CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

215515Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #: 215515

Drug Name: Vutrisiran

Indication(s): HTTR amyloidosis

Applicant: Alynlyam

Date(s): April 14, 2021

Review Priority: Standard

Biometrics Division: I

Statistical Reviewer: Tristan Massie, Ph.D.

Concurring Reviewers: Kun Jin, Ph.D., Team Leader
James (Hsien-Ming) Hung, Ph.D., Division Director

Medical Division: Division of Neurology

Clinical Team: Rainer Paine, M.D.
Laura Jawidzik, M.D., Team Leader

Project Manager: Anhtu Nguyen

Keywords:
External Control; Single Study; Rare Disease

Table of Contents

1 EXECUTIVE SUMMARY 5

2 INTRODUCTION 5

2.1 OVERVIEW 5
2.2 DATA SOURCES 6

3 STATISTICAL EVALUATION 6

3.1 DATA AND ANALYSIS QUALITY 6
3.2 EVALUATION OF EFFICACY 6
3.2.1 Study ALN-TTRSC02-002 6
3.2.1.1 Study Design and Endpoints 7
3.2.1.2 Statistical Methodologies 7
3.2.1.3 Patient Disposition, Demographic and Baseline Characteristics 16
3.2.1.4 Results and Conclusions 20
3.2.1.4.1 Sponsor's Results 20
3.2.1.4.2 Reviewer's Results 28
3.3 EVALUATION OF SAFETY 30

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 30

4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION 30
4.1.1.1 Individual Sites 31
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS 33

5 SUMMARY AND CONCLUSIONS 33

5.1 STATISTICAL ISSUES 33
5.2 COLLECTIVE EVIDENCE 33
5.3 CONCLUSIONS AND RECOMMENDATIONS 33

LIST OF TABLES

Table 1 Double Blind Phase 3 Sham Controlled Study Characteristics 6

Table 2 Comparison of Baseline Demographic Characteristics 18

Table 3 Comparison of Baseline Disease Characteristics 19

Table 4 Primary Analysis of mNIS+7 20

Table 5 mITT Analysis of mNIS+7 (no imputation of missing data) 21

Table 6 ITT analysis of Norfolk QoL-DN with multiple imputation 24

Table 7 mITT analysis of Norfolk QoL-DN 24

Table 8 Impact of COVID-19 Pandemic on HELIOS-A 26

Table 9 Comparison of Baseline mNIS+7 between current (HELIOS-A) and external control (APOLLO) studies 28

LIST OF FIGURES

Figure 1 Components of the Primary Endpoint, Modified NIS+7 10

Figure 2 Scoring algorithms for Nerve Conduction items in mNIS+7 12

Figure 3 mNIS+7 Primary Analysis 22

Figure 4 Serum Transthyretin comparison between Vutrisiran and Patisiran (HELIOS-A) 27

Figure 5 Sponsor's Subgroup Analyses 31

Figure 6 Primary Efficacy Comparison by Individual Sites 32

1 EXECUTIVE SUMMARY

The Division had agreed to allow the sponsor to compare the investigational treatment Vutrisiran, to the placebo arm from the sponsor's APOLLO trial involving the approved predecessor product, Patisiran, which was also used as an active control in the current HELIOS-A trial. The APOLLO placebo and HELIOS-A Vutrisiran arms seem to have several notable differences in patient characteristics. For example, the baseline measurement of the primary efficacy measure: baseline mNIS+7 mean: 74.6 vs. 60.6 and Race proportions are different (32 vs. 17% Asian, genotype: 52 vs 44% V30M, 75 vs. 65% Male, 74 vs 61% NIS $\geq$50). Comparison to an external control with such clear differences in patient composition is not likely to be reliable. It is not clear how best to account for the bias that might be attributed to these baseline differences between the investigation drug arm and the historical placebo. Vutrisiran was not statistically superior to the concurrent control arm Patisiran, but rather they appeared relatively similar in terms of change from baseline to Month 9 in mNIS+7 and a prespecified noninferiority margin for change in TTR was met . Considering the large effect Patisiran displayed in APOLLO compared to the APOLLO placebo and the apparent large effect of Vutrisiran (HELIOS-A) to the same external APOLLO placebo and similarity to Patisiran in HELIOS-A and relative similarity to the historical Patisiran APOLLO, -2.0 (S.E.=1.5) mean mNIS+7 change at 9 months may allay these concerns about baseline differences in this case.

2 INTRODUCTION

2.1 Overview

The development plan for Vutrisiran (IND 139086) builds upon learnings from the prior development of another siRNA that targets TTR, Patisiran. The completed patisiran Phase 3 APOLLO study, which enrolled a similar patient population on face and incorporated similar endpoints to HELIOS-A, is used as an external control for efficacy analyses. The primary endpoint in this study, change from baseline in mNIS+7 score at Month 9, and the key secondary endpoint of Norfolk QoL-DN total score at Month 9, are compared with the placebo group from the APOLLO study at Month 9 as the primary analysis. To allow for a robust comparison with APOLLO, cross-study differences were attempted to be minimized by using similar inclusion/exclusion criteria; overlapping study sites and global footprint; and the same efficacy assessments, methodologies, training, and central laboratory for mNIS+7.

Use of an external comparator (together with a concurrent active control group) may be an acceptable clinical study design for diseases occurring in small populations, with well understood natural history of disease course. The natural history of hATTR amyloidosis has

been well-characterized in several, large, randomized clinical trials, including the APOLLO study, as well as natural history studies.

The Patisiran reference comparator group helps to validate the use of the external control for the primary and secondary efficacy analyses, both by allowing descriptive comparison of the clinical efficacy endpoints between treatment groups within this study and by establishing that a similar (non-inferior) level of TTR reduction is achieved for Vutrisiran and Patisiran.

Table 1 Double Blind **Phase 3 Sham Controlled Study Characteristics**

Study Name	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
HELIOS-A/APOLLO	3	9 months primary	18 months 9 months (primary)	ITT: 77 Placebo(APOLLO)*, 122Vutrisiran(HELIOS); 42 Patisiran(HELIOS)	Hereditary transthyretin amyloidosis

*External placebo control

2.2 Data Sources

The mNIS7 primary clinical endpoint data is contained in the following analysis dataset file.

```
\\CDSESUB1\evsprod\nda215515\0002\m5\datasets\aln-ttrsc02-
002\analysis\adam\datasets\adef.xpt
```

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The submitted data and analysis quality appear adequate.

3.2 Evaluation of Efficacy

3.2.1 Study ALN-TTRSC02-002

HELIOS-A: A Phase 3 Global, Randomized, Open-label Study to Evaluate the Efficacy and Safety of ALN-TTRSC02 in Patients with Hereditary Transthyretin Amyloidosis (hATTR Amyloidosis)

First Subject Enrolled: 14-Feb-2019
Data Cut-off Date: 10-Nov-2020

3.2.1.1 Study Design and Endpoints

This was a global, Phase 3 randomized, open-label study designed to evaluate efficacy, safety, and pharmacokinetics (PK)/pharmacodynamics (PD) of vutrisiran (ALN-TTRSC02) in adult patients with hATTR amyloidosis. Patients were to be randomized 3:1 to vutrisiran or patisiran, a reference comparator. Randomization was to be stratified by TTR genotype (V30M vs. non-V30M) and baseline NIS score (<50 vs ≥50).

The study was to consist of a Screening Period of up to 42 days, an 18-month Treatment Period, an 18-month Treatment Extension Period which was to include collection of safety and efficacy in patients who switch from patisiran to vutrisiran treatment, and up to a 1-year Follow-up Period after the last dose of study drug.

After the Screening period, and at the start of the Treatment Period, eligible patients were to be randomized 3:1 on Day 1 to receive 25 mg of vutrisiran administered as a subcutaneous (SC) injection once every 3 months (q3M) or patisiran administered as an intravenous (IV) infusion once every 3 weeks (q3w). During the 18-month Treatment Period, patients were to undergo assessments for efficacy and/or safety (as outlined in the Schedule of Assessments), with key efficacy assessments being performed prior to first dose, at Month 9 (primary efficacy analysis time-point) and at Month 18; samples for TTR assessment were to be collected more frequently throughout the 18-month Treatment Period.

3.2.1.2 Statistical Methodologies

SAP location: \\CDSESUB1\evsprod\nda215515\0003\m5\53-clin-stud-rep\535-rep-effic-safety-stud\hattr-pn\5351-stud-rep-contr\aln-ttrsc02-002\aln-ttrsc02-002-statistical.pdf

For efficacy assessments, if the scheduled visit (e.g., Month 9) was not performed, the unscheduled and/or discontinuation visits performed within a ± 3-month window were to be grouped with the scheduled visit. In situations in which a Month 9 or Month 18 efficacy visit is unable to be completed due to the Coronavirus disease 2019 (COVID-19) pandemic limiting the patient's ability or willingness to access the study center or their ability to have received their scheduled doses of study drug, Month 9 and Month 18 efficacy assessments may be completed within 6 months after the intended time point (i.e, up to Study Month 15 or Month 24, respectively). For patients impacted by the COVID-19 pandemic, efficacy visits delayed up to 6 months after the end of the protocol-defined efficacy visit window were to be included in the analysis. The derived visits were to be used for all analyses.
Unless otherwise specified above, data collected at unscheduled visits were to be included in by patient data listings and figures, but no assignment to a study visit was to be made for the purpose of by-visit summary tabulations. However, unscheduled visits may be used in the

calculation of baseline values and for inclusion in any categorical shift summaries (e.g., shift from baseline to "worst" postbaseline value).

