



Kolloquium „Statistische Methoden in der empirischen Forschung“

Verblindete Fallzahlneuschätzung in einer konfirmatorischen Diagnosegütestudie

Maria Stark

7.02.2023

Gefördert durch **DFG** (ZA 687/1-1)

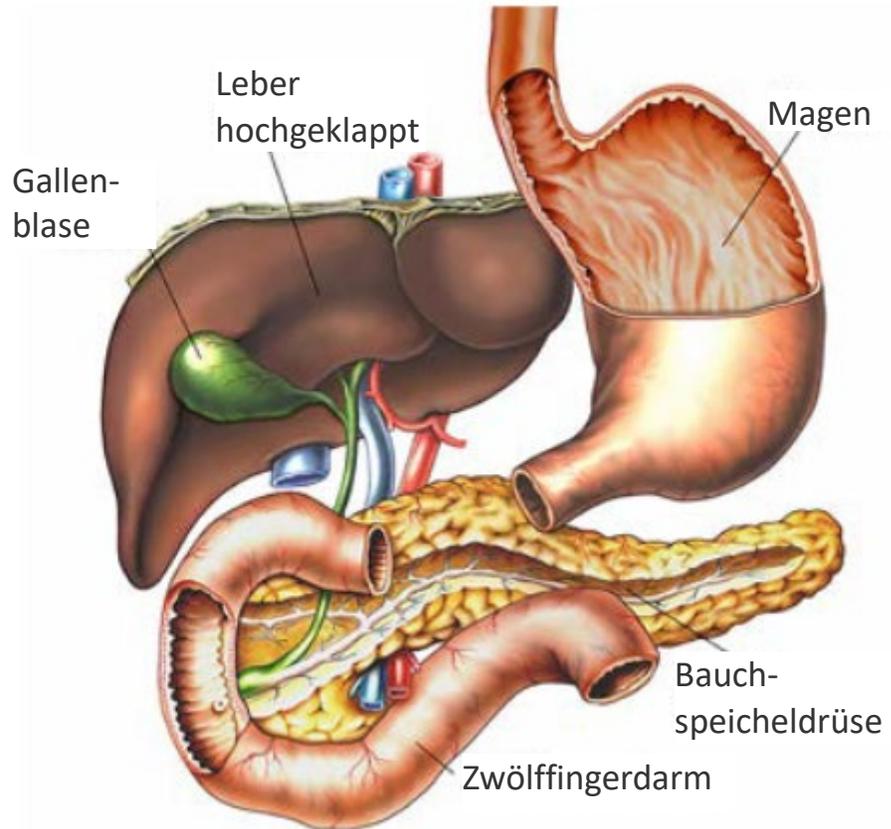


Universitätsklinikum Hamburg-Eppendorf





Beispielstudie



Diagnose:

Karzinom Bauchspeicheldrüse



Experimenteller Test:

Positronen Emissions Tomographie +
Computer Tomographie (PET/CT)



Komparatortest:

Computer Tomographie (CT)



Referenzstandard:

Biopsie

Konfirmatorische
Diagnosestudie

Mit passender Fallzahl

Mit passender Power

WOFÜR?

WAS?

WIE?

WARUM?

Fallzahlplanung

1

Konfirmatorische
Diagnosestudie

2

Anpassung der
Fallzahl im Laufe der
Studie

Fallzahl kann zu hoch oder zu niedrig sein

1 Optimale Fallzahlplanung

2 Verblindetes adaptives Design zur
Fallzahneuschätzung



1 Fallzahl ist Maximum aus beiden Endpunkten

2 Unsichere Annahmen für Fallzahlplanung

Entwicklung eines diagnostischen Tests

Phase I



Phase II



Phase III



Phase IV



Zulassung

Ziel

Technische Evaluation

Diagnosegüte bei bekanntem Gesundheitszustand

Diagnosegüte bei unbekanntem Gesundheitszustand

Konsequenzen der Diagnose für die Behandlung

Endpunkte

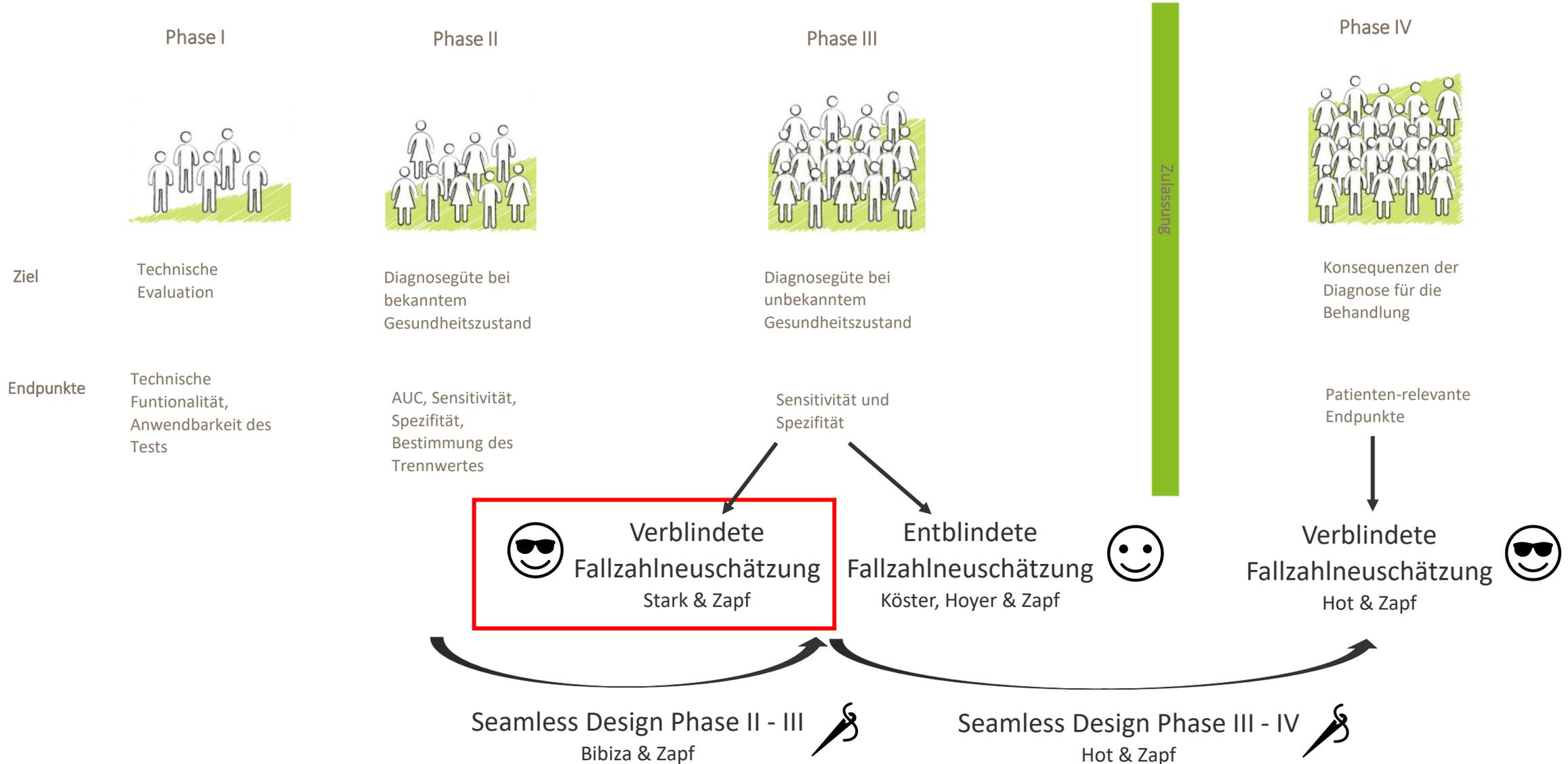
Technische Funktionalität, Anwendbarkeit des Tests

