Research OMICS International

Design Considerations for Clinical Trials Using the Cutaneous Dermatomyositis Disease Area and Severity Index as a Primary Endpoint

Julie A. Brevard[1*], Mark Hurtt[1], Joanna Horobin[1], Olga Polyanskaya[1], James Baker[1], Victoria P. Werth[2,3] and David F. Fiorentino[4]

[1]Idera Pharmaceuticals, Inc., Cambridge, MA, U.S.A

[2]Corporate Michael J. Crescenz (Philadelphia) Veterans Affairs Medical Center, Philadelphia, PA, U.S.A

[3]Department of Dermatology, University of Pennsylvania School of Medicine, Philadelphia, PA, U.S.A

[4]Department of Dermatology, Stanford University School of Medicine, Stanford, CA, U.S.A

*Corresponding author: Julie A. Brevard, Idera Pharmaceuticals, Inc., Cambridge, MA, U.S.A, Tel: 6176795567; E-mail: jbrevard@iderapharma.com

Abstract

Objective: Skin inflammation is a universal characteristic of the multisystem inflammatory myopathy dermatomyositis (DM). About 35% of DM patients have a polycyclic or chronic disease course, usually due to active skin disease, despite standard of care therapy, demonstrating a significant unmet need for novel treatment options. To test the efficacy of novel therapeutics, feasible prospective clinical trials must be designed that employ quantitative endpoints. The Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI) is a validated, reliable, and responsive measurement of cutaneous DM disease activity and damage. The primary objectives were to identify requisite design parameters (e.g., target trial population, duration, frequency of assessments, treatment difference, sample size) for a proofofconcept clinical trial of a novel therapy in patients with DM.

Methods: Trial simulations were conducted to determine sample size and statistical power. Simulation inputs were obtained from analyses of CDASI natural history data obtained from 115 adult DM patients followed at Stanford University.

Results: A population with baseline CDASI Activity scores ≥ 15 was simulated to eliminate a potential floor effect in patients with milder disease and to mimic a clinical trial population with moderate to severe skin disease. Given a 24 weeks trial with monthly disease assessments, simulations demonstrated that 45 patients per group were required to detect a 5 point treatment difference and 12 patients per group were required to detect a 10 point treatment difference with 80% statistical power using a 2 sided test.

Conclusion: This study provides evidence that the CDASI is a practical and feasible primary efficacy endpoint for DM clinical trials and lays out a framework for a clinical trial using CDASI as a primary endpoint.

Introduction

Dermatomyositis (DM) is a rare idiopathic inflammatory myopathy that impacts the skin, muscle, joints, heart, and lungs [1-3]. Approximately 35% of patients have a polycyclic or chronic continuous disease course, despite current standards of care, demonstrating a significant residual unmet need [4].

To promote the development of novel DM therapies, it is necessary to define feasible clinical development pathways with measurement methods capable of detecting clinically meaningful changes in disease activity. Because cutaneous manifestations are among the most severe disease components and have a large impact on quality of life, outcome measures evaluating changes in cutaneous disease activity may be attractive for use in interventional clinical trials [5,6].

The Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI) was developed as an outcome measure of skin disease for use in trials and clinical practice and includes assessments of disease activity and damage. The CDASI was recently developed and validated, and thus no randomized, controlled therapeutic clinical trials using the CDASI have been reported to date. Accordingly, there are limited data in the public domain to inform trial design.

The primary objective of this work is to offer a roadmap to investigators wishing to design a clinical trial in dermatomyositis. For the first time, we report data on the CDASI that can be used for trial design, including power and sample size calculations. In the current study we conducted computer simulations of a clinical trial of a hypothetical treatment to inform the design of a real-world clinical trial of a novel DM therapy. Trial simulations are often employed to assist trial design and guide sample size selection in situations more complex than standard designs, such as when repeated longitudinal assessments or regression-based analytical techniques are used. They are intended to mimic realworld situations and enable interrogation of key variables and assumptions. Here, we simulated a placebo-controlled clinical trial using the CDASI as the primary efficacy endpoint, with the objective of calculating the sample size required to detect a clinically relevant difference between the treatment and placebo arms.

To simulate data that mimicked expectations in a real trial, key variables and assumptions were required. We utilized CDASI natural history data from the Stanford University outpatient dermatology clinic, first characterized by Anyanwu et al. to provide this information for our simulations [7].

