8258(83)90111-7

8. Murali R, Soslow RA, Weigelt B. Classification of endometrial carcinoma: more than two types. *Lancet Oncol* 2014;15(7):e268-78. http://dx.doi.org/10.1016/s1470-2045(13)70591-6

9. Wilczyński M, Danielska J, Wilczyński J. An update of the classical Bokhman's dualistic model of endometrial cancer. *Prz Menopauzalny* 2016;15(2):63-8. http://dx.doi.org/10.5114/pm.2016.61186

10. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet* 2009;105(2):103-4. http://dx.doi.org/10.1016/j.ijgo.2009.02.012

11. National Institute for Health and Care Excellence (NICE). *Appendix C: Methodology checklist: randomised controlled trials. In: The guidelines manual [PMG6] appendices B-I*. 2012. URL: https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-appendices-bi-pdf-3304416006853 (Accessed 14 July 2021).

12. Critical Appraisal Skills Programme. *CASP Cohort Study Checklist*. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf (Accessed 20 May 2021).

13. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M*, et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj* 2016;355:i4919. http://dx.doi.org/10.1136/bmj.i4919

14. National Institute for Health and Care Excellence (NICE). *Appendix H: Appraisal checklists, evidence tables, GRADE and economic profiles*. 2015. URL: https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-appendices-2549710189/chapter/appendix-h-appraisal-checklists-evidence-tables-grade-and-economic-profiles (Accessed 14 July 2021).

15. Miller D, Scambia G, Bondarenkop I, Westermann A, Oaknin A, Oza A*, et al.* ZoptEC: Phase III randomized controlled study comparing zoptarelin with doxorubicin as second line therapy for locally advanced, recurrent, or metastatic endometrial cancer (NCT01767155). Abstract 5503. *Journal of Clinical Oncology* 2018;36(15; supplement). http://dx.doi.org/10.1200/JCO.2018.36.15_suppl.5503

16. ClinicalTrials.gov. *Zoptarelin doxorubicin (AEZS 108) as second line therapy for endometrial cancer (ZoptEC)*. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT01767155 (Accessed 10 June 2021).

17. McMeekin S, Dizon D, Barter J, Scambia G, Manzyuk L, Lisyanskaya A*, et al.* Phase III randomized trial of second-line ixabepilone versus paclitaxel or doxorubicin in women with advanced endometrial cancer. *Gynecol Oncol* 2015;138(1):18-23. http://dx.doi.org/10.1016/j.ygyno.2015.04.026

18. Rubinstein M, Halpenny D, Makker V, Grisham RN, Aghajanian C, Cadoo K. Retreatment with carboplatin and paclitaxel for recurrent endometrial cancer: A retrospective study of the Memorial Sloan Kettering Cancer Center experience. *Gynecol Oncol Rep* 2019;28:120-3. http://dx.doi.org/10.1016/j.gore.2019.04.002

19. Mazgani M, Le N, Hoskins PJ. Reuse of carboplatin and paclitaxel in patients with relapsed endometrial cancer--the British Columbia Cancer Agency experience. *Gynecol Oncol* 2008;111(3):474-7. http://dx.doi.org/10.1016/j.ygyno.2008.08.029

20. Julius JM, Tanyi JL, Nogueras-Gonzalez GM, Watkins JL, Coleman RL, Wolf JK*, et al.* Evaluation of pegylated liposomal doxorubicin dose on the adverse drug event profile and outcomes in treatment of recurrent endometrial cancer. *Int J Gynecol Cancer* 2013;23(2):348-54. http://dx.doi.org/10.1097/IGC.0b013e31827c18f3

21. Makker V, Hensley ML, Zhou Q, Iasonos A, Aghajanian CA. Treatment of advanced or recurrent endometrial carcinoma with doxorubicin in patients progressing after paclitaxel/carboplatin: Memorial Sloan-Kettering Cancer Center experience from 1995 to 2009. *Int J Gynecol Cancer* 2013;23(5):929-34. http://dx.doi.org/10.1097/IGC.0b013e3182915c20

22. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R*, et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45(2):228-47. http://dx.doi.org/10.1016/j.ejca.2008.10.026

23. Mathews C, Im E, Alfaya L, Travers K, Gibson CJ. Review of evidence for predictive value of microsatellite instability/mismatch repair status in response to non–anti-PD-(L)1 therapies in patients with advanced or recurrent endometrial cancer. In: 34th Annual Meeting & Pre-Conference Programs of the Society for Immunotherapy of Cancer (SITC 2019): part 1. *Journal for ImmunoTherapy of Cancer* 2019;7 (Suppl 1):282. http://dx.doi.org/10.1186/s40425-019-0763-1

24. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET*, et al.* Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;5(6):649-55.

25. Karnofsky D, Burchenal J. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod C, editor. *Evaluation of Chemotherapeutic Agents* New York, NY: Columbia University Press; 1949:191–205.

26. The ECOG-ACRIN Cancer Research Group. *ECOG Performance Status*. 2020. URL: https://ecog-acrin.org/resources/ecog-performance-status (Accessed 4 June 2021).

27. National Cancer Institute. *Common Terminology Criteria for Adverse Events (CTCAE)*. URL: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm (Accessed 26 May 2021).

28. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. *National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU) Technical Support Document (TSD) 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE*. 2016. URL: http://nicedsu.org.uk/technical-support-documents/population-adjusted-indirect-comparisons-maic-and-stc/ (Accessed 19 June 2021).

29. Remiro-Azócar A, Heath A, Baio G. Conflating marginal and conditional treatment effects: Comments on 'Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study'. *Stat Med* 2021;40(11):2753-8.

30. Expert Panel on Radiation Oncology–Gynecology. *Management of recurrent endometrial cancer*. American College of Radiology (ACR) Appropriateness Criteria; 2016. URL: https://acsearch.acr.org/docs/3094112/Narrative/ (Accessed 17 June 2021).

31. Makker V, Rasco D, Vogelzang NJ, Brose MS, Cohn AL, Mier J*, et al.* Lenvatinib plus pembrolizumab in patients with advanced endometrial cancer: an interim analysis

of a multicentre, open-label, single-arm, phase 2 trial. *Lancet Oncology* 2019;20(5):711-8. http://dx.doi.org/https://dx.doi.org/10.1016/S1470-2045(19)30020-8

32. Makker V, Taylor MH, Aghajanian C, Oaknin A, Mier J, Cohn AL*, et al.* Lenvatinib Plus Pembrolizumab in Patients With Advanced Endometrial Cancer. *Journal of Clinical Oncology* 2020;38(26):2981-92. http://dx.doi.org/https://dx.doi.org/10.1200/JCO.19.02627

33. Ott PA, Bang YJ, Berton-Rigaud D, Elez E, Pishvaian MJ, Rugo HS*, et al.* Safety and Antitumor Activity of Pembrolizumab in Advanced Programmed Death Ligand 1-Positive Endometrial Cancer: Results From the KEYNOTE-028 Study. *Journal of Clinical Oncology* 2017;35(22):2535-41. http://dx.doi.org/10.1200/JCO.2017.72.5952

34. Barlesi F, Steins M, Horn L, Ready N, Felip E, Borghaei H*, et al.* Long-term outcomes with nivolumab versus docetaxel in patients with advanced NSCLC: checkmate 017 and checkmate 057 2-year update. Paper presented at: COSA's 43rd and ANZBCTG's 38th Annual Scientific Meetings. Partners for Progress in Breast Cancer Research and Care. 15–17 November 2016; Gold Coast Convention and Exhibition Centre, Queensland. http://dx.doi.org/10.1111/ajco.12646

35. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE*, et al.* Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373(17):1627-39. http://dx.doi.org/10.1056/NEJMoa1507643

36. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E*, et al.* Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373(2):123-35. http://dx.doi.org/10.1056/NEJMoa1504627

37. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L*, et al.* First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet* 2021;398:27-40. http://dx.doi.org/10.1016/s0140-6736(21)00797-2

38. National Institute for Health and Care Excellence (NICE). *Pembrolizumab for untreated metastatic or unresectable recurrent head and neck squamous cell carcinoma [TA661]*. 2020. URL: https://www.nice.org.uk/guidance/ta661 (Accessed 16 June 2021).

39. National Institute for Health and Care Excellence (NICE). *Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy [TA520]*. 2018. URL: https://www.nice.org.uk/guidance/ta520 (Accessed 16 June 2021).

40. National Institute for Health and Care Excellence (NICE). *Durvalumab for treating locally advanced unresectable non-small-cell lung cancer after platinum-based chemoradiation [TA578]*. 2019. URL: https://www.nice.org.uk/guidance/ta578 (Accessed 16 June 2021).

41. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nat Commun* 2020;11(1):3801. http://dx.doi.org/10.1038/s41467-020-17670-y

42. Schadendorf D, Wolchok JD, Hodi FS, Chiarion-Sileni V, Gonzalez R, Rutkowski P*, et al.* Efficacy and Safety Outcomes in Patients With Advanced Melanoma Who Discontinued Treatment With Nivolumab and Ipilimumab Because of Adverse Events: A Pooled Analysis of Randomized Phase II and III Trials. *J Clin Oncol* 2017;35(34):3807-14. http://dx.doi.org/10.1200/jco.2017.73.2289

9 Appendix

9.1 *ERG assessment of the company SLR and included clinical studies*

Table 57: ERG ROBIS assessment of risks of bias of the CS systematic review of clinical effectiveness

ROBIS domain, and signalling questions	ERG's assessment of whether criteria met, with comments
DOMAIN 1: STUDY ELIGIBILITY CRITERIA	
1.1 Did the review adhere to pre-defined objectives and eligibility criteria?	**Probably Yes**. No pre-published protocol, unclear if the changes made to searches (CS Appendix D.2) at the update were made *a priori*. The same eligibility criteria was used for the original and update clinical SLR (CS Appendix D.1 Table 1). A date limit was not set for the eligibility criteria; however, a date limit was applied to the MEDLINE and Embase update searches which means that older (pre-2018) may have been missed (particularly for paclitaxel studies and some systematic reviews study designs which were included search terms in the update SLR). However, this is mitigated by the fact that the Cochrane Library update and trials register searches were not date limited.
1.2 Were the eligibility criteria appropriate for the review question?	**Probably Yes**. The criteria presented in CS Appendix D.1 Table 1 are appropriate for the review question. The company did not consider hormone therapy (which was within the NICE final scope) as one of the comparators included in the original or update clinical SLR. However, the company provided a targeted literature review (TLR) for hormone therapy (CS Appendix L). The review followed the same eligibility criteria of the clinical SLR. No studies from the hormone therapy TLR was included in the economic evaluation. No additional relevant studies were identified by the ERG. A scenario analysis was conducted with hormone therapy, using the UK RWE study as a proxy to validate the base-case.
1.3 Were eligibility criteria unambiguous?	**Yes**. Eligibility criteria were unambiguous.
1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate?	**Probably Yes**. Most of the restrictions were appropriate. However, a reason for limiting sample size to ≥ 20 patients for observational studies was not provided. It is unclear whether this is appropriate.
1.5 Were any restrictions in eligibility criteria based on sources of information appropriate?	**Yes**. No language restrictions were applied.

Domain 1 risk of bias	**Low concern**
DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES	
2.1 Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?	**Yes.** Searched MEDLINE, Embase and Cochrane (CDSR and CENTRAL), ClinicalTrials.gov and EU clinical trials register (CS Appendix D.2).
2.2 Were methods additional to database searching used to identify relevant reports?	**Yes.** Hand searching of some relevant conferences and websites was undertaken.
2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	**Yes**. Full searches are reported for database sources. A variety of terms were used for each concept and these were combined correctly. However, the update search strategy is different in parts to the original searches, focussing on fewer interventions/comparators (reflecting those in the CS decision problem, except for hormone therapy), but with a broader population (EC rather than recurrent or advanced EC) and with other study types included. Terms for paclitaxel and systematic reviews are included in the update searches, but not in the original searches. The date limit applied to the MEDLINE and Embase update searches means that older (pre-2018) paclitaxel studies and some systematic reviews may have been missed, although this is mitigated by the fact that the Cochrane Library update and trials register searches were not date limited.
2.4 Were restrictions based on date, publication format, or language appropriate?	**Yes**. There are no restrictions on publication format or language in the search strategies. Date limits are appropriate, but note issue of older (pre 2018) paclitaxel studies and systematic reviews in 2.3 above
2.5 Were efforts made to minimise errors in selection of studies?	**Yes.** Title/abstract screening and full text screening for the wider SLR were undertaken by two reviewers. For title/abstract screening, where there was disagreement about the relevance of a study, it was progressed to full text screening. For full text screening, where there was disagreement about the relevance of a study, reasons for inconsistencies were discussed, if an agreement was not reached, a third reviewer was invited to make a judgment.
Domain 2 risk of bias	**Low concern**
DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL	
3.1 Were efforts made to minimise error in data collection?	**Yes**. Pre-defined extraction form used, extraction by two reviewers and verification by a third reviewer.
3.2 Were sufficient study characteristics available for both review authors	**Yes**. Characteristics of the thirteen studies meeting the eligibility criteria were presented by the company (CS Appendix D.4.3).

and readers to be able to interpret the results?	
3.3 Were all relevant study results collected for use in the synthesis?	**Yes**. CS Appendix D.4 (table 20 to table 28).
3.4 Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	**Probably Yes.** Methodological quality was assessed using Appendix C of PMG6 refers - methodology checklist for randomised controlled trials in the old NICE guidelines manual) for RCTs, CASP check list for Non-RCTs, and ROBINS I assessment tool for the UK RWE study (CS section B.2.3.1.4 and Appendix D.7). These are not the tools preferred by NICE. The CS does not justify using a non-preferred checklist. The ERG quality assessed the studies using both the company's preferred checklist and the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) checklist for RCTs; and both the company's preferred checklist and the Institute of Health Economics checklist for Non-RCTs (including the UK RWE study). The ERG had some differences with the company's judgements (CS section B.2.3.1.4 and Appendix D.7 (see ERG report section 3.1).
3.5 Were efforts made to minimise error in risk of bias assessment?	**Yes**. Assessments by two reviewers, and reasons for disagreements were discussed and verified. Individual study authors were contacted for missing or incomplete information.
Domain 3 risk of bias	**Low concern**
DOMAIN 4: SYNTHESIS AND FINDINGS	
4.1 Did the synthesis include all studies that it should?	**No**. One study (Lissoni *et al* 1996) was excluded from the SLR (CS Appendix D.4.2 Table 14), but the ERG does not consider it as a previously identified paper and recommends its inclusion. However, the study is not eligible for the MAICs as PFS or OS data is not reported and therefore not important for economic evaluation.
4.2 Were all predefined analyses followed or departures explained?	**No information**. Pre-defined analyses not specified in the CS.
4.3 Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?	**Probably Yes**. Results from different studies were described narratively. No meta-analysis was undertaken.
4.4 Was between-studies variation (heterogeneity) minimal or addressed in the synthesis?	**Probably No**. Heterogeneity was not explicitly discussed by the CS. However, the company aimed to generate a dataset for the comparator that was aligned with the patient population of GARNET.

4.5 Were the findings robust, e.g., as demonstrated through funnel plot or sensitivity analyses?	**Probably Yes**. Results from different studies were described narratively. The author discussed some studies that may be problematic.
4.6 Were biases in primary studies minimal or addressed in the synthesis?	**Probably No**. Biases were assessed using the company's preferred tools (CS section B.2.3.1.4 and Appendix D.7). Most of the Non-RCT studies had a high risk of bias, while the RCTs mostly had a low risk of bias. The ERG had some differences with the company's judgements (CS section B.2.3.1.4 and Appendix D.7 (see ERG report section 3.1). The quality of the studies were highlighted in the findings or conclusions of the review.
Domain 4 risk of bias	**Unclear concern**
Overall risk of bias in the review	
A. Did the interpretation of findings address all of the concerns identified in Domains 1 to 4?	**No.** The company did not addressed heterogeneity in study results in their analysis. In addition, the company did not provide information on predefined analyses in an explicitly referenced protocol or in the submission.
B. Was the relevance of identified studies to the review's research question appropriately considered?	**Yes**. Studies included were relevant to the objective and inclusion / exclusion criteria. The relevance of the included studies was also highlighted as part of the individual study quality assessment.
C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?	**Probably No**. There was no bias in the reporting of the findings from the review.
Risk of bias in the review	**Low risk of bias with some concern**

ERG summary assessment of risks of bias of the clinical effectiveness, Hormone Therapy Targeted Literature Review

The company did not initially consider hormone therapy (which was within the NICE final scope) as one of the comparators, and thus was not included in the original or update clinical SLR. However, the company provided a targeted literature review (TLR) for hormone therapy (CS Appendix L). The review followed the same eligibility criteria of the clinical SLR; however, efforts made to minimize errors in selection of studies during title/abstract or full text screening were not reported. All articles screened at full text stage were initially excluded and then re-evaluated with a relaxed set of inclusion criteria, in the effort to identify for hormone therapy (CS Appendix L.5). The ERG notes that it is uncertain whether any other records excluded at title/abstract screening would have met these relaxed criteria. No studies from the hormone therapy TLR were found relevant by the company for this submission; thus, none was included in the cost-effectiveness analysis. A scenario analysis was conducted with hormone therapy, using the UK RWE study as a proxy to validate the base-case.

