



VEMS-Manifest für eine wissenschaftliche Zweckmässigkeitsforschung und -beurteilung medizinischer Leistungen

Olten, März 2014

Impressum

Autor: Flavian Kurth, Michel Romanens

Inhaltliche Begleitung
und Projektverantwortung: Michel Romanens

Nutzungsrecht: Verein Ethik und Medizin Schweiz VEMS
Verwendung der Texte, auch auszugsweise, nur mit
Quellenangabe

Zusammenfassung

Das vorliegende Manifest geht der Frage nach, weshalb die Steuerungsmassnahmen unseres Gesundheitswesens der letzten Jahre so wenig Wirkung hatten und haben oder sogar zu gegenteiligen Effekten führen. Hierzu werden die wissenschaftlichen Grundlagen untersucht, auf denen sie fussen. Dabei zeigen sich fünf Problemfelder:

1. das Abhängigkeitsproblem
2. das Koordinationsproblem
3. das Datenbasisproblem
4. das Voreingenommenheits- und Methodikproblem
5. das Kommunikationsproblem

1. das Abhängigkeitsproblem

Die wissenschaftlichen Grundlagen der Steuerungsentscheide unseres Gesundheitswesens sind Public-Health-Studien. Die finanzielle Unabhängigkeit der meisten Schweizer Public-Health-Institute ist allerdings nicht gegeben. Deren Geschäftsmodelle sehen eine mindestens teilweise Finanzierung durch Auftragsarbeiten vor. Neben Interessenverbänden figurieren hier als Auftraggeber vor allem die Krankenkassen und deren Dachverbände, namentlich die santésuisse. Dies wäre insofern nicht zwingend problematisch, als sich auch Institute, die klinische Forschung betreiben, in einer gewissen Abhängigkeit von der Pharmazeutischen Industrie befinden, weshalb deren Arbeiten entsprechend streng überwacht werden. Die dafür zuständigen Stellen wie EDI, BAG, SAMW und Ethikrat der Statistiken der Schweiz messen hier aber nicht mit gleichen Ellen: Viele unserer Public-Health-Studien genügen gleich in mehreren Punkten den Auflagen der GEP (good epidemiological practice) (2) bezüglich Planung, Durchführung, Analyse und Beurteilung der Resultate nicht, ohne dass die Autoren dafür gerügt würden. Im Gegenteil werden mangelhafte Arbeiten in gängigen Verfahren unserer Behörden angewendet, von diesen Stellen teilweise sogar zitiert und von unserer Regierung dazu verwendet, Steuerungsentscheide zu promovieren, die so opportun wie fragwürdig sind.

Dass man hier auf einem Auge blind ist, liegt an eben diesem Opportunismus, der auf einem stillschweigenden Konsens bezüglich der Herangehensweise an die Herausforderungen

unseres Gesundheitswesens fusst: Es gilt, die Leistungserbringer in die Schranken zu weisen, um Kosten zu sparen. Aufgabe der Forschung wäre es nun aber gerade, diese Herangehensweise und ihre wissenschaftliche Legitimation kritisch zu hinterfragen, um festzustellen, wo und wie tatsächlich und auch langfristig Geld zu sparen ist. Dies ist unabdingbar, wollen wir zu wirkungsvollen Steuerungsentscheiden gelangen, insbesondere heute, da sich abzeichnet, dass die derzeitige Herangehensweise problematisch ist und die prognostizierten Erfolge weitgehend ausbleiben – ja dass sich sogar ein gegenteiliger Effekt zeigt: eine stetige Verteuerung bei sinkender Qualität, bzw. Verfügbarkeit medizinischer Versorgung. Nur eine Finanzierung frei von Interessen kann die momentane Situation verbessern, denn nur dann können unsere Public-Health-Institute unabhängig forschen, und nur unabhängiges Forschen ist wissenschaftlich und zielführend.

2. das Koordinationsproblem

Public-Health-Studien sind teuer, da sie in vergleichsweise grossen Kohorten vergleichsweise komplexe Zusammenhänge untersuchen. Die Schweiz als kleines Land stösst hier schnell mal an ihre Grenzen. Bemühungen, die Arbeit der verschiedenen Public-Health-Institute zu koordinieren, scheitern aber am «Kantönligeist». Eine solche Koordination und die Erstellung nationaler Register wären indes Voraussetzung, um in absehbarer Zeit nur schon die Voraussetzungen dafür zu schaffen, professionelle Public-Health-Studien zu erstellen. Dies zeigt ein Blick auf Studien, die in Deutschland erstellt wurden und bei vergleichbaren Voraussetzungen teilweise zu Schlüssen kommen, die den Schlüssen von Schweizer Studien diametral entgegengesetzt sind. Hier ist der Bund gefordert, auf nationaler Ebene Strukturen aufzubauen, die so überzeugend sind, dass sich die Kantone daran beteiligen.

3. das Datenbasisproblem

Die Datenbasis einer Studie bestimmt, was mit ihr überhaupt gesagt werden kann und welchen Evidenzgrad ihre Aussagen haben werden. Die Auflagen der GEP (good epidemiological practice) sehen deshalb vor, die Datenbasis aufgrund des Studiendesigns zu erstellen. Eine Forderung, der viele Public-Health-Studien der Schweiz nicht genügen. Oftmals wird mit Datenbasen gearbeitet, die eben gerade billig verfügbar sind. In diesen

Datenbasen sind dann nicht selten wichtige, die behaupteten Sachverhalte begründende Erklärungsvariablen nicht erfasst. Oder es wurden in den Datenbanken Subgruppen gebildet, die mit unterschiedlicher Gewichtung in die Gesamtbeurteilung fallen, was zu einer komplett falschen Sichtweise auf die Sachverhalte führen kann und unter dem Begriff «Simpson Paradox» im wissenschaftlichen Diskurs als Problem erkannt ist. Auf der Basis solcher Datenbanken Studien zu erstellen ist nicht nur eine Verschwendung wissenschaftlicher Ressourcen, es hat auch ein gefährliches Missbrauchspotenzial, indem dann Studien vorliegen, die wissenschaftlich überzeugend wirken, die behaupteten Sachverhalte aber aufgrund ihrer Datenbasis gar nicht beweisen können.

4. das Voreingenommenheits- und Methodikproblem

Kein Mensch ist frei von Vorurteilen. Diese schränken unseren Blick auf den Gegenstand der Betrachtung mehr oder weniger ein. Die Regeln der Wissenschaftlichkeit schreiben deshalb vor, eine möglichst unvoreingenommene Haltung einzunehmen, also zuerst hinzuschauen, was ist, und erst dann Schlüsse daraus zu ziehen. Dieser fast schon banalen Forderung genügen viele unserer Public-Health-Studien und Beurteilungen der Medizin nicht. Die Voreingenommenheit besteht hier in einer einseitig ökonomisch-utilitaristischen Haltung, welche sich mit dem Argument des dringenden Handlungsbedarfs zur Eindämmung der Kosten begründet. Diesen stellt der VEMS nicht in Abrede, die Haltung und die daraus resultierende Methodik hingegen schon.

Die Ökonomie ist eine Mittelwert-Disziplin, sie vergleicht, errechnet Durchschnitte, wägt Kosten und Nutzen gegeneinander ab und projiziert auf das Individuum, was im Durchschnitt das bestmögliche Kosten-Nutzen-Verhältnis aufweist. Die Medizin geht den umgekehrten Weg: Sie rechnet ihre Erkenntnisse aus Einzelbeobachtungen hoch auf eine Metaebene und gelangt so zu ihren Schlüssen, die dann in der Behandlung auf das Individuum projiziert und individuell angepasst werden. Sieht man im Indikations- und Behandlungsentscheid hingegen in erster Linie eine Kosten-Nutzen-Abwägung, wie dies heute der Fall ist, wird man der Medizin nicht gerecht. Sie fragt nach der Zweckmässigkeit, und dies mit gutem Grund: Würde sie nach der Nützlichkeit fragen, blieben viele Patienten unbehandelt, ja viele Krankheiten wären heute gar nicht behandelbar, weil man sie aufgrund schlechter Nutzwertanalysen nie behandelt, ergo auch nie untersucht und erforscht hätte. Die

Frage der Medizin ist nicht, ob eine Intervention nützt, sondern ob sie zur Erfüllung des Patientenauftrags nötig, ob sie zweckmässig ist. So muss sie fragen, anders darf sie nicht fragen, will sie Anwältin ihrer Patientinnen und Patienten sein, die krank sind und gesund werden wollen, die leiden und für ihr Leiden einer Linderung bedürfen.

Die Medizin wird nun aber durch diese inadäquate Beurteilung und Behandlung, mit der man ihr begegnet, nicht eine andere; sie sucht sich vielmehr Wege, innerhalb der neuen Rahmenbedingungen, die ihr die Ökonomie setzt, zu überleben und dabei so gut wie möglich Medizin zu bleiben. Dies kann dann auf der Mesoebene (Spitäler, Pflegeheime, Institutionen) und auf der Mikroebene (Arbeit am Krankenbett) dazu führen, dass ein Spital einem Patienten aufgrund seines Alters einen Eingriff verwehrt, der zweckmässig wäre, aber in diesem Fall als nicht nützlich gilt, beispielsweise eine Hüftgelenkoperation, um auf der anderen Seite den so verlorenen Umsatz damit wettzumachen, dass an einem jüngeren Patienten ein Eingriff vorgenommen wird, der nicht zweckmässig ist, aber in diesem Fall als nützlich gilt, beispielsweise eine präventive Blinddarmentfernung oder eine präventive Knieoperation.

Die Steuerungsentscheide unseres Gesundheitswesens auf der Makroebene (Gesundheitspolitik- und -ökonomie) zwingen die Leistungserbringer in ein Verhalten, das nicht zu kosteneindämmenden, sondern zu gegenteiligen Effekten führt: unser Gesundheitswesen wird aufgrund einer immer schlechteren Qualität der Indikations- und Behandlungsentscheide immer ineffizienter, ergo immer teurer. Diese Steuerungsentscheide sollten deshalb überdacht werden. Der Wissenschaft fällt dabei die Aufgabe zu, Grundlagen zu liefern, die die Sachverhalte wirklich zu erklären versuchen, anstatt, wie dies heute der Fall ist, Bestätigungen vorgefertigter Meinungen zu liefern. Hierzu muss ein Paradigmenwechsel von der Nützlichkeits- und Wirtschaftlichkeits- hin zu einer Zweckmässigkeitsbeurteilung und -forschung stattfinden. Dies ist erst möglich, sieht man den Teufelskreis und durchbricht ihn: Weil man der Medizin voreingenommen entgegentritt, trifft man Steuerungsentscheide, welche diese in ein Verhalten zwingen, das die vorgefertigten Meinungen dann bestätigt.

