Fallzahlplanung

Dirk Hasenclever

Klinische Fragestellung

- Ich hätte mal eine kleine Frage für die Planung einer Studie. **Dauert nicht lange.**

- Ich möchte wissen, ob Metoclopramidgabe zusätzlich zu Dexamethason (i.v. intraoperativ) die Häufigkeit postoperativer Übelkeit und postoperativen Erbrechens (PONV) verringert.
- Wieviele Patienten bräuchte ich für so eine Studie?

Illusion Nr. 1:

Der Biometriker rechnet kurz nach und nennt eine Zahl. **DEM IST NICHT SO !**

Hintergrundinformationen

- Indikation
  - OP mit balancierter oder Regionalanästhesie
  - PONV nach Schmerz zweitstärkste post-OP Belastung

- Therapien zur PONV-Prophylaxe
  - Dexamethason, Setrone, Dimenhydrinat: effektiv, OR≈0.5
  - Metoclopramid: umstritten; Datenlage widersprüchlich aus Studien mit verschiedenen Dosierungen und Applikationszeitpunkten
  - Nebenwirkungen von MCP unklar (z.B. extrapyramidale Symptome)

- Motivation aus Pilotstudie
  - n=200, 8mg Dexamethason + 10 mg MCP, PONV-Rate nur 10%

Odds ratios

$\frac{p}{1-p}$ heißt Odds

Quantifikation von Unterschieden als Odds Ratio

$$\frac{\frac{p1}{1-p1}}{\frac{p2}{1-p2}}$$

Logistische Regression

$$g(x) = \frac{\exp(x)}{1+\exp(x)}; \qquad g^{-1}(p) = \ln\left(\frac{p}{1-p}\right) \text{ (Logodds)}$$

Mögliche Studienfragestellungen

(Leitlinien: Dexamethason erste Wahl für PONV-Prophylaxe.)

- Bringt die zusätzliche Gabe von MCP zu 8 mg Dexamethason eine zusätzliche Verringerung der PONV-Rate?

- Wenn ja, welche MCP-Dosis (10/25/50 mg) sollte man nehmen?

- Könnte es sein, dass sich Dexamethason und MCP in der Wirkung verstärken? (superadditive Wirkung)

- Sind die Nebenwirkungsraten akzeptabel?

Studiendesign

randomisiert, Parallelgruppen, doppelblind

- **erste Frage erfordert Vergleich Dexa vs. Dexa+MCP**
- **Dosisfragestellung: 4 Arme**
  Dexa als Basismedikation für alle + 0 / 10 / 25 / 50 mg MCP
- **Interaktion (bei 1 Dosisstufe MCP) erfordert faktorielles Design:**
  Kontrolle, nur Dexa, nur MCP, Dexa+MCP
- **Interaktion + Dosisfragestellung: 8 Arme**

Dexa:	**0**	**0**	**0**	**0**	**8**	**8**	**8**	**8**
MCP:	**0**	**10**	**25**	**50**	**0**	**10**	**25**	**50**

Welche Endpunktdaten werden erhoben?

- Angaben zu Erbrechen / Übelkeit innerhalb 24h post OP
  - überhaupt aufgetreten
  - was (Erbrechen/Übelkeit/Beides)
  - Zeitpunkt
  - Intensität (Skala 1-10)
  - Dauer der Episoden (Definition der Episoden!)
  - Verbrauch an Rescue-Medikation

- unerwünschte Ereignisse
  - intra-OP nach Gabe der Medikation
  - post-OP (was, wann, Dauer, Intensität)

Definition des primären Endpunkts

- Primärer Endpunkt: **PONV ja/nein**, weil ...
- ... enthält zwar wenig Information,
- ... erfasst jedoch alle Ereignisse,
- ... ist weniger subjektiv als PONV-Intensität,
- ... ist noch zuverlässig zu erheben, auch wenn ein Follow-up verspätet durchgeführt wird oder Detailangaben fehlen.

Größe des entdeckenswerten Unterschieds

- angenommenes Szenario für PONV-Raten (8 Dexa + x MCP):
  x=0: **30%** x=10: **25%** x=25: **22.5%** x=50: **20%**
  (ziemlich pessimistisch: Odds Ratio > 0.5 für die größte Dosis)

- Aufdeckung seltener Ereignisse: 4 vs. 5 vs. 6 vs. 8%

- in einem faktoriellen Design könnte man für PONV annehmen:

Prophylaxe	PONV	PONV:nicht-PONV	Odds Ratio vs. Kontr.
Kontrolle:	**50%**	**1:1**	
Dexa:	**33%**	**1:2**	**0.5**
MCP:	**40%**	**2:3**	**0.66**
Dexa+MCP additiv:	**25%**	**1:3**	**0.33**
Dexa+MCP superadditiv:	**20%**	**1:4**	**0.25**

Interaktionseffekt

Diagnostische Tests ...

