Use of External Controls for the Primary Analysis of a Randomized Phase 3 Study to Evaluate a Higher Dose Regimen of Nusinersen for the Treatment of Spinal Muscular Atrophy (SMA)

Richard Foster*

Biogen

* Joint work with Peng Sun, Yingying Liu, Stephanie Fradette, Rachel Farewell (All Biogen), Bo Lu (Ohio State University), and Ming-Hui Chen (University of Connecticut)

Disclosures

Richard Foster is an employee of Biogen and holds stock and/or stock options in Biogen.

Outline

- Background:
  - Nusinersen for the treatment of SMA
  - The DEVOTE Study: A global 3-part (A, B, and C), Phase 2/3 study to evaluate a higher dose regimen of nusinersen administered intrathecally
- Statistical approaches for historical data borrowing in the DEVOTE study
- Results from the DEVOTE study
- Conclusion

Nusinersen for the Treatment of SMA

- SMA: a rare neuromuscular disorder that results in loss of motor neurons and progressive muscle wasting
- Nusinersen:
  - Since the December 2016 FDA approval, it became the first approved drug for the treatment of SMA
  - Nusinersen has shown clinically meaningful and sustained efficacy across the SMA spectrum, with a well-established safety profile of the 12 mg regimen from over 10 years of study[1–3]
- The ENDEAR study:
  - Completed pivotal study in infantile onset SMA patients
  - Patients were randomized to 12 mg regimen of nusinersen and sham (untreated)
  - Large treatment was observed comparing 12 mg regimen of nusinersen with sham (untreated)
  - Basis of initial FDA approval of nusinersen in 2016

1. Finkel RS, *et al. N Engl J Med.* 2017;377(18):1723–1732. 2. Szabó L, *et al. Eur J Paediatr Neurol.* 2020;27:37–42. 3. Maggi L, *et al. J Neurol Neurosurg Psychiatry.* 2020;91:1166–1174.

DEVOTE Study Design

Key Inclusion/Exclusion Criteria

- Genetic documentation of 5q SMA
- Later-onset SMA
- Age 2 to ≤ 15 years, inclusive, at the time of informed consent

PART A: Nusinersen 28mg/28mg safety group *(n = 6)*

Key Inclusion/Exclusion Criteria

- Genetic documentation of 5q SMA

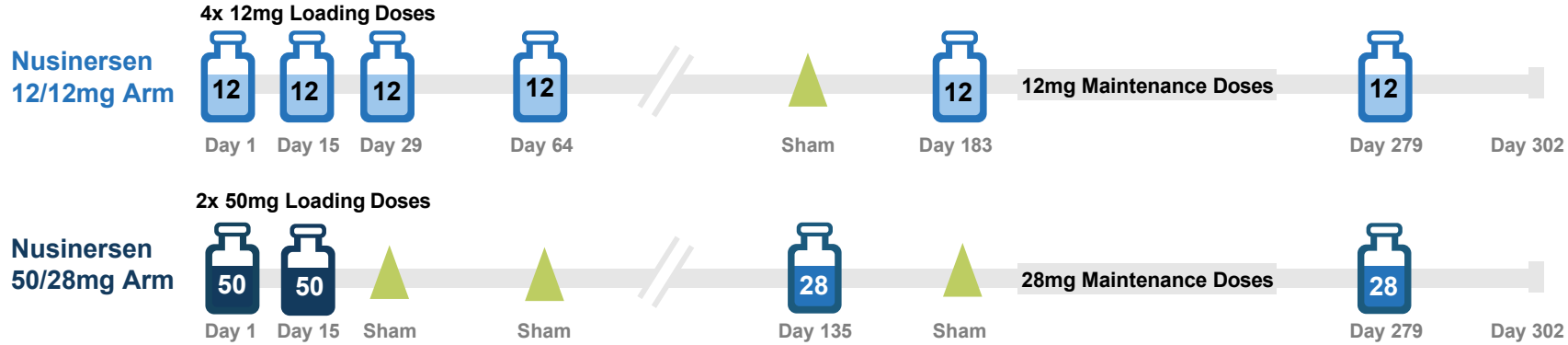
Infantile-onset cohort

- ≤7 months of age at informed consent
- 2 copies of SMN 2 gene

Later-onset cohort

- Age 2 to <10 at time of informed consent
- Can sit independently but has never had the ability to walk independently
- HFMSE score ≥10 and ≤54 at screening

PART B: Randomized (1:2), 2 Cohorts: infantile-onset *(n = 75)*, Later Onset *(n=24)*



Key Inclusion/Exclusion Criteria

- Prior 12mg nusinersen treatment for ≥1 year
- Any age
- Ambulatory or non ambulatory

PART C: 12mg to 50/28mg Transition *(n = 40)*

Pivotal: Part B infantile onset cohort

Pivotal Infantile Cohort in DEVOTE

DEVOTE pivotal infantile cohort: 75 subjects 2:1 to higher dose regimen (N = 50) and 12 mg regimen (N = 25)

- Population: Treatment naïve (challenging to enroll with 3 drugs on the market: 12 mg nusinersen, Zolgensma, and Risdiplam)
- Inclusion/exclusion criteria + schedule: Consistent with those in the completed Phase 3 study ENDEAR.

