

Redaktion

A. Zink, Berlin
 J. Sieper, Berlin

J. Gerß · M. Eveslage · A. Faldum · R. Schmidt

Institut für Biometrie und Klinische Forschung (IBKF), Westfälische Wilhelms-Universität Münster

Moderne Studiendesigns und Analysemethoden in der klinischen Forschung

Der medizinische Fortschritt verlangt die Entwicklung neuer statistischer Methoden, die über klassische Studiendesigns hinausgehen. Moderne Studiendesigns wie mehrstufige adaptive Designs und Bayes-Verfahren werden diesen Herausforderungen gerecht.

Klinische Studien, bei denen nur eine geringe Anzahl von Patienten zur Verfügung steht, stellen methodisch eine große Herausforderung dar. Dies gilt nicht nur für seltene rheumatische Erkrankungen wie Vaskulitiden oder Kollagenosen. Auch moderne medizinische Entwicklungen tragen dazu bei, dass zunehmend mehr klinische Studien mit kleinen Patientenzahlen durchgeführt werden. Die Ursache hierfür sind u. a. neue diagnostische Verfahren zur Krankheits- und Patientencharakterisierung, mit denen individuelle Risikoprofile und Therapiekonzepte erstellt werden können („personalisierte Medizin“).

Klinische Studien mit kleiner Patientenzahl verlangen die Entwicklung neuer statistischer Methoden, die über die klassischen Studiendesigns hinausgehen, um unter Wahrung der etablierten Qualitätskriterien mit diesen Herausforderungen umgehen zu können.

Im Folgenden sollen nach einem einleitenden Verweis auf klassische Studiendesigns die modernen Ansätze adaptiver Designs und Bayes-Designs dargestellt werden.

Klassische Designs in konfirmatorischen klinischen Studien

In konfirmatorischen klinischen Studien gilt es, aus den gesammelten Informationen Schlüsse zu ziehen, um beispielsweise die Wirksamkeit eines neuartigen Medikaments gegenüber einer Vergleichstherapie nachzuweisen. Dabei erfolgt der Wirksamkeitsnachweis durch das Anwenden von statistischen Tests. Soll in einer Studie etwa die Wirkung einer neuartigen Therapie gegenüber Placebo bei der Behandlung von Patienten mit rheumatoider Arthritis untersucht werden, kann dazu eine zweiarmlige randomisierte Studie durchgeführt werden, in der eine Hälfte der Patienten mit der neuartigen Therapie und die andere Hälfte mit Placebo behandelt wird. Als primäre Zielgröße soll im folgenden Beispiel die Änderung des Disease Activity Scores (DAS28) gegenüber Baseline dienen (Δ_{DAS28}). Dabei wird die Zielgröße Δ_{DAS28} hier vereinfachend als normalverteilt mit bekannter Standardabweichung $\sigma=2,4$ angenommen.

Für den statistischen Test werden die Nullhypothese H_0 und die Alternativhypothese H_1 aufgestellt:

H_0 : Die neuartige Therapie wirkt bezüglich der erwarteten Veränderung des DAS28 schlechter oder höchstens genauso gut wie Placebo (formal: H_0 : $E[\Delta_{\text{DAS28}}(\text{neu})] \leq \Delta_{\text{DAS28}}(\text{Placebo})$),

H_1 : Die neuartige Therapie wirkt bezüglich der erwarteten Veränderung des DAS28 besser als Placebo (formal: H_1 : $E[\Delta_{\text{DAS28}}(\text{neu})] > \Delta_{\text{DAS28}}(\text{Placebo})$).

Dabei bezeichnet $E[\Delta_{\text{DAS28}}(\text{neu})]$ die erwartete Änderung des DAS28 unter der neuartigen Therapie und $E[\Delta_{\text{DAS28}}(\text{Placebo})]$ den entsprechenden Erwartungswert unter Placebo.

Das in diesem Beispiel betrachtete Testproblem ist einseitig, d. h. der Test beschränkt sich auf den Überlegenheitsnachweis der neuartigen Therapie. Regulatorische Vorgaben verlangen für gewöhnlich zusätzlich die Prüfung der Unterlegenheit (zweiseitiges Testproblem). Dieser zweite Teil des Tests lässt sich analog zum ersten Teil durchführen und wird im Folgenden nicht gesondert dargestellt.

