



LEADER 3D: A Natural History Study External Control as Confirmatory Evidence for a Rare Disease Drug Approval

Omaveloxolone (Skyclarys)

Before reading this case study, please refer to the [LEADER 3D Case Study User Guide](#) as an informational resource. FDA recognizes that certain aspects of drug development that are feasible for common diseases may not be feasible for rare diseases and that development challenges are often greater with increasing rarity of the disease. The small population affected by a rare disease presents additional considerations that must be addressed and also calls for appropriate flexibility. **Please note this case study is not intended or designed to provide specific strategies for obtaining product approval.** Rare disease drug development is not one-size-fits-all. The kind and quantity of data in each rare disease application will be different based on the unique considerations of each development program and must therefore be assessed on a case-by-case basis.

Learning Objectives

After reading this case study you will:

- Understand the key considerations for determining that the external control group is sufficiently comparable to the study population with regards to relevant prognostic factors, (e.g., disease severity, age of disease onset).
- Understand the key considerations for selecting an appropriate natural history study as an external control, including site overlap, standardized outcome measures, and availability of prognostic baseline variables for matching.
- Have a greater understanding of how propensity score matching can be used to analyze treatment and natural history external control group data in rare disease drug development.
- Understand how data from natural history studies can successfully support substantial evidence of effectiveness.
- Understand how the kind and quantity of confirmatory evidence is dependent on the results of the adequate and well-controlled trial, e.g., a less-persuasive adequate and well-controlled clinical investigation may require a greater quantity of compelling confirmatory evidence to allow for a conclusion of substantial evidence of effectiveness.



To access FDA guidance documents for rare disease drug development organized by topic, please scan this QR code.

Case Study Summary

This case study describes the use of data from a natural history study (NHS) as confirmatory evidence to support the substantial evidence of effectiveness (SEE) of omaveloxolone (Skyclarys) for the treatment of Friedreich's ataxia (FA). The Applicant submitted data from a single adequate and well-controlled (AWC) investigation; however, FDA determined that the trial results were not sufficiently robust to provide SEE due to limited support from secondary endpoints. As a result, FDA required additional data to establish SEE. To address this need, the Applicant submitted a propensity-matched analysis comparing data from a three-year, open-label extension (OLE) study with data from an FA NHS used as an external control. This additional support contributed to the confirmatory evidence, substantiating the efficacy finding.



**U.S. FOOD & DRUG
ADMINISTRATION**

<https://www.fda.gov/drugs/CDERARC>

FDA Guidance Corner

Note: The FDA Guidance Corner includes excerpts from draft FDA guidance documents, which, when final, represent the Agency's current thinking on topics in the case study. For up-to-date guidance documents please search: [Guidance Documents for Rare Disease Drug Development | FDA](#).

In this case study, the approval of omaveloxolone illustrates the use of natural history study data as an external control and component of the confirmatory evidence package in the setting of a serious rare and progressive disease with an unmet medical need. This guidance document outlines key considerations, including the selection of appropriate controls, to support efficient and successful rare disease drug development programs.

Guidance for Industry [Rare Diseases: Considerations for the Development of Drugs and Biological Products](#) (December 2023)

A critical element of an adequate and well-controlled study is the use of an appropriate control to enable reliable and unbiased, to the degree possible, efficacy assessments. Typically, use of a randomized concurrent control group (e.g., placebo, no treatment, active treatment) is recommended to distinguish changes occurring because of the drug from those changes occurring because of other factors, such as natural disease progression.

For serious rare diseases with unmet medical need, interest is frequently expressed in using an external control. In clinical investigations with external controls, outcomes in participants receiving the test treatment according to a protocol are compared with outcomes in a group of people external to the clinical investigation who had not received the same treatment. However, the lack of blinding and inability to eliminate systematic differences between treatment groups, given the nonrandomized nature of the comparison, are limitations to the use of an external control group. For example, in the case of a historical external control, there may be systematic differences between the nonconcurrent treatment groups attributable to changes in standard of care or diagnostic approaches over time.

Given the limitations, external control designs are usually reserved for specific circumstances, such as clinical investigations where the drug effect can be demonstrated in diseases with well-understood and characterized natural history, high and predictable mortality or progressive and predictable morbidity, and clinical investigations in which the drug effect is large and persuasive. The suitability of an externally controlled clinical investigation design warrants a case-by-case assessment, and early discussion with the relevant review division is recommended.

This guidance, which, when final will represent the Agency's current thinking, describes factors to consider when assessing whether a single adequate and well-controlled clinical investigation and confirmatory evidence are sufficient to demonstrate substantial evidence of effectiveness, and provides examples of types of data that could be considered confirmatory evidence.

Draft Guidance for Industry [Demonstrating Substantial Evidence of Effectiveness With One Adequate and Well-Controlled Clinical Investigation and Confirmatory Evidence](#) (September 2023)

When one adequate and well-controlled clinical investigation and confirmatory evidence are considered together to assess effectiveness, the quality and quantity of confirmatory evidence are also important considerations. Confirmatory evidence should be evidence generated from quality data derived from an appropriate source.

The quantity (e.g., number of sources) of confirmatory evidence necessary to support effectiveness may vary across development programs. Importantly, the quantity of confirmatory evidence needed in a development program will be impacted by the features of, and results from, the single adequate and well-controlled clinical investigation that the confirmatory evidence is intended to substantiate. It may be possible for a highly persuasive adequate and well-controlled clinical investigation to be supported by a lesser quantity of confirmatory evidence, whereas a less-persuasive adequate and well-controlled clinical investigation may require a greater quantity of compelling confirmatory evidence to allow for a conclusion of substantial evidence of effectiveness.

