Computational Statistics and Data Analysis 199 (2024) 108015

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Optimizing designs in clinical trials with an application in treatment of Epidermolysis bullosa simplex, a rare genetic skin disease

Joakim Nyberg [a,*], Andrew C. Hooker [a], Georg Zimmermann [b,e], Johan Verbeeck [c], Martin Geroldinger [b,e], Konstantin Emil Thiel [b,e], Geert Molenberghs [c,d], Martin Laimer [f,g], Verena Wally [f]

[a] Department of Pharmacy, Uppsala University, SE-751 24, Uppsala, Sweden
[b] Team Biostatistics and Big Medical Data, Intelligent Data Analytics (IDA) Lab Salzburg, Paracelsus Medical University, A-5020, Salzburg, Austria
[c] Data Science Institute (DSI), Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, BE-3500, Hasselt, Belgium
[d] Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), KULeuven, BE-3000, Leuven, Belgium
[e] Research and Innovation Management, Paracelsus Medical University Salzburg, A-5020, Salzburg, Austria
[f] EB House Austria, Research Program for Molecular Therapy of Genodermatoses, Department of Dermatology and Allergology, University Hospital of the Paracelsus Medical University Salzburg, A-5020, Salzburg, Austria
[g] Department of Dermatology and Allergology, University Hospital of the Paracelsus Medical University, A-5020, Salzburg, Austria

ARTICLE INFO

Keywords:
Epidermolysis bullosa simplex
Optimal design
D-optimal
Clinical trial design

ABSTRACT

Epidermolysis bullosa simplex (EBS) skin disease is a rare disease, which renders the use of optimal design techniques especially important to maximize the potential information in a future study, that is, to make efficient use of the limited number of available subjects and observations. A generalized linear mixed effects model (GLMM), built on an EBS trial was used to optimize the design. The model assumed a full treatment effect in the follow-up period. In addition to this model, two models with either no assumed treatment effect or a linearly declining treatment effect in the follow-up were assumed. The information gain and loss when changing the number of EBS blisters counts, altering the duration of the treatment as well as changing the study period was assessed. In addition, optimization of the EBS blister assessment times was performed. The optimization was utilizing the derived Fisher information matrix for the GLMM with EBS blister counts and the information gain and loss was quantified by D-optimal efficiency. The optimization results indicated that using optimal assessment times increases the information of about 110-120%, varying slightly between the assumed treatment models. In addition, the result showed that the assessment times were also sensitive to be moved ± one week, but assessment times within ± two days were not decreasing the information as long as three assessments (out of four assessments in the trial period) were within the treatment period and not in the follow-up period. Increasing the number of assessments to six or five per trial period increased the information to 130% and 115%, respectively, while decreasing the number of assessments to two or three, decreased the information to 50% and 80%, respectively. Increasing the length of the trial period had a minor impact on the information, while increasing the treatment period by two and four weeks had a larger impact, 120% and 130%, respectively. To conclude, general applications of optimal design methodology, derivation of the Fisher information matrix for GLMM with count

* Corresponding author.
E-mail address: joakim.nyberg@farmaci.uu.se (J. Nyberg).

https://doi.org/10.1016/j.csda.2024.108015
Received 1 June 2023; Received in revised form 25 June 2024; Accepted 25 June 2024

Available online 2 July 2024
0167-9473/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

data and examples on how optimal design could be used when designing trials for treatment of the EBS disease is presented. The methodology is also of interest for study designs where maximizing the information is essential. Therefore, a general applied research guidance for using optimal design is also provided.

1. Introduction

Epidermolysis bullosa simplex (EBS) is a rare, genetic disease characterized by hyperfragility of epithelial-lined tissues and surfaces leading to recurrent blistering particularly of skin and mucous membranes (Has et al., 2020). As there is currently no cure and available treatments are limited to symptom alleviation and wound care, a growing number of innovative therapeutic compounds are evaluated in clinical trials. One of these trials was a randomized, placebo-controlled, double-blind, 2-period cross-over phase 2/3 trial, which assessed the therapeutic potential of the immunomodulatory 1% diacerein cream compared to placebo to reduce the number of blisters (Wally et al., 2018).

Sixteen patients in this trial were randomly assigned to either the placebo or the diacerein treatment according to a 1:1 ratio, and were treated daily for 4 weeks, followed by a 3-month follow-up. After a washout period, patients were crossed over to the opposite treatment, following an identical treatment schedule. In each study period, blisters in the treated body surface area were counted (in the clinics) at the start of the treatment period, after 2 and 4 weeks of treatment, and after follow-up (week 16). The primary endpoint was the proportion of patients with more than 40% reduction from baseline in the number of blisters after 4 weeks of treatment. In this work, the longitudinal raw blister counts (0-16 weeks) are modeled with a generalized linear mixed model (Verbeeck et al., 2024). An illustration of the design of the EBS study is presented in Fig. 1.

The work presented in this manuscript is conducted within the EBStatMax project of the European Joint Program on Rare Diseases, which aims to explore and establish innovative statistical methods for rare diseases. The goal of this paper is to utilize optimal design methodology (Fedorov, 2010; Atkinson and Donev, 1992) to enhance the expected information (i.e. outcome) of the EBS trial and to investigate how sensitive the information in the designs is with respect to different design choices, such as the number of blister assessments, treatment duration, study period length and timing of the blister assessments.