Historical data from ENDEAR:

- N = 37: sham (untreated) subjects
- N = 49: 12 mg subjects

Primary endpoint: Change from baseline in CHOP-INTEND in infants who received higher dose regimen as compared to a prespecified matched sham (untreated control group from the ENDEAR study)

CHOP-INTEND = Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders

Key Consideration [1]

Subjects were randomized to higher dose regimen and 12 mg regimen in DEVOTE, though the primary analysis compares to matched sham patients

Nusinersen 12 mg regimen is already highly efficacious:

- Uncertain on effect size of higher dose regimen on top of 12 mg regimen
- Not feasible to conduct an adequately powered study given difficulty in enrolling patients

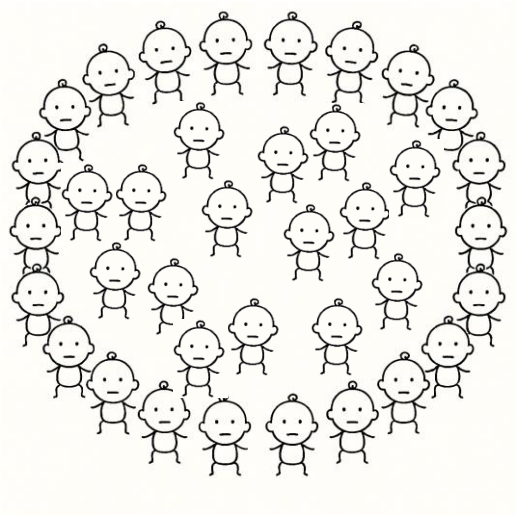
Key Considerations [2]

- For primary analysis, the high dose arm in DEVOTE is viewed as a hypothetical arm in ENDEAR:
  - In a hypothetical study with two active doses, both active doses will be compared to the sham (untreated) arm for hypothesis testing purposes
  - Unmeasured confounding effect would not alter the highly progressive nature of infantile onset patients. Effect size of higher dose regimen versus sham (untreated) is expected to be large and valid conclusion will still be drawn even if unmeasured confounding effect is present.
- It is valuable to have internal 12 mg control data since the effect size is expected to be smaller.

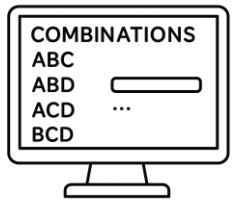
Identifying Matched Sham (Primary Analysis)

Exhaustive search algorithm:

N=37 Endear untreated subjects



In the range of N = 37 to N = 20 ENDEAR sham (untreated) subjects, identify the largest subgroup such that the absolute standardized mean difference (ASMD) between the subgroup and DEVOTE high dose is <= 0.3 for two key covariates



If N reaches 20 and yields no matched sample, then keep N = 20 and implement the following adjustment in sequential order:

- Loosen the criteria to be ASMD <= 0.35
- Loosen the criteria to be ASMD <= 0.4
- Smallest robust Mahalanobis distance

Outcome: At N = 20, a matched sham group satisfying the ASMD criteria of <=0.3 for two key covariates were identified.

Matched Sham control arm

Conditional Borrowing of ENDEAR 12 mg Regimen Data as a Supplementary Analysis

Conditional borrowing algorithm:

- In the range of N = 49 to N = 10 ENDEAR standard dose subjects, first match on two key baseline covariates using the same exhaustive search algorithm.
- A matched set is accepted provided it meets the pre-specified criteria comparing to the efficacy of DEVOTE 12 mg regimen arm

Some notes:

- The efficacy comparability criteria aims to control unobserved confounding effect.
- The conditional borrowing approach has been published *

Outcome: No matched group from the 12 mg ENDEAR arm could be identified that satisfies the pre-specified criteria

* Liu, Yingying, et al. "Matching design for augmenting the control arm of a randomized controlled trial using real-world data." *Journal of Biopharmaceutical Statistics* 32.1 (2022): 124-140.

Hierarchical Testing

- Statistical comparisons for the primary and key secondary endpoints were performed sequentially with each successive analysis requiring statistical significance of the preceding comparison (see table below)
- Once an endpoint in the hierarchy was found to be not statistically significant, subsequent endpoints with p-values < 0.05 were reported as nominally significant

Endpoint	Time Point (Study Day)	Comparator (DEVOTE 50/28mg Vs.)
Primary endpoint:		
CHOP-INTEND	183	*ENDEAR matched sham*
Key secondary endpoints:		
HINE-2 (responder,[a] change)	183	*ENDEAR matched sham*
Plasma NfL	183	*ENDEAR matched sham*
CHOP-INTEND	302	*DEVOTE 12/12mg*
HINE-2 (change)	302	*DEVOTE 12/12mg*
Plasma NfL	64	*DEVOTE 12/12mg*
EFS	NA	*ENDEAR matched sham*
OS	NA	*ENDEAR matched sham*
EFS	NA	*DEVOTE 12/12mg*
OS	NA	*DEVOTE 12/12mg*

