

TU DORTMUND

FALLSTUDIEN I - SOMMERSEMESTER 2025

Projekt 2: Vergleich zweier Verteilungen: Reduktion von Epilepsie-Symptomen

Dozenten:

JProf. Dr. Dennis Dobler

Loreen Sabel

Dr. Susanne Frick

Prof. Dr. Philipp Doeblen

Verfasser: Quang Vinh Nguyen

Gruppenmitglieder:

Alain Kauth, Ibrahim Rahman

11. Mai 2025

Inhaltsverzeichnis

1. Einleitung	1
2. Problemstellung	1
2.1. Datenmaterial	1
2.2. Ziele des Projekts	2
3. Statistische Methoden	3
3.1. Grundlagen des Hypothesentestens	3
3.2. U-Test nach Wilcoxon, Mann und Whitney	4
3.3. Q-Q-Plot	6
3.4. Nichtparametrische Maßzahl der Effektstärke nach Cliff	6
4. Statistische Auswertung	7
4.1. Deskriptive Statistik	7
4.2. Hypothesentests	8
5. Zusammenfassung	11
Literatur	13
A. Zusätzliche Tabellen	14
B. Zusätzliche Grafiken	14

1. Einleitung

Epilepsie ist eine weit verbreitete neurologische Erkrankung, die durch wiederkehrende Krampfanfälle gekennzeichnet ist. Die Behandlung von Epilepsie erfolgt in der Regel pharmakologisch, wobei Antiepileptika eine zentrale Rolle spielen. Dennoch ist die Wirksamkeit verschiedener Medikamente unterschiedlich und nicht alle Patienten sprechen auf eine Behandlung gleich gut an.

Im Rahmen dieses Projekts wird die Wirksamkeit des Antiepileptikums Progabid im Vergleich zum Placebo basierend auf von Thall und Vail (1990) berichteten Daten überprüft. Dabei werden auch die empirischen Verteilungen der relevanten Variablen berücksichtigt, insbesondere die Annahme einer Normalverteilung. Die Testergebnisse zeigen, dass die Wirkung von Progabid statistisch nicht signifikant ist.

Der vorliegende Bericht gliedert sich in fünf Kapitel. In Kapitel 2 erfolgt zunächst eine Beschreibung der Studie sowie der erhobenen Daten. Darüber hinaus wird eine formale Vorstellung der konkreten Ziele des Berichts gegeben. In Kapitel 3 werden dann der QQ-Plot und der U-Test erläutert. In Kapitel 4 wird der Analyseprozess diskutiert. Die Ergebnisse der Testverfahren und Schlussfolgerungen werden schließlich in Kapitel 5 zusammengefasst.

2. Problemstellung

In diesem Kapitel werden die Projektziele sowie das Datenmaterial ausführlich dargelegt.

2.1. Datenmaterial

Der Datensatz wurde im Bericht von Thall und Vail (1990) als Beispiel verwendet, stammt jedoch aus einer klinischen Studie von Leppik et al. (1985). Es gab 59 Epilepsie-Patienten, die randomisiert auf zwei Gruppen aufgeteilt wurden. Davon erhielten 31 Patienten das Antiepileptikum Progabid, während die restlichen 28 Patienten ein Placebo erhielten. Im Verlauf der Behandlung besuchten die Patienten die Klinik viermal, wo die Anzahl von Krämpfen innerhalb der vorangegangenen zwei Wochen berichtet wurden. Darüber hinaus wurden im Datensatz auch die Anzahl von Krämpfen innerhalb von acht Wochen vor der Behandlung (Baseline) und das Alter jedes Patienten erfasst. Alle Angaben sind in der Tabelle 1 aufgelistet.

Tabelle 1: Übersicht über Variablen im Datensatz

Variable	Bedeutung	Skalenniveau	Ausprägungen
unabhängige Variablen			
Patient	Identifikationsnummer der behandelten Person	nominal	{1, ..., 59}
Alter	Alter in Jahren	metrisch	natürliche Zahlen
Base	Anzahl der Krämpfe innerhalb der Baseline	metrisch	natürliche Zahlen
Behandlung	Progabid erhalten	dichotom	{1=ja, 0=nein}
Besuch	Klinikbesuch (von 1 = erster bis 4 = letzter)	nominal/ordinal	{1, 2, 3, 4}
abhängige Variablen			
Anzahl	Anzahl der Krämpfe zwischen zwei Besuchen	metrisch	natürliche Zahlen

Der Datensatz beinhaltet die Anzahl der Anfälle zu verschiedenen Zeitpunkten. So lässt sich die Wirksamkeit der Behandlung im Zeitverlauf sinnvoll beobachten. In den zugänglichen Dokumenten wird auf die Datenerhebung nicht hingewiesen. Deswegen erfolgt in diesem Bericht keine Beurteilung bezüglich der Zuverlässigkeit bzw. Genauigkeit. Allerdings werden keine Auffälligkeiten bei den Einträgen festgestellt. Es liegen auch keine fehlenden Daten vor. Die beiden Behandlungsgruppen waren relativ gleich groß. Trotzdem ist die Größe der Placebo-Gruppe mit nur 28 immer noch nicht groß genug und schränkt somit die Auswahl geeigneter Testmethoden ein.

2.2. Ziele des Projekts

Das vorliegende Projekt hat zum Ziel, die Wirksamkeit von Progabid bei der Reduktion epileptischer Symptome zu untersuchen. Dementsprechend werden die Anzahl der Anfälle zwischen beiden Gruppen für jeden Klinikbesuch sowie der Mittelwert pro Besuch auf Signifikanz getestet. Zusätzlich werden andere unabhängige Variablen wie Alter und Anzahl der Anfälle während des Baseline-Zeitraums ebenfalls nach Gruppen ausgewertet, um mögliche Unterschiede vor der Behandlung zu identifizieren. Von Interesse sind auch die Verteilungen der Anzahl der Anfälle (während der Baseline bzw. nach jedem Klinikbesuch), vor allem, ob dabei eine Normalverteilung vorliegt.