AUC, Sensitivität, Spezifität, Bestimmung des Trennwertes

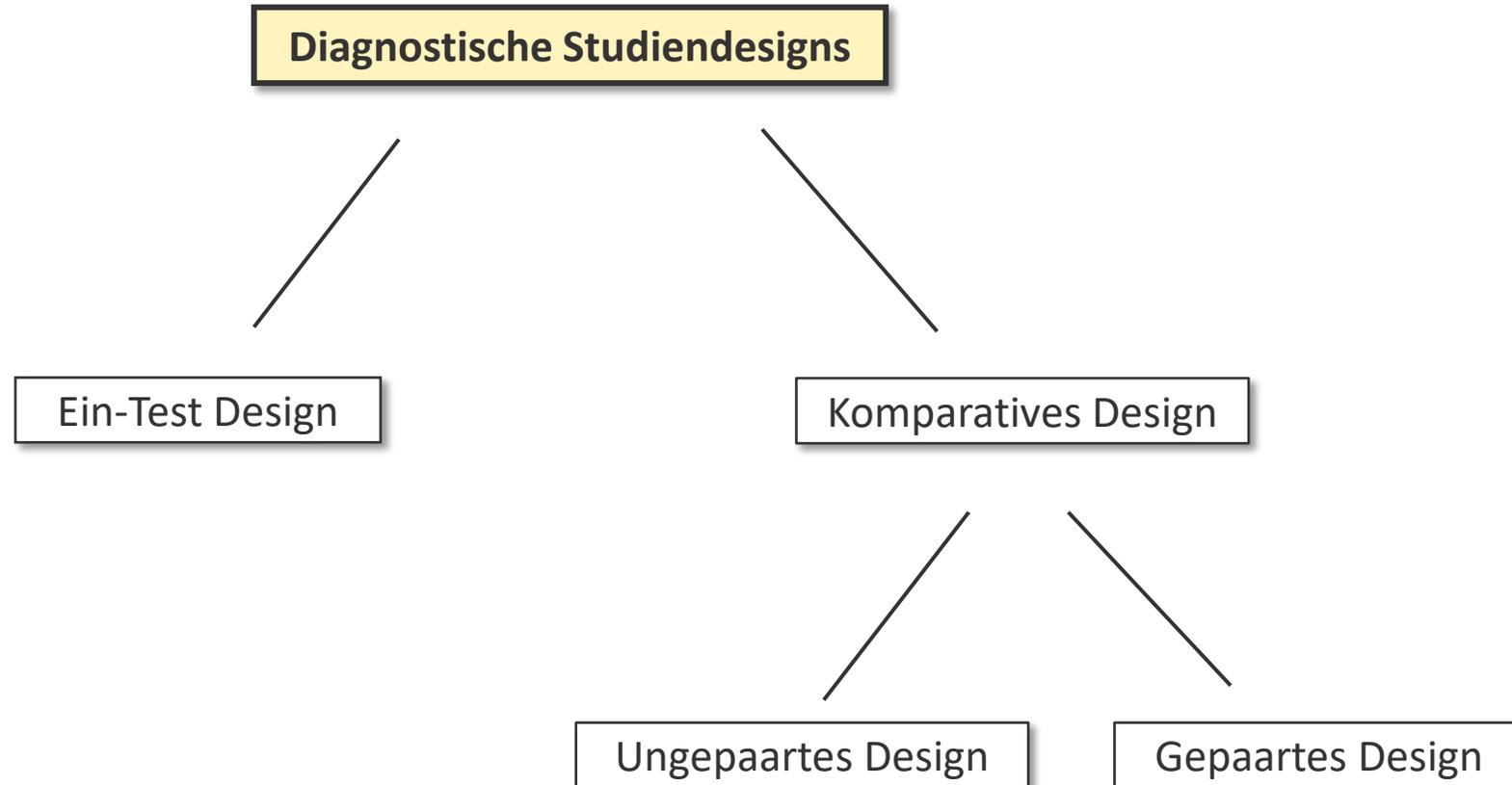
Sensitivität und Spezifität

Patienten-relevante Endpunkte

Entwicklung eines diagnostischen Tests



Diagnostische Studiendesigns

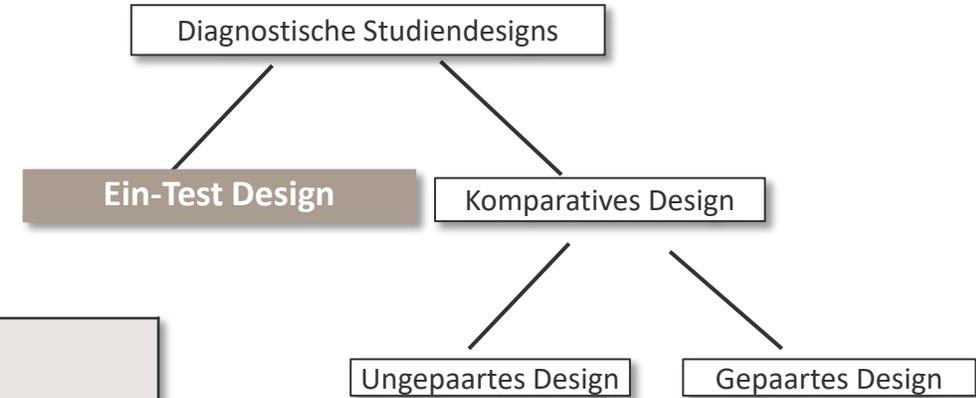


Ein-Test Design

		Referenzstandard	
		Krank (n_D)	Gesund (n_{ND})
Experimenteller Test	Positiv	Richtig Positiv	Falsch Positiv
	Negativ	Falsch Negativ	Richtig Negativ

Sensitivität Se_E
 $= RP/n_D$

Spezifität Sp_E
 $= RN/n_{ND}$

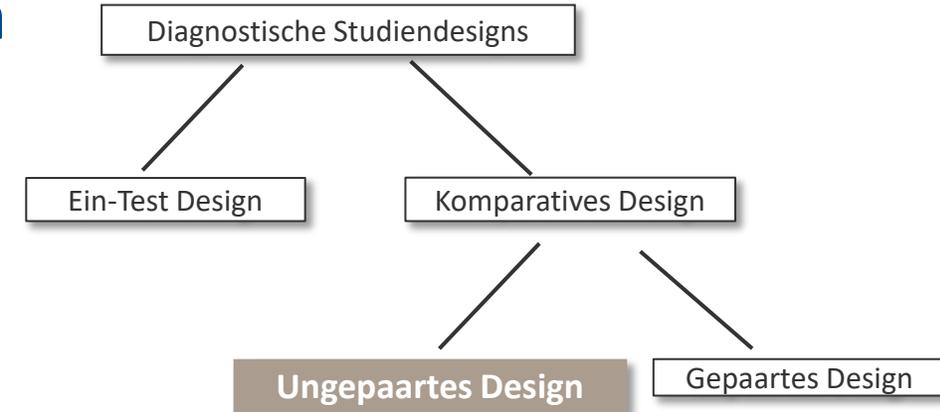


Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Ungepaartes Design



Experimenteller Test

	Referenzstandard		
	Krank (n_{D_E})	Gesund (n_{ND_E})	
Experimenteller Test	Positiv	Richtig Positiv RP_E	Falsch Positiv FP_E
	Negativ	Falsch Negativ FN_E	Richtig Negativ RN_E

Sensitivität Se_E
 $= RP_E / n_{D_E}$

Spezifität Sp_E
 $= RN_E / n_{ND_E}$

Komparatortest

	Referenzstandard		
	Krank (n_{D_C})	Gesund (n_{ND_C})	
Komparator-test	Positiv	Richtig Positiv RP_C	Falsch Positiv FP_C
	Negativ	Falsch Negativ FN_C	Richtig Negativ RN_C