Table 58: ERG and company assessment of RCT risk bias (Appendix C of PMG6 refers - methodology checklist for randomised controlled trials in the old NICE guidelines manual)

		ZoptEC ERG		ZoptEC CS	
A. Selection bias (systematic differences between the comparison groups)					
A1	An appropriate method of randomization was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)	Yes	Patients were randomized in a 1:1 ratio to receive treatment with either AEZS-108 (Arm A) or doxorubicin (Arm B).	Yes	Centrally randomised
A2	There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)	Yes	Yes, participants were randomly allocated by central randomisation.	Unclear	Centrally randomised but otherwise not reported
A3	The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	Baseline demographics were similar for treatment groups.	YES/PARTIAL	Similar for age, race, ECOG and stage
Likely direction of effect			**Low risk of bias**	**Low risk of bias,** There appeared to be low risk for systematic differences between comparison groups	
B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)					
B1	The comparison groups received the same care apart from the intervention(s) studied	Unclear	It is unclear whether concurrent treatment administration was balanced across intervention groups.	Yes	The groups appeared to receive the same care
B2	Participants receiving care were kept 'blind' to treatment allocation	No	Open label study.	No	Open label.
B3	Individuals administering care were kept 'blind' to treatment allocation	No	Open label study.	No	Open label.
Likely direction of effect			**High risk of bias**	**High risk of bias,** Although this was an open label trial, the outcomes were objective.	
C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants)					
C1	All groups were followed up for an equal length of time (or analysis was adjusted	Yes	There is a standard follow-up protocol for all patients.	Yes	The final analysis, which was event-based, was conducted

	to allow for differences in length of follow-up)				after approximately 384 randomised patients had died
C2	a. How many participants did not complete treatment in each group?		Not reported.		
C2	b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Unclear	No information on dropouts.	Yes	13/256 vs. 15/255 did not complete
C3	a. For how many participants in each group were no outcome data available?		There were missing outcome reports for PFS for 10 participants (that is patients allocated to a treatment but never treated). 4 patients in AEZS-108 /Zoptarelin Doxorubicin group, and 6 patients in Doxorubicin group.		
C3	b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available).	Unclear	It is unclear whether there are any important differences between those with and without outcome data in both intervention groups.	Yes	Analysis performed in the ITT or mITT (Excluding patients allocated to a treatment but never treated)
Likely direction of effect			**Unclear**	**Low risk of bias,** There was a low risk of attrition bias	
D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)					
D1	The study had an appropriate length of follow-up	Yes	Yes, 3.87 years overall study follow-up period from start date – August 2013 to completion – January 30,2017.	Yes	Yes, the OS and PFS data were mature
D2	The study used a precise definition of outcome	Yes		Yes	The primary endpoint was OS, and other endpoints included PFS, ORR and CBR
D3	A valid and reliable method was used to determine the outcome	Yes	Standardised measurements were used to assess the outcomes. Response and progression were	Yes	Standard outcomes for OS, PFS and ORR were evaluated

			evaluated using the international criteria proposed by the revised Response Evaluation Criteria in Solid Tumors (RECIST) guideline (version 1.1). Adverse events (AEs) were assessed according to the National Cancer Institute CTCAE version 4.03 or subsequent ones.		
D4	Investigators were kept 'blind' to patients' exposure to the intervention	No	Open label study.	No	Open label study.
D5	Investigators were kept 'blind' to other important confounding and prognostic factors	No	Open label study.	No	Open label study.
Likely direction of effect			**Unclear**	**Low risk of bias,** Although this was an open label trial, the outcomes were objective	

		McMeekin et al 2015 ERG		**McMeekin et al 2015 CS**	
A. Selection bias (systematic differences between the comparison groups)					
A1	An appropriate method of randomization was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)	Yes	Patients were randomized in a 1:1 ratio to ixabepilone or either paclitaxel or doxorubicin, depending on prior therapy received.	Unclear	Not reported.
A2	There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)	Unclear	It is unclear whether treatment group allocation was concealed.	Unclear	Not reported.
A3	The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	Baseline demographics were similar for randomised patients in the ixabepilone and control arms.	Yes	Prognostic factors appear balanced at baseline.
Likely direction of effect			**Unclear**	**Unclear,** Methods of randomisation were unclear	
B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)					

B1	The comparison groups received the same care apart from the intervention(s) studied	Unclear	It is unclear whether concurrent administration of hormone replacement therapy were balanced across intervention groups.	Yes	
B2	Participants receiving care were kept 'blind' to treatment allocation	No	Open label study.	No	Open label study.
B3	Individuals administering care were kept 'blind' to treatment allocation	No	Open label study.	No	Open label study.
Likely direction of effect			**Unclear**	**High risk of bias,** Although this was an open label trial, the outcomes were objective	
C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants)					
C1	All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)	Unclear	All participants were followed up for at least 6 months.	Yes	An interim analysis was conducted after 176 deaths had been observed or 300 patients had been randomized and followed for 6 months, whichever came earlier. If the follow-up on 300 patients occurred first, a minimum number of 160 deaths were required before conducting the futility analysis
C2	a. How many participants did not complete treatment in each group?		As at database lock date, a very high number of participants had dropped out of the study treatments (209 in the ixabepilone arm and 210 in the control arm).		
C2	b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Unclear	It is unclear if there are any significant differences between those who dropped out and those who stayed on treatment.	Yes	At the time of database lock (DBL; February 8, 2012), 419 patients were off study treatment, 209 in the ixabepilone arm and 210 in the control arm. The most common reason for treatment discontinuation was

					disease progression (52% of ixabepilone patients and 53% of control patients) and study drug toxicity (14% of ixabepilone patients and 7% of control patients).
C3	a. For how many participants in each group were no outcome data available?		There were missing outcome reports for PFS and ORR for 25 participants in each group.		
C3	b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available).	Unclear	Outcome reports for PFS and ORR was not reported for equal number of patients in both groups - 25 participants in each group. It is unclear whether there are any important differences between those with and without outcome data in both intervention groups.	Yes	Efficacy was reported in all randomised patients
Likely direction of effect			**Unclear**	**Low risk of bias**	
D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)					
D1	The study had an appropriate length of follow-up	Yes	Yes, 29.6 months overall study follow-up period from start date – August 17, 2009.	Yes	Yes, the OS and PFS data were mature
D2	The study used a precise definition of outcome	Unclear	The definition of outcomes were not noted.	Yes	The primary endpoint was OS, and other endpoints included PFS and ORR
D3	A valid and reliable method was used to determine the outcome	Yes	Standardised measurements were used to assess the outcomes. Response and disease progression were evaluated using Response Evaluation Criteria in Solid Tumors, version 1.1. Adverse events (AEs), laboratory values, and other signs and symptoms were assessed according to the National Cancer Institute CTCAE version 3.0.	Yes	Standard outcomes for OS, PFS and ORR were evaluated

D4	Investigators were kept 'blind' to patients' exposure to the intervention	No	Open label study.	No	Open label study.
D5	Investigators were kept 'blind' to other important confounding and prognostic factors	No	Open label study.	No	Open-label
Likely direction of effect			**Unclear**	**Low risk of bias,** Although this was an open label trial, the outcomes were objective.	

Table 59: ERG and company assessment of non-RCT risk of bias (CASP cohort study checklist)

Section A: Are the results of the study valid?	GARNET ERG	GARNET CS
1. Did the study address a clearly focused question issue?	Yes The objective of the study was "to evaluate the antitumor activity of dostarlimab in participants with recurrent and advanced dMMR/MSI-H EC, in terms of ORR and DOR by blinded independent central review (BICR) using Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1)"".	Yes Objective: To evaluate the antitumour activity of dostarlimab in subjects with recurrent and advanced dMMR/MSI-H EC, in terms of ORR and DOR by BICR using RECIST v1.1
2. Was the cohort recruited in an acceptable way?	Can't Tell CS section B.2.3.1.1, and Appendix P.1	Yes Patients were recruited from 117 sites in 9 countries as part of this multicentre, global clinical trial according to pre-defined eligibility criteria
3. Was the exposure accurately measured to minimise bias?	Yes Patient Baseline Characteristics were accurately classified, such as FIGO disease stage at diagnosis, histology, type and number of prior lines of therapy were presented.	Yes Standard, validated, objective measurements were evaluated including ORR, DOR, DCR, PFS.
4. Was the outcome accurately measured to minimise bias?	Yes CS section B.2.3.1	Yes Outcomes were assessed by BICR according to RECIST criteria
5a. Have the authors identified all important confounding factors?	Not applicable	Yes/Partial Predefined subgroup data cross some factors
5b. Have they taken account of the confounding factors in the design and/or analysis? Or Could there	Not applicable Descriptive statistics	Yes/Partial Predefined subgroup data cross some factors

be confounding factors that haven't been accounted for?		
6a. Was the follow up of subjects complete enough?	Can't Tell CS Appendix D.6 table 64	Yes Follow-up was sufficiently recorded: The most common reason for treatment discontinuation was PD; Most of the study discontinuations were because of death
6b. Was the follow up of subjects long enough?	No Not for whole sample. As at the time of Cut-off for analysis March 1, 2020, the median in study follow-up time was 16.3 months; median Duration of response (DOR) and median OS was not reached.	No Median OS was immature; however the follow-up was long enough to determine the other outcomes
Section B: What are the results?	**GARNET ERG**	**GARNET CS**
7. What are the results of this study?	- CS section B.2.4	
8. How precise are the results?	- Confidence intervals (CIs) reported for all outcomes except adverse events.	Yes/Partial 95% CIs were generally within a reasonable range; some of the smaller subgroups are large intervals
9. Do you believe the results?	Can't tell Study was unblinded, and single-arm with small sample size. The results should be interpreted with caution.	Yes Evaluated by BICR under clinical trial conditions
Section C: Will the results help locally?	**GARNET ERG**	**GARNET CS**
10. Can the results be applied to the local population?	Can't Tell Study was unblinded, and single-arm with small sample size. The results must be interpreted with caution.	Yes/Partial A global multicentre study with generally good generalisability; however, the majority of patients were White so may not be relevant to some populations

11. Do the results of this study fit with other available evidence?	Can't Tell No published studies for dostarlimab in recurrent or advanced EC is available.	Unclear No other published studies for dostarlimab in EC
12. What are the implications of this study for practice?	Can't Tell Implications for Practice "Study results from this interim analysis (IA-2) demonstrate that dostarlimab treatment results in durable responses in a substantial proportion of participants with recurrent or advanced dMMR or dMMR/MSI-H EC". Study was unblinded, and single-arm with small sample size. Difficult to draw conclusion because of study design.	Unclear Clinical trial evidence for dostarlimab in EC; however not an RCT & therefore the extent of benefit vs. other treatments is not clear

Section A: Are the results of the study valid?	Makker et al 2013 ERG comments	Makker et al 2013 CS comments
1. Did the study address a clearly focused question issue?	Yes The objective of the study was "To determine the efficacy of second-line doxorubicin in the treatment of advanced/recurrent endometrial carcinoma that has progressed after adjuvant paclitaxel/carboplatin (TC) therapy among patients treated at MSKCC between 1995 and 2009."	Yes Objective: To investigate the activity of doxorubicin in the second-line setting in patients who progressed after paclitaxel/carboplatin adjuvant treatment
2. Was the cohort recruited in an acceptable way?	No Retrospective study. Participants were recruited from electronic medical records.	No Single centre, retrospective study
3. Was the exposure accurately measured to minimise bias?	Yes CS Appendix D.4.3	Yes Standard, validated, objective measurements were evaluated including ORR by RECIST criteria, OS and PFS
4. Was the outcome accurately measured to minimise bias?	Yes RECIST v1.1 was used. Toxicity was assessed version 4.0 of Common Terminology Criteria for Adverse Events (CTCAE).	Unclear Response was defined according to standard RECIST criteria; not clear if blinded
5a. Have the authors identified all important confounding factors?	Not applicable	Unclear NR
5b. Have they taken account of the confounding factors in the design and/or analysis? Or Could there be confounding factors that haven't been accounted for?	Not applicable Descriptive Statistics.	Unclear NR

6a. Was the follow up of subjects complete enough?	Can't Tell It is unclear if the excluded 8 subjects will have different outcomes than those assessed.	No Follow-up not reported
6b. Was the follow up of subjects long enough?	Can't Tell Only noted the follow-up duration of the one patient alive after receiving the doxorubicin treatment (49.4 months).	Yes The follow-up was sufficient for the outcomes assessed
Section B: What are the results?	**Makker et al 2013 ERG comments**	**Makker et al 2013 CS comments**
7. What are the results of this study?	- CS Appendix D.4	
8. How precise are the results?	- Confidence intervals (CIs) reported for all outcomes except adverse events. CIs were large for all outcomes.	Yes 95% CIs were generally within a reasonable range
9. Do you believe the results?	Can't tell Study was unblinded, retrospective, with patient selection bias and small sample size. The results should be interpreted with caution.	Yes/Partial Results appear reliable, although small population <30
Section C: Will the results help locally?	**Makker et al 2013 ERG comments**	**Makker et al 2013 CS comments**
10. Can the results be applied to the local population?	Can't Tell Study was unblinded, and single-arm with a small sample size. The results must be interpreted with caution.	Unclear Patients from single centre in US
11. Do the results of this study fit with other available evidence?	Yes CS Appendix D.4	Unclear Only a few similar studies

12. What are the implications of this study for practice?	Yes Implications for Practice “Doxorubicin may be considered inactive as second-line therapy in this endometrial carcinoma population.”	No Single arm, single centre, retrospective study; small patient population

Section A: Are the results of the study valid?	**Mazgani et al 2008 ERG comments**	**Mazgani et al 2008 CS comments**
1. Did the study address a clearly focused question issue?	Yes The objective of the study was “To evaluate the efficacy of reusing carboplatin and paclitaxel (taxol) in women with relapsed endometrial cancer.”	Yes Objective: To evaluate the efficacy of reusing carboplatin and taxol in women with relapsed EC
2. Was the cohort recruited in an acceptable way?	No Participants were selectively recruited.	No Single centre, retrospective study
3. Was the exposure accurately measured to minimise bias?	Yes Patient were accurately classified into endometroid and papillary serous histology groups.	Yes Objective measurements were evaluated including response, OS, PFS
4. Was the outcome accurately measured to minimise bias?	Can’t Tell Response Confirmatory measure deviated from RECIST criteria.	Unclear Response was defined according to standard RECIST criteria; not clear if blinded
5a. Have the authors identified all important confounding factors?	Can’t Tell Not reported.	Unclear Not reported.
5b. Have they taken account of the confounding factors in the design	Can’t Tell	Unclear

and/or analysis? Or Could there be confounding factors that haven't been accounted for?	Not reported.	Not reported.
6a. Was the follow up of subjects complete enough?	Can't Tell Not reported.	No Follow-up not reported
6b. Was the follow up of subjects long enough?	Can't Tell Not reported.	Yes The follow-up was sufficient for the outcomes assessed
Section B: What are the results?	**Mazgani et al 2008 ERG comments**	**Mazgani et al 2008 CS comments**
7. What are the results of this study?	- CS Appendix D.4.4, D.4.5, D.4.6	
8. How precise are the results?	- Confidence intervals (CIs) were reported for most outcomes. CIs were wide for most outcomes, particularly for outcomes related to papillary serous histology subgroup.	Unclear 95% CIs not reported for all outcomes; reasonable range for some but large for serous histology subgroup (small population size)
9. Do you believe the results?	Can't tell Study was unblinded, retrospective, with patient selection bias and small sample size. The results should be interpreted with caution.	Yes/Partial Results appear reliable, although small population <30
Section C: Will the results help locally?	**Mazgani et al 2008 ERG comments**	**Mazgani et al 2008 CS comments**
10. Can the results be applied to the local population?	Can't Tell Study was unblinded, and single-arm with a small sample size. The results must be interpreted with caution.	Unclear Patients from single centre in Canada/baseline characteristics NR
11. Do the results of this study fit with other available evidence?	Yes	Unclear

