



# Der Goldstandard medizinischer Studien

Ein Workshop der Österreichischen Cochrane Zweigstelle an der Donau-Uni Krems befasste sich mit der kritischen Evaluierung von randomisierten, kontrollierten Studien. Von M. Strausz

Klinische Forschung ist die Grundlage für medizinischen Fortschritt und Voraussetzung für evidenzbasierte Medizin. „Um die Wirksamkeit und Sicherheit einer Intervention zu untersuchen, sind randomisierte, kontrollierte Studien (RCTs: randomised controlled trials) der Goldstandard“, erklärte Prof. Dr. Gerald Gartlehner, Leiter des Departments für Evidenzbasierte Medizin und Klinische Epidemiologie der Donau-Universität Krems im Rahmen eines Workshops in Krems.

RCTs verfügen über das passende Studiendesign, um kleine, aber klinisch wesentliche Behandlungseffekte zu untersuchen, da sie am besten in der Lage sind, Bias (Verzerrungen) und Confounding (Störfaktoren) zu minimieren. Mit diesen Studien kann beispielsweise die Überlegenheit einer neuen Therapie im Vergleich zu einer Standardtherapie oder Placebo nachgewiesen werden. In RCTs werden die Studienteilnehmer durch unterschiedliche Randomisierungsfaktoren per Zufallsprinzip einer Interventions- oder Kontrollgruppe zugeteilt. Die Unvorhersehbarkeit der Zuteilung führt zu annähernd gleichen Studiengruppen mit gleichmäßig verteilten prognostischen Faktoren, die in nicht-randomisierten Studien als Confounder Ergebnisse verzerren können. Confounding liegt vor, wenn ein Störfaktor sowohl mit der Intervention/Exposition als auch mit dem Studienausgang in Zusammenhang steht und diesen beeinflusst bzw. verfälscht.

RCTs werden immer prospektiv durchgeführt.

Eine Schwäche der RCTs ist, dass durch die strenge Selektion der Studienpopulation die Anwendbarkeit der Ergebnisse auf die Normalbevölkerung (externe Validität) eingeschränkt ist. Daher versuchen pragmatische RCTs durch geringere Selektion, praxisrelevanter Durchführung und den Einsatz patientenrelevanter Endpunkte eine höhere externe Validität als explanatorische RCTs (Studien mit stark selektierten Populationen unter hochkontrollierten Bedingungen) zu erzielen.

## Bewertungskriterien von RCTs

Ein Hauptkriterium ist die Frage nach der adäquaten Form der Randomisierung für die Zuteilung der Studienteilnehmer in die Behandlungsgruppen. „Wichtig ist, dass das Zufallsprinzip und die Unvorhersehbarkeit der Zuteilung gewahrt werden“, betonte Gartlehner. Es gibt unterschiedliche Arten der Randomisierung (beispielsweise einfache Randomisierung, Blockrandomisierung, stratifizierte Randomisierung u.a.), die abhängig von Größe und Design der jeweiligen Studie eingesetzt werden können.

Die besten Randomisierungen liefern computergenerierte Zuteilun-

gen. Beispiele für inkorrekte Randomisierungsverfahren sind hingegen die Sortierung nach Geburtsdaten, nach personenbezogenen Daten wie Ambulanz- oder Sozialversicherungsdaten oder abwechselnde Zuordnung (ABAB) – die ersten fünf etc.

Häufig wird das Auswahlverfahren als „Quasi-Randomisierung“ bezeichnet. „Hier sollten die Alarmglocken läuten, denn dieser Begriff wird zumeist bei einer unvollständig durchgeführten Randomisierung verwendet, bei der das Zufallsprinzip oder die Unvorhersehbarkeit nicht sichergestellt wurden“, warnte Gartlehner.

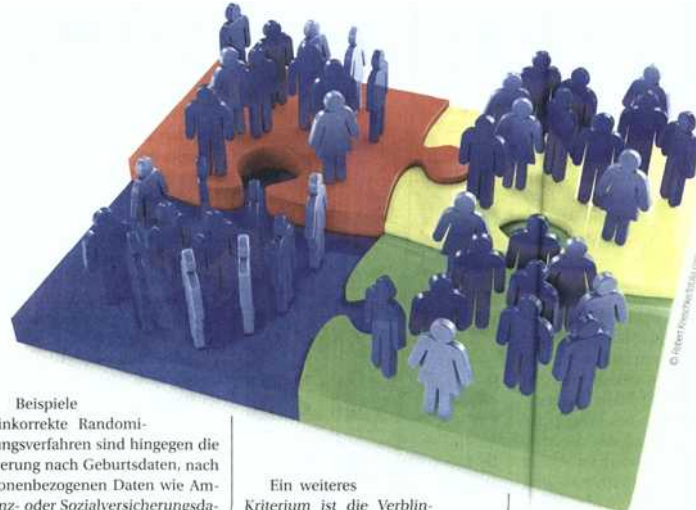
Neben dem auf dem Zufallsprinzip basierenden Erstellen einer Reihenfolge (Randomisierungssequenz) ist die Geheimhaltung dieser Sequenz (Allocation Concealment) ein zweiter, bedeutender Randomisierungsaspekt. Bei korrekt durchgeführter Geheimhaltung kann der Untersucher nicht erkennen, welcher Studiengruppe der nächste Proband angehört wird. Kennt der Untersucher schon im Vorfeld das Zuordnungsprinzip, könnte er darauf Einfluss nehmen, was das Studienergebnis verfälschen könnte. Methodenstudien haben gezeigt, dass mangelnde Geheimhaltung zu einer Überschätzung von Behandlungseffekten führt.

