



---

**AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN**

---

# Telemonitoring bei Herzinsuffizienz und die Grenzen der "Real-World"-Evidenz aus nicht-randomisierten Studien

**VON PROF. DR. STEFAN SAUERLAND UND DR. LARS G. HEMKENS IM AUFTRAG DES DEUTSCHEN NETZWERKS EVIDENZ-BASIERTE MEDIZIN E.V. (DNEBM – WWW.EBM-NETZWERK.DE)**



Am 17.12.2020 hat der Gemeinsame Bundesausschuss (G-BA) beschlossen, ein Telemonitoring für Patientinnen und Patienten mit Herzinsuffizienz einzuführen [1].

Das Behandlungskonzept besteht darin, dass Vitalparameter und weitere Messwerte entweder durch kardiale Implantate oder durch externe Geräte erfasst und mindestens einmal täglich an ein telemedizinisches Zentrum weitergegeben werden. Dort kann eine behandlungsbedürftige Verschlechterung frühzeitig erkannt werden, um dann durch geeignete Interventionen eine Dekompensation der Herzinsuffizienz im Ansatz zu verhindern. Telemonitoring soll den Drehtür-Effekt, also sich häufig wiederholende ungeplante Krankenhausaufnahmen aufgrund von Herzinsuffizienz, verhindern und letztlich nicht nur die kardiale Morbidität, sondern auch die Mortalität reduzieren.

Organisatorisch und ärztlich führt Telemonitoring in Neuland. Der bzw. die primär behandelnde Arzt/

Ärztin (PBA) behält zwar die Kontrolle über die Behandlung. Parallel aber überwacht ein/e Kardiologe/ in im Telemedizin-Zentrum (TMZ) den Erkrankungsverlauf aus der Distanz und weist gegebenenfalls auf Änderungsbedarf hin.

Ein solches Dreiecksverhältnis zwischen PBA, TMZ und Patient braucht gute Organisation, klare Spielregeln und viel Vertrauen. Der Erfolg des Telemonitorings hängt neben den technischen Voraussetzungen von schnellem, zielgerichtetem, kooperativem Handeln ab.

Telemonitoring als Ergänzung zur ambulanten Präsenzmedizin von Haus- und Facharzt wird vor allem in ländlichen Regionen, wo das ärztliche Versorgungsnetz dünn ist, sehr begrüßt werden. Aufgrund der COVID-19-Pandemie steigt das Interesse an Telemedizin und Telemonitoring aber insgesamt deutlich an. Mittelfristig ist damit zu rechnen, dass auch bei anderen Krankheiten telemedizinische Konzepte zunehmen werden.

Die wissenschaftliche Bewertung von Telemonitoring ist nicht ganz einfach, weil es sich um eine komplexe Intervention handelt, bei der die verschiedenen Elemente der Behandlung sehr variabel gestaltet werden können, sodass sich Nutzen und Schaden deutlich zwischen verschiedenen Varianten von Telemonitoring unterscheiden können.

Überall dort, wo durch geschicktes Optimieren der Behandlung und geschicktes Auswählen der Studienteilnehmer die Ergebnisse einer Studie beeinflussbar erscheinen, kommt die Frage auf, ob nicht ein „hochwissenschaftliches“, aber praxisfermes Studiendesign

Probleme damit hat, die Wirklichkeit abzubilden. Traditionell verwenden RCTs (randomisiert kontrollierte Studien) meist wissenschaftlich fokussierte, zum Teil sogar artifizielle Behandlungsprotokolle und liefern so zwar eine korrekte Antwort unter den jeweiligen Studienbedingungen, sie beantworten aber nicht die eigentlich relevante Frage für die Behandlungsentscheidung.

So geht es beim Telemonitoring weniger darum, ob bei sorgfältig ausgewählten Patienten mit Herzinsuffizienz eine sehr detaillierte und aufwendige Überwachung hilfreich ist, sondern ob ein realistisch umsetzbares Telemonitoring für die große Mehrzahl der Patienten im Behandlungsalltag vorteilhaft ist.

In der Vergangenheit wurde dieses Spannungsfeld oft als Gegensatz zwischen richtigen Antworten auf weniger relevante Fragen und weniger richtigen Antworten auf relevante Fragen überhöht und missverstanden.

Auch wurde dies als Spannungsfeld zwischen RCTs und nicht-randomisierten Studien beschrieben [2]. Hierbei wurde RCTs eine hohe interne Validität (d. h. richtige Antworten), aber eine niedrige externe Validität (d. h. praktisch irrelevante Antworten) zugeschrieben, während für nicht-randomisierte Studien genau das Gegenteil gelten sollte.

Hierbei wurde ignoriert, dass es weniger ein Unterschied von Studienmethoden (randomisierte Zuteilung oder nicht) an sich ist, sondern vielmehr ein Unterschied in der Planung und Durchführung der Studien.

