

Blick ins Auditorium des 8. MVF-Fachkongresses

Highlights des 8. Fachkongresses „Extrapolation“ von „Monitor Versorgungsforschung“ in Berlin

Der „Risk of Bias“ muss das Leitprinzip werden

Wundern Sie sich als Arzt manchmal, warum die Ein- und Ausschlusskriterien von RTC-Studien ausgerechnet viele ihrer Patienten negieren? Denken Sie als Gesundheitspolitiker oft, dass es für ihre Entscheidungen mehr an Evidenz geben muss, als man in rein klinischen Studiensettings herausfinden kann? Fragen Sie sich als Pharmamanager, welchen Wert Register, Real World-Daten und Patient Reported Outcome-Studien in der frühen Nutzenbewertung wirklich haben? Hinter all diesen Fragen steht das Metathema der Übertragbarkeit von Studienergebnissen, die im Fokus des 8. MVF-Fachkongresses mit dem Titel „Extrapolation 2018“ stand.

>> Zum achten Mal veranstaltete „Monitor Versorgungsforschung“ seinen jährlichen wissenschaftlichen Fachkongress, der sich der wichtigen Themen der Versorgungsforschung und des Versorgungsmanagements widmet und diese aus den Blickwinkeln der unterschiedlichen Stakeholder beleuchtet. In diesem Jahr stand das Thema „Extrapolation“ im Mittelpunkt. Dies geschah auf Anregung von Prof. Dr. Jürgen Windeler, dem Leiter des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), der im Titelinterview mit „Monitor Versorgungsforschung“ (MVF 01/18) erklärte: „RCT sind nicht die Grundlage von EBM, sondern sie sind für die Frage nach dem Nutzen die fehlerärmsten Instrumente – nicht mehr und nicht weniger.“

Windeler verwandte im MVF-Interview ausdrücklich den Begriff „fehlerärmste“, weil er genau weiß, dass jedes RCT, egal welches Design man wählt, seine Limitationen hat. So können pragmatische Studien, die mit möglichst wenig Einschlusskriterien arbeiten, um ein möglichst breites Anwendungsspektrum abdecken zu können, sinnvoll sein, jedoch laut Windeler „sehr problematisch werden, wenn sich z. B. Ärzte und Patienten nicht an die ursprüngliche Randomisierung halten und alles fröhlich durcheinander geht.“ Dann gehe am Ende die Aussagekraft gegen Null. Gleiches kann man aber auch bei klassischen RCT diskutieren, weil nach Windelers Ansicht eine scharfe Trennung zwischen „Ideal“ und „Alltag“ nicht existiere, genauso wenig wie der oft betonte Unterschied zwischen Efficacy und Effectiveness, den er einfach für Quatsch hält. Auch deshalb, weil keiner genau wüsste, ob eine Studie, die einen „Alltag“ abzubilden versuche, nicht viele andere „Alltage“ ignoriert. Er möchte darum „gerne wissen, was in der Versorgung eigentlich genau passiert.“ Das wiederum ist eine Frage der Extrapolation, der sich der MVF-Fachkongress 2018 widmete. Nachfolgend lesen Sie die reaktionelle Zusammenfassung des ersten Teil des Fachkongresses, der sich mit dem Status Quo der Übertragbarkeit von Studienergebnissen befasste.

Dr. Ilona Köster-Steinebach: verstärkte Ausrichtung der Forschung an Patientenpräferenzen



„Wie gestalten wir ein Gesundheitssystem evidenzbasiert?“ So lautete die Eingangsfrage von Dr. Ilona Köster-Steinebach, acht Jahre lang für den Verbraucherzentrale-Bundesverband Patientenvertreterin im G-BA, nun designierte hauptamtliche Geschäftsführerin des Aktionsbündinis Patientensicherheit (APS) und weil noch nicht im Amt bestätigt, in einer Transistionsphase, weshalb sie in ihrem Startvortrag des MVF-Fachkongresses „Extrapolation 2018“ ihre private Meinung vertrat. Aber auch deshalb, weil „solche komplizierten Evidenzfragen“ in der Patientenvertretung noch nicht so richtig hundertprozentig in allen Facetten ausdiskutiert seien.

Um sich ihrer Kernfrage – „Was meint Evidenz für Patienten?“ – zu nähern, bemühte sie eingangs eine Definition, die aus der Qualitätssicherung stammt. Diese besagt, dass Qualität das Ausmaß sei, in dem Gesundheitsleistungen für Individuen oder Populationen die Wahrscheinlichkeit erwünschter gesundheitlicher Behandlungsergebnisse erhöhen und mit dem gegenwärtigen professionellen Wissensstand übereinstimmen. Positiv sei an dieser Definition, dass hier tatsächlich die Präferenzen der Patienten und auch der Begriff Wunsch explizit genannt seien. Köster-Steinebach: „Aber irgendwie scheinen die Präferenzen in dieser Definition als bekannt vorausgesetzt.“ Wobei sich ihr die Frage stelle: Sind sie das denn eigentlich? Und: Ist der Patientenwunsch nur als Ergebnis von Behandlung zu realisieren? Oder gibt es auch andere Wege diesem Patientenwunsch nachzukommen? Gleich danach schließen sich die Kernfragen an: Welche Verbindung hat der professionelle Wissensstand mit den Patienten-



wünschen?“ Ist der professionelle Wissensstand der Standard, an dem sich die Präferenzen der Patienten zu orientieren haben? Oder geben die Patientenpräferenzen vor, welcher Wissensstandard durch Forschung zu erarbeiten wäre?

