

Methodische Lösungsansätze zur Generierung belastbarer(er) Evidenz bei seltenen Erkrankungen

Ralf-Dieter Hilgers

IQWIG Herbstsymposium

25 November 2023



This research is part of the EU-FP7 **IDeAI project** (GA No. 602552) embedded in DH's EU Horizon 2020 fundings from **EJP RD** (GA No. 825575) and **ERICA** (GA No 964908)



© Ralf-Dieter Hilgers, RWTH Aachen University



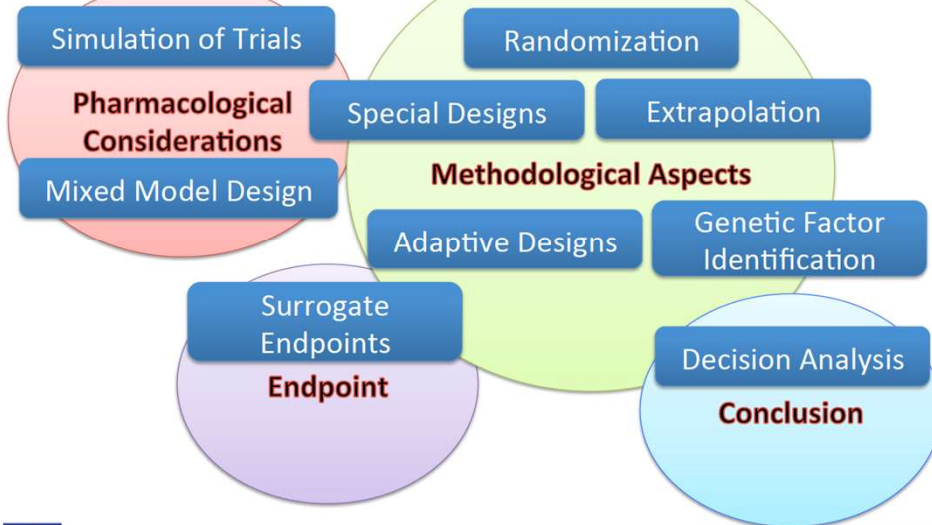
Gliederung

- Rahmenbedingungen
 - IDEAL / EJP-RD
- Randomisierung in Klinischen Studien
- Fill-it Up Design
- Zusammenfassung



Integrated DEsign and AnaLysis of small population group trials

2013-2017, FP7 EU no 602552



- etwa 100 begutachtete Artikel (www.ideal.rwth-aachen.de)
- IRDiRC-Arbeitsgruppe zu klinischen Studien mit kleinen Populationen
- Teilnahme am EMA-Workshop zu klinischen Studien mit kleinen Populationen (März 2017)
- Fortsetzung in EJP-RD (<https://www.ejprardiseases.org>)

Hilgers et al. *Orphanet Journal of Rare Diseases* (2018) 13:77
<https://doi.org/10.1186/s13023-018-0820-8>

Orphanet Journal of Rare Diseases

REVIEW

Open Access



Lessons learned from IDeAI — 33 recommendations from the IDeAI-net about design and analysis of small population clinical trials

- **Wissenschaftlicher Nachweis (Paper)**
- **Software**
- **Webinare**
- **Beispiele (in EJP-RD)**