FDA Guidance Corner

In this case study, the Applicant used propensity score matching to create comparable patient populations by matching open-label extension study participants to natural history study participants in a 1:1 ratio. In addition, the Applicant utilized a mixed model repeat measures analysis method to evaluate the primary efficacy endpoint (mFARS change). This guidance document describes important considerations when analyzing data from externally controlled trials.

Draft Guidance for Industry [Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products](#) (February 2023)

FDA does not recommend a particular approach to analyzing data from externally controlled trials. No single statistical or analytical method will be suitable for all trials involving external control arms, and potential approaches should be discussed with the appropriate FDA review division. Sponsors should provide a justification for the analytic methods selected as well as a description of the strengths and limitations of the methods used to assess the effect of treatment. In general, the analytic method used should identify and manage sources of confounding and bias, including a strategy to account for differences in baseline factors and confounding variables between trial arms.

Various statistical methodologies may be appropriate for these types of comparisons, each with a corresponding level of complexity regarding approaches to account for bias. The assumptions involved should be made explicit, and sensitivity analyses as well as model diagnostics should be conducted to examine such assumptions. Importantly, however, adding complexity to an analytical framework usually requires making additional assumptions, which often cannot be substantiated and may impair the interpretability of results.

Various aspects of clinical care may change over time, such as the standard of care for the condition of interest, types of treatments, supportive care regimens, and criteria for determining disease response or progression. Such temporal differences are difficult to address using statistical analyses alone. It is important to consider whether and how different time frames in the treatment arm and the external control arm impact the interpretability of study findings.

Consideration should also be given, based on available scientific data, to the anticipated effect size for analyses of the primary endpoint. Especially when the anticipated effect size is modest, an externally controlled trial may not be an appropriate study design because of concerns for bias affecting the results. In addition, sponsors should develop a priori plans for assessing the impact of confounding factors and sources of bias, with quantitative or qualitative bias analyses used to evaluate these concerns. Such prespecified analyses can assist in the interpretation of study results.

Key Concept: Propensity Score Matching

Clinical investigations aim to determine whether a drug meaningfully improves a patient's condition by separating the drug's effect from other influences, such as the natural course of disease, placebo responses, or biased assessments. Randomized controlled trials (RCT) achieve this by creating treatment and control groups that are as similar as possible at baseline, allowing differences in outcomes to be attributed to the therapy being studied. However, in rare diseases, randomization may be impractical or ethically challenging, and FDA has long recognized that well-selected external controls, such as natural history cohorts, can sometimes provide an acceptable comparator for evaluating effectiveness. When using an external control, it becomes essential to ensure that the treated and comparison groups are closely aligned on factors known to influence outcomes.

Propensity score matching provides a structured way to achieve this alignment. A propensity score is the estimated probability that a patient would receive the treatment based on a set of observed baseline characteristics.¹ By matching patients with similar scores, researchers can balance many confounders simultaneously and create comparison groups that approximate the balance achieved through randomization.² When matching is successful:

1. Propensity scores are estimated using all relevant confounders available in both datasets.
2. Each treated patient is matched to an external control patient with a comparable score.
3. The resulting matched cohort exhibits balanced baseline characteristics, enabling analyses that resemble those of an RCT, although the balance is achieved statistically rather than through randomization.

¹ [Propensity score matching methodology: why and how it is used. Video Journal of Biomedicine. \(2022\)](#)

² [Kanamori S, Takeuchi M. Treatment effect estimation using the propensity score in clinical trials with historical control. BMC Med Res Methodol. \(2024\)](#)

Key Concept: Propensity Score Matching (continued)

This approach reduces bias that would otherwise arise from differences in disease severity, demographics, or other prognostic variables. Although propensity score methods cannot adjust for unmeasured confounding, they offer a transparent and practical strategy for constructing comparable groups in rare disease settings where sample sizes are small and traditional multivariable matching is not feasible.³

Case Study Relevance to Key Concept: In the omaveloxolone analysis, propensity scores were used to match participants in the Part 2 OLE study with comparable individuals from the Friedreich's ataxia-Clinical Outcome Measures (FA-COMS) Natural History Study using optimal 1:1 matching. This produced a balanced comparator group that reflected the clinical profile of the treated population, supporting a more reliable assessment of omaveloxolone's effect on FA disease progression while meeting methodological expectations for externally controlled studies. The propensity score matching was based on sex, baseline age, age of FA onset, baseline modified Friedreich's Ataxia Rating Scale (mFARS) score, and baseline gait score. The objective of this analysis was to approximate the randomization effect by creating groups that were equally distributed in terms of significant variables that could potentially impact the outcome of the efficacy analysis.

Introduction to the Rare Disease: Friedreich's Ataxia

Friedreich's ataxia (FA) is a rare, genetic, progressive neurodegenerative disorder caused by a trinucleotide guanine-adenine-adenine (GAA) repeat expansion in the first intron of the frataxin (*FXN*) gene. Approximately 1 in 30,000 to 1 in 50,000 individuals are affected by FA.⁴ The *FXN* gene encodes an essential mitochondrial protein of the same name. The repeat expansion impairs gene transcription, causing a deficiency of the frataxin protein which plays an important role in the formation of iron-sulfur clusters within mitochondria. Decreased frataxin results in mitochondrial dysfunction—including decreased ATP production and increased oxidative stress—as well as neuroinflammation and progressive neurodegeneration associated with neurologic manifestations, including ataxia. In addition to ataxia, patients with FA develop balance issues, loss of motor skills, and non-neurological features such as hypertrophic cardiomyopathy, kyphoscoliosis, and pes cavus, an abnormally high curvature of the foot. Symptom onset in FA is typically between 5 and 15 years of age, although age of onset can range from 2 years to >70 years. Progressive symptoms include loss of ambulation and wheelchair dependency by mid-20s. Individuals with FA have a shortened life expectancy with the average age of death occurring at approximately 37.5 years of age.