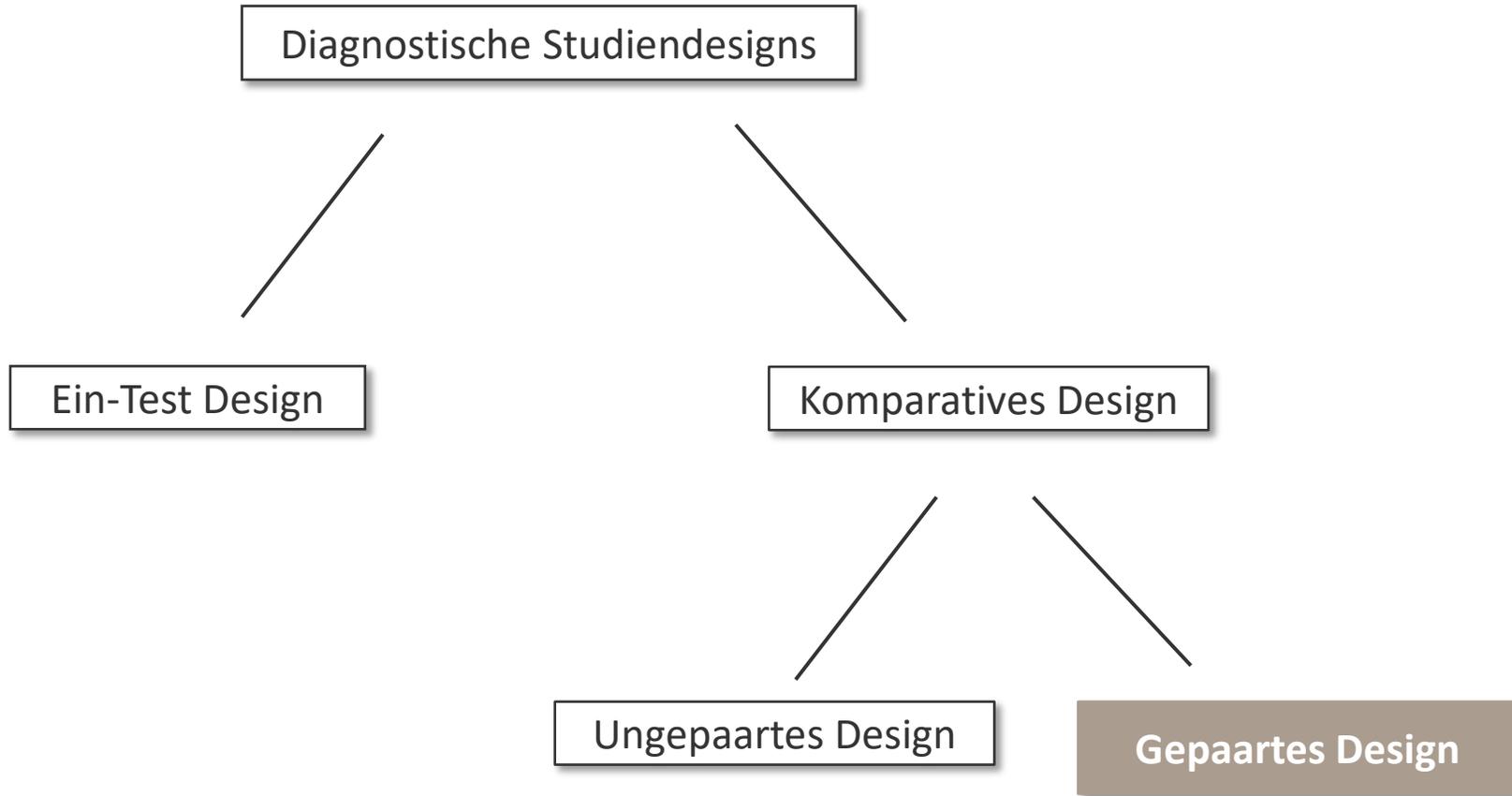
Sensitivität Se_C
 $= RP_C / n_{D_C}$

Spezifität Sp_C
 $= RN_C / n_{ND_C}$

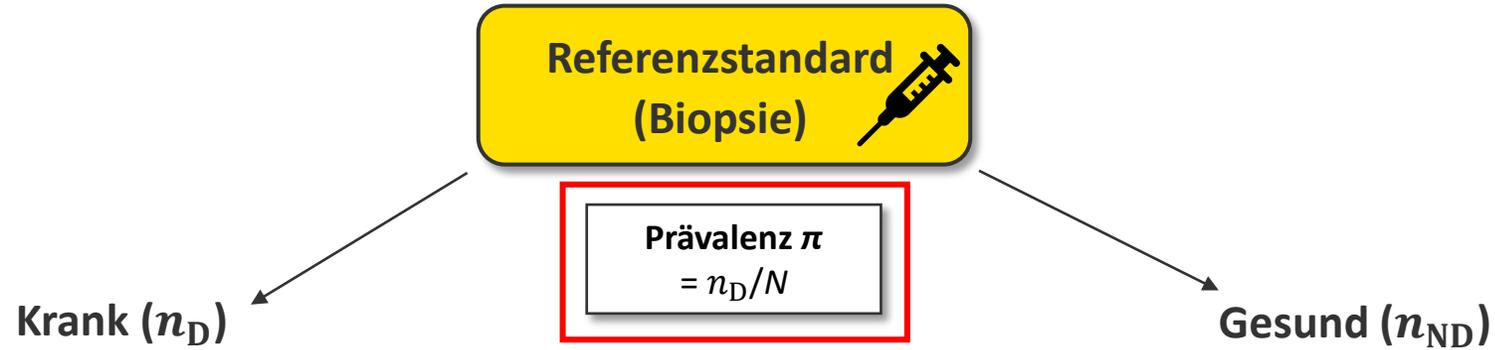
Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde



Gepaartes Design

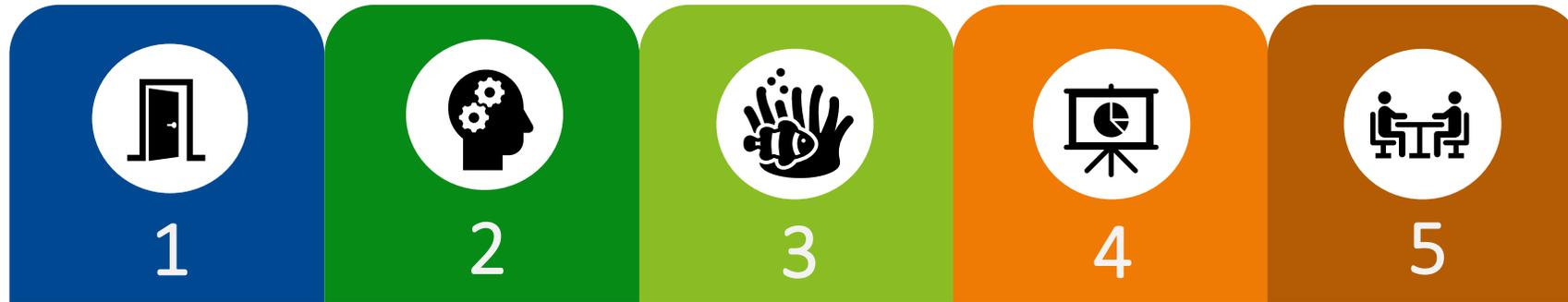


		Komparatortest (CT)				Komparatortest (CT)	
Exp. Test (PET/CT) ⊕	Richtig Positiv RP_E	Richtig Positiv RP_C	Falsch Negativ FN_C	Exp. Test (PET/CT) ⊕	Falsch Positiv FP_E	Richtig Negativ RN_C	
	Falsch Negativ FN_E	ψ_D	--		Richtig Negativ RN_E	ψ_{ND}	--
Sensitivität $Se_E = RP_E/n_D$		Sensitivität $Se_C = RP_C/n_D$		Spezifität $Sp_E = RN_E/n_{ND}$		Spezifität $Sp_C = RN_C/n_{ND}$	

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde



Einleitung

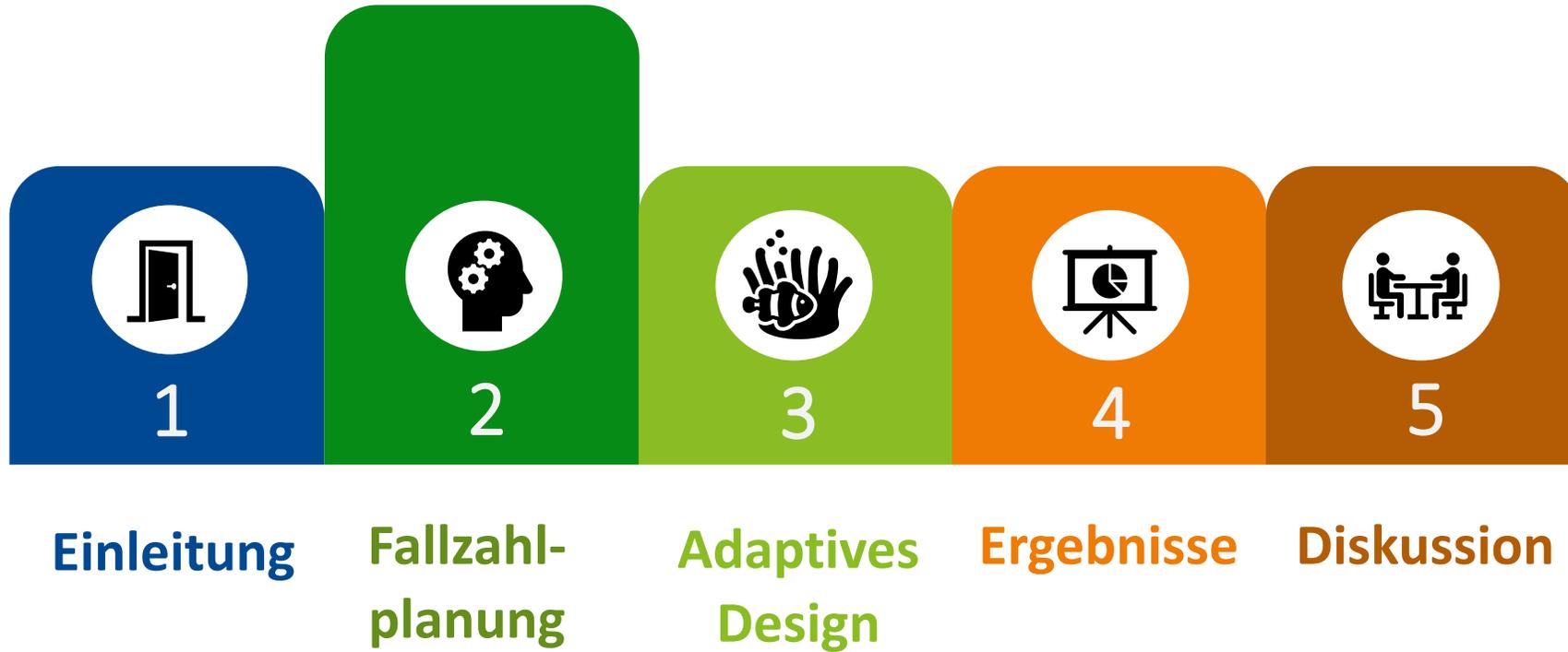
**Fallzahl-
planung**

**Adaptives
Design**

Ergebnisse

Diskussion

 Gepaartes
Design



 Gepaartes
Design

Hypothese gepaartes Design

Sensitivität und Spezifität sind **unabhängige co-primäre Endpunkte**

Intersection-Union-Test:

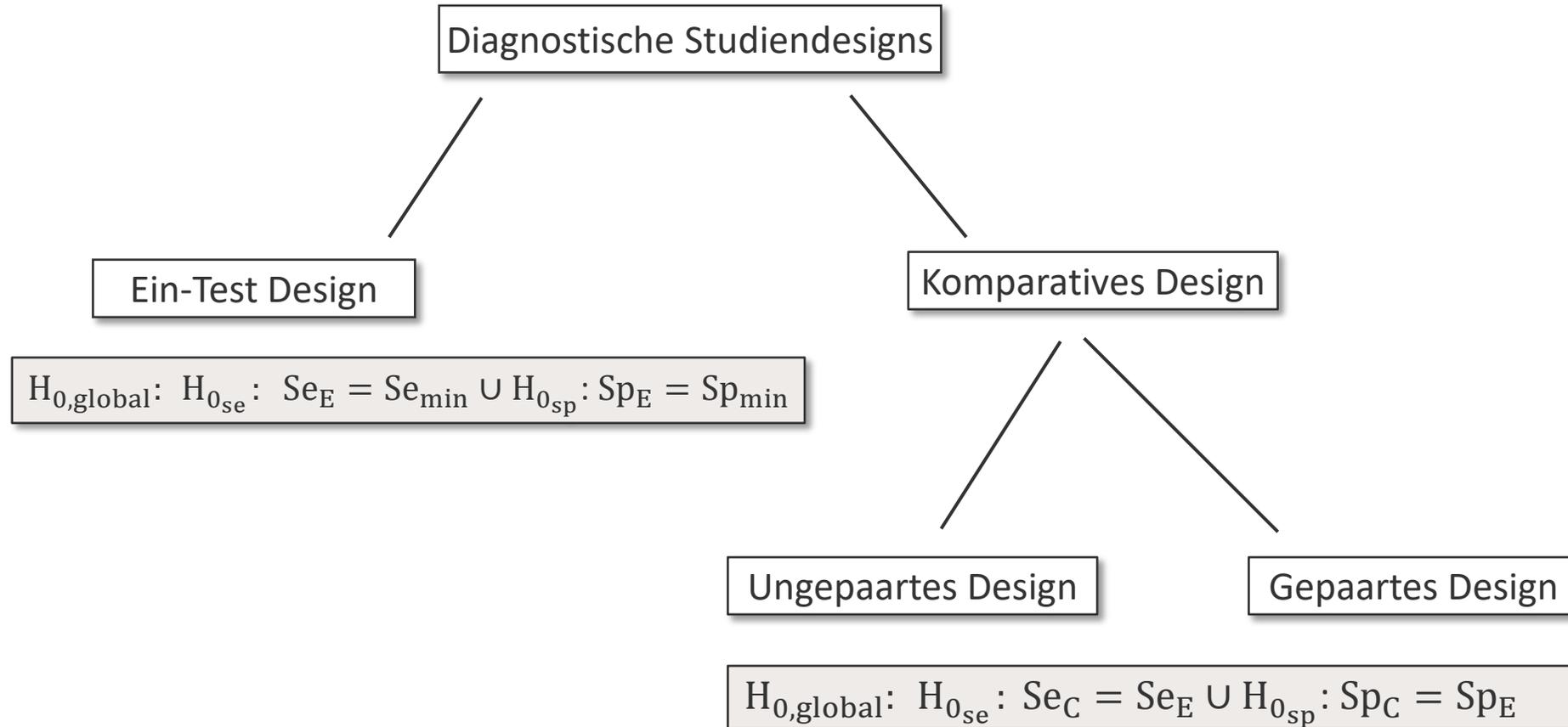
$$H_{0_{Se}}: Se_C = Se_E \text{ and } H_{0_{Sp}}: Sp_C = Sp_E$$