Aus ihrer Sicht nehme die Forschung viel stärker die Frage der Wahrscheinlichkeit – welche Variante, welches Behandlungsergebnis tritt wahrscheinlich ein – in den Fokus, als die der eigentlichen Präferenzen der Patienten und der Frage, ob ein erzieltes Behandlungsergebnis denn überhaupt mit dem übereinstimmt, was sich Patienten von ihrer Behandlung wünschen. Ihr eindeutiges Playdoyer: „Eine verstärkte Ausrichtung der Forschung an Patientenpräferenzen und ein stärkerer Einbezug von Patientenvertretern, wobei man auch über neue politische Rahmenvorgaben nachdenken sollte. Und nicht zuletzt auch eine Methodendiskussion in der Wissenschaft, die eine bessere Übertragung der Prinzipien der EBM ermöglichen sollte, was aber nicht heiße, „dass alles RCT sein muss“, wohingegen das grundsätzliche Misstrauen bezüglich der Frage bestünde, ob und wann ein Bias vorliegt, weshalb sie für die „Rückkehr zur Methode rigoroser Falsifikation“ statt der beim G-BA zu beobachteten Beliebigkeit eintritt.

Dr. Antje Behring: Evidenztransfer statt Extrapolation



Dazu kam gleich ein vernehmliches Veto von Dr. Antje Behring, Referentin in der Abteilung Arzneimittel des G-BA, die nicht glaubt, dass ihr Arbeitgeber, der Gemeinsame Bundesausschuss „so sehr beliebig“ arbeite, wobei sie aber nur einen Einblick in den Part Arzneimittel, die Bereiche rund um den § 35a SGB V – die frühe Nutzenbewertung und die dazugehörigen Beratungen – habe. Sie sprach in ihrem Vortrag an, welche Evidenz der G-BA tatsächlich nutzt, um seine Entscheidungen zu treffen, wozu auch Patientenpräferenzdaten zählen würden. Und sich hier immer in dem Konflikt befinde, wie die Aussagen von Patienten so gut wie möglich objektivierbar seien. Das beginne bei der Validierung der verwandten Instrumente, vor allem Fragebögen und deren generellen Nutzbarkeit für eine entsprechende Arzneimitteltherapie oder Indikation.

Doch insgesamt hätte der G-BA bei Patientenpräferenzen recht wenig Erfahrungen, gleichwohl sich der Bundesausschuss bei der PREFERE-Studie engagiert hätte, um Patientenpräferenzen zu schärfen. Doch seien „bis jetzt noch keine guten Erfahrungen damit gemacht worden“, auch könnte man im Bereich der Arzneimitteltherapien bei Patientenpräferenzen „noch nicht so richtig gute, positive Beispiele bringen“.

Daher bleibe im Bereich Arzneimittel der Goldstandard die RCT, nicht nur bei der Erstbewertung, sondern auch für befristete Auflagen würden weiterhin randomisierte kontrollierte Studien gefordert. Abweichungen von RCT seien hingegen nur in Ausnahmefällen wie bei dramatischen Effekten möglich, bei denen auch einarmige Studien und Registerdaten „in den ganz seltensten Fällen akzeptabel“ seien. Hier beginnt die Methodik: „Solange es nur einarmige Studien gibt, sind nur nicht quantifizierbare Aussagen möglich“, erklärte Behring, und spricht damit einen Konflikt an, der durchaus

Kongress-Programm „Extrapolation 2018“

von	bis	Thema	
		Status Quo: Übertragbarkeit von Studienergebnissen	
09:30	10:00	Check-in	
10:00	10:10	Begrüßung	Prof. Dr. Roski, MVF Christian Koch, Vilua
10:10	10:40	Evidenz für Patienten	Dr. Köster-Steinebach
10:40	11:10	Stellenwert von RCT, Registerdaten und weiteren Datenquellen für Entscheidungen des G-BA	Dr. Behring, G-BA
11:10	11:40	Evidenz aus Studien – Kriterien und Standards	Prof. Dr. Antes, Cochrane, Freiburg
11:40	12:10	Studien aus dem Ausland als Entscheidungsgrundlage in Deutschland	Prof. Dr. Wensing, Universität Heidelberg
12:10	12:40	Was Register leisten können und was nicht	Prof. Dr. Prof. h.c. Dr. h.c. Neugebauer, Neuruppin, (DNVF)
		Perspektiven der Stakeholder	
13:40	14:10	Evidenz für die Primärversorgung	Corinna Schaefer, ÄZQ
14:10	14:40	RCT vs. Real World? Klinische Studien und onkologischer Alltag	Prof. Dr. Ludwig, AKdÄ
14:40	15:10	Welche Evidenz für welche Entscheidungen? – Die Sicht der Kassen	Dr. Haas, GKV-Spitzenverband, Berlin
15:10	15:40	Real World Evidenz — Was ist das und was kann sie uns sagen?	Prof. Dr. Paar, Sanofi, Berlin
		Methoden	
16:00	16:30	Frühe Nutzenbewertung: Kombination RCT mit Beobachtungsdaten	Dr. Happich, Lilly, Bad Homburg
16:30	17:00	Trials und Real World Data: Unterschiede und Kombination	Prof. Dr. Wilke, Wismar
17:00	17:30	Kausalität, Zusammenhang, Koinzidenz. Alles kann uns weiterbringen	Prof. Dr. Häussler, IGES Institut, Berlin
17:30	18:00	Standpunkt: Gute Forschung statt Pseudo-Kontroverse	Prof. Dr. Schrappe, Köln
18:00	18:30	Podiumsdiskussion: Wie muss richtige Extrapolation vorgehen?	Prof. Dr. Ludwig, AKdÄ
18:30	18:40	Verabschiedung	Prof. Dr. Roski, MVF





