Use of advanced analytical approaches to optimize measurement of treatment effects

Sian Ratcliffe Smethurst, PhD

SVP, Head of Quantitative Sciences & Development Operations

Biogen

February 2025

Disclosures

- Dr Sian Ratcliffe Smethurst is an employee of Biogen and holds stock and/or stock options in Biogen
- Dr Sian Ratcliffe Smethurst is a former employee of Pfizer and holds stock in Pfizer
- The opinions expressed in this presentation are those of Dr Sian Ratcliffe Smethurst

Presentation Outline

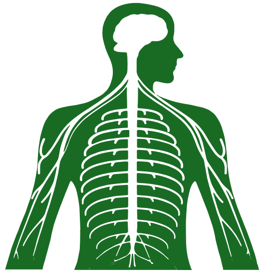
- Leveraging disease biology understanding to build a Causal Mediation Model
  - Tofersen (Qalsody) example of Plasma Neurofilament Light Chain (NfL) as a surrogate for treatment effect in SOD1 ALS
- External data matching, a critical step for the successful approval of Omaveloxolone (Skyclarys)
- The opportunity to utilize Prognostic Scores to reduce sample size in study designs - PROCOVA methods and digital twins

All examples illustrate the value that prior historical data and robust frameworks for regulatory acceptance

Causal Mediation Model

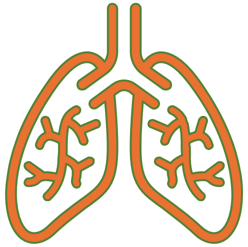
Tofersen (Qalsody) Example

Amyotrophic Lateral Sclerosis (ALS): a rare, fatal neurodegenerative disease characterized by loss of upper and lower motor neurons



ALS is a progressive, adult-onset disease[1]

Weakness leads to difficulty breathing, swallowing, moving limbs, walking



ALS is uniformly fatal

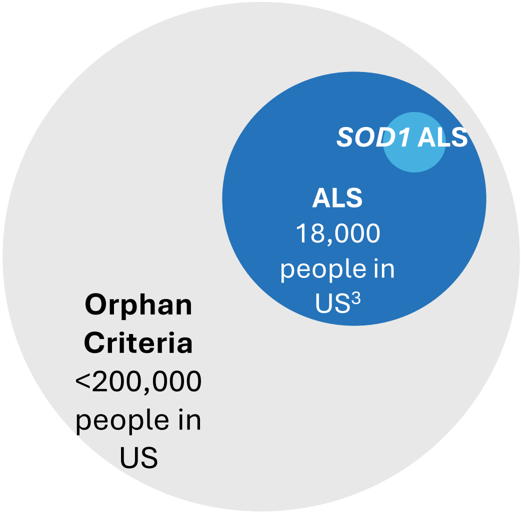
typically due to respiratory failure within 3 to 5 years from symptom onset[2]

SOD1-ALS is a rare, progressive and fatal disease

Caused by a mutation in the superoxide dismutase-1 (*SOD1*) gene

Affects ~330 people in the US[4,5]

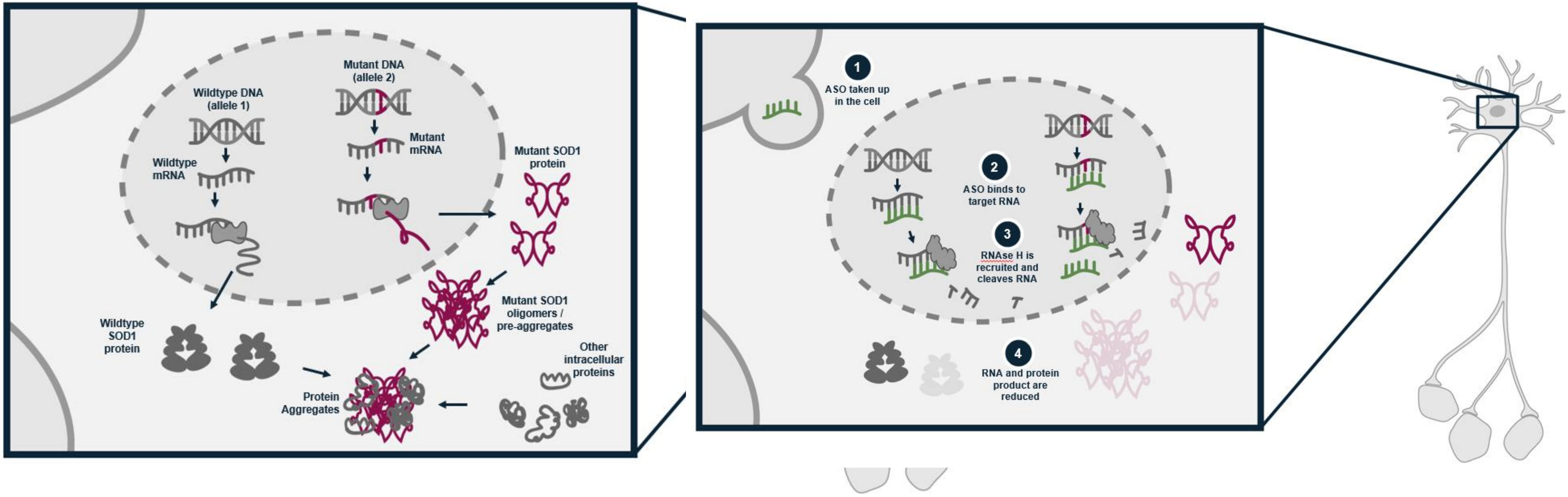
Median survival 2.7 years from diagnosis[6]



1. Al-Chalabi A, Hardiman O. *Nat Rev Neurol.* 2013;9(11):617-628. 2. Brown RH, Al-Chalabi A. *N Engl J Med*. 2017;377(2):162-172.

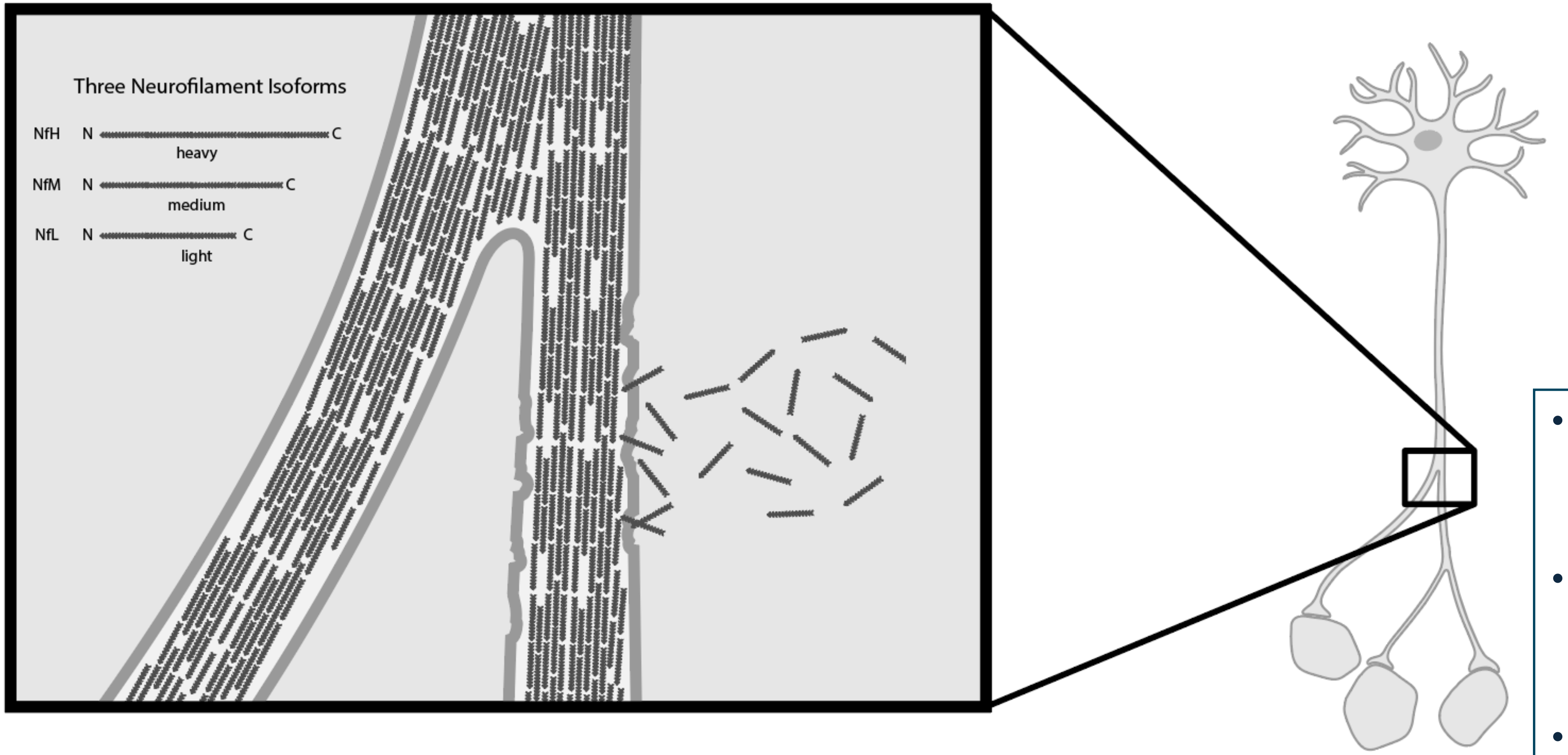
3. Mehta P, et al. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2023;24:108-116. 4. Zou ZY, et al. *J Neurol Neurosurg Psychiatry*. 2017;88:540-549. 5. Brown CA, et al. *Neuroepidemiology*. 2021;55:342-353. 6. Bali T, et al. *J Neurol Neurosurg Psychiatry*. 2017;88:99-105.

Mutations in the *SOD1* gene lead to production of a mutated form of SOD1 protein - Tofersen mediates degradation of *SOD1* mRNA to reduce synthesis of SOD1 protein



ASO, antisense oligonucleotide, RNA, ribonucleic acid; RNase H, ribonuclease H.
Based on Robberecht W, Philips T. *Nat Rev Neurosci.* 2013;14:248-264.

