Esketamine for treatment-resistant depression

Produced by Kleijnen Systematic Reviews Ltd (KSR)

Authors
Robert Wolff, Deputy Director, KSR
Nigel Armstrong, Health Economist, KSR
Steve Ryder, Health Economist, KSR
Titas Buksnys, Health Economist, KSR
Debra Fayter, Systematic Reviewer, KSR
Stephanie Swift, Systematic Reviewer, KSR
Gill Worthy, Statistician, KSR
Caro Noake, Information Specialist, KSR
Jos Kleijnen, Director, KSR, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to
Robert Wolff, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 10/09/2019

Source of funding:
This report was commissioned by the NIHR HTA Programme as project number 12/78/96.

Declared competing interests of the authors
None.

Acknowledgements
Annette Chalker contributed to the data extraction and critique.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.
Academic in confidence (AiC) data are highlighted in yellow throughout the report.

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

This report should be referenced as follows:
Wolff R, Armstrong N, Ryder S, Buksnys T, Fayter D, Swift S, Worthy G, Noake C, Kleijnen J. Esketamine for treatment-resistant depression: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2019.

Contributions of authors
Robert Wolff acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Nigel Armstrong acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Titas Buksnys and Steve Ryder acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Debra Fayter and Stephanie Swift acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Caro Noake critiqued the search methods in the submission and contributed to the writing of the report. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

AAP	Atypical antipsychotic
ACP	American College of Physicians
AD	Antidepressant
AE	Adverse event
AiC	Academic in confidence
ANCOVA	Analysis of covariance
APA	American Psychiatry Association
Aug	Augmentation
BSC	Best supportive care
CBT	Cognitive behavioural therapy
C-SSRS	Columbia-Suicide Severity Rating Scale
CADTH	Canadian Agency for Drugs and Technologies in Health
CDSR	Cochrane Database Systematic Reviews
CEA	Cost effectiveness analysis
CFB	Change from baseline
CG	Clinical guideline
CGI-S	Clinical Global Impression – Severity
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CiC	Commercial in confidence
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
DARE	Database of Abstracts of Reviews of Effects
DBS	Deep brain stimulation
DSM-5	Diagnostic and Statistical Manual of Mental Disorders – Fifth edition
DSU	Decision Support Unit
EBM	Evidence-based medicine
epCS	Epidural cortical stimulation
ECT	Electroconvulsive therapy
EMA	European Medicines Agency
EQ-5D	European Quality of Life-5 Dimensions
EQ-5D-3L	European Quality of Life-5 Dimensions – 3 levels
EQ-5D-5L	European Quality of Life-5 Dimensions – 5 levels
ERG	Evidence Review Group
ESK	Esketamine
ESK-NS	Esketamine nasal spray
EU	European Union
EUCTR	European Union Clinical Trials Register
FAS	Full analysis set
GAD-7	Generalised Anxiety Disorder – 7-item scale
GAD	Generalised anxiety disorder
GP	General practitioner
HAM-D	Hamilton Depression Rating Scale
HCP	Healthcare professional
HCRU	Healthcare resource use
HDRS	Hamilton Depression Rating Scale
HR	Hazard ratio
HRQoL	Health-related quality of life
HSI	Health status index
HTA	Health technology assessment
IDMC	Independent data monitoring committee
ICER	Incremental cost effectiveness ratio

ICTRP	International Clinical Trials Registry Platform
INAHTA	International Network of Agencies for Health Technology Assessment
IND	Induction phase
IQR	Interquartile range
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IWRS	Interactive web response system
KM	Kaplan–Meier
KSR	Kleijnen Systematic Reviews
LS	Least squares
LYG	Life years gained
MA	Maintenance phase
MADRS	Montgomery-Åsberg Depression Rating Scale
MAOI	Monoamine oxidase inhibitor
MDD	Major depressive disorder
MDE	Major depressive episode
MECIR	Methodological Expectations of Cochrane Intervention Reviews
MeDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical subject headings
NA	Not applicable
NCPE	National Centre for Pharmacoeconomics
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
nortrip	Nortriptyline
NR	Not reported
NRI	Norepinephrine reuptake inhibitor
NS	Nasal spray
OAD	Oral antidepressant
OC	Observed cases
OP	Optimisation phase
OR	Odds ratio
PAS	Patient access scheme
PBAC	Pharmaceutical Benefits Advisory Committee
PBO	Placebo
PBO-NS	Placebo nasal spray
PGI-S	Patient Global Impression – Severity
PHQ-9	Patient Health Questionnaire – 9 questions
PRIMA	Preliminary Independent Model Advice
PSA	Probabilistic sensitivity analyses
PSS	Personal social services
PWC-WS	Physicians Withdrawal Checklist- Withdrawal Symptoms- subscale
QALY	Quality-adjusted life year
QIDS	Quick Inventory of Depressive Symptomatology
QLDS	Quality of life in depression scale
RCT	Randomised controlled trial
RePEc	Research papers in economics
SAD	Seasonal affective disorder
SAE	Serious adverse events
SARI	Serotonin antagonist and reuptake inhibitor
SD	Standard deviation
SDS	Sheehan disability scale
SLaM	South London and Maudsley
SLR	Systematic literature review
SMC	Scottish Medicine Consortium

SNRI	Serotonin–norepinephrine reuptake inhibitor
SR	Systematic review
SSRI	Selective serotonin reuptake inhibitor
TA	Technology appraisal
TADS	Tavistock adult depression study
TAU	Treatment as usual
TCA	Tricyclic antidepressant
TeCA	Tetracyclic antidepressant
TMS	Transcranial magnetic stimulation
TRD	Treatment-resistant depression
UK	United Kingdom
US	United States (of America)
USA	United States of America
VNS	Vagal nerve stimulation
WHO	World Health Organization
WPA	World Psychiatric Association
WTP	Willingness-to-pay
XR	Extended release

Table of Contents

Table of Tables

Table of Figures

1. Executive summary

1.1 Critique of the decision problem in the company's submission

The final scope issued by the National Institute for Health and Care Excellence (NICE) defined the decision problem as follows:

- Adults with treatment-resistant depression (TRD)
- Esketamine nasal spray (ESK-NS) in addition to established clinical management
- Comparators including
    - Selective serotonin reuptake inhibitors (SSRIs)
    - Tricyclic antidepressants (TCAs)
    - Monoamine oxidase inhibitors (MAOIs)
    - Serotonin–norepinephrine reuptake inhibitors (SNRIs)
    - Vortioxetine
    - Combination or augmentation treatments (with lithium or an antipsychotic)
    - Electroconvulsive therapy (ECT)
    - Best supportive care (BSC)
- Outcomes of interest
    - Response to treatment (including response rate and time to response)
    - Relapse (including relapse rate and time from remission to relapse)
    - Severity of depression
    - Cognitive dysfunction
    - Remission of symptoms
    - Anxiety
    - Sleep quality
    - Hospitalisation
    - Functioning and associated disability
    - Mortality
    - Adverse effects of treatment (including adverse effects of treatment discontinuation)
    - Health-related quality of life (HRQoL)

Regarding the population, the evidence presented in the company submission (CS) was broadly in line with the NICE scope. However, there are some important discrepancies with the scope, which include the specification of moderate to severe depression. Also, the main clinical and cost effectiveness evidence was of questionable applicability to those in the age group 65 years and above, although the company did perform some mitigatory adjustment to the cost effectiveness model, see section 1.3.

The intervention defined and presented in the CS was ESK-NS co-administered with a newly initiated oral antidepressant (OAD). This is in line with the expected label indication (*"ESK-NS in combination with an SSRI or SNRI for treatment resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode"*).

The comparators in the decision problem of the CS were in line with the NICE scope. However, the company suggested adding mirtazapine, a tetracyclic OAD, to the list of comparators. Subsequently, mirtazapine was a comparator in network meta-analyses (NMAs) presented in the CS.

The outcomes investigated in the CS reflected the NICE scope. However, some outcomes were not reported for the main trials in the CS, namely cognitive dysfunction, hospitalisation and sleep quality.

No patient access scheme (PAS) was presented in the CS. Given the method of administration of ESK-NS requiring supervision by a healthcare professional, it will be important to ensure that access to healthcare support will not inappropriately discriminate against individuals for whom geography may pose a challenge.

1.2 Summary of clinical effectiveness evidence submitted by the company

The CS and response to clarification provided sufficient details for the Evidence Review Group (ERG) to appraise the searches for eligible studies. A good range of resources were searched and the majority of searches were well documented making them transparent and reproducible. Additional searches of HTA agencies, clinical trials registries and conference proceedings were reported. However, the ERG was concerned about the language bias of restricting searches to English language only as this is not in line with current best practice.

Six studies formed the evidence base for ESK-NS. Four of these were randomised controlled trials (TRANSFORM-1, TRANSFORM-2, TRANSFORM-3, SUSTAIN-1) and two were open label extension studies (SUSTAIN-2, SUSTAIN-3). The two main trials which informed the economic modelling were TRANSFORM-2 and SUSTAIN-1.

TRANSFORM-2 enrolled adults with a history of non-response to at least two antidepressants in the current episode with one antidepressant assessed prospectively while SUSTAIN-1 assessed maintenance of effect (prevention of relapse). Both trials compared ESK-NS plus a newly initiated OAD to a newly initiated OAD plus placebo and both involved flexible dosing of 56 mg/ 84 mg of ESK-NS. ESK-NS was given for four weeks in TRANSFORM-2 and patients were either followed-up for 24 weeks or joined SUSTAIN-1. SUSTAIN-1 also enrolled patients directly who had not taken part in TRANSFORM-2. In SUSTAIN-1, ESK-NS was given until relapse or trial termination.

In TRANSFORM-2, ESK-NS + OAD in comparison to placebo nasal spray (PBO-NS) + OAD showed a statistically significant reduction on the Montgomery-Åsberg Depression Rating Scale (MADRS) at day 28 (difference in least squares means -4.0, 95% confidence interval (CI) -7.31 to -0.64). Of note, there are differences between the type of OAD for remission rates after 28 days, e.g. within the SSRI group: sertraline (odds ratio (OR) 1.38, 95% CI 0.26 to 7.22) vs. escitalopram (OR 4.71, 95% CI 1.08 to 20.63). The trial also showed differences in response rate and remission rate, respectively, between the two groups. Other reported outcomes were in favour of the intervention (see Table 1.1).

In SUSTAIN-1, the percentage of relapse was lower in the ESK-NS + OAD (stable remitters: 26.7%, stable responders: 25.8%) group in comparison to participants receiving PBO-NS + OAD (45.3% and 57.6%, respectively). The trial also showed time to relapse to be in favour of the intervention group for both, stable remitters (hazard ratio (HR) 0.49, 95% CI 0.29 to 0.84) and stable responders (HR 0.30, 95% CI 0.16 to 0.55). Other reported outcomes were in favour of the intervention (see Table 1.2).

In the induction phase of TRANSFORM-2, more adverse events were observed in patients treated with ESK-NS + OAD compared to those receiving PBO-NS + OAD (85.2% vs. 60.6%, see Table 1.3). In SUSTAIN-1 more adverse events were seen in the ESK-NS + OAD group in the maintenance phase (82.2% vs. 45.5%) and the follow-up phase (11.0% vs. 7.8%), see Table 1.4. Potential adverse events, especially psychiatric disorders (47.8% vs. 19.3% in TRANSFORM-2), need to be considered before considering ESK-NS as a treatment option for patients with TRD.

The main limitation of these trials in terms of this appraisal is that they only included patients aged 18 to 64 years of age. Furthermore, the trials in the CS excluded patients with moderate/severe alcohol abuse according to Diagnostic and Statistical Manual of Mental Disorders – Fifth edition (DSM-5)

criteria. The committee will need to consider whether evidence in the CS on effectiveness and safety of ESK-NS can be generalised to those with a dual diagnosis of depression and alcohol misuse.

The trials in the CS also excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on Columbia-Suicide Severity Rating Scale (C-SSRS) or a history of suicidal behaviour in the 12 months prior to screening. Again, the committee will need to consider if the evidence in the CS on effectiveness and safety of ESK-NS can be generalised to this vulnerable population.

As discussed in section 1.3, the ERG noted a lack of clarity on dosing in the included trials which might impact on the generalisability of these trials.

Furthermore, the ERG noticed the short-term nature of the trials which is a concern, especially for safety-related outcomes. SUSTAIN-3, when reported in full, should give a fuller picture of any potential longer-term risks with ESK-NS including those related to withdrawing from treatment.

The company stated that the NMA was not considered sufficiently robust to inform the cost effectiveness analysis (CEA). The ERG could run the NMA and obtained results which were very close to those provided by the company so they have no concerns about the NMA analysis methods. However, the main concerns about the NMA results are due to the clinical and methodological differences between the studies included in each network.

Table 1.1: Summary of efficacy results of TRANSFORM-2

Outcome	ESK-NS + OAD	OAD + PBO-NS
MADRS[a,b]		
Change from baseline (observed cases)		
MMRM (difference in LS means, SE, 95% CI)[d]	-4.0 (1.69, -7.31 to -0.64)	
Onset of clinical response (FAS)		
Generalised Cochran-Mantel-Haenszel test[e]	OR 1.79 (95% CI 0.57 to 5.67)	
Response and remission (observed cases)		
Response rate[f]	69.3%	52.0% (unadjusted)[g]
		34.0% (adjusted)[g]
Remission rate[h]	52.5%	31.0% (unadjusted)[g]
		18.0% (adjusted)[g]
CGI-S (observed cases)[i]		
MMRM (difference in LS means, SE, 95% CI)[d]	-0.4 (0.17, -0.72 to -0.04)	
PHQ-9 (observed cases)[i]		
MMRM (difference in LS means, SE, 95% CI)[d]	-2.4 (0.88, -4.18 to -0.69)	
GAD-7 (observed cases)[j]		
ANCOVA (difference in LS means, SE, 95% CI)[k]	-1.0 (0.67, -2.35 to 0.28)	

Outcome	ESK-NS + OAD	OAD + PBO-NS
SDS (observed cases)[l]		
MMRM (difference in LS means, SE, 95% CI)[b]	-4.0 (1.17, -6.28 to -1.64)	
EQ-5D (observed cases)[b,m]		
Change from baseline to day 28 (mean, SD)	N=104, 0.310 (0.2191)	N=100, 0.235 (0.2525)
Other outcomes defined in the final scope		
Cognitive dysfunction	NR	NR
Hospitalisation	NR	NR
Sleep quality	NR	NR

Based on Tables 7, 19, 21, 23, 24, 26, 45 and Figure 15 of the CS as well as the CSR

[a] Related to response, severity of depression, and remission; [b] Used in the economic model; [c] = Table 19 of the CS reported this as "109". Error corrected by the ERG; [d] Change from baseline was the response variable and fixed effect model terms for treatment, day, country, class of OAD (SNRI or SSRI), treatment-by-day, and baseline value were covariates; [e] Adjusted for region and class of OAD (SNRI or SSRI); [f] ≥50% reduction from baseline in MADRS total score; [g] See details in section 5.2.6.1; [h] MADRS total score of ≤12; [i] Related to severity of depression; [j] Related to anxiety; [k] Change from baseline was the response variable and treatment, country, class of OAD (SNRI or SSRI), and baseline GAD-7 value were covariates; only ANCOVA reported; [l] Related to functioning and associated disability; [m] = Related to health-related quality of life

ANCOVA = analysis of covariance; CGI-S = Clinical Global Impression; CI = confidence interval; CS = company submission; CSR = clinical study report; EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; ESK = esketamine; FAS = full analysis set; GAD-7 = Generalised Anxiety Disorder – 7-item scale; HR = hazard ratio; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; NR = not reported; NS = nasal spray; OAD = oral antidepressant; OR = odds ratio; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Table 1.2: Summary of efficacy results of SUSTAIN-1

Outcome	ESK-NS + OAD	OAD + PBO-NS
Time to relapse		
Stable remitters[a]		
Time to relapse	HR 0.49 (95% CI 0.29 to 0.84)	
Stable responders[b]		
Time to relapse	HR 0.30 (95% CI 0.16 to 0.55)	
MADRS (LOCF)[c,d]		
Change from baseline		
ANCOVA (difference in LS means, SE, 95% CI)[e]	Stable remitters[a]: -5.2 (1.82, -8.7 to -1.58)	
	Stable responders[b]: -7.4 (1.95, -11.30 to -3.55)	
Response/remission		
Responder at end of maintenance phase[f]	Stable remitters[a]: 67/89 (75.3%)	Stable remitters[a]: 48/86 (55.8%)
	Stable responders[b]: 41/62 (66.1%)	Stable responders[b]: 20/59 (33.9%)
Remitter at end of maintenance phase[f]	Stable remitters[a]: 58/89 (65.2%)	Stable remitters[a]: 36/86 (41.9%)

Outcome	ESK-NS + OAD	OAD + PBO-NS
	Stable responders[b]: 29/62 (46.8%)	Stable responders[b]: 15/59 (25.4%)
CGI-S (LOCF)[g]		
ANCOVA (difference in LS means, SE, 95% CI)[e]	Stable remitters[a]: P value 0.055[h]	
	Stable responders[b]: P value 0.002[h]	
PHQ-9 (LOCF)[g]		
Change from baseline		
ANCOVA (difference in LS means, SE, 95% CI)[e]	Stable remitters[a]: -2.4 (0.90, -4.20 to -0.65)	
	Stable responders[b]: -3.0 (0.93, -4.87 to -1.18)	
Response/remission		
Responder at end of maintenance phase	Stable remitters[a]: 72/89 (80.9%)	Stable remitters[a]: 57/86 (66.3%)
	Stable responders[b]: 48/61 (78.7%)	Stable responders[b]: 40/58 (69.0%)
Remitter at end of maintenance phase	Stable remitters[a]: 51/89 (57.3%)	Stable remitters[a]: 38/86 (44.2%)
	Stable responders[b]: 23/61 (37.7%)	Stable responders[b]: 12/58 (20.7%)
GAD-7 (LOCF)[i]		
ANCOVA (difference in LS means, SE, 95% CI)[e]	Stable remitters[a]: -1.7 (0.72, -3.12 to -0.28)	
	Stable responders[b]: -1.1 (0.72, -2.56 to 0.31)	
SDS (LOCF)[g]		
Change from baseline		
ANCOVA (difference in LS means, SE, 95% CI)[e]	Stable remitters[a]: -2.9 (1.30, -5.51 to -0.38)	
	Stable responders[b]: -4.7 (1.31, -7.30 to -2.10)	
Response/remission		
Responder at end of maintenance phase[f]	Stable remitters[a]: 58/83 (69.9%)	Stable remitters[a]: 43/78 (55.1%)
	Stable responders[b]: 42/60 (70.0%)	Stable responders[b]: 23/53 (43.4%)
Remitter at end of maintenance phase[f]	Stable remitters[a]: 48/83 (57.8%)	Stable remitters[a]: 30/78 (38.5%)
	Stable responders[b]: 25/60 (41.7%)	Stable responders[b]: 11/53 (20.8%)
EQ-5D (HIS score)[h]		
Change from baseline to end of maintenance phase (mean, SD)[f]	Stable remitters[a]: N=88, -0.067 (0.1180)	Stable remitters[a]: N=86, -0.096 (0.1484)
	Stable responders[b]: N=61, -0.023 (0.0753)	Stable responders[b]: N=58, -0.073 (0.1383)
Other outcomes defined in the final scope		
Cognitive dysfunction	NR	NR
Hospitalisation	NR	NR

Outcome	ESK-NS + OAD	OAD + PBO-NS
Sleep quality	NR	NR

Based on Tables 7, 8, 27, 28, 29, 30 of the CS

[a] Patients who were in stable remission at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase; [b] Patients who were stable responders (who were not stable remitters) at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase; [c] Related to relapse, severity of depression, and remission; [d] Used in the economic model; [e] Change from baseline was the response variable and treatment, country, and baseline value were covariates; [f] Variable duration (until relapse or study termination); [g] Related to severity of depression; [h] No further information reported, [i] Related to anxiety

ANCOVA = analysis of covariance; CGI-S = Clinical Global Impression; CI = confidence interval; CS = company submission; EQ-5D = European Quality of Life-5 Dimensions; ESK = esketamine; GAD-7 = Generalised Anxiety Disorder – 7-item scale; HSI = health status index; HR = hazard ratio; LOCF = last observation carried forward; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; NR = not reported; NS = nasal spray; OAD = oral antidepressant; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error

Table 1.3: Safety results of TRANSFORM-2

	ESK-NS + OAD	OAD + PBO-NS
Induction phase, n (%)	**N=115**	**N=109**
Overall summary		
AE	98 (85.2)	66 (60.6)
AE possibly related to nasal spray drug[a]	90 (78.3)	39 (35.8)
AE possibly related to OAD[a]	39 (33.9)	26 (23.9)
AE leading to death	1 (0.9)	0
≥1 serious AE	1 (0.9)	1 (0.9)
AE leading to nasal spray drug being withdrawn[b]	8 (7.0)	1 (0.9)
AE leading to OAD being withdrawn[b]	4 (3.5)	0
Follow-up phase, n (%)	**N=34**	**N=52**
Overall summary		
AE	9 (26.5)	12 (23.1)
AE possibly related to nasal spray drug[a]	0	1 (1.9)
AE possibly related to OAD[a]	1 (2.9)	3 (5.8)
AE leading to death	0	0
≥1 serious AE	1 (2.9)	0
AE leading to OAD being withdrawn[b]	0	0

Based on Tables 37 and 38 of the CS

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] An AE that started in the double-blind induction phase and resulted in discontinuation in the follow-up phase was counted as treatment-emergent in the double-blind induction phase; [redacted]

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

Table 1.4: Safety results of SUSTAIN-1 (overall)

	Induction phase	Optimisation phase	Maintenance phase		Follow-up phase	
	ESK-NS + OAD (N=437)	ESK-NS + OAD (N=455)	ESK-NS + OAD (N=152)	OAD + PBO-NS (N=145)	ESK-NS + OAD during any phase (N=481)	OAD + PBO-NS for all phases (N=64)
AE, n (%)	336 (76.9)	335 (73.6)	125 (82.2)	66 (45.5)	53 (11.0)	5 (7.8)
AE possibly related to nasal spray drug, n (%)[a]	301 (68.9)	281 (61.8)	106 (69.7)	37 (25.5)	7 (1.5)	0
AE possibly related to OAD, n (%)[a]	71 (16.2)	61 (13.4)	13 (8.6)	9 (6.2)	3 (0.6)	0
AE leading to death, n (%)	0	0	0	0	0	0
≥1 serious AE, n (%)	13 (3.0)	11 (2.4)	4 (2.6)	1 (0.7)	3 (0.6)	0
AE leading to nasal spray drug being withdrawn, n (%)	22 (5.0)	5 (1.1)	4 (2.6)	3 (2.1)	NA[b]	NA[b]
AE leading to OAD being withdrawn, n (%)[c]	8 (1.8)	2 (0.4)	3 (2.0)	0	0[c]	0[c]

Based on Table 39 of the CS

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] Patients did not receive nasal spray during the follow-up phase; [c] An AE that started in the induction phase and resulted in discontinuation in a subsequent phase was counted as treatment-emergent in the induction phase.

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

1.3 Summary of the key issues in the cost effectiveness evidence

Given that the NICE scope has no upper age limit, the ERG considers that a new version of the base-case model, submitted at the clarification stage should be used as an updated company base-case. It includes acute response and remission transition probabilities and utilities for major depressive episode (MDE), response and remission/recovery states from both TRANSFORM-2 and TRANSFORM-3, weighted by percentage in each age group such that if set to 0% for age >65 years one gets the same result as in the original base-case. This forms the starting point for the ERG base-case.

Regarding the intervention, ESK-NS + OAD, the ERG is concerned with the lack of clarity on dosing in TRANSFORM-2 and TRANSFORM-3 trials plus the complex dose changes in SUSTAIN-1 and SUSTAIN-2, which mean that it is difficult to know how applicable to clinical practice the transition probabilities estimated from the trials would be. The ERG recognises that adopting a mix of OADs as concomitant and comparator treatment is not ideal, given possible differences in effectiveness between individual OADs. There is the possibility that ESK-NS might be cost effective in combination with one OAD and not another. However, the ERG did not have the data to implement the required variation in all parameter estimates required for the model. The ERG is convinced that the limitations of the NMA are sufficient to exclude any other comparator except in a scenario analysis. However, the applicability to clinical practice of results would be highest in those patients who might be switched to one of the OADs prescribed in the trials. The ERG could find no errors or violations of modelling convention in the model. The only other key issues, which were substantial, were addressed as matters of judgement as much as was feasible by the ERG in forming the ERG base-case (issues 1 to 5) and three additional scenarios (issues 6 to 8):

1) Time horizon: although a lifetime time horizon is usually required, the company base case of five years in the company base case is longer than that in the previous appraisal in a similar population, technology appraisal (TA) 367. The ERG discovered that 20 years appeared to be the minimum to ensure no continued difference in cost or quality-adjusted life years (QALYs) in the model, see section 5.2.5.
2) Adjustment for placebo effect to the acute response or remission transition probabilities only for the comparator. This introduces a bias in favour of ESK-NS + OAD. The ERG considers that the company made a case for stating that some of the placebo response might be due to the effect of additional clinic visits in the trials, but is not convinced that this is the only factor and that it could only apply to ESK-NS + OAD in clinical practice, see section 5.2.6.1.
3) Discontinuation for reasons other than loss of efficacy. There was a lack of evidence to support there being no loss of efficacy on discontinuing ESK-NS and remaining only on OAD. This problem had already been identified in the NICE Preliminary Independent Model Advice (PRIMA) scientific report, see section 5.2.6.3.
4) Effect on mortality of ESK-NS + OAD. There was an absence of evidence for a treatment effect on mortality and no such treatment effect was applied in TA367, see section 5.2.6.7.
5) Cost of clinic visit for ESK-NS + OAD based on patient to nurse ratio of 1:6. This was believed by the ERG to be implausible in clinical practice. It was also a finding of the NICE PRIMA scientific report, see section 5.2.8.2.
6) The considerable difference between ESK-NS + OAD and OAD in the loss of response and relapse transition probabilities. There was a lack of comparative evidence to inform these parameters, the

values being derived from different sources. Such a difference is also inconsistent with the judgement of the committee in TA367, see section 5.2.6.2.

7) The probabilities of response and remission at each line of subsequent therapy appeared to be too low when considering how they were implemented in the model and by comparison to the values in what was purported to be the data source, i.e. the STAR*D trial. There was also a lack of clarity in the method of calculation of these probabilities. It also seemed to be inconsistent with the method recommended by the committee in TA367, see sections 5.2.6.4 and 5.2.6.5.

8) Although the ERG is not convinced that the placebo response is explained entirely by the effect of additional clinic visits in the trials, it does consider that it is reasonable to attribute some of the effect on response and remission to be attributable to the extra clinic sessions. Therefore, it might be that the correct comparator should be OAD plus additional clinic sessions.

Searches were undertaken to identify economic evaluations and United Kingdom (UK) based resource use and HRQoL evidence. The CS provided sufficient details for the ERG to appraise the searches. An extensive range of databases and additional resources was searched.

1.4 Summary of the ERG's preferred assumptions and resulting ICER

The ERG base-case was created based on the preferred assumptions of the ERG regarding the issues 1) to 5), as listed in section 1.3. The results are shown in Table 1.5.

Table 1.5: ICER resulting from ERG's preferred assumptions (cumulative effect)

Preferred assumption		Section in ERG report	Cumulative ICER £/QALY
	Company base-case using 'adults and elderly' model		£7,699
1	Time horizon 20 years	5.2.5	£4,774
2	No adjustment for placebo effect to OAD Acute response or remission transition probabilities	5.2.6.1	£12,743
3	No discontinuation for reasons other than loss of efficacy	5.2.6.3	£53,254
4	No effect on mortality of ESK-NS + OAD	5.2.6.7	£55,478
5	Cost of clinic visit for ESK-NS + OAD based on patient to nurse ratio of 1:1	5.2.8.2	£62,566

ERG = Evidence Review Group; ESK = esketamine; ICER = incremental cost effectiveness ratio; NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year

1.5 Summary of exploratory and sensitivity analyses undertaken by the ERG

Two scenario analyses were created and added to the ERG base-case, based on the preferred assumptions of the ERG regarding issues 6) and 7), as listed in section 1.3. A further scenario was a response to the idea that the placebo effect might be the result of extra clinic visits, the cost for which should be equal in both the intervention and the comparator. The results are shown in Table 1.6.

Table 1.6: Exploratory analyses undertaken by the ERG (non-cumulative)

ERG assumption		Section in ERG report	ICER £/QALY
5	ERG's base-case using 'adults and elderly' model		£62,566
6	No difference between ESK-NS + OAD and OAD in the loss of response and relapse transition probabilities	5.2.6.2	£97,396
7	A decrease in response and remission was applied at each line of subsequent therapy (including BSC) by multiplying the values for OAD by a factor equal to the ratio of values in Step 3 versus Step 4 in STAR*D. These ratios are: 13.7/13.0 and 16.8/16.3 for remission and response respectively. Values estimated by the company from STAR*D were, for loss response, 22.2% for first line TRD and 22.8% for second line TRD and, for relapse, of 6.8% for first line TRD and 12.8% for second line TRD.	5.2.6.4, 5.2.6.5	£148,650
8	Cost of clinic visits for OAD set equal to that for ESK-NS + OAD	5.2.8.3	£53,911

BSC = best supportive care; ERG = Evidence Review Group; ESK = esketamine; ICER = incremental cost effectiveness ratio; NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year, TRD = treatment-resistant depression

In conclusion, the result of the adjustments to the company base-case produced an ERG base-case with an ICER that was considerably higher that the company base-case, i.e. £62,566 instead of £7,699. Scenario analyses showed it could be as low as £53,911 and as high as £148,650. The approach taken to form the ERG base-case contrasts very strongly with the assumptions made in the CS, which at every stage enhanced the treatment effect on the basis of unclear justification, i.e. no or very little comparative evidence and rather opaque exposition. In particular, no data were provided to support the lack of impact on effectiveness of discontinuing ESK and all of the evidence to inform the company base case came from differential data sources for the intervention and the comparator beyond the acute phase. Despite a request for clarification, it remains unclear why more data from the SUSTAIN studies could not have been used to inform the relapse and loss of response rates for OAD.

Finally, the method of estimating all transition probabilities beyond the acute phase is unclear, both the precise data used from SUSTAIN-1 to inform those for ESK-NS + OAD and the calculations used to transform the data from STAR*D to inform those for OAD.

2. Background

2.1 Introduction

In this report, the Evidence Review Group (ERG) provides a review of the evidence submitted by Janssen in support of esketamine nasal spray (ESK-NS), trade name SPRAVATO®, for patients with treatment-resistant depression (TRD). In this section, the critique of the company's description of the underlying health problem and the overview of current service provision is outlined. The information is taken from section B.1.3 of the company submission (CS) with subsections referenced as appropriate.[1] The ERG also received a submission from the United Kingdom (UK) mental health charity SANE which presented the views of those with TRD.[2] The views were largely taken from a survey of 100 patients and 90 carers where patients had not responded to at least two different anti-depressants in the current depressive episode.

2.2 Critique of company's description of underlying health problem

The underlying health problem of this appraisal is TRD which the company described as '*major depressive disorder (MDD) that has not responded to at least two different treatments with OADs* [oral antidepressants] *in the current moderate to severe depressive episode*'.[1]

The company described MDD (also known simply as 'depression') as a '*severely debilitating and potentially life-threatening psychiatric disorder. MDD is characterised by recurrent episodes of persistent low mood and / or loss of interest or pleasure in (almost) all activities*'.[1] The accompanying symptoms as '*profound sleep disturbance, fatigue, change in appetite/weight, agitation or slowness of speech/action, diminished concentration, decreased libido, inability to enjoy life, and feelings of worthlessness*'.[1] The CS provided details of the diagnosis and the psychological, physical and social symptoms of MDD and TRD.

Regarding burden of disease, the company stated that '*around 3% of the UK population, about 2 million people are affected by MDD at any given time*'.[1] The CS identified that there could be over 130,000 patients in the UK who do not achieve remission with currently available OADs and therefore have TRD. The company further stated that '*the total estimated societal burden of TRD is £3.9 billion, the majority of which (80%) is due to carer burden and lost productivity*'[1] and that depression '*can develop at any age, but disproportionally effects* [sic] *people of working age*'.[1]

The company made several statements to illustrate the seriousness and impact on patients of TRD in relation to non-TRD: '*Episodes of depression in patients with TRD are typically three times longer than in patients with non-treatment resistant MDD* [CS reference 37] *and are associated with increased all-cause mortality* [CS reference 38], *mainly due to a seven times increased risk of suicide relative to MDD* [CS reference 39].[1] The company added that '*at least 30% of patients with TRD attempt suicide at least once during their lifetime*'.[1]. In the survey conducted by SANE, 80% of patients reported having had suicidal thoughts in the previous 12 months.[2]

In the CS, the company further stated that '*The impact of TRD on patient health-related quality of life (HRQoL) is profound; patients with TRD have around 35% greater reductions in HRQoL compared with non-treatment resistant MDD, and report impairment in HRQoL in the range of metastatic cancer or acquired blindness* [CS reference 40].'[1] In the survey by SANE, 89% of patients reported TRD as having a major impact on their quality of life with 93% having a loss of interest or pleasure in all or almost all activities most of the day.[2]

The company described the negative impact on work activity of TRD. The survey by SANE commented that 45% of those with TRD had to stop work completely.[2]

The company concluded that '*there is a large unmet need for a safe, well-tolerated treatment with a rapid onset of action and durable efficacy*'.[1] This was supported by the SANE submission which stated that just 56% of patients and carers considered their current treatment to be effective and that 57% believed the benefits of antidepressants outweighed the adverse effects.[2]

ERG comment: The company provided a good overview of the underlying health problem of treatment-resistant depression, illustrating the seriousness of the condition and its impact on patients and their families. The ERG checked the references provided to support the statements in the CS. In general, these were appropriately referenced. However, some points should be noted:

- The ERG noted that TRD was not explicitly defined in the National Institute for Health and Care Excellence (NICE) scope and, as mentioned in the CS, is not consistently defined in clinical practice. The definition used by the company ('*major depressive disorder (MDD) that has not responded to at least two different treatments with OADs in the current moderate to severe depressive episode*')[1] reflects the expected licence for ESK-NS as well as the European Medicines Agency (EMA) guidance and therefore appears reasonable.[3]
- The ERG could not verify the estimate of 130,000 people with TRD, but given differences in the definition of TRD, this will be difficult to determine with certainty. The estimated societal burden of TRD of £3.9 billion was taken from a retrospective analysis of service use and costs of 129 Tavistock Adult Depression Study (TADS) patients.[4] In current (2015/16) prices the authors stated that costs would be approximately £25,000 per person. The authors acknowledged that costs in their study were higher than other studies using a different definition of TRD.
- The statement that those with TRD are at seven times increased risk of suicide relative to MDD was based on a Medicare analysis of with 4,639 patients with TRD and 7,524 with managed depression.[5] In this study, 7% of those with TRD and 1% with managed depression made a suicide attempt or self-inflicted injury. Although those with TRD are at increased risk of suicide, the exact difference between groups should be treated with some caution. In this context, it is important to note that in the main trials of this submission the following patients were excluded: '*Suicidal ideation/intent within 6 months prior to screening per the investigator's clinical judgements and/or based on C-SSRS* [Columbia-Suicide Severity Rating Scale], *or a history of suicidal behaviour in the 12 months prior to screening*'[1]
- The company stated that '*in clinical practice semi-structured interviews are usually used to diagnose and monitor the level of depressive symptoms. Scoring systems for depression are rarely used in NHS* [National Health Service] *clinical practice*'.[1] The clinical trials for ESK-NS used the clinician-reported Montgomery-Åsberg Depression Rating Scale (MADRS) and the patient-reported outcome Patient Health Questionnaire-9 questions (PHQ-9) to measure the severity of depressive episodes. The company reported that '*feedback from NICE early scientific advice was that "the MADRS score is appropriate to measure outcomes in the ESK-NS clinical trials"*'.[1]

2.3 Critique of company's overview of current service provision

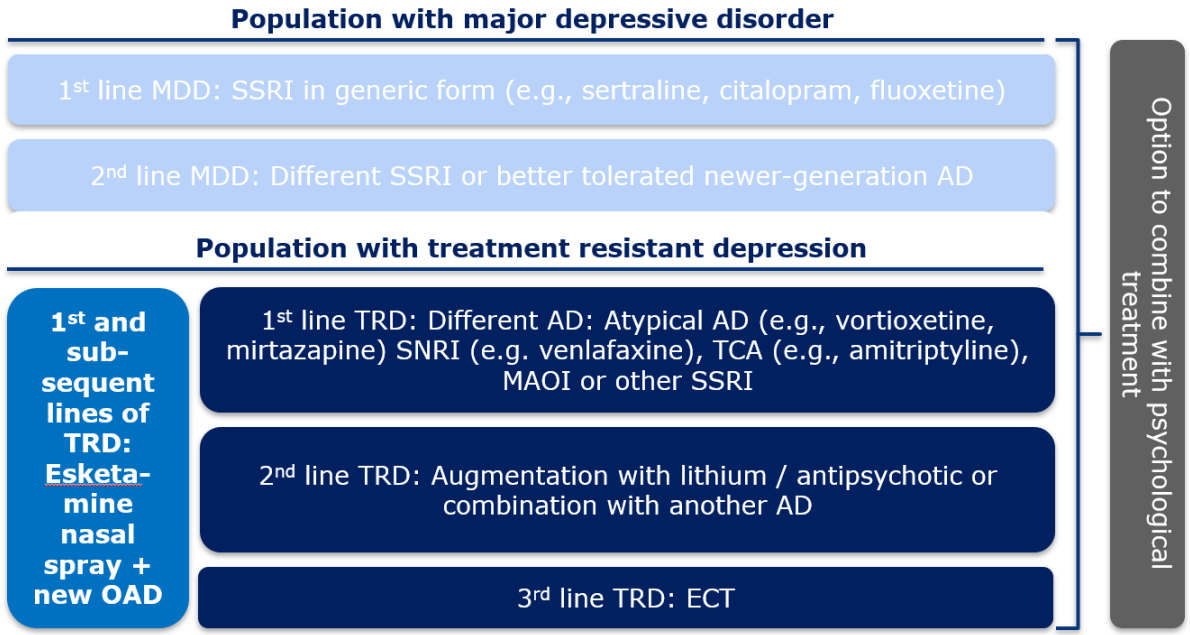
The company correctly stated that there are no UK guidelines specific to TRD. The main relevant guideline is NICE clinical guideline CG90 which covers the recognition and management of depression in adults.[6] This guideline along with that of the British Association for Psychopharmacology were

described in the CS.[7] The company also referenced American Psychiatry Association (APA) Practice Guidelines for the treatment of Patients with Major Depressive Disorder.[8]

Currently, the first-line treatment for MDD is an OAD, typically a selective serotonin reuptake inhibitor (SSRI). After four weeks, if response is inadequate or due to patient preference, a switch to another OAD is recommended. NICE recommends initially a different SSRI or a better tolerated newer-generation antidepressant but recognises the weakness of the evidence of any advantage switching either within or between classes. NICE subsequently advises an antidepressant of a different pharmacological class that may be less well tolerated, for example venlafaxine, a tricyclic antidepressant (TCA) or an monoamine oxidase inhibitor (MAOI).[6]

It is at this third-line and beyond (or first-line treatment-resistant and beyond) that ESK-NS is to be placed and should be taken alongside a new OAD according to the CS, see Figure 2.1.[1] In response to request for clarification, the company advised that the label indication is expected to change to ESK-NS in combination with an SSRI or SNRI for treatment-resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode.[3] At this stage, ESK-NS is a comparator for other treatments including atypical antidepressants (ADs), serotonin–norepinephrine reuptake inhibitor (SNRI), TCA, MAOI or other SSRI and for augmentation with either lithium or other antipsychotic and electroconvulsive therapy (ECT).

Figure 2.1: Proposed future MDD and TRD treatment pathway



Based on Figure 6 of the CS[1]
AD = antidepressant; ECT = electroconvulsive therapy; MAOI = monoamine oxidase inhibitor; MDD = major depressive disorder; NICE = National Institute for Health and Care Excellence; OAD = oral antidepressant; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TCA = tricyclic antidepressant; TRD = Treatment-resistant depression

The company summarised the currently used treatments for patients with TRD and their limitations. The problem of delay in response to OADs (four to six weeks) was particularly highlighted by the company. In this context, the CS described ESK-NS as having a '*unique mechanism of action which results in a rapid onset of action (within 24 hours)*'.[1]

Regarding the introduction of ESK-NS, the company noted that '*it is expected to diminish the need for combination and augmentation strategies in addition to invasive non-pharmacological treatments that are associated with an increased side effect burden in later lines*'.[1]

The CS advised that '*during and after ESK-NS administration at each treatment session, patients should be monitored for sedation and dissociation until the patient is stable and ready to leave the clinic based upon clinical judgement. While ESK-NS could potentially be used in all lines of treatment, the suitability should be addressed by a specialist in mental health and the setting needs to be appropriate to allow for the required observation and monitoring period*'.[1]

The company highlighted geographic access as a consideration in relation to equality.

ERG comment: The overview of the current pathway for TRD, presented in the CS, was appropriate. The pathway shows that there are a number of possible comparators. The ERG noted in particular that a NICE appraisal of vortioxetine has been conducted (technology appraisal (TA) 367) and that vortioxetine is recommended for adults with major depressive episodes whose condition has responded inadequately to two OADs within the current depressive episode, see Figure 2.1.[9] The company were unable to conduct a direct or indirect comparison of ESK-NS and vortioxetine to inform the decision problem.

NICE recommends that for relapse prevention patients who respond to treatment should continue to take their OAD for at least six months after remission. For those at high risk of relapse, OAD should be continued for at least two years with a re-evaluation to assess if maintenance should continue.[6]

The company was asked to provide a breakdown of how long people in clinical practice might be expected to take esketamine in an acute phase and in the maintenance phase. In response, the company stated that '*in the acute treatment phase, patients are expected to receive ESK-NS + OAD for 4 weeks, and patients who do not respond and / or reach remission at that time point are expected to discontinue treatment*'.[3] They further stated that '*SUSTAIN-1 data on relapse among stable remitters indicated that a patient with TRD needed to be in relapse-free remission for 36 weeks (approximately nine months) to achieve recovery. (…) Once entering the maintenance phase, a benefit of ESK-NS is that it can be discontinued while patients can still receive OAD for recurrence prevention. A total of 35.4% of patients were assumed to stop ESK-NS immediately upon achieving recovery (…) For the remainder of patients, treatment with ESK-NS + OAD will be continued during the maintenance phase and discontinued over time. Based on UK expert opinion, a 4-week discontinuation risk of 25% for ESK-NS + OAD was used during recovery. (…) Patients who achieve response (without remission) are assumed to continue ESK-NS + OAD as long as they are in the response health state and have not reached remission, as they are assumed to be at high risk of relapse'*.[3] The company stated that their assumptions were discussed with UK clinical experts and considered to be representative of clinical practice. The implications of these assumptions are discussed within this report.