Die Variablen werden mithilfe der deskriptiven Analyse ausgewertet. Sofern bei einer Variable die Angabe hinsichtlich der Normalverteilung von Interesse ist, wird der QQ-Plot ergänzt. Zur Überprüfung der Wirksamkeit von Progabid im Vergleich zu Placebo wird der U-Test herangezogen. Später wird in Kapitel 4 gezeigt, dass die Annahme einer Normalverteilung bei Variablen der Anzahl der Anfälle nicht erfüllt ist. Da keine Normalverteilung bei den zu vergleichenden Variablen vorausgesetzt wird, zeigt sich der U-Test hierfür als eine geeignete Methode.

3. Statistische Methoden

In diesem Kapitel werden der QQ-Plot und der U-Test vorgestellt, die im Rahmen des vorliegenden Projekts Anwendung finden. Ergänzend zum U-Test werden zudem auch die Grundlagen des Hypothesentests erörtert. Die Datenverarbeitung und die Anwendung der hier aufgeführten Methoden erfolgt unter Verwendung der Software *R* mit Version 4.4.2 (Team, 2024) sowie des Pakets *ggplot2* (Wickham et al., 2016).

Der Kürze und Konsistenz halber werden in der Beschreibung aller Methoden folgenden Notationsregeln verwendet, ohne Erläuterungen explizit zu wiederholen. Zufallsvariablen (inkl. Teststatistik) werden in Großbuchstaben und Realisierungen in Kleinbuchstaben dargestellt. Z. B. sind x_i die entsprechenden Realisierungen der Variablen $X_i, i = 1, \dots, n$. Ferner lässt sich die Verteilungsfunktion der Variable X durch F_X ausdrücken, wobei es gilt $F_X(x) = Pr(X \leq x)$. Für die Standardnormalverteilung wird die Verteilungsfunktion als Φ bezeichnet.

3.1. Grundlagen des Hypothesentestens

Der Hypothesentest ist ein Verfahren zur Überprüfung einer Aussage, das auf den gesammelten Daten basiert. Die zu überprüfende Aussage muss dabei in Form von zwei Hypothesen, nämlich der Nullhypothese H_0 und der Alternativhypothese H_1 , ausgedrückt werden, die sich gegenseitig ausschließen und den Definitionsbereich des interessierenden Parameters vollständig abdecken (Genschel und Becker, 2005, S.175). Der Test besteht darin, zwischen zwei Hypothesen zu entscheiden.

Teststatistik: Die Testentscheidung wird auf Basis einer Prüfgröße getroffen, die in der Regel als Teststatistik bezeichnet wird. Die Teststatistik, definiert als eine Funktion der Stichprobenvariablen, ist ebenfalls eine Zufallsvariable (Genschel und Becker, 2005, S.180). Die konkrete Formulierung der Funktion für die Teststatistik erfolgt dabei in Abhängigkeit vom jeweiligen Testproblem. Ihr empirischer Wert (auch Realisierung genannt) wird dann aus Beobachtungen der vorliegenden Stichprobe errechnet.

Dabei ist zu betonen, dass jede Testentscheidung nur auf einer begrenzten Datenmenge aus der Grundgesamtheit bzw. der Teststatistik basiert, die als Zufallsstichprobe bzw. -variable betrachtet wird (Genschel und Becker, 2005, S.181). Deshalb lassen sich (z. B. aufgrund von Extremwerten der Stichprobe) Fehlentscheidungen nicht ausschließen.

Fehler 1. und 2. Art: Beim Hypothesentest lassen sich Fehlentscheidungen in Fehler 1. und 2. Art unterscheiden (siehe Tabelle 2).

Tabelle 2: Übersicht über mögliche Fälle beim Hypothesentest

	H_0 wahr	H_0 falsch
H_0 beibehalten	richtig	Fehler 2. Art
H_0 abgelehnt	Fehler 1. Art	richtig

Falls die Nullhypothese H_0 verworfen wird, obwohl sie wahr ist, spricht man hierbei vom Fehler 1. Art. Falls H_0 nicht verworfen wird, obwohl sie falsch ist, wird vom Fehler 2. Art gesprochen (Genschel und Becker, 2005, S.182-183).

Signifikanzniveau: Um die Wahrscheinlichkeit für den Fehler 1. Art gering zu halten, ist bei einem vorgegebenen Stichprobenumfang eine hohe Wahrscheinlichkeit für den Fehler 2. Art in Kauf zu nehmen und umgekehrt (Hedderich und Sachs, 2020, S.459). In der Regel wird nur die Wahrscheinlichkeit für den Fehler 1. Art kontrolliert, sodass sie einen bestimmten Schwellenwert nicht überschreitet. Dieser Wert wird als das Signifikanzniveau α bezeichnet (Genschel und Becker, 2005, S.186-187).

p-Wert und Entscheidungsregel: Der p-Wert gibt die Wahrscheinlichkeit für den empirischen Wert und noch extremere Werte der Teststatistik (in Richtung der Alternativhypothese) bei Gültigkeit der Nullhypothese an (Genschel und Becker, 2005, S.189). Hierbei wird der p-Wert abhängig von der Formulierung des Testproblems berechnet. Ein kleiner p-Wert deutet darauf hin, dass unter der Nullhypothese die Realisierung der Teststatistik wenig wahrscheinlich ist. Daraus ergibt sich die Entscheidungsregel, dass die Nullhypothese verworfen wird, wenn der p-Wert kleiner als das Signifikanzniveau ist.