Key Terminology: Mixed Model Repeat Measures

Mixed model repeated measures (MMRM) is a statistical approach commonly used to analyze outcomes that are measured repeatedly over time in the same patients. Rather than relying on a single before-and-after comparison, MMRM incorporates all available longitudinal data to estimate how patients change over time. The method accounts for the natural correlation among repeated measurements within an individual and can provide reliable estimates without requiring imputation of missing values, as long as missing data are missing for reasons unrelated to the outcome being measured. MMRM provides accurate estimates of treatment differences at specific timepoints (typically the study endpoint) while using all available patient data and accounting for the natural correlation between repeated measurements from the same patient. Unlike simpler methods, MMRM does not require patients to complete all scheduled visits to be included in the analysis.

For more information on MMRM, please refer to [Analyzing Repeated Measurements Using Mixed Models](#) and [Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials](#).

Case Study Relevance to Key Terminology: In this case study, MMRM was used to evaluate the primary efficacy endpoint by comparing the trial participants with their propensity score matched FA-COMS counterparts. By leveraging the full set of repeated mFARS assessments, MMRM provided a robust estimate of differences in disease progression between the omaveloxolone-treated and external control groups.

 Prior to approval of omaveloxolone in 2023, there were no approved drugs for the treatment of FA.

³ [Kurz, C.F., Krzywinski, M. & Altman, N. Propensity score matching. Nat Methods 21, 1770–1772 \(2024\).](#)

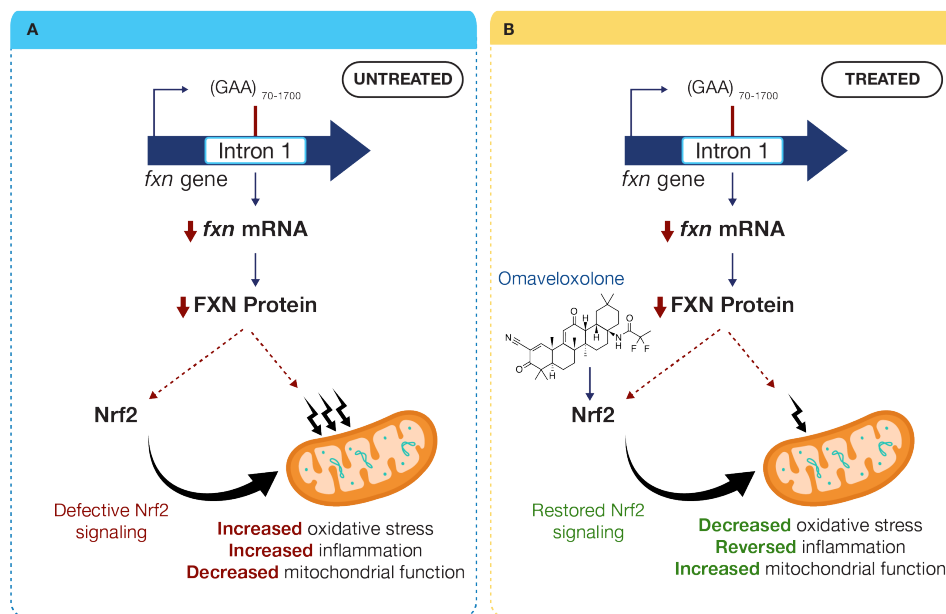
⁴ [Friedreich Ataxia - StatPearls - NCBI Bookshelf](#)

Omaveloxolone Mechanism of Action

It is postulated that impairment of nuclear factor erythroid 2-related factor 2 (Nrf2) signaling is associated with frataxin deficiency and mitochondrial dysfunction and underlies the pathogenesis and progression of FA. Omaveloxolone is a semisynthetic oleanane triterpenoid with antioxidant and anti-inflammatory properties. Although the precise mechanism by which omaveloxolone exerts its therapeutic effect in patients with FA is unknown, it is thought to include activation of the transcription factor Nrf2, which regulates the cellular response to oxidative stress and improves mitochondrial bioenergetics, thereby restoring redox balance and reducing inflammation (**Figure 1**).

Figure 1: Mechanism of Action of Omaveloxolone: FA is caused by a trinucleotide guanine-adenine-adenine (GAA) repeat expansion in the FXN gene, which impairs transcription and leads to a deficiency of the frataxin protein. This deficiency causes impaired iron-sulfur cluster biogenesis, resulting in mitochondrial dysfunction, increased oxidative stress, and chronic inflammation, which drive the pathogenesis and progression of FA. Furthermore, frataxin deficiency is associated with impaired Nrf2 signaling, thus decreasing the cell's natural ability to manage oxidative stress.

Omaveloxolone, a synthetic triterpenoid derivative of oleanolic acid, acts as an activator of the Nrf2 pathway and promotes its accumulation and translocation to the nucleus. This restored Nrf2 signaling increases the expression of antioxidant genes, reduces inflammation, and improves mitochondrial bioenergetics. By addressing these underlying cellular deficits, omaveloxolone reduces oxidative stress and improves neurological function in patients with FA.



The Single Adequate and Well-Controlled Investigation

Study Design

The effectiveness of omaveloxolone for the treatment of FA was evaluated in the single, adequate, well-controlled Study 1402 (MOXIe), followed by an OLE study with up to three years of omaveloxolone exposure data. Study 1402 was a two-part study evaluating the pharmacokinetics (PK), pharmacodynamics (PD), safety, and efficacy of omaveloxolone in the treatment of patients with FA.

Study 1402 Part 1: Part 1 of the 1402 study was a randomized, double-blind, placebo-controlled, dose-finding study that evaluated the safety, efficacy, PK and PD of omaveloxolone at different doses.