Beispiele für adäquate Geheimhaltung der Randomisierungssequenz:

- sequenziell nummerierte, verschlossene, undurchsichtige Umschläge (sequentially numbered, opaque, sealed envelopes: SNOSE),
- seriell nummerierte oder kodierte identische Container,
- zentrale Randomisierung (Internet oder telefonisch),
- jeder andere Prozess, der verhindert, dass die Reihenfolge der Gruppenzuordnung vorhergesehen werden kann.

Beispiele für inadäquate Geheimhaltung:

- durchsichtige, unverschlossene Briefumschläge,
- durchgehend nummerierte Briefumschläge,
- offene Randomisierungslisten,
- spezifische, personenbezogene Daten.



Ein weiteres Kriterium ist die Verblindung, die dazu dient, eine eventuelle unabsichtliche Beeinflussung der Studienergebnisse durch Untersucher oder Teilnehmer auszuschalten. Verblindung ist besonders dann maßgeblich, wenn Zielparameter gemessen werden, die durch Subjektivität beeinflussbar sind (z.B. Schmerzen). Bei harten Endpunkten (Mortalität) ist eine Verblindung hingegen nicht nötig.

Studien können einfach oder doppelt verblindet sein. Bei einer einfach verblindeten Studie weiß der Patient nicht, in welcher Studiengruppe er sich befindet, bzw. der Untersuchende nicht, welcher Intervention der Patient zugeteilt wurde. Bei einer doppelt verblindeten Studie haben beide keine Kenntnis, welcher Interventionsgruppe der Patient angehört. Doppel-Verblindung ist nicht immer möglich, beispielsweise bei chirurgischen Eingriffen, in solchen Fällen sollte derjenige, der die Endpunkte beurteilt (Evaluator blinding), verblindet sein.

## Die Drop-out -Rate

Als Ausfallrate (Attrition, Loss to follow-up) bezeichnet man den Prozentsatz an Personen, die randomisiert wurden, jedoch nicht bis zum Ende der Studie teilgenommen haben. Dabei wird zwischen der allgemeinen und der differenziellen Drop-out-Rate unterschieden. Erste misst den Prozentsatz aller Studienteilnehmer, deren Daten nach erfolgreicher Randomisierung aus diversen Gründen (Nebenwirkungen, mangelnde Adherence, ...) nicht mehr verfügbar sind. „Es gibt keine klaren Richtlinien, wie hoch die allgemeine Drop-out-Rate sein darf. Der Richtwert wird häufig mit 20 Prozent angenommen, aus pragmatischen Gründen setzt man bei speziellen Populationen (psychiatrisch Erkrankte) ein höherer Grenzwert an“, so Gartlehner. Bei der differenziellen Drop-out-Rate, die den Unterschied im Verhältnis der Drop-out-Raten der einzelnen Behandlungsgruppen misst, wird als Grenzwert

15 Prozent angegeben. Allgemein gilt, dass Drop-out-Raten ein hohes Verzerrungspotenzial haben, wenn der Verhältniswert von Personen ohne Enddaten zu Personen, die bis zum Studiende teilgenommen haben, hoch ist.

## Studienendpunkte und Analyse

Eine gute RCT zeichnet sich dadurch aus, dass alle erhobenen Endpunkte auch in der Studie berichtet werden. Bei publizierten Studien gilt es zu überprüfen, ob die primären Endpunkte, die in den Methoden angeführt wurden, auch in den Resultaten adäquat behandelt wurden. Denn Publikationsbias und Outcome Reporting Bias sind ein häufiges Problem in der medizinischen Literatur, können aber nur schwer überprüft werden. Falls das Studienprotokoll registriert wurde, kann man im Register die a priori definierten Endpunkte einsehen. Um Messbias zu vermeiden, müssen die Endpunkte in allen Behandlungsgruppen zum selben Zeitpunkt gemessen werden.

„Eine Intention-to-treat (ITT)-Analyse gilt als beste statistische Auswertung für RCTs, weil sie die Vorteile einer Randomisierung erhält“, so Gartlehner. Diese bedeutet: Idealerweise sollten alle Patienten, die in einer Gruppe randomisiert wurden, auch bei der Studienanalyse als dieser Gruppe zugehörig mit einberechnet werden. Andere statistische Auswertungen wie Per-protocol-Analysen, Completers-only-Analysen oder As-treated-Analysen sollten nur gemeinsam mit ITT-Analysen durchgeführt werden.

Workshop: „Kritische Bewertung medizinischer Studien“ 5. bis 6. März, Donau-Universität Krems. Veranstalter waren die Österreichische Cochrane Zweigstelle (ÖCZ) in Kooperation mit dem Department für Evidenzbasierte Medizin und Klinische Epidemiologie.