3.2. U-Test nach Wilcoxon, Mann und Whitney

Der U-Test ist ein Hypothesentest zum Vergleich zweier unabhängiger Stichproben, ohne Voraussetzungen über Parameter der Verteilung zu machen (nichtparametrisch).

Es sei von zwei unabhängigen Stichproben auszugehen. Die erste Stichprobe bildet sich aus m Zufallsvariablen X_1, \dots, X_m , die unabhängig und identisch gemäß F_X verteilt sind. In ähnlicher Weise umfasst die zweite Stichprobe n Zufallsvariablen Y_1, \dots, Y_n , die unabhängig und identisch gemäß F_Y verteilt sind.

Zum Vergleich zwischen F_X und F_Y lassen sich die drei möglichen Testprobleme nach Hartung et al. (2009) wie folgt formulieren (siehe S.517):

- | | | | |
|-----|--|-------|---|
| (1) | $H_0 : F_X(x) \geq F_Y(x)$
für alle x | gegen | $H_1 : F_X(x) \leq F_Y(x)$ für alle x und
$F_X(x) < F_Y(x)$ für mindestens ein x |
| (2) | $H_0 : F_X(x) = F_Y(x)$
für alle x | gegen | $H_1 : F_X(x) \neq F_Y(x)$
für mindestens ein x |
| (3) | $H_0 : F_X(x) \leq F_Y(x)$
für alle x | gegen | $H_1 : F_X(x) \geq F_Y(x)$ für alle x und
$F_X(x) > F_Y(x)$ für mindestens ein x |

Dann ergibt sich für alle drei Hypothesentest die Teststatistik als

$$U(X, Y) = \sum_{i=1}^m \sum_{j=1}^n (I[Y_j < X_i] + \frac{1}{2} I[Y_j = X_i]).$$

Für die Teststatistik wird an dieser Stelle die Formel von Hedderich und Sachs (2020) vorgestellt (siehe S.570), da diese Formel mit dem gelieferten Wert der Funktion **wilcox.test** in *R* übereinstimmt. Dabei nimmt die Indikatorfunktion $I[\text{Bedingung}]$ den Wert 1 an, falls die Bedingung darin erfüllt ist. Sonst hat sie den Wert 0. Solange es keine Bindungen gibt (keine Beobachtungen mit gleichen Werten), kann die Verteilung der Teststatistik exakt berechnet werden. Allerdings erfolgt bei der Berechnung mittels **wilcox.test** standardmäßig eine Normal-Approximation, wenn die Stichproben Bindungen oder mindestens 50 Beobachtungen aufweisen. Dementsprechend wird aus der Realisierung von U der neue Wert

$$\hat{z} = \frac{u - \frac{mn}{2}}{\sqrt{\frac{mn}{12} \left(N + 1 - \frac{1}{N(N-1)} \sum_{k=1}^r (b_k^3 - b_k) \right)}} \quad \text{mit } N = m + n$$

abgeleitet. Dabei ist r die Anzahl der Bindungsgruppen und b_k die Vielfachheit der Beobachtungen in der k -ten Gruppe. Die p-Werte werden nun anhand der Standardnormalverteilung ermittelt.

Entsprechend jedem der oben genannten Testprobleme, wird H_0 zum Niveau α verworfen, falls gilt

$$(1) \quad 1 - \Phi(\hat{z}) < \alpha; \quad (2) \quad 2 \cdot (1 - \Phi(\hat{z})) < \alpha \quad \text{bzw.} \quad (3) \quad \Phi(\hat{z}) < \alpha.$$

3.3. Q-Q-Plot

Das Q-Q-Plot bietet eine Visualisierungsmethode, um zu überprüfen, ob die vorliegenden Daten aus einer bestimmten Verteilung mit der Verteilungsfunktion F_{theor} stammen. Der i -te Wert $x_{(i)}$ der aufsteigend geordneten Datenreihe $x_{(1)}, \dots, x_{(n)}$ vom Umfang n wird hierbei als das $i/(n+1)$ -Quantil der empirischen Verteilung betrachtet (Hartung et al., 2009, S.847). Daneben lassen sich die entsprechenden Quantilwerte der theoretischen Verteilung mithilfe der Quantilsfunktion $Q_{theor} = F_{theor}^{-1}$ ebenfalls ermitteln. Indem die Punkte $\left(x_{(i)}, Q_{theor}\left(i/(n+1)\right)\right), i = 1, \dots, n$ in einem xy -Koordinatensystem abgetragen werden, kann man die empirischen Quantilwerte mit den theoretischen vergleichen. Liegen die Datenpunkte alle auf der Linie $y = x$, bedeutet dies, dass die empirischen und theoretischen Quantilen jeweils den gleichen Wert aufweisen. Somit stimmen die tatsächliche und die theoretische Verteilung perfekt miteinander überein.

Ist die Annahme einer Normalverteilung zu prüfen, können die Quantile der Standardnormalverteilung eingesetzt werden. Standardmäßig in R sowie im Paket *ggplot2* wird eine Diagonale durch das erste und dritte Quartil gezogen (nicht notwendigerweise die Linie $y = x$). Je näher die Datenpunkte an dieser Diagonale liegen, desto eher wird die Annahme einer Normalverteilung unterstützt und umgekehrt.

3.4. Nichtparametrische Maßzahl der Effektstärke nach Cliff

Neben der statistischen Signifikanz ist in der Regel auch die praktische Bedeutung eines Effekts von Interesse. Zur Beurteilung der Effektstärke hat Cliff (1993) eine Maßzahl δ vorgestellt. Wenn die Daten nicht normalverteilt ist, werden nichtparametrische Effektstärken, wie Cliffs δ , insbesondere empfohlen, da sie robuste Ergebnisse liefern. Auch bei Erfüllung der Normalverteilungsannahmen hat die Verwendung nichtparametrischer Effektstärken keine Nachteile (Cliff, 1993, S.494-509).

