


## RESEARCH ARTICLE

# Combined Assessment of Function and Survival to Demonstrate the Effect of Treatment on Progressive Supranuclear Palsy

Massimiliano Germani, PhD,<sup>1</sup> Irene Rebollo Mesa, PhD,<sup>1\*</sup>  Tim J. Buchanan, PhD,<sup>1</sup> Steven De Bruyn, MD,<sup>1</sup> Teresa Gasalla, MD,<sup>2</sup> Hans Lieve G. Van Tricht, BSN,<sup>1</sup> Colin Ewen, PhD,<sup>3</sup> Lawrence I. Golbe, MD,<sup>4</sup> Adam Boxer, MD, PhD,<sup>5</sup> and Günter Högl, MD<sup>6,7,8</sup>

<sup>1</sup>UCB, Braine-l'Alleud, Belgium

<sup>2</sup>UCB, Morrisville, North Carolina, USA

<sup>3</sup>UCB, Slough, United Kingdom

<sup>4</sup>Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA

<sup>5</sup>Department of Neurology, University of California, San Francisco, USA

<sup>6</sup>Department of Neurology, LMU University Hospital, Ludwig-Maximilians-Universität (LMU), Munich, Germany

<sup>7</sup>Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

<sup>8</sup>German Center for Neurodegenerative Diseases (DNZE), Munich, Germany

**ABSTRACT: Background:** Progressive supranuclear palsy (PSP) is a rare and fatal neurodegenerative disorder for which there are currently no disease-modifying treatments. Recent trials of potential therapies had durations of 12 months, which may be insufficient because of non-random missingness due to death. Longer durations, incorporating PSP Rating Scale and survival, can reduce the potential for type II error. Selecting efficacy measures more sensitive to disease modification may facilitate identification of treatment effect.

**Objective:** The objective of this study was to evaluate the simulated phase 3 PSP trial assessing the effect of disease-modifying intervention on a novel combined

primary endpoint comprising function (PSP Rating Scale) and survival, the Combined Assessment of Function and Survival (CAFS), and to determine operating characteristics of the CAFS.

**Methods:** To simulate PSP progression in the trial population, we developed models of PSP Rating Scale and survival using data from published clinical studies. These models were used to define operating characteristics of the CAFS for use in a phase 3 trial.

**Results:** The sample size determined (N = 384; 1:1 randomization) would provide >80% power to detect significant treatment effects on the CAFS compared with placebo. The CAFS provides good operating

Current address for Dr. Massimiliano Germani: Chiesi España, Barcelona, Spain

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

\*Correspondence to: Dr. Irene Rebollo Mesa, UCB, Chemin du Foriest B, 1420 Braine-l'Alleud, Belgium; E-mail: [irene.rebollomesa@ucb.com](mailto:irene.rebollomesa@ucb.com)

**Relevant conflicts of interest/financial disclosures:** M.G., I.R.M., T.J.B., S.D.B., T.G., H.L.G.V.T., and C.E. were employees of UCB during this study and may hold/have access to stock options. L.I.G. has consultancies with AI Therapeutics, Amylyx, Apellis, Aprinolia, Ferrer, IQVIA, Mitochon, Mitsubishi Tanabe, P3Lab, Roche, Switch, UCB, and Woolsey; participated in advisory boards with Amylyx, Roche, Rossy Centre (University of Toronto), and Springer; board of directors: CurePSP (travel expenses only); inventions: PSP Rating Scale and A Video Guide to the PSP Rating Scale (IP of Rutgers University). A.B. has consultancies with AGTC, Alector, Alzprotect, Amylyx, Arkuda, Arrowhead, Arvinas, Aviado, Eli Lilly, GSK, Humana, Merck, Modalis, Muna, Oligomerix, Oscotec, Pfizer, Roche, Switch, Transposon, and UnlearnAI; institution's research support with Biogen, Eisai, and Regeneron; research support with National Institute on Aging: NIH U19AG063911, R01AG078457, R01AG073482, R56AG075744, R01AG038791, RF1AG077557,

P01AG019724, R01AG071756, and U24AG057437; Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Association for Frontotemporal Degeneration, Bluefield Project to Cure FTD, Gates Ventures, GHR Foundation, Rainwater Charitable Foundation, UCSF Parkinson's Spectrum Disorders Center, and the University of California Cures AD Program. G.H. has employment with LMU University Hospital; ongoing research collaborations: AbbVie, Alzprotect, Amylyx, Aprinolia, Asceneuron, Bayer, Bial, Biogen, Biohaven, Epidarex, Ferrer, Kyowa Kirin, Lundbeck, Novartis, Retrotope, Roche, Sanofi, Servier, Takeda, Teva, and UCB; honoraria: AbbVie, Bayer, Bial, Biogen, Bristol Myers Squibb, Kyowa Kirin, Pfizer, Roche, Teva, UCB, and Zambon.

**Funding agencies:** This research was funded by UCB. G.H. was supported by the European Joint Programme on Rare Diseases (Improve-PSP) and the Deutsche Forschungsgemeinschaft DFG, (German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy ID 390857198).

Full financial disclosures and author roles may be found in the online version of this article.

**Received:** 5 April 2024; **Revised:** 9 August 2024; **Accepted:** 12 September 2024

Published online 29 October 2024 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)). DOI: 10.1002/mds.30027

characteristics and increased power to detect moderate treatment effects on the PSP Rating Scale. We propose a trial design allowing potential detection of treatment effects at a preplanned interim analysis after participants complete 12 months of treatment, with assessment of effects of treatment ( $\leq 24$  months) on survival.

