in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Metreleptin for treating lipodystrophy

Produced by Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam and Maastricht University

Authors
Marie Westwood, Review Manager, KSR Ltd, UK
Nasuh Büyükkaramikli, Health Economics Researcher, Erasmus School of Health Policy & Management, EUR, the Netherlands
Isaac Corro Ramos, Health Economics Researcher, EUR
Regina Leadley, Systematic Reviewer, KSR Ltd
Rob Riemsma, Systematic Reviewer, KSR Ltd
Nigel Armstrong, Health Economics Manager, KSR Ltd
Marscha Holleman, Health Economics Researcher, EUR
Gill Worthy, Statistician, KSR Ltd
Janine Ross, Information Specialist, KSR Ltd
Maiwenn Al, Health Economics Researcher, EUR
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

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

Date completed 06/04/2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number GID-HST10011.

Declared competing interests of the authors
None.

Acknowledgements
The authors are grateful for the valuable comments on the DCE of Dr Esther de Bekker-Grob.

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:
Westwood M, Büyükkaramikli N, Corro Ramos I, Leadley R, Riemsma R, Armstrong N, Holleman M, Worthy G, Ross J, Al M, Kleijnen J. Metreleptin for treating lipodystrophy: a Highly Specialised Technology Evaluation. York: Kleijnen Systematic Reviews Ltd, 2018.

Contributions of authors
Marie Westwood 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. Nasuh Büyükkaramikli acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Isaac Corro Ramos, Marscha Holleman and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Regina Leadley and Rob Riemsma acted as systematic reviewer, 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. Janine Ross critiqued the search methods in the submission and contributed to the writing of the report. Maiwenn Al critiqued the company's economic evaluation, contributed to the writing of the report and provided general guidance on the health economics part of the project. 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

AACE	American Association of Clinical Endocrinologists
ADA	antidrug antibodies
AE	Adverse Events
AGL	acquired generalised lipodystrophy
ALT	alanine aminotransferase
APL	acquired partial lipodystrophy
AST	aspartate aminotransferase
BI	budget impact
BIC	Bayesian information criterion
BMI	body mass index
BSCL	Berardinelli-Seip congenital lipodystrophy
CE	Cost effectiveness
CEA	Cost effectiveness analysis
CEAC	Cost effectiveness acceptability curve
CFAS	controlled concomitant medication full analysis set
CGL	congenital generalised lipodystrophy
CHD	coronary heart disease
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company submission
CSR	Clinical study report
CUH	Cambridge University Hospitals
DCE	Discrete choice experiment
DM	diabetes mellitus
EAP	early access programme
EMA	European Medicines Agency
EMR	electronic medical record
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
ESRD	end-stage renal disease
EUR	Erasmus University Rotterdam
FAS	full analysis set
FDA	Food and Drug Administration
FFA	free fatty acid
FPL	familial partial lipodystrophy
GI	gastrointestinal
GL	generalised lipodystrophy
GPRD	General Practice Research Database
HbA_{1c}	glycated haemoglobin
HDL-C	high density lipoprotein cholesterol
HIV	human immunodeficiency virus
HRG	Healthcare resource groups
HRQoL	Health-related quality of life
HST	Highly specialised technologies
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio

ITT	Intention to treat
IV	Intravenous
KM	Kaplan-Meier
KSR	Kleijnen Systematic Reviews
LD	lipodystrophy
LDL-C	low density lipoprotein cholesterol
LOCF	Last observation carried forward
LS	least squares
LYG	Life year gained
LYS	Life year saved
MAA	marketing authorisation application
MedDRA	Medical Dictionary for Regulatory Activities
MI	myocardial infarction
MHRA	Medicines and Healthcare Products Regulatory Agency
MMRM	mixed-effect model repeated measures
MPGN	membranoproliferative glomerulonephritis
MRU	Medical resource utilisation
NA	Not applicable
NAFLD	non-alcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NR	Not reported
PCOS	polycystic ovary syndrome
PL	partial lipodystrophy
PRESS	Peer review of electronic search strategies
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PRO	Patient-reported outcome
PROMIS	Patient Reported Outcomes Measurement Information System
PSA	Probabilistic sensitivity analyses
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY(s)	Quality-adjusted Life Year(s)
QoL	quality of life
RCT	Randomised controlled trial
REMS	Risk evaluation management strategy
SAS	safety analysis set
SAE	Serious Adverse Events
SC	subcutaneous
SD	Standard deviation
SEM	standard error on the mean
SmPC	Summary of product characteristics
SoC	Standard of care
TEAEs	Treatment-emergent adverse events
UK	United Kingdom
UKPDS	UK Prospective Diabetes Study
ULN	upper limit of normal
USA	United States of America

Table of Contents

TABLE OF TABLES

TABLE OF FIGURES

1. SUMMARY

1.1 Background

The term lipodystrophy describes a heterogeneous group of rare disorders, which are characterised by a deficiency of adipose tissue (body fat) without underlying nutritional deprivation. Lipodystrophy syndromes are associated metabolic abnormalities, including diabetes mellitus (DM), hypertriglyceridemia and steatohepatitis, and with organ damage consequent upon ectopic lipid storage.

Lipodystrophy syndromes are categorised by aetiology (genetic or acquired) and by the distribution of adipose tissue deficiency (generalised, affecting the entire body, or partial). This results in four major categories: congenital generalised lipodystrophy (CGL), acquired generalised lipodystrophy (AGL), familial partial lipodystrophy (FPL) and acquired partial lipodystrophy (APL).

Congenital generalised lipodystrophy, also known as Berardinelli-Seip syndrome, is an autosomal recessive disorder with multiple genetic causes, which is characterised by an almost complete lack of body fat and prominent muscularity starting at birth or in infancy. Soon after birth, patients with CGL exhibit insatiable hunger and accelerated linear growth rates. Infants may also develop hepatosplenomegaly and umbilical prominence or hernia. Additionally, patients may have phlebomegaly and acanthosis nigricans later in childhood. Acquired generalised lipodystrophy, also known as Lawrence syndrome, is more common in females (female:male ratio 3:1) and appears usually before adolescence (but may develop at any time in life) with progressive loss of fat affecting the whole body including palms and soles of the feet. Familial partial lipodystrophy is a group of extremely rare, usually autosomal dominant disorders, characterised by loss of fat affecting the limbs, buttocks and hips. Patients also have fat accumulation in the face, neck, and intra-abdominal areas, causing a Cushingoid appearance. Acquired partial lipodystrophy, also known as Barraquer-Simons syndrome, usually has a childhood or adolescent onset and is more common in females (female:male ratio 4-5:1). APL is distinguishable from other lipodystrophy (LD) syndromes by the unique cephalocaudal progression of subcutaneous fat loss that is observed. Subcutaneous adipose tissue loss begins in the face and subsequently spreads to the neck, upper extremities, thorax and abdomen.

1.2 Summary of submitted evidence on the nature of the condition and the impact of the new technology

The company submission (CS) states that interviews with patients with lipodystrophy were conducted, at the US National Institutes of Health, on behalf of Aegerion, and that these interviews demonstrate the negative impact of lipodystrophy. This interview study is referenced as: 'Agerion Pharmaceuticals Ltd. Lipodystrophy patient research (NIH). Data on file. 2017.' Neither these data, nor a description of the interview study, were provided.

The CS includes the following statements, summarising the findings of the interview study:

- Patients are highly constrained by food access issues, impacting on many aspects of their daily lives including attending school, work and social situations. Patients also

suffer from mood and sleeping problems. The extreme level of food seeking additionally creates stress on families/carers. Carers may need to provide 24/7 supervision, especially as patients may also consume inappropriate or non-food items.

- Female lipodystrophy patients can suffer reproductive dysfunction as a result of leptin deficiency and severe insulin resistance. The adverse impact of reproductive dysfunction in females in the general population, including polycystic ovary syndrome (PCOS), infertility and miscarriage are well documented. For example, the spectrum of the symptoms of PCOS such as hirsutism, skin problems, menstrual problems and finally infertility has a huge negative impact on the individuals' psychological and interpersonal functioning. The interviews with patients with lipodystrophy confirm the impact of reproductive dysfunction in the context of lipodystrophy.
- Patients with LD can experience anxiety and depression due to the clinical burden of the disease including impaired physical appearance (which can be associated with bullying and low self-esteem), hyperphagia, reproductive dysfunction, fatigue and chronic pain.
- Other symptoms such as fatigue and frequent infection/illness, in addition to hyperphagia and anxiety/depression, can lead to impaired or complete inability to work or attend school, as well as to social isolation. In turn members of the family may not be able to work or socialise due to caring responsibilities.

The CS (pages 44-48) presents selected quotes from patients with lipodystrophy and their carers, in support of the above points.

The CS also states that: 'Metreleptin treatment is effective at improving metabolic abnormalities associated with LD, both in the short-term and long-term. Many of these changes have the potential to substantially improve the QoL of patients and their carers.'

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

The remit of this appraisal, as defined in the final agreed NICE scope, is to evaluate the benefits and costs of metreleptin within its licensed indication for treating lipodystrophy for national commissioning by NHS England.

At the time of submission of the ERG report, metreleptin did not have a marketing authorisation in the UK for the treatment of lipodystrophy. The latest available information (09/03/2018) is that:

[redacted]

[redacted]

The ERG notes some deviations from the final agreed NICE scope. Briefly, these include:

- The CS (section 12.1.2, page 153) states that the comparator for the cost effectiveness analysis was standard clinical management without metreleptin, as defined in the NICE scope, (including lifestyle modifications such as diet and exercise, use of lipid lowering drugs; and medications for diabetes). However, no data for the comparator were included in the clinical effectiveness section of the CS.
- The clinical effectiveness section of the CS focuses primarily on metabolic outcome measures; the CS includes no data or only very limited data for the clinical or patient-perceived outcomes specified in the NICE scope. No data are provided on liver cirrhosis, complications of diabetes, organ damage (including heart and kidneys) or effects on appearance. Mortality and pancreatitis are only reported where these are considered to be adverse effects of treatment or, in the case of pancreatitis, discontinuation of treatment.

The ERG recognises that no comparative studies of metreleptin versus standard care are available and that, in such cases, cost effectiveness analysis requires an indirect comparison between treatment and comparator studies. However, where indirect comparisons are used, it is essential that the same rigorous approach to identifying, selecting and reporting studies is applied for both intervention and comparator studies. There are serious problems with the identification, selection and reporting of comparator data in the CS. No systematic attempts to identify comparator studies and no selection criteria for such studies are reported. Parameters for the standard of care arm, in the cost effectiveness analysis, were informed by a single natural history study, which was not included in the CS.

The ERG has extracted additional data on clinical/patient-perceived outcomes from a short report of a follow-up study to the main study included in the CS, which was provided in response to clarification questions. This study was used in the cost effectiveness analyses, but was not included in the clinical effectiveness section of the CS.

1.4 Summary of clinical effectiveness evidence submitted by the company

Single arm, observational studies of metreleptin treatment found improvements in metabolic abnormalities from baseline to month 12 of treatment in patients with GL and in the subgroup of patients with PL who had similar metabolic disturbances to those seen in patients with GL (PL patients with leptin level <12 ng/ml with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L).

- In study NIH 991265/20010769, mean actual change in HbA_{1c} to Month 12/LOCF was -2.2% (95% CI: -2.7 to -1.6, p<0.001) for GL patients and -0.9% (95% CI: -1.4 to -0.4, p<0.001) for patients in the PL subgroup.
- In study FHA101, mean actual change from baseline to Month 12/LOCF for HbA_{1c} was -1.2% (95% CI: -4.3 to 2.0) for GL patients and -0.9% (95% CI: 95% CI: -1.4 to -0.4) for patients in the PL subgroup.

- In study NIH 991265/20010769, mean percent change in triglycerides to Month 12/LOCF was -32.1% (95% CI: -51.0 to -13.2, p=0.001) for the GL group and -37.4% (95% CI: -57.2 to -8.6, p<0.001) in the PL subgroup excluding the one outlying noncompliant patient.
- In study FHA101, mean percent change from baseline to Month 12/LOCF for triglycerides was similar in the GL group as -26.9% (95% CI: -124.1 to 70.4); however, for the PL subgroup, the mean percent change was lower at -8.5%. (95% CI: -36.4 to 19.5) Five of the seven patients in the PL subgroup in this study showed reductions from baseline to Month 12/LOCF in triglycerides ranging from -5.7% to -52.3%.
- Mixed model repeated measures (MMRM) analyses, from study NIH 991265/20010769, indicate that these effects persist to month 36.

With respect to safety and adverse events, the CS concludes that the known side effects of metreleptin can be managed as part of the normal clinical practice for patients with this complex condition. The ERG notes that the CS does not report the safety concerns as highlighted in the Centre for Drug Evaluation and Research Report (not included in the CS) nor the associated risk evaluation management strategy (REMS). The summary of safety in this report states: 'The principal safety concerns with metreleptin are T-cell lymphoma and anti-metreleptin antibodies with neutralizing activity. These concerns are of sufficient magnitude to require REMS. Other safety findings that warrant inclusion in the Warning and Precautions section of the metreleptin labelling include hypoglycemia, autoimmunity, and hypersensitivity.'

The clinical effectiveness sections of the CS did not include any results from control/comparator studies.

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

The company submission and response to clarification provided sufficient details for the ERG to appraise the literature searches. The CS states that a SLR was conducted to search for trials of metreleptin and trials of relevant comparators. However, the ERG is concerned that the search strategies did not contain any terms for comparators and only studies for the intervention will have been retrieved.

The key issue limiting the robustness of the efficacy data presented in the CS is the lack of any comparative studies; estimates of treatment effects are based on changes from baseline in single arm metreleptin treatment studies. This problem is compounded as the CS does not include any attempt to draw indirect comparisons through studies of the effects of established clinical management (diet, lifestyle modifications, lipid lowering drugs and anti-diabetic medications). The natural history study, used to provide comparator data for the cost effectiveness analysis, is not included in the clinical effectiveness sections of the CS and has a population which is not comparable to those included in the metreleptin intervention studies.

A further substantive issue concerns the nature of the treatment effects reported. The CS focuses primarily on changes in surrogate outcome measures (e.g. HbA_{1c}, triglycerides, hepatic enzymes) and includes very little information about any effects of treatment on patient-

perceived symptoms and clinical outcomes (e.g. hyperphagia, organ damage). Further data were available from a follow-up study, which was used in cost effectiveness modelling, but was not reported in the clinical effectiveness sections of the CS.

1.6 Summary of the evidence submitted to support the value for money of the treatment and cost to the NHS and PSS

The CS states that a systematic review was undertaken for economic, cost and resource use and health-related quality of life (HRQoL) evidence using a combined search for all of these areas. The cost effectiveness searches in the company submission were reported in enough detail for the ERG to appraise them. Three economic evaluation studies were identified by the company. However, none of these studies were eligible for inclusion in the review of economic evaluations of metreleptin, since the scope of all studies was not relevant to the CS.

A patient-level model was developed, aiming to assess the cost effectiveness of metreleptin versus standard of care (SoC) for patients with lipodystrophy. The model had a cycle length of one year and a time horizon of 60 years. A UK NHS PSS perspective was used in the model. Base case outcomes were incremental costs per quality-adjusted life-year (QALY) gained for metreleptin compared to standard of care. Both costs and effects are discounted at rate of 3.5%.

Two identical cohorts with 112 patients were used to populate the model. Individual patient data was obtained from the NIH follow-up study. Where individual patient data were not available, a Markov approach was used. A patient's survival probability is affected by abnormalities in a patient's heart, liver, kidney, or pancreas, i.e., the more organs with abnormalities, the higher the mortality for patients. Expected utilities and medical costs were based on the number of organ abnormalities. Health states were defined by the values of a set of attributes such as organ abnormalities, retinopathy, neuropathy, amputation, impaired physical appearance, hyperphagia, and female reproductive dysfunction. Metreleptin discontinuation was based on patient data or was assumed to be 2.05% per year when patient data were not available.

All patients in the NIH follow-up study were treated with metreleptin until death. A time-varying Cox proportional hazards model was used to estimate the relation between organ abnormality and mortality. Different parametric curves were fitted to the survival data from the trial, where the exponential curve showed the best fit.

Health utility estimates were derived from a discrete choice experiment (DCE) within the general population. These estimates were used to estimate QALYs associated with lipodystrophy.

Metreleptin is available in 11.3 mg vials (10 mg dose). However, the availability of smaller vial sizes (5.8 mg and 3 mg) is expected within the next three months. Given the anticipated availability of smaller vials, an average price per patient for metreleptin was assumed in the base case analysis. Resource use was based on questionnaires completed by two clinical advisers who treat lipodystrophy at Addenbrooke's Hospital. Health-state costs were based on

NHS reference costs. Only the cost of hypoglycaemic events was included in the model as adverse event.

Several assumptions were assessed in the sensitivity analysis, i.e., a price fall of 90% of metreleptin after 10 years, reduced initial price, elimination of mortality benefit of metreleptin for PL patients, changes to assumptions regarding organ abnormality progression, alternate survival extrapolation methods, and earlier treatment initiation. A deterministic one-way sensitivity analysis was conducted for the key clinical and economic variables in the model. A probabilistic sensitivity analysis (PSA) was also conducted.

When only 11.3 mg vials were included in the cost effectiveness analysis, the incremental cost effectiveness ratio (ICER) was £1,316,932 for metreleptin compared to SoC. The ICER was £671,927 per QALY gained for metreleptin compared to SoC when multiple vial sizes of metreleptin are available. When a PAS was applied to the scenarios of only 11.3 mg vials available and multiple vial sizes available, ICER yielded [redacted] and [redacted] per QALY gained respectively for metreleptin versus SoC. These values are higher than the thresholds used by NICE in HST appraisals.

1.7 Summary of the ERG's critique of the value for money evidence submitted

The ERG identified several critical issues with the company's economic analysis. Some of these issues were partially addressed in the revised electronic model submitted by the company in its response to the clarification letter. One of the most important concerns related to the organ impairment progression and matching methodology, which contributed directly or indirectly to a potential bias in favour of metreleptin treatment compared to SoC. The ERG requested that the company conduct *de novo* statistical analyses, in order to try to resolve these concerns. However, the company stated that they could not finalise this request given the timelines. There were also serious concerns surrounding the survival analyses conducted by the company and the implementation of these analyses in the model. There were several additional issues with the extrapolation of other attributes not related to organ damage and metreleptin discontinuation, which create potential bias in favour of metreleptin.

Overall, the ERG has serious concerns about the validity and reliability of the disutility weights reported by the company, and therefore considers these disutility weights to be speculative. The key concern is that the use of DCE to directly obtain disutility values for heath states is still in its infancy. The most striking issue relates to the fact that DCE classifies health states far more often below zero than TTO (time-trade-off) and produces lower average health state values. In addition, various major flaws in the design of the DCE and the analysis of the data were identified, leading to a negative assessment of the way QALYs are currently estimated.

The ERG also had several concerns about the resource use and costs included in the model. Furthermore, the ERG considered the validation of the model to be insufficient.

Given the many critical issues described above, it proved impossible for the ERG to give any indication on the cost-effectiveness of metreleptin. The uncertainty around the ICERs

presented by the company far exceeds that created by parameter uncertainty and reported in the CS.

1.8 Summary of the evidence submitted on the impact of the technology beyond direct health benefits and on the provision of specialised services

The CS includes a budget impact model to estimate the total costs to the NHS for a period of five years of adopting metreleptin for LD patients in the UK. The results presented by the company suggest that the net budget impact of implementing metreleptin will be £18,762,893 in year 1, and will rise to £34,114,350 in five years. The cumulative net budget impact over the first five years will be £133,045,965. Additionally, the estimated total number of LD patient eligible for metreleptin treatment after five years is 44 and the uptake of metreleptin rises from 85% in year 1 to 90% in year 5.

The CS also includes estimates of the impact of metreleptin on (i) inability to work or attend school for patients and carers; (ii) estimates of out-of-pocket costs for patients and carers including costs related to diabetes, transportation, fertility and cosmetic treatment; and (iii) other carer costs.

1.9 Summary of the ERG's critique on the evidence submitted on the impact of the technology on non-health-related benefits

In general, the assumptions made in the budget impact analysis could be considered as plausible. However, there are some concerns about the expected uptake rate of metreleptin. The ERG considers the high expected uptake rate as reliable, but the reason behind the rising uptake rate from 85% in year 1 to 90% in year 5 is unclear since the company did not provide further details on these assumptions. Furthermore, the validity of the estimated discontinuation rate provided by the company remains unclear since detailed information on these assumptions were also not provided.

The ERG has some concerns related to the impact of metreleptin beyond direct health benefits. No costs associated with inability to work or attending school were calculated in the analyses. However, as part of the NIH follow-up study, data on these attributes were collected. The ERG notes that, while there were data collected on these attributes, the company did not find it possible to estimate associated costs; the reason for this is unclear. The ERG also questions the assertion in which the company stated that metreleptin will mitigate the costs of hospitalisation and fertility and cosmetic treatments, since no evidence was provided to support this assertion. No indirect health care costs, due to additional life-years after receiving metreleptin, were reported in the CS and the company stated that they expected that these costs would not influence the cost effectiveness results. In the opinion of the ERG, these costs should be included in the model for completeness. Finally, the CS does not include costs related to informal care and productivity loss for the caregiver. The company states that it is currently conducting research to gain more details of these issues, but the ERG considers it to be inadequate that the impact of lipodystrophy on informal carers was not identified prior to the CS.

1.10 ERG commentary on the robustness of evidence submitted including strengths, weaknesses and areas of uncertainty

Strengths: The ERG believes that the following represent strengths within the CS:

- The company's submission provided sufficient details for the ERG to appraise the searches, which were on the whole clear, transparent and reproducible. An adequate range of resources were searched.
- Despite the rarity of LD syndromes, the company has presented data from a large, multinational study of metreleptin treated patients.
- The ERG considers that the budget impact model is generally based on plausible assumptions.

Weaknesses: The following are the main weaknesses of the CS, observed by the ERG:

- The CS lacks information about the long-term effects of metreleptin treatment.
- The CS (section 9.9.2, page 121) states that: 'Over 85% of the 107 patients in study NIH 991265/20010769 received >1 year of metreleptin, 72% received >2 years, 54% received >3 years, and 28% received 6 or more years of metreleptin in this study. The maximum duration of therapy was 14 years.' Despite this, the reporting of long-term clinical effectiveness outcomes, in the CS, was limited to information on the persistence (up to 36 months) of changes in HbA_{1c} and triglycerides on metreleptin treatment.
- Where long-term outcomes were available (in the NIH follow-up study, not included in the CS), these were either inferred from changes in surrogate outcome measures (e.g. hepatic enzymes, 24-hour protein excretion, blood pressure), or lacked any definition (e.g. hyperphagia recorded in notes).
- The CS lacks information about UK lipodystrophy patients; only one patient in the metreleptin treatment studies and one patient in the natural history study used in the cost effectiveness analysis, were UK patients.
- Despite the existence of an early access programme (EAP), which includes UK patients and has been running for more than 10 years, no results from the EAP were included in the CS and no justification/explanation for this was provided.
- The study details and results for the NIH follow-up study and the GL/PL natural history study, which were used to inform cost effectiveness modelling, were not included in the clinical effectiveness section of the CS.
- Participants in the NIH follow-up study and the GL/PL natural history study were not comparable and it is not clear that the matching exercise reported in the CS was adequate to account for the apparent differences.
- The clinical effectiveness section of the CS does not include any assessment of the comparative effectiveness of metreleptin vs. standard care (either direct or indirect).
- The process used to identify and select comparator/natural history studies remains unclear; the company's response to clarification questions stated that: 'The clinical SLR was carried out to search for trials of both metreleptin and trials of relevant comparators (see Section 9.1 of the submission).' However, the searches reported in the relevant sections of the CS were specific to metreleptin/leptin replacement interventions and did

not include any terms to search for comparator studies; these searches would not have reliably retrieved studies of comparator interventions or natural history studies.

- There are several concerns related to the estimation of organ impairment progression. Due to these issues, the ERG has substantial concerns about the appropriateness of the statistical methods used by the company. Therefore, the ERG requested that the company conduct de novo statistical analyses, however, the company stated that they were not able to finalise this request due to the given timelines.
- Serious concerns regarding the survival analyses conducted by the company and the implementation of these analyses in the model were identified.
- There were also several issues related to the matching methodology conducted by the company.
- The ERG considers the derivation of the utility decrement from the company's DCE as invalid.
- The ERG considers the validation of the model to be inadequate.

Areas of uncertainty: There is considerable uncertainty about the long-term effects of metreleptin treatment, particularly in relation to patient-perceived symptoms and clinical outcomes. The clinical effectiveness section of the CS includes only very limited evidence about patient perceived symptoms (hyperphagia) and clinical outcomes (liver damage) and data are limited to one year. The 'post-metreleptin improvements' reported in the follow-up study, but not in the CS, are frequently based on measures taken at one year and use definitions based on changes in surrogate outcome measures (e.g. improvement in liver abnormality is defined as 20% reduction in alanine transaminase/aspartate transaminase ratio (ALT/AST) at year one in a patient who had elevated ALT/AST at baseline) or provide no definition at all. The follow-up study also included some information on newly emergent (on metreleptin treatment) lipodystrophy characteristics in patients with no evidence of these characteristics prior to metreleptin initiation. However, no indication of the timeframe of observation was provided. Broadly, these data indicate that new incidences of organ abnormalities (liver, kidney and heart) and female reproductive dysfunction continue to occur, in all categories of LD patient, on metreleptin treatment. However, the data presented are insufficient to allow an adequate assessment of how the rate of development of new abnormalities on metreleptin treatment would compare with that seen in patients on standard care.

There remains some uncertainty regarding the long-term effects of metreleptin on metabolic measures. The CS includes some information on the persistence (up to 36 months) of changes in HbA_{1c} and triglycerides on metreleptin treatment, however, these data indicate that the apparent effect of metreleptin on triglyceride levels may not be applicable to the overall PL population. The potential effects of neutralising antibodies on the long-term efficacy of metreleptin treatment remain unclear. In clinical trials (studies NIH 991265/20010769 and FHA101) included in the CS, most patients (95%) developed antibodies to metreleptin. Overall, in patients where antibody data were available, neutralising anti-drug antibody activity was observed in 38/102 patients (37%) and, of these 38 patients, 58% achieved resolution of neutralising antibodies; these data were not linked to information about long-term efficacy or any withdrawals from treatment due to lack of efficacy.

The observed effects of metreleptin are all based on changes from baseline in single arm metreleptin treatment studies. The lack of comparative studies means that the extent to which any observed effects may be attributed to metreleptin remains unclear.

The significance of pancreatitis, as an adverse event following withdrawal from treatment, remains unclear. The CS describes six incidences of pancreatitis as an adverse event, across the 148 lipodystrophy patients in the two included studies. The mechanism for pancreatitis in these patients was presumed to be return of hypertriglyceridemia and therefore increased risk of pancreatitis in the setting of discontinuation of effective therapy for hypertriglyceridemia. With reported non-compliance rates of between 9% and 19% the extent of the pancreatitis risk, for these patients, remains unclear and would appear to warrant further consideration.

There is no mention in the CS of possible stopping rules for metreleptin. Given the many differences between and within groups of patients with different lipodystrophy syndromes, it cannot be expected that the treatment works equally well or even at all in all patients and the effectiveness of the treatment might diminish over time. Therefore, stopping rules should be considered.

Currently, only 11.3 mg vials of metreleptin are available. However, the company expects the availability of smaller vial sizes (i.e., 5.8 mg and 3.0 mg) within three months after submission.

1.11 Summary of exploratory sensitivity analyses undertaken by the ERG

The ERG identified some programming errors and corrected these programming errors to obtain a corrected version of the CS model. Even though these errors had a significant impact in total QALYs, the incremental results and ICER did not change significantly due to these corrections.

Given the many critical issues described earlier (Section 1.7), it proved impossible for the ERG to give any indication on the cost-effectiveness of metreleptin, thus no ERG base case was estimated. Based on the corrected company base case, the ERG conducted additional exploratory scenario analyses, challenging some of the structural assumptions of the model as well as some key input parameter choices.

It appears that the cost effectiveness results are very much dependent on the dosage assumptions of metreleptin (multiple vial size or single vial size), the treatment effect of metreleptin on disease attributes and utility input choice.

The ERG does not consider the cost-effectiveness model as reliable and trustworthy enough to inform decision making on the cost-effectiveness of metreleptin. The uncertainty around the company-reported ICERs is much larger than suggested by the PSA, which only addresses parameter uncertainty. However, the ERG still expects decision uncertainty to be rather low, as the ICER values, even in the best cases that the company presented, are significantly above the accepted thresholds.

2. BACKGROUND

2.1 Introduction

This chapter presents an overview of Lipodystrophy (LD) and its management. The content of this chapter is based on relevant literature, information provided by clinical advisors to the Evidence Review Group (ERG) and information presented in the background sections of the company's submission (CS).[1] For additional information on the aetiology, epidemiology, health impact, prognosis and management of LD, please see the CS (pages 32 to 61).

2.2 Description of health problem

2.2.1 Disease overview

The term lipodystrophy describes a heterogeneous group of rare disorders, which are characterised by a deficiency of adipose tissue (body fat) without underlying nutritional deprivation.[1, 2] Lipodystrophy syndromes are associated metabolic abnormalities, including diabetes mellitus (DM), hypertriglyceridemia and steatohepatitis,[1, 2] and with organ damage consequent upon ectopic lipid storage.[2]

Lipodystrophy syndromes are categorised by aetiology (genetic or acquired) and by the distribution of adipose tissue deficiency (generalised, affecting the entire body, or partial). This results in four major categories: congenital generalised LD (CGL), acquired generalised LD (AGL), familial partial LD (FPL) and acquired partial LD (APL).[1, 2]

Congenital Generalised Lipodystrophy

Congenital generalised lipodystrophy, also known as Berardinelli-Seip syndrome, is an autosomal recessive disorder with multiple genetic causes, which is characterised by an almost complete lack of body fat and prominent muscularity staring at birth or in infancy.[1-4] Soon after birth, patients with CGL exhibit insatiable hunger and accelerated linear growth rates. [1-3] Infants may develop hepatosplenomegaly and umbilical prominence or hernia.[3] Additionally, patients may have phlebomegaly and acanthosis nigricans later in childhood.[2, 3] A few patients develop DM during infancy, but development of DM most frequently occurs during the teenage years or later.[3] Diabetes, hypertriglyceridemia and hepatic steatosis can lead to the development of diabetic complications (nephropathy, neuropathy and retinopathy), recurrent attacks of acute pancreatitis, cirrhosis of the liver, and heart disease (cardiomyopathy, heart failure, myocardial infarction, arrhythmia), which are major causes of morbidity and mortality.[2, 3]

Acquired Generalised Lipodystrophy

Acquired generalised lipodystrophy, also known as Lawrence syndrome, is more common in females (female:male ratio 3:1) and appears usually before adolescence (but may develop at any time in life) with progressive loss of fat affecting the whole body including palms and soles of the feet.[1, 2, 5] The pattern and extent of fat loss in AGL is variable; most patients have generalised fat loss, but in a few cases some areas of the body (e.g. intra-abdominal and bone marrow fat) are spared.[3] As with CGL, AGL patients are highly likely to develop DM, hypertriglyceridemia and hepatic steatosis.[3, 4] Approximately 25% of AGL cases are associated with panniculitis (which presents clinically as subcutaneous inflammatory nodules), 25% with

autoimmune disease, and 50% are of idiopathic origin.[1, 3, 6] Autoimmune disorders that have been associated with AGL include juvenile-onset dermatomyositis, rheumatoid arthritis, systemic lupus erythematosus, and Sjögren syndrome.[1, 3, 7]

Familial Partial Lipodystrophy

Familial partial lipodystrophy is a group of usually autosomal dominant disorders, characterised by loss of fat affecting the limbs, buttocks and hips.[2, 3] The various forms of FPL are extremely rare.[1, 4] Numerous genetic mutations have been identified for FPL including the LMNA gene in familial PL type 2 (FPLD2).[1, 8]The most prevalent form of FPL is FPLD2, also known as the Dunnigan-Variety.[1, 4] FPLD2 develops during puberty, resulting in gradual atrophy of subcutaneous fat in the extremities followed by fat loss in the anterior abdomen and chest, giving the appearance of increased muscularity.[1, 4] Patients also have fat accumulation in the face, neck, and intraabdominal areas, causing a Cushingoid appearance.[1, 2, 9]Metabolic complications are common in adulthood,[10] with associated increased risk of heart disease.[11]

Acquired Partial Lipodystrophy

Acquired partial lipodystrophy, also known as Barraquer-Simons syndrome, usually has a childhood or adolescent onset and is more common in females (female:male ratio 4-5:1).[1, 2]APL is distinguishable from other LD syndromes by the unique cephalocaudal progression of subcutaneous fat loss that is observed.[1, 2, 4] Subcutaneous adipose tissue loss begins in the face and subsequently spreads to the neck, upper extremities, thorax and abdomen.[1, 2, 4] The lower extremities, lower abdomen and gluteal region do not exhibit lipoatrophy but rather accumulate excess adipose tissue.[1, 2, 4, 12] With the exception of hepatomegaly, metabolic complications are rarely seen in association with APL.[1, 12] APL is associated with autoimmune disease, particularly membranoproliferative glomerulonephritis (MPGN), in approximately 20% of cases.[2, 12]

2.2.2 Epidemiology

The CS states that there are limited published data available on the incidence and prevalence of LD in England. One recent study is cited, Chiquette et al. 2017,[13] however, the CS states that this study 'was not deemed accurate or generalisable for a UK population and the anticipated metreleptin licence.' Chiquette et al. used two approaches, one based on identification of cases from five electronic medical record (EMR) databases including the United Kingdom General Practice Research Database (GPRD), and one based on searches of the published literature, to estimate the prevalence of all LD.[13] The estimated worldwide prevalence of all LD, based on EMR database searches of four USA databases and the UK GPRD, was 3.07 cases/million (95% CI: 2.30 to 4.02).[13] No separate estimate was reported for the UK. The estimated European Union (EU) prevalence estimate, based on the total number of LD cases identified from searches of the published literature adjusted for underreporting and extrapolated to the total EU population, was 2.63 cases/million.[13] The study authors state that their estimates are at the lower end of the range of previously published numbers and that their approach may have underestimated prevalence.

The CS (CS, section 6.2, page 42-43) states that: 'More relevant and accurate estimates are available based on early access programme (EAP) data from a decade of metreleptin use in UK

clinical practice at Addenbrooke's. There are currently [redacted] LD patients receiving metreleptin at Addenbrooke's under the EAP – [redacted]. Of these patients, some may have initiated metreleptin over a decade ago since the beginning of the EAP. As the EAP has been running for over 10 years it is expected that the number of patients on the programme is a good indicator of the number of eligible patients in England. Clinicians from Addenbrooke's Hospital in England who are involved in the UK EAP have been consulted to provide an estimate of the number of new GL and PL patients each year who would be eligible for metreleptin. Based on expert clinical opinion, it is assumed that [redacted] new patients each year would be eligible for metreleptin treatment ([redacted]).' The estimates in table D58 (CS, page 199) give an indication of the expected number of UK patients who will be eligible for metreleptin treatment over the next five years, increasing from 26 in year 1 to 44 in year 5; these estimates were based on Addenbrooke's EAP data and expert opinion.

ERG comment: The CS estimates of the numbers of UK patients eligible for metreleptin treatment appear low when compared to published estimates of the prevalence of LD; the number of patients currently treated divided by the estimated total population for England and Wales 26/58.38 million gives an estimated prevalence of approximately 0.45 cases/million. The reason for this discrepancy is unclear. Given that only some of the patients in England and Wales, who have LD, are currently eligible for treatment with metreleptin under the UK EAP at Addenbrooke's Hospital:

'Recombinant leptin is specifically indicated for patients with severe LD and low leptin levels (<10 µg/L). The national service will select and treat patients with leptin as is clinically indicated. The cost of leptin is expressly excluded from the funding for this service.'[1]

It is possible that approval by NICE based on the licenced indication may result in a higher proportion of patients with LD being eligible/considered for metreleptin treatment. This is a particular concern if the licensed indication follows the outline suggested in the latest available information (09/03/2018) i.e [redacted] Whilst the EAP at Addenbrooke's Hospital and associated criteria for treatment are well established (>10 years duration), the ERG notes that there is uncertainty around the issue of future patient numbers.

2.2.3 Aetiology

Lipodystrophy syndromes can be inherited or acquired. Autosomal recessive CGL and autosomal dominant FPL are the two most common types of genetic LD. Mutations in the *AGPAT2*, *BSCL2*, *CAV1* and *PTRF* have been reported in patients with CGL, and mutations in *LMNA*, *PPARG*, *AKT2* and *PLIN1* have been reported in patients with FPL.[3] Acquired LD can be caused by autoimmune disease, drug or vaccine injections, and panniculitis; around 50% of acquired LD is of unknown origin.[3] An important sub-type of acquired LD occurs with prolonged exposure to protease-inhibitor-containing antiretroviral therapy in HIV-infected patients.[3]

ERG comment: The CS reports the exclusion of specific aetiologies of acquired LD (table C11, page 69 of the CS):

- HIV-associated LD
- LD secondary to drug administration (insulin growth hormone, steroids, antibiotics and vaccinations)
- LD secondary to systemic diseases such as uncontrolled diabetes mellitus, thyrotoxicosis, anorexia nervosa, malnutrition, malignancy and chronic infections

The scope issued by NICE does not exclude these sub-types of LD. Furthermore, the search strategies reported in the CS, for both clinical evidence (Appendix 1, page 220-223 of the CS) and economic evidence (Appendix 3, page 225-227 of the CS) included terms for HIV-associated LD.

2.2.4 Pathogenesis

Subcutaneous adipose tissue loss is a primary feature of LD, regardless of the sub-type and hence levels of the adipocyte-secreted hormone leptin are very low in these patients.[1, 14] Leptin promotes satiety (the feeling of feeling full), leading to decreased food intake,[1, 15] and also decreases gluconeogenesis in the liver and adipose tissue and increases glucose utilisation in skeletal muscle by activating signalling pathways which overlap with, but are not identical to, those of insulin.[1, 16] Leptin may also protect peripheral tissues from lipotoxicity by stimulating fatty acid oxidation.[1, 17] A deficiency in leptin can therefore result in insatiable hunger, increased gluconeogenesis and reduced fatty acid oxidation.[1] People with LD syndromes often have severe hypertriglyceridemia, with serum levels in the range of 1,000 mg/dL [11.29 mmol/L] compared with normal levels of 150 mg/dL [1.69 mmol/L]) being reported.[1, 18] The accumulation of ectopic fat throughout the body is associated with severe insulin resistance, resulting in the development of hyperglycaemia and HbA_{1c} levels consistent with a diagnosis of DM.[19, 20] These metabolic complications are drivers of the morbidity and mortality associated with LD syndromes.[2, 5]

2.2.5 Clinical features

Micro- and Macro-vascular complications

Elevated triglyceride levels have been found to be independently predictive of myocardial infarction, ischaemic heart disease and death, in large general population studies.[21] Two small studies have reported increased prevalence cardiovascular disease in patients with FPL, compared to unaffected family controls. One study reported atherosclerotic vascular disease in 12/39 (31%) of FPL patients compared to 6/45 (13%) of unaffected controls, however, it should be noted that rates of cigarette smoking were also higher in the FPL group, 13/39 (33%), than in unaffected controls, 9/45 (20%).[22] A second study, compared metabolic and clinical outcomes in *LMNA* mutation carriers with FPL and insulin resistance to matched family controls;[11] 8/23 (35%) of FLP patients had coronary heart disease (CHD), compared to 1/17 (6%) of controls, and all FLP patients had developed CHD before the age of 55 years.[11]

With respect to cardiomyopathy, a study of 44 GL patients reported that they found echocardiographic evidence of LV hypertrophy, as well as ECG abnormalities in 'more than half of patients,' with rates varying by type of GL.[23] This study also reported that, 'Although

cardiomyopathy was a frequent finding in our lipodystrophy patients, we found severe heart failure in only 2 patients.'[23] Review articles have reported that heart disease (cardiomyopathy, heart failure, myocardial infarction and arrhythmia) is a major cause of mortality in people with LD.[2, 3]

ERG comment: The CS tends to overstate the evidence about hypertriglyceridemia and heart disease in LD. For example, section 6.1.3.1, pages 37-38 of the CS, states: 'In the Copenhagen City Heart Study, which was initiated in 1976 and has followed 19,329 subjects, each 1 mmol/L increase in triglycerides is associated with a 40% increase in risk for myocardial infarction (MI), a 25% increase in risk for ischemic heart disease, and an 18% increase in risk of death in women, and 16%, 12%, and 10% increased risks, respectively, in men, when adjusted for age and HDL-C.'[1] These numbers are not reported in the cited study and are not consistent with the multifactorially adjusted hazard ratios (HRs) which are reported: For women these were 1.20 (95% CI: 1.05 to 1.37) for MI, 1.10 (95% CI: 0.99 to 1.21) for ischaemic heart disease and 1.18 (95% CI: 1.10 to 1.27) for total death; for men the corresponding values were 1.04 (95% CI: 0.98 to 1.11) for MI, 1.00 (95% CI: 0.95 to 1.06) for ischaemic heart disease and 1.08 (95% CI: 1.03 to 1.13).[21] It is also important to note that estimates of the risks associated with elevated triglyceride levels, which are derived from general population studies, should not be assumed to be directly transferable to patients with LD syndromes.

Renal failure and pancreatitis

Review articles have reported that patients with LD syndromes and hypertriglyceridemia and severe insulin resistance are pre-disposed to developing acute pancreatitis, cirrhosis, ESRD requiring renal transplantation and blindness due to diabetic retinopathy.[1-3] Chronic renal disease and membranoproliferative glomerulonephritis (MPGN) can occur in patients with GL and PL due to longstanding, suboptimally controlled DM.[1] Approximately one-fifth of patients with APL will develop MPGN, which can be fatal in some patients.[1, 3, 12] The CS (section 6.1.3.2 of the CS, page 38) states that, in the pivotal study NIH 991265/20010769, 31% of patients reported a history of pancreatitis (33 of 107).[1]

ERG comment: The CS (section 6.1.3.2, page 38) also includes the following statement: 'Additionally, one of the primary concerns with hypertriglyceridemia, especially when triglyceride levels exceed 1,000 mg/dL (11.29 mmol/L), is the risk for acute pancreatitis which can be life-threatening with a high mortality rate of 40% to over 50% when accompanied by complications like infection or organ failure.'[1] However, no reference is provided to support this statement.

Liver disease

Ectopic fat distribution in LD can lead to reduced liver function, and the development of cirrhosis and non-alcoholic fatty liver disease (NAFLD).[1, 2, 24] Liver failure, gastrointestinal haemorrhage, hepatocellular carcinoma have also been identified as causes of mortality amongst patients with LD.[2] An open-label, prospective study of metreleptin therapy in 27 patients with inherited and acquired forms of LD reported a reduction in mean NAFLD activity score, from 4.3 at baseline to 2.4 on treatment; patients who had fibrosis at baseline remained stable on treatment.[24]

A review of 63 cases of AGL from the literature and report of an additional 16 cases found hepatomegaly in approximately 72% of patients.[6] In this review, 50% of patients with AGL had elevated alanine aminotransferase (ALT) levels.[6]

ERG comment: The CS (section 6.1.3.3, page 38) incorrectly reports the results from the review of AGL patients, described above, as 84% with hepatomegaly and 60% with elevated ALT; a different study, by the same authors is erroneously cited.[12]

The CS (section 6.1.3.3, page 38) also states that: 'Non-alcoholic steatohepatitis (NASH) is highly prevalent in patients with LD, and there are no treatment options current available to treat this condition.' A study which makes no mention of NASH[9] is cited in support of this statement.

Hyperphagia

Low leptin levels act on the brain as a starvation signal, and therefore patients with LD can experience insatiable hunger and hyperphagia.[1, 25] As described above (section 2.2.4), hyperphagia due to leptin deficiency is also a key driver of morbidity associated with LD syndromes.[1] Patients with LD cannot store excess calories in their adipose tissue, and instead they are deposited as ectopic fat in the liver and muscle, causing severe insulin resistance, diabetes mellitus, hypertriglyceridemia, and steatohepatitis.[1, 4, 9]

Hyperphagia can also affect the management of LD. Dietary modifications are required to manage the metabolic complications of LD, however, dietary restriction may be challenging to achieve in some patients due to hyperphagia.[2, 25, 26] In addition, in children food restriction must be balanced by requirements for growth.[1, 2]

Fatigue and pain

Patients with LD syndromes may experience fatigue and pain due as part of their disease course. In a review of 16 case reports of patients with AGL treated at a single treatment centre in the US, patients presented with pain at diverse sites. While no quantitative data were gathered, pain was reported in knee joints, abdomen, calf muscle and skin by one patient each.[6] The case descriptions suggested that pain could be attributed to a number of different underlying causes. For example, one patient presented with pain in the calf muscle, which was suggestive of intermittent claudication.[6] Another patient developed painful skin lesions over her legs and thighs alongside abdominal pain.[6] An additional patient had pain in both knee joints, while loss of plantar fat in the feet was associated with the development of "painful" callosities, which limit movement.[6] In addition, one patient reported general fatigue.[6]

ERG comment: The scope issued by NICE[27] does not include pain in the list of specified outcomes and the clinical effectiveness section of the CS (CS, pages 67 to 123) does not include any evidence about effects of metreleptin treatment on pain.

Physical appearance

The partial and generalised loss of subcutaneous fat as well as abnormal fat distribution can have a marked effect on the physical appearance of patients with GL and PL. In CGL, patients may have prominent muscles, phlebomegaly, acanthosis nigricans, and umbilical prominence.[1,

[2] In AGL, patients may also have severe adipose tissue loss from the palms, soles, and intraabdominal area.[1, 4] The loss of subcutaneous adipose tissue in FPL can affect the appearance of the limbs, buttocks and hips. Additionally, excess fat accumulation, which varies by FPL subtype, may result in a Cushingoid appearance (including facial roundness).[1, 2] The distinguishing physical features of APL include cephalocaudal progression of fat loss, beginning in the face and subsequently spreading to the neck, upper extremities, thorax and abdomen.[1, 2] The CS includes anonymised patient photographs illustrating the morphology of generalised (Figure B3, page 40 of the CS) and acquired (Figure B4, page 41 of the CS) LD syndromes.[1]

Depression and neurological affects

The CS (section 6.1.3.7, pages 41-42) states that the disease course of LD may have negative consequences for patients' psychological health, and that physical dysmorphia, insatiable hunger and hyperphagia, infertility, fatigue and pain may contribute to depression in patients.[1] A 2016 practice guideline on the diagnosis and management of LD syndromes states that: 'Patients should be assessed for distress related to lipodystrophy and referred as necessary to mental health professionals and/or plastic surgeons.'[2]

Additionally, neurological deficits may also occur in GL and PL.[1] A 2017 systematic review reported rates of intellectual disability of 50% in patients with AGL, 47% in patients with CGL, 43% in patients with FPL and 8% in patients with APL, respectively.[5]

ERG comment: The CS (section 6.1.3.7, pages 41-42) also includes the statement: 'In a survey of LD experts in Europe, depression was considered to be of clinical importance and, anecdotally, occurs at a medium-high frequency amongst patients with GL and PL.' [1] An article about fertility and obstetric complications in women with FPL, which makes no mention of depression or anxiety, was erroneously cited in support of this statement.[1]

The scope issued by NICE[27] does not include depression or anxiety in the list of specified outcomes and the clinical effectiveness section of the CS (CS, pages 67 to 123) does not include any evidence about effects of metreleptin treatment on depression and anxiety.

Infertility and PCOS

Hypogonadotropic hypogonadism leading to delayed puberty, infertility, and abnormalities in the menstrual cycle, hirsutism and polycystic ovary syndrome (PCOS) in women, have are common in patients with LD syndromes.[2, 10, 28, 29]

A study comparing fertility and obstetric complications in women who had FPL due to LMNA to the general population and unaffected familial controls, found that 54% of the women with LMNA mutations exhibited clinical PCOS phenotypes, 27% had infertility, 50% experienced at least one miscarriage, 36% developed gestational diabetes and 14% experienced eclampsia and foetal death.[10] In the general population, 4.8% of women have PCOS, 10% have infertility, 10.1% experience at least one miscarriage, 5–10% have gestational diabetes and 2.6% experience eclampsia and foetal death.[10]

2.2.6 Diagnosis

Firm diagnostic criteria have not been established for LD.[2] The American Association of Clinical Endocrinologists (AACE) and a 17 member committee of nominees from worldwide endocrine societies have both attempted to develop consensus recommendations for the detection and diagnosis of LD.[2, 30]

The differential diagnosis should include conditions presenting with severe weight loss (malnutrition, anorexia nervosa, uncontrolled DM, thyrotoxicosis, adrenocortical insufficiency, cancer cachexia, HIV-associated wasting, chronic infections).[2] Differentiating between LD syndromes and uncontrolled DM is particularly difficult as both may present with extreme hypertriglyceridemia, however, restoration of glycaemic control in non-LD DM leads to restoration of body fat.[2] Generalised LDs can be confused with mutations of the insulin receptor or acromegaly, and FPL can be confused with Cushing's syndrome, truncal obesity and multiple symmetric lipomatosis.[2]

The multi-society practice guideline on the diagnosis and management of LD syndromes, which was published in 2016,[2] recommends that diagnosis be initially be based on history, physical examination, body composition and metabolic status, and further states that confirmatory genetic testing is helpful in suspected familial LD and should also be considered in at-risk family members.[2] The guideline also states that serum complement levels and autoantibodies may support the diagnosis of acquired lipodystrophy syndromes, and that there is no defined serum leptin level that can be used to establish a diagnosis of LD.[2] In patients with LD, the guideline recommends screening for comorbidities associated with the disease including diabetes, dyslipidaemia, NAFLD and cardiovascular and reproductive dysfunction.[2]

Differentiation of genetic and acquired LD can be hampered by the heterogeneity of subcutaneous adipose tissue loss between LD types. With CGL, patients typically have a lack of subcutaneous adipose tissue from infancy, whereas adipose tissue may appear as normal in infancy in patients with AGL.[2] The presence of autoimmune disease increases the suspicion of an acquired subtype.[1, 2]

AACE have conducted a MEDLINE literature search and panel discussion to inform their consensus statement on the detection of LD.[30] Although it does not have the structure of a guideline, the content of this statement is consistent with the practice guideline described above.

2.2.7 Prognosis

A recently published systematic review of the clinical features and management of non-HIV-related LD in children included 351 studies (including 219 case reports) of 1,141 patients; adult patients identified were excluded if the onset of LD had occurred after 18 years of age.[5] The review included 519 patients with CGL, 86 patients with AGL, 124 patients with FPL and 124 patients with APL.[5] The geographic distribution of the studies included in this review is not clear, however, the review does report some mortality data.

Of the 502 patients with CGL whose mortality status was known at the time of being reported (mean age at reporting, 12.6 years), 33 were dead; the mean age at death was 12.5 years (range, 0.4 to 46.0 years), with respiratory infection the most frequently reported cause of death, followed by cardiac failure.[5] Donohue syndrome resulted in a high mortality rate of 50% (21 of 42 patients dead at reporting) and a relatively early mean age at death (1.2 years; range, 0.03 to 8.3 years), with respiratory infection the most common cause.[5] Nine AGL patients were dead at the time of reporting and the mean age at death for these patients was 32.2 years, range 4.0 to 82.0 years.[5] For partial lipodystrophy, seven FPL patients were dead at the time of reporting and the mean age at death was 27.8 years (range 1.0 to 77.0 years), and three APL patients were dead at the time of reporting, with the mean age at death being 22.7 years (range 12.0 to 44.0 years).[5]

ERG comment: The CS (section 6.3, page 43) states that there are no natural history studies of LD patients in England (or the UK) to inform on the life expectancy of people with the disease in England. However, the CS does not present any search strategies used to identify natural history studies. In addition, no information was provided about survival/age at death for patients diagnosed during adulthood; it is likely that considering only patients diagnosed during childhood (as above) will result in lower estimates for mean age at death.

2.2.8 Impact on patients' health-related quality of life (HRQoL)

The CS (section 7.1, page 44) states that there is a paucity of published studies evaluating health-related quality of life (HRQoL) in patients with LD and their families. A literature review conducted to inform the CS (described in section 10.1.5, pages 132-135) identified one conference abstract reporting an evaluation of HRQoL in LD patients from the Lipodystrophy Connect Register, a global registry which collects self-reported data from both patients and care givers.[31] The study used a QoL questionnaire, which included items from the Patient Reported Outcomes Measurement System (PROMIS) short forms, questions on financial impact and impact of pain; 58/126 (48%) of participants responded to the QoL questionnaire.[31] Of the responders, 97% were female and 84% had partial LD.[31] EQ-5D scores were estimated from PROMIS global health items.[31] The estimated mean EQ-5D score for the LD syndromes population was 0.67, compared to a general population estimate of 0.866.[31] The abstract also noted that patients with LD syndromes reported some impairment in QoL on domains of physical health, mental health, social isolation and stigma, compared to the general population, however, no domain-specific data were presented.[31]

ERG comment: The CS also states that: 'Interviews with patients with LD conducted at the NIH in the US on behalf of Aegerion demonstrates the negative impact of LD.' (CS section 7.1, page 44). This statement is referenced as 'Aegerion Pharmaceuticals Ltd. Lipodystrophy patient research (NIH). Data on file. 2017.' These data were not provided; selected quotes from patients and carers are presented (CS: Figure B5, page 45; Figures B6 and B7, page 46; Figures B8 and B9, page 48).

2.3 Current service provision

The CS states that Aegerion are not aware of any NICE clinical guidelines, NICE pathways or published national guidelines on the management and treatment of LD. Metreleptin is the only

drug specifically for the treatment of LD. In the UK, treatment with metreleptin is currently provided, as part of an early access programme (EAP), under the National Severe Insulin Resistance Service at Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust. An overview of the NHS service specification (A03/S(HSS)/b)[32] is provided in the CS (CS, section 8.1.1, Table B9).

The CS states that: 'There is currently no standard clinical pathway for the treatment of LD in England.'[1] Standard care comprises an energy-restricted diet to lower triglycerides and glucose, which can be supplemented by treatments aimed at reducing complications such as DM (oral antidiabetic drugs including oral medications such as metformin, and injectable therapies including GLP-1 agonists in some patients and/or insulin) and hypertriglyceridemia (fibrates, statins).[1] Sections 8.1.2 and 8.2 of the CS (pages 58-61) provide a description of the various management options.