EUROPEAN JOINT PROGRAMME
RARE DISEASES



© Ralf-Dieter Hilgers, RWTH Aachen University



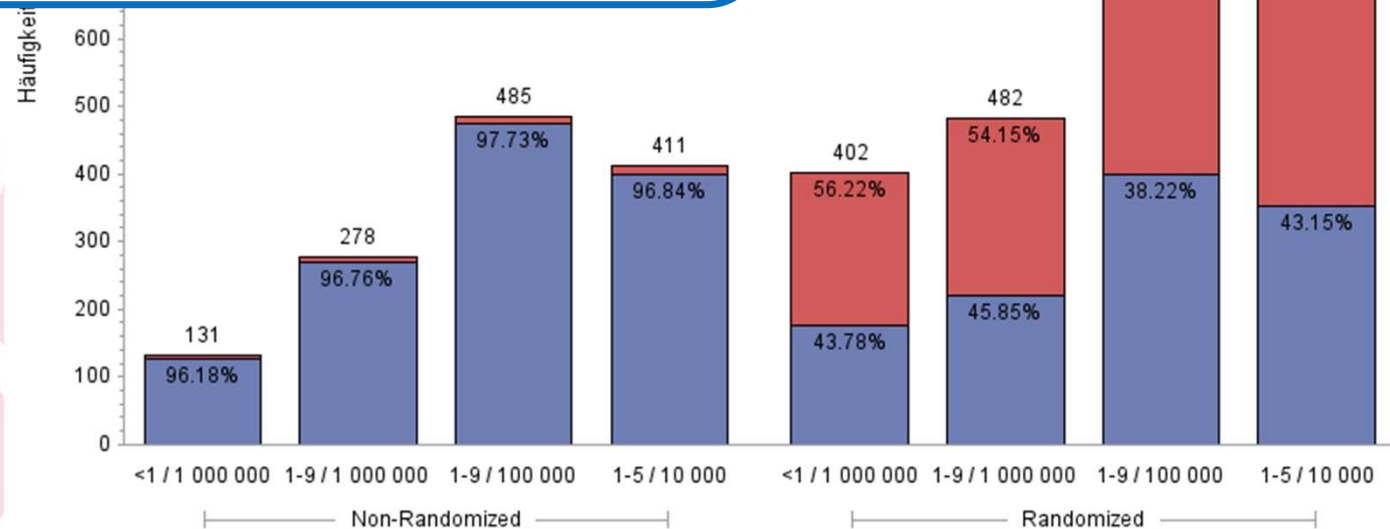
Klinische Studien in Rare Diseases

- im Allgemeinen ein hoher Standard bei der Methodik klinischer Prüfungen
- mit "gewöhnlichen" RDs keine spezifischen methodischen Probleme:
 - *Geographische Verteilung*
 - *Anzahl der Patienten (Diagnostik)*
 - *Heterogenität*
 - *behandelnde Kliniken*

Methodik klinischer Studien im Bereich Rare Diseases :

(Clinicaltrials.gov & Orphanet)

Wied 2023



Randomisierung – Impact auf den Evidenzlevel

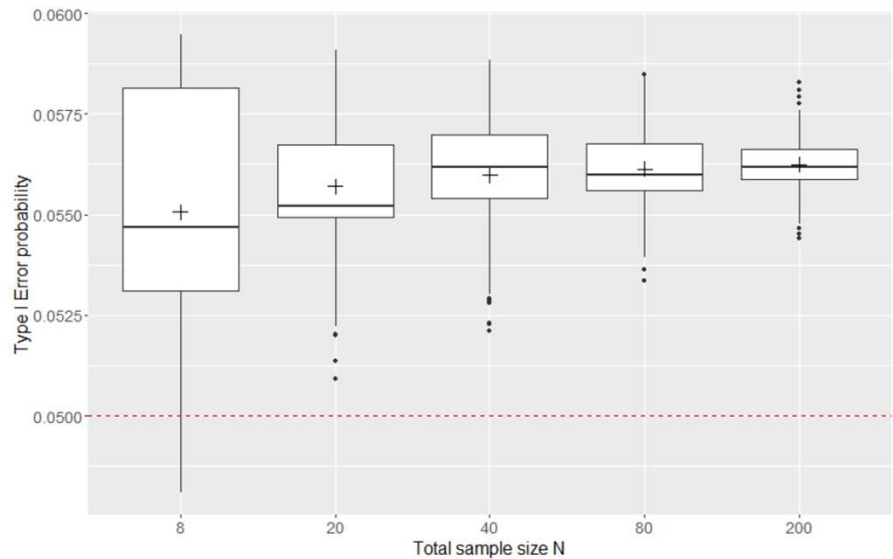
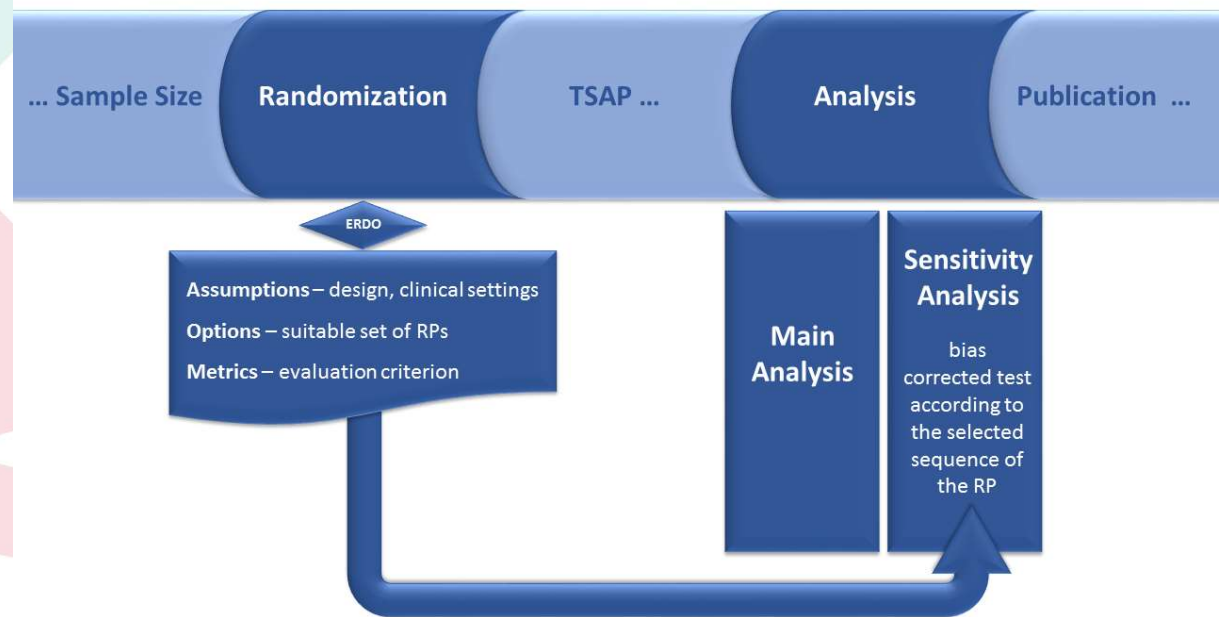


Figure 1: Quantification of Selection Bias depending on the total sample size for PBR(4) for Selection Bias effect of 10% of the treatment effect.

