ELSEVIER
URBAN & FISCHER

Available online at www.sciencedirect.com
ScienceDirect

ZEFQ
www.elsevier.de/zefq

Z. Evid. Fortbild. Qual. Gesundh. wesen (ZEFQ) ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Schwerpunkt

Cluster-randomisierte Studien: eine wichtige Methode in der allgemeinmedizinischen Forschung

Jean-François Chenot*

Abteilung Allgemeinmedizin, Universitätsmedizin Göttingen, Humboldtallee 38, 37073 Göttingen

Zusammenfassung

Interventionsstudien in der Allgemeinmedizin untersuchen häufig organisatorische Veränderungen oder edukative Interventionen z.B. im Rahmen von Leitlinienimplementierungen. Die Randomisierung findet dann aus pragmatischen Gründen meist nicht auf Patientenebene, sondern auf Praxisebene statt. Die Studienteilnehmer aus einer Praxis bilden ein sog. Cluster (Gruppe), womit die Grundannahme der unabhängigen Patientenstichprobe nicht mehr gegeben ist. Dies erfordert eine höhere Probandenzahl und eine komplexere Auswertung. Die Vor- und Nachteile der Cluster-Randomisierung, sowie die praktische Bedeutung bei der Planung und Auswertung dieser Studien werden in diesem Artikel am Beispiel von zwei Cluster-randomisierten Studien vermittelt.

Schlüsselwörter: Allgemeinmedizinische Forschung, Cluster-Randomisierung

Cluster randomised trials: an important method in primary care research

Summary

In primary care research interventional studies often address organisational changes or educational interventions, for example, in the context of guideline implementation. For pragmatic reasons randomisation is often conducted at practice level instead of at the individual patient level. Patients from one practice form a cluster, thus violating the basic assumption of independent patient samples. Hence an increased number of participants and a more complex analysis are required. Using the example of two cluster randomised trials the present article provides insights into the advantages and disadvantages of cluster randomisation as well as its practical significance for the planning and analysis of cluster randomised trials.

Key words: primary care research, cluster randomisation

Einleitung

Randomisiert kontrollierte Studien (RCT) werden zunehmend auch in Allgemeinmedizinpraxen in der Primärversorgung durchgeführt [1]. Sie gelten als Goldstandard der evidenzbasierten Medizin, insbesondere in der Pharmakotherapie. Studien in der Primärversorgung untersuchen häufig organisatorische Veränderungen, edukative Interventionen, z.B. im Rahmen von Leitlinienimplementierungen. Da solche Interventionen typischerweise in einer Organisationseinheit stattfinden findet auch die

*Korrespondenzadresse: Priv. Doz. Dr. med. Jean-François Chenot, MPH, Abteilung Allgemeinmedizin, Universitätsmedizin Göttingen, Humboldtallee 38, 37073 Göttingen. Tel.: 0551-396599; fax: 0551-399530.
E-Mail: jchenot@gwdg.de

Randomisierung nicht individuell, sondern auf Praxisebene statt. Die Studienteilnehmer aus einer Praxis bilden ein sog. Cluster (Gruppe) und man spricht dann von einer „Cluster-Randomisierung". Mit der zunehmenden Forschung in der Allgemeinmedizin ist die Anzahl cluster-randomisierter Studien (cRCT) seit dem Beginn der neunziger Jahre stark angestiegen [2]. Andere Settings in dem cRCTs eingesetzt werden sind z.B. edukative Interventionen in Schulen oder im Public Health-Bereich, bei denen z.B. Schulklassen oder Betriebe ein Cluster bilden.

In diesem Artikel werden Vor- und Nachteile der Cluster-Randomisierung im Vergleich zur individuellen Randomisierung sowie die sich daraus ergebenden speziellen Anforderungen bei deren Planung und Auswertung vermittelt. Zur Erläuterung werden hierzu zwei aktuelle deutsche cluster-randomisierte Studien im hausärztlichen Setting in Grundzügen beispielhaft vorgestellt [3,4].

