



santésuisse

santésuisse Wirtschaftlichkeitsprüfung: Dokumentation der Umsetzung des Regressionsmodells

14.11.2023

Autoren:

Andreas Kohler, santésuisse

Carla Leber, santésuisse

Manuel Langhart, santésuisse

Philip Moline, Newindex

Juerg B. Reust, Mandatnehmer FMH

Sabrina Steffen, FMH

Inhalt

1.	Einleitung.....	3
2.	Modellspezifikation	4
•	2.1. Erste Stufe	4
•	2.2. Zweite Stufe	7
•	2.3. Indexberechnung	8
•	2.4. Berechnung Unsicherheitsfaktor	8
3.	Datengrundlage	9
•	3.1. Zeitperiode	9
•	3.2. Abhängige Variablen (Kostenarten)	9
•	3.3. Unabhängige Variablen	10
4.	Datenbereinigung und Transformationsschritte.....	12
•	4.1. Kostenzuweisung	12
•	4.2. Medikamentendaten für PCGs.....	12
•	4.3. Auswahl der Beobachtungen	13
5.	Literatur	14

1. Einleitung

santésuisse, curafutura und FMH haben im Vertrag vom 20.3.2018 die Screening-Methode als ersten Teil der Kontrolle der Wirtschaftlichkeit der ärztlichen Leistungserbringer festgelegt. Diese Screening-Methode dient der Identifikation der statistisch auffälligen Ärzte. Weist ein Arzt auffällige Kosten auf, so heisst dies nicht per se, dass der Arzt unwirtschaftlich arbeitet. Um näher abzuklären, ob ein im Sinne der angewendeten Methode auffälliger Arzt unwirtschaftlich arbeitet oder nicht, erfolgt eine Einzelfallprüfung durch santésuisse. Die Einzelfallprüfung ist eine individuelle Beurteilung des statistisch auffälligen Arztes, in deren Rahmen der Arzt Gelegenheit hat sein Kostenbild zu begründen. Die Einzelfallprüfung ist nicht zu verwechseln mit der sogenannten analytischen Methode. Es werden in der Einzelfallprüfung durch santésuisse keine Patientendossiers eingesehen.

Die vorliegende Dokumentation beschreibt die Umsetzung der Screening-Methode.

Sie beschreibt insbesondere die verwendete Modellspezifikation, die Datengrundlage die Datenbereinigung sowie die Transformationsschritte, mit welchen die Werte des Regressionsindexes ermittelt werden.

Die Dokumentation basiert auf dem Methodenbericht (B,S,S. 2016) und dem Schlussbericht (Polynomics 2017).

2. Modellspezifikation

Das Regressionsmodell besteht aus zwei Stufen. Das Ziel des Regressionsmodells ist es, denjenigen Teil der Kosten zu identifizieren, welcher nicht durch die Alters- und Geschlechterstruktur sowie die Morbidität der Patienten erklärt wird. Dieser Teil der Kosten ist ein Mass für den individuellen Praxisstil des Arztes. Er ist die Basis für das statistische Screening potentiell unwirtschaftlich behandelnder Leistungserbringer.

2.1. Erste Stufe

Auf der ersten Stufe des Arztes i und der Alters- und Geschlechtergruppe j ist ein Fixed-Effects-Modell spezifiziert. Das Fixed-Effects-Modell wird für jede Facharztgruppe gemäss Definition des SASIS AG Zahlstellenregisters (ZSR) und Leistungsart separat mit der Methode der kleinsten Quadrate (ordinary least squares – OLS) geschätzt.