**Conclusions:** Use of the CAFS could provide a comprehensive and robust estimate of the clinical benefit of

future therapies. © 2024 UCB. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

**Key Words:** joint rank test; clinical rating scale; item response theory; progressive supranuclear palsy; neurodegenerative disease

Progressive supranuclear palsy (PSP) is a rare, progressive neurodegenerative disease.<sup>1,2</sup> The different clinical phenotypes of PSP share a common pathophysiology and may converge to PSP-Richardson's syndrome over time.<sup>3,4</sup> PSP leads to death after an average of 6 to 7 years.<sup>5,6</sup> Neuropathologically, PSP is an archetypical primary four-repeat tauopathy that affects neurons, oligodendrocytes, and astrocytes.<sup>7,8</sup>

There are currently no approved disease-modifying treatments that halt or delay PSP progression.<sup>9</sup> Recent studies have elucidated the pathogenesis of tauopathies, facilitating the development of disease-modifying interventions.<sup>10</sup> However, earlier clinical trials (CTs) that evaluated the efficacy and disease modification potential of investigational therapies for PSP were unsuccessful.<sup>11,12</sup>

To date, most PSP studies have used the PSP Rating Scale (PSPRS) as a primary outcome measure and defined change in the PSPRS within a 12-month observation period as the primary endpoint.<sup>11-13</sup> The PSPRS is a prospectively validated physician-rated measure of disease severity in PSP, measured by 28 items within six categories (daily activities, mentation, bulbar, ocular motor, limb motor, gait/midline examination).<sup>14</sup> The rater assigns a score of 0 to 2 or 0 to 4 points for each item, yielding a total PSPRS score from 0 to 100. Higher PSPRS scores indicate greater disease severity, with more functional impairment.<sup>14</sup> Previous studies that used the PSPRS indicate annual increases in total PSPRS score in patients with PSP between 10 and 12.<sup>14-19</sup> A reduction in progression of PSP scores by ~40% is considered clearly clinically meaningful<sup>20</sup> and would likely increase survival time.

Results from past CTs suggest the investigational therapies evaluated were not efficacious. Optimization of trial protocols is still warranted to increase sensitivity of trials to changes in functional decline, obtain additional information on potential effects on mortality, and detect delayed effects of treatment (by increasing the duration of trials). Previously, data missing after patient death were treated as missing at random.<sup>11</sup> This could bias estimates of treatment effect, reduce power to detect efficacy, and may cause potential reductions in mortality to be overlooked. This could lead to a lack of definitive data to support the approval of novel therapies.

Based on the Joint Rank Test (JRT),<sup>21</sup> the Combined Assessment of Function and Survival (CAFS) was developed for analysis of function adjusted for mortality in patients with amyotrophic lateral sclerosis.<sup>22</sup> Data from recent PSP trials, as well as recent FDA guidance for amyotrophic lateral sclerosis,<sup>23</sup> prompted us to apply this methodology to PSP. We describe simulation work used to define operating characteristics of the CAFS and enable use of the CAFS in PSP trials.

## Subjects and Methods

Two reference datasets were used to obtain the relevant parameters to inform trial simulations. Data from a phase 2/3 CT of davunetide<sup>24</sup> were used to determine parameters for assessment of disease progression measured by PSPRS, and parameters for survival were assessed using data from the research conducted to develop and validate the PSPRS.<sup>14</sup> Data from trials of riluzole<sup>25</sup> and tideglusib<sup>16</sup> were also used to calculate sample sizes required for assessment of impact on overall survival.

### Phase 3 Clinical Trial Design

The proposed design of a hypothetical phase 3 PSP trial was a double-blind, placebo-controlled study with up to 24 months treatment duration, assessing efficacy, safety, and tolerability of a disease-modifying therapy. The final proposed design included a preplanned interim efficacy analysis after the last participant reached 12 months post-randomization, to allow early stopping for a positive result. The proposed population comprised participants  $\geq 40$  years of age with probable/possible PSP-Richardson's syndrome (meeting Movement Disorder Society-PSP diagnostic criteria).<sup>2</sup> The primary endpoint was the CAFS, comprising change from baseline to month 24 in the signs and symptoms of PSP (assessed using PSPRS) and time to death to month 24.

### The Item Response Theory Model to Simulate Longitudinal PSPRS Data

An Item Response Theory (IRT) model was developed based on the complete PSPRS dataset from a trial of davunetide in PSP to simulate disease progression in

the trial population.<sup>24</sup> The IRT model is a psychometric approach based on the relationship between the disease severity of each individual, measured by a test item, and the overall measure of disease progression that the item was designed to measure.<sup>26</sup> The model was used to simulate a CT in a two-step approach and was run using the final parameters estimated based on the overall davunetide dataset (313 participants with PSPRS who received davunetide [ $n = 157$ ] or placebo [ $n = 156$ ]), including 28 items for each participant.<sup>24</sup> For technical details of the IRT model, see the Supporting Information Methods. The IRT model was used to stimulate the PSPRS of 5000 placebo-treated participants (disease duration  $\leq 7$  years) and 5000 active-treated participants for each of three clinical scenarios of treatment effect, depending on the extent of expected efficacy:

- Absence of efficacy or null: annual change from baseline in PSPRS total score as observed in the davunetide trial<sup>24</sup> ( $\sim 11.1$  points).
- Clinically significant effect: treated participants annual change from baseline reduced by 40% compared with the null.
- Moderate effect: treated participants annual change from baseline reduced by 20% compared with the null.

The following visit schedule was assumed: baseline, six visits during year 1, and four visits during year 2.

The IRT model developed was qualified for use in simulations using the Visual Predictive Check approach. The Visual Predictive Check approach compared the predicted versus the observed mean score for each item, the predicted versus observed probabilities for each score item, and the predicted versus total score.

In neurodegenerative disease, moderate treatment effects (ie, 20% to 35% reduction in disease progression) often have no impact on cognitive impairment.<sup>27,28</sup> The IRT model allowed estimation of disease progression without assuming the relationship between time and increase in the total PSPRS score. The only parametric assumption in the IRT model links time with progression of disease (disease severity), which defines the latent variable.

Scores generated by the IRT model, representing the virtual dataset, were analyzed using a linear mixed-effects model to compute the mean percentage change in total PSPRS score between treatment groups. This two-step simulation process ensures the use of different models for virtual data creation and analysis.

### Simulation of Overall Survival

The largest CT to evaluate survival for up to 3 years in people with PSP-Richardson's syndrome was the

riluzole phase 3 trial,<sup>25</sup> which estimated the probability of survival for 3 years from enrollment to be approximately 50%. A 2-year mortality rate of approximately 30% is expected with placebo or ineffective treatment. A strong association between higher PSPRS scores and higher probability of death has previously been demonstrated,<sup>14</sup> and a score of  $>70$  equated to  $>50\%$  probability of mortality within 1 year.

Data from research completed to develop the PSPRS<sup>14</sup> indicated that a higher baseline (median baseline scores  $>40$ ) and a greater increase in PSPRS score during the first year (exceeding the median 1 year slope of 9.55) were correlated with a reduction in median survival of  $\geq 1$  year. The correlation between PSPRS and survival was incorporated in the simulation design, with a higher simulated baseline, or faster progression correlated with a higher risk of death.