Der statistische Hypothesentest ist ein Verfahren, um anhand der Studiendaten eine Entscheidung zu treffen, die entweder in Form der Ablehnung oder Beibehaltung der Nullhypothese besteht. Ziel ist die Widerlegung der Nullhypothese und damit die Annahme der Alternativhypothese. Dabei gilt es, 2 Arten möglicher Fehlentscheidungen zu berücksichtigen. Der Fehler 1. Art wird begangen, wenn die Nullhypothese abgelehnt wird, obwohl sie der Wahrheit entspricht. Die maximal erlaubte Wahrscheinlichkeit des Fehlers 1. Art wird durch das Signifikanzniveau α vorgegeben. Diese Irrtumswahrscheinlichkeit beträgt bei zweiseitigen Nullhypothesen üblicherweise 5%, bei einseitigen Nullhypothesen 2,5%. Ein Fehler 2. Art wird begangen, wenn die Nullhypothese fälschlicherweise beibehalten wird. Die Wahrscheinlichkeit für diese Art von Fehler wird mit β bezeichnet. Das Komplement $1-\beta$ entspricht der Wahrscheinlichkeit, die Nullhypothese abzulehnen,

Hier steht eine Anzeige.



wenn die Alternativhypothese tatsächlich wahr ist, und wird Power oder Güte eines Tests genannt.

Gemäß guter klinischer Praxis sowie aufgrund regulatorischer Vorgaben werden Studien so geplant, dass die Wahrscheinlichkeit für den Fehler 1. und 2. Art kontrolliert ist. Dazu legt man bei der Planung der Studie das Signifikanzniveau α fest. Die Power lässt sich über eine geeignete Fallzahlplanung sichern. Dazu ist es zunächst erforderlich, das spätere Studienergebnis oder zumindest den minimalen als klinisch relevant zu erachtenden Effekt abzuschätzen. So könnte man in unserem Beispiel davon ausgehen, dass es unter Placebo zu keiner Änderung des DAS28 kommt, d. h. $E[\Delta_{DAS28}(\text{Placebo})] = 0$. Unter der neuartigen Therapie dagegen erwartet man im Mittel eine Senkung des DAS28 um 0,6 Einheiten, d. h. $E[\Delta_{DAS28}(\text{neu})] = 0,6$. Der Therapieeffekt δ beträgt

$$\delta := E[\Delta_{DAS28}(\text{neu}) - E[\Delta_{DAS28}(\text{Placebo})]] = 0,6$$

Falls dieser erwartete Therapieeffekt tatsächlich besteht, so soll er mit einer Power von 80% nachgewiesen werden können.

In einem klassischen Design wird die Fallzahl im Vorfeld im Studienprotokoll festgelegt. Im Beispiel ergibt sich daraus bei einer Wahl von $\alpha=2,5\%$ eine notwendige Fallzahl von 252 Patienten pro Therapiegruppe. Nach Erhebung aller Daten wird die finale statistische Auswertung gemäß dieser verbindlichen Vorgaben durchgeführt. Der aus den Daten berechnete p-Wert liefert die Entscheidung: Wenn der p-Wert kleiner oder gleich dem vorgegebenen Signifikanzniveau α ist, so kann die Nullhypothese abgelehnt werden, und das Ziel der Studie ist erreicht. Ein solches Design ohne Zwischenauswertung wird als einstufig bezeichnet.

Adaptiv-sequenzielle Designs

Die eingangs dargelegte Tendenz zur Untersuchung kleinerer Patientenpopulationen bedingt wegen des geringeren Rekrutierungspotenzials oft längere Studienlaufzeiten. Lange Studiendauern sind jedoch aus ethischer und ökonomischer Sicht problematisch: Im Fall einer eviden-

ten Überlegenheit einer neuen Therapie möchte man diese den Patienten möglichst zeitnah verfügbar machen. Umgekehrt erscheint es sinnvoll, im Falle eines nicht vorhandenen oder nicht relevanten Effekts die Untersuchung eines Therapieansatzes zeitnah zu beenden. Dem kann durch eine adaptiv-sequenzielle Studienplanung adäquat entsprochen werden, die es ermöglicht, in der laufenden Studie Zwischenanalysen durchzuführen und Entscheidungen über den weiteren Studienverlauf zu treffen. Für die Durchführung von Zwischenanalysen bedarf es jedoch strikter Regeln, denn in einer mehrstufigen Studie wird die Nullhypothese nicht nur einmal, sondern in jeder Zwischenanalyse getestet. Es besteht somit ein multiples Testproblem. Fehler 1. Art können in jeder Zwischenauswertung begangen werden und addieren sich schlimmstenfalls auf. Dann ist bei unsachgemäßer Durchführung die Gesamtfehlerwahrscheinlichkeit nicht mehr kontrolliert. Man spricht von α -Inflation. Darüber hinaus ist es problematisch, wenn Ergebnisse einer Zwischenauswertung bekannt werden und so den Studienablauf beeinflussen. Zum Beispiel könnte der scheinbare Vorteil einer Studientherapie zu Änderungen im Rekrutierungsmuster und so zum Scheitern der Studie führen. Dem muss im statistischen Design sowie bei der Organisation der Studie Rechnung getragen werden.