MVF-Herausgeber Prof. Dr. Reinhold Roski (li.) und Christian Koch von vilua begrüßten die Teilnehmer des 8. Fachkongresses von „Monitor Versorgungsforschung" zum fünften Mal in der Bertelsmann Repräsentanz in Berlin.

recht schwierig zu verstehen ist: Denn der G-BA muss bei Orphan Drugs diese Zulassungsstudien, aber auch andere Zulassungsarten wie zum das „Conditional Approval" oder „Exceptional Circumstances" akzeptieren, und damit auch die mit ihnen einhergehenden Ausnahmeregelungen und Unsicherheiten.

Dahinter steht die Divergenz, warum einerseits die Zulassung in ihrer Ja/Nein-Entscheidung beispielsweise eine gute Qualität bescheinigt, und auch die Sicherheit und Wirksamkeit hinreichend belegt, doch aufgrund der Tatsache, dass es sich um einarmige Studien handelt, der G-BA möglicherweise „keinen Zusatznutzen" bescheinigt. Behrings Antwort: Das sei alleine darin begründet, dass der Bundesausschuss einen Vergleich anstellen müsse, der vergleichende Daten bedingt, um quantifizieren zu können, wie hoch der Zusatznutzen ist. Nun sei es aber so, dass im Falle von „Conditional Marketing Authorizations" und bei „Exceptional Circumstances Authorizations" erst nach der Zulassung Evidenz generiert werden müsse. Behrings Frage: „Welche Evidenz ist wie zu generieren? Sind das tatsächlich Registerstudien? Oder ist es einfach nur die Fortführung von bereits begonnenen RCT?" Dies sei die Erklärung für viele Befristungen, die der G-BA in seine Entscheidungen einbaue: „Vorläufige Daten zum Zeitpunkt der Nutzenbewertung ergeben eigentlich auch nur eine vorläufige Bewertung."

Doch wie steht es um die wirkliche Güte der in RCT gewonnenen Daten und deren Übertragbarkeit? „Extrapolationen von Daten haben wir bis jetzt noch nicht so sehr häufig gemacht", gibt Behring offen zu, wobei sie hier PUMA-Arzneimittel ausnimmt. Und auch sei der Evidenztransfer so unüblich nicht, was sie am Beispiel von Ledipasvir und Sofosbuvir im Bereich der Hepatitis C zeigt. Hier habe man sich ganz genau angeschaut, welche Evidenz für Erwachsene und welche Evidenz für Kinder vorgelegen hätte, um die Frage zu beantworten: „Kann ich den beträchtlichen Anhaltspunkt bei Erwachsenen auf Kinder transferieren oder nicht?" Anders liegt die Sache, wenn bei den betrachteten Studien Patientengruppen fehlen. Dann würden Ärzte, die im G-BA tätig seien, sagen: „Finde ich meinen Patienten eigentlich, den ich behandeln möchte, auch in dieser Studie wieder?" Und: „Wie geht man damit um?"

Prof. Dr. Gerd Antes: Risk of Bias als Leitprinzip



Das kann niemand besser als Prof. Dr. Gerd Antes, der wissenschaftliche Leiter und stellvertretende Direktor von Cochrane Deutschland, beantworten, der auf die scheinbare so heile Welt der RCT („das ist sie nicht") und auf „Real World XXL", die Big Data, einging. Er stellte zu Beginn seines Vortrags den Patienten, von ihm Herr Neumann genannt, vor, der nach Shared Decision Making der Therapieemfehlung seines Arztes folgt, aber sich letzten Endes selbst fragt: „War das eigentlich richtig?"

Es gebe nur einen Weg, die Alternative festzustellen, indem der Patient „zeitgleich das Gegenteil oder eben das Nichtstun wählt", was aus zwei Gründen rein fiktiv wäre: Erstens gibt es die Möglichkeit außer in der Dermatologie nicht, bei der man zum Beispiel nur den linken Arm mit einer Salbe bestreicht. Und zweitens hilft die so gewonnene Erkenntnis dem Patienten selbst nicht, eventuell aber der nächsten Generation. „Alles, worüber wir hier reden, ist das Ersetzen der beiden Hälften von Neumann durch Gruppen", verdeutlicht Antes sein pragmatisches Gedankenspiel: „Alles, was wir über die letzten 100 Jahre gelernt haben, ist, diese Gruppen so zu optimieren, dass tatsächlich ein Ergebnis mit minimalem Bias rauskommt."