	CS Appendix D.4 According to the Authors “Other than in the case series of Markman et al. who described 3 patients who had relapsed metastatic endometrial cancer with persistent chemosensitivity to platinumand/or paclitaxel,we were unable to find any other data on the reuse of carboplatin–taxol in relapsed endometrial cancer in the English language literature”. However, this study fits well with studies including other types of chemotherapies.	Author states there are no other studies about reuse of carboplatin–taxol in relapsed EC in the English language literature
12. What are the implications of this study for practice?	Yes Implications for Practice “Carboplatin–taxol regimen is an efficacious treatment. Due to the patient selection these outcomes reported are likely to be an overstatement of what could be achieved in practice.” Study was unblinded, retrospective, with patient selection bias and small sample size. Difficult to draw conclusion because of study design.	No Single arm, single centre, retrospective study

Section A: Are the results of the study valid?	**Rubinstein et al 2019 ERG comments**	**Rubinstein et al 2019 CS comments**
1. Did the study address a clearly focused question issue?	Yes The objective of the study was “To determine the efficacy of second-line doxorubicin in the treatment of advanced/recurrent endometrial carcinoma that has progressed after	Yes Objective: To examine the clinical outcomes of EC patients who received PC in the adjuvant setting and who were specifically re-treated with PC in the recurrent or metastatic disease setting

	adjuvant paclitaxel/carboplatin (TC) therapy.”	
2. Was the cohort recruited in an acceptable way?	No Retrospective study. Participants were recruited from an institutional database.	No Single center, retrospective study
3. Was the exposure accurately measured to minimise bias?	Yes CS Appendix D.4.3.	Yes Standard, validated, objective measurements were evaluated including response (RECIST), OS and PFS
4. Was the outcome accurately measured to minimise bias?	Yes RECIST v1.1 was used.	Yes An independent radiologist, blinded to patients' clinical details assessed response per RECIST 1.1 criteria
5a. Have the authors identified all important confounding factors?	Not applicable	Unclear Some baseline prognostic factors are not reported & data not reported by prognostic/confounders
5b. Have they taken account of the confounding factors in the design and/or analysis? Or Could there be confounding factors that haven’t been accounted for?	Not applicable Descriptive Statistics.	Unclear Not reported
6a. Was the follow up of subjects complete enough?	Can’t Tell It is unclear if the excluded 5 subjects will have different outcomes than those assessed.	No Follow-up not reported
6b. Was the follow up of subjects long enough?	Can’t Tell No information.	Yes The follow-up was sufficient for the outcomes assessed

Section B: What are the results?	Rubinstein et al 2019 ERG comments	Rubinstein et al 2019 CS comments
7. What are the results of this study?	- CS Appendix D.4	
8. How precise are the results?	- Confidence intervals reported for most outcomes. Reported CIs were wide.	No 95% CIs not reported for all outcomes; quite large range for some outcomes
9. Do you believe the results?	Can't tell Study was unblinded, retrospective, with patient selection bias and small sample size. The results should be interpreted with caution.	Yes/Partial Results appear reliable, although small population <30
Section C: Will the results help locally?	**Rubinstein et al 2019 ERG comments**	**Rubinstein et al 2019 CS comments**
10. Can the results be applied to the local population?	Can't Tell Study was unblinded, and single-arm with a small sample size. The results must be interpreted with caution.	Unclear Patients from single centre in Canada/baseline characteristics NR
11. Do the results of this study fit with other available evidence?	Yes CS Appendix D.4.	Unclear Only a few similar studies
12. What are the implications of this study for practice?	Yes Implications for Practice "selected patients with recurrent endometrial cancer (EC) who are >6 months from completion of paclitaxel and carboplatin (PC) derive benefit from retreatment with PC with a response rate of 50%."	No Single arm, single centre, retrospective study, small population

Section A: Are the results of the study valid?	Julius et al 2013 ERG comments	Julius et al 2013 CS comments
1. Did the study address a clearly focused question issue?	Yes The objective of the study was "To determine factors which may increase the likelihood of adverse drug events (ADEs) in recurrent endometrial cancer patients treated with pegylated liposomal doxorubicin (PLD) as well as this agent's impact on clinical outcomes."	Yes Objective: To determine factors which may increase the likelihood of ADEs in recurrent EC patients treated with pegylated liposomal doxorubicin
2. Was the cohort recruited in an acceptable way?	No Retrospective study. Participants were recruited from a medical records database.	No Single center, retrospective study
3. Was the exposure accurately measured to minimise bias?	Yes CS Appendix D.4.3	No Objective measures OS and PFS & TTP evaluated; response was an outcome but was not reported
4. Was the outcome accurately measured to minimise bias?	Can't Tell More detail is needed on the methods of outcomes assessment. E.g. what was the criteria used to assess radiographic evidence of response to therapy.	Unclear Limited details of evaluations.
5a. Have the authors identified all important confounding factors?	No Platinum sensitivity status was identified as an important confounding factor. However, other factors could have been noted, such as age, BMI, comorbidities, and race/ethnicity, number of prior	Unclear Some baseline prognostic factors are not reported & data not reported by prognostic/confounders

	chemotherapy, cycles of chemotherapy prior to receiving PLD, stage of disease, type of endometrial cancer histology classification, ECOG status e.t.c.	
5b. Have they taken account of the confounding factors in the design and/or analysis? Or Could there be confounding factors that haven't been accounted for?	No Other factors could have been noted, such as age, BMI, comorbidities, and race/ethnicity, number of prior chemotherapy, cycles of chemotherapy prior to receiving PLD, stage of disease, type of endometrial cancer histology classification, ECOG status e.t.c.	Unclear Not reported
6a. Was the follow up of subjects complete enough?	Can't Tell No information.	Yes Follow-up was sufficiently reported
6b. Was the follow up of subjects long enough?	Can't Tell Not reported.	Yes The follow-up was sufficient for the outcomes assessed
Section B: What are the results?	**Julius et al 2013 ERG comments**	**Julius et al 2013 CS comments**
7. What are the results of this study?	- CS Appendix D.4. Median overall PFS for all doses combined was not reported.	
8. How precise are the results?	- Confidence intervals were not reported.	Unclear 95% CIs not reported
9. Do you believe the results?	Can't tell Study was unblinded, retrospective, with patient selection bias and small sample size. The results should be interpreted with caution.	Yes/Partial Results appear reliable, although small population <30

Section C: Will the results help locally?	Julius et al 2013 ERG comments	Julius et al 2013 CS comments
10. Can the results be applied to the local population?	Can't Tell Study was unblinded, and single-arm with small sample size, heterogeneity of patients, and lack of dose diversity. The results must be interpreted with caution.	Unclear Patients from single centre in US
11. Do the results of this study fit with other available evidence?	Yes CS Appendix D.4	Unclear Only a few similar studies
12. What are the implications of this study for practice?	Yes Implications for Practice "this is one of the first studies to demonstrate benefit of PLD in recurrent endometrial cancer as well as that dose level did not significantly influence efficacy. This study confirmed cumulative dose/cycles did increase risk of toxicity with PLD, which is common with most cytotoxic agents. PLD remains a viable option for patients with recurrent or progressive endometrial cancer..."	No Single arm, single center, retrospective study; small patient population

Table 60: ERG and company assessment of UK RWE study risk of bias (The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool)

ROBINS-I tool (Stage I): At protocol stage		
Specify the review question CS		**ERG assessment**
Participants	English patients diagnosed with advanced or recurrent endometrial cancer who have progressed on or after first-line platinum doublet therapy, specifically a GARNET trial-like cohort i.e. application of the inclusion and exclusion criteria as per the GARNET TRIAL where possible	England residents with at least one incident primary diagnosis of advanced or recurrent endometrial cancer between 01/01/2013 and 31/12/2018 who must have received exactly one prior platinum doublet therapy for recurrent or advanced disease.
Experimental intervention	Current UK treatment paradigms as a basket of treatments, in the line directly post-platinum	Basket of common chemotherapy regimens.
Comparator	Not applicable	None
Outcomes	Survival outcomes – overall survival, time to next treatment and time to treatment discontinuation	Time to next treatment (TTNT) as a proxy for Progression free survival (PFS) FS and Overall survival (OS)
List the confounding domains relevant to all or most studies		**ERG assessment**
1. dMMR status 2. Race/ethnicity (Black, Others, Unknown vs. White) 3. Age category (≥65 years vs. <65 years) 4. ECOG status at treatment initiation (1 vs. 0) 5. Histology at initial diagnosis (Non-endometrioid, Unknown vs. Endometrioid) 6. Federation of Gynecology and Obstetrics (FIGO) Stage (Stage III/IV vs. Stage I/II) 7. Grade of disease at diagnosis (Grade 3/4, Unknown vs. Grade 1/2) 8. Number of prior platinum-based therapies (0 or 1 vs. ≥2) 9. Prior surgery for study indication (Yes vs. No)		1. Race/ethnicity (black, others, unknown versus white) 2. Age category (≥65 years versus <65 years) 3. ECOG PS status at treatment initiation (1 versus 0) 4. Histology at initial diagnosis (non-endometrioid, unknown versus endometrioid) 5. FIGO stage at initial diagnosis (Stage III/IV versus Stage I/II) 6. Grade of disease at diagnosis (Grade 3/4, unknown versus Grade 1/2) 7. Number of prior platinum-based therapies (0 or 1 versus ≥2) 8. Prior surgery for study indication (yes versus no) 9. dMMR/MSHI status
List co-interventions that could be different between intervention groups that could impact on outcomes		**ERG assessment**

• The systemic anti-cancer therapy (SACT) database collects data on systemic anti-cancer therapies only. No other pharmacological interventions would be captured within the study, which could impact on outcomes. • The study would also capture surgery and radiotherapy interventions.	• No co-intervention recorded in the patient-level UK health data available through the NCRAS where the UK RWE information was obtained.

ROBINS-I tool (Stage II): For each study		
Specify a target randomised trial specific to the study		**ERG assessment**
Design	Individually randomised – the trial would be designed as per the GARNET trial cohort 2A	Individually randomized study design
Participants	English patients diagnosed with advanced or recurrent endometrial cancer who have progressed on or after first-line platinum doublet therapy, specifically a GARNET trial-like cohort i.e. application of the inclusion and exclusion criteria as per the GARNET trial	England residents with primary diagnosis of advanced or recurrent endometrial cancer who must have received exactly one prior platinum doublet therapy for recurrent or advanced disease
Experimental intervention	Current UK treatment paradigms, as a basket of treatments and for each individual relevant treatment, in the line directly post-platinum	Basket of common chemotherapy regimens.
Comparator	Placebo	Placebo
Is your aim for this study…?		
To assess the effect of *assignment to* intervention	Yes	Yes
To assess the effect of *starting and adhering to* intervention	No	No
Specify the outcome *Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.*		**ERG assessment**
• Progression free survival (PFS) • Overall survival (OS) • These are a proposed benefit of the intervention		• Progression free survival (PFS) • Overall survival (OS) • Proposed benefit of the intervention
Specify the numerical result being assessed *In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.*		**ERG assessment**

• PFS - The time from date of first dose to the earlier date of assessment of progression or death by any cause in the absence of progression based on: (1) the time of first documentation of PD per RECIST v1.1 • OS - The time from date of first dose of study treatment to the date of death by any cause.	• PFS – time from the date of the first dose to the earlier date of assessment of disease progression or death by any cause in the absence of disease progression based on the time of first documentation of disease progression per RECIST v1.1. • OS - defined as the time from the date of the first dose of study treatment to the date of death by any cause.

ROBINS-I tool (Stage II): preliminary consideration of confounders

Confounding domain CS	Confounding domain ERG assessment	Measured variable(s) CS	Measured variable(s) ERG assessment	Is there evidence that controlling for this variable was unnecessary?* CS	Is there evidence that controlling for this variable was unnecessary?* ERG assessment	Is the confounding domain measured validly and reliably by this variable (or these variables)? CS	Is the confounding domain measured validly and reliably by this variable (or these variables)? ERG
dMMR status	dMMR/MSHI status	No – was not available in the data set	No - not reported	Influence of controlling for this variable was not explored in this descriptive study. Descriptive statistics on this variable were captured in the study	No - Not reported	No – dMMR/MSI-H biomarker data are available within the NCRD. Although the dMMR/MSI-H biomarker is not prognostic, not having complete information on this biomarke	No information

						r is a limitation.	
Race/ethnicity	Race/ethnicity	Yes - (Black, Others, Unknown vs. White)	Yes - Black, Others, Unknown vs. White	As above	No - CS section CS B.2.7.1; Appendix D.5.1	Yes	Yes – CS section CS B.2.7.1; Appendix D.5.1
Age	Age	Yes - (≥65 years vs. <65 years)	Yes - ≥65 years vs. <65 years	As above	Yes - CS section CS B.2.7.1; Appendix D.5.1	Yes	Yes - CS section CS B.2.7.1; Appendix D.5.1
ECOG status at treatment initiation	ECOG PS status at treatment initiation	Yes - (1 vs. 0)	Yes - 1 vs. 0	As above	Yes - CS section CS B.2.7.1; Appendix D.5.1	Partially - ECOG status is recorded at diagnosis in the database. This may introduce bias to the ECOG status of recurrent patients; the ECOG status recorded at stage I and II EC diagnosis may not represent the ECOG status experienced at recurrence.	No - CS section B.2.7.1; Appendix D.5.1. ECOG PS is recorded at registry diagnosis. This may not be appropriate for those with recurrent disease.
Histology at initial diagnosis	Histology at initial diagnosis	Yes - (Non-endomet	Yes - Non-endomet	As above	No - CS section CS B.2.7.1;	Yes	Yes - CS section CS

		rioid, Unknown vs. Endometrioid)	rioid, Unknown vs. Endometrioid		Appendix D.5.1		B.2.7.1; Appendix D.5.11
Federation of Gynecology and Obstetrics (FIGO) Stage	FIGO stage at initial diagnosis	Yes - (Stage III/IV vs. Stage I/II)	Yes - Stage III/IV vs. Stage I/II	As above	No - CS section CS B.2.7.1; Appendix D.5.1	Yes	Yes - CS section CS B.2.7.1; Appendix D.5.1
Grade of disease at diagnosis	Grade of disease at diagnosis	Yes - (Grade 3/4, Unknown vs. Grade 1/2)	Yes - Grade 3/4, Unknown vs. Grade 1/2	As above	Yes - CS section CS B.2.7.1; Appendix D.5.1	Yes	Yes - CS section CS B.2.7.1; Appendix D.5.1
Number of prior platinum-based therapies	Number of prior platinum-based therapies	Yes - (0 or 1 vs. ≥2)	Yes - 0 or 1 vs. ≥2	As above	No - CS section CS B.2.7.1; Appendix D.5.1	Yes	Yes - CS section CS B.2.7.1; Appendix D.5.1
Prior surgery for study indication	Prior surgery for study indication	Yes - (Yes vs. No)	Yes - Yes vs. No	As above	No - CS section CS B.2.7.1; Appendix D.5.1	Partially - Beyond the cancer registry, diagnosis and procedure recording in HES is poor for all but inpatient settings owing to limited clinician capacity. Accordingly, the reporting of factors dependent upon	No information

						hospital data, such as surgery, may select on more acute or serious healthcare events and fail to present a full picture of surgical treatment .	