- CHOP-INTEND = Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders; EFS = event-free survival; HINE-2 = = Hammersmith Infant Neurological Exam section 2; OS = overall survival; NfL= neurofilament light chain
- Descriptive statistics were calculated for additional secondary endpoints comparing 50/28mg to DEVOTE 12/12mg, including number and duration of hospitalizations and number of serious respiratory events.
- [a]HINE-2 responders were defined using the 7 motor milestone categories (excluding voluntary grasp) as follows: (1) the participant shows at least 2-point improvement in the ability to kick or an increase to the maximum score in this category (touching toes). Alternatively, the participant shows a 1-point improvement in any of the following categories: head control, rolling, sitting, crawling, standing, or walking, AND (2) across the 7 motor milestone categories (excluding voluntary grasp), the participant demonstrates improvement in more categories than worsening.

Highlight of Results from DEVOTE

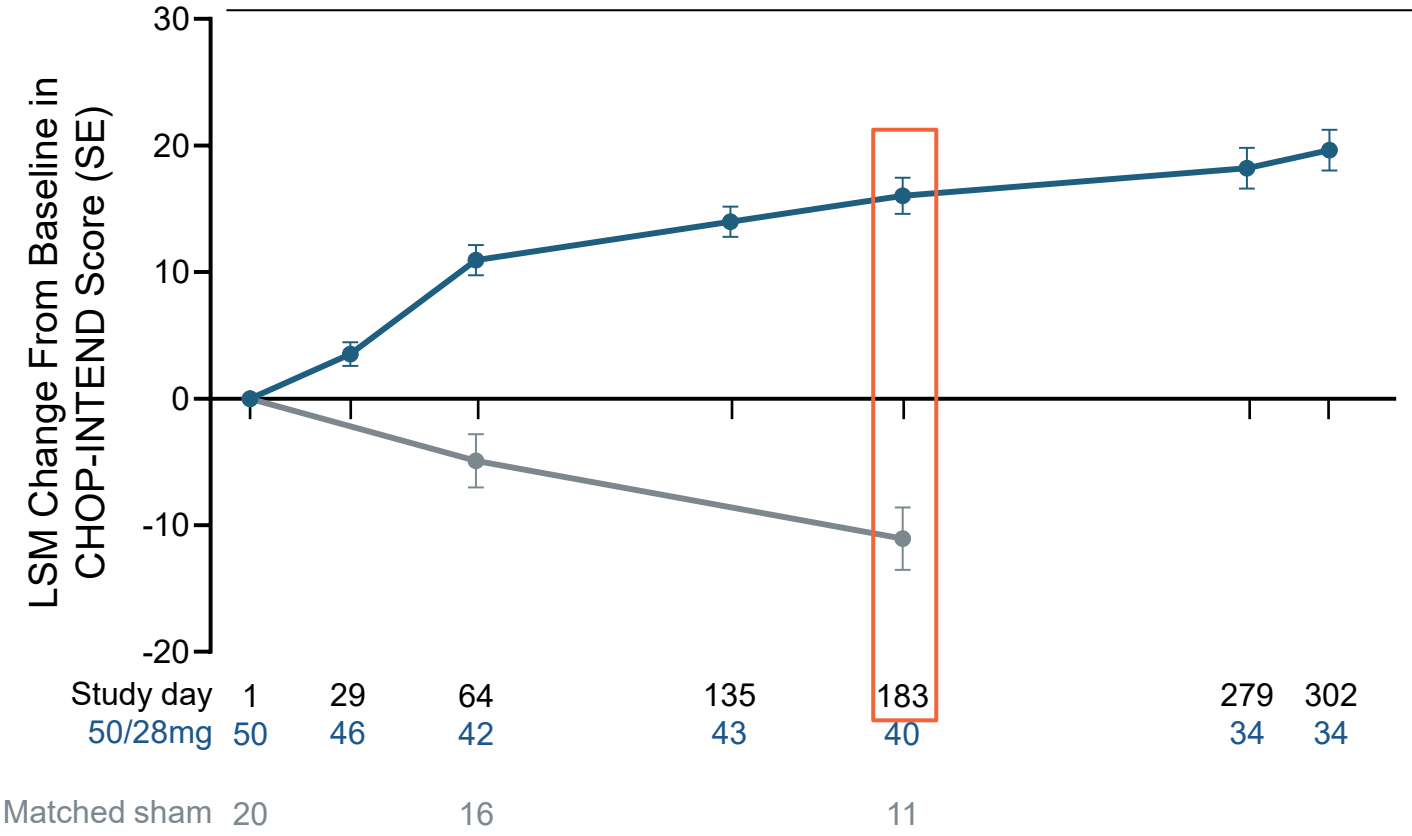
The study met its primary endpoint at six months, achieving a statistically significant improvement in motor function in infants who received the higher dose regimen as compared to a prespecified matched sham (untreated) control group from the ENDEAR study.