Wichtig sei hier ein „maximaler Schutz gegen systematische Fehler", neudeutsch Bias, wozu Co-Faktoren ausbalanciert, randomisiert und verblindet wird oder eben auch nicht. Das Schlagwort auf dem Weg zur Wahrheit lautet für ihn: „Counterfactual Thinking", was nichts anderes heißt als: „Was passiert eigentlich, wenn ich das Gegenteil machen würde?" Das sei ein weit über die Medizin hinausgehender philosophischer Ansatz. Dabei gebe es eigentlich nur






AOK
Die Gesundheitskasse.

Mit Herz und Verstand - die AOK Nordost setzt auf innovative Betreuungsprogramme

Als regionale Versorgerkasse für die Regionen Berlin, Brandenburg und Mecklenburg-Vorpommern machen wir uns für Projekte stark, die unseren Versicherten im Alltag mit ihrer Erkrankung konkret helfen und die Betreuung durch den behandelnden Arzt sinnvoll ergänzen.

AOK-Curaplan Herz Plus

Mittlerweile nutzen rund 7.000 Versicherte der AOK in Berlin und Brandenburg das Programm AOK-Curaplan Herz Plus. Zeichen einer drohenden kritischen Verschlechterung einer Herzinsuffizienz rechtzeitig zu erkennen und sofort intervenieren zu können ist das Ziel dieses Versorgungsprogramms. So werden die Patienten noch in der regulären ambulanten Versorgung wieder stabilisiert und Krankenhausbehandlungen vermieden. Eine wissenschaftliche Studie des Instituts für Community Medicine der Universitätsmedizin Greifswald belegt die positiven Auswirkungen sowohl auf die Überlebenswahrscheinlichkeit als auch die gesundheitsökonomischen Effekte der Teilnehmer. Das Konzept besteht dabei aus zwei wichtigen Säulen: der individuellen telefonischen Betreuung der Patienten sowie der telemedizinischen Überwachung von Zeichen und Beschwerden der Herzinsuffizienz. Das Programm wird seit 2006 mit der Gesellschaft für Patientenhilfe und dem Unfallkrankenhaus Berlin durchgeführt.

HerzEffekt MV

In Mecklenburg-Vorpommern leistet HerzEffekt MV einen weiteren Beitrag zur Herzgesundheit. Ziel des Programms ist der Aufbau eines digitalen Versorgungszentrums – dem sogenannten Care-Center. Herzkranken Patienten wird damit ein wohnortnaher Zugang zu spezialisierter medizinischer Versorgung angeboten. Per zufallsgesteuertem Auswahlverfahren werden die Teilnehmer einer Studiengruppe der Interventionsgruppe (Gruppe mit medizinischen Messgeräten) oder der Kontrollgruppe (Gruppe ohne medizinische Messgeräten) zugeordnet. Alle Teilnehmer am Programm erhalten medizinische Untersuchungen. Teilnehmer der Interventionsgruppe erhalten je Krankheitsbild telemedizinische Geräte und übermitteln täglich gemessene Parameter an das Care-Center. Medizinische Fallmanager können bei Veränderungen in den Messergebnissen, die auf eine Verschlechterung des Gesundheitszustandes hindeuten, schnell reagieren. Das durch den Gemeinsamen Bundesausschuss (G-BA) geförderte Innovationsfondsprojekt HerzEffekt MV wird mit der Universitätsmedizin Rostock Versorgungsstrukturen GmbH als Konsortialführer, sowie den Konsortialpartnern AOK Nordost, Techniker Krankenkasse und Philips GmbH Market DACH durchgeführt.

Gesundheit in besten Händen

www.aok.de/nordost

zwei Fehlerarten zu berücksichtigen, davon aber jede Menge: systematische und zufällige.

Systematische Fehler versucht man in Studien klein zu halten, weg bekomme man sie nicht, sagt Antes, und „wenn man Pech hat, erwischt einen auch noch der Zufall". Antes: „Diese beiden stehen im engen Zusammenhang, sie zu beherrschen ist die große Kunst." Daher lautet seine Botschaft: „Der Goldstandard ist nicht der RCT", sondern: „Was ist für Neumann am besten?" Dies sei zum großen Teil das RCT, wobei man genauso gut auf der ganzen Welt Informationen von Patienten in sehr ähnlichen Situationen einsammeln und die dem einzelnen Patienten als Entscheidungsgrundlage zur Verfügung stellen könnte. Dem aber steht bis auf wenige Regionen wie Skandinavien der Datenschutz entgegen, warum dann eben doch der effizienteste Weg Studien seien. Antes: „Wenn wir jetzt noch Register dazunehmen und diese Beobachtungsstudien nennen, stimmt das Bild sogar."

Das Kernproblem aber sei ungelöst. Obwohl unglaublich viele methodische Studien gemacht würden, müsse man verstehen lernen, die wichtigen Schräubchen zu finden, an denen gedreht werden muss, um den Output zu verbessern. „Was wir sehr gut wissen, ist leider, dass 50 Prozent von dem, was oben angefangen wird, beim Patienten nicht ankommt", bedauert Antes. Das sei fast so etwas wie eine Naturkonstante.

Doch an der schieren Menge an Studien liege es gewiss nicht, meint Antes, der Medline bemüht, um einen extrem hohen Zuwachs von rund 30.000 Studien pro Jahr zu beweisen. Wobei noch einmal 50 Prozent gar nicht veröffentlicht würden, was rund 60.000 neue Studien pro Jahr ausmache: „Wenn man über Big Data spricht, kann man hier schon anfangen."