2.4 Description of the technology under assessment

Metreleptin is a leptin replacement therapy administered to address the effects of leptin deficiency in the population of LD patients with low leptin levels. It is a recombinant human leptin analogue produced in Escherichia coli cells by recombinant DNA technology to form recombinant methionyl-human leptin.[1, 33]

3. CRITIQUE OF THE COMPANY'S INTERPRETATION OF THE DECISION PROBLEM

3.1 Introduction

The remit of this appraisal, as defined in the final agreed NICE scope,[27] is to evaluate the benefits and costs of metreleptin within its licensed indication for treating lipodystrophy for national commissioning by NHS England. The final NICE scope outlines the agreed population, intervention, comparators and outcomes for the appraisal.[27] The NICE scope also sets out wider considerations relating to the impact of the technology beyond direct health benefits and on the delivery of the specialised service, the nature of the condition, costs to the NHS and PSS and value for money.

At the time of submission of the ERG report, metreleptin did not have a marketing authorisation in the UK for the treatment of lipodystrophy.

3.2 Adherence to the decision problem

Table 1 presents a summary of the decision problem as set out in the NICE scope[27] and the company's adherence to this (based on information presented on pages 19-23 of the CS).[1]

Table 1: Adherence to the agreed decision problem, as reported in the CS

	Final scope issued by NICE	Deviations of submission from the scope
Population	People with generalised or partial lipodystrophy	The original indication being sought from the European Medicines Agency (EMA) was as an adjunct to diet as a replacement therapy to treat the complications of leptin deficiency: • in patients with congenital or acquired GL, in adults and children 2 years of age and above • in patients with familial or acquired PL, characterised by leptin level <12 ng/ml with triglycerides ≥5.65 mmol/l and/or HbA_{1c} ≥6.5%, in adults and children 2 years of age and above uncontrolled on standard therapy Clinical efficacy and safety data from the clinical trials included a subgroup of PL patients related to the original indication, in addition to all eligible PL and GL patients. Of note, the definition of the PL subgroup and the age thresholds is currently under discussion in the regulator process and is likely to change prior to approval. The following indication is based on Day 180 questions: • in patients with congenital or acquired GL, in adults and children 6 years of age and above; • in patients with familial or acquired PL, characterised by leptin level <12 ng/ml with triglycerides ≥5.65 mmol/l and/or HbA_{1c} ≥8%, in adults and children 12 years of age despite optimized standard treatment
Intervention	Metreleptin	No deviations from scope
Comparator(s)	Established clinical management without metreleptin (including diet and lifestyle	No deviations from scope

	Final scope issued by NICE	Deviations of submission from the scope
	modifications, lipid lowering drugs and medications for diabetes)	
Outcomes	The outcome measures to be considered include: • Improvement in metabolic abnormalities • Liver function (including cirrhosis) • Glucose control and diabetes (including complications of diabetes and need for diabetes therapies) • Satiety • Pancreatitis • Use of other drugs • Organ damage including heart and kidneys • Growth and development • Reproductive dysfunction • Infection • Mortality • Adverse effects of treatment • Health-related quality of life (for patients and carers; including effects on appearance)	The outcome measures considered in the cost effectiveness assessment base case include: • improvement in metabolic abnormalities (e.g. triglycerides) • liver function (including cirrhosis) • glucose control and diabetes • satiety / hyperphagia • pancreatitis • organ damage to liver, heart and kidneys • reproductive dysfunction • mortality (linked to level of organ abnormalities) • adverse effects of treatment • Ability to perform school or work • health-related quality of life (for patients and carers; including effects on appearance) Other outcomes considered but not included in cost effectiveness assessment base case • improvement in other metabolic abnormalities (e.g. beyond triglycerides) • use / discontinuation of other drugs (including diabetes therapies such as insulin) • organ damage beyond liver, heart and kidneys • growth and development • infections • direct mortality benefit of treatment (e.g. beyond impact on organ abnormalities)

	Final scope issued by NICE	Deviations of submission from the scope
		• anxiety/depression • chronic pain and muscle spasms • complications of diabetes including retinopathy, neuropathy, and amputation (e.g. toes, limb) • impact on family and caregivers including ability to perform work • adverse effects of treatment • female infertility Potential adverse effects of treatment such as hypoglycaemia, the development of neutralising antibodies, and lymphoma were considered and their impact on patient preferences was assessed. However, due to the lack of robust information on their prevalence and the incremental role of metreleptin on their occurrence, their impact was not included in the base case cost effectiveness analyses.
Nature of the condition	• Disease morbidity and patient clinical disability with current standard of care • Impact of the disease on carer's quality of life • Extent and nature of current treatment options	No deviations from scope
Impact of the new technology	• Overall magnitude of health benefits to patients and, when relevant, carers • Heterogeneity of health benefits within the population • Robustness of the current evidence and the contribution the guidance might make to strengthen it	No deviations from scope
Cost to the NHS and Personal Social Services (PSS), and Value for Money	• Cost effectiveness using incremental cost per quality-adjusted life year	No deviations from scope

	Final scope issued by NICE	Deviations of submission from the scope
	• Patient access schemes and other commercial agreements • The nature and extent of the resources needed to enable the new technology to be used	
Impact of the technology beyond direct health benefits, and on the delivery of the specialised service	• Whether there are significant benefits other than health • Whether a substantial proportion of the costs (savings) or benefits are incurred outside of the NHS and personal and social services • The potential for long-term benefits to the NHS of research and innovation • The impact of the technology on the overall delivery of the specialist service • Staffing and infrastructure requirements, including training and planning for expertise	No deviations from scope
Other considerations	• If the evidence allows, subgroups according to whether the lipodystrophy is generalised or partial, or congenital or acquired, and according to the presence of complications associated with lipodystrophy (including diabetes and hypertriglyceridemia) will be considered • Guidance will only be issued in accordance with the marketing authorisation • Guidance will take into account any Managed Access Arrangements	Subgroups included in the model were identified based on the labelled indication. The following subgroups were included in the economic analysis: GL; PL; CGL; all NIH patients including those who do not meet the label indication
Related NICE recommendations and NICE Pathways	None	None

	Final scope issued by NICE	Deviations of submission from the scope
Related National Policy	NHS England. *Manual for Prescribed Specialised Services 2017/18. Chapter 62: highly specialist metabolic disorder services (adults and children), 2016 [Internet]*, 2017 [accessed 4.4.18]. 382p.[34] Available from: https://www.england.nhs.uk/wp-content/uploads/2017/10/prescribed-specialised-services-manual-2.pdf Department of Health. *The national service framework for long-term conditions [Internet]*. Leeds, 2005 [accessed 4.4.18]. 106p.[35] Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198114/National_Service_Framework_for_Long_Term_Conditions.pdf Department of Health. *NHS Outcomes Framework: at-a-glance [Internet]*, 2016 [accessed 4.4.18]. 5p.[36] Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/513157/NHSOF_at_a_glance.pdf	None

ERG comment: The latest available information (09/03/2018) is that: [redacted]

3.3 ERG critique of the company's adherence to the decision problem as set out in the NICE scope

3.3.1 Population

The population included in the clinical effectiveness sections of the CS relates to people with generalised and partial lipodystrophies.

A subgroup of the partial lipodystrophy population is also described (patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L). The CS describes this subgroup as related to the original EMA licenced indication, which was for adults and children over two years of age with CGL or AGL, and adults and children over two years of age with FPL or APL characterised by leptin levels <12 ng/ml with triglycerides ≥5.65 mmol/L and/or HbA_{1c} ≥6.5%, uncontrolled on standard therapy.

The CS (Table A1, pages 19-20) describes a further population of interest, based on EMA day 180 questions: adults and children aged six years and over, with CGL or AGL; adults and children aged 12 years and over, with FPL or APL characterised by leptin levels <12 ng/ml with triglycerides ≥5.65 mmol/l and/or HbA_{1c} ≥8%. The studies included in the clinical effectiveness section of the CS appear to have included GL patients <2 years of age and some patients in the PL subgroup with leptin levels >12 ng/ml, triglyceride levels <5.65mmol/ml and HbA_{1c} <6.5%. Five of the 66 GL patients included in the NIH 991265/20010769 were under six years of age and one was under two years of age, 40/66 (60.6%) of GL patients and 16/31 (51.6%) of PL subgroup patients had triglyceride levels <5.65 mmol/L, and 17/66 (25.8%) of GL patients and 2/31 (6.5%) of PL subgroup patients had HbA_{1c} <6.5%. None of the patients in the FH101 study were under six years of age, however, 6/9 (66.7%) of GL patients and 6/7 (85.7%) of PL patients had triglyceride levels <5.65 mmol/L, and 3/9 (33.3%) of GL patients and 1/7 (14.3%) of PL patients had HbA_{1c} <6.5%.[37, 38]

The clinical effectiveness section of the CS did not include any subgroup data for genetic and acquired LD syndromes.

ERG comment: The extent to which the population included in the clinical effectiveness sections of the CS is consistent with licenced indication for metreleptin remains unclear; at the time of submission of the ERG report, metreleptin does not yet have a UK licence for the treatment of LD syndromes. The latest available information (09/03/2018) suggests that: [redacted]

[redacted]

Of further note is the following information, provided in the company's response to clarification questions:[39] 'In NIH 991265/20010769 there was one patient from the UK (patient 901-026; 51 years, male, with AGL) who received metreleptin for 248 days (24/10/2003 to 27/06/2004). The patient was discontinued early because ineligibility was determined. Study FHA101 only included patients from the US. The NIH Follow-Up study also includes information for the same UK patient included in NIH 991265/20010769 (patient NIH-026). The Natural History study collected data for patients with lipodystrophy who were not treated with metreleptin at five locations: two in the US, one in Turkey, and two in Brazil (data collection in Brazil is ongoing). One patient from the UK, a female with APL diagnosed at age 42, was cared for at NIH and is included in the study.' This information raises concerns about the applicability, to the UK NHS, of information used in the CS.

3.3.2 Interventions

It is unclear whether the studies included in the CS describe metreleptin use in line with its licenced indication; at the time of submission of the ERG report, metreleptin does not yet have a UK licence for the treatment of LD syndromes.

In the CS (Table A2, pages 24-25), the recommended starting dose for metreleptin is reported as:

- Males and females ≤40 kg: 0.06 mg/kg
- (injection volume: 0.012 ml/kg)
- Males >40 kg: 2.5 mg (0.5 ml)
- Females >40 kg: 5 mg (1 ml)

With dose adjustments based on clinical response (e.g. inadequate metabolic control) or other consideration (e.g. tolerability issues, excessive weight loss especially in paediatric patients:

- Males and females ≤40 kg: maximum 0.13 mg/kg (0.026 ml/kg)
- Males >40 kg: maximum 10 mg (2 ml)
- Females >40 kg: maximum 10 mg (2 ml)

The recommended dosing frequency was once daily.

Participants in the studies included in the clinical effectiveness section of CS were treated with metreleptin, with the recommended dose ranges, given once daily or BID.

3.3.3 Comparators

The CS (section 12.1.2, page 153) states that the comparator for the cost effectiveness analysis was standard clinical management without metreleptin (including lifestyle modifications such as diet and exercise, use of lipid lowering drugs; and medications for diabetes). However, no data for the comparator were included in the clinical effectiveness section of the CS.

ERG comment: There are serious problems with the identification, selection and reporting of comparator data in the CS. No systematic attempts to identify comparator studies and no selection criteria for such studies are reported. Parameters for the standard of care arm, in the cost effectiveness analysis, were informed by a single natural history study, which was not included in the CS.

The company's response to clarification questions[39] states that: 'A review of the literature was conducted and leading lipodystrophy experts in the US, Brazil and Turkey were consulted.' However, no details of the search strategies used or inclusion/exclusion criteria for such a review were provided. In addition, it is unclear why only lipodystrophy experts in the US, Brazil and Turkey were contacted. The response to clarification questions[39] separately states that: 'The clinical SLR was carried out to search for trials of both metreleptin and trials of relevant comparators.' However, the search strategies described in section 17.1, appendix 1 of the CS[1] include lipodystrophy terms, which are combined with metreleptin terms using the AND function, i.e. these searches are not suitable for the identification of studies of the natural history of lipodystrophy syndromes or studies about interventions other than leptin replacement. In addition, the CS did not provide details of how unpublished studies were sought, for example was the UK treatment centre at Addenbrooke's Hospital approached form information? This information was requested in the clarification questions, but was not provided.

The company's response to clarification questions[39] included 23 spreadsheets and a document describing the natural history study.[40] The response to clarification questions includes the statement: 'Patients in the untreated sample were followed from birth while patients in the treated sample were first observed at the time of treatment. Additionally, two of the centers in the Natural History study also offered metreleptin treatment and appear to have preferentially selected patients with more severe symptoms for treatment. Therefore, the treated patients were, on average, at a more advanced stage of the disease at the start of observation compared to the untreated patients.' The baseline characteristics tables from the included metreleptin studies[37, 38] and the report of the natural history study[40] appear to support the view that patients in the treatment studies were at a more advanced stage of disease (see Tables 5, 6 and 8). However, the lack of clear information about which patients and results from the natural history study were used, in the CS means that it is impossible to adequately assess the extent to which it can provide a reliable comparison with data from the intervention studies.

The ERG recognises that no comparative studies of metreleptin versus standard care are available and that, in such cases, cost effectiveness analysis requires an indirect comparison between treatment and comparator studies. However, where indirect comparisons are used, it is essential that the same rigorous approach to identifying, selecting and reporting studies is applied for both intervention and comparator studies.

This is a major weakness of the CS which limits the interpretation of the available evidence.

3.3.4 Outcomes

The clinical effectiveness section of the CS focuses primarily on metabolic outcome measures; the CS includes no data or only very limited data for the clinical or patient-perceived outcomes specified in the NICE scope.[27] The protocols for both of the two studies included in the clinical effectiveness section of the CS list only metabolic and adverse events outcome measures;[37, 38] all other outcomes data appear to have been derived from publications of outcome data collected *ad hoc* by study investigators. No data are provided on liver cirrhosis, complications of diabetes, organ damage (including heart and kidneys) or effects on appearance. Mortality and pancreatitis are only reported where these are considered to be adverse effects of treatment or, in the case of pancreatitis, discontinuation of treatment.

3.3.5 Cost to the NHS and PSS, and value for money

The CS includes a cost effectiveness model in which the primary health outcome is valued in terms of incremental QALYs gained. In general, the scope was followed when assessing the costs of metreleptin to the NHS and the value for money it provides.

4 IMPACT OF THE NEW TECHNOLOGY – CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

4.1.1 Searches

Section 9.1.1 of the CS states that a systematic literature review was undertaken to search for trials of metreleptin and trials of relevant comparators. Search strategies were reported in detail in Appendix 17.1. The search was conducted on 10 March 2017. The selection of databases searched was adequate (Ovid Medline and Medline in Process, EMBASE, and the Cochrane Library Databases) and all searches were clearly reported and reproducible, the database name, database date span, and date searched was provided for the majority of the searches. The service provider used to search the Cochrane Library was not provided, and the strategy for this database appeared incomplete, however, a complete version was provided in the company's response to clarification questions. No language or date limits were applied and the searches were not limited by study design so would capture both RCTs and non-randomised studies.

Additional searches in key international HTA websites (limited to Europe only), a number of relevant conferences and clinical trials registries were also undertaken, however more specific details of these searches were not provided in the CS (i.e. search terms, website details and results retrieved).

Internal sources at Aegerion Pharmaceuticals were also used to source ongoing clinical studies and unpublished clinical study reports.

The ERG ran a test strategy to investigate recall from searching for epidemiology and natural history studies along with more sensitive terms for the condition. The search retrieved 1,540 results. More details of this can be found in Appendix 1.

ERG comment:

- The search strategies did not include any search terms for comparators. Only studies for the intervention metreleptin would have been retrieved, natural history studies may have also been missed.
- The search strategies were well constructed with condition and intervention facets and contained a combination of subject heading index and free text terms. The majority of subject heading terms were unnecessarily exploded but this would not impact on results retrieved. The ERG also notes that there were broad search terms used for endocrine disease.
- The ERG noted that there were some additional terms for the condition that could have been added to the strategies to increase sensitivity, such as disease acronyms (FLP, FPLD2 etc.). The inclusion criteria lists additional condition terms not used in the search strategies such as the rare lipodystrophy syndromes, Donohue Syndrome, Wiedermann Rautenstrauch syndrome and Berardinelli-Seip Syndrome.
- The ERG feels that a search of additional grey literature sources such as the FDA could have retrieved further information of value, particularly regarding safety information published by the FDA regarding metreleptin.

- The grey literature searches (CS Appendix 17.1.5) in the company submission did not provide full details of the search terms used, the precise date of the searches or the number of records. It's not clear if the company searched for the condition or intervention or both in these resources, the ERG cannot therefore comment on the robustness of these searches.

4.1.2 Inclusion criteria

The eligibility criteria for the review are described in Table 2 (CS, Table C11, pages 68-69). The inclusion criteria are generally broad and aim to include all relevant intervention studies. The main problem, as described in section 3.3.3 above, is that no systematic process is reported for the identification and selection of comparator studies. In addition, a number of exclusion criteria are listed for population (HIV-associated LD, LD secondary to drug administration, LD secondary to systemic diseases such as uncontrolled diabetes mellitus, thyrotoxicosis, anorexia nervosa, malnutrition, malignancy and chronic infections), which are not consistent with either the NICE scope.[27]

Table 2: Eligibility criteria

Inclusion criteria	
Population	Patients with congenital or acquired GL, in adults and children 2 years of age and above Patients with familial or acquired PL, characterised by leptin level <12 ng/ml with triglycerides ≥5.65 mmol/l and/or HbA_{1c} ≥6.5 %, in adults and children 2 years of age and above Patients with rare LD syndromes (e.g. Donohue syndrome, mandibuloacral dysplasia (type A and type B) and Wiedemann Rautenstrauch syndrome), in adults and children 2 years of age and above
Interventions	Studies considering an interventional treatment
Outcomes	Clinical outcomes, including (not limited to): distribution of fat (% fat loss across face and neck, abdomen, thorax, upper limbs and lower limbs and number of fat sparing across face and neck abdomen, upper limb, lower limb, palms and soles), menstrual irregularities (polycystic ovaries etc.), hirsutism, growth, treatment related adverse events and mortality associated with LD and comorbidities associated with underlying disease Metabolic outcomes, including (not limited to): blood glucose (fasting glucose mg/dl), serum insulin (insulin (uIU/ml), HbA_{1c} %, lipid profile (triglycerides mg/dl, total cholesterol mg/dl, HDL-C mg/dl and LDL-C mg/dl), liver function tests (AST U/L, ALT U/L), alkaline phosphatase (U/L), blood urea nitrogen (mg/dl), creatinine (mg/dl) and leptin (ng/ml) Metabolic complications, including (not limited to): diabetes, hypertriglyceridemia, insulin resistance and acute pancreatitis Quality of life outcomes if measured within the trial, including standardised and non-standardised outcomes

Inclusion criteria	
Study design	RCTs, non-RCTs (e.g. single arm trials, real world/observational studies), pooled analyses, retrospective analyses, long-term extension phase studies, systematic reviews/meta-analyses Ongoing clinical studies and unpublished reports available internally at Aegerion Pharmaceuticals (unpublished)
Language restrictions	None
Search dates	Journal articles, reports and summaries: No restrictions Conference abstracts published within the last four years (January 2013-January 2017, inclusive)
Exclusion criteria	
Population	HIV-associated LD LD secondary to drug administration (insulin growth hormone, steroids, antibiotics and vaccinations) LD secondary to systemic diseases such as uncontrolled diabetes mellitus, thyrotoxicosis, anorexia nervosa, malnutrition, malignancy and chronic infections LD in children <2 years of age
Interventions	Studies considering a non-interventional treatment
Outcomes	Studies reporting symptoms or short-term outcomes only Key search terms including: anatomy, histology, diagnosis, genetics, preclinical and reaction time
Study design	Phase 1 RCTs Study protocols Abstract with more recent existing full text publication Abstract or paper with insufficient reporting on population, study type or outcomes Healthy volunteer studies Animal studies Editorials/letters General reviews (other than systematic reviews)
Language restrictions	NA
Search dates	Conference abstracts published before 2013
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; HbA_{1c}, glycated haemoglobin; HDL-C, high density lipoprotein cholesterol; HIV,Human immunodeficiency virus; LD, lipodystrophy; LDL-C, low density lipoprotein cholesterol; RCT, randomised controlled trial	

4.1.3 Critique of data extraction

The CS states that the process of study selection was made according to specifications in the protocol.[41] The following statement about study selection and data extraction methods is given in appendix 1 (CS, section 17.1.7, pages 223 to 224): 'All abstracts were reviewed by two experienced systematic review researchers; any difference in opinion regarding eligibility was resolved through discussion, using a third reviewer if necessary. The same process was applied to the subsequent review of full papers. Data were extracted from eligible publications into pre-defined tables by a researcher and verified against the original source paper by a second

researcher.'[1] This statement was repeated 10 times in succession, but no further details (e.g. a list of items to be extracted) were provided.

ERG comment: Although not clearly reported in the main body of the CS, the data extraction process seems to have been performed using standard systematic review methodology.[42]

4.1.4 Quality assessment

Each included study was critically appraised using criteria which the CS states were 'adapted from Critical Appraisal Skills Programme (CASP): Making sense of evidence 12 questions to help you make sense of a cohort study.'[1] No reference was provided and the critical appraisal presented (CS, Tables C20 and C21, pages 88 to 90) included only seven questions. When assessing methodological quality, it is generally preferable to use a published, validated risk of bias tool, appropriate to the study design being considered. In this case, the new Cochrane tool for assessing risk of bias in non-randomised intervention studies (ROBINS-I)[43] would have been an appropriate choice or, alternatively, the Newcastle-Ottawa scale for assessing the quality of non-randomised studies[44] could have been used. Further problems were that no information was provided about the number of reviewers involved in the critical appraisal process. Table C20, critical appraisal of study NIH 991265/20010769,[37] was incomplete in the CS;[1] a corrected version was supplied in the company's response to clarification questions.[39]

Economic evaluations were assessed using a checklist adapted according to Drummond and Jefferson (1996).[45]

ERG comment: There was a lack of information about the quality assessment process and published, validated Risk of Bias tools were not used to assess studies included in the clinical effectiveness section of the CS.

4.1.5 Evidence synthesis

The CS does not include any information about synthesis methods, however, the protocol for the systematic review linked to the CS[41] includes the following statement: 'The review will consist of data extraction and a narrative synthesis. No formal statistical analysis is planned.'

ERG comment: The ERG agrees with this approach.

4.2 Critique of trials of the technology of interest, their analysis and interpretation

4.2.1 Studies included in/excluded from the submission

The systematic review conducted by the company identified 29 publications relating to metreleptin treatment for LD syndromes, which met the inclusion criteria listed in Table 2 above (CS, figure C14, page 70).

In total, the CS listed 16 publications relating to two eligible metreleptin interventional open label studies, the CSRs for which formed the basis of the clinical effectiveness section of the CS.[37, 38] The methodology and baseline participant characteristics for these two studies are described in detail in the CS (pages 73-85). Tables from the CS, describing study methods (Table 4) and baseline study characteristics (Tables 5 and 7), are reproduced below.

Study NIH 991265/20010769 (NCT00025883)[37] was an open-label, single-arm, investigator-sponsored study conducted at the NIH in the US between 2000 and 2014, with continuous enrolment and variable duration of follow-up; a follow-up study is ongoing.[46] The study aimed to investigate whether treatment with metreleptin could improve the metabolic sequelae, including pathological derangements in glucose and lipid homeostasis, found in patients with LD syndromes. Patients were enrolled from the US, countries in Europe including the UK, and other countries.

ERG comment: The response to clarification questions indicated that the CSR included only one patient from the UK.

Study NIH 991265 was a pilot, dose-escalation study to determine the safety and efficacy of short-term leptin replacement (up to eight months) and NIH 20010769 was conducted to determine the long-term safety and efficacy of metreleptin treatment for patients with LD. Study NIH 20010769 allowed for the rollover of patients from the pilot study, as well as for direct enrolment of new patients. Although conducted as separate studies, NIH 991265 and NIH 20010769 are treated, in the CS, as a single extended study since the two studies employed a similar protocol and all but one of the patients studied under the pilot study continued long-term treatment in the second study.[1, 47] Patients received self-administered or caregiver administered, subcutaneous metreleptin injections in one to two daily doses ranging from 0.06 to 0.24 mg/kg/day in study NIH 20010769 (0.01 to 0.08 mg/kg/day in study NIH 991265). Starting doses were dependent on age and gender, and doses were adjusted to achieve metabolic control and avoid excessive weight loss. Anti-hyperglycaemic and lipid-lowering regimens were modified if clinically indicated.[1, 37] The co-primary efficacy endpoints in this study were: actual change from baseline in HbA_{1c} at Month 12, and percent change from baseline in fasting serum triglycerides at Month 12.[1, 37] The study was conducted in the US where metreleptin was approved by the FDA in 2014. As of December 2014, all patients were either off metreleptin treatment or had transitioned to commercial product or free-drug programmes.[1, 37] The CSR for this study was based on all available data from the final integrated analysis on all patients (n=107) over the 14-year development period of metreleptin.[1, 37]

Study FHA101 was an open-label, expanded access study designed to provide metreleptin for the treatment of patients with diabetes mellitus and/or hypertriglyceridemia associated with LD. The study was initiated in 2008 in the US and all patients were enrolled from the US. As with study NIH991265/ 20010769, as of December 2014, all patients were either off metreleptin treatment or had transitioned to commercial product or free-drug programmes.[1, 38] Patients or caregivers injected metreleptin subcutaneously at 0.02 mg/kg twice daily (BID) for one week, modified to one month in June 2009, followed by 0.04 mg/kg BID.[1, 38] Dosage adjustments were allowed based on patient response. Dose titration up to 0.08 mg/kg BID was allowed if there were no improvements in metabolic parameters, and a reduction in target dose was permitted if tolerability became an issue. If metabolic parameters were stabilised after one year of treatment, then a decrease in dosing frequency from BID to once daily was allowed. Patients continued concomitant glucose-and lipid-lowering medications after the baseline visit, and further adjustments were permitted at the discretion of the treating physician.[1, 38] Patients

met with their treating physician one week after the first treatment and monthly for the first three months, followed by every three months throughout the first year. Following one year of treatment, patient visits were scheduled every six months or more frequently as deemed appropriate by the investigator. [1, 38]

The NIH991265/ 20010769 study included a much higher proportion of participants with GL, 66/107 (62%) than the FH101 study, 9/41 (22%). In study NIH 991265/20010769 the median age of the GL group was 15 years with 68% of patients <18 years of age; patients in the PL subgroup were older (median age 38 years) than those in the GL group, with 84% ≥18 years of age.[1] In study FHA101 most patients in both groups were ≥18 years of age at the time of enrolment.[1] In general, the baseline metabolic measures for patients in study FHA101 were not as elevated as those for patients in study NIH 991265/20010769 (see Tables 5 and 7 below).

Nine publications[48-56] were listed in Table C13 (CS, page 72) as 'excluded published studies.' The reason given for exclusion was: 'These studies were not included in the EMA (or the FDA) application; they only include a small number of patients and/or a population not relevant to this submission e.g. Japanese patients and/or PL patients who are not specific to the sought after indicated population.'[1]

ERG comment: The number of studies listed in tables C12 and C13 (CS, pages 71-72), does not match the total given in the PRISMA flow diagram (figure C14, CS, page 70). In addition, the exclusion of the studies listed in Table C13 (CS, page 72) is not consistent with the NICE scope[27] or with the pre-specified inclusion/exclusion criteria for the systematic review (see Table 2 above). The arbitrary exclusion of studies, based on small sample size, is particularly problematic in the context of summarising the evidence about an ultra-rare condition. Of particular note is the study by Simha et al. 2012,[50] which assessed the effects of leptin therapy in 24 female patients with Dunnigan variety FPL and moderate or severe hypoleptinemia and found no significant change from baseline to six months in fasting glucose, insulin, glucose tolerance, or HbA_{1c} levels.

The company provided a revised table of included/excluded studies in their response to clarification questions (Table 3, below). Although this table provides some further information on the reasons for excluding studies, it does not provide any reasons that are consistent with the pre-specified inclusion/exclusion criteria. The two publications, relating to one systematic review and listed in Table 3, were mentioned in the CS (section 9.2.2, page 69), but no references were provided; copies of the articles were not provided in either the CS or the response to clarification questions.

One included article, Oral et al. 2006[57] reported outcomes (circulating lymphocytes and cytokine response) which were not listed in the CSR for NIH 991265/20010796.[37]

Based on the PRISMA flow diagram (Figure C14, CS, page 71), 31 articles were excluded at the full text screening stage; details of these articles were not provided.

The CS does not include a description of the methods or baseline participant characteristics of the 'GL/PL natural history study', which was used to provide comparator data for the cost effectiveness modelling. A summary of the study protocol and baseline participant characteristics were provided in the company's response to clarification questions, and these are reproduced in Tables 8 and 9, below.[40] Table 9 provides details of those baseline participant characteristics that were also reported in the CS for the two metreleptin studies, NIH 991265/20010796 and FH101 or which were available from the NIH follow-up study,[46] (see Table 6). We have included these details in our report in order to allow a crude comparison to be made between the treatment studies included in the CS and the GL/PL study. As noted in the CS, participants in the GLPL natural history study had generally lower levels of HbA_{1c} and triglycerides than those in the metreleptin treatment studies. Of further note is the high proportion (approximately 50%) of participants in the GL/PL natural history study who were of Turkish ethnicity. The matching exercise outlined in section 17.6.2, Appendix 6, pages 270-271 of the CS, does not indicate that either ethnicity or baseline metabolic measures were considered when matching participants from the NIH follow-up study[46] to participants from the GL/PL natural history study.[40] Definitions of organ damage differed between the NIH follow-up study[46] and the GL/PL natural history study,[40] and the proportion of patients with liver, kidney or heart damage at baseline, or with a history of pancreatitis was generally lower in the GL/PL natural history study than in the NIH follow-up study. This may be because the metreleptin intervention study included patients who were at a later stage of LD than the GL/PL natural history study, where the baseline period is defined as the time before first GL/PL diagnosis.[40]

Table 3: Publications identified by the systematic literature view and their inclusion or exclusion in the submission (reproduced from the company's response to clarification questions

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Metreleptin studies				
NIH 991265/20010796 (NCT00025883)				
Oral et al. 2002[58] Full publication	Prospective, open-label, single arm (4 months)	Patients with various forms of LD (N=9)	To determine whether leptin replacement improves the insulin resistance, diabetes, and hypertriglyceridemia of patients with LD	Study NIH 991265/20010769 was used to inform the clinical effectiveness and safety of metreleptin. Overall 16 published studies relating to this study were identified in the SLR However, the studies were (mostly) not specifically described in the submission. They were published while the study was ongoing and thus report on fewer patients than in an integrated CSR, which has been provided by Aegerion. The integrated CSR includes data from 107 LD patients (GL=66; PL=41; PL subgroup=31) and therefore is more statistically robust than these individual studies. A follow-up to this study (NIH-follow-up study) was used to inform the economic model.
Petersen et al. 2002[59] Full publication	Case control (3-8 months)	Patients with severe GL (fasting leptin concentration less than 4 ng/ml) associated with diabetes (N=3)	To examine whether or not leptin treatment might improve insulin sensitivity in LD patients	
Javor et al. 2005a[60] Full publication	Prospective, open-label, single arm (12 months)	GL patients (N=15)	To determine the long-term effects of leptin replacement in a cohort of LD subjects	
Oral, et al. 2006[57] Full publication	Prospective, open-label, single arm (4-8 months)	Patients with various forms of LD (N=10)	To study lymphocyte subpopulations and in vitro peripheral blood mononuclear cell activation during a study evaluating the effects of leptin on metabolic functions in severe LD (serum leptin levels <4 ng/ml).	
Musso, et al. 2005[61] Full publication	Prospective, open-label, single arm (8-12 months)	Patients with various forms of LD (N=14)	(a) Investigated the role of recombinant leptin therapy on the hyperandrogenic state and menstrual dysfunction of patients up to 1 year of treatment; (b) evaluated the effect of	

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
			metreleptin on the growth hormone (GH) and insulin-like growth factor 1 (IGF-1) axis; (c) evaluated the pituitary-adrenal and thyroid axis over a 1-year period of metreleptin therapy; and (4) evaluated the effect of metreleptin therapy on the pituitary gonadal axis in a few male subjects to complement recent studies in male normal volunteers	
Park et al. 2007[62] Full publication	Prospective, open-label, single arm (12 months)	Patients with FPLD (N=6)	To investigate the role of low-dose recombinant leptin therapy in patients with FPLD to determine (1) the response of metabolic parameters to treatment, (2) the safety and tolerability of treatment over the long term, and (3) the differences of metabolic parameters at baseline and in response to treatment in patients with FPLD and GL.	
Chan et al. 2011[37] Full publication	Prospective, single-arm, open-label (12 months, but ongoing. Some patients have received up to 9 years of treatment up to July 2009 data cut)	Patients with acquired or inherited LD (N=55)	Evaluate the safety and effectiveness of leptin replacement therapy in patients with LD	

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Joseph et al. 2014[63] Full publication	Prospective, single-arm, open-label (24 months)	Patients with various forms of LD (N=82)	To study the effects of metreleptin in TGs and HDL in LD in contrast to changes in TGs and HDL in interventions for the obesity-associated metabolic syndrome	
Christensen et al. 2014[64] Full publication	Prospective, single-arm, open-label (96-120 months)	Patients with CGL (N=31)	To study the effects of metreleptin on bone mineral content and mineral metabolism	
Chong et al. 2010[65] Full publication	Prospective, single-arm, open-label (96 months: metabolic outcomes at 12 months reported)	Patients with GL or PL (acquired or inherited) (N=48)	To determine whether leptin replacement in LD patients ameliorates their metabolic abnormalities over an extended period of time and whether leptin therapy is effective in the different forms of LD	
Brown et al. 2013[66] Abstract	Prospective, single-arm, open-label (12 months but on-going; as of a July 2011 data cut, treatment duration was 2 month to 11 years including 64 patients treated for approximately	Patients with various LD subtypes (CGL, FPL, AGL, APL) (N=64)	To examine the effect of metreleptin on achieving commonly accepted therapeutic targets for HbA_{1c} and TG reduction at a 12-month treatment time point	

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
	12 month or more)			
Muniyappa et al. 2013[67] Full publication	Prospective, single-arm, open-label (16-20 weeks)	Congenital or acquired LD (N=13)	To examine the early effects (16–20 weeks) of leptin replacement on B-cell function in patients with LD	
Diker-Cohen et al. 2015[19] Full publication	Prospective, open-label, single arm (12 months, but ongoing. Some patients have received up to 9 years of treatment up to July 2009 data cut)	GL or PL (N=86)	Evaluate the safety and effectiveness of leptin replacement therapy in patients with GL and PL	
Moran, et al. 2004[68] Full publication	Prospective, open-label, single arm (12 months)	Patients with various forms of LD (N=14)	To determine the effect of leptin replacement therapy in patients with LD on (1) body composition, comprising changes in fat and lean body mass and (2) bone density and serum markers of bone metabolism. In addition, the effects on liver volume and resting energy expenditure were determined	The study by Moran was used in Section 9.6.1.4.4 Effect of metreleptin on hyperphagia (CS, page 99) "As reported by Moran and colleagues from the NIH, metreleptin treatment of 14 patients with LD (12 with GL and 2 with PL) dramatically decreased food intake at 4 months from 3,170 kcal/day to 1,739 kcal/day."

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Safar Zadeh et al. 2013[24] Full publication	Prospective, single-arm, open-label (Mean: 26 months; median 15 months, range 4–68 months)	Patients with GL or PL (N=27)	To study the spectrum of liver disease in LD and the effects of leptin replacement	The study by Safar-Zaheh was used in Section 9.6.1.4.3: Effect of metreleptin on hepatic enzymes, liver volume, and liver pathology (CS, page 98)
Javor et al. 2005b[69] Full publication	Prospective, open-label, single arm (Mean 6.6 [range: 4-18] months)	GL (8 patients) or FPLD (2 patients) (N=10)	To examine the prevalence of NASH in LD patients with steatosis and to assess the histological changes in the context of biochemical and radiographic changes seen with metreleptin therapy.	The results of the study by Javor were not specifically included in the submission; however it showed that metreleptin significantly reduced triglycerides, transaminases, hepatomegaly, and liver fat content. These reductions were associated with significant reductions in steatosis and the hepatocellular ballooning injury seen in NASH.
FHA101 (NCT00677313)				
Ajluni et al. 2016[70] Full publication	Prospective, single-arm, open-label (expanded access) (12 months)	Patients with PL and diabetes and/or hypertriglyceridemia with no pre-specified leptin level (N=23)	To determine the efficacy and safety of metreleptin among patients with PL using an expanded-access model	Study FHA101 was used, in the CS, as supportive evidence of the clinical effectiveness and safety of metreleptin. One publication relating to FHA101 was identified.[70] However, the study not specifically described in the submission. Instead the integrated CSR, provided by Aergerion was used. includes data from 41 patients (GL= 9; PL=32; PL subgroup=7)

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Metreleptin studies identified in the SLR but not included in the submission (with reason for exclusion)				
Beltrand et al. 2007[48] Full publication	Prospective, open-label, single arm (4 months)	Children with BSCL (N=7)	To test safety and efficacy of metreleptin treatment in children with BSCL before development of severe metabolic disease	Small sample size, short duration (4 months) study, only conducted in children (age range: 2.4-13.6 years)
Beltrand, et al. 2010[49] Full publication	Prospective, open-label, single arm (28 months)	Children with BSCL (N=8)	To assess the long-term efficacy and safety of leptin-replacement therapy to correct for the metabolic disorders.	Small sample size, only conducted in children (included 7 children from the above, short term trial).
Simha, et al. 2012[50] Full publication	A parallel group, open-label, observational study (6 months)	FPLD2 patients (N=24)	To compare efficacy of leptin therapy in FPLD patients with SH (serum leptin 7th percentile of normal) vs. those with moderate hypoleptinaemia (MH; serum leptin in 7th to 20th percentiles).	Small sample size only in patients with familial PL
Asthana, et al. 2015[51] Abstract	Prospective, open-label, single arm (16-32 weeks [4-8 months])	GL (N=9) or PL (N=8) (N=17)	To compare plasma angiopoietin-like protein 3 (ANGPTL3) and 4 in patients with LD and healthy controls and b) to examine the effects (16–32 weeks) of leptin replacement on ANGPTL 3 and 4	Small sample size, only an abstract (lack of information)
Brown, et al. 2015[52] Abstract	Non-randomised crossover study (19 days)	Previously leptin-treated (N=5, all GL, treatment duration 1-12y) and leptin-naïve (N=10, 9 PL) subjects (N=15)	To determine if leptin improves glucose and lipid metabolism in LD, independent of its effects on food intake.	Small sample size, only an abstract (lack of information)

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Ebihara, et al. 2007[53] Full publication	Prospective, open-label, single arm (36 months)	GL patients (Japanese) (N=7)	To evaluate the efficacy and safety of long-term leptin-replacement therapy on seven Japanese patients with generalised LD.	Small sample size in Japanese patients (i.e different ethnic population than expected in the UK)
Schlogl, et al. 2016[54] Full publication	Prospective, open-label, single arm (52 weeks [12 months])	Patients with GL or PL (N=9)	Resting state functional MRI scans and extensive behavioural testing assessing changes in hunger/satiety regulation were performed during the first 52 weeks of metreleptin treatment in nine patients with LD	Small sample size
Vatier, et al 2016[55] Full publication	Prospective, open-label, single arm (compassionate therapeutic programme) (12 months)	Patients with various forms of LD (N=16)	To evaluate the effect of metreleptin on insulin sensitivity and insulin secretion using dynamic IV clamp procedures in 16 patients with genetic LD syndromes, included in a compassionate therapeutic programme	Small sample size
Araujo-Vilar, et al. 2015[56] Full publication	Retrospective, open-label study, single arm (Median 3 years [range 9 months to 5 years, 9 months])	Patients with genetic LD syndromes (N=9)	To determine the effectiveness of recombinant methionyl leptin (metreleptin) for improving glucose metabolism, lipid profile, and hepatic steatosis in patients with genetic lipodystrophy syndromes	Small sample size

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Rodriguez, et al. 2014[71] Full publication	SLR and meta-analysis	LD not associated with the use of HIV protease inhibitors	A systematic review of the MEDLINE and Cochrane Library databases was conducted to identify studies assessing the effect of metreleptin on metabolic and hepatic endpoints of patients with lipodystrophy not associated with the use of HIV protease inhibitors	Systematic reviews were an inclusion criteria in the clinical SLR. Two publications from the same group reported the results of a systematic review and meta-analysis into the effects of metreleptin on metabolic and hepatic endpoints in patients with lipodystrophy syndromes not associated with the use of HIV protease inhibitors. In the full-text article by Rodríguez et al. 2014, 12 studies were included after full-text review of the papers identified in their literature search of Medline and the Cochrane library. All of these papers have been included in the current SLR reported here i.e. Beltrand et al. 2007 and 2010; Chan et al. 2011; Chong et al. 2009; Ebihara et al. 2007; Javor et al. 2005b; Moran et al. 2004; Oral et al. 2002; Park et al. 2007; Petersen et al. 2002; Safar Zadeh et al. 2013; and Simha et al. 2012. In the abstract by Paz-Filho et al. 14 studies were identified (the details were not reported). The results of the systematic review and meta-analysis were not considered relevant to the submission due to some limitations. In Rodríguez et al. a meta-analysis of results (N=226 patients across the studies)showed
Paz-Filho, et al. 2014[72] Abstract	SLR and meta-analysis	LD not associated with the use of HIV protease inhibitors	A systematic review of the MEDLINE and Cochrane Library databases was conducted to identify studies assessing the effect of metreleptin on metabolic and hepatic endpoints of patients with LD not associated with the use of HIV protease inhibitors	

Publication	Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
				that metreleptin decreased FPG (0.75 standardised mean differences [SMD] units [range 0.36-1.13], P = 0.0001), HbA_{1c} (0.49 [0.17-0.81], P = 0.003), triglycerides (1.00 [0.69-1.31], P < 0.00001), total cholesterol (0.62 [0.21-1.02], P = 0.003), liver volume (1.06 [0.51-1.61], P = 0.0002) and AST (0.41 [0.10-0.73] P =0.01). However, the review has several limitations, particularly that several of the studies from NIH 991265/20010796 were included individually but they may have included some of the same patients. In Paz-Filho et al. a meta-analysis of results from clinical studies in 243 patients showed that metreleptin decreased FPG [0.76 SMD units (range 0.40-1.12), P < 0.0001], HbA_{1c} [0.55 (0.23-0.86), P = 0.0006], triglycerides [1.12 (0.81-1.43), P < 0.00001], total cholesterol [0.62 (0.21-1.02), P = 0.003), liver volume [0.98 (0.52-1.43), P < 0.0001], liver fat [0.67 (0.44-0.89), P < 0.0001], ALT [0.44 (0.07-0.80), P = 0.02] and AST [0.45 (0.17-0.73) P = 0.002].

Publication		Study design (treatment duration)	Population	Objective	Inclusion or exclusion in the submission
Comparator study					
29	Dantas de Medeiros Rocha, et al. 2010[73] Full publication	Prospective, open-label, single arm	BSCL patients (N=10)	To evaluate the effect of diet intervention and oral zinc supplementation on the metabolic control of BSCL patients	This study was not considered suitable for the submission because oral zinc supplementation is not established clinical management for the treatment of LD, together with the study limitations i.e small sample size and short treatment duration.

Abbreviations: AGL = acquired generalised lipodystrophy; APL = acquired partial lipodystrophy; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BSCL = Berardinelli-Seip congenital lipodystrophy (also known as CGL); CGL = congenital generalised; CSR = clinical study report; FPG = fasting plasma glucose; FPL = familial partial lipodystrophy; FPLD = familial partial lipodystrophy, Dunnigan variety; GL = generalised lipodystrophy; HDL = high-density lipoprotein-cholesterol; IV = intravenous; LD = lipodystrophy; MRI = magnetic resonance imaging; MH = moderate hypoleptinaemia (serum leptin in 7th to 20th percentiles); NASH = non-alcoholic steatohepatitis; PL = partial lipodystrophy; Pts = patients; SD = standard deviation; SH = severe hypoleptinaemia (serum leptin 7th percentile of normal); SMD = standardised mean differences; TG = triglycerides

Table 4: Summary of study methods, reproduced from Table C15 (CS, pages 77-80)

Study name	NIH 991265/20010769
Objective	To evaluate the safety and efficacy of recombinant methionyl human leptin (metreleptin) replacement in patients with GL and PL
Location	The studies were conducted at the NIH, however patients were also enrolled from countries outside the US: GL: 59% were from the US; 20% from Europe/Eastern Mediterranean (Belgium, UK, Germany, Italy, Lithuania, Spain, Turkey, Albania, Israel, and Serbia); 18% from other countries.* PL: 78% from the US, 7% from Europe/Eastern Mediterranean; 15% from other countries*
Design	Open-label, single-arm
Duration of study	Continuous enrolment over 14 years (2000-2014): NIH 991265: 8 months NIH 20010769: Primary endpoint evaluated at 12 months; longer-term efficacy data presented at 36 months
Patient population	Patients with GL or PL
Sample size	N=107 (GL=66; PL=41; PL subgroup=31)*
Inclusion criteria	Age: Study NIH 2001769: 6 months; Study NIH 991265: >5 years Clinically significant LD identified as an absence of fat outside the range of normal variation and/or identified as a disfiguring factor by the patient Circulating leptin levels: Study NIH 2001769: <12.0 ng/mL in females, <8.0 ng/mL in males and <6 ng/mL in children 6 months- 5 years; Study NIH 991265: ≤8.0 ng/mL in females and ≤6.0 ng/mL in males Presence of at least 1 of the following metabolic abnormalities: • Presence of diabetes mellitus • Fasting insulin concentration >30 μU/mL (208.4 pmol/L) • Fasting triglyceride concentration >200 mg/dL (>2.26 mmol/L), or postprandially elevated triglyceride concentrations Triglyceride concentration >500 mg/dL (>5.65 mmol/L) when fasting is not clinically indicated (e.g., infants)b
Exclusion criteria	General: Pregnant women, women in their reproductive years who did not use an effective method of birth control, and women who were nursing or who were lactating within 6 weeks of having completed nursing. Exclusions for underlying disease likely to increase side effects or to hinder objective data collection: • Known infectious liver disease (in Study NIH 99165, known liver disease due to causes other than NASH) • Known human immunodeficiency (HIV) infection • Current alcohol or substance abuse • Psychiatric disorder impeding competence or compliance • Active tuberculosis • Use of anorexigenic drugs

	• Other condition(s) that in the opinion of the clinical investigators would impede completion of the study • Patients who have a known hypersensitivity to Escherichia coli-derived proteins
Statistical tests*	The primary and key secondary efficacy analyses were performed primarily on the FAS (all patients who received at least 1 dose of study drug and had either primary efficacy parameter of interest measured at baseline and at least one post-baseline visit). The primary and key secondary endpoints were summarised using descriptive statistics and 95% CIs. P values for the primary endpoints were computed using paired t-tests to determine if the change from baseline to Month 12 was significantly different from 0, at a one-sided α-level of 0.025. The LOCF method was used to impute any missing Month 12 results. The imputation only took into account results that were at least 6 months (180 days) post-baseline. Thus, the analysis included all patients that had baseline and at least Day 180 measurements. A MMRM analysis was used to assess changes over time for the entire duration of the study.
Primary outcomes	• Actual change from baseline in HbA_{1c} at Month 12 • Percent change from baseline in fasting serum triglycerides at Month 12
Key secondary outcomes	Proportion of patients achieving target actual decreases of: • ≥1% decrease in HbA_{1c} or ≥30% decrease in fasting serum triglycerides at Month 12 • ≥1.5% decrease in HbA_{1c} or ≥35% decrease in fasting serum triglycerides at Month 12 • ≥2% decrease in HbA_{1c} or ≥40% decrease in fasting serum triglycerides at Month 12 • Actual and percent change from baseline in fasting plasma glucose levels at Month 12
Other relevant secondary outcomes	• Actual change from baseline in HbA_{1c} at each post-baseline visit • Percent and actual change from baseline in fasting serum triglycerides at each postbaseline visit • Actual and percent change from baseline in fasting lipids (total cholesterol, LDL-C, HDL-C) through Month 12 • Actual change from baseline in ALT and AST at each post-baseline visit through Month 12 • Actual change from baseline in liver volume at each post-baseline visit through Month 12
Other endpoints of relevance	• Assessment of concomitant medications • Adverse events (including deaths, and cases of pancreatitis and infections) • Growth and pubertal status • Liver volume and pathology: Ultrasound of the liver and, if abnormalities are found, possibly liver biopsies
Study name	**FHA101**
Objective	To provide expanded access to metreleptin to patients with LD and associated metabolic disorders such as diabetes mellitus and/or hypertriglyceridemia and to test the safety and efficacy of metreleptin in this population of patients.
Location	Six centres in the US*

Design	Open-label, expanded-access
Duration of study	Continuous enrolment over 6 years (2008-2014)*: Primary endpoint evaluated at 12 months; longer-term efficacy data presented at 36 months
Patient population	Patients with GL or PL (including subgroup of PL patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L)
Sample size	N=41 (GL= 9; PL=32; PL subgroup=7)*
Inclusion criteria	Male or female ≥5 years old Physician-confirmed LD as defined by evidence of generalised (whole body) or partial (limbs) loss of body fat outside the range of normal variation Diagnosed with at least 1 of the following 2 metabolic disorders: • Diabetes mellitus • Hypertriglyceridemia as defined by fasting triglyceride concentrations >2.26 mmol/L (>200 mg/dL)
Exclusion criteria	Diagnosed with human immunodeficiency virus (HIV) infection Clinically significant medical condition that could potentially affect study participation and/or personal well-being, as judged by the Investigator Acquired LD and clinically significant haematologic abnormalities (such as neutropaenia and/or lymphadenopathy) Known infectious liver disease Known allergies to E. coli-derived proteins or hypersensitivity to any component of study treatment
Statistical tests*	The primary and key secondary efficacy analyses were performed primarily on the FAS (all patients who received at least 1 dose of study drug and had either primary efficacy parameter of interest measured at baseline and at least one post-baseline visit). The primary and key secondary endpoints were summarised using descriptive statistics and 95% CIs. P values for the primary endpoints were computed using paired t-tests to determine if the change from baseline to Month 12 was significantly different from 0, at a one-sided α-level of 0.025. The LOCF method was used to impute any missing Month 12 results. The imputation only took into account results that were at least 6 months (180 days) post-baseline. Thus, analysis of primary efficacy endpoints included all patients that have baseline and at least Month 6 measurements. A MMRM analysis was used to assess changes over time for the entire duration of the study.
Primary outcomes	• Actual change from baseline in HbA_{1c} at Month 12 • Percent change from baseline in fasting serum triglycerides at Month 12

Key secondary outcomes	Proportion of patients achieving target actual decreases of: • ≥1% actual decrease in HbA_{1c} or ≥30% decrease in fasting triglycerides at Month 12 • ≥1.5% decrease in HbA_{1c} or ≥35% decrease in fasting triglycerides at Month 12 • ≥2% actual decrease in HbA_{1c} or ≥40% decrease in fasting triglycerides at Month 12 • Actual and percent change from baseline for fasting glucose levels at Month 12
Other relevant secondary outcomes	• Actual change from baseline in HbA_{1c} at each post-baseline visit • Percent and actual change from baseline in fasting serum triglycerides at each postbaseline visit • Actual change from baseline in ALT and AST at each post-baseline visit through Month 12

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; FAS, full analysis set; FFA, free fatty acid; GL,generalised lipodystrophy; HbA_{1c}, glycated haemoglobin; HDL-C, high density lipoprotein cholesterol; HIV, human immunodeficiency virus; LD, lipodystrophy; LDL-C, low density lipoprotein cholesterol; LOCF, last observation carried forward; MMRM, Mixed-effect Model Repeated Measures; NASH, non-alcoholic steatohepatitis; NIH, National Institutes of Health; PL, partial lipodystrophy; UK, United Kingdom; US, United States

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

[b] Inclusion criteria for study NIH 20010769 (but not NIH 991265)

Table 5: Baseline characteristics for study NIH 991265/20010769, reproduced from Table C16 (CS, page 82)

Characteristic	GL (N = 66)	PL (N = 41)	
		PL subgroup[a] (N = 31)	Overall (N = 41)
Female, n (%)	51 (77.3)	30 (96.8)	40 (97.6)
Race, n (%)			
Caucasian	31 (47.0)	26 (83.9)	36 (87.8)
Black	16 (24.2)	0	0
Asian/Native American/Hispanic/Other	3 (4.5)/ 2 (3.0)/ 11 (16.7)/ 3 (4.5)	1 (3.2)/ 0 / 2 (6.5)/ 2 (6.5)	1 (2.4)/ 0/ 2 (4.9)/ 2 (4.9)
Age, years, median (range)	15.0 (1.0, 68.0)	38.0 (15.0, 64.0)	34.0 (10.0, 64.0)
<18 years	45 (68.2)	5 (16.1)	8 (19.5)
≥18 years	21 (31.8)	26 (83.9)	33 (80.5)
LD type, n (%)			
Acquired	21 (31.8)	4 (12.9)	6 (14.6)
Congenital/Familial	45 (68.2)	27 (87.1)	35 (85.4)
Fasting leptin, ng/ml, median (range)	1.0 (0.2, 5.3)	5.9 (1.6, 16.9)	5.9 (1.0, 16.9)
BMI, kg/m^2, median (range)	20.5 (14.0, 29.5)	25.1 (18.6, 33.3)	25.3 (17.7, 33.3)
HbA_{1c}, %			
Median (range)	8.7 (4.5, 13.7)	8.6 (5.7, 13.3)	7.8 (4.6, 13.3)
≥6.5, n (%)	49 (74.2)	29 (93.5)	29 (70.7)
≥8.0, n (%)	42 (63.6)	19 (61.3)	19 (46.3)
Fasting plasma glucose, mmol/L, median (range)	10.3 (5.04)	9.9 (4.33)	8.7 (4.35)
Fasting triglycerides, mmol/L			
Median (range)	14.5 (25.29)	14.8 (25.72)	12.0 (22.85)
≥2.26 mmol/L	50 (75.8)	27 (87.1)	34 (82.9)
≥5.65 mmol/L	26 (39.4)	15 (48.4)	15 (36.6)
ALT, >ULN, n (%)	49 (74.2)	9 (29.0)	14 (34.1)
AST, >ULN, n (%)	36 (54.5)	7 (22.6)	10 (24.4)
Anti-diabetic medications at baseline, n (%)	53 (80.3)	30 (96.8)	37 (90.2)
Lipid-lowering medications at baseline, n (%)	34 (51.5)	26 (83.9)	34 (82.9)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; GL, generalised lipodystrophy; HbA_{1c}, glycated haemoglobin; LD, lipodystrophy; PL, partial lipodystrophy; ULN, upper limit of normal

[a] PL subgroup, patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

ERG comment: Additional baseline lipodystrophy characteristics were reported in the NIH follow-up study,[46] for the 107 patients originally included in the NIH 991265/20010769 study

and an additional five patients. These data were not included in the CS, but are recorded in Table 6, below.