Methods

Simulation plan

Computer simulations of a clinical trial of a hypothetical treatment were performed to generate datasets that mimicked real-world trial results. Information was obtained from the Stanford dataset to use as inputs to the simulations, including the baseline CDASI score mean and Standard Deviation (SD), the amount of change expected over time, and how strongly CDASI scores were correlated between baseline and the end of study. Additional assumptions about the clinical trial were required for the simulation including the target population, treatment duration, and frequency of assessments. After simulating data, the data were analysed to determine the requisite sample size for a specified treatment difference at a specified power level.

Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI)

The CDASI is an outcome measure designed to assess cutaneous DM disease activity and damage. Disease involvement is rated in 15 different anatomical locations using multiple activity and damage measures. Activity and Damage Subscale scores range from 0 to 100 and 0 to 32, respectively, where higher scores indicate greater disease severity.

The CDASI activity subscale (CDASI-a) was evaluated in this study since disease activity may be reversible with treatment whereas damage from DM is not expected to be directly modified by treatment.

In multiple studies, the CDASI-a has demonstrated excellent inter- and intrarater reliability and responsiveness to change on the Physician Global Assessment, the current gold standard measure of skin-related disease activity [7,8]. These studies have shown that patients with CDASIa scores of 14 to 19 are considered to have moderatetosevere skin disease. In addition, a 4 to 5 point reduction in CDASI-a score was shown to represent a minimal clinically important difference (MCID) in individual patients. The MCID was based on a responsiveness analysis using the definition of response (≥ 2 cm improvement) on the Physician Global Activity Assessment visual analogue scale [7].

CDASI natural history dataset

The Stanford University dataset included CDASI scores recorded from 115 adults with clinical and histological evidence of DM. Data were collected longitudinally during routine clinical management between May 2007 and November 2012. Patients were allowed to rotate on and off standard topical and systemic therapies as deemed clinically appropriate. Data were collected in accordance with the Declaration of Helsinki principles and with the approval of the Stanford University Institutional Review Board.

CDASI natural history analyses

The CDASI natural history dataset was analysed to identify inputs for the trial simulations. First, the overall population was evaluated to assess CDASI-a scores at baseline. The CDASIa scores were plotted longitudinally over a period of 200 days and grouped by baseline CDASI-a score (0 to 14, 15 to 24, and ≥ 25) (Figure 1). Patients with baseline CDASI-a scores <15 or ≥ 15 and followup assessments at 12 or 24 weeks (± 28 days) were assessed to evaluate change at these time points. The correlation between CDASI assessment at baseline with CDASI at Week 12, and Week 24 was calculated for the subset of these subjects that had assessments at all three time points. The first assessment for a given patient was defined as the baseline score, and thus may represent data from patients already undergoing medical therapy. In addition, if a patient had multiple visits occurring in a single visit window, the last visit was analysed.
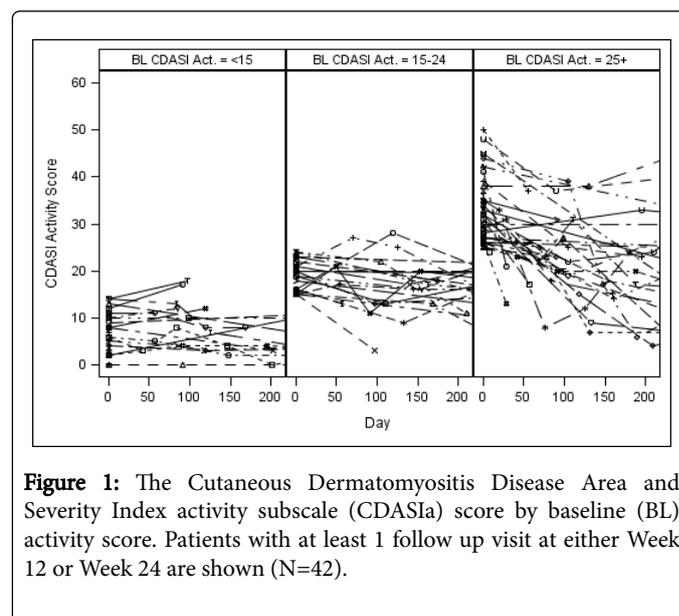

Figure 1: The Cutaneous Dermatomyositis Disease Area and Severity Index activity subscale (CDASIa) score by baseline (BL) activity score. Patients with at least 1 follow up visit at either Week 12 or Week 24 are shown (N=42).

