Statistical leadership in establishing surrogacy: applications and reflections

Tianle Chen, Luan Lin

Joint work with Ying Tian, Peng Sun and Brian Millen

Biogen Inc., Cambridge, MA, USA

Disclosures

Tianle Chen and Luan Lin are employees of Biogen and holds stock and/or stock options in Biogen.

Tofersen was granted accelerated approval by FDA in the U.S. and is marketed under the brand name QALSODY™. In the rest of the world, tofersen is an investigational drug where safety and efficacy have not been evaluated or established.

Aducanumab is approved for use in the United States, United Arab Emirates, and Qatar. In the rest of the world, it is an investigational drug.

In the U.S., aducanumab-avwa is indicated for the treatment of Alzheimer's disease. This indication is approved under accelerated approval based on reduction in amyloid beta plaques observed in patients treated with aducanumab-avwa. Continued approval for this indication may be contingent upon verification of clinical benefit in confirmatory trial(s).

Outline

- Background of surrogate endpoint
  - Definition, rationale and examples
  - History and regulatory position
- Statistical validation of surrogate endpoint
  - Prentice criterion and proportion of treatment effect explained
  - Principal stratification causal inference approach
  - Correlation-based approaches
  - Meta-analytical approach
- Surrogate endpoint examples in neuroscience
  - Amyloid reduction in Alzheimer's disease
  - Plasma NfL in SOD1-ALS
- Concluding remarks

What is a surrogate endpoint?

- A marker, such as a laboratory measurement, radiographic image, physical sign, or other measure, that is not itself a direct measurement of clinical benefit but is known to predict clinical benefit or is reasonably likely to predict clinical benefit

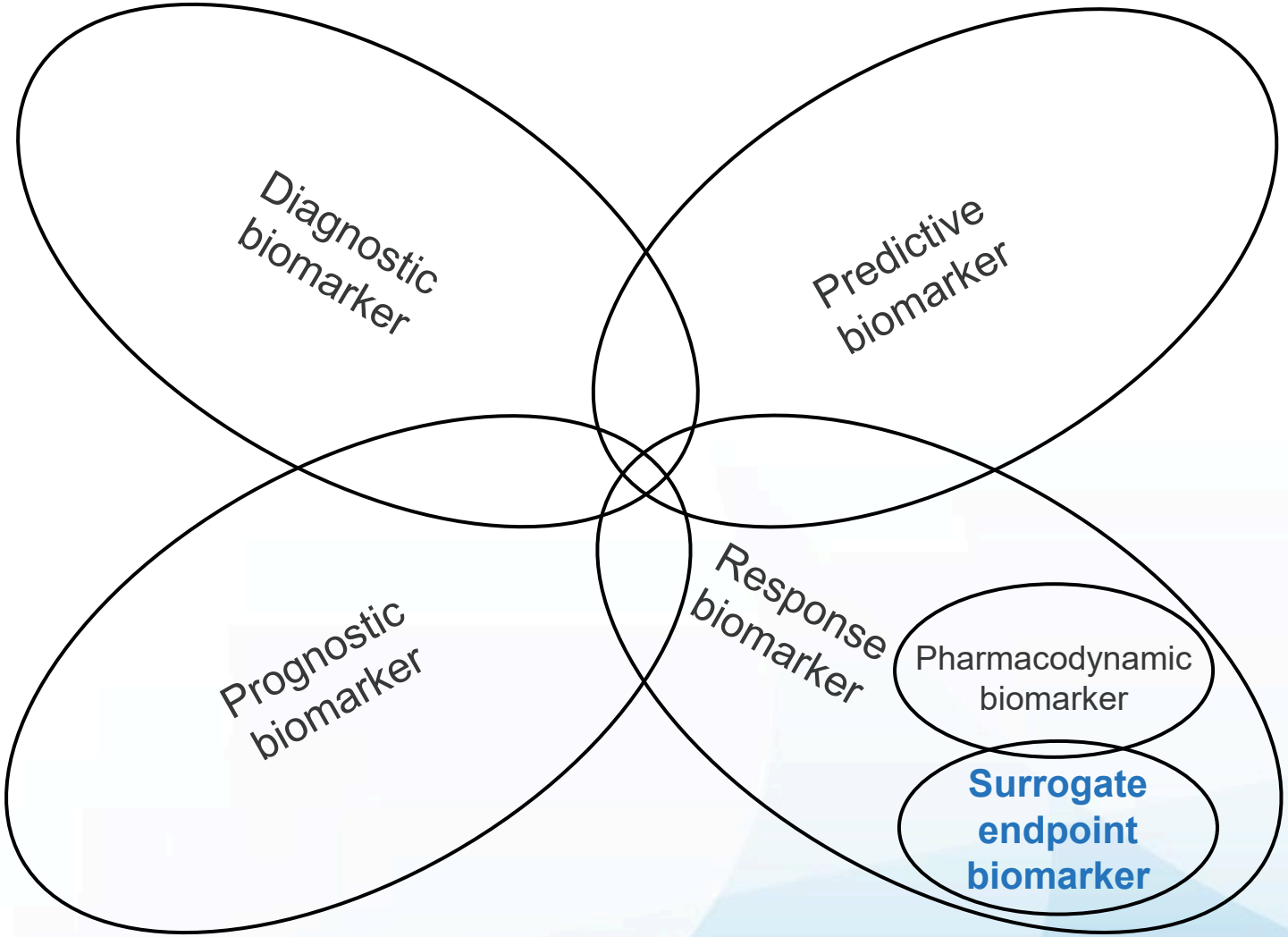
Why do we need surrogate endpoints?

- Shorten study duration
- Reduce sample size
- Lower cost
- Accelerate treatment decision making

Examples of surrogate endpoints

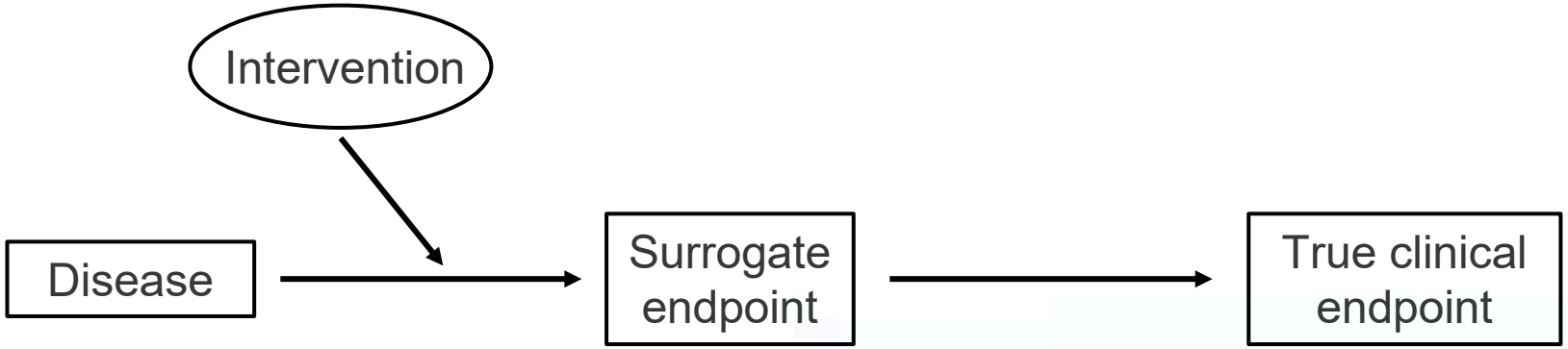
- Tumor shrinkage and progression free survival for cancers
- Hemoglobin A1c (HbA1c) reduction for diabetes mellitus
- HIV-RNA reduction for HIV
- Blood pressure reduction for stroke prevention

Biomarkers in clinical development[1]



[1]BEST (Biomarkers, EndpointS, and other Tools) Resource, FDA-NIH Biomarker Working Group, 2021

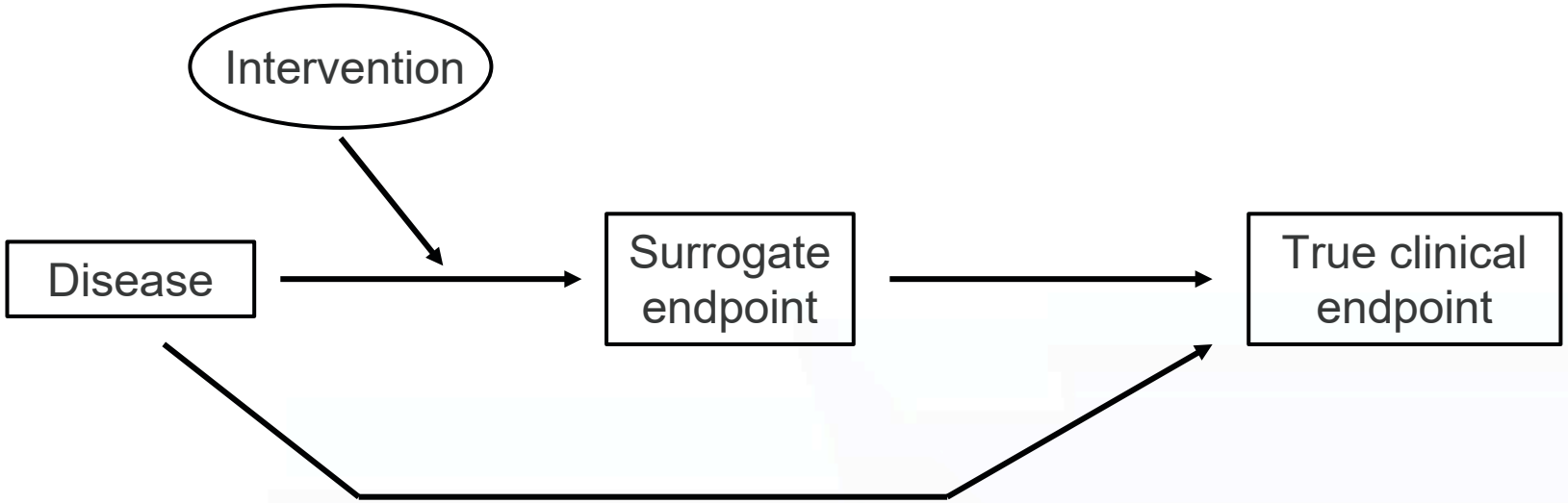
Hypothetical Perfect Case for Surrogate Endpoint[1]



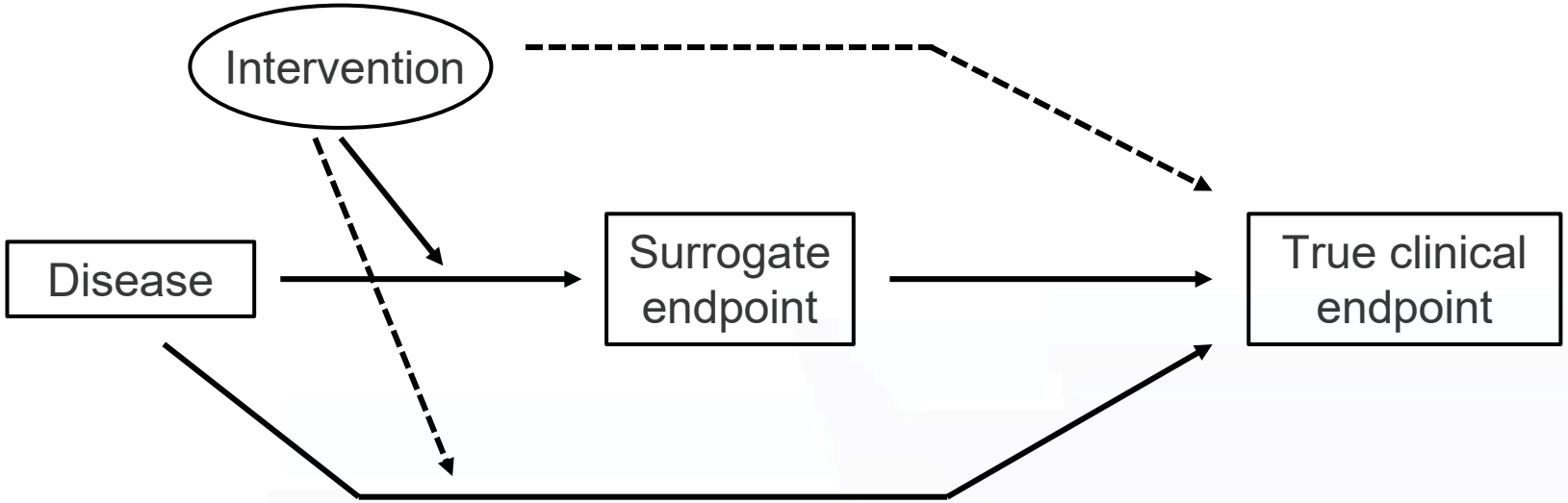
- The surrogate endpoint is on the sole casual pathway of the disease progression
- Drug effect is fully mediated through the surrogate endpoint

[1]Fleming TR and DeMets DL, Annals of Internal Medicine 1996, 125: 605-613

Other possible cases



Other possible cases



History of surrogate endpoint with FDA

- The history of surrogate endpoint started in 1970s accompanied with controversies.
- FDA defined 3 levels of SEs with different levels of clinical validation.

3 levels of surrogate endpoints

Level of Clinical Validation	Feature	Type of Approval	Additional Evidence
Validated	Known to predict clinical benefit with a clear mechanistic rationale	Traditional	No additional efficacy information is required pre- or post- approval
Reasonably likely	Reasonably likely to predict a drug's intended clinical benefit	Accelerated	Additional trial data assessing the effect of the intervention on the clinical benefit of endpoint of interest needs to be collected in the post-marketing setting to verify effect
Candidate	Still under evaluation as there is insufficient evidence	N/A	

(FD&C Act, section 507(e)(9)); BEST (Biomarkers, Endpoints, and other Tools) Resource, FDA-NIH Biomarker Working Group, 2021

History of surrogate endpoint with FDA

- The history of surrogate endpoint started in 1970s accompanied with controversies.
- FDA defined 3 levels of SEs with different levels of clinical validation.
- FDA instituted its Accelerated Approval Program to allow for earlier approval of drugs that treat serious conditions, and that fill an unmet medical need based on a surrogate endpoint.
- FDA maintains a list of surrogate endpoints with approval.
- FDA issued "Considerations for Discussion of a New Surrogate Endpoint(s) at a Type C PDUFA Meeting Request" and encourage early consultation.

Guiding questions on the qualification of surrogate endpoint (SE) from FDA

Relationship of the SE with the Clinical Outcome

- Rationale for Using the SE as a Primary Endpoint
  - What is the clinical outcome the SE is proposed to predict?
  - What is the rationale for using an SE rather than the clinical outcome measure (e.g., feasibility, study duration, sample size, etc.)?
  - What evidence exists to support the relationship between the SE and the clinical outcome of interest (e.g., epidemiologic studies, randomized controlled trials, data generated from therapeutic products from the same class)?
  - …….