This Phase 3 study was to use the APOLLO study placebo arm as an external control. Patient-level data from this study were to be compared with patient-level data from APOLLO for efficacy analyses. Except for TTR endpoints, analysis models were to include only the 2 treatment groups compared, vutrisiran and placebo (APOLLO), and only simple descriptives were to be presented for patisiran (HELIOS-A). For TTR endpoints, vutrisiran and patisiran (HELIOS-A) were to be compared unless otherwise specified.

Sample Size Determination

Approximately 160 patients were to be enrolled in this study, with a 3:1 randomization ratio to either vutrisiran or patisiran. The sample size was chosen to enable an adequate characterization of the long-term safety profile, as well as the efficacy of vutrisiran in this patient population. For the primary efficacy endpoint of mNIS+7 and the secondary endpoint of Norfolk QoL-DN total score, the vutrisiran group in the Phase 3 study was to be compared to the placebo group from the APOLLO study. For the mNIS+7 change from baseline at 9 months, the observed mean (±standard deviation [SD]) was 15 ± 17 points for the placebo group from the APOLLO study. Assuming a mean change of 0 points for the vutrisiran group, there was >90% power to establish the superiority over placebo using a 2-sided t-test with a significance level of 0.05. For the Norfolk-QoL DN total score change from baseline at 9 months, the observed mean (±SD) was 11.5 ± 19.2 points for the placebo group from the APOLLO study. Assuming a mean change of -4 points for the vutrisiran group, there was >90% power to establish the superiority over placebo using a 2-sided t-test with a significance level of 0.05.

Analysis Populations

- Modified Intent-to-Treat (mITT) population: All randomized patients who received any amount of study drug. Patients were to be analyzed according to the treatment to which they were randomized.
- TTR Per-protocol (PP) Population: All mITT population patients with a nonmissing TTR assessment at baseline and ≥1 trough TTR assessment between Months 6 (Week
- 24) and Month 18 [Week 72]) that meets the requirements described in sponsor Table 3. Patients were to be analyzed according to the treatment to which they were randomized.
- Month 9 Efficacy PP Population: All mITT population patients treated with vutrisiran or placebo meeting the following criteria:
    - Month 9 efficacy visit date within 3 calendar months of protocol-planned Month 9 efficacy visit window
    - No serious or severe COVID-19 custom query AE terms or reported on or before Month 9 efficacy visit date
    - For vutrisiran-treated patients, received all planned vutrisiran doses up to and including Week 36 with ≤ 28 day delay

- Patients were to be analyzed according to the treatment to which they were randomized.
- Month 18 Efficacy PP Population: All mITT population patients treated with vutrisiran or placebo meeting the following criteria:
    - Month 18 efficacy visit date within 3 calendar months of protocol-planned Month

Initiation of Local Standard Treatment for hATTR Amyloidosis

In the APOLLO study, there were placebo-treated patients who discontinued study drug, but remained on study and received local standard treatment. For the primary analysis of mNIS+7 and Norfolk QoL-DN, assessments were censored (excluded from analysis) after initiation of any of the following:

- Orthotopic liver transplant
- Use of TTR stabilizing agents (eg, tafamidis, diflunisal) for >14 days

For consistency of data handling, the placebo group from the APOLLO study was to follow the same censoring rule as the APOLLO study.

For this study, APOLLO censoring rules were to be applied. Additionally, assessments were to be censored after initiation of any of the following recently approved treatments:

- Any use of TTR-targeting anti-sense oligonucleotides (e.g, inotersen)
- Any use of patisiran (applicable for the vutrisiran treatment group only)

These data were to be included and flagged in efficacy listings. These assessments from either study were to be included in sensitivity analyses as specified.

For TTR percent reduction, TTR assessments collected after initiation of local standard treatment for hATTR amyloidosis were to be excluded from the analysis. For all other efficacy endpoints, data from either study collected after initiation of local standard treatment for hATTR amyloidosis were to be included in analyses.

A separate listing was to be provided for patients who initiated local standard treatment for hATTR amyloidosis while on study.

Modified Neuropathy Impairment Score (mNIS+7) and Neuropathy Impairment Score (NIS)

Note: the mNIS+7 and NIS measurements are conducted in duplicate per time point. The average of 2 complete duplicate values was to be reported, except in cases of missing or partially missing data as described in the table below.

Figure 1 Components of the Primary Endpoint, Modified NIS+7

Assessment Tool	Total Points	Components (maximum points)
Modified NIS+7	304	• NIS-W: Weakness (192) • NIS-R: Reflexes (20) • Quantitative sensory testing by body surface area including touch pressure (TP) and heat as pain (HP): QST-BSA$_{TP+HP5}$ (80) • ∑5 nerve conduction studies (10) – Ulnar compound muscle action potential (ulnar CMAP) – Ulnar sensory nerve action potential (ulnar SNAP) – Sural sensory nerve action potential (sural SNAP) – Tibial compound muscle action potential (tibial CMAP) – Peroneal compound muscle action potential (peroneal CMAP) • Postural blood pressure (BP) (2)

There are 5 components within mNIS+7 total score including NIS-W, NIS-R, QST, Σ5 NC, and postural BP, as described in detail below.

1. NIS-W is the sum of the cranial nerve components (3rd nerve, 6th nerve, facial weakness, palate weakness, tongue weakness) and muscle weakness components (respiratory, neck flexion, shoulder abduction, elbow flexion, brachioradialis, elbow extension, wrist flexion, wrist extension, finger flexion, finger spread, thumb abduction, hip flexion, hip extension, knee flexion, knee extension, ankle dorsiflexors, ankle plantar flexors, toe extensors, toe flexors). Assessments are performed separately for the right- and left-hand side of the body. Scoring for each component is 0 (normal), 1 (25% weak), 2 (50% weak), 3 (75% weak), 3.25 (move against gravity), 3.5 (movement, gravity eliminated), 3.75 (muscle flicker, no movement), and 4 (paralysis). The maximum total score for NIS-W is 192.
2. NIS-R is the sum of the reflex components (biceps brachii, triceps brachii, brachioradialis, quadriceps femoris, and triceps surae). Assessments are performed separately for the right- and left-hand side of the body. Scoring for each component is 0 (normal), 1 (decreased) and 2 (absent). Adjustments are made for the age of the patient (e.g., absent reflexes in a patient older than 60 is assessed as 0, or normal). The maximum total score for NIS-R is 20.
3. QST measures heat pain and touch pressure from among 10 distributed anatomical sites on 1 side of the body. Scoring is based on point abnormality (< 95th = 0 point, ≥ 95th – < 99th = 1 point and ≥ 99th = 2 points). The test score is 2 times the total scores for QST of heat pain by body surface area (QST-BSAHP) and QST of touch pressure by body surface area (QST-BSATP). The maximum total score for QST is 80. Missing values for QST- BSAHP and QST-BSATP will be imputed separately (see below), but only the total score will be summarized.
4. ∑5 nerve conductions (NCS) include ulnar CMAP, ulnar SNAP, sural SNAP, tibial CMAP, and peroneal CMAP. Their measured values were transformed to normal deviates from percentile values correcting for applicable variables of age, sex, height or weight as based on earlier studies of a large healthy patient reference cohort; the algorithms for transforming to normal deviates and points will be described in programming specifications. Additionally, these percentile values were expressed as points from obtained percentile values (i.e., > 5th =

0 points; ≤ 5th – > 1st = 1 point and ≤ 1st = 2 points (and similarly when abnormality is in the upper tail of the normal distribution). The total score is calculated as the mean point scores of the nonmissing 5 nerve tests after multiplying this value by 5. The maximum total score for ∑5 NCS is 10.

5. Postural blood pressure (PBP) test measures autonomic function to address risk of orthostasis. The points are assigned based on change in PBP (> -20 mmHg = 0 points;

   -30 – ≤ -20 mmHg = 1 point; and ≤ -30 mmHg = 2 points).

If at least 1 value at a time point and (replicate) assessment is missing, then do the following:

- Missing items within the ∑5 NCS component (e.g., missing tibial CMAP) are handled in step 4 above.

If 1 of the following components or subcomponents of mNIS+7 (NIS-W, NIS-S,

QST-BSATP, QST-BSAHP, the entire ∑5 NCS, or postural BP) is missing from 1 of the assessments at a time point ("replicate A"), then impute via substituting the component from the other assessment at that time point ("replicate B")

recover 2 complete replicate mNIS+7 measures A and B at that particular time point.

- If both replicates A and B are incomplete at different components, insert replicate A components into replicate B and replicate B components into replicate A as necessary allow recovery of 2 complete replicate mNIS+7 measures A and B.
- If NIS-W component is missing from both replicates A & B, mNIS+7 score is considered as missing. If 1 of the other mNIS+7 components or subcomponents (NIS-R, the entire ∑5 NCS, QST-BSATP, QST-BSAHP, or postural BP is missing from both replicates A and B, impute this component as the average of the component using data from any of the patients who had nonmissing data for that component at the time point (within study group).
- If 1 entire mNIS+7 replicate is missing, do not impute the missing replicate; just use the singular (other) replicate rather than the average of the 2 in subsequent calculations.
- If no single complete mNIS+7 measure can be found or recovered, mNIS+7 score is considered as missing.

Algorithms for Setting Normal Deviates and Points

For nerve conductions, raw values are provided by the (b) (4). Each raw value is first converted to a z-score which is then used to set either normal deviate or point score.