Simulation procedures and software

Once inputs were defined, we repeatedly simulated the trial placebo and treatment arms to account for a range of potential outcomes with varying patient numbers per arm. We then compared the results of each trial using our algorithm, estimating power by tracking the number of times a significant result was obtained. This approach enabled identification of the requisite sample size at 80% power with pre specified statistical methods. Next, we repeated the simulations at varying target treatment differences between the placebo and treatment arms to determine respective requisite sample sizes. In total, 3,500 trial simulations were performed and analysed.

Simulations were conducted with an adapted standard software package (mvtnorm package [9] in R software [10]), which has been used to conduct similar simulations [11]. The actual simulation code used is provided in Appendix 1.

Statistical methods for the analysis of the simulated data

The simulated data were analysed using a repeatedmeasures mixed model (RMMM) *via* SAS PROC MIXED (version 9.4, Cary, NC). The CDASI-a score was the dependent variable, and baseline score, week, treatment (drug or placebo), and treatment by drug interaction were entered as covariates in the model. Patient was modelled as a random effect. The power obtained with multiple targeted treatment differences

in CDASI-a score between the placebo and treatment group at Week 24 were then tested with a 2 tailed test with α=0.05.

Results

CDASI natural history descriptive statistics

To obtain simulation inputs, descriptive statistics for the overall Stanford population and stratification subgroups were computed. The simulation inputs and their source data are summarized in Table 1.

Description	Data	Source
Baseline CDASI-a	21.2 points	All natural history subjects
Baseline CDASI-a standard deviation	11.77 points	All natural history subjects
CDASI-a correlation between baseline and Week 24	r=0.60	Natural history subjects with data at baseline, Week 12, and Week 24
Population	Baseline ≥ 15	Assumption
Average difference between treated and placebo arms	5, 7, and 10 points	Assumption
Duration of trial	6 months	Assumption
Frequency of assessments	Monthly	Assumption

Table 1: Summary of simulation inputs (CDASI-a, Cutaneous Dermatomyositis Disease Area and Severity Index activity subscale).

Patient Set	Time point	n	Mean (SD)	Median	Range
All Patients	Baseline	115	21.2 (11.77)	21.0	0, 50
	Week 12	34	17.4 (9.74)	17.5	0, 39
	CFB Week 12	34	-4.1 (6.72)	-3.0	-27, 6
	Week 24	20	15.2 (7.59)	16.5	2, 33
	CFB Week 24	20	-7.0 (9.77)	-4.5	-38, 5
Patients with BL CDASI ≥ 15	Baseline	81	27.0 (8.65)	26.0	15, 50
	Week 12	23	21.3 (8.86)	22.0	3, 39
	CFB Week 12	23	-6.7 (6.75)	-5.0	-27, 6
	Week 24	16	17.9 (5.77)	17.0	6, 33
	CFB Week 24	16	-8.4 (10.49)	-6.0	-38, 5
Patients with BL CDASI <15	Baseline	34	7.4 (4.38)	8	0, 14
	Week 12	11	9.3 (5.68)	10	0, 18
	CFB Week 12	4	1.1 (2.07)	0	-1, 5
	Week 24	11	4.5 (2.52)	4	2, 8
	CFB Week 24	4	-1.8 (2.63)	-2.5	-4, 2

Table 2: Descriptive statistics for CDASI activity subscale score (BL: Base Line; CDASI: Cutaneous Dermatomyositis Disease Area and Severity Index; CFB: Change from Baseline; SD: Standard Deviation).

For the overall population, the baseline mean (SD) CDASI-a score was 21.2 (11.77) points (Table 2). Baseline CDASI-a scores were similar between clinically amyopathic and classic patients (classic mean=22 [n=81], amyopathic mean=24 [n=23]). A total of 34 patients had a visit occurring within the 12 weeks visit window, with a mean (SD) change from baseline of -4.1 (6.72) points. A total of 20 patients had a visit occurring within the 24 weeks visit window, with a mean (SD) change from baseline of 7.0 (9.77) points. There were 42 patients with visits occurring in either the 12th or 24th week window, or 12 patients with visits occurring within both the 12th and 24th week visit windows (data not shown).

