

Beiträge und Analysen

Gesundheitswesen

aktuell 2023

herausgegeben von Uwe Repschläger,
Claudia Schulte und Nicole Osterkamp

Christian Hans

Ein Subsamplingverfahren zur Hochrechnung einer Stichprobe auf Grundlage
des Morbiditätsorientierten Risikostrukturausgleichs
Seite 66–79

doi: 10.30433/GWA2023-66

Christian Hans

Ein Subsamplingverfahren zur Hochrechnung einer Stichprobe auf Grundlage des Morbiditätsorientierten Risikostrukturausgleichs

Der Morbiditätsorientierte Risikostrukturausgleich (MRSA) wird regelmäßig reformiert, was die Zuweisungen für jede Krankenkasse verändert. Das hier vorgestellte Subsamplingverfahren kann den wirtschaftlichen Effekt dieser Reformen auf Basis einer Stichprobe der GKV-Grundgesamtheit vorhersagen. Das Verfahren verbessert die Hochrechnung der Stichprobe, indem neben den bekannten Verteilungen der Grundgesamtheit zusätzliche Informationen verwendet werden: Das Subsampling verwendet die Koeffizientenvektoren aus den offiziellen MRSA-Regressionen des Bundesamtes für Soziale Sicherung (BAS). Nach Subsampling wird die MRSA-Regression durch Ensemblebildung angenähert, was repräsentativere Ergebnisse für die Zuweisungen vor und nach Einführung von Reformen liefert. Dadurch können die gesetzlichen Krankenkassen Reformvorschläge besser bewerten, was ihrerseits fundiertere Reformvorschläge für den MRSA ermöglicht und ihre finanzielle Planungssicherheit erhöht.

Einleitung

Der MRSA wird in der Gesetzlichen Krankenversicherung (GKV) genutzt, um Schwächen des in einer langen Historie gewachsenen Versicherungsmarktes auszugleichen. Dabei werden Finanzmittel zwischen allen gesetzlichen Krankenkassen umverteilt, um den Wettbewerb gerechter zu gestalten. Ursache des Problems sind die unterschiedlichen Versichertenstrukturen der Krankenkassen und die Begrenzung der Beitragssätze: Die gesetzlichen Krankenkassen können sich weder ihre Versicherten aussuchen noch deren individuelle Zusatzbeitragssätze unbegrenzt an die zu erwartenden Leistungsausgaben anpassen. Durch den im Jahr 2009 eingeführten MRSA soll dieses Problem auf folgende Weise gelöst werden: Die Einnahmen der Krankenkassen werden im Gesundheitsfonds gesammelt und über den MRSA neu verteilt, sodass jede Krankenkasse Zuweisungen in Abhängigkeit von der Risikostruktur ihrer Versicherten erhält. Das Verfahren basiert auf einer gewichteten Kleinstquadrate-Schätzung (englisch: WLS), die für alle rund 73 Millionen Versicherten der GKV Alter, Geschlecht, Wohnort, zuschlagsfähige Krankheiten und einige weitere Merkmale enthält. Zielvektor der Regression sind

die berücksichtigungsfähigen Leistungsausgaben, die für jede versicherte Person angefallen sind. Als Gewicht dienen die Versichertentage im Kalenderjahr. Das Ergebnis der Regression sind somit die bedingten Mittelwerte der berücksichtigungsfähigen Leistungsausgaben pro Tag für jede enthaltene Variable (Sozialgesetzbuch Fünftes Buch, SGB V, §§ 266 ff sowie Risikostruktur-Ausgleichsverordnung).

Der Gesetzgeber hat den MRSA als lernendes System implementiert (IGES et al. 2005). Es wird jedes Jahr vom BAS durch kleinere oder größere Veränderungen nach Anhörung des dort ansässigen Wissenschaftlichen Beirats angepasst. Daneben werden in gesetzgeberischen Verfahren in unregelmäßigen Abständen Reformen des MRSA durchgeführt. Zu den häufigsten Anpassungen gehören Änderungen in der Erfassung und Gruppierung der über 15.000 ICD-10-Kodes für Krankheiten zu etwa 400 bis 500 hierarchischen Morbiditätsgruppen (HMG), die in der WLS verwendet werden. Zu jedem dieser Anpassungsvorschläge geben die Krankenkassen oder Kassenverbände Stellungnahmen ab, um diese zu bewerten, Probleme aufzuzeigen oder eigene Vorschläge einzubringen. Dadurch wird der MRSA fortwährend verändert und verbessert.

