Points to consider in the design and analysis of clinical trials in small populations

Version:	1.0
Date:	25 Nov 2022
Authors:	Francois Aubin, Andreas Kaiser, Frances Lynn, Tal Otiker, Carol Reid, Aysun Cetinyurek Yavuz on behalf of SIG "Small populations"

Contents

1 Introduction / Abstract

Diseases that occur in populations with a low prevalence or in vulnerable populations lead to limitations in the number of available patients for drug development. In such a setting, when sample sizes for clinical development are limited, one usually speaks of small populations. Drug development in small populations poses some special challenges. This applies above all to the statistical planning of the clinical studies, their statistical evaluation and interpretation as well as decision-making. The challenges can often only be addressed using innovative statistical methods and data analysis approaches. This ***points-to-consider*** document is aimed at statisticians who are going to work in this exciting field intending to give them an initial introduction and to sensitize them to the peculiarities of small populations. Basic points to consider in clinical development for small populations are addressed and potential useful statistical approaches are proposed.

2 Characteristics of small populations

2.1 Few patients

Conducting clinical trials in small populations can result from four main categories of causes: a low prevalence of the condition studied, the need to limit the number of subjects exposed to the investigational treatment(s), the heterogeneity of the disease and external/contextual factors. In addition, two or more causes from these categories of causes are frequently associated.

- Low prevalence:
  - Rare disease (a disease is considered as rare if fewer than 1 out of 2000 in the EU, or fewer than 200,000 people affected by the condition in the US, corresponding to about 1 out of 1500 people). It has also been proposed to name “ultra-rare diseases” diseases that have a prevalence of <1 per 50 000 persons.
  - Unique study population (e.g., studying the effect of exposure to microgravity in astronauts)
  - Remote or isolated environments
- Need to limit the number of subjects exposed:
  - constraints regarding the particular protection of individuals (paediatric population, pregnant women)
  - toxicity of the product (cancer treatment)
  - potential lack of efficacy (placebo, small doses…)
- Heterogeneity
  - of the disease, making it difficult to obtain a clinically relevant study population
  - of the treatment due to, e.g., individually tailored therapy
- External/contextual factors:
  - Emergency situations
  - Public health urgency
  - Restricted resources coupled with a high level of need

2.2 Disease characteristics

Disease characteristics would impact the design of any clinical trial, as there is no "one design fits all", and this is even more relevant when designing trials in small populations. The following disease characteristics should be considered:

- Rare disease and non-rare disease – does the disease meet the criteria for a "rare" or "ultra-rare" disease (see above)
- Onset of disease – is it possible to know the time of disease onset? Do patients need to be recruited at time of diagnosis?
- Disease progression and clinical outcome – are there standard endpoints used to measure disease progression? Is there regulatory guidance on which endpoints can be used for registration? Can disease progression lead to intercurrent events that need to be considered in the estimand framework?
- Severity of disease – is there a standard way to measure disease severity? Should the study limit eligibility by severity of disease at study entry? Should the randomization be stratified by severity?
- Heterogeneity of disease and symptoms – how heterogenous is the patient experience with the particular disease? Should eligibility criteria be tightened to reduce heterogeneity in the trial population?
- Co-morbidities – what are the common co-morbidities in the target population and how may they impact the trial in terms of collection of endpoints, dropout rate, subjective PRO and missing data?
- Available treatment options – is there a treatment indicated for the disease? What is the standard of care? Does it differ based on geographical region? Are off-label treatments commonly prescribed for the disease? What are1st line, 2nd line, 3rd line therapies? Is there benefit to recruitment of patients who failed certain other therapies vs a treatment naïve population? What would be the impact of rescue medications? Can patients be asked to avoid those for the duration of the trial and if not how to take that into account in the estimand framework?
- New and innovative treatment options – are there other clinical trials running concurrently for the same disease testing new treatments that may become available to patients during the course of the study?

2.3 Feasibility and operational constraints

2.3.1 Specialized centres

For rare diseases, frequently there are only a limited number of centres that treat patients with these diseases. These centres are quite often academic centres unused to working on clinical trials, which can lead to issues with compliance and data quality. Often, the data collected may relate to unusual endpoints which are difficult for monitors to check adequately.

The centres will often treat patients who have been referred from all over the world. This can lead to some logistical problems due to the need for travel and accommodation in addition to problems introduced due to language difficulties. In the case of childhood diseases, family members will also have to travel and find accommodation which can add costs and complication to the trials.