5. das Kommunikationsproblem

Wissenschaftliches Forschen verpflichtet sich einer transparenten Kommunikation. Nur so ist gewährleistet, dass nicht nur wissenschaftlich gearbeitet, sondern auch effektiv geforscht wird. Denn die Forschung lässt sich nur weiterbringen, wenn sie anschlussfähig ist, wenn also

an Studien angeknüpft werden kann. Dies ist nicht gegeben, sind die Arbeiten intransparent und widersprüchlich. Der VEMS nennt solche Arbeiten Janusstudien und -beurteilungen. Wir halten diese für eine grosse Gefahr, der nur begegnet werden kann, indem die involvierten Wissenschaftler offen und ehrlich kommunizieren, welche Aussagen ihre Studien machen und mit welcher Evidenz.

So wäre die derzeit gegebene Situation, dass wir über die Zusammenhänge des Gesundheitswesens wenig wirklich wissen, sukzessive zu verbessern. Nur Ehrlichkeit führt uns der Wahrheit näher. Dies geschieht allerdings viel zu wenig, herrschende Methodik, Arbeitshaltung und Arbeitsweise schaffen eine negative Wissenschaftskultur und untergraben die Glaubwürdigkeit der Forschung unseres Gesundheitswesens. Um es mit Friedrich Nietzsche zu sagen: «Unser Zeitalter, so viel es von Ökonomie redet, ist ein Verschwender: es verschwendet das Kostbarste, den Geist» (Nietzsche, Morgenröte, drittes Buch, Vers 179). Die Herausforderungen unseres Gesundheitswesens sind nur zu meistern, wenn sich die Wissenschaft wieder auf ihre Rolle als Wahrheitsfinderin besinnt und nicht lediglich Argumente für Entscheide liefert, die praktikabel und opportun sind, weil sie den Partikularinteressen der stärksten Kräfte im System zudienen.

Inhaltsverzeichnis

| | | |
|---|-------|----|
| Ausgangslage | Seite | 8 |
| Wie steht es um die Wissenschaftlichkeit im Gesundheitswesen? | Seite | 12 |
| Public-Health-Studien und Versorgungsforschung | Seite | 17 |
| HTA, QALY, Missbrauch der Empfehlungen der Fachgesellschaften | Seite | 19 |
| Fehlende oder mangelhafte Begleitforschung | Seite | 21 |
| Was ist Auditforschung und was bringt sie? | Seite | 23 |
| Von der Janusforschung zur Auditforschung | Seite | 25 |
| Dreh- und Angelpunkt Indikationsentscheid | Seite | 29 |
| Drei Schritte zur Zweckmässigkeitsforschung | Seite | 31 |
| Referenzen | Seite | 37 |

Ausgangslage

Im Gesundheitswesen zahlen die Nachfrager (die Patienten) nicht direkt für die bezogenen Dienstleistungen, sondern indirekt über ihre Versicherungsbeiträge. In einem solchen System (third-party payment) kommt es naturgemäss sowohl zu einer Überproduktion als auch zu einem Überkonsum, weshalb ohne regulierende Steuerungsmassnahmen die Gefahr einer unkontrollierten Kostenausweitung bestünde. Ziel regulativer Steuerungsmassnahmen ist die optimale Allokation der zur Verfügung stehenden Mittel. Dies ist ein Auftrag des Souveräns an Politik und Behörden, ihm gilt es, Partikularinteressen der verschiedenen Akteure des Gesundheitswesens unterzuordnen, der Industrie, der Leistungserbringer und der Krankenkassen. Es gibt drei Möglichkeiten, mit Steuerungsentscheiden auf die Gesundheitskosten einzuwirken:

Kontrolle des Preises

Preisverhandlungen mit der Industrie und den Leistungserbringern sind im Gesundheitswesen die direkteste Massnahme der preisregulierenden Kostenkontrolle. Daneben können aber auch Auflagen an die Leistungserbringer helfen, den Preis zu drücken. Dies wurde mit den Fallpauschalen versucht, die die Spitäler in einen Kostenrahmen pro Behandlung zwingen. In England, dessen Gesundheitswesen planwirtschaftlichere Züge aufweist als unseres, kann auf den Preis noch direkter eingewirkt werden, indem Auflagen an Abläufe, Prozesse und Strukturen der Leistungserbringer auf eine Verbesserung der Effizienz zielen.

Kontrolle der Nachfrage:

Indem für den Konsumenten Anreize geschaffen werden, die angemessenen Konsum belohnen und übermässigen bestrafen, soll eine Regulierung der Nachfrage gelingen. Die Krankenkassen bieten beispielsweise Konzepte wie Managed-Care oder Telemedizin an, bei welchen der Konsument von einer vergünstigten Prämie profitiert und im Gegenzug bereit ist, bezüglich seines Konsums im Krankheitsfall bestimmte Abstriche in Kauf zu nehmen. Auch wenn Patienten, die sich bemühen, ihre Nachfrage nach medizinischen Leistungen durch einen gesunden Lebensstil gering zu halten, mit Anreizen wie einem Beitrag an das Fitnessabo belohnt werden, zielt das auf eine Kontrolle der Nachfrage.

Kontrolle des Angebots:

Da jeder Arzt, der zugelassen ist, seine Leistungen den Krankenkassen verrechnen kann und diese gezwungen sind, sie zu begleichen (Vertragszwang), besteht ein verständliches Bedürfnis, das Angebot medizinischer Leistungen zu kontrollieren. Dies geschieht mit den Wirtschaftlichkeitsverfahren, mit denen die frei praktizierenden Ärzte überprüft werden, ebenso mit Zulassungsbeschränkungen für bestimmte Ärzte. Auch Standards, wann der Arzt eine bestimmte Leistung erbringen darf, wirken sich regulierend auf das Angebot aus. Health Technology Assessments (HTA) beispielsweise versuchen, auf der Basis von Kosten-Nutzen-Analysen Grundlagen für Indikationsentscheide zu liefern, was darauf hinausläuft, das Angebot der OKP-Leistungen zu überprüfen und gegebenenfalls zu beschneiden.

Untersuchen wir die Steuerungsmassnahmen im Schweizer Gesundheitswesen auf ihre Absicht und ihre Wirkung hin, so stellen wir eine gewisse Einseitigkeit fest:

Kontrolle des Preises:

Bezüglich der Kontrolle des Preises waren die Bemühungen hinsichtlich der Medikamentenpreise bisher wenig fruchtbar; in der Schweiz sind diese nach wie vor wesentlich höher als in anderen Ländern. Auch ob TARMED, der Einzelleistungstarif für sämtliche in der Schweiz erbrachten ambulanten ärztlichen Leistungen, tatsächlich kostendämpfende Wirkung hat, ist seit seiner Einführung im Januar 2004 umstritten. Bei den Fallpauschalen, die 2012 in der Schweiz eingeführt wurden, zeichnet sich heute ab, dass der Druck auf den Preis nicht zu einer Dämpfung der Kosten, sondern vielmehr zu einer Verschiebung und für bestimmte Patienten zu einer Einschränkung des Angebots führt: Der Kostenrahmen legt fest, welche Leistungen und Patienten sich für ein Spital lohnen und welche nicht, was zur Folge hat, dass das Angebot entsprechend angepasst und unter Umständen beschnitten wird. Neben der Frage, ob sich so insgesamt tatsächlich Kosten sparen lassen, stellen sich gerechtigkeitsethische Fragen, die kontrovers diskutiert werden. Klar ist, dass das Fallpauschalensystem der Schweiz in der heutigen Ausgestaltung keine eigentliche Preiskontrolle mehr ist, sondern eine Kontrolle des Angebots.

Kontrolle der Nachfrage:

Ärztetzwerke als nachfrageeinschränkende Massnahme sind insofern fraglich, als der Konsument wohl eher selten freiwillig ein solches Versicherungsmodell wählt, sondern in der Regel aus finanzieller Not. Dies führt zu einer Zweiklassenmedizin, die den sozialen Charakter des Gesundheitswesens prinzipiell untergräbt. Der wesentliche Kritikpunkt bei diesen Modellen ist aber der, dass sie bis heute nicht nachweisen können, ihre Einsparungen allein durch ein verbessertes Management und eine optimierte integrierte Versorgung zu realisieren. Solange dies nicht belegt ist, steht der Verdacht im Raum, dass die Einsparungen durch Herauszögern und Vorenthalten von Behandlungen und Weiterverweisungen realisiert werden. Die geheimen Verträge jedenfalls, die Ärztenetzwerke in diesen Modellen mit Krankenkassen abschliessen, schaffen dazu einen starken Anreiz, indem Überschreiten des Budgets für Arzt und/oder Netzwerk eine Strafe zur Folge hat und Unterschreiten einen Bonus, unabhängig von der gegebenen Patientensituation. Damit haben wir nicht mehr eine nachfrageeinschränkende Kostenkontrolle, sondern de facto und hinter dem Rücken der Versicherten und Patienten eine Angebotsbeschneidung.

Kontrolle des Angebots:

Die Wirtschaftlichkeitsverfahren mit ihren Quervergleichen der Durchschnittskosten nötigen den Arzt, diese im Rahmen des Mittelwerts eines Vergleichskollektivs zu halten, unabhängig von seinem Patientengut. Es bleibt ihm nichts anderes übrig, als entweder sein Krankengut entsprechend anzupassen (preistreibende Patienten meiden) oder sein Angebot einzuschränken (teure Behandlungen meiden). Damit haben wir de facto ein Globalbudget und mit diesem all die von Gesundheitssystemen mit Globalbudget bekannten und gut dokumentierten Probleme. Die gleiche Tendenz ist bezüglich der in den letzten Monaten vor allem vom Swiss Medical Board SMB stark vorangetriebenen Health Technology Assessments HTA festzustellen. Wenn HTAs Behandlungspfade nicht mehr analysieren, um sie zu optimieren, sondern umpflügen, um wo immer möglich Behandlungen als im Verhältnis zu ihren Kosten zu wenig nutzbringend zu monieren, dann ist das primäre Ziel auch hier nicht Optimierung, sondern Beschneidung des Angebots innerhalb eines de facto Globalbudgets.

Es ist also festzustellen, dass sich die Bestrebungen, die Kosten unseres Gesundheitswesens in den Griff zu bekommen, vornehmlich auf die Kontrolle des Angebots konzentrieren und dies mit der Tendenz, die Ausgaben in den Rahmen eines Globalbudgets zu drücken. Wenn heute von einer Kostenobergrenze geredet wird, ab welcher unser Gesundheitswesen nicht mehr bezahlbar sei, dann ist das ja nichts anderes als der Ruf nach einem Globalbudget. Das Problem eines solchen ist, dass es einen Kostenrahmen festlegt, ohne die effektive Patientensituation zu berücksichtigen. Die Folge ist die schlimmste, weil ungerechteste Art von Rationierung: Eine als Rationalisierung getarnte Rationierung schafft ein willkürliches System der Angebotsbeschränkung, das den benachteiligt, der das Pech hat, zur falschen Zeit im falschen Versicherungsmodell zu sein, zur falschen Zeit im falschen Spital, beim falschen Arzt, oder dessen Körper sich weigert, seine Krankheitssymptome gemäss den Mittelwerten zu entwickeln.

