

**Prüfung der Wirtschaftlichkeitsprüfung von santésuisse:  
Rationale und Fragestellung**  
(Autor der Fragen: Dr. Michel Romanens  
Verein Ethik und Medizin Schweiz)

**Kurzgutachten zu den Fragen des Modul 1 (16 Fragen)**

**Erstellt von** HDoz. Dr. rer. nat. habil. **Walter Warmuth** (Gesundheitsforen Leipzig) unter  
Beteiligung von **Helmut Dahl** und Dr. **David Matusiewicz** (ForBiG Forschungsnahe  
Beratungsgesellschaft im Gesundheitswesen)

Zitate habe ich sprachlich nicht verändert, wie auch die Fragen.

Es geht im Gutachten wesentlich um die Überprüfung der Anwendbarkeit zumindest der Nützlichkeit<sup>1</sup> mathematischer Modelle in Gesundheitswesen.

Behinderungen ergeben sich daraus, dass

1. die empirische Sozialforschung (auch Sozialstatistik) gewöhnlich von empirischen Mittelwerten von Anzahlen, Dauern und Kosten ausgeht und annimmt, dass sich Beobachtungswerte, die um diese schwanken, „gleichberechtigt“ ausgleichen,
2. der Fundus der mathematischen Methodenstatistik fast ausschließlich ein mathematisch normalverteiltes Schwanken um Mittelwerte bedingt und es so zu einem überirdischen Konsens angewandter mathematischer Statistik und empirischer Sozialforschung kommt,
3. risikomathematische Herangehensweisen (versicherungsmathematische, finanzmathematische, extremwerttheoretische, ...) Sozialforschern und Statistikern<sup>2</sup> weitgehend unbekannt sind.

Problematisch bleibt, dass ein Gesundheitswesen konstituierendes System als Organisationssystem nicht von der Finanzierung eines Gesundheitswesens getrennt wird. Versorgung und Versorgungsforschung könnten nach der Theorie sozialer Systeme ihre Heimat im Organisationssystem haben. Gesundheitsökonomie meint sich mit der Finanzierung sozialer Systeme abgeben zu müssen und zu können, ohne die unterschiedliche bis fehlende Ausdifferenzierung zu beachten.

**Frage 1.** *Wie beurteilen Sie einen fixen RSS-Index > 130% betreffend dem Beweisgrad für Überarztung generell (also für alle Aerzte gleichbleibend)? (Anm.: hier geht es darum, ob eine fixe Grenze über alle Facharzttrichtungen hinweg aus Ihrer Sicht je zielführend war, mithin, ob ganz grundsätzlich die Methode überhaupt je als juristische Beweisgrenze validiert und damit verwendbar war). Siehe auch link A (ZSRKanton.pdf) und link B (ZSRKanton2.pdf).*

Die seit den 1970er Jahren in der Schweiz praktizierte „statistische Durchschnittsmethode“ wird als zweckrational ausgegeben, sie ist es nicht. Sie orientiert sich am Mittelwert und unterstellt, dass es Abweichungen gibt, nach oben und nach unten, die sich letztendlich ausgleichen. Einer Asymmetrie des Ausgleiches meint man mit einem Rechnungsstellerstatistik-Index zu begegnen, der in einer Vorstellung von einem Konfidenzintervall hoher Sicherheit bis zu 130 % des Indexes reicht.

Kostenverteilungen in Gesundheitswesen haben eine spezifische Gestalt. Zunächst gebe ich ein Beispiel nach der Arztgruppe an.

---

<sup>1</sup> „Alle mathematischen Modelle sind falsch, manche sind nützlich.“ (Sachs, L., Statistische Methoden, Springer-Verlag, Berlin 1993, S. 146.)

<sup>2</sup> Die ETH verfügt auf diesem Gebiet über besondere Traditionen; ich möchte auf den Schweizer Paul Embrechts und den Deutschen Hans Föllmer verweisen. Meine liebe Frau Elke hat mit dem Schweizer ETH-Absolventen Moritz Adelmeyer das für die Lehre in der Schweiz und Deutschland „gültige“ Buch „Finanzmathematik für Einsteiger“ geschrieben.



Abbildung 1. Heavy-tail bei der Gestation (Mittelwert der Kosten 4.455,28 €, Standardabweichung 4.355,72 rund 100% des Mittelwertes), 62 % der Fälle liegen unter dem Mittelwert, 71 % der Fälle liegen unter dem Mittelwert + 10% des Mittelwertes), veröffentlicht in Degenkolbe, B.; Gette, E.; Seiffert, A.; Warmuth, W.: Justierung durch Vergleich, Predictive Modeling News, (2011)1. S. 12-14. Die Anwendung des Modells der Normalverteilung schließt den Kostenausgleich der Kosten bis ca. 20.000 € durch negative Kosten ab ca. -10.000 € ein. Die hohen Kosten sind medizinischen Notwendigkeiten geschuldet, die Konzentration vor dem Mittelwert resultiert aus der extremen Überarztung durch medizinisch nicht notwendige Kaiserschnitte (mehr als das 10-fache des medizinisch Notwendigen). Bezüglich dieser Kostengruppe gleichen sich das deutsche und das schweizerische Gesundheitswesen schon seit etwa fünf Jahren.

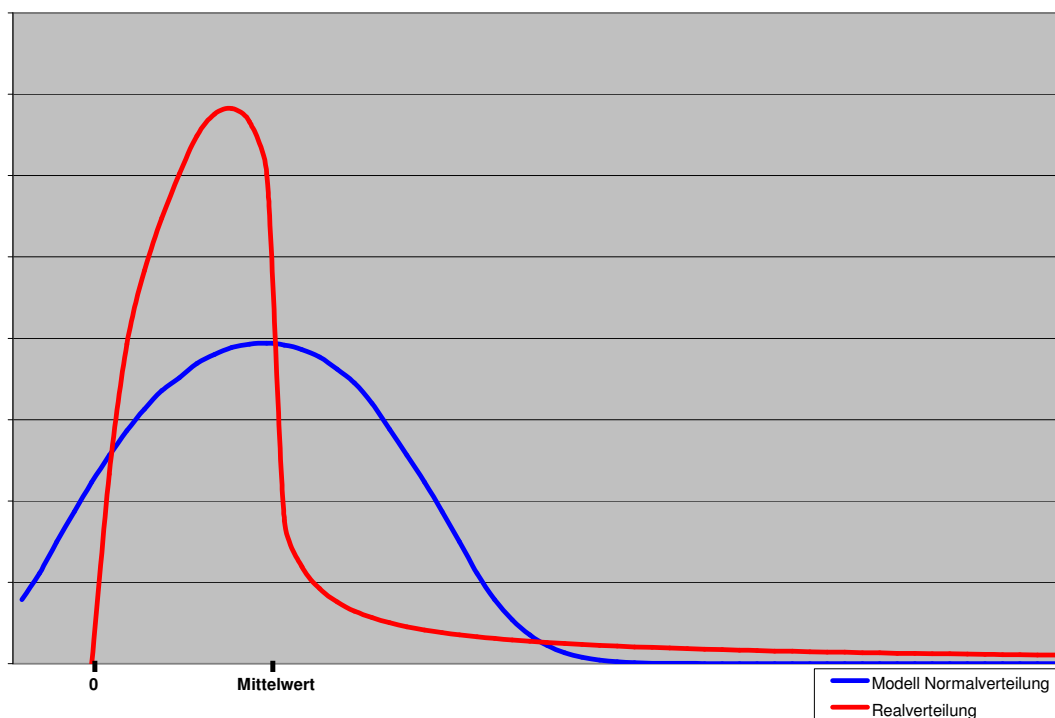
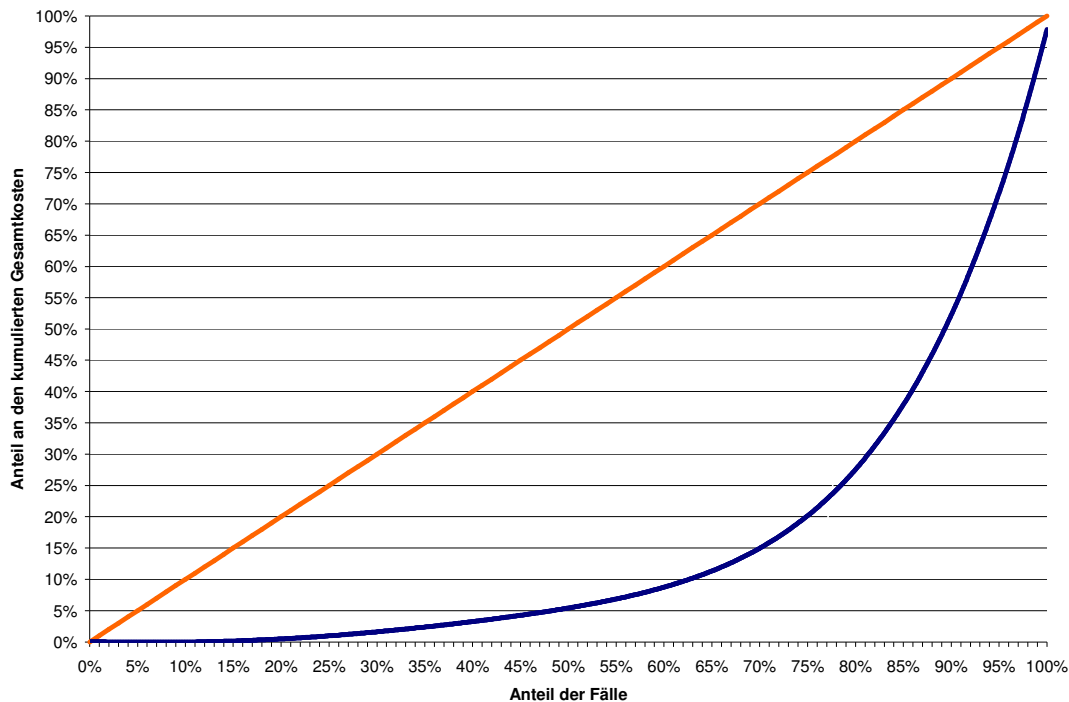