Study 1402 Part 2: Part 2 of the 1402 study served as the AWC trial in support of the effectiveness of omaveloxolone in FA patients (**Figure 2**). It was a randomized, double-blind, placebo-controlled, parallel-group study that evaluated 150 mg of omaveloxolone given once daily in FA patients for 48 weeks. The dose of 150 mg for Part 2 was selected based

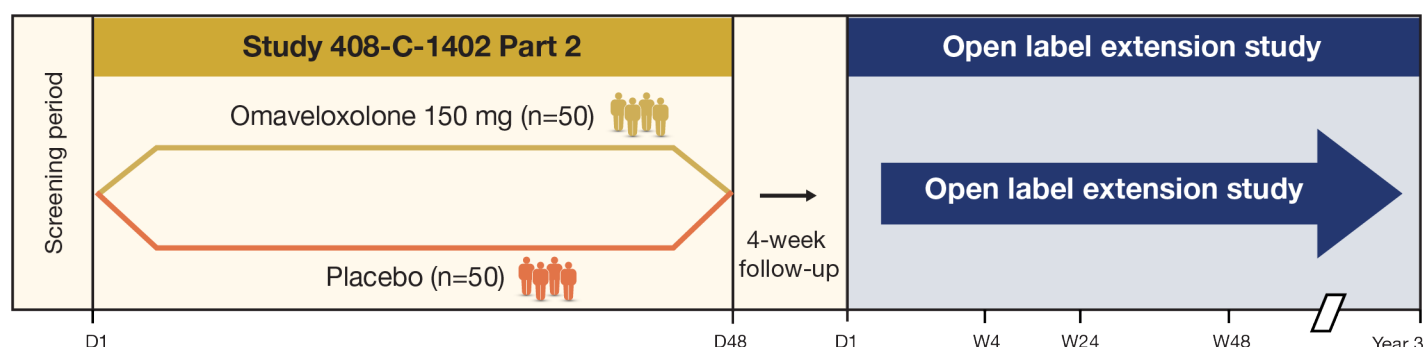
on the dose-finding data from Part 1. Patients enrolled in Part 2 were randomized 1:1 to receive omaveloxolone or placebo. Randomization was stratified by the presence of pes cavus (i.e., pes cavus compared to no pes cavus). Presence of pes cavus (musculoskeletal foot deformity characterized by high arch of the foot that does not flatten with weight-bearing) can affect patients' ability to use their legs, walk, and perform neurologic and exercise testing independent of their ataxia. Patients with pes cavus comprised only 20% of the study population. Therefore, the Applicant amended the protocol to limit enrollment of pes cavus patients in the study after a post hoc analysis demonstrated a greater treatment effect for patients without pes cavus.

Study 1402 OLE: The Open-Label Extension Study

The purpose of the Study 1402 OLE was to assess long-term safety and tolerability of omaveloxolone in FA participants and provide confirmatory evidence to support the effectiveness of the drug. Study 1402 OLE allowed all participants from Parts 1 and 2 of Study 1402 to enroll. Upon entering the OLE study, all participants remained blinded to their previous treatment in Part 1 and 2.

The study dosing regimen for the OLE was the same as in Part 2 (150 mg) and all participants had in-person assessments during treatment in the clinic at Day 1, Week 4, Week 24, and every 24 weeks thereafter for three years ([Figure 2](#)).

Figure 2: Schematic for Part 2 of Study 1402 and open-label extension study; (Adapted from Figure 1 of NDA-216718 and Information in the Integrated Review for omaveloxolone [Skyclarys])



Efficacy Endpoints

Study 1402 Part 2, the AWC trial, included the following primary and secondary endpoints:

Primary Endpoint:

- **The change in the mFARS scores from Baseline to Week 48.** mFARS is a clinically validated neurological assessment comprised of four components that focus on functional abilities. mFARS removed section D of the FARS, which contains items that are less clinically meaningful, and consists of FARS components A, B, C, and E ([Table 1](#)). A higher score on the mFARS examination signifies more severe physical impairment, and a reduction in mFARS score signifies improvement in functioning.

Secondary Endpoints:

- **The proportion of patients who are a treatment success at Week 48, based on the Patient Global Impression of Change (PGIC).** A treatment success is defined as "Much Improved" or "Very Much Improved" on the PGIC. PGIC is a 7-point scale that requires the patient to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention.
- **The proportion of patients who are a treatment success at Week 48, based on the Clinical Global Impression of Change (CGIC).** A treatment success is defined as "Much Improved" or "Very Much Improved" on the CGIC. CGIC is a 7-point scale that requires the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention.

Table 1: Modified Friedreich's Ataxia Rating Scale (mFARS)⁵

mFARS measures	Assessed Function	Assessments and Scoring	Total Points	Clinical Extrapolations of Possible Effects on Patient Abilities
FARS A Bulbar function	Speech clarity and strength and volume of coughing	Cough (2) Speech (3)	5	Affects the ability to communicate clearly. Patients may also be at increased risk for respiratory infection.
FARS B Upper limb coordination	Fine motor coordination	Finger-finger (6) Nose-finger (8) Dysmetria (8) Rapid movements (6) Finger taps (8)	36	Ability to complete activities of daily living, such as getting dressed, eating, brushing teeth, typing, pointing, reaching, and turning a doorknob.
FARS C Lower limb coordination	Coordination of legs and feet	Heel-along-shin slide (8) Heel-to-shin tap (8)	16	Closely related to upright stability, also an important contributor to decline in ambulatory patients. Affects activities like putting on socks and shoes.
FARS E Upright Stability	Sitting, standing, and walking	Sitting Position (4) Stance feet apart (4) Stance feet apart (eye-closed) (4) Stance feet together (4) Stance feet together (eye-closed) (4) Tandem stance (4) Stance dominant foot (4) Tandem walk (3) Gait (5)	36	Assessment of individual ambulatory ability. Affects activities like walking, sitting in a car, standing in a line, and showering.