Survival data were simulated based on the hazard function determined in the placebo arm of the riluzole trial,<sup>25</sup> including the simulated individual PSPRS scores as covariates in the model. The *simSurv* package from R was used to perform the simulation, and hazard ratios (HR) for survival were used to compare survival between the treated and placebo groups (eg, HR of 2 would indicate risk of death in the treated group was twice that in the control group).<sup>29</sup> The baseline hazard was obtained using a flexible spline model with three knots, based on the time-to-death data in the riluzole study. At each simulation replication, the simulated PSPRS baseline and 1-year individual slope were included as proportional-hazard covariates in the survival simulation model.

Survival probabilities observed in the riluzole trial,<sup>25</sup> ie, 97%, 87%, 78%, and 70%, at months 6, 12, 18, and 24, respectively, were used to define the null condition (HR = 1). From there, three treatment conditions were simulated:

- HR = 1.1: Treatment increased mortality risk.
- HR = 0.8: Treatment moderately reduced mortality risk.
- HR = 0.5: Treatment greatly reduced mortality risk.

The indirect treatment effect was introduced via covariate adjustment with the individual baseline and annual slope of simulated PSPRS data.

### Calculation of the CAFS

Each participant's outcomes were compared with all other participants in the trial in  $N-1$  comparisons. A score of  $-1$  (worse),  $0$  (equal), or  $1$  (better) was assigned to each pairwise comparison (Fig. 1). The pairwise comparison was first based on survival status last known at the time of the analysis, essentially resulting in no missing data:

- If person A died and the comparative participant (person B) was alive,  $-1$  was assigned to the former and  $1$  to the latter.
- If both participants died, the comparison would be based on time to death: the longer survival time would be assigned a score of  $1$  and the shorter  $-1$ ; if both had the same survival time, a score of  $0$  would be assigned to both.
- If both participants were alive at their last visit, the changes in PSPRS score between baseline and the last common visit would be compared.
- The smaller increase would be assigned a score of  $1$  and the larger  $-1$ ; if both had the same change in PSPRS, a score of  $0$  would be assigned to both.

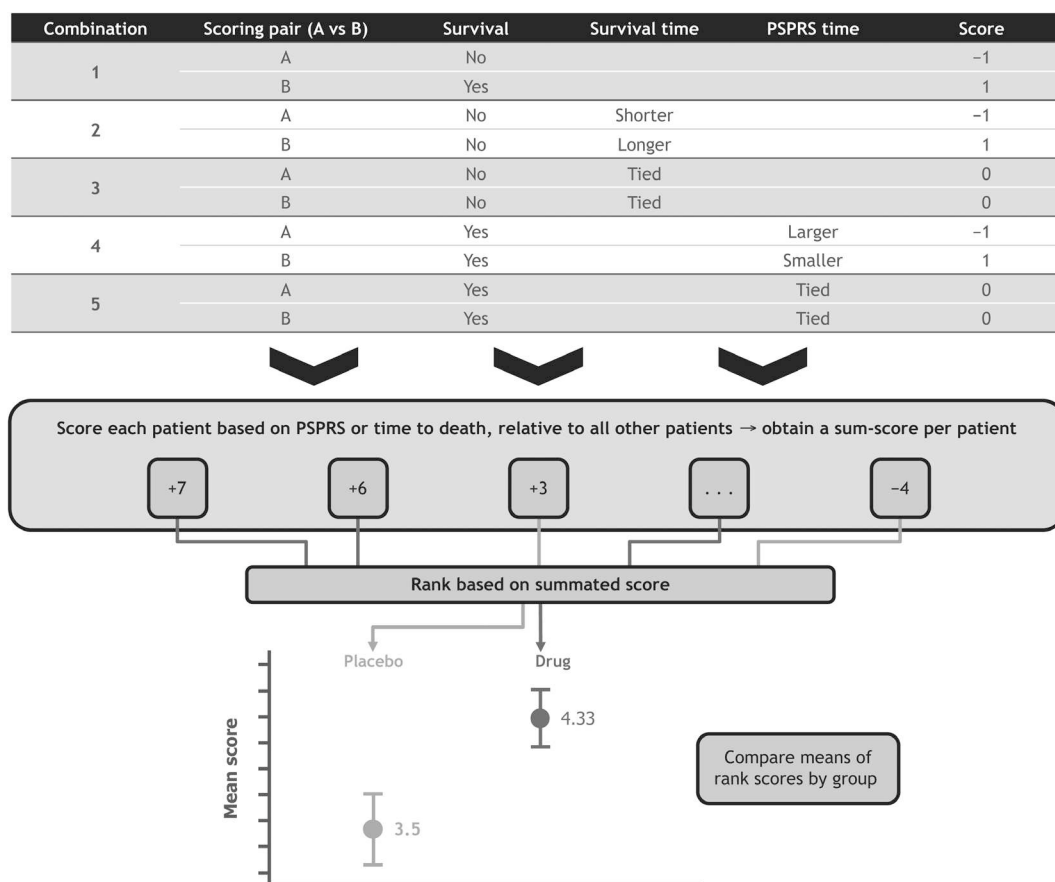
The  $N-1$  scores are summated, and the corresponding rank is their CAFS.

### Sample Size

The simulation was based on participants with PSP-Richardson's syndrome being randomized to investigational therapies or placebo (1:1). A total of 384 participants were required to provide  $>80\%$  power to detect significant treatment effects on either of the

