



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

Olten, Februar 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. 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 auch Institute, die klinische Forschung betreiben, sich 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) bezüglich Planung, Durchführung, Analyse und Beurteilung der Resultate nicht, ohne dass die Autoren dafür gerügt würden. Im Gegenteil werden mangelhafte Arbeiten 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. Aufgabe der Forschung wäre es aber gerade, diese

Herangehensweise und ihre wissenschaftliche Legitimation kritisch zu hinterfragen, um zu wirkungsvollen Steuerungsentscheiden zu gelangen, insbesondere heute, da sich abzeichnet, dass diese Herangehensweise problematisch ist und die prognostizierten Erfolge weitgehend ausbleiben. Nur eine Finanzierung frei von Interessen kann die momentane Situation in diese Richtung 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 «Kantönlicheist». Eine solche Koordination und die Erstellung nationaler Register wären indes Voraussetzung, um in absehbarer Zeit nur schon die Möglichkeit 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-Paradoxon» 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 bei 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 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 Blinddarm- oder Knieoperation.

Die Steuerungsentscheide unseres Gesundheitswesens auf der Makroebene (Gesundheitspolitik- und -ökonomie) zwingen die Leistungserbringer in ein Verhalten, das zu gegenteiligen Effekten führt: unser Gesundheitswesen wird immer ineffizienter. 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- zur 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 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. 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
Was ist Wissenschaftlichkeit	Seite	12
Janusstudien und -beurteilungen	Seite	13
Auditforschung	Seite	18
Janusforschung versus Auditforschung	Seite	19
Nationalfond-Projekt FP45	Seite	22
Dreh- und Angelpunkt Indikationsentscheid	Seite	23
Drei Schritte zur Zweckmässigkeitsforschung	Seite	24
Referenzen	Seite	30

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. Regulative Massnahmen haben drei Möglichkeiten, die Kosten zu kontrollieren:

Kontrolle des Preises

Regulative Massnahmen, die auf die Kontrolle des Preises zielen, sind die Fallpauschalen, welche die Spitäler in einen Kostenrahmen pro Behandlung zwingen, sowie die Preisverhandlungen mit der Industrie und den Leistungserbringern.

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 kontrolliert 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) versuchen, auf der Basis von Kosten-Nutzen-Analysen Grundlagen für Indikationsentscheide zu liefern, was letztlich darauf hinausläuft, das Angebot der OKP-Leistungen zu kontrollieren und gegebenenfalls zu beschneiden.

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 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 ihre Nachfrage nach medizinischen Leistungen durch einen gesunden Lebensstil gering zu halten versuchen, mit einem Beitrag, beispielsweise an das Fitnessabo, belohnt werden, zielt dies auf die Kontrolle der Nachfrage.

Es ist festzustellen, dass die Bestrebungen, die Kosten unseres Gesundheitswesens in den Griff zu bekommen, sich vor allem auf die Kontrolle des Angebots konzentrieren. Auch haben gewisse preiskontrollierende Massnahmen eine angebotseinschränkende Wirkung. Der Kostenrahmen der Fallpauschalen beispielsweise legt mit dem Preis auch 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. Gleiches gilt für die Wirtschaftlichkeitsverfahren, die sich in der Praxis als reine Kostenkontrolle ausgestalten und den Arzt nötigen, sein Angebot im Rahmen des Mittelwerts eines Vergleichskollektivs zu halten, unabhängig von seinem Patientengut. Überdies sind gewisse nachfrageeinschränkende Massnahmen insofern fraglich, als der Konsument wohl eher selten freiwillig ein solches Versicherungsmodell wählt, sondern in der Regel aus finanzieller Not. Die Folge ist eine Zweiklassenmedizin, die den sozialen Charakter des Gesundheitswesens untergräbt. Insbesondere die Tatsache, dass die Versicherten bis heute nicht genügend über die Konsequenzen dieser Modelle, vor allem über die in diesen Modellen getroffenen Bonusvereinbarungen zwischen Ärztenetzwerken und Krankenkassen informiert worden sind, wirft ein schlechtes Licht auf ihre gerechtigkeitsethische Dimension. Solche nachfrageeinschränkende Modelle sind eindeutig dazu geeignet, den Arzt dazu anzureizen, das Angebot über den Rahmen der mit dem Patientenwunsch einvernehmlichen nachfragekontrollierenden Wirkung hinaus einzuschränken, sprich verdeckte Rationierung zu betreiben.