Das Regressionsmodell ist auf der ersten Stufe wie folgt spezifiziert:

$$\begin{aligned}\ln(y_{ij}) = & \alpha_i \\ & + \beta_1 AGG_{i1} + \dots + \beta_J AGG_{iJ} \\ & + \gamma_1 HFR_{ij} + \gamma_2 SPV_{ij} \\ & + \delta_{11} PCG_{11,ij} + \dots + \delta_{1G} PCG_{1G,ij} \\ & + \dots \\ & + \delta_{P1} PCG_{P1,ij} + \dots + \delta_{PG} PCG_{PG,ij} \\ & + \varepsilon_{ij}\end{aligned}$$

<i>Variablen</i>	<i>Beschreibung</i>
y_{ij}	Bruttokosten pro Erkrankter
AGG	Alters- und Geschlechtergruppe
HFR	Hohe Franchise
SPV	Spital- oder Pflegeheimaufenthalt im Vorjahr
PCG	Pharmaceutical Cost Group

Das Regressionsmodell auf der ersten Stufe wird in Stata mit dem Befehl **areg** implementiert (StataCorp 2019). Die Beobachtungen werden mit der Anzahl Erkrankten des Arztes in jeder Alters- und Geschlechtergruppe gewichtet.

2.1.1. Abhängige Variable

$\ln(y_{ij})$ bezeichnet die logarithmierten Bruttokosten pro Erkrankter y_{ij} von Arzt i in der Alters- und Geschlechtergruppe j . Die Bruttokosten sind die abhängige Variable und werden nach Leistungsarten (Kostenarten) Arzt (TARMED), Medikamente, Labor, MiGeL, und Physiotherapie unterschieden.

2.1.2. Fixed-Effect

Der Fixed-Effect wird im Modell mit α_i bezeichnet. Der Fixed-Effect ist ein Mass für den individuellen Praxisstil eines Arztes, weil für die Morbidität seiner Patienten kontrolliert wird. Er gibt Auskunft über die durchschnittlichen Kosten pro Erkrankter bei, mit seiner Facharztgruppe, vergleichbarer Patientenstruktur (B,S,S. 2016, S. 5). Der Fixed-Effect bildet nicht alle Praxisbesonderheiten ab.

2.1.3. Unabhängige Variablen

Die unabhängigen Variablen bilden die Alters- und Geschlechterstruktur sowie die Morbidität der Patienten ab.

Alters- und Geschlechterstruktur

Die Dummy-Variablen $AGG_{i1}, \dots, AGG_{ij}$ bilden die Struktur der Alters- und Geschlechtergruppen ab.¹ Die Variable AGG_{ij} nimmt den Wert 1 an, wenn die Bruttokosten von Arzt i in der Alters- und Geschlechtergruppe $j = 1, \dots, J$ angefallen sind und andernfalls den Wert Null. So wird berücksichtigt, dass Patienten in jeder Alters- und Geschlechtergruppe im Durchschnitt unterschiedlich hohe Bruttokosten verursachen, unabhängig von ihrer Morbidität und vom Praxisstil des behandelnden Arztes.

Morbidität

Die Morbidität der Patienten wird über die Indikatoren Franchise HFR , Spitalaufenthalt im Vorjahr SPV sowie die pharmazeutischen Kostengruppen PCG abgebildet. Dies sind Indikatoren, welche den Krankheitszustand der Patienten eines Arztes beschreiben und somit unabhängig von dessen Behandlungsstil sind.

Die Variable Franchise HFR_{ij} bildet für jeden Arzt i in jeder Alters- und Geschlechtergruppe j den Anteil der Grundleistungen (Konsultationen und Hausbesuche) bei Patienten mit einer hohen Franchise ab (über CHF 500 bei Erwachsenen und über CHF 100 bei Kindern). Dies ist eine kontinuierliche Variable, welche Werte zwischen 0 und 1 annimmt.

Die Variable SPV_{ij} bildet für jeden Arzt i in jeder Alters- und Geschlechtergruppe j den Anteil der Patienten mit einem Spital- oder Pflegeheimaufenthalt im Vorjahr (d. h. mehr als 3 aufeinanderfolgende Nächte) ab. Dies ist, wie die Variable Franchise, eine kontinuierliche Variable, welche Werte zwischen 0 und 1 annimmt.