- Results favored the higher dose regimen relative to sham across secondary endpoints and trended in favor of the higher dose regimen over the currently approved 12 mg regimen on key biomarker and efficacy measures.
- The higher dose regimen was generally well tolerated, with reported adverse events generally consistent with SMA and the known safety profile of nusinersen. The percentage of serious adverse events was lower in the higher dose regimen (60%; 30) as compared to the 12 mg group (72%; 18).
- Global submissions of this investigational higher dose regimen are currently ongoing

Change in CHOP-INTEND
(Part B Infantile-Onset) – Primary Endpoint

- Scores improved by 15.1 points from baseline at Day 183 in the 50/28mg group vs. a decline of 11.1 points in the sham group

Time Point	LSM Change		LSM Difference (95% CI); p-Value
Day 183	***50/28mg:*** +15.1	***Sham:*** −11.1	26.19 (20.7, 31.7); **p < 0.0001** (JRT)

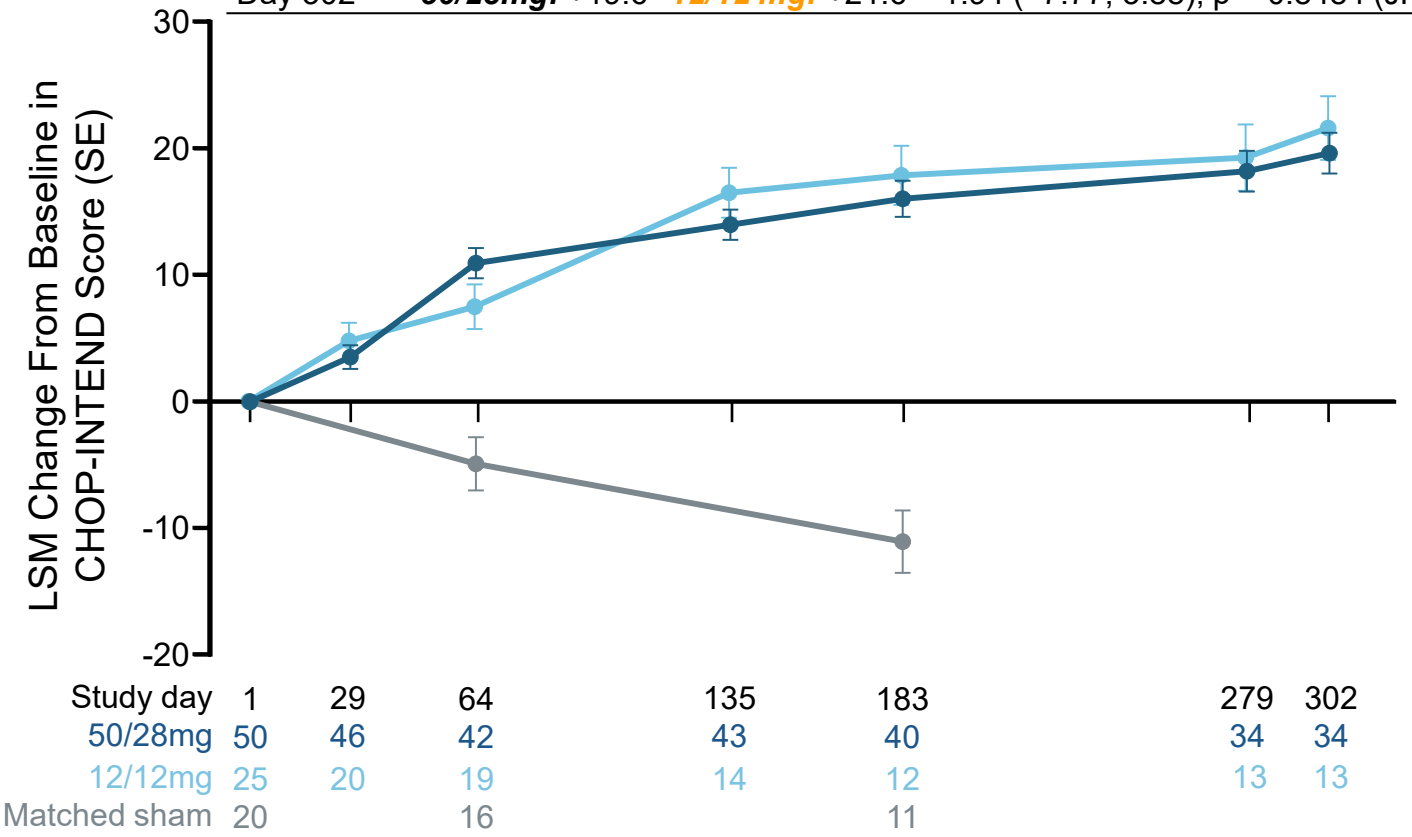


- ANCOVA = analysis of covariance; CHOP-INTEND = The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders; CI = confidence interval; JRT = joint rank test; LSM = least-squares mean; SE, standard error
Study Day 1 is baseline. Results shown are from multiple imputation and an ANCOVA model with adjustment for participants' disease duration and baseline CHOP-INTEND score. The comparisons between 50/28mg and 12/12mg groups, and between 50/28mg and matched sham groups were performed as separate analyses. LSM difference from ANCOVA; p-value from JRT