... in der Medizin

		Patient	
		krank	**gesund**
Test	**+**	☺ *Sensitivität*	falsch pos. ☹
	–	falsch neg. ☹	☺ *Spezifität*

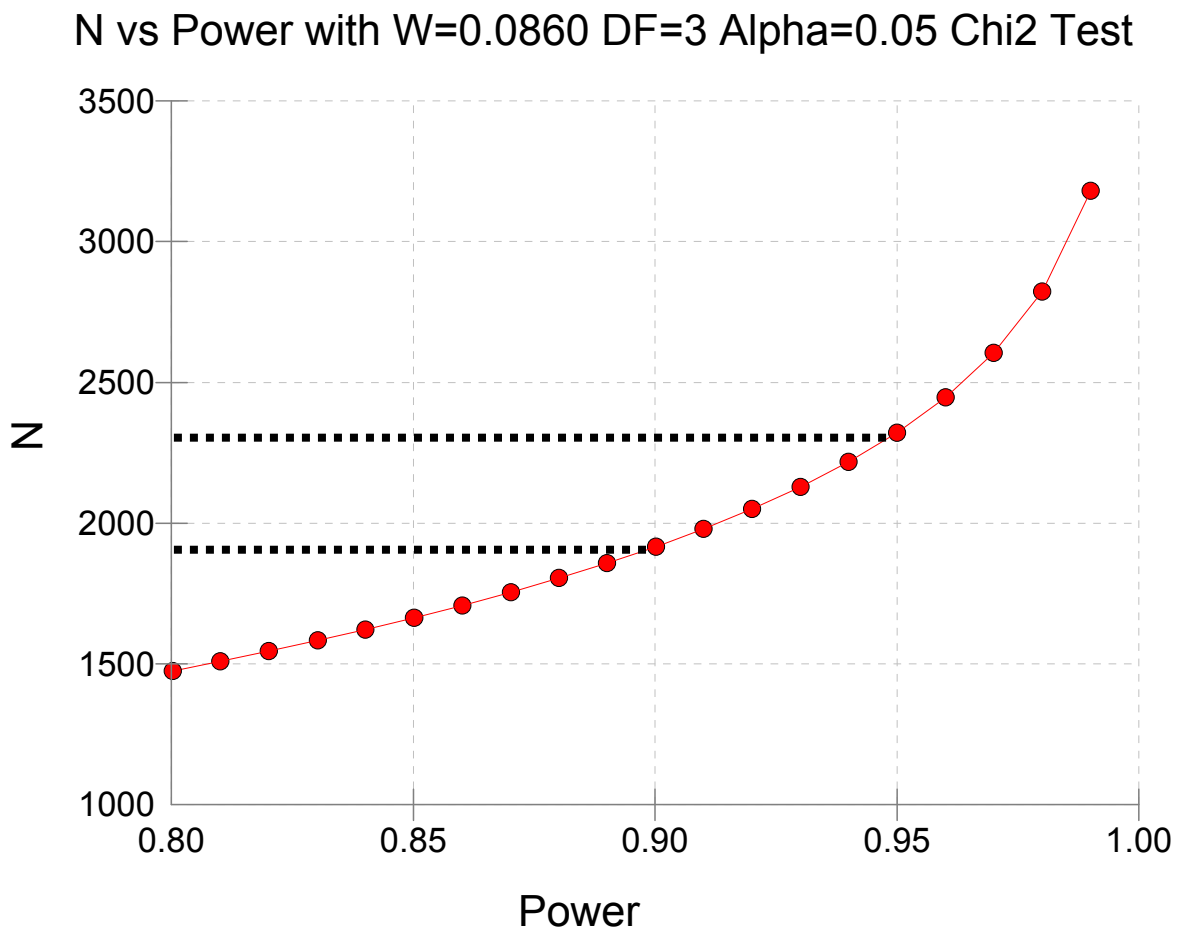
... in der Statistik

		Nullhypothese („null Unterschied“)	
		falsch	**richtig**
Test	**+**	☺ *Power*	F. 1. Art ☹ α
	–	F. 2. Art ☹ β	☺ *Konfidenzniveau*