Die Variablen PCG_{pg} bezeichnen die pharmazeutischen Kostengruppen (PCG). Der Index $p = 1, \dots, P$ indexiert die P pharmazeutischen Kostengruppen gemäss der aktuellen PCG-Liste des BAGs. Der Index $g = 1, \dots, G$ indexiert die G Gruppen, welche auf Basis der empirischen Verteilung der Wirkstoffmengen (gemessen anhand der Defined Daily Doses – DDDs), für jede PCG gebildet werden. Die PCG p wird in einer Facharztgruppe (FAG) berücksichtigt, wenn bei mehr als 30 Ärzten je mindestens 1.8 DDD pro Erkrankter abgerechnet wurden. Die Herleitung von 1.8 DDD entspricht einer Halbjahresdosis (180 definierte DDD) pro 100 Patienten (Polynomics 2017, S.33). Diese Definition ist notwendig da nur aggregierte Daten und keine Individualdaten zur Verfügung stehen. Wird eine PCG berücksichtigt, werden wie folgt G Gruppen gebildet:

- Zehn Gruppen nach Dezilen, wenn $\geq 10\,000$ Beobachtungen mit einem Wert grösser Null. Plus eine Gruppe, wenn DDD-Menge gleich Null.
- Vier Gruppen nach Quartilen, wenn $\geq 1\,000$ bis $< 10\,000$ Beobachtungen mit einem Wert grösser Null. Plus eine Gruppe, wenn DDD-Menge gleich Null.
- Zwei Gruppen nach Median, wenn > 0 bis $< 1\,000$ Beobachtungen mit einem Wert grösser Null. Plus eine Gruppe, wenn DDD-Menge gleich Null.

Jede Gruppe wird als Dummy-Variable in dem Modell abgebildet.

Die Modellierung der PCGs hat folgende zwei Gründe. Erstens, die PCG haben eine endogene Komponente, welche vom Verschreibungsverhalten oder eben vom individuellen Praxisstil eines Arztes abhängt. Würden die PCG als Anzahl DDD pro Erkrankte im Modell berücksichtigt, ist diese, vom Arzt beeinflussbare, endogene Komponente stärker. Zweitens, würde unterstellt, dass der Effekt der PCG auf die Kosten pro Erkrankte linear ist. Die gewählte Umsetzung der PCG

¹ Als Dummy-Variablen werden Variablen bezeichnet, welche nur die Werte Null oder Eins annehmen.

berücksichtigt, wie viele DDD pro Erkrankter ein Arzt im Verhältnis abgibt. Das ist von einem einzelnen Arzt wenig bis kaum beeinflussbar. Zudem erlaubt die Spezifikation einen nicht-linearen Effekt der PCG auf die Kosten pro Erkrankter.

Beispiel: Die PCG ADH (ADHS) wird zum Beispiel im Geschäftsjahr 2018 in der FAG Ophthalmologie im Modell nicht berücksichtigt, aber in der FAG Allgemeine Innere Medizin (AIM). Es haben also in der FAG AIM mehr als 30 Ärzte je mindestens 1.8 DDD pro Erkrankter abgerechnet. Da in der FAG AIM zwischen 1 000 und 10 000 Beobachtungen einen Wert grösser null aufweisen, werden vier Gruppen nach den Quartilen gebildet und eine Gruppe für DDD-Mengen gleich Null. Im Modell werden also in der FAG AIM die vier Dummy-Variablen $PCG_{11,ij}$, $PCG_{12,ij}$, $PCG_{13,ij}$, $PCG_{14,ij}$ eingeschlossen. Die Dummy-Variable $PCG_{11,ij}$ nimmt den Wert 1 an wenn Arzt i in der Alters- und Geschlechtergruppe j im 1. Quartil der Verteilung der Anzahl DDD des PCG ADH Medikamente abgerechnet hat und andernfalls den Wert Null. Das 1. Quartil kann interpretiert werden als „geringe Anzahl DDD“, das 2. und 3. Quartil als „mittlere Anzahl DDD“ und das 4. Quartil als „hohe Anzahl DDD“. Rechnet ein Arzt in einer Alters- und Geschlechtergruppe also eine vergleichsweise geringe Anzahl DDD ab, so nimmt die Dummy-Variable für das 1. Quartil den Wert 1 an und die Dummy-Variablen für alle anderen Quartile den Wert Null.