Testfall Risikopool

Neben den jährlichen Anpassungen werden aber auch größere Reformen wie beispielsweise die Wiedereinführung eines Risikopools für besonders kostenintensive Versicherte im Jahr 2021 umgesetzt. Allerdings soll ein noch vom Wissenschaftlichen Beirat zu erstellendes Gutachten zeigen, welche Auswirkungen der Risikopool für Versicherte, die im jeweiligen Ausgleichsjahr Kosten von mehr als 100.000 Euro verursacht haben, entfaltet.

Vor diesem Hintergrund ist zurzeit eine Veränderung des Risikopools in der Debatte, die auf einen Vorschlag von Schillo et al. (2016) zurückgeht. Aktuell ist der Risikopool so implementiert, dass jeder Versicherte, der im jeweiligen Ausgleichsjahr Kosten von mehr als 100.000 Euro verursacht hat, eine Anpassung seiner Kosten erfährt: Ein Anteil von 80 Prozent der Kosten über 100.000 Euro werden über den Risikopool direkt erstattet, der Rest geht in die MRSA-Regression ein. Diese Ausgestaltung des Risikopools adressiert somit Hochkostenfälle. Der Reformvorschlag, der von Schillo et al. (2016)

unterbreitet wurde, adressiert stattdessen hohe Unterdeckungen: Versicherte, deren Kosten die Zuweisung aus dem MRSA nach der Regression um mehr als 30.000 Euro übersteigen, werden einer Hochkostengruppe zugeordnet. Weitergehende Vorschläge verwenden mehrere Hochkostengruppen, zum Beispiel für 30.000 bis 50.000, 50.000 bis 100.000 und mehr als 100.000 Euro Unterdeckung. Mit diesen zusätzlichen Variablen wird die WLS-Schätzung wiederholt. Diese Art von Reformvorschlag stellt eine grundlegende Veränderung dar: Es werden nicht nur neue Variablen eingeführt, sondern auch die verwendeten zuschlagsfähigen Kosten und das Schätzverfahren angepasst, indem eine zweite Regression eingeführt wird.

Das vorgestellte Verfahren ist in der Lage, die verschiedenen Ausgestaltungen eines solchen Reformvorschlags zu simulieren und Vorhersagen über die Veränderung der Gesamtzuweisungen in Abhängigkeit von der Versichertenstruktur zu machen. Für die Krankenkasse bedeutet dies, den Effekt einer solchen Reform auf ihre Zuweisungen besser simulieren zu können. Zur Evaluation der Methodik dient der Fall eines Reformvorschlags, der bereits vom BAS auf der Grundgesamtheit simuliert wurde. Der wahre Effekt dieser Reform ist somit bekannt und kann als Benchmark dienen.

Das Problem, das durch das Verfahren des Subsampling gelöst wird, ist die korrekte Schätzung der Reformeffekte, die für die GKV-Grundgesamtheit zu erwarten sind, wenn man nur einen Teildatensatz zur Verfügung hat. Aufgrund der Summentreue der Zuweisungen wäre eine Simulation auf den eigenen Daten nicht aussagekräftig – die Gesamtzuweisung der Krankenkasse würde sich nicht ändern. Daher ist eine Hochrechnung zwingend notwendig. Jede Krankenkasse hat mit den Daten ihrer eigenen Versicherten eine Stichprobe der GKV, die aber nicht repräsentativ für die Grundgesamtheit ist. Dies ist aufgrund der hohen Dimensionalität der Daten (laut Risikostruktur-Ausgleichsverordnung 2021 werden 623 Variablen in der WLS-Schätzung verwendet) quasi unmöglich. Bisher hat die Krankenkasse, deren Daten für dieses Projekt verwendet wurde, ein klassisches Hochrechnungsverfahren angewendet, um ihre Stichprobe der Grundgesamtheit anzunähern. Diese Versuche waren nur zu einem gewissen Grad erfolgreich, was der Verfügbarkeit von Informationen über die Grundgesamtheit geschuldet ist: Lediglich die Randverteilungen der einzelnen Variablen wurden vom BAS in Form der Grundlagenbescheide (GLB) I, II und III zur Verfügung gestellt. Damit sind zwar

Informationen über die Verteilung aller Variablen individuell bekannt – das bedeutet die Alters- und Geschlechtsstruktur, die Verteilung der Morbidität auf die HMGs etc. –, über die Interaktion der Variablen ist jedoch praktisch nichts bekannt. Informationen über die Komorbidität und die Verteilung der Krankheiten nach Altersgruppen oder Wohnort wurden für den MRSA bislang nicht veröffentlicht. Da diese Informationen aber bei der Schätzung der WLS von Bedeutung sind, ist die Hochrechnung nur unvollständig möglich.

