



**Bundesamt
für Soziale Sicherung**

Bestimmung

des Bundesamtes für Soziale Sicherung

nach

§ 273 Absatz 2 Satz 3 und Absatz 3 Satz 4 SGB V

**zur kassenübergreifenden Auffälligkeitsprüfung für
das Berichtsjahr 2013**

**im Rahmen der Sicherung der Datengrundlagen für den
Risikostrukturausgleich (RSA)**

in der Fassung vom 19.02.2021

Inhaltsverzeichnis

VORBEMERKUNG	5
1 INHALT UND ANWENDUNGSBEREICH	6
2 RAHMENBEDINGUNGEN DER AUFFÄLLIGKEITSPRÜFUNG	7
2.1 Prüfgegenstand	7
2.2 Vergleichskenngrößen und -zeitpunkte.....	7
2.3 Versicherten-Klassifikationsmodell als Berechnungsgrundlage	7
3 ANALYSESCHWERPUNKT UND PRÜFAUFBAU	9
4 GRUPPIERUNG	10
5 REGRESSIONSANALYSE	12
5.1 Vorbemerkungen	12
5.2 Kurzerklärung des Modells	13
5.3 Regressionsgleichung.....	14
5.4 Analysegröße.....	17
6 ERHEBLICHKEITSPRÜFUNG.....	18
6.1 Vorbemerkung	18
6.2 Analysegröße.....	18
7 SCHWELLENWERTE	22
7.1 Schwellenwerte für die Regressionsanalyse.....	22
7.2 Schwellenwert für die Erheblichkeitsprüfung	22
8 MITTEILUNGEN	24
8.1 Dokumentation für den GKV-Spitzenverband.....	24
8.2 Bekanntgabe der Ergebnisse.....	25
9 PRÜFKONSEQUENZEN, GESONDERTE BESTIMMUNGEN	26
10 ANHANG: AUSFÜHRLICHE DARSTELLUNG DES VERWENDETEN REGRESSIONSMODELLS.....	27

10.1	Generalized Linear Models	27
10.2	Poisson-Verteilung.....	28
10.3	Linearer Prädiktor und Link-Funktion	29
10.4	Hauptmodell.....	29
10.5	Interpretation der Regressionskoeffizienten.....	32

Gemäß § 273 Absatz 2 Satz 3 und Absatz 3 Satz 4 SGB V

bestimmt das Bundesamt für Soziale Sicherung

im Benehmen mit dem GKV-Spitzenverband

das Nähere über die

Prüfung zur Feststellung einer monetär erheblichen Auffälligkeit als kassenübergreifende

Vergleichsanalyse

wie folgt:

Vorbemerkung

Nach § 273 SGB V überprüft das Bundesamt für Soziale Sicherung (BAS) die Daten nach § 267 Absatz 1 Satz 1 SGB V auf ihre Rechtmäßigkeit. Gegenstand dieser Überprüfung ist insbesondere die Zulässigkeit der Meldung von Diagnosedaten entsprechend den Vorschriften zum Einwirkungsverbot und zur Übermittlung und Aufbereitung von Leistungsdaten.

Die kassenübergreifende Auffälligkeitsprüfung nach § 273 Absatz 2 und Absatz 3 SGB V ist als standardisiertes Prüfverfahren wie folgt konzipiert:

Im ersten Schritt werden die Diagnosedaten der Krankenkassen bezüglich Häufigkeit und Schwere auf Ebene der Hierarchisierten Morbiditätsgruppen (HMGs) des Versichertenklassifikationsmodells auf eine statistische Auffälligkeit untersucht (§ 273 Absatz 2 SGB V).

Im zweiten Schritt wird bei Krankenkassen mit statistischen Auffälligkeiten untersucht, ob die gefundenen Auffälligkeiten auf Ebene der HMGs auch zu erheblich erhöhten Zuweisungen geführt haben können (sog. Erheblichkeitsprüfung, § 273 Absatz 3 SGB V).

Als Ergebnis dieser Prüfung wird den betroffenen Krankenkassen gemäß § 273 Absatz 3 Satz 3 SGB V mitgeteilt, dass eine monetär erhebliche Auffälligkeit festgestellt wurde, was eine Einzelfallprüfung gemäß § 273 Absatz 4 SGB V zur Folge hat.

Widerlegt die betroffene Krankenkasse im Zuge der Einzelfallprüfung gemäß § 273 Absatz 4 SGB V nicht die Vermutung, dass ein Rechtsverstoß die Auffälligkeit verursacht hat, erfolgt im letzten Schritt nach § 273 Absatz 6 SGB V die Ermittlung eines Korrekturbetrages.

1 Inhalt und Anwendungsbereich

Nach § 273 Absatz 2 Satz 3 und Absatz 3 Satz 4 SGB V bestimmt das BAS im Benehmen mit dem Spitzenverband Bund der Krankenkassen (GKV-Spitzenverband) das Nähere zur Durchführung der kassenübergreifenden Vergleichsanalyse, insbesondere die Schwellenwerte für die Feststellung einer statistischen und einer monetär erheblichen Auffälligkeit.