Descriptive statistics were analyzed for patients with baseline CDASI-a scores <15 (N=34) and ≥ 15 (N=81). The baseline mean (SD) score was 7.4 (4.38) and 27.0 (8.65) points for patients with baseline CDASI-a scores <15 and ≥ 15, respectively (Table 2). Among patients with baseline CDASI-a scores <15, 11 patients had a visit within the 12 weeks window and 4 patients had a visit within the 24 weeks window, with mean (SD) changes from baseline of 1.1 (2.07) and -1.8 (2.63) points, respectively. Among patients with baseline CDASI-a scores ≥ 15, 23 patients had a visit within the 12 weeks window and 16 patients had a visit within the 24 week window, with mean (SD) changes from baseline of -6.7 (6.75) and -8.4 (10.49) points, respectively.

The correlation of CDASI-a scores between visits was also required for the trial simulations to model the stability of the CDASI-a score within a subject over the time period of the trial. In the subset of patients with baseline CDASI-a scores ≥ 15 and who had data at baseline, Week 12, and Week 24, Pearson's correlations between baseline CDASI-a scores and Week 12 or Week 24 scores were 0.81 and 0.60, respectively.

Target population

A target trial population with a baseline CDASI-a score of ≥ 15 was simulated as this represents patients with moderate-to-severe skin disease [7]. In addition, patients with a low CDASI-a score at baseline (<15) reported small magnitudes of change on the CDASI over time, while patients with higher baseline CDASI-a scores (≥ 15) had greater score improvement over time (Table 2 and Figure 1). The limited raw score improvement observed in patients with baseline CDASI-a scores <15 suggested a potential floor effect of the assessment and a compelling reason to enrich the clinical trial with patients having higher disease activity.

Duration of treatment and frequency of assessments

A trial duration of 24 weeks was selected for the simulations, consistent with expert reviews that have shown common standard of care treatments often require at least two to four months to generate improvements in disease manifestations [12]. Within the 24 weeks treatment period, we assumed monthly clinic visits, for a total of 6 follow-up assessments. The 4 weeks interval between assessments was assumed to be a reasonable burden on the patient and clinic, with the multiple assessments allowing for analysis of repeated measurements.

Target difference between treatment arms

The assumed target treatment difference (i.e., advantage of drug over placebo) is a critical parameter for this analysis. Prior studies reported a MCID in CDASI-a score of 4 to 5 points in individual patients [7]. Therefore, we evaluated whether targeting a higher level of improvement might be warranted. To explore the proportion of patients who experienced a 4 to 5 point improvement under treatment regimens that are assumed to be as efficacious as a test medication, we plotted the cumulative frequency of CDASI-a improvement over 24 weeks for patients with a baseline CDASI-a score ≥ 15 and fit a cumulative distribution function line to the data. The analysis showed that 65%, 53% and 47% of patients had at least a 5, 7 and 10 points change in CDASIa score from baseline, respectively (Figure 2). Of note, the patients in this study did not receive standardized or controlled treatment, which likely increased data variability. Overall, the analysis suggested a 5 to 10 point range was a reasonable target treatment

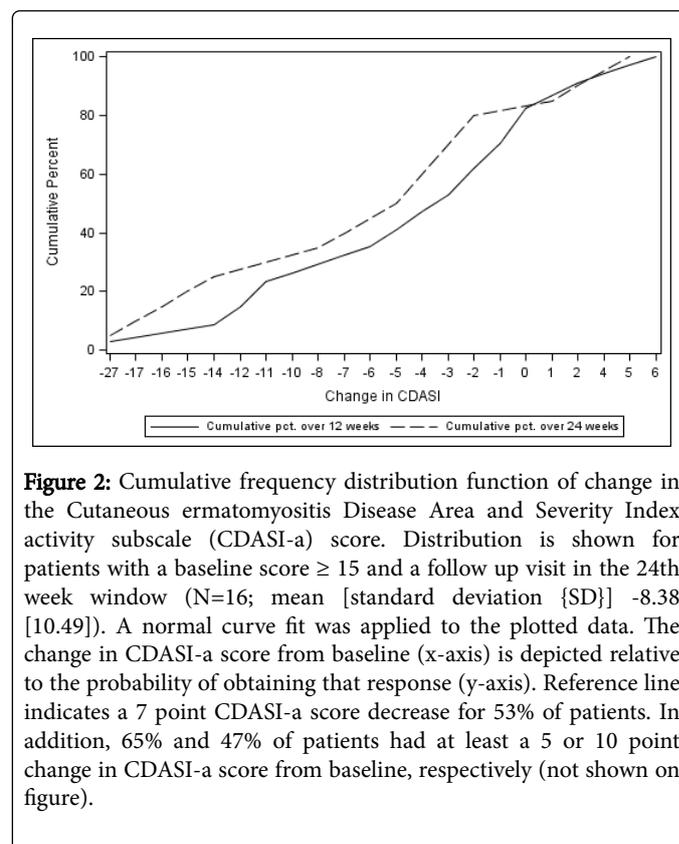difference for evaluation in the trial simulations in a population with moderate to severe DM.