Die hier vorgestellte Methodik versucht diese Informationslücke zu schließen, indem sie Regressionskoeffizienten des MRSA annähert. Die verwendeten Regressionskoeffizienten wurden im Laufe der Stellungnahmen zum MRSA vom BAS veröffentlicht, sind also repräsentativ für die Grundgesamtheit. Das Subsamplingverfahren macht sich diese Information zunutze, um die Frage beantworten zu können, ob durch Kombination von Subsamples einer Stichprobe eine bessere Vorhersage der Regressionsergebnisse auf der Grundgesamtheit erreicht werden kann als durch Hochrechnung der Stichprobe selbst?

Methodisches Vorgehen

Der hier verwendete Datensatz enthält rund neun Millionen der etwa 73 Millionen Versicherten in der GKV. Es handelt sich also um eine Stichprobe, die etwa zwölf Prozent der Grundgesamtheit darstellt. Diese Stichprobe ist allerdings nicht repräsentativ. Durch die regelmäßigen Meldungen der Grundlagenbescheide (GLB) I, II, III und IV, die im Laufe des Jahres für die Berechnung des MRSA an die Krankenkassen verschickt werden, sind Vergleichswerte verfügbar. Allerdings gilt das nur für die Randverteilung der im MRSA enthaltenen Variablen. Sowohl die Altersstruktur als auch die geografische Verteilung der Versicherten weichen stark von derjenigen der gesamten GKV ab. Noch größere und wichtigere Abweichungen zeigen sich allerdings bei der Betrachtung der Morbidität: Vor allem seltene Krankheiten sind stark über- oder unterrepräsentiert.

Diese Informationen werden in jedem herkömmlichen Hochrechnungsverfahren verwendet, um die Grundgesamtheit durch Gewichtung der einzelnen Beobachtungen zu approximieren. Auch im hier vorgestellten Verfahren werden diese Informationen verwendet, allerdings nur zur Beschleunigung des Verfahrens. Theoretisch könnte es

auch ohne Randverteilungen angewendet werden, der Rechenaufwand wäre aber wesentlich größer, da die Anzahl der benötigten Subsamples deutlich größer wäre, um eine ähnliche Schätzgüte zu erreichen.

Da das Subsamplingverfahren die Regressionsergebnisse aus der MRSA-Schätzung verwendet, muss der Datensatz so aufbereitet werden, dass diese Schätzung repliziert werden kann. Das bedeutet, dass der Datensatz genau die gleichen Variablen enthalten muss wie der Koeffizientenvektor des BAS. Für die Schätzung der Reformeffekte muss der Datensatz zusätzlich in der veränderten Spezifikation aufbereitet werden, sodass die Zuweisungen von Prä- und Post-Reform miteinander verglichen werden können. Beispielhaft bedeutet dies für die in der Einleitung erwähnte Risikopoolreform, dass jede Person zusätzliche Variablen für eine Hochkostengruppe erhält: Jede Person mit Unterdeckungen über 30.000 Euro wird einer Hochkostengruppe zugeordnet. Im Zuge dessen muss auch der Kostenvektor angepasst werden. Nach der vorgeschlagenen Risikopoolreform werden Kosten über die MRSA-Regression verteilt, die zuvor über einen separaten Topf, den Risikopool, ausgezahlt wurden. Trotz dieser grundlegenden Veränderung der Zuweisungsmodalitäten lässt sich mittels Subsampling ein Reformeffekt berechnen. Die gesamte Reform wird dabei durch eine Veränderung der Regressionskoeffizienten abgebildet. Multipliziert man diese mit der Matrix der Versichertenmerkmale, erhält man die Gesamtzuweisungen, die der Krankenkasse vor und nach der Reform zustehen. Die Differenz – korrigiert um die Auszahlungen aus dem abgeschafften Risikopool – bildet dann den Reformeffekt für diese Krankenkasse. Für die gesamte GKV ist der Reformeffekt stets Null, da die Finanzmittel lediglich anders zugewiesen werden, der zu verteilende Betrag aber konstant bleibt, sofern sich die tatsächlichen Leistungsausgaben nicht ändern.