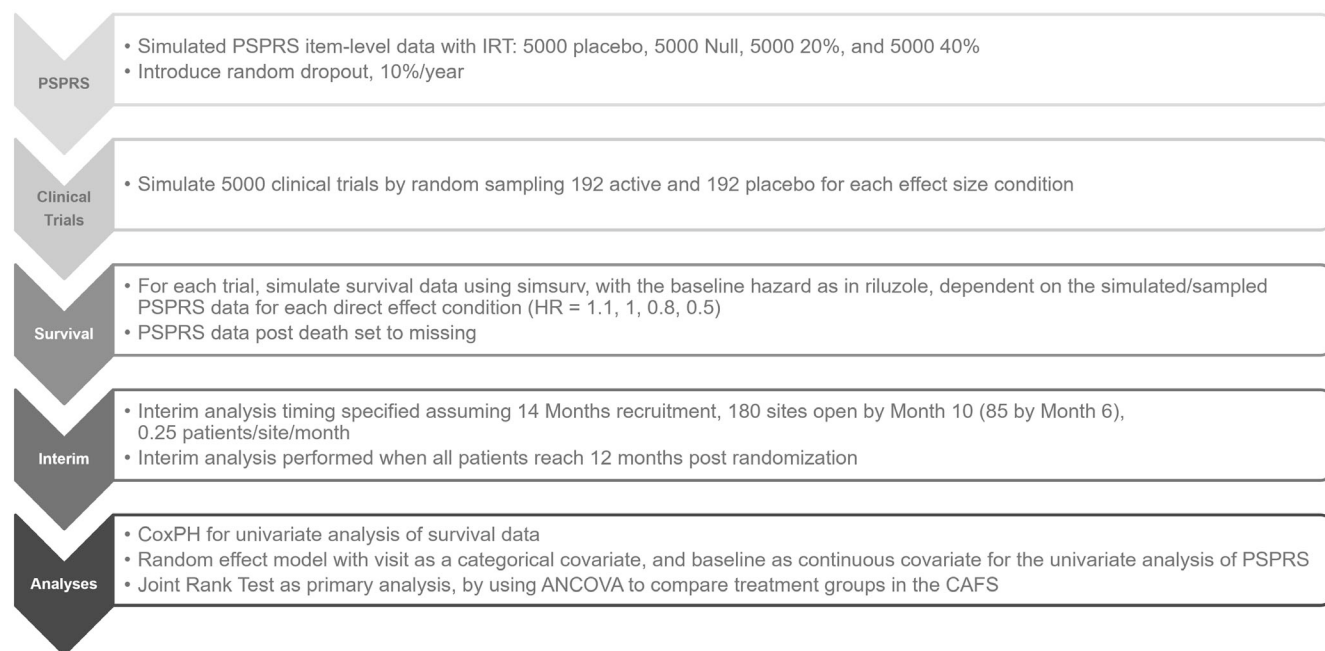
subcomponents of the CAFS. Based on the survival probabilities determined in the riluzole trial placebo arm, the simulation identified that 384 participants would provide  $>80\%$  power to detect a minimum  $50\%$  reduction in relative risk of death ( $HR = 0.5$ ) in the investigational therapies group compared with placebo, after 24 months of treatment. This sample size would account for an expected  $20\%$  random censoring by the end of the 2-year double-blind treatment period, as observed in the riluzole trial. Additionally, assuming a mean annual change from baseline in PSPRS score of  $11.24$  (standard deviation:  $9.95$ ),<sup>10</sup> as in the davunetide and tideglusib trials,<sup>16,24</sup> 384 participants would provide  $>90\%$  power to detect  $\geq 40\%$  reduction ( $4.6$ -point difference) in annual disease progression, with  $1\%$  type I error (assuming up to  $10\%$  annual random dropout and  $\sim 15\%$  annual mortality).

The difference between treatment groups in the CAFS was tested using analysis of covariance (ANCOVA), adjusted for baseline PSPRS score, with treatment group as a fixed effect. Statistical significance of the treatment effect was evaluated using the  $P$  value corresponding to the Wald-test statistic of the fixed effect of treatment.



**FIG. 1.** Schema of the Joint Rank Score. For all comparisons, scores of  $0$ ,  $1$ , or  $-1$  are assigned, the process is repeated for all participants, resulting in a summated score that is ranked. Participants are assigned by treatment group, then mean summated score is calculated per group. The mean ranked score per treatment group is compared using ANOVA. PSPRS, Progressive Supranuclear Palsy Rating Scale; ANOVA, analysis of variance.





**FIG. 2.** Simulation study sequence. Missing data are introduced via two mechanisms. First, a random annual dropout of 10% is introduced to the PSPRS data, assuming equal probability of dropout at each study month. Subsequently, once survival data are simulated, PSPRS data generated after (simulated) date of death are made missing. PSPRS, Progressive Supranuclear Palsy Rating Scale; IRT, Item Response Theory; HR, hazard ratio; CoxPH, Cox proportional hazards model; ANCOVA, analysis of covariance; CAFS, Combined Assessment of Function and Survival.

## Clinical Trial Simulations and Analysis

Effect sizes (for PSPRS and survival) were simulated; Figure 2 shows the sequence followed in the simulation study. For both variables, a null scenario (no difference between placebo and the investigational therapy) was generated, and two levels of effectiveness corresponding to minimum detectable differences in endpoints (survival HR = 0.5, and 40% reduction in PSPRS progression) and intermediate effect (survival HR = 0.8, and 20% reduction in PSPRS progression). The intermediate effect is meant to understand power of the JRT when combining moderate effects. A further effect size was generated to simulate a slightly deleterious effect of treatment on survival (HR = 1.1) and to evaluate the corresponding loss of power when using the CAFS in this scenario.

A total of 5000 participants per scenario provided stable results, and each step was run in a loop 5000 times. At each of the 5000 replications, a new CT sample was drawn from the IRT model-generated PSPRS “population data” comprising 192 placebo participants and 192 participants treated with investigational therapies. At each new replication of the 24-month trial, a new bootstrap sample was drawn for each effect size, with replacement. Based on the randomly drawn sample of PSPRS data, survival data were generated. An interim dataset was created after all participants completed the 12-month treatment period to evaluate performance of the preplanned interim analysis. Each virtual trial was analyzed considering both interim and

final simulated data, according to the stopping rules, and alpha spending strategy (Fig. 3).

To compare operating characteristics of the CAFS with the univariate analyses of its subcomponents, we used a linear mixed-effects model to analyze treatment effects on the PSPRS total score, and a Cox proportional hazards model was used to analyze treatment effects on overall survival.