The methodology applied in this paper is neither novel in terms of the statistical methods used nor with respect to the study design in general, but it demonstrates how to practically combine knowledge regarding a drug and a disease to provide a rationale for the optimal design of future clinical trials using a parametric model, with the aim of extracting a maximum amount of information. This is of high relevance as trials for rare diseases may fail not because of a lack of efficacy of a treatment, but due to a lack of appropriate study design for the study objectives and hence for lack of power (Day et al., 2018).

When using optimal design in a clinical trial, the level of information contained in the trial is quantified with the expected Fisher information (Fedorov, 2010; Atkinson and Donev, 1992). The idea of optimal design is to change certain aspects of the trial and to evaluate the respective influence on the expected information about model parameters, particularly on parameters related to the treatment effect (efficacy) of an experimental treatment. The derivation of the expected Fisher information based on an estimated GLMM is described in Section 2, which also contains information on the methodology used to assess various design aspects. Section 3, describes how the information derived from a study changes in response to the different design parameters investigated. Finally, Section 4 discusses results and contemplates how the methods presented here are generalizable to other rare diseases.

2. Methods

The starting point of this work is a reanalysis of a diacerein versus placebo phase 2/3 study, in which changes in blister counts were modeled with a modeling average (MA) approach (Wally et al., 2018; Verbeeck et al., 2024; Aoki et al., 2017). The MA approach used a pool ($n = 16$) of different generalized linear mixed effect models (GLMMs), including different sources of variability, such as

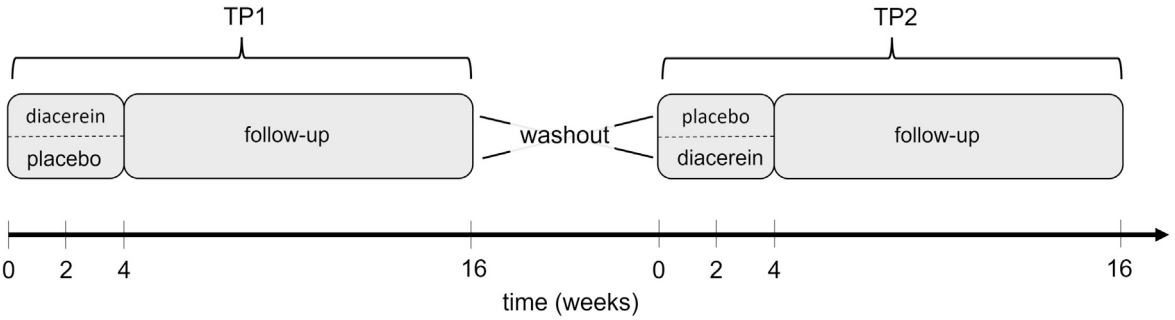

Fig. 1. An illustration of the standard design used in the placebo-controlled, double-blind, 2-period cross-over phase 2/3 trial described by Wally et al. (Wally et al., 2018). The left-hand box represents the first trial period (TP1), which includes a treatment arm (diacerein) and a placebo arm with a 1:1 ratio. The next trial period (TP2) starts after a washout period, upon which the two treatment groups switch treatments (cross-over design). The x-axis indicates the blister count assessment times relative to the start of each trial period.

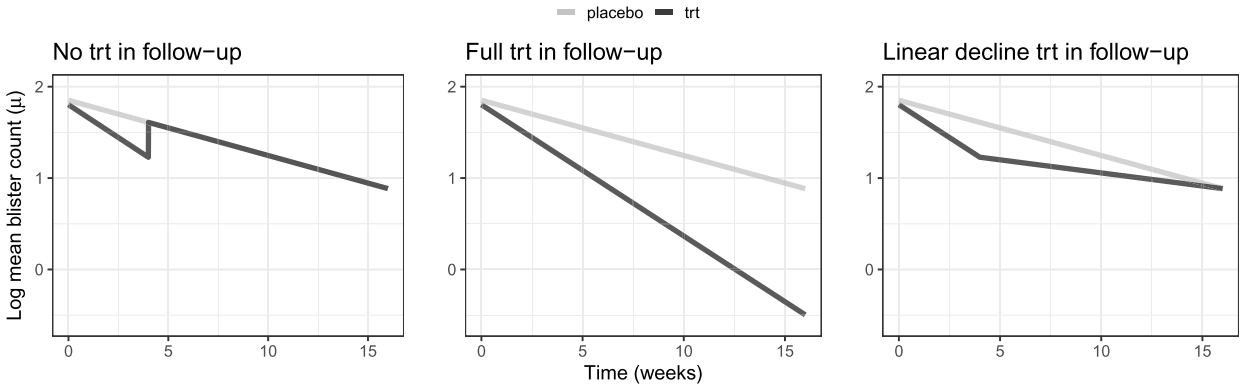

Fig. 2. Illustration of the different treatment effect models used in the optimal design. The models are illustrated on the log link function scale, i.e. log mean blister counts, versus study period time (in weeks). The illustration assumes both placebo and treatment in study period 1 (TP1). No treatment effect (trt) in the follow-up versus placebo (left panel), equal (full) treatment effect in follow-up period and the treatment period versus placebo (mid panel, the best MA model) and linear declining treatment effect in the follow-up period versus placebo (right panel).