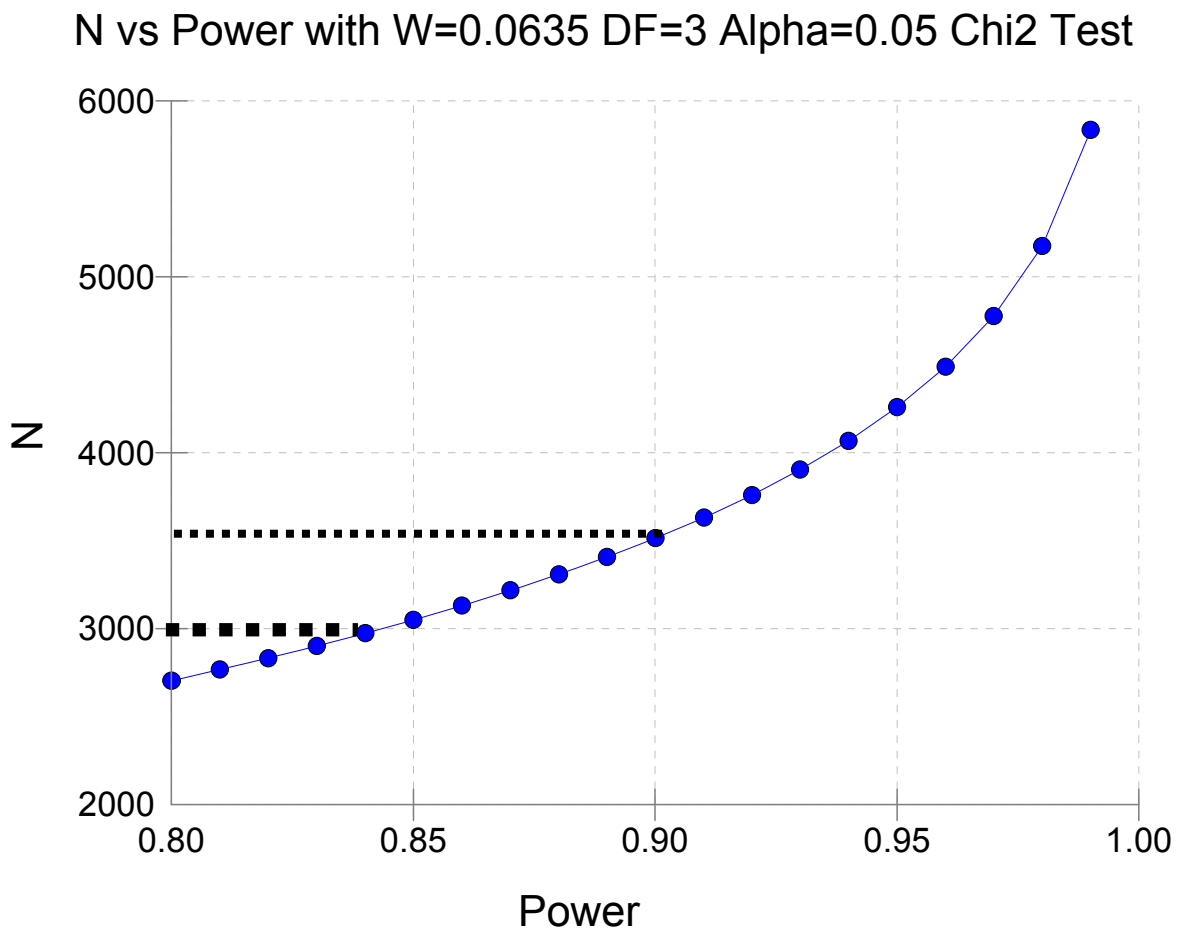
Test, Signifikanzniveau und Power

- Chi-Quadrat-Test für >2 Ereignisraten
- Signifikanzniveau: 5%
- Power für PONV: mindestens 90%, besser 95%
- Power für wichtige sekundäre Endpunkte: mindestens 80%

Fallzahlkalkulation: PONV



Fallzahlkalkulation: seltene Ereignisse



Fallzahlfestsetzung

- Eine Gesamtfallzahl von N=3000 ...
  - ... liefert >98% Power für PONV
  - ... liefert 84% Power für das Szenario für seltene Ereignisse
  - ... ist in vertretbarer Zeit erreichbar.

- Ein faktorielles Design zum Erkennen der Wechselwirkung ...
  - ... erfordert selbst für 80% Power eine Fallzahl von n>7000 ...
  - ... das ist nicht erreichbar.
  - Auf die Beantwortung dieser Fragestellung wurde verzichtet.

