

RESEARCH ARTICLE

Clinical trials for Wolfram syndrome neurodegeneration: Novel design, endpoints, and analysis models

Guoqiao Wang^{1,2*}, Zhaolong Adrian Li³, Ling Chen², Heather Lugar^{1,3}, Tamara Hershey^{1,3,4}

1 Department of Neurology, Washington University in St Louis School of Medicine, St Louis, Missouri, United States of America, **2** Division of Biostatistics, Washington University in St Louis School of Medicine, St Louis, Missouri, United States of America, **3** Department of Psychiatry, Washington University in St Louis School of Medicine, St Louis, Missouri, United States of America, **4** Department of Radiology, Washington University in St Louis School of Medicine, St Louis, Missouri, United States of America

* guoqiao@wustl.edu, tammy@wustl.edu



OPEN ACCESS

Citation: Wang G, Li ZA, Chen L, Lugar H, Hershey T (2025) Clinical trials for Wolfram syndrome neurodegeneration: Novel design, endpoints, and analysis models. PLoS One 20(5): e0321598. <https://doi.org/10.1371/journal.pone.0321598>

Editor: Miriam Ann Hickey, University of Tartu, ESTONIA

Received: October 21, 2024

Accepted: March 8, 2025

Published: May 9, 2025

Copyright: © 2025 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The data used to compare the models were simulated and can be reproduced using the sample SAS code. The raw data used to obtain the parameters described in this manuscript cannot be made available in the manuscript, supplemental files, or a public repository because the sample

Abstract

Objective

Wolfram syndrome, an ultra-rare condition, currently lacks effective treatment options. The rarity of this disease presents significant challenges in conducting clinical trials, particularly in achieving sufficient statistical power (e.g., 80%). The objective of this study is to propose a novel clinical trial design based on real-world data to reduce the sample size required for conducting clinical trials for Wolfram syndrome.

Methods

We propose a novel clinical trial design with three key features aimed at reducing sample size and improve efficiency: (i) Pooling historical/external controls from a longitudinal observational study conducted by the Washington University Wolfram Research Clinic. (ii) Utilizing run-in data to estimate model parameters. (iii) Simultaneously tracking treatment effects in two endpoints using a multivariate proportional linear mixed effects model.

Results

Comprehensive simulations were conducted based on real-world data obtained through the Wolfram syndrome longitudinal observational study. Our simulations demonstrate that this proposed design can substantially reduce sample size requirements. Specifically, with a bivariate endpoint and the inclusion of run-in data, a sample size of approximately 30 per group can achieve over 80% power, assuming the placebo progression rate remains consistent during both the run-in and randomized periods. In cases where the placebo progression rate varies, the sample size increases to approximately 50 per group.

size of our Wolfram syndrome patient group is so small, and the disease so rare, that human participant characteristics such as sex, age, and number of visits could be used to identify individuals even after de-identification of the data. The corresponding author, Dr. Hershey, or Beth Beato bbeato@wustl.edu, Project/Study Manager in Department of Psychiatry at Washington University in St Louis, can be contacted to request data. As per the Human Research Protection Office (HRPO) at Washington University, a preface to data sharing agreement and a data sharing agreement reviewed by the research office will be employed prior to data sharing. HRPO regulations permit access to potentially identifiable data only to research personnel on our study protocol and approved through the University.

Funding: This work was supported by the National Institutes of Health (NIH) under grant number HD070855, The Snow Foundation, and the American Diabetes Association (ADA).

Competing interests: The authors have declared that no competing interests exist.

List of Abbreviations: ER: endoplasmic reticulum; WURS: Wolfram Unified Rating Scale; MRI: magnetic resonance imaging; logMAR: logarithm of the minimum angle of resolution; eTIV: estimated total intracranial volume; LME: linear mixed effects.

Conclusions

For rare diseases like Wolfram syndrome, leveraging existing resources such as historical/external controls and run-in data, along with evaluating comprehensive treatment effects using bivariate/multivariate endpoints, can significantly expedite the development of new drugs.

1 Introduction

Wolfram syndrome is a rare (estimated to affect 1 in 770,000–160,000 people), autosomal recessive, multisystem disease first defined in 1938 as the combination of childhood-onset insulin-dependent diabetes, optic nerve atrophy, diabetes insipidus, and hearing loss [1,2]. The major causative gene *WFS1* encodes for wolframin [3], an endoplasmic reticulum (ER) transmembrane glycoprotein involved in preventing cell death potentially through suppressing ER stress [4], regulating intracellular calcium homeostasis [5], and modulating mitochondrial function [6]. It is now understood that pathogenic mutations in *WFS1* can result in death or dysfunction of insulin-producing pancreatic β cells [3], causing insulin-dependent diabetes. *WFS1* mutations are also presumed to impact other organ systems [7], leading to a wide range of symptomatology. Several agents proposed to protect cells from ER stress-mediated apoptosis are being tested [5,8–12]; however, currently there are no approved treatments that alter Wolfram syndrome progression.

The identification of causative *WFS1* genetic mutations has allowed for a genetic rather than purely clinical diagnosis of Wolfram syndrome. Observations of genetically diagnosed Wolfram syndrome individuals indicate that the disease phenotype is indeed much more variable than classically understood, extending to loss of taste and smell [13–15], bladder and bowel dysfunction [16–18], gait and balance issues [19,20], and mental health disorders [13,21]. Wolfram syndrome is associated with early brain structural alterations such as reduced or stalled white matter development [22–25]. In addition, patterns of explicit neurodegeneration can be detected in gray matter regions over 2–3 years, regardless of age [23]. As these abnormalities progress and brain-mediated symptoms worsen, quality of life decreases and life-threatening conditions can develop (e.g., respiratory failure from brainstem atrophy) [26].

Identifying biomarkers of brain degeneration is thus a critical step towards designing efficient and effective clinical trials for candidate pharmaceutical agents or therapies. We have conducted the world's largest and longest natural history study of early Wolfram syndrome, focusing on quantification of neurological features of the disease and their changes over time. Over the 13 years of the Washington University Wolfram Research Clinic (<https://wolframsyndrome.wustl.edu/items/research-clinic/>; ClinicalTrials.gov Identifier NCT03951298), we validated a standardized clinical severity rating scale for Wolfram syndrome (the Wolfram Unified Rating Scale, WURS) [27], described unexpectedly early neurological symptoms and neuropathological differences, and characterized the rate of change in these symptoms and in regional brain volumes. We found that visual acuity and the volume of the thalamus (based on

magnetic resonance imaging [MRI]) deteriorate the most consistently and rapidly in our cohort [24,28]. This information has been used to justify outcome measures and predict power for safety [8] and clinical efficacy studies using standard designs (e.g., ClinicalTrials.gov Identifier: NCT03717909).