The Division was concerned that the mFARS did not directly measure clinical benefit and therefore encouraged the sponsor to include straightforward patient-reported outcomes (PRO) or performance-based outcomes (PerfO) for this purpose. The Applicant proposed PGIC and CGIC as key secondary endpoints, which was considered acceptable to provide clinical meaningfulness of the results on mFARS.

Results: Study 1402 Part 2, the AWC Trial

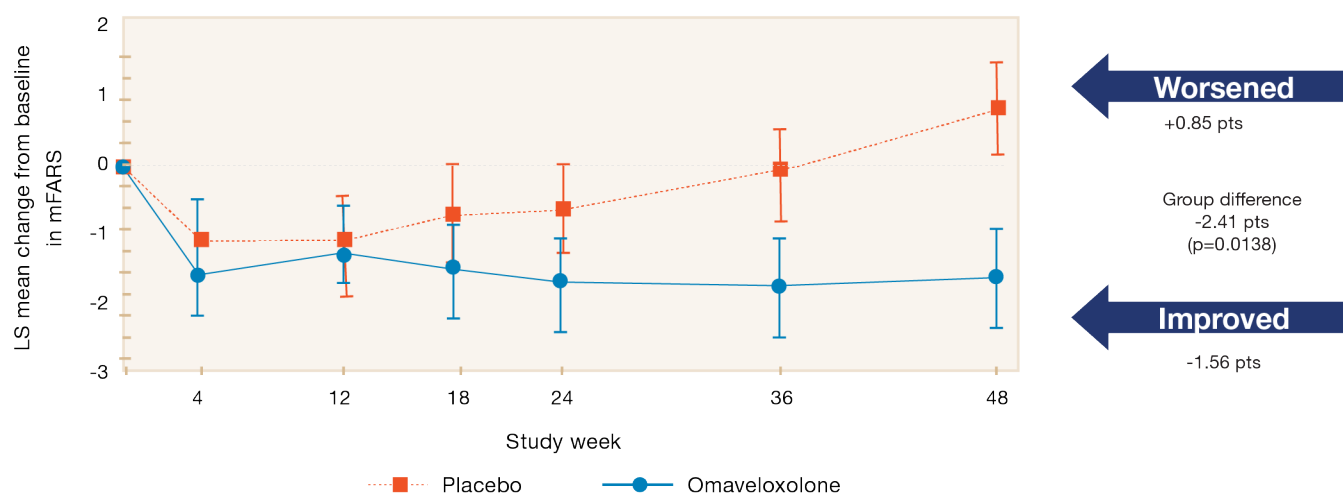
Primary analysis of efficacy was based on the Full Analysis Set (FAS, N = 82), which consisted of participants without pes cavus while the All-Randomized Population (ARP, N = 103) included both participants with pes cavus (20% of subjects) and without pes cavus.

Efficacy results for the primary endpoint (mFARS change) in the FAS population are shown in [Figure 3](#). In the FAS, patients randomized to omaveloxolone demonstrated a mean improvement from baseline in mFARS of -1.56 points, while patients randomized to placebo experienced a mean worsening from baseline in mFARS of 0.85 points at Week 48. This resulted in a statistically significant treatment difference of -2.41 points in favor of omaveloxolone, (p=0.0138). In the ARP, a nominally significant improvement in mFARS at Week 48 compared to placebo was also observed (-1.94 points, p=0.0331).

⁵ <https://www.christopherreeve.org/wp-content/uploads/2024/04/Instructions-for-administering-the-mFARS.pdf>

NHS External Control as Confirmatory Evidence

Figure 3: mFARS Changes from Baseline by Visit (Full Analysis Set)



The key secondary endpoints, PGIC and CGIC, favored the omaveloxolone arm in the FAS but were not statistically different from the placebo. PGIC was nominally significant in the ARP, but CGIC remained non-significant in this population (Table 2). The treatment group did not show a statistically significant improvement in the other secondary endpoints such as 25-foot timed walk test, frequency of falls, or peak work relative to placebo, although frequency of falls favored the treatment group.

Because the secondary endpoints, which were considered to provide clinical meaningfulness of the mFARS, did not reach statistical significance in the FAS or ARP for CGIC, and only reached nominal significance in the ARP for PGIC, FDA required additional data from the Applicant to establish substantial evidence of effectiveness.⁶

Table 2: Patient Global Impression of Change (PGIC) and Clinical Global Impression of Change (CGIC) Responses at Week 48

	Full Analysis Set (FAS)		All Randomized Population (ARP)	
	Placebo (n=42)	Omaveloxolone (n=40)	Placebo (n=52)	Omaveloxolone (n=51)
PGIC				
LS mean	4.32	3.89	4.47	3.91
LS mean difference between treatment	-	-0.43 (p=0.1300)	-	-0.56 (p=0.0282)
CGIC				
LS mean	4.06	3.92	4.18	3.90
LS mean difference between treatment	-	-0.13 (p=0.5259)	-	-0.28 (p=0.1328)

⁶ For more information on the confirmatory evidence of effectiveness of omaveloxolone, please refer to page 100 of the [Integrated Review for omaveloxolone \(Skyclarys\)](#), NDA 216718

The Confirmatory Evidence Package

To address the need to provide additional data to support the effectiveness of omaveloxolone, the Applicant conducted post hoc propensity-matched analyses comparing long-term mFARS data from the Study 1402 OLE to mFARS data from the FA Clinical Outcome Measures Study (FA-COMS).

Highlights

Factors Facilitating Successful Use of FA-COMS Data as Confirmatory Evidence

- Systematic collection of baseline characteristics, which included diagnostic criteria and prognostic indicators for disease progression.
- Inclusion of a large number of participating patients.
- Long duration of follow-up.
- Absence of confounding factors such as off-label treatments.
- Standardized case report forms across all treatment centers.
- Shared overlapping treatment centers and geographic regions with the clinical investigations.
- Standardized assessment metrics identical to the ones of the clinical trial.