Gegenüber der Bestimmung zur kassenübergreifenden Auffälligkeitsprüfung in der Fassung vom 23.10.2017 ergeben sich z. T. weitreichende inhaltliche Änderungen. Diese sind im Wesentlichen auf die Neufassung des § 273 SGB V durch das GKV-FKG¹ zurückzuführen, das am 1. April 2020 in Kraft getreten ist. Die Änderungen betreffen die Beschränkung der Prüfung auf die sog. Korrekturmeldung sowie die Einführung der o.g. Erheblichkeitsprüfung, welche methodisch die Hinwendung zu einem Schwellenwert-Verfahren mit einem monetären Parameter zur Folge hat. Darüber hinaus ergeben sich weitere Anpassungen, die aus der Weiterentwicklung des Versichertenklassifikationsmodells im morbiditätsorientierten Risikostrukturausgleich (Morbi-RSA) resultieren und eine Abkehr von der getrennten Betrachtung ambulanter und stationärer Diagnosen mit sich bringen.

Im Folgenden werden die Rahmenbedingungen der Auffälligkeitsprüfung (Prüfgegenstand, relevante Vergleichsgrößen und -zeitpunkte, verwendetes Klassifikationsmodell, vgl. Abschnitt 2) sowie die maßgeblichen Analyseverfahren (Abschnitt 4 bis 6) und Schwellenwerte (Abschnitt 7) inklusiver der damit verbundenen Mitteilungs- und Dokumentationspflichten (Abschnitt 8) für die Prüfung der Daten des Berichtsjahres 2013 gemäß § 273 Absatz 2 Satz 3 und Absatz 3 Satz 4 SGB V spezifiziert.

¹ Gesetz für einen fairen Kassenwettbewerb in der gesetzlichen Krankenversicherung (siehe BGBl. I S.622, 623).

2 Rahmenbedingungen der Auffälligkeitsprüfung

2.1 Prüfgegenstand

Gegenstand der Auffälligkeitsprüfung sind jeweils die gemäß § 7 Absatz 4 Satz 2 RSAV gemeldeten Morbiditätsdaten (Satzart 400, Satzart 500 und Satzart 600 der Korrekturmeldung) eines Berichtsjahres nach Anwendung des BAS-Datenbereinigungskonzeptes.

Hinweis:

Gemäß § 273 Absatz 1 i. V. m. Absatz 2 SGB V ist die Prüfung der ambulanten Diagnosedaten obligatorisch, wohingegen die Prüfung der übrigen Morbiditätsdaten fakultativ ist. Diese werden jedoch durch die Betrachtung der gesamten Gruppierung (vgl. Abschnitt 4) in die Prüfung einbezogen. Auf diese Weise erfolgt eine genaue Abbildung der gesamten Morbiditätsdaten der Gesetzlichen Krankenversicherung (GKV). Außerdem ist der Zuordnungsalgorithmus des Morbi-RSA im Zuge seiner Weiterentwicklung stetig sektorübergreifender geworden, weswegen die separate Betrachtung der früheren Prüfbereiche der stationären Diagnosen (Satzart 500) und ambulanten Diagnosen (Satzart 600) nicht mehr sachgemäß ist. Die Entwicklung der verordneten Arzneimittel (Satzart 400) wird indirekt bei der Prüfung aller Diagnosedaten mit einbezogen. Eine Interpretation der Ergebnisse wird demnach einfacher und bildet die Bedingungen des Versichertenklassifikationsmodells des Morbi-RSA näher ab.

2.2 Vergleichskenngrößen und -zeitpunkte

Die Prüfung der zuweisungsrelevanten Daten beruht nach § 21 Absatz 1 RSAV auf dem Vergleich der zu prüfenden Korrekturmeldung des Berichtsjahres mit der Korrekturmeldung des vorangegangenen Berichtsjahres (im Folgenden Prüf- und Referenzmeldung bzw. Prüf- und Referenzjahr genannt). Wegen der fehlenden monetären Bedeutung für den Jahresausgleich entfällt die Prüfung der Erstmeldung.

2.3 Versicherten-Klassifikationsmodell als Berechnungsgrundlage

Prüf- und Referenzmeldung werden für die Prüfung auf Basis des Versicherten-Klassifikationsmodells gemäß § 8 Absatz 4 (§ 31 Absatz 4 a. F.) RSAV aufbereitet, um die Versicherten einer bzw. mehreren HMG(s) zuzuordnen (sog. Gruppierung, vgl. Abschnitt 4).

Für jedes Ausgleichsjahr existiert eine eigene Version des Klassifikationsmodells. Somit richtet sich die Wahl des Klassifikationsmodells nach dem Ausgleichsjahr, auf das sich die

Prüfmeldung bezieht. Dies ist – aufgrund der prospektiven Modellausrichtung – stets das Jahr, das auf das Berichtsjahr der Prüfmeldung folgt.

Für die Prüfmeldung dokumentiert Anlage 1 das jeweils verwendete Klassifikationsmodell mit genauen Angaben zu Ausgestaltungsdetails und technischen Spezifikationen.