Table 6: Baseline lipodystrophy characteristics for the NIH follow-up study population, including the 107 participants in the NIH 991265/20010769 study

Characteristic	All patients N=112 (93 F, 19 M)	GL patients N=68 (51F, 17M)	PL patients N=44 (42 F, 2M)
Impaired physical appearance	86 (77%)	56 (82%)	30 (68%)
Disruption to female reproductive system	45 (80%)	21 (78%)	24 (83%)
Heart abnormality	50 (45%)	36 (53%)	14 (32%)
Hyperphagia	88 (79%)	57 (84%)	31 (70%)
Kidney abnormality	71 (63%)	46 (68%)	25 (57%)
Liver abnormality	105 (94%)	63 (93%)	42 (95%)
Pancreatitis	44 (39%)	21 (31%)	23 (52%)
Unable to attend school or perform work	48 (43%)	39 (57%)	9 (20%)

Impaired physical appearance is determined by the presence of acanthosis nigricans, hyperkeratosis, or hirsutism.
Disruption to female reproductive function is determined by the presence of irregular menstruation or polycystic ovary syndrome (PCOS).
Heart abnormality includes hypertrophy, any dilation, any regurgitation, cardiomyopathy, and tachycardia.
Hyperphagia is determined by notes in the medical charts.
Kidney abnormality includes proteinuria, enlarged kidneys, nephropathy, hydronephrosis, renal disease, nephromegaly, renal failure, renal calculus, and glomerulosclerosis.
Liver abnormality includes hepatomegaly, any form of fatty liver or steatosis, fibrosis, cirrhosis, and hepatitis.
A patient is considered to have pancreatitis at baseline if the patient has ≥1 episodes of pancreatitis in the one year prior to metreleptin initiation.
Loss of ability to perform work/school work is defined as incomplete school attendance due to disease symptoms for school age patients or not working/working part-time due to disease symptoms

Table 7: Baseline characteristics for study FH101, reproduced from Table C17 (CS, page 83)

Characteristic	GL (N = 9)	PL (N = 32)	
		PL subgroup[a] (N = 7)	Overall (N = 32)
Female, n (%)	8 (88.9)	7 (100.0)	31 (96.9)
Race n (%)			
Caucasian	8 (88.9)	5 (71.4)	22 (68.8)
Black	1 (11.1)	2 (28.6)	3 (9.4)
Asian/Native American/Hispanic/Other	0/0/0/0	0/0/0/0	1 (3.1)/ 2 (6.3)/ 1 (3.1)/ 3 (9.4)
Age, median (range)	25.0 (9.0, 67.0)	42.0 (23.0, 57.0)	44.5 (23.0, 67.0)
<18 years	3 (33.3)	0	0
≥18 years	6 (66.7)	7 (100.0)	32 (100.0)
LD type			

Characteristic	GL (N = 9)	PL (N = 32)	
		PL subgroup[a] (N = 7)	Overall (N = 32)
Acquired	6 (66.7)	1 (14.3)	3 (9.4)
Congenital/Familial	2 (22.2)	6 (85.7)	29 (90.6)
BMI, kg/m^2, median (range)	21.3 (13.9, 38.4)	27.6 (20.9, 30.5)	30.3 (19.1, 41.2)
HbA_{1c}, %			
Median (range)	8.4 (5.1, 10.2)	7.6 (5.7, 11.1)	8.0 (5.6, 12.8)
≥6.5, n (%)	6 (66.7)	6 (85.7)	27 (84.4)
≥8.0, n (%)	5 (55.6)	2 (28.6)	16 (50.0)
Fasting plasma glucose, mmol/L, median (range)	10.4 (4.2, 23.3)	7.4 (5.1, 13.4)	7.8 (2.0, 15.0)
Fasting triglycerides, mmol/L,			
Median (range)	3.3 (1.5, 119.9)	2.9 (0.7, 14.0)	3.2 (0.7, 50.4)
≥2.26 mmol/L	6 (66.7)	4 (57.1)	23 (71.9)
≥5.65 mmol/L	3 (33.3)	1 (14.3)	7 (21.9)
ALT, >ULN, n (%)	5 (55.6)	5 (71.4)	23 (71.9)
AST, >ULN, n (%)	4 (44.4)	2 (28.6)	9 (28.1)
Anti-diabetic medications at baseline, n (%)	2 (22.2)	6 (85.7)	19 (59.4)
Lipid-lowering medications at baseline, n (%)	2 (22.2)	6 (85.7)	19 (59.4)

Abbreviations: ALT = alanine aminotransferase; AST = aspartate aminotransferase; BMI = body mass index; GL = generalised lipodystrophy; LD = lipodystrophy; HbA_{1c} = glycated haemoglobin; PL = partial lipodystrophy; ULN = upper limit of normal

[a]PL subgroup = patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

Table 8: Protocol synopsis for the GL/PL natural history study, reproduced from an unpublished report included in the company's response to clarification questions

Study rationale	Generalized lipodystrophy (GL) and partial lipodystrophy (PL) are ultra-rare conditions associated with partially or fully absent adipose tissue, respectively. With fat accumulating in non-adipose tissue, GL and PL can lead to physical irregularities, organ damage. More research is needed to understand the natural history, including organ damage and mortality, of patients with GL and PL.
Objectives	1. To describe the demographic and clinical characteristics of metreleptin-naïve patients with GL and PL 2. To describe time to organ damage and time to disease progression of metreleptin-naïve patients with GL and PL 3. To describe the overall survival of metreleptin-naïve patients with GL and PL
Study Measures and Outcomes	Study measures included: - Patient demographic characteristics as of diagnosis of GL or PL (**Objective 1**) - Type of lipodystrophy diagnosed (i.e., phenotype and genotype) (**Objective 1**) - Patient physical characteristics and vital signs during patient's lifetime (**Objective 1**) - Laboratory values during patient's lifetime (**Objective 1**) - Organ damage during patient's lifetime (**Objective 2**)

	- Complications and comorbidities during patient's lifetime (**Objective 1**) - Mortality and causes of death (**Objective 3**) Disease progression was defined as the onset of a second organ damage following prior damage in a different organ. (**Objective 2**)
Data Sources	Data extracted from medical charts from five leading treatment centers for GL and PL across three countries (Brazil, Turkey, and the United States). These include: - United States (data collection complete) 1. National Institutes of Health (Rebecca Brown, MD, MHSc) 2. University of Michigan (Elif Oral, MD) - Turkey (data collection complete) 3. Dokuz Eylul University Medical School (Baris Akinci, MD) - Brazil (ongoing data collection) 4. Universidade de São Paulo – Campus Ribeirão Preto (Maria Cristina Foss de Freitas, MD) 5. Universidade Federal do Ceará (Renan Montenegro Junior, MD)
Data Collection Procedures	Retrospective, non-interventional, observational, closed cohort, longitudinal study design based on medical charts of metreleptin-naïve patients diagnosed with GL or PL prior to January 1, 2015. De-identified data for this study were collected from each site (e.g., investigators, research nurses, research assistants) into a single electronic database.
Data Analysis	All analyses were conducted for the entire sample, and by type of lipodystrophy (i.e., GL and PL) separately. **Objective 1**: Continuous variables were described in terms of means, standard deviations, and medians. Categorical were reported through frequencies and proportions. Standard errors for count variables were reported. **Objective 2**: Time to first organ damage and time to progression were analyzed through Kaplan-Meier analyses. Progression in the number of damaged organs (i.e., from 0 to 1, 1 to 2, 2 to 3, and 3 to 4) was also described using Kaplan-Meier analyses. **Objective 3**: Overall survival was described from the appearance of first evidence of GL or PL (i.e., first of appearance of symptoms or diagnosis) and from birth. Time to death was described using Kaplan-Meier analyses.
Privacy and Ethics	All patient data were de-identified prior to analysis. This study is non-interventional, no specific drug was investigated, and no prospective data were collected. This study was approved by local institutional review boards across all sites.
Note about PL patients: Not all included PL patients meet the criteria of low leptin levels, elevated A1c, and/or elevated triglycerides which have been proposed for the metreleptin EMA labelling. **Note about participating sites:** As of February 2018, data collection for Brazil was not yet complete. Data for the 178 patients from sites in the US and Turkey are shown.	

Table 9: Baseline characteristics for the GL/PL study, taken from an unpublished report included in the company's response to clarification questions

Characteristic	GL (N = 56)	PL (N = 122)	All (N=178)
Female, n (%)	33 (58.9)	86 (70.5)	119 (66.9)
Race, n (%)			
Caucasian	11 (19.6)	63 (51.6)	74 (41.6)
Black	11 (19.6)	2 (1.6)	13 (7.3)
Asian/Native American/Hispanic/Other[$]	0 (0)/0 (0)/1 (1.8)/33 (58.9)	0 (0)/0 (0)/5 (4.1)/53 (43.4)	0 (0)/0 (0)/6 (3.4)/86 (46.3)
Age at diagnosis, years, median (IQR)	11 (4, 21)	34 (24, 48)	29 (13, 43)
<18 years (%)	37 (66.1)	20 (16.4)	57 (32.0)
≥18 years (%)	19 (33.9)	102 (83.6)	121 (68.0)

Characteristic	GL (N = 56)	PL (N = 122)	All (N=178)
LD type, n (%)			
Acquired	5 (8.9)	26 (21.3)	31 (17.4)
Congenital/Familial	49 (87.5)	96 (78.7)	145 (81.5)
Fasting leptin, ng/ml			
n (%)	1 (5.9)	14 (25.9)	15 (21.1)
mean (SD)	1.2 (0)	8.8 (7.7)	8.3(7.7)
BMI, kg/m^2, median (range)	NR	NR	NR
HbA_{1c}, %			
n (%)	6 (35.3)	40 (74.1)	46 (64.8)
Mean (SD)	8.1 (3.4)	7.4 (2.0)	7.5 (2.2)
≥6.5, n (%), n (%)	3 (50.0)	22 (55.0)	25 (54.3)
≥8.0, n (%), n (%)	3 (50.0)	15 (37.5)	18 (39.1)
Fasting plasma glucose, mmol/L			
n (%)	12 (70.6)	33 (61.1)	45 (63.4)
mean (SD)	150.0 (116.6)	163.7 (71.5)	160.0 (84.6)
Fasting triglycerides*, mmol/L			
n (%)	13 (76.5)	46 (85.2)	59 (83.1)
Mean (SD)	5.4 (3.7)	5.1 (6.9)	5.1 (6.3)
≥2.26 mmol/L, n (%)	10 (76.9)	25 (54.3)	35 (59.3)
≥5.65 mmol/L, n (%)	6 (46.2)	10 (21.7)	16 (27.1)
ALT			
n (%)	16 (94.1)	49 (90.7)	65 (91.5)
ALT, >ULN, n (%)	5 (31.3)	13 (26.5)	18 (27.7)
AST			
n (%)	16 (94.1)	47 (87.0)	63 (88.7)
AST, >ULN, n (%)	3 (18.8)	5 (10.6)	8 (12.7)
Anti-diabetic medications at baseline, n (%)	NR	NR	NR
Lipid-lowering medications at baseline, n (%)	NR	NR	NR
Liver damage	15 (26.8)	27 (22.1)	42 (23.6)
Kidney damage	4 (7.1)	14 (11.5)	18 (10.1)
Heart damage	8 (14.3)	10 (8.2)	18 (10.1)
Pancreatitis	2 (3.6)	8 (6.6)	10 (5.6)

*Fasting triglycerides converted from reported units (mg/dL) to mmol/L
[$]Of those participants who's ethnicity was classified as 'other', 80/86 were Turkish
Liver damage includes chronic hepatitis, mild to severe fibrosis, cirrhosis, hepatic steatosis, hepatomegaly, transplant and other types of liver disease (n=5)
Kidney damage includes albuminuria, nephropathy, proteinuria, kidney failure requiring dialysis or transplant, transplant and other kidney disease (n=7)
Heart damage includes angina, atherosclerosis, atrial fibrillation, cardiac arrhythmia, cardiomyopathy, heart failure, ischemia, left ventricular hypertrophy, myocardial infarction, transplant and other heart abnormalities (n=10)

4.2.2 Details of relevant studies not included in the submission

As noted in section 4.2.1, nine studies[48-56] which met the pre-specified inclusion criteria[1] and were consistent with the NICE scope[27] were inappropriately excluded from the submission. In addition, details of the methods and results of the two main studies (the GL/PL natural history study and the NIH follow-up study) used to inform the cost effectiveness analysis were not included in the submission; study reports[40, 46] were provided in the company's response to clarification questions and, as far as possible, we have included information from these documents in our report.

The company's response to clarification questions acknowledged that: 'One of the primary objectives of the NIH Follow-Up study was to build on the NIH pivotal trial and extend it in two ways: a) increase the patient sample size (from 107 to 112), and b) expand the outcomes evaluated from biomarkers such as HbA_{1c} and triglycerides to more direct measures of clinical burden for patients including hyperphagia, organ abnormalities, physical appearance, ability to perform work/school, mortality, etc.'[39] No justification was provided for not reporting results for patient perceived outcomes from the NIH follow-up study in the CS, beyond a statement that: 'The NIH Follow-Up study included many of these clinical outcomes and they are incorporated into the CE model.'[39]

4.2.3 Summary and critique of company's analysis of validity assessment

The company provided an appraisal of the validity of the two metreleptin intervention studies included in the CS,[37, 38] using seven criteria based on the 12 CASP questions for cohort studies (see Section 4.1.4):

- Was the cohort recruited in an acceptable way?
- Was the exposure accurately measured to minimise bias?
- Was the outcome accurately measured to minimise bias?
- Have the authors identified all important confounding factors?
- Have the authors taken account of the confounding factors in the design and/or analysis?
- Was the follow-up of patients complete?
- How precise (for example, in terms of confidence interval and p values) are the results?

The validity assessment performed by the company (Section 9.5.1, CS pages 87-90, and corrected in the response to clarification questions) is reproduced in Tables 10 and 11 below.

Table 10: Critical appraisal of study NIH 991265/20010769

Study name: NIH 991265/20010769		
Study question	**Response yes/no/not clear/N/A)**	**How is the question addressed in the study?**
Was the cohort recruited in an acceptable way?	Yes	The patient population was representative of a defined population. The patients had low leptin levels (<12.0 ng/mL in females, <8.0 ng/mL in males and <6 ng/mL in children 6 months- 5 years) and at least 1 metabolic abnormality out of diabetes mellitus; fasting insulin concentration >30 μU/mL, and/or fasting triglyceride concentration >2.26 mmol/L or postprandially elevated triglycerides >5.65 mmol/L when fasting was clinically not indicated (e.g., in infants); these are the hallmarks of this syndrome, i.e., insulin resistance with diabetes mellitus and hypertriglyceridemia. Patients were recruited from different regions across the world.
Was the exposure accurately measured to minimise bias?	Yes	Exposure was clearly defined and accurately measured. The measurement of exposure was objective i.e. dose and duration, including average (mean [SD], median and range) for daily dose (mg/day), and weighted average dose (mg/kg).
Was the outcome accurately measured to minimise bias?	Yes	The study's efficacy endpoints were objective measurements, including the co-primary endpoints of HbA_{1c} and triglycerides. These measurements were primarily obtained at a single laboratory and thus treatment effects could be appropriately evaluated. The efficacy endpoints were clinically relevant to the patient and the progression of disease.
Have the authors identified all important confounding factors?	Yes	Potential confounding factors included: concomitant medication use, sex, race, age, weight, height, body weight category, BMI, region, LD subtype (CGL, AGL, FPL, APL), gene mutation (LMNA, PPARg, Seipin, AGPAT-2, ZMPSTE24, Other, and not applicable), baseline laboratory values.

Study name: NIH 991265/20010769		
Study question	**Response yes/no/not clear/N/A)**	**How is the question addressed in the study?**
Have the authors taken account of the confounding factors in the design and/or analysis?	Yes	In addition to the FAS, efficacy was analysed on the CFAS, which included all patients in the FAS who have controlled concomitant medication use, described as no change or a decrease in baseline concomitant medications (anti-diabetic or lipid lowering therapies), prior to Month 12. Data for all anti-diabetic or lipid lowering therapies, including type, dose, regimen, and route of administration, underwent medical review and patients who had these types of medications added or doses increased that may have had an impact on the efficacy endpoints were excluded from the CFAS. Patients were excluded separately based on the type of medication that was added or increased, e.g., patients with potentially confounding anti-diabetes medications were excluded from the analyses of HbA_{1c} and those with potentially confounding lipid-lowering therapies were excluded from analyses of triglycerides. In general, the results for the efficacy analyses were consistent for the FAS and the CFAS. In addition, subgroup analyses were conducted based on a number of baseline characteristics to show whether treatment effects were observed consistently across relevant populations. including: LD subtype (AGL, CGL, FPL, and APL); age (age categories <6, ≥6 to <12, ≥12 to <18, < 18, and ≥18 years old); region (US, EU, EU and Eastern Mediterranean, and Other); presence of metabolic abnormalities at baseline (HbA_{1c} [<6.5 and ≥6.5%], ≥7%, ≥8% and fasting triglycerides [<2.26 mmol/L and ≥2.26 mmol/L / <200 and ≥200 mg/dL, ≥5.65 mmol/L / ≥500 mg/dL; and between ≥2.26 and ≤5.65 mmol/L / ≥200 and ≤500 mg/dL]); concomitant insulin, anti-diabetic medications and lipid-lowering medications at baseline; baseline leptin levels (<12 ng/mL / ≥12 ng/mL, primary efficacy analysis only) (see Section 9.6.1.5)
Was the follow-up of patients complete?	Yes	Only one patient was lost to follow-up (see Section 9.4.7)

Study name: NIH 991265/20010769		
Study question	**Response yes/no/not clear/N/A)**	**How is the question addressed in the study?**
How precise (for example, in terms of confidence interval and p values) are the results?	Yes, the precision of the results is reasonable	The following results with 95% CIs were reported were reported: GL patients: mean change from baseline to Month 12/LOCF for HbA_{1c} was -2.2% (95% CI: -2.7, -1.6) and the mean percent change in triglycerides was -32.1% (-51.0, -13.2) PL subgroup[a] patients (excluding outlier patient): mean change from baseline to Month 12/LOCF for HbA_{1c} was -0.9% (95% CI: -1.4, -0.4) and the mean percent change in triglycerides was -37.4% (-57.2, -8.6). The majority of patients in both the GL group and the PL subgroup achieved meaningful reductions in both HbA_{1c} and triglycerides.
Adapted from Critical Appraisal Skills Programme (CASP): Making sense of evidence 12 questions to help you make sense of a cohort study		
Abbreviations: AGL = acquired generalised lipodystrophy; APL = acquired partial lipodystrophy; BMI = body mass index; CFAS = Controlled Concomitant Medication Full Analysis Set; CGL = congenital generalised lipodystrophy; CI = confidence interval; EU = European Union; FAS = full analysis set; FPL = familial partial lipodystrophy; GL = generalised lipodystrophy; HbA_{1c} = glycated haemoglobin; LD = lipodystrophy; LOCF = last observation carried forward; PL = partial lipodystrophy; SD = standard deviation; US = United States		

Table 11: Critical appraisal of study FH101

Study name: FHA101		
Study question	Response yes/no/not clear/N/A)	How is the question addressed in the study?
Was the cohort recruited in an acceptable way?	Yes	The patient population was representative of a defined population. Patients had to have been diagnosed with at least 1 of the following 2 metabolic disorders: diabetes mellitus and/or hypertriglyceridemia as defined by fasting triglyceride concentrations >2.26 mmol/L (>200 mg/dL), which are the hallmark of this syndrome
Was the exposure accurately measured to minimise bias?	Yes	Exposure was clearly defined and accurately measured. The measurement of exposure was objective i.e. dose and duration, including average (mean [SD], median and range) for daily dose (mg/day), weighted average dose (mg/kg).
Was the outcome accurately measured to minimise bias?	Yes	The study's efficacy endpoints were objective measurements, including the co-primary endpoints of HbA_{1c} and triglycerides. These measurements were primarily obtained at a single laboratory and thus treatment effects could be appropriately evaluated. The efficacy endpoints were clinically relevant to the patient and the progression of disease.
Have the authors identified all important confounding factors?	Yes	Potential confounding factors included: concomitant medication use, sex, race, age, weight, height, body weight category, BMI, region (US, EU, EU and Eastern Mediterranean, other), LD subtype (CGL, AGL, FPL, APL), gene mutation (LMNA, PPARg, Seipin, AGPAT-2, ZMPSTE24, Other, and Not Applicable), baseline laboratory values
Have the authors taken account of the confounding factors in the design and/or analysis?	Partially	As in study NIH 991265/20010769 efficacy was analysed on the FAS and the CFAS, which included all patients in the FAS who have controlled concomitant medication use, described as no change or a decrease in baseline concomitant medications (anti-diabetic or lipid lowering therapies), prior to Month 12. In general, the results for the efficacy analyses were consistent for the FAS and the CFAS.
Was the follow-up of patients complete?	Yes	Only two patients were lost to follow-up (see Section **Error! Reference source not found.**)
How precise (for example, in terms of confidence interval and p values) are the results?	Due to the small sample sizes, the 95% CIs were wide	The following results with 95% CIs were reported were reported: GL patients: mean change from baseline to Month 12/LOCF for HbA_{1c} was -1.2 % (95% CI: -4.3, 2.0) and the mean percent change in triglycerides was -26.9% (-124.1, 70.4) PL subgroup patients (excluding outlier patient): mean change from baseline to Month 12/LOCF for HbA_{1c} was -0.9% (95% CI: -1.4, -0.4) and the mean percent change in triglycerides was -8.5% (-36.4, 19.5).
Adapted from Critical Appraisal Skills Programme (CASP): Making sense of evidence 12 questions to help you make sense of a cohort study		

Abbreviations: AGL, acquired generalised lipodystrophy; APL = acquired partial lipodystrophy; BMI = body mass index; CFAS = Controlled Concomitant Medication Full Analysis Set; CGL = congenital generalised lipodystrophy; CI = confidence interval; EU = European Union; FAS = full analysis set; FPL = familial partial lipodystrophy; GL = generalised lipodystrophy; HbA_{1c} = glycated haemoglobin; LD = lipodystrophy; LOCF = last observation carried forward; PL = partial lipodystrophy; SD = standard deviation; US = United States

The ERG agrees with the content of the critical appraisals provided, but does not consider this to be an adequate approach to assessing risk of bias in a cohort study (see Section 4.1.4).

No critical appraisal or risk of bias assessment was provided for the GL/PL natural history study.

4.2.4 Summary and critique of results

For the evaluation of clinical effectiveness of any treatment, a comparison between treated and untreated patients, who are similar with respect to characteristics other than treatment, is needed. Clinical or 'patient-perceived' outcomes, such as organ damage or hyperphagia, are more relevant than biochemical markers of 'surrogate outcome measures', such as triglyceride levels or HbA_{1c}. The CS (pages 90-95 and 103-104) focuses on change from baseline, in triglyceride levels or HbA_{1c}, in metreleptin treated patients. These results, along with any results for clinical outcomes included in the CS (pages 98-100) are reproduced and critiqued below.

We have added further results for clinical outcomes, which were not included in the CS, including results from the NIH follow-up study and the GL/PL natural history study, for which no results were reported in the CS.

Efficacy

Change in HbA_{1c} and triglycerides

The single arm metreleptin treatment study, NIH 991265/20010769, found statistically significant reductions in both HbA_{1c} and triglyceride levels in both GL and PL.[37] The mean (SD) actual change in % HbA_{1c}, from baseline to month 12 of treatment, LOCF, was -2.2 (2.15) for GL patients, -0.9 (1.23) for the PL subgroup and -0.6 (1.22) for all PL patients. The corresponding values, for % change in triglyceride levels, were -32.1 (71.28) for GL patients, -37.4 (30.81) for the PL subgroup and -20.8 (47.93) for all PL patients. Full results for markers of glycaemic control and lipid metabolism are provided in Table 12 below, reproduced from the CS (CS, Table C22, pages 90-92).[1]

Additional data were presented in the CS (pages 96-97) to support the persistence of these effects to 36 months. The CS also includes some subgroup data for changes in percentage HbA_{1c} and triglycerides. In general, greater mean decreases from baseline to the primary time point of Month 12/LOCF were observed amongst patients who had higher baseline percentage HbA_{1c} and triglyceride levels. Similarly, patients with the acquired forms of LD generally achieved larger mean decreases from baseline compared with patients who had the congenital/familial form. Subgroup data for markers of glycaemic control and lipid metabolism are provided in Table 13 below, reproduced from the CS (CS, Table C23, pages 101-102).[1]

ERG comment: Subgroup data were not provided for the overall PL population.

The smaller, single arm metreleptin treatment study, FH101, reported decreases in percentage HbA_{1c} and triglyceride levels, from baseline to month 12 of treatment, in all patient groups. However, these decreases were not statistically significant. Full results for markers of

glycaemic control and lipid metabolism are provided in Table 14 below, reproduced from the CS (CS, Table C24, pages 103-105).[1]

ERG comment: One study, which met the pre-specified inclusion criteria but was excluded from the CS (see section 4.2.1,[50] assessed the effects of leptin therapy in 24 female patients with Dunnigan variety FPL and moderate or severe hypoleptinemia and found no significant change from baseline to six months in fasting glucose, insulin, glucose tolerance, or HbA_{1c} levels.

The GL/PL natural history study[40] did not report any information about changes in markers of glycaemic control or lipid metabolism over time.

Table 12: Glycaemic control and lipid metabolism results from NIH 991265/20010769 study

Study name		NIH 991265/20010769		
Size of study groups	**Treatment**	GL = 62 PL subgroup[a] = 30 PL overall = 40		
Study duration	**Time unit**	12 months		
Type of analysis	**Intention-to - treat/per protocol**	FAS: all patients who received at least 1 dose of study drug and who had either primary efficacy parameter of interest measured at baseline and at least one post-baseline visit		
Co-primary endpoint: Change from baseline in HbA_{1c} (%) using LOCF (FAS population, excluding outlier patient[b])				
		GL N = 62	PL subgroup N = 29[a.b]	PL overall N = 39[b]
Baseline value	**n**	62	29	39
	Mean (SD)	8.6 (2.33)	8.8 (1.91)	8.0 (2.18)
Month 12 value, LOCF	**n**	59	27	36
	Mean (SD)	6.4 (1.68)	8.0 (1.83)	7.5 (1.84)
Effect size: actual change from baseline	**n**	59	27	36
	Mean (SD)	-2.2 (2.15)	-0.9 (1.23)	-0.6 (1.22)
	95% CI	-2.7, -1.6	-1.4, -0.4	-1.0, -0.2
Statistical test	**Type**	P values computed using paired t-tests		
	p value	<0.001	<0.001	0.005
Co-primary endpoint: Change from baseline in triglycerides (mmol/L) using LOCF (FAS population, excluding outlier patient[b])				
		GL N = 62	PL subgroup N = 29[a, b]	PL overall N = 39[b]
Baseline value	**n**			
	Mean (SD)	14.7 (25.66)	15.7 (26.42)	12.5 (23.35)
Month 12 value, LOCF	**n**			
	Mean (SD)	4.5 (6.10)	6.0 (8.41)	5.4 (7.37)
Effect size: percent change from baseline	**n**	57	27	36
	Mean (SD)	-32.1 (71.28)	-37.4 (30.81)	-20.8 (47.93)
	95% CI	-51.0, -13.2	-57.2, -8.6	-51.0, -13.2

Statistical test	**Type**	P values computed using paired t-tests		
	p value	0.001	<0.001	0.013
Key secondary endpoint: Actual and percent change from baseline in fasting plasma glucose levels at Month 12 (mmol/L) using LOCF (FAS population)				
		GL N = 62	PL subgroup N = 30[a]	PL overall N = 40
Baseline value	**n**			
	Mean (SD)	10.2 (5.05)	10.0 (4.36)	8.8 (4.39)
Month 12 value, LOCF	**n**	59	28	37
	Mean (SD)	7.0 (3.40)	8.1 (3.55)	7.5 (3.28)
Effect size: actual change from baseline	**n**	59	28	37
	Mean (SD)	-3.0 (4.72)	-1.8 (2.83)	-1.2 (2.69)
	95% CI	-4.2, -1.7	-2.9, -0.7	-2.1, -0.3
Statistical test	**Type**	P values computed using paired t-tests		
	p value	<0.001	0.003	0.012
Effect size: percent change from baseline	**n**	59	28	37
	Mean (SD)	-19.7 (37.21)	-13.2 (28.99)	-6.1 (29.59)
	95% CI	-29.4, -10.0	-24.4, -1.9	-16.0, 3.8
Statistical test	**Type**	P values computed using paired t-tests		
	p value	<0.001	0.023	0.219
Key secondary endpoint: Responder analysis: patients who met target reductions in HbA_{1c} or triglycerides at Month 12/LOCF (FAS population)				
		GL N = 62	PL subgroup N = 30[a]	PL overall N = 40
≥1% actual decrease in HbA_{1c} or ≥30% decrease in triglycerides				
Month 12 value, LOCF	**n/N1 (%)**	47/59 (79.7)	19/28 (67.9)	19/37 (51.4)
	95% CI[c]	(67.2, 89.0)	(47.7, 84.1)	(34.4, 68.1)
≥1.5% actual decrease in HbA_{1c} or ≥35% decrease in triglycerides				
Month 12 value, LOCF	**n/N1 (%)**	44/59 (74.6)	14/28 (50.0)	14/37 (37.8)
	95% CI[c]	61.6, 85.0	30.7, 69.4	22.5, 55.2
≥2% actual decrease in HbA_{1c} or ≥40% decrease in triglycerides				

Month 12 value, LOCF	**n/N1 (%)**	39/59 (66.1)	12/28 (42.9)	12/37 (32.4)
	95% CI[c]	52.6, 77.9	24.5, 62.8	18.0, 49.8
Other secondary endpoints: Change from baseline to Month 12/LOCF in fasting lipids (FAS population)				
		GL N = 62	PL subgroup N = 30[a]	PL overall N = 40
Total cholesterol (mmol/L)				
Baseline	**n**	62	30	40
	Mean (SD)	5.9 (3.66)	6.4 (2.80)	5.9 (2.62)
Actual change from **baseline**	**n**	41	21	30
	Mean (SD)	-2.3 (2.91)	-0.9 (1.52)	-0.6 (1.45)
LDL-C (mmol/L)				
Baseline	**n**	37	17	24
	Mean (SD)	2.6 (1.35)	2.8 (1.02)	2.6 (1.01)
Actual change from **baseline**	**n**	22	12	18
	Mean (SD)	-0.9 (1.29)	-0.3 (0.66)	-0.1 (0.62)
HDL-C (mmol/L)				
Baseline	**n**	56	25	35
	Mean (SD)	0.7 (0.25)	0.8 (0.23)	0.8 (0.21)
Actual Change from **BL**	**n**	35	17	26
	Mean (SD)	-0.0 (0.24)	0.0 (0.14)	0.0 (0.14)

Abbreviations: CI, confidence interval; FAS, full analysis set; GL, generalised lipodystrophy; HbA_{1c}, glycated haemoglobin; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; LOCF, last observation carried forward; PL, partial lipodystrophy; SD, standard deviation

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

[b] Excluding results for Patient 901-080 who had an outlier value for percent increase from baseline in triglycerides of >1000% at Month 12/LOCF. The patient was terminated from treatment by the Investigator for noncompliance with dosing

Table 13: Glycaemic control and lipid metabolism subgroup results from NIH 991265/20010769 study

	GL				PL subgroup[a,b]			
	HbA_{1c}		Triglycerides		HbA_{1c}		Triglycerides	
	N	**Mean (SD) actual Δ to Month 12**	N	**Mean (SD) percent Δ to Month 12**	N	**Mean (SD) actual Δ to Month 12**	N	**Mean (SD) percent Δ to Month 12**
Baseline HbA_{1c} (%):								
<6.5	14	-0.1 (0.35)	14	-4.1 (55.58)	2	0.1 (0.64)	2	-40.8 (27.29)
≥6.5	45	-2.8 (2.08)	43	-41.2 (73.97)	25	-1.0 (1.24)	25	-37.1 (31.57)
≥7	45	-2.8 (2.08)	43	-41.2 (73.97)	23	-1.1 (1.28)	23	-37.2 (32.95)
≥8	39	-3.0 (2.13)	37	-38.6 (78.36)	18	-1.3 (1.33)	18	-43.6 (33.60)
Baseline triglycerides (mmol/L):								
<2.26	13	-1.6 (1.71)	13	6.7 (44.20)	3	-0.9 (0.36)	3	-20.7 (28.33)
≥2.26	45	-2.3 (2.28)	45	-42.5 (73.87)	24	-0.9 (1.31)	24	-39.5 (31.03)
≥5.65	24	-3.3 (2.56)	24	-72.0 (25.09)	15	-1.0 (1.62)	15	-53.7 (25.21)
LD type								
Congenital/ Familial	40	-1.8 (1.92)	39	-22.2 (80.54)	23	-0.7 (0.88)	23	-37.4 (26.64)
Acquired	19	-2.9 (2.47)	18	-53.5 (39.09)	4	-2.0 (2.42)	4	-37.0 (54.98)
Age (years)								
< 6	5	0.2 (0.60)	5	-10.5 (58.18)	0	NA	0	NA
≥6 to <12	11	-1.1 (1.51)	11	-14.1 (49.74)	0	NA	0	NA
≥12 to <18	24	-2.6 (1.89)	23	-42.9 (45.55)	5	-0.6 (1.24)	5	-50.6 (33.62)
≥18	19	-2.8 (2.46)	18	-35.3 (106.23)	22	-1.0 (1.25)	22	-34.4 (30.15)
Region[c]								
US	34	-1.9 (2.02)	34	-23.2 (85.87)	20	-1.0 (1.32)	20	-41.8 (27.97)
EU and EM	11	-2.6 (1.96)	11	-52.1 (41.84)	2	-0.7 (0.28)	2	13.3 (38.20)
EU	7	-1.5 (1.45)	7	-38.7 (48.04)	1	-0.5 (NA)	1	40.3 (NA)
Other	12	-2.6 (2.81)	11	-39.5 (39.99)	5	-0.8 (1.23)	5	-39.8 (26.45)

Abbreviations: Δ**,** change; EU, European Union, EM, Eastern Mediterranean; FAS, Full Analysis Set; GL, generalised lipodystrophy; HbA_{1c}, glycated haemoglobin; LOCF, last observation carried forward; NA, non-applicable; PL, partial lipodystrophy; SD, standard deviation; US, United States

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

[b] Excluding results for Patient 901-080 who had an outlier value for percent increase from baseline in triglycerides of >1000% at Month 12/LOCF. The patient was terminated from treatment by the Investigator for noncompliance with dosing (Study NIH 991265/20010769, Listing 16.2.1.1)

[c] EU includes Belgium, UK, Germany, Italy, Lithuania, and Spain; EM includes Turkey, Albania, Israel, and Serbia; Other includes Argentina, Canada, India, Madagascar, Pakistan, Peru, and Saudi Arabia

Table 14: Glycaemic control and lipid metabolism, results from FH101 study

Study name		FHA101		
Size of study groups	**Treatment**	GL = 9 PL subgroup[a] = 7 PL overall = 29		
Study duration	**Time unit**	12 months		
Type of analysis	**Intention-to -treat/per protocol**	FAS: all patients who received at least 1 dose of study drug and who had either primary efficacy parameter of interest measured at baseline and at least one post-baseline visit		
Co-primary endpoint: Change from baseline in HbA_{1c} (%) using LOCF (FAS population)				
		GL N = 9	PL subgroup[a] N = 7	PL overall N = 29
Baseline value	**n**	9	7	29
	Mean (SD)	7.7 (1.99)	7.8 (1.71)	8.1 (1.77)
Month 12 value, LOCF	**n**	5	7	26
	Mean (SD)	6.2 (1.96)	7.0 (0.76)	7.8 (1.76)
Effect size: actual change from baseline	**n**	5	7	26
	Mean (SD)	-1.2 (2.53)	-0.8 (1.85)	-0.4 (1.49)
	95% CI	-4.3, 2.0	-2.5, 0.9	-1.0, 0.2
Statistical test	**Type**	P values computed using paired t-tests		
	p value	0.360	0.289	0.210
Co-primary endpoint: Change from baseline in triglycerides (mmol/L) using LOCF (FAS population)				
		GL N = 9	PL subgroup[a] N = 7	PL overall N = 29
Baseline value	**n**	8	7	29
	Mean (SD)	19.9 (40.90)	4.0 (4.54)	8.5 (12.37)
Month 12 value, LOCF	**n**	6	7	26
	Mean (SD)	7.6 (11.10)	3.6 (3.57)	6.4 (10.06)
Effect size: percent change from baseline	**n**	5	7	26
	Mean (SD)	-26.9 (78.32)	-8.5 (30.22)	8.7 (93.39)
	95% CI	-124.1, 70.4	-36.4, 19.5	-29.1, 46.4

Statistical test	**Type**	P values computed using paired t-tests		
	p value	0.486	0.485	0.640
Key secondary endpoint: Actual and percent change from baseline in fasting plasma glucose levels at Month 12 (mmol/L) using LOCF (FAS population)				
		GL N = 9	PL subgroup[a] N = 7	PL overall N = 29
Baseline value	**n**	9	7	29
	Mean (SD)	11.4 (6.03)	8.0 (2.83)	8.5 (3.45)
Month 12 value, LOCF	**n**	6	7	27
	Mean (SD)	10.2 (7.58)	6.9 (2.16)	8.3 (2.99)
Effect size: actual change from BL	**n**	6	7	27
	Mean (SD)	-1.5 (9.90)	-1.1 (2.95)	-0.2 (4.14)
	95% CI	-11.9, 8.8	-3.8, 1.6	-1.8, 1.5
Statistical test	**Type**	P values computed using paired t-tests		
	p value	0.719	0.358	0.838
Effect size: percent change from baseline	**n**	6	7	27
	Mean (SD)	-7.3 (53.71)	-9.0 (26.45)	13.9 (69.14)
	95% CI	-63.6, 49.1	-33.4, 15.5	-13.4, 41.3
Statistical test	**Type**	P values computed using paired t-tests		
	p value	0.754	0.403	0.304
Key secondary endpoint: Responder analysis: patients who met target reductions in HbA_{1c} or triglycerides at Month 12/LOCF (FAS population)				
		GL N = 9	PL subgroup[a] N = 7	PL overall N = 29
≥1% actual decrease in HbA_{1c} or ≥30% decrease in triglycerides				
Month 12 value, LOCF	**n/N1 (%)**	3/6 (50.0)	2/7 (28.6)	9/26 (34.6)
	95% CI[b]	11.8, 88.2	3.7, 71.0	17.2, 55.7
≥1.5% actual decrease in HbA_{1c} or ≥35% decrease in triglycerides				
Month 12 value, LOCF	**n/N1 (%)**	3/6 (50.0)	2/7 (28.6)	9/26 (34.6)
	95% CI[b]	11.8, 88.2	3.7, 71.0	17.2, 55.7
≥2% actual decrease in HbA_{1c} or ≥40% decrease in triglycerides				

Month 12 value, LOCF	**n/N1 (%)**	3/6 (50.0)	1/7 (14.3)	7/26 (26.9)
	95% CI[b]	11.8, 88.2	0.4, 57.9	11.6, 47.8
	Mean (SD)	-104.1 (74.18)	-0.3 (7.21)	-3.6 (24.81)

Abbreviations: CI, confidence interval; FAS, full analysis set; GL, generalised lipodystrophy; HbA_{1c}, glycated haemoglobin; LOCF, last observation carried forward; PL, partial lipodystrophy; SD, standard deviation

[a] PL subgroup = patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

[b] 95% CI based on the 2-sided exact binomial proportions

Persistence of change in HbA_{1c} and triglycerides over time

The CS[37] reports some information about longer term (up to 36 months) changes in HbA_{1c} and triglycerides in patients on metreleptin treatment. Least-squares mean (LS mean) changes from baseline in HbA_{1c} in the GL group based on a mixed model repeated measures (MMRM) analysis were -2.3%, -2.1% and -1.5% at Months 12, 24 and 36, respectively.[1, 37] The overall MMRM analysis showed a statistically significant decrease from baseline for GL patients with an LS mean change of -1.4% ($p<0.001$). Results were similar in the PL subgroup with LS mean changes in HbA_{1c} of -0.9%, -1.3%, and -1.0% at Months 12, 24, and 36 and an overall LS mean change of -0.6% ($p<0.001$).[1, 37]

In the GL group, LS mean percent changes from baseline in triglycerides were -48.3%, -22.6% and -40.6% at Months 12, 24, and 36, respectively; based on the overall MMRM analysis, the LS mean change in triglycerides was -22.4% ($p<0.001$). For the PL subgroup (excluding data from the 'outlier' patient described previously), LS mean percent changes in triglycerides were -36.2%, -31.7%, and -13.7% at Months 12, 24 and 36, respectively, with an overall LS mean change of -18.6% ($p=0.004$).[1, 37]

ERG comment: Data for the overall PL population (not included in the CS) indicated no statistically significant change in triglyceride levels over time. The LS mean (SEM) percentage change values were as follows: month 12 = -16.7 (8.62), p = 0.054; month 24 = -9.4 (16.41), p = 0.566; month 36 =4.4 (17.53), p = 0/801; overall MMRM = -8.3 (5.46), p=0.131.[37]

Liver function (hepatic enzymes), liver pathology

Data from the NIH 991265/20010769 study,[1, 37] suggest that metreleptin treatment may be associated reductions in hepatic enzymes. In the 41 GL patients with hepatic data available, the mean (SD) changes, in ALT and AST, from baseline to month 12 of treatment were -53.1 (126.56) U/L and -23.8 (142.38) U/L, respectively. Reductions were smaller for the PL subgroup (-5.0 (11.95) and -6.0 (14.77) for ALT and AST, respectively) and for the overall PL group (-0.4 (26.95) and -5.1 (21.06) for ALT and AST, respectively. Full results for hepatic enzymes are provided in Table 15 below, reproduced from the CS (CS, Table C22, pages 90-92).[1] No assessments of statistical significance were presented.

Table 15: Hepatic enzymes results from NIH 991265/20010769 study

Change from baseline to Month 12 in liver transaminase levels (FAS Population)				
		GL N = 62	PL subgroup N = 30[a]	PL overall N = 40
ALT (U/L)				
Baseline	**n**	62	30	40
	Mean (SD)	111.9 (112.62)	39.2 (28.02)	54.8 (57.99)
Actual change from **baseline**	**n**	41	21	30
	Mean (SD)	-53.1 (126.56)	-5.0 (11.95)	-0.4 (26.95)
AST (U/L)				
Baseline	**n**	62	30	40
	Mean (SD)	75.0 (71.07)	31.9 (19.64)	38.4 (33.46)
Actual change from **baseline**	**n**	41	21	30
	Mean (SD)	-23.8 (142.38)	-6.0 (14.77)	-5.1 (21.06)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; FAS, full analysis set; GL, generalised lipodystrophy; PL, partial lipodystrophy; SD, standard deviation

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

ERG comment: The full CSR for the NIH 991265/20010769 study,[74] provided in response to clarification questions, also reports median (range) values for change in hepatic enzymes. These values show a wide range and are not clearly supportive of a treatment effect: The median (range) change in ALT (U/L) from baseline to 12 months of treatment was -35.0 (-368.0 to 293.5) for GL patients, -3.0 (-36.0 to12.0) for the PL subgroup and -0.5 (-56.0 to 80.0) for all PL patients; the median (range) change in AST (U/L) from baseline to 12 months of treatment was -20.5 (-331.0 to 734.0) for GL patients, -2.0 (-51.0 to 12.0) for the PL subgroup and -1.5 (-65.0 to 54.0) for all PL patients.

Similar results were reported for the smaller FH101 study (see Table 16).[1, 38]

Table 16: Hepatic enzymes results from FH101 study

Change from baseline to Month 12 in liver transaminase levels (FAS population)				
		GL N = 9	PL subgroup[a] N = 7	PL overall N = 29
ALT (U/L)				
Baseline	**n**	9	7	29
	Mean (SD)	122.1 (140.47)	35.3 (16.64)	40.7 (34.37)
Actual change from **baseline**	**n**	4	5	19
	Mean (SD)	-191.5 (167.27)	-5.1 (12.94)	-7.4 (25.80)
AST (U/L)				
Baseline	**n**	9	7	29
	Mean (SD)	76.0 (72.52)	27.7 (8.98)	35.9 (28.44)
Actual change from **baseline**	**n**	4	5	19
	Mean (SD)	-104.1 (74.18)	-0.3 (7.21)	-3.6 (24.81)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; FAS, full analysis set; GL, generalised lipodystrophy; PL, partial lipodystrophy; SD, standard deviation

[a] PL subgroup = patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

The CS states that a total of 21 patients with GL and eight patients in the PL subgroup had liver volume assessed at baseline and at least one post-baseline assessment,[1, 37] 20 of 21 patients with GL and six of the eight patients in the PL subgroup had hepatomegaly (liver volume >2000 mL). Reductions in liver volume were observed at all post-baseline assessments in 15 (71%) of the 21 patients with GL who could be assessed for changes from baseline and an additional four patients had reductions at all assessments on or after Month 12. Reductions in liver volume for these 19 patients ranged from 7% to 71%, with most patients (12 of 19) having reductions of ≥30%. Among the eight patients in the PL subgroup, four (50%) had reductions observed at all post-baseline assessments and an additional patient had reductions at all assessments on or after Month 12. Reductions in liver volume for these five patients ranged from 8% to 51%. Among paediatric patients, reductions from baseline were observed at all assessments in 10 (77%) of 13 patients with data available, all with GL; the remaining three patients had reductions at all assessments after Month 12. Reductions ranged from 7% to 64% with most of these paediatric patients (eight of 13) having reductions ≥30%.[1, 37]

ERG comment: The median (range) of observed change in liver volume (mL) from baseline to month 12 of treatment, taken full CSR for the NIH 991265/20010769 study[74] provided in response to clarification questions, was -34.8 (-53.9 to -10.0) for GL patients (n=12), -15.8 (-21.2 to 4.4 for the PL subgroup (n=7) and -16.7 (-21.2 to 4.4) for all PL patients (n=8).

Results of paired liver biopsies from 27 patients in Study NIH 991265/20010769 were reported in the publication by Safar-Zadeh et al.2013[24] Of these 27 patients, 86% had borderline or definite NASH at baseline and 33% had NASH after leptin replacement for 25.8 ± 3.7 months

(p = 0.0002).[24] Significant improvements were observed in steatosis grade and ballooning injury scores with a reduction in the NAFLD activity score during long-term treatment with metreleptin in patients with NASH.[1, 24, 69] Patients with liver fibrosis at baseline remained stable on metreleptin.[1, 24, 69]

ERG comment: The CS lacks long-term data about the effects of metreleptin on the development and progression of liver disease. The ongoing studies section of the CS (CS, page 27) states that: 'The NIH Follow-Up study has allowed for consideration of longer history and follow-up across a range of outcomes not fully studied in the clinical trial. While the retrospective and observational nature of this single-arm study is acknowledged, a wealth of information about these patients' experiences with LD both before and after initiation with metreleptin has been reported, including outcomes such as hyperphagia, female reproductive dysfunction, damage to key organ systems, and death.'[1] However, no results from this study are reported in the clinical effectiveness sections of the CS; a study report was provided in response to clarification questions.[46] This report defined an improvement in liver abnormality as at least a 20% reduction in AST/ALT at one year post-metreleptin initiation, in patients who had evidence of pre-treatment liver abnormalities, and no additional emergent liver abnormalities during that year; liver abnormalities included hepatomegaly, any form of fatty liver or steatosis, fibrosis, cirrhosis, and hepatitis (see section 4.2.1, Table 6) and only 56/105 (53%) of patients who were classified as having pre-treatment liver abnormalities also had elevated hepatic enzymes. Of the 63 GL patients with evidence of pre-metreleptin liver abnormalities, 32 (51%) were classified as having post-metreleptin improvement, compared to 6/42 (14%) for PL patients; no data were reported for the PL subgroup.[46] It should be noted that, whilst these data appear to be evidence that metreleptin treatment is associated with improvements in liver function, a decrease in AST/ALT levels, set at an apparently arbitrary threshold of 20%, is a surrogate outcome measure and is unlikely to be an adequate indicator of long term clinical outcomes. Of the five GL patients who had no evidence of liver abnormalities before metreleptin treatment, two (40%) had emergent liver abnormalities after metreleptin initiation; there were no emergent liver abnormalities in the PL population.[46] No indication of mean/median length of follow-up was provided.

ERG comment: The CS did not report any comparator results for development and progression of liver disease (from the GL/PL natural history study); a study report was provided in response to clarification questions.[40] This report included information on the number of patients with liver damage (including chronic hepatitis, mild to severe fibrosis, cirrhosis, hepatic steatosis, hepatomegaly, and transplant) at baseline; the baseline period was defined as birth to first known date of GL or PL diagnosis (see Section 4.2.1, Table 6) and the number of patients with liver abnormalities over the whole observation period, including baseline and follow-up period (time from first known date of GL or PL diagnosis to date of chart abstraction, death or loss to follow-up). The mean follow-up period for GL patients was 8.8 years and the mean follow-up period for PL patients was 5.7 years.[40] Over the whole observation period, 50/56 (89.3%) of GL patients and 79/122 (64.8%) of PL patients were found to have liver damage.[40] Using the reported data for the baseline period and the whole observation period, it is possible to calculate the proportion of patients who did not have liver damage at baseline,

but developed liver damage during the follow-up period (after GL/PL diagnosis). Of the 41 GL patients who did not have liver damage at baseline 35 (85.4%) developed liver damage during follow-up and 52/95 (54.7%) of PL patients who did not have liver damage at baseline developed damage during follow-up.

Other organ damage (heart and kidneys)

The clinical effectiveness section of the CS does not include any evidence about the effects of metreleptin treatment on the development or progression of heart or kidney damage.[1]

ERG comment: In the study report for the NIH follow-up study[46] a patient's heart abnormality was considered to have improved at one year post-metreleptin initiation if they were classified as pre-hypertensive (systolic <140 or ≥120 or diastolic <90 or ≥80) at baseline and normal (systolic <120 and diastolic <100) at one year and had no additional emergent heart conditions during that year.[46] Based on these criteria, 11/36 (31%) of GL patients and 1/14 (7%) of PL patients were classified as having experienced an improvement in their heart abnormality over one year of metreleptin treatment. However, it should be noted that heart abnormalities included hypertrophy, any dilation, any regurgitation, cardiomyopathy, and tachycardia and only 27/50 (54%) of patients with a pre-treatment heart abnormality were also classified as hypertensive or pre-hypertensive; one year changes in blood pressure alone are unlikely to provide an adequate indicator of long term clinical improvement/progression for the conditions listed. Of the 32 GL patients who had no evidence of heart abnormalities before metreleptin treatment, nine (28%) had emergent heart abnormalities after metreleptin initiation, and 6/30 (20%) of PL patients who had no evidence of heart abnormalities before treatment had emergent abnormalities after metreleptin initiation. [46] No indication of mean/median length of follow-up was provided.

Similarly, the study report for the NIH follow-up study[46] defined one year post-metreleptin improvement in kidney abnormalities as a 20% reduction in 24 hour protein excretion, where elevated 24 hour protein excretion was present at baseline, and no additional emergent kidney conditions. Based on these criteria, 19/46 (41%) of GL patients and 4/25 (16%) PL patients were classified as having experienced an improvement in their kidney abnormality over one year of metreleptin treatment. However, it should be noted that kidney abnormalities included proteinuria, enlarged kidneys, nephropathy, hydronephrosis, renal disease, nephromegaly, renal failure, renal calculus, and glomerulosclerosis and only 38/74 (51%) of patients with a pre-treatment kidney abnormality also had elevated 24 hour protein excretion; one year changes in 24 hour protein excretion alone are unlikely to provide an adequate indicator of long term clinical improvement/progression for the conditions listed. Of the 22 GL patients who had no evidence of kidney abnormalities before metreleptin treatment, eight (36%) had emergent kidney abnormalities after metreleptin initiation, and 4/19 (21%) of PL patients who had no evidence of heart abnormalities before treatment had emergent abnormalities after metreleptin initiation.[46] No indication of mean/median length of follow-up was provided.

ERG comment: The CS did not report any comparator results for development and progression of heart or kidney damage (from the GL/PL natural history study); a study report was provided in response to clarification questions.[40] This report included information on the

number of patients with kidney damage (including albuminuria, nephropathy, proteinuria, kidney failure requiring dialysis or transplant, and transplant) and heart damage (including angina, atherosclerosis, atrial fibrillation, cardiac arrhythmia, cardiomyopathy, heart failure, ischemia, left ventricular hypertrophy, myocardial infarction, and transplant) at baseline, (see Section 4.2.1, Table 6). Over the whole observation period, 28/56 (50.0%) of GL patients and 49/122 (40.2%) of PL patients were found to have kidney damage, and 22/56 (39.3%) of GL patients and 37/122 (30.3%) of PL patients were found to have heart damage.[40] Using the reported data for the baseline period and the whole observation period, it is possible to calculate the proportion of patients, who did not have organ damage at baseline, but developed kidney or heart damage during the follow-up period (after GL/PL diagnosis). Of the 52 GL patients who did not have kidney damage at baseline 24 (46.2%) developed kidney damage during follow-up and 35/108 (32.4%) of PL patients who did not have kidney damage at baseline developed damage during follow-up. Using the same approach, of the 48 GL patients who did not have heart damage at baseline 14 (29.2%) developed heart damage during follow-up and 27/112 (24.1%) of PL patients who did not have heart damage at baseline developed damage during follow-up.

Hyperphagia

The CS reports results from an additional publication of the NIH 991265/20010769 study, by Moran et al. 2004[68] This article reports food intake data for 8/14 metreleptin-treated patients LD; mean (SD) food intake in these patients decreased from 3,170 (436) kcal/day at baseline to 1,739 (162) kcal/day at four months.[68]

ERG comment: This study also reported mean (SD) food intake at 12 months (n=6) and these data indicated a subsequent increase in food intake to 2,015 (410) kcal/day (not significantly different from baseline.

The CS also reports results from a further publication, by McDuffie et al. 2004,[26] which assessed satiation (the time to voluntary cessation of eating from a standardised food array after a 12-hour fast) and satiety (the time to hunger sufficient to consume a complete meal after consumption of a standardised preload) in eight female patients with hypoleptinemia, from the NIH 991265/20010769 study. Metreleptin treatment mean (SD) decreased satiation time from 41.2 (18.2) to 19.5 (10.6) min, increased mean (SD) satiety time from 62.9 (64.8) to 137.8 (91.6) min, decreased mean (SD) energy consumed to produce satiation from 2034 (405) to 1135 (432) kcal, and decreased the amount of food desired in the post-absorptive state.[26]

This study is not listed in the included publications provided by the company (see Section 4.2.1, Table 3). The ERG has added the numerical results from this study (not provided in the CS) to the above text.