Abbildung 2. Behandlungskostenverteilungen in Gesundheitswesen (aus Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.)



**Abbildung 3.** Konzentration der Behandlungskosten in Gesundheitswesen (aus Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.)

Da die Kosten in der Abbildung 3 der Größe nach sortiert sind, entspricht der RSS<sup>3</sup>-Index von 130 % gerade 65 % der kumulierten Kosten, das sind etwa 94 % der gesamten Fälle. Es ist also „normal“, dass 6 % der Fälle noch 35 % der Kosten ausmachen. Kein einziger der 6 % teuersten Fälle steht a priori für eine Überarztung.

*Überarztung im niedrigen Indexbereich?* Von einer Überarztung in erster Linie kann ausgegangen werden, wenn 94 % der kleinsten Fallkosten kumuliert merklich die prognostizierten 65 % der kumulierten Kosten übersteigen. Es gibt Methoden, die eine auffällige Konzentration im niedrigen Indexbereich belegen. Diese Methoden habe ich in großen Projekten im Zusammenhang mit der Betrugserkennung bei Versicherern (Allianz Versicherungs-AG) und Gewerbesteuererklärungen (Landesamt für Steuern Bayern) benutzt. Es wird dabei jedoch nur die Vermutung des Faktes gestärkt, der einzelne Fall oder die einzelnen Fälle sind nicht belegbar, da sie in der Masse maskiert sind. Veröffentlicht sind viele Methoden in Warmuth, W.: Und sie betrügt uns doch, Versicherungswirtschaft, 62(2007)3. S. 200-203 und Warmuth, E.; Warmuth, W.: Was haben Aktienkurse und Fibonacci-Zahlen gemeinsam? Der Mathematikunterricht, 58(2012)1. S. 34-38 (mit weiteren ETH-Zitaten zu Methoden bei der Betrugserkennung).

*Überarztung bei höheren Kosten?* Eine Überarztung in zweiter Linie bei den 6 % höheren Kosten ablesen zu können, halte ich für schwierig bis unmöglich. Diese Kosten sind nicht unbedingt allen Arztgruppen, allen Regionen, ... zuordenbar. Diese höheren Kosten sind teilweise managebar, teilweise poolbar, ganz sicher aber rückversicherbar u. U. sind diese Risiken verbriefbar.

Ein Realkostenausgleich für die 94 % der Fälle ist äquivalent zu einem Ausgleich der Kosten für die 94 % nur zu je 65 % von den mittleren Kosten. Das gilt aber nur in Bezug auf große Bereiche des Anfallens der Kosten, wo auch die höheren Kosten bereichsanteilig vorkommen. Vermutlich ist das unterhalb der Schweiz nicht praktikierbar, für den einzelnen Arzt gilt das keinesfalls. Die Abbildung 3 variiert nur leicht bei Abrechnungen der Leistungserbringer gegenüber deutschen Krankenkassen,

<sup>3</sup> Rechnungsstellerstatistik RSS

- wenn auf Versichertengruppen fokussiert wird: Bestandsversicherte, Kündigung, Verstorbene, Neugeborene, Neuversicherte (ein Jahr oder älter)
- wenn auf Leistungsbereiche der Krankenversicherer fokussiert wird: stationäre Leistungen, ambulante Leistungen, Arzneimittel über Apothekenabrechnungszentren, sonstige Leistungen (z. B. Heil- und Hilfsmittel, aber auch Arzneimittel bei z. B. Enzyersatztherapien, außerbudgetäre Leistungen, Rehabilitations-Leistungen, Pflege-Leistungen nach dem SGB V, ...), zahnärztliche Leistungen, Arbeitsunfähigkeits-, Entgeltfortzahlungs- und Krankengeldleistungen (in Deutschland werden diese versicherungsfremden Leistungen aus den Krankenversicherungsbeiträgen quersubventioniert)

Kassenbezogen spielt dabei die Alters- und Geschlechtsstruktur, die Beschäftigtenstruktur (dabei nur wenig die Erwerbminderungsrentenstruktur), die Dauermedikations- und Nachbehandlungsstruktur eine Rolle.

Eine fixe Grenze über alle Facharzttrichtungen hinweg ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen war von Anfang an nicht zielführend, sie kann so auch nicht evaluiert worden sein. Wenn die spezifische Behandlungskostenstruktur auch weiterhin nicht einbezogen werden soll, dann ist es unerheblich, ob eine facharztspezifische Verfeinerung gewählt wird oder nicht. Die hohen Kosten (6 % der Fälle mit 35 % der Kosten) lassen sich kaum auf die Arzttrichtungen oder auf Ärztegruppen mit spezifischer Behandlungskostenstruktur aufteilen, auf einzelne Ärzte überhaupt nicht. Es ist deshalb auch nicht mit arztgruppenspezifischen Absenkungen für die abrechenbaren Fälle zur Stützung der Fälle oberhalb der „Toleranzschwelle“ zu rechnen. Die Ungleichbehandlung der Fachrichtungen, Arztgruppen und Ärzte ist beträchtlich. Wenn die Ärzte zur Selbstregulierung aufgefordert werden, dann können sie das nur, indem sich z. B. Ärzte verschiedener Arzttrichtungen zusammenschließen, indem z. B. Leistungsangebote ein- oder ausgeschlossen werden, indem z. B. nur gut abrechenbare Versichertenrisiken zur Behandlung selektiert werden usw. Unter Berücksichtigung der spezifischen Behandlungskostenstruktur sind sogar einige Spezifikationen unterhalb der Arzttrichtungen sinnvoll. Sie können den Ausrichtungscharakter der Arztgruppe bzw. die Bestandszusammensetzung der Patienten bzw. Versicherten berücksichtigen, die hohen Kosten bleiben aber auch hier außen vor.

**Frage 2.** *Halten Sie, und falls ja, warum, eine Anpassung der Beweisgrenze (nach oben oder nach unten) für eine unabdingbare Vorbedingung (Flexibilisierung der Beweisgrenze) für ein sachgerechtes Beurteilungsverfahren? (Anm.: hier geht es um die Frage, anschliessend an Frage 1, welche Argumente ganz grundsätzlich dafür sprechen, dass eine Flexibilisierung eine Grundvoraussetzung für ein korrektes Verfahren darstellt, und zwar unabhängig davon, ob mit oder ohne Vergleichsgruppe gerechnet wird).*

Eine fixe obere Grenze über alle Facharzttrichtungen hinweg ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen war von Anfang an nicht zielführend.