Studienhintergrund

Beispielstudie 1 [3]: In einer dreiarmigen kontrollierten Studie wird in zwei Interventionsarmen die DEGAM-Leitlinie „Kreuzschmerzen" in Allgemeinmedizinpraxen implementiert. Im zweiten Interventionsarm wird zusätzlich eine Beratung (motivational counselling) zur körperlichen Aktivität durch Medizinische Fachangestellte angeboten. Studienendpunkt ist die Funktionskapazität gemessen mit „Funktionsfragebogen Hannover".

Beispielstudie 2 [4]: In einer dreiarmigen kontrollierten Studie wird in zwei Interventionsarmen ein Arthrosemanagement für Hüfte oder Knie in Allgemeinmedizinpraxen implementiert. Im zweiten Interventionsarm wird zusätzlich eine Case Management durch Medizinische Fachangestellte durchgeführt. Studienendpunkt ist die Lebensqualität gemessen mit der Kurzform der „Arthritis Impact Measurement Scale".

Was sind die Vor- und Nachteile der Cluster-Randomisierung?

Das wichtigste Argument für die Cluster-Randomisierung ist, dass es aus pragmatischen Gründen innerhalb einer Praxis organisatorisch oft nicht möglich ist, Patienten unterschiedliche Interventionen anzubieten. Es wird eine sog. **Kontamination** befürchtet, d.h. dass z.B. Patienten der Kontrollgruppe doch einen Teil oder die ganze Intervention des Interventionsarms erhalten oder umgekehrt [5]. Alternativen Studiendesigns bei organisatorischen Interventionen können unter Umständen dass sog. stepped wedge design oder cross over trial sein [6].

Patienten aus einer Praxis teilen mit hoher Wahrscheinlichkeit Eigenschaften wie z.B. Exposition derselben Umweltfaktoren oder einen ähnlichen sozioökonomischen Standard. Es besteht die erhöhte Wahrscheinlichkeit einer Korrelation für den Studienendpunkt relevanter Faktoren innerhalb eines Clusters. Der Statistiker spricht dann von einer reduzierten Varianz (Streuung). Damit ist die Grundannahme der unabhängigen Patientenstichprobe nicht mehr gegeben (Abb. 1). Dies kann durch Adjustierungen im Studiendesign und bei der Analyse zumindest teilweise berücksichtigt werden (siehe unten). Für systematische Verzerrungen (Bias) ist eine Korrektur dagegen nur in Ausnahmefällen möglich.

Durch Verblindung und Randomisierung der Patienten sollen bekannte und noch wichtiger unbekannte Einflussfaktoren (confounder) auf die Studienendpunkte (outcome) gleichmäßig in Kontroll- und Interventionsarm verteilt werden.

Da sich organisatorische Veränderungen oder edukative Interventionen meist weder für teilnehmende Praxen noch für Patienten verblinden lassen, sind cRCTs besonders anfällig für **Selektionsbias** [7]. Der Bias kann auf zwei Ebenen, sowohl bei der Praxenrekrutierung als auch bei der Patientenrekrutierung in den Praxen entstehen.

Üblicherweise werden Praxen erst nach der Rekrutierung randomisiert um zu verhindern, dass z.B. besonderst motivierte Praxen in den Interventionsarm kommen **(Allocation bias)**. Praxen in den oft als weniger attraktiv empfunden Kontrollarmen scheiden mit höherer Wahrscheinlichkeit aus. Wird die Intervention von den Praxen als zu aufwändig empfunden kann es in den Interventionsarmen zu einer erhöhten ausscheiden aus der Studie kommen. Die durch das Ausscheiden von Praxen entstehenden leeren Cluster können bei der Auswertung nicht berücksichtigt werden, was eine Intention to treat–Analyse unmöglich macht [8].