The CS does not include any data on hyperphagia from the NIH follow-up study. The study report for the NIH follow-up study,[46] provided in response to clarification questions, states only that 'hyperphagia is determined by notes in the medical charts'; no objective measures (e.g. calorie intake or standardised measures of satiation) are reported.[46] At baseline, 57/68 (84%) of GL patients and 31/44 (70%) of PL patients were classified as having hyperphagia.

Similarly, the NIH follow-up study states that 'improvement in hyperphagia is determined by improvement as indicated in post-metreleptin notes' and specifies that patients must have at least one year of post-metreleptin data in order to be included in the improvement count.[46] Based on this definition, 47 (89%) of the 53 GL patients and 25/26 (96%) of PL patients who had hyperphagia at baseline and who had at least one year of post-metreleptin data were classified as having experienced improvements in hyperphagia.[46] Whilst these results appear to indicate that metreleptin treatment is associated with improvements in hyperphagia, it should be noted that no objective measures of hyperphagia were reported and no details were provided about the nature of the hyperphagia information recorded in notes.

ERG comment: The CS did not report any comparator results for hyperphagia and the GL/PL natural history study did not report any information about hyperphagia.[40]

Concomitant medication use

The CS included some information, from the NIH 991265/20010769 study, about discontinuation of insulin, oral antidiabetics, or lipid-lowering therapies following initiation treatment with metreleptin.[1, 37] Sixteen (41%) of 39 patients with GL who were receiving insulin at baseline were able to discontinue insulin use after starting metreleptin and seven (22%) of 32 patients who were receiving oral antidiabetic medications at baseline were able to discontinue use of these drugs. Among the 34 patients who were receiving lipid-lowering therapies at baseline, eight (24%) were able to discontinue these medications.[1, 37] In the PL subgroup, one patient was able to discontinue the use of oral antidiabetic medications and one was able to discontinue the use of lipid-lowering therapies.[1, 37]

ERG comment: The CS also states that: 'Many of these patients could discontinue the use of these therapies within the first 12 months of metreleptin treatment.' However, no times to discontinuation are reported.

The CS does not include any data on concomitant medication use from the NIH follow-up study. The study report for the NIH follow-up study,[46] reported that 57/68 (83.8%) of GL patients and 43/44 (97.7%) of PL patients were on anti-diabetic medication (insulin or oral anti-diabetics) at baseline.[46] A new anti-diabetic medication was initiated (defined as two or more fills of a medication not present at baseline), after the start of metreleptin treatment, in 54/68 (79.4%) of GL patients and 36/44 (81.8%) of PL patients.[46] The equivalent data for lipid lowering medication showed that 28/68 (41.2%) of GL patients and 30/44 (68.2%) of PL patients were on lipid-lowering medication (statin and/or fibrates) at baseline.[46] A new lipid-lowering medication was initiated (defined as two or more fills of a medication not present at baseline), after the start of metreleptin treatment, in 18/68 (26.5%) of GL patients and 27/44 (61.4%) of PL patients.[46] Medication discontinuation was defined as a 12-month period without any medication prescription fills and included both baseline medications and medications initiated after the start of metreleptin treatment; 41/64 (64.1%) of GL patients and 15/44 (34.1%) of PL patients were able to discontinue antidiabetic medications.[46] Most discontinuations were for bolus insulin or metformin, only two GL patients discontinued basal insulin or insulin + metformin.[46] With respect to lipid-lowering medication, 19/35 (54.3% of GL patients and 16/38 (68.2%) of PL patients were able to discontinue lipid lowering

medications.[46] The majority of discontinuations, 26/35, were for fibrates, with few patients discontinuing statin use.[46]

Growth and development

The CS includes some information, from the NIH 991265/20010769 study, about growth and development in metreleptin treated patients.[1, 37] This study assesses stature at screening/baseline and at least one post-baseline time point in 40 children (<18 years of age), including 36 patients with GL and four patients with PL (two in the PL subgroup). Among the 36 GL patients, 22 were reported to have normal stature at study entry, 10 had tall stature for their age, and four had short stature. Overall 16 (44%) of the 36 patients were reported to have had growth complete or near complete prior to entry. Among the other 20 patients, 10 were reported to have normal growth (including five with normal stature, three who were tall and two who were short at baseline), two had growth acceleration (one with normal stature and one with short stature), and eight had growth deceleration (five with normal stature and three who were tall). Among the four PL patients with data available, two patients (in the PL subgroup) had growth complete or near complete at study entry. Among the other two patients, one had short stature at baseline with growth deceleration reported on metreleptin and one had tall stature at baseline with normal growth on metreleptin.[1, 37]

Overall 33 patients <18 years of age had pubertal status assessed at baseline, including 27 patients with GL and six patients with PL (five in the PL subgroup); 26 of these patients had puberty complete, near complete, or probably complete (based on growth data) prior to metreleptin. Among the other seven patients, all with GL, four had delayed puberty prior to metreleptin and three had precocious puberty; follow-up was available for three of these patients, all with delayed puberty at entry (two had normal development on metreleptin and one continued to have delayed puberty). Among the 14 patients without baseline data reported who were not prepubertal (normal for age), 13 reported normal pubertal onset and/or progression on metreleptin at a post-baseline assessment and one had delayed onset reported.[1, 37]

ERG comment: The NIH follow-up study[46] did not report any additional information about the growth and development of metreleptin-treated patients. The GL/PL natural history study[40] does not include any information about growth and development.

Reproductive dysfunction

The clinical effectiveness section of the CS does not include any evidence about the effects of metreleptin treatment on reproductive dysfunction.[1]

ERG comment: Two publications,[58, 61] listed as included publications related to the NIH 991265/20010769 study (see Section 4.2.1, Table 3) reported assessments of the effects of metreleptin treatment on female reproductive dysfunction. In one study, 10 female patients with GL showed a mean (SD) decrease in serum free testosterone from 39.6 (11) to 18.9 (4.5) ng/dL following metreleptin treatment; ovarian ultrasound showed a polycystic ovarian disease pattern in all patients that did not change after therapy, and eight of the 10 patients had amenorrhea prior to therapy and all eight developed normal menses after therapy.[61] The second

study included seven female patients with severe LD; five of these patients had intact reproductive systems and only one was cycling normally at the start of metreleptin treatment, but all five had normal menses by the fourth month of treatment.[58] The results from these two publications were not included in the CS.

The NIH follow-up study[46] also reports information about the effects of metreleptin treatment on female reproductive dysfunction. The report defined disruption to the female reproductive system as the presence of irregular menstruation or polycystic ovary syndrome (PCOS). Female patients are not considered to have disruption to female reproductive function if they are experiencing menopause, are prepubescent, or had surgical removal of reproduction organs. At baseline, 21/27 (78%) of relevant female GL patients and 24/29 (83%) of relevant female PL patients were classified as experiencing reproductive dysfunction.[46] Twelve (57%) of the 21 effected GL patients and eight (33%) of the 24 effected PL patients were reported as having post-metreleptin improvement ('improvement in any of irregular menstruation or PCOS').[46] However, no definition of the criteria used to determine improvement was provided.

The CS did not report any comparator results for reproductive dysfunction (from the GL/PL natural history study); a study report was provided in response to clarification questions and this report includes information about female reproductive dysfunction in LD patients.[40] This report included information on the number of female patients with reproductive dysfunction (including amenorrhea, menstruation <6 times per year, pregnancy loss, infertility or subfertility, ovarian cysts, and PCOS) at baseline, (see Section 4.2.1, Table 6). Over the whole observation period, 2/15 (13.3%) of female GL patients and 15/41 (36.6%) of female PL patients were found to have reproductive dysfunction.[40] Using the reported data for the baseline period and the whole observation period, it is possible to calculate the proportion of patients, who did not have reproductive dysfunction at baseline, but developed problems during the follow-up period (after GL/PL diagnosis). Of the 13 female GL patients who did not have reproductive dysfunction at baseline, nine (69.2%) developed reproductive dysfunction during follow-up and 19/26 (73.1%) of female PL patients who did not have reproductive dysfunction at baseline developed problems during follow-up.

Pancreatitis

The clinical effectiveness section of the CS does not include any information about the effects of metreleptin treatment on pancreatitis; pancreatitis is only reported as an adverse event occurring subsequent to metreleptin withdrawal (CS, section 9.7.2.5, page 114).

ERG comment: The NIH follow-up study[46] reports information about the effects of metreleptin treatment on pancreatitis. A patient was considered to have pancreatitis at baseline if they had ≥1 episodes of pancreatitis in the one year prior to metreleptin initiation.[46] At baseline, 21/63 (31%) of GL patients and 23/44 (52%) of PL patients had a history of pancreatitis.[46] Improvement in pancreatitis was defined as no recorded episodes of pancreatitis post-metreleptin initiation or only episodes of pancreatitis which were due to non-compliance.[46] Based on these criteria, 20/21 (95%) of effected GL patients and all effected PL patients experienced improvements in pancreatitis on metreleptin treatment. These data were not included in the CS, but are of particular importance given the identified risk of pancreatitis

following metreleptin withdrawal; it is important to consider the extent to which this risk may be balanced by any reduction in the incidence of pancreatitis that may occur in patients on treatment.

The CS did not report any comparator results for pancreatitis (from the GL/PL natural history study).[40] This report included information on the number of patients with pancreatic damage (all pancreatitis) at baseline, (see Section 4.2.1, Table 6). Over the whole observation period (including baseline and follow-up), 7/56 (12.5%) of GL patients and 20/122 (16.4%) of PL patients experienced at least one episode of pancreatitis.[40] Five (71.4%) of the 7 effected GL patients and 12/20 (60.0%) of effected PL patients experienced pancreatitis during the follow-up period (after GL/PL diagnosis).

Health-related quality of life including effects on appearance and activities of daily living

The clinical effectiveness section of the CS does not include any information about the effects of metreleptin treatment on measures of health-related quality of life.

ERG comment: The NIH follow-up study[46] reports information about the effects of metreleptin treatment on impaired physical appearance and ability to perform work/school work. Impaired physical appearance was defined as the presence of acanthosis nigricans, hyperkeratosis, or hirsutism; at baseline, 56/68 (82%) of GL patients and 30/44 (68%) of PL patients were classified as having impaired physical appearance.[46] Thirty-eight (68%) of the 56 effected GL patients and 14 (47%) of the 30 effected PL patients were reported as having post-metreleptin improvement ('improvement in any of acanthosis nigricans, hyperkeratosis, or hirsutism').[46] However, no definition of the criteria used to determine improvement was provided. Loss of ability to perform work/school work was defined as incomplete school attendance due to disease symptoms for school age patients or not working/working part-time due to disease symptoms; at baseline 39/68 (57%) of GL patients and 9/44 (20%) of PL patients were effected.[46] Improvement in loss of ability to perform work/school work is defined as complete school attendance for school-age patients or the ability for a patient to work, even if the patient has chosen not to work; 31/39 (79%) effected GL patients and 5/9 (56%) of effected PL patients experienced improvements in their ability to perform work or school work whilst on metreleptin treatment.[46]

The CS did not report any comparator results for impaired physical appearance or ability to perform activities of daily living (from the GL/PL natural history study).[40] This report included information on the numbers of patients characteristics of physical appearance associated with lipodystrophy; only one of the three characteristics included in the NIH follow-up study definition of impaired physical appearance (acanthosis nigricans) was also recorded in the GL/PL natural history study. Acanthosis nigricans was present in 29 (56.9%) of the 51 GL patients and 39 (37.7%) of the 105 PL patients in the GL/PL natural history study, for whom information was available.[40] The GL/PL natural history study did not include any information about the ability of LD patients to perform activities of daily living.

Mortality

Survival data for LD patients, from the GL/PL natural history study[40] and for patients on metreleptin treatment, from the NIH follow-up study[46] are used in the cost effectiveness analyses presented in the CS,[1] but no survival data are presented in the clinical effectiveness section of the CS.

ERG comment: For information, we have reproduced the mortality tables from both the NIH follow-up[46] and GL/PL natural history[40] studies (Tables 17 and 18 below).

Table 17: Mortality and cause of death data from the NIH follow-up study

	All Patients (n=112)	GL Patients (n=68)	PL Patients (n=44)
Age at metreleptin initiation			
Mean (SD)	24.3 (15.4)	17.5 (11.4)	34.6 (15.2)
Median (IQR)	18.2 (14.0, 34.6)	15.4 (11.9, 20.2)	34.6 (18.9, 45.9)
Years from metreleptin initiation to last known status*			
Mean (SD)	8.4 (4.5)	8.8 (4.7)	7.7 (4.2)
Median (IQR)	7.6 (4.5, 11.7)	8.1 (5.3, 12.3)	5.6 (4.3, 10.8)
Age at last known status*			
Mean (SD)	32.6 (16.2)	26.3 (12.9)	42.4 (16.2)
Median (IQR)	27.1 (20.5, 44.7)	24.3 (18.9, 29.2)	42.6 (28.7, 56.2)
Patients still alive, n (%)$			
Yes	94 (83.9)	55 (80.9)	39 (88.6)
No	13 (11.6)	12 (17.6)	1 (2.3)
Uncertain	5 (4.5)	1 (1.5)	4 (9.1)
Years from first GL/PL symptoms to death			
Kaplan-Meier Mean (SE)	15.4 (0.5)	14.7 (0.7)	16.7 (0.3)
Patients who died, n	13	12	1
Age at metreleptin initiation			
Mean (SD)	24.2 (15.3)	23.9 (16.0)	27.7 (NA)
Median (IQR)	17.7 (15.1, 27.7)	17.4 (14.9, 27.7)	27.7 (NA)
Years from metreleptin initiation to death			
Mean (SD)	6.3 (4.9)	6.5 (5.0)	3.4 (NA)
Median (IQR)	4.3 (1.9, 10.6)	4.8 (1.8, 11.2)	3.4 (NA)
Age at death			
Mean (SD)	30.5 (15.6)	30.4 (16.2)	31.2 (NA)
Median (IQR)	25.3 (20.1, 31.2)	24.5 (19.7, 37.4)	31.2 (NA)
Potential contributing factors, n (%)			
End stage liver disease	4 (30.8)	4 (33.3)	0 (0.0)
End stage renal disease	2 (15.4)	2 (16.7)	0 (0.0)
Cardiac failure	2 (15.4)	2 (16.7)	0 (0.0)
Cardiac failure and kidney failure	1 (7.7)	1 (8.3)	0 (0.0)
Hepatorenal failure	1 (7.7)	1 (8.3)	0 (0.0)
Lymphoma	1 (7.7)	1 (8.3)	0 (0.0)
Respiratory failure	1 (7.7)	0 (0.0)	1 (100)
Unknown	1 (7.7)	1 (8.3)	0 (0.0)

GL, generalized lipodystrophy; IQR, interquartile range; NIH, National Institutes of Health; NA, not applicable; PL, partial lipodystrophy; SD, standard deviation; SE, standard error

*Last known status is the latest date in which patient status is known

$Status of patient as of 12/18/2017

Table 18: Mortality and cause of death data from the GL/PL natural history study

	All Patients (n=178)	GL Patients (n=56)	PL Patients (n=122)
Years from first GL/PL symptoms to end of observation period*			
Mean (SD)	14.7 (13.3)	12.7 (10.5)	15.5 (14.4)
Median (IQR)	10.0 (3.9, 21.4)	10.1 (3.5, 18.0)	9.6 (4.0, 23.1)
Years from first GL/PL symptoms to diagnosis			
Mean (SD)	8.0 (11.4)	3.9 (7.4)	9.8 (12.5)
Median (IQR)	2.5 (0.0, 12.0)	0.0 (0.0, 3.8)	5.0 (0.0, 15.9)
Patients still alive, n (%)			
Yes	142 (79.8)	37 (66.1)	105 (86.1)
No	14 (7.9)	8 (14.3)	6 (4.9)
Unknown	22 (12.4)	11 (19.6)	11 (9.0)
Years from first GL/PL symptoms to death**			
Kaplan-Meier Mean (SE)	48.0 (2.2)	29.8 (1.8)	52.5 (1.9)
Median (IQR)	56.3 (34.5, NR)	31.7 (30.7, NR)	56.3 (56.3, NR)
Patients who died, n	14	8	6
Age at first GL/PL symptoms			
Mean (SD)	24.9 (21.2)	16.9 (20.6)	35.6 (18.3)
Median (IQR)	20.5 (6.5, 49.5)	8.3 (2.3, 32.1)	29.7 (26.0, 55.4)
Age at death			
Mean (SD)	45.3 (17.2	38.2 (16.0)	54.8 (14.8)
Median (IQR)	42.0 (31.3, 62.5)	31.7 (28.2, 52.4)	57.9 (39.5, 69.0)
Death related to lipodystrophy, n (%)			
Yes	7 (50.0)	6 (75.0)	1 (16.7)
No	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	7 (50.0)	2 (2.0)	5 (83.8)
Patients who died, n	14	8	6
Cause of death reported,$ n (%)	10 (71.4)	8 (100)	2 (33.3)
Method of assessing cause of death, n (%)			
Per practice health records	3 (21.4)	1 (12.5)	2 (33.3)
Per physician recollection	5 (35.7)	4 (50.0)	1 (16.7)
From death certificate	2 (14.3)	2 (25.0)	0 (0.0)
Not confirmed	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	4 (28.6)	1 (12.5)	3 (50.0)
Potential contributing factors, n (%)			
Bone marrow/hematologic abnormalities	0 (0.0)	0 (0.0)	0 (0.0)
Cancer	0 (0.0)	0 (0.0)	0 (0.0)
Cardiovascular event	4 (28.6)	3 (37.5)	1 (16.7)
Cerebrovascular disease	3 (21.4)	1 (12.5)	2 (33.3)
Immunosuppression	1 (7.1)	1 (12.5)	0 (0.0)
Infection (viral)	0 (0.0)	0 (0.0)	0 (0.0)
Infection (bacterial)	1 (7.1)	1 (12.5)	0 (0.0)
Liver disease	3 (21.4)	2 (25.0)	1 (16.7)
Pancreatitis	1 (7.1)	1 (12.5)	0 (0.0)
Pneumonia	1 (7.1)	1 (12.5)	0 (0.0)
Renal failure	0 (0.0)	0 (0.0)	0 (0.0)
Sepsis	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	5 (35.7)	1 (12.5)	4 (66.7)

	All Patients (n=178)	GL Patients (n=56)	PL Patients (n=122)
Other[$$]	1 (7.1)	1 (12.5)	0 (0.0)
Location where patient died, n (%)			
At home	1 (7.1)	1 (12.5)	0 (0.0)
At the hospital	7 (50.0)	5 (62.5)	2 (33.3)
Unknown	5 (35.7)	1 (12.5)	4 (66.7)
Other	1 (7.1)	1 (12.5)	0 (0.0)

[*]The end of the observation was defined as the earliest of: date of chart abstraction; death; loss to follow-up
[**]In order to account for censoring due the end of data availability, the average time to death was calculated using the Kaplan-Meier estimate
[$]Causes of death included mentions of cardiac arrest, death following coronary artery bypass graft, diabetic foot infection, heart failure related to valvular stenosis, hospitalisation for kidney failure, multiple diagnoses (atypical interstitial pneumonitis, progressive CGL with insulin resistance, hepatosplenomegaly, thrombocytopenia, polycythemia, acanthosis nigricans, hypertriglyceridemia), myocardial infarction, possible cardiac episode, probable end stage liver disease, and stroke
[$$]Other potential contributing factors of death included mentions of pancytopenia, steatohepatitis, and chronic renal insufficiency

Safety and tolerability

The CS states that the safety profile of metreleptin in patients with LD is consistent with that of a patient population with significant co-morbidities.[1] The CS further states that long-term exposure available from clinical trials across a relatively large population of patients with this ultra-rare disease provides guidance on the expected safety profile of this agent intended for chronic therapy in patients with GL and in a subgroup of patients with PL who have more significant baseline metabolic disturbances of HbA_{1c} ≥6.5% and triglycerides ≥5.65 mmol/L.[1]

The CS refers to further data from the post-marketing period from 138 patients who have been exposed worldwide to commercially available metreleptin (including 116 in the US and 22 in Japan) has shown a safety profile that is consistent with that observed in clinical trials with no new safety signals identified. The identified risks of hypersensitivity, acute pancreatitis associated with metreleptin discontinuation, and hypoglycaemia with concomitant use of insulin and insulin secretagogues can be managed with risk communication in labelling and educational activities.[33, 37, 38] The CS states that in conclusion, the known side effects of metreleptin can be managed as part of the normal clinical practice for patients with this complex condition.

ERG comment: The CS does not include any mention of the safety concerns highlighted in the Centre for Drug Evaluation and Research Report (not included in the CS) or the associated Risk Evaluation Management Strategy (REMS).[75] The summary of safety in this report states: 'The principal safety concerns with metreleptin are T-cell lymphoma and anti-metreleptin antibodies with neutralizing activity. These concerns are of sufficient magnitude to require REMS. Other safety findings that warrant inclusion in the Warning and Precautions section of the metreleptin labelling include hypoglycemia, autoimmunity, and hypersensitivity.'

The CS provides no reference for the data described as post-marketing period from 138 patients who have been exposed worldwide to commercially available metreleptin (including 116 in the US and 22 in Japan).[1]

Adverse events

Study NIH 991265/20010769

A summary of treatment-emergent adverse events (TEAEs) is shown in Table 19, below (reproduced from the CS, Table C25, pages 108-109).[1] In the GL group, 59 (89%) of the 66 patients reported at least one TEAE; drug-related TEAEs were reported in 32 (49%) of these patients.[37] Compared with the GL group, the overall incidence of TEAEs was similar in the PL subgroup with 27 (87%) of the 31 patients experiencing at least 1 TEAE; the incidence of drug-related TEAEs was lower (23%).

TEAEs of severe intensity were reported in 29 (44%) of the 66 GL patients and in 13 (42%) of the 31 patients in the PL subgroup; most severe TEAEs were assessed as unrelated to study treatment.[37]

Overall, 23 (35%) of the 66 GL patients and 7 (23%) of the 31 patients in the PL subgroup experienced a treatment-emergent SAE.[37] The types of SAEs were consistent with the underlying LD disease, and primarily included reports of abdominal pain and pancreatitis, infections, and worsening liver function. Drug-related SAEs were not common, reported in three GL patients, including one case of hypertension, one of respiratory distress and one case of anaplastic large-cell lymphoma. None of the patients in the PL subgroup experienced a drug-related SAE.

Discontinuations due to TEAEs were reported in five patients with GL (8%) and one patient in the PL subgroup (3%). In four of these six patients, the events leading to withdrawal led to death.[37]

The majority of the commonly reported events in the GL group were consistent with the expected pharmacologic effects of metreleptin, including weight loss, hypoglycaemia, and decreased appetite, or were gastrointestinal (GI) disorders or constitutional symptoms, including abdominal pain and headache.[37] Other commonly reported GI disorders in patients with GL included nausea and constipation. The most commonly reported drug-related TEAEs in GL patients were weight decreased (15 patients, 23%) and hypoglycaemia (eight patients, 12%).

In general, the safety profile in the PL subgroup was consistent with that observed in the overall GL group. The most common TEAEs reported in the PL subgroup were abdominal pain, hypoglycaemia, nausea, fatigue, alopecia and constipation. The most commonly reported drug-related TEAEs in patients in the PL subgroup were hypoglycaemia and fatigue (each three patients, 10%).[37]

Table 19: Adverse events: study NIH 991265/20010769 (safety analysis set)

	GL (N = 66)	PL subgroup[a] (N = 31)	PL overall (N = 41)
Overall Summary			
TEAE	59 (89.4)	27 (87.1)	35 (85.4)
Drug-related TEAE	32 (48.5)	7 (22.6)	8 (19.5)
Severe TEAE	29 (43.9)	13 (41.9)	16 (39.0)
Drug-related severe TEAE	7 (10.6)	0	0
Treatment-emergent SAE	23 (34.8)	7 (22.6)	10 (24.4)
Drug-related treatment emergent SAE	3 (4.5)	0	0
TEAE leading to study drug discontinuation	5 (7.6)	1 (3.2)	1 (2.4)
On-study deaths	3 (4.5)	1 (3.2)	1 (2.4)
Most common (≥5% Incidence overall) TEAE			
Weight decreased	17 (25.8)	2 (6.5)	2 (4.9)
Abdominal pain	11 (16.7)	6 (19.4)	6 (14.6)
Hypoglycaemia	10 (15.2)	6 (19.4)	7 (17.1)
Decreased appetite	8 (12.1)	1 (3.2)	1 (2.4)
Headache	8 (12.1)	0	0
Nausea	6 (9.1)	5 (16.1)	6 (14.6)
Fatigue	6 (9.1)	3 (9.7)	3 (7.3)
Ear infection	6 (9.1)	0	0
Arthralgia	6 (9.1)	2 (6.5)	3 (7.3)
Upper respiratory tract infection	5 (7.6)	1 (3.2)	2 (4.9)
Back pain	5 (7.6)	2 (6.5)	2 (4.9)
Anxiety	5 (7.6)	0	1 (2.4)
Proteinuria	5 (7.6)	0	1 (2.4)
Ovarian cyst	5 (7.6)	0	1 (2.4)
Depression	4 (6.1)	1 (3.2)	3 (7.3)
Alopecia	3 (4.5)	3 (9.7)	3 (7.3)
Constipation	3 (4.5)	3 (9.7)	3 (7.3)
Pain in extremity	3 (4.5)	2 (6.5)	3 (7.3)

Abbreviations: GL = generalised lipodystrophy; PL = partial lipodystrophy; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment-emergent adverse event

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

ERG comment: The CS states that the total patient-years of exposure for GL patients was 328.3 years and the median actual duration of treatment (excluding dose interruptions) was 47.2 months.[1, 37] The total patient-years of exposure for PL subgroup patients was 121.3 years and the median actual duration of treatment (excluding dose interruptions) was 29.3 months.[1, 37] The CSR for the NIH 991265/20010769 study also notes that: All but one (>99%) of the 107 patients in the safety analysis set (SAS) received six months or more of metreleptin treatment, with 87% (93 patients) receiving >1 year, 72% (77 patients) receiving >2 years, and 54% (58 patients) receiving >3 years of metreleptin. More than one quarter of patients (28%, 30 patients), received more than six years of treatment with metreleptin with 13 (12%) on treatment for 10 years or more.[74] The timescale over which adverse events was reported is not explicitly stated in the CS, but CSR indicates that patients in the SAS received ongoing at six month exposure intervals.[74]

The CS states that overall, 23 (35%) of the 66 GL patients and 7 (23%) of the 31 patients in the PL subgroup experienced at least one serious adverse event (SAE). This appears to be an error as the numbers refer to treatment-emergent SAE not overall SAE. The CS states that in general, the safety profile in the PL subgroup was consistent with that observed in the overall GL group.[1] The ERG group disagrees. In the GL group weight decrease was a TEAE in 25.8% whereas it was 6.4% in the PL subgroup. Similarly, decreased appetite was a TEAE in 12.1% of the GL group and in 6.4% of the PL subgroup. In addition, the ERG would argue that weight decrease in 25.8% of GL group is an undesirable adverse event given the loss of adipose tissue associated with the condition.

Study FHA101

A summary of TEAEs is shown in Table C26 (pages 111-112 of the CS) and replicated in Table 20, below[1].

In the GL group, seven (78%) of the nine patients reported at least one TEAE; drug-related TEAEs were reported in six (67%) of these patients.[38] All seven patients in the PL subgroup experienced at least one TEAE, and TEAEs were assessed as drug-related in six (86%) of these seven patients.

In six (67%) of the nine patients with GL, events of severe intensity were reported. All TEAEs in the PL subgroup were mild to moderate in severity.[38] Among the PL patients not included in the PL subgroup, events of severe intensity were reported in nine (36%) of the 25 patients.

Overall, six (67%) of the nine GL patients experienced at least one SAE, none of which was assessed as related to study treatment.[38] There were no SAEs reported in patients in the PL subgroup. Ten patients with PL who were not in the PL subgroup experienced SAEs.

Discontinuations due to TEAEs were reported in the two patients who died and in two additional patients with PL (not in the PL subgroup).[38]

In general, when considering the difference in sample size, the types and incidence for commonly reported TEAEs in study FHA101 were similar to those reported in the pivotal study

NIH 991265/20010769. Among the nine patients with GL in Study FHA101, the most commonly reported TEAEs, all reported in two patients (22%), were hypoglycaemia, upper respiratory tract infection, abdominal pain, increased liver function tests, and ear infection.[38] For the seven patients in the PL subgroup, the most commonly reported TEAEs were hypoglycaemia, upper respiratory tract infection, and urinary tract infection (each three patients, 43%), and nausea, anxiety, and sinusitis (each two patients, 29%). The only drug-related TEAE reported in more than one GL patient was hypoglycaemia (two patients, 22%). In the PL subgroup, the only drug-related TEAEs reported in more than one patient were hypoglycaemia and nausea (each two patients, 29%).

Table 20: Adverse events: Study FHA101 (safety analysis set)

	GL (N = 9)	**PL subgroup[a] (N = 7)**	**PL overall (N = 32)**
Overall summary			
TEAE	7 (77.8)	7 (100.0)	27 (84.4)
Drug-related TEAE	6 (66.7)	6 (85.7)	22 (68.8)
Severe TEAE	6 (66.7)	0	9 (28.1)
Drug-related severe TEAE	0	0	2 (6.3)
Treatment-emergent SAE	6 (66.7)	0	10 (31.3)
Drug-related treatment emergent SAE	0	0	1 (3.1)
TEAE leading to study drug discontinuation	1 (11.1)	0	3 (9.4)
On-study deaths	1 (11.1)	0	1 (3.1)
Most common (≥5% incidence overall) TEAE (MedDRA preferred term)			
Hypoglycaemia	2 (22.2)	3 (42.9)	11 (34.4)
Upper respiratory tract infection	2 (22.2)	3 (42.9)	6 (18.8)
Urinary tract infection	1 (11.1)	3 (42.9)	6 (18.8)
Nausea	1 (11.1)	2 (28.6)	12 (37.5)
Anxiety	1 (11.1)	2 (28.6)	2 (6.3)
Sinusitis	0	2 (28.6)	2 (28.6)
Liver function test increased	2 (22.2)	1 (14.3)	1 (3.1)
Abdominal pain	2 (22.2)	1 (14.3)	5 (15.6)
Vomiting	1 (11.1)	1 (14.3)	4 (12.5)

	GL (N = 9)	PL subgroup[a] (N = 7)	PL overall (N = 32)
Headache	1 (11.1)	1 (14.3)	4 (12.5)
Injection site bruising	1 (11.1)	1 (14.3)	4 (12.5)
Lymphadenopathy	1 (11.1)	1 (14.3)	3 (9.4)
Dizziness	0	1 (14.3)	3 (9.4)
Muscle spasms	0	1 (14.3)	6 (18.8)
Myalgia	0	1 (14.3)	3 (9.4)
Viral infection	0	1 (14.3)	3 (9.4)
Ear infection	2 (22.2)	0	1 (3.1)
Dyspnoea	1 (11.1)	0	2 (6.3)
Vertigo	0	0	4 (12.5)
Injection site pruritus	0	0	3 (9.4)

Abbreviations: GL = generalised lipodystrophy; PL = partial lipodystrophy; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment-emergent adverse event

[a] PL subgroup = PL subgroup = patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

ERG comment: The CS describes the total patient-years of exposure for GL patients was 11.3 years and the median actual duration of treatment (excluding dose interruptions) was 21.3 months.[1, 37] The total patient-years of exposure for PL subgroup patients was 28.4 years and the median actual duration of treatment (excluding dose interruptions) was 51.3 months.[1, 37] The CSR for the FH101 study also notes that: Overall, 35 (88%) of the 40 patients with data available for exposure received six months or more of metreleptin treatment with 70% (28 patients) receiving >1 year, 45% (18 patients) receiving >2 years, and 35% (14 patients) receiving >3 years of metreleptin in this study. Overall, four patients (10%), received more than five years of treatment with metreleptin.[76] The timescale over which adverse events was reported is not explicitly stated in the CS, but CSR indicates that patients in the safety population received ongoing at six month exposure intervals.[76]

The CS states that overall, six (67%) of the nine GL patients experienced at least one serious adverse event (SAE)[1]. This appears to be an error as the numbers refer to treatment-emergent SAE not SAE.

Paediatric population

The CS reported safety and tolerability with respect to the paediatric population.[1] The CS states that across the two completed clinical studies (NIH 991265/20010769 and FHA101), there were 50 paediatric subjects (five in the PL subgroup and 45 with GL) enrolled and exposed to metreleptin. Limited clinical data exists in children less than six years old.[33]

The CS reports that the overall, the safety and tolerability of metreleptin are similar in children and adults.[1] In GL patients, the overall incidence of drug related adverse reactions was similar

regardless of age. SAEs were reported in 15 paediatric patients, primarily reports of abdominal pain and pancreatitis (each three patients), and pneumonia and liver disorder (each two patients).[33] The only common TEAEs reported at a higher incidence (≥10% difference) in patients ≥6 to <18 years compared to adults were abdominal pain (25% vs 5%) and nausea (15% vs 0%).[33] In PL patients, assessment across age groups is limited, due to the small sample size.[33] However, there were no apparent differences in the overall incidence or the incidence of common adverse events between age categories.[33]

ERG comment: The CS only mentions the paediatric population from the NIH 991265/20010769 study (five in the PL subgroup and 45 with GL). It omits the three paediatric patients who have PL but do not meet the subgroup criteria (patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L). The CS also omits the paediatric population from the FHA101 study. A further three paediatric subjects (in GL) were enrolled and exposed to metreleptin in FHA101.[38]

The CS includes additional information concerning 'selected adverse reactions' (CS, section 9.7.2.5, pages 114-116).[1]

Pancreatitis

Across the 148 patients included in LD studies, six (4%) patients (four with GL and two with PL), experienced treatment emergent pancreatitis.[33, 37, 38] All patients had a history of pancreatitis and hypertriglyceridemia.[33, 37, 38] One of the patients who developed septic shock concurrent with pancreatitis died; the other five patients recovered and continued on treatment.[33, 37, 38] Abrupt interruption and/or non-compliance with metreleptin dosing was suspected to have contributed to the occurrence of pancreatitis in several of these patients. The mechanism for pancreatitis in these patients was presumed to be return of hypertriglyceridemia and therefore increased risk of pancreatitis in the setting of discontinuation of effective therapy for hypertriglyceridemia.[33]

ERG comment: The CS describes abrupt interruption and/or non-compliance with metreleptin dosing was suspected to have contributed to the occurrence of pancreatitis in several of these patients. Tables C18 and C19 (pages 86-87 of the CS) describe the number of premature discontinuations in study NIH 991265/20010769 and study FHA101 respectively.[1] In Table C18 23/66 (34.8%) GL patients; 15/41 (36.6%) PL patients and 11/31 (35.5%) PL subgroup patients prematurely discontinued. In Table C19 4/9 (44.4%) GL patients; 20/32 (62.5%) PL patients and 2/7 (28.6%) PL subgroup patients prematurely discontinued. The numbers of patients who discontinue treatment are alarmingly high given that discontinuation of treatment appears to be associated with an increased risk of pancreatitis.

The Centre for Drug Evaluation and Research Report (not included in the CS),[75] includes the following statement: 'The sponsor argues that the patients who developed pancreatitis were either non-compliant or they discontinued metreleptin therapy too rapidly and induced a rebound in serum TG levels. Dr Golden was unable to confirm the sponsor's assertion and rightly points out that the lack of a control group and the small size of the lipodystrophy database leave unanswered the question of metreleptin's role in the cases of pancreatitis.'

Serious infections

A significant number of patients with acquired forms of LD have low C3 levels and the presence of polyclonal immunoglobulin C3 nephritic factor, increasing the risk for recurrent bacterial infections.[6, 12]

A review of available literature was undertaken to understand the propensity as well as the rate of development of serious infection in patients with LD. The conclusion of this review was that the natural history of patients with LD with low leptin levels is to experience higher rates of infection than the general population.[29, 77-80]

In study NIH 991265/20010769, serious infections were reported in seven (11%) of 66 patients with GL and in two (7%) of 31 patients in the PL subgroup.[37]. The only serious infections reported in more than one patient in the GL group were sepsis and pneumonia, each reported in two patients (3%). In the PL subgroup, serious infections included cellulitis, streptococcal infection, and pharyngitis in one patient and osteomyelitis and cellulitis in the other. All serious infections were assessed as unrelated to study treatment and none led to treatment discontinuation. In study FHA101, no serious infections were reported in the GL group or in the PL subgroup.[38]

ERG comment: The CS[1] states that the natural history of patients with LD with low leptin levels is to experience higher rates of infection than the general population and cites Mancuso 2002 amongst others.[78] Mancuso 2002 is a study of leptin-deficient mice and cannot be cited as evidence in humans.[78] Moon 2013 is also cited in support of patients with LD with low leptin levels who experience higher rates of infection than the general population.[79] Moon 2013 describes leptin's Role in lipodystrophic and nonlipodystrophic Insulin-Resistant and Diabetic Individuals and does not contain any direct evidence in support of this claim.[79]

Hypoglycaemia

Metreleptin may decrease insulin resistance in diabetic patients, resulting in hypoglycaemia in patients with LD and co existing diabetes.[33] Hypoglycaemia, deemed as related to metreleptin treatment, occurred in 13.3% of patients studied. All reports of hypoglycaemia in patients with GL and in the PL subgroup, have been mild in nature with no pattern of onset or clinical sequelae.[33] Generally the majority of events could be managed by food intake with only relatively few modifications of anti-diabetic medicine dosage occurring.[33]

T cell lymphoma

Three cases of T cell lymphoma have been reported while taking metreleptin in clinical studies.[33] All three patients had acquired GL. Two of these patients were diagnosed with peripheral T cell lymphoma while receiving the medicinal product. Both had immunodeficiency and significant haematological abnormalities including severe bone marrow abnormalities before the start of metreleptin treatment. A separate case of anaplastic large cell lymphoma was reported in a paediatric patient receiving the medicinal product who did not have haematological abnormalities before treatment.

ERG comment: The Centre for Drug Evaluation and Research Report (not included in the CS), notes that *in-vitro* and *in-vivo* data indicate that leptin, through activation of tumor-associated leptin receptors, can influence the growth and progression of malignant cells, and includes the following statement: 'According to our colleagues in the Division of Hematology Products, the incidence of T-cell lymphoma in the general population from the United States is 2.3 per 100,000 for males and 1.4 per 100,000 for females. While the incidence of lymphoma in patients from the NIH and FHA101 clinical studies was 5,900 per 100,000 in males and 1,900 per 100,000 in females, these crude estimates are based on a very small sample of patients and therefore have very wide confidence intervals. Moreover, in addition to not knowing if lipodystrophy itself may be associated with an increased risk for lymphoma, two of the three cases of lymphoma were confounded by histories of neutropenia and treatment with G-CSF. Nevertheless, the clinical review team considers the T-cell lymphoma data sufficient to warrant a boxed warning.'[75]

Immunogenicity (neutralising antibodies)

In clinical trials (studies NIH 991265/20010769 and FHA101), the rates of antidrug antibodies (ADAs) for GL patients and the PL subgroup patients were 96% (51 out of 53 patients) and 93% (27 out of 29 patients), respectively.[33]

Overall, in patients where antibody data was available, neutralising ADA activity was observed in 38/102 patients (37%): 25/53 (47%) with GL and 6/29 patients (21%) within the PL subgroup. An attenuation (typically denoted by initial improvement and then decline of both HbA_{1c} and triglyceride levels) and worsening (denoted by decline from baseline in both HbA_{1c} and triglycerides) of metreleptin effect was reported in patients with PL and GL, both with and without neutralising ADAs. In the majority of patients with neutralising activity and apparent attenuation or worsening of metreleptin effect, this effect was transient and without clinical impact.

Serious and/or severe infections that were temporally associated with neutralising activity occurred in five GL patients.[33] These events included one episode in one patient of serious and severe appendicitis, two episodes in patients of serious and severe pneumonia, a single episode of serious and severe sepsis and non-serious severe gingivitis in one patient and six episodes of serious and severe sepsis or bacteraemia and one episode of non-serious severe ear infection in one patient. One serious and severe infection of appendicitis was temporally associated with neutralising activity in a patient with PL who was not in the PL subgroup (i.e. not the indicated population but with a similar safety profile). None of these temporally associated infections were considered related to metreleptin treatment by the study investigators. LD patients with neutralising antibodies and concurrent infections responded to standard of care treatment.

Of the 38 patients with neutralising activity 58% achieved resolution of neutralising antibodies, including 15 patients with GL and seven patients with PL, and 87% (33/38) received uninterrupted metreleptin dosing throughout the period of neutralising activity.[33]

ERG comment: The Centre for Drug Evaluation and Research Report (not included in the

CS), included the following text concerning immunogenicity:

'Metreleptin is highly immunogenic; almost all patients, including those from the obesity development programs, treated with this protein developed binding antibodies. Of greatest immunogenic concern is the potential development of neutralizing antibodies, with resultant inhibition of endogenous leptin activity or loss of efficacy in patients with lipodystrophy.

The sponsor used the following categorization for neutralizing activity from their in-vitro assay: Category A: result is less than the assay cut-point on initial testing; Category B: result is higher than the assay cut-point on initial testing, but less than assay cut-point on repeat testing; Category C: result is higher than the assay cut-point on initial testing and re-testing, but is less than the assay cut-point after additional 1:10 dilution; Category D: result is higher than the assay cut-point on initial testing and re-testing and after additional 1:10 dilution but not after 1:100 dilution; and Category E: result is higher than the assay cut-point on initial testing and re-testing and after additional 1:10 and 1:100 dilutions. Categories D and E represent high potency neutralizing activity to metreleptin. Seven patients from the NIH and FHA101 studies developed neutralizing antibody activity (categories D or E). One of these patients had loss of efficacy, as indicated by an increase in HbA_{1c} concentrations, and five hospitalizations due to bacterial infections. A second patient, also with a history of hospitalization for sepsis and worsening glycaemic control, was recently reported to have developed neutralizing activity. These cases raise concern that development of neutralizing antibodies to metreleptin could impair metabolic control and immune function.

The clinical ramifications of developing neutralizing antibodies are not well characterized; yet, the potential risks of worsening metabolic control and/or severe infections in metreleptin treated patients with lipodystrophy led the clinical review team to recommend that this information be included in a boxed warning.'[75]

Injection site reactions

Injection site reactions were reported in 3.5% of patients with LD treated with metreleptin.[33] All events reported in clinical studies in patients with LD have been mild or moderate in severity and none have led to treatment discontinuation. Most events occurred during the initial 12 months of initiation of metreleptin.

All events reported in clinical studies in patients with LD have been mild or moderate in severity and none have led to treatment discontinuation.

ERG comment: The Centre for Drug Evaluation and Research Report (not included in the CS),[75] included additional information on immune-related adverse reactions (hypersensitivity): 'In the NIH trials, 15% of patients experienced 13 reactions that could be considered immune-related. These included urticaria (2.8%), anaphylactic reaction (1.4%), and papular rash (1.4%). In study FHA101, 32% of patients experienced 13 reactions that could be considered immune-related. These included urticaria, swelling face, rash, pruritus, injection site inflammation, injection site pruritus, and injection site urticaria.'

Deaths

Study NIH 991265/20010769

A summary of treatment emergent deaths is shown in Table C25 of the CS (page 109) and replicated in Table 21, below[1].

The CS states[1] that over the 14-year study duration, treatment-emergent deaths were reported in four (4%) of the 107 patients, including three patients with GL and one patient in the PL subgroup.[37] TEAEs leading to death included renal failure, cardiac arrest (concurrent with pancreatitis and septic shock), progressive end-stage liver disease (chronic hepatic failure), and hypoxic-ischaemic encephalopathy. None of the deaths were assessed as drug-related.

Discontinuations due to TEAEs were reported in five patients with GL (8%) and one patient in the PL subgroup (3%). In four of these six patients, the events leading to withdrawal led to death.[37]

Table 21: On-study deaths, study NIH 991265/20010769 (safety analysis set)

	GL (N = 66)	PL subgroup[a] (N = 31)	PL overall (N = 41)
On-study deaths	3 (4.5)	1 (3.2)	1 (2.4)

Abbreviations: GL = generalised lipodystrophy; PL = partial lipodystrophy; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment-emergent adverse event

[a] PL subgroup = patients with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

Study FHA101

A summary of treatment emergent deaths is shown in Table C26 (page 111 of the CS) and replicated in Table 22, below[1].

Two (5%) of the 41 patients died during study FHA101, including one patient with GL and one with PL (not in the PL subgroup).[38] The cause of death was progression of pre-existing adenocarcinoma in one patient and loss of consciousness following a fall in her home in another. Neither of the deaths was assessed as drug-related.

Discontinuations due to TEAEs were reported in the two patients who died and in two additional patients with PL (not in the PL subgroup).[38]

Table 22: On-study deaths, study FH101 (safety analysis set)

	GL (N = 9)	PL subgroup[a] (N = 7)	PL overall (N = 32)
On-study deaths	1 (11.1)	0	1 (3.1)

Abbreviations: GL = generalised lipodystrophy; PL = partial lipodystrophy; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment-emergent adverse event

[a] PL subgroup = PL subgroup = patients with baseline leptin <12 ng/mL and HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L

4.3 Summary of evidence presented in other submissions

No other scientific evidence was submitted by other consultees. This ERG report does not include a detailed discussion of non-scientific opinion submitted by other consultees or expert testimony provided by other consultees to the appraisal process. Only one submission, from diabetes UK, was made to NICE.

4.4 Additional work on clinical effectiveness undertaken by the ERG

Additional work on clinical effectiveness undertaken by the ERG has been included in Section 4.2.4 of this report. No further additional work was undertaken by the ERG.

4.5 Conclusions of the clinical effectiveness section

4.5.1 Completeness of the CS with regard to relevant clinical studies and relevant data within those studies

The ERG is confident that all relevant published studies of metreleptin were identified in the CS, however there were some weaknesses in the methods used to identify unpublished data. However, not all of the relevant studies identified were included in the CS and some relevant outcomes from the studies that were included were not reported.

Importantly, the follow-up study (NIH follow-up) to the main study used in the CS (NIH 991265/20010769) was not included in the CS, even though this study was used in the cost effectiveness analyses presented.

The search strategies reported in the CS did not include any terms for comparators and would only have retrieved studies of the intervention metreleptin. The ERG is, therefore, not confident that the all relevant comparator and natural history studies were identified and considered for inclusion in the CS. Comparator data for the cost effectiveness analyses were based on a single un-published study (GL/PL natural history study) which, as with the NIH follow-up study, was not included in the CS. From the information provided, the ERG cannot be confident that this study represents the best available source of comparator data.

4.5.2 Interpretation of treatment effects reported in the CS in relation to relevant population, interventions, comparator and outcomes

The key issue limiting the robustness of the efficacy data presented in the CS is the lack of any comparative studies; estimates of treatment effects are based on changes from baseline in single arm metreleptin treatment studies. This problem is compounded as the CS does not include any attempt to draw indirect comparisons through studies of the effects of established clinical management (diet, lifestyle modifications, lipid lowering drugs and anti-diabetic medications). The natural history study, used to provide comparator data for the cost effectiveness analysis, is not used in the clinical effectiveness sections of the CS and has a population which is not comparable to those included in the metreleptin intervention studies. It is therefore not possible to assess the extent to which any apparent treatment effects are attributable to metreleptin, or whether similar effects could be achieved using standard care.

A further substantive issue concerns the nature of the treatment effects reported. The CS focuses primarily on changes in surrogate outcome measures (e.g. HbA_{1c}, triglycerides, hepatic enzymes) and includes very little information about any effects of treatment on patient-perceived symptoms and clinical outcomes (e.g. hyperphagia, organ damage). The report of the NIH follow-up study,[46] provided in response to clarification questions states that:

‘The National Institutes of Health (NIH) Follow-Up study serves as a follow up to the metreleptin clinical trial. This study has allowed for consideration of longer history and follow-up across a range of outcomes not fully studied in the clinical trial. The study is intended to describe the patients who have taken metreleptin at the NIH experiences with lipodystrophy both before and after metreleptin initiation, including outcomes such as hyperphagia, female reproductive dysfunction, damage to key organ systems, and death, as well as trial reported outcomes such as leptin, triglyceride, and glycated haemoglobin (HbA_{1c}) levels.’

and includes the stated objective:

‘Describe the outcome of metreleptin on patient health, such as organ damage, hyperphagia, female reproductive dysfunction, death, and metabolic status measures such as leptin, triglyceride, and HbA_{1c} levels.’

However, the ‘post-metreleptin improvements’ reported in this study are frequently based on measures taken at one year and use definitions based on changes in surrogate outcome measures; for example, improvement in liver abnormality is defined as 20% reduction in ALT/AST at year in a patient who had elevated ALT/AST at baseline. Since changes in ALT/AST from baseline to one year are reported in the main NIH 991265/20010769,[37] the presentation of these data in the NIH follow-up study does not provide additional information about organ damage, but is rather a different way of presenting the same data.

Whilst it may appear reasonable to assume that improvements in surrogate outcomes, such as HbA_{1c}, triglycerides and hepatic enzymes, are likely to predict long-term impacts on future health (e.g. in terms of development of cardiovascular disease, liver cirrhosis and pancreatitis). It should be noted that improvements in these measures are not, in themselves, evidence of a treatment effects on long-term health outcomes. Furthermore, where links between these measures and long-term health outcomes are generally accepted, the evidence underpinning such links was derived from populations very different from the LD population.

4.5.3 Uncertainties surrounding the clinical effectiveness

There is considerable uncertainty about the long-term effects of metreleptin treatment, particularly in relation to patient-perceived symptoms and clinical outcomes. A limited report of the NIH follow-up study,[46] provided in response to clarification questions, included some information on newly emergent (on metreleptin treatment) lipodystrophy characteristics in patients with no evidence of these characteristics prior to metreleptin initiation. However, no indication of the timeframe of observation was provided. The ERG has added these data to the results section of this report (see section 4.2.4). Broadly, these data indicate that new incidences of organ abnormalities (liver, kidney and heart) and female reproductive dysfunction continue to occur, in all categories of LD patient, on metreleptin treatment. The data presented are insufficient to allow an adequate assessment of how the rate of development of new abnormalities on metreleptin treatment would compare with that seen in patients on standard care.

Despite the statement in the CS (section 6.2, page 42) that the EAP, which includes some UK patients at Addenbrooke's Hospital, has been running for over 10 years, no data from the EAP were included in the CS and no explanation for this was provided in either the CS or the company's response to clarification questions.

The CS includes some information on the persistence (up to 36 months) of changes in HbA_{1c} and triglycerides on metreleptin treatment (see section 4.2.4). These data indicate that the apparent effect of metreleptin on triglyceride levels may not be applicable to the overall PL population.

The CS (section 9.7.2.5, pg 114) describes incidences of pancreatitis as an adverse event, following withdrawal from treatment: **'**Across the 148 patients included in LD studies, 6 (4%) patients (4 with GL and 2 with PL), experienced treatment-emergent pancreatitis. All patients had a history of pancreatitis and hypertriglyceridemia. One of the patients who developed septic shock concurrent with pancreatitis died; the other 5 patients recovered and continued on treatment. Abrupt interruption and/or non-compliance with metreleptin dosing was suspected to have contributed to the occurrence of pancreatitis in several of these patients. The mechanism for pancreatitis in these patients was presumed to be return of hypertriglyceridemia and therefore increased risk of pancreatitis in the setting of discontinuation of effective therapy for hypertriglyceridemia.'[1] Non-compliance rates of between 9% and 19% were reported,[1] and the extent of the pancreatitis risk, for these patients, remains unclear. Similarly, the results for the NIH 991265/20010 study,[37] described in the CS, note the exclusion of an 'outlier' patient in whom an increase from baseline in triglycerides of >1000% at Month 12/LOCF was observed. This increase was attributed to non-compliance; the extent to which such large increases in triglycerides may be seen in patients who withdraw abruptly from metreleptin is unclear, and similarly the persistence and long-term consequences of any such increases is unknown.

With respect to safety and adverse events, the CS concludes that the known side effects of metreleptin can be managed as part of the normal clinical practice for patients with this complex condition. The CS does not report the safety concerns as highlighted in the Centre for Drug Evaluation and Research Report (not included in the CS) or the associated REMS.[75] The summary of safety in this report states: 'The principal safety concerns with metreleptin are T-cell lymphoma and anti-metreleptin antibodies with neutralizing activity. These concerns are of sufficient magnitude to require REMS. Other safety findings that warrant inclusion in the Warning and Precautions section of the metreleptin labelling include hypoglycemia, autoimmunity, and hypersensitivity.'[75]

5. VALUE FOR MONEY FOR THE NHS AND PSS

5.1 Introduction

This chapter aims to provide an assessment of whether or not metreleptin for lipodystrophy represents value for money for the NHS in England. The main source of evidence used to inform this assessment is the CS to NICE, which includes a cost effectiveness model and description of the methods and results of an economic analysis using the submitted model. This chapter first looks at other economic analyses of metreleptin or other treatments for lipodystrophy available either from the literature or elsewhere in the public domain. This is followed by a detailed exposition and critique of the submitted model and accompanying economic analysis. Due to the concerns of the ERG with respect to the credibility of the submitted model, Chapter 6 includes some exploratory analyses undertaken using a new model developed by the ERG.

5.2 Review of existing economic analyses

The company conducted a systematic review of studies of the cost effectiveness of metreleptin or other treatments for lipodystrophy, and studies of costs, resource use and HRQoL associated with lipodystrophy. The details of the search strategy were provided in the Section 17.3, Appendix 3, of the company submission.[1]

5.2.1 Searches

A single combined search strategy was used for all of these areas. Searches were carried out during February 2017 and up to 7 March 2017. Details of the search strategies were provided in Appendix 17.3 of the CS. The selection of databases searched was adequate (Ovid Medline and Medline in Process, EMBASE, and the Cochrane Library Databases) and all searches were clearly reported and reproducible, the database name, database date span, and date searched was provided for the majority of the searches. The service provider used to search the Cochrane Library was not provided. Section 17.3, Appendix 3 of the CS,[1] listed Econlit being searched, no Econlit strategy was provided but following the company clarified that this was a mistake and Econlit was not searched. A search of key International HTA websites and disease specific conference websites was also undertaken but more specific details of these searches were not provided in the CS (i.e. search terms, website details and results retrieved).

ERG comment:

- The search strategies were generally well constructed and contained a combination of subject heading index and free text terms. The majority of subject heading terms were unnecessarily exploded but this would not impact on results retrieved.
- A good range of economic, costs and HRQoL subject heading terms and keywords were used in the strategies but a specific filter was not referenced.
- There were some discrepancies in the translation of the strategies between databases, for example, animal limits included in the EMBASE strategy but not the Medline strategy. However, these were minor and no significant errors in translation that would result in studies being missed.