Die Modellierung der PCG, d. h. die berücksichtigten PCGs in einer FAG sowie die Anzahl gebildete Gruppen, passt sich den Daten an und kann deshalb in jedem Geschäftsjahr anders sein.

2.1.4. Fehlerterm

Die Variable ε_{ij} bezeichnet den idiosynkratischen Fehlerterm (Residuum) des Regressionsmodells auf der ersten Stufe.

2.2. Zweite Stufe

Der Fixed-Effect α_i aus der ersten Stufe bildet den individuellen Praxisstil ab, welcher auf der zweiten Stufe Arzt i um die Einflüsse des Standortkantons und der FAG bereinigt wird. Damit werden kantonspezifische Faktoren (z. B. unterschiedliche Taxpunktwerte) sowie facharztgruppenspezifische Faktoren (z. B. Komplexität der Behandlungen) berücksichtigt, welche kostenrelevant sind.

Das Regressionsmodell auf der zweiten Stufe ist wie folgt spezifiziert:

$$\begin{aligned}\hat{\alpha}_i = & \eta \\ & + \theta_1 KT_{1i} + \dots + \theta_K KT_{Ki} \\ & + \phi_1 FAG_{1i} + \dots + \phi_F FAG_{Fi} \\ & + u_i\end{aligned}$$

Das Regressionsmodell auf der zweiten Stufe wird in Stata mit dem Befehl **reg** implementiert (StataCorp 2019). Die Beobachtungen werden mit der totalen Anzahl Erkrankten des Arztes gewichtet.

2.2.1. Abhängige Variable

Der in der ersten Stufe geschätzte Fixed-Effect wird mit $\hat{\alpha}_i$ bezeichnet und ist die abhängige Variable auf der zweiten Stufe.

2.2.2. Unabhängige Variablen

Die unabhängigen Variablen bilden die Einflüsse des Standortkantons und der FAG des Arztes auf seinen individuellen Praxisstil ab. Dies sind arzt spezifische und nicht patientenspezifische Variablen. Sie variieren also nicht über die Alters- und Patientengruppen innerhalb einer Arztpraxis und können deshalb nicht auf der ersten Stufe in das Modell eingeschlossen werden.

Konstante

Der Parameter η bezeichnet die Konstante im Regressionsmodell.

Standortkanton

Die Dummy-Variablen $KT_{1i}, \dots, KT_{Ki}$ bilden die Einflüsse des Standortkantons des Arztes ab. Die Variable KT_{ki} nimmt den Wert 1 an wenn die Praxis von Arzt i in Kanton $k = 1, \dots, K$ ist und andernfalls den Wert Null. Damit werden kantonsspezifische Faktoren wie kantonale Unterschiede in den Taxpunktwerten oder Miet- und Personalkosten berücksichtigt.

Facharztgruppe

Die Dummy-Variablen $FAG_{1i}, \dots, FAG_{Fi}$ bilden die Einflüsse der Facharztgruppe des Arztes ab. Die Variable FAG_{fi} nimmt den Wert 1 an wenn Arzt i in der FAG $f = 1, \dots, F$ eingeteilt ist und andernfalls den Wert Null. Damit werden facharztspezifische Faktoren wie der Komplexität der Behandlung Rechnung getragen.

2.2.3. Fehlerterm

Die Variable u_i bezeichnet den idiosynkratischen Fehlerterm (Residuum) des Regressionsmodells auf der zweiten Stufe.

2.3. Indexberechnung

Aus den geschätzten Residuen der zweiten Stufe $\hat{u}_i$ wird der Indexwert Index_i von Arzt i in FAG f wie folgt berechnet:

$$\text{Index}_i = s_f \times \exp(\hat{u}_i)$$

wobei $s_f = \left(\frac{1}{\frac{1}{N} \sum_{i=1}^N \exp(\hat{u}_i)} \right) \times 100$ den Skalierungsfaktor bezeichnet, der so gewählt ist, dass der Mittelwert aller Indizes innerhalb von FAG f den Wert 100 hat. N bezeichnet die Anzahl Ärzte pro FAG f . In dem, im Skalierungsfaktor s_f enthaltenen Durchschnitt $\frac{1}{N} \sum_{i=1}^N \exp(\hat{u}_i)$, wird das Residuum $\hat{u}_i$ von Arzt i mit der totalen Anzahl Erkrankten des Arztes i gewichtet.

