RheinAhrCampus Remagen

Gesundheitsökonomie (Master)

Skript

Prof. Dr. Olaf Winkelhake

3. März 2026

Die Arbeitspläne zu den Kapiteln finden Sie unter *https://winkelhake.myrac.de*

Inhaltsverzeichnis

1 Vorbemerkungen 3

I Rationierung 7

2 Formen von Rationierung 8

3 Rationierung und Ethik 12

II Methoden der Effizienzmessung 24

4 Kostendefinitionen 25

5 Ergebnisdefinitionen 28

6 Messung von Lebensqualität 39

7 Lebenserwartung 46

III Epidemiologische Studientypen 60

8 Kohortenstudien 62

9 Fall-Kontroll-Studien 71

10 Standardisierung 75

11 Konfidenzintervalle 82

IV Incentives für Versicherte und Leistungserbringer 95

12 Selbstbehalte und Boni 97

13 Einzelleistungsvergütung 110

14 Pauschalhonorierung 113

15 Pauschalen und Risiko **119**

16 Fixgehalt **131**

V Steuerung über Organisationsstrukturen 136

17 Das Gesundheitssystem der USA **137**

18 Krankenversicherungskonzepte in den USA **144**

19 Das Gesundheitssystem in Großbritannien **151**

20 Die Krankenhausreform in Dänemark **158**

21 Mögliche Lehren für Deutschland **160**

Kapitel 1

Vorbemerkungen

1.1 Prüfung und Prüfungsvorbereitung

Wenn Sie Ihren Bachelor am RAC gemacht haben, wissen Sie, wie die Prüfungen in *Einführung Ökonomie* und *Grundlagen der Gesundheitsökonomie* abgelaufen sind. Die gleichen Spielregeln gelten auch für die Prüfung dieser Veranstaltung.

- Es wird eine Open Book Klausur geben, bei der Sie alles, was aus Papier ist, mitnehmen dürfen. Ihr Taschenrechner muss nicht aus Papier sein.
- Weil Sie in der Prüfung das Skript vor der Nase haben, wird es keine Auswendiglernfragen geben, sondern nur Anwendungs- und Übertragungsfragen.
- Bei Ihrem Studiengang handelt es sich um einen *konsekutiven* Master. Das bedeutet, dass er auf den Inhalten des Bachelors aufbaut. Für diese Veranstaltung bedeutet das, dass ich in der Veranstaltung und in der Prüfung auf die Inhalte meiner Bachelor-Veranstaltungen zurückgreife. Ich versuche dann nicht, Wissenslücken aufzudecken, habe aber keine Skrupel z.B. die Funktionsweise des Gesundheitsfonds, die im zweiten Semester Thema war, vorauszusetzen. Daher ist es klug, die Skripte dieser Veranstaltungen mit in die Prüfung zu nehmen und nicht alles vergessen zu haben, was drin steht.[1]

1.2 Ablauf der Sitzungen

Die Sitzungen finden im flipped classroom Verfahren statt, d.h. ich setze voraus, dass Sie die Arbeitspläne, die Sie von http://winkelhake.myrac.de herunterladen können, jeweils *vorher* abgearbeitet haben.

Als Anhaltspunkt für den Zeitbedarf können Sie sich an folgender Rechnung orientieren. Die Veranstaltung ist mit 5 Credit Points gewichtet, wobei jeder CP mit einem Workload von 30 Stunden verbunden ist, also insgesamt 150 Stunden. Bei 22 Kapiteln

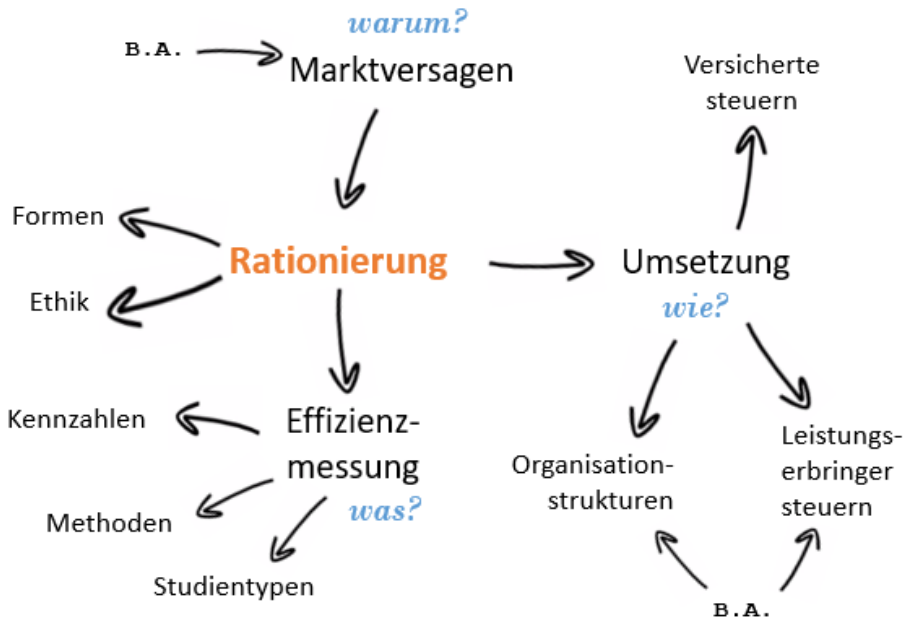
[1] Falls Sie Ihren Bachelor an einer anderen Hochschule gemacht haben, bedeutet das Extraarbeit für Sie, weil Sie diese Skripte noch einmal (zumindest oberflächlich) durchgehen müssen, um Lücken zu erkennen und schließen. Das gilt natürlich auch für den Fall, dass Sie die Prüfungen bestanden, aber inzwischen alles vergessen haben. Der Großteil dieses Hintergrundwissens wird erst in der zweiten Semesterhälfte relevant. Mein Ratschlag ist daher, eventuell vorhandenen Lücken in der ersten Semesterhälfte zu schließen.

ergibt das $^{150}/_{22} = 6,8$ Stunden je Kapitel/Sitzung, wobei wir zwei Sitzungen je Woche haben. Das ist also einiges an Zeit.

Die wöchentlichen Sitzungen sind ein *Angebot* und zwar eines, dass Sie auch ablehnen können.[2] Ich versuche, Ihnen das Wegbleiben so einfach wie möglich zu machen. Alle Zettel, die ich in der Sitzung beschrifte, lade ich zeitnah als Protokoll hoch, so dass Sie sich auch in Eigenregie auf die Prüfung vorbereiten können. Sie sollten aber nicht davon ausgehen, dass Sie auf diese Weise unglaublich viel Zeit sparen, denn die Inhalte der Sitzungen sind natürlich auch prüfungsrelevant. 10-20% gesparte Zeit halte ich aber für denkbar, wenn Sie diszipliniert in Eigenregie arbeiten. Sie können die Protokolle und Ihre Aufzeichnungen natürlich mit in die Prüfung nehmen.

1.3 Rationierung als Querschnittsthema

Den roten Faden der Veranstaltung habe ich in der folgenden Concept-Map dargestellt.



Rationierung bedeutet in einer ersten groben Definition, dass nicht jeder das bekommt, was er gern hätte. Da das ein unangenehmes Thema ist und vermutlich niemand gern rationiert, ist die Frage, warum man überhaupt rationieren sollte.

Inhalte des Bachelors

Im Bachelor haben wir uns mit der Frage herumgeschlagen, warum das (deutsche) Gesundheitssystem so kompliziert ist. Die ganz kurze Antwort war, dass das daran liegt, dass das Gesundheitssystem sofort zusammenbrechen würde, wenn man es nach der Standard-Marktlogik organisieren würde. Weil 10% der Bevölkerung 80% der Kosten verursachen,[3] wären diese 10% schnell tot, wenn sie die Kosten selbst tragen müssten. Also brauchen wir

[2] Anspielung auf „Der Pate“ von Francis Ford Coppola. Don Vito Corleone. Mein großes Idol!

[3] Winkelhake, Olaf et al., "Einkommen, Gesundheitszustand und Inanspruchnahme in Deutschland 1992", Sozial- und Präventivmedizin 1997. S.3-10.

einen Ersatzmechanismus. In Deutschland ist das die Versicherung, die die Finanzierung von der Inanspruchnahme entkoppelt.

Ökonomische und ethische Aspekte von Rationierung

Mit der Entkoppelung von Finanzieurng und Inanspruchnahme handeln wir uns aber andere Probleme ein. Eines davon ist die Überinanspruchnahme des Gesundheitssystems. Wenn wir nichts tun, frisst uns das Gesundheitssystem die Haare vom Kopf. Das ist nicht so schlimm, wie viele tote chronisch Kranke, aber auch nicht erstrebenswert. Also müssen wir, um die Finanzierung des Systems nicht zu gefährden, irgendwelche Mechanismen verwenden, die diese Überinanspruchnahme verhindern oder zumindest eindämmen. Über diese Mechanismen werden wir im zweiten Kapitel nachdenken.

Man könnte vielleicht einwenden, dass ich hier den zweiten Schritt vor dem ersten mache. Sollte man nicht zuerst klären, ob Rationierung ethisch überhaupt vertretbar ist? Wenn das nicht so wäre, würde sich doch vieles erübrigen.

Ich möchte trotzdem bei dieser Reihenfolge bleiben, denn es ist schwierig, eine ethische Frage zu beantworten, wenn gar nicht klar ist, worum es in dieser Frage überhaupt geht. Dann geht es letztlich nur um inhaltsleere Empörung.

Effizienz als Kriterium

Wir werden (Achtung Spoiler!) sehen, dass wir um Rationierung nicht herumkommen. Dann ist die Frage, nach welchem Kriterium man rationieren soll. Nach der alphabetischen Sortierung der Krankheiten sicherlich nicht und auch nicht, ob ein Promi diese Krankheit hat oder nicht. Ein sinnvolles Kriterium ist die Effizienz und (das haben die Bachelor bei mir in der ersten Erstsemesterwoche gehört) wir Ökonomen sind *Effizienzhandwerker*. Wir werden uns also mit der Frage befassen, mit welchen Kennzahlen und Instrumenten man das „Preis-Leistungs-Verhältnis“ medizinischer Leistungen messen kann.

Steuerung von Versicherten und Leistungserbringern

Wenn wir eine Idee haben, wie wir medizinische Leistungen identifizieren, die relativ viel bringen und dabei relativ wenig kosten, ist die Frage, ob wir die unmittelbar Betroffenen, d.h. die Patienten und die Ärzte, dazu bringen können, mehr oder minder freiwillig auf Leistungen zu verzichten, die ein schlechtes Preis-Leistungs-Verhältnis bieten.

Wir werden sehen, dass wir keinen guten Hebel haben, um bei den Versicherten anzusetzen, obwohl die Idee von Selbstbehalten in der öffentlichen Diskussion immer wieder hoch gehandelt wird. Das wird sich aber als Illusion entpuppen.

Vielversprechender ist der Ansatz bei den Leistungserbringern/Ärzte, bzw. bei den „Spielregeln“ nach denen die Ärzte bezahlt werden. Wir schauen uns da verschiedene Konzepte an.

Steuerung über Organisationsstrukturen

Im zweiten Semester haben wir uns die Organisation des deutschen Gesundheitssystems angeschaut. Vielleicht neigen wir dazu, zu denken, dass das auch so sein muss. Wenn wir uns in diesem Teil der Veranstaltung die Gesundheitssysteme anderer Länder anschauen,

werden wir sehen, dass das überhaupt nicht so ist und andere Länder ganz andere Organisationsstrukturen für das gleiche „Problem“ Gesundheit haben. So wie wir ein ganzes Semester mit dem deutschen System verbracht haben, könnten wir auch ein ganzes Semester z.B. mit dem britischen System verbringen. Die Zeit haben wir aber nicht und so wichtig ist es für Sie mit Sicherheit nicht, die Ecken und Winkel dieses Systems kennenzulernen. Wir werfen also nur jeweils einen groben Blick auf das System und picken uns ein interessantes Detail heraus.

Teil I

Rationierung

Rationierung ist ein Thema, das sich wie ein roter Faden durch die Gesundheitsökonomie zieht. Rationierung bedeutet, Patienten gezielt medizinische Leistungen vorzuenthalten. Etwas verbal weichgespülter spricht man auch von *Priorisierung* und meint damit Verfahren, mit denen man „wichtigere“ von „weniger wichtigeren“ Leistungen unterscheiden kann. Letztlich bedeutet das aber das gleiche, denn „weniger wichtig“ bedeutet letztlich „kann man als erstes weglassen.“

Wenn man sich mit Rationierung befasst, sind folgende Fragen zentral:

1. Warum muß man überhaupt rationieren? Geht es nicht auch ohne Rationierung?
2. Welche Formen von Rationierung gibt es?
3. Ist Rationierung ethisch überhaupt vertretbar?
4. Wenn man rationiert – nach welchen Spielregeln sollte das geschehen?

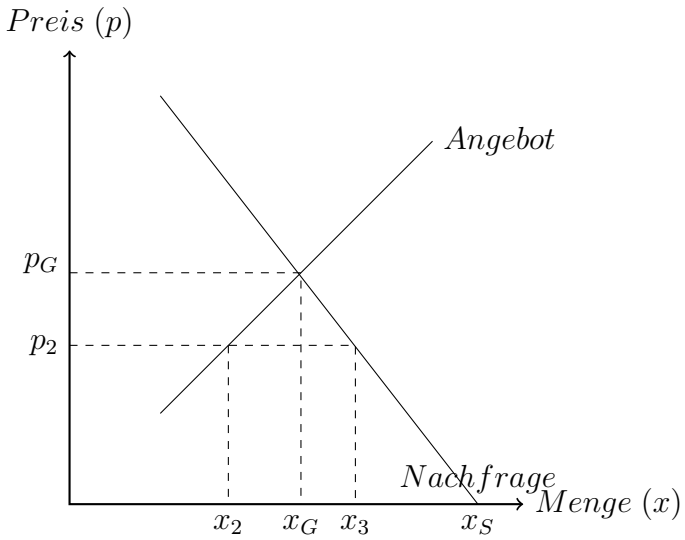
Mit den ersten drei Fragen werden wir uns in diesem Kapitel „Rationierung“ befassen. Die Beantwortung der vierten Frage wird einen großen Teil der ersten Semesterhälfte und die Kapitel „Effizienzmessung“ und „Epidemiologie“ ausmachen. Zudem stellt sich die vierte Frage nur dann, wenn wir die erste und dritte Frage bejahen.

Kapitel 2

Formen von Rationierung

2.1 Rationierung auf Märkten

Ein Umstand, den wir leicht übersehen, ist, dass Rationierung auf jedem Standard-Markt stattfindet. Dazu habe ich einmal eine geometrische Darstellung eines solchen Standard-Markts gezeichnet:



Wir wissen, dass der Gleichgewichtspreis bei p_G liegt. Nehmen wir an, wir sprechen über iPhones und der Preis eines iPhones liegt bei 1.000€. Hat jeder, der ein iPhone haben will, eins? Nein. Jeder, der bereit ist, 1.000€ zu bezahlen, hat eins. Bob kann oder will die 1.000€ nicht bezahlen und hat deswegen keins. Der Preis funktioniert also als Rationierungsmechanismus und führt dazu, dass diejenigen, die am meisten zu zahlen bereit sind, eher zum Zug kommen als die, die weniger zu zahlen bereit sind. Aus dem Bachelor kennen wir die Sättigungsmenge, die hier bei x_S liegt. Genau genommen wird die ganze Nachfrage zwischen x_G und x_S rationiert. Es gibt also viele Menschen, die gern ein iPhone hätten, aber keins haben.

Dieses einfache Diagramm macht uns auch deutlich, was das Problem an Versicherungsmärkten ist. Wenn die Nachfrage von der Finanzierung entkoppelt ist, hat das Gut aus Sicht des Versicherten einen Preis von Null und dann wird die Sättigungsmenge nachgefragt.

Bei vielen Gesundheitsgütern müssen wir nicht fürchten, dass die Versicherten unbedingt eine Dialyse haben wollen, obwohl ihre Nieren funktionieren. Einfach nur, weil Dialyse sie nichts kostet. Dafür sind medizinische Leistungen häufig zu unangenehm und zeitraubend. Wir können aber festhalten, dass die Nachfrage wegen des fehlenden Preissignals relativ hoch sein wird und damit auch die Kosten, die zu refinanzieren sind.

2.2 Harte und weiche Rationierung

Das Begriffspaar „harte vs. weiche Rationierung“ befasst sich mit der Frage, ob die Angebotsmenge eines rationierten Gutes ausgeweitet werden *kann* oder nicht. Diese Unterscheidung ist schwammig, weil der Zeitbezug unklar ist. Angenommen, es gibt einen Unfall mit sehr vielen Schwerverletzten. Die Anzahl der Notärzte, die schnell vor Ort sein können, ist fix. Man kann nicht kurzfristig welche herbeibeten. Die Rationierung ist also hart. Was man aber tun kann ist, die Zahl von Notärzten insgesamt zu verändern. Wäre es in der Vergangenheit möglich gewesen, mehr Notärzte einzustellen? Ja. Man hat sich aber dagegen entschieden. Das wäre weiche Rationierung. Man könnte sich auch jetzt, unter dem Eindruck des schweren Unfalls dazu entscheiden, mehr Notärzte einzustellen. Das nutzt aber den aktuell Schwerverletzten nichts.

In solchen Fällen gibt es keine Alternative zu (harter) Rationierung. Das ist für *Katastrophenmedizin* typisch. Was dann passiert, wird *Triage* genannt und besteht in der Aufteilung der Verletzten in

1. Aussichtslos. Wird nicht überleben. Nicht behandeln.
2. Leichter Fall. Überlebt auch ohne sofortige Hilfe
3. Überlebt nur mit sofortiger Hilfe.

Die Regel ist, zuerst die Verletzten des dritten Typs zu behandeln.[1]

Wenn man das bedenkt, sprechen wir in der Gesundheitsversorgung im Regelfall über kurzfristig harte und langfristig weiche Rationierung.

2.3 Starke und schwache Rationierung

Die Unterscheidung zwischen starker und schwacher Rationierung bezieht sich auf die Frage, ob es einen *Parallelmarkt* gibt, auf dem man das Gut zu dem dort herrschenden Marktpreis kaufen kann. Bei schwacher Rationierung ist das möglich, bei starker nicht.

Ein Beispiel dafür ist das Mounjaro von Eli Lilly. Es ist als Diabetesmedikament entwickelt worden und wird in dieser Funktion auch von den gesetzlichen Kassen bezahlt. Ein großer Teil der Patienten haben keinen Diabetes, sondern nutzen es als „Abnehmspritze“. Dann wird es (Stand 2026) nicht von den Kassen bezahlt. Man kann es aber (mit einem Rezept des Arztes) in der Apotheke kaufen. Auf eigene Kosten.

Starke Rationierung ist also relativ selten, weil es wenige Gründe gibt, warum ein solcher Parallelmarkt für Selbstzahler nicht entstehen sollte. Wenn das aber passiert, sind es im Regelfall ethische Überlegungen, wie (in den 1970er Jahren) das Abtreibungsverbot oder heute einen Markt für Spenderorgane.

[1] Löwenstein, Stephan, "Wir müssen jeden Tag triagieren", Frankfurter Allgemeine Zeitung 10.3. 2021.

2.4 Direkte und indirekte Rationierung

Direkte Rationierung setzt an einer *Personengruppe* an, für die man Rationierungsregeln festlegt. Trifft man die Entscheidung, Patienten ab 75 Jahren keine künstlichen Hüftgelenke mehr einzupflanzen, dann ist das eine *direkte* Rationierung, weil klar ist, welche Personengruppe betroffen ist und welche nicht. Altersrationierung ist daher eine Form direkter Rationierung

Indirekte Rationierung bildet keine Untergruppen, sondern verändert das Rationierungsrisiko für die Gesamtgruppe. Verringert man die Anzahl der Notärzte, erhöht man das Sterberisiko für Notfälle. Ob Herr Müller, ein Notfall, schnellstmöglich behandelt wird oder ob er warten muß, ist Zufall. Je nachdem, ob gerade ein Notarzt frei ist oder nicht.

2.5 Offene und verdeckte Rationierung

Offene Rationierung liegt dann vor, wenn der Rationierte weiß, dass er rationiert wird. Bei verdeckter Rationierung ist das nicht der Fall.

Wie bei den anderen Formen von Rationierung ist auch hier die Unterscheidung nicht trennscharf. Ist eine Rationierung verdeckt, wenn die Information *grundsätzlich* verfügbar ist, der Patient diese Information aber nicht einholt? Wie leicht müsste dann diese Information verfügbar sein, damit man von offener Rationierung sprechen kann?

2.6 Warteschlangen

Warteschlangen stellen einen besonderen Fall von Rationierung dar, da es im Regelfall nicht darum geht, eine Leistung nicht zu erbringen, sondern bei zeitlichen Schwankungen der Nachfrage die Kapazitäten besser auszulasten. Dazu habe ich mir eine fiktive Verteilung von Hüft-OPs ausgedacht.

	Coxarthrose	Notfall	Summe
Quartal I	100	80	180
Quartal II	100	30	130
Quartal III	100	30	130
Quartal IV	100	60	160
Summe	400	200	600

Ein Notfall liegt vor, wenn jemand stürzt, sich die Hüfte bricht und schnell behandelt werden muß. In vielen Fällen liegt aber eine sogenannte Coxarthrose vor, d.h. eine Gelenkabnutzung, deren Behandlung zeitlich etwas verschoben werden kann.

Die höchste Fallzahl liegt im Beispiel im ersten Quartal mit 180 Fällen vor. Wenn man sich an dieser Nachfrage orientiert, wird man eine Jahreskapazität von 720 Fällen vorhalten, obwohl es insgesamt nur 600 Fälle gibt, d.h. der Auslastungsgrad wird bei 600/720=83,3% liegen. Nimmt man in Kauf, dass nicht jeder Coxarthrose-Fall in dem Quartal, in dem er diagnostiziert wird, behandelt wird, sondern einige Fälle ein Quartal warten müssen, wäre eine Kapazität von 600 ausreichend.

Diese 600 wären ein theoretischer Idealfall. Die Rechnung geht dann auf, wenn die Kapazitäten zu 100% immer verfügbar sind und die Zahl der Fälle auch exakt der Verteilung

folgt. Es darf also (um in unserem Beispiel zu bleiben) keine Jahre mit besonders viel Glatteis geben, in denen mehr Menschen stürzen und sich die Hüfte brechen. Um solche ungeplanten Schwankungen aufzufangen, ist es sinnvoll, keine Auslastung von 100% zu planen, sondern vielleicht 5% Leerkapazitäten zu kalkulieren.

2.7 Ewige Warteschlangen

Warteschlangen, die der Kapazitätsauslastung dienen, bauen sich in Phasen hoher Nachfrage auf und in Phasen niedriger Nachfrage ab. Man kann sie daran erkennen, daß es regelmäßig Zeitpunkte gibt, bei denen es *keine* Warteschlange gibt.

Davon unterscheiden muß man Warteschlangen, die sich nicht abbauen, sondern immer vorhanden sind. Eine solche Warteschlange könnte auf einen Kapazitätsengpass in der Vergangenheit zurückgehen, der über lange Zeit mitgeschleppt wird. Das wäre ein Fall schlechten Managements, weil man eine solche Warteschlange durch zeitlich begrenzte Kapazitätsausweitung beseitigen könnte.

Wenn diese Schlange durch zu geringe Kapazitäten verursacht wird, müßte die Schlange immer länger werden. Wenn es pro Jahr 600 Hüft-OP-Fälle gibt, aber nur 500 OP-Termine, müßte die Schlange jedes Jahr um 100 Fälle wachsen.

Wenn man dauerhaft Warteschlangen beobachtet, diese Schlangen aber mehr oder minder konstant bleiben, liegt dieser Schlange vermutlich folgende Idee zugrunde:

> *Wenn Personen mit einer Diagnose sterben, während sie sich in der Warteschlange befinden, kann man mit weniger OP-Terminen als Diagnosen auskommen. Man spekuliert also darauf, dass nicht jeder, der wartet, auch behandelt wird.*

Dieses Kalkül ist nicht ganz so zynisch, wie es zunächst scheint. Angenommen Coxarthrose-Patienten müssen grundsätzlich ein Quartal warten, bis sie behandelt werden. Viele dieser Patienten sind betagt und von 400 Fällen versterben (nehmen wir an) 100, während sie warten. Das bedeutet, daß man statt 600 Fällen nur 500 Fälle pro Jahr behandeln muß. Diesen Einsparungen von 100 Fällen steht ein Schaden gegenüber, der den Rationierten entstanden ist.

1. 100 Personen sind gar nicht behandelt worden, weil sie in der Schlange gestorben sind. Das sind die Personen, die von einer Behandlung aber auch nicht mehr viel gehabt hätten, da sie sich sowieso am Ende ihres Lebens befunden haben. Der Schaden, der dieser Gruppe entstanden ist, ist also relativ klein.

2. 300 Personen mit Coxarthrose mußten ein Quartal warten, bis sie behandelt wurden. Der Schaden, der dieser Gruppe entstanden ist, besteht in der Wartezeit.

Diese Form von Warteschlange ist eine Variante indirekter Rationierung, die die Patienten rationiert, die nur sehr wenig Nutzen aus der Behandlung ziehen würden.

Kapitel 3

Rationierung und Ethik

3.1 Erst beschreiben, dann bewerten

In der vorangegangenen Einheit haben wir uns mit verschiedenen Formen von Rationierung beschäftigt. Wir haben uns nicht die Frage gestellt, ob man das „darf“, ob das „gerecht“ ist usw. Das haben wir auf diese Einheit verschoben. Ich halte diese Reihenfolge für sinnvoll, weil dann klarer ist, *worüber* wir sprechen, wenn wir von „Rationierung“ sprechen. Wir können dann immer noch zu dem Ergebnis kommen, dass wir das nicht wollen, aber dann wissen wir zumindest, *was* wir nicht wollen.

3.2 Eine Arbeitsdefinition von Ethik

Als Ökonomen sind wir gewohnt, mit sehr klaren Begriffen zu arbeiten. Wenn wir z.B. von *Deckungsbeitrag* sprechen, dann gibt es keine große Diskussion darüber, was mit diesem Begriff gemeint ist. Der Deckungsbeitrag ist $DB = U - K_v$ also die Differenz zwischen Umsatz und variablen Gesamtkosten. Es gibt da keine rivalisierenden Schulen, die eine andere Definition verwenden. Wenn es Uneinigkeit gibt, dann vielleicht darüber, in welchen Situationen der DB eine besonders relevante Kennzahl ist und in welchen nicht.

In der Ethik ist das anders. Es gibt eine größere Anzahl konkurrierender Ansätze, die „Ethik“ oder „Gerechtigkeit“ jeweils anders definieren.

Vor diesem Hintergrund möchte ich Ethik als *gesellschaftliche Spielregeln für Gerechtigkeit* definieren. Das ist keine besonders weitreichende Definition, aber für unsere Zwecke hier ausreichend. Ich möchte auch „Ethik“ und „Moral“ als Synonyme verwenden. Teilweise wird zwischen diesen Begriffen unterschieden, teilweise nicht. Für unsere Zwecke müssen wir sie nicht unterscheiden.

3.3 Utilitarismus als Ausgangspunkt

Die Grundidee

Die Ausgangsthese des Utilitarismus ist, dass jeder Mensch nach einer möglichst hohen Lebensqualität strebt. Das ist nicht besonders spannend und neu. Meist wird ein solcher Lebensstil als *Hedonismus* bezeichnet.

Was neu ist, ist die Überlegung, dass dies nicht nur das Ziel eines einzelnen Menschen ist, sondern auch das einer ganzen Gesellschaft sein sollte. Es geht also um eine möglichst hohe Lebensqualität der gesamten Gesellschaft. Die klassische Formulierung für dieses Ziel ist

Das größte Glück der größten Zahl

Dieses Ziel ergibt nur dann Sinn, wenn man Lebensqualität *personenübergreifend* miteinander vergleichen kann. Dann kann man auch den Schaden von Alice mit dem Nutzen von Bob vergleichen und saldieren. Diese Denkweise liegt Ökonomen sehr nahe, weil es dann letztlich um die Maximierungsregel des Effizienzprinzips aus *Einheit* **??** geht.

Fallbeispiel Corona - Impfung

Anfang 2021 wurden die ersten Corona-Impfstoffe zugelassen. Weil es viel mehr Impfwillige als Impfstoff gab und das Impfen selbst auch Zeit brauchte, stellte sich die Frage, wer *zuerst* geimpft werden sollte. Damals fiel die Entscheidung zugunsten der Hochbetagten, weil die ein höheres Mortalitätsrisiko hatten.

Diese Begründung ist eine Anwendung des Effizienzprinzips: *Minimiere die Zahl der Toten bei gegebener Impfkapazität.*

Das Problem an dieser Vorgehensweise ist, dass es kein Nachdenken darüber gab, ob „wenig Tote“ ein besonders gutes Ziel ist.

Ein genauso denkbares Ziel wäre *„minimiere die Zahl der durch Corona verlorenen Lebensjahre“* gewesen. Dieses Ziel ist nicht mit dem ersten identisch. *Einheit Lebenserwartung* befasst sich mit der Berechnung der Lebenserwartung. Das müssen wir hier nicht vertiefen, aber ich greife auf Ergebnisse dieser Methode zurück: Im Durchschnitt lebt ein 80jähriger in Deutschland noch 9 Jahre und ein 20jähriger noch etwa 62 Jahre.

	A	B	C	D	E
1	Alter		**20**	**80**	**Summe**
2	Anzahl		1.000	1.000	2.000
3	Restlebenserwartung		62	9	
4	Szenario "Keine Impfung"				
5		Sterbewahrscheinlichkeit	0,1%	0,5%	
6		Tote	1	5	6
7		verlorene Lebensjahre	62	45	107
8	Szenario "Alte impfen"				
9		Sterbewahrscheinlichkeit	0,1%	0,0%	
10		Tote	1	-	1
11		gerettete Leben	-	5	5
12		gerettete Lebensjahre	-	45	45
13	Szenario "Junge impfen"				
14		Sterbewahrscheinlichkeit	0,0%	0,5%	
15		Tote	-	5	5
16		gerettete Leben	1	-	1
17		gerettete Lebensjahre	62	-	62

Mit diesen dürren Zahlen können wir eine Überschlagsrechnung machen: Ich habe eine Sterbewahrscheinlichkeit im Falle einer Corona-Infektion bei den Jungen von einem Promille angenommen und bei den Alten eine fünffach höhere. Weiterhin habe ich angenommen, dass eine Impfung eine Erkrankung zuverlässig verhindert. Diese Zahlen sind grobe Schätzungen, aber bessere Zahlen wären damals nur erhältlich gewesen, wenn man ein paar tausend Personen hätte sterben lassen, um dann Wahrscheinlichkeiten zu erhalten. Vielleicht nicht die beste Idee.

Bei *diesen* Zahlen sehen wir, dass wir, wenn wir die Alten zuerst impfen, *weniger Tote* haben werden, aber auch *weniger Lebensjahre retten* werden.[1] Dieses Beispiel macht uns deutlich, dass Utilitarismus nicht über die *Inhalte* von Moral spricht, sondern über die *Methode* mit der man moralisches Handeln identifizieren kann. Weil das bedeutet, dass man Handlungen nach ihren Konsequenzen beurteilt, ist eine Alternativbezeichnung für Utitlitarismus auch *Konsequentialismus*.

Fingerübung: Nach meiner Rechnung wäre es unter dem Aspekt „rette möglichst viele Lebensjahre“ besser gewesen, die Jungen zuerst zu impfen. Weil die Mortalitätszahlen damals wackelig waren, hätte man auch anders an die Frage herangehen können: Es war klar, dass die Mortalitätsrate bei den ungeimpften Alten höher sein würde als bei den Jungen. Man könnte im Rahmen der Tabelle, die ich zusammengebastelt habe, ausrechnen, wie viel höher diese Rate sein müsste, damit die Strategie „die Alten zuerst“ die bessere gewesen wäre. Machen Sie das mal.

Einkommenstransfers

Glücklicherweise geht es im Utilitarismus nicht immer gleich um Leben und Tod. Das zweite Beispiel soll die Frage sein, ob es nicht sinnvoll sein könnte, Einkommen über Steuern umzuverteilen.

Stellen wir uns zwei Menschen vor. Einen Armen und einen Reichen und stellen uns vor, was passieren würde, wenn wir dem Reichen 500€ über Steuern wegnehmen und dem Armen einen Transfer in gleicher Höhe geben.

	der Arme	der Reiche
Einkommen vor Steuern in €	1.000	100.000
Umverteilung in €	Transfer: 500	Steuern: 500
Änderung der Lebensqualität	hoher Gewinn	geringer Verlust

Das Easterlin-Paradox *(Einheit **??**)* sagt uns nun, dass der Verlust der 500€ für den Reichen weniger schwer wiegt als der Gewinn für den Armen. Ein Utilitarist würde daher diese Umverteilung befürworten, weil die Lebensqualität in Summe gestiegen ist.

Anwendung auf das Gesundheitssystem

Auf das Gesundheitssystem bezogen bedeutet Utilitarismus bzw. *das größte Glück der größten Zahl*, daß es kein grundsätzliches Anrecht eines Patienten auf eine Maximalversorgung gibt, sondern der Umfang der Versorgung sich danach bemessen muss, welchen

[1] Was wir hier tun, ist in *Opportunitätskosten* zu denken: *Wenn wir die Alten impfen, werden Junge sterben.* Dieses Konzept haben Sie bereits in *Einheit **??*** kennengelernt.

Nutzen er bei welchen Kosten stiftet. Das kann im Extremfall dazu führen, daß lebensrettende Leistungen nicht erbracht, d.h. rationiert werden.

In den *Grundlagen der Gesundheitsökonomie* haben wir uns ausführlich mit der *Einnahmeseite* der GKV befasst. Sie basiert (mit Abstrichen) auf dem Leistungsfähigkeitsprinzip, das Sie in *Einheit* **??** kennengelernt haben.

> *Wer mehr Geld hat, soll höhere Beiträge zahlen.*

Dieses Prinzip lässt sich ethisch nur dann begründen, wenn man utilitaristisch denkt und annimmt, dass der Schaden der bei den Höherverdienenden entsteht, kleiner ist als der Nutzen bei den Geringverdienern.

Wenn aber die *Einnahmeseite* der Krankenversicherung ein utilitaristisches Denken voraussetzt, muss man es konsequenterweise auch auf der Ausgabenseite anwenden.

Anders herum: Wenn man dieses Denken auf der Ausgabenseite nicht anwenden will, müsste man es auch auf der Einnahmeseite unterbinden.

Diese Überlegung bringt mich dazu, auf der Ausgabenseite nicht nur nach *„was bringt es?“* zu fragen, sondern auch nach *„was kostet es?“*.

3.4 Endlichkeit

Außer in Sciencefiction-Romanen ist das Leben endlich. Man kann spekulieren, ob die Grenzen der Lebenserwartung im Wesentlichen erreicht sind, oder noch erheblicher Spielraum besteht.[2]

In jedem Fall ist der Tod unausweichlich. Das bedeutet, daß medizinische Versorgung, auf lange Frist gesehen, keinen Erfolg hat.

> *Medizin rettet kein Leben, sondern verschiebt im Erfolgsfall den Tod weiter in die Zukunft.*

[2]Diese Frage werden wir in *Einheit Lebenserwartung* am Rand anschneiden.

Wenn die Hauptkonfliktlinie der ethischen Diskussion um Rationierung ist, ob es legitim sein kann, lebenswichtige Leistungen zu rationieren, dann sind die Alternativen genau genommen nicht „Patient lebt“ vs. „Patient stirbt“, sondern „Patient stirbt später“ vs. „Patient stirbt früher“.

Das ist zwar keine angenehme Überlegung, aber sie macht den Zugang zu der Frage leichter, welche Kosten für welche Lebensverlängerung legitim sind, weil *„Leben retten“* kein erreichbares Ziel ist.

Diese leider betrübliche Aussicht spielt z.B. bei der Entwicklung neuer Medikamente eine Rolle. In *Einheit* **??** haben wir eine ökonomische Erklärung gefunden, warum viele Pharmafirmen einen Forschungsschwerpunkt in der Onkologie setzen. Viele dieser Medikamente werden in der Spätphase der Erkrankung eingesetzt, in der es nicht mehr um Heilung geht, sondern um den Gewinn an zusätzlicher Lebenszeit. Studien zeigen, dass man diesen Zugewinn in einigen Fällen eher in Tagen als in Wochen messen sollte. Egal, wie wir das Problem drehen und wenden: Irgendwann wird dieser Zugewinn so gering werden, dass wir als Gesellschaft nicht mehr bereit sind, sagen wir 100.000€ für, sagen wir, zwei Tage zu zahlen. Auf der anderen Seite wäre es komisch, wenn wir nicht bereit wären, 5€ auszugeben und damit 50 Jahre Lebenszeit zu gewinnen. Wir müssen schauen, dass wir Regeln für einen sinnvollen Mittelweg finden.

3.5 Konsum, Gesundheit und schwache Rationierung

Budgetgerade

Rationierung im Gesundheitssystem bedeutet (nach der Systematik, die wir in der letzten Einheit kennengelernt haben) im Regelfall *weiche* Rationierung. Ein Patient erhält eine medizinische Leistung nicht, weil entschieden worden ist, daß er sie nicht bekommen *soll*. Grundsätzlich wäre es möglich, die Leistung zu erbringen.

Die Begründung einer solchen Rationierung kann nur darin bestehen, daß die Ressourcen, die für die Behandlung aufgebracht werden müssten, an anderer Stelle mehr Nutzen stiften.

Da Ressourcen knapp sind, muß entschieden werden, wofür sie alternativ eingesetzt werden sollen. Diese Überlegung kann man mit dem Instrumentarium der mikroökonomischen Haushaltstheorie darstellen.

Als Haushalt oder Volkswirtschaft steht ein gegebenes Budget zur Verfügung. Dieses Budget kann man theoretisch komplett für Gesundheit ausgeben oder komplett für andere Dinge, die in der Abbildung mit „Konsum“ zusammengefasst werden.

Die gestrichelte Linie ist die *Budgetgerade*, die alle theoretisch möglichen Aufteilungen zwischen diesen zwei Verwendungszwecken abbildet.

Falls Sie mit diesem Instrument nicht vertraut sind, habe ich Ihnen eine kleine Tabelle mit Zahlen, die grob den Werten für Deutschland 2021 (in Mrd. €) entsprechen aufgelistet.[3]

Ausgaben Gesundheit	Ausgaben Konsum	$\sum$
1.724	0	1.724
1.700	24	
1.600	124	
862	862	
465	*1.259*	
124	1.600	
24	1.700	
0	1.724	

Wenn Sie diese Punkte in ein Diagramm auf einem Blatt Millimeterpapier übertragen, werden Sie sehen, dass diese Punkte alle auf einer Linie liegen. Der Budgetgerade, weil ich nur acht denkbare Aufteilungen aufgelistet habe.

Jeder Punkt auf dieser Gerade mit Ausnahme des Extrempunktes 100% Ausgaben für Gesundheit stellt eine Form weicher Rationierung dar. Man könnte mehr Geld für Gesundheit ausgeben, tut es aber nicht, sondern gibt das Geld lieber für Konsumzwecke aus.

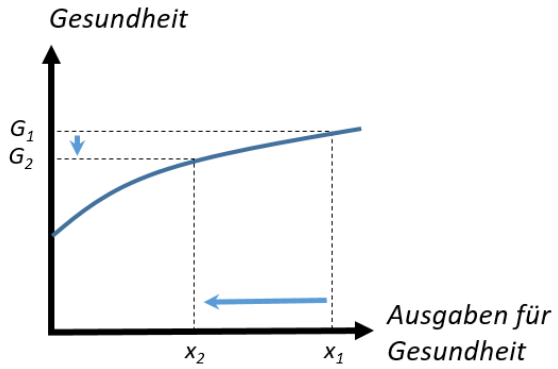
Abnehmender Ertragszuwachs

Das Gesetz[4] des abnehmenden Ertragszuwachses besagt, dass eine Erhöhung des Inputs zwar zu einer Erhöhung des Outputs führt, aber nur zu einer unterproportionalen Erhöhung. In unserem Kontext bedeutet es, dass der Nutzen der medizinischen Versorgung zwar mit den Gesundheitsausgaben steigt, aber dieser Zuwachs unterproportional ist. Wenn wir doppelt so viel Geld für Gesundheit ausgeben, sind wir nicht doppelt so gesund.

Unser Thema ist aber nicht *mehr Geld für Gesundheit*, sondern *Rationierung*. Wir schauen uns diese Grafik also quasi „von rechts“ an, also von einem Punkt $x_1$aus, an dem wir relativ viel Geld für Gesundheit ausgeben.

[3] Die Zahlen sind grob gerundet und die Abgrenzung ist nicht ganz sauber.

[4] Wie immer darf man den Begriff „Gesetz“ in den Sozialwissenschaften nicht wörtlich nehmen. Gesetz bedeutet *Regel, die in den allermeisten Fällen gilt. Aber nicht überrascht sein, wenn es Gegenbeispiele gibt.*



Wenn wir von diesem Punkt ausgehend Leistungen rationieren und die Ausgaben auf x_2 halbieren, führt das ganz klar zu einem Rückgang der Gesundheit. Die Frage ist nur, wie groß dieser Rückgang ist. So wie ich (etwas manipulativ) die Kurve gezeichnet habe, ist dieser Rückgang von $G_1 \rightarrow G_2$ ziemlich gering. Verallgemeinert bedeutet das, dass Rationierung umso weniger Schaden anrichtet, je höher das ursprüngliche Leistungsniveau des Gesundheitssystems ist.

Anders herum: Wenn wir bereits sehr viel Geld für Gesundheit ausgeben, dann wird weiteres Geld für Gesundheit wahrscheinlich nicht sehr viel bringen.

Opportunitätskosten

In *Einheit* **??** haben Sie das Konzept der Opportunitätskosten kennengelernt. Auch in unserem Zusammenhang hier ist es nützlich, in dieser Kategorie zu denken. Wenn wir relativ viel Geld für Gesundheit ausgeben, bleibt relativ wenig Geld für andere (Konsum)zwecke. Das Gesetz des abnehmenden Ertragszuwachses gilt natürlich auch für Konsumausgaben.

Vor diesem Hintergrund wird klar, dass es weder sinnvoll ist, sein Geld komplett für Konsumzwecke auszugeben, aber auch nicht komplett für Gesundheit. Die Frage nach der sinnvollsten Aufteilung könnte man etwas technisch so beantworten:

> *So lange die Ausgabe von einem Euro für Gesundheit die Lebensqualität mehr steigert als bei einer Ausgabe dieses Euros für Konsumzwecke, sollte man den Euro für Gesundheit ausgeben.*

Schwache Rationierung

Diese Regel kann man bei individuellen Kaufentscheidungen ziemlich gut anwenden. Alice steht vor der Entscheidung, ob sie sich ein teures Handy kaufen soll oder eine Urlaubsreise machen möchte. Als Ökonomen gehen wir davon aus, dass sie die Alternative wählen wird, die ihr den größeren Nutzen stiftet.

Leider funktioniert diese Regel bei Gesundheit nicht sehr gut, weil die Finanzierung von Gesundheitsleistungen im Wesentlichen nicht über privat finanzierte Nachfrage stattfindet, also Alice, die das Handy im Laden bezahlen müsste, sondern über Versicherungen.[5]

[5]Eine Ausnahme davon ist das, was ich den *Dacia-Effekt* nennen möchte. Das Geschäftsmodell der Automarke Dacia besteht darin, sehr günstige Autos anzubieten, deren Preis vorwiegend dadurch so niedrig ist, weil das Sicherheitsniveau der Autos unterdurchschnittlich ist. Die deutschen Hersteller achten penibel darauf, dass ihre Modelle durchgängig 5/5 NCAP-Sicherheitspunkte bekommen. Die Dacia - Modelle haben durchgängig nur 3 Punkte. Etwas dramatisch formuliert sparen Dacia Fahrer an ihrer Gesundheit und an ihrer Überlebenswahrscheinlichkeit, weil ihnen das gesparte Geld wichtiger ist als ein

In *Einheit* **??** haben wir darüber nachgedacht, warum diese Entkoppelung von Nachfrage und Finanzierung sinnvoll ist.

In *Einheit* **??** haben wir das Konzept von *Versicherungen* kennengelernt, mit dem man diese Entkoppelung organisieren kann. Eine Alternative wäre eine *Steuerfinanzierung*. Großbritannien macht das so und in *Einheit Das Gesundheitssystem in Großbritannien* werden wir einen Blick auf dieses System werfen. In beiden Fällen handelt es sich um eine institutionalisierte Solidarität der Versicherten/Steuerzahler miteinander. Etwas naiv formuliert legen alle Teilnehmer Geld in einen großen Topf aus dem dann die Leistungen für die finanziert werden, die sie brauchen. Diese Solidarität ist aber nicht grenzenlos. Es gibt daher Spielregeln, für *was* man Geld aus diesem Topf nehmen kann und für was nicht. Ich fokussiere hier auf die Geldebene, weil es, wenn wir über Rationierung sprechen, im Regelfall tatsächlich um *Geld* geht und nicht um die *Leistung*, die mit dem Geld finanziert werden soll. Genauer betrachtet bedeutet Rationierung dann

> *Das Gesundheitssystem bezahlt eine medizinische Leistung nicht, weil entschieden worden ist, dass diese Leistung nicht vom Gesundheitssystem bezahlt werden soll. Wenn der Patient sie trotzdem haben möchte, kann er das tun, wenn er selbst die Kosten trägt.*

3.6 Schleier des Nichtwissens bei Rawls

Der Schleier des Nichtwissens ist ein Gedankenexperiment, das aus der politischen Philosophie stammt. Wir schauen uns zunächst die Originalidee an und stricken Sie dann auf unsere Frage der Rationierung im Gesundheitssystem um.

Gerechte Einkommensverteilung

Eine zentrale Frage, mit der sich der politische Philosoph John Rawls in seinem Buch *Eine Theorie der Gerechtigkeit* aus dem Jahr 1971 befasste, war, welche Eigenschaften eine gerechte Einkommensverteilung haben würde.

Wenn man Menschen fragt, liegt es nahe, dass man die Antworten bekommt, die dem Befragten selbst am angenehmsten sind. Um dieses Eigeninteresse zu neutralisieren, dachte er sich das Gedankenexperiment des *Schleiers des Nichtwissens* aus.

Der Schleier des Nichtwissens

Man solle sich vorstellen, daß man über die Einkommensverteilung abstimmen müsse, ohne daß man wüsste, welchen Platz man in dieser Verteilung einnehmen würde. Man wüsste nicht, ob man arm, reich, Mann, Frau, Europäer, Afrikaner, ... ist.

Die Spielregel, für die sich die Menschen unter diesen Bedingungen entscheiden würden, wäre die „richtige“, weil sie durch die freiwillige Einigung aller Menschen zustande kommen würde.

Die Lösung

Rawls war der Auffassung, daß sich die Menschen in diesem Gedankenexperiment bei einer Wahl zwischen zwei Verteilungen immer für die Verteilung entscheiden würden, bei

noch höheres Sicherheitsniveau.

der die ärmste Person das höhere Einkommen hätte. Diese Regel hat er *Differenzprinzip* genannt.

	alle Anderen	der Ärmste
A	100	10
B	500	x

Wenn wir von der Einkommensverteilung A ausgehen, ist Verteilung B nur dann gerecht, wenn x mindestens 11 ist. Wenn x bei 10 bleibt, ist es unerheblich, dass alle anderen ihr Einkommen verfünffachen. Relevant ist die Frage, ob es dem Ärmsten besser geht.

Die Kritik

Er hat diese These nicht empirisch getestet, sondern angenommen, dass ja klar wäre, dass jeder Mensch in der von ihm konstruierten Situation für diese Regel stimmen würde. Unter Ökonomen hat das Stirnrunzeln erzeugt.[6]

Die Grundidee von Rawls war, dass die Menschen das Differenzprinzip allein aus Eigeninteresse verfolgen würden. Dann bräuchte man keinen Rückgriff auf übergeordnete moralische Prinzipien oder auf Mitleid.

Als Ökonom fällt es einem nicht schwer, den Schleier des Nichtwissens als Wette darauf zu sehen, dass man der Ärmste der gesamten Gesellschaft sein wird, bzw. eben *nicht* der Ärmste.

Spielen wir den Fall für die kleinstmögliche Gesellschaft durch: Es gibt *zwei* Personen und das Risiko, der Ärmste zu sein liegt damit bei 50%. Wenn wir uns dann die beiden Verteilungen A und B anschauen und $x = 9,9$ ist, ist es schwer vorstellbar, dass irgendjemand das Differenzprinzip wählen würde. Stattdessen nimmt er für die Chance auf ein viel höheres Einkommen das Risiko eines kleinen Einkommensverlusts in Kauf.

Anwendung auf Gesundheits - Leistungskataloge

Ich finde, dass dieses Gedankenexperiment eigentlich auf die Frage nach einem gerechten Gesundheitssystem noch viel besser passt als auf die Frage nach der gewünschten Einkommensverteilung.

In *Ihrem* Alter (Anfang/Mitte 20) ist ihr langfristiger Gesundheitszustand tatsächlich viel unklarer als ihre langfristige Einkommenshöhe. Der Schleier der Unwissenheit ist also für Gesundheit viel weniger abstrakt als für Einkommen.

Um den „Schleier des Nichtwissens“ für unsere Frage zu operationalisieren, habe ich mir sechs alternative Leistungskataloge (A-F) für ein Gesundheitssystem ausgedacht, über die abgestimmt werden soll.

[6] Harsanyi, John C. Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory, 1975.

	Versorgung gut	Versorgung nicht gut	Rationierungsrisiko	Prämie je Monat
A	90%	10%	$^1/_{10}$	50€
B	99%	1%	$^1/_{100}$	160€
C	99,9%	0,1%	$^1/_{1.000}$	300€
D	99,99%	0,01%	$^1/_{10.000}$	500€
E	99,999%	0,001%	$^1/_{100.000}$	1.000€
F	100%	0%		2.000€

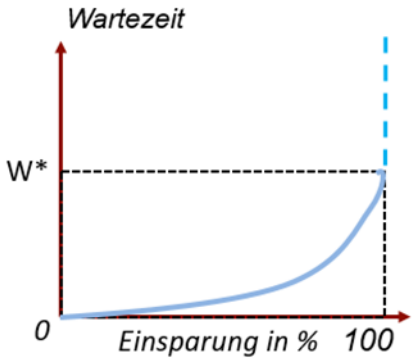
Für A-E gilt, daß es einen Teil der Bevölkerung gibt, der nicht gut versorgt werden wird, d.h. dem theoretisch verfügbare Leistungen vorenthalten werden, d.h. der rationiert werden würde. In den Zahlen der Tabelle steckt die Annahme, daß das Gesetz des abnehmenden Ertragszuwachses auch in der medizinischen Versorgung gilt, d.h. daß es immer teurer wird, einen noch etwas größeren Anteil der Bevölkerung gut zu versorgen.

Vergleichen wir jetzt die Szenarien *E* und *F*. In *E* ist der Leistungsumfang so gestaltet, dass eine von 100.000 Personen rationiert wird. Die Prämie für dieses Leistungspaket liegt bei 1.000€/Monat. Wenn man dieses Risiko von $^1/_{100.000}$ nicht tragen will, muss man für Paket *F* stimmen, das 1.000€ mehr kostet als *E*.

Dieses fiktive Beispiel macht deutlich, daß es nicht plausibel ist, anzunehmen, daß die Entscheider bereit sind, *jeden* Betrag zu zahlen, um Ihr Risiko der Rationierung auf Null zu drücken. Das bedeutet aber, daß die Personen in Unkenntnis ihres künftigen Gesundheitszustands bereit sind, das Risiko einer möglichen Rationierung zu tragen, wenn sie dadurch hinreichend hohe Einsparungen bei den Versicherungsbeiträgen erhalten können.

3.7 Ehemalige Prüfungsaufgaben

Warteschlange II



Die Kontrollfrage im letzten Kapitel befasste sich mit der geometrischen Darstellung von Wartezeit und prozentualen Einsparungen bei Warteschlangen, in denen Patienten während des Wartens versterben. Bei dieser Darstellung sind zwei Punkte klar: Die Kurve beginnt im Ursprung und verläuft ab einer Wartezeit von W* und einer Einsparung von 100% (alle Patienten versterben in der Schlange) senkrecht. Über den Verlauf zwischen diesen Punkten kann man nur spekulieren.

Die Darstellung bei den Lösungen nahm einen linearen Verlauf an. In der folgenden Abbildung ist der Verlauf nichtlinear. Wenn man sich dafür entschieden hat, Warteschlangen einzuführen, damit man Geld sparen kann, indem Patienten während des Wartens versterben, ist der nebenstehende Verlauf ethisch eher zu begründen als der lineare Verlauf. Warum?

Warteschlange II (Lösung)

Der Verlauf der Kurve unterstellt, daß ein Großteil der möglichen Einsparung durch Mortalität schon nach sehr kurzer Wartezeit eintritt. Das bedeutet, daß ein großer Teil der Patienten nur noch eine geringe Restlebenserwartung hat und der Verlust an Lebenszeit durch die Schlange nicht sehr hoch ist.

Teil II

Methoden der Effizienzmessung

In den letzten Einheiten haben wir uns mit der Legitimität von Rationierung befasst. Das aus meiner Sicht stärkste Argument ist das, dass wir uns als Versicherte freiwillig selbst rationieren, wenn wir unseren künftigen Gesundheitszustand nicht kennen. Dann werden wir eine Art Lotterie auf unsere eigene Gesundheit spielen. Das bedeutet nicht, dass wir russisches Roulette spielen werden, aber es bedeutet, dass wir keinen Leistungsumfang finanzieren wollen, der auch absurd unwahrscheinliche Fälle abdeckt.

Das hilft uns insofern weiter, als dass es bei der Frage nach Rationierung nicht um das *„ob“* gehen wird, sondern um das *„was“*. Als Ökonomen liegt uns die Antwort nahe: Als erstes sollten wir den Rotstift bei Leistungen ansetzen, die viel kosten und wenig bringen. Wenn das unsere (zunächst sehr schwammige) Regel sein soll, reden wir über Effizienz.

Kapitel 4

Kostendefinitionen

F4.1 Der rote Faden

Die Frage nach *Effizienz* steht ganz am Anfang eines Ökonomiestudiums. Effizienz ist als $\frac{Ergebnis}{Aufwand}$ definiert. In den folgenden Abschnitten werden wir sehen, dass es nicht einfach ist, festzulegen, wie man das *Ergebnis* mißt. Daher wollen wir uns in dieser kleinen Einheit kurz mit dem einfacheren Teil des Bruchs, dem *Aufwand* befassen.