ROBINS-I tool (Stage II): preliminary consideration of co-interventions

Co-intervention CS	Co-intervention ERG assessment	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? CS	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)? ERG assessment	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator (CS)	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator (ERG assessment)
Surgery	Not applicable	Influence of controlling for this intervention was not explored in this descriptive study. Descriptive statistics on this intervention were captured in the study	Not applicable	Favour experimental	No information
Radiotherapy					

ROBINS-I tool (Stage II): risk of bias assessment

Signalling questions	Description CS	Description ERG	Response options CS	Response options ERG
Bias due to confounding				
1.1 Is there potential for confounding of the effect of intervention in this study? **If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered **If Y/PY to 1.1:** determine whether there is a need to assess time-varying confounding:	• All the pre-intervention prognostic factors listed above could impact intervention received at start of follow up. • The study looks at a basket of chemotherapies, therefore the outcomes of patients will be captured regardless of intervention received. • The study only captures patient's post-platinum treatment, confounding factors could influence what patients received platinum treatment first line.	Only counts available (no adjustment for confounders).	Y	Y
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	Analysis of survival outcomes in the post-platinum setting were not split according to intervention received, all chemotherapies were included in the basket.	Analyses were not split according to intervention received; all chemotherapy treatments were grouped together in one basket.	N	N
1.3. Were intervention discontinuations or				Not applicable

switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)				
Questions relating to baseline confounding only				
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	This observational descriptive study did not control for any confounding prognostic factors, it described the prognostic factors within the cohort and captured the entire cohort's survival outcomes.	Descriptive statistics. The study did not control for any confounding factors.	N	N
1.5. **If <u>Y/PY</u> to 1.4:** Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?			Not applicable	Not applicable
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	This observational descriptive study did not control for any post-intervention variables, it described the entire cohort's survival outcomes.	Descriptive statistics. The study did not control for any confounding factors.	<u>N</u>	<u>N</u>
Questions relating to baseline and time-varying confounding				
1.7. Did the authors use an appropriate analysis method that	This observational descriptive study did not control for any	Descriptive statistics. The study did not	N	N

controlled for all the important confounding domains and for time-varying confounding?	confounding prognostic factors, it described the prognostic factors within the cohort and captured the entire cohort's survival outcomes.	control for any confounding factors.		
1.8. **If <u>Y/PY</u> to 1.7:** Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?				Not applicable
Risk of bias judgement				
Optional: What is the predicted direction of bias due to confounding?	This observational descriptive study did not control for any confounding prognostic factors. The study was designed to capture a real work UK advanced recurrent endometrial cancer population, adjusting for any prognostic variables within this cohort would decrease the generalizability of the cohort to a typical UK cohort. An indirect treatment comparison using matched adjusted indirect comparison methodology has been used to control for confounding, when comparing the outcomes described in this study versus the outcomes observed for patients treated with dostarlimab in the GARNET trial.	Descriptive statistics. The study did not control for any confounding factors.	No information Unpredictable	No information Unpredictable
Bias in selection of participants into the study				

2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? **If N/PN to 2.1:** go to 2.4	All patients with survival outcomes data who received a treatment in the line directly post-platinum were included in the study cohort. Survival outcomes were tracked from the chemotherapy given directly post-platinum. No patient characteristics observed after the start of the intervention affected patient selection.	CS section B.3.2	N	N
2.2. **If Y/PY to 2.1:** Were the post-intervention variables that influenced selection likely to be associated with intervention?				Not applicable
2.3 **If Y/PY to 2.2:** Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?				Not applicable
2.4. Do start of follow-up and start of intervention coincide for most participants?	Start of follow up for the patient begin at entry into the NCRAS database, based on the date of endometrial cancer diagnosis; therefore, in advance of start of intervention for participants.	CS section B.3.2	Y	Y
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4:** Were adjustment techniques used that are likely to correct for the presence of selection biases?				Not applicable
Risk of bias judgement				

Optional: What is the predicted direction of bias due to selection of participants into the study?	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Low	Low
Bias in classification of interventions				
3.1 Were intervention groups clearly defined?	SACT contains detailed systemic treatment data for patients treated or funded by the National Health Service (NHS). All treatments captured in SACT aligned to this patient population and tumour of interest were included.	CS section B.3.2	Y	Probably Yes
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Yes. All SACT therapies were to be included.	CS section B.3.2	Y	Y
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	No. All interventions are included within the SACT database, separate to information regarding outcomes.	All of the basket of chemotherapy recorded were obtained independent of the pre-defined outcomes.	N	N
Risk of bias judgement				
Optional: What is the predicted direction of bias due to classification of interventions?	Treatments in primary care are not included. As such, some oral and hormone therapies may be underreported, as is perhaps evident in the near total absence of hormone therapy delivery identified for the GARNET-like population Favours experimental / Favours comparator	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Moderate	Moderate

	/ Towards null /Away from null / Unpredictable			
Bias due to deviations from intended interventions				
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2				
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	All patients captured within the post-platinum patient cohort were required to receive a post-platinum treatment as recorded in SACT.	There is insufficient information on the administration of the basket of therapies.	N	No information
4.2. **If Y/PY to 4.1:** Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?				Not Applicable
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6				
4.3. Were important co-interventions balanced across intervention groups?	Co-interventions, surgery and radiotherapy, was captured as a descriptive statistic for the entire patient cohort. The co-interventions are therefore used by some participants in the cohort and not others.		N	Not Applicable
4.4. Was the intervention implemented successfully for most participants?	All patients captured within the post-platinum patient cohort were required to receive a post-platinum treatment, for any duration, as recorded in SACT. Time on treatment was recorded also.		Y	Not Applicable
4.5. Did study participants adhere to the assigned intervention regimen?	All patients captured within the post-platinum patient cohort were required to receive a post-		Y	Not applicable

	platinum treatment, for any duration, as recorded in SACT. Time on treatment was recorded also.			
4.6. **If N/PN to 4.3, 4.4 or 4.5:** Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?				Not applicable
Risk of bias judgement				
Optional: What is the predicted direction of bias due to deviations from the intended interventions?	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Moderate	No information
Bias due to missing data				
5.1 Were outcome data available for all, or nearly all, participants?	Although data completeness is high for most core items available within the NCRD, staging data were absent for around 9% of the 45,494 EC patients diagnosed between 2013 and 2018. Given that staging information was central to the derivation of the advanced or recurrent disease cohort, these patients could not be included. It is unlikely that tumour staging is missing completely at random, thereby introducing some degree of selection bias; missing staging data will typically relate to older patients with advanced disease and short survival from diagnosis, such that	CS section B. 2.4 Grade of disease at diagnosis and ECOG PS status were not reported for a substantial number of patients.	N	N

	pathology was never completed. ECOG status was not recorded for a large number of patients. Scenario analysis was completed to include patients ECOG≤1 only (to match the GARNET trial criteria) or include patients ECOG≤1 and not recorded patients.			
5.2 Were participants excluded due to missing data on intervention status?	Intervention status was available for all patients; SACT collect all data for intravenous chemotherapies administered in the NHS.	Intervention status was reported for all patients.	N	N
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Patients were excluded due to lack of staging data.	CS section B. 2.4 Participants were excluded based on no recorded stage at diagnosis.	Y	Y
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3:** Are the proportion of participants and reasons for missing data similar across interventions?			NA	Not applicable
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3:** Is there evidence that results were robust to the presence of missing data?	Scenario analysis was completed to include patients ECOG≤1 only (to match the GARNET trial criteria) or include patients ECOG≤1 and not recorded patients; survival outcomes form both groups were similar. The impact of including patients with missing staging data was not explored.	CS section B.2.7 and Appendix D.5.1 - scenario analyses	PN	Probably Yes

Risk of bias judgement				
Optional: What is the predicted direction of bias due to missing data?	The exclusion of patients with missing staging data was required to align the study cohort with the GARNET trial cohort, where disease stage was an inclusion criteria. Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	It is uncertain if the scenario analyses have removed the risk of bias arising from the missing data.	Moderate	Moderate Unpredictable
Bias in measurement of outcomes				
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Outcomes were mortality and time to next treatment as recorded in the databased, which would not be influenced by knowledge of intervention.	Outcome measures were retrieved as recorded in the database.	N	N
6.2 Were outcome assessors aware of the intervention received by study participants?	Outcomes were assessed using time to event data from the NCRAS data base.	Yes, no blinding	N	Y
6.3 Were the methods of outcome assessment comparable across intervention groups?	Outcomes were assessed using time to event data from the NCRAS data base for all patients.		Y	Not applicable
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Outcomes were assessed using time to event data from the NCRAS data base for all patients.	There may be notable errors in measurement Progression free survival (PFS). CS section 2.7.2 “progression is not recorded within the NCRAS database, time to next therapy	N	Y

		(TTNT) was used as a proxy for PFS. TTNT was defined as the time from the start of line of therapy until failure (the earliest of all-cause death or the start of a new line of treatment). Patients lost to follow-up or still in same line of treatment at the end of the study period were censored." In established literature, PFS is often defined as the time from the date of the first dose to the earlier date of assessment of disease progression or death per RECIST v1.1.		
Risk of bias judgement				
Optional: What is the predicted direction of bias due to measurement of outcomes?	There is an absence of routine data concerning progression, remission or recurrence within the cancer registry. Accordingly, there is a need to use proxy measures (e.g. TTNT for disease progression). The reliability of results from such approaches will be dependent on their validity. TTNT may overestimate the time to progression.	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Low	Moderate

	TTD has been captured in the study and could be used as an alternative, lower bout, time to progression survival outcome. Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable			
Bias in selection of the reported result				
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome *measurements* within the outcome domain?	No. Outcomes were assessed using time to event data from the NCRAS data base.		N	N
7.2 ... multiple *analyses* of the intervention-outcome relationship?	No. Outcomes were assessed using time to event data from the NCRAS data base.	Descriptive statistics. The study did not control for any confounding factors.	N	N
7.3 ... different *subgroups*?	The study captures a broad UK advanced/recurrent endometrial cancer population, with a wide range of patient characteristics. Specific subgroups within this population would have different outcomes when treated with the intervention.	The results are from a large cohort available from a national database. The results may be different if specific chemotherapies were analysed within the basket of chemotherapy or if patients had received more than one line of prior platinum doublet therapy (but having platinum doublet therapy as the last line)	Y	Y

Risk of bias judgement				
Optional: What is the predicted direction of bias due to selection of the reported result?	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Moderate	Moderate
Overall bias				
Risk of bias judgement				
Optional: What is the overall predicted direction of bias for this outcome?	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable	Moderate	Moderate

Abbreviations: EC: endometrial cancer; ECOG: Eastern Cooperative Oncology Group; N: no; NCRAS: National Cancer Registration and Analysis Service; NCRD: National Cancer Registry Dataset; NHS: National Health Service; PN: partial no; PY: partial yes; SACT: Systemic Anti-Cancer Therapy; TTD: time-to-discontinuation; TTNT: time-to-next treatment; Y: yes.

Comparison of the ERG and company quality assessments using the company's preferred tools

A comparison of ERG and company appraisal of study quality for ZoptEC and McMeekin *et al* (2015), using the Appendix C of PMG6 methodology checklist for randomised controlled trials in the old NICE guidelines manual for RCTs is provided in ERG report Appendix Table 58 and Figure 39. For ZoptEC, overall, the ERG agreed with the company on a "no" rating for 4/14 (28.6%) of items, all domains combined (selection, performance, attrition, and detection risk of bias); thus, a low or unclear risk of bias was reported for most of the domains. The ERG agreed with the company on 2/4 (50%) applicable risk of bias domains – with "low risk of bias" ratings. These domains were related to selection and performance bias. For McMeekin *et al* (2015), overall, the ERG agreed with the company on a "no" rating for 4/14 (28.6%) of items, all domains combined (selection, performance, attrition, and detection risk of bias); thus, a low or unclear risk of bias was reported for most of the domains. The ERG agreed with the company on 1/4 (25%) applicable risk of bias domains – with an "unclear risk of bias" rating. The domain was related to selection bias. The ERG and company quality assessment for ZoptEC is more comparable than that to McMeekin *et al* (2015).

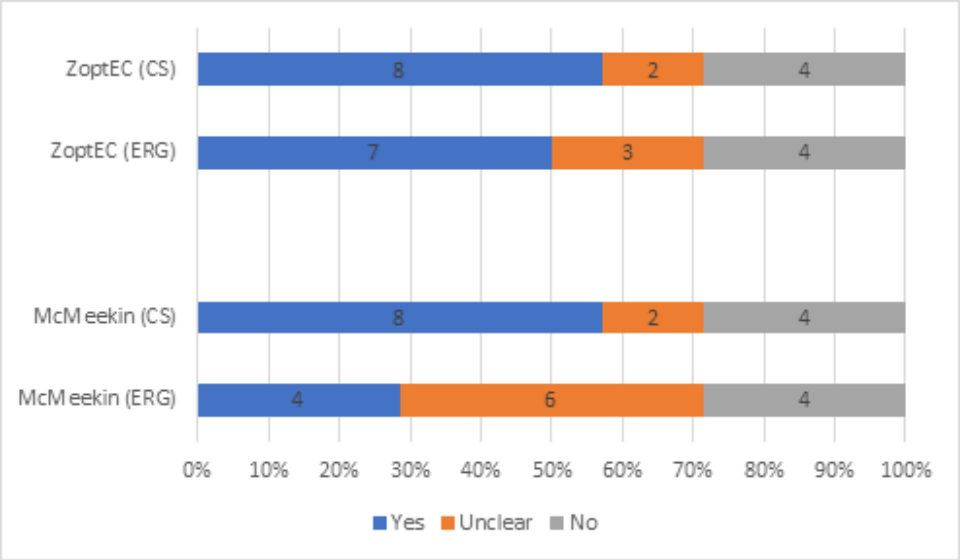


Figure 39: Comparison of ERG and company appraisal of RCTs Appendix C of PMG6 methodology checklist for randomised controlled trials in the old NICE guidelines manual

A comparison of ERG and company appraisal of study quality for GARNET, Rubinstein *et al*. (2019), Mazgani *et al*. (2008), Julius *et al*. (2013), and Makker *et al*. (2013), using the CASP check list for Non-RCTs is provided in ERG report Appendix Table 59 and Figure 40. A 'no' rating on the checklist was reported as a high risk of bias, and a 'yes' was reported as a low risk of bias. There were differences between the ERG and company judgements for overall risk of bias in most of the studies, except Julius *et al*. (2013). The ERG noted an overall moderate risk of bias, while the company noted a low risk of bias for the GARNET trial. For Rubinstein *et al*. (2019), Mazgani *et al*. (2008), and Makker *et al*. (2013), the ERG noted an overall moderate risk of bias, while the company noted a high risk of bias. For Julius *et al*. (2013), the ERG agrees with the company's judgment of an overall high risk of bias.

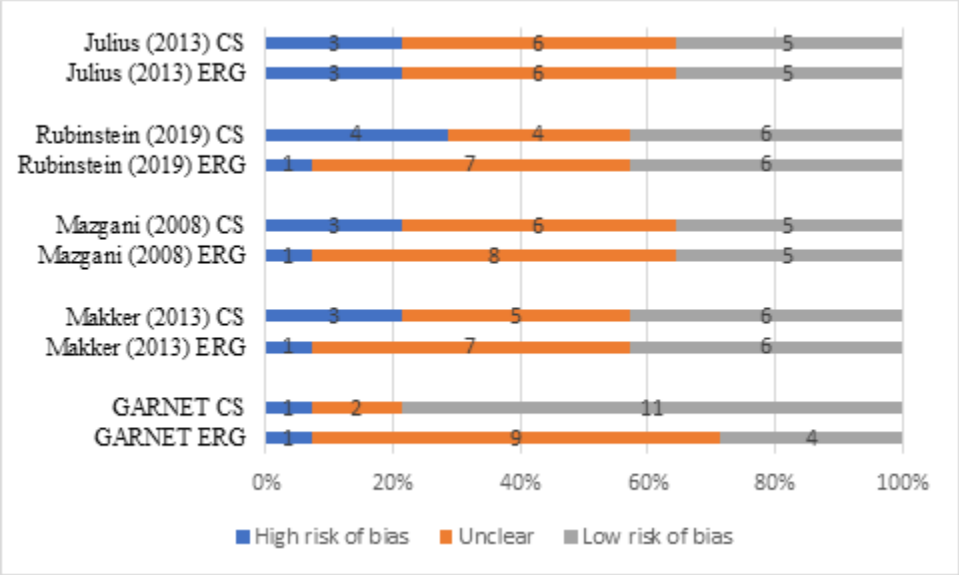


Figure 40: Comparison of ERG and company appraisal of Non-RCTs using the CASP check list

A comparison of ERG and company appraisal of the study quality for the UK RWE study, using the ROBINS I assessment tool is provided in ERG report Appendix Table 60 and Figure 41. The ROBINS I tool evaluates the risk of bias using seven domains (including: confounding, participants selection, the classification of intervention, deviation of intervention, missing data, outcome measurements, and bias in the selection of the reported results). Concerning the bias due to the confounding, participants selection, the classification of intervention, missing data, and bias in the selection of the reported results, the ERG agrees with the company's judgments. For bias due to deviations from intended interventions, the ERG and the company's judgements differ on all items. The ERG notes that there was insufficient information on the administration of the basket of therapies. It is therefore unknown if any deviations would lead to bias in the effect estimate. For bias due to measurement of outcomes, there was a difference in 3/4 (75%) of the items. The ERG agrees with the company with an overall "moderate risk of bias".

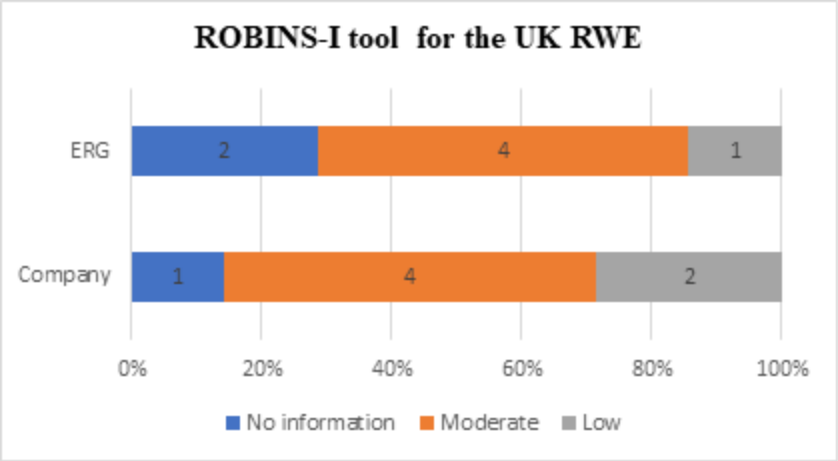


Figure 41: Comparison of ERG and company appraisal of included studies using the ROBINS I assessment tool for the UK RWE study

ERG quality assessments using the NICE preferred tools

For the more applicable Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) checklist, the overall risk of bias for studies with low risk of bias in all domains was judged as “low risk of bias”, while some concerns in multiple domains or a high risk of bias in at least one domain was judged as “high risk of bias”. Both the ZoptEC trial and McMeekin *et al* (2015) study had an overall judgment of “high risk of bias”. A summary of results is presented in Figure 42.