Die nicht veröffentlichten Studien seien eine gigantische Verschwendung, ein Betrug an den Studienteilnehmern, weil die damit generierten Informationen nicht der nächsten Generation zur Verfügung stünden. Zudem werde als Grundlage der drei Hauptachsen – HTA, klinische Leitlinie und Patienteninformation – ein nicht reparabler, grauenhafter Bias eingeführt. „Alle Leitliniengruppen, G-BA, IQWiG oder IQTiG oder Patienteninformationen hängen auf dieser falschen Basis", sagt Antes, daran sei auch kurzfristig trotz aller Bemühungen nichts zu machen. Aber man könne das Problem angehen, zum Beispiel mit einem Verfahren, das die Kanadier vor Jahren einmal „Knowledge Refinery" genannt hätten und die einem Fünferschema folge:

1. Formulieren der Fragestellung
2. Systematische Suche in der Literatur
3. Qualitätsbewertung der Funde
4. Zusammenfassung der Evidenz
5. Interpretation der Ergebnisse

Wer dies mache, bekomme einen sogenannten „Forest Plot" oder auch „kumulative Forest Plots", in denen sich die Evidenz verdichtet. Doch: Bis wohin ist die Frage, weil es nach der statistischen Logik nun einmal irgendwo aufhören muss, spätestens, wenn das Konfidenzintervall der gesammelten Studien – die Summe des kumulierten Wissens – die 1 nicht mehr trifft, wäre Schluss. Antes: „Das ist aber mathematisch falsch", weil auch Faktoren wie „Repeated Testing" und „Alpha-Korrektur" zu berücksichtigen seien. Dennoch gebe es bis heute „keine akzeptierte logische, mathematisch-statistische Stoppregel". Doch selbst wenn es die gäbe, ist sich Antes sicher, wären die deutschen Ethikkommissionen – und auch die vieler anderer Länder – nicht in der Lage, diese Regel umzusetzen.

Dies alles sind rein RCT getriggerte Probleme. Doch nun kämen auch noch Real World-Beobachtungsstudien dazu, die in einer rein deutschen Fehlentwicklung sofort in eine Frontenstellung münden würde. Dabei sei Versorgungsforschung versus RCT kein Widerspruch, denn pro Fragestellung bräuchte man nun einmal die jeweils beste Methode. Antes: „Der Stein des Anstoßes ist die Evidenzhierarchie von randomisierten Studien über Kohortenfallkontrolle bis zu nicht transparenten Experten." Und weiter bis hin zu Big Data, Digitalisierung, personalisierte Medizin, Translation, Innovation und künstlicher Intelligenz, nach Antes „ein Mischmasch von wirklich starker Begriffsverwirrung", beschrieben in den beiden Büchern „The End of Theory" und „Big Data". Diese würden eine schiere Datensintflut beschreiben und im Endeffekt postulieren, dass damit wissenschaftliche Methoden überflüssig würden. Und noch schlimmer, dass alleine durch viele Daten aus Korrelationen Kausalität entstünde, was „einfach falsch" sei. Zwar beanspruche Big Data für sich, unstrukturierte Daten – weit über Beobachtungsstudien hinaus – zu analysieren, doch fuße das Ganze auf einer wissenschaftlich falschen Basis, denn, so Antes: „Mehr Daten sind nicht zwangsläufig besser, sondern erfordern überproportional viel mehr Aufwand, um das Richtige heraus zu bekommen."

Dies alles führt ihn zu seiner ganz persönlichen „Evidenzhierarchie 2025": Oben experimentelle Studien, gefolgt von Beobachtungsstudien und ganz unten Big Data, Stand heute. Antes: „Suchen Sie bei Big Data den Qualitätsbegriff. Suchen Sie ein Konfidenzintervall bei Big Data. Suchen Sie das Ausmaß der Unsicherheit von den Aussagen: Sie werden dies alles nicht finden." Doch alleine darum, weil ein Informatiker alles ausrechnen kann, was richtig ist, wenn nur die Datenmenge stimmt. Der Grund: Bei unendlich vielen Daten werde zwangsläufig jeder Test signifikant. Das Problem: Es gibt nach Antes noch lange keine Übereinstimmung zwischen statistischer Signifikanz und klinischer Relevanz. Das müsse schon jeder lernen, der im ersten Semester Medizin studiert. Auch darum sei er zum MVF-Fachkongress gerne gekommen, um zu versuchen, Wege wieder zusammenzuführen und nicht nur darüber zu sprechen, wo es nicht zusammenpasst. Sein Fazit zum ersten: „Qualität auf allen Stufen des Wissensprozesses ist die oberste Messlatte, an der wir alles entwickeln müssen." Und zum zweiten: „Der Risk of Bias muss als Leitprinzip zur völligen Normalität werden."