One of the primary challenges for clinical trials in rare diseases such as Wolfram syndrome is enrolling an adequate number of participants to ensure sufficient statistical power. Here, we leverage our unique longitudinal natural history dataset to explore the added power obtained (and reduced sample sizes required) by using three different innovative design features. First, we consider the impact of including historical data from patients who were not treated in the clinical trial. This approach allows more participants to receive active treatment without compromising statistical power compared to the traditional 1:1 randomization trial [29,30]. For example, a trial with 2:1 randomization can be analyzed as if it had a 1:1 ratio by borrowing the same number of historical/external controls as the randomized placebo controls. Second, we consider how best to use data collected from Wolfram syndrome patients prior to their participation in a clinical trial (i.e., 'run-in' data [31,32]). Run-in data aids in estimating important model parameters such as the rate of change of the placebo group or the variances, thereby enhancing statistical power [31,32]. Depending on whether the disease progression remains relatively stable during a moderate follow-up duration (e.g., a total of 4–6 years), the disease progression rate for participants on placebo can be assumed to be the same or different during the run-in period and the randomized period (Fig 1).

Finally, a single primary endpoint may not fully capture the important effects of an intervention, particularly for diseases with diverse manifestations such as Wolfram syndrome. In such cases, the use of multiple primary endpoints presents an appealing solution [33]. The multiple endpoints can be modeled in several ways: as a multivariate endpoint [34], as a composite endpoint (e.g., average [30,35] or sum [36]), or by analyzing each endpoint individually in a sequential manner. The use of an integrated scale such as a composite endpoint [37,38] has occurred in Alzheimer's disease trials and can

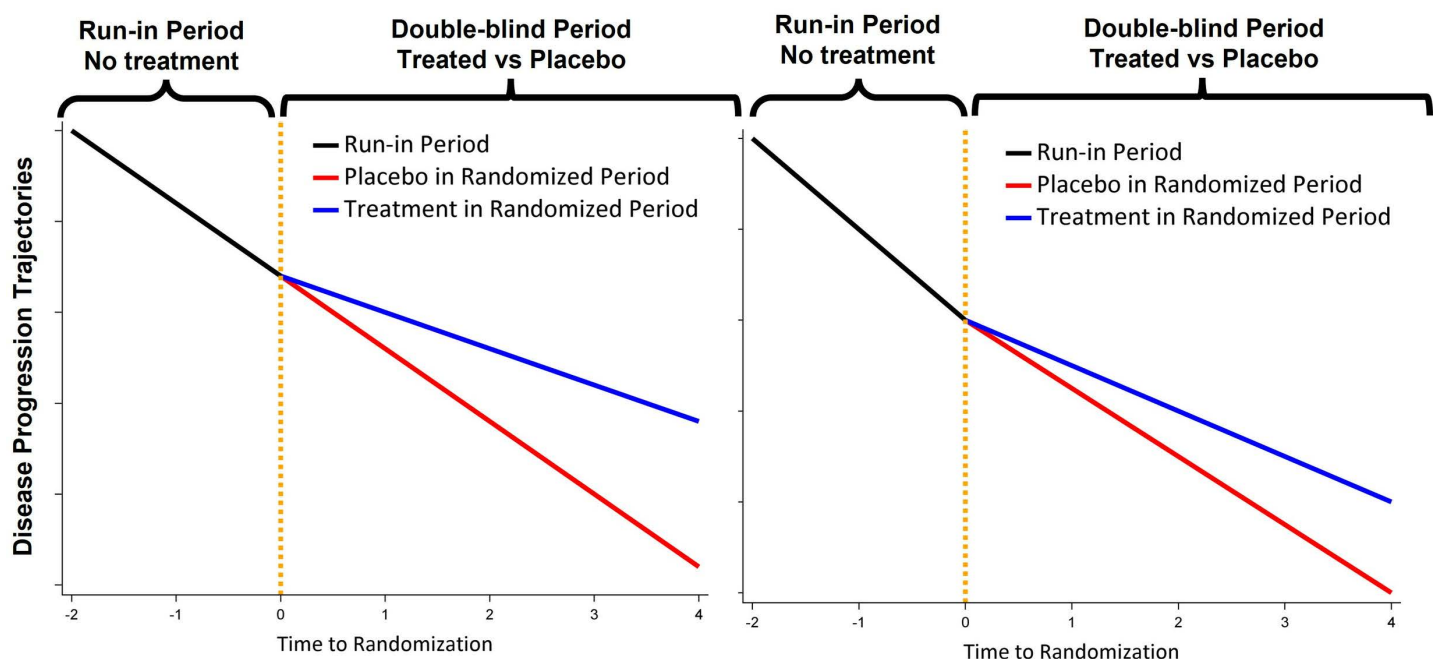


Fig 1. Demonstration of how clinical trials might use run-in data. Left: a trial design that uses the same disease progression rate for placebo during the run-in period and the randomized period; Right: a trial design that uses different disease progression rates for placebo during the run-in period and the randomized period.

<https://doi.org/10.1371/journal.pone.0321598.g001>

be acceptable to the FDA [39]. However, a composite endpoint may impose challenges in interpretation and may inadvertently assign disproportionate weight to a particular endpoint upon averaging or summing [40]. Analyzing multiple endpoints sequentially requires careful control of type I error and can sometimes result in a loss of statistical power [41]. To address these limitations, we propose a direct modeling approach that considers multiple endpoints simultaneously using proportional mixed models for repeated measures [42]. This approach allows for the identification of a single proportional reduction in disease progression across multiple endpoints, effectively capturing the treatment effect across a range of measures.

In this paper, we present the impact of these three design features on the predicted power and required sample sizes in clinical trials of interventions that target neurodegeneration in Wolfram syndrome. By using our unique longitudinal dataset from a natural history study of Wolfram syndrome, we can empirically determine the design that optimally balances the power to detect slowing of neurodegeneration while minimizing the required sample size. Future clinical trials targeting neurodegeneration in Wolfram syndrome may then use this information to optimize their trial designs and justify their proposed sample sizes.

2 Methods

Study Oversight: This study does not involve human subjects and is based on simulated data using information obtained from de-identified group-level data. Therefore, it does not require IRB approval. Consequently, participant consent is not required for this study.

Ethics approval and consent to participate do not apply to this study as it is primarily based on simulated data.

2.1 Data

Participants with genetically confirmed *WFS1* mutations, and under the age of 30 at enrollment, were assessed annually at the Washington University Wolfram Syndrome Research Clinic [23]. Briefly, at each annual visit participants were evaluated by a pediatric ophthalmologist or optometrist who measured best corrected visual acuity by Snellen optotype [28]. In the following analyses, we used the logarithm of the minimum angle of resolution (logMAR) value for best-corrected visual acuity with both eyes open; higher logMAR value indicates worse visual acuity. Participants also underwent magnetic resonance imaging where a T1-weighted high-resolution magnetization-prepared rapid gradient-echo (MPRAGE) sequence was acquired. Scans from each annual visit were processed longitudinally using the semi-automatic segmentation program FreeSurfer v5.3 [43], and regional brain volumes were extracted and corrected for estimated total intracranial volume (eTIV) [44]. Data acquired between 2010 and 2017 were included in these analyses.