Figure 2: Cumulative frequency distribution function of change in the Cutaneous ermatomyositis Disease Area and Severity Index activity subscale (CDASI-a) score. Distribution is shown for patients with a baseline score ≥ 15 and a follow up visit in the 24th week window (N=16; mean [standard deviation {SD}] -8.38 [10.49]). A normal curve fit was applied to the plotted data. The change in CDASI-a score from baseline (x-axis) is depicted relative to the probability of obtaining that response (y-axis). Reference line indicates a 7 point CDASI-a score decrease for 53% of patients. In addition, 65% and 47% of patients had at least a 5 or 10 point change in CDASI-a score from baseline, respectively (not shown on figure).

Clinical trial simulations

Once the inputs for the power simulation were obtained, the simulation was conducted. Data for a DM patient population were simulated based on the descriptive statistics obtained from the natural history data (baseline mean CDASI-a score=21.2, SD=11.77, r=0.6). Out of this population of simulated data, patients with a baseline CDASI-a score ≥ 15 were "selected" for analysis, as these data represented patients with moderate to severe DM.

We then explored the requisite sample size needed to detect a given treatment difference in a scenario with repeated measures analysis and 6 follow-up assessments conducted at 4 week intervals. To obtain 80% power with a 2 sided test, 12 patients per group were required to detect a 10 point treatment difference, 23 patients were required per group to detect a 7 point difference between groups, and 45 patients were required per group to detect a 5 point treatment difference (Figure 3).

To test the sensitivity of the model to the assumption of the strength of the correlation, we conducted a second simulation in a similar manner to the first while varying the correlation of the CDASI-a score between baseline and month six, considering r values of 0.5, 0.6, and 0. The power and sample size required for each of these conditions were then examined, assuming a treatment difference of 7 points. The sample size required for a t-test (rather than RMMM) was also calculated to examine the power gained by repeated measurements of the CDASI (Figure 4). Approximately 16 patients, 22 patients, or 27 patients were required to obtain 80% statistical power assuming

correlations of r=0.7, r=0.6, or r=0.5, respectively. A summary of the simulations performed are presented in Table 3.
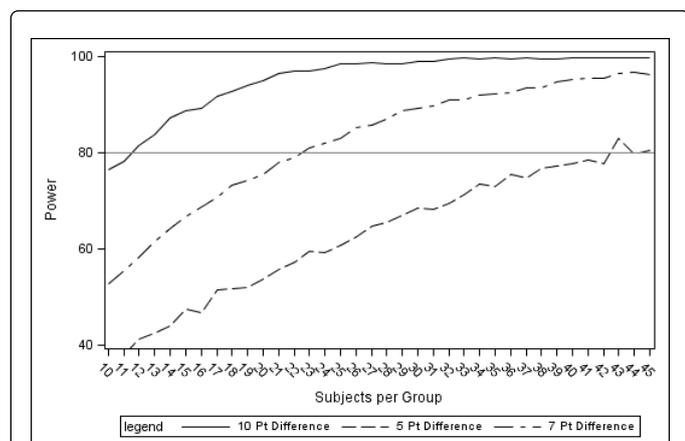

Figure 3: Power and sample size for Cutaneous Dermatomyositis Disease Area and Severity Index activity subscale score by assumed mean difference between baseline and Week 24 scores. The simulated correlation between baseline and Week 24 was r=0.6 with a first order autoregressive correlation structure.

	Treatment difference		
Correlation	5 points	7 points	10 points
0.5		X	
0.6	X	X	X
0.7		X	

Table 3: Simulations performed to facilitate clinical trial design (Within-subject correlations between Baseline and Week 24 Cutaneous Dermatomyositis Disease Area and Severity Index activity subscale score. Power simulations that were performed are indicated by "x").