Figure 2 Scoring algorithms for Nerve Conduction items in mNIS+7

Parameter (abbreviation, units)	**Z-score Equation**
Ulnar CMAP (UMAE, mV)	UMAEz = UMAE - (12.34105776660360 + -0.04413566229394*age)
Peroneal CMAP (PMAK, mV)	PMAKz = PMAK - (7.26194271992764 + -0.04324792361150*age)
Tibial CMAP (TMAK, mV)	TMAKz = TMAK - (11.91330602787780 + 0.79541113237729*age + -0.01622568048284*age^2 + 0.00008722708244*age^3 + -5.49838000227032*bsa)
Sural SNAP (SSAB, uV)	y = ln(1+SSAB) - (5.58389110852732 + 0.01546531508256*age + -0.01693664229181*height + -0.00034198271691*age^2) if y > 0 then SSABz = y/(0.7001926549587 + -0.20662093194825*bsa) else if y ≤ 0 then SSABz = y
Wrist-Fifth Ulnar SNAP (USAW, uV)	y = ln(1+USAW) - (3.38461614473185 + -0.01702085210501*age + 0.59463420483099*sex) if y > 0 then USAWz = y else if y ≤ 0 then USAWz = y/(0.16695910921534 + 0.00291051555416*age)

1) sex is coded 1 = male, 2 = female.
2) height is in cm.
3) weight is in kg.
4) bsa = body surface area (m^2) = weight$^{0.425}$ x height$^{0.725}$ x 0.007184.
5) age is an integer rounded down to the nearest year.

For a given parameter, the calculated z-score is then compared to 2 lookup tables (see below) using the following procedure to assign the normal deviate score:
1. If the calculated z-score value is less than the lower z-score value (ZLOWER) in the Extended Percentiles (EP) table, then the normal deviate value is the maximum of (-3.72, (z-score – ZLOWER + WLOWER)).
2. If the calculated z-score value is >= ZLOWER and less than the minimum z-score value in the All Percentiles (AP) table, then the normal deviate value = -2.75.
3. If the calculated z-score value is greater than upper z-score value (ZUPPER) in the EP table, then the normal deviate value is the minimum of (3.72, (z-score – ZUPPER + WUPPER)).
4. If the calculated z-score value is <= ZUPPER and greater than the maximum z-score value in the AP table, then the normal deviate value = 2.75.
5. If the calculated z-score falls between 2 values in the AP table, select the z-score with the normal deviate value closer to 0 (i.e., the 50th percentile). The normal deviate value is the value from the Normal Deviate column corresponding to the selected z-score. For example, if UMAEZ = -3.3 then the UMAE normal deviate (UMAEND) value = -1.65. If UMAEZ = 5.1 then UMAEND = 2.05.
6. If the calculated z-score ≤ the z-score corresponding to the 50th percentile and exactly matches a z-score for that parameter in the AP table, then select the row with the next highest z-score value. The normal deviate value is the value from the Normal Deviate column corresponding to the selected z-score. For example, if UMAEZ = -4.8488 then UMAEND = -2.05.

7. If the calculated z-score > the z-score corresponding to the 50th percentile and exactly matches a z-score for that parameter in the AP table, then select the row with the next lowest z-score value. The normal deviate value is the value from the Normal Deviate column corresponding to the selected z-score. For example, if UMAEZ = 5.1216 then UMAEND = 2.05.
The points for the parameters are calculated as follows:

- For UMAE, PMAK, TMAK, SSAB, and USAW: normal deviates greater than -1.65 are assigned 0 points, normal deviates equal to or less than -1.65 but greater than -2.33 are assigned 2 points.

ANCOVA/MI

ANCOVA incorporating MI was to be the default analysis for most continuous efficacy endpoints at Month 9. MI is a broadly applicable technique for handling missing data. Missing data are imputed multiple times using a regression method. Each imputed data set is analyzed using the same analysis model, and the point estimates and standard errors are combined to provide inferences that reflect the uncertainty about the missing values. MI assumes the missing data are missing at random (MAR).
For a given endpoint, missing endpoint values were to be multiply imputed separately for each treatment group using a regression procedure, with baseline information including baseline score and KPS as covariates and genotype, age at hATTR symptom onset, prior tetramer stabilizer use, region, FAP stage (I vs. II/III), Cardiac subpopulation, sex, and baseline NIS (<50 vs. =50) as factors. For NIS-related endpoints, the categorical baseline NIS score was not to be included in the regression procedure.
One hundred imputed datasets (per treatment group) were to be generated from the MI regression procedure using SAS PROC MI. Each of the imputed datasets were then to be analyzed using an ANCOVA model, including a covariate (baseline value) and factors (treatment group; genotype; age of disease onset, baseline NIS score [<50 vs ≥50]), unless otherwise specified. For NIS related endpoints, the categorical baseline NIS score was not to be included in the model. The resulting estimates (LS mean differences and standard errors) from the 100 imputed datasets Were to be combined using SAS PROC MIANALYZE to produce inferential results (difference in LS means, 95% CI for the difference, and the P value from the test that the difference is zero)[Rubin 1996]. Combined LS mean estimates were to be calculated as the average of the 100 complete-data estimates. A total variance estimate was to be calculated as a weighted sum of within-imputation variance, which is the average of the complete-data variance estimates, and a between-imputation variance term. Complete details may be found in the SAS documentation for the MIANALYZE procedure (see Combining Inferences from Imputed Data Sets under Details: http://support.sas.com/documentation/onlinedoc/stat/131/mianalyze.pdf).

Primary Efficacy Evaluations

The primary endpoint is change from baseline at Month 9 for mNIS+7. The primary comparison was to be conducted at Month 9. The primary endpoint was to be analyzed using the general ANCOVA/MI methods. Additionally, change from baseline at Month 18 for mNIS+7 was to be analyzed as a secondary endpoint using the general MMRM methods, and Month 9 and 18 LS mean estimates from this MMRM were to be presented graphically as well.

Sensitivity Analysis: Including Data Post Local Standard Treatment for hATTR amyloidosis or Post Serious COVID-19 AE

The primary analysis was not to include assessments performed after the initiation of local standard treatment for hATTR amyloidosis or on or after onset of a serious COVID-19 AE. Sensitivity analysis of mNIS+7 change from baseline including data post local standard treatment for hATTR amyloidosis or post serious COVID-19 AE from either study was to be conducted using the ANCOVA method at Month 9 and MMRM method at Month 18.

Sensitivity Analysis: Propensity Score

To allow some control of more factors and covariates without saturating the model, a propensity score approach was to be used to reduce the predictors to a single propensity score. The propensity score is defined as the probability of being treated with vutrisiran as obtained from a logistic regression model of treatment group [vutrisiran; placebo (APOLLO)]. The logistic regression model was to include the following baseline variables:

• Continuous variables
− NT-proBNP (log-transformed)
− mNIS+7
− Norfolk QoL-DN total score
• Categorical variables
− Previous tetramer stabilizer use (tafamidis/diflunisal) [Yes; No]
− Karnofsky Performance Status (KPS) [60; 70-80; 90-100]
− Cardiac Subpopulation [Yes; No]
− PND score [I; II; IIIA; IIIB/IV]
− Age at hATTR Symptom onset [< 50; ≥ 50]
− Neuropathy Impairment Score (NIS) [< 50; ≥ 50]
− Genotype [V30M; non-V30M]
− FAP stage [I; II/III]

The primary endpoint was to be analyzed in this sensitivity analysis using the ANCOVA method at Month 9 and MMRM method at Month 18, including the propensity score covariate in addition to the default model factors and covariates.

Sensitivity Analysis: Pattern-Mixture Model

The primary analysis ANCOVA/MI method addresses data under missing at random (MAR) assumptions. To assess the robustness of the primary analysis results under missing not at random (MNAR) assumptions, a sensitivity analysis using a pattern-mixture model (PMM) was to be conducted at Month 9 using a modified ANCOVA/MI method.

The model was to be based on the following assumptions:

1. Patients who have missing data due to COVID-19, including patients who have missing assessments, who have data censored because a serious COVID-19 AE was reported before Month 9, or who die due to COVID-19:

a. Under the hypothetical estimand of interest where the COVID-19 pandemic did not occur, these assessments should have been obtained with no COVID-19 impact. Therefore, for patients meeting either of these criteria, assessments were to be considered MAR, and were to be imputed using MI estimated from all non-missing data collected on treatment from the vutrisiran group.

2. Patients who have missing data unrelated to COVID-19 and are alive before Month 9:

a. Placebo-treated patients who have missing data: The missing data are considered MAR and were to be imputed using MI estimated from placebo-treated patients. The imputation is done

regardless of whether a patient was on-treatment or discontinued treatment before the scheduled Month 9 efficacy assessment.
b. Vutrisiran-treated patients who have missing data while on treatment: Patients are expected to continue to show benefit from treatment similar to that observed at the scheduled time point. Therefore, missing data during the on-treatment period (within 126 days of the patient's last dose before the scheduled Month 9 efficacy assessment) are considered MAR and were to be imputed using MI estimated from all non-missing data collected on treatment from the vutrisiran group. The 126-day window was selected given the long PD effect of vutrisiran.
c. Vutrisiran-treated patients who have missing data after stopping their study treatment: Patients will no longer benefit from treatment in the future and would be expected to have trajectory similar to placebo-treated patients. Therefore, missing data after treatment discontinuation (more than 126 days after the patient's last dose of study drug before the scheduled Month 9 efficacy assessment) were to be imputed using the data from placebo-treated patients.
3. Patients who have missing data and who die before Month 9 unrelated to COVID-19:
a. Assuming deaths observed in the study were likely to be related to worsening of disease, the missing data were to be imputed by taking random samples from the worst 10% mNIS+7 change from baseline scores among vutrisiran- and placebo-treated patients at Month 9. The imputation was to be done for patients from both vutrisiran and placebo groups.
Following the procedure described above 100 imputed datasets were to be generated and each dataset was to be analyzed and estimates combined as for the ANCOVA/MI described above.

