in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Naltrexone-bupropion for managing overweight and obesity

Produced by Kleijnen Systematic Reviews (KSR) Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors
Debra Fayter, Systematic Reviewer, KSR Ltd, UK
Bram Ramaekers, Health Economist, Maastricht UMC
Sabine Grimm, Health Economist, Maastricht UMC
Nigel Armstrong, Health Economist, KSR Ltd
Caro Noake, Information Specialist, KSR Ltd
Ching-Yun Wei, Health Economist, KSR Ltd
Gill Worthy, Statistician, KSR Ltd
Sofia Carrera, Systematic Reviewer, KSR Ltd
Rob Riemsma, Reviews Manager, KSR Ltd
Manuela Joore, Health Economist, Maastricht UMC
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

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

Date completed 09/03/2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number STA 14/55/01.

Declared competing interests of the authors
None.

Acknowledgements
None.

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:
Fayter D, Ramaekers BLT, Grimm S, Armstrong N, Noake C, Wei C-Y, Worthy G, Carrera S, Riemsma R, Joore MA, Kleijnen J. Naltrexone-bupropion for managing overweight and obesity: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2017.

Contributions of authors
Rob Riemsma 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. Bram Ramaekers acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Sabine Grimm, Nigel Armstrong and Ching-Yun Wei acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Debra Fayter and Sofia Carrera 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. Manuela Joore acted as health economist on this assessment, critiqued the company's economic evaluation, contributed to the writing of the report and provided general guidance. 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

ACM	All-cause mortality
ACMM	Adjusted censored mixture model
AE	Adverse Events
BCOF	Baseline observation carried forward
BMI	Body mass index
BNF	British National Formulary
BP	Blood pressure
BI	budget impact
BMOD	Behaviour modification
BSC	Best supportive care
CE	Cost effectiveness
CEA	Cost effectiveness Analysis
CEAC	Cost effectiveness acceptability curve
CHD	Coronary heart disease
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
COE	Control of Eating questionnaire
COR	Contrave obesity research
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company's submission
CSR	Clinical study report
CVD	Cardiovascular disease
CVOT	Cardiovascular outcomes trial
DM	Diabetes mellitus
EMA	European Medicines Agency
EPAR	European public assessment report
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
EUR	Erasmus University Rotterdam
FCI	Food Craving Inventory
FDA	Food and Drug Administration
GP	General practitioner
HOMA-IR	Homeostasis model assessment of insulin resistance
HR	Hazard ratio
HRQL	Health Related Quality of Life
hs-CRP	High-sensitivity C reactive protein
HTA	Health Technology Assessment
IC	Indirect Comparison
ICER	Incremental Cost Effectiveness Ratio
IDS-SR	Inventory of Depressive Symptoms – Subject related
ITC	Indirect treatment comparison
ITT	Intention to Treat
IWQOL-Lite	Impact of Weight on Quality of Life-Lite version
KSR	Kleijnen Systematic Reviews
LOCF	Last observation carried forward
LYS	Life Year Saved
MACE	Major adverse cardiovascular events
MD	Mean Difference
MeSH	Medical Subject Headings
MHRA	Medicines and Healthcare Products Regulatory Agency
mITT	Modified ITT analysis
MTC	Mixed Treatment Comparison

NA	Not applicable
NB32	Naltrexone 32mg plus bupropion 360mg
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NR	Not Reported
OR	Odds Ratio
ORL	Orlistat
OS	Overall survival
PRESS	Peer Review of Electronic Search Strategies
PSA	Probabilistic Sensitivity Analyses
PSS	Personal Social Services
QALY(s)	Quality-adjusted Life Year(s)
QoL	Quality of life
RCT	Randomised Controlled Trial
RR	Relative Risk; Risk Ratio
SAE	Serious Adverse Events
SD	Standard deviation
SM	Standard management
SPC	Summary of product characteristics
STA	Single Technology Appraisal
UMC	University Medical Centre
T2DM	Type 2 Diabetes Mellitus
TEAEs	Treatment-emergent adverse events
TESAEs	Treatment-emergent serious adverse events
TIA	Transient Ischaemic Attack
UK	United Kingdom
VAS	Visual Analogue Scale
WHO	World Health Organisation

Table of Contents

Table of Tables

Table of Figures

1. SUMMARY

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

The NICE scope describes the decision problem as naltrexone-bupropion prolonged release (32mg daily) or NB32 for managing overweight (≥ 27 kg/m2 to < 30 kg/m^2; in the presence of one or more weight-related co-morbidities) and obesity (≥ 30 kg/m^2). The comparators are described as: standard management without naltrexone-bupropion and orlistat (360 mg/day). Standard management is not defined in the NICE scope.

NB32 is indicated, as an adjunct to a reduced-calorie diet and increased physical activity, for the management of weight in adult patients (≥18 years). NB32 treatment should be discontinued after 16 weeks if patients have not lost at least 5% of their initial body weight. Likewise, treatment with orlistat should be stopped after 12 weeks if patients have been unable to lose at least 5% of their body weight since the start of treatment. In most trials NB32 and orlistat are continued throughout the trial, usually one year's duration.

The main question regarding the decision problem is the appropriateness of the intervention and comparator insofar as what constitutes standard management in clinical practice. The company assumed that standard management was more like that in the COR-I and COR-II trials in which patients received advice on diet and exercise but were not allowed to participate in a weight loss programme. However, if those who are prescribed NB32 or orlistat would engage in a concomitant weight loss programme then the intervention might be more like that, referred to as 'intensive behaviour modification', in the COR-BMOD trial. Similarly, if those who are eligible for either NB32 or orlistat would otherwise engage in a weight loss programme then the comparator might be more like that in COR-BMOD.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company's submission included data from four main trials comparing NB32 to placebo as an adjunct to standard management: COR-I, COR-II (general overweight and obese population), COR-BMOD (intensive behaviour modification) and COR-DM (diabetes population). Mean BMI across the trials was 36 to 37. Approximately 2% of participants were overweight and 98% obese.

All trials were multicentre and all were conducted in the US. All had a joint primary outcome of percentage change in total body weight and proportion of patients with >5% decrease in total body weight. Three trials measured outcomes at week 56. One trial, COR-II measured the primary outcome at 28 weeks. In COR-II, NB32 patients who had lost less than 5% of their body weight at visits between weeks 28 and 44 were re-randomised to continue with NB32 or escalate to NB48. The four trials included 4,536 patients. Of these 2,510 patients were randomised to NB32, 578 to NB16 (in COR-I) and 1,448 randomised to placebo.

The main results presented in the company submission (CS) were based on a modified intention-to-treat (mITT) analysis. According to the CS this was defined as "*all randomised patients with a post-baseline body weight measurement obtained while the patient remained on study medication.*" In this modified ITT analysis, approximately 20% of randomised patients were not included in the analyses.

Direct evidence from the four main trials showed that the mean difference in percentage weight change at week 56 from baseline was -3.3 (95% CI: -4.3 to -2.2) for COR-DM, favouring NB32 compared with placebo; -4.2 (95% CI: -5.6 to -2.9) for COR-BMOD; -4.6 (95% CI: -5.2 to -3.9) for COR-II (at 28 weeks); and -4.8 (95% CI: -5.6 to -4.0) for COR-I. Analyses for the number of patients with ≥ 5% decrease in weight at week 56 also significantly favoured NB32 over placebo in all four trials (Odds

ratios: 3.4 (95% CI: 2.2 to 5.5) for COR-DM; 2.9 (95% CI: 2.0 to 4.1) for COR-BMOD; 6.6 (95% CI: 5.0 to 8.8) for COR-II (at 28 weeks); and 4.9 (95% CI: 3.6 to 6.6).

The percentages of overweight patients (BMI < 30 kg/m^2) in the trials are too small to present meaningful subgroup analyses.

Adverse events occurred in 83.1% to 93.7% of treatment groups and 68.5% to 88.0% of placebo groups. Approximately 58% to 76% of these were attributed to the drug in NB32 groups across the trials. Serious adverse events occurred at similar rates in treatment and placebo groups across the trials. However, a larger number of patients discontinued due to adverse events across the trials (19.5% to 29.4% for treatment groups) versus 9.8% to 15.4% in placebo groups).

The main category of adverse event occurring more frequently in treatment groups across the trials was gastrointestinal disorders. Nausea, in particular, occurred frequently and more often in treatment groups. Across the trials, rates of nausea ranged from 29.2% to 42.3% in treatment groups. Rates ranged from 5.3% to 10.5% in placebo groups. Vomiting, constipation and dry mouth also occurred more frequently in treatment groups although at a lower rate than that of nausea. Nervous system disorders such as headache, dizziness and tremor occurred more frequently in treatment groups.

The incidence of events of particular concern (serious cardiovascular disorders and suicidality measured on IDS) was extremely small and any differences between groups could not be ascertained in view of the small numbers in both groups.

No trials were identified that compared NB32 directly with orlistat or with different types of behavioural interventions. Therefore, the company performed indirect comparisons to compare NB32 with orlistat using placebo as the common comparator. Twenty trials were included in the indirect treatment comparison (ITC), four for NB32 and 16 for orlistat.

Results for mean percentage weight change from baseline at one year showed that there were no significant differences between NB32 and orlistat for people with diabetes and for all patients combined. There was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients are excluded (MD 1.13 (95% CrI: 0.44, 1.80)). The difference is most significant for the third sensitivity analysis, where studies with 'intensive' behaviour modification (BMOD and XENDOS) were also excluded (MD 2.98 (95% CrI: 1.60, 4.36)).

Results for ≥5% reduction in weight at one year showed that there were no significant differences between NB32 and orlistat for people with diabetes and for all patients combined. There was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients are excluded (OR 0.77 (95% CrI: 0.61, 0.96)). The difference is most significant for the third sensitivity analysis, where studies with 'intensive' behaviour modification (BMOD and XENDOS) were also excluded (OR 0.44 (95% CrI: 0.23, 0.84)).

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The company submission (CS) and response to clarification provided sufficient details for the ERG to appraise the searches for eligible trials. Searches were carried out in accordance with the NICE guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4 using a good range of databases. Additional searches of conference proceedings and organisational websites were reported, along with the checking of reference lists of existing systematic literature reviews (SLRs) and meta-analyses.

The four main trials comparing NB32 to placebo are of high quality. However, there are a number of limitations when applying them to clinical practice. There are very little data on ethnic groups relevant

to the UK (particularly people from Asia) within the NB32 trials, therefore it is not possible to make any firm conclusions for that group. There are very few overweight as opposed to obese participants in the trials. The majority of the participants in the NB32 trials are female. Trials do not measure weight loss beyond 56 weeks. The large dropout from the NB32 trials (up to 50%) is relevant to practice. The US setting may reflect a different patient profile and differing approaches to standard care than in a UK setting.

A comparison between NB32 (plus standard management) versus intensive behaviour modification is missing. Furthermore, comparisons between NB32 and orlistat are based on indirect comparisons only.

The company used modified ITT data from NB32 trials, but this is misleading. The mITT population in the NB32 trials is very different from mITT populations in the orlistat trials. In the NB32 trials, 21.9% of patients receiving NB32 were randomised but excluded from the analyses against 1.6% of patients receiving orlistat.

Comparison with orlistat may be biased in favour of NB32. NB32 trials were published in 2010 or later; most of the trials with orlistat were published before 2005, so caution should be exercised when making indirect comparisons; this is particularly true for conditions such as diabetes where background standard therapy (for glucose and lipids especially) may be very different now.

We have reproduced the company's indirect analyses comparing orlistat and NB32 using full ITT data from the NB32 trials. The results show that the positive effects of NB32 when compared to orlistat have all disappeared. For the first outcome (≥5% reduction in weight at one year), there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded using mITT data. However, in both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: OR = 1.09 (95% CrI: 0.87 to 1.36), ITT-BOCF: OR = 1.06 (95% CrI: 0.84 to 1.33). Moreover, although none of the differences are statistically significant, all results now favour orlistat.

For the second outcome (mean percentage weight change at one year), using mITT data there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded. However, in both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: MD= -0.09 (95% CrI: -0.77 to 0.58), ITT-BOCF: MD = -0.54 (95% CrI: -1.21 to 0.12). Moreover, although most of the differences are not statistically significant, most results now favour orlistat.

Standard management in the UK might be better reflected by COR-BMOD; therefore, we have included a new analysis: an indirect comparison of NB32 plus intensive behaviour modification (COR-BMOD) versus orlistat plus intensive behaviour modification (XENDOS). The results show that both outcomes significantly favour orlistat over NB32 (≥5% reduction in weight at one year: OR 1.86 (95% CI: 1.30 to 2.66); Mean percentage weight change from baseline (CFB) at one year: MD -2.09 (95% CI: -3.53 to -0.65)).

Finally, we performed our preferred analyses, i.e. using full ITT data and no pooling of NB32 trials (using only COR-I ITT data for non-diabetics, instead of COR-I, COR-II and COR-BMOD combined). The results for 'obese patients with T2DM' and 'intensive behaviour modification' are the same as before, but results for 'obese patients without T2DM' have changed considerably again, and are almost the same as in the company's original analyses. Both outcomes show no significant difference between NB32 and orlistat, but both favour NB32.

The table below shows the main results for obese people with diabetes, obese people without diabetes and NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification.

Table 1.1: Company results versus ERG results

Population		Company analyses (mITT data)*	Company analyses (ITT-BCFA data)**	ERG preferred analyses**
		Orlistat vs NB32	Orlistat vs NB32	Orlistat vs NB32
Obese people with T2DM				
≥5% reduction in weight at 1 year	OR	1.09 (0.63 to 1.88)¶	1.59 (0.89 to 2.79)¶	1.59 (0.89 to 2.79)¶
Mean % weight CFB at 1 year	MD	0.21 (-0.87 to 1.30)†	-1.21 (-2.30 to -0.11)¶	-1.21 (-2.30 to -0.11)¶
Obese people without T2DM				
≥5% reduction in weight at 1 year	OR	0.77 (0.61 to 0.96)†	1.06 (0.84 to 1.33)¶	0.61 (0.31 to 1.22)†
Mean % weight CFB at 1 year	MD	1.13 (0.44 to 1.80)†	-0.54 (-1.21 to 0.12)¶	1.11 (-0.39 to 2.63)†
Intensive behaviour modification				
≥5% reduction in weight at 1 year	OR	1.22 (0.84 to 1.77)¶	1.86 (1.30 to 2.66)¶	1.86 (1.30 to 2.66)¶
Mean % weight CFB at 1 year	MD	-0.21 (-1.28 to 1.70)¶	-2.09 (-3.53to -0.65)¶	-2.09 (-3.53to -0.65)¶

Results are OR with 95% CI/CrI for ≥5% reduction in weight at 1 year and mean difference (MD) with 95% CI/CrI for mean % weight CFB at 1 year.
An OR less than one favours NB32 over orlistat and a CI including 1 is not significant. A MD of >0 favours NB32 over orlistat and indicates greater % weight reduction and a CI including 0 is not significant.
¶ = Favours orlistat; † = Favours NB32.
*) Bayesian NMA (OR, 95% CrI) using mITT data; **) Using the Bucher method for indirect comparisons and ITT-BCFA data.
FE = fixed effect; ITT-BCFA = all randomised patients with baseline-carried-forward analysis; MD = Mean Difference; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; OR = Odds Ratio; T2DM = Type 2 diabetes mellitus;

Which of the estimates of treatment effect is more applicable to clinical practice depends on the definition of standard management. If individuals who are eligible for NB32 would also engage in a weight loss programme when prescribed NB32 then the so-called intensive behaviour modification estimate might be more applicable. If this is not the case, then an estimate excluding intensive behaviour modification might be more appropriate. Of course, the estimate of 1.06 (0.84 to 1.33) is based on pooling both the trials with and without intensive behaviour modification and it is therefore tempting to infer that this represents clinical practice, where some do and some do not engage in weight loss programmes. This must be regarded with caution for a number of reasons, which include uncertainty as to the precise proportion who would engage in a weight loss programme and the degree of resemblance between such a programme and the intensive behaviour modification in COR-BMOD.

1.4 Summary of cost effectiveness evidence submitted by the company

The company conducted systematic reviews to identify relevant cost effectiveness studies, health-related quality of life (HRQoL) studies, resources and costs studies. The company did not identify any study investigating the cost effectiveness of NB32 adjunct to standard non-pharmacological

management in the population of interest for the current decision problem, and hence developed a *de novo* model with a lifetime horizon.

The company developed an economic model using an individual-level approach, more specifically a discrete event simulation (DES). It was argued that an individual-level approach is better suited than a cohort-level approach to capture the chronic implications of both weight and weight-related health events in a heterogeneous group of overweight and obese patients. The DES model was implemented in Excel using the "discretely integrated condition event" (DICE) principles and structure. The company used an economic evaluation by Ara et al. (also an individual-level model) as a starting point, which is from a 2012 Health Technology Appraisal comparing different pharmacological treatments for obesity. The following events are considered in the economic model:

- treatment discontinuation;
- development of T2DM;
- first cardiovascular event (either stroke or MI);
- second cardiovascular event (either stroke or MI) and;
- death.

Upon model entry, a simulated patient is assigned a profile of sampled baseline characteristics that are explanatory factors for risks, costs and/or utility in the model (sampled baseline characteristics as well as random numbers for the sampled patient are equal across all three treatments). The baseline profile characterises the individual patients by:

- age (years);
- gender (male, female);
- height (meters);
- BMI (kg/m^2);
- T2DM status (yes, no);
- smoker status (current, previous, never);
- receive insulin, if diabetic (yes, no);
- receive statins (yes, no);

The company stated that the economic analysis aimed to reflect the patient group for which the drug is licensed: adult patients who are obese (BMI ≥30kg/m2), or overweight (BMI ≥27kg/m2 and <30kg/m2) in the presence of one or more weight-related comorbidities (e.g., T2DM, dyslipidaemia, or controlled hypertension). The company assumed that no patients would have a history of angina or diabetes other than T2DM and no patients received anti-hypertensive medication and/or aspirin.

In line with the final scope and licensed indications, the company considered orlistat as an adjunct to standard management and standard management alone as comparators for NB32 as an adjunct to standard management. NB32 is implemented as per its European Medicines Agency (EMA) Summary of Product Characteristics (SmPC) posology and method of administration, incorporating a four week escalation period, after which the maximum recommended daily dose of 32mg naltrexone hydrochloride and 360mg bupropion hydrochloride is assumed. Orlistat is similarly implemented as per its EMA SmPC posology and method of administration, a 360mg daily dose. The company specified standard management as implemented in the analysis to reflect the non-pharmaceutical dietary and lifestyle management treatment received in UK NHS practice.

Treatment effectiveness estimates (i.e. time to treatment discontinuation data, proportion of responders, and change in body weight) were mainly derived from the COR trial programme, including the COR-

I, COR-II, COR-BMOD and COR-DM trials. All the analyses were based on the company's modified ITT analysis, which reflects only those patients who have a post-baseline measurement whilst on the study drug. Time to treatment discontinuation was estimated based on the COR trial programme and extrapolated after one year using the NB-CVOT study. All patients were assumed to have discontinued after treatment duration data were unavailable in this study. It should be noted that the company used the same time to treatment discontinuation Kaplan-Meier curves for both NB32 and orlistat. The company justified this by stating that data were lacking for orlistat. Both the proportions of responders and change in body weight were obtained from the COR trial programme for NB32 and standard management; an ITC was used to calculate this for orlistat. The changes in body weight were used to predict development of T2DM, cardiovascular event (either stroke or MI) and death using parametric time-to-event models (Weibull distribution) retrieved from the report by Ara et al. Also the natural history of BMI model, to predict BMI over time, was retrieved from this report. The company stated that it was unable to make trial data comparisons of AEs associated with NB32 and orlistat because details from clinical literature and regulatory documents on orlistat were insufficient. Therefore, the company assumed equal AE related costs for NB32 and orlistat. The impact of AE on utility scores was not incorporated by the company.

The company applied a Tobit model to estimate utility values based on the Public Health England weight management economic assessment tool v2 (Health Survey for England EQ-5D data analysis). This model includes explanatory variables for BMI, age, gender, and obesity-related conditions (stroke, MI, cancer and T2DM).

Costs in the model consisted of drug acquisition costs, non-drug costs related to standard management (applicable to all treatments considered), obesity-related comorbidity costs and adverse event costs. Drug acquisition costs for NB32 and orlistat were based on the list price and Monthly Index of Medical Specialities respectively. The non-drug resource use items comprising standard management in the model consisted of GP visits, nurse visits and blood tests which were informed by the COR trials, literature and clinical opinion. Moreover, obesity related comorbidity costs were retrieved from the literature and for AE the costs of one GP visit were assumed.

The company's model uses 1,000 patient profiles for their deterministic analysis. For the probabilistic sensitivity analysis (PSA), the company used only 500 patient profiles and 100 PSA simulations. Moreover, not all model parameters were incorporated in the PSA.

In the base-case deterministic analysis, NB32 was associated with an incremental QALY gain of 0.0765 QALYs versus standard management, and 0.0192 QALYs versus orlistat. The incremental costs of NB32 were £1,044 versus standard management and £750 versus orlistat. The incremental cost effectiveness ratio (ICER) of NB32 versus standard management was £13,647 per QALY. The estimated ICER versus orlistat was £32,084 per QALY. Subgroup analyses performed by the company indicated that the ICERs of NB32 compared with standard management and orlistat were £5,059 and £72,069 respectively for T2DM patients and £6,283 and £28,291 respectively for non-T2DM patients.

The deterministic sensitivity analyses performed by the company showed that the most influential parameters were the parameters of the Tobit model for utilities and the discount rate for QALYs, as well as parameters related to the measures of relative efficacy from the ITC. The company performed scenario analyses on the following model aspects: the time period over which weight is regained, the cost of T2DM, the utility estimates, costs of AEs, discounting, and the time horizon. The most influential scenarios were shortening the time period for weight regain from three to two years (ICER £41,016), and shortening the time horizon from lifetime to 15 years (£53,514).

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The majority of the cost effectiveness searches in the CS were well documented and easily reproducible, and were carried out in line with the NICE guide to the methods of technology appraisal.

The ERG considered it reasonable to use the economic model by Ara et al. as a starting point for the current analysis. It should be noted that the company deviated from the assumption made by Ara et al., that patients would have regained weight to obtain the *baseline* BMI within three years in a linear fashion and assumed instead that patients would have regained weight to obtain the *age/sex predicted* BMI in three years. The company did not provide justification for why their deviation from Ara et al.'s assumption was 'logical' and plausible. Hence, to be consistent with Ara et al., the ERG preferred to assume weight regain to the baseline BMI in its base-case. Furthermore, the linear weight regain over the time-span of three years was implemented incorrectly in the model in that the weight regain occurs instantaneously at the end of the three year period. The ERG also questioned the (justification for the) assumption of equivalent weight loss at similar assessment times. The company's model assumed weight loss for orlistat patients at weeks 12 and 52 to be comparable to weight loss for NB32 patients at weeks 16 and 56. This was not justified besides stating that this assumption was also upheld within the ITC. The model only includes the possibility of two subsequent cardiovascular events (i.e. either two strokes, two MIs or one stroke and one MI), implicitly assuming that the impact of the third cardiovascular event on costs, quality of life and survival, is negligible. It can, however, be questioned whether having a stroke after having experienced two MIs is indeed unimportant.

The population aimed to reflect the scope. However, patient characteristics in the model were sampled from estimates that were based on a variety of sources. It is questionable whether this is reflective of UK clinical practice. The ERG agrees with using the COR trial programme patient-level data to inform baseline patient characteristics in the model (as done for age, gender and height). This follows from a) that the effectiveness estimates are derived from this population and b) that the company stated, based on clinical opinion, that patient characteristics in the COR trial programme are a fair reflection of the typical patient group that would receive NB32 in UK NHS clinical practice. However, the appropriateness of other baseline characteristics is less clear. The ERG considered the BMI sampled in the model and compared it with the baseline BMI in the COR trial programme and concluded that baseline BMI is vastly underestimated in the economic model. This is also reflected in the average baseline weight of 92kg in the model, while the averages ranged between 99kg and 105kg in the COR trial programme. Given that BMI is included as a predictive factor for utility, T2DM, cardiovascular events and death, the utility values and the time to these events in the model are overestimated, likely inducing bias in favour of NB32.

Other baseline characteristics are also potentially underestimated:

- Proportion of current smokers
- Proportion of patients receiving anti-hypertensive medication
- Proportion of patients with a history of angina and/or diabetes other than T2DM
- Proportion of patients receiving aspirin

In contrast to the above, the proportion of patients receiving statins and patients with T2DM might have been overestimated. Moreover, correlations between covariates were not incorporated in the sampling of the patient characteristics, leading to counter-intuitive patient profiles. For instance, based on the patient characteristics of the COR-I, COR-II, COR-BMOD and COR-DM trials, it becomes clear that the patients without T2DM (COR-I, COR-II and COR-BMOD trials) have different patient characteristics (e.g. regarding age, sex, hypertension status and statin use) than patients with T2DM (COR-DM trial). This is neglected in the sampling of the patient population. To address these issues,

the ERG adjusted the baseline characteristics used in the model. This included calibrating the natural history model to predict BMI over time.

The company did assume no patients had a history of angina and/or diabetes other than T2DM. This assumption was made as no data were identified on these characteristics for overweight/obese patients. The ERG agrees with this statement and would therefore argue that it can be questioned whether the results of the economic analyses are representative for patients with a history of angina and/or diabetes other than T2DM.

One major limitation of the model is the inability to incorporate re-treatment, behaviour modification treatment and/or bariatric surgery (for which patients become eligible over time once their BMI is/increases to >40kg/m2 in the model).

The ERG considers that the use of the ITT population (instead of the mITT) to inform treatment response and weight loss would have been both more appropriate and more conservative. Using the true ITT data, NB32 would achieve a smaller mean percentage of weight loss and smaller proportion of responders compared to the mITT data. It is also the ERG's view that it was inappropriate to pool from all COR studies, including COR-BMOD and COR-II. Effectiveness estimates derived from the COR-BMOD trial where NB32 was administered in combination with intensive behavioural modification are substantially different when compared to effectiveness estimates derived from studies in which NB32 was administered together with standard management only. Likewise, the ERG considers the use of COR-II for the derivation of treatment effectiveness beyond 28 weeks as inappropriate because NB32 participants with <5% weight loss at visits between Weeks 28 and 44 were re-randomised. The ERG therefore considers that NB32 treatment effectiveness estimates should only be derived from the COR-I and COR-DM trials.

Because of the following reasons, the ERG believes time to discontinuation (TTD) is underestimated for all treatments in the model but in particular for orlistat:

(1) TTD estimates for the period after the one year assessment were derived from the NB-CVOT study in which patients had characteristics associated with an increased risk of CV outcomes, potentially leading to a shorter TTD.
(2) The end of the NB-CVOT study was used as the maximum TTD, whether patients in that study had discontinued or not.
(3) The company claims that the most reasonable and conservative assumption was to assume that TTD for orlistat would follow a similar trajectory to NB32, given that patient-level data for orlistat were unavailable. However, the ERG found publications reporting TTD for orlistat, which reveal that orlistat TTD was longer than the 12.29 months estimated by the model, with many studies reporting that the proportion of patients still receiving orlistat at 12 months was >50%.
(4) For the derivation of the orlistat TTD, the KM estimates for NB32 TTD for the first 16 weeks were linearly scaled to fit the first 12 weeks of orlistat treatment.

The ERG considers the company's claim that not accounting for a HRQoL impact of AEs in the economic model is conservative as highly questionable. The company provided no systematic overview of evidence that showed that the AE profile of orlistat was indeed worse than that of NB32. There is no direct evidence comparing the two drugs and indirect treatment comparisons between the drugs focused on efficacy but not on safety outcomes. Therefore the company's assertion of the likely superiority of NB32 in relation to orlistat in terms of AE remains speculative. Upon request, the company provided a scenario analysis in their response to clarification question B13, in which "pragmatic application of on-

treatment disutilities has been provided", assuming all AEs to be associated with a utility decrement of 0.05 for the duration of one week. This analysis increased the company's base-case ICERs against orlistat and SM by £188 and £87 per QALY gained, respectively.

The ERG is concerned that the regression model that informs the utility estimates does not appear to be published in a peer-reviewed journal. As a consequence, given the limited amount of details, the validity of these regression models to estimate utility values cannot be assessed by the ERG. However, upon request from the ERG, the company assessed the face validity of the utility estimates. The company stated that the utility values predicted by the Tobit model for the healthy population resembled the ones from the general UK population and that the remainder of the predicted utilities lay below these, demonstrating face validity.

The ERG considered it plausible to use Ara et al. to inform healthcare resource use assumptions. Regarding the costs of standard management, it is unclear to the ERG why the company added a GP visit for the 52 week assessment for patients receiving standard management only. Therefore, the ERG removed this GP visit for patients receiving standard management only.

The ERG ran the deterministic CS base-case model with 1,000 individual sampled patients, which resulted in an ICER of NB32 versus orlistat ~£3,000 higher than the base-case results reported in the CS. In the ERG's further analyses, there was substantial variation in the ICERs obtained in model runs when a different set of random numbers was used and a new set of patients were sampled. Based on the ERG's findings, and the uncertainty that the company's diagnostic exercises truly reflected the stability of the model, the ERG believes that the model should ideally be evaluated using a much larger number of sampled patients (more than the 1,000 that are used in the CS base-case). However, model run times were prohibitive (six hours on average per model run with 1,000 patient profiles) and the model was restricted to incorporate a maximum of 1,000 patients. Moreover, the ERG believes the PSA results in the CS are flawed for multiple reasons: 1) the low number of individual sampled patients (500) included in the PSA; 2) the low number (100) of PSA simulations and; 3) the exclusion of key input parameters from the PSA (e.g. TTD, natural history of BMI model, obesity-related events).

The structure and technical implementation of the company's model caused long run times (6 hours on average), and caused the model to crash on multiple computers. This hampered the company's and the ERG's ability to perform an appropriate PSA and the ERG's ability to check the model's validity and perform further scenario analyses (other than those that were described below). It should be considered whether simpler approaches (e.g. an individual-level state transition model) would have been more appropriate to reflect this decision problem, given the gain in transparency and that it would have been possible to reflect the condition-specific events in such a model. An individual-level state transition approach would potentially resolve most of the validity issues (e.g. the fact that BMI was not accurately reflected at each time period).

The ERG considered the internal validity of the model (e.g. checking formulae in the DICE sheet, examining the implementation of TTD in the model, examining available intermediate outcomes). However, the ERG was unable to examine the internal validity of the model according to its usual standards. This was mainly a consequence of the long model run times for one single deterministic analysis (six hours) and the inability to examine intermediate outcomes. For instance, the nature of the model hampered the ERG's ability to do sensitivity analysis; extreme value analysis; trace analysis/analysis of intermediate outcomes which are recommended by the ISPOR taskforce on model transparency and validation. Therefore, the ERG wishes to note that it cannot be guaranteed that there are no modelling errors (in addition to the methodological flaws described below). In this light, the

ERG considers it troublesome that the company did not provide the results of the internal validation it performed (as requested in response to clarification question B19).

One of the main validity issues or methodological flaws the ERG encountered was the inaccurate reflection of patients' BMI and consequently health-related quality of life. After the first year, patients have on average only three events in 32.8 years, equalling to an average of one event per 10.6 years. This entails that BMI after the first year is only updated on average once every 10.6 years (implicitly assuming a stable BMI in the periods between events), while this should be updated at least annually to reflect the increasing BMI due to its correlation with age (as reflected in the natural history model predicting BMI over time). This could have been solved by an annual updating event, the integration of the BMI function or the use of a different model structure. Apart from the annual updating or integration of BMI (and the impact on associated risks and utility values), the lack of model updating also affects other assumptions in the model. For example, the assumption regarding weight regain after treatment discontinuation for NB32 and orlistat was intended to reflect linear weight regain for a period of three years after which the BMI is obtained (predicted by the natural history model). However, if there is no event in this three year weight regain period, which is more likely than not (based on the average of one event per 10.6 years), the BMI estimated at the time of treatment discontinuation is maintained for this weight regain period of three years after which the weight is regained instantly. It should be noted that, if the death event were to be excluded from this calculation, the average time until one event would increase to 17.2 years. According to the ISPOR taskforce on DES, it would have been recommended to incorporate 'time checks' (i.e. 'update events'). Given that BMI is underestimated as a consequence of this methodological flaw, the utility values and the time to the events in the model are overestimated, likely inducing bias in favour of NB32. Moreover, assuming stable BMI for long periods of time also limits the face validity of the model.

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

The majority of searches for eligible studies in the CS were well documented and easily reproducible. Searches were carried out on a good range of databases. The strategies utilised recognised study design filters. Supplementary searches of conference proceedings and organisational websites, and the checking of references lists were undertaken by the company in order to identify additional studies not retrieved by the main searches.

Four good quality large RCTs for NB32 and 16 comparator trials were included in the submission. Analyses were presented for all patients and people with and without T2DM, including a large number of sensitivity analyses.

The economic model structure is similar to the assessment by Ara et al., which is a Health Technology Appraisal report (2012) comparing different pharmacological treatments for obesity.

1.6.2 Weaknesses and areas of uncertainty

There were limitations with the use of indexing terms on Embase.com searches, as strategies only used EMTREE. Although some mapping between indexing terms does take place on Embase.com it is possible that relevant MEDLINE indexing terms (MeSH) will not have been included in the search, and potentially relevant records could have been missed

The main weakness of the CS is the use of mITT populations for the NB32 trials. These data overestimate the benefits of NB32 over placebo or orlistat when compared to the true ITT data.

Uncertainty remains surrounding the effectiveness of NB32 for patients who are overweight with comorbidities as opposed to obese; ethnic groups relevant to a UK setting and those who have previously used orlistat. Further uncertainties include any further weight loss and maintenance of weight loss after 56 weeks, and retreatment with NB32. The relative benefit of NB32 in comparison to orlistat is uncertain when all data are taken into account. The benefit of NB32 when compared to an optimally delivered intensive intervention in practice is unclear as is NB32 treatment discontinuation in clinical practice.

The interpretation and validity of the results are particularly hampered given that the company's model did underestimate TTD, did not incorporate behaviour modification interventions (e.g. weight loss programmes), bariatric surgery and re-treatment nor an updating event or integration of the BMI function that was required to accurately reflect patients' expected quality of life and costs associated with resource use. The lack of an updating event or integration of BMI could significantly bias the results in favour of NB32. The model structure and technical implementation of the model hampered the assessment of validity of all parts of the model in the given time-frame. It should be considered whether simpler approaches (e.g. an individual-level state transition model) would have been more appropriate to reflect this decision problem, given the gain in transparency and given that it would have been possible to reflect the condition-specific events in such a model. An individual-level state transition approach would potentially resolve most of the validity issues (e.g. the lacking updating event).

Furthermore, the ERG considers the model as unfit for purpose, due to its extremely long run times, the fact that it crashes on many computers, and the inability to perform PSA.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

Numerous issues were identified by the ERG. The ERG was able to adjust/correct some of these issues in its base-case. The ERG base-case ICERs (deterministic) of NB32 compared with standard management and orlistat ranged between £9,813-£10,510 and £38,871-£45,694 per QALY gained respectively. Subgroup analyses performed conditional on the ERG base-case, indicated that the ICERs (deterministic) of NB32 compared with standard management and orlistat were £10,535 per QALY gained and dominated respectively for T2DM patients and £9,594 and £25,744 per QALY gained respectively for non-T2DM patients.

In conclusion, the large variation around the ICERs when different random numbers and sampled patient profiles are used is of particular concern. In two different model runs of the ERG base-case, the ICER varied by as much as £7,000 per QALY gained. It is therefore the ERG's view that the company's model is of very limited value for the current decision problem and that results are to be interpreted with extreme caution.

2. BACKGROUND

In this section the ERG provides a review of the evidence submitted by Orexigen in support of naltrexone plus bupropion (NB32), trade name Mysimba® as a centrally acting anti-obesity product. We outline and critique the company's description of the underlying health problem and the overview of current service provision. The information is taken from Chapter 3 of the company submission (CS) with sections referenced as appropriate.

2.1 Critique of company's description of underlying health problem

The underlying health problem of this appraisal is overweight and obesity. According to Section 3.1 of the CS *"In clinical practice, body fatness is generally assessed by the BMI, calculated as body weight (kg) divided by height squared (m^2). The BMI range for normal weight is 18.5–24.9kg/m^2; overweight is 25–29.9kg/m^2; obese is 30–40kg/m^2 and morbidly obese is defined as >40kg/m^2"*.[1]

In Section 3.4 of the CS the prevalence of overweight and obesity is reported *"Based on the 2014 Health Survey for England, a total of 11,126,000 adults (aged ≥16) were obese (BMI ≥30kg/m^2). In addition 15,825,000 adults are overweight[2] with around 30% or 4,747,500 having a BMI ≥27kg/m^2. Of these, an estimated 16% will have one or more weight-related comorbidity, equivalent to 779,680 patients. Therefore, a total of 11,905,680 adults in England are overweight or obese with one or more weight-related comorbidities"*.[1]

In Section 3.1 of the CS it is noted that *"Men are more likely to be overweight; however women are more likely to be obese."* It is also noted that *"those aged 55–64 years are the most likely to be obese, while 16–24 year olds are least likely."*[1]

The CS states that **"***For both overweight and obesity, the fundamental cause is an energy imbalance between calories consumed from food and drink and calories expended through exercise and energy expenditure; over time, this imbalance results in abnormal or excessive fat accumulation.*" The submission also highlights increased intake of foods that are high in fat and a decrease in physical activity levels as the most influential factor in increasing the prevalence of obesity. The CS also references a number of other factors influencing obesity.[1]

The CS describes how a number of health problems are associated with being overweight or obese and that the available literature focuses on those associated with obesity. They also state that *"because many people who are overweight will become obese in their lifetime, it is reasonable to assume the comorbidities listed are relevant to both populations"*. These include T2DM, hypertension, heart disease, dyslipidaemia, coronary artery disease and stroke, respiratory effects, cancers, reproductive function and osteoarthritis.[1]

In Section 3.2 of the CS the company states *"Overweight and obesity also have a substantial mental health burden and can be associated with sleep apnoea and severe depression"*.[1]

Section 3.4 of the CS states that *"In 2004, research by a House of Commons Select Committee estimated that 34,100 deaths were attributable to obesity. This equates to 6.8% of all deaths in England"*.[3]

The economic burden of obesity is highlighted in the CS. *" A report from 2007 estimated that NHS costs attributed to elevated BMI were £4.2 billion, with indirect costs amounting to £15.8 billion.[4] This was expected to rise to £6.3 billion in 2015, £8.3 billion in 2025 and £9.7 billion in 2050.[3, 4]"*.[1]

ERG comment: The ERG checked the references provided to support the statements in the submission. In general these were found to be appropriate. However the ERG noted a number of discrepancies:

- Although BMI measures of overweight and obesity cited in the CS match NICE guidelines,[5] the guidelines also emphasise that BMI should be interpreted with caution and that waist circumference in people with a BMI < 35kg/m2 should be considered. The guidelines also state that "*The use of lower BMI thresholds (23 kg/m2 to indicate increased risk and 27.5 kg/m2 to indicate high risk) to trigger action to reduce the risk of conditions such as type 2 diabetes, has been recommended for black African, African-Caribbean and Asian (South Asian and Chinese) groups*."[5]
- It was unclear how exactly numbers of adults who are overweight or obese with weight-related comorbidities in England quoted in the CS were derived. No source was cited for the estimated 16% with a weight-related comorbidity.
- The statement that women are more likely to be obese is incorrect. Twenty-seven percent of both genders are obese.[2] Women are more likely to be morbidly obese (BMI>40) than men (3.6% vs 2.2%) 68% of men were overweight or obese in 2015 compared to 58% of women.[3]
- Important variations for the prevalence of obesity have also been linked with social class. It has been suggested that this is associated with the degree of relative social inequality.[4]
- The studies supporting the link between excess weight and depression report an association only for those who are severely obese and/or have a chronic health condition.
- The report cited by the company on deaths associated with obesity referenced data collected in 2001.[3] According to the World Health Organisation, an estimated 9.6% of deaths among men and 11.5% of women are due to overweight and obesity in developed countries.[6] Applying these to England (2001 data) gives 52,500 not 34,100 deaths attributable to obesity as cited by the company.

2.2 Critique of company's overview of current service provision

The CS notes that in England "*Treatment is based upon a patient's BMI and what, if any, comorbidities are present, as outlined in Table 8"* (duplicated below).[1]

Table 2.1: Summary of treatment options for overweight and obese patients

BMI classification (kg/m2)	Waist circumference[a]			Comorbidities present
	Low	High	Very high	
Overweight (25-29.9)	1	2	2	3
Obesity I (30-34.9)	2	2	2	3
Obesity II (35-39.9)	3	3	3	4
Obesity III (40 or more)	4	4	4	4
Treatment options				
1	General advice on healthy weight and lifestyle			
2	Diet and physical activity			
3	Diet and physical activity, consider drugs			
4	Diet and physical activity, consider drugs; consider surgery			

Source: Table 8 of the CS[1]
Footnote: [a] for men, waist circumference of less than 94cm is low, 94–102cm is high and more than 102cm is very high. For women, waist circumference of less than 80cm is low, 80–88cm is high and more than 88cm is very high.
BMI = body mass index

The CS states that "*in NHS England, the initial standard of care is to advise lower-energy diets, increased physical activity and behavior modification. The exact nature of these treatments can vary in both style and intensity throughout NHS England and may be delivered by either dieticians, GPs or WeightWatchers®. For patients who have not achieved adequate weight loss (who have not reached their target weight loss, or who have reached a plateau) on such standard management, pharmacological treatment should be considered.*"[1]

In Section 3.3 of the CS it is stated that "*Currently in the EU, orlistat is the only available, orally effective, pharmacological product for weight management on the market; this is especially problematic given the complex aetiology of the disease across individuals (…)Due to its mechanism of action, orlistat is associated with several limitations, as detailed in Section 3.6. Therefore, the potential benefits of the addition of pharmacotherapy to standard management are not generally observed, as use of orlistat remains low."*[1]

In Section 3.3 of the CS it is reported that *"Surgery is only indicated for patients with a BMI ≥40kg/m^2 or between 35kg/m^2 and 40kg/m^2 with other significant disease, and who have failed all non-surgical measures, including intensive management in a Tier 3 service. Therefore, surgery should be considered a last resort for patients who have exhausted all other treatment options as seen by the limited number of surgeries conducted each year, and is therefore not considered an appropriate comparator to NB32, in line with the final scope for this submission."*[1]

The company states that *"NB32 can be used as an alternative first-line pharmacological treatment in patients for whom orlistat is contraindicated or is not utilized due to physician / patient choice, and patients who persevere with standard management despite the expected lack of effectiveness. NB32 should also be considered for patients who have not achieved adequate weight loss with orlistat treatment, or who did not comply with dietary requirements associated with orlistat, or were unable to tolerate orlistat treatment and who would otherwise revisit standard management measures."*[1]

ERG comment: The company provides an appropriate overview of the current provision of services in relation to overweight and obesity. However the following should be noted:

- Although the limitations of orlistat in terms of gastrointestinal adverse effects are appropriately highlighted, the CS does not provide data to support that the use of orlistat remains low. In England in 2014, pharmacies dispensed just over half a million items for treating obesity with a net ingredient cost of £15.3 million. All of these prescriptions were for orlistat.[3]
- Surgery provides better long-term outcomes for the morbidly obese (BMI>40).[5] A total of 6,032 bariatric surgery procedures (1,444 in male and 4,588 in women) were carried after a diagnosis of obesity in the year 2014-2015.[3]
- The ERG notes that NB32 is placed at first line in the clinical pathway as an alternative to orlistat and at second line in the pathway for those who have previously taken orlistat unsuccessfully. However in none of the main trials have patients previously taken orlistat.

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 3.1: The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG Comment
Population	Adults who have a BMI of: ≥30kg/m^2 (obese) or ≥27kg/m^2 to <30kg/m^2 (overweight) in the presence of one or more weight-related co-morbidities	Adults who have a BMI of: ≥30kg/m^2 (obese) or ≥27kg/m^2 to <30kg/m^2 (overweight) in the presence of one or more weight-related co-morbidities	-	In line with the scope of the decision problem.
Intervention	Naltrexone-bupropion prolonged-release	Naltrexone-bupropion prolonged-release	-	In line with the scope of the decision problem. Note also that, in fact the intervention is an add-on to standard management.
Comparator (s)	Standard management without naltrexone-bupropion Orlistat (prescription dose)	Standard management without naltrexone-bupropion Orlistat (prescription dose)	-	In line with the scope of the decision problem. However, it is not clear what is meant by "Standard management without naltrexone-bupropion".
Outcomes	BMI Weight loss Percentage body fat Waist circumference Incidence of Type 2 diabetes Cardiovascular events Mortality Adverse effects of treatment Health-related quality of life	Weight loss Percentage body fat Waist circumference Incidence of Type 2 diabetes Cardiometabolic parameters Mortality Adverse effects of treatment Health-related quality of life	Key outcomes captured in pivotal trial programme	BMI and percentage body fat are not reported in the CS. The data on cardiovascular events are limited.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG Comment
Economic analysis	The cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared Costs will be considered from an NHS and Personal Social Services perspective	The cost effectiveness of treatments is expressed in terms of incremental cost per quality-adjusted life year The time horizon for estimating clinical and cost effectiveness reflects the lifetime of patients Costs are considered from an NHS and Personal Social Services perspective	-	In line with the scope of the decision problem.
Subgroups to be considered	People with Type 2 diabetes	People with Type 2 diabetes; the COR-DM study provides data for this subgroup	-	In line with the scope of the decision problem.
Special considerations including issues related to equity or equality	None specified	None specified	-	-

Source: CS[1]

BMI = body mass index; CV = cardiovascular; NB32 = naltrexone 32mg plus bupropion; NICE = National Institute for Health and Care Excellence; T2DM = type 2 diabetes mellitus

3.1 Population

The population is described in the scope as "Adults who have a BMI of:

- ≥ 30kg/m2 (obese) or
- ≥ 27kg/m2 to < 30kg/m2 (overweight) in the presence of one or more weight-related co-morbidities."[7]

The population in the Company Submission (CS) matches the scope.