Das Subsamplingverfahren erfolgt in vier Schritten:

- 1) Ziehen von Subsamples
- 2) Berechnen von MRSA-Regressionen auf jedem Subsample
- 3) Bilden von Ensembles mittels Linearkombination mehrerer Subsamples
- 4) Berechnung des Reformeffekts mit den besten Ensembles

Zunächst werden zufällige Subsamples aus dem bestehenden Datensatz gezogen. Im folgenden Schritt wird auf jedem dieser Subsamples eine WLS-Regression in der aktuellen Spezifikation des MRSA (Prä-Reform) und in der Spezifikation des zu untersuchenden Reformvorschlages durchgeführt (Post-Reform). Im dritten Schritt werden die errechneten Regressionskoeffizienten verwendet, um ein Ensemble aus mehreren Subsamples zu bilden, das die Grundgesamtheit besser repräsentiert. Im letzten Schritt werden die besten Ensembles verwendet, um die Differenz der Gesamtzuweisungen „Prä-“ und „Post-Reform“ zu berechnen. Die Anzahl der gezogenen Subsamples und der daraus erstellten Ensembles ist theoretisch unbeschränkt. Eine Begrenzung erfolgt in der Praxis nur aus Zeitgründen. Im Folgenden werden die einzelnen Schritte erläutert.

Ziehen von Subsamples

Im ersten Schritt werden aus dem Datensatz, der eine Stichprobe, also ein Sample der Grundgesamtheit darstellt, Subsamples gezogen. Die Ziehung der Subsamples erfolgt mit Zurücklegen und zufällig. Um die Ziehung der Subsamples zu verbessern, werden Gewichte verwendet, die die Subsamples in Richtung bekannter Merkmale der Grundgesamtheit bewegen. Dazu wird der Datensatz anhand der bekannten Randverteilungen mittels Entropy Balancing (Hainmueller 2012) hochgerechnet. Dieser Schritt ist durch die zur Verfügung stehenden Informationen eingeschränkt, stellt aber eine deutliche Verbesserung der Repräsentativität der Zellenbesetzungen der MRSA-Variablen dar.

Durch das Entropy Balancing wird jeder versicherten Person ein Hochrechnungsgewicht zugeordnet. Dieses Hochrechnungsgewicht würde im Normalfall zur Approximation der Grundgesamtheit dienen; für das hier verwendete Verfahren dient es stattdessen als Wahrscheinlichkeit, mit der die Beobachtung gezogen wird. Andere Hochrechnungsverfahren wären daher ebenfalls denkbar, man könnte sogar auf die Hochrechnung verzichten, da die Ziehung der Subsamples aus dem verfügbaren Datensatz erfolgt und nicht aus der Hochrechnung. Das Gewicht jeder Beobachtung hätte dann einen Wert von Eins. Der Vorteil der gewichteten Ziehung ist die schnellere Annäherung an den Zielvektor: Die Wahrscheinlichkeit dafür, ein Sample mit hoher Repräsentativität zu ziehen, ist ohne Hochrechnung so gering, dass deutlich mehr Zeit für das Ziehen von Stichproben verwendet werden müsste, was bei den jährlichen Stellungnahmen aufgrund der Fristen problematisch wäre.

Beim Subsampling werden aus dem bestehenden Datensatz wiederholt Stichproben mit Zurücklegen gezogen, deren Größe zwischen 60 und 90 Prozent des Datensatzes liegt. In diesem Schritt ist die Größe der Subsamples durch die Eigenschaften des MRSA bedingt. Kleinere Samples sind theoretisch möglich, führen in der Praxis aber zu Problemen bei der Schätzung, da es in den HMGs seltener Krankheiten dazu kommen kann, dass keine Beobachtungen mit dieser Ausprägung gezogen werden, wodurch leere Spalten entstehen. Die Durchführung der MRSA-Regression ist auf so einem Subsample nicht möglich, da dessen Versichertenmatrix mit leeren Spalten während der Regression nicht invertierbar ist, was die Berechnung unmöglich macht. Da diese Methode darauf abzielt, Reformen des MRSA zu bewerten, müssen die Samples sowohl Prä-Reform als auch Post-Reform für die WLS-Regression verwendet werden. Das bedeutet: Auch wenn sich die Gruppierung durch die Reform ändert, muss die Matrix invertierbar bleiben. Deshalb funktionieren Subsamples mit weniger als drei Millionen Beobachtungen in der Regel nicht, was ihre Verwendung in der Praxis einschränkt.

Wie bereits erläutert, wird die Hochrechnungsinformation in diesem Schritt als Ziehwahrscheinlichkeit verwendet. Individuen, die laut Hochrechnung repräsentativer sind, werden mit höherer Wahrscheinlichkeit gezogen. Da es sich um eine Ziehung mit Zurücklegen handelt, ist die Verteilung der Personen-IDs nicht mehr zufällig, sondern um die Hochrechnungsergebnisse des Entropy Balancing normalverteilt. Die bedingten Mittelwerte aller Subsamples nähern sich den bekannten Randverteilungen des BAS-Datensatzes an, variieren dabei aber in den unbekanntem Verteilungen aller MRSA-Variablen. Ein Nachteil dieses Vorgehens besteht, wie bei jedem Samplingverfahren, in der Abhängigkeit vom vorliegenden Datensatz: Variablenkombinationen, die in der Stichprobe nicht vorhanden sind, können durch Subsampling nicht synthetisiert werden.