Es ist zweckmäßig, im Bruch $Effizienz = \frac{Ergebnis}{Aufwand}$, den Aufwand in Geldeinheiten zu messen. Die Frage ist aber, welche Kosten einer Behandlung zuzurechnen sind und welche nicht. Hierzu hat sich eine Systematik, die der Kostenrechnung entliehen ist, etabliert.

F4.2 Direkte und indirekte Kosten

1. Direkte Kosten sind die Kosten, die unmittelbar mit der Behandlung zusammenhängen, also z.B. Kosten der OP, Rehabilitation und evtl. anschließende Dauermedikation.

2. Indirekte Kosten sind die Kosten, die im Umfeld der Behandlung entstehen, also z.B. Einkommensausfall.

F4.3 Zeitliche Dimension

Eine komplexere Frage ist, welchen Zeithorizont die Ermittlung der Kosten haben soll. Nehmen wir den Fall einer Herztransplantation. Es gibt Kosten, die unmittelbar für die Operation anfallen. Anschließend muß der Patient aber lebenslang Medikamente nehmen, um zu verhindern, daß das Organ abgestoßen wird. An dieser Stelle ist es sinnvoll, zwischen dem Zeitraum zu unterscheiden, an dem der Patient ohne Behandlung gestorben wäre und dem Zeitraum, der die gewonnene Lebenszeit umfasst. Diese Überlegung führt zu folgender Matrix.

	direkte K.	indirekte K.	andere Ges.ausg.
während der Behandlung	I	IV	VII
bis zum Tod ohne Behandlung	II	V	VIII
in der gewonnenen Lebenszeit	III	VI	IX

Es ist klar, daß die Kosten in den Feldern *I-VI* den Kosten der Behandlung zählen und es ist auch klar, daß die Felder *VII* und *VIII* nicht zu den Kosten der Behandlung zählen.[1]

Der strittige Punkt ist, wie mit Feld *IX* umgegangen werden soll.

Man kann sich auf den Standpunkt stellen, daß die Behandlung heute zu einem Lebenszeitgewinn von (sagen wir) 10 Jahren führt und der Patient in diesen 10 Jahren auch Kosten verursachen wird, die ohne die Behandlung nicht angefallen wären, weil er dann schon tot wäre. Also sind das Kosten, die auf die Entscheidung zur Behandlung heute zurückzuführen sind und den Kosten der Behandlung zuzurechnen sind. Die Gegenposition (*IX* gehört nicht zu den Kosten der Behandlung) wird ebenfalls vertreten. Vergleicht man die Kosten verschiedener Behandlungen ist es daher wichtig, daß mit den Kosten in Feld *IX* einheitlich umgegangen wird.

F4.4 Friktionskosten

Als wir über indirekte Kosten gesprochen haben, war ein Beispiel für diesen Kostentyp der Ausfall an Einkommen. An dieser Überlegung möchte ich anknüpfen.

Bob hat einen Herzinfarkt erlitten und überlebt und befindet sich jetzt in der Reha. Die Herz-OP hat Geld gekostet, die Reha jetzt kostet Geld und bis er wieder arbeitsfähig ist, bekommt er Einkommensersatzleistungen, d.h. Lohnfortzahlung vom Arbeitgeber und anschließend Krankengeld von der Krankenkasse. So weit so gut. Das kann man alles addieren.

Die Friktionskostenmethode[2] stellt aber nun folgende Überlegung an: Durch die Arbeitsunfähigkeit von Bob gibt es Produktionsausfälle, d.h. auf volkswirtschaftlicher Ebene sinkt das BIP. Auch das sind Kosten, so der Ansatz, die berücksichtigt werden müssen. Die Frage ist nur, wie. Das Ergebnis ist $\frac{medizinisches\,Ergebnis + vermiedene\,Kosten}{Kosten}$. Der Bruch wirkt etwas verwirrend, weil in Zähler wie Nenner Kosten auftauchen. Im Zähler sind es allerdings *vermiedene* Kosten, also Opportunitätskosten. Wenn Bob schneller (oder überhaupt) wieder arbeiten kann, dann sind die Produktionsausfälle niedriger. Dann muß dieser Betrag aber auf der *„was-bringt-es“* Seite des Bruchs stehen und nicht auf der *„was-kostet-es“* Seite.

[1] Wenn sich der Patient jedes Jahr gegen Grippe impfen läßt, fallen diese Kosten unabhängig von der Herztransplantation an. Es ist also nicht sinnvoll, diese Kosten der Behandlung zuzurechnen.

[2] Der klassische Aufsatz dazu ist: Koopmanschap, Marc A. und Ineveld, B. Martin van, "Towards a New Approach for Estimating Indirect Costs of Disease", Social Science and Medicine 1992. S.1005-1010.

Diese Einbeziehung indirekter Opportunitätskosten ist umstritten, weil nach dieser Rechnung die Effizienz der gleichen Operation bei einem Erwerbstätigen höher ist als bei einem Rentner. Unabhängig von der Restlebenserwartung. Allein durch den Umstand, dass der Rentner keine Produktivitätsausfälle erzeugt.

Diese Bevorzugung von Erwerbstätigen hört sich asozialer an, als sie ist, denn der Erwerbstätige ist ja die Finanzierungsquelle des ganzen Systems. Das bedeutet nicht, „wer zahlt, bestimmt“, sondern es bedeutet, dass es volkswirtschaftlich unklug wäre, wenn die Erwerbstätigen keine Vorzugsbehandlung bekämen, denn ohne gesunde Erwerbstätige, die fleißig BIP erzeugen, kann das Gesundheitssystem nicht finanziert werden und damit auch nicht die Versorgung der Nichterwerbstätigen.

Kapitel 5

Ergebnisdefinitionen

F5.1 Der rote Faden

Wenn $Effizienz = \frac{Ergebnis}{Aufwand}$ ist, haben wir uns in der letzten Einheit mit der Frage nach dem *Nenner* des Bruchs befasst. Das war verhältnismäßig einfach. Komplizierter ist die Frage, was das Ergebnis einer medizinischen Leistung ist, also der *Zähler*. Hierzu gibt es mehrere konkurrierende Alternativen.

1. Die Kosten-Nutzen-Analyse versucht, das Ergebnis monetär zu bestimmen. In der Finanzwissenschaft ist dieses Instrument verbreitet, in der Gesundheitsökonomie ist es eher ungebräuchlich. Wir werden uns trotzdem etwas ausführlicher mit diesem Ansatz befassen, weil er für die Frage nach dem Umfang medizinsicher Infrastruktur hilfreich ist.

2. Die Kosten-Wirksamkeits-Analyse misst das Ergebnis in medizinischen Einheiten. Das ist bei der Therapiewahl nützlich, hilft aber bei der ökonomischen Frage, auf welche Behandlungen am ehesten verzichtet werden kann, nicht weiter.

3. Die Kosten-Nutzwert-Analyse versucht, das Ergebnis über die gewonnene Lebensqualität zu messen. Dieser Ansatz ist am erfolgversprechendsten, hat aber methodische Probleme.

In dieser Einheit werden wir diese drei Ansätze näher kennenlernen.

F5.2 Kosten-Nutzen-Analyse

Finanzwissenschaft

Die Finanzwissenschaft (engl.: public economics oder public finance) ist das Teilgebiet der VWL, das sich mit der Frage befasst, welches Volumen an öffentlichen Gütern bereitgestellt werden soll und wie dieses Volumen finanziert werden soll.

Nun sind Gesundheitsgüter im Regelfall keine öffentlichen Güter, sondern private Güter. Die Versicherungslösung für die Finanzierung der Gesundheitsgüter führt aber dazu, dass die eigentlich privaten Güter eine Reihe von Eigenschaften gewinnen, die für öffentliche Güter typisch sind. Wenn das so ist, könnte es hilfreich sein, im Werkzeugkasten der VWL nachzuschauen, welche Instrumente die Finanzwissenschaft entwickelt hat, um mit dem Problem der Frage nach der richtigen *Menge* der richtigen öffentlichen Güter umzugehen.

Messung des Nutzens

Die Grundidee der Kosten-Nutzen-Analyse ist sehr einfach: Sie vergleicht die Kosten einer Maßnahme mit der Zahlungsbereitschaft der Begünstigten. Diese Zahlungsbereitschaft wird (genauso wie die Kosten) in Euro gemessen. Sie gibt den *monetären* Nutzen der Empfänger des öffentlichen Gutes an.

Strategisches Verhalten

Bei privaten Gütern ist die Messung der Zahlungsbereitschaft relativ einfach. Wenn Sie sich ein iPhone für 800 € kaufen, weiß ich, daß Ihre Zahlungsbereitschaft mindestens 800€ beträgt, weil Sie selbst den Preis zahlen. Wenn man wissen will, wie hoch der Nutzen städtischer Grünflächen ist, kann man die Quadratmetermieten vergleichbarer Wohnungen in der Nähe öffentlicher Parks mit den Preise von Wohnungen ohne nahe Grünflächen vergleichen.[1]

Bei öffentlichen Gütern oder Gesundheitsgütern ist das anders, weil Konsum und Finanzierung entkoppelt sind.[2] Aus diesem Grund haben die Befragten einen Anreiz, den Nutzen, den sie angeben, bewusst zu hoch anzusetzen. Weil man den Betroffenen nicht in das Gehirn schauen kann, ist die Frage, wie man die „wahre“ Zahlungsbereitschaft ermitteln kann, von zentralem Interesse.

F5.3 Kosten-Wirksamkeits-Analysen

Während die Kosten-Nutzen-Analyse ein allgemeiner Ansatz ist, den man prinzipiell u.a. auch auf Gesundheitsgüter anwenden kann, ist die Kosten-Wirksamkeits-Analyse in gewisser Weise das genaue Gegenteil. Ihr Fokus liegt ausschließlich auf der Bewertung medizinischer Güter.

Indikationsspezifische Messung

Die Idee der Kosten-Wirksamkeits-Analyse ist es, das Ergebnis einer medizinischen Leistung in einer relevanten medizinischen Größe zu messen. Der Grundgedanke dieses Ansatzes ist, daß das unmittelbare Ziel medizinischer Leistungen darin besteht, einen medizinisch relevanten Wert günstig zu verändern.

- Der Blutdruck soll gesenkt werden,
- die Beweglichkeit des Hüftgelenks soll erhöht werden
- ...

Für diese Werte gibt es medizinische Kennzahlen in Form von Blutwerten, Gelenkwinkeln usw. Mit diesem Ansatz kann man unter verschiedenen Therapien die effizienteste identifizieren. Im Folgenden ist das für zwei fiktive Blutdrucksenker dargestellt.

[1] Wambach, Achim. Klima muss sich lohnen, Freiburg: Herder, 2022, S. 84.

[2] Bei dem eben angesprochene Park-Beispiel handelt es sich zwar auch um einen öffentlichen Park, aber der nutzt Ihnen wenig, wenn er öffentlich ist, aber 2 km. entfernt liegt.

	A	B
ø Senkung mm/Hg	4	7
Therapiekosten in €	55	145
€ je mm/Hg	13,75	20,71

Medikament A wirkt nicht so gut wie B. B ist allerdings unverhältnismäßig viel teurer als A. Das Kosten-Wirksamkeits-Verhältnis von B ist schlechter. Zu diesem Ergebnis kommt man, indem man die Kosten durch die Wirksamkeit dividiert. Wenn man die Wahl zwischen beiden Therapien hat, würde man A als die effizientere Therapie bevorzugen.

Compliance, Effectiveness und Efficacy

In vielen Fällen wird die Wirksamkeit von Therapien in kontrollierten Studien unter *Idealbedingungen* ermittelt. In der Praxis können die Ergebnisse stark von den Idealergebnissen abweichen. Ein Grund dafür ist, daß in den kontrollierten Studien sichergestellt wird, daß die Patienten die Medikamente auch wirklich nehmen. Da viele Medikamente Nebenwirkungen haben, ist das in der Versorgungspraxis nicht unbedingt der Fall. Eine häufigere Nebenwirkung von Blutdrucksenkern bei Männern ist Impotenz. Wenn die Patienten diese Nebenwirkung erleiden und das Medikament nicht mehr nehmen und ihren Arzt nicht informieren, weil ihnen das peinlich ist, dann kann der Arzt nicht zwischen *„das Medikament wirkt bei diesem Patienten nicht gut“* und *„das Medikament würde gut wirken, wenn es genommen werden würde“* unterscheiden.

Der Begriff *Compliance* bezeichnet die Therapietreue des Patienten, d.h. das Ausmaß, in dem sich der Patient an die Anweisungen des Arztes hält. In diesem Zusammenhang wird zwischen der *Effectiveness*, d.h. der theoretischen Idealwirkung in kontrollierten Studien und der *Efficacy*, d.h. der Wirkung in der medizinischen Praxis unterschieden.[3]

Einschränkung: Indikationsbezogenheit

Die Kosten-Wirksamkeit kann nur zwischen alternativen Therapien der gleichen Indikation verglichen werden, also z.B. nur zwischen verschiedenen Blutdrucksenkern.[4] Die Methode ist nicht geeignet, einen Blutdrucksenker mit einem Cholesterinsenker zu vergleichen, weil die Wirksamkeit eine jeweils andere Kennzahl verwendet. Das bedeutet, daß es möglich ist, eine effiziente Blutdrucktherapie zu identifizieren und daneben eine effiziente Cholesterintherapie, aber nicht, ob eine der beiden Therapien „wichtiger“ oder „effizienter“ ist als die andere. Aus gesundheitsökonomischer Sicht ist das unbefriedigend. Mit dieser Methode kann man also letztlich nur die Therapieform ermitteln, die am wenigsten ineffizient ist, weiß aber nicht, ob die so ermittelte Therapieform insgesamt ein brauchbares Preis-Leistungs-Verhältnis hat.

[3] Für dieses Begriffspaar gibt es keine „amtlichen“ deutschen Übersetzungen. Häufig wird von „klinischer Wirksamkeit“ und „Alltagswirksamkeit“ gesprochen. Diese Alltagswirksamkeit ist aber schwer zu ermitteln, weil man die Daten als Nebenprodukt aus der Versorgungspraxis gewinnen muß, die Versorgungsforschung in Deutschland aber kaum stattfindet.

[4] Das gilt auch für den Ansatz der Value Based Healthcare, der im Wesentlichen eine Variante der Kosten-Wirksamkeits-Analyse ist. Vgl. Porter, Michael E. und Guth, Clemens. Chancen für das deutsche Gesundheitssystem - von Partikularinteressen zu mehr Patientennutzen, Berlin: Springer Gabler, 2012.

F5.4 Kosten-Nutzwert-Analyse

Der dritte Ansatz, die Effizienz medizinischer Leistungen zu messen ist die Kosten-Nutzwert-Analyse (cost-utility-analysis). Sie versucht, die Kritikpunkte, die an den anderen Verfahren gemacht werden können, zu vermeiden. Dieses Verfahren ist das in der Gesundheitsökonomie am weitesten entwickelte und eingesetzte Analyseinstrument zur Effizienzmessung.[5] Aus diesem Grund werden wir uns mit diesem Werkzeug am intensivsten befassen.

QALY (Quality Adjusted Life Year)

Die Grundidee der Kosten-Nutzwert-Analyse ist ziemlich einfach. Eine medizinische Leistung ist nur dann sinnvoll, wenn sie entweder das Leben des Patienten verlängert oder seine Lebensqualität steigert – oder beides tut. Eine bloße Verbesserung medizinischer Indikatoren wie bei der Kosten-Wirksamkeits-Analyse, die keine Verbesserung der Lebensqualität oder Lebensverlängerung bringt, ist Geldverschwendung.

Das Quality Adjusted Life Year (QALY) ist die Kennzahl, die Lebensqualität und Lebensdauer in einen Wert zusammenfasst. Ein QALY bezeichnet ein (fiktives) Lebensjahr bei einer Lebensqualität von 100%.[6]

Beispiel: Erhöhung der Lebensqualität bei gleicher Lebensdauer

In der Tabelle ist ein Beispiel für Patienten mit einer Restlebenszeit von drei Jahren aufgelistet.

Jahr	1	2	3	4	$\sum$
LQ ohne Behandlung	0,8	0,8	0,7	-	2,3
LQ mit Behandlung	1,0	1,0	0,8	-	2,8
Gewinn	0,2	0,2	0,1	-	0,5

Die Lebensqualität (LQ) nimmt während dieser Zeit ab. Diese drei Jahre sind also weniger „wert“ als drei Jahre bei 100% Lebensqualität. Den Wert erhalten wir durch die Aufsummierung der Jahreswerte. Er beträgt 2,3. Die Maßeinheit dieses Werts ist das QALY.

Die Behandlung, um die es geht, verlängert das Leben nicht. Auch Behandelte werden im 4. Jahr tot sein. Die Behandlung hebt aber die Lebensqualität an, so daß insgesamt 0,5 QALY gewonnen werden.

Ein Beispiel für solche Behandlungen sind Stents nach Herzinfarkten. Sie erhöhen die Lebensdauer nicht, steigern aber die Lebensqualität. Deshalb könnte es sinnvoll sein, solche Stents zu setzen.

[5] In *Einheit Das Gesundheitssystem in Großbritannien* werden wir sehen, wie dieses Konzept im britischen Gesundheitssystem eingesetzt wird.

[6] Wie man Lebensqualität misst, ist Thema der nächsten Einheit. An dieser Stelle nehmen wir einfach an, dass wir wissen, wie hoch sie ist.

Beispiel: Lebensverlängerung bei geringerer Lebensqualität

Das zweite Beispiel stellt den Fall dar, daß die Behandlung das Leben verlängert, diese Verlängerung aber mit einem Verlust an Lebensqualität erkauft werden muß.[7]

Jahr	1	2	3	4	$\sum$
LQ ohne Behandlung	0,9	0,85	-	-	1,75
LQ mit Behandlung	0,85	0,8	0,75	-	2,4
Gewinn	-0,05	-0,05	0,75	-	0,65

In Summe bringt die Behandlung immer noch einen Gewinn von 0,65 QALY. In diesem Beispiel haben wir also beide relevanten Effekte (Lebensdauer und Lebensqualität) mit einer Kennzahl (QALY) abgedeckt.

€ je QALY

Die Effizienz einer medizinischen Leistung kann über € je QALY gemessen werden. Wenn die Behandlung 0,65 QALYs bringt und 5.000€ kostet, dann kostet jedes über diese Behandlung gewonnenes QALY $\frac{5.000€}{0{,}65 QALY} = 7.692{,}32$ €/QALY . Diese Kennzahl kann man für jede medizinische Behandlung ermitteln und miteinander vergleichen, also auch Blutdrucksenker und Blutfettsenker, Krebstherapie und Schulzahnarzt.

Priorisierung

Mit der Kennzahl €/QALY kann man die Effizienz beliebiger Therapien miteinander vergleichen, wie im Folgenden für die fiktiven Therapien A bis D.

Therapie	€	QALY	E/QALY
A	100	0,05	2.000
B	10.000	0,50	20.000
C	5.000	0,20	25.000
D	100.000	2,00	50.000

Im Vergleich zu B bringt A sehr wenig, kostet aber auch wenig, so daß die Effizienz von A höher als von B ist. Usw.

Solche Listen kann man auf zweierlei Weise verwenden, um den Leistungsumfang eines Gesundheitssystems unter Effizienzgesichtspunkten zu gestalten.

[7] Diese Tabelle macht noch einmal deutlich, wie die im letzten Abschnitt beschriebenen Complianceprobleme entstehen. Der Patient merkt unmittelbar, daß die Behandlung seine Lebensqualität senkt. Die Lebensqualität ist sogar noch niedriger als ohne Behandlung. Dass dies ein Preis ist, den er für ein verlängertes Leben zahlt, ist ihm vielleicht nicht klar. Zudem hat er das Risiko, dass er zwar jetzt den Preis in Form niedrigerer Lebensqualität zahlt, den Nutzen eines längeren Lebens aber nicht einstreichen kann, weil er vorher vielleicht doch stirbt, denn die Restlebensdauer ist ja nur ein Durchschnittswert. Vor diesem Hintergrund ist es nicht sehr überraschend, wenn ein Patient die Mitarbeit verweigert. Er muss dazu nicht dumm oder willensschwach sein, sondern einfach nur starke Gegenwartspräferenzen haben.

1. Man kann die Liste „von oben“ abarbeiten, bis ein vorher definiertes Budget erschöpft ist.

2. Man kann einen Schwellenwert definieren, bis zu dem Leistungen in den Katalog aufgenommen werden.

Priorisierung ist in diesem Kontext aber nur eine nette Umschreibung für *Rationierung*, denn die Mühe der Listenerstellung macht nur dann Sinn, wenn man Leistungen identifizieren will, deren Preis-Leistungs-Verhältnis so schlecht ist, daß man sie nicht in den Leistungskatalog aufnimmt. Wenn am Ende sowieso alle Leistungen finanziert werden, kann man sich die Arbeit sparen. Priorisierung bedeutet immer „Dinge nicht tun“.

Diskontierung

Eine Aussage der volkswirtschaftlichen Zinstheorie ist, daß Menschen eine Gegenwartspräferenz haben und Konsum nur dann zeitlich aufschieben, wenn man sie über Zinsen kompensiert, so daß einem Konsum heute ein größerer Konsum morgen gegenübersteht. Bezogen auf den Gewinn an Lebensqualität bedeutet das, daß Gewinne an Lebensqualität umso weniger stark ins Gewicht fallen, je weiter sie in der Zukunft liegen. Aus diesem Grund müssen sowohl die Kosten als auch die Gewinne an Lebensqualität diskontiert werden.

Diese Rechnung ähnelt dem Kapitalwert aus der betriebswirtschaftlichen Investitionsrechnung aus dem Bachelor.

Das britische Gesundheitssystem setzt diese Methode ein und verwendet einen kalkulatorischen Zinssatz von 1,5 %.

Fehlende Puzzleteile

Am Anfang dieser Einheit habe ich angekündigt, daß wir uns intensiv mit der Kosten-Nutzwertanalyse befassen wollen. Und jetzt ist das Unterkapitel schon vorbei. Das ist deswegen so, weil Sie jetzt das Konzept kennen, ich mich aber um zwei zentrale Fragen herumgeschummelt habe

1. Wie ermittelt man die Lebensqualität? Wir haben einfach mit Zahlen hantiert, die vom Himmel gefallen sind. Da müssen wir noch genauer hinschauen.

2. Wie berechnet man Lebenserwartung? Wir beobachten, daß manche Menschen relativ früh sterben und manche sehr alt werden. Für die Berechnung der QALYs ist die Frage zentral, wie lange die Menschen etwas von der Behandlung haben. Wie kommt man da zu Prognosewerten? Das ist kniffeliger als „einfach den Durchschnitt berechnen“.

Diese beiden fehlenden Puzzleteile sind zu umfangreich, als daß man sie als Unterabschnitt in dieser Einheit besprechen könnte. Daher gibt es jeweils eine eigene Einheit zu diesen Bausteinen.

F5.5 Ehemalige Klausuraufgaben

Kosten je QALY

1. Bitte ermitteln Sie im Rahmen einer Kosten-Nutzwert-Analyse den Zugewinn an qualitätsadjustierten Lebensjahren der untenstehenden Behandlung bei einem Diskontierungssatz von 10%.

2. Wie hoch sind die Kosten, die ein gewonnenes qualitätsadjustiertes Lebensjahr unter diesen Bedingungen verursacht?

Kosten und Nutzen fallen jeweils am Ende des Jahres an.

Jahr	1	2	3	4
LQ ohne Behandlung	0,8	0,8	0,7	-
LQ mit Behandlung	0,7	0,9	0,8	0,7
Kosten €	5.000	2.000	2.000	2.000

Diskontsatz

Sie haben gelernt, daß die Berechnung der Kosten eines QALYs den konventionellen Investitionsrechenverfahren ähnelt. Hier wie dort gibt es einen Diskontsatz oder kalkulatorischen Zinssatz. In der Standard-BWL richtet sich dieser Satz nach der Rendite alternativer Anlagemöglichkeiten. Wenn man meint, mit anderen Investitionen mindestens eine Rendite von 10% zu erzielen, dann gibt einem der Kapitalwert mit einem kalkulatorischen Zinssatz von 10% an, wieviel mehr Gewinn (in €) man über diese Standardverzinsung hinaus mit einem Investitionsprojekt erzielen kann.

Bei den QALYs gibt es solche alternativen Mindestrenditen nicht, auf die man zur Festlegung des kalkulatorischen Zinses zurückgreifen kann. Es ist daher wenig überraschend, das verschiedene Organisationen, die Kosten-Nutzwert-Analysen durchführen, verschiedene Diskontsätze verwenden.

Der National Health Service (NHS) in Großbritannien verwendet einen Diskontsatz von 3,5 % p.a. Die Weltbank 3 % und das US-amerikanische Center for Disease Control 5 %. Es ist klar, daß beim Vergleich zweier Leistungen für beide der gleiche Zinssatz verwendet werden muß. Trotzdem hat die Wahl des Zinssatzes eine Auswirkung auf die Frage, welcher Typ von Therapie tendenziell bei höheren Diskontsätzen besser abschneidet als bei niedrigen.

Beschreiben Sie diesen Therapietyp.

Tipp: Vielleicht hilft es, wenn Sie sich vorstellen, daß die Zinssätze nicht so nahe beieinanderliegen, wie in der Praxis, sondern ganz stark verschieden sind, also 1% bei Organisation X und 1000 % bei Organisation Y.

€ je QALY einer Herzkatheteruntersuchung [8]

Ein Patient wird mit Akutsymptomen ins Krankenhaus eingeliefert. Im Herzkatheterlabor wird eine Verengung der Herzkranzgefäße an kritischer Stelle diagnostiziert. Mit einem Ballonkatheter und dem Setzen eines Stents wird die Verengung behoben. Ohne Behandlung wäre der Patient verstorben.

Der Patient ist anschließend beschwerdefrei. Die Herzleistung ist nicht eingeschränkt. In Fällen dieser Art wird zur Vermeidung eines Reinfarkts der Patient standardmäßig dauerhaft mit folgenden Medikamenten behandelt:

1. Simvastatin zur Stabilisierung der Herzkranzgefäße.
2. Aspirin zur Hemmung der Blutgerinnung.
3. Metoprolol zur Senkung des Pulses.

Zusätzlich muß der Patient einmal jährlich zur Ultraschall-Nachuntersuchung.

Da es sich um einen leichten Herzinfarkt ohne Folgeschäden handelt, kommt es zu keiner Reduktion der Restlebenserwartung von 40 Jahren. Da die Medikation Nebenwirkungen hat und eine psychische Belastung darstellt, sinkt die Lebensqualität aber auf 95%. Der Krankenhausaufenthalt anschließender einmaliger Nachuntersuchung im Herzkatheterlabor kostet 10.000€. Die Rehabilitation 5.000€. Die jährlichen Ultraschall-Nachuntersuchungen kosten jeweils 1.000€. Die Medikamente sind als Generika verfügbar und kosten 300€/Jahr.

Aus den Angaben kann man die €/QALY berechnen. Es sei angenommen, daß die Kosten und Nutzen zu Periodenende anfallen. Der Diskontsatz ist 3%.

Mit der Hand wird die Berechnung bei 40 Jahren mühsam. Entwerfen Sie eine Excel-Tabelle zur Lösung dieser Aufgabe.

Ein Durchbruch in der Brustkrebstherapie

Supertinib ist ein (fiktiver) neuer Wirkstoff in der Behandlung von Brustkrebs in fortgeschrittenem Stadium. Es geht also nicht mehr um eine Heilung, sondern darum, den Tod bei guter Lebensqualität hinauszuschieben. Supertinib greift gezielt Krebszellen an und zerstört sie. Leider wirkt das Medikament nur eine gewisse Zeit. Dann werden die Krebszellen Resistenzen entwickeln und die Krankheit wird weiter fortschreiten. Patientinnen, die sehr gut auf das Medikament ansprechen, haben noch eine Perspektive von maximal 3,5 Jahren. Supertinib hat Nebenwirkungen, so daß die Lebensqualität während der Einnahme des Medikaments bei 90% liegt. Sobald klar ist, daß Supertinib nicht (mehr) wirkt, wird es abgesetzt. Der Gesundheitszustand, in dem sich die Patientinnen dann befinden, ist nicht mehr so gut, so daß die Lebensqualität noch bei 80% liegt. Ohne Therapie, bzw. nach Absetzen des Medikaments versterben die Patientinnen nach durchschnittlich 6 Monaten.

Ergebnisse einer Studie: Bei 20% der Patientinnen schlägt das Medikament überhaupt nicht an. Das ist nach wenigen Tagen klar. Bei 30% Patientinnen, die auf das Medikament

[8] Diese Aufgabe ist keine ehemalige Prüfungsaufgabe, weil sie faktisch nur mit Excel lösbar ist, das Ihnen in der Klausur nicht zur Verfügung steht.

ansprechen, wirkt das Medikament ein Jahr und dann nicht mehr. Bei 50% der Patientinnen, die zwei Jahre erfolgreich behandelt worden sind, wirkt das Medikament nach 2 Jahren nicht mehr. Nach 48 Monaten wirkt das Medikament bei keiner Patientin mehr.

Wie viele zusätzliche QALYs bringt die Therapie im Durchschnitt bei einem Diskontsatz von 3% und nachschüssiger Rechnung? Unterjährige Verzinsung können Sie ignorieren. Geben Sie den Wert in Monaten mit einer Nachkommastelle an.

Kosten je QALY (Lösung)

Der Lebensqualitätsgewinn am Ende des ersten Jahres beträgt 0,7-0,8=-0,1, d.h. die Behandlung mindert die Lebensqualität. Dieser Wert muß auf die Gegenwart abgezinst werden, weil zwischen heute und den Kosten und Nutzen 12 Monate liegen. Es ist eine nachschüssige Rechnung.

Erstes Jahr: $\frac{0,7-0,8}{1,1} = -0,0909$
Zweites Jahr $\frac{0,9-0,8}{1,1^2} = 0,0826$
Drittes Jahr $\frac{0,8-0,7}{1,1^3} = 0,0751$
Viertes Jahr $\frac{0,7-0,0}{1,1^4} = 0,4781$
In Summe beträgt das 0,5449 QALYs.

In gleicher Weise müssen die Kosten der Behandlung der vier Jahre auf den heutigen Tag nachschüssig diskontiert werden. Die Werte sind 4545,46€, 1652,89€, 1502,63€ und 1366,03€. In Summe sind es 9067,01€ Die Kosten je QALY betragen somit $\frac{9067,01€}{0,5449\,QALY} = 16.639,77€/\text{QALY}$

Diskontsatz (Lösung)

- Zuerst einmal kann man aussortieren, welche Dinge nicht relevant sind
  - Die Antwort kann nichts mit der Schwere der Krankheit an sich zu tun haben, weil geringe Gewinne jeweils mit dem gleichen Satz diskontiert werden wie hohe. Egal, wie hoch der Diskontsatz ist.
  - Die Antwort kann nichts mit den Kosten an sich zu tun haben, weil auch die mit dem jeweils gleichen Satz diskontiert werden. •
- Wozu führt ein sehr hoher Diskontsatz?
  - 1.000% kalkulatorischer Zins würde dazu führen, daß Lebensqualitätsgewinne im Folgejahr und noch später mit einem sehr geringen Gewicht in die Rechnung eingehen, also wenig „wert“ sind.
  - Ein hoher Diskontsatz begünstigt also medizinische Leistungen, die sehr kurzfristig Nutzen bringen. Die mittel- und langfristigen Wirkungen der Therapien sind dann weitgehend bedeutungslos.

Herzkatheter (Lösung)

Das QALY kostet 1987,40€. Die Excel-Tabelle am Seitenende ist ein Beispiel dafür, wie man dieses Ergebnis berechnen kann.

Ein Durchbruch in der Brustkrebstherapie (Lösung)

- Man fängt mit einer fiktiven großen Zahl von Patientinnen an. Z.B. 1.000 und schätzt ab, wann wieviele Patientinnen sterben.
- In den ersten 12 Monaten passiert folgendes: 200 Frauen sterben nach 6 Monaten mit einer Lebensqualität von 0,8, 800 Frauen leben 12 Monate mit einer Lebensqualität von 0,9. Insgesamt sind das 960+8.640=9.600 QALYs. Mit 3% auf heute diskontiert: 9.600/1,03=9320,38 Monate

	A	B	C	D	E	F	G
1		Kosten		QALY		disk. Kosten	disk. QALY
2	Jahr	10000	Untersuchung				
3		5000	Reha			Zins	3,0%
4		300	Medikamente				
5	1	15300	Summe	0,95		14.854,37 €	0,92
6	2	1300	Medikamente + Ultraschall	0,95		1.225,37 €	0,90
7	3	1300		0,95		1.189,68 €	0,87
8	4	1300		0,95		1.155,03 €	0,84
9	5	1300		0,95		1.121,39 €	0,82
44	40	1300		0,95		398,52 €	0,29
45							
46			Summe			43.641,44 €	21,96
47							
48			€ je QALY	1.987,40 €			

Abbildung 5.1: Herzkatheter (Lösung)

- In den Monaten 12-24 passiert folgendes: Von den 800 Frauen nach 12 Monaten sterben 30% (240) nach 6 Monaten mit 80% Lebensqualität und 560 leben das ganze Jahr mit 90%. Insgesamt sind das 1.152+6048=7.200 QALYs. Diskontiert auf heute $7.200/1{,}03^2$=6.788,69 Monate
- In den Monaten 24-36 passiert folgendes: Von den 560 Frauen, die zwei Jahre erfolgreich therapiert worden sind, sterben 50% (280) nach 6 Monaten bei einer Lebensqualität von 0,8 und 280 leben 12 Monate bei 0,9. Insgesamt sind das 1.344+3.024=4.368 Monate. Diskontiert auf heute $4368/1{,}03^3$=3.997,39 Monate
- In den Monaten 36-48 passiert folgendes: Alle 280 Frauen, die drei Jahre erfolgreich therapiert worden sind, sterben nach durchschnittlich 6 Monaten. Insgesamt sind das 1.344 Monate. Diskontiert auf heute 1.194,13 Monate.
- Die Summe aller diskontierten Monate ist 9320,38+6.788,69+3.997,39+1.194,13= 21.300,59 Monate. Pro Frau sind das 21,3 Monate. Diese 21,3 Monate sind die durchschnittlich gewonnenen QALYs einer Therapie.
- Die Frage war aber, was die Therapie bringt. Wenn man nichts tut, sind die Patientinnen ja nicht sofort tot, sondern „erst“ nach 6 Monaten bei 80% Lebensqualität, die man mit 3% diskontieren muß. Bei 1000 Frauen wären das 4660,2 Monate, bzw. gerundet durchschnittlich 4,7 Monate.
- Diesen Betrag muß man von 21,3 abziehen und erhält 16,6 Monate.

Kapitel 6

Messung von Lebensqualität

F6.1 Glücksforschung

Der Begriff

Bereits im Bachelor haben Sie die Grenzen des Ansatzes kennengelernt, den Wohlstand eines Landes über das BIP pro Kopf zu messen. Was man auf diese Weise misst, ist der *materielle* Wohlstand. Was man eigentlich gern messen würde ist die *Lebensqualität*, aber die ist schwerer zu messen als das BIP. In den letzten Jahren haben sich daher einige Ansätze entwickelt, mit denen man versucht, Lebensqualität zu messen.

In der Ökonomie hat sich der (etwas kitschig klingende) Begriff der *Glücksforschung* etabliert.[1]

Das Königreich Bhutan weist das Bruttonationalglück aus und es gibt Kennzahlen wie den Happy Planet Index, der auch ökologische Aspekte berücksichtigt. Das bedeutet, daß die Frage nach der Messung von Lebensqualität keine spezifisch gesundheitsökonomische Frage ist und wir mit unseren QALYs auch auf Instrumente zurückgreifen können, die nicht unmittelbar aus dem Gesundheitsbereich stammen.

Befragung

Die methodisch einfachste Möglichkeit, die Lebensqualität zu messen, besteht darin, die Menschen direkt zu fragen. Das *sozioökonomische Panel* (SOEP) ist eine Studie, die seit den 1980er Jahren durchgeführt wird und in der die gleiche Personengruppe jährlich befragt wird.[2] Die Schlussfrage ist in jedem Jahr die folgende: *Wie zufrieden sind Sie, gegenwärtig, alles in allem, mit Ihrem Leben?*

Man kann nun lange über das „wording“ streiten und über Unterschiede zwischen „Glück“, „Zufriedenheit“ und „Lebensqualität“, aber mit solchen Fragen kommt man in Bezug auf unsere Lebensqualitätsmessungen schon relativ weit.

[1] Siehe Binswanger, Mathias. Die Tretmühlen des Glücks - wir haben immer mehr und werden nicht glücklicher ; was können wir tun?, Freiburg im Breisgau [u.a.]: Herder, 2013. Für eine kritische Einordnung siehe Weimann, Joachim et al. Geld macht doch glücklich - wo die ökonomische Glücksforschung irrt, Stuttgart: Schäffer-Poeschel, 2012.

[2] Auf die Frage, wie mit der „Panelsterblichkeit“, d.h. dem Wegzug oder dem tatsächlichen Versterben von Teilnehmern umgegangen werden soll, gehen wir hier nicht ein.

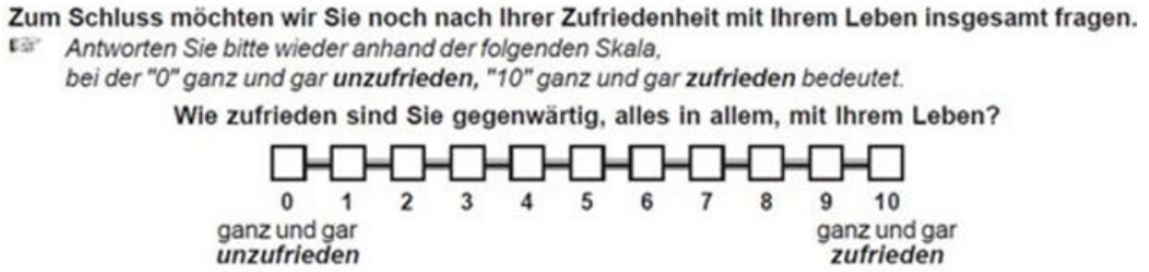


Abbildung 6.1: Sozioökonomisches Panel - Abschlussfrage

Standard-Matrizen

Ein anderer Ansatz ist es, Kriterien für eine Standard-Matrix zu entwickeln, die Patienten nach diesen Kriterien zu befragen und den einzelnen Zellen der Matrix Lebensqualitätswerte zuzuordnen. Ein Beispiel für diesen Ansatz ist die (nach ihrem Verfasser benannte) Rosser-Matrix.[3]

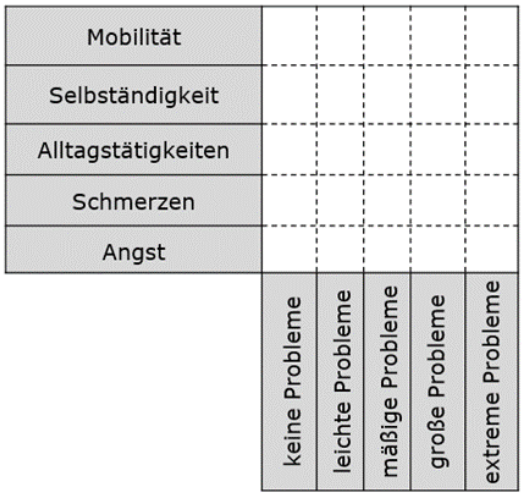
		Schmerz			
		A	B	C	D
Behinderung	I	1,000	0,995	0,990	0,967
	II	0,990	0,986	0,973	0,932
	III	0,980	0,972	0,956	0,912
	IV	0,964	0,956	0,942	0,870
	V	0,946	0,935	0,900	0,700
	VI	0,875	0,845	0,680	0,000
	VII	0,677	0,564	0,000	-1,486
	VIII	-1,028	-	-	-

Diese Matrix ist auf Gesundheitszustände zugeschnitten und verwendet die Kriterien Schmerz und Behinderung. Eine stärkere soziale Beeinträchtigung (die Definition von Behinderung Stufe III) verbunden mit mittleren Schmerzen (Klasse C) stellt eine Lebensqualität von 95,6 % dar. Hierbei ist es irrelevant, welche Krankheit die stärkere soziale Beeinträchtigung und die mittleren Schmerzen hervorruft.

Einen differenzierteren Ansatz stellt der Fragebogen *EQ-5D* dar, der fünf Kriterien und fünf Stärkegrade verwendet, so dass 5^5=3125 verschiedene Kombinationen möglich sind.[4]

[3]Schöffski, Oliver und von der Schulenburg, Matthias. Gesundheitsökonomische Evaluation, Berlin, Heidelberg, New York: Springer, 2008. S. 116

[4]Dieser Fragebogen existiert in mehreren Varianten, die aber alle der gleichen Idee folgen. Siehe https://euroqol.org/



Jeder dieser Kombinationen wird ein Lebensqualitätswert zugewiesen. Noch ausführlicher ist der SF (Short Form) 36, der 36 Fragen beinhaltet.[5]

Lotterieverfahren

Ein anderer Zugang besteht darin, Personen zu fragen, welche Risiken sie bereit sind, auf sich zu nehmen, um ihren derzeitigen Gesundheitszustand zu verbessern. Das *standard gamble Verfahren* arbeitet mit der Frage, welches Sterberisiko eine Person zu tragen bereit wäre, um vollständig zu gesunden. Der Befragte spielt also eine Art Lotterie zwischen „tot“ und „vollkommen gesund“.

- Wenn die akzeptable Sterbewahrscheinlichkeit sehr niedrig ist (sagen wir 0,1%) dann ist der Gesundheitszustand, um den es geht, aus der Sicht des Befragten nicht sehr einschränkend und die Lebensqualität entsprechend hoch.

- Ist die akzeptable Sterbewahrscheinlichkeit sehr hoch (sagen wir 80%) dann ist das ein Hinweis darauf, daß der Befragte seine Lebensqualität als sehr gering einschätzt, weil er bereit ist, ein so hohes Risiko einzugehen, um wieder gesund zu werden.

Tatsächlich kann man die akzeptierte Sterbewahrscheinlichkeit als prozentualen Lebensqualitätsverlust interpretieren. Ein Befragter, der bereit ist, eine Sterbewahrscheinlichkeit von 5% auf sich zu nehmen, um wieder gesund zu werden, hat demnach eine Lebensqualität von 95%.

Eine Variante des standard gamble ist das *time tradeoff Verfahren*. Dort geht es aber nicht um eine Sterbewahrscheinlichkeit, sondern um die Bereitschaft, auf Lebenszeit zu verzichten. Ein gesundheitlich beeinträchtigter Befragter, der bereit ist von 8 verbleibenden Lebensjahren eines herzugeben, wenn er dafür keine gesundheitlichen Einschränkungen mehr hätte, hätte dementsprechend einen Lebensqualitätsverlust von 1/8=12,5%, bzw. eine Lebensqualität von 87,5%.

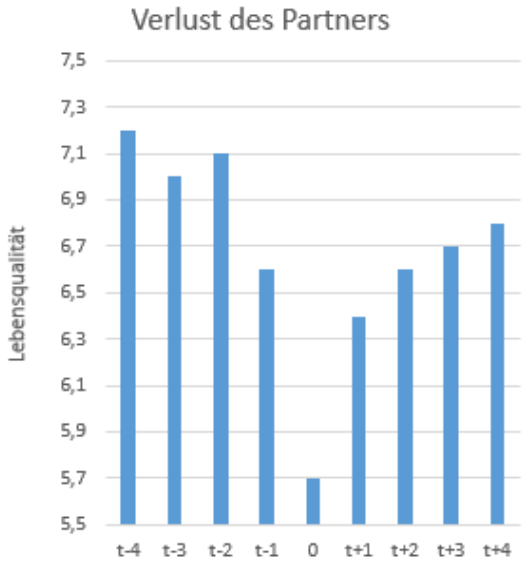
[5]Der SF 36 zeigt aber auch die Grenzen solcher Instrumente auf. Ein Beispiel ist die Messung der Mobilität über die Frage „I can walk several blocks“. In der deutschen Variante ist das mit „Straßenkreuzung“ übersetzt, was aber eine deutlich andere Strecke ist als der us-amerikanische „block“. Das bedeutet, daß es problematisch sein kann, Daten aus verschiedenen Ländern zu vergleichen.

F6.2 Methodische Probleme

Happiness Set Point

Ein Problem der Messung von Lebensqualität besteht in der Frage, wen man fragt und *wann* man ihn fragt. So könnte es sein, daß die Aussagen über „Lebensqualität von Querschnittsgelähmten“ sich sehr unterscheiden, je nachdem, ob man Querschnittsgelähmte fragt oder Leute nicht querschnittsgelähmt sind.

Das bereits erwähnte SOEP bietet als Längsschnittstudie die Möglichkeit, die Veränderung der Lebenszufriedenheit einzelner Personen zu beobachten. Im SOEP werden noch viele weitere Daten abgefragt, so daß es möglich ist, den Einfluss einschneidender Lebensereignisse auf die Lebenszufriedenheit zu untersuchen. Mit den Daten war das medizinische natürlich nicht sehr differenziert möglich, sondern nur für Ereignisse wie „Heirat“, „Geburt des ersten Kindes“, „Tod des Partners“ und „Arbeitslosigkeit“. Bis auf „Arbeitslosigkeit“ wiesen alle Ereignisse die folgende Struktur auf: Die Lebensqualität verändert sich ein bis zwei Jahre vor dem Ereignis und bewegt sich ein bis zwei Jahre nach dem Ereignis in etwa wieder auf das alte Niveau.[6]Das bedeutet, daß die Betroffenen (im Verlustfall) ihren Verlust bewältigen[7] bzw. daß die Freude über ein positives Ereignis nicht dauerhaft ist.[8]



Für die gesundheitsrelevante Messung von Lebensqualität bedeutet das, daß es einen Unterschied machen könnte, ob die Gesundheitseinschränkung erst kurz oder schon länger besteht und inwiefern man die noch kommenden Bewältigungsstrategien bei „frisch“ Erkrankten in die aktuelle Einschätzung der eigenen Lebensqualität einbeziehen sollte.[9]

[6]Clark, Andrew E. et al. Lags and Leads in Life Satisfaction - A Test of the Baseline Hypothesis, Berlin: DIW-Diskussionspapier, 2003.

[7]Im Englischen „coping“ genannt, im Deutschen „Resilienz“

[8]In Bezug auf materielle Gewinne wird hier auch der Begriff der „hedonistischen Tretmühle“ verwendet.

[9]So ermitteln Schulz, Richard und Decker, Susan, "Long-term adjustment to physical disability: The role of social support, perceived control, and self blame", Journal of Personality and Social Psychology, 48 (5) 1985. S.1162-1172. für Personen, die seit 20 Jahren querschnittsgelähmt sind, nur geringfügig niedrigere Lebensqualitätswerte als bei Nicht-Querschnittsgelähmten gleichen Alters.

Identische Nutzenfunktionen

Die existierenden Methoden zur Messung von Lebensqualität in Gesundheitskontexten arbeiten mit aggregierten Daten und Durchschnittswerten. Im Abschnitt über den Happiness Set Point haben wir gesehen, wie Menschen *im Durchschnitt* den Verlust ihres Partners verarbeiten. Mit solchen Werten können wir umgehen. Aber im Einzelfall könnte es sein, dass der Sarg noch nicht unter der Erde ist und schon ein Tinder-Konto eröffnet wird. Oder – am anderen Ende des Spektrums - dass die Person nie über diesen Verlust hinwegkommt. Der Durchschnitt gibt die Streuung nicht wieder.

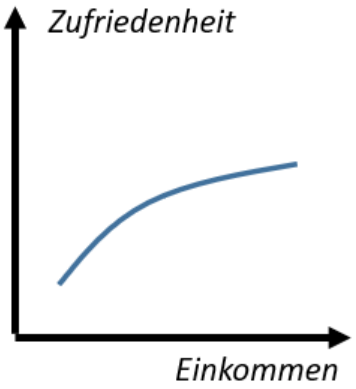
F6.3 Was bleibt - trotz der Einschränkungen?

Es gibt noch mehr Detailprobleme, aber die zentralen methodischen Probleme der Messung von Lebensqualität kennen Sie jetzt. Es gibt Kritiker, die diese Probleme für so gravierend halten, dass sie die ganze QALY-Methode ablehnen. Ich kann die Einwände gut verstehen, denn die Probleme sind ja wirklich da und sie sind nicht nebensächlich.

Easterlin - Paradox

Wenn wir die Lebensqualität nicht direkt messen können, könnten wir versuchen, eine Ersatzgröße, die mit Lebensqualität korreliert - und besser messbar - ist, zu verwenden.

Im Bachelor haben Sie das Easterlin-Paradox kennengelernt. Es beschreibt die Beobachtung, dass die Zufriedenheit (Alternativbegriff: Lebensqualität) unterproportional mit dem Einkommen steigt und ab einer überschaubaren Höhe fast gar nicht mehr zunimmt.



Wir haben diese Beobachtung als Ausprägung des *Gesetzes des abnehmenden Ertragszuwachses*, das wir im Bachelor kennengelernt haben, identifiziert.

In gleicher Manier können wir unterstellen, dass der Zusammenhang zwischen *Gesundheit* und Lebensqualität ähnlichen Spielregeln folgt: Der Unterschied zwischen guter und *sehr* guter Gesundheit könnte unter dem Aspekt der Lebensqualität marginal sein. Beim ökonomischen Wachstum haben wir gesehen, dass die Regel

> *Maximiere das BIP pro Kopf und du maximierst damit die Lebensqualität*

nicht funktioniert und hier die Regel

> *Maximiere die individuelle Gesundheit und du maximierst damit die Lebensqualität*

auch nicht. Wir tappen dann in die gleiche Falle, wenn wir versuchen, eine Ersatzgröße zu maximieren, weil die Größe, die wir eigentlich maximieren wollen, zu schwer messbar ist. Bei einem linearen Zusammenhang scheint mir das eine gute Idee zu sein. Bei einem unterproportionalen Zusammenhang nicht.

Eigenschaften einer befriedigenderen Lösung

Ich fürchte auch, dass das Standardargument *„noch 10 Jahre Forschung und wir werden diese Probleme gelöst haben“* Augenwischerei ist. Zu dieser Position bin ich gekommen, indem ich versucht habe, die Eigenschaften einer Methode zu benennen, die nicht mehr der genannten Kritik unterliegen würde.

Wir könnten in der Zukunft noch ein paar mehr Befragungen über Lebensqualität im Rollstuhl durchführen, indem wir die Gruppen präziser zuschneiden und an den Fragen etwas herumfeilen und die Befragungszeitpunkte variieren, aber das Grundproblem bleibt bestehen. Es ist eben eine Befragung und ob *„meine Lebensqualität liegt bei 80%“* für Alice das gleiche bedeutet wie für Bob, wird immer unklar bleiben.

Was man haben wollen würde, wäre ein *Hedometer*. Jeremy Bentham, einer der Begründer des Utilitarismus stellt sich, Anfang des 19. Jhr. vor, dass findige Ingenieure ein Messinstrument entwickeln könnten, das Glück wie Temperatur, Gewicht oder Luftdruck messen könnte. Objektiv.

Mein Eindruck ist, dass man sich von den *Neurowissenschaften* etwas in der Art verspricht. Mit bildgebenden Verfahren kann man die Aktivität des Gehirns und die Reaktion auf Reize beobachten. Wenn ich Alice und Bob, die beide im Rollstuhl sitzen, das Bild von einem laufenden Menschen zeige und derweil ihre Hirnaktivitäten beobachte und feststelle, dass elektrochemisch die gleichen Dinge in beiden Hirnen passieren, weiß ich immer noch nicht, ob Alice und Bob das gleiche fühlen. Das wissen nicht mal Alice und Bob übereinander.

Gottfried Willhelm Leibnitz, Philosoph und Namensvater des Butterkekses hat das *„Monaden haben keine Fenster*“ genannt, wobei „Monaden“ hier (vereinfacht gesagt) eine Metapher für Menschen sind. Wir können nicht in den Kopf des andern schauen und fühlen, was er fühlt.

So lange nicht geklärt ist, wie man nachweisen will, dass das Gefühl von Alice exakt dem von Bob entspricht, scheint mir die Idee einer exakten Lebensqualitätsmessung unmöglich zu sein.

Was wäre die Alternative?

Die Frage, die ich mir an diesem Punkt aber stelle ist, worin die Alternative besteht. Das ist eine sehr grundsätzliche Frage. Ich neige zu der pragmatischen Antwort, dass ich lieber mit einem schlechten Instrument arbeite, als gar keins zu haben.

Warum müssen wir denn überhaupt Lebensqualität quantifizieren? Ist das nicht nur eine Marotte von Ökonomen, die alle Dinge mit Zahlen messen müssen? Ich glaube tatsächlich, dass Ökonomen zu dieser Marotte neigen, aber im gesundheitsökonomischen Kontext sehe ich keine Alternative.

Die Überlegung, dass Lebensqualität etwas rein Subjektives ist und jeder für sich entscheiden soll, was ihm wichtig ist, funktioniert so lange, wie nur die Person selbst betroffen ist. Wir haben aber gesehen, dass ein Gesundheitssystem nur dann funktioniert, wenn wir

Inanspruchnahme und Finanzierung entkoppeln.Das bedeutet dann aber, dass Alice erwartet, dass Bob, Carol, Dave und Eve ihre Behandlung bezahlt, weil Alice finanziell gar nicht in der Lage dazu ist. Dazu werden die Nettozahler nur bereit sein, wenn irgendwie geklärt werden kann, dass der Nutzen dieser Behandlung in einem akzeptablen Verhältnis zu den Kosten steht, die sie für Alice tragen. Dazu muss der Gewinn an Lebensqualität, den Alice durch die Behandlung hat, in irgendeiner Form personenübergreifend nachvollziehbar sein, weil es eben nicht nur um Alice geht, sondern auch um das Geld von Bob, Carol, Dave und Eve.

Die QALYs sind nicht besonders toll. Darüber muss man nicht lange streiten. Aber die Alternative zu den QALYs ist – *nichts*. Bessergesagt: Blanke Willkür.

Kapitel 7

Lebenserwartung

F7.1 Der rote Faden

In *Einheit Ergebnisdefinitionen*haben wir uns mit der Kosten-Nutzwertanalyse befasst und das Instrument der QALYs kennengelernt. Unser Problem war, dass wir dafür Lebensqualität und Lebensdauer prognostizieren mussten. In der letzten Einheit haben wir uns mit der Frage nach der Messung von Lebens*qualität* befasst. Jetzt werden wir uns mit der Prognose der Lebens*dauer* befassen.