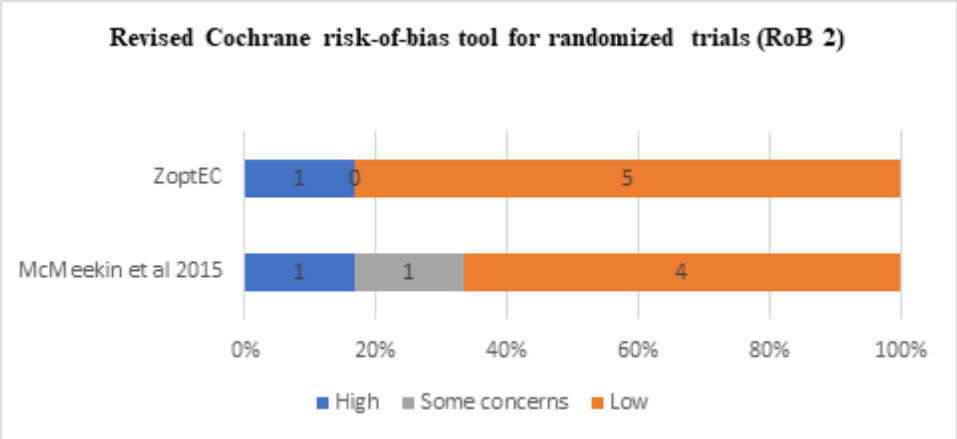


Figure 42: ERG appraisal of included studies using the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) Appraisal checklist

For the Institute of Health Economics Quality Appraisal Checklist for Non-RCTs, a ‘no’ rating on the checklist was reported as a high risk of bias, and a ‘yes’ was reported as a low risk of bias. A summary of results is presented in Figure 43. Most of the studies had

low or partial risks of bias. The ERG considers the GARNET trial of better quality than the primary comparator evidence (UK RWE study), GARNET had about 90% of the items rated as low or partial/unclear risks of bias, while the UK RWE study had about 65% of the items rated as low or partial/unclear risks of bias.

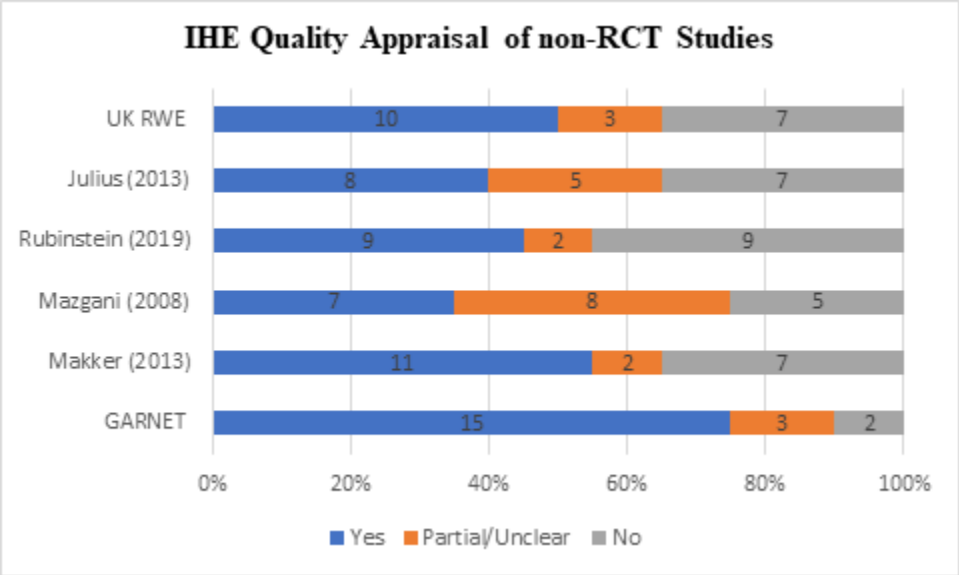
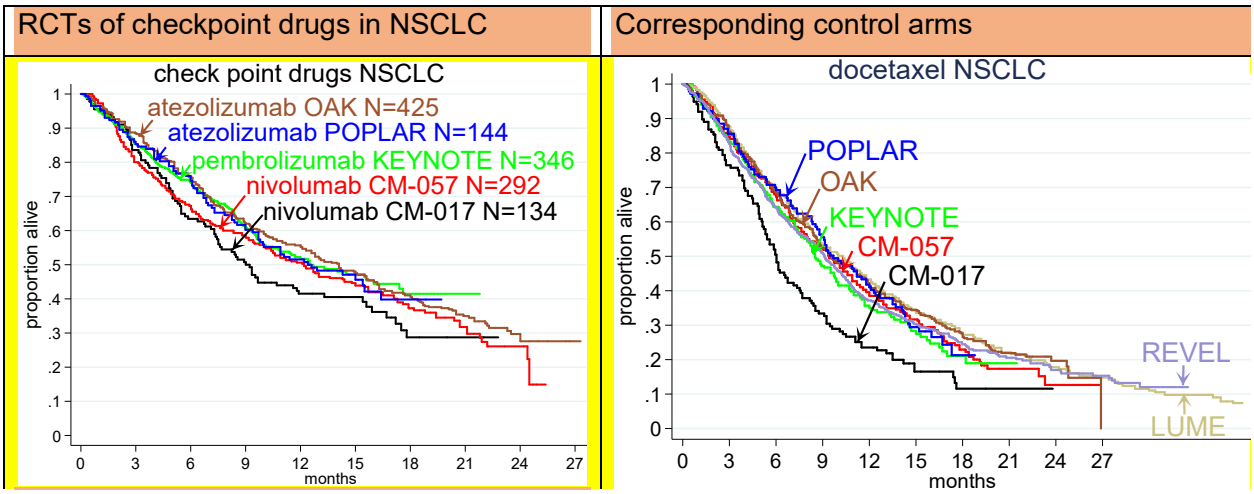


Figure 43: ERG appraisal of included studies using the IHE Quality Appraisal Checklist

9.2 *Trajectories of KM OS curves from trials of checkpoint inhibitors for treating non-small cell lung cancer (NLSCLC)*

The ERG looked at the trajectory of KM OS plots in RCTs of checkpoint drugs in non-small cell lung cancer (NLSCLC).[34-36] The trajectories exhibit a gradually decreasing slope without the pronounced long flat tail seen in the GARNET single arm study; a similar trajectory is seen in the large recent gastro-oesophageal cancer CHECKMATE 649 RCT[37] and in the NICE STA ID 1019 for pembrolizumab in previously treated advanced / metastatic urethral cancer (based on a single arm study). The control arms in such RCTs show a similar trajectories. The ERG consider it likely the pronounced flat tail in GARNET is contributed by small patients numbers and immature follow up more than by extended treatment effects.



CM=CheckMate

Figure 44: Trajectory of KM plots for patients in trials of checkpoint inhibitors for non-small cell lung cancer (NSCLC)

9.3 *ERG critique of company's approaches to OS and PFS modelling with extrapolation to 40 years*

The company's approach, described on CS page 137, entailed the following elements:

- Assessment of proportional hazard assumption (dostarlimab vs. RWE)
- Use of information criteria to judge goodness of fit of parametric models
- Visual inspection of extrapolated parametric curves versus observed KM curves
- Clinical plausibility of short and long term survival estimates based on discussion with and survey of UK clinical experts opinions

The observed OS and PFS KM plots for dostarlimab (GARNET: CS Figures 11 and 13) are characterised by changes in trajectory of the curve (especially for OS) and long flat tails from about 18 months to 32 months during which few patients were at risk and there was a sparsity of events (Figure 45). Most parametric models are unlikely to fit well to changing trajectory in the observed data and for some models the flat tail is likely to strongly weight extrapolations extending to 40 years (Figure 45). These features may be contingent on the small number (N=129) and possible heterogeneity of patients and immaturity of observation.

9.3.1 OS dostarlimab (GARNET)

The company rejected the assumption of proportional hazards and explored a complement of nine parametric models extrapolated to 40 years. The ERG has reservations about the extended time horizon and the potential influence on modelling of the flat tail in the KM data. On extrapolated to 40 years the CS parametric models other than exponential, Weibull, and gamma provided implausibly generous survival predictions with significant survivors well beyond 40 years (Figure 46). It can be noted that at 5 years the ggamma model generates easily the best survival of eight models other than the nearest rival (Gompertz model) that predicts about 40% patients as immortal.

[redacted]

45 [redacted]

[redacted]

[redacted] 46

On the basis AIC/BIC scores (CS Table 55) the company selected the ggamma model as best fit. This model generates clearly implausible >20% survivors after 40 years and about 18% after 55 years. The CS justifies the choice of the ggamma model by pointing to the correspondence between the treatment-waning adjusted ggamma model and the mean of seven expert clinicians' opinions about survival at 3, 5, 10,15 and 20 years; to the ERG this seems to be a teleological construction. The ERG consider that a more plausible parametric model would be better selected before any waning adjustment is applied, and point out that the mean of seven clinicians' predictions ignores the range inherent in clinicians' opinions and also the uncertainty associated with the estimation of a proportion. The ERG think that for an average clinician value to be useful a survey involving a larger number of experts may be required. Figure 47 summarises the individual clinician predictions at 3, 5, 10, 15 and 20 years (based on CS Table 56); these are predicted proportions for 129 patients and the ERG have attached binomial 95% CIs. Considerable variation is evident. It is unclear to the ERG if clinicians were appraised or not appraised of the possibility of waning when making their estimates. As an approximation of full variation associated with the predictions the ERG takes the range from the lowest 95% CI to the highest 95% CI at each of the years predicted. These ranges are represented as vertical bars in Figure 46 (for unadjusted OS) and in Figure 48 (for waning-adjusted OS).

Figure 46 indicates that all models (unadjusted for treatment-waning effect) are encompassed within clinicians' range of predictions (the only exception being the Gompertz model at 20 years). Figure 48 indicates that this is still the case at most years after waning adjustment of most models. The ERG suggest that the clinical predictions may be associated with too much uncertainty to strongly support any particular choice of parametric model.

47

[redacted]

48

The company's treatment waning-adjusted ggamma model predicts an implausible 4% survivors at 40 years; however in the CS economic model this treatment-waning adjusted ggamma model is further adjusted by "capping" from 20 to 40 years so that survival does not exceed that for a matched UK general population. That capping is required from 20 years onward implies a time horizon of 40 years might be too extended. Two STAs of PD1 drugs quoted by the company employ shorter time horizons of 20 years[38] and 25 years;[39] a 40 year horizon was used in TA578,[40] but sensitivity analysis with shorter time horizons increased the ICER substantially. At 20 years the waning adjusted ggamma model suggests about [redacted] of patients are cured of endometrial cancer and will suffer the same mortality from other causes (other cancers, heart disease etc) as the matched general population.

9.3.1.1 Influence upon parametric models of the flat tail in the observed data

Some of the CS parametric models extrapolation to 40 years may be sensitive to the flat tail seen in the KM plots. To monitor this potential influence, particularly in regard to the CS-selected ggamma model, the ERG split the KM plot at various time points so as to reduce the size of the flat tail; the "reduced data" was then modelled using standard parametric models. The data was split at 14.6, 18.5 and 20.64 months and compared with models using the complete KM plot (no split). The results (Figure 49) indicate that the Gompertz and ggamma models are sensitive to the extent of the flat tail and at each split time the ggamma models generate implausible proportions of survivors when extrapolated to 40 years.

9.3.1.2 Adjustment for treatment waning

The company's justification for applying treatment-waning is stated as follows: "*treatment waning assumptions were applied in line with UK clinical expert feedback and previous appraisals of I-O therapies*" (CS section B.3.3.4; page 137). Information supplied in clarification identified one of the questions to be posed for clinical experts as "[redacted]?" The ERG have been unable to identify the clinicians' quantitative responses among clinical responses supplied in clarification.

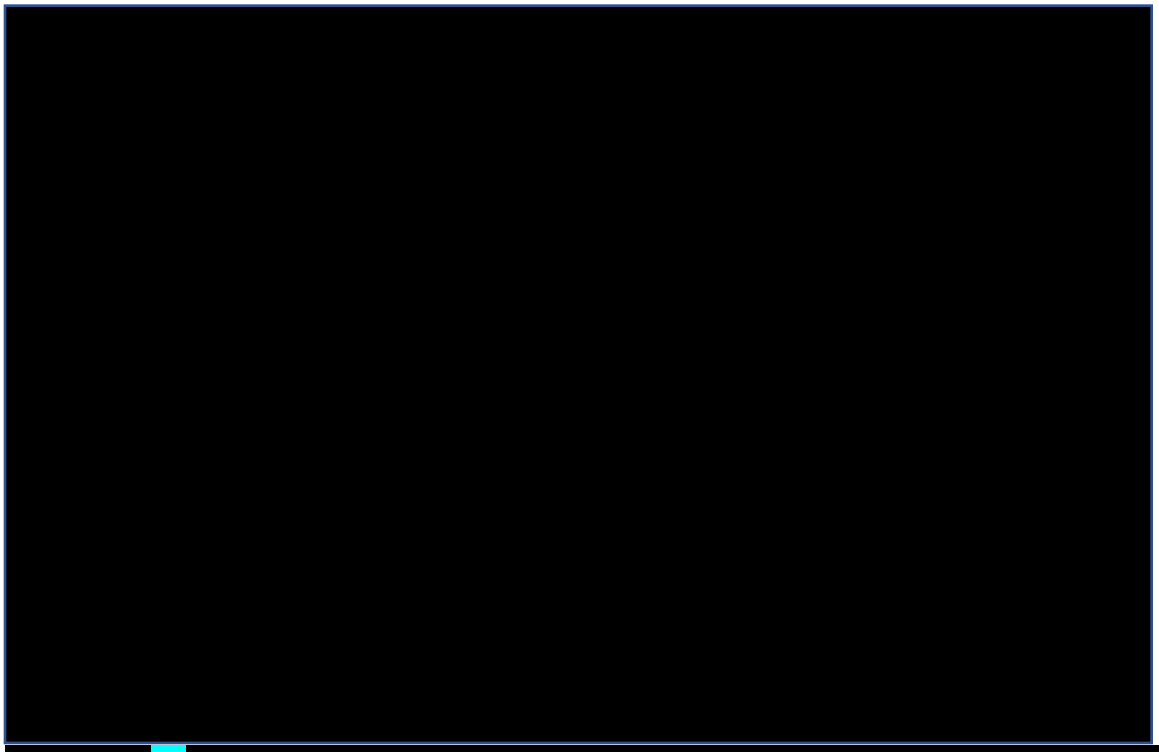
In the company base case waning adjustment the unadjusted ggamma model was used for the phase 0 to [redacted], the phase from [redacted] to 40 years was fully waning-adjusted so that "*efficacy associated with dostarlimab was assumed to be equal to the efficacy associated with current clinical management*", this was achieved by applying the MAIC HR of 0.35 (95% CIs: 0.22 - 0.55; CS Table 24) from [redacted] onward; for

the phase between [redacted] the waning effect gradually changed from zero to full waning (a linear change in hazard).

[redacted] shows the hazards (left) and OS (right) for the company's treatment-waning adjusted and unadjusted ggamma models.

Figure 50 right shows that the company's waning adjustment exerts a large influence on ggamma modelled OS, reducing the predicted proportion alive at 40 years from >20% to ~ 4%. Using alternative HR values (from within the MAIC 95% CIs) of 0.25 and 0.5 (rather than 0.35) indicates considerable sensitivity of the ggamma model to the HR applied. Even with HR of 0.25 there remain predicted survivors beyond 40 years.

Corresponding results for the company's exponential and Weibull models are summarised in Figure 51 and Figure 52 (see Appendix 9.6 for waning-adjusted hazards for all CS models). The influence of waning on exponential and Weibull modelled OS seems muted relative to that seen for the ggamma model.



51

52

The company's unadjusted spline models generate survivors beyond 40 years but with waning adjustment survivors at 20 years are reduced to <1% producing curves very similar to the unadjusted exponential and Weibull models. Loglogistic and lognormal models do not support proportional hazards and using the MAIC HR to for these models does not seem appropriate.

The ERG consider unadjusted and adjusted exponential and Weibull models and possibly waning adjusted spline models represent more plausible extrapolated survival than the company's unadjusted or waning-adjusted ggamma models.

