

Biometrische Methodik, Fallzahlplanung und statistische Power
Kurze Hinweise zur Erstellung von Studienprotokollen klinischer Studien
zur Einreichung an die Ethikkommission

Zu Anspruch und Notwendigkeit

Wichtige Voraussetzungen für die ethische Vertretbarkeit von Experimenten an Menschen sind eine hinreichende **Erfolgschance** der Studie und absehbar ausreichende **Belastbarkeit** der Studienergebnisse bei der Verallgemeinerung auf eine Population.

Die Erfolgschance wird biometrisch mit der **statistischen Power** gemessen, ausreichende Belastbarkeit wird durch ein von allen beteiligten Disziplinen konsensualisiertes

Studienprotokoll angestrebt. Das Studienprotokoll ist eine gedankliche Vorwegnahme aller Studienschritte und stellt die eigentliche wissenschaftliche Leistung dar.

Es gibt ausführliche Guidelines zur Erstellung von Studienprotokollen, aus biometrischer Sicht z.B. Schäfer u.a. 1999 (www.tu-dresden/medek).

Ziel biometrischer Verfahren ist die Verallgemeinerung der empirischen

Beobachtungsergebnisse auf eine Population, d.h. der Schluß von den konkret untersuchten Patienten auf eine (fiktive) Gesamtheit z.B. aller zukünftigen Patienten unter kontrollierten inferentiellen Gütekriterien.

Mit anderen Worten: Ziel ist die Abgrenzung des Versuchsergebnisses vom Zufall.

Hauptsächlich verwendet man statistische Tests (zur Prüfung von Hypothesen, z.B. für inter- und intraindividuelle Vergleiche) und statistische Schätzungen (für Effektmaße, z.B. relative Wirksamkeit, Anteil Responder usw.).

Studien ohne biometrische Planung (Pilotstudien) sind ethisch vertretbar, z.B. wenn

- Vorversuche an sehr wenigen Patienten durchgeführt werden sollen, um neue experimentelle Methoden oder Meßprinzipien zu erproben, ohne auf eine Population schließen zu wollen,
- Planungsgrundlagen für eine spätere Hauptstudie zu entwickeln sind.

In diesem Fall muß die notwendige Fallzahl nichtbiometrisch begründet werden, und konfirmatorische (=„beweisende“) Signifikanzaussagen müssen vermieden werden.

Beispiele für die Zielstellung einer klinischen Studie:

- Nachweis der Überlegenheit einer Testtherapie gegenüber einer Standardtherapie (einseitiger Unterschiedsnachweis)
- Nachweis der Wirksamkeit einer Therapie (Unterschiedsnachweis Verum versus Plazebo)
- Nachweis der Nichtunterlegenheit einer Testtherapie gegenüber einer Standardtherapie (einseitiger Äquivalenznachweis)
- Nachweis der Äquivalenz zweier Behandlungen (zweiseitige Äquivalenz)

Ausreichende Erfolgschance einer Studie bedeutet i.A. eine **Mindestpower von 80%**. Damit ist das Risiko einer erfolglosen Studie (d.h. nichtsignifikantes Ergebnis) immer noch 20%. In Ausnahmefällen darf die Power auch geringer sein. Dazu ist eine ethische Güterabwägung notwendig.

Die **Anzahl notwendiger Patienten** bzw. Studienteilnehmer hängt zusammen mit der Power und der Größe des zu erwartenden „Effektes“: große Effekte sind mit wenigen Patienten

objektivierbar (z.B. 6 bis 10 Patienten), kleine Effekte erfordern immensen Aufwand (mehrere hundert oder tausend Patienten). Die zu erwartende und in der Studie nachzuweisende Effektgröße muß aus Vorversuchen oder aus der Literatur entnommen werden.

Alle weiteren Überlegungen zur Fallzahlplanung gliedern sich nach der **Skalierung der Hauptzielvariablen** der Studie, hier unterschieden in die beiden wichtigsten Szenarien: „Prozentzahl“ (=binärer Outcome) und „Mittelwert“ (=quantitativer Outcome), sowie danach, ob ein Test oder eine Schätzung das Hauptziel darstellt.

A Vergleich von Wahrscheinlichkeiten („Prozenten“):

Die notwendige Fallzahl hängt ab

- von der nachzuweisenden Effektgröße = Mindestunterschied der beiden zu vergleichenden Prozentzahlen gemessen in Prozentpunkten und
- von der ungefähren Lage der beiden Prozentzahlen (wenn unbekannt, dann verwende „worst case“ = symmetrisch um 50%)

Beispielszenarien

Zwei zu vergleichende Gruppen im Überlegenheitsnachweis
Signifikanzniveau $\alpha=0,05$ einseitig

	Effektgröße						
	80%Pkte.	40%Pkte.		20%Pkte.		10%Pkte.	
Gruppe 1	≤ 5%	≤ 5%	≤ 30%	≤ 5%	≤ 40%	≤ 5%	≤ 45%
Gruppe 2	≥ 85%	≥ 45%	≥ 70%	≥ 25%	≥ 60%	≥ 15%	≥ 55%
notwendige Fallzahl pro Gruppe	4	14	19	39	77	111	309

Falls mehrere Gruppen zu vergleichen sind und/ oder mehrere Zielkriterien gleichzeitig und gleichgewichtig benutzt werden, erhöht sich die Fallzahl durch die notwendigen Testadjustierungen beträchtlich. Deshalb sollte man die Anzahl der statistischen Tests möglichst beschränken.