Guiding questions on the qualification of surrogate endpoint (SE) from FDA – continued

Relationship of the SE with the Clinical Outcome

- Relationship of the SE with the Causal Pathway(s)
  - What is known about the causal pathway(s) for the intended disease? What is the relationship of the SE to this pathway(s)?
  - Does the intended disease or use have multiple causal pathways? If so, what is the evidence that the specific pathway the SE monitors is the primary pathway leading to the outcome being assessed?
  - ……
- Threshold for Change Required to Demonstrate Clinical Relevance
- Consistency of SE Response under Various Conditions

Notations

- Randomized clinical trials
- Z – treatment indicator (placebo, dose group 1, dose group 2, …)
- S – candidate surrogate endpoint (biomarker)
- T – true clinical endpoint

Prentice criterion[1]

	Criterion	Model	Test
1.	Z have a significant effect on S	$S_i = \mu_s + \alpha Z_i + \epsilon_{S_i}$	$H_0: \alpha = 0$
2.	Z have a significant effect on T	$T_i = \mu_T + \beta Z_i + \epsilon_{T_i}$	$H_0: \beta = 0$
3.	S and T significantly correlated	$T_i = \mu + \gamma S_i + \epsilon_i$	$H_0: \gamma = 0$
4.	The full effect of Z on T explained by S	$T_i = \tilde{\mu}_T + \beta_s Z_i + \gamma_Z S_i + \tilde{\epsilon}_{T_i}$	$H_0: \beta_s \neq 0$

Problem:

- Applicable to a perfect surrogacy case, which is unrealistic
- Impossible to prove the null of $\beta_s \neq 0$ in finite samples
- Does not quantify the predictive ability of a surrogate endpoint

[1]Prentice RL, Statistics in medicine 1989, 8 (4): 431-440

Freedman approach[1]: Proportion of treatment effect explained (PTE)

- An extension to Prentice approach
- Define and estimate the proportion of treatment effect on T explained by S
- Examine the change in the regression coefficient for Z with and without S added

$$T_i = \mu_T + \beta Z_i + \epsilon_{T_i}$$

$$T_i = \tilde{\mu}_T + \beta_s Z_i + \gamma_Z S_i + \tilde{\epsilon}_{T_i}$$

$$PTE = \frac{\beta - \beta_S}{\beta} = 1 - \frac{\beta_S}{\beta}$$

Problem:

- Not a proportion, not bounded between 0 and 1
- Very wide confidence intervals

[1]Freeman LS, Graubard BI and Schatzkin A, Statistics in medicine 1992, 11: 167-178

Principal stratification causal inference approach[1,2]

- Application of principal stratification on surrogate marker evaluation
- The missing biomarker data in the unobserved cells were imputed using baseline variables predictive of the longitudinal biomarker data
- Each subject has the pair of biomarker data under placebo and under treatment after the imputation

Treatment	Patient Index	Biomarker Outcomes	
		$S(\text{pbo})$	$S(\text{trt})$
Placebo	1 $\dots$ m	$s_1\ (\text{pbo})^{\ obs}$ $\dots$ $s_m\ (\text{pbo})^{\ obs}$	Missing
Treatment	$m+1$ $\dots$ n	Missing	$s_{m+1}\ (\text{trt})^{\ obs}$ $\dots$ $s_n\ (\text{trt})^{\ obs}$

[1]Frangakis C and Rubin DB, Biometrics 2002, 58: 21-29; [2]Gilbert PB and Hudgens MG, Biometrics 2008, 64: 1146-1154

Principal strata and principal surrogacy

Principal surrogacy criteria

- Average Causal Necessity:

  No average causal treatment effect on T

  in strata with $S(trt) = S(pbo)$

- Average Causal Sufficiency:

  Beneficial average treatment effect on T

  in strata with $S(trt) > S(pbo)$

Principal Stratum	Definition
Causal Necessity Stratum 00	$S(pbo) = S(trt) = 0$
Causal Sufficiency Stratum 01	$S(pbo) = 0,\ S(trt) = 1$
Doomed Stratum 10	$S(pbo) = 1,\ S(trt) = 0$
Causal Necessity Stratum 11	$S(pbo) = S(trt) = 1$

- Subjects within each stratum share the same $\{S(pbo), S(trt)\}$
- The construction of principal strata is independent of treatment assignment

Problem:

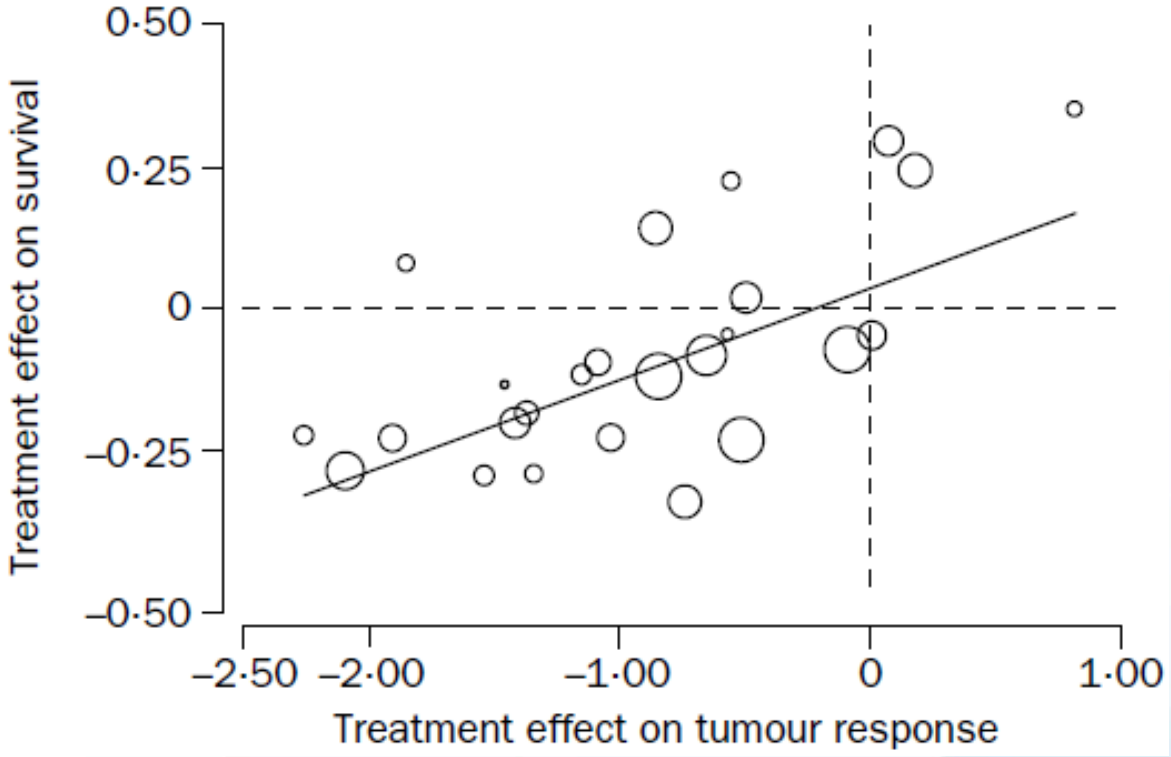
- The principal stratum is not observable in most cases.
- Need a good model to predict the unobserved biomarker response.

Correlation-based approaches

- Subject-level correlation
  - Individual's change from baseline in S vs change from baseline in T
- Group-level correlation
  - Group-mean level treatment benefit on S vs treatment benefit on T from active dose groups

Group-level meta-analysis

Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis[1]



[1]Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G and Piedbois P. Lancet 2000, 356: 373-378

Meta-Analytic Approach[1]

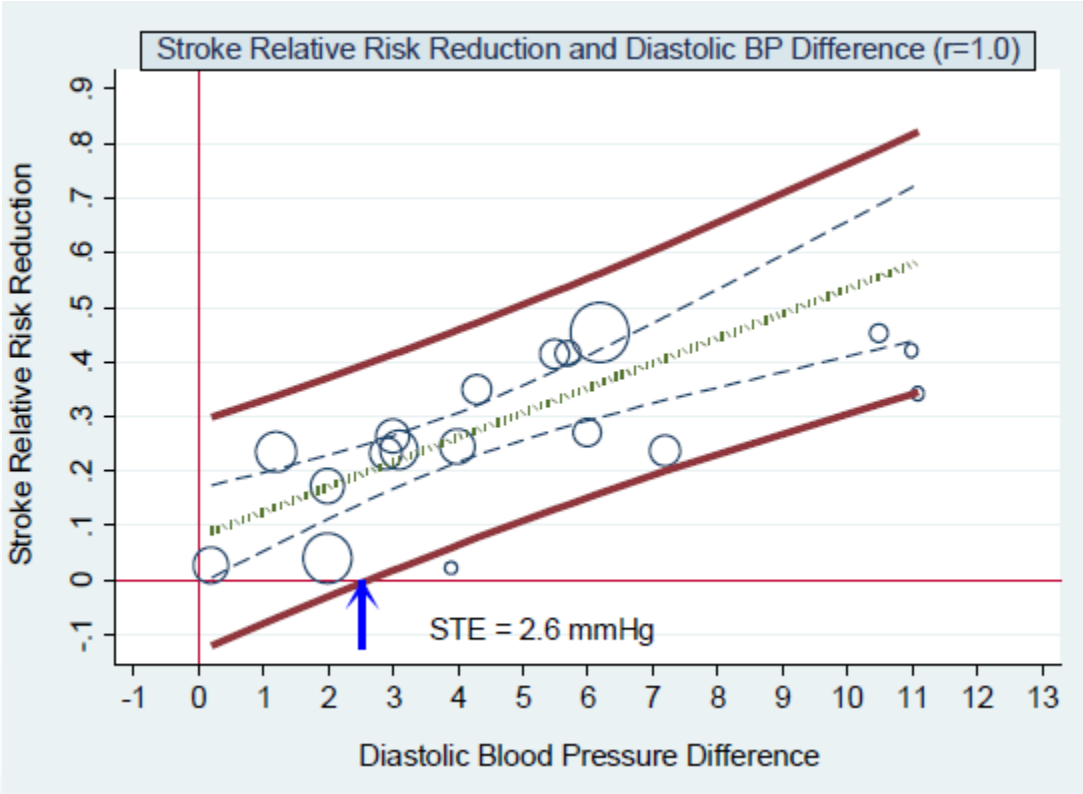
- Model the subject level data of S and T jointly
  - the treatment effects on S and T are estimated after accounting for their covariance
  - the variation of the treatment effects is modelled

- Measure the association of S and T at two levels
  - At individual level, R^2_{ind} quantifies the strength of association between S and T after adjustment for trial and treatment effects
  - At trial level, R^2_{trial} quantifies the strength of association between the treatment effects on S and on T

[1]Buyse M, Molenberghs G, Burzykowski T, Renard D and Geys H, Biostatistics 2000, 1(1):49-67.

Surrogate threshold effect (STE)[1]

- Can be obtained from the meta-analytic approach
- STE is the minimum treatment effect on S necessary to predict a non-zero treatment effect on T.
- STE allows for a more direct clinical evaluation of the appropriateness of a candidate surrogate endpoint.
- The value of STE largely depends on the prediction variance.



Lassere *et al*. BMC Medical Research Methodology 2012, 12:27

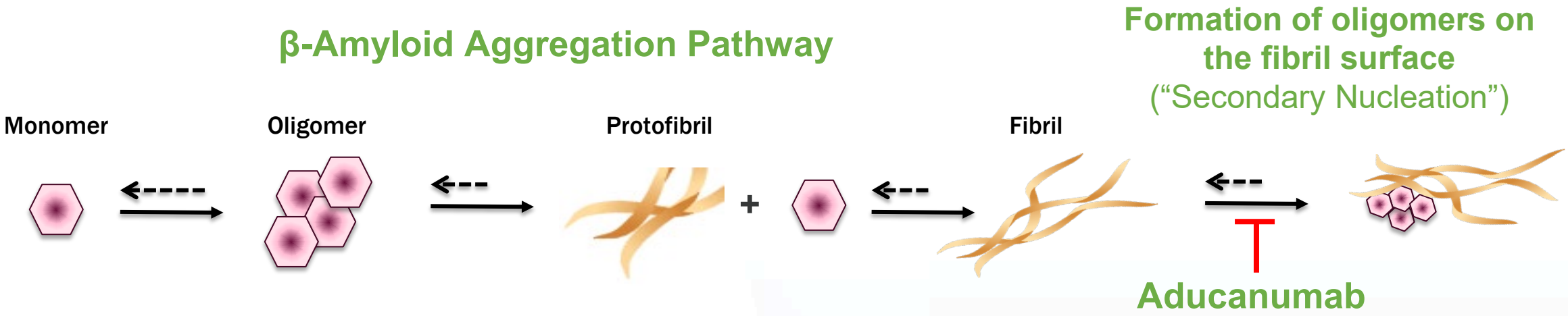
[1]Burzykowski T, Buyse M, Pharmaceutical statistics 2006, 5. 173-86. 10.1002/pst.207

Alzheimer's disease and amyloid cascade hypothesis

- Alzheimer's disease is a progressive neurodegenerative disorder characterized by insidious and unrelenting cognitive and functional decline.

- Amyloid cascade hypothesis – the driving force behind the disease process is the accumulation of Aβ resulting from an imbalance between Aβ production and Aβ clearance in the brain

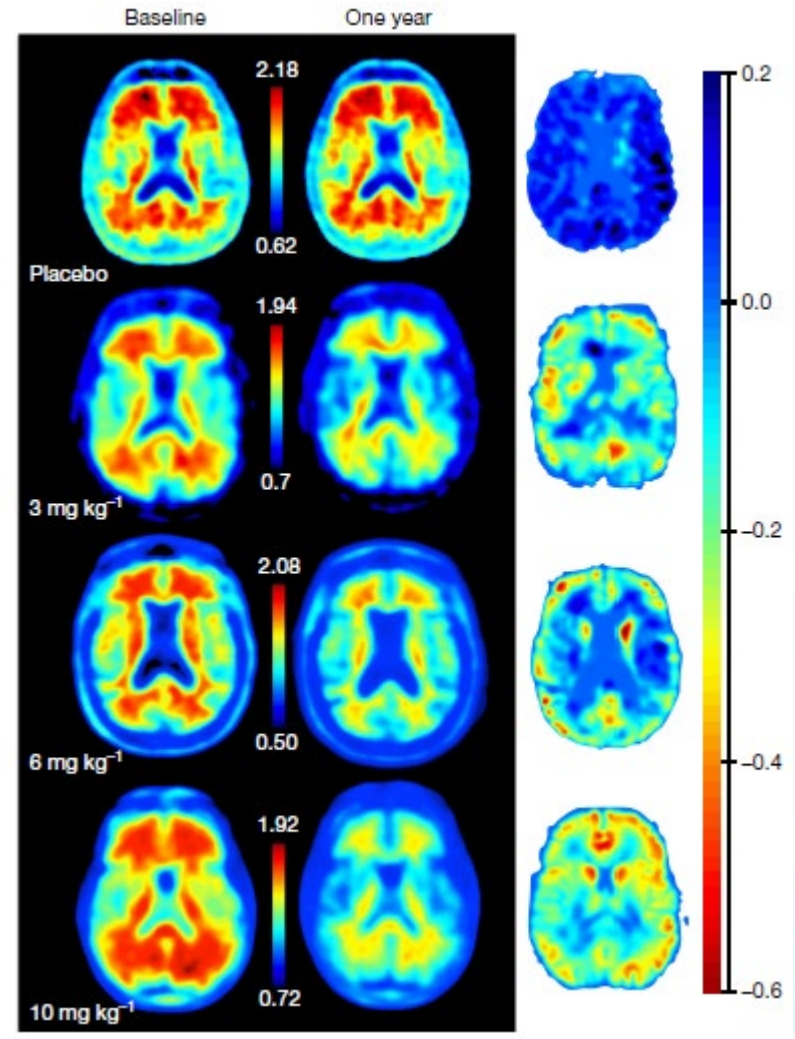
β-amyloid is a key pathological hallmark of Alzheimer's disease



- Aducanumab-avwa is a human monoclonal antibody that is directed at aggregated β-amyloid

Amyloid PET

- Amyloid PET imaging was used to provide:
  - Qualitative assessment (visual interpretation) of brain Aβ plaque at screening
  - Quantitative assessment of the effect of aducanumab on brain Aβ plaque longitudinally



Sevigny J, *et al*. Nature 2016, 537: 50-56

PET = positron emission tomography

Aducanumab study design

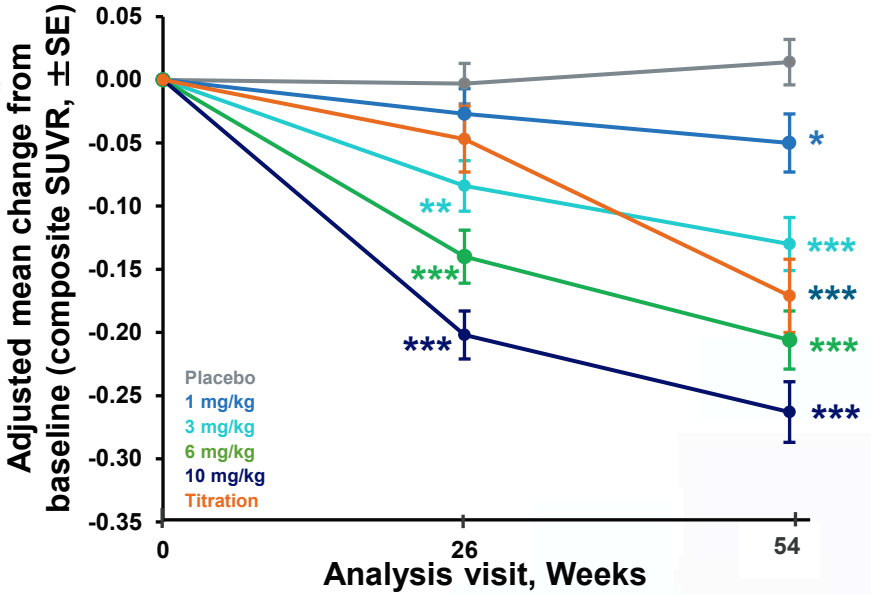
PRIME	A 12-month, randomized, double-blind, placebo-controlled, Phase 1b study
Population	Mild cognitive impairment & mild AD dementia with **positive amyloid PET** at baseline
Doses	Placebo, 1 mg/kg, 3 mg/kg, 6 mg/kg, 10 m/kg fixed dose, 10 mg/kg titration
sample size	197 patients, ~ 30 per arm
Primary endpoint	Safety and tolerability
Other endpoints include	▪ Secondary: **Amyloid PET** ▪ Exploratory: **CDR-SB, MMSE**