Neurofilaments are a marker of motor neuron integrity

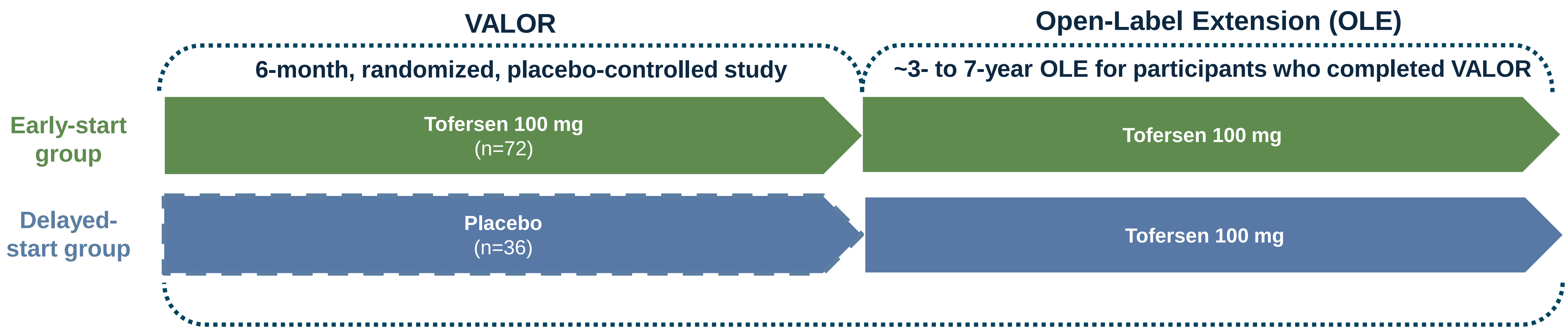


Following axonal injury, neurofilaments leak into the interstitial fluid before passing into the blood and CSF

- Neurofilament levels are elevated in ALS exceeding levels in nearly all other neurodegenerative disease
- Neurofilament levels are prognostic for decline in clinical function in ALS
- Neurofilament levels are prognostic for survival in ALS

Based on Gagliardi D, et al. *Int J Mol Sci.* 2019;20:4152.

VALOR and its open-label extension were conducted to evaluate tofersen in adults with *SOD1*-ALS



Data prospectively integrated to evaluate early- vs. delayed-start tofersen

Population (n=108)

- Adults with weakness attributable to ALS and a confirmed *SOD1* mutation

Primary analysis population

- Composed of n=60 participants predicted to have faster progressing disease based on *SOD1* mutation type and/or pre-randomization ALSFRS-R slope

Primary endpoint

- ALSFRS-R total score

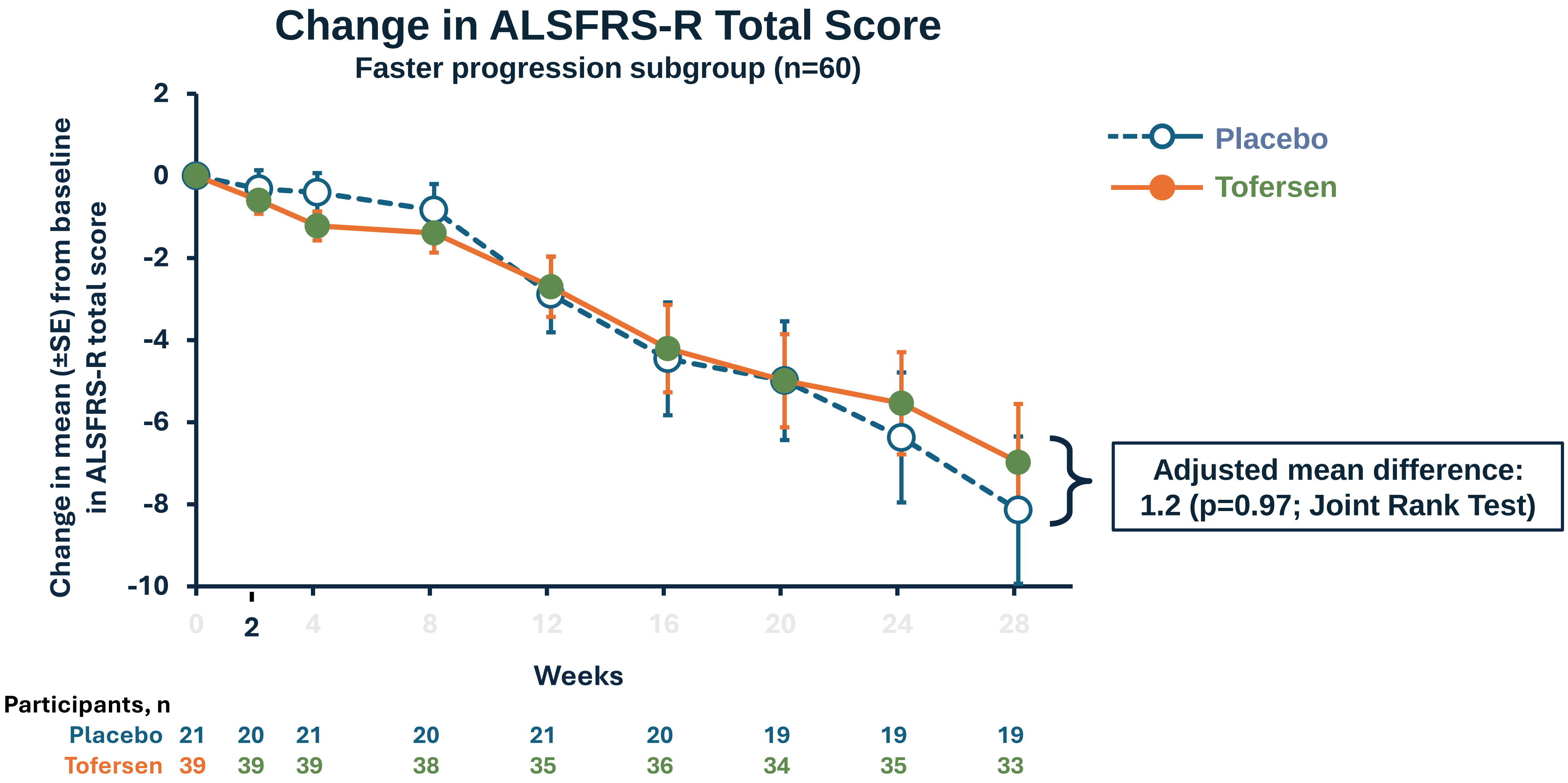
Secondary endpoints (in order of testing)

- Total SOD1 protein
- Plasma NfL
- Percent-predicted slow vital capacity (SVC)
- HHD megascore
- Ventilation assistance-free survival
- Overall survival

VALOR did not achieve statistical significance on its primary endpoint at 6 months. Other clinical secondary outcomes also did not reach statistical significance.

Statistical significance was not achieved on the primary analysis in VALOR

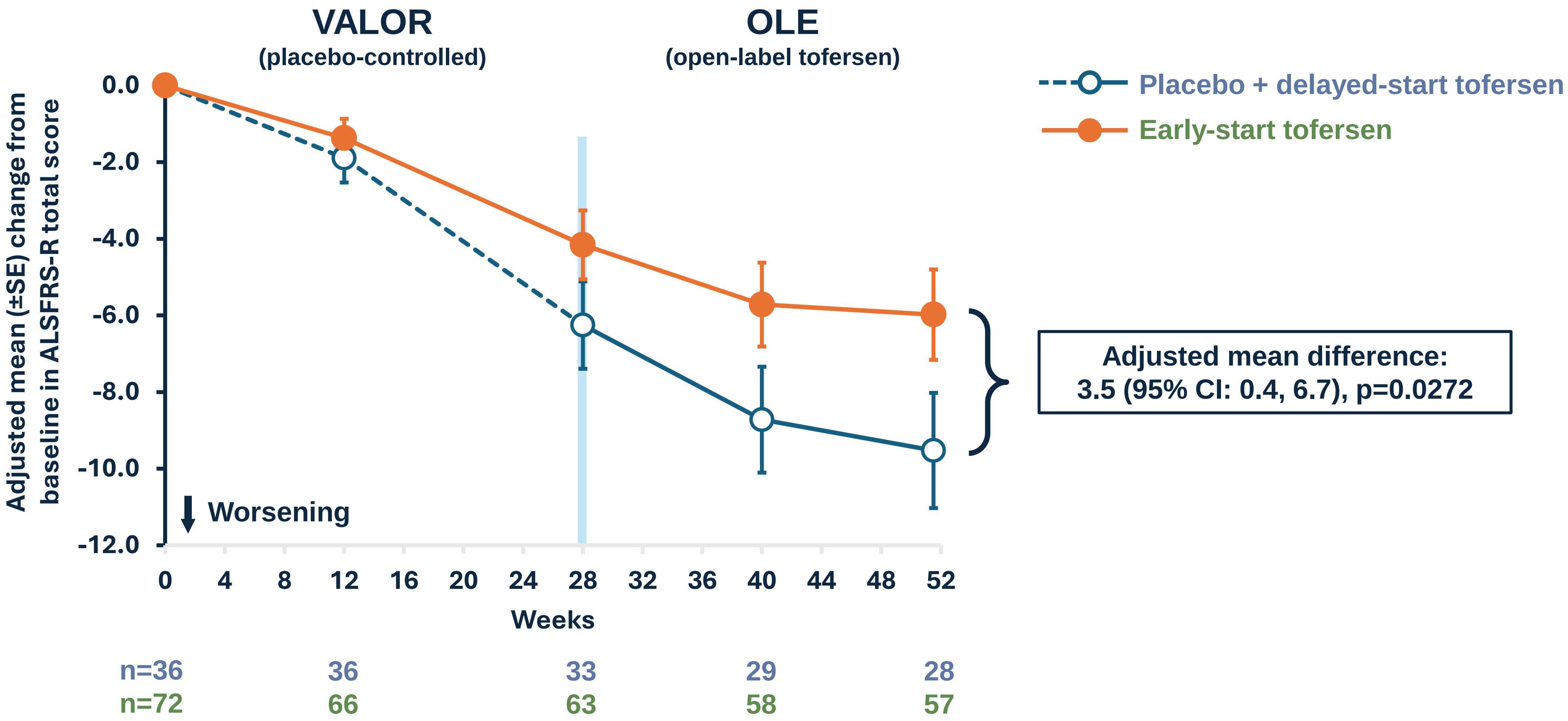
VALOR; Primary Analysis Population (N=60)



Source: VALOR ALSFRS excel data mITT

Effect on clinical function (ALSFRS-R) was observed based on VALOR + Extension Phase

VALOR + OLE; ITT population (N=108)