Impact on Efficacy

Per protocol amendment 3 (17 July 2020), efficacy assessments that may be missed due to COVID-19 may be delayed up to 6 months after the scheduled timepoint to minimize missed endpoint ascertainment. Such delayed efficacy assessments due to COVID-19 were to be included in analyses.

Additional analyses of the primary and key secondary endpoints were to be conducted based on the Month 9 Efficacy PP population using an ANCOVA model at Month 9, and based on the Month 18 Efficacy PP population using an MMRM at Month 18. These analyses represent the initial visit windows for the efficacy assessments in place prior to the COVID-19 pandemic.

Additional descriptive summaries for the primary and secondary endpoints were to be provided by pandemic phase. Pandemic phase definitions may vary over time given the evolving nature of the COVID-19 pandemic; potential definitions may include the following:

- Before and during pandemic, where assessments will be considered during the pandemic if the event occurs on or after first confirmed case of COVID-19 based on the country where the study site is located.
- Before March 2020, between March 2020 and June 2020, and after June 2020.

Multiple Comparisons Procedure (US/Japan/Brazil)

In the US, Japan, and Brazil, the overall familywise error rate was to be controlled at α=0.05 for the primary and secondary endpoint hypothesis tests as follows:

1 Modified Neuropathy Impairment Score +7 (mNIS+7) change from baseline at Month 9
2 Norfolk Quality of Life-Diabetic Neuropathy (NorfolkQoL-DN) total score change from baseline at Month 9
3 10-MWT gait speed change from baseline at Month 9

For the US filing, results for both the primary endpoint, mNIS+7 change from baseline at Month 9, and key secondary endpoint, Norfolk QoL-DN total score change from baseline at Month 9, must be statistically significant to declare a positive trial.

Key secondary endpoint: Norfolk QoL-DN Total Score

Change from baseline in Norfolk QoL-DN total score (primary analysis incorporating COVID-19 pandemic impact questions was to be analyzed using an ANCOVA/MI model at Month 9, and using an MMRM at Month 18. Sensitivity, binary, and Efficacy PP Population analyses for Norfolk QoL-DN total score were also to be conducted.

10-meter Walk Test Speed, mBMI, and R-ODS

For 10-meter walk test speed, mBMI, and R-ODS, change from baseline was to be analyzed using an ANCOVA/MI model at Month 9 (with mBMI and R-ODS analyzed at Month 9 as exploratory endpoints), and using an MMRM at Month 18. Binary analyses for 10-meter walk test speed were also to be conducted.

TTR Percent Reduction

TTR percent reduction through Month 18 is defined as the average trough (i.e., predose) TTR percent reduction from Month 6 to 18, which is the steady state period for both vutrisiran and patisiran. Only trough TTR assessments meeting requirements described in the TTR PP population definition were to be included. The Hodges-Lehmann method [Hodges and Lehmann 1962], stratified by previous TTR stabilizer use (yes vs no), where values within each stratum were to be first aligned by the within-stratum 1-sample Hodges-Lehmann median, and then used to estimate the 95% CI for the median difference between the vutrisiran and patisiran groups in this study. Non-inferiority of vutrisiran (versus patisiran) was to be declared if the lower limit of the 95% CI for the median treatment difference in TTR percent reduction (vutrisiran - patisiran) in this study is greater than –10%.

Sensitivity analyses using the same analysis method were to be conducted to compare the TTR percent reduction through Month 18 between the vutrisiran group from this study and the pooled patisiran group from this study and the APOLLO study.

3.2.1.3 Patient Disposition, Demographic and Baseline Characteristics

Patient disposition based on the 10 November 2020 data cutoff date for CSR1 is as follows.

Of the 189 patients screened, 164 patients were randomized to either vutrisiran (122 patients) or patisiran (42 patients). All randomized patients were treated with study drug.
Patients were randomized and treated at 57 study centers in 22 countries. Countries that randomized ≥10 patients included the US (33 patients [20.1%]), France (20 patients [12.2%]), Bulgaria (13 patients [7.9%]), Portugal (13 patients [7.9%]), and Australia (12 patients [7.3%]).
Of the 122 patients in the vutrisiran group, 3 (2.5%) patients discontinued study drug prior to Month 9: 2 study drug discontinuations were due to death of the patient unrelated to study treatment and 1 study drug discontinuation was due to physician decision for a patient who did not comply with study visits and was considered lost to follow up. These same 3 patients stopped study participation. None of the patients in the vutrisiran group discontinued study drug after Month 9. As of the data cutoff date, 119 (97.5%) patients in the vutrisiran group were still participating in the study (note: primary timepoint is Month 9 but double blind treatment period follow-up is 18 months).
Of the 42 patients in the patisiran group, 2 (4.8%) patients discontinued study drug prior to Month 9, both due to the death of the patient unrelated to study treatment.
In addition, 2 (4.8%) patients in the patisiran group discontinued study drug after Month 9: 1 (2.4%) due to the death of the patient unrelated to study treatment and 1 (2.4%) due to an unrelated AE. All 4 of these patients stopped study participation. As of the data cutoff date, 38 (90.5%) patients in the patisiran group were still participating in the study.

Baseline Demographics and Disease Characteristics

In the vutrisiran group, the mean age was 57.8 years (range, 26 to 85 years) and the majority of patients were white (70.5%) and male (64.8%) (Table 2). Demographic characteristics were similar in the patisiran group, with a mean age of 58.0 years (range, 31 to 81 years), 69.0% of patients were white, and 64.3% of patients were male. A comparison of the characteristics between APOLLO and HELIOS-A exhibited some notable between-study differences (e.g., Age, Sex, and Race differences, as well as baseline mNIS+7 differences).

Table 2 Comparison of Baseline Demographic Characteristics

Demographic	APOLLO	HELIOS-A		
	Placebo (N=77)	Vutrisiran (N=122)	Patisiran (N=42)	Total (N=164)
Age at informed consent (years)				
Mean (SD)	62.2 (10.8)	57.8 (13.2)	58.0 (10.5)	57.9 (12.5)
Median (min, max)	63.0 (34, 80)	60.0 (26, 85)	60.0 (31, 81)	60.0 (26, 85)
Age group (years), n (%)				
18 to 64	44 (57.1)	76 (62.3)	31 (73.8)	107 (65.2)
65 to 74	24 (31.2)	39 (32.0)	9 (21.4)	48 (29.3)
≥75	9 (11.7)	7 (5.7)	2 (4.8)	9 (5.5)
Sex, n (%)				
Male	58 (75.3)	79 (64.8)	27 (64.3)	106 (64.6)
Female	19 (24.7)	43 (35.2)	15 (35.7)	58 (35.4)
Race, n (%)				
White	50 (64.9)	86 (70.5)	29 (69.0)	115 (70.1)
Asian	25 (32.5)	21 (17.2)	8 (19.0)	29 (17.7)
Black or African American	1 (1.3)	4 (3.3)	4 (9.5)	8 (4.9)
Other	0	10 (8.2)	1 (2.4)	11 (6.7)
More than one race	0	1 (0.8)	0	1 (0.6)
Unknown	1 (1.3)	0	0	0
Ethnicity, n (%)				
Not Hispanic or Latino	65 (84.4)	109 (89.3)	38 (90.5)	147 (89.6)
Hispanic or Latino	11 (14.3)	12 (9.8)	4 (9.5)	16 (9.8)
Unknown	1 (1.3)	1 (0.8)	0	1 (0.6)
Region, n (%)				
North America	10 (13.0)	27 (22.1)	8 (19.0)	35 (21.3)
Western Europe	36 (46.8)	43 (35.2)	20 (47.6)	63 (38.4)
Rest of the World	31 (40.3)	52 (42.6)	14 (33.3)	66 (40.2)

Abbreviations: max=maximum; min=minimum; SD=standard deviation.