EMERGE & ENGAGE	Two 18-month, randomized, double-blind, placebo-controlled, Phase 3 studies
Population	Mild cognitive impairment & mild AD dementia with **positive amyloid PET** at baseline
Doses	Placebo, low dose (3 or 6 mg/kg depending on ApoE), high dose (10 mg/kg)
sample size	3285 patients, ~500 per arm
Primary endpoint	**CDR-SB** at 18 months
Other endpoints include	▪ Secondary: 3 other clinical endpoints ▪ Tertiary: **amyloid PET, correlation between CDR-SB and amyloid PET**

PET = positron-emission tomography; CDR-SB = Clinical Dementia Rating–Sum of Boxes

Aducanumab reduced amyloid PET in a dose- and time-dependent manner

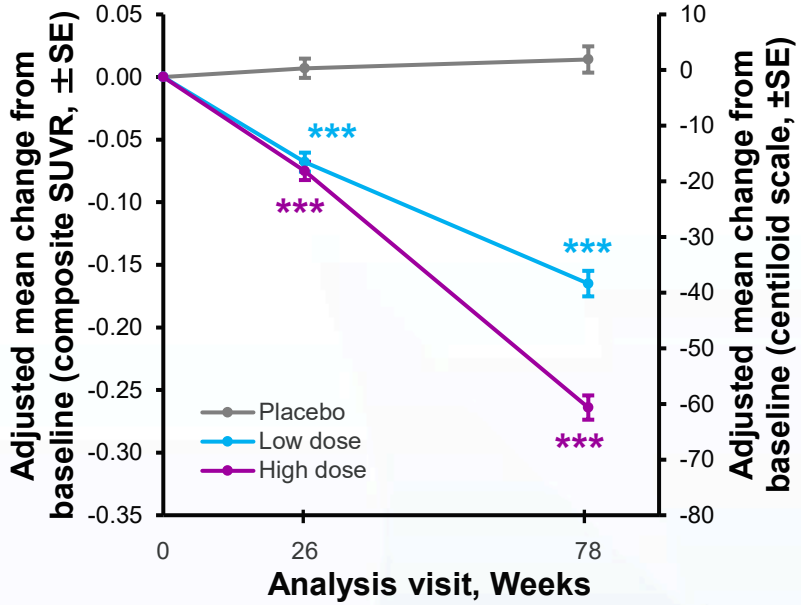
PRIME[1]



	1 mg/kg	3 mg/kg	6 mg/kg	10 mg/kg	Titration to 10 mg/kg
Difference from Placebo (Week 54)	-0.064	-0.145	-0.220	-0.277	-0.185

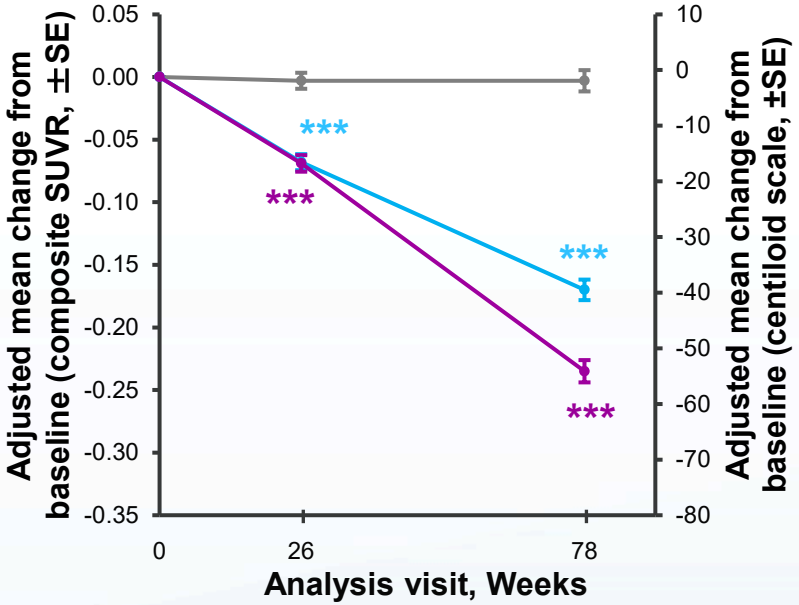
*p<0.05, **p<0.01, *** p<0.001 (nominal)

EMERGE[2,3]



Placebo	n=159	129	93
Low-dose adu	n=159	129	100
High-dose adu	n=170	138	109

ENGAGE[2,3]



	n=204	168	124
	n=198	169	138
	n=183	156	112

1. Budd Haeberlein S, et al. Data presented at CCFDIE 2021; 2. Budd Haeberlein S, et al. *J Prev Alzheimers Dis*. 2022;9(2):197-210. Figure 1; 3. Budd Haeberlein S, et al. Data presented at ADPD 2021.

Validation of amyloid reduction as a surrogate endpoint in Alzheimer's disease

- Multiple statistical frameworks were undertaken to assess the relationship between amyloid reduction and clinical outcome:
  - Approaches based on a single study as introduced
  - Meta-analytic approach using subject-level data across aducanumab studies
  - Group-level correlation leveraging data from other anti-amyloid compounds with similar mechanism
- Various analyses coherently point in the same direction: both within-study and between-study associations link amyloid reduction and clinical benefit.
- Will introduce the correlation-based analyses at 2 levels with learning points in this presentation

Biomarkers at different disease stage have unique features[1]

	Class I biomarker Example: Amyloid PET	Class II biomarker Example: Tau PET
Overview	• Captures pathological changes at earliest stage of the disease process (prior to symptom manifestation)	• Captures pathological changes during symptom manifestation of the disease process
Trajectory	• Rate of change highest prior to onset of clinical symptoms • Change during clinical presentation limited due to plateau	• Rate of change highest following onset of clinical symptoms
Biomarker progression	• Minimal progression during a typical AD clinical trial window	• Moderate progression during a typical AD clinical trial window

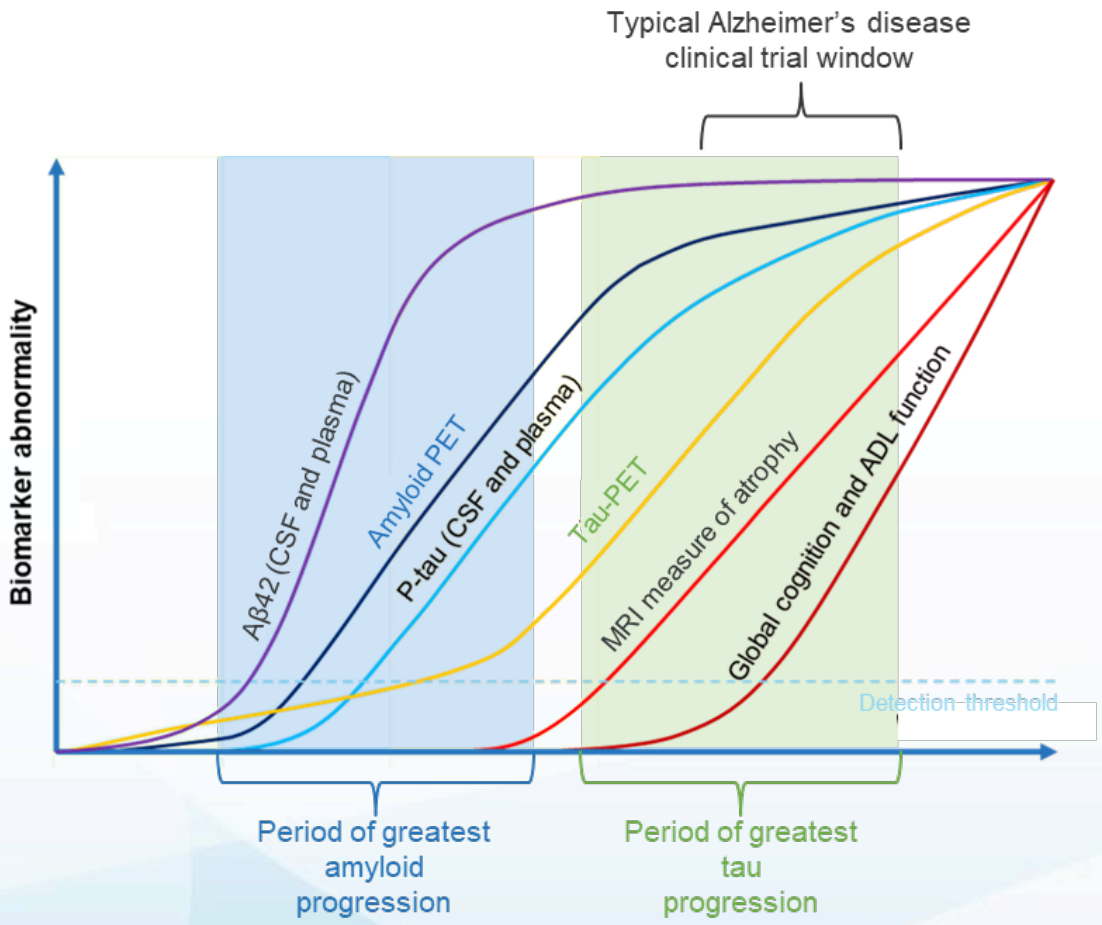
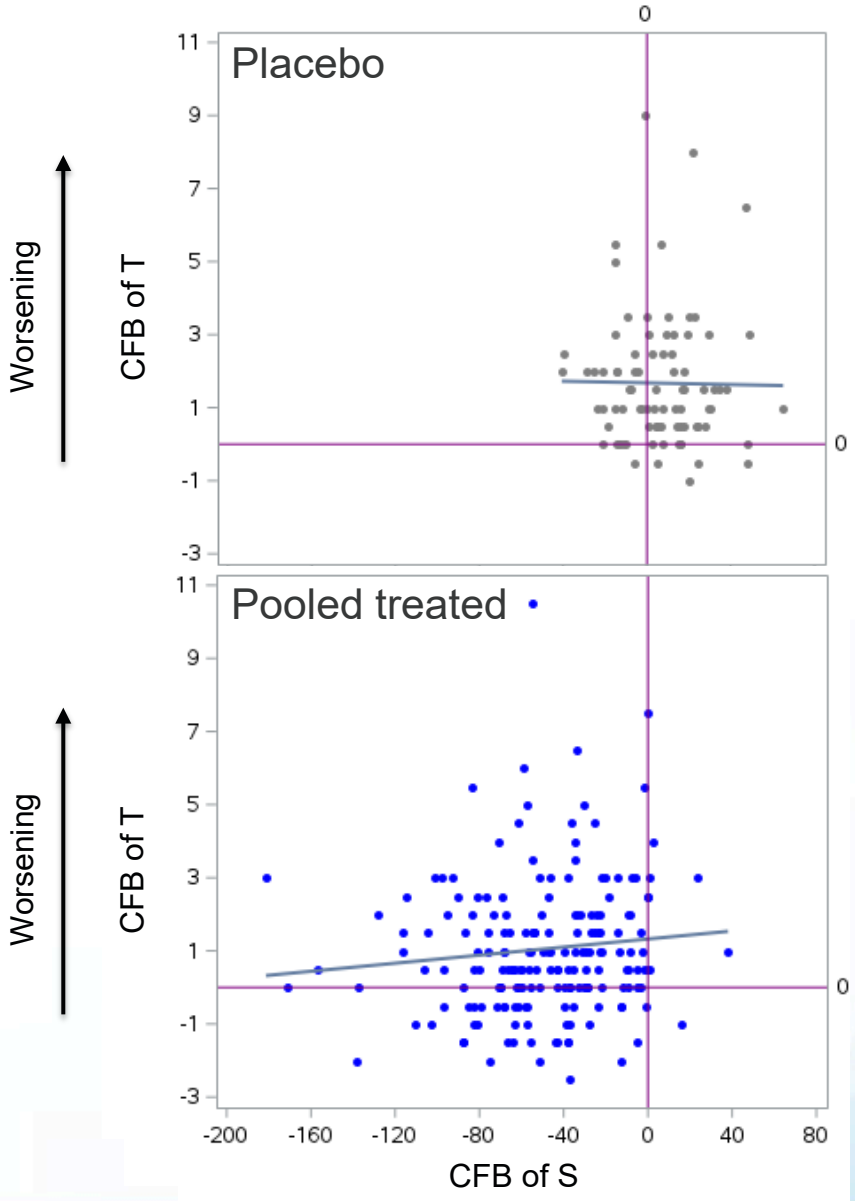


Figure adapted from Hansson O. *Nat Med*. 2021;27:954-963.

[1]Chen T, et al, data presented at ADPD 2023. Aβ = amyloid beta; ADL = activities of daily living; CSF = cerebrospinal fluid; PET = positron emission tomography

Class I Biomarker[1]



Placebo Arm Features:

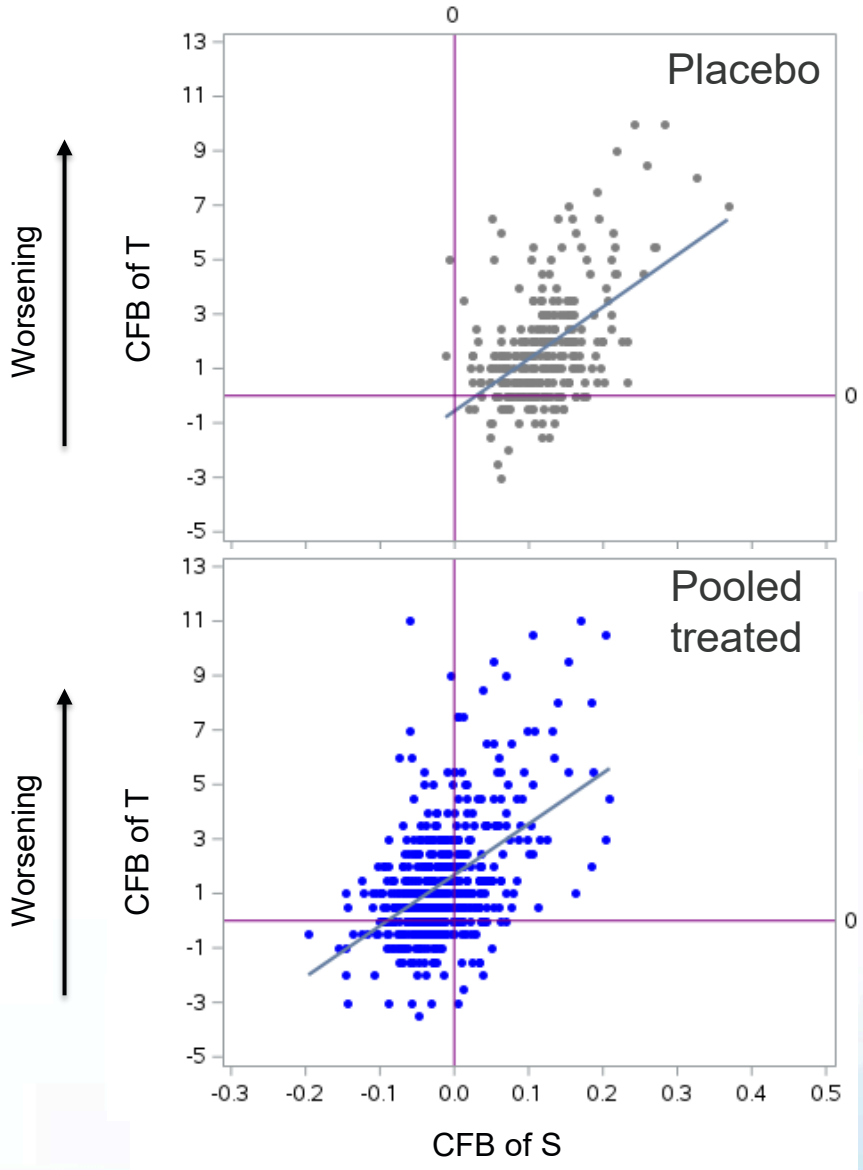
- S level remains stable (CFB scatter around 0)
- T progresses/worsens naturally (CFB mostly above 0)
- No correlation in placebo arm as expected

Treated Arm Features:

- S level reduced from baseline (CFB shift to left, mostly below 0)
- T progresses slower compared to placebo (CFB shift towards bottom)
- **Treatment-induced** correlation observed in treated subjects

[1]Chen T, et al, data presented at ADPD 2023. CFB = change from baseline. Least-square regression lines are displayed.