Prof. Dr. Michel Wensing: vom Scale-up- zum Scale-out-Transfer



Prof. Dr. Michel Wensing, der – bevor er in der Abteilung für Allgemeinmedizin und Versorgungsforschung an der Universität Heidelberg die Professur für Implementierungswissenschaft übernahm – 25 Jahre im Universitätsklinikum Nijmegen tätig war, ging in seinem Vortrag vor allem auf die Problematik der Übertragbarkeit von Studien aus dem Ausland ein. Auf die Frage „sind Forschungsergebnisse aus dem Ausland relevant für Deutschland?" hat er drei potenzielle Antworten: „Ja, wenn die Situation ähnlich ist", „nein, nur Forschung in Deutschland ist relevant" und zum dritten: „Es hängt davon ab."

Diese breite Streuung der Antworten rührt übrigens daher, dass es bislang nur wenig Literatur gibt, die hier weiterhelfen würde. Eine davon stammt von Aarons, der 2017 den Unterschied zwischen Scale-up- und dem Scale-out-Transfer beschrieben hatte, eine weitere von Stirman, der 2013 systematisch inventarisiert, dokumentiert und über 250 Kategorien von Änderungen erstellt hat, die dazu führen können, dass eine Intervention am Ende vielleicht doch nicht so effektiv ist, wie sie in der ursprünglichen Studie beschrieben wurde. Das habe, so Wensing, mit „Program Drift" oder „Voltage Drop" zu tun.

Während „Program Drift" bedeute, dass die Einschlusskriterien in der Realität anders als im Vergleich zur ursprünglichen Studie sind, beschreibt „Voltage Drop" einen möglichen Einfluss von weniger gut ausgebildeten Ärzten oder Pflegekräften. Dies könne wiederum dazu füh-

Zitationshinweis

Stegmaier, P.: „Der Risk of Bias muss das Leitprinzip werden", in: „Monitor Versorgungsforschung" 03/18, S. 16-23, doi: 10.24945/MVF.03.18.1866-0533.2077

ren, dass die Intensität einer Intervention in der Realität nicht so hoch wie in der ursprünglichen Studie ausfällt. Wensing: „Aus diesem Grund muss man lernen, wie Kontext-Mechanismen funktionieren und welchen Einfluss sie auf Outcomes nehmen können."

Um komplexe Interventionen zu erforschen, müsse man daher immer zuerst logische Modelle aufbauen, die erklären, wie ein Interventionsprogramm eigentlich funktioniert. Dazu gehören unter anderem Modifizierungen und Einfluss nehmende Kontextfaktoren, um überhaupt die grundlegende Frage klären zu können: „Welche Auswirkungen haben diese Modifizierungen und diese Kontextfaktoren, die vielleicht für Deutschland speziell sind, auf die Effekte?"

Dazu gehört aber ebenso die „Benefit-Harm-Ratio" („Ist diese nach dem Transfer nach Deutschland immer noch positiv?), die Kosteneffektivität sowie die Kosten nach Implementierung („Ist die Kosteneffektivität nach Transfer und nach Implementierung immer noch attraktiv?") sowie der „Budget Impact" („Weil manchmal Kosten vor allem von bestimmten Leistungserbringern oder Stakeholdern getragen werden müssen"). All das seien Schwerpunkte des von ihm vertretenen Fachgebiets der Implementierungswissenschaft, die beschreibt, welche Maßnahmen und Strategien man anwenden kann, um die Implementierung zu fördern.

Prof. Dr. Neugebauer: „über den Wert von RCT"



Wer erwartet hatte, dass Prof. Dr. Prof. h.c. Dr. h.c. Edmund A.M. Neugebauer, der Dekan der Medizinischen Hochschule Brandenburg (MHB) in Neuruppin und als langjähriger Vorstand des DNVF sprechend für die AG Register des Deutschen Netzwerks Versorgungsforschung e. V., einen „brennenden Vortrag für Register und gegen RCT" halten würde, wurde enttäuscht. Neugebauer, der selbst aus der klinischen Forschung kommt und viele randomisierte kontrollierte Studien durchgeführt hat, muss eigenen Bekundungen zufolge „über den Wert von RCT" nicht weiter aufgeklärt werden, obwohl es – dies sei Inhalt seines Beitrages – Limitierungen gebe, für RCT einerseits, für Register auf der anderen Seite.

Dabei konzidiert er durchaus, dass auch an Register bestimmte Qualitätskriterien gestellt werden müssen, die jedoch zum heutigen Zeitpunkt nicht so viele Register erfüllen würden. Wobei es auch positive Ausnahmen wie das Krebs-, das Trauma- oder die StuDoQ-Register der Deutschen Gesellschaft für Allgemein- und Viszeralchirurgie gebe. Generell gelte, das wie bei RCT auch an Register bestimmte Qualitätskriterien anzulegen und zu erfüllen sind, wie bereits im Jahr 2010 im Register-Memorandum des DNVF definiert:

1. Systematik/Angemessenheit eines Registers
2. Standardisierung
3. Validität der Stichprobengewinnung
4. Validität der Datenerhebung
5. Validität der statistischen Analysen und Berichte
6. Übergreifende Qualitätsanforderungen

„All das muss erfüllt sein, um von einem guten Register sprechen zu können", erklärte Neugebauer. Darum gebe es ähnlich wie bei einer klinischen Studie, bei der ein Studienprotokoll existiere, natürlich auch ein Registerprotokoll.