Für zwei Stichproben vom Umfang m und n mit den jeweiligen Zufallsvariablen X_1, \dots, X_m und Y_1, \dots, Y_n , wird Cliffs δ berechnet als

$$\delta = \frac{\sum_{i=1}^m \sum_{j=1}^n (I[Y_j < X_i] - I[Y_j > X_i])}{mn}.$$

Cliffs δ hat den Wertebereich $[-1, 1]$, wobei Extremwerte auf keine Überlappung zwischen den beiden Verteilungen und 0 auf eine vollständige Überlappung hindeuten. Nach den üblichen Faustregeln wird der absolute Wert dieser Statistik als geringer (0.15), mittlerer (0.33) bzw. starker (0.47) Effekt interpretiert (Meissel und Yao, 2024, S.2-3).

4. Statistische Auswertung

Der Teil „Statistische Auswertung“ besteht aus zwei Abschnitten. Der erste Abschnitt befasst sich mit der deskriptiven Statistik des Datensatzes. Zunächst wird die Variable „Base“ analysiert, um einen Einblick in die Verteilung der Anzahl der Krämpfe vor der Behandlung zu erhalten. Da das Alter oder der vorbestehende Epilepsiestatus das Behandlungsergebnis beeinflussen können, werden anschließend potenzielle Unterschiede in diesen Variablen zwischen den beiden Patientengruppen identifiziert. Weiterhin wird der zeitliche Verlauf des Epilepsiestatus während der Behandlung betrachtet.

Im zweiten Abschnitt wird die Wirksamkeit des Antiepileptikums Progabid im Vergleich zu Placebo durch Hypothesentests überprüft. Dabei wird nicht nur die Anzahl der Krämpfe bei jedem Besuch berücksichtigt, sondern auch deren Mittelwert. Als Ergänzung zu Hypothesentests werden die Effektstärken ermittelt.

4.1. Deskriptive Statistik

Die Verteilung der Anzahl der Krämpfe im Basiszeitraum wird grafisch mit Hilfe eines Histogramms und eines Q-Q-Plots untersucht. Das Histogramm in Abbildung 1 zeigt eine deutlich rechtsschiefe Verteilung. Die meisten Beobachtungen liegen im Bereich niedriger Anfallszahlen, während einige sehr hohe Werte auftreten.

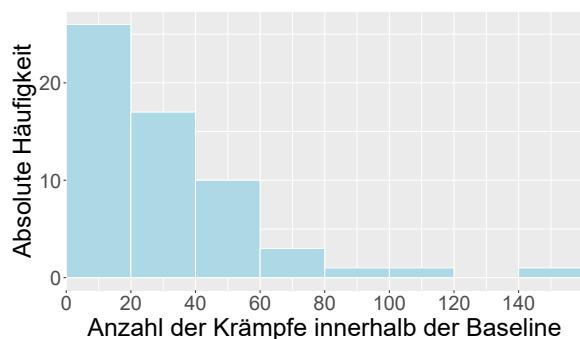


Abbildung 1: Histogramm der Baseline-Krampfanzahl

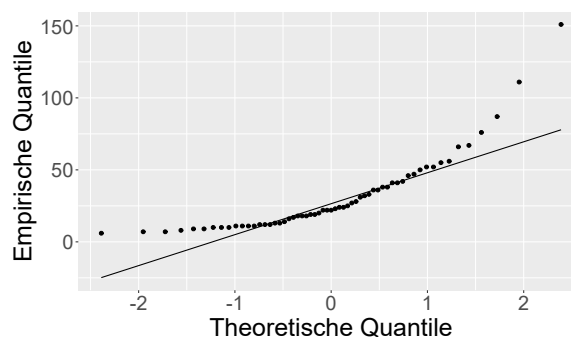


Abbildung 2: Q-Q-Plot der Baseline-Krampfanzahl

Im Q-Q-Plot (siehe Abbildung 2) ist leicht zu erkennen, dass die Datenpunkte einen aufsteigenden Bogen bilden. Die Datenpunkte, insbesondere in den unteren und oberen Quantilen, weichen somit stark von der Diagonalen ab. Sowohl die Asymmetrie im Histogramm als auch die systematische Abweichung im Q-Q-Plot widersprechen der Annahme einer Normalverteilung.

Der Vergleich von Alter und Anzahl der Anfälle vor der Behandlung erfolgt auf Basis von Boxplots (siehe Abbildung 3 für die Baseline-Krampfanzahl bzw. Abbildung 4 für das Alter) und statistischen Kennzahlen in Tabelle 5. Hinsichtlich der Baseline-Krampfanzahl weist die Placebo-Gruppe einen niedrigeren Median von 19 Krämpfen auf, verglichen mit 24 Krämpfen in der Progabid-Gruppe. Die Mittelwerte liegen mit 30.79 (Placebo) und 31.61 (Progabid) sehr nahe beieinander, was auf eine ähnliche zentrale Tendenz hindeutet. Bemerkenswert ist der größere Interquartilsabstand in der Placebo-Gruppe (36.75 vs. 24.50), was eine höhere Variabilität im mittleren Datenbereich anzeigt.