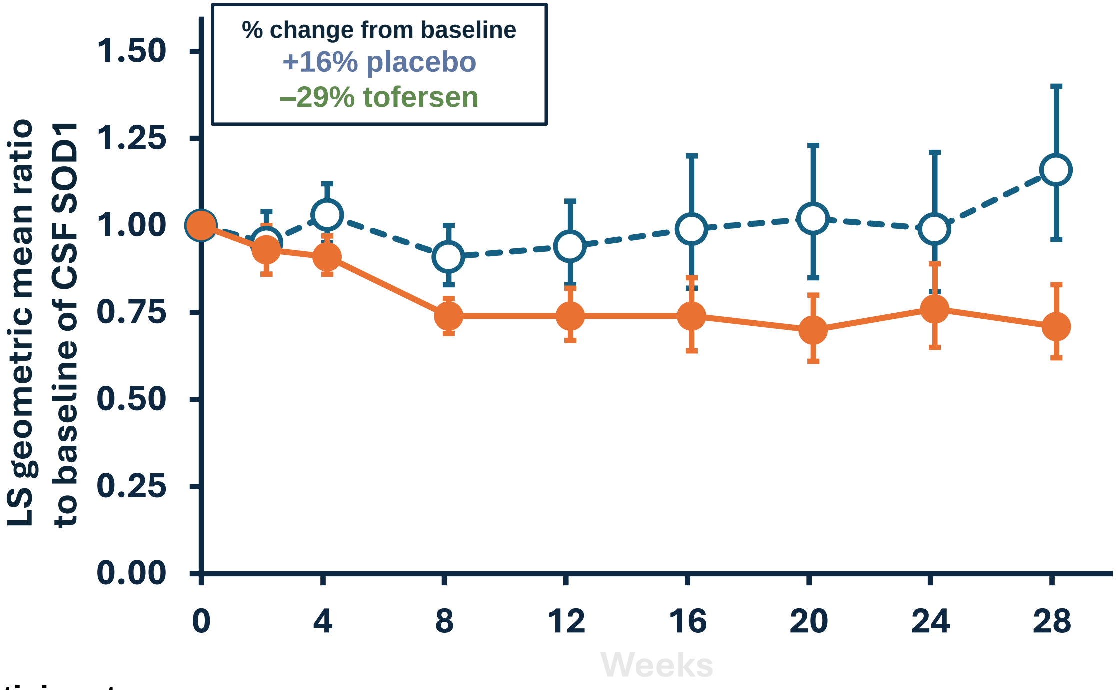
ALS, amyotrophic lateral sclerosis; ALSFRS-R, ALS Functional Rating Scale–Revised; OLE, open-label extension.
Analysis is based on ANCOVA model in conjunction with multiple imputation for missing data. The model includes covariates for the corresponding baseline value, baseline plasma NfL, and use of riluzole or edaravone.

Effects were seen on Biomarkers: CSF SOD1 and plasma NfL

VALOR; Primary Analysis Population (N=60)

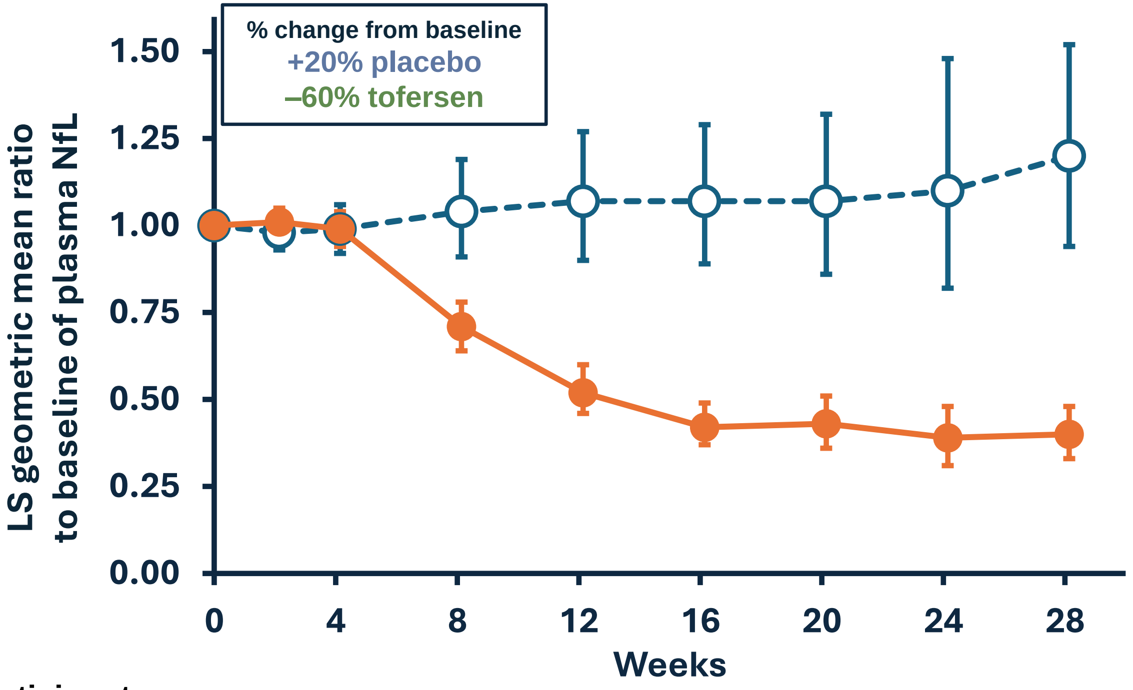
Placebo
Tofersen

Total SOD1 protein in CSF



Participants, n								
Placebo	21	20	18	21	19	19	19	17
Tofersen	38	37	34	33	34	32	32	31

Plasma NfL



Participants, n								
Placebo	21	20	17	20	19	19	18	17
Tofersen	38	37	34	32	33	32	34	27

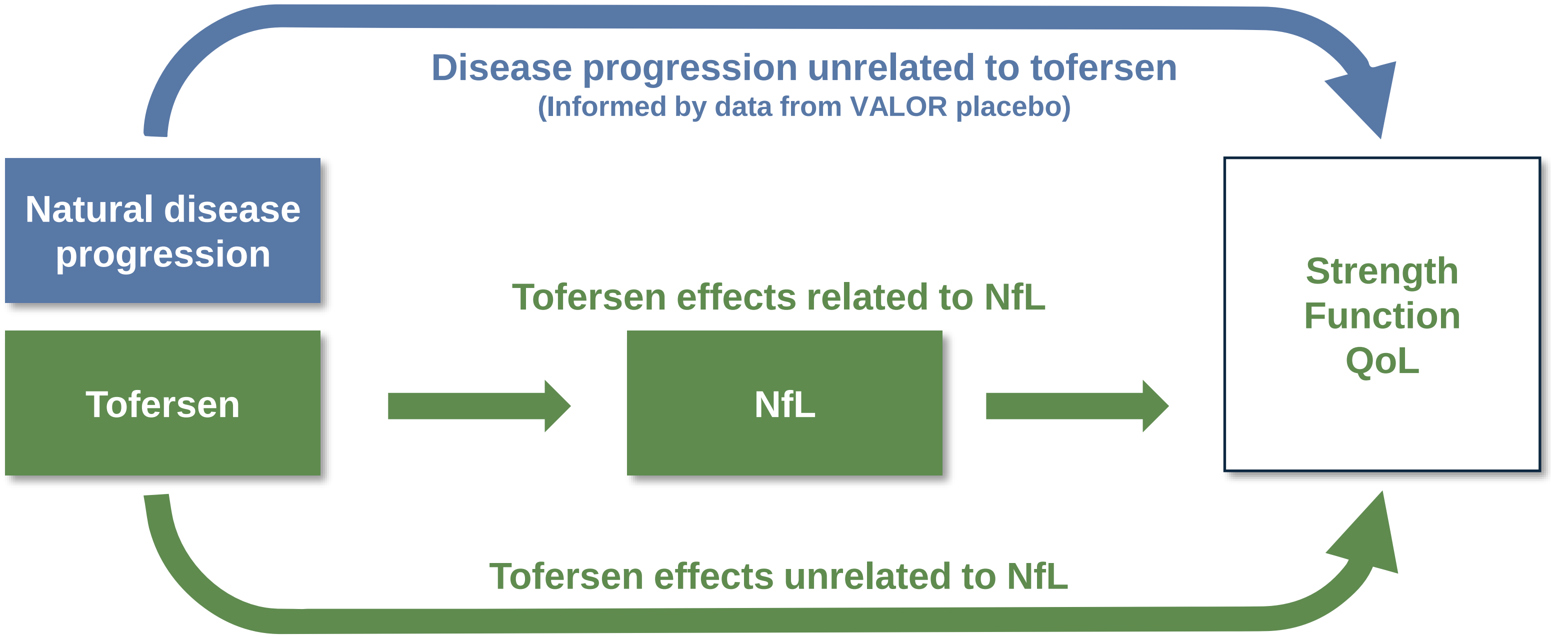
CSF, cerebrospinal fluid; FPS, faster progression subgroup; NfL, neurofilament light; SOD1, superoxide dismutase-1.

Higher baseline NfL level in the tofersen arm predicting more pronounced natural disease progression

	VALOR (ITT; N=108)	
	VALOR: Placebo (n=36)	VALOR: Tofersen (n=72)
Most common *SOD1* mutations		
p.Ile114Thr	10 (27.8)	10 (13.9)
p.Ala5Val	6 (16.7)	11 (15.3)
p.Gly94Cys	2 (5.6)	4 (5.6)
p.His47Arg	4 (11.1)	1 (1.4)
Riluzole use n (%)	22 (61)	45 (63)
Edaravone use n (%)	3 (8)	6 (8)
Time from symptom onset (m)		
median (Q1, Q3)	14.6 (6.6, 32.0)	11.4 (7.2, 28.9)
min, max	2.4, 103.2	1.7, 145.7
% predicted SVC at baseline		
mean (SD)	85.1 (16.5)	82.1 (16.6)
min, max	54.8, 120.4	46.7, 134.7
ALSFRS-R baseline total score		
mean (SD)	37.3 (5.8)	36.9 (5.9)
min, max	24, 47	15, 48
ALSFRS-R pre-randomization slope		
mean (SD)	–1.2 (1.2)	–1.1 (1.4)
min, max	–4.9, –0.02	–8.3, 0.0
ALSFRS-R run-in slope		
mean (SD)	–0.7 (3.3)	–1.0 (2.2)
min, max	–11, 10	–9, 4
Plasma NfL (pg/mL)		
mean (SD)	89.7 (86.5)	100.4 (82.8)
median (min, max)	64.6 (8, 370)	78.5 (5, 329)

ALSFRS-R, ALS Functional Rating Scale-Revised; ITT, intent-to-treat; OLE, open-label extension; SOD1, superoxide dismutase-1; SVC, slow vital capacity.

Mediation model built to evaluate NfL as a potential surrogate biomarker reasonably likely to predict clinical benefit



Model deconstructs the observed effect in a tofersen-treated participant into three components:

Change due to:

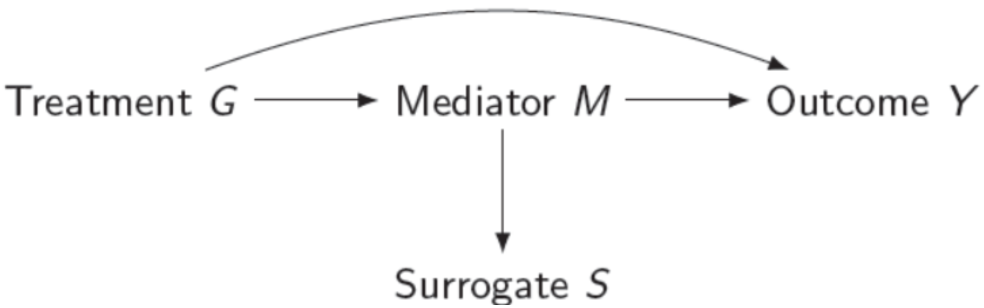
1. Natural disease progression
2. Effect of tofersen through the NfL pathway
3. Effect of tofersen through non-NfL pathway/factors

NfL, neurofilament light; QoL, quality of life.