2.2 Models

2.2.1 Univariate linear mixed effects model. The univariate linear mixed effects (LME) model that estimates a linear disease progression rate can be written as [42,45]:

$$y_{ijk} = (\mu_0 + u_{0i}) + (\beta_k + u_{1i}) t_{ij} + \varepsilon_{ij} \quad (1)$$

where y_{ijk} denotes the longitudinal assessments for subject i at time j for group k , $i = 1, 2, \dots, n$, $j = 0, 1, \dots, n_i$, with $j = 0$ representing the baseline visit, and $k = 1, 2$ representing the placebo group and the treatment group, respectively; μ_0 is the baseline intercept, β_k is the disease progression rate for group k where time t_{ij} represents subject-specific assessment time; u_{0i} , u_{1i} are the random effects for the intercept and the progression rate (i.e., slope) and are assumed to follow the same bivariate normal distribution for both groups: $\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{u_{0i}}^2 & \sigma_{u_{0i}u_{1i}} \\ \sigma_{u_{0i}u_{1i}} & \sigma_{u_{1i}}^2 \end{bmatrix}\right)$; the within-subject error is assumed to follow the same normal distribution for both groups $\varepsilon_{ij} \sim N(0, \sigma_e^2)$.

2.2.2 Treatment effect represented by difference in disease progression rates or proportional/percent reduction. When analyzing the primary endpoint using model (1), the treatment effect is commonly expressed as the difference between the annual disease progression rates, i.e., $\beta_1 - \beta_2$ for the two subject groups. An alternative approach is to model the treatment effect as the proportional/percent reduction in the disease progression rate relative to the placebo rate, as described below [29,42]:

$$y_{ijk} = (\mu_0 + u_{0i}) + ((1 - \theta_T I(k = 2)) \beta_1 + u_{1i}) t_{ij} + \varepsilon_{ij} \quad (2)$$

where θ_T is the proportional/percentage reduction treatment effect; I is the indicator function with $I(k = 2) = 1$ and $I(k = 1) = 0$; β_1 is the disease progression rate of the placebo group, and all the other parameters have the same meanings as those in model (1).

2.2.3 Multivariate proportional linear mixed effects model. Utilizing a proportional/percent reduction to represent the treatment effect offers several advantages: 1) enhances the interpretability of the treatment effect compared to solely focusing on the difference in slope of the absolute values; 2) enables the totality evaluation of the overall treatment effect across multiple endpoints; and 3) allows for a reduction in sample size by simultaneously modeling multivariate endpoints. Such multivariate endpoints can be modeled using a multivariate proportional mixed effects model as below, which accounts for the correlation among different endpoints through the use of random effects [42]:

$$y_{ijk} = (\mu_{0c} + u_{0ic}) + ((1 - \theta_T I(k = 2)) \beta_{1c} + u_{1ic}) t_{ij} + \varepsilon_{ijc} \quad (3)$$

Where, i represents the subject; j represents the visit/time where $j = 0$ at baseline; c is the index for the endpoint included in the multivariate endpoint; $k = 1, 2$ represents the placebo group and the treatment group, respectively; I is the indicator function with $I(k = 2) = 1$ and $I(k = 1) = 0$; μ_{0c} and β_{1c} represent the baseline intercept and the progression rate of the placebo group for endpoint c , respectively; u_{0ic} , u_{1ic} are the random effects for the intercept and the progression rate (slope) for subject i endpoint c , and are assumed to follow a multivariate normal distribution $N(0_{2c}, \Sigma_{2c \times 2c})$ (e.g., 4 by 4 unstructured covariance matrix if there are 2 endpoints); θ_T is the treatment effect represented as a proportional/percentage reduction of the placebo disease progression; and ε_{ijc} is the within-subject residual and has different variances for different endpoints.

3 Results

Characteristics of participants in the Wolfram syndrome longitudinal observational study have been previously described [23]. In the current study, data from a total of 37 patients genetically diagnosed with Wolfram syndrome were included. Among them, 21 individuals (57%) were female. The mean (SD) baseline age of the participants was 13.3 (5.5) years, and the mean (SD) follow-up duration was 2.9 (1.2) years.

3.1 Estimated annual disease progression rates and the associated variance/covariance

A bivariate linear mixed-effects model was employed to estimate the annual progression rates and the associated variance/covariance components for both visual acuity and thalamus volume simultaneously in participants with Wolfram syndrome. These two biomarkers were found to deteriorate the most consistently and rapidly in previous longitudinal observational data [24,28], suggesting usefulness as disease progression endpoints. To facilitate model convergence, thalamus volumes were normalized to z-scores using the baseline mean (SD) of healthy controls. This adjustment made the range of normalized thalamus volume and visual acuity more compatible. Estimates are presented in Table 1 and Table 2. For interpretability, the estimated intercept for visual acuity of 0.540 logMAR corresponded to about 20/69 ft (6/21 m; 3rd line) on the Snellen scale, and the estimated slope of 0.062 logMAR/y translated to approximately 7.4 years to deteriorate to 20/200 ft (6/60 m; 1st line).

3.2 Evaluation of model performance by study designs and endpoints

To evaluate model performance for each study design (i.e., historical controls, run-in data) and each type of endpoints (i.e., univariate, multivariate), we conducted simulations for various scenarios. The simulations were configured as follows:

- Individual baseline values and annual progression rates for visual acuity and thalamus volume were generated using a multivariate normal distribution with the mean and variance/covariance components specified in [Table 1](#) and [Table 2](#).
- Individual longitudinal data were simulated by combining the individual baseline values, annual progression rates, and within-subject error, according to model 3.
- Treatment effects were introduced as a proportional reduction in the placebo rate of progression at various levels: 0% (Type I error), 30%, and 40%.
- Time was set to 0 at baseline, negative before baseline (i.e., the run-in period), and positive after baseline.
- The run-in period lasted for 2 years, and the treatment period lasted for 3 years.
- Assessments were scheduled every 6 months.
- The per-visit dropout rate was set at 4.5%.
- The randomization ratio was 2:1 with pooled historical/external controls to make the power estimation based on a randomization ratio of 1:1, the sample sizes range from 30 per group to 120 per group.
- For each model, 1000 data sets were simulated, allowing for estimation of power with precision up to three decimal places [46].
- Type I error and power were calculated as the proportion of the 1000 simulated trials per scenario that yielded P-values < 0.05.
- Disease progression rates during both the run-in and randomized periods are the same vs different (slower progression during the randomized period due to the placebo and/or trial effect).