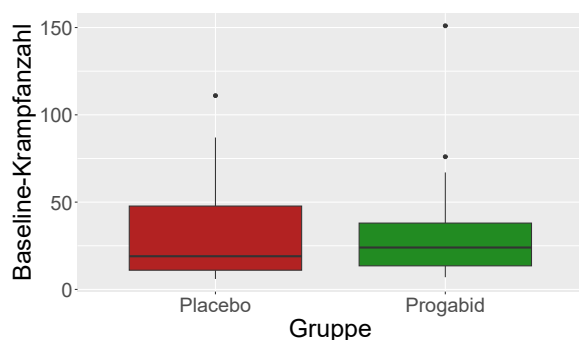


Abbildung 3: Boxplots der Baseline-Krampfanzahl nach Behandlungsgruppe

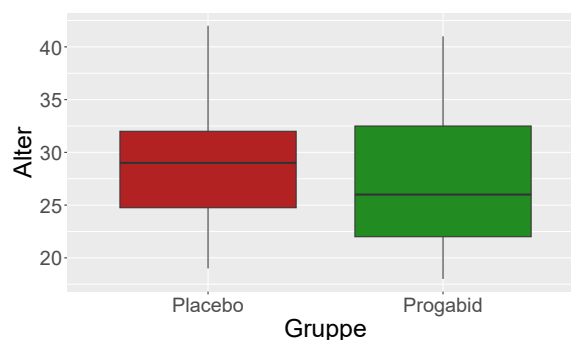


Abbildung 4: Boxplots des Patientenalters nach Behandlungsgruppe

Bezüglich des Alters zeigt die Placebo-Gruppe ein höheres Medianalter (29 vs. 26 Jahre) sowie einen höheren Mittelwert (29.00 vs. 27.74 Jahre). Die Altersverteilung ist in der Progabid-Gruppe etwas heterogener, was durch einen größeren Interquartilsabstand (10.50 vs. 7.25) belegt wird, obwohl die Standardabweichungen ähnlich sind (6.60 vs. 6.00). Im Allgemeinen weisen das Alter und der Epilepsiestatus der Patienten in den beiden Gruppen einige auffällige Unterschiede in der Lage und Streuung der Verteilung auf. Bei der Interpretation der Behandlungseffekte sollten diese Unterschiede berücksichtigt werden.

4.2. Hypothesentests

Die Krampfanzahl zu den vier einzelnen Besuchszeitpunkten ermöglicht eine zeitliche Verlaufsbeobachtung und kann unterschiedlich schnell eintretende oder temporäre Therapieeffekte identifizieren. Der erste Besuch kann frühe Effekte aufzeigen, während spätere Besuche potenzielle Langzeitwirkungen oder Gewöhnungseffekte erfassen.

Der Mittelwert über alle Besuchszeitpunkte bietet hingegen eine zusammenfassende Beurteilung der Gesamtwirksamkeit über den gesamten Studienzeitraum und reduziert die Variabilität einzelner Variablen. Für einen umfangreichen Bericht werden alle fünf Variablen untersucht. Anhand der in Tabelle 6 zusammengefassten Kennzahlen ist zu erkennen, dass es während des Behandlungszeitraums zu keinem großen Unterschied zwischen den beiden Gruppen kam. Darüber hinaus lässt sich in beiden Gruppen ebenfalls kein eindeutiger Auf- oder Abwärtstrend feststellen. Zur Visualisierung werden Boxplots in Abbildung 5 erstellt.

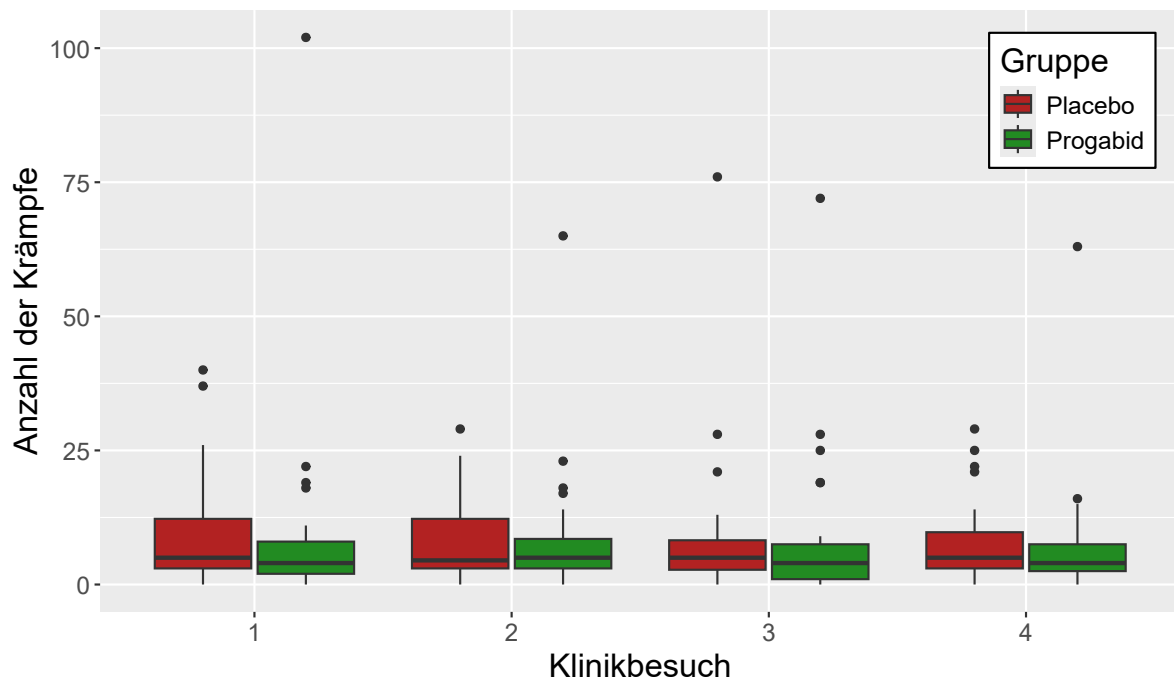


Abbildung 5: Verteilung von Krampfanzahl über die vier Zeitpunkte in den beiden Gruppen

Für die meisten der verglichenen Variablen weichen die Datenpunkte in den Q-Q-Plots der Diagonale stark ab (siehe Abbildung 6 bis Abbildung 13). Die Annahme einer Normalverteilung ist somit nicht erfüllt. Da die Annahme einer Normalverteilung nicht für beide Verteilungen gemacht werden kann, bietet sich hier der Einsatz nichtparametrischer Verfahren wie U-Test oder Cliffs Delta an. Es ist jedoch zu beachten, dass diese Methoden immer noch unabhängige Zufallsvariablen erfordern, was in Studien wie dieser in der Regel erfüllt wird.