Note: This table was copied from page 54 of the HELIOS-A study report

Table 3 Comparison of Baseline Disease Characteristics

Demographic	APOLLO	HELIOS-A		
	Placebo (N=77)	Vutrisiran (N=122)	Patisiran (N=42)	Total (N=164)
Modified Neuropathy Impairment Score +7 (mNIS+7)				
Mean (SD)	74.61 (37.04)	60.55 (35.99)	57.69 (33.71)	59.82 (35.34)
Median (min, max)	71.50 (11.0, 153.5)	63.50 (2.5, 158.0)	53.44 (7.0, 137.6)	59.94 (2.5, 158.0)
Norfolk Quality of Life-Diabetic Neuropathy (Norfolk QoL-DN) total score				
Mean (SD)	55.5 (24.3)	47.1 (26.3)	47.3 (29.9)	47.2 (27.2)
Median (min, max)	53.5 (8, 111)	44.0 (-1, 105)	41.0 (1, 125)	42.0 (-1, 125)
10-meter Walk Test (10-MWT, m/s)				
Mean (SD)	0.790 (0.319)	1.006 (0.393)	1.011 (0.400)	1.007 (0.393)
Median (min, max)	0.800 (0.00, 1.53)	1.049 (0.08, 1.87)	1.000 (0.11, 1.93)	1.025 (0.08, 1.93)
Modified Body Mass Index (mBMI, kg/m^2)				
Mean (SD)	989.9 (214.2)	1057.5 (234.0)	1060.3 (226.6)	1058.2 (231.4)
Median (min, max)	959.7 (569, 1508)	1047.2 (589, 1723)	1029.1 (646, 1636)	1038.1 (589, 1723)
Rasch-built Overall Disability Scale (R-ODS)				
Mean (SD)	29.8 (10.8)	34.1 (11.0)	34.0 (10.4)	34.1 (10.8)
Median (min, max)	30.5 (3, 48)	35.0 (5, 48)	35.0 (9, 47)	35.0 (5, 48)
Polyneuropathy Disability (PND) Score				
I	20 (26.0)	44 (36.1)	15 (35.7)	59 (36.0)
II	23 (29.9)	50 (41.0)	17 (40.5)	67 (40.9)
IIIA	22 (28.6)	16 (13.1)	7 (16.7)	23 (14.0)
IIIB	11 (14.3)	12 (9.8)	3 (7.1)	15 (9.1)
IV	1 (1.3)	0	0	0
Neuropathy Impairment Score (NIS)				
Mean (SD)	57.02 (32.04)	43.02 (28.63)	43.11 (28.23)	43.04 (28.44)
Median (min, max)	53.88 (7.0, 125.5)	36.00 (5.0, 127.0)	38.00 (5.5, 115.6)	36.50 (5.0, 127.0)
Serum TTR (mg/L)				
Mean (SD)	198.84 (58.08)	206.11 (61.03)	206.47 (65.28)	206.20 (61.94)

Note: This table was copied from page 58 of the HELIOS-A study report

3.2.1.4 Results and Conclusions

3.2.1.4.1 Sponsor's Results

The primary endpoint was the difference between vutrisiran (HELIOS-A) and placebo (APOLLO) treatment in the change from baseline at Month 9 in mNIS+7, analyzed using general ANCOVA/MI methods in the mITT population.
The mNIS+7 is a composite measure of neurologic impairment. The mNIS+7 is scored from 0 (no impairment) to 304 points (maximum impairment). Higher scores represent a greater severity of disease. The mNIS+7 was assessed at baseline and Month 9. A decrease from baseline in mNIS+7 is indicative of an improvement in neuropathy, whereas an increase in mNIS+7 suggests worsening of neuropathy.
At Month 9, the vutrisiran group showed an improvement in neuropathy compared to baseline (least-squares [LS] mean change from baseline: -2.24 points) while the placebo group showed a worsening of neuropathy (LS mean change from baseline: +14.76 points) (Table 4 and Figure 3). This represents a statistically significant improvement in neuropathy at 9 months for patients in the vutrisiran group compared to the placebo group (LS mean difference between groups: -17.00 points, p<0.0001).

Table 4 Primary Analysis of mNIS+7

	APOLLO	HELIOS-A
Statistic*	Placebo (N=77)	Vutrisiran (N=122)
Month 9 LS mean (SE)	14.76 (2.00)	-2.24 (1.43)
95% CI	(10.84, 18.68)	(-5.04, 0.57)
LS mean difference (SE) (vutrisiran – placebo)	-	**-17.00 (2.44)**
95% CI	-	**(-21.78, -12.22)**
p-value	-	**3.542E-12**

Abbreviations: AE=adverse event; ANCOVA= analysis of variance; CI=confidence interval; COVID-19=coronavirus disease 2019; hATTR=hereditary transthyretin-mediated (amyloidosis); LS=least squares; mITT=modified intent-to-treat; mNIS+7=modified neuropathy impairment score +7; SAP=statistical analysis plan; SE=standard error.
Note: mNIS+7 score = mean of 2 nonmissing assessments planned to be performed ≥24 hours to ≤7 days apart at each scheduled visit, after component imputation; Higher scores represent a greater severity of disease (range = 0 to 304). Assessments after initiation of local standard treatment for hATTR amyloidosis or on or after onset of a serious COVID-19 AE excluded from analysis (considered missing before multiple imputation).
* Multiple imputation estimates and p-value derived per combining least squares (LS) estimates per Rubin's rules based on 100 datasets where missing Month 9 values were imputed using a regression procedure including select baseline variables (see SAP). LS estimates derived from analysis of covariance model, controlling for categorical factors (treatment, genotype, age of disease onset) and continuous covariate (baseline value).

Note: copied from page 65 of Sponsor's study report

A descriptive summary of mNIS+7 at baseline and Month 9, including data from the placebo, vutrisiran, and patisiran groups, is provided in Table 5. Overall, the results in the vutrisiran and patisiran groups at Month 9 appear reasonably comparable.

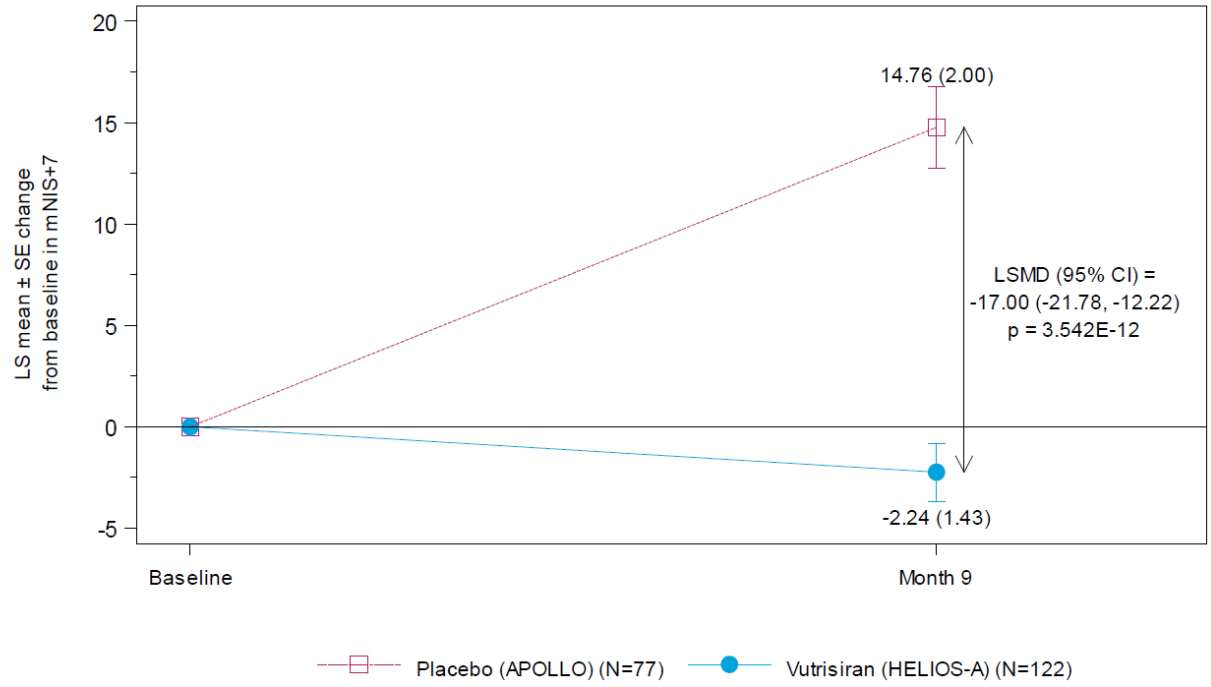
Table 5 mITT Analysis of mNIS+7 (no imputation of missing data)

Visit	Actual/ Change	Statistic	APOLLO	HELIOS-A	
			Placebo (N=77)	Vutrisiran (N=122)	Patisiran (N=42)
Baseline	Actual	n	77	122	42
		Mean (SD)	74.61 (37.04)	60.55 (35.99)	57.69 (33.71)
		Median	71.50	63.50	53.44
		Min, Max	11.0, 153.5	2.5, 158.0	7.0, 137.6
Month 9	Actual	n	67	114	37
		Mean (SD)	90.99 (41.31)	57.50 (37.98)	52.56 (33.25)
		SE	5.05	3.56	5.47
		Median	91.50	57.00	50.25
		Min, Max	19.0, 167.5	1.0, 160.1	6.0, 152.3
	Change from baseline	n	67	114	37
		Mean (SD)	15.22 (17.18)	-1.70 (13.22)	-1.41 (17.23)
		SE	2.10	1.24	2.83
		Median	13.00	-0.44	-2.00
		Min, Max	-16.6, 72.0	-35.0, 26.0	-47.0, 62.9

Abbreviations: AE=adverse event; COVID-19=coronavirus disease 2019; hATTR=hereditary transthyretin-mediated (amyloidosis); max=maximum; min=minimum; mITT=modified intent-to-treat; mNIS+7=modified neuropathy impairment score +7; SD=standard deviation; SE=standard error.
Note: mNIS+7 score = mean of 2 nonmissing assessments planned to be performed ≥24 hours to ≤7 days apart at each scheduled visit, after component imputation; Higher scores represent a greater severity of disease (range = 0 to 304). Assessments after initiation of local standard treatment for hATTR amyloidosis or on or after onset of a serious COVID-19 AE excluded from analysis (considered missing before multiple imputation).