Dieses Unterfangen ist kniffeliger, als es zunächst scheinen mag. Zum ersten können wir keine *individuellen* Prognosewerte liefern, sondern nur *Durchschnittswerte* für eine Personengruppe. Wenn die Lebenserwartung, sagen wir, bei 80 Jahren liegt, bedeutet das nicht, dass jeder irgendwie ein „Recht“ darauf hat, 80 zu werden, sondern dass Viele früher sterben werden und Viele später. Zum anderen müssen wir irgendwie mit dem Problem umgehen, dass wir von Sterbedaten der Vergangenheit nur sehr schwer auf die Zukunft schließen können, weil es einen medizinisch-technischen Fortschritt gegeben hat und weiterhin gibt. Es deutet sich schon an: Ein perfektes Prognoseintstrument werden wir nicht finden, sondern nur eines, das besser ist als andere, die noch schlechter sind.

F7.2 Rückblick: Berechnung eines PKV-Tarifs

Worum es im Bachelor ging

Im Bachelor hat die Kalkulation eines PKV-Tarifs einen recht großen Platz eingenommen. Darauf möchte ich jetzt aufbauen. . Damals war unser Problem, dass wir wissen mussten, wann wie viele Versicherte sterben, d.h. keine Prämien mehr zahlen, aber auch keine Rechnungen mehr stellen.

Wenn wir das nicht wissen, können wir keine Prämie berechnen, bzw. wir haben dann das Problem, dass die Versicherungen einen großen Anreiz haben, unseriös zu kalkulieren.

In dieser Einheit geht es um die Berechnung von künftiger Lebenszeit in Jahren. Das ist sehr ähnlich, aber doch etwas Anderes. Die Zahlen, mit denen wir in dieser früheren Einheit hantiert haben, werde ich in dieser Einheit immer wieder aufgreifen. Daher ist es vermutlich schlau, in der Vorbereitung die PKV-Einheit parallel zu dieser aufgeschlagen zu haben.

Überlebensraten

Eine leicht zu verstehende Kennzahl ist die Überlebensrate. Sie gibt an, wie groß der Anteil der Personen ist, die nach einer definierten Zeitspanne t noch lebt.

In der Onkologie ist z.B. die *Fünf-Jahre-Überlebensrate* eine häufig verwendete Größe, die besagt, wie viele Patienten fünf Jahre nach Erstdiagnose mit und ohne Behandlung noch leben.

	A	B	E	F
6			Personen	
7	Jahr	Alter	Anzahl	Sterbe-W.
8	2019	30	10.000,0	0,05%
9	2020	31	9.994,6	0,06%
10	2021	32	9.988,6	0,06%
11	2022	33	9.982,3	0,07%
12	2023	34	9.975,0	0,07%
13	2024	35	9.967,8	0,08%
14	2025	36	9.960,3	0,09%
77	2088	99	110,7	38,16%
78	2089	100	68,5	

Aus Zeile 13 des Excel-Screenshots kann ich ablesen, dass von 10.000 Männern, die 2019 30 Jahre alt waren, nach 5 Jahren, also 2024 statistisch noch 9.967,8 Männer leben werden. Das bedeutet, das die 5-Jahres-Überlebensrate eines 30jährigen Mannes bei 99,68% liegt. Solche Raten sind aber nur für Menschen sinnvoll, deren Lebenszeit absehbar bald endet.

F7.3 Erwartungswerte

Nicht alle 80jährigen fallen an ihrem Geburtstag tot um

Aus den Medien wissen Sie, daß die Lebenserwartung in Deutschland bei etwa 80 Jahren liegt. Nun kennen Sie aber sicherlich Menschen, die älter sind und sie wissen von Menschen, die weit vor ihrem 80 Geburtstag gestorben sind. Also kann die Lebenserwartung nicht so etwas wie eine „garantierte Mindestlebensdauer“ sein oder irgendeine Zahl, die sich auf eine konkrete Person bezieht, sondern auf eine Gruppe von Personen.

Beispiel: Masernimpfung

Lassen Sie uns diese Überlegung an einer Überschlagsrechnung am Beispiel der öffentlich breit diskutierten *Masernimpfung* vertiefen. Die WHO geht davon aus, daß von 1.000 in Deutschland an Masern erkrankten Kindern *eines* stirbt.[1] Ohne Impfung würde (da Masern extrem ansteckend sind) *jedes* Kind irgendwann Masern bekommen. Bei derzeit etwa 750.000 Geburten pro Jahr in Deutschland wären das 750 tote Kinder je Jahr.

Nehmen wir an, ein maserntotes Kind hätte (wäre es geimpft gewesen) noch 70 Jahre weitergelebt, dann rettet die Impfung 750*70=52.500 Lebensjahre pro Jahr. Da man aber

[1] Schwere Behinderungen als noch häufigere Folge ignorieren wir hier.

	A	B	C	D	E	F	G
1	**Lebenserwartung 2019**						
2			**Personen**			**gelebte Jahre**	
3	**Jahr**	**Alter**	**Anzahl**	**Sterbe-W.**	**Tote**	**d. J.**	**insg.**
4	2019	30	10.000,0	0,05%	5,4	164,7	164,7
5	2020	31	9.994,6	0,06%	6,0	190,3	355,0
6	2021	32	9.988,6	0,06%	6,2	203,1	558,1
72	2087	98	172,6	35,84%	61,8	6.091,9	782.199,3
73	2088	99	110,7	38,16%	42,3	4.204,3	786.403,6
74	2089	100	68,5	100,00%	68,5	6.880,8	793.284,4
75					Durchschnitt		79,3
76					Restlebenserwartung		49,3
77	**Quellen**						
78	Statistisches Bundesamt, Sterbetafeln 2017/19 (männlich)						

Tabelle 7.1: Sterbetafel

nicht weiß, *welches* von jeweils 1.000 erkrankten Kindern stirbt, kann man nur die 52.500 geretteten Jahre auf die 750.000 Kinder verteilen, was einen Durchschnitt von 0,07 Jahren oder knapp vier Wochen je Kind bedeutet. Das ist kein Wert, den *eine konkrete Person* erreichen kann. Entweder man stirbt (ziemlich unwahrscheinlich) oder man überlebt.

In meiner Rechnung habe ich etwas geschummelt. Ich habe behauptet, daß ein Maserntoter noch 70 Jahre leben würde. Wahrscheinlich ist das sogar ein ganz guter Wert, aber woher habe ich den? Und warum ist Ihnen das vermutlich nicht aufgefallen?

Wir haben uns an der Lebenserwartung in Deutschland, die bei etwa 80 Jahren liegt, orientiert. Das ist aber geschummelt, weil es in diesem Kapitel ja genau darum gehen soll, wie diese Zahl berechnet wird. Ich berechne die Lebenserwartung also mit einer Rechnung, die voraussetzt, daß ich die Lebenserwartung schon kenne. So geht das nicht.

F7.4 Sterbetafeln

Im Bachelor brauchten wir für jedes Jahr die Zahl der voraussichtlich sterbenden Versicherten. Dazu haben wir die Sterbetafeln des StatBA verwendet.

Im Screenshot ist das die Spalte D.

In der Tarifkalkulation haben wir mit fiktiven 10.000 Versicherten im Alter von 30 begonnen. (Zeile 4) In Zeile Spalte F können Sie sehen, daß bei der Todeswahrscheinlichkeit von 0,05% (gerundet dargestellt) laut Prognose 5,4 Menschen sterben werden, bevor sie 31 werden. Das wären 5,4*31=167,4 gelebte Jahre.

Unterjähriger Todeszeitpunkt

Wenn wir aber für diese 5,4 Menschen 167,4 Jahre veranschlagen, unterstellen wir, dass alle unmittelbar vor ihrem 31 Geburtstag gestorben sind. Das ist unrealistisch. Das andere Extrem wäre, wenn wir annehmen, dass alle Toten direkt nach ihrem 30 Geburtstag gestorben sind. Dann hätten sie zusammen genau 30 Jahre gelebt. Wenn wir genauere und zuverlässigere Daten haben, ist das schön. In der Praxis geht man davon aus, dass

die Todesfälle gleichmäßig über das Jahr verteilt sind. Bei 5,4 Toten ist das schlecht zu rechnen. Bei 12 Toten wäre einer im Januar, einer im Februar usw. gestorben. Wenn wir diese Annahme treffen, können wir davon ausgehen, dass die Toten eines Jahres im Durchschnitt noch 6 Monate gelebt haben, also 30,5 Jahre alt geworden sind, also insgesamt 167,4 Jahre.[2]

Für das nächste Jahr (von 31 bis 32 Jahren) gibt es ebenfalls eine Überlebenswahrscheinlichkeit. Jetzt ist die Prognose, dass 6,0 Menschen sterben werden, die jeweils 31,5 Jahre alt geworden sind, also zusammen 190,3 Jahre. Alle Toten, die 2019 gelebt haben, aber vor 2021 gestorben sind, haben dann zusammen 335 Jahre gelebt.

Dieses „wie viele Leute werden in diesem Jahr vermutlich sterben“ und „wie lange werden sie gelebt haben“ und das „wie lange werden alle Toten zusammen gelebt haben“ wird bis zum Alter von 100 durchgerechnet und aufsummiert. Das Ergebnis steht in Zeile 74. Dann sind, laut Prognose, noch 68,5 Personen am Leben. Die werden nicht alle vor dem 101. Geburtstag sterben, aber das ignoriert man, weil das das Ergebnis nur auf einer hinteren Nachkommastelle ändert. In einer Rechnung, die sowieso wackelig ist, weil sie sich auf das Jahr 2089 bezieht.

Restlebenserwartung

In G74 sehen Sie, dass alle 10.000 Personen zusammen 793.284,4 Jahre gelebt haben werden, also im Durchschnitt 79,3 Jahre. Dieser Wert ist die Lebenserwartung eines 30jährigen. Etwas ausführlicher: Ein 30jähriger wird im Durchschnitt 79,3 Jahre alt werden. Weil sie schon 30 sind, bedeutet das, dass die Restlebenserwartung 49,3 Jahre beträgt.

Das bedeutet nicht, daß alle 30jährigen in diesem Alter auf einen Schlag tot umfallen. Wir haben gesehen, daß bereits im ersten Jahr ein paar Menschen sterben und eine Reihe von Menschen deutlich älter als 79,3 werden. Dieser Wert ist eben nur ein erwarteter Durchschnittswert oder Erwartungswert.

Querschnitt und Längsschnitt

Alle Zahlen, die der Rechnung zugrunde liegen stammen aus dem Jahr 2019. Die einjährige Sterbewahrscheinlichkeit eines 99jährigen im Jahr 2019 lag bei 38,16% (Zelle D73). Der Trick an der Rechnung ist nun, daß man unterstellt, daß diese 38,16% auch im Jahr 2089 gelten, wenn die 30jährigen 99 sind. Man nimmt also Querschnittszahlen (ein Jahr, viele Altersklassen) und benutzt diese Zahlen als Längsschnittzahlen (viele Jahre, ein Jahrgang).

Das ist natürlich eine wackelige Rechnung. Aus diesem Grund werden die Sterbetafeln auch ständig aktualisiert.

[2]Wenn Sie das mit dem Taschenrechner nachrechnen und 164,7 herauskommt, ist weder Ihr Taschenrechner kaputt noch meine Rechnung falsch. Es ist die alte Leier mit den Excel-Screenshots: Die 0,05% Sterbewahrscheinlichkeit und die 5,4 Toten werden von Excel gerundet dargestellt, aber mit allen Nachkommastellen berechnet.

	1900	1920	1940	1960	1980	2000
0 Jahre	46,32	57,52	68,67	75,5	81,24	83,77
6 Jahre	58,73	64,34	69,57	73,01	76,61	78,25
10 Jahre	55,47	60,82	65,95	69,17	72,68	74,27
20 Jahre	47,41	51,79	56,51	59,72	62,97	64,4
40 Jahre	30,98	34,35	38,22	41,08	43,78	44,9
60 Jahre	15,36	18,14	21,68	23,73	25,77	26,39
80 Jahre	5,64	7,35	8,3	9,4	10,18	10,18

Tabelle 7.2: historische Entwicklung der Restlebenserwartung

F7.5 Restlebenserwartung

Die 2019er Daten kamen zu einer Lebenserwartung von 79,3 Jahren für Menschen, die schon 30 sind. Das bedeutet, daß die Restlebenserwartung 49,3 Jahre ist. Diese Zahl ist interessant, weil man mit ihr analysieren kann, wie die Veränderung der Lebenserwartung im historisch zustande kommt. Die Daten der Tabelle stammen ebenfalls vom StatBA und umspannen 100 Jahre.

Neugeborene

Die Lebenserwartung der Neugeborenen (Also die Restlebenserwartung beim Alter von Null) ist in diesem Zeitraum von 46 auf 84 Jahre, also um 38 Jahre gestiegen. Wenn man sich die Perspektiven eines Erstklässlers anschaut, halbiert sich dieser Zugewinn in etwa. Der 6jährige hatte 1900 eine Restlebenserwartung von insgesamt 59 Jahren, heute von 78, was ein Zugewinn von „nur noch“ 19 Jahren ist. Das klingt seltsam. Der Grund ist, dass damals so viele Kinder *vor* dem sechsten Lebensjahr gestorben sind. Etwa 30%. Heute liegt diese Quote unter 0,5%.

Rentner und Hochbetagte

Vergleicht man die Restlebenserwartung eines Rentners von 65 bzw. eines Hochbetagten von 80 am Anfang und Ende unseres Betrachtungszeitraums, dann sieht man, wie überschaubar der Zugewinn in diesem Zeitraum von 100 Jahren für diese Altersgruppe ist: 10 bzw. 5 Jahre. Schaut man bei diesen Altersgruppen noch genauer hin, kann man sehen, wann diese 10 bzw. 5 Jahre entstanden sind: Im Wesentlichen in der ersten Hälfte des 20. Jhr. Die Veränderungen danach Jahren sind sehr gering.

Einfluss der gesunkenen Kindersterblichkeit

Zusammengefasst lässt sich konstatieren, daß der Zuwachs an Lebenserwartung zum Großteil durch eine drastische Senkung der Kinder- und Säuglingssterblichkeit entstanden ist. Dieser Effekt ist aber ausgeschöpft, da schon seit 40 Jahren mehr als 99% der Neugeborenen auch das Grundschulalter erreichen.

Das bedeutet, daß eine weitere Steigerung der Lebenserwartung durch eine geringere Sterblichkeit in den höheren Altersklassen erfolgen müsste und die Daten zeigen, wie mühsam dieses Geschäft ist. Daher verstehe ich auch nicht, welche Grundlage Aussagen wie *wir werden bald alle 120 Jahre alt* haben. Selbst wenn wir es künftig schaffen würden,

alle Todesfälle vor dem Alter von 80 zu vermeiden, sieht es nicht so aus, als würde die Lebenserwartung dann deutlich über 90 Jahre steigen.

Effekt der Hochleistungsmedizin

Die Entwicklung der Restlebenserwartung in mittlerem und höherem Alter deutet daraufhin, dass der Effekt der modernen Hochleistungsmedizin nicht allzu groß ist. Würden wir die medizinische Versorgung auf den Stand von 1980 zurückschrauben, würden wir das an der Restlebenserwartung der Rentner kaum merken.

F7.6 Anwendung auf die Berechnung der € je QALY

In der folgenden Tabelle finden Sie eine Überschlagsrechnung für eine Hüft-TEP. Die OP-Kosten liegen bei etwa 7.600€ in Deutschland. Die anschließende Reha schlägt mit etwa 2.000€ zu Buche. Es sei angenommen, daß es keine weiteren Folgekosten gibt und die Prothese bis zum Lebensende nicht ausgewechselt werden muß. Der Lebensqualitätsgewinn sei (die Zahl ist gegriffen) mit 10% veranschlagt, der Diskontsatz sei 3%.

Mit diesen Zahlen kann man für jede Restlebensdauer (Spalte A) die €/QALY (Spalte E) berechnen. Ich zäume das Pferd also von hinten auf: Wie lange muss der Patient noch leben, damit ein QALY x Euro kostet.

Aus dieser Rechnung ergibt sich, daß eine OP bei einem 65jährigen, der (laut der Restlebenserwartungs-Tabelle) noch durchschnittlich 22 Jahre leben wird, Kosten von 6.023,75€/QALY verursachen wird (Zeile 28). Die gleiche OP bei einem 80jährigen kostet genauso viel, spendet dem Patienten aber nur durchschnittlich 10 Jahre lang Nutzen. Dadurch steigen die Kosten je QALY auf 11.254,13€ (Zeile 16).

	A	B	C	D	E
1	Kosten OP		7.600,00 €		
2	Kosten Reha		2.000,00 €		
3	Gewinn LQ		0,10		
4	Diskontsatz		3,0%		
5					
6	Jahr	Disk.faktor	disk. LQ	kumul. LQ	€/QALY
7	1	1,030	0,097	0,097	98.880,00 €
8	2	1,061	0,094	0,191	50.170,64 €
9	3	1,093	0,092	0,283	33.938,91 €
10	4	1,126	0,089	0,372	25.826,60 €
11	5	1,159	0,086	0,458	20.962,04 €
12	6	1,194	0,084	0,542	17.721,36 €
13	7	1,230	0,081	0,623	15.408,61 €
14	8	1,267	0,079	0,702	13.675,81 €
15	9	1,305	0,077	0,779	12.329,65 €
16	10	1,344	0,074	0,853	11.254,13 €
17	11	1,384	0,072	0,925	10.375,43 €
18	12	1,426	0,070	0,995	9.644,36 €
19	13	1,469	0,068	1,063	9.026,84 €
20	14	1,513	0,066	1,130	8.498,53 €
21	15	1,558	0,064	1,194	8.041,59 €
22	16	1,605	0,062	1,256	7.642,64 €
23	17	1,653	0,061	1,317	7.291,44 €
24	18	1,702	0,059	1,375	6.980,03 €
25	19	1,754	0,057	1,432	6.702,13 €
26	20	1,806	0,055	1,488	6.452,71 €
27	21	1,860	0,054	1,542	6.227,69 €
28	22	1,916	0,052	1,594	6.023,75 €

Der Preis von 11.254,13€/QALY für den 80jährigen ist nicht allzu hoch.[3] Das ändert sich auch nicht grundlegend, wenn man etwas pessimistischer rechnet. Der Spruch *„keine künstliche Hüfte ab 75“* ist also in diesem Fall gesundheitsökonomischer Blödsinn, aber die Rechnung macht deutlich, welche Bedeutung das Alter der Patienten hat. Bei einer Restlebenserwartung von zwei Jahren würde man wahrscheinlich anders denken.

F7.7 Die Lebenserwartung als gesundheitsökonomischer Global-Indikator

Es liegt nahe, die Lebenserwartung als eine Art Grobindikator für die Leistungsfähigkeit eines Gesundheitssystems zu verwenden. Man könnte dann zwar nicht einzelne Leistungen beurteilen, aber immerhin das System als Ganzes.

Diese Idee ist nur bedingt sinnvoll, da nur ein Teil einer Veränderung der Lebenserwartung auf das Gesundheitssystem zurückzuführen ist. Wir haben schon über das Beispiel der Masernimpfung gesprochen. Ohne den Impfgegnern das Wort zu reden, muß man konstatieren, daß die Zahl der Maserntoten schon in den 1950er Jahren um 90% gesunken war. Die Impfstoffe sind aber erst in den 1960er Jahren entwickelt worden. Der Kampf gegen die Masern ist also in erster Linie erfolgreich gewesen, weil sich die Lebensbedingungen der Kinder verbessert haben und nicht, weil man sensationelle medizinische Fortschritte

[3] In *Einheit Das Gesundheitssystem in Großbritannien* werden wir sehen, dass selbst die für ihre Sparsamkeit in Gesundheitsdingen bekannten Briten mit dieser Summe überhaupt kein Problem hätten. Bei den gewählten Werten würden sie erst bei einer Restlebenserwartung von weniger als vier Jahren die Augenbrauen runzeln.

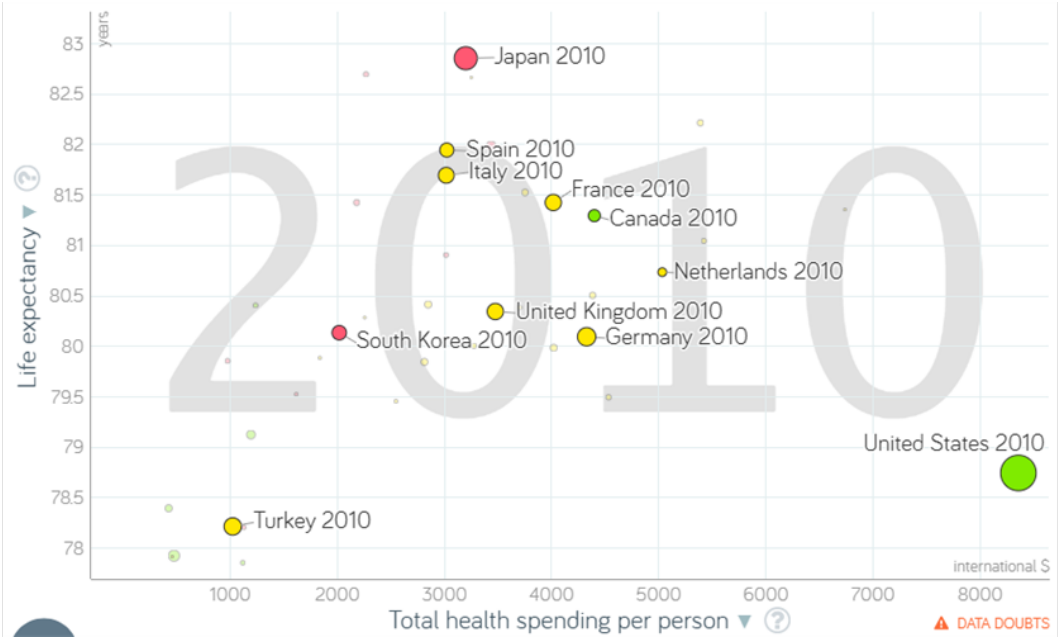


Abbildung 7.1: Gesundheitsausgaben und Lebenserwartung

gemacht hat.[4] Es wäre also eine grobe Überschätzung, wenn man die Lebensverlängerung ausschließlich auf das Gesundheitssystem zurückführen würde.

Wenn die Entwicklung der Lebenserwartung im Längsschnitt ein wackeliger Indikator für die Leistungsfähigkeit des Gesundheitssystems ist, so kann die Lebenserwartung im Querschnitt, d.h. im internationalen Vergleich aufschlussreich sein. Die Weltbank-Daten für das Jahr 2010 zeigen, daß die Lebenserwartung in den westlichen Industrieländern sich um ein bis zwei Jahre unterscheiden.[5] Auf der horizontalen Achse sind die Pro-Kopf Ausgaben abgetragen. Auch hier zeigt sich kein großer Unterschied zu unseren Nachbarländern. Dieser Vergleich zeigt uns, daß Deutschland sowohl was die Lebenserwartung als auch die Gesundheitsausgaben betrifft, im Mittelfeld liegt.

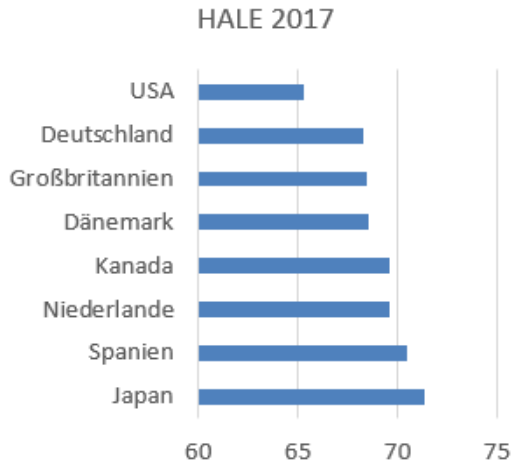
Insgesamt lässt sich aber die Einschätzung, das deutsche Gesundheitssystem sei das beste der Welt, nicht aufrechterhalten. Es ist aber auch nicht so grottenschlecht, so daß man von grundlegenden Reformen gravierende Verbesserungen oder Einsparungen erwarten könnte.

F7.8 Healthy Life Expectancy

Die Healthy Life Expectancy (HALE) ist eine Modifikation des Konzepts der Lebenserwartung. Ähnlich wie das QALY berücksichtigt HALE, daß nicht nur die Lebensdauer, sondern auch die Lebensqualität bedeutsam ist. HALE ist die erwartete Lebensdauer, die eine Person in gutem Gesundheitszustand lebt.

[4] Illich, Ivan. Die Nemesis der Medizin - die Kritik der Medikalisierung des Lebens, München: Beck, 2007, S. 18. Illich ist ein fast immer polemischer Autor und nicht immer zuverlässig. An dieser Stelle schon.

[5] Auf http://www.gapminder.org finden Sie ein ganz hervorragendes kostenloses Visualisierungtool für gesundheitsökonomische Daten mit dem die Grafik gemacht ist. Es ist eng mit dem Buch Factfulness von Hans Rosling verbunden, das aus meiner Sicht ein ganz hervorragender Lesetipp ist. Das Buch gibt es (gleicher Titel) auch auf Deutsch.



Die Grafik zeigt die Ergebnisse für einige ausgewählte Länder.[6] Das Ergebnis der Studie ist, das in fast allen Ländern HALE und die reine Lebenserwartung um etwa 8 Jahre auseinanderlagen.

F7.9 Ehemalige Klausuraufgaben

Alternative Berechnungsmethoden der Lebenserwartung

Die Methode des Statistischen Bundeamts über Sterbetafeln ist kompliziert und nicht sehr intuitiv. Könnte es nicht einen viel einfacheren Weg geben? Folgende Kandidaten wären denkbar.

1. Man wählt einen Geburtsjahrgang aus, der inzwischen komplett verstorben ist. Für diese Personen ermittelt man das Sterbejahr und erhält so die Lebensdauer. Die durchschnittliche Lebensdauer verwendet man als Schätzung der Lebenserwartung.

2. Man ermittelt das Alter aller im letzten Jahr Verstorbenen. Den Durchschnitt verwendet man als Schätzung für die Lebenserwartung

Bewerten Sie diese Konzepte vor dem Hintergrund, daß es medizinischen Fortschritt gibt.

Lebenserwartung des Wolpertingers

Der Wolpertinger ist ein in Bayern weltbekanntes Fabeltier. Leider ist es nicht sehr langlebig. Das älteste bekannte Exemplar ist am Tag vor seinem dritten Geburtstag gestorben, ist also knapp 36 Monate alt geworden. Die Gesellschaft der Freunde des Wolpertingers führt eine Studie durch. Aus dem zentralen Zuchtregister werden alle am 1.4.

[6] Kyu, Hmwe Hmwe et al. , "Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017", The Lancet, 10159 (392) 2018. S.1859.

geborenen Wolpertinger aus den Jahren 2012, 2013 und 2014 gezogen. Die Besitzer werden am 1.4. 2014 angerufen und gefragt, ob der Wolpertinger noch lebt.

Auf diese Weise können die Daten von 2.225 Tieren erhoben werden. Nach einem Jahr, also am 1.4.2015 werden die Besitzer ein zweites Mal angerufen und gefragt, ob der Wolpertinger noch lebt. Das ist bei insgesamt 1.000 Tieren der Fall.

	A	B	C	D	E
1			geboren am		
2		01.04.2012	01.04.2013	01.04.2014	Summe
3	Lebend am 1.4.2014	425	950	850	2225
4	Lebend am 1.4.2015	0	400	600	1000

Wie hoch ist die Lebenserwartung eines Wolpertingers in Monaten? Geben Sie den Wert mit einer Nachkommastelle an.

Natürlich ist die Frage albern, aber die Berechnung ist die gleiche wie beim Menschen. Nur ist der Wolpertinger schneller tot, so daß man noch „mit der Hand“ rechnen kann.

Zähe Hundertjährige

Das statistische Bundesamt, das die Sterbetafeln herausgibt, berechnet die einjährigen Überlebenswahrscheinlichkeiten bis zum Alter von 100. Nun ist aber klar, daß es einige Hundertjährige gibt, die noch älter werden. In dem Abschnitt, in dem wir die Restlebenserwartung von 30jährigen berechnet haben, haben wir das ignoriert. Die folgende Tabelle ist noch einmal ein Ausschnitt mit den wichtigsten Zahlen.

	A	BS	BT
1	**Alter**	**99**	**100**
2	P(Überl.)	0,638	0,618
3	P(Tod)	0,362	0,382
4	n	49	30
5	Tote	28	19
6	gelebte Jahre	2.755	1.873
7	Summe gelebte Jahre	754.694	756.567

Wie gravierend dieses Ignorieren ist, sollen Sie nun anhand eines Extremszenarios überprüfen. Nehmen wir an, daß alle Personen, die 100 Jahre alt geworden sind, auch 120 Jahre alt werden und dann schlagartig sterben. Ich denke, das ist ein Szenario, das einen sicherlich zu hohen Wert liefern wird. Wie würde sich das auf die Restlebenserwartung der 30jährigen auswirken?

Historische Entwicklung der Lebenserwartung in Deutschland

In der Grafik „Entwicklung der Lebenserwartung. . .“ unten auf der Seite, stellen Sie fest, daß die Zuwächse zu Beginn der Zeitreihe viel größer sind als gegen Ende. Dort flacht die Kurve ab.

Es gibt eine sehr einfache Erklärung dafür. Welche?

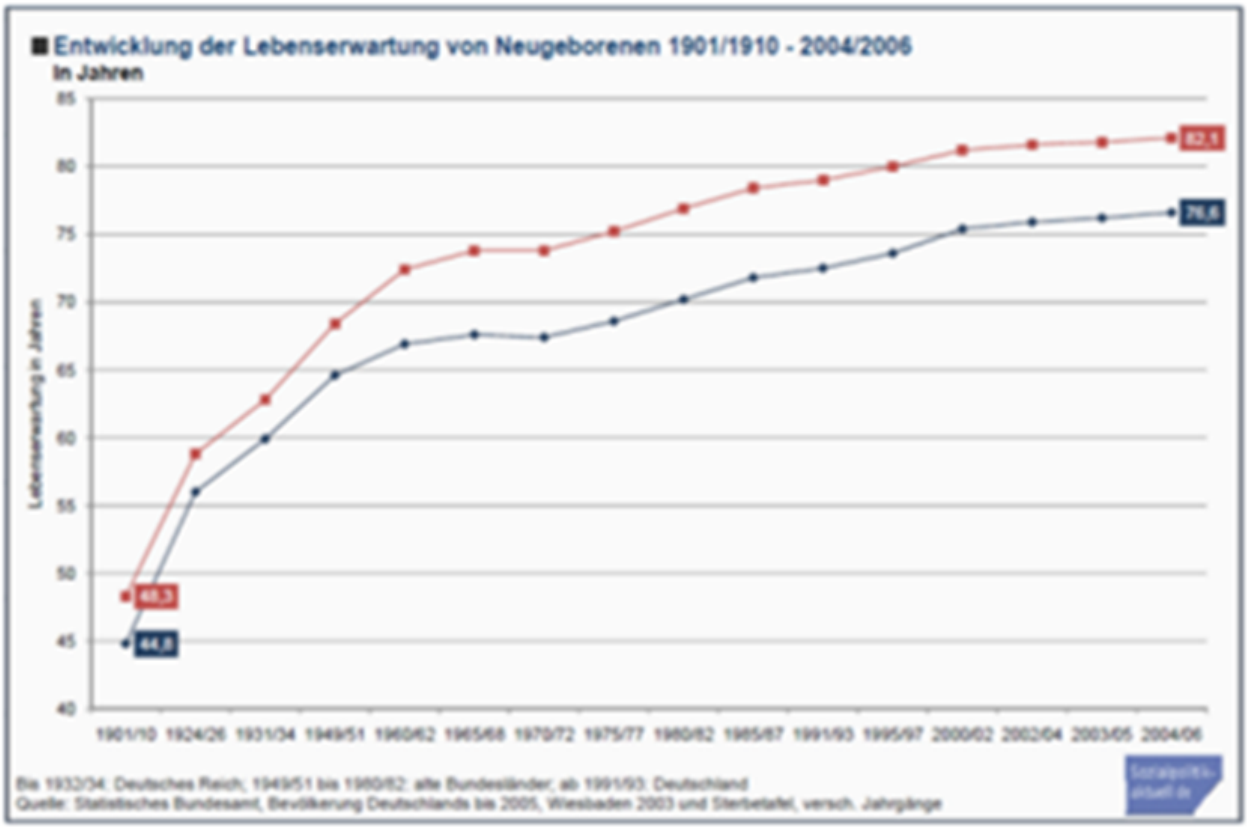


Abbildung 7.2: Historische Entwicklung

Alternative Berechnungsmethoden (Lösung)

- In dem komplett verstorbenen Jahrgang gibt es Personen, die in der Vergangenheit an Dingen gestorben sind, an denen sie heute nicht mehr sterben würden. Sie sind also aus heutiger Sicht „zu jung“ gestorben.
- Diese „zu jung“ Gestorbenen sind vor vielleicht 20 Jahren gestorben. Hätten Sie damals die heutige medizinische Versorgung gehabt, wären sie erst letztes Jahr gestorben. Sie fehlen also bei den im letzten Jahr Verstorbenen, weil sie schon lange tot sind, aus heutiger Sicht aber nicht hätten sterben müssen.

Beide Verfahren führen also dazu, daß die Gewinne an Lebenserwartung durch medizinischen Fortschritt sich erst mit sehr großem Zeitverzug in der Kennzahl niederschlagen.

Lebenserwartung des Wolpertingers (Lösung)

- Von den 850 Neugeborenen leben nach 12 Monaten noch 600. Die Überlebenswahrscheinlichkeit beträgt 600/850=0,706.
- Von den 950 Einjährigen leben nach 12 Monaten noch 400. Die Überlebenswahrscheinlichkeit beträgt 400/950=0,421.
- Von den 425 Zweijährigen lebt nach 12 Monaten keiner mehr. Die Überlebenswahrscheinlichkeit ist Null.
- Jetzt startet man mit einer fiktiven großen Zahl Neugeborener. Z.B. 100.000 Im Jahr 2015 leben von den 100.000 Neugeborenen 70.600 Tiere jeweils 12 Monate (die Überlebenden) 29.400 Tiere jeweils durchschnittlich 6 Monate (die gleichmäßig über das Jahr Gestorbenen
- Insgesamt sind das 70.600*12 + 29.400*6 = 1.023.600 Monate
- Im Jahr 2016 leben anfangs noch 70.600 Tiere
- 42,1% dieser Tiere überleben auch 2015, d.h. 29.723 Tiere leben 12 Monate 57,9% der Tiere versterben, d.h. 40.877 leben durchschnittlich 6 Monate
- Insgesamt sind das 29.723*12 + 40.877*6 = 601.938 Monate
- Im Jahr 2017 leben anfangs noch 29.723 Tiere Alle Tiere sterben. Im Durchschnitt nach 6 Monaten Insgesamt sind das 29.723*6 = 178.338 Monate
- In allen Jahren zusammen haben die 100.000 Tiere insgesamt 1.023.600+601.938+178.338=1.803.876 Monate gelebt pro Tier sind das im Durchschnitt 18,03 Monate.

Historische Entwicklung (Lösung)

Die Zeitabstände auf der horizontalen Achse sind nicht gleich. Zwischen den ersten beiden Werten liegen über 20 Jahre, zwischen den letzten beiden ein Jahr. Überschlägt man den jährlichen Zuwachs, ist die Kurve relativ linear.

So würde ich an die Aufgabe herangehen:

Gesucht ist eine Erklärung für die dargestellten Daten. Was für Erklärungen könnten das sein? Als erstes suche ich Typen von Erklärungen. Mir sind drei eingefallen:

- Irgendetwas hat sich in den Lebensumständen der Menschen so geändert, daß sich die Lebenserwartung erhöht hat.
- Die Daten stimmen nicht
- Wenn man genauer hinsieht, sagen die Daten nicht das, was sie zu sagen scheinen.

Bevor ich mich auf die erste Gruppe stürze und viel Text produziere, denke ich besser noch mal nach, weil mir sonst die Zeit für die anderen Gruppen fehlt. Bei den Lebensumständen fallen mir eine ganze Reihe plausibler Faktoren ein. Hygiene, Kriege, medizinischer Fortschritt. Könnte das die Lebenserwartung erhöht haben? Könnte. Aber welcher Faktor welches Gewicht hat, kann ich ohne weitere Recherche nicht sagen.

Ich kann jetzt lang und breit improvisieren, versuchen, mich an Beispiele zu erinnern und irgendetwas Schwammiges in Richtung *„all diese Faktoren spielen eine Rolle, aber man kann nicht sagen, . . . “* zu Papier bringen.

Ist das das, was ich tun soll? Dann müsste die Frage aber anders gestellt sein. Irgendetwas in Rich-tung *„Identifizieren Sie mögliche Faktoren und diskutieren Sie die Bedeutung. . . “* Steht da aber nicht. Da steht *eine einfache Erklärung.* „eine“ ist auch noch kursiv, also handelt es sich wohl um eine genau benennbare Erklärung.

Die Fragestellung winkt mit einem Zaunpfählchen. Mit meinem Hygiene-Krieg-Fortschritt-Medley komme ich also nicht ans Ziel.

Könnte es sein, daß die Daten falsch sind? Könnte sein, aber habe ich eine Chance, das zu erkennen? Säße ich an meinem Schreibtisch, würde ich kurz googlen, ob in anderen Quellen andere Zahlen auftauchen, aber in einer Klausur kann ich das nicht. Wenn die Daten falsch wären, ich das aber nicht erkennen kann, das aber die Lösung wäre, wäre das keine sinnvolle Fragestellung. Also kann ich auch diese Interpretation verwerfen.

Kann es also sein, daß ich eine Aussage in die Daten hineinlese, die gar nicht da ist? Wie könnte es dazu kommen?

- Ich habe sehr schlampig hingeschaut.
- Der Ersteller der Grafik führt mich bewusst irre.
- Der Ersteller der Grafik führt mich unabsichtlich irre.

Es könnte sein, daß ich schlampig hingeschaut habe. Ich habe gedacht, daß die Grafik etwas über die Höhe der Lebenserwartung im 20 Jhr. sagt, aber wenn ich noch einmal konzentriert hinschaue, was in der Überschrift und auf den Achsen steht, könnte ich feststellen, daß es gar nicht um die Lebenserwartung, sondern (sagen wir) den Anteil von Rentnern an der Bevölkerung geht. Das hätte grob etwas miteinander zu tun, wäre aber etwas deutlich Verschiedenes. Vielleicht geht es auch gar nicht um das 20. Jhr. sondern um das Mittelalter. Also schaue ich mir die Achsen gründlich an. . . und stolpere über die

angegebenen Jahreszahlen. Die zeitlichen Abstände zwischen den Datenpunkten sind gar nicht gleich. Die Grafik arbeitet mit Datenpunkten. Sie gibt also an, daß es nicht für jeden Zeitpunkt Daten gibt, sondern nur in gewissen Abständen. Dadurch, daß die Grafik die Punkte aber verbindet, suggeriert sie mir einen kontinuierlichen Verlauf und suggeriert mir, daß ich es mit einem Koordinaten-kreuz (einer XY-Grafik) zu tun habe. Habe ich aber gar nicht. Die Grafik führt mich also in die Irre. Entweder ist das Absicht (glaube ich nicht. Warum sollte der Autor das tun?) oder der Mensch, der die Grafik erstellt hat, hat an dieser Stelle nicht genug nachgedacht. Eine Säulengrafik wäre besser gewesen.

Zähe Hundertjährige (Lösung) Es gibt noch 30 Personen, die 100 Jahre alt sind (Zelle BT4) Diese Personen würden an-nahmegemäß noch 20 Jahre leben.

Insgesamt wären das 600 zusätzlich gelebte Jahre. Statt 756.567 gelebten Jahren (Zelle BT7) wären es jetzt 757.167 Jahre. Bei einer Ausgangspopulation von 10.000 Personen wären das 75,72 Jahre. Da die Population bereits 30 Jahre alt ist, beträgt die Restlebenserwartung 45,72 Jahre

Teil III

Epidemiologische Studientypen

Drei wichtige Begriffe

Für diesen Teil der Veranstaltung, der sich mit epidemiologischen Studientypen befasst, gibt es drei wichtige Begriffe, die ich kurz skizzieren möchte.

Ätiologie

Wenn sich ein Kind einen Arm bricht oder Masern bekommt, die die *Entstehung* (Ätiologie) dieser Krankheit relativ offensichtlich. Bei den typischen Volkskrankheiten (Bluthochdruck, Diabetes, ...) ist das nicht so. Es lassen sich statistische Zusammenhänge ermitteln, die z.B. darin bestehen, daß übergewichtige Menschen mit einer höheren Wahrscheinlichkeit einen Diabetes Typ II entwickeln werden als Normalgewichtige, aber es gibt Normalgewichtige mit Diabetes und Übergewichtige ohne Diabetes.

Epidemiologie

Hier kommt die Epidemiologie ins Spiel. Dieser Begriff klingt nach „Epidemie“ und „Corona“. Tatsächlich ist die Analyse der Infektionskrankheiten der Ursprung der Epidemiologie. In diesem Grenzbereich zwischen Medizin, Statistik und Sozialwissenschaften geht es aber meist nicht um virologische und medizinische Forschung, sondern um die Frage, wie massenhaft auftretende Krankheiten sich verbreiten. Diese massenhaft auftretenden Krankheiten waren eben klassischerweise Infektionskrankheiten und das Thema von Epidemiologie war, wie man die Verbreitung verhindern – oder noch besser – den Ausbruch insgesamt verhindern könnte.

Die Strategien, die im Rahmen der Corona-Pandemie diskutiert wurden (Isolation der Risikogruppen, Isolation der Erkrankten, Impfung, Herdenimmunität,...) sind Strategien dieser Infektions-Epidemiologie.

Die spanische Grippe am Ende des ersten Weltkriegs war die letzte große Epidemie. Die Opferzahlen lagen (je nach Zählung) weltweit bei etwa 20-40 Mio.[7] In den folgenden 100 Jahren spielten Infektionskrankheiten auf globaler oder auch nur nationaler Ebene keine große Rolle mehr, so dass andere Krankheiten in den Fokus der Epidemiologie kamen. Weiterhin ging es um massenhaft auftretende Erkrankungen, aber jetzt nicht mehr um

[7] Spinney, Laura. 1918 - die Welt im Fieber - wie die Spanische Grippe die Gesellschaft veränderte, München: Carl Hanser, 2018.

Infektionskrankheiten, sondern die sogenannten „Volkskrankheiten“ wie Bluthochdruck und Diabetes.

Natürlich gab es weiterhin Epidemiologen, die sich mit Infektionskrankheiten befasst haben, aber vor Corona waren das relativ wenige. Es könnte sein, dass sich das jetzt än-dert und die Infektionskrankheiten wieder stärker in den Mittelpunkt kommen.

Risikofaktoren

Wie bereits erwähnt, gibt es keine direkte Ätiologie bei Volkskrankheiten, wie bei den Infektionskrankheiten. Es lassen sich nur statistische Korrelationen finden. Um hier weiterzukommen braucht man ein anderes statistisches Instrumentarium.

Die Volkskrankheiten-Epidemiologie versucht, Risikofaktoren zu identifizieren, d.h. Umstände, die eine Erkrankung wahrscheinlicher machen. Dazu sind Risikokennzahlen notwendig, mit denen man die Bedeutsamkeit eines Faktors quantifizieren kann und man muss Studien durchführen, um die Daten zu erhalten, anhand derer man die Kennzahlen überhaupt erst berechnen kann. In diesem Teil der Veranstaltung lernen Sie Studientypen und Risikokennzahlen kennen, die in der Epidemiologie verwendet werden.

Kapitel 8

Kohortenstudien

F8.1 Das Konzept

Die Grundidee einer Kohortenstudie besteht darin, eine Personengruppe (die Kohorte) zu definieren, die über einen langen Zeitraum beobachtet wird. Die Beobachtung besteht in regelmäßigen Untersuchungen und Befragungen der Studienteilnehmer (Probanden). Zu Beginn der Studie sind die Probanden normalerweise gesund und man wartet den Studienverlauf ab, und beobachtet, ob und bei wem die Krankheit auftritt.

Die bekannteste Kohortenstudie wurde 1948 in Framingham, Massachusetts durchgeführt. Sie läuft immer noch mit den Kindern und Enkeln der ersten Probanden und untersucht Einflussfaktoren für koronare Herzerkrankungen. Ein Großteil des Wissens über die Bedeutung von Blutdruck, Blutfetten und Blutzucker für Herzerkrankungen stammt aus dieser Studie.

F8.2 Datensparsamkeit

Die regelmäßig ermittelten Daten werden nicht „ins Blaue“ hinein erhoben. Wenn eine Studie nach Faktoren suchen, die die Entwicklung von Tumoren in der Leber begünstigen, dann werden die Probanden nicht nach ihrer sexuellen Orientierung befragt werden, weil das ja vielleicht spannend sein könnte. Sie werden nur über die Dinge befragt und auf die Dinge hin untersucht, die im Verdacht stehen, die Krankheit zu begünstigen und deren Erhebung für die Fragen der Studie sinnvoll sind.

Bevor eine solche Studie durchgeführt wird, braucht man daher Hypothesen, die man durch die Studie überprüft. Wenn man sich nicht an dieses Prinzip der Datensparsamkeit hält, werden unnötige Erhebungskosten verursacht und die Bereitschaft der Probanden, sich wiederholt befragen und untersuchen zu lassen, wird gering sein, wenn offensichtlich ist, daß die Datenerhebung ziellos ist und einfach möglichst viele Daten abgefragt werden sollen.

Die Relevanz des Studiendesigns

Umgekehrt bedeutet das, dass man Versäumnisse, die man zu Beginn der Studie begangen hat, nur schwer wieder ausbügeln kann, weil man die Daten nicht hat. Zusammengefasst bedeutet das, daß die anspruchsvolle Arbeit im Design der Studie steckt. Die eigentliche Durchführung ist dann zum Großteil nur Fleißarbeit. Sollten Sie eine empirische Thesis ins Auge fassen, ist das eine wichtige Überlegung für Sie.

Diese Ausführungen gelten grundsätzlich für jeden Studientyp. Das, was man am Anfang nicht fragt, wird man am Ende nicht wissen. Bei der Kohortenstudie ist dieses Problem aber noch einmal größer, weil die Laufzeit viel länger ist und man die Suppe auslöffeln muss, die einem der eigene Vorvorvorgänger vor 20 Jahren eingebrockt hat.

Gegenmodell Panel

Im letzten Absatz habe ich gesagt, dass Datensparsamkeit ein Prinzip ist, das für *jeden* Studientyp gilt. Ich möchte kurz auf eine Datenquelle eingehen, bei der das nicht so ist: Das Sozioökonomische Panel (SOEP), das seit 1984 vom Deutschen Institut für Wirtschaftsforschung jährlich durchgeführt wird.[1]

Etwa 15.000 Haushalte mit 30.000 Personen werden mit einem standardisierten Fragebogen befragt, der durch kleine Zusatzfragen ergänzt wird. Ein *Panel* ist eine Befragung der gleichen Personengruppe mit den gleichen Fragen über einen längeren Zeitraum. Das hört sich ähnlich wie eine Kohorte an. Der Unterschied ist, dass das Panel keine eigene Fragestellung hat, sondern Daten zur Verfügung stellen will, auf deren Basis man eigene Auswertungen fahren kann.

In *Einheit Messung von Lebensqualität* haben wir über den *Hapiness Set Point* nachgedacht. Dort haben wir uns die Studie von Clark et al angesehen, die untersucht hat, wie Menschen mit großen (positiven wie negativen) Veränderungen in ihrem Leben zurechtkommen.[2] Diese Studie hat keine eigenen Daten erhoben, sondern auf vorhandene Daten des SOEPs zurückgegriffen. Die Frage, die dann zu dieser Studie geführt hat, war, *was ist da und was kann man daraus machen.* Wenn jemand in irgendeinem Jahr seit 1984 einen Arbeitsplatz hatte und im Jahr darauf arbeitslos war, konnte man messerscharf schließen, dass er arbeitslos war. Wenn er im Jahr darauf immer noch arbeitslos war und danach immer noch, konnte man sich an einem Finger abzählen, dass er langzeitarbeitslos war. Das mag 10 Jahre her gewesen sein, aber egal. Zu der gleichen Person hatte man Lebenszufriedenheitsdaten und jetzt konnte man diese Lebenszufriedenheitsdaten mit der Arbeits(losigkeits)biographie der Person koppeln. Die wichtigste Frage war: *Hat das schon jemand gemacht?* Nein.[3] War nicht so. Also weiter...

Bei diesem Studientyp braucht man also sehr geduldige Probanden, weil man ins Blaue fragt und heute noch nicht weiß, welche Fragestellung irgendeinem Forscher in zehn Jahren zu den Daten wohl einfallen mag. Der Forscher muss dann aber mit den Daten in der Form leben, die sie haben. Wenn er meint, dass die Lebensqualitätsabfrage (wir haben sie in *Einheit Messung von Lebensqualität* angeschaut) blöd ist, hilft das nichts. Das sind die Daten, die da sind.

---

[1] https://www.diw.de

[2] Clark, Andrew E. et al. Lags and Leads in Life Satisfaction - A Test of the Baseline Hypothesis, Berlin: DIW-Diskussionspapier, 2003.

[3] Weil diese Frage so wichtig ist, da das SOEP öffentlich zugänglich ist, gibt es eine Datenbank mit allen Studien, die auf SOEP-Daten basieren und Clark hätte auch Pech haben können und feststellen müssen, dass jemand vor ihm auf die gleiche Idee gekommen war.

F8.3 Probleme von Kohortenstudien

Kosten

Kohortenstudien sind relativ teuer, weil man eine relativ große Gruppe von Personen braucht, da die Probanden anfangs gesund sind und man viele Jahre abwarten muß, bis einige Teilnehmer die Krankheit entwickeln. Das bedeutet, daß man auf diese Weise nur relativ häufig auftretende Krankheiten untersuchen kann, da bei seltenen Krankheiten, die vielleicht nur bei einer von fünfzigtausend Personen auftritt, die Kosten zu hoch wären.

Ethische Probleme

Kohortenstudien setzen voraus, daß man Vermutungen über krankheitsauslösende Faktoren hat. Im Sinne der Datensparsamkeit wird man die Werte, die man erhebt, auf die Vermutungen, die man belegen oder verwerfen will, ausrichten. Diese Vermutungen dürfen aber keinen Einfluss auf Untersuchung und Behandlung der Probanden haben. Wenn man vermutet, daß ein hoher Blutdruck das Risiko für einen Herzinfarkt erhöht, dürfen die Mitarbeiter der Studie einem Probanden mit hohem Blutdruck nicht anraten, Blutdrucksenker zu nehmen, da sie so das Ziel der Studie gefährden.

Da man Menschen so in ein vermutetes Unglück laufen läßt, gibt es ethische Kritik an diesem Studientyp.

F8.4 Das relatives Risiko als Kennzahl

Sie haben schon Zigarettenpackungen mit einem Warnhinweis gesehen. Mit den epidemiologischen Vorkenntnissen, die wir jetzt haben, wissen wir, dass diese Aussage, genau genommen, falsch ist. Rauchen *verursacht* keinen Lungenkrebs, sondern macht die Erkrankung *wahrscheinlicher*. Das ist inzwischen Allgemeinwissen, aber woher stammt diese Erkenntnis? Eine wichtige Datenquelle ist die bereits erwähnte Framingham-Studie.

Eine fiktive Raucherstudie

Nehmen wir an (die Daten sind fiktiv), dass wir 20.000 Leute über 40 Jahre beobachten. Von diesen 20.000 sind 12.500 Nichtraucher und 7.500 Raucher. Am Anfang der Studie sind alle 20.000 gesund. Am Ende der Studie sind 500 Raucher an Lungenkrebs erkrankt und 50 Nichtraucher. Bei den 50 Rauchern ist klar, dass Rauchen nicht die Ursache sein kann, weil sie ja gar nicht geraucht haben. Also gibt es eine Art „Grundrisiko“, aber uns ist (so wie ich die Zahlen gewählt habe) klar, dass das Risiko bei Rauchern viel höher

	A	B	C	D	E
1			**Outcome**		
2			**kein Lungenkrebs**	**Lungenkrebs**	Summe
3	**Exposition**	**Raucher**	7000	500	7500
4		**Nichtraucher**	12450	50	12500
5		Summe	19450	550	20000
6					
7	durchschnittliches Risiko		2,75%		
8	Risiko der Exponierten		6,67%		
9	Risiko der nicht Exponierten		0,40%		
10	relatives Risiko		16,67		

Abbildung 8.1: fiktive Raucherstudie

ist, weil die Zahl der Raucher kleiner ist, aber die Zahl der Krebskranken höher. Dieses *„das Risiko ist höher“* kann man quantifizieren. Dazu habe ich die Daten in eine Tabelle übertragen.

In dieser Tabelle bedeutet *Exposition* in den Zeilen in Epidemiologendeutsch, die Unterscheidung zwischen Probanden, die den Risikofaktor „haben“, also rauchen (das sind die „Exponierten“) und die, die den Risikofaktor nicht haben (das sind die „Nichtexponierten“).

In den Spalten findet man das *Outcome*, d.h. die Unterscheidung, was am Ende passiert ist, also „kein Lungenkrebs“ bzw. „Lungenkrebs“.

Wenn 550 von 20.000 Personen erkranken, dann ist das durchschnittliche Risiko einer Person 2,75%. Im Hinterkopf haben wir aber die Vermutung, dass Rauchen eine Rolle spielt. Wir waren so klug, die Probanden danach zu fragen, ob sie rauchen oder nicht und können daher diese Risiken für die beiden Gruppen getrennt berechnen.

Wir stellen fest, dass Raucher ein Risiko von 6,67% haben und Nichtraucher von 0,4%. Wenn wir diese beiden Risiken $\left(\frac{Risiko\,der\,Exponierten}{Risiko\,der\,Nichtexponierten}\right)$ durcheinander dividieren, d.h. $\frac{6{,}67}{0{,}4} = 16{,}675$ dann erhalten wir das relative Risiko. Also ist das Risiko der Raucher 16,675 mal so hoch wie das der Nichtraucher.[4]

Abgrenzungsprobleme

Ich habe mit erfundenen Zahlen gearbeitet, damit klar ist, dass das Ergebnis so nicht stimmen wird. Wenn wir aber über unser einfaches Beispiel etwas nachdenken, werden wir schnell merken, dass es auch mit besseren Zahlen nicht einfacher wird.

Ich habe unterstellt, dass es zwei durchgängige Gruppen gibt. Raucher und Nichtraucher. Aber was bedeutet „Nichtraucher“? Heißt das „hat noch nie eine Zigarette geraucht, auch als Teenie nicht“? Dann werden wir wenige Nichtraucher finden. Häufig wird *„weniger als 400 Zigaretten“* als Nichtraucher definiert. Aber diese „400“ ist etwas willkürlich und wir müssten testen, ob 200 Zigaretten mehr oder weniger einen Unterschied machen und wenn ja, wie wir das interpretieren sollen.