Table 61 summarises ERG comments regarding the steps taken by the company to justify its selection of the waning-adjusted ggamma model.

Table 61: Summary of company's selection procedure for waning-adjusted ggamma model of dostarlimab OS

Step	Company's modelling	ERG Comment 1	ERG Comment 2
1	Make selection of preferred model by comparing extrapolated models with the mean of clinicians' predictions of survival (at 3, 5,10,15, 20 years)	The mean of clinicians' predictions fails to reflect the wide variation between predictions of individual clinicians; when this variation is accounted for all models fall within range of clinicians' predictions.	The clinicians' predictions are too various to strongly indicate superiority of any model over an alternative. A survey of opinions of a larger number of experts would seem desirable.
2	Select ggamma model on basis that the waning-adjusted ggamma model conforms to the mean of clinicians' predictions	This seems teleological (selection to serve purpose). The selected model should conform to the mean of clinicians' predictions before waning-adjustment as well as after. The ggamma model requires treatment waning adjustment to conform to the mean of clinicians' predictions, but still generates implausible survivors at 40 years.	Only exponential Weibull and gamma models generate reasonable extrapolation without waning adjustment; other models predict survivors beyond 40 years. The unadjusted ggamma model generates very implausible extrapolation (>20% alive at 40 years). The waning-adjusted ggamma model generates more modestly implausible extrapolation (>4% survivors at 40 years).
3	The gamma model has reasonable AIC/BIC scores (rank1 on AIC; rank 3 on BIC)	The differences between models in IC score is fairly trivial for a KM curve with multiple changes in trajectory.	AIC/BIC scores can be influenced strongly by the long flat tail in the observed KM plot. This is seen particularly with CS Gompertz and ggamma models
4	The waning-adjusted ggamma model requires capping so that survival rate does not exceed that of the matched general population.	For the waning-adjusted ggamma model capping was required from year 20 to year 40.	The capping requirement implies that the waning-adjusted ggamma model may be over-generous in survivors upon extrapolation and that a time horizon of say 20 or 25 year used in other PD1 STAs may be appropriate.

9.3.1.3 OS Conclusion/summary

The company's parametric models either fit poorly to the observed data (according AIC/BIC scores and or visual inspection) or predict implausible survival in extrapolation with decreasing hazard to 40 years that seems inconsistent with an ageing population, likely due to the influence of the long flat tail in the observed data. The company considers that "*the extended tail of the KM curves (that) is the hallmark of I-O therapies" , and point out that "in other cancers, I-O therapies have been shown to result in extended treatment benefits and long-term remission even after treatment discontinuation, offering a substantially improved prognosis for many patients.* [41] *Indeed, the long-term benefits of I-O therapies have been demonstrated across multiple indications including melanoma, lung, head and neck, where patients who discontinued therapy had durable responses that extended beyond the end of treatment.* [42] *Given this trend, it is reasonable to believe some patients who respond to dostarlimab may continue to experience extended treatment benefits and long-term remission beyond the two-year follow-up in the GARNET trial to date*".

Summary of time to event evidence from GARNET

The single arm phase 2 GSK study GARNET provided time to event analysis evidence in the form of Kaplan Meier plots about overall survival, progression free survival and time on treatment for 129 patients treated with dostarlimab with maximum follow up of ~30 months. To varying degrees the plots exhibit multiple changes in trajectory and long flat tails where few events occur and few patients are at risk. The company's position is that the plots are typical of PD inhibitors and that they indicate particularly long term benefit in a sub population of good-responders. The ERG position is that these characteristics may depend on a too small and heterogeneous population being followed up for too short a time in the absence of a comparator. The company's position might be supported if these KM characteristics were uniquely and universally found for this class of drug in both endometrial and other cancers but were not seen with alternative therapies for endometrial cancer. The company has not presented data that support their position other than a few references and an interpretation of clinicians' opinions. The ERG has done a rapid analysis of some available relevant studies.

Further modelling in company submission

The company explored the use of observed KM data until the flat tail was reached, followed by parametric modelling thereafter. Unsurprisingly this modelling resulted in very unrealistic extrapolations due to the lack of events and flatness in the tail of the observed data, and was rejected by the company.

The ERG have explored the influence of the flat tail on extrapolation of parametric models by splitting the observed data at several time points in the flat tail and extrapolating thereby to generate models that encompass reduced influence of the flat tail (see Appendix section 9.3.1.1).

9.3.2 PFS dostarlimab (GARNET)

The company used the same procedures as for OS. Treatment waning was applied to parametric models of PFS; the ERG have not previously encountered such application to PFS and are unsure of the company's justification for doing so. The KM plot has an even more extensive flat tail than seen for OS; consequently no parametric models fit the KM well. Models with superior AIC/BIC aggregate scores generate extrapolations predicting that after about 14 months many un-progressed patients ([redacted]) will remain without progression to 40 years.

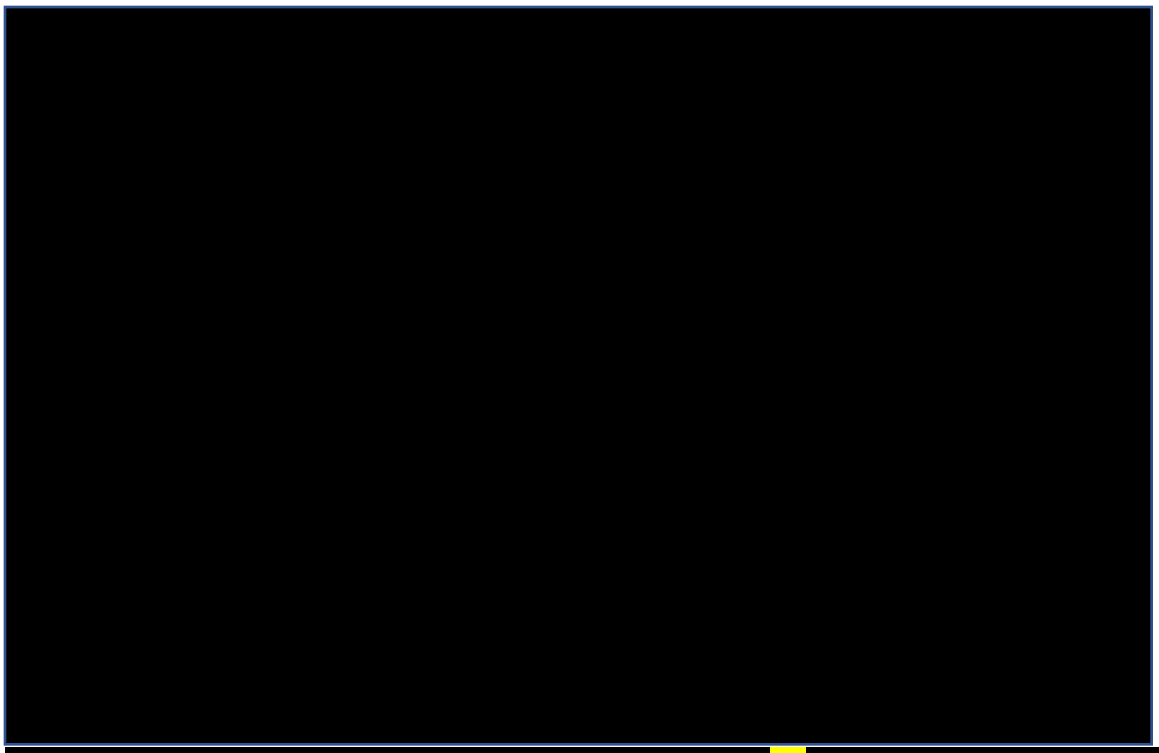
According to AIC/BIC values (CS Document B Table 50) the best ranking parametric fits were supplied by spline, ggamma and Gompertz models; however with or without application of treatment waning these generate unrealistic extrapolations to 40 years and were rejected by the company. As base case model the company selected the lognormal model. This generated more plausible extrapolation to 40 years but provided a poor fit to the PFS KM (CS Document B Figure 41) and tallied poorly with clinicians' predictions (Figure 53 and Figure 54). In choosing the lognormal model the company disregard the clinicians' PFS predictions stating "*however, based on plausibility considering the OS extrapolations, a more conservative survival curve, the lognormal, was identified for use in the base case*". The ERG find this teleological and do not consider this a sound argument since the plausibility of the company's ggamma model

for OS is far from obvious (see section above). The impact of selecting the lognormal model for PFS in conjunction with the ggamma model for OS is to greatly promote the accrual of post-progression survival benefit even though dostarlimab treatment has long ceased (see following section).

The PFS predictions of seven clinical experts were far less variable than for OS (Figure 55), with one respondent tending to be an outlier that influences mean values. The outlier predictions at 15 and 20 years are very different to those of the other six clinicians. ERG assessment of the variation in clinicians' predictions has discounted the outlier and, as for OS, has taken the range from lowest to highest 95% CI. These are plotted as vertical bars in Figure 53 and Figure 54.

The mean and range of clinicians' predictions seem somewhat unrealistic in that patients remain without progression after 15 to 20 years even though clinical opinion is that treatment would cease after [redacted]. The CS is inconsistent in the use and weight given to clinicians' predictions, accepting those for OS but rejecting those for PFS. As stated above and elsewhere the ERG find clinicians' predictions are associated with too much uncertainty to be used as a sound guide for modelling.

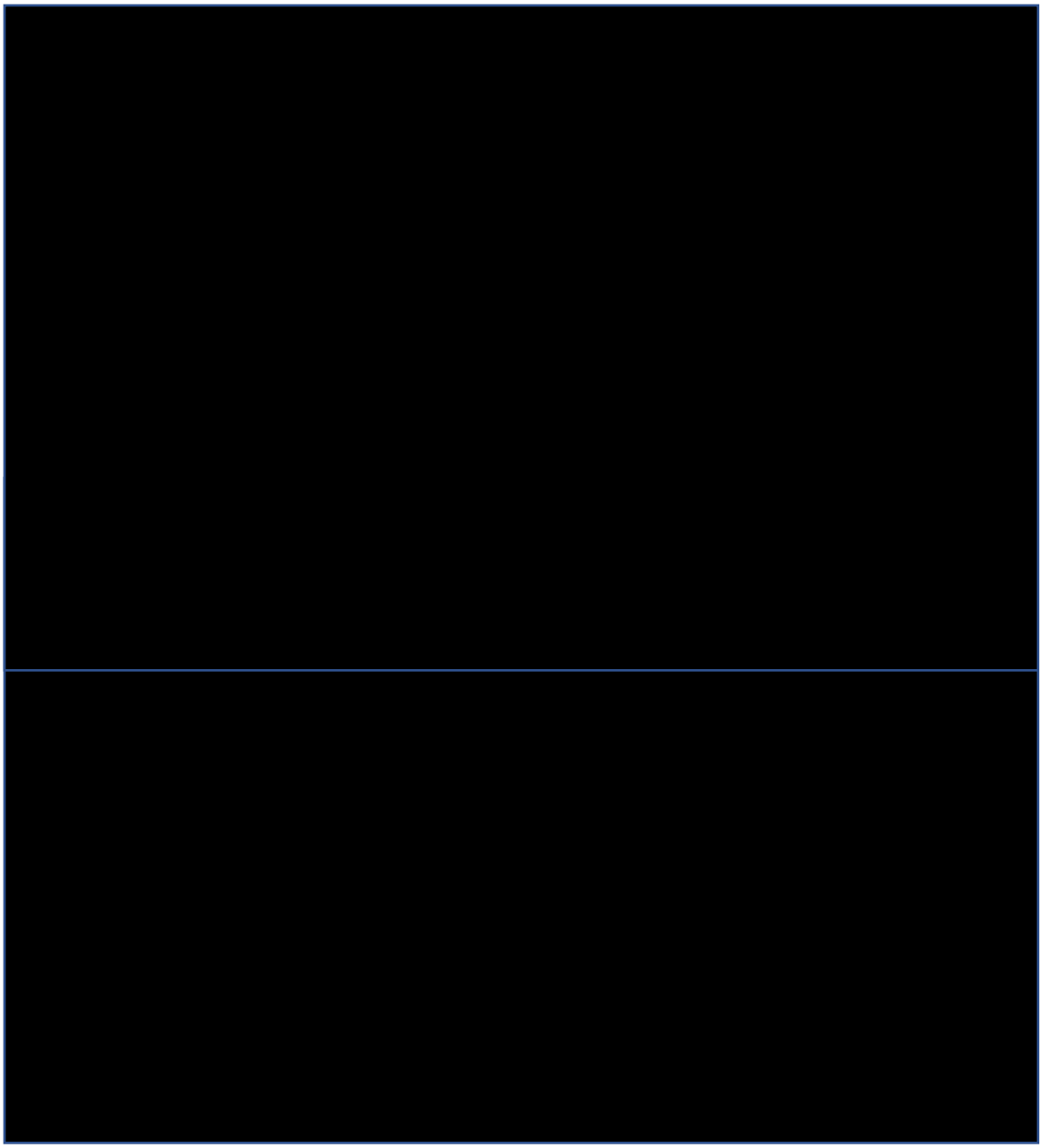
Because of the influence of the accentuated flat tail of the PFS KM plot and the seemingly optimistic clinicians' predictions of progression it is difficult to select a suitable parametric model and to decide if treatment waning represents a valid adjustment. The ERG explored additional models that might fit clinicians' predictions more consistently than seen in the CS. In particular bathtub and Rayleigh models of OS and PFS failed to generate superior models to those generated by the company.



53

54

55



9.3.2.1 Impact of CS selection ggamma and lognormal models for OS and PFS

Modelling PFS and OS partitions LY and QALY accrual between pre-progression benefit (estimated from the area under the PFS curve) and post-progression benefit (estimated from the area between OS and PFS curves). When time on treatment ([redacted]) is short compared to the modelled time horizon (40 years) accrual of pre-progression benefit is generally expected to be greater than that for post-progression benefit since any treatment effect will terminate and / or wane relatively early.

Figure 56 shows the accrual of LY benefit in pre-progression and post-progression during the KM phase of ~32 months (left) and during the CS models extrapolated to 20 years (right) and compared to expert clinicians' mean estimates. During the KM phase pre-progression gain (brown) is much larger than post-progression gain (green), whereas after extrapolation using the company's models of OS and PFS the reverse is the case (pale green area is much greater than pale brown area). Further extrapolation beyond twenty years (240 months) perpetuates this trend. This result is reflected in the output of the company's economic model where 61% of total life years for dostarlimab accrues in post-progression. In contrast to this the mean of clinician's predictions for OS and PFS implies that on average the clinician's do not think there would be post-progression gain after 120 months. The company base case selection of model for OS seems overgenerous relative to clinicians' opinion while in contrast the base case model for PFS greatly underestimates PFS relative to clinicians' opinion. This results from the company's inconsistent use of clinical opinion (see above).

9.4 *ERG alternative modelling of OS and PFS*

For reasons explained in previous sections the ERG think the base case models proposed by the company are likely inappropriate.

9.4.1 GARNET

The relatively small number of patients (N=129) in GARNET and the single arm nature of the GARNET study, together with the changes in the trajectory of the KM plot for OS, and the pronounced flat tails seen in both the KM plots for OS and PFS, means that any modelling for extrapolation will unavoidably be associated with considerable imprecision; this will also apply for the results of the MAIC analyses, that were undertaken by the company as supporting evidence, and in which the dostarlimab sample size was further reduced.