Hinweis:

HMGs stellen jene Zuordnungsebene im Klassifikationsmodell dar, auf deren Basis die morbiditätsorientierten Zuschläge (HMG-Zuschläge) je Versicherten vergeben werden. Es ist daher davon auszugehen, dass sich manipulationsbedingte Abweichungen in den zu überprüfenden Daten auf dieser Ebene am ausgeprägtesten manifestieren. Gleichzeitig bedeutet die Aggregation der Diagnose- und auch der Arzneimitteldaten entsprechend dem für das Ausgleichsjahr gültigen Zuordnungsalgorithmus eine Verdichtung der aus den rohen Morbiditätsdaten zu gewinnenden Informationen. Dabei stellen die HMGs als Analyseebene insbesondere sicher, dass die Auffälligkeitsprüfung als Kriterium auch die unterschiedlichen Schweregrade der Diagnosen berücksichtigt, vgl. § 273 Absatz 2 Satz 1 SGB V.

Seit dem Ausgleichsjahr 2012 werden stationäre Nebendiagnosen einer strikteren Validierung unterzogen, um die Manipulationsresistenz des Klassifikationsmodells zu erhöhen. Stationäre Nebendiagnosen werden seither in den meisten Fällen den Diagnosen aus der vertragsärztlichen Versorgung gleichgestellt. Dadurch werden stationäre Nebendiagnosen prinzipiell bei der Prüfung der ambulanten Aufgreifkriterien M2Q sowie den Arzneimittelkriterien mitberücksichtigt. Ausnahmen hiervon sind in den jeweils gültigen Festlegungen nach § 8 Absatz 4 (§ 31 Absatz 4 a. F.) RSAV dargelegt.

3 Analyseschwerpunkt und Prüfaufbau

Der statistische Analyseschwerpunkt liegt auf der Entwicklung der HMG-Besetzung (Anzahl der Versicherten mit einer HMG-Zuordnung) von der Referenz- zur Prüfmeldung auf Einzelkassenebene. Die Zuordnung der Versicherten zu den HMGs wird aus der sog. Prüfgruppierung gewonnen, die auf Grundlage des jeweils gültigen Versichertenklassifikationsmodells (vgl. Abschnitt 2.3) und mit den speziell für die Auffälligkeitsprüfung geltenden Spezifikationen (vgl. Abschnitt 4) durchgeführt wird.

Aus dem Vergleich der Gruppierungsergebnisse für Prüf- und Referenzmeldung lassen sich die für die statistische Auffälligkeitsprüfung relevanten Kennzahlen erzeugen. Dies geschieht auf Basis einer Poisson-Regression, die für sämtliche HMGs des betreffenden Jahresausgleichs durchgeführt wird (vgl. Abschnitt 5).

Anschließend wird gemäß § 273 Absatz 3 Satz 1 SGB V geprüft, ob die festgestellten statistischen Auffälligkeiten auf Kassenebene zu einer erheblichen Steigerung der Zuweisungen aus dem Gesundheitsfonds geführt haben können. Diese Vorgabe wird durch die Erheblichkeitsprüfung, welche in Abschnitt 6 beschrieben ist, umgesetzt.

Die Schwellenwerte für die statistischen Auffälligkeiten und die Erheblichkeitsprüfung werden in Abschnitt 7 bestimmt.

4 Gruppierung

Prüf- und Referenzmeldung werden nach den Vorgaben des unter 2.3 spezifizierten Klassifikationsmodells gruppiert. Dabei gelten folgende Abweichungen von der regulären Gruppierung:

- 1) Als Stammdaten-Datei wird ausschließlich die zugehörige SA100 verwendet.
- 2) Die Versichertenzuordnung erfolgt auf Grundlage des kassenindividuellen Pseudonyms.
- 3) Die Zuordnung der Versicherten zu den Morbiditätsgruppen erfolgt zeitgleich.
- 4) Versicherte mit null Versichertentagen werden nicht berücksichtigt, d. h. ihnen werden zwar AGGs, aber keine HMGs zugeordnet.
- 5) Kostenerstatter und Auslandsversicherte (Zuordnung erfolgt ebenfalls über das kassenindividuelle Pseudonym und auf Basis der SA100) werden nicht berücksichtigt, d. h. ihnen werden zwar AGGs, aber keine HMGs zugeordnet.
- 6) Daten, für die nach § 21 Abs. 2 Satz 1 Nr. 1 (§ 39a Abs. 3 a.F.) RSAV ein Korrekturbetrag ermittelt worden ist, werden von der Gruppierung ausgeschlossen. Diese Daten werden über eine Liste mit den auszuschließenden Datensätzen (sog. „Diagnose-Streichliste“) entfernt oder durch eine bereinigte Datenmeldung (sog. „Austauschsatzart“) ersetzt.

Die Gruppierungen beziehen das gesamte HMG-Spektrum mit ein.

Da bei der Prüfung einer Korrekturmeldung die zugehörige Referenzmeldung i. d. R. einen unterschiedlichen Fusionsstand aufweist, wird die betroffene Referenzmeldung stets auf den gleichen Fusionsstand wie die zu prüfende Korrekturmeldung gebracht. Dies geschieht *nach* der Gruppierung der Daten (eine Aggregation der Datenmeldungen auf Versichertenebene findet folglich fusionsbedingt nicht statt).

Zur Validierung wird wahlweise die Gruppierung der Referenz- oder Prüfmeldung auf Ebene der unbereinigten Daten (d.h. ohne Berücksichtigung von „Diagnose-Streichlisten“ und „Austauschsatzarten“) mit dem GKV-Spitzenverband abgeglichen.