Wird an der nicht sinnvollen Interpretation „Beweisgrenze“ ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen festgehalten, dann erscheint das Finden von oberen Grenzen zumindest nach Arzttrichtungen als Verbesserung, es ist für diesen Zweck notwendig. Die „Toleranzschwellen“ in den einzelnen Arzttrichtungen stellen u. U. ein Konstrukt dar, das mit einer großen Menge von Abrechnungen problemlos umgehen kann. Von den normalen hohen Kosten sind die Arzttrichtungen aber nur sehr vage vorhersehbar betroffen, die Volatilität nimmt mit dem Abnehmen der Anzahl der Ärzte sehr stark zu. Der einzelne Arzt ist schicksalhaft „Toleranzgrenzen“ ausgeliefert.

Auch die Ergänzung durch untere Grenzen ist methodisch sinnvoll, aus der Sicht der „Steuerungswirkung“ sollte das aber wohlüberlegt werden. Aus den Erfahrungen bei der Weiterentwicklung des DRG-Systems gibt es sehr wohl Anlässe, wo auf eine untere Grenze zu verzichten ist. Mit einer Wahl einer unteren Grenze werden bei den dominanten symptomatischen Anlässen für Arztkontakte m. M. n. die Aufwendungen bei der Diagnostik zunehmen.

Ohne Berücksichtigung der spezifischen Behandlungskostenstruktur in Gesundheitswesen ist ein „korrektes“ Verfahren unmöglich, anders ausgedrückt: ein Verteilmechanismus, der ausschließlich auf das Kompensieren von Überschreitungen durch das Unterschreiten vom Mittelwert setzt, ist für einen Kostenausgleich in Gesundheitswesen nicht nützlich.<sup>4</sup>

Ein kleiner Index deutet auf eine „selektive“ Versorgung hin. Das kann dann ein Indiz für eine Unterversorgung sein, mehr noch wird es sich aber um eine „nicht zu Ende erfolgte“ Versorgung bei diesem Arzt handeln, die Weiterversorgung erfolgt bei einem anderen Arzt. Eine vorsichtige Annäherung an die Bewertung solcher Situationen setzt das Liefern von patientenbezogenen Abrechnungen an die santésuisse voraus. Aber auch dann würde versteckt die Vorstellung von der Möglichkeit einer generellen „korrekten“ Behandlung eine Rolle spielen.

**Frage 3.** *Ist die Prüfung sämtlicher Aerzte zielführend oder wäre auch eine zufällige Stichprobe (z.B. 5% pro Facharztgruppe) ausreichend? (Anm.: hier geht es um die Frage, ob ein geringerer Screening-Aufwand Gelder frei machen würde, welche für eine verbesserte Screening Methode verwendet werden könnten bzw. was das Rationale in Deutschland ist, pro Jahr nur 5% pro Facharztgruppe zu screenen).*

Zufällige Referenzstichproben sind ausreichend. Dabei bezieht sich die Referenz aber nicht allein auf Arztgruppen, sondern auch auf Konstellationen von Arztgemeinschaften, die mit einer gruppenübergreifenden Ausstattung (z. B. zur Diagnostik) versehen sind (Nutzensvorteile vs. Verwertungsdruck), auf Patientengruppen, auf Regionen – ganz wichtig ist aber dabei auch die Abbildung des Ausmaßes, das durch die Patienten bedingt ist, die zum Überschreiten der „Toleranzschwellen“ bei den Ärzten führen. Solche Referenzkonstellationen sind sicherlich vereinbar. Hierbei ist eine Zufallsstichprobe in dem Sinne zu stratifizieren, dass über der „Toleranzschwelle“ keine „realen“ Ärzte in der Referenz verbleiben, Referenzkonstellationen übergreifend werden z. B. virtuelle Ausgleichpools gebildet.

Wie bei den DRG-Weiterentwicklungen könnte eine Referenzkonstellation für ein Kalenderjahr gelten. Die Zufallsauswahl in den Schichten ist Bedingung.

**Frage 4.** *Ist die ANOVA Korrektur anhand der ZSR<sup>5</sup> Variable eine ausreichend erklärende Variable im ANOVA Modell betreffend der Problematik des „tautology bias“ und wie ist diese Variable in der Literatur begründet, bzw. was beschreibt sie eigentlich? (Anm.: anhand der ZSR Nummer ist jeder Arzt mit seinen Kosten identifizierbar. In der ANOVA Korrektur beginnt das Modell mit der ZSR Nummer als fixed effect variable zu rechnen und ist nichts anderes als der RSS<sup>6</sup>-Index des Arztes. In der Gleichung  $RSS\text{-Index} = ANOVA[ZSR]$  ist der RSS-Index enthalten, man erklärt also die Arztkosten durch die Arztkosten, was wir als Tautology Bias bezeichnen. Wichtig zu wissen: die ZSR Variable erklärt als fixed effect variable je nach Facharzttrichtung ebenfalls ca. 40% der Arztkosten<sup>5</sup>).*

Mittelwertunterschiede in den ärztlichen Leistungen sollen mit der ANOVA Korrektur ausgeglichen werden. Kosten werden zu korrigierten Kosten indem die Einflussfaktoren Facharztgruppe, Patientengut und der Standort-Kanton „herausgerechnet“ werden. Die Korrektur erfolgt so, dass die korrigierten Mittelwerte a posteriori hypothetisch als gleich anzusehen sind. Diese Methode bedingt bei der Schätzung von Parametern zumindest annähernd symmetrisch um die Mittelwerte schwankende ärztliche Leistungen, und das bei gleicher Varianz, letztendlich muss es sich um Normalverteilungen handeln. Auch wenn beide Annahmen nicht erfüllt sind, ist die Herangehensweise in Teilen bei Sozialwissenschaftlern üblich, bei Gesundheitsdaten eher selten hilfreich, bei Gesundheitskostendaten nur bedingt

---

<sup>4</sup> Ein Blick in den stationären Sektor lehrt, dass Verweildauern bei Abrechnungen nach DRG nahezu symmetrisch verteilt sind. Hier wirkt sich aus, dass das DRG-System (an deutschen Daten beobachtet) durch die Weiterentwicklungen gut aus den Kinderschuhen erwächst. Für die Kostenverteilung gilt die Symmetrie nicht, obwohl „blutige Entlassungen“ aus dem Krankenhaus sowohl die Verweilzeit als auch die Kosten senken. Die starke Asymmetrie zu hohen Kosten bleibt bestehen, sie ist aus Behandlungssicht „normal“.

<sup>5</sup> Zahlstellenregister ZSR

<sup>6</sup> Rechnungsstellerstatistik RSS

interpretierbar. Die ANOVA Korrektur sagt nur etwas über die Kosten aus, die um einen von höheren Kosten bereinigten Mittelwert schwanken und sich dabei ausgleichen. Im Rahmen der in der Schweiz letztendlich nicht zur Diskussion stehenden „statistischen Durchschnittsmethode“ stellt die ANOVA Korrektur eine Veränderung dar, die, dort wo sie anwendbar ist, weder gut noch schlecht ist. Die bedingte Anwendbarkeit bezieht sich nur auf den sich ausgleichenden Kostenbereich. Es wird an keiner Stelle untersucht, ob eine Varianzanalyse überhaupt anwendbar ist.

