

„N-of-1“-Studie – Ein mögliches Studienkonzept für die Allgemeinmedizin?

T. Fischer
E. Hummers-Pradier

“N-of-1”-Study: A Concept of Primary Care Research?

Literaturbeispiel

Im Februar 2004 erschienen 2 klinische Berichte im British Medical Journal bezüglich einer 25-jährigen schwangeren Patientin, die sich mit ausgeprägter Übelkeit und Erbrechen in einer hausärztlichen Praxis vorstellte [1, 2]. Nachdem ein Behandlungsversuch mit Prochlorperazine erfolglos geblieben war, fragte die Patientin nach Alternativen zur konventionellen Therapie. Aufgrund einer komplementärmedizinischen Ausrichtung der behandelnden Ärztin, schlug diese einen Therapieversuch mit Vitamin-B6 (Pyridoxin) vor. Eine Literaturrecherche in Medline, Cochrane Library etc. verwies auf 3 plazebokontrollierte Kontrollstudien, die zeigten, dass Vitamin-B6 keinen Einfluss auf das Erbrechen, wohl aber auf die Übelkeit habe. Um eine mögliche Wirkung des Präparates auf die schwangere Patientin zu überprüfen, entschieden sich die britischen Kollegen zu einem ungewöhnlichen Vorgehen: Einem doppelblinden, plazebokontrollierten Behandlungsversuch bei dieser Patientin, einer so genannten „N-of-1“-Studie. Nach Zustimmung der Patientin (und einer Ethikkommission!) erhielt die Patientin 10 mg Vitamin-B6 3× täglich, bzw. Plazebo in jeweils 5 Behandlungsphasen (2 Tage aktive Behandlung, 1 Tag „Auswaschphase“, dann 2 Tage Plazebo) über insgesamt 25 Behandlungstage. Die Erfassung der Symptome erfolgte mittels eines täglich auszufüllendem validierten Selbsterfassungsbogens. Die Patientin selbst gab subjektiv an, dass sich die Übelkeit während der Behandlung nicht relevant verändert habe. Die statistische Auswertung (gepaarte t-Tests für jedes Symptom) ergab keinen signifikanten Unterschied der Symptome im Vergleich Verum vs. Plazebo. Die Patientin selbst entschied sich die Einnahme abzubrechen, nachdem sie über die Ergebnisse informiert worden war.

Hintergrund

Die Darstellung der beiden Artikel aus dem BMJ soll hier keineswegs zu einer Diskussion der Therapieoptionen der Schwangerschaftsübelkeit führen, sondern steht vielmehr exemplarisch für den methodisch interessanten Ansatz. Gerade in der hausärztlichen Praxis steht man häufig vor der Problematik, ob die bekannte wissenschaftliche Evidenz auf den konkreten, individuellen Fall übertragbar ist. Allzu oft beruht diese wissenschaftliche Evidenz auf fallzahlstarken randomisierten Kontrollstudien mit Patienten, deren Beschwerdespektrum und biometrische Daten weit von der hausärztlichen Realität entfernt sind (z. B. zu junge, selten multimorbide Patienten). Zudem stellen die „harten“ Endpunkte vieler Studien wie Infarkt oder Tod eine eher abstrakte, nicht selten sogar negierte Bedrohung für viele, gerade ältere Patienten dar. Demgegenüber erwarten die Patienten vom (Haus-)Arzt vielmehr eine „fühlbare“, symptomorientierte Therapie [3]. Für symptomorientierte Therapeutika liegen jedoch deutlich weniger methodisch hochwertige Kontrollstudien vor. Aus hausärztlicher Sicht bleibt so häufig nur der Behandlungsversuch. Die Patienten werden im Therapieverlauf zu ihrer eigenen Einschätzung zur Symptomatik befragt und die Therapie ggf. modifiziert. Dieses Vorgehen entspricht dem einer Beobachtungsstudie. Behandlungsversuche (bzw. Beobachtungsstudien) weisen jedoch eine Reihe von Nachteilen auf [4, 5]:

- Plazebo-Effekt (ein bedeutsamer Confounder!)
- sekundäre Effekte der Medikation (z. B. kann die häufig beobachtete Euphorie unter Kortisonmedikation Studienergebnisse beeinträchtigen)
- der natürliche Krankheitsverlauf kann gerade bei phasenhaft verlaufenden Krankheiten die Ergebnisse beeinflussen

Institutsangaben

Abteilung Allgemeinmedizin, Georg-August-Universität Göttingen

Korrespondenzadresse

Dr. med. Thomas Fischer · Abteilung Allgemeinmedizin · Universität Göttingen · Humboldtallee 38 · 37073 Göttingen · E-mail: tfische@gwdg.de

