

Schlüsselstudien zur Herzinsuffizienz: Zweifel an ihrer externen Validität

Bei Analysen und Schlussfolgerungen aus Ergebnissen randomisierter kontrollierter Studien (RCTs) muss auf die *interne* und *externe Validität* geachtet werden. Die *interne Validität* beschreibt, inwieweit die Ergebnisse der Studie für die gewählte Stichprobe valide und „reliabel“ (methodisch zuverlässig) sind. Häufige Fehler (Verzerrungen, engl. „Bias“) sind: „Attrition“- , „Detection“- , „Observer“- und „Performance“-Bias. Hinzukommen viele mögliche statistische Verzerrungen (Übersicht bei ^[1]; vgl. auch ^[2]). Die *externe Validität* beschreibt, inwieweit die Ergebnisse einer Studie auch auf Patienten außerhalb der Studie, also in die „real world“ übertragen werden können. Idealerweise repräsentiert die in der Studie gewählte Stichprobe die allgemeine Bevölkerung bzw. die betreffenden Patienten im klinischen Alltag. Dass dies sehr häufig nicht der Fall ist, haben wir am Beispiel der Unterrepräsentation von Frauen oder älteren Menschen in RCTs schon häufig kritisiert (vgl. ^[3]).

Forscher aus Graz haben kürzlich die externe Validität von 6 Schlüsselstudien („pivotal trials“) zur Herzinsuffizienz mit erhaltener bzw. mittelgradig reduzierter linksventrikulärer Ejektionsfraktion (HFpEF bzw. HFmrEF) untersucht ^[4]. Sie verglichen die Ein- und Ausschlusskriterien („eligibility“) aus diesen Studien mit den klinischen Charakteristika einer eigenen, unselektierten Kohorte („all comers“) und das Gesamtüberleben.

Die zum Vergleich herangezogenen randomisierten, kontrollierten Studien (RCT) waren: CHARM-Preserved und I-PRESERVE mit den Angiotensin-II-Rezeptorblockern (AT-II-RB) Candesartan und Irbesartan; PARAGON-HF mit dem Angiotensin-Rezeptor-Nepriylisin-Inhibitor (ARNI) Valsartan/Sacubitril; TOPCAT mit dem Mineralokortikoid-Rezeptor-Antagonisten (MRA) Spironolacton; EMPEROR-Preserved und DELIVER mit den „Sodium Glucose Linked Transporter 2“-Hemmern (SGLT2-Hemmer) Empagliflozin und Dapagliflozin (vgl. ^[5] sowie die Liste der Studiennamen auf unserer Webseite mit den „Links“ zu den Artikeln in DER ARZNEIMITTELBRIEF; ^[6]).

Die Therapieempfehlungen aus den aktuellen Herzinsuffizienz-Leitlinien der europäischen kardiologischen Gesellschaft (ESC; ^[7]) basieren auf diesen RCTs. Bei der HFmrEF und der HFpEF haben SGLT2-Hemmer eine Klasse-I-(A) und AT-II-RB, ARNI und MRA eine Klasse-IIb-Empfehlung („may be considered“) erhalten.

Methodik: Die Forscher durchsuchten ihre Klinikdatenbank nach Patienten, die sich innerhalb eines Jahres (8/2018-7/2019) in der internen Notaufnahme des Univ.-Klinikums Graz vorstellten und bei denen anhand der klinischen Angaben sowie der Echokardiografie- und Laborbefunde (proBNP-Werte) die Diagnosen HFpEF oder HFmrEF gestellt wurden. Patienten mit Demenz, Palliativ- und Heimpflege wurden von der Analyse ausgeschlossen (allgemeine Ausschlusskriterien für RCTs).

Die klinischen Charakteristika der übrigen Patienten wurden abgeglichen mit den Ein- und Ausschlusskriterien aus den genannten RCTs. Es wurde so eine „eligibility-rate“ (in Prozent) für jede der 6 RCTs berechnet und in einem weiteren Schritt der Endpunkt Überleben in Abhängigkeit von der „eligibility“. Hierzu wurde der Vitalstatus der Kohorte zum Stichtag 31.12.2020 vom Statistischen Bundesamt Österreich abgerufen.