Table 1. Estimated annual disease progression rates and 95% CI for each endpoint.

Parameter	Estimate	SE	P-value	95% Confidence Limits	
Intercept of Visual Acuity (logMAR)	0.540	0.069	<.0001	0.399	0.681
Slope of Visual Acuity(LogMAR/year)	0.062	0.013	<.0001	0.035	0.089
Intercept of Thalamus Volume (z-score)	-1.116	0.165	<.0001	-1.451	-0.781
Slope of Thalamus Volume (z-score/year)	-0.139	0.023	<.0001	-0.185	-0.093

<https://doi.org/10.1371/journal.pone.0321598.t001>

Table 2. Estimated variance/covariance components for each endpoint.

	Intercept of Visual Acuity	Slope of Visual Acuity	Intercept of Thalamus Volume	Slope of Thalamus Volume
Intercept of Visual Acuity	0.174	0.020	-0.178	-0.020
Slope of Visual Acuity	0.020	0.004	-0.042	-0.002
Intercept of Thalamus Volume	-0.178	-0.042	0.811	0.025
Slope of Thalamus Volume	-0.020	-0.002	0.025	0.007
Residual Variance of Visual Acuity	0.025			
Residual Variance of Thalamus Volume	0.005			

<https://doi.org/10.1371/journal.pone.0321598.t002>

3.2.1 Type I error/power estimation with vs. without run-in data for a single primary endpoint (Visual acuity). Fig 2 illustrates the power of various design and analysis models. The inclusion of run-in data significantly enhances power compared to trials without run-in data, regardless of whether the placebo progression rates remain the same or not throughout both run-in and randomized periods. Nevertheless, when the placebo progression rates are the same during run-in and randomized periods (Fig 2A and 2B), utilization of run-in data results in a greater increase in power compared to scenarios where the placebo progression rates change (Fig 2C and 2D). Estimating the treatment effect as a percent reduction demonstrates higher power compared to estimating it as the difference in disease progression rates. Moreover, such gain in power is more pronounced when the disease progression rates for placebo

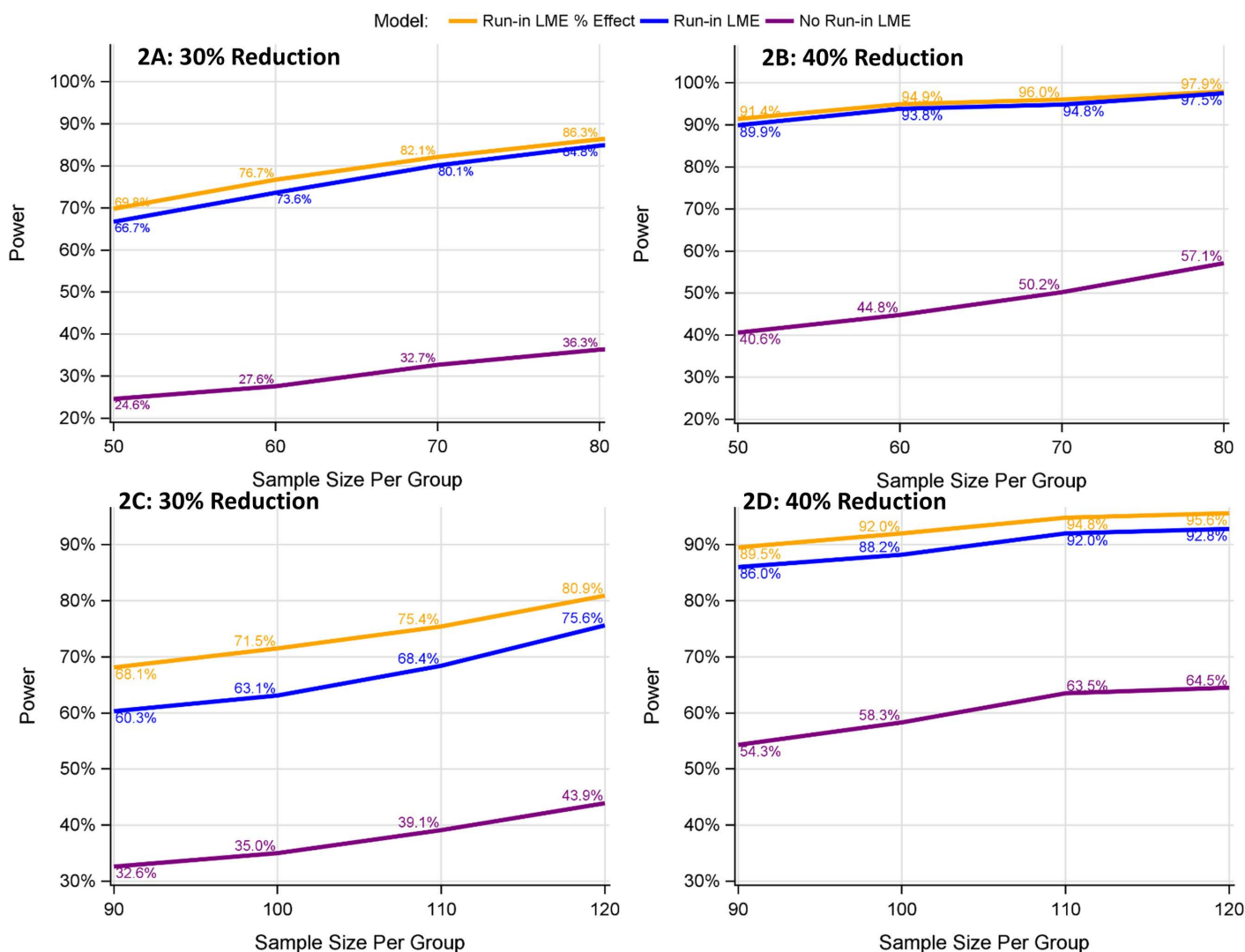


Fig 2. Power comparison between models with vs. without run-in data for the univariate endpoint of visual acuity. A and B: The disease progression rate for placebo participants is assumed to be the same throughout both the run-in and randomized periods (as in left panel in Fig 1). C and D: The disease progression rate for placebo participants is assumed to be slower in the randomized than the run-in period (as in right panel in Fig 1). 30%, 40% reduction: 30%, 40% reduction in the disease progression relative to the placebo group. No Run-in LME (purple lines): without run-in data analyzed using LME model 1. Run-in LME (blue lines): with run-in data analyzed using LME model 1. Run-in LME % Effect: (yellow lines) with run-in data analyzed using LME model 2.

<https://doi.org/10.1371/journal.pone.0321598.g002>

participants differ between the run-in and randomized periods. The type I error for all designs/models generally remains within 2% of the nominal 5% level (S1 Fig).