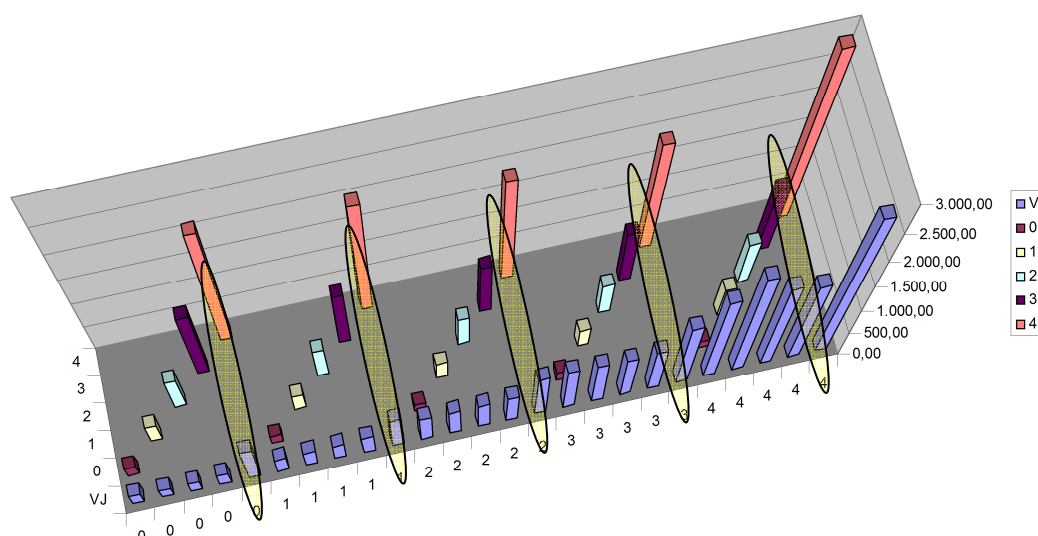
Die Methode gründet sich auf eine Zerlegung der Varianz in als identisch postulierte Schwankungen um die zunächst noch unterschiedlichen Mittelwerte und die durch die Unterschiede der Mittelwerte bedingten Schwankungen. Die unterschiedlichen Mittelwerte werden zu gleichen Mittelwerten „korrigiert“. Methodisch ist das im gesetzten Rahmen soweit in Ordnung. Nun ist die Kostenverteilung aber weit entfernt davon, normalverteilt zu sein. Wenige hohe Kosten in Arztreferenzkonstellationen beeinflussen sehr stark den Mittelwert, die Varianzen sind ebenfalls hochsensibel betroffen. Von identischen Varianzen auszugehen ist nicht lauter.

Die Berücksichtigung der Facharztgruppe, des Patientengutes und des Standort-Kantons erscheint zunächst als Verbesserung. Es ist nicht ermittelbar, zu welchen Lasten die Standardisierung führt:

Beispiel: Eine Überarztung aus einem Überangebot von Leistungserbringern führt zu geringeren Kostenmittelwerten und größeren Varianzen. Die geringeren Kostenmittelwerte resultieren aus der Mengenausweitung bei den eher gesunden Versicherten, bei den intensiv zu behandelnden Patienten kommt es in der Regel nicht zu einer Mengenausweitung. Die größeren Varianzen entstehen vor allem durch die Verschiebung des Mittelwertes weg von den höheren Kosten. Die Anzahl der Ärzte über der „Toleranzschwelle“ nimmt bei geringeren mittleren Kosten zu.

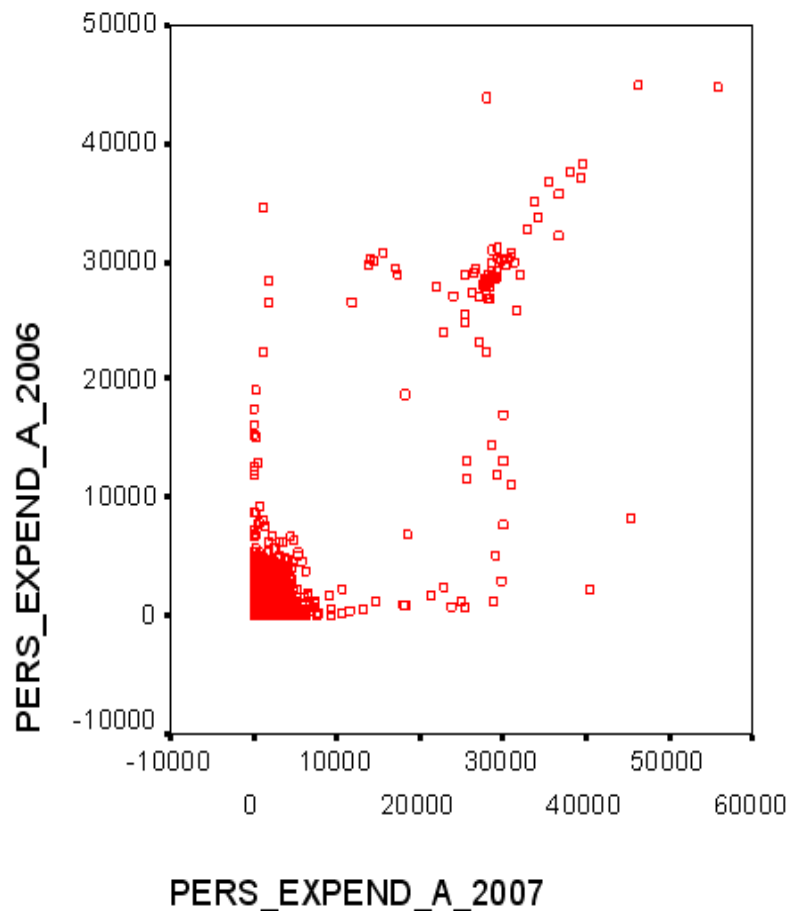
Beispiel: Die Tätigkeiten der verschiedenen Facharztgruppen variieren nicht nur sehr stark, es sind die gruppenspezifischen Behandlungskosten, die stark unterschiedlich streuen. Die präziseren Mittelwerte sind als Vorteile bei der Standardisierung anzusehen, die großen sehr sensibel auf wenige Fälle reagierenden Streuungen lassen Auswirkungen schlecht vorhersagen. Einen Tautologie-Bias vermuten Sie aber m. M. n. zu Unrecht. Es ist sehr wohl so, dass bestimmte hohe Kosten spätere hohe Kosten erklären, dass Kostenentwicklungen sich in Vorzeiten andeuten.

Ich führe zwei Beispiele aus der Untersuchung deutscher Gesundheitskosten an. Zunächst betrachte ich mittlere Jahreskosten gekoppelt an die Anzahl der Arztkontakt-Quartale aufeinanderfolgender Jahre.



**Abbildung 4.** Hohe Arztkosten kündigen sich im Vorjahr an. Die Achsen bezeichnen die Anzahl der Arztkontakt-Quartale. Die unterste Säulenzeile gibt die mittleren Jahresgesamtkosten im Vorjahr an. Die von den Ellipsen noch erreichten Säulen sind die Jahreskosten der „Hochnutzer“ (4 Quartale) im Folgejahr.

Im Weiteren betrachte ich nur die ambulanten Jahreskosten in aufeinanderfolgenden Jahren.



**Abbildung 5.** Hohe Kosten sind im ambulanten Sektor relativ selten. Das Aufeinanderfolgen von hohen Kosten ist nicht selten. Auffällig ist das „Nest“ der Kosten im Bereich von 30.000€ pro Jahr (Dialyse, Strahlentherapie, künstliche Befruchtung, ...).

**Frage 5.** Ist die ANOVA Korrektur mittels aggregierten Alters-/ und Geschlechtsklassen statistisch korrekt? (Anm.: hier geht es um das Ausmass des Aggregation Bias, wonach rund 40% der Kosten durch Alter/Geschlechts Klassen erklärt werden, obwohl nur 4% anhand von Individualdaten zu erwarten wären, und was das für Effekte auf das Endergebnis haben wird, das heisst, wie weit die ANOVA Korrektur die fixed effect variable ZSR korrekt oder eben verzerrt korrigiert).

Wenn alle Patientenkosten eines Arztes berücksichtigt werden sollen, dann ist die ANOVA Korrektur keinesfalls statistisch korrekt. Es können aber für bestimmte Therapien, denen ein ähnlicher Diagnostikaufwand gegenübersteht, u. U. Kostenbereiche gefunden werden, wo sich Kosten um einen spezifischen Wert alters- und geschlechtsabhängig ausgleichen. Hier wäre die Anwendung der ANOVA Korrektur nützlich. Nur hier könnte gutgläubig eine Beziehung zum Aggregation Bias hergestellt werden.

**Frage 6.** Sollte eine Korrektur für die Arztkosten auf Individualdaten der Patienten erfolgen, um einen Aggregation Bias zu vermeiden? (Anm.: dies ist eine wichtige grundsätzliche Frage, da santésuisse nur über aggregierte Alters/Geschlechts-Klassen verfügt, im Gegensatz zu den Versicherern, welche sich weigern, die Einzeldaten aus „Datenschutzgründen“ [Zitat santésuisse] zu liefern).