2. Vermeide so weit wie möglich irgendwelche Restriktionen, wie Stratifizierung, finale Balance im Randomisierungsprozess



1. Beachte die Wahl des Randomisierungsverfahrens insbesondere bei SPG Studien



ERDO Framework

Quantifizierung der Auswirkungen von Bias auf die Typ I Fehlerwahrscheinlichkeit für verschiedene Designs (Layout, Art des primären Endpunkts)

	Parallel group design	Endpoint	Hypothesis	Test	Selection Bias	Chronological Bias
Hilgers et al. [8] (2017)	2-arm single center	continuous	$H_0 : \mu_E = \mu_C$	t -Test	additiv	stepwise, linear, logarithmic
Hilgers et al. [9] (2019)	2-arm multicenter	continuous	$H_0 : \mu_E = \mu_C$	t -Test	additiv	stepwise, linear, logarithmic
Uschner et al. [10] (2018)	multi-arm single center	continuous	$H_0 : \mu = 0$	ANOVA	additiv	-
Rückbeil et al. [11] (2017)	2-arm single center	time to event	$H_0 : \lambda_E / \lambda_C = 1$ ($H_0 : S_E = S_C$)	F -/Logrank-Test	multiplicative	stepwise, linear, logarithmic

Table 2: Overview of already evaluated study designs with respect to randomization procedures and bias.

- **Software: randomizeR**
- **Situationen: multiple endpoints (S. Schoenen iSTORE), group sequential (Bodden, Muellenmeister),**
- **In der Entwicklung: crossover trials (Hilgers), allocation ratio (Hilgers) und response adaptive randomization (Ihl), binary endpoints (Reugels), LMEM, platform trials,**

Fazit 1: Randomisierung

- **Auswahl des Randomisierungsverfahrens** kommt gerade in kleinen Studien eine große Bedeutung zu.
- Vorschlag: **Auswahl analog zur Fallzahlplanung** in concealed Report
- Sind **bisherige Allokationen bekannt**, resultiert Bias. (Verblindung hilft!)
- Bei der Wahl des Randomisierungsverfahrens sollte auf die **Auswirkung auf den Evidenzlevel** geachtet werden.
- Das Biasmodell kann zur Planung und Analyse verwendet werden, d.h. um ein **Bias-korrigiertes Testergebnis** zu erzielen.

RESEARCH

Does Randomization matter in Rare Disease Clinical Trials

Stephanie Wied^{1*}, Denis Kirilov Razsolkov¹, Nicole Heussen^{1,2} and Ralf-Dieter Hilgers¹

Unbalanzierte Studie – Single Arm Studien

- By fostering the use of alternative innovative designs for randomised clinical trials, patients will have a higher probability of being assigned to active treatment, whether in Phase 2 and/or in Phase 3 registrational trials (especially critical for rare paediatric genetic diseases where the window of therapeutic intervention may be relatively narrow).

Extremfall:

- benchmark new clinical trial designs (i.e. basket, platform CTs, shared control arm trials between different sponsors...) that should be assessed and compared to the existing 'gold standard' CT model for rare diseases (i.e. single arm);



1 17 April 2023
2 EMA/CHMP/564424/2021
3 Committee for Medicinal Products for Human Use (CHMP)

- 4 Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation
- 5
- 6
- 7 Considerations on evidence from single-arm trials



© Ralf-Dieter Hilgers, RWTH Aachen University

IHI Call 4 topic 3



Sind Single Arm Studien der Gold Standard?

RESEARCH

Open Access



Randomized controlled trial data for successful new drug application for rare diseases in the United States

RCT, or SAT (Table 1). Approximately two-thirds of orphan drugs (151 of 233) had a pivotal RCT, and one-third (82 of 233) had a pivotal SAT in the clinical data package. Most drugs (190 of 233) were first approved

Conclusions Our results indicated that the presence or absence of RCT data in the clinical data package for successful new drug application in the US was associated with three factors: severity of disease outcome, type of drug usage, and type of primary endpoint. These results highlight the importance of selecting target diseases and potential efficacy variables to optimize orphan drug development.