2.3.2 Slow recruitment

Small numbers of patients can lead to very slow recruitment, which will lengthen the time to submission although this may be mitigated by identifying referral centres that will send their patients for treatment. Conversely, in areas of unmet need, it may be that patients are lining up for treatment and the difficulties are in deciding who will be treated.

3 Designing studies for small populations

3.1 What do we know about the population to design a study?

Often in small populations, there is limited data available to inform the study design. There may not be accepted endpoints and, in many cases, there will be a need to develop novel endpoints including composites. Understanding of the expected outcomes may be limited.

3.2 Interaction with stakeholders

3.2.1 Interactions with patient groups and clinical networks

Interactions with patient groups are essential in diseases in small populations. They often have extensive knowledge of the disease course and a passion for new therapies giving them a willingness to collaborate with companies who are developing new treatments. In rare diseases the clinical network will be very small and those working in the area will all be familiar with each other's work. Different sites will often have their own way of doing things that may have been developed over years and it can be challenging to develop a protocol that suits all of the sites involved in the study.

3.2.2 Limited knowledge of disease by regulators

For rare diseases, the regulators frequently have little understanding of the nature and course of the disease. This can make it very difficult to gain their agreement on study designs. There will often be novel endpoints used and it can be difficult to gain the acceptance of the regulators for an endpoint with which they are unfamiliar. The desire of the regulators for a RCT in place of a single arm study can often lead to unrealistic sample sizes being proposed. The need for statistical significance in these situations will often require some very creative thought regarding the primary endpoints proposed. Often in rare diseases with unmet need, there will be a need to compare the disease course following treatment with the therapy under development with the natural history of the disease. In this situation, it can be difficult to convince regulators of the comparability of the treated and natural history populations, particularly when those in the natural history populations are identified at a more advanced stage of the disease. This is a situation in which patient groups are invaluable, as they are able to educate the regulators on the disease course that would be expected if the patients were not treated.

3.3 Eligibility criteria

Trade-off between broad criteria and power: when setting eligibility criteria for a clinical trial, sponsors may wish to keep these as broad as possible to allow more patients to enrol into the trial and speed up timelines. This is likely to be even more applicable to small populations, where enrolment is almost always a challenge. However, the broader the eligibility criteria, the greater the chance of diluting the treatment effect by including subjects who are potentially less likely to benefit

from the new treatment and by introducing heterogeneity into the study population which will increase the variability. The outcome would be a possible reduction in the power of the study.

Impact on applicability of extrapolation: when planning to use extrapolation between two populations, the key question is how similar the studies conducted in each population are. Hence, if extrapolation is considered, it is important to design the studies to maximise the similarity between the source and target populations. For example, when designing a trial in a paediatric population, it would make sense to try and use the same eligibility criteria used in the adult studies to enable future extrapolation.

3.4 Limitations of „traditional" designs

A primary purpose of many clinical trials is evaluation of the efficacy of an experimental intervention. In a well-designed trial, the data that are collected and the observations that are made will eventually be used to overturn the equipoise. Central principles in proving efficacy, and thereby eliminating equipoise, are avoiding bias and establishing statistical significance. This is ideally done through the use of controls, randomization, blinding of the study, credible and validated outcomes responsive to small changes, and a sufficient sample size.

The recruitment of a sufficient sample-size may be prevented in situations like

- Small populations: low prevalence (rare diseases), unique study population, remote or isolated environments
- Need to limit the number of subjects exposed: paediatric population, pregnant women, highly toxic products, potential lack of efficacy in severe/life threatening diseases
- Highly heterogeneous disease population
- Time constraints (emergency situations, public health urgency, high unmet medical need)

As a consequence, it may become difficult to conduct a randomized, internally controlled and adequately powered trial in a reasonable timeframe [1], [2].

In addition, particularly in the context of rare (orphan) diseases, a traditional clinical development is made impossible by the addition of several of the above factors: many rare diseases affect children, are heterogeneous in their presentation, correspond to an unmet clinical need rendering long term follow-up to assess hard clinical endpoints like survival impossible.

Orphan diseases are often poorly characterized and under-researched, making it difficult to choose a relevant primary endpoint and to estimate the expected effect-size to allow an adequate sample-size computation.

Consequently, pivotal studies for orphan drug approvals are more likely to be smaller, single arm, nonrandomized, open label, and to use surrogate endpoints to assess efficacy.

