Ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene

ERRATUM

Replacement pages following the factual accuracy check

10th September 2015

Produced by ERG: Warwick Evidence

Similar rates of adverse events were experienced by patients receiving ataluren and placebo. No deaths were reported from either study. A cumulative summary of serious adverse events from four ongoing and five completed company-sponsored clinical trials appeared to suggest that serious cardiac disorders, infections and infestations, injury poisoning and procedural complications and total number of serious adverse events are more common with ataluren than placebo, however it is not clear from the information provided whether this is due to longer exposure in the ataluren group.

Outcomes from the six patient submissions and the patient organisations Muscular Dystrophy UK and Action Duchenne were highly positive in nature and no known disadvantages to the treatment were reported. However, a reverse of benefits after stopping treatment was observed in one case. Key themes identified by the ERG included the emotional and social impacts of DMD, the anticipated effects of treatment, and the importance to carers of self-reliance and reduced burden No details on how generalisable these views are to the wider UK nmDMD community were reported.

Summary of evidence submitted on value for money

The Company's submission included a decision analytical semi-Markov model to compare the costs and benefits of ataluren with best supportive care versus best supportive care for people with nonsense mutation Duchenne Muscular Dystrophy. The model starts with a hypothetical cohort of children age 8.5 years and weighing approximately 25kg and simulates the clinical pathway for people with nmDMD. In each three-monthly cycle people incur costs and benefits depending on their health state and the cost consequences are assessed. The model time horizon was set at the time at which the last individual leaves the ambulant health state. The discount rate was 3.5% per annum. Results are presented in terms of mean costs and mean benefits, measured in QALYs. Information required to populate the model was obtained from various sources, with data on the treatment benefit of ataluren versus best supportive care mainly drawn from Study 007. One-way sensitivity analyses and scenario analyses were undertaken to determine the impact of changes in parameter values and assumptions on the base case results.

The initial model submitted by the Company estimated mean costs for ataluren and best supportive care of £5,092,540 and £235,207, with equivalent mean QALYs of 6.152 and 2.385, giving incremental costs and QALYs of £4,857,333 and 3.767. A revised model was subsequently submitted by the Company, which included improvements in the distributions used to extrapolate data forward over time. This model was found to have an error, but after

- The inclusions of the costs of 6 months of ataluren treatment post loss of ambulation, in line with clinical advice.
- Refitting of survival curves to the various sets of Kaplan-Meier data, using a log-normal distribution for time to loss of ambulation, and flexible parametric distributions for other transitions.

The ERG ran a number of different models, using different assumptions for the distributions used to extrapolate trial results over time. These generated incremental cost estimates ranging from £4,295,464 to £5,544,981 with a range of associated QALY estimates of 1.722-3.924. The ERG's best estimate of cost and QALYs, which uses a log-normal distribution for loss of ambulation, and the statistically best fitting models for all other events, includes treatment with ataluren for 6 months post loss of ambulation and a life time horizon, giving incremental mean costs of £5,544,981 and associated QALYs of 3.049. The ERG undertook additional analyses of budget impact taking account of the expected weight of patients with nmDMD likely to be eligible for ataluren use leading to estimates of an average annual budget impact of £19,069,166, as compared to the £12,223,821 reported in the initial Company submission.

1.7. Effects of technology beyond direct health benefits and on provision of specialised services

The ERG considered that the company presented appropriate wider societal costs and some potential savings for ataluren. However the ERG were concerned about the heavy reliance on the Landfeldt study for this and were concerned that these wider societal costs might be either under- or overestimated. Because of the uncertainty it was not possible to assess quantitatively which, if any, of these costs would be alleviated by the use of ataluren. The likely impact of ataluren on the delivery of the specialised services for DMD and for nmDMD in particular is not yet clear in a number of respects. The most important potential impact is the likely need for clinical input for additional monitoring and decisions on continuation and stopping of treatment.