Der Anreiz zur Überversorgung ist fraglos ein zentrales Problem des Gesundheitswesens; ihm muss mit Angebotseinschränkung begegnet werden. Diese Einschränkung darf sich aber nur auf jenen Teil des Angebots konzentrieren, der nachweislich medizinisch nicht notwendig ist. Geht man darüber hinaus und beschränkt auch Angebote, die medizinisch angezeigt sind, so wird der Rahmen der Rationalisierung gesprengt und – meist verdeckt – rationiert. Die alles entscheidende Frage ist also die, wie man möglichst objektiv feststellen kann, ob eine Massnahme medizinisch angezeigt ist oder nicht. In jüngster Zeit hat sich hierzu die Anwendung von Kosten-Nutzen-Analysen etabliert. Deren Einführung im Gesundheitswesen hat allerdings zu einer Reihe von Problemen geführt, die der VEMS in seiner Evaluation «Rationierung, QALY und die mathematische Maschinerie» (1) untersucht.

Grundsätzlich ist der Begriff des Nutzens der Medizin fremd, sie beurteilt ihre Interventionen nach dem Kriterium der Zweckmässigkeit. Dies mit gutem Grund: Ob eine medizinische Massnahme der Heilung nützt, ist nicht immer bestimmbar, ob sie dem Zweck der Heilung und Linderung angemessen ist, ob sie also zur Erfüllung des Patientenauftrags nötig ist, hingegen schon. Kosten-Nutzen-Analysen andererseits arbeiten mit Durchschnittsvergleichen. Dies kann zur Folge haben, dass für bestimmte Patientengruppen medizinische Angebote eingeschränkt werden, die in ihrem Fall nützlich und zweckmässig wären, aufgrund des Mittelwertvergleichs aber als nicht nützlich erscheinen. Sie verschwinden also quasi vom Radar der Betrachtung und haben das Nachsehen.

Die Wissenschaft ist gefordert, solide Grundlagen für angebotseinschränkende Steuerungsmassnahmen zu liefern. Sie soll jene medizinischen Interventionen identifizieren, die zweifelsfrei nicht angebracht, also nicht zweckmässig, unnötig waren oder sind. Irrt sie, identifiziert sie falsch, so werden die aus ihren Grundlagen resultierenden Einschränkungen von den Patienten eben kompensiert: indem diese auf einen anderen – unter Umständen teureren – Angebotskanal wechseln (z.B. von der Praxis des Spezialisten ins Spitalambulatorium) oder indem die vorenthaltenen Leistungen verspätet und mit teilweise erheblichen Mehrkosten eingefordert werden – ganz einfach, weil sie dann eingefordert werden müssen, um das Leben des Patienten nicht zu gefährden. Eine unprofessionelle Bedarfsforschung leistet also einen wesentlichen Beitrag zu einer Verteuerung des Gesundheitswesens. Mit diesem Manifest soll aufgezeigt werden, welche wissenschaftlichen Fehler der Forschung unseres Gesundheitswesens zu welchen Fehlschlüssen führen und was deren Folgen sein können oder bereits sind.

Wie steht es um die Wissenschaftlichkeit im Gesundheitswesen?

Das Ziel der Wissenschaft ist es, Dinge zu erklären, um daraus Handlungen abzuleiten, die dem Menschen zugutekommen. Keine wissenschaftliche Theorie kann für sich beanspruchen, die letzte Wahrheit gefunden zu haben, denn die Wissenschaft entwickelt sich laufend weiter. Eine Theorie muss hierzu nicht nur bezüglich der Fälle überprüfbar sein, für die sie zutrifft (Verifizierbarkeit), sondern auch bezüglich jener, in denen sie dies nicht ist (Falsifizierbarkeit). Ist dies nicht möglich, droht der nicht überprüfte Fall, für welchen sie nicht zutrifft, überraschend einzutreffen, wodurch Handlungen, die sich allein auf die Verifizierbarkeit stützen, ungewünschte Folgen haben können. Auch kann eine Theorie nur weiterentwickelt werden, wenn beides gegeben ist, denn nur in diesem Fall ist sie anschlussfähig.

Die Newtonschen Axiome beispielsweise vermögen die Bewegung von Körpern auf der Erde zu erklären, befinden sich die Körper aber im Raum, greifen sie nicht. Die Relativitätstheorie vermag diese Phänomene zu erklären, versagt dann aber bei der Erklärung der Phänomene im subatomaren Bereich, wozu eine neue Theorie erforderlich ist, die Quantenphysik. Die Newtonschen Axiome mussten also falsifizierbar sein, damit die Relativitätstheorie überhaupt entwickelt werden konnte, und bevor diese entwickelt war, wäre

es verantwortungslos gewesen, bemannte Raumflüge durchzuführen. Nur wenn beides, Verifizierbarkeit und Falsifizierbarkeit, gegeben ist, redet man von empirischer Forschung. Die Aussage «Heute regnet es» ist sowohl verifizierbar als auch falsifizierbar. Die Aussage «Heute regnet es oder nicht» hingegen ist nicht falsifizierbar. Solche Aussagen nützen nicht nur nichts, sie haben hohes Potential, Schaden anzurichten. Das Resultat empirischer Forschung andererseits sind Erkenntnisse, die abzuwägen helfen, welche Handlungen sinnvoll sind, weil sie mit sehr hoher Wahrscheinlichkeit die gewünschten Folgen haben werden. Ist die Forschung hingegen nicht empirisch, so ist sie keine Hilfe beim Abwägen, lässt sich aber sehr gut zu manipulativen Zwecken missbrauchen, womit sie Schaden anrichtet. Damit empirisches Forschen möglich ist, sind folgende Auflagen zu erfüllen:

Unvoreingenommenheit

Die Arbeit darf die in ihr postulierten Thesen nicht zur Herleitung des zu erbringenden Beweises voraussetzen. Sie muss als Erstes unvoreingenommen hinschauen und beobachten, was ist, bevor sie dazu ansetzt, die beobachteten Phänomene zu erklären.

Transparenz

Die Arbeit muss beim Erklären des Festgestellten ehrlich sein. Sie muss offenlegen, wo sie weshalb welche Schlüsse zieht, andere Schlüsse einräumen und Auskunft darüber geben, weshalb diese nicht weiterverfolgt werden.

Wissenschaftliche Logik und Methodik

Die Arbeit muss diskursiv sein, also so ausgestaltet, dass sie es jedem Leser ermöglicht, die in ihr formulierten Schlussfolgerungen logisch nachzuvollziehen, damit er, falls er andere Schlüsse zieht, diese begründen und die Theorie so weiterentwickeln kann.

Viele Arbeiten, auf die sich Steuerungsentscheide des Schweizer Gesundheitswesens stützen, und viele Methoden, mit denen operativ bei der Durchsetzung dieser Entscheide gearbeitet wird, genügen diesen Auflagen an die Wissenschaftlichkeit nur teilweise oder gar nicht. Sie sind indes internationaler wissenschaftlicher Standard und werden auch in den Empfehlungen der GEP (good epidemiological practice) (2) bezüglich Planung, Durchführung, Analyse und

Beurteilung der Resultate festgehalten. Die Forschung im Gesundheitswesen Schweiz hält sich aber nicht daran.

In der Regel werden in den Schweizer Public-Health-Studien auf der Basis von Mittelwertvergleichen Schlussfolgerungen gezogen, ohne die effektiven Ursachen der beobachteten Unterschiede untersucht zu haben. Der Aussagewert einer solchen Beurteilung oder Studie ist in manchen Fällen so gering, dass aus ihr auch ein anderer Schluss gezogen werden kann, unter Umständen sogar der gegenteilige. Der VEMS nennt solche Arbeiten deshalb Janusstudien und -beurteilungen. Die Theorien der Gesundheitsökonomie funktionieren, wie jede ökonomische Theorie, nur im Modell, in einer Idealwelt. Die Wirklichkeit ist komplexer, und an dieser ist jede Theorie zu werten: verbessert sie eine gegebene Situation oder verschlechtert sie sie? Ja ist sie nur schon in der Lage, die gegebene Situation zu erkennen und akkurat abzubilden oder verleitet sie vielmehr zu Fehlschlüssen und -Entscheidungen? Dass bei unseren Studien tendenziell Letzteres der Fall ist, soll im Folgenden anhand einiger Beispiele aufgezeigt werden.

Wirtschaftlichkeitsverfahren WZW

Der Dreh- und Angelpunkt bei der Frage, ob tatsächlich rationalisiert oder nicht vielmehr rationiert wird, ist die Beurteilung ärztlicher Arbeit. Mit welcher Methodik werden Ärztinnen und Ärzte identifiziert, die zu ihrer persönlichen Bereicherungen Behandlungen vornehmen, welche medizinisch gar nicht nötig sind? Wie ahndet man diese schwarzen Schafe, wie rügt und wie büsst man sie? Und vor allem: wer tut dies mit welchem wissenschaftlichen Hintergrund und mit welcher Rechtspraxis?

Die Antworten auf diese Fragen sind alarmierend. Seit rund 20 Jahren prüft der Dachverband der Schweizer Krankenkassen santésuisse Schweizer Ärztinnen und Ärzte gestützt auf die im KVG verankerte WZW-Regel daraufhin, ob ihre Behandlungen wirksam, zweckmässig und wirtschaftlich sind. Der Arzt wird betreffend Jahreskosten pro Patient mit dem Kollektiv einer Gruppe Ärzte verglichen. Eine Margenüberschreitung von 30% wird gemäss Bundesgericht als Beweis für Überarztung betrachtet.

Die Methode operiert allerdings anhand von Mittelwertvergleichen, wobei die Variablen des Vergleichs eindeutig ungenügend sind. So erklärt die Variable «Standort-Kanton» in einem Kanton wie beispielsweise Bern zu wenig, da hier aufgrund der Grösse ein

Arzt auf dem Land ganz andere Voraussetzungen hat (mehr Hausbesuche) als ein Arzt in der Stadt (mehr chronisch kranke Patienten). Auch die Variable «Facharztgruppe» hat zu wenig Erklärungskraft. So bilden zum Beispiel Kardiologen und Internisten eine Vergleichsgruppe, obwohl sie sich stark unterscheiden können (Kardiologe höhere Behandlungskosten, Internist höhere Medikamentenkosten). Wichtige Variablen mit hoher Erklärungskraft wie die Morbidität andererseits werden gar nicht erst erhoben.