Justification for Use of Mediation Model

- A strong mediator can be a surrogate endpoint

- A good surrogacy endpoint may not be on the direct causal path



- The model can be reframed with marginal structure models with NDE (natural direct effect; tofersen effect through non-NfL pathway), NIE (natural indirect effect = tofersen effect through NfL pathway), and ATE (average treatment effect = sum of NDE + NIE) derived

Model Fitting Results: Correlations observed for Tofersen-driven reductions on NfL at week 16 with ALSFRS at Week 28 were consistent across other measures in function, strength and QoL

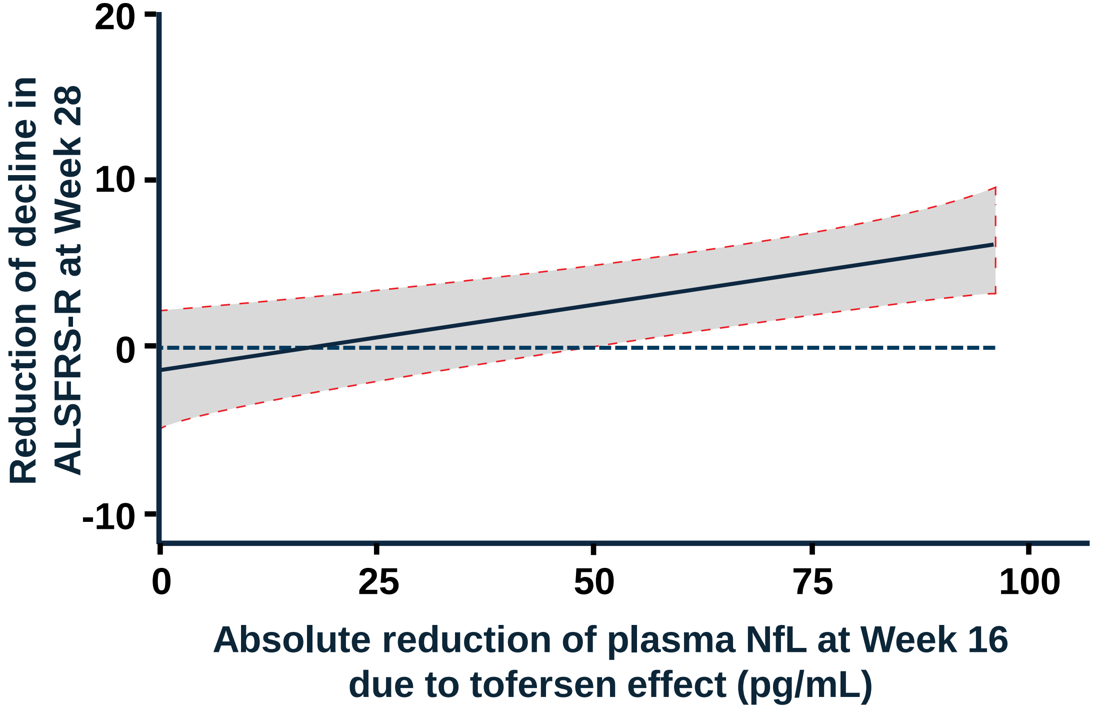
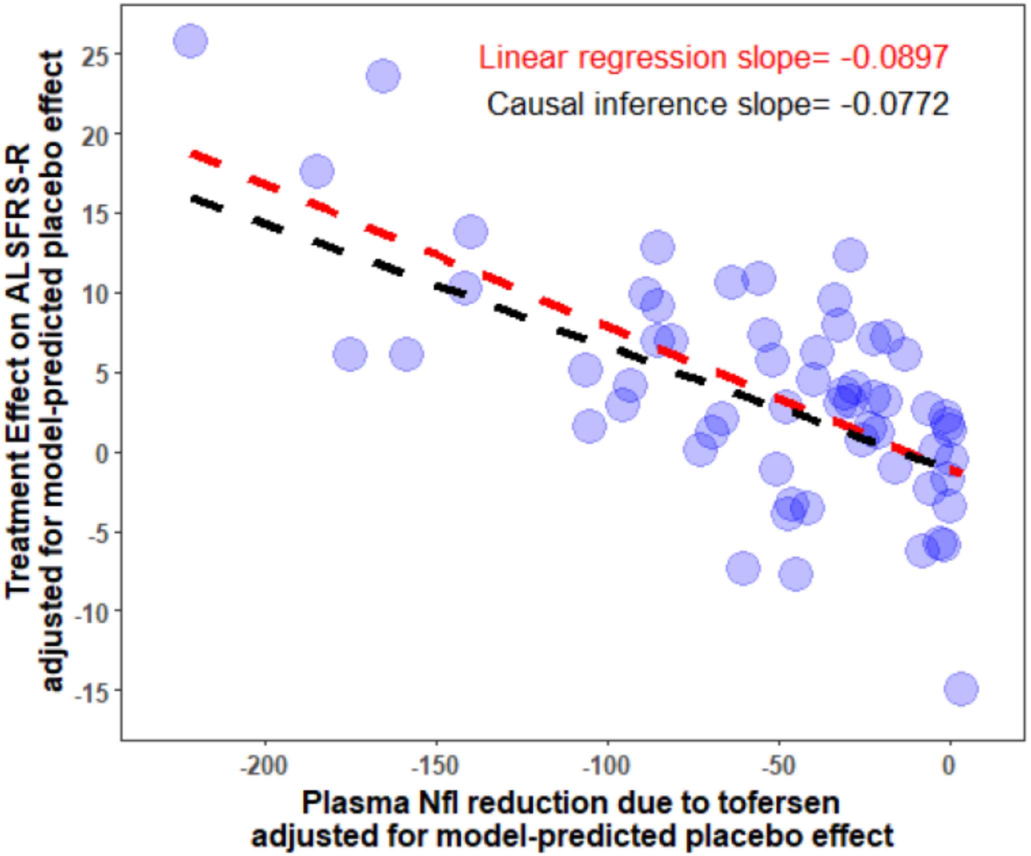


Figure 10: Relationship between Plasma NfL reduction due to tofersen and treatment effect on ALSFRS-R changes from baseline after adjusting for natural ALSFRS-R and NfL progression in tofersen-treated subjects



Source: Clinical Pharmacology Reviewer's Analysis

1. FDA briefing book for the tofersen advisory committee meeting on March 22, 2023.

Tofersen USPI Indication Statement

Tofersen is indicated for the treatment of amyotrophic lateral sclerosis (ALS) in adults who have a mutation in the superoxide dismutase 1 (*SOD1*) gene.

This indication is approved under accelerated approval based on reduction in plasma neurofilament light chain observed in patients treated with tofersen.

Continued approval for this indication may be contingent upon verification of clinical benefit in confirmatory trials.

Propensity matching to external natural history data

Omaveloxolone (Skyclarys) Example

Regulatory Flexibility and Omaveloxolone

- Rare, genetic, progressive, neurodegenerative disorder caused by a deficiency in Frataxin
- Unmet need with no approved treatments

Omaveloxolone
Novel
Nrf2 activator
and
First therapy in FA

MOXie Study 1402 Part 2 demonstrated a statistically significant treatment benefit of omaveloxolone compared to placebo in mFARS (both FAS and ITT)

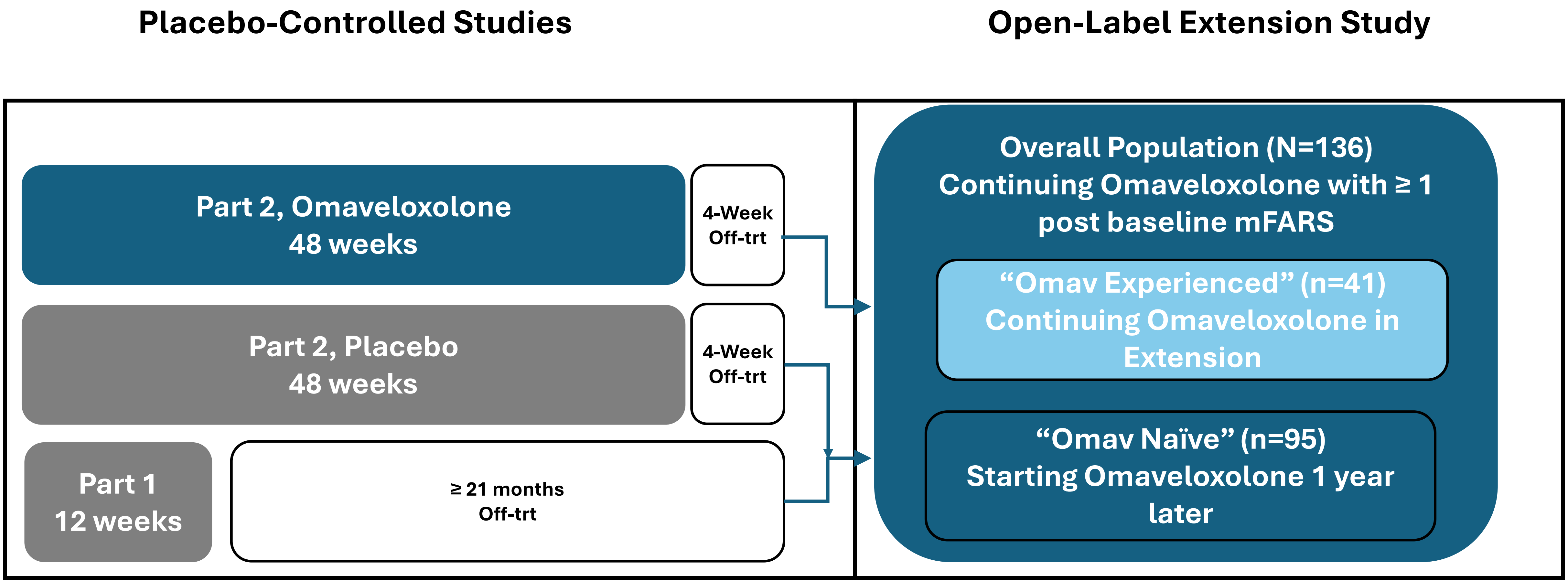
Confirmatory evidence from the natural history comparison

Pharmacodynamic data supporting the biologic plausibility of the treatment effect

Acceptable and manageable safety profile

mFARS=modified Friedreich's ataxia rating scale

MOXie Study 1402 Extension: Overall Population and Subpopulations Based on Prior Omaveloxolone Experience



MOXie Study 1402 Extension and FA-COMS[1] (Ongoing Prospective Natural History Study) External Control Highly Comparable