3.5 Randomization & Blinding

Although randomized controlled trials are the gold standard, in severe diseases it may not be ethical to include a placebo arm and there may not be a suitable active comparator to include in a randomized trial. RCTs with equal allocation are generally the most efficient designs to assess effectiveness; however, depending on the circumstances, alternatives such as unequal allocation (i.e., more patients receive the new drug than the control), which can provide increased safety

experience and reduce the use of placebo, or a dose-comparison design (i.e., randomization to more than one dose, with or without placebo) could be considered. In the case of unequal randomization, the control arm could be supplemented with external control data. Other points of consideration for small populations are

- Stratification: usually matching or stratification may improve power [3] but the effectiveness of stratified randomisation in small clinical trials is questionable [4].
- Response adaptive methods [5]
- Randomisation based inference should be considered, particularly for complex statistical models and when response adaptive randomisation is used [4]

In some situations, for example in a life-threatening disease with no available treatments a single arm study with comparison to external control data may be acceptable. Bias can be minimized if the natural history of the disease is well understood and the primary endpoint is an objective endpoint. Video-recording and the use of centralized raters can also be used to limit bias.

3.6 Sample size determination and powering

In rare diseases it may not be possible to recruit the number of patients required to sufficiently power a study. Recruitment may be slow and a large number of sites might be required with only one or two patients recruited at each site which can result in less experienced sites and increased variability. Continuous variables usually give higher precision & smaller sample sizes than dichotomous variables, e.g., 'responder' & 'non-responder'. This is particularly true if the baseline value is accounted for in an appropriately pre-specified model like in analysis of covariance (ANCOVA) model. Other baseline measurements may also increase the efficiency of such models. The selection of experienced sites and/or training of outcome assessors can help to reduce variability and increase power. In some rare disease settings, a one-sided 5% significance level may be acceptable and powering a study at 80% rather than 90% could be considered as ways of reducing the sample size.

As mentioned above it may be possible in some situations to have a single arm study or to supplement a small placebo arm with external control data.

3.7 Utilizing external data sources for analysis

When a fully randomized adequately powered study is not possible external control data can be used to supplement a placebo arm using Bayesian borrowing, or in place of an internal control arm [6]. Methodologies such as inverse probability treatment weighting based on propensity scores or matching can be used to generate a control arm that is similar to the study treatment arm in terms of prognostic variables. See Sec. 4 for further details.

3.8 Useful designs

"*No methods exist that are relevant to small studies that are not also applicable to large studies. However, it may be that in conditions with small and very small populations, less conventional and/or less commonly seen methodological approaches may be acceptable if they help to improve the interpretability of the study results*" [5].

While randomised clinical trials with concurrent control are still considered the gold standard [5] [7], it is recommended to systematically take into consideration alternative trial design options when studying treatments for a rare disease:

- Group-sequential designs [5] [7]: The possibility of early stopping provides a potential reduction of sample size but they may also increase study size in some circumstances.
- Adaptive designs [7] [4] [1], among which inferentially seamless adaptive designs
- Externally controlled trials (historical or concurrent control), see Sec. 4 for further details: a comparative trial will usually be preferable but may not always be possible.
- Designs that allow subjects to be used more than once [7] [1]
    - multiple n-of-1 trials [5],
    - crossover trial designs (For stable diseases with relatively short treatment duration, and where there are sufficient data to determine an appropriate washout period, cross-over designs may allow potential large reductions in sample size)
    - factorial designs,
    - repeated measurements (longitudinal trials)
    - randomised withdrawal designs
- Risk based allocation designs. Because the design is nonrandomized, its use should be considered only in situations in which an RCT would not be possible.
- Bayesian designs and methods [7] [4] [1]

3.9 Considerations on some other design features

While hard clinical endpoints, such as survival or serious morbidity, are preferred, surrogate endpoints may be an option although evaluation of the surrogates might be challenging [5] [7] [4]. Patient-centred endpoints could be an alternative when hard clinical endpoints cannot be observed in sufficient patients within a reasonable timeframe.

Inefficient use of baseline measurements like change scores, percentage change from baseline should be avoided and approaches like ANCOVA should be used instead [8].

Hierarchical models [1] provide a natural framework for combining information from a series of small clinical trials conducted within ecological units (e.g., space missions or clinics). They also provide a foundation for analysis of longitudinal studies, which are necessary for increasing the power of research involving small clinical trials.

Inclusion of covariates measured after the start of treatment should be avoided when developing statistical models.

Dichotomisation of continuous variables (responder analysis) leads to information loss and should be avoided as well if not clinically justified [5] [8].