The company advised that suitability for ESK-NS should be addressed by a specialist in mental health. However in the CS, the company stated that *'only an estimated 10% of patients with TRD are referred to specialist mental health services (generally those deemed to be at risk of suicide)*'.[1] Furthermore the survey by SANE commented that just over a half of respondents had been seen by a psychiatrist with an average of a three year wait.[2]

Administration of ESK-NS requires observation by a healthcare professional due to potential adverse effects and driving is not permitted until the next day after a restful sleep. This has implications for resourcing and for patients. Implications of resourcing are discussed within this report. In terms of

patients, the ERG received the following information from the mental health charity SANE in their submission: '*The main advantage of the administration method would be contact with healthcare practitioners who might be able to give additional support in managing the patient's depression and encourage greater compliance with medication. The disadvantages to patients include the costs of travel and the time involved and difficulties in accessing clinic for patients such as those with mobility problems, agoraphobia or those in a care home. Further disadvantages could be the risk of disassociation after administration thus requiring input from carers*'.[2]

3. Critique of company's definition of decision problem

Table 3.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the CS and rationale	Rationale if different from the final NICE scope	ERG comment
Population	Adults with treatment-resistant depression	The population would be more appropriately defined as: *"Adults with treatment resistant MDD who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode"*.	The proposed wording reflects the expected population in the marketing authorisation.	In line with the scope. However, trial results might not be applicable regardless of severity. Also, the trials in the economic model included only those aged 18 to 64 years. The trials in the CS excluded patients with moderate/severe alcohol abuse according to DSM-5 criteria. The trials in the CS also excluded patients who had suicidal/ homicidal ideation/intent within 6 months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening.
Intervention	ESK-NS in addition to established clinical management	ESK-NS co-administered with a newly initiated OAD.	In response to clarification, the company advised that the label indication is expected to change to *'ESK-NS in combination with an SSRI or SNRI, is indicated for adults with treatment-resistant major depressive disorder, who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode'*.[3]	In line with the scope. However, the potential impact of a number of issues of applicability discussed within this report. These include the effectiveness of different types of OADs, the complex dosing of ESK-NS and the assumption that at some point patients can discontinue ESK-NS with no reduction in efficacy.

	Final scope issued by NICE	Decision problem addressed in the CS and rationale	Rationale if different from the final NICE scope	ERG comment
Comparator(s)	• SSRIs • TCAs • MAOIs • SNRIs • Vortioxetine • Combination or augmentation treatments (with lithium or an antipsychotic) • ECT • Best supportive care	As per the scope, plus the tetracyclic OAD mirtazapine.	Mirtazapine is currently not included in the final scope. Mirtazapine should be included as a comparator as two retrospective database analyses conducted by 1) King's College London, using secondary data from the South London and Maudsley (SLaM) Trust, and 2) IQVIA, using *Longitudinal Patient Data*, a primary care prescription data set, which show that mirtazapine is amongst the five most frequently prescribed treatments for TRD.[10, 11] NICE stated in their early scientific advice in 2013 and at the NICE Scoping Workshop for ESK-NS in TRD held on 17 September 2018 that RWE will determine which comparators are the most relevant ones.[12] Figure 5 [of the CS] shows the most frequently used OAD therapies for TRD in the UK. Of the list of comparators in the final scope, it shows that SSRIs, TCAs, SNRIs, and mirtazapine are the most relevant comparators.	The trials included in the CS compared ESK-NS + OAD + placebo and OAD. The implications of adjustments made for the high placebo response are discussed within this report.
Outcomes	• Response to treatment (including response rate and time to response)	As per the scope, with the addition of the impact of ESK-NS on indirect costs and carer HRQoL.	TRD-associated disability has been associated with substantial indirect costs. In a systematic literature review, Johnston et al. 2019[13] found that increasing treatment resistance	

	Final scope issued by NICE	Decision problem addressed in the CS and rationale	Rationale if different from the final NICE scope	ERG comment
	• Relapse (including relapse rate and time from remission to relapse) • Severity of depression • Cognitive dysfunction • Remission of symptoms • Anxiety • Sleep quality • Hospitalisation • Functioning and associated disability • Mortality • Adverse effects of treatment (including adverse effects of treatment discontinuation) • HRQoL		was associated with higher costs, reduced HRQoL and decreased health status.[13] In addition, McCrone et al. 2018 showed that 80% of the total UK society burden of TRD was due to lost productivity and carer burden.[14] NICE CG90 states that *"depression incurs significant non-healthcare costs such as social service costs, direct costs to patients and their families, and lost productivity costs due to morbidity or premature mortality"*.[6] Consideration of the wider indirect cost impact is in line with NICE social values which state that: *"Decisions about whether to recommend interventions should not be based on evidence of their relative costs and benefits alone. NICE must consider other factors when developing its guidance, including the need to distribute health resources in the fairest way within society as a whole"*.[15] Additionally, the feedback from NICE at the early scientific advice meeting was that *"Workplace productivity and occupational functioning should not currently be included in the base-case of the economic model however such data*	

	Final scope issued by NICE	Decision problem addressed in the CS and rationale	Rationale if different from the final NICE scope	ERG comment
			could be presented as supporting evidence".[12]	
Subgroups to be considered	If evidence allows the following subgroups will be considered by severity of the condition in people with treatment-resistant depression. In addition, the clinical and cost effectiveness of ESK-NS may be considered in different positions in the treatment pathway.	No subgroup analyses based on level of severity at baseline or ESK-NS in different positions in the treatment pathway.	There is insufficient comparative evidence to evaluate the effectiveness of ESK-NS by level of severity or positioning in the treatment pathway. Therefore, ESK-NS plus OAD has been considered in the full label population, as per the clinical trials and anticipated license indication.	Some subgroup data on severity of disease were provided in response to the request for clarification, see section 4.2.6.
Special considerations including issues related to equity or equality		In relation to equality, Janssen would like to highlight geographic access as a key consideration. Additionally, there may be an equality consideration for patients aged ≥65 years.		The ERG agrees that, given the method of administration of ESK-NS requiring supervision by a healthcare professional, it will be important to ensure that access to healthcare support will not inappropriately discriminate against individuals for whom geography may pose a challenge. The main trials included only those aged 18 to 64 years. The main trials in the economic model included only those aged 18 to 64 years.

Based on Table 1 of the CS[1]

C-SSRS = Columbia-Suicide Severity Rating Scale; CG = clinical guideline; CS = company submission; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders – Fifth edition; ECT = electroconvulsive therapy; ERG = Evidence Review Group; ESK-NS = esketamine nasal spray; HRQoL = health-related quality of life; MAOI = monoamine oxidase inhibitor; MDD = major depressive disorder; NICE = National Institute for Health and Care Excellence; OAD = oral antidepressant; RWE = real-world evidence; SLaM = South London and Maudsley; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TCAs = tricyclic antidepressants; TRD = treatment-resistant depression; UK = United Kingdom

3.1 Population

The population defined in the scope is adults with treatment-resistant depression, i.e. people who do not respond to at least two therapies. The population is broadly consistent with the NICE scope and the expected marketing authorisation.[3, 16] However, the scope does not specify severity. Also, subgroup analysis reveals that severity as measured by functional impairment in terms of Sheehan Disability Scale (SDS) does seem to have an impact on the effectiveness of ESK + OAD (see Appendix E of the CS).[17]

The two main trials which informed the economic modelling were TRANSFORM-2 and SUSTAIN-1. TRANSFORM-2, an acute treatment study, enrolled adults with a history of non-response to at least two antidepressants in the current episode with one antidepressant assessed prospectively. SUSTAIN-1 assessed maintenance of effect (prevention of relapse).

The main limitation of these trials in terms of this appraisal is that they only included patients aged 18 to 64 years of age. A four-week trial in adults aged 65 and over (TRANSFORM-3) was included in the CS only as supporting evidence and did not inform the economic model. The ERG was, therefore, concerned as to the relevance of evidence to the older population. The company was asked to clarify if they considered the trials to be applicable to patients aged 65 years and over. In response, the company presented results of patients aged 65 to 74 years from TRANSFORM-3 showing them to be similar in magnitude to those in the younger adult population; the lower effect noted in those aged 75 years and over was considered to be an artefact of the low number of patients (n=22).[3] However, for response and remission, the results for TRANSFORM-3 were much lower. Day 28 risks of remission and response (ESK + OAD vs. OAD + PBO-NS) were: 69.3% vs. 52.0% and 52.5% vs. 31.0% for TRANSFORM-2. For TRANSFORM-3 these were: 27.0% vs. 13.3% and 17.5% vs. 6.7%, respectively. As can be seen, the risk differences were also lower for TRANSFORM-3 suggesting that, although ESK + OAD was still effective, its effectiveness was not only lower in absolute terms, but lower relative to OAD. The dose of ESK was also lower in TRANSFORM-3. Indeed, whilst the company argued that TRANSFORM-2 was representative of the population in the scope, they also argued in Section B 3.5.1 of the CS that TRANSFORM-2 and TRANSFORM-3 could not be pooled, partly because of differential efficacy which they explained in terms of difference in age and dose.[1] On this basis, the ERG questions the applicability of TRANSFORM-2 to the whole population. Also, there is no equivalent study to SUSTAIN-1 in the older age group by which comparisons might be made. SUSTAIN-2 included older patients, but relapse was not measured and no separate subgroup analysis was provided.[1, 17]

The company also submitted a new version of the base-case model to include acute response and remission transition probabilities and utilities for MDE, response and remission/recovery states from both TRANSFORM-2 and TRANSFORM-3, weighted by percentage in each age group such that if set to 0% for age >65 years one gets the same result as in the original base-case. Section 5.2.3 discusses this in more detail.

The trials in the CS excluded patients with moderate/severe alcohol abuse according to Diagnostic and Statistical Manual of Mental Disorders – Fifth edition (DSM-5) criteria. The committee will need to consider whether evidence in the CS on effectiveness and safety of ESK-NS can be generalised to those with a dual diagnosis of depression and alcohol misuse.

The trials in the CS also excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening. Again, the committee will need to consider if

the evidence in the CS on effectiveness and safety of ESK-NS can be generalised to this vulnerable population.

3.2 Intervention

The intervention in the NICE scope is ESK-NS in addition to established clinical management. In the trials ESK-NS is co-administered with a newly initiated OAD according to the expected licence. According to the CS, a Committee for Medicinal Products for Human Use (CHMP) positive opinion is expected in September 2019 with marketing authorisation anticipated to be granted by the European Commission in November 2019.[1] The anticipated indication was given in the CS is as follows:

- ESK-NS is indicated for treatment-resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode.
- ESK-NS must be co-administered with a newly initiated OAD therapy.

In response to request for clarification, the company advised that the label indication is expected to change to ESK-NS in combination with an SSRI or SNRI for treatment-resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode.[3] In the main trials TRANSFORM-2 and SUSTAIN-1 over 60% were prescribed a SNRI and the remainder a SSRI. The OAD as assigned by the investigator could be one of four: duloxetine, escitalopram, sertraline or venlafaxine XR. The company also confirmed that no OADs are contraindicated with ESK-NS. However, when ESK-NS is to be given with MAOIs blood pressure may be increased and would require close monitoring.[3]

The company stated that '*ESK-NS comes as a single-use device that delivers a total of 28 mg of esketamine in two sprays (one spray per nostril). ESK-NS is self-administered and is to be used under the supervision of a healthcare professional. One device (for a 28 mg dose), two devices (for a 56 mg dose), or three devices (for an 84 mg dose), are to be used, with a five-minute interval between each nasal spray self-administration*'.[1]

The company provided the following information on dosing:

'Induction phase dosing: In weeks 1–4, patients start on 56 mg (<65 years) or 28 mg (≥65 years) on Day 1. Subsequent doses are 56 or 84 mg twice a week. Dose adjustments should be made based on efficacy and tolerability. Evidence of therapeutic benefit should be evaluated at the end of the induction phase to determine need for continued treatment.

Maintenance phase dosing: It is recommended to maintain the dose the patient receives at the end of the induction phase in the maintenance phase. In weeks 5-8, 56 mg or 84 mg once weekly. From Week 9, 56 mg or 84 mg every 2 weeks or once weekly.

The need for continued treatment should be re-examined periodically'.[1]

The issue of how well dosing in the trials might reflect dosing in clinical practice is discussed in section 5.2.4.

The company did not use the TRANSFORM-1 trial in the economic modelling of this submission. This trial was similar to TRANSFORM-2 in that ESK-NS + OAD were compared to OAD + PBO-NS twice weekly for four weeks. However the company stated that '*ESK-NS was administered as a fixed dose which is not in line with the expected licence*'.[1]

The company stated that after depressive symptoms improve, treatment is recommended for at least six months. The ERG asked the company to provide a breakdown of how long people in clinical practice might be expected to take esketamine in an acute phase and in the maintenance phase.[18] In response, the company stated that '*in the acute treatment phase, patients are expected to receive ESK-NS + OAD for 4 weeks, and patients who do not respond and / or reach remission at that time point are expected to discontinue treatment*'.[3] The response further stated that '*SUSTAIN-1 data on relapse among stable remitters indicated that a patient with TRD needed to be in relapse-free remission for 36 weeks (approximately nine months) to achieve recovery (…) Once entering the maintenance phase, a benefit of ESK-NS is that it can be discontinued while patients can still receive OAD for recurrence prevention. A total of 35.4% of patients were assumed to stop ESK-NS immediately upon achieving recovery (…) For the remainder of patients, treatment with ESK-NS + OAD will be continued during the maintenance phase and discontinued over time. Based on UK expert opinion, a 4-week discontinuation risk of 25% for ESK-NS + OAD was used during recovery.* (…) *Patients who achieve response (without remission) are assumed to continue ESK-NS + OAD as long as they are in the response health state and have not reached remission, as they are assumed to be at high risk of relapse*'.[3] The company stated that their assumptions were discussed with UK clinical experts and considered representative of clinical practice. The implications of these assumptions are discussed within this report, see section 5.2.6.2. The company's advisory board agreed that [redacted].[3]

A key difference between ESK-NS and other antidepressants is that, although it is self-administered, this needs to be done under the supervision of a healthcare professional. The company stated that *'during and after ESK-NS administration at each treatment session, patients should be observed for sedation and dissociation until the patient is stable based on clinical judgment. In the SUSTAIN-2 trial, approximately 60% of individuals were ready to leave after 1 hour, with approximately 95% ready to leave after 90 minutes'*.[1] The company's own advisors agreed that [redacted].[3]

In addition to this supervision patients will need to be aware that after taking ESK-NS according to the CS '*driving is not permitted until the next day after a restful sleep*'.[1]

The company acknowledged the potential of ESK-NS for abuse, misuse, and diversion due to its similar pharmacologic profile to ketamine. They stated that the controlled distribution model was intended to limit diversion. They further stated that *'during clinical development trials of ESK-NS, the percentage of nasal spray kits that were not returned from the clinical sites was 0.004% (5 of 141,561 kits)'* and that *'there were no reports of overdose, drug abuse, or confirmed diversion of drug product across the clinical development programme'*.[1]

3.3 Comparators

The main trials in this appraisal compared ESK-NS + OAD to placebo nasal spray (PBO-NS) and OAD. The company submission stated that "*efficacy estimates (response and remission) for the OAD + PBO-NS arm of the TRANSFORM-2 trial were high compared with other studies in TRD*" and on this basis the response rate was adjusted down for PBO-NS.[1] The company attributed this to the high number of clinic visits.

The ERG is concerned that any placebo effect (due to clinic visits or for any other reason) was likely to be present in both trial arms. Therefore, only removing the placebo effect for OAD + PBO while not

removing it for ESK would likely overestimate the ESK treatment benefit. The company was asked to use the unadjusted estimates of response for OAD + PBO-NS for the model base case or perform the same adjustment to ESK-NS + OAD.[18] The company provided these data which are detailed in this report, see section 4.2.5 for these results and section 5.2.6.1 for a detailed discussion of this issue.[3]

3.4 Outcomes

The NICE final scope listed the following outcomes:

- Response to treatment (including response rate and time to response)
- Relapse (including relapse rate and time from remission to relapse)
- Severity of depression
- Cognitive dysfunction
- Remission of symptoms
- Anxiety
- Sleep quality
- Hospitalisation
- Functioning and associated disability
- Mortality
- Adverse effects of treatment (including adverse effects of treatment discontinuation)
- Health-related quality of life (HRQoL)

The outcomes investigated in the CS reflected the scope. However, some outcomes defined in the final scope issued by NICE were not reported for the main trials in the CS, namely cognitive dysfunction, hospitalisation and sleep quality.

3.5 Other relevant factors

No patient access scheme was presented in the CS. The ERG agrees that, given the method of administration of ESK-NS requiring supervision by a healthcare professional, it will be important to ensure that access to healthcare support will not inappropriately discriminate against individuals for whom geography may pose a challenge.

4. Clinical effectiveness

4.1 Critique of the methods of review(s)

4.1.1 Searches

Appendix D of the CS[17] reported search methods for two systematic literature reviews (SLRs):

- Systematic literature review of acute management of patients with TRD
- Systematic literature review of ongoing maintenance treatment of patients with TRD

Section D1.1 of the CS details a systematic search of the literature used to identify evidence reporting on the efficacy and safety of esketamine and its comparators. Searches were undertaken on 14 July 2017 and updated on 10 May 2019. A summary of the sources searched is provided in Table 4.1.

Table 4.1: Data sources for the clinical effectiveness systematic review for the acute management of patients with TRD

Resource	Host/Source	Date Range	Date searched
Electronic databases			
Medline	OVID	1990-2017/07/14	14/7/17 (Updated 10/5/19)
Epub ahead of print[a]			
Medline In-Process & Other Non-Indexed Citations			
Medline Daily Update			
Embase		1990-2017/07/14	
PsycINFO		1990-2017/07/14	
Cochrane CENTRAL	EBM Reviews via OVID	Up to 14th July 2017	
CDSR			
DARE			
HTA Database			
ACP Journal Club			
Cochrane clinical answers			
Cochrane methodology register			
NHS EED			
Conference proceedings[b]			
Anxiety and Depression Association of America Conference		2016-2019	31/10/18 (updated 24/5/19) Unable to access abstracts
International Conference on Management of Depression		2016-2019	31/10/18 (updated 24/5/19) Unable to access abstracts

Resource	Host/Source	Date Range	Date searched
American Psychiatry Association Annual Meeting		2016-2019	1/11/18 (updated 23/5/19)
European Congress of Psychiatry		2016-2019	5-6/11/18 (updated 23/5/19)
The Royal College of Psychiatrists International Congress		2016-2019	6/11/18 (updated 24/5/19) Unable to access abstracts
WPA World Congress of Psychiatry		2016-2019	6/11/18 (updated 23/5/19) Unable to access abstracts for 2017-19
ISPOR (USA/Europe)		2016-2019	23/5/19
HTA agencies[b]			
NICE, SMC, PBAC, CADTH, NCPE			30/05/2019
Trials registries[b]			
ClinicalTrials.gov			Not reported
EUCTR			Not reported
WHO ICTRP			10/5/19

[a] Whilst Medline epub ahead of print was included in the resources listed for the 2019 update, it was unclear if it had been included in the original searches; [b] Studies identified were considered for inclusion in either the acute or maintenance treatment categories, respectively
ACP = American College of Physicians; CADTH = Canadian Agency for Drugs and Technologies in Health; CDSR = Cochrane Database Systematic Reviews; DARE = Database of Abstracts of Reviews of Effects; EBM = evidence-based medicine; EED = Economic Evaluation Database; EUCTR = European Union Clinical Trials Register; HTA = Health Technology Assessment; ICTRP = International Clinical Trials Registry Platform; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NCPE = National Centre for Pharmacoeconomics; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; PBAC = Pharmaceutical Benefits Advisory Committee; SMC = Scottish Medicine Consortium; TRD = treatment-resistant depression; USA = United States of America; WHO = World Health Organization; WPA = World Psychiatric Association

Section D1.2 of the CS details a systematic search of the literature used to identify evidence reporting on the efficacy and safety of therapies used in the maintenance treatment of TRD.[17] Searches were undertaken on 01 February 2017 and updated on 23 May 2019. A summary of the sources searched is provided in Table 4.2 below.

Table 4.2: Data sources for the clinical effectiveness systematic review for the ongoing maintenance of patients with TRD

Resource	Host/Source	Date Range	Date searched
Electronic databases			
Medline	OVID	1946-2017/02/1	1st Feb 2017 (Updated 23/5/19)
Epub ahead of print*			
Medline In-Process & Other Non-Indexed Citations*			
Medline Daily Update*			

Resource	Host/Source	Date Range	Date searched
Embase		1974-2017/01/30	
Cochrane CENTRAL	EBM Reviews via OVID	Up to 2017/02/1	
CDSR			
DARE			
HTA Database			
ACP Journal Club			
Cochrane clinical answers			
Cochrane methodology register			
NHS EED			

* Whilst listed in the 2019 update searches, it was not clear from reporting whether these additional Medline in process resources were included in the original searches.
CDSR = Cochrane Database Systematic Reviews; DARE = Database of Abstracts of Reviews of Effects; EBM = evidence-based medicine; EED = Economic Evaluation Database; HTA = Health Technology Assessment; NHS = National Health Service; TRD = treatment-resistant depression

ERG comment:

- During clarification, the company confirmed that both sets of searches (Tables 4.1 and 4.2) were screened for papers relevant to both SLRs: *"…during screening for either the acute or maintenance treatment SLRs, any studies that were potentially relevant for inclusion in the other review were flagged and assessed for eligibility"*.[3]
- The selection of databases searched was comprehensive, and the majority of searches were clearly reported and reproducible. The database name, host and date searched were provided. An extensive range of resources additional to database searching were included in the SLR to identify further relevant studies and grey literature. Missing data regarding the clinical trials registry searches were queried at clarification.[18] The ERG noted that searches were reported in sections D1.1 and 1.2 for Clinical Trials.gov and the EU Clinical Trials registry and asked for full details of all search dates and search strategies used.[18] In their response, the company failed to provide full details for the searches listed above but instead provided search dates and strategy for an additional search of the WHO ICTRP (World Health Organization - International clinical trials registry platform).[3] Although this omission may affect reproducibility, it is unlikely to affect the overall recall of results.
- The ERG noted that a randomised controlled trial (RCT) filter was applied to the Cochrane library searches. The MECIR (Methodological Expectations of Cochrane Intervention Reviews) Manual advises *"…do not use filters in pre-filtered databases e.g. do not use a randomized trial filter in CENTRAL or a systematic review filter in DARE"*.[19] The inclusion of these filters may result in unnecessarily restricting the results retrieved. However, given the breadth of the searches reported, this is unlikely to have impacted on the overall recall of results.
- There were some limitations with the use of MeSH indexing terms in the Embase search for acute management of TRD. Although some automated mapping between indexing terms does take place it is possible that relevant Emtree indexing terms were not included in the search, and potentially relevant records could have been missed. Given the additional use of free text terms this is unlikely to have affected the overall recall of results.
- The ERG was concerned that limiting the searches reported in sections D1.1 and 1.2 to English language may have introduced potential language bias.[17] Current best practice states that *"whenever possible review authors should attempt to identify and assess for eligibility all possibly relevant reports of trials irrespective of language of publication"*.[20]

- Whilst not reported in the submission, the company confirmed that reference checking was performed as part of both the original and update searches.[3]
- The ERG noted that the strategies in section D1.2 (Ongoing maintenance treatment of patients with TRD) appeared to include a reduced interventions facet compared with that used in section D1.1 for acute treatment, further to this not all of the drugs listed in Table 5 of the CS (Eligibility criteria) appeared in the strategies (missing drugs included reboxetine, butriptyline, clomipramine etc.). Whilst there were some limited free text terms for the drug types of interest (see Embase strategy line #72), the ERG was unsure of the rationale behind this decision and what impact it may have had on the overall recall of results. Whilst this omission was not directly addressed in their response, the company did clarify that both sets of searches reported in section D were screened for papers relevant to both SLRs.[3] Without rerunning the searches, the ERG is unable to confirm what impact this may have had on the overall recall of results, however this approach may have mitigated against some loss of recall.
- The ERG queried whether any additional searches were conducted for non-RCTs, in response the company reported that *"an SLR was conducted (December 2018) interrogating the same electronic databases as the clinical SLRs. A bespoke search strategy using a validated search filter to identify observational studies was employed"*.[3] Full search strategies were provided for the resources listed in Table 4.3.
- The company confirmed at clarification that the searches reported in sections D1.1 and 1.2 were intended the inform section B2.10 (adverse events).[3] While the searches outlined would have retrieved some relevant information in these areas, the addition of a trials filter may have resulted in relevant references being missed. Guidance by the Centre for Reviews and Dissemination (CRD) recommends that if searches have been limited by a study design filter, additional searches should be undertaken to ensure that adverse events that are long-term, rare or unanticipated are not missed.[21] The searches for observational studies sent at clarification may have mitigated against this loss of recall, although it is unclear whether these searches were screened for adverse events.

Table 4.3: List of resources for which full search strategies were provided

Resource	Host/Source	Date Range	Date searched
Electronic databases			
Medline	OVID	1990-2018/12/17	18/12/18
Epub ahead of print			
Medline In-Process & Other Non-Indexed Citations			
Medline Daily Update			
Embase		1990- 2018/12/17	
Cochrane CENTRAL	EBM Reviews via OVID	Up to 2018/12/17	
CDSR			
DARE			
HTA Database			
ACP Journal Club			
Cochrane clinical answers			
Cochrane methodology register			
NHS EED			

Resource	Host/Source	Date Range	Date searched
PsycINFO	OVID	1990- 2018/12/wk2	19/12/18
ACP = American College of Physicians; CDSR = Cochrane Database Systematic Reviews; DARE = Database of Abstracts of Reviews of Effects; EBM = evidence-based medicine; EED = Economic Evaluation Database; HTA = Health Technology Assessment; NHS = National Health Service			

4.1.2 Inclusion criteria

The company reported on two different SLRs performed to identify evidence reporting on data relevant for: 1) the acute management of patients with TRD; and 2) the ongoing maintenance treatment of patients with TRD.

Population:

For the acute management SR, the company reported that the population of interest was adults (18 years or older) with TRD (defined as unipolar MDD with failure to respond to ≥2 antidepressant treatment regimens of adequate dose and duration *in the current episode*). However, due to inconsistent reporting in this research field, the definition of current or prior episode was not used as an inclusion or exclusion criteria for study selection. A step-wise procedure was used at the full-text screening stage. At first pass, studies that included patients with ≥1 treatment failure were included, with no exclusions as to whether treatment failures occurred during the current or prior episode in anticipation of subgroup results for the required ≥2 treatment failure population. However, no information was provided as to how or at what stage of the process the second (or more) pass selection process was applied. This was not the population defined in the scope, which had a broader definition of *'adults with treatment-resistant depression'*.[16] The company justified their use of this narrower population by indicating that this reflects the expected marketing authorisation of esketamine.

For the ongoing maintenance SR, the company reported that the population of interest was adults (18 years or older) with TRD (by any definition). This was a broader population than reported for the acute management SR, and was in line with the scope.

Interventions and comparators:

For the acute management SR, the company included all the classes of medications indicated in the scope (SSRIs, TCAs, MAOIs, SNRIs, vortioxetine, augmentation treatments (with anti-psychotics), combination treatments (with lithium), electroconvulsive therapy (ECT) and best supportive care (BSC). They also included three additional classes of comparators that were not specified in the scope: SARIs (serotonin antagonist and reuptake inhibitors; trazodone), NRIs (norepinephrine reuptake inhibitor; reboxetine) and TeCAs (tetracyclic antidepressants; amoxapine, maprotiline, mianserin, mirtazapine, setiptiline). Several of these are not considered common OAD medications in the UK, and as such the inclusion of such comparators may skew any resulting data away from the standard UK perspective.

For the ongoing maintenance SR, the company included all the classes of medications indicated in the scope (SSRIs, SNRIs, TCAs, MAOIs, vortioxetine, augmentation treatments (with anti-psychotics), augmentation treatments (with lithium), ECT and BSC. They also included two additional classes of comparators that were not specified in the scope: SARIs (trazodone) and NRIs; and also included no therapy as a comparator. The company did not include TeCAs, amoxapine, maprotiline, mianserin or septiptiline, which were named drugs included in the acute management SR.

Outcomes:

For the acute management SR, the company included depressive symptoms (based on change in any depression rating scale, such as Hamilton Depression Rating Scale (HAM-D) or Montgomery-Åsberg Depression Rating Scale (MADRS)), response rate, relapse rate, remission rate, time to response, time to remission, mortality and discontinuation due to adverse events, all of which were in line with the scope.

Additionally, the CS included recurrence rate, suicide behaviour/ attempts and suicidal ideation. Conversely, the company failed to include some outcomes specified in the scope, namely cognitive dysfunction, anxiety, sleep quality, hospitalisation, functioning and associated disability, adverse events related to treatment discontinuation and HRQoL.

Studies that reported only on adverse events were excluded under the inclusion/exclusion criteria of the systematic review, which means that some relevant studies may have been missed. In the decision problem, the company also state that they have included two additional outcomes: impact of ESK-NS on costs and carer-related HRQoL. These outcomes did not appear to have been included nor identified within the framework of the SR.

For the ongoing maintenance systematic review, the company included: depressive symptoms (based on change in one of five named depression rating scales: MADRS, Quick Inventory of Depressive Symptomatology (QIDS-SR14), Clinical Global Impression – Severity (CGI-S), Patient Global Impression – Severity (PGI-S) and HAMD/HDRS), onset of clinical response, remission, relapse and HRQoL (PHQ9 and Quality of Life in Depression Scale (QLDS)), all of which were in line with the scope. They additionally included: recurrence, discontinuation, discontinuation due to adverse events, European Quality of Life-5 Dimensions (EQ-5D) and health resource utilisation information. However, the company failed to include some outcomes specified in the scope, namely response rate, time from remission to relapse, cognitive dysfunction, anxiety, sleep quality, hospitalisation, functioning and associated disability, mortality, adverse events and adverse events related to treatment discontinuation.

Study design:

For the acute management SR, the company only included RCTs that reported on the efficacy and safety of acute interventions with ≤4 weeks of follow-up data. This restriction based on follow-up time was considered by the ERG to be inappropriate. While several of the company's own trials reported a core treatment period of four weeks, the maintenance and post-treatment follow-up phases are much longer than this (up to 24 weeks), and therefore other trials with longer follow-up periods may represent relevant comparator datasets.

For the ongoing maintenance SR, the company only included RCTs with either >4 weeks of treatment or maintenance treatments >4 weeks or treatment explicitly for relapse prevention (presumably of any duration, since this was not specified).

Study selection:

Across both SRs, two reviewers were involved in study selection, and any discrepancies were resolved by the intervention of a third reviewer. This was considered sufficient to minimise bias in study selection. In the ongoing maintenance SR, studies were restricted based on language (only English language studies were included), meaning relevant studies may have been missed. The company was asked to clarify why a date limit of 1990 had been applied to searching/screening for the systematic review.[18] The response stated that this was '*based on internal clinical expert opinion that TRD-related publications started in the early 1990s. The 1990 date limit was therefore applied to ensure that the*

current standard of depression treatment was captured'.[3] No references were supplied to support this perspective. It is not normally recommended in systematic reviews to set arbitrary date limits in case relevant studies are missed.

4.1.3 Critique of data extraction

No information was provided on the number of reviewers involved in the data extraction process, therefore reviewer error and bias cannot be ruled out.

4.1.4 Quality assessment

Quality was assessed for the two RCTs that informed the economic model (TRANSFORM-2 and SUSTAIN-1) and two further RCTs (TRANSFORM-1 and TRANSFORM-3) using the NICE recommended tool.[22] This was considered a sufficient tool to use.

The open-label extension study, SUSTAIN-2, was assessed using a different set of signalling questions to the four RCTs, which was appropriate given the difference in study design; however, the company did not report the tool that was used. It appeared to the ERG that most of the signalling questions were based on a reporting guideline rather than a risk of bias assessment, and as such, this was probably an inappropriate tool to use.

No information was provided on the number of reviewers who were involved in the quality assessment, therefore reviewer error and bias cannot be ruled out.

4.1.5 Evidence synthesis

The company performed a feasibility assessment of the n=68 citations (Figure 1 and Figure 2 in Appendix D[17]) identified by their acute phase systematic literature searches and concluded that limited NMA could be conducted.

The company also performed a feasibility assessment of the n=49 citations (Figure 3 and Figure 4 in Appendix D[17]) identified by their maintenance phase systematic literature searches, and concluded that a network meta-analysis could not be conducted. However, it was not clear why this was the case, and no supporting network diagrams or study details were provided. Further details of the NMA are provided in section 4.4.

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

4.2.1 Details of included studies

Six studies formed the evidence base for ESK-NS (Table 4.4). Four of these were randomised controlled trials (TRANSFORM-1, TRANSFORM-2, TRANSFORM-3, SUSTAIN-1) and two were open label extension studies (SUSTAIN-2, SUSTAIN-3).[1]

Table 4.4: Summary of clinical effectiveness evidence for esketamine

	TRANSFORM-1	TRANSFORM-2	TRANSFORM-3	SUSTAIN-1	SUSTAIN-2	SUSTAIN-3
In economic model	No	Yes	No	Yes	No	No
Rationale for use/non-use in economic model	ESK-NS was administered as a fixed dose which is not in line with the anticipated licence	ESK-NS was administered via flexible dosing in line with the anticipated licence	Patients aged ≥65 years, who, for tolerability reasons, were started on an initial dose of 28 mg ESK-NS which is below the minimum effective dose of 56 mg	ESK-NS was administered via flexible dosing in line with the anticipated licence	A non-comparative study primarily designed to assess long-term safety (with minimal efficacy data)	An ongoing, non-comparative study primarily designed to assess long-term safety (with minimal efficacy data). Only interim data are available
Study design	Randomised, double-blind, parallel-group, active-controlled, Phase 3				Open-label, long-term, Phase 3	
Population	Adults (aged 18–64 years) with recurrent or single-episode TRD		Adults (aged ≥65 years) with recurrent or single-episode TRD	Adults (aged 18–64 years) with recurrent or single-episode TRD. Patients either directly entered the study or transferred from TRANSFORM-1/2 (having completed double-blind induction phase and demonstrated treatment response at end of 4-week double-blind induction phase of these transfer studies)	Adults (aged ≥18 years) with recurrent or single-episode TRD	

	TRANSFORM-1	TRANSFORM-2	TRANSFORM-3	SUSTAIN-1	SUSTAIN-2	SUSTAIN-3
Intervention	Fixed dose ESK-NS (56 mg OR 84 mg) twice weekly for 4 weeks (starting dose for all patients: 56 mg) PLUS newly initiated OAD	Flexibly-dosed ESK-NS (56 mg/84 mg) twice weekly for 4 weeks (starting dose for all patients: 56 mg) PLUS newly initiated OAD	Flexibly-dosed ESK-NS (28 mg/56 mg/84 mg) twice weekly for 4 weeks (starting dose for all patients: 28 mg) PLUS newly initiated OAD	Flexibly-dosed ESK-NS (SUSTAIN-1: 56 mg/84 mg; SUSTAIN-2/3: 28 mg/56 mg/84 mg in patients aged ≥65 years) twice weekly, weekly, or every other week (depending on efficacy and tolerability) until relapse or study termination PLUS newly initiated OAD		
Comparator	Newly initiated OAD plus PBO-NS twice weekly for 4 weeks			Newly initiated OAD plus PBO-NS twice weekly, weekly, or every other week (depending on efficacy and tolerability) until relapse or study termination	NA	

Based on Table 7 of the CS[1]
CS = company submission; ESK-NS = esketamine nasal spray; NA = not applicable; OAD = oral antidepressant; PBO-NS = placebo nasal spray; TRD = treatment-resistant depression

ERG comment*:* The company included two trials in the economic model (TRANSFORM-2, SUSTAIN-1) and these two alongside the TRANSFORM-3 and SUSTAIN-2 trials will be discussed in this section. The remaining trials TRANSFORM-1 (the fixed dosing study) and the ongoing non-comparative study SUSTAIN-3 will be discussed briefly in sections 4.2.8 and 4.2.9, respectively.

The two trials included in the initial economic model were TRANSFORM-2 and SUSTAIN-1, see Table 4.5. These were randomised, double-blind controlled trials targeting adults aged 18 to 64 years with recurrent or single episode depression. Both trials compared ESK-NS plus a newly initiated OAD to a newly initiated OAD plus placebo and both involved flexible dosing of 56 mg/ 84 mg of ESK-NS. ESK-NS was given for four weeks in TRANSFORM-2 and patients were either followed up for 24 weeks or joined SUSTAIN-1. SUSTAIN-1 also enrolled patients directly who had not taken part in TRANSFORM-2. In SUSTAIN-1, ESK-NS was given until relapse or trial termination.

The focus of the two trials was also different. TRANSFORM-2 aimed to treat patients with TRD in the acute phase of depression. Hence in TRANSFORM-2 the primary outcome was response as measured by the change in the 10-item clinician administered MADRS total score from baseline to the end of the four-week double-blind induction phase. SUSTAIN-1 aimed to delay relapse of depressive symptoms in patients with TRD who were in stable remission. The primary outcome for this trial was relapse defined as the time between patient randomisation into the maintenance phase and the first documentation (earliest date) of a relapse event (based on MADRS) during the maintenance phase among patients in stable remission (based on MADRS) at the end of the optimisation phase following treatment with ES-NS plus an OAD. Further outcomes in each trial relevant to the appraisal are given in Table 4.5.

TRANSFORM-2 enrolled 227 patients whereas SUSTAIN-1 enrolled 705 patients. See Table 4.5 for further details of the methodology of the two trials.

Table 4.5: Summary of study methodology for RCTs included in economic model

	TRANSFORM-2	**SUSTAIN-1**
Study objective	To evaluate the efficacy, tolerability and safety of flexibly-dosed ESK-NS (56 mg/ 84 mg) plus a newly initiated OAD (ESK-NS + OAD) versus a newly initiated OAD plus PBO-NS (OAD + PBO-NS) for the treatment of TRD in adults aged 18–64 years	To evaluate the efficacy, tolerability and safety of flexibly-dosed ESK-NS (56 mg/84 mg) plus a newly initiated OAD (ESK-NS + OAD) versus a newly initiated OAD + PBO-NS in delaying relapse of depressive symptoms in adults aged 18–64 years with TRD who are in stable remission following an induction (4 weeks) and optimisation (12 weeks) course of ESK-NS plus an OAD
No of patients	227	705
Study phases	Screening/prospective observational phase: • 4 weeks • Antidepressant taper period: ≤3 weeks (optional) • Double-blind induction phase: 4 weeks	**Direct-entry patients only:** • Screening/prospective observational phase, with an optional taper of ≤3 weeks for OAD(s): 4 weeks • Open-label induction phase: 4 weeks

	TRANSFORM-2	SUSTAIN-1
	• Follow-up phase: ≤24 weeks (only for those patients ineligible or unwilling to participate in subsequent long-term study SUSTAIN-1 following double-blind induction phase)	**Direct-entry and transferred-entry (from TRANSFORM-1/2) responder patients:** • Optimisation phase: 12 weeks (open-label for direct-entry patients, double-blind for transferred-entry patients) • Maintenance phase: variable duration (until relapse or study termination) • Follow-up phase: 2 weeks
Outcomes	• **Response (MADRS)** • Severity of depression (MADRS, CGI-S, PHQ-9) • **Remission (MADRS)** • Anxiety (GAD-7) • Functioning and associated disability (SDS) • Mortality (Safety outcome) • **Adverse effects of treatment (including adverse effects of treatment discontinuation)** • **Health-related quality of life (EQ-5D)**	• **Relapse (MADRS)** • Severity of depression (MADRS, CGI-S, PHQ-9) • Remission (MADRS) • Anxiety (GAD-7) • Functioning and associated disability (SDS) • Mortality (Safety outcome) • **Adverse effects of treatment (including adverse effects of treatment discontinuation)** • Health-related quality of life (EQ-5D)
Based on Tables 6 and 7 of the CS[1] Outcomes marked in bold are used in the model. CGI-S = Clinical Global Impression – Severity; EQ-5D = European Quality of Life-5 Dimensions; ESK-NS = esketamine nasal spray; GAD-7 = Generalised Anxiety Disorder – 7-item scale; MADRS = Montgomery-Åsberg Depression Rating Scale; OAD = oral antidepressant; PBO-NS = placebo nasal spray; PHQ-9 = Patient Health Questionnaire – 9 questions; RCT = randomised controlled trial; SDS = Sheehan Disability Scale; TRD = treatment-resistant depression		

ERG comment: The main trials in the CS and the economic model were randomised. Evidence is available for both acute treatment of treatment-resistant depression and for maintenance of effect after remission.

The above trials included only patients aged 18 to 64 years. A separate trial of those aged 65 and over with different dosing (TRANSFORM-3) and an open-label trial in adults aged 18 years or over (SUSTAIN-2) were initially not included in the model but are described below.

In response to clarification, the company advised that the label indication is expected to change to ESK-NS in combination with an SSRI or SNRI for treatment-resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode.[3] This reflects the trials where patients received either a SNRI or SSRI in conjunction with ESK-NS.

The company was asked to justify the use of the MADRS and PHQ-9 scales to determine depression severity. The PHQ-9 definition of response used was defined as ≥50% reduction from baseline in the PHQ-9 total score. A patient was defined as a responder at any given time point if the percentage improvement (decrease) in MADRS total score from baseline was ≥50%. A patient was defined as being

in remission at any time point if their MADRS total score was ≤12. The ERG noted that in technology appraisal 367 (TA367), remission was defined as MADRS total score of 10 or less.[9] The company stated that the difference was to account for the fact that remote raters were used instead of face-to-face raters.[3] Remote raters were used as the dissociative effects of ESK-NS might have resulted in unblinding if face-to-face MADRS raters were used.

The two main trials not included in the initial economic model were TRANSFORM-3 and SUSTAIN-2.

TRANSFORM-3 (138 participants) was a randomised, double-blind controlled trial targeting adults aged 65 years or over with recurrent or single episode depression. TRANSFORM-3 compared ESK-NS plus a newly initiated OAD to a newly initiated OAD plus placebo (28 mg/56 mg/84 mg) twice weekly for four weeks (starting dose for all patients: 28 mg). This lower dosage reflected the older population in the trial. Patients were either followed-up for two weeks or joined SUSTAIN-2.

SUSTAIN-2 (802 participants) also enrolled patients directly who had not taken part in TRANSFORM-3. SUSTAIN-2 was a one year long non-comparative study to assess long-term safety and tolerability of ESK-NS with selected efficacy outcomes also evaluated.

Further details of the methodology of the two trials are presented in Table 4.6.

A discussion of the role of TRANSFORM-3 in the economic model can be found in section 5.2. As SUSTAIN-2 was an open label study with no comparator, it is most useful as supporting evidence for longer-term safety outcomes.