Zusammenfassend ist festzuhalten, dass die Arztkosten hier durch die Arztkosten selbst erklärt werden. Als seien die Kosten von Arzt A für die Kosten von Arzt B ursächlich. Unterschiede der Durchschnittskosten wären gemäss Auftrag doch nicht zu monieren und zu büssen, sondern vielmehr zu erklären, und zwar dadurch, wie sie effektiv entstehen: durch die Patienten und die Arbeit, die der Arzt an ihnen leistet. Die gewählten Vergleichsvariablen erklären deshalb gerade einmal 6 bis 11% der Arztkosten. Ob ein Arzt als Überarzt angezeigt wird, hängt folglich wesentlich davon ab, welches Patientengut er hat und in welcher Vergleichsgruppe er ist, nicht oder nur bedingt aber davon, ob er tatsächlich unnötige Leistungen erbringt. Würden Schüler so beurteilt, so würde ein mathematisch begabtes Kind in einer Klasse von musisch begabten als Schulversager angezeigt. Ein Beweis für Überarztung ist so jedenfalls nicht zu erbringen. Dies ist nur auf der Ebene der Krankengeschichten des angeschuldigten Arztes möglich, was aufgrund des Aufwands kaum je getan wurde. Hierzu könnte ein Vergleich der Durchschnittskosten mit denen von Kolleginnen und Kollegen allenfalls als Prescreening im Sinne eines Hinweises dienen, wo genauer hinzuschauen ist. Dieses Hinschauen muss dann aber in Stichproben von Einzelfällen bestehen, denn nur auf dieser Ebene lassen sich erbrachte unnötige medizinische Interventionen oder unterlassene nötige Interventionen nachweisen. Die heutige Methode hingegen zeigt sowohl Nichtüberarzt als Überarzt an, als auch Überarzt als Nichtüberarzt. Dieser Grobvergleich wird dann zwar durch Studium der Rechnungssteller-Statistik (RSS) des jeweiligen Arztes verfeinert und relativiert. Dies aber erstens basierend auf einer mangelhaften Datengrundlage (wichtige Variablen fehlen in der RSS, Verzerrungen durch Biases wie «confounding», «aggregation», «simpson's paradox», «collinearity», «tautology», «endogeneity from omitted variable bias») und zweitens von der falschen Stelle: der *santésuisse*, die nicht objektiv sein kann, da sie finanzielle Partikularinteressen hat.

Aus der heute gängigen Praxis resultieren willkürliche Entscheide, die beim Arzt das Gefühl einer Machtlosigkeit zurücklassen. Die Bestrafung etlicher korrekt arbeitender Ärzte

hat ein allgemeines Klima der Angst mit der Folge einer restriktiven Verschreibungspraxis auch bei nötigen Behandlungen zu verantworten. Dies verursacht weitaus höhere Folgekosten. Andererseits finden geschickte Überarzte Wege, sich mit Tricks aus der Statistik zu stehlen, indem sie ihr Verhalten entsprechend anpassen, auch wenn das dann vielleicht nicht im Interesse des Patienten ist: Loswerden der teuersten Risiken, Labor umgehen, indem man nicht alles abrechnet, möglichst viele junge, nicht hohe Kosten verursachende Patienten im Patientenkollektiv aufnehmen.

Wir haben hier also den klaren Fall einer wissenschaftlich mangelhaften Beurteilung, deren Folge ein Fehlverhalten seitens der Leistungserbringer ist, welches Schaden und Mehrkosten verursacht. Der VEMS behandelt dieses Problem in seinem Dossier «Wirtschaftlichkeitsverfahren» (www.vems.ch/projekt-wzw) und handelt es wissenschaftlich in einem entsprechenden Grundlagenpapier ab, welches er im Rahmen seiner Arbeit in der Arbeitsgruppe WZW der FMH ausgearbeitet hat (3). Die in diesem Papier erarbeiteten und begründeten Forderungen lassen sich im Wesentlichen auf vier Punkte zusammenfassen:

- Komplette Neuausrichtung der WZW Verfahren bis Ende 2014 im Sinne der Professionalisierung der Methode unter primärer Verwendung der Patientenmorbidity.
- Damit dies geschehen kann: Einsetzung einer paritätischen Steuerungsgruppe mit Weisungsbefugnis, welche die notwendigen Anpassungen rasch umsetzt.
- Bis dies geschehen ist: Einstweiliges Prüfmoratorium; Anpassen der RSS-Statistik betreffend Sicherheitsmargen (höherer Index als Cutoff für den Überarztungsverdacht, z.B. > 200%).
- Grundsätzlich: keine Kostenvergleiche unter Ärzten.

Die derzeitige Situation können nur die Krankenversicherer wirklich wollen. Allen anderen Akteuren, vornehmlich den Versicherten und Patienten, erwachsen daraus nur Nachteile. Hier ist von allen in aller Entschiedenheit gegen die einseitige Ausgestaltung des Gesundheitswesens zugunsten der Versicherer zu kämpfen. Dies ist ein Kampf, den in erster Linie die Ärzteschaft selbst führen muss, was sie erst kann, wenn sie sich vom Nützlichkeitsdiktat distanziert und zurück zum Zweckmässigkeitsdiktat der Medizin findet. Damit wäre der Weg frei für eine adäquate Beurteilung der Ärzte.

Public-Health-Studien und Versorgungsforschung

Studien, die einen Zusammenhang von Ärztedichte und Überarztung als Folge eines Überangebots beweisen wollen (4), haben ihren Weg in die Publikumsmedien gefunden und das Bild des Arztes als das eines potentiellen Abzockers gefestigt. Zurecht? In Einzelfällen sicher, ja. Doch ist das auch statistisch relevant? Kann nachgewiesen werden, dass ein Überangebot zwingend zu einer Überversorgung führt? Der Arzt kann auf eine schlechte Auslastung seiner Praxis ja auf zwei Art reagieren: Er erbringt Leistungen, die nicht angebracht sind, um seine Praxis auszulasten, passt also die Nachfrage dank seinem Wissensvorsprung gegenüber dem Patienten dem Angebot an. Oder er redimensioniert, schafft Teilzeitstellen, entlässt Personal, passt sein Angebot also der Nachfrage an. Ist beweisbar, dass er in statistisch signifikantem Mass Ersteres tut?

Wäre dies der Fall, so wären entsprechende Steuerungsmassnahmen gerechtfertigt: Zulassungsstopps für neue Ärztinnen und Ärzte, Erschweren der Bedingungen, eine Praxis zu eröffnen, Verschärfen der Vorschriften, Erhöhen der Kontrollen etc. Ist dies jedoch nicht der Fall, schaffen dieselben Bedingungen eine Situation, die dem Auftrag einer adäquaten Versorgung zuwiderläuft: die Situation einer Unterversorgung. Insbesondere in Kantonen mit vielen älteren, behinderten und schwer kranken, polymorbiden Patienten wäre das fatal. Dies soll an einem Beispiel aus dem Kanton Tessin veranschaulicht werden: Nachdem ein grosses kardiologisches Zentrum eröffnet wurde, stiegen die entsprechenden Eingriffe sprunghaft an. Daraus kann man folgern, dass dieses Zentrum, um ausgelastet zu sein, Behandlungen machte, die gar nicht nötig gewesen wären. Oder man kann folgern, dass im Kanton Tessin, in welchem viele ältere Personen aus anderen Kantonen ihren Lebensabend verbringen, zuvor diesbezügliches ein Unterangebot geherrscht hat, weshalb diese Patienten sich früher in anderen Kantonen behandeln liessen, beispielsweise in ihrem früheren Wohnkanton, wo sie den Arzt noch kennen. Folgert man Ersteres und ist dies nicht tatsächlich der Fall, droht eine Unterversorgung. Was ist also richtig, was lässt sich nachweisen? Und wie steht es im Kanton Baselstadt mit seinem angeblichen Überangebot psychiatrischer Dienstleistungen?

Ein Hinweis darauf könnte der Wohnkanton der behandelten Patienten sein: stammen sie aus dem Kanton ihres Arztes oder reisen sie an? So banal dies erscheint: Unsere Public-Health-Studien arbeiten mit Datenbanken, in denen diese Variable gar nicht erfasst ist, vornehmlich ist dies in den Datenbanken des Dachverbands der Schweizer Krankenkassen santésuisse der Fall. Was kann auf dieser Grundlage überhaupt bewiesen werden, was muss

vermutet bleiben, was ist Fakt, was Hypothese und mit welcher Begründung? Ehrliche Antworten würden hier Ordnung der Dinge schaffen und helfen, zusammen mehr Wissen über die effektive Situation der medizinischen Versorgung zu erarbeiten. Dies ist allerdings nicht opportun. Im derzeitigen Klima der Angst vor explodierenden Gesundheitskosten werden auf Teufel komm raus Studien gesucht, bestellt und verfasst, die einen Zusammenhang zwischen Ärztedichte und Überversorgung belegen sollen. Ob dieser Zusammenhang realiter besteht, ist auf der Grundlage der verwendeten Datenbanken aber gar nicht beurteilbar. Nur verdient man damit eben kein Geld, und so werden weiter fleissig wissenschaftlich wertlose Studien produziert, die nichts nützen, aber viel Schaden anrichten. Das Problem ist also, dass unsere Public-Health-Institute gar nicht die wirtschaftliche Unabhängigkeit haben, wirklich wissenschaftliche Studien zu erstellen, also auch die Datengrundlagen hierfür zu schaffen, anstatt auf eben gerade billig verfügbare Daten zurückzugreifen, die so billig sind wie nutzlos, die Voreingenommenheit der Auftraggeber der Studien aber bestätigen. Wenn man sie entsprechend interpretiert ...

Als gravierendes Beispiel einer Public Health-Studie, die den Anforderungen an die Wissenschaftlichkeit eindeutig und in mehrerer Hinsicht nicht genügt, soll hier die 2011 publizierte Studie von André Busato und Pius Matter mit dem Titel «Geographic variation in the costs of ambulatory care in Switzerland» (5) betrachtet werden. Diese postuliert, ein höheres Angebot medizinischer Leistungen führe zu einer Kostenausweitung. Dabei geht die Arbeit von einer im oben beschriebenen Sinn mangelhaften Datengrundlage aus und nimmt eine mangelhafte Risikostrukturadjustierung vor, indem Faktoren wie Alter und Morbidität ebenso wenig berücksichtigt werden wie der Migration bias (Verzerrungen durch Patienten, die sich nicht am Wohnort behandeln lassen). Die Studie liesse deshalb auch den gegenteiligen Schluss zu, und andere Arbeiten sind unter Einschluss vertiefender Erklärungsvariablen auch zum gegenteiligen Schluss gekommen (6). Die Busato-Studie räumt dies sogar selber ein, postuliert dann aber dennoch, ein höheres Angebot medizinischer Leistungen führe zwingend zu einer Kostenausweitung mit der Tendenz unnötiger Behandlungen, ohne nachvollziehbar zu begründen, wieso.

Die Bedarfsanalysen, die sich auf solche Studien stützen, sind mangelhaft, folglich sind auf ihnen fussende Steuerungsentscheide nicht zielführend: Eine Beschränkung des Angebots (z. B. Zulassungsstopp für neue Ärzte) ändert den Bedarf nicht, nur den Angebotskanal. Die Spezialisten arbeiten in diesem Fall dann einfach vermehrt in den

Ambulatorien der Spitäler – wo sie im Übrigen höhere Anreize zur Überarztung haben und diesbezüglich nicht kontrolliert werden. Mit der Folge einer Kostenausweitung. Der VEMS thematisiert dieses Problem im Dossier «Public Health» (www.vems.ch/public-health) und in seinem Thesenpapier zur Kostenexplosion (7).