Class II Biomarker[1]



Placebo Arm Features:

- S accumulates naturally (CFB mostly above 0)
- T progresses naturally (CFB mostly above 0)
- Positive correlation observed

Treated Arm Features (one possible case):

- Reduced accumulation of S compared to placebo (CFB shift towards left)
- Slower progression of T compared to placebo (CFB shift towards bottom)
- Similar positive correlation as in placebo observed

Other possible cases exist!

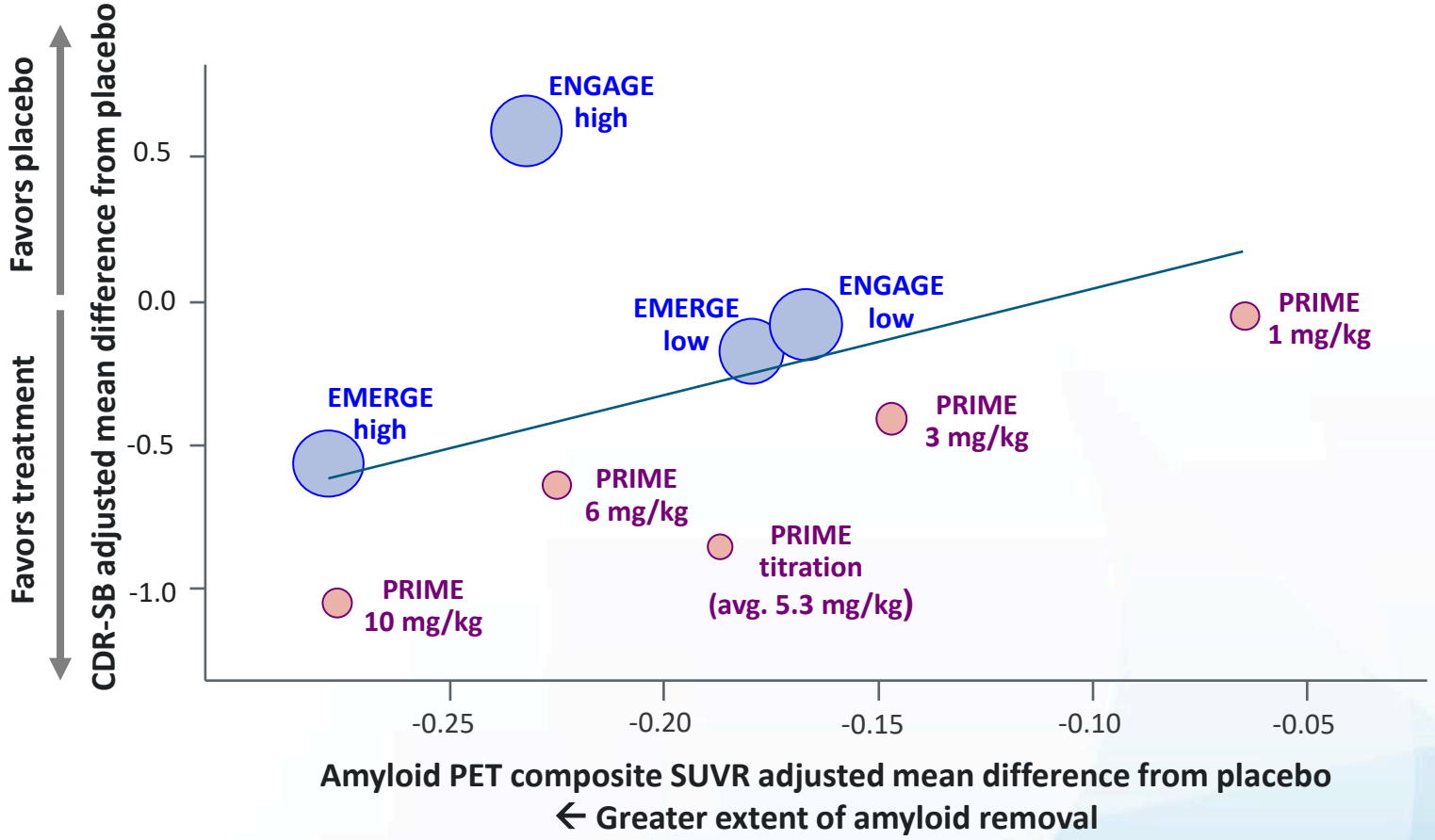
[1]Chen T, et al, data presented at ADPD 2023. CFB = change from baseline. Least-square regression lines are displayed.

Subject-level correlation

- Interpretation varies among biomarkers at different disease stage.
- For most Class II biomarkers, a correlation between the CFBs of 2 endpoints can exist purely as a biological relationship independent of treatment.
- For Amyloid PET as a Class I biomarker, the treatment-induced correlation provides evidence of association between the treatment effect on amyloid (amyloid reduction) and the clinical benefit (slower clinical worsening).

- Limitations:
  - True correlation might be obscured due to the narrow range of CFB (eg, single dose study) and large subject heterogenicity in clinical measures.
  - It may not be conducted appropriately.

Group-Level Correlation – Aducanumab Studies

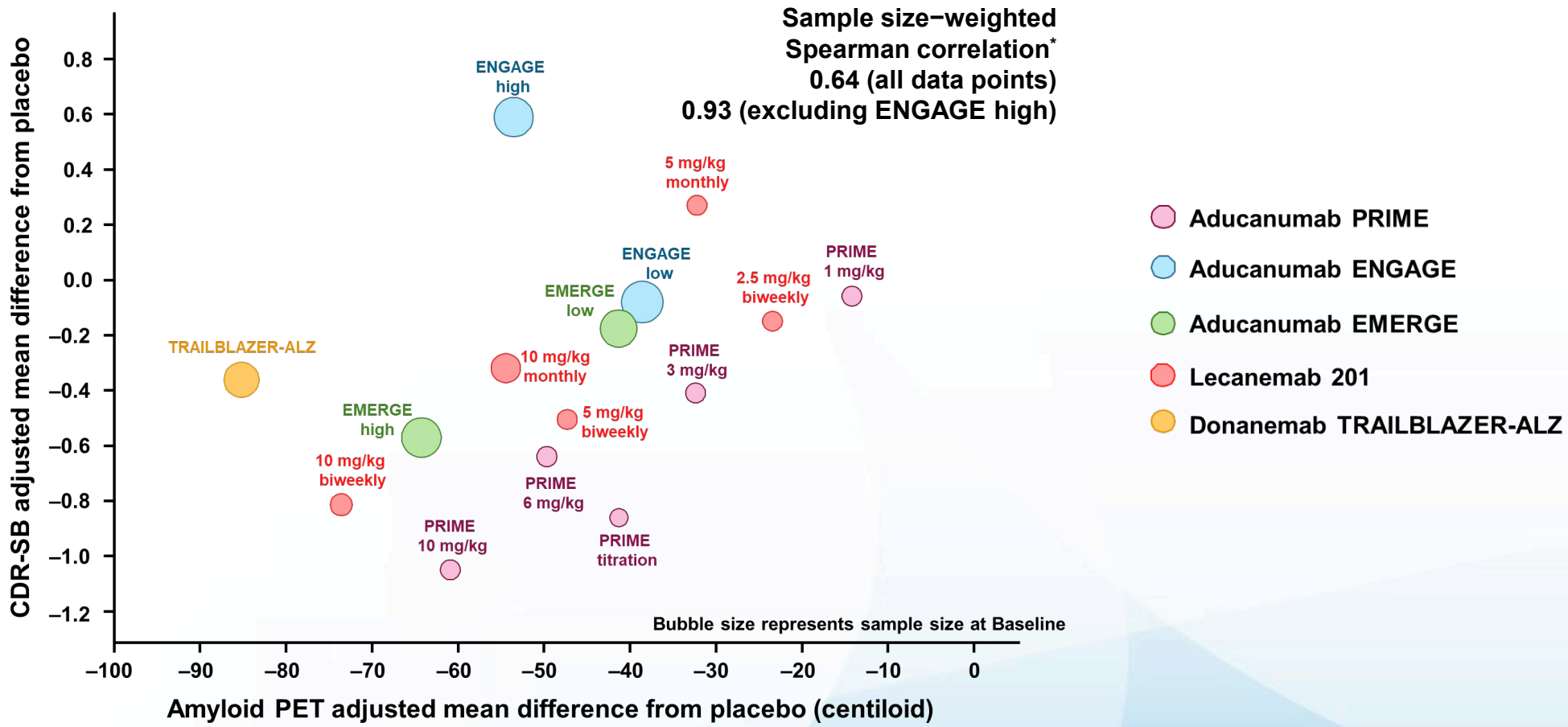


Sample-size weighted Spearman correlation	0.47 (all data points) 0.94 (excluding ENGAGE high)

- Bubble size represents sample size at baseline.
- Sample size weighted regression line using Study PRIME, ENGAGE [low dose] and EMERGE data

Aβ=amyloid beta; CDR-SB=Clinical Dementia Rating Scale-Sum of Boxes; PET=positron emission tomography; SUVR=standardized value uptake ratio

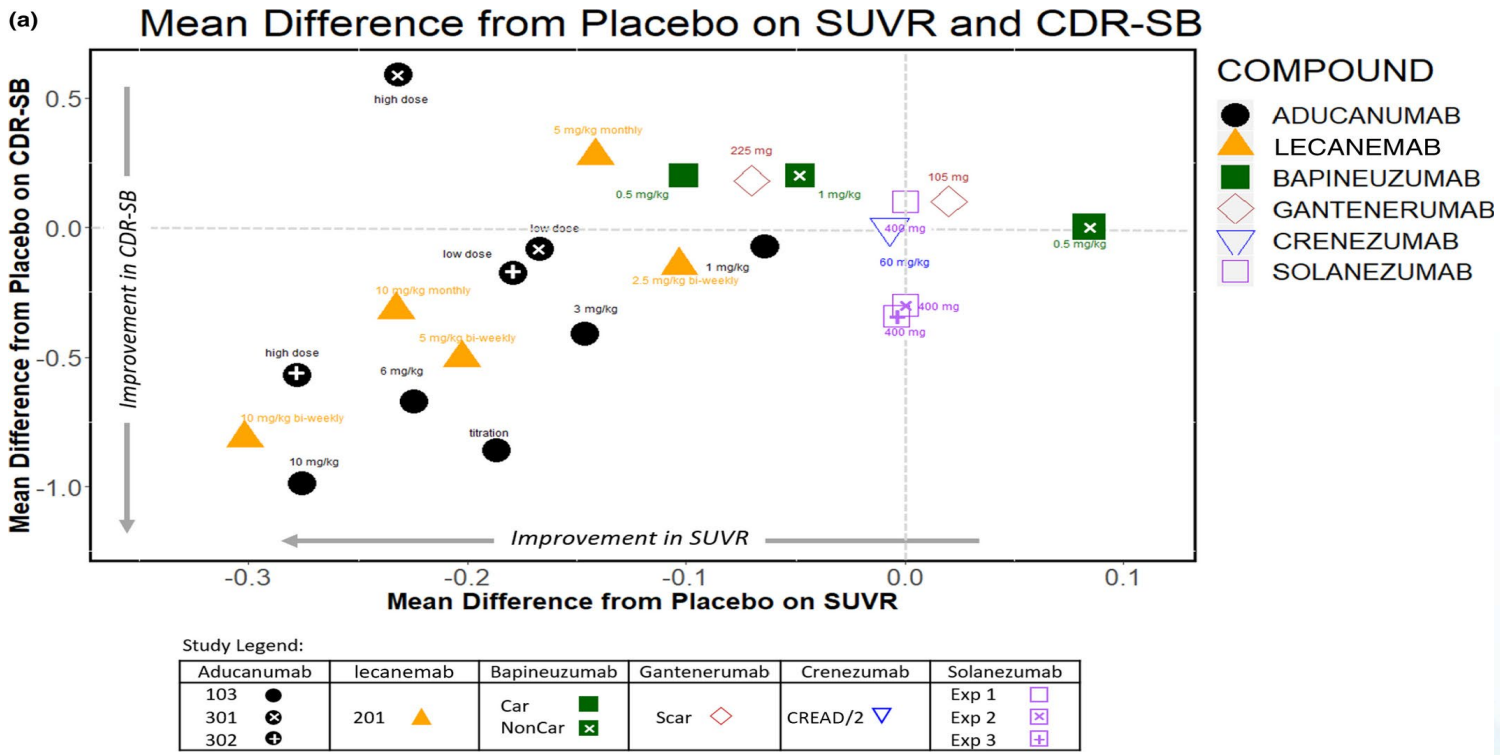
Group-Level Correlation – Second generation Anti-Aβ Drugs (at the time of accelerate approval)[1]



*Sample-size weighted partial Spearman correlation adjusting for study indicator of Aducanumab PRIME, Aducanumab phase 3, Donanemab, and Lecanemab. Sample size and clinical results are based on the (sub)population with amyloid PET assessments.

[1]Chen T, et al, data presented at CTAD 2023. Aβ = amyloid beta; PET = positron emission tomography; CDR-SB = Clinical Dementia Rating–Sum of Boxes

Group-Level Correlation – FDA analysis[1]



Study Legend:

Aducanumab	lecanemab	Bapineuzumab	Gantenerumab	Crenezumab	Solanezumab
103 ●	201 ▲	Car ■	Scar ◇	CREAD/2 ▽	Exp 1 □
301 ⊗		NonCar ⊠			Exp 2 ⊠
302 ⊕					Exp 3 ⊞

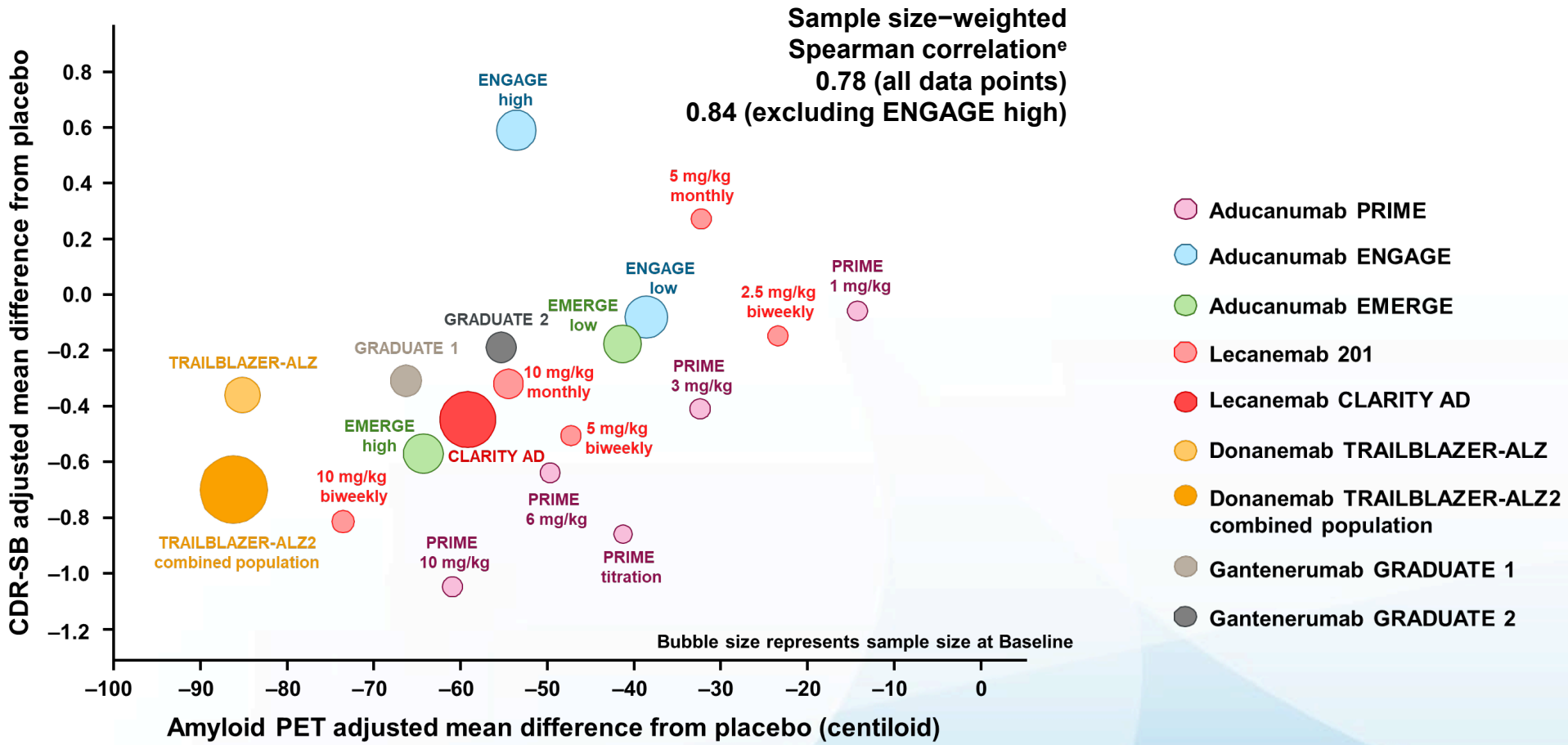
Notes:

[1] Mean baseline-corrected, placebo-adjusted values used to allow for cross-program comparison

[2] Donanemab also shows similar results in centiloids (CL)

[1]Zhu H, Mehta M, Huang SM and Wang Y. Clinical Pharmacology & Therapeutics 2022. 111(4): 728-731. Aβ=amyloid beta; CDR-SB=Clinical Dementia Rating scale Sum of Boxes; PET=positron emission tomography; SUVR=standardized value uptake ratio

Group-Level Correlation – Second generation Anti-Aβ Drugs (with most up-to-date information)[1]



*Sample-size weighted partial Spearman correlation adjusting for study indicator of Aducanumab PRIME, Aducanumab phase 3, Donanemab, Lecanemab and Gantenerumab. Sample size and clinical results are based on the (sub)population with amyloid PET assessments, except for Lecanemab CLARITY AD and Gantenerumab GRADUATE 1 and 2 where ITT clinical results were used.