Table 4.6: Summary of study methodology for TRANSFORM-3 and SUSTAIN-2

Trial no. (acronym)	**ESKETINTRD3005 (TRANSFORM-3)**	**ESKETINTRD3004 (SUSTAIN-2)**

Study objective	• To evaluate the efficacy of flexibly-dosed esketamine nasal spray plus a newly initiated OAD (ESK-NS + OAD) versus a newly initiated OAD plus placebo nasal spray (OAD + PBO-NS) for the treatment of TRD in elderly adults aged ≥65 years
	• To evaluate the safety and tolerability of each treatment regimen
		• To evaluate the long-term safety and tolerability of flexibly-dosed esketamine nasal spray plus a newly initiated OAD (ESK-NS + OAD) in adults aged ≥18 years with TRD, with special attention to the following:
		  • Potential effects on cognitive function
		  • Potential treatment-emergent symptoms of cystitis and/or lower urinary tract symptoms
		  • Potential withdrawal and/or rebound symptoms following cessation of esketamine treatment

Number of patients enrolled	N=138	N=802
Study phases	• Screening/prospective observational phase: 4 weeks • Antidepressant taper period: ≤3 weeks (optional) • Double-blind induction phase: 4 weeks	Direct-entry patients only: • Screening phase: 4 weeks Direct-entry and transferred-entry (from TRANSFORM-3) non-responder[a] patients only: • Open-label induction phase: 4 weeks

Trial no. (acronym)	ESKETINTRD3005 (TRANSFORM-3)	ESKETINTRD3004 (SUSTAIN-2)
	• Follow-up phase: • **TRANSFORM-3:** 2 weeks (only for those patients ineligible or unwilling to participate in subsequent long-term safety study SUSTAIN-2 following double-blind induction phase)	Direct-entry and transferred-entry (from TRANSFORM-3) responder[a] patients: • Optimisation/maintenance phase: 48 weeks • Follow-up phase: 4 weeks
Reported outcomes specified in the decision problem[b]	• Response (MADRS) • Severity of depression (MADRS, CGI-S, PHQ-9) • Remission (MADRS) • Anxiety (GAD-7) • Functioning and associated disability (SDS) • Mortality (Safety outcome) • Adverse effects of treatment (including adverse effects of treatment discontinuation) • Health-related quality of life (EQ-5D)	• Response (MADRS, PHQ-9) • Severity of depression (MADRS, CGI-S, PHQ-9) • Remission (MADRS, PHQ-9) • Anxiety (GAD-7) • Functioning and associated disability (SDS) • Mortality (Safety outcome) • Adverse effects of treatment (including adverse effects of treatment discontinuation) • Health-related quality of life (EQ-5D)
Based on Table 74 of the CS[1] [a] Response was defined as a ≥50% reduction from baseline in the MADRS total score; [b] Severity of depressive symptoms assessed using the MADRS score CGI-S = Clinical Global Impression – Severity; EQ-5D = European Quality of Life-5 Dimensions; ESK-NS + OAD = esketamine nasal spray (flexibly-dosed) plus a newly initiated oral antidepressant; GAD-7 = Generalised Anxiety Disorder – 7-item scale; MADRS = Montgomery-Åsberg Depression Rating Scale; OAD = oral antidepressant; OAD + PBO-NS = newly initiated oral antidepressant plus placebo nasal spray; PHQ-9 = Patient Health Questionnaire – 9 questions; SDS = Sheehan Disability Scale; TRD = treatment-resistant depression		

4.2.2 Statistical analysis of the studies included in the economic model

Table 4.7 summarises details on the statistical analysis for TRANSFORM-2 and SUSTAIN-1. These trials are used in the economic model.

Table 4.7: Statistical analysis of TRANSFORM-2 and SUSTAIN-1

	TRANSFORM-2	SUSTAIN-1
Sample size, power calculation	The maximum sample size planned was calculated assuming a treatment difference for the double-blind induction phase of 6.5 points in MADRS total score between ESK-NS + OAD and the OAD + PBO-NS arms, an SD of 12, a one-sided significance level of 0.025, and a drop-out rate of 25%. The treatment difference and SD used in this calculation were based on results of Panel A of the ESKETINTRD2003 study and on clinical judgment.	The maximum number of relapses (in patients with stable remission) required was 84, which would provide 90% power to detect a hazard ratio of 0.493 at the one-sided significance level of 0.025 for a fixed-sample design to detect superiority of ESK-NS plus an OAD over OAD plus PBO-NS in delaying relapse of depressive symptoms in patients with TRD who were in stable remission.

	TRANSFORM-2	SUSTAIN-1
	About 98 patients were required to be randomised to each treatment group to achieve 90% power using a fixed design assuming no interim analysis.	Calculation of sample size assumed that the time to the first relapse follows an exponential distribution, with a median time of 6 months for an OAD plus PBO-NS and 12.17 months for ESK-NS plus an OAD (corresponding 6-month relapse rates: 50% for OAD plus PBO-NS and 28.95% for ESK-NS plus an OAD). Accounting for assumptions made for accrual period and rate, maximum study duration, and dropout rate, a total of approximately 211 patients in stable remission needed to be randomised (1:1) to obtain 84 relapses.
Interim analysis for sample size re-estimation or stopping for futility	An interim analysis was planned to re-estimate sample size or to stop the study due to futility. Due to recruitment dynamics, a sample size re-estimation was not recommended after the study started, and the interim analysis was removed from the planned analyses in the second protocol amendment.	To evaluate the assumptions used in the sample size calculation, relapse rates were to be monitored sequentially during the maintenance phase. In particular, a two-stage group sequential design was adopted, with one interim analysis to be performed when at least 33 relapse events had occurred in stable remitters with at least 30 relapses from patients treated with ESK-NS plus an OAD in the optimisation phase. The interim analysis was conducted according to a separate statistical analysis plan. The IDMC reviewed the interim analysis results and made a recommendation to either stop the study for efficacy or provide the sample size adjustment based on the rules defined in the interim analysis statistical analysis plan.
Statistical testing sequence and levels of significance	A fixed sequence, serial gatekeeping procedure was applied to adjust for multiplicity and to strongly control type I error across the primary and the three key secondary efficacy endpoints. Testing of the endpoints was performed sequentially in the following order: change in MADRS total score, onset of clinical response by Day 2 (24 hours), change in SDS total score, and change in PHQ-9 total score. Testing of the endpoints was performed sequentially in the order indicated above and were considered statistically significant at the one-sided 0.025 level only if the endpoint was individually significant at the one-sided 0.025 level	A two-stage group sequential design, with one interim analysis was adopted as described above. In either case of stopping at the interim analysis or continuing with sample size re-estimation, control of overall type I error would thereby be maintained. The final efficacy analysis was performed at a significance level of 0.046 (two-sided). If the result of the final efficacy analysis was significant ($Z_f \geq 1.998$), ESK-NS plus an OAD would be declared superior to an OAD plus PBO-NS in delaying relapse.

	TRANSFORM-2	SUSTAIN-1
	and previous endpoints in the hierarchy were significant at the one-sided 0.025 level.	
Hypothesis objective	The hypothesis for TRANSFORM-2 was that, in adult patients with TRD, switching from a failed OAD to ESK-NS plus a newly initiated OAD would be superior to switching to a newly initiated OAD treatment (active comparator) plus PBO-NS in improving depressive symptoms.	ESK-NS plus an OAD is more effective than treatment with an OAD plus PBO-NS in delaying relapse of depressive symptoms in patients with TRD in stable remission.
Statistical analysis (primary outcome)	The primary endpoint was: • Change from baseline to Day 28 in the MADRS total score reported as the difference in treatment means. The primary analysis was based on the full analysis set and the MADRS total scores collected during the double-blind induction phase. Different analysis methods were used dependent on the regulatory needs of specific regions: ANCOVA (EU) and MMRM (non-EU). ***ANCOVA*** Change from baseline in MADRS total score at Day 28 of the double-blind induction phase was analysed based on LOCF data. The model included factors for treatment, country, and class of OAD (SNRI or SSRI), and baseline MADRS total score as a covariate. ***MMRM*** Change from baseline in MADRS total score at Day 28 of the double-blind induction phase was analysed based on observed data. The model included baseline MADRS total, and treatment, class of OAD (SNRI or SSRI), day, day-by-treatment interaction, and country as fixed effects. The within-patient covariance between visits was estimated via an unstructured variance-covariance matrix.	The primary endpoint was: Time to relapse during the maintenance phase, while on their initially randomised treatment. The primary analysis was based on the full (stable remitters) analysis set and relapse (based on MADRS total score, defined in Table 10 of the CS) collected during the maintenance phase. The treatment groups were compared using the weighted log-rank test. Time to relapse was summarised (number of events, number of censored patients and quartiles of time to relapse). The cumulative distribution function of the time to relapse was estimated by the Kaplan-Meier method.
Statistical analysis (key secondary outcomes)	• Analysis of the proportion of patients showing onset of clinical response by Day 2 (24 hours) that was maintained for the duration of the double-blind induction phase in the ESK-NS plus an OAD arm versus the OAD plus PBO-NS arm was planned using a Cochran-Mantel-Haenszel chi square test adjusting for country and class of antidepressant (SSRI or SNRI).	• For time to relapse in stable responders (who were not stable remitters), time to relapse was summarised and the cumulative distribution function of time to relapse was estimated by the Kaplan-Meier method. The difference in time to relapse between treatment groups was evaluated using a two-sided log-

	TRANSFORM-2	SUSTAIN-1
	• Change from baseline in SDS total score and change from baseline in PHQ-9 total score at Day 28 in the double-blind induction phase were analysed using the same models described for the primary efficacy analysis.	rank test and the hazard ratio and 95% CI were estimated based on the Cox proportional hazards model with treatment as a factor. • For MADRS, PHQ-9, CGI-S, GAD-7, and SDS, change from baseline (for the maintenance phase) at each visit, including observed case and LOCF data, were analysed using the ANCOVA model with factors for treatment and country and baseline score as covariates. The proportion of patients with response and remission based on MADRS, PHQ-9 or SDS were summarised over time.
Data management, patient withdrawals	**Imputation for missing timepoints:** For endpoints using ANCOVA, the LOCF method was applied to the MADRS total score, SDS total score, PHQ-9 total score, and CGI-S for the double-blind induction phase. The last post-baseline observation during the double-blind induction phase was carried forward as the endpoint for that phase. In addition to the observed cases and the endpoint assessments, the LOCF values were created for intermediate post-baseline timepoints as well.	**Imputation for missing timepoints:** For the MADRS, CGI-S, PHQ-9, GAD-7 and SDS, both observed case and LOCF values were determined for the induction, optimisation and maintenance phases. The last post-baseline observation during each phase was carried forward as the "Endpoint." In addition to the observed cases and endpoint assessment, the LOCF values were created for intermediate post baseline timepoints.
	Imputation for missing items: For MADRS total score, if two or more items were missing, no imputation was performed, and the total score was left missing. Otherwise, the total score was calculated as a sum of the non-missing items multiplied by the ratio of the maximum number of items (i.e., 10) to the number of non-missing items. For all other scales where multiple items were summed to create a total, if any item of the scale was missing at a visit, the total score for that scale at that visit was left blank.	

Based on Table 17 of the CS[1]

ANCOVA = analysis of covariance; CGI-S = Clinical Global Impression – Severity; CI = confidence interval; CS = company submission; ESK = esketamine; EU = European Union; GAD-7 = Generalised Anxiety Disorder – 7-item scale; IDMC = independent data monitoring committee; LOCF = last observation carried forward; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = Mixed-Effects Model using Repeated Measures; NS = nasal spray; OAD = oral antidepressant; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TRD = treatment-resistant depression

ERG comment: The ERG has no concerns regarding the appropriateness of the statistical methods described in Table 4.7.

4.2.3 Trial inclusion criteria and participant characteristics

Details of the full inclusion criteria for TRANSFORM-2 and SUSTAIN-1 are provided in section B.2.3.4 of the CS and for TRANSFORM-3 and SUSTAIN-2 in Appendix M3 and are not reproduced in this report.[1, 17]

ERG comment: The ERG notes the following in relation to these inclusion criteria:

- TRANSFORM-2 and SUSTAIN-1 only included patients aged 18 to 64 years of age. TRANSFORM-3 was conducted in patients aged over 65 years only. SUSTAIN-2 included a wider age range but was an open label trial. TRANSFORM-3 was included in the CS only as supporting evidence and did not inform the economic model. The ERG was, therefore, concerned as to the relevance of evidence to the older population. The company was asked to clarify if they considered the trials to be applicable to patients aged 65 years and over.[18] The company presented results of patients aged 65 to 74 years from TRANSFORM-3 showing them to be similar in magnitude to those in the younger adult population. The lower effect noted in those aged 75 years and over was considered to be an artefact of the low number of patients (n=22).[3]
- The trials in the CS excluded patients with moderate/severe alcohol abuse according to DSM-5 criteria. The committee will need to consider whether evidence in the CS on effectiveness and safety of ESK-NS can be generalised to those with a dual diagnosis of depression and alcohol misuse.
- The trials in the CS also excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening. Again, the committee will need to consider if the evidence in the CS on effectiveness and safety of ESK-NS can be generalised to this vulnerable population.
- In the trials, the patients had to be adherent to current OAD treatment (without adjustment in dosage) throughout screening/prospective observational phases. In clinical practice patients may not adhere to OAD medication, i.e. this might limit the generalisability of the findings.
- The trials excluded patients who had not responded to an adequate course of treatment with ECT in the current major depressive episode. This appears to be in line with the proposed pathway for ESK-NS. The committee will need to consider if ESK-NS is likely to be offered to patients who have not responded to ECT.
- Details of selected baseline characteristics across the four main trials (TRANSFORM-2, SUSTAIN-1, TRANSFORM-3 and SUSTAIN-2) are shown in Table 4.8.

Table 4.8: Selected demographic baseline characteristics of the main trials: TRANSFORM-2, SUSTAIN-1, TRANSFORM-3 and SUSTAIN-2

Characteristic	TRANSFORM-2 (N=223)	SUSTAIN-1 (N=705)	TRANSFORM-3 (N=138)	SUSTAIN-2 (N=802)
Age, mean years (SD)	45.7 (11.89)	46.1 (11.10)	70.0 (4.52)	52.2 (13.69)
Age category, n (%)				
18–44 years	94 (42.2)	292 (41.4)	NA	225 (28.1)
45–64 years	129 (57.8)	413 (58.6)	NA	399 (49.8)
65–74 years	NA		116 (84.7)	159 (19.8)
≥74 years	NA		21 (15.3)	19 (2.4)
Sex, n (%)				
Male	85 (38.1)	248 (35.2)	52 (38.0)	300 (37.4)
Female	138 (61.9)	457 (64.8)	85 (62.0)	502 (62.6)
Race, n (%)				
American Indian or Alaskan Native	NA	1 (0.1)	NA	
Asian	2 (0.9)	3 (0.4)	NA	81 (10.1)
Black or African American	11 (4.9)	31 (4.4)	NA	15 (1.9)
White	208 (93.3)	635 (90.1)	130 (94.9)	686 (85.5)
Multiple	2 (0.9)	4 (0.6)	4 (2.9)	8 (1.0)
Not reported	NA	9 (1.3)	2 (1.5)	4 (0.5)
Other	NA	22 (3.1)	NA	8 (1.0)
Unknown	NA		1 (0.7)	
Employment status, n (%)[a]				
Any type of employment	131 (58.7)	448 (63.5)	24 (17.5)	450 (56.1)
Any type of unemployment	69 (30.9)	180 (25.5)	8 (5.8)	175 (21.8)
Other	23 (10.3)	77 (10.9)	105 (76.6)	177 (22.1)
Region, n (%)				
Europe	134 (60.1)	411 (58.3)	59 (43.1)	322 (40.1)

Characteristic	TRANSFORM-2 (N=223)	SUSTAIN-1 (N=705)	TRANSFORM-3 (N=138)	SUSTAIN-2 (N=802)
North America	89 (39.9)	195 (27.7)	70 (51.1)	147 (18.3)
Other	NA	99 (14.0)	8 (5.8)	333 (41.5)
Class of OAD, n (%)				
SNRI	152 (68.2)	440 (62.9)	61 (44.5)	407 (50.8)
SSRI	71 (31.8)	259 (37.1)	76 (55.5)	394 (49.2)
OAD, n (%)				
Duloxetine	121 (54.3)	323 (46.2)	48 (35.0)	251 (31.3)
Escitalopram	38 (17.0)	128 (18.3)	50 (36.5)	237 (29.6)
Sertraline	32 (14.3)	130 (18.6)	25 (18.2)	157 (19.6)
Venlafaxine XR	32 (14.3)	118 (16.9)	14 (10.2)	156 (19.5)
MADRS total score, mean (SD)	37.1 (5.67)	37.9 (5.50)	35.2 (6.16)	31.4 (5.39)
PHQ-9 total score, mean (SD)	20.3 (3.68)	19.9 (4.18)	17.5 (5.65)	17.3 (5.01)
Screening C-SSRS lifetime, n (%)[b]				
No event	126 (56.5)	407 (57.7)	73 (54.1)[j]	474 (59.3)
Suicidal ideation	74 (33.2)	193 (27.4)	43 (31.9)[j]	203 (25.4)
Suicidal behaviour	23 (10.3)	105 (14.9)	19 (14.1)[j]	123 (15.4)
Screening C-SSRS past 6 or 12 months, n (%)				
No event	151 (67.7)	499 (70.8)	86 (63.7)	583 (72.9)
Suicidal ideation (past 6 months)	71 (31.8)	205 (29.1)	48 (35.6)	215 (26.9)
Suicidal behaviour (past 12 months)	1 (0.4)[c]	1 (0.1)	1 (0.7)	2 (0.3)
Duration of current episode, mean weeks (SD)	114.6 (157.96)	132.2 (209.18)	215.8 (341.71)	160.5 (261.80)
Number of previous antidepressant medications, n (%)[d]	[e]			
2	136 (61.0)	248 (57.7)	68 (49.6)	452 (58.0)
3	53 (23.8)	111 (25.8)	34 (24.8)	182 (23.4)

Characteristic	TRANSFORM-2 (N=223)	SUSTAIN-1 (N=705)	TRANSFORM-3 (N=138)	SUSTAIN-2 (N=802)
4	20 (9.0)	39 (9.1)	17 (12.4)	83 (10.7)
≥5	9 (4.0)	20 (4.7)	7 (5.1)	49 (6.3)
Number of major depressive episodes including current episode, n (%)				
1	29 (13.0)	83 (11.8)	18 (13.1)	111 (13.9)
2–5	159 (71.3)	454 (64.5)	86 (62.8)	534 (66.7)
6–10	31 (13.9)	122 (17.3)	20 (14.6)	121 (15.1)
>10	4 (1.8)	45 (6.4)	13 (9.5)	35 (4.4)

Based on Tables 12 and 13 of the CS[1] and Tables 77 and 78 of the CS appendices[17]

[a] Any type of employment included: any category containing "employed," sheltered work, housewife or dependent husband, and student. Any type of unemployment included: any category containing "unemployed." Other included: retired and no information available; [b] C-SSRS category: No event = 0; Suicidal ideation = 1, 2, 3, 4, 5; Suicidal behaviour = 6, 7, 8, 9, 10; [c] Due to a data collection error, one patient in TRANSFORM-2 reported suicidal behaviour in the 12 months prior to screening. The suicidal behaviour for this patient actually occurred more than 12 months prior to screening.; [d] Referring to the number of antidepressant medications with non-response (defined as ≤25% improvement in MGH-ATRQ) taken for ≥6 weeks during the current episode; [e] All of the five patients not accounted for in this baseline measure TRANSFORM-2 were determined to have failed at least two OADs based on other data in the database

C-SSRS = Columbia – Suicide Severity Rating Scale; MADRS = Montgomery-Åsberg Depression Rating Scale; NA = not applicable; OAD = oral antidepressant; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

In line with the different age inclusion criteria, the mean age of patients in TRANSFORM-2 and SUSTAIN-1 was approximately 46 years (52 years in SUSTAIN-2) compared to 70 years in TRANSFORM-3. Both male and female participants were represented across the trials and women formed over 60% of the population in the trials. Most participants (85.5% to 94.9%) identified as white. Most participants were employed (56.1% to 63.5%) except for TRANSFORM-3 where most participants were not of working age. Trial participants were mainly from Europe in TRANSFORM-2 (60.1%) and SUSTAIN-1 (58.3%) and from North America in TRANSFORM-3 (51.1%).

ERG comment: The larger number of women in the trials reflects the higher prevalence of women with depression. The ERG notes that Black and Asian people appear to be underrepresented across the two trials.

The company stated that **'***neither TRANSFORM-2 or* [sic] *SUSTAIN-1 enrolled any patients in the UK. (One UK patient was enrolled in the supporting trial, TRANSFORM-3, and 12 UK patients were enrolled in the long-term safety study, SUSTAIN-2). Although subgroup analyses conducted on the primary outcomes in TRANSFORM-2 and SUSTAIN-1 did suggest minor effects of patient region, country, and/or ethnicity on ESK-NS treatment response, drawing conclusions from these results is cautioned due to the small numbers of patients in these subgroups and the resulting wide confidence intervals*'.[1] The lack of UK patients in the main trials included in the economic model is a limitation particularly given the mode of delivery of this intervention. There is a lack of evidence in how well ESK-NS might work in the NHS setting.

ESK-NS patients across the trials received either a SNRI or SSRI. In TRANSFORM-2 and SUSTAIN-1, used in the model, most patients (68.2% and 62.9% respectively received a SNRI). The most frequently prescribed OAD in these trials was duloxetine (54.3% and 46.2%, respectively). Patients had an average score on MADRS of 37.1 in TRANSFORM-2 and 37.9 in SUSTAIN-1 indicating severe depression. Over 40% had a lifetime score on C-SSRS indicating suicide ideation or behaviour. As mentioned before, the trials in the CS excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening. For most patients (approximately 87%) this was not their first major depressive episode. Most patients had received two prior OADs in this episode (61% and 57.7% in TRANSFORM-2 and SUSTAIN-1, respectively).

ERG comment: The committee will need to consider how well the OADs prescribed as co-interventions across these trials reflect those prescribed at this stage of the pathway in the NHS setting.

There is evidence available in the trials on those given ESK-NS after over two previous OADs to inform later stages of the proposed pathway, but participant numbers are smaller. Across the trials between 49.6% and 61.1% had received two previous OADs. The committee is referred to the subgroup analysis described in section 4.2.6.

4.2.4 Risk of bias assessments of included trials

The company's quality assessment of the two ESK-NS trials supporting the economic model used the NICE recommended tool.[22] The quality assessment was reported in the main submission and in the appendices of the CS and is shown in Table 4.9 for TRANSFORM-2 and SUSTAIN-1 and in Tables 4.10 and 4.11 for TRANSFORM-3 and SUSTAIN-2, respectively.[1, 17] The open-label extension study, SUSTAIN-2, was assessed using a different set of signalling questions to the four RCTs, but the

company did not report the tool that was used. As stated in section 4.1.4, it was not clear how many reviewers were involved in the quality assessment process.

ERG comment: All three RCTs (TRANSFORM-2 and -3 and SUSTAIN-1) were judged by the company to have met all of the relevant quality criteria. The ERG re-assessed the studies against the specified criteria and agrees that the RCTs were well conducted with appropriate procedures of randomisation and allocation concealment.

However, the question regarding the blinding of care providers, participants and outcome assessors has been answered in the affirmative (i.e. that all three populations were adequately blinded). The ERG queries whether blinding (specifically of care providers and participants) could be maintained in a clinical situation where the dissociative effects of the esketamine intervention were so much more overt than the comparator that they required the use of remote, independent raters to assess the primary outcome.

The ERG agrees that the observational study (SUSTAIN-2) met all of the relevant criteria on the tool used for assessment by the company. However, it appeared to the ERG that most of the signalling questions were based on a reporting guideline rather than a risk of bias assessment, and as such, this was probably an inappropriate tool to use. Although SUSTAIN-2 appeared to be a well conducted observational study, it is a non-comparative open-label study and as such will be open to bias. It is best viewed as supporting evidence for ESK-NS and indeed the company did not include it in economic modelling stating that its primary aim was to assess long-term safety.

Table 4.9: Company quality assessment of TRANSFORM-2 and SUSTAIN-1

	TRANSFORM-2	SUSTAIN-1
Was randomisation carried out appropriately?	Yes. Patients were randomised in a 1:1 ratio based on a computer-generated randomisation schedule prepared before the study by or under the supervision of the sponsor.	Yes. At the start of the maintenance phase patients were randomised in a 1:1 ratio based on a computer-generated randomisation schedule prepared before the study under the supervision of the sponsor.
Was the concealment of treatment allocation adequate?	Yes. IWRS was used to assign a unique treatment code, which dictated the treatment assignment and matching medication kits for the patient.	Yes. An IWRS was used to assign a unique treatment code, which dictated the treatment assignment and matching medication kits for the patient.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Demographics and disease characteristics were balanced between the groups. Randomisation was balanced by using randomly permuted blocks (block size=4) and was stratified by country and class of OAD (SNRI or SSRI) initiated in the double-blind induction phase.	Yes. Demographics and disease characteristics were balanced between the groups. Both randomisations were balanced by using randomly permuted blocks (block size=4) and were stratified by country.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes. This was a double-blind study. The IWRS was used to manage study agent inventory while ensuring that no one at the site had to be unblinded. The investigator was not provided with the treatment randomisation codes. The investigators and the site personnel were blinded to the treatment assignment until all patients completed study participation through the follow-up phase. To maintain the blinding of intranasal study medication, the esketamine and placebo intranasal devices were indistinguishable (via use of a bittering agent added to the placebo solution to simulate the taste of the intranasal solution with active drug). To ensure an unbiased efficacy evaluation, independent, remote (by phone), blinded MADRS raters were used to assess the antidepressant treatment response.	Yes. This was a double-blind study. The IWRS was used to manage study agent inventory while ensuring that no one at the site had to be unblinded. The investigator was not provided with the unique treatment randomisation codes. The blind was not to be broken until all patients completed the study and the database was finalised. To maintain the blinding of intranasal study medication, the esketamine and placebo intranasal devices were indistinguishable (via use of a bittering agent added to the placebo solution to simulate the taste of the intranasal solution with active drug). To ensure an unbiased efficacy evaluation, independent, remote (by phone), blinded MADRS raters were used to assess the antidepressant treatment response.
Were there any unexpected imbalances in drop-outs between groups?	No. The overall drop-outs were generally well-balanced between treatment arms.	No. The overall drop-outs during the randomised maintenance phase were generally well-balanced between treatment arms and the primary reasons for treatment discontinuation were also well-balanced between treatment arms.

	TRANSFORM-2	SUSTAIN-1
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. Based on the clinical study report all outcomes are reported in detail	No. Based on the clinical study report all outcomes are reported in detail.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. Efficacy analyses in the double-blind induction phase were performed on the FAS, defined as all randomised patients who received at least 1 dose of intranasal study medication and 1 dose of OAD medication. The safety analysis set was defined as all randomised patients who received at least 1 dose of intranasal study medication or 1 dose of OAD medication. For the MADRS, if 2 or more items were missing, no imputation was performed and the total score was left missing. For all other scales where multiple items were summed to create a total, if any item of the scale was missing at a visit, the total score for that scale at that visit was left blank.	Yes. There were 2 FAS defined for the maintenance phase: • Full (stable remitters): used to perform primary and secondary efficacy evaluations on randomised patients who were in stable remission at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase. • Full (stable responders): used to perform secondary efficacy evaluations on randomised patients who were stable responders (who were not stable remitters) at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase. For the MADRS, if 2 or more items were missing, no imputation was performed, and the total score was left missing. For all other scales where multiple items were summed to create a total, if any item of the scale was missing at a visit, the total score for that scale at that visit was considered missing.
Based on Table 18 of the CS[1] CS = company submission; FAS = full analysis set; IWRS = interactive web response system; MADRS = Montgomery-Åsberg Depression Rating Scale; OAD = oral antidepressant; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor		

Table 4.10: Company quality assessment of TRANSFORM-3

	TRANSFORM-3
Was randomisation carried out appropriately?	Yes. Central randomisation was implemented. Patients were randomised in a 1:1 ratio based on a computer-generated randomisation schedule prepared before the study by or under the supervision of the sponsor.

	TRANSFORM-3
Was the concealment of treatment allocation adequate?	Yes. An IWRS was used to assign a unique treatment code, which dictated the treatment assignment and matching medication kits for the patient.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Demographics and disease characteristics were balanced between the groups. Randomisation was balanced by using randomly permuted blocks (block size=4) and was stratified by country and class of oral antidepressant (SNRI or SSRI) initiated in the double-blind induction phase.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes. This was a double-blind study. The IWRS was used to manage study agent inventory while ensuring that no one at the site had to be unblinded. The investigator was not provided with randomisation codes. Randomisation codes were disclosed fully only after the study was completed and the clinical database was closed. To maintain the blinding of intranasal study medication, the esketamine and placebo intranasal devices were indistinguishable (via use of a bittering agent added to the placebo solution to simulate the taste of the intranasal solution with active drug). To ensure an unbiased efficacy evaluation, independent, remote (by telephone), blinded MADRS raters were used to assess the antidepressant treatment response
Were there any unexpected imbalances in drop-outs between groups?	No. The overall drop-outs were generally well-balanced between treatment arms and the primary reasons for treatment discontinuation were also well-balanced between treatment arms.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. Based on the clinical study report all outcomes are reported in detail.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. Efficacy analyses in the double-blind induction phase were performed on the FAS, defined as all randomised patients who received at least 1 dose of intranasal study medication and 1 dose of oral antidepressant medication. The safety analysis set was defined as all randomised patients who received at least 1 dose of intranasal study medication or 1 dose of oral antidepressant medication. For the MADRS, if 2 or more items were missing, no imputation was performed and the total score was left missing. For all other scales where multiple items were summed to create a total, if any item of the scale was missing at a visit, the total score for that scale at that visit was considered missing.

Based on Table 50 of the CS[1]

CS = company submission; FAS = full analysis set; IWRS = interactive web response system; MADRS = Montgomery-Åsberg Depression Rating Scale; OAD = oral antidepressant; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Table 4.11: Company quality assessment of SUSTAIN-2

	SUSTAIN-2
Was the hypothesis/aim/objective of the study clearly stated?	Yes, the objective was to evaluate the long-term safety and tolerability of flexibly-dosed esketamine nasal spray plus a newly initiated oral anti-depressant in adults aged ≥18 years with treatment resistance depression.
Was the study conducted prospectively?	Yes, this was an open-label prospective study to investigate the long-term safety and tolerability of esketamine.
Were the cases collected in more than one centre?	Yes, patients were enrolled at multiples sites across Europe, South America and Asia.
Were patients recruited consecutively?	Yes, the study recruited both direct entry and transferred entry subjects from a previous study (ESKEINTRD3005), based on clearly defined eligibility criteria.
Were the characteristics of the patients included in the study described?	Yes, demographics and baseline disease characteristics were reported for patients in the study.
Were the eligibility criteria (i.e. inclusion and exclusion criteria) for entry into the study clearly stated?	Yes. Exclusion criteria were clearly stated for direct entry patients. For transferred entry, patients had to have completed the double-blind induction phase of ESKEINTRD3005.
Did patients enter the study at a similar point in the disease?	Yes. All patients (direct-entry and transferred-entry) had TRD, defined as non-response to at least 2 OADs.
Was the intervention of interest clearly described?	Yes, all of the most relevant characteristics of esketamine were reported (including dosage, frequency, duration and administration methods). Details for the induction and optimisation/maintenance phases were clearly defined.
Were additional interventions (co-interventions) clearly described?	Yes, all patients received one of four OADs from 2 classes, with dosing according to local prescribing guidelines.
Based on Table 51 of the CS[1] CS = company submission; OAD = oral antidepressant; TRD = treatment-resistant depression	

4.2.5 Main efficacy results

Tables 4.12 and 4.13 summarise the efficacy results of TRANSFORM-2 and SUSTAIN-1, the RCTs used to inform the economic model. Tables 4.14 and 4.15 summarise the efficacy results of TRANSFORM-3 and SUSTAIN-2 which the CS included as supporting evidence.

Table 4.12: Summary of efficacy results of TRANSFORM-2

Outcome	ESK-NS + OAD	OAD + PBO-NS
MADRS[a,b]		
Change from baseline (observed cases)		
Baseline (mean, SD)	N=114, 37.0 (5.69)	N=109, 37.3 (5.66)
Day 28 (mean, SD)	N=101, 15.5 (10.67)	N=100, 20.6 (12.70)
Change from baseline to day 28 (mean, SD)	N=101, -21.4 (12.32)	N=100[c], -17.0 (13.88)
MMRM (difference in LS means, SE, 95% CI)[d]	-4.0 (1.69, -7.31 to -0.64)	
Onset of clinical response (FAS)		
Achieved onset of clinical response by day 2 (n, %)	N=114, 9 (7.9%)	N= 109, 5 (4.6%)
Generalised Cochran-Mantel-Haenszel test[e]	OR 1.79 (95% CI 0.57 to 5.67)	
Response and remission (observed cases)		
Response rate[f]	69.3%	52.0% (unadjusted)[g]
		34.0% (adjusted)[g]
Remission rate[h]	52.5%	31.0% (unadjusted)[g]
		18.0% (adjusted)[g]
CGI-S (observed cases)[i]		
Baseline (mean, SD)	NR	NR
Day 28 (mean, SD)	NR	NR
Change from baseline to day 28 (mean, SD)	N=101, -2.1 (1.33)	N=97, -1.6 (1.38)
MMRM (difference in LS means, SE, 95% CI)[d]	-0.4 (0.17, -0.72 to -0.04)	
PHQ-9 (observed cases)[i]		
Baseline (mean, SD)	N=114, 20.2 (3.63)	N=109, 20.4 (3.74)
Day 28 (mean, SD)	N=104, 7.3 (5.74)	N=100, 10.2 (7.68)
Change from baseline to day 28 (mean, SD)	N=104, -13.0 (6.42)	N=100, -10.2 (7.80)
MMRM (difference in LS means, SE, 95% CI)[d]	-2.4 (0.88, -4.18 to -0.69)	
GAD-7 (observed cases)[j]		
Baseline (mean, SD)	N=114, 13.2 (5.12)	N=109, 13.1 (4.83)
Day 28 (mean, SD)	N=110, 5.2 (5.46)	N=102, 6.2 (5.17)

Outcome	ESK-NS + OAD	OAD + PBO-NS
Change from baseline to day 28 (mean, SD)	N=110, -7.9 (6.12)	N=102, -6.8 (5.75)
ANCOVA (difference in LS means, SE, 95% CI)[k]	-1.0 (0.67, -2.35 to 0.28)	
SDS (observed cases)[l]		
Baseline (mean, SD)	N=111, 24.0 (4.07)	N=104, 24.2 (4.38)
Day 28 (mean, SD)	N=86, 10.1 (7.71)	N=86, 14.8 (9.07)
Change from baseline to day 28 (mean, SD)	N=86, -13.6 (8.31)	N=85, -9.4 (8.43)
MMRM (difference in LS means, SE, 95% CI)[b]	-4.0 (1.17, -6.28 to -1.64)	
EQ-5D (observed cases)[b,m]		
Baseline (mean, SD)	N=114, 0.530 (0.2081)	N=109, 0.501 (0.2143)
Day 28 (mean, SD)	N=104, 0.843 (0.1407)	N=100, 0.732 (0.2325)
Change from baseline to day 28 (mean, SD)	N=104, 0.310 (0.2191)	N=100, 0.235 (0.2525)
Difference in LS means, SE, 95% CI	NR	NR
Other outcomes defined in the final scope		
Cognitive dysfunction	NR	NR
Hospitalisation	NR	NR
Sleep quality	NR	NR

Based on Tables 7, 19, 21, 23, 24, 26, 45 and Figure 15 of the CS as well as the CSR[1, 23]

[a] Related to response, severity of depression, and remission (Table 4.5); [b] Used in the economic model; [c] = Table 19 of the CS reported this as "109". Error corrected by the ERG; [d] Change from baseline was the response variable and fixed effect model terms for treatment, day, country, class of OAD (SNRI or SSRI), treatment-by-day, and baseline value were covariates; [e] Adjusted for region and class of OAD (SNRI or SSRI); [f] ≥50% reduction from baseline in MADRS total score; [g] See details in section 5.2.6; [h] MADRS total score of ≤12; [i] Related to severity of depression (Table 4.5); [j] Related to anxiety (Table 4.5); [k] Change from baseline was the response variable and treatment, country, class of OAD (SNRI or SSRI), and baseline GAD-7 value were covariates; only ANCOVA reported; [l] Related to functioning and associated disability (Table 4.5); [m] = Related to health-related quality of life (Table 4.5)

ANCOVA = analysis of covariance; CGI-S = Clinical Global Impression; CI = confidence interval; CS = company submission; CSR = clinical study report; EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; ESK = esketamine; FAS = full analysis set; GAD-7 = Generalised Anxiety Disorder – 7-item scale; HR = hazard ratio; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; NR = not reported; NS = nasal spray; OAD = oral antidepressant; OR = odds ratio; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Table 4.13: Summary of efficacy results of SUSTAIN-1

Outcome	ESK-NS + OAD	OAD + PBO-NS
Time to relapse		
Stable remitters[a]		
Number of relapses	24/90 (26.7%)	39/86 (45.3%)
Time to relapse	HR 0.49 (95% CI 0.29 to 0.84)	

Outcome	ESK-NS + OAD	OAD + PBO-NS
Stable responders[b]		
Number of relapses	16/62 (25.8%)	34/59 (57.6%)
Time to relapse	HR 0.30 (95% CI 0.16 to 0.55)	
MADRS (LOCF)[c,d]		
Change from baseline		
Change from baseline to end of maintenance phase (mean, SD)[e]	Stable remitters[a]: N=89, 7.5 (11.59)	Stable remitters[a]: N=86, 12.5 (13.63)
	Stable responders[b]: N=62, 4.4 (11.38)	Stable responders[b]: N=59, 11.4 (12.00)
ANCOVA (difference in LS means, SE, 95% CI)[f]	Stable remitters[a]: -5.2 (1.82, -8.7 to -1.58)	
	Stable responders[b]: -7.4 (1.95, -11.30 to -3.55)	
Response/remission		
Responder at beginning of maintenance phase	Stable remitters[a]: 90/90 (100.0%)	Stable remitters[a]: 86/86 (100.0%)
	Stable responders[b]: 62/62 (100.0%)	Stable responders[b]: 59/59 (100.0%)
Responder at end of maintenance phase[e]	Stable remitters[a]: 67/89 (75.3%)	Stable remitters[a]: 48/86 (55.8%)
	Stable responders[b]: 41/62 (66.1%)	Stable responders[b]: 20/59 (33.9%)
Remitter at beginning of maintenance phase	Stable remitters[a]: 90/90 (100.0%)	Stable remitters[a]: 85/86 (98.8%)
	Stable responders[b]: 37/62 (59.7%)	Stable responders[b]: 38/59 (64.4%)
Remitter at end of maintenance phase[e]	Stable remitters[a]: 58/89 (65.2%)	Stable remitters[a]: 36/86 (41.9%)
	Stable responders[b]: 29/62 (46.8%)	Stable responders[b]: 15/59 (25.4%)
CGI-S (LOCF)[g]		
Change from baseline to end of maintenance phase (median, range)[e]	Stable remitters[a]: N=89, 0.0 (-3 to 4)	Stable remitters[a]: N=86, 1.0 (-2 to 5)
	Stable responders[b]: N=62, 0.0 (-2 to 4)	Stable responders[b]: N=58, 1.0 (-3 to 5)
ANCOVA (difference in LS means, SE, 95% CI)[f]	Stable remitters[a]: P value 0.055[h]	
	Stable responders[b]: P value 0.002[h]	
PHQ-9 (LOCF)[g]		
Change from baseline		
Change from baseline to end of maintenance phase (mean, SD)[e]	Stable remitters[a]: N=89, 3.3 (5.58)	Stable remitters[a]: N=86, 5.9 (7.09)
	Stable responders[b]: N=61, 1.7 (5.02)	Stable responders[b]: N=58, 4.7 (5.48)
ANCOVA (difference in LS means, SE, 95% CI)[f]	Stable remitters[a]: -2.4 (0.90, -4.20 to -0.65)	
	Stable responders[b]: -3.0 (0.93, -4.87 to -1.18)	

Outcome	ESK-NS + OAD	OAD + PBO-NS
Response/remission		
Responder at beginning of maintenance phase	Stable remitters[a]: 88/90 (97.8%)	Stable remitters[a]: 86/86 (100.0%)
	Stable responders[b]: 60/62 (96.8%)	Stable responders[b]: 56/59 (94.9%)
Responder at end of maintenance phase	Stable remitters[a]: 72/89 (80.9%)	Stable remitters[a]: 57/86 (66.3%)
	Stable responders[b]: 48/61 (78.7%)	Stable responders[b]: 40/58 (69.0%)
Remitter at beginning of maintenance phase	Stable remitters[a]: 83/90 (92.2%)	Stable remitters[a]: 76/86 (88.4%)
	Stable responders[b]: 25/62 (40.3%)	Stable responders[b]: 32/59 (54.2%)
Remitter at end of maintenance phase	Stable remitters[a]: 51/89 (57.3%)	Stable remitters[a]: 38/86 (44.2%)
	Stable responders[b]: 23/61 (37.7%)	Stable responders[b]: 12/58 (20.7%)
GAD-7 (LOCF)[i]		
Change from baseline to end of maintenance phase (mean, SD)[e]	Stable remitters[a]: N=89, 2.2 (4.45)	Stable remitters[a]: N=86, 4.0 (5.93)
	Stable responders[b]: N=61, 1.4 (3.76)	Stable responders[b]: N=58, 2.6 (4.26)
ANCOVA (difference in LS means, SE, 95% CI)[f]	Stable remitters[a]: -1.7 (0.72, -3.12 to -0.28)	
	Stable responders[b]: -1.1 (0.72, -2.56 to 0.31)	
SDS (LOCF)[g]		
Change from baseline		
Change from baseline to end of maintenance phase (mean, SD)[e]	Stable remitters[a]: N=82, 4.7 (7.34)	Stable remitters[a]: N=77, 7.2 (10.44)
	Stable responders[b]: N=58, 2.2 (6.63)	Stable responders[b]: N=53, 6.8 (7.64)
ANCOVA (difference in LS means, SE, 95% CI)[f]	Stable remitters[a]: -2.9 (1.30, -5.51 to -0.38)	
	Stable responders[b]: -4.7 (1.31, -7.30 to -2.10)	
Response/remission		
Responder at beginning of maintenance phase	Stable remitters[a]: 84/89 (94.4%)	Stable remitters[a]: 74/84 (88.1%)
	Stable responders[b]: 45/60 (75.0%)	Stable responders[b]: 48/57 (84.2%)
Responder at end of maintenance phase[e]	Stable remitters[a]: 58/83 (69.9%)	Stable remitters[a]: 43/78 (55.1%)
	Stable responders[b]: 42/60 (70.0%)	Stable responders[b]: 23/53 (43.4%)
Remitter at beginning of maintenance phase	Stable remitters[a]: 72/89 (80.9%)	Stable remitters[a]: 63/84 (75.0%)