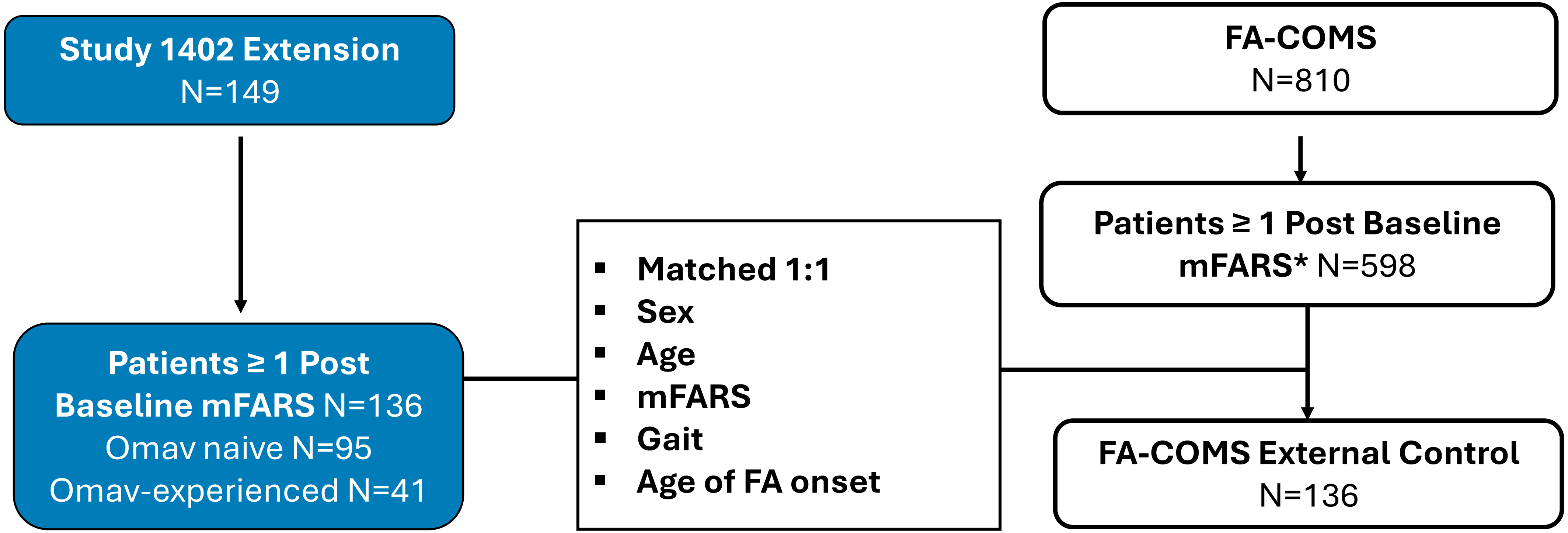
	Pooled Omav N=136	Matched FA-COMS N=136
Female, n (%)	70 **(51.5%)**	70 **(51.5%)**
Male, n (%)	66 **(48.5%)**	66 **(48.5%)**
Age (years), mean (SD)	**26.6** (7.3)	**26.2** (13.7)
Age of FA onset, mean (SD)	**15.5** (5.3)	**15.3** (10.6)
mFARS score, mean (SD)	**42.2** (12.6)	**41.5** (16.9)
Gait (Question 7 from Upright Stability), mean (SD)	**2.8** (1.4)	**2.8** (1.7)

Diagnostics confirmed high level of comparability

1. ClinicalTrials.gov Identifier: NCT03090789

Primary Analysis Population

- Moxie Extension: all patients in the Study 1402 Extension population
- Matched FA-COMS: the corresponding matched natural history patients from the NH population



*Having baseline, at least one post-baseline mFARS within 3 years, and all baseline characteristics used for propensity score calculation

Considerations for 1:1 Matching: Rationale for inclusion of covariates to match on and balance in the algorithm

Logistic Regression Covariate	Rationale for Inclusion	Reference	Number (%) of FA-COMS Patients with Data (n=810)	Number (%) of Study 1402 Extension Patients with Data (n=149)
Age	Age is the primary determinant of phenotypic severity	Patel, 2016	807 (99.6%)	149 (100%)
Age of FA onset	Surrogate for relative rate of progression and GAA repeat length	Patel, 2016	801 (98.9%)	149 (100%)
Sex	Sexual dimorphisms inconsistently observed in ataxia studies	Klockgether, 1998; Friedman, 2010	810 (100%)	149 (100%)
Gait score at baseline	Allows matching of patients at the same level of function	Rummey, 2022	790 (97.5%)	149 (100%)
mFARs score at baseline	Allows matching of patients at the same level of function	Rummey, 2022	789 (97.4%)	149 (100%)
Other covariates considered but not included				
GAA1 Repeat Length	Not included		745 (92%)	131 (87.9%)
Pes cavus[a]	Not included		432 (53.3%)	149 (100%)

a. The definition of pes cavus between the 2 studies was not consistent. Pes cavus was based on clinical judgment in FA-COMS; however, Study 1402 Extension defined a flashlight test such that if light was visible under the arch of the foot while standing the patient was deemed as having pes cavus.

Propensity-Matched Diagnostics Confirming Highly Comparable Match on Demographics and Stage of Disease

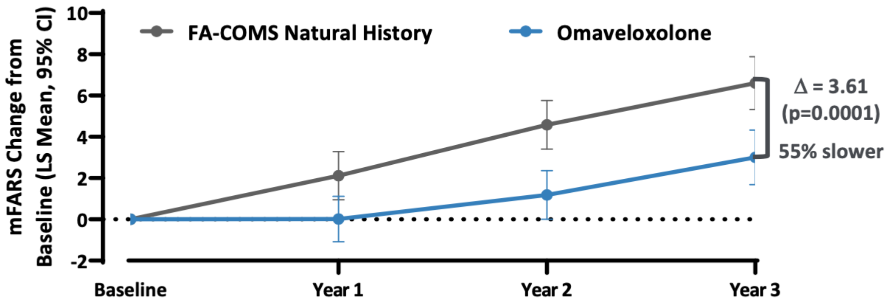
Diagnostic	Score	Good Fit Criterion
Standardized Difference of Means of Propensity Score	0.0055	<0.1
Standardized Difference of Means of Covariates		
Sex	0.0000	<0.1
Baseline gait	0.0672	<0.1
Baseline mFARS	0.0826	<0.1
Age at baseline	0.0375	<0.1
Age at FA onset	0.0292	<0.1
Ratio of the Variances of the Propensity Score	1.0243	~1; >0.8 and <1.25
Ratio of the Variances of the Residuals for Covariates		0.5 to 2
Sex	0.9999	0.5 to 2
Baseline gait	0.5751	0.5 to 2
Baseline mFARS	0.6068	0.5 to 2
Age at baseline	0.3428	0.5 to 2
Age at FA onset	0.3194	0.5 to 2

Post-hoc propensity matched results: Patients in MOXie Study 1402 Extension progress slower vs FA-COMS external control

Propensity-Matched Analysis
Primary Pooled Population

Patients in the matched FA-COMS group progressed 6.61 mFARS points at Year 3 vs. 3.00 mFARS points for patients treated with omaveloxolone in MOXIe Extension (nominal p=0.0001)

mFARS progression 55% slower in MOXIe Extension patients compared to matched FA-COMS patients at Year 3



	Baseline	Year 1	Year 2	Year 3
MOXIe Extension (n)	136	133	102	77
FA-COMS Natural History (n)	136	108	103	83

mFARS	Baseline	Year 1	Year 2	Year 3
	Mean (SD)	LS Mean (SE)	LS Mean (SE)	LS Mean (SE)
MOXIe Extension	42.2 (12.60)	0.015 (0.56)	1.18 (0.59)	3.00 (0.66)
Matched FA-COMS	41.0 (16.10)	2.11 (0.59)	4.58 (0.59)	6.61 (0.65)
Difference	-	-2.10 (0.81) p=0.0101	-3.41 (0.84) p< 0.0001	**-3.61 (0.93) p=0.0001**

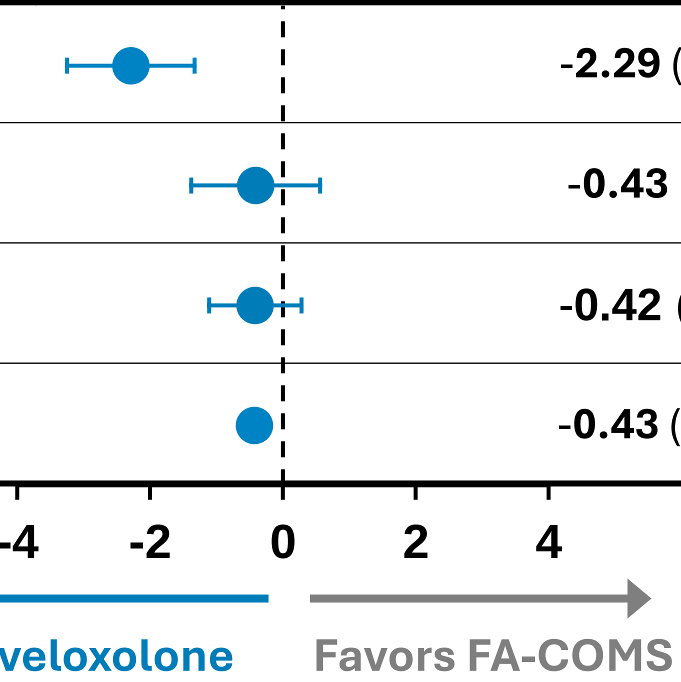
MOXIe
a study in Friedreich's ataxia

REATA PHARMACEUTICALS

Lynch, D. R., Goldsberry, A., Rummey, C., Farmer, J., Boesch, S., Delatycki, M. B., et al. (2024). Propensity matched comparison of omaveloxolone treatment to Friedreich ataxia natural history data. *Annals of Clinical and Translational Neurology, 11*(1), 4–16.