3.2.2 Type I error/power estimation for bivariate endpoint (visual acuity and thalamus volume). Fig 3 presents an illustration of power associated with bivariate endpoints, with or without run-in data. Consistent with the simulations conducted for the univariate endpoint models, the inclusion of run-in data enhances power considerably. Further, the

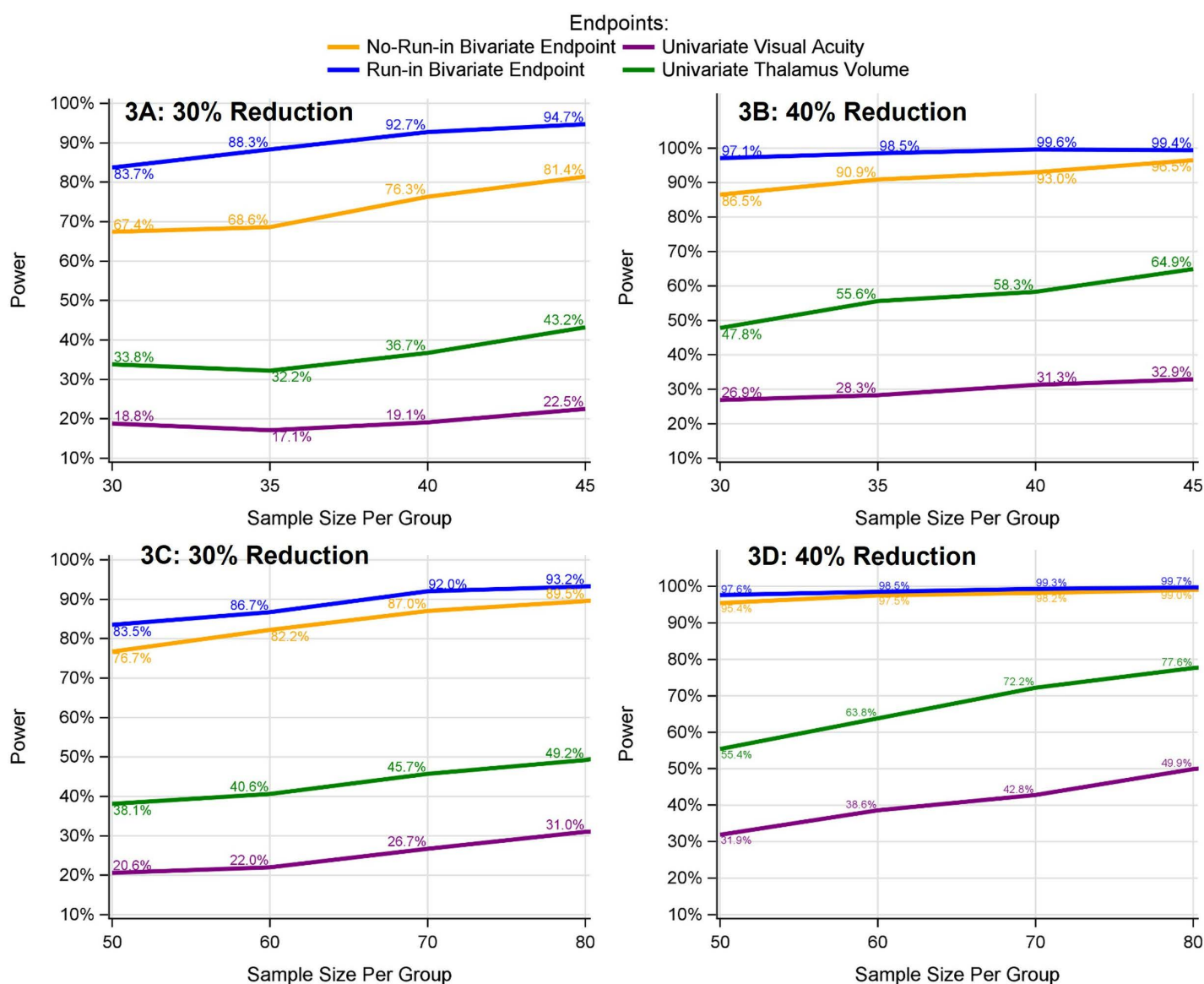


Fig 3. Power comparison between models with bivariate vs. univariate endpoints and with vs. without run-in data. A and B: The disease progression rate for placebo participants is assumed to be the same throughout both run-in and randomized periods (as in left panel in Fig 1). C and D: The disease progression rate for placebo participants is assumed to be slower during the randomized than run-in period (as in right panel in Fig 1). 30%, 40% reduction: 30%, 40% reduction in the disease progression relative to the placebo group. Univariate Visual Acuity: visual acuity analyzed by LME model 1. Univariate Thalamus Volume: thalamus volume analyzed by LME model 1; No-Run-in Bivariate Endpoint: bivariate endpoint of visual acuity and thalamus volume analyzed by model 3 without run-in data; Run-in Bivariate Endpoint: bivariate endpoint of visual acuity and thalamus volume analyzed by model 3 with run-in data.

<https://doi.org/10.1371/journal.pone.0321598.g003>

bivariate endpoint provides a significant gain in power, effectively doubling or even tripling the power observed with the univariate endpoint. Overall, the type I error for all endpoints/designs typically stays within 2% of the expected 5% level (S1 Fig).

4 Discussion

Our results demonstrate the range of power that can be obtained in clinical trials targeting neurodegeneration in Wolf-Parkinson-White syndrome using practical and data-driven approaches. By leveraging our unique existing natural history dataset, we demonstrate the considerable potential gain in power achieved by using run-in and historical/external control data and by choosing a bivariate primary endpoint, assuming a 3-year follow-up with 6-month assessment intervals. After taking into account the assumptions and caveats associated with these models, investigators will be able to use this information and maximize existing resources to guide and standardize clinical trial protocols going forward, leading to more efficient trial designs for this ultra-rare disease.

The three features of the proposed design that contributed to increased power were the borrowing of historical/external controls, use of run-in data, and a bivariate primary endpoint of visual acuity and thalamus volume. Each of these design features boosted our power estimates, although there are some additional issues that need to be discussed. First, by leveraging historical data from our natural history study, we can employ an unequal randomization ratio, such as 2:1 with two times the number of patients on the active treatment. Our simulations assumed that we could borrow enough historical/external controls so that the efficacy analysis would have a 1:1 randomization ratio. In a real-world trial, however, the availability of historical/external controls will vary. Second, although each individual had 2 years of run-in data in the simulation, it is not necessary that the run-in data have equal duration. Any amount of run-in data with various durations can help [31]. In any trial that includes run-in data, it is recommended to conduct tailored simulations to evaluate the specific power gain from these run-in data. Unlike run-in data, which originate from participants in the trial, external controls are often collected from patients not directly involved in the ongoing trial. To ensure their appropriateness and comparability with concurrent controls, several measures should be implemented: (i) Apply the same inclusion and exclusion criteria to both the external control group and the concurrent control group to maintain a similar baseline population. (ii) Use techniques such as propensity score matching to align external controls with concurrent controls based on demographic, clinical, and other relevant baseline characteristics. (iii) Perform adjustments during the analysis to account for any residual differences in baseline characteristics between external and concurrent controls. (iv) Select external controls from studies or datasets with comparable conditions, including similar disease stages, treatment environments, and diagnostic criteria. These steps ensure that the target population for treatment effect estimation is consistent between trial participants and external controls, thereby reducing the risk of bias and enhancing the validity of the study results.