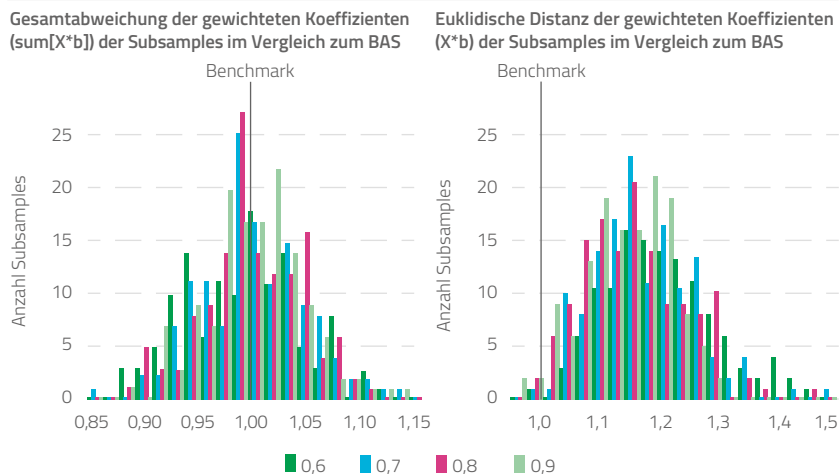
In der Praxis hat es sich bewährt, zur weiteren Berechnung mindestens 200 Subsamples zu ziehen, da sich mit kleineren Subsamplesätzen nur geringfügige Verbesserungen gegenüber der herkömmlichen Hochrechnung ergeben.

MRSA-Regression auf Prä- und Post-Subsamples durchführen

Im zweiten Schritt werden auf jedem der Subsamples die MRSA-Regressionen durchgeführt. Die Regression wird auf jedem Subsample in der Konfiguration „vor der Reform, (Prä)“ und in der Konfiguration „nach der Reform, (Post)“ berechnet. Es werden alle Samples aussortiert, auf denen entweder Prä- oder Post-Reform kein Ergebnis der WLS-Regression berechnet werden kann. Wie im vorangegangenen Abschnitt erläutert, ist dieses Problem auf perfekte Multikollinearität zurückzuführen, wenn beispielsweise für seltene Regressionsgruppen – normalerweise HMGs – keine Beobachtungen gezogen wurden. Dies führt in Abhängigkeit des Datensatzes zu einem kleinen Bias bei seltenen Krankheiten, da manche Diagnosen nur von wenigen Personen – im vorliegenden Datensatz waren es 20 HMGs, die weniger als 50 Personen enthielten – repräsentiert werden. Subsamples, die diese Personen nicht enthalten, werden aussortiert, weshalb bei diesen Krankheiten die Menge der zur Verfügung stehenden Informationen reduziert ist. Die Wahrscheinlichkeit, dass mehrere Samples genau die gleichen Repräsentanten für seltene HMGs haben, ist hoch. Daraus folgt, dass die Interaktion der seltenen HMG mit anderen HMGs nur bedingt erfasst werden kann. Dies ist aber durch den Datensatz bedingt und kann durch die Limitation des Verfahrens – es können nur bestehende Beobachtungen gezogen werden – nicht ausgeglichen werden.

Abbildung 1 zeigt die Abweichung der Regressionsergebnisse vom BAS-Vektor auf einzelnen Subsamples verglichen mit der Abweichung der Regressionsergebnisse vom BAS-Vektor auf der hochgerechneten Stichprobe, das in diesem Schritt als Benchmark dient. Die Färbung zeigt die Größe der einzelnen Subsamples als Anteil der Grundgesamtheit (60 bis 90 Prozent). Links wird die Abweichung in den Gesamtzuweisungen dargestellt, die wie zu erwarten die Form einer Normalverteilung annimmt. Die Spannweite der Abweichungen beträgt etwa 30 Prozent, mehr als 80 Prozent der Subsamples befinden sich im Bereich von plus/minus fünf Prozent um die Benchmark-Linie. Auf der rechten Seite wird die Abweichung der euklidischen Distanz des berechneten Koeffizientenvektors zum BAS-Vektor gezeigt. Wie man sieht, stellen die meisten Subsamples eine Verschlechterung dar, nur etwa fünf Prozent der Subsamples sind bereits allein eine Verbesserung gegenüber der Hochrechnung. Dieser Zugewinn wird durch den Kombinationsschritt weiter verbessert.