period effects P_i (TP1 ($i = 1$) versus TP2 ($i = 2$)), time effects t_k (blister count assessment time t for patient k) and treatment effects G_{ik} (placebo versus diacerein in period i for patient k). The most informative model from the MA approach, i.e. the model with the highest MA weight ($w = 0.58$) and hence the most influential GLMM, was reused in this work on the optimal blister count assessment. The conditional expectation of the blister outcome, y_{ikt}, given the patient-specific random effects $\boldsymbol{b}_k$, is described by the following GLMM:

$$\mu_{ikt} = E(y_{ikt}|\boldsymbol{b}_k) = g^{-1}(\boldsymbol{x}'_{ikt}\boldsymbol{\beta} + \boldsymbol{z}'_{ikt}\boldsymbol{b}_k), \tag{1}$$

with $\boldsymbol{x}_{ikt}$ and $\boldsymbol{\beta}$, the vectors of the fixed effects variables and parameters, $\boldsymbol{z}_{ikt}$ and $\boldsymbol{b}_k$ vectors of the random effects variables and parameters and g^{-1} an inverse link function. The random effects are assumed to follow a multivariate normal distribution with mean zero and covariance matrix Ω, i.e., $\boldsymbol{b}_k \sim \mathcal{N}(0, \Omega)$. The blister counts are assumed Poisson distributed and hence a log link function was used. The model with the highest weight ($w = 0.58$) from the MA approach (Verbeeck et al., 2024) is:

$$\log(\mu_{ikt}) = g(E(y_{kt}|\boldsymbol{b}_k)) = \beta_{intercept} + \beta_{trt}G_{ik} + \beta_{time}t_k + \beta_{period}P_i + \beta_{trt:time}G_{ik}t_k + \beta_{period:time}P_i t_k + \boldsymbol{z}'_{kt}\boldsymbol{b}_k,$$

with G_{ik} a treatment group indicator (diacerein versus placebo), P_i a trial period indicator (TP1 versus TP2) and t the continuous period time (in weeks). The random effect models used were

$$\boldsymbol{z}'_{kt}\boldsymbol{b}_k = b_{0,k}P_i + b_{1,k}(1 - P_i),$$

where

$$\begin{pmatrix} b_{0k} \\ b_{1k} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right).$$

The fixed effects parameters of this GLMM estimated from the EBS trial data are:

$$\widehat{\boldsymbol{\beta}'} = \left(\hat{\beta}_{intercept}, \hat{\beta}_{trt}, \hat{\beta}_{time}, \hat{\beta}_{period}, \hat{\beta}_{trt:time}, \hat{\beta}_{period:time}\right)$$
$$= (1.8510, -0.0482, -0.0604, -0.3494, -0.0833, 0.0730)$$

and the inter-patient variability parameters for each period were estimated as

$$\hat{\Omega} = \begin{pmatrix} 0.182 & 0.179 \\ 0.179 & 0.502 \end{pmatrix}.$$

Note that β_{trt} as well as $\beta_{trt:time} * t$ are accounted for when calculating the total treatment effect at time t. Moreover, the estimated model assumes a continuation of the drug's effect throughout the treatment period and the follow-up. This is a quite strong assumption, which arguably impacts the optimal design assessment. Therefore, two additional drug effects models were investigated: (1) a linearly declining drug effect in the follow-up period, and (2) a model assuming no drug effect in the follow-up period. The three evaluated treatment models are illustrated in Fig. 2.

2.1. Optimal design

In this work, the expected information from various designs of the EBS trial were classified using an optimal design approach. In optimal design evaluation, the expected Fisher information matrix (FIM) is often used to quantify the level of information in a design.

Furthermore, a single criterion, i.e., the determinant of the Fisher information matrix, was used as a single metric of the information, referred to as the D-criterion. Other criteria, such as A-optimal, common for linear models, C-optimal etc. are also available (Atkinson and Donev, 1992). However, the D-optimal approach is one of the most adopted criteria for nonlinear models and was used in this paper since the approximate FIM is derived for nonlinear random effect models even though the model with the highest weight from the MA approach was a linear model. The paragraph below describes a linearized version of a FIM that is used for GLMMs with Poisson distributed data, based on the work by Longford (Longford, 1994) and Ogungbenro et al. (Ogungbenro and Aarons, 2011). See also Stroup, Bolker and Madden et al. (Stroup, 2012; Bolker, 2015; Madden et al., 2002) for more information about GLMMs and design of experiments.

The patient-specific FIM_k is dependent on the model f, the patient-specific elementary design $\boldsymbol{\xi}_k$, and the model parameters $\boldsymbol{\beta}$ (fixed effects) and $\boldsymbol{b}_k \sim \mathcal{N}(0, \Omega)$ (random effects):

$$FIM(\boldsymbol{\xi}_k) = \begin{pmatrix} J_k^T V_k^{-1} J_k & 0 \\ 0 & \frac{1}{2} F_k \end{pmatrix},$$

where

$$\begin{aligned} F_k &= \mathrm{tr}\left(V^{-1} \frac{\partial V}{\partial \omega_r} V^{-1} \frac{\partial V}{\partial \omega_s} \right), \\ \boldsymbol{J}_k &= \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}}, \\ V_k &\approx Z_k \Omega Z_k^T + W_k, \\ W_k &= \mathrm{diag}(\mu_{k,1}, \ldots, \mu_{k,n_k}), \\ Z_k &= \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{b}_k}, \\ log(\mu_k) &= f(\boldsymbol{\xi}_k, \boldsymbol{\theta}, \boldsymbol{b}_k) \end{aligned} \tag{2}$$

and $\boldsymbol{\theta}$ encompasses all model parameters. Once a patient specific FIM_k is calculated, a FIM representing the complete study cohort can be defined by a summation over the information from each elementary (patient's) design:

$$FIM = \sum_{k=1}^{N} FIM_k(\xi_k),$$

where N is the total number of patients in the study. Note that assuming the same elementary design in each patient enables further simplification to the study-level FIM:

$$FIM = \sum_{j=1}^{J} n_j \cdot FIM_j(\xi_j)$$

where J is the number of unique elementary designs $\boldsymbol{\xi}_j$ and n_j are the number of patients with the elementary design $\boldsymbol{\xi}_j$, and evidently $N = \sum_{j=1}^{J} n_j$. This notation also allows for usage of nonlinear models given a first order approximation of the model around $\boldsymbol{b}_k = 0$, although in some cases a first-order expansion is insufficient to reach acceptable accuracy (Molenberghs and Verbeke, 2005). In the context of nonlinear mixed effects models with continuous data this assumption has shown to predict the expected information well (Nyberg et al., 2015).