[1]Chen T et al, data presented at CTAD 2023. Aβ = amyloid beta; PET = positron emission tomography; CDR-SB = Clinical Dementia Rating–Sum of Boxes

Group-level correlation

- Leverage the fundamental aspects of randomized placebo-controlled trials:
  - Directly associate the treatment benefit of biomarker with the treatment benefit of clinical
  - Individual heterogeneity addressed by randomization and the use of model-adjusted group means.
- Apply to various biomarkers, regardless of different trajectories
- Apply to different types of endpoints (continuous, count, time-to-event, etc)
- Require a lot of data from trials with clinical effect!

Reflections from Alzheimer's disease example

- There is no single golden standard approach for validating surrogate endpoint. Consistent results from multiple statistical approaches provide compelling substantiating evidence for validating the surrogate endpoint.
- Biomarkers in Alzheimer's disease follow various trajectories over the disease process, each requiring special statistical considerations.
- Group-level correlation analysis is an appropriate way for validating surrogate endpoint. It is instrumental in Alzheimer's disease given the emerging data from multiple clinical trials and the use of a standardized scale for the surrogate endpoint, which allowed cross-program comparison.

A Causal Mediation Model to Evaluate the Individual Surrogacy of Plasma Neurofilament Light Chain (NfL) for SOD1 ALS using Tofersen Data

- Background: Accelerated approval of tofersen for the treatment of *SOD1*-ALS
- Causal inference model for individual surrogacy
- Conclusion

Tofersen

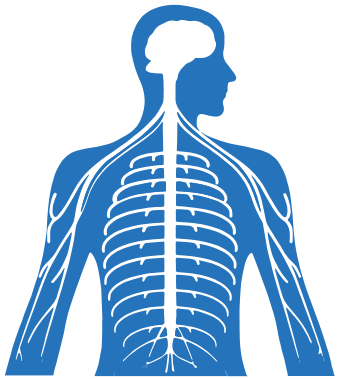
USPI Indication Statement

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.

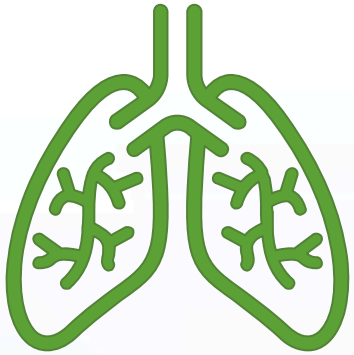
Continued approval for this indication may be contingent upon verification of clinical benefit in confirmatory trials.

ALS is 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

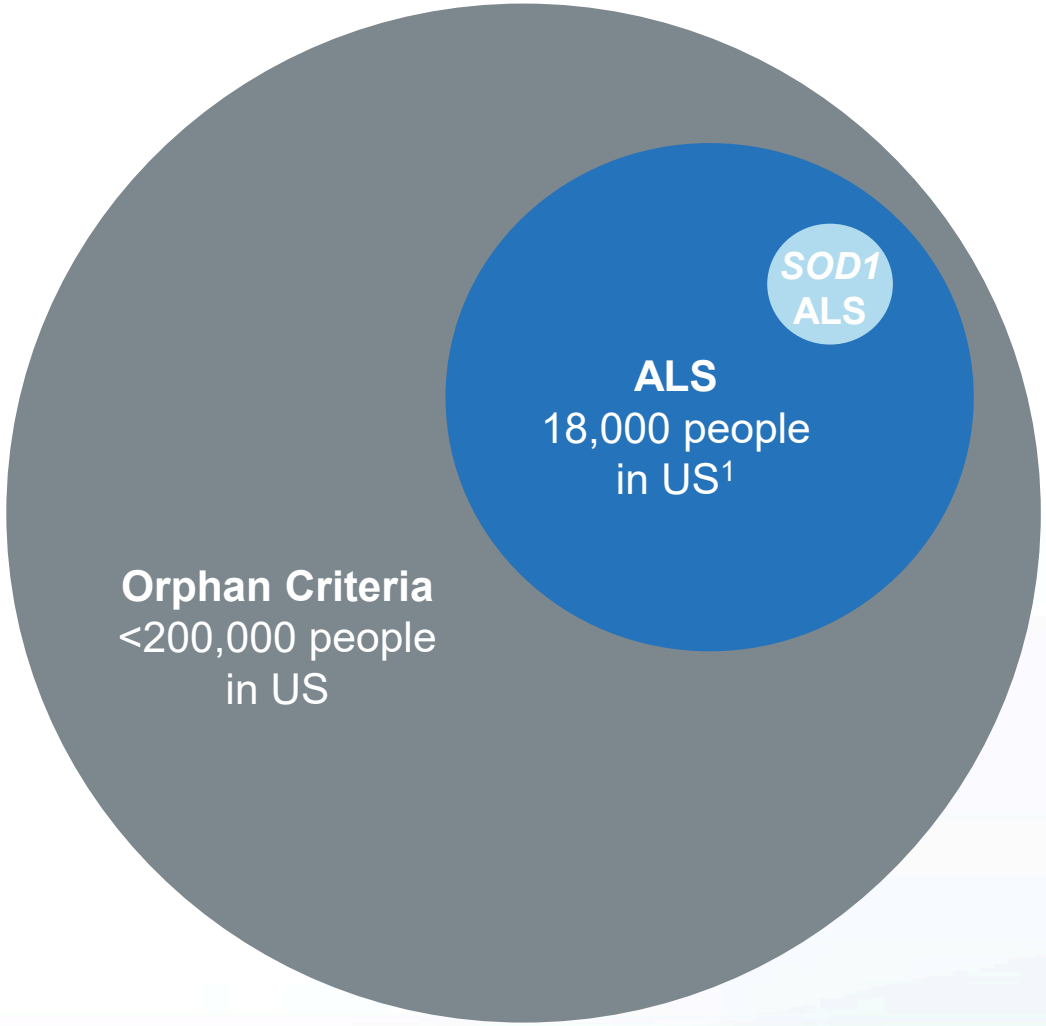
3 to 5 years

from symptom onset[2]

ALS, amyotrophic lateral sclerosis

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.

SOD1-ALS is a rare, progressive, and fatal disease



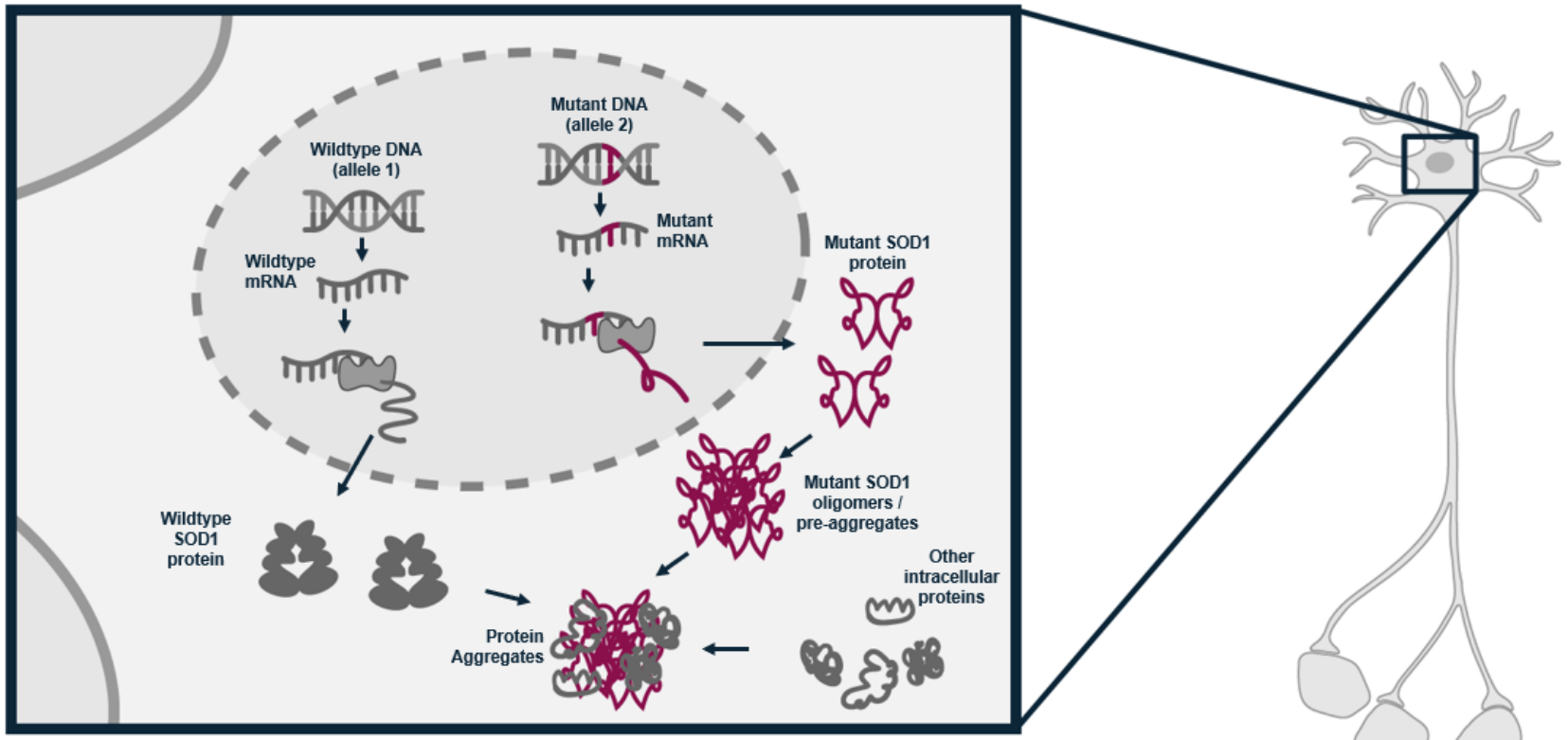
SOD1-ALS

- Caused by a mutation in the superoxide dismutase-1 (*SOD1*) gene
- Affects ~330 people in the US[2,3]
- Median survival 2.7 years from diagnosis[4]

Remains a progressive, fatal disease with a high unmet medical need

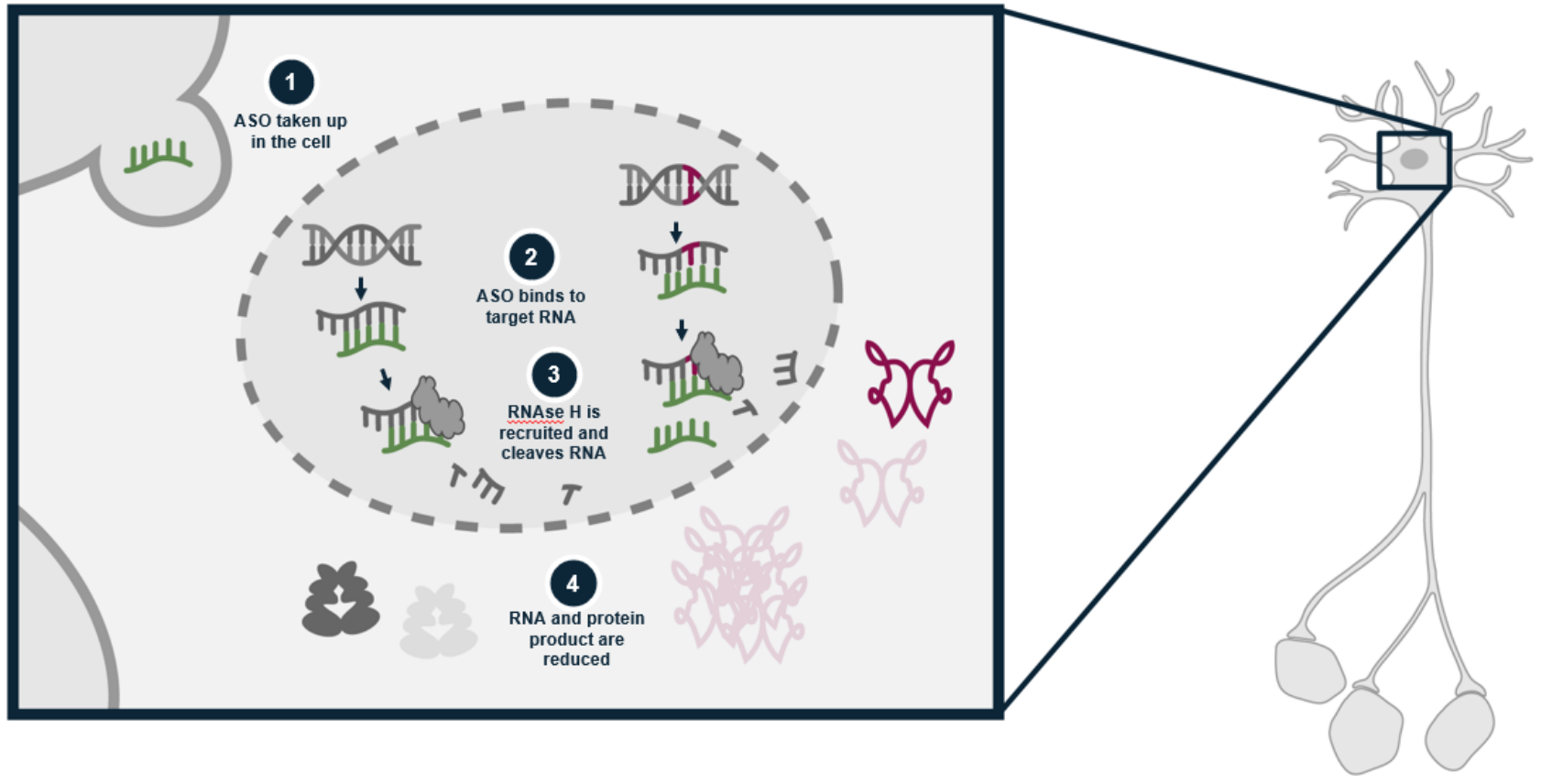
1. Mehta P, et al. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration.* 2023;24:108-116. 2. Zou ZY, et al. *J Neurol Neurosurg Psychiatry.* 2017;88:540-549. 3. Brown CA, et al. *Neuroepidemiology.* 2021;55:342-353. 4. Bali T, et al. *J Neurol Neurosurg Psychiatry.* 2017;88:99-105.

Mutations in the *SOD1* gene lead to production of a mutated form of SOD1 protein



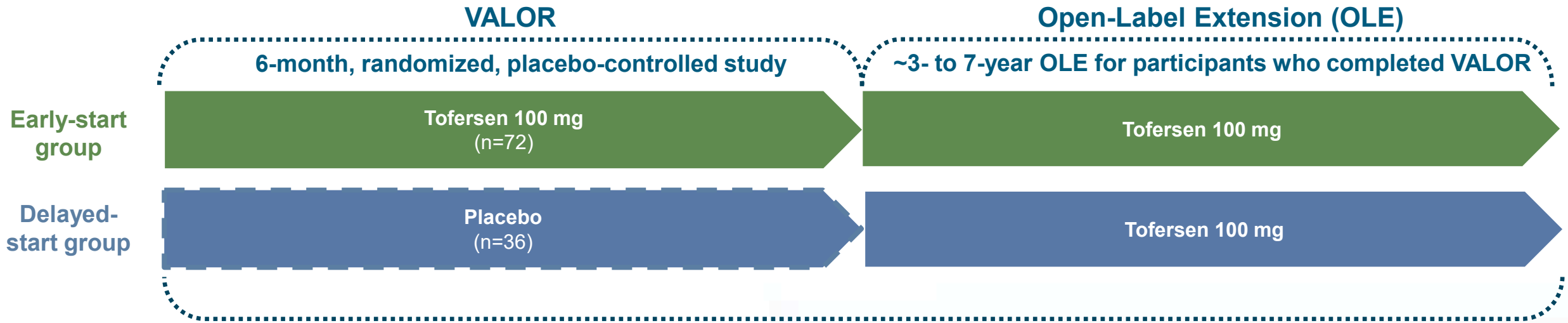
Based on Robberecht W, Philips T. *Nat Rev Neurosci.* 2013;14:248-264.