Hinweis:

Die Zuordnung zu den unter 5) genannten Gruppen weicht von der regulären Gruppierung ab, da eine Aggregation der Versichertentage auf Ebene des GKV-Pseudonyms nicht stattfindet. Es kann also theoretisch bei Krankenkassenwechslern vorkommen, dass diese – anders als bei der regulären Gruppierung – den Status „Kostenerstatter“ oder „Auslandsversicherte“ nicht erhalten und in der Folge für diese

Versicherte eine Zuordnung zu HMGs erfolgt. Es handelt sich hierbei jedoch in der Praxis allenfalls um eine äußerst seltene Randerscheinung ohne zahlenmäßige Relevanz.

5 Regressionsanalyse

5.1 Vorbemerkungen

Der statistische Analyseschwerpunkt liegt auf der Veränderung der Besetzung in den einzelnen HMGs beim Übergang von Referenz- zu Prüfmeldung auf Ebene der Einzelkasse. Die Veränderung der Besetzungszahlen wird in Form von Relativen Risiken dargestellt. Diese und weitere Kennzahlen zur Identifikation der kassenbezogenen statistischen Auffälligkeiten werden mit einer Regressionsanalyse ermittelt.

Dieses statistische Verfahren dient dazu, auffällige Abweichungen sichtbar zu machen, die von einem zum nächsten Jahr in bestimmten HMG-Bereichen bei einzelnen Krankenkassen auftreten. Eine Regressionsanalyse zeichnet sich vor allem dadurch aus, dass sie den Einfluss mehrerer unabhängiger Einflussgrößen auf eine Zielvariable darzustellen vermag. So lässt sich mit einer multiplen Regression bspw. berücksichtigen, welchen Einfluss die Faktoren Alter und Geschlecht auf das erhöhte Auftreten einer bestimmten HMG bei einer Krankenkasse haben. Indem die GKV-weit ermittelten HMG-Besetzungszahlen standardmäßig als Referenzkategorie in die Regression mit einbezogen werden, ist außerdem der Bezug zur GKV-Entwicklung automatisch gegeben.

Die über einen Regressionsansatz zugänglichen Rechen- und Modellierungsoptionen sind für die hier angestrebte Analyse daher sehr wertvoll. Gleichzeitig machen sie das Verfahren deutlich komplexer und schwerer darstellbar. Den Abschnitten 5.3 und 5.4, die Modell und Analysegrößen definieren, werden daher eine verbale Erläuterung unter 5.2 und ein ausführlicher Anhang zur Seite gestellt, um das Verfahren nachvollziehbar und verständlich zu machen.

Hinweis:

Indem Alters- und Geschlechtsmerkmale der Versicherten sowie die GKV-weiten Besetzungszahlen in der Regression berücksichtigt werden, wird die Bedingung umgesetzt, dass gemäß § 273 Absatz 2 Satz 1 nur jene auffälligen Steigerungen untersucht werden sollen, die sich nicht auf demografische Veränderungen des Versichertenbestandes beziehen. Der Einfluss von Alter und Geschlecht auf die abhängige Variable wird dadurch herausgerechnet.

5.2 Kurzerklärung des Modells

Die Regressionsanalyse wertet die Unterschiede in den Gruppierungsergebnissen von Prüf- und Referenzmeldung aus. Für die Prüfung werden für sämtliche HMGs einzelne Regressionen durchgeführt.

Das Vorliegen einer HMG bildet die abhängige Variable in der Regression. Als unabhängige Variablen gehen die 40 Altersgruppen², die in der Prüfmeldung vorhandenen Krankenkassen und die zugrundeliegende Meldung (Prüf- vs. Referenzmeldung) in die Regression ein. Um die GKV-weite Entwicklung zu berücksichtigen, wird die Gesamt-GKV als zusätzliche Krankenkasse modelliert. Die Gesamt-GKV wird somit zur Referenzkategorie in der Dimension „Krankenkasse“. Die Ergebnisse der Gruppierung werden für die Durchführung der Regression nach den genannten Merkmalen gruppiert. Genau genommen ist somit die HMG-Anzahl in der jeweiligen Gruppe die Zielvariable.

Ausschlaggebend für die Analyse ist ein weiterer Parameter, der in die Regression mit aufgenommen wird: ein Interaktionsterm der beiden Variablen „Krankenkasse“ und „Meldung“. Über ihn lässt sich die Wechselbeziehung dieser beiden Einflussfaktoren bestimmen. Er erlaubt also eine Aussage darüber, in welchem Maße der Einfluss des Meldewechsels auf die HMG-Entwicklung von der einzelnen Krankenkasse abhängt. Die allgemeine GKV-Entwicklung bildet den Maßstab für die Stärke des Effekts; die relativierenden Effekte von Alter und Geschlecht sowie Größe und Morbiditätsniveau der Krankenkasse sind über die restlichen Einflussfaktoren in der Regression berücksichtigt.

Als relevante Analysegröße werden einzig die Regressionskoeffizienten verwertet, die aus dem Interaktionsterm Krankenkasse * Meldung für jede einzelne Krankenkasse resultieren. Mathematisch transformiert, ergibt sich aus den Regressionskoeffizienten das sogenannte Relative Risiko (RR). Darüber lässt sich eindeutig ablesen ($RR < > 1$), ob der Zusammenhang zwischen Meldewechsel und HMG-Entwicklung bei einer Krankenkasse unter oder über dem GKV-Trend liegt. Als Kennzahlen ergänzt werden die RRs durch die zugehörigen p-Werte. Sie geben das statistische Signifikanzniveau der jeweiligen Regressionskoeffizienten an.