Table 2 Results of univariable and multivariable logistic regression analyses

Factor	Single-arm trial (n=82)	Randomized controlled trial (n=151)	Total (n=233)	Univariable logistic regression analysis			Multivariable logistic regression analysis		
				Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Prevalence									
< 1/100,000	13 (15.9%)	26 (17.2%)	39	Reference			Reference		
≥ 1/100,000	69 (84.1%)	125 (82.8%)	194	0.91	0.44–1.88	0.7899	1.31	0.52–3.34	0.561
Severity of the disease outcome									
High mortality	67 (81.7%)	67 (44.4%)	134	Reference			Reference		
Others	15 (18.3%)	84 (55.6%)	99	5.60	2.94–10.68	<0.0001*	5.63	2.64–12.00	<0.0001*
Drug usage									
Mono therapy	78 (95.1%)	128 (84.8%)	206	Reference			Reference		
Combination therapy	4 (4.9%)	23 (15.2%)	27	3.50	1.17–10.51	0.0253*	2.95	1.80–18.57	0.0032*
Target age segment									
Children with or without adult	24 (29.3%)	55 (36.4%)	79	Reference			Reference		
Adult only	58 (70.7%)	96 (63.6%)	154	0.72	0.40–1.29	0.2714	1.72	0.75–3.97	0.2027
Primary efficacy endpoint									
Pharmacodynamic/Response biomarker	69 (84.1%)	67 (44.4%)	136	Reference			Reference		
Clinical outcome	13 (15.9%)	84 (55.6%)	97	6.65	3.39–13.06	<0.0001*	5.57	2.57–12.06	0.0001*
Therapeutic classification									
Oncology	63 (76.8%)	46 (30.5%)	109	Reference					
Nononcology	19 (23.2%)	105 (69.5%)	124	7.57	4.08–14.06	<0.0001*			N/A

*P value < 0.05

Verwendung Historischer Kontrollen

E10, a hybrid approach of using external control data to add to a concurrent randomized control arm in a clinical trial may sometimes be useful.

FDA, 2019

1. The **standard treatment** has to be **precisely defined** and must be the same treatment for randomized controls.
2. The **historical control group** must have been part of a **clinical study** with the same requirements for patient eligibility.
3. The **methods of treatment evaluation** have to be the same.
4. **Patient characteristics** have to be comparable.
5. The study must have been **performed in the same organization** with the same investigators.
6. There should be **no indications** leading one to **expect a difference**.

Rare Diseases: Natural History Studies for Drug Development Guidance for Industry

Pocock's (1976) Kriterien

Methoden:

- Bennett (2021) Bayesianische Methoden (Power Prior),
- Schmidli (2014) Meta-Analytic-Predictive Priors
- Viele (2014): Pool than Test Approach

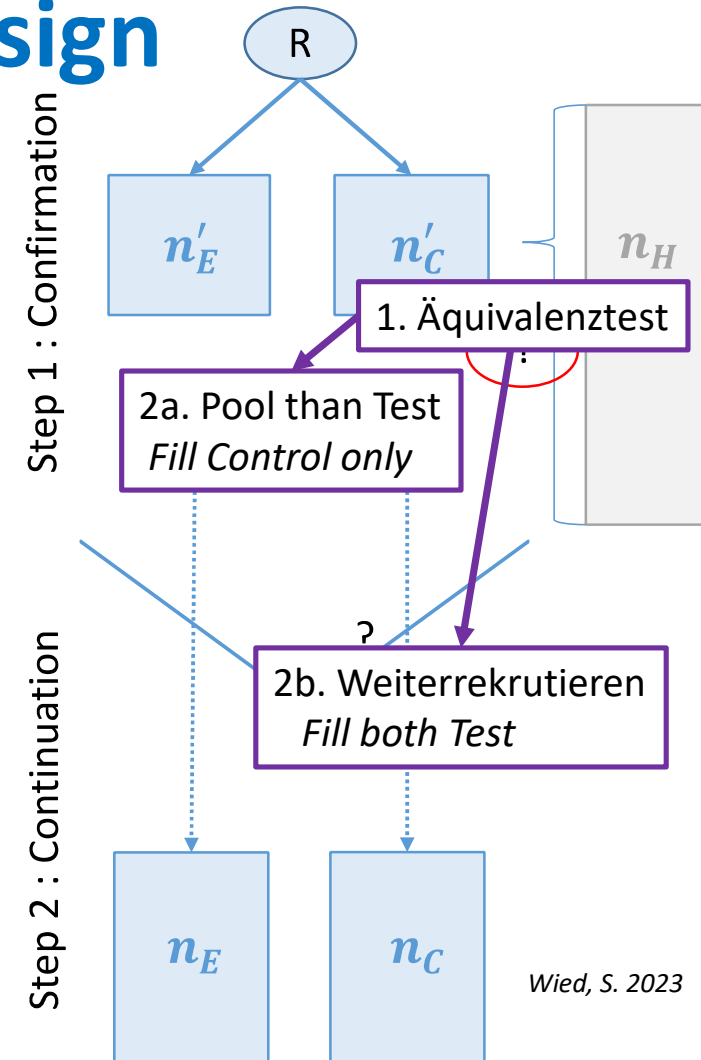
Historical Control– Fill-it-Up Design

Motivation

- Überlegenheitsnachweis von E (experimental) gegenüber C (control)
- Reduziere die erforderliche RCT Fallzahl, indem (eine große Anzahl) Historische Kontrollen verwendet werden (unbalanced Design um die Powerforderung zu gewährleisten).