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.

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

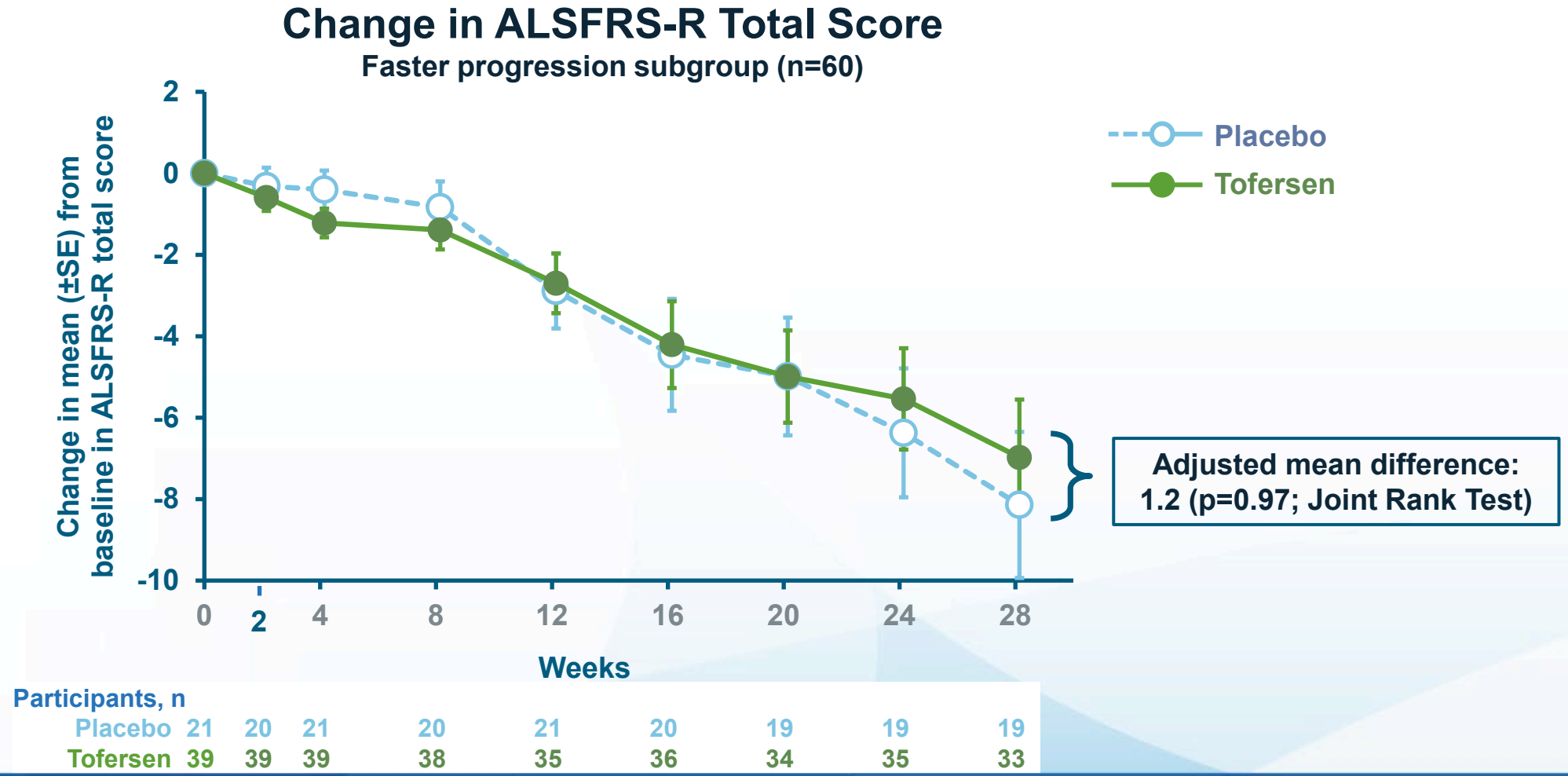
- ALSFRS-R total score

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

Statistical significance was not achieved on the primary analysis in VALOR

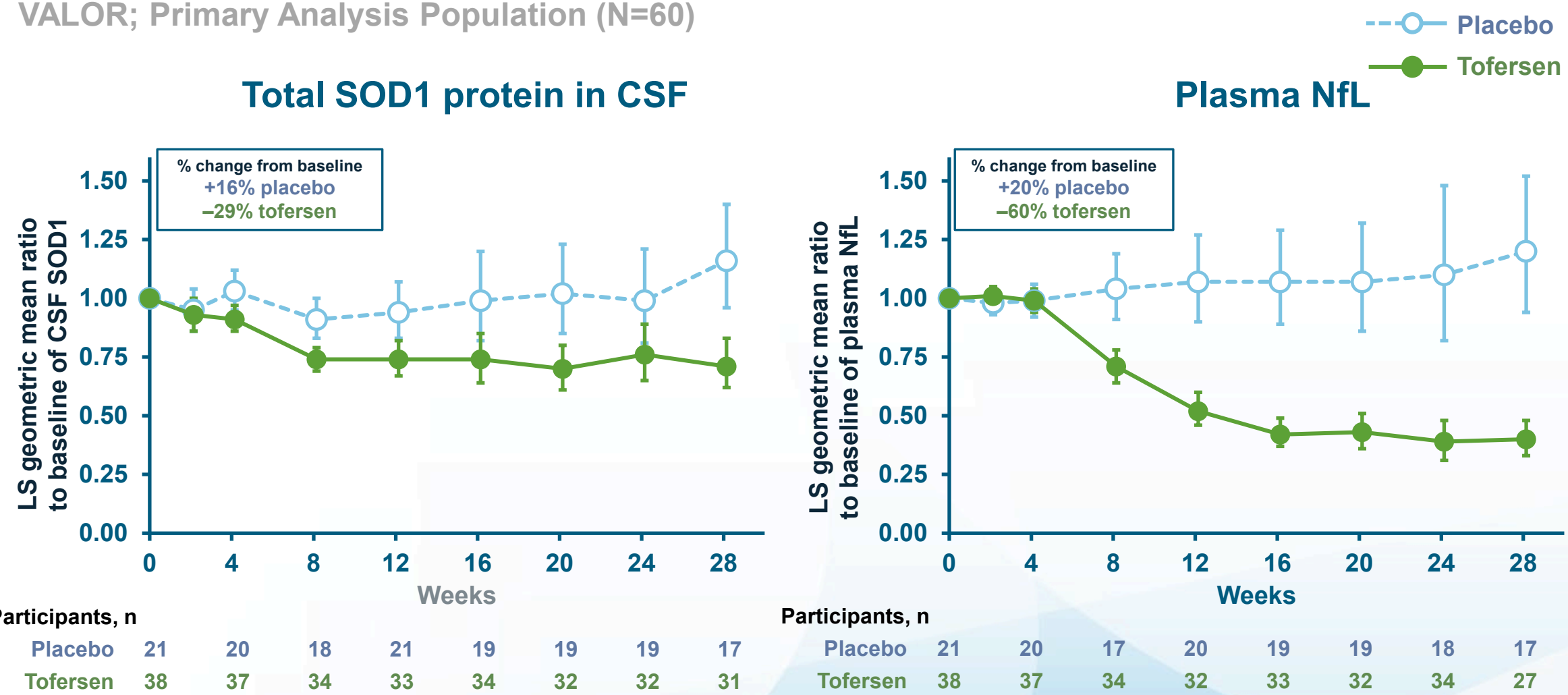
VALOR; Primary Analysis Population (N=60)



Source: VALOR ALSFRS excel data 10.0

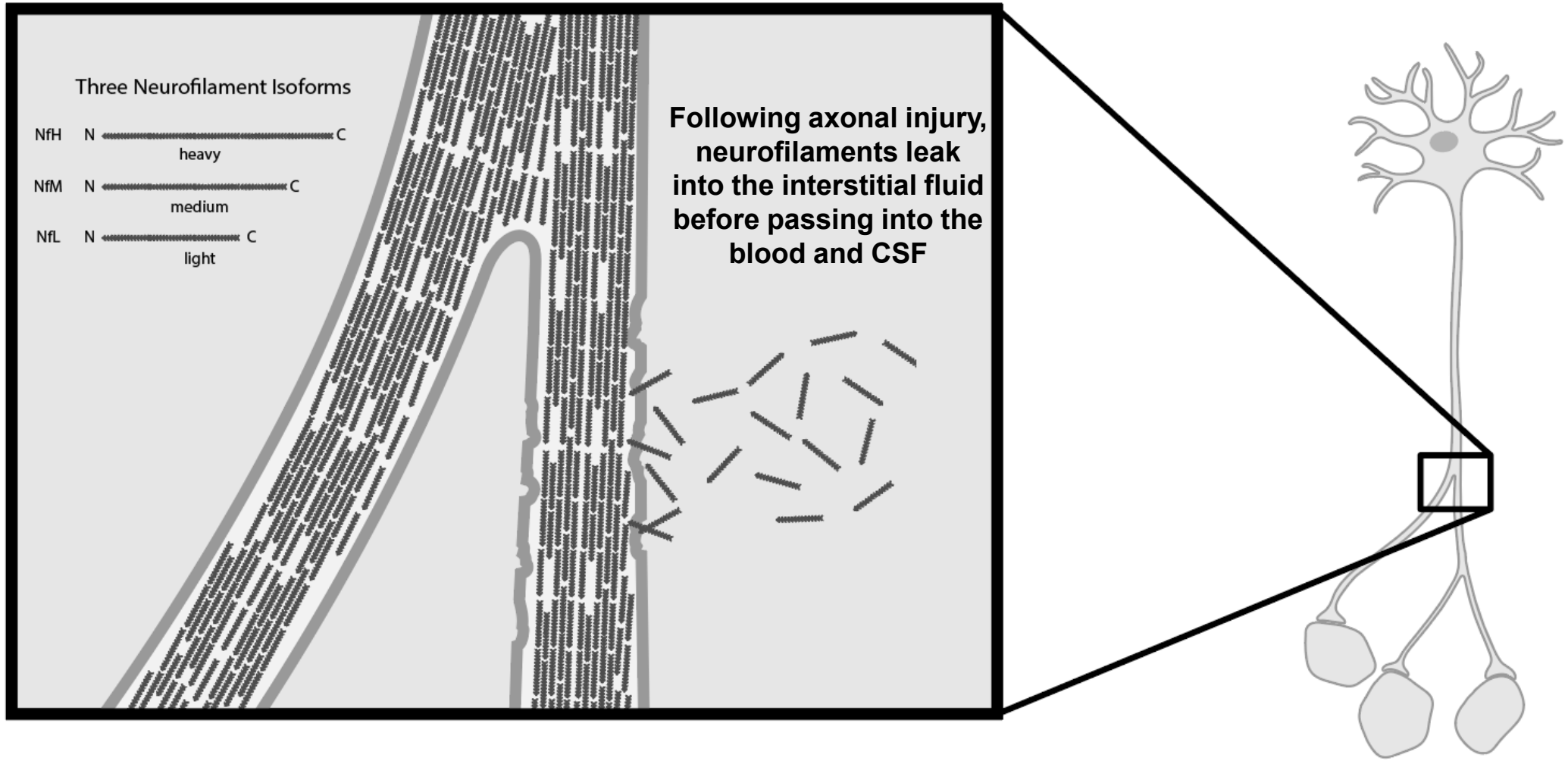
Tofersen levels of CSF SOD1 and plasma NfL

VALOR; Primary Analysis Population (N=60)



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

Neurofilaments are a marker of motor neuron integrity



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

The behavior of neurofilament is well characterized in the ALS literature

1990	2000	2010	2020		
Rosengren 1996	Norgren 2003	Ganesalingam 2011	Poesen 2017	Abu-Rumeileh 2020	Haji 2022
	Brettschneider 2006	Steinacker 2011	Steinacker 2017	Benatar 2020	Halbgebauer 2022
	Zetterberg 2007	Tortelli 2012	Andrés Benito 2018	Delaby 2020	Heckler 2022
	Boylan 2009	Boylan 2013	Benatar 2018	De Schaepdryver 2020	Kmezic 2022
	Rejin 2009	Gaiottino 2013	De Schaepdryver 2018	Huang 2020	Masrori 2022
		Ganesalingam 2013	Feneberg 2018	Khalil 2020	Sferruzza 2022
		Lehnert 2014	Gong 2018	Sugimoto 2020	Shi 2022
		Gonçalves 2015	Illán-Gala 2018	Sun 2020	Thompson 2022
		Lu 2015	Khalil 2018	Thouvenot 2020	Yildiz 2022
		McCombe 2015	Li 2018	Yang 2020	Zecca 2022
		Menke 2015	Rossi 2018	Behzadi 2021	Zhang 2022
		Tortelli 2015	Scarafino 2018	Bjornevik 2021	De Shaepdryver 2023
		Chen 2016	Schreiber 2018	Brodovitch 2021	Meyer 2023
		Oeckl 2016	Bridel 2019	Gagliardi 2021	Smith 2023
		Steinacker 2016	De Schaepdryver 2019	Kojima 2021	
		Weydt 2016	Forgrave 2019	Simonini 2021	
		Wilke 2016	Gille 2019	Vacchiano 2021	
		Xu 2016	Kasai 2019	Verde 2021	
		Gaiani 2017	Verde 2019a	Zhou 2021	
		Gendron 2017	Verde 2019b	Escal 2022	
		Kaiserova 2017		Falzone 2022	

- **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**

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

VALOR; ITT population (N=108)

	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.

Factors affecting the primary analysis [1]

- Mechanisms to control for disease heterogeneity based on clinical features (mutation type and/or pre-randomization ALSFRS-R slope for defining the primary analysis population, N=60) is not adequate
- Longer study duration (>6 months) is needed to:
  - Reliably detent a decline in the control arm
  - Allow sufficient time for biological activity to translate to clinical benefit
- Baseline imbalances in NfL: NfL are prognostic for decline in clinical function and survival in ALS
- The primary analysis does not include baseline NfL as a covariate
- The joint rank test is sensitive to imbalance in death when the number of death is small and not informative of the underlying survival distributions

Factors affecting the primary analysis [2][1]

- The VALOR study design was typical of many Phase 3 ALS trials, with a 28-week duration and an acceptable primary endpoint ALSFRS-R.

- The selection of enrichment factors was based on available scientific knowledge at the time the study was designed and pre-randomization slope on ALSFRS-R is commonly used as an enrichment criteria in ALS trials.