Change in CHOP-INTEND
(Part B Infantile-Onset)

- At Day 302, mean improvement was higher in the 12/12mg group vs. the 50/28mg group

Time Point	LSM Change		LSM Difference (95% CI); p-Value
Day 183	***50/28mg:*** +15.1	***Sham:*** −11.1	26.19 (20.7, 31.7); **p < 0.0001** (JRT)
Day 302	***50/28mg:*** +19.6	***12/12 mg:*** +21.6	−1.94 (−7.77, 3.88); p = 0.8484 (JRT)

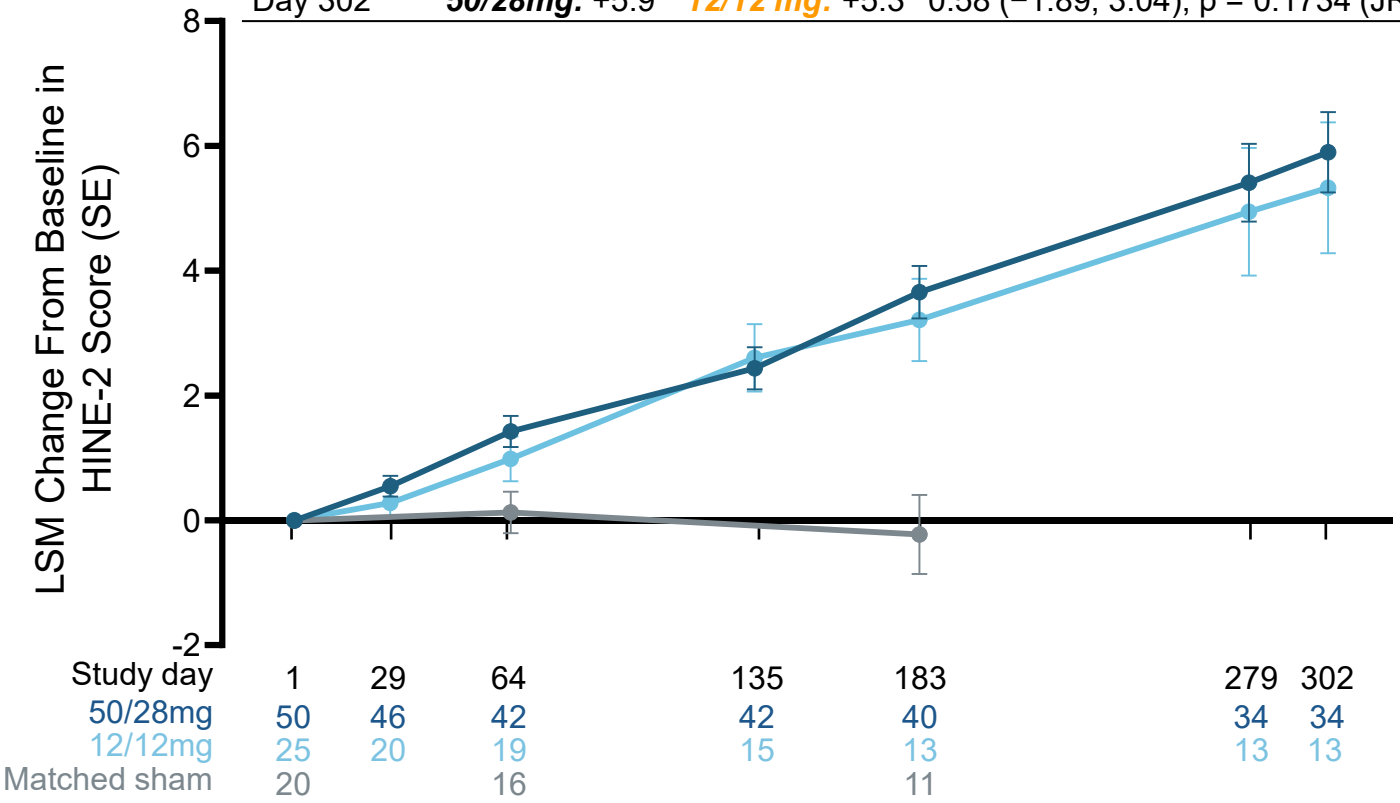


- ANCOVA = analysis of covariance; CHOP-INTEND = The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders; CI = confidence interval; JRT = joint rank test; LSM = least-squares mean; SE, standard error Study Day 1 is baseline. Results shown are from multiple imputation and an ANCOVA model with adjustment for participants' disease duration and baseline CHOP-INTEND score. The comparisons between 50/28mg and 12/12mg groups, and between 50/28mg and matched sham groups were performed as separate analyses. LSM difference from ANCOVA; p-value from JRT

Change in HINE-2
(Part B Infantile-Onset)