Wer über Angebotskontrolle redet, sollte das Kind auch beim Namen nennen: Es geht um Rationierung. Diese findet in der Schweiz weitgehend verdeckt und als Rationalisierung deklariert statt. Dabei hat sie, da sie eben gerade nicht zu einer Kürzung des Angebots der Leistungserbringer führt, sondern lediglich zu einer Verschiebung desselben in Richtung

Wertschöpfungsoptimierung für den Leistungserbringer, ironischerweise einen preistreibenden Effekt, was der VEMS in seiner Evaluation «Rationierung, QALY und die mathematische Maschinerie» (1) untersucht. Angebotseinschränkende Massnahmen zur Steuerung unseres Gesundheitswesens sind also mit Bedacht und Nachsicht einzusetzen, ihre Grundlagen haben wissenschaftlich fundiert zu sein, ihre Wirkung muss kritisch überprüft werden, die Entscheide entsprechend überdacht. Dies ist Aufgabe der Public-Health-Wissenschaft.

Andererseits sind die Möglichkeiten preis- und nachfragekontrollierender Massnahmen bei weitem nicht ausgeschöpft. Hier hätte die Ökonomie einen fruchtbaren Beitrag zu leisten, sei es, wenn es darum geht, in zähen Preisverhandlungen mit der Industrie bessere Konditionen auszuhandeln, sei es beim Optimieren von Abläufen, Strukturen und Prozessen der Leistungserbringer zur Optimierung der Wertschöpfung. Auch die Herausforderung, die Nachfrage nach medizinischen Leistungen einzuschränken, hat die Ökonomie noch nicht beherzt angenommen. Ist von dieser Seite wirklich nicht mehr zu erwarten als Modelle, die vordergründig die Nachfrage einschränken, um de facto zu rationieren? Hier ist die Ökonomie bei der Ausarbeitung innovativer Lösungen gefragt, die den Versicherten motivieren, auf den Bezug eindeutig unnötiger medizinischer Leistungen freiwillig zu verzichten oder den Bezug zwar nötiger, aber vermeidbarer medizinischer Leistungen zu verhindern, zum Beispiel durch präventive Massnahmen wie Ändern des Lebensstils. Die Ökonomie bindet ihre Kräfte aber zu sehr, indem sie sich auf die Kontrolle des Angebots konzentriert. Und hier reibt sie sich auf, denn hier reibt sie sich an der Medizin, die nicht mit Kosten-Nutzen-Abwägungen arbeitet, sondern ihre Entscheide nach dem Kriterium der Zweckmässigkeit fällt und prüft.

So lange sich die Ökonomie so in unser Gesundheitswesen einbringt, dass sie bei ihrem Ansinnen, das Angebot zu kontrollieren, den Indikations- und Behandlungsentscheid der Medizin unter ihre Kontrolle bringen will, wird sie mehr schaden als nützen. Ihre Aufgabe ist eine andere, und diese sollte sie endlich wahrnehmen. Der Indikations- und Behandlungsentscheid der Medizin indes ist durch diese selbst zu kontrollieren, seine Qualität zu optimieren. Dies kann nur geschehen, indem seine Zweckmässigkeit überprüft wird, was nur auf der Ebene interinstitutioneller Expertenaudits möglich ist, in welchen nach dem Zufallsprinzip stichprobenartig Fälle auf ihre Zweckmässigkeit hin überprüft werden, um gegebenenfalls in Gesprächen mit den behandelnden Leistungserbringern zu klären, ob eine

unzweckmässige Intervention vorliegt, welche dann verworfen oder im Wiederholungsfall gebüsst wird.