Here, a D-optimal design approach was used to maximize the determinant of FIM or equally to minimize the joint expected model parameter uncertainty. The numerical optimizations were performed in the statistical programming software R, version 4.2, using a simplified version of a repeated linear search algorithm with the aim to approximate a global optimum, similar to the line search algorithm implemented in the optimal design software PopED (Foracchia et al., 2004; Nyberg et al., 2012b). Furthermore, the level of information between two designs (for example A and B) is compared using the D-efficiency:

$$D_{\text{eff}} = \frac{|FIM_A|^{1/p_A}}{|FIM_B|^{1/p_B}},$$

where p_A and p_B (here $p_A = p_B$) is the number of estimated parameters with design A and design B respectively. A D_{eff} of 100% means that design A and B are similarly informative, while a D_{eff} of 200% and 50% means that design A is twice as informative as design B and design A contains half of the information of design B, respectively.

Based on the data from the study by Wally et al. (Wally et al., 2018) and the modeling work presented by Verbeeck et al. (Verbeeck et al., 2024) the elementary design, referred to as the standard design, corresponds to 8 patients starting with diacerein treatment and 8 patients starting with placebo treatment with equal blister assessment times at $t = \{0, 2, 4, 16\}$ weeks. Further, a treatment duration of 4 weeks and a follow-up of 12 weeks, repeated in a cross-over design after a wash-out phase, were assumed, analogously to the original study. Note that the number of assumed patients in an elementary design will not change the D-optimal design but will affect the expected uncertainty.

2.2. Design space

The design space for the four blister count assessment times was defined as $t_{min} = 0$ weeks and $t_{max} = 16$ weeks where the optimal period assessments times t_{opt} were optimized within t_{min} to t_{max}. In the scenarios investigating the study period length, see subsection 2.5 below, two additional t_{max} boundaries were investigated; $t_{max} = 20$ weeks and $t_{max} = 24$ weeks. The optimal period assessment times were forced to be the same in the two treatment periods as well as between the different treatment groups (placebo or diacerein treatment). Hence, each patient had 8 blister count observations in total and roughly 128 (2*8*8) observations were expected in the full study except for the scenario with fewer/additional assessment times (see Section 2.4).

2.3. Sensitivity of the standard design

Sampling windows (assessment time windows) were determined to investigate how sensitive the standard design is due to changes in the blister assessment times. Two different magnitudes of shifts were investigated: ± maximum one week and ± maximum one day from the standard design. For the first time point and last time point, an assessment time within 0–2 (day or weeks) and 14–16 weeks or 15.71 (16 weeks - 2 days) –16 weeks were tested. The sampling windows were calculated using 1000 uniformly distributed samples of blister assessment times within the defined sampling windows, corresponding to 1000 different designs. It is assumed that all patients use the same assessment times, in each of the 1000 sampled studies.
From a clinical point of view, some flexibility with respect to the assessment times might considerably decrease the trial burden for the study participants: Due to the disease, every additional visit at the clinic is burdensome (e.g., even seemingly routine activities like changing clothes or additional bandage are very painful for the patients). Therefore, the possibility of combining a study visit with a regularly scheduled monitoring visit at the clinic – given that both visits are sufficiently close to each other – has the potential of decreasing the frequency of study withdrawal substantially and lowering traveling burden for the patient.

2.4. Changing the number of blister assessments

The information of the design with respect to the number of blister assessments was investigated by optimizing designs with either 2, 3, 5, 6 assessments (optimized between $t_{\min} = 0$ weeks and $t_{\max} = 16$ weeks) in each study period and compared to the standard design with 4 assessments per study period by calculating the D-efficiency for each optimal design relative to the standard design.

2.5. Increasing the study period length

The standard design with a 4-week treatment period (using the three different treatment model's optimal assessment times within the 16 weeks period length) and the sensitivity of the standard design with respect to the study period length were investigated by optimizing assessment times for a period duration of 16 weeks (standard design), 20 weeks or 24 weeks, still with a fixed treatment period length of 4 weeks. Optimal designs were calculated for the three different treatment models and the D-efficiency was calculated relative to the standard design.

2.6. Increasing the treatment duration

The standard design with a 4-week treatment period and the sensitivity of the standard design with respect to the treatment duration was investigated by altering the treatment duration in the model to either 6 weeks or 8 weeks while keeping the study period time fixed to a totality of 16 weeks. For each scenario (4, 6, and 8 weeks) and with the three different treatment models, the design was optimized with respect to the blister count assessment times and the D-efficiency was calculated relative to each standard design (with optimized assessment times with a 4-week treatment duration).