Finally, our simulations indicated that relying on a single endpoint is not ideal for such a rare disease with a wide spectrum of symptomatology. A bivariate endpoint (visual acuity and thalamus volume) not only significantly improved statistical power, but also would provide a more comprehensive evaluation of treatment effects across the disease spectrum without compromising type I error control. However, the bivariate endpoint method assumes that the proportional treatment effect is approximately the same across endpoints. Although this assumption might seem strong, it has been implicitly used in various composite endpoints [29,30,47]. When the proportional treatment effect differs between endpoints, the model will estimate an average treatment effect across all endpoints, similar to the use of a composite score of averaging multiple endpoints [29,30,47]. We recommend conducting a sensitivity bivariate analysis to estimate the proportional treatment effect for each endpoint separately. By comparing the average treatment effect to endpoint-specific treatment effects, a comprehensive evaluation of the treatment effect can be achieved. In this study, we compare various scenarios based on the power differences for a given sample size. Alternatively, we can quantify the sample size savings required to achieve 80% power. For example, in Fig 2B, our simulations show that the traditional LME model requires a sample size of 140 per group to achieve approximately 81% power. In contrast, the run-in LME % Effect model achieves about 80%

power with only 40 participants per group, resulting in a savings of about 100 participants per group. Researchers can adapt the provided sample SAS code to evaluate sample size savings for other scenarios in a similar manner.

While not directly addressed in our simulations, it may be beneficial for the field to consider a platform trial with a master protocol for early Wolfram syndrome. A platform trial is a method for evaluating multiple targeted therapies within a single disease setting, using a perpetual framework where therapies can enter or exit the platform based on predefined decision algorithms [48]. Building platform trials with master protocols for Wolfram syndrome will enable more efficient utilization of resources and accelerate the pace of drug discovery. Such trials facilitate the establishment of a trial network infrastructure that encourages extensive collaboration among researchers in the field of rare diseases [48]. Currently, a group of investigators across Europe are conducting a multi-site clinical trial (ClinicalTrials.gov Identifier: NCT03717909) with a master protocol. We would advocate for a similar approach within the US, ideally calibrating measures across sites, clinics and regions.

Our study has several limitations. First, our natural history data were obtained from a non-random group of participants. Enrolled participants were generally young, in an early disease state, and able to travel to St. Louis for annual several day-long visits. In addition, they were able to perform visual acuity tests (i.e., were not already blind) and did not have any contraindications for MRI scans (e.g., no cochlear implants, bladder stimulators, etc.). As a result, older, more impaired participants were not included in the study. However, the younger and less impaired individuals are the ones that presumably would benefit the most from slowing neurodegeneration and avoiding irreversible consequences of Wolfram syndrome. Second, there may be other, unexplored measures of neurodegeneration that could be more precise or more easily collected than MRI-based regional brain volumes. For example, fluid biomarkers are becoming more available and acceptable in clinical trials for other neurodegenerative conditions [37,49] and may have promise in Wolfram syndrome [50]. Third, the proposed multivariate proportional linear mixed-effects model assumes linearity in the disease progression trajectory to estimate an overall comparison between treatment and placebo groups. However, when linearity is a concern, alternative methods can be employed. For example, non-parametric approaches like the mixed model for repeated measures (MMRM) do not rely on a linearity assumption [47]. Additionally, parametric methods, such as models using spline-based time variables, can capture more complex, nonlinear progression patterns [38]. These methods can also incorporate a proportional treatment effect framework, allowing multiple endpoints to be combined, which may enhance statistical power compared to testing single endpoints independently [29,42,51,52]. Lastly, our study aims to enhance the feasibility of clinical trials for rare diseases by reducing sample size requirements through the use of alternative bivariate endpoints and proportional linear mixed-effects models. However, the selection of specific endpoints for inclusion in the bivariate endpoint, as well as the assumptions underlying the model, should be carefully evaluated using available data, including historical datasets or results from phase 2 trials, before implementation in a phase 3 trial.

In this paper, we propose a novel paradigm for conducting clinical trials in Wolfram syndrome, focusing on optimizing available resources and statistical approaches to enhance the feasibility of trials in this rare disease. The use of historical/external controls, run-in data, and a bivariate endpoint can greatly reduce the sample size required, thus facilitating the progress towards identifying truly effective drugs for this rare disease.

Supporting information

S1 Fig. Type I error by sample size for different endpoints/models. A: Single primary endpoint, placebo progression rates remain the same during both the run-in and randomized periods; B: Single primary endpoint, placebo progression rates are different during the run-in and randomized periods; C: Placebo progression rates remain the same during both the run-in and randomized periods; D: Placebo progression rates are different during the run-in and randomized periods. (DOCX)

Acknowledgments

None

Author contributions

Conceptualization: Guoqiao Wang, Zhaolong Adrian Li, Ling Chen, Heather Lugar, Tamara Hershey.

Data curation: Guoqiao Wang.

Formal analysis: Guoqiao Wang.

Methodology: Guoqiao Wang, Zhaolong Adrian Li, Ling Chen, Heather Lugar, Tamara Hershey.

Supervision: Tamara Hershey.

Writing – original draft: Guoqiao Wang, Zhaolong Adrian Li, Ling Chen, Heather Lugar, Tamara Hershey.

Writing – review & editing: Guoqiao Wang, Zhaolong Adrian Li, Ling Chen, Heather Lugar, Tamara Hershey.