There is no ‘one size fits all’ solution. The right combination of design features and techniques have to be identified for each clinical trial in a small population.

4 Utilizing external data

4.1 Historical trial data and real-world data

Randomized clinical trials (RCTs) provide the best quality evidence. For small populations, however, traditional RCTs may not always be feasible and obtaining adequate power may not be possible with relevant endpoints. Use of historical data might be helpful, but a careful consideration is needed. Leveraging historical data from previous or concurrent clinical trials can be done by elicitation of informative Bayesian prior distributions [9], [10], [6], [11], [12]. There are well developed methods for different types of analysis used by industry. However, these approaches are still novel.

Real world data (RWD) and real-world evidence (RWE) are another option which are particularly important for rare diseases. Real world evidence is the evidence obtained outside randomized clinical trials. FDA has released RWD/RWE framework [13] and they are actively exploring ways to optimize the use of RWD to support regulatory decision making. Due to the demand coming from patient populations, increased awareness of stakeholders and regulators, pharma companies have more interest in using RWD to support their efforts for rare diseases. RWD in rare diseases has other challenges: the populations are diverse, and may be followed up by different specializations, there are different data sources, etc. Different types of data sources, such as electronic medical records, patient registries, natural history registries might be helpful to answer important research questions for small populations. The data obtained via these routes might not always be of best quality. The focus should be on quality of the data, not on quantity. While attempting to use RWD to generate evidence, it is essential to improve the quality of data and standardize the data collection to increase their utility. Furthermore, due to the small patient populations for rare diseases, it is also recommended that cooperation within and across countries should be done to collect good quality data.

The use of historical data or RWD data can support single arm trials serving as an "external" control group for small populations. The FDA draft guidance on rare disease [14] lists the use of external control data as an option for a serious disease when there is (1) an unmet medical need, (2) a highly predictable disease course which can be objectively measured and (3) an expected drug effect which is large, self-evident and temporally close to the interventions. When using external data, the external control group needs to be similar enough to the investigated group. This can be achieved by matching techniques, use of propensity scores or inverse probability treatment weighting. Alternatively, the size of the control group can be diminished by using an informative prior distribution constructed from "external" control arm from previous clinical trials and/or RWD. Extrapolation from adults to paediatric populations might be considered to reduce the size of a clinical trial in paediatric populations.

4.2 Applicability of extrapolation

Extrapolation is an approach where knowledge about efficacy but also pharmacology and/or safety is extended from the so-called source population to the target population. Extrapolation is largely discussed and applied in paediatric drug development. There the paediatric patients represent the target population and the source population can be adult patients or paediatric patients from another development program where e.g. sufficient evidence for treatment effect has been established. More details on extrapolation can be found in regulatory guidelines like the EMA's

reflection paper on extrapolation in paediatrics [3] and the draft ICH guideline E11A on paediatric extrapolation [15]. Although these guidelines are designed for paediatric drug development, most of the outlined framework and considerations can be used as starting point for applying extrapolation in small populations in general.

While extending knowledge from source to target population by extrapolation, the development program in small population can focus on generating data to fill knowledge gaps or to increase the evidence to a degree as required for decision making. This offers the opportunity for studies with smaller sample sizes to conclude on benefit-risk.

The scientific basis for extrapolation lies in a sufficient similarity of the disease between target and source population. Relevant factors for evaluating similarities are

- pathophysiology and aetiology of the disease like clinical presentation and manifestation of the disease, biochemical, genetic/epigenetic, cellular, tissue, organ system, and epidemiologic factors, sub-type of disease and sub-populations, biomarkers that are common in the pathophysiology of the disease and disease progression,
- disease progression, in particular similar endpoints or biomarkers, endpoint measurements, disease onset, short-term and long-term outcome, available treatments and treatment effects
- pharmacology of the drug like underlying absorption, distribution, metabolism, and excretion (ADME) of the drug, mechanism of action (MOA), dose/exposure-pharmacodynamic relationship (PK/PD), relationship between biomarker end clinical endpoint, dose/exposure-response relationship, similarity of response measurements.

An exact similarity is usually not required, and some differences are accepted to a degree which does not preclude extrapolation.