Outcome	ESK-NS + OAD	OAD + PBO-NS
	Stable responders[b]: 28/60 (46.7%)	Stable responders[b]: 30/57 (52.6%)
Remitter at end of maintenance phase[e]	Stable remitters[a]: 48/83 (57.8%)	Stable remitters[a]: 30/78 (38.5%)
	Stable responders[b]: 25/60 (41.7%)	Stable responders[b]: 11/53 (20.8%)
EQ-5D (HSI score)[h]		
Start of maintenance phase (mean, SD)	Stable remitters[a]: N=90, 0.925 (0.0440)	Stable remitters[a]: N=86, 0.918 (0.0422)
	Stable responders[b]: N=62, 0.877 (0.0664)	Stable responders[b]: N=59, 0.875 (0.0796)
End of maintenance phase (mean, SD)[e]	Stable remitters[a]: N=88, 0.857 (0.1275)	Stable remitters[a]: N=90, 0.822 (0.1442)
	Stable responders[b]: N=61, 0.855 (0.0880)	Stable responders[b]: N=58, 0.802 (0.1292)
Change from baseline to end of maintenance phase (mean, SD)[f]	Stable remitters[a]: N=88, -0.067 (0.1180)	Stable remitters[a]: N=86, -0.096 (0.1484)
	Stable responders[b]: N=61, -0.023 (0.0753)	Stable responders[b]: N=58, -0.073 (0.1383)
Other outcomes defined in the final scope		
Cognitive dysfunction	NR	NR
Hospitalisation	NR	NR
Sleep quality	NR	NR

Based on Tables 7, 8, 27, 28, 29, 30 of the CS[1]

[a] Patients who were in stable remission at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase; [b] Patients who were stable responders (who were not stable remitters) at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase; [c] Related to relapse, severity of depression, and remission (Table 4.5); [d] Used in the economic model; [e] Variable duration (until relapse or study termination); [f] Change from baseline was the response variable and treatment, country, and baseline value were covariates; [g] Related to severity of depression (Table 4.5); [h] No further information reported, [i] Related to anxiety (Table 4.5)

ANCOVA = analysis of covariance; CGI-S = Clinical Global Impression; CI = confidence interval; CS = company submission; EQ-5D = European Quality of Life-5 Dimensions; ESK = esketamine; GAD-7 = Generalised Anxiety Disorder – 7-item scale; HR = hazard ratio; HSI = health status index; LOCF = last observation carried forward; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; NR = not reported; NS = nasal spray; OAD = oral antidepressant; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error

Table 4.14: Summary of efficacy results of TRANSFORM-3

Outcome	ESK-NS + OAD	OAD + PBO-NS
MADRS		
Change from baseline (observed cases)		
Baseline (mean, SD)	N=72, 35.5 (5.91)	N=65, 34.8 (6.44)
Day 28 (mean, SD)	N=63, 25.4 (12.70)	N=60, 28.7 (10.11)
Change from baseline to day 28 (mean, SD)	N=63, -10.0 (12.74)	N=60, -6.3 (8.86)

Outcome	ESK-NS + OAD	OAD + PBO-NS
MMRM (difference in LS means, SE, 95% CI)[a]	-3.6 (NR, -7.20 to -0.07)	
Response and remission (observed cases)		
Response rate	17/63 (27.0%)	8/60 (13.3%)
Remission rate	11/63 (17.5%)	4/60 (6.7%)
CGI-S (observed cases)		
Baseline (mean, SD)	N=72, 5.1 (0.76)	N=65, 4.8 (0.80)
Day 28 (mean, SD)	N=64, 3.9 (1.33)	N=60, 4.3 (1.20)
Change from baseline to day 28 (mean, SD)	N=64, -1.2 (1.30)	N=60, -0.5 (1.03)
MMRM (difference in LS means, SE, 95% CI)[a]	-0.7 (0.21, -1.10 to -0.27)	
PHQ-9 (observed cases)		
Baseline (mean, SD)	N=72, 17.6 (4.99)	N=65, 17.4 (6.33)
Day 28 (mean, SD)	N=64, 11.6 (7.04)	N=57, 13.5 (6.81)
Change from baseline to day 28 (mean, SD)	N=64, -6.4 (7.24)	N=57, -4.1 (6.36)
MMRM (difference in LS means, SE, 95% CI)[a]	-2.8 (1.16, -5.08 to -0.48)	
GAD-7 (observed cases)		
Baseline (mean, SD)	NR	NR
Day 28 (mean, SD)	NR	NR
Change from baseline to day 28 (mean, SD)	NR	NR
ANCOVA (difference in LS means, SE, 95% CI)	NR	
SDS (observed cases)		
Change from baseline		
Baseline (mean, SD)	N=45, 21.8 (5.90)	N=44, 22.9 (4.74)
Day 28 (mean, SD)	N=36, 14.3 (9.33)	N=37, 19.2 (7.25)
Change from baseline to day 28 (mean, SD)	N=29, -7.5 (8.24)	N=37[b], -3.8 (5.57)
MMRM (difference in LS means, SE, 95% CI)[a]	-4.6 (1.82, -8.21 to -0.94)	
Response and remission		
Response rate	15/44 (34.1%)	10/44 (22.7%)
Remission rate	7/44 (15.9%)	2/44 (4.5%)
EQ-5D (observed cases)		
Baseline (mean, SD)	N=72, 0.581 (0.2258)	N=65, 0.635 (0.2276)
Day 28 (mean, SD)	N=65, 0.658 (0.2608)	N=59, 0.680 (0.1918)
Change from baseline to day 28 (mean, SD)	N=65, 0.086 (0.2674)	N=59, 0.041 (0.2074)

Outcome	ESK-NS + OAD	OAD + PBO-NS
Difference in LS means, SE, 95% CI	NR	NR
Other outcomes defined in the final scope		
Cognitive dysfunction	NR	NR
Hospitalisation	NR	NR
Sleep quality	NR	NR

Based on Tables 30 to 35 of the response to request for clarification[3]

[a] Change from baseline was the response variable and fixed effect model terms for treatment, day, country, class of OAD (SNRI or SSRI), treatment-by-day, and baseline value were covariates; [b] Table 32 of the response to request for clarification[3] reported this as "85". Error corrected by the ERG

CGI-S = Clinical Global Impression; CI = confidence interval; EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; ESK = esketamine; GAD-7 = Generalised Anxiety Disorder – 7-item scale; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; NR = not reported; NS = nasal spray; OAD = oral antidepressant; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Table 4.15: Summary of efficacy results of SUSTAIN-2

Outcome	ESK-NS + OAD
MADRS	
Change from baseline (LOCF)	
Baseline (mean, SD)	N=779, 31.2 (5.29)
End of induction (mean, SD)	N=756, 14.8 (8.83)
Change from baseline to end of induction (mean, SD)	N=756, -16.4 (8.76)
Response/remission (observed cases)	
Responder at beginning of induction	NR
Responder at end of induction	581/688 (84.4%)
Remitter at beginning of induction	NR
Remitter at end of induction	349/688 (50.7%)
CGI-S (LOCF)	
Baseline (median, range)	N=779, 5.0 (1 to 7)
End of induction (median, range)	N=763, 3.0 (1 to 7)
Change from baseline to end of induction (median, range)	N=763, -2.0 (-6 to 2)
PHQ-9 (LOCF)	
Baseline (mean, SD)	N=779, 17.3 (5.00)
End of induction (mean, SD)	N=746, 8.4 (5.80)
Change from baseline to end of induction (mean, SD)	N=746, -8.9 (6.67)
GAD-7 (LOCF)	
Baseline (mean, SD)	N=771, 11.3 (5.45)
End of induction (mean, SD)	N=732, 5.3 (NR)
Change from baseline to end of induction (mean, SD)	N=724, -5.9 (5.85)

Outcome	ESK-NS + OAD
SDS (LOCF)	
Change from baseline	
Baseline (mean, SD)	N=709, 22.2 (5.45)
End of induction (mean, SD)	N=648, 12.8 (7.89)
Change from baseline to end of induction (mean, SD)	N=626, -9.3 (7.86)
Response/remission	
Responder at beginning of induction	NR
Responder at end of induction	295/571 (51.7%)
Remitter at beginning of induction	NR
Remitter at end of induction	132/571 (23.1%)
EQ-5D (HSI score)	
Start of induction (mean, SD)	N=779, 0.601 (0.2056)
End of induction (mean, SD)	N=745, 0.792 (0.1725)
Change from baseline to end of induction phase (mean, SD)	N=745, 0.190 (0.2138)
Other outcomes defined in the final scope	
Cognitive dysfunction	NR
Hospitalisation	NR
Sleep quality	NR

Based on Tables 39 to 45 of the response to request for clarification[3]

CGI-S = Clinical Global Impression; EQ-5D = European Quality of Life-5 Dimensions; ESK = esketamine; GAD-7 = Generalised Anxiety Disorder – 7-item scale; HSI = health status index; LOCF = last observation carried forward; MADRS = Montgomery-Åsberg Depression Rating Scale; NR = not reported; NS = nasal spray; OAD = oral antidepressant; PBO = placebo; PHQ-9 = Patient Health Questionnaire – 9 questions; SD = standard deviation; SDS = Sheehan Disability Scale

ERG comment: Tables 4.12 and 4.13 summarise the efficacy results of TRANSFORM-2 and SUSTAIN-1, respectively, which are the RCTs used to inform the economic model. However, some outcomes defined in the final scope issued by NICE have not been reported in the CS, namely cognitive dysfunction, hospitalisation and sleep quality (see Table 3.1).[1]

Both of these trials report on a number of outcomes, however, it should be noted that according to Table 7 of the CS (see Table 4.5), only response and remission based on MADRS (TRANSFORM-2) and relapse (SUSTAIN-1) are used in the economic model.[1]

In TRANSFORM-2 (Table 4.12), ESK-NS + OAD in comparison to PBO-NS + OAD showed a statistically significant reduction of MADRS at day 28 (difference in LS means -4.0, 95% CI -7.31 to 5.67). The trial also showed differences in response rate and remission rate, respectively, between the two groups. For the control arm of the trial, adjusted and unadjusted estimates are reported. As discussed in section 3.3, the ERG prefers the use of unadjusted estimates. Other reported outcomes (CGI-S, PHQ-9, GAD-7, SDS and EQ-5D) were in favour of the intervention (see Table 4.12 for details).

SUSTAIN-1 reported results separately for participants considered stable remitters (defined as *"patients who were in stable remission at the end of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase"*) and stable responders (defined as *"patients who were stable responders (who were not stable remitters) at the end*

of the optimisation phase and who received at least 1 dose of intranasal study drug and 1 dose of OAD during the maintenance phase"). As shown in Table 4.13, the percentage of relapse was lower in the ESK-NS + OAD (stable remitters: 26.7%, stable responders: 25.8%) group in comparison to participants receiving PBO-NS + OAD (45.3% and 57.6%, respectively). The trial also showed time to relapse to be in favour of the intervention group for both, stable remitters (HR 0.49, 95% CI 0.29 to 0.84) and stable responders (HR 0.30, 95% CI 0.16 to 0.55). Other reported outcomes (CGI-S, PHQ-9, GAD-7, SDS and EQ-5D) were in favour of the intervention (see Table 4.13 for details). However, it should be noted that these results are based on last observation carried forward (LOCF) which fails to acknowledge uncertainty in the imputed values and results, typically, in confidence intervals that are too narrow.[24]

The results for TRANSFORM-2 and SUSTAIN-1 are in line with those of TRANSFORM-3 and SUSTAIN-2 which have been summarised in Tables 4.14 and 4.15.

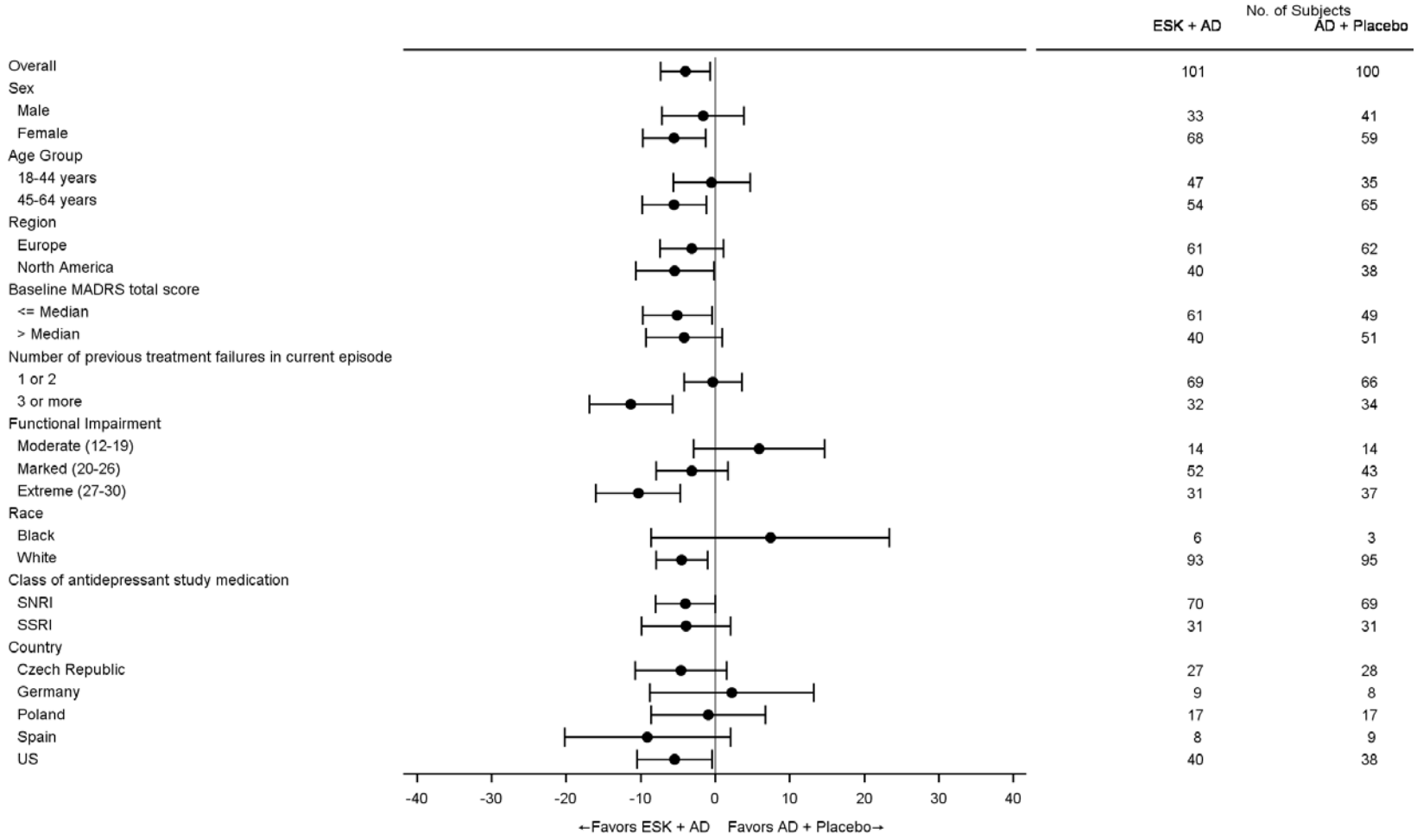
4.2.6 Subgroup analysis

Table 8 of the CS[1] listed pre-planned subgroups for TRANSFORM-2 and SUSTAIN-1:

- *Gender; race (White, Black, Other); country; number of previous treatment failures in current episode (based on MGH-ATRQ); class of OAD study medication (SNRI or SSRI)*
- *Functional impairment based on baseline SDS total score: not impaired (0–3), mild (4–11), moderate (12–19), marked (20–26), extreme (27–30)*
- *Age group (18–44 years, 45–64 years)*
- *Region (North America, Europe, Other)*
- *Baseline MADRS total score (≤/> median) (TRANSFORM-2 only)*
- *Consented protocol (pre-/post-protocol amendment 4) (SUSTAIN-1 only)*
- *Study entry route (direct-entry, transferred-entry) (SUSTAIN-1 only)*
- *OAD (duloxetine, escitalopram, sertraline, venlafaxine XR) (SUSTAIN-1 only)*

Figure 4.1 shows the differences by subgroup for TRANSFORM-2 in a forest plot. Table 4.16 gives further details. Based on information received in response to the request for clarification, Table 4.16 also includes details on unadjusted response and remission rates by OAD class and type as well as by disease severity.[3] Similarly, a forest plot for SUSTAIN-1 is presented in Figure 4.2 (no further details were provided).

Figure 4.1: Forest plot of LS mean treatment difference (95% CI) in change in MADRS total score from baseline to Day 28 by subgroup (MMRM; full analysis set) – TRANSFORM-2



Based on Figure 15 of the CS appendices[17]

AD = antidepressant; CI = confidence interval; CS = company submission; ESK = esketamine; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; US = United States (of America)

Table 4.16: MADRS total score: change from baseline to the end of induction by subgroup (observed cases MMRM and LOCF ANCOVA; full analysis set) – TRANSFORM-2

Subgroup	ESK-NS + OAD N=114	OAD + PBO-NS N=109
Gender		
Male		
Mean (SD) CFB to Day 28 (OC)	–20.5 (11.85) (n=33)	–18.3 (13.19) (n=41)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–1.7 (2.80; –7.17 to 3.86)	
Mean (SD) CFB to endpoint (LOCF)	–17.9 (13.92) (n=39)	–17.1 (13.95) (n=46)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–2.2 (2.65; –7.45 to 3.01)	
Female		
Mean (SD) CFB to Day 28 (OC)	–21.9 (12.61) (n=68)	–16.1 (14.38) (n=59)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–5.5 (2.13; –9.71 to –1.31)	
Mean (SD) CFB to endpoint (LOCF)	–20.4 (13.42) (n=73)	–15.7 (14.53) (n=63)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–4.4 (2.09; –8.52 to –0.26)	
Age group		
18–44 years		
Mean (SD) CFB to Day 28 (OC)	–23.1 (11.01) (n=47)	–2.5 (12.64) (n=35)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–0.5 (2.62; –5.64 to 4.69)	
Mean (SD) CFB to endpoint (LOCF)	–20.7 (13.14) (n=54)	–22.0 (13.05) (n=40)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–0.6 (2.52; –5.53 to 4.41)	
45–64 years		
Mean (SD) CFB to Day 28 (OC)	–20.0 (13.30) (n=54)	–14.0 (13.69) (n=65)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–5.5 (2.19; –9.82 to –1.18)	
Mean (SD) CFB to endpoint (LOCF)	–18.5 (14.01) (n=58)	–13.0 (13.95) (n=69)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–4.8 (2.13; –8.99 to –0.58)	
Region		
Europe		
Mean (SD) CFB to Day 28 (OC)	–22.3 (12.83) (n=61)	–19.4 (13.93) (n=62)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–3.2 (2.16; –7.42 to 1.09)	
Mean (SD) CFB to endpoint (LOCF)	–20.1 (14.08) (n=68)	–18.2 (14.63) (n=65)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–2.9 (2.18; –7.18 to 1.42)	
North America		
Mean (SD) CFB to Day 28 (OC)	–20.1 (11.54) (n=40)	–13.1 (13.07) (n=38)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–5.4 (2.66; –10.69 to –0.18)	
Mean (SD) CFB to endpoint (LOCF)	–18.7 (12.90) (n=44)	–13.5 (13.31) (n=44)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–4.6 (2.68; –9.88 to 0.66)	
Baseline MADRS total score		
≤37		
Mean (SD) CFB to Day 28 (OC)	–17.7 (11.17) (n=61)	–12.6 (12.75) (n=49)

Subgroup	ESK-NS + OAD N=114	OAD + PBO-NS N=109
Diff in LS means (SE; 95% CI) (MMRM)[a]	–5.1 (2.36; –9.74 to –0.43)	
Mean (SD) CFB to endpoint (LOCF)	–16.2 (12.35) (n=65)	–11.3 (12.84) (n=55)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–5.7 (2.33; –10.27 to –1.10)	
>37		
Mean (SD) CFB to Day 28 (OC)	–27.2 (11.90) (n=40)	–21.2 (13.72) (n=51)
Diff in LS means (SE; 95% CI) (MMRM)[a]	–4.2 (2.59; –9.27 to 0.94)	
Mean (SD) CFB to endpoint (LOCF)	–24.2 (13.97) (n=47)	–21.4 (13.90) (n=54)
Diff in LS means (SE; 95% CI) (ANCOVA)[b]	–1.5 (2.54; –6.54 to 3.45)	
Number of previous treatment failures in the current episode of depression (induction phase)		
2[c]		
Mean (SD) CFB to Day 28 (OC)	–20.4 (11.91) (n=59)	–21.0 (12.89) (n=64)
Diff in LS means (SE; 95% CI) (MMRM)[d]	0.5 (2.08; –3.60 to 4.59)	
Mean (SD) CFB to Day 28 (LOCF)	–19.0 (12.54) (n=64)	–19.8 (13.61) (n=70)
Diff in LS means (SE; 95% CI) (ANCOVA)[d]	–0.1 (2.06; –4.15 to 3.98)	
≥3		
Mean (SD) CFB to Day 28 (OC)	–22.7 (12.77) (n=38)	–10.3 (12.95) (n=35)
Diff in LS means (SE; 95% CI) (MMRM)[d]	–11.5 (2.70; –16.85 to –6.22)	
Mean (SD) CFB to Day 28 (LOCF)	–19.9 (15.02) (n=44)	–10.3 (13.33) (n=38)
Diff in LS means (SE; 95% CI) (ANCOVA)[d]	–9.1 (2.65; –14.30 to –3.84)	
Disease severity		
Remission		
Moderate (MADRS total score at baseline 18-34) (n=65)	56.25	30.30
	OR 2.96 (95% CI 1.07 to 8.20)	
Severe (MADRS total score at baseline >34) (n=136)	50.72	31.34
	OR 2.26 (95% CI 1.12 to 4.54)	
Response		
Moderate (MADRS total score at baseline 18-34) (n=65)	59.38	36.36
	OR 2.56 (95% CI 0.94 to 6.96)	
Severe (MADRS total score at baseline >34) (n=136)	73.91	59.70
	OR 1.91 (95% CI 0.93 to 3.95)	
Functional impairment (assessed by SDS)		
Mild (SDS: 4–11)		
Mean (SD) CFB to Day 28 (OC)	–22.0 (-) (n=1)	–9.0 (-) (n=1)
Diff in LS means (SE; 95% CI) (MMRM)[e]	–15.6 (17.00; –49.07 to 17.97)	
Mean (SD) CFB to endpoint (LOCF)	–22.0 (-) (n=1)	–9.0 (-) (n=1)
Diff in LS means (SE; 95% CI) (ANCOVA)[f]	–10.5 (17.37; –44.77 to 23.74)	
Moderate (SDS: 12–19)		
Mean (SD) CFB to Day 28 (OC)	–14.9 (11.25) (n=14)	–22.8 (13.62) (n=14)

Subgroup	ESK-NS + OAD N=114	OAD + PBO-NS N=109
Diff in LS means (SE; 95% CI) (MMRM)[e]	5.9 (4.45; –2.89 to 14.64)	
Mean (SD) CFB to endpoint (LOCF)	–13.9 (11.50) (n=15)	–19.8 (14.13) (n=17)
Diff in LS means (SE; 95% CI) (ANCOVA)[f]	1.4 (4.37; –7.24 to 10.01)	
Marked (SDS: 20–26)		
Mean (SD) CFB to Day 28 (OC)	–20.8 (11.88) (n=52)	–16.8 (13.27) (n=43)
Diff in LS means (SE; 95% CI) (MMRM)[e]	–3.1 (2.44; –7.96 to 1.67)	
Mean (SD) CFB to endpoint (LOCF)	–18.4 (13.53) (n=58)	–16.9 (13.30) (n=45)
Diff in LS means (SE; 95% CI) (ANCOVA)[f]	–2.7 (2.43; –7.47 to 2.10)	
Extreme (SDS: 27–30)		
Mean (SD) CFB to Day 28 (OC)	–26.2 (12.16) (n=31)	–14.8 (14.88) (n=37)
Diff in LS means (SE; 95% CI) (MMRM)[e]	–10.3 (2.87; –16.00 to –4.66)	
Mean (SD) CFB to endpoint (LOCF)	–24.3 (13.68) (n=35)	–13.9 (15.65) (n=41)
Diff in LS means (SE; 95% CI) (ANCOVA)[f]	–7.6 (2.84; –13.22 to –2.03)	
Race		
Black		
Mean (SD) CFB to Day 28 (OC)	–16.8 (9.60) (n=6)	–18.3 (17.21) (n=3)
Diff in LS means (SE; 95% CI) (MMRM)[g]	7.4 (8.13; –8.63 to 23.41)	
Mean (SD) CFB to endpoint (LOCF)	–16.8 (9.60) (n=6)	–17.6 (16.89) (n=5)
Diff in LS means (SE; 95% CI) (ANCOVA)[h]	4.3 (7.38; –10.22 to 18.86)	
White		
Mean (SD) CFB to Day 28 (OC)	–21.8 (12.41) (n=93)	–17.0 (13.99) (n=95)
Diff in LS means (SE; 95% CI) (MMRM)[g]	–4.5 (1.76; –7.96 to –1.03)	
Mean (SD) CFB to endpoint (LOCF)	–19.7 (13.79) (n=104)	–16.2 (14.30) (n=102)
Diff in LS means (SE; 95% CI) (ANCOVA)[h]	–3.8 (1.69; –7.11 to –0.44)	
Other		
Mean (SD) CFB to Day 28 (OC)	–18.0 (19.80) (n=2)	–16.5 (9.19) (n=2)
Diff in LS means (SE; 95% CI) (MMRM)[g]	–5.5 (12.25; –29.69 to 18.62)	
Mean (SD) CFB to endpoint (LOCF)	–18.0 (19.80) (n=2)	–16.5 (9.19) (n=2)
Diff in LS means (SE; 95% CI) (ANCOVA)[h]	–8.6 (12.19; –35.59 to 15.46)	
Class of OAD		
SNRI		
Mean (SD) CFB to Day 28 (OC)	–22.0 (11.99) (n=70)	–18.1 (13.88) (n=69)
Diff in LS means (SE; 95% CI) (MMRM)[i]	–4.0 (2.04; –8.02 to 0.03)	
Mean (SD) CFB to endpoint (LOCF)	–20.8 (12.92) (n=76)	–17.0 (14.40) (n=75)
Diff in LS means (SE; 95% CI) (ANCOVA)[j]	–4.0 (1.97; –7.87 to –0.11)	
SSRI		
Mean (SD) CFB to Day 28 (OC)	–20.1 (13.13) (n=31)	–14.6 (13.81) (n=31)
Diff in LS means (SE; 95% CI) (MMRM)[i]	–3.9 (3.04; –9.91 to 2.08)	

Subgroup	ESK-NS + OAD N=114	OAD + PBO-NS N=109
Mean (SD) CFB to endpoint (LOCF)	–16.8 (14.72) (n=36)	–14.8 (13.97) (n=34)
Diff in LS means (SE; 95% CI) (ANCOVA)[j]	–2.3 (2.90; –8.03 to 3.38)	
Type of OAD		
Day 28 Remission rates (%)		
SSRI	51.61 (n=36)	25.81 (n=34)
	OR 3.07 (95% CI 1.05 to 8.93)	
Sertraline	33.33 (n=15)	26.67 (n=16)
	OR 1.38 (95% CI 0.26 to 7.22)	
Escitalopram	63.16 (n=21)	26.67 (n=17)
	OR 4.71 (95% CI 1.08 to 20.63)	
SNRI	52.86 (n=76)	33.33 (n=75)
	OR 2.24 (95% CI 1.13 to 4.45)	
Duloxetine	50.00 (n=59)	32.73 (n=61)
	OR 2.06 (95% CI 0.95 to 4.47)	
Venlafaxine XR	62.50 (n=17)	33.33 (n=15)
	OR 3.33 (95% CI 0.76 to 14.58)	
Day 28 Response rates (%)		
SSRI	67.74 (n=36)	45.16 (n=34)
	OR 2.55 (95% CI 0.91 to 7.17)	
Sertraline	58.33 (n=15)	33.33 (n=16)
	OR 2.80 (95% CI 0.58 to 13.48)	
Escitalopram	73.68 (n=21)	53.33 (n=17)
	OR 2.45 (95% CI 0.58 to 10.33)	
SNRI	70.00 (n=76)	55.07 (n=75)
	OR 1.90 (95% CI 0.95 to 3.82)	
Duloxetine	70.37 (n=59)	60.00 (n=61)
	OR 1.58 (95% CI 0.72 to 3.51)	
Venlafaxine XR	68.75 (n=17)	40.00 (n=15)
	OR 3.30 (95% CI 0.75 to 14.47)	
Country		
Czech Republic		
Mean (SD) CFB to Day 28 (OC)	–26.8 (10.78) (n=27)	–21.8 (15.34) (n=28)
Diff in LS means (SE; 95% CI) (MMRM)[k]	–4.6 (3.11; –10.72 to 1.54)	
Mean (SD) CFB to endpoint (LOCF)	–24.8 (13.25) (n=29)	–21.8 (15.34) (n=28)
Diff in LS means (SE; 95% CI) (ANCOVA)[l]	–3.7 (3.21; –10.00 to 2.65)	
Germany		
Mean (SD) CFB to Day 28 (OC)	–10.2 (12.43) (n=9)	–13.5 (9.09) (n=8)
Diff in LS means (SE; 95% CI) (MMRM)[k]	2.2 (5.60; –8.81 to 13.25)	

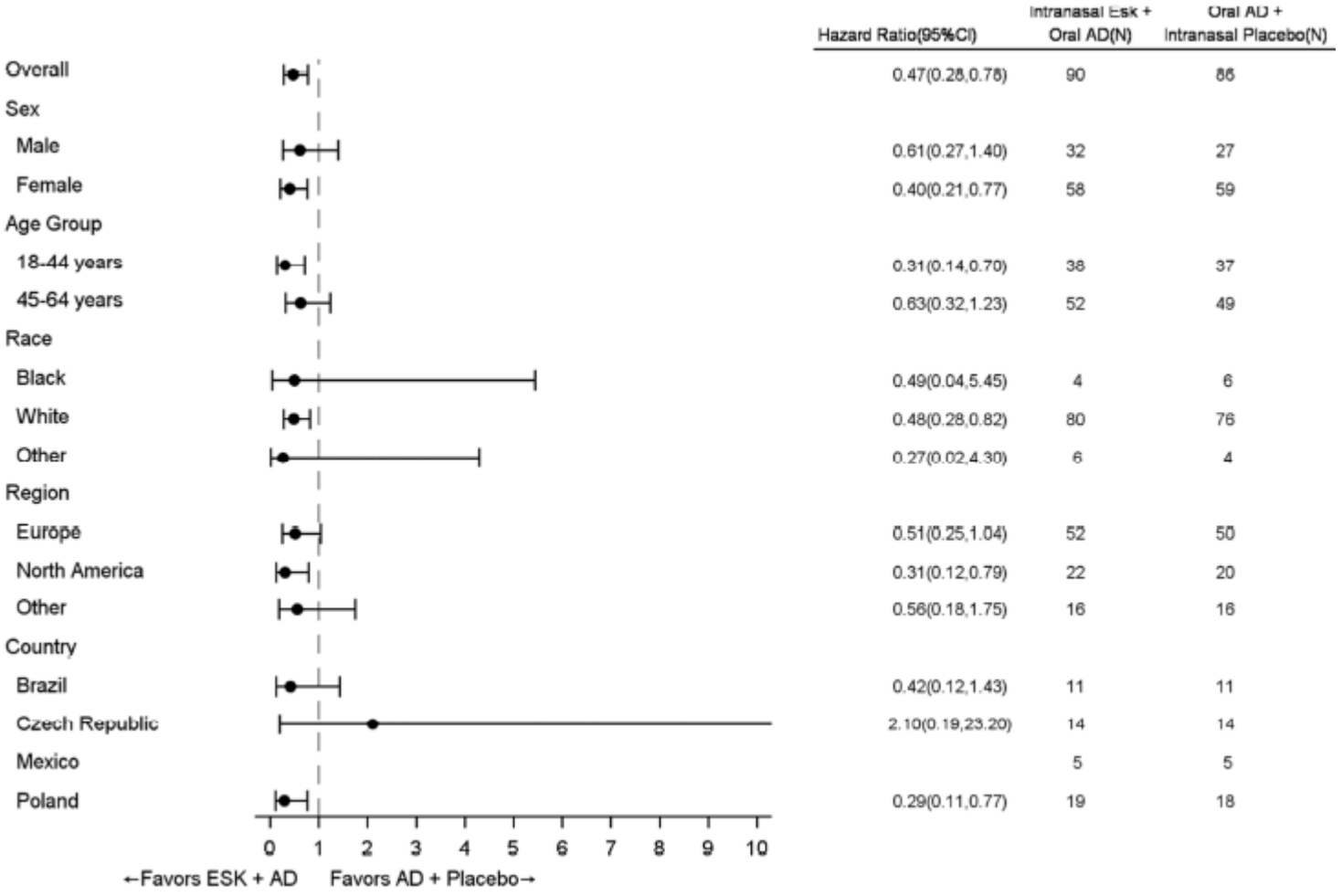
Subgroup	ESK-NS + OAD N=114	OAD + PBO-NS N=109
Mean (SD) CFB to endpoint (LOCF)	–9.2 (12.15) (n=10)	–10.1 (10.96) (n=10)
Diff in LS means (SE; 95% CI) (ANCOVA)[l]	–0.6 (5.42; –11.30 to 10.07)	
Poland		
Mean (SD) CFB to Day 28 (OC)	–24.8 (9.59) (n=17)	–23.6 (8.48) (n=17
Diff in LS means (SE; 95% CI) (MMRM)[k]	–0.9 (3.89; –8.57 to 6.78)	
Mean (SD) CFB to endpoint (LOCF)	–21.6 (12.12) (n=20)	–21.7 (11.42) (n=18)
Diff in LS means (SE; 95% CI) (ANCOVA)[l]	0.8 (3.94; –6.93 to 8.60)	
Spain		
Mean (SD) CFB to Day 28 (OC)	–15.5 (16.04) (n=8)	–9.1 (16.09) (n=9)
Diff in LS means (SE; 95% CI) (MMRM)[k]	–9.1 (5.63; –20.18 to 2.00)	
Mean (SD) CFB to endpoint (LOCF)	–13.8 (15.86) (n=9)	–9.1 (16.09) (n=9)
Diff in LS means (SE; 95% CI) (ANCOVA)[l]	–9.4 (5.75; –20.71 to 1.96)	
United States		
Mean (SD) CFB to Day 28 (OC)	–20.1 (11.54) (n=40)	–13.1 (13.07) (n=38)
Diff in LS means (SE; 95% CI) (MMRM)[k]	–5.5 (2.56; –10.52 to –0.44)	
Mean (SD) CFB to endpoint (LOCF)	–18.7 (12.90) (n=44)	–13.5 (13.31) (n=44)
Diff in LS means (SE; 95% CI) (ANCOVA)[l]	–4.6 (2.58; –9.72 to 0.46)	

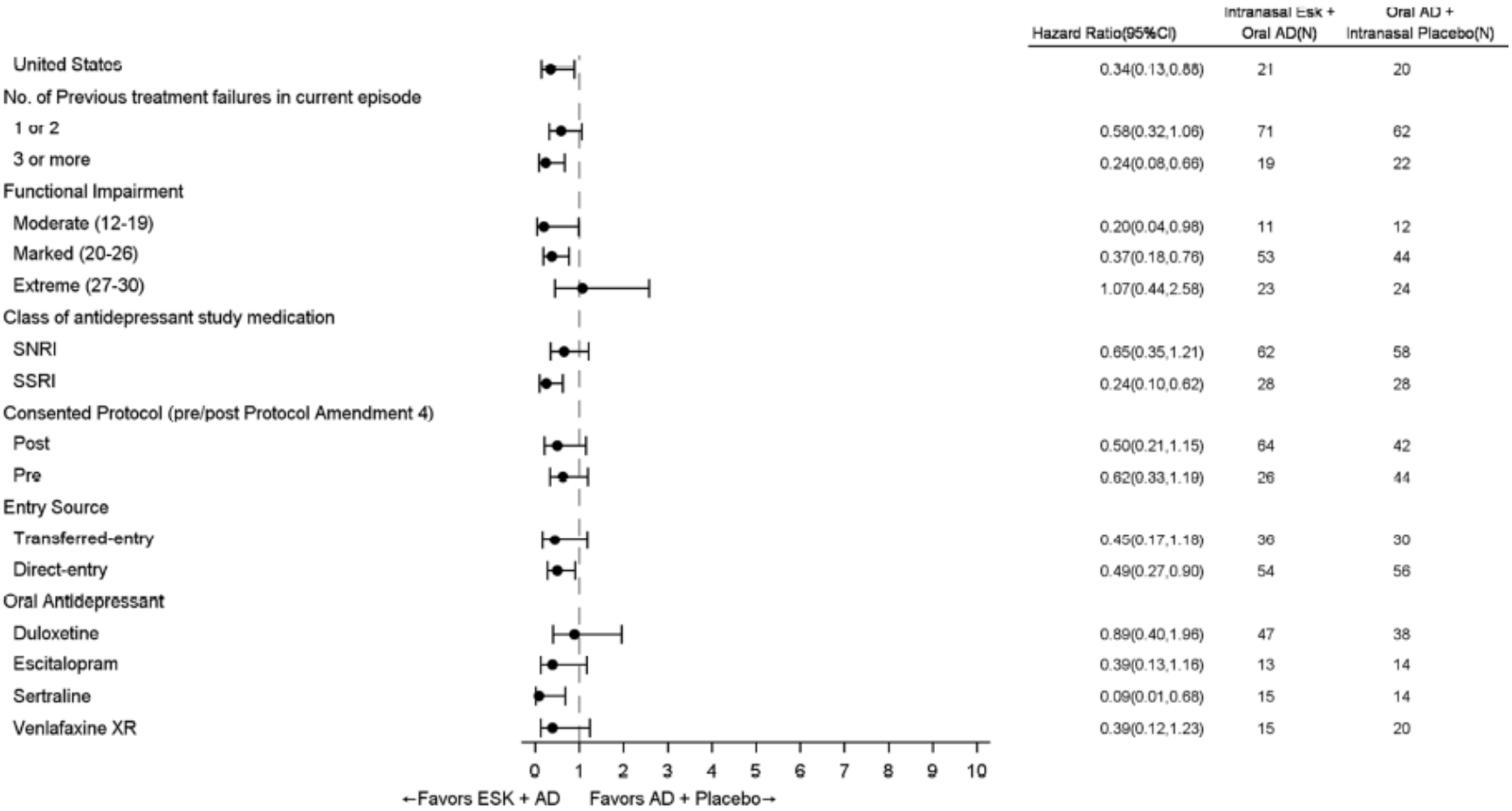
Based on Table 52 of the CS appendices[19] and Tables 3 and 8 of the response to the request for clarification[3]

[a] Fixed effect model adjusted for treatment, day, country, OAD (SNRI or SSRI), sex, treatment-by-day, treatment-by-sex, treatment-by-day-by-sex, and baseline value; [b] Adjusting for treatment, country, OAD , treatment-by-sex, and baseline MADRS value were covariates; [c] The minimum number of prior OADs to which patients could have not responded to at the beginning of induction was two since patients had to demonstrate non-response to one OAD during the screening/prospective observation phase; [d] Fixed effect model adjusted for treatment, day, country, OAD, number of previous treatment failures in current episode, treatment-by-day, treatment-by-number of previous treatment failures in current episode, treatment-by-day-by-number of previous treatment failures in current episode, and baseline value; [e] Fixed effect model adjusted for treatment, day, country, OAD , functional impairment, treatment-by-day, treatment-by-functional impairment, treatment-by-day-by-functional impairment, and baseline value; [f] Adjusted for treatment, country, OAD , functional impairment, treatment-by-functional impairment, and baseline value; [g] Fixed effect model adjusted for treatment, day, country, OAD , race, treatment-by-day, treatment-by-race, treatment-by-day-by-race, and baseline value; [h] Adjusted for treatment, country, OAD , race, treatment-by-race, and baseline MADRS value; [i] Adjusted for treatment, day, country, OAD, treatment-by-day, treatment-by- OAD, treatment-by-day-by- OAD, and baseline value; [j] Adjusted for treatment, country, OAD , treatment-by- OAD, and baseline MADRS value; [k] Adjusted for treatment, day, country, OAD, treatment-by-day, treatment-by-country, treatment-by-day-by-country, and baseline value; [l] Adjusted for treatment, country, OAD , treatment-by-country, and baseline MADRS value

ANCOVA = analysis of covariance; CFB = change from baseline; CI = confidence interval; CS = company submission; Diff = difference; ESK-NS = esketamine nasal spray; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; OAD = oral antidepressant; OC = observed cases; OR = odds ratio; SD = standard deviation; SDS = Sheehan Disability Scale; SE = standard error; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Figure 4.2: Forest plot of LS mean treatment difference (95% CI) in change in MADRS total score from baseline to Day 28 by subgroup (MMRM; full analysis set) – SUSTAIN-1





Based on Figure 16 of the CS appendices[17]

AD = antidepressant; CI = confidence interval; CS = company submission; ESK = esketamine; LS = least squares; MADRS = Montgomery-Åsberg Depression Rating Scale; MMRM = mixed-effects model using repeated measures; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

ERG comment: Due to the small numbers of participants in each arm, any differences in the subgroup analyses need to be interpreted with some caution.

A visual inspection of TRANSFORM-2 (Figure 4.1) indicate some differences between gender, number of previous treatment failures in current episode, functional impairment, and race. Furthermore, the advantage of esketamine compared to the control group seems bigger in remission rather than relapse, see Table 4.16 for details. Of note, there are differences between the type of OAD for remission rates after 28 days, e.g. within the SSRI group: sertraline (odds ratio (OR) 1.38, 95% CI 0.26 to 7.22) vs. escitalopram (OR 4.71, 95% CI 1.08 to 20.63). This might indicate a clinically relevant limitation of the basket approach used in the economic model.

No relevant differences were noted when visually inspecting the forest plot provided for SUSTAIN-1, replicated in Figure 4.2. However, no further details were provided, preventing a closer examination.

4.2.7 Safety results

Safety results for TRANSFORM-2 (Table 4.17) and SUSTAIN-1 (Tables 4.18 and 4.19), the trials used in the economic model, are reported below. Furthermore, safety results for TRANSFORM-3 (Table 4.20) and SUSTAIN-2 (Tables 4.21 and 4.22) are presented.