Zusammenfassend ist festzustellen, dass eine unwissenschaftliche Art, Forschung zu betreiben, in den Public Health Grundlagen für Steuerungsentscheide liefert. Dies einerseits, indem die Entscheidungsträger sich aus Überzeugung auf diese Grundlagen stützen, und andererseits, weil diese von den Medien so weit verbreitet wurden, dass sie geglaubt werden und als Wahrheiten verinnerlicht sind, weshalb es für die Politik opportun ist, sie zu bedienen. Exemplarisch sehen wir hier, wie problematisch reine Vergleiche von Mittelwerten ohne oder unter ungenügendem Einschluss von Erklärungsvariablen sind. Weitere Beispiele für wenig aussagekräftige Janusstudien sind die Datensammlung der ANQ, die Mortalitätsvergleiche des BAG, die Performance-Vergleiche von comparis.ch und generell Studien, welche lediglich Häufigkeitsunterschiede messen. Bleibt sie auf dem heutigen Stand stehen, sind von der Wissenschaft keine fruchtbaren Beiträge zum Gesundheitswesen zu erwarten. Entwickelt sich unsere Public-Health-Forschung hingegen im hier aufgezeigten Sinne wissenschaftlich weiter, so nähert sie sich der Medizin an und versteht sie folglich auch besser. Und daraus resultieren dann auch wieder Entscheide, die ihre Wirkung nicht verfehlen oder Gegenteiliges bewirken.

HTA, QALY, Missbrauch der Empfehlungen der Fachgesellschaften

Ein zentrales Anliegen der Gesundheitsökonomie ist es, objektiv beurteilen zu können, welche medizinischen Interventionen welchen Nutzen haben und in welchem Verhältnis dieser Nutzen zu den Kosten steht. Die Medizin macht solche Abwägungen auch, wenn sie die Zweckmässigkeit einer medizinischen Intervention beurteilt. Es sind dies jedoch Abwägungen von medizinischem Nutzen mit medizinischen Kosten, mit gesundheitlichen Beeinträchtigungen des Patienten also. Die Medizin definiert Zweckmässigkeit so, dass sie fragt, ob eine Intervention rein medizinisch gesehen nötig ist oder unnötig, während die Ökonomie es so sieht, dass Zweckmässigkeit bei gutem Kosten-Nutzen-Verhältnis im monetären Sinn gegeben sei.

Jede Sichtweise kreiert ihre Realität, und die dominierende Sichtweise auf unser Gesundheitswesen ist die ökonomische. Die Folge davon sind Health Technology Assessments, die Anwendung des QALY-Konzepts innerhalb derselben und der Missbrauch von Empfehlungen der Fachgesellschaften. Das Bestreben ist es hier, die Frage, wann eine bestimmte Behandlung angezeigt ist und wann nicht, zu standardisieren. Health Technology Assessments (HTA) auf der Basis von Quality adjusted life years (QALY's) und die Empfehlungen der Fachgesellschaften sollen nicht Empfehlungen bilden, sondern verbindliche Regeln. Eine Behandlung resultiert aber aus dem Indikationsentscheid. Während die Behandlung objektivierbar und standardisierbar ist, bleibt der Indikationsentscheid immer subjektiv. Es stellt die eigentliche Leistung des Arztes dar, aus einer Vielzahl von Parametern seines Patienten zusammen mit diesem den Indikationsentscheid zu fällen. Hierzu braucht der Arzt einen Ermessensspielraum.

Die medizinische Zweckmässigkeit eines Indikationsentscheids kann nur auf der Ebene der Betrachtung des Einzelfalls beurteilt werden. Der Einbezug des Faktors «Lebensqualität» bei den QALY's zeigt dies exemplarisch: diese kann nur vom Patienten beurteilt werden, nicht von aussen, nicht objektiv, nur subjektiv. Eine Indikation, die bei alleiniger Betrachtung des Standards zweckmässig erscheint, kann im Einzelfall medizinisch unzweckmässig sein und vice versa. Ist es dem Arzt nicht mehr erlaubt, den Raum für seinen Indikationsentscheid auszunützen, zwingt ihn dies erstens, auch Entscheide zu fällen, von denen er weiss, dass sie seinem Patienten schaden. Zweitens ist eine Verfeinerung der Standards so nicht möglich. Dies kann nur auf der Auditebene geleistet werden, indem Abweichungen nicht gebüsst, sondern erhoben werden, um der Komplexität der Sachverhalte gerecht zu werden und tatsächliche Ursachen zu finden.

Auch hier werden also Instrumente angewandt und promoviert, deren effektive Auswirkung eine Kostensteigerung sein dürfte, indem sie geeignet sind, die Qualität und damit die Effizienz des Indikations- und Behandlungsentscheids zu verschlechtern. Diese Zusammenhänge wären vorgängig zu untersuchen, und dies sollte im internationalen wissenschaftlichen Kontext geschehen, unter Berücksichtigung der Eigenheiten unseres Gesundheitswesens. Im Mindesten sind ihre Effekte, werden sie trotz Bedenken eingeführt, laufend zu überprüfen und zu erforschen.

Fehlende oder mangelhafte Begleitforschung

Massnahmen, die eingeführt oder promoviert werden, sollten im Sinne eines Monitorings daraufhin überprüft werden, ob sie ihre Wirkung auch tatsächlich zeigen. Ist dies nicht der Fall oder zeichnen sich sogar gegenteilige Effekte ab, sind die betreffenden Massnahmen umgehend zu überdenken, zu überarbeiten und dann weiter zu überprüfen. Insbesondere bei der gegebenen Situation, dass mangelhafte Studien die Grundlagen für viele dieser Massnahmen bilden, sollte wenigstens genau und kritisch hingeschaut werden, was sie effektiv bewirken. Dass dies ungenügend der Fall ist, soll an zwei Beispielen gezeigt werden.

Im Zuge der Einführung des Fallpauschalensystems in der Schweiz war der Ruf nach einer begleitenden Forschung so laut wie zwecklos; diese starteten per Januar 2012 zu ihrem Blindflug. Als sich dann abzeichnete, dass dieser Blindflug wohl eher zu einer Verteuerung denn zu einer Kosteneindämmung führen würde, wurden Studien in Auftrag gegeben und ausgearbeitet, die den Beweis erbringen sollten, dass sich das System insgesamt lohne oder dass die Mehrheit der Leistungserbringern sich zumindest zufrieden zeigten mit dem neuen System. Eine tatsächliche und wissenschaftliche Begleitforschung indes war dies nicht. Diese würde helfen, Klarheit zu schaffen. Gerade weil die Wahrheiten, die sie zutage fördern würde, möglicherweise nicht opportun sind.

Klar ist: die Fallpauschalen bilden nur gerade 50% der Spitalkosten ab, welche zudem von Spital zu Spital aufgrund ganz unterschiedlicher Grundvoraussetzungen stark variieren. Die Spitäler sind deshalb gezwungen, sich mit Tricks die wirtschaftlichen Bedingungen zu erhalten, damit sie ihren Patientenauftrag weiterhin so gut wie möglich erfüllen können. Manchmal gelingt das besser, manchmal schlechter, ob besser oder schlechter, hängt auch von vielen Faktoren ab, die mit dem Patienten selbst nichts zu tun haben. Die ungenaue Standardisierung schafft auf diese Weise schlechte Anreize: Vorzeitige Entlassungen mit Folgekosten in der Rehabilitation und bei spitalexternen Dienstleistungen (Spitex), Vermeiden unwirtschaftlicher Patienten, Quersubventionierung durch unnötige Eingriffe. Überdies und gravierender: Das Spital erhält Anreize zur Unredlichkeit. Dies schafft eine Kultur der Korruption, die insgesamt wohl kaum kosteneindämmende Wirkung haben dürfte.

Ein weiteres Beispiel ist die Begleitforschung zu Hausarztmodellen. Als der Bundesrat die Förderung dieser Modelle und deren Promotion in der Managed-Care-Vorlage vom Sommer 2012 beschloss (8), wird er sich wohl auf wissenschaftlich fundierte Grundlagen gestützt haben. Würde man annehmen. Tatsächlich legen viele Untersuchungen eher nahe,

dass sich solche Modelle nur dort bewähren, wo es sich um Kontrakte zwischen Arbeitgebern und Ärztenetzwerken handelt. Dies ist auch verständlich, ist dem Arbeitgeber doch an einer raschen, zielführenden Behandlung gelegen, damit der Arbeitnehmer möglichst schnell wieder eingesetzt werden kann. Besteht diese Zusammenarbeit hingegen zwischen Krankenkassen und Ärztenetzwerken, lohnt sich die Sache insgesamt wohl eher nicht, denn hier besteht der Anreiz nicht darin, möglichst gut zu behandeln, sondern möglichst nicht zu behandeln, um Kosten zu sparen.

Ob dies auch in der Schweiz der Fall ist, wäre spätestens heute zu untersuchen. Auf welche Grundlagen stützen sich die Betreiber von Ärztenetzwerken und die Krankenkassen bei der Promovierung derselben, wenn sie vollmundig behaupten, so liesse sich Geld sparen bei gleichzeitiger Wahrung, wenn nicht gar Verbesserung der Qualität? Was die Kosteneinsparung betrifft, legt sich jedenfalls eher Gegenteiliges nahe. Sind die Kantone AG, BE, SG, und ZH, in denen Ärztenetzwerke besonders viele PatientInnen behandeln, auch jene Kantone, in denen die OKP-Kosten abnehmen? Nur dann wäre ja die Einsparung nicht einfach auf andere Institutionen verlagert, sondern tatsächlich realisiert. Wir haben die OKP-Kosten in diesen Kantonen mit den Kosten der übrigen Kantone für die Jahre 2001-2012 verglichen (9). Im Ergebnis finden wir: In den vier Kantonen mit mehr als 10 Managed-Care-Netzwerken im Jahr 2012 steigen die OKP Kosten pro Kopf über die Jahre mehr an als in den Kantonen mit weniger MC-Netzwerken. Wenn die MC-Netzwerke aber über die ganze Behandlungskette Einsparungen zur Folge hätten, müsste die gegenteilige Beobachtung sichtbar werden.

Dieser Vergleich besagt natürlich auf der wissenschaftlichen Ebene nichts, zeigt aber doch auf, dass die Sache in die verkehrte Richtung laufen könnte. Es besteht somit Abklärungsbedarf betreffend die Versorgungsqualität über die ganze Behandlungskette. Es ist nicht zu verstehen, warum solche zentralen Fragen nicht schon längst durch unabhängige wissenschaftliche Studien beantwortet wurden und warum niemand solche Studien in Auftrag gibt – allen voran die Netzwerke selbst, die damit ja ihre Arbeit besser verstehen und verbessern könnten, weshalb sie sie dann auch nicht dauernd zu rechtfertigen bräuchten. Das eigens und sich selbst beurteilende EQAM-Label und die diversen internen Qualitätszirkel der Ärztenetzwerke jedenfalls helfen wenig, in dieser Sache Klarheit zu schaffen und entsprechend zu korrigieren.

Was ist Auditforschung und was bringt sie?