3.2 Intervention

The intervention is described in the scope as naltrexone-bupropion prolonged-release. This is the same in the CS.

The indication for naltrexone-bupropion prolonged-release (32mg daily) or NB32 (UK brand name: Mysimba) is as follows:

"Mysimba is indicated, as an adjunct to a reduced-calorie diet and increased physical activity, for the management of weight in adult patients (≥18 years) with an initial Body Mass Index (BMI) of

- *≥30kg/m^2 (obese), or*
- *≥27kg/m^2 to <30kg/m^2 (overweight) in the presence of one or more weight-related co-morbidities (e.g., Type 2 diabetes, dyslipidaemia, or controlled hypertension)"*[1]

Treatment with Mysimba should be discontinued after 16 weeks if patients have not lost at least 5% of their initial body weight.[1] In most trials NB32 is continued throughout the trial, usually one year duration. The company states that "*For patients continuing treatment post 16 weeks, treatment should be continued as long as clinical benefit is observed*."[1] It is unclear how long treatment duration would be in clinical practice.

Mysimba is orally administered. Each tablet contains 8mg naltrexone and 90mg bupropion hydrochloride. Dose should be escalated for the first four weeks as follows:

- Week 1: One tablet in the morning
- Week 2: One tablet in the morning and one tablet in the evening
- Week 3: Two tablets in the morning and one tablet in the evening
- Week 4 and onwards: Two tablets in the morning and two tablets in the evening

In Section 2.3 Table 6 of the CS there is a statement "*Retreatment with NB32 is not routinely anticipated and thus not modelled.*"[1] The company was asked to justify why patients would not be retreated with naltrexone-bupropion for any subsequent weight gain after a successful treatment with the drug.[8] The company replied "*There are no data to indicate the effectiveness of retreatment with NB32 following successful treatment with NB32 and subsequent discontinuation and weight regain. If NICE thinks this is likely to happen in practice, an option for NICE is to consider that the current cost-effectiveness model assumes the same analysis for patients independent of whether they have received previous NB32 or not. Clinical rationale can inform the likelihood of retreatment success until evidence merges.*"[9]

3.3 Comparators

The comparators described in the scope are 'Standard management without naltrexone-bupropion' and 'Orlistat (prescription dose)'. These are the same in the submission.

However, the NICE scope does not specify what is meant by 'Standard management without naltrexone-bupropion'.

According to the CS, standard management consisted of customary diet and behaviour modification in three of the four main trials (COR-I, COR-II and COR-DM; CS, page 16-17).[1] In these three trials at baseline, weeks 12, 24, 26 and 49 (4, 16, 28 and 40 for COR-DM) patients received instructions to follow a hypocaloric diet (500 kcal/day deficit) and increase physical activity, and written behaviour modification advice.

In response to the ERG, the company stated that in the COR-I and COR-II studies "*Patients were encouraged to increase physical activity, with a prescription for walking starting with at least 10 minutes on most days of the week, and increasing this gradually to 30 minutes on most days of the week throughout the study. They were encouraged to lose weight and maintain weight loss, and were encouraged to follow the prescribed programme (as described). Participation in any other weight loss programme was not permitted. The use of meal replacements (such as Slim Fast® or Weight Watchers®) was discouraged, but occasional use did not necessitate withdrawal from the study. The prescribed exercise could be performed in a gymnasium or health club.*"[9]

In COR-DM "*Patients were encouraged to increase physical activity, with a prescription for walking at least 30 minutes three times per week. Patients were encouraged to follow the prescribed programme. Participation in any other weight loss programme was not permitted. The use of meal replacements (such as Slim Fast® or Weight Watchers®) was discouraged, but occasional use despite contrary instructions did not necessitate withdrawal from the study. The prescribed exercise could be performed in a gymnasium."*[9]

In COR-BMOD standard management consisted of intensive behaviour modification. According to information provided by the company, it included "*three components: dietary instruction, closed group sessions, and prescribed exercise*".

BMOD consisted of group meetings lasting 90 minutes weekly for the first 16 weeks, every other week for the next 12 weeks and monthly thereafter. They included instructions to consume a balanced deficit diet and to increase to 180 min/week of planned, moderately vigorous, physical activity (CS, page 57).

In the COR-I and COR-II trials participants were not permitted to engage in weight loss programmes other than the prescribed programme of diet modification and exercise advice. This represents a more minimal approach to standard management than might be expected in practice. The COR-BMOD trial could be seen as best practice for standard management in that a more intensive intervention was delivered. Group sessions were included as well as dietary instruction and prescribed exercise. The choice of standard management has implications for the effectiveness and cost effectiveness of NB32 and these will be highlighted in this report.

The marketing authorisation for orlistat (UK brand name: Xenical) states that "The Committee for Medicinal Products for Human Use (CHMP) decided that Xenical's benefits are greater than its risks in conjunction with a mildly hypocalorific diet for the treatment of obese patients with a BMI greater or equal to 30 kg/m2, or overweight patients (BMI ≥28 kg/m2) with associated risk factors. The Committee recommended that Xenical be given marketing authorisation."[10] Orlistat comes as a capsule (120mg) to be taken three times a day.

The marketing authorisation further states that: "Xenical is given as one capsule taken with water just before, during, or up to one hour after each main meal. If a meal is missed or contains no fat, Xenical should not be taken. The patient should be on a diet in which about 30% of the calories come from fat,

and which is rich in fruit and vegetables. The food in the diet should be spread over three main meals. Treatment with Xenical should be stopped after 12 weeks if patients have been unable to lose at least 5% of their body weight since the start of treatment."[10]

In response to the draft scope, the Royal College of Physicians (RCP) pointed out that "The comparators seem reasonable but there is no direct head to head comparison with orlistat. Most of the trials with orlistat were conducted over 20 years ago, so caution should be exercised when making indirect comparisons; this is particularly true for conditions such as diabetes where background standard therapy (for glucose and lipids especially) may be very different now."[11] In addition, the RCP stated that "There is very little data on some ethnic groups (particularly people from Asia) within the trials with Naltrexone-Bupropion, so it may not be possible to make any firm conclusions for that group."[11]

3.4 Outcomes

None of the NB32 trials report BMI. In the CS this is explained as follows (CS, page 51):

"*Of note, change in BMI was not a pre-defined endpoint. Although this is an adequate research tool, it is limited in the assessment of an individual, as it does not consider different body morphologies (e.g. muscle vs adipose) and may be skewed by very high muscle mass.[12] In addition, some population groups, such as people of Asian family origin and older people, have comorbidity risk factors that are of concern at different BMIs (lower for adults of an Asian family origin and higher for older people).*[5] *Therefore, alternative methods to measure body fatness, such as waist circumference, were utilised in the trials.*"[1]

However, NICE Clinical Guideline (NICE CG189, 2014[5]) states that:

- BMI should be used as a practical estimate of adiposity in adults
    - BMI should be interpreted with caution; waist circumference may be used in addition for patients with BMI <35kg/m2
    - Bioimpedance should not be used
- BMI should be interpreted with caution in muscular adults
    - Other populations, such as Asians and older patients, have comorbidity risk factors that are of concern at different BMIs
- Assessment of health risks associated with being overweight or obese should be based on BMI and waist circumference

Furthermore, BMI could easily have been calculated from data available in the trials.

In addition, 'cardiovascular events' are not reported in the CS. Instead the CS reports 'cardiometabolic parameters'. The FDA requested a trial: NB-CVOT to examine the risk of cardiovascular events, but this trial was terminated early. Where cardiovascular events are reported in the CSRs we will add them to this report.

3.5 Other relevant factors

No special considerations, including issues related to equity or equality, were specified (CS, page 14). A patient access scheme is not mentioned in the submission.

4. CLINICAL EFFECTIVENESS

The company conducted a systematic review to identify studies of NB32 and potential comparator therapies to treat adults who are overweight or obese. In Section 4.1 we critique this review.

4.1 Critique of the methods of review(s)

The systematic review conducted by the company formed the source of studies for both the NB32 direct evidence and the indirect treatment comparison between NB32 and orlistat. It was used to inform both efficacy and adverse event data.

4.1.1 Searches

The following paragraphs contain summaries and critiques of all searches related to clinical effectiveness presented in the company submission. The Canadian Agency for Drugs and Technologies in Health (CADTH) evidence based checklist for the Peer Review of Electronic Search Strategies (PRESS), was used to inform this critique.[13] The submission was checked against the Single Technology Appraisal (STA) specification for company/sponsor submission of evidence.[14] The ERG has presented only the major limitations of each search strategy in the report.

The company submission stated that systematic review searches were undertaken in May 2016. Search strategies were reported in Appendix 2 of the CS for the following databases: Embase, MEDLINE, MEDLINE in-Process, Cochrane's CENTRAL, DARE and CDSR.

Additional searches of the following conference proceedings were reported for the last two years: International Congress on Obesity (ICO), European Congress on Obesity by the European Association for the Study of Obesity (ECO), American Diabetes Association (ADA), International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Annual European Congress and ISPOR Annual International Congress. In their response to clarification the company confirmed that the conference searches were conducted in June 2016 and provided the search terms.[9]

The CS also reported that the reference lists of existing systematic literature reviews and meta-analyses were checked for additional studies not identified by the main searches.

Searches utilised study design filters based on the Scottish Intercollegiate Guidelines Network (SIGN) Embase filters for RCTs, Observational Studies and Systematic Reviews.[15]

ERG comment:

- The database searches were clearly structured and documented. No language limits were applied.
- In their response to clarification the company confirmed that they searched Embase and MEDLINE simultaneously using a single database provider (Embase.com) and search strategy. This approach has limitations when using subject heading terms which could affect recall of results. Embase subject heading terms (Emtree) were used in the search strategy, and although simultaneous searching of Embase.com should automatically identify and search for equivalent MEDLINE subject heading terms (MeSH), it is not clear if this is the case for all potentially useful MeSH terms. Given the potential limitations of this approach, the ERG considered it preferable to search each database separately, or at least to ensure inclusion of both Emtree and MeSH terms in the search strategy.
- The ERG noted the use of study design filters in the Cochrane Library searches of CDSR, DARE and CENTRAL. It was considered that this was an overly restrictive approach given that these resources are already filtered by study design. Of particular concern was the search

of CENTRAL, which when rerun by the ERG yielded approximately 65 additional results without the study design filters. The ERG requested that the company rerun this search and screen these additional papers to confirm that no relevant papers had been missed. In their response to clarification the company responded "*Searches were conducted again by applying the CENTRAL limit in the Cochrane Library instead of using the study design filters, as was done originally. This found only five additional unique papers from which three were deemed relevant. However, these three potentially relevant studies were published after June 2016, when the original searches were conducted. As such, no additional studies were included from this approach*."[9]

- Section 4.10.1 stated "The search strategy used to identify RCT evidence for NB32 and orlistat 120mg TID is described in Section 4.1."[1], therefore the same limitations as described above will have applied.
- No mention was made in Section 4.12 of the company submission with regard to how adverse events data were identified. The ERG queried this omission and asked for confirmation that the results of searches reported in Appendix 2 of the CS were screened for adverse events. Guidance by the Centre for Reviews and Dissemination (CRD)[16] 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. The company responded: "*No additional searches to those reported in Section 4.1 and Appendix 2 were conducted to identify adverse event (AE) data, but results retrieved were screened for AEs*." This issue is further discussed in Section 4.1.2 of this report.

4.1.2 Inclusion criteria

The eligibility criteria used in the search strategy of the review for randomised controlled trials (RCTs) and non-RCTs is presented in Table 4.1.

Table 4.1: Eligibility criteria for trials to be included in the systematic review

Criteria	Inclusion criteria	Exclusion criteria
Population	Adults who are obese (BMI ≥30kg/m2) or overweight, according to one of the following definitions: • 25kg/m2 to 29.9kg/m2 • ≥27kg/m2 to <30kg/m2 • >28kg/m2 with one or more weight-related comorbidity (T2DM, dyslipidaemia and/or controlled hypertension)	Healthy volunteers Children (age <18 years) Diseases other than that specified in inclusion criteria
Study design	RCTs Non-RCTs Systematic reviews and meta-analyses of RCTs[a]	*In vitro* studies Preclinical studies Comments, letters, editorials Case reports, case series Non-systematic reviews Observational studies
Intervention	Studies assessing at least one of the following interventions will be included: Naltrexone-bupropion Orlistat	Studies that do not assess at least one of the included interventions will be excluded

Criteria	Inclusion criteria	Exclusion criteria
Comparator	Comparator therapies may include one of the following: Behavioural interventions Lifestyle or dietary modifications Any treatment listed under the interventions Any other pharmacological treatments for obesity or weight management	Studies will not be excluded on comparator therapy if it includes at least one of the treatments listed under the interventions
Study duration	All trials with total randomised phase duration >1 year are included	Studies with <1-year duration
Language	Studies published in English were included Studies published in non-English languages were flagged	Studies will not be excluded on the basis of publication language
Source: Table 10 of the CS[1] Footnote: a, Systematic reviews and meta-analyses of RCTs were identified and flagged. Bibliographies of these systematic reviews will be screened to check if literature searches have missed any potentially relevant studies. BMI = body mass index; RCT = randomised controlled trial; T2DM = Type 2 diabetes mellitus.		

ERG comment:

- Although non-RCT studies were eligible for inclusion, they were not considered further once sufficient RCTs were found. The company was asked to clarify the exclusion of non-RCTs. They responded that "*Both RCTs (randomised controlled trials) and non-RCTs were identified through SLR (systematic literature review), and screened for AEs. However, non-RCT evidence was not formally considered as part of comparative safety assessments as RCT data were available for the intervention and comparators of interest to the decision problem. This included longer-term safety data to that available from the pivotal trial programme*." Although this may be acceptable for effectiveness data, it is not normally acceptable for adverse events. Non-RCT studies should have been assessed for long-term follow-up and reporting of rare adverse events.[16] Additionally, bibliographic details of the nine non-RCTs should have been provided. However, in the case of this technology assessment, the ERG did not find any relevant non-RCT studies of NB32 that were missed or inappropriately excluded.
- The inclusion criteria state that "*Studies published in non-English languages were flagged*."[1] The company was asked to clarify the methods for dealing with these studies and responded "*Non-English language studies were to be included if sufficient evidence from English language articles was not available. In light of the completeness of English language RCTs, all non-English language studies were excluded*."[9] The ERG noted that 44 full text articles were excluded but a complete list of these articles was not provided so it was not possible to ascertain if any relevant non-English language studies had been excluded.
- Studies that do not assess at least one of the included interventions (naltrexone-bupropion or orlistat) were excluded. This means that studies comparing a behavioural intervention with placebo (or waiting list control) have been excluded as were studies comparing different types of behavioural interventions. According to the scope, these studies should have been included; this would have allowed an indirect comparison of naltrexone-bupropion versus different types of behavioural interventions.

4.1.3 Critique of data extraction

The company stated that "*All relevant data were extracted from the included full text of articles by one reviewer and quality checked against the original source by a second reviewer*".[1]

ERG comment: Although the company stated that two reviewers were involved in the data extraction of included studies, it was unclear how discrepancies were resolved (e.g. use of a third reviewer). Although it is good practice to include this detail when reporting a systematic review, we believe that overall the data extraction was carried out appropriately.

4.1.4 Quality assessment

The CS stated that quality assessment of included studies was done "in accordance with the NICE-recommended checklist for RCT assessment of bias".[1] Elements assessed were randomisation, allocation concealment, comparability of groups, blinding of care providers, patients and outcome assessors and drop out, selective reporting of outcomes and use of intention to treat analysis and appropriate methods for dealing with missing data.

ERG comment: Study quality appeared to have been assessed using appropriate tools.

4.1.5 Evidence synthesis

Two types of evidence synthesis are described in the CS: a meta-analysis of the NB32 trials and an indirect comparison comparing NB32 with orlistat.

The meta-analysis

To compare and pool the relative treatment effects between the four trials comparing NB32 and placebo (COR-I, COR-II, COR-BMOD and COR-DM), a frequentist pairwise meta-analysis was performed to assess the following outcomes:

- At least a 5% reduction in weight at one year from baseline (the one year time point ranged from 52 to 57 weeks). This was a dichotomous outcome.
- Mean percentage weight change from baseline at one year (the one year time point ranged from 52 to 57 weeks). This was a continuous outcome.

The ERG asked for clarification as to why these outcomes had been selected for the meta-analysis. The company responded that "*The outcome of 5% reduction in weight from baseline was incorporated as per the European Medicines Agency (EMA) licence and associated treatment stopping rules; whereas the mean % weight change from baseline was incorporated to account for the overarching treatment effect of each regimen. Meta-analysed results for alternate outcomes were not required for the de novo model, and were therefore not produced*."[9]

The NB-CVOT study was excluded from all meta-analyses, due to the trial design, objective, and patient population, being different from the other studies.

The frequentist pairwise meta-analysis was performed using R (version 3.3.1) using the metafor package.[17, 18] The pairwise meta-analysis, presents relative treatment effects per trial, and an overall 'pooled' relative treatment effect for placebo versus NB32 which was calculated using a random effects model.[19] To further evaluate the trial-heterogeneity, sensitivity analyses were performed for the three non-T2DM trials, and for the non-T2DM trials excluding the COR-BMOD trial, as patients received intensive behaviour modification. The statistical heterogeneity of the pairwise meta-analysis was assessed using I^2, where the I^2 value describes the percentage of total variation across studies that is due to heterogeneity rather than chance.[20] The mITT populations were used in the meta-analyses. Results for the number with ≥ 5% reduction in weight (binary outcome) were reported as odds ratios (OR) and

results for the mean percentage change in weight from baseline (continuous outcome) were reported as mean differences (MD) both with 95% confidence intervals (CI).

ERG comment: The meta-analyses used appropriate statistical methods. Only two outcomes were included in the meta-analysis and both were measures of weight loss, these were also the two co-primary outcomes in the COR trials. The company stated that other outcomes were not meta-analysed as they were not needed for the economic model. This seems to be reasonable as the other outcomes were reported for the COR trials, and given the heterogeneity in terms of populations and background therapy (see below) additional meta-analyses may not have been appropriate.

Subgroup analyses were performed to explore the heterogeneity by splitting the studies into those containing only type 2 DM (T2DM) patients and those excluding type 2 DM patients. The standard management received in the COR-BMOD trial was more intensive than in the COR-I and II trials. The CS states that "*these differences (the presence or absence of T2DM and the intensity of the diet and exercise programme) between the trial designs are likely to explain the heterogeneity in results between the four trials*" (CS, Section 4.91, page 111). The ERG agrees that there was clinical and statistical heterogeneity between the four COR trials and that because of this the results from the separate analyses for T2DM and no T2DM should be used.

The indirect comparison

An indirect treatment comparison (ITC) was performed to compare NB32 with orlistat (120mg TID), using placebo as a common comparator.

ITC were performed to compare NB32 and orlistat for the following outcomes:

- Mean percentage weight change from baseline at one year (the one year time point ranged from 52 to 57 weeks [continuous outcome])
- At least 5% reduction in weight at one year from baseline (the one year time point ranged from 52 to 57 weeks [dichotomous outcome])

Some data imputations were required to maximise inclusion of evidence in the analyses, and the methods of imputation are described in Appendix 10 of the CS. The analysis used the mITT populations.

Odds Ratios (OR) were used as the effect size for ≥ 5% reduction in weight and mean differences (MD) for the mean percentage change in weight from baseline.

To investigate the effect of T2DM and to populate the economic model (in which results from the ITC were applied according to individual patient T2DM status), if data were available then all the analyses and sensitivity analyses were performed separately for:

- Trials where T2DM is part of the trial inclusion criteria (T2DM analysis)
- Trials where T2DM is part of the trial exclusion criteria (no T2DM analysis)
- All trials regardless of T2DM (any T2DM analysis)

To assess the effects of weight loss in trials where a large proportion of patients had comorbidities, a sensitivity analysis was performed excluding trials where ≥75% of patients ≥1 comorbidity (hypertension, dyslipidaemia, or T2DM). Due to anticipated heterogeneity with respect to the duration of and therapies received during the lead-in periods, sensitivity analyses were also performed excluding those trials incorporating lead-in periods.

The specific type and intensity of standard management varied between the trials, although treatment arms within the same trial received the same standard management. For the analysis, it was therefore

assumed that the additional treatment benefit from the standard management was additive but that the relative treatment effect between treatment arms would be unaffected. Further sensitivity analyses were performed excluding studies with 'intensive' behaviour modification, and excluding trials with lead-in periods or 'intensive' behaviour modification.

A Bayesian NMA was performed for each outcome using the available data (CS, Table 31 and Table 32). Markov Chain Monte Carlo (MCMC) methods were used which combine prior distributions with the data to construct a posterior distribution of parameters of interest upon which to base summary results. All models were fitted using WinBUGS (version 14),[21] via R (version 3.3.1).[17] An initial 50,000 iterations were discarded as the 'burn-in' period, which was assessed by running two chains using different starting values and assessing convergence using Brooks-Gelman-Rubin plots.[22] Then, 10,000 samples (posterior distribution) were used for obtaining summary estimates. In total, 10,000 samples were deemed sufficient for each of the different analyses as the Monte Carlo error was less than 5% of the standard deviation.[23] Therefore, the samples could be used directly in the economic model, preserving the correlation between treatment effects and avoiding the need to make assumptions regarding the shape of the posterior distribution. Autocorrelation was assessed to determine whether samples were highly correlated, a thinning interval of five was applied to ensure that the chain was mixing well and was representative of the posterior distribution. The goodness-of-fit was assessed using the total residual deviance and tested using a chi-squared test. Random effects and fixed effect models were used; however, random effect results are only presented for the 'any T2DM' analysis. Random effects results are not presented for the T2DM only and non-T2DM analyses, as the models failed to update effectively using the recommended priors, likely due to the low number of studies. The models and prior distributions used for the two outcomes were those described in the NICE Decision Support Unit (DSU) Technical Support Document (TSD) 2.[24]

ERG comment: The ERG re-ran the Bayesian NMA for both the binary and continuous weight loss outcomes and reproduced the results reported in the CS for the three analysis groups: T2DM, no T2DM and any T2DM patients. The modelling used the code supplied in the NICE Decision Support Unit (DSU) Technical Support Document (TSD) 2[24] and the analysis methods were appropriate. Model fit was tested and the results were reported. The decision to present only fixed effect model results for the T2DM and no T2DM subgroups was correct as the ERG also found that there were problems with model convergence for these models, especially for the T2DM analyses. The fixed effect model provided the best fit to the data and results that are likely to be more reliable. There was no need to evaluate inconsistency in the analyses as they were straightforward indirect comparisons between NB32 and orlistat using placebo as the common comparator. Appropriate sensitivity and subgroup analyses were used to explore differences resulting from the inclusion or exclusion of patients with T2DM and those trials using intensive behaviour modification as background therapy.

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

4.2.1 Overview of the evidence in the submission

The company identified 36 relevant RCTs. The CS stated that "*Of the 36 included RCTs, 5 studies investigated treatment with NB32…., while the remaining 31 studies investigated treatment with orlistat.*"[1] The five studies of NB32 will be discussed in this section and are listed in Table 4.2. The studies of orlistat were used to form an indirect comparison with NB32 and will be discussed in Section 4.4 of this report.

Table 4.2: List of relevant RCTs

Trial name	Population	Intervention[a]	Comparator
COR-I[25]	Adults with uncomplicated obesity or who were overweight with dyslipidaemia or hypertension	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32) OR Naltrexone 16 mg per day + bupropion 360 mg per day (NB16)	Placebo
COR-BMOD[26]	Adults with uncomplicated obesity or who were overweight with dyslipidaemia or hypertension	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32) + BMOD	Placebo + BMOD
COR-II[27]	Adults with uncomplicated obesity or who were overweight with dyslipidaemia or hypertension	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32)	Placebo
COR-DM[28]	Adults with T2DM and BMI $\geq$ 27 and $\leq$ 45 kg/m^2	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32)	Placebo
NB-CVOT[29]	Adults with a BMI of 27 to 50 and who had characteristics associated with an increased risk of CV outcomes [b]	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32)	Placebo

Source: Tables 11 and 12 and text of section 4.2 of the CS[1]
Footnote: a) Two tablets of NB32 or placebo were taken twice a day (each tablet contained 8mg naltrexone hydrochloride and 90mg bupropion hydrochloride) b) terminated early (after 50% interim analysis)
BMI = body mass index; BMOD = intensive behaviour modification; COR = Contrave obesity research; CV = cardiovascular; CVOT = cardiovascular outcomes trial; DM = diabetes mellitus; RCT = randomised controlled trial; T2DM = type 2 diabetes mellitus

A further trial, IGNITE, unpublished at the time of the systematic review, was identified by the company and presented as supporting evidence.[30]

The company stated that there were no relevant ongoing trials. However in the background section of the CS two trials were mentioned "*a further Phase IV study to assess the effect of NB32 on the occurrence of MACE in overweight and obese patients was requested. Data from this trial are due in 2022. The CHMP also requested additional assessment of the pharmacokinetics of NB32 in patients with renal impairment and in patients with hepatic impairment, as the submitted trials did not collect such data, nor did the Phase III programme allow a direct evaluation of safety in these patient groups. Such a trial is ongoing.*"[1]

The company was asked to provide details of these studies and to indicate if any interim data were available.[8] The company replied regarding the MACE study that "*Study synopsis is provided as an attachment. No information related to the new MACE study has been published or is available on any bibliographic database as it is currently still in the planning stage*".[9] The study is a multicentre,

randomised, double-blind, placebo-controlled study of the effect of NB32 on the occurrence of major adverse cardiovascular events (MACE) in overweight and obese adults with cardiovascular disease. Based in the US, the trial will aim to enrol 8,000 patients. It will have a lead-in period of two weeks, and a treatment period estimated to last for up to six years until the targeted number of adjudicated MACE events (378) has been reached. The primary MACE composite comprises the first occurrence of CV death, nonfatal myocardial infarction (MI), and nonfatal stroke.[31]

Regarding the trial of patients with renal impairment and patients with hepatic impairment, the company replied "*Study synopsis are provided as an attachment. As both the renal and hepatic impairment studies are small phase I studies requested by regulatory agencies, no information related to these studies have been published or made available on clinical study databases, such as clinicaltrials.gov.*"[9] Both studies aimed to enrol 32 to 48 participants. One was "*to evaluate the effect of hepatic impairment on the PK of naltrexone, bupropion, and their major active metabolites following a single oral dose of NB in subjects with varying degrees of hepatic function.*"[32] And the other was "*To evaluate the effect of renal impairment on the PK of naltrexone, bupropion, and their major active metabolites following a single oral dose of NB (total dose of 16 mg naltrexone and 180 mg bupropion) in subjects with varying degrees of renal function*."[33]

The company stated that "*This submission focuses on data from the four pivotal RCTs: COR-I, COR-II, COR-BMOD, and COR-DM with only longer-term efficacy and safety data used to predict maintenance of pivotal trial outcomes presented from the NB-CVOT study and supported with data from the IGNITE study*."[1] Accordingly, the four pivotal RCTs: COR-I, COR-II, COR-BMOD, and COR-DM will be discussed in some detail in Section 4.2.2 of this report whilst NB-CVOT and IGNITE will be discussed more briefly in Section 4.2.7.

All trials included patients who were obese or overweight with comorbidities. COR-I, COR-II and COR-BMOD excluded patients with diabetes but in COR-DM all patients had type two diabetes.

None of the trials compared NB to orlistat, a comparator specified in the NICE scope.[7] All the main trials compared NB32 to placebo. COR-I also included a treatment arm where patients received NB16.[25]

In both arms of the trials patients received customary diet and behaviour modification. According to the CS "*This included a hypocaloric diet (500 kilocalorie [kcal] per day deficit based on the World Health Organization [WHO] algorithm for calculating resting metabolic rate) as well as instructions on increasing physical activity (COR-I and COR-II), or more intensive behaviour modification counselling (COR-BMOD).*" This represents '*standard management without naltrexone-buproprion*' as specified in the NICE scope.[7] More detail is provided in Section 3.3 of this report.

ERG comment:

- The CS appropriately focuses on the four main NB32 RCTs. However these all compare NB32 to placebo with both arms receiving standard care. The ERG draws to the attention of the committee that no trials directly compare NB32 to orlistat as specified in the NICE scope.[7]
- The ERG also notes that standard care varies between the trials in that COR-BMOD has a more intensive form of behavioural management.
- The ERG confirms that evidence from the ongoing trials could not have been incorporated into the CS. However the ERG draws the attention of the committee to the ongoing MACE trial.[31]
- The ongoing investigations into patients with renal or hepatic impairment are drawn to the attention of the committee. Currently as stated in the CS, "*Patients with end-stage renal failure*

or severe renal or hepatic impairment are listed as a contraindicated patient population in the Mysimba SmPC."[1]

4.2.2 Overview of the direct evidence

This section focuses on the four main trials: COR-I, COR-II, COR-BMOD and COR-DM. Further details of their design can be found in Table 4.3.

Table 4.3: Trial designs of included NB32 studies

Trial name	Location	Number of participants	Trial design and duration	Primary outcome
COR-I[25]	34 study sites in the US	1,742	Phase III, multicentre, randomised, double-blind placebo-controlled 56 week study	Percentage of change in total body weight and proportion of patients with ≥ 5% decrease in total body weight at week 56
COR-BMOD[26]	9 study sites in the US	793	Phase III, multicentre, randomised, double-blind placebo-controlled 56 week study	Percentage of change in total body weight and proportion of patients with ≥ 5% decrease in total body weight at week 56
COR-II[27]	36 study sites in the US	1,496	Phase III, randomised, parallel-arm, double-blind, placebo-controlled, 56 week study	Percentage of change in total body weight and proportion of patients with ≥ 5% decrease in total body weight at week 28
COR-DM[28]	53 study sites in the US	505	Phase III, multicentre, randomised, double-blind placebo-controlled 56 week study	Percentage of change in total body weight and proportion of patients with ≥ 5% decrease in total body weight at week 56

Source: Table 12 of the CS[1]
BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus; US = United States

All trials were multicentre and all were conducted in the US. All had a joint primary outcome of percentage change in total body weight and proportion of patients with >5% decrease in total body weight. Three trials measured outcomes at week 56.[25, 26, 28] One trial, COR-II measured the primary outcome at 28 weeks. In COR-II, NB32 patients who had lost less than 5% of their body weight at visits between weeks 28 and 44 were re-randomised to continue with NB32 or escalate to NB48.

The four main trials included 4,536 patients. Of these 2,510 patients were randomised to NB32, 578 to NB16 (in COR-I) and 1,448 randomised to placebo.

ERG comment:

- As all of the trials were conducted in the US, participant characteristics may not reflect a UK population particularly in terms of ethnicity. Patient characteristics will be discussed later in this section.
- As all of the trials were conducted in the US, standard care may differ from a UK setting. Differences between the trials in terms of standard care have already been highlighted in Section 3.3 of this report.
- It is also possible that standard care varied within the trials given the number of centres (34 centres for COR-I, 36 for COR-II, nine for COR-BMOD and 53 for COR-DM).
- Three trials measure the primary outcome at 56 weeks. Although this is acceptable in terms of weight loss, there is no information on maintenance of weight loss after this time. The CS states that "*For patients continuing treatment post 16 weeks, treatment should be continued as long as clinical benefit is observed*."[1] It is unclear how long patients would continue to take the drug in practice.
- The licensing for NB32 indicates that it should be discontinued after 16 weeks if patients have not lost at least 5% of their initial body weight. However the main trials do not incorporate this stopping rule as the licensing was based on evidence found in the trials. The company stated "*As a result of pooled, post-hoc analyses of the COR trials that showed a strong relationship between early weight loss and clinically meaningful longer term weight loss, the license terms for NB32 include more prescriptive discontinuation rules*."[9]
- The ERG notes that the primary outcome includes ≥ 5% decrease in total body weight. The CHMP recommended investigation of ≥10% and this outcome is presented in the submission for individual trials but not for the meta-analyses. The results section of this report will also present results for patients with ≥ 10% weight loss.

Participant inclusion and exclusion criteria are shown in Table 4.4.

Table 4.4: Participant inclusion and exclusion criteria in the NB32 trials

Trial name	Patient age	Patient BMI	Includes patients with diabetes?	Main exclusion criteria relating to obesity
COR-I[25] COR-II[27] COR-BMOD[26]	18 to 65	BMI 30 to 45 kg/m2 and uncomplicated obesity OR BMI 27 to 45 kg/m2 and controlled hypertension and / or dyslipidaemia	No	Any anorectic or weight loss agents Participated in a weight loss management program concurrent to trial (COR-I and II) or within one month prior to randomisation(COR-BMOD) Weight change of > 4 kg within 3 months prior to randomisation Obesity of known endocrine origin History of surgical or device intervention for obesity History of treatment with, hypersensitivity or intolerance to bupropion or naltrexone
COR-DM[28]	18 to 70	BMI 27 to 45 kg/m2	Yes, all had T2DM	Type 1 diabetes Any anorectic or weight loss agents Obesity of unknown endocrine origin other than DM Loss or gain of > 5 kg within 3 months prior to screening Participated in a weight loss management program within one month prior to randomisation History of surgical or device intervention for obesity Treatment with bupropion or naltrexone within 12 months prior to screening
Source: Table 12 of the CS[1] BMI = body mass index; BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus; T2DM = type 2 diabetes mellitus				

Participant inclusion criteria for age and BMI are similar across the four main trials. As previously mentioned, one trial was conducted exclusively in patients with type 2 diabetes mellitus[28] whilst the other three excluded patients with diabetes. All trials included patients with a relatively stable weight and excluded obesity of endocrine origin. Other exclusions were patients were prior use of any anorectic or weight loss agents and those with a history of surgery or device intervention.

ERG comment:

- The ERG notes that evidence for diabetic patients was based on one trial of 505 participants.
- Inclusion and exclusion criteria appear to be reasonable for the main trials. The ERG draws to the attention of the committee that prior use of orlistat was an exclusion criterion in all four COR trials. Therefore the effect of NB32 on those who have failed on orlistat has not been examined.

Participant characteristics are displayed in Table 4.5.

Table 4.5: Participant characteristics in the NB32 trials

	COR-I[25]		COR-BMOD[26]		COR-II[27]		COR-DM[28]	
	NB32	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
Age, mean years (SD)	44.4 (11.1)	43.7 (11.1)	45.9 (10.4)	45.6 (11.4)	44.3 (11.2)	44.4 (11.4)	54.0 (9.1)	53.5 (9.8)
Age range (min, max)	19, 65	18, 66	19, 65	19, 64	18, 65	18, 65	20, 72	27, 70
Sex, female, n (%)	496 (85)	496 (85)	528 (89.3)	185 (91.6)	847 (84.6)	420 (84.8)	195 (58.2)	90 (52.9)
Ethnicity, n (%)								
White	440 (75)	440 (76)	405 (68.5)	149 (73.8)	835 (83.4)	414 (83.6)	261 (77.9)	140 (82.4)
Black	106 (18)	110 (19)	145 (24.5)	44 (21.8)	133 (13.3)	72 (14.5)	63 (18.8)	18 (10.6)
Asian	6 (1.0)	4 (0.7)	6 (1.0)	2 (1.0)	12 (1.2)	4 (0.8)	7 (2.1)	5 (2.9)
Other	31 (5.4)	27 (4.6)	35 (6.0)	7 (3.5)	21 (2.1)	5 (0.8)	4 (1.2)	7 (4.1)
BMI, mean kg/m^2 (SD)	36.1 (4.4)	36.2 (4.0)	36.3 (4.2)	37.0 (4.2)	36.2 (4.5)	36.1 (4.3)	36.4 (4.8)	36.4 (4.5)
Obesity class, n (%)								
BMI < 30 kg/m^2	18 (3.1)	5 (0.9)	8 (1.4)	1 (0.5)	25 (2.5)	14 (2.8)	18 (5.4)	11 (6.5)
BMI ≥30 and <35 kg/m^2	224 (38.4)	217 (37.3)	207 (35.0)	64 (31.7)	398 (39.8)	186 (37.6)	111 (33.1)	49 (28.8)
BMI ≥35 and <40 kg/m^2	204 (35.0)	229 (39.4)	230 (38.9)	79 (39.1)	316 (31.6)	191 (38.6)	110 (32.8)	64 (37.6)
BMI ≥40 kg/m^2	137 (23.5)	130 (22.4)	146 (24.7)	58 (28.7)	262 (26.2)	104 (21.0)	96 (28.7)	46 (27.1)
Other, n (%)								
Weight, mean kg (SD)	99.7 (15.9)	99.5 (14.3)	100.2 (15.4)	101.9 (15.0)	100.3 (16.6)	99.2 (15.9)	104.2 (18.9)	105.1 (17.0)
Smoker, n (%)	65 (11)	65 (11)	0*	0*	108 (10.8)	52 (10.5)	38 (11.3)	15 (8.8)
Hypertension, n (%)	130 (22)	113 (19)	86 (14.6)	37 (18.3)	212 (21.2)	106 (21.4)	212 (63.3)	103 (60.6)
Dyslipidaemia, n (%)	284 (49)	288 (50)	270 (45.7)	81 (40.1)	560 (55.9)	263 (53.1)	280 (83.6)	145 (85.3)
Alcohol use, n (%)	254 (43.6)	244 (42)	251 (42.5)	100 (49.5)	462 (46.2)	217 (43.8)	96 (28.7)	69 (40.6)
History of depression	66 (11.3)	73 (12.6)	83 (14.0)	31 (15.3)	131 (13.1)	76 (15.4)	29 (8.7)	14 (8.2)
History of anxiety	29 (5.0)	18 (3.1)	19 (3.2)	7 (3.5)	47 (4.7)	30 (6.1)	10 (3.0)	9 (5.3)
Statin use	11.5	8.6	9.1	8.4	11.7[$]	13.1	49.3	45.9

Source: Table 15 of the CS[1] and CSRs for COR-I[34], COR-BMOD[35], COR-II[36] and COR-DM[37]

Footnote: *Only non-smokers were eligible for the COR-BMOD trial. $ Includes NB48

BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus

Mean age was approximately 44 years apart from in COR-DM where participants were older (mean age 54)[28] The majority of participants were female although COR-DM had a more even distribution of female and male participants.[28] The majority of participants across the trials were of white ethnicity. Approximately 15% of participants were Black or African American. Just 1% were of Asian origin. Participants in the trials tended to be obese rather than overweight with an average BMI of 36 to 37. Approximately 2% had a BMI of < 30 (overweight). Hypertension was present in approximately 20% of patients across the COR trials although in COR-DM as expected over 60% had hypertension. Similarly, dyslipidaemia was present in approximately half of participants in the COR trials but in approximately 84% of the COR-DM patients.

ERG comment:

- Overweight patients in addition to obese patients were included in the NICE scope.[7] However there is only a very small percentage of patients who are overweight in the trials. Therefore the ERG draws to the attention of the committee that this population is not well represented.
- The majority of participants in the trials are female. The ERG draws to the attention of the committee that this does not reflect the distribution of obesity according to gender. Men in England are more likely to be overweight or obese (68% vs 58% in 2015).[3]
- The ERG draws to the attention of the committee that Asian patients are not well represented in the trials so results may not be applicable to these ethnic groups.

4.2.3 Direct evidence: Quality assessment

Table 4.6 presents the company's quality assessment of the four main trials with comments from the ERG.

Table 4.6: Quality assessment of included NB32 trials

Study question	Company's assessment of risk of bias				ERG comments
	COR-I[25]	COR-BMOD[26]	COR-II[27]	COR-DM[28]	
Was randomisation carried out appropriately?	Low	Low	Low	Low	Methods in all trials were appropriate.
Was the concealment of treatment allocation adequate	Low	Low	Low	Low	Methods in all trials were appropriate.
Were the groups similar at the outset of the study in terms of prognostic factors, for example severity of disease?	Low	Low	Low	Low	Methods in all trials were appropriate.
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Low	Low	Low	Low	Methods in all trials were appropriate.
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Low	Low	Low	Low	Company noted that more patients dropped out of NB32 groups due to adverse effects.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	Low	Low	Low	Low	Methods in all trials were appropriate.
Did the analysis include an intention-to-treat-analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Low	Low	Low	Low	Analyses were performed on the modified ITT for all trials. ITT was included as a sensitivity analysis. All those in the ITT analysis had to have a post-baseline body weight measurement during the treatment phase and for the mITT the measurement was while on the drug.
Was statistical powering such to detect a significant difference between treatment groups?	Low	Low	Low	Low	Methods in all trials were appropriate.
Source: CS Appendix tables 3 to 6[38] BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus					

ERG comment:

- Apart from the mITT analyses, the ERG agrees that the four main trials were of high quality and attempts were made to lower the risk of bias. Methods of analysis relating to intention-to-treat will be discussed below. In terms of dropout, the ERG noted that more patients dropped out of NB32 groups due to adverse events. The ERG was concerned that higher rates of adverse events (especially nausea – see Section 4.2.5 of this report) in the intervention arm could have resulted in un-blinding of participants.

The main results presented in the CS were based on a modified intention-to-treat (mITT) analysis. According to the CS this was defined as "*all randomised patients with a post-baseline body weight measurement obtained while the patient remained on study medication*". Missing data were imputed using the LOCF method for primary analysis.[1]

Table 4.7 presents the numbers of patients randomised and the numbers included in the mITT analysis for each trial.

Table 4.7: Randomisation and analysis sets

	No randomised		**Company's 'Modified ITT analysis set'¶**	
Trial name	NB32	Pbo	NB32 n (% of randomised)	Pbo n (% of randomised)
COR-I[25]	583	581	471 (80.8)	511 (88.0)
COR-BMOD[26]	591	202	482 (81.6)	193 (95.5)
COR-II[27]	1001	495	825 (82.4)	456 (92.1)
COR-DM[28]	335	170	265 (79.1)	159 (93.5)
Source: Section 4.5 of the CS[1] and Appendix 5 of the CS[38] Footnote: ¶ All randomised patients with a post-baseline body weight measurement obtained while on study drug BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus				

It can be seen that using the modified ITT analysis loses approximately 19% of the patients randomised to NB32 and 9% of those allocated to placebo.

ERG comment:

- The main results presented in the company submission were based on a modified intention-to-treat (mITT) analysis. This analysis includes only those patients who have a baseline and at least one post-baseline measurement whilst on the study drug. Patients who discontinued without providing follow-up weight assessments were excluded. The use of the mITT population is likely to be biased as the reasons why a patient discontinued trial treatment or failed to return for post-baseline weight assessments could be related to the efficacy or safety of the drug. Patients who were not seeing a satisfactory weight loss or experiencing side effects are more likely to stop taking the study drug. Results for the true intention-to-treat analysis should be the main data presented in the submission as this includes all patients in the treatment arms to which they were originally randomised. In our report we present the ITT results in addition to the mITT results. The only study where full ITT results were not available was COR-BMOD.

4.2.4 Direct evidence: Efficacy results

The main results of the modified intention-to-treat analysis presented in the CS are shown in Table 4.8. Tables 4.9 and 4.10 compare the mITT results for the primary outcomes with the two methods of ITT analysis **(**weight regain imputation method and using baseline-carried forward analysis).

Table 4.8: Main results of NB32 trials (mITT analysis)

	COR-I[25]		**COR-II[27]***		**COR-BMOD[26]**		**COR-DM[28]**	
	NB32	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
N	471	511	825	456	482	193	265	159
Baseline weight, mean kg (SD)	100.2 (16.3)	99.3 (14.3)	100.7 (16.7)	99.3 (16.0)	100.7 (15.4)	101.9 (15.0)	104.2 (18.9)	105.0 (17.1)
End of study weight, mean kg (SD)	94.2 (17.4)	98.0 (15.2)	94.2 (17.6)	97.2 (16.2)	91 (17.1)	96.4 (17.1)	101.0 (19.7)	103.0 (17.3)
Percent change from baseline at end of study, LS mean (SE)	-6.1 (0.3)	-1.3 (0.3)	-6.5 (0.2)	-1.9 (0.3)	-9.3 (0.4)	-5.1 (0.6)	-5.0 (0.7)	-1.8 (0.4)
NB32 – placebo, Difference of LS mean	-4.8 (-5.6 to -4.0)		-4.6 (-5.2 to -3.9)		-4.2 (-5.6 to -2.9)		-3.3 (-4.3 to -2.2)	
No of patients with ≥ 5% decrease in weight, n (%)	226 (48.0)	84 (16.4)	459 (55.6)	80 (17.5)	320 (66.4)	82 (42.5)	118 (44.5)	30 (18.9)
Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	4.9 (3.6 to 6.6)		6.6 (5.0 to 8.8)		2.9 (2.0 to 4.1)		3.4 (2.2 to 5.5)	
No of patients with ≥ 10% decrease in weight, n (%)	116 (24.6)	38 (7.4)	225 (27.3)	32 (7.0)	200 (41.5)	39 (20.2)	49 (18.5)	9 (5.7)
Patients with ≥ 10% decrease in weight, NB32 vs placebo, OR (95% CI)	4.2 (2.8 to 6.2)		5.4 (3.6 to 8.0)		2.9 (2.0 to 4.4)		3.8 (1.8 to 7.9)	
Source: Section 4.7 of the CS[1] Footnote: *Week 28 results BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus								

Table 4.9: Comparison of mITT and ITT results: percentage weight loss

Type of analysis	Outcome	COR-I[25]		COR-II[27]*		COR-BMOD[26]		COR-DM[28]	
		NB32	Pbo	NB32	Pbo	NB32	Pbo	NB32	Pbo
mITT	Percent change from baseline at end of study, LS mean (SE)	-6.1 (0.3)	-1.3 (0.3)	-6.5 (0.2)	-1.9 (0.3)	-9.3 (0.4)	-5.1 (0.6)	-5.0 (0.7)	-1.8 (0.4)
	NB32 – placebo, Difference of LS mean	-4.8 (-5.6 to -4.0)		-4.6 (-5.2 to -3.9)		-4.2 (-5.6 to -2.9)		-3.3 (-4.3 to -2.2)	
ITT using weight regain imputation method (ITT Imp)	Percent change from baseline at end of study, LS mean (SE)	-4.6 (0.3)	-1.2 (0.3)	-5.2 (0.2)	-1.9 (0.3)	NR	NR	-3.5 (0.3)	-1.7 (0.4)
	NB32 – placebo, Difference of LS mean	-3.4 (-4.1 to -2.7)		-3.4 (-3.9 to -2.8)		NR		-1.9 (-2.8 to -0.9)	
ITT using baseline-carried forward analysis (ITT BOCF)	Percent change from baseline at end of study, LS mean (SE)	-4.0 (0.3)	-0.9 (0.3)	-4.8 (0.2)	-1.5 (0.3)	-5.9 (0.4)	- 4.0 (0.6)	-3.1 (0.3)	-1.3 (0.4)
	NB32 – placebo, Difference of LS mean	-3.1 (-3.8 to -2.4)		-3.3 (-3.9 to -2.7)		-1.9 (-3.2 to -0.6)		-1.7 (-2.7 to -0.8)	

Source: Section 4.7 of the CS[1] and CS appendices[38]

Footnote *Week 28 results

BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus, Imp = weight regain imputation, BOCF = baseline observation carried forward

Table 4.10: Comparison of mITT and ITT results: patients with ≥ 5% decrease in weight

Type of analysis	Outcome	COR-I[25]		COR-II[27]*		COR-BMOD[26]		COR-DM[28]	
		NB32	Pbo	NB32	Pbo	NB32	Pbo	NB32	Pbo
mITT	No of patients with ≥ 5% decrease in weight, n (%)	226 (48.0)	84 (16.4)	459 (55.6)	80 (17.5)	320 (66.4)	82 (42.5)	118 (44.5)	30 (18.9)
	Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	4.9 (3.6 to 6.6)		6.6 (5.0 to 8.8)		2.9 (2.0 to 4.1)		3.4 (2.2 to 5.5)	
ITT Imp	No of patients with ≥ 5% decrease in weight, n (%)	203 (34.8)	78 (13.4)	446 (44.6)	79 (16.0)	NR	NR	104 (31.0)	27 (15.9)
	Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	3.6 (2.7 to 4.9)		4.7 (3.5 to 6.2)		NR			
ITT BOCF	No of patients with ≥ 5% decrease in weight, n (%)	180 (30.9)	67 (11.5)	421 (42.1)	69 (13.9)	NR	NR	94 (28.1)	24 (14.1)
	Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	3.6 (2.6 to 4.9)		4.9 (3.7 to 6.6)		NR		2.4 (1.4 to 3.9)	

Source: Section 4.7 of the CS[1] and CS appendices[38]

Footnote: *Week 28 results

BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus, Imp = weight regain imputation, BOCF = baseline observation carried forward

Subgroup analyses

Subgroup analyses were conducted for the co-primary efficacy variable (percentage of change in total body weight and proportion of patients with ≥5% decrease in total body weight at week 56 or week 28). Subgroups included study centre, sex, race, age, age group, BMI category, and tobacco use inter alia. The company did not provide results data on subgroups in the main submission document but stated that for all four trials results were 'generally consistent' or 'consistent' with the main findings. The percentages of overweight patients (BMI < 30 kg/m^2) in the trials are too small to present meaningful subgroup analyses. As the ERG had identified that the majority of participants in the trials were women we present the subgroup results for males and females separately.