Abbildung 1: Distanzmaße von 200 Beispielsubsamples



Quelle: eigene Darstellung

Ensembles mittels Linearkombination mehrerer Subsamples bilden

Im dritten Schritt werden die Regressionsergebnisse auf den Prä-Reform-Subsamples mit den Regressionsergebnissen des BAS verglichen, die im Vorfeld der Stellungnahmen veröffentlicht werden. Der Vergleich der auf den Subsamples berechneten Regressionsvektoren mit dem BAS-Vektor stellt den eigentlichen Zugewinn an Information gegenüber der Hochrechnung dar.

Aus den Koeffizientenvektoren auf den gezogenen Prä-Reform-Subsamples wird eine gewichtete Linearkombination gebildet, um ein repräsentatives Ensemble zu erstellen. Die dabei berechneten Gewichte werden dann auch für die jeweiligen Post-Reform-Samples verwendet. Die Subsamples werden so zu einem Ensemble kombiniert, dass der Abstand zu den repräsentativen Koeffizienten des BAS-Vektors minimiert wird. Bei der Bildung dieser Ensembles hat sich eine Einschränkung der einzelnen Gewichte bewährt: Die Gewichtung der einzelnen Subsamples im Intervall $[-1,1]$ ermöglicht die Abdeckung einer großen Bandbreite möglicher Repräsentationen. In der Praxis zeigt sich, dass bei großen Ensembles aus mehr als 200 Subsamples die meisten der Subsamples

aussortiert werden, indem ihnen ein Gewicht von Null zugewiesen wird. Da eine große Zahl der Samples aufgrund der zufälligen Auswahl der enthaltenen Personen eine schlechte Repräsentation der Grundgesamtheit darstellt, ist dies auch zu erwarten. Dieser Umstand stellt aber kein Problem dar. In diesem Verfahren entsteht die Verbesserung durch Linearkombination von vielen Subsamples zu einem Ensemble, nicht durch die einzelnen Subsamples.

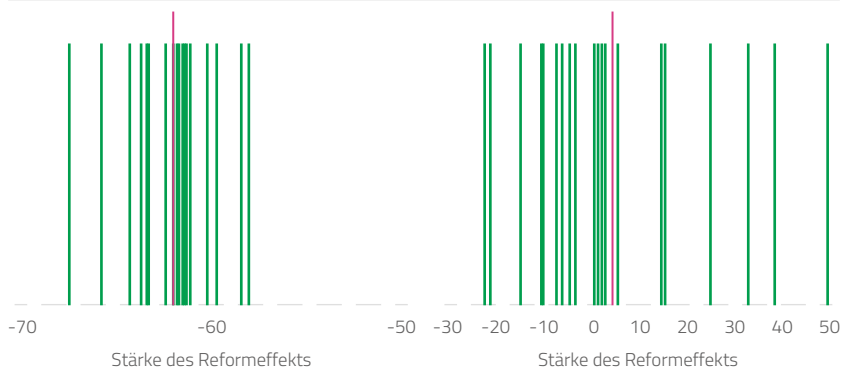
Die Berechnung der Ähnlichkeit der Koeffizientenvektoren kann auf mehrere Arten geschehen: Man könnte entweder die Distanz der einzelnen Koeffizienten vergleichen, die Distanz der Zuweisungen, die für jede Gruppe entstehen, wenn man Koeffizienten mit der Anzahl der Versicherungsjahre multipliziert, oder die Gesamtzuweisungen, die für alle Versicherten der Krankenkasse ausgezahlt werden. Für die Annäherung an die Grundgesamtheit sind allerdings die Zuweisungen der Gruppen am wichtigsten, da sie die strukturelle Ähnlichkeit der Stichprobe zur Grundgesamtheit am besten abbilden – sie beachten sowohl die Höhe der Koeffizienten als auch die Anzahl der von ihnen betroffenen Personen.