Schon in den 1960er Jahren wurde erkannt [3], dass RCTs entweder eher auf die informierte Entscheidung fokussieren können und entsprechend in einem alltäglichen Setting durchgeführt werden und realistische Behandlungsalternativen vergleichen (sogenannte pragmatische Studien), oder eher unter strikt kontrollierten Rahmenbedingungen, einem Laborexperiment vergleichbar, durchgeführt werden,

um bestmögliche Bedingungen zur Untersuchung der vor allem medikamentösen Therapeutika zu schaffen (sogenannte explanatorische Studien).

Strikte Rahmenbedingungen, wie z. B. der Ausschluss von komorbiden Personen oder Schwangeren, sollen dabei helfen, die Variabilität in den Daten zu reduzieren, sodass der interessierende Effekt klarer erkennbar ist. Auch sollen vulnerable Patientengruppen entsprechend den ethisch-rechtlichen Vorgaben besonders geschützt werden.

Bei pragmatischen RCTs werden Placebo-Kontrollen und artifizielle Zusatzuntersuchungen vermieden, und Teilnehmer werden nicht ausgeschlossen, wenn es klinisch nicht sinnvoll ist. Die Sicht auf RCTs ist dennoch vielfach geprägt von den explanatorischen Studien, wie sie meist für die Arzneimittelzulassung verwendet werden.

Auch der Begriff der "Real-World"-Evidenz als „Label“ für alltagsnahe Forschungsergebnisse ist weitverbreitet, wobei oft vergessen wird, dass dies sowohl randomisierte wie nicht-randomisierte Studien umfasst, so sie denn auf alltäglichen Bedingungen und Daten aus der Behandlungsroutine beruhen.

Problematisch ist, dass hierdurch suggeriert wird, RCTs würden eine „unreal world“, also eine artifizielle Situation, abbilden. Auch wird auf diese Weise nahegelegt, dass interne Validität (als Kerneigenschaft von RCTs) und externe Validität (als mutmaßliche Kerneigenschaft von Beobachtungsstudien) den gleichen Stellenwert besäßen.

In Wahrheit aber ist ein richtig ermittelter Therapieeffekt für eine bestimmte Patientengruppe viel wichtiger für eine Entscheidung als ein unklarer Therapieeffekt, der möglicherweise auch bei anderen Patientengruppen oder unter anderen Rahmenbedingungen erzielt werden kann.

Wenn RCTs sehr alltagsnah konzipiert und durchgeführt werden, wird die externe Validität manch →



→ einer pragmatischen RCT über jeden Zweifel erhaben sein und die Frage der Übertragbarkeit stellt sich bei dieser randomisierten „Real-World“-Evidenz nicht.

Dass nicht-randomisierte „Real-World“-Evidenz sich so großer Beliebtheit erfreut, liegt zum Großteil auch daran, dass Beobachtungsdaten heute breit verfügbar sind.

Wo früher manch ein/e Doktorand/in mühselig das Krankenblattarchiv eines Krankenhauses nach Daten durchsucht hat, können heute mit einem Maus-Klick alle relevanten Patientendaten zu einer Forschungsfrage herausgefiltert werden.

Auf Bundesebene existieren sogenannte Routine- und Abrechnungsdaten und die Digitalisierung der Medizin wird weitere Datenquellen verfügbar machen. Auch wenn die schnelle Verfügbarkeit von Daten eigentlich für jede Form einer Studie (auch die RCT) hilfreich ist, profitieren hiervon derzeit vor allem retrospektive Kohortenstudien [4]. Gerade große Datenmengen verführen zu schnellen Analysen.

Das Beispiel Telemonitoring bei Herzinsuffizienz zeigt, welche Risiken ein unkritischer Umgang mit nicht-randomisierter „Real-World“-Evidenz birgt: In einer systematischen Analyse wurden kürzlich die Ergebnisse von RCTs und großen „Real-World“-Beobachtungsstudien zum Telemonitoring verglichen [5].

Neben 17 RCTs – darunter auch solche aus Deutschland – fanden sich 6 Beobachtungsstudien mit der geforderten Mindestgröße von 100 Personen. Die Studien untersuchten vergleichbare Telemonitoring-Konzepte, und auch die Nachbeobachtungszeit war vergleichbar. Während die RCTs in Summe 10.130 Teilnehmer rekrutiert hatten, umfassten die nicht-randomisierten „Real-World“-Analysen insgesamt 343.520 Personen – also 30-fach mehr Daten.

Die Metaanalyse der RCTs zeigte allenfalls eine Tendenz zu einer geringeren Gesamtsterblichkeit (Odds Ratio 0,92; 95%-Konfidenzintervall 0,81 bis 1,05;  $P = 0,73$ ).

Die Beobachtungsstudien dagegen berichteten deutlich größere Interventionseffekte: Die Effektschätzer in diesen 6 Analysen lagen zwischen 0,19 und 0,67.

Diese drastische Diskrepanz zwischen randomisierter und nicht-randomisierter Evidenz ist umso bemerkenswerter, wenn man berücksichtigt, dass alle 6 Beobachtungsstudien eigentlich gute Methoden angewendet hatten, um Bias durch Störvariablen (Confounding) statistisch zu reduzieren.