Eine Berücksichtigung von Individualdaten der Patienten trägt zum besseren Verständnis bei den Arztkosten bei. Da der RSS-Index und die ANOVA Korrektur jedoch nicht auf die Entstehung



von Kosten durch Behandlung von Patienten ausgerichtet ist, kann diese Berücksichtigung von Individualdaten unterbleiben. Ein Entgeltsystem, das nur zur Kenntnis nimmt, dass ein „Mittelwert“ von Kosten steigt (was nicht einmal etwas mit einer Überarztung zu tun hat), bestenfalls noch das Steigen der Kosten bei einem „mittleren Patienten“ konstatiert, ist durch die Berücksichtigung von Individualdaten nicht zu verbessern.

Der mittlere Patient wird älter, ältere Patienten benötigen in der Regel höhere Behandlungsaufwendungen, für mehr ältere Patienten sind auch mehr Ärzte vonnöten. Ärzte behandeln Patientenkollektive konkreter Patienten, die sich weder im Mittel im Alter, in der Geschlechtszusammensetzung, in der Morbidität, ... unmittelbar bei ihm ausgleichen, noch ist der Arzt mit einem mittleren Behandlungsbedarfsgebiet versehen, noch ist die Facharztmischung, noch der Umfang der Ärzteschaft bedarfsvergleichbar.

Beispiel: Die Fallpauschalen bezogene Krankenhausstatistik in Deutschland (DRG-Statistik – Diagnosen und Prozeduren der vollstationären Patienten und Patientinnen in Krankenhäusern) als Vollerhebung mit „weitgehend ... vollständiger ... Erfassung“ (Selbsteinschätzung des Statistischen Bundesamtes) weicht erheblich von Statistiken ab, in denen Merkmale der Krankenhäuser und der Krankenhauspatienten in Bezug auf z. B. Diagnosen kumuliert erfasst werden.

Die letztendliche Ursache hierfür ist, dass nach ICD-10-Diagnosen keine Abrechnung von Leistungen durch die Leistungserbringer erfolgt, auch überhaupt nicht erfolgen kann. Die Behandlung (und dann auch Abrechnung) orientiert sich zunächst an Krankheitsgruppen MDG/DRG (erste drei Stellen) und die gruppeneigenen CCL, im zweiten Schritt kommen die (eine weitere) Diagnostik betreffenden OPS, therapeutische OPS und die PCCL hinzu. Es gibt nur wenige Kodierzusammenhänge (zu ICD-10), es gibt keinen Kodierzusammenhang, der nicht durch die vom Controlling benutzte Software bedient wird. Schon bei der Aufnahme eines Patienten erstellt das nach DRG abrechnende Krankenhaus eine eigene „Einlieferungsdiagnose“ (die assoziiert ist mit der für möglich gehaltenen DRG-Abrechnung nach MDG/DRG [erste drei Stellen]), diese ist in der Regel nicht die Einlieferungsdiagnose eines einweisenden Arztes.

Beispiel: Wird ein aus dem Bett gestürzter älterer Patient mit einer (möglichen) Kopfverletzung in ein Krankenhaus zu einer Tageszeit eingeliefert, in der kein CT verfügbar ist, dann wird entweder an ein anderes Krankenhaus verwiesen oder auf die Abrechnung nach B80\* (Kopfverletzung) fokussiert (B80Z, Risikofaktor 0,305, mittlere Verweildauer 2,3 Tage, Erlös weniger als 1.000 €).

Die gleiche Einweisung in ein Krankenhaus mit 24-Stunden-CT-Verfügbarkeit und Stroke Unit (OPS 8-981 Neurologische Komplexbehandlung) orientiert auf B39\* (Schlaganfall) (B39A, Risikofaktor 5,593, mittlere Verweildauer 16,0 Tage, Erlös fast 20.000 €, im Bundesland Rheinland-Pfalz fast 25.000 €).

Auch bei einer nicht risikoaversen Arztvergütung sind Individualdaten der Patienten nicht leicht und in beliebiger Tiefe zur Korrektur hinzuziehbar.

Beispiel: Als eine Stichprobe der DRG-Statistik kann die Datenbank beim InEK (Institut für das Entgeltsystem im Krankenhaus) gelten. Freiwillige Kalkulationskrankenhäuser ergänzen fallbezogene Kostendaten zu den Grunddaten der DRG-Statistik. Die Struktur der in der „Stichprobe“ abgebildeten Kostenträger, Leistungserbringer, Regionen, Patienten, Behandlungen, Kodierungen, ... weicht ebenfalls erheblich von DRG-Abrechnungen in Deutschland nach der DRG-Statistik ab. Diese nicht zufällige Stichprobe des InEK wird zwar zur Weiterentwicklung des DRG-Systems in Deutschland benutzt, für einen nicht so durchdacht geregelten Bereich wie die ambulante Versorgung sind aus einer solchen Datenbank abgeleitete Kennwerte angreifbar.

Arztkosten und Individualdaten sollten aus der gleichen Datenquelle gezogen werden, am geeignetsten ist hier die Datenquelle des Kostenträgers (die Kostenträger stellen der santésuisse nur nach Ärzten verdichtete Daten bereit, ein Patient kann in mehreren Abrechnungen vorkommen).

Durch ein Mehr an Individualdaten, die von den Patienten „preisgegeben“ werden, steigt die Regressgefahr und damit die Ausuferung von Behandlungsleistungen zur Verringerung dieser Gefahr. Der Arzt unterbricht den diagnostischen Prozess in der Regel mit seiner Entscheidung, die Therapie zu beginnen. „Dabei besteht die Entscheidung nicht in der Feststellung, ob die

Krankheit nun tatsächlich vorliegt oder nicht. In der Praxis ist vielmehr von Bedeutung, ob das Maß der Wahrscheinlichkeiten für therapeutische Entscheidungen ausreichend ist.“<sup>7</sup>

**Frage 7.** Die ANOVA Korrektur anhand der ZSR Variable und den Alters- und Geschlechtsklassen erklärt rund 80% der Arztkosten. Lässt sich diese Beobachtung mathematisch und gemäss Literatur begründen? (Anm.: dies ist eine Anschlussfrage an Frage 5, mit Fokus auf die Gesamtpformance der aktuellen ANOVA-Indexes, wie weit dies überhaupt in der Literatur beschrieben wurde und ob die ETH Gutachter hier mit der Erklärung, der ANOVA Index könne so gerechnet werden, aus gesundheitsökonomischer Sicht nicht nachvollziehbare Empfehlungen abgegeben haben, mithin, ob das Seminar für Statistik der ETH Zürich in seinem Gutachten 2005 überhaupt die Voraussetzungen mitbringt, die damalige Fragestellung adäquat zu beantworten, wenn man das Ergebnis betrachtet [hohes R<sup>2</sup> des ANOVA-Indexes]).

Es gibt keine Literatur, es wird eine solche Literatur auch nie geben, die das Vorgehen des Seminars für Statistik der ETH Zürich im Zusammenhang mit Kosten bei der Behandlung<sup>8</sup> empfiehlt.

Dreist sind die in dem Gutachten von 2005 auf Seite 12 abgegebenen Bewertungen. Diese Bewertungen sind alle falsch bzw. unnütz.

*Kriterium Methode:* Wenn nur auf einen Mittelwert aller Kosten in den Gruppen abgestellt wird, dann ist die Berücksichtigung von Alter, Geschlecht und Kanton sogar ein massiver Rückschritt. Die hohen Kosten verteilen sich nicht gleichmäßig auf die Arztpraxen, nicht entsprechend Alter und Geschlecht der Patienten. Würden die höheren Kosten a priori eliminiert, dann wären die neuen Restmittelwerte u. U. eine Verbesserung.