Bibliografie

Z Allg Med 2005; 81: 180–182 · © Georg Thieme Verlag KG Stuttgart · New York
DOI 10.1055/s-2005-836353
ISSN 0014-336251

- unkontrollierte Interaktionen (z.B. mit nichtverschreibungspflichtigen Medikamenten, über deren Einnahme der behandelnde Arzt möglicherweise keine Informationen verfügt)
- die Erwartungen sowohl von Patienten aber auch Ärzten können dazu führen, dass beide einander nicht enttäuschen wollen und Effekte daher überbewertet werden
- nur selten werden Auslassversuche unternommen, um die Wirkung zu überprüfen.

Das kontrollierte Vorgehen bei „N-of-1“-Studien soll diesen Einflüssen begegnen. Dieses Studienkonzept hat eine lange Tradition in der psychologischen Forschung [6]. In die primärärztliche Forschung haben „N-of-1“-Versuche vor knapp 20 Jahren Einzug gehalten [4], jedoch ohne wesentliche Verbreitung gefunden zu haben [7]. „N-of-1“-Studien unterscheiden sich dabei von der Intention wesentlich von randomisierten Kontrollstudien (RCTs). Während Letztere die Frage beantworten, ob „eine Behandlung x verglichen mit der Behandlung y in einer definierten Population hilfreich ist“, untersuchen „N-of-1“-Studien die Frage, ob „eine Therapie bei einem individuellen Patienten hilfreich ist“ [8].

Methodisches Vorgehen

Wichtige Voraussetzung für „N-of-1“-Studien ist die doppelte Verblindung, was eine große logistische Herausforderung darstellen kann. Die Behandlungssequenzen (Plazebo – Verum) sollten zudem randomisiert werden [5]. Üblicherweise werden dabei nur 2 Therapieoptionen miteinander verglichen. Theoretisch sind mehrere Optionen überprüfbar, was jedoch die Anzahl der für eine valide Aussage notwendigen Zyklen und damit die logistischen Probleme deutlich erhöht. Mindestens 3 Behandlungszyklen werden als notwendig angesehen, um die Wahrscheinlichkeit zu reduzieren, dass nicht-therapiebedingte Effekte zu falschen Schlussfolgerungen verleiten¹. Dabei werden mehr Behandlungszyklen benötigt, wenn der zu erwartende Effekt der Therapie im Verhältnis zur normalen Fluktuation des Krankheitsverlaufs moderat ausfällt (vergleichbar einer Power-Abschätzung bei randomisierten Kontrollstudien) [9].

Die Ergebnis-Kriterien sollten reliabel, valide und vor dem Versuch definiert sein; sie sollten in sinnvollem zeitlichen Zusammenhang zur Intervention gemessen werden. Dabei müssen Auswaschphasen von Medikamenten ebenso berücksichtigt werden wie mögliche Effekte (bzw. Nebenwirkungen) durch den Beginn oder das Absetzen der Therapie. Die Befragung der Patienten, unter welchen Symptomen sie am meisten leiden, kann dabei helfen die für Patienten (!) relevanten Zielkriterien zu bestimmen. Dabei können validierte Instrumente wie der „Patienten-generierte Index“ [10] oder „Schedule for the evaluation of Individual Quality of Life (SEIQoL)“ [11] eingesetzt werden. Bei stark fluktuierenden Symptomen können Symptom-Tagebücher Informationen über symptomfreie Zeiträume als mögliche Zielgröße liefern. Die Schwere von Symptomen wird in „N-of-1“-Studien zumeist über visuelle Skalen erfasst.

¹ Bei 3 Behandlungszyklen und einer qualitativen Messung des Outcomes z.B. im Sinne von „besser“ oder „schlechter“ beträgt die Wahrscheinlichkeit für eine durch den Zufall bedingten, konsistenten Überlegenheit des Verums $1 : 2^3 = 0,125$ [9].

Die Darstellung der Ergebnisse erfolgt bei „N-of-1“-Studien häufig grafisch. Nachteile derartige Darstellungen sind – neben der Subjektivität – die Anfälligkeiten gegenüber Confoundern und die fehlende Quantifizierung möglicher falsch-positiver Ergebnisse (Typ-1-Fehler) [12].

Sinnvolle statistische Analyseverfahren für „N-of-1“-Studien sind seit einiger Zeit Gegenstand intensiver Diskussionen [5, 13]. So lassen sich parametrische Tests nur bedingt auf diese Daten anwenden; nicht-parametrische Tests weisen hier Vorteile auf. Zeitreihen-Analysen sind eine weitere Option [14].