A key criterion for the appraisal, and for the evaluation undertaken in the RCT and the CS was the definition ambulatory. The NICE scope does not provide a clear definition. The RCT states that for inclusion in the study, patients had to be able to walk ≥75 metres. However, the Company's economic model adopted a definition of loss of ambulation (i.e. inability to walk >0 metres). Inevitably the different definitions may influence the outcomes of the assessment and it remains unclear which definition should be used in clinical practice. This is of importance as the suggested stopping rule for ataluren is based on the

using the technology over and above usual clinical practice. In summary no additional tests are believed to be required to identify patients eligible for treatment with ataluren.

Monitoring of ataluren treated patients is considered in section 8.2.3.

Currently NHS England[32] has a policy statement which suggests that since ataluren is being considered by NICE as a Highly Specialised Technology Evaluation to test the benefits and costs, it will not be commissioned until the outcome is known. NHS England also state that '*Where an individual's clinician believes that there may be exceptional clinical circumstances that might warrant consideration of funding outside of this policy, an application can be made under NHS England's Individual Funding Request (IFR) procedure'.*

2.5. Critique of background information provided in the CS

The ERG consider the background information provided by the Company to be fair, comprehensive and appropriate, and the ERG clinical advisors agree that this is an accurate overview of the condition relevant to the decision problem.

The Company provide a detailed coverage of the underlying nature of DMD, the prevalence as well as the epidemiology of DMD and a concise coverage of the underlying aetiology of DMD.

The information provided directly related to nmDMD was limited and it is unclear to what extent the information on DMD is applicable to nmDMD.

The CS did not discuss diagnosis of DMD in the background but touches on the benefits of early diagnosis to maximise the treatment effect of novel treatments, i.e. ataluren if approved.

The CS provided some relevant information about the impact of the DMD on the carers' QoL. The specific impact on carers' quality of life in nmDMD specifically remains unclear. No QoL data for carers was presented.

A concise overview of the impact of DMD on the health related quality of life (HRQoL) in boys was provided. However, it is unclear whether the impact of DMD on the QoL in girls, which make up a more diverse group with a variable degree of disability, is the same to that reported in boys with this condition and whether this can be extended to patients with nmDMD.

Finally, the Company could have referred to the North Star Clinical Network which was set up in 2003 to help improve services and set national standards of care for children living with DMD.[39] The

supplement appendix of the Bushby paper 2014, [41] states that: *"For these reasons the permutation test provides a more accurate assessment than the pre-specified rank test of the treatment differences in this study*".

f) Post-hoc analysis

Additional analyses were carried out in a sub-population of subjects in the decline phase (>7 years of age, treated with corticosteroids, 6MWD ≥150 m, <80% predicted 6MWD) as this group of patients was believed to be the most likely to display the greatest measureable effect with ataluren treatment. While this analysis was believed to be clinically and scientifically justified according to the CHMP, the EMA also noted that: "…*the patients in the decline phase of their ambulation constituted of a subset of the study 007 population and the analysis should be seen as exploratory*."

g) Adjustment for multiplicity

"*The p-values of the primary and secondary outcome measures were adjusted for comparison of two dose levels against placebo*"[41] (p. 479). The method for adjustment was not reported. Reported nominal p-values were not adjusted for multiplicity. The ERG noted that the reported nominal p-values were generally lower than the adjusted values and that the values for the MMRM analyses were lower than for the observed data. The outcomes table C9.14 on page 90 in the CS does not report any p values for the observed differences, but reports p-values for the MMRM model which for all comparisons except the ITT analysis suggests that the difference was statistically significant. The analysis does not state whether these are nominal or adjusted p-values, but the text on page 94 clarifies that these are nominal p-values. The p-values for the cITT MMRM analysis (the corrected analysis reporting 31.7m (95% CI 5.1-58.3) treatment effect of ataluren in the CS (nominal p=0.0197, adjusted p=0.0367) appear to include the only adjusted p-value reported in the CS. The analysis sources of the p-values are unclear in the CS. Please refer to section 4.2.5 for further detail.