*Kriterium Ausreisserproblematik:* Hier zeigt sich das absolute Unverständnis beim Seminar für Statistik. Erstens handelt es sich bei einer Heavy-tail Verteilung<sup>9</sup> (eine solche haben die Gesundheitskosten) nicht um Ausreißer, sondern um einen Bereich höherer Kosten, der zweitens aber nur wenige Patienten betrifft, und so nicht bei jedem Arzt gemittelt oder gewichtet zu Buche schlägt. Bei als normalverteilt angenommenen Daten könnte das Seminar für Statistik so schließen, wie sie geschlossen haben. Es wird an keiner Stelle darauf hingewiesen, dass trotz des Nichterfülltseins von Voraussetzungen abgewogen wurde, dennoch die Methode für nützlich zu erachten.

Der Hinweis des Seminars für Statistik auf eine „... Logtransformation der Daten“ ist gut, die andere Kostenverteilungsart wird somit m. M. n. zur Kenntnis genommen, und zwar nicht nur die Schiefe sondern auch die Heavy-tail-Eigenschaft. Hiermit wird eine andere Betrachtungsrichtung eröffnet. Die ANOVA-Korrektur wird damit aber final beerdigt, weil höhere Kosten durch die Transformation heruntergewichtet, die Varianzkomponenten damit auch geschwächt und das Vorhandensein von höheren Behandlungskosten bei allen Ärzten Bedingung wäre.

*Kriterium Vergleichsgruppengrösse:* „Grössere Vergleichsgruppen bilden stabilere Vergleichswerte“ (Zitat aus dem Gutachten), dieses gilt nur unter der Annahme, dass es sich um

---

<sup>7</sup> Köbberling, J.: Diagnoseirrtum, Diagnosefehler, Befunderhebungsfehler, Bewertungen und Vermeidungsstrategien, Verlag Versicherungswirtschaft, Karlsruhe 2013, S. 5.

<sup>8</sup> „Kosten bei der Behandlung“ steht hier für den „Leistungsumfang der [Krankenversicherungen], der Prävention, Diagnostik, Symptomatik, Kuration (Therapie von Krankheiten), Rehabilitation und Palliative Therapien“ betrifft, nach Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.

<sup>9</sup> Bei allen Verteilungen von Behandlungskosten in Deutschland handelt es sich um Heavy-tail Verteilungen, hohe Kosten sind hier „normal“. Ein gutes Modell stellt die Pareto-Verteilung dar. Real sind mit 80 % der Behandlungen rund 20 % der Kosten gedeckt, im 95 % Konfidenzintervall liegen 65 % der Kosten, auf wenige nicht korrekt als Hochkostenfälle [aus gesundheitsökonomischer Sicht sind sie es] bezeichnete Fälle entfallen noch 35 % der Kosten. Siehe z. B. Brammer, K., Schmidt, A., Warmuth, W., „Morbiditätsorientierte Simulation des Ertrags bei der gesetzlichen Krankenversicherung auf der Basis einer kassenspezifischen zufälligen und marktabhängigen Zu- und Abgangs-Modellierung“, 9. Data Mining Cup-Anwendertage, 04./05.06.2008; auch Klemd, F. „Modellierung des Heavy-tail Kostenbereiches für Krankenversicherungen – unter besonderer Berücksichtigung der Ausbreitung, der Abgrenzung zu ‚Ausreißern‘ und der Nähe zu Extremverteilungen –“, BA-Arbeit, 2012.

Patienten bei den Ärzten handelt, deren homogene Kosten sich jeweils im Mittel ausgleichen. Wenn die höheren Kosten eliminiert würden, dann geschähe dieser Ausgleich um einen kleineren Mittelwert auch nur hypothetisch, die Unterscheidung nach Arztgruppen wäre hier jedoch nützlich. Nach dem propagierten Inhalt des Gutachtens 2005 (Vergleichbarkeit sichern) müssten die Vergleichsgruppen so groß gebildet werden, dass alle Patienten, Arztgruppen, Kantone mit ihren niedrigeren und höheren Kosten repräsentiert wären, an ein gleichgroßes Gebilde könnte die Kostenerstattung als Block erfolgen. Ich vermute, dass ein solches Konstrukt die Größe der Schweiz hätte (nur die Schweiz ist sich vermutlich ähnlich, Selbstähnlichkeit), damit würde es zu einem Realkostenausgleich kommen.

*Kriterium Berücksichtigung von Besonderheiten:* Die benannte Berücksichtigung von Alter und Geschlecht betrifft einen „mittleren Patienten“, hier verweise ich auf die Antwort zur Frage 6, wo das schon in seinen risikoaversen Auswirkungen dargestellt wurde.

*Kriterium Spezifität:* Gesundheitsökonomisch wird auf wenige „falsch-positive“ Fälle gesetzt, wodurch hohe Gesamtkosten vermieden werden. Die Spezifität („richtig-negative“ Fälle zu erkennen) wird allein hoch gewertet. Die Sensitivität („richtig positiv“ Fälle zu erkennen) ist zumindest für Leid, das Patienten unnötig zugefügt wird, verantwortlich. Die Ethik bleibt hier aus gesundheitsökonomischer Sicht auf der Strecke, auch Ärzte klären ihre Patienten hier nicht immer genügend auf. Im Falle des ELISA-Tests auf HIV kommt es regelmäßig aus diesem Grunde zum Suizid<sup>10</sup>.

*Kriterium Weiterentwicklung:* Die Methode steckt in ihrer praktizierten Anwendung in einer sich selbst einengenden Sackgasse.

*Kriterium Abdeckungsgrad:* Aus den gleichen Daten wird mit vorgeblichen Innovationen ein schlechteres Ergebnis erzielt, es sei denn es wird auf einen Realkostenausgleich abgestellt, wie er sich aus dem „Kriterium Vergleichsgruppengröße“ vermutlich ergibt.

**Frage 8.** *Wie beurteilen Sie die Herangehensweise im ETH Gutachten Stahel/Roth 2005 und die Erkenntnisse aus dem ETH Gutachten Nr. 2 für den Kanton Tessin? (Anm.: hier geht es darum, in welchem Ausmass Sie feststellen können, dass die Gutachter sich nun aufgrund des zweiten Gutachtens widersprechen).*

Insbesondere nach der Selbsteinschätzung im Gutachten 2005 vom Kriterium Methode und vom Kriterium Ausreisserproblematik her hätte das Ergebnis im Kanton Tessin überhaupt nicht auftreten können. Es ist einfach das eingetreten, was zu erwarten war, dass einfach etwas anderes herauskommt.

Ich vermute, dass durch die Herauslösung der höheren Kosten eine größere Ähnlichkeit der Kantone verbleiben würde.

Der Ausflug der Autoren in dem neueren Gutachten zu „mässig bis stark korrelierten“ Indexwerten für die beiden Methoden zeigt, dass sie sich selbst von ihrem Gutachten 2005 distanzieren.

**Frage 9.** *Wie beurteilen Sie ganz prinzipiell ein Modell, das primär die Arztkosten als erklärende Variable (fixed effect model) mit einbezieht? (Anm.: hier geht es darum, wie weit ein solches Modell für die Erklärung der Arztkosten in der Gesundheitsökonomie überhaupt diskutiert wurde, und falls ja, ob es validiert wurde).*

Zur Prognose von *Gesamtaufwendungen* ist die Einbeziehung von Arztkosten als erklärende Variable hilfreich. Das habe ich in einer Reihe von Projekten demonstriert (z. B. Vorhersage von Krankenversicherungskosten eines Distrikts in Indien, z. B. Vorhersage von Programmkosten in speziellen Versorgungsinitiativen einer deutschen Krankenkasse). Für einen einzelnen Arzt kann mit hoher Wahrscheinlichkeit (>80 %) geschlossen werden, dass seine „Hochkostenpatienten“ auch im Folgejahr diese hohen Kosten verursachen (die Abbildung 4 gehört zu diesen Untersuchungen). Für ein Indexverfahren auf Vergleichsgruppenbasis wäre

---

<sup>10</sup> Gerd Gigerenzer, J. A. Muir Gray, Bessere Ärzte, bessere Patienten, bessere Medizin, Medizinisch Wissenschaftliche Verlagsgesellschaft

hierfür die Repräsentativität dieser Patienten mit höheren Kosten vonnöten, diese Vergleichsgruppe hätte vermutlich die Größe der Schweiz.