--- Ende der Fallzahlplanung für die PONV-Studie ---

Planungsvorgaben

- Signifikanzniveau α, zu dem der Fehler 1. Art kontrolliert werden soll. [α = pr (H0 ablehnen | H_0 wahr)]
- Spezifische Alternativhypothese H_A
- Powerforderung (1-β) für H_A : β ist der maximal tolerierte Fehler 2. Art. [β = pr (H_0 beibehalten | H_A wahr)]
- Zusatzinformation über Varianz durch Störgrößen und Messfehler.

- Allereinfachstes Beispiel: Fallzahl für den Gauss Test.
- Fallzahlformeln, Fallzahlplanungssoftware. Eventuell Simulation.

Frequentistische Fallzahlplanung I

Einfachster Fall: $X \sim N(\theta, \sigma^2)$; σ^2 bekannt.

Einseitiges Testproblem: $H_0: \theta \leq 0$ gegen $H_1: \theta > 0$

Stichprobe des Umfangs n

Testgröße: $Z = \frac{\bar{x}_{(n)}}{\sigma}\sqrt{n} \sim N(0,1)$ unter der Nullhypothese. ("Gauss - Test")

Sei $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung N(0,1):

$$\Phi(z) = \int_{-\infty}^{z} \phi(\varsigma) d\varsigma; \quad \text{mit der Dichtefunktion} \quad \phi(\varsigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varsigma^2}{2}\right)$$

Sei $Z_{1-\alpha}$ das $1-\alpha$ Quantil der Standardnormalverteilung: $\Phi(Z_{1-\alpha}) = 1-\alpha$

$\Rightarrow$ Lehne H_0 ab, falls $Z > Z_{1-\alpha}$

(N.B.: Falls Test zweiseitig sein soll: $|Z| > Z_{1-\alpha/2}$)

Frequentistische Fallzahlplanung II

Planungsvorgaben: $\Delta =$ relevanter Unterschied, für den der Fehler zweiter Art kontrolliert, bzw. die Power garantiert werden soll.

$1-\beta =$ Powervorgabe

Verteilung der Teststatistik Z unter der konkreten Alternative H_Δ: $X \sim N(\Delta, \sigma^2)$

$$Z \sim N\left(\frac{\Delta}{\sigma}\sqrt{n},\quad 1\right)$$

$\left(\frac{\Delta}{\sigma}\right)$ heißt auch der effektive Unterschied (Δ in Einheiten der Standardabweichung).

Frequentistische Fallzahlplanung III

Powerbedingung:

$$1\text{-}\beta = \text{pr}(\text{Ablehnen } H_0 \mid H_\Delta) = \text{pr}(Z > Z_{1\text{-}\alpha} \mid H_\Delta)$$

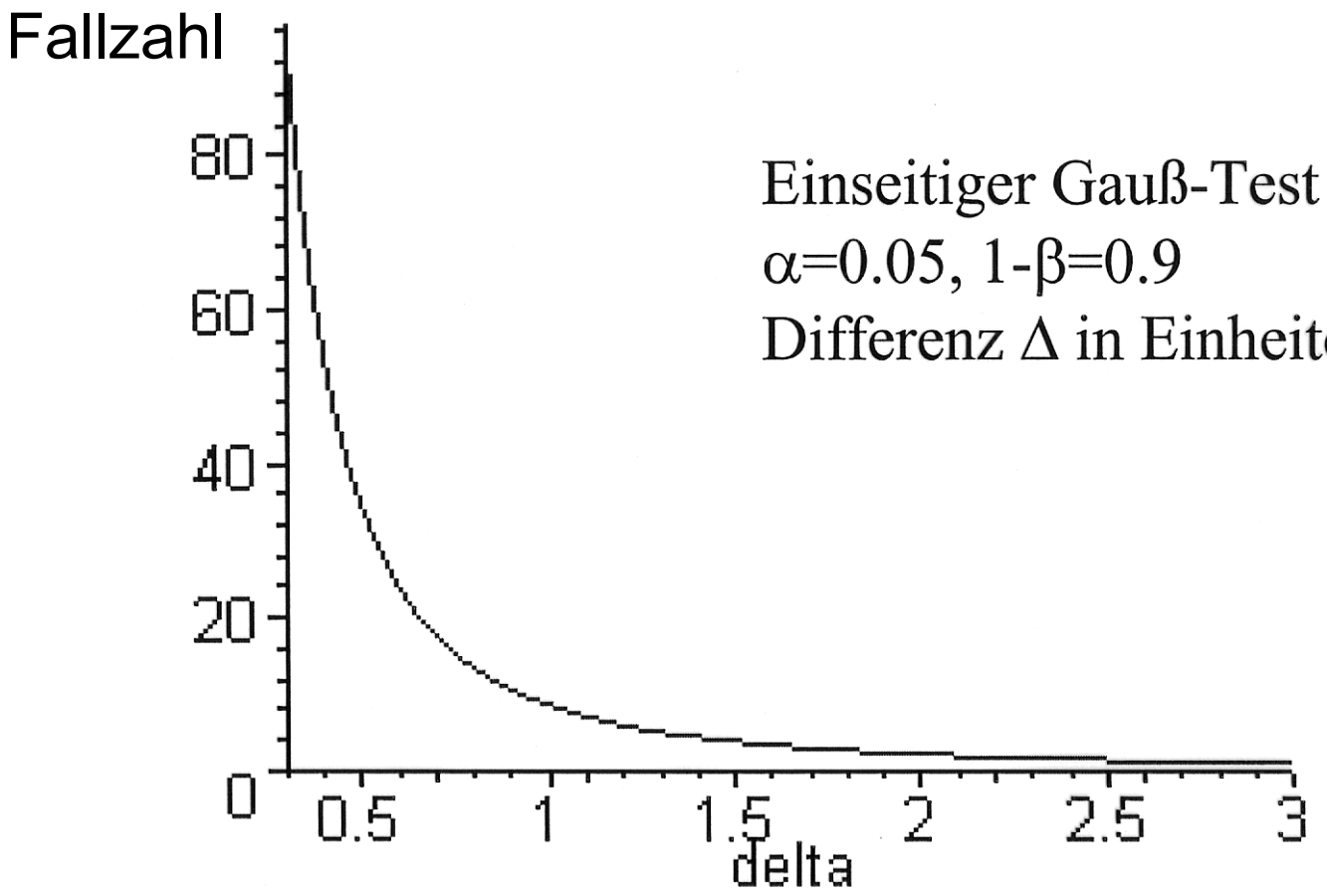
$$= 1 - \Phi\left(Z_{1\text{-}\alpha} - \frac{\Delta}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{\Delta}{\sigma}\sqrt{n} - Z_{1\text{-}\alpha}\right)$$