$$H_{0_{global}}: H_{0_{Se}} \cup H_{0_{Sp}}$$

$$\text{Globaler Fehler 1. Art} = \alpha_{Se} \cdot \alpha_{Sp}$$

$$\text{Gesamtpower} = \text{Power}_{Se} \cdot \text{Power}_{Sp}$$

Hypothesen



Stark & Zapf, 2020; Stark et al., 2022

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Konventionelle Fallzahlplanung

$\alpha = 5\%$ (zweiseitig), Gesamtpower = 80% Prävalenz: $\pi = 47\%$	Komparator-test (CT)	Experimenteller Test (PET/CT)	Diskordante Testergebnisse
Erkrankte Population	$Se_C = 81\%$	$Se_E = 90\%$	$\psi_D = 9\%$
Gesunde Population	$Sp_C = 66\%$	$Sp_E = 80\%$	$\psi_{ND} = 14\%$

1

Anzahl erkrankter Personen (Miettinen et al., 1968):

$$n_D = \frac{\left(z_{1-\alpha/2} \cdot \psi_D + z_{1-\beta_{Se}} \sqrt{\psi_D^2 - \frac{1}{4} (Se_C - Se_E)^2 (3 + \psi_D)} \right)^2}{\psi_D (Se_C - Se_E)^2} = 74$$

Anzahl gesunder Personen:

$$n_{ND} = \frac{\left(z_{1-\alpha/2} \cdot \psi_{ND} + z_{1-\beta_{Sp}} \sqrt{\psi_{ND}^2 - \frac{1}{4} (Sp_C - Sp_E)^2 (3 + \psi_{ND})} \right)^2}{\psi_{ND} (Sp_C - Sp_E)^2} = 47$$

McCray et al., 2017; Stark et al., 2022

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparator-test

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Konventionelle Fallzahlplanung

$\alpha = 5\%$ (zweiseitig), Gesamtpower = 80% Prävalenz: $\pi = 47\%$	Komparator-test (CT)	Experimenteller Test (PET/CT)	Diskordante Testergebnisse
Erkrankte Population	$Se_C = 81\%$	$Se_E = 90\%$	$\psi_D = 9\%$
Gesunde Population	$Sp_C = 66\%$	$Sp_E = 80\%$	$\psi_{ND} = 14\%$

1

Anzahl erkrankter Personen:

$$n_D = 74$$

Anzahl gesunder Personen:

$$n_{ND} = 47$$

2

Fallzahl um Sensitivität zu zeigen:

$$N_{Se} = \frac{n_D}{\pi} = \frac{74}{0.47} = 157$$

Fallzahl um Spezifität zu zeigen:

$$N_{Sp} = \frac{n_{ND}}{1 - \pi} = \frac{47}{1 - 0.47} = 88$$

3

Finale Fallzahl:

$$N = \max(N_{Se}, N_{Sp}) = 157$$

Optimale Fallzahlplanung

3 $N_{Se} \stackrel{!}{=} N_{Sp}$

2 $\frac{n_D}{\pi} \stackrel{!}{=} \frac{n_{ND}}{1 - \pi}$



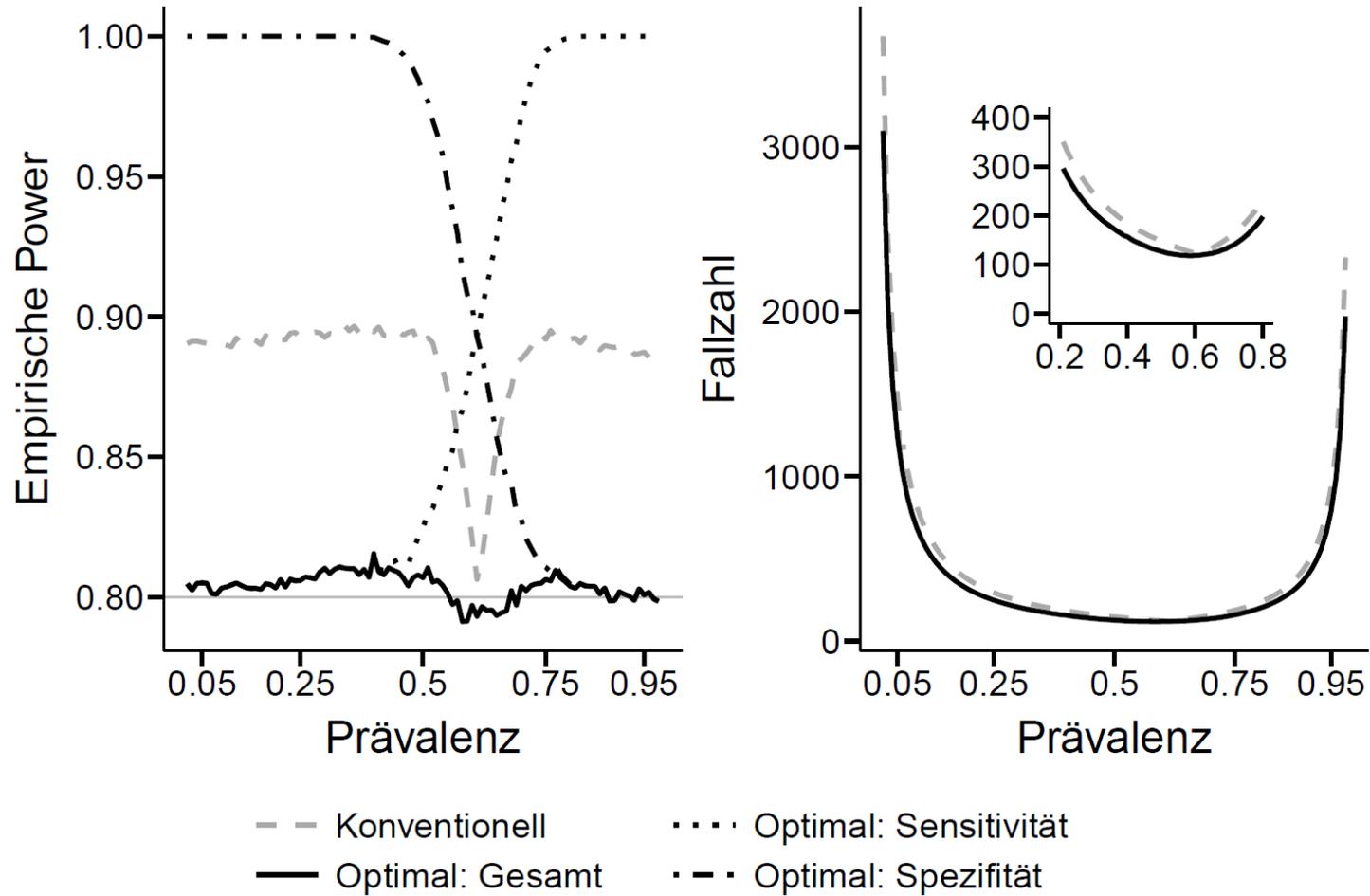
$$\frac{\left(z_{1-\alpha/2} \cdot \psi_D + z_{1-\beta_{Se}} \sqrt{\psi_D^2 - \frac{1}{4} (Se_C - Se_E)^2 (3 + \psi_D)} \right)^2}{\psi_D (Se_C - Se_E)^2 \pi} \stackrel{!}{=} \frac{\left(z_{1-\alpha/2} \cdot \psi_{ND} + z_{1-\beta_{Sp}} \sqrt{\psi_{ND}^2 - \frac{1}{4} (Se_C - Se_E)^2 (3 + \psi_{ND})} \right)^2}{\psi_{ND} (Se_C - Se_E)^2 (1 - \pi)}$$