2.4. Berechnung Unsicherheitsfaktor

Der Unsicherheitsindikator $UI(\hat{a}_i)$ für den Fixed-Effects auf der ersten Stufen wird wie folgt berechnet:

$$UI(\hat{a}_{1st,i}) = \sqrt{\frac{1}{J_i J_i - 1} \sum_{j=1}^{J_i} \frac{P_{ij}}{P_i} \varepsilon_{ij}^2}$$

wobei J_i die Anzahl Beobachtungen pro Arzt i (über alle Alters- und Geschlechtergruppen J), P_{ij} die Anzahl Erkrankten von Arzt i in Alters- und Geschlechtergruppe j , P_i die totale Anzahl Erkrankte von Arzt i und ε_{ij} den Fehlerterm auf der ersten Stufe bezeichnet. $\hat{a}_{1st,i}$ bezeichnet den berechneten Fixed-Effekt aus der 1. Stufe.

Der Unsicherheitsindikator wird dazu verwendet ein 95%-Konfidenzintervall (lower und upper bound) für den Fixed-Effects aus der ersten Stufe zu berechnen:

$$\begin{aligned} \hat{a}_{1st,i}^{low} &= \hat{a}_{1st,i} - 1.96 \times UI(\hat{a}_{1st,i}) \\ \hat{a}_{1st,i}^{up} &= \hat{a}_{1st,i} + 1.96 \times UI(\hat{a}_{1st,i}) \end{aligned}$$

Mit Hilfe des lower und upper bounds des Fixed-Effects aus der ersten Stufe wird der bereinigte Fixed-Effects aus der zweiten Stufe korrigiert:

$$\begin{aligned} \hat{u}_i^{low} &= \hat{a}_{1st,i}^{low} - \hat{a}_{2nd,i} \\ \hat{u}_i^{up} &= \hat{a}_{1st,i}^{up} - \hat{a}_{2nd,i} \end{aligned}$$

Schlussendlich wird der lower und upper bound des Regressionsindexes wie folgt berechnet:

$$\begin{aligned} \text{Index}_i^{low} &= s_f \times \exp(\hat{u}_i^{low}) \\ \text{Index}_i^{up} &= s_f \times \exp(\hat{u}_i^{up}) \end{aligned}$$

wobei s_f den Skalierungsfaktor von oben bezeichnet.

3. Datengrundlage

Die Datengrundlage umfasst die folgenden Daten.

Eine Beobachtungseinheit ist auf der Ebene ZSR-Nr., Alters- und Geschlechtergruppe und Geschäftsjahr eineindeutig identifiziert.

3.1. Zeitperiode

Die Zeitperiode bezieht sich auf die Geschäftsperiode (Geschäftsjahr) der abgerechneten und verarbeiteten Leistungen.

3.2. Abhängige Variablen (Kostenarten)

Tabelle 1 führt die abhängigen Variablen des Regressionsmodells mit entsprechender Datenquelle auf.

Tabelle 1: Abhängige Variablen

Variable	Definition	Einheit	Typ	Datenquelle
Arztkosten	Arztbehandlung TARMED	Taxpunkte	Numerisch	DP SASIS AG
Medikamentenkosten	Medikamente bei Arzt (direkt) sowie bei Apotheken (veranlasst)	CHF	Numerisch	DP SASIS AG
Laborkosten	Labor bei Arzt (direkt) sowie bei Laboratorien (veranlasst)	CHF	Numerisch	DP SASIS AG
MiGeL-Kosten	MiGeL bei Arzt (direkt) sowie bei Apotheken und Abgabestellen MiGeL (veranlasst)	CHF	Numerisch	DP SASIS AG
Physiotherapiekosten	Physiotherapie bei Arzt (direkt) sowie bei Physiotherapeuten (veranlasst)	CHF	Numerisch	DP SASIS AG
Übrige Kosten	Alle Kosten, die keiner obigen Leistungsart zugeordnet werden	CHF	Numerisch	DP SASIS AG
Totale Kosten	Summe aller obigen Leistungsarten	CHF	Numerisch	DP SASIS AG
Totale direkte Kosten	Summe der Arzt-, Labor- (direkt), Medikamente- (direkt), MiGeL- und übrigen Kosten	CHF	Numerisch	DP SASIS AG