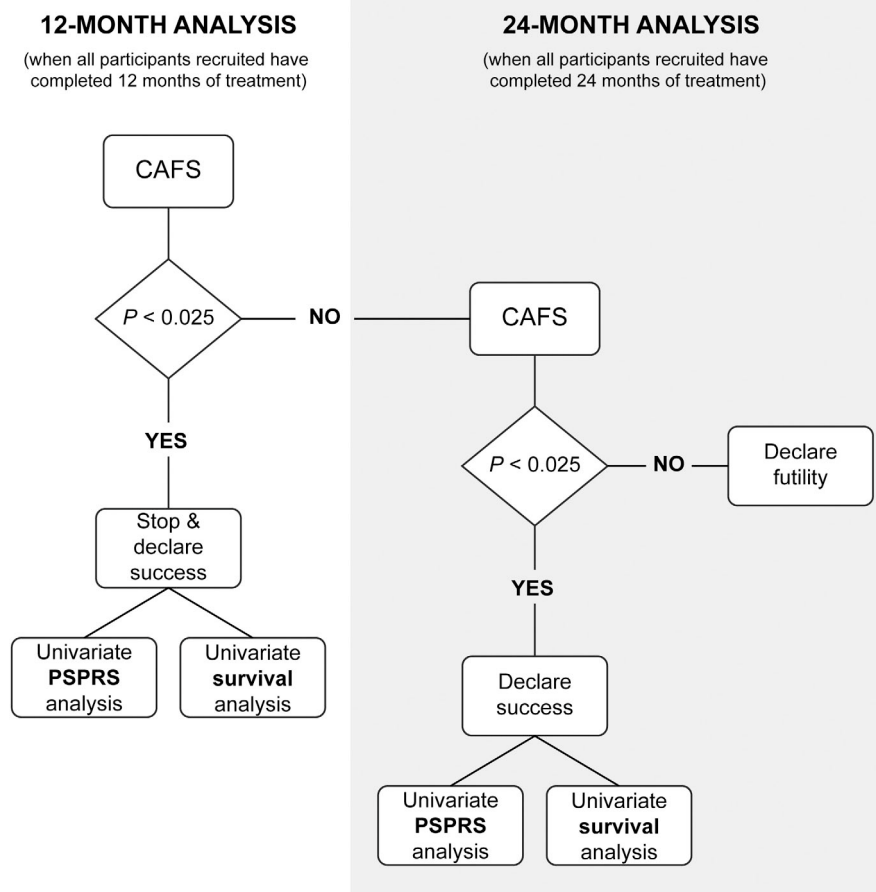
## Results

### The Item Response Theory Model of Longitudinal PSPRS Data

The davunetide CT did not show statistically significant differences between treatment arms for any clinical endpoints; therefore, IRT analysis was performed in the complete sample without compromising the validity of the results. The predicted versus observed PSPRS total scores comparing PSPRS disease progression with IRT modeling are presented in Supporting Information Figure S1. The IRT model and the derived final parameters were used to simulate the PSPRS total score for 384 virtual participants.

### Clinical Trial Simulation

The proposed trial design (Supporting Information Fig. S2) included an interim analysis after the last participant reached 12 months post-randomization, to



**FIG. 3.** Testing sequence and alpha spending. A Bonferroni split of alpha will be applied in the trial to maintain the familywise 2-sided type I error at 5% between the interim and final analysis. CAFS, Combined Assessment of Function and Survival; PSPRS, Progressive Supranuclear Palsy Rating Scale.

allow early stopping for efficacy. A futility analysis was not planned because all patients would be enrolled and treated for a long time. Continuation to the final analysis would be of interest to provide definitive evidence of efficacy and comprehensive safety data. Assuming 384 participants, a recruitment rate of 0.25 participant per site per year (Supporting Information Methods) and 180 sites open by month 10 (~90 sites by month 6), a total duration of ~38 months is expected (10% random annual dropout), meaning that interim analysis could be projected at month 26 (~12 months after 14 months of recruitment) and 1 year earlier than the completion date. According to this strategy, success could be declared at the interim analysis when the  $P$  value of the JRT (ANCOVA) was  $<0.025$ . If success was declared at the interim, the trial was not “analyzed” at month 24. Final analyses would be completed only for trials where the JRT  $P$  value for treatment effects was  $\geq 0.025$  at the interim.

Operating characteristics of the JRT in the final trial design across conditions indicate type I error was well controlled by the Bonferroni correction, keeping the familywise type I error rate  $<5\%$  (Table 1). The trial

design is characterized by high power at the interim analysis, given PSPRS effects of 40% are present irrespective of survival effects. Improved survival rates in the active arm increased power significantly ( $>90\%$ ) and were particularly valuable when the effect of treatment on the PSPRS was moderate (20%). Reduced survival rates in the active arm, even when mild, led to reduced power, reflecting the protective properties of the JRT, indicating comprehensive evaluation of treatment effect.

The current simulation design showed that a small number of trials would be expected to be negative at the interim analysis, but positive at the final analysis. This number, however, was much larger when the PSPRS effects were null, and effects on survival were large ( $HR = 0.5$ ). This pattern of results was partly due to the properties of the JRT, but largely due to the assumptions of the simulation design. The final analysis is expected to provide additional value when assumptions fail, eg, delayed treatment effects, slower rate of death, faster recruitment, and/or larger variability at interim analysis.

Table 2 shows operating characteristics of the univariate analyses of the PSPRS for total score, as a function

**TABLE 1** Operating characteristics of the Joint Rank Test (ANCOVA) analysis performed with the PSPRS total score

| HR <sup>a</sup> | PSPRS ES (%) | Success at Interim (%) | Success at Final (%) <sup>b</sup> | Power <sup>c</sup> |
|-----------------|--------------|------------------------|-----------------------------------|--------------------|
| 0.5             | Null         | 30.1                   | 23.24                             | 0.53               |
| 0.5             | 20           | 89.8                   | 6.56                              | 0.96               |
| 0.5             | 40           | 99.84                  | 0.16                              | 1.00               |
| 0.8             | Null         | 8.48                   | 3.36                              | 0.12               |
| 0.8             | 20           | 63.18                  | 6.56                              | 0.69               |
| 0.8             | 40           | 96.6                   | 1.22                              | 0.97               |
| 1.0             | Null         | 3.4                    | 1.02                              | 0.04 <sup>d</sup>  |
| 1.0             | 20           | 40.56                  | 2.78                              | 0.43               |
| 1.0             | 40           | 89.02                  | 1.24                              | 0.90               |
| 1.1             | Null         | 2.64                   | 1.62                              | 0.04               |
| 1.1             | 20           | 30.06                  | 1.26                              | 0.31               |
| 1.1             | 40           | 82.64                  | 1.08                              | 0.84               |

Success is determined by  $P < 0.025$ .

<sup>a</sup>True direct effect simulated on survival.

<sup>b</sup>Success at final analysis shows the percentage of trials with a negative result at interim analysis and positive result at final analysis.

<sup>c</sup>Power reflects the proportion of trials that are either successful at the interim or at final analysis, simulating a real trial; thus, trials successful at the interim are not reanalyzed at the final analysis.

<sup>d</sup>Two-sided type I error.

Abbreviations: ANCOVA, analysis of covariance; PSPRS, Progressive Supranuclear Palsy Rating Scale; HR, hazard ratio; ES, effect size.

of the outcome in the primary analysis of the JRT. For each clinical scenario defined by survival HR and effect size in PSPRS, probability of each potential outcome was defined, with higher probability resulting in the most likely scenario. Although the rank score cannot be interpreted by itself, descriptive statistics of the CAFS for the 10-item and total PSPRS scores across simulation conditions are available in Supporting Information Tables S2 and S3.