Die Ergebnisse der U-Tests zeigen durchgehend keine statistisch signifikanten Unterschiede zwischen der Progabid- und der Placebo-Gruppe auf dem konventionellen 5%-Signifikanzniveau (siehe Tabelle 3). Die p-Werte für alle fünf untersuchten Variablen liegen über 0.05, wobei die Nullhypothese, dass die Verteilung der Krampfanzahl in der Placebo-Gruppe größer oder gleich der in der Progabid-Gruppe ist, in allen Fällen beibehalten wird. Bemerkenswert ist jedoch, dass der p-Wert beim ersten Besuch mit 0.053 nur knapp über der Signifikanzgrenze liegt. Ähnlich verhält es sich mit dem Mittelwert (0.098), der ebenfalls einen tendenziellen Unterschied zwischen den Behandlungsgruppen andeutet. Die p-Werte für den zweiten (0.451), dritten (0.246) und vierten Besuch (0.103) fallen hingegen deutlich höher aus, wobei zum vierten Besuch wieder eine tendenzielle Verbesserung der p-Werte zu beobachten ist. Insgesamt deuten die Ergebnisse darauf hin, dass Progabid keinen signifikanten Effekt hat.

Tabelle 3: Ergebnisse der Überprüfung der Wirkung des Antiepileptikums Progabid im Vergleich zu Placebo mittels U-Test

Krampfanzahl	Nullhypothese	Teststatistik	p-Wert	H_0
1. Besuch	$H_0^1 : F_{pla}(x) \geq F_{pro}(x)$	540.5	0.053	beibehalten
2. Besuch	$H_0^2 : F_{pla}(x) \geq F_{pro}(x)$	442.5	0.451	beibehalten
3. Besuch	$H_0^3 : F_{pla}(x) \geq F_{pro}(x)$	479.5	0.246	beibehalten
4. Besuch	$H_0^4 : F_{pla}(x) \geq F_{pro}(x)$	517.5	0.103	beibehalten
Mittelwert	$H_0^m : F_{pla}(x) \geq F_{pro}(x)$	519.5	0.098	beibehalten

Anmerkung: Wenn die Nullhypothese nicht abgelehnt wird, sagt man auch, dass die Krampfanzahl bei der Progabid-Gruppe stochastik nicht kleiner als bei der Placebo-Gruppe. Die Auswirkung von Progabid wird somit nicht bestätigt.

Die Ergebnisse der Effektstärkenanalyse in Tabelle 4 liefern ergänzende Informationen zur Beurteilung der Wirksamkeit von Progabid im Vergleich zum Placebo.

Tabelle 4: Effektstärken der Wirkung des Antiepileptikums Progabid im Vergleich zu Placebo mittels Cliffs Delta

Krampfanzahl	Effektstärke	Interpretation
1. Besuch	0.245	gering
2. Besuch	0.020	unerheblich
3. Besuch	0.105	unerheblich
4. Besuch	0.192	gering
Mittelwert	0.197	gering

Die Analyse zeigt, dass zum ersten Besuchszeitpunkt mit einem Cliffs Delta von 0.245 ein geringer Effekt vorliegt, was mit dem beinahe signifikanten p-Wert von 0.053 korrespondiert. Für den zweiten und dritten Besuch werden mit Werten von 0,020 bzw. 0,105 lediglich unerhebliche Effekte festgestellt, was die deutlich höheren p-Werte (0.451 und 0.246) widerspiegelt. Interessanterweise zeigt sich zum vierten Besuch wieder ein geringer Effekt (0.192). Der Mittelwert über alle Besuche weist mit einem Cliffs Delta von 0.197 ebenfalls einen geringen Effekt auf, was die tendenziellen Ergebnisse des statistischen Tests (0.098) untermauert. Die Effektstärken der fünf Variablen werden allen nur als gering oder nicht signifikant betrachtet. Es ist jedoch zu beachten, dass diese Beurteilung lediglich auf allgemeinen Konventionen beruht.

5. Zusammenfassung

Die vorliegende Studie untersuchte die Wirksamkeit des Antiepileptikums Progabid im Vergleich zu Placebo bei der Reduktion epileptischer Anfälle. Basierend auf den von Thall und Vail berichteten Daten wurden Epilepsie-Patientinnen nach einer achtwöchigen Beobachtungsphase ohne Behandlung (Baseline-Zeitraum) randomisiert zwei Gruppen zugewiesen. Die Progabid-Gruppe erhielt das Antiepileptikum, während die Kontrollgruppe ein Placebo erhielt. Bei vier aufeinanderfolgenden Klinikbesuchen wurde jeweils die Anzahl der Krämpfe in den vorangegangenen zwei Wochen erfasst. Das Datenmaterial umfasste 59 Patientinnen mit Informationen zu Alter, Baseline-Krampfanzahl sowie der Krampfanzahl zu den vier Besuchszeitpunkten.

Die Analyse der Baseline-Charakteristika zeigte weitgehend vergleichbare Gruppen hinsichtlich des Alters, wobei die Placebo-Gruppe ein etwas höheres Medianalter aufwies (29 vs. 26 Jahre). Bei der Baseline-Krampfanzahl wurden trotz ähnlicher Mittelwerte (30.79 vs. 31.61) unterschiedliche Verteilungscharakteristika festgestellt, mit einem niedrigeren Median in der Placebo-Gruppe (19 vs. 24) und einem größeren Interquartilsabstand (36.75 vs. 24.50).