- A significantly greater proportion of participants receiving 50/28mg nusinersen vs. sham met the definition of a HINE-2 responder at Day 183 (58% vs. 0%; p < 0.0001; data not shown)
- The 50/28mg group had significantly greater improvements in change in HINE-2 score from baseline at Day 183 vs the sham group
- At Day 302, mean improvement was higher in the 50/28mg group vs. the 12/12mg group

Time Point	LSM Change		LSM Difference (95% CI); p-Value
Day 183	***50/28mg:*** +3.7	***Sham:*** −0.2	3.9 (2.5, 5.4); **p < 0.0001** (JRT)
Day 302	***50/28mg:*** +5.9	***12/12 mg:*** +5.3	0.58 (−1.89, 3.04); p = 0.1734 (JRT)

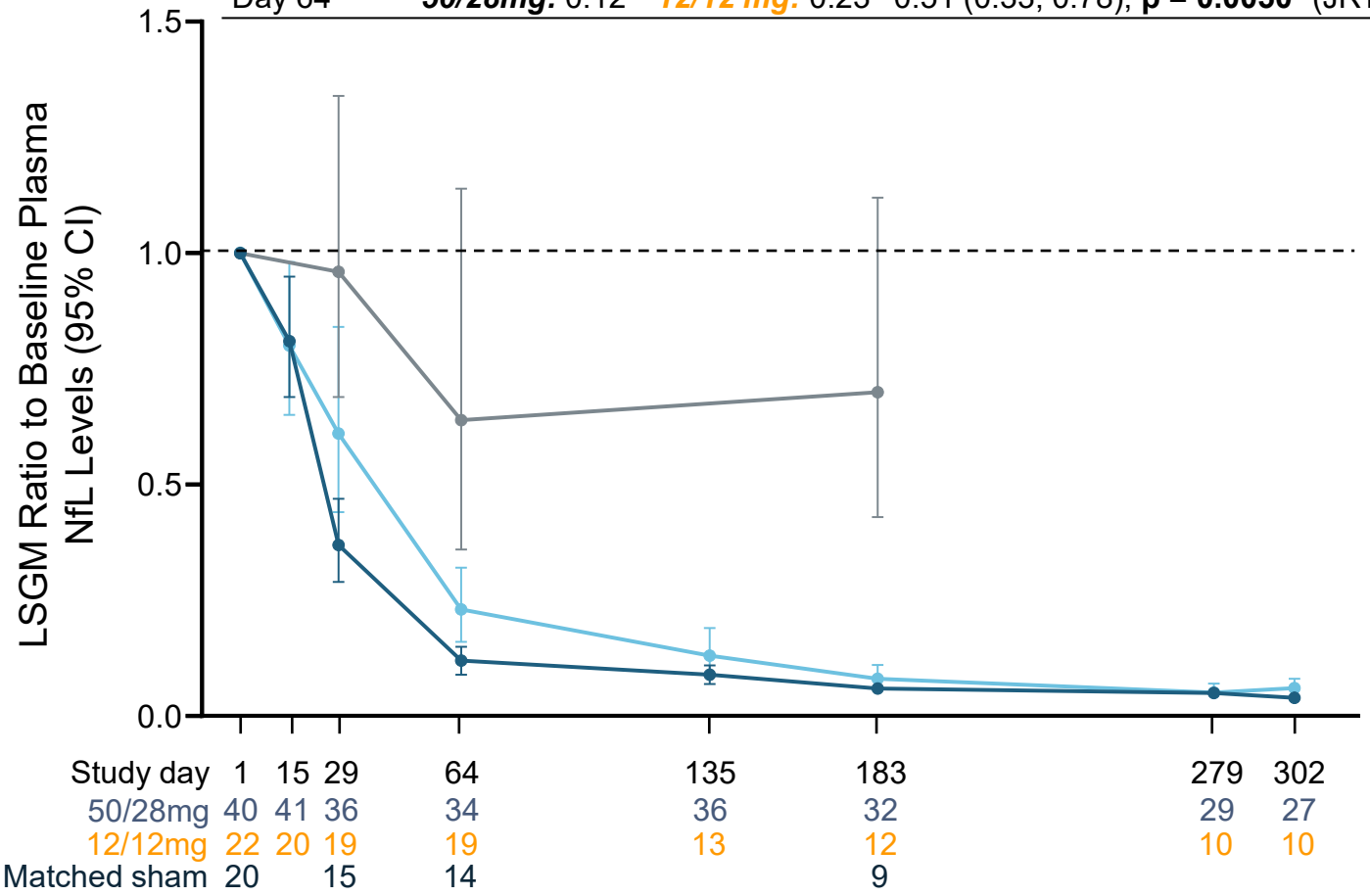


- ANCOVA = analysis of covariance; CI = confidence interval; HINE-2 = Hammersmith Infant Neurological Exam section 2; JRT = joint rank test; LSM = least-squares mean; SE, standard error
Study Day 1 is baseline. Results shown are from multiple imputation and an ANCOVA model with adjustment for participants' disease duration, baseline HINE-2 score, and baseline CHOP-INTEND score. The comparisons between 50/28mg and 12/12mg groups, and between 50/28mg and matched sham groups were performed as separate analyses. LSM difference from ANCOVA; p-value from JRT