Was ist mit denen, die ihren „Status“ wechseln? Also während der Studiendauer mit dem Rauchen aufhören oder anfangen oder erst aufhören und nach x Jahren wieder an-

[4]Das ist der Wert für unsere erfundenen Zahlen. Das tatsächliche relative Risiko ist etwa doppelt so hoch.

Geschätzte Zahl der HIV-Neuinfektionen in Deutschland im Jahr 2019[5)]		
	Gesamtzahl	**2.600 (2.400 – 2.900)**
	Männer	2.200 (2.000 – 2.400)
	Frauen	480 (400 – 550)
Nach Infektionsweg	Sex zwischen Männern	1.600 (1.500 – 1.800)
	Heterosexuelle Kontakte	650 (570 – 740)
	i.v. Drogengebrauch	360 (290 – 450)
	Mutter-Kind-Transmission[4)]	< 10

Abbildung 8.2: RKI-Daten zur HIV-Infektionen

fangen? Werfen wir diese Wechsler aus der Studie heraus oder berechnen wir sie getrennt oder versuchen wir ihre Daten doch irgendwie in einen großen Datenpool aufzunehmen, indem wir „Jahre mit Rauchen“ als Merkmal nehmen? Müssen wir zwischen starken und schwachen Rauchern unterscheiden usw.

Das sind alles Fragen, auf die man eine methodisch halbwegs saubere Antwort finden kann, aber diese Antwort erfordert einiges Nachdenken und Statistik. Aus diesem Grund ist Epidemiologie auch ein eigenes (Master-) Studium.

F8.5 Fallbeispiel HIV-Infektionen in Deutschland

Nachdem wir das Konzept des relativen Risiko mit fiktiven Daten verstanden haben, möchte ich ein Fallbeispiel mit Echtdaten durchrechnen.

Die Daten

Das Robert-Koch-Institut (RKI) gibt jedes Jahr die Zahl der HIV-Neuinfektionen bekannt. HIV-Infektionen sind in Deutschland meldepflichtig, aber es gibt noch unentdeckte Infektionen. Daher enthält die folgende RKI-Tabelle auch Schätzwerte. Wir wollen mit der mittleren Schätzung arbeiten (den Zahlen, die jeweils an der ersten Stelle stehen)

Wir brauchen die Gesamtzahlen: Zu diesem Zeitpunkt lebten 42 Mio. Frauen und 41 Mio. Männer in Deutschland. Das ist einfach. Schwieriger ist es, die Anzahl homosexueller Männer zu ermitteln. Im Folgenden gehe ich von einer Quote von 5% aus, die den Anteil vermutlich überschätzt.[5] Es soll also 2.050.000 homosexuelle Männer in Deutschland geben.

Längs- vs. Querschnittsdaten

Bevor wir weitermachen, müssen wir noch einmal über die Daten nachdenken. Wir haben das relative Risiko als Kennzahl für Kohortenstudien kennengelernt, also für Daten einer Population, die über einen längeren Zeitraum betrachtet wird. Das bedeutet *Längsschnittdaten*. Das RKI liefert uns aber Daten für 2019, also für *ein Jahr*. Das sind *Querschnittsdaten*. Können wir da mit einem relativen Risiko rechnen? Ja – und zwar weil unsere

[5]https://www.queer.de gibt eine Quote von 5% aller Männer an, die mindestens einmal in ihrem Leben gleichgeschlechtliche Sexualkontakte hatten. Auch hier taucht sofort wieder das Abgrenzungsproblem auf. Zudem ist unklar, wie queer.de zu diesen Zahlen gelangt. Ist das eine Schätzung mit viel Bauchgefühl oder Ergebnis einer Umfrage, bei der man annimmt, dass sich dort *alle* Schwulen „outen“?

Population keine Stichprobe der deutschen Bevölkerung ist, sondern die Grundgesamtheit selbst. Wenn wir nur Daten aus Bayern oder Köln hätten und die auf Deutschland hochrechnen würden, würden wir in Probleme laufen, aber hier haben wir Daten über die Gesamtbevölkerung.

Relatives Risiko der homosexuellen Männer

	A	B	C	D	E
1			**Outcome**		
2			**kein HIV**	**HIV**	Summe
3	**Exposition**	**homosexuell**	2.048.400	1600	2.050.000
4		**heterosexuell**	38.949.679	321	38.950.000
5		Summe	40.998.079	1921	41.000.000
6					
7	durchschnittliches Risiko		0,005%		
8	Risiko der Exponierten		0,078%		
9	Risiko der nicht Exponierten		0,001%		
10	relatives Risiko		94,70		

Wenn wir die Daten in unsere Excel-Tabelle einfüttern, erhalten wir ein relatives Risiko von 94,7. Bis auf Zelle D4 ergeben sich alle Werte direkt aus der RKI-Tabelle. Auf die 321 bin ich über folgende Überlegung gekommen: Wenn 650 Infektionen über heterosexuelle Kontakte entstanden sind und von 83 Mio. Einwohnern 41 Mio. Männer sind, dann entfallen 41/83 der Infektionen auf Männer. Damit ist klar, dass unsere Rechnung sich definitiv nicht auf Rock n‘ Roll bezieht und auch nicht auf Drugs, sondern nur auf Sex.

Schlussfolgerungen

Mit dem Instrument des relativen Risikos, ein paar frei verfügbaren Zahlen und ein paar mäßig problematischen Annahmen haben wir die homosexuellen Männer als Hochrisikogruppe identifiziert. Das ist jetzt nicht überaus überraschend, aber vielleicht ist die Höhe des relativen Risikos höher als Sie vermutet haben.

Welche Schlussfolgerungen aus diesem Ergebnis zu ziehen sind, ist (zumindest mir als HIV-Präventions-Laien) nicht ganz klar. Klar ist, dass es eine eindeutig umrissene Höchstrisikogruppe gibt. Bei der dürfte aber *Information* nicht das Problem sein *(„Achso... Das wusste ich gar nicht... Jetzt werde ich aber...“)* sondern Wahrnehmungsverzerrung oder Wurstigkeit oder noch etwas anderes.

F8.6 Ehemalige Prüfungsaufgaben

Winkelhak´sches Drüsenfieber

Diese grausame Erkrankung ist nicht ansteckend. Erkrankte mutieren spontan zu Zombies (siehe Bild) und verenden innerhalb von Minuten.

Glücklicherweise gibt es seit letztem Jahr eine Impfung, die aber jährlich aufgefrischt werden muß und 12,50€ kostet.

Im letzten Jahr haben sich 10 Mio. Personen in Deutschland impfen lassen. Von diesen 10 Mio. ist nur einer an dem Fieber gestorben. Von den 70 Mio. Nichtgeimpften sind 14 gestorben.

Der Hersteller des Impfstoffs möchte, daß die Krankenkassen die Impfkosten übernehmen, was bisher nicht der Fall ist. Da die Krankheit meldepflichtig ist, sind unsere Daten keine Stichprobe, sondern eine Vollerhebung. In diesem Fall können wir die Daten wie eine Kohortenstudie verwenden. Als Risikofaktor können wir „nicht geimpft“ verwenden und als Outcome „Erkrankung“.

1. Berechnen Sie die überzeugendsten Kennzahlen für eine Pressemitteilung und machen Formulierungsvorschläge.

2. Welche Zahlen sollten in der Pressemitteilung vielleicht besser nicht stehen?

Mammographie-Screening

Brustkrebs ist die häufigste Krebserkrankung bei Frauen. Aus diesem Grund gibt es das Mammographie-Screening, dessen Idee es ist, daß Frauen (ohne familiäre Vorbelastung) ab 50 flächendeckend alle zwei Jahre zur Früherkennung gehen sollen. Berechnen Sie anhand des folgenden Textausschnitts das relative Risiko der Nichtteilnahme am Mammographie-Screening. Überlegen Sie, wie sie mit den 39 Todesfällen umgehen sollen. Interpretieren Sie das Ergebnis.

> *Von 1.000 Frauen im Alter von 50 Jahren sterben binnen zehn Jahren statistisch 39 an anderen Ursachen als Brustkrebs und vier an Mammakarzinomen, sofern sie am Screening teilgenommen haben. Statistisch nur eine Frau mehr, nämlich fünf von 1.000, sterben an Brustkrebs ohne Teilnahme am Screening.*

Winkelhak´sches Drüsenfieber (Lösung)

- Das Sterberisiko der Geimpften (nicht Exponierten) ist 1/10 Mio.
- Das Sterberisiko der Nichtgeimpften (Exponierten) ist 14/70 Mio.
- Das relative Risiko beträgt (14/70.000.000)/(1/10.000.000)=2

Das wäre eine gute Zahl für eine Pressemitteilung

- Würden alle Einwohner geimpft, gäbe es 8 Tote. Derzeit gibt es 15.
- Ein Durchimpfen der Bevölkerung würde also 8 Leben retten und 12,5*80 Mio = 1 Mrd. € kosten bzw. 125 Mio. € je Leben.

Das sind zwei Zahlen, die man vielleicht nicht allzu laut aussprechen sollte, wenn man seinen Impfstoff verkaufen will.

Mammographie-Screening (Lösung)

So würde ich an die Frage herangehen:

Als erstes würde ich darüber nachdenken, ob die 39 Todesfälle, die nicht mit Brustkrebs zusammenhängen, in die Rechnung aufgenommen werden sollen oder nicht. Für beides gibt es Gründe. Die Motivation, am Screening teilzunehmen ist nicht, nicht an Brustkrebs zu sterben, sondern insgesamt möglichst nicht zu sterben. Dann würde man die 39 Frauen in die Rechnung einbeziehen und berechnen, wie stark das Screening die Wahrscheinlichkeit mindert, nicht zu sterben. Auf der anderen Seite will man ja wissen, wie stark das Screening die krankheitsspezifische Mortalität beeinflußt, Das spräche dafür, die 39 Toten nicht in die Rechnung einzubeziehen.

Wenn beide Wege begründbar sind und keiner als eindeutig falsch erkennbar ist, muß ich wohl beide Alternativen berechnen.

Ohne die 39 Toten ist das relative Risiko (5/1000)/(4/1000)=1,25

Mit den 39 Toten ist das relative Risiko (44/1000)/(43/1000)=1,023

Ich denke jetzt über diese beiden Werte nach. Der erste Wert ist deutlich von 1 verschieden, der zweite nahe 1. Was sagt mir das? Der zweite Wert sagt mir, daß die Überlebenswahrscheinlichkeit durch das Screening kaum ansteigt, das Screening also insgesamt wenig bringt.

Woran liegt das?

Es könnte sein, daß das Screening nichts taugt. Dann müßte aber der erste Wert ebenfalls sehr niedrig sein. Ist er aber nicht.

Also bleibt als Interpretation, daß das Screening selbst zwar gut ist, aber ein Risiko testet, das sehr gering ist. Das wird auch deutlich, wenn ich die Absolutzahlen betrachte. Von allen Frauen, die in dem Alter sterben, stirbt nur etwa jede zehnte an Brustkrebs. Wenn ich an dieser Stelle die Vorsorge verbessere, bezieht sich das ja nur auf diese 10%.

Kapitel 9

Fall-Kontroll-Studien

F9.1 Konzept

Der Nachteil von Kohortenstudien ist, daß diese Studien erst nach vielen Jahren zu Ergebnissen führen und sehr teuer sind. Sie ergeben auch nur bei Krankheiten Sinn, die relativ häufig sind.[1]

Während Kohortenstudien mit Gesunden beginnen und warten, bis ein Teil der Gesunden erkrankt ist, gehen Fall-Kontroll-Studien den umgekehrten Weg. Sie untersuchen Personen die schon krank sind (Fälle) und stellen diesen Personen Gesunde gegenüber (Kontrollen) und vergleichen, in welchem Umfang beide Gruppen einem Risikofaktor exponiert waren.

F9.2 Fallbeispiel: Eine weitere Lungenkrebsstudie

In *Einheit Kohortenstudien* haben wir schon eine fiktive Studie über den Zusammenhang zwischen Rauchen und Lungenkrebs betrachtet. Dieses Beispiel möchte ich jetzt wieder aufgreifen, aber eben nicht im methodischen Rahmen der Kohortenstudie, sondern als Fall-Kontroll-Studie.

Unser Alltagswissen sagt uns, dass die meisten Raucher nicht an Lungenkrebs erkranken werden und das es ziemlich lange dauert, bis das, wenn es passiert, passieren wird. Wenn wir das Geld und die Zeit nicht haben, könnten wir eine Fall-Kontroll-Studie entwerfen, die die gleiche Frage untersucht. Nehmen wir an, wir haben die Mittel, um 300 Teilnehmer zu untersuchen und befragen. Von diesen 300 Teilnehmern sollen 100 Personen Lungenkrebs haben, 200 nicht.

Ein Merkmal von Kohortenstudien ist, dass diese Aufteilung zwar nicht vollkommen willkürlich ist, aber nichts mit der Verteilung in der Gesamtbevölkerung zu tun hat. Die Gesamtzahl von 300 könnte auch eine andere sein und die 1/3-Quote von Krebskranken könnte auch 50:50 sein.

Unsere fiktive Studie besteht nun darin, dass wir die Teilnehmer, deren Outcome wir kennen, fragen, ob sie exponiert waren, d.h. geraucht haben oder rauchen.

Die Ergebnisse befinden sich in der folgenden Tabelle.

[1] Bei sehr seltenen Krankheiten ist vorher absehbar, dass es keine (oder zu wenige) Erkrankungen in der Kohorte geben wird, als dass man Schlüsse ziehen könnte.

	A	B	C	D	E	F
1			Outcome			
2			+	-	Σ	
3		+	30	12	42	
4	Exposition	-	70	188	258	
5		Σ	100	200	300	
6						
7	Odds Exponierte			2,500		
8	Odds nicht Exponierte			0,372		
9	Odds Ratio			6,714		
10						
11	*Anteil ("Risiko") Exponierte*					*0,714*
12	*Anteil ("Risiko") nicht Exponierte*					*0,271*
13	*Relatives "Risiko"*					*2,633*

F9.3 Odds Ratio

Der Begriff „odds“

Eine Risikokennzahl, die wir noch nicht kennengelernt haben ist, das Odds Ratio. Der Begriff hat nur indirekt etwas mit „odd“ = langweilig zu tun. Er kann mit „Chancen“ übersetzt werden. Wenn Sie eine Münze haben und bei „Zahl“ gewinnen, könnten Sie (in Risiko-Deutsch) sagen, dass das Risiko zu verlieren bei 50% liegt. Sie könnten aber auch sagen, dass die Wahrscheinlichkeit zu gewinnen bei 1:1 liegt. Die Odds sind die Wahrscheinlichkeit in diesem zweiten Sinn, wobei es in der Epidemiologie im Regelfall nicht um das Gewinnen, sondern um das Verlieren geht.

Die Berechnung der Odds und des Odds Ratio

Das Odds eines Rauchers in unserer Studie ist also odds=(kranke Raucher)/(gesunde Raucher)=30/12. Entsprechend ist das Odds der Nichtraucher 70/188. Das Odds Ratio ist der Quotient aus dem Odds der Exponierten und dem der nicht Exponierten, also

$$\frac{\frac{kranke\ Raucher}{gesunde\ Raucher}}{\frac{kranke\ Nichtraucher}{gesunde\ Nichtraucher}} = \frac{\frac{30}{12}}{\frac{70}{188}} = 6,714$$

Nachteil des Odds Ratio

Das relative Risiko ist intuitiv verständlich. In unserer HIV-Rechnung haben wir ermittelt, dass das Risiko der Homosexuellen 94mal so hoch ist. Das ist zwar abstrakt, aber mit dieser Zahl kann man etwas anfangen. Das Odds Ratio sieht sehr ähnlich aus, aber man kann es nicht als „das soundsovielfache“ interpretieren. Die Aussage dieser Kennzahl ist lediglich „ein OR von 1 bedeutet keinen Einfluss des Faktors“ und „je höher das OR, umso größer der Einfluss“. Als Multiplikator eines Standardrisikos taugt die Kennzahl nicht.

Gründe für die Nichtverwendung des Relativen Risikos

Wenn wir uns die Nachteile des OR anschauen und sie mit dem relativen Risiko vergleichen, drängt sich die Frage auf, warum man diese schwer interpretierbare Kennzahl

verwendet, wo es doch mit dem relativen Risiko eine viel eingängigere Alternative gibt. Der Grund ist, dass das relative Risiko bei Fall-Kontroll-Studien nicht funktioniert.

Wenn Sie zurückblättern, sehen Sie, dass ich das RR für die Raucher-Fall-Kontroll-Studie einfach einmal berechnet habe, den Begriff Risiko aber in Gänsefüßchen gestellt habe, damit klar ist, dass irgendwas in dieser Rechnung nicht stimmt.

Mathematisch stimmt das Ergebnis von 2,66. Das ist nicht das Problem. Das Problem ist die Auswahl der Teilnehmer. In einer *Kohortenstudie* wählen wir eine repräsentative Kohorte aus und beobachten, was diese Kohorte im Laufe der Jahre so treibt und ob das Verhalten zu unterschiedlichen Ergebnissen führt. In der *Fall-Kontroll-Studie* wissen wir ja schon, was passiert, d.h. ob ein Mensch Krebs bekommen wird oder nicht. Das ist ja genau das Auswahlkriterium. Wir wollten 100 Krebskranke und 200 Gesunde. Wie auch immer wir auf dieses Verhältnis von 100:200 gekommen sind.

Spielen wir einmal folgende Geschichte durch: Kurz bevor wir unsere Fall-Kontroll-Studie starten, tut sich noch eine Drittmittelquelle auf. Wir können jetzt 400 statt 200 Kontrollen einbeziehen. Ich nehme einmal an, dass sich bei einer Verdoppelung der Kontrollen auch die Zahlen der Exponierten und nicht Exponierten verdoppeln. Die Zahlen sehen dann so aus:

	A	B	C	D	E	F
1			Outcome			
2			+	-	Σ	
3		+	30	24	54	
4	Exposition	-	70	376	446	
5		Σ	100	400	500	
6						
7	Odds Exponierte			1,250		
8	Odds nicht Exponierte			0,186		
9	Odds Ratio			6,714		
10						
11	*Anteil ("Risiko") Exponierte*					*0,556*
12	*Anteil ("Risiko") nicht Exponierte*					*0,157*
13	*Relatives "Risiko"*					*3,540*

Der Unterschied ist, dass die Werte in den Zellen D3-D5 verdoppelt sind. In der kleinen Studie war das Relative Risiko 2,633. Hier in der großen Studie ist es jetzt 3,540. Dann ist diese Kennzahl ja in diesem Kontext Blödsinn, weil sie vom Verhältnis von Fällen zu Kontrollen abhängt. Wenn wir die Drittmittel dazu verwenden würden, 10 zusätzliche Fälle in die Studie aufzunehmen, kommt noch eine andere Zahl heraus. Sie können dieses Spiel beliebig permutieren. Für jedes unterschiedliche Fall-Kontroll-Verhältnis kommt ein anderes relatives Risiko heraus.

Weil es bei diesem Studientyp aber kein „richtiges“ Verhältnis zwischen Fällen und Kontrollen gibt, ist die Berechnung des relativen Risikos sinnlos.

Gründe für die Verwendung des Odds Ratio

Werfen Sie noch einmal einen Blick auf die Varianten der Studie. Das relative Risiko war immer ein anderes. Deswegen können wir es nicht verwenden. Wenn Sie aber das Odds Ratio vergleichen, sehen Sie, dass es in jeder Variante immer den gleichen Wert von 6,714 hat. Das bedeutet, dass das Odds Ratio nicht sensibel auf das Verhältnis von Fällen zu Kontrollen reagiert. Und diese Eigenschaft brauchen wir bei diesem Studientyp.

Wieso ist das so? Schauen wir uns dazu noch einmal die Definition des Odds Ratio an. $\frac{\frac{kranke\,Raucher}{gesunde\,Raucher}}{\frac{kranke\,Nichtraucher}{gesunde\,Nichtraucher}} = \frac{\frac{30}{12}}{\frac{70}{188}} = 6,714 = OR$

Die Mittelstufenmathe-Regel ist „Brüche dividiert man, indem man den Zähler mit dem Kehrwert des Nenners multipliziert. Dann ist

$OR = \frac{kranke\,Raucher}{gesunde\,Raucher} * \frac{gesunde\,Nichtraucher}{kranke\,Nichtraucher}$

Wenn wir jetzt z.B. die Zahl der Kontrollen (Nichterkrankte) verdoppeln, dann müsste sich (wenn wir alles richtig machen) die Zahl der gesunden Raucher und die der gesunden Nichtraucher auch verdoppeln. Für unsere Rechnung bedeutet das:

$OR = \frac{kranke\,Raucher}{2*gesunde\,Raucher} * \frac{2*gesunde\,Nichtraucher}{kranke\,Nichtraucher}$

Die „2“ kürzt sich also weg. Das gleiche gilt natürlich auch für eine Änderung der Zahl der Fälle. Durch diesen Mathe-Kniff hat man also eine robuste Zahl bekommen, die nur eben nicht so intuitiv ist wie das Relative Risiko.

Kapitel 10

Standardisierung

F10.1 Ausgangsproblem

In *Einheit Kohortenstudien* habe ich kurz über die Framingham-Studie berichtet. Ziel der Studie war es, Risikofaktoren für koronare Herzkrankheiten zu identifizieren. Nun könnte man denken *„zieh die Studie durch und dann weißt du es“*. Wenn man aber einschlägige Fachzeitschriften durchblättert, stößt man ständig auf Studien, die dieser Framingham-Studie ähneln. Warum macht man das?

Ein Grund sind andere Definitionen. Wir haben gesehen, wie kniffelig es ist, zu definieren, wer „Raucher“ ist. Wenn man eine andere Definition verwendet, was ja legitim sein kann, wird vermutlich zu anderen Ergebnissen kommen. Vielleicht ändern sich die Rauchgewohnheiten, indem Zigaretten weniger Schadstoffe enthalten. Oder Raucher von Zigaretten auf Verdampfer umsteigen, weil die vielleicht weniger schädlich sind. Oder vielleicht hat man auch nur die Vermutung, dass eine US-Studie nur mäßig gut auf deutsche Verhältnisse passt und will das mit deutschen Daten überprüfen.

Wir haben also einen Haufen von Studien, deren Ergebnisse wir miteinander vergleichen wollen. Dieses Vergleichen kann aber, selbst wenn wir keine Definitionsprobleme wie *„wer ist ein Raucher“* haben, schwierig sein.

F10.2 Zwei fiktive Fallbeispiele

Maisbrei und Alzheimer

Unsere Hypothese ist, dass es für das Auftreten von Alzheimer relevant ist, aus welcher Quelle die Kohlehydrate stammen. Dazu haben wir eine Studie in zwei Ländern durchgeführt: Deutschland und Malawi. Wir stellen fest, dass der Anteil der Menschen mit Alzheimer in Deutschland, bei dem die Kohlehydrate primär aus Weizen und Kartoffeln stammen, bei 2% liegt (echte Zahl) und in Malawi bei 0,1% (erfundene Zahl, die vermutlich nicht allzu falsch ist). In Malawi nehmen die Menschen Kohlehydrate in Form des Nationalgerichts Nsima (eine Form von Maisbrei) zu sich. Also sinkt offensichtlich das Risiko, an Alzheimer zu er-kranken, wenn man mehr Maisbrei isst.

Vermutlich ist Ihnen schon klar, dass da noch irgendwas fehlt. Bevor wir also eine *„Esst mehr Maisbrei“*- Kampagne starten und *„Weizen tötet“* auf jede Mehlpackung drucken lassen, denken wir noch einmal nach. Das Beispiel habe ich absichtlich dämlich gewählt, weil der Fehler offensichtlich wird, sobald Sie „Malawi“ in Google eingeben. Malawi ist eines der ärmsten Länder der Welt. Das Pro-Kopf-BIP liegt bei etwa $400. Pro Jahr. Da überrascht

es uns nicht, dass die Lebenserwartung bei etwa 63 Jahren liegt. Unsere epidemiologische Allgemeinbildung reicht aus, dass uns klar ist, dass Alzheimer eine Krankheit ist, die typischerweise in höherem Alter auftritt. Wir wissen seit *Einheit Lebenserwartung* auch, dass es auch in Malawi Menschen gibt, die über 63 Jahre alt sind, aber vermutlich nicht so viele wie in Deutschland. Wenn die Altersstruktur für die Ergebnisse relevant sind, sich die Altersstruktur zweier Studien unterscheiden, dann müssen wir versuchen, diese Altersstruktur irgendwie aus den Daten herauszurechnen.

Vielleicht ist Maisbrei wirklich eine gute Idee (ich glaube nicht, ich hab mir das ja einfach ausgedacht), aber die 0,1% Alzheimerkranke im Vergleich zu 2% in den „Rohdaten" belegen das nicht.

Reis und Gebärmutterhalskrebs

Inspiriert von der ersten Studie stellen wir folgende Hypothese auf: Das Essen von Reis verringert die Wahrscheinlichkeit, an Gebärmutterhalskrebs zu erkranken. Für den Vergleich wählen wir Deutschland und Südkorea aus. Erst wollten wir China nehmen, aber dann wären wir in eine Variante des Malawi-Problems gelaufen. Also Südkorea. Tatsächlich (in meiner erfundenen Studie) stellen wir fest, dass es in Südkorea viel weniger Fälle pro Mio. Einwohner gibt als in Deutschland. Also *„mehr Reis essen"*.

Das Problem ist versteckter, weil sie vermutlich nicht wissen, dass es in Südkorea eine sehr liberale Gesetzgebung bei Abtreibungen gibt und die Koreaner in den letzten Jahrzehnten eine spürbare Präferenz für Jungen hatten. In den 1990er Jahren gab es bei etwa 650.000 Geburten 350.000 Jungen und 300.000 Mädchen. Frauen sind also in Südkorea spürbar unterrepräsentiert. Unsere Zahlen sind also auch unbrauchbar, weil sie die gleiche Geschlechtsaufteilung unterstellen, die aber nicht vorliegt. Um die Zahlen zu vergleichen, müssen wir also die unterschiedliche Geschlechtsstruktur der Länder herausrechnen. Das ist alles nicht so wahnsinnig schwer, aber wir müssen es tun.

F10.3 Fallbeispiel: Corona in Sachsen

Eine Vergleichstabelle

Ein Thema, das um die Jahreswende 2020/21 in den Medien starke Aufmerksamkeit erregte, war die hohe Zahl von Coronainfektionen in Sachsen. Das Robert-Koch-Institut veröffentliche die Fall- und Todeszahlen täglich im Internet. Für dieses Fallbeispiel habe ich mir die zu der Zeit (13.1.2021) aktuellen Zahlen besorgt.[1]

In das größte Fettnäpfchen sind wir mit dieser Falle nicht getappt, weil wir die Absolutzahlen nicht miteinander verglichen haben. Dann wäre das Saarland eine Insel der Seligen, weil es dort wenig Fälle gibt. Aber uns ist klar, wie wenig Leute dort wohnen. Also ist es sinnvoll, die Zahlen auf die Bevölkerung zu beziehen. Hier habe ich das in Fälle je 1 Mio. Einwohner getan. Wenn man diese Rechnung für alle Bundesländer durchführt, ist Sachsen der traurige Spitzenreiter.

Die Hypothesenmaschine in den Medien ist sofort angesprungen und hat diverse Erklärungen angeboten:

[1] Falls Sie die Zahlen nachrechnen wollen: Die Angaben des RKI unterscheiden sich von Tabelle zu Tabelle jeweils um ein paar Fälle. Meine Vermutung: Die Zahlen sind nicht zum gleichen Zeitpunkt aktualisiert worden. Für die Methodik, um die es geht, und das Ergebnis sind diese kleinen Abweichungen aber irrelevant.

		Fälle	
	Einwohner	in den letzten 7 Tagen	je 1 Mio. Einwohner
Deutschland	83.020.000	142.827	1720,4
Sachsen	4.010.000	12.395	3091,0
Saarland	990.500	1.741	1757,7

Tabelle 10.1: Corona-Fälle in Sachsen

- Es gibt viele Nazis und Querdenker in Sachsen, die Corona für eine Verschwörung halten und deswegen keine Vorsichtsmaßnahmen treffen.
- Der Sachse an sich ist stur und tut grundsätzlich nicht, was ihm „die da oben“ sagen.
- usw.

Die Altersstruktur

Diese Erklärungen mögen alle zutreffen oder auch nicht. Der misstrauisch epidemiologisch Vorbelastete vermutet aber irgendeinen relevanten Hintergrundfaktor wie den aus unseren beiden fiktiven Fallbeispielen, der das Ergebnis verfälscht.

Mein Hintergrundwissen ist, dass im Laufe der Wiedervereinigung viele junge Menschen aus den Ost-Bundesländern in den Westen gezogen sind, so dass die Altersstruktur in diesen Ländern anders ist als im Westen. Jetzt weiß ich auch noch, dass das Infektionsrisiko für Kinder niedriger ist. Wenn es also (überspitzt gesagt) in Sachsen nur Rentner und keine Kinder gibt, wäre es nicht so überraschend, wenn dort die Infektionszahlen höher sind. Also braucht man die Altersstruktur von Sachsen im Vergleich zu Deutschland. Die kann man im Internet finden.

Der obere (blaue) Balken ist der Jungendquotient, der untere (rote) Balken der Altenquotient. Für Sachsen bedeutet dass, das auf 100 Personen im Alter von 20-64 28,2 Personen unter 20 Jahren und 43,0 Personen im Alter von 65 und älter entfallen, wärend dieses Verhältnis in Gesamtdeutschland bei 30,3 : 100 : 34,7 liegt. Diese Grafik liefert uns eine Altersstruktur in drei Klassen. Das ist nicht allzu exakt, aber wir wollen hier in erster Linie das Instrument verstehen und nicht Nachkommastellen quälen.

Mit den Jugend- und Altersquotienten kann man die Anteile der drei Gruppen an der Bevölkerung berechnen:

	B	C	D	E	F
1		Sachsen		Deutschland	
2		Quotienten	Anteile	Quotienten	Anteile
3	unter 20	28,2	16,5%	30,3	18,4%
4	20-64	100,0	58,4%	100,0	60,6%
5	65 und älter	43,0	25,1%	34,7	21,0%
6	Summe	171,2	100,0%	165,0	100,0%

Wenn in Sachsen auf 100 Personen im Alter zwischen 20 und 64 28,2 Personen unter 20 Jahre und 43 Personen im Alter von 65 oder höher, dann sind das insgesamt 171,2 Personen und die Anteile, die in Spalte D stehen. Für Deutschland ist die Rechnung analog aufgebaut. Es wohnen zwar nicht nur Rentner in Sachsen, aber die Altersstruktur ist schon unterschiedlich. Das könnte zumindest einen Teil der Mehr-Infektionen erklären.

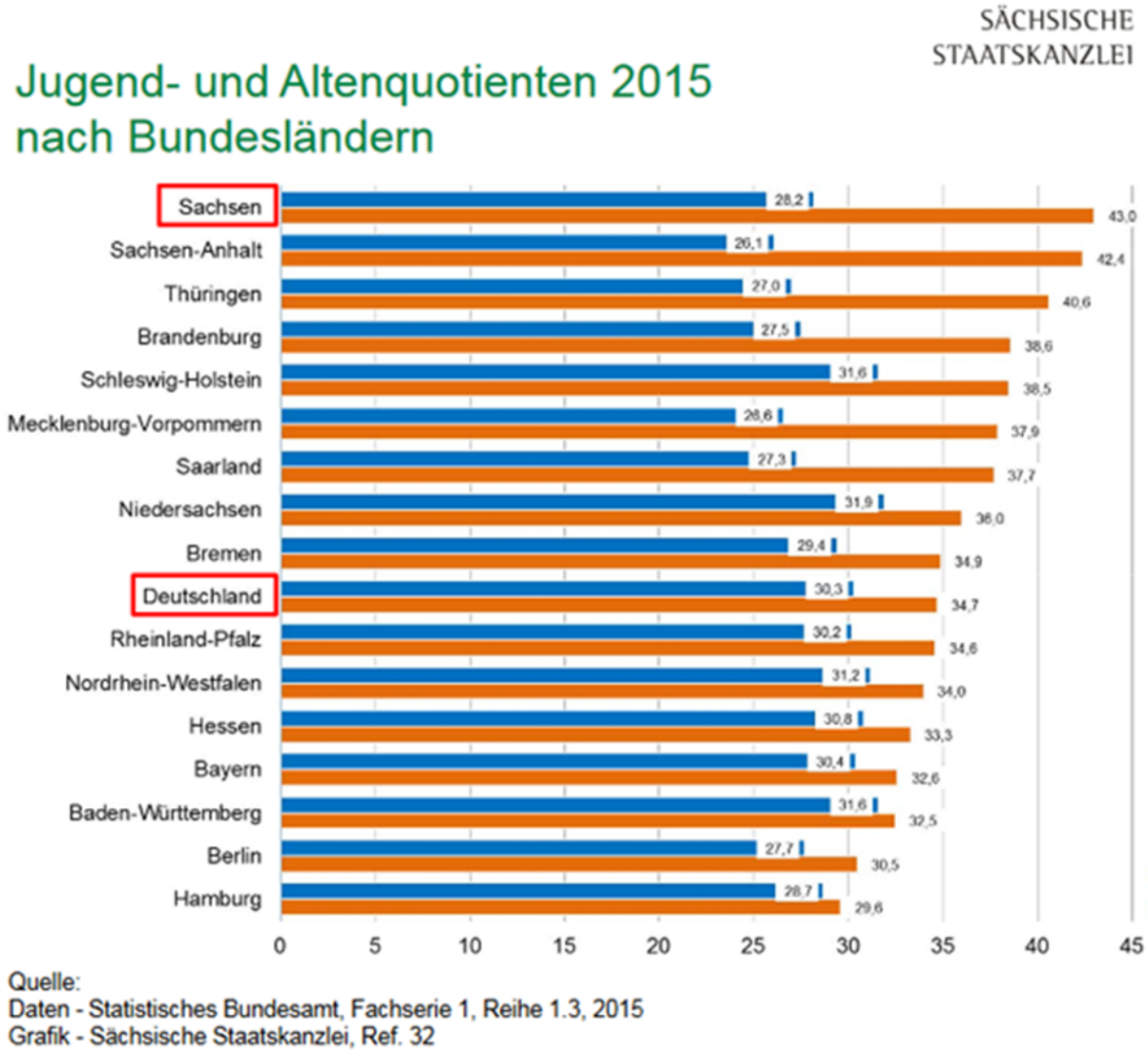


Abbildung 10.1: Altersstruktur nach Bundesländern

Die Inzidenzraten

Das Robert-Koch-Institut (RKI) gibt wöchentlich die Zahl der neu Corona-Infizierten nach Altersklassen differenziert heraus. Der epidemiologische Begriff für „neue Fälle“ ist *Inzidenz*, der für „die Fälle, die es schon gibt“ *Prävalenz*. Die Aufschlüsselung nach Alter ist differenzierter als unsere Drei-Klassen-Gesellschaft. Daher habe ich die Fallzahlen entsprechend eingeordnet. Das Ergebnis ist sieht so aus.[2]

	A	B	C	D	E
8		Deutschland			
9		Anteile			Anteile
10		Bevölkerung	Personen	neue Fälle	Infizierte
11	unter 20	18,4%	15.245.491	13.752	0,090%
12	20-64	60,6%	50.315.152	94.775	0,188%
13	65 und älter	21,0%	17.459.358	34.300	0,196%
14	Summe	100,0%	83.020.000	142.827	0,172%
15	Pro Mio.		1.721		

Die Zahlen in der Spalte B habe ich aus der sächsischen Alterststruktur-Grafik für Gesamtdeutschland, die Zahlen in Spalte C ergeben sich aus der Gesamtbevölkerung von 83.020.000. Spalte D gibt die 7-Tage-Inzidenz des RKI in den drei Altersgruppen an und Spalte E ist der Anteil der Infizierten in jeder der drei Gruppen.

Die magische Kennzahl steht in Zelle C15: Die Inzidenz bezogen auf 1 Mio. Einwohner.

Eine Erkenntnis, die wir quasi als Beifang erhalten ist, dass die Inzidenzrate bei den Jungen mit 0,09% tatsächlich niedriger ist als bei den anderen Altersgruppen. Allerdings ist diese Rate bei den Alten nicht dramatisch viel höher als bei den Mittelalten.

Direkte Standardisierung

Wir wissen, dass die Altersstruktur in Sachsen „ungünstig“ (im Sinne von „mehr Erkrankungen“ ist. Wir könnten uns nun fragen, wie viele Erkrankungen es in Sachsen gäbe, wenn Sachsen die gleichen altersspezifischen Inzidenzraten hätte. Wir wissen, dass deutschlandweit 0,196% der Alten sich in den letzten 7 Tagen infiziert haben. Für Sachsen kennen wir diesen Wert nicht.[3] Wir könnten aber einfach durchrechnen, was herauskäme, wenn dieser Wert in Sachsen genauso hoch wäre. Wir wissen jetzt schon, dass Sachsen viele Erkrankungen haben wird, weil es viele Alte gibt, aber wie viele Mehrerkrankungen das sind, kann man schwer schätzen.

	A	B	C	D	E
17		Sachsen mit der gesamtdeutschen Infektionsstruktur			
18		Anteile			Anteile
19		Bevölkerung	Personen	neue Fälle	Infizierte
20	unter 20	16,5%	660.526	596	0,090%
21	20-64	58,4%	2.342.290	4.412	0,188%
22	65 und älter	25,1%	1.007.185	1.979	0,196%
23	Summe	100,0%	4.010.000	6.986	0,174%
24	Pro Mio		1.742		

Diese Tabelle ist eine Mischung aus der „echten“ Bevölkerungsstruktur Sachsens (4,01 Mio Einwohner, davon 25,1% Alte usw.) und den Inzidenzzahlen Gesamtdeutschlands

[2] Wenn Sie genau hinschauen, ist die Gesamtzahl der Neuinfektionen eine etwas andere als in der ersten Tabelle, da dies die Zahlen der ersten Kalenderwoche 2021 sind und die erste Tabelle die 7-Tage-Inzidenz vom 13.1. auflistet.

[3] Ich habe ihn jedenfalls nicht gefunden, aber auch nicht stundenlang gesucht.

(0,196% der Alten infizieren sich usw.) Die blauen Zahlen in der Spalte D sind also fiktive Zahlen. Wenn sich von den „echten“ 1,007 Mio. Rentnern in Sachsen 0,196% infizieren würden (so wie in ganz Deutschland) dann wären das 1.979 infizierte Rentner.

Die Rechnung für die beiden anderen Gruppen ist analog.

Nach dieser Mischung gesamtdeutscher Indzidenzraten und sächsischer Bevölkerung hätte Sachsen 6.986 Infizierte haben müssen. Bezogen auf eine Millionen Einwohner wären das 1.742 Infizierte gewesen. Das sind 21 mehr als im Bundesdurchschnitt.

Diese 21 Mehrinfizierten sind also die Infizierten, die auf die „schlechte“ Altersstruktur zurückzuführen sind. Mehr nicht. Tatsächlich waren es aber über 3.000 Infizierte. Das bedeutet also, dass der Unterschied in den Infiziertenzahlen nur zu einem sehr kleinen Anteil auf die Altersstruktur zurückzuführen sind und im Wesentlich anderen Gründe haben muß. Ein paar Kandidaten habe ich ja schon aufgelistet.

Dieses Verfahren der Übertragung der Struktur einer Population auf eine andere nennt man direkte Standardisierung. Es gibt auch eine indirekte Standardisierung. Die schauen wir uns aber hier nicht an.

Multiple Standardisierung

Wenn es mehrere relevante Faktoren gibt, in denen sich die Struktur der Teilnehmer in den Studien unterscheiden (könnten), muß man diese Faktoren in die Standardisierung mitein-beziehen. Wenn das Geschlecht noch eine Rol-le spielen könnte, müsste man nicht nur Alters- sondern Alters/Geschlechts-Gruppen bilden und die Anteilswerte für diese Gruppen berechnen.

F10.4 Ehemalige Prüfungsaufgaben

Standardisierung nach Tierkreiszeichen

Im Auftrag des Bundesgesundheitsministeriums sollen Sie eine Studie zum Thema Übergewicht entwickeln. Der Abteilungsleiter im Ministerium ist esoterisch angehaucht. Er ist fest davon überzeugt, daß das Tierkreiszeichen einer Person sein Gewicht beeinflußt. (Der Zwilling neigt zum Hüftgold) Dieser Faktor muss in der Studie unbedingt berücksichtigt werden. Sie halten die Tierkreiszeichen-Idee für Blödsinn und wollen sich nicht blamieren, indem Sie mit Ihrem Namen für tierkreiszeichenstandardisierte Studien einstehen sollen. Das Tierkreiszeichen müssen Sie bei der Untersuchung nicht extra erheben, weil es über das Geburtsdatum, das ohnehin erhoben wird, ermittelbar ist. Um den Kunden zufriedenzustellen, werden Sie nach der Datenerhebung eine gesonderte Auswertung durchführen, in der auch nach dem Tierkreiszeichen standardisiert wird. Diese Auswertung wollen Sie mit dem Auftraggeber durchsprechen, dann allerdings nicht veröffentlichen.

Ihre Erwartung ist, daß das Tierkreiszeichen irrelevant ist und daß Sie dies dem Auftraggeber vermitteln können Woran erkennen Sie an Ihren Auswertungen die Irrelevanz? Gleiche Frage, anders herum: Es könnte theoretisch sein, daß hier ein Einfluss besteht und der esoterische Abteilungsleiter überraschenderweise Recht hat. Wenn das so wäre, wären Sie auf einen Schlag berühmt. Das wollen Sie sich natürlich auch nicht entgehen lassen. Wie würden Sie diesen Zusammenhang an Ihren Auswertungen erkennen?

Standardisierung nach Tierkreiszeichen (Lösung)

Wenn das Tierkreiszeichen für das Übergewicht keine Rolle spielt, dann müßte der Anteil der Übergewichtigen in allen Tierkreiszeichen gleich hoch sein. Wenn Sie dann standardisieren, sind die nichtstandardisierten Daten mit den standardisierten identisch. Sollte das nicht der Fall sein, scheinen Sie auf einen Zusammenhang gestoßen zu sein.

Kapitel 11

Konfidenzintervalle

F11.1 Das Konzept der Stichprobe

Massenproduktion

Im Bachelor haben Sie das Konzept der Massenproduktion kennengelernt. Ein wichtiger Bestandteil dieses Konzepts ist, dass die Bauteile so präzise hergestellt werden konnten, dass sie ohne weitere Bearbeitung zusammengebaut werden konnten. Nehmen wir an, es geht um eine Schraube.[1]

Diese Schraube muss mindestens 19mm lang sein und höchstens 21mm. Außerhalb dieser Fertigungstoleranzen wird es Probleme geben.

Nun war eine Erfahrung, die die Ingenieure sehr früh gemacht haben die, dass auch stabile Produktionsprozesse ein klein wenig „aus dem Ruder laufen“ konnten, d.h. dass irgendwann die Schrauben zu lang oder zu kurz abgesägt wurden. Das hat natürlich kein Mensch gemacht, sondern eine Maschine, die anfänglich auf 20 mm eingestellt war, aber irgendwann die Schrauben zu kurz oder zu lang abgesägt hat. Die Frage war nun, wie man in einem solchen Produktionsprozess möglichst früh merken kann, dass der Prozess beginnt, aus dem Ruder zu laufen. Man könnte jedes einzelne Teil prüfen. Das ist aber teuer.

Die Alternative ist, eine Stichprobe aus der Tagesproduktion ziehen und prüfen. Man hat dann, sagen wir, 1% der Produktion geprüft und versucht aus dieser Stichprobe auf die 99% rückzuschließen, die man *nicht* untersucht hat. Das statistische Instrumentarium das man hier einsetzt, heißt *Stichprobentheorie* und die frühe Zeit der Massenproduktion vor gut 100 Jahren hat die Stichprobentheorie stark vorangetrieben.

Gesundheitsökonomie

In der Gesundheitsökonomie wird ebenfalls mit Stichproben gearbeitet. Die epidemiologischen Studientypen sind ja durchgängig Stichproben. Neue Medikamente werden auch an Stichproben getestet, bevor sie auf den Markt kommen. Das bedeutet, dass auch in

[1] Quelle: Wikipedia

unserem Bereich Stichproben allgegenwärtig sind. Überall will man von der Stichprobe auf die Grundgesamtheit zurückschließen. Aber man hat halt nur eine Stichprobe.

F11.2 Die Belastbarkeit von Stichprobenergebnissen

Repräsentativität

Anfang 2021 waren 614.000 Menschen gegen Corona geimpft worden. Davon sind 7 im Zusammenhang mit der Impfung gestorben.[2] Diese 614.000 könnte man als Stichprobe für die Gesamtbevölkerung ansehen. Bei einer Mortalitätsrate[3] von 1:87.714 und dem Ziel, 60% der Bevölkerung, d.h. 50 Mio. Menschen zu impfen, würde man mit 570 Impftoten rechnen.

Das Problem ist nun aber, dass die 614.000 Geimpften überhaupt nicht repräsentativ für die Gesamtbevölkerung sind, weil im Wesentlichen die Hochbetagten zuerst geimpft werden. Daher ist es nicht überraschend, dass der Jüngste der 7 Verstorbenen 79 Jahre alt war. Die 614.000 taugen also als Stichprobe für die Gesamtbevölkerung überhaupt nichts, bzw. nur als Stichprobe für die Hochbetagten, die noch nicht geimpft worden sind.

Der Stichprobenfehler

Ich habe mir gerade einen Spielwürfel zur Hand genommen. Wir wissen, dass bei einem nicht gezinkten Würfel jede Punktzahl gleich häufig sein muß. Ich würfele jetzt sechsmal. Das (nicht geschummelte) Ergebnis ist:

$4, 2, 6, 3, 1, 4.$

Die 4 ist zweimal gewürfelt worden, die 5 gar nicht. Also ist der Würfel gezinkt. Oder? Aus Ihrer eigenen Mensch-Ärgere-Dich-nicht-Erfahrung sagt Ihnen Ihre Intuition, dass es sogar ziemlich unwahrscheinlich ist, dass man bei sechs Würfen alle Zahlen einmal würfelt. Also gilt diese Regel bei sechs Würfen noch nicht so richtig. Wenn ich aber 600 mal würfele und noch immer keine 5 gewürfelt habe, würden ihnen Zweifel kommen, ob der Würfel nicht doch manipuliert worden ist. [4]

Das *Gesetz der großen Zahl* bedeutet, dass die Häufigkeitsverteilung einer Stichprobe sich der der Grundgesamtheit umso mehr annähert, je größer die Stichprobe ist. Das ist letztlich nur eine Umformulierung des Stichprobenfehlers.

Gesundheitsökonomische Relevanz

Das Gesetz der großen Zahl bedeutet für uns, dass eine Stichprobe umso exaktere Rückschlüsse auf die Grundgesamtheit ermöglicht, je größer sie ist. Das klingt nach einer Binsenweisheit und ist vermutlich auch eine.

Es gilt dann aber auch der Umkehrschluss, dass eine Stichprobe umso unzuverlässiger ist, je kleiner sie ist. Das trifft uns hart, weil unsere Studien im Regelfall nur ein paar hundert Probanden umfasst, wir aber wissen wollen, was in der Gesamtbevölkerung passieren würde, wenn wir diese Intervention zum Standard machen würden.

---

[2] Geinitz, Christian, "Mögliche Nebenwirkungen des Impfstoffs", FAZ 16.1. 2021.

[3] Geimpfte je Todesfall

[4] Die Wahrscheinlichkeit, dass bei 6 Würfen keine 5 gewürfelt wird, ist $(5/6)^6 = 33,5$ also gar nicht so unwahrscheinlich. Bei 600 Würfen ist die Wahrscheinlichkeit 0,00000000000000000000000000000000000000003%

Die gesundheitsökonomische Fragestellung ist also eine etwas andere als die bei der Schraubenherstellung. Uns geht es darum, dass medizinische Maßnahmen nicht bei allen Patienten das gleiche Ergebnis bringen. Wir haben also keinen standardisierten Produktionsprozess, sondern eine standardisierte Maßnahme.

Ein Medikament wirkt bei einigen Patienten gar nicht, bei anderen schwach, mittel oder gut. Das wissen wir aber erst, wenn wir es an diesem Patienten ausprobiert haben. Das können wir aber nicht, weil es teuer ist und das wollen wir auch nicht, weil wir ja dann regelmäßig Millionen von Patienten mit einem Wirkstoff behandeln, von dem wir dann irgendwann feststellen *„ist doch nicht so doll. Das alte war besser“.*

Um noch einmal das Würfel-Beispiel zu verwenden: Wir haben nicht 600 Würfe, sondern nur 6. Mehr können wir nicht bezahlen. Trotzdem sollen wir uns dann ein Urteil über den Würfel bilden. So gut das eben mit sechs Würfen möglich ist.

F11.3 Fiktives Fallbeispiel: Superprolol

Als einfaches Beispiel habe ich mir einen neuen Blutdrucksenker ausgedacht. Ein (real existierender) Wirkstoff, der sehr häufig eingesetzt wird, ist *Metoprolol*. Also soll mein fiktiver neuer Blutdrucksenker, der eine Verbesserung dieses Wirkstoffs darstellt, *Superprolol* heißen.

Ergebnisse der Ministudie

Die Wirksamkeit eines Blutdrucksenkers wird in „Millimeter Quecksilbersäule“ (mm/Hg) gemessen. Um die Wirksamkeit von Superprolol zu messen, haben wir eine Studie mit 6 Probanden durchgeführt.

Sie ahnen es: Ich nehme einfach die Zahlen, die ich eben gerade gewürfelt habe als Ausmaß der Senkung, also

$4, 2, 6, 3, 1, 4$.

Wir können jetzt den Mittelwert berechnen. Das ist 3,33. Uns ist aber klar, dass diese Zahl extrem wackelig ist. Wenn wir noch einmal eine Studie mit sechs anderen Teilnehmern durchführen würden, wäre es sehr überraschend, wenn dann das gleiche Ergebnis herauskommt. Ich würfele also noch einmal:

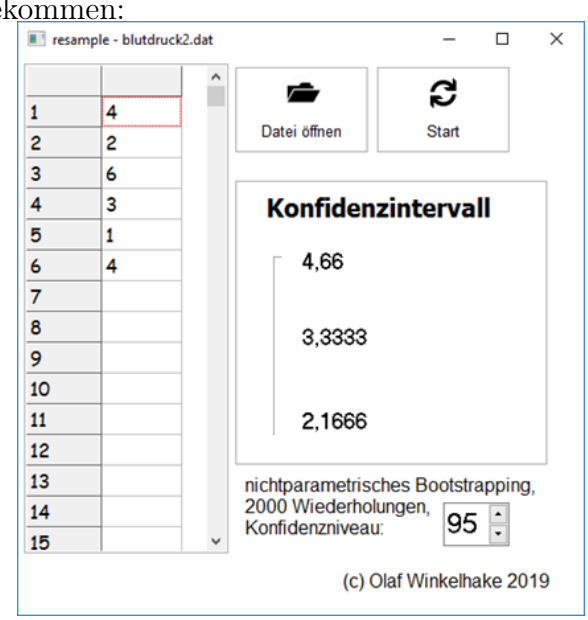
$5, 1, 2, 4, 4, 1$

Der Mittelwert dieser Studie wäre 2,83. Also sollten wir mit der 3,33 nicht an die Öffentlichkeit gehen. Mit der 2,83 auch nicht. Beide Zahlen sind unzuverlässig.

Weil ich einen Spielwürfel genommen habe, liegt die Idee nahe, gar nicht zu würfeln, sondern gleich den Erwartungswert von 3,5 zu verwenden. Das können wir genau dann machen, wenn wir die Wahrscheinlichkeitsverteilung des Würfeln kennen. Dann kann ich es mir auch sparen, viele tausend Mal zu würfeln, weil ich schon vorher weiß, was rauskommen müsste. Das dumme ist nun, dass ich an dieser Stelle etwas geschummelt habe, weil wir in den Studien (um in der Metapher zu bleiben) mit einem vermutlich gezinkten Würfel würfeln, aber nicht wissen, *wie* er gezinkt ist. Das bekommen wir nur durch Ausprobieren heraus und weil wir nicht tagelang würfen wollen, würfeln wir eben nur ein paar Mal und fangen dann an, herumzurechnen.

Das Konfidenzintervall

Man kann die (Un)zuverlässigkeit der 3,33 einer Stichprobe quantifizieren. Das statistische Instrument dazu wird Konfidenzintervall genannt. Ich möchte erst einmal die Ergebnisse zeigen, dann diese Ergebnisse interpretieren und erst zum Schluss zeigen, wie es berechnet wird. Damit Sie mit diesem Instrument herumspielen können, habe ich eine kleine Software für Windows geschrieben, die Sie von der Homepage herunterladen können. In diese Software habe ich unsere 6 Werte eingelesen, Start gedrückt und folgendes Ergebnis bekommen:



Der Mittelwert ist 3,33. Das wussten wir schon. Durch großen statistischen Zauber hat das Programm aber zwei Werte (2,166 und 4,66) berechnet, die die Enden des Konfidenzintervalls darstellen.

Wir wissen, dass unser Ergebnis von 3,33 nicht belastbar ist, weil es nur auf einer (sehr kleinen) Stichprobe beruht. Das Konfidenzintervall sagt uns nun, dass, wenn wir den Blutdrucksenker flächendeckend einsetzen, die durchschnittliche Wirkung irgendwo zwischen 2,166 und 4,66 liegen wird. Das ist ein ziemlich breites Konfidenzintervall, aber wir haben ja auch nur 6 Probanden gehabt.

Konfidenzniveau

Können wir wirklich sicher sein, dass die durchschnittliche Wirklung zwischen 2,166 und 4,66 liegen wird? Leider nicht. In dem Screenshot wird ein Konfidenzniveau angezeigt, das in der Berechnung bei 95% liegt. Das bedeutet, dass, wenn wir Pech haben, die durchschnittliche Wirkung außerhalb des Intervalls liegen wird.

Wir können „Pech“ auch quantifizieren. Hier ist es 5%. Man nennt diesen Wert auch *Irrtumswahrscheinlichkeit*. Die Gegenwahrscheinlichkeit ist das Konfidenzniveau. In der Statistik hat sich ein Konfidenzniveau von 95% als Standard etabliert. Wir werden noch sehen, in welchen Fällen es attraktiv ist, von diesem Niveau abzuweichen.