Summary

The statistical methods used in the 007 trial were appropriate, however, a number of post-hoc adjustments as well as post-hoc analyses were undertaken all of which appeared to favour the intervention (ataluren) arm of the trial. Both trial 007 and the CS were transparent about adjustments and justifications; however, the ERG considers that the reporting of outcomes was selective. The ERG would have expected clear reporting of outcomes separately according to pre-specified analyses using rank-transformed data with post-hoc analyses using permutation. The ERG would have also expected reporting of both adjusted and nominal p-values throughout with p-values for differences of observed data in table C9.14 on page 90 of the CS. While the observed difference between ataluren and placebo might be clinically significant, the statistical significance of some reported outcomes should be viewed with extreme caution as this was derived following several post-hoc adjustments. The

adjustments seem to be methodologically appropriate but reporting as sensitivity analyses might have been more appropriate. This should be considered when assessing the evidence of the reported treatment effect in the primary and secondary outcomes in section 4.2.5.

4.2.4. Summary of selected outcomes measures

The NICE scope listed 11 outcome measures to be considered. Some of these outcomes were not adequately measured or reported by the CS (described below). The relevant results are all from the single eligible RCT (trial 007), other than for adverse effects. The CS refers to outcomes of myometry and timed function tests from study 004 but no data are reported.

4.2.4.1. Ambulation

The primary outcome in the CS is 6MWD, a measure of ambulation, which was also the primary outcome in the 007 trial. The CS states on p. 62 and 125 that prior to this trial there were no established primary or secondary endpoints for studies in DMD patients.

The 6MWD test is a measure of exercise tolerance and functional status where the individual is asked to walk on a flat surface for 6 minutes. It is a reliable measure and shows only small variation at individual level over short periods of time. However a recent systematic review looking at nine chronic paediatric conditions, which included three studies in DMD, found evidence that the measurement properties of the 6MWD test varied between studies.[50] The authors concluded that caution is recommended in the interpretation of changes in 6MWD in children with chronic conditions. The CS states on p.125 that a 30 metre change in 6MWD versus placebo is in the range in which other drugs have been approved in multiple inherited conditions. The 6MWD test is known to be at risk of inter-operator bias through encouragement,[51] however the Company confirmed that the clinical evaluator was blinded to allocation. In addition, de Groot et al (2011)[52] discuss potential variations that can occur in the administration of the 6MWD test, for example differences in the distance between turning points, the choice of circuit layout (e.g. circle, squares or use of a treadmill), and instructions given. They note that guidelines for the standardised administration of the test are available. Standardisation between different centres is therefore important. In response to a clarification question the Company provided details of the standardisation of the 6MWD test across study centres, which appear appropriate.

The CS also reported the proportion of patients who experienced at least 10% worsening in 6MWD compared with baseline. The rational for the 10% cut-off was not provided.

[redacted] This indicates selective reporting of results.

The results of the 6MWD test from trial 007 were used as for the measure of time to loss of ambulation in the CS economic evaluation.

4.2.4.2. Muscle function

Muscle function was measured by four timed function tests, stand from supine, 4-stair ascend, 4-stair descent, 10 metre run/walk. The CS states that timed function tests are established clinical assessments in DMD. The CS does not report details of these tests or how these were standardised between centres. However the ERG consider that standardised administration of the test between different centres is an important consideration. The ERG is not aware of any evidence for the validity of these tests as measures of muscle function. Minimal clinically important differences (MCIDs) have been published for these outcomes, based on trial 007.[53] In response to a clarification question the Company confirmed that a clinical evaluator training group developed standardised procedures for timed function tests and training and a manual were provided to all study sites, including refresher training after approximately one year.

In the North Star group, standard annual assessment of ambulatory patients with DMD includes measurement of 10m walk/run, time to stand from supine and stair climb. These tests have been validated by the North Star group for use in clinical monitoring and their measurements are included in other trials. The ERG requested information on the MCID for the timed function tests. The Company response stated that for the 10 metre walk/run the MCID is 0.76 seconds,[54] but that estimates of the MCID for the other timed function tests could not be identified.