Table 4.17: Safety results of TRANSFORM-2

	ESK-NS + OAD	OAD + PBO-NS
Induction phase, n (%)	**N=115**	**N=109**
Overall summary		
AE	98 (85.2)	66 (60.6)
AE possibly related to nasal spray drug[a]	90 (78.3)	39 (35.8)
AE possibly related to OAD[a]	39 (33.9)	26 (23.9)
AE leading to death	1 (0.9)	0
≥1 serious AE	1 (0.9)	1 (0.9)
AE leading to nasal spray drug being withdrawn[b]	8 (7.0)	1 (0.9)
AE leading to OAD being withdrawn[b]	4 (3.5)	0
AEs reported in ≥5% of patients		
Nervous system disorders, n (%)	**72 (62.6)**	**39 (35.8)**
Dysgeusia	28 (24.3)	13 (11.9)
Dizziness	24 (20.9)	5 (4.6)
Headache	23 (20.0)	19 (17.4)
Somnolence	15 (13.0)	7 (6.4)
Paraesthesia	13 (11.3)	1 (0.9)
Dizziness postural	8 (7.0)	1 (0.9)
Hypoaesthesia	8 (7.0)	1 (0.9)
Psychiatric disorders, n (%)	**55 (47.8)**	**21 (19.3)**
Dissociation[c]	30 (26.1)	4 (3.7)
Anxiety	12 (10.4)	5 (4.6)
Insomnia	11 (9.6)	5 (4.6)

	ESK-NS + OAD	OAD + PBO-NS
Gastrointestinal disorders, n (%)	**52 (42.5)**	**26 (23.9)**
Nausea	30 (26.1)	7 (6.4)
Vomiting	11 (9.6)	2 (1.8)
Diarrhoea	10 (8.7)	10 (9.2)
Dry mouth	9 (7.8)	3 (2.8)
Hypoaesthesia oral	9 (7.8)	1 (0.9)
Paraesthesia oral	9 (7.8)	1 (0.9)
Ear and labyrinth disorders, n (%)	**34 (29.6)**	**6 (5.5)**
Vertigo	30 (26.1)	3 (2.8)
General disorders and administration site conditions, n (%)	**30 (26.1)**	**13 (11.9)**
Feeling drunk	9 (7.8)	1 (0.9)
Fatigue	5 (4.3)	6 (5.5)
Respiratory, thoracic and mediastinal disorders, n (%)	**24 (20.9)**	**15 (13.8)**
Throat irritation	9 (7.8)	5 (4.6)
Nasal discomfort	8 (7.0)	2 (1.8)
Eye disorders, n (%)	**18 (15.7)**	**3 (2.8)**
Vision blurred	14 (12.2)	3 (2.8)
Investigations, n (%)	**14 (12.2)**	**4 (3.7)**
Blood pressure increased	11 (9.6)	0
Follow-up phase, n (%)	**N=34**	**N=52**
Overall summary		
AE	9 (26.5)	12 (23.1)
AE possibly related to nasal spray drug[a]	0	1 (1.9)
AE possibly related to OAD[a]	1 (2.9)	3 (5.8)
AE leading to death	0	0
≥1 serious AE	1 (2.9)	0
AE leading to OAD being withdrawn[b]	0	0

Based on Tables 37 and 38 of the CS[1]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] An AE that started in the double-blind induction phase and resulted in discontinuation in the follow-up phase was counted as treatment-emergent in the double-blind induction phase; [redacted]

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

Table 4.18: Safety results of SUSTAIN-1 (overall)

	Induction phase	Optimisation phase	Maintenance phase		Follow-up phase	
	ESK-NS + OAD (N=437)	ESK-NS + OAD (N=455)	ESK-NS + OAD (N=152)	OAD + PBO-NS (N=145)	ESK-NS + OAD during any phase (N=481)	OAD + PBO-NS for all phases (N=64)
AE, n (%)	336 (76.9)	335 (73.6)	125 (82.2)	66 (45.5)	53 (11.0)	5 (7.8)
AE possibly related to nasal spray drug, n (%)[a]	301 (68.9)	281 (61.8)	106 (69.7)	37 (25.5)	7 (1.5)	0
AE possibly related to OAD, n (%)[a]	71 (16.2)	61 (13.4)	13 (8.6)	9 (6.2)	3 (0.6)	0
AE leading to death, n (%)	0	0	0	0	0	0
≥1 serious AE, n (%)	13 (3.0)	11 (2.4)	4 (2.6)	1 (0.7)	3 (0.6)	0
AE leading to nasal spray drug being withdrawn, n (%)	22 (5.0)	5 (1.1)	4 (2.6)	3 (2.1)	NA[b]	NA[b]
AE leading to OAD being withdrawn, n (%)[c]	8 (1.8)	2 (0.4)	3 (2.0)	0	0[c]	0[c]

Based on Table 39 of the CS[1]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] Patients did not receive nasal spray during the follow-up phase; [c] An AE that started in the induction phase and resulted in discontinuation in a subsequent phase was counted as treatment-emergent in the induction phase.

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

Table 4.19: Safety results of SUSTAIN-1 (AEs reported in ≥5% of patients)

	ESK-NS + OAD	OAD + PBO-NS
Induction phase (Safety [IND] analysis set)	**N=437**	**NA**
Total number of patients with an AE, n (%)	**336 (76.9)**	**-**
Nervous system disorders, n (%)	**248 (56.8)**	**-**
Dizziness	97 (22.2)	-
Dysgeusia	90 (20.6)	-
Somnolence	65 (14.9)	-
Headache	60 (13.7)	-
Paraesthesia	48 (11.0)	-
Sedation	44 (10.1)	-
Dizziness postural	33 (7.6)	-
Hypoaesthesia	30 (6.9)	-
Psychiatric disorders, n (%)	**163 (37.3)**	**-**
Dissociation	82 (18.8)	-
Anxiety	31 (7.1)	-
Gastrointestinal disorders, n (%)	**150 (34.3)**	**-**
Nausea	94 (21.5)	-
Hypoaesthesia oral	32 (7.3)	-
Vomiting	29 (6.6)	-
Ear and labyrinth disorders, n (%)	**108 (24.7)**	**-**
Vertigo	99 (22.7)	-
Respiratory, thoracic and mediastinal disorders, n (%)	**88 (20.1)**	**-**
Nasal discomfort	29 (6.6)	-
Throat irritation	26 (5.9)	-
Eye disorders, n (%)	**63 (14.4)**	**-**
Vision blurred	45 (10.3)	-
Investigations, n (%)	42 (9.6)	-
Blood pressure increased	34 (7.8)	-
Optimisation phase (Safety [OP] analysis set)	**N=455**	**NA**
Total number of patients with an AE, n (%)	**335 (73.6)**	**-**
Nervous system disorders, n (%)	**212 (46.6)**	**-**
Dysgeusia	79 (17.4)	-
Somnolence	63 (13.8)	-
Dizziness	61 (13.4)	-
Headache	57 (12.5)	-
Dizziness postural	26 (5.7)	
Hypoaesthesia	24 (5.3)	-
Paraesthesia	24 (5.3)	-

	ESK-NS + OAD	OAD + PBO-NS
Psychiatric disorders, n (%)	**136 (29.9)**	**-**
Dissociation	73 (16.0)	-
Gastrointestinal disorders, n (%)	**116 (25.5)**	**-**
Nausea	48 (10.5)	-
Hypoaesthesia oral	34 (7.5)	-
Ear and labyrinth disorders, n (%)	**101 (22.2)**	**-**
Vertigo	91 (20.0)	-
Respiratory, thoracic and mediastinal disorders, n (%)	**73 (16.0)**	**-**
Nasal discomfort	26 (5.7)	-
Investigations, n (%)	**47 (10.3)**	**-**
Blood pressure increased	26 (5.7)	-
Eye disorders, n (%)	46 (10.1)	-
Vision blurred	30 (6.6)	-
Maintenance phase (Safety [MA] analysis set)	**N=152**	**N=145**
Total number of patients with an AE, n (%)	**125 (82.2)**	**66 (45.5)**
Nervous system disorders, n (%)	**83 (54.6)**	**30 (20.7)**
Dysgeusia	41 (27.0)	10 (6.9)
Somnolence	32 (21.1)	3 (2.1)
Dizziness	31 (20.4)	7 (4.8)
Headache	27 (17.8)	14 (9.7)
Paraesthesia	11 (7.2)	0
Dizziness postural	10 (6.6)	3 (2.1)
Sedation	10 (6.6)	1 (0.7)
Hypoaesthesia	9 (5.9)	0
Psychiatric disorders, n (%)	**60 (39.5)**	**15 (10.3)**
Dissociation	35 (23.0)	0
Anxiety	12 (7.9)	5 (3.4)
Confusional state	9 (5.9)	0
Gastrointestinal disorders, n (%)	**53 (34.9)**	**11 (7.6)**
Nausea	25 (16.4)	1 (0.7)
Hypoaesthesia oral	20 (13.2)	0
Vomiting	10 (6.6)	1 (0.7)
Paraesthesia oral	8 (5.3)	1 (0.7)
Ear and labyrinth disorders, n (%)	**43 (28.3)**	**9 (6.2)**
Vertigo	38 (25.0)	8 (5.5)
Eye disorders, n (%)	**32 (21.1)**	**1 (0.7)**
Vision blurred	24 (15.8)	1 (0.7)
Diplopia	9 (5.9)	0

	ESK-NS + OAD	OAD + PBO-NS
Infections and infestations, n (%)	**32 (21.1)**	**25 (17.2)**
Viral upper respiratory tract infection	11 (7.2)	12 (8.3)
Respiratory, thoracic and mediastinal disorders, n (%)	**29 (19.1)**	**11 (7.6)**
Nasal discomfort	11 (7.2)	4 (2.8)
Throat irritation	8 (5.3)	1 (0.7)
Investigations, n (%)	**19 (12.5)**	**10 (6.9)**
Blood pressure increased	10 (6.6)	5 (3.4)

Based on Table 40 of the CS[1]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] Patients did not receive nasal spray during the follow-up phase; [c] An AE that started in the induction phase and resulted in discontinuation in a subsequent phase was counted as treatment-emergent in the induction phase.

AE = adverse event; CS = company submission; ESK = esketamine; IND = induction phase; MA = maintenance phase; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; OP = optimisation phase; PBO = placebo

Table 4.20: Safety results of TRANSFORM-3

	ESK-NS + OAD	OAD + PBO-NS
Induction phase, n (%)	**N=72**	**N=65**
Overall summary		
AE	51 (70.8)	39 (60.0)
AE possibly related to nasal spray drug[a]	42 (58.3)	22 (33.8)
AE possibly related to OAD[a]	13 (18.1)	11 (16.9)
AE leading to death	0	0
≥1 serious AE	3 (4.2)	2 (3.1)
AE leading to nasal spray drug being withdrawn[b]	4 (5.6)	2 (3.1)
AE leading to OAD being withdrawn[b]	1 (1.4)	1 (1.5)
AEs reported in ≥5% of patients		
Total number of patients with an AE, n (%)	51 (70.8)	39 (60.0)
Psychiatric disorders, n (%)	26 (36.1)	11 (16.9)
Dissociation	9 (12.5)	1 (1.5)
Dysphoria	4 (5.6)	0
Insomnia	4 (5.6)	3 (4.6)
Anxiety	2 (2.8)	5 (7.7)
Nervous system disorders, n (%)	24 (33.3)	16 (35.8)
Dizziness	15 (20.8)	5 (7.7)
Headache	9 (12.5)	2 (3.1)
Dysgeusia	4 (5.6)	3 (4.6)
Hypoaesthesia	4 (5.6)	1 (1.5)
Paraesthesia	4 (5.6)	2 (3.1)

	ESK-NS + OAD	OAD + PBO-NS
Gastrointestinal disorders, n (%)	19 (26.4)	8 (12.3)
Nausea	13 (18.1)	3 (4.6)
Hypoaesthesia oral	4 (5.6)	0
Vomiting	4 (5.6)	1 (1.5)
General disorders and administration site conditions, n (%)	14 (19.4)	8 (12.3)
Fatigue	9 (12.5)	5 (7.7)
Investigations, n (%)	14 (19.4)	6 (9.2)
Blood pressure increased	9 (12.5)	3 (4.6)
Ear and labyrinth disorders, n (%)	10 (13.9)	4 (6.2)
Vertigo	8 (11.1)	2 (3.1)
Infections and infestations, n (%)	8 (11.1)	6 (9.2)
Urinary tract infections	6 (8.3)	1 (1.5)
Follow-up phase, n (%)	**N=12**	**N=3**
Overall summary		
AE	1 (8.3)	1 (33.3)
AE possibly related to nasal spray drug[a]	0	1 (33.3)
AE possibly related to OAD[a]	1 (8.3)	0
AE leading to death	0	0
≥1 serious AE	0	0
AE leading to OAD being withdrawn[b]	0	0

Based on Tables 37 and 38 of response to request for clarification[3]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] An AE that started in the double-blind induction phase and resulted in discontinuation in the follow-up phase was counted as treatment-emergent in the double-blind induction phase.

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

Table 4.21: Safety results of SUSTAIN-2 (overall)

	ESK-NS + OAD
Induction phase, n (%)	**N=779**
AE	653 (83.8)
AE possibly related to nasal spray drug[a]	586 (75.2)
AE possibly related to OAD[a]	177 (22.7)
AE leading to death	0
≥1 serious AE	17 (2.2)
AE leading to nasal spray drug being withdrawn	53 (6.8)
AE leading to OAD being withdrawn	20 (2.6)
Optimisation/maintenance phase, n (%)	**N=603**
AE	516 (85.6)

	ESK-NS + OAD
AE possibly related to nasal spray drug[a]	402 (66.7)
AE possibly related to OAD[a]	110 (18.2)
AE leading to death	2 (0.3)
≥1 serious AE	38 (6.3)
AE leading to nasal spray drug being withdrawn[b]	23 (3.8)
AE leading to OAD being withdrawn[b]	14 (2.3)
Follow-up phase, n (%)	**N=357**
AE	55 (15.4)
AE possibly related to nasal spray drug[a]	9 (2.5)
AE possibly related to OAD[a]	5 (1.4)
AE leading to death	0
≥1 serious AE	8 (2.2)
AE leading to OAD being withdrawn[b]	1 (0.3)

Based on Table 47 of response to request for clarification[3]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] An AE that started in the previous phases and resulted in discontinuation in the follow-up phase was counted as treatment-emergent in the previous phase

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

Table 4.22: Safety results of SUSTAIN-2 (AEs reported in ≥5% of patients)

	ESK-NS + OAD (N=802)
Total number of patients with an AE, n (%)	**723 (90.1)**
Nervous system disorders, n (%)	**528 (65.8)**
Dizziness	264 (32.9)
Headache	200 (24.9)
Somnolence	134 (16.7)
Dysgeusia	95 (11.8)
Hypoaesthesia	95 (11.8)
Sedation	71 (8.9)
Dizziness postural	67 (8.4)
Paraesthesia	58 (7.2)
Psychiatric disorders, n (%)	**384 (47.9)**
Dissociation	221 (27.6)
Anxiety	72 (9.0)
Insomnia	63 (7.9)
Gastrointestinal disorders, n (%)	**373 (46.5)**
Nausea	201 (25.1)
Vomiting	87 (10.8)

	ESK-NS + OAD (N=802)
Hypoaesthesia oral	73 (9.1)
Diarrhoea	60 (7.5)
Infections and infestations, n (%)	**279 (34.8)**
Viral upper respiratory tract infection	82 (10.2)
Urinary tract infections	65 (8.1)
influenza	43 (5.4)
General disorders and administration site conditions, n (%)	**187 (23.3)**
Fatigue	63 (7.9)
Musculoskeletal and connective tissue disorders, n (%)	**154 (19.2)**
Back pain	41 (5.1)
Investigations, n (%)	**143 (17.8)**
Blood pressure increased	75 (9.4)
Ear and labyrinth disorders, n (%)	**126 (15.7)**
Vertigo	88 (11.0)
Eye disorders, n (%)	**105 (13.1)**
Vision blurred	60 (7.5)

Based on Table 48 of response to request for clarification[3]

Notes: 1) Incidence was based on the number of patients experiencing ≥1 AE, not the number of events; 2) AEs were coded using MedDRA version 20.0

[a] Study drug relationships of possible, probable, and very likely were included in this category; [b] An AE that started in the previous phases and resulted in discontinuation in the follow-up phase was counted as treatment-emergent in the previous phase.

AE = adverse event; CS = company submission; ESK = esketamine; MedDRA = Medical Dictionary for Regulatory Activities; NA = not applicable; NS = nasal spray; OAD = oral antidepressant; PBO = placebo

ERG comment: In the induction phase of TRANSFORM-2, more adverse events were observed in patients treated with ESK-NS + OAD compared to those receiving PBO-NS + OAD (85.2% vs. 60.6%, see Table 4.17). In SUSTAIN-1 more adverse events were seen in the maintenance phase (82.2% vs. 45.5%) and the follow-up phase (11.0% vs. 7.8%), see Table 4.18. Potential adverse events, especially psychiatric disorders (47.8% vs. 19.3%, see Table 4.17), need to be considered before considering ESK-NS as a treatment option for patients with TRD.

The company reported seven deaths among 1,861 patients treated with ESK-NS across the six phase 2 and 3 studies, three of which were completed suicides.[1] The company stated that, based on the severity of patients' underlying illness and the lack of a consistent pattern the suicides were considered unrelated to ESK-NS treatment. In this context it is important to note that the trials in the CS excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening.[1] The committee will need to consider if the evidence in the CS on effectiveness and safety of ESK-NS can be generalised to this vulnerable population.

The company was asked to provide any additional data pertaining to the development of addiction or addiction-related issues (e.g. withdrawal) during any of the identified studies.[18] In response, the company stated that '*across all clinical studies there were no cases of overdose or reports of drug*

abuse. Furthermore, there were no reports from the investigational sites of any patients engaging in drug-seeking behaviour or requesting an increase in the frequency of treatment sessions (as a potential early indicator of drug-seeking behaviour)'.[3] They further clarified that '*all Phase 3 studies included the PWC-20 to systematically assess the risk of dependence with short- and long-term use of esketamine nasal spray (…) Based on the PWC-20 results, there was no evidence suggestive of a distinct withdrawal syndrome in the longer-term studies* (…) *Levels of esketamine in the circulation do not accumulate with twice-weekly or lower dosing frequency. The steady state for physical dependence is not achieved therefore a drug withdrawal is not expected, as suggested by the PWC-20 results*'.[3]

While this appears reasonable, the company did note at clarification that Physicians Withdrawal Checklist- Withdrawal Symptoms- subscale (PWC-WS) results were higher in non-responders to ESK-NS. The ERG considers that it will be important to monitor these patients as they move to further treatments.

4.2.8 Supporting evidence

TRANSFORM-1 was regarded as a supporting trial in the CS and was not been included in the base case economic model. The company stated the rationale for this decision: '*In TRANSFORM-1, with the exception of the first dose (56 mg for all patients) ESK-NS was administered at fixed doses of either 56 mg or 84 mg which is not reflective of the anticipated esketamine licence*'.[1]

A total of 346 patients aged 18 to 64 years were randomised to treatment during the double-blind induction phase with either esketamine nasal spray 56 mg (fixed dose) plus a newly initiated OAD or esketamine nasal spray 84 mg (fixed dose) plus a newly initiated OAD or a newly initiated OAD plus placebo nasal spray. Of the 346 patients randomly assigned to treatment, 315 (91%) patients completed the 28-day double-blind induction phase, and 31 (9%) patients withdrew. There was a higher early withdrawal rate in the ESK-NS-84 + OAD arm (n=19; 16.4%) compared with the ESK-NS-56 + OAD (n=6; 5.1%) and OAD + PBO-NS (n=6; 5.3%) arms. Improvement in depressive symptoms, as assessed by the change in MADRS total score from baseline to Day 28 of induction numerically favoured the ESK-NS-56 + OAD and ESK-NS-84 + OAD arms over OAD + PBO-NS. However, these improvements did not reach statistical significance.

ERG comment: As the licence for ESK-NS is expected to be for flexible dosing, it is appropriate to treat TRANSFORM-1 as supporting evidence only. However, it is important to consider the implications of the higher withdrawal rate in the higher dosage group of ESK-NS which was mainly due to adverse events or patient choice. The company stated that withdrawals in the ESK-NS-84 + OAD arm were not due to any new or dose-related safety finding, and that 11 of the 19 early withdrawal patients (58%) withdrew after their first esketamine nasal spray dose which was 56 mg as stipulated by the fixed titration study design. The withdrawal rate could explain the lack of statistically significant results, but the ERG remains concerned that TRANSFORM-1 does not provide convincing evidence of the efficacy or safety of ESK-NS.

4.2.9 Ongoing trials

The CS included details of a long-term non-comparative safety study of ESK-NS which is ongoing (SUSTAIN-3). The study population includes those who have previously participated in completed or ongoing trials, including TRANSFORM-1/2/3 and SUSTAIN-1/2. The company provided interim safety results from a cut-off of 31 December 2018 which included data from 1,140 patients treated for a mean of 13.7 months.[26] They stated that '*the interim analysis has revealed no unexpected safety findings, with a safety and tolerability profile that is consistent with the previous Phase 3 clinical*

studies'.[1] SUSTAIN-3 is expected to complete in the third quarter of 2021, when final safety and efficacy data will be available.

ERG comment: The ERG noted that there were three deaths in this SUSTAIN-3 (0.3%). These were detailed in the interim CSR as follows: [redacted] This study, when reported in full, will give a fuller picture of any potential longer-term risks with ESK-NS including those related to withdrawing from treatment.

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

The company conducted a Bayesian NMA to assess the relative effectiveness of ESK-NS plus a newly initiated OAD versus the comparators in the NICE scope. Feasibility assessment of the studies identified in the systematic review identified that an NMA could only be conducted for the acute phase of treatment. However, the company considered the NMA of acute treatment comparisons not to be robust and it was only used to inform scenario analyses in the analysis of cost effectiveness.[1]

Nineteen trials were used to inform the network. The outcomes investigated were change from baseline in MADRS total score, response rates based on MADRS, remission rates based on MADRS and discontinuations due to adverse events. The company stated that the NMA was not considered sufficiently robust to inform the CEA so no quality assessment of the trials was performed. The trials used in the NMA are listed in Table 4.23.

Table 4.23: Overview of the 19 trials included in the best-case scenario evidence network

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
ADMIRE[27] Double-blind RCT Multicentre, Japan NCT00876343.	• Adults aged 20–65 years • DSM-5 diagnosis of MDD • HAM-D-17[a] ≥18 • Duration of current episode ≥8 weeks without adequate response to 1-3 OADs of ≥6 weeks duration Patients received an SSRI/SNRI during an 8-week single blind prospective treatment phase and those with an inadequate response were randomised	Augmentation SSRI/SNRI Aripiprazole 3-5 mg/day (flexible dose)	197	38.1 (9.6)	101 (52.1)	25.3 (7.3)	17.5 (26.1) months	1; 119 (61.3%) 2; 54 (27.8%) 3; 21 (10.8%) 4+; 0 (0%)	6 (plus 28-day screening phase and 8-week prospective treatment phase)
		Augmentation SSRI/SNRI Aripiprazole 3 mg/day	194	39.2 (9.1)	124 (62.9)	25.2 (7.2)	15.7 (21.6) months	1; 130 (66.0%) 2; 53 (26.9%) 3; 14 (7.1%) 4+; 0 (0%)	
		Augmentation SSRI/SNRI Placebo	195	38.7 (9.2)	115 (59.0)	25.5 (7.4)	15.6 (16.4) months	1; 124 (63.6%) 2; 49 (25.1%) 3; 22 (11.3%) 4+; 0 (0%)	
Bauer 2013[28] Open-label, RCT Multicentre, international	• Adults aged 18–65 years • DSM-5 diagnosis of MDD • Duration of current episode ≥42 days and ≤18 months	Augmentation SSRI/SNRI + quetiapine XR (target dose 300 mg/day)	229	NR	NR	33.2 (5.34)	190.7 (119.3) days	NR	6

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
NCT00789854	• MADRS ≥25 • Stage I TRD with an inadequate response to an SSRI/ venlafaxine or stage II TRD with an inadequate response to two ADs from two different classes-most recent of which must have been an SSRI or venlafaxine	Augmentation SSRI/SNRI + lithium (target plasma level 0.6-1.2 mmol/l)	221	NR	NR	32.9 (5.20)	180.3 (119.6 days)	NR	
		Switch quetiapine XR (target dose 300 mg/day)	225	NR	NR	33.70 (5.60)	175.2 (110.8) days	NR	
Berman 2007[29] Double-blind RCT Multicentre, USA	• Adults aged 18–65 years • DSM-4 diagnosis for major depressive episode that had lasted ≥8 weeks with an inadequate response 1-3 OAD trials (>6 weeks duration) • HAM-D-17[a] ≥18 All patients received SSRI/SNRI for 8 weeks in an open label prospective treatment phase; those with an incomplete response were eligible for randomisation	Augmentation SSRI/SNRI Placebo	176	44.2 (10.9)	63 (35.8)	25.9 (6.5)	43.6 (53.8) months	1; 117 (66.5%) 2; 45 (25.6%) 3; 18 (8.0%)	6 (plus 8-week prospective treatment phase)
		Augmentation SSRI/SNRI Aripiprazole 5-20 mg/day	182	46.5 (10.6)	70 (38.5)	26.0 (6.1)	38.6 (59.0) months	1; 121 (66.5%) 2; 45 (24.7%) 3; 16 (8.8%)	
Berman 2009[30] Double-blind RCT Multicentre, USA	• Adults aged 18–65 years • DSM-4 diagnosis for major depressive episode that had lasted ≥8 weeks with	Augmentation SSRI/SNRI Placebo	172	45.6 (11.3)	55 (32.0)	27.1 (5.8)	Median 17.2 (1.6-236.5) months	0; 2 (2.9%) 1; 117 (68%)	6 (plus 8-week prospective

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
	an inadequate response 1-3 OAD trials (>6 weeks duration) • HAM-D-17[a] ≥18 All patients received SSRI/SNRI + placebo for 8 weeks in a single-blind label prospective treatment phase; those with an inadequate response were eligible for randomisation							2; 45 (26.2%) 3; 3 (1.7%) 4; 2 (1.2%)	treatment phase)
		Augmentation SSRI/SNRI Aripiprazole 5-20 mg/day	177	45.1 (10.6)	39 (22.0)	26.6 (5.8)	Median 18.8 (2.1-433.1) months	0; 3 (21.7%) 1; 127 (71.8%) 2; 38 (21.5%) 3; 9 (5.1%) 4; 0	
Corya 2006[31] Double-blind RCT Multicentre, 16 countries	• Adults ≥18 years • DSM-5 diagnosis of MDD, single episode or recurrent, without psychotic features • Nonresponse to of ≥6 weeks SSRI Patients received venlafaxine in an open-label 7-week lead-in phase; those displacing less than a partial response entered the double-blind taper phase and then proceeded to the 12-week double-blind phase	Switch fluoxetine 25/50 mg/day Olanzapine 6/12 mg/day	243	45.7 (10.8)	(27.5)	30 (6.8)	186 days	Mean 4.1	12-week acute phase (plus 7-week lead-in phase and 5–9-day taper phase)
		Switch olanzapine 6 or 12 mg/day	62						
		Switch fluoxetine 25 or 50mg/day	60						
		Augmentation venlafaxine	59						

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
		75-375 mg/day							
		Switch fluoxetine 5 mg/day Olanzapine 1 mg/day [arm serves as a pseudo placebo]	59						
Dunner 2007[32] Open-label RCT	• Adults aged 21–65 years • Nonresponse to ≥1 course of ≥4 weeks SSRI/SNRI • MADRS ≥20 Patients were assigned a prospective open-label 6-week lead-in treatment with sertraline; those failing to respond were eligible for randomisation	Augmentation sertraline 50 mg/day-200 mg/day	20	46.3 (10.4)	(45)	30.7 (5.4)	NR	2≥ SSRI/SNRI; 65 %	8 (plus 6-week lead-in period)
		Augmentation Sertraline 50 mg/day-200 mg/day Ziprasidone 80 mg/day	22	43.1 (9.4)	(45.5)	30.2 (5.7)	NR	2≥ SSRI/SNRI; 63.6 %	
		Augmentation Sertraline, 50 mg/day-200 mg/day Ziprasidone 1600 mg/day	19	42.6 (13.3)	(52.6)	28.9 (5.4)	NR	2≥ SSRI/SNRI; 63.2 %	

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
Luzny 2013[33] Open-label RCT Single centre, Czech Republic Abstract publication	• Adults aged 65+ years • Fulfil diagnostic criteria for MDD and failing prior OAD with two different OADs in monotherapy	ECT BW up to 8 electro convulsions	8	67.3 (3.9)	7	NR	NR	2; 8 (100%)	6
		Seropram (Citalopram) 20-40 mg/day	12	68.2 (4.1)	8	NR	NR	2; 12 (100%)	
Lenze 2015[34] Double-blind RCT 3 centres, USA & Canada NCT00892047	• Adults aged ≥60 years • DSM-5 diagnosis of MDD with at least moderate symptoms • MADRS ≥15 • Although prior treatment failure not explicitly stated 74% of patients were reported to have not responded to ≥1 OAD trialled during the present episode Patients were assigned to a 12-week prospective open label venlafaxine extended release-patients who did not achieve remission were eligible for randomisation	Augmentation venlafaxine Aripiprazole	91	Median 66 (IQR: 62.8, 70.5)	39 (43)	Median 24 (IQR: 18, 29)	Median 118 (IQR: 45, 364)	≥1, 73%	12 (plus 12-week prospective treatment phase)
		Augmentation venlafaxine Placebo	90	Median 65.7 (IQR: 62.8, 69.8)	39 (43)	Median 23 (IQR: 18, 26)	Median 104 (IQR: 28, 317)	≥1, 75%	
Marcus 2008[35] Double-blind RCT Multicentre, USA	• Adults aged 18–65 years • DSM-4 diagnosis of major depressive episode that lasted ≥8 weeks • Inadequate response to previous OAD (1-3 OAD trials of >6 weeks duration) Patients were assigned to an 8-week prospective single-blind treatment phase	Augmentation SSRI/SNRI Aripiprazole 5-20 mg/day	191	44.6 (11.0)	65 (34)	25.2 (6.2)	43.7 (68.0) months	1; 135 (71.1%) 2; 49 (25.8%) 3; 5 (2.6%) 4; 1 (0.5%)	6 (plus 8-week prospective treatment phase)
		Augmentation SSRI/SNRI	190	44.4 (10.7)	62 (32.6)	27.0 (5.5)	48.5 (88.8) months	1; 128 (67.7%)	

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
	of SSRI/SNRI-patients who did not response were eligible for randomisation	Placebo						2; 51 (27.0%) 3; 10 (5.3%)	
Nierenberg 2003[36] Double-blind RCT Single centre, UK	• Adults aged 18–70 years • DSM-3 diagnosis of MDD • HAMD-D-17[a] ≥18 • Treatment-resistant depression defined as at least 1 but no more than 5 failed medication trials during the current episode Patients were assigned to a 6-week prospective open-label treatment phase of nortriptyline. Non-responders were eligible for randomisation	Augmentation nortriptyline Lithium	18	37.2 (8.3)	9	NR	97.3 months (111.8)	Mean failed trials during current episode 1.9 (SD 1.2)	6 (plus 6-week prospective treatment phase)
		Augmentation nortriptyline Placebo	17	39.7 911.9)	10	NR	84.5 months (94.9)	Mean during current episode 2.5 (SD 1.6)	
OPERATION[37] Double-blind RCT Multicentre, China	• Adults aged 18–65 years • MDD • Stage 2 TRD criteria described by Thase and Rush. • HRSD-17[a] ≥17	Switch venlafaxine XR 225 mg/day	50	40.5 (11.5)	NR	NR	4.7 (4.6) years	NR	8
		Switch mirtazapine, 45 mg/day	55		NR	NR	5.5 (6.6) years	NR	
		Switch paroxetine	45		NR	NR	7.5 (6.5) years	NR	
POLARIS[38] Double blind, phase III RCT	• Adults aged 18–65 years	Augmentation SSRI/SNRI Placebo	221	46.6 (11.0)	75 (33.9)	26.3 (5.3)	16.9 (35.0) months	1; 170 (78%) 2; 44 (20.2%)	6 (plus 8-week prospective

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
Multicentre, international NCT01360632	• DSM-4 diagnosis of MDD, single episode or recurrent, without psychotic features of ≥8 weeks duration • Reporting an inadequate response to 1-3 OADs including the most recent drug treatment • HADRS-17[a] ≥18 Patients were assigned to an 8-week prospective single-blind placebo as an adjunctive to standard OAD (SSRI/ SNRI)-patents with an inadequate response were eligible for randomisation							3; 4 (1.8%)	treatment phase)
		Augmentation SSRI/SNRI Brexpiprazole 1 mg/day	226	45.7 (11.6)	68 (30.1)	26.7 (5.6)	18.7 (43.0) months	1; 177 (78.7%) 2; 42 (18.7%) 3; 6 (2.7%)	
		Augmentation SSRI/SNRI Brexpiprazole 3 mg/day	230	44.5 (11.2)	74 (32.2)	26.4 (5.2)	17.7 (33.0) months	1; 184 (81.4%) 2; 34 (15.0%) 3; 7 (3.1%)	
PYXIS[39] Double blind, phase III RCT Multicentre, USA, Canada and Europe NCT01360645	• Adults aged 18–65 years • DSM-5 diagnosis of MDD, single episode or recurrent, without psychotic features of ≥8 weeks duration • Reporting an inadequate response of 1-3 OADs including the most recent drug treatment • HADRS-17[a] ≥18 Patients were assigned to an 8-week prospective single-blind placebo as an adjunctive to standard OAD (SSRI/SNRI)-patents with an inadequate response were eligible for randomisation	Augmentation SSRI/SNRI Placebo	191	45.2 (11.3)	52 (28.3)	27.1 (5.6)	13.7 (17.1) months	NR	6 (plus 8-week prospective treatment phase)
		Augmentation SSRI/SNRI Brexpiprazole 2 mg/day	188	44.1 (11.6)	25 (30.9)	26.6 (5.8)	13.5 (14.2) months	NR	
Shelton 2005[40] Double blind RCT	• DSM-5 diagnosis of MDD	Switch fluoxetine 25 - 50 mg/day	146	42.5 (10.7)	(32.9)	28.5 (7.5)	NR	NR	8 (plus 7-week dose-

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
Multicentre, USA and Canada	• ≥1 past treatment failure with 4 weeks of SSRI Patients entered a 7-week nortriptyline dose-escalation period to demonstrate treatment failure for eligibility for randomisation	Olanzapine 6-12 mg/day							escalation period)
		Switch olanzapine 6-12 mg/day	144	43.4 (11.0)	(35.4)	28.4 (7.4)	NR	NR	
		Switch fluoxetine 25 - 50 mg/day	142	41.7 (11.0)	(27.5)	28.4 (7.3)	NR	NR	
		Augmentation nortriptyline up to 175 mg/day	68	41.5 (10.1)	(32.4)	28.8 (6.5)	NR	NR	
STAR*D (step 3b)[41] Open-label RCT	• Eligible participants for third-step treatment entered Level 3 if they had not achieved remission or were unable to tolerate Level 2 or Level 2A treatments • Patients were not required to meet MDD criteria at the time of entry into Level 3, as long as they had MDD criteria at entry into Level 1 and had not adequately responded or been able to tolerate previous levels	Switch mirtazapine 15-60 mg/day	114	44.8 (11.6)	66 (57.9)		34.8 (70.4) months	NR	16
		Switch nortriptyline up 25-150 mg/day	121	45.1 (12.2)	59 (48.8)		32.5 (59.6) months	NR	
STAR*D (step 4)[42] Open-label RCT	• Eligible participants for fourth-step treatment entered Level 4 if they had not achieved remission or were unable	Switch tranylcypromine 10-60 mg/day	58	46.6 911.6)	25 (43.1)	NR	33.1 (67.9) months	NR	14

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
	to tolerate the first-three levels of treatment	Switch venlafaxine 37.5-300 mg Mirtazapine 15-45 mg/day	51	45.3 (10.6)	28 (54.9)	NR	55.7 (92.2) months	NR	
Tanghe 1997[43] Double-blind RCT Single centre	• Hospitalised patients with therapy resistant depression • DSM-2 R criteria for MDD • Resistant to ≥2 separate OADs	Switch moclobemide 200-600 mg/day	19	43 (12)	13	41 (7)	NR	NR	4
		Switch amitriptyline up to 280 mg/day	29				NR	NR	
		Switch moclobemide 200-600 mg/day Switch amitriptyline up to 280 mg/day	20				NR	NR	
Thase 2007[44] Double-blind RCT Canada and USA	• Adults aged 18–65 years • HAM-D-17[a] ≥22 • DSM-5 diagnosis of MDD • Failure to achieve a response to an OAD (except fluoxetine) after ≥6	Augmentation fluoxetine 50 mg/day Olanzapine 6 mg/day	200	44.3 (10.2)	68 (34)	30.1 (6.7)	415.4 (555) days		6 (8-week lead-in period)

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
2 concurrent identical studies (pooled results extracted) NCT00035321	weeks within the current episode of MDD Patients received fluoxetine in an 8-week open label lead-in phase to establish fluoxetine resistance	Augmentation fluoxetine 50 mg/day	206	44.6 (10.0)	78 (37.9)	29.9 (6.4)	428.6 (603.3) days		
		Switch olanzapine 6 mg/day	199	44.3 (10.8)	76 (38.2)	29.9 (6.7)	366.5 (544.4) days		
TRANSFORM-2[23, 45, 46] Double-blind Phase III RCT Multi-centre Europe and USA	• Adults aged 18–64 years • DSM-5 • MDD with no response to ≥1 but ≤5 in current episode The prospective observational phase patients take a different OAD for ≥2 weeks-non-responders eligible for randomisation	Switch SSRI (escitalopram or sertraline) or SNRI (duloxetine or venlafaxine XR) according to local prescribing guidelines (open label) Esketamine nasal spray 56 mg or 84 mg BW for 4 weeks	116	44.9 (12.58)	39 (34.2)	37.0 (5.69)	111.4 (124.28)	1; 9 (7.9%) 2; 69 (60.5%) 3; 24 (21.1%) 4; 7 (6.1%) 5; 3 (2.6%) 6; 1 (0.9%) 9; 1 (0.9%)	4
		Switch SSRI (escitalopram or sertraline) or SNRI (duloxetine or venlafaxine	111	46.4 (11.14)	46 (42.2)	37.3 (5.66)	118.0 (187.37)	1; 18 (16.5%) 2; 54 (49.5%) 3; 22 (20.2%)	

Trial	Inclusion criteria and study design prior to randomisation	Randomised interventions (acute phase)	N	Mean age (SD)	Male, n (%)	Mean MADRS total score (SD)	Duration of current episode, mean (SD)	Previous OAD use, n (%)	Duration of trial, weeks
		XR) according to local prescribing guidelines (open label) Placebo nasal spray BW for 4 weeks						4; 13 (11.9%) 5; 1 (0.9%) 6; 1 (0.9%)	

Based on Table 10 of the CS appendices[17]

[a] HAM-D may also be referred to as HAM-D-17, HRSD, HADRS-17, and HSRD in the literature. Predecessor versions of the HAM-D contained only 17 items.

AD = antidepressant; BW = bi-weekly; CS = company submission; DSM-2/3/4/5 = Diagnostic and Statistical Manual of Mental Disorders – 2/3/4/5; ECT = electroconvulsive therapy; HAM-D-17, Hamilton Depression Rating Scale 17-item; IQR = interquartile range; MADRS = Montgomery-Åsberg Depression Rating Scale; MDD = major depressive disorder; NR = not reported; OAD = oral antidepressant; RCT = randomised controlled trial; SD = standard deviation; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TRD = treatment-resistant depression; UK = United Kingdom; USA = United States of America; XR = extended release

4.4 Critique of the indirect comparison and/or multiple treatment comparison

In the acute treatment NMA, the company included comparator therapies based on switch or augmented treatments (i.e. where patients were randomised to switch or continue with their current OAD, respectively, with or without an additional OAD). The ERG felt that studies where patients received multiple OADs were outside the scope (patients receiving esketamine in the background of a single OAD), and should therefore not be included in the network.

The NMA assumed comparability between SSRIs and SNRIs, which the company indicated was supported by subgroup analyses in TRANSFORM-2 (Appendix E and Table 3 of the clarification letter) and NICE guidance[6, 9]; this is also in line with the proposed changes to the CHMP marketing authorisation, which were stated by the company to be, *"SPRAVATO®, in combination with an SSRI or SNRI, is indicated for adults with treatment-resistant major depressive disorder, who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode"*.[3] However, as noted in section 4.2.6, there are differences between the type of OAD for remission rates after 28 days, e.g. within the SSRI group: sertraline (odds ratio (OR) 1.38, 95% CI 0.26 to 7.22) vs. escitalopram (OR 4.71, 95% CI 1.08 to 20.63).

The NMA was based on a best-case scenario evidence network. Out of the 49 citations (42 trials) identified by the acute treatment SR, 19 studies were included. The remaining 23 trials were excluded due to lack of relevant outcomes or comparators, dose issues (specifically for esketamine) and being unable to be connected in the network. Of note, the SUSTAIN-1 trial, which was included in the economic analysis, was not included in the NMA which was appropriate as patients from TRANSFORM-1 and -2 could enter SUSTAIN-1 so they would not be independent trial population. TRANSFORM-3, which was not included in the economic analysis due to age and dose restrictions, was also excluded from the NMA.

Full details of the NMA methodology including the feasibility assessment, included trials and the assessment of their clinical similarity were provided in Appendix D of the company submission. NMA could be performed for acute treatments for the following outcomes: change from baseline in MADRS, MADRS response, MADRS remission and discontinuations due to AE. The NMA used standard Bayesian models as recommended in NICE DSU TSD 2.[47] WinBUGS code and some data were provided in the response to clarification but not for all the reported analyses. Change from baseline in MADRS for the base-case, response for scenario 2 and remission for scenario 1 were provided. The ERG could run the NMA and obtained results which were very close to those provided by the company so they have no concerns about the NMA analysis methods.

The main concerns about the NMA results are due to the clinical and methodological differences between the studies included in each network. This was highlighted in the submission *"clinical trial heterogeneity in terms of overall study design, inclusion criteria and patient population meant treatment comparisons could not be undertaken (in either acute or maintenance treatment settings"*.[1] However, they still performed and presented results for an acute treatment NMA but in order to perform this analysis had to relax the inclusion criteria. Relaxing the inclusion criteria by including MADRS results from more variable timepoints (four to eight weeks rather than just four weeks) increased the clinical heterogeneity of the NMA making the results less reliable as the submission states that data suggest that relative treatment effects change over time after four weeks. The submission stated that the MADRS and HAM-D scales were combined in the NMA although the clarification response indicated that this had not been done in any of the NMA. The company also reports that there were differences in the comparator arms regarding whether they were switch SSRI/SNRI or switch SSRI. Based on these

differences the ERG agrees with the company that there are considerable uncertainties in the NMA and the results should be interpreted cautiously.

A further issue of concern for the NMA is the use of the adjusted OAD + placebo arm in TRANSFORM-2 which was adjusted to account for the effect of additional clinic visits. This used results from a paper by Posternak and Zimmerman which found that additional visits increased the treatment effect for patients on placebo, and estimated the size of the reduction in HAM-D score with additional follow-up assessments.[48] The high placebo effect seen in TRANSFORM-2 was considered by clinicians to be related to the use of a nasal-spray treatment and the increased level of healthcare contact during the twice-weekly clinical visits. However, although Posternak and Zimmerman state that it was a meta-analysis, it does not report any details of the statistical methods used nor any details of the methods or results of the individual studies so it is not possible to verify whether the reported reductions in HAM-D scores were reliable. The study by Posternak and Zimmerman used HAM-D whereas the trials presented in the CS used MADRS score as the primary outcome measure so estimates of improvements in HAM-D were converted to MADRS scores using a method reported by Leucht et al.[49] The numbers applied in the adjustment were therefore based on two sets of estimates from single studies, one of which did not report any statistical methods and therefore may be unreliable.