1

Mit der Bedingung: $Power_{Se} \cdot Power_{Sp} = \text{Gesamtpower} \Leftrightarrow (1 - \beta_{Se}) \cdot (1 - \beta_{Sp}) = \text{Gesamtpower}$

$$\beta_{Sp} = 1 - \frac{\text{Gesamtpower}}{1 - \beta_{Se}}$$

Optimale Fallzahlplanung





Einleitung

**Fallzahl-
planung**

**Adaptives
Design**

Ergebnisse

Diskussion

 Gepaartes
Design

 Optimale
Fallzahl-
planung



Einleitung

**Fallzahl-
planung**

**Adaptives
Design**

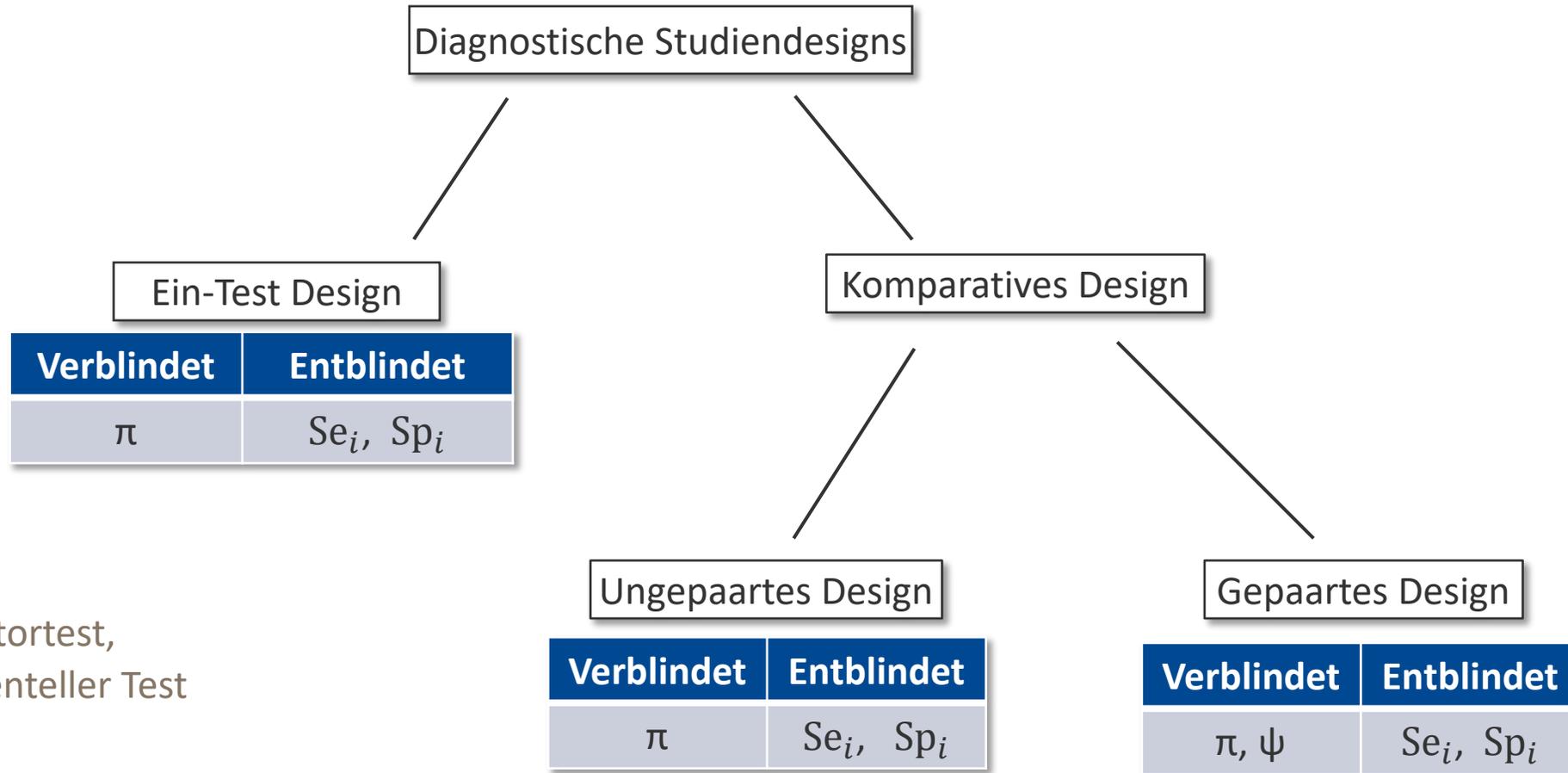
Ergebnisse

Diskussion

 Gepaartes
Design

 Optimale
Fallzahl-
planung

Parameter für Fallzahlneuschätzung



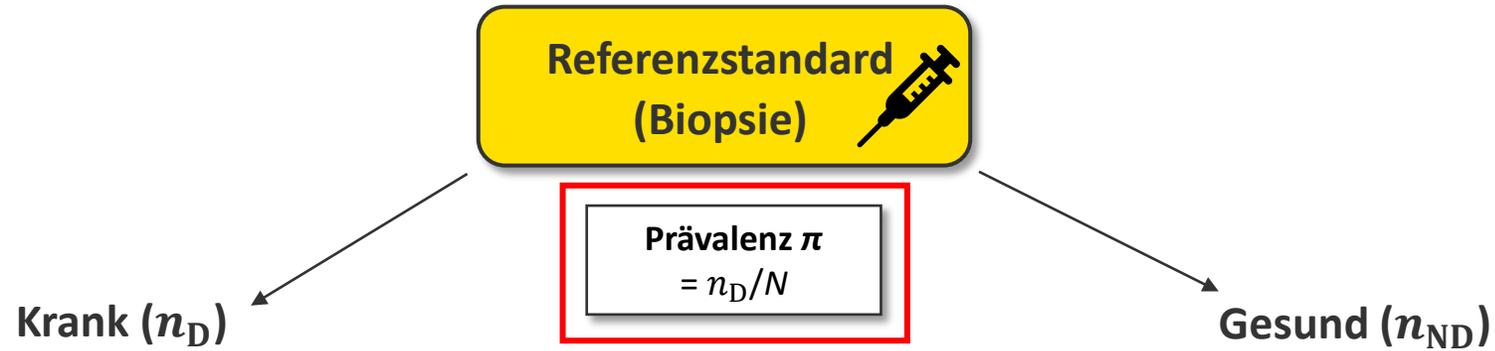
i = Komparatortest,
Experimenteller Test

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Darstellung des gepaarten Designs



		Komparatortest (CT)				Komparatortest (CT)	
Exp. Test (PET/CT) ⊕	Richtig Positiv RP_E	Richtig Positiv RP_C	Falsch Negativ FN_C	Exp. Test (PET/CT) ⊕	Falsch Positiv FP_E	Richtig Negativ RN_C	
	Falsch Negativ FN_E	ψ_D	--		Richtig Negativ RN_E	ψ_{ND}	--
Sensitivität $Se_E = RP_E/n_D$		Sensitivität $Se_C = RP_C/n_D$		Spezifität $Sp_E = RN_E/n_{ND}$		Spezifität $Sp_C = RN_C/n_{ND}$	

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Optimale Fallzahlplanung

③ $N_{Se} \stackrel{!}{=} N_{Sp}$

② $\frac{n_D}{\pi} \stackrel{!}{=} \frac{n_{ND}}{1 - \pi}$

$$\frac{\left(z_{1-\alpha/2} \cdot \psi_D + z_{1-\beta_{Se}} \sqrt{\psi_D^2 - \frac{1}{4} (Se_C - Se_E)^2 (3 + \psi_D)} \right)^2}{\psi_D (Se_C - Se_E)^2 \pi} \stackrel{!}{=} \frac{\left(z_{1-\alpha/2} \cdot \psi_{ND} + z_{1-\beta_{Sp}} \sqrt{\psi_{ND}^2 - \frac{1}{4} (Se_C - Se_E)^2 (3 + \psi_{ND})} \right)^2}{\psi_{ND} (Se_C - Se_E)^2 (1 - \pi)}$$

①

Mit der Bedingung: $Power_{Se} \cdot Power_{Sp} = \text{Gesamtpower} \Leftrightarrow (1 - \beta_{Se}) \cdot (1 - \beta_{Sp}) = \text{Gesamtpower}$

$$\beta_{Sp} = 1 - \frac{\text{Gesamtpower}}{1 - \beta_{Se}}$$

Ablauf des adaptiven Designs

Fixes Design

1. Berechnung der initialen Fallzahl mit Annahmen über Störparameter π , ψ_D and ψ_{ND}
2. Rekrutierung der initial berechneten Fallzahl
3. Auswertung der Studie

Verblindetes adaptives Design

1. Berechnung der initialen Fallzahl mit minimalen ψ_D and ψ_{ND} und Annahme über π
2. Rekrutierung der initial berechneten Fallzahl
3. Schätzung der Störparameter $\hat{\pi}$, $\hat{\psi}_D$ und $\hat{\psi}_{ND}$ 
4. Neuschätzung der Fallzahl \hat{N}
5. Rekrutierung weiterer Personen bis zur finalen Fallzahl
6. Auswertung der Studie