Diese Beispiele zeigen, dass ein reiner Vergleich durchschnittlicher Kosten und Häufigkeiten keine wirklichen Erkenntnisse bringt. Was macht nun die Auditforschung besser? In der Medizin können sehr viele Variablen ein beobachtetes Phänomen begründen. Im Gesundheitswesen kommen zu diesen medizinischen Faktoren ökonomische und politische hinzu, was eine hohe Komplexität schafft. Vergleiche auf der Basis von Prävalenzen und daraus berechneten Mittelwerten zeigen Unterschiede aber lediglich auf, sie begründen sie nicht. Dies kann nur gelingen, wenn die richtigen, die ursächlichen Variablen in den Vergleich mit einbezogen werden. Die Begründungskraft der gewählten Vergleichsvariablen bestimmt die Aussagekraft einer Studie. Diese Variablen zu finden ist allerdings ein hoher Aufwand, der nur auf der Ebene der Einzelbeobachtung mit Audits zu leisten ist. Die Auditforschung also geht über den Mittelwertvergleich hinaus, indem die beobachteten Unterschiede mit stichprobenartigen Audits an Einzelfällen auf ihre Ursache hin untersucht werden, um so begründende Variablen zu finden und das Urteil entsprechend zu verfeinern und gegebenenfalls zu revidieren.

Als Beispiel sei hier exemplarisch die Arbeit «Assessing the Association of Appropriateness of Coronary Revascularization and Clinical Outcomes for Patients With Stable Coronary Artery Disease» (10) aus Kanada genannt, welche die Beziehung zwischen Zweckmässigkeitskriterien für die Behandlung von eingeeengten Herzkranzgefässen (Revaskularisation) und dem Auftreten von Ereignissen bei Patienten mit einer stabilen koronaren Herzkrankheit behandelt. Die Studie untersuchte die Angemessenheit der koronaren Revaskularisation in Ontario, Kanada, und analysierte die Assoziation mit Ereignissen in der Zukunft. Damit will die Studie die rationale Nutzung anhand der Kriterien für eine koronare Revaskularisation überprüfen. Denn es ist nicht bekannt, ob diese Zweckmässigkeits-Richtlinien die Prognose widerspiegeln.

Die Methode beruht auf einer zufällig ausgewählten Gruppe von Patienten aus einer Kohortenstudie, die sämtliche Patienten erfasst, die mit stabiler koronarer Herzkrankheit zwischen April 2006 bis 31. März 2007 mittels Herzkatheter untersucht wurden. Die Eignung für koronare Revaskularisation bei der Koronarangiographie wurde rückwirkend beurteilt unter Verwendung der Zweckmässigkeitskriterien. Im Ergebnis fand die Studie bei 1625 Patienten mit stabiler koronarer Herzkrankheit, dass die Revaskularisation (PTCA oder Bypass-Operation) nur in 69% zweckmässig war. Zweckmässige Eingriffe waren mit einem

statistisch signifikant niedrigeren Risiko für tödlichen oder nicht-tödlichen Herzinfarkt von 39% (95%, CI 0,42 bis 0,88) im Vergleich mit der medikamentösen Therapie verbunden. Sowohl in der Gruppe, in welcher die Zweckmässigkeit unsicher war, als auch in der Gruppe, in welcher ein Eingriff als unzweckmässig beurteilt wurde, war das Auftreten von Herzinfarkt im Verlauf der nächsten drei Jahre gegenüber einer medikamentösen Therapie ohne Eingriff statistisch betrachtet nicht verbessert, aber auch nicht verschlechtert worden. Es zeigte sich indes auch, dass unterlassene Eingriffe mit einer schlechteren Prognose verbunden waren. Bei den medizinisch-konservativ behandelten Personen (38% der gesamten Kohorte) wären 50% dieser Personen mit einer Revaskularisation zweckmässig behandelt worden. Die Autoren folgern, dass die Zweckmässigkeitskriterien sowohl eine Über-, als auch eine Unterversorgung betreffend Revaskularisationen feststellen konnten und dass die Unterversorgung mit erheblichen Risiken verbunden war.

Die Arbeit hat zwar auch ihre Schwächen: So erfolgte beispielsweise die Beurteilung der Zweckmässigkeit mittels standardisiertem Vorgehen und retrospektiv. Trotzdem ist dies eine sehr wichtige Studie, denn sie zeigt auf, wie eine zweckmässige Versorgungsforschung durchgeführt werden könnte. Ein weiteres Beispiel, in welche Richtung diese gehen könnte, kommt aus dem US-Staat Maryland, wo ein Team um Dr. Sam Goldberg, Dr. Roger Leonard und Dr. Marc Mugmon bei der Legislative und den Aufsichtsbehörden hartnäckig für systemexterne peer reviewers geworben hat, um so stichprobenartige externe Beurteilungen aller Koronareingriffe im Staat Maryland zu erhalten. Die Notwendigkeit eines Systems zur vertieften Untersuchung von Erklärungsvariablen einerseits und zur Kontrolle der Ärzte durch die Ärzte selbst ist also auch hier erkannt, und das Beispiel zeigt, dass dies durchaus auch auf der legislativen Ebene durchsetzbar ist.

Zum Schluss, dass dies unabdingbar ist, kommt auch das vom Schweizerischen Nationalfond finanzierte Projekt FP45 mit dem Titel «Probleme des Sozialstaats NF 45: (Projekt-No: 4045-059698), Gesundheitswesen Schweiz: Gibt es Unter- oder Überversorgung? Die Bedeutung von Sozialschicht, Wohnregion, Nationalität, Geschlecht und Versicherungsstatus.» Darin ist nachzulesen, dass regionale Unterschiede eben nicht mit dem Ansatz des Public Health erklärt werden können: «Die regionalen Unterschiede der Inanspruchnahme von Gesundheitsleistungen lassen vermuten, dass ein Teil der Unterschiede bei den kurativen Leistungen nicht ausschliesslich medizinische Gründe – wie unterschiedliche Häufigkeit von Beschwerden, Krankheiten und Unfällen – haben. Sie können

auch durch das Angebot (Ärztedichte) bzw. ein regional unterschiedliches Inanspruchnahmeverhalten mit beeinflusst sein. Somit sind sowohl Massnahmen auf Seite Patientenschaft als auch Ärzteschaft zu postulieren. Aber auch die Forschung ist weiterhin gefordert. Die auffälligen regionalen Unterschiede der Inanspruchnahme sollten vertieft untersucht werden, wenn möglich auf der Basis klinisch abgesicherter Daten.»

Diese Forderung gilt es ernst zu nehmen! Die BERNCA-Studien (11) sind ein Schritt in diese Richtung. Sie sind zwar auch Mittelwertvergleichsstudien, schliessen aber Erklärungsvariablen ein, indem sie Unterschiede betreffend Komplikationen bei hospitalisierten Patienten mit vertieften Fragen besonders beim Pflegepersonal auf ihre Ursache hin untersuchen. Der Mittelwertvergleich schafft eine Hypothese, der Auditprozess überprüft diese anhand der Krankengeschichten und validiert die Hypothese.

Von der Janusforschung zur Auditforschung

Ein Vergleich der beiden Arten zu forschen macht klar, weshalb allein eine Audit- und Zweckmässigkeitsforschung wirklich wissenschaftliche Verbesserungen bringen kann. Janusstudien wählen die Vergleichsvariablen aufgrund falscher Kriterien:

- einfache Verfügbarkeit der Daten
- mangelnde Ressourcen zur vertieften Forschung
- Voreingenommenheit (Variablen werden nach erwarteten Begründungen gewählt, nicht nach beobachteten)

Die Audit-Forschung hingegen geht nicht einfach von einer Datengrundlage aus, die eben gerade verfügbar ist, sondern fordert und schafft diese Grundlage. Dies führt überhaupt erst zu validierbaren Aussagen, denn solange Aussagen ambivalent sind, kann von ihnen auch nicht gesagt werden, ob und unter welchen Bedingungen sie wahr oder falsch sind. Dies erst ermöglicht es, die Studie empirisch zu überprüfen, weiterzuentwickeln und abzuwägen, welche Entscheide aus ihr abgeleitet werden können und vertretbar sind. Wo reine Mittelwertvergleiche allenfalls Hypothesen generieren, vermag eine Auditstudie Beweise zu erbringen. Ein Vergleich auf einen Blick:

Janusforschung und -beurteilung

- Mittelwert- und Prävalenzvergleich
- Datengrundlage gibt Studie vor
- geringe Aussagekraft, ambivalent
- Weiterentwicklung kaum möglich
- empirische Überprüfung kaum möglich
- Hypothesen generierend
- reaktante Reaktion

Auditforschung und -beurteilung

- differenzierter Vergleich
- Studie schreibt Datengrundlage vor
- hohe, eindeutige Aussagekraft möglich
- Weiterentwicklung möglich
- Empirische Überprüfung möglich
- Beweise generierend
- konziliante Reaktion

Steuerungsentscheide, die sich auf Janusstudien und -beurteilungen stützen, sind logisch schwer zu vertreten, weshalb dies emotional geschieht. Ebenso emotional ist die Reaktion: Wer sich ungerecht beurteilt fühlt, reagiert mit Reaktanz – er sucht Wege, den Massnahmen auszuweichen. Die Folge ist oftmals ein Erreichen relativer Einsparungen bei einer effektiven Verteuerung durch Erhöhung der absoluten Kosten. Das Ziel der optimalen Allokation der Mittel kann so nicht erreicht werden. Steuerungsentscheide hingegen, die sich auf Auditforschung und -beurteilung stützen, sind logisch begründbar. Die aus ihnen folgenden Entscheide mögen hart sein, sind aber plausibel, weshalb die entsprechenden Auflagen eher eingehalten werden. Das Ziel der optimalen Allokation der Mittel ist so schrittweise zu erreichen. Dies könnte wie folgt bedeuten:

Für die Wirtschaftlichkeitsverfahren

Medizinische Massnahmen auf ihre medizinische Zweckmässigkeit hin überprüfen kann nur die Ärzteschaft selbst, und dies nur auf der Auditebene, also auf der Ebene von stichprobenartig untersuchten Einzelfällen. Dies könnte ein Dysfunction Board als Institution der FMH mit Sanktionsrecht leisten. Dieses stellt auf der Auditebene fest, ob ein Indikations- und Behandlungsentscheid sachgemäss, ob die Behandlung medizinisch nötig ist. Auffällige Ärzte sollen dem Board gemeldet werden, durch Patientinnen und Patienten, aber auch durch die Ärzte selbst. Das Board trifft dann aufgrund des institutionalisierten Vorgehens entsprechende Abklärungen und führt gegebenenfalls Gespräche mit dem Arzt, dem Team oder dem betreffenden Spital oder Heim. Dies ist auch der Ansatz der Initiative Qualitätsmedizin (IQM) (www.initiative-qualitaetsmedizin.de), in welcher sich Krankenhäuser aus Deutschland, Österreich und

der Schweiz für mehr medizinische Qualität bei der Behandlung ihrer Patienten engagieren. IQM stellt den medizinischen Fachexperten aus den teilnehmenden Krankenhäusern innovative und anwenderfreundliche Instrumente zur Verfügung, die auf einer offenen Fehlerkultur und kollegialer Unterstützung beruhen, wobei nach drei Grundsätzen der IQM-Mitglieder gearbeitet wird:

1. Qualitätsmessung auf Basis von Routinedaten
2. Transparenz der Ergebnisse durch deren Veröffentlichung
3. Qualitätsverbesserungen durch Peer-Review-Verfahren

Der VEMS verfolgt diesen Ansatz im Dossier «Dysfunction Assessment» (www.vems.ch/dysfunction-assessment).