Natural History Study (FA-COMS) as an External Control for Study 1402 OLE

FA-COMS (NCT03090789⁷) is a multicenter NHS that started in 2001 with an estimated enrollment goal of 2,000 FA patients worldwide, to be assessed annually for up to 15 years. It was considered an appropriate external comparator because it was a large, multicenter, longitudinal cohort in which mFARS was prospectively and systematically assessed using standardized methods. Substantial overlap in participating sites with Study 1402 OLE, collection of key prognostic baseline variables enabling propensity score matching, and the absence of approved disease-modifying therapies in FA support the clinical relevance and comparability of the matched populations for evaluating deviation from the expected natural history trajectory. The FA-COMS study aimed to expand the network of clinical research centers in FA and to provide a framework for facilitating therapeutic interventions. The FA-COMS has led to the development of clinical outcome measures for FA, including the mFARS.

FDA Guidance Corner

In this case study, the Applicant used natural history data as an external control and compared it with data from an open-label long-term extension study to generate confirmatory evidence supporting the efficacy findings from the single adequate and well-controlled trial. This draft guidance document describes the broad potential uses of a natural history study in all phases of drug development for rare diseases, as well as important considerations for designing an externally controlled trial using natural history study data as an external control.

Draft Guidance for Industry [Rare Diseases: Natural History Studies for Drug Development](#) (March 2019)

To demonstrate effectiveness, sponsors must provide substantial evidence from adequate and well-controlled investigations, including clinical investigations, that include (among other factors) a valid comparison to a control. The sponsor uses data collected from an adequate control group to discriminate patient outcomes caused by the investigational drug from outcomes caused by other factors (i.e., what would have happened if similar patients had not received the investigational drug).

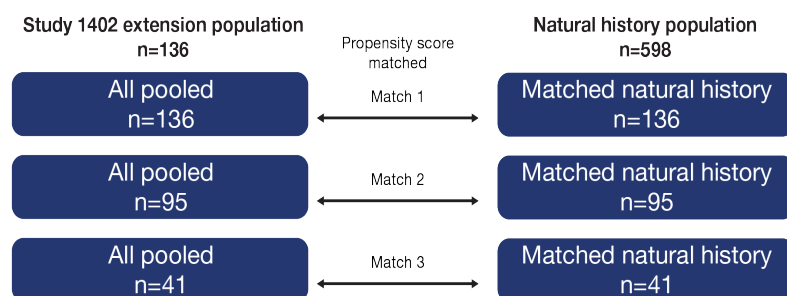
FDA regulations recognize historical controls as a possible control group (usually reserved for special circumstances); however, inability to control for certain biases could limit the ability of externally controlled trials to demonstrate substantial evidence of effectiveness. However, bias may be mitigated in certain situations where the disease course is predictable and the treatment effect dramatic.

⁷ <https://clinicaltrials.gov/study/NCT03090789>

To obtain further supportive evidence of effectiveness of omaveloxolone, the Applicant conducted a post hoc propensity matched analysis on the mFARS data from the Study 1402 OLE study compared to the mFARS data from the FA-COMS natural history study using a three-year cutoff. Three separate populations of pooled participants from the Study 1402 OLE were analyzed (**Figure 4**):

1. Patients who had been on omaveloxolone or placebo in Study 1402 Part 2 (pooled analysis),
2. Patients who had been on placebo in Part 2 (placebo-omav); and
3. Patients who had been on omaveloxolone in Part 2 (omav-omav).

Figure 4: Primary Analysis Populations for Propensity Matched Analysis



The FA-COMS study enrolled more than 1,000 patients, and mFARS was assessed annually for up to 19 years in some participants. In Study 1402 OLE, outcomes were assessed every 24 weeks. In addition, the centers participating in the FA-COMS study overlapped with the centers in the Study 1402 OLE. A total of eight sites, accounting for 84% of participants enrolled in Study 1402 OLE, were also active sites for the FA-COMS NHS. This overlap implied a likely consistent standard of care as well as consistency in clinical assessments between the NHS and Part 2 OLE study.

The primary efficacy endpoint was determined utilizing a mixed model repeat measures (MMRM) analysis that compared Study 1402 OLE participants to propensity score matched FA-COMS participants by measuring the change in mFARS score at Year 3 compared to baseline. The post hoc propensity matched analysis of data from Study 1402 OLE for three years compared to the natural history data demonstrated a statistically significant difference in mFARS change from baseline in favor of the Study 1402 OLE for the primary pooled analysis ($p=0.0002$), the primary placebo-omav population analysis ($p=0.0001$), and the primary omav-omav population analysis ($p=0.0190$) (**Table 3**).

FDA Guidance Corner (continued)

The use of external controls requires careful planning and assessment, including the following considerations:

- The external control group needs to be very similar to the treated group in all respects, including disease severity, duration of illness, prior treatments, and any other aspects of the disease that could affect outcomes and the timing of outcomes.
- Selection bias is a major concern when using external controls because there is no randomization and unrecognized baseline differences can affect outcomes. Points to consider include the following:
 - Critical patient disease characteristics may not have been assessed or may have been assessed differently based on historical approaches, resulting in a lack of comparability (e.g., disease definitions, diagnostic techniques, and approaches to safety monitoring may have evolved).
 - Aspects of standard of care may have changed.
 - Data collection intervals and quality may lack consistency and not be comparable.
- Use of an external control group is especially challenging if the outcome assessments used in the external control group are not well defined and unreliable and, therefore, not suitable for regulatory use.