Schätzung der Störparameter

Maximum Likelihood Schätzer Binomialverteilung:

$$\hat{\pi} = \frac{n_D}{N}$$

Maximum Likelihood Schätzer der Multinomialverteilung:

$$\hat{\psi}_D = \frac{n_{D_{10}} + n_{D_{01}}}{n_D}$$

$$\hat{\psi}_{ND} = \frac{n_{ND_{10}} + n_{ND_{01}}}{n_{ND}}$$

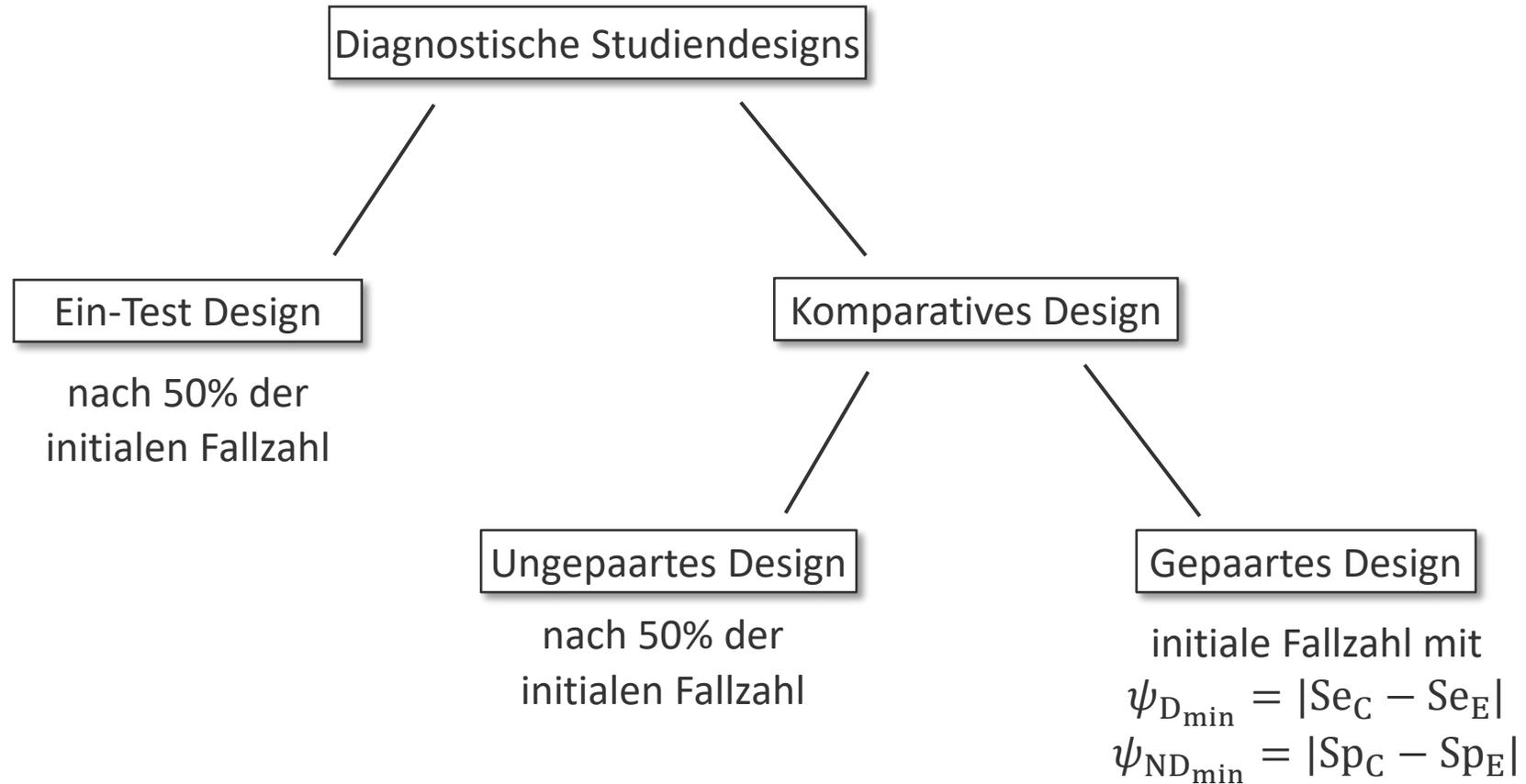
Stark et al., 2022

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

Zeitpunkt der Neuschätzung



Stark & Zapf, 2020; Stark et al., 2022

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde



Einleitung

 Gepaartes Design

Fallzahl- planung

 Optimale Fallzahl-
planung

Adaptives Design

 Anpassung
der Fallzahl

Ergebnisse

Diskussion



Einleitung

 Gepaartes Design

Fallzahl- planung

 Optimale Fallzahl-
planung

Adaptives Design

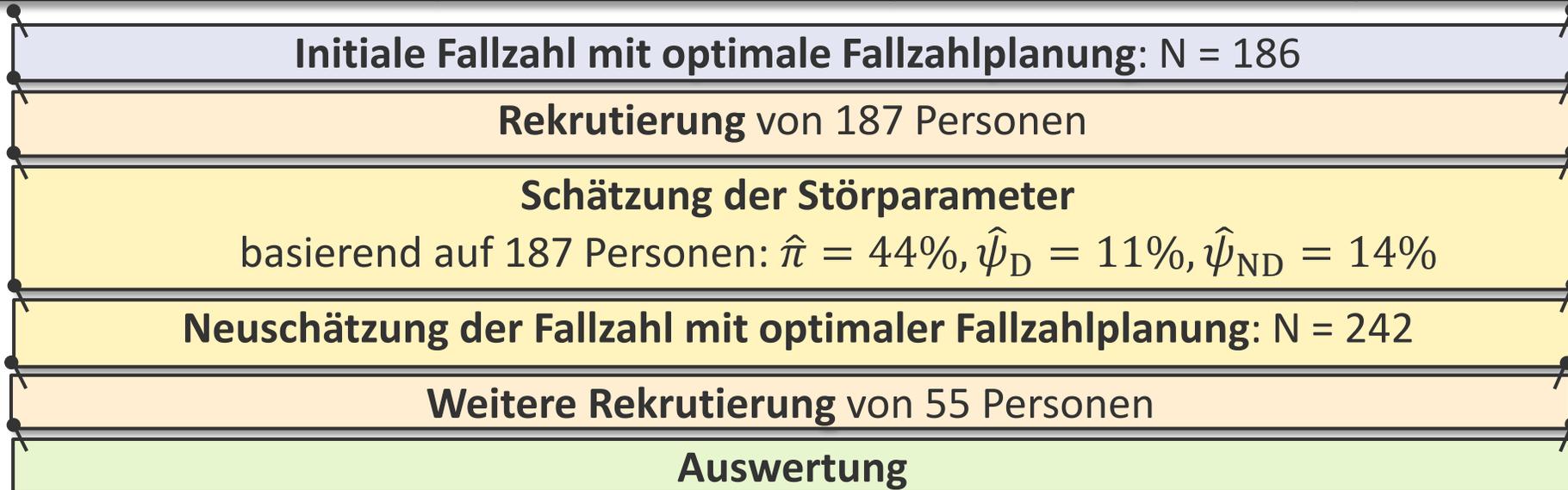
 Anpassung
der Fallzahl

Ergebnisse

Diskussion

Anwendung des adaptiven Designs

$\alpha = 5%$ (zweiseitig), Gesamtpower = 80% Prävalenz: $\pi = 47\%$	Komparatortest (CT)	Experimenteller Test (PET/CT)	Diskordante Testergebnisse
Erkrankte Population	$Se_C = 81\%$	$Se_E = 90\%$	$\psi_D = 9\%$
Gesunde Population	$Sp_C = 66\%$	$Sp_E = 80\%$	$\psi_{ND} = 14\%$



McCray et al., 2017; Stark et al., 2022

Notation Se_E, Sp_E = Sensitivität, Spezifität exp. Test
 Se_C, Sp_C = Sensitivität, Spezifität Komparatortest

π = Prävalenz
 ψ_D = Anzahl diskordante Testergebnisse bei Kranken
 ψ_{ND} = Anzahl diskordante Testergebnisse bei Gesunden

N = gesamte Fallzahl
 N_{Se}, N_{Sp} = Fallzahl Sensitivität, Spezifität
 n_D, n_{ND} = Anzahl Kranke, Gesunde