Note: This table was copied from Page 67 of the Sponsor's study report

Figure 3 mNIS+7 Primary Analysis



N evaluable

Placebo	77	67
Vutrisiran	122	114

Note: This figure was copied from page 66 of sponsor's study report

Of the 122 Vutrisiran patients, 114 (93.4%) had evaluable baseline and Month 9 assessments included in the analysis:

- 101 (82.8%) patients completed their Month 9 assessments within the original protocol window (3 months).
- 13 (10.7%) patients completed their Month 9 assessments within the extended window allowed for assessments delayed due to the COVID-19 pandemic (between 3 and 6 months).

8 (6.6%) patients were either missing their Month 9 assessments or had their Month 9 assessment censored:

- 7 (5.7%) patients were missing their Month 9 assessments:
  - 4 (3.3%) patients were missing their Month 9 assessment for reasons related to the COVID-19 pandemic:
    - Patient [redacted] (b) (6) died due to an SAE of COVID-19 pneumonia before their Month 9 assessment.

        - 3 patients ([redacted]) were unable to complete their assessments due to COVID-19 restrictions at the time of data cut-off.

    - 3 (2.5%) patients were missing their Month 9 assessment for reasons unrelated to the COVID-19 pandemic:
        - Patient [redacted] died due to an SAE before their Month 9 assessment (occlusion of common internal and external iliac artery);
        - 2 patients ([redacted]) did not complete their assessments within the original protocol window.
    - 1 (1.3%) patient had their Month 9 assessment censored due to an SAE of COVID-19 pneumonia (Patient [redacted])

Of the 77 patients from the APOLLO placebo group:

- All had baseline assessments
- 67 (87.0%) patients had evaluable baseline and Month 9 assessments (completed within the 3-month protocol window) included in the analysis.

Of the 42 Patisiran patients in HELIOS-A, 37 (88%) had a Month 9 mNIS+7 assessment.

Norfolk QoL-DN, Key Secondary Endpoint, Change from Baseline to Month 9

The Norfolk QoL-DN is standardized QoL questionnaire designed to measure the perception of the effects of polyneuropathy by the patient. The range of possible scores for Norfolk QoL-DN is -4 (best possible quality of life) to 136 (worst possible quality of life). Lower scores indicate a higher quality of life. Norfolk QoL-DN was assessed at baseline and Month 9. A decrease from baseline in Norfolk QoL-DN total score represents improvement in quality of life, and an increase from baseline in total score represents worsening.

At Month 9, the vutrisiran group showed an improvement in quality of life compared to baseline (LS mean change from baseline: -3.3 points) while the placebo group showed a worsening in quality of life (LS mean change from baseline: +12.9 points) (Table 6). This represents a statistically significant improvement in quality of life at 9 months for patients in the vutrisiran group compared to the placebo group (LS mean difference between groups: -16.2 points, $p<0.0001$).

Table 6 ITT analysis of Norfolk QoL-DN with multiple imputation

Statistic*	APOLLO Placebo (N=77)	HELIOS-A Vutrisiran (N=122)
Month 9 LS mean (SE)	12.9 (2.2)	-3.3 (1.7)
95% CI	(8.5, 17.3)	(-6.6, -0.1)
LS mean difference (SE) (vutrisiran – placebo)	-	-16.2 (2.8)
95% CI	-	(-21.7, -10.8)
p-value	-	5.426E-09

Abbreviations: AE=adverse event; ANCOVA= analysis of variance; CI=confidence interval; COVID-19=coronavirus disease 2019; hATTR=hereditary transthyretin-mediated (amyloidosis); LS=least squares; mITT=modified intent-to-treat; NIS=neuropathy impairment score; Norfolk QoL-DN=Norfolk Quality of Life-Diabetic Neuropathy; SAP=statistical analysis plan; SE=standard error.
Notes: Norfolk QoL-DN score = composite of 35 quality of life questions; lower scores indicate higher quality of life (range = -4 to 136). Nonmissing scores for items 27-33 imputed for patients who reported an impact due to COVID-19 pandemic (see SAP). Assessments after initiation of local standard treatment for hATTR amyloidosis or on or after onset of a serious COVID-19 AE excluded from analysis (considered missing before multiple imputation).
* Multiple imputation estimates and p-value derived per combining least squares (LS) estimates per Rubin's rules based on 100 datasets where missing Month 9 values were imputed using a regression procedure including select baseline variables (see SAP). LS estimates derived from analysis of covariance model, controlling for categorical factors (treatment, genotype, age of disease onset, baseline NIS) and continuous covariate (baseline value).

Note: Table copied from page 74 of sponsor's study report

A descriptive summary of Norfolk QoL-DN at baseline and Month 9, including data from the placebo, vutrisiran, and patisiran groups, is provided in Table 7. Overall, the results in the vutrisiran and patisiran groups at Month 9 appear reasonably comparable.

Table 7 mITT analysis of Norfolk QoL-DN

Visit	Actual/ Change	Statistic	APOLLO Placebo (N=77)	HELIOS-A Vutrisiran (N=122)	 Patisiran (N=42)
Baseline	Actual	n	76	121	42
		Mean (SD)	55.5 (24.3)	47.1 (26.3)	47.3 (29.9)
		Median	53.5	44.0	41.0
		Min, Max	8, 111	-1, 105	1, 125
Month 9	Actual	n	66	115	38
		Mean (SD)	66.2 (27.6)	41.8 (26.6)	44.2 (27.1)
		SE	3.4	2.5	4.4
		Median	68.0	40.0	39.0
		Min, Max	5, 109	-4, 102	1, 115

Note: Table copied from page 75 of sponsor's study report

Impact of COVID-19 on Norfolk QoL-DN

The potential impact of visit delays related to COVID-19 restrictions on Norfolk QoL-DN was evaluated using the Month 9 Efficacy PP Population. The results with this analysis are consistent with the primary analysis. As such, the data does not suggest that the COVID-19 pandemic had an impact on the analysis of Norfolk QoL-DN scores at Month 9.

In addition, due to the potential for the COVID-19 pandemic to impact aspects of quality of life in multiple ways (e.g., infection, anxiety and stress from the pandemic, the potential for loss of employment, and the disruptions in physical activity and social interactions due to social distancing and the closure of public gathering places), additional information on specific impacts on quality of life associated with the COVID-19 pandemic were collected.

For the primary analysis, the derivation of Norfolk QoL-DN total score and Physical Functioning/Large Fiber domains were modified for patients reporting any impacts on quality of life due to COVID-19. Full details of these modifications are presented in the SAP. Following protocol amendment 3, patients will be instructed to respond to 3 questions on the impact of the COVID-19 pandemic on work and regular daily activities, social activities, and general health. For the primary analysis of Norfolk QoL-DN, for assessments associated with a COVID-19 pandemic impact questionnaire among patients who indicate they are impacted by the pandemic, nonmissing values to items 27-33 will be replaced with the respective average item score (mean of nonmissing item scores) by treatment group among patients whose item scores are not impacted by the pandemic as follows:

- Replace items 27-30 if COVID-19 pandemic impact question 1 response = Yes
- Replace items 31-32 if COVID-19 pandemic impact question 2 response = Yes
- Replace item 33 if COVID-19 pandemic impact question 3 response = Yes

Following item-level replacement due to the COVID-19 pandemic, domain scores are calculated as the average scores of nonmissing included items multiplied by the number of items, rounded to the nearest integer, if at least 50% of the items are nonmissing. A domain score is missing if more than 50% of the included items are missing.

Month 9 Patient Data Received After the Data Cut-off

Despite protocol modifications and risk mitigation steps related to the COVID-19 pandemic, 2 vutrisiran patients ([redacted]) and 4 patisiran patients ([redacted]) were unable to complete their full Month 9 efficacy visit assessments by the time of the data cutoff for the sponsor's study report. In the context of ongoing uncertainties related to the COVID-19 pandemic and when these assessments would be able to be performed due to regional lockdowns and patient restrictions, the database was locked as planned. The patients subsequently completed their Month 9 efficacy assessments within the extended 6-month window allowed for the pandemic.

The two affected Vutrisiran patients ([redacted]) had mNIS+7 changes from baseline of -0.5 and +19.3, respectively, after allowing the extended Month 9 window for the Month 9 assessment.