- The Cochrane Library search strategy was much simplified and only searched for the term lipodystrophy using word variations. The ERG feels it would have been better to search for the different terms for the condition individually.
- The grey literature searches (CS Appendix 17.3.5) in the company submission did not provide full details of the search terms used, the precise date of the searches or the number of records. The ERG cannot therefore comment on the robustness of these searches.

5.2.2 Review process and results

The company conducted a systematic review of studies of the cost effectiveness of metreleptin or other treatments for lipodystrophy. The details of the search strategy were provided in the Section 17.3, Appendix 3, of the company submission.[1]

The selection criteria used for the health economic evidence were reported in Table D31 of the company submission (CS, page 138).[1] A total of 2,109 publications were identified from the electronic searches. After removal of duplicates, 1,005 publications remained. After title and abstract screening, 1,083 publications were excluded as these were not of relevance to the research question. A total of 21 articles were assessed in full for further evaluation. Of these, 18 were excluded based on population (n=7), study type (n=5), date (n=5) and outcome (n=1). Manual searches of key international HTA websites and disease specific conference websites identified no additional papers. This left three papers for data extraction; two papers considering HIV-related lipoatrophy and one paper considering HIV-related lipodystrophy and lipoatrophy. The flow of studies through the identification and selection processes is depicted in Figure D23 of the CS (CS, page 140).[1]

None of the three studies identified were considered relevant to the economic evaluation of metreleptin. A summary of the key characteristics of each of the identified studies is provided in the CS (CS, Table D32, page 142).[1] Quality assessments for two of the three identified health economic studies, based on an adapted assessment criteria list from Drummond and Jefferson (1996),[81] are also provided in the CS (CS, Table D33, pages 143-148).[1]

The studies were deemed by the company to meet most of the criteria for a well-reported, high-quality economic evaluation, but the scope of all three studies was not considered to be relevant to the submission of the metreleptin owing to the population studied.

ERG comment:

The ERG identified some inconsistencies between the inclusion criteria used to select cost effectiveness studies and those used to select studies for the clinical effectiveness section of the CS. For instance, studies of HIV-related LD were included in the cost effectiveness review, but not in the clinical effectiveness sections of the CS. The company, in the response to the clarification letter, stated that metreleptin is not indicated for the HIV-related LD population. Although the FDA prescribing guidelines state that metreleptin is not indicated for the treatment of HIV-related lipodystrophy,[82] this is not clear from either the NICE scope,[27] or the regulatory information provided in the CS (CS, section 3.1, pages 25-26).[1]

Also, in the CS, quality assessments for only two of the three identified studies were provided. It was not clear to the ERG, why no quality assessment was provided for the remaining study.

The models identified from the review seem to mainly focus on the HIV-related disease attributes and their cost/QALY impacts (e.g. CD4+ T-Cell count), and therefore do not provide relevant information on LD related disease attributes. Thus, the ERG concurs that none of the studies are relevant to the economic evaluation of metreleptin.

5.3 Exposition of the company's model

5.3.1 Economic evaluation scope

The company's submission to NICE presents a model-based cost effectiveness analysis using QALYs as the main health outcome in the comparison of metreleptin versus standard of care (SoC) for the treatment of patients with lipodystrophy. The model considers the patient population from the NIH follow-up study are representative of the lipodystrophy patients in the UK. The lipodystrophy patients in the company base-case include the following subgroups, which fall under the expected licensed indication for metreleptin, described in the CS.[1]

- adults and children who are six years old or older, with congenital or acquired GL
- adults and children who are 12 years old or older, with familial or acquired PL, characterised by leptin level < 12 ng/ml with triglycerides > 500 mmol/l and/or HbA_{1c} > 8 %, while on standard therapy

The intervention, metreleptin, is a recombinant analogue of the human hormone leptin, administered through subcutaneous injection. The comparator, SoC for lipodystrophy is considered to be the standard clinical management without metreleptin, including lifestyle modifications such as diet and exercise, use of lipid lowering drugs, and medications for diabetes.

The analysis takes the perspective of the NHS in England but some of the potential costs (like day care costs) which may fall under Personal Social Services (PSS) appear not to have been included.

The model simulates the metreleptin-eligible patients (according to the expected licensed indication) in the NIH follow-up trial, with and without metreleptin and estimates cost and health consequences over a 60-year time horizon. The cycle length of the model is one year. The primary model outcomes are the estimated incremental QALY gain, the incremental costs and incremental cost effectiveness ratio (ICER) associated with the use of metreleptin compared to SoC. Costs and health outcomes are discounted at a rate of 3.5%.

For those patients receiving metreleptin, an annual treatment acquisition cost of £852,859 is used for all patients, assuming that the treatment is administered in 10 mg doses. Based on the anticipated availability of smaller vial sizes, the company assumed an annual treatment acquisition cost of £434,633. The company submitted a simple PAS discount of [redacted] on the list price to PASLU. The annual costs for SoC were assumed to be £3,000. Upon starting treatment with metreleptin, it is expected that patients will remain fully adherent on metreleptin for the

rest of their lives, unless they discontinue the treatment. The model assumes the observed discontinuation from the patient level data of NIH follow-up trial and once the data are unavailable, an annual discontinuation rate of 2.05% is assumed.

ERG comment:
The scope of the economic evaluation is generally in line with the NICE scope, and the deviations in the company's decision problem are discussed in Section 3.3 of the ERG report. The ERG assessed the adherence of the scope of the economic evaluation to the NICE reference case, which is shown in Table 23 below.

Table 23: Adherence to the reference case principles relevant to highly specialised technologies

Element of economic analysis	Reference case	ERG comment
Defining the decision problem	The scope developed by NICE	The scope of the economic analysis is generally in line with the scope developed by NICE, deviations already discussed in Section 3 of the ERG report.
Comparator	Therapies routinely used in the NHS, including technologies regarded as the current best practice	Standard of care (SoC) is considered the only comparator, it is the established clinical management without metreleptin (including diet and lifestyle modifications, lipid lowering drugs and medications for diabetes).
Perspective on costs	NHS and PSS	NHS perspective was adopted.
Perspective on outcomes	All health effects on individuals.	Yes, patient health benefits are included in the model. Benefits to other afflicted individuals (e.g. caregivers) are not included in the model but discussed qualitatively in the submission
Type of economic evaluation	Cost-effectiveness analysis	Yes
Time horizon	Sufficient to capture differences in costs and outcomes	No, lifetime horizon should have been considered, but the time horizon was chosen as 60 years, and not all patients were dead at the end of the time horizon.
Synthesis of evidence on outcomes	Based on a systematic review	Meta-analysis was not used, as there is no connected network and no RCT available. Some adjustment methods were used to obtain relative comparative clinical-effectiveness estimates for metreleptin vs. SoC from the non-randomized evidence obtained from separate studies.
Measure of health effects	QALYs and life years	Health outcomes are valued in terms of life years and QALYs gained.

Element of economic analysis	Reference case	ERG comment
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	The disutility values associated with disease attributes in the model were derived from a discrete choice experiment, within a sample that is argued to reflect the general population (1000 respondents). The valuation was based on some QALY estimation techniques in the literature (Bansback et al., 2012[83]; Viney et al., 2014[84])
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	
Discount rate	An annual rate of 3.5% on both costs and health effects.	Costs and outcomes were discounted at 3.5%.
Equity weighting	An additional weighting can be applied for incremental QALYs above 10 years.	No additional equity weighting is applied to QALY gains.

In terms of the population, it is not clear that the NIH follow-up study trial population (used as the baseline population in the cost effectiveness model) is representative of UK lipodystrophy patients (i.e. in terms of GL/PL, female/male, congenital/acquired ratios etc.). Only one UK patient was included in the NIH follow-up study. If the characteristics of the LD patients in the UK are expected to differ from the NIH follow-up study, these differences should be reflected in the cost effectiveness model, as well.

It is unclear to the ERG if the treatment eligibility criteria, for the expected licensed indication reported in the CS, is considered only once at baseline or at every consultation whilst the patient is on the medication. For instance, a PL patient might have a high HbA_{1c} value at baseline and therefore be eligible for the metreleptin treatment. However, during the course of the disease her HbA_{1c} value might decrease to a value that is below 8%. It is uncertain from the CS, if the metreleptin treatment would be stopped for this hypothetical patient or not.

In addition, the latest available information (09/03/2018) suggests that: [redacted]

The choice of the time horizon (60 years) seemed to be unsuitable, since at the end of the time horizon of the model, a substantial part of the population (e.g. 26.7% of the metreleptin arm)

was still alive. The time horizon and the mortality calculations should be adjusted in such a way that a negligible number of patients is alive at the end of the time horizon.

5.3.2 Model structure

The CS states that an Excel-based patient level Markov model was developed to perform the cost effectiveness analysis of metreleptin in GL/PL patients. In the CS, the justification of the modelling approach was provided using the taxonomy and the checklist given in Brennan et al. 2006[85] The model intends to simulate the disease progression of lipodystrophy, with and without metreleptin (i.e. SoC with metreleptin vs. SoC only), by using the patient level data from the NIH Follow-up study and GL/PL Natural History study. Patients are modelled for a maximum of 60 years from the start of the treatment. The health state of a patient is determined by the set of attributes listed below, which indicates the level of impairment due to the disease.

- Organ impairment related attributes
    - Heart, kidney, pancreas and liver abnormalities (list of conditions that would fall under an organ abnormality is given in Figure 34 of the CS)
- Lab related attributes
    - HbA_{1c} levels (partial/ no response), triglyceride (partial/ no response) levels
- Other attributes
    - Hyperphagia, ability to work/ perform at school, physical appearance, fast disease progression

In addition to the attributes above, hypoglycaemia events for each patient throughout his/her lifetime are also simulated in the model. The baseline values for these attributes at the start of the model are derived from the NIH follow-up study (including all 112 patients) for both metreleptin and SoC treatment arms.

For the metreleptin treatment arm, real-world data from the NIH follow-up study is used to populate the model for the attributes of heart, kidney, pancreas and liver impairment until the end of the data availability. Once real-world data are no longer available for a given patient, organ abnormality progression is extrapolated at an aggregate level (i.e. in terms of cumulative number of impaired organs), following a Markov process. For SoC, the cumulative number of impaired organs is extrapolated directly from the start of the time horizon, since the company stated that there were no patient level data on organ abnormality. The conditional organ-specific impairment probability weights are applied onto the extrapolated cumulative number of impaired organs to get an estimate for the organ-specific abnormality costs and disutilities accrued at a given period, when organ-specific impairment data are not available. The details of this extrapolation exercise for both the metreleptin and SoC arms (e.g. how the transition probabilities for the Markov process are derived from the NIH follow-up study for metreleptin and from the GL/PL natural history study for SoC) will be discussed in Sections 5.3.3.1 and 5.3.3.2.

For the blood lab attributes (i.e. HbA_{1c} and triglycerides), real-world data are used to populate the model for each patient, until data availability for the metreleptin arm. A last observed

carried forward (LOCF) approach is followed for extrapolating these attributes beyond data availability, until the end of the time horizon.

For the other attributes, in the NIH follow-up study, real world data seem to be recorded, at most, twice; one measure at baseline and a second measure a year (or more) later. The latter value for the attribute is applied from the first cycle and onwards for the patients receiving metreleptin.

For patients receiving SoC, for all attributes other than organ impairment related attributes, the baseline values from the NIH follow-up study are assumed to remain the same until the end of the time horizon.

Only age, cumulative number of impaired organs and type of the lipodystrophy (i.e. GL or PL) are assumed to have a direct impact on a patient's mortality, whereas all attributes listed above as well as the hypoglycaemia events are assumed to have a direct impact on a patient's QoL and costs.

In the base-case, the average of the model outcomes from the NIH follow-up study patients who fell within the original expected licensed indication reported in the CS (80 out of 112 patients) were presented. Similarly, in the subgroup analyses, the average of the model outcomes from those NIH Follow-up trial patients who were in the considered subgroup (e.g. for the PL subgroup, the average model outcomes from the 17 PL patients from the NIH Follow-up trial) were presented.

ERG comment:
The ERG agrees with the company that a patient level modelling approach would be more appropriate for the modelling of the course of the disease for lipodystrophy. However, it is not clear to what extent the potential additional advantages of a patient level modelling approach in comparison to Markov cohort approach were realised in the CS model.

Firstly, the first order stochastic uncertainty (i.e. random variability in outcomes between identical patients) was not explored in the CS model. Instead, each patient in the NIH follow-up study was modelled as an individual cohort, and the model outcomes of that patient were not taken into account if that patient did not fall with in the category for the analysed population (e.g. for expected licensed indication population, results from GL patients with baseline age smaller than 6 were not taken into account). No sampling procedures such as bootstrapping were employed. This modelling approach might underestimate the overall uncertainty of the course of the disease and might overemphasise the dependence on the assumption of the representativeness of the NIH follow-up study for LD patients. This might be problematic, since some of the subgroup results are based on the model outcomes from only a small number of patients (e.g. PL subgroup results depend on model outcomes from only 17 patients).

In the CS, the formal selection criteria for the attributes that are modelled for each patient were not clearly explained. Not all of the disease attributes identified in Section 17.6, Appendix 6, of the CS were included in the model (e.g. depression, neuropathy, amputation, retinopathy,

neutralising antibody risk etc.) It is not clear to the ERG how the final list of attributes included in the model were selected, furthermore it is unclear whether any other relevant and important attributes for lipodystrophy patients were not included in the model.

The current extrapolation approach used in the model for disease attributes ignores all possible interdependencies between disease attributes. All disease attributes are modelled/extrapolated independently of each other. The ERG considers this approach highly questionable, as in other metabolic disease models (e.g. diabetes) most disease attributes are interlinked, for instance the current value of an attribute is used as an input while estimating the future value of another attribute (e.g. cardiovascular disease risk in the next period might be associated with this period's HbA_{1c} and triglyceride levels).

Besides overlooking possible interdependencies in disease attribute extrapolation, the model also applied the extrapolation from different time points in the metreleptin and SoC arms. For patients in the metreleptin arm, the extrapolation of disease attributes is applied from the last observation point (of the available real-world data for each patient) onwards until the end of the time horizon. However, for the patients in the SoC arm, the extrapolation of disease attributes is always applied from the baseline (since there are no real-world data for SoC). This difference in the start times for the extrapolation in the model might lead to an underestimation of the uncertainty for the patients under metreleptin.

In the model, the cumulative number of organ impairments was considered as the primary disease progression surrogate. The ERG has serious concerns about this approach, which will be elaborated in the next section.

5.3.3 Evidence used to inform the company's model parameters

Table 24, below, presents a summary of the evidence sources used to inform the company's model parameters. A more detailed list of model parameter values and sources is presented in the CS (CS, Table D37, pages 162-163).

Table 24: Summary of evidence sources used to inform key parameter groups in the company's model

Parameter group	Source of parameter values
Initial patient distribution (disease attributes, age, sex, disease type)	Based on the baseline from the NIH Follow-up study, both for SoC and metreleptin arms.
Transition probabilities for the organ impairment (metreleptin)	The real-world data from the NIH Follow-up study is used to populate the model until data is available. When there is no real-world data available, disease progression (in terms of total number of organs impaired) is extrapolated by a Markov process, based on transition probabilities that are estimated from the transitions of the number of impaired organs in the whole population of the NIH Follow-up study.
Transition probabilities for the organ impairment (SoC)	From the start of the time horizon, disease progression (in terms of total number of organs impaired) is extrapolated by a Markov process, based on transition probabilities that are estimated by the transitions from a subset of the GL/PL Natural History study. The subset is

Parameter group	Source of parameter values
	selected based on a matching method to make the baseline characteristics of the two studies, NIH Follow-up study and the subset of the GL/PL Natural History study, similar (in terms of age, gender and the initial organ damage)
Transition probabilities for blood-lab attributes (HbA_{1c} and triglycerides)	For the metreleptin arm, the real-world data from the NIH Follow-up study is used directly, to populate the model until data is available. When the real-world data becomes unavailable, the LOCF method is used to extrapolate the blood-lab attributes and the last observed data is assumed for all the periods until the end of the time horizon. For the SoC arm, the baseline blood-lab attribute values from the NIH Follow-up study are assumed to remain unchanged throughout the whole time horizon.
Transition probabilities for other attributes (e.g. hyperphagia, ability to work/study, reproduction, physical appearance and fast progression)	In the metreleptin arm, for some of the patients, some of the disease attributes are assumed to improve from the baseline value. This improvement is assumed from the first cycle and onwards until the end of the time horizon. It is stated that these improvements were based on the observed patterns in the NIH Follow-up study. For the patients in the SoC arm, all these disease attributes are assumed to remain unchanged from their baseline values.
Adverse events (hypoglycaemia)	In the metreleptin arm, the real-world data from the NIH Follow-up study is used directly, to populate the model until data is available. When the real-world data becomes unavailable, the mean imputation method is used to extrapolate the number of hypoglycaemia events per year until the end of the time horizon. For the SoC arm, it is assumed that the patients do not experience hypoglycaemia events.
Treatment discontinuation	In the metreleptin arm, the patients are at risk of discontinuation from the metreleptin treatment. The observed treatment discontinuation data (i.e. the proportion of the time each patient is on metreleptin treatment in each period) from the NIH Follow-up study is used to populate the model until the data is available. A weighted overall average value of 2.047% for the discontinuation rate is applied to the patients who are still on the treatment at the last observation point, at each cycle until the end of the time horizon. The discontinuation from the metreleptin treatment has implications on the drug acquisition costs and organ impairment progression transition probabilities (for discontinued patients, related parameters from the SoC arm are applied).
Mortality	A Cox proportional hazard model is fitted to the GL/PL Natural History data, with number of impaired organs as the only independent, time-varying covariate. The resulting hazard ratio from this model represents the proportional change in the hazard rate due to an additional impaired organ.

Parameter group	Source of parameter values
	To derive the survival probabilities based on a given number of organ abnormalities, this hazard ratio is applied to: 1) for GL patients, to the survival curve fitted to the patient level survival data from the GL sub-population of the NIH Follow-up study; 2) for PL patients, to the gender/age adjusted mortality figures from the UK life table (based on the sex ratio in the PL sub-population of the NIH Follow-up study). Both the GL and PL curves above are adjusted based on the baseline number of average number of the impaired organs from the NIH Follow-up study data.
Utility decrements for the lipodystrophy complications	A discrete choice experiment (DCE) is used to provide an estimate of health disutilities for the key lipodystrophy attributes selected by the CS. An additive approach is followed while implementing the disease attribute disutilities simultaneously. Perfectly healthy individual was assumed to have a utility of 1.
Metreleptin treatment costs	Data on file from Aegerion.
Standard of care treatment cost	Assumption
Costs for lipodystrophy related complications and other resource use	KOL input and NHS reference costs.

5.3.3.1 Extrapolation of organ impairment progression

Abnormalities in four organs (heart, kidney, liver and pancreas) are considered in the model and the conditions that are categorised as organ abnormalities for each of the four organs are listed in Table 25 below.

Table 25: List of conditions that are categorised as organ abnormalities

Organ	Condition(s)
Liver	Ectopic fat deposit on liver Hepatomegaly Hepatic steatosis Steatohepatitis Cirrhosis Liver failure
Heart	Cardiomyopathy Heart failure Myocardial infarction Arrhythmia
Kidney	Chronic kidney disease Nephropathy Kidney failure
Pancreas	Pancreatitis
Source: Table 34 in the CS.[1]	

Organ impairment progression in the metreleptin arm

In the NIH follow-up study, real-world data pertaining to each patient's organ-specific abnormality were available for a limited time. When real-world data was no longer available,

for each patient the total number of abnormal organs was extrapolated using a Markov process. The progression probabilities (i.e. transition probability for developing the next organ impairment) were estimated by fitting exponential parametric survival functions to each of the four KM curves given in Figure 1, derived from the NIH follow-up study. The first KM curve in Figure 1 below represents time to develop the first organ abnormality; the second KM curve represents time to develop the second organ abnormality (given one abnormality at the baseline); the third KM curve represents time to develop the third organ abnormality (given two abnormalities at the baseline) and the last KM curve represents time to develop the fourth organ abnormality, given three abnormalities at the baseline.

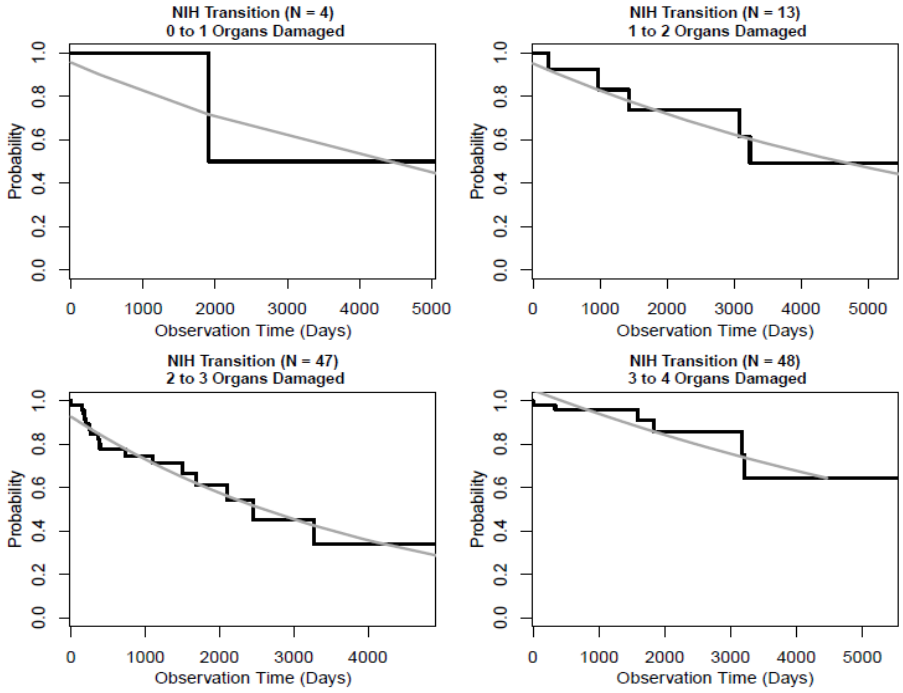
The KM and the fitted exponential curves for disease progression from the NIH follow-up study and the resulting progression probabilities obtained from the fitted exponential curves are given in Figure 1 and Table 26, respectively.

Table 26: Estimated annual progression probabilities from the NIH follow-up data (N=112*)

Progression event	Estimated progression probability	Number of patients at risk	Number of progressions
0 to 1	5.4%	4	1
1 to 2	5.0%	13	5
2 to 3	8.3%	47	17
3 to 4	3.9%	48	7

Source: Table 70 in the CS.[1]
*NIH follow-up study included 114 patients, but sufficient data after baseline is available for only for 112

Figure 1: NIH follow-up study organ abnormality progression



Source: Figure 1 in the second response to the CL.[39]

Organ impairment progression for SoC in the unmatched cohort

The same extrapolation approach (Markov process for the total number of abnormal organs) is followed for organ impairment progression under SoC. The estimated transition probabilities that are derived from the GL/PL natural history study data are applied to patients from baseline until the end of the time horizon. Note that at baseline, the patients are assumed as identical in both the SoC and metreleptin arms in the electronic model. However, if the extrapolated number of impaired organs of a SoC patient led to fewer impaired organs than for that patient on metreleptin (this can happen since real-world data are being used in the metreleptin arm), then the extrapolated number of impaired organs in the SoC arm was replaced by that from the metreleptin arm.

The KM and the fitted exponential curves for disease progression from the GL/PL natural history study and the resulting progression probabilities are given in Section 17.6.2.1, Appendix 6, of the CS (CS, Figure 35 and Table 70, pages 256-257).[1] Note that these probabilities are from the original, "unmatched" population of the GL/PL natural history study, and they are not used in the model. The matching exercise and the consequential "matched" transition probabilities will be explained further in Section 5.3.3.3.

ERG comment:

The ERG has several concerns surrounding the modelling of the disease progression in terms of the number of organ abnormalities, the categorisation of the clinical conditions to organ abnormality types (and resolving of the organ impairment in the metreleptin arm), the data updates in the evidence submitted by the company after the CS, differences between the NIH follow-up study and GL/PL natural history studies, and some other methodological concerns

regarding the estimation of the transition probabilities related to organ impairment. These issues are listed, in summary, below:

1. Level of aggregation while modelling the impacts of the lipodystrophy progression on different organs
2. Difficulties in the interpretation of the real-world data on the organ impairments provided in the CS
3. Data updates delivered after the original CS
4. Differences between the NIH follow-up study and GL/PL natural history study in terms of participant baseline characteristics and inherent structural censoring (patients were observed from their enrolment time and onwards in the NIH follow-up trial, whereas in the GL/PL natural history study, the retrospective patient records were collected to the earliest possible time point)
5. Staggering method (i.e. assuming one day in between two or more organ impairments that were observed simultaneously)
6. Lack of clarity regarding the approach of the incorporation of the time to event data from the NIH follow-up study and from the GL/PL natural history study
7. A patient's simulated number of impaired organs under SoC is forced to be higher than that patient's simulated number of impaired organs under metreleptin in each cycle
8. Lack of details and justification for the methods followed and the assumptions taken while estimating the transition probabilities for the number of organs impaired:
    a. The statistical modelling of the organ impairment process is not in line with the observed organ impairment progression from the real-world data
    b. The current approach implicitly assumes that the organ impairment process possesses the Markov memoryless property
    c. Patient characteristics have no impact on the transition probabilities for the number of impaired organs.
    d. The plausibility of the selected method used in the company submission for the estimation of the transition probabilities from longitudinal data.

<u>Level of aggregation while modelling the impacts of lipodystrophy progression on different organs</u>

In the extrapolation of organ impairment progression, only the cumulative number of organ impairments (out of four organs) was taken into account, based on a non-transparent categorisation applied to the clinical conditions identified from the real-world data that were collected/recorded in an ad-hoc manner. It is not clear why the type of affected organ (pancreas, kidney, heart and liver) and the severity of an organ abnormality (e.g. ectopic fat deposit on an organ or an organ failure) were not taken into consideration in the analysis. Based on this assumption in the CS, the cost and health outcomes from an ectopic fat deposit around the liver are assumed the same as those from a myocardial infarction or those from a kidney failure. Furthermore, if a patient has two conditions affecting the same organ (e.g. heart failure and myocardial infarction), the cost and QALY impacts of the second condition affecting the same organ would be ignored. These implications were deemed to be highly unrealistic and unjustifiable by the ERG.

The company, in its response to the clarification letter, provided three arguments to justify this high aggregation level for organ impairment. These arguments were 1) evidence from other cost effectiveness models in the literature which have very simple model structures 2) traceability of the cost effectiveness model from the CS and 3) data constraints.[39]

The ERG disagrees with the first argument, related to evidence from other cost effectiveness models in the literature, as the provided examples were in other, unrelated disease indications (e.g. late stage oncology or aortic aneurysm surgery). With respect to the traceability concern, the ERG considers that this cannot be a justification argument, since this issue would be resolved with transparent programming and reporting practices. Finally, the ERG can understand the company's argument on data constraints; if additional states were considered for the type and severity of organ impairment, the data from the NIH follow-up trial and the GL/PL natural history study might be insufficient to populate the necessary transition probabilities between the additional states. However, the ERG considers that there may be alternative options to incorporate organ impairment severity/type, other than incorporating additional states; for instance, a clinically plausible cumulative organ impairment severity index could have been developed and incorporated as a time-dependent disease attribute in the simulation. Using this approach, the difference in severity among patients having the same number of organ impairments could have been partially reflected (e.g. the cumulative organ impairment severity index of a patient having arrhythmia and ectopic fat deposit around liver would be lower than a patient having myocardial infarction and kidney failure).

Difficulties in the interpretation of the real-world data on the organ impairments provided in the CS

The ERG had considerable difficulties in tracing and interpreting the real-world organ impairment data provided. The ERG had the impression that the conditions which are categorised as an organ impairment in Table 25 above were considered to be permanent, non-reversible conditions; this was how organ impairment was extrapolated in the model, as the number of impaired organs can only stay the same or increase over time. However, from the real-world data provided in the electronic model of the CS, it became clear to the ERG that these conditions could actually be reversible (i.e. in some of the cycles, the previously existent abnormalities of the kidney, pancreas and liver had resolved). When asked about these improvements, the company gave more details in its response to the clarification letter:

"Improvement in kidney and liver abnormalities were assigned to patients with proteinuria (kidney) or impaired hepatic function (liver) based on a reduction of at least 20% of previously abnormal laboratory readings for protein excretion (kidney) and ALT/AST (liver) in the year after metreleptin treatment (…)

As laboratory data for protein excretion and ALT/AST were not available as a time series in the natural history data, we chose to only track the development of organ abnormalities and not subsequent resolution in the organ progression and survival analysis (…)

The only type of pancreatic abnormality included in either the organ progression / survival analysis or the CE model was pancreatitis. An NIH nurse reviewed patient records for evidence

of pancreatitis prior to metreleptin initiation and identified which patients experience no re-occurrence of pancreatitis after metreleptin initiation." (Response to clarification letter, page 27)).[39]

Considering organ impairment improvements only for the metreleptin patients and not for the patients on SoC may well lead to a bias in favour of the metreleptin. Also, whilst the ERG acknowledges that an improvement in a blood-lab value might be an indicator of improvement in organ function, we do not agree that an arbitrary level of improvement in blood-lab values can automatically be considered to be synonymous with the resolution of an impairment. Within the given time constraints, the ERG cannot audit whether or not the categorisation of organ impairment conditions was conducted consistently. Hence, the ERG cannot judge the reliability of the real-world data used in the estimation of the clinical input parameters.

Data updates delivered after the original CS

The company updated the real-world data on organ abnormalities, used in the statistical analyses, twice following the original submission. In its first response to the clarification letter, the company stated that "…, data for the NIH Study were updated upon further validation. Specifically, one patient for whom the latest survival status was uncertain is now confirmed to be alive. Moreover, the end of study period has been updated from January 22, 2017 to December 18, 2017. Pancreatitis data have also been updated to reflect validation of pancreatitis incidence by an NIH clinician. We also corrected an inconsistency in which heart conditions were considered abnormalities in the NIH data used for this analysis relative to the definition used in the Natural history data and the definition used in the CE model. Specifically, hypertension was included as an abnormality in the previous version of this analysis. The revised data is consistent with the definition of heart abnormality used in the CE model." (Response to clarification letter, page 57).[39]

In its second response to the clarification letter, the company stated: "We have additionally corrected some inconsistencies in the definition of organ abnormalities between the NIH Follow-Up Study and Natural History Study and have excluded patients with certain missing data prior to treatment from organ abnormality progression and survival analyses.", and in the footnote mentioned that: "corrections to the NIH Follow-Up Study data are described in the NIH Follow-Up Study summary report. Additionally, this analysis was previously completed using an older version of pancreatitis data for NIH patients and now uses the current, validated version (consistent with other analyses and the data used for the CE model)." (Response to clarification letter, page 21).[39]

Since these data updates appear to have been conducted in an *ad-hoc* manner, i.e. the recording of the organ abnormalities and its categorisation was not specified in a pre-determined protocol and the changes were not transparent, the ERG cannot audit the provided real-world data on organ abnormalities and cannot judge the reliability of these data.

The impact of these data updates on the transition probabilities used in the electronic model will be explained in Section 5.3.4.

Differences between the NIH follow-up trial and GL/PL natural history study

The company noted that the patients from the GL/PL natural history study have data from birth, whereas for patients in the NIH follow-up study, data are only available from the start of their treatment. The company also noted that the resulting truncated data from the NIH follow-up study may lead to biased estimates. Upon a request for clarification on this expected bias, the company provided the following argument in its response to the clarification letter: "Patients with truncated histories are more likely to transition once they are observed than those patients whose prior histories are fully observed. This is because patients with truncated histories are likely to have already spent some time in the state in which they are first observed. Patients whose entire history is observed, on the other hand, spend a longer amount of time in the observed state before transitioning even if they transition at the same rate. This implies that we would estimate higher transition probabilities for those patients with truncated data (NIH patients) than those with full data (GL/PL patients)".(Response to clarification letter, page36).[39]

The ERG considers that the potential bias resulting from this asymmetry of truncation can be partially corrected by statistical matching methods. Furthermore, this argument of bias from the company conflicts with the company's current modelling approach that is built on the "memoryless" assumption, which presumes that a transition from one state to another does not depend on the time spent in the former state. This assumption will be further discussed in point 7b.

"Staggering" method applied to the multiple organ impairments diagnosed in the same visit

The ERG requested an explanation for the steep declines observed in the KM curves near t=0, in all sub-figures depicted in Figure 35 of the CS (CS, Section 17.6.2.1, Appendix 6, page 256).[1] In its response to the clarification letter, the company stated that the information on organ abnormalities was collected during patients' physician visits, and that sometimes patients were diagnosed with abnormalities to multiple organs at the same visit.

The company stated that they dealt with these simultaneous multiple organ diagnoses by "staggering" the diagnoses so that they are one day apart from each other. This resulted in the current KM curves, where some patients seem to spend only one day in an abnormality state before transitioning to the next.

The company provided the "staggered" number of instances in which patients in the NIH follow-up and GL/PL natural history studies were diagnosed with abnormalities to multiple organs on the same visit, as reflected on the transition KM curves:

"-18 natural history patients develop abnormalities to two organs after having had no prior abnormalities
-12 natural history patients and 1 NIH patient develop abnormalities to two organs when they already have one afflicted organ
-10 natural history patients and 1 NIH patient develop abnormalities to two organs when they already have two other afflicted organs

-4 natural history patients and 2 NIH patients develop abnormalities to 3 organs after having had no prior abnormalities
-2 natural history patients and 1 NIH patient develop abnormalities to 3 organs when they have previously had one afflicted organ
-1 natural history patient develops abnormalities to all four organs at the same time" (Response to clarification letter, page 38)[39]

This "staggering" approach would overestimate the speed of progression of the organ abnormality and, since it was applied primarily in the GL/PL natural history study records, this overestimation affected mostly the speed of organ impairment progression probabilities in the SoC arm. Therefore, the ERG anticipates the actual transition probabilities for organ impairment progression in the SoC arm to be smaller than the estimates provided in the CS. However, the ERG cannot quantify this, given the time limitations, and given that the data and the statistical codes provided by the company were not transparent and clear.

Lack of clarity regarding the approach to the incorporation of the time to event data from the NIH follow-up study and from the GL/PL natural history study

In the CS, while generating the KM curves from the "*time to next organ impairment*" data from the NIH follow-up and GL/PL natural history studies, it was not clear whether a death event was considered as a censor or an organ impairment event.

In their first response to the clarification letter, the company provided some results (Table 4 to Table 7 in the first tier of the response to the clarification letter, page 32-34), where the impact of death event categorisation was explored in *de novo* Cox proportional hazard model analyses conducted on several pooled datasets of NIH follow-up and GL/PL natural history studies (original pooled datasets, original matched pooled datasets, updated pooled datasets and updated and matched pooled datasets).[39] From these results, it can be seen that the categorisation of the death event (as an organ impairment event or as a censoring event) has a considerable impact on the hazard ratios (hazard rate for the organ impairment under metreleptin vs. under SoC); considering the death event as a censoring event seems to decrease the hazard ratios. The company did not explain the reasons for this effect of censoring in detail and more importantly, the company did not state which categorisation approach was chosen in the analyses that yielded the organ progression transition probabilities that were used in the electronic model.

Furthermore, the ERG has doubts about the compatibility of the time to event data used for the NIH Follow-up study and for the GL/PL Natural History study.

In Figure 36 from the CS (CS, Section 17.6.2.1, Appendix 6, page 257), the numbers at the top of each subfigure (N=4 for 0 to 1 organ damaged, N=13 for 1 to 2 organs damaged, N=47 for 2 to 3 organs damaged, N=48 for 3 to 4 organs damaged) sum to 112, which is the total number of patients in the NIH Follow-up study.[1] This suggests that the Kaplan-Meier (KM curves) in Figure 36, were incorrectly based on time-to-event data that were not contingent on number of organs already damaged, i.e. not all of the patients who developed the n^{th} organ impairment were considered in the next KM curve, which is analysing the time to develop the $(n+1)^{th}$ organ

impairment. For example, it seems highly unlikely that no one who progressed from 0 to 1 subsequently progressed from 1 to 2 organs damaged and that no one who progressed from 1 at the start to 2 subsequently progressed from 2 to 3 organs damaged. This implies that the rate of progression has mostly likely been underestimated.

The company seems to follow a different approach when analysing the time to next organ impairment from the GL/PL Natural History study. In Figure 35 from the CS (CS, Section 17.6.2.1, Appendix 6, page 256), the numbers at the top of each subfigure (N=142 for 0 to 1 organ damaged, N=151 for 1 to 2 organs damaged, N=120 for 2 to 3 organs damaged, N=77 for 3 to 4 organs damaged) sum to 490, which is larger than the total number of patients in the GL/PL Natural History study (N=178).[1] This suggests that the KM curves in Figure 35, were based on time-to-event data contingent on number of organs already failed, i.e. all of the patients who developed the n^{th} organ impairment were taken into account in the baseline number at risk of the next KM curve, which is analysing the time to develop the $(n+1)^{th}$ organ impairment. This is confirmed by Figure 4 of the short report of on the GL/PL Natural History study supplied in response to the clarification letter.[40]

Overall, the approaches used in the incorporation of the time to event data from the NIH Follow-up study and from the GL/PL Natural History study appear to be incompatible. This would cause a bias, which favours metreleptin. However, the data and the codes provided by the company regarding the NIH Follow-up study were not transparent and therefore the ERG cannot scrutinise them adequately.

<u>A patient's simulated number of impaired organs under SoC is forced to be higher than that patient's simulated number of impaired organs under metreleptin in each cycle</u>
In the electronic model, there is a logical formula that forces the simulated number of impaired organs of a patient under the SoC to be always larger than or equal to the simulated number of impaired organs of that patient under metreleptin.

Even though the organ impairment transition probabilities are higher for SoC, sometimes the extrapolation under the SoC arm can result in fewer organs being impaired compared to the metreleptin arm, since in the metreleptin arm, real-world data is used as an input until the data stops being available. In the instances, where the number of organ impairments of a patient under SoC is lower compared to the same patient in the metreleptin arm, the logical formula takes the higher number from the metreleptin arm to use for SoC.

The ERG deems the use of this formula to be problematic, since it creates a bias favouring metreleptin. The treatment effect and the potential benefit of metreleptin was already reflected in the transition probability estimations. Adding a formula that forbids the simulated number of impaired organs of a patient under the SoC from being smaller than the simulated number of impaired organs of that patient under metreleptin, cannot be considered as an evidence-based modelling practice, but is rather a reflection of the company's expectation bias in the electronic model. In Section 6, the impact of relaxing this programming constraint on the cost effectiveness results will be presented in the exploratory analyses.

The statistical modelling of the organ impairment process is not in line with the observed organ impairment progression from the real-world data

In the statistical modelling approach of the organ impairment process, it was assumed that the cumulative number of impaired organs can stay the same or increase by one. The observations from the real-world data were not in line with this assumption. As discussed previously, in the NIH follow-up study, it was observed that sometimes organ abnormalities resolved over time. Also, from the real-world data it was sometimes observed that multiple organ impairments developed in a given year.

The company, in its first response to the clarification letter, argued that the simplification of allowing only one organ impairment in a year would result in a conservative estimate of the benefit of metreleptin treatment, because with metreleptin, patients would experience multiple organ impairments less frequently. The ERG considers this deduction as speculative, without any formal analysis.

The current approach implicitly assumes that the organ impairment process possesses the Markov memoryless property

The statistical approach the company followed assumed that the probability distribution for the total number of impaired organs would possess Markov memoryless property (e.g. transition from one state to another does not depend on the time spent in the former state). The ERG asked the company to justify the plausibility of this assumption.

In its first response to the clarification letter, in Table 8 (Response to clarification letter, page 38), the company provided the results of the linear regression models conducted on the matched GL/PL Natural History cohort data, where the time to develop the n^{th} organ impairment was the dependent variable, and the time spent with *(n-1)* organ impairment was the only independent variable for *n=1,2,3* and *4*.[39] The company interpreted the results as indicating that there is no strong evidence for a consistent, significant correlation between time spent in the former state and time to progression, for the matched control patients from the GL/PL natural history study. This test was not conducted for NIH follow-up study, since the patients in this study were not followed from their birth.

The ERG considers that there could be other available tests for the Markov memoryless property, however the ERG also considers that the memoryless assumption is not the assumption that is driving the final results that affect decision making.

Patient characteristics have no impact on the transition probabilities for the number of impaired organs

The current modelling approach implicitly assumes that patient characteristics, such as age, gender, type of lipodystrophy, type of organ impairment and its severity, time on metreleptin treatment, blood glucose/triglyceride levels have no impact on the transition probabilities for the number of impaired organs.

In its first response to the clarification letter (Response to clarification letter, question B3.e3, pages 40-43), the company presented the results of some adjusted Cox proportional models,

where the treatment, type of lipodystrophy, gender, baseline age and type of organ impairment at baseline were added as covariates, applied on the pooled dataset of NIH follow-up study and matched GL/PL natural history study population. These results indicated that, when the covariates were adjusted, the treatment seemed not to have a significant impact on the estimated time to next organ impairment, whereas other baseline patient characteristics, such as the baseline organ impairment type seemed to have a substantial impact, even though the direction of the impact was not always consistent and in line with the *a priori* expectations of the ERG (sometimes positive and sometimes negative).[39]

The company acknowledged that these characteristics were important contributors to survival and progression. However, they stated that they did not anticipate that using the estimated transition probabilities in the original CE model would be biased by systematic differences in these attributes across groups, as the goal of the matching was to balance several of these attributes across the NIH (treated) patients and natural history study (control) patients.

The ERG does not agree with the company's anticipation that there would not be any bias by not including these patient characteristics in the statistical analysis of organ impairment, because of the matching between the NIH follow-up study and the GL/PL natural history study populations. Firstly, the matching exercise conducted by the company took only age, gender, type of lipodystrophy and the initial number of organ impairments into account. Secondly, without the data on the baseline organ impairment type of the matched populations from the two studies, the size and the direction of the potential bias arising from not incorporating these covariates cannot be known.

<u>The plausibility of the selected method used, in the company submission, for the estimation of the transition probabilities from longitudinal data</u>

Due to the issues discussed above, the ERG had doubts about the appropriateness of the statistical approach selected by the company, especially given the fact that other standard methods for estimating Markov chain transition probabilities (e.g. multi-state models or maximum likelihood estimation methods) are existent in the literature and are commonly used.[86, 87]

Therefore, the ERG asked the company to conduct a *de novo* statistical analysis for the estimation and the extrapolation of organ abnormality progression, using commonly accepted methods, on the pooled dataset (including label-eligible patients from both NIH follow-up study as well as the natural history study), including all the relevant covariates. The company stated that they could not complete this request given the timelines. The ERG therefore cannot assess the direction and the size of the potential bias caused by not following standard statistical methods, as opposed to the approach followed by the company, whose major flaws are described above.

5.3.3.2 Derivation of mortality inputs for the model

Real-world survival data from the NIH follow-up study are used to populate the model for the survival of the metreleptin arm patients as long as there are data available. Beyond the follow-up period, each patient's survival is extrapolated using the corresponding fitted survival

distribution, depending on that patient's lipodystrophy type (i.e. PL or GL), adjusted according to the total number of organ abnormalities. For patients receiving only SoC, as there are no real-world data available, survival is extrapolated using the fitted distributions directly from the baseline.

Extrapolation of the survival of the GL patients

To provide mortality inputs for the GL patients in the model, the KM curve pertaining to the GL patients from the NIH follow-up study is extrapolated beyond the end of available data. The company declared that the approach described in Latimer et al. 2013[88] and Williams et al. 2017[89] is followed, while selecting the most appropriate fitted parametric curves (exponential, Weibull, lognormal and log-logistic) to the available KM data. The company considered that the exponential distribution would be the best fit based on the statistical fit (AIC) and visual inspection, which are depicted in Figure 38 and in Table 72 in the CS (CS, Section 17.6.2.2, Appendix 6, page 260).[1] The final baseline GL survival curve in the electronic model used the observed survival probabilities for years 0 to 16, and afterwards extrapolated survival probabilities from the exponential distribution.

Extrapolation of the survival of the PL patients

The company stated that there is no excess mortality due to PL, as these patients experience milder symptoms compared to GL patients, and the observed deaths in the NIH Follow-up study among PL patients were extremely low (only one death). Hence, the survival of the PL patients was extrapolated using the age and gender specific mortality figures from the latest (2014-2016) UK lifetables. The final baseline survival curve (based on the female to male ratio and average baseline age from the PL patient subgroup of the NIH Follow-up study) is presented in Figure 39 of the CS (CS, Section17.6.2.2, Appendix 6, page 261).[1]

Relationship between the organ abnormality progression and mortality

In the CS, it is assumed that the survival in a period is determined by the type of lipodystrophy and the number of organs impaired in that period. Other attributes such as the type(s) of organ impairment(s) or the length of time spent with a given organ impairment are assumed to have no impact on mortality.

The assumed relationship between mortality and the number of impaired organs was tested with a Cox proportional hazards model fitted to the GL/PL natural history study data. The number of impaired organs as a time-varying covariate is included in the Cox proportional hazards models to predict mortality for the full GL/PL population, GL subpopulation and PL subpopulation. The regression coefficients from these analyses for the full, GL and PL samples are given in Table 27 below.

Table 27: Cox proportional hazards model on GL/PL natural history study with number of impaired organs as time-varying covariate

Independent Variable	Regression Coefficient (Beta)	Exponential of Regression Coefficient (Hazard Ratio)	Standard Error	p-value
FULL SAMPLE				
Number of Impaired Organs	1.2839*	3.6108	0.3329	0.000115
GL SAMPLE				
Number of Impaired Organs	1.0897*	2.9734	0.4155	0.00873
PL SAMPLE				
Number of Impaired Organs	1.5237*	4.5892	0.5302	0.00406

Source: Table 73 in the CS.[1]
*Statistically significant at 1%

The company used Schoenfeld residual tests for the proportional hazards assumption for the number of impaired organs for the GL subpopulation, PL subpopulation and the whole patient population from the GL/PL natural history study. Results of these tests are provided in Table 74 of the CS (CS, Section 17.6.2.3, Appendix 6, page 264),[1] which suggested that there is insufficient statistical evidence to reject the null hypothesis that the slope of the residuals in time is approximately 0; this is interpreted by the company as indicating that there is no statistically significant correlation between time and the Schoenfeld residuals.

The company provided some alternative proportional hazards models fitted to the GL/PL natural history study data, by including additional covariates in the baseline model such as the gender, country of origin, age and lab values (HbA_{1c}, triglycerides and leptin levels). The results of the additional models are provided in Table 75 of the CS (CS, Section 17.6.2.3, Appendix 6, pages 265-267).[1]

Model 1 included squared and cubed versions of the main independent variable, number of impaired organs, to test for non-linear effects. Model 2 included additional demographic covariates such as age, gender and country of origin. Model 3 included additional blood-lab covariates such as HbA_{1c}, triglycerides and leptin. Model 4 included both blood-lab values and demographic values as additional covariates, both in the GL subpopulation, PL subpopulation and the whole patient population of the GL/PL natural history study. In all of these models, except for Model 1, the number of impaired organs was the only significant covariate.

Eventually the company chose to use the Cox proportional hazard model with the number of impaired organs as the only independent variable. The formal goodness of fit test results were

not provided and the reasons for the selection of the model to use in the base-case were not explained.

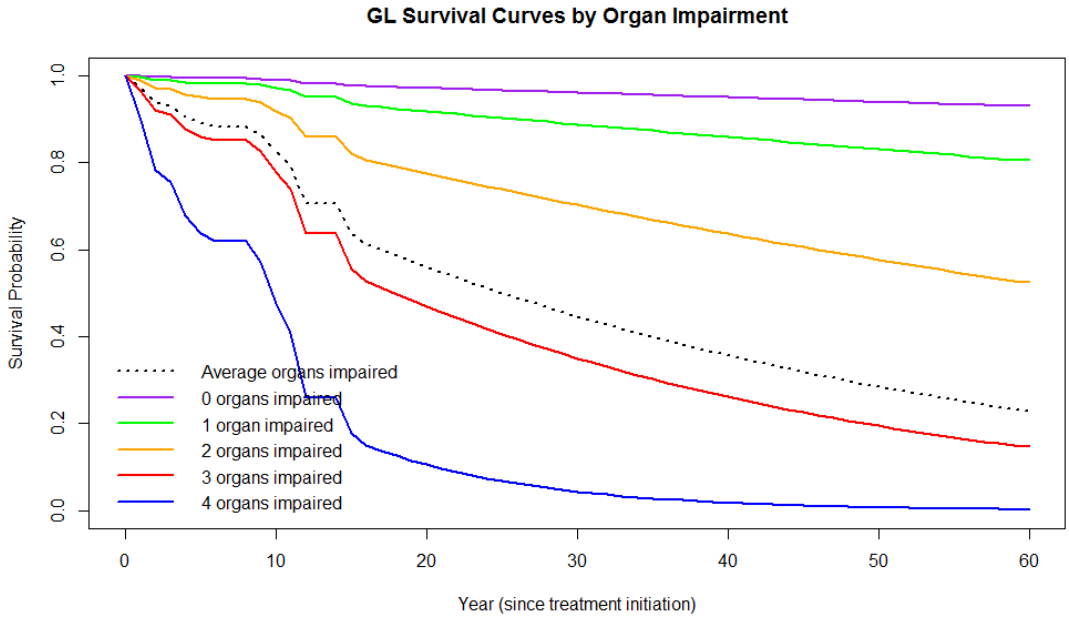
Organ abnormality specific survival curves

The company generated survival curves conditional on the number of organ impairments for the GL and PL patients, to use in the survival extrapolation in the electronic model. To construct these survival curves, baseline GL and PL survival curves obtained from the NIH follow-up data and from the UK population life table respectively were scaled by the coefficient obtained from the Cox model, whose results are given in Table 27 above.

The GL baseline survival curve was interpreted as the survival of patients with the average number of impaired organs among GL patients in the NIH follow-up study. Similarly, the PL baseline survival curve was interpreted as the survival of patients with the average number of impaired organs among PL patients in the NIH follow-up study.

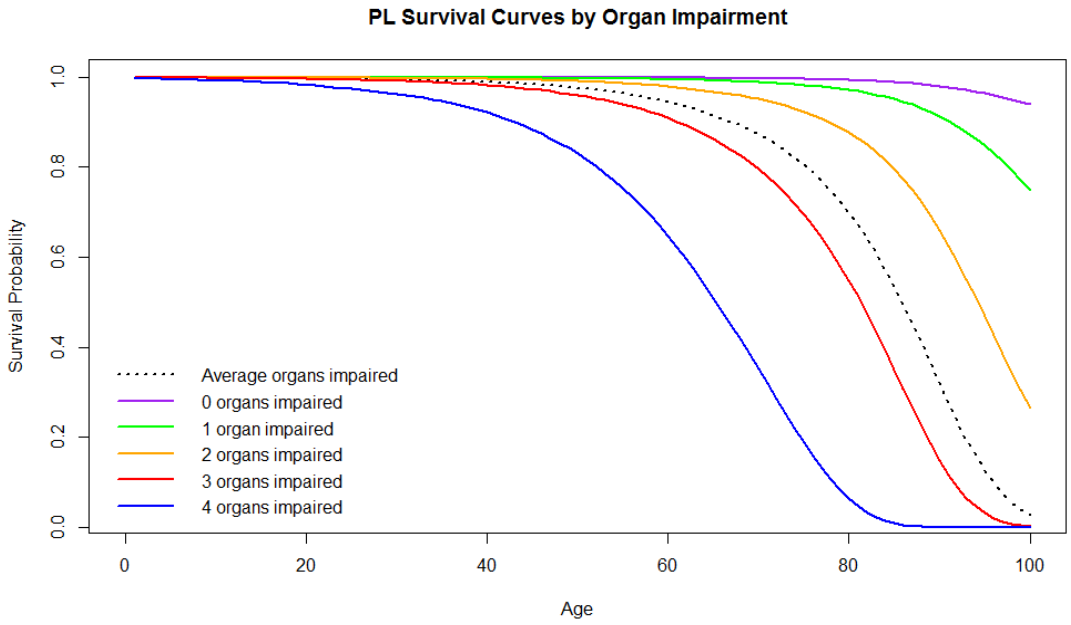
For both GL and PL patients, first the survival curves for the patients with 0 impaired organs were derived; then the survival curves with 0 impaired organs were scaled, by the Cox model coefficient, to derive the survival curves for patients with 1, 2, 3, and 4 impaired organs. This yielded five survival curves for the GL population and five survival curves for the PL population. Each curve corresponding to each of the possible levels of organ impairment (e.g. 0, 1, 2, 3, and 4). These curves for the GL and PL patients are shown in Figure 2 and Figure 3 below, respectively.

Figure 2: GL survival curves by organ impairment levels



Source: Figure 40 in the CS.[1]

Figure 3: PL survival curves by organ impairment levels



Source: Figure 41 in the CS.[1]

ERG comment:

The ERG has several concerns regarding the survival analyses conducted by the company and how the results from these analyses were implemented in the electronic model. The main issues are listed below:

1. Data updates delivered after the original CS
2. Estimation of the survival components from different datasets and synthesising the survival analysis results in a non-systematic manner
3. Lack of face validity for the GL/PL patient's survival extrapolation results (some GL/PL patients have a more favourable life expectancy than the general UK population)
4. Having a substantial number of patients alive (above 25%) at the end of the time horizon
5. Not checking the clinical plausibility of the GL survival extrapolation.
6. The assumption that survival is affected only by age, gender, type of lipodystrophy and number of organs impaired.
7. Wrong derivation of the conditional survival curves given a fixed number of organ impairments.

<u>Data updates delivered after the original CS</u>

As described in Section 5.3.3.1, the data used in the statistical analyses were updated twice after the company submission. Similar to the organ abnormality data, survival data from the NIH follow-up data were also updated. The ERG cannot audit these changes within the time available.

Estimation of the survival components from different datasets and synthesising the survival analysis results in a non-systematic manner

The survival analyses reported in Section 17, Appendix 6 of the CS included an extrapolation exercise (Section 17.6.2.2 of the CS) for the survival of the GL/PL patients using parametric models and national life tables, followed by an estimation exercise (Section 17.6.2.3 of the CS) for the relationship between organ abnormality and mortality.[1] While the extrapolation exercise was conducted on the patients from the NIH follow-up study, the estimation exercise was conducted on the patients from the GL/PL natural history study. The hazard ratio coefficient from the estimation exercise is applied to the parametric/life table survival curves obtained from the extrapolation exercise.