Overall, power of the JRT was largely driven by effects on the PSPRS; thus, a significant JRT was commonly accompanied by a significant effect on the univariate PSPRS analysis. When treatment effects on the PSPRS are moderate (ie, 20% reduction in the annual progression), the linear mixed-effects model is always underpowered (Table 2). In contrast, the JRT allows the detection of treatment effects that are moderate in the PSPRS, if accompanied by moderate (HR = 0.8) or strong (HR = 0.5) treatment effects on survival, with powers of 0.69 and 0.96, respectively (Table 1). Greater discrepancies between results of the JRT and of the PSPRS change alone appeared when significant JRT was driven by treatment effects on survival, accompanied by moderate treatment effect (20%) on the PSPRS. In the absence of treatment effect on the PSPRS, when direct effects on survival were strong (HR = 0.5), many trials (30% to 40%) result in a nonsignificant JRT (Table 3). However, the Cox proportional hazards

analysis of survival would be statistically significant at the end of month 24 in >80% of trials. Although this simulated scenario is artificial and unlikely, these findings highlight that the operating characteristics of the JRT are largely driven by effects on function (PSPRS), and survival is taken into consideration but is not driving the trial outcome.

## Discussion

Clinical scores are key to assessing disease progression in neurodegenerative conditions, because they represent outcomes that cannot be assessed directly. Based on the simulated results, a phase 3 trial design has been recommended, with 384 participants with PSP-Richardson's syndrome, randomized 1:1, over a treatment duration of 12 to 24 months. A trial design was proposed based on the CAFS and included an interim analysis after all participants completed 12 months of treatment. The sample size determined would provide >80% power to detect significant treatment effects on either of the two subcomponents of the CAFS (change from baseline in PSPRS and time to death).

In this new phase 3 trial design concept, two major design changes from previous studies in PSP<sup>11-13</sup> were suggested to optimize decision making. First, the option to increase the treatment period from 12 to 24 months. This will detect potentially delayed treatment effects

**TABLE 2** Operating characteristic of PSPRS total score analysis

| HR <sup>a</sup> | PSPRS ES (%) | Result of JRT                  | Interim %-sig PSPRS 12 M Effect | Interim %-sig PSPRS 24 M Effect | Final %-sig PSPRS 12 M Effect | Final %-sig PSPRS 24 M Effect | Probability of JRT Result (%) |
|-----------------|--------------|--------------------------------|---------------------------------|---------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 0.5             | 20           | NS <sup>b</sup>                | 0.3                             | 0.5                             | 0.3                           | 0.6                           | 3.6                           |
| 0.5             | 20           | Significant interim            | 58.0                            | 19.3                            | 59.8                          | 67.4                          | 89.8                          |
| 0.5             | 20           | Significant final <sup>c</sup> | 0.8                             | 0.4                             | 0.9                           | 2.0                           | 6.6                           |
| 0.5             | 40           | NS <sup>b</sup>                |                                 |                                 |                               |                               | 0                             |
| 0.5             | 40           | Significant interim            | 97.9                            | 45.8                            | 98.5                          | 99.6                          | 99.8                          |
| 0.5             | 40           | Significant final              | 0.1                             | 0.0                             | 0.1                           | 0.1                           | 0.2                           |
| 0.8             | 20           | NS                             | 8.7                             | 3.6                             | 9.6                           | 11.6                          | 30.3                          |
| 0.8             | 20           | Significant interim            | 47.6                            | 14.3                            | 49.4                          | 50.9                          | 63.2                          |
| 0.8             | 20           | Significant final              | 2.8                             | 0.8                             | 3.1                           | 4.3                           | 6.6                           |
| 0.8             | 40           | NS                             | 1.8                             | 0.4                             | 1.9                           | 2.0                           | 2.2                           |
| 0.8             | 40           | Significant interim            | 94.7                            | 41.6                            | 95.8                          | 96.3                          | 96.6                          |
| 0.8             | 40           | Significant final              | 1.0                             | 0.1                             | 1.0                           | 1.2                           | 1.2                           |
| 1.0             | 20           | NS                             | 24.1                            | 7.3                             | 26.2                          | 28.1                          | 56.7                          |
| 1.0             | 20           | Significant interim            | 33.8                            | 9.7                             | 35.0                          | 35.1                          | 40.6                          |
| 1.0             | 20           | Significant final              | 1.3                             | 0.3                             | 1.5                           | 2.1                           | 2.8                           |
| 1.0             | 40           | NS                             | 8.6                             | 2.1                             | 9.0                           | 9.4                           | 9.7                           |
| 1.0             | 40           | Significant interim            | 86.9                            | 37.0                            | 88.6                          | 88.8                          | 89.0                          |
| 1.0             | 40           | Significant final              | 1.1                             | 0.3                             | 1.1                           | 1.2                           | 1.2                           |

Significance of PSPRS evaluated at  $P < 0.05$  because no multiple testing correction will be applied to supplementary analysis. All calculations assume that once the interim is significant, the trial stops (and therefore there is no final analysis). A negligible amount of models between 0% and 1.6% failed to converge at the interim analysis because of missing follow-up data.

<sup>a</sup>True direct effect simulated on survival.

<sup>b</sup>Not significant either at the interim or the final analyses.

<sup>c</sup>Significant JRT at final analysis when the interim JRT was not significant.

Abbreviations: PSPRS, Progressive Supranuclear Palsy Rating Scale; HR, hazard ratio; ES, effect size; JRT, Joint Rank Test; sig, significant; M, month; %-sig PSPRS 12 M effect, percentage of significance for the Progressive Supranuclear Palsy Rating Scale at 12 months; %-sig PSPRS 24 M effect, percentage of significance for the Progressive Supranuclear Palsy Rating Scale at 24 months; NS, not significant.

and provide additional time for biological effects to be translated into clinical benefit, boosting power when combined with treatment effects on PSPRS. Second, ~20% mortality would be expected over 24 months, resulting in nonrandom missing data and resulting bias. We proposed to switch the primary endpoint from PSPRS to a combined assessment of function and survival using the CAFS. The CAFS acknowledges death as an important outcome (rather than a random event and loss of information if using PSPRS alone). This approach would also allow detection of moderate treatment effects and is currently the FDA-recommended approach to amyotrophic lateral sclerosis, including for ongoing phase 3 studies.<sup>23,30</sup>

The CAFS would be evaluated initially after all participants reached 12 months of treatment. If positive, the double-blind treatment period would stop, otherwise

treatment would continue until all participants reached 24 months, when the CAFS would be evaluated again. Overall, the trial is designed to maximize the probability for success at the interim analysis but retain alpha and power for the final analysis, in case of violation of assumptions. Continuation of the trial to 24 months if the interim analysis is negative is expected to be of value if:

- The effect on PSPRS is moderate and an effect on survival is present.
- Treatment effects are delayed (eg, differences in PSPRS progression are observed only from month 6 onward).
- Participant survival is higher than predicted.
- Significant differences in PSPRS are required/desirable at 24 months, regardless of the outcomes at 12 months.