Für die Public-Health-Studien und die Versorgungsforschung

Die Generierung der Datengrundlage ist der zentrale und teuerste Teil einer Studie. Die kleinen, dezentral arbeitenden Public-Health-Institute der Schweiz haben nicht die Mittel, eigene Daten zu erheben, weshalb sie auf bestehende Daten zurückgreifen. Mit der Wahl der Datengrundlage sind die Erkenntnismöglichkeiten aber, wie aufgezeigt, weitgehend bestimmt. Der Bund sollte deshalb den Auftrag für ein Nationales Forschungsprogramm erteilen, welches auf der Auditebene bei den Leistungserbringern vertiefende Erklärungsvariablen erhebt, um diese Datengrundlagen dann den Public-Health-Instituten zur Verfügung zu stellen. Dies kann mit anonymisierten Fragebögen und Stichproben (Zensusmethode, z.B. Tageszensus) geschehen. Der VEMS thematisiert dies im Dossier «Public Health» (www.vems.ch/public-health).

Für HTAs, QALY und die Empfehlungen der Fachgesellschaften

Die gute Absicht, einen Rahmen zu schaffen, um dem Arzt seine Arbeit zu erleichtern, wird in der Hand der lediglich an den Kosten, nicht aber an der medizinischen Zweckmässigkeit eines Indikationsentscheids interessierten Kassen zu einem Kontrollinstrument, das diese Arbeit erschwert oder gar verunmöglicht. Indikationsentscheide, die zweckmässig sind, den Empfehlungen aber zuwiderlaufen,

müssen dem Arzt nicht nur erlaubt sein, sie sollten auf der Auditebene auch erfasst werden, um diese Empfehlungen laufend zu verfeinern. Eine so ausgebaute Datenbasis hilft dem Arzt enorm, indem sie der Ärzteschaft ermöglicht, gemeinsam ihr Wissen zu erweitern. In keiner Weise ist es akzeptabel, dass – was das Swiss Medical Board SMB bei der Ausarbeitung seiner Empfehlungen tut – Experten angefragt und im Editorial der Empfehlung entsprechend aufgeführt werden, wenn deren Input gar nicht oder kaum beachtet wurde, bzw. deren Kritik an der Empfehlung innerhalb derselben kein Platz eingeräumt wird. Das QALY-Konzept als Grundlage für Behandlungsentscheide ist entschieden abzulehnen, was der VEMS in seiner Evaluation «Rationierung, QALY und die mathematische Maschinerie» (1) ausführlich begründet.

Für die Begleitforschung zu den Fallpauschalen und den Ärztenetzwerken

Aufgabe einer Begleitforschung zu den Fallpauschalen und den Ärztenetzwerken ist es, zu evaluieren, ob diese Systeme und Modelle ihre Wirkung tatsächlich zeigen (die Fallpauschalen die einer positiven Beeinflussung des Preises, die Ärztenetzwerke die einer Nachfrageeinschränkung). Mit anderen Worten: ob sie tatsächlich rationalisieren, indem medizinisch eindeutig nicht nötige Leistungen unterbunden werden, oder ob sie nicht vielmehr rationieren, also eindeutig nötige und medizinisch zweckmässige Leistungen unterbinden.

Alsdann sind ihre Effekte auf die tatsächliche Entwicklung der Gesundheitskosten zu untersuchen, was nur die Betrachtung der Spitäler auf der Ebene ihrer behandelten Einzelfälle zu leisten vermag. Dabei könnte sich möglicherweise ergeben, dass das Fallpauschalensystem nur für bestimmte Indikationen eine kostendämmende Wirkung hat (so, wie es in Kanada ursprünglich ja auch entwickelt wurde) und ausgedehnt auf den gesamten Leistungskatalog eines Spitals (wie dies in der Schweiz der Fall ist) insgesamt eher kostentreibend wirkt. Entsprechend wäre das Fallpauschalen-System zu korrigieren und so sukzessive zu verbessern, wozu auch die Verfeinerung der Qualitätsanalysen des BAG einen wesentlichen Beitrag leisten könnte.

Hausarztmodelle mit Budgetmitverantwortung andererseits bergen aufgrund ihres Anreizsystems mit Boni und Bussen die Gefahr verdeckter Rationierung. Hier

kann es mehr als in anderen Versicherungsmodellen zu gefährlicher Unterversorgung mit nicht nur monetären, sondern gesundheitlichen Folgen für die Versicherten kommen. Entsprechend sollten diese Modelle sich im Sinne des Verursacherprinzips einer Kontrolle durch Drittinstanzen (nicht durch von ihnen selbst ins Leben gerufene Qualitätslabels) stellen. Und diese Kontrollen sollten dann eben nicht Nutzwertanalysen auf der Basis von Durchschnittswerten sein. Vielmehr sollten sie die Zweckmässigkeit der medizinischen Entscheide beurteilen, was nur auf der Basis der stichprobenartigen Einzelfallbetrachtung durch medizinische Fachpersonen in Audits möglich ist. Wurden wissentlich oder fahrlässig medizinisch notwendige und zweckmässige Behandlungen und Weiterverweisungen hinausgezögert oder ganz unterbunden, so ist dieses Verhalten in jedem Fall, nicht nur im nachweislichen Schadensfall, zu büssen. Es ist nicht verständlich, weshalb die Betreiber dieser Netzwerke nicht selber entsprechende Strukturen aufgebaut haben, wären diese Audits doch der beste Weg, Rationalisierung nicht nur glaubhaft zu machen, sondern auch effektiv zu leisten. Werden solche Stichproben hingegen verweigert, so legt sich nahe, wie gespart wird: durch verdeckte Rationierung. Diese gehört indes geahndet und gebüsst, nicht mit dem Argument der Vermeidung von unnützen Behandlungen gefördert, wie dies heute geschieht.

Behandlungskosten lassen sich geringfügig optimieren, die Hauptkosten verursachen falsche Indikationsentscheide. Diese werden heute tendenziell gefördert: mit den falschen Beurteilungen der Wirtschaftlichkeitsverfahren der santésuisse, mit einer einseitigen, unwissenschaftlichen Public-Health-Forschung, die Grundlagen liefert für Fehlentscheide, mit den Fallpauschalen und in den Ärztenetzwerken, deren preisdämpfende Effekte immer wieder vollmundig behauptet werden, nicht aber bewiesen oder nur schon nahegelegt wurden. Im Gegenteil.

Dreh- und Angelpunkt Indikationsentscheid

Allen Steuerungsmassnahmen, die sich auf Janusstudien und -vergleiche stützen, ist gemeinsam, dass sie den Indikationsentscheid des Arztes untergraben, indem sie vom Arzt verlangen, in diesen ökonomische Erwägungen mit einzubeziehen. Würde man im

Bildungswesen gleich vorgehen, psychologische Beurteilungstests würden so anlegt, dass man bei Schülern, die zur Entwicklung ihrer Potentiale eine kostenintensive Förderung benötigen, den Schulpsychologen vorschreibe, diese Potentiale möglichst nicht festzustellen. Es ist klar, dass dies nicht zielführend sein kann, weil auf der Basis eines Mittelwertvergleichs nur die Wirtschaftlichkeit, nicht aber die medizinische Zweckmässigkeit einer Behandlung beurteilt wird. Auf der Auditebene hingegen lässt sich diese Zweckmässigkeit messen, weil:

- der Arzt die Hoheit über den Indikationsentscheid behält.
- Fälle und nicht Mittelwerte verglichen werden.
- Ärzte von Ärzten beurteilt werden, nicht von fachfremden Instanzen.
- Wissen für alle generiert wird, nicht nur Macht für die Krankenkassen.

Die Janusforschung ist eine Wirtschaftlichkeitsforschung, die Auditforschung tut den Schritt zur Zweckmässigkeitsforschung und -beurteilung. Die beiden Fragen, ob eine Massnahme nötig ist, um den Patientenauftrag zu erfüllen, und ob sie bezahlbar ist, dürfen nicht vermischt werden. Die erste Frage kann der Arzt beantworten, die zweite der Gesundheitsökonom. Wissensgewinn bezüglich Zweckmässigkeit ist nur möglich, wenn diese Fragen getrennt behandelt werden.

Dass die Medizin sich nicht nach dem Kriterium der Nützlichkeit beurteilt und beurteilen lassen darf, lässt sich an einem Vergleich mit dem Verkehrswesen zeigen: Verkehrskonzepte werden Zweckmässigkeitsanalysen und -beurteilungen unterworfen, nicht Nützlichkeitsanalysen. Dies mit gutem Grund: Ob eine bestimmte Verkehrsverbindung nützlich ist, lässt sich nicht sagen, denn es hängt von vielen Faktoren ab, auf die kein Einfluss genommen werden kann: Wird die Verbindung dann auch tatsächlich genutzt, oder ziehen es die Bürgerinnen und Bürger trotz ÖV vor, die fragliche Strecke mit dem Auto zu fahren? Was nützt es überhaupt, dass die fragliche Strecke befahrbar ist? Wer bestimmt, was an der Befahrbarkeit dieser Strecke nützlich sein soll? Deshalb fragt man hier anders: Wir haben einen Bedarf eruiert, ist das fragliche Verkehrskonzept diesem Bedarf nun angemessen oder nicht, erfüllt es seinen Zweck, ist es zweckmässig?

Aus vergleichbaren Gründen fragt auch die Medizin nach der Zweckmässigkeit. Würde sie sich fragen, ob eine Intervention nützt, sie würde sich anmassen, mit Sicherheit sagen zu können, dass die Behandlung wirken wird. Dies liegt aber nur teilweise in ihrer

Hand. Nimmt der Patient die Medikamente auch tatsächlich? Macht er mit, will er überhaupt geheilt werden? Treten unabsehbare Faktoren ein? Dies alles kann sie nicht absehen, und zu beurteilen, ob die Heilung über sich hinaus einen Nutzen hat, würde bedeuten, ganze Patientengruppen zu diskriminieren. Deshalb fragt sie einzig danach, ob die Behandlung dem Zweck der Heilung oder Linderung angemessen, ob sie nötig, ob sie medizinisch zweckmässig ist, denn diese Frage lässt sich eindeutig beantworten.