² Die Zuordnung zu den 40 Altersgruppen (AGGs) erfolgt für alle Versicherten mit mehr als null Versichertentagen in der jeweiligen Satzart 100.

5.3 Regressionsgleichung

Die Zielvariable Y_i , also die Anzahl der Versicherten in der HMGx, wird mit einer **log-linearen Poisson-Regression** modelliert³. Bei einer Anzahl von r Einflussgrößen und n Beobachtungen⁴ lautet deren allgemeine Form:

$$\log(E(Y_i)) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir}, \text{ mit } i = 1, \dots, n, \quad (1)$$

wobei $E(Y_i)$ für den Erwartungswert der Zielvariablen der i -ten Beobachtung,

x_1 bis x_r für die unabhängigen Variablen (bzw. deren Realisierungen x_{i1} bis x_{ir}) und

β_1 bis β_r für die dazugehörigen Regressionskoeffizienten stehen⁵.

Mit den folgenden Spezifikationen für die rechte Seite der Regressionsgleichung ist das Modell für den hier vorgesehenen Zweck vollständig definiert⁶:

- 1) Die unabhängigen Variablen werden bestimmt. Sie sind im hier betrachteten Fall sämtlich kategorial, werden also durch 0/1-kodierte Dummy-Variablen abgebildet, genau durch:
 - die unabhängige Variable AGG „A“, mit 40 Ausprägungen und 40-1 Dummy-Variablen,

³ Für Näheres zur Poisson-Regression und deren Einordnung in die Klasse der GLM-Modelle s. die Abschnitte 10.1 und 10.2 im Anhang.

⁴ Eine Beobachtung entspricht dabei den nach den Merkmalen AGG, Krankenkasse und Meldung aggregierten Werten der Gruppierungsdaten auf Einzelversichertenebene. Siehe unter Punkt 1) die Erläuterung zu den unabhängigen Variablen.

⁵ Vgl. hierzu die Formel (1), Abschnitt 10.3 im Anhang. Letztere ist mit dieser identisch bis auf den Achsenabschnitt β_0 . Auf diesen kann hier und im Weiteren verzichtet werden, da das verwendete Modell keinen Achsenabschnitt vorsieht.

⁶ Eine Abbildung der vollständigen Regressionsgleichung findet sich im Anhang, Formel (2) im Abschnitt 10.4, dort sind auch die einzelnen Regressionsparameter näher erläutert.

- die unabhängige Variable Krankenkasse „ K “, mit $m+1$ Ausprägungen⁷ und m Dummy-Variablen (wobei m der Anzahl der Krankenkassen in der Prüfmeldung entspricht) und
 - die unabhängige Variable Meldung „ M “ mit 2 Ausprägungen und einer Dummy-Variablen.
- 2) Der sog. Offset wird ergänzt, der bei aggregierten Daten den Populationsbezug herstellt, hier also:

$\beta_G \log(G_i)$, wobei G_i die Gesamtversichertenanzahl je Krankenkasse, AGG und Meldung darstellt, mit einem Regressionskoeffizienten β_G von 1.

- 3) Der Interaktionsterm I wird als zusätzliche unabhängige Variable ergänzt, hier also:

die unabhängige Variable $I = M * K$, mit ebenfalls $m+1$ Ausprägungen und m Dummy-Variablen.

Hinweis:

Die Poisson-Regression ist das wissenschaftlich gebräuchlichste und gängigste Modell bei der Analyse von Zähldaten (nicht-negative, diskrete Zielgröße). Dies gilt insbesondere bei gruppierten Zähldaten, wie im hier betrachteten Fall. Es ist darauf hinzuweisen, dass das Poisson-Modell hier als Standardverfahren angewandt wird. Damit geht natürlicherweise einher, dass die Annahme einer Poisson-verteilter Zufallsvariable nicht in allen betrachteten Fällen exakt zutreffen kann. Dies ist jedoch unerheblich, solange die Aussagekraft der Ergebnisse hiervon nicht in nennenswertem Maße beeinträchtigt wird. Das BAS hat sich daher durch zahlreiche Voranalysen – insbesondere durch die Prüfung alternativer Verteilungsannahmen und eine Untersuchung der Randbereiche (gering besetzte HMGs) – vergewissert, dass die generalisierte Annahme einer Poisson-verteilter Zufallsvariable methodisch unbedenklich ist. Das theoretisch größte Problem einer in den Verteilungsannahmen unpräzisen Modellspezifikation sind unterschätzte Standardfehler. Es gibt jedoch keine Anhaltspunkte, dass es im hier betrachteten Kontext zu systematischen Verzerrungen dieser Art kommt. Als zusätzliche vorbeugende Maßnahme wird von vornherein ein strengeres Signifikanzniveau als generell üblich bei der

⁷ Die Anzahl der Krankenkassen beträgt $m + 1$, weil die Gesamt-GKV als zusätzliche Krankenkasse mit in die Regression eingeht, vgl. Abschnitt 10.4 im Anhang.

Bewertung der Regressionsschätzer angesetzt (vgl. Abschnitt 7), um die Validität der Ergebnisse sicherzustellen.