3. Results

The optimal designs were compared to the standard design by means of the D-efficiency as well as the expected relative standard error (RSE), i.e. the uncertainty, of the model parameters. The optimal designs were $\boldsymbol{\xi}_j = (0, 0, 4, 16)$ weeks, $\boldsymbol{\xi}_j = (0, 0, 16, 16)$ weeks and $\boldsymbol{\xi}_j = (0, 0, 3.88, 16)$ weeks for the model without treatment effect in follow-up, the model with continuing treatment effect in follow-up, and the model with linearly declining treatment effect in follow-up respectively. The RSE's and respective D-efficiency are presented in Table 1. The table also shows that some parameters are difficult to estimate with high precision, such as the treatment effect (β_{trt}) because of the large RSE's. On the other hand, the treatment effect interaction with time ($\beta_{trt:time}$) has a stronger impact on the blister counts and hence might be more informative from an optimal design perspective.

3.1. Sensitivity of the standard design

The sensitivity of the standard design with respect to variations in blister assessment times ($t = 0, 2, 4, 16$) for all simulated follow-up treatment effects is presented in Fig. 3 and shows that the standard design with no treatment effect in follow-up is very sensitive, i.e. shows a large variability in expected information, to ± weekly spreads of assessment times, spreading from an D-efficiency of slightly above 100% for some designs all the way down to <5% D-efficiency for other designs. The mean D-efficiency of the sampled

Table 1
Expected Relative standard errors of the model parameters and D-efficiency for the optimal designs (OD) relative to the standard design (Std) which is assumed the same for all treatment effect (trt) models.

Parameter	No trt in follow-up		Full trt in follow-up		Linear trt in follow-up	
	Std	OD	Std	OD	Std	OD
$\beta_{intercept}$	7.64	7.40	7.65	7.64	7.58	7.39
β_{trt}	320.93	277.16	287.54	285.68	321.01	276.99
β_{time}	15.94	15.03	24.01	17.44	15.82	15.02
β_{period}	61.82	57.57	57.89	57.95	61.82	57.56
$\beta_{trt:time}$	57.55	52.07	22.88	16.9	71.17	53.40
$\beta_{time:period}$	65.01	56.93	23.57	17.58	70.51	58.61
ω^2_{TP1}	38.93	38.22	41.43	43.00	40.22	38.19
ω^2_{TP2}	38.57	38.02	39.07	40.17	38.94	38.00
D-Efficiency (%)	100.00	109.48	100.00	120.91	100.00	111.97

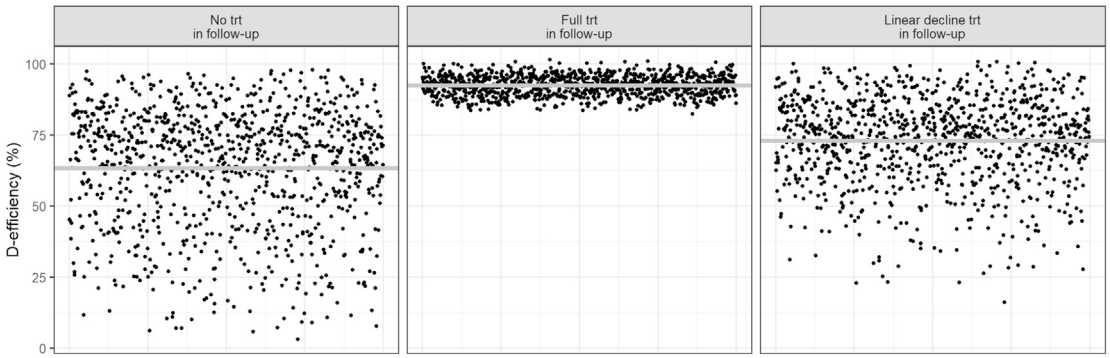

Fig. 3. Illustration of D-efficiency based on 1000 sampled designs from the standard design ± one week. The figure is stratified by the different treatment effect models in follow-up. The standard designs have blister assessments at week 0,2,4 and 16. The grey solid lines represent the mean D-efficiency.

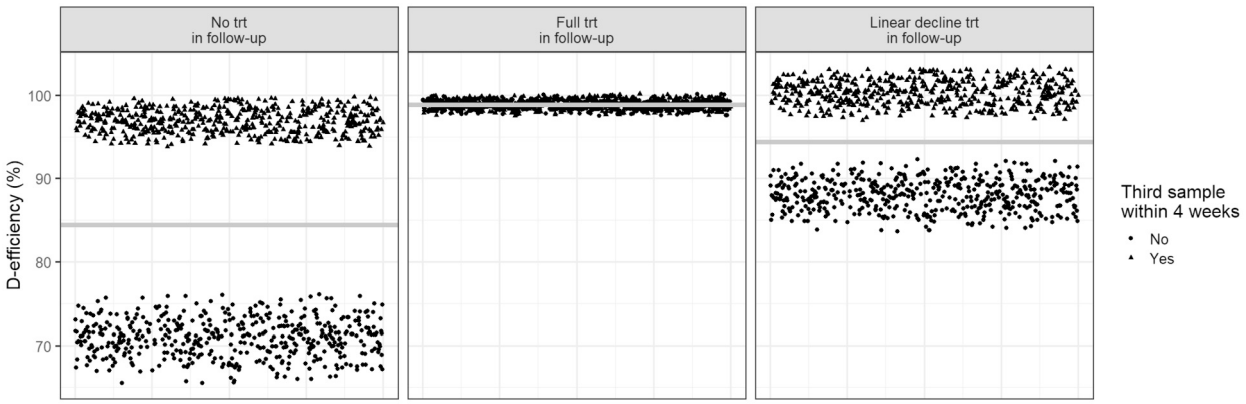

Fig. 4. Illustration of D-efficiency based on 1000 sampled designs from the standard design ± one day. The figure is stratified by the different treatment effect models and the third sample in the standard design around week 4 is marked with different symbols if the assessment time is within or outside of the treatment period. The standard designs have blister assessments at week 0,2,4 and 16. The grey solid lines represent the mean D-efficiency.