NHS External Control as Confirmatory Evidence

Table 3: Change in mFARS Score at Year 3 for Primary Populations (Matches 1-3 populations).

mFARS change from Baseline at Year 3			
	Pooled Population (Match 1)	Primary Placebo-Omav Population (Match 2)	Primary Omav-Omav Population (Match 3)
	LS Mean (±SE)	LS Mean (±SE)	LS Mean (±SE)
Study 1402 Extension	2.983 (0.6433)	3.238 (0.7604)	2.208 (1.3122)
Matched FA-COMS	6.337 (0.6183)	7.334 (0.7212)	6.448 (1.2136)
Difference	-3.354 (0.8923) p=0.0002	-4.096 (1.0480) p=0.0001	-4.240 (1.7888) p=0.0190

Despite the general limitations associated with NHS data as an external control, such as patient selection biases for comparison and missing information, FDA noted that the consistency of the treatment effect (i.e., similar treatment effect across each treatment group) in the mFARS between the 1402 OLE and FA-COMS study data was a major point of consideration. Every omaveloxolone treatment group in Study 1402 OLE appeared to show slower worsening in mFARS for each of the three years compared to propensity score matched individuals in the FA-COMS NHS.⁸ Not only did the data confirm that the patients with FA treated with omaveloxolone deviated from the natural history, the natural history data also confirmed that the placebo arm in Study 1402 performed similarly over 48 weeks to what would have been expected in the absence of treatment.

Pharmacodynamic Data as Confirmatory Evidence

The Applicant also included mechanistic data from pharmacodynamic (PD) biomarkers that provided additional support for the omaveloxolone treatment effects observed in Study 1402. The Applicant noted ferritin light chain (FTL), ferritin heavy chain (FTH1), and γ -glutamyl transferase (GGT1) are targets of Nrf2. Nrf2 activation regulates the iron metabolism pathway, resulting in increased expression of iron-metabolizing genes such as FTL and FTH1, which combine to form ferritin. Data from Part 1 of Study 1402 showed a dose-dependent increase in ferritin levels in participants during omaveloxolone treatment. FDA agreed that the Applicant's in vitro data supported the conclusion from animal models that omaveloxolone is an Nrf2 activator. Furthermore, the mechanistic increase in ferritin can be considered a PD effect of omaveloxolone suggestive of Nrf2 activation, providing additional confirmatory evidence of omaveloxolone's effectiveness.

Conclusion

The substantial evidence of effectiveness of omaveloxolone in patients with FA was provided by a single AWC trial with positive results on a clinically meaningful primary endpoint, as well as confirmatory evidence from a natural history used as an external control and pharmacodynamic data supporting the biological plausibility of the treatment effect.

This case study demonstrates that natural history study data can serve as valuable confirmatory evidence of efficacy when methodological challenges are appropriately addressed. The Applicant successfully mitigated common limitations of external natural history comparisons through rigorous propensity score matching to minimize selection bias. Additionally, the magnitude of difference in mFARS progression between the natural history cohort and Study 1402 population was similar for each matched group and provided consistent support for omaveloxolone's effectiveness in patients with FA.

⁸ Patel M et al. (2016) Progression of Friedreich ataxia: quantitative characterization over 5 years. *Ann Clin Transl Neurol* 3(9):684–94

Key Takeaways

- Generally, the type and quantity of confirmatory evidence is dependent on the persuasiveness of the single adequate and well-controlled clinical investigation. In this case, the lack of support from the secondary endpoints necessitated a stronger confirmatory evidence package.
- Because the secondary endpoints, which were considered to provide clinical meaningfulness of the mFARS, did not reach statistical significance in the FAS or ARP for CGIC, and only reached nominal significance in the ARP for PGIC, FDA required additional data from the Applicant to establish substantial evidence of effectiveness. To meet this need, additional data from the NHS contributed to the confirmatory evidence package through post hoc propensity score-matched analyses comparing the long-term OLE study (Study 1402 OLE) to the FA-COMS and provided further evidence to establish SEE.
- Multiple factors enabled successful use of the FA-COMS as an NHS external control, including systematic baseline data collection, large participant numbers, long follow-up duration, absence of confounding factors, standardized case report forms and assessment metrics across treatment centers, and overlapping treatment centers between the clinical trial and NHS.
- Propensity score matching identified comparable patient populations by matching Study 1402 OLE participants 1:1 with FA-COMS natural history participants based on sex, baseline age, age of FA onset, baseline mFARS score, and baseline gait score, with all three matched populations showing statistically significant improvements in mFARS scores at Year 3 favoring omaveloxolone treatment.
- MMRM was used to evaluate the primary efficacy endpoint (mFARS change from baseline to Year 3) by comparing Study 1402 OLE participants with their propensity score matched FA-COMS counterparts, providing a robust estimate of differences in disease progression between the omaveloxolone-treated and external control groups.
- Pharmacodynamic biomarker data contributed to the confirmatory evidence used to support the biological plausibility of omaveloxolone's mechanism of action.



Critical Thinking Questions for a Rare Disease Drug Development Program

Rare disease drug developers should engage FDA early to discuss clinical study design and the statistical analysis plan. When preparing for and designing a clinical investigation for a rare disease medical product, consider the following questions:

- What is the development plan to demonstrate substantial evidence of effectiveness?
 - If conducting two adequate and well-controlled clinical investigations is not feasible, what is the scientific justification for the proposed development approach?
- If planning to design a single adequate and well controlled trial for demonstrating substantial evidence of effectiveness, what characteristics of the disease, endpoint, and treatment effect would strengthen or weaken this approach (Consider the clinical meaningfulness and robustness of the primary and secondary endpoints and the predicted treatment effect size)?
- What data can contribute to the confirmatory evidence package? Please note that the quantity of confirmatory evidence may be impacted by the design and results of the one adequate and well-controlled clinical investigation.
 - Can mechanistic and pharmacodynamic biomarkers play a role in the confirmatory evidence package?
 - Is there a plan to evaluate how biomarker data can support biological plausibility, dose selection, and interpretation of clinical findings?
- When planning to use natural history study data as an external control, consider what may be key limitations and biases in the natural history comparator, and how these limitations will be mitigated.
 - Have the following been addressed between the treatment and external control groups?
 - sample size
 - selection bias
 - missing data
 - temporal changes in standard of care
 - data collection rigor differences
 - follow-up duration
 - standardization of assessments and data collection methods
 - overlap with trial sites
 - other confounding factors
- What statistical methods are being considered in the analysis plan? Have you consulted with the FDA review division on the appropriateness of the analysis plan?