Mögliche „Indikationen“ für „N-of-1“-Studien

- „N-of-1“-Studien sind besonders geeignet (modifiziert nach [8, 9]),
- wenn wesentliche Zweifel am Behandlungseffekt bei individuellen Patienten vorliegen
 - wenn für individuelle Patienten keine Evidenz in Studien vorliegt (z.B. bei alten, multimorbiden Patienten, die häufig in RCTs ausgeschlossen werden, aber auch bei Kindern)
 - wenn eine längerfristige Therapie mit teuren Medikamenten erwogen wird
 - bei chronischen Krankheiten mit stabiler Symptomatik
 - wenn kurzfristig erkennbare, patientenrelevante Zielgrößen gemessen werden können
 - wenn geklärt werden soll, ob individuelle Patienten zu Respondern oder Non-Respondern einer etablierten Therapie gehören
 - wenn Patienten Nebenwirkungen erwarten und daher einer Therapie skeptisch gegenüber stehen (vor allem bei einer geplanten Dauermedikation). Hier kann ein randomisiert-kontrollierter Versuch eine Option darstellen, Patienten von einer Einnahme zu überzeugen.

Ein geeignetes Beispiel ist die Anwendung von Methylphenidat bei Kindern, die am Aufmerksamkeitsmangel-Syndrom leiden. So wurde Eltern in einer Studie angeboten, ihren auffälligen Kindern die Medikation unter randomisiert-kontrollierten Bedingungen zu verabreichen. Dieser Ansatz wurde von der großen Mehrzahl der Eltern begrüßt, da er den Familien die Chance bieten konnte, den individuellen Effekt der Therapie auf die Kinder und die Auswirkungen auf den Alltag der Familien wissenschaftlich fundiert zu beurteilen. Immerhin 37 von initial 43 teilnehmenden Familien beendeten den aufwändigen Crossover-Versuch, der die Wirkung der Substanz deutlich bestätigte [15]. Nichtsdestotrotz entschied sich ein Teil der Familien trotz positiver statistischer Wirkung gegen die Einnahme, da sich die Familien selbst ein Bild von Wirkungen und Nebenwirkung machen konnten.

Ethische Aspekte

Einige Aspekte von „N-of-1“-Studien entsprechen sehr wohl den Charakteristika von „klassischen“ RCTs (Verblindung, Randomisation der Therapie), daher erscheint die Zustimmung durch eine Ethikkommission gerade bei nicht etablierten, experimentellen Therapieformen sinnvoll. Zumeist dienen diese Studien jedoch eher der Verbesserung der Versorgung eines individuellen Patienten, häufig mit bereits zugelassenen Therapieformen, hier

sollte die Einbeziehung von Ethikkommissionen individuell entschieden werden.

Die Anwendung von Placebos oder alternativ einer Standardtherapie als Vergleichsmaßstab sollte mit den teilnehmenden Patienten offen besprochen werden. Die Einholung eines „informed consent“ ist obligat.

Nachteile von „N-of-1“-Studien und Hindernisse bei der Durchführung

„N-of-1“-Studien haben insgesamt wenig Verbreitung gefunden. Als wesentliche Ursache wird allgemein der erhebliche logistische Aufwand angenommen: Geeignete Patienten müssen identifiziert und aufgeklärt werden, man benötigt externe Unterstützung bei der Verblindung und Randomisation sowie durch Pharmazeuten bei der Herstellung optisch möglichst identischer Verum- und Placebomedikamente. Auch kann außerhalb von Universitäten die fehlende Anbindung an Ethikkommissionen ein Hindernis sein.

Eine psychologische Barriere für Ärzte, ihren Patienten diese Studienform anzubieten, kann die Veränderung der vertrauten Arzt-Patienten-Beziehung hin zu einer Arzt/Wissenschaftler-Patient/„Studienobjekt“-Beziehung sein. Ärzte müssen sich dabei ihren Patienten gegenüber rechtfertigen, warum sie die Standard-Behandlung verlassen, bzw. infrage stellen. Dies kann – vor allem wenn die Patienten ihrerseits hohe Erwartungen an die Medikation haben – sehr hinderlich sein.