The ERG considers that for the sake of consistency, the estimation and extrapolation exercises should have been conducted on the same dataset and requested clarification from the company regarding the rationale of their approach.

The company stated that the estimation of the relationship between organ impairment and mortality was conducted using only the GL/PL natural history study because of the data limitations of the NIH follow-up study. They noted that, in the NIH follow-up study, information about the early stage of patients' disease was lacking and the observation window in the study was much shorter compared to the GL/PL natural history study. Nevertheless, the company provided the results of the same Cox proportional hazards model to estimate the effect of number of organ impairments on mortality, but using only NIH follow-up study in Table 9 (response to clarification letter, page 48) in their first response to the clarification letter.[39] The company dismissed these results as they were not statistically significant, and the estimated HRs for the GL population from the NIH follow-up study was lower compared to that from the GL/PL natural history study in Table 27 above (NIH follow-up: 1.46 for GL, 4.60 for PL population; GL/PL natural history: 2.97 for GL, 4.59 for PL population).

In addition, the ERG asked the company to provide the results from a *de novo* extrapolation and estimation exercise, using data from a pooled dataset including label-eligible patients from both NIH follow-up and natural history studies, incorporating the study ID as a separate covariate.

The company, in its first response to the clarification letter, stated that a time varying Cox proportional hazard model relating mortality to number of organs with abnormalities (as well as additional covariates) on pooled data was conducted.[39] First, a Cox proportional hazard model was run on the pooled dataset with all NIH follow-up study patients along with matched GL/PL natural history patients, based on the Mahalanobis matching method, using the latest available data. In the second analysis, all NIH follow-up study and GL/PL natural history study patients were combined.

The results of these analyses were presented in the company's first response to the clarification letter (Response to clarification letter, question B10.b, pages 49-53).[39]

In these analyses, conducted on pooled datasets of the GL patients, both the number of organs impaired and the patient's age at the start of the observation were significant covariates. For the GL patients, the HR for the number of impaired organs from these covariate-adjusted analyses on the pooled datasets (HR=1.99 when matched GL/PL natural history study population is used and HR=2.21 when all patients from the GL/PL natural history patients are incorporated) were between the HR obtained from NIH follow-up study only and the HR obtained from the GL/PL natural history study only.

In these analyses conducted on pooled datasets of the PL patients, the number of organs impaired was the only significant covariate. For the PL patients, the HR for the number of impaired organs from these covariate-adjusted analyses on the pooled datasets (HR=6.77 when matched GL/PL natural history study population is used and HR=5.25 when all patients from the GL/PL natural history patients are incorporated) were higher than the HRs obtained from the NIH follow-up study only and the GL/PL natural history study only.

The ERG has difficulty in interpreting these results as they are based on multiple changes implemented at the same time (i.e. covariate adjustment and combining data from both trials as well as the survival data update due to the latest follow-up). The company stated that these *de novo* survival analyses had been implemented in the economic model, and reported some ICER results, however the ERG cannot judge the correctness of the implementation, since hardcoded numbers were used in the implementation of the *de novo* survival models, and the values cannot be traced back to the results of the *de novo* statistical analyses.

Lack of face validity for the GL/PL patient's survival extrapolation results (some GL/PL patients have a more favourable life expectancy than the general UK population)

The ERG considers that some of the survival estimates in the submission lack face validity. For instance, in the model, PL patients who have zero or one impaired organ at baseline have a better life expectancy than the UK general population. Therefore, the ERG asked the company to provide alternative clinically plausible mortality estimates, which cannot be lower than the UK general mortality figures, even if the patient has no organ abnormality.

The company confirmed that the mortality estimates used in the original submission were not clinically plausible and implemented a cap for the survival estimates used in the electronic model that was attached to its response to the clarification letter. In the updated version, the model uses the annual survival probability from the UK life table if the survival probability estimates based on the analyses on the NIH follow-up and the GL/PL natural history studies were more favourable.[39] The ERG considers that this solution is an artificial one. Ideally, the company should have explored the reasons underlying the quite high survival outcomes from the model and should have chosen a plausible survival extrapolation that would not lead to implausible mortality estimates.

Having a substantial number of patients alive (above 25%) at the end of the time horizon

In the company's original model, the '*percentage of people alive*' at the end of the time horizon (60 years) is considerably higher than zero (e.g. average probability of being alive at the end of the time horizon is 26.7% in the metreleptin arm). This seems implausible to the

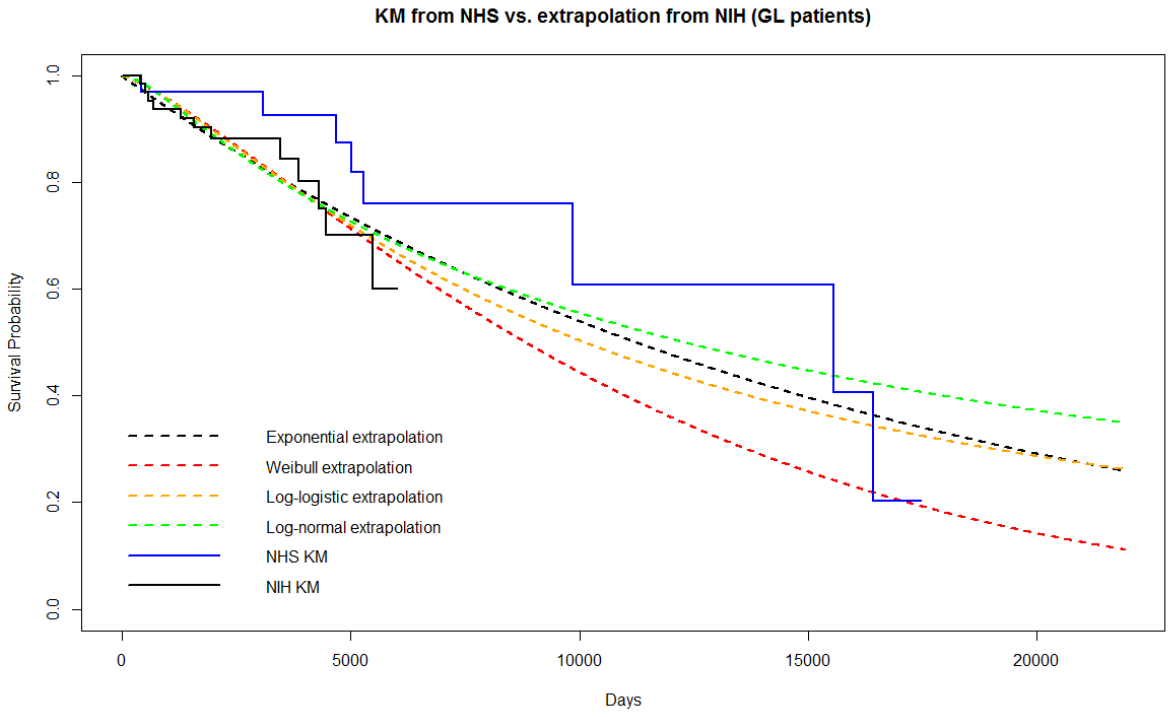
ERG, considering that the time horizon of the model was stated by the company to be lifetime. Therefore, the ERG asked the company to extend the time horizon, such that the average percentage of people alive at the end of the time horizon is almost zero. In its response to the clarification letter, the company provided an updated version of the model with a time horizon of 90 years.[39]

Not checking the clinical plausibility of the GL survival extrapolation

For the mortality of GL patients, data from the NIH follow-up was used, and in the extrapolation of that data, the approach as outlined by Latimer et al. 2013 was followed, but it appears that a crucial step mentioned in Latimer et al. was not included, i.e. checking the clinical plausibility of the extrapolated part of the curve.[88] Hence, the ERG asked the company to provide external data or expert opinion to assess whether another parametric function than the exponential should be used in the base case.

The company, in its response to the clarification letter, presented the results from a validation exercise using survival data from the GL/PL natural history study. The validation exercise compared the KM curve from the GL patients from the NIH follow-up study with that from the GL/PL natural history trial after an age-based adjustment procedure had been applied. The resulting KM curves can be seen in Figure 4 below.

Figure 4: Extrapolation validation for GL patients



Source: Figure 1 in the first response to the CL.[39]

The ERG had difficulty in interpreting the results of this validation exercise. Firstly, the age-adjustment procedure applied to the GL/PL natural history study patients was not clear.

Secondly, Figure 4 above suggests that patients receiving SoC live longer and the additional KM curve says nothing about the relevance of choosing an exponential distribution for the extrapolation. Therefore, the ERG disagrees with the company's interpretation of this graph, which states: "The graph in Figure 1 shows that the exponential extrapolation is in line with this constructed KM curve from the Natural History study".(Response to clarification letter, page 47) [39]

The assumption that survival is affected only by age, gender, type of lipodystrophy and number of organs impaired

The results from Table 75 (CS, page 266) suggest that the number of impaired organs is a significant covariate,[1] but the ERG questions whether this is the only significant covariate, noting that p-values alone might not be the only decision criteria for which covariates to include.

Therefore the ERG asked the company to provide all relevant details (dataset used, statistical codes compiled as well as all statistical outputs from the analyses including all relevant goodness of fit results) for the survival analysis exercises conducted (base case and sensitivity analyses in Table 75 from the CS), with their explanations, and to provide other prognostic survival models with additional covariates (for example type of LD, treatment received and any other relevant covariates), on the GL/PL natural history dataset, NIH follow-up study dataset and the pooled dataset, including only label-eligible patients.

The company, in its response, provided only the outputs of the sensitivity analyses conducted in Table 75 of the CS, on the full GL/PL natural history dataset. The company did not conduct any additional analyses.

The ERG considers that the concordance, R^2, and other goodness of fit statistic results provided by the company seem to compare the model in consideration with the null model. The model analysed in sensitivity analysis 4 (CS, Section 17.6.2.3, Appendix 6, pages 265-267) seemed to provide a better fit than a model based on number of organ impairments only.[1] However, the ERG could not check the statistical codes and the original data in detail and acknowledges that this analysis was not conducted on a pooled dataset, given the timelines. Therefore, the ERG is not certain if the function in sensitivity analysis 4 would be the most plausible predictive survival function that can be ever constructed from the data available from NIH follow-up study and GL/PL natural history study datasets.

Wrong derivation of the conditional survival curves given a fixed number of organ impairment

In the CS, the conditional survival curves given a number of organ impairment were derived from the final baseline GL and PL survival curves (Figure 38 and 39 in the CS, Section 17.6.2.2, Appendix 6, pages 260-261). In these derivations, it was assumed that these baseline survival curves correspond to the survival of patients that were having a fixed number (2.76) of organ impairments. This fixed number, 2.76, was stated as the average number of impaired organs in the NIH follow-up study and was used (together with the hazard ratios of an additional organ impairment for PL and GL patients as given in Table 27 above) while scaling the baseline

survival curves to conditional survival curves for PL and GL patients having zero, one, two and four organ impairments in the baseline.

The ERG considers that this approach is implausible, since the number of organs is not a fixed number throughout a patient's life, but rather a time variant parameter. The average number of impaired organs was 2.76 at the start of the NIH trial, but it was probably much higher (close to four), after 10/20 years. Therefore, the baseline survival curves do not represent a patient population whose number of organ impairments stayed fixed, hence scaling these curves based on this assumption, to conditional survival curves in Figures 5.2 and 5.3, probably overestimated the difference in survival at later time points in the conditional survival curves (i.e. it is expected that after many years, the number of impaired organs will be similar in all patients, independent from the number of organs impaired at the baseline).

5.3.3.3 Matching

The transition probabilities from the GL/PL natural history study (Table 70, CS, Section 17, Appendix, page 257) were not used in the economic model, because the company argued that the baseline characteristics of the GL/PL natural history and the NIH follow-up studies differ substantially (Table 76, CS, Section 17.6.2.4, Appendix 6, page 270), and the patients who were treated with metreleptin were, on average, at a more advanced stage of disease at the start of observation compared to the untreated (under SoC) patients.[1] Therefore, the company used *de novo* organ impairment progression transition probabilities for the SoC arm, derived from the same analysis, described in 5.3.3.1, conducted on a matched subset obtained from the GL/PL natural history study.

Matching methodology

The matching exercise created pairs of patients from both studies. For each treated patient from the NIH follow-up study, an untreated patient at a particular age from the GL/PL natural history study was found, whose reference age matched the treated patient's age at the start of treatment and whose level of organ abnormality at that age was close to that from the matched treated patient, was identified. *A priori* determined weights (α, β) were also assigned to the age and initial number of organ impairments, and gender ($1-\alpha-\beta$) of the patients; patients of the same gender were matched, as far as possible.

Treated GL patients were only matched to untreated GL patients and similarly, treated PL patients were only matched to untreated PL patients. For each treated patient in the NIH follow-up study, the algorithm searched through each pseudo patient generated from the GL/PL natural history dataset (each pseudo patient was generated by specifying a reference age). The pseudo-patient that minimised the weighted sum of the distances from the corresponding treated patient's baseline characteristics (*Diff*) was selected and that pseudo untreated patient was matched to the corresponding treated patient. The same untreated pseudo-patient can be matched with multiple treated patients in the NIH follow-up trial. The algorithm used in pairing the treated and untreated patients is reproduced in Box 1 below.

Box 1: The algorithm used in pairing the treated and untreated patients

Description of the matched cohort

1.) Subset GL/PL patients in the treated and untreated groups so that patients are only matched GL to GL and PL to PL.
2.) Create pseudo-patients with different starting ages.
    - For example, a patient who died or was censored at age 27 is split into 27 different "pseudo-patients," with a starting age of 0, 1, 2 … 24, 25, and 26.
3.) Find the difference (*Diff*) of each parameter (age, gender, initial number of organs impaired) between each treated patient and each untreated pseudo-patient. (For gender, males were coded to be 1 and females 0.)

Diff = (Absolute difference between the treated and untreated individuals) / (Standard deviation of the absolute difference between the treated and untreated individuals)

4.) Match each treated patient without replacement to the untreated pseudo-patient that minimizes an objective function (a weighted average of the differences in age, gender, and initial number of organs impaired).
    - The objective function took the form:

$\alpha * Diff(Age) + \beta * Diff(Initial\ Organ\ Impairment) + (1 - \alpha - \beta) * Diff(Gender)$

Being able to set the weights α, β allows for a flexible approach where changes to the relative importance of each characteristic for measuring the distance between treated and untreated patients can be made.
The weights were set as $\alpha = 0.35$ and $\beta = 0.35$ in the final version of the analysis.

Description of the matched cohort

The company's matching approach resulted in a list of pairs of treated patients and untreated pseudo-patients. The sample statistics of the treated and untreated patients are provided in Table 28, below.

Table 28: Sample statistics of treated and matched untreated pseudo-patients

	Treated patients (from the NIH Follow-up trial)	**Untreated matched pseudo patients (generated from the GL/PL Natural History study)**
Age at first symptoms (mean)	13.33	13.94
Age at start of treatment (mean)	24.28	25.51
Number of impaired organs at start of treatment (mean)	2.52	2.36
Number of mortality events (count)	13	31
% male	16.96	16.96
Source: Table 10 in the first response to the clarification letter, page 60[39]		

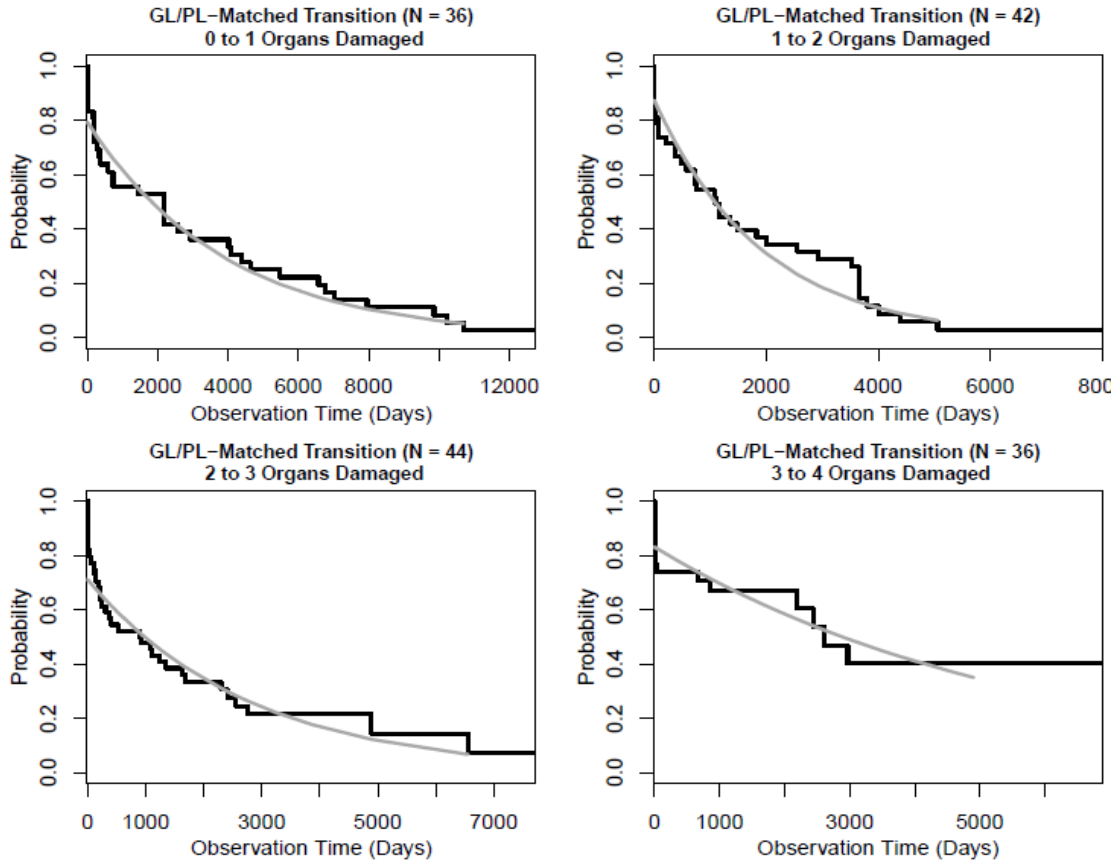
Extrapolation of organ impairment progression based on the matched untreated patient population

The same methods of analyses as described in Section 5.3.3.1 for SoC were applied by the company, but on the matched untreated pseudo-patients.

The KM and the fitted exponential curves for disease progression from the matched untreated pseudo-patients and the resulting progression probabilities obtained from the fitted exponential curves are given in Figure 5 and Table 29, respectively. In Table 29, the transition probability results obtained from the full GL/PL natural history study population are also presented, in order to show the impact of the matching exercise on the probability estimations.

The economic model uses the transition probabilities from the matched untreated pseudo-patients given in Table 29 as input. As can be seen from Table 29, the matched population's transition probabilities were higher in comparison to the results from the full GL/PL natural history study population, for transitions from 0 to 1 organ impairment, from 1 to 2 organ impairments, from 2 to 3 organ impairments. The transition from 3 to 4 organ impairments remained more or less unchanged.

Figure 5: Organ abnormality progression among matched natural history patients



Source: Figure 42 in the CS.[1]

Table 29: Estimated progression probabilities for the full GL/PL natural history study population (N=178) and for the matched untreated pseudo patients (N=47)

Full GL/PL Natural History study population			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	6.7%	142	127
1 to 2	13.3%	151	112
2 to 3	11.0%	120	76
3 to 4	6.4%	77	30
Untreated matched pseudo patients (generated from the GL/PL Natural History study)			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	8.9%	36	36
1 to 2	17.3%	42	39
2 to 3	12.3%	44	36
3 to 4	6.2%	36	16
Source: Tables 70 and 78 in the CS.[1]			

ERG comment:

The ERG has several concerns surrounding the matching exercise conducted by the company and how the results from these analyses were implemented in the electronic model. The main issues are listed below:

1. Appropriateness of the company's approach to the use of data to inform the estimates of treatment effectiveness
2. Lack of clarity regarding the matching algorithm used by the company
3. Independent estimation of the organ impairment transition probabilities from the treated and the matched untreated patient datasets
4. Lack of interpretation of the results

<u>Appropriateness of the company's approach to the use of data to inform the estimates of treatment effectiveness</u>

The ERG disagrees with the company on the appropriateness of the approach followed for analysing the evidence from the observational studies. In the NICE DSU TSD 17, some guidance has been provided for the selection of methods. In particular, a summary overview of the method selection algorithm as depicted in Figure 1, Figure 2 and Figure 3 in the TSD 17 document (p37-39).[90]

The company stated that the matching method employed in the CS was in line with NICE TSD 17, as it resembled the "nearest neighbour matching method", which was, according to the company, one of the two recommended matching methods (together with the propensity score matching) in NICE TSD 17. In the nearest neighbour matching method, a multivariate measure

of distance (typically the Mahalanobis distance) is minimised between the matched pairs. Since Mahalanobis distance was mentioned in the NICE TSD 17 as a typical example, the company, in its response to the clarification letter, provided results for an additional matching exercise, which minimises the distance between the treated and untreated cohorts based on the Mahalanobis distance.[39] In the latest submitted electronic model, the company used the transition probabilities derived from the matched untreated population based on the Mahalanobis distance minimisation method. The impact of this method and data updates on the transition probabilities used in the electronic model will be explained in Section 5.3.4.

The ERG considers that the NICE TSD 17 recommendations were misinterpreted by the company. Firstly, the nearest neighbour and propensity score matching methods (using distance measures such as Mahalanobis distance) were only mentioned as the most popular methods and they are not explicitly recommended *per se*.[39] In order to follow the actual recommendations in NICE TSD 17, the algorithm illustrated in Figure 2 and Figure 3 of that report should have been considered.[39]

The ERG notes that all of the steps depicted in Figure 2 from NICE TSD 17 were omitted. No discussion on the reasonability of the "no unobserved confounding" assumption was provided.

Furthermore, even after skipping all the necessary steps in Figure 2, the company employed some of the steps given in Figure 3 in an *ad-hoc* manner. The overlap between the treated and untreated groups before the matching and the balancing of the covariates after the matching were not assessed in a systematic way. No multivariate regression was conducted on the matched sample to estimate the treatment effect.

The selection of the covariates used in the matching (age, gender and number of organ impairments) was not based on a systematic selection procedure. Some of the influential observed confounders (e.g. the type of the organ impaired) were not included in the matching analysis. This might be problematic, since in the statistical analyses provided in the first response to the clarification letter document (Question B3.e.3, Response to clarification letter, pages 40-43), it can be seen that the type of the organ impairment had a significant impact on the transition probability estimates for the number of impaired organs.[39]

<u>Lack of clarity regarding the matching algorithm used by the company</u>
In the matching algorithm used by the company, for each patient died/censored in the GL/PL natural history study, pseudo patients that died/censored patient were created. It is not clear to the ERG how these pseudo patients were generated. The code provided by the company gave some errors and the ERG is especially concerned if the starting number of impaired organs for these pseudo patients remains the same as their starting ages increase. Omitting to update the starting number of impaired organs while updating the starting age of a pseudo patient would create a bias in favour of the metreleptin arm.

Furthermore, it was not obvious why a weight of 0.35 was chosen for the starting age and the initial number of impaired organs in the base-case. The ERG considers this choice to be

arbitrary, and the weights should reflect the relative impact of each of the covariates on the estimated treatment effect.

Independent estimation of the organ impairment transition probabilities from the treated and the matched untreated patient datasets

The organ impairment transition probabilities for the treated and the matched untreated patients were estimated from different datasets, independently. The ERG noted that the CS did not include any sort of justification of this approach, and questions why the treatment effect was not estimated from a pooled dataset.

Lack of interpretation of the results

The ERG considers that insufficient interpretation of the matching results was provided. The size of the untreated matched dataset (N=47) is approximately one third of the treated patients' dataset (N=112); this suggests that an untreated patient is matched to multiple treated patients from the NIH follow-up trial. The implications of this were not discussed sufficiently in the CS.

Furthermore, it is not clear if the treatment shows a benefit for patients with a low number of organ impairments. In the covariate adjusted analyses conducted on the pooled dataset (NIH follow-up and the matched untreated) provided in B3.e.3 (Question B3.e.3, Response to clarification letter, pages 40-43),[39] the treatment was not a significant covariate in most of the analyses.

Given the lack of discussion on the "no unobservable confounding" assumption, the arbitrary selection of covariates (omitting many other observable confounders such as the type of organ impaired), the arbitrary selection of the methods, and how the treatment effect is estimated from the matched datasets, the ERG considers that the clinical inputs (resulting from the matching and the corresponding survival and organ impairment transition probability estimation exercises) used in the cost-effectiveness part of the submission are not trustworthy.

5.3.3.4 Other attributes (blood-lab and attributes other than organ damage)

In the extrapolation of blood-lab attributes (i.e. HbA_{1c} and triglyceride values), for the metreleptin arm, real-world data from the NIH follow-up study are used directly, to populate the model until the last time data are available. When real-world data become unavailable, the last observation carried forward (LOCF) method is used to extrapolate blood-lab attributes and the last observed data is assumed for all the periods until the end of the time horizon. For the SoC arm, the baseline blood-lab attribute values from the NIH follow-up study are assumed to remain unchanged throughout the whole time horizon.

In the extrapolation of the remaining attributes other than blood-lab and organ damage (i.e. hyperphagia, ability to work, reproduction, physical progression and fast progression), for the metreleptin arm, in some of the patients, some of the disease attributes are assumed to improve from the baseline value. This improvement is assumed from the first cycle and onwards until the end of the time horizon. It is stated that these improvements were based on the observed

patterns in the NIH follow-up study. For the patients in the SoC arm, all these disease attributes are assumed to remain unchanged from their baseline values until the end of the time horizon.

ERG comment:
The ERG has several concerns surrounding the extrapolation of blood-lab and other attributes (other than organ damage), conducted by the company in the electronic model. The main issues are listed below:

1. Lack of clarity regarding the real-world data from the NIH follow-up trial used for the attributes
2. Lack of clarity about the attributes that were included in the model
3. The extrapolation method assumed for the blood lab attributes
4. The extrapolation method assumed for the other attributes

Lack of clarity regarding the real-world data from the NIH follow-up trial used for the attributes
In the economic model, for each patient, a maximum of two measurements were provided for the following attributes: hyperphagia, ability to work, reproduction, physical progression and fast progression. For each of these attributes, the values under the "0" column were used for the SoC arm patients and the values under the "1" column were used for metreleptin arm patients. It is stated, in the company submission, that the values under the "1" column indicate the improvement from the baseline, however, details on the size/definition of these improvements were not provided. Therefore, the ERG requested detailed information on these attributes.

The company provided the following details in its response to the clarification letter:

"Hyperphagia and Impaired ability to work/attend school were coded directly from clinician's notes indicating the presence or absence of these attributes before metreleptin treatment and the improvement of the condition after metreleptin treatment. Improvement in impaired physical appearance was determined by improvement in any of acanthosis nigricans, hyperkeratosis, or hirsutism by the last NIH visit date. Improvement in disruption to female reproductive function is determined by improvement in any of irregular menstruation or polycystic ovary syndrome (PCOS) by the last NIH visit date. For an underlying issue to be improved as of the last visit date, the patient must have had the issue at baseline, and cannot have experienced any new emergent issues in the follow-up period specifically for that issue. In the case that one underlying issue present at baseline did not improve, while another issue present at baseline did improve, the patient is considered to have improved."(Response to clarification letter, page 28)[39]

The company's explanation lacks any objective, measurable definition of a clinical improvement for these attributes. The ERG cannot judge the reliability of the improvement data on these attributes, based on the information supplied.
Furthermore, in the electronic model, where real-world data were missing, the missing value was automatically assumed to be "0". The ERG asked whether this was a programming error or a deliberate assumption. The company acknowledged that it was a deliberate assumption,

stating that they expect that any impairment would be likely to be indicated in the patient's medical data. Thus, when there is no evidence of an attribute being present, it was typically assumed that it was absent.

The company stated that the only exception would be hyperphagia, stating that this was unlikely to be documented unless physicians were prospectively asked to assess it, whether or not it was present.

The company corrected the electronic model in the new version submitted, together with its response to the clarification letter. In the corrected model, patients with no hyperphagia data in period 1 were considered to experience the average treatment effect of metreleptin for their relevant group (i.e. patients with hyperphagia at baseline who lack metreleptin treatment data at period 1, will be assumed to have a hyperphagia with a probability of 0.09 in period 1 and onwards, since 9% of patients in the real-world data who suffer from hyperphagia at baseline continued to have hyperphagia in period 1).

The ERG deemed these imputation approaches as speculative, since they were not based on evidence, but rather on assumptions/expectations.

Lack of clarity about the attributes that were included in the model

In the CS, neuropathy, amputation and retinopathy were named in the list of attributes used in the electronic model, which characterised an individual patient's health (CS, Section 12.1.6, page 158).[1] However, in the electronic model, the ERG was unable to find these attributes.

The company confirmed that these attributes were not included in the cost effectiveness model and admitted the reporting error in the CS. They explained that these attributes were included in the discrete choice experiment (and thus utility decrements estimated), however, since the data on these attributes were not systematically available in the NIH follow-up study, the company decided not to include them in the model.

The extrapolation method assumed for the blood lab attributes

It was not clear to the ERG why only a "last observed carried forward" approach was followed in the extrapolation of HbA_{1c} and triglyceride levels. Therefore, the ERG asked the company to justify their choice of extrapolation approach and explore other methods for HbA_{1c} (e.g. regression imputation or assuming a linear increase) and triglyceride (e.g. mean imputation) extrapolation.

In the updated version of the electronic model submitted with the company's response to the clarification letter, a scenario analysis is conducted where each patient under metreleptin was assumed to experience the same annual change in his/her blood-lab values that s/he experienced during the period when real world data were collected. On the other hand, for patients under SoC, a 0.01 percentage point increase of HbA_{1c} and a 1 mg/dL increase in triglyceride level were assumed each year. The ERG considers these scenario analyses uninformative, as the extrapolation parameters for the blood-lab values were arbitrarily chosen.

The company stated that the NIH follow-up study suggested some improvements in the blood-lab values, but there was variation between patients. They further stated that no specific trend was observed in the GL/PL natural history study. The company rationalised its extrapolation choice by claiming that the LOCF approach would be conservative, however, the ERG questions the validity of this claim, since substantiating such a claim requires a comparison of these longitudinal blood-lab values from both studies in a statistical analysis.

In general, the ERG does not agree with the assumptions of the company base-case and in the additional scenario analysis. Assuming that the blood-lab values would remain constant or keep on decreasing in the metreleptin arm cannot be considered as conservative, given the outstanding uncertainties about the anti-drug antibodies and long-term efficacy.

The extrapolation method assumed for the other attributes

The "Progression Speed" attribute has an impact on QoL and cost calculations but it has no influence on the disease progression probabilities in the model. The ERG had the impression that this attribute was related to the speed of disease progression, and hence the disease progression probabilities would be affected by this attribute. Therefore, the ERG requested additional details on the "fast progression" attribute and justification for the exclusion of this attribute's impact on the disease progression probabilities.

The company stated that the progression speed attribute was included to illustrate the disutility associated with living with an aggressive and progressive disease. Patients were categorised as experiencing fast progression at baseline if they developed more than one organ abnormality per nine years of age prior to metreleptin initiation. Patients were categorised as continuing to experience fast progression after metreleptin initiation if the next organ abnormality was observed within three years of metreleptin initiation.

The ERG considers this categorisation to be problematic, since the time frame used to define improvement was shorter than the time frame used to identify the existence of the attribute at base line (three years vs. nine years). Furthermore, the ERG remains unconvinced about the validity of excluding the impact of the "fast progression" attribute on the disease progression probabilities. The ERG expects that patients having this attribute would have different transition disease progression probabilities than patients without the attribute.

It is not clear to the ERG how the ability to work data and improvement in ability to work data were categorised in the NIH follow-up trial. Also, the ERG notes that the probabilities for being unemployed, partially employed and being retired were not incorporated in the calculations. The ERG is not certain if an improvement in the employment status of a patient would be directly attributable to the intervention.

Given the uncertainties and the lack of reliability of the collected data, the ERG requested alternative scenario analyses from the company, such as a scenario where the baseline and follow-up attribute values are the same in both metreleptin and SoC arms. In addition, another scenario analysis was requested, where these attributes do not stay constant but change over time. The company provided these scenarios embedded in the updated version of the electronic

model submitted together with the company's response to the clarification letter.[39`] The impact of the same non-organ/non blood-lab attribute levels was also examined in the ERG exploratory analyses in Section 6.

5.3.3.5 Adverse events

Only hypoglycaemia was incorporated in the economic model as an adverse event. In the metreleptin arm, the real-world data from the NIH follow-up study were used directly, to populate the model until data were no longer available. When real-world data became unavailable, mean imputation (for that specific patient until that specific time) was used to extrapolate the number of hypoglycaemia events per year until the end of the time horizon.

For the SoC arm, it was assumed that patients do not experience hypoglycaemia events. The justification of this extrapolation approach was not given in the CS.

ERG comment:

It was not clear to the ERG, if all hypoglycaemia events that the NIH follow-up patients experienced were collected systematically.

In addition, the ERG cannot understand why no adverse events, other than hypoglycaemia, were incorporated in the model (such as neutralising antibodies, fatigue, injection site issues, decreased weight, lymphoma, or impact of pancreatitis following discontinuation). It should be noted that the lymphoma risk was subject to a REMS in the FDA appraisal.[75]

The company stated that, beyond the prevalence of an adverse event, the following considerations affected the decision on the inclusion of an adverse event in the cost effectiveness analysis: i) whether these AEs were likely to be caused by metreleptin (vs. were a feature of lipodystrophy, since no control arm was available), ii) the availability of control data (e.g. baseline or pre-baseline information) and iii) whether the potential impact on cost effectiveness could be significant (e.g. vs. marginal).

The company stated that fatigue accounted for 7.3%-9.1% of total treatment-emergent AEs within lipodystrophy subgroups in the NIH 991265/20010769 study. However, their discussions with one of the clinical experts (Dr Brown at NIH), suggested that there was no significant increase in fatigue associated with the use of metreleptin.[39] They further stated that adequate information on fatigue prior to treatment with metreleptin was not available from chart data at NIH, thus a decision was made not to include of fatigue in the cost effectiveness assessment.

Based on the present neutralising antibody assay, the company noted that neutralising antibodies accounted for up to 6.1% of all AEs reported in GL patients, and 0% of all AEs reported in PL patients, and for the majority of these patients the impact on efficacy was transient. The company believes that further inclusion of neutralising antibody considerations, though potentially important clinically, would not have a large impact on the cost effectiveness assessment, since other markers for clinical efficacy were incorporated in the model already. The ERG disagrees with the company's argument, since the loss of efficacy was not captured

in the model. The real-world data from the NIH follow-up study was used in populating the model, but loss of efficacy was obviously not considered for the extrapolations of the blood-lab attributes and of the other attributes (e.g. hyperphagia, ability to work, etc.). Note that the anti-drug antibodies and the potential implications for long-term efficacy were the subject of the second REMS in the FDA appraisal.[75]

The company stated that all injection site issues in the NIH 991265/20010769 study were moderate, non-serious, and did not lead to treatment withdrawal. According to the company, the prevalence of such issues was low, occurring in between 6-7% of patients, depending on the lipodystrophy subgroup (GL vs PL) analysed. Consequently, their impact on cost effectiveness considerations was seen as likely to be marginal and they were not included in the analyses.

The company stated that weight decrease occurred commonly in the NIH 991265/20010769 study: accounting for 25.8% of total AEs reported in GL patients, and 4.9% of total AEs reported in PL patients. However, according to the company, excessive weight loss concerns were generally addressed by dose modification/reduction.

In the clinical effectiveness part of the CS, acute pancreatitis was listed as a treatment emergent adverse event and the company stated that abrupt interruption or non-compliance with metreleptin dosing was suspected to have contributed to the occurrence of pancreatitis. The treatment emergent acute pancreatitis risk due to metreleptin was not directly incorporated in the cost effectiveness analysis. When the ERG requested for a clarification, the company stated that the increased risk of pancreatitis due to metreleptin discontinuation was incorporated indirectly in the electronic model, by applying the organ impairment risks from the SoC arm for the patients who discontinue metreleptin. Even though the organ impairment risks from the SoC arm are higher than those from the metreleptin arm, the ERG considers that this increase in organ impairment risks is attributable to the situation of not receiving metreleptin treatment in the long-run and therefore does not actually represent the risk of acute pancreatitis as a treatment emergent adverse event, which might be due to abrupt interruption or non-compliance as well as other reasons.

The ERG partially agrees with some of the justifications provided by the company on the exclusion of some of the adverse events (e.g. injection site issues), but some of the assertions by the company were not evidence based and solely based on beliefs or expert opinions. Furthermore, the ERG has the impression that some critically important adverse events (e.g. neutralising antibodies and treatment emergent acute pancreatitis) were overlooked in the cost effectiveness analysis, which created a bias in favour of metreleptin. Compared to the many other issues in this economic evaluation however, the impact of this bias may be rather small.

5.3.3.6 Discontinuation

In the metreleptin arm, the patients are at risk of discontinuation from the metreleptin treatment.

The real-world discontinuation data from the NIH follow-up trial were used in the cost effectiveness analysis until data were available. After the point, where data were no longer

available, a weighted overall average value of 2.047% for the discontinuation rate is applied to the patients who are still on treatment at the last observation point, at each cycle until the end of the time horizon.

Discontinuation from metreleptin treatment has implications for drug acquisition costs and organ impairment progression transition probabilities (for discontinued patients, related parameters from the SoC arm are applied).

ERG comment:
In the calculation of the overall average discontinuation value of 2.047%, the discontinuations in the first period were excluded. The company justified this exclusion by the fact that the observed discontinuation data were available for period 1 for all patients and because the pattern of discontinuation in the short term (<1 year) may be substantially different than the discontinuation in the long run. The ERG considers that this exclusion might lead to bias if no statistically testing is conducted on the difference between short term and long-term discontinuation trends.

In addition, besides the drug acquisition costs, the model only reflects the impact of discontinuation in the organ impairment progression (i.e. when a patient discontinues, metreleptin, organ progression transition probabilities for SoC will be used for that patient). The ERG considers that the impact of discontinuation should also be reflected in other disease attributes, (e.g. blood-lab values, hyperphagia, ability to work etc.). Not including the impact of discontinuation on these attributes created a bias in favour of metreleptin. In Section 6, in the exploratory analyses, the impact of discontinuation on other attributes than organ impairment will be investigated.

5.3.3.7 Health-related quality of life

The company conducted a discrete choice experiment (DCE) on a large sample of the general population with the aim to provide reliable estimates of HRQoL "disutilities" associated with key lipodystrophy attributes. In this section, first the DCE study conducted by the company is summarised and critiqued. After the summary of the DCE study, the incorporation of the DCE disutility estimates to the economic model is explained.

Discrete choice experiment study

Details about the study methods and results were provided in Section 17, Appendix 5 of the CS.[1] The main features of the DCE study are summarised below.

Study design

The study analysed data generated by a DCE in which respondents had to choose between two hypothetical health profiles that differed in levels of organ impairment, disease attributes and life expectancy.

Sample selection

A market research firm, Survey Sampling International (SSI), surveyed 1,000 respondents from six countries: the US (250), UK (150), France (150), Germany (150), Italy (150) and Spain

(150). In the US, quotas were set in such a way that the final sample matched the US census on gender, age, region (Northeast, Midwest, South, West), and education. In each of the five European countries, quotas were set for the final sample to match Eurostat demographic characteristics for each country.

Survey

The survey consisted of three main components: (1) a demographic questionnaire, (2) a tutorial informing respondents of the disease and its associated attributes and (3) a conjoint survey in which participants had to choose their most preferred health profile from two choice cards. Choice cards were used to represent hypothetical patients and were constructed by assigning values to disease attributes and varying these values across the two cards.

The tutorial consisted of two parts whose topics are summarized in Table 30 below. The tutorials are fully presented in Appendix 17 – Section 5.4 of the CS.[1] After watching the tutorials, the participants answered a diagnostic question following each part. Those participants who spent less than four minutes reviewing the first part, or less than two minutes reviewing the second part were excluded from proceeding onto the conjoint survey and were not counted towards the respondent quota. Respondents were also excluded from the survey (and not counted towards the respondent quota) when incorrect responses to both diagnostic questions were given.

Table 30: Topics in each part of the survey tutorial

Part 1	Part 2
* Instructions for undertaking the survey * Description of survey pages * Example comparison screen (different for male or female respondents) * List of patient situation attributes * Lipodystrophy – An introduction * Organ damage * Heart damage * Liver damage * Kidney damage * Pancreas damage * Uncontrolled constant hunger (hyperphagia)	* Impaired ability to perform work/school work * Impaired physical appearance (different for male or female respondents) * Disruption to female reproductive functioning (female respondents only) * Depression * Chronic pain * Eye damage (retinopathy) * Nerve damage (neuropathy) * Amputation (e.g., toes, limb) * Impaired triglyceride (blood fat) control * Impaired blood sugar control * Risk of developing neutralizing antibodies
Source: Table D66 in the CS[1]	

The conjoint survey consisted of 14 choice tasks. For each task, participants had to choose between two choice cards consisting of 12 (out of a possible 20) attributes as indicated in Table 31 below. Attributes were shown in random order across respondents but in the same order for each respondent across tasks. Age and life expectancy were always at the top of the choice card and the position of organ abnormality attributes were randomised as a cluster.

Table 31: Summary of attributes and levels for discrete choice experiment

Features	Levels
Age	5 / 25 / 45
Life expectancy (expected age at death)	If age is 5: 15, 25, 45, 65 If age is 25: 35, 45, 65, 85 If age is 45: 55, 65, 85, 105
Remaining life years	= Life expectancy – age
Heart damage	Present / Absent
Liver damage	Present / Absent
Kidney damage	Present / Absent
Pancreas damage	Present / Absent
Progression of organ damage	No change / Slow / Fast
Ability to perform work / school work	Able / Unable
Uncontrollable constant hunger (hyperphagia)	Present / Absent
Impaired physical appearance	Present / Absent
Disruption to female reproductive functioning (Shown to women only)	No damage / Polycystic ovary syndrome / Infertility
Depression	Present / Absent
Chronic pain	Present / Absent
Eye damage (retinopathy)	Present / Absent
Nerve damage (neuropathy)	Present / Absent
Amputation (e.g., toes, limb)	Present / Absent
Triglycerides (blood fat) control	No response or worsening / Partial response / Achieved goal
Impaired blood sugar control	No response or worsening / Partial response / Achieved goal / Achieved goal with hypoglycemia
Risk of loss of response to treatment / Development of neutralizing antibodies	Standard risk / Increased risk due to development of neutralizing antibodies
Lymphoma (a type of blood cancer)	Standard risk / Increased risk
Source: Table D67 in the CS[1]	

QALY estimation

The data obtained from the conjoint survey was used to estimate a multinomial logit model, under the assumption that individuals derive utility from spending time in particular health states as in Bansback et al, 2012 and Viney et al, 2014.[83, 84] In particular, the utility function to be maximised based on the respondents' choices was the following:

$$U = T \times \left(\beta_0 + \sum_i \beta_i x_i \right) + \varepsilon$$

where T denotes the remaining life, β_0 denotes the coefficient quantifying how much utility was associated to one year of perfect health, β_i denotes the coefficient that quantifies the disutility generated by attribute i, x_i denotes an indicator variable which values "1" when attribute i is impaired, and ε denotes the error term.

For the two fertility attributes considered in the DCE, an additional indicator variable (taking a value "1" for females) was also included that multiplied the product of coefficient and attribute indicator variable.

Under a multinomial logit model, it is assumed that, when the utility of choice card A was greater than the utility of choice card B, it is more likely that A is chosen by the respondent. Choice cards also contained information about the current age of the hypothetical patient. This variable allowed stratification of QALY weights by patient's age, which potentially implied different weights for paediatric patients. However, in the utility function described above, age was not included, thereby introducing the potential for omitted variable bias. When age was included in the utility function, some coefficients (i.e. QALY weights) were significantly different (statistically) between patients of different ages. According to the company, excluding age from the utility function "implied that the analysis effectively calculated the QALY weights for a hypothetical patient of average age".[1]

Another assumption was to exclude the intercept coefficient from the utility function. This was done by the company for the sake of consistency, i.e. without intercept, the utility could be interpreted as that obtained from spending T years in a health state characterised by an attribute profile. Moreover, the utility of death was then equal to zero (since a health profile in which a patient dies implies that T = 0). This approach was also followed in Viney, et al 2014,[84] where the impact on the calculated QALY weights of including an intercept on the utility function was deemed negligible. The company indicated that the same happened in their case. The main difference was observed in the coefficients for the progression of organ abnormality, which changed by 20% across the two estimation approaches. However, the contribution of this single coefficient to the overall study conclusions was deemed negligible by the company.

After estimating the coefficients of the utility function as described above, QALY weights associated with each disease attribute were generated. These weights can be interpreted as the decrease in utility associated with attribute impairment as a fraction of the utility from spending one year in perfect health or simply:

$$QALY\ weight\ of\ attribute\ i = \frac{\beta_i}{\beta_0}$$

DCE Results

QALY weights obtained following the approach described above ranged from −0.27 for amputation to +0.03 for slow progression of organ damage. When the analysis considered only respondents from the UK, they ranged from −0.27 for amputation to -0.01 for slow progression of organ damage. All QALY weights obtained from all respondents and from UK respondents only can be seen in Table 32 below. Point estimates and confidence intervals shown in Table 32 were calculated by bootstrapping the QALY weights obtained from the multinomial model. Most of the point estimates based on UK respondents were different from those based on all respondents, and all confidence intervals are much wider, as a result of the smaller sample size.

Table 32: Per-period disutility toll from lipodystrophy-related complications

	All samples		UK only	
Health State	**Utility Value**	**95% Confidence Interval**	**Utility Value**	**95% Confidence Interval**
Heart abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Liver abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Kidney abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Pancreas abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Slow progression of organ abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Fast progression of organ abnormality	[redacted]	[redacted]	[redacted]	[redacted]
Unable to perform work/school work	[redacted]	[redacted]	[redacted]	[redacted]
Uncontrolled constant hunger (hyperphagia)	[redacted]	[redacted]	[redacted]	[redacted]
Impaired physical appearance	[redacted]	[redacted]	[redacted]	[redacted]
Disruption to female reproductive functioning - Polycystic Ovary Syndrome	[redacted]	[redacted]	[redacted]	[redacted]
Disruption to female reproductive functioning - Infertility	[redacted]	[redacted]	[redacted]	[redacted]
Depression	[redacted]	[redacted]	[redacted]	[redacted]
Chronic Pain	[redacted]	[redacted]	[redacted]	[redacted]
Eye damage (Retinopathy)	[redacted]	[redacted]	[redacted]	[redacted]
Nerve damage (Neuropathy)	[redacted]	[redacted]	[redacted]	[redacted]
Amputation (e.g. toes, limb)	[redacted]	[redacted]	[redacted]	[redacted]
Triglyceride (blood fat) control – No response or worsening	[redacted]	[redacted]	[redacted]	[redacted]
Triglyceride (blood fat) control – Partial response	[redacted]	[redacted]	[redacted]	[redacted]
Impaired blood sugar control – No response or worsening	[redacted]	[redacted]	[redacted]	[redacted]
Impaired blood sugar control – Partial response	[redacted]	[redacted]	[redacted]	[redacted]
Impaired blood sugar control – Achieved goal with hypoglycemia	[redacted]	[redacted]	[redacted]	[redacted]
Increased risk of loss of response to treatment/development of	[redacted]	[redacted]	[redacted]	[redacted]

	All samples		UK only	
Health State	**Utility Value**	**95% Confidence Interval**	**Utility Value**	**95% Confidence Interval**
neutralizing antibodies (e.g. with additional medication)				
Increased risk of lymphoma (a type of blood cancer)	[redacted]	[redacted]	[redacted]	[redacted]
Source: Table 68 and 69 in CL[1]				

Validation

Three UK lipodystrophy clinical experts (Dr Rebecca Brown, Dr David Savage, and Dr Anna Stears, from the Addenbrooke primary treatment centre in UK) provided input for the survey and commented on the results. Input from the experts helped identify and prioritise the disease attributes included in the survey. The experts also provided input on the tutorial materials used in the second module of the survey.

Some utility decrement estimates from the DCE were compared to estimates from Ara 2011.[91] According to the company, this comparison "generally validated the settings of the new study" although some differences were observed as shown in Figure 6.[1]

Figure 6: Validation of utility decrement estimates vs published literature

Source: Figure 33 in the CS[1]

ERG comment:

Overall, the ERG has serious concerns about the validity and reliability of the QALY weights reported by the company. The key issue is that the use of DCE to directly obtain disutility values for heath states is still in its infancy. Whilst in the past years several important methodological issues have been resolved, several still remain. The most striking issue relates to the fact that DCE classifies health states far more often below zero than TTO and produces lower average health state values. Given the important consequences for cost effectiveness analyses, and the broad acceptance of health state values derived using TTO, the question inevitably arises: what is the explanation for this difference? Suggestions from literature concern issues related to anchoring (either at death = 0 or worst possible health state = 0), framing issues, or even that choices between cards may be driven not only by differences in utilities but also by how easy it is to compare alternatives.[92]

As long as these differences are not fully understood, the use of DCE disutilities to estimate QALYs remains highly speculative.

Besides this key methodological issue other major issues can be observed both regarding the design of the experiment and the analysis of the resulting data.

Choice cards

The attributes that were used in the DCE were selected based on interviews with clinical experts in the UK and the USA. However, it is common practice to include various stakeholders in the selection of relevant attributes, which in this case would have been for example the patients besides the already included clinicians.

Despite a direct question of the ERG in the clarification letter (Question B13.c) the company did not provide details regarding the potential for overlap and/or correlation between attributes. For example, uncontrolled lab values for blood glucose will lead over time to retinopathy and if respondents are aware of this, it may create correlation between the two attributes.

The three levels of age that were used in the choice cards pose a problem, as it groups children's age and adult ages together. In general, different instruments should be used for these two groups, as respondents tend to choose differently in children.

Another issue with the use of age as an attribute in the choice cards relates to the fact that the two options could state different ages. The choice question ought to be answered conditional on a certain age of the patient; it is impossible to judge what impact the use of two different ages in the choice cards may have had on the choices made.

Life expectancy as presented in the choice card (age + remaining years to live) is possibly subjected to misinterpretation, as it is not clear if respondents were fully aware that the life expectancy indicated time of death.

The colour coding used in the choice cards as illustrated in Appendix 17.5 of the CS[1] appears to be problematic. On each card, red is used for the least favourable level of an attribute *on that card* and green for the most favourable. Thus, colours are not fixed for specific attribute levels, but may be green in one comparison and red in the next. Even if the text would be removed from the coding cards, one could still get a preferred option just based on the colour coding.

Respondent selection from six countries

Combining preference results from six different countries raises the question to what extend this is methodological sound. The fact that for EQ-5D country-specific tariffs have been developed suggests that this may not be the case. Considering this, it is unclear to the ERG why the company has not opted to use the disutilities based on UK respondents only.

It is unclear whether a scaling parameter was included to account for pooling the data from six different countries.

It is also not evident whether the meaning of the attributes and levels is guaranteed in all countries after translation, by for instance doing a forward translation and back translation. With the current information, the ERG cannot properly assess whether this represents an issue or not.

Experimental design

The survey is very long and complex, with 12 attributes being shown per card. This raises questions regarding the respondent cognitive burden of the task. From the information provided by the company it is not clear if a check for respondent burden was included, through a pre-test for example, or post-hoc by checking consistency between the first six choice sets and the last six.

In the survey, choice cards presented to women included an extra attribute for disruption to female reproductive functioning compared to the choice cards for men. It is then unclear how this influences the results of the pooled analysis of male and female respondents. The systematic omission of certain DCE designs for men potentially leads to bias, as there is a risk that men may have never seen certain attribute levels if they only occurred in combination with the 'women's fertility attribute'. It appears to the ERG that it might have been better to use

different disutility weights for male and female patients, given the current design of the choice cards.

In the company submission, no information was provided regarding the experimental design of the DCE. Thus, the ERG asked for additional information in the clarification letter (Question B13.b). In their response, the company explained that a Partial Profile Design was used, to allow for the option of not showing all attributes on each choice card, but rather a subset of 12 attributes. However, no further information was provided. So, it is not evident if a (Bayesian) D-efficient design was used? Neither is it clear whether priors were used and if so, why and which. The ERG would also have preferred to receive details on the correlation matrix as the question may be raised to which extent the DCE-values are based on preference values or are (partially) a product of correlation in the design itself.

Using a sound experimental design for a DCE is of key importance to find valid preference values and the lack of details provided by the company make it very difficult to assess the design used by the company.

MNL model

The company explained in the CS that a multinomial logit model was used to analyse the choice data. As the choices were always between two alternatives, this reduces to a logit model. These models have three strong assumptions: independence from irrelevant alternatives (or IIA) assumption, the identical and independent distribution (IID) assumption for the error terms and preference homogeneity. No information was provided in the CS or in the response to the clarification letter regarding any formal testing to check if these assumptions are satisfied. A mixed logit model which allows for preference heterogeneity should at the very least have been tested. It is quite possible that this alternative model would have had a substantial impact on the results. Thus, the model used by the company is most likely too simplistic for decision making.

The company decided to use a model that did not include age of the hypothetical patient as attribute. Most likely, age had an impact on the weights of other attributes (through at least a two-way interaction) and thus the ERG does not agree with the interpretation given by the company: “Excluding age implied that the analysis effectively calculated the QALY weights for a hypothetical patient of average age.”[1]

The company used a simple additive model to estimate the QALY weights. In this model, the intercept was excluded, and the company referred to Viney et al. 2014 as justification.[1, 84] However, whilst Viney et al. indeed report that the impact of including an intercept on the calculated QALY weights was negligible in their study, this does not provide any justification for omitting the intercept in general. Instead, the validity of such choice should have been tested separately in the current study.

Attribute and level selection

The selection of attributes and levels has not been determined with the target population. A pilot testing or at least asking patients which key symptoms are deemed important would have

been of great value. Clinicians' preferences are often not the same as patients or general population preferences.

Validity of QALY estimates

In the result section (Table D46), the company showed both the life-years accumulated in both treatment groups as well as the QALYs, without discounting. It is striking to see in that in the metreleptin group 35.7 life years were accumulated, translating into 15.3 QALYs, whereas for the SoC group 24.7 life years were accumulated, translating into a mere 0.65 QALYs. A simple division shows that this implies for the metreleptin group that patients experience on average a QoL utility of 0.43, whereas for SoC patients this value is 0.03. The latter implies that the average patient with lipodystrophy not receiving metreleptin values his/her health state as very close to death, which seems highly unlikely.

In conclusion, given all the major flaws in the design of the DCE and the analysis of the data, the ERG considers the disutility weights presented by the company as speculative. This assessment is further confirmed by the highly unlikely model results regarding life years and QALYs.