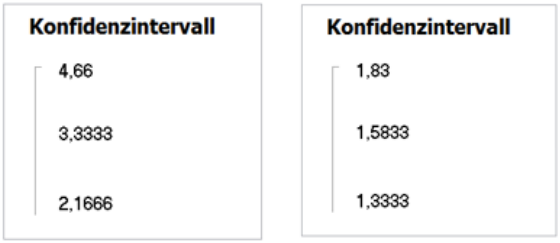
Eine beliebte Fehlinterpretation des Konfidenznivaus ist: *95% der Patienten werden innerhalb des Intervalls liegen.* Das ist falsch. Wo die einzelnen Patienten liegen, ist völlig unklar. Es geht nur um den Mittelwert aller Patienten und der liegt mit einer Wahrschein-

lichkeit von 95% innerhalb des Intervalls.[5]

F11.4 Fiktives Fallbeispiel: Vergleich mit etablierten Wirkstoffen

Statistische Signifikanz

Für die blutdrucksenkenden Wirkstoffe, die schon auf dem Markt sind, gibt es natürlich auch Studien und Konfidenzintervalle. Die folgenden Grafiken vergleichen den neuen Wirkstoff Superprolol (links) mit Looserprolol (rechts), bei dem das Patent schon lange ausgelaufen ist.



Sie sehen, dass die beiden Intervalle überlappungsfrei sind. Das obere Ende des Konfidenzintervalls von Looserprolol (1,83) ist kleiner als das untere Ende des Konfidenzintervalls von Superprolol (2,16). Das Konfidenzinveau (ist im Screenshot nicht enthalten) liegt bei 95%.

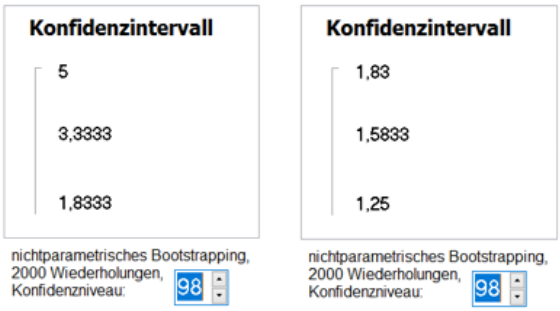
In Statistik-Sprech heißt das, dass Superprolol bei einem 95% Konfidenzniveau *statistisch signifikant besser* wirkt als Looserprolol.

Kritisches Signifikanzniveau

Das Superprolol-Team ist zufrieden, weil es zeigen kann, dass Looserprolol im Durchschnitt nicht so gut wirkt.

Nun ist aber noch eine etwas größere Lücke zwischen dem unteren Ende des Intervalls von Superprolol und dem oberen Ende des Intervalls von Looserprolol. Man könnte jetzt am Konfidenzniveau drehen. Wenn man es von 95% anhebt, werden die Intervalle breiter. Wir werden gleich überlegen, warum.

Das kritische Signifikanzniveau ist das Niveau, bei dem sich die beiden verglichenen Intervalle gerade nicht mehr überschneiden.



Meine Software kann das nicht automatisch. Ich habe herumgedreht und bin auf 98% gekommen. Das bedeutet, dass das Superprolol-Team die Aussage machen kann, dass die

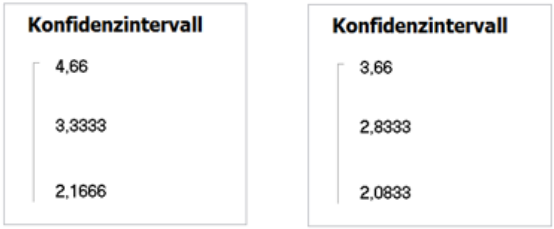
[5]Falls es mir nicht gelungen ist, Ihnen den Unterschied zwischen der richtigen und falschen Interpretation des Konfidenzniveaus zu vermitteln, müssen Sie den Absatz noch einmal ganz genau lesen oder aus einer anderen Quelle recherchieren. Dieser Unterschied ist wichtig.

Irrtumswahrscheinlichkeit für „Superprolol ist besser als Looserprolol“ nicht bei 5% liegt, sondern nur bei 2%. Das ist ein gutes Argument.

Zurück zur Frage, warum die Intervalle bei steigendem Signifikanzniveau breiter werden. Dazu möchte ich kurz das Beispiel wechseln, weil mir das intuitiver erscheint. Wir fragen 1000 Passanten, welche Partei sie bei der nächsten Bundestagswahl wählen werden. 8,3% der Befragten geben an, die FDP zu wählen. Wie wahrscheinlich ist es, dass das tatsächliche Ergebnis zwischen 8,25% und 8,35% liegen wird? Nicht sehr wahrscheinlich. Wir haben also eine relativ hohe Irrtumswahrscheinlichkeit. Wie können wir diese Wahrscheinlichkeit senken? Indem wir das Intervall breiter machen. Ich kann Ihnen, ohne auch nur eine einzige Person zu befragen, mit Sicherheit sagen, wie groß der Stimmenanteil der FDP sein wird: Er wird zwischen 0% und 100% liegen. Irrtumswahrscheinlichkeit Null. Aussage aber auch Null. Das bedeutet, dass die Breite des Intervalls und die Höhe der Irrtumswahrscheinlichkeit zusammenhängen.

Fehlende Signifikanz

Wir haben Superprolol mit Looserprolol verglichen. Superprolol wirkt statistisch signifikant besser, da sich die Konfidenzintervalle nicht überlappen. Jetzt wollen wir Superprolol (links) mit Babboprolol (rechts), dem bisherigen Marktführer vergleichen.



Wenn wir die Intervalle vergleichen, sehen wir, dass sie sich überlappen. Der Durchschnittswert von Superprolol ist zwar höher als der von Babbostatin, aber wir wissen, wie wenig belastbar dieser Wert ist. Also ist das belastbarste, was wir haben, die Konfidenzintervalle. Und die sind nicht überlappungsfrei. Und damit ist Superprolol eben *nicht* „statistisch signifikant besser“ als Babbostatin. Es ist auch nicht statistisch signifikant schlechter, sondern wir können auf der Basis der vorliegenden Daten keinen statistisch signifikanten Unterschied bei der Wirkung der beiden Wirkstoffe ermitteln.

Hierbei ist es statistisch irrelevant, ob sich die beiden Intervalle stark oder nur ganz wenig überlappen.

> *Überlappen heißt „statistisch nicht signifikant verschieden“.*

Aus dem letzten Abschnitt über das kritische Signifikanzniveau wissen wir, dass wir an der Irrtumswahrscheinlichkeit drehen können. In diesem Fall könnten wir auf die Idee kommen, die Irrtumswahrscheinlichkeit so lange hochzudrehen, bis es keine Überlappung mehr gibt. Das kann man machen, aber zusammen mit dem „statistisch-signifikant-Spruch“ gibt man immer das Signifikanzniveau an. Und wenn das unter 95% liegt, hört einem keiner mehr zu, weil das, siehe oben, unter den Mindestanforderungen liegt, die als Standard gelten.

Umgangssprachlicher Missbrauch

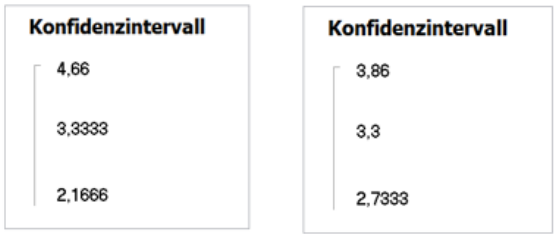
Der Begriff *statistisch signifikant* wird häufig aus Unkenntnis oder Manipulationsabsicht falsch verwendet. Hin und wieder hört man Aussagen wie: „Die durchschnittliche Wirkung

von 3,33 bei Superprolol ist statistisch signifikant höher als die von Babbostatin, die nur bei 2,83 liegt.“ Der Begriff „signifikant“ wird hier hin der Bedeutung „ich finde, der Unterschied der Mittelwerte ist ziemlich groß“ verwendet.

In den meisten Fällen werden solche Aussagen vermutlich von Personen getroffen, die zu wenig Ahnung von Statistik haben und die Aussage „der Mittelwert in beiden Studien ist verschieden“ ein bisschen beeindruckender und nach schlauer Wissenschaft klingen lassen wollen, aber dabei ins Fettnäpfchen treten. Es könnte aber auch sein, dass jemand, der weiß, dass er Unfug erzählt, das tut, weil er hofft, dass *Sie* nicht wissen, was Signifikanz überhaupt ist. In beiden Fällen sollten Sie diese Studie nur mit ganz spitzen Fingern anfassen, denn entweder haben die Macher keine Ahnung oder sie spekulieren auf die Uninformiertheit der Adressaten.

F11.5 Die Bedeutung der Stichprobengröße

Als Ökonomen sind wir diejenigen, die bei Studien als erstes an die Kosten denken. Je kleiner die Probandenzahl ist, umso billiger wird das Ganze. Bei meiner Superprolol-Studie war ich so geizig, dass ich nur 6 Probanden hatte. OK: ich war noch geiziger und habe einfach einen Würfel geworfen. Was passiert nun, wenn wir nicht ganz so geizig sind, und mehr Probanden in unsere Studie aufnehmen. Ich habe das simuliert, indem ich (ohne Schummeln. Ehrlich!) noch 24 Mal gewürfelt habe, so dass meine Superprolol-Studie jetzt 30 Probanden hat. Die Daten habe ich wieder in meine Software gefüttert und folgende Ergebnisse (links: n=6; rechts: n=30) bekommen:



Sie können sehen, dass das Intervall kleiner wird. Der Grund ist, dass das Gesetz der großen Zahl anfängt zu wirken. Nun ist 30 nicht wirklich eine große Zahl und das Intervall immer noch breit wie ein Scheunentor, aber es ist geschrumpft. Wenn ich jetzt noch 540 Mal gewürfelt hätte (wozu ich aber gar keine Lust hatte), wäre das Intervall noch weiter geschrumpft. Wir nehmen das erst einmal einfach so hin. Aber warum ist das so?

Ich möchte noch einmal auf meine FDP-Wahlprognose zurückkommen. Es gibt in Deutschland etwa 62 Mio. Wahlberechtigte. Wenn meine Stichprobe 62 Mio. Teilnehmer umfasst, habe ich die bestmögliche Prognose, wie groß der Stimmenanteil der FDP bei der nächsten Bundestagswahl sein wird. Mein Konfidenzintervall ist Null. Wenn ich jetzt 2 Mio. Wahlberechtigte nicht frage, sondern nur 60 Mio, sinkt die Präzision ein wenig und der Wert, den ich herausbekomme, könnte ein leicht anderer sein. Mein Konfidenzintervall wird also breiter. In der Realität frage ich aber vielleicht nur 2.000 Wahlberechtigte und mein Intervall ist relativ breit. Stocke ich diese Zahl auf, d.h. gebe mehr Geld aus, werden die Ergebnisse zuverlässiger und das Intervall schmaler.

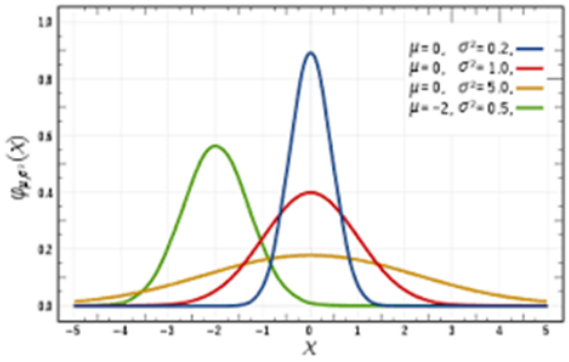
F11.6 Berechnung der Intervalle

Bisher haben wir uns nur mit der Interpretation von Konfidenzintervallen befaßt, aber nicht, wie man sie *berechnet*. Es gibt zwei grundsätzlich verschiedene Arten, wie man das tun kann.

Parametrische Verfahren

Die weniger aufwändigen Verfahren werden parametrische Verfahren genannt. Diese Verfahren setzen voraus, dass man den *Verteilungstyp* der Grundgesamtheit kennt.

Excel besitzt eine Funkktion zur Berechnung von Konfidenzintervallen. Diese Funktion setzt (ohne darauf hinzuweisen) eine *normalverteilte* Grundgesamtheit voraus, also etwas, das so aussieht:

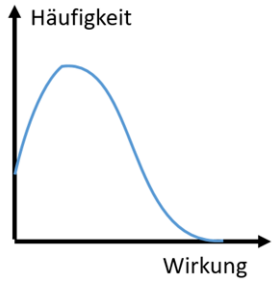


Grob gesagt ist die Normalverteilung eine eingipfelige, symmetrische Verteilung d.h. der häufigste Wert liegt in der Mitte aller Werte. Sehr kleine und sehr große Werte sind selten.

Ein parametrisches Konfidenzintervall geht nun nach folgender Überlegung vor: Ich habe hier eine Stichprobe in der es kleine und große Werte gibt. Diese Stichprobe steht stellvertretend für eine Grundgesamtheit, die normalverteilt ist. Also versuche ich, das Intervall so zu berechnen, dass diese Werte halbwegs in diese Normalverteilung passen.

Das Problem ist nun, dass die Ergebnisse eines Konfidenzintervalls für eine normalverteilte Grundgesamtheit völlig unbrauchbar sind, wenn gar keine Normalverteilung in der Grundgesamtheit vorliegt, denn das Verfahren quetscht die Daten ja in diese unterstellte Verteilungsform.

Sicherlich gibt es viele Bereiche, in denen die Unterstellung einer normalverteilten Gesundheit ganz plausibel ist. In gesundheitsökonomischen Fragen ist das aber definitiv nicht so. Es ist nicht nur nicht plausibel, dass eine Normalverteilung vorliegt, sondern definitiv höchst unplausibel. In Pharmastudien erwartet man eher Verteilungen, die etwa wie folgt aussehen:



Es gibt einen relativ hohen Anteil von Patienten, bei denen das Medikament nicht wirkt. Mit etwas gutem Willen könnte man diesen Verteilungstyp als Normalverteilung

bezeichnen, bei dem links ein gutes Stück der Verteilung abgeschnitten ist.

Weil es aber keine Normalverteilung ist, ergibt die Berechnung eines Konfidenzintervalls, das dies voraussetzt, keinen Sinn. Natürlich bekommt man Werte, aber diese Werte sind unbrauchbar.

Es gibt auch parametrische Verfahren für andere Verteilungstypen, z.B. lognormale Verteilung, Poisson-Verteilungen, Weibull-Verteilungen usw. Die Berechnung hinter diesen Alternativen ist jeweils eine andere, aber die Verfahren haben gemein, dass immer klar sein muß, dass die Grundgesamtheit den jeweiligen Verteilungstyp des verwendeten Verfahrens hat. Ist das nicht so, sind die Ergebnisse Müll. In der Statistik-Veranstaltung lernen Sie den Shaprio-Wilk-Test kennen, mit dessen Hilfe man entscheiden kann, ob eine Stichprobe normalverteilt ist oder nicht. Wie immer gibt es einen statistischen Pferdefuß. Der Test liefert eine „ja/nein“ Aussage, aber mit einer Irrtumswahrscheinlichkeit, die man „einstellen“ kann.

Nichtparametrische Verfahren

Nichtparametrische Verfahren arbeiten ohne Verteilungsannahme. Sie arbeiten computerbasiert mit dem Monte-Carlo Prinzip. Dieses Prinzip bedeutet, dass man Zufallsprozesse einfach mit einem Computer durchsimuliert und dann schaut, was herauskommt. Aus der Stichprobe zieht die Software zufällig Datensätze nach dem Prinzip „ziehen mit zurücklegen“.

Nehmen wir an, das Originalsample (deutsch: die Originalstichprobe) ist

Beobachtungs-nummer	1	2	3	4	$\bar{x}$
Originalsample	2	6	3	1	3,0

Der Mittelwert $\bar{x}$ ist 3,0.

Jetzt baue ich einen (computerbasierten) Zufallsgenerator, der mir mit gleicher Wahrscheinlichkeit eine Zahl zwischen 1 und 4 (der Anzahl der Beobachtungen) auswirft. Diesen Generator lasse ich vier Mal laufen. Das Ergebnis sei

$1; 3; 2; 2$.

Mit diesem Ergebnis erstelle ich ein sog. synthetisches Sample. Der Wert 1 des synthetischen Samples ist die Beobachtungsnummer 1 des Origialsamples, also 2. Der Wert 2 des synthetischen Samples die Beobachtungsnummer 3 des Originalsamles, also 3. Der Wert 3 des synthetischen Samples ist Beobachtungsnummer 2, also 6. Der Wert 4 des synthetischen Samples ist ebenfalls Beobachtungsnummer 2, also 6.[6]

Beobachtungs-nummer	1	2	3	4	$\bar{x}$
Originalsample	2	6	3	1	3,0
syn. Sample 1	2	3	6	6	4,25
syn. Sample 2					

Dieses synthetischen Sample hat den Mittelwert von 4,25.

Ich ziehe nun nicht nur *ein* synthetisches Sample, sondern 2.000.[7] Am Ende hat man dann 2.000 Mittelwerte von 2.000 synthetischen Samples. Ziemlich wahrscheinlich gibt

[6] Es handelt sich also um die Struktur „Ziehen mit Zurücklegen“

[7] Diese Wiederholungszahl ist wieder eine statistische Konvention. Die Erfahrung lehrt, dass die Ergebnisse sich nicht mehr nennenswert ändern, wenn man noch mehr Samples generiert.

es mehrere synthetische Samples, in denen viermal der kleinste Wert vorkommt und der Mittelwert dann 1,0 ist. Es werden auch mehrere Samples erzeugt werden, bei denen viermal der größte Wert vorkommt und der Mittelwert 6,0 ist. Diese 2.000 Mittelwerte sortiert man nun nach der Größe. Wenn man nun die 50 (2,5%) größten und die 50 (2,5%) kleinsten Werte wegwirft, erhält man die Grenzen des 95% Konfidenzintervalls.

Die nichtparametrischen Verfahren sind eine Mörderrechnerei, aber *Sie* müssen ja nicht rechnen, sondern der Computer. Jede vernünftige Statistiksoftware bietet diese Funktionen an.

Meine Software habe ich aus didaktischen Motiven an einem Nachmittag zusammengebastelt. Sie hat 120 Zeilen Code und braucht für ein Originalsample mit 1.000 Werten auf meinem schon etwas älteren Mittelklasse-Notebook 2 Sekunden.

Sie sollten bei eigenen Auswertungen daher nur dann ein *parametrisches* Verfahren wählen, wenn Sie sicher sind, dass die Grundgesamtheit entsprechend verteilt ist. Sonst sollten Sie tunlichst zu diesen nichtparametrischen Verfahren greifen.

F11.7 Ehemalige Prüfungsfragen

Extreme synthetische Samples

Wie hoch ist die Wahrscheinlichkeit, dass bei einem Originalsample von 2; 6; 3; 1. ein synthetisches Sample 1; 1; 1; 1 ist?

Empirische Abschlußarbeiten[8]

Viele Abschlußarbeiten beinhalten einen empirischen Teil. Der Autor konzipiert einen Fragebogen, verschickt ihn, wartet auf den Rücklauf und wertet ihn aus.

Ich möchte Ihnen für diesen Typ von Abschlußarbeit folgenden Vorschlag machen: Vergessen Sie das Verschicken und die Auswertung der Daten. Wenn das Unternehmen, für das Sie die Auswertung fahren sollen, wirklich Interesse hat, soll es Sie einstellen und ordentlich Geld auf den Tisch legen! Stecken Sie Ihre Energie *komplett* in die Entwicklung des Fragebogens und den Auswertungsplan, d.h. die Frage, mit welchen Instrumenten Sie die Variablen, die Sie im Fragebogen haben, anschließend auswerten wollen. Schreiben Sie die SPSS-Programme.

Aber bitte warten Sie nicht auf die Daten, weil Sie die Fragebögen gar nicht losschicken. Wenn ihr Betreuer Zahlen sehen will, erfinden Sie welche. Aber bitte: kündigen Sie das offen an und schreiben es auch in der Arbeit ;)

1. Welche Überlegung aus dem Kontext „Konfidenzintervall“ steht hinter dem seltsamen Vorschlag?

2. Einigen Unternehmen ist diese Überlegung nicht klar. Warum haben einige Unternehmen trotzdem (obwohl ihnen das klar ist) Interesse an Abschlussarbeiten in denen echte Daten erhoben werden?

[8] Ausnahmsweise mal keine ehemalige Prüfungsfrage.

Die kaputte Festplatte

Sie haben eine Blutdrucksenkerstudie für einen neuen Wirkstoff mit 300 Probanden durchgeführt und wollen nun mit Hilfe eines nichtparametrischen Verfahrens das 95% Konfidenzintervall für die Ergebnisse berechnen.

Gerade als Sie das Programm starten, hören Sie ein kirschendes Geräusch. Die Festplatte Ihres Rechners hat gerade den Geist aufgegeben und ihr Rechner bricht die Berechnungen ab. Es gibt Sicherheitskopien, aber es ist Donnerstagmittag und der Kollege, der das Passwort für die die Kopie hat, ist bis Montag im Urlaub und nimmt bekanntermaßen kein Handy mit.

Nun sollen Sie morgen, am Freitag, unternehmensintern erste Ergebnisse der Studie präsentieren und das Konfidenzintervall sollte ein wichtiger Punkt sein. Sie haben keine Existenzängste, da Sie der IT-Abteilung im Wochentakt Emails geschickt habe, daß Ihr Rechner komische Geräusche macht und sind immer wieder vertröstet worden. Wenn Sie morgen bei der Präsentation die „Festplatte kaputt“-Geschichte erzählen, werden nicht Sie einen Einlauf bekommen, sondern der IT-Mensch.

Gerade als Sie den nicht mehr arbeitenden Rechner ausschalten wollen, sehen Sie, daß das Programm während es gelaufen ist, die Mittelwerte der synthetischen Samples auf dem Bildschirm aufgelistet hat. Sie erkennen, daß der Rechner bei der Berechnung des 17. Synthetischen Samples abgebrochen hat. Sie fotografieren den Bildschirm mit dem Handy. Vielleicht ist das besser als nichts.

	A
1	4,2
2	6,3
3	7,2
4	2,8
5	3,1
6	6,8
7	6,8
8	2,9
9	3,2
10	4,3
11	4,4
12	6,5
13	4,4
14	6,8
15	6,7
16	3,0
17	

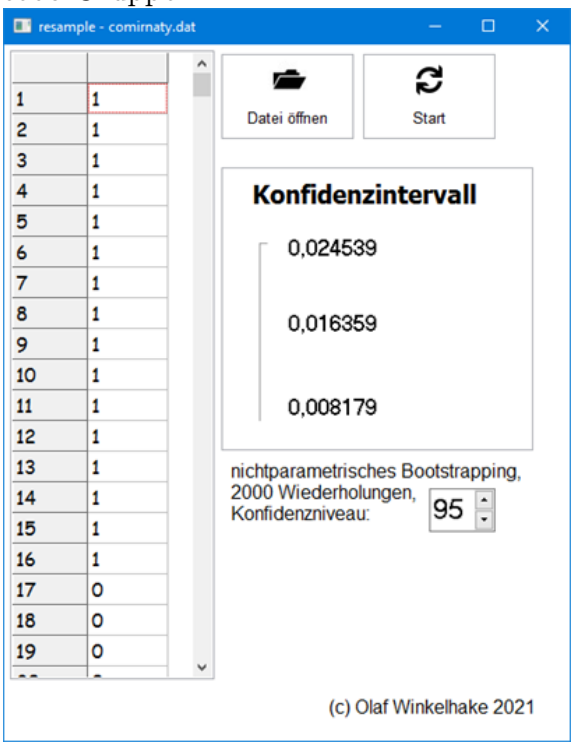
Gibt es irgendwelche Informationen, die Sie aus diesem Screenshot ziehen können und morgen als „besser als nichts – und ich bin auch nicht Schuld“ Ergebnis präsentieren können?

Konfidenzintervall der Comirnaty-Studie

Eine aktuell (Juli 2021) intensiv diskutierte Frage ist, ob Kinder zwischen 12 und 16 Jahren mit BioNTech gegen Corona geimpft werden sollen oder nicht. Den folgenden Text habe ich aus der derzeit einzigen Studie, die zu dieser Frage durchgeführt worden ist. Die Studie wurde im März 2021 beendet. Die Beobachtungsdauer lag bei 2 Monaten nach der zweiten Impfung. „Comirnaty“ ist der Handelsname des BioNTech-Impfstoffs.

> *The efficacy of Comirnaty was calculated in close to 2,000 children from 12 to 15 years of age who had no sign of previous infection. These received either the vaccine or a placebo (a dummy injection), without knowing which one they were given. Of the 1,005 children receiving the vaccine, none developed COVID-19 compared to 16 children out of the 978 who received the dummy injection. This means that, in this study, the vaccine was 100% effective at preventing COVID-19.*

Mit jeweils etwa 1.000 Probanden ist die Größe der Comirnaty-Studie mittel. Das bedeutet, dass man sich den Stichprobenfehler und die Konfidenzintervalle anschauen sollte. Weil „infiziert/nicht infiziert“ eine binäre Variable (0=gesund; 1=infiziert) ist und die Zahlen vorliegen, konnte ich eine zur Studie passende Daten-Datei erstellen, die aus 16 mal einer 1 und 962 mal einer 0 bestand und durch meine Software laufen lassen konnte, die ein nichtparametrisches Konfidenzintervall berechnet. Hier ist das Ergebnis für die Placebo-Gruppe:



1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	0
18	0
19	0

Das bedeutet, dass das Infektionsrisiko in der Studie bei 1,6% lag und die Grenzen des 95% Konfidenzintervalls bei 0,8% und 2,5%. Können Sie (ohne Software) eine mehr oder minder genaue Aussage darüber machen, wie das Ergebnis dieser Software für die Gruppe, die den echten Impfstoff erhalten hat, aussehen würde?

Kann man Aussagen über statistische Signifikanz der Wirkung des Impfstoffs machen?

Extreme synthetische Samples (Lösung)

Dieses Ergebnis würde dann entstehen, wenn mit einem 4er Würfel viermal hintereinander eine „4“ gwürfelt wird, denn die „1“ ist der vierte Originalwert.

Die Wahrscheinlichkeit, dass bei einmaligem Würfeln eine „4“ gewürfelt wird, ist 0,25. Die Wahrscheinlichkeit für *„viermal eine 4 würfeln“* ist dementsprechen $0,25^4 = 0,0039$

Empirische Abschlussarbeiten (Lösung)

- Sowohl die Zeit als auch die finanziellen Mittel führen dazu, daß die Stichprobe in einer empirischen Abschlussarbeit sehr klein sein wird.
- Das bedeutet, daß die Konfidenzintervalle relativ breit sein werden.
- Das bedeutet, daß es sehr unwahrscheinlich ist, auf statistisch signifikante Unterschiede zu stoßen.
- Man kann damit rechnen, daß die Mittelwerte deutlich verschieden sein werden, aber ob das nur ein Zufallseffekt ist oder nicht, wird man wahrscheinlich nicht erkennen. Also bedeutet die Arbeit, die man in die Erhebung der Daten steckt, vergebene Liebesmüh. Man sollte die Zeit besser für das Studiendesign verwenden.
- Auftraggebern, denen die mangelnde statistische Signifikanz der Ergebnisse klar ist, können solche Ergebnisse verwenden, um in unternehmensinternen Diskussionen statistische Laien hereinzulegen.

Die kaputte Festplatte (Lösung)

Bei 16 sythetischen Samples kann man den obersten und untersten Wert abschneiden. Das wären dann jeweil 6,25% unten und oben, bzw. eine Irrtumswahrscheinlichkeit von 12,5% bzw. ein Konfidenzniveau von 87,5%. Die Grenzen des Invervalle erhält man, indem man den kleinsten und größte Wert abschneidet und somit den zweitkleinsten und zweitgrößten Wert als Grenzen des Intervalls erhält. Das sind 2,9 bzw. 6,8.

Konfidenzintervall der Comirnaty-Studie (Lösung)

Das Sample umfasst 1005 Datensätze, von denen kein einziger infiziert ist. In einem synthetischen Sample *kann* es keine Infizierten geben. Die Infektionswahrscheinlichkeit ist daher in allen 2000 synthetischen Samples Null. Die untere und obere Grenze des Intervalls ist damit auch Null.

Die untere Grenze der Placebo-Stichprobe ist positiv. Also ist die Wirkung der Impfung statistisch signifikant.

Teil IV

Incentives für Versicherte und Leistungserbringer

Halbzeit

Die erste Halbzeit dieser Veranstaltung ist nun vorbei. In der folgenden Concept-Map möchte ich kurz skizzieren, wo wir uns befinden.

Dazu möchte ich den sozialkompatibleren Begriff *Priorisierung* verwenden, also die Unterscheidung zwischen wichtigeren und weniger wichtigen medizinischen Leistungen.

Bisher war unsere Frage, wie wir zwischen diesen beiden Typen unterscheiden können. Es ging also um das „was“. Wir haben uns mit der Frage der ethischen Zulässigkeit befasst, wie wir Effizienz messen und mit welchen Kennzahlen wir das tun.

Mit diesen Fragen sind wir nun „durch“, soweit wir das im zeitlichen Rahmen dieses Moduls können.

Ausblick

Wenn das *„was“* klar ist, ist die nächste Frage *„wie machen wir das denn dann“*? In *Einheit* **??** haben wir am Rand schon über Mechanismen gesprochen, z.B. die Warteschlange. Auf diesen „how to“ Aspekt wollen wir uns aber nun konzentrieren.

Incentives

Der erste Aspekt, den wir in diesem Teil abhandeln wollen, ist die Frage, welche Incentives (Anreize) man den Beteiligten geben kann, damit sie sich aus eigenem Interesse effizienter verhalten, als dies bisher der Fall war. Das ist eine viel weniger drastische Form von „Rationierung“, weil sie letztlich eine freiwillige Selbstrationierung ist, wie wir sie ja von jedem Markt kennen.

Unter „Beteiligten“ will ich hier zunächst die Versicherten verstehen und dann die Leistungserbringer. Die Versicherten werden wir relativ knapp behandeln, weil (Achtung: Spoiler) die Instrumente ziemlich stumpf sind. Wir werden mehr Zeit mit den Leistungserbringern verbringen, weil das Steuerungspotential dort größer ist.

Organisationsstrukturen

Das letzte Viertel der Veranstaltung (die zweite Hälfte der zweiten Hälfte) wollen wir uns mit der Frage befassen, in welchen Organisationsstrukturen wir agieren. Dazu ist es hilfreich, sich die Gesundheitssysteme anderer Länder anzusehen, denn die Effizienz- und Priorisierungsprobleme, mit denen das deutsche System konfrontiert ist, sind ja nur zu einem Teil „deutsche“ Probleme. Die Systeme in anderen Ländern müssen häufig mit den gleichen Problemen fertigwerden und es ist klug, zu schauen, ob sie für diese Probleme andere, vielleicht bessere Lösungen gefunden haben.

Kapitel 12

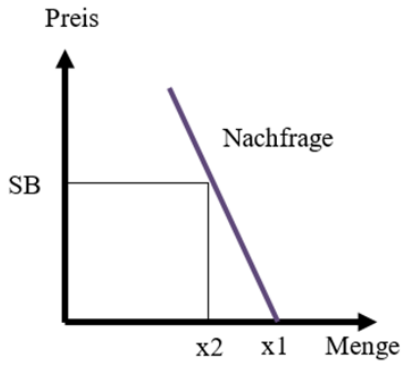
Selbstbehalte und Boni

In der Sitzung

- Wir klären Ihre Fragen zu dieser Einheit. Damit Sie nicht hektisch in Ihren Zetteln wühlen müssen: Schreiben Sie sich Ihre Fragen während des Durcharbeitens auf einem separaten Zettel auf.
- Wir setzen die Eigenbeteiligungs-Idee aus dem Geinitz-Artikel mit Excel um. Bringen Sie Ihr Notebook mit.
- Wir experimentieren mit dem Excel-Selbstbehaltsrechner und schauen an, welche (Verteilungs)wirkungen unterschiedlich hohe Sebstbehalte haben.

F12.1 Das Konzept der Selbstbehalte

Aus der Sicht eines Patienten ist seine medizinische Versorgung eine Flatrate. Er zahlt einen monatlichen Beitrag und kann dann alle Leistungen kostenlos in Anspruch nehmen. Durch dieses Arrangement hat er keinen Anreiz, sparsam zu sein.



In einer geometrischen Darstellung des Nachfrageverhaltens würde der Versicherte die Sättigungsmenge x_1 nachfragen. Mit Selbstbehalten wird der Versicherte an den Kosten seiner medizinischen Versorgung beteiligt. Auf diese Weise führt man den Preismechanismus teilweise wieder in das Gesundheitssystem ein. Der Versicherte muß entscheiden, ob die Leistung aus seiner Sicht für ihn den Preis „wert“ ist. Wenn das nicht der Fall ist und der Versicherte die Leistung für verzichtbar hält, wird er aus eigenem Interesse auf die Nachfrage verzichten. In der Abbildung führt ein Selbstbehalt in Höhe von SB zu einer Reduktion der Nachfrage auf x_2.

F12.2 Formen von Selbstbehalten

Es gibt drei Formen von Selbstbehalten.

1. Bei einer *absoluten Abzugsfranchise* zahlt der Versicherte bis zu einer jährlichen Obergrenze (Priorität) 100% der anfallenden Kosten. Bei einer Priorität von 1.000€ und Kosten von 600€ zahlt er die 600€ komplett. Bei Kosten von 1.200€ zahlt er die Kosten bis zur Priorität von 1.000€ selbst und die Versicherung trägt die verbleibenden 200€.

2. Bei einer *relativen Abzugsfranchise* zahlt der Versicherte einen Anteil von $x\%$ der Kosten. Auch hier gibt es eine Priorität. Bei einer Priorität von 1000€, einem Selbstbehaltsanteil von 20% und Kosten von 600€ zahlt er 120€, die Versicherung 480. Bei Kosten von 1.200€ zahlt er 240€. Bei Kosten von 10.000 zahlt er seine Priorität von 1.000€ komplett und die Versicherung den Rest.

3. Etwas komplexer ist die *Beitragsrückerstattung*. Ein Versicherter, der keine Leistungen in Anspruch genommen hat, erhält eine Beitragsrückerstattung (BRE) am Ende des Jahres. Hat er Leistungen in Anspruch genommen, werden diese von dem Betrag abgezogen. Bei einer BRE von 1.000 und Kosten von 600 erhält er 400€, bei Kosten von 1.200€ nichts. Die BRE ähnelt somit der absoluten Abzugsfranchise mit dem Unterschied, daß der Versicherte dort bei Inanspruchnahme Geld zahlen

muß, während er hier die BRE im Vorhinein über höhere Beiträge zahlt und bei Nichtinanspruchnahme zurückbekommt.[1]

F12.3 Der Wirkungsbereich von Selbstbehalten

Die übliche Höhe von Selbstbehalten

Allen Formen von Selbstbehalten gemein ist, dass es eine finanzielle Deckelung (Priorität) gibt. Jenseits dieser Priorität wirken Selbstbehalte nicht mehr. Erreicht ein Versicherter im Laufe des Jahres die Priorität, wird er sich für den Rest des Jahres nicht sparsamer verhalten als ohne Selbstbehalt. Weiß der Versicherte bereits zu Anfang des Jahres, daß er seine Priorität erreichen wird, ist der Selbstbehalt komplett wirkungslos.

Viele PKV-Tarife und einige GKV-Tarife bieten Selbstbehaltsregelungen an. Die Prioritäten liegen hier im Regelfall bei ca. 300-500€ pro Jahr. Anders als in der Schweiz gibt es in Deutschland keine Selbstbehalts-Pflicht. Dort müssen alle Tarife mindestens eine Priorität von 1.000€ beinhalten, die sich auf eine festgelegte Mischung aus absoluter und relativer Abzugs-franchise verteilen.

Die Wirkung bei Hochkostenversicherten

Im Bachelor haben wir über die Verteilung der Gesundheitsausgaben auf die Versicherten gesprochen. Die Daumenregel war, das 80% der Kosten auf 10 % der Versicherten entfallen. In Deutschland sind das für diese Gruppe weit über 20.000€ pro Kopf. Bei diesen Zahlen ist klar, dass die Kosten eines Hochkostenpatienten die Priorität schon im Januar übersteigen werden. Das bedeutet, dass dem Hochkostenpatienten klar ist, dass sich freiwillige Einschränkungen für ihn nicht auszahlen werden. Er wird sein Verhalten nicht ändern. Es gibt keine Steuerungswirkung.

Das bedeutet, dass bei dieser Kostenkonzentration beim Löwenanteil von 80% der Kosten Selbstbehalte völlig wirkungslos sind.

Die Wirkung bei Nullkostenversicherten

Ein weiteres Ergebnis der Bachelor-Einheit war, dass 40% der Versicherten gar keine Kosten verursachen. Das bedeutet, dass Selbstbehalte auch bei dieser Gruppe nicht kostendämpfend wirken, weil es gar keine Kosten gibt.

Die Wirkung bei Mittelkostenversicherten

Die einzige Gruppe, bei der über Selbstbehalte Geld einsparbar wäre, sind also die Mittelkostenversicherten, weil Null- und Hochkostenversicherte ihr Verhalten nicht ändern werden.

[1] Die meisten Krankenversicherer vereinfachen dieses Konzept noch weiter. Die Beitragsrückerstattung gibt es nur bei kompletter Leistungsfreiheit, d.h. wenn *gar keine* Kosten bei dem jeweiligen Versicherten angefallen sind. Das hört sich drakonischer an als es ist, weil diese Versicherungen nach dem Kosterstattungsprinzip arbeiten, das Sie im Bachelor kennengelernt haben. Die Idee ist, dass die Versicherten erst dann Rechnungen einreichen, wenn der Betrag dieser Rechnungen die Beitragsrückerstattungen übersteigt. Auf diese Weise spart sich die Versicherung die Überprüfung der Rechnungen, weil diese Rechnungen erst gar nicht zur Erstattung eingereicht werden und senkt so die Verwaltungskosten.

	A	B	C	D	E	F	G
1			hoch	mittel	null	Σ	Ø
2	Anzahl		10	50	40	100	
3	Kosten pro Kopf		24.000 €	1.200 €	- €	300.000 €	3.000 €
4	Priorität			1200			
5	Einsparpotential						
6		%	20%	20%	20%		
7		€/Kopf	4.800 €	240 €	- €	60.000 €	600 €
8	tatsächliche Einspar. €/Kopf		- €	240 €	- €	12.000 €	120 €
9	Pro-Kopf-Kosten nach SB		24.000 €	960 €	- €	288.000 €	2.880 €
10	Pro-Kopf-SB		1.200 €	960 €	- €	60.000 €	
11	Prämie nach SB		2.280 €	2.280 €	2.280 €	228.000 €	
12	Prämie + SB		3.480 €	3.240 €	2.280 €	288.000 €	
13	Veränderung		480 €	240 €	- 720 €	- 12.000 €	

Abbildung 12.1: Verteilungswirkung von Selbstbehalten

Aus den sehr groben Zahlen, die wir kennen, können wir die Größenordnung abschätzen, die theoretisch über Selbstbehalte einsparbar sind. Wenn die Mittelkostenversicherten 20% der Kosten verursachen wird ein Selbstbehalt ja nicht dazu führen, dass diese 20% komplett vermieden werden. Nehmen wir an, dass wir 20% der Kosten auf diese Weise einsparen können. Das sind 4% der Gesamtkosten. Das ist in etwa der Wert, um die die GKV-Ausgaben in jedem Jahr steigen. [2]Das bedeutet, dass wir, wenn es um das Einsparpotential geht, nicht mit allzu großen Erwartungen an die Selbstbehalte herangehen sollten. Dazu aber später mehr.

F12.4 Die Verteilungswirkungen von Selbstbehalten

Unsere Überschlagsrechnung hat uns vorgewarnt: Wir sollten nicht allzu große Wirkungen erwarten. Wenn es aber eine Wirkung, sei sie auch klein, gibt, warum soll man dann groß weiter über die Einführung von Selbstbehalten nachdenken. Man sollte doch dann einfach „machen“. Oder?

In solchen Fällen habe ich es mir zur Gewohnheit gemacht, trotzdem darüber nachzudenken, ob es unterschiedliche Personengruppen gibt, die unterschiedlich betroffen sind. Meine Erfahrung ist, dass man ziemlich häufig feststellt, dass der *Durchschnittseffekt* gering ist, es aber identifizierbare Gruppen gibt, die stark profitieren bzw. verlieren. Wenn Sie eine Hand in Eiswasser stecken und die andere in kochendes Wasser, dann ist die Temperatur im *Durchschnitt* ganz angenehm.

Um die Wirkungen auf identifizierbare Untergruppen zu untersuchen, habe ich ein paar halbwegs realistische Zahlen in eine Modellrechnung gestopft, die Sie in der Excel-Tabelle weiter unten sehen können.

[2]Falls Ihnen der Unterschied nicht geläufig ist: Ich habe hier von zwanzig *Prozent* und nicht zehn *Prozentpunkten* gesprochen. 20% von 20% sind eben 4%.

Vor dem Selbstbehalt

Nehmen wir an, die Pro-Kopf-Ausgaben einer Krankenversicherung liegen bei 3.000€. Das entspricht in etwa den GKV-Werten von 2017. Nach der 10:80-Regel bedeutet das, dass die 10% Hochkostenpatienten 24.000€ pro Kopf Kosten und die 50% Mittelkostenpatienten 1.200€.

Nehmen wir weiterhin an, dass 20% der Kosten vermeidbare Kosten wären. Das ist ein Wert, der in diesem Zusammenhang gern verwendet wird, für den es aber nur anekdotische Evidenz gibt. Aber nehmen wir an, dieser Anteil wäre so hoch.

Bei 100 Versicherten[3] betragen die Kosten insgesamt 300.000€, von denen annahmegemäß 60.000€ vermeidbar wären. Nun soll ein Selbstbehalt von 1.200€ die Patienten motivieren, diese vermeidbaren Kosten von 60.000€ auch tatsächlich zu vermeiden.

Wirkung des Selbstbehalts

Nach den bisherigen Überlegungen ist klar, dass die Hochkostenversicherten keinen Anreiz haben, das zu tun und die Nullkostenversicherten gar keine Kosten haben, die weiter reduzierbar wären.

Also bleiben die Mittelkostenversicherten, für die es finanziell attraktiv ist, die 240€ vermeidbare Kosten tatsächlich zu vermeiden, weil sie sie selbst zahlen müssen, da sie innerhalb der Priorität liegen.

Die Modellrechnung zeigt, dass von dem Einsparpotential in Höhe von 60.000€ (Zelle F7) über Selbstbehalte nur 12.000€ beeinflussbar sind. In Zeile 9 sehen Sie die Reaktionen der Versicherten, die (D9) darin besteht, dass die 50 Mittelkostenpatienten ihre Nachfrage um 240€/Kopf auf 960€ senken, so dass die Gesamtkosten von 300.000€ auf 288.000€ sinken.

In Zeile 10 sind die Selbstbehalte der Versicherten aufgelistet. Die Hochkostenpatienten tragen die kompletten 1.200€ und die Mittelkostenpatienten die 960€, die sie immer noch kosten und die komplett in der Priorität liegen. Insgesamt kommen so 60.000€ an Selbstbehalten zusammen (F10), so dass über die Prämien noch 228.000€ finanziert werden müssen, bzw. 2.880€/Kopf.

In Zeile 12 finden Sie die Gesamtbelastung der drei Gruppen über die neuen Beiträge plus die gezahlten Selbstbehalte und in Zeile 12 den Vergleich mit der alten Prämie von 3.000€.

Es ist nicht überraschend, dass die Hochkostenpatienten stark belastet werden. Man könnte sich auf den Standpunkt stellen, dass das auch in Ordnung ist, da sie 4.800€ vermeidbare Leistungen haben, die sie nicht vermeiden und dafür jetzt 480€ mehr bezahlen müssen.

Was aber überhaupt nicht intuitiv ist, ist die Position der *Mittelkostenpatienten*. Diese Gruppe ist die einzige, die tatsächlich Kosten vermieden hat. Trotzdem wird sie durch den Selbstbehalt mit 240€ stärker belastet. Wie kann das sein?

Der Grund ist, dass die *Hauptprofiteure* des Selbstbehalts die *Nullkostenversicherten* sind, weil sie davon profitieren, dass die beiden anderen Gruppen einen Teil ihrer Versorgung selbst bezahlen. Sie werden um 720€/Kopf entlastet, obwohl sie ihr Verhalten gar nicht geändert haben.

[3]Sie können auch jede andere Zahl verwenden. Hundert liegt nur nahe, weil wir ja mit prozentualen Anteilen rechnen.

Weil das so kontraintuitiv ist, habe ich die Modellrechnung ausführlich als Screenshot abgebildet. Das Excel-Modell können Sie von der Homepage herunterladen und ein wenig an den Zahlen drehen. Das Ergebnis bleibt das, dass *die Gruppe, die Kosten spart, zu den Verlierern gehört.*

F12.5 Die Hidden Agenda von Selbstbehalten

Mit „hidden agenda“ bezeichnet man Motive, die man hinter vorgeschobenen Motiven versteckt, weil die versteckten Motive abschreckend sind.

In diesem Kontext stellt sich folgende Frage: Wenn die Steuerungsmöglichkeiten über Selbstbehalte absehbar marginal sind, warum bieten Krankenversicherer sie dann an?

Eine stillschweigende Annahme in der Modellrechnung war, dass die Versicherten keine Wahl haben, ob sie einen Selbstbehaltstarif abschließen wollen oder nicht. Der Selbstbehalt von 1.200€ war Pflicht. Das ist in der Schweiz so. In Deutschland steht das nicht zur Debatte.

Das bedeutet, dass Mittel- und Hochkostenpatienten die Wahl haben, sich auf einen solchen Tarif einzulassen oder im Standard-Null-Selbstbehalts-Tarif zu bleiben. Auch wenn diese Personengruppen nicht ein so schönes Excel-Modell haben, werden sie sehr schnell merken, dass sie in einem Selbstbehaltstarif stärker belastet werden.

Nach ein bis zwei Jahren können wir davon ausgehen, dass kein Mittel- und Hochkostenpatient, der bis drei zählen kann, noch in einem Selbstbehaltstarif versichert sein wird.

Für die Nullkostenversicherten sieht die Rechnung anders aus. Sie werden entlastet, ohne dass sie ihr Verhalten ändern müssen. Das macht den Selbstbehaltstarif für sie attraktiv. Nach ein bis zwei Jahren sind die verirrten Mittel- und Hochkostenpatienten zwar wieder aus dem Tarif ausgeschieden, aber die Kosten im Selbstbehaltstarif sind trotzdem niedrig, weil die Personen, die sich immer noch dort versichern, sehr gesund sind. Da sind niedrige Ausgaben keine große Leistung.

Im Bachelor haben wir schon über Strategien von *Risikoselektion* nachgedacht und den Gesundheitsfonds als Instrument kennengelernt, solche Strategien für die Gesetzlichen Krankenkassen unattraktiv zu machen. Hier haben wir es mit einer *passiven* Risikoselektion zu tun. Da passive Risikoselektion *(„wir ekeln die Kranken weg“)* nie gut klingt, verkauft man diesen Effekt als *„der Selbstbehalt führt zu kostenbewusstem Verhalten der Versicherten“.*

Der Risikostrukturausgleich sorgt dafür, dass von dieser Risikoselektion nicht mehr viel übrig bleibt. Aus diesem Grund forcieren die Kassen die Selbstbehaltstarife auch nicht besonders.

F12.6 Beispiele für Bonus- und Malussysteme

Die Epidemiologie liefert Hinweise auf Risikofaktoren für Volkskrankheiten. Diese Volkskrankheiten sind zu einem großen Anteil verhaltensabhängig. Übergewicht und Bewegungsarmut stellen für viele Krankheiten ein erhöhtes Risiko dar. Darüber hinaus werden Vorsorgeuntersuchungen häufig nicht in Anspruch genommen, so dass Erkrankungen relativ spät entdeckt werden. Das wirkt sich häufig ausgabensteigernd aus.

Über Prämien für risikosenkendes Verhalten (Bonus) bzw. Strafzahlungen für risikoerhöhendes Verhalten (Malus) kann die Versicherung den Versicherten einen Anreiz geben,

sich risikoarm zu verhalten und so die Häufigkeit des Auftretens der Krankheit zu verringern.

Impfverweigerer

Ein aktuelles[4] Beispiel aus der Corona-Pandemie ist, dass Boris Palmer, OB von Tübingen, fordert, dass Ungeimpfte einen höheren Krankenversicherungsbeitrag zahlen sollen, da sie ein größeres Risiko für eine teure Krankenhausbehandlung haben.[5]

GKV-Bonusheft beim Zahnarzt

Für GKV-Versicherte ist das Zahnarzt-Bonusheft die bekannteste Form eines Bonussystems. Die Techniker-Krankenkasse[6] macht folgende Rechnung für eine Zahnkrone auf: Die Regelversorgung kostet 321,35. Wenn der Patient für fünf Jahre eine lückenlose Vor-sorge-Historie hat, muß er 32,14€ weniger zuzahlen.

An diesem Beispiel wird deutlich, dass Boni und Malus[7] schwer auseinanderzuhalten sind, denn die 32,14€ könnte man nicht als „Belohnung“ für brave Versicherte sehen, sondern genauso gut als „Strafe“ für säumige Versicherte.

Nichtraucherprogramme

Ein anderes Beispiel sind Bonusprogramme, die Raucher dazu motivieren sollen, mit dem Rauchen aufzuhören. In einer US-Studie wurden den (rauchenden) Teilnehmern insgesamt $750 angeboten, wenn sie für 12 Monate mit dem Rauchen aufhören.[8] Hier ist ganz klar, dass es sich um einen Bonus handelt, also eine Belohnung für erwünschtes Verhalten, denn die Versicherten, die weiterhin rauchen, müssen ja nicht *mehr* zahlen, sondern bekommen die das Geld eben *nicht*.

F12.7 Wirkungsabschätzung

Abschreckungspotential der Zahlungen

Die Versicherten, die man über Bonus/Malus-Zahlungen beeinflussen will, stehen den Alternativen, um die es geht, nicht gleichgültig gegenüber.

- sie haben Angst vorm Zahnarzt,
- sie sind nikotinabhängig
- sie fürchten sich vor Nanobots in den Impfstoffen usw.

[4] Ich schreibe dies Ende 2021.

[5] Geinitz, Christian, "Palmer: Ungeimpfte sollen zuzahlen", Frankfurter Allgemeine Zeitung 28.12. 2021. S.15.

[6] https://www.tk.de/techniker/leistungen-und-mitgliedschaft/informationen-versicherte/leistungen/zaehne/zuschuss-zahnersatz-2002872

[7] Der Plural von Malus ist Malus. Musste ich auch erst googeln.

[8] Volpp, Kevin G et al, "A randomized, controlled trial of financial incentives for smoking cessation", New England Journal of Medicine, 7 (360) 2009. S.699.

Für ihr Verhalten, dass die Krankenkassen viel Geld kostet, geben die versicherten teilweise erhebliche Summen freiwillig aus eigener Tasche aus. Ganz offensichtlich ist das bei Rauchern so.

Wenn ein Raucher eine Schachtel Zigaretten pro Tag raucht und diese Schachtel 7€ kostet und ein Monat 30 Tage hat, sind das 210€/Monat. Die kann er jederzeit einsparen, wenn er mit dem Rauchen aufhört. Wenn aber die mögliche Einsparung von 210€/Monat den Raucher nicht zum Aufhören bringt, ist es unwahrscheinlich, dass ein kleiner zusätzlicher Bonus oder Malus das bewirkt.

Die Summen, um die es bei Bonus/Malus-Konzepten geht, haben also eine eher symbolische Höhe.

Die Verwechselung von Strafe und Rechteerwerb

Für die Zielgruppe kann es schwer sein, zwischen einer Strafe und dem Erwerb eines Rechts, sich nicht regelkonform zu verhalten, zu unterscheiden. Es gibt eine Studie über den Versuch eines Kindergartens, die Eltern zum pünktlicheren Abholen der Kinder zu motivieren, indem bei einer zu späten Abholung eine Strafe zu zahlen war.[9]

Das überraschende Ergebnis der Studie war, dass der Anteil der verspätet abgeholten Kinder nicht sank, sondern *stieg*. Die Eltern haben die Zahlung nicht als Strafe für Fehlverhalten interpretiert, sondern als Preis für ein bequemeres Verhalten. So lange man bezahlt, ist es OK, wenn man seine Kinder später abholt.

Nudging

Nudging (engl. für „Anstupsen“) bezeichnet ein Konzept, mit dem man versucht, die Entscheidung von Individuen subtil zu beeinflussen.[10]

Beispiele

Das Paradebeispiel ist die Reihenfolge und Platzierung von Speisen in Kantinen. Es lässt sich beobachten, dass dies die Wahl der Gäste beeinflusst. Wenn der Salat vor den Pommes angeboten wird und die Pommes in etwas erhöhter Position, wird mehr Salat und weniger Pommes gegessen. Dazu muß man die Pommes nicht von der Karte nehmen.

Ein anderes Beispiel ist die Frage, wie man „Standardentscheidungen“ definiert, z.B. bei der Organspende. Die aktuelle Regel ist eine opt-in Regel, d.h. der Standard ist, *kein* Organspender zu sein. Wenn man Organspender sein will, muss man etwas tun: Einen Spenderausweis ausfüllen. Viele Nachbarländer Deutschlands haben „ich bin Spender“ als Standardentscheidung definiert. Wenn man das nicht will, kann man einen „Nichtspenderausweis“ ausfüllen. Dieses „ich muss etwas tun, um nicht dazuzugehören“ wird opt-out genannt.

[9] Vgl. Bowles, Samuel. The moral economy, New Haven and London: Yale University Press, 2016, S. 4. Die Originalstudie ist Gneezy, Uri und Rustichini, Aldo, "A fine is a price", The journal of legal studies, 29 (1) 2000. S.1-17.

[10] Die zentrale Publikation zu diesem Ansatz ist Thaler, Richard H. und Sunstein, Cass R.. Nudge - wie man kluge Entscheidungen anstößt, Berlin: Ullstein, 2014. Für sein Werk hat Tha-ler 2017 den Nobelpreis erhalten.

Libertaristischer Paternalismus

Die Vertreter der Ansatzes bezeichnen Nudging als *libertaristischen Paternalismus*. Das ist ein Oxymoron, also eine art „schwarzer Schimmel“. Was sie damit zum Ausdruck bringen wollen ist, dass Nudging eine Art Bevormundung (Paternalismus) ist, den der Bevormundete aber leicht umgehen kann. Er hat die Freiheit (das ist der libertaristsiche Aspekt), sich anders zu verhalten und dieses andere Verhalten ist mit sehr geringem Aufwand verbunden.

Anwendungsgebiete

Das Konzept der Nudges funktioniert gut, wenn es um eine Entscheidung geht, bei der die meisten weitgehend indifferent sind. Wenn es mir ein wenig egal ist, ob ich Organspender bin oder nicht, mache ich das, was weniger Aufwand erfordert, indem ich die Standardentscheidung übernehme. Wenn ich starke Präferenzen für Pommes habe und keinen Salat mag, wird Nudging keine Verhaltensänderung bringen.