Die abhängigen Variablen werden als Kosten / Taxpunkte pro Erkrankter in die erste Stufe des Regressionsmodells eingeschlossen.² Die Kosten in jeder Alters- und Geschlechtergruppe in der obenstehenden Tabelle werden also durch die Anzahl Erkrankte in der jeweiligen Alters- und Geschlechtergruppe dividiert. Der Erkrankte ist im DP SASIS AG wie folgt definiert: Als statistischer Behandlungsfall gilt der einmal oder mehrmals erkrankte Versicherte pro ZSR-Nr., der innerhalb des gelieferten Geschäftsjahres Zahlungen verursacht. Hinweis: für die Kostenart Übrige Kosten wird kein Regressionsindex berechnet. Allfällige Anpassungen finden sich in der aktuellen Version des Anhangs zu dieser Dokumentation.

² Damit die Arztkosten gemessen in Taxpunkten zusammen mit den Kosten der anderen Leistungsarten zu den Totalen Kosten aufsummiert werden können, müssen diese in CHF umgerechnet werden. Dazu wird ein CH-TPW ermittelt, indem die gesamten Arztkosten in CHF aller Ärzte in der Schweiz dividiert werden durch die gesamten Taxpunkte aller Ärzte in der Schweiz. Danach werden die Taxpunkte eines Arztes mit dem CH-TPW multipliziert und die daraus resultierenden Kosten in CHF zu seinen Kosten der anderen Leistungsarten addiert.

Die Kosten pro Erkrankter von Arzt i in Alters- und Geschlechtergruppe j werden im Modell mit y_{ij} bezeichnet (siehe Abschnitt 2.1.1, Seite 4 ff.). Die abhängige Variable wird auf der ersten Stufe des Modells logarithmiert (sogenanntes log-lineares Modell). Da der Logarithmus von Null nicht definiert ist, wird bei allen Kosten der Wert 1 dazu addiert, so dass Beobachtungen mit Nullwerten nicht verloren gehen.

Die abhängige Variable auf der zweiten Stufe des Regressionsmodells entspricht dem Fixed-Effect, der auf der ersten Stufe geschätzt wird (siehe Abschnitt 2.1.2, Seite 4). Die abhängige Variable wird auf der zweiten Stufe nicht transformiert.

3.3. Unabhängige Variablen

Tabelle 2 führt die unabhängigen Variablen des Regressionsmodells mit entsprechender Datenquelle auf.

Tabelle 2: Unabhängige Variablen

Variable	Definition	Typ	Datenquelle
Alter	Lebensalter der Versicherten basierend auf den 25 BAG-Altersgruppen: 0-5 J., 6-10 J., 11-15 J., 16-18 J., 19-20 J., 21-25 J., ..., 96+ J.	Integer	DP SASIS AG
Geschlecht	Männlich, Weiblich	Binär	DP SASIS AG
Anzahl Grundleistungen bei Erkrankten mit hoher Franchise	Anzahl Grundleistungen (Konsultationen und Hausbesuche) bei Erkrankten mit einer Jahresfranchise über CHF 500 (Erwachsene) oder CHF 100 (Kinder)	Integer	DP SASIS AG
Anzahl Erkrankte mit Spitalaufenthalt im Vorjahr	Anzahl Erkrankte mit 3 oder mehr aufeinanderfolgenden Nächten in einem Spital- oder Pflegeheim im Vorjahr	Integer	DP SASIS AG
PCGs	Anzahl Positionen pro Pharmacode DDD pro Pharmacode (gemäss aktueller PCG-Liste)	Integer	TP SASIS AG
Kanton	Kanton	Kategorial	BAG ZSR SASIS AG
FAG	Facharztgruppe gemäss Einteilung im Zahnstellenregister (ZSR) der SASIS AG	Kategorial	ZSR SASIS AG

Aus den Variablen Alter und Geschlecht werden die Dummy-Variablen $AGG_{i1}, \dots, AGG_{ij}$ gebildet, welche in das Regressionsmodell auf der ersten Stufe einfließen (siehe Abschnitt 2.1.3, Seite 4 ff.). Die Altersgruppen über 96 Jahre werden zu einer Altersgruppe zusammengefasst. Es sind dies alle Kombinationen von Alter und Geschlecht.