**TABLE 3** Operating characteristic of survival analysis

| HR <sup>a</sup> | PSPRS ES (%) | Result of JRT (Total)          | % Survival Significant (P < 0.05) Interim | % Survival Significant (P < 0.05) Final | Probability of JRT Outcome (%) | Power Survival Interim | Power Survival Final |
|-----------------|--------------|--------------------------------|---|---|--------------------------------|------------------------|----------------------|
| 0.5             | Null         | NS                             | 27.7                                      | 38.9                                    | 47                             | 0.59                   | 0.83                 |
| 0.5             | Null         | Significant interim            | 29  | 29.8                                    | 30                             | 0.96                   | 0.99                 |
| 0.5             | Null         | Significant final              | 20.6                                      | 23.2                                    | 23                             | 0.88                   | 1                    |
| 0.5             | 20           | NS <sup>b</sup>                | 1.5                                       | 2.5                                     | 4                              | 0.4                    | 0.7                  |
| 0.5             | 20           | Significant interim            | 83.1                                      | 88.5                                    | 90                             | 0.92                   | 0.99                 |
| 0.5             | 20           | Significant final <sup>c</sup> | 4.4                                       | 6.4                                     | 7                              | 0.66                   | 0.98                 |
| 0.5             | 40           | Significant interim            | 94.9                                      | 99.2                                    | 100                            | 0.95                   | 0.99                 |
| 0.5             | 40           | Significant final              | 0.1                                       | 0.1                                     | 0                              | 0.38                   | 0.88                 |
| 0.8             | Null         | NS                             | 9.3                                       | 13.1                                    | 88                             | 0.11                   | 0.15                 |
| 0.8             | Null         | Significant interim            | 5.3                                       | 5.8                                     | 8                              | 0.62                   | 0.69                 |
| 0.8             | Null         | Significant final              | 1.5                                       | 2.8                                     | 3                              | 0.44                   | 0.85                 |
| 0.8             | 20           | NS                             | 1.7                                       | 3.2                                     | 30                             | 0.06                   | 0.11                 |
| 0.8             | 20           | Significant interim            | 29.1                                      | 38.1                                    | 63                             | 0.46                   | 0.6                  |
| 0.8             | 20           | Significant final              | 1.3                                       | 4                                       | 7                              | 0.2                    | 0.61                 |
| 0.8             | 40           | NS                             | 0   | 0.1                                     | 2                              | 0                      | 0.03                 |
| 0.8             | 40           | Significant interim            | 47.6                                      | 65.2                                    | 97                             | 0.49                   | 0.67                 |
| 0.8             | 40           | Significant final              | 0.1                                       | 0.3                                     | 1                              | 0.07                   | 0.26                 |

All calculations assume that once the interim is significant, the trial stops, and therefore no final analysis is performed.

<sup>a</sup>True direct effect simulated on survival.

<sup>b</sup>Not significant either at the interim or the final analyses.

<sup>c</sup>Significant at final analysis when the interim was not significant.

Abbreviations: HR, hazard ratio; PSPRS, Progressive Supranuclear Palsy Rating Scale; ES, effect size; JRT, Joint Rank Test; NS, not significant.

- Data collected at interim analysis have higher variability than predicted.
- Recruitment occurs at an accelerated rate, in which case, direct treatment effects on survival would be key to a significant result.

IRT modeling was useful in simulating a CT without any underlying hypothesis about the shape of the disease progression profile versus time. In addition, IRT models can be used as psychometric tools to describe performance of the PSPRS in assessing disease progression. Based on IRT modeling results, different items in the PSPRS can be ranked indicating which items are most sensitive when describing disease severity.

The design was conceived as a single pivotal study, to provide definitive evidence of efficacy and a comprehensive safety package, in the absence of any prior proof of concept/efficacy study, and assuming no parallel phase 3 study would be feasible due to the rarity of PSP. The study would provide strong evidence to support a regulatory submission in the context of high unmet need in a rare disease. Our proposal is an innovative and

scientifically robust approach to provide PSP patients accelerated access to a much-needed treatment.

In general, IRT modeling results confirmed expected limitations of the PSPRS scoring approach to measure disease progression. Intermediate scores were often problematic, including many of the most sensitive items, demonstrating tendency of the investigators to consider these items as dichotomous. This reinforces the requirement to revise the answer categories, or at least scoring of those items. Revisions to the PSPRS were not an objective of this trial; however, the presented JRT approach was also tested on a smaller scale, focusing on the 10 items considered more clinically relevant by the FDA.<sup>31</sup> Although the IRT model confirmed that the selected items were among the most informative, the 10-item scale had a negligible impact on operating characteristics compared with the 28-item scale (Supporting Information Table S4). The expected mortality rate was derived from the Natural History and riluzole trial, which is 20 years old.<sup>25</sup> There have since been advances in medical treatment and earlier recognition of possible PSP.<sup>32</sup> Therefore, referral to a trial

would occur at a point in the disease course with lesser mortality, resulting in more prolonged post-enrollment survival than the model anticipates.

In addition, distortion in mortality rate could occur if another pandemic heavily affects the elderly or disabled, therefore affecting only part of each patient's clinical course, injecting statistical noise, and reducing the statistical power of the CAFS approach.

Switching the primary endpoint from change in the PSPRS to the CAFS mitigates against nonrandom "mis-singness" due to participant mortality and consequential bias. The CAFS could facilitate a comprehensive and robust estimate of the clinical benefit of potential future therapies for people living with PSP and optimize the probability of clinical success (if the drug is effective). These results provide guidance for the PSP community around design and interpretation of future CTs. ■

**Acknowledgments:** Medical writing support for the development of this manuscript, under the direction of the authors, was provided by Dolapo Odunjinrin, PhD Researcher, and Sarah Hibbert, PhD, of Ashfield MedComms, an Inizio company. Medical writing support was funded by UCB in accordance with Good Publication Practice (GPP 2022) guidelines (<http://www.ismpp.org/gpp-2022>).