5.4 Analysegröße

Der relevante Interaktionsterm I bildet die multiplikative Verknüpfung der beiden Variablen Meldung und Krankenkasse ab, also

$$M * K_1 \text{ bis } M * K_m, \text{ bzw. } I_1 \text{ bis } I_m. \quad (2)$$

Dies führt bei m Krankenkassen zu m Regressionskoeffizienten, also $\beta I_1 - \beta I_m$.

Werden die βI -Parameter exponiert, berechnet man also $\exp\{\beta I_1\}$ bis $\exp\{\beta I_m\}$, ergibt sich das relative Risiko⁸:

$$RR(I_1) - RR(I_m), \text{ bzw. } RR(I_k), \text{ mit } k = 1, \dots, m. \quad (3)$$

Für jeden der βI -Parameter geben die zugehörigen p-Werte das jeweilige Signifikanzniveau an, $\beta I_1 - \beta I_m$ werden also ergänzt durch:

$$p(\beta I_1) - p(\beta I_m), \text{ bzw. } p(\beta I_k) \text{ mit } k = 1, \dots, m. \quad (4)$$

Die Kennzahlen aus (1) und (2) werden für sämtliche HMGs erzeugt (für jede HMG_x wird also eine separate Regression durchgeführt).

Hinweis:

Die RR_s geben an, wie sich die Besetzungszahlen in der jeweiligen HMG bei der betreffenden Krankenkasse gemessen am GKV-Schnitt beim Übergang von Vergleichs- zu Prüfmeldung entwickelt haben. Diese Entwicklung kann überproportional ($RR(I_k) > 1$) oder unterproportional ($RR(I_k) < 1$) sein. Die $p(\beta I_k)$ -values liefern hierzu den Signifikanzwert und geben somit an, ob es sich um eine statistisch signifikante Abweichung handelt.

⁸ Für Näheres zum Relativen Risiko und der Interpretation der Regressionskoeffizienten s. Abschnitt 10.5 im Anhang.

6 Erheblichkeitsprüfung

6.1 Vorbemerkung

Der neu gefasste § 273 Absatz 3 Satz 1 SGB V führt eine „Erheblichkeitsprüfung“ ein, die die statistische Auffälligkeitsfeststellung um eine zusätzliche monetäre Dimension erweitert. Zu diesem Zweck wird für jede Krankenkasse der mit der Prüfmeldung gegenüber der Referenzmeldung erzielte „HMG-Zuweisungsgewinn“ je Versicherten ermittelt, der sich für statistisch auffällige HMGs ergibt. Dieser (in Euro je Versicherten) ausgedrückte HMG-Zuweisungsgewinn wird anschließend einem vorab definierten monetären Schwellenwert gegenübergestellt.

Dabei gilt es herauszustellen, dass die Zuweisungsberechnung unter gewissen (standardisierten) Annahmen erfolgt, die zu einer besseren Vergleichbarkeit von Prüf- und Referenzmeldung führen sollen.

Zum einen ist der Ausgangspunkt für die hier berechneten Zuweisungen die *zeitgleiche* Gruppierung für Prüf- und Referenzjahr nach Abschnitt 4 dieser Bestimmung und die daraus resultierenden HMG-Besetzungszahlen bzw.- Prävalenzen auf Kassen- und GKV-Ebene.

Zum anderen werden bei der Berechnung der HMG-Zuweisungen vollständige Versichertenepisoden angenommen. Das bedeutet, dass durchgängig der Jahreswert (entspricht dem Pro-Tag-Zuweisungsbetrag je HMG, multipliziert mit 365 bzw. 366 Tagen) den Berechnungen der HMG-Zuweisungen zugrunde liegt, welcher einheitlich ein und demselben Jahresausgleichsbescheid entnommen wird; Bezugspunkt ist jenes Ausgleichsjahr, auf das sich die Prüfmeldung bezieht. Weiterhin werden Änderungen in der Alters- und Geschlechtsstruktur zwischen Referenz- und Prüfmeldung berücksichtigt. Der allgemeinen Entwicklung der HMG-Besetzungen auf GKV-Ebene wird ebenfalls Rechnung getragen. Diese Herangehensweise führt im Ergebnis dazu, dass die für die Erheblichkeitsprüfung berechneten Zuweisungen von den tatsächlichen (d.h. für den Jahresausgleich ermittelten) HMG-Zuweisungen abweichen.

6.2 Analysegröße

Die für die Erheblichkeitsprüfung herangezogene Analysegröße ist die Differenz der für Prüf- und Referenzjahr jeweils aufsummierten HMG-Zuweisungen je Krankenkasse – beschränkt auf HMGs mit auffälligen Steigerungsraten. Dieser Zuweisungsdifferenzbetrag wird anschließend durch die Gesamtversicherten der Krankenkasse geteilt, wodurch sich der „Zuweisungsgewinn“ je Versicherten im Prüfwahljahr ergibt.

Die Berechnung der HMG-Zuweisungen setzt auf den HMG-Besetzungszahlen für Referenz- und Prüfjahr auf. Die HMG-Besetzungszahlen aus dem Prüfjahr je Krankenkasse können direkt aus der Prüfgruppierung gewonnen werden.