designs from ± one week within the standard design is around 65%. The model with full treatment effect in follow-up is much less sensitive to the spread in assessment time (mean D-efficiency of around 90%) while the linear decline model is somewhere in between with an average D-efficiency of 75%. Fig. 4 shows that the standard design is very robust, i.e. small loss of expected information, to designs ± one day from the standard design, especially if a full treatment effect in the follow-up is assumed. If a linear decline treatment effect or no treatment effect is assumed, the D-efficiency could be as low as 85% and 65%, respectively. However, the sensitivity of these designs depends heavily on the number of assessment times with a full treatment effect, favoring more assessment times in the full treatment effect period. On average the D-efficiency is quite high in all assumed treatments models, 98%, 94% and 85%, for the full treatment effect, the linear decline treatment effect and no treatment effect in follow-up, respectively.

Table 2
Expected D-efficiency (%) of the optimal design using different assessment times compared to the optimal design using 4 assessment times for each treatment effect model. The D-efficiency of the three different treatment effect models in the follow-up is tabulated; No = no treatment effect in follow-up period, Full = full treatment effect in follow-up period and Linear = Linear decay treatment effect in follow-up period.

# assessment times	Follow-up trt effect		
	No	Full	Linear
6 assessment times	132.19%	131.75%	133.17%
5 assessment times	117.99%	115.69%	118.75%
4 assessment times	100.00%	100.00%	100.00%
3 assessment times	79.21%	79.26%	80.25%
2 assessment times	48.06%	59.56%	48.69%

Table 3
Expected D-efficiency (%) of the optimal design assuming different study period lengths compared to the optimal design with a study period duration of 16 weeks. The D-efficiency of the three different treatment effect models in the follow-up is tabulated; No = no treatment effect in follow-up period, Full = full treatment effect in follow-up period and Linear = Linear decay treatment effect in follow-up period.

Period duration	Follow-up trt effect		
	No	Full	Linear
16 weeks	100.00%	100.00%	100.00%
20 weeks	102.39%	104.59%	101.94%
24 weeks	103.78%	104.74%	103.78%

3.2. Changing number of assessments

The D-efficiency of the optimal designs using different numbers of blister assessments are presented in Table 2, which indicates that the three different models give comparable D-efficiency. Furthermore, Table 2 shows that a 15–20% increase or decrease in D-efficiency is seen for each added/removed assessment time with slightly more loss of information for fewer assessment times in the full treatment effect model. The optimal designs for the different scenarios are not presented here but they all add additional replicated assessment times to the optimal designs for the respective follow-up treatment models.

3.3. Increasing the study period length

The D-efficiency is not very sensitive to increases in the study period duration from 16 weeks (standard design) to either 20 weeks or 24 weeks, (Table 3). As expected, the design is more informative with longer duration, but the D-efficiency gain is less than 5% for the three models with different treatment effects in the follow-up. Again, the optimal designs (not shown) are similar to the optimized designs for the 16-week duration but the late assessment time is pushed towards the end of the study period (at week 20 or week 24).

3.4. Extending the treatment period

Table 4 shows that treatment period length does have a major impact on the D-efficiency for the 'no effect' and 'linear decay effect' model, with an increase of efficiency of roughly 20% for and additional treatment duration of 2 weeks and roughly 30% more information if the treatment duration is doubled, i.e., to 8 weeks. This is expected since the parameter related to the treatment duration ($\widehat{\beta}_{trt:time}$) has a relative large impact on the blister count. The full treatment effect in the follow-up model is not affected by the change of treatment duration since the treatment effect is similar throughout the entire follow-up period.

4. Discussion

In the examples provided, optimal design methodology was able to increase the information in several aspects of the EBS designs, by optimizing blister assessment times, number of assessments, study period, and treatment length. In addition to optimizing designs, sensitivity of the standard design with respect to the blister assessment times was also investigated and shows that the design is not very sensitive towards switching the optimal design ± one day, as long as the number of assessments within the treatment period is similar to in the standard design (3 assessments within the treatment period). Shifting the design ± one week from the standard design could reduce the information to less than 5% of the information in the standard design (D-efficiency <5%). More combinations

Table 4
Expected D-efficiency (%) of the optimal design assuming different treatment period lengths compared to the optimal design with a treatment period duration of 4 weeks. All models assume a study period length of 16 weeks. The D-efficiency of the three different treatment effect models in the follow-up is tabulated; No = no treatment effect in follow-up period, Full = full treatment effect in follow-up period and Linear = Linear decay treatment effect in follow-up period.

Treatment duration	Follow-up trt effect		
	No	Full	Linear
4 weeks	100.00%	100.00%	100.00%
6 weeks	117.92%	100.00%	119.46%
8 weeks	131.55%	100.00%	133.27%

of optimizing the design space could be considered, e.g. simultaneously optimizing the period duration, the number of assessments and the assessment time. However, for simplicity and run-times reasons we only considered two dimensions at a time.