» Zwischenanalysen sind aus ethischen und ökonomischen Gründen sinnvoll

Historisch gesehen gehören gruppensequenzielle Designs zu den ersten validen mehrstufigen Designs [12, 13]. Dabei wird die Gesamtfallzahl in Stufen aufgeteilt, sog. Sequenzen. Nach jeder Sequenz findet eine Zwischenanalyse statt, in der sich entscheidet, ob die Studie weitergeführt oder frühzeitig mit Erfolg oder Misserfolg beendet wird. Diese Entscheidungsregeln, der Stichprobenumfang pro Sequenz und die maximale Anzahl der Sequenzen sind im Vorfeld der Studie festzulegen und können in der laufenden Studie nicht mehr verändert werden. Die genaue Fallzahl einer mehrstufig geplan-

ten Studie ergibt sich erst im Studienverlauf und hängt von der Anzahl der durchgeführten Sequenzen ab. Sie kann im Einzelfall entweder größer oder kleiner sein als die Fallzahl eines entsprechenden klassischen einstufigen Designs. Eine zentrale Kenngröße gruppensequenzieller Designs ist deshalb die durchschnittlich benötigte Fallzahl. Dabei handelt es sich um diejenige Fallzahl, die im Mittel benötigt würde, wenn die gruppensequenzielle Studie in identischer Weise in großer Zahl wiederholt durchgeführt würde. Es stellt sich heraus, dass durch Verwendung eines mehrstufigen Designs im Vergleich zum klassischen einstufigen Design im Mittel Fallzahl eingespart wird. Würde in dem im Abschnitt „Klassische Designs in konfirmatorischen klinischen Studien“ geschilderten klinischen Beispiel die Studie beispielsweise nicht einstufig, sondern in 2, 3 bzw. 5 gleich großen Stufen nach Pocock [13] geplant, so reduziert sich die Fallzahl im Mittel von 252 auf 215, 206 bzw. 201 Patienten pro Gruppe. Pläne mit mehr als 5 Stufen führen in praktisch relevanten Situationen zu keiner weiteren Fallzahlreduktion. Somit begegnet eine sequenzielle Studiendurchführung den Herausforderungen kleiner Fallzahlen in zweierlei Hinsicht: Zum einen stellt sie eine methodisch valide Möglichkeit dar, während der laufenden Studie Zwischenanalysen durchzuführen. Zum anderen geht sie im Mittel mit einer Verkürzung der Studienlaufzeit einher.

Adaptive Designs stellen eine Weiterentwicklung der gruppensequenziellen Designs dar. Hierbei ist es zusätzlich möglich, die Studie in einer Zwischenanalyse unter Verwendung der bisher beobachteten Ergebnisse neu zu planen, ohne die Integrität der Studie zu gefährden.

Eine typische Form der Adaption ist die datenbasierte Fallzahlneuplanung. So kann die Fallzahl und damit die Studienlaufzeit während der laufenden Studie reduziert/erhöht werden, wenn der beobachtete Therapieeffekt deutlich stärker/schwächer ausfällt als ursprünglich angenommen. Die genaue Größenordnung der Fallzahlanpassung ergibt sich datenabhängig auf Grundlage der Ergebnisse der Zwischenanalyse und muss deshalb durch ein unabhängiges Gremium festgestellt werden, um auszuschließen, dass

vorläufige Studienergebnisse den weiteren Studienablauf beeinflussen. Da Fallzahl und Studienlaufzeit bei adaptiver Studienplanung variabel sind, ist zu beachten, dass sich der genaue finanzielle Bedarf ebenfalls erst im Laufe der Studie ergibt. Das verlangt sowohl dem Sponsor als auch potenziellen Förderern der Studie eine flexible Finanzplanung ab. Um die finanzielle Planungssicherheit zu erhöhen, wird im Studienprotokoll üblicherweise eine maximale Fallzahl n_{\max} festgeschrieben, die bei Fallzahlanpassungen nicht überschritten werden darf. So ist gewährleistet, dass ein im Vorfeld festgelegter maximaler finanzieller Bedarf nicht überschritten wird.

» Flexible Studiendurchführung erhöht die Erfolgsaussichten der Studie

Insgesamt können durch das Instrument der adaptiven Fallzahlanpassung anfängliche Planungsunsicherheiten im Laufe der Studie ausgeglichen werden. Dies erhöht die Erfolgsaussichten der Studie und ermöglicht auch eine laufende Feasibility-Beurteilung bezüglich vorhandener Ressourcen. In klassischen adaptiven Designs [1, 11] wird angenommen, dass die Stichproben der einzelnen Sequenzen voneinander unabhängig sind und dass die Anzahl der Zwischenanalysen im Vorfeld festgelegt ist. Diese Forderung wird von Brannath et al. [4] abgeschwächt, bei deren Ansatz auch eine flexible Anzahl von Zwischenanalysen möglich ist. Adaptive Designs mit abhängigen Stufen wurden untersucht von Hommel et al. [10], Götte et al. [8] und Schmidt et al. [15]. Das gewählte adaptive Design legt fest, wie die Ergebnisse der Sequenzen miteinander kombiniert werden und welche Anpassungen möglich sind. Das grundlegende Konzept ist dabei bereits bei der Planung der Studie festzulegen, um die Einhaltung der Gesamtfehlerwahrscheinlichkeit zu garantieren.