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

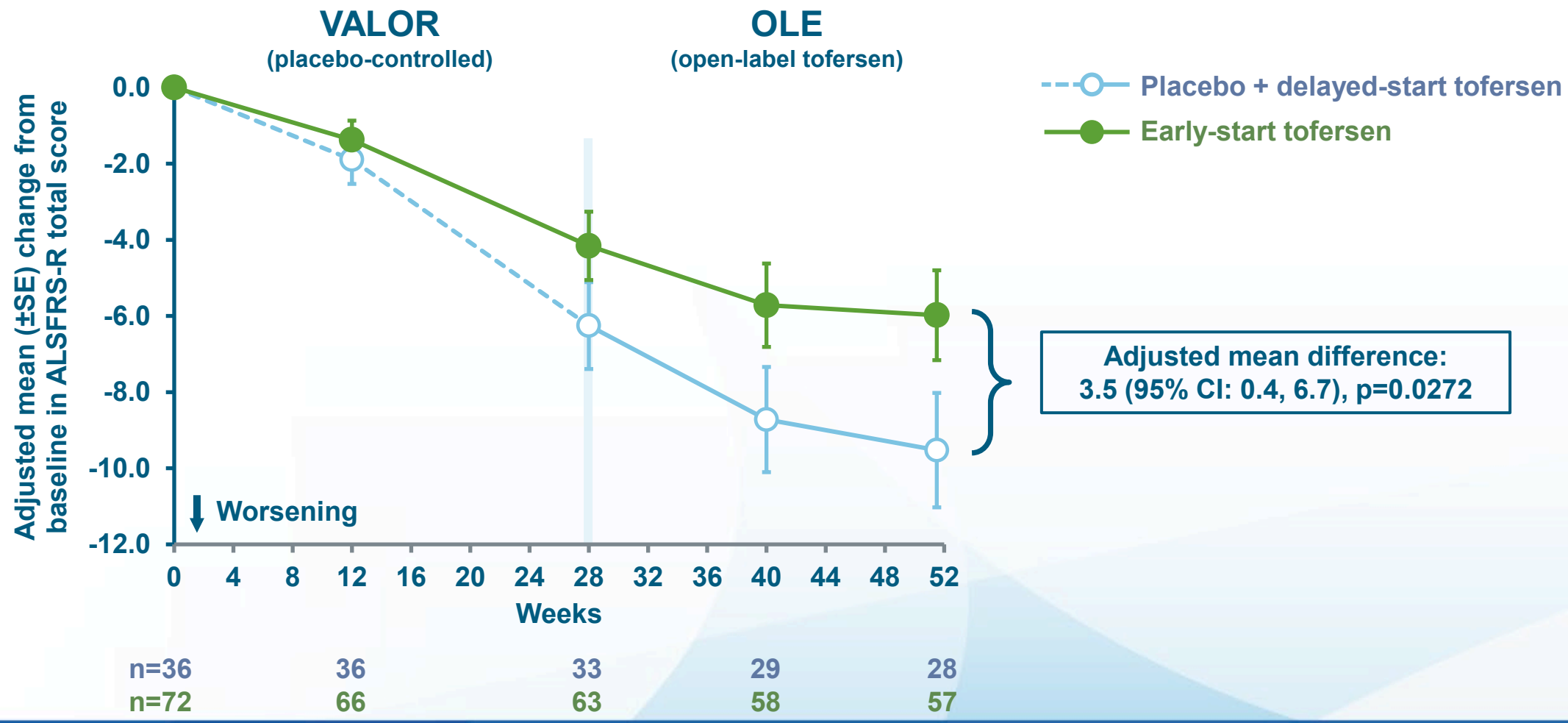
Modified analyses

- Used the full ITT population (N =108)
- Integrated VALOR and open label extension (OLE) data
  - Following VALOR readout, patients, site investigators, and study operation team for the OLE study remain blinded to treatment assignment in VALOR
  - Since placebo patients switched to tofersen in OLE, the VALOR + OLE analyses provides a conservative estimate of treatment effect
- Incorporated baseline plasma NfL as a covariate
- Investigated an alternative method for implementing the joint rank test

These exploratory analyses should be interpreted with caution given the limitations of data collected outside of a controlled study, which may be subject to confounding.

Effect on clinical function (ALSFRS-R)

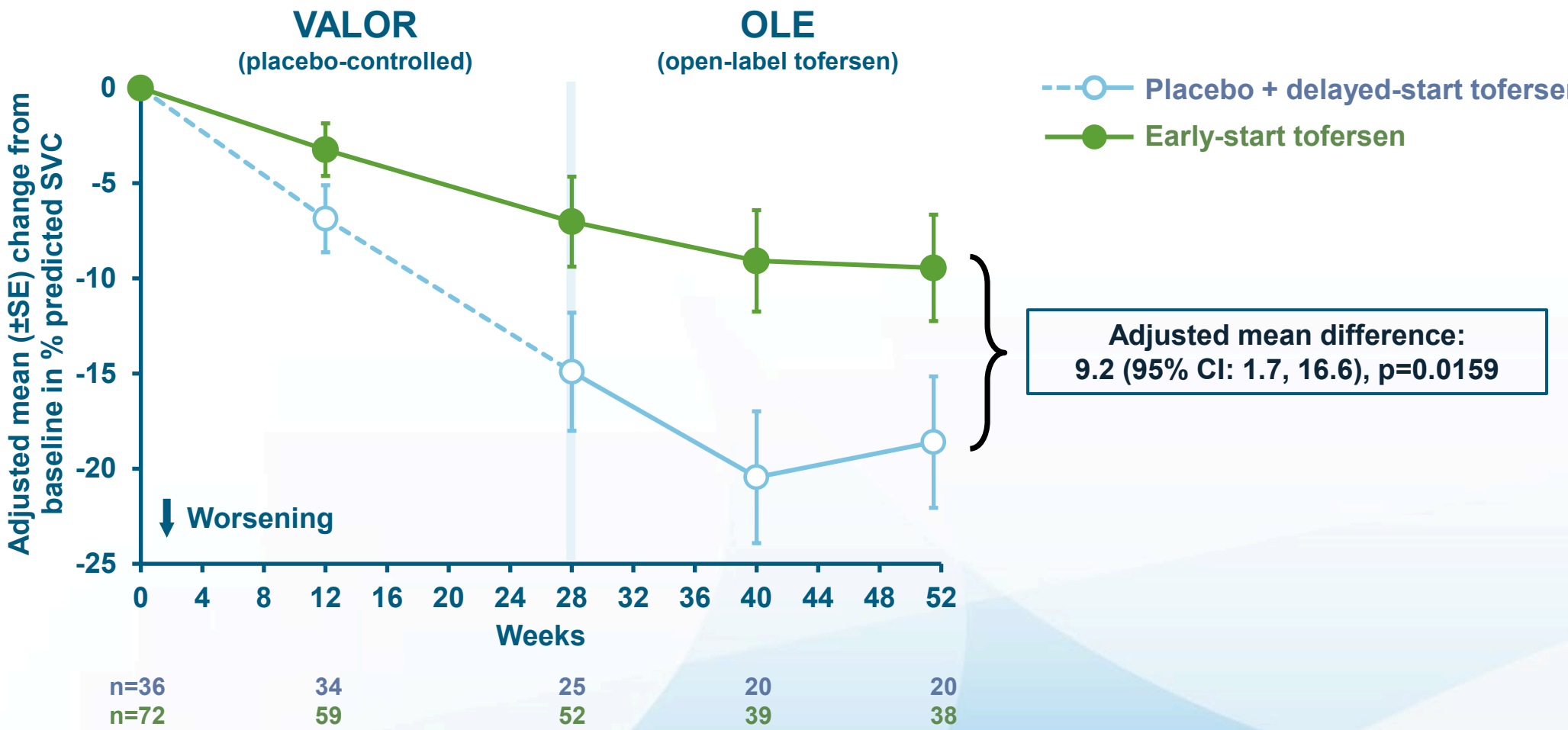
VALOR + OLE; ITT population (N=108)



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.

Effect on respiratory strength (SVC)

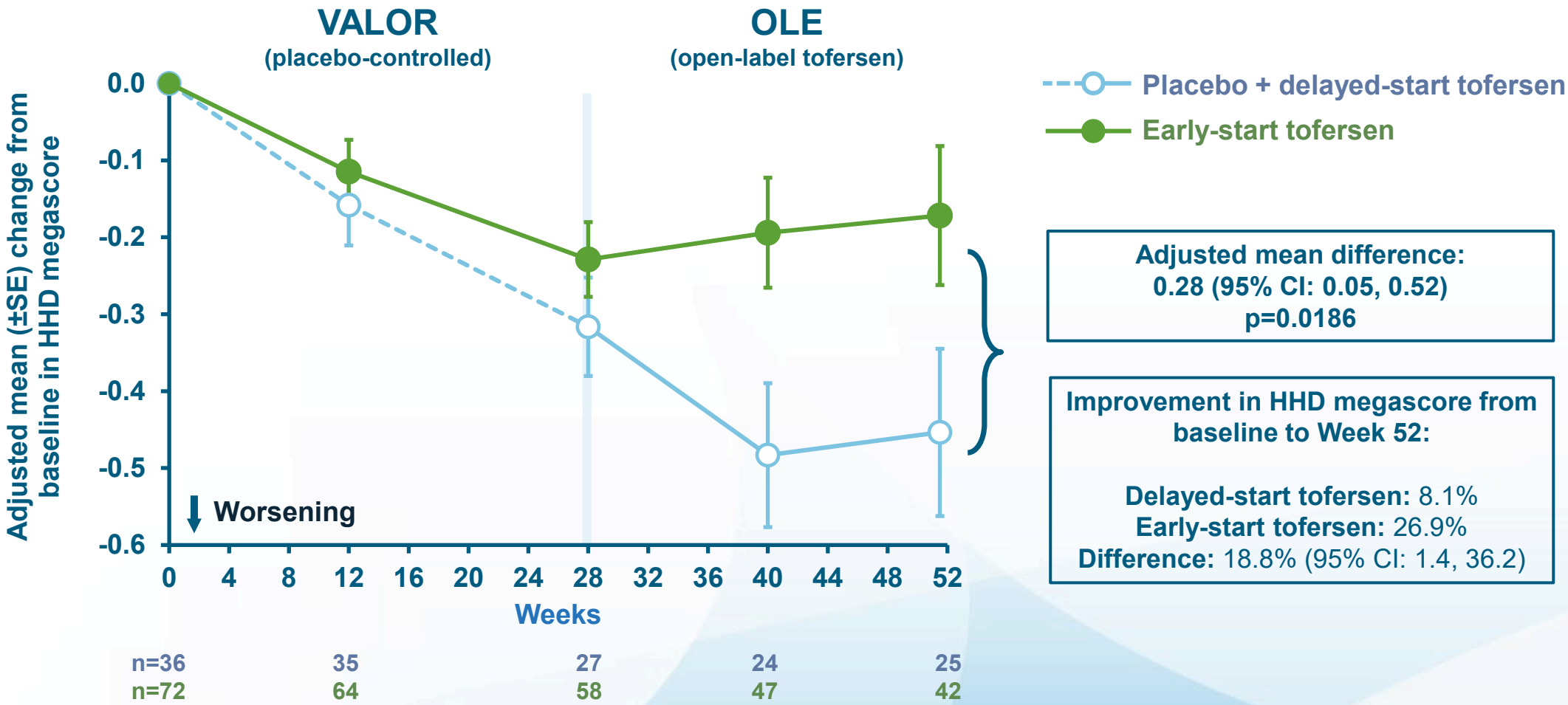
VALOR + OLE; ITT population (N=108)



OLE, open-label extension; SVC, slow vital capacity.
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.

Effect on muscle strength (HHD megascore)

VALOR + OLE; ITT population (N=108)



Adjusted mean difference:
0.28 (95% CI: 0.05, 0.52)
p=0.0186

Improvement in HHD megascore from baseline to Week 52:

Delayed-start tofersen: 8.1%
Early-start tofersen: 26.9%
Difference: 18.8% (95% CI: 1.4, 36.2)

HHD, handheld dynamometry; 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.

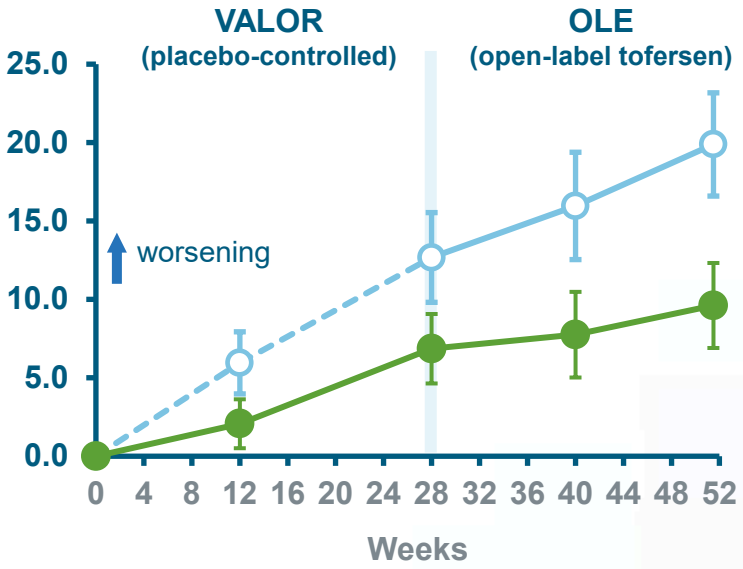
Effect on patient-reported outcome measures

VALOR + OLE; ITT population (N=108)

Placebo + delayed-start tofersen

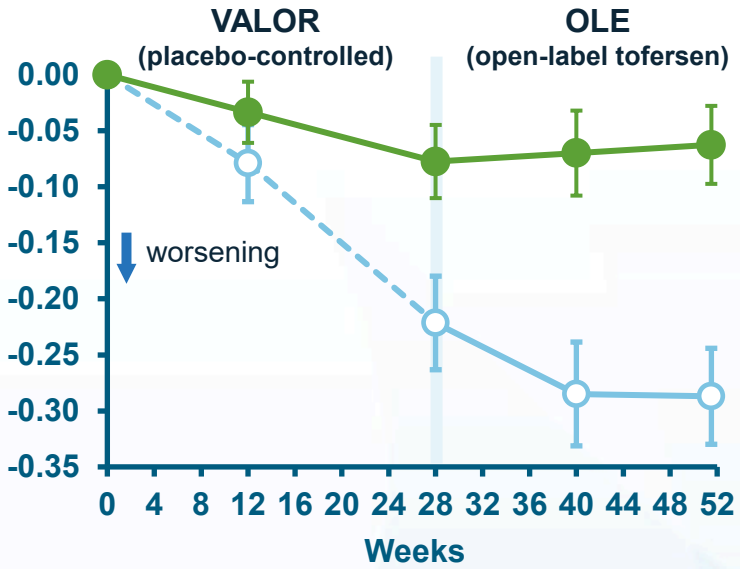
Early-start tofersen

Change in ALSAQ-5



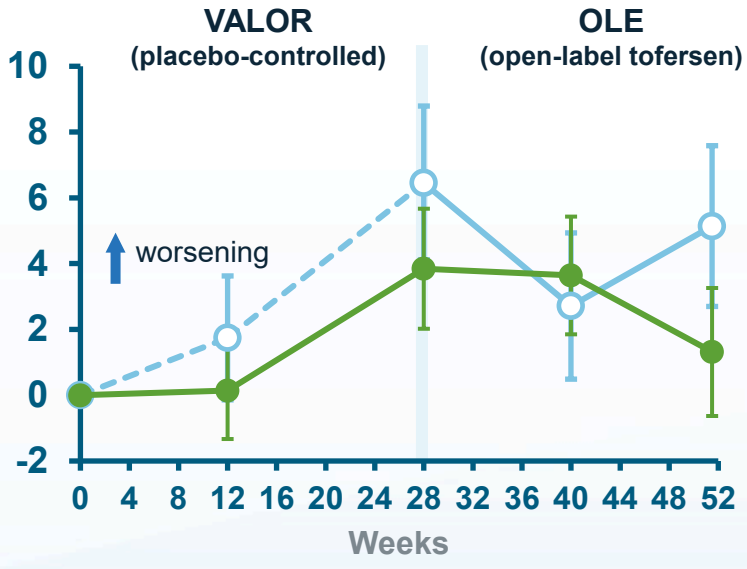
Difference: –10.3 (95% CI: –17.3, –3.2)
p=0.0044

Change in EQ-5D-5L Utility Score



Difference: 0.2 (95% CI: 0.13, 0.32)
p<0.0001

Change in Fatigue Severity Scale



Difference: –3.8 (95% CI: –9.0, 1.38)
p=0.1493

ALSAQ-5, 5 Item ALS Assessment Questionnaire; EQ-5D, EuroQOL-5 Dimension 5-Level Questionnaire; FSS, Fatigue Severity Scale; OLE, open-label extension; PRO, patient-reported outcome; 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.