The ERG therefore explored several alternative modelling options that seem more appropriate in extrapolation than those selected by the company. In particular OS was

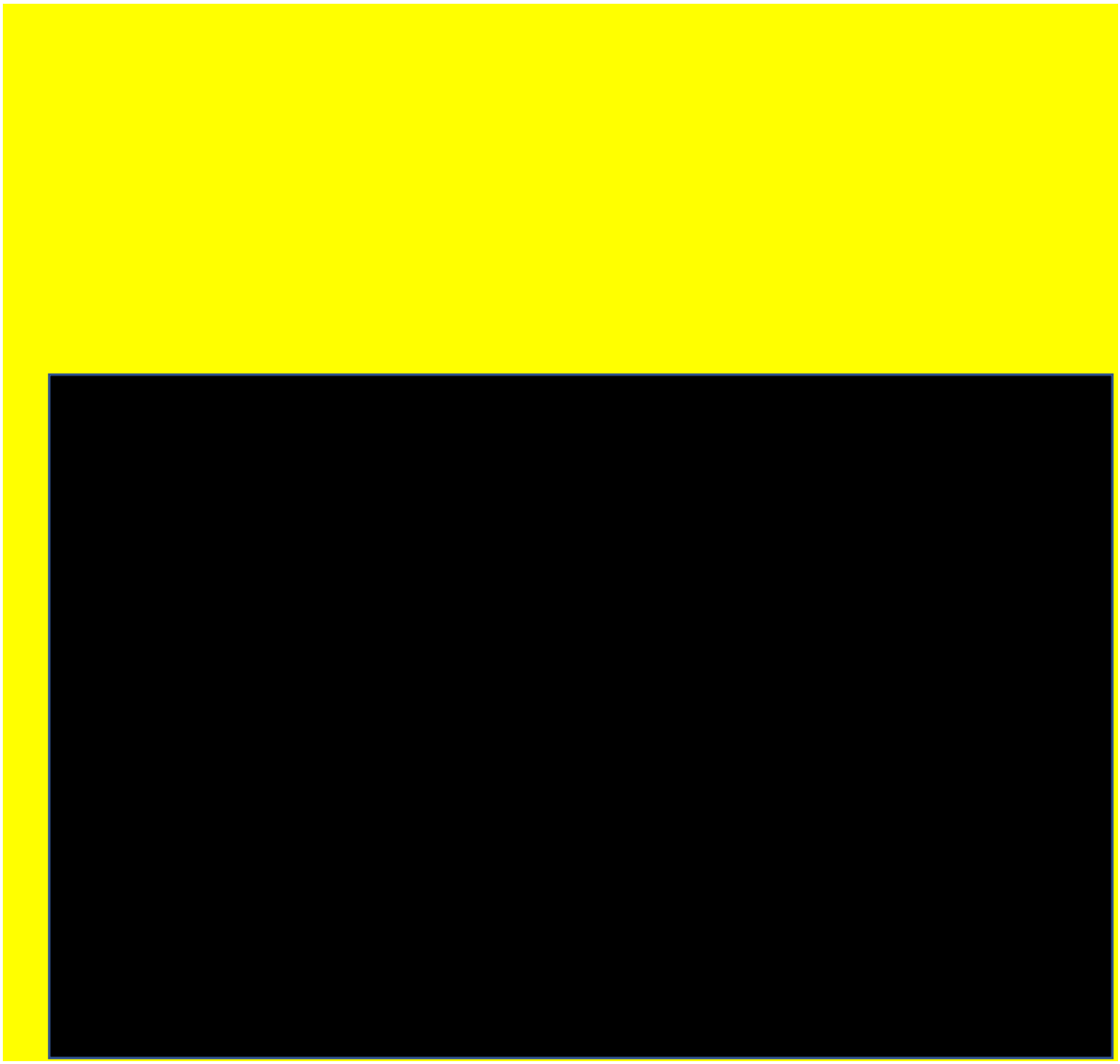
modelled with the treatment waning-adjusted Weibull distribution rather than the company's over-generous waning adjusted ggamma model, and PFS by adjusted and unadjusted Weibull models rather than the company's lognormal model. The results are summarised in Figure 57.

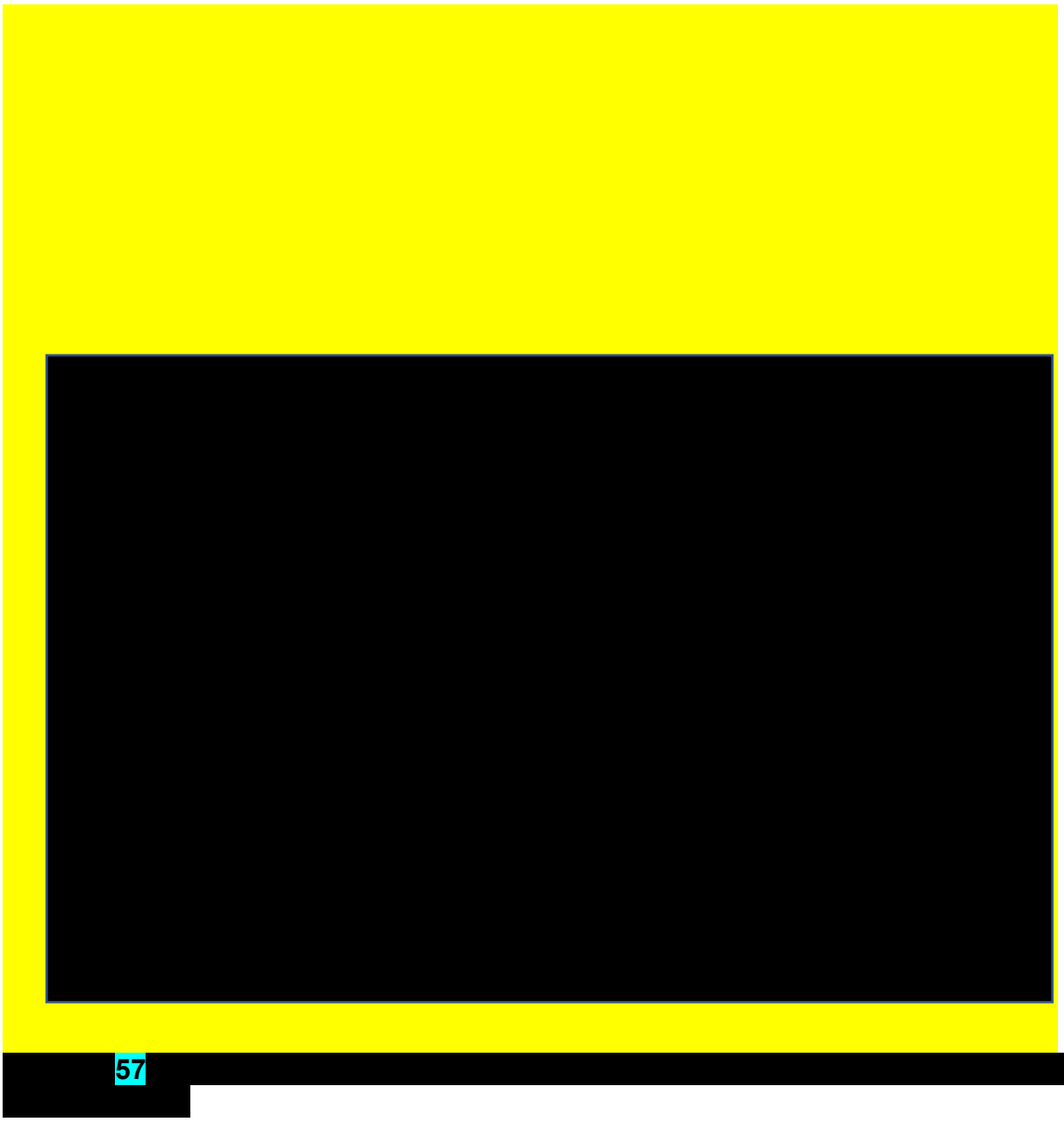
With the ERG models the area under the curve (AUC) estimates of pre-progression survival benefit accrual is greater than that for post-progression benefit. The unadjusted Weibull model for PFS requires capping to equal OS at a late stage of extrapolation so as to avoid predicting progression of dead patients. The adjusted PFS curve does not encounter this problem.

9.4.2 RWEQ

In contrast to the GARNET study the RWE KM data for PFS and for OS was mature (survival less than 15% at end of follow up), exhibited internally consistent trajectory, and was based on a large number of participants (N=[redacted]).

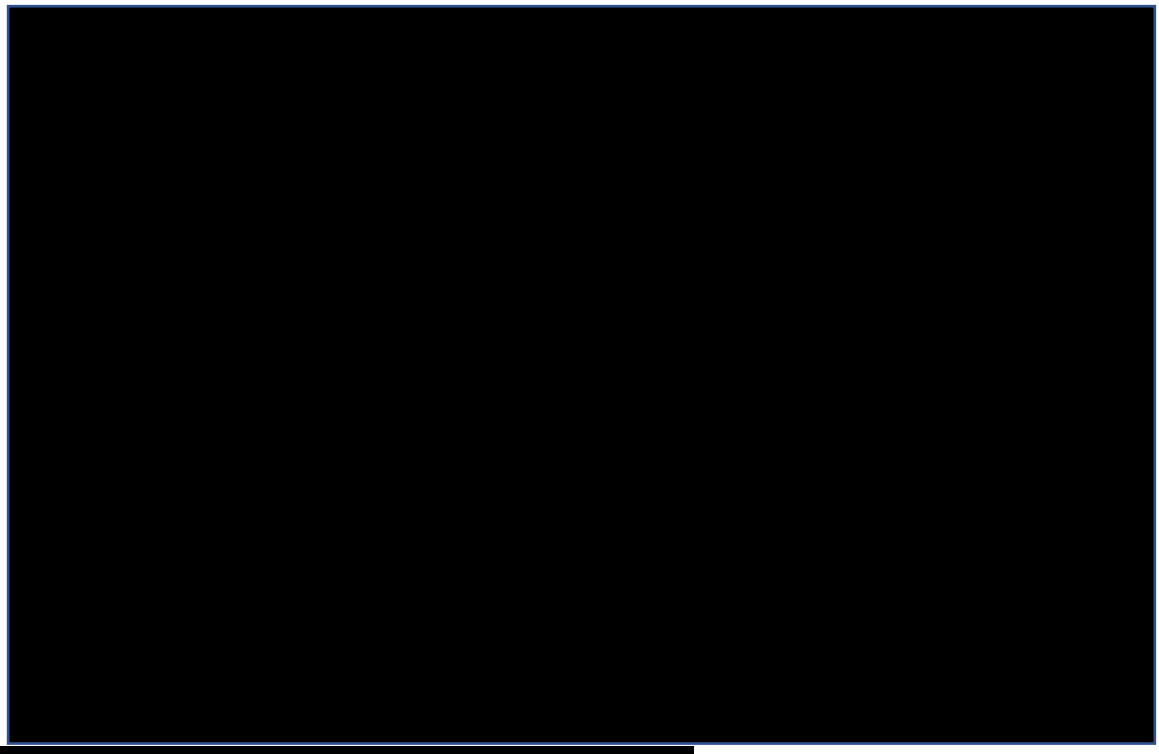
Figure 58 summarises observed OS and PFS (KM plots), the company's selected loglogistic (OS) and lognormal (PFS) models and clinicians' predicted PFS and OS at 5, 10, 15, and 20 years. The area under the curves allows estimation of accrual of observed pre-progression and post-progression LYs benefit. Model fit to KM OS and PFS is good, and clinicians' predictions, although slightly optimistic align well with both observed and modelled results. Gained LYs are more balance between pre-progression and post-progression than seen with the company's models for the population receiving dostarlimab. The ERG note that a model that more closely matches the clinicians' predictions would be more consistent with the company's position of clinician-led modelling.





57

[redacted]58[redacted]



[redacted]

The ERG therefore explored additional models using IPD developed from the KM details supplied by the company in clarification. A cubic spline model with 3 knots generated superior AIC values than loglogistic and in extrapolation more closely aligned with the clinicians' predictions than did the company's loglogistic model (Figure 59) but eventually flattens dramatically and seems less suitable than the CS loglog model.

Figure 59: RWEQ KM, loglogistic, and cubic spline models, compared to clinicians' predictions

9.5 *GARNET data on time on treatment with dostarlimab*

The company's model of time on dostarlimab treatment involved the following stages: [i] parametric models were fit to the observed KM plot for ToT; [ii] selection of the loglogistic model from among candidate parametric models (CS Figure 53); [iii] operation of the loglogistic model for [redacted] at which time point the proportion of patients continuing treatment was reduced to [redacted] of the starting population; [iv] continuation of the loglogistic model from [redacted] to [redacted] (reducing the proportion in treatment from [redacted] to about [redacted]); [v] discontinuation of treatment at [redacted]. The resulting model is shown in CS Figure 54. The justifications for this model were expert clinical opinion sought by GSK during a consultancy exercise said to support the [redacted] changes in the loglogistic model, the fact that this procedure had been

judged appropriate in the NICE appraisal of an analogous I-O therapy (avelumab) for Merkel cell carcinoma, and the appropriateness of the loglogistic model.

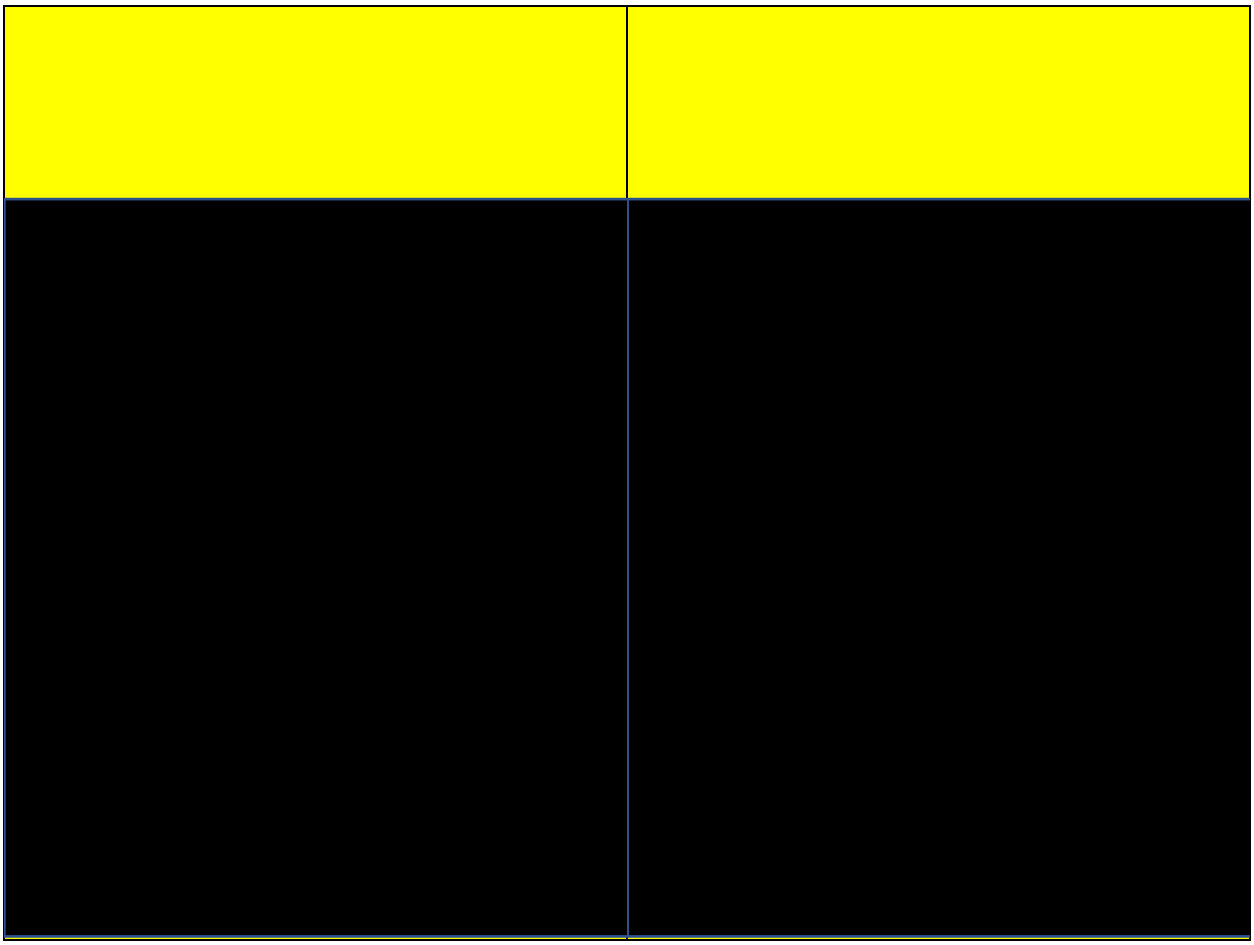
The ERG's critique of the company model includes the following points: the ERG found that the company model used the efficacy ToT KM but referred to this as ITT; since the observed data is only referenced for the first 2 years the ERG believe the parametric fit selected should be that which best fits the 0 to 2 year observed data; the amongst the clarification material supplied about the GSK clinical expert consultation the ERG failed to find supporting quantitative clinical expert opinion regarding the [redacted] cuts or the reduction to [redacted] introduced at [redacted]; the ToT model accepted by the avelumab appraisal committee included a 2 year reduction to 33% in treatment rather than to [redacted] These points are explained in more detail in the following section.