Eine Methodik, die weitreichende, auch ungeplante Designänderungen ermöglicht, ist das Conditional Rejection Principle [14]. Eine universelle Methodik für den simultanen adaptiven Test mehrerer Nullhypothesen liefert Hommel

Z Rheumatol 2015 · 74:106–112 DOI 10.1007/s00393-014-1450-8
© Springer-Verlag Berlin Heidelberg 2015

J. Gerß · M. Eveslage · A. Faldum · R. Schmidt

Moderne Studiendesigns und Analysemethoden in der klinischen Forschung

Zusammenfassung

Hintergrund. Der medizinische Fortschritt wird in Zukunft zu einer veränderten Studienlandschaft führen. Es gilt, kleine Fallzahlen und knappe Ressourcen optimal zu nutzen.

Methoden. Klassische einstufige Designs klinischer Studien mit fest vorgegebener Fallzahl erlauben keine Zwischenauswertungen oder Modifikationen des Designs. Adaptive Verfahren hingegen ermöglichen Flexibilität unter Wahrung der etablierten Gütekriterien. Neben Fallzahlanpassungen und Auswahl von Subkollektiven sind auch weitere reichende Designmodifikationen möglich, um die Studie an Veränderungen im Studienumfeld anzupassen. So kann die Erfolgsaussicht der Studie erhöht werden. Eine weitere Alternative oder Ergänzung zu klassischen Designs

stellen Bayes-Verfahren dar, mit denen bestehendes Vorwissen effizienter als bisher in klinische Studien eingebracht werden kann. Dies führt zu einem deutlichen Informationsgewinn. Daneben erlauben es Bayes-Verfahren insbesondere, Ergebnisse einer statistischen Auswertung transparent darzustellen. **Schlussfolgerung.** Zur Lösung neuer Herausforderungen an klinische Studien bieten sich wegen ihrer großen Flexibilität und Effizienz moderne Verfahren wie adaptive Designs und Bayes-Designs an. Beide Ansätze können auch miteinander verknüpft werden.

Schlüsselwörter

Hypothesentest · Zwischenanalyse · Adaptive Design · Bayes · Klinische Studie

Modern study designs and analysis methods in clinical research

Abstract

Background. Progress in the field of medical research requires further development of clinical trial methodology to overcome the challenges resulting from small patient populations and restricted resources.

Methods. Classical single-stage designs with fixed sample sizes do not allow for interim analyses or design modifications. In contrast, adaptive designs adhere to established quality criteria while providing flexibility when conducting a clinical trial. In the face of new discoveries or information collected in the course of a trial, sample size adjustment, the selection of the target population and further design modifications can be performed. This enhances the chance of success of a clinical trial. Besides adaptive designs, classical

approaches may be replaced or complemented by Bayesian methods. In a Bayesian approach prior knowledge can be efficiently included and hence the amount of information utilized in statistical analyses is increased. Furthermore, Bayes procedures allow the results of a statistical evaluation to be displayed very clearly.

Conclusion. Modern approaches, such as adaptive designs and Bayesian designs overcome the challenges in clinical research due to enhanced flexibility and efficiency. In addition, both approaches can be combined.

Keywords

Hypothesis testing · Interim analysis · Adaptive design · Bayes · Clinical trial

[9], wobei in der Zwischenanalyse Hypothesen fallengelassen und neue hinzugenommen werden können. Die Änderung eines primären Endpunktes erscheint jedoch nur in Ausnahmefällen vertretbar und bedarf grundsätzlich einer besonderen Rechtfertigung, da gewährleistet sein muss, dass die Interpretierbarkeit der konfirmatorischen Studienergebnisse uneingeschränkt möglich ist. Als vertretbarer Eingriff mag die Adaption des multiplen Testproblems angesehen werden, wenn etwa im Rahmen des hierar-

chischen Testens in der Zwischenanalyse die hierarchische Reihenfolge der Hypothesen datengesteuert geändert wird und diese Änderung inhaltlich begründbar ist.

Seamless-Designs ermöglichen die datenabhängige Auswahl von Behandlungsarmen in Zwischenanalysen und stellen somit eine Möglichkeit zur Kombination verschiedener klinischer Phasen in einem einheitlichen Studienkonzept dar (Phase-II/III-Studien, [2]). Die Kombination von Phase-II- und -III-Studiendesigns erscheint einerseits sinnvoll, da

sich sowohl die Fallzahl als auch der organisatorische Aufwand idealerweise verringern lassen. Dem steht jedoch gegenüber, dass bei getrennt durchgeführten Studienphasen etwaige neue Erkenntnisse aus einer Phase II Eingang in die Studienplanung der Phase III finden können, was wiederum das Risikopotenzial für Patienten im Vergleich zur kombinierten Studiendurchführung verringern kann, da die vorhandenen Phase-II-Informationen umfassender genutzt werden. Die Entscheidung zu entweder kombinierter oder zu klassischer, phasenweise getrennter Studiendurchführung ist somit stark kontextabhängig und im Einzelfall zu diskutieren.