Der Selektionsbias bei der Patientenrekrutierung in den Praxen wird auch als **subsampling-Bias** (auch recruitment bias) bezeichnet (Abb. 2). Optimalerweise sollten alle für die Studie geeigneten Patienten vor der Randomisierung identifiziert werden, z.B. sturzgefährdete Personen in Pflegeheimen [7]. Das ist aber z.B. bei akuten Zielerkrankungen oft nicht möglich. Werden nicht alle für die Studie in Frage kommenden Patienten rekrutiert kann dies

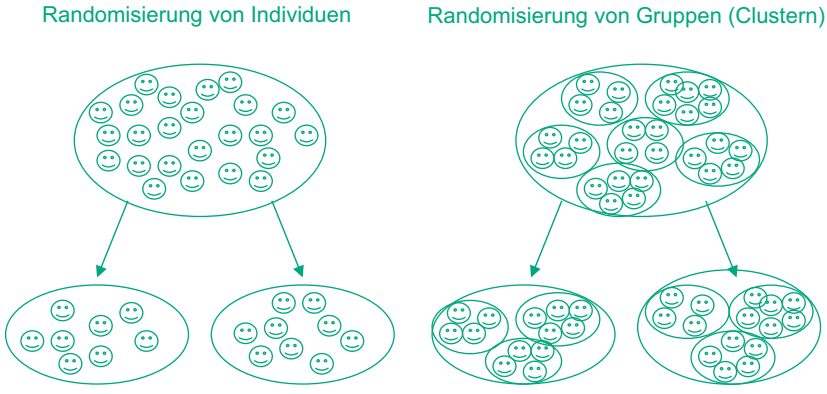
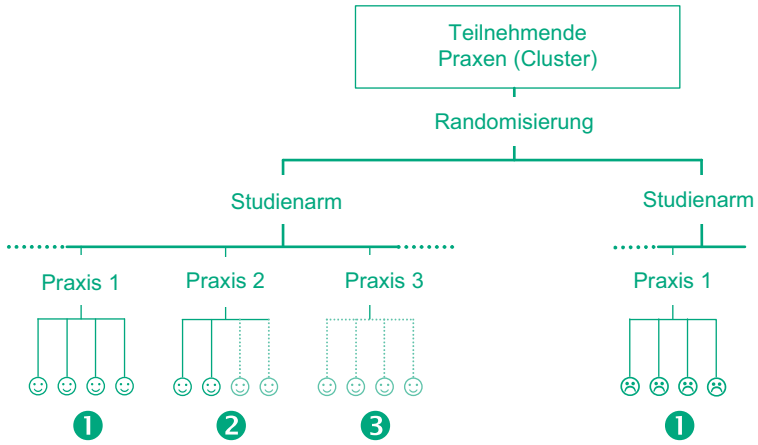


Abb. 1. Unterschied bei der Randomisierung von Individuen oder Clustern.



❶ Praxis die dem Vergleichstudienarm unähnliche Patienten rekrutiert
❷ Praxis die geeignete Patienten nicht rekrutiert
❸ Praxis die keine Patienten rekrutiert

Abb. 2. Mögliche Quellen für Bias in clusternrandomisierten Studien (adaptiert nach Giradeau & Ravaud 2009).

Tabelle 1. Vor- und Nachteile der Cluster-Randomisierung.

Vorteile	Nachteile
• Pragmatisch • Beste Möglichkeit zur Evaluation organisatorischer Veränderungen • Soll „Kontamination" verhindern	• Ähnlichkeiten in einem Cluster reduzieren die Variabilität • Annahme der unabhängigen Patientenstichprobe ist nicht mehr gegeben • verminderte effektive Stichprobengröße • Verblindung meist nicht möglich • Biasanfällig • Ethische Probleme • Analyse aufwendig

die Stichprobe verzerren [8]. Durch eine konsekutive Rekrutierung versucht man diesen Bias zu minimieren, was in der Praxis aber oft nicht gelingt. Praxen in den Kontrollarmen sind meist weniger motiviert und schließen weniger Patienten ein. Weitere Ursachen für Verzerrungen sind, dass die Intervention von Patienten als attraktiv oder anstrengend empfunden wird und so eher Patienten mit einen höheren oder niedrigen Ausprägungsgrad der Zielerkrankung in die Studie eingeschlossen werden.

Die Vor- und Nachteile der Cluster-Randomisierung sind in Tabelle 1 zusammengefasst. Wegen der vielen Nachteile durch die Clusterung müssen starke Gründe für dieses Studiendesign sprechen [5,9].