Although the adjustment was made to the treatment effect observed for the placebo + OAD arm this was a double-blind, randomised trial so the effects of the use of a nasal-spray treatment and the increased number of visits also applied to the esketamine + OAD arm. The paper by Posternak and Zimmerman also analysed the effect of additional visits in the active treatment arm and found a similar reduction in HAM-D with one extra visit (0.76 for active treatment vs. 0.86 for placebo) and concluded that *"a comparable therapeutic effect was also found in participants receiving active medication"*.[48] Any improvements in MADRS as a result of increased clinic visits would apply to both treatment arms in the trial so the post-hoc adjustment should have been made to both the esketamine and placebo arms. Due to concerns with both, NMA and the adjusted TRANSFORM-2 results, the ERG does not consider them to be reliable sources of treatment estimates.

4.5 Additional work on clinical effectiveness undertaken by the ERG

No additional work on clinical effectiveness was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

The CS included a systematic review of the evidence for ESK-NS. From this review the company identified and presented evidence from six studies of ESK-NS. Four of these were randomised controlled trials (TRANSFORM-1, TRANSFORM-2, TRANSFORM-3, SUSTAIN-1) and two were open label extension studies (SUSTAIN-2, SUSTAIN-3). SUSTAIN-3 is still ongoing.

Randomised evidence is thus available for both the acute treatment of treatment-resistant depression and for maintenance of effect after remission. The two main trials included in the economic model (TRANSFORM-2 and SUSTAIN-1) were in adults aged 18 to 64 years with recurrent or single episode depression. Both trials compared ESK-NS plus a newly initiated OAD to a newly initiated OAD plus placebo and both involved flexible dosing of 56 mg/84 mg. A separate trial of those aged 65 years and over with lower dosing (TRANSFORM-3) and an open-label trial in adults aged 18 years or over (SUSTAIN-2) were included in the CS but not in the initial model. A further trial, TRANSFORM-1, was regarded as a supporting trial in the CS and was not included in the base case economic model due to its fixed rather than flexible dosing which does not reflect the expected licence for ESK-NS.

In response to clarification, the company advised that the label indication is expected to change to ESK-NS in combination with an SSRI or SNRI for treatment-resistant major depressive disorder in adults who have not responded to at least two different treatments with antidepressants in the current moderate to severe depressive episode.[3] This reflects the trials where patients received either a SNRI or SSRI in conjunction with ESK-NS. Most patients in the trials had received two prior OADs in this episode (61% and 57.7% in TRANSFORM-2 and SUSTAIN-1, respectively). The committee will need to consider how well the OADs prescribed as co-interventions across these trials reflect those prescribed at this stage of the pathway in an NHS setting.

The trials were multinational. However, TRANSFORM-2 and SUSTAIN-1 did not enrol any patients in the UK. One UK patient was enrolled in the supporting trial, TRANSFORM-3, and 12 UK patients were enrolled in the long-term safety study, SUSTAIN-2. The lack of UK patients in the main trials is a limitation particularly given the mode of delivery of this intervention. Therefore, there is a lack of evidence in how well ESK-NS might work in the NHS setting.

In TRANSFORM-2, ESK-NS + OAD in comparison to PBO-NS + OAD showed a statistically significant reduction on the Montgomery-Åsberg Depression Rating Scale (MADRS) at Day 28 (difference in least squares means -4.0, 95% confidence interval (CI) -7.31 to -0.64). Of note, there are differences between the type of OAD for remission rates after 28 days, e.g. within the SSRI group: sertraline (odds ratio (OR) 1.38, 95% CI 0.26 to 7.22) vs. escitalopram (OR 4.71, 95% CI 1.08 to 20.63). The trial also showed differences in response rate and remission rate, respectively, between the two groups. Other reported outcomes were in favour of the intervention (see Table 4.12).

In SUSTAIN-1, the percentage of relapse was lower in the ESK-NS + OAD (stable remitters: 26.7%, stable responders: 25.8%) group in comparison to participants receiving PBO-NS + OAD (45.3% and 57.6%, respectively). The trial also showed time to relapse to be in favour of the intervention group for both, stable remitters (hazard ratio (HR) 0.49, 95% CI 0.29 to 0.84) and stable responders (HR 0.30, 95% CI 0.16 to 0.55). Other reported outcomes were in favour of the intervention (see Table 4.13).

In the induction phase of TRANSFORM-2, more adverse events were observed in patients treated with ESK-NS + OAD compared to those receiving PBO-NS + OAD (85.2% vs. 60.6%, see Table 4.17). In SUSTAIN-1 more adverse events were seen in the maintenance phase (82.2% vs. 45.5%) and the follow-up phase (11.0% vs. 7.8%), see Table 4.18. Potential adverse events, especially psychiatric disorders (47.8% vs. 19.3% in TRANSFORM-2), need to be considered before considering ESK-NS as a treatment option for patients with TRD.

A number of other restrictions in inclusion criteria limit the generalisability of the trials to NHS practice. The trials in the CS excluded patients with moderate/severe alcohol abuse according to DSM-5 criteria. The committee will need to consider whether evidence in the CS on effectiveness and safety of ESK-NS can be generalised to those with a dual diagnosis of depression and alcohol misuse. The trials also excluded patients who had not responded to an adequate course of treatment with ECT in the current major depressive episode. This appears to be in line with the proposed pathway for ESK-NS. The committee will need to consider if ESK-NS is likely to be offered to patients who have not responded to ECT.

It is not clear if ESK-NS can reduce incidences of suicidal behaviour or if conversely there may be greater risk of suicide. The company reported seven deaths among 1,861 patients treated with ESK-NS across the six phase 2 and 3 studies, three of which were completed suicides.[1] The company stated that, based on the severity of patients' underlying illness and the lack of a consistent pattern the suicides were considered unrelated to ESK-NS treatment. In this context it is important to note that the trials in

the CS excluded patients who had suicidal/homicidal ideation/intent within six months prior to screening per the investigator's clinical judgements and/or based on C-SSRS or a history of suicidal behaviour in the 12 months prior to screening.[1] The committee will need to consider if the evidence in the CS on effectiveness and safety of ESK-NS can be generalised to this vulnerable population.

The company stated that there were no cases of overdose or reports of drug abuse across all the clinical studies. However they did note at clarification that measures of withdrawal according to PWC-WS were higher in non-responders to ESK-NS.[3] The ERG considers that it will be important to monitor these patients as they move to further treatments.

SUSTAIN-3, when reported in full, should gve a fuller picture of any potential longer-term risks with ESK-NS including those related to withdrawing from treatment.

5. Cost effectiveness

5.1 ERG comment on company's review of cost effectiveness evidence

5.1.1 Searches performed for cost effectiveness section

Appendices G, H and I of the CS detailed systematic searches of the literature used to identify cost effectiveness (appendix G), HRQoL (appendix H) and cost and healthcare resource identification, measurement and valuation studies (appendix I).[17] The same search was reported for both resource use in appendix I and cost effectiveness in appendix G, therefore the same limitations will apply. Searches were undertaken in July 2018. A summary of the sources searched is provided in Tables 5.1 and 5.2. Reference lists of included studies were checked for further relevant studies.

Table 5.1: Data sources for published cost effectiveness studies and cost and healthcare resource identification, measurement and valuation (Appendices G and I)

Resource	Host/Source	Date Range	Date Searched
Electronic databases			
Medline, Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Medline Daily	OVID	1946-Present	4/7/18 (updated 4/4/19)
Embase		1974-2018/07/03	4/7/18 (updated 4/4/19)
HTA Database	EBM Reviews via OVID	Up to 4th Quarter 2016	4/7/18
NHS EED		Up to 1st Quarter 2016	
Econlit	OVID	1886-2018/6/21	4/7/18 (updated 4/4/19)
PsycINFO	OVID	1987-2018/07/wk1	4/7/18 (updated 4/4/19)
Conference proceedings[a]			
Anxiety and Depression Association of America Conference		2016-2019	31/10/18 (updated 24/5/19) Unable to access abstracts
International Conference on Management of Depression		2016-2019	31/10/18 (updated 24/5/19) Unable to access abstracts
American Psychiatry Association Annual Meeting		2016-2019	1/11/18 (updated 23/5/19)

Resource	Host/Source	Date Range	Date Searched
European Congress of Psychiatry		2016-2019	5-6/11/18 (updated 23/5/19)
The Royal College of Psychiatrists International Congress		2016-2019	6/11/18 (updated 24/5/19) Unable to access abstracts
WPA World Congress of Psychiatry		2016-2019	6/11/18 (updated 23/5/19) Unable to access abstracts for 2017-19
ISPOR (USA/Europe)		2016-2019	23/5/19
HTA agencies[a]			
NICE, SMC, PBAC, CADTH, NCPE			30/8/18 (updated 23/5/19)
Additional resources (cost effectiveness)[a]			
CEA Registry, RePEc, INAHTA, NIHR HTA database, ICER, Google Scholar, EuroQoL website, ScHARRHUD database			31.8.18 (updated 24/5/19)

[a] Where appropriate, searches were also used to inform both HRQoL (Appendix H) and cost and healthcare resource identification, measurement and validation (Appendix I)
CADTH = Canadian Agency for Drugs and Technologies in Health; CEA = cost effectiveness analysis; EBM = evidence-based medicine; EED = Economic Evaluation Database; HTA = Health Technology Assessment; INAHTA = International Network of Agencies for Health Technology Assessment; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NCPE = National Centre for Pharmacoeconomics; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; NIHR = National Institute for Health Research; PBAC = Pharmaceutical Benefits Advisory Committee; RePEc = Research Papers in Economics; SMC = Scottish Medicine Consortium; USA = United States of America; WPA = World Psychiatric Association

ERG comment:

- The majority of searches were clearly structured and documented. Missing data regarding the supplementary searches were provided at clarification.[3]
- There were limitations with the use of MeSH (Medical subject headings) indexing terms in the Embase searches. Although some automated mapping between indexing terms does take place it is possible that relevant Emtree indexing terms were not included in the search, and potentially relevant records could have been missed.

Table 5.2: Data sources for health-related quality of life studies (Appendix H)

Resource	Host/Source	Date Range	Date Searched
Electronic databases			
Medline, Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Medline Daily	OVID	2016-Present	5/7/18 (updated 4/4/19)
Embase		2016-2018/07/03	5/7/18 (updated 4/4/19)
HTA Database	EBM Reviews via OVID	2016- 2016 4th Quarter 2016	5/7/18 (updated 4/4/19)
NHS EED		2016- 2016 1st Quarter 2016	
CENTRAL		2016-2018/06	
CDSR		2016-2018/06/28	
DARE		2016- 2016 1st Quarter 2016	
PsycINFO	OVID	2016-2018/07/wk1	5/7/18 (updated 4/4/19)
CDSR = Cochrane Database Systematic Reviews; DARE = Database of Abstracts of Reviews of Effects; EBM = evidence-based medicine; EED = Economic Evaluation Database; HTA = Health Technology Assessment; NHS = National Health Service			

ERG comment:

- The majority of searches were clearly structured and documented. Missing data regarding the supplementary searches were provided at clarification.[3]
- All searches for health-related quality of life studies were limited to papers published after 2016, these searches were intended to identify any evidence published since the draft update of NICE clinical guideline CG90.[6]

5.1.2 Inclusion/exclusion criteria

The eligibility criteria used for inclusion in the economic evaluation reviews are presented in Table 5.3.

Table 5.3: Eligibility criteria for systematic review of cost-effectiveness analyses

Criteria	Include	Exclude
Population	Adult patients with MDD (with a particular focus on patients who have progressed to TRD)	Paediatric patients (<18 years), patients with related conditions (dysphoria, dysthymia, melancholia, SAD, mood disorder, GAD), and patients with comorbid depression
Intervention(s)/ comparator(s)	Antidepressant drugs, including: • Esketamine	Interventions not listed in inclusion column

Criteria	Include	Exclude
	• SSRIs (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline) • SNRIs (desvenlafaxine, duloxetine, levomilnacipran, milnacipran, venlafaxine) • Vortioxetine • Trazodone • Reboxetine • Tricyclics • Tetracyclics • Monoamine oxidase inhibitors • Atypical antipsychotics • Risperidone • Other pharmacological agents (agomelatine, tianeptine, lithium, amineptine, bicifadine, bupropion, lamotrigine, mazindol, sibutramine, olanzepine/fluoxetine) Augmentation and adjunctive strategies Non-pharmacological interventions, including: • Behavioural activation • CBT and other types of psychotherapy • Combined CBT + antidepressant • Deep brain stimulation • ECT • Interpersonal psychotherapy • Magnetic seizure therapy • Repetitive TMS • Transcranial direct current stimulation • Transcranial magnetic simulation • VNS	
Outcomes	Outcomes of interest included: • Model summary and structure • Sources of model inputs • Assumptions underpinning model structures • Discounting of costs and health outcomes • Total costs and health outcomes • ICERs	Outcomes not listed in inclusion column
Study design	Eligible study designs included: • Cost-utility analyses • Cost-effectiveness analyses • Cost-benefit analyses • Cost-minimisation analyses	Reviews/editorials Budget impact analyses
Territory of interest	No restriction – although primary focus was UK	-
Date of publication	Original review: no restriction	Original review: NA

Criteria	Include	Exclude
	April 2019 update: post-July 2018	April 2019 update: pre-July 2018
Language of publication	English language publications or foreign language publications with an English abstract	Foreign language publications without an English abstract

Based on Table 53 of the CS appendices[17]
CBT = cognitive behavioural therapy; CS = company submission; ECT = electroconvulsive therapy; GAD = generalised anxiety disorder; ICER = incremental cost-effectiveness ratio; MDD = major depressive disorder; NA = not applicable; SAD = seasonal affective disorder; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; TMS = transcranial magnetic stimulation; TRD = treatment-resistant depression; UK = United Kingdom; VNS = vagal nerve stimulation

ERG comment: The ERG noted that interventions such as amitriptyline and mirtazapine are ignored in the SLR. All other criteria seem appropriate.

5.1.3 Conclusions of the cost effectiveness review

The electronic database searches identified a total of 3,132 citations. Following removal of 431 duplicates, 2,701 citations were screened on the basis of title and abstract. A total of 341 citations were considered to be potentially relevant and were obtained for full text review. At this stage, a further 181 citations were excluded. Hand searching yielded 20 additional relevant citations. Therefore, a total of 16 publications (economic evaluations n=12, previous HTA submissions n=4) were identified for final inclusion in the review during the original search on July 2018. The updated systematic review on April 2019 found one additional economic evaluation. However, according to the company, none of the economic evaluations identified by the SLR evaluated the cost effectiveness of ESK-NS + OAD and were therefore not directly generalisable to the NICE decision problem.

ERG comment: The ERG agrees with the conclusions of the company's cost effectiveness review.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.4: 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	Adverse events were not adequately included: only as a scenario analysis and assuming only a GP contact cost.
Perspective on costs	NHS and PSS	Included
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Included, although assuming a mix of treatments as comparator and comparison to separate treatments only in a scenario.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Probably inadequate as not a lifetime horizon

Element of health technology assessment	Reference case	ERG comment on company's submission
Synthesis of evidence on health effects	Based on systematic review	Included
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.	Included
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Included
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Included
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Included
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	Included
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Included

EQ-5D = European Quality of Life-5 Dimensions; ERG = Evidence Review Group; GP = general practitioner; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; PSS = personal social services; QALY = quality-adjusted life year; UK = United Kingdom

5.2.2 Model structure

As reported in Section B 2.2.2 of the CS, the model is a state transition model with a cycle length of four weeks and, in addition to death, four health states, which are summarised in Table 5.5.[1]

Table 5.5: Health state definitions

Health state	Health state definition
MDE	Patients experience moderate to severe symptoms of major depressive disorder with a MADRS ≥28 and failed to respond to at least two different OAD treatments of adequate dosage and duration.
Response	Patients experience a 50% or greater MDD symptom improvement from patient's baseline MADRS score but did not achieve the threshold for remission (MADRS ≤12).
Remission	Associated with a period during which the patient is either symptom-free or has only minimal symptoms. The threshold used in the model for achieving remission was MADRS ≤12.
Recovery	Represents an extended asymptomatic phase, achieved after a patient remains in relapse-free remission for 36 weeks in a row (or approximately nine months).

Based on Table 42 of the CS[1]
CS = company submission; MADRS = Montgomery-Åsberg Depression Rating Scale; MDD = major depressive disorder; MDE = major depressive episode; OAD = oral antidepressant.

Patients enter the model in the major depressive episode (MDE) health state, after having failed to achieve a *"…clinically meaningful improvement…"* (page 160, CS) after treatment with at least two OADs *"prescribed in adequate dosages for adequate time"* (page 160, CS).[1] During each four-weekly Markov cycle, patients can occupy MDE, response, remission, recovery or death health states. Transition to recovery can only occur from remission and only after nine months (36 weeks) in the remission state and then with certainty.

Cycles in the model allow for up to three subsequent treatments, switching to a new treatment following:

- a non-response to acute treatment (at four weeks),
- a loss of response i.e. relapse from the response or remission health states respectively (5–40 weeks), or
- experience a recurrence of the MDE during the recovery health state (41 weeks+).

After three subsequent treatments, patients enter the MDE state from which they can still respond or go into remission, whilst being treated with best supportive care (BSC). The company stated that this structure had been validated by NICE Preliminary Independent Model Advice (PRIMA), which indicated that it was an improvement on the last NICE appraisal for depression, i.e. TA367.[9, 12]

Transitions between health states are governed by treatment phase:

- **Acute phase (weeks 1 to 4):**
    - Patients remain in MDE state for one cycle. They can then:
        - Transition to response or remission,
        - Remain in MDE state, but move to subsequent treatment,
        - Remain in MDE state, but discontinue treatment, or
        - Die.
- **Continuation phase (weeks 5 to 40):**
    - Patients in the response state can:
        - Continue treatment and remain in the same health state,
        - Improve their depressive symptoms further and transition into the remission health state,
        - Lose treatment response, return to the MDE health state, and begin the next treatment in the sequence,
        - Discontinue treatment and remain in the same health state, or
        - Die.
    - Patients in the remission state can:
        - Continue treatment and remain in the same health state,
        - Enter the recovery health state after 36 weeks (approximately nine months) of relapse-free remission,
        - Relapse (i.e. return to the MDE health state) and begin the next acute treatment in the sequence,
        - Discontinue treatment and remain in the same health state, or
        - Die.
- **Maintenance phase (weeks 41+):**
    - Patients in the recovery health state could:
        - Experience a recurrence event (i.e. return to the MDE health state) and move on to the next treatment in the sequence,
        - Continue treatment and remain in the current recovery health state, or

- Die

Transition probabilities are reported in section 5.2.6 of this report.

ERG comment: The model structure seems plausible and responds appropriately to the critique in TA367.[9]

5.2.3 Population

The population was described in the CS as adults with TRD with a moderate to severe depressive episode.[1] A moderate to severe episode of TRD was assumed to have minimum duration of two years. Treatment-resistant MDD was defined as non-response to two or more OADs prescribed at an adequate dose and for an adequate duration in the current episode.

ERG comment: The population is broadly consistent with the NICE scope and the expected marketing authorisation.[16] However, there are some issues of concern, as described in section 3.1.

The company did perform a subgroup CEA for the 65 years+ age group using TRANSFORM-3 to estimate transition probabilities for remission and response (the equivalent of those from TRANSFORM-2 presented in Section 5.2.6 of the CS). They also used utilities and dosing from TRANSFORM-3, but transition probabilities beyond the acute phase appear to have come from SUSTAIN-1.

Therefore, given that the NICE scope has no upper age limit, in the clarification letter the ERG requested that the main cost effectiveness analysis (CEA), i.e. for age <65 years, informed by TRANSFORM-2 and SUSTAIN-1 be combined with that for age 65 years+, using TRANSFORM-3 as well as SUSTAIN-2. The company responded by submitting a new version of the base-case model to include acute response and remission transition probabilities and utilities for MDE, response and remission/recovery states from both TRANSFORM-2 and TRANSFORM-3, weighted by percentage in each age group such that if set to 0% for age >65 years one gets the same result as in the original base-case.[3] This forms the starting point for the ERG base-case, see Section 7.2.

5.2.4 Interventions and comparators

The intervention in the analysis was ESK-NS co-administered with a newly initiated OAD (ESK-NS + OAD), see Section B.3.2.7 of the CS).[1] As stated in Section B.3.2.11.1, ESK-NS comes as a single-use device that delivers a total of 28 mg of esketamine in two sprays (one spray per nostril). One device (for a 28 mg dose), two devices (for a 56 mg dose), or three devices (for an 84 mg dose) are to be used, with a five-minute interval between each device. The average number of sessions per week and devices per session in the acute phase were derived from TRANSFORM-2, while for subsequent time-points they were derived from SUSTAIN-1. In TRANSFORM-2 on Day 1 of the induction phase, all patients randomised to receive esketamine nasal spray started with a dose of 56 mg weekly. Thereafter, esketamine could be dosed flexibly (56 or 84 mg) based on efficacy and tolerability up until Day 15 (or Day 18 if the Day 15 treatment session did not occur). Beyond Day 15, the esketamine nasal spray dose was to remain unchanged (see Figure 1, CS).[1] The precise rules of determining efficacy and tolerability were not reported in the CS or Appendix M.[1, 17] In TRANSFORM-3, the starting dose was 28 mg which could also be increased to 84 mg by Day 25 without any specification of the precise rules.

SUSTAIN-1 had the same dosing as TRANSFORM-2 in the first four weeks for direct entry patients. These patients then joined those who had been transferred from TRANSFORM-1 and TRANSFORM-2 to enter the optimisation phase where the dose could be adjusted at either week eight or 12:

- At week eight, reduce from weekly to every other week if MADRS total score ≤12; otherwise continue weekly until week 16,
- At week 12, increase to weekly if MADRS total score was >12; otherwise continue every other week until week 16.

In SUSTAIN-1, from week 16 onwards, the following rules applied:

- At week 16, if every other week AND MADRS total score >12 then frequency was increased to weekly; otherwise continue every other week,
- At week 16, if weekly then continue for four weeks.
- At week 20 or later:
    - if weekly AND MADRS total score ≤12 for last four weeks then reduce to every other week,
    - if every other week AND MADRS total score >12 then increase to weekly,
    - otherwise continue either weekly or every other week.
- Maximum of three changes permitted such that, if a given patient was unable to sustain, improvement on every other week dosing, they were to remain on a weekly dosing regimen for the remainder of the maintenance phase.

SUSTAIN-2 had the same weekly dosing as TRANSFORM-2 (aged <65 years)/ TRANSFORM-3 (aged ≥65 years) in the first four weeks for direct-entry patients. These patients joined those who had transferred from TRANSFORM-3 and then remained on the same weekly dose for the next four weeks. For direct-entry patients only, from week nine dosing could decrease to every other week and then switch back to weekly at four-weekly intervals according to the MADRS 12 threshold. Down titration was also possible for tolerability. For those who had been transferred from TRANSFORM-3 no change in dose or frequency was allowed from week nine except a reduction for tolerability.

Neither the concomitant OAD nor the comparator OAD were specified in the CEA: instead OAD was expressed as a mix of eight OADs, according to UK market share (See Section 5.2.8.3). The company did perform a scenario analysis (See Section B.3.4.4.9) based on an NMA using data from TRANSFORM-2 of response and remission presented in Appendix D, which compared ESK-NS + OAD with various other comparators in the form of drug classes.[1, 17] Table 5.6 shows the list of comparators as well as the remission and response probabilities. The NMA was based on an adjustment for the placebo effect (see Section 5.2.6.1 for more detail on the method of estimating those for ESK-NS + OAD and OAD).

Table 5.6: Response and remission rates at the end of the acute treatment phase

Treatment	Remission, %[a]	Response (but not remission), %[b]	Remission, %[c]	Response (but not remission), %[d]
ESK-NS + OAD (Switch SSRI/SNRI)	52.48	16.83	52.48	16.83
OAD (Switch SSRI/SNRI)	17.71	4.36	30.81	8.79
Aug tricyclic (nortrip) ± PBO	22.70	4.71	37.78	9.49
Aug SSRI/SNRI + AAP	27.65	4.04	44.45	8.15

Treatment	Remission, %[a]	Response (but not remission), %[b]	Remission, %[c]	Response (but not remission), %[d]
Aug SSRI/SNRI + lithium	21.98	2.57	36.88	5.24
Aug SSRI/SNRI ± PBO	16.25	2.05	28.80	4.19
Switch tetracyclic (mirtazapine)	13.28	3.26	24.09	6.67
Switch SSRI + AAP	22.38	4.04	37.51	8.15

Based on Table 79 of the CS[1]

[a] MADRS ≤12 with adjustment for 6 clinic visits; [b] ≥50% reduction in MADRS from baseline but MADRS score >12 with adjustment for 6 clinic visits; [c] MADRS ≤12 with no adjustment; [d] ≥50% reduction in MADRS from baseline but MADRS score >12 with no adjustment

AAP = atypical antipsychotic; Aug = augmentation; CS = company submission; ESK-NS + OAD = esketamine nasal spray; MADRS = Montgomery-Åsberg Depression Rating Scale; nortrip = nortriptyline; OAD = oral antidepressant; PBO = placebo; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

For all other parameters, equivalence with OAD was assumed given that these parameters were estimated from STAR*D and the company stated that this study included OAD and other augmentation strategies in 1st and 2nd line TRD.The results of the analysis are shown in Section 6.2. The company argued that the NMA was not robust enough to include these comparators in the base-case.

No non-pharmacological treatments, such as psychological therapy, were included as comparators (without concomitant pharmacological treatment).

ERG comment: The ERG requested clarity on the criteria by which dose was determined in TRANSFORM-2 (applicable also TRANSFORM-3) to which the company responded by stating that *"the intention was to emulate real-world clinical practice, thus there was no prescriptive algorithm"*.[3, 18] The continued lack of clarity on dosing in TRANSFORM-2 and TRANSFORM-3 trials plus the complex dose changes in SUSTAIN-1 and SUSTAIN-2 mean that it is difficult to know how applicable to clinical practice the transition probabilities estimated from the trials would be (see Section 5.2.6). This basis for questionable applicability is in addition to that in terms of whether the data to inform those transition probabilities were from patients were direct-entry or transferred-entry (see Section 5.2.6).

The ERG is convinced that the limitations of the NMA (see section 4.4) are sufficient to exclude those included comparators except in a scenario analysis, although the results should be re-calculated based on the NMA results unadjusted for the placebo effect, which the company provided in response to request for clarification (see Table 5.7).[3]

The ERG recognises that adopting a mix of OADs as concomitant and comparator treatment is not ideal. Indeed, there is some evidence of variability of effectiveness between the four OADs, see section 4.2.6. There is therefore the possibility that ESK-NS might be cost-effective in combination with one OAD and not another. However, the ERG did not have the data to implement the required variation in all parameter estimates required for the model. Therefore, it seems reasonable to not differentiate between specific OADs as either an add-on to ESK or a comparator. However, applicability to clinical practice of results would be highest in those patients who might be switched to one of the OADs prescribed in the trials.

5.2.5 Perspective, time horizon and discounting

As stated in Section B.3.2.4 of the CS, the base-case time horizon was five years.[1] This was justified by all of the treatment related benefits having been accounted for, see Figures 21 and 22 of the CS which were of the Markov trace for ESK-NS + OAD and OAD + PBO-NS, respectively.[1]

As stated in Section B3.2.5 of the CS, the base-case analysis took the perspective of the National Health Service (NHS) and personal social services (PSS) in England. Both costs and outcomes (life years and QALYs) were discounted at 3.5%, in line with the NICE Guide to the Methods of Technology Appraisal 2013.[50] The impact of discounting at 0% and 6% was assessed in sensitivity analyses.

ERG comment: The ERG asked the company to justify the choice of five years as a time horizon, given that it is longer than the time horizon used in TA367 and Edwards et al. 2013, but shorter than a lifetime horizon.[9, 18, 51] The ERG also requested the company to extend the time horizon to a lifetime given that this is according to the NICE reference case and to capture the chronic recurrent or episodic nature of the condition and to account for the effect on mortality associated with suicide.[18] In response, the company presented a sensitivity analysis that showed that the incremental cost effectiveness ratio (ICER) up to a time horizon of 50 years was lower than the base-case, e.g. £4,314 at 50 years.[3] The ERG also notes that by 20 years the percentages of the cohort in the response, remission or recovery health states in the cohort treated with ESK-NS + OAD are equal to those in the cohort treated with OAD + PBO-NS. Therefore, from this point onwards there can be no further difference in cost or QALYs and thus no need to extend the time horizon beyond this point. The ERG therefore has adopted 20 years in the ERG base-case, see section 7.2).

5.2.6 Treatment effectiveness and extrapolation

5.2.6.1 Acute phase

The transition probabilities (in the form of percentages) for response and remission are presented in Table 5.7. Response and remission values were estimated from TRANSFORM-2 (see Section 4.2.5), with the adjustment then applied to the OAD + PBP-NS arm only. Response (but not remission) was calculated by subtraction.

Table 5.7: Response and remission rates at the end of the acute treatment phase

Treatment	Remission, % (SE)[a]	Response (but not remission), % (SE)[b]	Response[c]
ESK-NS + OAD	52.48% (4.97)	16.83% (3.72)	69.31%
OAD + PBO-NS (unadjusted)	31.00% (4.26)	21.00% (4.07)	52.00%
OAD + PBO-NS (adjusted for six visits[d])	18.00% (3.84)	16.00% (3.67)	34.00%

Based on Table 45 of the CS[1]

[a] MADRS ≤12; [b] ≥50% reduction in MADRS from baseline but MADRS score >12; [c] ≥50% reduction in MADRS from baseline; [d] Base-case

CS = company submission; ESK-NS = esketamine nasal spray; MADRS = Montgomery-Åsberg Depression Rating Scale OAD = oral antidepressant; PBO-NS = placebo nasal spray; SE = standard error

The company argued that an adjustment was justified because:

1. there is a positive effect on outcome in both arms of the trial due to clinic visits such that the more visits the bigger the effect

2. this positive effect would continue to be observed in clinical practice only for ESK-NS and not for standard care

The adjustment is a reduction in the rates of response and remission estimated as the effect of a reduction in the number of clinic visits from eight in the trial to two in clinical practice. The size of the adjustment was estimated in multiples steps:

1. use the lower value of 0.67 from a range of 0.67 to 0.86 estimated as the improvement in the Hamilton Depression Rating Scale (HAM-D) of an extra clinic visit. The study also showed that this effect doubled with two extra visits.
2. 0.67 HAM-D points per follow-up clinic visit was converted to MADRS using a 1-point improvement on HAM-D being equivalent to ~1.2 points improvement on MADRS. This was based on a study that showed that 10, 20 and 25 points on the HAM-D corresponded to 12, 26 and 34 points on the MADRS.

This implied an adjustment of 0.804 MADRS points per clinic visit, i.e. 0.67 * 1.2. The number of extra clinic visits in clinical practice was estimated base on there being only two as opposed to eight visits in the first four weeks i.e. the acute phase. This then implies a decrease in the MADRS of 4.842, i.e. 6 * 0.804, which was applied to each patient in order to recalculate percentage remission and response.

ERG comment: The values for ESK-NS + OAD are appropriate. Only the unadjusted values for OAD + PBO-NS are valid in comparing with ESK-NS + OAD. This is because there is insufficient reason for believing that the values observed in the TRANSFORM-2 trial in the placebo arm have been overestimated *relative to* those in the intervention arm. There are several grounds for this argument:

1. Only the *treatment effect*, i.e. the difference/contrast between intervention and comparator in an RCT is unbiased. This is the fundamental basis of having a comparator arm. The company erroneously claim that *"high placebo rates"*, i.e. the outcome in only the placebo arm make estimating the *"true relative treatment effect"* a challenge (page 17, CS).[1] However, it is precisely because of outcomes that might be changed and often inflated beyond that due to the intervention itself that a placebo control arm is included, i.e. the so-called 'placebo effect' applies to the intervention as well as the control arm. Therefore, removing this placebo effect from the control arm means that it is retained in the intervention arm. The treatment effect, i.e. the difference between intervention and comparator thus becomes biased.
2. The company argues that one explanation of the placebo effect is *"high frequency and intensity of patient-health care professional interaction due to twice-weekly visits (of considerable length)"*, although erroneously applying the term *"treatment effect"* to the outcome in only the placebo arm of the trial (page 49, CS).[1] While the ERG would agree that this explanation of the placebo effect possesses some face validity, such an effect would still apply to both arms. This is acknowledged by the company in the response to clarification.[3] However, they claim that the effect of increased clinic visits would continue in clinical practice only for ESK-NS + OAD and that therefore the outcome is only elevated beyond what would be expected in clinical practice for the placebo arm and not the ESK-NS + OAD arm. They argue, on this basis, that removing from the placebo arm does not create a bias, but instead nullifies the bias of the placebo effect in the placebo arm. This implies that the technology in the decision problem is not ESK-NS + OAD, but it is ESK-NS + OAD + 8 clinic visits and that standard care, to which it should be compared, is OAD + 2 clinic visits. However, if, as the company claim, efficacy does improve with clinic visit frequency then standard care would also be improved by increasing the number of visits from 2 to 8. Therefore, the comparator for this ESK + OAD + 8 clinic visits would be OAD + 8 clinic visits which is the comparator in the trial, thus negating the need for any adjustment.

3. The evidence for the size of the effect of number of clinic visits is weak. There is no evidence from the trial itself that the basis of the placebo effect is the number of clinic visits, since everyone receives the same number of visits in TRANSFORM-2. The evidence provided by the company is from other studies and involves questionable assumptions regarding the relationship between clinic visit number and HAM-D and between HAM-D and MADRS. It is also unclear what the number of additional clinic visits might be.
4. The evidence for the placebo effect being the result of number of clinic visits as opposed to any other source is weak. As the company state, there are other plausible explanations of the placebo effect, two of which are listed by the company: *"Use of a nasal spray delivery system leading to patient expectation of 'something novel'"* and *"High patient expectation of benefit due to the portrayal in the media of esketamine as a 'promising' new treatment option for depression"*. (page 49, CS).[1] There is no evidence that these would play any less of a role than clinic visit frequency in mediating the placebo response. The company might also argue that, just as for clinic frequency, these factors would also apply the use of esketamine in clinical practice. On this basis one might regard the intervention to be ESK-NS + OAD + 8 clinic visits + patient expectations. Of course, it would be difficult to conceive of a suitable comparator in clinical practice that comprised partly of such expectations without actually giving the drug itself. Perhaps ironically therefore, these factors might be more of a reason for an adjustment than clinic visit frequency. However, as with clinic frequency, it is impossible to estimate the size of the effect from the trial data given that it applies equally to both arms.

The conclusion of the ERG therefore is that, whilst it might be the case that some of the placebo effect, however mediated, might continue into clinical practice, it is possible to reproduce it by increasing clinic visits even without esketamine and it is impossible to have confidence as to the size of any effect that might only apply to esketamine in clinical practice. On this basis the ERG requested that the company either use the unadjusted estimates of response for OAD + PBO-NS for the model base-case or perform the same adjustment to ESK-NS + OAD to which the company responded that a scenario had been presented without the adjustment (see Section 6.2.3.1). They also reiterated the justification employed within the CS, which has been critiqued by the ERG (as above). On this basis, the ERG base-case removes this adjustment and assumes an increase the cost of clinic visits for OAD to be identical to the monitoring cost of OAD in a scenario analysis (See Section 7.2).

5.2.6.2 Continuation and maintenance phases

Continuation phase:

The transition probabilities of response to remission are shown in Table 5.8. These were estimated by Poisson regression analysis of the SUSTAIN-1 data on patients who were initially were 'stable responders' and followed up over time to identify those who had a MADRS ≤12 for at least three of the last four weeks (three out of any four consecutive weeks during follow-up). In SUSTAIN-1, stable response was defined as a ≥50% reduction in the MADRS total score from baseline in each of the last two weeks of the optimisation phase (weeks 15 and 16) without meeting the criteria for stable remission.

Table 5.8: Four-week transition of moving from response to remission (MADRS ≤12) state

Treatment	Response to remission (SE)
ESK-NS + OAD	19.93% (4.98)
OAD + PBO-NS	12.39% (3.10)

Based on Table 46 of the CS[1]

CS = company submission; ESK-NS = esketamine nasal spray; MADRS = Montgomery-Åsberg Depression Rating Scale OAD = oral antidepressant; PBO-NS = placebo nasal spray; SE = standard error

The transition probabilities for loss of response (response to MDE) and relapse (remission to MDE in weeks 5 to 40) are shown in Table 5.9. Loss of response and relapse were stated to have been estimated from SUSTAIN-1 for ESK-NS + OAD and from STAR*D for OAD.[1]

Table 5.9: Four-week risk of relapse, loss of response and recurrence

Treatment	Relapse (SE)	Loss of response (SE)	Recurrence (SE)
ESK-NS + OAD	5.57% (4.98)	4.19% (2.55)	2.88% (1.80)
OAD + PBO-NS	9.24% (3.10)	22.43% (5.43)	2.88% (1.80)

Based on Table 47 of the CS[1]

CS = company submission; ESK-NS = esketamine nasal spray; OAD = oral antidepressant; PBO-NS = placebo nasal spray; SE = standard error

For loss of response on ESK-NS + OAD, as for response to remission, follow-up was also from then end of the optimisation phase (week 16). For relapse on ESK-NS + OAD, data from SUSTAIN-1 came from those who were 'stable remitters'. 'Stable remission' was defined as a MADRS total score of ≤12 for the last two weeks of the optimisation phase plus for at least three of the last four weeks of the optimisation phase with one excursion of the MADRS total score >12 or one missing MADRS assessment permitted at Week 13 or 14 of the optimisation phase only. Only those patients who relapsed during the first 24 weeks were counted: this corresponded to weeks 5 to 40, i.e. the continuation phase.

For loss of response on OAD, the company argued that SUSTAIN-1 could not be used because the only patients randomised to a placebo arm were those who had already been in 'stable response' or in 'stable remission' whilst on ESK-NS + OAD. Therefore, the probability was calculated as the weighted average of two risks, 22.2% for first-line TRD and 22.8% for second-line TRD, each estimated by fitting an exponential distribution to digitised Kaplan-Meier (KM) plots from STAR*D data.[3, 52] The weights were the percentages of those patients who had had two versus three or more previous treatment failures in SUSTAIN-1. The same method was used for relapse with 6.8% for first-line TRD and 12.8% for second-line TRD.[3]

Maintenance phase:

The transition probabilities for recurrence (remission to MDE in weeks 41+) are shown above in Table 5.9 (third column). For both, ESK-NS + OAD and OAD, the data pooled from both study arms of the double-blind phase of SUSTAIN-1 was used. All stable remitters who relapsed after 24 weeks of maintenance treatment (equal to 36 weeks post-acute treatment) were counted for the calculation of the recurrence rates.

ERG comment: It was unclear to the ERG how data were chosen from SUSTAIN-1 in order to estimate the transition probability of response to remission given that patients appear to enter SUSTAIN-1 from various sources, including either of the TRANSFORM-1 or TRANSFORM-2 or by direct entry. The company also specified that response and remission were defined more restrictively than in

TRANSFORM-2 in the sense that they had to have been "stable" and data were only analysed from the end of the optimisation phase (week 16). The company were therefore asked to confirm that the data sources for each of the transition probabilities appropriately reflect the starting health state, as defined by the MADRS, the treatment pathway and timing.[18] If this is not the case then they were asked to re-estimate the transition probabilities using the correct data.[18] In spite of an ERG request for clarification, the company did not provide any further details.[3, 18]

It was also unclear to the ERG why STAR*D was chosen given that at least some patients who entered SUSTAIN-1 were originally randomised to OAD + PBO-NS in TRANSFORM-1 or TRANSFORM-2. Therefore, there should have been some patients who had been observed to have lost response or relapsed whilst on OAD + PBO-NS. Indeed, the CONSORT diagram (Figure 11 in Appendix D of the CS) shows that 86 patients (including 48 from TRANSFORM-2) continued to be followed-up and, of these, 55 (33 from TRANSFORM-2) became stable remitters and responders during the optimisation phase with only one loss to follow-up beyond this phase.[17] The company did not provide any additional clarification.[3] Also, the loss or response value for OAD is much higher than those for ESK-NS + OAD, by a factor of over five, which is much higher than the relative risk in the acute phase. The company did conduct a scenario analysis (Section B.3.4.4.8) that was reported to have used SUSTAIN-1 to inform response and relapse. However, the precise data used was not clear, appearing to have been from only those patients who had received ESK-NS + OAD and then been randomised for a second time to OAD only. This contrasts very strongly with TA367, where the probability of relapse was assumed to be the same for all treatments.[9] The committee for TA367 also noted that, although STAR*D data provided the best available evidence, it might impose a poorer prognosis on patients than would be observed in the index trial.[9] The ERG believes that the problem with this submission is similar in that all ESK-NS transition probabilities have been estimated from the company trials, but that those for OAD beyond the acute phase have been estimated from a completely different source This probably incorporates a bias in favour of ESK-NS, not least because of the "placebo effect". The company in TA367 took a more conservative approach to relapse in that it assumed there to be no difference between intervention and comparator, using 14.2% from Limosin 2004 for second-line (one line prior to TRD) and 25.0% for third-line (first-line TRD) and 42.6% for fourth- (second-line TRD) and fifth-lines (third-line TRD) from STAR*D for all subsequent lines.[9] For this STA, the same values from STAR*D could be used as in TA367, but the ERG could not locate the values used in TA367 in the STAR*D paper.[53] Therefore, in a scenario, the ERG have assumed the same probability of relapse and loss of response for OAD as ESK-NS + OAD, see Section 7.2.

5.2.6.3 Discontinuation (for reasons other than loss of efficacy)

It was assumed that patients would not discontinue OAD in any phase for any reason other than lack of response. Discontinuation for any other reason from ESK-NS + OAD is presented in Table 5.10 for the acute, continuation, and maintenance phases.

Table 5.10: Risk of discontinuation following initial treatment

Comparator	Acute		Continuation		Maintenance	
	Risk	SE	Risk	SE	Risk*	SE
ESK-NS + OAD	0.00%	0.00%	1.69%	0.42%	24.89%	6.22%

Based on Table 48 of the CS[1]

* Based on assumptions

CS = company submission; ESK-NS = esketamine nasal spray; OAD = oral antidepressant; SE = standard error

Acute phase
It was assumed that patients would not discontinue ESK-NS + OAD in the acute phase for any reason other than lack of response.

Continuation phase
A discontinuation risk for other reasons was derived from SUSTAIN-1 by fitting an exponential distribution to the pooled data from the ESK-NS + OAD arm from stable responders and stable remitters. Relapse was counted as a censoring event. The estimated four-week risk was 1.69% (20% annually) and is presented in Table 5.10.

Maintenance phase
It was also assumed that 35.4% of patients were assumed to stop ESK-NS immediately upon achieving recovery, i.e. on being in the remission state after 40 weeks of treatment. This was the percentage of patients in SUSTAIN-1 who had ≤2 total number of MDD episodes, including the current episode.[54] The conceptual basis was that *"…a benefit of ESK-NS is it can be discontinued while patients can still receive OAD for recurrence prevention"* (p.175, CS).[1] For those patients who did not discontinue immediately, a four-week discontinuation risk of 25% for ESK-NS + OAD was stated to have been used during recovery. However, given that the percentage in Table 5.10 is lower than this and that Figure 24 of the CS shows the percentage remaining on ESK-NS to be 0% at two years, it appears that 24.89% was estimated in order to imply 0% at two years. These assumptions were stated by the company to have been validated by expert clinical opinion, although no reference to any report was cited. However, the minutes of an Advisory Board, dated 4th June 2019, presented as Appendix F in the response to clarification, revealed that there appeared to have been general agreement with a figure of 35% discontinuing on reaching recovery and a further 25% monthly risk.[3] Patients in the response state during the maintenance phase could not discontinue, this being justified by being *"at high risk of relapse"* (p.175, CS).[1]

The impact of discontinuation in either the continuation or the maintenance pahse was to stop incurring the cost of ESK-NS and only incur the cost of OAD whilst having no effect on QALYs (because patients were assumed to remain in the remission or recovery state until loss of response, relapse or recurrence). This was argued by the company to be conservative.