Die Kontrolle der Nachfrage können also nur die Krankenkassen leisten, indem sie entsprechende Modelle und Konzepte ausarbeiten und den Versicherten beliebt machen. Hier haben sie direkten Einfluss, und diesen sollten sie auch wahrnehmen. Übernehmen Politik und Behörden diese Aufgabe, kommt dies einer Bevormundung der Bürger gleich. Die Kontrolle des Angebots hingegen ist Aufgabe von Politik und Behörden, denn hier geht es um die Frage der Rationierung, und diese ist auf der Gesellschaftsebene, nicht im Rahmen wirtschaftlicher Erwägungen eines einzelnen Akteurs des Gesundheitswesens zu verhandeln. Wird diese Aufgabe an die Krankenkassen delegiert, droht eine Steuerung, die einseitig deren wirtschaftliche Interessen bedient und nicht zu einer optimalen Allokation der Mittel auf gesellschaftlicher Ebene führen kann.

Diese Rollenverteilung ist in unserem Gesundheitswesen in jüngster Zeit durcheinandergeraten: Wäre die Managed Care-Vorlage vom Sommer 2012 durchgekommen, hätte dies einem Quasiobligatorium entsprochen. Anstatt den Krankenkassen die Aufgabe zu überlassen, dieses Modell den Konsumenten beliebt zu machen, wollte man, dass die Politik regulierend eingreift; die Politik hat also Krankenkassenaufgaben übernommen. Bei den Wirtschaftlichkeitsverfahren andererseits übernehmen die Krankenkassen eindeutig Behördenaufgaben: Der Dachverband der Schweizer Krankenkassen santésuisse kontrolliert die Ärzte bezüglich Einhaltung der WZW-Regel des Krankenversicherungsgesetzes KVG (wirtschaftlich, zweckmässig, wirksam).

Diese Rollenverschiebung hat natürlich Auswirkungen auf die Qualität der Grundlagen der Steuerungsentscheide unseres Gesundheitswesens. Wenn sich die Machtverhältnisse solcherart verschieben, dass die Krankenkassen die Hoheit innehaben über die Beurteilung von Bedarf und Qualität medizinischer Leistungen, so werden diese einseitig aus deren Sicht definiert auf Wirtschaftlichkeit reduziert. Grundsätzlich ist zu bedenken, wie ein Gesundheitswesen der Zukunft aussehen wird, wenn es nach den Vorstellungen der Versicherungsindustrie ausgestaltet wird, die – notabene – ein Teil der Finanzindustrie ist. Die Vision eines Gesundheitswesens als Spekulationsobjekt, eines Casinospiels mit unserer medizinischen Grundversorgung, sollte uns jedenfalls einkehren und den gewählten Weg überdenken lassen.

Die heute bereits erfahrbare Folge dieser Tendenz: Die Bedarfs- und Qualitätskontrolle ist zu einer reinen Kostenkontrolle geworden. Dabei werden die Anforderungen an die Wissenschaftlichkeit, wie sie die Auflagen der GEP (good epidemiological practice) bezüglich Planung, Durchführung, Analyse und Beurteilung der Resultate (2) vorsehen, teilweise in mehreren Punkten missachtet, ohne dass dies von den zuständigen Stellen gerügt wird. Namentlich sind dies folgende Stellen:

- Schweizerische Akademie der Medizinischen Wissenschaften SAMW
- Akademien der Wissenschaften Schweiz
- Bundesamt für Gesundheit BAG
- Eidgenössisches Departement des Innern EDI
- Ethikrat der öffentlichen Statistik der Schweiz