Table 4.11: Results in male subgroups (mITT data)

	COR-I[25, 34]		COR-II[27, 36]		COR-BMOD[26, 35]		COR-DM[28, 37]	
	NB32	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
N	76	80	NR		56	17	121	75
Baseline weight, mean kg (SD)	115.36 (18.58)	112.16 (14.35)			118.11 (16.09)	122.12 (18.40)	116.83 (17.72)	111.89 (16.33)
End of study weight, mean kg (SD)	109.16 (18.10)	110.09 (15.33)			107.96 (18.99)	116.94 (20.35)	111.27 (18.58)	110.09 (15.95)
Percent change from baseline at end of study, LS mean (SE)	-5.20 (0.68)	-1.83 (5.92)			-8.75 (0.93)	-4.75 (1.70)	-4.79 (0.47)	-1.51 (0.60)
NB32 – placebo, Difference of LS mean (SE)	-3.34 (0.94)				-4.00 (1.94)		-3.28 (0.77)	
No of patients with ≥ 5% decrease in weight, n (%)	29 (38.16)	16 (20.00)			39 (69.64)	7 (41.18)	51 (42.15)	10 (13.33)
Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	2.36 (1.14, 4.86)				3.12 (0.99, 9.80)		4.69 (2.19, 10.05)	

Source: Trial CSRs
BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus

Table 4.12: Results in female subgroups (mITT data)

	COR-I[25, 34]		COR-II[27, 36]		COR-BMOD[26, 35]		COR-DM[28, 37]	
	NB32	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
N	395	431	NR		426	176	144	84
Baseline weight, mean kg (SD)	97.24 (14.02)	96.90 (13.00)			98.40 (13.81)	99.55 (13.19)	97.55 (15.49)	98.82 (15.47)
End of study weight, mean kg (SD)	91.29 (15.72)	95.80 (14.11)			88.79 (15.58)	94.40 (15.39)	92.31 (16.09)	96.71 (16.11)
Percent change from baseline at end of study, LS mean (SE)	-6.23 (0.34)	-1.15 (0.32)			-9.83 (0.41)	-5.66 (0.64)	-5.41 (0.46)	-2.15 (0.60)
NB32 – placebo, Difference of LS mean (SE)	-5.08 (0.46)				-4.17 (0.76)		-3.25 (0.76)	
No of patients with ≥ 5% decrease in weight, n (%)	197 (49.87)	68 (15.78)			281 (65.96)	75 (42.61)	67 (46.53)	20 (23.81)
Patients with ≥ 5% decrease in weight, NB32 vs placebo, OR (95% CI)	5.37 (3.87, 7.44)				2.59 (1.81, 3.71)		2.77 (1.52, 5.05)	

Source: Trial CSRs

BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus

It can be seen from Table s 4.11 and 4.12 that both men and women taking NB32 have statistically significantly better results than those taking placebo.

ERG comment:

- Based on the mITT data presented by the company NB32 results in greater weight loss and in a higher number reporting 5% or more weight loss. However the ERG draws the attention of the committee to the superior results of the BMOD trial. NB32 together with a more intensive behaviour modification programme resulted in 66.4% of patients losing 5% or more weight compared to 44 to 55% in the other three trials without such an intensive intervention. Moreover, in the BMOD trial the placebo and behaviour modification arm achieved results approaching the medication arms in the other trials.
- Both men and women appear to benefit from the intervention when comparing subgroups.
- Using the true ITT data, NB32 also results in a greater mean percentage of weight loss compared to placebo groups. The proportion of patients losing 5% or more weight is also higher in treatment groups. However the results, as expected, are lower than the mITT data. It is these data that should be used to determine effectiveness and to ascertain the clinical importance of the results.

Table 4.13 shows the reasons for treatment discontinuation across the trials.

Table 4.13: Reasons for treatment discontinuation

	COR-I[25]		**COR-II[27]**		**COR-BMOD[26]**		**COR-DM[28]**	
	NB32	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
Number randomised	583	581	1001	495	591	202	335	170
Number discontinued (%)	287 (49%)	291 (50%)	462 (46%)	226 (46%)	249 (42%)	84 (42%)	160 (48%)	70 (41%)
Reasons for discontinuation								
Adverse event	112	56	241	68	150	25	98	26
Lost to follow up	65	66	77	48	22	17	22	15
Withdrew consent	60	90	75	56	43	24	21	15
Enrolled but did not meet criteria	0	1	0	0	0	0	2	0
Study drug not dispensed	0	3	0	0	3	0	0	1
Participant judged weight loss insufficient	12	40	19	33	3	6	5	6
Drug non-compliance	17	15	31	13	13	5	8	3
Protocol non-compliance	9	7			4	0	3	4
Death	1	0	0	0	0	0	0	0
Other	11	13	19	8	11	7	1	0

Source: Figures 3 to 6 of the CS and accompanying text[1]
BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus

In COR-I, 870 (50%) of patients completed 56 weeks of treatment. 287 of 583 (49%) in the NB32 group discontinued whilst 291 of 581 (50%) in the placebo group discontinued. The company stated that '*Rates of discontinuation were similar across treatment groups*.'[1] However they noted (as can be seen in Table 4.13) more patients in the NB32 group discontinued due to adverse events than in the placebo group ($p < 0.0001$). More patients in the placebo group discontinued due to insufficient weight loss (6.9% vs. 2.1%, $p < 0.0001$) and withdrawal of consent (15.5% vs. 10.3%, $p = 0.0126$). The company stated that rates of discontinuation were higher during the first 16 weeks of the study in both treatment and placebo groups. A similar pattern was observed in COR-II where 46% of patients in each treatment group discontinued during 56 weeks of treatment. More NB-32 treated patients discontinued due to an adverse event (24.1% vs. 13.7%, $p < 0.001$) whilst more placebo group patients discontinued due to insufficient weight loss (6.7% vs. 1.9%, $p < 0.001$) and withdrawal of consent (11.3% vs. 7.5%, $p < 0.05$). In COR-BMOD 41.6% of the placebo group and 42.1% of the NB32 group discontinued treatment. Again a greater percentage of those in the NB32 group discontinued due to an adverse event (25.4% vs. 12.4%, $p < 0.001$). A greater number of placebo patients discontinued due to withdrawal of consent (11.9% vs. 7.3%, $p = 0.042$), loss to follow up (8.4% vs. 3.7%, $p = 0.008$) or self-perceived insufficient weight loss (3.0% vs. 0.5%, $p = 0.004$). In COR-DM 47.8% of the NB32 treatment group discontinued treatment compared with 41.2% in the placebo group. Again, a greater percentage of patients who received NB32 compared with placebo discontinued due to an adverse event (29.3% vs. 15.3%) A greater percentage in the placebo group withdrew as they were lost to follow up (8.8% vs.6.6%), withdrew consent (8.8% vs. 6.3%) or had self-perceived insufficient weight loss (3.5% vs. 1.5%).

ERG comment:

- Across the four main trials treatment discontinuation rates ranged from approximately 41 to 50%. This suggests that in practice up to half of patients may complete a year's treatment with NB32. Rates of discontinuation were found to be similar between NB32 and placebo in all trials.
- Reasons for discontinuation varied between treatment groups. In all four trials a higher number of patients in NB32 groups discontinued due to an adverse event. As more NB32 patients in each trial discontinued due to an adverse event this indicates that the mITT population is likely to be biased as these patients would be more likely to be missing a post-baseline weight assessment and be excluded from the mITT analysis. Adverse events will be discussed further in Section 4.2.6 of this report. Although the placebo groups in all trials had more participants discontinuing due to insufficient weight loss, percentages of patients citing this reason were relatively low (approximately 7% in COR-I and COR-II, 3% in COR-BMOD and 3.5% in COR-DM).

4.2.5 Direct evidence: Meta-analysis results

The company reported the results of meta-analyses for the NB32 trials in Chapter 4.9.3 of the CS (CS, Figures 16 and 17, pages 115-117) and results of sensitivity analyses are presented in Appendix 8 of the CS.

The four trials COR-I, COR-II, COR-BMOD and COR-DM were pooled in random effects meta-analyses for the two weight loss outcomes. However the moderate to high levels of statistical heterogeneity observed ($I^2 = 66.6\%$ for $\geq 5\%$ reduction in weight and 70.1% for percentage weight change) indicate variation in the results between the trials. Sensitivity analyses pooled the T2DM trial (COR-DM) and non-T2DM trials (COR-BMOD, COR-I and COR-II) separately which reduced this

heterogeneity to I^2 = 0% for the percentage weight change analysis but not for the ≥ 5% reduction in weight analysis where the heterogeneity remained high. Further analyses then excluded the COR-BMOD trial which removed the observed heterogeneity from the ≥ 5% reduction in weight analysis.

ERG comment: The trials were conducted in two different populations (with and without T2DM) and one trial used a more intensive behaviour modification as the background therapy (COR-BMOD) compared to the other three trials. As these differences in populations and interventions appear to be linked to the statistical heterogeneity between the results they should be pooled separately. The COR-I and COR-II trials are clinically similar and could in theory be pooled for the no T2DM analysis. However, as COR-II assessed results at 28 weeks and patients were re-randomised after this it should not be pooled with COR-I. COR-DM is the only trial available for the T2DM analysis and COR-BMOD should be considered separately due to the more intensive behaviour management therapy. Therefore, the ERG believes none of the NB32 trials should be pooled.

4.2.6 Direct evidence: Safety results

Safety results were based on the four main trials: COR-I, COR-II, COR-BMOD and COR-DM. According to the CS treatment-emergent adverse events were defined as "*events that first occurred or worsened during double-blind treatment (i.e. a new event or an exacerbation of a pre-existing condition) with an onset date after study drug administration and within 7 days of the last confirmed dose date. AEs with an onset date before the first dose of study drug were recorded under medical history*."[1] Events were categorised across the trials as mild, moderate or severe and relationship to the study drug was investigated.

The company further stated that "*Safety data are presented for the safety analysis set, defined as all randomised patients who were administered at least one tablet of study treatment and had at least one investigator contact/assessment at any time after the start of study treatment, regardless of whether they discontinued the study. Patients were grouped in the safety analysis set according to which study treatment was administered on the first day of treatment following randomisation.*"[1]

An overview of adverse events is presented in Table 4.14.

Table 4.14: Overview of adverse events

	COR-I[25]		**COR-II[27]**		**COR-BMOD[26]**		**COR-DM[28]**	
	NB32	**Pbo**	**NB32 /48**	**Pbo**	**NB32**	**Pbo**	**NB32**	**Pbo**
Safety analysis set								
All TEAEs, n (%)	476 (83.1)	390 (68.5)	852 (85.9)	370 (75.2)	547 (93.7)	176 (88.0)	301 (90.4)	144 (85.2)
Drug-related TEAEs, n (%)	336 (58.6)	167 (29.3)	630 (63.5)	189 (38.4)	447 (76.5)	108 (54.0)	238 (71.5)	57 (33.7)
Severe TEAEs, n (%)	51 (8.9)	34 (6.0)	110 (11.1)	33 (6.7)	98 (16.8)	15 (7.5)	61 (18.3)	19 (11.2)
TESAEs, n (%)	9 (1.6)a	8 (1.4)a	21 (2.1)	7 (1.4)	22 (3.8)	1 (0.5)	13 (3.9)	8 (4.7)
Discontinued due to AEs, n (%)	112 (19.5)	56 (9.8)	241 (24.3)	68 (13.8)	150 (25.7)	25 (12.5)	98 (29.4)	26 (15.4)
Deaths	1	0	0	0	0	0	0	0

	COR-I[25]	COR-II[27]	COR-BMOD[26]	COR-DM[28]
Source: Tables 42, 44, 46 and 48 of the CS[1] and Trial CSRs[34-37] Footnote: a) none found to be related to the drug AE = adverse event; BMOD = intensive behaviour modification; COR = Contrave obesity research; DM = diabetes mellitus; TEAE = treatment emergent adverse event; TESAE = treatment emergent serious adverse event				

Adverse events occurred in 83.1% to 93.7% of treatment groups and 68.5% to 88.0% of placebo groups. Approximately 58% to 76% of these were attributed to the drug in NB32 groups across the trials. Serious adverse events occurred at similar rates in treatment and placebo groups across the trials. However a larger number of patients discontinued due to adverse events across the trials (19.5% to 29.4% for treatment groups vs. 9.8% to 15.4% in placebo groups).

The main specific adverse events are listed in Table 4.15.

Table 4.15: Specific adverse events (≥ 5% in at least one treatment arm of an included trial)

	COR-I[25]		COR-II[27]		COR-BMOD[26]		COR-DM[28]	
	NB32 (n = 573)	**Pbo (n = 569)**	**NB32/ 48 (n = 992)**	**Pbo (n = 492)**	**NB32 (n = 584)**	**Pbo (n =200)**	**NB32 (n = 333)**	**Pbo (n = 169)**
Adverse Event, n (%)								
Gastrointestinal disorders	**292 (51.0)**	**136 (23.9)**	**532 (53.6)**	**131 (26.6)**	**380 (65.1)**	**78 (39.0)**	**215 (64.6)**	**53 (31.4)**
Nausea	171 (29.8)	30 (5.3)	290 (29.2)	34 (6.9)	199 (34.1)	21 (10.5)	141 (42.3)	12 (7.1)
Vomiting	56 (9.8)	14 (2.5)	84 (8.5)	10 (2.0)	64 (11.0)	13 (6.5)	61 (18.3)	6 (3.6)
Constipation	90 (15.7)	32 (5.6)	189 (19.1)	35 (7.1)	141 (24.1)	28 (14.0)	59 (17.7)	12 (17.1)
Dry mouth	43 (7.5)	11 (1.9)	90 (9.1)	13 (2.6)	47 (8.0)	6 (3.0)	21 (6.3)	5 (3.0)
Diarrhoea	26 (4.5)	28 (4.9)	55 (5.5)	18 (3.7)	43 (7.4)	15 (7.5)	52 (15.6)	16 (9.5)
Abdominal pain upper	NR	NR	NR	NR	32 (5.5)	3 (1.5)	17 (5.1)	3 (1.8)
Infections and infestations	**203 (35.4)**	**200 (35.1)**	**359 (36.2)**	**205 (41.7)**	**188 (32.2)**	**63 (31.5)**	**121 (36.3)**	**77 (45.6)**
Upper respiratory tract infection	57 (9.9)	64 (11.2)	86 (8.7)	55 (11.2)	38 (6.5)	18 (9.0)	26 (7.8)	16 (9.5)
Sinusitis	30 (5.2)	34 (6.0)	51 (5.1)	35 (7.1)	16 (2.7)	6 (3.0)	16 (4.8)	14 (8.3)
Nasopharyngitis	29 (5.1)	31 (5.4)	82 (8.3)	40 (8.1)	36 (6.2)	15 (7.5)	28 (8.4)	23 (13.6)
Musculoskeletal and connective tissue disorders	**72 (12.6)**	**90 (15.8)**	**159 (16.0)**	**96 (19.5)**	**104 (17.8)**	**46 (23.0)**	**58 (17.4)**	**40 (23.7)**
Nervous system disorders	**167 (29.1)**	**95 (16.9)**	**326 (32.9)**	**81 (16.5)**	**263 (45.0)**	**60 (30.0)**	**129 (38.7)**	**32 (18.9)**

	COR-I[25]		COR-II[27]		COR-BMOD[26]		COR-DM[28]	
Headache	79 (13.8)	53 (9.3)	174 (17.5)	43 (8.7)	139 (23.8)	35 (17.5)	46 (13.8)	15 (8.9)
Dizziness	54 (9.4)	15 (2.6)	68 (6.9)	18 (3.7)	85 (14.6)	9 (4.5)	390 (11.7)	9 (5.3)
Tremor	12 (2.1)	1 (0.2)	35 (3.5)	3 (0.6)	34 (5.8)	2 (1.0)	22 (6.5)	4 (2.4)
Psychiatric disorders	**85 (14.8)**	**62 (10.9)**	**205 (20.7)**	**75 (15.2)**	**145 (24.8)**	**45 (22.5)**	**75 (22.5)**	**20 (11.8)**
Insomnia	43 (7.5)	29 (5.1)	97 (9.8)	33 (6.7)	51 (8.7)	12 (6.0)	37 (11.1)	9 (5.3)
Anxiety	9 (1.6)	12 (2.1)	48 (4.8)	21 (4.3)	30 (5.1)	7 (3.5)	18 (5.4)	2 (1.2)
Vascular disorders	**51 (8.9)**	**22 (3.9)**	**(7)**	**(3.3)**	**46 (7.9)**	**7 (3.5)**	**40 (12.0)**	**12 (7.1)**
Hot flush	30 (5.2)	7 (1.2)	42 (4.2)	6 (1.2)	28 (4.8)	1 (0.5)	7 (2.1)	4 (2.4)
Tinnitus	15 (2.6)	6 (1.1)	29 (2.9)	1 (0.2)	31 (5.3)	1 (0.5)	8 (2.4)	1 (0.6)
Hypertension	17 (3.5)	14 (2.5)	(1.9)	(1.6)	14 (2.4)	4 (2.0)	33 (9.9)	7 (4.1)
Source: Tables 43, 45, 47 and 49 of the CS[1] and Trial CSRs[34-37] Footnote: Adverse event categories in bold								

The main category of adverse event occurring more frequently in treatment groups across the trials was gastrointestinal disorders. Nausea, in particular, occurred frequently and more often in treatment groups. Across the trials rates of nausea ranged from 29.2% to 42.3% in treatment groups. Rates ranged from 5.3% to 10.5% in placebo groups. Vomiting, constipation and dry mouth also occurred more frequently in treatment groups although at a lower rate than that of nausea. Nervous system disorders such as headache, dizziness and tremor occurred more frequently in treatment groups.

The incidence of events of particular concern (serious cardiovascular disorders and suicidality measured on IDS) was extremely small and any differences between groups could not be ascertained in view of the small numbers in both groups.

ERG comment:

- The ERG draws to the attention of the committee the greater proportion of gastrointestinal events, particularly nausea, in NB32 groups across the trials. Although the majority of events were not serious, more participants withdrew as a result of adverse events in treatment groups. This finding is relevant to implementation of the intervention in clinical practice.
- The ERG notes that the NB-CVOT trial (described in Section 4.2.7 of the report below) was primarily designed to investigate the cardiovascular safety of NB32 in weight management. However the study was terminated earlier than originally planned (after the 50% interim analysis), after interim data were made public in a US patent (and related Orexigen security filings) and by the EMA in the Mysimba EPAR.
- A further trial on occurrence of MACE in overweight and obese patients with cardiovascular disease receiving NB32 was requested by CHMP. Based in the US, this randomised trial will

aim to enrol 8,000 patients and is estimated to last for up to six years until the targeted number of adjudicated MACE events (378) has been reached. The primary MACE composite comprises the first occurrence of CV death, nonfatal myocardial infarction (MI), and nonfatal stroke.[31]

4.2.7 Overview of the supporting RCTs

As previously stated, two trials (NB-CVOT and IGNITE) were used to provide data on long-term safety and efficacy only. NB-CVOT was presented in detail in the submission. IGNITE was presented only briefly and the company stated that '*At the time of database searches, this study was not yet published and was therefore not identified or included in the SLR*'.[1] This section will give an overview of each trial in turn.

NB-CVOT

The CS stated that "*The NB-CVOT study was a Phase IIIb, multicentre, randomised, double-blind, placebo-controlled trial to assess the occurrence of MACE in overweight or obese patients*."[1] Study details are given in the table below.

Table 4.16: Overview of NB-CVOT

Participants	**Intervention** (n =4,454)	**Control** (n =4,456)	**Trial design and duration**	**Primary outcome**
Patients aged 45 (men) or 50 (women) years or older, with a BMI 27–50kg/m^2 and a waist circumference of 88cm (women) or 102cm (men) or more. Patients had characteristics associated with an increased risk of adverse CV outcomes.	Naltrexone 32 mg per day + bupropion 360 mg per day (NB32) + customary diet and behaviour modification Patients were also encouraged to participate in an internet-based weight management program as well as having access to a personal weight loss coach and a low-fat, low-calorie meal plan.	Placebo + customary diet and behaviour modification Patients were also encouraged to participate in an internet-based weight management program as well as having access to a personal weight loss coach and a low-fat, low-calorie meal plan.	Lead-in period 4 week dose escalation period Maintenance period At 16 weeks, if patients did not lose ≥2% of their initial body weight or experienced a sustained (at ≥2 visits) increase in blood pressure (systolic or diastolic) of 10mmHg or greater they were discontinued	Time from treatment randomisation to the first confirmed occurrence of a MACE, defined as CV death, nonfatal stroke, or nonfatal myocardial infarction. Only SAEs and AEs leading to study drug discontinuation were collected. No subgroup analyses
Source: Table 11 and section 4.3 of the CS[1] AE = adverse event; CV = cardiovascular; MACE = major adverse cardiovascular events; SAE = serious adverse event				

The main differences between the NB-CVOT trial and the COR trials is that participants were all at increased risk of adverse cardiovascular outcomes. Furthermore, the trial incorporated a lead-in period. During the lead in period 1,490 patients discontinued. Of these 543 discontinuations were due to adverse events.

The trial incorporated a stopping rule at 16 weeks (unlike the COR trials). The company stated that "*A large decrease in the number of patients receiving the study drug occurred after the 16-week assessment, with 44% of placebo patients and 17.8% of NB32 patients discontinued by investigators.*

Most discontinuations were due to a failure to lose 2% of body weight, but 230 placebo patients and 154 NB32 patients discontinued treatment because of a greater than 10mmHg increase in blood pressure. A high percentage of patients who discontinued treatment remained in follow-up for MACE and contributed to the ITT analysis set."[38]

NB-CVOT was terminated early (after the 50% interim analysis), after 25% interim data were made public in a US patent (and related Orexigen security filings) and by the EMA in the Mysimba EPAR.

Patient characteristics are given in Table 4.17.

Table 4.17: NB-CVOT patient characteristics

Patient characteristics		
	NB32	**Pbo**
Age, mean years (SD)	61.1 (7.27)	60.9 (7.38)
Age range (min, max)	45, 86	45, 85
Sex, female, n (%)	2437 (54.7)	2419 (54.4)
Ethnicity, n (%)		
White	3738 (83.9)	3698 (83.1)
Black	656 (14.7)	648 (14.6)
Asian	19 (0.4)	27 (0.6)
Other	41 (0.8)	75 (1.6)
BMI, mean kg/m^2 (SD)	37.2 (5.26)	37.4 (5.44)
Obesity class, n (%)		
BMI < 30 kg/m^2	299 (6.7)	311 (7.0)
BMI ≥30 and <35 kg/m^2	1393 (31.3)	1408 (31.6)
BMI ≥35 and <40 kg/m^2	1476 (33.2)	1383 (31.1)
BMI ≥40 kg/m^2	1284 (28.8)	1348 (30.3)
Weight, mean kg (SD)	105.6 (19.09)	106.3 (19.18)
Smoker, n (%)	405 (9.1)	416 (9.3)
Hypertension, n (%)	4162 (93.4)	4117 (92.5)
Dyslipidaemia, n (%)	4100 (92.0)	4070 (91.5)
Type 2 diabetes	3784 (84.9)	3803 (85.5)
Alcohol use, n (%)	NR	NR
History of depression	1031 (23.1)	995 (22.4)
History of anxiety	NR	NR
Source: CS[1] and NB-CVOT CSR[31]		

The mean age of randomised patients was 61 years compared to approximately 44 years in the COR trials (54 in COR-DM). Just over half are female (similar to COR-DM) whereas the majority are female in COR-I, COR-II and COR-BMOD. Ethnicity was similar to the COR trials with the majority of participants being white.

In terms of comorbidities, T2DM was present in 85.2% of patients (0 in the COR trials except for COR-DM (100%)). 32.1% had cardiovascular disorders but the company described cardiovascular risk factors as "*well-controlled*".[1] Nearly all patients in NB-CVOT had hypertension or dyslipidaemia, Concomitant

medications included statins in 79.2% of patients, anti-hypertensive medications in 92.0%, and glucose lowering agents in 75.1%, and anti-depressant medication in 23.1% of patients. The company stated that there was a higher instance of depression in NB-CVOT than in the COR trials.[1]

The CS stated that "*The 50% interim analysis was completed on 3 March 2015, based on 192 adjudicated major adverse cardiovascular events (MACE) (from a database lock on 3 February 2015). Additional outcomes accumulated after the February 2015 database lock are included in a sensitivity analysis, which reports results after 64% of planned events."* Results are presented in the Table below for the 50% analysis.

Table 4.18: NB-CVOT effectiveness outcomes

	NB32 (n=4,455)	**Placebo (n=4,450)**	**HR (99.7% CI)**
Primary outcome, n (%)			
MACE	90 (2.0)	102 (2.3)	0.9 (0.6–1.3)
CV death	17 (0.4)	34 (0.8)	0.5 (0.2–1.2)
Nonfatal stroke	21 (0.5)	19 (0.4)	1.1 (0.4–2.8)
Nonfatal myocardial infarction	54 (1.2)	54 (1.2)	1.0 (0.6–1.8)
Secondary outcomes, n (%)			
MACE + hospitalisation for unstable angina	133 (3.0)	142 (3.2)	0.9 (0.7–1.3)
Fatal or nonfatal stroke	22 (0.5)	21 (0.5)	1.0 (0.4–2.6)
Fatal or nonfatal myocardial infarction	55 (1.2)	57 (1.3)	1.0 (0.6–1.7)
Other outcomes, n (%)			
All-cause mortality	43 (1.0)	51 (1.1)	0.8 (0.5–1.5)
Hospitalisation for unstable angina	47 (1.1)	43 (1.0)	1.1 (0.6–2.0)
Coronary revascularisation	132 (3.0)	145 (3.3)	0.9 (0.6–1.3)
All-cause mortality + nonfatal myocardial infarction + nonfatal stroke	114 (2.6)	119 (2.7)	1.0 (0.7–1.4)
MACE + hospitalisation for unstable angina + coronary revascularisation	188 (4.2)	205 (4.6)	0.9 (0.7–1.2)

Source: CS[1]
CV = cardiovascular; MACE = major adverse cardiovascular events

For the 50% interim analysis, time to first MACE, occurred in 192 patients; 102 (2.3%) in the placebo group and 90 (2.0%) in the NB32 group (HR: 0.88; 99.7% CI: 0.57, 1.34). The components of the primary composite outcome included CV death (0.8% of placebo patients and 0.4% of NB32 patients [HR: 0.50; 99.7% CI: 0.21, 1.19]), nonfatal stroke (0.4% of placebo patients and 0.5% of NB32 patients [HR: 1.10; 99.7% CI: 0.44, 2.78]) and nonfatal myocardial infarction (1.2% in both placebo and NB32 patients [HR: 1.00; 99.7% CI: 0.57, 1.75]). The company stated that "*In general, final end-of-study analyses support these data.*" (CS, page 109).

The company further stated that "*At trial completion, body weight decreased by a mean of 3.9kg (95% CI: -4.1, -3.7kg) in the NB32 group compared to a mean decrease of 1.2kg (95% CI: -1.3, -1.0kg) in the placebo group, corresponding to reductions of 3.6% and 1.1%, respectively (p<0.001).*" (CS, page 109).

Adverse events identified in NB-CVOT are detailed in Table 4.19.

Table 4.19: NB-CVOT adverse events

	NB32 (n=4455)	Placebo (n=4450)
Drug-related TEAEs, n (%)	982 (22.0)	174 (3.9)
Severe TEAEs, n (%)	217 (4.9)	108 (2.4)
TESAEs, n (%)	463 (10.4)	386 (8.7)
DC due to AEs, n (%)	1292 (29.0)	400 (9.0)
Deaths, n (%)	65	72
Source: CS[1] AE = adverse event; DC = discontinuation; TEAE = treatment emergent adverse event; TESAE = treatment emergent serious adverse event.		

It can be seen from Table 4.19 that more patients in the NB32 group experienced events that were considered by the investigator to be study drug–related (22.0% vs. 3.9% with placebo). In both groups, most TEAEs leading to discontinuation were considered mild or moderate in intensity. TESAEs, (defined as any AE occurring at any dose of study drug that resulted in death, life-threatening adverse drug experience, inpatient hospitalisation or prolongation of existing hospitalisation, persistent of significant disability or incapacity, important medical events or congenital anomaly or birth defect) were reported for 849 patients (9.5%) overall, 10.4% in the NB32 group and 8.7% in the placebo group. A total of 137 deaths occurred during the study, 65 patients in the NB32 group and 72 in the placebo group, although no deaths in this study were related to the study drug.

Discontinuations due to adverse events most commonly included gastrointestinal AEs, which occurred in 14.2% of NB32 patients and 1.9% of placebo-treated patients and central nervous system symptoms, which occurred in 5.1% of NB32 patients and 1.2% of placebo patients. Psychiatric symptoms resulted in study drug discontinuation in 3.1% of NB32 patients and 0.9% of placebo patients ($p<0.001$).

IGNITE

IGNITE was described in the CS as a "*Phase IIIb, randomised, open label, controlled study in which patients received NB32 plus comprehensive lifestyle intervention (CLI) or usual care (standard diet and exercise advice) for 26 weeks.*"[1] Patients in the NB32 + CLI group not achieving 5% weight loss at week 16 were discontinued. After week 26 patients in the usual care arm began NB32 + CLI and were assessed up to week 78. The primary endpoint was percentage change in weight from baseline to week 26 in the per protocol (PP) population. Other endpoints included percentage of patients achieving ≥5%, ≥10% and ≥15% weight loss, percent change in weight at week 78 and AEs necessitating study discontinuation. No subgroup analyses were conducted.

A total of 242 patients were randomised; 153 to NB32 + CLI and 89 patients to usual care for a total of 26 weeks. It was noted in the CS that although the trial was of 78 weeks' duration, patient numbers beyond 52 weeks were low; 61 NB32 patients were followed from week 52 onwards.

The CS stated that *"Patients assigned to treatment with NB32 plus standard management lost significantly more weight than patients treated with usual care at Week 26 (-9.4% vs -0.94% respectively; p<0.0001). For patients who remained on treatment, the initial weight loss observed at 26 weeks was sustained throughout Week 78, further supporting the maintained effectiveness of NB32 treatment."*

The CS further stated that *"In the IGNITE study, the safety profile shown was consistent with that seen in the previous, pivotal trials; most patients tolerated NB32 well, and those who developed AEs did so early in the treatment protocol. The most common AE leading to NB discontinuation was nausea (7.0% of all subjects), which is consistent with the rate in the Phase III trials (6.3%)."*

No further information was provided on IGNITE.

ERG comment:

- The NB-CVOT study has the potential to provide information on performance of NB32 in an older population with cardiovascular disease when compared to the COR trials. Most of the patients in NB-CVOT are diabetic, and many are depressed. However a number of problems were identified with NB-CVOT. These include the use of a lead-in period where large numbers of patients discontinued primarily due to adverse events. This implies that those continuing to the treatment period who were re-randomised were better able to tolerate the drug. The adverse event profile will be an overestimate of the tolerability of the drug. In addition only SAEs and AEs leading to study drug discontinuation were collected. Even so, an elevated number of gastrointestinal events were noted in the NB32 group. A further limitation is that the trial was terminated early (after the 50% interim analysis), after 25% interim data were made public in a US patent (and related Orexigen security filings) and by the EMA in the Mysimba EPAR. The trial was not able to provide a definitive answer to the cardiovascular risk of NB32 and a further trial has been instigated. The reliability of the final data on weight loss is also questionable.
- The IGNITE trial was described only briefly in the submission. The main limitation of this trial was that intervention and control groups were not directly comparable. This trial is not able to assess the unique effect of NB32 but only the combined effect of NB32 and a comprehensive lifestyle intervention. Furthermore the trial was only randomised for 26 weeks rather than 56 for most of the COR trials.

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

No trials were identified that compared NB32 directly with orlistat. Therefore, the company performed indirect comparisons to compare NB32 with orlistat using placebo as the common comparator.

As described in Section 4.1.2 of this report, the search was not aimed at finding studies of behavioural or lifestyle interventions. In addition to the inclusion criteria described in Section 4.1.2 of this report, the following studies/treatment arms were excluded:

- Treatment arm is not of interest
- Treatment group is not administered at recommended dosage
- Trial reduces to single treatment arm once other arms are pooled or excluded
- Trial reports no relevant outcome data
- Trial excludes patients during a lead-in period due to weight loss criteria or treatment compliance
- Trial has a wait list control group as a comparator arm in which patients receive no pharmaceutical treatment or standard management

For the analyses performed in the CS, NB and orlistat were evaluated at their recommended doses detailed in the summary of product characteristics for each treatment[39, 40]:

- NB – naltrexone 32mg/day prolonged release plus bupropion 360mg/day prolonged release (NB32)

- Orlistat – 120mg three times a day (TID)

ITC were performed to compare NB32 and orlistat for the following outcomes:

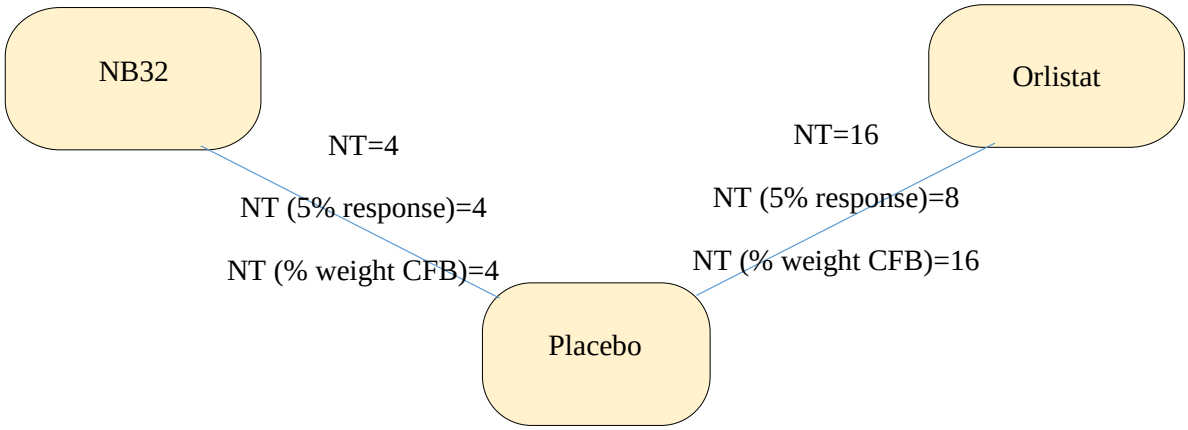
- Mean percentage weight change from baseline at one year (the one year time point ranged from 52 to 57 weeks [continuous outcome])
- At least 5% reduction in weight at one year from baseline (the one year time point ranged from 52 to 57 weeks [dichotomous outcome])

According to the CS, 36 RCTs were included (CS, page 46) in the systematic literature review, five studies investigated treatment with NB32 (COR-I, COR-II, COR-BMOD, COR-DM and NB-CVOT), while the remaining 31 studies investigated treatment with orlistat.

As explained in Section 4.2 of this report, four NB32 studies were used in the indirect comparisons (COR-I, COR-II, COR-BMOD and COR-DM); NB-CVOT was excluded from the analyses due to the trial design, objective, and patient population, being different from the other NB32 studies and patients were excluded during the lead in period. In addition, 16 out of 31 orlistat studies were used in the indirect comparisons. Reasons for exclusion of the 15 orlistat trials not used in the analyses are explained in Appendix 9 of the CS.

Twenty trials were included in the indirect treatment comparison (ITC). The list of trials, along with treatments and available outcome data that were included in the analyses are presented in Table 4.20. The maximum evidence base for each outcome, following the additional exclusion is given in the network of evidence presented in Figure 4.1.

Figure 4.1: Network of evidence



Source: CS, Figure 18, page 124[1]
Notes**:** 5% response defined as ≥5% reduction in weight from baseline at 1 year.
CFB = change from baseline; NB32 = naltrexone 32mg plus bupropion; NT = number of trials.

Table 4.20: Evidence base: trials, treatments and outcomes included in the ITC analyses

Trial (NT=20)	Trial duration	Treatment		Analysis population					Outcome	
		Arm 1	Arm 2	Trials without a lead-in period	Trial excludes T2DM patients[a]	T2DM is part of trial inclusion criteria	Trials without a high proportion of comorbidities[b]	StMan without intensive BMOD[c]	≥5% reduction in weight	Mean % weight CFB
Apovian 2013[27] (COR-II; NCT00567255)[d]	56 weeks	PBO	NB32	✓	✓	-	✓	✓	✓	✓
Greenway 2010[25] (COR-I; NCT00532779)	56 weeks	PBO	NB32	✓	✓	-	✓	✓	✓	✓
Hollander 2013[28] (COR-DM; NCT00474630)	56 weeks	PBO	NB32	✓	-	✓	-	✓	✓	✓
Wadden 2011[26] (COR-BMOD; NCT00456521)	56 weeks	PBO	NB32	✓	✓	-	✓	-	✓	✓
Astrup 2012[41] (NN8022-1807 study group; NCT00422058 [extension study: NCT00480909)	54 weeks (2-week lead-in period and 52-week treatment phase [weeks 20–52 were part of an extension study])	PBO	ORL 120mg TID	-	✓	-	✓	✓	✓	✓
Bakris 2002[42]	52 weeks	PBO	ORL 120mg TID	✓	-	-	-	✓	✓	✓
Berne 2005[43] (OST2D study group)	54 weeks (2-week lead-in period and 52-week treatment period)	PBO	ORL 120mg TID	-	-	✓	-	✓	✓	✓
Broom 2002[44] (UKM study group)	54 weeks (2-week lead-in period and 52-week treatment period)	PBO	ORL 120mg TID	-	-	-	✓	✓	✓	✓
Derosa 2003[45]	56 weeks (4-week lead-in period and 52-week treatment period)	PBO[e]	ORL 120mg TID[f]	-	-	-	-	✓	-	✓
Derosa 2010[46]	52 weeks	PBO	ORL 120mg TID	✓	-	✓	-	✓	-	✓

Trial (NT=20)	Trial duration	Treatment		Analysis population					Outcome	
		Arm 1	Arm 2	Trials without a lead-in period	Trial excludes T2DM patients[a]	T2DM is part of trial inclusion criteria	Trials without a high proportion of comorbidities[b]	StMan without intensive BMOD[c]	≥5% reduction in weight	Mean % weight CFB
Gotfredsen 2001[47] (EM Study-I)	52 weeks (4-week lead-in period and 48-week treatment period)	PBO	ORL 120mg TID	-	-	-	✓	✓	-	✓
Karhunen 2000[48] (EM Study-II)	108 weeks (4-week lead-in period and two 52-wk treatment periods)	PBO	ORL 120mg TID	-	✓	-	✓	✓	-	✓
Kelley 2002[49]	54 weeks (2-week screening and 52-week treatment phase)	PBO	ORL 120mg TID	✓	-	✓	-	✓	✓	✓
Lindgarde 2000[50]	54 weeks (2-week lead-in period and 52-week treatment period)	PBO	ORL 120mg TID	-	-	-	-	✓	✓	✓
Lucas 2003[51]	56 weeks (4-week lead-in period and 52-week treatment period)	PBO	ORL 120mg TID	-	-	-	-	✓	-	✓
Mathus-Vliegen 2006[52]	56 weeks (4-week lead-in period and 52-week treatment period)	PBO	ORL 120mg TID	-	✓	-	✓	✓	-	✓
Miles 2002[53]	54 weeks (2-wkk screening period and 52-week treatment phase)	PBO	ORL 120mg TID	✓	-	✓	-	✓	✓	✓
Reaven 2001[54]	56 weeks (4-week lead-in period and 52-week treatment period)	PBO[g]	ORL 120mg TID[g]	-	✓	-	✓	✓	-	✓
Swinburn 2005[55]	56 weeks (4-week lead-in period plus 52-week treatment period)	PBO	ORL 120mg TID	-	-	-	✓	✓	-	✓
Torgerson 2004[56] (XENDOS)	208 weeks	PBO	ORL 120mg TID	✓	✓	-	✓	-	✓	✓
Total NB32 trials				4	3	1	3	3	4	4

Trial (NT=20)	Trial duration	Treatment		Analysis population					Outcome	
		Arm 1	Arm 2	Trials without a lead-in period	Trial excludes T2DM patients[a]	T2DM is part of trial inclusion criteria	Trials without a high proportion of comorbidities[b]	StMan without intensive BMOD[c]	≥5% reduction in weight	Mean % weight CFB
Total ORL trials				5	5	4	8	15	8	16
Total trials				9	8	5	11	18	12	20

Source, CS, Table 29, pages 120-123[1]

Notes: [a]) As per the trial exclusion criteria; [b]) High proportion of comorbidities were defined as in Section 4.10.3 of the CS, [c]) Intensive BMOD defined as in Section4.10.4 of the CS; [d]) Non-responders in the Apovian 2013 trial were re-randomised to either NB32 or NB48. Non-responders who received NB48 after 32 weeks were not included in the analysis, and patients who received NB32 were double weighted in the analysis; [e]) PBO and PBO+FV have been pooled together; [f]) ORL 120mg TID and ORL120mg TID+FV have been pooled together; [g]) Trial presents arm data split by whether patients had syndrome X, and patients with/without syndrome X were pooled for each treatment.

BMOD = behaviour modification; CFB = change from baseline; COR = Contrave® obesity research; DM = diabetes mellitus; EM = European multicentre; FV = fluvastatin; NB = naltrexone plus bupropion; NB16 = naltrexone 16mg plus bupropion; NB32 = naltrexone 32mg plus bupropion; NB48 = naltrexone 48mg plus bupropion; NT = number of trials; ORL = orlistat 120mg TID; PBO = placebo; SM = Swedish Multimorbidity; StMan = Standard Management; T2DM = Type 2 diabetes mellitus; TID = three times a day; UKM = UK Mulitmorbidity; XENDOS = Xenical in the prevention of diabetes in obese subjects.

In total 24 analyses were performed (see CS Table 33, page 131), six base-case analyses for patients with T2DM, patients without T2DM and all patients (for each population a Bayesian and a Frequentist pairwise MA was performed). In addition four sensitivity analyses were performed for each population using Bayesian and Frequentist methods. Therefore, a total of 30 analyses would have been possible, but six analyses could not be performed due to insufficient data or because the evidence base was the same as in a previous analysis.

The four sensitivity analyses were: SA1-excluding trials where ≥75% of patients had at least one comorbidity; SA2-trials incorporating lead-in periods were excluded; SA3-studies with 'intensive' behaviour modification (BMOD and XENDOS) were excluded; SA4-trials with lead-in periods or 'intensive' behaviour modification were excluded.

ERG comment: The company did not actively search for trials comparing different types of behavioural interventions. Therefore, the CS includes only comparisons of NB32 plus standard management versus placebo plus standard management and NB32 plus intensive behaviour modification versus placebo plus intensive behaviour modification. There is no comparison of NB32 plus standard management versus intensive behaviour modification. In addition, the company did not include an analysis of NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification. We have added this analysis in Section 4.5.2 of this report (using data from COR-BMOD for NB32 and XENDOS for orlistat).

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

Baseline characteristics for the four NB32 trials are reported in Table 4.5 in Section 4.2. Baseline characteristics for the 16 orlistat trials are reported in the table below.