Vor dem Hintergrund individualisierter Medizin erscheinen adaptive Enrichment-Designs vielversprechend [5]. Hier ist es möglich, Subgruppen, in denen eine besonders ausgeprägte Wirksamkeit erwartet wird, zu untersuchen und in der Zwischenauswertung datenabhängig die erfolgversprechendste Zielpopulation auszuwählen.

In der Praxis erfolgt die Umsetzung eines adaptiven Designs nach strengen Regeln: Das in der Studie Anwendung findende adaptive Design ist im Studienprotokoll zu spezifizieren und mit publizierter Literatur zu belegen. Der Entscheidungsbaum, wie die Studie in Abhängigkeit von den Ergebnissen der Zwischenanalyse adaptiert werden soll, ist datenunabhängig im Vorfeld der eigentlichen Zwischenanalyse festzulegen und ist während der Studienlaufzeit vertraulich zu behandeln, um die Integrität der Studie zu wahren. Die sich aus der Zwischenanalyse ergebenden Konsequenzen für den Studienablauf werden durch ein unabhängiges Gremium festgestellt und daraufhin Sponsor und Studienkommission mitgeteilt. Dabei werden jedoch keinerlei Analyseergebnisse kommuniziert, die Rückschluss auf etwaige Therapieeffekte zulassen. Der Sponsor informiert schließlich die zuständigen Ethik-Kommissionen und Behörden, nach deren Votum die Adaptationen in einem Amendment zum Studienprotokoll festgehalten und in der Studie entsprechend implementiert werden. Eine besondere Herausforderung stellt die adäquate Berücksichtigung von Patienten dar, die während einer laufenden Zwischenanalyse rekrutiert werden [6]. Der

logistische Aufwand in adaptiven Studien wird durch eine höhere Erfolgsaussicht und Flexibilität der Studien kompensiert.

Aufgrund ihrer zunehmenden Verbreitung haben adaptive Designs große Aufmerksamkeit sowohl im statistischen, pharmazeutischen als auch regulatorischen Umfeld gewonnen. In ihrem Draft der „Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics“ erkennt die U.S. Food and Drug Administration adaptive Designs als valides und nützliches Werkzeug für klinische Studien an und liefert eine umfassende Diskussion der sich ergebenden Chancen und Herausforderungen. Für die Planung und Analyse gängiger adaptiver Designs stehen in der Praxis geeignete Softwarepakete zur Verfügung.

Bayes-Designs

Bayes-Statistik

Statistische Verfahren wie der klassische Hypothesentest fallen unter den Oberbegriff der sog. frequentistischen Statistik. Diese Bezeichnung geht auf die Vorstellung zurück, dass beispielsweise im Fall einer Studie zum Signifikanzniveau $\alpha=2,5\%$ bei einer hypothetischen mehrfachen Wiederholung dieser Studie in 2,5% der Fälle ein Fehler 1. Art begangen würde. Das heißt, eine Wahrscheinlichkeit ist eine relative Häufigkeit bei einer mehrfachen Wiederholung eines Experiments. Im Rahmen der Bayes-Statistik werden Wahrscheinlichkeiten in einem anderen Sinn verstanden und genutzt, um das vorliegende Wissen über einen bestimmten Sachverhalt auszudrücken [7].

Im klinischen Beispiel aus dem Abschnitt „Klassische Designs in konfirmatorischen klinischen Studien“ wird der Therapieeffekt durch die Größe δ beschrieben. Man kann sich nun Gedanken machen, welches Wissen über den Therapieeffekt δ vorliegt, und zwar zunächst zu einem Zeitpunkt, bevor die eigentliche Studie durchgeführt wird. Das Vorwissen könnte auf früheren Studien oder Registerdaten beruhen, in denen ein gleiches oder auch unterschiedliches Patientenkollektiv wie in der aktuellen Studie mit einer ähnlichen Therapie behandelt wurde, oder es könnte aus einer Befra-

gung von klinischen Experten über das erwartete Studienergebnis stammen. Das vorliegende Vorwissen über den Therapieeffekt δ wird in Form einer Wahrscheinlichkeitsverteilung dargestellt. Die **Abb. 1** zeigt eine mögliche sog. A-priori-Verteilung des Therapieeffekts δ . Wenn nun eine neue klinische Studie durchgeführt wird, so wird das Vorwissen über den Therapieeffekt δ ergänzt durch zusätzliche Informationen, die im Rahmen der aktuellen Studie gesammelt werden. Die zusätzlichen Informationen führen in der Regel dazu, dass sich das Wissen über den Therapieeffekt δ verbessert. Dieses verbesserte Wissen wird ebenfalls durch eine Wahrscheinlichkeitsverteilung dargestellt, die sog. A-posteriori-Verteilung (**Abb. 1**).