Discussion

Clinical development pathways are needed to promote the development and regulatory approval of novel DM therapies. Unfortunately, there is limited precedent in DM as only two therapies have been approved by the U.S. Food and Drug Administration (prednisone and repository corticotropin injection), with both approvals occurring more than 60 years ago. Compounding this limited precedent, few well-controlled therapeutic trials and natural history studies have been reported. This paucity of published information hampers the development, proper design and powering of clinical trials in DM. In the RIM trial of rituximab in DM [13], a significant difference between treatment groups was not found and the manuscript discusses the barriers of study design and proper inputs to the power calculation as contributing to the negative result. Proper powering and design may have resulted in a statistically significant result.

In this study, we used computer simulations and available natural history data to explore various design and power options for a proof-of-concept trial of a novel DM therapy using the CDASI-a as the primary efficacy endpoint. To conduct the trial simulations, we made several important assumptions. First, we assumed that a clinical trial of an experimental therapy for DM would need to enroll a population with moderate to severe disease. The natural history analyses showed that a baseline CDASI-a score ≥ 15 is a good definition for such a cohort, consistent with the lower end of the 14 to 19 cut-off ranges between mild and moderate to severe disease reported in the literature [7].
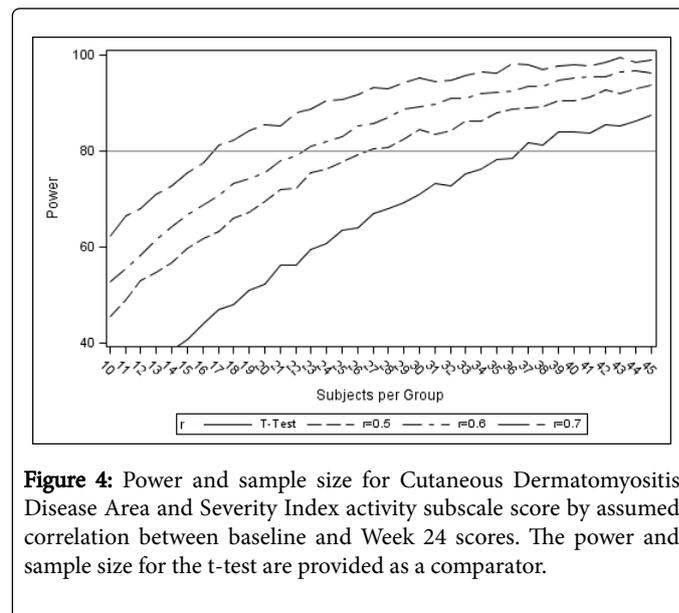

Figure 4: Power and sample size for Cutaneous Dermatomyositis Disease Area and Severity Index activity subscale score by assumed correlation between baseline and Week 24 scores. The power and sample size for the t-test are provided as a comparator.

Second, we assumed trial duration of 24 weeks with six monthly assessments. Based on expert reviews of common standard of care treatments, it was assumed that 24 weeks was a reasonable time frame to observe the depth and durability of a response in patients with recurrent or resistant DM. Ultimately, the power and sample size calculations from the trial simulations supported the assumed trial duration, indicating that a reasonably sized trial population (12 to 45 patients per group) was sufficient to detect clinically relevant changes on the CDASI-a over 24 weeks.

Third, we assumed a target treatment difference between the treatment and placebo arms of 5 to 10 points. The lower limit of this difference was anchored by the MCID of 4 to 5 points demonstrated in individual patients. Importantly, the CDASI natural history analyses showed that patients with a baseline CDASI-a score ≥ 15 had a mean decrease of 8.4 points over 24 weeks (Table 3) and 65% experienced at least a 5 point improvement (Figure 2). That led us to believe that an average CDASI improvement of greater than 5 points in a moderate to severe population is reasonable and that such a treatment would be an important addition to the treatment armamentarium.

Finally, we assumed statistical methods included a 2-sided RMMM analysis. At 80% power, the trial simulations showed that sample sizes of 45 patients per group, 23 patients per group, or 12 patients per group were required to detect a 5 point, 7 point, or 10 point treatment difference, respectively (Figure 3). In trials of rare diseases, it is important to maximize power given the practical limitations on sample size, and the use of repeated assessments and a mixed model analysis can increase power. The simulation demonstrated that use of a RMMM yielded savings between 10 to 20 patients per group (27% to 54%) compared to a student's t-test, depending on the correlation assumption used (Figure 4).