Change in Plasma NfL
(Part B Infantile-Onset)

- The 50/28mg group experienced **94% reduction** in plasma NfL from baseline at Day 183, as compared with a 30% reduction in the sham group
- The 50/28mg group experienced greater reductions in plasma NfL at Day 64 relative to the 12/12mg group

Time Point	LSGM Ratio to Baseline		LSGM Ratio (95% CI); p-Value
Day 183	***50/28mg:*** 0.06	***Sham:*** 0.70	0.08 (0.05, 0.14); **p < 0.0001** (JRT)
Day 64	***50/28mg:*** 0.12	***12/12 mg:*** 0.23	0.51 (0.33, 0.78); **p = 0.0050*** (JRT)

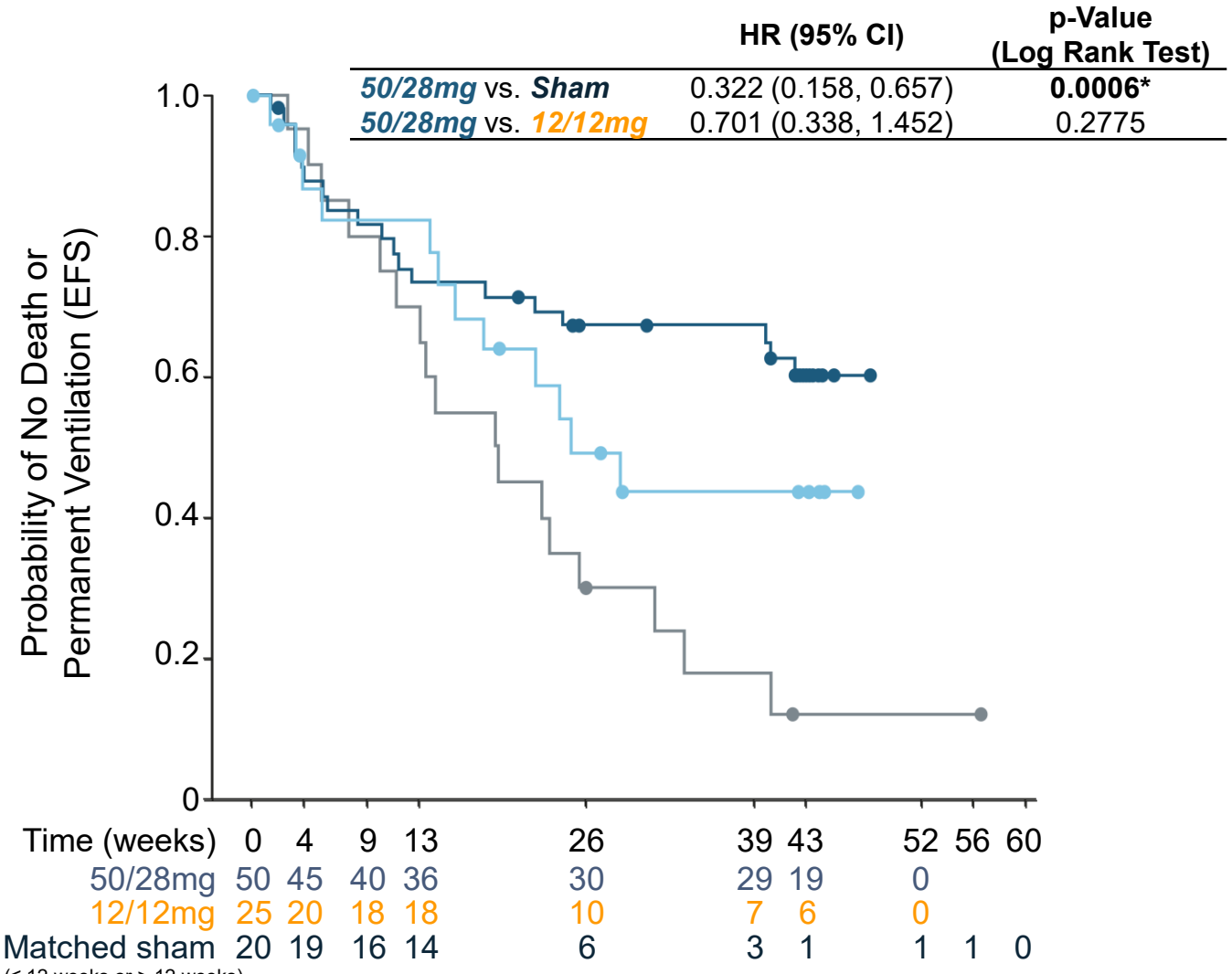


- ANCOVA = analysis of covariance; CHOP-INTEND = Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders; CI = confidence interval; JRT = joint rank test; LSGM = least-squares geometric mean; NfL= neurofilament light chain