Application of the disease attribute disutility estimates in the economic model

In the model, health states for each individual patient are characterised by the combination of a set of attributes, which serve as indicators of impairment. These attributes include organ abnormality (liver, heart, kidney and pancreas), hyperphagia, female reproductive dysfunction/infertility, loss of ability to perform at work/school, impaired physical appearance and metabolic abnormalities (such as failing to control triglycerides and HbA_{1c} levels). Each attribute level is associated with a utility decrement obtained from the discrete choice experiment study described above. These attribute levels are valued at every model cycle (1 year) to define an overall health state utility per patient. Table 33 shows the utility decrements used by the company in the economic model. Deterministic sensitivity analyses considered a 50% deviation from the mean value for the lower and upper limits. In the PSA, every utility decrement was assumed to follow a Beta distribution with the mean and standard error shown in Table 33.

Table 33: Utility decrements used in the cost effectiveness analyses

Attribute	Mean value	Standard error	Source
Heart Abnormality	-0.19	0.047	Company DCE and assumptions[1]
Liver Abnormality	-0.15	0.038	
Pancreas Abnormality	-0.13	0.032	
Kidney Abnormality	-0.13	0.028	
Hyperphagia	-0.11	0.015	
Disruption to female reproductive function	-0.06	0.064	
Loss of ability to perform work / school	-0.25	0.047	
Impaired Physical Appearance	-0.10	0.025	
Triglycerides: Achieved Goal (<=200 mg/dL)	0.00	NA	
Triglycerides: Partial Response (>200 mg/dL, <=500 mg/dL)	-0.05	0.012	
Triglycerides: No Response (>500 mg/dL)	-0.11	0.028	
HbA_{1c}: Hypoglycemia	-0.01	0.004	
HbA_{1c}: Achieved Goal (>4.0, <=7.0)	0.00	NA	
HbA_{1c}: Partial Response (>7.0%, <=8.0%)	-0.08	0.02	
HbA1C: No Response > 8.0%	-0.18	0.045	
Source: Table D37 and the electronic model in the CS[1]			

ERG comment:

The utility decrements derived from the company's DCE were used in the economic analyses since the characteristics valued by the DCE were similar (but not identical) to those collected in the NIH study. The effect of changes in utility decrement values was explored via sensitivity analyses. However, there are several attributes that the company mentioned as having impact on the patient's quality of life, which were not included in the economic analyses without further justification. These include pain, depression, retinopathy, neuropathy and amputation.

Despite the significant number of adverse events described in Section 5.3.3.5, only hypoglycaemia was included in the cost effectiveness analysis as an adverse event (with an associated utility decrement). No effort has been made to quantify the possible impact of other adverse events on patients' quality of life.

The CS (CS, Section 12.1.3, page 151) states that the true utility decrement associated with hyperphagia is likely to be underestimated since, according to the company, the "DCE cannot fully encompass the patient experience of such a unique aspect of the disease".[1] To quantify the impact of the utility decrement associated with hyperphagia on the cost effectiveness analyses, the company presented a scenario where this decrement was doubled. For further discussion on the utility decrement associated with hyperphagia the company refers to Section

Error! Reference source not found., Appendix 5 in the CS.[1] The ERG considers that a similar discussion and (when deemed necessary) scenario analyses on the remaining utility decrements should have been provided by the company.

The ERG identified some inconsistencies and programming errors in the cost effectiveness model submitted by the company. Firstly, the cell formula used in assigning disutilities to organ impairments in the metreleptin arm patients was different from that in the SoC arm, each formula followed different approaches with differing underlying assumptions. Secondly, both of the formulae used were not clear and not explained in the company submission. Finally, the ERG identified errors and logical inconsistencies in both of them.

The formula used in the metreleptin arm seemed to calculate the organ impairment associated disutilities from the real-world data (on the specific organ type impairment) until the data became no longer available. After that, the estimated cumulative number of organ impairments in each cycle was translated to the conditional probabilities for having a specific type of organ impairment at that cycle (e.g. probability of having a kidney impairment at a cycle given that the estimated total number of organ impairments is three at that cycle). In this translation, for each patient, the organ type assignment weights provided in Table D37 in the CS (CS, Section 12.2.6, page 164) were applied to the estimated cumulative number of organ impairments at each cycle independently, e.g. the probability that a patient has a kidney abnormality at a given cycle does not depend on whether or not that patient had a kidney probability in the previous cycle.[1] Furthermore, the ERG identified some errors in calculating the conditional probabilities (i.e. conditional probability of having specific type of organ impairment given a cumulative number of organ impairment). These errors led to inconsistent results, for instance, if the number of organ impairments of a patient at a given cycle is 4, the conditional probability for having a pancreas impairment would be equal to 1 (as well as having a heart, a liver or a kidney impairment). However, the formula used in the metreleptin arm, due to the errors in conditional probability calculations, provides incorrect estimates, for instance for some organs a probability value that is less than 1 and for the others a probability value that is more than 1.

In the formula used in the SoC arm, it was assumed that the type(s) of the organ(s) impaired at baseline stays impaired until the end of the time horizon. Therefore, the knowledge on the specific type of organ impairment at baseline was taken into account, while estimating the conditional probability for a specific organ impairment, given a cumulative number of impaired organs at a cycle. This seemed to be a more plausible approach, since some of the organ impairments are permanent conditions. However, the cell formula in the electronic model was not clear and not transparent and the ERG suspected some programming errors in this formula, such as using weights related to pancreas while calculating heart impairment related disutilities etc.

The ERG considers that the formula used in both arms to assign disutilities should be consistent. Therefore, in the corrected version of the company submission model, the ERG implemented the corrected version of the formula applied in the SoC arm to both arms. The impact of the correction of this error (together with the other programming errors) on the cost effectiveness results can be seen in the corrected CS base-case analyses in Section 6. In one of

the additional scenario analysis in Section 6, the ERG explored the impact of applying the alternative corrected formula from the metreleptin arm in both arms. Note that the same formulae were used while assigning organ impairment associated costs in the model, as well.

The systematic literature review conducted by the company identified only one study reporting on HRQoL in LD patients.[31] This study from Dhankar et al. 2015 collected data from the Lipodystrophy Connect Registry and reported an average estimated EQ-5D score associated with LD of 0.67. The ERG agrees with the company that EQ-5D domains might not provide an adequate perspective on quality of life for LD patients, and therefore the value reported by Dhankar et al. might not be fully appropriate.[31] However, given the lack of additional HRQoL data, and given the issues with the utility scores obtained by the DCE study as discussed previously, we present the results of some exploratory scenario analyses in Section 6, where the utility estimate from Dhankar et al. is multiplied by the life years gained obtained from the model, in order to get another estimate of QALYs gained (metreleptin vs. SoC).

5.3.3.8 Resources use and costs included in the model

Resource use associated with metreleptin treatment estimated using resource use questionnaires completed by two clinical advisers who treat lipodystrophy at Addenbrooke's Hospital.

Currently, only 11.3 mg vials (10 mg dose) metreleptin are available at a list price of £2,335 per vial. The availability of smaller vial sizes is expected within three months of submission of the variation to marketing authorisation, at a list price of £1,167.50 for a 5.8 mg vial (5 mg dose) and £583.80 for a 3 mg vial (2.5 mg dose). Based on the distribution (11.54% 10 mg dose; 69.23% 5 mg dose; 19.23% 2.5 mg dose) of observed current doses in the UK early access programme (EAP), an average annual per patient price of £434,633 is assumed in the analysis. Due to a loss of drug exclusivity after 10 years, a decrease of 90% of the list price of metreleptin was assumed in the model in one of the scenario analyses.

The costs related to standard of care treatment was estimated at £3,000 per patient per year.

In the CS, it was stated that the costs of home delivery and self-administration training will be funded by the company at no additional cost to patients or the NHS. Additional resource use costs, such as laboratory tests and office visits, are assumed to occur equally for both metreleptin and standard of care treatment and are assumed to be reflected in the nominal 'standard of care' costs. Standard of care costs were thus assigned to all patients in the model at each cycle.

A patient's health state is characterised by the presence or absence of abnormalities of the heart, kidney, liver, and/or pancreas. For each lipodystrophy-related complication, a patient's periodical costs are estimated based on their probability of occurrence of the complication and probability of survival in that period (Table 34). Unit medical costs for each complication were estimated based on NHS reference costs (Table 35). In the CS, it was stated that the following formula was used to estimate the cost per patient with organ abnormality:

estimated cost per patient with abnormality =

*(number of lipodystrophy-related inpatient stays per annum per patient / fraction of patients with abnormality) * cost per inpatient stay.*

In the model, no costs for hyperphagia, PCOS, inability to perform school or work, impaired physical appearance, or abnormal laboratory levels were included. Only adverse event cost of hypoglycaemia was included in the model at a price of £1,087.07 per hypoglycaemia-hospital admission.

Table 34: Estimated cost per patient with abnormality

Disease attribute	Estimated cost per patient with abnormality
Per-period medical costs from lipodystrophy-related complications	
Heart abnormality	£1,093.94
Renal abnormality	£590.04
Liver abnormality	£527.97
Pancreas abnormality	£44.28
Hyperphagia	£0
PCOS (Females Only)	£0
Unable to Perform School or Work	£0
Impaired Physical Appearance	£0
Per-period medical costs from non-achievement of triglyceride and/or glucose HbA_{1c} response	
Triglycerides Control	
Triglycerides: Achieved Goal (<=200 mg/dL)	£0
Triglycerides: Partial Response (>200 mg/dL, <=500 mg/dL)	£0
Triglycerides: No Response (>500 mg/dL)	£0
Glucose Control	
HbA_{1c}: Achieved Goal (<=7.0)	£0
HbA_{1c}: Partial Response (>7.0%, <=8.0%)	£0
HbA_{1c}: No Response > 8.0%	£0
Source: Table D40 in the CS[1]	

Table 35: National schedule of reference costs associated with lipodystrophy-related complication

Lipodystrophy-related complications	HRG currency codes
Heart abnormality	Weighted cost of total HRGs currency codes relating to coronary artery bypass: ED22A, ED22B, ED22C, ED23A, ED23B, ED23C, ED24A, ED24B, ED24C, ED25A, ED25B, ED25C, ED26A, ED26B, ED26C, ED27A, ED27B, ED27C, ED28A, ED28B, ED28C - NHS Ref costs relating to coronary artery bypass
Renal abnormality	Total of pre-transplant costs, transplant costs, and follow up outpatient costs. Total of LA10Z £232.52, + weighted cost of pre-transplantation workup costs LA11Z LA12A LA12B £373.44, + weighted costs of examination post-transplantation £233.69, + weighted cost of kidney transplant = £15716.14, + outpatient attendances for service code 102 £307.09
Liver abnormality	Weighted cost of total HRGs currency code GA01A, GA01B, GA01C, + outpatient attendances for service code 102 £307.09
Pancreas abnormality	Weighted average cost per FCE of elective inpatients, non-elective long stays, non-elective short stays for endocrine disorders KA08A, KA08B, KA08C
Source: Table D39 in the CS[1]	

ERG comment:

Currently, only 11.3 mg vials of metreleptin are available. In the submission, the availability of different vial sizes (5.9 mg and 3 mg) was assumed. The company confirmed, in their response to the clarification letter, that only 11.3 mg vials will be available at the time of marketing authorisation, but the approval of the other smaller vial sizes is expected within three months of submission of the variation.[39] All three vial sizes were used in the calculation of a weighted average annual drug acquisition costs for metreleptin (£434,633). This weighted average was based on the number of patients in Addenbrooke's Hospital expected to be treated with each vial size. The company adjusted the current dose mix at Addenbrooke's Hospital for potential increase. Therefore, they considered that six patients on 2.5 mg would be switched on 5 mg over time. The adjusted proportion of patients receiving each vial size is reported in Table 36. The detailed information on the adjusted vial use was not provided by the company (e.g. patient characteristics of the EAP patients were missing). Since the considered vial sizes are still not available yet, and the generalisability of the patients from the Addenbrooke's Hospital to the UK LD population, the ERG considers that there is a substantial amount of uncertainty on the drug acquisition costs for metreleptin.

Table 36: Proportion of EAP patients receiving each vial size

Vial	Proportion	EAP data
11.3 mg vial (administers up to a 10 mg dose)	11.54%	based on n=3
5.8 mg vial (administers up to a 5 mg dose)	69.23%	based on n=18
3 mg vial (administers up to a 2.5 mg dose)	19.23%	based on n=5
Source: Table 22 in the first response to the CL[39]		

The costs associated with standard of care are estimated at £3,000 and were applied to patients in both treatment arms. The ERG requested from the company an explanation how this estimate was calculated. In their response to the clarification letter, the company stated that the cost of standard of care was more like a nominal figure. Furthermore, the company stated that the SoC costs can be set to zero in the model with minimal impact on the ICER.[39] The ERG considers that for the SoC annual cost input for the model, rather than a nominal figure, an evidence-based figure should have been used, which is based on the expected health resource use of LD patients in the UK. In Section 6, results from the exploratory scenario analyses will be presented, where the annual cost for the SoC is varied to different values.

In the CS, it was stated that the estimated cost per patient with an abnormality was based on costs associated with an inpatient hospital stay, fraction of patients with that abnormality, and the number of lipodystrophy-related inpatient stays per patient.

The ERG requested from the company to provide details of the estimation of the abnormality costs per patient. In their response to the clarification letter, the company stated that costs per inpatient hospital stay for each organ were computed using the Health Resource Group (HRG) currency codes on Table 35, which yielded values of £11,888 for heart, £16,556 for kidney, £22,104 for liver, and £1,301 for pancreas abnormality.[39] However, it was still not clear to the ERG, how these values were derived from the HRGs.

Similar to the organ impairment associated disutility calculations explained in Section 5.3.3.7, the ERG identified the same programming errors while calculating the expected costs caused by the organ impairments. These errors will be corrected and the impact of the correction of programming errors (and using alternative formulae) will be explored in Section 6 of this report.

The company stated that no costs were included for hyperphagia, PCOS, inability to perform school or work, impaired physical appearance, or abnormal laboratory levels, because costs for these attributes were hard to quantify and varied substantially. Furthermore, the company argued, based on the NIH follow-up study, that these attributes were more likely to occur prior to metreleptin treatment than after metreleptin treatment. Therefore, setting the costs equal to £0 was deemed to be conservative.

It is a limitation that these costs were not included and no estimate was provided from the observed resource use from the literature or NIH follow-up study or other studies. However, the impact of ignoring these costs seems to have negligible impact on incremental costs. In Section 6, results from the exploratory scenario analyses will be presented, where these attribute costs are varied to different values.

Since a large number of assumptions and data were based on the expert opinion from two clinical advisors who treat lipodystrophy at Addenbrooke's Hospital, the ERG asked the company to provide all details of the communication between the company and these clinical experts. Furthermore, details on the justifications for clinical assumptions used in the model were requested. However, very little information on these requested items was provided by the company to the ERG. Therefore, the validity of some assumptions remains unclear.

The only adverse event costs to be incorporated in the analyses were those of hypoglycaemia. Other treatment emergent adverse events, such as fatigue, neutralising antibodies, injection site issues, and weight decrease were not deemed likely to have a large impact on the cost-effectiveness analysis by the company. It is likely that AEs like fatigue, neutralising antibodies, and injection site issues involve a certain amount of adverse event costs. The ERG is of the opinion that, although the impact of these AEs on the cost-effectiveness analysis can be marginal, the costs related to these AEs should have been included in the model, for completeness.

5.3.4 Model evaluation

The results of the health economic analysis are presented in terms of the incremental QALYs and incremental costs for metreleptin versus standard of care. The CS also included the results of the deterministic sensitivity analysis (DSA) and a probabilistic sensitivity analysis (PSA). In the PSA, alternative parameter values were simulated while not varying the set of patients included in the model. The following groups of parameters were sampled in the PSA:

- Costs of treatment
- Utility decrements of lipodystrophy-related complications
- Organ abnormality transition probabilities
- Discontinuation rate
- Probability estimates for number of organ abnormalities
- Discount rate

For the PSA, a value of 25% of the base value of the parameters was used as the standard error of many of the parameters, since many parameter inputs were not from taken from literature or estimated from clinical data, but assumption-based. In addition to the PSA, the results of a number of deterministic one-way and multi-way sensitivity analyses and scenario analyses were also presented in the CS (see Box 2).

Box 2: Sensitivity and scenario analyses presented within the CS

Deterministic one-way sensitivity analyses

- Utility decrements
- Annual cost of lipodystrophy-related complications
- Annual treatment costs per patient
- Model specifications
    - Discount rate costs
    - Discount rate life years and QALYs
    - Annual medical cost increase
    - Annual pharmacy cost increase
- Organ progression probabilities
- Relationship between organ abnormality and survival
- Time horizon of 30 years

Deterministic multi-way sensitivity analyses

- Assumes a lower price for metreleptin
- Doubles the hyperphagia utility decrement
- Incorporates resolution of heart abnormalities for some patients who experience a resolution of hypertension

Scenario analyses

- Future price changes
- Reduced initial price
- Elimination of mortality benefit of metreleptin for PL patients
- Changes to assumptions regarding organ abnormality progression
- Alternate survival extrapolation methods
- Earlier treatment initiation

ERG comment:

The company, in its response to the clarification letter, submitted an updated electronic model. The following changes were implemented to the original model in the updated version.

- A longer time horizon was used (90 years instead of 60 years)
- A mortality cap is implemented, which will take the corresponding age and gender adjusted mortality figure from the general UK population, if the survival estimate for a GL or PL patient generates a lower mortality estimate (hence LD patients will always have higher mortality than the UK general population)
- Transition probabilities for organ impairment were changed for both metreleptin arm and SoC arm patients due to the updates of the data from the NIH Follow-up study as well as the change of the matching method used (organ impairment progression probabilities estimated for the metreleptin and SoC arms from both the original and the updated models are given in Table 37 below).

- The imputation approach for the hyperphagia was updated. Previously in the original model, if there was no real-world data on hyperphagia in the second visit (during which hyperphagia was assessed), it was assumed that the patient had no hyperphagia. In the original model, if there is no real-world data, the patient is assumed to have a 9% probability of having hyperphagia (average baseline incidence of hyperphagia in the NIH Follow-up study).
- Some of the PSA and DSA settings were adjusted (upper and lower bounds for the metreleptin drug acquisition costs and the clinical inputs from the NIH follow-up and GL/PL natural history studies were updated, and the transition probabilities were sampled from Beta distribution in the new version, in comparison to the Normal distribution in the previous version)

Table 37: The estimated organ impairment progression probabilities in the original and in the updated versions in the electronic model

Estimated progression probabilities for the updated model - NIH Follow-up study updated data			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	0.0393	2	1
1 to 2	0.0555	14	4
2 to 3	0.0652	44	20
3 to 4	0.0219	52	5
Estimated progression probabilities for the original model - NIH Follow-up study original data			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	0.054	4	1
1 to 2	0.050	13	5
2 to 3	0.083	47	17
3 to 4	0.039	48	7
Estimated progression probabilities for the updated model - Matched GL/PL Natural History Patients (using Mahalanobis matching)			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	0.0896	33	33
1 to 2	0.1305	41	35
2 to 3	0.0860	36	22
3 to 4	0.0047	22	4

Estimated progression probabilities for the original model - Matched GL/PL Natural History Patients (N=47)			
Progression event	**Estimated progression probability**	**Number of patients at risk**	**Number of progressions**
0 to 1	0.089	36	36
1 to 2	0.173	42	39
2 to 3	0.123	44	36
3 to 4	0.062	36	16
Sources from top to the bottom: from Table 71 from the CS; Table 7 from the second tier of the response to the clarification letter; from Table 78 from the CS and Table 9 from the second tier of the response to the clarification letter.[1, 39]			

It should be noted that in the updated electronic model, for the SoC arm, the ERG noticed that the company used the wrong transition probability for estimating the risk of developing the 4th organ impairment. Instead of using 0.47% obtained from the matched untreated population from the GL/PL natural history study, the company used the estimate for the metreleptin patients from the NIH follow-up study (2.19%) in the model.

Furthermore, the ERG identified another programming error, which affected the company submission base-case. Due to the eligibility criteria of the original expected licensed indication, the company should have taken severe PL patients with triglycerides > 500 mmol/l and/or HbA_{1c} > 8% into account. However, the company applied the thresholds in a wrong way and applied these minimum thresholds as maximum thresholds. This wrong implementation of the license indication had excluded several severe PL patients from the base-case analysis. The ERG corrected these errors and present the corrected CS base-case analyses in Section 6.

5.4 Headline results reported within the company's submission

This section summarises the results of the economic analyses as presented by the company in its latest response to the clarification letter with the updated electronic model.[39] The company considered four different base case scenarios depending on the size of the vial and the price used for metreleptin. Thus, the results of the first base case scenario (BC1) are based on metreleptin list price and on a 10 mg vial size, which is currently being considered for marketing authorisation. However, it is expected that vials of 2.5 mg, 5 mg and 10 mg will be approved within three months after marketing authorisation. Therefore, the results of the second base case scenario (BC2) are based on metreleptin list price and on all available vial sizes. The results of the third and fourth base case scenarios (BC3 and BC4) are obtained from BC1 and BC2 after applying a [redacted]PAS price discount to metreleptin since the company expects this to be approved by PASLU.

5.4.1 Headline total QALYs and total costs for metreleptin versus standard care

Table 38 summarises the results of the economic analyses conducted for the four base case scenarios described above. Note that only discounted results are presented and that the difference in scenarios is only on the costs side of the analysis.

Table 38: Summary economic analyses results – company base case scenarios (discounted)

	LYs	QALYs	Costs BC1	Costs BC2	Costs BC3	Costs BC4
Metreleptin	18.36	8.56	£11,014,034	£5,652,808	[redacted]	[redacted]
SoC	14.71	0.25	£67,809	£67,809	£67,809	£67,809
Incremental	3.65	8.31	£10,946,226	£5,585,000	[redacted]	[redacted]
ICER	--	--	£1,316,932/ QALY	£671,927/ QALY	[redacted]/ QALY	[redacted]/ QALY

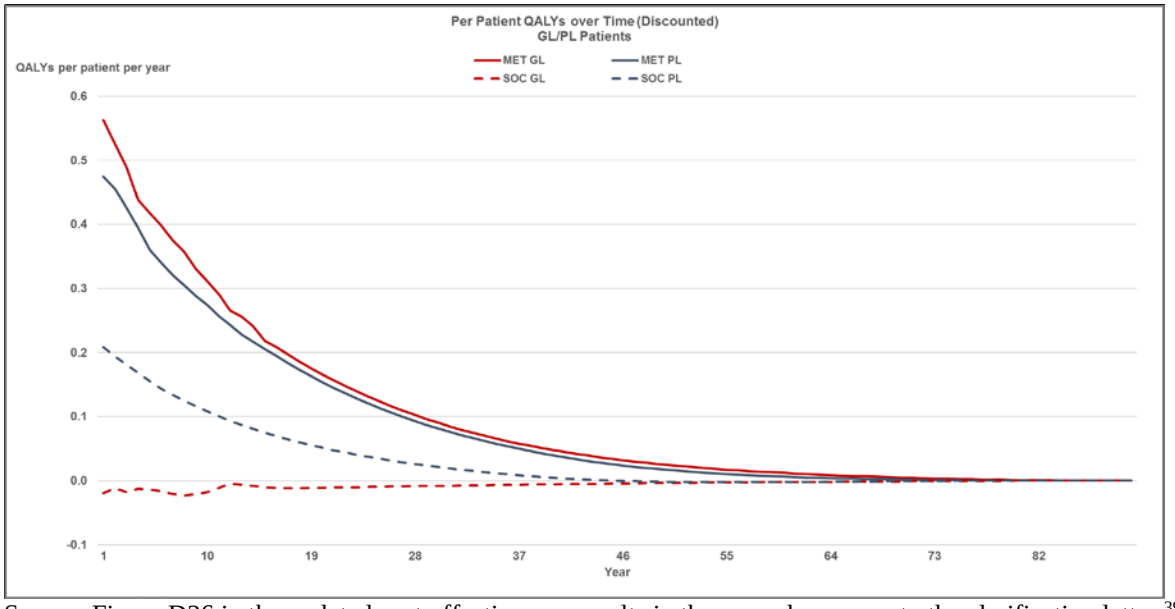
Sources: Table D44, D45 in the updated cost-effectiveness results in the second response to the clarification letter and Table 3 and 4 in the updated PAS submission template in the second response to the clarification letter.[39]

BC1: metreleptin list price and 10 mg vial size, BC2: metreleptin list price and multiple vial size, BC3: metreleptin PAS price and 10 mg vial size, BC4: metreleptin PAS price and multiple vial size.

Abbreviations: BC = base case, ICER = incremental cost-effectiveness ratio, LYs = life-years, QALYs = quality-adjusted life years, SoC = standard of care

In all scenarios, more than 99% of the total costs for the metreleptin arm are due to the cost of the therapy. Other medical costs are £26,156 in the four scenarios (less than 1% of the total costs). In the standard of care arm 65% of the total costs are due to therapy and 35% due to other medical costs. Life years and QALYs are accrued over a time horizon of 90 years. On average, metreleptin resulted in 39.04 (undiscounted) life years and 16.52 QALYs, whereas the standard of care arm resulted in 28.79 life years but negative (-0.19) QALYs. After discounting was applied, metreleptin resulted in 18.36 life years and 8.56 QALYs, and the standard of care arm resulted in 14.71 life years and 0.25 QALYs. The distribution of the QALYs per patient per year for both treatment arms and PL and GL patients separately is presented in Figure 7. In particular, this figure shows that for GL patients in the SoC arm the number of QALYs per year are always negative or zero suggesting that (from the general public point of view) these patients would rather die (at any time) than living with the disease.

Figure 7: QALYs per patient per year for metreleptin and SoC for PL and GL patients



Source: Figure D26 in the updated cost-effectiveness results in the second response to the clarification letter.[39]

ERG comment:
Results are generally well presented, although a discussion of the main results is missing in the company submission. In particular, the ERG considers that the face validity of the results regarding LYs and QALYs gained should have been explored. As mentioned in Section 5.3 of this report, the ERG has serious concerns about the validity of the QALYs presented by the company. Despite the significant amount of (undiscounted) life years predicted by the model in both arms, the number of QALYs was relatively low, especially in the SoC arm, which this was close to zero (or even negative when no discount was applied). Although the limitations of the study by Dhankar et al. 2015 were also discussed in Section 5.3 of this report,[31] this paper represents the only relevant source of utilities reported by the company. A naïve calculation using the average estimated EQ-5D score in Dhankar et al. (0.67) and the life years predicted by the company for the SoC arm would result in 19.29 and 9.86 undiscounted and discounted QALYs, respectively. These values are completely different to those presented by the company. Additional scenarios on utilities were explored by the ERG, and their results will be presented in Section 6 of this report.

Note that the ERG identified programming errors in the company base-case analyses, which are corrected and the impact of the corrections on cost-effectiveness of metreleptin is presented in Section 6.

5.4.2 Sensitivity analyses presented within the company's submission

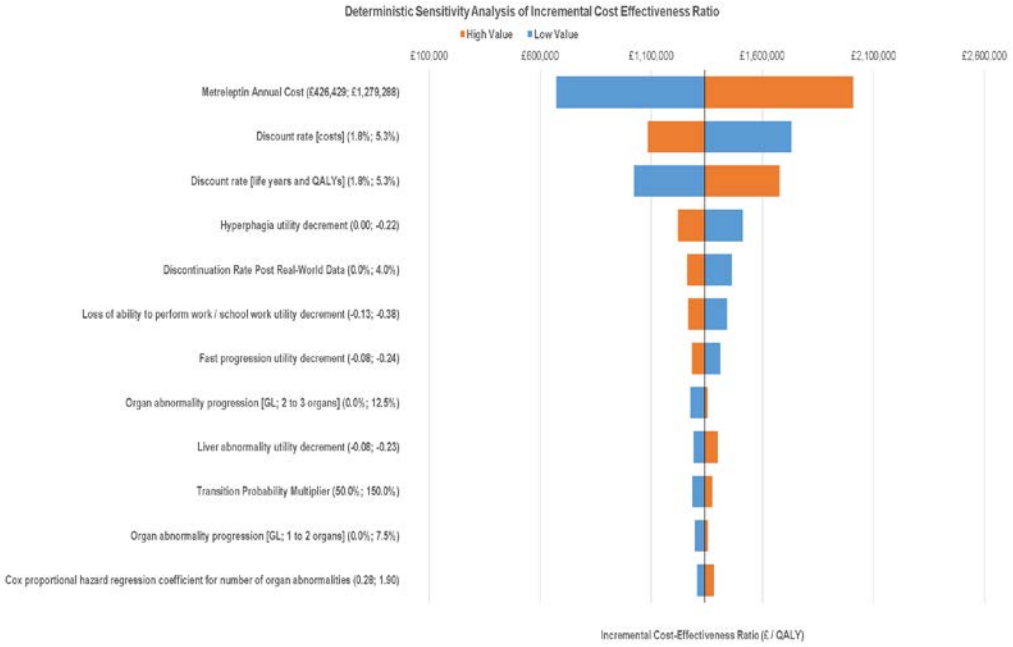
The company conducted a number of sensitivity, scenario a subgroup analyses. The results of all these analyses are summarised below. Only discounted results are presented here.

5.4.2.1 Sensitivity analyses

Sensitivity analyses included deterministic (DSA) and probabilistic sensitivity analyses (PSA). Univariate sensitivity analysis was performed on all single parameters of the model.

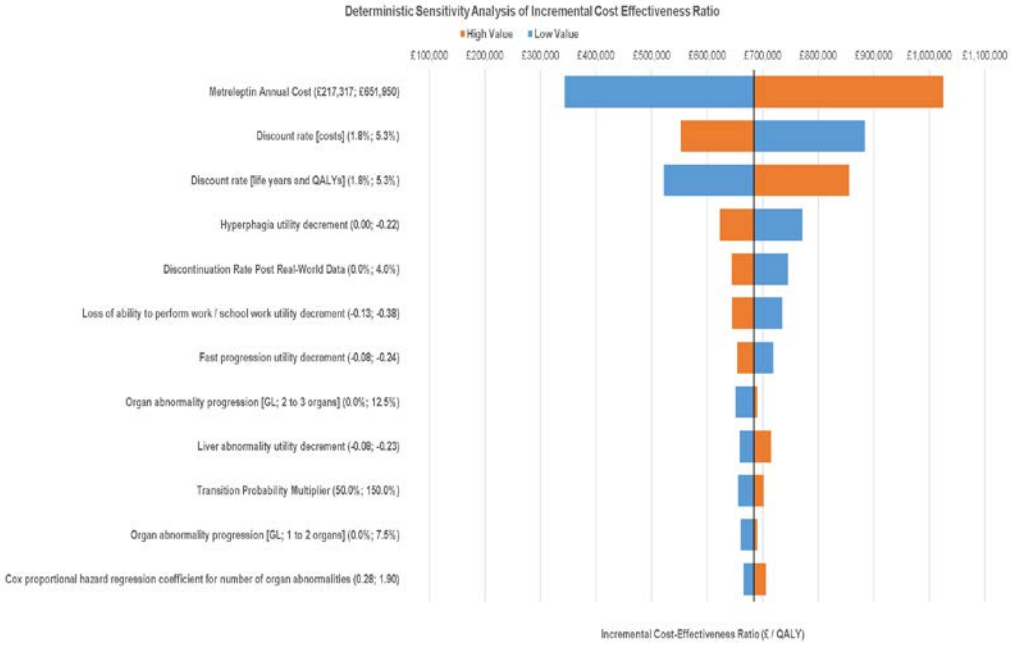
The results of the univariate DSAs were presented by the company as tornado diagrams and they are shown (for the four base case scenarios mentioned above) in the figures below. It was observed that in the four base case scenarios the metreleptin annual cost and the discount rates were the parameters for which the ICER was most sensitive. However, it should be noted that these parameters are typically not included in a DSA since they refer to structural/methodological uncertainty rather than parameter uncertainty. Besides these, the ICER was most sensitive to changes in the utility decrement due to hyperphagia and discontinuation rate.

Figure 8: Tornado diagram for BC1 – metreleptin list price and 10 mg vial size



Source: Figure D29 in the updated cost-effectiveness results in the second response to the clarification letter.[39]

Figure 9: Tornado diagram for BC2 – metreleptin list price and multiple vial sizes



Source: Figure D30 in the updated cost-effectiveness results in the second response to the clarification letter.[39]

Figure 10: Tornado diagram for BC3 – metreleptin PAS price and 10 mg vial size

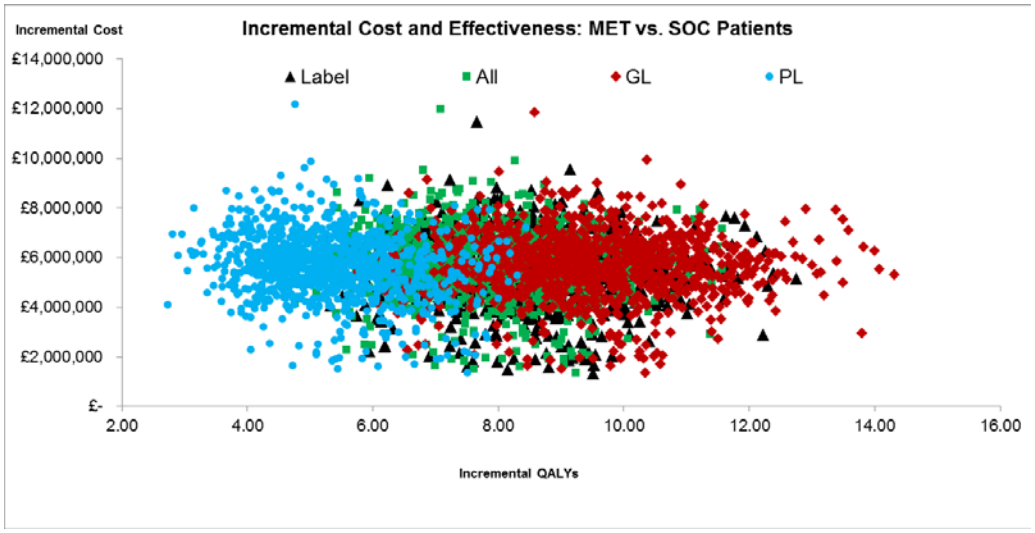
Source: Figure 1 in the updated PAS submission template in the second response to the clarification letter.[39]

Figure 11: Tornado diagram for BC4 – metreleptin PAS price and multiple vial sizes

Source: Figure 2 in the updated PAS submission template in the second response to the clarification letter.[39]

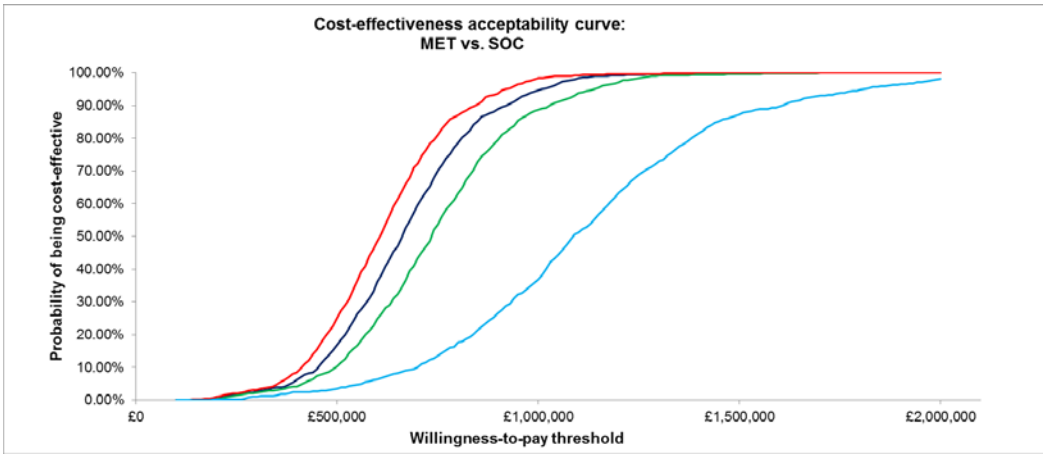
PSA was conducted using 1,000 model runs. The company presented results of the PSA as scatter plots of the total incremental costs and incremental QALYs on the CE plane and as cost effectiveness acceptability curves (CEACs). The PSA results were presented by the company for BC2 and BC4 only. The results of the two scenarios are presented in the figures below. Note that for BC1 and BC3, the only difference is on the cost side compared to BC2 and BC4. Therefore, the shape of the scatter plot of the PSA outcomes for BC2 and BC4 would be the same as that in BC2 and BC4, respectively, but shifted up on the incremental cost (Y) axis, which would result in less favourable CEACs for metreleptin.

Figure 12: PSA results on the CE plane – BC2: metreleptin list price and multiple vial sizes



Source: Figure 31 in the updated cost-effectiveness results in the second response to the clarification letter.[39]

Figure 13: CEACs – BC2: metreleptin list price and multiple vial sizes



Source: Figure 32 in the updated cost-effectiveness results in the second response to the clarification letter.[39]

Figure 14: PSA results on the CE plane – BC4: metreleptin PAS price and multiple vial sizes

Source: Figure 3 in the updated PAS submission template in the second response to the clarification letter.[39]

Figure 15: CEACs – BC4: metreleptin PAS price and multiple vial sizes

Source: Figure 4 in the updated PAS submission template in the second response to the clarification letter.[39]

ERG comment:

As in the base case analyses, the CS did not provide any interpretation of the results of the sensitivity analyses.

Parameters like time horizon, discount rates or the treatment costs are usually not included in the sensitivity analyses. The impact of changing these parameters on the ICER is usually assessed in scenario analyses. This is the approach followed by the ERG when presenting the results of their own analyses. In response to the clarification letter, the company indicated that the metreleptin cost per patient was included in the sensitivity analyses due to the uncertainty about the average per patient dose. However, the ERG considers that metreleptin cost should

not be explored in sensitivity analyses. If there are factors that impact annual metreleptin acquisition costs (such as patient dose), they should be varied independently from metreleptin price. In the updated version of the model submitted with the response to the clarification letter, the company did not include the time horizon in the sensitivity analyses as requested by the ERG. However, the discount rates were still included in the DSA and PSA. The analyses conducted by the ERG considered the discount rates fixed to 3.5% for both costs and effects.

The ERG found it unclear how the upper and lower limits for the parameters included in the DSA were obtained. The company indicated in the response to the clarification letter that since many parameters were assumption-based, ranges were selected to illustrate a wide set of reasonable values and that the bounds were updated to more clearly reflect the source of uncertainty. However, the ERG considers this still unclear since no discussion on the validity of these limits was provided. For those parameters that were derived from analysis of the NIH follow-up or natural history data, the updated version of the model included 95% CI limits in the DSA and the PSA. The ERG agrees with this latter choice. The ERG also identified some implausible values for some input parameters (e.g. negative standard deviations) and inappropriate probability distributions assigned to some parameters (e.g. normal distribution for disease progression or discontinuation rates, which might lead to negative estimates). The company corrected this in the updated version of the model.

PSA results were presented as scatter plots of total incremental costs and QALYs in the CE plane and CEACs with no further explanation. It is unclear why four different subgroups were presented in the CE plane and CEACs, as this was not the approach used in the base case scenarios or the DSAs. This makes the interpretation of the results more difficult.

5.4.2.2 Scenario analyses

The results of the scenarios run by company are shown in Table 39.

Table 39: Scenario analyses results

Scenario	Assumptions	QALYs gained	ICER BC1	ICER BC2	ICER BC3	ICER BC4
Base case	List price, with multiple vial sizes	8.31	£1,316,932	£671,927	[redacted]	[redacted]
Base case plus assume [redacted] lower price for Metreleptin	List price with [redacted] discount, with multiple vial sizes	8.31	[redacted]	[redacted]	--	--
Base case plus alternate inputs	Doubles hyperphagia disutility, incorporates heart abnormality improvement measured by hypertension	9.78	£1,132,896	£577,988	--	--
Base case plus alternative inputs assume [redacted] lower price for Metreleptin	List price with [redacted] discount, with multiple vial sizes, doubles hyperphagia disutility, incorporates heart abnormality improvement measured by hypertension	9.78	[redacted]	[redacted]	[redacted]	[redacted]
Future Price Changes: Loss of Metreleptin exclusivity	Metreleptin list price falls 90% after 10 years	8.31	£731,131	£373,391	[redacted]	[redacted]
Elimination of mortality benefit of Metreleptin for PL patients	PL patient survival is predicted from the general population curve based on patient age, regardless of less of organ abnormality.	8.31	£1,321,485	£674,235	[redacted]	[redacted]
Changes to assumptions regarding organ abnormality progression: Slower or faster organ progression risk for both metreleptin and standard of care patients	all organ progression probabilities increased by 50%	8.03	£1,346,604	£687,076	[redacted]	[redacted]
	all organ progression probabilities decreased by 50%	8.68	£1,276,347	£651,156	[redacted]	[redacted]
Changes to assumptions regarding organ abnormality progression: Alternative standard of care progression rates	Unadjusted natural history study organ abnormality progression probabilities used for standard of care patients (See Table 1 in appendix 17.6.1)	8.26	£1,326,825	£676,952	[redacted]	[redacted]

Scenario	Assumptions	QALYs gained	ICER BC1	ICER BC2	ICER BC3	ICER BC4
Alternate survival extrapolation methods: GL curve parameterisation	Weibull	8.67	£1,292,851	£659,609	[redacted]	[redacted]
	Log Normal	8.52	£1,302,991	£664,820	[redacted]	[redacted]
	Logit	8.32	£1,315,472	£671,192	[redacted]	[redacted]
Alternate survival extrapolation methods: GL organ abnormalities	GL organ abnormality cox regression coefficient: [Lower DSA bound, 0.275]	8.42	£1,276,963	£651,353	[redacted]	[redacted]
	GL organ abnormality cox regression coefficient: [Upper DSA bound, 1.904]	8.07	£1,360,883	£694,567	[redacted]	[redacted]
Alternate survival extrapolation methods: PL organ abnormalities	Observed general population curve corresponds to an average of 1 abnormal organ (2.76 in base case)	8.28	£1,266,105	£646,143	[redacted]	[redacted]
Early treatment initiation at age 1 (CGL)	List price, multiple vial sizes	12.35	--	£865,667	[redacted]	[redacted]
Early treatment initiation at age 1 (CGL) plus alternate inputs	List price, multiple vial sizes plus double hyperphagia decrement, plus parental disutility of -0.05 per period	14.51	--	£736,750	[redacted]	[redacted]

Sources: Table D51, D52 in the updated cost-effectiveness results in the second response to the clarification letter and Table 5 and 6 in the updated PAS submission template in the second response to the clarification letter.[39]
BC1: metreleptin list price and 10 mg vial size, BC2: metreleptin list price and multiple vial size, BC3: metreleptin PAS price and 10 mg vial size, BC4: metreleptin PAS price and multiple vial size.
Abbreviations: BC = base case, ICER = incremental cost-effectiveness ratio, LYs = life-years, QALYs = quality-adjusted life years, SoC = standard of care

ERG comment:

In general, the ICER is rather stable across all scenarios (per base case). The lowest ICER ([redacted]) was found for the scenario with [redacted] discount on metreleptin list price, assuming multiple vial sizes, doubled hyperphagia disutility and incorporating heart abnormality improvement measured by hypertension. The company argued that this scenario reflected the true metreleptin benefit. However, the ERG does not agree with that statement because there is no evidence that hyperphagia disutility should be twice as high from its DCE study estimate and also the argument that hypertension improvement is a surrogate for heart organ abnormality is deemed to be not convincing by the ERG.

5.4.2.3 Subgroups analyses

The following four subgroups were included in the economic analyses: generalised lipodystrophy (GL) patients (including those who do not meet the labelled indication), partial lipodystrophy (PL) patients (including those who do not meet the labelled indication), all NIH patients (including those who do not meet the labelled indication) and congenital generalised lipodystrophy (CGL) patients (including those who do not meet the labelled indication). A detailed description of these subgroups can be found in Section 2.2 of this report. The subgroup analyses were conducted by selecting the model results from those patients who meet the subgroup criteria. Discounted results are presented in Table 40.

Table 40: Summary results of the company subgroup analyses (discounted)

Subgroup	Number of patients	LYs		QALYs		ICER BC1	ICER BC2	ICER BC3	ICER BC4
		MET	SoC	MET	SoC				
All NIH	112	19.31	16.39	8.42	0.74	£1,469,868	£749,758	[redacted]	[redacted]
GL	68	17.98	13.61	8.87	-0.52	£1,202,792	£613,793	[redacted]	[redacted]
PL	44	21.37	20.68	7.73	2.68	£2,237,881	£1,140,745	[redacted]	[redacted]
CGL	48	19.27	14.77	9.57	-0.91	£1,170,263	£597,107	[redacted]	[redacted]

Sources: Table D54 (BC1), D56 (BC2) in the updated cost-effectiveness results in the second response to the clarification letter and economic model (BC3 and BC4).[39]
BC1: metreleptin list price and 10 mg vial size, BC2: metreleptin list price and multiple vial size, BC3: metreleptin PAS price and 10 mg vial size, BC4: metreleptin PAS price and multiple vial size.
Abbreviations: BC = base case, CGL = congenital generalised lipodystrophy, GL = generalised lipodystrophy, ICER = incremental cost-effectiveness ratio, LYs = life-years, MET = Metreleptin, NIH = National Institute of Health, PL = partial lipodystrophy, QALYs = quality-adjusted life years, SoC = standard of care

ERG comment:
The subgroups are in line with the scope of the NICE.[27] Subgroup analysis results show that the lowest ICER was obtained for the CGL subgroup, which is also very similar to the ICER for the GL subgroup. The highest ICER was found for the PL subgroup, which approximately doubled the ICER for the CGL subgroup.

In all subgroup analysis, for each subgroup, the average results of the patients that fall into the corresponding subgroup are calculated. This approach assumes that there is no difference in terms of transition probabilities (for disease progression or survival), health care resource utilisation and utilities among all subgroups. The ERG asked the company to check the plausibility of this assumption based on the patient level data from the NIH follow-up and natural history studies. Due to the small size of both the NIH follow-up study (n=112) and the natural history study (N=178), the company deemed not feasible to estimate transition probabilities (and hazard ratios) for each subgroup. Survival however was significantly different for GL and PL patients. Therefore, survival curves and the mortality hazard ratio associated with organ abnormalities was were estimated separately for GL and PL patients. The company considered that organ abnormality progression in the natural history study was not associated with lipodystrophy sub-type, in particular after an initial organ abnormality was observed. Thus, the company consider it plausible to use a single set of transition probabilities for both groups. Nevertheless, the company's model is set up to accommodate different transition probabilities for GL and PL. Hence, the impact of this assumption on the model results could be tested, should additional data become available in the future.

5.4.3 Validation

The whole of Section 12.7 (Validation) in the CS (CS, page 190) is the following sentence: "The approach to the model has been validated with leading lipodystrophy clinical experts including Dr. Rebecca Brown, Dr. David Savage and Dr. Anna Stears, and additional meetings to review findings are underway."[1] This sentence is provided under the company submission template heading "12.7.1 Describe the methods used to validate and cross-validate (for example with external evidence sources) and quality-assure the model. Provide references to the results produced and cross-reference to evidence identified in the clinical and resources sections". The ERG requested that the company provide all details of the validation methods, using the AdvisHE validation tool.[93] In the response to the clarification letter,[39] the company stated that the validation exercise reported in Section 12.7 of the CS specially involved discussing the conceptual mode, assumptions, and inputs with the clinical experts. Additional validation efforts were also completed, which were reported in the AdvisHE template submitted with the response to the clarification letter. However, not all types of validation were feasible due to the rare nature of lipodystrophy and lack of prior cost effectiveness analyses.

ERG comment:
The model was validated with leading lipodystrophy clinical experts and the validation tool was completed in response to the request for clarification. However, the ERG has some concerns regarding the model validation. With respect to face validity, the company stated that experts were asked to judge the appropriateness of the conceptual model, the input data, and

the model outcomes. However, the findings of the clinical experts were not reported. Furthermore, in Section 12.5.2 of the CS (CS, page178),[1] the company stated: "The outcomes from the model were not compared with the clinical trial results as no randomised controlled trial of Metreleptin in lipodystrophy patients has been conducted, largely due to the extreme rarity and severity of the condition". Thus, cross validation was not possible, as lipodystrophy is a rare disease and there are no existing cost effectiveness models.

Although the company provided more details of the validation of the model, most parts of the completed AdvisHE document were vague and not transparent. Therefore, the validation section is clearly inadequate.

5.5 Discussion of the available evidence relating to value for money for the NHS and PSS

This chapter focuses on the economic evidence about metreleptin for the treatment of LD syndromes, submitted to NICE by the company. The analysis from the company is a QALY-based cost effectiveness model comparing metreleptin versus SoC. In BC1 (metreleptin list price and 10 mg vial size), metreleptin is expected to result in 16.71 additional QALYs compared to SoC. The undiscounted incremental cost of metreleptin versus SoC is estimated to be £19,923,178 per patient. When discounted at a rate of 3.5%, the estimated QALYs gained were 8.31 for metreleptin treatment versus SoC. The discounted incremental cost of metreleptin versus SoC was £10,946,226 per patient, yielding an ICER of £1.3Million per QALY gained. The [redacted] ICER was reported for BC4 (metreleptin PAS price and multiple vial sizes), at [redacted] per QALY gained.

Several major problems relating to the company's submission were identified by the ERG. One of the most important concerns relates to the estimation of organ impairment progression. In the analysis, the type of affected organ and the severity of an organ abnormality were not taken into account. Organ impairment improvements were only considered for metreleptin treatment and organ impairment progression in the SoC arm was overestimated by the use of a staggering approach. Furthermore, the approaches used to incorporate time to event data from the NIH follow-up study and from the GL/PL natural history study were incompatible. The simulated number of impaired organs was biased in favour of metreleptin by use of an implausible formula in the electronic model. In addition, patient characteristics had no impact on the transition probabilities for the number of impaired organs. Due to the issues outlined above, the ERG has substantial concerns about the appropriateness of the statistical methods used by the company. In order to address these concerns, the ERG requested that the company conduct de novo statistical analysis, using more generally accepted methods in line with the guidance provided in NICE DSU TSD 17,[90] however, the company stated that they were not able to finalise this request given the timelines.

There are also serious concerns surrounding the survival analyses conducted by the company and the implementation of these analyses in the model. The estimation and extrapolation of the survival analyses from different datasets results in inconsistencies. There is also a lack of face validity for the survival extrapolation as the survival model estimates that after 65 years, over 23% of the patients are still alive. Considering that the average baseline age was 24 years, these survival estimates might not be valid for LD patient population. Survival is extrapolated by a

function based only on age, gender, type of lipodystrophy, and number of organs impaired, and it is questionable whether this is the most plausible survival function and whether other important covariates were missed.

The ERG identified several issues related to the matching methodology. The first issue is about the appropriateness of the company's approach to the use of data to inform the estimates of treatment effectiveness. Moreover, there is a lack of clarity regarding the matching algorithm used by the company. The ERG also had problems with the independent estimation of the organ impairment transition probabilities from the treated and the matched untreated patient datasets. Furthermore, insufficient interpretation of the matching results was provided.

There are also several issues identified by the ERG, which relate to the extrapolation of blood-lab measures (HbA_{1c} and triglycerides) and other attributes not related to organ damage conducted by the company in the model. Furthermore, while metreleptin discontinuation is only applied for organ impairment, the impact of discontinuation is not reflected in other disease attributes, which creates a bias in favour of metreleptin.

The ERG has several vital concerns about the derivation of the utility decrement from the company's DCE. The key issue is that the use of DCE to directly obtain disutility values for heath states is still in its infancy. The most striking unresolved methodological issue relates to the fact that DCE classifies health states far more often below zero than TTO and produces lower average health state values. This was indeed observed in the results of the current DCE study. In addition, various major flaws in the design of the DCE and the analysis of the data were identified, hence, the ERG considers the disutility weights presented by the company as speculative.

There are also a few issues related to resource use and costs included in the model, which lead to incompleteness of the model.

Finally, the ERG also has concerns about the sensitivity analyses and the validation of the model. Parameters like treatment costs and discount rates were included in the sensitivity analysis, although these parameters are usually not included in a DSA. It was unclear why the PSA results are presented in four different subgroups. The ERG considered the validation of the model to be inadequate and the information provided about the validation to be very vague and not transparent.

Given the level of evidence submitted by the company, it proved impossible for the ERG to give an indication on the cost-effectiveness of metreleptin. The CE model is based on non-reliable evidence and unjustified assumptions. More specifically, the RWD data used to estimate important inputs for the model is not reliable (e.g. twice data updates without being able to track what was been updated and how, vague definitions of organ impairment were applied). Additionally, both the methods used in quantifying the treatment effect and the DCE methodology used were not transparently reported but more importantly not credible.

The next chapter outlines the additional analyses conducted by the ERG, with the aim of addressing some of the problems identified in the critical appraisal of the economic analysis.

6. IMPACT ON THE COST-CONSEQUENCE ANALYSIS OF ADDITIONAL EXPLORATORY CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

6.1 Introduction

In this chapter the additional analyses performed by the ERG are presented. As described in Chapter 5, the ERG identified some programming errors in the model and some critical issues related to the input evidence used in populating the company's model.

First, the results are presented of a re-analysis of the company's economic analysis base-cases, following the correction of technical programming errors by the ERG.

Next, the results of several exploratory scenario analyses done by the ERG to explore areas of uncertainty will be presented.

6.2 Re-analysis of the company's economic analysis following the correction of technical programming errors

The ERG identified the following errors in the company model:

- Wrong transition probability is used for the fourth organ impairment annual probability for SoC
- The minimum HbA_{1c} and triglyceride thresholds for the PL eligibility were applied as maximum thresholds for PL patients
- The costs and disutilities associated with organ impairments were wrongly calculated, and different formulae were used for SoC and metreleptin arms

The base-case model's results after correcting these errors can be seen in Table 41 below.

Table 41: Summary economic analyses results – corrected company base case scenarios (discounted)

	LYs	QALYs	Costs BC1	Costs BC2	Costs BC3	Costs BC4
Metreleptin	18.47	9.12	£11,400,639	£5,850,224	[redacted]	[redacted]
SoC	14.99	0.43	£66,712	£66,712	£66,712	£66,712
Incremental	3.48	8.68	£11,333,927	£5,783,512	[redacted]	[redacted]
ICER	--	--	£1,305,355/ QALY	£666,101/ QALY	[redacted]/ QALY	[redacted]/ QALY

BC1: metreleptin list price and 10 mg vial size, BC2: metreleptin list price and multiple vial size, BC3: metreleptin PAS price and 10 mg vial size, BC4: metreleptin PAS price and multiple vial size.