Table 8 Impact of COVID-19 Pandemic on HELIOS-A

Visit Impact	HELIOS-A		
	Vutrisiran (N=122)	Patisiran (N=42)	Total (N=164)
Patients impacted[a], n (%)	87 (71.3)	20 (47.6)	107 (65.2)
Patients with missed, delayed, or partially completed visit[a], n (%)	87 (71.3)	20 (47.6)	107 (65.2)
Patients with any location change[a], n (%)	82 (67.2)	15 (35.7)	97 (59.1)
Patients with missed or delayed doses[a], n (%)	14 (11.5)	18 (42.9)	32 (19.5)

Abbreviations: COVID-19=coronavirus disease 2019.
[a] Patients whose participation in the study was impacted by the COVID-19 pandemic (eg, missed, delayed or partially completed visit, missed/delayed study drug dose, or visit location change such as phone visit, etc.).
Source: Table 14.5.1.1

Note: This table was copied from page 63 of the sponsor's study report

Multiple Comparisons/Multiplicity

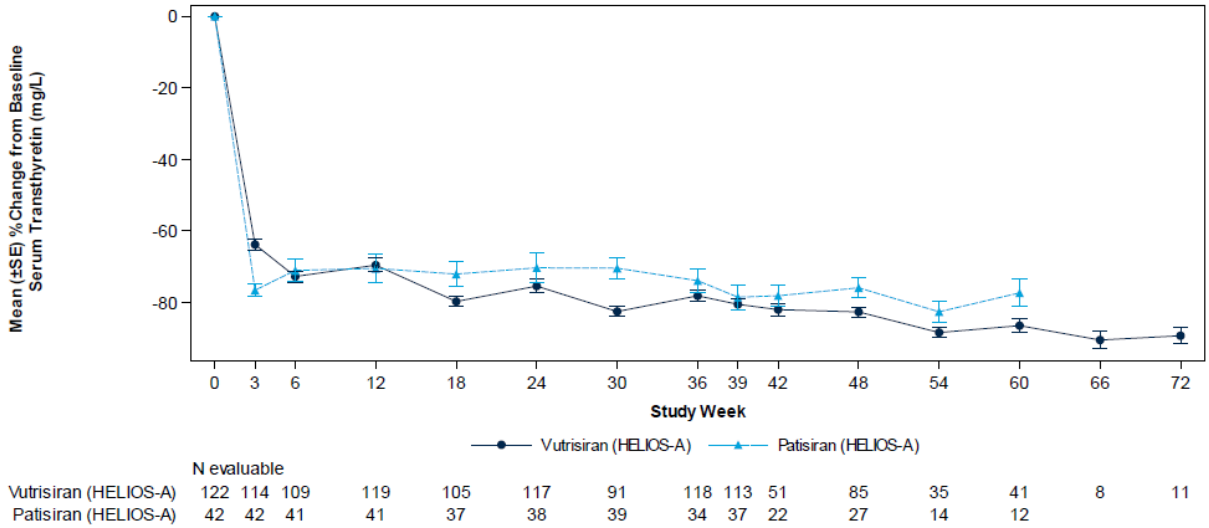
Type I error control for secondary endpoints was achieved by a hierarchical ordering procedure. Endpoints were tested in the following pre-specified hierarchy, as outlined in the SAP if the primary endpoint was significant at a 2-sided 0.05 significance level:

- Norfolk QoL-DN total score change from baseline at Month 9
- 10 MWT gait speed change from baseline at Month 9

Serum Transthyretin

At baseline, the median TTR level in the vutrisiran group was 203.2 mg/L (range, 58.4 to 343.2 mg/L and the median TTR level in the patisiran group was 203.9 mg/L (range, 71.0 to 353.2 mg/L). TTR reduction was similar between vutrisiran and patisiran as seen in Figure 4.

Figure 4 Serum Transthyretin comparison between Vutrisiran and Patisiran (HELIOS-A)



Abbreviations: mITT=modified intent-to-treat; SE=standard error; TTR=transthyretin.
Note: Month 9 and Month 18 nontrough TTR assessments presented at Weeks 39 and 81, respectively.

Note: Figure copied from Figure 14 on page 112 of the Sponsor's study report

Reviewer's Comment: The reviewer verified that the Vutrisiran group at Month 9 was non-inferior to Patisiran with respect to the prespecified 10 % margin using a 90% confidence interval. However, whether there was adequate prespecified clinical justification for the choice of a 10% margin is not clear. Nevertheless, this may give some support to the interpretability of the study (the external control comparison is questionably interpretable in light of the baseline differences across studies).

3.2.1.4.2 Reviewer's Results

The groups are not very comparable at baseline across studies APOLLO and HELIOS-A on the primary outcome measure, mNIS+7 (among other baseline characteristics as mentioned above). For reference, the standard error of the difference between Vutrisiran-HELIOS-A and Placebo APOLLO is estimated as 5.30 and standard error of the difference of the two Patisiran groups is estimated as 6.98, thus the cross study baseline mean differences are sizeable, even after accounting for their standard error (Table 9).

Table 9 Comparison of Baseline mNIS+7 between current (HELIOS-A) and external control (APOLLO) studies

Study	APOLLO		HELIOS-A	
Group	Placebo(N=77)	Patisiran(N=148)	Patisiran (N=42)	Vutrisiran (N=122)
Baseline mNIS+7 Mean(SD)	74.61(37.04)	80.9(41.5)	57.69(33.71)	60.55(35.99)

Based on the information in Table 9 the Placebo Vutrisiran pooled variance estimate would be ((77-1)*37.04**2+ (122-1)*35.99**2)/(122+77-2) and the Patisiran(APOLLO) Patisiran(HELIOS) pooled variance estimate would be ((148-1)*41.5**2+ (42-1)*33.71**2)/(148+42-2).

The 95% confidence interval for the difference of Patisiran - Placebo in APOLLO at Month 9 was -20.7, -11.27. The corresponding LSMean difference was -15.98(2.39). The Month 9 LSMean for Patisiran (APOLLO) [n=141] was -2.04 (S.E.=1.50) compared to 13.95 (S.E.=2.10) for Placebo (APOLLO). In HELIOS-A the Patisiran -Vutrisiran difference was -0.06 (S.E.=2.75).

Therefore, an exploratory noninferiority test upper 95% confidence limit would be (-15.98 - -0.06) +1.96*sqrt(2.39^2 + 2.75^2)=-8.77. However, formal non-inferiority testing would have required prespecification of the appropriate margin for the primary endpoint, mNIS+7, and the assumption of constancy across trials. For the sake of completeness the LSMean for Patisiran at 9 Months in Helios was -2.91 (S.E.=2.94) and the estimated LS Mean difference between Patisiran (HELIOS-A) and Placebo (APOLLO) was -16.77 (S.E.=3.59).

The APOLLO placebo and HELIOS-A Vutrisiran arms seem to have several other differences in patient characteristics as well. For example, Race proportions are different (32 vs. 17% Asian, genotype: 52 vs 44% V30M, 75 vs. 65% Male, 74 vs 61 NIS ≥50, Cardiac subpopulation 47% vs. 29%). Comparison to an external control with such clear differences in patient composition is not likely to be reliable. Vutrisiran was not statistically superior to the concurrent control arm Patisiran, but rather they appeared reasonably similar in terms of change from baseline to Month 9 in mNIS+7.

One hundred and sixty four (164) patients were randomized to either Vutrisiran (122 patients) or Patisiran (42 patients) in HELIOS-A. All randomized patients were treated with study drug. Patients were randomized and treated at 57 study centers in 22 countries. Countries that randomized ≥10 patients included the US (33 patients [20.1%]), France (20 patients [12.2%]), Bulgaria (13 patients [7.9%]), Portugal (13 patients [7.9%]), and Australia (12 patients [7.3%]). Of the 122 patients in the Vutrisiran group, 3 (2.5%) patients discontinued study drug prior to Month 9: 2 study drug discontinuations were due to death of the patient unrelated to study treatment (both Vutrisiran deaths were before month 9 and so are missing the month 9 mnis+7) and one study drug discontinuation was due to physician decision for a patient who did not comply with study visits and was considered lost to follow up. These same 3 patients stopped study participation.
There were two deaths in the Vutrisiran group, 1 loss to follow up, and 2 missing assessments attributed to the COVID-19 pandemic. Note that there had also been two deaths in the external APOLLO placebo.

Even if the 13 delayed Month 9 Vutrisiran assessments that were delayed between 3 and 6 months due to COVID-19 pandemic issues were imputed with poor outcomes (i.e., a mNIS+7 change of +20) it would not alter the rejection of the null hypothesis, i.e., superiority of Vutrisiran to the APOLLO placebo would still hold.

If the death rate is not very small or small but all in one arm, as in this trial, then an analysis ignoring deaths may be misleading and biased. In this case an alternative approach which takes deaths into account assigns the worst rank for deaths in the primary analysis when the primary endpoint is a function scale (“joint rank analysis” which is typically the recommended primary analysis in ALS). This analysis approach may be more reasonable because it’s not appropriate to equate death to a specific level on a functional scale, e.g., missing functional data caused by death is not meaningful, i.e., is obviously important and shouldn’t simply be treated like other missing data.

A joint rank test (joint rank of survival time and mNIS7+ changes given survival) analysis was conducted by this reviewer as a sensitivity analysis to avoid the bias of treating deaths missing mNIS7+ changes as ignorable (or equivalent to survivors non-missing changes). This analysis (N=186 t=6.11, p<.0001) also suggests that Vutrisiran (HELIOS-A) is superior to the external placebo (APOLLO) [however, note that again we are ignoring the group baseline differences here which could potentially invalidate the comparison].

Note that there were relatively high rates of patients missing mNIS+7 Heat Pain component items at Month 9. These missing item percentages were also considerable at baseline (~40%) and not too different between groups. Overall, the high proportions of certain mNIS+7 items being missing is not too concerning since there was a prespecified imputation plan and the groups were comparable on almost all specific items. Missing items was less of an issue for the Norfolk QOL-DN efficacy measure.

3.3 Evaluation of Safety

Safety in general is not addressed in this review. Please see the Clinical safety review. In the Vutrisiran group, SAEs were reported for 17.2% of patients; 2 patients had SAEs considered related to study drug. AEs leading to treatment discontinuation were reported for 2 patients (1.6%), which were both fatal, and neither death was considered related to study drug by the sponsor.

In the Patisiran group, 95.2% of patients had at least 1 AE, 40.5% of patients had SAEs, and 7.1% of patients had AEs leading to treatment discontinuation. Three patients in the Patisiran group died during the study (7.1%); the deaths were considered not related to study drug by the sponsor.