Zusätzlich zur untersuchten Variable stellt sich die Frage des Abstandsmaßes. Im Projekt wurden die euklidische Distanz, die quadrierte euklidische Distanz, der absolute Abstand und verschiedene Kombinationen der Abstandsmaße getestet. Die Methode funktioniert mit allen Abstandsmaßen und Variablen, in der Praxis hat sich aber die Zuweisung pro Gruppe, gemessen mit quadrierter euklidischer Distanz bewährt. Dieses Abstandsmaß garantiert eine hohe strukturelle Ähnlichkeit der Linearkombination, da die zugrundeliegende Struktur durch die Zuweisungen am besten abgebildet wird und durch die quadrierte euklidische Distanz ein konvexes Maß für die Abweichung genutzt werden kann, das große Abweichungen stark bestraft. Durch Konvexität des Maßes ist die Existenz eines globalen Minimums gewährleistet, das für jedes Ensemble eindeutige Kombinationsgewichte ergibt. Das Abstandsmaß impliziert auch eine stärkere Gewichtung von Gruppen mit hoher Auszahlung, da die Abweichung in diesen Gruppen sehr viel stärker bestraft wird als bei Gruppen mit kleiner Beobachtungszahl oder kleinen Koeffizienten.

Berechnung des Reformeffekts mit den besten Ensembles

In diesem Schritt werden Reformeffekte aus einzelnen Ensembles berechnet: Für jedes Ensemble, das Prä-Reform auf Ähnlichkeit zur Grundgesamtheit optimiert wurde, werden die Zuweisungen vor und nach der Reform berechnet. Dabei werden jeweils die gleichen Subsamples und Ensemblegewichte benutzt. So ist sichergestellt, dass Prä- und Post-Reform die gleichen Personen mit gleicher Gewichtung verglichen werden. Die Differenz der Gesamtauszahlung stellt den Reformeffektschätzer für das Ensemble dar. Durch die Bildung von verschiedenen Ensembles aus den verfügbaren Subsamples lässt sich eine ganze Reihe von Reformeffektschätzern produzieren. Da die Optimierung der Linearkombination deterministisch ist, entstehen die Unterschiede zwischen den Ensembles durch die Verwendung unterschiedlicher Subsamples. Die Verteilung der Reformeffektschätzer ergibt nicht nur ein Intervall von möglichen Reformeffekten, ihre Streuung gibt auch Auskunft über die Schätzungenauigkeit. In einzelnen Fällen, zum Beispiel bei der eingangs beschriebenen Risikopoolreform, müssen weitere Änderungen an den Zuweisungen berücksichtigt werden.

Abbildung 2: Beispielhafte Darstellung von Reformeffektschätzern (in Millionen Euro)



Quelle: eigene Darstellung

Abbildung 2 zeigt das Ergebnis zweier Reformeffektschätzungen auf Basis von jeweils 500 Subsamples. Jeder Strich ist der Reformeffekt (in Millionen Euro), der auf einem Ensemble berechnet wurde. Für beide Schätzungen wurden jeweils 20 Ensembles gebildet, hier dargestellt sind die Reformeffektschätzer auf den 20 besten Ensembles und ihr Durchschnitt als rote Linie. Die Ausgestaltung der Reform, die Anzahl der Subsamples, die Wahl des Abstandsmaßes für die Ensemblebildung und die Größe der Ensembles beeinflussen die Streuung der Schätzer.

Zur Evaluation der Methodik dient der Fall eines Reformvorschlags, der bereits vom BAS auf der Grundgesamtheit simuliert wurde. Für diesen Fall sind daher der Reformeffekt und die damit verbundene Veränderung der Koeffizienten sowie die Veränderung der Zuweisungen an die Krankenkassen bekannt. Er stellt somit das ideale Benchmark dar, um die besten Kombinationsgewichte und Abstandsmaße für das hier vorgestellte Verfahren zu finden. Nichtsdestoweniger muss erwähnt werden, dass dies das einzige zur Verfügung stehende Benchmarking war: Unter Umständen hätten andere Reformvorschlüsse zu einer anderen Wahl des Abstandsmaßes geführt. Es ist daher empfehlenswert, bei der Verwendung dieser Methodik weitere verfügbare Benchmarks zu verwenden, um die besten Abstandsmaße für den vorliegenden Anwendungsfall zu finden. Als Benchmarks können offizielle Schätzungen des BAS zu aktuellen Reformvorschlüssen verwendet werden. In Zukunft können auch die Ergebnisse von Reformen am aktuellen Modell verwendet werden, beispielsweise die Auswirkungen der Manipulationsbremse, die jedes Jahr mehrere HMGs streichen wird. Das Verfahren kann dann so kalibriert werden, dass die beobachteten Veränderungen in der Simulation getroffen werden.