» Bayes-Verfahren erlauben die effektive Nutzung von verfügbarem Vorwissen bei der Auswertung

Anhand der A-posteriori-Verteilung können entsprechend dem aktuellen Wissen nach Durchführung der Studie folgende Einschätzungen über den Therapieeffekt δ abgegeben werden:

- Der DAS28 wird unter der neuartigen Therapie im Schnitt um 0,82 Einheiten mehr gesenkt als unter Placebo [A-posteriori-Erwartungswert $E_{\text{post}}(\delta)=0,82$].
- Mit einer Wahrscheinlichkeit von 96% ist die neuartige Therapie bezüglich der Senkung des DAS28 besser als eine Behandlung mit Placebo.
- Ein Therapieeffekt δ in einem Bereich von $\pm 0,6$ um 0 wird als nicht klinisch relevant angesehen (s. oben). Dann gilt, dass beide Therapien in diesem Sinn mit einer Wahrscheinlichkeit von 32% gleichwertig sind.

Die letzten beiden Punkte zeigen eine besondere Eigenschaft von Bayes-Verfahren:

- ▣ Sie erlauben es, das vorliegende Wissen über einen interessierenden Sachverhalt messbar zu quantifizieren und die Ergebnisse transparent darzustellen.

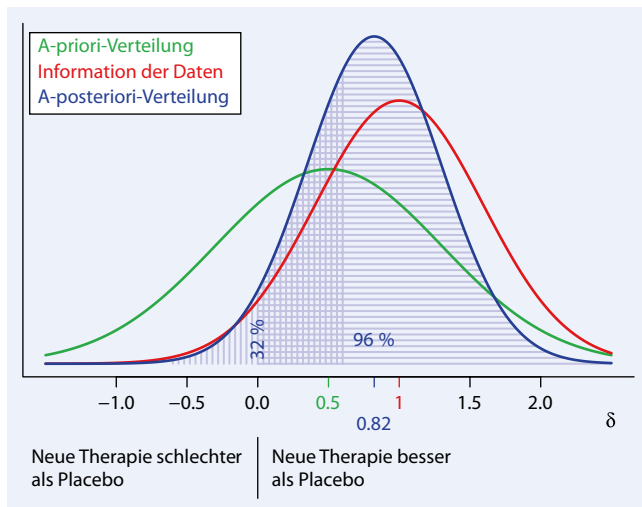


Abb. 1 ◀ A-priori-Verteilung (Erwartungswert = 0,5), Information der Daten der aktuellen Studie (Mittelwert = 1) und resultierende A-posteriori-Verteilung des Therapieeffekts δ

Diese Eigenschaft, die frequentistische Verfahren in dieser Form nicht aufweisen, eröffnet einige Möglichkeiten, Bayes-Verfahren in klinischen Studien gewinnbringend einzusetzen.

Bayes-Verfahren in klinischen Studien

Die oben dargestellten grundlegenden Prinzipien der Bayes-Statistik können in klinischen Studien in ganz unterschiedlicher Weise eingesetzt werden. Dabei kann man 2 alternative Strategien verfolgen.

- Eine erste Strategie besteht in einer statistischen Schlussweise, in der man vollständig und ausschließlich auf Verfahren der Bayes-Statistik zugreift, die anstelle eines frequentistischen Hypothesentests oder anderer klassischer Verfahren eingesetzt werden.
- Alternativ kann man eine Strategie verfolgen, in der die primäre Auswertung einer klinischen Studie mit einem Hypothesentest durchgeführt wird, ggf. mit einem mehrstufigen Studiendesign. Das statistische Konzept wird ergänzt durch Bayes-Verfahren, um zu bestimmten Zeitpunkten während der Durchführung der Studie das aktuelle Wissen bezüglich der Wirksamkeit oder Sicherheit der Studientherapie angemessen zu quantifizieren und daraus entsprechende Konsequenzen zu ziehen.

Im Folgenden werden zunächst einige vollständige Bayes-Designs klinischer

Studien und anschließend einige Konzepte vorgestellt, in denen Bayes-Verfahren ergänzend eingesetzt werden [3, 16].