4.2.4.3. Muscle strength

Force exerted during knee flexion and extension, elbow flexion and extension, and shoulder abduction was measured using myometry. The CS states on p. 101 (Results section) that "*myometric evaluation of limb strength is less sensitive to changes in disease status compared to TFTs, and muscle strength, although severely affected in ambulatory patients with DMD, deteriorates at a much slower rate than muscle function.*" The CS also justifies the inclusion of post hoc subgroup analysis in patients aged 5 to 6 by stating that "*myometry can only be adequately evaluated in younger patients*" (CS p. 102). The applicability of myometry in the whole trial population is therefore uncertain.

4.2.4.4. Ability to undertake activities of daily living

'Activities of daily living' were not evaluated by a specific validated tool, however the CSR states that the timed function tests (stand from supine, 4-stair ascend, 4-stair descent, 10 metre run/walk) measure the ability of patients to perform brief activities that are typical of patients' activities of daily living in a home, school, or community setting (CSR p.124, also confirmed in the response to clarifications). The ERG notes that there are other activities of daily living that are not captured in

Analysis	Placebo Baseline	Placebo Δ At week 48	Ataluren 40 mg/kg/day Baseline	Ataluren 40 mg/kg/day Δ At week 48	Difference between groups	Difference between groups (95% CI)
ITT All patients Placebo n=57, ataluren, n=57	359.6 m (87.7)	-42.6 m (90.1)	350.0 m (97.6)	-12.9 m (72.0)	29.7 m	26.4 m (-4.2, 57.1) p=0.0905
cITT All patients Placebo n=57, ataluren, n=57	361.1 m (87.5)	-44.1 m (88.0)	350.0 m (97.6)	-12.9 m (72.0)	31.3 m	31.7 m (5.1, 58.3) p=0.0197

Reproduced from CS Table C9.14 p. 90. Δ: change from baseline; MMRM: Mixed Model Repeated Measures; cITT: corrected intention to treat (post hoc analysis); ITT: Intention to treat.

Statistical significance can only be inferred for the modelled difference using MMRM from Table 8. P-values for the observed difference are not reported in the CS. The ERG was unclear why the reported p-values for the modelled difference (MMRM column) in the CS are different to the p-value for the main outcome reported in the EMA report (p= 0.0281) for the nominal (unadjusted) p value. The EMA also reported the adjusted p-value = 0.0561 which suggests lack of statistical significance of the difference between ataluren and placebo in 6MWD. The CSR was consulted to investigate this discrepancy. The following table (Table 9) was reproduced from Table 28 on page 100 of the CSR with the following outcomes reported for the ataluren 10, 10, 20 mg/kg vs placebo comparison.

Table 9 Post hoc MMRM Analysis of Change in Untransformed 6MWD Based on

Analysis	**Ataluren 10, 10, 20 mg/kg vs Placebo**			
	Difference		**p-value**	
	mean	**95% CI**	**nominal**	**adjusted**
MMRM[a]	31.7	5.1, 58.3	0.0197	0.0367[b]
Permutation test[c]	--	--	0.0281	0.0561[d]

[a] MMRM model: 6MWD = baseline 6MWD (covariate) + arm + visit + visit*arm + baseline 6MWD*visit + age group (<9 vs =9 years) + corticosteroid (yes vs no); unstructured variance/covariance matrix.
[b] Dunnett's test was applied to adjust for the comparison of 2 dose levels vs placebo.
[c] Permutation test of 10,000 re-randomizations. For each re-randomization, patients were dynamically re-randomized in the same order as they originally entered the study (starting seed = 14576).
[d] Based on the proportion of the 10,000 permutations in which the maximum effect size among the 2 comparisons (10, 10, 20 mg/kg vs placebo and 20, 20, 40 mg/kg vs placebo) exceeded the observed maximum

The submissions testify to a reduction in emotional and psychological burden of the condition with treatment. No submissions report whether there is a reduction in the practical burden, for example, if carers are able to return to work as a result of the greater independence of the child owing to treatment.