Die genaue Berechnung der **HMG-Besetzungen einer Krankenkasse im Referenzjahr** stellt sich wie folgt dar:

Aus der zeitgleichen Gruppierung wird eine HMG-AGG-Matrix gebildet, in der die Anzahl der Köpfe je HMG nach AGG aggregiert sind.

KK	AGG	N_AGG	HMG001	HMG301
KK_1	AGG01	$N_{(KK_1,AGG01)}$	$N_{(KK_1,AGG01,HMG001)}$	$N_{(KK_1,AGG01,HMG301)}$
..	AGG02	$N_{(KK_1,AGG02)}$	$N_{(KK_1,AGG02,HMG001)}$	$N_{(KK_1,AGG02,HMG301)}$
..
KK_N	AGG40	$N_{(KK_N,AGG40)}$	$N_{(KK_N,AGG40,HMG001)}$	$N_{(KK_N,AGG40,HMG301)}$

KK = Krankenkasse

N = Anzahl Versicherte

Anschließend wird die Prävalenz für jeden Wert der HMG-AGG-Matrix berechnet, um das kassenindividuelle Risiko für jede HMG auf AGG-Basis zu berechnen⁹.

$$P_{(AGG,HMG,KK,RJ)} = \frac{N_{(AGG,HMG,KK,RJ)}}{N_{(AGG,KK,RJ)}} \quad (5)$$

$P_{(AGG,HMG,KK,RJ)}$ = Prävalenz für das Auftreten einer HMG in der betreffenden AGG einer Krankenkasse im Referenzjahr

$N_{(AGG,HMG,KK,RJ)}$ = Anzahl Versicherte einer HMG in der betreffenden AGG einer Krankenkasse im Referenzjahr

$N_{(AGG,KK,RJ)}$ = Anzahl Versicherte einer AGG einer Krankenkasse im Referenzjahr

Diese Prävalenzen werden dann mit der Gesamtzahl der Versicherten der AGGs des Prüfjahres multipliziert, um die Anzahl der HMG-Zuordnungen je AGG zu erhalten, die die kassenindividuellen Erwartungswerte abbilden. Dieser Schritt dient der Übertragung der veränderten Versichertenstruktur aus dem Prüfjahr auf das Referenzjahr (Alters- und Geschlechtsstandardisierung):

$$EW(N)_{(AGG,HMG,KK,RJ)} = P_{(AGG,HMG,KK,RJ)} \times N_{(AGG,KK,PJ)} \quad (6)$$

$EW(N)_{(AGG,HMG,KK,RJ)}$ = Erwartungswert der Anzahl der Versicherten einer HMG in der betreffenden AGG einer Krankenkasse bezogen auf das Referenzjahr

⁹ Für AGGs, die mit null Versicherten besetzt sind, wird eine Prävalenz von null festgelegt.

$P_{(AGG,HMG,KK,RJ)}$ = Prävalenz für das Auftreten einer HMG in der betreffenden AGG einer Krankenkasse im Referenzjahr
 $N_{(AGG,KK,PJ)}$ = Anzahl Versicherte einer AGG einer Krankenkasse im Prüffahr

Die so zu erwartende Anzahl an Versicherten in der HMG-AGG-Matrix bezogen auf das Referenzjahr wird auf HMG-Ebene aggregiert:

$$\sum_{AGG_{01}}^{AGG_{40}} EW(N)_{(AGG,HMG,KK,RJ)} = N_{(HMG,KK,RJ)}. \quad (7)$$

$EW(N)_{(AGG,HMG,KK,PJ)}$ = Erwartungswert der Anzahl der Versicherten mit HMG in der betreffenden AGG einer Krankenkasse bezogen auf das Referenzjahr
 $N_{(HMG,KK,RJ)}$ = Anzahl der Versicherten mit HMG einer Krankenkasse im Referenzjahr (nach AGG-Bereinigung)

Aus den alters- und geschlechtsbereinigten HMG-Besetzungszahlen für das Referenzjahr und den HMG-Besetzungszahlen für das Prüffahr werden kassenindividuelle HMG-Veränderungsraten gebildet:

$$\frac{N_{(HMG,KK,PJ)}}{N_{(HMG,KK,RJ)}} - 1 = Vrate_{(HMG,KK)}, \quad (8)$$

$N_{(HMG,KK,PJ)}$ = Anzahl der Versicherten einer HMG einer Krankenkasse im Prüffahr
 $N_{(HMG,KK,RJ)}$ = Anzahl der Versicherten einer HMG einer Krankenkasse im Referenzjahr (nach AGG-Bereinigung)

um anschließend die Veränderungsrate der zu betrachtenden HMG der Krankenkasse mit der der GKV zu vergleichen. Überschreitet die Veränderungsrate der Krankenkasse die der GKV, wird die GKV-Veränderungsrate auf die HMG-Besetzung der Krankenkasse im Referenzjahr angewendet¹⁰:

$$N_{(HMG,KK,RJ)} \times (1 + Vrate_{(HMG,GKV)}) = N_{adj(HMG,KK,RJ)}. \quad (9)$$

$N_{adj(HMG,KK,RJ)}$ = Anzahl der Versicherten einer HMG einer Krankenkasse im Referenzjahr nach AGG-Anpassung und GKV-Veränderungsrate

Hinweis:

Die Berechnung der GKV-HMG-Veränderungsrate erfolgt auf dieselbe Weise wie für die Einzelkasse; d. h., die veränderte Versichertenstruktur zwischen den Berichtsjahren wird berücksichtigt, um ausschließlich den Teil der HMG-Steigerungsrate zu berechnen, der nicht auf Veränderungen bei Alter und Geschlecht zurückzuführen ist.