All mFARS Sections Favored Omaveloxolone vs Matched FA-COMS External Control

	Number of Patients Contributing Data			
	Omav	FA-COMS	Year 3 mFARS Treatment Difference (95% CI)	p-value
Upper Limb Coordination	136	136	**-2.29** (-3.25, -1.33)	<0.0001
Upright Stability	136	136	**-0.43** (-1.40, 0.53)	0.3798
Lower Limb Coordination	136	136	**-0.42** (-1.11, 0.28)	0.2374
Bulbar Function	136	136	**-0.43** (-0.61, -0.25)	<0.0001



Difference between treatment groups is omaveloxolone – FA-COMS

Prognostic Scores

Underpinning Digital Twin Methodology

Regulatory Guidance on Adjusting for Covariates

Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

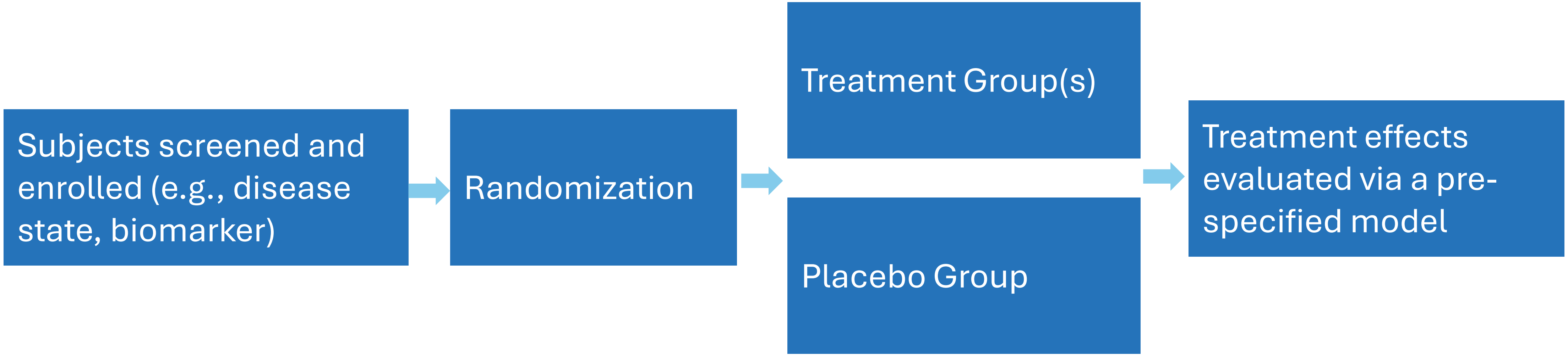
- Covariate adjustment leads to efficiency gains when the covariates are prognostic for the outcome of interest in the trial. Therefore, FDA recommends that sponsors adjust for covariates that are anticipated to be most strongly associated with the outcome of interest. In some circumstances these covariates may be known from the scientific literature. In other cases, it may be useful to use previous studies (e.g., a Phase 2 trial) to select prognostic covariates or form prognostic indices.

- Covariate adjustment can still be performed with covariates that are not prognostic, but there may not be any gain in precision (or may be a loss in precision) compared with an unadjusted analysis.

This guidance is consistent with the ICH guidance for industry *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials*. After specifying the treatment condition of interest, target population, and endpoint variable — the treatment effect estimated by covariate adjustment is a population summary measure defining an estimand.

Covariate Adjustment

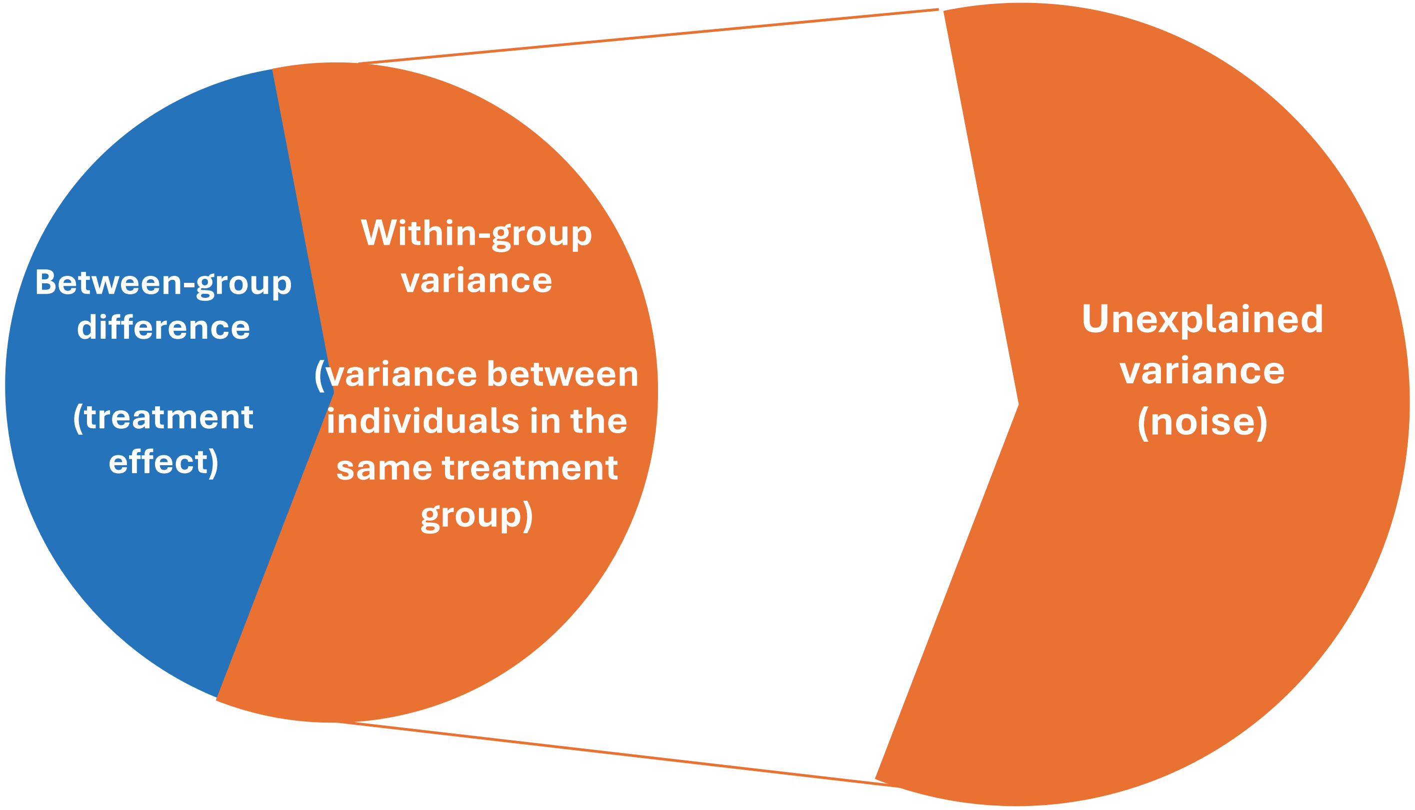
A statistical technique that accounts for baseline patient characteristics (called **covariates**) when estimating population-level treatment effects



Unadjusted model: endpoint change ~ treatment
Adjusted model: endpoint change ~ treatment + **age + gender + disease state + genotype + ...**

Covariate Adjustment Can Increase Power

Unadjusted, all within-group variance is unexplained

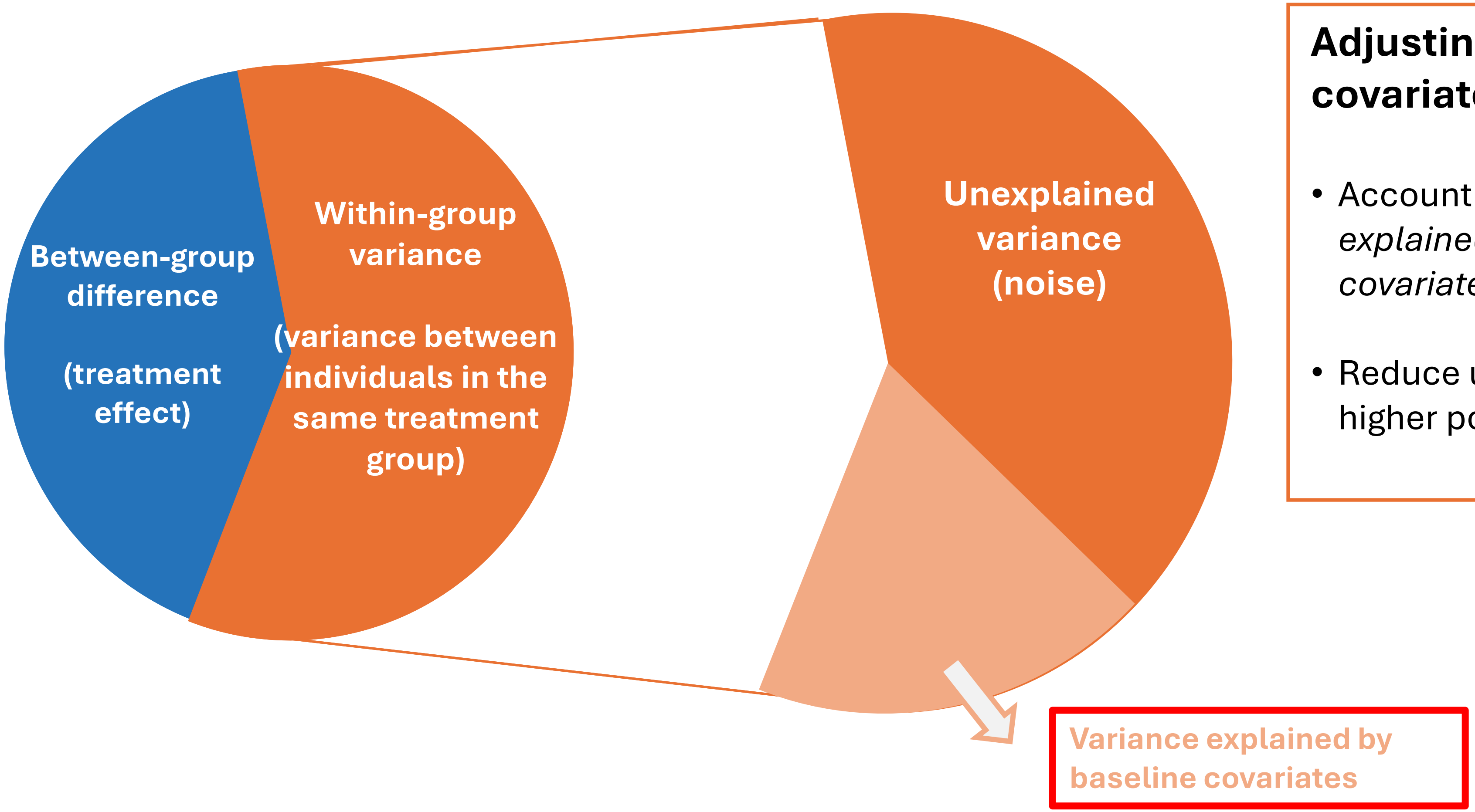


In an unadjusted model:

- All within-group variance is unexplained and attributed to random error
- Higher unexplained variance means the study has lower power

Covariate Adjustment Can Increase Power

By accounting for otherwise unexplained variance in disease progression



Adjusting for standard baseline covariates can:

- Account for some within-group variance *explained by prognostic baseline covariates*
- Reduce unexplained variance, leading to higher power (or narrower CI)

Prognostic Score (PS) constructed using models trained on historical data – from simple to ensemble machine learning models

Baseline Information

- Demographics
- Baseline Disease Status
- Biomarkers

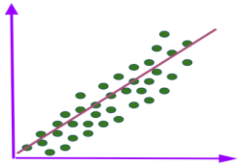
Anything else we can measure....