Selektions- und Subsampling-Bias

Beispielstudie 1: In den beiden Interventionsarmen die eine Schulung an mehreren Terminen erfordertet sind nach der Randomisierung jeweils 5 Praxen ausgeschieden, im Kontrollarm nur eine. Umgekehrt fordert eine Praxis aus dem Kontrollarm bei der Rekrutierung in eine andere Studie „Wir machen diesmal nur mit wenn wir in den Interventionsarm kommen". Die 42 Praxen im Kontrollarm ohne Implementierungsaktivitäten waren bei der Rekrutierung weniger motiviert und schlossen ca. 15% weniger Patienten mit Rückenschmerzen in die Studie ein, die Patienten waren im Schnitt etwas kränker als in den Interventionsarmen. Arzthelferinnen spielen bei der Rekrutierung in Hausarztpraxen eine besondere Rolle. Sie sprachen Interventionsarm mit Motivierender Beratung durch Arzthelferinnen häufiger Frauen und jüngere Patienten an.

Ethische Probleme

Eine Besonderheit von cluster-randomisierten Studien in Allgemeinmedizinpraxen ist, dass statt der potentiellen Studienteilnehmer die Praxis in die Studie einwilligt. Es ist umstritten, ob bei Studien in der Routineversorgung ohne zusätzliche Datenerfassung auf eine Aufklärung und Einverständniserklärung der individuellen Patienten evt. verzichtet werden kann [9]. Werden zusätzliche Patientendaten erhoben ist auf jeden Fall eine doppelte Einwilligung (Praxis und Patienten) notwendig. Patienten in der Standardgruppe (treatment as usual) müssen im Regelfall nicht über Intervention in der Interventionsgruppe, zu der sie keinen Zugang haben, aufgeklärt werden. Über die Notwendigkeit und Umfang einer Aufklärung und Einverständniserklärung muss eine Ethikkommission im Einzelfall entscheiden [10].

Der Intraclusterkorrelationskoeffizient

Der Anteil der Gesamtvarianz des Studienendpunkts, der durch die Clusterzugehörigkeit aufgeklärt wird kann durch den sog. Intraclusterkorrelationskoeffizient (ICC oder manchmal auch ρ) ausgedrückt werden [11]. Die Varianz ist die Streuung oder Verteilung des Studienendpunkts in einem Studienarm oder Cluster [9]. Es gibt mehrere Formeln zur Berechnung des ICC. Eine häufig verwendete Variante ist der Quotient aus der Varianz (S) innerhalb eines Clusters (Sw w=within) und der

Gesamtvarianz. Letztere ist die Summe der Varianz innerhalb des Clusters und der Varianz zwischen den Clustern (S_b b= between) **(Formel 1)**. Er nimmt Werte zwischen 0 und 1 an.

Formel 1: Berechnung des Intraclusterkorrelationskoeffizient (Erläuterung der Abkürzungen im Text)

$$ICC = \frac{S_b^2}{(S_b^2 + S_w^2)}.$$

In den meisten primärärztlichen Studien liegt der ICC zwischen 0,05–0,15 [11]. Cluster-randomisierte Studien sollten ihren post-hoc ermittelten ICC publizieren, damit er für ähnliche Studien zu Abschätzung zur Verfügung steht [12]. Auch der ICC ist nur eine Schätzung und hat einen Konfidenzintervall [13]. Bei kleinen Studien mit wenigen Clustern lässt sich der ICC nicht zuverlässig abschätzen.

Stichprobenberechnung

Die Stichprobenberechnung (power calculation) dient dazu, die Anzahl der Studienteilnehmer zu schätzen, die benötigt werden um einen bestimmten Effekt nachzuweisen [14]. Bei der Berechnung spielt die Varianz des Studienendpunkts und der erwartete Unterschied zwischen den Studienarmen eine entscheidende Rolle. Je höher die Varianz (Hintergrundrauschen), umso mehr Studienteilnehmer werden gebraucht, um im Rauschen den Effekt der Intervention zu ermitteln. Dabei muss eine Irrtumswahrscheinlichkeit β gewählt werden einen Effekt zu übersehen (Fehler zweiter Art). Üblicherweise wird ein β zwischen 0,1 bis 0,2 gewählt. Als Power (1-β) ausgedrückt ist die Aussagekraft einer Studie einen vorhanden Effekt mit 80 bis 90% Wahrscheinlichkeit aufzudecken. Im Allgemeinen wird eine Power von mindestens 80% angestrebt [14].