References

1. DJ W. Diabetes mellitus and simple optic atrophy among siblings: report of 4 cases. *Mayo Clin Proc.* 1938
2. Barrett TG, Bunday SE, Macleod AF. Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome. *The Lancet*, 1995. 346(8988): 1458-1463.
3. Inoue H, Tanizawa Y, Wasson J, Behn P, Kalidas K, Bernal-Mizrachi E, et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet.* 1998;20(2):143–148. <https://doi.org/10.1038/2441> PMID: [9771706](https://pubmed.ncbi.nlm.nih.gov/9771706/)
4. Shang L, Hua H, Foo K, Martinez H, Watanabe K, Zimmer M et al., beta-cell dysfunction due to increased ER stress in a stem cell model of Wolfram syndrome. *Diabetes*, 2014. 63(3): 923-33.
5. Lu S, Kanekura K, Hara T, Mahadevan J, Spears LD, Osowski CM, et al. A calcium-dependent protease as a potential therapeutic target for Wolfram syndrome. *Proc Natl Acad Sci U S A.* 2014;111(49):E5292-301. <https://doi.org/10.1073/pnas.1421055111> PMID: [25422446](https://pubmed.ncbi.nlm.nih.gov/25422446/)
6. Zmyslowska A, Kuljanin M, Malachowska B, Stanczak M, Michalek D, Wlodarczyk A, et al., Multiomic analysis on human cell model of wolfram syndrome reveals changes in mitochondrial morphology and function. *Cell Commun Signal*, 2021. 19(1): 116.
7. Urano F. Wolfram syndrome: diagnosis, management, and treatment. *Curr Diab Rep.* 2016;16(1):1-8. PMID: [26742931](https://pubmed.ncbi.nlm.nih.gov/26742931/)
8. Abreu D, Stone SI, Pearson TS, Bucelli RC, Simpson AN, Hurst S, et al. A phase Ib/IIa clinical trial of dantrolene sodium in patients with Wolfram syndrome. *JCI Insight.* 2021;6(15):e145188. <https://doi.org/10.1172/jci.insight.145188> PMID: [34185708](https://pubmed.ncbi.nlm.nih.gov/34185708/)
9. Abreu D, Urano F. Current landscape of treatments for Wolfram syndrome. *Trends Pharmacol Sci.* 2019;40(10):711–714. <https://doi.org/10.1016/j.tips.2019.07.011> PMID: [31420094](https://pubmed.ncbi.nlm.nih.gov/31420094/)
10. Gorgogietas V, Rajaei B, Heeyoung C, Santacreu BJ, Marín-Cañas S, Salpea P, et al. GLP-1R agonists demonstrate potential to treat Wolfram syndrome in human preclinical models. *Diabetologia.* 2023;66(7):1306–1321. <https://doi.org/10.1007/s00125-023-05905-8> PMID: [36995380](https://pubmed.ncbi.nlm.nih.gov/36995380/)
11. Mahadevan J, Morikawa S, Yagi T, Abreu D, Lu S, Kanekura K, et al. A soluble endoplasmic reticulum factor as regenerative therapy for Wolfram syndrome. *Lab Invest.* 2020;100(9):1197–207. <https://doi.org/10.1038/s41374-020-0436-1> PMID: [32366942](https://pubmed.ncbi.nlm.nih.gov/32366942/)
12. Nguyen LD, Fischer TT, Abreu D, Arroyo A, Urano F, Ehrlich BE. Calpain inhibitor and ibudilast rescue β cell functions in a cellular model of Wolfram syndrome. *Proc Natl Acad Sci U S A.* 2020;117(29):17389–98. <https://doi.org/10.1073/pnas.2007136117> PMID: [32632005](https://pubmed.ncbi.nlm.nih.gov/32632005/)
13. Bischoff AN, Reiersen AM, Buttlare A, Al-Lozi A, Doty T, Marshall BA, et al. Selective cognitive and psychiatric manifestations in Wolfram syndrome. *Orphanet J Rare Dis.* 2015. 10(1):66.
14. Alfaro R. Taste and smell function in Wolfram syndrome. *Orphanet J Rare Dis.* 2020;15(1):57. <https://doi.org/10.1186/s13023-020-1335-7> PMID: [32087739](https://pubmed.ncbi.nlm.nih.gov/32087739/)
15. Alfaro R, Nicanor-Carreón JG, Doty T, Lugar H, Hershey T, Pepino MY. Enhancement of taste by retronasal odors in patients with Wolfram syndrome and decreased olfactory function. *Chem Senses.* 2023:48.
16. Doty T, Foster ER, Marshall B, Ranck S, Hershey T. The effects of disease-related symptoms on daily function in Wolfram syndrome. *Transl Sci Rare Dis.* 2017;2(1–2):89–100. <https://doi.org/10.3233/TRD-170012> PMID: [29130034](https://pubmed.ncbi.nlm.nih.gov/29130034/)
17. Marshall B.A., Permutt M.A., Paciorkowski A.R., Hoekel J., Karzon R., Wasson J., et al. Phenotypic characteristics of early Wolfram syndrome. *Orphanet J Rare Dis.* 2013;8:64. <https://doi.org/10.1186/1750-1172-8-64> PMID: [23981289](https://pubmed.ncbi.nlm.nih.gov/23981289/)
18. Rove KO, Vricella GJ, Hershey T, Thu MH, Lugar HM, Vetter J, et al. Lower urinary tract dysfunction and associated pons volume in patients with Wolfram syndrome. *J Urol.* 2018.