Historische Reformen können ebenfalls verwendet werden, der Erkenntnisgewinn ist allerdings eingeschränkt: Durch die Reformen, die im Jahr 2020 im Zuge des Fairer-Kassenwettbewerb-Gesetzes (GKV-FKG 2020) beschlossen wurden, wurde die Anzahl im Modell abgebildeter Krankheiten stark erhöht, weshalb nun deutlich mehr HMGs in der Regression verwendet werden. Die Struktur der alten Daten hat daher deutliche Lücken. Viele Informationen, die heute relevant sind, wurden vor dem Jahr 2020 nicht für den MRSA erfasst.

Hauptaussagen und Ergebnisse

Die Annäherung an die BAS-Regression ist mit dem neuen Verfahren besser und erlaubt damit eine bessere Schätzung von Reformeffekten. Das Verfahren produziert allerdings keine neue Stichprobe, die für andere Auswertungen verwendet werden könnte, sondern repräsentativere Regressionsergebnisse. Außerdem bleibt die Unsicherheit bei der Erfassung von Beziehungen zwischen Variablengruppen, die im vorliegenden Datensatz nicht repräsentiert werden, bestehen.

Das Verfahren erlaubt es, eine Reihe von Reformeffektschätzern zu produzieren, deren Verteilung die Unsicherheit der Schätzung quantifizieren kann. Bei dem so produzierten Reformintervall handelt sich aber nicht um ein echtes Konfidenzintervall, da trotz Annäherung an den Koeffizientenvektor des BAS weiterhin unsicher ist, welche Reformeffektschätzer die Reform am besten abbilden. Insgesamt erlaubt das Verfahren einzelnen Krankenkassen bessere Vorhersagen darüber zu treffen, ob ein Reformvorschlag positive oder negative Effekte auf ihre Zuweisungen hätte.

Fazit

Das Verfahren ist ein deutlicher Fortschritt bei der Simulation von Reformeffekten. Die einzelnen Reformeffektschätzer bieten eine bessere Annäherung an die real zu erwartenden Effekte als eine Hochrechnung. Durch die Bildung eines Reformeffektintervalls aus mehreren Schätzern lässt sich auch die Schätzungsgenauigkeit quantifizieren, was bislang nicht möglich war. Das Verfahren ist aber auf die Regressionsergebnisse aus dem MRSA angewiesen und lässt sich daher nur in einem ähnlichen Kontext verwenden: Ohne Regressionsergebnisse würde das Abstandsmaß für die korrekte Kombination der Subsamples fehlen, was die Berechnung von Reformeffekten unmöglich machen würde.

Durch das hier vorgestellte Subsamplingverfahren wird die zukünftige Entwicklung des MRSA für die beteiligten Krankenkassen transparenter, da es jeder Krankenkasse eine bessere Einschätzung von Reformvorschlägen ermöglicht. Die daraus gewonnenen Erkenntnisse können im Prozess der Stellungnahmen dazu genutzt werden, Reformen zu identifizieren, die den MRSA nachhaltig verbessern. Reformen, die bestehende

Unter- und Überdeckungen der Krankenkassen abbauen, können unterstützt werden, während Reformen, die in höherer Ungleichverteilung resultieren, frühzeitig identifiziert und vermieden werden können. Weiterhin trägt eine korrekte Schätzung der zu erwartenden Reformeffekte zur finanziellen Stabilisierung der Krankenkassen bei. Das Subsampling leistet einen wesentlichen Beitrag dazu, das Vertrauen in die individuellen MRSA-Ergebnisse zu erhöhen.

Literatur

- Gesetz für einen fairen Kassenwettbewerb in der gesetzlichen Krankenversicherung (Fairer-Kassenwettbewerb-Gesetz – GKV-FKG) vom 22. März 2020.
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. In: *Political Analysis* 2012 (20). S. 25–46.
- IGES, Lauterbach, K., Wasem, J., Reschke, P., Sehlen, S., Schiffhorst, G. und W. F. Schröder (2005). Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Untersuchung zur Auswahl geeigneter Gruppenbildungen, Gewichtungsfaktoren und Klassifikationsmerkmale für einen direkt morbiditätsbezogenen Risikostrukturausgleich in der gesetzlichen Krankenversicherung. Bundesministerium für Gesundheit, Berlin.
- Schillo, S., Lux, G., Wasem, J. und F. Buchner (2016). High cost pool or high cost groups - How to handle high(est) cost cases in a risk adjustment mechanism? doi: 10.1016/j.healthpol.2016.01.003.
- Sozialgesetzbuch Fünftes Buch – SGB V. Paragraphen 265 bis 271 (Festlegungen zum morbiditätsorientierten Risikostrukturausgleich).
- Verordnung über das Verfahren zum Risikostrukturausgleich in der gesetzlichen Krankenversicherung in der Fassung vom 11. Juli 2021.