Aus der Anzahl Grundleistungen bei Erkrankten mit einer Jahresfranchise über CHF 500 (Erwachsene) oder CHF 100 (Kinder) wird die Variable HFR_{ij} gebildet, indem die Anzahl in jeder Alters- und Geschlechtergruppe durch die Gesamtanzahl des Arztes dividiert wird (siehe Abschnitt 2.1.3, Seite 4 ff.).³ Diese bildet den Anteil der Grundleistungen ab, welche bei Erkrankten mit einer hohen Jahresfranchise anfallen. Die Variable ist kontinuierlich und nimmt Werte zwischen 0 und 1 an.

³ Die Kategorie tiefe Franchise umfasst die Franchisestufen CHF 300 und CHF 500 bei Erwachsenen und CHF 100 bei Kindern. Die Kategorie hohe Franchise umfasst die Franchisestufen CHF 1000, CHF 1500, CHF 2000 und CHF 2500 bei Erwachsenen und CHF 200, CHF 300, CHF 400, CHF 500 und CHF 600 bei Kindern.

Aus der Anzahl Erkrankte mit Spitalaufenthalt im Vorjahr wird die Variable SPV_{ij} gebildet, indem die Anzahl in jeder Alters- und Geschlechtergruppe durch die Gesamtanzahl des Arztes dividiert wird (siehe Abschnitt 2.1.3, Seite 4 ff.). Die Variable bildet den Anteil Erkrankter mit Spitalaufenthalt im Vorjahr ab. Sie ist kontinuierlich und nimmt Werte zwischen 0 und 1 an.

Aus der abgerechneten Anzahl Positionen (Packungen) pro Pharmacode aus dem TP SASIS AG und den hinterlegten DDD pro Pharmacode auf der jeweils aktuellen BAG PCG-Liste wird die Anzahl DDDs pro PCG in jeder Alters- und Geschlechtergruppe berechnet. Auf Basis der Anzahl DDD pro PCG und Alters- und Geschlechtergruppe werden die Dummy-Variablen PCG_{pg} gemäss Polynomics (2017) gebildet (siehe Abschnitt 2.1.3, Seite 4 ff.).

Aus den Variablen Kanton und FAG werden die Dummy-Variablen $KT_{1i}, \dots, KT_{Ki}$ und $FAG_{1i}, \dots, FAG_{Fi}$ gebildet, welche auf der zweiten Stufe in das Regressionsmodell eingeschlossen werden.

4. Datenbereinigung und Transformationsschritte

Die Datenaufbereitung umfasst die folgenden Bereinigungs- und Transformationsschritte.

4.1. Kostenzuweisung

Kosten aller Leistungsarten, die bei einem ausführenden Leistungserbringer der Partnerobergruppe (PAOG) Arzt anfallen, werden dem ausführenden Leistungserbringer zugewiesen. Kosten aller Leistungsarten, bei denen der ausführende Leistungserbringer der Partnerobergruppe Apotheken, Laboratorien, Physiotherapie oder Abgabestellen MiGeL angehört, werden dem veranlassenden Leistungserbringer zugeordnet. Tabelle 3 illustriert die Kostenzuweisung mit Beispielen.

Tabelle 3: Beispiele Kostenzuweisung

PAOG Veranlassender Leistungserbringer		PAOG Ausführender Leistungserbringer	Leistungsart (LAR)
Arzt	→	Arzt	Arztkosten TARMED
Spital	→	Arzt	Arztkosten TARMED
...			
Arzt	←	Apotheke	Medikamente
Arzt	←	Abgabestelle MiGeL	MiGeL
...			