Der VEMS hat Regelverstösse immer wieder diesen Stellen gemeldet und immer wieder erfahren müssen, dass sie sich nicht zuständig fühlen, bzw. keinen Handlungsbedarf sehen, auch dann nicht, wenn unabhängige, renommierte Gutachter die monierten Verstösse belegen. Damit ist eine Situation geschaffen, die über kurz oder lang zu einer einseitigen Ausgestaltung des Gesundheitswesens nach den Interessen der Krankenkassen führen muss. Dem entgegenwirken kann nur eine unabhängige Wissenschaft, welche es folglich mit allem Nachdruck zu fordern und durchzusetzen gilt. Wie dies zu leisten ist, soll im Folgenden aufgezeigt werden.

Was ist Wissenschaftlichkeit?

Das Ziel der Wissenschaft ist es, Dinge zu erklären, um daraus Handlungen abzuleiten, welche dem Menschen nützen. Keine wissenschaftliche Theorie kann für sich beanspruchen, die 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 dies nicht der Fall 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, stimmen sie nicht. Die Relativitätstheorie vermag diese Phänomene zu erklären, versagt jedoch 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. Das Resultat empirischer Forschung sind Erkenntnisse, die abzuwägen helfen, welche Handlungen sinnvoll sind, weil sie mit sehr hoher Wahrscheinlichkeit die gewünschten Folgen haben werden. Ist die Forschung nicht empirisch, so ist sie keine Hilfe beim Abwägen, lässt sich aber sehr gut zu manipulativen Zwecken verwenden. Eine wissenschaftliche Arbeit hat folgende Kriterien 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 offen und 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.

Sind diese Voraussetzungen nicht gegeben, können höchstens Hypothesen generiert werden, und diese müssen dann als solche bezeichnet werden.

Janusstudien und -beurteilungen

Viele Arbeiten, auf die sich Steuerungsentscheide des Schweizer Gesundheitswesens stützen, und viele Methoden, mit denen operativ bei der Durchsetzung dieser Entscheide gearbeitet wird, genügen den Auflagen an die Wissenschaftlichkeit nicht. Ihnen allen ist gemeinsam, dass sie auf der Basis von Mittelwertvergleichen Schlussfolgerungen ziehen, 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 diese Studien und Beurteilungen aus diesem Grund Janusstudien und -beurteilungen. Beispiele hierzu sind:

Wirtschaftlichkeitsverfahren

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, weil zum Beispiel Kardiologen und Internisten eine Vergleichsgruppe bilden, obwohl sie sich stark unterscheiden können (Kardiologe höhere Behandlungskosten, Internist höhere Medikamentenkosten). Wichtige Variablen mit hoher Erklärungskraft wie die Morbidität andererseits werden nicht erhoben. Die gewählten Vergleichsvariablen erklären 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 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 nicht zu erbringen, sondern nur auf der Ebene der Krankengeschichten des angeschuldigten Arztes, was aufgrund des Aufwands kaum je getan wurde. Die Methode zeigt sowohl Nichtüberarztter als Überarztter an, als auch Überarztter als Nichtüberarztter. 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 Aggregation) und zweitens von der falschen Stelle: der *santésuisse*, die nicht objektiv sein kann, da sie finanzielle Partikularinteressen hat. Das Resultat sind 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 Überarztter 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. Der VEMS behandelt dieses Problem im Dossier «Wirtschaftlichkeitsverfahren» (www.vems.ch/projekt-wzw) und in seinem Grundlagenpapier, welches er im Rahmen seiner Arbeit in der Arbeitsgruppe WZW der FMH ausgearbeitet hat (3).

HTA, QALY, Missbrauch der Empfehlungen der Fachgesellschaften

Die Frage, wann eine bestimmte Behandlung angezeigt ist und wann nicht, soll standardisiert werden. 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, ist 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 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 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.