One of the five NB32 trials and four out of the 16 orlistat trials included people with type 2 diabetes (T2DM) only. Participants in these trials were generally older and had more often hypertension and dyslipidaemia.

Table 4.21: Participant characteristics in the orlistat trials

	Astrup 2012[41]		Bakris 2002[42]		Berne 2005[43]		Broom 2002[44]	
	ORL	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**
n	95	98	267	265	111	109	259	263
Age, mean years ± SD	45.9 ± 9.1	45.9 ± 10.3	53.2 ± 0.5	52.5 ± 0.5	58.9 ± 9.1	59.3 ± 8.5	46.7 ± 11.4	45.3 ± 11.5
Age range, min-max	NR	NR	NR	NR	NR	NR	22-73	20-74
Female, n (%)	73 (77)	73 (75)	169 (63)	156 (59)	50 (45)	50 (46)	202 (78)	207 (79)
Weight, mean Kg ± SD	96.0 ± 1.7	97.3 ± 12.3	101.2 ± 1.0	101.5 ± 1.0	95.3 ± 12.6	95.7 ± 12.5	100.9 ± 20.5	101.8 ± 19.8
BMI, mean Kg/m^2 ± SD	34.1 ± 2.6	34.9 ± 2.8	35.8 ± 3.9	35.4 ± 4.0	32.6 ± 3.1	32.9 ± 3.0	37.1 ± 6.4	37.0 ± 6.2
Waist circumference, cm ± SD	108 ± 9.7	108 ± 10.0	108.6 ± 12.2	110.8 ± 12.5	108.0 ± 9.0	109.0 ± 9.3	107.8 ± 15.6	108.6 ± 16.4
White, n	NR	NR	CAU 226 HIS 10	228 CAU 3 HIS	NR	NR	NR	NR
Black, n	NR	NR	AF-AM 20	AF-AM 31	NR	NR	NR	NR
Asian, n	NR	NR	NR	NR	NR	NR	NR	NR
Other, n	NR	NR	1	2	NR	NR	NR	NR
T2DM, n (%)	3 (3)	4 (4)	23 (8)	22 (8)	111 (100)	109 (100)	IGT 11 (4)	IGT 15 (5)
Hypertension, n (%)	NR	NR	267 (100)	265 (100)	AHD 50 (45)	AHD 49 (45)	54 (20)	59 (22)
Dyslipidaemia, n (%)	NR	NR	LDL/HDL, mean ± SD 3.18 (1.1)	LDL/HDL, mean ± SD 3.3 (1.2)	LLD 13 (12)	LLD 17 (16)	114 (43)	120 (45)

Key: AF-AM = African-American, AHD = Participants taking antihypertensive drugs, CAU = Caucasian, HIS = Hispanics, IGT = Impaired Glucose Tolerance, LLD = Lipid Lowering Drugs, NR = Not Reported, ORL = Orlistat, PBO = Placebo, T2DM = Type 2 Diabetes Mellitus

Notes: No data were reported for BMI ranges, smokers, alcohol use, history of depression or anxiety

Table 4.21: Participant characteristics in the orlistat trials (continued)

	Derosa 2003*[45]		Derosa 2010[46]		Gotfredsen 2001[47]		Karhunen 2000[48]	
	ORL	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**
n	52	47	126	128	16	14	36	36
Age, mean years ± SD	52.3 ± NR	51.5 ± NR	53.0 ± 6	52.0 ± 5	42.2 ± 11.7	40.2 ± 9.6	42.9 ± NR	44.4 ± NR
Female, n (%)	26 (50)	25 (53)	62 (49)	64 (50)	13 (81)	13 (93)	NR	NR
Weight, mean Kg ± SD	95.1	95.3	94.5 ± 9.6	91.7 ± 8.7	107.6 ± 17.7	99.4 ± 9.2	98.1 ± 12.2	97.3 ± 14.8
BMI, mean Kg/m^2 ± SD	32.2	31.9	33.1 ± 2.9	32.5 ± 2.3	36.9 ± 3.9	36.6 ± 3.9	35.7 ± 3.4	36.1 ± 4.4
Waist circumference, cm ± SD	102.1	102.2	102. 0 ± 6.0	101.0 ± 5.5	NR	NR	106.8 ± 10.5	106.2 ± 11.2
T2DM, n (%)	NR	NR	126 (100)	128 (100)	NR	NR	0	0
Hypertension, n (%)	NR	NR	93 (86.1)	89 (80.2)	NR	NR	NR	NR
Dyslipidaemia, n (%)	100	100	23 (21.3)	21 (18.9)	NR	NR	NR	NR
Smoker, n (%)	NR	NR	41 (32.5)	46 (35.9)	NR	NR	NR	NR

Key: NR = Not Reported, ORL = Orlistat, PBO = Placebo, T2DM = Type 2 Diabetes Mellitus, * ORL 120mg TID and ORL120mg TID+FV have been pooled and PBO and FV arms have been pooled.

Notes: No data were reported for age range, ethnicity, BMI ranges, alcohol use, history of depression or anxiety

Table 4.21: Participant characteristics in the orlistat trials (continued)

	Kelley 2002[49]		Lindgarde 2000[50]		Lucas 2003[51]		Mathus-Vliegen 2006[52]	
	ORL	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**
n	266	269	190	186	256	188	14	14
Age, mean years ± SD	57.8 ± 8.1	58 ± 8.2	53.7 ± 9.4	53.2 ± 9.9	48.0 ± NR	48.0 ± NR	42.0 ± 11.7	45.5 ± 9.3
Age range, min-max	NR	NR	27-74	28-75	NR	NR	NR	NR
Female, n (%)	150 (56)	151 (56)	124 (65)	115 (62)	199 (78)	158 (84)	NR	NR
Weight, mean Kg ± SD	102.0 ± 1	101.8 ± 1	96.1 ± 13.7	95.9 ± 3.5	98.6	99.2	102.6 ± 12.3	109.3 ± 16.4
BMI, mean Kg/m^2 ± SD	35.8 ± 0.2	35.6 ± 0.3	33.2 ± 3.0	33.2 ± 3.1	35.7	36.2	35.7 ± 3.8	37.6 ± 3.9
Waist circumference, cm ± SD	113.1 ± 0.7	113.9 ± 0.8	106.0 ± 10.8	106.0 ± 11.0	NR	NR	NR	NR
T2DM, n (%)	100	100	54 (28.0)	44 (24.0)	NR	NR	0	0
Hypertension, n (%)	NR	NR	143 (82.0)	137 (74.0)	NR	NR	NR	NR
Dyslipidaemia, n (%)	NR	NR	HC 75 (39.0)	HC 75 (40.0)	100	100	NR	NR

Key: HC= hypercholesterolemia, NR = Not Reported, ORL = Orlistat, PBO = Placebo, T2DM = Type 2 Diabetes Mellitus
Notes: No data were reported for ethnicity, BMI ranges, smokers, alcohol use, history of depression or anxiety

Table 4.21: Participant characteristics in the orlistat trials (continued)

	Miles 2002[53]		Reaven 2001[54]		Swinburn 2005[55]		Torgerson 2004[56]	
	ORL	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**	**ORL**	**PBO**
n	250	254	156	91	170	169	1640	1637
Age, mean years ± SD	52.5 ± 0.4	53.7 ± 0.4	45.1 ± NR	44.1 ± NR	52.0 ± 7.5	52.5 ± 7.4	43.0 ± 8.0	43.7 ± 8.0
Female, n (%)	120 (48)	122 (48)	109 (70)	65 (71)	104 (61)	89 (52)	905 (55)	905 (55)
Weight, mean Kg ± SD	102.1 ± 1.0	101.1 ± 1.1	101.0 ± NR	101.1 ± NR	103.3 ± 17.8	106.9 ± 17.8	110.4 ± 16.3	110.6 ± 16.5
BMI, mean Kg/m^2 ± SD	35.6 ± 0.3	35.2 ± 0.2	35.6 ± NR	35.3 ± NR	37.6 ± 5.1	38.0 ± 4.9	37.3 ± 4.2	37.4 ± 4.5
Waist circumference, cm ± SD	NR	NR	NR	NR	112.4 ± 12.8	114.8 ± 13.1	115.0 ± 10.4	115.4 ± 10.4
White, n	CAU 211	CAU 201	NR	NR	NR	NR	NR	NR
Black, n	24	36	NR	NR	NR	NR	NR	NR
Asian, n	NR	NR	NR	NR	NR	NR	NR	NR
Other, n	15	17	NR	NR	NR	NR	NR	NR
T2DM, n (%)	100	100	0	0	14 (8)	14 (8)	0	0
Hypertension, n (%)	NR	NR	NR	NR	25 (15)	31 (18)	NR	NR
Dyslipidaemia, n (%)	NR	NR	NR	NR	HC 51 (30)	HC 49 (29)	NR	NR

Key: CAU = Caucasian, HC= hypercholesterolemia, NR = Not Reported, ORL = Orlistat, PBO = Placebo, T2DM = Type 2 Diabetes Mellitus
Notes: No data were reported for age range, ethnicity, BMI ranges, smokers, alcohol use, history of depression or anxiety

The orlistat trials were well conducted. Four trials were small (fewer than 100 patients) (Derosa 2003, Gotfredsen 2001, Karhunen 2000, Mathus-Vliegen 2006). The orlistat trials were older than the NB32 trials (2000 to 2012 with only two of 16 conducted in the last 10 years as opposed to 2010 to 2013 for the COR programme). Whilst the NB32 trials were conducted exclusively in the US, the orlistat trials were conducted across the world including the UK (Broom 2002), Italy (Derosa 2003 and Derosa 2010), The Netherlands (Mathus-Vliegen 2006), Sweden (Berne 2005, Lindgarde 2000, Torgersen 2004), Denmark (Gotfredsen 2001), Finland (Karhunen 2000), Europe-wide (Astrup 2012), the US (Bakris 2002, Kelley 2002, Lucas 2003, Reaven 2001), the US and Canada (Miles 2002) and Australia and New Zealand (Swinburn 2005).

Mean age of participants varied across the orlistat trials from 41 to 59 years. The NB32 trials ranged from 44 to 54 years for COR-DM. Where reported in the orlistat trials, percentages of female participants varied from 45% to 87%. Most trials had a reasonable proportion of male participants. In contrast 85% to 90% of participants in NB32 trials were female with only COR-DM recruiting 44% males. Only COR-DM recruited patients with diabetes but half of the orlistat trials included at least some patients with diabetes.

According to the CS, 11 orlistat trials had a lead-in period prior to randomisation in which no patients were excluded due to lack of efficacy or treatment compliance. A sensitivity analysis was performed excluding these trials. The NB32 trials did not have a lead-in period.

Across the orlistat trials and between the orlistat and NB32 trials there was variation in the components and delivery of standard care. Standard care is generally not reported in sufficient detail to assess comparability between trials.

Overall, the COR trials and orlistat trials appear comparable. The main difference appears to be that most of the orlistat trials have a more even gender balance than the NB32 trials which are conducted predominantly in women.

For the first outcome (mean percentage weight change from baseline at one year), the analyses performed are shown in Table 4.22, below.

Table 4.22: Number of studies reporting data for ≥5% reduction in weight at one year

Analysis	Trials with patients with T2DM only		Trials excluding patients with T2DM		All trials regardless of T2DM	
	NB32	ORL	NB32	ORL	NB32	ORL
Base case: All trials included	1	3	3	2	4	8
SA1: Trials with 'high' co-morbidities were excluded	0[a]	0[a]	3[b]	2[b]	3	3
SA2: Trials with lead-in periods were excluded	1	2	3	1	4	4
SA3: Trials with intensive BMOD were excluded	1[b]	3[b]	2	1	3	7
SA4: Trials with lead-in periods or intensive BMOD were excluded	1[c]	2[c]	2[a]	0[a]	3	3

Source: CS, Table 34, page 132

Notes: a, Insufficient data available to perform analysis; b, Analysis not performed as evidence base the same as the base case analysis; c, Analysis not performed as evidence base the same as SA2.

BMOD = behaviour modification; NB32 = naltrexone 32mg plus bupropion; NMA = network meta-analysis; SA = sensitivity analysis T2DM = Type 2 diabetes mellitus.

Results are presented as ORs with 95% CI for the direct meta-analyses and as ORs with 95% CrI for the Bayesian NMA (see Table 4.23). An OR < 1 favours NB32 over orlistat or placebo.

For patients with T2DM only, results of sensitivity analyses were either similar to the base case or not performed. Therefore, we will only present base case results.

For patients without T2DM, results from SA3 (excluding trials with intensive BMOD) were the only sensitivity analysis with results different from the base case analysis. Therefore, we will only present base case and SA3 results.

For all patients combined, results of sensitivity analyses were similar to the base case. Therefore, we will only present base case results.

As can be seen from Table 4.23, the Bayesian NMA found no significant differences between NB32 and orlistat for T2DM patients and for all patients combined. There is a statistically significant difference favouring NB32 over orlistat in the analyses excluding studies with T2DM patients, which indicates that more patients receiving NB32 had a ≥ 5% reduction in weight at one year compared to those receiving orlistat. The largest difference was seen in the third sensitivity analysis, where studies with ‘intensive’ behaviour modification (BMOD and XENDOS) were also excluded.

Table 4.23: Results for ≥5% reduction in weight at one year

	PLA	NB32	Placebo vs NB32 (OR, 95% CI)*	PLA	ORL	Placebo vs Orlistat (OR, 95% CI)*	Orlistat vs NB32 (OR, 95% CrI)**
T2DM only	30/159	118/265	0.29 (0.18, 0.46)	87/632	236/617	0.25 (0.17, 0.36)	FE1.09 (0.63, 1.88)
No T2DM - Base case - SA3	 244/1160 162/967	 901/1655 581/1173	 0.25 (0.17, 0.36) 0.21 (0.17, 0.25)	 765/1735 27/98	 1236/1735 42/95	 0.35 (0.23, 0.52) 0.48 (0.26, 0.87)	 FE 0.77 (0.61, 0.96) FE 0.44 (0.23, 0.84)
All patients	274/1319	1019/1920	0.26 (0.19, 0.34)	1052/3101	1841/3068	0.32 (0.26, 0.39)	RE 0.80 (0.51, 1.28)

*) Frequentist Odds Ratio (Non-event) (M-H, Random, 95% CI)
**) Bayesian NMA (OR, 95% CrI)
An OR < 1 favours the second treatment over the first. There are small differences with the results presented in CS because the company presented fixed effect results and we present random effects results for the direct meta-analysis .
RE = results from random effects NMA models which were presented for all patients, FE = results from fixed effect NMA models which were presented for the type 2 DM and no type 2 DM groups due to problems with Bayesian model convergence

For the second outcome (Mean percentage weight change from baseline at one year), the analyses performed are shown in Table 4.24, below.

Table 4.24: Number of studies reporting data for mean percentage weight CFB at one year

Analysis	Trials with patients with T2DM only		Trials excluding patients with T2DM		All trials regardless of T2DM	
	NB32	ORL	NB32	ORL	NB32	ORL
Base case: All trials included	1	4	3	5	4	16
SA1: Trials with 'high' comorbidities were excluded	0[a]	0[a]	3[b]	5[b]	3	8
SA2: Trials with lead-in periods were excluded	1	3	3	1	4	5
SA3: Trials with intensive BMOD were excluded	1[b]	4[b]	2	4	3	15
SA4: Trials with lead-in periods or intensive BMOD were excluded	1[c]	3[c]	2[a]	0[a]	3	4

Source: CS, Table 35, page 133

Notes: a, Insufficient data available to perform analysis; b, Analysis not performed as evidence base the same as the base case analysis; c, Analysis not performed as evidence base the same as SA2.

BMOD = behaviour modification; CFB = change from baseline; NB32 = naltrexone 32mg plus bupropion; NMA = network meta-analysis; SA = sensitivity analysis T2DM = Type 2 diabetes mellitus.

Results are presented as MDs with 95% CIs for the direct meta-analyses and as MDs with 95% CrIs for the Bayesian NMA (see Table 4.25). A MD > 0 favours NB32 over orlistat or placebo and indicates greater % weight reduction.

For patients with T2DM only, results of sensitivity analyses were either similar to the base case or not performed. Therefore, we will only present base case results.

For patients without T2DM, results from SA3 (excluding trials with intensive BMOD) were the only sensitivity analysis with results different from the base case analysis. Therefore, we will only present base case and SA3 results.

For all patients combined, results of sensitivity analyses were similar to the base case. Therefore, we will only present base case results.

As can be seen from Table 4.25, the Bayesian NMA found no significant differences between NB32 and orlistat for T2DM patients and for all patients combined. There is a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients are excluded. The largest difference was seen in the third sensitivity analysis, where studies with ‘intensive’ behaviour modification (BMOD and XENDOS) were also excluded.

Table 4.25: Results for mean percentage weight CFB at one year

	Placebo vs NB32 (MD, 95% CI)*	Placebo vs Orlistat (MD, 95% CI)*	Orlistat vs NB32 (MD, 95% CrI)**
T2DM only	3.20 (2.22, 4.18)	3.63 (2.30, 4.96)	FE 0.21 (-0.87, 1.30)
No T2DM - Base case - SA3	 4.88 (4.34, 5.42) 5.00 (4.41, 5.59)	 2.83 (1.41, 4.25) 2.01 (0.75, 3.27)	 FE 1.13 (0.44, 1.80) FE 2.98 (1.60, 4.36)
All patients	4.39 (3.49, 5.29)	3.00 (2.31, 3.69)	RE 1.39 (-0.08, 2.82)

*) Frequentist Mean Difference (IV, Random, 95% CI)
**) Bayesian NMA
A MD > 0 favours the second treatment over the first and indicates greater % weight reduction.
There are small differences with the results presented in CS because the company presented fixed effect results and we present random effects results for the direct meta-analyses.
RE = results from random effects NMA models which were presented for all patients, FE = results from fixed effect NMA models which were presented for the type 2 DM and no type 2 DM groups due to problems with Bayesian model convergence

ERG comment: Our main problem with these analyses is the use of the mITT populations for the NB32 trials, which we think produce biased results (see Section 4.2.3 of this report). Therefore, we have added analyses using true ITT populations from the NB32 trials in Section 4.5.1 of this report.

4.5 Additional work on clinical effectiveness undertaken by the ERG

In this section we will present two additional analyses:

- Results based on the ITT populations for the NB32 trials
- A comparison of studies with 'intensive' behaviour modification (BMOD vs. XENDOS)
- ERG preferred analyses, including full ITT data and no pooling of NB32 trials

4.5.1 Results based on the ITT populations for the NB32 trials

The company submission used modified intention-to-treat (mITT) data in their analyses. This is common in obesity trials. In fact, most orlistat trials used mITT data in their analyses, which usually included all randomised participants who had a valid baseline measurement and at least one valid measurement after randomisation. The definition of the mITT population in the NB32 trials is quite similar: 'all randomised patients with a post-baseline body weight measurement obtained while the patient remained on study medication'.

We agree that patients should have a baseline weight because otherwise there would be no possibility to calculate weight change. But including only patients that also have at least one post-baseline measurement can introduce bias, because the reason for missing post-baseline measurements could be related to the effectiveness of the treatment. Therefore, ideally the investigators should also present ITT results with some sensitivity analyses looking at different methods of imputing missing follow-up weights.

Additionally, the modified ITT population used in the NB32 trials is different from the mITT population used in the orlistat trials. The term mITT population is therefore misleading. This becomes clear when we look at the difference in the numbers of patients randomised and analysed in the trials. In the NB32 trials, 3,239 patients were analysed out of 3,958 randomised (81.8%); while in the orlistat trials 7,640 patients were analysed out of 7,754 randomised (98.5%). In the intervention arms this was 1,960 patients analysed out of 2,510 randomised (78.1%) for NB32 and 3,884 patients analysed out of 3,946 randomised (98.4%) for orlistat. In other words, in the NB32 trials, 21.9% of patients receiving NB32

were randomised but excluded from the analyses against 1.6% of patients receiving orlistat. Therefore, results of the mITT analyses in the orlistat are more or less the same as the ITT analyses; but in the NB32 trials there may be considerable differences between the types of analyses.

We will present an overview of results for the two main outcome measures (≥5% reduction in weight and percentage weight change from baseline) based on three different analyses: the mITT analysis as presented in the CS, and two ITT analyses (ITT-BOCF = all randomised patients with baseline-observation-carried-forward analysis; and ITT-Imp = all randomised patients with weight regain imputation method analysis). Results presented here are based on the same trial inputs as in the company submission. Therefore, any differences in results are a consequence of ITT versus mITT analyses. That means we have included all four NB32 trials for 'all patients' and three NB32 trials (COR-I, COR-II and COR-BMOD) for 'No type 2 DM'.

For the first outcome, Table 4.26 presents the data used from the four NB32 trials for each analysis and Table 4.27 presents the results for NB32 versus placebo and orlistat for each analysis and population.

Table 4.26: Data synthesised in analyses for ≥5% reduction in weight at one year

Study name	Arm 1	Arm 2	n1	r1	n2	r2
Greenway 2010 (COR-I)						
mITT	NB32	PBO	471	226	511	84
ITT-Imp	NB32	PBO	583	203	581	78
ITT-BOCF	NB32	PBO	583	180	581	67
Apovian 2013 (COR-II)						
mITT	NB32	PBO	702	355	456	78
ITT-Imp	NB32	PBO	878	337	495	73
ITT-BOCF	NB32	PBO	878	308	495	58
Hollander 2013 (COR-DM)						
mITT	NB32	PBO	265	118	159	30
ITT-Imp	NB32	PBO	335	104	170	27
ITT-BOCF	NB32	PBO	335	94	170	24
Wadden 2011 (COR-BMOD)						
mITT	NB32	PBO	482	320	193	82
ITT-Imp	NB32	PBO	591	NR	202	NR
ITT-BOCF	NB32	PBO	591	NR	202	NR
Sensitivity analysis*	NB32	PBO	565	321	196	84

BMOD = behaviour modification; COR = Contrave® obesity research; DM = diabetes mellitus; ITT-BOCF, all randomised patients with baseline-observation- carried-forward analysis; ITT-Imp, all randomised patients with weight regain imputation method analysis; mITT , modified intention-to-treat analysis; n = number of patients; NB32 = naltrexone 32mg plus bupropion; NR = not reported; r = number of patients achieving ≥5% reduction in weight; PBO = placebo.

*) Post-hoc sensitivity analysis for all randomised patients with a baseline and at least one post-baseline body weight measurement (see results on page 72 of CSR[35])

Table 4.27: Bayesian NMA results for ≥5% reduction in weight at one year

Population	Analysis	Orlistat vs NB32	Placebo vs NB32
All patients, mITT	RE	0.80 (0.51, 1.28)	0.25 (0.18, 0.37)
Type 2 DM, mITT	FE	1.09 (0.63, 1.88)	0.29 (0.18, 0.46)
No type 2 DM, mITT	FE	0.77 (0.61, 0.96)	0.24 (0.20, 0.29)
All patients, ITT-Imp	RE	1.14 (0.70, 1.91)	0.36 (0.25, 0.55)
Type 2 DM, ITT-Imp	FE	1.58 (0.91, 2.73)	0.41 (0.25, 0.66)
No type 2 DM, ITT-Imp	FE	1.09 (0.87, 1.36)	0.34 (0.29, 0.40)
All patients, ITT-BOCF	RE	1.11 (0.67, 1.91)	0.36 (0.24, 0.55)
Type 2 DM, ITT-BOCF	FE	1.59 (0.89, 2.79)	0.42 (0.25, 0.68)
No type 2 DM, ITT-BOCF	FE	1.06 (0.84, 1.33)	0.33 (0.28, 0.40)

Results are OR with 95% credible intervals (CrI). An OR < 1 favours the second treatment over the first.
Note: FE model results were presented for the Type 2 DM group and no type 2 DM groups due to problems with model convergence, RE model results were presented for all patients.
DM = diabetes mellitus; FE = fixed effect; ITT-BOCF = all randomised patients with baseline-observation-carried-forward analysis; ITT-Imp = all randomised patients with weight regain imputation method analysis; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; RE = random effects.

For the second outcome, Table 4.28 presents the data used from the four NB32 trials for each analysis and Table 4.29 presents the results for NB32 versus placebo and orlistat for each analysis and population.

Table 4.28: Data synthesised in analysis for percentage weight change from baseline at one year

Study name	Arm 1	Arm 2	n1	M1	SE1	n2	M2	SE2
Greenway 2010 (COR-I)								
mITT	NB32	PBO	471	-6.10	0.30	511	-1.30	0.30
ITT-Imp	NB32	PBO	583	-4.6	0.3	578	-1.2	0.3
ITT-BOCF	NB32	PBO	583	-4.0	0.3	578	-0.9	0.3
Apovian 2013 (COR-II)								
mITT	NB32	PBO	702	-6.40	0.30	456	-1.20	0.30
ITT-Imp	NB32	PBO	878	-4.9	6.5 (SD)	495	-1.2	6.7 (SD)
ITT-BOCF	NB32	PBO	878	-4.4	0.2	495	-0.8	0.3
Hollander 2013 (COR-DM)								
mITT	NB32	PBO	265	-5.00	0.30	159	-1.80	0.40
ITT-Imp	NB32	PBO	335	-3.5	0.3	170	-1.7	0.4
ITT-BOCF	NB32	PBO	335	-3.1	0.3	170	-1.3	0.4
Wadden 2011 (COR-BMOD)								
mITT	NB32	PBO	482	-9.30	0.40	193	-5.10	0.60
ITT-Imp	NB32	PBO	591	NR	NR	202	NR	NR
ITT-BOCF	NB32	PBO	591	-5.9	0.4	202	-4.0	0.6

BMOD = behaviour modification; COR = Contrave® obesity research; DM, diabetes mellitus; M = mean; n = number of patients; NB32 = naltrexone 32mg plus bupropion; NR = not reported; PBO = placebo; SD = standard deviation; SE = standard error.

Table 4.29: Bayesian NMA results for percentage weight change from baseline at one year

Population	Analysis	Orlistat vs NB32	Placebo vs NB32
All patients, mITT	RE	1.39 (-0.08, 2.82)	4.38 (3.15, 5.63)
Type 2 DM, mITT	FE	0.21 (-0.87, 1.30)	3.21 (2.23, 4.21)
No type 2 DM, mITT	FE	1.13 (0.44, 1.80)	4.88 (4.35, 5.43)
All patients, ITT-Imp	RE	0.26 (-1.23, 1.71)	3.25 (1.98, 4.51)
Type 2 DM, ITT-Imp	FE	-1.21 (-2.30, -0.11)	1.80 (0.83, 2.79)
No type 2 DM, ITT-Imp	FE	-0.09 (-0.77, 0.58)	3.65 (3.15, 4.17)
All patients, ITT-BOCF	RE	-0.31 (-1.81, 1.09)	2.68 (1.38, 3.89)
Type 2 DM, ITT-BOCF	FE	-1.21 (-2.30, -0.11)	1.80 (0.83, 2.79)
No type 2 DM, ITT-BOCF	FE	-0.54 (-1.21, 0.12)	3.20 (2.70, 3.71)

Results are mean difference with 95% credible intervals (CrI).
A MD > 0 favours the second treatment over the first and indicates greater % weight reduction.
Note: FE model results were presented for the Type 2 DM group and no type 2 DM groups due to problems with model convergence, RE model results were presented for all patients.
DM = diabetes mellitus; FE = fixed effect; ITT-BOCF = all randomised patients with baseline-observation-carried-forward analysis; ITT-Imp = all randomised patients with weight regain imputation method analysis; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; RE = random effects.

As can be seen from Tables 4.28 and 4.29, the positive effects of NB32 when compared to orlistat have all disappeared. For the first outcome (≥5% reduction in weight at one year), there was a statistically significant difference using mITT data favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: OR = 1.09 (95% CrI: 0.87 to 1.36), ITT-BOCF: OR = 1.06 (95% CrI: 0.84 to 1.33). Moreover, although none of the differences are statistically significant, all results now favour orlistat.

For the second outcome (mean percentage weight change at one year), there was a statistically significant difference using mITT data favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: MD= -0.09 (95% CrI: -0.77 to 0.58), ITT-BOCF: MD = -0.54 (95% CrI: -1.21 to 0.12). Moreover, although most of the differences are not statistically significant, most results now favour orlistat.

4.5.2 Comparison of intensive trials BMOD and XENDOS

One trial of orlistat[56] and one of NB32[26] were considered to include 'intensive' behaviour therapy. Brief details of the criteria used to define 'intensive' behaviour modification were provided in section 4.10.4 of the CS. However exact details of the criteria used were not provided. These two trials were excluded in sensitivity analysis 3 of the network meta-analysis to assess the robustness of the treatment effect.

XENDOS was a four year trial of orlistat conducted in Sweden. COR-BMOD was a 56 week trial of NB32 conducted in the US. XENDOS randomised 3,305 patients and COR-BMOD 793. A comparison of the participants, interventions and comparators, outcomes and study designs is given in Table 4.30.

Table 4.30: Comparison of intensive trials: COR-BMOD and XENDOS

	COR-BMOD	XENDOS
Participants	• Age 18 to 65 with • BMI 30 to 45 kg/m2 and uncomplicated obesity OR • BMI 27 to 45 kg/m2 and controlled hypertension and / or dyslipidaemia • Patients with diabetes excluded	• Age 30 to 60 with • BMI >= 30 • Patients with diabetes excluded
Intervention and Comparator	NB32+Behaviour modification(BMOD) vs. Placebo + BMOD BMOD consisted of group meetings lasting 90 minutes weekly for the first 16 weeks, then every other week for the next 12 weeks and monthly thereafter. They included instructions to consume a balanced deficit diet and to increase to 180 min/week of planned, moderately vigorous, physical activity (CS, page 57).	Orlistat +Lifestyle changes vs. Placebo + Lifestyle changes All patients prescribed a reduced calorie diet (approx. 800 kcal/day deficit) 30% from fat and no more than 300mg cholesterol per day. Readjusted every six months to account for weight loss during preceding months. Dietary counselling every 2 weeks for first 6 months and monthly thereafter. All kept physical activity diaries.
Primary outcome	Percentage of change in total body weight and proportion of patients with ≥ 5% decrease in total body weight at week 56 using modified intention-to-treat data.	Time to onset of type 2 diabetes and change in body weight after 4 years' treatment using intention-to-treat data.
Study design	RCT	RCT
Source: CS[1] and Torgerson 2004[56] (XENDOS)		

Details of the participant characteristics in the two trials can be found in Table 4.31.

Table 4.31: Comparison of participants in COR-BMOD and XENDOS

	COR-BMOD		XENDOS	
	NB32	Pbo	ORL	Pbo
No randomised	591	202	1640	1637
Age, mean years (SD)	45.9 (10.4)	45.6 (11.4)	43.0 (8.0)	43.7 (8.0)
Sex, female, n (%)	528 (89.3)	185 (91.6)	905 (55.2)	905 (55.3)
BMI, mean kg/m^2 (SD)	36.3 (4.2)	37.0 (4.2)	37.3 (4.2)	37.4 (4.5)
Weight, mean kg (SD)	100.2 (15.4)	101.9 (15.0)	110.4 (16.3)	110.6 (16.5)
Source: CS[1] and Torgerson 2004[56] (XENDOS)				

In XENDOS significantly more patients in the orlistat group (72.8%) than in the placebo group (45.1%) achieved weight loss ≥ 5% after one year of treatment. In BMOD 66.4% of patients in the NB32 group and 42.5% in the placebo group had a weight loss ≥ 5% (a statistically significant result, based on 482 and 193 patients in the mITT analysis). In terms of ≥ 10% weight loss, in XENDOS significantly more patients in the orlistat group were successful (41.0% of orlistat patients vs. 20.8% of placebo patients). In BMOD 41.5% of patients in the NB32 group and 20.2% in the placebo group had a weight loss ≥ 10% (a statistically significant result, based on 482 and 193 patients in the mITT analysis).

During the first year of treatment, the proportion of patients experiencing at least one gastrointestinal event with orlistat or placebo in XENDOS was 91% vs. 65%, respectively. In COR-BMOD 65.1% of patients experienced gastrointestinal disorders in the NB32 group and 39% in the placebo group. Overall, 4% of placebo patients and 8% of orlistat patients withdrew from XENDOS because of adverse events or laboratory abnormalities; the difference was primarily due to gastrointestinal events. In COR-BMOD a greater percentage of those in the NB32 group discontinued due to an adverse event (25.4% vs. 12.4%, $p < 0.001$).

ERG comment:

- Although interventions in COR-BMOD and XENDOS could both be considered intensive, the nature of the co-intervention delivered varied in terms of delivery, intensity and advice components.
- Participant inclusion criteria were similar and both trials excluded patients with diabetes. COR-BMOD had a greater proportion of female participants than XENDOS. Participants in XENDOS were, on average approximately 10kg heavier.
- XENDOS specifically considered time to onset of type 2 diabetes in addition to change in body weight as a primary outcome.
- Both trials found active treatment with a drug to be superior to lifestyle management alone in terms of 5% or 10% weight loss. Although there were a large number of gastrointestinal events in the XENDOS trial, discontinuation due to adverse events was lower than that noted in the COR-BMOD trial.

The results of the indirect comparison of NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification, using data from COR-BMOD versus XENDOS, are presented in the Table below.

Table 4.32: Indirect comparison results for COR-BMOD versus XENDOS

Population	NB 32 vs placebo	Orlistat vs placebo	Orlistat vs NB32
≥5% reduction in weight at 1 year			
mITT	2.67 (1.90, 3.77)	3.26 (2.82, 3.77)	1.22 (0.84, 1.77)
ITT-Imp	NR		NR
ITT-BOCF	1.75 (1.26, 2.44)	3.26 (2.82, 3.77)	1.86 (1.30, 2.66)
Mean % weight CFB at 1 year			
mITT	-4.20 (-5.62, -2.78)	-3.99 (-4.46, -3.52)	-0.21 (-1.28, 1.70)
ITT-Imp	NR		NR
ITT-BOCF	-1.9 (-3.27, -0.53)	-3.99 (-4.46, -3.52)	-2.09 (-3.53, -0.65)

Results are OR with 95% CI for ≥5% reduction in weight at 1 year and mean difference (MD) with 95% CI for mean % weight CFB at 1 year. The analysis uses the Bucher method for indirect comparisons.
An OR < 1 favours the second treatment over the first. A MD > 0 favours the second treatment over the first and indicates greater % weight reduction.
CFB = Change from baseline; ITT-BOCF = all randomised patients with baseline-observation-carried-forward analysis; ITT-Imp = all randomised patients with weight regain imputation method analysis; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; NR = Not reported.

Results in Table 4.32 show that results are different dependent on which dataset is used. When using the mITT results for ≥ 5% reduction in weight at one year there is no significant difference between NB32 and orlistat. However when using the ITT BOCF results (ITT-Imp results were not available for

COR-BMOD) the results are statistically significant and favour orlistat (OR 1.86, 95% CI 1.30 to 2.66). For the percentage weight change from baseline there was also no significant difference between NB32 and orlistat when using the mITT results. However, when using the ITT-BOCF results there was a statistically significant difference which favoured orlistat (MD -2.09, 95% CI -3.53 to -0.65).

4.5.3 ERG preferred analyses

In Section 4.5.1 we have used the same trial inputs as in the company submission. Therefore, any differences in results were a consequence of ITT versus mITT analyses. That means we have included all four NB32 trials for 'all patients' and three NB32 trials (COR-I, COR-II and COR-BMOD) for 'No type 2 DM'.

As explained in Section 4.2.5, we would prefer not to pool any of the NB32 trials. That means, we will not present any results for 'all patients', because this would be a mix of diabetes patients in the COR-DM trial and non-diabetes patients in the other three COR trials. In addition, different interventions (standard management and intensive behaviour modification) would be pooled.

Therefore, we would preferably use only trials that include T2DM patients for obese patients with diabetes, and only trials that do not included T2DM patients for obese patients without diabetes in our analyses. In addition, we will not include trials with intensive behaviour modification (COR-BMOD for NB32, and XENDOS for orlistat) in these analyses. That means that the analyses for obese patients with diabetes are the same as before, but results for obese patients without diabetes will change as now only COR-I for NB32 and Astrup et al. (2012)[41] for orlistat are included.

Table 4.33 shows that the results for 'obese patients with T2DM' and 'intensive behaviour modification' are the same as in Sections 4.5.1 and 4.5.2, respectively. However, results for 'obese patients without T2DM' have changed considerably again, and are almost the same as in the company's original analyses. Both outcomes show no significant difference between NB32 and orlistat, but both favour NB32.

Table 4.33: ERG preferred analyses compared to other results

Population		Company analyses (mITT data)*	Company analyses (ITT-BCFA data)**	ERG preferred analyses**
		Orlistat vs NB32	Orlistat vs NB32	Orlistat vs NB32
Obese people with T2DM				
≥5% reduction in weight at 1 year	OR	1.09 (0.63 to 1.88)	1.59 (0.89 to 2.79)	1.59 (0.89 to 2.79)
Mean % weight CFB at 1 year	MD	0.21 (-0.87 to 1.30)	-1.21 (-2.30 to -0.11)	-1.21 (-2.30 to -0.11)
Obese people without T2DM				
≥5% reduction in weight at 1 year	OR	0.77 (0.61 to 0.96)	1.06 (0.84 to 1.33)	0.61 (0.31 to 1.22)
Mean % weight CFB at 1 year	MD	1.13 (0.44 to 1.80)	-0.54 (-1.21 to 0.12)	1.11 (-0.39 to 2.63)
Intensive behaviour modification				
≥5% reduction in weight at 1 year	OR	1.22 (0.84 to 1.77)	1.86 (1.30 to 2.66)	1.86 (1.30 to 2.66)
Mean % weight CFB at 1 year	MD	-0.21 (-1.28 to 1.70)	-2.09 (-3.53to -0.65)	-2.09 (-3.53to -0.65)

Population		Company analyses (mITT data)*	Company analyses (ITT-BCFA data)**	ERG preferred analyses**
		Orlistat vs NB32	Orlistat vs NB32	Orlistat vs NB32
Results are OR with 95% CI/CrI for ≥5% reduction in weight at 1 year and mean difference (MD) with 95% CI/CrI for mean % weight CFB at 1 year. An OR less than one favours NB32 over orlistat and a CI including 1 is not significant. A MD of >0 favours NB32 over orlistat and indicates greater % weight reduction and a CI including 0 is not significant. *) Bayesian NMA (OR, 95% CrI) using mITT data; **) Using the Bucher method for indirect comparisons and ITT-BCFA data. FE = fixed effect; ITT-BCFA = all randomised patients with baseline-carried-forward analysis; MD = Mean Difference; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; OR = Odds Ratio; T2DM = Type 2 diabetes mellitus;				

4.6 Conclusions of the clinical effectiveness section

The company conducted a systematic review to identify studies comparing NB32 to the comparators outlined in the NICE scope.[7] Relevant direct evidence comparing NB32 and placebo has been presented. However no trials directly comparing NB32 to orlistat were identified. Indirect comparisons were made between NB32 and orlistat.

The company submission focused on data from the four pivotal RCTs: COR-I, COR-II, COR-BMOD, and COR-DM. All of these RCTs compare NB32 to placebo with both arms receiving standard care. Standard care varies between the trials in that COR-BMOD has a more intensive form of behavioural management. In addition, COR-DM focused exclusively on patients with diabetes whilst the other trials exclude patients with diabetes. The ERG agrees that there was clinical and statistical heterogeneity between the four COR trials and that because of this the results from the separate analyses for patients with and without diabetes should be preferred and BMOD may not be suitable to be pooled with the other COR trials.

The NB-CVOT study was included in the submission as a supporting study as it presented longer term outcomes. NB-CVOT represents an older population with cardiovascular disease when compared to the COR trials. Most of the patients in NB-CVOT are diabetic, and many are depressed. A number of problems with the study were identified. NB-CVOT used a lead-in period where large numbers of patients discontinued primarily due to adverse events. This implies that those continuing to the treatment period who were re-randomised were better able to tolerate the drug. The adverse event profile will be an overestimate of the tolerability of the drug. In NB-CVOT only SAEs and AEs leading to study drug discontinuation were collected. Even so, an elevated number of gastrointestinal events were noted in the NB32 group. NB-CVOT was terminated early (after the 50% interim analysis), after 25% interim data were made public. The trial was not able to provide a definitive answer to the cardiovascular risk of NB32 and a further trial has been instigated. The reliability of the final data on weight loss is also questionable.

The COR trials were of high quality. However more patients dropped out of NB32 groups due to adverse events. Higher rates of adverse events (especially nausea) could have resulted in un-blinding of participants. The modified intention-to-treat analysis presented in the submission reflects only those who have a post-baseline measurement whilst on the study drug. Any discontinuations before the post-baseline weight assessment are discounted. Reasons for discontinuation could relate to efficacy or safety of the drug. Using the true ITT data, NB32 is still superior to placebo in terms of weight loss but results are more modest.

A number of points should be borne in mind when applying the results of the NB32 trials to clinical practice:

- Overweight patients in addition to obese patients were included in the NICE scope. However only a very small percentage (approximately 2%) of patients who are overweight are in the COR trials. Therefore this population is not well represented. Mean BMI in the trials is 36 to 37 which is severely obese.
- All of the COR trials were conducted in the US so participant characteristics and the nature of standard care may differ from a UK setting.
- Prior use of orlistat was an exclusion criterion in all four COR trials. Therefore the effect of NB32 on those who have failed on orlistat has not been examined.
- The majority of participants in the COR trials are female. The ERG draws to the attention of the committee that this does not reflect the distribution of obesity according to gender. Men in England are more likely to be overweight or obese (68% vs 58% in 2015).
- Asian patients are not well represented in the COR trials so results may not be applicable to these ethnic groups.
- Three of the four COR trials measure the primary outcome at 56 weeks. Although this is acceptable in terms of weight loss, there is no information on maintenance of weight loss after this time.
- The CS states that "*For patients continuing treatment post 16 weeks, treatment should be continued as long as clinical benefit is observed*."[1] It is unclear how long patients would continue to take the drug in practice.
- There are no data on the effectiveness of retreatment with NB32 following successful treatment with NB32 and subsequent discontinuation and weight gain.
- There were large dropout rates across the COR trials (up to 50%). This suggests that in practice up to half of patients may complete a year's treatment with NB32 which is relevant when considering transferability to clinical practice.
- Based on the mITT data presented by the company NB32 results in greater weight loss and in a higher number reporting 5% or more weight loss. However the superior results regardless of arm in the BMOD trial are of interest. NB32 together with a more intensive behaviour modification programme resulted in 66.4% of patients losing 5% or more weight compared to 44 to 55% in the other three trials without such an intensive intervention. In the BMOD trial the placebo and behaviour modification arm achieved results approaching the medication arms in the other trials.
- A greater proportion of gastrointestinal events, particularly nausea, were noted in NB32 groups across the trials. Although the majority of events were not serious, more participants withdrew as a result of adverse events in treatment groups.

A comparison between NB32 (plus standard management) versus intensive behaviour modification is missing. In its response to the clarification letter (Question A12, page 13), the company stated that "*the anticipated positioning of NB32 in the treatment pathway is for patients eligible for pharmacological treatment (alongside standard management)"*. Therefore, the company considered different types of behaviour modification not relevant to the decision problem. However, the NICE scope clearly mentions 'standard management without naltrexone-bupropion' as a comparator and this may very well include more intensive forms of behaviour modification than patients receiving instructions to follow a diet and increase physical activity, and written behaviour modification advice. Moreover, results from the COR-BMOD trial show that more intensive behaviour modification is still quite effective in patients eligible for pharmacological treatment. At first glance it seems that intensive behaviour modification in

the COR-BMOD trial (percentage change from baseline: -5.1 (SE: 0.6)) has similar effects as NB32 in the COR-I trial ((percentage change from baseline: -6.1 (SE: 0.3)). Therefore, a comparison of NB32 vs. intensive behaviour modification would have been of interest.

Regarding the comparison of NB32 with orlistat, the company used modified ITT data from NB32 trials, but this is misleading. The mITT population in the NB32 trials is very different from mITT populations in the orlistat trials. In the NB32 trials, 21.9% of patients receiving NB32 were randomised but excluded from the analyses against 1.6% of patients receiving orlistat.

The comparison with orlistat may be biased in favour of NB32. NB32 trials were published in 2010 or later; most of the trials with orlistat were published before 2005, so caution should be exercised when making indirect comparisons; this is particularly true for conditions such as diabetes where background standard therapy (for glucose and lipids especially) may be very different now.

We have reproduced the company's indirect analyses using full ITT data from the NB32 trials and we have included a new analysis: an indirect comparison of NB32 plus intensive behaviour modification (COR-BMOD) versus orlistat plus intensive behaviour modification (XENDOS). The results show that the positive effects of NB32 when compared to orlistat have all disappeared. For the first outcome (≥5% reduction in weight at one year), there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded when using mITT data. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: OR = 1.09 (95% CrI: 0.87 to 1.36), ITT-BOCF: OR = 1.06 (95% CrI: 0.84 to 1.33). Moreover, although none of the differences are statistically significant, all results now favour orlistat.

For the second outcome (mean percentage weight change at one year), there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded when using mITT data. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: MD= -0.09 (95% CrI: -0.77 to 0.58), ITT-BOCF: MD = -0.54 (95% CrI: -1.21 to 0.12). Moreover, although most of the differences are not statistically significant, most results now favour orlistat.

The results of the indirect comparison of NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification, using data from COR-BMOD versus XENDOS, show that both outcomes significantly favour orlistat over NB32 (≥5% reduction in weight at one year: OR 1.86 (95% CI: 1.30 to 2.66); Mean percentage weight CFB at one year: MD -2.09 (95% CI: -3.53 to -0.65)). This is particularly relevant, as the committee might assume that those who are prescribed NB32 or orlistat might want to participate in a weight loss programme. In that case, the BMOD trial might provide a better estimate of the effect of NB32 as an adjunct to standard management.

Finally, we performed our preferred analyses, i.e. using full ITT data and no pooling of NB32 trials. The results for 'obese patients with T2DM' and 'intensive behaviour modification' are the same as before, but results for 'obese patients without T2DM' have changed considerably again, and are almost the same as in the company's original analyses. Both outcomes show no significant difference between NB32 and orlistat, but both favour NB32.