In einer frühen Phase der Arzneimittelentwicklung werden Dosisescalationsstudien durchgeführt, um die maximale tolerierte Dosis (MTD) eines neu entwickelten Medikaments zu bestimmen. In einem verbreiteten klassischen Ansatz werden dazu einzelnen Probanden oder kleinen Kohorten von Probanden schrittweise immer größere Dosen des Medikaments verabreicht, um diejenige Dosis zu ermitteln, bei der die Wahrscheinlichkeit einer dosislimitierenden Toxizität (DLT) eine gewisse angestrebte Größe von z. B. 30% hat. Wenn dabei die ersten verabreichten Dosen noch weit unterhalb der MTD liegen, dauert es relativ lange, bis die gesuchte MTD erreicht wird. Dieses Problem wird mit einem Bayes-Verfahren verhindert, der „Continual Reassessment Method“. Hier werden ggf. verfügbare Vorinformationen sowie die im Laufe der Studie anfallenden Informationen genutzt, um das aktuelle Wissen über die MTD kontinuierlich zu aktualisieren. Mit jedem neuen Probanden wird die A-posteriori-Verteilung der MTD ermittelt. So kann man den folgenden Probanden Dosen verabreichen, die sich deutlich schneller der MTD annähern.

In anderen Studien der frühen Phase einer Arzneimittelentwicklung besteht das Ziel darin, aus mehreren alternativen Varianten einer Therapie (beispielsweise mehreren Dosen oder Darreichungsformen eines Medikaments) diejenige mit der größten Erfolgsaussicht in einer spä-

teren konfirmatorischen Studie auszuwählen (Play-the-winner-Studien). Bayes-Verfahren bieten hier nicht nur die Möglichkeit, Vorwissen einzubringen. Darüber hinaus kann man eine direkte Aussage darüber treffen, welche Variante der Therapie nach aktuellem Wissen die beste ist und mit welcher Wahrscheinlichkeit dies der Fall ist. Ist die Sicherheit groß genug, so wird die entsprechende Variante der Therapie für eine sich anschließende konfirmatorische Studie ausgewählt.

Im klinischen Beispiel aus dem Abschnitt „Klassische Designs in konfirmatorischen klinischen Studien“ wurde ein frequentistischer Hypothesentest zu einem Signifikanzniveau von $\alpha=2,5\%$ angewendet. Alternativ zu einem solchen klassischen Ansatz könnte man die Studie in einem entsprechenden vollständigen Bayes-Design durchführen. Dazu definiert man die Studie als erfolgreich, wenn gemäß der A-posteriori-Verteilung des Therapieeffekts δ die Wahrscheinlichkeit für die Überlegenheit der neuartigen Therapie gegenüber Placebo mindestens 97,5% beträgt. Andernfalls gilt die Studie als Misserfolg und der Nachweis der Überlegenheit der neuartigen Therapie als misslungen – auch wenn eine gewisse Überlegenheit gegenüber Placebo besteht, beispielsweise mit einer Wahrscheinlichkeit von 70%. Ein Bayes-Verfahren als Ersatz für einen klassischen Hypothesentest hat unter anderem den Vorteil, dass man neben den in einer klinischen Studie neu erhobenen Daten Vorwissen einbringen kann. Andererseits besteht das Problem, dass frequentistische Gütekriterien wie der Fehler 1. Art unter Umständen weniger streng kontrolliert werden als in einem Hypothesentest. In frequentistischem Sinn besteht somit die Gefahr eines Anstiegs der Fehlerwahrscheinlichkeit, und es können falsch positive Ergebnisse entstehen. Dies führt dazu, dass gemäß aktueller Guidelines für konfirmatorische Phase-III-Studien Bayes-Verfahren nur in Ausnahmesituationen angewandt werden.

Neben den bisher beschriebenen vollständigen Bayes-Designs können Bayes-Verfahren auch ergänzend in einer klinischen Studie eingesetzt werden, in der die primäre Auswertung mit einem frequentistischen Hypothesentest erfolgt. In einer

klassischen randomisierten Studie erfolgt die Randomisierung der Patienten mit einem festen Verhältnis (häufig 1:1), das während der gesamten Rekrutierungszeit konstant bleibt. Stattdessen kann mit einer „Bayesian response adaptive randomization“ ein dynamisches Randomisierungsverfahren angewendet werden. Dabei beobachtet man im Laufe einer Studie kontinuierlich, wie die einzelnen Patienten auf die Studientherapie ansprechen. So ergibt sich jeweils ein aktuelles Bild darüber, welche Studientherapie die erfolgreichste ist, inklusive einer Aussage über die Wahrscheinlichkeit ihrer Überlegenheit. Die (aktuell ermittelten) Wahrscheinlichkeiten der Überlegenheit jeder einzelnen Therapie werden nun so genutzt, dass der nächste rekrutierte Studienpatient mit einer höheren Wahrscheinlichkeit eine überlegene Studientherapie zugewiesen bekommt als eine unterlegene Therapie. Wird das Verfahren in dieser Form lange fortgesetzt, so verschiebt sich das Randomisierungsverhältnis erwartungsgemäß nach und nach immer mehr zugunsten der überlegenen Therapie. Ab einem bestimmten Zeitpunkt werden unterlegene Therapien gar nicht mehr verabreicht. Am Ende der Studie kann die statistische Auswertung mit einem angemessen adaptierten frequentistischen Hypothesentest erfolgen.