There is little discussion of the longer-term effects of treatment with ataluren. One submission discusses the impact that stopping treatment between trials had on the child, where there was a reverse of many of the positive benefits that had been seen.

The ERG notes that there are no details on how generalisable these views are to the wider UK nmDMD community. It is expected that there is a positive response bias to these submissions.

4.6.2 Summary of main conclusions from the EMA

Another additional piece of work undertaken by the ERG was consideration of The European Medicines Agency report (2015).[1] This report identified a need for input from a specialist Scientific Advisory Group (SAG) Neurology on three specific questions which are pertinent to this HST. Since our clinical experts advising the ERG on this HST have declared conflicts (e.g. reimbursement from PTC, advisor to PTC,) the ERG decided to summarise these points made by the SAG to gain a broader consideration of the evidence base.

a) Question 1: Does the SAG consider that the evidence for the mechanism of action of ataluren (nonsense mutation read-through) is convincing, and the results on dystrophin production could be seen as supportive of the pharmacodynamics of ataluren?

"The SAG considered that mechanism of action seemed plausible, but the experts felt that the provided data were still not convincing enough, and that they would need more information in order to be certain. The same was true for the data provided on dystrophin production in this case, that at least the data from the available biopsies, limited as they may be, should be provided. Thus the SAG considered that presently the available data on dystrophin production cannot be used as supportive of the pharmacodynamics of ataluren." (*page 49-50 of EMA).*

In agreement with the evaluation made by the SAG, the ERG noted that there was limited data available, even when considering the more recent available evidence published since the EMA report.

b) Question 2: Does the SAG agree that the presented pre-clinical and clinical evidence supports the bell shaped dose-response curve and hence, the absence of efficacy at the higher dose studied?

"The SAG considered that the proposed hypothesis for the bell shaped dose response curve seemed likely, but once again the experts felt that additional information was needed. More specifically, it was noted that while evidence on the bell-shape dose-response curve was available in several pre-clinical models, no data were generated in the mdx mouse model, relating the production of dystrophin to the levels of ataluren in the muscle fibres. Such evidence would be considered of relevance, as the available data describe only the relationship between plasmatic levels of ataluren and dystrophin production.

Overall, the SAG was of the view that no clear-cut conclusions could be derived on the bell-shaped dose-response hypothesis and the absence of efficacy in the higher dose studied in the Ph II trial." (*page 50 of EMA).*

c) Question 3: Does the SAG consider, based on the data presented by the Applicant, that the observed effects are sufficiently robust and clinically meaningful taking into account the results on the primary and secondary endpoints?

"The SAG considered that although the results were not sufficiently robust, the demonstrated effects were encouraging. The robustness of the results was challenged because of the observed variability in the primary efficacy data, the fact that many of the important conclusions supporting the efficacy of the drug were derived from the performed post hoc analyses, and the fact that there was little supportive evidence of effect from the data on the secondary endpoints. At the same time it was recognized that at the time the study was designed the knowledge of the natural history of the disease was different from what we now know. It was agreed that the applicant has performed the post hoc analyses in line with the most current knowledge about the natural history of the disease, and in this respect the definition of the sub-groups in these analyses is clinically and scientifically justified. The SAG experts considered that the results derived from these may be considered clinically relevant, especially in the sub-group of patients with more advanced disease. Additionally it was considered that the lack of effect on the secondary endpoints could be explained by the expected mechanism of action of the drug i.e. partial restoration of dystrophin production. Most of the secondary endpoints are of such nature that any effect will have to be driven by an increase in strength, rather than an improvement of function. The experts were presented with the latest available data, showing that minimal increase in dystrophin production could lead to functional improvement, but not to improvement of strength, and for the latter to occur, levels of dystrophin close to the ones in normal

muscular fibres must be achieved. The SAG experts agreed that this could be a valid explanation of the lack of concordance between the primary and secondary endpoints' efficacy data. It was also the position of the group that despite the fact that efficacy was most prominently shown in the subgroup of patients with more advanced disease, there were trends of efficacy in all the sub-groups by severity,

creatine kinase expression), no statistically significant differences were reported between ataluren and placebo in either study. On sub-groups defined by condition severity, it was reported that results favoured ataluren over placebo though no statistical tests were reported.