Anmerkungen: PAOG = Partnerobergruppe

4.2. Medikamentendaten für PCGs

Die Kosten der abgerechneten Medikamente auf Ebene ZSR-Nr., Alters- und Geschlechtergruppe und Pharmacode werden dem ausführenden Leistungserbringer (Arzt) zugewiesen falls der veranlassende Leistungserbringer der Partnerobergruppe Arzt angehört. Mit anderen Worten werden die Kosten der Medikamente, welche von einer Apotheke abgegeben werden, dem veranlassenden Arzt zugewiesen. Die Kosten der Medikamente, welche von einem Arzt abgegeben werden, werden diesem Arzt zugewiesen und nicht dem Arzt, der die Medikamente veranlasst hat.

Die abgerechnete Anzahl Positionen (Packungen) pro Pharmacode im TP SASIS AG wird wie folgt plausibilisiert. Der Fakturbetrag wird durch den Preis des Medikaments auf der BAG Spezialitätenliste (Stand Dezember) dividiert. Weicht die so berechnete Anzahl Positionen von der Anzahl Positionen im TP SASIS AG um mehr als 25 Prozent ab, wird die berechnete Anzahl Positionen verwendet (dies betraf zum Beispiel im Geschäftsjahr 2018 nur ca. 3 Prozent aller Beobachtungen).

4.3. Auswahl der Beobachtungen

Beobachtungen, welche die folgenden Kriterien erfüllen, werden gelöscht und fliessen somit nicht in das Regressionsmodell ein:

- Beobachtungen von Ärzten mit unbekanntem Kanton oder Praxisstandort Ausland
- Beobachtungen von Ärzten mit totalen Kosten kleiner als CHF 100 000 und insgesamt weniger als 50 Erkrankten (totale Kosten $\leq$ 100000 & Anzahl Erkrankte $\leq$ 50)
- Beobachtungen in Altersgruppen, bei denen das Geschlecht oder das Alter unbekannt ist
- Beobachtungen, bei denen die Anzahl Erkrankten fehlt
- Beobachtungen mit totalen Kosten und Anzahl Erkrankten kleiner oder gleich Null (z. B. Stornos)
- Beobachtungen mit Konsultationen oder Hausbesuchen kleiner Null
- Beobachtungen mit totalen Kosten pro Erkrankte kleiner Eins oder fehlenden Werten (z. B. Rundungsdifferenzen)
- Unbekannte bei den Variablen «Anteil Konsultationen mit hoher Franchise» und «Anteil Erkrankte mit Spitalaufenthalt im Vorjahr» werden bei der Anteilsberechnung nicht berücksichtigt. Enthält eine Beobachtungen nur Unbekannte wird der Wert der entsprechenden Variable auf 0 gesetzt (tritt nur bei vereinzelter Beobachtungen mit sehr wenigen Erkrankten auf)
- Vor der Berechnung der DDD werden Beobachtungen mit negativer Packungsanzahl oder negativen Kosten (Stornos) im SASIS Tarifpool gelöscht
- Da der SASIS Tarifpool im Gegensatz zum Datenpool nur die elektronisch erfassten Rechnungen umfasst, wird ein Abdeckungsgrad (Tarifpool an Datenpool) pro ZSR berechnet. Vor der Bildung der PCG-Kategorien wird die DDD-Menge mit dem Abdeckungsgrad hochgerechnet
- Beobachtungen von Facharztgruppen mit weniger als 50 ZSR werden gelöscht. Die Anzahl Beobachtungen ist zu gering um statistisch belastbare Ergebnisse zu erhalten

5. Literatur

B,S,S. 2016. „Methodische Weiterentwicklung der Wirtschaftlichkeitsprüfung. Methodenbericht.“ B,S,S. *Volkswirtschaftliche Beratung AG*. Methodenbericht im Auftrag von tarifsuisse ag.

Polynomics. 2017. „Weiterentwicklung der statistischen Methode zur Prüfung der Wirtschaftlichkeit. Schlussbericht.“ *Polynomics AG*. Studie im Auftrag von FMH, santésuisse und curafutura.

StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LP.