Some Mathematical Function

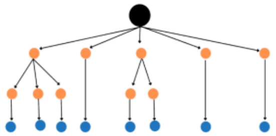
Learned Using:

Historical Disease Progression Dataset(s)

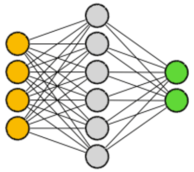
+



Traditional Statistical Models e.g., Linear Regression



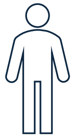
Traditional Machine Learning



Deep Learning

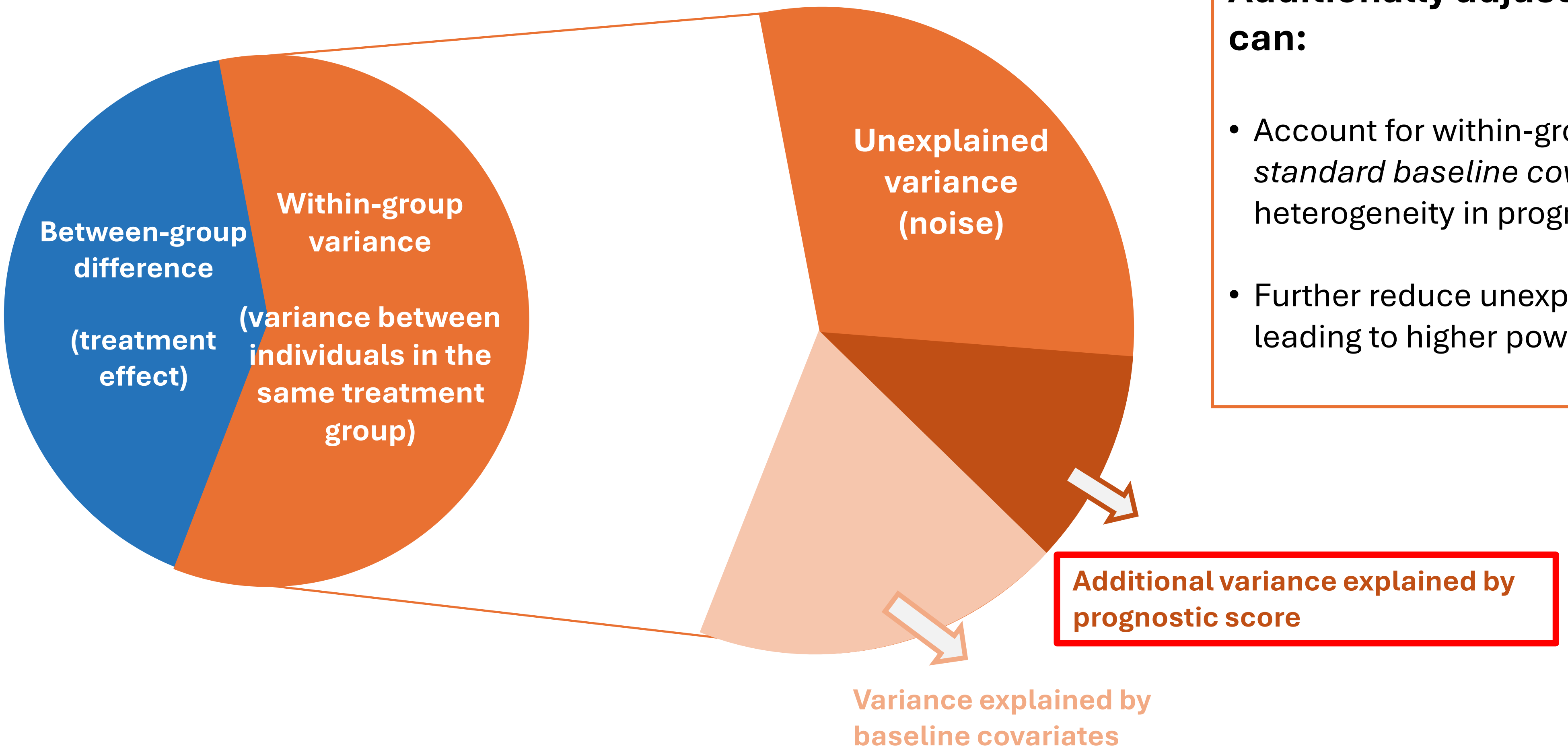
Prognostic Score

- A single quantity per participant that effectively predicts future disease progression (e.g., CDR-SB change).
- Sometimes prognostic scores can be called “digital twins”



Adjusting for a Prognostic Score can further increase power

By accounting for additional unexplained variance



Additionally adjusting for a highly PS can:

- Account for within-group variance *on top of standard baseline covariates* (e.g., heterogeneity in progression)
- Further reduce unexplained variance, leading to higher power

How Prognostic Scores can enhance the ability to detect treatment effects

% Variance Reduction

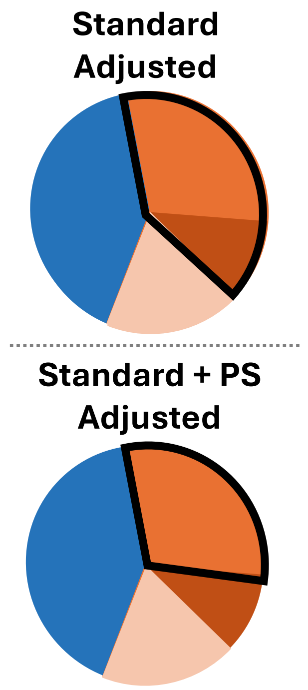
- Reduction in unexplained variance (i.e., noise) with PS adjustment
- Relative sizes of the outlined slices

Power Increase[1]

- Reducing unexplained variance boosts RCT power by enhancing signal detection

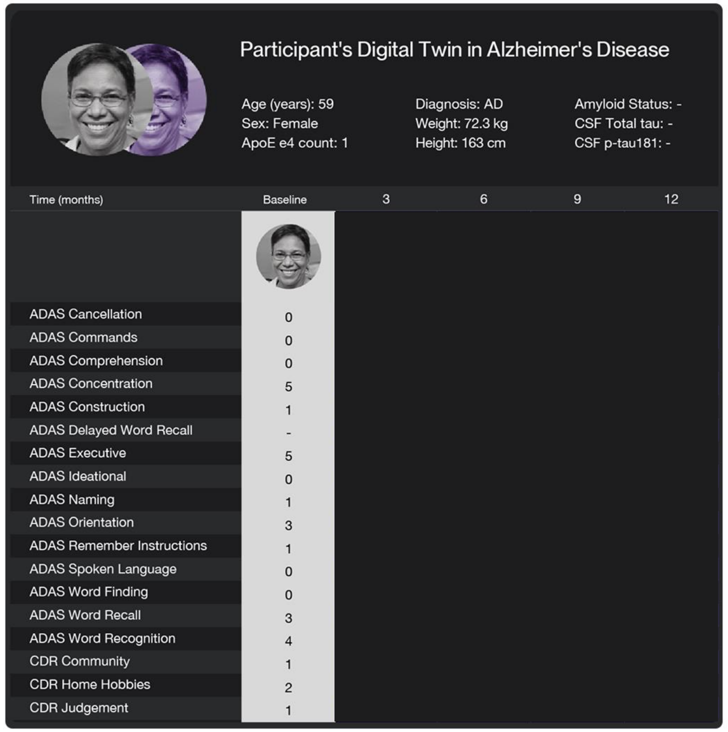
Effective Sample Size Increase (ESSI)

- Reducing unexplained variance also increases the ESSI by allowing the same data to yield more precise estimates



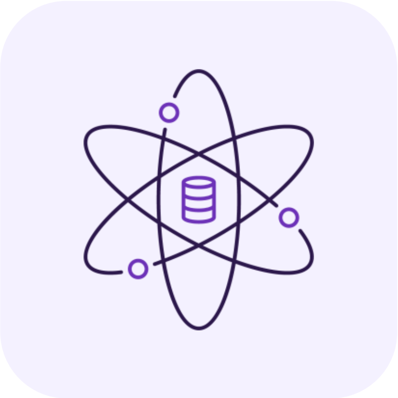
[1] Power increase comes from keeping sample size/costs the same combined with variance reduction due to PS adjustment

Leveraging prognostic scores as Digital Twins - A patient's digital twin is a comprehensive forecast of their future clinical outcomes

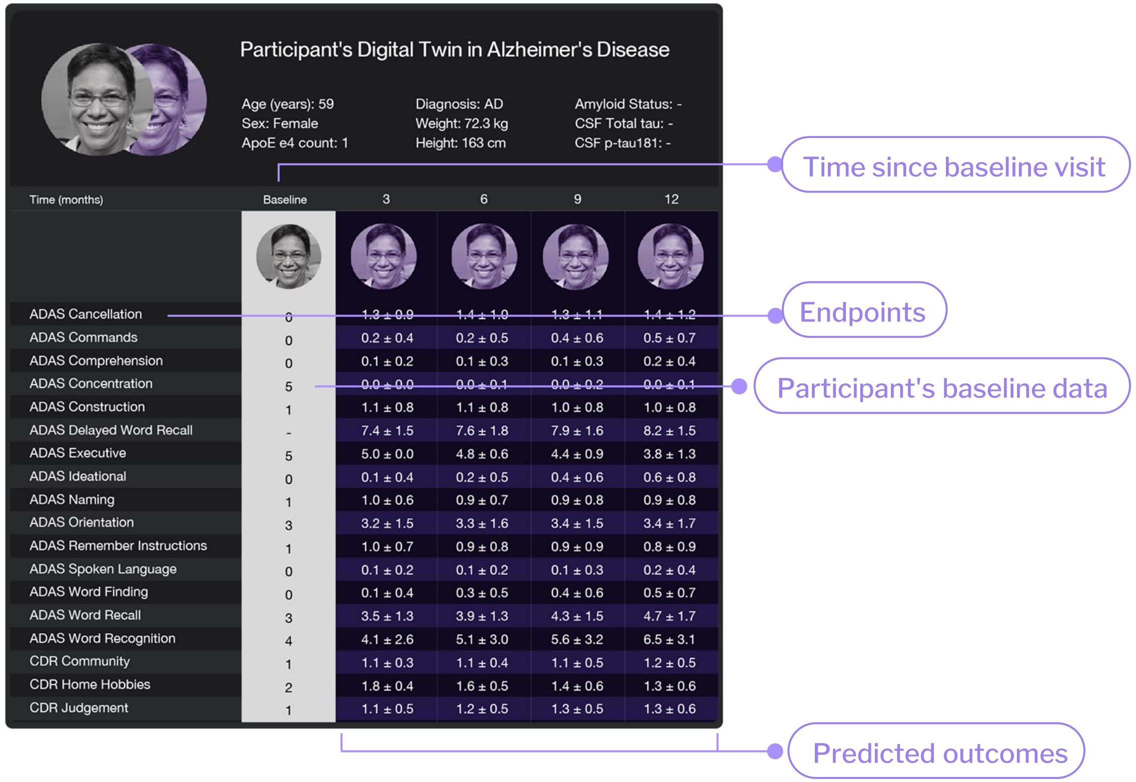


Time (months)	Baseline	3	6	9	12
ADAS Cancellation	0				
ADAS Commands	0				
ADAS Comprehension	0				
ADAS Concentration	5				
ADAS Construction	1				
ADAS Delayed Word Recall	-				
ADAS Executive	5				
ADAS Ideational	0				
ADAS Naming	1				
ADAS Orientation	3				
ADAS Remember Instructions	1				
ADAS Spoken Language	0				
ADAS Word Finding	0				
ADAS Word Recall	3				
ADAS Word Recognition	4				
CDR Community	1				
CDR Home Hobbies	2				
CDR Judgement	1				

