

Wie sinnvoll sind Responderanalysen in klinischen Studien?

Peter Kleist

GlaxoSmithKline AG, Münchenbuchsee

Quintessenz

- Responderanalysen sind mit diversen Schwächen verbunden, die die statistische Effizienz und Aussagekraft einer klinischen Studie einschränken.
- Stetige Variablen sollten grundsätzlich anhand der erhobenen Originaldaten ausgewertet werden.
- Unsere heute üblichen Studiendesigns erlauben keine Aussage zur individuellen Therapieresponse.

In der medizinischen Praxis werden aus stetigen (bzw. kontinuierlichen) Messwerten oftmals Kategorien gebildet, um die Diagnosestellung oder Therapieentscheidungen zu erleichtern. Beispiele für stetige Variablen sind das Körpergewicht, der Blutdruck und die Plasmaglukose. Definierte Schwellenwerte grenzen gesunde Individuen von Übergewichtigen, (Prä-)Hypertonikern oder (Prä-)Diabetikern ab.

Auch bei der Auswertung klinischer Studien begegnen wir häufig einer Aufteilung von stetigen Daten in Kategorien. Im Extremfall ergeben sich nur zwei Gruppen von Daten bzw. von Patienten (Dichotomisierung). Ein klassisches Beispiel für eine solche Dichotomisierung sind sogenannte Responderanalysen. Der Anteil der Patienten, der einen als klinisch relevant angesehenen Effekt aufzeigt (Responderrate), soll die vereinfachte Beurteilung der Wirksamkeit einer untersuchten Therapie ermöglichen.

Im Rahmen von Studien zu Antiepileptika ist es beispielsweise üblich, Responder von Nonrespondern auf Basis einer mindestens 50%igen Reduktion der Anfallshäufigkeit in einem definierten Zeitraum abzugrenzen. Und bei der Bewertung der antihypertensiven Wirksamkeit stuft man Patienten als Responder ein, wenn am Ende der Studie zum Beispiel ein diastolischer Blutdruck von ≤ 90 mm Hg oder eine Senkung des Ausgangswerts um mindestens 15 mm Hg erreicht werden konnte.

Die Frage ist, ob diese Vereinfachung – schliesslich wird eine Vielzahl von erhobenen Daten zu einer einfachen Ja/Nein-Information reduziert – im Rahmen der Auswertung einer klinischen Studie wirklich sinnvoll ist. Nachfolgend werden einige kritische Randbemerkungen zu Responderanalysen sowohl aus statistischer als auch aus klinischer Sicht gemacht.

1. Der bereits angesprochene Informationsverlust durch eine Responderanalyse vermindert die Teststärke im Vergleich zu einer Auswertung auf Basis der Originaldaten ganz erheblich – etwa so, als ob die Daten von mindestens einem Drittel der Patienten einfach igno-

riert würden. Anders ausgedrückt: Der erforderliche Stichprobenumfang wächst um mindestens 40% an. Als Folge davon nimmt die Aussagekraft kleiner Studien noch mehr ab, dagegen steigt das Risiko für ein falsch positives Ergebnis [1, 2].

2. Der Wert, der einen Responder von einem Nonresponder abgrenzt, ist häufig arbiträr: Warum qualifiziert eine 51%ige Abnahme von Krampfanfällen einen Patienten als Responder, eine 48%ige Reduktion dagegen nicht? Oder wenn sich die auf Basis der *Hamilton Depression Rating Scale* charakterisierte Depressionssymptomatik bei einem Patienten von 30 auf 15 Punkte und bei einem anderen von 30 auf 16 Punkte verbessert: Warum ist der eine ein Responder und der andere nicht? Hinzu kommt, dass eine vorhandene intraindividuelle Variabilität der Messwerte (Messfehler, zeitliche Fluktuationen) unzureichend berücksichtigt wird – insbesondere dann, wenn der durchschnittliche Effekt aller Patienten in einer Gruppe grösser ist als die als klinisch relevant betrachtete Veränderung. Für einen nicht unerheblichen Teil der Patienten ist die Studie dann nämlich nicht mehr in der Lage, eine klinisch relevante Veränderung zu zeigen, obwohl diese in Wirklichkeit vorhanden ist. Denn Messfehler bzw. intraindividuelle Schwankungen sorgen mit dafür, dass bei manchen Patienten der Grenzwert der klinischen Relevanz überschritten wird – bei anderen hingegen nicht [2, 3].
3. In einer Vielzahl klinischer Studien werden keine klinischen Endpunkte, sondern Surrogatendpunkte untersucht. Ein Beispiel hierfür ist wiederum der Blutdruck als Surrogat für Herzinfarkte und Schlaganfälle. Der Vorhersagewert für das Auftreten klinischer Ereignisse beruht jedoch auf dem Blutdruck als stetige Variable, d.h. auf einer linearen Beziehung zwischen der Ausprägung des Surrogats und den erwarteten klinischen Ereignissen. Eine Dichotomisierung, wie sie mit einer Responderanalyse vorgenommen wird, drückt diese Beziehung zwischen Blutdruckeinstellung und dem Krankheitsverlauf dagegen nur noch unzureichend aus. Darüber hinaus gibt es Beispiele dafür, dass selbst bei Nonrespondern in Bezug auf einen Biomarker oder Intermediärendpunkt die Therapie mit einem klinischen Nutzen verbunden sein kann (Beispiel in [4]).
4. Responderanalysen lassen zudem die hohe Variabilität *innerhalb* der beiden Gruppen, d.h. der Responder und der Nonresponder, ausser Acht. Diese Variabilität ist in der Regel deutlich höher als bei den Patienten, die sich nahe am Grenzwert bewegen. Ein Beispiel: Liegt der Trennpunkt zur Beurteilung einer gewichtsreduzierenden Massnahme bei 5% Gewichtsverlust, dann unterscheidet sich ein Patient mit einer 5,1%igen Gewichtsreduktion wenig von dem mit einer