The table below shows the main results for obese people with diabetes, obese people without diabetes and NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification.

Table 4.34: Company results versus ERG results

Population		Company analyses (mITT data)*	Company analyses (ITT-BCFA data)**	ERG preferred analyses**
		Orlistat vs NB32	Orlistat vs NB32	Orlistat vs NB32
Obese people with T2DM				
≥5% reduction in weight at 1 year	OR	1.09 (0.63 to 1.88)	1.59 (0.89 to 2.79)	1.59 (0.89 to 2.79)
Mean % weight CFB at 1 year	MD	0.21 (-0.87 to 1.30)	-1.21 (-2.30 to -0.11)	-1.21 (-2.30 to -0.11)
Obese people without T2DM				
≥5% reduction in weight at 1 year	OR	0.77 (0.61 to 0.96)	1.06 (0.84 to 1.33)	0.61 (0.31 to 1.22)
Mean % weight CFB at 1 year	MD	1.13 (0.44 to 1.80)	-0.54 (-1.21 to 0.12)	1.11 (-0.39 to 2.63)
Intensive behaviour modification				
≥5% reduction in weight at 1 year	OR	1.22 (0.84 to 1.77)	1.86 (1.30 to 2.66)	1.86 (1.30 to 2.66)
Mean % weight CFB at 1 year	MD	-0.21 (-1.28 to 1.70)	-2.09 (-3.53to -0.65)	-2.09 (-3.53to -0.65)

Results are OR with 95% CI/CrI for ≥5% reduction in weight at 1 year and mean difference (MD) with 95% CI/CrI for mean % weight CFB at 1 year.
An OR less than one favours NB32 over orlistat and a CI including 1 is not significant. A MD of >0 favours NB32 over orlistat and indicates greater % weight reduction and a CI including 0 is not significant.
*) Bayesian NMA (OR, 95% CrI) using mITT data; **) Using the Bucher method for indirect comparisons and ITT-BCFA data.
FE = fixed effect; ITT-BCFA = all randomised patients with baseline-carried-forward analysis; MD = Mean Difference; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; OR = Odds Ratio; T2DM = Type 2 diabetes mellitus;

5. COST EFFECTIVENESS

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

This section pertains mainly to the review of cost effectiveness analysis studies. However, the search section (5.1.1) also contains summaries and critiques of other searches related to cost effectiveness presented in the company submission. Therefore, the following section includes searches for the cost effectiveness analysis review, measurement and evaluation of health effects as well as for cost and healthcare resource identification, measurement and valuation.

5.1.1 Objectives of cost effectiveness searches and reviews

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the company submission.

Objectives of cost effectiveness analysis search and review

The CS reported that searches were carried out in May 2016. Searches contained a 10 year date limit, but were not limited by language. Searches were carried out on the following databases: Embase, MEDLINE, MEDLINE in-Process, HTA and NHS EED via the Cochrane library and Econlit. Searches were carried out in line with the NICE 2013 guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4.[57] Supplementary searches of the following conference proceedings were reported: International Congress on Obesity (ICO), European Congress on Obesity by the European Association for the Study of Obesity (ECO), International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Annual European Congress and ISPOR Annual International Congress. Along with searches of both the NICE and Public Health England websites, the CS also reported that "*bibliographic searches of published systematic reviews, economic models and health technology assessments (HTAs) were conducted.*"[38]

ERG comment: The majority of searches in Appendix 15 were well reported and easily reproducible. Table 22 reported the use of the SIGN study design filter for economics.[38] Unlike the clinical effectiveness section the filter devised for MEDLINE was used for the joint MEDLINE/Embase search, however the remaining condition and interventions facets of the strategy employed only Emtree terms and no MeSH. As stated in Section 4.1.1, although some mapping between indexing terms does take place on Embase.com it is possible that in this case some relevant Emtree/MeSH indexing terms will not be included in the search, and potentially relevant records could have been missed.

The ERG also noticed a number of areas for concern relating to the Econlit search reported for this section. Firstly the ERG asked the company to confirm that this search was conducted using the EBSCO platform as stated in Table 22, Appendix 15. The ERG noted the inclusion of the MH field tag in Table 26, which the ERG understands is not supported by EBSCO as field in Econlit. The company responded that the MH search functionality was incorrectly presented in Table 22. Further to this the ERG noted that the strategy appeared to contain an error in the line numbers being combined in lines S60 and S61. The line above (S59) had the combination "S11 AND S25 AND S58" which appeared to be correct; however the following two lines had the combination "S11 AND S22 AND S58. Line S25 was a combination of all listed interventions where line S22 was for "TI (lorcaserin OR belviq) OR AB (lorcaserin OR belviq)". The company confirmed that this was also due to a reporting error and provided a full revised strategy in their response to clarification. Finally despite being a pre-filtered specialist resource, as with the Cochrane strategies reported in the clinical effectiveness section, the Econlit strategy contained a redundant economics filter, which may have unnecessarily restricted the results

retrieved. However given other searches reported for this section, this is unlikely to have impacted on the overall recall of results.

Objectives of search and review for measurement and valuation of health effects

Searches were conducted to "*identify utility values associated with overweight (with at least one comorbidity) and obese conditions and their associated treatments*". [38]

Searches were carried out in June 2016 across a good range of databases. No date or language limits were applied. The company reported that the supplementary database and conference websites searched for modelling studies were also searched for utility studies.

ERG comment: Searches were well reported and easily reproducible. As with the previous sections the ERG had some concerns regarding the use of only Emtree indexing terms. Despite some mapping between indexing terms on Embase.com the same limitations as described in Section 4.1.1 will apply. The Econlit strategy for this section also contained the unsupported use of the MH field tag which the company reported as a presentation error in their response to clarification.

Objectives of search and review for cost and healthcare resource identification, measurement and valuation

A systematic literature review was conducted to "*identify the economic burden of obesity and associated treatments, in terms of healthcare resource utilisation as well as direct and indirect costs.*"[38]

Searches were carried out in June 2016 on a good range of databases. As with the previous sections supplementary searches of conference proceedings and other relevant websites were carried out in order to identify cost and resource use studies. As with the economics section, searches were limited to the last 10 years and for this section only data from the UK was sought.

ERG comment: Searches were well reported and easily reproducible. The same errors regarding the use of the unsupported MH field tag in the Econlit search and limited indexing terms on Embase.com searches appeared in these searches as for earlier sections, with regard to the latter the same limitations will apply.

5.1.2 Inclusion/exclusion criteria used in the study selection

The pre-specified eligibility are shown as a PICOS table in Table 52 of the CS[1] for cost effectiveness analysis studies, in Table 59 of the CS[1] for measurement and valuation of health effects studies, and in Table 61 of the CS[1] for cost and healthcare resource use studies.

ERG comment: The ERG was satisfied that the company's inclusion and exclusion criteria used in the study selection were appropriate for the three searches.

5.1.3 Included/excluded studies in the cost effectiveness review

The search identified 1,792 citations, of which 1,781 were identified through database searching, one additional study through bibliographic searching and 10 abstracts were identified from conference proceedings. After screening and eligibility assessment, 81 references were deemed eligible for full-text evaluation. Nineteen studies from 22 included publications met the inclusion criteria. Table 27 in Appendix 15 of the CS[1] provides a tabular overview of the included studies.

The following is an overview of the company's findings from the review, as reported in the CS[1]:

- None of the included studies considered NB32 as an intervention

- Four studies were set in the UK (Ara et al., 2012[58], Davies et al., 2012[59], Burch et al., 2009[60], Beaudet et al., 2011[61])
- Pharmacological treatment for obesity has the potential of being cost effective
- Results were particularly sensitive to uncertainty surrounding assumptions concerning duration of weight maintenance after initial weight loss and the effect of a reduction in BMI on health-related quality of life (HRQoL)
- A variety of model types and structures was used across the included studies, with most studies using timed cohort models, and one study using an individual-level timed model,[58] which was deemed most appropriate by the company.

The NICE and PHE website search identified only one result: the "Weight Management Economic Assessment Tool". The company reports that it "*helps healthcare professionals assess existing or planned weight management interventions and to allow commissioners to compare the costs of an intervention for English patients with potential cost savings*.[62]" The tool has been developed by PHE in conjunction with a panel of experts.[62]

ERG comment: The ERG considered that the searches and review were unlikely to miss any important studies and considers the company's conclusions as appropriate.

5.1.4 Conclusions of the cost effectiveness review

The CS provides an overview of the included studies but no specific conclusion regarding the cost effectiveness of NB32, or other pharmacological treatments, is formulated.

ERG comment: Since the identified studies did not consider NB32 as an intervention, the ERG agrees that no specific conclusion could be drawn from this review.

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

Table 5.1: Summary of the company's economic evaluation (with signposts to CS)

	Approach	Source / Justification	Signpost (location in CS)
Model	A DES model was implemented in Excel using the "discretely integrated condition event" (DICE) principles and structure	It was argued that an individual-level approach is better suited than a cohort-level approach to capture the chronic implications of both weight and weight-related health events in a heterogeneous group of overweight and obese patients.	Sections 5.2.2.1 and 5.2.2.2
States and events	Events: - treatment discontinuation - development of T2DM - first cardiovascular event - second cardiovascular event - death	The company used the economic evaluation by Ara et al.[58] as a starting point.	Sections 5.2.2.1 and 5.2.2.3
Comparators	- orlistat as an adjunct to standard management and;	Consistent with the scope and licensed indications	Section 5.2.4

	Approach	Source / Justification	Signpost (location in CS)
	- standard management alone		
Population	The company stated that the model aimed to reflect adult patients who are obese (BMI ≥30kg/m2) or overweight (BMI ≥27kg/m2 and <30kg/m2) in the presence of one or more weight-related comorbidities (e.g., T2DM, dyslipidaemia, or controlled hypertension).	NB32 is licensed as an adjunct to standard non-pharmacological management for this population.	Section 5.2.1
Treatment effectiveness	Treatment effectiveness is estimated based on reduced weight / BMI (retrieved from COR trial programme) and subsequent reduced risk of obesity-related comorbidities (based on the economic evaluation by Ara et al.[58]).		Sections 5.3.2 to 5.3.4
Adverse events	Costs were considered for AEs that occurred in at least 5% of patients (either treatment arm) in the COR-I trial. No disutilities related to adverse events were considered.	The 5% threshold was selected to reflect the British National Formulary criteria of all very common (> 1 in 10) and the majority of common (1 in 100 to 1 in 10) AEs. Moreover, the company stated that quality of life implications of adverse events were deemed to be too poorly understood to incorporate disutilities associated with adverse events.	Sections 5.4.3, 5.4.4 and 5.5.4
Health related QoL	The HRQL data used in the cost-effectiveness analysis are estimated based on Tobit regression analysis of EQ-5D individual-level data from a recent Health Survey for England.	The Tobit model includes explanatory variables for BMI, age, gender, and the obesity-related conditions in the economic model as well as cancer, and are therefore well suited to inform utility assumptions in the model.	Section 5.4
Resource utilisation and costs	Costs in the model consisted of drug acquisition costs, non-drug costs related to standard management (applicable to all	Considering the studies identified in the review, the company stated that the level of reporting was generally poor across studies, to the extent that it was difficult to	Section 5.5

	Approach	Source / Justification	Signpost (location in CS)
	treatments considered), obesity-related comorbidity costs and adverse event costs. These costs were primarily based on Ara et al.,[58] NHS reference costs and PSSRU.	elicit useful resource use estimates for this analysis. A notable exception to this was the study by Ara et al..[58] Hence the company used this study to inform healthcare resource use assumptions.	
Discount rates	Discount rate of 3.5% for utilities and costs	As per NICE reference case	Table 54
Sub groups	Stratified based on T2DM.	As per NICE scope	Section 5.9
Sensitivity analysis	Both DSA and PSA were performed as well as scenario analyses		Section 5.8

Source: CS
Abbreviations: DSA, deterministic sensitivity analysis; PSA, probabilistic sensitivity analysis; PSSRU, Personal Social Services Research Unit; T2DM, type 2 diabetes; DES, discrete event simulation; BMI, body mass

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.2: NICE reference case checklist

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Population	As per NICE scope	Yes	
Comparator(s)	Therapies routinely used in the National Health Service (NHS), including technologies regarded as current best practice	Yes	
Type of economic evaluation	Cost effectiveness analysis	Yes	
Perspective on costs	NHS and Personal Social Services (PSS)	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	
Synthesis of evidence in outcomes	Systematic review	Yes	

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Measure of health effects	Quality adjusted life years (QALYs)	Yes	
Source of data for measurement HRQoL	Described using a standardised and validated instrument	Yes	
Source of preference data for valuation of changes in HRQoL	Time-trade off or standard gamble	Yes	
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic modelling	Partly	The number of simulated patients (1,000) is too low to provide stable results The PSA does not incorporate all relevant parameters (e.g. the uncertainty surrounding time to treatment discontinuation, a key parameter in the model, was neglected). The number of PSA simulations (100) is too low to provide stable results.
Source: CS Abbreviation: PSA, probabilistic sensitivity analysis			

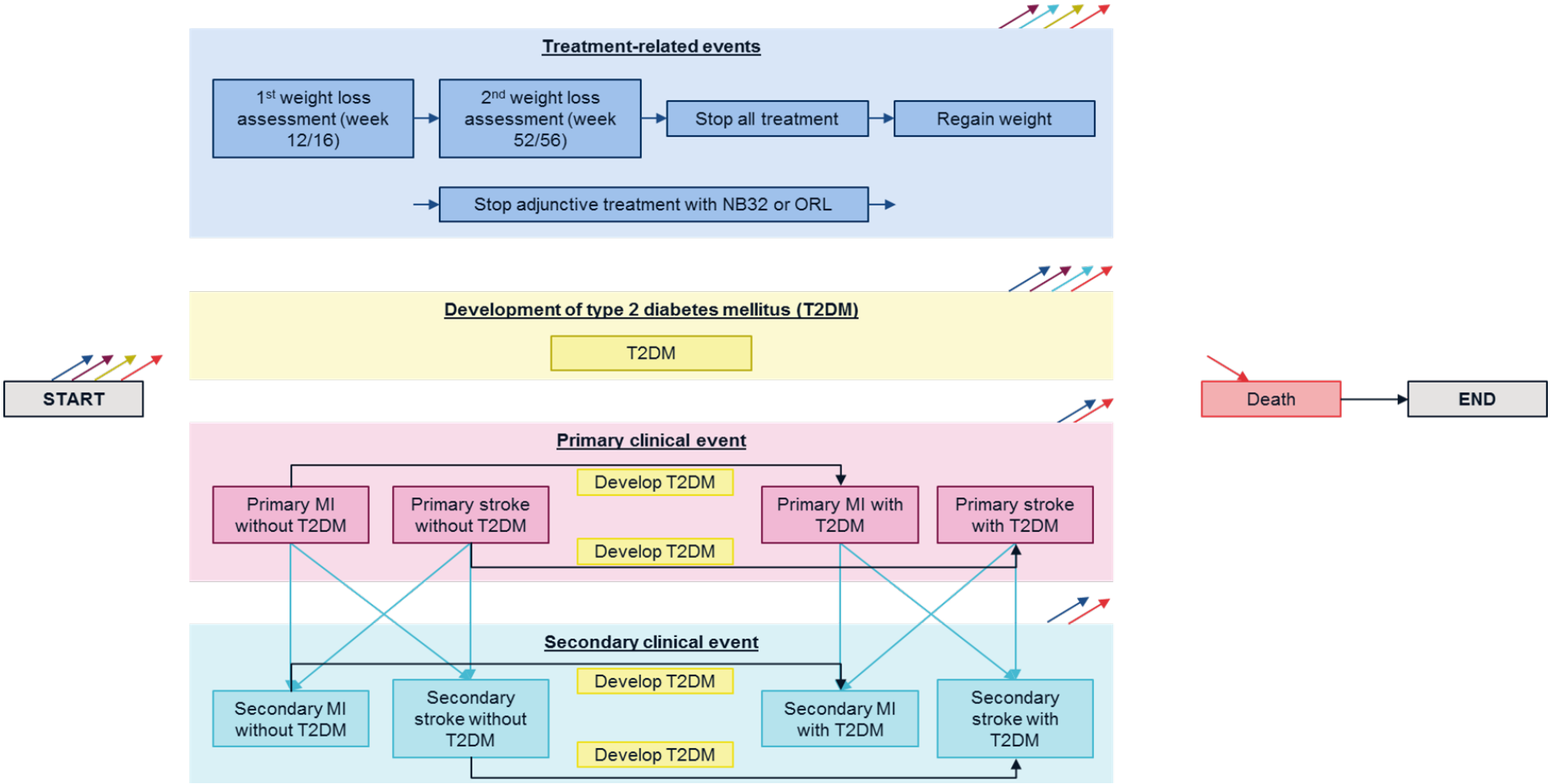
5.2.2 Model structure

The company developed a de novo economic model using an individual-level approach, more specifically a discrete event simulation (DES). It was argued that an individual-level approach is better suited than a cohort-level approach to capture the chronic implications of both weight and weight-related health events in a heterogeneous group of overweight and obese patients. The DES model was implemented in Excel using the "discretely integrated condition event" (DICE) principles and structure proposed by Caro.[63] In addition, the company used the economic evaluation by Ara et al.[58] (also an individual-level model) as a starting point, which is a Health Technology Appraisal report (2012) comparing different pharmacological treatments for obesity.

An overview of the de novo model structure is shown in Figure 5.1. This Figure aims to describe the logic and assumptions underpinning the model, by depicting the process of a simulated individual's progress through the model, from model entry ("START"), through the various treatment and disease

events that may occur in the model and have consequences for patient utility and/or health and social care costs, to death and model exit (“END”).

Figure 5.1: Model structure



Source: CS Figure 25
Abbreviations: MI, myocardial infarction, NB32, naltrexone 32mg plus bupropion; ORL, orlistat; T2DM, Type 2 diabetes mellitus.
Notes: Arrows demonstrate the possible transitions to each type of event.

As depicted in Figure 5.1, the following events are considered in the economic model:

- treatment discontinuation (either based on time to treatment discontinuation (TTD) for NB32 and orlistat, or based on weight loss assessment; see Sections 5.2.4 and 5.2.6 for more details);
- development of T2DM (based on model by Ara et al.[58]);
- first cardiovascular event (either stroke or MI, based on model by Ara et al.[58]);
- second cardiovascular event (either stroke or MI, based on model by Ara et al.[58]) and;
- death (first 15 years in the model based on model by Ara et al.[58]; afterwards based on general population mortality estimates from the Office for National Statistics National Life Tables).

Upon model entry a simulated patient is assigned a profile of sampled baseline characteristics that are explanatory factors for risks, costs and/or utility in the model (sampled baseline characteristics as well as random numbers for the sampled patient are equal across all three treatments). The baseline profile characterises the individual patients by:

- age (years);
- gender (male, female);
- height (meters);
- BMI (kg/m^2);
- T2DM status (yes, no);
- smoker status (current, previous, never);
- receive insulin, if diabetic (yes, no);
- receive statins (yes, no);

In addition to the characteristics listed above, history of angina, diabetes mellitus other than T2DM and whether patients receive anti-hypertensive medication and/or aspirin were implemented in the model. However, these characteristics did not play a role, because the company assumed that no patients would have a history of angina or diabetes other than T2DM and no patients received anti-hypertensive medication and/or aspirin (see Section 5.2.3 for more details on baseline patient characteristics in the model).

If a patient experiences an event, the patient condition (or attribute as it is often called in DES terminology) for this event is updated. For example, if a patient is predicted to stop adjunct NB32 treatment before the first scheduled response assessment, a condition is used to record that the individual is no longer receiving adjunct treatment. Following updating of conditions, time to event (TTE) estimates are updated for any events affected by condition changes from the first event. For example, if an event changes BMI, times to obesity-related-disease events (for which BMI is an explanatory factor) are re-estimated.

In addition to the main modelling assumptions that are highlighted in Table 5.3 (retrieved from CS Table 53), the company's model assumed weight loss for orlistat patients at weeks 12 and 52 to be comparable to weight loss for NB32 patients at weeks 16 and 56. More specifically, the company assumed that the percentage weight loss (compared with baseline weight) for NB32 at weeks 16 and 56 can be combined with the mean difference between NB32 and orlistat (obtained from the ITC, see Sections 4.4 and 4.5 of his report for more details) to estimate the percentage weight loss (compared with baseline weight) for orlistat at weeks 12 and 52, respectively. This is similar for the proportion of responders at the week 16 and week 12 weight assessments (response criterion of ≥5% weight loss from baseline) for NB32 and orlistat, respectively. The company stated that the assumption of equivalent

weight loss at similar assessment times was also upheld within the ITC. No further justification is provided for this assumption.

Table 5.3: Main modelling assumptions utilised in the CS economic model

Assumption made	Rationale
Treatment discontinuation	
If a patient discontinues treatment with NB32 or orlistat, it is assumed that the patient is eligible to continue to receive non-pharmacological standard management (dependent on their sampled TTD).	Clinical expert consultation suggested that standard management would continue beyond cessation of adjunctive pharmacological therapy.
Weight regain	
Weight regain begins immediately after a patient discontinues all treatment (that is, adjunctive pharmacological treatment as well as standard management).	This assumption was made in the model built by Ara et al.[58] For patients who discontinue adjunctive therapy but continue to receive non-pharmacological standard management, weight regain was assumed to only commence when standard management was discontinued. Clinical expert opinion was sought to validate this assumption.
Weight is regained linearly over a 3-year period.	This assumption was made in the model built by Ara et al.[58]
The regained weight is reflective of the BMI expected as predicted by the natural history model for BMI over time.	BMI was assumed to revert to the natural history model predicted BMI given the intrinsic correlation known between age and BMI. This setting was included as a scenario analysis within the report by Ara et al.,[58] but was considered the most appropriate setting within the de novo model for incorporating BMI over time.
Obesity-related clinical events	
Within the model, it is possible for patients to experience a primary and secondary cardiovascular event (MI or stroke), as well as developing T2DM.	This assumption was made in the model built by Ara et al.[58] It is expected that the incremental clinical impact of further cardiovascular events would be negligible, as the proportion of patients who would experience more than two cardiovascular events in clinical practice is small.
Source: CS Table 53 Abbreviations: MI, myocardial infarction; T2DM, Type 2 diabetes mellitus; NB32, naltrexone 32mg plus bupropion.	

ERG comment: It is unclear to the ERG why a DES approach is preferred over for instance an individual-level state transition model. However, the ERG considered it reasonable to use the economic model by Ara et al.,[58] (comparing different pharmacological treatments for obesity) as a starting point for the current analysis. Based on their analyses, Ara et al.,[58] considered assumptions regarding weight regain to be key drivers of cost effectiveness. In this context, it should be noted that the company deviated from the assumption made by Ara et al.,[58] that patients would have regained weight to obtain their baseline BMI in three years and assumed instead that patients would have regained weight to obtain the predicted BMI in three years (using the natural history model predicting BMI over time by Ara et al.;[58] see Section 5.2.6 for more details). In response to clarification question B1a, the company responded that this was a 'logical' assumption for a simulated patient's BMI to be consistent with their characteristics.[9] However, also based on the responses to clarification question B1, it is illustrated that

this is not a conservative assumption for NB32 versus orlistat; the ICER vs. orlistat increased by £1,536 (Table 6 in the clarification response).[9] Moreover, the company did not provide justification for why their deviation from Ara et al.'s[58] assumption was 'logical' and more plausible than assuming weight regain to baseline BMI. After weight is regained to reach the baseline BMI, the BMI increases using the annual increase based on age (according to the correlation between age and BMI as reflected in the natural history model predicting BMI over time). Hence, to be consistent with Ara et al.,[58] and to be conservative, the ERG preferred to assume weight regain to the baseline BMI in its base-case. Furthermore, the linear weight regain over the time-span of three years was implemented incorrectly in the model where, in fact, the weight regain occurs instantaneously at the end of the three year period. The ERG incorporated adjustments in its base-case to reflect a linear weight regain over three year.

The ERG also questioned the (justification for the) assumption of equivalent weight loss at similar assessment times for NB32 and orlistat. The company's model assumed weight loss for orlistat patients at weeks 12 and 52 to be comparable to weight loss for NB32 patients at weeks 16 and 56. This was not justified besides stating this assumption was also upheld within the ITC (see Section 5.2.6 for more details).

The model only includes the possibility of two subsequent cardiovascular events (i.e. either two strokes, two MI's or one stroke and one MI). Implicitly assuming that the impact of the third cardiovascular event, on costs, quality of life and survival, is negligible. It can however be questioned whether having a stroke after having experienced two MIs is indeed unimportant. However, as the company argues in response to clarification question B9, this assumption is most likely conservative and hence considered reasonable by the ERG.

5.2.3 Population

NB32 is licensed as an adjunct to standard non-pharmacological management (i.e. reduced-calorie diet and increased physical activity) in adult patients who are obese (BMI ≥30kg/m2) or overweight (BMI ≥27kg/m2 and <30kg/m2) in the presence of one or more weight-related comorbidities (e.g., T2DM, dyslipidaemia, or controlled hypertension).[40] The company stated that the economic analysis aimed to reflect this patient group. Table 5.4 provides an overview of mean values for sampling baseline characteristics for individual patients in the model and used as explanatory factors for risks, costs or utility. According to the company these baseline patient characteristics were derived from a range of sources to best represent patients in UK clinical practice.

Table 5.4: Main modelling assumptions utilised in the CS economic model

Parameter	Mean value reported in CS	Mean value in economic model (calculated by ERG)	Justification	ERG value (if differently); based on section 4.2.2 (or stated if different)
Age	47 years	47 years	COR trial programme patient-level data	T2DM: 53.8 Non-T2DM: 44.7
Female	79.0%	76.7%		T2DM: 52.9% Non-T2DM: 86.7%
Height	Female: 1.64 m Male: 1.78 m	Female: 1.64 m Male: 1.78 m Total: 1.67		
Weight		Female: 90.3 kg		

Parameter	Mean value reported in CS	Mean value in economic model (calculated by ERG)	Justification	ERG value (if differently); based on section 4.2.2 (or stated if different)
	Derived from model	Male: 98.3 kg Total: 92.2 kg	Calculated by ERG based weight sampled in the model	
BMI	Derived from BMI trajectory model by Ara et al.[58] (see Section 5.3.4.3)	Female: 33.57 kg/m^2 Male: 31.05 kg/m^2 Total: 32.98 kg/m^2	Calculated by ERG based on height and weight sampled in the model	See Table 5.21 for BMI sampled in the ERG base-case
T2DM at baseline	33.2%	33.3%	Ara et al.[58]	
Insulin use for T2DM patients	33.3%	T2DM: 29.4% Total: 9.8%	Clinical opinion[64]	
Smoking status	Current: 7.0% Previous: 54.0% Never: 39.0%	Current: 5.7% Previous: 52.5% Never: 41.8%	Dare et al.[65]	Current: 10.6% Previous: 54.0% Never: 35.4%
Statin use	79.3%	80.4%	NB-CVOT study[29]	T2DM: 47.6% Non-T2DM: 10.4%
History of angina	0.0%	0.0%	Assumption – set to 0.0% as no data were identified for overweight/ obese patients	
History of diabetes other than T2DM	0.0%	0.0%		
receive anti-hypertensive medication	0.0%	0.0%	Assumption – set to 0.0% as it did cause counter-intuitive results	T2DM: 47.9% Non-T2DM: 15.0% (assuming antihypertensive medication in 77.3%[66])
receive aspirin	0.0%	0.0%		10.9% [58]

Source: CS Table 55 and economic model submitted by company
Abbreviations: BMI, body mass index; DM, diabetes mellitus; T2DM, Type 2 diabetes mellitus.

The clinical data used for NB32 and standard management during the first year in the economic evaluation are mainly retrieved from the four multicentre, randomised, double-blinded, placebo-controlled studies comprising the COR trial programme (COR-I, COR-II, COR-BMOD and COR-DM). In three of these studies (COR-I, COR-II, COR-BMOD) participants were adults with BMI 30–45kg/m2 or BMI 27–45kg/m2 and dyslipidaemia or controlled hypertension. In the fourth study (COR-DM), participants were adults with T2DM and BMI 27–45kg/m2. The company stated that, although no UK centres were included and the mean BMI of 36kg/m2 was slightly higher than usually seen in clinical trials, patient characteristics in the COR trial programme are a fair reflection of the typical patient group that would receive NB32 in UK NHS clinical practice.[64] This was based on clinical opinion (Professor

John Wilding, physician with extensive experience of treating overweightness and obesity in the NHS; JW).

Given the relatively limited follow-up period (56 weeks) of the trials in the COR trial programme and the necessity to project lifetime outcomes, the company used the NB-CVOT trial to estimate the outcomes (i.e. TTD) beyond the first year in the economic evaluation (748 patients receiving NB32 were followed beyond 52 weeks). The NB-CVOT study was a Phase IIIb, multicentre, randomised, double-blind, placebo-controlled trial to assess the occurrence of MACE in overweight or obese patients (randomising patients to receive treatment with NB32 or placebo).[29] Patients were eligible for inclusion if they were aged 45 (men) or 50 (women) years or older, had a BMI 27–50kg/m2 and a waist circumference of 88cm (women) or 102cm (men) or more. The company stated that for the NB-CVOT trial BMI inclusion criteria (BMI 27–50kg/m2 and a minimum waist circumference of 88cm (women) or 102cm (men)) were slightly different compared with the COR trial programme, patients in the NB-CVOT study were older than those in the COR trial programme (with inclusion restricted to men over 45 years and women over 50 years), and enrolment was restricted to patients with increased risk of cardiovascular outcomes.[31]

In addition to the COR trial programme and the NB-CVOT trial, the natural history model predicting BMI over time and risk equations developed by Ara et al.,[58] which predict lifetime BMI, risks for the development of key weight-related disease (i.e. stroke, MI and T2DM) and death, were used in the economic model. This was based on adult patients from the GPRD (General Practice Research Database; accessed in January 2011) who had three or more BMI readings of over 27kg/m2 (see Section 5.2.6 for more details).

ERG comment: The population aimed to reflect the scope.[7] However, patient characteristics in the model were sampled from estimates that were based on a variety of sources. It is questionable whether this is reflective of UK clinical practice. The ERG agrees with using the COR trial programme patient-level data to inform baseline age, gender and height in the model. This follows from a) that the effectiveness estimates are derived from this population and b) that the company stated, based on clinical opinion (JW), that patient characteristics in the COR trial programme are a fair reflection of the typical patient group that would receive NB32 in UK NHS clinical practice.[64] However, the other baseline characteristics can be questioned.

The ERG compared the baseline BMI sampled in the model with the baseline BMI in the COR trial programme (Table 5.5). This comparison indicates that baseline BMI is vastly underestimated in the economic model, compared to the COR trial programme and as such also compared to UK clinical practice (as clinical opinion indicated that patient characteristics in the COR trial programme are a fair reflection of the typical patient group that would receive NB32 in UK NHS clinical practice). This is also reflected in the average baseline weight of 92kg sampled in the model, while the averages ranged between 99kg and 105kg in the COR trial programme (see Section 4.2.2 for more details). Given that BMI is included as a predictive factor for utility, T2DM, cardiovascular events and death (see sections 5.2.6 and 5.2.8 for more details), the utility values and the time to these events in the model are overestimated, likely inducing bias in favour of NB32.

Other baseline characteristics are also potentially underestimated:

- Proportion current smoker of 7% (sampled 6%) while the averages ranged between 9% and 11% in the COR trial programme (excluding COR-BMOD as this trial included only non-smokers; see Section 4.2.2 for more details).

- Proportion receiving anti-hypertensive medication of 0% while the averages of hypertensive patients ranged between 15% and 63% in the COR trial programme (see Section 4.2.2 for more details). Moreover, after reviewing the time to obesity-related events, it is unclear why the company indicated that setting this to >0% would lead to counter-intuitive results.
- Proportion of patients with history of angina and/or diabetes other than T2DM of 0%. Although there were hospitalisations for unstable angina (see Section 4.2.7), the company stated that there were no data to inform these proportions.
- Proportion receiving aspirin of 0% while based on Ara et al.[58] this can be calculated to be >10%. Moreover, after reviewing the time to obesity-related events, it is unclear why the company indicated that setting this to >0% would lead to counter-intuitive results.

In addition to this, the GPRD population from Ara et al.[58] had three or more BMI readings of over 27kg/m2, but this population did not consider whether patients had one or more weight-related comorbidities while NB32 is licensed for patients with a BMI between 27-30kg/m2 only in the presence of one or more weight-related comorbidities. Hence, it is the ERG's view that both the baseline patient characteristics and the risk equations developed by Ara et al.[58] to predict lifetime BMI and risk for the development of key weight-related diseases are based on a less severe population than the licensed indication for NB32.

In contrast to the above, the proportion of patients with diabetes might have been overestimated. The value of 33.3% was obtained from Ara et al., but this was not validated against the population in the scope i.e. those with BMI ≥30kg/m2) or overweight (BMI ≥27kg/m2 and <30kg/m2) in the presence of one or more weight-related comorbidities (e.g., T2DM, dyslipidaemia, or controlled hypertension). According to Health Survey England data for 2013, the percentage is between 14 and 15% depending on sex.[67]

Also, the proportion of patients receiving statins might have been overestimated as this is 8% to 13% in the COR-I, COR-II and COR-BMOD trials and 46% to 49% for diabetic patients in the COR-DM trial (see Section 4.2.2) while it was 79% (sampled 80%) in the economic model independent of T2DM status. Related to this, correlations between covariates were not incorporated in the sampling of the patient characteristics, leading to counter-intuitive patient profiles. For instance, based on the patient characteristics table of the COR-I, COR-II, COR-BMOD and COR-DM trials (in Section 4.2.2), it becomes clear that the patients without T2DM (COR-I, COR-II and COR-BMOD trials) have different patient characteristics (e.g. regarding age, sex, hypertension status and statin use) than patients with T2DM (COR-DM trial). This is neglected in the sampling of the patient population. To address these issues, the ERG adjusted the baseline characteristics used in the model (Table 5.4). This included calibrating the natural history model to predict BMI over time (see Section 5.3 for more details).

The company assumed no patients had a history of angina and/or diabetes other than T2DM. This assumption was made as no data were identified on these characteristics for overweight/obese patients. The ERG agrees with this statement and would therefore argue that it can be questioned whether the results of the economic analyses are representative for patients with a history of angina and/or diabetes other than T2DM.

Table 5.5: Distribution of BMI (patients sampled in the model and across the COR trial programme)

Obesity class	Model			COR-I		COR-II		COR-BMOD		COR-DM		NB-CVOT	
	Overall	Female	Male	Placebo	NB32	Placebo	NB32	Placebo	NB32	Placebo	NB32	Placebo	NB32
BMI<30kg/m2	8.0%	3.3%	23.6%	0.9%	3.1%[a]	2.8%	2.5%	0.5%	1.4%	6.5%	5.4%	7.0%	6.7%
BMI ≥30 and ≤35 kg/m2	74.6%	74.2%	76.0%	37.3%	38.4%	37.6%	39.8%	31.7%	35.0%	28.8%	33.1%	31.6%	31.3%
BMI ≥35 and <40 kg/m2	17.1%	22.2%	0.4%	39.4%	35.0%	38.6%	31.6%	39.1%	38.9%	37.6%	32.8%	31.1%[a]	33.2%[a]
BMI ≥ 40 kg/m2	0.3%	0.4%	0.0%	22.4%	23.5%	21.0%	26.2%	28.7%	24.7%	27.1%	28.7%	30.3%	28.8%

Source: Economic model submitted by the company and response to clarification question A17

[a]Original value in response to clarification question A17 contained incorrect proportions, this is corrected by the ERG.

5.2.4 Interventions and comparators

In line with the final scope and licensed indications, the company considered the following comparators for NB32 as an adjunct to standard management:

- orlistat as an adjunct to standard management and;
- standard management alone

NB32 is implemented as per its EMA Summary of Product Characteristics (SmPC) posology and method of administration, incorporating a four week escalation period, after which the maximum recommended daily dose of 32mg naltrexone hydrochloride and 360mg bupropion hydrochloride is assumed.[40] Orlistat is similarly implemented as per its EMA SmPC posology and method of administration, a 360mg daily dose.[39]

The company specified standard management as implemented in the analysis to reflect the non-pharmaceutical dietary and lifestyle management treatment received in UK NHS practice (see Section 5.2.9 for more details). The company stated based on clinical opinion (JW) that although standard management varies by geography, the non-pharmaceutical treatment administered in the COR-I and COR-II studies is a good reflection of the treatment patients are likely to receive in NHS England.[64] According to the NB32 license, standard management includes a reduced-calorie diet and increased physical activity.[40]

Stopping rules for both NB32 and orlistat are implemented in the model, as per their license terms:[39]

- NB32: patients who fail to meet the response criterion of ≥5% weight loss from baseline after 16 weeks after treatment initiation (12 weeks post-escalation period) discontinue pharmacological treatment.
- orlistat: patients who fail to meet the response criterion of ≥5% weight loss from baseline after 12 weeks after treatment initiation, discontinue pharmacological treatment.

Based on clinical opinion (JW),[64] the same stopping rule was applied 56 and 52 weeks after treatment initiation for NB32 and orlistat, respectively. It should be noted that these stopping rules only apply to pharmacological treatment (not necessarily to standard management that is provided in addition to NB32/orlistat), see Section 5.2.6 for more details regarding TTD.

ERG comment: The ERG considered whether, given that it is not required according to the license terms,[39, 40] the stopping rule at the secondary assessment, i.e. at 56 and 52 weeks after treatment initiation for NB32 and orlistat, would be reflective of clinical practice. The ERG found that in NICE clinical guideline 189 regarding obesity,[5] it is recommended (Section 1.9.9) that there will be a discussion regarding drug treatment longer than 12 months after discussing potential benefits and limitations. It should however be noted that this recommendation does not consider an objective response criterion such as the ≥5% weight loss from baseline used in the model.

One major limitation of the model is the inability to incorporate re-treatment, behaviour modification treatment (e.g. a weight loss programme) and or bariatric surgery (for which patients become eligible over time once their BMI is/increases to >40kg/m2[68] in the model). The company stated (in response to clarification question B3) that re-treatment is clinically plausible. However, the company did not incorporate this justified by a stated lack of data to inform re-treatment in the model.

5.2.5 Perspective, time horizon and discounting

The analysis was conducted from the perspective of the payer, i.e. the NHS England and Wales, over a lifetime horizon. Costs and outcomes were discounted by 3.5%.

ERG comment: This is in line with the NICE reference case.

5.2.6 Treatment effectiveness and extrapolation

i) Treatment effectiveness and extrapolation overview

In the CS,[1] clinical parameters and variables are reported as falling into the following four categories:

- Baseline patient characteristics,
- Treatment duration,
- Treatment effectiveness,
- Epidemiological models of natural history.

In this report, the baseline patient characteristics were presented and discussed in Section 5.2.3. Time to treatment discontinuation is discussed in Section 5.2.6 ii). Treatment effectiveness is discussed under the headings iii) Proportion of patients with weight loss ≥ 5% and iv) Mean change in body weight. Finally, obesity-related events and epidemiological BMI models are discussed in Section v) Risk of obesity-related events and natural history of BMI.

Treatment effectiveness estimates (i.e. time to treatment discontinuation data, proportion of responders, and mean change in body weight) were derived from the COR trial programme, including the COR-I, COR-II, COR-BMOD and COR-DM trials. All the analyses were based on the company's mITT analysis, which reflects only those patients who have a post-baseline measurement whilst on the study drug.

ERG comment:

The ERG questions the company's approach 1) using the mITT analysis for estimating the proportion of responders and mean change in weight and; 2) pooling across all COR studies for estimating the time to treatment discontinuation, proportion of responders and mean change in weight.

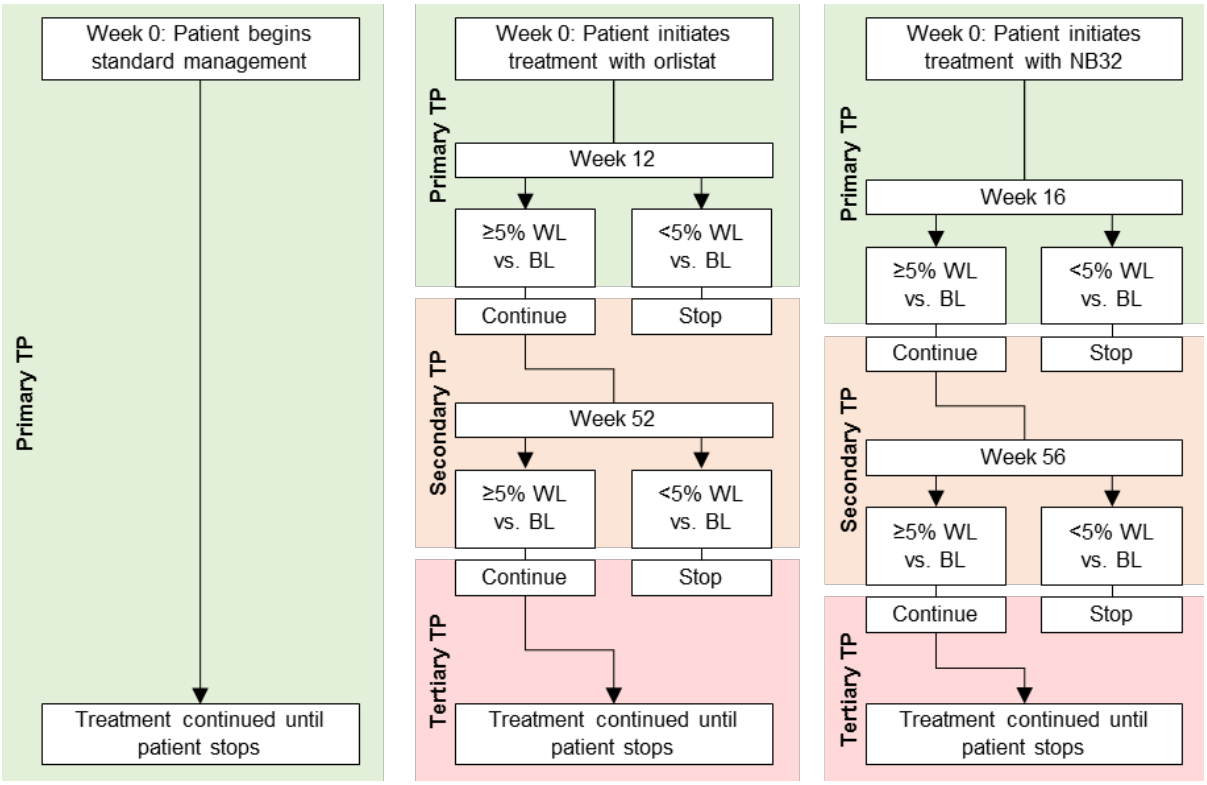
The ERG considers that the use of the ITT population would have been both more appropriate and more conservative. The mITT analysis includes only those patients who have a baseline and at least one post-baseline measurement whilst on the study drug. Patients who discontinued without providing follow-up weight assessments were excluded. Reasons for discontinuation could relate to efficacy or safety (i.e. AEs) of the drug. Using the true ITT data, NB32 would achieve a smaller mean percentage of weight loss and smaller proportion of responders compared to the mITT data. This is discussed in more detail in Section 4.2. Following the ERG's request for scenario analyses using data on clinical effectiveness and treatment discontinuation derived from the ITT population (Question B6), the company refused to carry out these analyses, stating that these were "*irrelevant to de novo model, due to the nature in which weight loss outcomes are derived*".[9] Whilst the company justified this by clarifying that the safety population, not the mITT population, was used to estimate time to treatment discontinuation (TTD) up to week 16, no further justification was provided for not presenting the scenario analyses using ITT estimates for proportion of responders and mean change in weight. The issue of using this population and the bias that it introduces are discussed further in Sections 5.2.6 iii) and iv).

It is the ERG's view that it was inappropriate to pool from all COR studies, including COR-BMOD and COR-II. In the CS, it is stated that, according to the company's criteria, the COR-BMOD study considered 'intensive' behaviour modification. This was based on *"the number of follow-up appointments with a medical/dietary professional; detail and severity regarding the prescription of dietary recommendations; and the level of physical activity participants were encouraged to follow"*.[1] In the response to clarification question B8[9], the company states that based on clinical expert opinion, the effects of intense behavioural modification and pharmacological treatment would be expected to be additive. This might suggest that the difference between pharmacologic treatments would remain the same irrespective of the intensity of non-pharmacological treatment. However, effectiveness estimates derived from COR-BMOD, where NB32 was administered in combination with intensive behavioural modification were substantially different when compared to effectiveness estimates derived from the studies in which NB32 was administered together with standard management only. Pooling clinical effectiveness data from all COR trials, including the COR-BMOD study, is therefore inappropriate. The ERG notes that if effectiveness estimates included intense behavioural modification, then this should also be reflected in the cost. In the absence of cost estimates, the ERG was unable to perform analysis including intense behavioural modification. Likewise, the ERG considers the use of COR-II for the derivation of treatment effectiveness beyond 28 weeks as inappropriate because NB32 participants with <5% weight loss at visits between weeks 28 and 44 were re-randomised. The ERG therefore considers that NB32 treatment effectiveness estimates, assuming no concomitant behaviour modification (e.g. weight loss programme), should only be derived from the COR-I and COR-DM trials. The issue of pooling from all COR studies and the bias that it introduces are discussed further in Sections 5.2.6 ii)-iv).

ii) Time to treatment discontinuation

Time to treatment discontinuation was estimated separately for patients receiving standard management, NB32 and orlistat. For patients receiving standard management (alone or in combination with adjunctive pharmacological therapy), treatment is given from week 0 until the patient stops treatment. For patients receiving pharmacological therapy, treatment duration is considered in three phases: phase 1 includes the time to primary assessment (conducted at week 16 for NB32 and at week 12 for orlistat); phase 2 is the time from primary to secondary assessment (which is conducted at week 56 for NB32 and at week 52 for orlistat); and phase 3 covers the time after the secondary assessment. It is stated within the CS that *"patients must cease to receive adjunctive therapy ahead of discontinuing standard management, after which they may either immediately discontinue standard management or continue to receive standard management alone"*.[1] Kaplan–Meier (KM) data and the proportion of responders are used to inform the duration of adjunctive therapy and standard management within the model. The expected pathway of care is illustrated in Figure 26 of the CS,[1] printed here in Figure 5.2.