**Frage 10.** *Ist die Weiterentwicklung der Methode anhand weiterer erklärender Variablen mit der ANOVA Methode möglich (Literatur dazu?) oder wäre ein Regressionsmodell hilfreicher, falls ja, warum? (Anm.: hier geht es darum, ob künftige Variablen eben auch nicht kategorischer Art sein werden und damit die ANOVA Methode nicht mehr adäquat eingesetzt werden kann; zweitens um die Frage, ob überhaupt ein ANOVA Modell als primäre statistische Methode nicht primär mit einem REGRESSION-Modell hätte konzipiert werden müssen, mithin die Frage, warum die Erfinder des ANOVA-Indexes auf die ANOVA Methode gesetzt haben und was für Vorteile sie damit allenfalls davon erwartet haben könnten).*

Wenn die höheren „normalen“ Kosten den Mittelwert bei einzelnen Ärzten unvergleichbar verzerren, dann sind Weiterentwicklungen im Rahmen einer Varianzanalyse nicht möglich, dann ist eher ein Rückbau oder Nichteinsatz angebracht. Es ist nicht erkennbar, dass die ANOVA-Methode bei der Arztvergütung nützlich ist.

Ein Regressions-Modell zur Arztvergütung ist überhaupt ungeeignet.

Während für die Anwendung des Regressions-Modells die Normalverteiltheit (im mathematischen strengen Sinne) Bedingung ist, kann die ANOVA-Methode auch teilweise auf Schätzungen zurückgreifen, die nur eine Nähe zu einer symmetrischen Verteilung aufweisen. In diesem Sinne ist der Versuch über die ANOVA-Methode vermutlich gestartet worden. Hier haben die „Erfinder“ des ANOVA-Indexes vermutlich ihre Hoffnungen gesetzt. Zur Schätzung der Größe, des Gewichtes und der Kosten risikoloser geborener Einlinge, oder zur Schätzung der Schwangerschaftsdauer bei Ausschluss der Frühgeburten und auch der Kaiserschnittgeburten wäre die ANOVA-Methode ein guter Griff.

Der Nützlichkeitsnachweis der ANOVA-Methode bei der Arztvergütung fehlt, er kann auch nicht erbracht werden, wenn auf den Mittelwert aller Kosten abgestellt wird.

**Frage 11.** *Ist eine verlässliche Beurteilung der Überarztung ohne Morbiditätsvariable möglich (RSS-Index) und ist ein Surrogatmarker Alter/Geschlecht für die Morbidität ausreichend? (Anm.: hier geht es um die Frage, ob der ANOVA-Index vom Konzept her und aufgrund bekannter Tatsachen nie hätte als statistische Verbesserung des RSS-Index bezeichnet werden dürfen, wie die Gutachter des Seminars für Statistik dies ja getan hatten).*

Überarztung ist aus Mittelwerten wegen der Spezifik von Gesundheitskosten nicht ablesbar. Die Tatsachenvermutungsgrenze kann an einer Obergrenze von Kosten, relativ von einem Mittelwert abgeleitet, nicht festgemacht werden. Eine Überarztung wird durch gesundheitsökonomische Handlungen gefördert, das Vorgehen nach dem RSS-Index verschlechtert die Grund-Versorgung nicht hochmobiler Versichertengruppen, fördert selektive Versorgungsangebote für junge, gesundheitsbewusste Versicherte, verschiebt Behandlungen regional und zeitlich zu Lasten von Versicherten. Der ANOVA-Index, unter anderen Prämissen erdacht, war nie eine Verbesserung, sicher eine Veränderung, teilweise eine Verschlechterung. Beispiel: Geringere Mittelwerte entstehen bei einer Mengenausweitung auf weniger morbide Versicherte, z. B. sind in zwei Regionen in Deutschland (Westfalen, Nordbaden) wegen eines Überangebotes an Leistungserbringern mehr Patienten anzutreffen, die dafür im Mittel nicht so hohe Kosten verursachen. Die Gesamtkosten fallen auf alle Versicherten bezogen höher aus, der Durchschnitt bei den Patienten ist geringer als in anderen Regionen.

Beispiel: Der Vorlaufzeit-Bias (z. B. bei Prävention) führt zu einer Vorverlegung des Zeitpunktes der Diagnose, damit zu einer Fallzahlerhöhung.

Der Überdiagnose-Bias erhöht durch das Aufgreifen nichtprogressiver Erkrankungen die Fallzahl.

Der Intervall-Bias überschätzt langandauernde Phänomene bei Zeitpunkt Betrachtungen.

Alle diese Verzerrungen sind Ausdruck von gesundheitsökonomisch verursachter Überarztung.

Beispiel: Der Selektion-Bias drückt aus, dass bekannte und unbekannt

Störgrößen/Einflussgrößen nicht gleichmäßig in z. B. der Interventions- und z. B. der

Kontrollgruppe verteilt sind (in Bezug auf die höheren Kosten hatte ich das schon benannt). Hier kann ganz sicher nicht eine Überarztung belegt werden, auch wenn sie vorhanden sein sollte. Beispiel: Die von mir in vielen Projekten bei Krankenkassen evaluierten Disease Management Programme in Deutschland führen zu einer Überarztung und Übermedikalisierung eher „gesunder“ Versicherter. Ihr durchschnittliches Kostenniveau ist niedriger als bei Versicherten mit höherer Morbiditätslast und höherer notwendiger Behandlungsintensität, soviel auch hier zum Mittelwertvergleich. Die in Deutschland seltene Integrierte Versorgung in den gleichen Krankheitsgruppen ist erfolgreich.

Beispiel: Die Einführung ambulanter Operationen führte zu einer beträchtlichen Mengenausweitung bei dieser Art Operationen. Im Durchschnitt sind diese Operationen günstiger als vergleichbare stationäre Eingriffe, da auch weniger komplizierte und unnötige Fälle operiert werden. Summarisch sind die Kosten merklich höher. Eine Sonderidee des Spitzenverbandes Bund der Krankenkassen führt dann noch zu weiteren Kosten: „Blutige Entlassungen“ aus ambulanten Operationen werden außerbudgetär (also nicht im Rahmen des Budgets der Kassenärztlichen Vereinigungen) direkt ungedeckt bei den Krankenkassen abgerechnet.

Beispiel: Ein geringerer Einzelkostensatz erfordert aus Ertragsgesichtspunkten eine Erhöhung der Patientenanzahl, dieses führt zu einem Anwachsen der mittleren Wartezeit der Patienten.<sup>11</sup>

**Frage 12.** *Basierend auf dem Überarztungsmodell Beck/Kunze<sup>10</sup> ist trotz Einbezug von Morbiditätsvariablen<sup>11</sup> entweder eine Flexibilisierung der Beweisgrenze notwendig oder es ist sogar je nach Facharztgruppe nicht möglich, verlässlich Kostenüberschreitungen zu erfassen<sup>12</sup>. Was hat dies für Implikationen für den RSS-Index und den ANOVA-Index, welche beide ohne Morbiditätsvariable(n) auskommen? (Anm.: hier geht es um die Frage, dass die Beobachtungen Beck /Kunze eigentlich dazu führen müssten, die Validität des RSS-Indexes erheblich zu bezweifeln, dass also mit dieser Publikation durchaus Konsequenzen für ein Moratorium des RSS-Indexes und des ANOVA-Indexes hätten plausibel hergeleitet werden können).*

In dem Artikel von S. von Rotz/U.Kunze/K.Beck wird ein „Instrument zur Beurteilung der Wirtschaftlichkeit von Grundversorgern“ vorgestellt. Die zur Anwendung des Modells notwendigen Voraussetzungen werden nicht geprüft, sie sind auch nicht erfüllt. Dennoch handelt es sich um ein bedingt nützlich Herangehen, da nur Variablen einbezogen werden, die in einem Zusammenhang zu den Arztkosten stehen können/könnten (nicht etwa wie in der MorbiRSA-Regression in Deutschland, wo Kodierungen betrachtet werden, die überhaupt keinen Kostenbezug bei Ärzten haben). Das in dem Artikel verwandte  $R^2$  ist natürlich nur eine Krücke, die nur unter normalverteilten Kosten betrachtet werden könnte. Es ist jedoch nach diesem Artikel überhaupt nicht zu erwarten, dass einzelne Ärzte mit einem RSS-Index in wirtschaftliche und unwirtschaftliche Ärzte eingeteilt werden können. Vielleicht gelänge das für große Gruppen ähnlicher Ärzte, die dann als Gruppe sicherlich aber geschlossen merklich unter der 130 %-Schranke liegen würde. Der RSS-Index bleibt ein nicht zu rechtfertigender Rahmen. Eine vernünftige Grenze ist auch nicht verhandelbar. Da auch bei einer Berücksichtigung einer Facharzt-, Kanton-, Versicherten-, Leistungserbringerdichte-, ... Spezifik die höheren Kosten über diese Gruppen nicht gleichmäßig gewichtet verteilt sind, ist auch eine anteilige Bemessung bis zu einer spezifischen Schranke bei großen Gruppen schwer möglich. Bei einzelnen Ärzten scheidet sie aus.