Oben habe ich eine Rauch-Entwöhnungsstudie erwähnt, die mit Boni gearbeitet hat. Aus ökonomischer Sicht würde man eine Erfolgsquote von Null prognostizieren, weil jeder Raucher schon über 200€ je Monat einsparen kann, wenn er aufhört, zu rauchen. Da sind die Boni lächerlich gering.

Vor dem Hintergrund dessen, was wir jetzt über Nudging wissen, sind wir sensibilisiert. Tatsächlich[11] war das Ergebnis der Studie

> *The incentive group had significantly higher rates of smoking cessation than did the information-only group 9 or 12 months after enrollment (14.7% vs. 5.0%, P<0.001) and 15 or 18 months after en-rollment (9.4% vs. 3.6%, P<0.001).*

In diesen Ergebnissen haben wir als Kontrollgruppe leider nicht die Raucher, die komplett selbständig mit dem Rauchen aufgehört haben, sondern nur die „information-only group“, die bereit waren, an einer Informationsveranstaltung teilzunehmen. Das relativiert den Erfolg ein wenig, denn 90,6% der Teilnehmer, die versucht haben, den Bonus einzustreichen, haben dann doch wieder mit dem Rauchen angefangen. Nimmt man dann noch die Absolutzahlen hinzu, war das Ergebnis, dass von anfänglich 436 Teilnehmern an dem Bonus-Programm letztlich 41 mit dem Rauchen aufhörten.

Wir müssen also annehmen, dass diese 41 Personen schon vorher einen so starken Wunsch hatten, mit dem Rauchen aufzuhören, dass der Nudge sie bei diesem Wunsch nur unterstützt hat, bzw. umgekehrt, dass der Nudge für 90,9% der Teilnehmer offenbar nicht stark genug war. Wir haben auch keine Informationen darüber, wie viele Raucher das Angebot, an der Studie teilzunehmen, *nicht* angenommen haben. Das wären dann die Raucher, die keine Motivation haben, ihr Verhalten zu ändern.

Verlustaversion

Wenn es um die Vermeidung eines Risikofaktors, wie z.B. Übergewicht geht, kann man die Anreize zu risikoarmem Verhalten sowohl als Bonus wie als Malus formulieren.

1. Normalgewichtige achten auf ihre Gesundheit und werden dafür belohnt.

[11] Hier können Sie übrigens den korrekten Einsatz von Konfidenzintervallen und kritischen Signifikanzniveaus sehen, über den wir in *Einheit Konfidenzintervalle* gesprochen haben.

2. Übergewichtige verursachen vermeidbare Kosten für die Versichertengemeinschaft und werden dafür bestraft.

Psychologen beobachten in Experimenten eine sogenannte *Verlustaversion.*[12] Ein Bonus von 500€ und ein Malus von 500€ werden nicht in gleicher Weise wahrgenommen, sondern der Malus wiegt schwerer.

Der sogenannte *Endowment-Effekt* beobachtet, dass der Besitz von etwas allein durch den Besitz nutzen stiftet. Wenn man etwas erworben hat, ist man nicht bereit, es zum Marktpreis wieder zu verkaufen, sondern erwartet einen höheren Preis. Nehmen wir an, ich schenke Ihnen einen Mensa-Kaffeebecher mit Aufdruck, der in der Mensa 5€ kostet. Wenn Sie diesen Becher jetzt wieder hergeben sollen, tun Sie das typischerweise nur, wenn man Ihnen 6€ zahlt.

Diese Überlegungen führen dazu, dass, wenn beides möglich ist, ein Malus besser funktionieren müsste als ein Bonus, weil der Versicherte ja etwas (Geld) hergeben muss, das er schon hat und nicht Geld bekommt, das er noch nicht hat.

F12.8 Zuammenfassung

Steuerungsinstrumente wie Selbstbehalt, Bonus- und Malussysteme erhalten große öffentliche Aufmerksamkeit, weil *alle* Versicherten von diesen Instrumenten betroffen sind.

Wir haben aber gesehen, dass das Einsparpotential der Instrumente sehr begrenzt ist, weil die finanziellen Anreize (positiv wie negativ) gering sind.

Sie *müssen* auch gering sein, damit die finanzielle Leistungsfähigkeit der Versicherten erhalten bleibt. Natürlich könnten wir eine absolute Abzugsfranchise von 100.000€ einführen. Die würde für 99,9% der Versicherten sehr gut wirken, aber dann brauchen wir auch keine Krankenversicherung mehr.

Weil die Versichertensteuerung nicht gut funktioniert, können wir diese Analyse kurz halten und jetzt Schluss machen. In der nächsten Einheit prüfen wir, ob die *Leistungserbringersteuerung* erfolgversprechender ist.

F12.9 Ehemalige Prüfungsaufgaben

Prävention als Wahlleistung

Der Gesetzgeber drängt die Krankenkassen dazu, mehr Präventionsleistungen als Wahlleistungen anzubieten.

1. Nach welchen Kriterien wird eine Kasse ihre Präventionsangebote auswäh-len?

2. Sind das die Kriterien, die aus gesamtgesellschaftlicher Sicht wichtig sind?

Die große Gesundheitsreform 2030

Die wichtigsten Punkte der (fiktiven) großen Gesundheitsreform 2030 sind:

- Das Neukundengeschäft der PKV wird eingestellt. Bisher PKV-Versicherte bleiben bis zu ihrem Lebensende dort versichert.

[12] Kahneman, Daniel. Schnelles Denken, langsames Denken, München: Siedler, 2011.

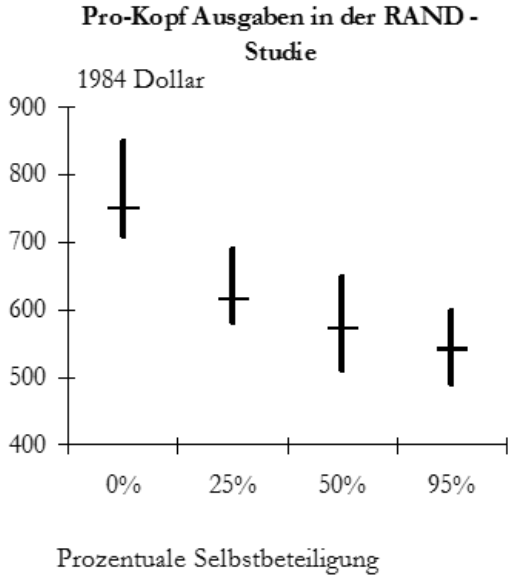
- Für alle GKV-Versicherten wird ein Selbstbehalt eingeführt.

Es gelten folgende Annahmen:

- Die Verteilung der Ausgaben entspricht der Aufteilung in drei Gruppen (Null-kosten, Mittelkosten, Hochkosten), mit der wir in der Veranstaltung immer gearbeitet haben.
- 2029 lagen die Durchschnittskosten bei 4.000€ pro Kopf
- Der Selbstbehalt reduziert die Durchschnittskosten der Mittelversicherten um 25%
-
- Die Durchschnittskosten der Hochkostenversicherten bleiben gleich.

Geben Sie die Änderung der durchschnittlichen Pro-Kopf Ausgaben für GKV-Versicherte in Euro an

Die RAND-Studie



Die RAND Corporation ist ein US-amerikanischer „Think Tank“. In den Jahren 1974-82 führte RAND ein sehr groß angelegtes Health Insurance Experiment (HIE) mit knapp 7.000 Teilnehmern durch, um der Regierung empirisch unterlegte Vorschläge für eine Gesundheitsreform unterbreiten zu können.

Das ist schon eine Weile her, aber immer noch das größte und wichtigste Experiment dieser Art.

Eine Frage, die man mit diesem Experiment beantworten wollte war, ob sich bei der Höhe der prozentualen Zuzahlung einen Einfluss auf die Inanspruchnahme medizinischer Leistungen beobachten lässt.

Dazu wurden vier Gruppen gebildet, die von keiner Zuzahlung (0%) bis zu 95% Zuzahlung reichten. Für Europäer klingt „95% Selbstbehalt“ komisch. Wenn wir gegen Ende der Veranstaltung das US-System näher betrachten, werden wir erkennen, daß es eine

Bevölkerungsgruppe gibt, für die das eine Verbesserung darstellt, weil sie *gar keine* Krankenversicherung haben, also 100% „zuzahlen“.

Die Grafik stellt die zentralen Ergebnisse mit 95% Konfidenzintervallen dar.

1. Welche möglichen Fehlinterpretationen sehen Sie?

2. Wie interpretieren Sie diese Grafik?

Prävention als Wahlleistung (Lösung)

- Gesundheitsökonomisch kann man Prävention als eine Investition in eine Person betrachten, die Kosten in der Zukunft vermeiden oder zumindest senken.
- Der gesamtgesellschaftliche Amortisationszeitraum ist also sehr lang.
- Wenn der Versicherte seine Kasse wechselt, dann hat die alte Kasse die Kosten der Prävention getragen, kommt aber nicht in den Genuss der Einsparungen.
- Eine Kasse wird Präventionsleistungen aber nur dann anbieten, wenn sie sich betriebswirtschaftlich rechnen.
  - Der medizinische Effekt stellt sich erst nach vielen Jahren ein und mit hoher Wahrscheinlichkeit bei einer anderen Kasse. Dieser Aspekt ist also irrelevant
  - Die Entscheidung über Präventionsleistungen hat also nichts mit dem medizinischen Effekt zu tun, sondern mit irgendetwas anderem, das sich betriebswirtschaftlich für die Kasse rechnet.
    * Dieser Effekt muss im Bereich Marketing oder Risikoselektion liegen.
    * Diese Effekte sind volkswirtschaftlich vollkommen irrelevant.

Die große Gesundheitsreform (Lösung)

Der Grund für die Annahmen über die PKV war, daß bei einer Einführung von Selbstbehalten in der GKV mit Wanderungsbewegungen freiwillig GKV-Versicherter hin zur PKV zu rechnen wäre. Welche Auswirkungen das auf die Kostenverteilung in der GKV hätte, wäre in der Größenordnung unklar. Durch die Annahme, daß die PKV für Neuzugänge gesperrt wird, habe ich dieses Problem eliminiert

1. Es gibt bei 10 Versicherten 4 Nullkostenversicherte, 5 Mittelkostenversicher-te und 1 Hochkostenversicherten. Bei 4.000€ Durchschnittskosten sind das 40.000€.
2. Der Hochkostenversicherte kostet 80% dieser Summe, d.h. 32.000
3. Die 5 Mittelkostenversicherten kosten zusammen 8.000€ bzw. 1.600€ pro Kopf
4. Wenn von dieser Summe 25% eingespart werden, sind das 0,25*1.600=400€ pro Kopf eines Mittelkostenversicherten
5. Bei 10 Versicherten und darunter 5 Mittekostenversicherten sind das 2.000€ bzw. 200€ pro Versicherten

Die RAND-Studie (Lösung)

Die Durchschnittsausgaben sind bei allen vier Gruppen bei höherer Selbstbeteiligung jeweils niedriger. Hieraus *könnte* man den Schluss ziehen, daß die Ausgaben umso geringer sind, je höher der Selbstbehalt. Das wäre eine Interpretation, die die Daten nicht zulassen, denn die drei Gruppen mit positiver Selbstbeteiligung (25%, 50% und 75%) haben zwar verschiedene Durchschnittswerte, aber überlappende Konfidenzintervalle, sind also nicht statistisch signifikant voneinander verschieden.

Statistisch signifikante Unterschiede gibt es nur zwischen „keine Selbstbeteiligung“ und „Selbstbeteiligung“.

Kapitel 13

Einzelleistungsvergütung

F13.1 Grundidee

Die Einzelleistungsvergütung (ELV) ähnelt einer klassischen Handwerkerrechnung. Ein Handwerker legt seine Abrechnungsstruktur offen. Bei einem Kostenvoranschlag schätzt er den Aufwand an Zeit und Material ab. Erhält er den Auftrag, dokumentiert er seinen Aufwand und stellt die Summe in Rechnung.

Diese Honorierungsform ergibt Sinn, wenn der Aufwand eines Auftrags schlecht abschätzbar ist, weil er kein Standardfall ist. Aus diesem Grund ist der Auftragnehmer nicht bereit, das Risiko einer Unterschätzung der Kosten zu tragen.

Im Bachelor-Skript finden Sie einen Überblick über die Regelungen, Probleme und Lösungsansätze für die ambulante Versorgung im deutschen System. Für uns sind diese Inhalt unter der Überschrift ELV insofern relevant, als dass die Finanzierung der ambulanten Versorgung zu einem Großteil über ELV abgewicket wird. Auf diese Inhalte greife ich im Folgenden zurück.

F13.2 Eigenschaften der ELV

Im Bachelor haben wir die ELV einfach als „so ist es“ hingenommen. Jetzt gehen wir einen Schritt weiter und überlegen uns, welche Eigenschaften diese Honorierungsform hat, d.h. welche Vor- und Nachteile.

Dazu möchte ich ein Raster mit drei Kriterien anwenden:

1. Effizienz in der Produktion
2. Qualität der Leistung
3. Leistungsvolumen

Wir schauen uns jetzt die ELV unter diesen Gesichtspunkten an und werden dieses Raster in den nächsten Einheiten in gleicher Weise auf die anderen Honorierungsformen anwenden.

Produktionseffizienz

Wenn wir den Leistungserbringer als Unternehmen betrachten und ein einfaches unternehmenstheoretisches Modell verwenden, dann ist der Gewinn die Differenz zwischen Umsatz und Kosten.

$G = U–K$ bzw.
$G = p * x–K$

Ich habe die erste Gleichung noch einmal beim Umsatz differenzierter dargestellt, weil uns das auf die Frage stößt, was denn das x ist.

x könnte der Zeitaufwand sein. Das ist aber in diesem Fall sehr umständlich. Bevor der Arzt die Blutdruckmanschette ansetzt, müsste er eine Stoppuhr starten, damit er messen kann, ob er in diesem konkreten Fall 90 oder 110 Sekunden braucht, um den Blutdruck zu messen.

Bei der einzelnen Leistung x kann es sich also nur um letztlich doch standardisierte Prozeduren handeln, das heißt, dass die Honorierung bei einem Patienten, der ein TShirt anhat und bei einem Patienten, der erst umständlich seinen Arm freimachen muss, die gleiche ist.

Wenn der Leistungserbringer nun eine Leistung x weniger effizient erbringt, als es ihm möglich ist, dann steigt K, was zu einem Sinken von G führt, weil p sich nicht nach K richtet. Das liegt nicht in seinem Interesse.

Daher wird die Produktionseffizienz in der ELV sehr hoch sein.

Qualität

Wenn $G = U–K$gilt, dann wirkt sich die Qualität der Leistung auf K aus, d.h. eine höhere Qualität ist mit höheren Kosten verbunden. Die ELV bietet dem Leistungserbringer keinen Anreiz, eine besonders hohe Qualität zu erbringen. Wenn er das tut, ist das eine Folge seines Berufsethos und nicht der Honorierungsform. Wir müssen daher damit rechnen, dass der Arzt ein Qualitätsniveau wählt, das „ausreichend“ ist und bei Kontrollen nicht beanstandet wird. Das bedeutet, dass wir bei der ELV mit einer eher mittleren Qualität rechnen können.

Leistungsvolumen

Wenn der Umsatz sich über den Preis und die Menge bildet, also
$U = p^*x$
ist und die variablen Stückkosten
$k_v < p$
sind, dann hat jede Leistung einen positiven Deckungsbeitrag und der Leistungserbringer wird im Rahmen einer Deckungsbeitragsmaximierung versuchen, seine Kapazitäten maximal auszulasten.

Auch hier müssen wir keine Raffgier des Leistungserbringers unterstellen, sondern nur einen positiven Zusammenhang zwischen intensiverer Behandlung und Gesundheitszustand des Versicherten. Selbst wenn der Gesundheitszustand nur unterproportional mit der Behandlungsintensität steigt, nutzt die Mehrbehandlung dem Patienten.

Wenn man die Kosten ignoriert, ist das *für Arzt und Patient* eine win-win Situation. Das bedeutet, dass wir mit einem insgesamt hohen Leistungsvolumen rechnen müssen.

Perspektive der Bewertung

Bis hierhin haben wir nur versucht, Eigenschaften des Konzepts ELV zu beschreiben. Wenn wir aber (in einem nächsten Schritt) andere Honorierungskonzepte mit der ELV vergleichen wollen, dann müssen wir diese Kriterien *bewerten*. Eine solche Bewertung setzt aber eine Perspektive voraus. Die ist da, auch wenn wir das gar nicht merken. Ein Beispiel:

Im Jahr 2021 hat Apple knapp $100 Mrd. Gewinn erzielt. Das ist doch toll, oder? Apple ist ein erfolgreiches Unternehmen, das gute Gewinne einfährt. Das kann man so sehen, wenn man die Perspektive z.B. eines Aktionärs einnimmt. Aus Sicht eines Kunden bedeutet das aber, das er ein überteuertes Produkt gekauft hat, weil in dem Preis, den er bezahlt hat, noch ein saftiger Gewinn für Apple enthalten war. Das hat er freiwillig getan. Niemand hat ihn gezwungen und er war offensichtlich der Meinung, dass das Produkt diesen überhöhten Preis trotzdem wert war. Nichtsdestotrotz ist für ihn „mehr Gewinn“ nicht positiv.[1] Offensichtlich beurteilen Aktionär und Kunde die gleiche Größe (Gewinn) unterschiedlich. Wenn wir die Höhe des Gewinns bewerten wollen, müssen wir uns daher für eine Perspektive entscheiden.

Bei den Kriterien, um die es uns in dieser (und den folgenden Einheiten) geht, ist es ganz ähnlich. Beim Kriterium „Leistungsvolumen“ haben wir das schon angeschnitten. Wir haben gesehen, dass der Arzt einen Anreiz zu einem möglichst großen Leistungsvolumen hat, weil das für Ihn mehr Einkommen bedeutet und er gleichzeitig seine Patienten besser behandeln kann. Er findet das aus beiden Gründen gut. Der Patient findet das vermutlich auch gut, weil er von der intensiven Behandlung profitiert. Die Krankenkasse findet das nicht so gut, weil sie die Leistungen ja bezahlen müssen und sich dann anschließend die Klagen der Versicherten anhören müssen, dass die Prämien so hoch sind.

Je nachdem, ob wir uns den Hut eines Leistungserbringers, eines chronisch Kranken, eines Gesunden, eines Kassenvertreters, ... aufsetzen, beurteilen wir die ELV anders. Diese Veranstaltung heißt *Gesundheitsökonomie*. Daher ist unsere Sicht nicht die eines Leistungserbringers, sondern eine *gesamtgesellschaftliche*. Diese Perspektive hat eine große Schnittmenge mit der Sicht der Krankenkassen, ist aber nicht mit ihr identisch. In *Einheit Selbstbehalte und Boni* haben wir gesehen, dass Kassen über Selbstbehalte oder Boni ihre eigenen Interessen verfolgen und versuchen, ihre Risikostruktur zu verbessern. Aus Kassensicht nachvollziehbar. Gesamtgesellschaftlich unerwünscht.

[1] „Bitte Apple, setzt eure Preise noch höher, damit ich noch mehr zahle und ihr mehr Gewinn macht“.

Kapitel 14

Pauschalhonorierung

F14.1 Der rote Faden

Da kommen wir her: Wir befassen uns mit der Frage, inwieweit man das Verhalten der Leistungserbringer über die Honorierungsform stärker in Richtung einer höheren Effizienz beeinflussen kann. In der letzten Einheit haben wir uns die Einzelleistungsvergütung angesehen.

So geht es weiter: Die nächste Honorierungsform auf unserer Liste ist die Honorierung über Pauschalen, die wir in Deutschland vor allem aus dem Krankenhausbereich kennen.

F14.2 Konzept

Eine Honorierung über Pauschalen bedeutet, dass der Anbieter für eine definierte Gesamtleistung ein definiertes Honorar bekommt, unabhängig davon, auf welche Weise er die Leistung erbringt und welcher tatsächliche Aufwand hinter der Leistungserbringung steckt. Liegen seine Kosten unterhalb des Honorars, kann er den Überschuss behalten. Im umgekehrten Fall muss er den Verlust tragen.

Abgrenzung zur ELV

Die Abgrenzung zur ELV ist schwammig, denn wir haben gesehen, dass die Leistung „Blutdruckmessen“ in der ELV ganz genau betrachtet auch eine Art Minipauschale ist, weil diese Leistung standardisiert honoriert wird und sich, genau genommen, aus mehreren Teilleistungen zusammensetzt. Der wesentliche Unterschied ist aber, das nicht die Prozessschritte honoriert werden, sondern das Ergebnis.

Kopf- und Fallpauschalen

In diesem Zusammenhang ist zwischen Kopf- und Fallpauschalen zu unterscheiden. Eine *Fallpauschale* liegt dann vor, wenn ein Leistungserbringer einen Betrag für eine definierte Behandlung erhält, z.B. eine Blinddarm-OP oder eine Spontangeburt.

Von einer *Kopfpauschale* spricht man, wenn ein Leistungserbringer eine Pauschale für eine Person (einen „Kopf“) erhält und sich verpflichtet, dafür die komplette medizinische Versorgung dieser Person zu erbringen. I*m* Bachelor haben Sie die Regelung des *Sicherstellungsauftrags* der KV kennengelernt. Das ist eine Art Kopfpauschale, mit dem

Unterschied, dass diese Kopfpauschale an eine Umverteilungsstelle (die KVen) gezahlt werden, die diese Pauschale in Form einer Einzelleistungsvergütung an die Vertragsärzte weitergeben.

F14.3 Rückgriff auf Bachelor - Einheiten

Ein großer Teil der Leistungen, die in deutschen Krankenhäusern erbracht werden, werden über Fallpauschalen (DRG) abgerechnet. Das war Inhalt des Bachelors. Wenn Sie diese Inhalte nicht mehr präsent haben (oder – anderer Studiengang – nie hatten) müssen Sie diese Einheit durcharbeiten. Dort steht alles, was wir für die weitere Analyse benötigen und was als prüfungsrelevantes Wissen von Ihnen erwartet wird.

F14.4 Ein Excel-Modell der Fallpauschalendynamik

Kostenverteilung

Im Bachelor haben wir uns kurz mit dem Umstand befasst, dass es effizientere und weniger effiziente Anbieter gibt. Nehmen wir an, es gibt insgesamt 1.900 Krankenhäuser, die bei den Kosten für einen Fall so aussehen:

	n	€
wenig effizient	450	2.000
Durchschnitt	1.000	1.500
hoch effizient	450	1.000

Diese Verteilung kann man noch mit mehr Gruppen bestücken, aber es geht uns nicht um konkrete Zahlen für einen konkreten Fall, sondern um die Dynamik der Fallpauschalen. Für diese Verteilung habe ich eine Modellrechnung erstellt, bei der am Anfang alle Häuser gleich viele Fälle haben. Auch das kann man variieren, muss man für unsere Zwecke nicht.

Die Modellrechnung

In Zelle F6 finden Sie die Fallzahl von insgesamt 380.000, die sich gleichmäßig auf die 1.900 Häuser verteilen.

Nun soll eine Fallpauschale eingeführt werden, die sich über die durchschnittlichen Fallkosten bestimmt, die in F10 stehen. Das bedeutet, dass die günstigen Häuser einen Deckungsbeitrag von 500€ erzielen (C12) und die teuren Häuser einen negativen DB von 500€. Die durchschnittlichen Häuser erzielen keinen DB.

	A	B	C	D	E	F
1						
2	**Fallkosten**		**1.000 €**	**1.500 €**	**2.000 €**	**Σ**
3	**Szenario "Heute"**					
4	Häuser					
5		Anzahl	450	1.000	450	1.900
6		Fälle insgesamt	90.000	200.000	90.000	380.000
7		Ø Fallzahl	200	200	200	
8	Kosten					
9		insgesamt	90.000.000 €	300.000.000 €	180.000.000 €	570.000.000 €
10		Ø				1.500 €
11	Deckungsbeitrag					
12		je Fall	500 €	- €	- 500 €	
13		je Haus	100.000 €	- €	- 100.000 €	
14	**Szenario "Morgen"**					
15	Häuser					
16		Anzahl	600	1.100	100	1.800
17		Fälle insgesamt	150.000	220.000	10.000	380.000
18		Ø Fallzahl	250	200	100	
19	Kosten					
20		insgesamt	150.000.000 €	330.000.000 €	20.000.000 €	500.000.000 €
21		Ø				1.316 €
22	Deckungsbeitrag					
23		je Fall	316 €	- 184 €	- 684 €	
24		je Haus	78.947 €	- 36.842 €	- 68.421 €	

Jetzt ist absehbar, was passiert:

- Die *teuren* Häuser haben die Option vom Markt zu gehen oder die Kosten zu senken. In Zeile 16 sehen Sie, dass es noch 100 Häuser gibt, die von den Negativdeckungsbeiträgen nicht heruntergekommen sind. 350 Häuser sind besser geworden oder nicht mehr am Markt.

- Die Zahl der *günstigen* Häuser ist größer geworden, weil alle Durchschnittshäuser jetzt nach der Einführung der Pauschale gesehen haben, dass sie DB erzielen können, wenn sie ihre Kosten senken. Alle haben das gesehen, 150 haben es geschafft und sind jetzt auch günstige Häuser.

- Gleichzeitig haben die günstigen Häuser ihre *Kapazitäten* ausgebaut, denn jeder zusätzliche Fall bringt DB.

- Die Versorgung hat sich somit insgesamt zu den günstigen Häusern verschoben. Statt bisher 90.000 Fällen werden jetzt 150.000 Fälle zu niedrigen Kosten behandelt.

- Die Gesamtzahl der Häuser hat sich um 100 verringert.

Als Globalkennzahl für die Wirkung kann man F20, die neuen Durchschnittskosten, verwenden. Sie sind um 184€ gefallen. Bitte beachten Sie, dass in dieser niedrigeren Zahl immer noch 316€ DB für die günstigen Häuser stecken.

Die zweite Runde Das Szenario „morgen“ zeigt, dass die Wirkung der Pauschale noch nicht zum Ende gekommen ist. Dadurch, dass die Durchschnittskosten gesunken sind – und damit die Pauschale – geraten jetzt die bisherigen Durchschnittshäuser unter Druck, weil sie jetzt (D23) 184€ Verlust je Fall machen.

Das ist besonders bitter für die bisher teuren Häuser, die sich am Durchschnitt orientiert haben und 500€ eingespart haben. Damit haben sie ihre Felle (Fälle) aber nicht ins Trockene gebracht, sondern schreiben, weil die anderen Häuser ebenfalls günstiger geworden sind, immer noch rote Zahlen.

Durch diesen Effekt kommen jetzt auch die Häuser unter finanziellen Druck, die bisher im Durchschnitt lagen. Auch diese Häuser müssen jetzt entscheiden, ob sie ihre Kosten hinreichend senken können, den Markt verlassen oder zunächst weiterhin Verluste schreiben – in der Hoffnung mittelfristig kostendeckend zu werden.

F14.5 Eigenschaften der Pauschalhonorierung

Analog zur Bewertung der ELV wollen wir nun die Pauschale nach den Kriterien Effizienz, Qualität und Sparsamkeit untersuchen. Dazu möchte ich auf das Krankenhausbeispiel zurückgreifen.

Produktionseffizienz

Die Anbieter haben einen Anreiz, die Leistungen, die erbracht werden, effizient zu erbringen, da vermeidbarer Ressourcenaufwand nicht zu einer höheren Pauschale führt, sondern zu einem niedrigeren DB.

Qualität

Die Qualität der Pauschalhonorierung ist schwer einzuschätzen. Der Begriff des *Drehtüreffekts* bezeichnet den Fall, dass die Qualität einer über eine Pauschale abgerechnete Leistung so schlecht ist, dass der Patient das Krankenhaus verlässt, und sehr schnell wieder eingeliefert werden muss, weil Komplikationen auftreten. Ein verwandter Begriff ist der der „blutigen Entlassung“.

Diese Strategie der schlechten Qualität funktioniert nur dann, wenn die Kontrolle, ob der Fall in der vereinbarten Qualität auch wirklich abgearbeitet worden ist, unzureichend ist. Funktioniert diese Kontrolle, liegt es im Eigeninteresse des Leistungserbringers, die definierte Qualität zu erbringen, weil er sonst die Nachbesserungen und Komplikationen aus der Fallpauschale, die er bereits erhalten hat, zahlen muss, weil der Fall ja nicht abgeschlossen ist. Dem Qualitätscontrolling der Kostenträger kommt daher eine Schlüsselrolle zu.

Wenn wir aber (erst einmal) annehmen, dass dieses Controlling funktioniert, können wir von einer hohen Qualität ausgehen.

Sparsamkeit

Auch das Kriterium „Sparsamkeit“ ist schwer einzuschätzen. In unserer Betrachtung der Fallpauschalendynamik im Krankenhaus gab es den Anreiz für die effizienten Häuser, ihre Kapazitäten auszuweiten. Das bedeutet aber, dass man diesen Häusern mehr Geld zahlt, als sie eigentlich brauchen. Sie erzielen ja Überschüsse mit jedem Fall. Theoretisch ist das ineffizient. Dieser finanzielle Anreiz wird allerdings umso kleiner, je länger die Fallpauschalen den aktuellen Durchschnittskosten angepasst werden. Gleichzeitig sinken ja, im Vergleich zum Ausgangsszenario die Kosten. Setzen wir also bei diesem Kriterium einen mittleren Wert an.

F14.6 Ehemalige Prüfungsaufgaben

Zentral festgelegte DRG-Pauschalen

Die Pauschalen im Krankenhausbereich werden zentral festgelegt, unterscheiden sich also zwischen den Krankenkassen nicht.

- Erhöht das das Wettbewerbsniveau bei den Krankenkassen?
- Mindert das das Wettbewerbsniveau bei den Krankenkassen?
- Kann man das nicht eindeutig sagen?

Zentral festgelegte DRG-Pauschalen (Lösung)

Die Höhe der Pauschalen ist kein Wettbewerbsfaktor der Kassen, weil sich eine Kasse gegenüber den Konkurrenten nicht durch niedrigere Kosten profilieren können. Größere Kassen können ihre Marktmacht nicht ausspielen. Insofern sinkt das Wettbewerbsniveau.

Nutznießer sind die kleinen Kassen und die Krankenhäuser, weil die großen Kassen sie nicht preislich ausquetschen können. Wenn Kassen Wettbewerbsvorteile haben wollen, müssen sie dies auf anderen Feldern erreichen, z.B. bei den Verwaltungskosten. Hier steigt das Wettbewerbsniveau.

Kapitel 15

Pauschalen und Risiko

F15.1 Der rote Faden

Da kommen wir her: In der letzten Einheit haben wir uns mit den Pauschalen ein Honorierungskonzept angesehen, dass unter Effizienzaspekten sehr attraktiv aussieht. Wenn wir ELV mit Pauschalen vergleichen, spricht aus gesamtwirtschaftlicher Sicht eigentlich alles dafür, möglichst viele Leistungen über Pauschalen zu honorieren. Ineffiziente Leistungserbringer werden dadurch unter Druck gesetzt und müssen eventuell aus dem Markt ausscheiden.

So geht es weiter: Können diese Leistungserbringer erwarten, dass andere ihre Ineffizienz durch höhere Preise bezahlen? Dafür muss es ziemlich gute Gründe geben. Die Anbieter führen als Hauptargument die wohnortnahe Versorgung an. Das ist ein Argument, dass uns zwar nicht jetzt, aber in *Einheit Die Krankenhausreform in Dänemark* beschäftigen wird.

In dieser Einheit wollen wir einen anderen Aspekt beleuchten. Wir schütten etwas Wasser in den Wein und schauen uns die finanziellen Nachteile diesese Konzepts an, die man vermutlich erst auf den zweiten Blick erkennt

F15.2 Fall- vs Kopfpauschalen

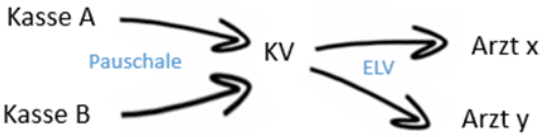
Drehtüreffekte

In der letzten Einheit haben wir uns die Fallpauschalen angesehen, die auch in der stationären Versorgung eingesetzt werden. Ein Problem, das wir dort ein wenig unter den Teppich gekehrt haben, war die Strategie, an der Qualität zu sparen, um die Kosten zu senken. Das haben wir unter dem Begriff *Drehtüreffekt* besprochen. Wenn die Versicherer nicht merken, dass ein zweiter Krankenhausaufenthalt von Dave eine Folge der schlechten Qualität des ersten Krankenhausaufenthalts war, kann das Krankenhaus *zwei* Fälle abrechnen, statt *einen* Fall ordentlich abzuarbeiten. Betriebswirtschaftlich ist das für das Krankenhaus eine sinnvolle Strategie. Wenn man nicht erwischt wird.

Sektoral getrennte Budgets

Ein ähnlicher Anreiz besteht bei getrennten Budgets. Aus dem Bachelor kennen wir die Funktion der KVen, die von den Kassen eine Kopfpauschale erhalten, die sie bei ELV an

die Vertragsärzte auszahlen.Wir haben es also mit einer zweistufigen Honorierung zu tun.



Mit dieser Pauschale übernimmt die KV den *ambulanten Sicherstellungsauftrag*. Das Problem ist, dass an den Rändern nicht ganz klar ist, worin dieser *ambulante* Sicherstellungsauftrag besteht. Das wird am Beispiel der Notfallambulanzen in den Krankenhäusers deutlich. Wenn es den niedergelassenen Ärzten gelingt, die an dieser Stelle zu leistende Arbeit stärker in den Krankenhausbereich zu verschieben, bekommt man das gleiche Geld für weniger Leistung. Die Krankenhäuser vertreten die nachvollziehbare Position, dass sie an dieser Stelle Arbeit übernehmen, die eigentlich in den ambulanten Bereich gehört und man deshalb auch ein Anrecht auf ein Stück des ambulanten Honorarkuchens hat.

Gatekeeping und Fundholding

In diesem Kontext sind zwei englische Begriffe wichtig.

Als Torwächter (gatekeeper) wird ein Arzt bezeichnet, der den Zugang des Versicherten zu ambulanten Spezialisten und Krankenhäusern regelt. In *Einheit Das Gesundheitssystem in Großbritannien* werden wir sehen, wie Großbritannien dies organisiert. Hier nur so viel: Die erste Anlaufstelle eines Versicherten ist sein Allgemeinarzt, der ihn entweder selbst behandelt oder innerhalb des Systems weiterleitet. In Notfällen kann der Versicherte natürlich direkt ins Krankenhaus gehen, wird aber bei „Spaß-Selbsteinweisungen“ sanktioniert.

Wenn die Geschichte an dieser Stelle zu Ende wäre, wäre dieses Konzept nicht besonders klug, weil die ökonomische Interessenlage des Gatekeepers darin besteht, mit dem Versicherten möglichst wenig Arbeit zu haben und ihn schnell weiterzuleiten.

Aus diesem Grund ist gatekeeping im Regelfall mit *fundholding* gekoppelt. Der Allgemeinarzt bekommt eine Pauschale für jeden Versicherten, der bei ihm eingeschrieben ist. Einen Großteil „seiner“ Versicherten bekommt er überhaupt nicht zu Gesicht. Die Versicherten *müssen* sich bei einem Gatekeeper einschreiben. Sie können den Gatekeeper wechseln. Nicht im Vierzehntagetakt, aber z.B. quartalsweise.

Die Pauschale, von der ich gerade gesprochen habe, deckt aber (anders als bei den KVen) nicht nur den ambulanten Bereich ab, sondern die *komplette* medizinische Versorgung und ist entsprechend höher

Make or Buy

In der *Einführung in die Ökonomie* haben wir uns mit der in der BWL zentralen Frage des „make or buy“ befasst. Fundholding bedeutet, dass der Gatekeeper auch Facharztbesuche und Krankenhausaufenthalte aus der Kopfpauschale bezahlen muß. Die Fachkollegen können ihm dann eine Rechnung stellen, die der Hausarzt bezahlt. An dieser Stelle sind zwei Aspekte relevant:

1. Zum einen ist der Hausarzt in der bestmöglichen Position, um als Außenstehender zu erkennen, ob der Fachkollege überflüssige Leistungen abrechnen will. Er hat die (medizinische) Kompetenz und den Einblick in den konkreten Fall.

2. Zum anderen wird der Hausarzt abwägen, ob er einen komplexeren Fall selbst behandeln will oder den Patienten überweist. Er wird die effizientere Alternative wählen, da er kein Interesse daran hat, an Fällen herumzupfuschen, für die er eigentlich nicht kompetent ist, weil ihm das vielleicht die Rechnung des Kollegen erspart, dafür aber seine Zeit und sein Geld kostet.

Für den Patienten bedeutet das allerdings, dass er Fachärzte nicht direkt, sondern nur nach Überweisung durch den Hausarzt aufsuchen kann, denn man kann natürlich von dem Hausarzt nicht verlangen, das finanzielle Risiko für Patienten zu tragen, die sofort und vielleicht überflüssigerweise einen (teuren) Facharzt aufsuchen.

Ein Grund für die Zurückhaltung gegenüber dieser Idee in Deutschland ist, dass der stationäre Sektor komplett getrennt organisiert und finanziert wird. Die Versicherung müsste bei einer Kopfpauschale damit rechnen, dass die Hausärzte ihre Patienten sehr schnell in ein Krankenhaus überweisen, weil die dort anfallenden Kosten nicht zu Lasten der Kopfpauschale gehen würden. Wenn ein Hausarzt also als Fundholder fungieren soll, wäre es angebracht, ihm die Verantwortung über das *gesamte* Kostengeschehen zu übertragen.

Dann steigt allerdings das finanzielle Risiko für den Hausarzt und muß durch ei-nen höheren Sicherheitszuschlag kompensiert werden. Gleichzeitig nehmen die Möglichkeiten des Hausarztes ab, die Versorgung kostengünstig bei akzeptabler Qualität zu gestalten, weil es nicht mehr nur um seine eigenen Leistungen geht und die von niedergelassenen Fachkollegen, sondern um den sehr großen Kostenblock der Krankenhauskosten.

F15.3 Unternehmerisches Risiko des Leistungserbringers

Die eingeschriebenen Versicherten als Stichprobe

In Einheit Konfidenzintervalle haben wir uns mit dem Problem befasst, dass wir Studien nur an einer relativ kleinen Stichprobe durchführen können und eine Übertragung der Stichprobenergebnisse auf die Grundgesamtheit nur eingeschränkt möglich ist. Beim Fundholding haben wir es mit dem gleichen Problem, nur quasi „andersrum“ zu tun.

Wir wissen aus systemweiten Daten, wie die Grundgesamtheit (d.h. die Kostenverteilung unter den Versicherten) verteilt ist. Wir arbeiten immer mit Regel aus dem Bachelor: 10% der Versicherten verursachen 80% der Kosten usw. Der Gatekeeper weiß, dass das für die Grundgesamtheit gilt, aber ob das bei den Versicherten, die sich bei ihm eingeschrieben haben, auch so ist, weiß er nicht.

Überschaden

Eine typische Praxisgröße für Fundholder umfasst 4.000 Versicherte. Nach unserer 10:80-Regel tauchen 40%, d.h. 1.600 Versicherte in einem Quartal überhaupt nicht auf und 400 Patienten sind extrem teuer.

Rechnen wir das kurz mit GKV-Zahlen aus dem Jahr 2022 durch. Die Durchschnittskosten pro Kopf lagen bei 3.729€.[1]Dann kosten 10 Versicherte 37.290 und der eine Hochkostenpatient 29.832€.

[1]https://www.gkv-spitzenverband.de/ krankenversicherung/ kv_grundprinzipien/ finanzierung/ finanzierung.jsp

Unser Fundholder kann bei 4.000 Versicherten statistisch mit 400 Hochkostenpatienten rechnen. Aus unseren Überlegungen zur Stichprobentheorie schwant es uns, dass die Wahrscheinlichkeit, dass es *genau* 400 Hochkostenversicherte sind, ziemlich gering ist. Etwas ungenau ist diese Wahrscheinlichkeit Null und die Wahrscheinlichkeit, dass es mehr als 400 Hochkostenpatienten gibt 50% und die, das es weniger gibt, auch 50%.

Wenn es, sagen wir, 402 Hochkostenpatienten gibt, spricht die Versicherungsmathematik von 2 Überschäden. Bei 396 Schäden wären es 4 Unterschäden.

Wenn wir dem Fundholder eine Kopfpauschale von 3.729€ zahlen, wird das also in 50% aller Fälle dazu führen, dass es mit 50% Wahrscheinlichkeit einen Überschaden geben wird, der den Fundholder richtig Geld kosten wird.

Sicherheitszuschlag

Im Bachelor haben wir ein wenig Versicherungsmathematik betrieben. Eine Größe, mit der wir damals hantiert haben, war der Sicherheitszuschlag, für den wir uns einfach einen Betrag für unsere Kalkulation ausgedacht haben. Jetzt ist der Zeitpunkt gekommen, über die Höhe dieses Zuschlags nachzudenken.

Spielen wir einmal kurz folgendes Szenario durch: Die Versicherung zahlt dem Fundholder $\frac{29.832€}{4.000} = 7,46€$ je Versicherten als Sicherheitszuschlag. Zusätzlich zu den 3.729€. Der Fundholder guckt jetzt nicht mehr ganz so ängstlich wie bisher, aber definitiv nicht entspannt, denn dieser Sicherheitszuschlag deckt exakt *einen* Überschaden ab. Wenn es zwei oder mehr Überschäden gibt, steht er wieder im Regen.

Wir können ohne viel Aufwand ausrechnen, wann der Fundraiser definitiv gut schlafen kann. Das schlimmste, was ihm passieren kann ist, dass alle 4.000 Versicherte Hochkostenpatienten sind. Wir können auch ausrechnen, wie wahrscheinlich das (nach unserer 80:10-Regel) ist: $0,1^{4000}$. Mein Taschenrechner rechnet nicht genau genug und gibt mir eine 0 als Wahrscheinlichkeit aus. Tatsächlich ist die Wahrscheinlichkeit minimal höher als Null.

Wenn der Fundholder dieses Risiko nicht tragen will, muss man ihm als Kopfpauschale die 29.832€ zahlen, was für die Krankenkassen natürlich sehr unattraktiv ist. A propos: Die Sicht der Krankenkassen haben wir ein wenig aus den Augen verloren. Bisher ist nicht klar, warum die sich auf die Idee des Fundholding einlassen sollten. Das holen wir jetzt nach.

F15.4 Attraktivität für die Versicherer

Grundvoraussetzung: Es muss billiger werden

In unserer Rechnung geben die Krankenkassen 3.729€ pro Kopf und Jahr aus. In den derzeitigen Honorierungsstrukturen. Damit die Kassen über einen Wechsel zum Fundholding-Konzept anfangen nachzudenken, muss am Ende eine Kopfpauschale stehen, die unter diesem Betrag liegt.

Notwendig: Einsparpotential

Damit die Rechnung für die Kassen aufgeht, muss es ein spürbares Einsparpotential geben. Nehmen wir einfach einmal an, 20% der medizinischen Kosten wären vermeidbar und nehmen wir an, der Fundholder könnte diese 20% identifizieren. Im bisherigen System

spart das niemand ein, weil niemand einen Anreiz hat, aber über „make or buy“ hätte der Fundholder einen Anreiz, den billigsten Weg zu finden, der auch „nichts tun“, also „komplett vermeiden“ sein könnte.

Diese 20% entsprechen einem Betrag von 745,80€ je Versicherten. Wenn der Fundholder im Durchschnitt diese Summe bei den Leistungen einspart, ist diese Summe die Manövriermasse zwischen Kasse und Fundholder

	Einsparung bei der Kasse
+	Sicherheitszuschlag
+	Bonus für den Fundholder
	745,80€

Bonus

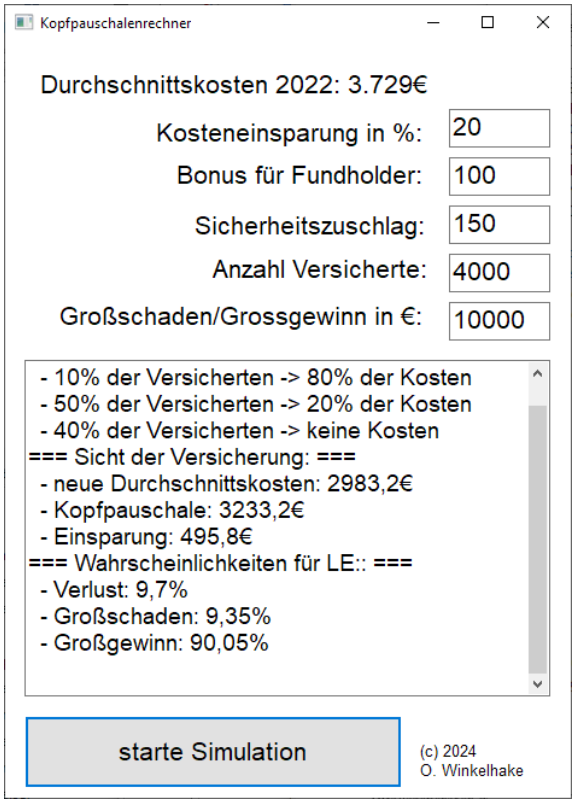
In meiner Rechnung habe ich über den Posen *Bonus für den Fundholder* noch nichts gesagt. Hinter diesem Posten verbirgt sich folgende Überlegung: Das Fundholding erzeugt zusätzlichen Aufwand, denn der Fundholder behandelt ja weiterhin die Fälle, die in seinen Kompetenzbereich fallen und organisiert und kontrolliert die Behandlung des Patienten bei anderen Leistungserbringern. Das kostet Zeit und Geld. Entweder er kann weniger Patienten behandeln oder er stellt Mitarbeiter ein, die ihn unterstützen (das wäre die britische Lösung). In jedem Fall muss man den Fundholder dafür honorieren. Diese Summe wird hier als *Bonus* bezeichnet.

F15.5 Das Simulationsmodell

Den Konfidenzintervallrechner aus *Einheit Konfidenzintervalle* habe ich in einen Kopfpauschaltenrechner umgebastelt. Ich simuliere mit einem Zufallsgenerator die 4.000 Versicherten und prüfe, ob die Zahlung der Kassen ausreicht, die Kosten des synthetischen Samples zu decken. Das mache ich 2.000 mal und zähle, in wie vielen Fällen der Fundholder einen Verlust machen würde.

Ausgangsszenario

Im folgenden Screenshot habe ich eine Kosteneinsparung von 20% angenommen, dem Fundholder eine Pauschale von 100€ gezahlt, sowie einen Sicherheitszuschlag von 150€.



Mit diesen Parametern liegt die Verlustwahrscheinlichkeit bei 9,7%. Mit dieser Wahrscheinlichket werden die Zahlungen der Kasse (den Bonus habe ich herausgerechnet) nicht ausreichen, die Versorgung der 4.000 Versicherten zu finanzieren. Trotz des Sicherheitszuschlags.

Um dieses Verlustszenario etwas differenzierter zu betrachten, habe ich die Kategorie des Großschadens bzw. Großverlusts definiert, denn wenn der Fundholder einen Verlust von 17,50€ macht, könnte man das ignorieren. Als (veränderbare) Schwelle für „spürbar viel Geld“ habe ich 10.000€ verwendet.

Mit 9,35% Wahrscheinlichkeit geht die Rechnung für den Fundholder spürbar nicht auf.

> *Fingerübung: Was ist mit der Differenz zu den 9,7% Verlustwahrscheinlichkeit? Was ist das?*

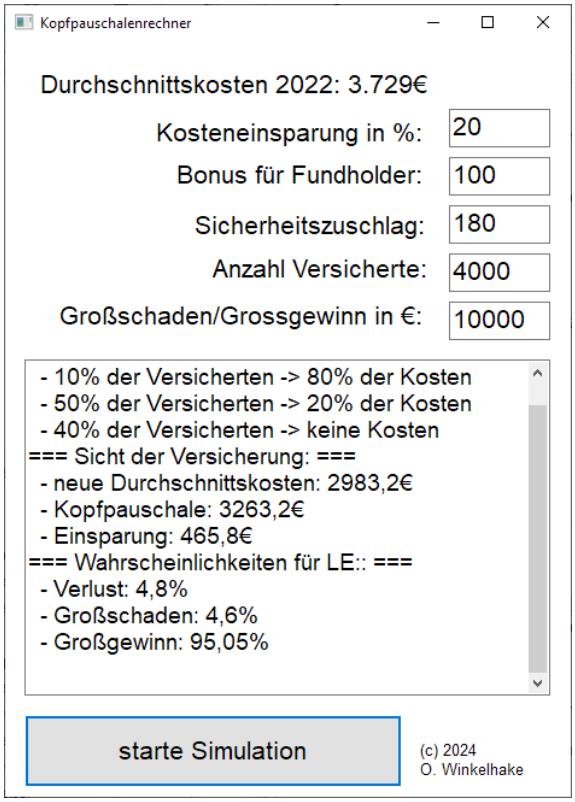
Auf der anderen Seite führt der Sicherheitszuschlag von 150€ pro Kopf mit einer Wahrscheinlichkeit von 90,05% zu einem spürbaren Gewinn, weil mehr als 10.000€ übrig bleiben.

Das kann man jetzt noch weiter differenzieren, aber die Botschaft ist klar: Der Fundholder trägt ein finanzielles Risiko, das nichts mit seinen Bemühungen zu tun hat, Kosten einzusparen. Das tut er ja. Bei jedem Patienten 20%. Er kann halt nur Pech bei den Überschäden haben.

In diesem Szenario spart die Kasse 495,80€ ein. Die theoretische Untergrenze bei den Kosten wären 2.983,20€, aber der Fundholder will für seinen Zusatzaufwand bezahlt und für sein Risiko kopensiert werden, was zusammen 250€ kostet.

Szenario: 5% Verlustwahrscheinlichkeit

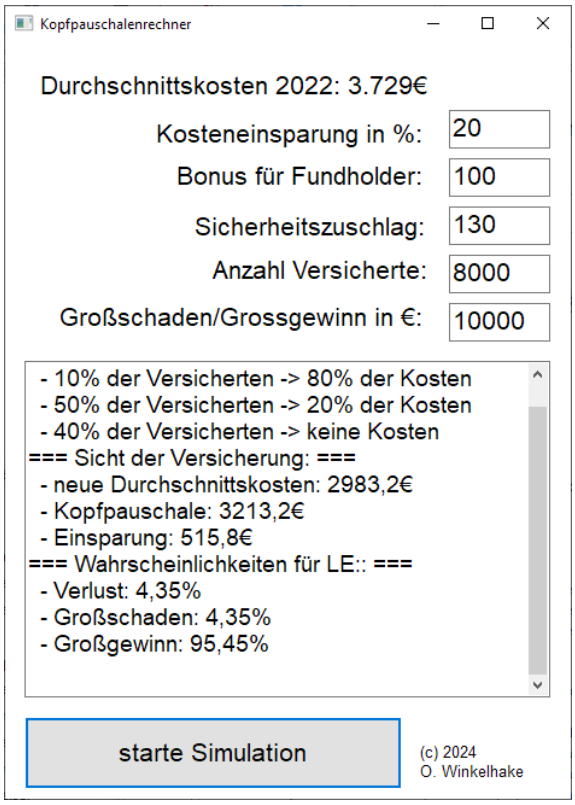
Die Verlustwahrscheinlichkeit im Ausgangsszenario ist relativ hoch. Wir können jetzt eine Variante finden, in der die Verlustwahrscheinlichkeit bei etwa 5% liegt. Nach etwas ausprobieren (mein Programm hat leider keine Solver-Funktion) bin ich auf die folgende Kombination gekommen. Mit einem Sicherheitszuschlag von 180€ kann die Kasse die Verlustwahrscheinlichkeit wie beschrieben drücken. Die Einsparungen sinken dementsprechend.



Szenario: Großpraxis

Im Kontext der Stichprobentheorie in *Einheit Konfidenzintervalle* haben wir gesehen, dass die Zuverlässigkeit von Stichprobenergebnissen mit der Stichprobengröße zunimmt, weil das Gesetz der großen Zahl besser wirkt. Diese Überlegung kann man auf unsere Fundholder übertragen. Je mehr Versicherte der Fundholder hat, umso stärker müsste sich die Kostenverteilung der 10:80-Regel annähern.

Das Äquivalent zur größeren Stichprobe ist hier eine größere Praxis. Nehmen wir einfach an, zwei bisher selbständige Fundholder schließen sich zu einer gemeinsamen Großpraxis mit 8.000 Versicherten zusammen und machen gemeinsame Kasse. Um dieses Szenario durchzumodellieren, muss ich nur die Zahl der Versicherten erhöhen und kann dann am Sicherheitszuschlag herumdrehen, bis wieder eine Verlustwahrscheinlichkeit von etwa 5% herauskommt.



Wenn man die Szenarien vergleicht, sieht man, dass die Großpraxis das gleiche Verlustrisiko bei einem um 50€ niedrigeren Sicherheitszuschlag tragen kann. Das hat nichts mit einer effizienteren Praxisführung zu tun (die es vielleicht zusätzlich auch noch gibt), sondern ausschließlich mit dem kleineren Stichprobenfehler.

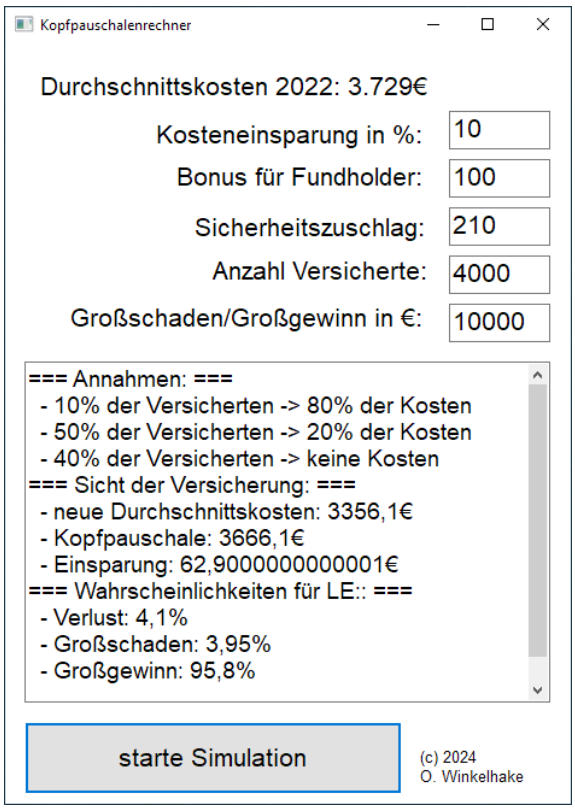
Szenario: Kleinvieh

Bevor Sie jetzt in großen Jubel ausbrechen, möchte ich ein letztes Szenario mit Ihnen durchgehen. Die respektablen Ergebnisse, die wir bisher ermittelt haben, sind das Ergebnis davon, dass wir „nette“ Zahlen in das Modell eingefüttert haben und dementsprechend nette Ergebnisse bekommen haben. Erinnern Sie sich: Wir haben hier kein Prognosemodell vor der Nase, sondern eine *Modellrechnung*.