Ziel

- Rechtfertige die Verwendung der historischen Kontrollen, durch Äquivalenznachweis von historischen and randomisierte Kontrollen
- Kombiniere historische and randomisierte Kontrollen
- Führe Überlegenheitsnachweise durch



Powergewinn (Fill-it-up versus randomized)

Einseitiger Test

Table 1: Overall power and sample sizes for different scenarios of the Fill-it-up-design depending on the choice of the equivalence margin Δ considering a medium effect size $\delta = 0.5$ when $n_H = 500$ are available and $\alpha_{S_1} = \alpha_{S_2} = 0.05$.

α_{Ept}	Δ	N_{FIU}	γN_{FIU}	AVN	$1 - \beta_{S_1}$ $= 1 - \beta_{S_2}$	$1 - \beta_{Ept, S_1}$	$1 - \beta_{Ept^c, S_2}$	$1 - \beta_{FIU}$
0.01	0.4596	100	54	100	0.80	0.9530	0.8471	0.8001
0.01	0.4798	100	54	100	0.80	0.9871	0.8130	0.8001
0.01	0.5000	116	62	116	0.85	0.9999	0.8001	0.8000
0.05	0.3250	102	54	100	0.81	0.9596	0.8403	0.8000
0.05	0.3901	138	74	136	0.90	0.9608	0.8392	0.8000
0.05	0.5000	112	60	110	0.84	0.8175	0.9825	0.8000
0.10	0.2303	124	66	120	0.87	0.9637	0.8368	0.8005
0.10	0.3652	124	66	120	0.87	0.8442	0.9558	0.8000
0.10	0.5000	100	54	96	0.80	0.9889	0.8113	0.8002
0.20	0.1663	102	54	94	0.81	0.9722	0.8279	0.8001
0.20	0.3331	102	54	94	0.81	0.9403	0.8599	0.8002
0.20	0.5000	132	72	120	0.89	0.8397	0.9607	0.8003

Gesamtstich-
probenumfang
für ein Design
ohne histori-
sche Kontrollen

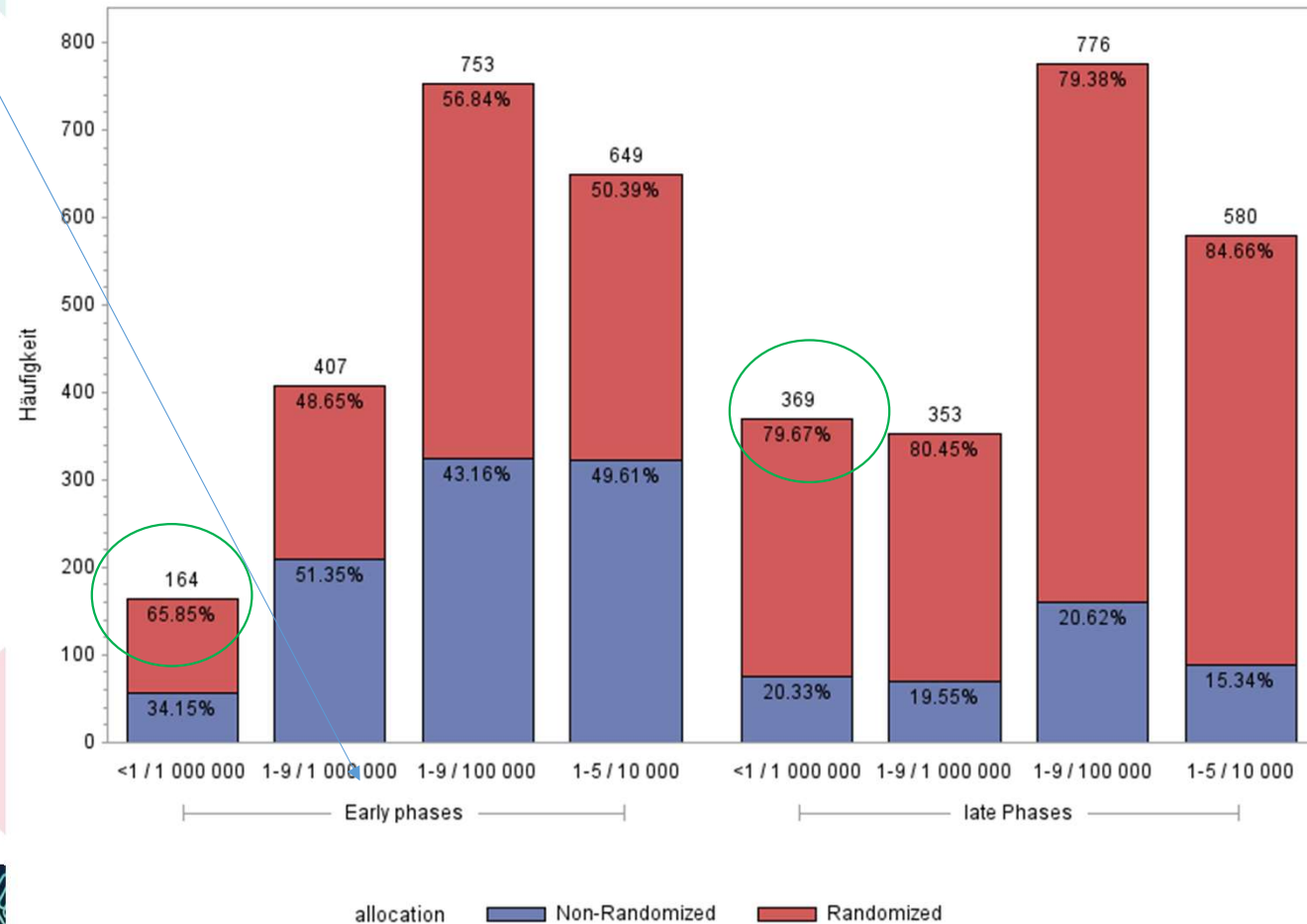
Stichproben-
umfang zum
Zeitpunkt des
Äquivalenztests