Depending on the type of extrapolation and identified similarities, study designs can be optimized for small populations. For example, if efficacy of the reference treatment can be extrapolated from historical studies or real-world data, the treatment effect of the investigational drug could be investigated in a single arm study and compared with the external control data. Also study design with an unbalanced randomization in favour of the investigational drug could be an option where the reference treatment is augmented by external data using a Bayesian approach. If similarity between a biomarker common to the pathophysiology and clinical outcome is established and efficacy has been demonstrated in the source population then the primary objective of a study in small populations could be to show treatment effects for this biomarker. On the other hand, if there is a lack of knowledge about the therapeutic dose but a general PK/PD relationship has been established, dose-finding could be done by investigating primarily the PD response.

Extrapolation can support decision making as well. If efficacy has been established in the source population it might be sufficient to demonstrate efficacy in the target, i.e., small population with less level of evidence. For example, a higher significance level like 5% for a one-sided test instead of a 2.5% could be applied or within a Bayesian framework a level of evidence of at least 95% posterior probability instead of 97.5%. Larger non-inferiority margins might be an acceptable option in some situations as well. Further, evidence of source and target could be synthesised using Bayesian dynamic borrowing [11], [6] and Bayesian meta-analytical approach [16], [17] for decision making supported by e.g., tipping point analysis [18].

5 Decision making

In small populations, clinical trials with few patients are usually the only feasible option. Accordingly, the data to be generated is limited and thus, the evidence for efficacy or futility of the drug but also for safety is also limited. Because often few treatment options are available and the number of eligible patients is limited, it is desirable to avoid approval of futile drugs while efficacious drugs with meaningful clinical efficacy should be identified with high confidence. Running a clinical trial which has low probability to detect a clinically meaningful treatment effect might even raise ethical concerns. This leads to additional challenges in decision making. The type I error and type II error have to be balanced and controlled to an acceptable degree taking into account the limitations of sample size.

Although studies in small populations are usually planned with focus on primary efficacy endpoints, for decision making the totality of data will be considered. Accordingly, generated evidence for all key efficacy endpoints, pharmacodynamic and/or biomarker makers related to efficacy and safety endpoints are assessed to determine the net clinical benefit.

5.1 Methods and rules for decision making and level of evidence

Because general guidance is lacking, decision making in small populations follows the same rules as for traditional drug development with modified requirements as a result of discussions with regulatory bodies. In paediatric development, the draft ICH guideline on paediatric extrapolation [15] explicitly mentioned hypothesis testing with a significance level of 2.5% for a one-sided test as the standard approach. EMA's reflection paper on paediatric extrapolation [3] suggests a liberation of the significance level in well-reasoned situations and 5% for a one-sided hypothesis test has been accepted.

In addition, determination of the treatment effect size is of major relevance for assessing benefit for the patients. Accordingly, the generated evidence should be sufficiently large to have an acceptable precision of the estimate and special attention should be put to this when planning and analysing a clinical trial. Because the precision, measured by the width of confidence intervals, decreases with increasing level of significance in case of a positive test, significance levels of larger than 5 % are usually rejected by regulatory bodies.

When comparing the investigational treatment with an active control, the reflection paper [3] proposes to use larger non-inferiority margins as alternative approach to an inflated significance level. Although both approaches yield similar results in decision making, increasing the non-inferiority margin might impair the interpretation of the data by formally accepting clinically meaningless treatment effects.

Bayesian methods are explicitly supported as well and the reflection paper [3] gives augmentation by external data as an additional possibility, too. Minimal posterior probabilities for, e.g., superiority evaluation of 97.5% and 95% correspond to the 2.5% and 5% significance level of one-sided hypothesis tests have been accepted in the past. Note that only for the so called non-informative prior distributions 97.5% and 95% minimal posterior probability implies a strict type I error control of 2.5% and 5%, respectively. Beside the synthesis of evidence, see Sec. 5.2, Bayesian methods offer more flexibility in defining decision rules and in quantification of evidence including the

determination of treatment effects. In addition, Bayesian methods yield a more intuitive interpretation of the results compared to frequentist approaches in case of limited data. This is highlighted in [19] where frequentist and Bayesian approaches for decision making are evaluated in context of rare diseases. In addition, the authors present a decision theoretical approach where, e.g., net-clinical benefit can be assessed by an appropriate cost function. Here, the decision is taken by comparing the mean of the resulting posterior distribution for the cost functions for the different potential alternatives in the decision process and the alternative with the largest mean is chosen.