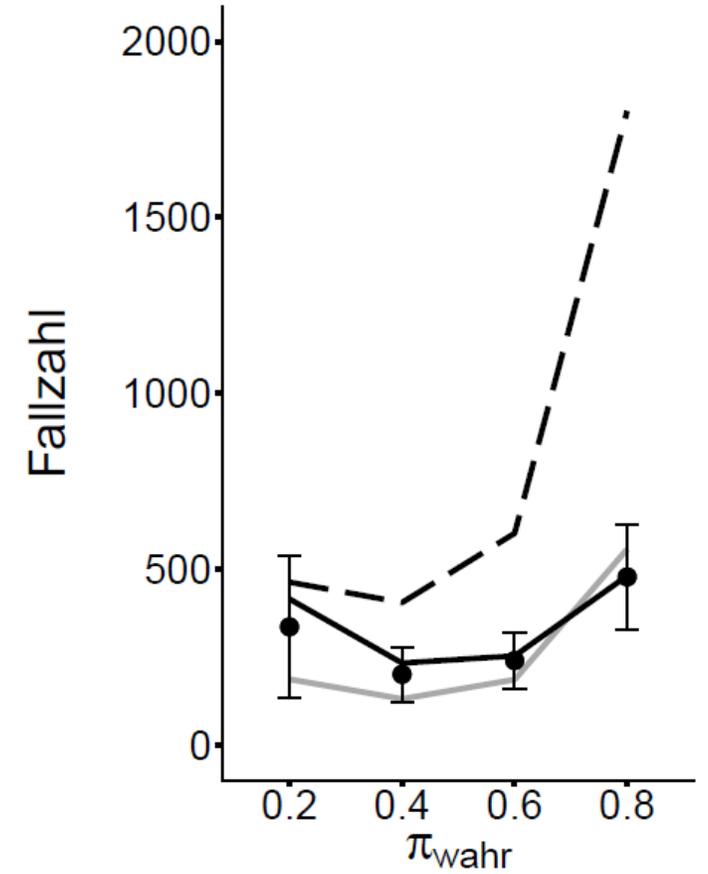
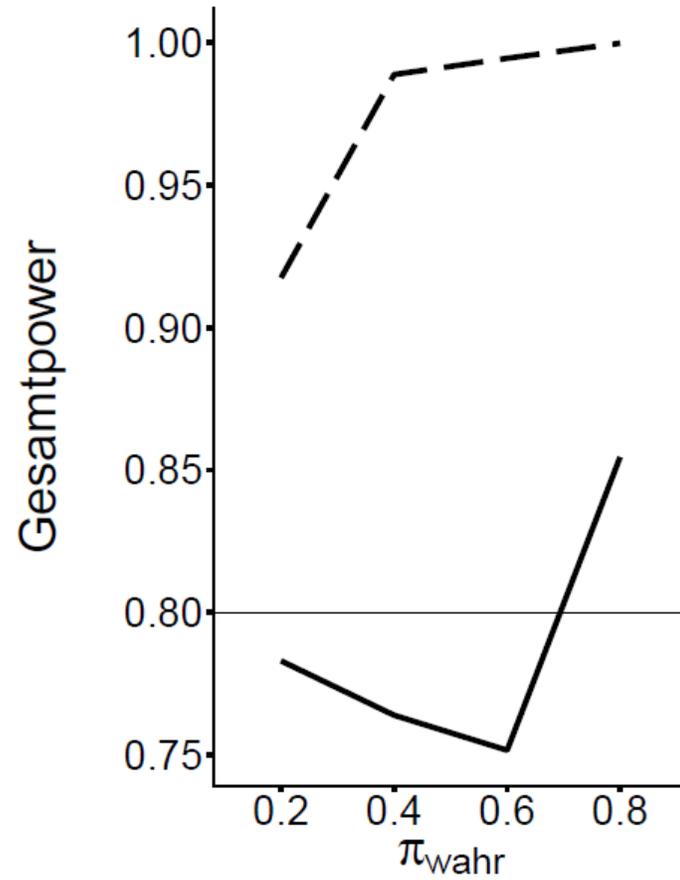
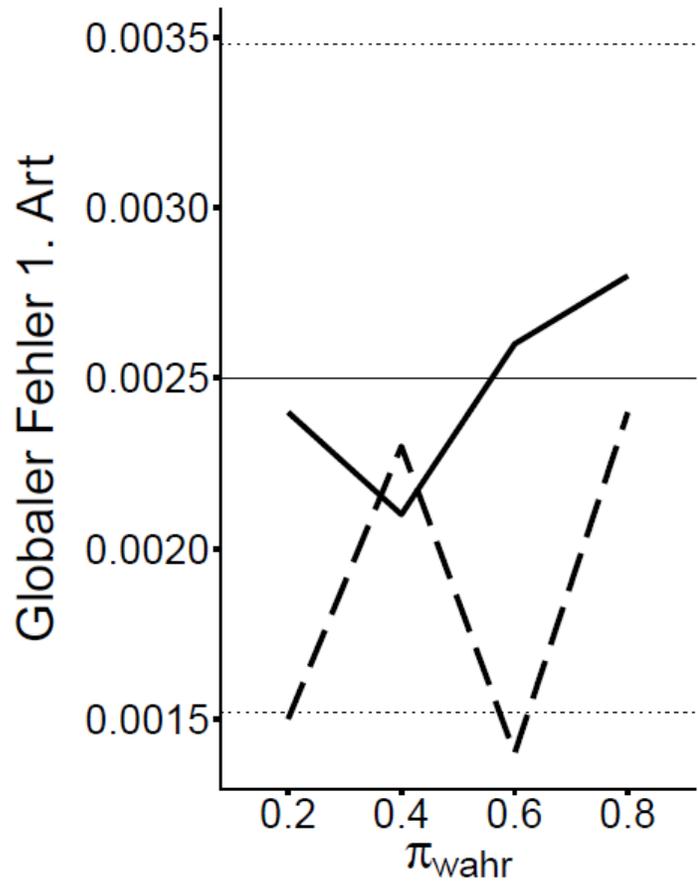
Simulationsstudie

$\alpha = 5\%$ (zweiseitig), Gesamtpower = 80% Wahre Prävalenz: $\pi_{\text{wahr}} = 20, 40, 60, 80\%$	Komparatortest (CT)	Experimenteller Test (PET/CT)	Wahre Diskordante Testergebnisse
Erkrankte Population	$Se_C = 80\%$	$Se_E = 90\%$	$\psi_D = 11\%$
Gesunde Population	$Sp_C = 70\%$	$Sp_E = 80\%$	$\psi_{ND} = 14\%$

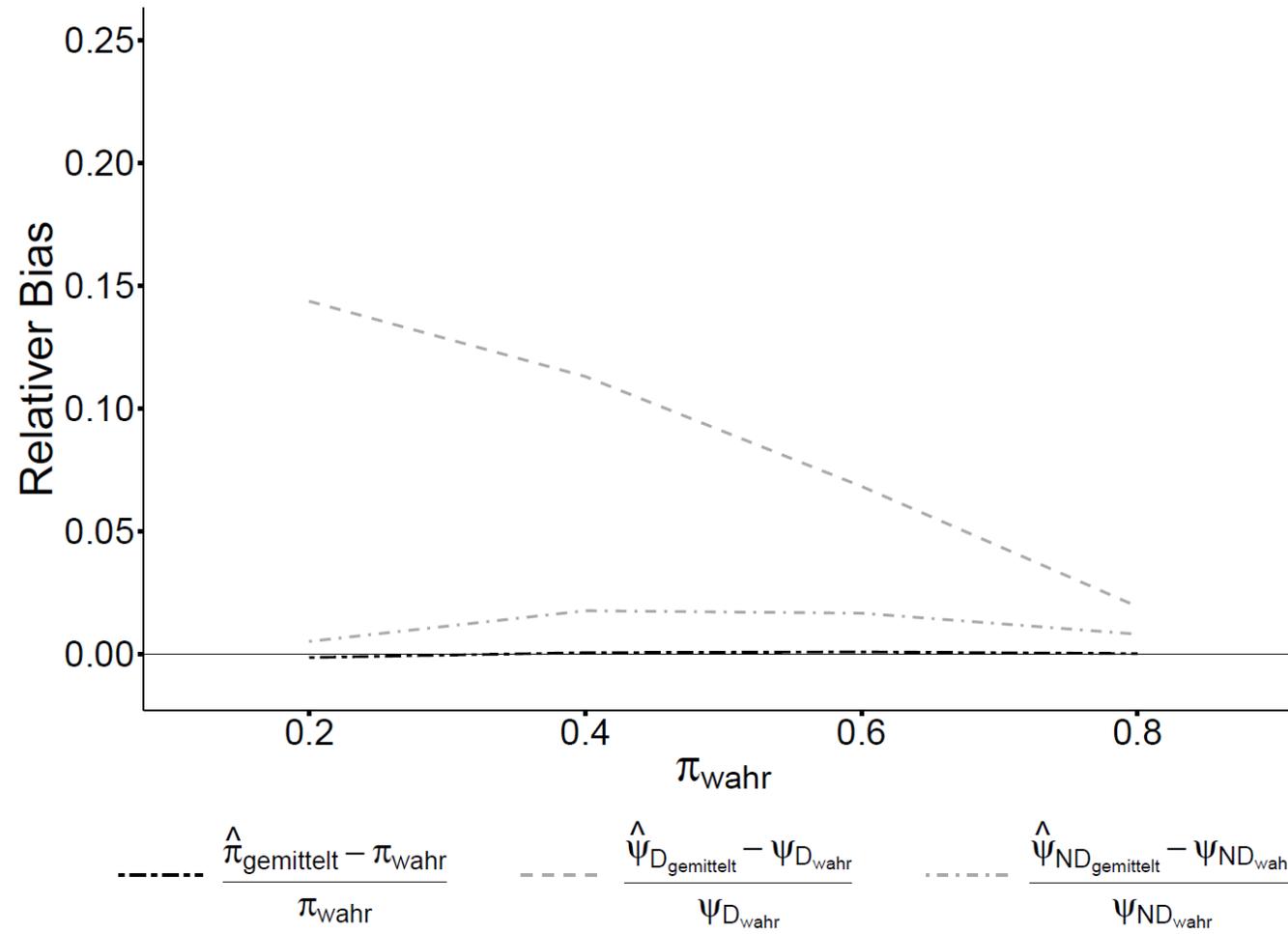
Variation der wahren Prävalenz
20%, 40%, 60%, 80%