Figure 5.2: Expected pathway of care across all treatment arms



Key: BL, baseline; NB32, naltrexone 32mg plus bupropion; TP, treatment phase; WL, weight loss.

Phase 1 (from treatment initiation to primary assessment):

For both NB32 and orlistat patients, treatment duration in phase 1 was based on KM estimates from NB32 treatment discontinuation data in the COR trial programme (illustrated in Figures 5.3 and 5.4). These data were also used for orlistat because *"there were no comparable duration of treatment data available to inform discontinuation ahead of primary assessment…"*.[1] However, because phase 1 was shorter for orlistat than for NB32 (12 weeks instead of 16 weeks), the KM data for NB32 patients were linearly scaled to fit the 12-week period to primary response assessment for orlistat. For NB32 patients, 67.2% continued treatment until 16 weeks. As a result of the linear scaling, this same proportion was also used for orlistat at 12 weeks.

Figure 5.3: NB32 adjunct therapy discontinuation from treatment initiation to 16 weeks (pooled COR trial programme data, all NB32 patients)

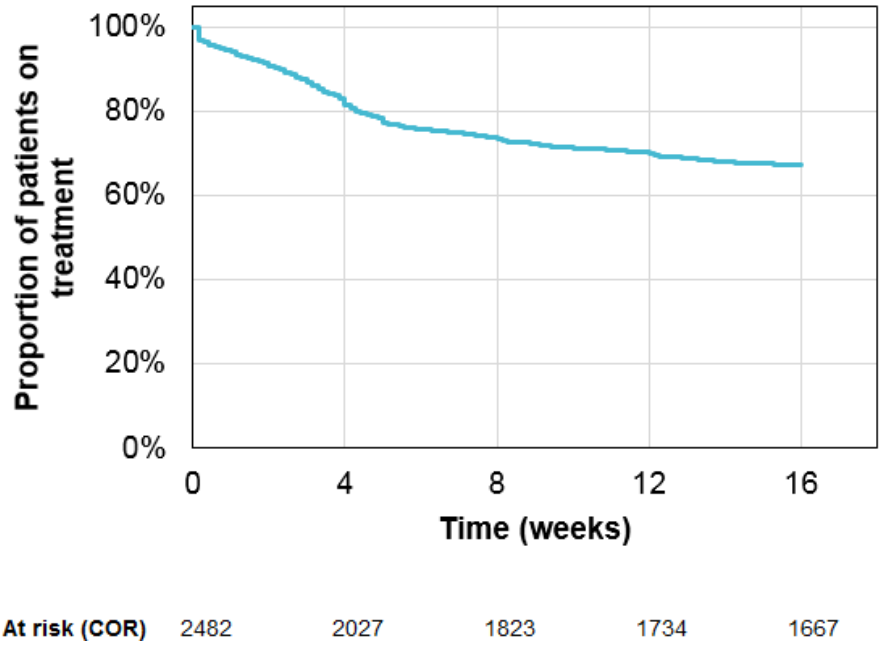
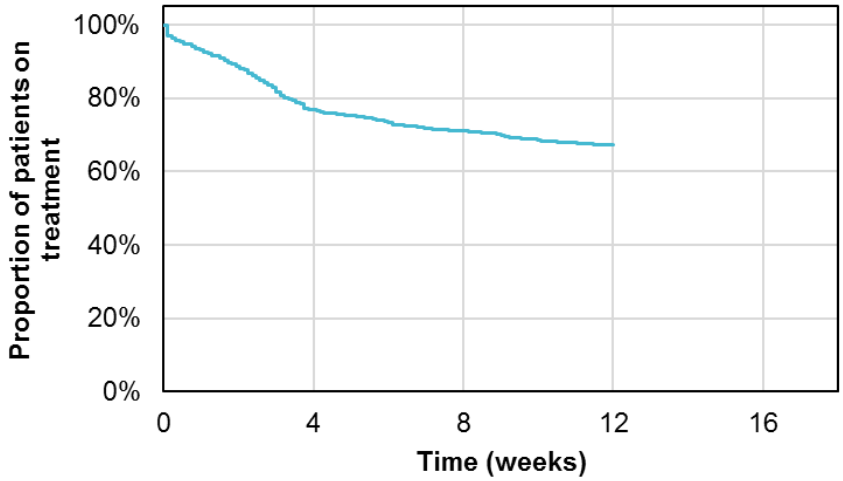


Figure 5.4: Orlistat adjunct therapy discontinuation from treatment initiation to 12 weeks (from pooled COR trial programme data, all NB32 patients)



Phase 2 (from primary assessment to secondary assessment):

For both NB32 and orlistat patients, treatment duration in phase 2 was based on KM estimates from NB32 treatment discontinuation data in the COR trial programme, with only those patients included in the analysis that had achieved response at their primary assessment date (i.e. a weight loss of at least 5% compared with baseline). Because treatment discontinuation data were not available for orlistat, the same NB32 treatment discontinuation KM data were used for orlistat, but shifted by four weeks to match the shifted time from primary to secondary assessment (12 to 52 weeks instead of 16 to 56 weeks). For NB32 patients, 86.1% of responding patients at week 16 continued treatment until 56 weeks. For orlistat patients, the same proportion of responding patients continued treatment until 52 weeks. This is illustrated in Figures 5.5 and 5.6 below.

Figure 5.5: NB32 adjunct therapy discontinuation from 16 to 56 weeks (from pooled COR trial programme data; NB32 16-week responders)

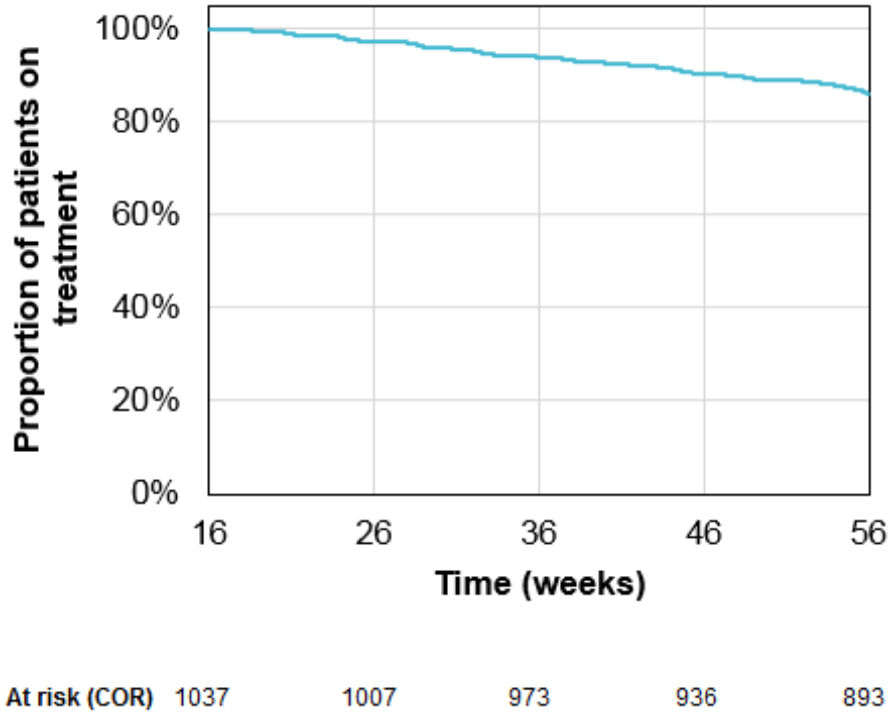
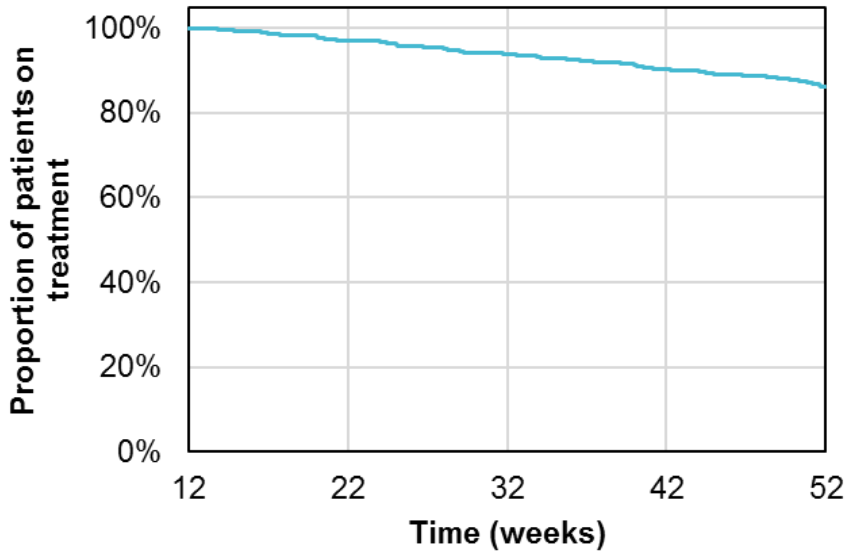


Figure 5.6: Orlistat adjunct therapy discontinuation from 12 to 52 weeks (from pooled COR trial programme data; NB32 16-week responders)



Phase 3 (from secondary assessment onwards):

For both NB32 and orlistat patients, treatment duration in phase 3 was based on KM estimates from NB32 treatment discontinuation data in the NB-CVOT study for the time period from 56 weeks to 158 weeks (end of study period). All patients were assumed to have discontinued after treatment duration data were unavailable (see Figure 5.7 and 5.8 below).

Figure 5.7: NB32 adjunct therapy discontinuation from 56 weeks (from NB-CVOT study data; NB32 56-week responders)

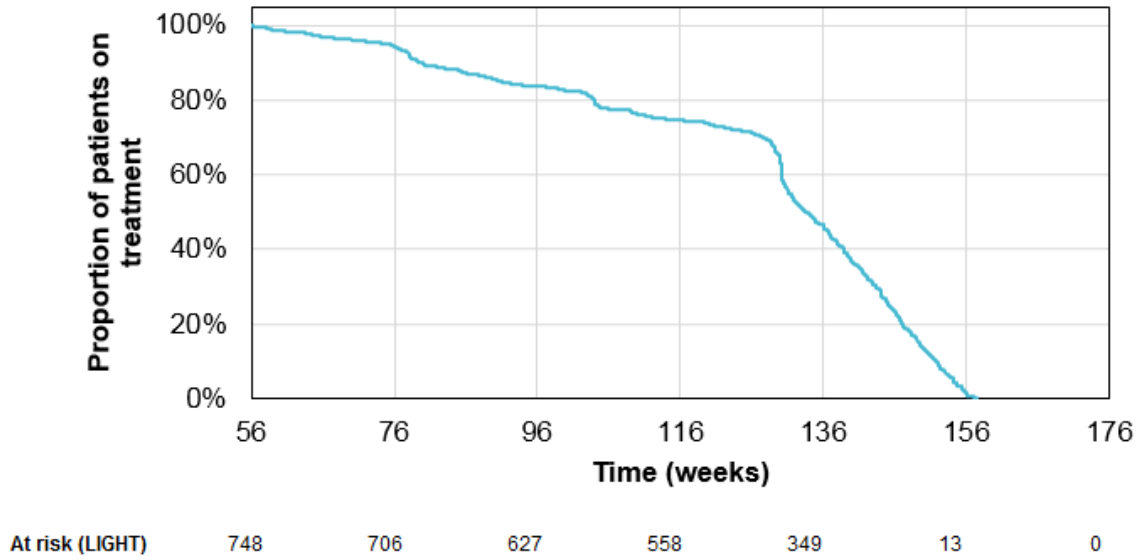
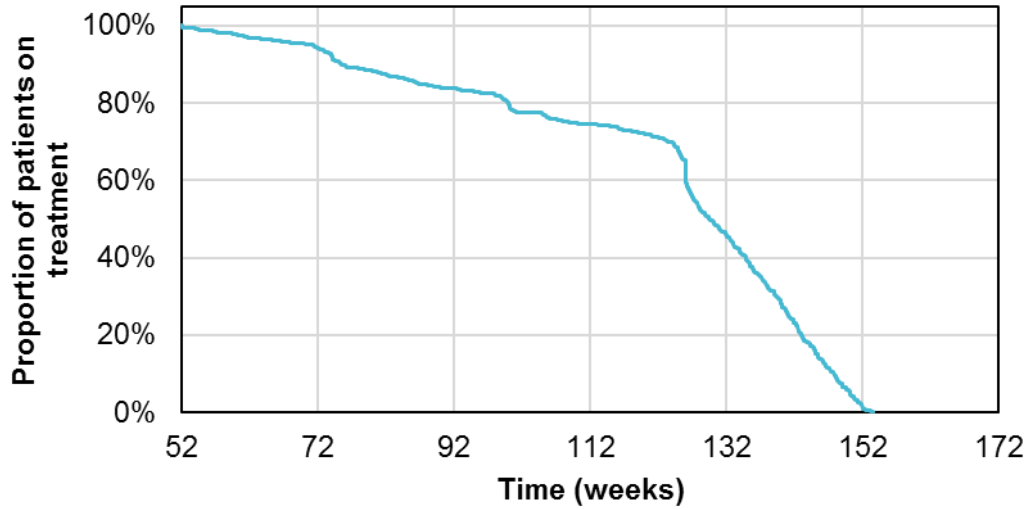


Figure 5.8: Orlistat adjunct therapy discontinuation from 56 weeks (from NB-CVOT study data; orlistat 52-week responders)

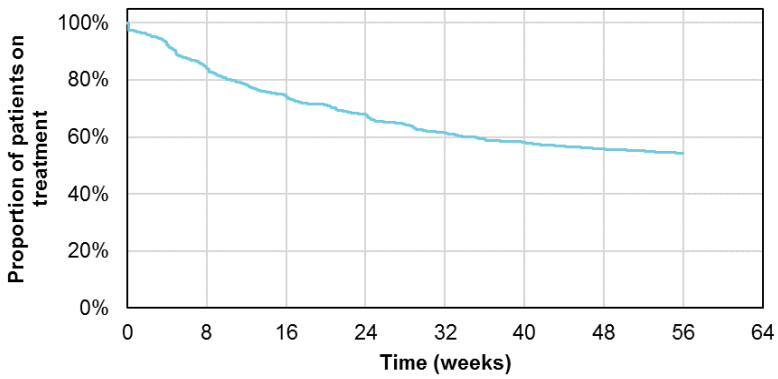


Treatment duration estimation for standard management:

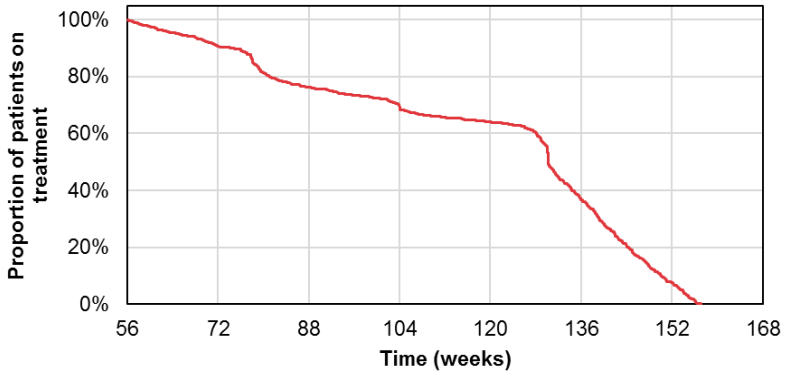
Patients receiving standard management are not subject to the same response-based treatment stopping rules as those receiving adjunct pharmacological treatment. Therefore all patient-level data from the COR trial programme could be used to inform TTD estimates in the first 56 weeks after treatment initiation. Treatment duration for standard management was then estimated using the available data from the COR trial programme up to 56 weeks and then joining the KM data from NB-CVOT to KM data from the COR trial programme by scaling the curve according to the proportion of patients who were still receiving standard management treatment at week 56 (see Figure 5.9 below).

Figure 5.9: Derivation of duration of standard management treatment

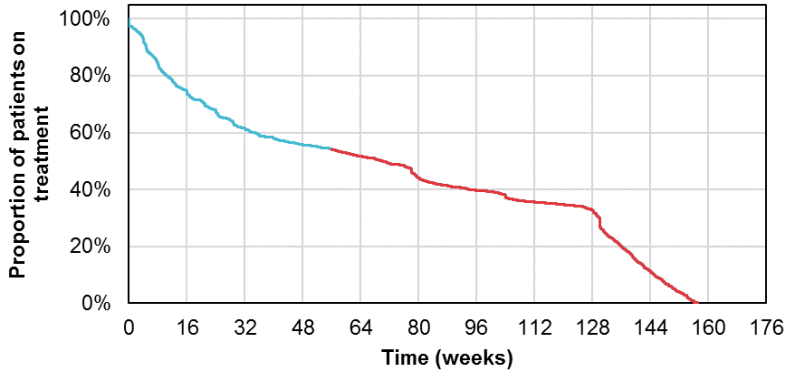
Kaplan–Meier data from the COR trial programme (Week 0 to Week 56)



Kaplan–Meier data from the NB-CVOT study (Week 56 onwards)



Combined Kaplan–Meier data from both studies (Week 0 onwards)



ERG comment: The ERG considers the use of the safety population for TTD as reasonable but believes that the TTD is underestimated in the model, in particular for orlistat.

Based on the CS,[1] it was unclear which population was used to estimate TTD. The company clarified in their response to Question A19 of the clarification letter[9] that the safety population was used to estimate TTD up to week 16. In the CS (Section 4.4), it is stated that the safety population "*included all randomised patients who were administered at least one tablet of study treatment and had at least one investigator contact/assessment at any time after the start of study treatment, regardless of whether they discontinue the study*".[1] The ERG wishes to highlight that the ITT definition commonly used in orlistat trials is closer to this definition of the safety population in COR trials than to the company's mITT population used for the COR trial programme.[42-44] The company argued in their response to clarification question A19.b that the KM function of ITT and safety populations would be the same except for the number of patients at risk at time 0. This difference would stem from untreated patients who would automatically be censored at time 0. According to the company, this would make the two KM functions equivalent. The ERG was satisfied that this was reasonable.

It is the ERG's view that TTD should not have been pooled from the four studies in the COR trial programme. The ERG considers it to be plausible that treatment duration is different in patients who also receive intense behavioural modification. Furthermore, TTD may be different in patients with or without T2DM. The ERG therefore considers that modelling TTD separately for both subgroups (T2DM and non-T2DM) may have been more appropriate.

The patient output from the company's model run revealed a mean TTD of 13.32 months, 12.29 months and 17.16 months for NB32, orlistat and for SM respectively. The ERG thinks that these may be under-estimates because:

(1) TTD estimates for the period after the one year assessment were derived from the NB-CVOT study in which patients had characteristics associated with an increased risk of CV outcomes, potentially leading to a shorter TTD. This is acknowledged by the company in the CS, in which it is stated that TTD is likely to be under-estimated by these data, given the age and comorbidity profile of NB-CVOT study patients[1].
(2) The end of the NB-CVOT study was used as the maximum TTD, whether patients in that study had discontinued or not.
(3) The company claims that the most reasonable and conservative assumption was to assume that TTD for orlistat would follow a similar trajectory to NB32, given that patient-level data for orlistat were unavailable. However, the ERG found publications reporting TTD for orlistat, which reveal that orlistat TTD was longer than the 12.29 months estimated by the model, with many studies reporting that the proportion of patients still receiving orlistat at 12 months was >50%.[42-44] However, these TTD estimates were not conditional on response to treatment (primary and secondary response assessments) and therefore have to be interpreted with caution, but reported response rates in two of these studies suggest that a significant proportion would still have continued treatment based on their response (45.7% response rate as measured by patients achieving >5% weight loss in Bakris et al.[42], 55.6% in Broom et al.[44]; Berne et al.[43] did not report response rates with the same level of weight loss). It is the ERG's view that the company should have validated their assumption for orlistat with these data. Furthermore, the company claimed that TTD may be shorter with orlistat than with NB32, given the known toxicity profile and association with treatment discontinuation in Question B2 of the clarification response.[9] This is, however, not supported by any evidence. It is the ERG's view

that TTD for NB32 and orlistat may be under-estimated. The ERG wishes to highlight that the under-estimation of TTD leads to an under-estimation of costs.

(4) For the derivation of the orlistat TTD, the KM estimates for NB32 TTD for the first 16 weeks were linearly scaled to fit the first 12 weeks of orlistat treatment. This was justified by the different time to primary assessment, and the fact that for NB32, the first four weeks include a titration period. The ERG believes that this linear scaling may further under-estimate orlistat TTD, resulting in worse effectiveness (patients will stop losing weight and start weight regain sooner), but also in decreased costs associated with orlistat and the effect of this is therefore ambiguous. The ERG therefore removed the linear scaling in its base-case analysis. The ERG furthermore considers there to be considerable uncertainty surrounding the TTD of orlistat estimation.

iii) Proportion of patients with weight loss ≥ 5%

The proportion of patients with weight loss ≥ 5% at primary response assessment (conditional on being on treatment) was obtained for NB32 by dividing the proportion of patients (50.8%) still on treatment by the total proportion of patients (65.2%) that had achieved a ≥ 5% weight loss at primary response assessment in the COR trial programme (COR-I, COR-II, COR-BMOD, COR-DM). The proportion of patients continuing treatment after this assessment, was thus estimated to be 75.7% of those still on treatment.

For orlistat, the proportion of patients with weight loss ≥ 5% at primary response assessment was not available. The company therefore used the relative effectiveness estimate for proportion of responders at secondary response assessment at one year (which was available as an odds ratio derived from the ITC) to obtain the proportion of responders at primary assessment. This yielded a proportion of 77.9% for the T2DM and 70.5% for non-T2DM groups, respectively.

At secondary response assessment, mean change in body weight estimated in the model determines the proportion of responders and non-responders.

ERG comment:

The ERG notes that there was a discrepancy between the mean OR for the proportion of responders of orlistat compared with NB32 used in the model and the one reported in the CS on page 19 and in a forest plot shown in Figure 19 of the CS.[1] In the CS, a mean OR of 1.09 is reported, whilst the model uses a mean OR of 1.13, which is based on the coda sample. It is important to note that both of these values would mean that a greater proportion of patients would achieve weight loss of ≥5% with orlistat compared to NB32 at the one year assessment. The company, however, notes that that difference based on the mean OR of 1.09 was not statistically significant. It was unclear whether a mistake was made in the report or within the model (the coda sample used) and the ERG was therefore unable to address the discrepancy. The ERG, however, notes that, if the mistake was in the model, then this would have likely caused a slight upwards bias to the ICER comparing NB32 with orlistat. Furthermore, this discrepancy is addressed in the ERG's base-case analysis where the ITT data and therefore a newly calculated OR is used.

The ERG's concerns about the derivation of proportion of responders for NB32 and comparators are presented in the following paragraphs for NB32 (1) and comparators (2):

(1) As was stated above in Section 5.2.6 i), it was inappropriate to pool the proportion of responders to NB32 treatment from all COR studies, including BMOD. By doing so, the proportion of responders to NB32 is over-estimated. This is supported by response rates for treatment with NB32 versus placebo as

presented in Table 4.8. As a result, it is the ERG's view that response rates to NB32 are likely to be over-estimated as a consequence of the pooling method.

Furthermore, the use of mITT data for the derivation of response rates would bias the estimates in favour of NB32. This is shown in Table 4.10 of the clinical effectiveness section, which shows that a smaller proportion of patients achieve a response when the ITT population is used, compared with the mITT population. The company was asked to provide an analysis using ITT populations but failed to do so.

(2) The ERG considers that the application of the base-case odds ratio derived from the ITC is also inappropriate because this was derived from all four COR studies, including the COR-BMOD and COR-II studies. The more appropriate estimation of both NB32 and orlistat rate of responders would be to use the rate of responders as pooled from COR-I and COR-DM and then apply the odds ratio derived from sensitivity analysis 3 in the ITC, which excludes the studies in which pharmacological treatment is combined with more intensive behavioural modification. The ERG also wishes to highlight that the estimation of the orlistat response rate at primary assessment was made based on the assumption that the one year odds ratio between orlistat and NB32 equally applies to the 12/16 week setting. The ERG was satisfied that, in the absence of other data, this was a reasonable assumption.

iv) Mean change in body weight

For NB32, mean change in body weight was estimated separately for responders and non-responders at the primary response assessment (16 weeks) and derived from the COR trial programme (COR-I, COR-II, COR-BMOD, COR-DM).

For orlistat, mean change in body weight compared with NB32 was derived from the ITC, assuming that weight loss at 16 weeks in NB32 patients was comparable with weight loss at 12 weeks in orlistat patients. This assumption was justified in the CS by the lack of a four week titration period for patients treated with orlistat. Moreover, due to lack of weight loss data for orlistat at 12 weeks and due to it not being possible to stratify weight loss by response status in the ITC, the relative weight loss (as in the mean difference in weight loss) of orlistat compared with NB32 at the one year assessment was used to estimate weight loss associated with orlistat treatment at 12 weeks (primary response assessment) for both responders and non-responders. For standard management patients, weight loss estimates were derived from the COR trial programme patient-level data and not stratified by response. For both orlistat and standard management patients, weight loss estimates were stratified by T2DM status, but for NB32 this was not done.

The average weight loss data used in the model are summarised in Table 5.6.

Table 5.6: Average weight loss at primary response assessment

Treatment	Outcome	Value	Source
NB32	Primary Week 16 assessment: Responders	9.4%	COR trial programme data
	Primary Week 16 assessment: Non-responders	1.9%	
ORL	Primary Week 12 assessment: Responders (all patients)	8.6%[a]	ITC[a]
	Primary Week 12 assessment: Responders (T2DM patients)	9.2%	ITC
	Primary week 12 assessment: Responders (non-T2DM patients)	8.3%	
	Primary Week 12 assessment: Non-responders (all patients)	1.1%[a]	ITC[a]

Treatment	Outcome	Value	Source
	Primary Week 12 assessment: Non-responders (T2DM patients)	1.7%	ITC
	Primary Week 12 assessment: Non-responders (non-T2DM patients)	0.8%	
SM	Week 12: All patients	2.3%	COR trial programme data
	Week 16: All patients	2.7%	
ITC, indirect treatment comparison; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; SE, standard error; SM, standard management; T2DM, Type 2 diabetes mellitus. Notes: [a], The derived proportion shown here is an estimate based upon the proportion of T2DM patients at baseline.			

At secondary response assessment, weight loss for NB32 patients was calculated as 11.7% for those who responded at primary response assessment and 8.8% for all patients combined. For orlistat, the 11.7% weight loss for NB32 patients was used together with the mean difference in weight loss of orlistat compared to NB32 derived from the ITC to estimate weight loss for orlistat responders at 52 weeks (weight loss for NB32 patients at week 56 was again assumed to be comparable to that of orlistat patients at week 52 given the lack of a four week titration period for orlistat). For standard management, weight loss was estimated based on the weight loss calculated for all NB32 patients in the COR trial programme regardless of response status (of 8.8%, as stated above) and the mean difference in weight loss based on the ITC, stratified by T2DM status. The average weight loss figures for the secondary response assessment are presented in Table 5.7.

Table 5.7: Average weight loss at secondary response assessment

Treatment	Outcome	Value	Source
NB32	Secondary Week 56 assessment	11.7%	COR trial programme data
ORL	Secondary Week 52 assessment (all patients)	10.9%	ITC[a]
	Secondary Week 52 assessment (T2DM patients)	11.5%	ITC
	Secondary Week 52 assessment (non-T2DM patients)	10.6%	
SM	Week 52/56 (all patients)	4.5%	ITC[a]
	Week 52/56 (T2DM patients)	5.6%	ITC
	Week 52/56 (non-T2DM patients)	3.9%	
ITC, indirect treatment comparison; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; SE, standard error; SM, standard management; T2DM, Type 2 diabetes mellitus. Notes: [a], The derived proportion shown here is an estimate based upon the proportion of T2DM patients at baseline.			

ERG comment: As was stated above in Sections 5.2.6 i) and iii), it was also inappropriate to pool the mean change in weight from all COR studies, including BMOD. By doing so, the proportion of responders to NB32 is over-estimated (see Table 4.8). Similarly, mean weight loss figures in the COR-BMOD study are larger in both the NB32 and placebo arms than in the other studies. As a result, it is the ERG's view that mean change in weight for patients treated with NB32 is likely to be over-estimated as a consequence of the pooling method (Table 4.8).

The use of mITT data for the derivation of mean change in weight introduces further bias in favour of NB32. This is shown in Table 4.9 of the clinical effectiveness section, which shows that a smaller mean change in weight is achieved when the ITT population is used, compared with the mITT population. The company was asked to provide an analysis using ITT populations but failed to do so, without providing adequate justification.

The ERG calls into question the assumption of weeks 12 and 52 on treatment with orlistat being comparable to weeks 16 and 56 for patients treated with NB32. The justification provided by the company was that the first four weeks of treatment with NB32 were a titration period. However, patients do lose weight even during this titration period, as shown in Figure 7 in the CS[1], where patients lose most weight in the first four weeks, followed by a phase where weight loss slows down. The ERG therefore considers this assumption as inappropriately justified. However, based on the same figure, patients may not lose significantly more weight in the last four weeks of a one year treatment period. Therefore the assumption of equivalence appears more valid for the one year assessment than for the primary response assessment.

The ERG wishes to highlight a further limitation in the company's analysis of weight loss for patients treated with orlistat at 12 weeks. This was derived by using the mean difference in weight loss of orlistat compared with NB32 as derived from the ITC at the one year assessment. The mean difference is an absolute measure, which would presumably vary according to the magnitude of weight loss achieved. With absolute weight loss being smaller at the primary than at the secondary response assessment at one year, applying the absolute mean difference derived from one year to NB32 weight loss data will result in an under-estimation of weight loss for patients treated with orlistat. The ERG therefore adjusted this in its base-case.

v) Risk of obesity-related events and natural history of BMI

In the CS, the risk of occurrence of obesity-related events is modelled conditional on changes in BMI, whereby BMI levels are assumed to change with age, based on a study by Ara et al.[58]. This is achieved in two parts:

1. Through the use of risk equations to estimate the time to stroke, MI, onset of T2DM and death.
2. Through a natural history model of BMI over time to estimate patients' BMI throughout their lifetime.

In the choice of the risk equations and BMI natural history model, the company heavily relied on the previously published HTA report by Ara et al.[58] The company states that the study by Ara et al.[58] identified limitations with existing studies of the relationship between the development of cardiovascular disease and weight. Those studies comprised cross-sectional studies identifying correlations between major clinical events and BMI, studies that categorised BMI and were therefore unable to capture changes within categories and other existing studies primarily conducted outside the UK. Ara et al.[58] therefore used large-scale GPRD data to estimate the risk of major cardiovascular events occurring at specific levels of BMI and age, controlling for confounding factors.

To establish both the risk equations for major clinical events and the natural history model of BMI, Ara et al.[58] drew 100,000 patients with three or more BMI readings over 27kg/m^2 from the GPRD database. For 1. the risk equations for obesity-related events, occurrence of all-cause mortality, MI, stroke and T2DM onset was identified for each patient. Separate patient cohorts were created for each outcome because complete patient data were not available. Except for the T2DM cohort, each of these cohorts were then subdivided into diabetic and non-diabetic cohorts (only including patients who were diabetic

or non-diabetic for the entire follow-up), resulting in seven cohorts for which TTE were estimated. Covariates included in the Ara et al.[58] model include baseline age; sex; use of aspirin, statins, blood-pressure lowering treatment; and smoking status. Diabetic cohorts also included a covariate dummy for insulin use. Only baseline BMI was used in TTE analysis. Weibull models were fitted to estimate TTE, with the Weibull scale parameter depending on each of the covariates, irrespective of statistical significance, and higher-order polynomial terms of BMI and age, based on significance at the 5% level. The Weibull shape parameter was only allowed to depend on a subset of prepared covariates, based on significance at the 5% level.

The company's model uses these TTE cohorts to inform the major clinical event estimates. However, general population data are used to inform all-cause mortality because only less than 10% of patients in the diabetic all-cause mortality cohort had follow-up beyond this point. Beyond 15 years, Weibull TTE estimates for MI, stroke and T2DM onset are applied over the company's model time horizon for obesity-related non-fatal events because it was not clear to the company what alternative assumptions were used in the models by Ara et al.[58] and data from the GPRD cohort were sparse.

To inform 2. the BMI trajectory of patients throughout their lives, Ara et al.[58] used multilevel modelling of the repeated measures of BMI in the GPRD cohort, with age as the timescale. Patients with BMI below 25kg/m^2 at any time were excluded from the analysis. Exploratory trajectory plots from random patients were used to inform model specification, before applying multilevel models. The model was adjusted for age and sex and the interaction between age and sex.

ERG comment: The ERG considered it appropriate to use risk equations for obesity-related events and the natural history model of BMI as reported in Ara et al.[58] The ERG is, however, concerned that the estimation of obesity-related events is based on a patient population that has a lower BMI (based on the Ara et al. BMI natural history model) than that of the population represented on the COR trial programme.

5.2.7 Adverse events

The company states that it was unable to make trial data comparisons between AE associated with NB32 and orlistat because details from clinical literature and regulatory documents on orlistat were insufficient. The company quotes the opinion of one clinical expert as *"NB32 patients have a HRQoL benefit over orlistat patients as a result of AE differences"*.[1] The company also refers to AEs in the lower digestive tract that can be particularly unpleasant for patients, referring to their own data on file. Lastly, the company claims that, *"while no NB32-related deaths were observed across the COR trial programme and the NB-CVOT and IGNITE studies, and orlistat mortality risk from increased liver reaction risk cannot be ruled out based on clinical study data"*.[1] As a result of quality of life implications of AE being poorly understood (especially in relation to whether the incidence of some AEs is treatment- or condition-related), the company only considers costs of AEs. The company adds that not accounting for HRQoL impact of AEs is conservative: 1) in comparison to orlistat, considering the relative AE profiles (for which no evidence was provided); and, 2) in comparison to standard management, for which the company claims that although the AE profile associated with NB32 together with standard management is less good than that of standard management alone, the treatment effectiveness HRQoL benefits outweigh any negative NB32-related AE effects on HRQoL. Furthermore, the company notes that in clinical trials and practice, treatment-related AEs are generally quickly resolved, with only short-term effects upon HRQoL.

The AE data used in the model are derived from the COR trial programme AE incidence data.

ERG comment: The ERG considers the company's claim that not accounting for HRQoL impact of AEs in the economic model is conservative as highly questionable. No systematic overview of evidence was provided that showed that the AE profile of orlistat was indeed worse than that of NB32. With regards to the comparative AE profile of NB32 vs. SM, it is clearly stated in the CS that the NB32 AE profile shows a higher incidence of AEs in the gastro-intestinal tract and nausea than that of SM.[1] The ERG does not consider the company's argument that these need not be reflected because treatment benefits outweigh any negative NB32-related AE effects on HRQoL a valid argument because the HRQoL effects of NB32 are captured in the model and AE effects are not. The ERG wishes to highlight that this omission leads to a downward bias in the ICER of NB32 compared with standard management. Upon request, the company provided a scenario analysis in their response to clarification question B13, in which *"pragmatic application of on-treatment disutilities has been provided"*,[9] assuming all AEs to be associated with a utility decrement of 0.05 for the duration of one week. This analysis increased the company's base-case ICERs against orlistat and SM by £188 and £87 per QALY gained, respectively. The ERG was satisfied that the impact of AEs on HRQoL was likely to be small.

5.2.8 Health-related quality of life

The company uses EQ-5D data from the literature to inform HRQoL in the economic model. This was because in the COR trial programme, only disease-specific QoL data were collected in all but the COR-II study, in which the SF-36 questionnaire was also administered. In the COR trials, HRQoL was assessed using the IWQOL-Lite questionnaire, which assesses the impact of weight on quality of life in the five domains of physical function, self-esteem, sexual life, public distress, and work.[1] However, according to the company, the requirement for a generic measure of HRQoL, the frequency of completion and limited follow-up of the COR trials limited the usefulness of these data for the purposes of the economic model.

The company therefore performed a systematic search for HRQoL studies and, after screening and eligibility assessment, 49 publications were identified from which a total of 39 studies were included in the review. Some of these studies examined the relationship between BMI and EQ-5D utility; others the relationship between weight-related comorbidities and utility. However, the inability to explain the impact of both weight and weight-related comorbidities on utility, limited the usefulness of most included studies. The company therefore explored the use of utility estimates derived from historic HSE patient EQ-5D data from Ara et al.[58] but discarded this option due to inconsistencies in the report.

The company used the PHE weight management economic assessment tool, which was identified through the review. It uses results from regression analysis of individual-level EQ-5D data drawn from HSE from 2011 to 2013. The model includes explanatory variables for BMI, age, gender, and obesity-related conditions (stroke, MI, cancer and T2DM). Both, Tobit model estimates and Ordinary Least Squares regression model estimates are presented. In the company's base-case, the Tobit model utility estimates are used, the OLS estimates are explored in a company's scenario analysis. No justification was provided for the preference of the Tobit over the OLS model. The company presented the relationships between BMI and related disease in an overview presented here in Table 5.8.

As stated in Section 5.2.7, adverse events were not assigned any dis-utilities in the economic model.

ERG comment: The PHE utility regression model was found by searching the NICE and PHE websites. The ERG is concerned that the regression model that informs the utility estimates does not appear to be published in a peer-reviewed journal. As a consequence, given the limited amount of details, the validity of these regression models to estimate utility values can therefore not be assessed by the ERG. Furthermore, the ERG was concerned that the presentation of the regression model used

to estimate patients' utilities did not include checking of the face validity associated with the health state utilities. The company, in their response to clarification question B11[9] provided a summary of example patient utilities, shown in Table 5.9 and compared these with published general population utilities.[69] The company noted that the utility values predicted by the Tobit model for the healthy population resembled the ones from the general UK population and that the remainder of the predicted utilities lay below these (Table 5.9), demonstrating face validity. The ERG questioned the company's preference for the Tobit model but was satisfied by the company's response to clarification question B11.b that Tobit models are generally more appropriate for the modelling of utilities than OLS models, particularly because the alternative OLS models disregard the lower and upper bounds commonly used for the estimation of utilities. The impact of AEs was not incorporated in the model, see Section 5.2.8.

Table 5.8: Public Health England weight management economic assessment tool v2 HSE EQ-5D data analysis

Covariate	Coeff.	Variance-covariance matrix									
		BMI	BMI^2	BMI^3	Age	Female	Stroke	MI	Cancer	T2DM	Const.
Tobit Model Estimates[a]											
BMI	0.05911	0.00008									
BMI^2	-0.00175	0.00000	0.00000								
BMI^3	0.00001	0.00000	0.00000	0.00000							
Age	-0.00440	0.00000	0.00000	0.00000	0.00000						
Female	-0.04054	0.00000	0.00000	0.00000	0.00000	0.00002					
Stroke	-0.18280	0.00001	0.00000	0.00000	0.00000	0.00001	0.00059				
MI	-0.16122	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00048			
Cancer	-0.16403	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00003	-0.00003	0.00028		
T2DM	-0.11093	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00002	0.00001	0.00001	0.00012	
Constant	0.67263	-0.00084	0.00002	0.00000	0.00000	-0.00008	-0.00010	0.00006	0.00002	-0.00001	0.00940
Ordinary Least Squares Regression Estimates											
BMI	0.03293	0.00003									
BMI^2	-0.00094	0.00000	0.00000								
BMI^3	0.00001	0.00000	0.00000	0.00000							
Age	-0.00219	0.00000	0.00000	0.00000	0.00000						
Female	-0.02258	0.00000	0.00000	0.00000	0.00000	0.00001					
Stroke	-0.12652	0.00000	0.00000	0.00000	0.00000	0.00000	0.00044				
MI	-0.11931	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00035			
Cancer	-0.10944	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00002	-0.00001	0.00017		
T2DM	-0.07800	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00001	0.00001	0.00000	0.00007	
Constant	0.65792	-0.00028	0.00001	0.00000	0.00000	-0.00002	-0.00004	0.00001	-0.00002	-0.00002	0.00311

BMI, body mass index; EQ-5D, EuroQol 5-dimensions; HSE, Health Survey for England; MI, myocardial infarction; OLS, ordinary least squares; T2DM, Type II diabetes mellitus.

Notes: [a], Censoring limits were -0.594 and 1; sigma 0.33898 (standard error 0.00365) (both to 5 decimal places).

Table 5.9: Summary of utilities estimated with the Tobit model compared with general population utility estimates

Patient characteristics	Male	Δ	Female	Δ
General population, 30 years	0.93		0.91	
Healthy, 30 years, BMI = 27	0.92	-0.01	0.90	-0.01
Diabetic, 30 years, BMI = 27	0.87	-0.06	0.85	-0.06
History of MI, 30 years, BMI = 27	0.85	-0.08	0.82	-0.09
History of stroke, 30 years, BMI = 27	0.83	-0.10	0.81	-0.10
History of MI and diabetic, 30 years, BMI = 27	0.78	-0.15	0.75	-0.16
History of stroke and diabetic, 30 years, BMI = 27	0.76	-0.17	0.73	-0.18
Healthy, 30 years, BMI = 35	0.88	-0.05	0.86	-0.05
Diabetic, 30 years, BMI = 35	0.83	-0.10	0.80	-0.11
History of MI, 30 years, BMI = 35	0.80	-0.13	0.77	-0.14
History of stroke, 30 years, BMI = 35	0.78	-0.15	0.75	-0.16
History of MI and diabetic, 30 years, BMI = 35	0.72	-0.21	0.69	-0.22
History of stroke and diabetic, 30 years, BMI = 35	0.70	-0.23	0.67	-0.24
General population, 50 years	0.88		0.86	
Healthy, 50 years, BMI = 27	0.88	0.00	0.86	0.00
Healthy, 50 years, BMI = 35	0.84	-0.04	0.82	-0.04
General population, 70 years	0.79		0.77	
Healthy, 70 years, BMI = 27	0.84	0.05	0.81	0.04
Diabetic, 70 years, BMI = 27	0.77	-0.02	0.74	-0.03
History of MI, 70 years, BMI = 27	0.73	-0.06	0.70	-0.07
History of stroke, 70 years, BMI = 27	0.72	-0.07	0.69	-0.08
History of MI and diabetic, 70 years, BMI = 27	0.65	-0.14	0.61	-0.16
History of stroke and diabetic, 70 years, BMI = 27	0.63	-0.16	0.59	-0.18
Healthy, 70 years, BMI = 35	0.79	0.00	0.76	-0.01
Diabetic, 70 years, BMI = 35	0.71	-0.08	0.68	-0.09
History of MI, 70 years, BMI = 35	0.67	-0.12	0.63	-0.14
History of stroke, 70 years, BMI = 35	0.65	-0.14	0.62	-0.15
History of MI and diabetic, 70 years, BMI = 35	0.57	-0.22	0.54	-0.23
History of stroke and diabetic, 70 years, BMI = 35	0.55	-0.24	0.52	-0.25

BMI, body mass index; MI, myocardial infarction.

5.2.9 Resources and costs

Costs in the model consisted of drug acquisition costs, non-drug costs related to standard management (applicable to all treatments considered), obesity-related comorbidity costs and adverse event costs.

Resource identification, measurement and valuation studies

In CS Appendix 17, the PRISMA flow diagram of studies identified for the cost and resource use review is presented. In total, 1,515 citations were identified through database searching, bibliographic searching and from conference proceedings. After screening and eligibility assessment, 22 publications were included in the review, which represented 20 unique studies (10 cost studies, four resource use studies, four resource use and cost studies and two cost effectiveness studies). A tabular summary of the characteristics of each included study is provided in CS Appendix 17.

The company stated that the level of reporting was generally poor across studies, to the extent that it was difficult to elicit useful resource use estimates for this analysis. A notable exception to this was the study by Ara et al.[58] Hence the company used this study to inform healthcare resource use assumptions.

Intervention and comparators drug acquisition costs

Table 5.10 summarises the drug acquisition costs for NB32 (8mg naltrexone/90mg bupropion) and orlistat. NB32 is associated with a four week titration period over which the dosage increases from one tablet daily (week 0), via two tablets daily (week 1) and three tablets daily (week 2) to four tablets per day (week 3 onwards).[40] The dosage for orlistat is three capsules daily (without titration period). As both NB32 and orlistat are oral medicines, it is anticipated that there are no costs associated with their administration.

There are no drug costs associated with standard management.

Table 5.10: Drug acquisition costs

Treatment	Pack size	Cost per pack	Cost per tablet	Source
NB32	112	£73.00	£0.65	List price submitted to the Department of Health
ORL	84	£18.44	£0.22	MIMS[70a]

Source: CS Table 62

Abbreviations: MIMS, Monthly Index of Medical Specialities; NB32, naltrexone 32mg plus bupropion; ORL, orlistat.

[a]The company argued that evidence shows that branded version of orlistat (Xenical) accounted for less than 1% of the total prescription items for orlistat. Hence, costs for Xenical are not included by the company.

Standard management costs

The non-drug resource use items comprising standard management in the model are GP visits, nurse visits and blood tests. The unit price of a GP and nurse visit were assumed to be £44.00 and £14.47 respectively (PSSRU (2015)[71]) while this was £3.01 for the costs of a blood test (NHS reference costs (2015) – Code DAPS05[72]). The resource use (i.e. expected frequencies) were estimated based on a combination of reporting in the COR studies,[28] the publication by Ara et al.[58] and UK clinical expert opinion. Moreover, the non-drug resource use and costs related to standard management were assumed different for the first 56 weeks and thereafter (see Table 5.11).

The company stated that, according to clinical opinion (JW),[64] non-drug resource use received alongside NB32 in the COR-I and COR-II clinical trials is a good reflection of the average diet and exercise regimens prescribed for obese and overweight patients in the UK. Therefore, the company included five GP visits during the first year consistently with these trials (though the timing of the 12/16 weeks GP visit for the response assessment differed between treatments). An additional GP visit was added for the reassessment (of the 5% weight loss) at 56 weeks for NB32 and 52 weeks for orlistat and standard management. Moreover, in line with the study by Ara et al.,[58] the company assumed monthly visits to

a healthcare professional for weight management (i.e. a GP or nurse visit at least every four weeks) and blood tests at baseline and three months for patients on active treatments (i.e. NB32 or orlistat). In addition, the company assumed, based on clinical opinion (JW), that all weight management patients would have an annual blood test to monitor blood glucose levels (either at week 52 or 56).