Drei Schritte zur Zweckmässigkeitsforschung

In drei Schritten ist ein Wechsel von der Wirtschaftlichkeits- zur Zweckmässigkeitsforschung zu schaffen:

Schritt eins - ehrliche Bestandesaufnahme

Heute wird von Krankenkassen, Politik und Medien für manche Studien und Beurteilungen ein Anspruch geltend gemacht, den diese schon von ihrer Voraussetzung her gar nicht leisten können. Anstatt ehrlich einzugestehen, welchen Evidenzgrad eine Studie überhaupt haben kann, wird dreist behauptet, die in ihr postulierten Hypothesen seien erwiesene Tatsachen, wie dies zum Beispiel von Ignatio Cassis getan wurde (12). Zitat: «Das Konzept der Public Health entspringt einer strengen Denkweise, die auf wissenschaftlichen Methoden gründet. Das Konzept der «Evidence-based Public Health» – also der «auf Wirksamkeitsnachweisen beruhenden Public Health» – lässt keinen Raum für Improvisationen oder Gefühlsentscheidungen und umfasst die Entwicklung, Umsetzung und Bewertung wirksamer Programme und Politiken durch Anwendung wissenschaftlicher Überlegungen und systematische Nutzung von Daten und Programmiermodellen». Dies ist Wunschdenken. Die folgende Systematik soll helfen, zu orten, welche Studien was vermögen und was nicht. Grundsätzlich gibt es zwei Typen von Studien:

Randomisierte Studie

Probanden erhalten auf zufälliger Basis den Wirkstoff und werden mit jenen verglichen, die an dessen Stelle ein Placebo erhielten. Das Ursachen-Wirkungs-Prinzip lässt sich gut untersuchen, vorausgesetzt, die Teilnehmer sind gut vergleichbar.

Beobachtungsstudie

Aus beobachteten Unterschieden werden Schlüsse gezogen, wobei der Erklärungsgehalt davon abhängt, wie gut der Ursachen-Wirkungs-Zusammenhang untersucht wurde. Beispiel: Männer mit Bart sterben im Schnitt früher als Männer ohne. Daraus die Forderung zu folgern, wer einen Bart trage, müsse höhere Krankenkassenprämien bezahlen, wäre dann jedoch zwar logisch, aber falsch. Denn der Ursachen-Wirkungs-Zusammenhang ist unter Einschluss einer weiteren Erklärungsvariable ein anderer: Männer, die in Berufen arbeiten, die schwere körperliche Arbeit verlangen, tragen im Schnitt häufiger einen Bart.

Randomisierte Studien sind für die Pharmazeutische Industrie Pflicht, im Bereich der Public-Health aber oft zu teuer. Hier haben wir drei Typen von Beobachtungsstudien:

Typ C, Mittelwertvergleichsstudien ohne Einschluss von Erklärungsvariablen: Janusstudien ohne wirklichen Erklärungsgehalt, anfällig für Manipulation und gefährlich, weil sie dazu benutzt werden können, Handlungen zu motivieren, die gerade gegenteilige Effekte haben und das Problem folglich nur verstärken.

Typ B, Mittelwertvergleichsstudien mit Einschluss von Erklärungsvariablen: Unterschiede werden erfasst, um mögliche Erklärungsvariablen zu finden, wie dies beispielsweise beim BERNCA-Index geschieht.

Typ A, Audit-Studien (Zweckmässigkeitsforschung):

Vermutete Ursachen werden auf der Ebene der Krankengeschichten systematisch untersucht, um die in Studien vom Typ B vermuteten Erklärungsvariablen zu validieren.

Daraus ergibt sich der Evidenzgrad einer Studie:

Level I: Mehrere randomisierte Studien haben den Ursachen-Wirkungs-Zusammenhang bewiesen, es gibt kaum noch Zweifel daran.

Level IIa: Es gibt noch zu wenige randomisierte Studien, die den Ursachen-Wirkungs-Zusammenhang belegen, aber doch schon so viele, dass mit hoher Wahrscheinlichkeit von diesem ausgegangen werden kann.

Level IIb: Es gibt keine randomisierten Studien, aber die Mehrheit der Experten ist sich einig darin, dass der Ursachen-Wirkungs-Zusammenhang überwiegt.

Level IIc: Es gibt noch zu wenige Beobachtungsdaten, die den Ursachen-Wirkungs-Zusammenhang belegen, von diesem auszugehen ist aber sicher nicht falsch.

Level III: Der Ursachen-Wirkungs-Zusammenhang ist nicht belegt, Handlungen, die sich auf ihn stützen, bringen nicht die gewünschte Wirkung oder schaden gar.

Die heute verfügbaren Beurteilungen und Studien des Gesundheitswesens Schweiz sind mehrheitlich reine Mittelwertvergleichsstudien von Typ C und kommen deshalb nicht über den Evidenzgrad Level III hinaus. Sie sind aus diesem Grund nicht nur unnützlich, sondern gefährlich. Mittelwertvergleichsstudien unter Einschluss von Erklärungsvariablen vom Typ B wie beispielsweise der BERNCA-Index können Evidenzgrad II erreichen. Mit Audit-Studien vom Typ A ist ein Evidenzgrad IIa

erreichbar, und dies sollte das Ziel der Public-Health-Forschung der Schweiz sein. Seien wir also realistisch! Politik, Behörden und Medien sollten vorsichtiger sein bei der Beurteilung des Evidenzgrads von Public Health-Studien und Wirtschaftlichkeitsbeurteilungen medizinischer Leistungen. Über die Resultate solcher Studien zu reden, als handelte es sich um erwiesene Tatsachen, wie sie höchstens randomisierte Studien liefern könnten, ist fahrlässig, irreführend und schädlich. Wenn überhaupt, dann ist mit diesen Beobachtungsstudien höchstens ein Evidenzgrad IIb, im besten Fall IIa erreichbar, in der Regel ist es heute Evidenzgrad III. Um das Ziel Zweckmässigkeitsforschung mit Resultaten vom Evidenzgrad IIa und IIb zu erreichen, muss die Situation, wie sie heute ist, eingestanden werden, damit Schritte zu ihrer Verbesserung eingeleitet werden können.

Schritt zwei – die richtigen Fragen stellen

Es stehen also einige offene Fragen im Raum. Anstatt weiter Behauptungen aufzustellen, sollten an erster Stelle diese Fragen beantwortet werden. Ihre Klärung hilft, zu verhindern, dass Studien, welche zwar interessant, aber mangelhaft sind, entsprechend beurteilt und wahrgenommen werden, um Irritationen und Fehlentscheide zu verhindern. Ihre Klärung hilft, Mängel zu erkennen und Schritte in die Wege zu leiten, um auf institutioneller Ebene schrittweise dem Ziel einer auditbasierten Zweckmässigkeitsforschung mit dem Resultat validierbarer und validierter Aussagen und verantwortungsvoller Entscheidungen mit der Möglichkeit einer Überprüfung ihrer effektiven Wirkung zu gelangen:

Weshalb wird die Beurteilung ärztlicher Leistungen in den Wirtschaftlichkeitsverfahren vom Dachverband der Krankenkassen santésuisse geleistet und nicht von einem Partner, der unvoreingenommen, weil frei von finanziellen Interessen ist? Weshalb werden die wissenschaftlich gut begründeten und im entsprechenden VEMS-Grundlagenpapier (3) ausformulierten Einwände gegen die Methodik und das Verfahren der santésuisse von santésuisse systematisch abgeblockt? Wieso lässt sich die FMH dies gefallen?

Gibt es eine Stelle, die sicherstellt, dass die Studien der Public Health-Institute der Schweiz den Anforderungen der GEP (good epidemiological practice) bezüglich Planung, Durchführung, Analyse und Beurteilung der Resultate (2) genügen?

Wieso sieht es der Ethikrat der öffentlichen Statistik der Schweiz gemäss seinem Tätigkeitsbericht 2011 als seine Aufgabe, dafür zu sorgen, dass auch private Anbieter und Institute den Anforderungen seiner Charta bezüglich der politischen Unabhängigkeit zu genügen haben, fühlt sich dann aber bei einer berechtigten Eingabe bezüglich der Studie eines Public Health-Instituts nicht zuständig, mit der Begründung, die Arbeit sei nicht von einem Auftraggeber, sondern auf eigene Verantwortung erstellt worden (13)?

Was spricht dagegen, dass die eher vielen und eher kleinen Public Health-Institute der Schweiz bei Studien, die geeignet sind, grundlegende Entscheide des Gesundheitswesens massgeblich zu beeinflussen, mit vereinten Kräften arbeiten?

Ist die finanzielle Unabhängigkeit der Public Health-Institute der Schweiz, vornehmlich von Krankenkassen unabhängigen Organisationen, gewährleistet?
Wer kontrolliert das?

Verfügen die Public Health-Institute überhaupt über die Datengrundlage, um aussagekräftige Studien erstellen zu können? Und wenn nicht: gibt es Bestrebungen, dies zu ändern?

Reichen die den Public-Health-Instituten zur Verfügung gestellten Forschungsgelder überhaupt, aussagekräftige Studien zu erstellen?

Gibt es eine Anlaufstelle für Politiker und Medienschaffende, die hilft, die Aussagekraft einer Public Health-Studie zu beurteilen?

Schritt drei – die richtigen Massnahmen einleiten

Aus der Beantwortung obiger Fragen ergeben sich Forderungen, die zwingend und frei von Partikularinteressen sind:

Die Evidenzgrade von Studien und Beurteilungen sind offenzulegen, diesbezüglich falsche Behauptungen sind von einem Ethikrat zu ahnden und mit Bussen zu belegen.

Die Finanzierung von Studien vom Evidenzgrad II ist nur durch Zusammenlegen von Budgets möglich und deshalb gemeinsam sicherzustellen (FMH, Pharmazeutische Industrie, Bund).

Massnahmen zur Förderung der Zweckmässigkeitsforschung auf Auditebene sind zu formulieren, Massnahmenpläne auszuarbeiten, damit dieser Paradigmenwechsel, der in den USA bereits stattgefunden hat, endlich auch in der Schweiz ankommt.

Referenzen

1.
www.physicianprofiling.ch/VEMSRationierung012014.pdf
2.
http://webcast.hrsa.gov/conferences/mchb/mcheipi_2009/communicating_research/Ethical_guidelines/IEA_guidelines.pdf
3.
<http://www.physicianprofiling.ch/WZWStatistik2014.pdf>
4.
<http://www.tagesanzeiger.ch/schweiz/standard/Warum-in-der-Schweiz-haeufiger-operiert-wird/story/15326826>
5.
<http://physicianprofiling.ch/GeographicVariationABusato2011.pdf>
6.
<http://www.physicianprofiling.ch/VersorgungOutcomeStillfried022012.pdf>
7.
<http://www.physicianprofiling.ch/VEMSThesenGesundheitskosten2013.pdf>
8.
http://www.parlament.ch/sites/kb/2000/Kommissionsbericht_SGK-S_00.3566_2001-07-10.pdf
9.
<http://physicianprofiling.ch/VEMSVersorgungssicherheit032014.pdf>
10.
<http://www.ncbi.nlm.nih.gov/pubmed/23062534>
11.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2771456/>
12.
<http://www.physicianprofiling.ch/PHYPublicHealthStellungnahmeCassis.pdf>
13.
<http://www.physicianprofiling.ch/VEMSEthikrat092012.pdf>