Ein wichtiges Feld für Register und darauf aufbauende Registerforschung sei es, Populationen evaluieren zu können, die nicht an Studien teilnehmen oder auch nicht teilnehmen können. So könne man eine größere Population mit der Teilpopulation in einer Studie vergleichen, wenn man parallel zur Studie ein Register mit all jenen mitlaufen lasse, die eben nicht in einer Studie eingeschlossen worden sind. Ebenso gut könne man mit Registern seltene oder zeitlich verzögerte Ereignisse erfassen, was bei Studienlaufzeiten von vielleicht ein, zwei oder drei Jahren bei randomisierten Studien selbst mit entsprechenden Follow-up-Zeiten kaum möglich sei.

Der Punkt, an dem sich die Geister scheiden, ist jedoch der Einsatz von Registern in der Evaluierung von Wirksamkeit und Nutzen im Versorgungsalltag. Darum stellte Neugebauer in seinem Vortrag die verschiedenen Merkmale von klinischen RCT auf der einen, und prospektiven Registern auf der anderen Seiten gegenüber.

Bei klinischen Studien könnten aufgrund der Fallzahlen nur limitierte Probanden- oder Patientenzahlen eingeschlossen werden, die obendrein „meistens sehr, sehr homogen" ausfallen würden. Gerade in Arzneimittelstudien würden diese so homogen gefasst, dass in einer solchen Studie eine fast gleiche Gruppe von Patienten eingeschlossen sind. Doch auch die Studien durchführenden Institute würden stark selektiert, seien damit ebenso homogen mit oft engen Ein- und Ausschlusskriterien.

Bei Register hingegen sei dies geradezu umgekehrt: Beobachtet würde eine unlimitierte Zahl von heterogenen Probanden in heterogenen Settings, was meint: normale Umgebung, offene Inklusion, verschiedene Arten von Beobachtungsstiefen etc. Andererseits sei wieder bei RCT die Vergleichbarkeit viel besser als bei Registern, doch „auf jeden Fall immer mit einem gewissen Bias" verbunden. Dies obwohl der Datenumfang in der Regel in der Tiefe bei randomisierten Studien viel größer als bei Registern sei, jedenfalls sei das Stand heute noch so. Was aber auch an der oft mangelnden Unterstützung für Register liege, obwohl Aufwand und Kosten bei RCT wesentlich größer als bei Register ausfielen. Die Stärken von Registern sieht Neugebauer in der Qualitätssicherung, der Epidemiologie und der Beschreibung von Änderungen in der Versorgung.

Bleibt immer noch die Frage nach dem Nutzennachweis. Bei Registern sei das auch heute noch die große Frage, bei RCT hingegen nehme man an, dass sie hier ihre große Stärke haben. Denn bei RCT gebe es durchaus Probleme mit vielen Verzerrungen. Dies beginne bei der eng gefassten, homogenen Studienpopulation mit geringer Co-Morbidität, minimaler Co-Medikation, engem Altersfenster und vielen anderen damit verbundenen Ausschlusskriterien. Dazu gehöre jedoch auch die Bereitschaft des Patienten, an Studien teilzunehmen, was eine weitere Patientenselektion mit sich bringe. Aber auch ein Versorgungs-Bias durch eine standardisierte, qualitätsgeprüfte Behandlung durch Experten sei zu beachten. Nicht zu vergessen sei schlussendlich auch der Beobachtungseffekt, der möglicherweise zu unrealistischen Ergebnissen führt.

Nur: Wie sieht die externe Validität von klinischen Studien aus? Neugebauers Antwort: „Bei klinischen Studien kann man nur Schlussfolgerungen aus den Studienpatienten für die Zielpopulation ziehen." Statistische Schlussfolgerungen seien nur auf die, in den Studien eingeschlossenen Patienten und damit nur auf die Studienpopulation möglich. Und ganz kategorisch: „Wenn die Population nicht die Zielpopulation ist, ist ein Schluss von der Studien- auf die Zielpopulation nicht ohne Weiteres möglich." Man könne jedoch schon rein praktisch nicht so viele Studien machen, damit alle möglichen Differenziertheiten von Patienten tatsächlich abgebildet würden. Doch gehe es bei der externen Validität weniger darum, ob die Patienten in Studien anders als die in der späteren Praxis seien, sondern darum, ob die Therapieeffekte andere sind, und wie deren Effekt modifiziert wird. Neugebauer: „Das Ziel der Versorgungsforschung sollte es sein, diese effektmodifizierenden Parameter zu identifizieren, zu analysieren und damit die Einflussfaktoren auf die externe Validität besser erklären zu können." <<

von: MVF-Chefredakteur Peter Stegmaier



„Die Situation, mit der wir uns bei der Leitlinienerstellung meistens herumschlagen, ist, dass die Evidenz, die wir haben, nicht gut genug ist, vor allem aber, dass wir keine haben, dass wir zu wenig oder oft gar nichts wissen. Das heißt, dass wir einmal in die Situation kämen, uns auszusuchen, nehmen wir jetzt Registerdaten oder lieber den guten RCT, haben wir ganz selten. Zweiter Punkt, den ich gerne ansprechen möchte: Wir haben nicht zu wenig RCT, wir haben die falschen. "

Corinna Schaefer, ÄZQ

„Es kommen viele onkologische Wirkstoffe auf den Markt, ohne dass wir die Evidenz haben, die wir eigentlich benötigen. Das ist die Ist-Situation, mit der wir umgehen und uns fragen müssen: Welche Evidenz generieren wir danach und mit welchen Studien? Die Frage lautet nicht: randomisierte kontrollierte Studien oder Real World-Evidenz? Es geht vielmehr darum, in welcher Situation wir welche Art von klinischen Studien beziehungsweise auch Registerdaten brauchen."