From week 60 onwards it is assumed that patients would continue to have nurse visits every four weeks.

Based on Table 5.11, it can be calculated that the costs of standard management, during the first 56 weeks, are £403.22 for standard management adjunct to NB32 and orlistat, while this is £397.21 for standard management alone. The company stated that this is in line with clinical opinion, as patients receiving standard management alone would incur approximately the same non-drug resource use costs as patients receiving NB32 or orlistat alongside standard management (excluding additional blood tests for patients receiving adjunctive therapy). After the first year, the costs of standard management are £14.47 (one nurse visit) every four weeks, independent of the treatment.

Table 5.11: Non-drug resource use related to standard management

Time (weeks)	NB32			ORL			SM		
	GP[a]	Nurse[a]	Blood[a]	GP[a]	Nurse[a]	Blood[a]	GP[a]	Nurse[a]	Blood[a]
0	1	0	1	1	0	1	1	0	0
4	0	1	0	0	1	0	0	1	0
8	0	1	0	0	1	0	0	1	0
12	0	1	0	1	0	1	1	0	0
16	1	0	1	0	1	0	0	1	0
20	0	1	0	0	1	0	0	1	0
24	1	0	0	1	0	0	1	0	0
28	0	1	0	0	1	0	0	1	0
32	0	1	0	0	1	0	0	1	0
36	1	0	0	1	0	0	1	0	0
40	0	1	0	0	1	0	0	1	0
44	0	1	0	0	1	0	0	1	0
48	1	0	0	1	0	0	1	0	0
52	0	1	0	1	0	1	1	0	1
56	1	0	1	0	1	0	0	1	0
60+[b]	0	1	0	0	1	0	0	1	0

Source: CS Table 66

Abbreviations: GP, general practitioner; NB, naltrexone 32mg plus bupropion; ORL, orlistat; SM, standard management.

[a]The costs of a GP visit (11.7 minutes) were £44.00 (PSSRU (2015)[71]), the costs of a nurse visit (15.5 minutes) were £14.47 (PSSRU (2015)[71]) and the costs of a blood test were £3.01 (NHS reference costs (2015) – Code DAPS05[72])

[b]These frequencies apply from Week 60 every 4 weeks while patients are still receiving treatment.

Obesity-related comorbidity costs

Costs associated with obesity-related comorbidities were retrieved from the literature (mainly from the literature review performed by Ara et al.[58] inflated from 2009 levels to 2015 levels[71]) and adapted

following UK clinical expert consultation (Table 5.12). Based on clinical opinion (JW),[64] it was assumed that the NHS costs associated with MI, stroke and T2DM are additive.

Table 5.12: Obesity-related comorbidity costs

Category	Cost	Source	Description (of primary) source
MI (Year 1)	£4,210.75	Literature review by Ara et al.[58a]	Economic evaluation of early high-dose lipid lowering therapy to avoid cardiac events, which used bottom-up costing methods and considered hospitalisation, procedural, medical resource use and drug costs.[73]
MI (Year 1+)	£345.91		
Fatal MI[b]	£1,390.80		HTA report evaluating the cost effectiveness of glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome.[74]
Stroke (Year 1)	£9,482.78		UK cost-of-illness model that aimed to estimate stroke-related costs over a 5-year period based on data from a randomised study comparing alternative strategies of stroke care.[75]
Stroke (Year 1+)	£2,664.16		
Fatal stroke[b]	£8,671.94		
T2DM (Year 1)	£347.57	Diabetes UK[76a]	It is not clear whether Ara et al.[58] incorporated T2DM costs after the first year of onset (the company interpreted the costs from Ara et al.[58] as the costs for the first year). Hence, it seemed inappropriate to the company to use the costs from Ara et al.[58] each year. Therefore, a more recent report summarised by Diabetes UK was used. This report estimated monitoring and medication costs to be between £300 and £370 per patient per annum. These costs based on a mix of Type 1 and Type 2 diabetic patients, were used in the absence of specific Type 2 data. However, it should be noted that Type 1 diabetics make up a small minority of cases.[77] An average of these two estimates (£335; at 2012 price level) was used in the model.
T2DM (Year 1+)	£347.57		

Source: CS Table 67 and CS section 5.5.3

Abbreviations: T2DM, Type 2 diabetes mellitus; MI, myocardial infarction;

[a]Costs inflated to 2015 levels using Personal Social Services Research Unit Hospital and Community Health Services inflation indices[71]

[b]Ara et al.[58] included a cost upon death, if the death was caused by MI or stroke. The figure for cardiovascular disease mortality as a proportion of overall mortality (31%) was taken from WHO 2016 data.[78] Of the deaths attributable to cardiovascular disease, the proportions of deaths caused by MI (43.1%), stroke (32.9%) or other causes (24.0%) were taken from WHO 2004 data.[79] From this information, mortality related to MI and stroke, as a proportion of overall mortality, was calculated as 13.4% and 10.2%, respectively.

Adverse event unit costs and resource use

AE rates for patients on NB32 and standard management were calculated based on the COR-I trial (Table 5.13, as the largest trial in the COR trial programme. Costs were considered for AEs that occurred in at least 5% of patients (either treatment arm). This threshold was selected to reflect the British National Formulary criteria of all very common (> 1 in 10) and the majority of common (1 in 100 to 1 in 10) AEs.[80] The company assumed that AEs are treated solely within primary care and costed £44.00, representing a single GP visit. This resulted in weekly AE costs, during treatment of £1.69 and £0.81 for NB32 and standard management, respectively.

According to the company, the level of reporting of AE data for orlistat (in the studies identified in Section 4.10 as well as EMA regulatory documents) was not sufficient to compare AE incidence in orlistat patients accurately to that in NB32 patients. Therefore, the company assumes the same weekly AE costs for patients treated with orlistat as calculated for NB32. The company indicated that it expected the safety profile of NB32 to be non-inferior versus orlistat and hence that this assumption is likely to be conservative.

Table 5.13: COR-I trial adverse event occurrences and rates

Adverse event	**NB32 (total N=573)**			**Standard management (total N=569)**		
	N	Probability (within study)	Instantaneous rate	N	Probability (within study)	Instantaneous rate
Anxiety	9	0.0157	0.00045	12	0.0211	0.00059
Constipation	90	0.157	0.00481	32	0.0562	0.00161
Depression	3	0.00524	0.00015	6	0.0105	0.00029
Diarrhoea	26	0.0454	0.00131	28	0.0492	0.0014
Dizziness	54	0.0942	0.00279	15	0.0264	0.00074
Dry mouth	43	0.075	0.0022	11	0.0193	0.00054
Headache	79	0.138	0.00418	53	0.0931	0.00271
Hot flush	30	0.0524	0.00151	7	0.0123	0.00034
Insomnia	43	0.075	0.0022	29	0.051	0.00145
Nasopharyngitis	29	0.0506	0.00146	31	0.0545	0.00155
Nausea	171	0.298	0.00998	30	0.0527	0.0015
Sinusitis	30	0.0524	0.00151	34	0.0598	0.00171
Upper respiratory tract infection	57	0.0995	0.00295	64	0.113	0.00331
Vomiting	56	0.0977	0.0029	14	0.0246	0.00069
Source: CS Tables 69 and 70						

ERG comment: The ERG considered it plausible to use Ara et al.[58] (identified in the review) to inform healthcare resource use assumptions. Regarding the costs of standard management, it is unclear to the ERG why the company added a GP visit for the 52 week assessment for patients receiving standard management only (in addition to the five GP visits during the first year which was considered to be reflective of UK clinical practice). Therefore, the ERG removed this GP visit for patients receiving standard management only.

Drug wastage associated with NB32 was not considered in the base-case model. However, when considering this in a scenario analysis (response to clarification question B15), it is illustrated that not

considering drug wastage is not conservative (ICER compared with orlistat increased by £3,426). Given the unavailability of data, the ERG was not able to consider drug wastage in the ERG base-case model.

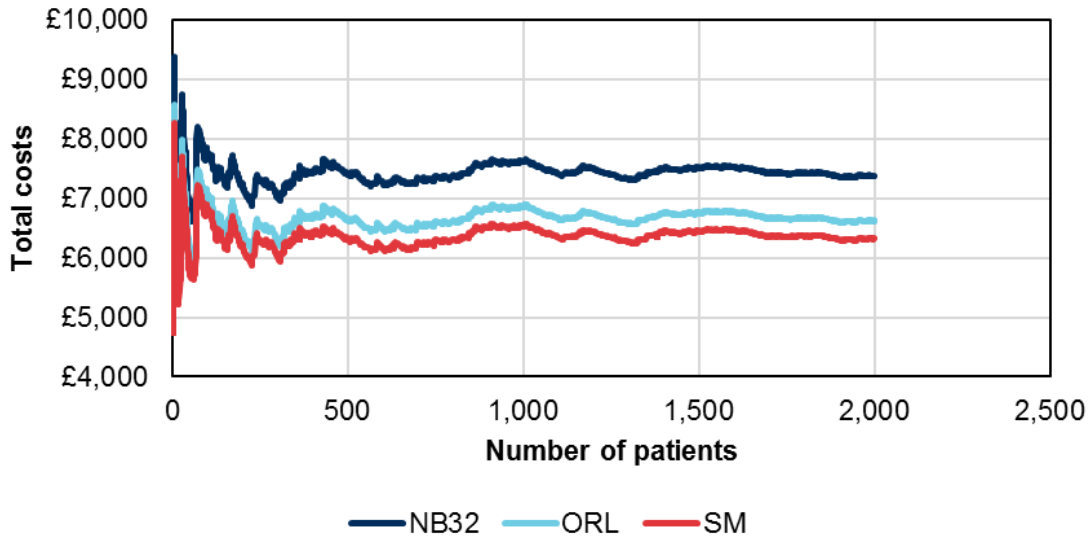
It was unclear to the ERG why the company considered it plausible to assume only a single GP visit for each adverse event. Assuming outpatient costs would increase the ICER of NB32 versus orlistat with £4,408 (CS Table 79). Moreover, it is unclear why the company expected the safety profile of NB32 to be non-inferior compared to the safety profile of orlistat. The company provided no systematic overview of evidence that showed that the AE profile of orlistat was indeed worse than that of NB32. There is no direct evidence comparing the two drugs and indirect treatment comparisons between the drugs focused on efficacy but not on safety outcomes. Therefore the company's assertion of the likely superiority of NB32 in relation to orlistat in terms of AE remains speculative. Therefore, it is questionable whether assuming the same AE costs for orlistat as calculated for NB32 is appropriate. Finally, it is unclear to the ERG why the company used the COR-I trial only to inform the rate of AE (e.g. why the COR-DM trial was not considered for T2DM specific AE rates).

5.2.10 Cost effectiveness results

Methodology for model analyses

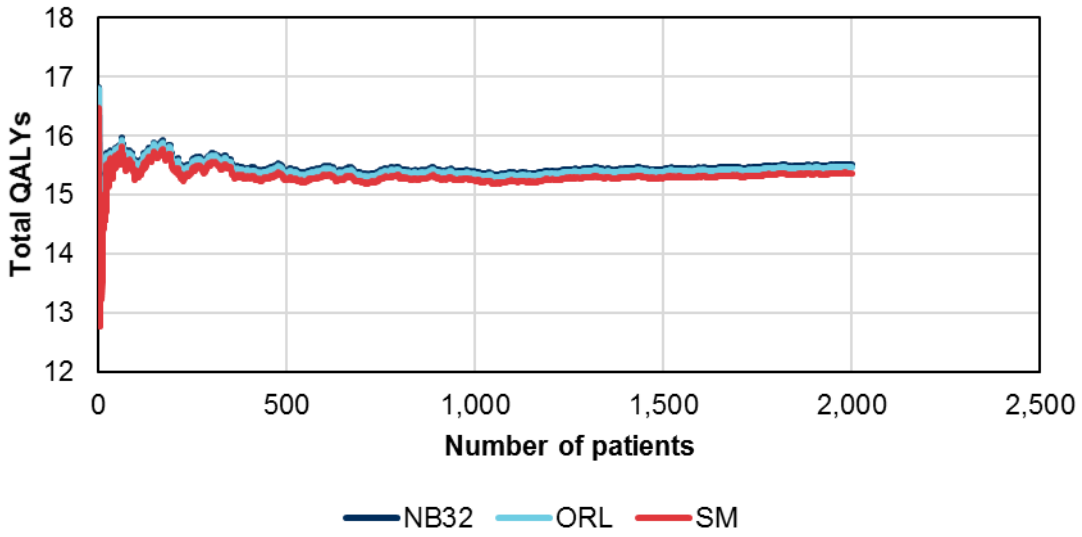
In order to obtain reliable results a sufficient number of individual randomly sampled random patient profiles need to be run such that the model results converge to a consistent value. In order to establish this number the company ran the model with 2,000 individual randomly sampled random patients and recorded total costs and total QALYs (Figures 5.10 and 5.11). Based on this exercise, the company decided that sampling 1,000 patients would a sufficient number to obtain a deterministic model result.

Figure 5.10: Diagnostic exercise – total costs



Source: CS, Figure 34
Key: NB32, naltrexone 32mg plus bupropion; ORL, orlistat; SM, standard management.

Figure 5.11: Diagnostic exercise – total QALYs



Source: CS, Figure 35
Key: NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management.

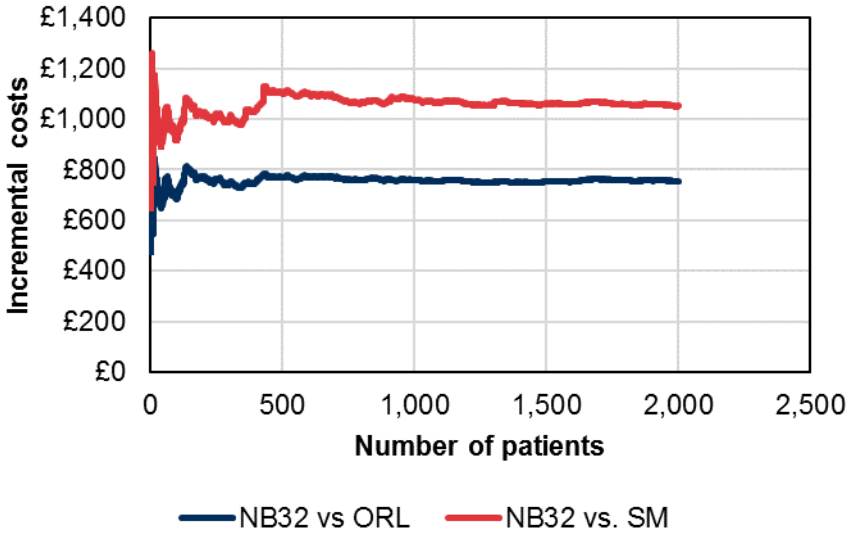
The probabilistic sensitivity analysis (PSA) took into account uncertainty surrounding cost estimates (except for drug cost and administration costs), utility estimates, BMI change in time, change in weight loss, and proportion of patients with response. Stochastic parameters not included in the PSA are TTD, administration costs (fixed zero), time until weight regain (fixed three years), and the probability of obesity related events (all cause mortality, non-fatal MI, non-fatal stroke, and onset of T2DM). Omitting to take into account uncertainty in the probabilities of all cause mortality and events was due to a lack of detail in the source.[58] An overview of the model inputs can be found in Appendix 18 of the CS.[1] For the PSA the company used 500 individual randomly sampled patients; the "*smallest number of patient profiles required after which model results appear to stabilise*". The number of PSA runs was chosen at 100, based on the run time required.

ERG comment: Ara et al.[58] used a cohort of 1,000,000 patients in their patient-level simulation and stated that, with a cohort size of 200,000 patients, there was still a small amount of variation in results, which stabilised after simulation of 400,000 patients. In contrast, a cohort of only 1,000 patients was used in the CS. The company provided two arguments to justify the lower number of sampled patients in comparison to Ara et al. First, the use of Excel instead of Simul8, which Ara et al. used, limited the number of sampled patients with regard to run-time. Second, the company argued that they "*were able to avoid the need to produce model results for a very large cohort (such as 1,000,000 patients) by controlling baseline characteristics for each model run. By controlling these characteristics, the only difference across patients was the treatment received.*"[9] The ERG finds this statement puzzling, as this is standard practice. In a patient-level simulation in each model run an identical individual randomly sampled patient should be evaluated for each of the comparators in the assessment.

The ERG asked the company to provide additional results from the diagnostic exercise to examine the minimum number of individual randomly sampled patients (incremental costs, QALYs and the incremental cost-effectiveness ratio (ICER)). The results indicate that for total and incremental QALYs 1,000 individual randomly sampled patients seems a sufficient number to obtain a reliable model result. For total and incremental costs and the ICER, the diagnostic exercise still shows fluctuations in results around 1,000 sampled patients (Figures 5.12, 5.13 and 5.14). The ERG ran the base case model with

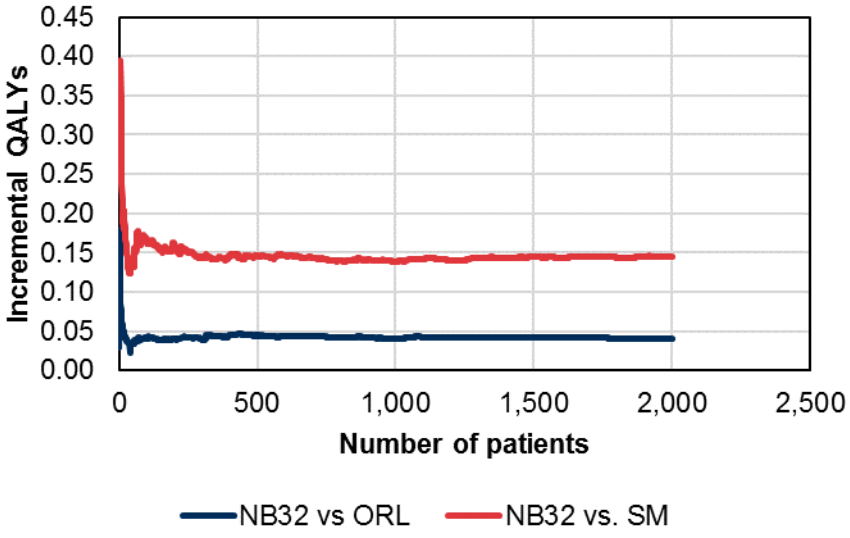
1,000 patients, which resulted in an ICER of NB32 versus orlistat ~£3,000 higher than the company's base case result. According to the ERG, the model should ideally be evaluated using at least 1,500 sampled patients.

Figure 5.12: Diagnostic exercise – incremental costs



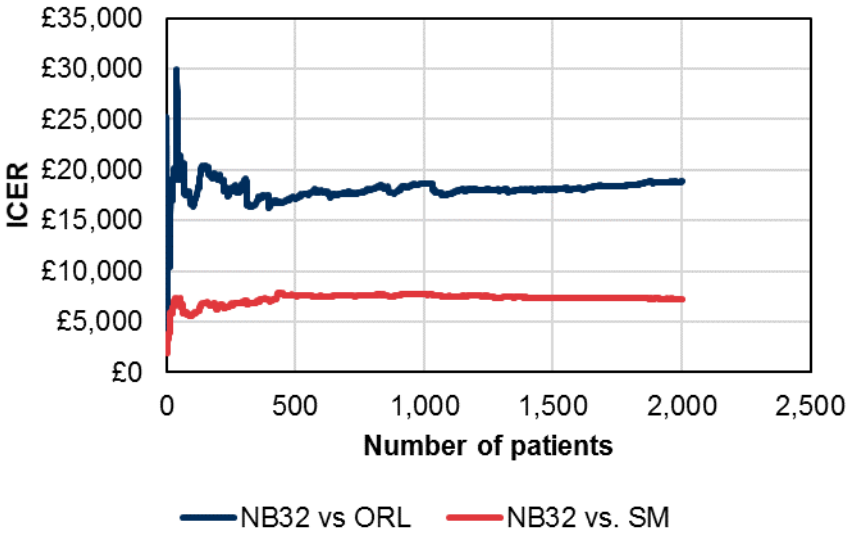
Source: Company response on clarification questions Figure 4
Key: NB32, naltrexone 32mg plus bupropion; ORL, orlistat; SM, standard management.

Figure 5.13: Diagnostic exercise – incremental QALYs



Source: Company response on clarification questions Figure 5
Key: NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management.

Figure 5.14: Diagnostic exercise – ICER



Source: Company response on clarification questions Figure 6
Key: ICER, incremental cost-effectiveness ratio; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management.

For the following stochastic parameters in the model uncertainty is not accounted for in the PSA: TTD, risk of obesity related events and the natural history of BMI model. The ERG asked the company to provide further clarification. The company stated that uncertainty of risk of obesity related events and natural history of BMI was not included in the PSA because Ara et al. 2012,[58] the source for key time to event equations, did not report variance-covariance matrices, and did not respond to email requests for these in time for the submission. According to the ERG, the company could have used standard errors, and/or have made assumptions to account for the uncertainty in these estimates in the PSA. The company also argued that "much of the key uncertainty regarding model results is structural and methodological, and based on the key conservative assumptions underpinning the analysis" and hence "the choice of the number of PSA iterations (be that 100, 100,000 or 100,000,000) does not demonstrate how conservative the structural model assumptions are.".[9] The ERG disagrees on that key methodological and structural assumptions are conservative (see Section 5.2.2). Moreover, the PSA is not only a method to show uncertainty around mean outcomes, but also the preferred method to obtain the mean outcomes.[9] Hence, if the PSA is flawed, so is the estimation of the mean outcomes of the model.

The PSA is performed with a smaller (500) number of individual randomly sampled patients. The ERG disagrees that this at this number of patients results appear to stabilise. As the figures of the diagnostic exercise show, at 500 patients the results for both QALYs and costs do not converge yet. As a result, the deterministic result on which the PSA runs are applied is unreliable. Another weakness of the PSA methodology the company used is the small number (100) of PSA runs. It is very unlikely this will result in a reliable probabilistic model estimate for an individual patient profile. Usually at least 1,000, and often much higher numbers of 5,000 to 10,000 PSA runs are required to obtain a reliable result. The ERG asked the company to provide a justification as to why the company believes that probabilistic sensitivity analysis using 100 simulations results in stable/plausible results. In their response, the company reiterated the choices made for an individual-level model *("better suited than a cohort-level approach to capture the chronic implications of both weight and weight-related health events in a heterogeneous group of overweight and obese patients"*), programmed in MS Excel *("applicable across various health technology assessment agencies internationally – some of which only consider*

models constructed within Microsoft Excel®. Microsoft Excel is also the preferred software package of NICE, and is typically considered more transparent than simulation models constructed in other software packages.").[9] The company further stated "*Regarding the number of PSA simulations, a trade-off between the number of patient profiles and the number of probabilistic draws was made. Given that within the PSA, 500 patient profiles are used for each PSA run, the number of PSA runs was chosen at 100. This number of PSA iterations was associated with a long run-time due to the limitations in processing power associated with Microsoft Excel*.".[9] Indeed, run-time of the model is relatively long. The ERG recorded run times between 450 and 600 hours of a PSA with 100 individual randomly sampled patients and 1,000 PSA runs. These numbers of sampled patients and PSA runs are still too low to obtain a reliable result. The ERG acknowledges that in any model study trade-offs are made between validity and reliability of the result and practical considerations. However, companies should provide a submission that is compliant to the NICE decision making process in which probabilistic model results are preferred, and the model is assessed by the ERG in a period of eight weeks. For this model, it was unfeasible for the ERG to perform an adequate assessment of the model's probabilistic results within the time frame of a NICE submission.

The ERG believes that the reliability of the probabilistic model results is severely compromised as a result of not accounting for uncertainty in some stochastic parameters, and instability due to too low a number of individual randomly sampled patients and too low a number of PSA runs.

Base-case model results

In the base-case deterministic analysis NB32 gains 0.0765 QALY versus standard management, and 0.0192 QALY versus orlistat. The incremental costs of NB32 are £1,044 versus standard management and £750 versus orlistat. The incremental cost-effectiveness ratio (ICER) of NB32 versus standard management is £13,647 per QALY. The estimated ICER versus orlistat is £32,084 per QALY. The latter ICER is also the ICER in a full incremental analysis. The probabilistic analysis shows a similar ICER for NB32 versus standard management (£13,958) and a higher ICER for NB32 versus orlistat (£36,084). See Tables 5.14 and 5.15 below.

Table 5.14: Base case deterministic results

Treatment	Total			Incremental			ICER (QALYs)	
	Costs	LYs[a]	QALYs	Costs	LYs[a]	QALYs	Versus baseline (SM)	Incremental
SM	£6,519	33.4768	15.3616					
ORL	£6,814	33.5151	15.4148	£294	0.0383	0.0531	£5,538	£5,538
NB32[b]	£7,563	33.5343	15.4381	£750	0.0192	0.0234	£13,647	£32,084

Key: ICER, incremental cost-effectiveness ratio; LY, life year; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management; T, technologies.
Source: Table 72 CS
Note: [a] LYs are undiscounted, costs and QALYs are discounted. [b] The ICER of NB32 versus SM amounts to £13,647

Table 5.15: Probabilistic base case model results

Treatment	Total			Incremental			ICER (QALYs)	
	Costs	LYs[a]	QALYs	Costs	LYs[a]	QALYs	Versus baseline (SM)	Incremental
SM	£6,411	33.5673	15.3664					
ORL	£6,667	33.6128	15.4176	£256	0.0455	0.0512	£4,993	£4,993
NB32[b]	£7,409	33.6242	15.4379	£742	0.0115	0.0204	£13,936	£36,405

Key: ICER, incremental cost-effectiveness ratio; LY, life year; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management; T, technologies.
Source: Adapted from Table 78 CS
Note: [a] LYs are undiscounted, costs and QALYs are discounted. [b] The ICER of NB32 versus SM amounts to £13,936

The company presented disaggregated results for costs. This shows that the cost differences between the comparators is caused by the treatment acquisition costs (Table 5.16)

Table 5.16: Summary of discounted costs by cost category

Technologies	Costs				
	Treatment acquisition	SM and CM	AEs	Death	Total
SM	£0	£5,982	£171	£367	£6,519
ORL	£238	£5,993	£216	£366	£6,814
NB32	£995	£5,983	£220	£366	£7,563

Key: AE, adverse event; CM, condition management; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; SM, standard management.
Source: CS Table 76

ERG comment: The deterministic total cost result should be interpreted with caution due a possible to small number of sampled patients to obtain a stable result. According to the NICE DSU guidance,[81] decision making should be based on probabilistic model results. However, in this submission, the PSA results are flawed, for reasons explained in the previous section.

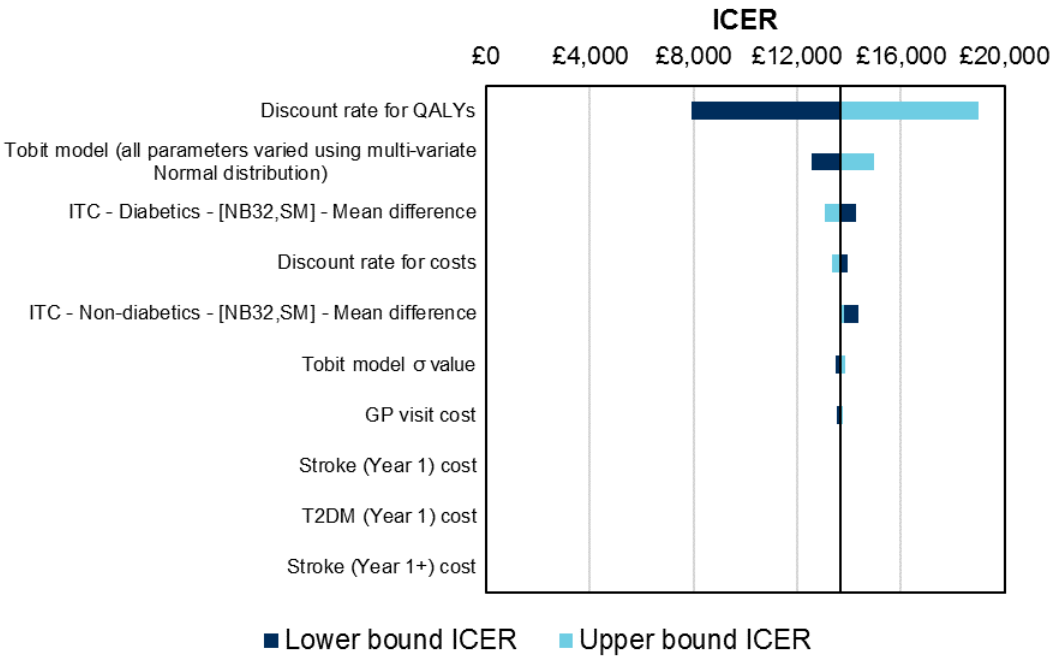
The ERG asked for more information on disaggregated outcomes of the model, such as costs associated with events, and time with events, but these were not provided by the company.

5.2.11 Sensitivity analyses

The company provided scatterplots and cost effectiveness acceptability curves based on the results of the PSA for NB32 versus standard management and NB32 versus orlistat separately. The cost effectiveness acceptability curves (CEAC) show that NB32 is associated with a 98% probability of being cost effective versus standard management and a 0% probability of being cost effective versus orlistat at a willingness to pay (WTP) threshold of £20,000 per QALY.

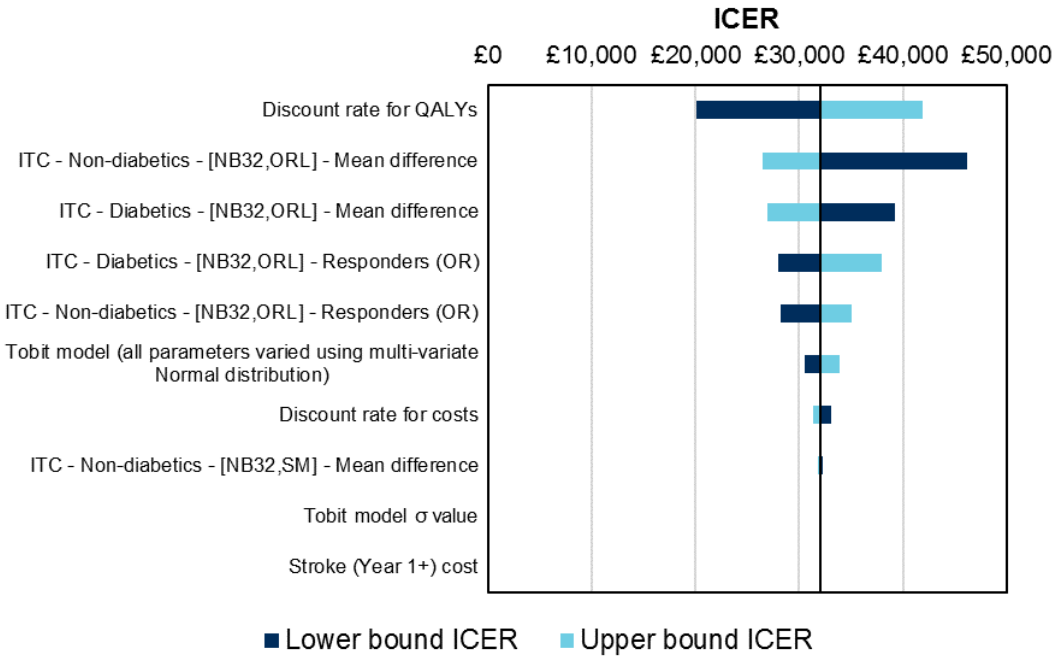
The deterministic sensitivity analyses performed by the company show that the most influential parameters are the parameters of the Tobit model for utilities and the discount rate for QALYs, as well as parameters related to the measures of relative efficacy from the ITC.

Figure 5.15: OWSA – NB32 versus SM



Key: GP, general practitioner; ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison; NB32, naltrexone 32mg plus bupropion; OLS, ordinary least squares; QALY, quality-adjusted life year; SM, standard management; T2DM, Type 2 diabetes mellitus. Source: CS Figure 42

Figure 5.16: OWSA – NB32 versus ORL



Key: GP, general practitioner; ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison; NB32, naltrexone 32mg plus bupropion; OLS, ordinary least squares; OR, odds ratio; QALY, quality-adjusted life year; SM, standard management; T2DM, Type 2 diabetes mellitus. Source: CS Figure 43

Note: The eighth parameter (ITC – Non-diabetics – [NB32, SM] – Mean difference] is not an error. This parameter is featured within the outcome of the analysis as patients who discontinue treatment with orlistat may continue treatment with standard management alone.

The company performed scenario analyses on the following model aspects: the time period over which weight is regained, the cost of T2DM, the utility estimates, costs of AEs, discounting, and the time horizon. The most influential scenarios were shortening the time period for weight regain from three to two years (ICER £41,016), and shortening the time horizon form lifetime to 15 years (£53,514).

Table 5.17: Scenario analysis results

Scenario				ICERs	
				NB32 vs	
n	**Model setting**	**Base case**	**Scenario tested**	**ORL**	**SM**
0	Base case			£32,084	£13,647
1	Weight regain	3 years	2 years	£41,016	£14,113
2	Weight regain	3 years	5 years	£29,739	£11,880
3	Cost of T2DM	£347.57	£175.86 in Year 1 only	£36,096	£13,764
4	Utility model	Tobit	OLS	£36,771	£10,285
5	AE costs	All GP	All outpatient	£36,492	£15,130
6	Discounting	3.5% for costs & effects	1.5% for costs & effects	£28,323	£9,969
7	Time horizon	Lifetime	15 years	£53,514	£22,763

Key: AE, adverse event; GP, general practitioner; ICER, incremental cost-effectiveness ratio; LY, life year; NB32, naltrexone 32mg plus bupropion; OLS, ordinary least squares; QALY, quality-adjusted life year; SM, standard management; T2DM, Type 2 diabetes mellitus. Source: CS Table 70

The company performed subgroup analyses for patients with and without T2DM at baseline. The results showed that the ICER of NB32 versus orlistat it higher in patients who have T2DM at treatment initiation (£72,069), compared to patients who do not have T2DM at that moment (£28,298). The company warns that the results are uncertain because the data regarding comparisons of NB32 to orlistat in patients with T2DM are limited as shown, in Section 4.10 of the CS.[1]

Table 5.18: Base case results – T2DM patients at baseline only

T	Total			Incremental			ICER (QALYs)	
	Costs	**LYs[a]**	**QALYs**	**Costs**	**LYs[a]**	**QALYs**	**Versus baseline (SM)**	**Incremental**
SM	£10,199	32.7296	14.3707					
ORL	£10,496	32.7583	14.4295	£297	0.0287	0.0588	£5,059	£5,059
NB32	£11,216	32.7656	14.4395	£720	0.0073	0.0100	£14,797	£72,069

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management.
Note: [a], LYs are undiscounted, costs and QALYs are discounted. Source CS Table 80

Table 5.19: Base case results – non-T2DM patients at baseline only

T	Total			Incremental			ICER (QALYs)	
	Costs	LYs[a]	QALYs	Costs	LYs[a]	QALYs	Versus baseline (SM)	Incremental
SM	£3,844	33.5497	15.7335					
ORL	£4,077	33.5854	15.7706	£233	0.0356	0.0371	£6,283	£6,283
NB32	£4,811	33.5944	15.7966	£734	0.0090	0.0259	£15,339	£28,291

Key: ICER, incremental cost-effectiveness ratio; LYG, life years gained; NB32, naltrexone 32mg plus bupropion; ORL, orlistat; QALY, quality-adjusted life year; SM, standard management; T, technologies.
Note: [a], LYs are undiscounted, costs and QALYs are discounted. Source: CS Table 81

ERG comment: For reasons described in the previous paragraph, the ERG believes the PSA results are flawed. As a result the CEACs should be interpreted with extreme caution. The company did not perform deterministic sensitivity analyses on all parameters that are uncertain. Most notably, some parameters that were left out of the PSA were also not varied in deterministic sensitivity analyses, such as TTD, the probability of obesity related events, and the BMI natural history model. For instance the uncertainty around TTD, influencing both treatment effects and costs, is likely to significantly affect model results. The subgroup analyses with T2DM and non-T2DM patients should be interpreted with caution, because in these subgroup analyses the baseline characteristics (which impact obesity-related comorbidities and utility values) are independent on T2DM status. As stated in section 5.2.3 this leads to counter-intuitive patient profiles.

5.2.12 Model validation and face validity check

Face validity

The company attempted to achieve face validity by using advice from a clinical expert (JW). The company indicated that advice from this expert was used to inform and validate key clinical assumptions in the analysis.[64]

Internal validity

An economist not involved in model adaptation reviewed the model for coding errors, inconsistencies and the plausibility of inputs. This included examining known modelling errors, and questioning of the assumptions. In addition, in response to clarification question B19, the company stated that it used a checklist of basic validity checks (e.g. setting all costs to zero and ensuring the model outputs zero costs), sheet by sheet check of model logic (e.g. checking DICE equation logic), module by module check of VBA logic, validity assessment of outcomes (e.g. comparing available trial data with the outcomes of the model), and editorial checks (e.g. performing a spell check of model content).

External validity

The company compared the estimated LYs in the model (range: 33.48 to 33.53 for the three treatments) with UK life expectancy for the general population at the age of 47 years (range 34 to 37 for males and females respectively) and considered this to be a validation of the LYs estimated in the model.

Cross validity

The company considered that the total QALYs from their model are similar to those reported by Ara et al.,[58] for standard management (15.36 versus 15.13) and orlistat (15.41 versus 15.30).

ERG comment: The ERG considered the internal validity of the model (e.g. checking formula's in the DICE sheet, examining the implementation of TTD in the model, examining available intermediate outcomes). However, the ERG was unable to examine the internal validity of the model according to its usual standards. This was mainly a consequence of the time available to the ERG in relation to the time the model requires to run one single deterministic analysis, and the inability to examine intermediate outcomes. For instance, the nature of the model hampered the ERG's ability to do sensitivity analysis; extreme value analysis; trace analysis/analysis of intermediate outcomes which are recommended by the ISPOR taskforce on model transparency and validation.[82] Therefore, the ERG cannot guarantee that there are no modelling errors (in addition to the methodological flaws described below). In this light, the ERG considers it as troublesome that the company did not provide the results of the internal validation it performed (as requested in response to clarification question B19).[9]

The ERG wishes to highlight that it considers the model submitted by the company as unfit for purpose. The implementation in DICE resulted in extremely slow runtimes (6 hours on average, but with occasional model run times of 10 hours, depending on computer specifications) for the deterministic analysis only. It should also be noted that the model crashed on most of the computers that it was tried on.

One of the main validity issues or methodological flaws the ERG encountered was the lack of an updating event or integration of BMI over time. The average time between model entry and death was 33.5 years (median: 35.1 years; interquartile range 6.2 years - 56.7 years). The ERG calculated that on average patients in the model have seven events (excluding the start and end events), on average four of which occur during the first year (i.e. until the second assessment or if not applicable, due to treatment discontinuation during the first year, until date of treatment discontinuation). Hence after the first year, patients have on average only three events in 32.8 years, equalling to an average of one event per 10.6 years (median: 10.0 years; interquartile range 1.7 years - 23.9 years). This entails that BMI after the first year is only updated on average once every 10.6 years (implicitly assuming a stable BMI in the periods between events), while this should be updated at least annually to reflect the increasing BMI due to its correlation with age (as reflected in the natural history model predicting BMI over time). An alternative would be discrete integration of the BMI function. Apart from the continuous development of BMI not being reflected (and the impact on associated risks and utility values), the lack of model updating also affects other assumptions in the model. For example, the assumption regarding weight regain after treatment discontinuation for NB32 and orlistat was intended to reflect linear weight regain for a period of three years after which the BMI is obtained (predicted by the natural history model). However, if there is no event in this three year weight regain period, which is more likely than not (based on the average of one event per 10.6 years), the BMI estimated at the time of treatment discontinuation is maintained for this weight regain period of three years after which the weight is regained instantly. It should be noted that if the death event would be excluded from this calculation (also excluding 0.7% of the patients with death as the only event), the average time until one event increases to 17.2 years (median: 14.1 years; interquartile range 2.9 years - 47.9 years). The death event, logically, does not provide an intermediate update of BMI. Additionally, given that BMI is not updated at the stop adjunct, stop treatment and death events, the average period without BMI update presented above is an underestimation of the actual period without BMI update in the model. The ISPOR taskforce on DES[83] states that in case "*the likelihood of discrete events is a function of the value of a continuous measure (e.g., diabetic complications are a function of Hb A1c, or clinical presentation is a function of tumor size), as described in the model structure and design section*", that *"time checks can be used to sample the likelihood of discrete events, conditional on the status of the continuous measure of disease progression (e.g., monthly time checks to update Hb A1c levels and define related probabilities of*

complications).” In the present model, the likelihood of events, as well as utility, is a function of BMI, which is predicted (by the natural history model) to change annually with increasing age. Hence, given the average time to event, it would have been recommended to incorporate ‘time checks’ (i.e. ‘update events’). Given that BMI is underestimated as a consequence of this methodological flaw, the utility values and the time to the events in the model are overestimated, likely inducing bias in favour of NB32. Moreover, assuming stable BMI for long periods of time also limits the face validity of the model.

Considering face validity, the ERG wishes to highlight that owing to the technical implementation of the model, it was difficult to assess the face validity of all parts of the model in the given time-frame. For example, the ERG identified one potential issue with the proportion of responders at secondary assessment. For both NB32 and orlistat this is directly determined by the estimated weight loss. Responders to these treatments at the primary assessment are therefore on average set to be responders at the one year assessment, too. As the company pointed out in their response to clarification question B5, not all patients are responders at the one year assessment in the model, due to the weight loss distribution being sampled from for each patient.[9] The weight loss distribution was a normal distribution with a mean weight loss of 11.7% and SD of 7.2% for NB32 and similarly for orlistat and SM based on ITC results, which means that only a small proportion of NB32 patients continuing on treatment after the primary assessment would be non-responders at one year (approximately 17%). When validating this assumption against the patient output, the ERG noted that proportions of non-responders at the one year assessment indicated a similar proportion of responders compared to the number of patients at baseline for patients treated with NB32, orlistat and only receiving standard management (57.3%, 55.6%, 56.6% of all patients at baseline, respectively). It is unclear how such similar proportions were obtained. These could be a result of the different events (treatment discontinuation, death, weight loss) but it was not possible to check whether these estimates truly exhibited face validity. It should therefore be considered whether simpler approaches (e.g. an individual-level state transition model) would have been more appropriate to reflect this decision problem, given the gain in transparency and given that it would have been possible to reflect the condition-specific events in such a model. A state-transition approach would potentially resolve most of the validity issues (e.g. the lacking updating event).

The ERG noted a small inconsistency in the implementation of the TTD estimation in the model. The TTD estimates for NB32 and orlistat were a result of 1. sampling from TTD KM estimates of patients up to the primary assessment (16 weeks for NB32 and 12 weeks for orlistat), 2. sampling from TTD KM estimates of patients up to the secondary assessment (56 weeks for NB32 and 52 weeks for orlistat), and 3. sampling from the TTD KM estimates for the remainder of time until there was no more available data from the NB-CVOT study (the maximum was a total of three years). The way that the sampling was done in each of these three steps was by using randomly generated numbers between 0 and 1, then finding the closest matching KM estimate (that is percentage of patients still on treatment) and then looking up the associated time to discontinuation. A vlookup function uses the random number and matches it to the largest value that it finds which is smaller than the random number, going through the lookup table from top to bottom (the table is sorted starting with the lowest percentage of patients still on treatment to the highest). The way the company sorted the values in lookup tables is different for NB32 and orlistat for the time period up to primary assessment – this results in very slightly smaller values of TTD for NB32 than for orlistat, but the impact of this inconsistency is expected to be minimal.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Table 5.20 summarises all main issues highlighted by the ERG in Section 5.2, indicates the expected direction of bias introduced by these issues and whether these are examined in any analyses/incorporated in the ERG base-case.