Da wie bereits erwähnt die Varianz durch die Clusterung abnimmt ist muss bei cRCTs die Stichprobengröße angepasst werden. Der benötigte Stichprobenumfang wird umso größer sein, je höher der erwartete ICC ist. Die Stichprobengröße muss mit dem Design-Effekt (DE) korrigiert werden **(Formel 2)**, dabei ist m die Anzahl der Patienten in einem Cluster. Der Design-Effekt ist im Regelfall eine Zahl zwischen 1 und 2 [11]. Ein Design-Effekt von 1,7 bedeutet z.B., dass die errechnete Stichprobe von 100 Patienten auf 170 erhöht werden sollte (100 × 1,7=170).

Formel 2: Berechnung des Design-Effekts (Erläuterung der Abkürzungen im Text)

$$\text{Design-Effekt DE} = 1 + \text{ICC}\ (m - 1).$$

Da der wahre ICC erst post-hoc ermittelt werden kann, muss für die Stichprobenberechnung auf eine Schätzung zurückgegriffen werden. Auch wenn in der Studienplanung meistens von gleichgroßen Clustern ausgegangen wird, ist dies in der Praxis fast nie der Fall. Besondere Anpassungen sind bei unterschiedlichen Gruppengrößen der einzelnen Cluster bei der Analyse notwendig [15].

Die Auswirkung der Zahl und Größe der Cluster auf die Power ist anschaulich in Tabelle 2 dargestellt [16]. Als effektive Stichprobengröße (ESS) bezeichnet man die für die Clusterung adjustierten Stichprobengröße **(Formel 3)**. Wobei k Zahl der Cluster mit der (durchschnittlichen) Zahl der Patienten m pro Cluster multipliziert die nichtadjustierte Stichprobengröße ist.

Formel 3: Effektive Stichprobengröße (Erläuterung der Abkürzungen im Text)

$$ESS = \frac{mk}{DE}.$$

Stichproberechnung

Beispielstudie 1: Für die Berechnung des Stichprobenumfangs wurde für eine Power 80% (ß=0,2), Signifikanzniveau α=0,05 bei einer angenommen kleinen Differenz von 0,1 (Unterschied 10% Funktionskapazität) und ein ICC von 0,03 errechnet, dass bei 40 Praxen in jedem Studienarm 16 Patienten pro Praxis rekrutiert werden müssen. Zusätzlich wurde eine angenommene Drop-out-Rate von 25% berücksichtigt.

Beispielstudie 2: Für die Berechnung des Stichprobenumfangs wurde für eine Power 90% (ß=0,1), Signifikanzniveau α=0,05 bei einer angenommen kleinen Differenz von 0,1 (Unterschied 10% Lebensqualität) und ein ICC von 0,03 errechnet, dass bei 25 Praxen in jedem Studienarm 14 Patienten pro Praxis rekrutiert werden müssen. Zusätzlich wurde eine angenommene Drop-out-Rate von 10% berücksichtigt.

Datenanalyse

Um eine mögliche Verzerrung der Ergebnisse durch die Clusterung bei der Analyse zu berücksichtigen, muss bei der Auswertung eine sog. Cluster-Adjustierung durchgeführt werden. Viele cRCTs werden ohne Adjustierung ausgewertet, man spricht dann von einer „naiven Analyse". Dies führt zur Schätzung von zu kleinen Konfidenzin-

Tabelle 2. Effektive Stichprobengröße und Power bei konstanter Stichpobe.

Praxen k	Patienten m	Gesamt (mk)	ρ= 0.017 DE	ESS	Power t-Test*
4	32	128	1.527	84	61
8	16	128	1.255	102	70
16	8	128	1.119	114	75
32	4	128	1.051	122	78
64	2	128	1.017	126	79
128	1	128	1.000	128	80

*Power in % bei Annahme einer Effektgröße von 0.5, wenn die Hälfte der Clusters in die Intervention bzw. die Kontrollgruppe randomisiert wird und der Studienendpunkt der Vergleich eines Mittelwertes mit einem zweiseitigen t-Test durchgeführt wird mit dem einem Signifikanzniveau von 0,05.
m=Anzahl der Patienten in einem Cluster; k=Anzahl der Praxen/Clusters; mk=Gesamtzahl der eingeschlossenen Patienten, DE=Designeffekt; and ESS=Effektive Stichprobengröße.

tervallen und erhöht die Wahrscheinlichkeit einen Signifikanten Unterschied zu finden obwohl keiner besteht (Fehler erster Art) [17,18]. In diesem Artikel können nur die Grundprinzipien der clusteradjustierten Auswertung beschrieben werden.