[a]Using UK valuation weights

Source: Integrated efficacy analysis excel files - tab labelled ALSAQ-5, EQ-5D-5L, FSS

Tofersen Safety Profile

Warnings and Precautions with the use of tofersen

Myelitis and/or Radiculitis
Serious adverse reactions of myelitis and radiculitis have been reported in patients treated with tofersen

Papilledema and Elevated Intracranial Pressure
Serious adverse reactions of papilledema and elevated intracranial pressure have been reported in patients treated with tofersen

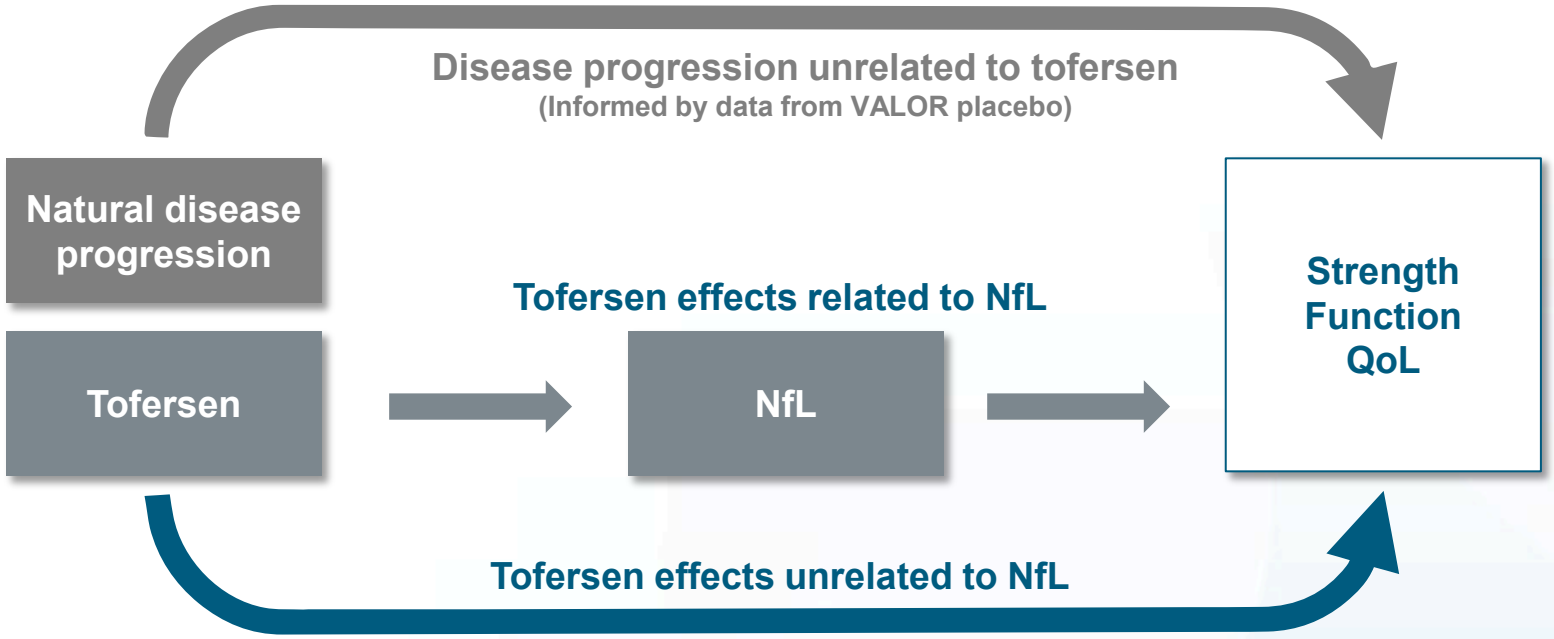
Aseptic Meningitis
Serious adverse reactions of aseptic meningitis (also called chemical meningitis or drug-induced aseptic meningitis) have been reported in patients treated with tofersen

In addition, nonserious adverse drug reactions of CSF white blood cell increased, and CSF protein increased have also been reported with tofersen

Adverse Reactions

The most common adverse reactions (≥ 10% of patients treated with tofersen and greater than placebo) were pain, fatigue, arthralgia, cerebrospinal fluid white blood cell increased, and myalgia.

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.

Equations for NfL Modeling

Observed data from Valor study

z_{0it_0}: Baseline NfL for Placebo

$z_{1it_0}, z_{1it_0,\text{std}}$: Baseline NfL for Tofersen (original scale and standardized)

z_{0it_1}: NfL level at week 16 for Placebo

z_{1it_1}: NfL level at week 16 for Tofersen

Δy_{0it_2}: Change from baseline in clinical outcome at week 28 for placebo

Δy_{1it_2}: Change from baseline in clinical outcome at week 28 for tofersen

$v_{0ij}, v_{1ij}, w_{0ij}, w_{1ij}, u_{1ij}$: Standardized covariates in the model

Parameter of Interest

λ_0, λ_j : Parameters for tofersen effect through for non-NfL pathway

γ_1 and γ_2 : Parameters for estimating for tofersen effect through NfL pathway

$$\log(z_{0it_1}) = \alpha_{0,z} + \beta_{0,z}\log(z_{0it_0}) + \sum_{j=1}^{m_1}\beta_{0,z}^{(j)}v_{0ij} + \epsilon_{0it_1,z} \quad (2.1)$$

$$\Delta y_{0it_2}|\Delta z_{0it_1} = \alpha_{0,y} + \beta_{0,y}z_{0it_0} + \gamma_{0,y}\Delta z_{0it_1} + \sum_{j=1}^{m_2}\beta_{0,y}^{(j)}w_{0ij} + \epsilon_{0it_2,y} \quad (2.2)$$

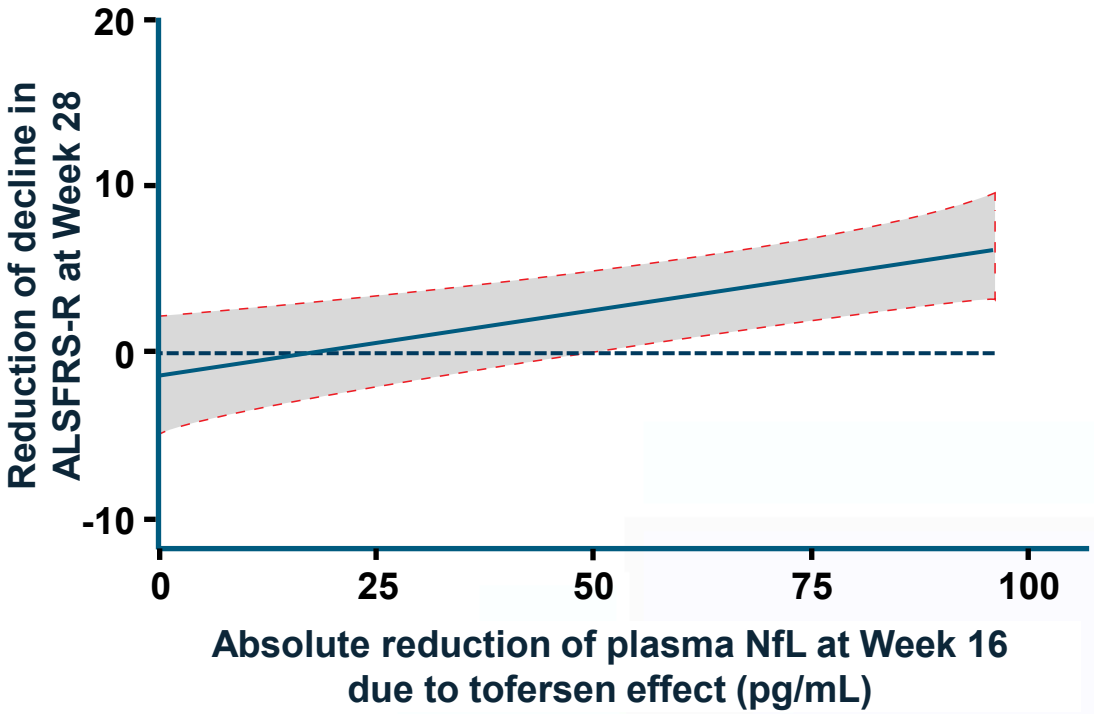
$$\begin{aligned}\Delta y_{1it_2}|\Delta z_{1it_1} = {} & \alpha_{0,y} + \beta_{0,y}z_{1it_0} + \gamma_{0,y}\left[e^{\alpha_{0,z}+\beta_{0,z}\log(z_{1it_0})+\sum_{j=1}^{m_1}\beta_{0,z}^{(j)}v_{1ij}} - z_{1it_0}\right] + \sum_{j=1}^{m_2}\beta_{0,y}^{(j)}w_{1ij} + \\ & \lambda_0 + \sum_{j=1}^{m_3}\lambda_j u_{1ij} + \\ & (\gamma_1 + \gamma_2 z_{1it_0,\text{std}})\left\{\Delta z_{1it_1} - \left[e^{\alpha_{0,z}+\beta_{0,z}\log(z_{1it_0})+\sum_{j=1}^{m_1}\beta_{0,z}^{(j)}v_{1ij}} - z_{1it_0}\right]\right\} + \\ & \epsilon_{1it_2,y}\end{aligned} \quad (2.3)$$

Expected clinical outcome worsening for tofersen patients under natural disease progression ($\Delta\tilde{y}_{1it_2}|\, z_{1it_0}, v_{1ij}, w_{1ij}$)

Tofersen effect due to unrelated to NfL ($\Delta y_{1it_2} - \Delta\hat{y}_{1it_2} - \Delta\tilde{y}_{0it_2}|u_{1ij}$)

Tofersen effect due to NfL reduction $\Delta\hat{y}_{1it_2}|z_{1it_0,\text{std}}, \Delta z_{1it_1} - \Delta\tilde{z}_{0it_1}, v_{1ij}$

Model Fitting Results [1]



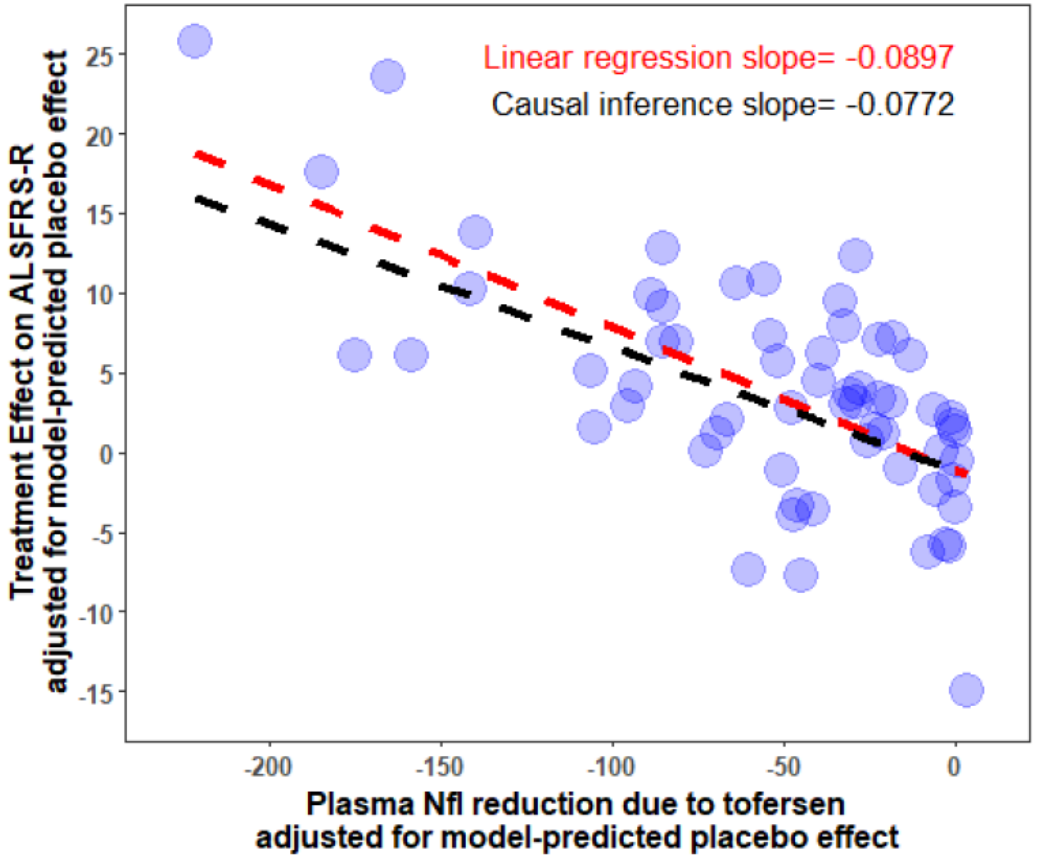
Predicted benefit on clinical outcomes (at Week 28) for each 10 pg/mL reduction of plasma NfL (at Week 16)*	
ALSFRS-R total score	0.77 (p=0.0038)
Percent-predicted SVC	1.45 (p=0.0706)
HHD overall megascore	0.029 (p=0.1303)
ALSAQ-5 total score	2.194 (p=0.0056)
EQ-5D-5L utility score	0.017 (p=0.0894)

*Example for a participant with a baseline plasma NfL level equivalent to the sample mean for ITT completers (96.78 pg/mL)

ALS, amyotrophic lateral sclerosis; ALSAQ-5, ALS Assessment Questionnaire; ALSFRS-R, ALS Functional Rating Scale-Revised; EQ-5D-5L, EuroQol 5 Dimension, 5 Level Questionnaire; HHD, handheld dynamometry; ITT, intent-to-treat; NfL, neurofilament light; QoL, quality of life; SVC, slow vital capacity.

Model Fitting Results [2][1]

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

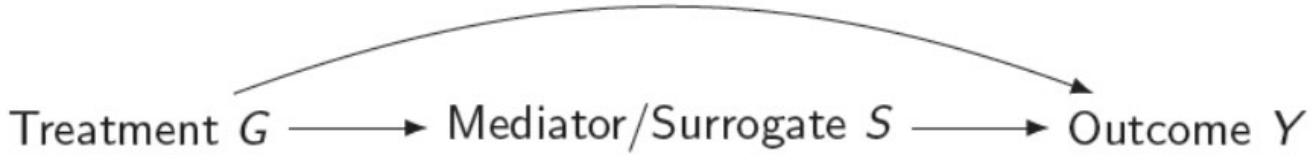
1. FDA briefing book for the tofersen advisory committee meeting on March 22, 2023.

The causal inference model relates to several existing statistical literatures

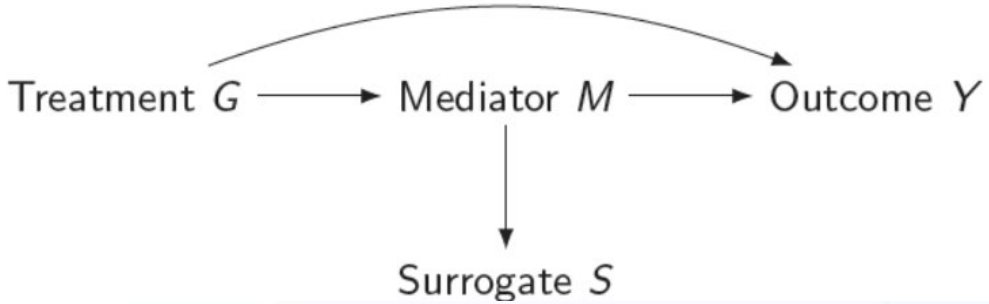
- Causal effect predictive curve proposed by Gilbert and Hudgens (2008):
  - Causal effect predictive curve can be constructed from the model that utilizes specific parametric form to estimate the underlying causal effects
- Heterogeneity of the utility of a surrogacy endpoint:
  - In parametric form, the model allows differentiating tofersen effect through the NfL pathway by baseline NfL level
- Precision medicine which focuses on identification of treatment by covariate interactions:
  - Covariates are robustly built into various component of the model: natural progression of NfL, natural progression of clinical function, and tofersen effect unrelated to NfL
- Mediation analysis: Surrogacy and mediation analysis are naturally linked together

Mediation Justification [1]

- 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), NIE (natural indirect effect), and ATE (average treatment effect) derived

Mediation Justification [2]

- NDE (natural direct effect): Tofersen effect through the non-NfL pathway

$$\frac{\sum_{i=1}^{n_1}\left(\lambda_0 + \sum_{j=1}^{m_3}\lambda_j u_{1ij}\right)}{n_1}$$

- NIE (natural indirect effect): Tofersen effect through the NfL pathway

$$\frac{\sum_{i=1}^{n_1}(\gamma_1 + \gamma_2 z_{1it_0,std})\left\{\Delta z_{1it_1} - \left[e^{\alpha_{0,z}+\beta_{0,z}\log(z_{1it_0})+\sum_{j=1}^{m_1}\beta_{0,z}^{(j)}v_{1ij}} - z_{1it_0}\right]\right\}}{n_1}$$

- ATE (average treatment effect): Sum of NDE and NIE

Reflection from Tofersen Example

- The modeling work relates to several existing statistical literatures
- The model can be extended to non-continuous endpoints (e.g., survival, binary, and ordinal) and it can also be extended to allow non-linear functional forms in its various components

Concluding remarks

- Surrogate endpoints have constantly been a challenging and controversial area. Early engagement with FDA and other regulatory agencies is recommended.
- Different statistical approaches have been proposed to validate surrogate endpoints while each has its limitations. Consistent results from multiple approaches and from multiple studies provide compelling substantiating evidence.
- Different disease areas have distinct features. Careful considerations are required to identify the appropriate statistical approaches in each case.
- Statisticians played a leadership role in establishing surrogacy in the Alzheimer's disease and SOD1-ALS examples. Collaborations with biomarker scientists, physicians and regulatory team are important.