Table 5.20: Main ERG critique of company's submitted economic evaluation

Issue	Bias introduced[a]	ERG analyses (analysis number in section 5.3)	Addressed in analysis?
Model structure (section 5.2.2)			
• Weight regain assumptions deviated from those in Ara et al.[58] in that the company modelled weight regain towards the predicted BMI instead of the baseline BMI.	+	Base-case (7)	Response to clarification question B1; ICER (NB32 vs orlistat) increased by £1,536.
• Weight regain is not implemented linearly in the economic model.	+	Base-case (1)	
• The model structure is restricted to only having two cardiovascular events. Experiencing a stroke after two MI's might have an impact on the outcomes and costs.	-		
Population (section 5.2.3)			
• Baseline BMI is vastly underestimated in the economic model, hence overestimating utility and time to T2DM, cardiovascular events and death.	+	Base-case (4)	
• The proportions of current smokers, patients receiving anti-hypertensive medication and/or statins are most likely underestimated.	+	Base-case (5)	
• Counter-intuitive patient profiles are generated as correlations between patient characteristics are not incorporated.	+/-	Base-case (5)	
• It is questionable whether the results of the economic analyses are representative for patients with a history of angina and/or diabetes other than T2DM.	+/-		
Interventions and comparators (section 5.2.4)			
• Behaviour interventions, bariatric surgery and re-treatment are not implemented.	+/-		
Treatment effectiveness and extrapolation (section 5.2.6)			
• The company used modified ITT analysis instead of ITT analysis for estimation of percentage of weight loss and response rates.	+	Base-case (2)	
• Pooling from all COR studies is inappropriate because: 1. BMOD uses different intensity of treatment- accompanying management; 2. COR-II data are only available up to 28 weeks.	+	Base-case (2)	

Issue	Bias introduced[a]	ERG analyses (analysis number in section 5.3)	Addressed in analysis?
• An assumption is made that weight loss is equivalent for NB32 and orlistat at different times (16 weeks and 12 weeks, respectively).	+/-		
• The mean change in weight for orlistat at primary assessment was derived using the mean difference in treatment effect at secondary assessment (for NB32 versus orlistat) applied to NB32 mean change in weight at primary assessment	+	Base-case (3)	
• TTD data for orlistat were obtained by linearly scaling the 16 weeks TTD curve for NB32 to fit into the 12 weeks. The company did not provide alternative analysis upon request.	+/-	Base-case (8)	
• TTD (after 56 weeks) is under-estimated because it was derived from a more severe patient population (from NB-CVOT study) and it was assumed that all patients discontinued after the trial period had ended.	+		
Adverse events (sections 5.2.7-5.2.9)			
• AE-related utility decrements were not included.	+		Response to clarification question B13; ICER (NB32 vs orlistat) increased by £188.
• Only the COR-I trial was used to inform AE rates; the COR-DM trial could have been used to obtain T2DM specific AE rates.	+/-		
• Questionable whether the assumption of equivalent AE costs for NB32 and orlistat is conservative.	+/-		CS Table 79; using outpatient costs would increase the ICER of NB32 versus orlistat with £4,408
Health-related quality of life (section 5.2.8)			
• Use of PHE weight management economic assessment tool for derivation of utilities may not be appropriate.	+/-		
Resources and costs (section 5.2.9)			
• An unnecessary GP visit, related to response assessment, is incorporated for standard management.	+	Base-case (6)	

Issue	Bias introduced[a]	ERG analyses (analysis number in section 5.3)	Addressed in analysis?
• Assuming only a singly GP visit for each adverse event without plausible justification.	+		
• NB32 drug wastage was not considered in the model	+		Response to clarification question B15; ICER (NB32 vs orlistat) increased by £3,426.
Cost-effectiveness analyses (sections 5.2.10 and 5.2.11)			
• The number of simulated patients (1,000) is too low to provide stable results; the ICER varies substantially with each model run.	+/-		The ICER presented by the company was ~£3,000 lower than the one obtained by the ERG after re-running the deterministic analysis.
• The PSA does not appropriately reflect uncertainty surrounding the most important parameters (e.g. the uncertainty surrounding TTD, a key parameter in the model, was neglected).	+/-		Implementation of PSA violates best practices and hence should not be used for decision making. Moreover, the model run times are prohibitive to appropriately run PSA.
• The number of PSA simulations is restricted to 100, which is too low to appropriately reflect uncertainty	+/-		
Validation (section 5.2.12)			
• The lack of an updating event or discrete integration of the BMI function overestimates utility and time to T2DM, cardiovascular events and death. Moreover, implicitly assuming a stable BMI for on average 17 years hampers the face validity of the model.	+		

Abbreviations: NA, not applicable
[a]Likely conservative assumptions (of the intervention versus all comparators) are indicated by '-'; while '+/-' indicates that the bias introduced by the issue is unclear to the ERG and '+' in indicates that the ERG believes this issue likely induces bias in favour of the intervention versus at least one comparator.

Based on all considerations from Section 5.2 (summarised in Table 5.20), the ERG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments made by the ERG form the ERG base-case and were subdivided into three categories (derived from Kaltenthaler 2016[84]):

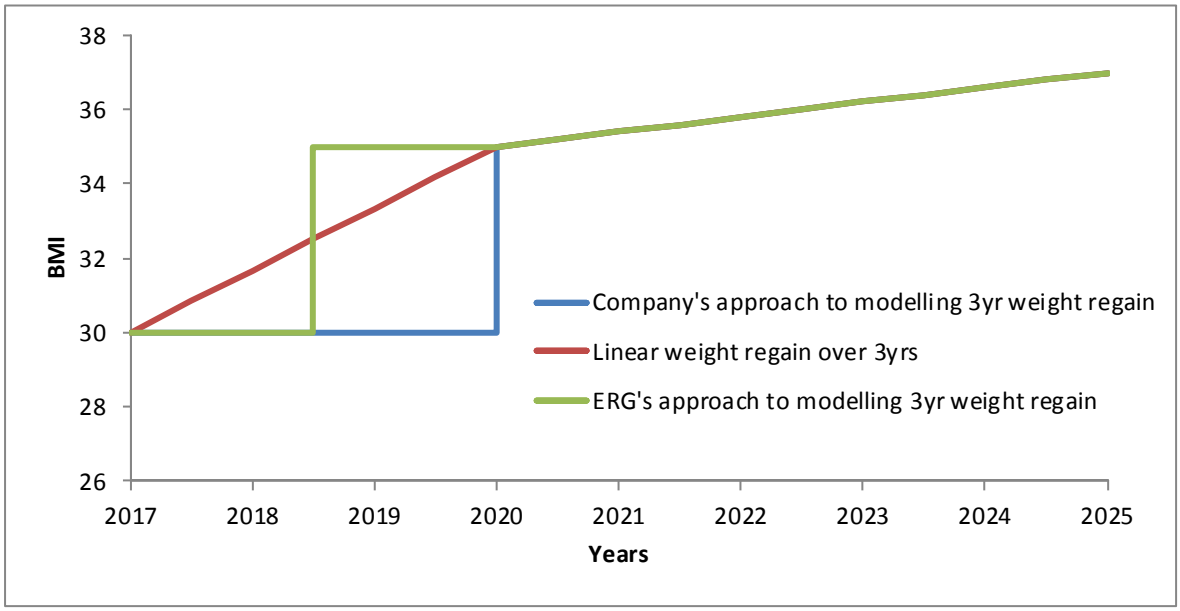
- Fixing errors (correcting the model were the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model were the ERG considers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

Additionally, exploratory sensitivity analyses were performed by the ERG to examine the potential impact of alternative assumptions on the cost effectiveness estimates.

Fixing errors

1. Fixing errors consisted of using a weight regain period of 1.5 years after which weight is instantly regained, to reflect the three year linear weight regain assumption made by the company (see Figure 5.17 for an illustration of this with an example assuming treatment discontinuation at start of 2017). In the company base-case the weight is regained instantly after 3.0 years (instead of linearly over three years' time), see Sections 5.2.2 and 5.2.6 for more details. The ERG's approach first under-estimates BMI, then over-estimates it. In contrast to this, the company's approach under-estimates BMI for the whole duration of three years.

Figure 5.17: Illustrative example of ERG implementation of weight regain to reflect the three year linear weight regain assumption



Fixing violations

2. Using the ITT data instead of mITT data based on the COR-I and COR-DM trial only.
   The ERG considered the usage of ITT data from the COR-I and COR-DM trials as most appropriate (see Section 5.2.6 for more details).

3. Using a relative risk instead of mean difference to extrapolate the difference between treatments in change from baseline weight from the secondary to the primary assessment.

   The ERG calculated a relative risk for the difference between treatments in change from baseline weight based on the secondary assessment (which was based on the ITC). The relative risk from the secondary assessment (instead of using the mean difference as done in the CS) was applied to the change from baseline weight at the primary assessment (instead of using the mean difference as done in the CS), see Section 5.2.6 for more details.

4. The natural history model to predict BMI is calibrated to reflect the baseline BMI distribution as observed in the COR trial programme.

   The patient characteristics in the COR trial programme were considered a fair reflection of the typical patient group that would receive NB32 in UK NHS clinical practice. To maintain consistency between effectiveness estimates and the population in which these were derived, the ERG preferred to reflect the baseline BMI distribution as observed in the COR-I (for non T2DM patients) and COR-DM (for T2DM patients) trials in the economic model (see Section 5.2.3 for more details). The calibration was performed using a minimisation of sum of squared error terms that was operationalized using Solver in Excel in two steps, and separately for T2DM patients and non-T2DM patients:

   a. Calibrate the constants of the natural history model to predict BMI (calibrated to reflect a mean BMI of 36 kg/m2, as observed in the COR trials).

   b. Calibrate the variance of the constants, to calibrate the distribution over the BMI groups (calibrated based on proportions in BMI categories, using the sum of squared differences compared with the COR-I/COR-DM trials).

5. Adjust the baseline age (dependent on T2DM status), proportions of females (dependent on T2DM status), proportion of smokers, proportion receiving statins (dependent on T2DM status), proportion receiving anti-hypertensive medication (dependent on T2DM status) and proportion receiving aspirin.

   The ERG preferred to use baseline characteristics from the COR trial programme and stratified for T2DM status, if applicable (see Section 5.2.3 for more details).

6. Removal of GP visit for standard management.

   The GP visit for the 52 week assessment for patients receiving standard management only (in addition to the five GP visits during the first year, which was considered to be reflective of UK clinical practice) was removed (see Section 5.2.9 for more details).

Matters of judgment

7. Weight regain towards baseline BMI was assumed.

   The ERG noted that the company deviated from the assumption made by Ara et al.,[58] that patients would have regained weight to obtain the baseline BMI in three years and assumed instead that patients would have regained weight to obtain the predicted BMI in three years. The company did not provide justification for why this deviation was 'logical' and more plausible. To be consistent with Ara et al.,[58] the ERG preferred to assume weight regain to the baseline BMI in its base-case (see Section 5.2.2 for more details).

8. Remove linear scaling assumption for TTD of orlistat.

   The ERG believes that the linear scaling of TTD estimates for NB32 to obtain orlistat TTD may result in underestimating TTD for orlistat (see Section 5.2.6 for more details).

5.3.1 Deterministic ERG base-case

Given the flaws highlighted for the PSA, the ERG was restricted to doing a deterministic analysis using 1,000 patient profiles (as the maximum number of patient profiles was restricted to 1,000) to obtain the ERG base-case incorporating all abovementioned adjustments (see Table 5.21 for the BMI distribution sampled in the ERG base-case). The ERG did calculate the ERG base-case two times, each time based on different random numbers and a different set of sampled patients. The ERG base-case ICERs (deterministic) of NB32 compared with standard management and orlistat ranged between £9,813-£10,510 and £38,871-£45,694 per QALY gained respectively (see Table 6.1).

Table 5.21: BMI distribution in ERG base-case

	T2DM			No-T2DM		
	CS	**ERG**	**COR-DM**	**CS**	**ERG**	**COR-I**
BMI<30kg/m2	9%	0%	6%	8%	0%	2%
BMI ≥30 and ≤35 kg/m2	75%	35%	31%	75%	35%	38%
BMI ≥35 and <40 kg/m2	16%	62%	35%	18%	64%	37%
BMI ≥ 40 kg/m2	0%	2%	28%	0%	1%	23%

5.3.2 Additional exploratory analyses performed based on the ERG base-case

Additional sensitivity analyses were performed to examine the potential impact of the following alternative assumptions on the cost effectiveness estimates:

1. Using an instantaneous weight regain at the point of three years
2. Using a lower proportion (15%) of T2DM patients

The exploratory analyses indicated that using an instantaneous weight regain at the point of three years and a lower proportion (15%) of T2DM patients decreased the ICERs (Table 6.1).

5.3.3 Subgroup analyses performed based on the ERG base-case

Subgroup analyses were performed for patients with and without T2DM based on the ERG base-case. For patients with T2DM, NB32 was dominated by orlistat while the ICER versus standard management was £10,535 per QALY gained. In the subgroup without T2DM, NB32 compared with standard management and orlistat resulted in ICERs of £9,594 and £25,744 per QALY gained respectively (see Table 6.1).

5.4 Conclusions of the cost effectiveness section

The majority of the cost effectiveness searches in the CS were well documented and easily reproducible, and were carried out in line with the NICE guide to the methods of technology appraisal.

Reviewing the overall evidence, the ERG confirmed that there was no existing cost effectiveness model for NB32 for the current indication, and thus that development of a de novo model was necessary. The economic model described in the CS is considered by the ERG to meet the NICE reference case on most points. However, the analyses performed by the company were flawed (too low a number of sampled patient profiles, too low a number of PSA simulations and key parameters were not incorporated in the PSA) and deviated from the NICE reference case stating that probabilistic model

results are preferred. The company developed a de novo economic model using an individual-level approach, more specifically a discrete event simulation (DES). It was argued that an individual-level approach is better suited than a cohort-level approach to capture the chronic implications of both weight and weight-related health events in a heterogeneous group of overweight and obese patients. The DES model was implemented in Excel using the DICE principles and structure proposed by Caro.[63] In addition, the company used the economic evaluation by Ara et al.[58] (also an individual-level model) as a starting point, which is a Health Technology Appraisal (2012) comparing different pharmacological treatments for obesity. The model considered the following events:

- treatment discontinuation;
- development of T2DM;
- first cardiovascular event (either stroke or MI);
- second cardiovascular event (either stroke or MI) and;
- death.

The company base-case ICERs (deterministic) of NB32 compared with standard management and orlistat were £13,647 and £32,084 respectively. The deterministic sensitivity analyses performed by the company show that the most influential parameters are the parameters of the Tobit model for utilities and the discount rate for QALYs, as well as parameters related to the measures of relative efficacy from the ITC. These analyses, as well as the PSA performed by the company should be interpreted with extreme caution given the flaws highlighted above. Subgroup analyses performed by the company indicated that the ICERs (deterministic) of NB32 compared with standard management and orlistat were £14,797 and £72,069 per QALY gained respectively for T2DM patients and £15,339 and £28,291 per QALY gained respectively for non-T2DM patients.

The main issue with the company's model was its structure and its technical implementation which caused long run times (6 hours on average), and which caused the model to crash on multiple computers. This hampered the company's and the ERG's ability to perform an appropriate PSA and the ERG's ability to check the model's validity and perform further scenario analyses (other than those that were described in Section 5.3). It should be considered whether simpler approaches (e.g. an individual-level state transition model) would have been more appropriate to reflect this decision problem, given the gain in transparency and that it would have been possible to reflect the condition-specific events in such a model. An individual-level state transition approach would potentially resolve most of the validity issues (e.g. the fact that BMI was not accurately reflected at each time period).

Apart from that, numerous issues were identified by the ERG, the most important of which are summarised in Table 5.20. The ERG was able to adjust/correct some of these issues in its base-case. The ERG base-case ICERs (deterministic) of NB32 compared with standard management and orlistat ranged between £9,813-£10,510 and £38,871-£45,694 per QALY gained respectively. Subgroup analyses performed conditional on the ERG base-case, indicated that the ICERs (deterministic) of NB32 compared with standard management and orlistat were £10,535 per QALY gained and dominated respectively for T2DM patients and £9,594 and £25,744 per QALY gained respectively for non-T2DM patients. However, it should be noted that several issues remained unexplored (of which several were expected to be non-conservative, see Table 5.20) and thus the results should be interpreted in this context (i.e. with extreme caution). The interpretation and validity of the results are particularly hampered given that the company's model did underestimate TTD, did not incorporate behaviour modification interventions, bariatric surgery and re-treatment nor an updating event that was required to accurately reflect patients' expected quality of life and costs associated with resource use. As

discussed in Section 5.2.12, the fact that BMI development was not reflected in the model could significantly bias the results in favour of NB32.

The large variation around the ICERs when different random numbers and sampled patient profiles are used is of particular concern. In two different model runs of the ERG base-case, the ICER varied by as much as £7,000 per QALY gained. It is therefore the ERG's view that the company's model is of very limited value for the current decision problem and that results are to be interpreted with extreme caution.

In conclusion, given that the deterministic ERG base-case ICER of NB32 versus orlistat is estimated to range between £38,871 and £45,694 per QALY gained (based on different random numbers and different samples of patients), and the remaining issues/methodological flaws highlighted above, uncertainty around the cost effectiveness estimates of NB32 remains substantial.

6. IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

In Section 5.3 the ERG base-case was presented, which was based on various changes compared to the company base-case. Table 6.1 shows the ERG replication of the company base-case, the ERG base-case, the exploratory analyses and subgroup analyses performed by the ERG (conditional on the ERG base-case). Appendix 1 contains technical details on the analyses performed by the ERG.

Table 6.1: ERG base-case, exploratory and subgroup analyses

	Technologies	Total costs	Total QALYs	NB32 Incremental costs	NB32 Incremental QALYs	NB32 ICER (£/QALY)
ERG base-case 1*	NB32	£7,017	15.21			
	Orlistat	£6,275	15.20	£742	0.02	£45,694
	SM	£5,964	15.11	£1,053	0.10	£10,510
ERG base-case 2*	NB32	£7,188	15.08			
	Orlistat	£6,455	15.06	£733	0.02	£38,871
	SM	£6,141	14.97	£1,047	0.11	£9,813
Company's base-case	NB32	£7,563	15.44			
	Orlistat	£6,814	15.41	£749	0.03	£32,084
	SM	£6,519	15.36	£1,044	0.08	£13,647
ERG replication of company's base-case	NB32	£6,948	15.36			
	Orlistat	£6,219	15.33	£729	0.02	£34,994
	SM	£5,974	15.29	£973	0.06	£15,568
Exploratory analyses						
1) Using instantaneous weight regain at 3 years	NB32	£7,048	15.19			
	Orlistat	£6,311	15.17	£737	0.02	£37,947
	SM	£6,007	15.09	£1,041	0.10	£10,021
2) Lower proportion (15%) of T2DM patients	NB32	£5,740	15.55			
	Orlistat	£4,992	15.53	£748	0.02	£28,687
	SM	£4,702	15.45	£1,038	0.10	£10,013
Subgroup analyses						
3) Subgroup non-T2DM patients	NB32	£4,603	15.77			
	Orlistat	£3,844	15.74	£759	0.03	£25,744
	SM	£3,565	15.66	£1,038	0.11	£9,594
4) Subgroup T2DM patients	NB32	£12,213	14.08			
	Orlistat	£11,527	14.09	£686	0.00	Dominated
	SM	£11,173	13.98	£1,040	0.10	£10,535

*These results are due to random variation between different model runs.

7. OVERALL CONCLUSIONS

7.1 Statement of principal findings

The four main trials comparing NB32 to placebo are of high quality. However there are a number of limitations when applying them to clinical practice. There are very little data on ethnic groups relevant to the UK (particularly people from Asia) within the NB32 trials, therefore it is not possible to make any firm conclusions for that group. There are very few overweight as opposed to obese participants in the trials. The majority of the participants in the NB32 trials are female. Trials do not measure weight loss beyond 56 weeks. The large dropout from the NB32 trials (up to 50%) is relevant to practice. The US setting may reflect a different patient profile and differing approaches to standard care than in a UK setting.

A comparison between NB32 (plus standard management) versus intensive behaviour modification is missing. Furthermore, comparisons between NB32 and orlistat are based on indirect comparisons only.

The company used modified ITT data from NB32 trials, but this is misleading. The mITT population in the NB32 trials is very different from mITT populations in the orlistat trials. In the NB32 trials, 21.9% of patients receiving NB32 were randomised but excluded from the analyses against 1.6% of patients receiving orlistat.

Comparison with orlistat may be biased in favour of NB32. NB32 trials were published in 2010 or later; most of the trials with orlistat were published before 2005, so caution should be exercised when making indirect comparisons; this is particularly true for conditions such as diabetes where background standard therapy (for glucose and lipids especially) may be very different now.

We have reproduced the company's indirect analyses comparing orlistat and NB32 using full ITT data from the NB32 trials and we have included a new analysis: an indirect comparison of NB32 plus intensive behaviour modification (COR-BMOD) versus orlistat plus intensive behaviour modification (XENDOS). The results show that the positive effects of NB32 when compared to orlistat have all disappeared. For the first outcome (≥5% reduction in weight at one year), there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded using mITT data. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: OR = 1.09 (95% CrI: 0.87 to 1.36), ITT-BOCF: OR = 1.06 (95% CrI: 0.84 to 1.33). Moreover, although none of the differences are statistically significant, all results now favour orlistat.

For the second outcome (mean percentage weight change at one year), there was a statistically significant difference favouring NB32 over orlistat in the analyses where studies with T2DM patients were excluded using mITT data. In both ITT analyses there is no significant difference between NB32 and orlistat for studies with T2DM patients excluded (ITT-Imp: MD= -0.09 (95% CrI: -0.77 to 0.58), ITT-BOCF: MD = -0.54 (95% CrI: -1.21 to 0.12). Moreover, although most of the differences are not statistically significant, most results now favour orlistat.

The results of the indirect comparison of NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification, using data from COR-BMOD versus XENDOS, show that both outcomes significantly favour orlistat over NB32 (≥5% reduction in weight at one year: OR 1.86 (95% CI: 1.30 to 2.66); mean percentage weight CFB at one year: MD -2.09 (95% CI: -3.53 to -0.65)).

Finally, we performed our preferred analyses, i.e. using full ITT data and no pooling of NB32 trials (using only COR-I ITT data for non-diabetics, instead of COR-I, COR-II and COR-BMOD combined).

The results for 'obese patients with T2DM' and 'intensive behaviour modification' are the same as before, but results for 'obese patients without T2DM' have changed considerably again, and are almost the same as in the company's original analyses. Both outcomes show no significant difference between NB32 and orlistat, but both favour NB32 in this subgroup.

The table below shows the main results for obese people with diabetes, obese people without diabetes and NB32 plus intensive behaviour modification versus orlistat plus intensive behaviour modification.

Table 7.1: Company results versus ERG results

Population		Company analyses (mITT data)*	Company analyses (ITT-BCFA data)**	ERG preferred analyses**
		Orlistat vs NB32	Orlistat vs NB32	Orlistat vs NB32
Obese people with T2DM				
≥5% reduction in weight at 1 year	OR	1.09 (0.63 to 1.88)	1.59 (0.89 to 2.79)	1.59 (0.89 to 2.79)
Mean % weight CFB at 1 year	MD	0.21 (-0.87 to 1.30)	-1.21 (-2.30 to -0.11)	-1.21 (-2.30 to -0.11)
Obese people without T2DM				
≥5% reduction in weight at 1 year	OR	0.77 (0.61 to 0.96)	1.06 (0.84 to 1.33)	0.61 (0.31 to 1.22)
Mean % weight CFB at 1 year	MD	1.13 (0.44 to 1.80)	-0.54 (-1.21 to 0.12)	1.11 (-0.39 to 2.63)
Intensive behaviour modification				
≥5% reduction in weight at 1 year	OR	1.22 (0.84 to 1.77)	1.86 (1.30 to 2.66)	1.86 (1.30 to 2.66)
Mean % weight CFB at 1 year	MD	-0.21 (-1.28 to 1.70)	-2.09 (-3.53to -0.65)	-2.09 (-3.53to -0.65)

Results are OR with 95% CI/CrI for ≥5% reduction in weight at 1 year and mean difference (MD) with 95% CI/CrI for mean % weight CFB at 1 year.
An OR less than one favours NB32 over orlistat and a CI including 1 is not significant. A MD of >0 favours NB32 over orlistat and indicates greater % weight reduction and a CI including 0 is not significant.
*) Bayesian NMA (OR, 95% CrI) using mITT data; **) Using the Bucher method for indirect comparisons and ITT-BCFA data.
FE = fixed effect; ITT-BCFA = all randomised patients with baseline-carried-forward analysis; MD = Mean Difference; mITT = modified intention-to-treat analysis; NB32 = naltrexone 32mg plus bupropion; OR = Odds Ratio; T2DM = Type 2 diabetes mellitus;

Which of the estimates of treatment effect is more applicable to clinical practice depends on the definition of standard management. If individuals who are eligible for NB32 would also engage in a weight loss programme when prescribed NB32 then the so-called intensive behaviour modification estimate might be more applicable. If this is not the case, then an estimate excluding intensive behaviour modification might be more appropriate. Of course, the estimate of 1.06 (0.84 to 1.33) is based on pooling both the trials with and without intensive behaviour modification and it is therefore tempting to infer that this represents clinical practice, where some do and some do not engage in weight loss programmes. This must be regarded with caution for a number of reasons, which include uncertainty as to the precise proportion who would engage in a weight loss programme and the degree of resemblance between such a programme and the intensive behaviour modification in COR-BMOD. Furthermore, costs of such intensive behaviour modification would also need to be considered in the economic model.

With regards to the economic model, one issue stood out: the structure and technical implementation of the company's model caused long run times (6 hours on average), and caused the model to crash on multiple computers. This hampered the company's and the ERG's ability to perform an appropriate PSA and the ERG's ability to check the model's validity and perform further scenario analyses (other than those that were described below). It should be considered whether simpler approaches (e.g. an individual-level state transition model) would have been more appropriate to reflect this decision problem, given the gain in transparency and that it would have been possible to reflect the condition-specific events in such a model. An individual-level state transition approach would potentially resolve most of the validity issues (e.g. the fact that BMI was not accurately reflected at each time period).

Apart from this, the ERG identified numerous issues of which the most important ones are summarised in Table 5.20. The ERG was able to adjust/correct some of these issues in its base-case. The ERG base-case ICERs (deterministic) of NB32 compared with standard management and orlistat ranged between £9,813-£10,510 and £38,871-£45,694 per QALY gained respectively. Subgroup analyses performed conditional on the ERG base-case, indicated that the ICERs (deterministic) of NB32 compared with standard management and orlistat were £10,535 per QALY gained and dominated respectively for T2DM patients and £9,594 and £25,744 per QALY gained respectively for non-T2DM patients. However, it should be noted that several issues remained unexplored (some of which were expected to be non-conservative) and thus the results should be interpreted in this context (i.e. with extreme caution). The interpretation and validity of the results are particularly hampered given that the company's model did underestimate TTD, did not incorporate behaviour modification interventions, bariatric surgery and re-treatment nor accurately reflected patients' expected quality of life and costs associated with resource use. As discussed in Section 5.2.12, the fact that BMI development was not accurately reflected in the model (due to lack of an updating event or integration of the BMI function) could significantly bias the results in favour of NB32. The large variation around the ICERs when different random numbers and sampled patient profiles are used is of particular concern. In two different model runs of the ERG base-case, the ICER varied by as much as £7,000 per QALY gained. It is therefore the ERG's view that the company's model is of very limited value for the current decision problem and that results are to be interpreted with extreme caution.

In conclusion, given that the deterministic ERG base-case ICER of NB32 versus orlistat is estimated to range between £38,871 and £45,694 per QALY gained (based on different random numbers and different samples of patients), and the remaining issues/methodological flaws highlighted above, uncertainty around the cost effectiveness estimates of NB32 remains substantial.

7.2 Strengths and limitations of the assessment

The majority of searches for eligible studies in the CS were well documented and easily reproducible. Searches were carried out on a good range of databases and carried out in accordance with the NICE 2013 guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4. The strategies utilised recognised study design filters. Supplementary searches of conference proceedings and organisational websites, and the checking of references lists were undertaken by the company in order to identify additional studies not retrieved by the main searches.

Four good quality large RCTs for NB32 and 16 comparator trials were included in the submission. Analyses were presented for all patients and people with and without T2DM, including a large number of sensitivity analyses.

The economic model structure is similar to the assessment by Ara et al.[58], which is a Health Technology Appraisal report (2012) comparing different pharmacological treatments for obesity.

The main weakness of the CS was the use of mITT populations for the NB32 trials. These data overestimate the benefits of NB32 over placebo or orlistat when compared to the true ITT data.

The validity issues highlighted by the ERG, the technical implementation of the model, as well as the assumptions regarding TTD, lack of reflection of behaviour modification interventions, bariatric surgery and re-treatment, and inaccurate reflection of BMI hamper the interpretation and therefore question the validity of the results.

Furthermore, the ERG considers the model as unfit for purpose, due to its extremely long run times, the fact that it crashes on many computers, and the inability to perform PSA.

7.3 Suggested research priorities

An ongoing randomised trial in the US will be available in up to six years' time to provide data concerning the effect of NB32 on the occurrence of major adverse cardiovascular events (MACE) in overweight and obese adults with cardiovascular disease.[31] Further research will also be needed to ascertain the role of NB32 in patients who are overweight with comorbidities and patients of Asian ethnicity. Long term weight loss and maintenance should be investigated and any additional benefits of NB32 over and above intensive behaviour management clarified.

8. REFERENCES

[1] Orexigen. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity [ID:757]: Submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA)*, January 2017 [accessed 11.1.17]. 267p.

[2] Moody A, Neave A. *Health survey for England 2015: adult overweight and obesity [Internet]*: Health and Social Care Information Centre, 2016 [accessed 3.2.17] Available from: http://www.content.digital.nhs.uk/catalogue/PUB22610/HSE2015-Adult-obe.pdf

[3] Baker C, Bate A. *Obesity statistics: briefing paper 3336 [Internet]*: House of Commons Library, Feb 2016 [accessed 31.10.16] Available from: researchbriefings.files.parliament.uk/documents/SN03336/SN03336.pdf

[4] Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, et al. *Foresight. Tackling obesities: future choices – project report [Internet]*: Government Office for Science, 2007 [accessed 11.10.16] Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287937/07-1184x-tackling-obesities-future-choices-report.pdf.

[5] National Institute of Health and Care Excellence. *Obesity: identification, assessment and management. NICE clinical guideline 189 [Internet]*. London: NICE, 2014 [accessed 12.10.16] Available from: https://www.nice.org.uk/guidance/cg189/chapter/1-recommendations

[6] World Health Organisation (WHO). *Obesity and overweight: fact sheet [Internet]*, June 2016 [accessed 10.10.16] Available from: http://www.who.int/mediacentre/factsheets/fs311/en/

[7] National Institute for Health and Clinical Excellence. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity: Final scope [Internet]*. London: NICE, October 2016 [accessed 31.10.16]. 4p. Available from: https://www.nice.org.uk/guidance/GID-TAG486/documents/final-scope

[8] National Institute for Health and Care Excellence. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity [ID757]: clarification letter*. London: NICE, 2017. 10p.

[9] Orexigen. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity [ID757]: response to request for clarification from the ERG*, 2017. 55p.

[10] European Medicines Agency (EMA). *Assessment report for Xenical (Procedure number: EMEA/H/C/154/A-20/0057) [Internet]*, 2012 [accessed 26.10.16] Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000154/WC500134995.pdf

[11] National Institute for Health and Clinical Excellence. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity: Response to consultee and commentator comments on the draft scope [Internet]*. London: NICE, October 2016 [accessed 31.10.16]. 4p. Available from: https://www.nice.org.uk/guidance/GID-TAG486/documents/scope-consultation-comments-and-responses

[12] Public Health England. Measurement of obesity [Internet]. 2016 [accessed 13.10.16]. Available from: http://www.noo.org.uk/NOO_about_obesity/measurement

[13] Canadian Agency for Drugs and Technologies in Health. *CADTH peer review checklist for search strategies [Internet]*. Ottawa: CADTH, 2013 [accessed 17.7.13]. 3p. Available from: http://www.cadth.ca/en/resources/finding-evidence-is

[14] National Institute for Health and Care Excellence. *Single Technology Appraisal: specification for manufacturer/sponsor submission of evidence [Internet]*. London: NICE, 2012 [accessed 31.1.17]. 76p. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Specification-for-manufacturer-sponsor-submission-of-evidence-June-2012.doc

[15] Scottish Intercollegiate Guidelines Network (SIGN). Search Filters [Internet]. Edinburgh: Healthcare Improvement Scotland, 2015 [accessed 19.1.17]. Available from: http://www.sign.ac.uk/methodology/filters.html

[16] 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

[17] R Core Team. The R project for statistical computing [Internet]. 2016 [accessed 22.11.16]. Available from: https://www.R-project.org/

[18] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010;36(3):1-48.

[19] Haslam D. Obesity: current and emerging drug treatments. *Pharm J* 2015;294(7866):20068647.

[20] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.

[21] Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;10(4):325-37.

[22] Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;7(4):434-55.

[23] Woodward P. *Bayesian analysis made simple: an excel GUI for WinBUGS.* Boca Raton, FL: CRC Press, 2011.

[24] Dias S, Welton N, Sutton A, Ades A. *NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials [Internet]*, 2011 (Updated: September 2016) [accessed 30.9.16] Available from: http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%202Sep2016v2.pdf

[25] Greenway FL, Fujioka K, Plodkowski RA, Mudaliar S, Guttadauria M, Erickson J, et al. Effect of naltrexone plus bupropion on weight loss in overweight and obese adults (COR-I): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* 2010;376(9741):595-605.

[26] Wadden TA, Foreyt JP, Foster GD, Hill JO, Klein S, O'Neil PM, et al. Weight loss with naltrexone SR/bupropion SR combination therapy as an adjunct to behavior modification: the COR-BMOD trial. *Obesity (Silver Spring)* 2011;19(1):110-20.

[27] Apovian CM, Aronne L, Rubino D, Still C, Wyatt H, Burns C, et al. A randomized, phase 3 trial of naltrexone SR/bupropion SR on weight and obesity-related risk factors (COR-II). *Obesity (Silver Spring)* 2013;21(5):935-43.

[28] Hollander P, Gupta AK, Plodkowski R, Greenway F, Bays H, Burns C, et al. Effects of naltrexone sustained-release/bupropion sustained-release combination therapy on body weight and glycemic parameters in overweight and obese patients with type 2 diabetes. *Diabetes Care* 2013;36(12):4022-9.

[29] Nissen SE, Wolski KE, Prcela L, Wadden T, Buse JB, Bakris G, et al. Effect of naltrexone-bupropion on major adverse cardiovascular events in overweight and obese patients with cardiovascular risk factors: a randomized clinical trial. *JAMA* 2016;315(10):990-1004.

[30] Halseth A, Shan K, Walsh B, Gilder K, Fujioka K. Method-of-use study of naltrexone sustained release (SR)/bupropion SR on body weight in individuals with obesity (In Press) [Data on file provided with the company's submission]. 2016.

[31] Orexigen. A multicenter, randomized, double-blind, placebo-controlled study assessing the occurrence of major adverse cardiovascular events (MACE) in overweight and obese subjects with cardiovascular risk factors receiving naltrexone SR/bupropion SR [Data on file provided with the company's submission]. 11 May 2015

[32] Orexigen. *Preliminary synopsis: single-dose pharmacokinetic study in hepatic impairment: a phase 1, open-label, parallel study to evaluate the pharmacokinetics of a single oral dose of extended-release combination of naltrexone and bupropion in subjects with normal hepatic function or varying degrees of impaired hepatic function (Naltrexone HCl / Bupropion HCl prolonged release) [PDF provided with the company's response to clarification]*, 2016

[33] Orexigen. *Preliminary synopsis: single-dose pharmacokinetic study in renal impairment: a phase 1, open-label, parallel study to evaluate the pharmacokinetics of a single oral dose of extended-release combination of naltrexone and bupropion in subjects with normal renal function or varying degrees of impaired renal function (Naltrexone HCl / Bupropion HCl prolonged release) [PDF provided with the company's response to clarification]*, 2016

[34] Orexigen. *A multicenter, randomized, double blind, placebo controlled study comparing the safety and efficacy of two doses of naltrexone sustained release (SR)/bupropion sustained release (SR) and placebo in obese subjects. (Clinical Study Report: NB-301) [PDF provided with the company's submission]*, 2010

[35] Orexigen. *A multicenter, randomized, double-blind, placebo-controlled study comparing the safety and efficacy of naltrexone sustained release (SR)/bupropion sustained release (SR) and placebo in subjects with obesity participating in a behavior modification program (Clinical Study Report: NB-302) [PDF provided with the company's submission]*, 2010

[36] Orexigen. *A multicenter, randomized, double-blind, placebo-controlled study comparing the safety and efficacy of naltrexone sustained release (SR)/bupropion sustained release (SR) and placebo*

in obese subjects (Clinical Study Report: NB-303) [PDF provided with the company's submission], 2010

[37] Orexigen. *A multicenter, randomized, double-blind, placebo-controlled study comparing the safety and efficacy of naltrexone 32 mg sustained release/bupropion 360 mg sustained release and placebo in obese subjects with type 2 diabetes mellitus (Clinical Study Report: NB-304) [PDF provided with the company's submission]*, 2009

[38] Orexigen. *Naltrexone-bupropion (prolonged release) for managing overweight and obesity [ID:757]: Appendices to accompany company evidence submission. Single technology appraisal (STA)*, January 2017 [accessed 11.1.17]. 268p.

[39] Roche. *Summary of product characteristics (SPC): Xenical 120 mg hard capsules [Internet]*, 2009 [accessed 26.7.16] Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000154/WC500058428.pdf

[40] Orexigen. Summary of product characteristics (SPC): Mysimba (Naltrexone/bupropion) [PDF provided with the company's submission]. 2016.

[41] Astrup A, Carraro R, Finer N, Harper A, Kunesova M, Lean ME, et al. Safety, tolerability and sustained weight loss over 2 years with the once-daily human GLP-1 analog, liraglutide. *Int J Obes (Lond)* 2012;36(6):843-54.

[42] Bakris G, Calhoun D, Egan B, Hellmann C, Dolker M, Kingma I. Orlistat improves blood pressure control in obese subjects with treated but inadequately controlled hypertension. *J Hypertens* 2002;20(11):2257-67.

[43] Berne C. A randomized study of orlistat in combination with a weight management programme in obese patients with Type 2 diabetes treated with metformin. *Diabet Med* 2005;22(5):612-8.

[44] Broom I, Wilding J, Stott P, Myers N. Randomised trial of the effect of orlistat on body weight and cardiovascular disease risk profile in obese patients: UK Multimorbidity Study. *Int J Clin Pract* 2002;56(7):494-9.

[45] Derosa G, Mugellini A, Ciccarelli L, Fogari R. Randomized, double-blind, placebo-controlled comparison of the action of orlistat, fluvastatin, or both an anthropometric measurements, blood pressure, and lipid profile in obese patients with hypercholesterolemia prescribed a standardized diet. *Clin Ther* 2003;25(4):1107-22.

[46] Derosa G, Maffioli P, Salvadeo SA, Ferrari I, Gravina A, Mereu R, et al. Comparison of orlistat treatment and placebo in obese type 2 diabetic patients. *Expert Opin Pharmacother* 2010;11(12):1971-82.

[47] Gotfredsen A, Westergren Hendel H, Andersen T. Influence of orlistat on bone turnover and body composition. *Int J Obes Relat Metab Disord* 2001;25(8):1154-60.

[48] Karhunen L, Franssila-Kallunki A, Rissanen P, Valve R, Kolehmainen M, Rissanen A, et al. Effect of orlistat treatment on body composition and resting energy expenditure during a two-year weight-reduction programme in obese Finns. *Int J Obes Relat Metab Disord* 2000;24(12):1567-72.

[49] Kelley DE, Bray GA, Pi-Sunyer FX, Klein S, Hill J, Miles J, et al. Clinical efficacy of orlistat therapy in overweight and obese patients with insulin-treated type 2 diabetes: a 1-year randomized controlled trial. *Diabetes Care* 2002;25(6):1033-41.

[50] Lindgarde F. The effect of orlistat on body weight and coronary heart disease risk profile in obese patients: the Swedish Multimorbidity Study. *J Intern Med* 2000;248(3):245-54.

[51] Lucas CP, Boldrin MN, Reaven GM. Effect of orlistat added to diet (30% of calories from fat) on plasma lipids, glucose, and insulin in obese patients with hypercholesterolemia. *Am J Cardiol* 2003;91(8):961-4.

[52] Mathus-Vliegen EM, van Ierland-van Leeuwen ML, Bennink RJ. Influences of fat restriction and lipase inhibition on gastric emptying in obesity. *Int J Obes (Lond)* 2006;30(8):1203-10.

[53] Miles JM, Leiter L, Hollander P, Wadden T, Anderson JW, Doyle M, et al. Effect of orlistat in overweight and obese patients with type 2 diabetes treated with metformin. *Diabetes Care* 2002;25(7):1123-8.

[54] Reaven G, Segal K, Hauptman J, Boldrin M, Lucas C. Effect of orlistat-assisted weight loss in decreasing coronary heart disease risk in patients with syndrome X. *Am J Cardiol* 2001;87(7):827-31.

[55] Swinburn BA, Carey D, Hills AP, Hooper M, Marks S, Proietto J, et al. Effect of orlistat on cardiovascular disease risk in obese adults. *Diabetes Obes Metab* 2005;7(3):254-62.

[56] Torgerson JS, Hauptman J, Boldrin MN, Sjostrom L. XENical in the prevention of diabetes in obese subjects (XENDOS) study: a randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care* 2004;27(1):155-61.

[57] National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 [Internet]*. London: NICE, 2013 [accessed 31.1.17]. 93p. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword

[58] Ara R, Blake L, Gray L, Hernandez M, Crowther M, Dunkley A, et al. What is the clinical effectiveness and cost-effectiveness of using drugs in treating obese patients in primary care? A systematic review. *Health Technol Assess* 2012;16(5):iii-xiv, 1-195.

[59] Davies MJ, Chubb BD, Smith IC, Valentine WJ. Cost-utility analysis of liraglutide compared with sulphonylurea or sitagliptin, all as add-on to metformin monotherapy in Type 2 diabetes mellitus. *Diabet Med* 2012;29(3):313-20.

[60] Burch J, McKenna C, Palmer S, Norman G, Glanville J, Sculpher M, et al. Rimonabant for the treatment of overweight and obese people. *Health Technol Assess* 2009;13 Suppl 3:13-22.

[61] Beaudet A, Palmer JL, Timlin L, Wilson B, Bruhn D, Boye KS, et al. Cost-utility of exenatide once weekly compared with insulin glargine in patients with type 2 diabetes in the UK. *J Med Econ* 2011;14(3):357-66.

[62] Copley V. *User guide: weight management economic assessment tool version 2 [Internet]*. Oxford: Public Health England, 2016 [accessed 11.8.16] Available from: http://www.noo.org.uk/visualisation/economic_assessment_tool

[63] Caro JJ. Discretely Integrated Condition Event (DICE) simulation for pharmacoeconomics. *PharmacoEcon* 2016;34(7):665-72.

[64] Orexigen. Clinical validation meeting for mysimba® NICE submission [Data on file provided with the company's submission]. 29 September 2016.

[65] Dare S, Mackay DF, Pell JP. Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PLoS One* 2015;10(4):e0123579.

[66] Gu Q, Burt VL, Dillon CF, Yoon S. Trends in antihypertensive medication use and blood pressure control among United States adults with hypertension: the National Health And Nutrition Examination Survey, 2001 to 2010. *Circulation* 2012;126(17):2105-14.

[67] Health and Social Care Information Centre. *Health survey for England 2013 (NS) [Internet]*, 2014 [accessed 2.3.17] Available from: http://content.digital.nhs.uk/catalogue/PUB16076

[68] National Institute of Health and Care Excellence. *Weight management: lifestyle services for overweight or obese adults. NICE public health guideline PH53 [Internet]*. London: NICE, 2014 [accessed 12.10.16] Available from: https://www.nice.org.uk/guidance/ph53

[69] Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value Health* 2011;14(4):539-45.

[70] MIMS. Orlistat [Internet]. 2016 [accessed 21.11.16]. Available from: http://www.mims.co.uk/drugs/nutrition/obesity/orlistat

[71] Curtis L, Burns A. *Unit Costs of Health and Social Care: 2015 [Internet]*. Canterbury: Personal Social Services Research Unit (PSSRU), 2015 [accessed 18.11.16] Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2015/index.php

[72] Department of Health. *NHS reference costs 2014-2015 [Internet]*. London: Department of Health, 2015 [accessed 22.9.16] Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015

[73] Ara R, Pandor A, Stevens J, Rees A, Rafia R. Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation. *Health Technol Assess* 2009;13(34):1-74, 75-118.

[74] Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, et al. Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. *Health Technol Assess* 2005;9(27):iii-iv, ix-xi, 1-158.

[75] Youman P, Wilson K, Harraf F, Kalra L. The economic burden of stroke in the United Kingdom. *PharmacoEcon* 2003;21 Suppl 1:43-50.

[76] Diabetes UK. Cost of diabetes [Internet]. [accessed 8.11.16]. Available from: http://www.diabetes.co.uk/cost-of-diabetes.html

[77] Diabetes UK. Number of people with diabetes reaches over 4 million [Internet]. 2016 [accessed 31.10.16]. Available from: https://www.diabetes.org.uk/About_us/News/Number-of-people-with-diabetes-reaches-over-4-million/

[78] World Health Organisation (WHO). Cardiovascular diseases (CVDs): fact sheet [Internet]. 2016 [accessed 11.10.16]. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/

[79] World Health Organisation (WHO). Types of cardiovascular disease [Internet]. 2004 [accessed 22.11.16]. Available from: http://www.who.int/cardiovascular_diseases/en/cvd_atlas_01_types.pdf

[80] Joint Formulary Committee. *British National Formulary: Adverse reactions to drugs [Internet]* London: BMJ Group and Pharmaceutical Press, 2016 [accessed 22.9.16]. Available from: http://www.evidence.nhs.uk/formulary/bnf/current/guidance-on-prescribing/adverse-reactions-to-drugs

[81] Davis S, Stevenson M, Tappenden P, Wailoo AJ. *NICE DSU technical support document 15: cost-effectiveness modelling using patient-level simulation [Internet]*, 2014 [accessed 7.3.17] Available from: http://www.nicedsu.org.uk

[82] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force 7. *Value Health* 2012;15(6):843-50.

[83] Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force 4. *Value Health* 2012;15(6):821-7.

[84] 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.

APPENDIX 1: DETAILS OF ERG ANALYSES (FOR VALIDATION PURPOSES)

Adjusted cells are printed in *Italics*

Fixing errors

1. Fixing errors consisted of using a weight regain period of 1.5 years after which weight is instantly regained, to reflect the three year linear weight regain assumption.
   Efficacy I 117

Fixing violations

2. Using the ITT data instead of mITT data; and based on the COR-I and COR-DM trials only
   Efficacy F35:M36, I53:56, I74, I88:92, AS6:BI10017
3. Using a relative risk instead of mean difference to extrapolate the difference between treatments in change from baseline weight from the secondary to the primary assessment.
   Efficacy I53:56
4. The natural history model to predict BMI is calibrated to reflect the baseline BMI distribution as observed in the COR trial programme.
   DICE equations W108, G118
5. Adjusting the baseline age (dependent on T2DM status), proportions of females (dependent on T2DM status), proportion of smokers, proportion receiving statins (dependent on T2DM status), proportion receiving anti-hypertensive medication (dependent on T2DM status) and proportion receiving aspirin.
   Controls J27:K42, DICE equations AX15, BA15
6. Removal of GP visit for standard management.
   Non-drug costs J90

Matters of judgment

7. Assuming weight regain towards baseline BMI instead of predicted BMI.
   DICE equations I 415, G:J415, G:J417, D419, D486, I486, D553, I553, D614, I614, D682, I682
8. Removing linear scaling assumption of TTD for orlistat.
   Treatment duration AO7:AP221,
   Range named td_first_response_lookup_int_b