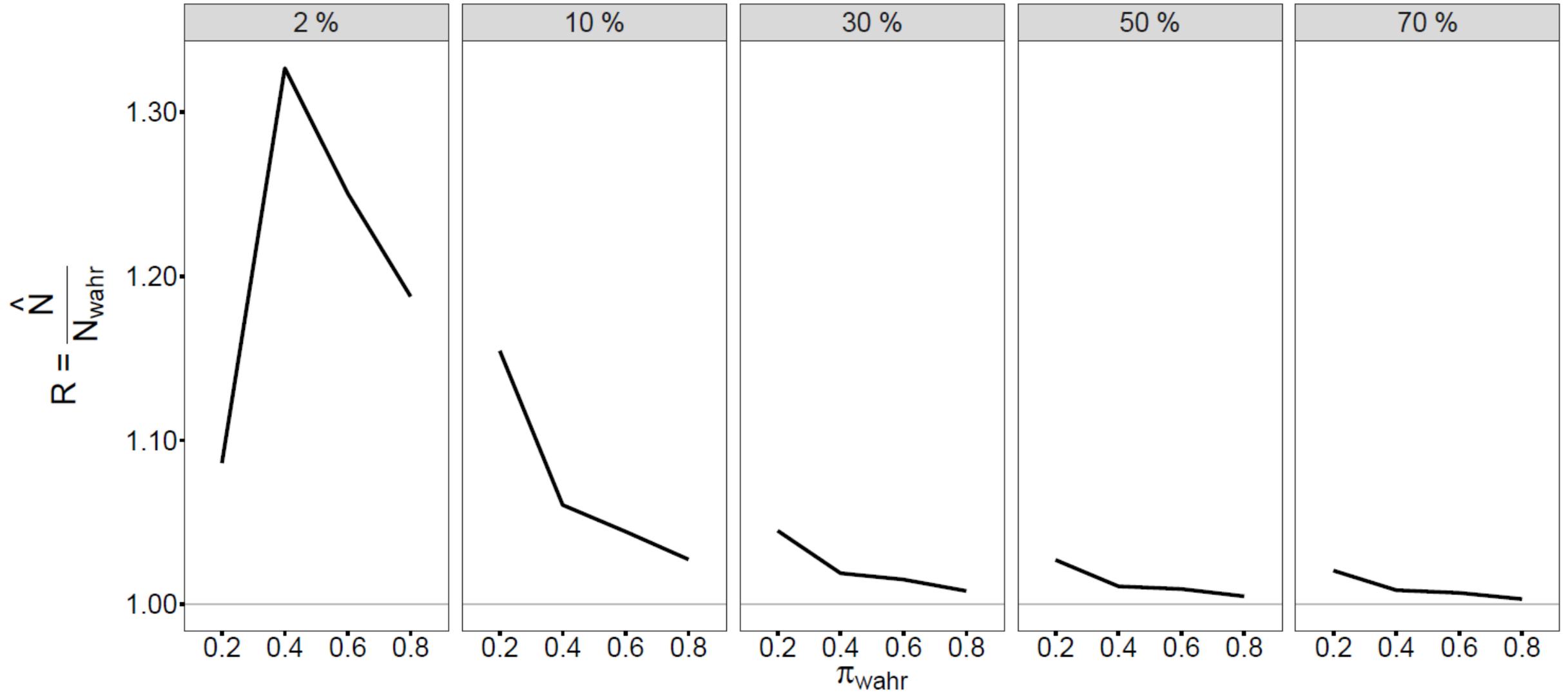
Angenommene diskordante Testergebnisse im fixen Design
 $\psi_D = 18\%$
 $\psi_{ND} = 24\%$

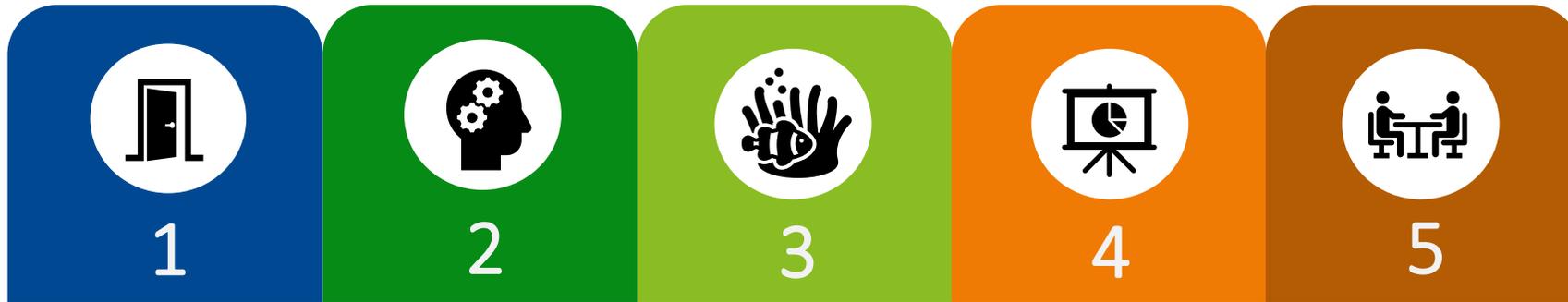
Angenommene Prävalenz
 $\pi_{\text{wahr}} + 10\%$



Interimanalyse
 adaptives Design
 fixes Design
 $N_{\text{wahr}} \pm \text{RMSE adaptives Design}$







Einleitung

 Gepaartes Design

Fallzahl- planung

 Optimale Fallzahl-
planung

Adaptives Design

 Anpassung
der Fallzahl

Ergebnisse

 Einhaltung
Fehler 1. Art

 Erreichen
der Power

Diskussion



Einleitung

 Gepaartes Design

Fallzahl- planung

 Optimale Fallzahl-
planung

Adaptives Design

 Anpassung
der Fallzahl

Ergebnisse

 Einhaltung
Fehler 1. Art

 Erreichen
der Power

Diskussion

Diskussion und Conclusion

Zwei Methoden, die unterstützen die gewünschte Power zu erreichen:

- 1 Optimale Fallzahlplanung
- 2 Verblindetes adaptives Design zur Fallzahlneuschätzung

Vorteile

Optimale Fallzahlplanung:

 Test auf Überlegenheit und/oder Nicht-Unterlegenheit

Adaptives Design:

 Power nahe der gewünschten Power

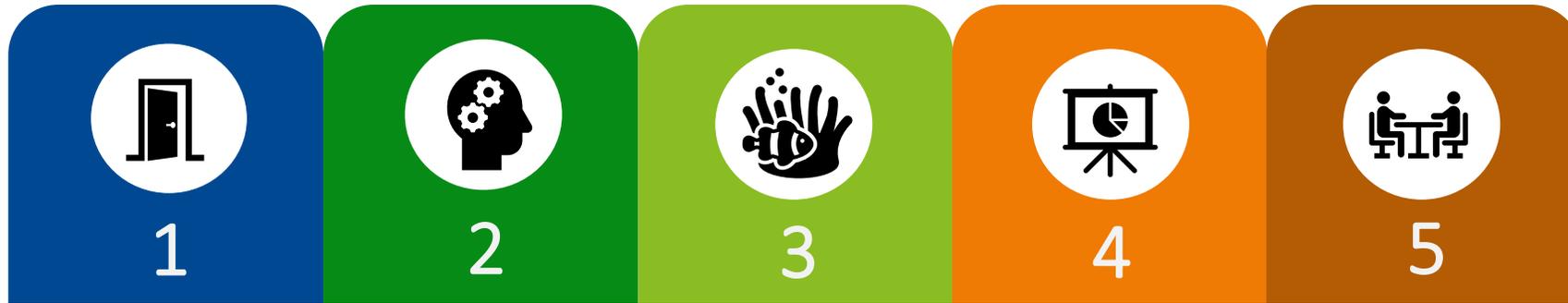
 Einhaltung des Fehlers 1. Art

Nachteile

Adaptives Design:

 Fallzahl zum Zeitpunkt der Interimsanalyse bereits höher als wahre Fallzahl

 Neugeschätzte Fallzahl unrealistisch hoch



Einleitung



Gepaartes
Design

Fallzahl- planung



Optimale
Fallzahl-
planung

Adaptives Design



Anpassung
der Fallzahl

Ergebnisse



Einhaltung
Fehler 1. Art



Erreichen
der Power

Diskussion



Bei unsicherer
Fallzahlplanung:
verblindete
Fallzahlneuschätzung

- Hot, A., Benda, N., Bossuyt, P. M., Gerke, O., Vach, W., & Zapf, A. (2022). Sample size recalculation based on the prevalence in a randomized test-treatment study. *BMC Medical Research Methodology*, 22(1), 205. <https://doi.org/10.1186/s12874-022-01678-7>
- Hot, A., Bossuyt, P. M., Gerke, O., Wahl, S., Vach, W., & Zapf, A. (2021). Randomized test-treatment studies with an outlook on adaptive designs. *BMC Medical Research Methodology*, 21(1), 110. <https://doi.org/10.1186/s12874-021-01293-y>
- McCray, G. P., Titman, A. C., Ghaneh, P., & Lancaster, G. A. (2017). Sample size re-estimation in paired comparative diagnostic accuracy studies with a binary response. *BMC Medical Research Methodology*, 17, Article 102. <https://doi.org/10.1186/s12874-017-0386-5>
- Miettinen, O. S. (1968). The Matched Pairs Design in the Case of All-or-None Responses. *Biometrics*, 24(2), 339-352. <https://doi.org/10.2307/2528039>
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press.
- Stark, M., Hesse, M., Brannath, W., & Zapf, A. (2022). Blinded sample size re-estimation in a comparative diagnostic accuracy study. *BMC Medical Research Methodology*, 22, Article 115. <https://doi.org/10.1186/s12874-022-01564-2>
- Stark, M., & Zapf, A. (2020). Sample size calculation and reestimation based on the prevalence in a single-arm confirmatory diagnostic accuracy study. *Statistical Methods in Medical Research*, 29(10), 2958-2971. <https://doi.org/10.1177/0962280220913588>
- Vach, W., Bibiza, E., Gerke, O., Bossuyt, P. M., Friede, T., & Zapf, A. (2021). A potential for seamless designs in diagnostic research could be identified. *Journal of clinical epidemiology*, 129, 51-59. <https://doi.org/10.1016/j.jclinepi.2020.09.019>
- Zapf, A., Stark, M., Gerke, O., Ehret, C., Benda, N., Bossuyt, P., Deeks, J., Reitsma, J., Alonzo, T., & Friede, T. (2020). Adaptive trial designs in diagnostic accuracy research. *Statistics in Medicine*, 39(5), 591-601. <https://doi.org/10.1002/sim.8430>
- Zhou, X.-H., McClish, D. K., & Obuchowski, N. A. (2011). *Statistical Methods in Diagnostic Medicine* (2nd ed.). John Wiley & Sons.

Maria Stark

Wissenschaftliche Mitarbeiterin
Universitätsklinikum Hamburg-Eppendorf
Institut für Medizinische Biometrie und Epidemiologie
E-Mail: m.stark@uke.de