Ein wichtiger Nachteil ist die fehlende Generalisierbarkeit der Ergebnisse: Die Resultate eines Patienten erlauben aufgrund der Vielzahl möglicher Confounder keine Übertragung der Ergebnisse auf Andere (umgekehrt ist aber gerade die Heterogenität der Patienten häufig Anlass für die Durchführung von „N-of-1“-Studien). Dieser fehlenden Möglichkeit zur Extrapolation der Ergebnisse wird von einigen Autoren das Konzept multipler „N-of-1“-Studien entgegengesetzt. Dabei werden die Ergebnisse vieler Studienteilnehmer zusammengefasst, bei denen jeweils randomisiert und kontrolliert mehrere Behandlungszyklen durchgeführt worden waren. Dieses Vorgehen wird derzeit vor allem im Bereich der Komplementärmedizin als Studienoption diskutiert [16].

Zusammenfassung

„N-of-1“-Studien stellen eine interessante Alternative dar, wenn für Therapieoptionen keine Evidenz vorhanden ist oder unklar ist, ob die vorhandene Evidenz auf individuelle Patienten anwendbar ist. Da im hausärztlichen Bereich viele ältere und/oder multimorbide Patienten behandelt werden, die häufig aus üblichen RCTs ausgeschlossen werden, können „N-of-1“-Studien gerade für die allgemeinärztliche Versorgung aber auch die Forschung eine sinnvolle Studienform sein.

Haben Sie Ideen für „N-of-1“-Studien?

Über konkrete Vorschläge würden wir uns freuen. Gerne sind wir auch bereit, Hilfe bei der logistischen Umsetzung anzubieten, falls wir sie jetzt zu einer konkreten Studie motivieren konnten. Vorschläge und Anfragen bitte an die Autoren.

Interessenkonflikte: keine angegeben

Literatur

- Harker N, Montgomery A, Fahey T. Treating nausea and vomiting during pregnancy: case progression. *BMJ* 2004; 328: 337
- Harker N, Montgomery A, Fahey T. Treating nausea and vomiting during pregnancy: case outcome. *BMJ* 2004; 328: 503–506
- Kane RL, Rockwood T, Philp I, et al. Differences in valuation of functional status components among consumers and professionals in Europe and the United States. *J Clin Epidemiol* 1998; 51: 657–666
- Guyatt G, Sackett D, Taylor DW, et al. Determining optimal therapy – randomized trials in individual patients. *N Engl J Med* 1986; 314: 889–892
- Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. *Clin Med Ass J* 1988; 139: 497–503
- Martin M. Single subject designs for assessment of psychotropic drug effects in children. *Child Psychiat Hum Dev* 1971; 2: 102–105
- Porta M, Bolumar F, Hernandez I, et al. N of 1 trials. Research is needed into why such trials are not more widely used. *BMJ* 1996; 313: 427
- Peloso PM. Are individual patient trials (N-of-1 trials) in rheumatology worth the extra effort? *J Rheumatol* 2004; 31: 8–11
- Price JD, Evans JG. N-of-1 randomized controlled trials: singularly useful in geriatric medicine. *Age and Ageing* 2002; 31: 227–232
- Ruta DA, Garratt AM, Leng M, et al. A new approach to the measurement of quality of life. The patient-generated index. *Med Care* 1994; 32: 1109–1126
- Browne JP, O'Boyle CA, McGee HM, et al. Individual quality of life in the healthy elderly. *Qual Life Res* 1994; 3: 235–244
- Bobrovitz CD, Ottenbacher KJ. Comparison of visual inspection and statistical analysis of single-subject data in rehabilitation research. *Am J Phys Med Rehabil* 1998; 77: 94–102
- Spiegelthaler DJ. Statistical issues in studies of individual response. *Scand J Gastroenterol* 1988; 147 (Suppl): 40–45
- Horne GP, Yang HCK, Ware WB. Time series analysis for single subject designs. *Psychol Bull* 1982; 91: 178–189
- Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials: practical, useful, and highly endorsed by families. *Arch Pediatr Adolesc Med* 1999; 153: 1292–1296
- Kawakita K, Suzuki M, Namura K, et al. A proposal for a simple and useful research design for evaluating the efficacy of acupuncture: multiple, randomized N-of-1 trials. *J Japan Soc Acupunct Moxibus* 2004; (online journal, www.jsam.jp)

Zur Person



Dr. med. Thomas Fischer
Jahrgang 1969, verheiratet, drei Kinder.
Facharzt für Allgemeinmedizin und Phlebologe, wissenschaftlicher Mitarbeiter der Abteilung Allgemeinmedizin der Universität Göttingen seit Juli 2001.
Forschungsschwerpunkte: Phlebologische Krankheitsbilder in der Allgemeinmedizin, Evaluation quantitativer Datenerhebungsverfahren in der hausärztlichen Versorgung und neue Formen der Lehre in der Allgemeinmedizin.