5.2 Evidence synthesis to support decision making

To increase the degree of evidence, the database can be enlarged by considering external (or historical) data from previous clinical trials, natural history studies or real-world databases. Such approaches have the advantage that data collection in the planned new study can be focussed on the investigational drug while evidence for, e.g., comparator treatment or placebo are borrowed from the external data source. This enables study designs which are more attractive to patients like single arm or highly unbalanced studies favouring randomization to the investigational treatment. For paediatric development in context of extrapolation, the use of external data to support decision making, i.e. by borrowing information, is given as an option in the ICH E11A [15] an EMA's reflexion paper on extrapolation [3]. Examples on how external data can be utilized are presented and discussed in [6].

Qualitative incorporation of external information for decision making can be done by extrapolations as well. For those similarities in, e.g., disease progression, pathophysiology of the disease and mode of action are considered [3], [15]. Further, evidence from other interventional treatments, population and/or indication can be synthesised to support decision making, see Sec. 4.

6 References

[1] Institute of Medicine, Small Clinical Trials: Issues and Challenges, Washington, DC: The National Academies Press, 2001.

[2] S. A. Bell und C. T. Smith, „A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov," *Orphanet Journal of Rare Diseases,* Bd. 9, p. 170, 2014.

[3] EMA/189724, *Reflection paper on the use of extrapolation in the development of medicines for paediatrics,* European Medical Agency, 2018.

[4] R.-D. Hilgers, F. König, G. Molenberghs und S. Senn, „Design and analysis of clinical trials for small rare disease populations," *Journal of Rare Diseases Research & Treatment,* Bd. 1, pp. 53-60.

[5] CHMP/EWP/83561, *Guideline on clinical trials in small populations,* Committee for Medical Products for Human Use (CHMP), EMA , 2006.

[6] H. Schmidli, D. A. Häring, M. Thomas, A. Cassidy, S. Weber und F. Bretz, „Beyond Randomized Clinical Trials: Use of External Controls," *Clinical Pharmacology & Therapeutics,* Bd. 107, pp. 806-816, 2020.

[7] S. Day, A. H. Jonker, L. P. L. Lau, R.-D. Hilgers, I. Irony, K. Larsson, K. C. Roes und N. Stallard, „Recommendations for the design of small population clinical trials," *Orphanet Journal of Rare Diseases,* p. 195, 2018.

[8] R.-D. Hilgers, K. Roes, N. Stallard und IDeAl, Asterix and InSPiRe project groups, „Directions for new developments on statistical design and analysis of small population group trials," *Orphanet Journal of Rare Diseases,* Bd. 11, p. 78, 2016.

[9] B. Neuenschwander, G. Capkun-Nigglia, M. Bransona und D. J. Spiegelhalter, „Summarizing historical information on controls in clinical trials," *Clinical Trials,* Bd. 7, pp. 5-18, 2010.

[10] H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter und B. Neuenschwander, „Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information," *Biometrics,* Bd. 70, pp. 1023-1032, 2014.

[11] M. Gamalo-Siebers, J. Savic, C. Basu, X. Zhao, M. Gopalakrishnan, A. Gao, G. Song, S. Baygani und L. Thompson, „Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation," *Pharmaceutical Statistics,* Bd. 16, pp. 232-249, 2017.

[12] M. Ghadessi, R. Tang, J. Zhou, R. Liu, C. Wang, K. Toyoizumi, C. Mei, L. Zhang, C. Q. Deng und R. A. Beckman, „A roadmap to using historical controls inclinical trials – by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG),“ *Orphanet Journal of Rare Diseases,* pp. 15-69, 2020.

[13] FDA, „Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products (draft guidance),“ *Guidance for Industry,* 2021.

[14] FDA, „Rare Diseases: Common Issues in Drug Development Guidance for Industry (draft guidance),“ *Guidance for Industry,* 2019.

[15] ICH E11A, „Pediatric extrapolation (draft),“ International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2022.

[16] C. Röver und T. Friede, „Dynamically borrowing strength from another study through shrinkage estimation,“ *Statistical Methods in Medical Research,* Bd. 29, pp. 293-308, 2020.

[17] C. Röver und T. Friede, „Bounds for the weight of external data in shrinkage estimation,“ *Biometrical Journal,* pp. 1-13, 2021.

[18] N. Best, R. G. Price, I. J. Pouliquen und O. N. Keene, „Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing,“ *Pharmaceutical Statistics,* Bd. 20, pp. 551-562, 2021.

[19] F. Miller, S. Zohar, N. Stallard, J. Madan, M. Posch, S. W. Hee, M. Pearce, M. Vågerö und S. Day, „Approaches to sample size calculation for clinical trials in rare diseases,“ *Pharmaceutical Statistics,* Bd. 17, pp. 214-230, 2018.