Clinical trial participant's baseline data



Participant's baseline data is run through a digital twin generator that was trained on historical data*



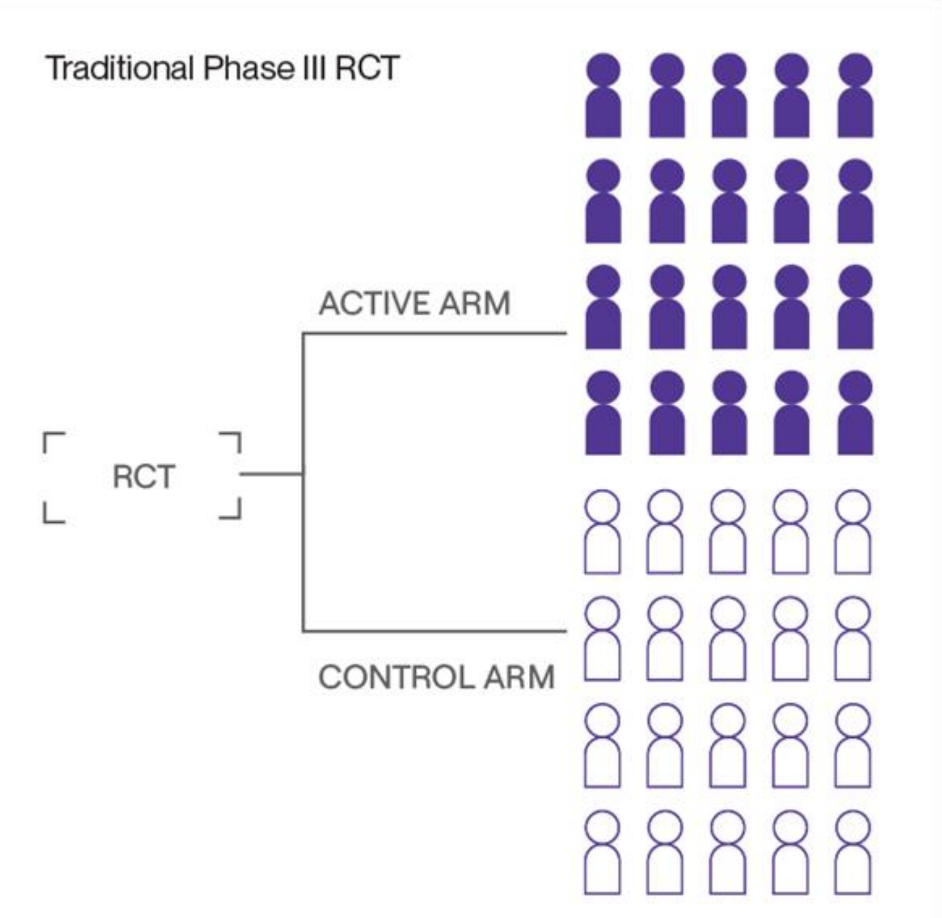
Time (months)	Baseline	3	6	9	12
ADAS Cancellation	0	1.3 ± 0.9	1.4 ± 1.0	1.3 ± 1.1	1.4 ± 1.2
ADAS Commands	0	0.2 ± 0.4	0.2 ± 0.5	0.4 ± 0.6	0.5 ± 0.7
ADAS Comprehension	0	0.1 ± 0.2	0.1 ± 0.3	0.1 ± 0.3	0.2 ± 0.4
ADAS Concentration	5	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.2	0.0 ± 0.1
ADAS Construction	1	1.1 ± 0.8	1.1 ± 0.8	1.0 ± 0.8	1.0 ± 0.8
ADAS Delayed Word Recall	-	7.4 ± 1.5	7.6 ± 1.8	7.9 ± 1.6	8.2 ± 1.5
ADAS Executive	5	5.0 ± 0.0	4.8 ± 0.6	4.4 ± 0.9	3.8 ± 1.3
ADAS Ideational	0	0.1 ± 0.4	0.2 ± 0.5	0.4 ± 0.6	0.6 ± 0.8
ADAS Naming	1	1.0 ± 0.6	0.9 ± 0.7	0.9 ± 0.8	0.9 ± 0.8
ADAS Orientation	3	3.2 ± 1.5	3.3 ± 1.6	3.4 ± 1.5	3.4 ± 1.7
ADAS Remember Instructions	1	1.0 ± 0.7	0.9 ± 0.8	0.9 ± 0.9	0.8 ± 0.9
ADAS Spoken Language	0	0.1 ± 0.2	0.1 ± 0.2	0.1 ± 0.3	0.2 ± 0.4
ADAS Word Finding	0	0.1 ± 0.4	0.3 ± 0.5	0.4 ± 0.6	0.5 ± 0.7
ADAS Word Recall	3	3.5 ± 1.3	3.9 ± 1.3	4.3 ± 1.5	4.7 ± 1.7
ADAS Word Recognition	4	4.1 ± 2.6	5.1 ± 3.0	5.6 ± 3.2	6.5 ± 3.1
CDR Community	1	1.1 ± 0.3	1.1 ± 0.4	1.1 ± 0.5	1.2 ± 0.5
CDR Home Hobbies	2	1.8 ± 0.4	1.6 ± 0.5	1.4 ± 0.6	1.3 ± 0.6
CDR Judgement	1	1.1 ± 0.5	1.2 ± 0.5	1.3 ± 0.5	1.3 ± 0.6

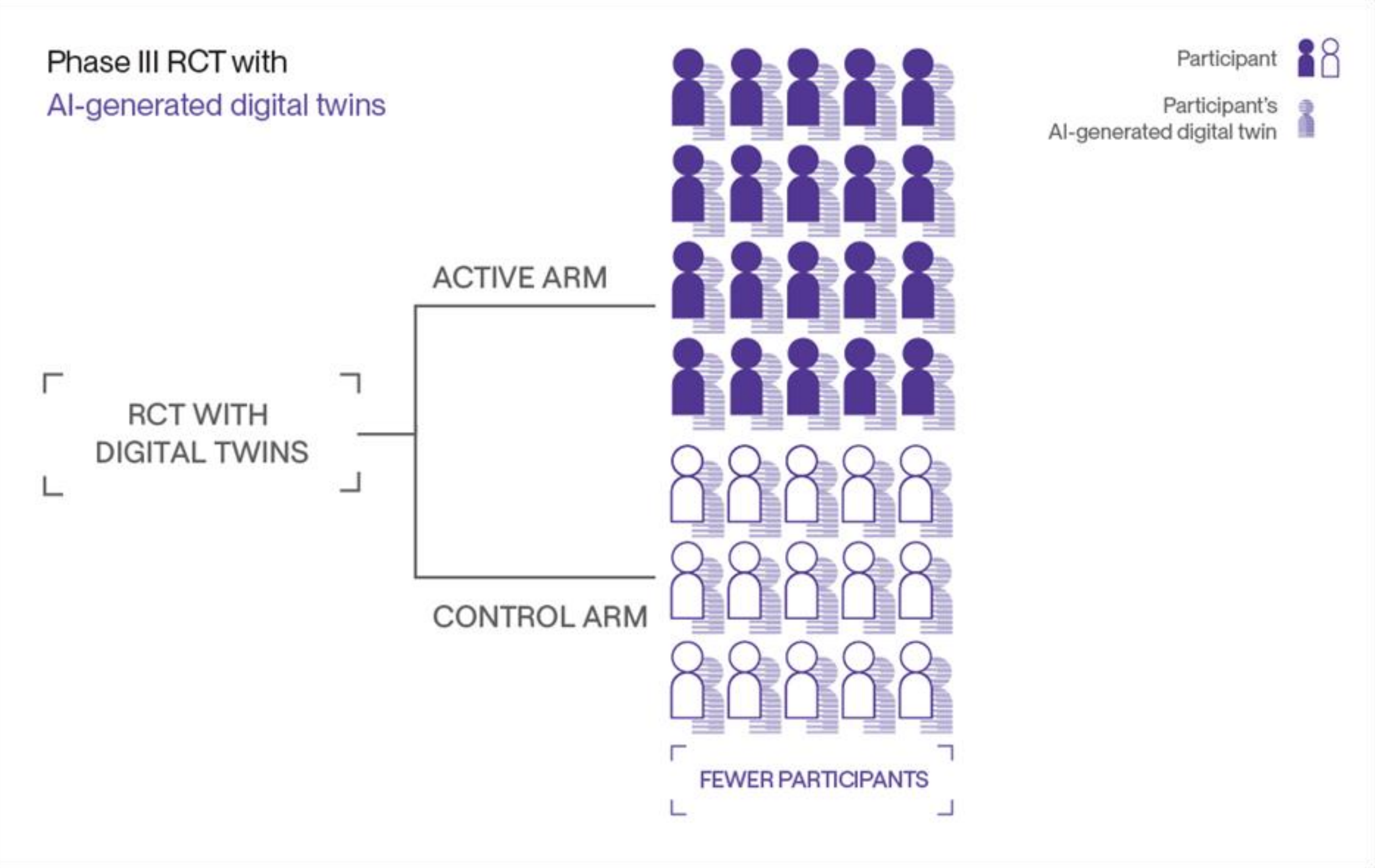
Participant's digital twin is then created

Historical data from large consortia databases/Nat Hx studies comprising multiple disease measures, endpoints and features

*images courtesy of unlearn.ai

Digital twins can used to run studies with fewer participants and high power to enable faster trials





*Unlearn's EMA Qualification Opinion; FDA's Concurring Statements

Concluding Remarks

- Robust analytical frameworks and data-driven approaches have been effective in supporting regulatory flexibility in rare disease drug development
- With deep understanding of disease biology and drug mechanism of action, causal mediation models can be constructed to assess potential surrogate biomarker predictive of clinical effect
- Propensity matching approaches to historical control data can support establishing confirmatory evidence from single arm trials
- Leveraging prognostic score methods to create digital twins can be used to run smaller studies with higher power

Acknowledgments

- Peng Sun (Biogen)
- Richard Foster (Biogen)
- Luan Lin (former Biogen)
- Susie Sinks (Biogen)
- Manjit McNeil (Biogen)
- Brian Millen (Biogen)
- Bo Lu (Ohio State University)
- Ming-Hui Chen (UConn)
- Shannon Rich (former Biogen)
- Angie Goldsberry (former Biogen)
- Phoebe Jiang (Biogen)
- Roland Brown (Biogen)
- Luis Olmos (Unlearn)