Studien der Fallpauschalen

Spitäler rechnen nach Fällen ab, für welche schweizweit einheitliche Tarife gelten. Das Argument, so liessen sich Kosten sparen, wurde allerdings nie wissenschaftlich belegt und wird inzwischen auch kaum mehr vertreten. Denn die Fallpauschalen bilden nur gerade 50% der Spitalkosten ab, welche zudem von Spital zu Spital aufgrund ganz unterschiedlicher Grundvoraussetzungen stark variieren. Die Spitäler sind gezwungen, sich mit Tricks die wirtschaftlichen Bedingungen zu erhalten, um ihren Patientenauftrag so gut als möglich zu erfüllen. 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 Beurteilung schafft 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.

Public Health-Studien und Versorgungsforschung

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» betrachtet werden. Diese postuliert, ein höheres Angebot medizinischer Leistungen führe zu einer Kostenausweitung. Dabei geht die Arbeit von einer 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. Die Studie selbst räumt dies sogar ein, postuliert dann aber dennoch, ein höheres Angebot medizinischer Leistungen führe zu einer Kostenausweitung mit der Tendenz unnötiger Behandlungen, ohne zu begründen, wieso. Die Bedarfsanalyse dieser Studie ist also mangelhaft, folglich der auf ihr fussende Steuerungsentscheid nicht zielführend: Die Beschränkung des Angebots (Zulassungsstopp) ändert den Bedarf nicht, nur den Angebotskanal. Die Spezialisten arbeiten vermehrt in den Ambulatorien der Spitäler, wo sie 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).

Weitere Beispiele für wenig aussagekräftige Janusstudien sind:

- die Datensammlung der ANQ
- die Mortalitätsvergleiche des BAG
- die Performance-Vergleiche von comparis.ch
- generell Studien, welche lediglich Häufigkeitsunterschiede messen

Die Tatsache, dass diese Arbeiten und Methoden aufgrund ihrer mangelhaften Aussagekraft zur Untermauerung ganz unterschiedlicher Behauptungen und Forderungen verwendet werden können, birgt hohes Potential des Missbrauchs. Janusstudien und -beurteilungen können je nach Bedarf als wissenschaftliche Begründung für Sanktionen und Forderungen missbraucht werden, deren Gründe anderer Natur sind und die nicht das Ziel der optimalen Allokation der Mittel verfolgen, sondern Partikularinteressen. Dass diese Studien und

Beurteilungen die Grundlage der Steuerung unseres Gesundheitswesens bilden, ist ein Missstand, der zwingend und dringend zu bekämpfen ist.

Auditforschung

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 geht über den Mittelwertvergleich hinaus, indem die beobachteten Unterschiede mit stichprobenartigen Audits an Einzelfällen auf ihre Ursache hin untersucht werden.

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» aus Kanada genannt, welche die Beziehung zwischen Zweckmässigkeitskriterien für die Behandlung von eingengten 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, wo die Zweckmässigkeit unsicher war, als auch in der Gruppe, wo 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 Schwächen: So erfolgte die Beurteilung der Zweckmässigkeit mittels standardisiertem Vorgehen und retrospektiv. Trotzdem ist dies eine sehr wichtige Studie, denn sie zeigt auf, wie künftig eine zweckmässige Versorgungsforschung durchgeführt werden könnte.

Ein weiteres Beispiel, in welche Richtung eine zweckmässige Versorgungsforschung 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.

Ein weiteres Problem der Datenbasen vieler Studien und Beurteilungen löst allerdings auch die Audit-Forschung nicht: das Problem des Aggregation Bias. Darunter versteht man eine Verzerrung, die entsteht, weil in einer Datenbasis Subgruppen gebildet wurden, die mit unterschiedlicher Gewichtung zur Gesamtbeurteilung beitragen und diese so verfälschen. Das Resultat kann dann beispielsweise sein, dass falsch angenommen wird, das Alter erkläre die Gesundheitskosten zu 93%, während es dies ohne diese Verzerrung und tatsächlich lediglich zu 10% tut. Mit diesem Problem, das im wissenschaftlichen Diskurs als Simpson-Paradox erkannt ist, setzt sich der VEMS in seiner «Charta Ethik und Medizin» (4) auseinander. Für

die nachfolgenden Gedanken ist es allerdings nicht wesentlich und soll hier deshalb nicht eingehend diskutiert werden.