4,9%igen Abnahme – obwohl der erste als Responder eingestuft wird und der zweite nicht. Dagegen ist ein Patient, dessen Gewicht um 10% ansteigt, wohl kaum mit dem Patient mit der 4,9%igen Abnahme zu vergleichen – obwohl beide Patienten der Nonresponder-Gruppe zugeordnet sind [5].

5. Eine beobachtete Response lässt sich nicht kausal auf die zugrunde liegende Intervention zurückführen. Studien sind oftmals so angelegt, dass bei vielen Patienten – unabhängig von der Therapie – eine Verbesserung gegenüber dem Ausgangsstatus auftritt. Beispiele hierfür sind Regression-to-the-mean- und Trainingseffekte, die spontane Variabilität der Erkrankung sowie Ko-Interventionen. Massgeblich ist daher nicht die intraindividuelle Veränderung im Studienverlauf, sondern der Vergleich zwischen einer Interventions- und einer Kontrollgruppe [5].

In Bereichen, in denen die sogenannte «Plazebo-Response» hoch ist (z.B. bei bipolaren Störungen oder bei der Schmerzbehandlung), ist es daher problematisch, überhaupt von Therapie-Respondern zu sprechen. Denn unterscheidet sich die Responserate unter der untersuchten Therapie nur unwesentlich von der unter Plazebo, lässt sich eben nur ein geringer Anteil der Wirkung auf die Intervention selbst zurückführen. Hinter Run-in-Phasen zur Identifikation von Therapie-Respondern bzw. zur Elimination von Plazebo-Respondern vor der Randomisierung, insbesondere bei Studien im Bereich der Psychiatrie, muss daher ebenfalls ein grundsätzliches Fragezeichen gesetzt werden [6, 7].

6. Weder unsere gängigen Parallelgruppen- noch Cross-Over-Studien erlauben uns, überhaupt Aussagen zur individuellen Response zu machen. Dazu wäre es notwendig, bei einem Patienten Interventions- und Kontrollbehandlung – in unregelmässiger Reihenfolge – immer wieder zu verabreichen und nach einem konsistenten Reaktionsmuster zu schauen. In diesem Sinn ist auch die Frage, ob es überhaupt «Plazebo-Responder» gibt, nicht eindeutig zu beantworten [8]. Selbst eine Verschlechterung unter Therapie könnte eine Response darstellen, da das Ausmass der Verschlechterung unter Kontrollbedingungen noch grösser sein könnte. Dazu wäre jedoch ein Vergleich zu «Was wäre, wenn?» notwendig.

Eine einzelne Studie reflektiert lediglich einen zeitlichen Ausschnitt und bietet Informationen zu einem durchschnittlichen Verhalten einer Patientenpopulation. Würde man eine bestimmte Studie mit denselben Patienten nochmals in exakt gleicher Weise durchführen, kann es zwar sein, dass in beiden Studien jeweils 80% der Patienten Responder sind; allerdings müssen das nicht die gleichen Patienten sein. In der Wiederholungsstudie zeigen vielleicht nur 80% der ursprünglichen Responder erneut eine Response, während jetzt aber 80% der ehemaligen Nonresponder zu Respondern werden. Das Einzige, was sich dann sicher sagen lässt, ist: zu jedem beliebigen Zeitpunkt hat jeder beliebige Patient eine 80%ige Chance, auf die Therapie anzusprechen [3, 9].