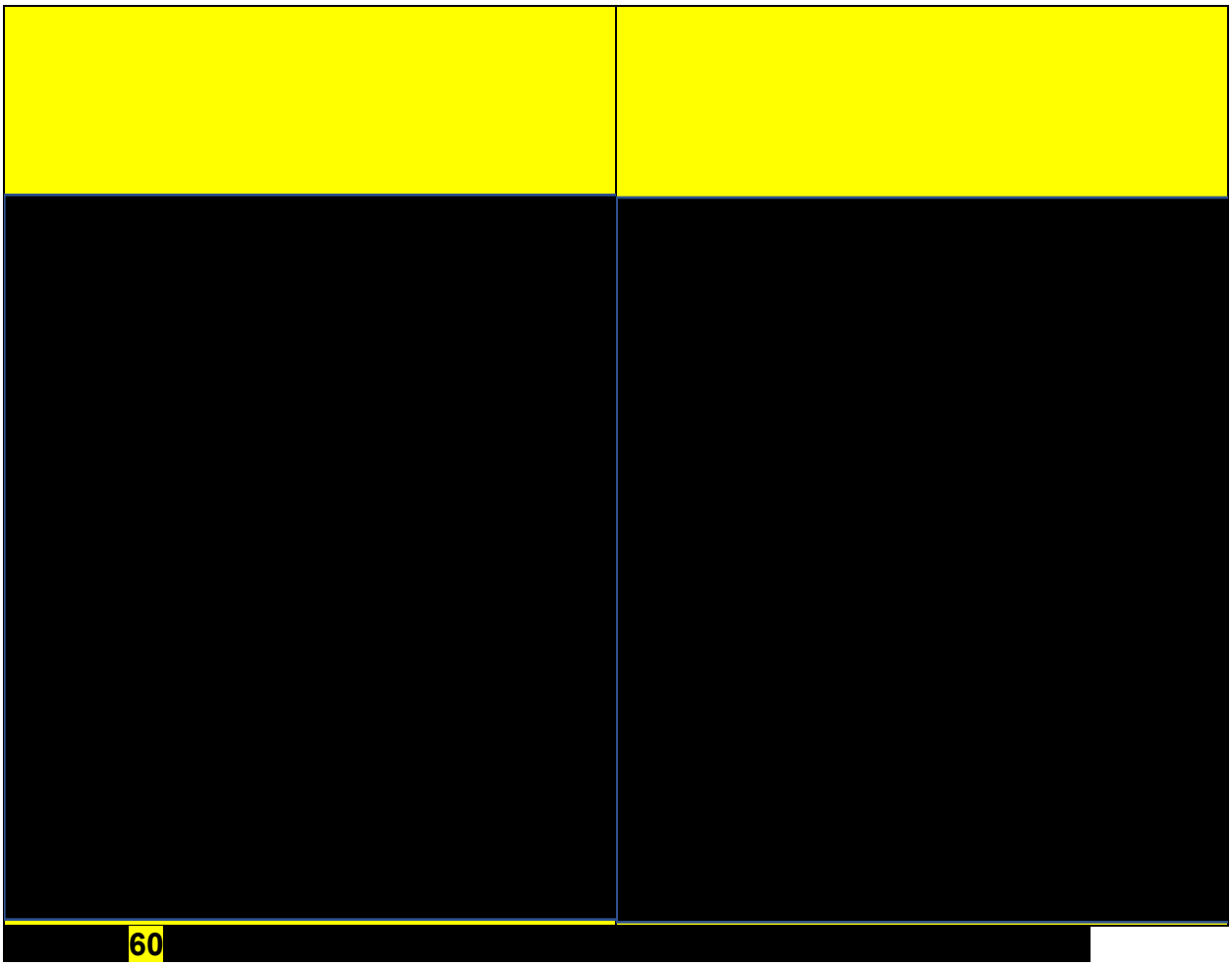
The company modelled treatment arms separately and stated (CS section B.3.3.7) "*standard parametric distributions described in CS Section B.3.3.3 were fitted to the ToT data for the ITT population (N=129) in GARNET to estimate ToT for dostarlimab within the model*". The modelled ToT KM plot exhibits several changes in trajectory and a long flat tail. In clarification the ERG received underlying data for ToT in the ITT population. This indicated that the first events occurred in [redacted] patients at [redacted] months (reproduced in Table 62).

62 [redacted]

Day	Month	Event	Censor	N at risk	S(t)
[redacted]					

This yields the KM plot shown left in Figure 60. However in the company economic model the ToT KM plot is for the efficacy population (N=[redacted]), the corresponding KM plot is shown in [redacted] right in which multiple events occur early at [redacted] months rather than [redacted] months; this plot corresponds to the KM shown in CS Figure 53. The difference between plots has potential implications for parametric modelling of ToT.





60

CS Table 60 (reproduced in Table 63 below for reference) presents AIC and BIC values entitled “Summary of goodness-of-fit data for dostarlimab ToT (GARNET ITT population) standard parametric and spline models”. However these values actually refer to models for the efficacy population detailed in the economic model. AIC/BIC values for the ERG’s parametric modelling of the ITT data supplied in clarification is shown in Table 64. These values and ranking differ somewhat from those in CS Table 60. Parametric models of ToT are summarised in Figure 61. Differences between ITT and efficacy models are modest but are most pronounced over the first 2 years of modelling (i.e. that part most relevant to the company’s modelling). Of the ERG models the best fit to the first one year and first two years of the ITT KM is provided by the ggamma model.

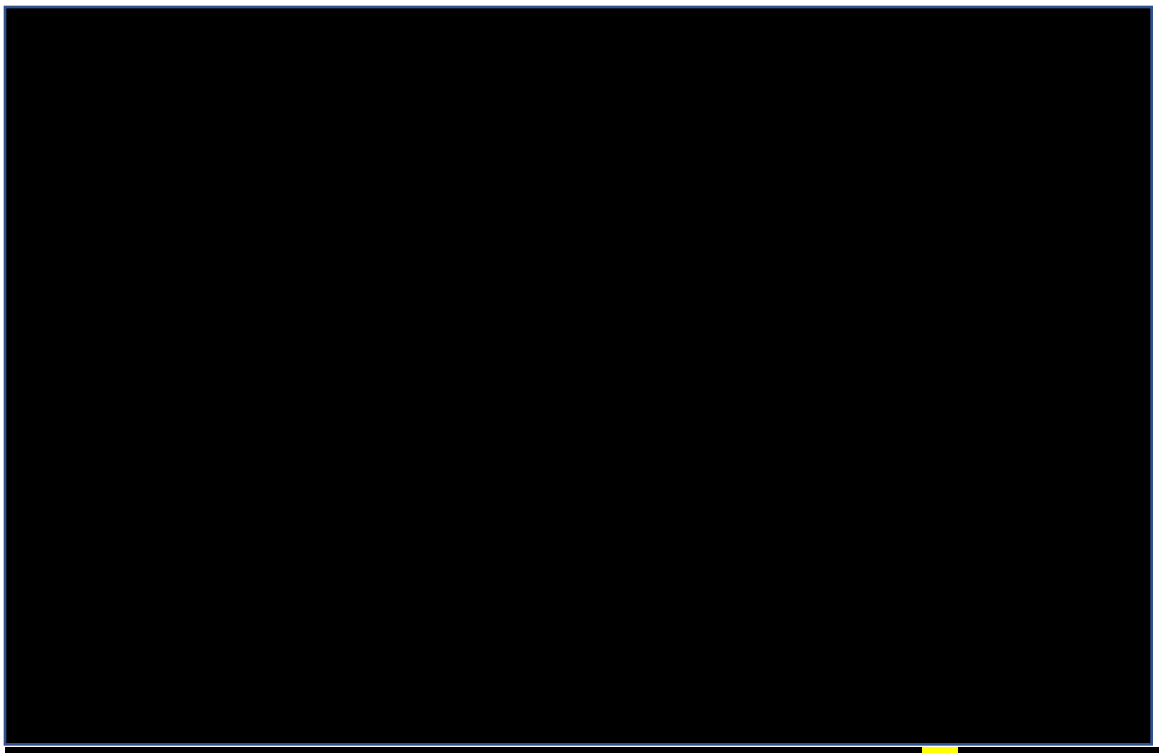
63

Distribution	AIC[a]	AIC Rank	BIC[a]	BIC Rank
Standard parametric models				
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Gamma	[redacted]	[redacted]	[redacted]	[redacted]
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Spline hazard with single knot	[redacted]	[redacted]	[redacted]	[redacted]
Spline hazard with two knots	[redacted]	[redacted]	[redacted]	[redacted]

[a] A small AIC or BIC value represents a better goodness of fit.

Table 64: AIC/BIC values for the ERG's parametric modelling of the GARNET ITT data for ToT

Model	AIC	BIC	AIC/BIC aggregate	rank	Observations
ggamma	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
lognormal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
R1P	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]



61

9.5.1 Expert clinical opinion on ToT

The company's justification for their method of modelling ToT is said to be supported by the opinion of clinical experts who undertook a GSK consultancy exercise; in clarification the ERG requested details of the exercise. The CS states: "*UK clinical expert opinion indicates that, regardless of whether patients are continuing to derive clinical benefit from dostarlimab, they would likely not receive dostarlimab any longer than* [redacted]*: any patients still receiving dostarlimab at* [redacted] *were assumed to discontinue treatment at this point in the base case cost-effectiveness analysis. Long-term extrapolations beyond* [redacted] *for ToT for dostarlimab were therefore not required*". However the ERG have been unable to identify clinicians' responses within the clarification consultation exercise details that can fully justify this statement.

In the exercise clinicians were shown a graph and a Table of data about time on treatment (shown in Figure 62 below); it is difficult to evaluate these because it is

described as “[redacted]” and how this data was constructed is unclear. In particular in shows ~[redacted] in treatment at ~ [redacted]

62 [redacted]

Clinicians were asked:

“[redacted]. The ERG believe the trajectory referred to is that used for “[redacted]” as shown in the graph above, however the graph trajectory stops at weeks 49-54 with about [redacted] in treatment (in contrast the GARNET trial data shows [redacted] in treatment at 52 weeks). The figure of [redacted] in the table at [redacted] appears to be a value “suggested” to the clinicians as a possibility. Extrapolation of the trajectory in the [redacted] would be difficult to gauge and the validity of any estimates doubtful because the origin and nature of the data is unclear. The ERG have not been able to identify clinicians’ quantitative responses to this question that might justify the company contention that in clinicians’ opinion patients would not receive dostarlimab beyond [redacted].

At time points after [redacted] the tabulated "[redacted]" provided to the clinicians does not align closely to the results observed in GARNET (e.g. KM plot shown in CS Figure 53) as shown below in Figure 63. The illustrative data shown to clinicians departs from the ToT KM at about [redacted]. The same considerations do not apply however when the company has analysed OS and PFS.

Figure 63: Comparison of GARNET ToT KM plot and data shown to clinical experts during elicitation exercise

The company further state that "*Accordingly, UK clinical experts indicated that based on their clinical experience with other I-O therapies, they would expect the real-world percentage of patients receiving dostarlimab after* [redacted] *would likely be between* [redacted]*% and* [redacted]*%, notably lower than the* [redacted]*% predicted by the GARNET ToT KM curve, and the percentages of patients on treatment at* [redacted] *predicted by all of the long-term extrapolations presented in [CS Figure 53]".* Again the ERG were unable to find clinicians' quantitative responses to support these values ([redacted] and [redacted]

Further questions posed for clinical experts were:

A]

[redacted]B]

"[redacted]The ERG note the wide range of clinicians' responses.

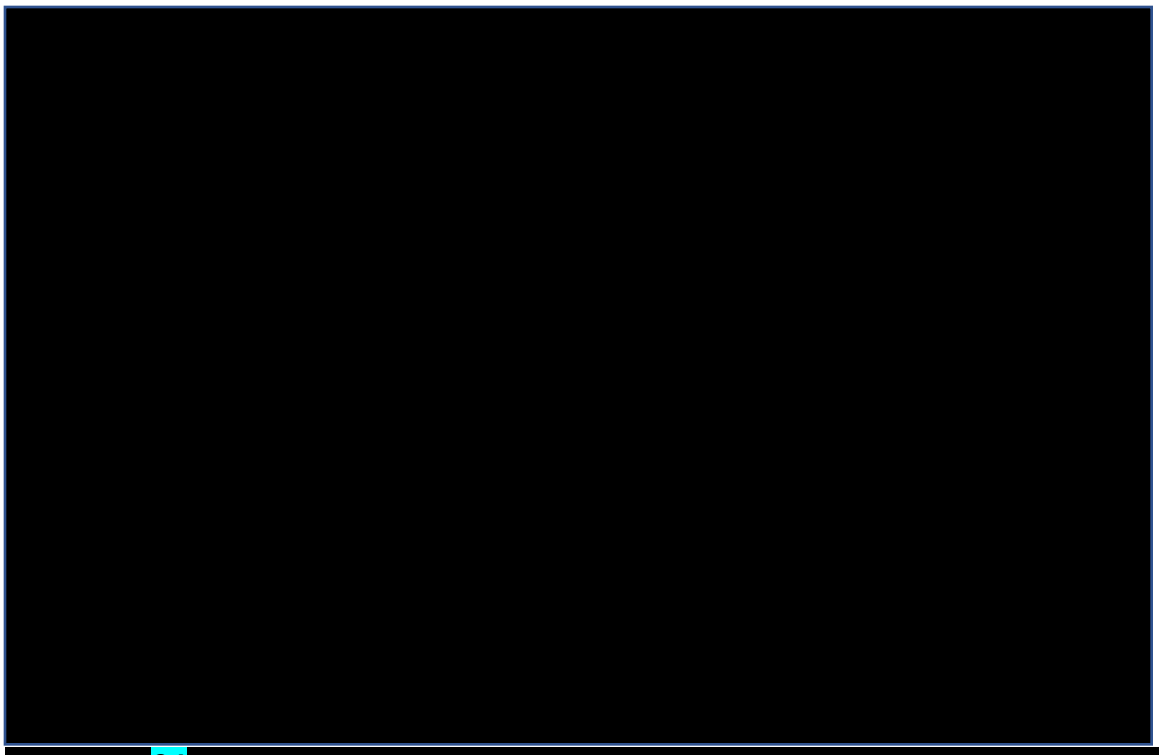
The ERG opinion is that the clinical experts' answers to structured questions posed in the consultancy exercise do not precisely support the company's modelling of ToT as shown in CS Figure 54, but may reflect the company's interpretation of clinicians narrative responses obtained during consultancy.

The company partly justify their ToT model on the basis that a similar clinician-opinion led model has been accepted by the avelumab appraisal committee. For the avelumab appraisal (TA517) a 2 year reduction to 33% remaining in treatment (rather than to [redacted] was implemented. The sponsors in that submission stated: "*Expert opinion was sought from three clinicians to establish how avelumab would be expected to be administered in practice, based on clinician experience of immunooncology therapies in other indications (such as ipilimumab, nivolumab and pembrolizumab)"* and "*In the model it has been assumed that the majority of patients cease treatment at 2 years. XXXXX XXXXX XXXXX both agreed that it was reasonable for a third of patients to remain on treatment after this time, with XXXXX XXXXX suggesting a realistic estimate would be between 30% and 40%. All clinicians agreed that a maximum treatment duration of 5 years, after which time all patients cease treatment, is reasonable. Furthermore, XXXXX XXXXX predicted that, based on melanoma data, continued treatment benefit would be observed* ". It is perhaps surprising that for modelling ToT with dostarlimab that GSK

applied such a large effect at [redacted] reducing proportion in treatment to only [redacted] The ERG consulted TA517 NICE documents but the observed PFS KM plot was one of many items completely redacted.

In view of this critique of the ToT modelling the ERG prefer a model that [a] is based on ITT population, rather than efficacy population; [b] uses the ggamma model for years [redacted], since this provides a superior fit to the KM data at times up to both [redacted] and [redacted]; [c] implements a reduction in the proportion in treatment to a larger value than the company's [redacted] at [redacted], since this seems more consistent with information available from the consultancy exercise; [d] continuation of the ggamma model to be in line with NICE appraisal committee for avelumab to year 5 years when all treatment is discontinued. To indicate the impact of selecting a higher value than [redacted] for exploratory illustration purposes the ERG looked at 27% at [redacted] (an arbitrary intermediate value between the company's [redacted]% and the 33% accepted by the TA 517 appraisal committee). Resulting models are shown in Figure 64 and compared with models with reduction to [redacted] on treatment at [redacted]. Weibull models generated in the same way are shown in Figure 65.

The economic impact of alternative ToT models to the base case CS model may be appreciable.



64

65

9.6 *Hazard plots for the company's parametric models of OS of patients receiving dostarlimab*

[redacted] left shows the hazards of the company's unadjusted parametric models extrapolated to 40 years. Taking modelled hazard as an indicator of risk of death for the ageing GARNET population it appears that, with the exception of Weibull, exponential and gamma models, risk of death continuously decreases through time. With treatment-waning adjustment for waning ([redacted] right) again with the exception of Weibull, exponential and gamma models, the risk of death decreases with time from [redacted] on. Decreasing hazard over such a 40 year extended period seems rather implausible in the context of ageing human populations which generally experience increasing risk of death with ageing over such extended time scales.

[redacted]

66 [redacted]

[redacted] below compares hazard of the CS treatment-waning adjusted ggamma model with hazard for the company's age-matched general population. To align with the

company’s ggamma model, the matched general population hazard is based on the well-fitting Gompertz parametric model for the matched population. Hazard from the ggamma model is more than ten times less than that for the general population from about [redacted] onward.

67 [redacted]