Die Wirksamkeitsanalyse mittels U-Tests ergab für keinen der untersuchten Zeitpunkte oder den Mittelwert statistisch signifikante Unterschiede auf dem 5%-Niveau. Dennoch zeigten der erste Besuch ($p=0.053$) und der Mittelwert über alle Besuche ($p=0.098$) tendenzielle Unterschiede zwischen den Behandlungsgruppen. Die ergänzende Effektstärkenanalyse mittels Cliff's Delta bestätigte geringe Effekte für den ersten Besuch (0.245), den vierten Besuch (0.192) und den Mittelwert (0.197), während für den zweiten und dritten Besuch lediglich unerhebliche Effekte festgestellt wurden.

Insgesamt deuten die Ergebnisse darauf hin, dass Progabid möglicherweise einen schwachen positiven Effekt auf die Reduktion epileptischer Anfälle hat, der jedoch in dieser Studie nicht mit ausreichender statistischer Sicherheit bestätigt werden konnte. Die Kombination aus beinahe signifikanten p-Werten und geringen Effektstärken, insbesondere zu Beginn der Behandlung und über den gesamten Studienzeitraum betrachtet, legt nahe, dass eine klinisch relevante Wirkung existieren könnte, die in größeren Studienpopulationen deutlicher nachweisbar wäre. Für eine abschließende Beurteilung der Wirksamkeit von Progabid wären weitere Studien mit größeren Stichproben erforderlich.

Literatur

- N. Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 1993. doi: 10.1037/0033-2909.114.3.494.
- U. Genschel und C. Becker. *Schließende Statistik: Grundlegende Methoden*. Springer, Berlin, Heidelberg, 2005.
- J. Hartung, B. Elpelt, und K. H. Klösener. *Statistik — Lehr- und Handbuch der angewandten Statistik*. Oldenbourg, München, 15. auflage edition, 2009.
- J. Hedderich und L. Sachs. *Angewandte Statistik: Methodensammlung mit R*. Springer Spektrum, Heidelberg, 2020. (17. Auflage),.
- I.E. Leppik et al. A double-blind crossover evaluation of progabide in partial seizures. *Neurology*, 35:285, 1985.
- K. Meissel und E. S. Yao. Using cliff’s delta as a non-parametric effect size measure: An accessible web app and r tutorial. *Practical Assessment, Research, and Evaluation*, 29(2), 2024. doi: 10.7275/pare.1977.
- R Core Team. R: A language and environment for statistical computing, 2024. URL <https://www.R-project.org/>.
- P. F. Thall und S. C. Vail. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46(3):657–671, 1990.
- Hadley Wickham, Danielle Navarro, und Thomas Lin Pedersen. *ggplot2: Elegant Graphics for Data Analysis (3e)*. Springer-Verlag, work-in-progress 3rd edition, 2016. URL <https://ggplot2-book.org/>.

A. Zusätzliche Tabellen

Tabelle 5: Vergleich der Baseline-Krampfanzahl und des Alters zwischen Placebo- und Progabid-Gruppen.

Gruppe	Variable	Min	Median	Mittelwert	Max	SD	IQR
Placebo	Baseline-Krämpfe	6	19	30.79	111	26.10	36.75
	Alter	19	29	29.00	42	6.00	7.25
Progabid	Baseline-Krämpfe	7	24	31.61	151	27.98	24.50
	Alter	18	26	27.74	41	6.60	10.50

Anmerkung: SD = Standardabweichung, IQR = Interquartilsabstand.

Tabelle 6: Statistische Zusammenfassung nach Behandlungsgruppe für die Anzahl der Krämpfe bei jedem Besuch und den Mittelwert

Gruppe	Anzahl der Krämpfe	Min	Median	Mittelwert	Max	SD	IQR
Placebo	beim 1. Besuch	0.00	5.00	9.36	40.00	10.14	9.25
	beim 2. Besuch	0.00	4.50	8.29	29.00	8.16	9.25
	beim 3. Besuch	0.00	5.00	8.71	76.00	14.61	5.50
	beim 4. Besuch	0.00	5.00	7.96	29.00	7.63	6.75
	Mittelwert pro Besuch	1.50	4.00	8.58	35.75	8.75	7.75
Progabid	beim 1. Besuch	0.00	4.00	8.58	102.00	18.24	6.00
	beim 2. Besuch	0.00	5.00	8.42	65.00	11.86	5.50
	beim 3. Besuch	0.00	4.00	8.13	72.00	13.89	6.50
	beim 4. Besuch	0.00	4.00	6.71	63.00	11.26	5.00
	Mittelwert pro Besuch	0.00	3.75	7.96	75.50	13.47	5.12

Anmerkung: SD = Standardabweichung, IQR = Interquartilsabstand.