## Data Availability Statement

Data from non-clinical studies is outside of UCB's data sharing policy and is unavailable for sharing.

## References

- National Institute of Neurological Disorders and Stroke (NINDS). Progressive supranuclear palsy (PSP); Accessed August 08, 2024 <https://www.ninds.nih.gov/health-information/disorders/progressive-supranuclear-palsy-pp>.
- Hoglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the Movement Disorder Society criteria. *Mov Disord* 2017;32(6):853–864.
- Boxer AL, Yu JT, Golbe LI, Litvan I, Lang AE, Hoglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. *Lancet Neurol* 2017;16(7):552–563.
- Respondek G, Hoglinger GU. The phenotypic spectrum of progressive supranuclear palsy. *Parkinsonism Relat Disord* 2016;22 (Suppl 1):S34–S36.
- dell'Aquila C, Zoccolella S, Cardinali V, et al. Predictors of survival in a series of clinically diagnosed progressive supranuclear palsy patients. *Parkinsonism Relat Disord* 2013;19(11):980–985.
- Lyons S, O'Dowd S, Walsh R, Lynch T. Progressive supranuclear palsy in 2022: recent developments and an eye to the future. United Kingdom: ACNR, *Advances in Clinical Neuroscience and Rehabilitation*; 2022.
- Irwin DJ. Tauopathies as clinicopathological entities. *Parkinsonism Relat Disord* 2016;22(1):S29–S33.
- Kovacs GG, Lukic MJ, Irwin DJ, et al. Distribution patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathol* 2020;140(2):99–119.
- Lamb R, Rohrer JD, Lees AJ, Morris HR. Progressive supranuclear palsy and corticobasal degeneration: pathophysiology and treatment options. *Curr Treat Options Neurol* 2016;18(9):42.
- Limorenko G, Lashuel HA. Revisiting the grammar of Tau aggregation and pathology formation: how new insights from brain pathology are shaping how we study and target Tauopathies. *Chem Soc Rev* 2022;51(2):513–565.
- Dam T, Boxer AL, Golbe LI, et al. Safety and efficacy of anti-tau monoclonal antibody gosuranemab in progressive supranuclear palsy: a phase 2, randomized, placebo-controlled trial. *Nat Med* 2021;27(8):1451–1457.
- Hoglinger GU, Litvan I, Mendonca N, et al. Safety and efficacy of tilavonemab in progressive supranuclear palsy: a phase 2, randomized, placebo-controlled trial. *Lancet Neurol* 2021;20(3):182–192.
- Stamelou M, Schope J, Wagenpfeil S, et al. Power calculations and placebo effect for future clinical trials in progressive supranuclear palsy. *Mov Disord* 2016;31(5):742–747.
- Golbe LI, Ohman-Strickland PA. A clinical rating scale for progressive supranuclear palsy. *Brain* 2007;130(Pt 6):1552–1565.
- Ghosh BC, Carpenter RH, Rowe JB. A longitudinal study of motor, oculomotor and cognitive function in progressive supranuclear palsy. *PLoS One* 2013;8(9):e74486.
- Tolosa E, Litvan I, Hoglinger GU, et al. A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. *Mov Disord* 2014;29(4):470–478.
- Apetauerova D, Scala SA, Hamill RW, et al. CoQ10 in progressive supranuclear palsy: a randomized, placebo-controlled, double-blind trial. *Neurol Neuroimmunol Neuroinflamm* 2016;3(5):e266.
- Leclair-Visonneau L, Rouaud T, Debilly B, et al. Randomized placebo-controlled trial of sodium valproate in progressive supranuclear palsy. *Clin Neurol Neurosurg* 2016;146:35–39.
- Nuebling G, Hensler M, Paul S, Zwergal A, Crispin A, Lorenzl S. PROSPERA: a randomized, controlled trial evaluating rasagiline in progressive supranuclear palsy. *J Neurol* 2016;263(8):1565–1574.
- Hewer S, Varley S, Boxer AL, Paul E, Williams DR, Investigators AL. Minimal clinically important worsening on the Progressive Supranuclear Palsy Rating Scale. *Mov Disord* 2016;31(10):1574–1577.
- Ramchandani R, Schoenfeld DA, Finkelstein DM. Global rank tests for multiple, possibly censored, outcomes. *Biometrics* 2016;72(3):926–935.
- Berry JD, Miller R, Moore DH, et al. The Combined Assessment of Function and Survival (CAFS): a new endpoint for ALS clinical trials. *Amyotroph Lateral Scler Frontotemporal Degener* 2013;14(3):162–168.
- Food and Drug Administration (FDA). Amyotrophic lateral sclerosis: developing drugs for treatment guidance for industry; Accessed August 08, 2024 <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/amyotrophic-lateral-sclerosis-developing-drugs-treatment-guidance-industry>.
- Boxer AL, Lang AE, Grossman M, et al. Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. *Lancet Neurol* 2014;13(7):676–685.
- Bensimon G, Ludolph A, Agid Y, et al. Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: the NNIPPS study. *Brain* 2009;132(Pt 1):156–171.
- Krekels E, Novakovic AM, Vermeulen AM, Friberg LE, Karlsson MO. Item response theory to quantify longitudinal placebo and paliperidone effects on PANSS scores in schizophrenia. *CPT Pharmacometrics Syst Pharmacol* 2017;6(8):543–551.
- Knopman DS. Lecanemab reduces brain amyloid-beta and delays cognitive worsening. *Cell Rep Med* 2023;4(3):100982.
- Ramanan VK, Day GS. Anti-amyloid therapies for Alzheimer disease: finally, good news for patients. *Mol Neurodegener* 2023;18(1):42.
- Brilleman SL, Wolfe R, Moreno-Betancur M, Crowther MJ. Simulating survival data using the *simSurv* R package. *J Stat Softw* 2021;3(97):1–27.
- Fournier CN. Considerations for amyotrophic lateral sclerosis (ALS) clinical trial design. *Neurotherapeutics* 2022;19(4):1180–1192.
- Yousefi E, Gewily M, König F, et al. Efficiency of multivariate tests in trials in progressive supranuclear palsy. *arXiv* 2023; arXiv:2312.08169 [stat.ME]. <https://arxiv.org/abs/2312.08169>

32. Mamarabadi M, Razjouvan H, Golbe LI. Is the latency from progressive supranuclear palsy onset to diagnosis improving? *Mov Disord Clin Pract* 2018;5(6):603–606.

## Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.