Abbreviations: BC = base case, ICER = incremental cost-effectiveness ratio, LYs = life-years, QALYs = quality-adjusted life years, SoC = standard of care

As observed in Table 41, these errors do not seem to have a major effect on the cost effectiveness results (comparing to the values in Table 38). The subgroup analysis with the corrected CS model can be seen in Table 42 below.

Table 42: Subgroup analyses results – corrected company base case scenarios (discounted)

Subgroup	Number of patients	LYs		QALYs		ICER BC1	ICER BC2	ICER BC3	ICER BC4
		MET	SoC	MET	SoC				
All NIH	112	19.39	16.60	9.42	1.82	£1,486,050	£758,164	[redacted]	[redacted]
GL	68	18.09	13.92	9.78	0.39	£1,203,175	£614,091	[redacted]	[redacted]
PL	44	21.40	20.74	8.87	4.03	£2,334,659	£1,190,374	[redacted]	[redacted]
CGL	48	19.40	15.16	10.70	0.05	£1,152,297	£588,002	[redacted]	[redacted]

BC1: metreleptin list price and 10 mg vial size, BC2: metreleptin list price and multiple vial size, BC3: metreleptin PAS price and 10 mg vial size, BC4: metreleptin PAS price and multiple vial size.

Abbreviations: BC = base case, CGL = congenital generalised lipodystrophy, GL = generalised lipodystrophy, ICER = incremental cost-effectiveness ratio, LYs = life-years, MET = Metreleptin, NIH = National Institute of Health, PL = partial lipodystrophy, QALYs = quality-adjusted life years, SoC = standard of care

Again, the impact of the errors was relatively small. As in the company subgroup analysis, CGL patients have the largest gain in QALYs with metreleptin.

Total QALYs seem to increase in the corrected model, however incremental costs and ICERs seem to be similar. The ERG did not repeat the PSA and the DSA of the corrected model, since the results and the main findings are not expected to change substantially and the company's model is extremely slow. To explore structural and input uncertainty, the ERG conducted various scenario analyses. These scenarios are presented only for BC2 and BC4, as the impact of having/not having multiple vial sizes available on ICER is already known from the previous analyses.

6.3 Exploratory scenario analyses conducted by the ERG

The ERG conducted six additional scenario analyses to explore structural and input parameter uncertainty. These scenarios are described below:

- Scenario 1: The impact of metreleptin discontinuation was reflected in not only in organ impairment progression, but also in the progression of other disease attributes. For instance, when a patient on metreleptin discontinues the treatment, the corresponding values from the SoC arm were assumed for discontinued patients' blood-lab and other attributes (e.g. hyperphagia, ability to work, etc.)
- Scenario 2: Abandoning the logical constraint imposed on the SoC arm patients, which never allowed them to have fewer number of organ impairments than metreleptin
- Scenario 3: Assuming that there is no difference between the SoC and metreleptin treatments in terms of the disease attributes other than organ impairment and blood-lab values (e.g. hyperphagia, ability to work, physical appearance, etc.) during a patient's lifetime
- Scenario 4: Using utility input from Dhankar et al. for all the years that a patient is alive
- Scenario 5: Except for the data at baseline, no real-world data is directly used in the simulation of the organ/blood-lab attributes for the metreleptin arm patients

- Scenario 6: For the disutility and cost calculations associated with the number of organs impaired, the corrected formula from the metreleptin arm (assuming independent application of the organ specific abnormality probability weights) is used in both arms.

6.3.1 Results of the ERG's scenario analyses

The results from these exploratory scenario analyses are given in Table 43 below.

Table 43: Exploratory scenario analyses from the ERG

Scenario	Assumptions	QALYs metreleptin	QALYs SoC	QALYs gained	ICER BC2	ICER BC4
Base case	Multiple vial sizes	9.12	0.43	8.68	£666,101	[redacted]
Scenario 1	The impact of metreleptin discontinuation in other attributes	6.78	0.43	6.34	£911,588	[redacted]
Scenario 2	Abandoning the logical constraint imposed on the SoC arm patients	9.12	0.45	8.66	£667,515	[redacted]
Scenario 3	No change between the SoC and metreleptin treatments in terms of attributes other than organ impairment and blood-lab values	2.82	0.43	2.39	£2,424,009	[redacted]
Scenario 4	Using utility input from Dhankar et al. for all the alive years of the patient	12.38	10.05	2.33	£2,480,754	[redacted]
Scenario 5	Except for the data at baseline, no real-world data is directly used in the simulation of the organ/blood-lab attributes for the metreleptin arm patients	6.53	0.45	6.08	£881,810	[redacted]
Scenario 6	Alternative organ impairment associated cost/disutility calculation	8.28	-0.43	8.71	£663,725	[redacted]

Scenarios 3 and 4 had the highest impact on the results since the ICERs in these scenarios are three-fold larger than the ICER from the base case(s).

In scenario 3, the treatment effect of metreleptin on attributes like hyperphagia, ability to work was assumed to be zero. The impact on the ICER suggests that the treatment effect of metreleptin on these attributes is one of the key drivers of the cost effectiveness. It should be noted that the evidence on the effectiveness of metreleptin for these attributes was rather weak, therefore future research can definitely reduce this uncertainty.

Since the ERG was concerned about the utility estimates provided by the company (including the overall methodological DCE approach), scenario analysis 4 demonstrated how different the utility estimates used in the submission were compared to the EQ5D values from the literature and how changing the utility input to the model can change the results substantially

6.4 Discussion

As discussed in the previous section, the ERG considers that the evidence base used in this cost effectiveness analysis is not reliable and trustworthy enough to inform decisions on metreleptin. However, the ERG expects that the decision uncertainty from the payer perspective related to metreleptin's value for money would be rather low, in view of the fact that the ICER estimates from all analyses, including the analyses with PAS discounts, are markedly above the acceptable thresholds considered for orphan drugs.

7. COST TO THE NHS AND PSS AND OTHER SECTORS

7.1 Summary of submitted evidence relating to the costs to the NHS and PSS

The CS includes a budget impact model to estimate the total costs to the NHS, for a period of five years, of adopting metreleptin in England. Published data on the incidence and prevalence of lipodystrophy relevant to the expected metreleptin license were lacking. Since EAP data from a decade of metreleptin use in UK clinical practice were deemed relevant and representative, these data were used to estimate patient numbers for the budget impact analysis. In December 2017, there were 26 patients in the UK receiving metreleptin (nine patients with GL and 17 with uncontrolled PL). Based on expert opinion, it was assumed that yearly six new patients (two for GL and four for PL) are eligible for metreleptin treatment. For mortality, it was assumed that one patient with PL will die every year and one patient with GL will die every two years. Based on these assumptions, the number of patients treated with metreleptin will rise from 22 in year 1 to 44 in year 5. The estimated numbers of patients eligible for metreleptin treatment over the next five years are presented in Table 44.

Table 44: Estimated eligible patient numbers for metreleptin

Patient group	Year 1	Year 2	Year 3	Year 4	Year 5
GL	9	11	12	14	15
PL	17	20	23	26	29
Total	26	31	35	40	44

Source: Table D58 in the CS[1]

GL, generalised lipodystrophy; PL, partial lipodystrophy

It is assumed that the uptake rate will rise from 85% in year 1 to 90% in year 5, based on clinical expert opinion. A discontinuation rate of 0% in the first five years was assumed for metreleptin. The expected uptake rate of metreleptin is shown in Table 45.

Table 45: Expected uptake rate of metreleptin over the next five years

	Year 1	Year 2	Year 3	Year 4	Year 5
Uptake rate	85%	85%	90%	90%	90%

Source: Table D59 in the CS[1]

The first budget impact analysis assumed the availability of only 10 mg dose vials at a list price of £2,335, resulting in annual per patient drug costs of £852,859. Since all start-up costs concerning the administration of metreleptin will be covered by Aegerion, supportive medicines costs are expected to be zero. This resulted into a net budget impact of £18,762,893 in year 1 rising to £34,114,350 in year 5 and a cumulative net budget impact over years 1-5 of £133,045,965.

In the second budget impact analysis, it is assumed that 18.23% of the patients with lipodystrophy will receive a 3 mg dose (at a list price of £583.80), 69.23% will receive a 5 mg dose (at a list price of £1,167.50), and 11.54% will receive a 10 mg dose of metreleptin. This resulted in a net budget impact of £9,561,936 in year 1 rising to £17,385,338 in year 5 and a net cumulative budget impact of £67,802,818.

In the third analysis, a PAS discount of [redacted] was assumed for 11.3 mg vial (10 mg dose). The anticipated PAS price was [redacted] per 11.3 mg vial, which equates to treatment costs of [redacted] per patient per annum. In year 1, the net budget impact was [redacted] and rising to [redacted] in year 5. The cumulative net budget impact was [redacted] for all patients with lipodystrophy.

Budget impact analysis 4 assumed the availability of all three vial sizes and a PAS discount of [redacted]. Based on EAP data, it was assumed that 11.54% of the patients with lipodystrophy receive the 10 mg dose vial, 69.23% of patients receive the 5 mg dose vial, and 19.23% of patients receive the 2.5 mg dose vial. This resulted in a net budget impact of [redacted] in year 1 and [redacted] in year 5 (net cumulative budget impact over years 1-5 was [redacted]).

7.2 ERG critique of the company's budget impact analysis

In general, the ERG considers the assumptions made in the budget impact analysis as plausible. However, there are some concerns about the expected uptake rate of metreleptin, which is assumed to rise over the next five years from 85% in year 1 to 90% in year 5. The ERG requested that the company to provide all details of data used for this assumption. The company stated that this assumption was based on company forecast assumptions. The uptake is expected to be high, but due to potential barriers, some patients may be unwilling or unable to receive metreleptin. The ERG considers the high expected uptake rate as reliable, but the reason behind the rising uptake rate from 85% in year 1 to 90% in year 5 is still unclear. Furthermore, discontinuation of metreleptin was only included to reflect mortality of LD patients. However, discontinuation due to patient preferences or clinical recommendation was considered as 0% in the first five years, because of the small estimated patient numbers in the budget impact. Since the estimated discontinuation rate is based on clinical expert opinion and no detailed information on this expert opinion was provided to the ERG, the validity of these assumptions remains unclear.

8. IMPACT OF THE TECHNOLOGY BEYOND DIRECT HEALTH BENEFITS AND ON THE DELIVERY OF THE SPECIALISED SERVICE

8.1 Summary of cost savings estimated within the CS

8.1.1 Nature of estimates presented

The CS includes estimates of impacts of metreleptin on (i) inability to work or attend school for patients and carers; (ii) estimates of out-of-pocket costs for patients and carers including costs related to diabetes, transportation, fertility and cosmetic treatment; and (iii) other carer costs.

8.1.2 Societal costs

A substantial number of patients with lipodystrophy are affected from birth, with symptoms such as hyperphagia and organ abnormalities manifesting in childhood. Due to hyperphagia, patients may be highly constrained by food access issues, which can heavily affect their daily lives including attending school and work. In the NIH Follow-Up study, of 50 adult patients treated with metreleptin, 48% did not work of which at least 1/3 was due to lipodystrophy. Over half (59.4%) of the 64 non-adult patients treated with metreleptin had impaired school attendance.

Patients may need 24/7 supervision from carers. Carers are mostly family members, typically the mother of the patient. Of 114 patients treated with metreleptin in the NIH follow-up study, 35% had a caregiver who was not working or who was working part time due to supporting the patient. When patients were treated with metreleptin, only 7% of these patients had a caregiver who was not working or only working part time, which is a reduction of 80%.

8.1.3 Costs borne by patients

Most patients with lipodystrophy have type 2 diabetes at a very young age. Indirect costs due to diabetes are considerably high, which are to a large extent costs for the patients and their carers.[94] These costs include loss of earnings by the patients and carers. A study from the UK estimated the earnings lost at £869 to £13,841 per patient and at £1,300 to £10,960 per carer.[95]

Other out-of-pocket costs for patients and carers are costs related to transportation to the hospital. About 20% of patients with lipodystrophy will need hospitalisation in a given year. In some patients, more than five hospitalisations per year were observed.[96] Fertility treatment and cosmetic treatment are further potential costs, which are not always reimbursed by the NHS. However, the company stated that effective management of lipodystrophy, including metreleptin treatment, is expected to mitigate these costs.

Patients treated with metreleptin would typically need to visit the specialist centre at Addenbrooke's twice a year. Thus, patients will have travel costs to Addenbrooke's in Cambridge and they probably also need an overnight stay in Cambridge.

8.1.4 Other carer costs

Two different surveys have described the substantial time burden for the majority of people living with a rare disease and their carers, with 42% spending over two hours a day on caring.

[97, 98] In the NIH follow-up study, 35% of the 114 patients treated with metreleptin had one caregiver who was not working or only working part time. After metreleptin treatment, only 7% of the patients had a carer not working or only working part time. Data about time spent on informal care by family members for patients with lipodystrophy are currently lacking. However, the company states that it is currently conducting market research in England to further understand the impact on caregivers in more detail.

8.1.5 Discussion of wider societal (non-health) benefits

A number of issues regarding the impact of metreleptin beyond direct health benefits are mentioned in the submission. However, no costs associated with inability to work or attend school were calculated in the analyses. The company admits that these attributes may impose costs, though the costs vary substantially and are hard to quantify. Furthermore, these attributes are more likely to be present in patients who receive standard of care. Therefore, the company considered including £0 in associated costs to be a conservative approach. The ERG requested that the company justify the plausibility of these assumptions. The company responded that very limited information is available about the economic burden of lipodystrophy. Moreover, the costs associated with these attributes are likely to be highly variable. As part of the NIH follow-up study, data about the extent to which patients experience each of these attributes prior and after metreleptin treatment were collected.[39] The ERG does not understand that, while there were data collected on these attributes, it was not possible to estimate associated costs. Although these attributes are more likely to be present in patients not treated with metreleptin, these attributes could still be present in some patients treated with metreleptin.

The ERG requested that the company provide more details and the source of the hospitalisation figures (20% of lipodystrophy patients are hospitalised at least once a year, with some hospitalised more than five times a year). However, the company did not respond to this request. Furthermore, the ERG has a problem with the assertion in which the company stated that metreleptin will mitigate the costs of hospitalisation and fertility and cosmetic treatment, since this is not based on any evidence.

No indirect health care costs, due to additional life-years after receiving metreleptin, were reported in the CS. The company was requested to provide estimates for these costs. The company responded that the model was not designed to include these costs. Furthermore, it was not expected that any indirect health care costs would influence the cost effectiveness results. Although the company expects the indirect health care costs due to additional life-years to be low, these costs should be included in the model for completeness.

The estimates related to informal caregivers were obtained from the NIH follow-up study. It was stated that there were 114 LD patients in the NIH follow-up study, however, this does not match any of the numbers in the studies reported elsewhere in the CS. A substantial number of informal caregivers (family members of the patient) does not work or work part time due to taking care of the patient with lipodystrophy before metreleptin treatment. After metreleptin, 7% of these caregivers are still not working or are working part time. The CS does not include costs related to informal care and productivity loss for the caregiver. Although the company states that it is currently conducting research to gain more details of these issues, the ERG

considers it as inadequate that the impact of lipodystrophy on informal carers was not identified prior to the CS.

8.2 Staffing and infrastructure requirements associated with the use of the technology

The company stated that, since metreleptin has been available for over 10 years in the UK through the EAP, there is already a lot of expertise within the NHS to support the safe and effective use of metreleptin treatment. Healthcare professionals are training the patients on the proper use of subcutaneous injections, through which metreleptin could be administered at home by the patient or carer.

Furthermore, it was stated in the CS that no additional facilities, technology, or infrastructure will be required for the introduction of metreleptin treatment on the NHS in England.

9. DISCUSSION

9.1 Statement of principal findings – clinical effectiveness

Single arm, observation studies of metreleptin treatment found improvements in metabolic abnormalities from baseline to month 12 of treatment in patients with GL and in the subgroup of patients with PL who had similar metabolic disturbances to those seen in patients with GL (PL patients with leptin level <12 ng/ml with baseline HbA_{1c} ≥6.5% and/or triglycerides ≥5.65 mmol/L).

- In study NIH 991265/20010769, mean actual change in HbA_{1c} to Month 12/LOCF was -2.2% ($p<0.001$) for GL patients and -0.9% ($p<0.001$) for patients in the PL subgroup.[1, 37]
- In study FHA101, mean actual change from baseline to Month 12/LOCF for HbA_{1c} was -1.2% for GL patients and -0.8% for patients in the PL subgroup.[1, 38]
- In study NIH 991265/20010769, mean percent change in triglycerides to Month 12/LOCF was -32.1% ($p=0.001$) for the GL group and -37.4% ($p<0.001$) in the PL subgroup excluding the 1 outlying noncompliant patient.[1, 37]
- In study FHA101, mean percent change from baseline to Month 12/LOCF for triglycerides was similar in the GL group as -26.9%; however, for the PL subgroup, the mean percent change was lower at -8.5%. Five of the 7 patients in the PL subgroup in this study showed reductions from baseline to Month 12/LOCF in triglycerides ranging from -5.7% to -52.3%.[1, 38]

Mixed model repeated measures (MMRM) analyses indicate that these effects persist to month 36; LS mean percent changes from baseline in HbA_{1c} were -2.3%, -2.1% and -1.5% at Months 12, 24 and 36, respectively.[1, 37] The overall MMRM analysis showed a statistically significant decrease from baseline for GL patients with an LS mean change of -1.4% ($p<0.001$). Results were similar in the PL subgroup with LS mean changes in HbA_{1c} of -0.9%, -1.3%, and -1.0% at Months 12, 24, and 36 and an overall LS mean change of -0.6% ($p<0.001$).[1, 37] In the GL group, LS mean percent changes from baseline in triglycerides were -48.3%, -22.6% and -40.6% at Months 12, 24, and 36, respectively; based on the overall MMRM analysis, the LS mean change in triglycerides was -22.4% ($p<0.001$). For the PL subgroup (excluding data from the 'outlier' patient described previously), LS mean percent changes in triglycerides were -36.2%, -31.7%, and -13.7% at Months 12, 24 and 36, respectively, with an overall LS mean change of -18.6% ($p=0.004$).[1, 37]

With respect to safety and adverse events, the CS concludes that the known side effects of metreleptin can be managed as part of the normal clinical practice for patients with this complex condition. The CS does not report the safety concerns as highlighted in the Centre for Drug Evaluation and Research Report (not included in the CS) or the associated REMS.[75] The summary of safety in this report states: 'The principal safety concerns with metreleptin are T-cell lymphoma and anti-metreleptin antibodies with neutralizing activity. These concerns are of sufficient magnitude to require REMS. Other safety findings that warrant inclusion in the Warning and Precautions section of the metreleptin labelling include hypoglycemia, autoimmunity, and hypersensitivity.'

9.2 Statement of principal findings – cost-consequence evaluation, NHS budget impact and societal analysis

A systematic review of economic evaluation studies of patients with lipodystrophy was included in the CS. Three economic evaluation studies were identified by the company. However, none of these studies were eligible for the economic evaluation of metreleptin, since the scope of all studies was not relevant to the CS.

A patient-level model was developed, aiming to assess the cost effectiveness of metreleptin versus standard of care for patients with lipodystrophy.

Individual patient data was obtained from the NIH follow-up study. A patient's survival probability is affected by abnormalities in a patient's heart, liver, kidney, or pancreas, i.e., the more organs with abnormalities, the higher the mortality for patients. Expected utilities and medical costs are based on the number of organ abnormalities. Each time point, health states are defined by the values of a set of attributes such as abnormalities of the liver, heart, kidney, and pancreas, retinopathy, neuropathy, amputation, impaired physical appearance, hyperphagia, and female reproductive dysfunction.

Health utility estimates were derived from a discrete choice experiment (DCE) within the general population. These estimates were used to estimate QALYs associated with lipodystrophy.

Metreleptin is available in 11.3 mg vials (10 mg dose). However, the availability of smaller vial sizes (5.8 mg and 3 mg) is expected within the next three months. Given the anticipated availability of smaller vials, an average per patient price of metreleptin was assumed in the base case analysis. Resource use was based on resource use questionnaires completed by two clinical advisers who treat lipodystrophy at Addenbrooke's Hospital. Health-state costs were based on NHS reference costs. Only the cost of hypoglycaemic events was included in the model as adverse event.

Several assumptions were assessed in the sensitivity analysis, i.e., a price fall of 90% of metreleptin after 10 years, reduced initial price, elimination of mortality benefit of metreleptin for PL patients, changes to assumptions regarding organ abnormality progression, alternate survival extrapolation methods, and earlier treatment initiation. A deterministic one-way sensitivity analysis was conducted for the key clinical and economic variables in the model. A probabilistic sensitivity analysis (PSA) was also conducted.

When only 11.3 mg vials were included in the cost effectiveness analysis, the incremental costs per QALY gained were £1,316,932 for metreleptin compared to SoC. The additional costs were £671,927 per QALY gained for metreleptin compared to SoC when multiple vial sizes of metreleptin are available. When a PAS was applied to the scenarios of only 11.3 mg vials available and multiple vial sizes available, ICER yielded [redacted] and [redacted] per QALY gained respectively for metreleptin versus SoC.

The ERG identified several critical issues with the company's economic analysis. One of the most important concerns related to the organ impairment progression, which led to bias in favour of metreleptin treatment compared to SoC. The ERG requested the company to conduct *de novo* statistical analyses. However, the company could not finalise this request given the timelines. The ERG also had serious concerns surrounding the survival analysis conducted by the company and the implementation of theses analyses in the model. There are also several issues identified by the ERG related to the extrapolation of other attributes not related to organ damage and metreleptin discontinuation, which created bias.

Furthermore, the ERG considers the disutility weights presented by the company as speculative. The key concern is that the use of DCE to directly obtain disutility values for heath states is still in its infancy. The most striking unresolved methodological issue relates to the fact that DCE classifies health states far more often below zero than TTO and produces lower average health state values. This was indeed observed in the results of the current DCE study. In addition, various major flaws in the design of the DCE and the analysis of the data were identified, leading to a negative assessment of the way QALYs are currently estimated.

The ERG also had several concerns about the resource use and costs included in the model. Furthermore, the ERG considered the validation of the model as insufficient.

Given the many critical issues described above, it proved impossible for the ERG to give any indication on the cost-effectiveness of metreleptin, and the uncertainty around the ICERs presented by the company goes far beyond that created by parameter uncertainty and reported in the CS.

The CS includes a budget impact model to estimate the total costs to the NHS for a period of five years of adopting metreleptin for LD patients in the UK. The budget impact analysis results presented by the company suggest that the net budget impact of implementing metreleptin will be £18,762,893 in year 1 and will rise to £34,114,350 in five years. The cumulative net budget impact over the first five years will be £133,045,965. Additionally, the estimated total number of LD patient eligible for metreleptin treatment after five years is 44 and the uptake of metreleptin rises from 85% in year 1 to 90% in year 5.

The CS also includes estimates of the impact of metreleptin on (i) inability to work or attend school for patients and carers; (ii) estimates of out-of-pocket costs for patients and carers including costs related to diabetes, transportation, fertility and cosmetic treatment; and (iii) other carer costs.

In general, the assumptions made in the budget impact analysis could be considered as plausible. However, there are some concerns about the expected uptake rate of metreleptin. The ERG considers the high expected uptake rate as reliable, but the reason behind the rising uptake rate from 85% in year 1 to 90% in year 5 is unclear since the company did not provide further details on these assumptions. Furthermore, the validity of the estimated discontinuation rate considered by the company remains unclear since detailed information on these assumptions were also not provided by the company.

The ERG has some concerns related to the impact of metreleptin beyond direct health benefits. No costs associated with inability to work or attending school were calculated in the analyses. However, as part of the NIH follow-up study, data on these attributes were collected. The ERG does not see that, while there were data collected on these attributes, it was not possible to estimate associated costs. The ERG also has a problem with the assertion in which the company stated that metreleptin will mitigate the costs of hospitalisation and fertility and cosmetic treatment, since this is not based on any evidence. No indirect health care costs, due to additional life-years after receiving metreleptin, were reported in the CS and the company expected that these costs would not influence the cost effectiveness results. In the opinion of the ERG, these costs should be included in the model for completeness. Finally, the CS does not include costs related to informal care and productivity loss for the caregiver. The company states that it is currently conducting research to gain more details of these issues, but the ERG considers it as inadequate that the impact of lipodystrophy on informal carers was not identified prior to the CS.

9.3 Strengths and limitations

9.3.1 Strengths of the CS

The ERG believes that the following represent strengths within the CS:

- The company's submission provided sufficient details for the ERG to appraise the searches, which were on the whole clear, transparent and reproducible. An adequate range of resources were searched.
- Despite the rarity of LD syndromes, the company has presented data from a large, multinational study of metreleptin treated patients.
- The ERG considers that the budget impact model is generally based on plausible assumptions.

9.3.2 Weaknesses of the CS

The following are the main weaknesses of the CS, observed by the ERG:

- The CS lacks information about the long-term effects of metreleptin treatment.
- The CS (section 9.9.2, page 121) states that: 'Over 85% of the 107 patients in study NIH 991265/20010769 received >1 year of metreleptin, 72% received >2 years, 54% received >3 years, and 28% received 6 or more years of metreleptin in this study. The maximum duration of therapy was 14 years.'[1] Despite this, the reporting of long-term clinical effectiveness outcomes, in the CS, was limited to information on the persistence (up to 36 months) of changes in HbA_{1c} and triglycerides on metreleptin treatment.
- Where long-term outcomes were available (in the NIH follow-up study, not included in the CS), these were either inferred from changes in surrogate outcome measures (e.g. hepatic enzymes, 24-hour protein excretion, blood pressure), or lacked any definition (e.g. hyperphagia recorded in notes).
- The CS lacks information about UK lipodystrophy patients; only one patient in the metreleptin treatment studies and one patient in the natural history study that was used in the cost effectiveness analysis, were UK patients.

- Despite the existence of an EAP, which includes UK patients and has been running for more than 10 years, no results from the EAP were included in the CS and no justification/explanation for this was provided.
- The study details and results for the NIH follow-up study and the GL/PL natural history study, which were used to inform cost effectiveness modelling, were not included in the clinical effectiveness section of the CS.
- Participants in the NIH follow-up study and the GL/PL natural history study were not comparable and it is not clear that the matching exercise reported in the CS was adequate to account for the apparent differences.
- The clinical effectiveness section of the CS does not include any assessment of the comparative effectiveness of metreleptin vs. standard care (either direct or indirect).
- The process used to identify and select comparator/natural history studies remains unclear; the company's response to clarification questions stated that: 'The clinical SLR was carried out to search for trials of both metreleptin and trials of relevant comparators (see Section 9.1 of the submission).'[39] However, the searches reported in the relevant sections of the CS were specific to metreleptin/leptin replacement interventions and did not include any terms to search for comparator studies; these searches would not have reliably retrieved studies of comparator interventions or natural history studies.
- There are several concerns related to the estimation of organ impairment progression. Due to these issues, the ERG has substantial concerns about the appropriateness of the statistical methods used by the company. Therefore, the ERG requested the company to conduct de novo statistical analysis, however, the company stated that they were not able to finalise this request due to the given timelines.
- Serious concerns regarding the survival analyses conducted by the company and the implementation of these analyses in the model were identified.
- There were also several issues related to the matching methodology conducted by the company.
- The ERG considers the derivation of the utility decrement from the company's DCE as invalid.
- The validation of the model is considered as inadequate and vague by the ERG.

9.4 Uncertainties

There is considerable uncertainty about the long-term effects of metreleptin treatment, particularly in relation to patient-perceived symptoms and clinical outcomes. The clinical effectiveness section of the CS includes only very limited evidence about patient perceived symptoms (hyperphagia) and clinical outcomes (liver damage) and data are limited to one year. The 'post-metreleptin improvements' reported in the NIH follow-up study,[46] but not in the CS, are frequently based on measures taken at one year and use definitions based on changes in surrogate outcome measures (e.g. improvement in liver abnormality is defined as 20% reduction in ALT/AST at year one in a patient who had elevated ALT/AST at baseline) or provide no definition at all. The NIH follow-up study[46] also included some information on newly emergent (on metreleptin treatment) lipodystrophy characteristics in patients with no evidence of these characteristics prior to metreleptin initiation. However, no indication of the timeframe

of observation was provided. Broadly, these data indicate that new incidences of organ abnormalities (liver, kidney and heart) and female reproductive dysfunction continue to occur, in all categories of LD patient, on metreleptin treatment. The data presented are insufficient to allow an adequate assessment of how the rate of development of new abnormalities on metreleptin treatment would compare with that seen in patients on standard care.

There remains some uncertainty regarding the long-term effects of metreleptin on metabolic measures. The CS includes some information on the persistence (up to 36 months) of changes in HbA_{1c} and triglycerides on metreleptin treatment (see Section 4.2.4). These data indicate that the apparent effect of metreleptin on triglyceride levels may not be applicable to the overall PL population. The potential effects of neutralising antibodies on the long-term efficacy of metreleptin treatment remain unclear. In clinical trials (studies NIH 991265/20010769 and FHA101), most patients (95%) developed antibodies to metreleptin.[33] Overall, in patients where antibody data was available, neutralising anti-drug antibody activity was observed in 38/102 patients (37%) and, of these 38 patients, 58% achieved resolution of neutralising antibodies.[33] Seven patients from the NIH and FHA101 studies developed high potency neutralizing activity to metreleptin.[75] One of these patients had loss of efficacy, as indicated by an increase in HbA_{1c} concentrations, and five hospitalisations due to bacterial infections.[75] A second patient, also with a history of hospitalisation for sepsis and worsening glycaemic control, was recently reported to have developed neutralising activity.[75] These cases raise concern that development of neutralising antibodies to metreleptin could impair metabolic control and immune function.[75]

The observed effects of metreleptin are all based on changes from baseline in single arm metreleptin treatment studies. The lack of comparative studies means that the extent to which any observed effects may be attributed to metreleptin remains unclear. This problem is compounded as the CS does not include any attempt to draw indirect comparisons through studies of the effects of established clinical management (diet, lifestyle modifications, lipid lowering drugs and anti-diabetic medications). The natural history study, used to provide comparator data for the cost effectiveness analysis, is not used in the clinical effectiveness sections of the CS and has a population which is not comparable to those included in the metreleptin intervention studies. It is therefore not possible to assess the extent to which any apparent treatment effects are attributable to metreleptin, or whether similar effects could be achieved using standard care.

The significance of pancreatitis, as an adverse event following withdrawal from treatment, remains unclear. The CS (section 9.7.2.5, pg 114) describes incidences of pancreatitis as an adverse event, following withdrawal from treatment: ‘Across the 148 patients included in LD studies, six (4%) patients (four with GL and two with PL), experienced treatment-emergent pancreatitis. All patients had a history of pancreatitis and hypertriglyceridemia. One of the patients who developed septic shock concurrent with pancreatitis died; the other five patients recovered and continued on treatment. Abrupt interruption and/or non-compliance with metreleptin dosing was suspected to have contributed to the occurrence of pancreatitis in several of these patients. The mechanism for pancreatitis in these patients was presumed to be return of hypertriglyceridemia and therefore increased risk of pancreatitis in the setting of

discontinuation of effective therapy for hypertriglyceridemia.'[1] Non-compliance rates of between 9% and 19% were reported,[1] and the extent of the pancreatitis risk, for these patients, remains unclear. The CS (section 9.9.1.1, page 120-121) states that: 'The identified risks of hypersensitivity, acute pancreatitis associated with metreleptin discontinuation, and hypoglycaemia with concomitant use of insulin and insulin secretagogues can be managed with risk communication in labelling and educational activities.'[1] However, no evidence is presented in support of this assertion. Similarly, the results for the NIH 991265/20010 study,[37] described in the CS, note the exclusion of an 'outlier' patient in whom an increase from baseline in triglycerides of >1000% at Month 12/LOCF was observed. This increase was attributed to non-compliance; the extent to which such large increases in triglycerides may be seen in patients who withdraw abruptly from metreleptin is unclear, and similarly the persistence and long-term consequences of any such increases is unknown.

There is no mention in the CS of possible stopping rules for metreleptin. The CS (Table A2, page 24-25) appears to assume that treatment will be ongoing for the full lifetime of the patient. However, given the many differences between and within groups of patients with different LD syndromes, it cannot be expected that the treatment works equally well or even at all in all patients and the effectiveness of the treatment might diminish over time. Therefore, stopping rules should be considered.

Currently, only 11.3 mg vials of metreleptin are available. However, the company expects the availability of smaller vial sizes (i.e., 5.8 mg and 3.0 mg) within three months after submission. This will impact the ICER significantly.

The ERG does not consider the cost-effectiveness model as reliable and trustworthy enough to inform decision making on the cost-effectiveness of metreleptin. The uncertainty around the company-reported ICERs is much larger than suggested by the PSA, which only addresses parameter uncertainty. However, the ERG still expects decision uncertainty to be rather low, as the ICER values, even in the best cases that the company presented, are significantly above the accepted thresholds.

10. REFERENCES

[1] Aegerion Pharmaceuticals Ltd. *Metreleptin for treating lipodystrophy [ID861]: submission to National Institute of Health and Care Excellence. Highly Specialised Technologies Programme (HST)*: Aegerion Pharmaceuticals Ltd., 2018. 280p.

[2] Brown RJ, Araujo-Vilar D, Cheung PT, Dunger D, Garg A, Jack M, et al. The diagnosis and management of lipodystrophy syndromes: a multi-society practice guideline. *J Clin Endocrinol Metab* 2016;101(12):4500-4511.

[3] Garg A. Clinical review: lipodystrophies: genetic and acquired body fat disorders. *J Clin Endocrinol Metab* 2011;96(11):3313-25.

[4] Fiorenza CG, Chou SH, Mantzoros CS. Lipodystrophy: pathophysiology and advances in treatment. *Nat Rev Endocrinol* 2011;7(3):137-50.

[5] Gupta N, Asi N, Farah W, Almasri J, Barrionuevo P, Alsawas M, et al. Clinical features and management of non-HIV-related lipodystrophy in children: a systematic review. *J Clin Endocrinol Metab* 2017;102(2):363-374.

[6] Misra A, Garg A. Clinical features and metabolic derangements in acquired generalized lipodystrophy: case reports and review of the literature. *Medicine (Baltim)* 2003;82(2):129-46.

[7] Pope E, Janson A, Khambalia A, Feldman B. Childhood acquired lipodystrophy: a retrospective study. *J Am Acad Dermatol* 2006;55(6):947-50.

[8] Savage DB, Soos MA, Powlson A, O'Rahilly S, McFarlane I, Halsall DJ, et al. Familial partial lipodystrophy associated with compound heterozygosity for novel mutations in the LMNA gene. *Diabetologia* 2004;47(4):753-6.

[9] Garg A. Acquired and inherited lipodystrophies. *N Engl J Med* 2004;350(12):1220-34.

[10] Vantyghem MC, Vincent-Desplanques D, Defrance-Faivre F, Capeau J, Fermon C, Valat AS, et al. Fertility and obstetrical complications in women with LMNA-related familial partial lipodystrophy. *J Clin Endocrinol Metab* 2008;93(6):2223-9.

[11] Hegele RA. Premature atherosclerosis associated with monogenic insulin resistance. *Circulation* 2001;103(18):2225-9.

[12] Misra A, Peethambaram A, Garg A. Clinical features and metabolic and autoimmune derangements in acquired partial lipodystrophy: report of 35 cases and review of the literature. *Medicine (Baltim)* 2004;83(1):18-34.

[13] Chiquette E, Oral EA, Garg A, Araujo-Vilar D, Dhankhar P. Estimating the prevalence of generalized and partial lipodystrophy: findings and challenges. *Diabetes Metab Syndr Obes* 2017;10:375-383.

[14] Haque WA, Shimomura I, Matsuzawa Y, Garg A. Serum adiponectin and leptin levels in patients with lipodystrophies. *J Clin Endocrinol Metab* 2002;87(5):2395.

[15] Cowley MA, Smart JL, Rubinstein M, Cerdan MG, Diano S, Horvath TL, et al. Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* 2001;411(6836):480-4.

[16] Kim JK, Gavrilova O, Chen Y, Reitman ML, Shulman GI. Mechanism of insulin resistance in A-ZIP/F-1 fatless mice. *J Biol Chem* 2000;275(12):8456-8460.

[17] Ceddia RB. Direct metabolic regulation in skeletal muscle and fat tissue by leptin: implications for glucose and fatty acids homeostasis. *Int J Obes (Lond)* 2005;29(10):1175-83.

[18] Conrad Stöppler M. Triglycerides. 2016 [accessed 26.1.18]. Available from: http://www.emedicinehealth.com/triglycerides/page2_em.htm#normal_triglyceride_values

[19] Diker-Cohen T, Cochran E, Gorden P, Brown RJ. Partial and generalized lipodystrophy: comparison of baseline characteristics and response to metreleptin. *J Clin Endocrinol Metab* 2015;100(5):1802-10.

[20] Garg A, Misra A. Lipodystrophies: rare disorders causing metabolic syndrome. *Endocrinol Metab Clin North Am* 2004;33(2):305-31.

[21] Nordestgaard BG, Benn M, Schnohr P, Tybjaerg-Hansen A. Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA* 2007;298(3):299-308.

[22] Garg A. Gender differences in the prevalence of metabolic complications in familial partial lipodystrophy (Dunnigan variety). *J Clin Endocrinol Metab* 2000;85(5):1776-82.

[23] Lupsa BC, Sachdev V, Lungu AO, Rosing DR, Gorden P. Cardiomyopathy in congenital and acquired generalized lipodystrophy: a clinical assessment. *Medicine (Baltim)* 2010;89(4):245-50.

[24] Safar Zadeh E, Lungu AO, Cochran EK, Brown RJ, Ghany MG, Heller T, et al. The liver diseases of lipodystrophy: the long-term effect of leptin treatment. *J Hepatol* 2013;59(1):131-7.

[25] Stears A, Hames C. Diagnosis and management of lipodystrophy: a practical update. *Clinical Lipidology* 2014;9(2):235-59.

[26] McDuffie JR, Riggs PA, Calis KA, Freedman RJ, Oral EA, DePaoli AM, et al. Effects of exogenous leptin on satiety and satiation in patients with lipodystrophy and leptin insufficiency. *J Clin Endocrinol Metab* 2004;89(9):4258-63.

[27] National Institute for Health and Care Excellence. *Highly specialised Technologies Evaluation. Metreleptin for treating lipodystrophy [ID:861]: final scope*. London: NICE, 2017. 4p.

[28] Oral EA, Ruiz E, Andewelt A, Sebring N, Wagner AJ, Depaoli AM, et al. Effect of leptin replacement on pituitary hormone regulation in patients with severe lipodystrophy. *J Clin Endocrinol Metab* 2002;87(7):3110-7.

[29] Ozata M, Ozdemir IC, Licinio J. Human leptin deficiency caused by a missense mutation: multiple endocrine defects, decreased sympathetic tone, and immune system dysfunction indicate new targets for leptin action, greater central than peripheral resistance to the effects of leptin, and spontaneous correction of leptin-mediated defects. *J Clin Endocrinol Metab* 1999;84(10):3686-95.

[30] Handelsman Y, Oral EA, Bloomgarden ZT, Brown RJ, Chan JL, Einhorn D, et al. The clinical approach to the detection of lipodystrophy - an AACE consensus statement. *Endocr Pract* 2013;19(1):107-16.

[31] Dhankhar P, Isupov T, Araujo-Vilar D, Brown R, Garg A, Jae DH, et al. Estimating quality of life of patients with lipodystrophy. *Value Health* 2015;18(3).

[32] NHS England. *NHS standard contract for insulin-resistant diabetes services (all ages). Service specification A03/S(HSS)/b [Internet]*, 2013 [accessed 26.1.18]. 10p. Available from: https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2016/01/a03-insulin-resist-diabetes.pdf

[33] Aegerion Pharmaceuticals Ltd. *Myalepta. Summary of product characteristics (DRAFT). Data on file*. 44p.

[34] NHS England. *Manual for Prescribed Specialised Services 2017/18. Chapter 62: highly specialist metabolic disorder services (adults and children), 2016 [Internet]*, 2017 [accessed 4.4.18]. 382p. Available from: https://www.england.nhs.uk/wp-content/uploads/2017/10/prescribed-specialised-services-manual-2.pdf

[35] Department of Health. *The national service framework for long-term conditions [Internet]*. Leeds, 2005 [accessed 4.4.18]. 106p. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198114/National_Service_Framework_for_Long_Term_Conditions.pdf

[36] Department of Health. *NHS Outcomes Framework: at-a-glance [Internet]*, 2016 [accessed 4.4.18]. 5p. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/513157/NHSOF_at_a_glance.pdf

[37] Aegerion Pharmaceuticals Ltd. *Clinical study report: protocol 991265 and 20010769. Metreleptin long term efficacy and safety of leptin replacement in the treatment of patients with lipodystrophy*, 2016. 252p.

[38] Aegerion Pharmaceuticals Ltd. *Clinical study report: protocol FHA101 (MB002-002). Metreleptin an open-label treatment protocol to provide metreleptin for the treatment of diabetes mellitus and/or hypertriglyceridaemia associated with lipodystrophy. Data on file*, 2016. 160p.

[39] Aegerion Pharmaceuticals Ltd. *Metreleptin for treating lipodystrophy [ID861]: response to request for clarification from the ERG*: Aegerion Pharmaceuticals Ltd., 2018. 98p.

[40] Tuttle E. *Burden of generalized and partial lipodystrophy among metreleptin-naive patients: a natural history study*. Menlo Park, CA: Analysis Group Inc., 2018. 48p.

[41] Tolley Health Economics. *Protocol for the systematic literature review (SLR) on the clinical evidence (efficacy and safety) associated with metreleptin and any comparator treatments for lipodystrophy (Clinical protocol A)*, 2017

[42] 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

[43] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

[44] Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. [accessed 27.2.18]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

[45] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;313(7052):275-83.

[46] Tuttle E. *Metreleptin for lipodystrophy: NIH follow-up study*. Menlo Park, CA: Analysis Group Inc., 2018. 15p.

[47] National Institute of Diabetes Digestive Kidney Diseases, National Institutes of Health Clinical Center. Leptin to treat lipodystrophy. 2015. Available from: https://ClinicalTrials.gov/show/NCT00025883

[48] Beltrand J, Beregszaszi M, Chevenne D, Sebag G, De Kerdanet M, Huet F, et al. Metabolic correction induced by leptin replacement treatment in young children with Berardinelli-Seip congenital lipoatrophy. *Pediatrics* 2007;120(2):e291-6.

[49] Beltrand J, Lahlou N, Le Charpentier T, Sebag G, Leka S, Polak M, et al. Resistance to leptin-replacement therapy in Berardinelli-Seip congenital lipodystrophy: an immunological origin. *Eur J Endocrinol* 2010;162(6):1083-91.

[50] Simha V, Subramanyam L, Szczepaniak L, Quittner C, Adams-Huet B, Snell P, et al. Comparison of efficacy and safety of leptin replacement therapy in moderately and severely hypoleptinemic patients with familial partial lipodystrophy of the Dunnigan variety. *J Clin Endocrinol Metab* 2012;97(3):785-92.

[51] Asthana A, Abel BS, Skarulis MC, Gorden P, Brown RJ, Muniyappa R. Effects of recombinant human leptin (Metreleptin) therapy on plasma angiopoietin-like proteins 3 and 4 in lipodystrophy patients. In: *Late-breaking Cardiovascular Endocrinology II*: Endocrine Society, 2015: LBF-091-LBF-091.

[52] Brown RJ, Muniyappa R, Walter PJ, Gharib AM, Ouwerkerk R, Mattingly M, et al. Effects of leptin on glucose and lipid metabolism during constant food intake. In: *GI Regulatory Peptides; Glucose Metabolism*: Endocrine Society, 2015: SAT-629-SAT-629.

[53] Ebihara K, Kusakabe T, Hirata M, Masuzaki H, Miyanaga F, Kobayashi N, et al. Efficacy and safety of leptin-replacement therapy and possible mechanisms of leptin actions in patients with generalized lipodystrophy. *J Clin Endocrinol Metab* 2007;92(2):532-41.

[54] Schlogl H, Muller K, Horstmann A, Miehle K, Puschel J, Villringer A, et al. Leptin substitution in patients with lipodystrophy: neural correlates for long-term success in the normalization of eating behavior. *Diabetes* 2016;65(8):2179-86.

[55] Vatier C, Fetita S, Boudou P, Tchankou C, Deville L, Riveline J, et al. One-year metreleptin improves insulin secretion in patients with diabetes linked to genetic lipodystrophic syndromes. *Diabetes Obes Metab* 2016;18(7):693-7.

[56] Araujo-Vilar D, Sanchez-Iglesias S, Guillin-Amarelle C, Castro A, Lage M, Pazos M, et al. Recombinant human leptin treatment in genetic lipodystrophic syndromes: the long-term Spanish experience. *Endocrine* 2015;49(1):139-47.

[57] Oral EA, Javor ED, Ding L, Uzel G, Cochran EK, Young JR, et al. Leptin replacement therapy modulates circulating lymphocyte subsets and cytokine responsiveness in severe lipodystrophy. *J Clin Endocrinol Metab* 2006;91(2):621-8.

[58] Oral EA, Simha V, Ruiz E, Andewelt A, Premkumar A, Snell P, et al. Leptin-replacement therapy for lipodystrophy. *N Engl J Med* 2002;346(8):570-8.

[59] Petersen KF, Oral EA, Dufour S, Befroy D, Ariyan C, Yu C, et al. Leptin reverses insulin resistance and hepatic steatosis in patients with severe lipodystrophy. *J Clin Invest* 2002;109(10):1345-1350.

[60] Javor ED, Cochran EK, Musso C, Young JR, Depaoli AM, Gorden P. Long-term efficacy of leptin replacement in patients with generalized lipodystrophy. *Diabetes* 2005;54(7):1994-2002.

[61] Musso C, Cochran E, Javor E, Young J, Depaoli AM, Gorden P. The long-term effect of recombinant methionyl human leptin therapy on hyperandrogenism and menstrual function in female and pituitary function in male and female hypoleptinemic lipodystrophic patients. *Metabolism* 2005;54(2):255-63.

[62] Park JY, Javor ED, Cochran EK, DePaoli AM, Gorden P. Long-term efficacy of leptin replacement in patients with Dunnigan-type familial partial lipodystrophy. *Metabolism* 2007;56(4):508-16.

[63] Cambridge University Hospitals. National Severe Insulin Resistance Service. Information for referring clinicians [Internet]. [accessed 26.1.18]. Available from: https://www.cuh.nhs.uk/national-severe-insulin-resistance-service/information-for-referring-clinicians

[64] Christensen JD, Lungu AO, Cochran E, Collins MT, Gafni RI, Reynolds JC, et al. Bone mineral content in patients with congenital generalized lipodystrophy is unaffected by metreleptin replacement therapy. *J Clin Endocrinol Metab* 2014;99(8):E1493-500.

[65] Chong AY, Lupsa BC, Cochran EK, Gorden P. Efficacy of leptin therapy in the different forms of human lipodystrophy. *Diabetologia* 2010;53(1):27-35.

[66] Brown RJ, Lutz K, Cochran E, Pratt J, Miller S, Blase E, et al. Metreleptin treatment for metabolic abnormalities associated with lipodystrophy: achieving A1C and triglyceride targets. In: *Diabetes: New Diagnostic & Treatment Modalities*: Endocrine Society, 2013: OR17-1-OR17-1.

[67] Muniyappa R, Brown RJ, Mari A, Joseph J, Warren MA, Cochran EK, et al. Effects of leptin replacement therapy on pancreatic beta-cell function in patients with lipodystrophy. *Diabetes Care* 2014;37(4):1101-7.

[68] Moran SA, Patten N, Young JR, Cochran E, Sebring N, Reynolds J, et al. Changes in body composition in patients with severe lipodystrophy after leptin replacement therapy. *Metabolism* 2004;53(4):513-519.

[69] Javor ED, Ghany MG, Cochran EK, Oral EA, DePaoli AM, Premkumar A, et al. Leptin reverses nonalcoholic steatohepatitis in patients with severe lipodystrophy. *Hepatology* 2005;41(4):753-60.

[70] Ajluni N, Dar M, Xu J, Neidert AH, Oral EA. Efficacy and safety of metreleptin in patients with partial lipodystrophy: lessons from an expanded access program. *J Diabetes Metab* 2016;7(3).

[71] Rodriguez AJ, Neeman T, Giles AG, Mastronardi CA, Paz Filho G. Leptin replacement therapy for the treatment of non-HAART associated lipodystrophy syndromes: a meta-analysis into the effects of leptin on metabolic and hepatic endpoints. *Arq Bras Endocrinol Metabol* 2014;58(8):783-97.

[72] Paz-Filho GJ, Neeman T, Mastronardi CA, Rodriguez AJ. A meta-analysis into the effects of leptin replacement therapy on metabolic and hepatic endpoints of patients with lipodystrophy syndromes not associated with the use of HIV protease inhibitors. Presented at Endocrine Society's 96th Annual Meeting and Expo; 21-24 June 2-14; Chicago: USA [Internet]. 2014 [accessed 7.3.18]. Available from: http://press.endocrine.org/doi/abs/10.1210/endo-meetings.2014.OABA.2.SUN-0877

[73] Dantas de Medeiros Rocha E, Dantas Leite L, de Fatima Paiva Baracho M, Goretti do Nascimento Santos M, de Araujo DM, Cavalcante Franca M, et al. Effect of diet intervention and oral zinc supplementation on metabolic control in Berardinelli-Seip syndrome. *Annals of Nutrition & Metabolism* 2010;57(1):9-17.

[74] Aegerion Pharmaceuticals Ltd. *Clinical study report: protocol 991265 and 20010769. Metreleptin long term efficacy and safety of leptin replacement in the treatment of patients with lipodystrophy. Full Report and appendices*, 2016. 5939p.

[75] FDA Center for Drug Evaluation and Research. *Summary Review for Regulatory Action: Myalept/Metreleptin [Internet]*, 2014 [accessed 28.3.18]. 22p. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125390Orig1s000SumR.pdf

[76] Aegerion Pharmaceuticals Ltd. *Clinical study report: protocol FHA101 (MB002-002). Metreleptin an open-label treatment protocol to provide metreleptin for the treatment of diabetes mellitus and/or hypertriglyceridaemia associated with lipodystrophy. Full study report and appendices*, 2016. 2505p.

[77] Rodriguez AJ, Mastronardi CA, Paz-Filho GJ. New advances in the treatment of generalized lipodystrophy: role of metreleptin. *Ther Clin Risk Manag* 2015;11:1391-400.

[78] Mancuso P, Gottschalk A, Phare SM, Peters-Golden M, Lukacs NW, Huffnagle GB. Leptin-deficient mice exhibit impaired host defense in gram-negative pneumonia. *The Journal of Immunology* 2002;168(8):4018-4024.

[79] Moon HS, Dalamaga M, Kim SY, Polyzos SA, Hamnvik OP, Magkos F, et al. Leptin's role in lipodystrophic and nonlipodystrophic insulin-resistant and diabetic individuals. *Endocr Rev* 2013;34(3):377-412.

[80] Farooqi IS, Matarese G, Lord GM, Keogh JM, Lawrence E, Agwu C, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest* 2002;110(8):1093-103.

[81] Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford University Press, 2005.

[82] Bristol-Myers Squibb. *Myalept. Prescribing information [Internet]*, accessed 26.1.18. 55p. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/125390s000lbl.pdf

[83] Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31(1):306-18.

[84] Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value eq-5d health states. *Health Econ* 2014;23(6):729-42.

[85] Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15(12):1295-310.

[86] Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. *Health Econ* 2002;11(1):33-42.

[87] Jackson C. *Multi-state modelling with R: the msm package [Internet]*, 2018 [accessed 28.3.18]. 57p. Available from: https://cran.r-project.org/web/packages/msm/vignettes/msm-manual.pdf

[88] Latimer NR. Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013;33(6):743-54.

[89] Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and markov decision-analytic modeling. *Med Decis Making* 2017;37(4):427-439.

[90] Faria R, Hernandez Alava M, Manca A, Wailoo AJ. *NICE DSU Technical Support Document 17: the use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: methods for comparative individual patient data*, 2015. 86p. Available from: http://www.nicedsu.org.uk

[91] 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.

[92] Robinson A, Spencer AE, Pinto-Prades JL, Covey JA. Exploring differences between TTO and DCE in the valuation of health states. *Med Decis Making* 2017;37(3):273-284.

[93] Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. *PharmacoEcon* 2016;34(4):349-61.

[94] Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med* 2012;29(7):855-62.

[95] Holmes J, Gear E, Bottomley J, Gillam S, Murphy M, Williams R. Do people with type 2 diabetes and their carers lose income? (T2ARDIS-4). *Health Policy* 2003;64(3):291-296.

[96] Aegerion Pharmaceuticals Ltd. *Lipodystrophy patient research (NIH). Data on file*, 2017

[97] EURORDIS. *Rare diseases Europe. Juggling care and daily life. The balancing act of the rare disease community. A Rare Barometer survey*. Paris: EURORDIS (Paris Office), 2017. 40p. Available from: https://www.eurordis.org/news/3000-rare-disease-patients-carers-voice-difficulties-balancing-care-life

[98] Rare Disease UK. *The Rare Reality – an insight into the patient and family experience of rare disease* London: Rare Disease UK, 2016. 44p. Available from: https://www.raredisease.org.uk/media/1588/the-rare-reality-an-insight-into-the-patient-and-family-experience-of-rare-disease.pdf

Appendix 1: ERG Search Strategies

ERG Epidemiology/Natural History Test Search

The following search was run to investigate additional condition terms identified by the ERG and to identify the number of records retrieved combining these with epidemiology/natural history terms. The ERG feels the number retrieved was a manageable number for the company to screen as part of their SLR to identify potential epidemiological and natural history studies.

Embase (OVIDSP): 1974 to 2018 March 07
Searched: 8.3.18

1 exp lipodystrophy/ (10776)
2 (lipodystrop$ or lipid dystroph$ or lipoatroph$ or FPLD or CGL2 or (Dunnigan adj syndrome$) or (lawrence adj syndrome$) or (Berardinelli$ adj syndrome$) or (wiedemann adj rautenstrauch) or (donohue adj syndrome$) or kobberling or koebberling).ti,ab,ot. (7234)
3 1 or 2 (13064)
4 incidence/ (299938)
5 standardized incidence ratio/ (2223)
6 Prevalence/ (570695)
7 standardized mortality ratio/ (2172)
8 demography/ (183246)
9 epidemiological data/ (29634)
10 mortality/ (689114)
11 disease progression/ (254412)
12 disease activity/ (69311)
13 morbidity/ (299793)
14 (occurrence$ or incidence$ or prevalence$ or episode$ or mortalit$ or morbidit$ or epidemiolog$ or demograph$ or (natural adj2 history) or (disease adj2 progres$) or (disease adj2 course)).ti,ab,ot. (3633979)
15 or/4-14 (4293380)
16 3 and 15 (2733)
17 limit 16 to yr="2008 -Current" (1540)