Sonderfall „100%“:

Es sei nachzuweisen, daß das betrachtete Zielereignis immer oder überhaupt nicht auftritt. In diesem Fall erfolgt der Nachweis der statistischen Evidenz für die Forschungshypothese durch Angabe des Konfidenzintervalles, das dann genügend eng um die 100% (bzw. 0%) liegen muß.

Beispiele für den Nachweis „100%“

(Versuchsergebnis sei, daß alle untersuchten Patienten das betrachtete Zielereignis haben):

	Untere Grenze des 95%-Konfidenzintervalles			
	90%	95%	99%	99,9%
Anzahl notwendiger Patienten	28	58	297	592

Schreibweise für das Versuchsergebnis dann z.B.:

„Mit einer Irrtumswahrscheinlichkeit von 5% beträgt die Wahrscheinlichkeit für das Zielereignis mindestens 90% (bzw. wie in der Tabelle als untere Grenze des Konfidenzintervalles angegeben)“.

B Schätzung von Wahrscheinlichkeiten:

Die Schätzung für eine gesuchte Populationswahrscheinlichkeit erfolgt mit Hilfe des Prozentanteiles der Patienten, die das Zielereignis aufweisen, ergänzt um den statistischen „Unschärfbereich“ = Konfidenzintervall.

Die maximal tolerierbare Breite des „Unschärfbereiches“ muß fachwissenschaftlich begründet werden! Aus ihr ergibt sich die notwendige Fallzahl.

Beispiele:

	Breite des 95%-Konfidenzintervalles			
	40%Pkte.	20%Pkte.	10%Pkte.	6%Pkte.
Beispiel 1: erwartet 10%	1% ... 41% n = 27	4% ... 24% n = 54	6% ... 16% n = 153	7% ... 13% n = 442
Beispiel 2: erwartet 50%	30% ... 70% n = 20	40% ... 60% n = 82	45% ... 55% n = 290	47% ... 53% n = 784

1. Zeile: erwünschtes 95%-Konfidenzintervall

2. Zeile: notwendige Anzahl Patienten

C Vergleich von Mittelwerten:

Die notwendige Fallzahl hängt ab

- von der erwarteten und nachzuweisenden relativen Effektgröße = Mindestunterschied der beiden zu vergleichenden Mittelwerte (= absolute Effektgröße) im Verhältnis zur Standardabweichung SD (SD in jeder Gruppe einzeln bei unabhängigen Gruppen; bzw. SD der individuellen Differenzen bei intraindividuellen Vergleichen), gemessen in Prozent

Beispielszenarien

Eine, zwei, drei oder vier mit einer Kontrolle zu vergleichende unabhängige Gruppen im Unterschiedsnachweis

Signifikanzniveau $\alpha=0,05$ zweiseitig

Anzahl Testgruppen	notwendige Fallzahl		
	relativer Mindesteffekt 80%	relativer Mindesteffekt 100%	relativer Mindesteffekt 120%
1	in Behandlungs- und Kontrollgruppe je 26, zusammen 52	in Behandlungs- und Kontrollgruppe je 17, zusammen 34	in Behandlungs- und Kontrollgruppe je 12, zusammen 24
2	in jeder Behandlungsgruppe 26, in der Kontrollgruppe 37, zusammen 89	in jeder Behandlungsgruppe 17, in der Kontrollgruppe 24, zusammen 58	in jeder Behandlungsgruppe 12, in der Kontrollgruppe 17, zusammen 41
3	in jeder Behandlungsgruppe 27, in der Kontrollgruppe 43, zusammen 124	in jeder Behandlungsgruppe 17, in der Kontrollgruppe 29, zusammen 80	in jeder Behandlungsgruppe 12, in der Kontrollgruppe 21, zusammen 57
4	in jeder Behandlungsgruppe 27, in der Kontrollgruppe 50, zusammen 158	in jeder Behandlungsgruppe 17, in der Kontrollgruppe 34, zusammen 102	in jeder Behandlungsgruppe 12, in der Kontrollgruppe 24, zusammen 72

Literatur z.B.:

- Machin D, Campbell MJ: Statistical Tables for the Design of Clinical Trials. Blackwell Publ. Oxford 1987
 Rasch D u.a.: Verfahrensbibliothek Versuchsplanung I + II. Oldenbourg Verlag, München 1998
 Schäfer H, Berger J u.a.: Empfehlungen für die Erstellung von Studienprotokollen (Studienplänen) für klinische Studien. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 30(1999), 141-154