Wie im Abschnitt „Adaptiv-sequenzielle Designs“ beschrieben, kann eine klinische Studie im Rahmen eines adaptiven Designs nach einer Zwischenauswertung (datenabhängig) neu geplant werden. Zum Beispiel wird die Zielpopulation neu definiert (Enrichment-Design), oder es erfolgt eine Neuplanung der Fallzahl. In welcher Weise die Studie in Abhängigkeit von den Ergebnissen der Zwischenauswertung adaptiert wird, ist vor Beginn der Studie festzulegen. Dabei hat man relativ große Freiheiten und muss keineswegs nur auf Verfahren der frequentistischen Statistik zurückgreifen. Stattdessen können die entsprechenden Algorithmen mithilfe von Bayes-Verfahren definiert werden. So kann man anhand der A-posteriori-Verteilung des Therapieeffekts entscheiden, welche Subgruppe die erfolgversprechendste Zielpopulation darstellt. Die Neuplanung der Fallzahl kann aufgrund der sog. prädiktiven Power erfolgen. Das

heißt, man nutzt im Rahmen eines Bayes-Verfahrens das gesamte zum Zeitpunkt der Zwischenauswertung vorliegende Wissen über den Therapieeffekt in Form der aktuellen A-posteriori-Verteilung.

Ein dritter großer Einsatzbereich ergänzender Bayes-Verfahren in einer klinischen Studie stellt die Auswertung der Sicherheit und Verträglichkeit der Studientherapie etwa durch das Data Monitoring Committee (DMC) dar. Hier geht es nicht darum, vorab Hypothesen aufzustellen und bei der Entscheidung über deren Ablehnung oder Beibehaltung den Fehler 1. Art zu kontrollieren. Stattdessen muss man Sicherheitsrisiken schon bei ersten Anzeichen aufdecken, und es ist hilfreich, wenn eine quantitative Aussage über die Wahrscheinlichkeit vorliegt, mit der ein bestimmtes Risiko besteht. Bayes-Verfahren bieten diese Möglichkeiten, indem man im Laufe der Studie regelmäßig die A-posteriori-Wahrscheinlichkeit der entsprechenden Sachverhalte ermittelt.

Fazit für die Praxis

- Der medizinische Fortschritt verlangt die Entwicklung neuer statistischer Methoden, die über klassische Studiendesigns hinausgehen.
- Mehrstufige adaptive Designs bieten die Möglichkeit einer flexiblen Studiendurchführung mit hohen Erfolgsaussichten.
- Bayes-Verfahren erlauben die effektive Nutzung von verfügbarem Vorwissen bei der Auswertung einer Studie sowie die transparente Darstellung der Ergebnisse.

Korrespondenzadresse



Dr. rer. nat. J. Gerß
Dipl.-Stat.
 Institut für Biometrie und
 Klinische Forschung (IBKF),
 Westfälische Wilhelms-
 Universität Münster
 Schmeddingstr. 56,
 48149 Münster
 joachim.gerss@ukmuenster.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. J. Gerß, M. Eveslage, A. Faldum und R. Schmid geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

Literatur

1. Bauer P (1989) Multistage testing with adaptive design. *Biom Inform Med Biol* 20(4):130–148
2. Bauer P, Kieser M (1999) Combining different phases in the development of medical treatments within a single trial. *Stat Med* 18:1833–1848
3. Berry SM, Carlin BP, Lee JJ, Müller P (2010) Bayesian adaptive methods for clinical trials. Chapman&Hall/CRC, Boca Raton
4. Brannath W, Posch M, Bauer P (2002) Recursive combination tests. *J Am Stat Assoc* 97:236–244
5. Brannath W, Zuber E, Branson M et al (2009) Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med* 28:1445–1463
6. Faldum A, Hommel G (2007) Strategies for including patients recruited during interim analysis of clinical trials. *J Biopharm Stat* 17:1211–1225
7. Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman&Hall/CRC, Boca Raton
8. Götte H, Hommel G, Faldum A (2009) Adaptive designs with correlated test statistics. *Stat Med* 28:1429–1444
9. Hommel G (2001) Adaptive modifications of hypotheses after an interim analysis. *Biom J* 43:581–589
10. Hommel G, Lindig V, Faldum A (2005) Two-stage adaptive designs with correlated test statistics. *J Biopharm Stat* 15:613–623
11. Lehman W, Wassmer G (1999) Adaptive sample size calculations in group sequential trials. *Biometrics* 55:1286–1290
12. O'Brien P, Fleming TR (1979) A multiple testing procedure for clinical trials. *Biometrics* 35:549–556
13. Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64:191–199
14. Schäfer H, Müller H (2001) Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Stat Med* 20:3741–3751
15. Schmidt R, Faldum A, Witt O, Gerß J (2014) Adaptive designs with arbitrary dependence structure. *Biom J* 56:1–21
16. Spiegelhalter DJ, Abrams KR, Myles JP (2004) Bayesian approaches to clinical trials and health-care evaluation. Wiley, New York