ERG comment: The ERG considers that it is reasonable to assume no discontinuation during the acute phase and the rate during the continuation phase also appears to be reasonable given that it was estimated from the trial data albeit based on an arbitrary definition of stable and choice of exponential distribution. However, the rates of discontinuation in the maintenance phase were not based on any observed data, but instead on assumptions. The company could have continued to use data from the SUSTAIN-1 study, which could have had a parametric curve fitted to extrapolate up to the time horizon. It is also not reasonable to assume that the treatment effect is maintained, i.e. no decrease in QALYs on discontinuing ESK-NS and continuing with only OAD. Indeed, the company themselves provide evidence that continuation of the treatment effect on discontinuing ESK-NS is not credible in Section B2.2 of the CS: *"…for ESK-NS, it was uncertain whether long-term treatment would be necessary as it was hypothesised that the antidepressant effect following short-term ESK-NS treatment could be maintained with an OAD alone. The maintenance study, SUSTAIN-1, however, showed this to not be the case: patients who discontinued ESK-NS demonstrated a significantly greater relapse rate than those who remained on ESK-NS…"* (p.51).[1] The ERG also question the assumption that discontinuation implies no decrease in QALYs. In the continuation phase, where the rate was estimated from the data, relapse was a censoring event, which implies that patients discontinued without relapsing. However, no evidence was presented as to the rate of relapse of those discontinuing. In the maintenance

phase, where the rate of discontinuation was assumed, it is completely opaque as to the rate of recurrence in those who have discontinued. In both phases, it is also unclear whether there might be a diminution in utility and thus a loss of QALYs even if relapse or recurrence do not occur. In the absence of any data as to the effect on relapse or recurrence or utility on discontinuation of ESK-NS, the ERG assumed no discontinuation for reasons other than loss of efficacy in the ERG base-case, see section 7.2.

5.2.6.4 Subsequent treatments

As reported in Section B.3.2.9.3, the company estimated the transition probabilities for each of three further lines of subsequent treatment (Table 5.11). They stated that they had been estimated from the STAR*D trial, as they were in TA367.[1, 9, 53] Although the final numbers were stated to have been validated by two advisory boards, it is unclear how any of the numbers reported in STAR*D were transformed to produce those in Table 5.11, other than that *"…data being converted to 4-week risks using standard formulae"* (p.177, CS).[55, 56]

The precise mix of OADs that formed subsequent treatment was not specified, but examination of the model revealed that the cost was identical to the OAD as employed in first-line treatment, see section 5.2.8.3).

Table 5.11: Health state transition probabilities – subsequent treatment

Treatment	MDE to Response*	MDE to Remission*	Response to Remission†	Loss of Response†	Relapse†	Recurrence†
TRD line 2	3.54%	0.86%	2.76%	12.79%	22.81%	2.88%
TRD line 3	2.75%	0.65%	2.76%	12.79%	22.81%	2.88%
TRD line 4	2.14%	0.49%	2.76%	12.79%	22.81%	2.88%

Based on Table 50 of the CS[1]

* Evaluated at the end of the acute phase; † Per 4-week cycle.

CS = company submission; MDE = major depressive episode; TRD = treatment-resistant depression.

ERG comment: Although the company stated that they used STAR*D, their methods were unclear and the resulting values were much lower than those in STAR*D.[53] If one assumes that TRD line 2 is equivalent to fourth-line since the onset of MDD then this might also be equivalent to Step 4 in STAR*D.[53] At this line in STAR*D, the probabilities of response and remission were reported to be 16.3% and 13.0%, which could therefore be compared to 3.54% + 0.86% (assuming that response is the sum of these two transition probabilities) and 0.86%, respectively. The company stated that values were converted to four-week risks, but it is not clear what the unconverted risks were, nor what the number of weeks for the unconverted risks was. Additionally, it is not clear what the basis of the conversion was. This is because such a conversion would only be required if the event could occur over multiple four-weekly cycles such that the cumulative risk is then equal to the unconverted one. According to the model structure, transition from MDE to either response or remission can only occur in the first cycle on starting any line of therapy. As soon as it is determined that the patient has failed to respond or remit, they then move to the next line, thus preventing any further response or remission. Of course, it might be that in STAR*D response or remission did occur later than four weeks after initiating treatment and the report of the STAR*D study does not report the mean number of weeks to response or remission, i.e. 8.3 and 7.4 respectively. Therefore, the model should allow transition from MDE to response or remission over more than one cycle. Otherwise, the full effectiveness of the treatment is underestimated. As a second-best solution the ERG assumed that all response or remission occurred in the first cycle on starting treatment in an ERG scenario, see section 7.3. This approach will overestimate the benefit and cost of treatment, but only because of the lower rate of discounting applied to the QALYs and cost due

to some response or remission occurring too early. However, given that patients were encouraged to switch treatment if no response and follow-up visits occurred every two months, it is likely that this was determined after no more than one more cycle.[53] As also recommended by the committee of TA367, a decrease in response and remission was applied at each line of therapy by multiplying the values for OAD by a factor equal to the ratio of values in Step 3 versus Step 4 in STAR*D.[9, 53] Specifically the FAD for TA367 states that *"…the Committee considered it more appropriate to apply a proportionate reduction in the rates of remission for fourth and subsequent lines of treatment, as seen in the STAR*D trial, to the remission data used for third-line treatment"* (p. 48).[9] These ratios are: 13.7/13.0 and 16.8/16.3 for remission and response, respectively. Therefore, the factors applied at second-, third- and fourth-line TRD are the ratios, the ratios to the power 2 and the ratios to the power 3 respectively.

The ERG used the same method of adjusting by line for loss of response and relapse in this ERG scenario. This was achieved by using the company estimated values, for loss of response, of 22.2% for first line TRD and 22.8% for second line TRD and, for relapse, of 6.8% for first line TRD and 12.8% for second line TRD.[3]

5.2.6.5 Best supportive care

As reported in Section B.3.2.9.4 of the CS, the company estimated the transition probabilities for the BSC treatment mix (Table 5.12). In the model, the BSC treatment phase applies to patients whose disease has failed all previous treatments (fifth-line TRD and onwards). In this phase, patients could achieve response or remission at every cycle, and those who had achieved response or remission could experience loss of response or relapse at every cycle.

The efficacy estimates (response and remission) during the BSC treatment phase were stated to have been based on the HTA monograph by Edwards 2013, which were estimated from expert UK clinical opinion based on available evidence.[51] The authors of the monograph were stated to have been contacted to confirm how clinical opinion was derived and they confirmed that the results of the STAR*D trial formed part of the available evidence considered by the clinical experts informing the Edwards 2013 publication.[51] The efficacy estimates from the study were further validated by clinical experts in June 2019.[56]

The CS then stated that standard calculations were used to convert the reported two-month probabilities to four-week probabilities. To avoid double counting, the transition probability for remission was subtracted from the probability for response to derive the transition probability for MDE to response (excluding remission) that was used in the current model.

For sensitivity analysis, a confidence interval of ±10% of the mean was assumed for all probabilities shown in Table 5.12.

The precise mix of OADs that formed subsequent treatment was not specified, but examination of the model revealed that the cost was identical to the OAD as employed in first-line treatment, see section 5.2.8.3.

Table 5.12: Health state transition probabilities – best supportive care treatment mix

Treatment	Response[†,*]	Remission[†]	Loss of Response[†]	Relapse[†]
Best supportive care treatment mix	0.83%	0.41%	10.38%	4.20%

Based on Table 51 of the CS[1]

[†] Per four-week cycle. [*] Response minus remission.

ERG comment: Transition probabilities are attributed to an HTA monograph by Edwards 2013 supplemented by methodological advice from the authors of the HTA as to how clinical opinion was derived and then further supplemented by a validation exercise and subsequent conversion of two-month probabilities to four-week probabilities.[51, 56] There is no way of validating whether the assumptions and adjustments are appropriate. Results of the STAR*D trial in terms of transition to BSC are not reported. The means of converting these non-reported two-month probabilities to four-week probabilities were also not provided. Given that the mix of drugs referred to as subsequent therapy (up to fourth-line TRD) is precisely the same as the mix referred to as BSC (fifth-line), in the absence of specific data, it seems logical to apply the same method of estimating the transition probabilities for all lines of therapy beyond first-line in the model, see section 5.2.6.4. Therefore, the factors applied for BSC are the ratios 13.7/13.0 and 16.8/16.3 for remission and response respectively, each to the power 4 in the ERG base-case, see section 7.3.

5.2.6.6 Adverse events

In TRANSFORM-2, AEs, defined as those first reported or worsening in severity after initiating study treatment, were of mild to moderate severity. There were 14 most commonly reported AEs, with incidence ≥5% and occurring more frequently in the ESK-NS + OAD over the OAD + PBO-NS arm. These include nausea/vomiting, dissociation, dizziness, headache, vertigo, dysgeusia (distortion of sense of taste), somnolence, sedation, insomnia, blurry vision, increased blood pressure, paraesthesia, hypoesthesia (reduced sense of touch or sensation), and fatigue (see section B.2.10.1.1 of the CS and Table 4.17). Over 90% of TEAEs resolved on the same day of nasal spray self-administration.[54] Patients receiving ESK-NS + OAD were monitored during self-administration and post-administration for one hour on average. It was therefore assumed that, in the base-case, there would be no cost or negative impact on quality of life associated with AEs.

For completeness, a scenario analysis including AEs was conducted based on the rates of AEs seen in TRANSFORM-2 (see Tables 37 and 38 in Section B.2.10.1.1 of the CS) and their associated disutility.[1]

ERG comment: The ERG would have preferred a more extensive search for adverse events beyond what was reported in TRANSFORM-2. Specifically, the company report adverse events from SUSTAIN-1 (see Tables 4.18 and 4.19) but do not use this evidence in the economic model. However, the ERG considers that most of the effect to AEs will be during the monitoring phase and notes that the effect of inclusion of AEs is minimal and therefore not change has been made to the ERG base-case.

5.2.6.7 Mortality

As reported in section B.3.2.9.6 of the CS, mortality effects were accounted for in the economic model based on two different sources.[1] These were all-cause mortality risk, specific to age and gender, and an excess annual mortality for TRD, associated with suicide, of 0.47% linked to the MDE health state.[57] It was assumed that half the excess mortality risk associated with suicide would still be present in the response state.

Gender and age-specific all-cause mortality were sourced from the Office of National Statistics life tables.[58] The model firstly derived a weighted mortality risk for each age. This was weighted according to the proportion of males and females in the cohort and the baseline age. The risk was applied to the number of patients alive at the beginning of the cycle in each health state:

$$n_{death\ all-cause\ cycle\ i} = n_{alive\ cycle\ i} \times p_{age},$$

where:

- i is the cycle under consideration,
- $n_{death\ all-cause\ cycle\ i}$ is the number of patients that die during cycle i, due to all-cause mortality,
- $n_{alive\ cycle\ i}$ is the number of patients alive at the beginning of cycle i, and
- p_{age} is the mortality risk (i.e. probability) at a specific age.

Additional mortality from suicide attempts was also stated to have been explicitly modelled, which was performed in two steps. First, for patients in each health state, the number of suicide attempts was calculated, and second, a proportion of these suicide attempts were considered fatal, giving the total of patients who died from suicide. The calculation was as follows: risk was applied to the number of patients alive at the beginning of the cycle in each health state:

n_(death suicide cycle i)=n_(alive cycle i) ⟦× SA⟧_hs ⟦× p⟧_fatal,

where:

i is the cycle under consideration,

n_(death suicide cycle i) is the number of patients that die during cycle i due to suicide,

n_(alive cycle i) is the number of patients alive at the beginning of cycle i,

⟦SA⟧_hs is the risk of suicide attempt (i.e., probability) at the current health state, and

p_fatal is the risk of a suicide attempt being fatal.

It is unclear how this calculation would be performed given that the risk of suicide attempt was not reported.

ERG comment: The ERG considers the use of gender and age-specific all-cause mortality tables to be appropriate but has concerns that trial-based data were ignored in favour of the results of a published meta-regression.

The meta regression itself is based on analysis of 28 small interventional studies which focus on a range of different interventions (capsulotomy, cognitive behavioural therapy (CBT), deep brain stimulation (DBS), electroconvulsive therapy (ECT), epidural cortical stimulation (epCS), ketamine, vagal nerve stimulation (VNS) and treatment as usual (TAU)). No evidence was provided to suggest that this mix of interventions is representative of standard care in the UK and no justification was made for reliance on published meta regression over trial-based evidence.

Also, an examination of the model reveals that the method described in the CS is not the way that excess mortality was incorporated. In fact, it was simply by treating the 0.47% as a hazard ratio such that the excess was independent of risk of suicide. As the ERG pointed out in the clarification letter, this appears to be methodologically correct given that the excess was estimated conditional on being depressed rather than attempting suicide, although the company did not provide any further clarification for the method stated in the CS.[3, 18, 57]

However, the main problem is the assumption by the company that risk of mortality will decrease when treating with ESK-NS, given its differential risk of response and remission. This presumes that all of

the excess mortality is removed by moving from the MDE to the remission state and half of it on moving to the response state. This is contrary to evidence of three suicides in trials all of which, whilst considered unrelated to ESK-NS treatment, occurred in patients treated with esketamine, see section 4.2.7. Also, as acknowledged by the committee, no mortality effect was included in TA367.[9] Therefore, the ERG assumed no effect on mortality of ESK-NS + OAD in the ERG base-case, see section 7.2.

5.2.7 Health-related quality of life

5.2.7.1 Health-related quality-of-life data from clinical trials

EQ-5D-5L was used to measure the quality of life of patients in the TRANSFORM-2 trial from which utility values could be derived:

- Data were retrospectively mapped to EQ-5D-3L based on the UK valuation set,[59] as described in Section B.3.2.10.2 of the CS[1].

The company suggests that this represents NICE's preference as per the NICE reference case.

Further details of the methodology used to derive the utilities are presented in Section B.3.2.10.2 of the CS.

ERG comment: The use of HRQoL (and utility) data reported directly from patients is in line with the NICE reference case. Mapping of European Quality of Life-5 Dimensions – 3 levels (EQ-5D-3L) data from European Quality of Life-5 Dimensions – 5 levels (EQ-5D-5L) data is also in line with the NICE reference case. The precise method used by the company has been criticised in a report by the Decision Support Unit 2017[60] – but the latest NICE position statement on the use of EQ-5D-5L does suggest that the mapping function developed by van Hout et al.[59] should be used for reference-case analyses, for consistency with the current guide to the methods of technology appraisal.

5.2.7.2 Mapping

Individual scores from the five dimensions were used to obtain a weighted health status index using the method from van Hout and colleagues[59], described below:

- Scores from each dimension were combined to obtain a 5L profile score or health state: e.g. a score of 1 for each dimension gives a 5L profile score of 11111. Dimension scores were combined in the following order: Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression.
- Utilities for each possible profile on the EQ-5D-3L were computed using the Dolan algorithm which is specific to the UK [61].
- Patients were assigned probabilities for each possible profile on the EQ-5D-3L based on their profile on the EQ-5D-5L.
- The utility score on the EQ-5D-5L for each patient was computed as a weighted average of the utilities, where weights were the above-mentioned probabilities.

In the model, the utilities are stratified by health state. The health state QALYs at each cycle are calculated by multiplying the user-specified utility by the duration of the Markov cycle (28 days) expressed in years.

As noted above, disutility due to adverse events (AEs) was included as a scenario in the CS. For each AE included in the model, treatment-dependent inputs were used to calculate the associated utility decrements by treatment: the incidence for each AE by treatment, the duration of each event, and the

specific utility decrements of each event. The per-cycle utility decrement is calculated for all AEs and then summed to give a per-cycle AE-associated utility decrement for each treatment. This decrement is "added" to the utility only for patients on treatment during the acute phase; it is assumed that patients who are not on treatment do not experience any AEs. AEs associated with treatment are assessed only in the acute treatment phase and not in the maintenance phase, as it is assumed that patients are likely to have adapted well to the treatment by this time. The inclusion of AE-associated utility decrements is likely to be a conservative assumption, as the impact of AE on quality of life may already be captured in the utility analysis for the health states. In other words, the inclusion of AE-associated utility decrements may be double counting the impact of AEs on quality of life.

After the patient utilities (and disutilities in the scenario) were calculated, the values were aggregated across the health states for each cycle to obtain QALYs over time.

Utility scores were estimated for all the following health states in the Markov model using data from the TRANSFORM-2 study:

- Baseline/Major Depressive Episode (MDE)
- Response at four weeks/each cycle
- Remission at four weeks/each cycle
- Recovery after 36 weeks in remission

The baseline utility data were used to inform the utility score for patients in MDE.

Remission was defined as having a total MADRS score of 12 or less at week 4 (Day 28).

Response was defined as an improvement of 50% or more in total MADRS score at week 4 (day 28) compared with baseline. In the economic model the health states "remission" and "response" are mutually exclusive, meaning that patients in the health state "response" are patients who showed response, but did not reach remission.

The utility score for patients achieving recovery was assumed to be the same as the utility score for patients achieving remission at four weeks.

A set of descriptive summaries, i.e. mean, standard deviation [SD], standard error [SE], minimum, lower quartile [Q1], median, upper quartile [Q3], and maximum was computed for all the corresponding utility scores.

Utility scores were assumed to depend only on the health state of the patient, and not to be treatment-specific. Data from both treatment arms in the TRANSFORM-2 study were pooled to increase the robustness and precision of estimates. Analyses were based on observed data only and no imputation for missing data was performed. The estimates used to populate the utilities per health state in the economic model are summarised in Table 5.13.

Table 5.13: Summary of utilities used in the model (by health state)

Health State	Utility	Standard deviation	SE	Source
MDE (baseline value in TRANSFORM-2)	0.417	0.233	0.016	TRANSFORM-2
Response (value at day 28 in TRANSFORM-2)	0.764	0.123	0.020	TRANSFORM-2
Remission (assumption)	0.866	0.122	0.013	TRANSFORM-2

Health State	Utility	Standard deviation	SE	Source
Recovery (assumption)	0.866	0.122	0.013	Assumption*
Based on Table 54 of the CS[1] * Assumed to be the same as remission CS = company submission; MDE = major depressive episode; SE = standard error				

ERG comment: In its clarification letter, the ERG asked the company to explain why it did not consider using the DSU EQG (EuroQoL) method when mapping utilities from EQ-5D-5L.[18] The company responded that the method they used was consistent with the *"NICE position statement on this topic"*.[3] The ERG is satisfied with this explanation by the company.

The ERG notes that the company originally intended to use data from several trials (not just TRANSFORM-2) to generate utility values. *"Utility values for the model will be derived using the patient reported EQ5D administered during the 3 clinical trials. Other values to populate the model will be sourced from the literature. In the acute trials, the EQ5D will be administered at Days 1, 4, 8, 15, 22 of the double-blind phase as well as at the end of the study. In the maintenance trial, the EQ5D score will be collected on a monthly basis, as well as at the time of treatment discontinuation"*.[12] It is not clear why the company chose to ignore EQ5D data from the maintenance trial, SUSTAIN-1, to inform the utility values of the remission/recovery state. Page 6,348 of 11,938 of the clinical study reports for SUSTAIN-1 suggests a mean EQ-5D-5L follow up value of 0.842 (SD 0.1146). If this value could be converted into an EQ-5D-3L equivalent it could have been used in the economic model.

The ERG has been unable to validate the utility values used in the model (as set out in Table 5.13). Furthermore, comparison with a previous STA, vortioxetine for treating major depressive episodes (TA367) reveals considerable variation in baseline utility for populations with major depressive episodes, namely 0.417 in the CS and 0.54 in TA367.[1, 9] However, the ERG do not believe that there is a better source and therefore decided not to change baseline utility in the ERG base-case.

5.2.7.3 AE disutilities

Disutility due to dry mouth was obtained from Revicki et al. 1998.[62] The study reported utilities for patients in North America with MDD who had completed at least eight weeks of treatment. The disutility due to vision blurred was derived from Sullivan et al. 2006[63] which reported EQ-5D index scores for chronic conditions in the United States of America (USA), estimated from the nationally representative Medical Expenditure Panel Survey pooled from 2000–2002 with 38,678 adults.[63] Other disutilities listed in Table 5.14 were from the study by Sullivan et al. 2004[64] a cost effectiveness study of eight OADs used as initial treatment for depression in the US.

Since the AEs related to ESK-NS observed in the ESK-NS + OAD arm of TRANSFORM-2 were transient and resolved within hours, the scenario analysis conservatively assumed a duration of one day for all AEs.

Table 5.14: AE disutilities for scenario analysis

AE	Disutility (SE)
Anxiety	–0.129 (0.032)
Blood pressure increased	0.000 (0.000)
Delusional perception	0.000 (0.000)
Derealisation	0.000 (0.000)

AE	Disutility (SE)
Diarrhoea	–0.044 (0.011)
Dissociation	0.000 (0.000)
Dizziness	–0.085 (0.021)
Dizziness postural	0.000 (0.000)
Dry mouth	–0.010 (0.003)
Dysgeusia	0.000 (0.000)
Fatigue	–0.085 (0.021)
Feeling abnormal	–0.085 (0.021)
Feeling drunk	–0.085 (0.021)
Headache	–0.115 (0.029)
Hypoaesthesia	0.000 (0.000)
Hypoaesthesia oral	0.000 (0.000)
Illusion	–0.085 (0.021)
Insomnia	–0.129 (0.032)
Nasal discomfort	0.000 (0.000)
Nausea	–0.065 (0.016)
Paraesthesia	0.000 (0.000)
Paraesthesia oral	0.000 (0.000)
Somnolence	–0.085 (0.021)
Throat irritation	–0.010 (0.003)
Vertigo	–0.085 (0.021)
Vision blurred	–0.050 (0.012)
Vomiting	–0.065 (0.016)

Based on Table 55 of the CS[1]
AE = adverse event; CS = company submission; SE = standard error

ERG comment: The ERG regards the approach to estimating the values and handling of AE disutilities as reasonable.

5.2.8 Resources and costs

The cost categories included in the model were costs associated with treatment (drug acquisition costs including subsequent therapies, cost of supervision of self-administration and post-administration monitoring), costs associated with disease management (costs of OAD), and costs associated with different health states.

5.2.8.1 Resource use and costs data identified in the SLR

According to Appendix I of the CS, the SLR performed in July 2018 (with an update in April 2019) identified 19 studies that considered MDD, but only two specifically considered patients with TRD.[17] The company stated that one of the eligible UK studies was not aligned with the definitions used in this appraisal (no formal definition of resistance was provided) and the second study did not contain the data granularity required to inform the analysis, as the study did not report data per health state.

5.2.8.2 Treatment costs

The cost per a single-use device that delivers a total of 28 mg of esketamine in two sprays (one spray per nostril) is £163, equating to a cost of £326 per 56 mg dose and £489 per 84 mg dose. The average number of sessions per week and devices per session in the acute phase were estimated based on TRANSFORM-2 trial, while for subsequent time-points they were derived from SUSTAIN-1 trial. These numbers were also tested in sensitivity analysis and a plausibility limit was applied to prevent the number of ESK-NS devices being less than two (56 mg) or greater than three (84 mg). Similarly, limitation was applied to the number of sessions (no less than 0.5 and no more than 2).

The average treatment administration cost of esketamine was based on assumption that two nurses (one band 5 and one band 4) are needed for the supervision of self-administration of ESK-NS and that a cohort of six patients will be concurrently supervised. In the CS it was also mentioned that patients will be observed for 60 minutes on average and 9.57% of patients might experience a blood pressure increase which might prolong supervision of the patients. All these assumptions resulted in an average cost per patient, per administration of £30.08 (see Table 5.15).

Table 5.15: Administration and observation resource use and costs

Item	Resource use	Cost per hour	Total duration HCP is required (hours)	Number of patients in cohort	Average cost per session per patient
Administration/ preparation	1x band 4 nurse	£28	0.25	6	£30.08
	1x band 5 nurse	£37	0.25		
Supervision of self-administration	1x band 4 nurse	£28	1		
	1x band 5 nurse	£90*	1		
Monitoring post self-administration	1x band 5 nurse	£37	1.25		
Based on Table 56 of the CS[1] CS = company submission; HCP = healthcare professional					

A summary of drug acquisition and resource costs through all treatment phases in the model is presented in Table 5.16.

Table 5.16: Acquisition and resource costs associated with ESK-NS administration

Items	Acute Weeks 1–4	Continuation (relapse prevention) Weeks 5–8	Continuation (relapse prevention) Weeks 9–40	Maintenance (recurrence prevention) Week 41 onwards
Average number of sessions per week	1.850	0.992	0.711	0.675
Average number of devices per session	2.530	2.605	2.605	2.571
Drug acquisition cost per 4-week cycle	£3,051.61	£1,684.73	£1,208.42	£1,131.00

Items	Acute Weeks 1–4	Continuation (relapse prevention) Weeks 5–8	Continuation (relapse prevention) Weeks 9–40	Maintenance (recurrence prevention) Week 41 onwards
Administration and observation costs	£222.60	£119.33	£85.60	£81.17
Total cost per 4-week cycle	£3,274.21	£1,804.06	£1,294.02	£1,212.17

Based on Table 57 of the CS[1]
CS = company submission; ESK-NS = esketamine nasal spray

5.2.8.3 Comparator cost

All OADs with a market share greater than 3% of all treatments were included in the analysis. The average cost of OADS per four-week cycle was estimated using prescription cost analysis and market share information from IQVIA data (see Table 5.17).[11] For specific drugs (duloxetine, escitalopram, sertraline, and venlafaxine) the daily doses were derived from TRANSFORM-2 trial, while a mid-point of the plausible dose ranges was chosen for other OADs. The analysis resulted in weighted average cost of £4.15 per four-week cycle. Following response to clarification, the company adjusted weighted average cost and included the elderly population in the analysis of the cost. This resulted in a revised weighted average cost of £4.06. Since ESK-NS is incremental to OADs, the associated cost was equal on both sides.

Table 5.17: Weighted average OAD cost

OAD	Market share (%)	Daily dose (mg)	Average cost per mg	Average cost per 4-weeks
Amitriptyline	13.78	100.00 mg	£0.0029	£8.00
Citalopram	17.89	30.00 mg	£0.0031	£2.57
Duloxetine	5.40	59.00 mg	£0.0052	£8.54
Escitalopram	2.42	18.15 mg	£0.0050	£2.56
Fluoxetine	13.38	40.00 mg	£0.0026	£2.93
Mirtazapine	19.66	30.00 mg	£0.0027	£2.28
Sertraline	18.53	129.70 mg	£0.0005	£1.71
Venlafaxine	8.94	210.17 mg	£0.0017	£10.12
Weighted average cost per 4 weeks				£4.15

Based on Table 58 of the CS[1]
CS = company submission; OAD = oral antidepressant

5.2.8.4 Health state costs

Resource use in the MDE, relapse, recurrence, and recovery states were based on a retrospective chart review of medical records of patients with TRD, since TRANSFORM-2, SUSTAIN-1 and the published literature have not reported such information.

The retrospective chart review included data from 295 patients with TRD in the UK from both primary and secondary care. Data were collected from nine GPs and 30 psychiatrists and provided information on numbers of GP visits, psychiatrist visits, psychotherapies, psychiatric hospitalisations (general ward/psychiatric hospital), A&E visits, length of stay when hospitalised, antidepressant treatment

history (including dosing, duration, line of therapy, adherence), other psychiatric medications prescribed (anxiolytics, hypnotics, and antipsychotics), ECT, medical devices, AEs, management of AEs, and suicides. The full report was made available for the ERG on 28 August 2019.

Health resource use costs, excluding drug treatment costs, for four-week cycle are shown in Table 5.18. The costs for response and remission, according to the CS, were based on a conservative assumption which is biased against ESK-NS, as patients in the OAD arm spend greater time in the response state, and it might be expected that patients in response have greater healthcare resource use (HCRU) costs compared with patients in remission.

Table 5.18: List of health states and associated costs in the economic model

Health states	Value (95% CI)
MDE	£980.08 (761.48, 1,198.67)
Response	£164.46 (102.81, 226.11)
Remission	£164.46 (102.81, 226.11)
Recovery	£83.75 (47.97, 119.53)
Based on Table 59 of the CS[1] CI = confidence interval; CS = company submission; MDE = major depressive episode	

5.2.8.5 Adverse event related costs

Adverse events related cost were not included in the base-case analysis. The company justified this assumption based on TRANSFORM-2 trial, where most AEs were transient and resolved during the post-administration observation phase. Only the cost of a GP contact (at £37 per contact) for all ESK-NS-associated AEs was considered in a scenario analysis.

ERG comment:

a) The assumption applied in the model that six patients will be concurrently supervised during self-administration seems to be not realistic. The ERG asked the company to conduct an additional sensitivity analysis for average cost per session, where the number of patients in a clinic varies between plausible levels. In the response to clarification, the company agreed to assess the impact on the average administration cost per session per patient of varying the number of patients seen in a clinic at any one time. The sensitivity analysis resulted in the ICER of £6,420 when patient to nurse ratio was set to 20:1 and the ICER was £9,252, when patient to nurse ratio was set to 1:1. The ERG believes that latter scenario would be the most plausible in clinical practice and should be used in the ERG base-case (see section 7.2).
b) Although there will be no adjustment to OAD for the placebo effect in the ERG base-case (See section 5.2.6.1), the ERG consider that it is reasonable to attribute some of the effect on response and remission to be attributable to the extra clinic sessions. On this basis, the correct comparator might actually be OAD plus additional clinic sessions. Therefore, the cost of clinic sessions for OAD is increased to the level for ESK-NS + OAD in an ERG scenario (see section 7.3).
c) The company, in their submission, had mentioned that at visit 8 (four weeks) a psychiatrist is required to assess response according to the pathway given. However, this cost was not included in the economic model. In the response to clarification, the company argued that this cost would cancel out in each treatment arm, since all patients, irrespective to their initial treatment, would be assessed at week 4.[3] Therefore, inclusion of the cost would not impact the base-case ICER. After the explanation provided by the company, the ERG is satisfied with this assumption.

d) The ERG received the full study of the retrospective chart review on 28 August 2019. The study describes research methods in detail, however, the ERG noted that it is unclear how monthly costs were calculated, given that information was provided only about the sources of the cost. Indeed, instead of providing information about the time period on which all the calculations were based in each health state, the company stated that the data has been standardised to a 28-day period. This shortcoming introduces an uncertainty into the findings.
e) The company did not include any cost of adverse events in their base-case analysis, given that patients in TRANSFORM-2 trial experienced only transient AEs which resolved during the post-administration observation phase. The ERG believes that this assumption is reasonable, since the cost of post-administration observation phase is included in the model. However, the ERG thinks that latter assumption only partially covers the cost of AEs. In the TRANSFORM-2 trial and in the company's submission it was reported that around 90% of TEAEs were resolved on the same day. Therefore, the ERG believes that some AEs will occur after the observation phase, but notes that the effect of inclusion of AEs is minimal and therefore not change has been made to the ERG base-case.

6. Cost effectiveness results

6.1 Company's cost effectiveness results

The base-case clinical and economic outcomes are presented in Table 6.1. Over a five-year time horizon, ESK-NS + OAD was associated with an additional 0.336 QALYs compared with OAD. The incremental drug cost for ESK-NS + OAD was £10,456; ESK-NS + OAD was estimated to have lower disease management costs, saving £8,243 compared with OAD Table 6.1. This resulted in an incremental cost difference of £2,213 and therefore a base-case incremental cost effectiveness ratio (ICER) of £6,582 per QALY.

Table 6.1: Base-case results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER incremental (£/QALY)
OAD	£48,478	4.508	2.239				
ESK-NS + OAD	£50,691	4.519	2.575	£2,213	0.011	0.336	£6,582

Based on Table 62 of the CS[1]

CS = company submission; ESK-NS; ICER = incremental cost-effectiveness ratio; LYG = life years gained; OAD = oral antidepressant; QALYs = quality-adjusted life years

ERG comment: Given that the NICE scope has no upper age limit, the ERG requested that the company conduct a cost effectiveness analysis for the whole population by adding data specific for those aged 65 years over, including TRANSFORM-3 and SUSTAIN-2 to the existing data for those aged 18-64 years. In response to this request for clarification, the company submitted a model for the combined 18–64 years and ≥65 years populations. The model includes the derived weighted averages for transition probabilities for response and relapse in the acute phase, utilities, and cost inputs of the two populations. The same model assumptions as previously submitted in the base-case model are applied. Based on the 2011 Census of the Office of National Statistics, 20.8% of patients with TRD are ≥65 years.[58] With this input, the ICER was revised to £7,699 per QALY.

6.2 Company's sensitivity analyses

6.2.1 Probabilistic sensitivity analysis

To determine the uncertainty surrounding the base-case ICERs, a probabilistic sensitivity analysis (PSA) was conducted with a total of 10,000 Monte Carlo simulations. A Beta distribution was assigned to probabilities, proportions, and utility and disutility data which take values between 0 and 1, while a Gamma distribution was assigned to costs, doses, and resource use, which take positive values and are likely to be positively skewed. Uncertainty was characterised by standard error. Results of PSA are shown in Table 6.2.

Table 6.2: Probabilistic sensitivity analysis results

Technologies	Total costs (95% CI)	Total QALYs (95% CI)	Incremental costs	Incremental LYG (95% CI)	ICER incremental (£/QALY))
OAD	£48,493 (£38,548, £59,404)	2.24 (2.10 to 2.38)			

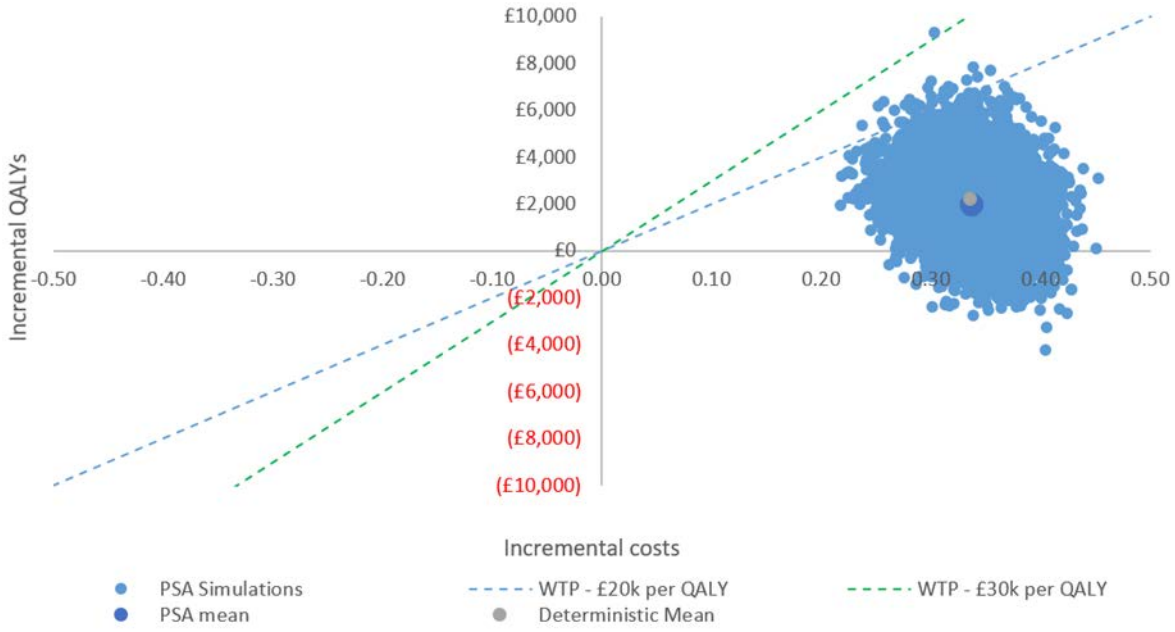
Technologies	Total costs (95% CI)	Total QALYs (95% CI)	Incremental costs	Incremental LYG (95% CI)	ICER incremental (£/QALY))
ESK-NS + OAD	£50,479 (£42,209, £59,389)	2.58 (2.43 to 2.72)	£1,987 (–£840, £4,822)	0.34 (0.27 to 0.40)	£5,903

Based on Table 63 of the CS[1]

CI = confidence interval; CS = company submission; ESK = esketamine; ICER = incremental cost-effectiveness ratio; LYG = life years gained; NS = nasal spray; OAD = oral antidepressant; QALYs = quality-adjusted life years

The incremental cost effectiveness plane and the corresponding cost effectiveness acceptability curves are shown in Figures 6.1 and 6.2.

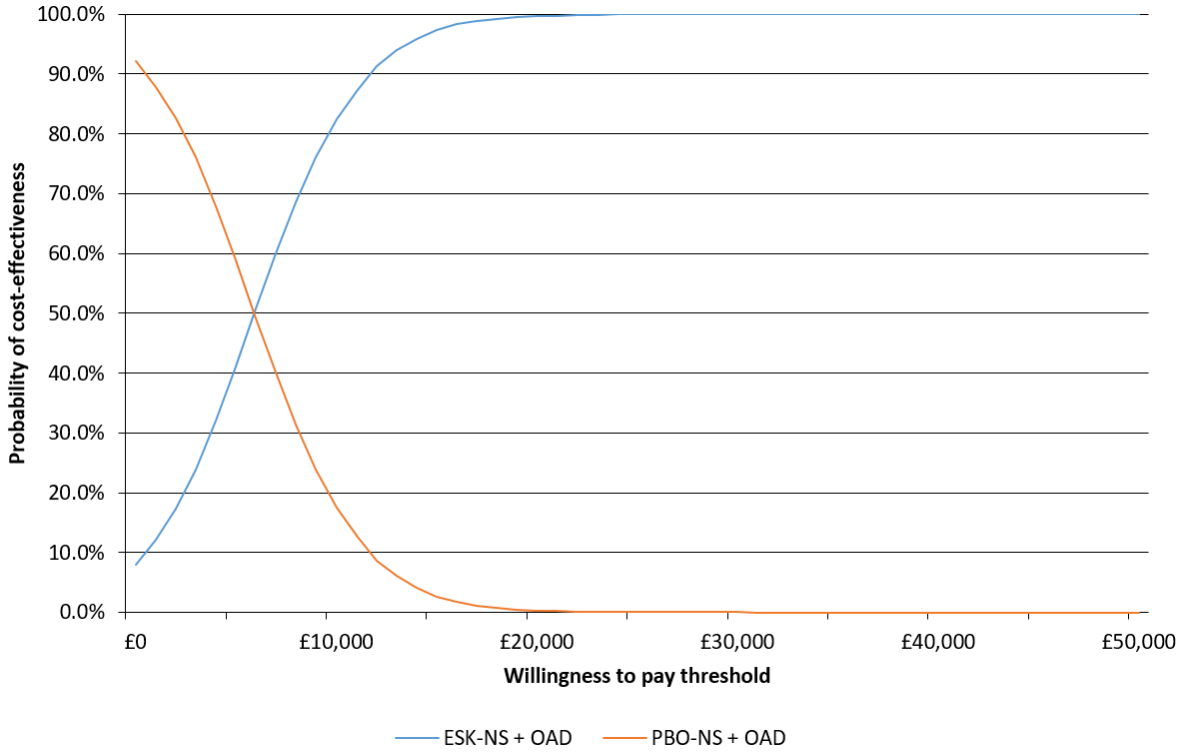
Figure 6.1: Cost effectiveness plane



Based on Figure 25 of the CS[1]

CS = company submission; PSA = probabilistic sensitivity analyses; QALY = quality-adjusted life year; WTP = willingness-to-pay

Figure 6.2: Cost effectiveness acceptability curve



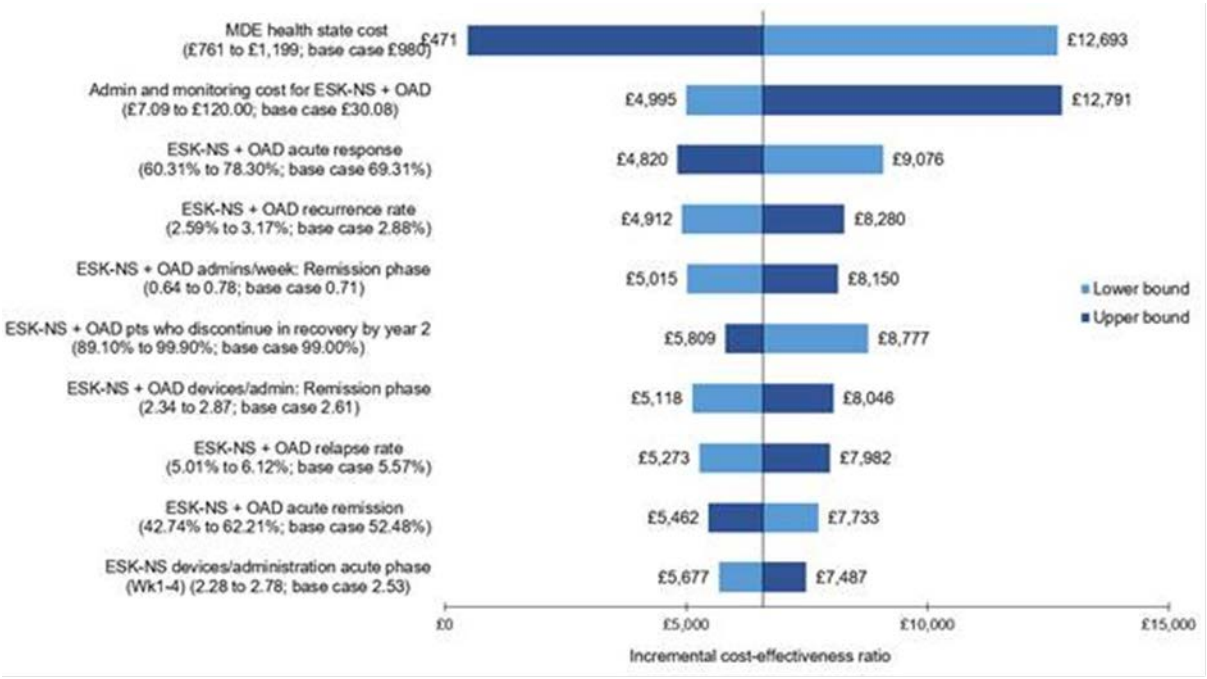
Based on Figure 25 of the CS[1]
ESK = esketamine; OAD = oral antidepressant; PBO = placebo

ERG comment: The PSA results are congruent with the deterministic analysis results and the most influential parameters (medical cost of the MDE state and the administration/observation cost associated with ESK-NS + OAD) seemed reasonable.

6.2.2 Deterministic sensitivity analysis

The results of the deterministic sensitivity analysis are shown below in the tornado diagram in Figure 6.3. All parameters were varied but the figure shows the 10 parameters with the greatest impact. Furthermore, no parameter tested in univariate sensitivity resulted in an ICER above £20,000 per QALY.