19. Pickett KA, Duncan RP, Hoekel J, Marshall B, Hershey T, Earhart GM, et al. Early presentation of gait impairment in Wolfram syndrome. *Orphanet J Rare Dis.* 2012;7:92. <https://doi.org/10.1186/1750-1172-7-92> PMID: [23217193](#)
20. Pickett KA, Duncan RP, Paciorkowski AR, Permutt MA, Marshall B, Hershey T, et al. Balance impairment in individuals with Wolfram syndrome. *Gait.* 2012;36(3):619–624. <https://doi.org/10.1016/j.gaitpost.2012.06.008> PMID: [22771154](#)
21. Reiersen AM, Noel JS, Doty T, Sinkre RA, Narayanan A, Hershey T. Psychiatric diagnoses and medications in Wolfram syndrome. *Scand J Child Adolesc Psychiatr Psychol.* 2022 10(1):163–174
22. Hershey T, Lugar HM, Shimony JS, Rutlin J, Koller JM, Perantie DC, et al. Early brain vulnerability in Wolfram syndrome. *PLoS One.* 2012;7(7):e40604. <https://doi.org/10.1371/journal.pone.0040604> PMID: [22792385](#)
23. Lugar HM, Koller JM, Rutlin J, Eisenstein SA, Neyman O, Narayanan A, et al. Evidence for altered neurodevelopment and neurodegeneration in Wolfram syndrome using longitudinal morphometry. *Sci Rep.* 2019;9(1):6010. <https://doi.org/10.1038/s41598-019-42447-9> PMID: [30979932](#)
24. Lugar HM, Koller JM, Rutlin J, Marshall BA, Kanekura K, Urano F, et al., Neuroimaging evidence of deficient axon myelination in Wolfram syndrome. *Scient Rep.* 2016. 6(1):21167.
25. Samara A, Rahn R, Neyman O, Park KY, Samara A, Marshall B, et al. Developmental hypomyelination in Wolfram syndrome: new insights from neuroimaging and gene expression analyses. *Orphanet J Rare Dis.* 2019;14(1):279. <https://doi.org/10.1186/s13023-019-1260-9> PMID: [31796109](#)
26. Minton JAL, Rainbow LA, Ricketts C, Barrett TG. Wolfram syndrome. *Rev Endocr Metab Disord.* 2003 4(1): 53–59
27. Nguyen C, Foster ER, Paciorkowski AR, Viehoveer A, Considine C, Bondurant A, et al. Reliability and validity of the Wolfram Unified Rating Scale (WURS). *Orphanet J Rare Dis.* 2012;7:89. <https://doi.org/10.1186/1750-1172-7-89> PMID: [23148655](#)
28. O'Bryhim B.E., Samara A., Chen L., Hershey T., Tychsen L., Hoekel J. Longitudinal changes in vision and retinal morphology in Wolfram syndrome. *Am J Ophthalmol.* 2022;243:10–18. <https://doi.org/10.1016/j.ajo.2022.07.003> PMID: [35850251](#)
29. Wang G, Berry S, Xiong C, Hassenstab J, Quintana M, McDade EM, et al. A novel cognitive disease progression model for clinical trials in autosomal-dominant Alzheimer's disease. *Stat Med.* 2018
30. Bateman RJ, Benzinger TL, Berry S, Clifford DB, Duggan C, Fagan AM, et al., The DIAN-TU next generation Alzheimer's prevention trial: adaptive design and disease progression model. *Alzheimer's & Dementia.* 2017. 13(1):8–19.
31. Wang G, Aschenbrenner AJ, Li Y, McDade E, Liu L, Benzinger TLS, et al. Two-period linear mixed effects models to analyze clinical trials with run-in data when the primary outcome is continuous: applications to Alzheimer's disease. *Alzheimer's Dementia: Transl Res Clin Interv.* 2019. 5:450–457.
32. Madsen K, Miller J, Province M. The use of an extended baseline period in the evaluation of treatment in a longitudinal Duchenne muscular dystrophy trial. *Stat Med.* 1986;5(3):231–241. <https://doi.org/10.1002/sim.4780050304> PMID: [3526500](#)
33. Hamasaki T, Evans SR, Asakura K. Design, data monitoring, and analysis of clinical trials with co-primary endpoints: a review. *J Biopharm Stat.* 2018;28(1):28–51. <https://doi.org/10.1080/10543406.2017.1378668> PMID: [29083951](#)
34. Salloway S, Farlow M, McDade E, Clifford DB, Wang G, Llibre-Guerra JJ, et al. A trial of gantenerumab or solanezumab in dominantly inherited Alzheimer's disease. *Nature Med.* 2021. 27(7):1187–1196.
35. Papp KV, Rofael H, Veroff AE, Donohue MC, Wang S, Randolph C, et al. Sensitivity of the preclinical Alzheimer's cognitive composite (PACC), PACC5, and repeatable battery for neuropsychological status (RBANS) to amyloid status in preclinical Alzheimer's disease-atabecestat phase 2b/3 EARLY clinical trial. *J Prevent Alzheimer's Disease.* 2022. 9(2): 255–261.
36. Mintun MA, Lo AC, Duggan Evans C, Wessels AM, Ardayfio PA, Andersen SW, et al. Donanemab in early Alzheimer's disease. *New England J Med.* 2021. 384(18): 1691–1704.
37. Salloway S, Farlow M, McDade E, Clifford DB, Wang G, Llibre-Guerra JJ, et al. A trial of gantenerumab or solanezumab in dominantly inherited Alzheimer's disease. *Nature Medicine.* 2021:1–10.
38. Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks J, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA.* 2023
39. Morant AV, Vestergaard HT, Lassen AB, Navikas V. US, EU, and Japanese regulatory guidelines for development of drugs for treatment of Alzheimer's disease: implications for global drug development. *Clin Transl Sci.* 2020;13(4):652–664. <https://doi.org/10.1111/cts.12755> PMID: [32043310](#)
40. Xiong C, van Belle G, Chen K, Tian L, Luo J, Gao F, et al. Combining multiple markers to improve the longitudinal rate of progression: application to clinical trials on the early stage of Alzheimer's disease. *Stat Biopharm Res.* 2013. 5(1):54–66.
41. Guidance D. Multiple endpoints in clinical trials guidance for industry. Center for Biologics Evaluation and Research (CBER). 2017.
42. Wang G, Liu L, Li Y, Aschenbrenner AJ, Bateman RJ, Delmar P, et al. Proportional constrained longitudinal data analysis models for clinical trials in sporadic Alzheimer's disease. *Alzheimer's & Dementia: Transl ResClin Interv.* 2022. 8(1):e12286.
43. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron.* 2002;33(3):341–355. [https://doi.org/10.1016/s0896-6273\(02\)00569-x](https://doi.org/10.1016/s0896-6273(02)00569-x) PMID: [11832223](#)
44. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage.* 2004. 23(2):724–38.

45. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. 998. John Wiley & Sons;2012.
46. Wang G, Kennedy RE, Cutter GR, Schneider LS. Effect of sample size re-estimation in adaptive clinical trials for Alzheimer's disease and mild cognitive impairment. *Alzheimers Dement (N Y)*. 2015;1(1):63–71. <https://doi.org/10.1016/j.trci.2015.03.002> PMID: [29854926](#)
47. Van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early Alzheimer's disease. *New England J Med*. 2023. 388(1):9-21.
48. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377(1):62–70. <https://doi.org/10.1056/NEJMr1510062> PMID: [28679092](#)
49. Wang G, Coble D, McDade EM, Hassenstab J, Fagan AM, Benzinger TLS, et al. Staging biomarkers in preclinical autosomal dominant Alzheimer's disease by estimated years to symptom onset. *Alzheimer's & Dementia*. 2019. 15(4):506-514.
50. Eisenstein SA, Boodram RS, Sutphen CL, Lugar HM, Gordon BA, Marshall BA, et al. Plasma neurofilament light chain levels are elevated in children and young adults with wolfram syndrome. *Front Neurosci*. 2022;16:795317. <https://doi.org/10.3389/fnins.2022.795317> PMID: [35495027](#)
51. Wang G, Wang W, Mangal B, Liao Y, Schneider L, Li Y, et al. Novel non-linear models for clinical trial analysis with longitudinal data: a tutorial using SAS for both frequentist and Bayesian methods. *Stat Med*. 2024
52. Wang G, Hassenstab J, Li Y, Aschenbrenner AJ, McDade EM, Llibre-Guerra J, et al. Unlocking cognitive analysis potential in Alzheimer's disease clinical trials: investigating hierarchical linear models for analyzing novel measurement burst design data. *Stati Med*. 2024