Eine, meiner Meinung nach, sehr optimistische Annahme war, dass der Einsatz des Fundholders 20% der bisherigen Kosten einspart (also knapp 750€) uns aber nur 280€

(Bonus + Sicherheitszuschlag) kostet. Diese 20% an vermeidbaren Kosten sind eine Hausnummer, die hin und wieder in der politischen Diskussion auftauchen und nie in irgendeiner Form empirisch belegt werden.

Unsere Rechnungen gehen erstens davon aus, dass diese Spekulation zutrifft und zweitens, dass die Fundholder in der Lage/Willens sind, diese 20% auch tatsächlich komplett einzusparen. Eine Denkfigur, die wir in unseren Modellrechnungen immer wieder durchgespielt haben ist, die Parameter zu verändern und zu schauen, ob eine Entscheidung irgendwann kippt. Das ist immer die große Stunde des Solvers in Excel gewesen. Hier können wir einmal einen etwas weniger optimistischen Wert einfüttern und schauen, was passiert. Eine Simulation für eine Kosteneinsparung von 10% sieht so aus:



Die Einsparung aus Sicht der Versicherung liegt jetzt bei 62,90€, also deutlich niedriger. Das hat zwei Gründe: Zum einen können die Fundholder nicht mehr so viel einsparen, zum anderen muss der Sicherheitszuschlag nach oben gehen, weil die Überschäden ja teurer sind.

Wenn ich an diesem Rädchen noch weiter drehe und die Einsparung auf 8% senke, zahlen die Kassen mehr als vorher.

Die Frage, ob Fundholding eine gute Idee ist, hängt also ganz stark davon ab, wie viel Verschwendung im System vorliegt. Wir sind durch unsere Modellrechnung ein wenig sensibilisiert und sehen, dass *ein bischen* Verschwendung wahrscheinlich nicht ausreicht, um die Risiken der Fundholder zu kompensieren. In jedem Fall ist Fundholding kein Selbstläufer.

F15.6 Principal - Agent - Theorie (PAT)

Die Grundidee

Die PAT (auf Deutsch häufiger „Sachwaltertheorie“) untersucht Beziehungen, in denen ein Auftraggeber (Principal) einen Sachwalter (Agent) damit beauftragt, seine Interessen zu vertreten, weil er das selbst nicht kann (oder will).

Dieser Beziehungstyp ist allgegenwärtig.

- Sie können Ihr Auto nicht selbst reparieren und schieben es daher in eine Autowerkstatt, damit ein Mechaniker das tut.
- Sie können als Aktionär nicht das Tagesgeschäft des Unternehmens, dessen Aktionär Sie sind, bestimmen. Also stellen Sie Manager ein, die das für Sie tun und Aufsichtsräte, die die Manager kontrollieren.
- Sie können als medizinischer Laie und Patient nicht entscheiden, welche Therapie für Sie sinnvoll ist und diese Therapie auch nicht selbst durchführen. Also beauftragen Sie einen Arzt, das zu tun.
- Ihnen fehlt die Lust und die Zeit, Ihre Wohnung zu putzen. Also beauftragen Sie eine Putzfrau, das zu tun.

In diesen Principal-Agent-Beziehungen ist es im Regelfall sinnvoll, dem Agenten einen relativ großen Spielraum zu lassen. Häufig ist im Voraus nicht im Detail klar, was zu tun ist oder es ist einfach zu aufwändig, alle Details im Voraus zu klären. Sie wollen nicht, dass der Automechaniker Sie zwanzigmal anruft, um abzuklären ob Sie dies oder das wollen oder nicht.

- Wenn Teil x noch in Ordnung ist, dann soll es nicht ausgetauscht werden, sonst schon. Weil Sie keine Ahnung haben, überlassen Sie dem Mechaniker die Entscheidung, ob es Sinn ergibt, das Teil auszutauschen.
- Wenn Ihr Auto aber schon sehr betagt ist, erwarten Sie vom Mechaniker aber auch, dass er Ihnen den Rat gibt, das Auto auf den Schrott zu schieben.

Eigeninteressen des Agents

An dieser Stelle deutet sich an, dass der Agent Eigeninteressen hat und diese Eigeninteressen zum Problem für den Principal werden können.

- Der Mechaniker führt Reparaturen durch, die eigentlich nicht sinnvoll sind, weil er an der Reparatur verdient
- Der CEO eines Unternehmens ignoriert langfristige Ziele, da sein Gehalt sich an kurzfristigen Ergebnissen orientiert.
- Der Arzt führt Behandlungen durch, die eigentlich nicht sinnvoll sind, weil er so mehr Umsatz erzielen kann.
- Die Putzfrau wird nach Stunden bezahlt und putzt langsamer als möglich wäre.

Incentives für den Agent

Die PAT lotet nun aus, mit welchen Instrumenten der Principal seinen Agenten in solchen Situationen „auf Linie“ bringen kann. Es ist relativ klar, dass der Agent Sanktionen befürchten muss, wenn er dabei erwischt wird, wenn er seine eigenen Interesse zu Lasten seines Principals verfolgt. Das bedeutet, dass es ein Element der *Kontrolle* in den meisten Principal-Agent-Beziehungen gibt.

Noch geschickter ist es allerdings, dem Agent zusätzliche Anreize zu setzen, die ihn stärker an die Interessen des Principals binden.

Einige AGs, die die Kurzfristorientierung des Managements unterbinden wollen, sind z.B. dazu übergegangen, Erfolgsprämien nicht ausschließlich an den aktuellen Ergebnissen zu orientieren, sondern auch Boni über mehrere Jahre zu verteilen. So könnte ein Teil des Bonus für 2020 erst im Jahr 2022 ausgezahlt werden, wenn auch im Jahr 2022 entsprechende Erfolge stattfinden. Auf diese Weise bekommt das Management einen Anreiz, den zeitlichen Horizont zu erweitern. Mit der Putzfrau könnte man einen Pauschalbetrag aushandeln, für den die Wohnung geputzt wird, unabhängig von der tatsächlich benötigten Zeit usw.

Anwendung auf die Pauschalhonorierung

Das Konzept der Pauschalhonorierung lässt sich mit der PAT recht gut darstellen, wenn man die Versicherung als Principal und den Fundholder als Agent betrachtet.

Der Fundholder bekommt einen Sicherheitszuschlag, um sein Risiko zu kompensieren, das er mit dem Fundholding übernimmt. Diesen versicherungsmathematischen Aspekt können wir an dieser Stelle ignorieren. Darüber hinaus bekommt er vom Principal einen Bonus und natürlich auch noch die zu erwartenden Durchschnittskosten. Seine Aufgabe ist es, die Versorgung der Versicherten möglichst kostengünstig bereitzustellen. Der Anreiz, das zu tun, besteht darin, dass er, wenn seine Ausgaben unterhalb der Durchschnittskosten bleiben, diese Differenz behalten kann. Mit dieser Mechanik bringt der Principal den Agent „auf Linie“.

F15.7 Interessenkonflikte

Wenn wir Fundholding durch die PAT-Brille betrachten, merken wir schnell, dass Fundholding ein extrem kompliziertes Konstrukt ist.

Wer ist der Principal?

In unserer Analyse sind wir ganz selbstverständlich davon ausgegangen, dass der Principal des Fundholders die Versicherung ist. So haben wir auch unsere Simulation aufgezogen. Vielleicht war das etwas kurz gedacht.

Aus Sicht des Patienten ist er, der Patient, der Principal. Er beauftragt den Arzt, seine Gesundheitsversorgung bereitzustellen. Weil er (o je!) diese Versorgung aber nicht direkt bezahlen muss, will er eine Maximalversorgung. Wenn die Versicherung den Arzt über ein Fundholding-Konzept „auf Linie“ bringt, ist das die Linie der Versicherung und nicht die die Patienten.

Es wird noch komplizierter. Der *Patient* ist ja die Person, die weiß, dass sie medizinische Leistungen braucht. In *Einheit Rationierung und Ethik* haben wir uns die Idee des

Schleiers des Nichtwissens von John Rawls angeschaut. Für den Patienten hat sich der Schleier schon gelüftet. Er ist krank. In der rawls'schen Denke müssen wir uns über den *Gesunden* Gedanken machen, der noch nicht weiß, ob und wann er krank werden wird. Für den ist der Schleier noch nicht gelüftet. Sie erinnern sich an unser Gedankenexperiment: Es ist unwahrscheinlich, dass die Gesunden für eine Maximalversorgung stimmen würden. Also würden die Gesunden vom Fundholder ebenfalls erwarten, dass er die Gesundheitsversorgung des Kranken ohne Firlefanz zu möglichst niedrigen Kosten bereitstellt. Die Gesundhen und die Kranken in der gleichen Kasse haben also ganz unterschiedliche Interessenlagen und können sich beide als Principals des Fundholders sehen.

Wir haben also eine sehr unübersichtliche Struktur mit mehreren Principals, die teilweise konträre Interessen haben.

Fachärzte und Krankenhäuser

Ganz am Anfang der Bachelor Veranstaltung haben wir darüber nachgedacht, dass im Gesundheitssystem die Kosten des Einen das Einkommen des Anderen sind.

Das Ziel von Fundholding besteht darin, die Kosten der medizinischen Versorgung zu senken. Das können wir direkt in

> *das Ziel ist, die Einkommen der Leistungserbringer zu senken*

übersetzen, denn anders geht es nicht. Wir brauchen nicht viele Finger, um abzählen zu können, dass das diesen Leistungserbringern nicht gefallen wird und wir wissen auch, wer diese Leistungserbringer sein werden: Die Fachärzte und die Krankenhäuser, also die Leistungserbringer, an die der Fundholder seine Patienten *nicht* weiterleitet, weil er die Behandlung selbst übernimmt und die Leistungserbringer, denen er genau auf die Finger schaut, denn jedes verzichtbare Röntgenbild, das im Krankenhaus gemacht wird, ist ein Röntgenbild, dass der Fundholder bezahlen muss. Wenn das Röntgenbild nicht gemacht wird, ist das Umsatz, den das Krankenhaus nicht macht und (indirekt) Einkommen, dass die Mitarbeiter im Krankenhaus nicht bekommen.

Aus Sicht dieser anderen Leistungserbringer ist der Fundholder also ein Scherge der Kassen und genauso sperrig, wie die Leistungserbringer sich den Kassen gegenüber verhalten, werden sie sich dem Fundholder gegenüber verhalten. Er wird also kein einfaches Leben haben und die Leistungserbringer werden nicht gleich kuschen, wenn er böse guckt.

Kapitel 16

Fixgehalt

- Das Fixgehalt ist die Honorierungsform, die am häufigsten verwendet wird. Wir werden uns überlegen, welche Vor- und Nachteile diese Honorierungsform hat.
- Mit dem Instrumentarium der PAT werden wir analysieren, wie der Arbeitgeber versuchen kann, die Interessenlage des fix entlohnten Arbeitnehmers zu beeinflussen.

F16.1 Konzept

In der letzten Einheit haben wir die PAT kennengelernt und ich möchte diesen Ansatz gleich verwenden, um das Konzept des Fixgehalts zu beschreiben. Natürlich ist uns allen grundsätzlich klar, was ein Fixgehalt ist, aber vielleicht hilft uns dieser neue Zugang zu einem umfassenderen Verständnis. Also:

> *Der Principal kauft einen Block an Arbeitszeit des Agents und erwirbt damit die Verfügungsrechte über diese Zeit.*

Kein unternehmerisches Risiko

Da der Preis vorher ausgehandelt worden ist, hängt er nicht von der Auslastung des Agents ab. Das Risiko von Leerkapazitäten trägt der Principal. Wenn in der ELV keine Patienten kommen, gibt es keine Umsätze. Wenn es in der Pauschalhonorierung keine Fälle oder Köpfe gibt, gibt es keine Umsätze. All diese Risiken gibt es beim Fixgehalt nicht.

Niedrigere Honorierung

In der letzten Einheit haben wir länger darüber nachgedacht, dass ein Fundholder, der ein unternehmerisches Risiko tragen *muss*, sich dafür honorieren lassen wird. Das war der Sicherheitszuschlag, der so hoch sein kann, dass er jegliche Einsparpotentiale auffrisst.

Wenn wir jetzt über ein Konzept nachdenken, in dem es dieses Risiko *nicht* gibt, sprechen wir über eine Honorierung, die *niedriger* ausfällt als bei der Pauschalhonorierung.

Kein Eigentum an Produktionsmitteln

Im Regelfall stellt der Principal die notwendige Infrastruktur zur Verfügung, so dass der Kapitalbedarf des Agents sehr gering ist. Karl Marx würde hier vom „freien“ Arbeiter sprechen, der in einem Doppelsinn frei ist. Er ist kein Leibeigener, also nicht irgendeinem

Feudalherren gegenüber verpflichtet, aber er ist auch frei von Produktionsmitteln, d.h. die Produktionsinfrastruktur gehört nicht ihm, sondern dem Kapitalisten. Wir würden den, ideologisch etwas weniger stark belastet, *Arbeitgeber* nennen.

In unserer Kostensystematik würden wir die Produktionsinfrastruktur als *Fixkosten* bezeichnen. Wenn der Arbeitnehmer die Fixkosten selbst tragen müsste, müsste er eine große Summe investieren und mit dieser Investition ja wieder ein unternehmerisches Risiko tragen. Genau die Abwesenheit dieses Risikos ist aber ja das Kennzeichen des Fixgehalts.

Für viele Absolventen eines Medizinstudiums ist die Arbeit als angestelter Arzt in einem MVZ u.a. deswegen attraktiv, weil eben genau diese Investition in die Fixkosten nicht geleistet werden muss. Als angestellter Jung-Zahnarzt muss man nicht erst 150.000€ für drei Behandlungsstühle investieren, sondern benutzt die, die da sind. Gleichzeitig gewinnt man Flexibilität, weil die Investition ja refinanziert werden muss, unabhängig davon, ob man persönlich eine Familenphase mit reduzierter Arbeitszeit plant oder nicht.

F16.2 Außen- und Innenfinanzierung

Fixgehalt ist die Honorierungsform, die im Gesundheitssystem am verbreitetsten ist.

An dieser Stelle ist es wichtig, zwischen *Außen-* und *Innenfinanzierung* zu unterscheiden, wie das Beispiel des *Krankenhauses* verdeutlicht.

Wie in dieser und der Bachelor-Veranstaltung skizziert, werden Krankenhäuser in Deutschland zu einem großen Teil über DRG-Fallpauschalen honoriert. Dieses Geld erhält das *Unternehmen*.

„Das Unternehmen“ behandelt die Patienten aber nicht, sondern die *Menschen* in dem Unternehmen. Diese Menschen (Ärzte, Pflegepersonal, kaufmännische Abteilung) werden aber nicht über Fallpauschalen bezahlt, sondern über Fixgehälter. Das bedeutet, dass man über die *Außenfinanzierung* eines Leistungserbringers bestimmte Anreize setzen kann. Das garantiert aber nicht, dass diese Anreize direkt auf die handelnde Person in dem Unternehmen durchschlagen, wenn diese Person in der *Innenfinanzierung* nach einem anderen Prinzip honoriert wird.[1]

Der gleiche Bruch in den Finanzierungsprinzipien liegt auch bei den niedergelassenen Ärzten in Deutschland vor. Der Arzt selbst ist selbständig und wird wie bereits beschrieben honoriert. Er erbringt die Leistung aber nicht allein, sondern wird von einem Team dabei unterstützt, das von ihm über Fixgehälter bezahlt wird. In der Arztpraxis haben wir also eine bunte Mischung von Honorierungsformen.

F16.3 Eigenschaften des Fixgehalts

Wie bei den anderen Honorierungsformen wollen wir nun versuchen, das Fixgehalt durch unser Bewertungsraster laufen zu lassen

- Effizienz: Im Vergleich zu ELV und Pauschalen hat der Agent kein ausgeprägtes Interesse daran, eine definierte Handlung möglichst effizient zu erbringen. Wenn er sie

[1] Ich reite ein wenig auf diesem Umstand herum, weil die Überlegung einerseits trivial ist, aber andererseits leicht übersehen wird. Ich erinnere mich an meine ersten Begegnungen mit den Fallpauschalen und meiner spontanen Erkenntnis, dass das ja zu einer sehr effizienten Produktion führt, weil ... Was ich übersehen hatte, war, dass „das Krankenhaus“ ja kein monolithischer Block ist, sondern eine Binnenstruktur hat. Vielleicht sind Volkswirte besonders von diesem Problem betroffen, weil sie eher zu einer *um-die-kleinteiligen-Details-kümmern-sich–dann-die-Betriebswirte* Denke neigen.

effizient erbringt, hat er keinen Vorteil dadurch, weil der Principal seine *Arbeitszeit* gekauft hat. Anders herum: Wenn er die Leistung ineffizient erbringt, hat er dadurch keine Nachteile.

- Qualität: Ein Fixgehalt setzt keinen Anreiz für den Agent, die Leistungen in hoher Qualität zu erbringen. Anders als bei der Pauschalhonorierung ist für seine Honorierung die Qualität irrelevant, da er über seine Arbeitszeit honoriert wird.
- Leistungsvolumen: Das Leistungsvolumen des Agents in Zeiteinheiten gemessen ist fix. Innerhalb dieses Zeitrahmens hat er aber keinen Anreiz, möglichst viele Leistungen zu erbringen.

Wenn wir diese Eigenschaften anschauen, sieht das Fixgehalt für den Principal nicht besonders attraktiv aus. Warum wird es dann überhaupt eingesetzt?

F16.4 Löhne, Leistung und Kontrolle

Lohnniveau

Wir haben schon festgehalten, dass bei einem Fixgehalt kein unternehmerisches Risiko besteht und daher die Honorierung niedriger ist. Das ist der wesentliche Pluspunkt dieser Honorierungsform sowohl für den Agent als auch für den Principal.

Wenn das Honorar niedrig genug ist, kompensiert das den fehlenden Anreiz zu Effizienz und hoher Qualität und die Rechnung geht für den Principal auf.

Diese Konstruktion wirft das Problem auf, dass für den Fixhonorierten der Unterschied zwischen Ausbeutung und angemessenem Abschlag für fehlendes Risiko schwer zu erkennen ist.

Wenn Bob mit einer halben Stelle als Arzt bei Alice, der Praxisinhaberin angestellt ist, könnte es sein, dass Alice fünfmal so viel verdient wie Bob. Ob Alice nun schlecht zahlt, oder ob dieser hohe Unterschied die Kompensation für Risiko, Investition und Flexibilität ist und ob Bob auch die Lohnnebenkosten im Blick hat, ist unklar. Es könnte aber sein, dass Bob sich unterbezahlt und benachteiligt fühlt, obwohl er das objektiv überhaupt nicht ist.[2]

Kontrollkosten

Wenn Alice befürchten muss, dass Bobs Leistungsniveau relativ gering ist, weil er mit seinem Fixgehalt keinen Anreiz zu Effizienz hat und weil er sich vielleicht (fälschlicherweise) für unterbezahlt hält, ist eine Standardstrategie die der Leistungskontrolle.

Wie immer die im Detail auch ausfällt (was hier nicht unser Thema ist) kostet diese Kontrolle Geld und diese Kosten verringern die Attraktivität des Fixgehalts.

Effizienzlöhne

Arbeitgeber zahlen (so die Effizienzlohnhypothese) freiwillig ein Gehalt über Marktniveau. Wenn Alice 500€ je Monat mehr zahlt als andere Ärzte, ist dieser Arbeitsplatz für Bob attraktiv. Vielleicht fühlt er sich immer noch unterbezahlt, aber ihm ist klar, dass die

[2] Bernau, Patrick und Bollmann, Ralph, "Republik der Neider", Frankfurter Allgemeine Sonntagszeitung 14.1. 2024. S.17.

anderen Ärzte noch schlechter zahlen. Aus diesem Grund wird er nicht riskieren wollen, eine Abmahnung und Kündigung zu erhalten, sondern seinen Arbeitseinsatz erhöhen.

Für Alice können sich diese 500€ rechnen, wenn sie dadurch ihre Kontrollkosten senken kann und Bob gleichzeitig sein Leistungsniveau anhebt. Wenn die Kombination dieser beiden Effekt größer als 500€ sind, hat sich der Effizienzlohn gerechnet.

Die Idee der Effizienzlöhne funktioniert natürlich nur so lange, bis genügend Arbeitgeber auf die gleiche Idee kommen. Wenn Carol, Dave und Eve auch 500€ mehr zahlen, weil sie sich den Effizienzlohneffekt bei Alice angeschaut haben, hat Bob keinen großen Anreiz mehr, seinen Arbeitsplatz zu behalten, weil ja alle jetzt mehr zahlen.

Boni

Der Nachteil von Effizienzlöhnen besteht aus Sicht des Arbeitgebers darin, dass er im Voraus gezahlt wird, d.h. in der Hoffnung, dass die überdurchschnittliche Bezahlung die beschriebene Wirkung entfaltet. Das kann so eintreten oder auch nicht. Vielleicht ist Bob einfach nur geschickt und nutzt das gesunkene Kontrollniveau aus, um eine ruhige Kugel zu schieben.

Unter diesem Aspekt sind Boni so etwas wie Effizienzlöhne, die retrospektiv gezahlt werden. Wenn Bob bestimmte Ziele erreicht hat, bekommt er den Bonus, also einen Effizienzlohn bei nachgewiesener Effizienz.

Diese Idee wirft eine Reihe von Problemen auf.

1. Bei individuellen Boni muss das Ziel sehr individualisiert ausgehandelt und überprüft werden. Das verursacht Kosten.

2. Bei Gruppen-Boni besteht die Gefahr von Trittbrettfahrern, die von der Mehrleistung der anderen Gruppenmitglieder profitieren, selbst aber nichts beitragen.

Korrumpierungseffekt

Ein drittes Problem von Boni ist etwas aufwändiger zu beschreiben. Daher widme ich ihm einen eigenen Abschnitt.

Die Idee der Boni basiert auf der Annahme, dass die Arbeitnehmer motiviert werden müssen, weil sie letztlich keine Lust haben, zu arbeiten. Wenn man nichts tut, schieben sie eine Kugel, die maximal ruhig ist. Dougas McGregor hat für solche Ansätze in den 1960er Jahren den Begriff *x-Theorie* geprägt. [3] Das Gegenstück dazu sind *y-Theorien*, die davon ausgehen, dass Menschen grundsätzlich motiviert sind und ihre Arbeit gern, freiwillig und engagiert tun. Diese Unterscheidung ist natürlich selbst wieder eine Theorie und heißt, mäßig originell, *xy-Theorie*.

Es sieht danach aus, als ob es Beispiele für beiden Theorien gibt und die Frage, welche Theorie im Einzelfall zutrifft, eben eine Frage des Einzelfalls ist. Der Erklärungswert dieser Theorie ist also etwas begrenzt, weil die Theorie mit jedem Verhalten kompatibel ist.

An einer Stelle hilft uns die Theorie aber weiter. Wir können uns überlegen, dass es folgende Szenarien gibt:

[3] Frischmuth, Carlos. New Work Bullshit, Frankfurt: Frankfurter Allgemeine Buch, 2021, S. 109

	Mitarbeitertyp	
	x-Mitarbeiter	y-Mitarbeiter
Bonus	hohe Leistung	?
kein Bonus	niedrige Leistung	hohe Leistung

Boni sind auf Mitarbeiter zugeschnitten, die der x-Theorie entsprechen, also im Wesentlichen nichts tun, wenn man sie nicht (durch Geld) motiviert. Hier spricht man auch von *extrinsischer* Motivation.

Auf der anderen Seite gibt es aber auch Mitarbeiter, die der y-Theorie entsprechen. Diese Mitarbeiter bringen auch ohne monetäre Anreize eine hohe Leistung (intrinsische Motivation) Man könnte nun sagen, dass die Boni in diesem Fall Verschwendung sind, weil die Leistung auch ohne diesen Anreiz erfolgen würde. Es könnte aber sein, dass die Boni tatsächlich nicht nur überflüssig sind, sondern Schaden anrichten.

Wenn intrinsische Motivation für den Agent eine große Rolle spielt und der Principal ihm extrinsische Anreize (Bonus) setzt, kann das zu einem Verlust der intrinsischen Motivation führt, weil der Bonus das Mitarbeiterbild des Unternehmens deutlich macht.

Das kann zu zwei Reaktionen führen

- Ohne Zahlung des Bonus leistet der bisher intrinsisch Motivierte weniger als bisher. Der Bonus wirkt dann wie eine Droge.[4]
- Der Agent wägt ab, ob die Mehrleistung, die er zur Erreichung des Bonus erbringen muss, mit der Höhe des Bonus in einer attraktiven Relation steht. Ist das nicht der Fall, senkt er sein Leistungsniveau, weil das Unternehmen ihm klargemacht hat, dass es gar keine volle Leistung von ihm erwartet.

Dieser Mechanismus wird Korrumpierungseffekt oder auch Motivationsverdrängung genannt.[5]

[4]Sprenger, Reinhard K.. Mythos Motivation - Wege aus einer Sackgasse, Frankfurt/Main: Campus-Verl, 1994.

[5]Dobelli, Rolf. Die Kunst des klugen Handelns - 52 Irrwege, die Sie besser anderen überlassen, München: dtv, 2015.

Teil V

Steuerung über Organisationsstrukturen

Im zweiten Teil dieser Veranstaltung haben wir uns die Frage gestellt, mit welchen Instrumenten man ein Gesundheitssystem stärker in Richtung „Effizienz“ entwickeln kann. Die Ergebnisse unserer Überlegungen waren:

- Auf eine freiwillige Verhaltensänderung der Versicherten sollten wir nicht bauen.
- Die verschiedenen Honorierungsform der Leistungserbringer setzen unterschiedliche Anreize. Es gibt aber keine „Gewinner-Honorierungsform“, die auf jeden Fall funktionieren wird.

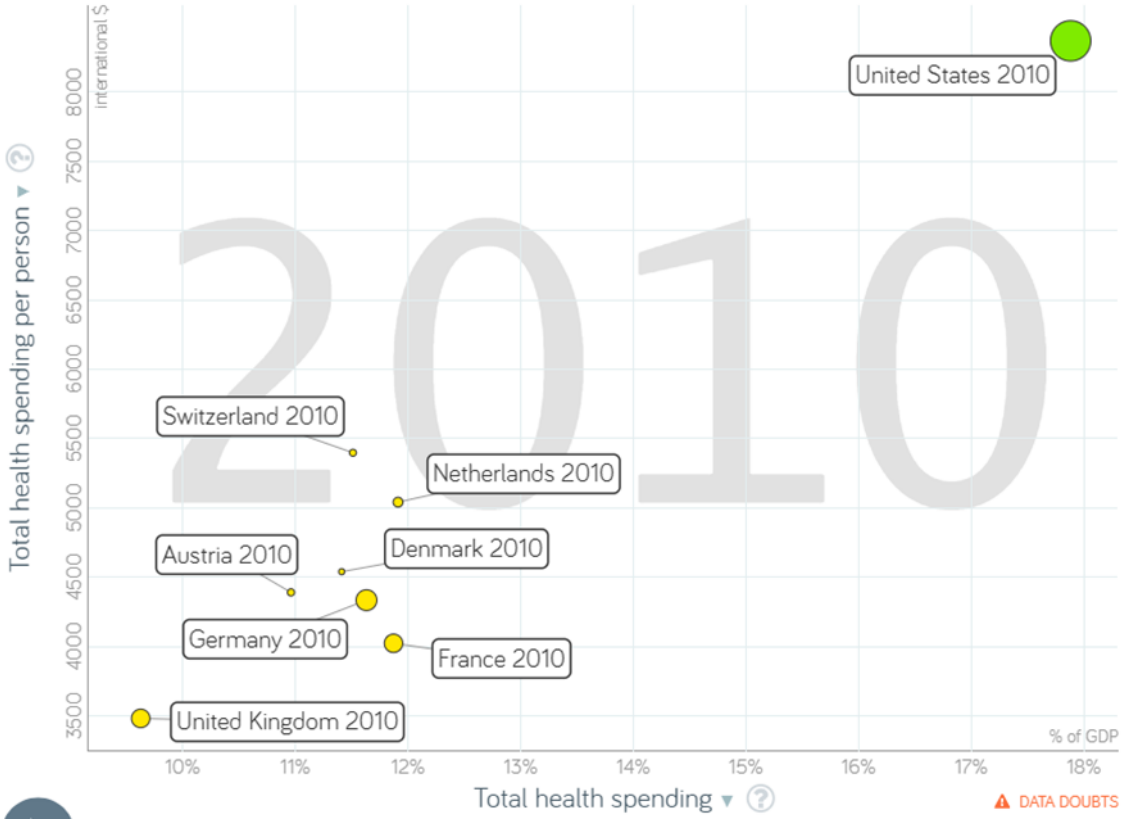
Eine weitere Ebene, auf der man versuchen kann, die Effizienz eines Gesundheitssystems zu erhöhen, besteht in einer Änderung der *Organisationsstrukturen*. Hierzu ist es hilfreich, sich den Aufbau des Gesundheitssystems in anderen Ländern anzuschauen. Im Folgenden werden die Grundstrukturen der Systeme in den USA und Großbritannien vorgestellt und wir werden uns den Krankenhaussektor in Dänemark anschauen. Der Schwerpunkt dieses Teils liegt auf den USA, da dort die Vielfalt an Geschäftsmodellen größer ist als in Großbritannien.

Kapitel 17

Das Gesundheitssystem der USA

F17.1 Kostenniveau

Das US-Gesundheitssystem ist im internationalen Vergleich ein extrem teures System, wie die folgende Grafik zeigt.[1]



Das betrifft sowohl das *absolute Ausgabenniveau* als auch der *Anteil der Gesundheitsausgaben am BIP*.

Diese doppelte finanzielle Ausreißerposition ist insofern wichtig, als dass es nicht überraschend ist, dass die Pro-Kopf-Ausgaben in den USA hoch sind, weil das Pro-Kopf-Einkommen auch hoch ist. Es ist aber insofern wichtig, als dass auch der Anteil am BIP, der in den USA für Gesundheit ausgegeben wird, im Vergleich sehr hoch ist.

[1] http://www.gapminder.org. Die Daten sind nicht ganz neu, aber die letzten, für die umfangreiche, vergleichbare OECD-Daten verfügbar waren. Die Größe der Kreise zeigt die Bevölkerungsgröße an.

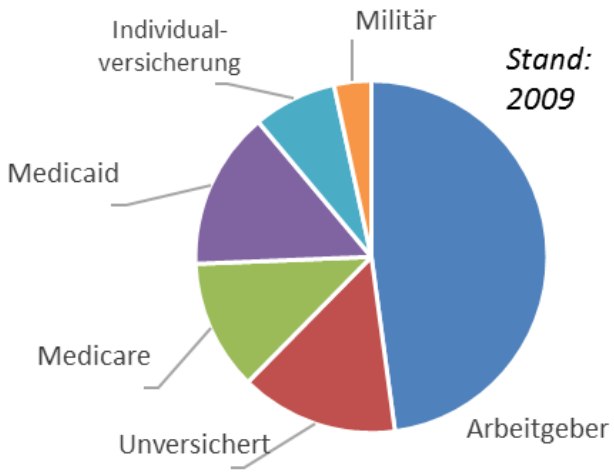
im Bachelor haben wir festgestellt, dass die Einkommenselastizität von Gesundheitsausgaben überproportional ist, also Gesundheitsgüter nach der Standarddefinition in die Kategorie der *Luxusgüter* fallen. Insofern ist es auch nicht überraschend, dass ein Land mit hohen Einkommen nicht nur mehr Geld für Gesundheit ausgibt, sondern auch einen höheren Anteil am Einkommen. Wir werden aber sehen, dass diese Erklärung für den extrem hohen Anteil am BIP zu einfach ist.

F17.2 Finanzierung der medizinischen Versorgung

In den meisten Industrieländern ist das Gesundheitssystem staatlich organisiert oder zumindest sehr staatsnah. In den USA ist das anders. Die folgende Grafik[2] zeigt Ihnen, wie die US-Amerikaner versichert sind.

Staatliche Systeme

Medicaid ist eine staatliche Minimalversicherung für Einwohner ohne Vermögen und mit geringem Einkommen, *Me*dicare für ältere Einwohner ohne Vermögen. Nimmt man noch die Versorgung des Militärs hinzu, ist etwa ein Viertel der Bevölkerung über diese staatlich finanzierten Systeme versichert.



Arbeitgeber

Die Hälfte der Bevölkerung ist über den Arbeitgeber versichert. Arbeitgeber *können* ihren Mitarbeitern eine Krankenversicherung als freiwillige Sozialleistung anbieten. Das tun sie im Regelfall nicht aus reiner Menschenliebe, sondern weil die Mitarbeiter dann geringere Gehaltsforderungen haben. Wenn die Einsparungen beim Gehalt höher sind als die Kosten der Krankenversicherung, ist das Angebot einer Krankenversicherung für den Arbeitgeber lukrativ.

Das bedeutet, dass der traditionelle Partner der Versicherung nicht der einzelne Arbeitnehmer ist, sondern der Arbeiteber. Arbeitgeber und Versicherung schließen einen

[2] Amelung, Volker Eric. Managed Care - Neue Wege im Gesundheitsmanagement, Wiesbaden: Gabler, 5. Auflage, 2012.

Gruppenvertrag ab. Dazu schätzt die Versicherung die die Morbiditätsstruktur der Belegschaft ein und kalkuliert eine einheitlichen Pro-Kopf-Pauschale je Mitarbeiter.[3]

Das bedeutet, dass die Krankenversicherung bei einem Arbeitgeberwechsel endet. Es bedeutet auch, dass Arbeitnehmer mit teuren Krankheiten einen Nachteil bei der Arbeitsplatzsuche haben, weil es zwar keine individuell höhere Prämie für diese Patienten gibt, aber die Pauschalen steigen.

Invididualversicherung

Die Krankenversicherungen sind in erster Linie auf die Gruppenverträge spezialisiert und nicht auf die Kalkulation individueller Prämien, wie die deutsche PKV. Der Anteil solcher Individualversicherungen ist relativ überschaubar, weil die Prämien relativ hoch sind und der Arbeitgeber sich nicht, wie in Deutschland, an der Finanzierung beteiligen muß.

Unversicherte

Etwa 20% der Bevölkerung hat *keine* Krankenversicherung. Der Anteil der Bevölkerung, die keine Versicherung haben *will*, dürfte ziemlich gering sein, aber wir haben bereits gesehen, dass man entweder ziemlich arm sein oder einen Arbeitsplatz mit Krankenversicherung haben muß, um krankenversichert zu sein.

Eine häufige Nebenhandlung in US-Filmen ist, dass eine Figur aus der Mittelschicht schwere Gesundheitsprobleme bekommt und Arztrechnungen nicht mehr bezahlen kann.[4] Das ist überhaupt nicht exotisch. Tatsächlich sind (kostspielige) Gesundheitsprobleme die Hauptursache für Privatinsolvenzen.[5]

Ein weiteres Detail aus US-Arztserien: Dort spielen Notfallambulanzen eine große Rolle, weil es dort eine Behandlungspflicht des Krankenhauses gibt, die andere Leistungserbringer nicht haben. Wenn ein Patient also keine Kreditkarte vorlegen kann, kann der Arzt die Behandlung ablehnen. Das bedeutet, dass die Notfallambulanzen der Krankenhäuser häufig auch die medizinische Versorgung von Nichtversicherten mitübernehmen.

F17.3 Affordable Care Act (Obamacare)

Ein wichtiges innenpolitisches Projekt der Obama-Administration war, einen flächendeckenden Versicherungsschutz für die gesamte Bevölkerung zu schaffen. Das hat zu dem Afforable Care Act (ACA), auch „Obamacare“ genannt, geführt.

Grundidee

Die Grundidee von ACA ist, standardisierte Einzelverträge mit drei verschiedenen standardisierten Leistungskatalogen einzuführen, die von privaten Krankenversicherungen angeboten werden. Die Kassen können die Prämien selbst kalkulieren, dürfen aber keine Gesundheitsprüfung durchführen.

[3] Im Detail ist das noch etwas komplizierter, weil es Regeln gibt, wie viele Alternativangebote ein Arbeitgeber vorhalten muss und welche Eigenschaften diese Angebote haben müssen. Für unseren Zweck reicht diese grobe Beschreibung.

[4] „Mittelschicht“ ist wichtig, weil diese Filmfiguren (zumindest am Anfang) noch Vermögenswerte haben, die sie auflösen müssen, bis sie Medicaid-berechtigt sind.

[5] Himmelstein, David U. et al. , "Medical bankruptcy in the United States, 2007: results of a national study", The American journal of medicine 2009. S.741-6.

Einwohner, die keine anderweitige Krankenversicherung haben, müssen mindestens einen Vertrag über das kleinste ACA-Versicherungspaket abschließen. Tun sie das nicht, müssen sie eine Strafe zahlen.

Man muß den ACA als Fehlschlag bezeichnen.[6]

- Die Hauptzielgruppe der Unversicherten wählt häufig lieber die Strafzahlung, da sie selbst die Prämien des Minimaltarifs nicht zahlen kann.
- Arbeitgeber, die ihren Mitarbeitern bisher eine Krankenversicherung angeboten haben, beenden dieses Angebot und verweisen auf die ACA-Tarife, d.h. diese nicht mehr krankenversicherten Mitarbeiter vergrößern erst einmal die Gruppe der Nichtversicherten.
- Die rechtlichen Rahmenbedingungen für die Versicherer sind unattraktiv, so dass viele Versicherungen keine ACA-Tarife (mehr) anbieten.

Widerstand

Ein Grund für diesen Fehlschlag war, dass der politische Widerstand der Republikaner sehr hoch war und eines der wichtigsten Wahlversprechen von Donald Trump war, den ACA sofort wieder abzuschaffen (was dann aber noch nicht passiert ist).[7]

F17.4 Libertarismus

Aus europäischer Sicht ist dieser weit verbreitete Widerwille gegen einen auch nur rudimentären Sozialstaat schwer verständlich. Ein Grund scheint mir die Idee des Libertarismus zu sein.

„tar“ versus „ral“

Parteien wie die FDP werden als *liberal* bezeichnet. Ihr Gesellschaftsbild wird *Liberalismus* genannt. Die Denkrichtung, um die es hier geht, heißt *Libertarismus*. In beiden Begriffen steckt die lateinische Wurzel *Libertas* (Freiheit). Ohne irgendwas über diese beiden Denkschulen zu wissen, können wir schon spekulieren, dass *Freiheit* eine ganz zentrale Rolle spielen wird.

In Deutschland gibt es kaum Vertreter des Libertarismus. Die Wahl des argentinischen Präsidenten Javier Milei 2023 hat dieser Denkrichtung aber auch bei uns mehr Aufmerksamkeit beschert. Die Wikipedia bezeichnet ihn als *ultraliberal* und *Anarchokapitalisten*. Präziser wäre tatsächlich *libertär*.

Sklaverei und Freiheit

Ein Ausgangspunkt des Libertarismus war die Auseinandersetzung mit der Sklaverei in den USA. Ein Sklave ist dadurch definiert, dass er arbeiten und das Produkt seiner Arbeit

[6] von Petersdorff, Winand, "Obamas Krankenversicherung gehen die Anbieter aus", Frankfurter Allgemeine Zeitung 18.8 2016. S.20.

[7] Über den Stellenwert der Krankenversicherung bei der demokratischen Partei siehe Steffens, Frauke, "Wie weiter mit Obamacare", Frankfurter Allgemeine Zeitung 11.3. 2021. Aktuell ist eine grundlegende Reform oder Weiterentwicklung des Systems nicht auf der politischen Agenda.

an seinen Besitzer abliefert muss. Da hat er keine Wahl, weil er dem Besitzer wie eine Sache gehört. Im Gegenzug dazu bekommt er Essen und Unterkunft in der Menge und Qualität, die sein Besitzer für richtig hält. Auch da hat er keine Wahl.

Staatsfinanzierung

Ein Staat muss seine Leistungen über *Steuereinnahmen* finanzieren. Bei den Steuern hat der Besteuerte keine Wahl, ob er die Steuern zahlen will oder nicht. Damit (so die Überlegung) nimmt der Staat seinen Bürgern gegenüber die gleiche Position ein, wie der Sklavenhalter gegenüber seinen Sklaven. Das ist nicht legitim

Der *Steuerzahlergedenktag* bringt diese Überlegung zum Ausdruck. Die durchschnittliche Steuerbelastung des Einkommens (Einkommensteuer, MwSt, Mineralölsteuer, ...) lag 2024 bei 52,6%, so dass man sagen könnte, das der Durchschnittsbürger bis zu, 11.7. für den Staat arbeitet.[8]

Etwas polemisch formuliert könnte man sagen, dass dieser Bürger ein Halbtagssklave ist, weil er die Hälfte seines Einkommens ohne dass er eine Wahl hat, abgeben muss. Natürlich bekommt er auch Gegenleistungen dafür, aber das ist beim Sklaven ja auch so.

Minimalstaat

Wenn man gegen Sklaverei ist, muss man (so die Überlegung) auch gegen Steuerzahlungen sein. Zumindest gegen alle Steuern, mit denen Dinge finanziert werden, die keine öffentlichen Güter sind.[9]

Die wesentliche öffentlichen Güter sind *Sicherheit* und *Eigentumsschutz*. Der Staat sorgt also dafür, dass die Freiheit, die sich ein Bürger nimmt, nicht zur Einschränkung der Freiheit anderer führt.

Weil ein Staat, der sich auf diese essentiellen Leistungen beschränkt, um kein Sklavenhalterstaat zu sein, ziemlich schlank sein wird, nennt man dieses Konzept *Minimalstaat*. Manchmal wird auch der Begriff *Nachtwächterstaat* verwendet, weil die Hauptaufgabe des Staates eben ist, aufzupassen, dass keiner übergriffig wird und Straftaten begeht.[10]

Selbstverantwortung

Der Preis, den die Menschen für die Abschaffung der Halbtagsskaverei zahlen müssen ist, dass sie sich selbst um ihre Interessen kümmern müssen, weil der Staat das nicht tut. Das bedeutet, dass die Kehrseite von *Freiheit* die *Selbstverantwortung* ist.

Ich habe die Freiheit zu tun und lassen, was ich will (so lange das keinen anderen in seiner Freiheit einschränkt). Ich habe also auch die Freiheit, großen Mist zu bauen. Dann muss ich die Konsequenzen aber auch selbst tragen und kann dann nicht nach Vater Staat rufen.

---

[8] o.V., "Bürger arbeiten bis zum 11. Juli für den Staat", Frankfurter Allgemeine Zeitung 10.7. 2024. S.19.

[9] Mit dieser Systematik haben wir uns in *Einheit* **??** befasst. Ein Ergebnis war, dass die meisten Güter, die umgangssprachlich als „öffentliche Güter“ bezeichnet werden, tatsächlich private Güter oder Clubgüter sind, die der Staat nur kostenlos anbietet.

[10] Der zentrale Text des Libertarisms ist Nozick, Robert. Anarchie, Staat, Utopia, München: Olzog, 2011.

Die Rolle von (Kapital) Märkten

Güter, die der Staat bisher kostenlos zur Verfügung gestellt hat, die aber keine echten öffentlichen Güter sind, verschwinden nicht zwangsläufig. Sie können von Unternehmen über Märkte angeboten werden und von den Menschen gekauft werden. Bei einigen Gütern, wie z.B. Bildung werden *Kredite* eine große Rolle spielen, d.h. Menschen investieren kreditfinanziert in ihr Humankapital und zahlen den Kredit über das höhere Einkommen dann zurück.

Zwangsversicherungen

Ich glaube nun nicht, dass der Durchschnittsamerikaner sich intensiv mit dem Libertarismus beschäftigt hat. Vielmehr scheint es so zu sein, dass allgemeine libertaristische Denkmuster den Mainstream bilden.

Das bedeutet nicht, dass die US-Amerikaner etwas gegen Krankenversicherung haben. Es bedeutet aber, dass sie etwas gegen *Zwangsversicherungen* haben, weil diese Zwangsversicherungen zum einen Zwang sind, zum anderen aber auch einen Umverteilungseffekt beinhalten. Im Bachelor haben wir uns mit dem Problem befasst, dass marktwirtschaftlich operierende Versicherungen risikoäquivalente Prämien kalkulieren müssen. Das würde dazu führen, dass chronisch Kranke keinen Krankenversicherungsschutz bezahlen können. Also geht es bei einer Pflicht zur Krankenversicherung immer um *Umverteilung*.

Umverteilung über einen Zwangsmechanismus einer Pflichtversicherung ist aber etwas, das aus libertaristischer Sicht nicht legitim ist. Wer freiwillig eine Krankenversicherung abschließen will, soll das tun, aber staatliche Zuschüsse (die ja über Steuern finanziert werden müssen) und Pflichtversicherungen sind abzulehnen.

F17.5 Konsequenzen für das Gesundheitssystem

Wenn die Freiheit von staatlichen Eingriffen eine Grundüberzeugung ist, legt das nahe, die Regelungsdichte im Gesundheitssystem so gering wie möglich zu halten. Für die Kassen, aber auch für die Leistungserbringer bedeutet das, daß sie großen rechtlichen Spielraum haben, neue Geschäftsmodelle zu entwickeln und auszuprobieren.

Der Anreiz, zu experimentieren ist für die Versicherungen groß, weil das System, wie gezeigt, extrem teuer ist und es für die Bevölkerung keinen Versicherungs*zwang* gibt.[11] Die Versicherungen sind also darauf angewiesen, dass sich genügend Arbeitgeber finden, die sie als Kunden gewinnen können, weil die Alternative für die Arbeitgeber ist, es mit der Krankenversicherung für die Mitarbeiter dann einfach bleiben zu lassen.

In Deutschland kann das Kostenniveau den Kassen ein Stück weit egal sein, da es einen Zwang zur Krankenversicherung gibt. In den USA gibt es dann einfach keine Nachfrage nach Versicherung mehr. Der Druck zur Kostendämpfung ist für die Versicherungen also viel höher.[12] Vor diesem Hintergrund ist es sinnvoll, sich einige dieser Geschäftsmodelle

[11] Hier ist es ähnlich wie mit einer Impfpflicht. Eine Impfpflicht beinhaltet eben eine Pflicht, bei der es die Option gibt, ihr nicht nachzukommen. Diese Pflichtverweigerung führt dann zu negativen Sanktionen, aber eine Impf*pflicht* besteht nicht darin, dass die Impf-Gestapo nachts die Tür eintritt und den Impfgegner ins Impfzentrum schleppt. Das wäre ein (drakonisch durchgesetzter) Impf*zwang*. Analog gibt es in den USA eben nur die Versicherungspflicht, von der man sich über die Strafzahlung quasi freikaufen kann.

[12] An dieser Stelle ist festzuhalten, dass es in erster Linie die *Versicherungen* sind, denen hier die Felle davonschwimmen, weniger die Leistungserbringer. Die Leistungserbringer können sich erst einmal

anzuschauen und auf mögliche Lehren für Deutschland zu prüfen. Das wird der Inhalt der nächsten Einheit sein.

halbwegs entspannt zurücklehnen und darauf spekulieren, dass ein Erkrankter irgendwann auch ohne Versicherung in seinem Wartezimmer sitzen wird.

Kapitel 18

Krankenversicherungskonzepte in den USA

F18.1 Kostenerstattung

Die traditionelle Krankenversicherung in den USA ist, wie die deutsche PKV nach dem Kostenerstattungsprinzip organisiert, das Sie aus dem Bachelor kennen. Allerdings gibt es in den USA kein Kapitaldeckungsverfahren.

Dieser Versicherungstyp wird *indemnity insurance* oder auch *fee for service plan* genannt. Im Bachelor haben wir gesehen, dass ein struktureller Nachteil dieses Konzepts darin besteht, dass es keine Vertragsbeziehung zwischen dem Leistungserbringer und der Versicherung gibt, sondern nur zwischen Patient und Versicherung und Patient und Leistungserbringer. Damit sind die Einflussmöglichkeiten der Versicherung auf die Kosten sehr gering.

F18.2 Managed Care

Die fehlenden Einflussmöglichkeiten der Versicherung in Kostenerstattungssystemen sind ein Grund für die hohen Versorgungskosten in den USA. Eine Idee, diese Kosten zu senken, besteht darin, dass die Versicherungen Organisationskonzepte entwickeln, die die diese Kosten senken sollen. Diese Konzepte werden unter dem Namen *Managed Care* subsumiert.[1] Es gibt eine große Zahl von Varianten, von denen ich Ihnen nur einige skizzieren will.

Im Bachelor haben wir über *managed competition* gesprochen. Das klingt sehr ähnlich, aber bei der „competition“ geht es um den *Wettbewerb* zwischen den Krankenversicherern, um Risikoselektion zu vermeiden und Anreize für Effizienzsteigerungen zu bieten.

Der Verbindungspunkt zwischen beiden Konzepten ist der, dass wir im Bachelor durchgerechnet haben, wie sich die Wettbewerbsposition einer Kassen mit schlechter Risikostruktur aber niedrigen Versorgungskosten darstellt. Auf die Frage, wie es *die Kasse* denn schaffen soll, die Kosten zu senken, bin ich damals immer mit *„das kommt noch im Master“* ausgewichen. Jetzt ist dieses *„das kommt noch“* gekommen. Wir haben vielleicht spontan Ideen, wie ein *Krankenhaus* Kosten senken kann - aber eine *Kasse*?

[1] Amelung, Volker Eric. Managed Care - Neue Wege im Gesundheitsmanagement, Wiesbaden: Gabler, 5. Auflage, 2012.

			rechtliche Beziehung zu den Versorgungseinrichtungen	
			gleiches Unternehmen	rechtlich selbständig, aber Exklusivvertrag
			staff model	group model
Kostenübernahme für externer Leistungserbringer	0%	closed panel		
	>0%	open panel		

Abbildung 18.1: HMO-Varianten

F18.3 Health Maintenance Organizations (HMO)

Grundidee

Die Grundidee der HMO ist, dass Versicherung und Leistungserbringung *aus einer Hand* kommen. Eine HMO betreibt eigene Versorgungszentren und Krankenhäuser. In diesen Einrichtungen werden nur eigene Versicherte behandelt.[2] Die Ärzte, die dort arbeiten, sind Angestellte der Versicherung.

Das Konzept wurde in den 1930er Jahren als eine Art integrierte Betriebskrankenkasse eines Bauunternehmens entwickelt und insbesondere in den 1970er Jahren politisch stark gefördert.

Der größte Anbieter dieses Versicherungskonzepts ist *Kaiser Permanente*. Das ist die Nachfolgeorganisation genau dieser BKK aus den 1930er Jahren.

HMOs ergeben nur in relativ dicht besiedelten Gebieten Sinn, weil sonst die Entfernungen zu den kasseneigenen Versorgungseinrichtungen zu groß werden. Aus diesem Grund sind HMOs in den USA ein Phänomen der Ballungsräume. In ländlichen Gebieten, z.B. im mittleren Westen gibt es keine HMOs. Die Relevanz der Bevölkerungsdichte wird uns noch in *Einheit Mögliche Lehren für Deutschland* beschäftigen.

Varianten

Das Grundkonzept der HMOs ist fast ein Jahrhundert alt. Daher haben sich auch Varianten der Grundidee entwickelt. Die wichtigsten davon habe ich in einer Matrix dargestellt.

Ein wichtiges Unterscheidungskriterium ist die Frage, zu welchen Bedingungen externe Leistungserbringer in Anspruch genommen werden können. In *closed panel HMOs* muß der Versicherte dann die kompletten Kosten selbst tragen, in *open panel HMOs* meist etwa 70%. Diese zweite Variante wird auch *point of service plan* genannt. In beiden Fällen ist der Anreiz, nur versicherungseigene Leistungserbringer aufzusuchen, hoch.

Das zweite Unterscheidungskriterium ist die Frage nach der rechtlichen Beziehung der Versicherung zu den Versorgungseinrichtungen. Im *staff model* sind die Ärzte direkt bei der Versicherung angestellt. Im *group model* sind die Ärzte bei einer *medical group*

[2]Natürlich gibt es Regelungen für Notfälle, aber solange man nicht plötzlich als Versicherter von Y vor einem Versorgungszentrum von X zusammenbricht, wird man dort nicht behandelt.

angestellt, die einen Exklusivvertrag mit der HMO hat. Ich muß gestehen, dass ich die Existenz dieser beiden Varianten nicht verstehe, bzw. keine *ökonomischen* Gründe dafür erkennen kann. Ich vermute daher, dass diese Unterscheidung *juristische* Gründe hat.

Diese Varianten existieren auch in Kombination, d.h. es gibt eine closed panel staff model HMO usw.

Strukturvereinfachung

Die HMO vereinfacht die Struktur des Systems, weil Finanzierung und Versorgung aus einer Hand kommen. Vergleichen wir eine HMO mit einem deutschen Krankenhaus. Die Kasse zahlt eine Fallpauschale an das Krankenhaus. Das Krankenhaus bezahlt das Personal, das den Fall abarbeitet.

Wir haben dann also eine in sich geschachtelte Principal-Agent-Beziehung, weil die Kasse Interessen hat, das Krankenhaus und auch das Personal. Wir haben gesehen, dass die finanziellen Rahmenbedingungen der Krankenhäuser einen großen Anreiz darstellen, viele Fälle zu generieren. Der Anreiz ist groß, den einzelnen Fall effizient abzuarbeiten, aber es werden eben *viele* Fälle sein.

Dieses Eigeninteresse des Krankenhauses (allgemein gesprochen: des Leistungserbringers) wird neutralisiert, weil er Teil des Versicherers ist. Der Anreiz zu *angebotsinduzierter Nachfrage* fällt also weg.

Attraktivität für die Beteiligten

1. Die *Versicherungen* haben einen Anreiz, in die Versorgungseffizienz zu investieren, weil sie eine eigene Versorgungsinfrastruktur haben, die sie nicht mit Konkurrenten teilen müssen. Sie können unternehmensinterne Leitlinien für die Behandlung ihrer Patienten entwickeln. Auf diese Weise können sie einen exklusiven Wettbewerbsvorteil bei den Prämien erreichen, der sie für ihre Kunden attraktiv macht.

2. Die *Patienten* müssen eine eingeschränkte Arztwahl hinnehmen und von Ärzten behandelt werden, die von einer Überversorgung nicht mehr finanziell profitieren, weil die Angestellte der Versicherung sind. Dafür sind die Prämien niedriger.