Fazit 2: Verwendung Historischer Kontrollen

- **Hybrider Ansatz** entsprechend der FDA Guideline jedoch unter Berücksichtigung des entstehenden **Multiplen Testproblems**.
 - letzter Test prüft **Non-Inferiority Hypothese!**
- Mindestens **50% des initialen Gesamtstichprobenumfangs** muß im ersten Schritt eingeschlossen werden.
- **Powergewinn ?**
- *Fazit: in der Situation, wenn man wenige Patienten rekrutieren kann, kann das Verfahren Power gewinnen, wenn man vorab gut geeignete Kontrollen (etwa aus einem Register) identifiziert hat.*

Randomisierte Studien in frühen Phasen

- Hier ist das Power-Requirement nicht bindend



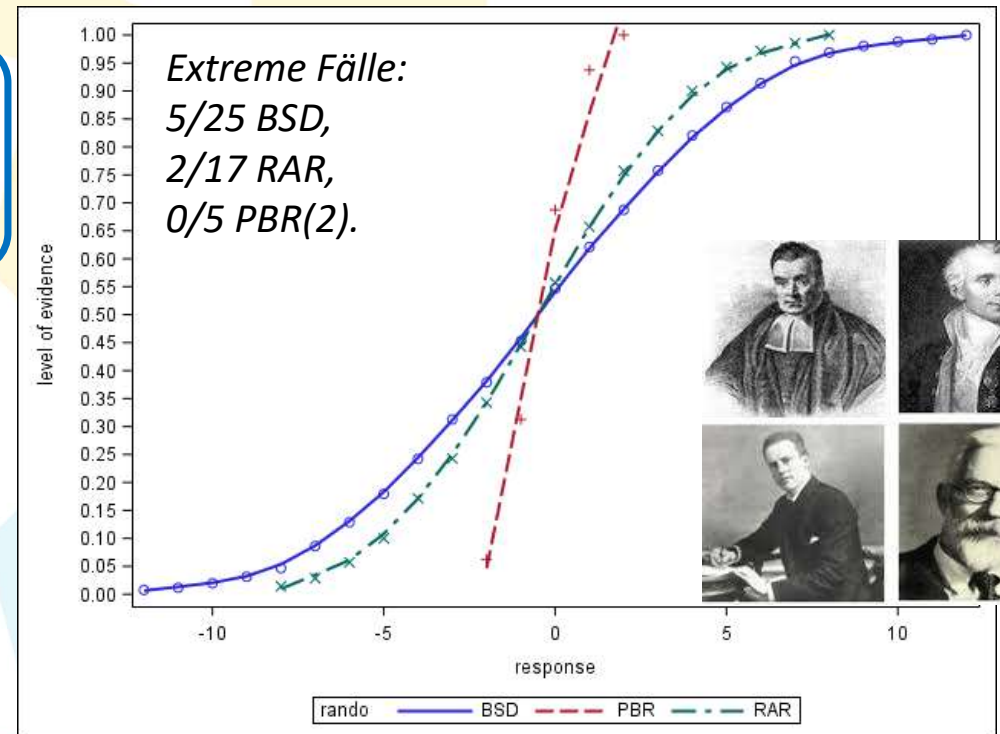
Clinical trials DB mit Orphanet DB

Wied 2023

Design Based Inference in finiten Populationen

Es stellt sich die Frage, sind die Ergebnisse mit dem Zufall vereinbar?
Welches Evidenzlevel kann man zeigen?