- Similar rates of severe adverse events were experienced by patients receiving ataluren and placebo but there were difference in types of event. Gastrointestinal disorders, vomiting, falls, investigations, weight decreases, metabolism and nutrition disorders, decreased appetite, musculoskeletal and connective disorders, back pain, headaches and nervous system disorders were more likely to occur with ataluren. In contrast, patients receiving placebo had higher rates of infections and infestations.
- No deaths were reported from either study.
- The Company presented a cumulative summary of serious adverse events from four ongoing and five completed Company-sponsored clinical trials. This appeared to suggest that serious cardiac disorders, infections and infestations, injury poisoning and procedural complications (specifically femur fractures) and total number of serious adverse events are more common with ataluren than placebo, however it is not clear from the information provided whether this is due to longer exposure in the ataluren group.

4.7.3 ERG assessment of uncertainties in clinical effectiveness

- A key criterion for the appraisal, and for the evaluation undertaken in the RCT and the CS was the definition of ambulatory. The NICE scope does not provide a clear definition. The RCT states that for inclusion in the study patients had to be able to walk ≥75 metres. The criteria used in the RCT are adopted by the company in the CS for the systematic review of clinical effectiveness. However, the CS economic model adopted a definition of loss of ambulation (i.e. inability to walk >0 metres). Inevitably the different definitions may influence the outcomes of the assessment.
- The comparator adopted in the RCT was best supportive care. Given that it was a multinational trial, it was felt that there may be heterogeneity in the comparator that may affect the outcome and influence its external validity.
- The selection of evidence through the search strategy and the selection process had the potential to affect the evidence reviewed in the systematic review of clinical

Table 1 Summary of key model input parameters and sources as reported in the Company's submission

Model inputs	Source(s)
Time to loss of ambulation: intervention	Derived based on information reported by Bushby et al. (2014)[41]
Time to loss of ambulation: best supportive care	Derived based on information reported by Ricotti et al. (2013)[5]
Non-ambulation to non-ambulation VA	Derived based on information reported by Humbertclaude et al. (2012)[58]
Non-ambulation to non-ambulation and scoliosis	
Non-ambulation to non-ambulation and scoliosis and VA	
Other cause mortality	ONS 2014
Death from nmDMD	Derived based on information reported in Norwood et al. (2009)[29]
Health state costs	Landfelt et al., 2014;[34] ONS 2015; OECD 2015[59]
Health state utility values	Landfeldt et al., 2014[34]
nmDMD, nonsense mutation Duchenne dystrophy; VA, ventilation assisted; ONS, Office of national statistics	

Information required to populate the model was obtained from Study 007 and published sources. Transition probabilities required for the transition to loss of ambulation health state were derived from Study 007. Transitions from the non-ambulant state to more severe health states were derived from Humbertclaude et al. (2012).[58] Information on costs was obtained from secondary sources and converted to UK pounds using UK 2012 purchasing power parity and inflated to 2014 costs using the consumer price index for health. In the ataluren group, treatment was dependent on the bodyweight of children until they reached 19 years old after which a constant weight of 70kg was assumed. Children in the intervention group received treatment until they progressed to the non-ambulatory stage. It was stated that children would continue to receive ataluren treatment for six months after loss of ambulation, but costs for this treatment were not included in the model. In the best supportive care group, children continued to receive the same treatment after loss of ambulation. Adverse events were not considered in the model.