In this work, an optimal design was sought for the EBS trial, using a pre-specified model with already assumed known parameters. This is an often-used approach, even for nonlinear models where the design depends on both the model and the model parameters. However, incorrect assumptions regarding the model structure and/or the model parameters could change the optimal design and the expected precision. Other approaches, such as robust optimal designs (Tod and Rocchisani, 1997; Nyberg and Hooker, 2012; Foo and Duffull, 2010; Dodds et al., 2005; D'Argenio, 1990) or even using the full pool of estimated models from the MA approach (Alhorn et al., 2019) could be applied instead. On the contrary, even though robust approaches can be considered, they are in general computationally more intensive and for the purpose of showing the benefits and potential with the optimal design approach, the pre-specified model and parameter estimates from the MA approach deemed appropriate. In addition, the used single model from the MA approach had much of the total model weight and therefore contributed with high importance to the overall MA model (Verbeeck et al., 2024). More mechanistic models, e.g. with slow onset treatment effects, could also be considered using the methods presented in the paper and may further improve the information gain with the optimal design approach.

An issue which is often present in small sample size studies, such as in rare disease populations, is that model parameters might suffer from large uncertainty. This is also the case for some parameters in the model that was used in this work e.g. β_{trt} parameter in Table 1. On the other hand, high uncertainty is expected in small studies and is therefore an argument for the importance of maximizing the available information.

Increasing the number of blister assessment times from the standard design which had 4 assessment times, increased the signal-to-noise ratio by adding additional replicated assessment times instead of additional unique assessment times. This is something which is seen in optimal design when there is a sufficiently large sample size to identify the model and estimate its parameters. This could be addressed by adding autocorrelation components to the model, e.g., AR(1) (autoregressive of order one) models where a blister count is dependent on the magnitude of the previous blister count and the distance in time between the assessments (Nyberg et al., 2012a). However, the magnitude of the auto-correlation might be difficult to assess and therefore a common assumption, which was also used in this work, is to ignore any potential autocorrelation.

Finally, the results presented in this paper are based on the assumed model and only considering optimal design theory of information while other aspects, such as therapeutic and side effects, are not part of our investigation. For example, it is likely that the period length should be extended when testing longer treatment period and that a longer treatment period comes with higher patient burden and cost. Nevertheless, such restrictions could be added to the optimal design evaluations and considered if they could be specified in a model and/or the design space. However, this does not mean that the considerations presented in this manuscript are purely theoretical. In fact, the opposite is true: Optimization of study designs is of crucial importance in Epidermolysis bullosa (as well as in other rare diseases), due to the high burden for the patients who participate in a clinical trial. Since the disease is rare, the distances between the patients' cities of residence and specialized centers are usually quite large. Moreover, seemingly routine activities of everyday life are very burdensome and painful for the patients. Therefore, from a clinical perspective, the statistical questions addressed in this manuscript are very important for designing a clinical trial. At the same time, however, for the reasons stated above, we must acknowledge that further refinements and more detailed methodological considerations are needed in order to align the results from optimal design theory with "reality", that is, actually planning and conducting a clinical trial. This does not only apply to the work presented in this manuscript but should be regarded as a key aim of research on optimal designs in general.

To conclude, this research illustrates how optimal design methodology can be used to improve and investigate clinical trial designs in rare diseases. Optimal design assessments are particularly important when designing studies in small populations, like rare disease studies, where it is crucial to maximize the information from the study population.

5. Applied researcher guidance

A recipe on how the research presented could be applied when planning for a study design is presented in Fig. 5. This is not intended to cover all aspects of study design and planning but gives an overview of the specifics using the optimal design theory

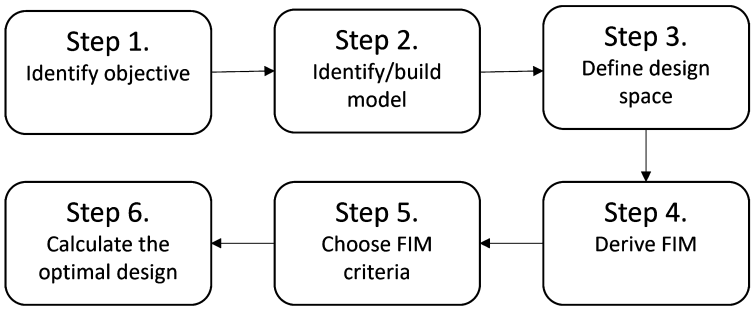

Fig. 5. Illustration of key steps needed to calculate an optimal study design. Step 1 - Identify the objectives of your optimal design, i.e. design a study which is maximizing the study power, estimate all model parameters with as high precision as possible, estimate the number of patients needed to reach 80% study power, etc. Step 2 - Build a model from historical data or use a model from literature which represents the relevant outcome of the study, e.g. in this paper the relevant outcome was blister counts. Step 3 - Identify what aspects of the design to optimize, i.e. number of patients, study length, number of assessments, etc. In addition, define the design boundaries based on e.g. clinical considerations; minimum and maximum number of assessments, minimum and maximum study length, minimum and maximum number of patients to include, balance in study arms, etc. Step 4 - Derive the fisher information matrix (FIM) for your model, i.e. for Poisson distributed data with GLMM the FIM presented in this paper could be used. For continuous normal distributed data with nonlinear mixed effect models, see e.g. (Foracchia et al., 2004). Step 5 - Choose the criteria to minimize/maximize, e.g. the joint parameter uncertainty (D-optimal design as used in this paper) or other criteria of relevance. These criteria should match the objectives. Step 6 - Calculate the optimal design using an optimal design software (Nyberg et al., 2012b) or by using any suitable optimization package.

which is used in this paper. Also note that the objectives, aspects of the design and the design criteria is very specific to the study and the drug indication. After applying the methodological steps (Fig. 5), the result will be a design (or multiple designs) that matches the objectives.