- $N_{\text{total}} = 8$ Zulassungsstudie für Increlex als Long-term Behandlung von Kindern und Erwachsenen mit schwerem primären Insulinähnlichem growth factor-1 deficiency (Joppi, 2013)
- Finite Population: Zufallsziehungsmodell macht keinen Sinn als Basis für die Inferenz
- *Design Based Analysis*



Schlußfolgerung für Design Based Analysis basierend auf dem *Randomization Test*
Bewerte den Evidenzlevel mit **randomization based level of inference** (ist möglich)



Fazit 3: Design Based Analysis

- Auch in **kleinen Studien**, etwa Phase I und Phase II Studien hat **Randomisierung** bereits seinen Platz
- **kein konfirmatorischer Beweis**
- Es stellt sich die Frage, sind die **Ergebnisse mit dem Zufall vereinbar**, oder welchen Evidenzlevel kann man zeigen?
- Ein **neuer** bisher **noch nicht verstandener Weg** zur Generierung der Evidenz.....

Zusammenfassung

- Randomisierung ist auch in Studien mit seltenen Erkrankungen mit kleinen Fallzahlen möglich
- Der Auswahl des Randomisierungsverfahrens kommt eine besondere Bedeutung zu.
- Es steht ein Tool zur Quantifizierung des Effektes von Bias auf die Studienergebnisse zur Verfügung.
- Fill-it-up Design ist ein hybrider Ansatz zur Verwendung Historischer Kontrollen
- Design Based Inferenz ist ein anderer Ansatz zur Quantifizierung der Evidenz

-> ERDERA // IHI

Coming together
is a beginning;
keeping together
is progress;
working together
is success.

Henry Ford

Backup



Coordination & transversal activities
(strategy, sustainability, ethics, communication)

Accelerating of research translation & clinical studies

Capacity building & empowerment



RD research funding

Access to data, tools & services – Virtual Platform

WP leaders: Rima NABBOUT, Ralf-Dieter HILGERS

WP20: Beschleunigung der Validierung, des Einsatzes und der Entwicklung innovativer Methoden, die auf klinische Prüfungen bei RDs zugeschnitten sind (3Ds)

Verbesserung der Methoden für klinische Studien über seltene Krankheiten durch Validierung und Demonstration theoretischer Methoden mit realen Daten aus ERNs

Ermittlung der wichtigsten Lücken bei klinischen Studien im Bereich der FTE und Vorschlag eines Fahrplans zur Priorisierung dieser Lücken, um schließlich Projekte zu innovativen Methoden einzuleiten

Verbesserung der Verbreitung erfolgreicher Methoden für klinische Studien im Hinblick auf die klinische Umsetzung durch aktive Unterstützung bei der Methodik und Umsetzung

Demonstrate

Develop/Innovate

Disseminate

Standard – Unbalanziertes Parallelgruppen Design

Two group t-test of equal means (unequal n's)

	1	2	3	4	5	6	7	8	9	10	11
Test significance level, α	0,050	0,050	0,050	0,050	0,050	0,050	0,050	0,050	0,050	0,050	
1 or 2 sided test?	1	1	1	1	1	1	1	1	1	1	
Group 1 mean, μ_1											
Group 2 mean, μ_2											
Difference in means, $\mu_1 - \mu_2$											
Common standard deviation, σ											
Effect size, $\delta = \mu_1 - \mu_2 / \sigma$	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	
Power (%)	80	80	80	80	80	80	80	80	80	80	
n_1	51	38	28	26	26	25	25	25	25	25	
n_2	51	76	280	1300	1950	2500	3125	3750	12500	2500	
Ratio: n_2 / n_1	1,000	2,000	10,000	50,000	75,000	100,000	125,000	150,000	500,000	100,000	
$N = n_1 + n_2$	102	114	308	1326	1976	2525	3150	3775	12525	2525	

Randomisierungsverfahren

- CR** (*Vollständige Randomisierung*) Wahrscheinlichkeit für jeden Patienten eine von zwei Behandlungen zu erhalten ist $\frac{1}{2}$
- BSD(a)** (*Big Stick design*) CR wobei die maximale Imbalance der Zuweisungen kleiner oder gleich A sein muß.
- RAR** (*Random Allocation Rule*) jeweils die Hälfte der Patienten erhält eine von zwei Behandlungen
- PBR(m_s)** (*Permuted Block Randomization*) Implementierung von RAR in k Blöcken der Länge m_s ; $1 \leq s \leq k$
- MP(a)** (*Maximal Procedure*) erlaubt die maximale Imbalance der Zuweisungen kleiner oder gleich A , forciert terminal balance und erzwingt eine Gleichverteilung der resultierenden Sequenzen.

Extrapolation of Treatment effect



www.ideal.rwth-aachen.de

- G. Hlavin, et al. [Evidence, eminence and extrapolation](#). 2016
- Dette, H. et al. [Equivalence of regression curves](#). 2018.
- Dette, H. et al. [Optimal designs for comparing regression models with correlated observations](#). 2016.
- Dette, H. et al. [Optimal Designs For Comparing Curves](#). 2016.
- Schorning, K. et al. [Model Selection versus Model Averaging in Dose Finding Studies](#). 2016.