Noch eine Randbemerkung: Responderanalysen können missbräuchlich durchgeführt werden, um ungerechtfertigte Aussagen zur Wirksamkeit einer Therapie zu

machen. Zum einen verbietet es sich, Auswertungen auf Basis von Responder- und Nonresponder-Subgruppen vorzunehmen, da Response keine unabhängige, vor der Randomisierung erhobene Untersuchungsvariable darstellt. Zum anderen sollte man offenen Verlängerungsstudien keinen Glauben schenken, die über Responderraten berichten, die deutlich höher sind als die in randomisierten Studien; solche Studien zeichnen sich nämlich in der Regel durch eine starke Positivauswahl von Patienten aus (Selektionsbias).

Schlussfolgerungen

Im gegenwärtigen Zeitalter der «stratifizierten Medizin» stehen wir entwicklungsgeschichtlich zwischen dem Ansatz «one size fits all» und einer zunehmend personalisierten Medizin. Ob eine verfügbare Therapie für einen Patienten geeignet ist oder nicht, ist eine absolut notwendige Frage, die uns auch noch Jahre beschäftigen wird. Doch Vorgehensweisen, die im klinischen Alltag Bedeutung haben, sind für die Auswertung klinischer Studien oftmals nicht sinnvoll. Responderanalysen in klinischen Studien, obwohl gängige Praxis, mit eigenen Befürwortern [10] und teilweise sogar von Arzneimittelbehörden gefordert [11, 12], sollten daher kritisch hinterfragt werden. Bei der Beurteilung eines Antibiotikums (geheilt / nicht geheilt) oder einer onkologischen Therapie (am Leben / gestorben) sind dichotome Auswertungen der logische Ansatz. Werden jedoch stetige Variablen untersucht (Blutdruck, Laborvariablen, *Rating-Scales*, Anzahl Ereignisse), sind Responderanalysen mit diversen Schwächen verbunden, welche die statistische Effizienz und die Aussagekraft einer klinischen Studie limitieren.

Auf keinen Fall versetzen uns Responderanalysen in die Lage, Aussagen zur individuellen Response im Sinne einer Kausalitätsbeziehung zur Intervention zu machen. Der mehrfach zitierte Statistiker Stephen Senn soll das Schlusswort behalten: «Für die Entwicklung eines Arzneimittels braucht man Hunderte, mitunter Tausende von Patienten, um überhaupt eine Wirksamkeit aufzuzeigen. Im Individualfall ist es jedoch unmöglich zu sagen, ob ein Patient einen Nutzen davongetragen hat oder nicht» (frei nach [9]).

Korrespondenz:

Dr. med. Peter Kleist
GlaxoSmithKline AG
Talstrasse 3–5
CH-3053 Münchenbuchsee
peter.m.kleist@gsk.com

Empfohlene Literatur

- Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
- Senn S. Disappointing dichotomies. *Pharmaceut Statist*. 2003;2: 239–40.
- Senn S. Individual response to treatment: is it a valid assumption? *BMJ*. 2004;329:966–8.

Die vollständige nummerierte Literaturliste finden Sie unter www.medicalforum.ch.

Wie sinnvoll sind Responderanalysen in klinischen Studien?

Literatur (Online-Version)

- 1 Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
- 2 Senn S. Disappointing dichotomies. *Pharmaceut Statist*. 2003;2:239–40.
- 3 Senn S. Individual response to treatment: is it a valid assumption? *BMJ*. 2004;329:966–8.
- 4 Llovet JM, Ricci S, Mazzaferro V, et al. for the SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med*. 2008;359:378–90.
- 5 Snapinn AM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials*. 2007;8:31. doi:10.1186/1745-6215-8-31.
- 6 Lee S, Walker JR, Jakul L, et al. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress Anxiety*. 2004;19:10–9.
- 7 Trivedi MH, Rush J. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology*. 1994;11:33–4.
- 8 Kaptchuk TJ, Kelley JM, Deykin A, et al. Do «placebo responders» exist? *Contemp Clin Trials*. 2008;29:587–95.
- 9 Senn S. Three things that every medical writer should know about statistics. *J Eur Med Writer Ass*. 2009;3:159–62.
- 10 Lewis JA. In defence of the dichotomy. *Pharmaceut Statist*. 2004;3:77–9.
- 11 European Agency for the Evaluation of Medicinal Products. Note for guidance on clinical investigation of medicinal products in the treatment of epileptic disorders. CPMP/EWP/566/98 rev 1. London, 16 November 2000. www.emea.europa.eu/pdfs/human/ewp/056698en.pdf.
- 12 US Food and Drug Administration. Draft guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071975.pdf.