Acknowledgement

Georg Zimmermann gratefully acknowledges the support of the WISS 2025 project 'IDA-Lab Salzburg' (20204-WISS/225/197-2019 and 20102-F1901166-KZP). All authors gratefully acknowledge the funding of the "EBStatMax Demonstration Project" by the European Joint Programme on Rare Diseases (EU Horizon 2020 research and innovation programme, grant agreement no. 825575). Verena Wally gratefully acknowledges the support of DEBRA Austria.

References

Alhorn, K., Schorning, K., Dette, H., 2019. Optimal designs for frequentist model averaging. Biometrika 106, 665–682.
Aoki, Y., Röshammar, D., Hamrén, B., Hooker, A.C., 2017. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. J. Pharmacokinet. Pharmacodyn. 44, 581–597. https://doi.org/10.1007/s10928-017-9550-0.
Atkinson, A., Donev, A., 1992. Optimum Experimental Designs. Oxford Statistical Science Series. Clarendon Press.
Bolker, B.M., 2015. Linear and generalized linear mixed models. In: Ecological Statistics: Contemporary Theory and Application, pp. 309–333.
D'Argenio, D.Z., 1990. Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. Math. Biosci. 99, 105–118.
Day, S., Jonker, A.H., Lau, L.P.L., Hilgers, R.D., Irony, I., Larsson, K., Roes, K.C., Stallard, N., 2018. Recommendations for the design of small population clinical trials. Orphanet J. Rare Dis. 195.
Dodds, M., Hooker, A., Vicini, P., 2005. Robust population pharmacokinetic experiment design. J. Pharmacokinet. Pharmacodyn. 32, 33–64. https://doi.org/10.1007/s10928-005-2102-z.
Fedorov, V., 2010. Theory of Optimal Experimental Design. Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, pp. 581–589.
Foo, L.K., Duffull, S., 2010. Methods of robust design of nonlinear models with an application to pharmacokinetics. J. Biopharm. Stat. 20, 886–902.
Foracchia, M., Hooker, A., Vicini, P., Ruggeri, A., 2004. Poped, a software for optimal experiment design in population kinetics. Comput. Methods Programs Biomed. 74, 29–46.
Has, C., Bauer, J., Bodemer, C., Bolling, M., Bruckner-Tuderman, L., Diem, A., Fine, J.D., Heagerty, A., Hovnanian, A., Marinkovich, M., Martinez, A., McGrath, J., Moss, C., Murrell, D., Palisson, F., Schwieger-Briel, A., Sprecher, E., Tamai, K., Uitto, J., Woodley, D., Zambruno, G., Mellerio, J., 2020. Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility. Br. J. Dermatol. 183, 614–627. https://doi.org/10.1111/bjd.18921.
Longford, N., 1994. Logistic regression with random coefficients. Comput. Stat. Data Anal. 17, 1–15.
Madden, L.V., Turechek, W.W., Nita, M., 2002. Evaluation of generalized linear mixed models for analyzing disease incidence data obtained in designed experiments. Plant Disease 86, 316–325. https://doi.org/10.1094/PDIS.2002.86.3.316.
Molenberghs, G., Verbeke, G., 2005. Models for Discrete Longitudinal Data. Springer.
Nyberg, J., Bazzoli, C., Ogungbenro, K., Aliev, A., Leonov, S., Duffull, S., Hooker, A., Mentré, F., 2015. Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics. Br. J. Clin. Pharmacol. 19, 6–17.
Nyberg, J., Hooker, A.C., 2012. The robustness of global optimal designs. In: Abstr 2529, Population Approach Group in Europe Conference, p. 21. https://www.page-meeting.org/?abstract=2529.
Nyberg, J., Höglund, R., Bergstrand, M., Karlsson, M.O., Hooker, A.C., 2012a. Serial correlation in optimal design for nonlinear mixed effects models. J. Pharmacokinet. Pharmacodyn. 39, 239–249.
Nyberg, J., Ueckert, S., Strömberg, E., Hennig, S., Karlsson, M., Hooker, A., 2012b. Poped: an extended, parallelized, nonlinear mixed effects models optimal design tool. Comput. Methods Programs Biomed. 108, 789–805.
Ogungbenro, K., Aarons, L., 2011. Population Fisher information matrix and optimal design of discrete data responses in population pharmacodynamic experiments. J. Pharmacokinet. Pharmacodyn. 38.
Stroup, W.W., 2012. Generalized Linear Mixed Models: Modern Concepts, Methods and Applications, 1st ed. CRC Press.

Tod, M., Rocchisani, J., 1997. Comparison of ed, eid, and api criteria for the robust optimization of sampling times in pharmacokinetics. J. Pharmacokinet. Biopharm. 25, 515–537.

Verbeeck, J., Geroldinger, M., Thiel, K., Hooker, A.C., Ueckert, S., Karlsson, M., Bathke, A.C., Bauer, J.W., Molenberghs, G., Zimmermann, G., 2024. How to Analyze Continuous and Discrete Repeated Measures in Small-Sample Cross-Over Trials? Biometrics 79, 3998–4011. https://doi.org/10.1111/biom.13920.

Wally, V., Hovnanian, A., Ly, J., Buckova, H., Brunner, V., Lettner, T., Ablinger, M., Felder, T., Hofbauer, P., Wolkersdorfer, M., Lagler, F., Hitzl, W., Laimer, M., Kitzmüller, S., Diem, A., Bauer, J., 2018. Diacerein orphan drug development for epidermolysis bullosa simplex: a phase 2/3 randomized, placebo-controlled, double-blind clinical trial. J. Am. Acad. Dermatol. 78, 892–901.