Dose
Response
Profiles

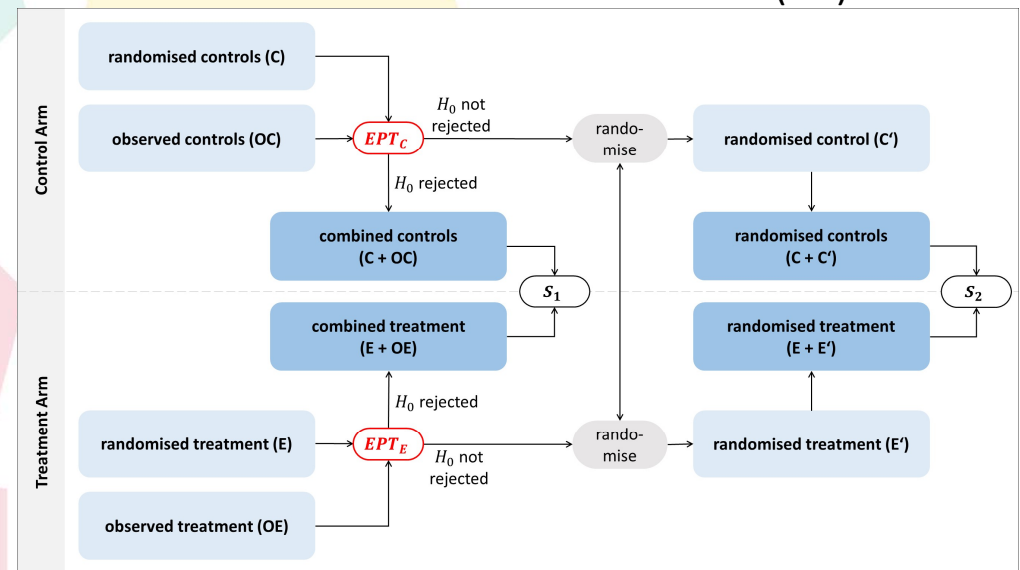
Motivation

- established treatment effect of E (experimental) versus C (control) in similar larger population
- *transfer of treatment effect possible?*
- *conduct small RCT in target population*

Procedure

- combine historical and randomized controls
- equivalence between controls
- equivalence between experimental groups
- equivalence between treatment effects
- all or none multiple testing

Historical data and Randomized data (HR)



Wied, 2023

Randomisierte Studien in frühen Phasen

		PrevalenceClass_name											
		1-5/10.000 (m = 2823)		1-9/100.000 (m = 3120)		1-9/1.000.000 (m = 2024)		<1/1.000.000 (m = 816)		Unknown (m = 19)		Gesamt (m = 8802)	
		m	%	m	%	m	%	m	%	m	%	m	%
Phases	Allocation												
keine Angabe	keine Angabe	558	6.34	650	7.38	254	2.89	94	1.07	11	0.12	1567	17.80
N/A	keine Angabe	18	0.20	7	0.08	12	0.14	5	0.06	0	0.00	42	0.48
	N/A	144	1.64	140	1.59	83	0.94	14	0.16	0	0.00	381	4.33
	Non-Randomized	42	0.48	56	0.64	32	0.36	12	0.14	0	0.00	142	1.61
	Randomized	189	2.15	157	1.78	62	0.70	37	0.42	0	0.00	445	5.06
Early Phase 1	keine Angabe	1	0.01	0	0.00	1	0.01	1	0.01	0	0.00	3	0.03
	N/A	11	0.12	10	0.11	6	0.07	1	0.01	0	0.00	28	0.32
	Non-Randomized	3	0.03	4	0.05	1	0.1	0	0.00	0	0.00	8	0.09
	Randomized	4	0.05	8	0.09	3	0.03	1	0.01	0	0.00	16	0.18
Phase 1	keine Angabe	90	1.02	37	0.42	120	1.36	8	0.09	0	0.00	255	2.90
	N/A	236	2.68	202	2.29	259	2.94	43	0.49	0	0.00	740	8.41
	Non-Randomized	145	1.65	145	1.65	83	0.94	12	0.14	1	0.01	386	4.39
	Randomized	70	0.80	111	1.26	49	0.56	33	0.37	0	0.00	263	2.99
Phase 1 Phase 2	keine Angabe	17	0.19	15	0.17	22	0.25	1	0.01	0	0.00	55	0.62
	N/A	76	0.86	71	0.81	63	0.72	10	0.11	0	0.00	220	2.50
	Non-Randomized	62	0.70	63	0.72	28	0.32	6	0.07	0	0.00	159	1.81
	Randomized	50	0.57	35	0.40	20	0.23	7	0.08	0	0.00	112	1.27
Phase 2	keine Angabe	78	0.89	45	0.51	99	1.12	14	0.16	0	0.00	236	2.68
	N/A	269	3.06	270	3.07	271	3.08	42	0.48	2	0.02	854	9.70
	Non-Randomized	115	1.31	117	1.33	98	1.11	38	0.43	2	0.02	370	4.20
	Randomized	207	2.35	282	3.20	129	1.47	68	0.77	0	0.00	686	7.79