Janusforschung versus Auditforschung

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 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. Und 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.

Janusforschung und -beurteilung

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

Auditforschung und -beurteilung

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

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 sie eher eingehalten werden. Das Ziel der optimalen Allokation der Mittel ist so schrittweise zu erreichen. Dies könnte konkret wie folgt aussehen:

Wirtschaftlichkeitsverfahren

Medizinische Massnahmen auf ihre Zweckmässigkeit hin prüfen kann nur die Ärzteschaft selbst auf der Auditebene. Dies könnte ein Dysfunction Board als Institution der FMH mit Sanktionsrecht leisten. Dieses stellt auf der Auditebene fest, ob ein Indikations- und Behandlungsentscheid sachgemäss 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. Der VEMS verfolgt diesen Ansatz im Dossier «Dysfunction Assessment» (www.vems.ch/dysfunction-assessment).

HTA, QALY's, 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 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.

Fallpauschalen

Behandlungskosten lassen sich geringfügig optimieren, die Hauptkosten verursachen aber falsche Indikationsentscheide. Und diese werden mit den Fallpauschalen

gefördert. Die Qualität des Indikationsentscheids ist von zentraler Bedeutung und kann nur auf der Auditebene verbessert werden, zum Beispiel mit einem Peer Review System als externem und freiwilligem System bei der Indikationsstellung für Abklärungen, Behandlungen und Eingriffe, indem beispielsweise Spitäler untereinander die Zweckmässigkeit von Entscheiden anhand zufälliger Reviews (Audits) kontrollieren. Der VEMS verfolgt diesen Ansatz im Dossier «Dysfunction Assessment» (www.vems.ch/dysfunction-assessment).

Public-Health-Studien

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 weitgehend bestimmt. Wurden in ihr wichtige Erklärungsvariablen nicht erhoben, ist auf dieser Basis nichts wirklich zu erklären. Da die Studie dennoch eine Aussage machen muss, macht sie eine janusartige, was zu Missbrauch und gravierenden Fehlentscheiden führen kann. Der Bund sollte deshalb den Auftrag für ein Nationales Forschungsprogramm erteilen, welches auf der Auditebene bei den Leistungserbringern vertiefende Erklärungsvariablen erhebt, um diese Datengrundlage 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).

Nationalfond-Projekt FP45

Zu diesen Schlüssen 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 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.

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 Zweckmässigkeit einer Behandlung beurteilt wird. Auf der Auditebene hingegen lässt sich die Zweckmässigkeit einer medizinischen Massnahme 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 und ob sie bezahlbar ist, dürfen nicht vermischt werden. Die erste Frage kann der Arzt beantworten, die zweite der Gesundheitsökonom. Werden diese beiden Fragen vermischt, ist kein Wissensgewinn bezüglich Zweckmässigkeit möglich.

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 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 Bestandaufnahme

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. 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 A, 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 C, 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 IIc: 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 IIa: 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 Public Health-Studien der Schweiz sind reine Mittelwertvergleichsstudien von Typ A 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 IIc erreichen. Mit Audit-Studien vom Typ C ist ein Evidenzgrad IIb 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. 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 IIc 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, um Schritte zu ihrer Verbesserung einzuleiten.

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:

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 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?

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 IIb und IIc 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/mchepi_2009/communicating_research/Ethical_guidelines/IEA_guidelines.pdf
3.
<http://www.physicianprofiling.ch/WZWStatistik2014.pdf>
4.
<http://physicianprofiling.ch/VEMSEthikCarta2014.pdf>