In the model the primary measure of effectiveness was quality-adjusted life-years (QALYs), gained

over the duration of the model. (The time horizon was set at 'until the last patient loses ambulation'). All costs and benefits were discounted at 3.5% per annum. The base care analysis was conducted from an NHS and PSS perspective (with a scenario analysis from a wider societal perspective), and results were presented in terms of disaggregated costs, life-years gained (LYG) and QALYs. In the submission, one-way sensitivity analyses were undertaken by varying direct costs of health states, and patient and caregiver utility values by ± 20%. Also, a number of scenario analyses were undertaken: increasing caregivers' disutilities; increasing costs and disutilities for people requiring ventilatory assistance; inclusion of direct and indirect non-medical costs; and increasing the time horizon of the model.

5.3.3.1. Relative treatment effects of ataluren versus standard care

The model uses clinical effectiveness estimates for ataluren and best supportive care versus best supportive care alone Study 007 (Bushby et al. (2014)[41]) and from other published sources. It is important to note that this approach assumes that the populations from the different studies are comparable. Information on the delay in reductions in ambulatory ability (measured using the 6MWD) with ataluren were obtained from Study 007, and information about loss of ambulation with best supportive care were obtained from Ricotti et al. (2013).[5] Transition probabilities from loss of ambulation to more severe health states were obtained from a study of the natural history of DMD (Humbertclaude et al., 2012).[58] Additional information on background all-cause mortality was obtained from the Office of National Statistics (2014).

5.3.3.2. Transition probabilities for standard care

Improvements in ambulation with ataluren, compared to best supportive care, were estimated based on a least squares regression of changes in 6MWD from week 24 to week 48 of Study 007. The regression analysis was undertaken on the data from Week 24 to Week 48 because it was deemed to be more representative of the long-term treatment effect of ataluren (Company submission: expert opinion). The authors suggested that this is a conservative assumption because ataluren has a greater benefit compared to best supportive care in improving 6MWD in the first 24 weeks of the study.

Results from the regression analysis based on information from Week 24 to 48 showed that there was a decrease in the 6MWD of 59.0m in the best supportive care arm compared to a decrease of 25.2m in the ataluren arm. (33.8m between treatment groups). These declines in 6MWD were linearly extrapolated (from a mean baseline 6MWD of 355.7m) to estimate mean time to loss of ambulation, defined as 6MWD = 0m. As a result of this linear extrapolation, loss of ambulation was assumed to occur in the best supportive care and ataluren arms at week 313 (6 years) and week 733 (14.1 years), respectively. This equated to a difference of 420 weeks/8.1 years. (Please see Section 5.5 of this report for a critique of this approach).

6.5 Discussion

The first four models all give relatively similar results, but the 5th and 6th are very different, due principally to the change in distribution used to extrapolate loss of ambulation in the best supportive care arm. The 5th model uses the distributions with the best statistical fit, but it is also important to consider whether the results it produces are deemed clinically plausible. Model 6 is based on re-digitisations of data undertaken by the ERG, together with the best statistically fitting models, adjusted for clinically plausibility (specifically time before loss of ambulation in the BSC model). Model 2 is the most recent analysis undertaken by the company, whilst model 7 is the ERG's "most plausible" scenario.

In addition to the elements of uncertainty which the ERG has been able to address quantitatively, there are a number of other areas of uncertainty it is important to consider. Some of these are related directly to a lack of underlying data, but others are as a result of choices made in the modelling process which have not been quantitatively considered in the Company submission. These include:

- The use of a cohort with a starting age of 8.5, rather than 5 years as specified in the scope.
- The assumption that the treatment benefit with ataluren is permanent, with the advantage over best supportive care found between weeks 24 and 48 of Study 007 continuing until people lose ambulation.
- The use of a linear extrapolation of mean difference in 6MWD which relies on the assumption of a homogeneous population following the same trajectory of progression. Such an approach is not valid if this assumption is not met.
- No additional treatment related adverse events with ataluren which engender costs or reductions in quality of life.
- Treatment adherence to ataluren is 100%, and no-one will discontinue treatment for any reason other than loss of ambulation.
- There are no additional costs for administration, training or monitoring related to ataluren treatment.

All these assumptions appear to be optimistic ones and it therefore seems appropriate to regard the results produced by the model as an optimistic upper bound on the possible benefits of ataluren treatment.