Prof. Dr. Wolf-Dieter Ludwig, AkdÄ





„Der Begriff Real World-Data ist eigentlich nur methodisch definiert und spielt mit der Suggestion, dass diese Daten – weil sie ja aus der realen Welt stammen – wenig Verzerrung und eine hohe Repräsentativität mitbringen. Damit wird etwas gemacht, was nicht zulässig ist, da der Begriff Real World Data diesem Heilversprechen nicht gerecht wird."

Dr. Antje Haas, Spitzenverband Bund



„Mein Herz schlägt für große Datensätze, mit denen man die gesamte Welt der Krankheiten abbilden und sämtliche Assoziationen darstellen und sehen kann, ob sie mit dem, was man weiß, übereinstimmen. Man kann Zusammenhänge zwischen Therapien, Arzneimitteln und Krankheitsbildern, auch sozial unerwünschte Wirkungen abbilden und sogar seltene Effekte darstellen, die man sonst nie darstellen könnte, ebenso eine ganze Reihe von Komorbiditäten. Natürlich können damit auch falsche Assoziationen oder Scheinassoziationen abgebildet werden. Doch letztendlich liegt in Big Data ein großer Schatz verborgen. Jedes Problem hat seine Lösung: RCT sind gut für den Nachweis von Wirkungen, doch nicht für beispielsweise die Aufdeckung von neuen Zusammenhängen. Darum: Alles kann uns weiterbringen. Mein Rat: Lasst uns das Kind nicht mit dem Bade ausschütten! Darum sage ich: Versorgungsforschung, öffne dich! Dazu bist du bisher nämlich nicht bereit gewesen."

Prof. Dr. Bertram Häussler, IGES Institut

https://www.ncbi.nlm.nih.gov/pubmed/23075287

„Eine wissenschaftliche Arbeit von Saunders et al zeigte, dass von insgesamt 180.590 Patienten mit Typ-2-Diabetes, die im Scottish Diabetes Research Network erfasst sind, durch den Einfuss vieler Einschluss- und Ausschlusskriterien höchstens 35,7, minimal aber nur 3,5% der Patienten in ein RCT eingeschlossen werden könnten. Das heißt: Die Studie gilt für 96% der Patienten nicht! Das soll kein Bashing gegen RCT sein, sondern einfach als ein Fakt dargestellt sein, dessen man sich einfach klarwerden muss. Deswegen sagen die Studienautoren ganz eindeutig: Wir fanden heraus, dass viele der großen Studien zur glykämischen Kontrolle bei Menschen mit Typ-2-Diabetes eine begrenzte externe Validität haben, wenn sie auf eine populationsbasierte Kohorte von Menschen mit Typ-2-Diabetes angewendet werden. Insbesondere entspricht die Altersverteilung der Studienteilnehmer häufig nicht der von Menschen mit Typ-2-Diabetes in einer zeitgenössischen britischen Bevölkerung."*

Prof. Dr. W. Dieter Paar, Sanofi





„Es wundert einen schon, dass im Rahmen eines HTA-Prozesses faktisch niemand daran interessiert ist, was später in der realen Welt an Effekten herauskommen könnte. Offensichtlich kapituliert man damit einfach vor den Herausforderungen einer zu Beantwortung dieser Frage nötigen, aber möglichen Simulation. Doch denke ich, die Gesellschaft als auch die Pharmaindustrie sollte durchaus ein Interesse an derartigen Simulationen haben, weil es durchaus Arzneimittel gibt, die ihr Potenzial nur oder hauptsächlich in der realen Welt ausspielen können."

Prof. Dr. Thomas Wilke, Ingress-Health/IPAM

„Mit dem Verfahren des Propensity Score-Machtings kann man Patienten identifizieren, die eher in einer klinischen Studie oder eher in einer Beobachtungsstudie zu finden waren. Wenn man diese nach Propensity Scores sortiert, steht ein höherer Propensity Score für einen Patienten, der sich typischerweise in einer klinischen Studie findet und ein niedriger Score für einen Patienten, der typischerweise in einem Versorgungsalltag zu finden ist. Wenn dem so ist, würden sich alle idealerweise relativ symmetrisch um eine Nulllinie verteilen, denn nur dann hat man in relativen Anteilen ungefähr so viele Patienten im Alltag, wie in der Klinik. Nur ist das bei vielen Studien offensichtlich nicht der Fall: Wir haben, um das an einem Beispiel darzustellen, viele Patienten in einer klinischen Studie, die relativ wenig Entsprechung im Alltag haben."

Dr. Michael Happich, Lilly





„Ich glaube, da sind wir uns alle einig: Patienteneigenschaften können wir abzählen. Wir scheitern immer dann, wenn wir Risikofaktoren in ein Modell einbauen wollen, die wir nicht kennen. Das ist ein alter Lehrsatz der klinischen Epidemiologie: Man kann um einen Faktor nicht risikoadjustieren, um dessen Existenz man überhaupt nicht weiß."

Prof. Dr. Matthias Schrappe, Universität Köln