Also (Φ^{-1} auf beiden Seiten): $Z_{1\text{-}\beta} := \Phi^{-1}(1-\beta) = \frac{\Delta}{\sigma}\sqrt{n} - Z_{1\text{-}\alpha}$

Auflösen nach n ergibt die erforderliche Fallzahl:

$$n = \frac{(Z_{1\text{-}\alpha} + Z_{1\text{-}\beta})^2}{\left(\frac{\Delta}{\sigma}\right)^2}$$

Frequentistische Fallzahlplanung III



Einseitiger Gauß-Test
α=0.05, 1-β=0.9
Differenz Δ in Einheiten von ☞

Zutaten für die Fallzahlplanung

Für die Fallzahlkalkulation ist festzulegen:

- Mit welchem **primären Endpunkt** soll der Erfolg einer Therapie gemessen werden?
- Mit welchem **statistischen Werkzeug** (Test) wird er ausgewertet?
- Welcher **Unterschied** soll aufgedeckt werden?
- Mit welcher **Varianz**, die diesen Unterschied „verwischt“, muß gerechnet werden?
- Wie groß dürfen die **Fehlerraten** der statistischen Diagnostik sein? (Signifikanzniveau+Power festlegen)
- Ggf.: Welcher Anteil von **Drop-outs** wird erwartet?

Quellen für die Informationen ...

... die der Biometriker braucht:

- Literatur
- vorhandene Daten aus alten Studien oder Patientenakten
- Kenntnis des Patientenguts (auch in Dingen, die nicht immer explizit in den Akten stehen, z.B. Compliance, soziale Aspekte)
- ggf. Daten, die zur Studienplanung probehalber erhoben werden
- ggf. gedankliche Planspiele zur möglichen Datenlage

Zusammenfassung: Der Weg zur Fallzahl

Biometrie = Anlegen eines Meßwerkzeugs an einen biologischen Sachverhalt

- präzise Formulierung der klinischen Fragestellung
- Prüfung möglicher Designs, Designoptimierung
- Festlegung des Regimes der Erhebung endpunkt-relevanter Informationen und der Werkzeuge dafür
- Definition des Endpunkts
- Festlegung geeigneter Prozeduren, welche die valide Erhebung des Endpunkts sichern
- Beschaffung von Informationen, welche die voraus-sichtliche Datenlage beim Endpunkt erahnen lassen

Statistik = Rechnen mit Wahrscheinlichkeiten

- Festlegung der Auswertungsmethode
- Zusammenstellung der quantitativen Parameter zur Fallzahlberechnung
- Durchführung der Berechnung