Die einfachste aber nicht sehr zu empfehlende Methode zur Adjustierung beim Vergleich von Intervention und Kontrolle ist, anstatt jeden individuellen Einzelwert den Clusterdurchschnitt als Datenpunkt in jeden Studienarm auszuwerten [19].

Alternativ kann durch Anpassung von univariaten Teststatistiken mit dem Designeffekt die individuellen Messwerte als einzelne Datenpunkte ausgewertet werden [20]. Zum Beispiel kann beim Vergleich kontinuierlicher Daten mit dem T-Test der T-Wert durch die Wurzel des Desing-Effekts geteilt werden, wodurch sich die Anforderung an die Signifikanz erhöhen.

Diese einfachen (univariaten) Auswertungen ermöglichen es aber nicht für den Einfluss anderer Faktoren (Confounder) auf den Studienendpunkt zu adjustieren. Für adjustierte multivariate Auswertung stehen mehrere statistische Methoden zur Verfügung.

Bei der Kovarianzanalyse (ANCOVA) wird die Clusterzugehörigkeit als Kovariate im Model berücksichtigt. Dieses Verfahren ist sinnvoll, wenn nur wenige Cluster vorliegen, bzw. wenn Unterschiede zwischen einzelnen Clustern (z.B. Praxen) interessieren. Im Gegensatz hierzu wird in der hierarchischen Mehrebenenanalyse (multilevel analysis), ein Zufallseffekt definiert, der die Varianz der Zielgröße (z.B. Therapieerfolg), ggfs. in Abhängigkeit von Kovariaten, über verschiedene Cluster beschreibt. Insbesondere wenn viele Cluster vorliegen, ist dieses Verfahren vorzuziehen. Die korrekte Formulierung eines solchen Modells kann komplex sein und erfordert einschlägige biometrische Expertise. Grundsätzlich ist es auch möglich, mehr als zwei Hierarchieebenen zu berücksichtigen [21]. Ein Beispiel dafür ist z.B. die Untersuchung der Wirksamkeit Sicherheitshinweise zur Verhütung von Haushaltsunfällen für Familien mit Kindern durch Hausärzte [22]. Hier liegt eine Clusterung sowohl auf Praxen – als auch auf Familienebene vor.

Datenanalyse
Beispielstudie 1: Die Datenanalyse für den Studienendpunkt (Funktionskapazität) erfolgte mit einem Mixed-Effect-Model mit Clusteradjustierung. Als Kovariable wurden unter anderen Schmerz in den letzten 6 Monaten berücksichtigt. Der ermittelte ICC wird nicht angegeben.
Beispielstudie 2: Die Datenanalyse für den Studienendpunkt (Lebensqualität) erfolgt mit Kovarianzanalyse (ANCOVA) mit Clusteradjustierung. Als Kovariable wurden unter anderen Depression berücksichtigt. Die für die Variablen ermittelten ICCs werden nicht angegeben.

Berichterstattung

Für die nachvollziehbare Berichterstattung (Reporting) von RCTs ist das sog. CONSORT-Statement entwickelt worden. Für Cluster-ranomisierte RCTs gibt es eine Erweiterung des CONSORT-Statements, das im Wesentlichen die zusätzliche Angabe der Anzahl der Cluster und der Clustergröße vorsieht [23].

Ergebnis
Beispielstudie 1: Die Intervention hatte einen geringen (statistisch signifikanten) Effekt auf die Funktionskapazität bei Rückenschmerzpatienten, der im Studienarm mit der Beratung durch die Arzthelferinnen ausgeprägter war. Patienten in den Interventionsarmen hatten weniger Schmerztage.
Beispielstudie 2: Die Intervention hatte keinen Effekt auf den primären Studienendpunkt Lebensqualität bei Arthrosepatienten. Nur bei einigen sekundären Endpunkten konnten positive Effekte nachgewiesen werden.