**Frage 13.** *Basierend auf der Tatsache, dass selbst Morbi-korrigierte Methoden (Beck/Kunze) nicht für alle Fachrichtungen anwendbar sind, müsste man da nicht davon ausgehen können, dass der RSS- und ANOVA-Index zuerst fachspezifisch auf die Validität geprüft werden müssen, bevor man*

---

<sup>11</sup> Dietrich, C. E. M., Walleitner, P.: Warteschlangen-Theorie und Gesundheitswesen, Einführende Betrachtungen für den Praktiker, Medizinische Informatik und Statistik 34, Springer Verlag Berlin Heidelberg New York 1982. Patienten, die im Mittel alle 10 Minuten eintreffen, von einem Arzt im Mittel in 9 Minuten behandelt werden, warten im Mittel 81 Minuten in der Praxis, siehe dort S. 43.

*überhaupt daran denken kann, dies als Beweismittel für Überarztung den Gerichten zu präsentieren?*

Die fachspezifische Prüfung sollte bei der Bestimmung der Verteilung der individuellen Fallkosten beginnen. Zu berücksichtigen wären das Alter, das Geschlecht und die Art (Beschäftigter, Rentner, Betreuungsbedürftiger, ...) der Versicherten zum Leistungsbeginnzeitpunkt, die aufgesuchte Erstarztgruppe, die Verfügbarkeit (zeitlich und räumlich) von bedingt für den Fall erforderlichen weiteren Leistungserbringern (u. U. im Netzwerk) unter Berücksichtigung weiterer Merkmale (z. B. muslimische Frauen gehen eher ungern zu einem Frauenarzt), die versorgungsrelevante Leistungserbringerdichte und die patientenbezogene Verfügbarkeit von Leistungserbringern und die Art der Leistung (Vorsorge, Diagnostik, Therapie, ...).

**Frage 14.** *Basierend auf dem deutschen Modell, erfolgt anhand des Morbi-RSA (Erklärungsgehalt im  $R^2$ ?) eine patientenseitige Kostenzuteilung, ferner eine Auditierung von Kostenausreissern, welche im Morbi-RSA Modell nicht erklärt werden können. Ist es vor dem Hintergrund dieser Vorgehensweise überhaupt noch sachgerecht, mit einem RSS-Index oder ANOVA-Index zu arbeiten, oder sollte diese Methode vollständig verlassen werden? (Anm.: hier geht es vor allem um die Frage, ob überhaupt noch eine Vergleichsgruppe für ein adäquates „physician profiling“ notwendig ist, mithin, ob auf die ZSR Variable in den statistischen Modellen komplett verzichtet werden könnte.*

Aus methodischer und datenstatistischer Sicht und vermutlich erst recht aus fachlicher Sicht ist es nicht sachgerecht nach einem RSS-Index oder ANOVA-Index zu arbeiten, das dabei benannte „physician profiling“ könnte jedoch im Rahmen eines „gemanagten“ Vollkostenausgleichs der Krankenversicherungen mit den Ärzten Regelbestandteil sein.

In die Vorgehensweise im Rahmen des Morbi-RSA in Deutschland können keine Hoffnungen gesetzt werden. Die für den Morbi-Teil des RSA verwendete Modellierung ist sinnwidrig<sup>12</sup>, über 60 Milliarden Euro werden Jahr für Jahr auf der Individualebene betrachtet falsch verteilt. Die Regression ist nicht anwendbar, in Bezug auf Alter, Geschlecht und Erwerbsminderung in Teilen jedoch nützlich. Davon betroffen sind aber nur die Krankenkassen. Bei den Ärzten erfolgt eine Vollkostenerstattung, die kassenärztlichen Vereinigungen verhandeln, mit Daten besser ausgestattet und nicht sozialwissenschaftlich methodisch beschränkt, unter Berücksichtigung von Alter und Geschlecht der Patienten, von weiteren soziodemographischen Merkmalen der Patienten, von regionalen Arztgruppen- und Versorgungsunterschieden, mit „Partnern“ eine vollkostendeckende Patientenpauschale.

Mit der Nichtanwendbarkeit der Regression bei Gesundheitskosten hat auch das  $R^2$  keinen Erklärungsgehalt.

**Frage 15.** *Schätzung erwarteter „predictive ratios“ in Abhängigkeit der Praxisgrösse: A) 500 Patienten, B) 1.000 Patienten, C) 15.000 für folgende Modelle: 1) RSS Index, ANOVA-Index, ANOVA-Index + PCG; 2) Morbi-Index, Morbi-Index + Ausschluss teure Patienten. Hier geht es darum abzuschätzen, ob ein Systemwechsel von Modell 1 zu Modell 2 für die meisten Praxen (Grösse um 1000 Pat./Jahr) erwartungsgemäss mit einer signifikanten Verbesserung der „predictive ratios“ verbunden sein wird.*

Solche Schätzungen sind nur möglich, wenn die normalen höheren Kosten ausgespart bleiben. Hierbei sind die Arztgruppen, die Versichertenstruktur, die Kantone mit einer Unterteilung in Gebiete mit einer soliden Grundversorgung, mit einer Überversorgung und mit einer Unterversorgung zu berücksichtigen. Letztendlich könnten solche Schätzungen aber an realen Gesamtkosten validiert werden. Auf den einzelnen Arzt kann das aber sicherlich nicht heruntergebrochen werden.

---

<sup>12</sup> Warmuth, W.: Sie wollen doch nur spielen: Gesundheitsökonomie in der GKV, Versicherungswirtschaft, 68(2013)13. S. 20-24.

**Frage 16.** *Ist die Homogenisierung von Vergleichsgruppen aus gesundheitsökonomischer Sicht zwecks Verbesserung des RSS- und ANOVA-Index je genügend trennscharf möglich, ohne dass erhebliche Confounders verbleiben? (Anm.: hier geht es um die Frage, ob auch gut erarbeitete Vergleichsgruppen innerhalb von Fachärzten nicht doch noch einen begründbar grossen (> 60%) Kostenspielraum belassen könnten, etwa in dem Sinne, dass beispielsweise Kardiologen mit Zuweisungspraxis sich sub-spezialisieren auf Herzinsuffizienz und andere auf Hypertonie, mit entsprechend differierender Kostenstruktur aufgrund der unterschiedlichen Morbiditäten / Co-Morbiditäten).*

Geht man von Mittelwerten aus, die alle Patienten-Kosten eines Arztes berücksichtigen, dann ist die Willkür des RSS- und ANOVA-Index nicht beseitigbar.

Wie bei der Frage 15 halte ich eine Homogenisierung von Vergleichsgruppen (nach Arztgruppen, nach der Versichertenstruktur, nach den Kantonen mit einer Unterteilung in Gebiete mit einer soliden Grundversorgung, mit einer Überversorgung und mit einer Unterversorgung, für normale Kosten) für möglich. An diesen Vergleichsgruppen könnte ein Sub-Spezialisieren simuliert werden (ohne dass es erst zur Praxis wird), es könnte die Ausweitung genauso erkundet werden, wie die Bewirtschaftung des verbleibenden „Restes“. Ebenso sind Confounder „fiktiv“ ermittelbar, die unter Beachtung einer Bilanztreue belegen, dass der Kostenspielraum zwischen den Ärzten dennoch enorm ist, womit sich die Beweiskraft von RSS- und ANOVA-Index pulverisiert.

Leipzig, 30.04.14



---

Dr. Walter Warmuth