Figure 6.3: Results of univariate sensitivity analysis (tornado diagram)



Based on Figure 27 of the CS[1]

CS = company submission; ESK = esketamine; MDE = major depressive disorder; NS = nasal spray; OAD = oral antidepressant

Table 6.3: Results of univariate analysis

Variable (lower bound to upper bound; base-case value)	ICER with lower bound	ICER with upper bound
MDE health state cost (£761 to £1,199; base-case £980)	£12,693	£471
Admin and monitoring cost for ESK-NS + OAD (£7.09 to £120.00; base-case £30.08)	£4,995	£12,791
ESK-NS + OAD acute response (60.31% to 78.30%; base-case 69.31%)	£9,076	£4,820
ESK-NS + OAD recurrence rate (2.59% to 3.17%; base-case 2.88%)	£4,912	£8,280
ESK-NS + OAD administrations/week continuation phase (0.64 to 0.78; base-case 0.71)	£5,015	£8,150
ESK-NS + OAD pts who discontinue in recovery by Year 2 (89.10% to 99.90%; base-case 99.00%)	£8,777	£5,809
ESK-NS + OAD devices/administration during continuation phase (2.34 to 2.87; base-case 2.61)	£5,118	£8,046
ESK-NS + OAD relapse rate (5.01% to 6.12%; base-case 5.57%)	£5,273	£7,982
ESK-NS + OAD acute remission (42.74% to 62.21%; base-case 52.48%)	£7,733	£5,462
ESK-NS devices/administration acute phase (Wk1-4) (2.28 to 2.78; base-case 2.53)	£5,677	£7,487

Based on Table 64 of the CS[1]

CS = company submission; ESK = esketamine; ICER = incremental cost-effectiveness ratio; MDE = major depressive disorder; NS = nasal spray; OAD = oral antidepressant

6.2.3 Scenario analyses

6.2.3.1 Treatment effect adjustment

The effect of removing the adjustment for the placebo effect, consistent with the values of remission and relapse for OAD of 31.0% and 52.0% as opposed to the adjusted values of 18.0% and 34.0%, was to increase the ICER to £16,209. The company also performed analyses of combinations of various percentages of the unadjusted response and remission.

ERG comment: As explained in section 5.2.6.1, the ERG believes that the adjustment should not be made to the placebo arm of the TRANSFORM-2 trial and there is no basis for any given percentage reduction in either response of remission. This view is reflected in the ERG base-case, see section 7.2. The ERG considered sensitivity analyses to be appropriate.

6.2.3.2 Other comparators

Based on the data from the NMA, and using the data from TRANSFORM-2 adjusted for the placebo effect, as presented in section 5.2.4, the company reported the following results (Table 6.4) for a set of comparators other than those in the company trials.

In response to the request for clarification, the company also presented the results base on the NMA estimates unadjusted for the placebo effect (see section 5.2.4), which are show in Table 6.5.[3]

Table 6.4: Scenario analysis considering all comparators, adjusted for placebo effect

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline* (£/QALY)	ICER incremental (£/QALY)	ICER versus ESK-NS + OAD (£/QALY)
Aug SSRI/SNRI + AAP	£48,059	4.5089	2.2597						£8,344
Aug tricyclic (nortrip) ± PBO	£48,634	4.5081	2.2358	£576	-0.0008	-0.0240	Dominated	Dominated	£6,058
Aug SSRI/SNRI + lithium	£48,837	4.5078	2.2268	£203	-0.0003	-0.0090	Dominated	Dominated	£5,320
OAD + PBO	£49,250	4.5072	2.2090	£413	-0.0006	-0.0177	Dominated	Dominated	£3,934
Aug SSRI/SNRI ± PBO	£49,580	4.5067	2.1958	£329	-0.0004	-0.0132	Dominated	Dominated	£2,929
Switch tetracyclic (mirtazapine)	£49,865	4.5063	2.1834	£285	-0.0004	-0.0124	Dominated	Dominated	£2,108
ESK + AD	£50,691	4.5188	2.5751	£826	0.0125	0.3917	£8,344	£2,108	

Based on Table 81 of the CS[1]

* Baseline in this analysis is Aug SSRI/SNRI + AAP.

AAP = atypical antipsychotic; Aug = augmentation; ESK-NS = esketamine nasal spray; CS = company submission; ICER = incremental cost-effectiveness ratio; LYG = life years gained; nortrip = nortriptyline; OAD = oral antidepressant; PBO-NS = placebo nasal spray; PBO = placebo; QALYs = quality-adjusted life years; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

Table 6.5: Scenario analysis considering all comparators, unadjusted for placebo effect

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER versus baseline* (£/QALY)	ICER incremental (£/QALY)	ICER versus ESK-NS + OAD (£/QALY)
Aug SSRI/SNRI + AAP	£45,709	4.5121	2.3569						£22,823
Aug tricyclic (nortrip) ± PBO	£46,445	4.5111	2.3261	£737	-0.0010	-0.0307	Dominated	Dominated	£17,049
Aug SSRI/SNRI + lithium	£46,804	4.5106	2.3105	£359	-0.0005	-0.0156	Dominated	Dominated	£14,686
OAD + PBO	£47,327	4.5098	2.2877	£523	-0.0008	-0.0228	Dominated	Dominated	£11,701
Aug SSRI/SNRI ± PBO	£47,870	4.5091	2.2661	£543	-0.0007	-0.0216	Dominated	Dominated	£9,124
Switch tetracyclic (mirtazapine)	£48,287	4.5085	2.2477	£416	-0.0006	-0.0184	Dominated	Dominated	£7,341
ESK + AD	£50,691	4.5188	2.5751	£2,404	0.0103	0.3274	£22,823	£7,341	

Based on Table 15 of the response to request for clarification[3]

* Baseline in this analysis is Aug SSRI/SNRI + AAP.

AAP = atypical antipsychotic; Aug = augmentation; ESK-NS = esketamine nasal spray; CS = company submission; ICER = incremental cost-effectiveness ratio; LYG = life years gained; nortrip = nortriptyline; OAD = oral antidepressant; PBO-NS = placebo nasal spray; PBO = placebo; QALYs = quality-adjusted life years; SNRI = serotonin–norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor

ERG comment: The ERG acknowledges that there are significant limitations to the NMA, as discussed in section 4.4. Also, even when the adjustment is removed, most comparators are dominated: only Aug SSRI/SNRI + AAP is not because it is the least costly. However, the ICER vs. Aug SSRI/SNRI + AAP is higher than the ICER vs. OAD and there is an unexplained discrepancy between the cost, life years and QALYs associated with OAD when OAD is the only comparator and when it is one of several. Nevertheless, the ERG has made the judgment not to include Aug SSRI/SNRI + AAP in any ERG analyses.

6.3 Model validation and face validity check

In Section B.3.6, it was stated that two independent senior health economic modellers, external to the model process, performed quality assurance, which entailed:[1]

- Review of modelling structural assumption and techniques chosen.
- Review of technical deployment (formulas, functionality).
- Review of data inputs and sources.
- Conducting extreme scenario analyses and validation of results.

The first review was conducted in 2018 and the second in 2019.

Two global advisory boards (in July 2017 and November 2018 (no citation in CS)) and two UK HTA advisory boards (in October 2018 and June 2019 (no citation in CS)) were also held to inform the development of the model.[1] The NICE Preliminary Independent Model Advice (PRIMA) also produced a report to advise the company.[12] The executive summary did identify that, although the model captures key clinical features for the short-term management of TRD it is less successful in reflecting the highly variable longer-term outcomes of people with TRD. This included the effect on retreatment, lack of incorporation of the 'negative effects of treatment discontinuation', and the extrapolation of the treatment effect. Indeed, the PRIMA report states that *"…the health economics expert was concerned that treatment discontinuation only serves to reduce healthcare costs"* (p. 17).[12] Another problem identified was the number of sessions for ESK being so much fewer than for OAD during the recovery phase, consistent with the simultaneous treatment of six patients (0.17 sessions per person).

ERG comment: It is commendable that the model has been checked and validated. However, some of the issues identified by PRIMA remain.[12] In particular, the long-term effect of retreatment has not been incorporated. However, the ERG considers that this could be considered to be outside the scope as the population would then be at a different line of therapy. What is of more concern is the continued lack of a negative effect of discontinuation, at least for reasons other than loss of efficacy. The company provided no data to show that those who discontinued treatment, even censoring for relapse, would not demonstrate any diminution in quality of life. The CS also provided no data to support the assumption that 35.4% of patients in the recovery phase would immediately discontinue with no loss of quality of life. The ERG also believes that the simultaneous monitoring of six patients that continues to be assumed is probably not feasible in clinical practice.

7. Evidence Review Group's additional analyses

7.1 Exploratory and sensitivity analyses undertaken by the ERG

Based on all considerations discussed in section 5.2 of this report, the ERG defined a new base-case and constructed three additional scenarios on this base-case. These scenarios included multiple adjustments to the company base-case submitted with the clarification response in order to include data that is suitable for adults of any age, i.e. 'ID1414 esketamine CEM adults and elderly GB 13082019 (ACIC)'. These adjustments are subdivided into three categories (derived from Kaltenthaler 2016[65]):

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the ERG considers that reasonable alternative assumptions are preferred)

Fixing errors

None identified.

Fixing violations

None identified.

Matters of judgment

1) Time horizon 20 years
2) No adjustment for placebo effect to OAD Acute response or remission transition probabilities
3) No discontinuation for reasons other than loss of efficacy
4) No effect on mortality of ESK-NS + OAD
5) Cost of clinic visit for ESK-NS + OAD based on patient to nurse ratio of 1:1
6) No difference between ESK-NS + OAD and OAD in the loss of response and relapse transition probabilities
7) A decrease in response and remission was applied at each line of subsequent therapy (including BSC) by multiplying the values for OAD by a factor equal to the ratio of values in Step 3 versus Step 4 in STAR*D.[53] These ratios are: 13.7/13.0 and 16.8/16.3 for remission and response respectively. The ERG used the same method of adjusting by line for loss of response and relapse in this ERG scenario. This was achieved by using the company estimated values, for loss of response, of 22.2% for first-line TRD and 22.8% for second-line TRD and, for relapse, of 6.8% for first-line TRD and 12.8% for second-line TRD.[3]
8) Cost of clinic visit for OAD set equal to that for ESK-NS + OAD

Issues (1) to (5) are all incorporated as the ERG's preferred model assumptions and thus form the ERG base-case, the results for which are shown in Tables 7.1 and 7.2. Table 7.3 shows how the individual adjustments of (6), (7) and (8) impact additionally as scenarios on the ERG base-case.

7.2 ERG's base-case analysis

Table 7.1: ERG's base-case analysis (deterministic)

	Total costs (£)	Total LYGs	Total QALYs	Incr. costs (£)	Incr. LYGs	Incr. QALYs	ICER versus baseline (£/QALY)
OAD	145,153.86	13.278	6.678	15,298	0.000	0.246	£62,078
ESK-NS + OAD	160,452.22	13.278	6.925				

ERG = Evidence Review Group; ESK = esketamine; ICER = incremental cost effectiveness ratio, Incr. = incremental, LYG = life year gained, NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year

Table 7.2: ERG's base-case: cumulative effect of each assumption

Preferred assumption		Section in ERG report	Cumulative ICER £/QALY
	Company base-case using 'adults and elderly' model		£7,699
1	Time horizon 20 years	5.2.5	£ 4,627
2	No adjustment for placebo effect to OAD Acute response or remission transition probabilities	5.2.6.1	£ 12,557
3	No discontinuation for reasons other than loss of efficacy	5.2.6.3	£ 52,872
4	No effect on mortality of ESK-NS + OAD	5.2.6.7	£ 55,027
5	Cost of clinic visit for ESK-NS + OAD based on patient to nurse ratio of 1:1	5.2.8	£ 62,078

ERG = Evidence Review Group; ESK = esketamine; ICER = incremental cost effectiveness ratio, Incr. = incremental, NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year

Table 7.3: ERG's base-case analysis (probabilistic, LYs not generated)

	Total costs (£)	Total QALYs	Incr. costs (£)	Incr. QALYs	ICER versus baseline (£/QALY)
OAD	£145,471.41	6.682	£15,367	0.247	£62,141
ESK-NS + OAD	£160,838.28	6.929			

ERG = Evidence Review Group; ESK = esketamine; ICER = incremental cost effectiveness ratio, Incr. = incremental, NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year

7.3 ERG's additional analyses

Table 7.4: ERG scenario analyses

ERG assumption		Section in ERG report	ICER £/QALY
5	ERG's base-case using 'adults and elderly' model	7.2	£ 62,078
6	No difference between ESK-NS + OAD and OAD in the loss of response and relapse transition probabilities	5.2.6.2	£97,396

ERG assumption		Section in ERG report	ICER £/QALY
7	A decrease in response and remission was applied at each line of subsequent therapy (including BSC) by multiplying the values for OAD by a factor equal to the ratio of values in Step 3 versus Step 4 in STAR*D.[53] These ratios are: 13.7/13.0 and 16.8/16.3 for remission and response respectively. Values estimated by the company from STAR*D were, for loss response, 22.2% for first line TRD and 22.8% for second line TRD and, for relapse, of 6.8% for first line TRD and 12.8% for second line TRD.[3]	5.2.6.4, 5.2.6.5	£ 148,376
8	Cost of clinic visits for OAD set equal to that for ESK-NS + OAD	5.2.8	£ 53,728
ERG = Evidence Review Group; ESK = esketamine; NS = nasal spray; OAD = oral antidepressant; QALY = quality-adjusted life year; TRD = treatment-resistant depression			

7.4 Conclusions of the cost effectiveness section

The company model is a state transition model with a cycle length of four weeks and, in addition to death, four health states, which are summarised in Table 5.5.[1] Patients enter the model in the major depressive episode (MDE) health state, after having failed to achieve a *"…clinically meaningful improvement…"* (page 160, CS) after treatment with at least two OADs *"prescribed in adequate dosages for adequate time"* (page 160, CS).[1] During each four-weekly Markov cycle, patients can occupy MDE, response, remission, recovery or death health states. Transition to recovery can only occur from remission and only after nine months (36 weeks) in the remission state and then with certainty. Patients can cycle through up to three subsequent treatments. After three subsequent treatments, patients enter the MDE state from which they can still respond or go into remission, whilst being treated with BSC. The company stated that this structure had been validated by NICE PRIMA, which indicated that it was an improvement on the last NICE appraisal for depression, i.e. TA367.[9, 12] Transitions between health states are governed by treatment phase:

- Acute phase (weeks 1 to 4)
- Continuation phase (weeks 5 to 40)
- Maintenance phase (weeks 41+)

The population was described in the CS as adults with TRD with a moderate to severe depressive episode.[1] A moderate to severe episode of TRD was assumed to have minimum duration of two years. Treatment-resistant MDD was defined as non-response to two or more OADs prescribed at an adequate dose and for an adequate duration in the current episode.

The intervention in the analysis was ESK-NS co-administered with a newly initiated OAD (ESK-NS + OAD), see Section B.3.2.7 of the CS).[1] The average number of sessions per week and devices per session in the acute phase were derived from TRANSFORM-2, while for subsequent time-points they were derived from SUSTAIN-1. In TRANSFORM-2 the precise rules of determining efficacy and tolerability were not reported in the CS or Appendix M.[1, 17] In TRANSFORM-3, the starting dose was 28 mg which could also be increased to 84 mg by day 25 without any specification of the precise rules. SUSTAIN-1 had the same dosing as TRANSFORM-2 in the first four weeks for direct entry patients. These patients then joined those who had been transferred from TRANSFORM-1 and TRANSFORM-2 to enter the optimisation phase where the dose could be adjusted at either week 8 or 12. Dosing was then determined according to a complex set of rules, whereby effectiveness measured in a variety of ways that depended on number of weeks on treatment determined whether treatment was administered

weekly of fortnightly. Neither the concomitant OAD nor the comparator OAD were specified in the CEA: instead it was as a mix of eight OADs, according to market share. The company did perform a scenario analysis (see section B.3.4.4.9 of the CS) based on an NMA using data from TRANSFORM-3 of response and remission presented in Appendix D, which compared ESK-NS + OAD with various other comparators in the form of drug classes.[1, 17] For all other parameters, equivalence with OAD was assumed given that these parameters were estimated from STAR*D and the company stated that this study included OAD and other augmentation strategies in 1st and 2nd line TRD.

As stated in Section B.3.2.4 of the CS, the base-case time horizon was five years the analysis took the perspective of the NHS and PSS in England. Both costs and outcomes (life years and QALYs) were discounted at 3.5%.[1]

In terms of the effectiveness of ESK-NS + OAD versus OAD in the acute phase, response and remission values were estimated from TRANSFORM-2 with the adjustment then applied to the OAD + PBP-NS arm only. The adjustment is a reduction in the rates of response and remission estimated as the effect of a reduction in the number of clinic visits from eight in the trial to two in clinical practice, i.e. a reduction of six. In the continuation phase, the transition probabilities of response to remission were estimated by analysis of the SUSTAIN-1 data. The transition probabilities for loss of response (response to MDE) and relapse (remission to MDE in weeks 5 to 40) were stated to have been estimated from SUSTAIN-1 for ESK-NS + OAD and from STAR*D for OAD.[1] In the maintenance phase, the transition probabilities for recurrence (remission to MDE in weeks 41+) for both, ESK-NS + OAD and OAD, the data pooled from both study arms of the double-blind phase of SUSTAIN-1 was used. The effect of discontinuation for reasons other than loss of efficacy (not loss of response, relapse or recurrence) was to stop incurring the cost of ESK-NS and only incur the cost of OAD and to have no effect on QALYs because patients were assumed to remain in the remission or recovery state until loss of response, relapse or recurrence. It was assumed that patients would not discontinue OAD in any phase for any reason other than lack of response. In the acute phase, it was assumed that patients would not discontinue ESK-NS + OAD in the acute phase for any reason other than lack of response. In the continuation phase, discontinuation risk for other reasons was derived from SUSTAIN-1. In the maintenance phase, it was also assumed that 35.4% of patients were assumed to stop ESK-NS immediately upon achieving recovery, i.e. on being in the remission state after 40 weeks of treatment. This was the percentage of patients in SUSTAIN-1 who had ≤2 total number of MDD episodes, including the current episode.[56] For those patients who did not discontinue immediately, a four-week discontinuation risk of 25% for ESK-NS + OAD was stated to have been used during recovery. The company estimated the transition probabilities for each of three further lines of subsequent treatment based on evidence from STAR*D. This was also stated to be one source for the estimation, using clinical expert opinion, of the transition probabilities for the best supportive care treatment mix, i.e. for patients whose disease has failed all previous treatments (fifth-line TRD and onwards). Mortality effects were accounted for in the economic model based on two different sources. These were all-cause mortality risk, specific to age and gender, and an excess annual mortality for TRD, associated with suicide, of 0.47% linked to the MDE health state. It was assumed that half the excess mortality risk associated with suicide would still be present in the response state.

EQ-5D-5L was used to measure the quality of life of patients in the TRANSFORM-2 trial from which utility values were derived, one for each health state, MDE, response, remission and recovery, with the

latter two being assumed to be the same. Disutilities due to adverse events (AEs) were included as a scenario and values were obtained from a variety of sources.

The cost categories included in the model were costs associated with treatment (drug acquisition costs including subsequent therapies, cost of supervision of self-administration and post-administration monitoring), costs associated with disease management (costs of OAD), and costs associated with different health states. The cost per a single-use device that delivers a total of 28 mg of esketamine in two sprays (one spray per nostril) is £163, equating to a cost of £326 per 56 mg dose and £489 per 84 mg dose. The average number of sessions per week and devices per session in the acute phase were estimated based on TRANSFORM-2 trial, while for subsequent time-points they were derived from SUSTAIN-1 trial. The average treatment administration cost of esketamine was based on assumption that two nurses (one band 5 and one band 4) are needed for the supervision of self-administration of ESK-NS and that cohort of six patients will be concurrently supervised. All OADs with a market share greater than 3% of all treatments were included in the analysis. The average cost of OADS per four-week cycle was estimated using prescription cost analysis (page 170 in the CS) and market share information from IQVIA data (page 2 in the CS).[11] For specific drugs (duloxetine, escitalopram, sertraline, and venlafaxine) the daily doses were derived from TRANSFORM-2 trial, while a mid-point of the plausible dose ranges was chosen for other OADs. The analysis resulted in weighted average cost of £4.15 per four-week cycle. Since ESK-NS is incremental to OADs, the associated cost was equal on both sides. Resource use in the MDE, relapse, recurrence, and recovery states were based on a retrospective chart review of medical records of patients with TRD. Only cost of a GP contact (at £37 per contact) for all ESK-NS-associated AEs was considered in a scenario analysis.

Over a five-year time horizon, ESK-NS + OAD was associated with an additional 0.336 QALYs compared with OAD. The incremental drug cost for ESK-NS + OAD was £10,456; ESK-NS + OAD was estimated to have lower disease management costs, saving £8,243 compared with OAD. This resulted in an incremental cost difference of £2,213 and therefore a base-case incremental cost effectiveness ratio (ICER) of £6,582 per QALY. The effect of removing the adjustment for the placebo effect, consistent with the values of remission and relapse for OAD of 31.0% and 52.0% as opposed to the adjusted values of 18.0% and 34.0%, was to increase the ICER to £16,209. The company also performed analyses of combinations of various percentages of the unadjusted response and remission. Based on the data from the NMA, and using the data from TRANSFORM-2 adjusted for the placebo effect, the company conducted a scenario analysis for a set of comparators other than those in the company trials. In response to the request for clarification the company also presented the results base on the NMA estimates unadjusted for the placebo effect.

The ERG believes that the model structure seems plausible and responds appropriately to the critique in TA367.[9]

The population is broadly consistent with the NICE scope and the expected marketing authorisation.[16, 66] However, the maximum age in the trials (TRANSFORM-2 and SUSTAIN-1) used to inform the CEA, which is 64 years.[1] The ERG questions the applicability of TRANSFORM-2 to the age 65 years+ age group. It is also therefore questionable what the applicability of SUSTAIN-1 would be to the 65 years+ age group: unfortunately, there is no equivalent study in the older age group by which a comparison might be made. SUSTAIN-2 included older patients, but relapse was not measured and no separate subgroup analysis was provided.[1, 17] Therefore, given that the NICE scope has no upper age limit, in the clarification letter the ERG requested that the main cost effectiveness analysis (CEA), i.e. for age <65 years, informed by TRANSFORM-2 and SUSTAIN-1 be combined with that for age 65 years+, using TRANSFORM-3 as well as SUSTAIN-2. The company responded by submitting a new version

of the base-case model to include acute response and remission transition probabilities and utilities for MDE, response and remission/recovery states from both TRANSFORM-2 and TRANSFORM-3, weighted by % in each age group such that if set to 0% for age >65 years one gets the same result as in the original base-case.[3] This forms the starting point for the ERG base-case.

In terms of the intervention, the lack of clarity on dosing in TRANSFORM-2 and TRANSFORM-3 trials plus the complex dose changes in SUSTAIN-1 and SUSTAIN-2 mean that it is difficult to know how applicable to clinical practice the transition probabilities estimated from the trials would be. This basis for questionable applicability is in addition to that in terms of whether the data to inform those transition probabilities derived from patients were direct-entry or transferred-entry. In terms of the comparators, the ERG is convinced that the limitations of the NMA are sufficient to exclude those included comparators except in a scenario analysis. However, applicability to clinical practice of results would be highest in those patients who might be switched to one of the four OADs prescribed in the trials.

The ERG also notes that by 20 years the percentages of the cohort in the response, remission or recovery health states in the cohort treated with ESK-NS + OAD are equal to those in the cohort treated with OAD + PBO-NS. Therefore, from this point onwards there can be no further difference in cost or QALYs and thus no need to extend the time horizon beyond this point. The ERG therefore has adopted 20 years in the ERG base-case.

With regards to the effectiveness of ESK-NS +OAD versus OAD, the ERG argue that, whilst it might be the case that some of the placebo effect, however mediated, might continue into clinical practice, it is possible to reproduce it by increasing clinic visits even without esketamine and, on that basis, it is impossible to have confidence as to the size of any effect that might only apply to esketamine in clinical practice. Accordingly, the ERG base-case removes this adjustment and assumes an increase the cost of clinic visits for OAD to be identical to the monitoring cost of OAD in a scenario analysis. However, it remains unclear to the ERG how data were chosen from SUSTAIN-1 in order to estimate the transition probability of response to remission given that patients appear to enter SUSTAIN-1 from various sources, including either of the TRANSFORM-1 or TRANSFORM-2 or by direct entry. It also remains unclear to the ERG why STAR*D was chosen given that at least some patients who entered SUSTAIN-1 were originally randomised to OAD + PBO-NS in TRANSFORM-1 or TRANSFORM-2. In line with TA367 and given the absence of any comparative trial evidence, the ERG assumed there to be no difference in the loss of response and relapse transition probabilities in an ERG scenario.[9] The ERG considers that it is reasonable to assume no discontinuation during the acute phase and the rate during the continuation phase also appears to be reasonable given that it was estimated from the trial data albeit based on an arbitrary definition of stable and choice of exponential distribution. However, the rates of discontinuation in the maintenance phase were not based on any observed data, but instead on assumptions, despite the availability of SUSTAIN-1 data, which could have had a parametric curve fitted to extrapolate up to the time horizon. It is also not reasonable to assume that the treatment effect is maintained, i.e. no decrease in QALYs on discontinuing ESK-NS and continuing with only OAD. Although some continuation of effect post-discontinuation is not impossible, in the absence of any data as to the effect on relapse or recurrence or utility on discontinuation of ESK-NS, the ERG assumed no discontinuation for reasons other than loss of efficacy in the ERG base-case. To estimate the transition probabilities of subsequent therapy, although the company stated that they used STAR*D, their methods were unclear and the resulting values were much lower than those in STAR*D.[53] Given that the values from STAR*D were stated to have been adjusted to a four-weekly risk and that the model did not allow transition from MDE to response or remission over more than one cycle, the full effectiveness must have been underestimated. Given that patients were encouraged to switch treatment, if no response and

follow-up visits occurred every two months, the ERG believe it reasonable to assume that all response or remission occurred in the first cycle on starting treatment in an ERG scenario.[53] Therefore, for the ERG scenario, as also recommended by the committee of TA367, a decrease in response and remission was applied at each line of therapy by multiplying the values for OAD by a factor equal to the ratio of values in step 3 versus step 4 in STAR*D.[9, 53] The ERG used the same method of adjusting by line for loss of response and relapse in this ERG scenario. This was achieved by using the company estimated values, for loss of response, of 22.2% for first line TRD and 22.8% for second line TRD and, for relapse, of 6.8% for first line TRD and 12.8% for second line TRD.[3] Logically, the same method was used for BSC, given that BSC (fifth-line TRD) appears to be the same OAD mix as for all other lines of therapy. ERG considers that most of the effect to AEs will be during the monitoring phase and notes that the effect of inclusion of AEs is minimal and therefore not change has been made to the ERG base-case. The ERG noted that the implementation of MDE-associated mortality in the model differed from that described in the CS. However, the main problem with the estimation of mortality is the assumption by the company that risk of mortality will decrease with by treating with ESK-NS, given its differential risk of response and remission. This presumes that all of the excess mortality is removed by moving from the MDE to the remission state and half of it on moving to the response state. This also needs to be considered in the context of only three suicides in all the esketamine trials, which whilst they were considered unrelated to ESK-NS treatment, all occurred in patients treated with esketamine, see section 4.2.7). Also, as acknowledged by the committee, no mortality effect was included in TA367.[9] Therefore, the ERG assumed no effect on mortality of ESK-NS + OAD in the ERG base-case.

Although the ERG has been unable to validate the utility values used in the model, they do not believe that there is a better source and therefore there is no change to utility in the ERG base-case.

The assumption applied in the model that six patients will be concurrently supervised during self-administration seems to be not realistic. Therefore, the patient to nurse ratio is set to 1:1 in the ERG base-case. Although there will be no adjustment to OAD for the placebo effect in the ERG base-case (see section 5.2.6.1), the ERG consider that it is reasonable to attribute some of the effect on response and remission to be attributable to the extra clinic sessions. On this basis, the cost of clinic sessions for OAD is increased to the level for ESK-NS + OAD in an ERG scenario.

The result of the adjustments to the company base-case produced an ERG base-case with an ICER that was considerable higher than the company base-case, i.e. £62,078 instead of £7,699 (revised company base-case). It is important to note that this was a reflection of relatively conservative assumptions by the ERG regarding the treatment effect (difference between ESK-NS + OAD and OAD) at every stage in the model. Indeed, scenario analysis revealed that the ICER could be much higher, i.e. £148,650. Also, these assumptions were deemed, by the ERG, to be based on sound methodology, such as the removal of the placebo effect adjustment to reduce only the effectiveness of OAD in the acute phase. They were also in line with the scientific advice from NICE PRIMA, such as the removal of ESK-NS discontinuation that implied only a reduction in cost with no loss of effectiveness.[12] The scenarios were also aligned with opinion of the TA367 committee informed by clinical expert opinion, such as the removal of the treatment effect on relapse and loss of response and the increase in the effectiveness of subsequent therapies.[9]

In conclusion, the approach taken to form the ERG base-case and scenarios contrast very strongly with the assumptions made in the CS, which, at every stage, enhanced the treatment effect on the basis of no or very little comparative evidence and rather opaque exposition. In particular, no data were provided to support the lack of impact on effectiveness of discontinuing ESK and all of the evidence to inform the company base case came from differential data sources for the intervention and the comparator

beyond the acute phase. Despite a request for clarification, it remains unclear why more data from the SUSTAIN studies could not have been used to inform the relapse and loss of response rates for OAD. Finally, the method of estimating all transition probabilities beyond the acute phase is unclear, both the precise data used from SUSTAIN-1 to inform those for ESK-NS + OAD and the calculations used to transform the data from STAR*D to inform those for OAD.

8. End of life

According to section B.2.13.3 of the CS, this is not applicable.[1]

9. References

[1] Janssen. *Esketamine for treatment-resistant depression [ID1414]. Document B: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA)*: Janssen, 2019 [accessed 11.7.19]. 237p.

[2] National Institute for Health and Care Excellence. *Patient organisation submission (SANE): Esketamine for treatment-resistant depression [ID1414]* [accessed 9.7.19]

[3] Janssen. *Esketamine for treatment-resistant depression [ID1414]: Response to request for clarification from the ERG*, 2019 [accessed 15.8.19]. 178p.

[4] Fonagy P, Rost F, Carlyle JA, McPherson S, Thomas R, Pasco Fearon RM, et al. Pragmatic randomized controlled trial of long-term psychoanalytic psychotherapy for treatment-resistant depression: the Tavistock Adult Depression Study (TADS). *World Psychiatry* 2015;14(3):312-21.

[5] Feldman RL, Dunner DL, Muller JS, Stone DA. Medicare patient experience with vagus nerve stimulation for treatment-resistant depression. *J Med Econ* 2013;16(1):62-74.

[6] National Institute for Health and Care Excellence. *Depression in adults: recognition and management. Clinical guideline 90 [Internet]*. London: NICE, 28 October 2009 [accessed 30.7.19]. 53p. Available from: https://www.nice.org.uk/guidance/cg90

[7] Cleare A, Pariante CM, Young AH, Anderson IM, Christmas D, Cowen PJ, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2008 British Association for Psychopharmacology guidelines. *J Psychopharmacol* 2015;29(5):459-525.

[8] Qaseem A, Barry MJ, Kansagara D. Nonpharmacologic versus pharmacologic treatment of adult patients with major depressive disorder: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2016;164(5):350-9.

[9] National Institute for Health and Care Excellence. *Vortioxetine for treating major depressive episodes: NICE Technology appraisal guidance 367 [Internet]*. London: NICE, 2015 [accessed 15.2.19] Available from: https://www.nice.org.uk/guidance/ta367/resources/vortioxetine-for-treating-major-depressive-episodes-pdf-82602733813189

[10] Janssen. *Data on File. Esketamine_DoF_12June2019_HEMAR_TM_002 [As referenced in CS]*

[11] Janssen. *Data on File. Esketamine_DoF_03June2019_HEMAR_TM_001 [As referenced in CS]*

[12] National Institute for Health and Care Excellence. *Scientific Advice Report. Product: esketamine, Indication: treatment resistant depression [Word document]*. London: NICE, April 2013 [accessed 30.7.19]. 26p.

[13] Johnston KM, Powell LC, Anderson IM, Szabo S, Cline S. The burden of treatment-resistant depression: a systematic review of the economic and quality of life literature. *J Affect Disord* 2019;242:195-210.

[14] McCrone P, Rost F, Koeser L, Koutoufa I, Stephanou S, Knapp M, et al. The economic cost of treatment-resistant depression in patients referred to a specialist service. *J Ment Health* 2018;27(6):567-73.

[15] National Institute for Health and Care Excellence. *Social value judgements: principles for the development of NICE guidance (second edition)*: NICE, (accessed 15.6.19) Available from:

https://www.nice.org.uk/Media/Default/About/what-we-do/Research-and-development/Social-Value-Judgements-principles-for-the-development-of-NICE-guidance.pdf

[16] National Institute for Health and Care Excellence. *Esketamine for treatment-resistant depression [ID1414]: Final scope [Internet]*. London: NICE, 2019 [accessed 6.6.19]. 5p. Available from: https://www.nice.org.uk/guidance/gid-ta10371/documents/final-scope

[17] Janssen. *Esketamine for treatment-resistant depression [ID1414]. Document B Appendices: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA)*: Janssen, 2019 [accessed 11.7.19]. 623p.

[18] National Institute for Health and Care Excellence. *Esketamine for treatment-resistant depression [ID1414]: Clarification letter*. London: NICE, 2019. 17p.

[19] Higgins J, Lasserson T, Chandler J, Tovey D, Churchill R. *Methodological Expectations of Cochrane Intervention Reviews (MECIR): standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates. Version 1.04 [Internet]*, 2017 [accessed 12.12.17] Available from: http://community.cochrane.org/mecir-manual

[20] Lefebvre C, Manheimer E, Glanville J. 6.4.9: Language, date and document format restrictions (Chapter 6: Searching for studies) [Internet]. In: Higgins JPT, Green S, editors. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 11.10.16]. Available from: http://handbook.cochrane.org/

[21] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 23.3.11] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[22] National Institute for Health and Care Excellence. *Methods for the development of NICE public health guidance (third edition). Process and methods (PMG4) [Internet]*. London: NICE, 2012 [accessed 06.09.19]. 269p. Available from: https://www.nice.org.uk/process/pmg4/resources/methods-for-the-development-of-nice-public-health-guidance-third-edition-pdf-2007967445701

[23] Janssen. *ESKETINTRD3002 Clinical Study Report [As referenced in CS]*, 2018

[24] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 23.3.11]. Available from: http://handbook.cochrane.org/

[25] Chen L, Li X, Lane R, Furey M, Singh J, Drevets W, et al. Poster M76. Relationship between the antidepressant effects of esketamine nasal spray and perceptual disturbances [Commercial in confidence]. *American College of Neurophsycopharmacology 57th Annual Meeting*. Hollywood, FL, USA.

[26] Janssen. *54135419TRD3008 Abbreviated Interim Clinical Study Report [As referenced in CS]*, 2019

[27] Kamijima K, Higuchi T, Ishigooka J, Ohmori T, Ozaki N, Kanba S, et al. Aripiprazole augmentation to antidepressant therapy in Japanese patients with major depressive disorder: a randomized, double-blind, placebo-controlled study (ADMIRE study). *J Affect Disord* 2013;151(3):899-905.

[28] Bauer M, Dell'osso L, Kasper S, Pitchot W, Dencker Vansvik E, Kohler J, et al. Extended-release quetiapine fumarate (quetiapine XR) monotherapy and quetiapine XR or lithium as add-on to

antidepressants in patients with treatment-resistant major depressive disorder. *J Affect Disord* 2013;151(1):209-19.

[29] Berman RM, Marcus RN, Swanink R, McQuade RD, Carson WH, Corey-Lisle PK, et al. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a multicenter, randomized, double-blind, placebo-controlled study. *J Clin Psychiatry* 2007;68(6):843-53.

[30] Berman RM, Fava M, Thase ME, Trivedi MH, Swanink R, McQuade RD, et al. Aripiprazole augmentation in major depressive disorder: a double-blind, placebo-controlled study in patients with inadequate response to antidepressants. *CNS Spectr* 2009;14(4):197-206.

[31] Corya SA, Williamson D, Sanger TM, Briggs SD, Case M, Tollefson G. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety* 2006;23(6):364-72.

[32] Dunner DL, Amsterdam JD, Shelton RC, Loebel A, Romano SJ. Efficacy and tolerability of adjunctive ziprasidone in treatment-resistant depression: a randomized, open-label, pilot study. *J Clin Psychiatry* 2007;68(7):1071-7.

[33] Luzny J. P.2.f.009 Treating pharmacoresistant depression in psychogeriatry - should electroconvulsive therapy or intravenous seropram be used? *Eur Neuropsychopharmacol* 2013;23(2):S400-1.

[34] Lenze EJ, Mulsant BH, Blumberger DM, Karp JF, Newcomer JW, Anderson SJ, et al. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: a randomised, double-blind, placebo-controlled trial. *Lancet* 2015;386(10011):2404-12.

[35] Marcus RN, McQuade RD, Carson WH, Hennicken D, Fava M, Simon JS, et al. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: a second multicenter, randomized, double-blind, placebo-controlled study. *J Clin Psychopharmacol* 2008;28(2):156-65.

[36] Nierenberg AA, Papakostas GI, Petersen T, Montoya HD, Worthington JJ, Tedlow J, et al. Lithium augmentation of nortriptyline for subjects resistant to multiple antidepressants. *J Clin Psychopharmacol* 2003;23(1):92-5.

[37] Fang Y, Yuan C, Xu Y, Chen J, Wu Z, Cao L, et al. Comparisons of the efficacy and tolerability of extended-release venlafaxine, mirtazapine, and paroxetine in treatment-resistant depression: a double-blind, randomized pilot study in a Chinese population. *J Clin Psychopharmacol* 2010;30(4):357-64.

[38] Thase ME, Youakim JM, Skuban A, Hobart M, Zhang P, McQuade RD, et al. Adjunctive brexpiprazole 1 and 3 mg for patients with major depressive disorder following inadequate response to antidepressants: a phase 3, randomized, double-blind study. *J Clin Psychiatry* 2015;76(9):1232-40.

[39] Thase ME, Youakim JM, Skuban A, Hobart M, Augustine C, Zhang P, et al. Efficacy and safety of adjunctive brexpiprazole 2 mg in major depressive disorder: a phase 3, randomized, placebo-controlled study in patients with inadequate response to antidepressants. *J Clin Psychiatry* 2015;76(9):1224-31.

[40] Shelton RC, Williamson DJ, Corya SA, Sanger TM, Van Campen LE, Case M, et al. Olanzapine/fluoxetine combination for treatment-resistant depression: a controlled study of SSRI and nortriptyline resistance. *J Clin Psychiatry* 2005;66(10):1289-97.

[41] Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, et al. A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. *Am J Psychiatry* 2006;163(7):1161-72.

[42] McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry* 2006;163(9):1531-41; quiz 1666.

[43] Tanghe A, Steeman J, Bollen E, van Renynghe de Voxvrie G, Dendooven M, Haazen L. Moclobemide and amitriptyline, alone or in combination, in therapy resistant depression. *Hum Psychopharmacol* 1997;12(5):509-10.

[44] Thase ME, Corya SA, Osuntokun O, Case M, Henley DB, Sanger TM, et al. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, and fluoxetine in treatment-resistant major depressive disorder. *J Clin Psychiatry* 2007;68(2):224-36.

[45] Popova V, Daly EJ, Trivedi M, Cooper K, Lane R, Lim P, et al. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. *Am J Psychiatry* 2019;176(6):428-38.

[46] Janssen. *Clinical Protocol ESKETINTRD3002; Phase 3 [As referenced in CS]*, 2015

[47] Dias S, Welton NJ, Sutton AJ, Ades AE. *NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials*. Sheffield: Decision Support Unit, ScHARR, 2011 (Updated: September 2016). 98p. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf

[48] Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry* 2007;190:287-92.

[49] Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Szegedi A. Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *J Affect Disord* 2018;226:326-331.

[50] National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 (2018 revision) [Internet]*. London: NICE, 2013 (revised 2018) [accessed 30.7.19]. 93p. Available from: https://www.nice.org.uk/process/pmg9/

[51] Edwards SJ, Hamilton V, Nherera L, Trevor N. Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(54):1-190.

[52] Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Med Res Methodol* 2011;11:139.

[53] Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* 2006;31(9):1841-53.

[54] Janssen. *Data on File. Esketamine_DoF_30Jun2019_HEMAR_TM_001 [As referenced in CS]*

[55] Briggs A, Sculpher M, et al. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press, 2006.

[56] Janssen. *Data on File. Esketamine_DoF_11June2019_HEMAR_TM_001 [As referenced in CS]*

[57] Bergfeld IO, Mantione M, Figee M, Schuurman PR, Lok A, Denys D. Treatment-resistant depression and suicidality. *J Affect Disord* 2018;235:362-367.

[58] Office for National Statistics. *National Life Tables: 2015-2017*: ONS Available from: https://www.ons.gov.uk/releases/nationallifetablesuk2015to2017

[59] van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708-15.

[60] Alava MH, Wailoo A, Pudney S. *NICE DSU: Methods for mapping between the EQ-5D-5L and the 3L for technology appraisal [Internet]*. Sheffield: Decision Support Unit, ScHARR, 2017 [accessed 18.7.19] Available from: http://nicedsu.org.uk/wp-content/uploads/2017/05/Mapping-5L-to-3L-DSU-report.pdf

[61] Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108.

[62] Revicki DA, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. *J Affect Disord* 1998;48(1):25-36.

[63] Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making* 2006;26(4):410-20.

[64] Sullivan PW, Valuck R, Saseen J, MacFall HM. A comparison of the direct costs and cost effectiveness of serotonin reuptake inhibitors and associated adverse drug reactions. *CNS Drugs* 2004;18(13):911-32.

[65] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

[66] European Medicines Agency. *Guideline on clinical investigation of medicinal products in the treatment of depression (EMA/CHMP/185423/2010 Rev 2).* 2013 (accessed 1/3/19) Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-depression_en.pdf

Appendix 1: ERG search strategies

Additional limitations of the CS searches not covered in the main body of the report:

Clinical effectiveness

- The ERG noted an error in line combinations in both the original and update search for Embase reported in Appendix D.1.1; lines #141-144 appeared to be missing from the final combination in line #145. In their response the company confirmed that they had rerun the strategy correcting the initial error resulting in an additional 610 hits from this *"These were screened for trials investigating zotepine or ECT and no further relevant trials were identified. The 610 additional studies were excluded on the grounds of study design (n=536), intervention (n=14), population (n=13), comparator not of interest/did not influence network (n=28), and duplicate (n=19)."*[3]
- The ERG noted that no synonyms for esketamine were included in the strategies for acute management or ongoing maintenance (Appendix D), although Emtree subject headings were included. A brief search on Medline and Embase for esketamine and treatment- resistant depression with the additional terms ("s-ketamin" or "s-ketamine" or vesierra or Ketanest or Spravato) yielded no additional relevant studies.