Future clinical trials in DM will likely need to test new regimens against the current standard of care. Patients being followed in the Stanford natural history study received standard of care according to a physician's judgement and were not randomised to treatment and the treatment regimen was not controlled or standardised. We used those data to represent standard of care treatment in a clinical trial but these data may have increased variability due to the fact that the regimen of background therapy varied within and between patients. In a well-controlled clinical trial setting, patients assigned to the standard of care arm may experience less variation. Thus, the analyses reported here may be enhanced in the future with the reporting of longitudinal data from prospective natural history studies and randomized controlled clinical trials.

A potential study limitation was the small number of patients who informed the estimate of correlation. Given the size of the dataset available for correlational analysis (only 12 patients had CDASI-a scores at baseline, Week 12, and Week 24), we could not test various correlational structures to see which one had the best fit. Instead we assumed that the data would follow an autoregressive [1] structure. This is a reasonable assumption for longitudinal data, and it is supported by the fact that the Week 12 correlation (r=0.81) was approximately equal to the square root of the Week 24 correlation (r=0.60). To address this limitation, we examined study power under multiple correlation strengths (Figure 4), which allows those planning future clinical trials to calibrate their power calculations as the situation dictates.

Overall, this study proposes a framework for a feasible clinical development pathway in cutaneous DM. Cutaneous disease manifestations are a major concern for patients and physicians due to their severity and impact on quality of life. The CDASI is validated, reliable, and responsive outcome measure of DM-related skin disease [7]. The sample size required to detect a clinically relevant change on the CDASI-a over 24 weeks is reasonable for a rare disease population.

In the future, trial design parameters may be further refined with the availability of additional CDASI data. Consistent with our objectives, the analyses reported here contributed to the design of a randomised, double-blind, placebocontrolled Phase 2 clinical trial of the investigational Tolllike receptor antagonist drug candidate IMO8400 in adult patients with DM (NCT Identifier: NCT02612857). It is our hope that the methods described in this study may also guide the design of future DM clinical trials using the CDASI as a primary endpoint.

Acknowledgements

Conflict of Interest:

Funding Sources:

References:

1. Iorizzo III LJ, Jorizzo JL (2008) The treatment and prognosis of dermatomyositis: an updated review. J Am Acad Dermatol 59: 99–112.
2. Luo YB, Mastaglia FL (2015) Dermatomyositis, polymyositis, and immunemediated necrotizing myopathies. Biochem Biophys Acta 1852: 622-632.
3. Rider LG, Werth VP, Huber AM, Alexanderson H, Rao AP, et al. (2011) Measures for Adult and Juvenile Dermatomyositis, Polymyositis, and Inclusion Body Myositis. Arthritis Care Res 63: S118-157.
4. Bronner IM, Van der Meulen MF, de Visser M, Kalmijn S, Van Venrooij WJ, et al. (2006) Long-term outcome in polymyositis and dermatomyositis. Ann Rheum Dis 65: 1456-1461.
5. Goreshi R, Chock M, Foering K, Feng R, Okawa J, et al. (2011) Quality of life in dermatomyositis. J Am Acad Dermatol 65:1107-1116.
6. Robinson ES, Feng R, Okawa J, Werth VP (2015) Improvement in the cutaneous disease activity of patients with dermatomyositis is associated with a better quality of life. Br J Dermatol 172: 169174.
7. Anyanwu CO, Fiorentino DF, Chung L, Dzuong C, Wang Y, et al. (2015) Validation of the Cutaneous Dermatomyositis Disease Area and Severity Index: characterizing disease severity and assessing responsiveness to clinical change. Br J Dermatol 173: 969-974.
8. Goreshi R, Okawa J, Rose M, Feng R, Lee LA, et al. (2012) Evaluation of reliability, validity, and responsiveness of the CDASI and the CAT-BM. J Invest Dermatol 132: 1117-1124.
9. Genz A, Bretz F, Miwa T (2016) mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-5.
10. R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
11. Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, et al. (2010) A versatile gene-based test for genome-wide association studies. Am J Hum Genet 87: 139-145.
12. Findlay AR, Goyal NA, Mozaffar T (2015) An overview of polymyositis and dermatomyositis. Muscle Nerve 51: 638-56.
13. Oddis CV, Reed AM, Aggarwal R, Rider LG, Ascherman DP, et al. (2013) Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 65: 314-324.