We recommend meeting with the Agency early⁹ in the drug development program to reach alignment regarding endpoint selection, trial design, and the approach to demonstrate substantial evidence of effectiveness.

⁹ Please refer to the Additional Resources section for guidance on how to engage early with FDA, including information on Critical Path Innovation Meetings, INTERACT Meetings, and Pre-IND Meetings

Additional Resources

1. Tai G, Corben LA, Woodcock IR, Yiu EM, Delatycki MB. 2021. Determining the Validity of Conducting Rating Scales in Friedreich Ataxia through Video. *Movement Disorders Clinical Practice*. 8(5):688–693. doi:<https://doi.org/10.1002/mdc3.13204>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8287168/#:~:text=The%20SARA%20is%20used%20as,s11095-020-02882-0>
2. Wellhagen GJ, Hamrén B, Kjellsson MC, Åstrand M. 2020. Dose-Response Mixed Models for Repeated Measures – a New Method for Assessment of Dose-Response. *Pharmaceutical Research*. 37(8). doi:<https://doi.org/10.1007/s11095-020-02882-0>
3. Pane C, Salzano A, Trinchillo A, Del Prete C, Casali C, Marcotulli C, Defazio G, Guardasole V, Vastarella R, Giallauria F, et al. 2020. Safety and Feasibility of Upper Limb Cardiopulmonary Exercise Test in Friedreich Ataxia. *European Journal of Preventive Cardiology*. 29(3). doi:<https://doi.org/10.1093/eurjpc/zwaa134>.
4. Patel M, Isaacs CJ, Seyer L, Brigatti K, Gelbard S, Strawser C, Foerster D, Shinnick J, Schadt K, Yiu EM, et al. 2016. Progression of Friedreich ataxia: Quantitative Characterization over 5 Years. *Annals of Clinical and Translational Neurology*. 3(9):684–694. doi:<https://doi.org/10.1002/acn3.332>.
5. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. *JAMA*. 2016;315(4):407–408. doi:10.1001/jama.2015.19394: <https://doi.org/10.1001/jama.2015.19394>
6. Mallinckrodt, C.H., Lane, P.W., Schnell, D. et al. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Ther Innov Regul Sci* 42, 303–319 (2008). <https://doi.org/10.1177/009286150804200402>
7. Guidance for Industry: Critical Path Innovation Meetings. <https://www.fda.gov/media/89497/download>
8. Draft Guidance for Industry: Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products. <https://www.fda.gov/media/172311/download>
9. Draft Guidance for Industry: Early Drug Development and the Role of Pre-IND Meetings: <https://www.fda.gov/media/117322/download>

Case Study References by Order of Appearance

Page 1

- See the LEADER 3D Case Study User Guide available at <https://www.fda.gov/media/185425/download>
- See the FDA webpage with a list of guidance documents for Rare Disease Drug Development available at <https://www.fda.gov/drugs/guidances-drugs/guidance-documents-rare-disease-drug-development>.

Page 2

- See the FDA webpage with a list of guidance documents for Rare Disease Drug Development available at <https://www.fda.gov/drugs/guidances-drugs/guidance-documents-rare-disease-drug-development>.
- See guidance for industry Rare Diseases: *Considerations for the Development of Drugs and Biological Products* (December 2023), available at <https://www.fda.gov/media/119757/download>
- See the draft guidance for industry: *Demonstrating Substantial Evidence of Effectiveness With One Adequate and Well-Controlled Clinical Investigation and Confirmatory Evidence* (September 2023), available at <https://www.fda.gov/media/172166/download>

Page 3

- See the draft guidance for industry: *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products* (February 2023), available at <https://www.fda.gov/media/164960/download>
- See Propensity score matching methodology: why and how it is used. Video Journal of Biomedicine (2022), available at <https://www.tandfonline.com/doi/abs/10.2217/vjbm-2022-0014>
- See Kanamori S, Takeuchi M. Treatment effect estimation using the propensity score in clinical trials with historical control. BMC Med Res Methodol. (2024), available at <https://pubmed.ncbi.nlm.nih.gov/38389058/>

Page 4

- See Kurz CF, Krzywinski M, Altman N. Propensity score matching. Nat Methods. (2024), available at <https://pubmed.ncbi.nlm.nih.gov/39284960/>
- See Friedreich Ataxia – StatPearls, available at <https://www.ncbi.nlm.nih.gov/books/NBK563199/>
- See Detry, MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models, JAMA. (2016), available at <https://pubmed.ncbi.nlm.nih.gov/26813213/>

Page 7

- See instructions for administering the mFARS, available at <https://www.christopherreeve.org/wp-content/uploads/2024/04/Instructions-for-administering-the-mFARS.pdf>

Page 8

- See page 100 of the Integrated Review document for Omaveloxolone (Skyclarys), available at https://www.accessdata.fda.gov/drugsatfda_docs/nda/2023/216718Orig1s000MedR.pdf

Page 9

- See study FA-COMS (NCT03090789), available at <https://clinicaltrials.gov/study/NCT03090789>
- See the draft guidance for industry: *Rare Diseases: Natural History Studies for Drug Development* (March 2019), available at <https://www.fda.gov/media/122425/download>

Page 11

- See Patel M et al. Progression of Friedreich ataxia: quantitative characterization over 5 years. Ann Clin Transl Neurol. (2016), available at <https://pubmed.ncbi.nlm.nih.gov/27648458/>