Die Analyse zeigt, dass Interventionseffekte in Beobachtungsstudien stark verzerrt und irreführend sein können. Vorausgesetzt wird bei diesem logischen Schluss, dass die RCTs sich auf etwa dieselben Patienten, Behandlungsumstände und Interventionen beziehen (was hier der Fall ist).

In einer „Real-World“-Situation ist es aus ärztlicher Perspektive gut nachvollziehbar, dass sich Anwendbarkeit und Nutzen von Telemonitoring sehr stark zwischen einzelnen Patienten unterscheiden. Gerade wenn man nicht nur den Patienten und die Krankheitsschwere, sondern auch sein häusliches und soziales Umfeld betrachtet, lassen sich Dutzende von Faktoren erkennen, die für ein effektives Telemonitoring im Einzelfall hinderlich oder förderlich sein können. In den nicht-randomisierten Studien führt dies zu starken Selektionsmechanismen, was ein valides Erfassen von den tatsächlichen Behandlungseffekten deutlich erschwert.

Ärztliche Sorgfalt, Erfahrung und vielleicht auch Intuition sind also für die reale Medizin unverzichtbar. Genau dieses zentrale Element ärztlicher Heilkunst, die richtige Indikationsstellung, entzieht sich aber oft einer Messbarkeit, sodass dieser Einfluss-

faktor aus einer nicht-randomisierten Studie auch mit bester statistischer Methodik nicht „herausgerechnet“ werden kann, allein dadurch, weil es keine Daten gibt, die dies ermöglichen.

Viel von dem, was als ärztliche Erfahrung die Behandlungsentscheidung beeinflusst, bleibt daher als sogenanntes „residuelles Confounding“ (überbleibende Störeinflüsse) übrig und kann die Ergebnisse nicht-randomisierter Studien stark verfälschen.

Der tatsächliche Einfluss mancher Details ärztlichen Handelns bleibt somit in Studien letztlich oft unklar. Je näher die Studien jedoch, unabhängig vom Design, an der Behandlungsrealität bleiben, desto eher lassen sich ihre Ergebnisse auch auf die Praxis übertragen.

Wenn man die Forderung nach „Real-World“-Evidenz in diesem Sinne versteht und RCTs in den Alltag integriert, kommt man der Idealvorstellung

einer Verbindung zwischen Forschung und Praxis schon sehr nahe.

Abschließend sei noch erklärt, warum im Jahr 2021 das Telemonitoring bei Herzinsuffizienz (Stadium II oder III) trotz des nicht signifikanten Gesamtmortalität-Effekts eingeführt wird: Die Metaanalysen zum engmaschigen, ärztlich überwachten Telemonitoring zeigten zumindest für die kardiovaskuläre Mortalität einen Vorteil [6]. Und immerhin 2 der 4 RCTs kamen aus Deutschland, sodass es für die „Real-World“-Anwendung nun direkte Vorbilder gibt. ■

**Prof. Dr. med. Stefan Sauerland**  
 Institut für Qualität und Wirtschaftlichkeit  
 im Gesundheitswesen (IQWiG)  
 Im Mediapark 8, D-50670 Köln;  
 Tel.: 0221 / 356 85 – 359  
 stefan.sauerland@iqwig.de

#### Literatur:

1. Gemeinsamer Bundesausschuss, 2020: G-BA führt neues datengestütztes Management für Patientinnen und Patienten mit Herzschwäche in die Versorgung ein (Pressemitteilung). <https://www.g-ba.de/presse/pressemitteilungen-meldungen/922/>
2. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM: Real-world evidence - What is it and what can it tell us? *N Engl J Med* 2016; 375: 2293-2297. <https://doi.org/10.1056/NEJMs1609216>
3. Schwartz D, Lellouch J: Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967; 20: 637-48 (erneuter Abdruck in: *J Clin Epidemiol* 2009; 62: 499-505) <https://doi.org/10.1016/j.jclinepi.2009.01.012>
4. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016; 188(8): E158-E164. <https://doi.org/10.1503/cmaj.150653>
5. Schell LK, Schulz A, Angelescu K, Glinz D, Knellen M, Sauerland S: Misleading results even after adjustment for a dozen covariates: a comparison between 'real world' analyses and randomized controlled trials (RCTs) in telecardiology. Abstracts of the 25th Cochrane Colloquium, Edinburgh, UK. *Cochrane Database of Systematic Reviews* 2018; (9 Suppl 1). <https://doi.org/10.1002/14651858.CD201801>.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitssystem (IQWiG), 2019: Datengestütztes, zeitnahes Management in Zusammenarbeit mit einem ärztlichen telemedizinischen Zentrum für Patientinnen und Patienten mit einer fortgeschrittenen Herzinsuffizienz (Rapid Report N19-01). <https://www.iqwig.de/de/projekte-ergebnisse/projekte/nichtmedikamentoese-verfahren/n-projekte/n19-01-datengestuetztes-zeitnahes-management-in-zusammenarbeit-mit-einem-aerztlichen-telemedizinischen-zentrum-fuer-patientinnen-und-patienten-mit-einer-fortgeschrittenen-herzinsuffizienz-rapid-report.11591.html>