Ergebnisse: Unter den 1.248 aufgenommenen Patienten mit der Diagnose Herzinsuffizienz waren 554, bei denen die spezielleren Diagnosen HFpEF oder HFmrEF gestellt wurden. Insgesamt hatten 147 allgemeine Ausschlusskriterien (s.o.). Die verbliebenen 407 bildeten die Studienkohorte. Diese war im Median 79 Jahre alt, und 52% waren Frauen. Im Median betrug die linksventrikuläre Auswurfraction (LVEF) 50% und der NT-proBNP-Wert 3.088 pg/ml. Bei 65% wurde die Diagnose HFpEF und bei 35% HFmrEF gestellt. Die Kenndaten stimmten weitestgehend mit denen in den RCTs überein. Wichtigster Unterschied: in der Grazer Kohorte befanden sich mehr akut dekompensierte Patienten: 61% hatten bei der Vorstellung in der Notaufnahme Beinödeme, 55% Pleuraergüsse, 39% feuchte Rasselgeräusche und 72% erhielten Diuretika. Von der Gesamtzahl dieser „allcomers“ wären folgende Anteile grundsätzlich „eligible“ für die RCTs gewesen: 88% für CHARM-Preserved, 51% für EMPEROR-Preserved, 49% für DELIVER, 40% für I-PRESERVE, 35% für TOPCAT sowie 28% für PARAGON-HF. Nach Anwendung der für die Studien jeweils spezifischen Ausschlusskriterien, wie z.B. chronisch obstruktive Lungenerkrankung (COPD) mit Heim-Sauerstofftherapie, schweres Vitium oder fortgeschrittene Nierenerkrankungen, sanken die korrespondierenden „eligibility rates“ auf 80%, 22%, 29%, 17%, 8% sowie 11%.

Somit erweisen sich die Studienpopulationen bei mindestens 5 dieser RCTs als hoch selektiert und keinesfalls repräsentativ für die meisten Patienten mit HFpEF oder HFmrEF. Die Ergebnisse aus den beiden RCTs mit SGLT2-Hemmern (EMPEROR-Preserved und DELIVER) sind beispielsweise nur auf jeden fünften Patienten mit HFpEF oder HFmrEF anwendbar. Trotzdem haben diese Wirkstoffe in der ESC-Leitlinie aus dem vergangenen Jahr den höchsten Empfehlungs- und Evidenzgrad erhalten. Die hohe Selektion bei der Auswahl der Patienten in den RCTs ist auch daran erkennbar, dass trotz der allgemein hohen Prävalenz dieser Krankheitsentitäten, die Rekrutierungsraten in den 6 RCTs nur bei 2,4-7,6 Patienten pro Jahr und Zentrum lagen.

Eine „Non-Eligibility“ wirkt sich statistisch erheblich auf den Endpunkt Überleben aus. In der Nachbeobachtungszeit von im Median 22 Monaten traten in der Grazer Kohorte insgesamt 121 Todesfälle auf (Sterberate 30%). Zum Vergleich: In den 6 RCTs betrug die Sterberate in einem Zeitraum zwischen 27-50 Monaten 11-16%. Während Patienten, die für die RCTs „eligible“ gewesen wären, etwa die gleichen Sterberaten hatten wie in den RCTs, lag die Sterbewahrscheinlichkeit bei „non-eligibility“ 2-3fach höher („Hazard Ratio“: 1,71-2,97).

Diskussion: Für die Durchführbarkeit von Studien sind enge Ein- und Ausschlusskriterien wichtig. Wenn jedoch dadurch überwiegend Patienten mit geringen Risiken eingeschlossen werden, entsteht ein falscher Eindruck, sowohl hinsichtlich der Wirksamkeit und Verträglichkeit bestimmter Therapien, als auch hinsichtlich der Prognose. Bevor in Leitlinien hochgradige Therapieempfehlungen abgegeben werden, wie etwa für die SGLT2-Hemmer bei HFpEF und HFmrEF, sollten Informationen zur externen Validität vorliegen und berücksichtigt werden. Dies sollte in Zeiten großer Datenbanken und smarterer

Analysesysteme mit überschaubarem Aufwand zu bewerkstelligen sein.

Fazit

Im Design der meisten Schlüsselstudien zur medikamentösen Behandlung von Herzinsuffizienz-Patienten mit erhaltener bzw. mittelgradig reduzierter linksventrikulärer Ejektionsfraktion wurden > 70% der Betroffenen ausgeschlossen. Daher ist eine vorschnelle Verallgemeinerung der Studienergebnisse unzulässig. Bevor in Leitlinien hochgradige Therapieempfehlungen abgegeben werden, sollten mehr Informationen zur externen Validität vorliegen, beispielsweise aus Registerstudien.

Literatur

1. Glossar des Deutschen Netzwerks Evidenzbasierte Medizin e.V.: <https://www.ebm-netzwerk.de/de/service-ressourcen/ebm-glossar> (Zugriff am 29.2.2024). ([Link zur Quelle](#))
2. AMB 2016, 50, 96DB01. AMB 2017, 51, 64DB01. AMB 2019, 53, 79b. AMB 2020, 54, 36DB01. ([Link zur Quelle](#))
3. AMB 1999, 33, 25. AMB 2017, 51, 64DB01. AMB 2020, 54, 69. ([Link zur Quelle](#))
4. Santner, V., et al.: Eur. J. Int. Med. 2023, 118, 73. ([Link zur Quelle](#))
5. AMB 2014, 48, 52. AMB 2020, 54, 03. AMB 2021, 55, 81. AMB 2023, 57, 81. ([Link zur Quelle](#))
6. <https://der-arzneimittelbrief.com/studien> ([Link zur Quelle](#))
7. McDonagh, T.A., et al.: Eur. Heart J. 2023, 44, 3627. ([Link zur Quelle](#))