3. Die *Leistungserbringer*... Siehe Frage am Ende der Einheit.

F18.4 Preferred Provider Organization

Konzept

Durch die staatliche Förderung der HMOs und die niedrigeren Prämien stieg der Marktanteil dieses Konzepts in den 1970er Jahren und setzte die selbständigen Leistungserbringer, die weiterhin nach dem Kostenerstattungsprinzip mit den Versicherungen abrechnen wollten, unter ökonomischen Druck. Dieser Druck führte dazu, dass sich Ärzte mit eigener Praxis zusammenschlossen um Verhandlungen mit Versicherungen zu führen.

Das Arrangement mit den Versicherungen besteht aus zwei Teilen:

1. Die teilnehmenden Leistungserbringer räumen der Versicherung bei der Abrechnung erhebliche Rabatte ein.

2. Im Gegenzug schafft die Versicherung für die Patienten große Anreize, einen der teilnehmenden Ärzte aufzusuchen. Dieser Anreiz besteht darin, dass die Behandlung durch einen teilnehmenden Arzt (einen Preferred Provider) für den Patienten kostenlos (bzw. günstiger) ist als die Behandlung durch einen Leistungserbringer der kein Preferred Provider ist.

Spielarten

- Eine *Preferred Provider Organization* (PPO) behält eine grundsätzlich freie Arztwahl bei. Bei den Leistungserbringern, die keine preferred providers sind, muss der Patient allerdings einen Großteil der Behandlungskosten selbst tragen.
- Eine *Exclusive Provider Organization* (EPO) beschränkt die Arztwahl auf die Ärzte, mit denen Rabatte vereinbart wurden. Der Unterschied zu einer closed panel HMO besteht darin, dass die Leistungserbringer in einer EPO immer noch selbständig sind und mit der Versicherung über Kostenerstattung abrechnen.

Diese beiden Varianten zeigen, dass der Wettbewerb der Versicherungskonzepte dazu führt, dass man Aspekte jeweils anderer Konzepte übernimmt. So ist die PPO eine Reaktion auf die HMOs und die open panel HMOs eine Reaktion auf die PPOs und die EPOs wiederum eine Reaktion auf die closed panel HMOs.

Attraktivität für die Beteiligten

1. Die *Versicherung* muß wenig an ihren Prozessen ändern, weil sie nach wie vor mit den Leistungserbringern über Kostenerstattung abrechnen. Die Honorare mit den Preferred Providers sind allerdings niedriger. Sie muß keine eigene Versorgungsstruktur aufbauen, ist also auch in der Lage, in Regionen mit geringer Bevölkerungsdichte ein Angebot machen zu können.

2. Die *Ärzte* können über die Gewährung von Rabatten ihre Auslastung sichern, ohne ihre Selbständigkeit aufzugeben. Auch sie müssen wenig an ihren Prozessen ändern, da das Honorierungskonzept sich (bis auf die Höhe) nicht ändert. Die Ärzte sind nicht exklusiv an die PPO gebunden, sondern können auch Patienten, die Kunden bei anderen Versicherungen sind, behandeln.

3. Der *Patient* hat (im Vergleich zur closed panel HMO) bessere Bedingungen für freie Arztwahl. Die Prämien der PPO sind niedriger als bei der Standard-Kostenerstattungs-Versicherung.[3] Dafür akzeptiert er die Einschränkungen der vollkommen freien Arztwahl.

Die Erfahrungen der PPOs sind, dass etwa 75% der Arztkontakte auf Anbieter auf der Liste entfallen. Das bedeutet, dass die verbesserten Möglichkeiten für eine freie Arztwahl eine *Option* sind, für die man (im Vergleich zur closed panel HMO) einen Prämienaufschlag akzeptiert, die aber selten genutzt wird.[4]

[3] Genauer gesagt: Die Prämien der PPO *müssen* niedriger sein, denn sonst wäre kein Versicherter bereit, sich bei seiner Arztwahl in erster Linie auf die Liste der preferred provider zu beschränken.

[4] Auch hier: Die Prämien der closed shop HMO *müssen* niedriger sein als bei der PPO, weil sonst kein Anreiz für die Versicherten besteht, weitere Einschränkungen bei der Arztwahl auf sich zu nehmen.

F18.5 Vorwärts- und Rückwärtsintegration

Make or buy

Mit diesen beiden Konzepten sind wir letztlich wieder bei der ganz grundlegenden Frage des *make or buy*.

Dort ist das Thema der vertikalen Integration angeschnitten worden, also der Frage, ob es für ein Unternehmen sinnvoll sein kann, zum eigenen Kunden oder Lieferanten zu werden.[5] Genau das ist, was bei HMOs und PPOs passiert.

Vorwärts- vs. Rückwärtsintegration

Das Interessante an PPOs und HMOs ist, dass bei solchen Integrationsprozessen überhaupt nicht ausgemacht ist, ob eine Vorwärtsintegration oder eine Rückwärtsintegration erfolgt, d.h. ob Versicherer eigene Versorgungsstrukturen aufbauen (HMO) oder Leistungsanbieter eigene Versicherungen (PPO).

Begriffsverwirrung

An dieser Stelle sei kurz auf eine mögliche Begriffsverwirrung hingewiesen. *„Versorgungsintegration“* und *„integrierte Versorgung“* bezeichnen in Deutschland im Regelfall keine *betriebswirtschaftlichen* Prozesse im Sinne einer porter´schen vertikalen Integration, sondern eine *verbesserte Kommunikation* zwischen den ambulanten und stationären Sektoren der Gesundheitsversorgung.

F18.6 Einschätzung

Risikoselektionseffekte

Es lässt sich empirisch zeigen, dass PPOs und HMOs niedrigere Prämien anbieten können als die klassischen Kostenerstatter. Die Frage ist allerdings, welchen Anteil *Risikoselektion* an den Einspareffekten haben. Für die Schweiz, deren Strukturen etwas anders sind, liegen die Schätzungen bei einem Anteil von 20%.[6] Diese Risikoselektion entsteht dadurch, dass die Einschränkungen bei der freien Arztwahl umso weniger relevant sind, je seltener die Versicherten zum Arzt gehen, sprich, je gesünder sie sind. Je gravierender die Einschränkungen der Inanspruchnahme, umso unattraktiver ist das Versicherungsangebot für Hochkostenpatienten. Die Prämien sind dann (zumindest teilweise) nicht deshalb so niedrig, weil die Versorgung so effizient ist, sondern weil der Anteil der Hochkostenpatienten sehr niedrig ist.

Rosinenpicken

Wir haben das US-System als experimentierfreudig kennengelernt. Das bedeutet aber andersherum, dass der Anspruch an die Versicherungen nicht ist, eine Lösung zu entwickeln, die für die gesamte Bevölkerung funktioniert.

---

[5] Porter, Michael E. Wettbewerbsstrategie, Frankfurt: Campus, 12. Auflage, 2013.

[6] Lehmann, Hansjörg. Managed Care - Kosten senken mit alternativen Versorgungsformen, Chur: Rüegger, 2003.

Die Versicherungen können also Lösungen für mehr oder minder kleine Teilgruppen entwickelt und müssen sich nicht um die „Versorgung in der Fläche“ kümmern. Die Konzepte betreiben also eine Art Rosinenpickerei.

Um den Personenkreis, der nicht zu dieser Teilgruppe der Rosinen gehört, muss ich mich nicht kümmern. Wenn dann 20% der Bevölkerung durch das Raster fällt, ist das nicht mein Problem.

F18.7 Ehemalige Prüfungsaufgaben

Sicht der Ärzte

Im Text des Kapitels ist die Sicht der Ärzte auf das Konzept der HMO ausgespart. Wie sehen die Ärzte dieses Konzept? Gibt es überhaupt „die“ Ärzte, oder gibt es da verschiedene Untergruppen?

Sicht der Ärzte (Lösung)

- Als Angestellter trägt der Arzt kein unternehmerisches Risiko.
  - Er muss nicht in die Ausstattung einer eigenen Praxis investieren
  - Er muss auch keine eigene Akquisition von Patienten betreiben
- Etablierte Ärzte haben dieses Risiko schon auf sich genommen
  - Sie haben in ihre Praxis investiert
  - Sie haben einen Patientenstamm aufgebaut
- Diese Investition etablierter Ärzte wird durch ein Angestelltenverhältnis entwertet
- Der Wegfall des unternehmerischen Risikos ist mit einer niedrigeren Honorierung verbunden
- Das Angestelltenverhältnis bedeutet einen Verlust an Autonomie Eine HMO ist also nur für Berufsanfänger attraktiv und für ältere Ärzte, deren Praxis nicht gut läuft.
- Dem HMO-Arzt muß klar sein, daß er relativ wenig verdienen wird und weniger Autonomie haben wird.

Kapitel 19

Das Gesundheitssystem in Großbritannien

F19.1 Volkswirtschaftliche Rahmendaten

In *Einheit Das Gesundheitssystem der USA* haben wir die Pro-Kopf-Ausgaben für Gesundheit und den Anteil der Gesundheitsausgaben am BIP für eine Reihe von Ländern miteinander verglichen. Die Pro-Kopf-Ausgaben in Großbritannien liegen um etwa 20% unterhalb des deutschen Niveaus, während das Pro-Kopf BIP in Großbritannien etwa 10 % unter dem deutschen liegt.

Das britische Gesundheitssystem, der *National Health Service* (NHS) ist steuerfinanziert und versorgt die gesamte Bevölkerung Großbritanniens. Es gibt kein paralleles, alternatives privates Versicherungsmodell wie in Deutschland.

Die Steuerfinanzierung setzt das Konzept des Leistungsfähigkeitsprinzips konsequent um, da es, anders als in Deutschland, keine Deckelung der Beiträge gibt, weil es gar keine für das Gesundheitssystem gesondert berechneten Beiträge gibt, sondern die Finanzierung des Systems aus dem allgemeinen Steuerhaushalt erfolgt.

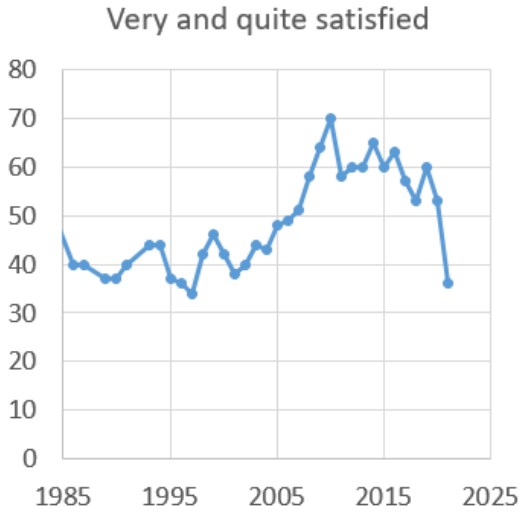
F19.2 Zufriedenheit mit dem System

Der NHS hat in Deutschland seit Jahrzehnten einen sehr schlechten Ruf und wird häufig als abschreckendes Beispiel staatlicher Mangelwirtschaft und Rationierung verwendet. Die Briten nehmen ihr eigenes Gesundheitssystem anders wahr.

Langzeitbetrachtung

Traditionell sind die Briten extrem zufrieden mit Ihrem Gesundheitssystem. Der Thinktank *King's Fund* führt seit Jahrzehnten Umfragen zur Zufriedenheit mit dem NHS durch. Die lange Reihe mit dem Anteil der „ziemlich und sehr Zufriedenen“ über knapp 40 Jahre[1] sieht so aus:

[1] https://www.kingsfund.org.uk/blog/2022/03/public-satisfaction-nhs-falls-25-year-low, eigene Berechnungen.



Diese Zahlen hängen natürlich ein wenig in der Luft, wenn man keine Vergleichszahlen hat. Vielleicht sind die 70% Zufriedenheit im Jahr 2010 ja vergleichsweise wenig, weil der Anteil von Menschen in ..., die zufrieden sind, noch viel höher liegt. Der *Commonwealth Fund*, ein US-Thinktank führt mit größeren Abständen ähnliche Umfragen in mehreren Ländern durch. Die letzte Befragung fand 2016 statt. Die ähnlich gelagert Einschätzung

> *The System Works Pretty Well and Only Minor Changes Are Necessary to Make It Work Better*[2]

teilten in Deutschland 60% und in den USA 19%. Für Großbritannien lagen für dieses Jahr leider keine Daten vor.

Bei allen Ungenauigkeiten lässt sich aber trotzdem festhalten, dass die Briten mit dem britischen System zufriedener sind als die deutschen mit dem britischen System. Und (ein Beifang) wir können sehen, wie unzufrieden die US-Amerikaner mit ihrem System sind.[3]

Größe des Parallelmarkts

Ein weiterer Indikator für die Zufriedenheit sind die Höhe der Ausgaben für private Krankenversicherung. Wären die Briten mit den Leistungen des staatlichen Gesundheitssystems großflächig unzufrieden, würde sich als eine Art Überdruckventil ein großer Parallel- und Zusatzmarkt privater Anbieter bilden, denn es gibt in Großbritannien keine *starke* Rationierung, d.h. ein solcher Privatmarkt wäre legal. Es gibt aber keinen nennenswerten Parallelmarkt. Die gesamten freiwilligen privaten Gesundheitsausgaben liegen etwa auf deutschem Niveau und der Anteil der Ausgaben für private Zusatzversicherungen liegt bei 2,3% aller Gesundheitsausgaben.[4]

[2]https://www.commonwealthfund.org/international-health-policy-center/system-stats/system-works-pretty-well

[3]Das beißt sich ein wenig mit dem starken Widerstand, den Obamacare bekommen hat. Es sieht so aus, als würde man mit dem aktuellen Stand unzufrieden sein, die möglichen Lösungen aber auch ablehnen. Das scheint also kein deutsches Phänomen zu sein.

[4]https://www.ons.gov.uk/peoplepopulation andcommunity/ healthandsocialcare/ healthcaresystem/ datasets/ healthaccountsreferencetables

Jüngere Entwicklungen

Wenn Sie sich die lange Reihe der Zufriedenheitswerte weiter oben anschauen, können Sie sehen, dass die jüngsten Werte (2020 und 2021) stark gefallen sind. Diese Entwicklung ist schwer zu interpretieren. Es könnte sein, dass die jahrzehntelange (deutsche) Kritik an dem System sich jetzt doch bewahrheitet.[5] Der NHS wäre dann ein unterfinanzierter Koloss, der seit langem auf tönernen Füßen steht und irgendwann umkippen muss. Das „irgendwann" ist eben jetzt. Die Briten selbst haben das spät gemerkt, aber immerhin jetzt. Wir in Deutschland wussten das vor den Briten.

Diese Lesart *könnte* zutreffen. Was mich aber skeptisch macht ist, dass dieser Einbruch an Zufriedenheit exakt in die Zeit der Corona - Epidemie fällt.

Wir haben wir das Konzept der *Warteschlange* kennengelernt. Der Ausgangspunkt der Idee von Warteschlangen war ja, dass die Nachfrage nach Gesundheitsleistungen zeitlich nicht konstant ist, sondern schwankt. Wenn man die Kapazitäten nun an einer theoretischen Maximalnachfrage ausrichtet, bedeutet das, dass man enorme Leerkapazitäten finanzieren muss. Man kann Geld sparen, indem man sich eben nicht an der Maximalnachfrage orientiert, sondern in Kauf nimmt, dass es Zeiträume gibt, in denen die Nachfrage größer als die Kapazitäten sind. Dann verschiebt man die Behandlung eines Teils der Patienten in ein Folgequartal. Das ist der Preis, den man dafür zahlen muss, die Kapazitäten etwas heruntergefahren zu haben, um den Auslastungsgrad zu erhöhen. Wir wissen: Das kann nicht permanent so sein, weil die Warteschlange dann immer länger werden würde.

Die gleiche Mechanik ist durch Corona in Bewegung gekommen. Der Unterschied ist aber, dass man die Zahl von Oberschenkelhalsbrüchen besser prognostizieren kann (mehr, wenn es draußen glatt ist) als den Ausbruch einen Virus. Mit Corona befinden wir uns mit einem Fuß in der *Katastrophenmedizin*, über die wir schon gesprochen haben.

Wenn wir (konservativ) annehmen, dass sich in einer Katastrophe die Nachfrage nach Leistung x *verdoppelt* und wir die Erwartung haben, dass das System dann nicht ins Schlingern gerät, müssen wir akzeptieren, dass unsere Auslastung in Nicht-Katastrophen-Zeiten *immer* unter 50% liegen wird. Diese Kapazitäten müssen wir natürlich finanzieren. Und zwar nicht nur dann, wenn wir sie brauchen, sondern auch dann, wenn wir sie *nicht* brauchen. Und wir müssen sie auch für Katastrophen vorhalten, deren Charakter wir noch gar nicht kennen.

Die Pinguingrippe, die (ich weiß das) 2032 ausbrechen wird, überträgt sich über das Wassser. Wir können uns schützen, indem wir spezielle Wasserfilter verwenden, die speziell den Pinguinvirus ausfiltern. Wir haben jetzt die Wahl, ob wir für jede Katastrophe, die sich irgendjemand ausdenkt, Vorsorge treffen, und Pinguinfilter, Gewehre gegen Zombies, Laserkanonen gegen Aliens, ... auf Lager halten oder uns 2032 das Geheul anhören, dass sich niemand um die Pinguine gekümmert hat, obwohl der Winkelhake dass schon vor Jahren angekündigt hat.

Bei den volkswirtschaftlichen Rahmendaten haben wir bereits gesehen, dass das britische System mit vergleichsweise wenig Ressourcen betrieben wird. Das scheint auch das zu sein, was die britische Bevölkerung will. Es ist nun nicht überraschend, dass ein System, dass im Normalbetrieb ziemlich gut ausgelastet ist, in einer Pandemie in größere Schwierigkeiten gerät als ein System, das größere Reserven hat. Dies führt mich zu einer möglichen alternativen Leseweise der Unzufriedenheit:

[5] Buchsteiner, Jochen, "Dreißig Krankenwagen in der Schlange", Frankfurter Allgemeine Zeitung 29.12. 2022. S.3.

> *Das System ist halbwegs gut und billig. Das finde ich gut. In der Krise bin ich aber unzufrieden, weil das System so billig ist, dass es keine Reserven gibt. Jetzt fände ich es aber gut, wenn wir früher mehr Geld ausgegeben hätten und mehr Kapazitäten vorgehalten hätten. Wenn man das jetzt aber als Lehre aus Corona machen würde, werde ich nächstes Jahr auch wieder unzufrieden sein, weil man mehr Geld ausgibt als unbedingt notwendig, weil wir dann Überkapazitäten haben, weil es keine Pandemie gibt.*

Vor diesem Hintergrund wird es spannend sein, die Zufriedenheitswerte in den kommenden Nicht-Pandemie-Jahren zu beobachten. Sollte die zweite Leseweise zutreffen, werden sie nach oben gehen, ohne dass sich irgendetwas am System ändern müsste.

F19.3 Hausarztversorgung

Jeder Einwohner muss sich bei einem Hausarzt (General Practitioner) einschreiben, der für ihn die erste Anlaufstelle ist. Ein direkter Zugang von Patienten zu Spezialisten und Krankenhäusern ist (außer in Notfällen) nicht zulässig, sondern nur nach Einweisung durch den Hausarzt. Der Hausarzt erhält für jede Person, die sich bei ihm einschreibt, eine Pauschale. Die Hausärzte sind teilweise selbständig, teilweise angestellt. Sie müssen eine fünfjährige Zusatzausbildung absolvieren. Ihr Einkommen ist typischerweise etwas höher als in Deutschland.

Die Hausarztpraxen sind größer als in Deutschland. In der medizinischen Versorgung wird der Arzt durch *practice nurses* unterstützt, die eine akademische Ausbildung haben und im Qualifikationsniveau zwischen der deutschen Arzthelferin und dem deutschen Arzt stehen. Sie können u.a. einfachere ambulante chirurgische Eingriffe selbst durchführen und Standardfälle in Eigenregie therapieren. Ein Patient wird daher nicht notwendigerweise vom Arzt selbst behandelt, sondern nur, wenn die Komplexität des Falls die Qualifikation der practice nurse übersteigt. Die practice nurse dient quasi als vorgelagerter Filter für den Arzt. Da ihr Einkommen geringer ist als das des Arztes und der Arzt nur noch die komplexen Fälle behandelt, sind die Versorgungskosten relativ niedrig.

F19.4 Warteschlangen

Wir haben Warteschlangen als Instrument besserer Kapazitätsauslastung und als Rationierungsinstrument bereits diskutiert. Ein großer Teil der (insgesamt geringen) Unzufriedenheit der Briten mit ihrem System bezieht sich auf diesen Punkt.

In seiner oben zitierten Umfrage aus dem Jahr 2010 untersuchte der Commonwealth Fund auch die Länge der Wartezeiten, die in der Abbildung unten dargestellt sind.

Bei den Facharztterminen sind die Wartezeiten geringer als in Deutschland. Ein Grund dafür ist, dass es keinen direkten Zugriff der Patienten auf Fachärzte gibt. Bei den geplanten chirurgischen Eingriffen sind die Wartezeiten etwas länger als in Deutschland.

Die Wartezeiten sind im Zeitablauf relativ konstant geblieben, so dass ein häufig vermuteter Ärztemangel nicht der Grund sein kann, da dann die Wartezeiten kontinuierlich länger werden müssten.

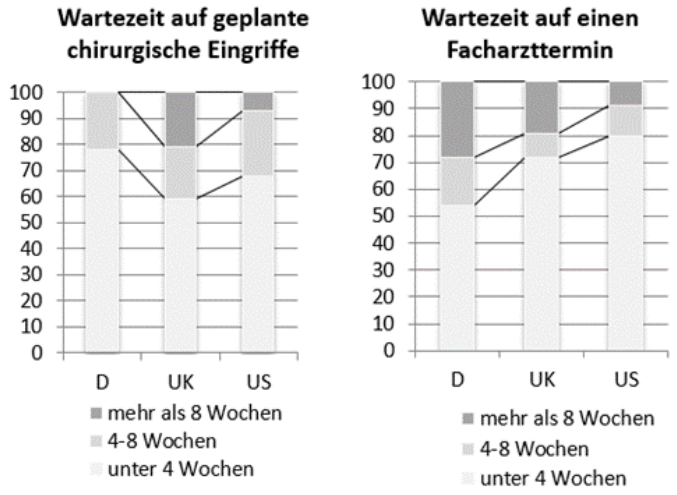


Abbildung 19.1: Warteschlangen im Ländervergleich

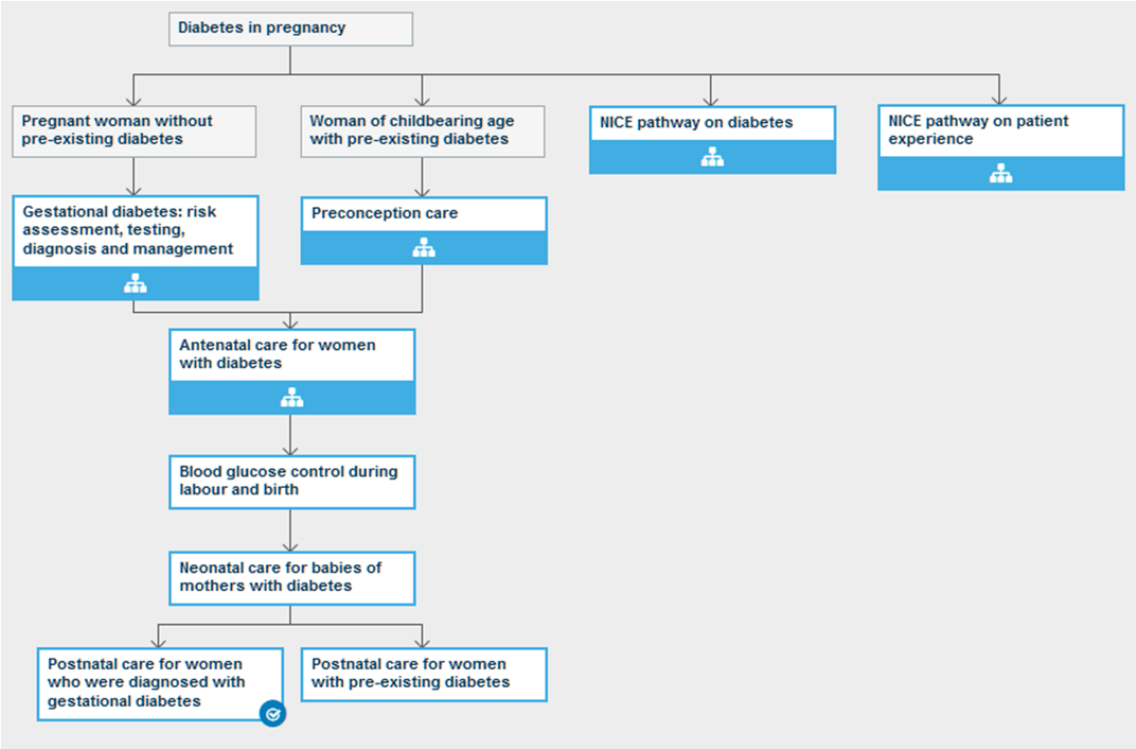


Abbildung 19.2: Therapierichtlinie „Schwangerschaftsdiabetes“

F19.5 Effiziente Leistungskataloge und Richtlinien

NICE

Da der NHS ein Einheitssystem ist, gibt es auch einen einheitlichen Leistungskatalog. Das *National Institute for Health and Care Excellence* (NICE) legt verbindliche Therapierichtlinien fest. Diese Therapierichtlinien sind über das Internet frei zugänglich, wie z.B. die folgende Richtlinie für Schwangerschaftsdiabetes, die über den Browser auch durch die Patienten „durchklickbar“ sind.

„Richtlinie“ bedeutet, dass es einen bestimmten Anfangspunkt gibt (die Schwangere hat einen erhöhten Blutzuckerspiegel), der dann weiter differenziert wird. War sie schon vorher Diabetikerin? Wie hoch ist der Blutzuckerspiegel? Wie hoch ist ihr BMI usw. weil es nicht *die* Universaltherapie für alle Patienten gibt.

QALYs

Zur Erstellung dieser Richtlinien verwendet das NICE das Instrument der QALYs, das in *Einheit Ergebnisdefinitionen* vorgestellt worden ist. Therapien, die bezüglich einer Aufnahme in den NHS-Leistungskatalog untersucht werden, müssen einen Schwellenwert von 30.000 £/QALY unterschreiten. Liegt der Preis darüber, wird die Therapie nur in begründeten Ausnahmefällen in den Leistungskatalog aufgenommen.

F19.6 Ehemalige Prüfungsaufgaben

Schwester Agnes

Schwester Agnes ist ein DDR-Film aus dem Jahr 1975, in dem es um eine Gemeindeschwester geht, die mit ihrem Moped über sächsische Dörfer fährt, Patienten betreut und sich mit sturen SED-Bonzen anlegt. Für einen DDR-Film war das sehr kritisch und ist daher im (ostdeutschen) Gedächtnis geblieben.

In Anlehnung an den Titel dieses Films gab es in Brandenburg ein Modellprojekt AGneS.[6] Die Idee dieses Modells ist, die Versorgung von Patienten in ländlichen Regionen mit Ärztemangel zu verbessern. Weitergebildete Krankenschwestern übernehmen Tätigkeiten, die in anderen (besser versorgten) Regionen durch Ärzte durchgeführt werden. Dieses Modell greift letztlich Strukturen auf, wie sie in der ambulanten Versorgung in Großbritannien durch die practice nurses seit Jahrzehnten existieren.

AGnES ist von Standesvertretern der Ärzte stark kritisiert worden.

Ist das nicht überraschend? Wir sind doch zu dem Schluss gekommen, daß Ärzte in Großbritannien ziemlich gut verdienen, weil sie Leistungen, die die practice nurse erbringen kann, auf diese abwälzen und sich auf die (besser bezahlten) Leistungen konzentrieren, die ein komplettes Medizinstudium erfordern. Die deutschen Ärzte müssten dann doch froh sein und schon die Dollarzeichen in den Augen haben. Warum sind sie gegen Schwester Agnes? Oder sind gar nicht alle Ärzte gegen Schwester Agnes?

[6] Arztentlastende, gemeindenahe, E-Health-gestützte Systemintervention. Ein schönes Beispiel dafür, daß man jedes Akronym, das man haben möchte, irgendwie hingebastelt bekommt.

Schwester Agnes (Lösung)

Wenn eine Teilversorgung durch ambulante Gemeindeschwestern gut funktioniert, ist die naheliegende Schlussfolgerung, daß die Hausärzte grundsätzlich für diesen Teil der medizinischen Versorgung überqualifiziert und überbezahlt sind und man die Arbeit genauso gut von (geringer bezahlten) Gemeindeschwestern erledigen lassen kann.

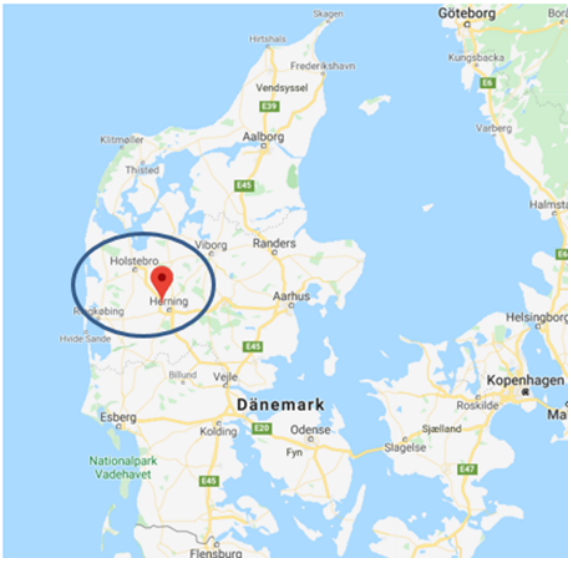
Das gilt natürlich nicht nur für die unterversorgten Regionen, sondern für alle Regionen. Die Nachfrage nach Hausärzten würde also stark zurückgehen. Weniger Mediziner würden sich auf diesen Bereich spezialisieren und sich eher als Facharzt spezialisieren, was den ohnehin überbesetzten Facharztsektor noch stärker unter Druck setzt.

Kapitel 20

Die Krankenhausreform in Dänemark

F20.1 Superkrankenhäuser

Dänemark hat 5,7 Mio. Einwohner. Bis 2025 soll die stationäre Versorgung komplett von etwa 20 sogenannten Superkrankenhäusern übernommen werden, die dann jeweils für etwa 300.000 Bewohner zuständig sind.[1] Ein Großteil dieser Superkrankenhäuser sind komplette Neubauten, wie z.B. in Godstrup.[2]



Dieser Neubau (mit eigener Autobahnabfahrt) soll West-Jütland versorgen und fünf bestehende Krankenhäuser ersetzen, die bis zu 70km entfernt liegen.[3] Im Bachelor haben Sie sich mit der Krankenhauslandschaft in Deutschland befasst. Daher wissen Sie, daß dieses Konzept sich ganz deutlich von der deutschen Strategie unterscheidet, die möglichst jedes Krankenhaus in strukturschwachen Regionen erhalten möchte, um eine *wohnortnahe* stationäre Versorgung zu gewährleisten.

Ausstattung und Wohnortnähe

Ein Grund für die starke Zentralisierung der Krankenhäuser ist, daß die Ausstattung der Superkrankenhäuser viel besser sein kann als die der kleinen Häuser in der Fläche.

[1] Zum Vergleich: In Deutschland liegt diese Quote bei etwas über 40.000 Einwohner/Krankenhaus.
[2] Einen guten Einblick bietet Baltzter, Sebastian, "Dänemarks eilige Retter", FAS 19.1 2020. S.20.
[3] Die Karte stammt aus Google Maps. Das Einzugsgebiet von Godstrup habe ich grob eingezeichnet.

Es gibt also einen negativen Trade-Off zwischen Nähe und Ausstattung. Je mehr Häuser es gibt, umso geringer die Möglichkeiten, teure Ausstattung zu refinanzieren, weil die Häuser zu klein für die Auslastung der Ausstattung sind. Die Versorgung ist dann zwar schneller erreichbar, aber qualitativ nicht besonders hoch. Die Frage ist, ob es in einem Notfall besser ist, schnell in einem schlecht ausgestatteten Krankenhaus eingeliefert zu werden oder 15 Minuten später in ein maximal ausgestattetes Krankenhaus.

Größe und Auslastung

Ein zweiter Vorteil größerer Krankenhäuser ist, dass die Auslastung besser geplant werden kann. Das hat nichts mit Professionalität zu tun, sondern ausschließlich mit Statistik und dem Gesetz der großen Zahl, das wir aus dem Kontext des Stichprobenfehlers kennen.

Die Vorteile aus der Zusammenfassung kleinerer Einheiten in eine größere finden wir an vielen Stellen außerhalb des Gesundheitssystems. Ein Beispiel dafür sind *Callcenter*. Erst einmal erscheint es fragwürdig, warum Unternehmen diese Funktion auslagern sollen. Sie könnten ihre Telefonhotline doch auch selbst betreiben. Das Callcenter ist doch bestenfalls genauso kompetent wie eigene Mitarbeiter. Der Vorteil ist, daß Callcenter mehrere Kunden haben und wenn die Hotline bei Kunden A heißläuft, ist bei anderen Kunden im Regelfall weniger los, so daß die Kapazitäten immer dorthin fließen, wo es viele Anrufe gibt. Über diesen Mechanismus können die Callcenter einen besseren Service mit weniger Ressourcen (und damit Kosten) anbieten als die Kunden dies jeweils selbst könnten. Der Effekt ist der gleiche wie beim Superkrankenhaus.

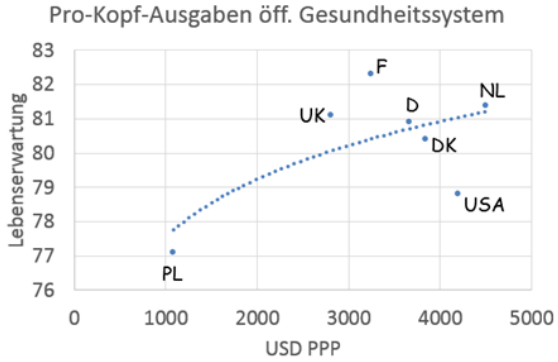
Kapitel 21

Mögliche Lehren für Deutschland

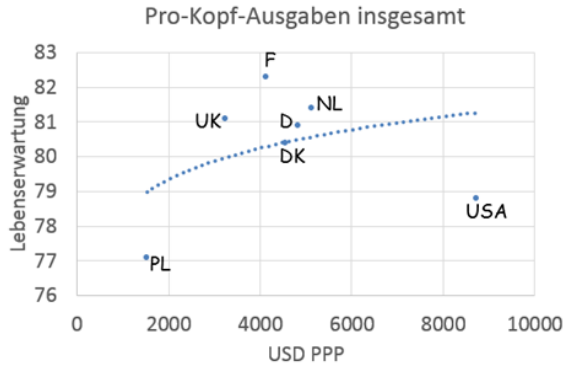
F21.1 Internationaler Vergleich

Kennzahlen wie die Lebenserwartung machen nur im Vergleich mit Vergangenheitswerten oder Werten anderer Länder Sinn. Bei den Kopfpauschalen in *Einheit Pauschalen und Risiko* haben wir uns schon überlegt, welche Einsparziele realistisch sein könnten. An dieser Stelle sind wir nun wieder - und zwar auf Ebene des Gesamtsystems eines Landes, denn es ist ja nicht sinnvoll, Systemveränderungen mit unrealistisch hohen Erwartungen zu überfrachten.

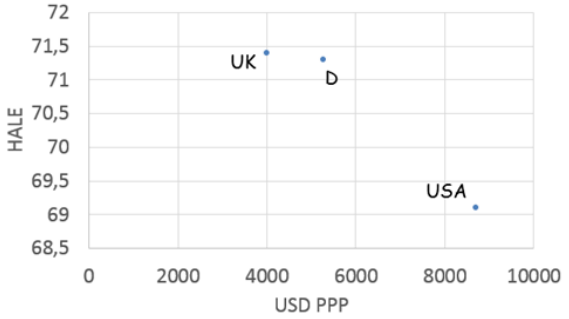
Die folgende Grafik stammt aus OECD-Daten aus dem Jahr 2015 und zeigt Lebenserwartung und Pro-Kopf-Ausgaben im Rahmen des öffentlichen Gesundheitssystems für eine Auswahl verschiedener Länder an.



Die westeuropäischen Nachbarländer Deutschlands liefern ähnliche Werte wie das deutsche System. Polen, als Beispiel für ein ehemaliges Ostblockland, gibt pro Kopf deutlich weniger aus, hat aber eine niedrigere Lebenserwartung. Die USA fallen aus diesem Schema heraus. Das System hat sehr hohe Ausgaben bei relativ niedriger Lebenserwartung. Dieses Ergebnis wird noch deutlicher, wenn man die Pro-Kopf Gesamtausgaben betrachtet, wie in der folgenden Grafik aus gleicher Quelle. Hier wird deutlich, wie groß der Anteil der privaten Zahlungen in den USA ist.



Es könnte nun sein, daß die numerische Lebenserwartung kein guter Leistungsindikator ist, weil sie die Lebensqualität nicht berücksichtigt. Mit dem Instrument der *Health Adjusted Life Expectancy* (HALE) hat die WHO ein Instrument entwickelt, das die Lebensqualität berücksichtigt. Für Deutschland, Großbritannien und die USA sieht der Zusammenhang zwischen HALE und gesamten Pro-Kopf-Gesundheitsausgaben für 2014/15 wie folgt aus.



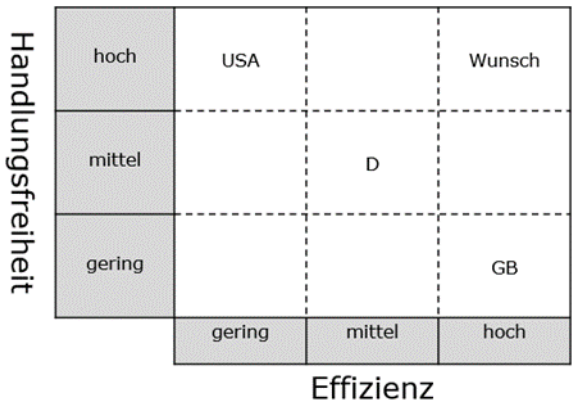
Bei allen Ländern, für die Daten verfügbar waren, zeigte sich, daß die HALE ziemlich konstant 8-10 Jahre unter der numerischen Lebenserwartung lag.

Wenn man HALE als Kennzahl für den Erfolg eines Gesundheitssystems interpretiert, dann liegt die Effizienz des deutschen Systems zwischen der des britischen und des US-amerikanischen Systems, wobei die Effizienz des US-Systems relativ gering ist.

Abschätzung von Effizienzreserven

Nimmt man Ausgaben und Lebenserwartung anderer westlicher Industrieländer als Benchmark, kommt man zu dem Ergebnis, dass möglicherweise vorhandene Effizienzreserven im deutschen Gesundheitssystem überschaubar sind. Die Nachbarländer Deutschlands geben ähnlich viel Geld aus und werden ähnlich alt. Das britische System ist europaweit spürbar effizienter. Es kostet (bei höherer HALE) etwa 1.000€ pro Kopf weniger als das deutsche. Diese 1.000€ könnte man daher als Obergrenze für mögliche Effizienzreserven ansehen.

Handlungsfreiheit, Effizienz und Zufriedenheit



Der Vergleich der Systeme der USA, Deutschlands und Großbritanniens legt nahe, daß es einen negativen Trade-Off zwischen der Handlungsfreiheit der Akteure und der Effizienz des Systems gibt.

Die hohen Zufriedenheitswerte des NHS sind ein Indikator dafür, daß die Briten die niedrige Handlungsfreiheit als Preis der hohen Effizienz akzeptieren und mit einem „no-frills“ Gesundheitssystem gut leben können. Die Mittelstellung des deutschen Systems gepaart mit der etwas größeren Unzufriedenheit ist ein Indiz dafür, daß eine gesellschaftliche Diskussion über die Ziele des Gesundheitssystems in Deutschland aussteht. Die Wunschvorstellung eines effizienten (d.h. kostengünstigen) Systems mit hoher Handlungsfreiheit der Akteure ist unrealistisch.

Der Sinn von Wettbewerb zwischen Leistungserbringern

Die Möglichkeit der Kunden, zwischen verschiedenen Anbietern zu wählen, besitzt in unserer Gesellschaft einen hohen Wert. Die Anbieter werden unter Druck gesetzt, sich stärker an den Wünschen der Kunden zu orientieren.

Dieses Bild von Wettbewerb setzt Überkapazitäten voraus. Andernfalls können die Anbieter ihre Leistungen auch ohne große Kundenorientierung absetzen.

Die Anbieter haben ein Interesse, ihre Kapazitäten auszulasten. Im Gesundheitssystem können sie das über eine Leistungsausweitung im Rahmen der angebotsinduzierten Nachfrage. Das bedeutet, dass die Anbieter vorhandene Überkapazitäten sukzessive abbauen. Bei freiem Zugriff auf die Fachärzte wird der nächste Facharzttermin *immer* acht Wochen in der Zukunft liegen. Unabhängig von der Zahl der Fachärzte.

Der Nutzen von Wettbewerb zwischen Krankenkassen

Derzeit gibt es in Deutschland noch etwa 100 Krankenkassen. Der derzeitige Nutzen dieser hohen Anzahl ist unklar. Die Kassen unterscheiden sich weder in Prämienkalkulation noch im Leistungskatalog. Beides ist gesetzlich fixiert. Die Kassen nutzen auch die gleiche Versorgungsinfrastruktur. Das bedeutet, dass es keine relevanten Wettbewerbsparameter für die Anbieter gibt.

Vor diesem Hintergrund gibt es m.E. nur zwei sinnvolle Entwicklungspfade:

1. Die Kassen fusionieren mittelfristig zu einer Einheitskasse nach dem Muster des NHS. Diese Einheitskasse muss nicht steuerfinanziert sein, aber eine Einheitskasse

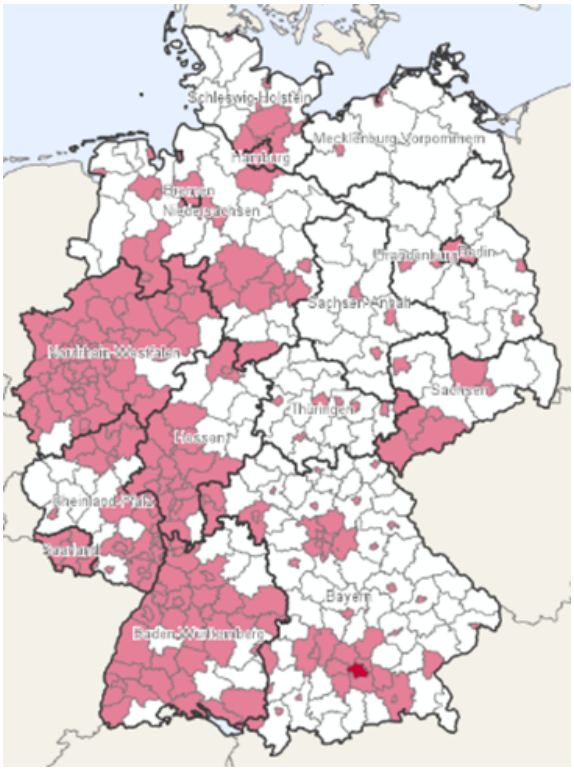


Abbildung 21.1: 160 Einwohner/km^2

2. Die Kassen erhalten zusätzliche Wettbewerbsparameter.

Denkbare Wettbewerbsparameter *könnten* sich auf die Prämienkalkulation und auf den Leistungskatalog beziehen. Beides wäre unklug, da dies vor allem zur Risikoselektion genutzt werden würde. Wir haben diese Aspekte im Bachelor untersucht.

Eine Wettbewerbsebene, die in Deutschland derzeit ausdrücklich unerwünscht ist, ist die der *Eigeneinrichtungen*, d.h. den Aufbau einer kasseneigenen Versorgungsstruktur im Sinne einer HMO. An dieser Stelle lassen sich aus dem insgesamt hoch ineffizienten US-System vielleicht nützliche Lehren ziehen.

Szenario: Kasseneigene Versorgungsstrukturen

Die Möglichkeiten für eine HMO sind in Deutschland günstiger als in den USA, da Deutschland dichter besiedelt ist und so die Wege zu einem kasseneigenen Versorgungszentrum nicht zu lang werden. Hierzu eine Überschlagsrechnung.

- Wenn ein Versorgungszentrum einen *Augenarzt* (als Beispiel für einen nicht allzu exotischen Facharzt) beschäftigen soll, muss es etwa 25.000 Versicherte versorgen, um diesen Augenarzt auszulasten.

- Das Einzugsgebiet soll maximal bei etwa 1.250 km^2 liegen. Ist das Einzugsgebiet quadratisch, beträgt die maximale Fahrstrecke zum Zentrum dann 25 km.

Unter diesen Bedingungen kann eine Kasse ein Versorgungszentrum aufbauen, wenn die Versichertendichte mindestens 20 Versicherte je km^2 beträgt. Beträgt der Marktanteil der Kasse 12,5 % oder mehr, ist das in Regionen mit einer Bevölkerungsdichte von 160

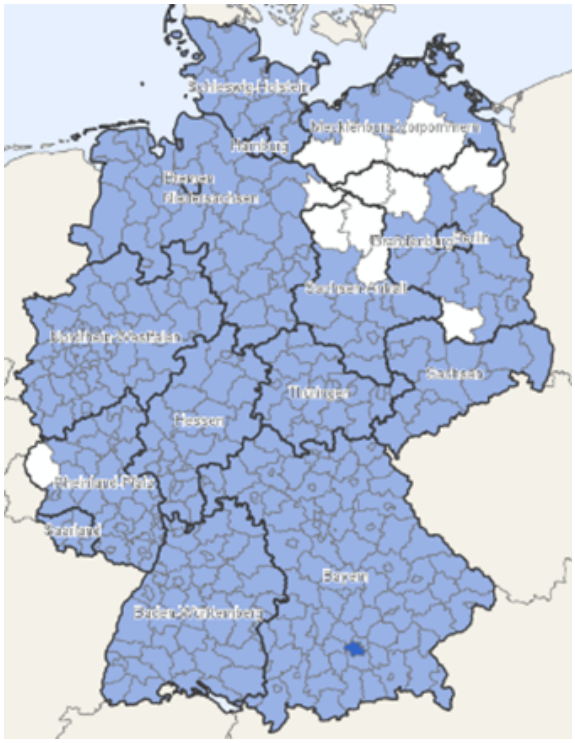


Abbildung 21.2: 60 Einwohner/km^2

Einwohnern/km^2 oder mehr der Fall. Die Karte von www-genesis.destatis.de zeigt die Landkreise, die im Jahr 2013 mindestens 160 Einwohner je km^2 aufwiesen.

Kassen, die Marktanteile in dieser Größenordnung haben sind die AOK, die Barmer GEK und die TK. Für diese Kassen ist das Konzept kasseneigener Versorgungszentren in den Ballungsräumen umsetzbar. In den strukturschwächeren Regionen nicht. Dort wohnen aber auch nicht sehr viele Menschen, so daß der *Anteil* der Bevölkerung, denen man dieses Angebot machen kann, viel größer ist als die vielen weißen Flächen vermuten lassen. Mit einem Marktanteil von etwa 1/3 kann die AOK dieses Konzept nahezu flächendeckend anbieten, wenn man die notwendige Versichertendichte auf mindestens 60 Einwohner/km^2 herunterschraubt.

Beide Szenarien zeigen, daß auch in Deutschland die Möglichkeiten dieses Konzepts begrenzt sind, bzw. nur sehr wenige Anbieter partiell oder flächendeckend in der Lage sind, ein entsprechendes Angebot zu machen.

Ehemalige Prüfungsaufgaben

TK-Versorgungszentrum

Bei der Überlegung, ob HMO-Versorgungszentren eine Option für die gesetzlichen Kassen in Deutschland wären, sind wir davon ausgegangen, daß das Zentrum 25.000 Versicherte versorgen muß, die sich auf maximal 1.250 km^2 verteilen.

Diese Rechnung arbeitet mit einem sehr dichten Netz von Versorgungszentren. Im Folgenden sollen Sie davon ausgehen, daß ein Zentrum auch dort eingerichtet wird, wo sich die Versicherten auf maximal 3.000 km^2 verteilen. Für die Rechnung ignorieren wir, daß ein Teil der Bevölkerung in der PKV versichert ist und gehen von 100% GKV-Versicherten aus.

- Angenommen, das Einzugsgebiet der Zentren wäre genau kreisförmig: Wie groß wäre die maximale Entfernung eines Versicherten zum Versorgungszentrum? Berechnen Sie den Wert in Kilometern auf eine Nachkommastelle genau.

- Die TK hat einen Marktanteil von 14%. Wie hoch müsste die regionale Bevölkerungsdichte mindestens sein, damit die TK unter den genannten Konditionen dort ein Zentrum aufbaut? Geben Sie den Wert auf ganze Einwohner gerundet an.

Ein neuer PKV-Anbieter aus Polen[1]

In der Veranstaltung haben wir uns ausführlich mit der Berechnung der Lebenserwartung befasst. Damit haben wir ein Problem wieder aufgegriffen, dass bereits im Bachelor eine Rolle gespielt hat: Die Berechnung einer Prämie für die PKV. Ein Detail, das dort keine Rolle gespielt hat, war die Frage, welche Sterbetafel die Versicherungen verwenden. Wir haben dort einfach die deutsche verwendet. Das ist nun nicht zwingend so, weil auch andere Länder diese Berechnungen für ihre Bevölkerung anstellen.

In dieser Aufgabe greife ich auf einen Fall zurück, der etwa 20 Jahre alt ist. Damit Sie diese Aufgabe im Rahmen der Prüfung lösen können, vereinfache ich das historische Beispiel, damit Sie zum einen Hintergrundinformationen aus dem Skript ziehen können und zum anderen die Idee des Versicherers deutlicher wird.

Die Dubiosa AG ist ein Versicherungsunternehmen mit Sitz in Polen. In der Simulation im Bachelor haben wir eine Prämie von knapp 220€ für einen fiktiven PKV-Durchschnitts-Volltarif berechnet. Die Dubiosa will (auf der Basis des gleichen Kapitaldeckungs-Kalkulationsprinzips) mit einer Prämie von 100€ an den Markt gehen. Diesen Tarif muss sie aber erst von der Aufsichtsbehörde (BaFin) genehmigen lassen. Nach einer ersten Durchsicht der Unterlagen durch die BaFin ist folgendes klar:

1. Die Dubiosa verwendet die polnischen Sterbetafeln.

2. Die Dubiosa verwendet die polnischen Alterskostenprofile.

[1] Prüfungsaufgabe im Sommersemester 2022. Bearbeitungszeit 30 Minuten. Vor der Prüfung habe ich darauf hingewiesen, dass es eine Frage zur PKV-Kalkulation geben würde, zu deren Vorbereitung es erforderlich sein würde, die einschlägigen Einheiten aus dem Bachelor (noch einmal) durchzuarbeiten. Im Aufgabentext habe ich darauf hingewiesen, dass der Anbieter aus *Polen* kommt, weil es in dieser Einheit Mögliche Lehren für Deutschland ein paar relevante Daten zu Polen gibt. Das ist auch der Grund, warum ich diese Aufgabe an das Ende dieser Einheit gesetzt habe und nicht an *Einheit Lebenserwartung*.

3. Der Tarif wird seit vielen Jahren in Polen angeboten. Die Prämien sind stabil. Der Tarif hat aber bisher kaum Kunden, da (grob gesagt) das polnische System mit dem britischen identisch ist und der Leistungsumfang des Tarifs nicht größer ist als der des „polnischen NHS“.

Polen hat eine eigene Währung (Zloty), aber wir nehmen einfach an, Polen hätte den Euro. Warum? Wir wollen uns an dieser Stelle nicht mit unter- oder überbewerteten Währungen bei den Alterskostenprofilen herumschlagen, sondern annehmen, dass die Kaufkraftparität zwischen Deutschland und Polen exakt gleich ist. Wenn wir mehr Zeit hätten (und Geldpolitik relevanter wäre) könnte man das noch thematisieren. Tun wir aber nicht.

- Wieso kann die Dubiosa die etablierten deutschen Anbieter bei den Prämien so massiv unterbieten? Dazu könnten wir uns viele Gründe ausdenken, die wir „an den Haaren“ herbeiziehen (geringere Verwaltungskosten, weil die Versicherung in Polen ist, ...) aber in Ihren Überlegungen soll es nur um die Aspekte gehen, die in der Problembeschreibung angesprochen werden.

- Welche der genannten Details müsste die Dubiosa ändern, damit die Seriosität der Kalkulation der der deutschen Anbieter entspricht?

TK-Versorgungszentrum (Lösung)

- Die Fläche eines Kreises ist πr^2. Wenn $3.000km^2 = 3,1415 * r^2$ ist, dann ist $r^2 = 954,96$ und $r = 30,9km$
- Auf 3000km^2 braucht die TK 25.000 Versicherte, also 8,33 Versicherte pro km^2.
- Bei einem Marktanteil von 14% wäre das bei einer Bevölkerungsdichte von $X*0,14 = 8,3$ bzw. $X = 59,29$ der Fall

Ein neuer PKV-Anbieter aus Polen (Lösung)

1. Die Grundlage für die Alterskostenprofile sind die polnischen Ausgaben. Die Grafik in Mögliche Lehren für Deutschland zeigt, dass die Pro-Kopf-Ausgaben in Polen deutlich unter denen in Deutschland liegen. Die polnischen Alterskostenprofile werden also nicht ausreichen, um die deutschen Alterskostenprofile zu finanzieren. Hierbei ist es unklar, ob diese niedrigeren Ausgaben durch einen geringeren Leistungsumfang entstehen oder durch eine niedrigere Honorierung oder durch eine Mischung von beidem.
2. Die Lebenserwartung in Polen (gleiche Grafik) liegt etwa vier Jahre unterhalb der deutschen. Der Kapitalstock wird nicht ausreichen, weil die deutschen Versicherten später sterben.
3. Die Dubiosa muss deutsche Alterskostenprofile verwenden, die das deutsche Leistungsniveau und die deutsche Honorierungshöhe berücksichtigen
4. Die Dubiosa muss zusätzlich deutsche Sterbetafeln verwenden, weil sonst auch bei deutschen Alterskostenprofilen der Kapitalstock nicht ausreichen wird, die Kosten zu decken.

Eine Strategie, die ich in der Bearbeitung häufiger beobachtet habe, war, dass die Prüflinge sich an Details aus dem Bachelor festgebissen hatten, wie z.B. dem Mindestrechnungszins, der ja vielleicht in Polen anders ist als in Deutschland. Das ist bedenkenswert, war aber durch die Aufgabenstellung als irrelevant abgetan. Dort war nur von Alterskostenprofilen und Sterbetafeln die Rede. Um die musste es also gehen.