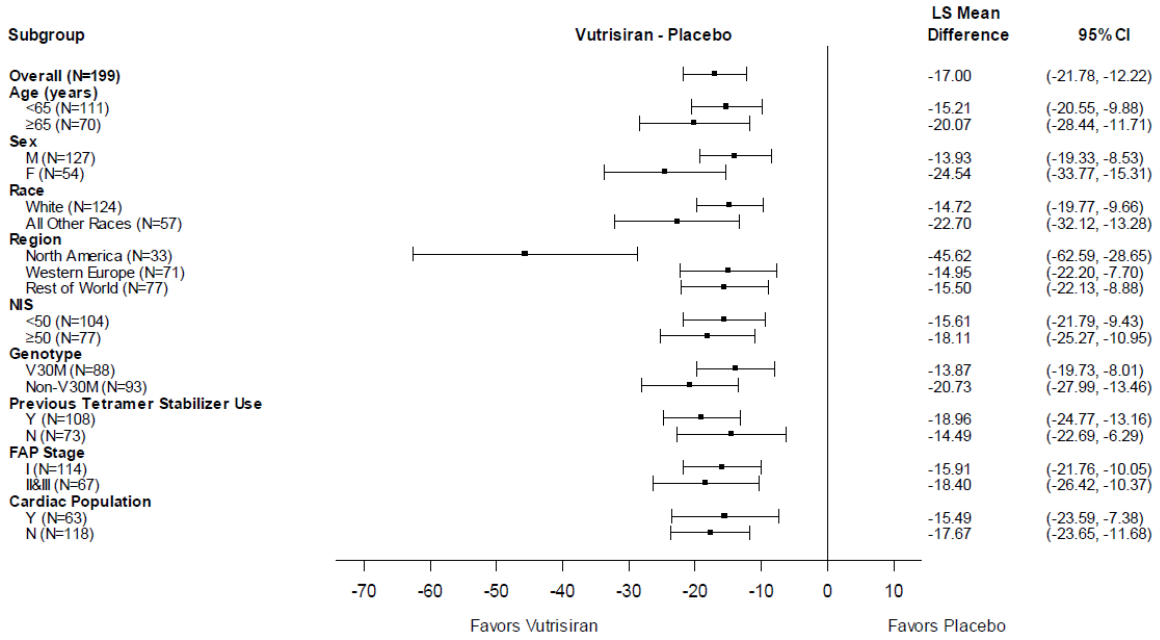
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The demographics of the comparison groups were: 69% were Male [65% Vutrisiran, 75% Placebo (APOLLO)], 68% were classified as White [70% Vutrisiran and 65% Placebo (APOLLO)], 23% as Asian [17% Vutrisiran and 32% Placebo (APOLLO)]and 9% as Other, 68% were Age<65 [62% for Vutrisiran and 57% for Placebo (APOLLO)]and 32% were Age≥65. For the most part, patients receiving Vutrisiran experienced similar improvements relative to placebo in mNIS+7 across all subgroups including age, sex, race, region, baseline NIS score, genotype, previous tetramer stabilizer use, disease stage, and cardiac subpopulation (Figure 5). However, there was some evidence of a smaller effect in Males than Females, but still there appears to be a nominally significant effect in Males compared to the APOLLO placebo

(quantitative Sex interaction test p=.0737). There was also significant variation between regions with North America appearing to have a bigger treatment difference than the other two regions.

Figure 5 Sponsor's Subgroup Analyses



Abbreviations: ANCOVA=analysis of variance; CI=confidence interval; FAP=Familial Amyloid Polyneuropathy; LS=least squares; mITT=modified intent-to-treat; mNIS+7=modified neuropathy impairment score +7; NIS=neuropathy impairment score; V30M=valine to methionine variant at amino acid 30.
Source: Figure 14.2.3.1.1

Note: This figure was copied from page 102 of the sponsor's study report

4.1.1.1 Individual Sites

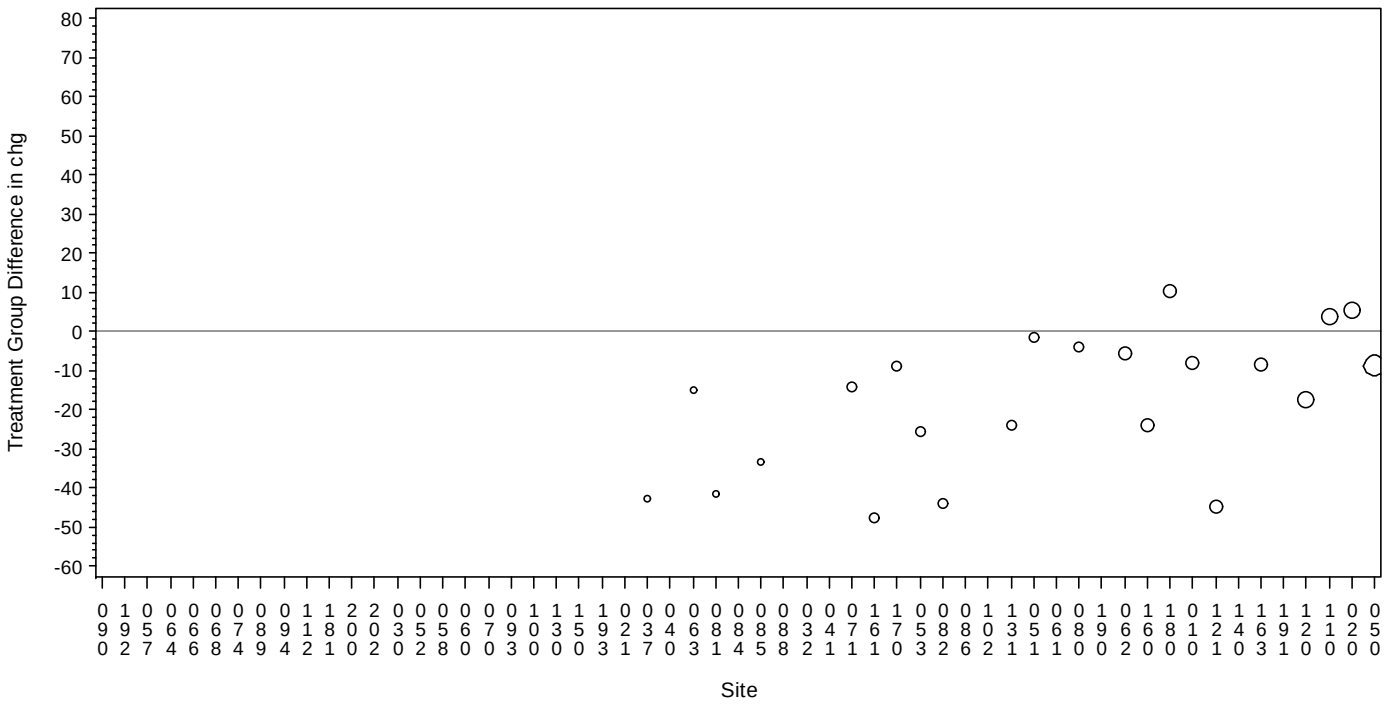
Investigation of Regional and Site effects

This study was conducted at 63 study centers (56 had at least one patient for Patisiran or Vutrisiran). Only 25 sites had patients for both Placebo (APOLLO) and Vutrisiran (HELIOS-A) and only 18 had both Patisiran (HELIOS-A) and Vutrisiran(HELIOS-A), but the HELIOS-A randomization was 1:3 and most sites had small sample sizes. The average total sample size was 4 between Placebo-APOLLO and Vutrisiran-HELIOS-A (average of 2 Vutrisiran with a maximum of 9 and 0.667 Patisiran with a maximum of 4). Analyses by site were not performed by the sponsor.

Figure 6 shows Vutrisiran (HELIOS-A) minus Placebo(APOLLO) LS Mean differences in mNIS+7 change from baseline at Month 9 by individual sites. As indicated by the preponderance

of sites with negative differences on the figure, the treatment difference was reasonably consistent across sites (note that due to the use of an external control the Vutrisiran and Placebo were not concurrent at the same site). No single sites appear to be excessively driving the overall result. The primary analysis result is rather insensitive to individual site exclusions. The largest site in the US was site 086 (Baltimore, N=5 Vutrisiran).

Figure 6 Primary Efficacy Comparison by Individual Sites



4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

5.2 Collective Evidence

Collective evidence is not considered in this review since there was only one trial.

5.3 Conclusions and Recommendations

The APOLLO placebo and HELIOS-A Vutrisiran arms seem to have several differences in patient characteristics. For example, Race proportions are different (32 vs. 17% Asian, genotype: 52 vs 44% V30M, 75 vs. 65% Male, 74 vs 61 NIS $\geq$50, Cardiac subpopulation). Comparison to an external control with such clear differences in patient composition is not likely to be reliable. It is not clear how best to account for the bias that might be attributed to these baseline differences between the investigation drug arm and the historical placebo. Vutrisiran was not statistically superior to the concurrent control arm Patisiran, but rather they appeared similar in terms of change from baseline to Month 9 in mNIS+7. Considering the large effect Patisiran displayed in APOLLO compared to the APOLLO placebo and the apparent large effect of Vutrisiran (HELIOS-A) to the same external APOLLO placebo and similarity to Patisiran in HELIOS-A may allay these concerns about baseline differences and lack of comparability. Therefore, although only one controlled efficacy study was submitted in this application for this drug in this rare and serious disease with an unmet need for effective treatments the study results seem sufficiently compelling.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

TRISTAN S MASSIE
02/14/2022 03:22:00 PM

KUN JIN
02/14/2022 03:52:22 PM
I concur with the review.

HSIEN MING J HUNG
02/14/2022 03:57:02 PM