Schlussfolgerung

Cluster-Randomisierung ist ein in der allgemeinmedizinischen Forschung häufiges und adäquates Studiendesign, wenn pragmatische Gründe oder ein erhöhtes Kontaminationsrisiko vorliegen. Sie stellt allerdings erhöhte methodische Anforderungen, da die Clustereffekte bei der Stichprobengrößenberechnung und der Analyse berücksichtigt werden müssen. Diese sollten Forscher in der Primärversorgung in Grundzügen verstehen.

Anmerkung

Das Manuskript basiert auf einen 2008 gehaltenen Prüfungsvortrag für das Zertifikat Epidemiologie der Deutschen Gesellschaft für Epidemiologie, der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie und der Deutschen Gesellschaft für Sozialmedizin und Prävention.

Interessenkonflikte

Es bestehen keine Interessenskonflikte.

Danksagung

Für kritische Durchsicht und konstruktive Kritik bedanke ich mich bei Prof. Dr. Martin Scherer, Dr. Anne Simmenroth-Nayda und Dr. Carsten Schmidt.

Literatur

[1] Ovhed I, van Royen P, Håkansson A. What is the future of primary care research? Probably fairly bright, if we may believe the historical development. Scand J Prim Health Care 2005;23:248–53.
[2] Bland JM. Cluster randomised trials in the medical literature: two bibliometric surveys. BMC Med Res Meth 2004;4:21.
[3] Becker A, Leonhardt C, Kochen MM, et al. Effects of two guideline implementation strategies on patient outcomes in primary care: a cluster randomized controlled trial. Spine 2008;33:473–80.
[4] Rosemann T, Joos S, Laux G, et al. Case management of arthritis patients in primary care: a cluster-randomized controlled trial. Arthritis Rheum 2007;57:1390–7.
[5] Torgensen DJ. Contamination in trials: is cluster randomisation the answer?. BMJ 2008;322:355–7.

[6] Bonell CP, Hargreaves JR, Cousens SN, et al. Alternatives to randomisation in the evaluation of public-health interventions: design challenges and solutions. J Epidemiol Community Health 2009 Feb 12; [Epub ahead of print].
[7] Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. BMC Med Res Methodol 2005;5:10.
[8] Giraudeau B, Ravaud P. Preventing Bias in Cluster Randomised Trials. PLoS Med 2009;6:e1000065.
[9] Winkens RA, Knottnerus JA, Kester AD, Grol RP, Pop P. Fitting a routine health-care activity into a randomized trial: an experiment possible without informed consent?. J Clin Epidemiol 1997;50:435–9.
[10] Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999;318:1407–9.
[11] Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. BMJ 1998;316:1455.
[12] Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. BMC Med Res Methodol 2004;4:9.
[13] Ukoumunne OC. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. Stat Med 2002;21:3757–74.
[14] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005;365:1348–53 [Deutsche Version Z Arztl Fortbild Qualitatssich 2006;100:129–35].
[15] Manatunga AK, Hudgens MG, Chen S. Sample size estimation in cluster randomized studies with varying cluster size. Biometrical Journal 2001;43:75–86.
[16] Killip S, et al. What is an intracluster correlation coefficient? Crucial concepts for primary care researches. Ann Fam Med 2004;2:204–8.
[17] Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. BMJ 1998;317:1171–2.
[18] Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. J Clin Epidemiol 2003;56:341–50.
[19] Kerry SM, Bland JM. Analysis of a trial randomised in clusters. BMJ 1998;316:54.
[20] Campbell MK, et al. Analysis of cluster randomized trials in primary care: a practical approach. BMJ 1998;316:1455.
[21] Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. Am J Epidemiol 1994;140:279–89.
[22] Clamp M, Kendrick D. A randomised controlled trial of general practitioner safety advice for families with children under 5 years. BMJ 1998;316:1576–9.
[23] Campbell MK, et al. CONSORT statement: extension to cluster randomised trials. BMJ 2004;328:702–8.