*Indicates nominal significance. Study Day 1 is baseline. Multiple imputation was performed based on log transformed plasma NfL. Results shown are from an ANCOVA model with adjustment for participants' disease duration, baseline log plasma NfL, and baseline CHOP-INTEND score. The comparisons between 50/28mg and 12/12mg groups, and between 50/28mg and matched sham groups were performed as separate analyses. LSGM difference from ANCOVA;
p-value from JRT

Event-Free Survival
(Part B Infantile-Onset)

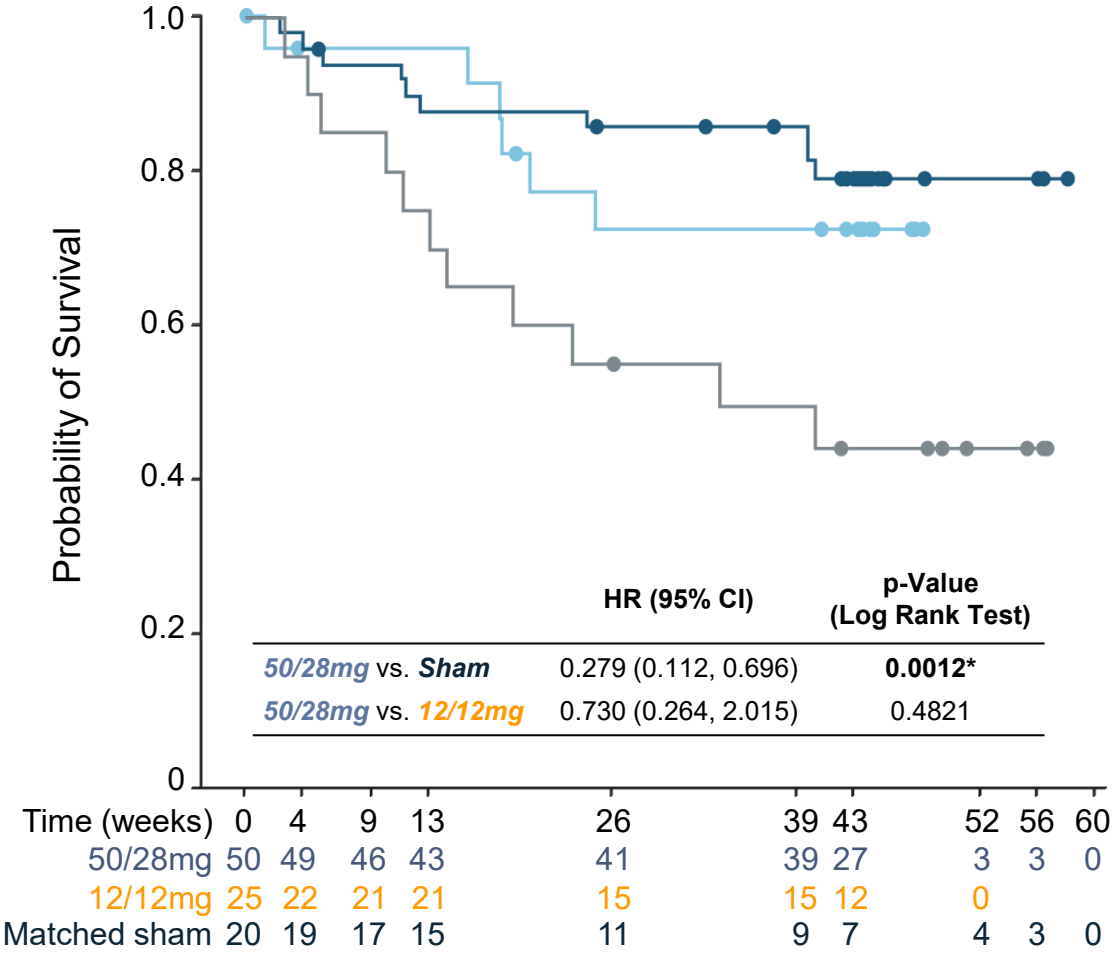
- Risk of death or permanent ventilation (EFS) was reduced by:
  - 67.8% in the 50/28mg group vs. sham group
  - 29.9% in the 50/28mg vs. the 12/12mg group.



- CI = confidence interval; EFS = event-free survival; HR = hazard ratio
  *Indicates nominal significance. p-value is from log rank test stratified by disease duration (≤ 12 weeks or > 12 weeks)

Overall Survival
(Part B Infantile-Onset)

- Risk of death was reduced by 72.1% in the 50/28mg group vs. sham group, and by 27.0% vs. the 12/12mg group



- CI = confidence interval; HR = hazard ratio
*Indicates nominal significance. Hazard ratio determined from Cox proportional hazards model adjusting for disease duration and baseline CHOP-INTEND. p-value is from log rank test stratified by disease duration (≤ 12 weeks or > 12 weeks)

Conclusion

- DEVOTE is an example in which comparison to historical data is the primary analysis in a randomized pivotal clinical trial

- Thoughtful considerations in setting up the statistical framework for the primary analysis a priori is critical for the success