B. Zusätzliche Grafiken

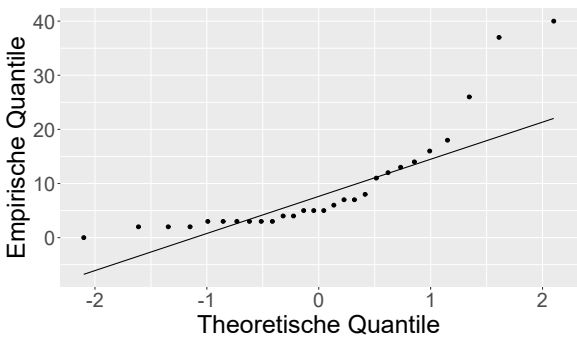


Abbildung 6: Q-Q-Plot Anzahl der Krämpfe der Placebo-Gruppe beim 1. Besuch

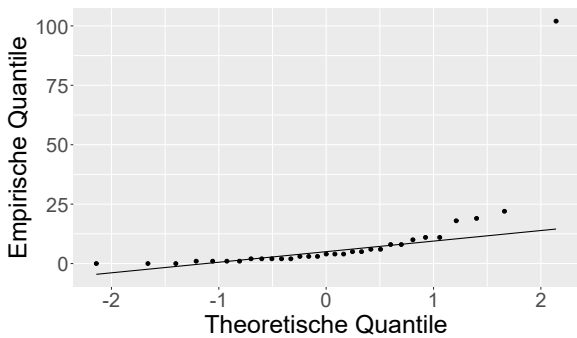


Abbildung 7: Q-Q-Plot Anzahl der Krämpfe der Progabid-Gruppe beim 1. Besuch

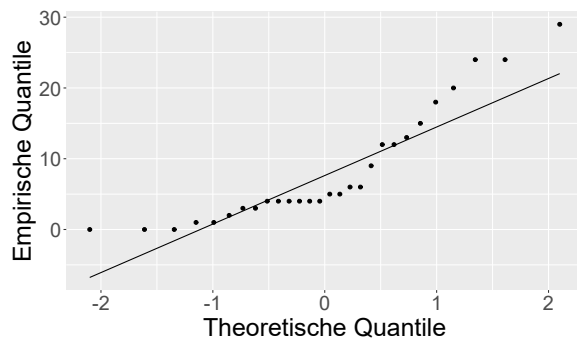


Abbildung 8: Q-Q-Plot Anzahl der Krämpfe der Placebo-Gruppe beim 2. Besuch

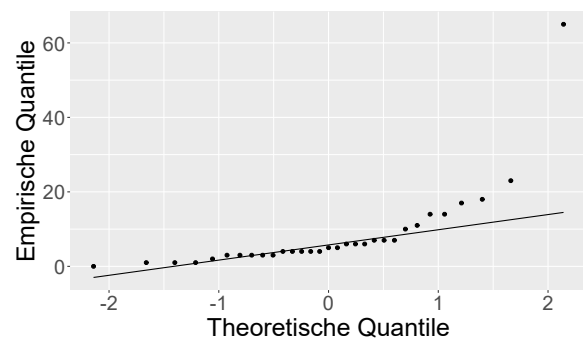


Abbildung 9: Q-Q-Plot Anzahl der Krämpfe der Progabid-Gruppe beim 2. Besuch

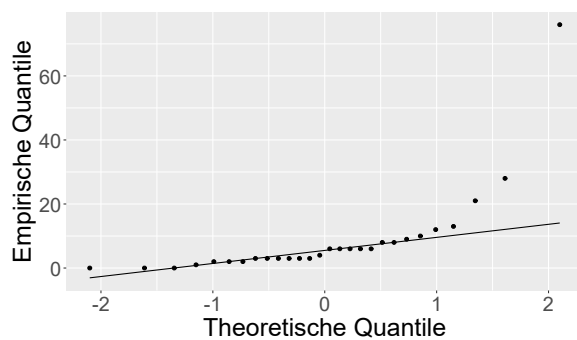


Abbildung 10: Q-Q-Plot Anzahl der Krämpfe der Placebo-Gruppe beim 3. Besuch

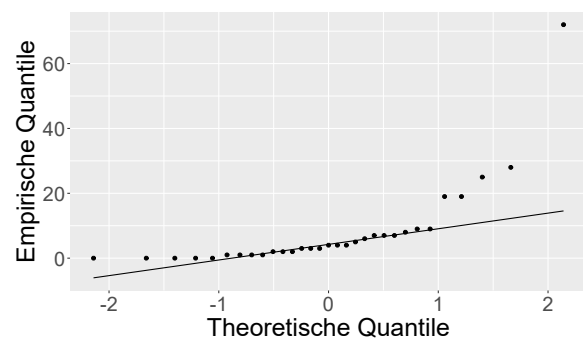


Abbildung 11: Q-Q-Plot Anzahl der Krämpfe der Progabid-Gruppe beim 3. Besuch

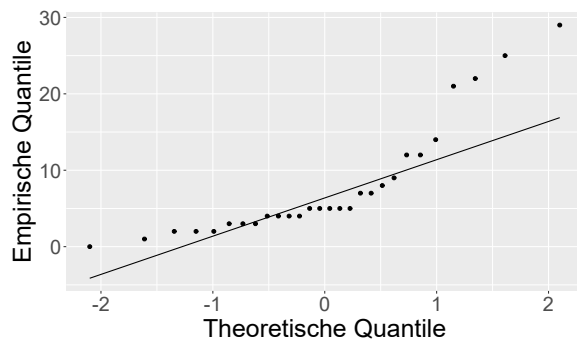


Abbildung 12: Q-Q-Plot Anzahl der Krämpfe der Placebo-Gruppe beim 4. Besuch

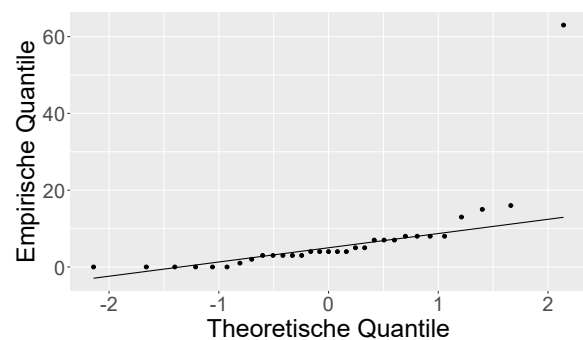


Abbildung 13: Q-Q-Plot Anzahl der Krämpfe der Progabid-Gruppe beim 4. Besuch