¹⁰ Anderenfalls wird die Veränderungsrate der Krankenkasse angewendet. Bei der Anwendung der kassenindividuellen HMG-Veränderungsrate handelt es sich jedoch nur um einen formalen Berechnungsschritt, da in der anschließenden Betrachtung nur noch auf statistisch auffällige HMGs aus Schritt 1 und damit auf HMGs mit überdurchschnittlichen Steigerungsraten abgestellt wird.

Für jede HMG mit auffälligen Steigerungsraten wird die Differenz der kassenindividuellen HMG-Besetzungszahlen im Prüffjahr und der adjustierten Besetzungszahlen (nach Bereinigung um die kassenindividuelle Alter- und Geschlechtsstruktur sowie die GKV-Veränderungsrate) mit dem entsprechenden annualisierten Zuschlag des Ausgleichsjahres multipliziert und über alle auffälligen HMGs aufsummiert. Anschließend wird der Kassengröße Rechnung getragen, indem der berechnete Zuweisungsdifferenzbetrag durch die Gesamtversichertenzahl der Krankenkasse im Prüffjahr dividiert wird:

$$\frac{\sum[(N_{(HMG, KK, PJ)} - N_{adj(HMG, KK, RJ)}) \times Z_{(HMG)}]}{N_{(KK, PJ)}} = ZuwG_{(KK)}. \quad (10)$$

$N_{(HMG, KK, PJ)}$ = Anzahl der Versicherten einer HMG einer Krankenkasse im Prüffjahr

$N_{adj(HMG, KK, RJ)}$ = Anzahl der Versicherten einer HMG einer Krankenkasse im Referenzjahr nach kassenindividueller AGG-Anpassung und GKV-Veränderungsrate

$N_{(KK, PJ)}$ = Anzahl der Versicherten einer Krankenkasse im Prüffjahr

$Z_{(HMG)}$ = Annualisierter Zuschlag (Jahreswert) einer HMG im zugehörigen Ausgleichsjahr

$ZuwG_{(KK)}$ = Zuweisungsgewinn je Versicherten einer Krankenkasse im Prüffjahr.

Im Ergebnis resultiert der Zuweisungsgewinn je Versicherten einer Krankenkasse im Prüffjahr.

7 Schwellenwerte

Die im Rahmen der Auffälligkeitsprüfung berechneten Analysegrößen geben für sich alleine noch keine Auskunft darüber, wann eine Datenmeldung als auffällig eingestuft werden muss. Es bleiben unvermeidlich Grenzwerte bzw. Toleranzbereiche zu definieren, die den Normbereich der Untersuchung bilden.

Eine Krankenkasse gilt gemäß § 273 Absatz 2 SGB V als auffällig, wenn sie den statistischen Schwellenwert (vgl. Abschnitt 7.1) in einer oder mehreren HMGs übersteigt (sog. statistische Auffälligkeit).

Eine Krankenkasse gilt gemäß § 273 Absatz 3 SGB V als auffällig, wenn sie zusätzlich zu den Kriterien von 7.1 auch den monetären Schwellenwert (vgl. Abschnitt 7.2) überschreitet (sog. monetär erhebliche Auffälligkeit). Nur in diesem Fall leitet das BAS eine Prüfung nach § 273 Absatz 4 SGB V ein.

7.1 Schwellenwerte für die Regressionsanalyse

Für die Schwellenwerte der Regressionsanalyse gelten folgende Bedingungen für eine statistische Auffälligkeit (Schritt 1):

Grundsätzlich gilt als notwendige Bedingung für eine statistische Auffälligkeit, dass

$$1.) \quad RR(I_k)_{HMGx} > 1, \quad (2)$$

bei einem kritischen Wert für den korrespondierenden α -Fehler von 0,01, wenn also zusätzlich gilt, dass:

$$2.) \quad p(\beta I_k)_{HMGx} < 0,01. \quad (3)$$

Hinweis:

Dies gilt nur für alle Fälle in denen der RR-Wert und der p-Wert darstellbar sind. Nicht darstellbar ist der p-Wert dann, wenn im Referenzjahr eine HMG nicht belegt ist, im Prüfljahr aber eine HMG-Belegung vorhanden ist. Nicht darstellbare p-Werte werden zusammen mit den Ergebnissen der Regressionsanalyse in Anlage 2 (vgl. Abschnitt 8) ausgewiesen.

7.2 Schwellenwert für die Erheblichkeitsprüfung

Der nach Abschnitt 6.2 berechnete Zuweisungsgewinn je Versicherten wird mit einem vorab definierten Schwellenwert S(ZuwG) abgeglichen.

Ist

$$\mathbf{ZuwG}_{(KK)} > \mathbf{S}(\mathbf{ZuwG}) \quad (4)$$

liegt eine monetär erhebliche Auffälligkeit vor.

Der Schwellenwert für die Erheblichkeitsprüfung wird gemeinsam mit den pseudonymisierten Ergebnissen der Erheblichkeitsprüfung in Anlage 3 zu dieser Bestimmung bekannt gegeben (vgl. Abschnitt 8). Dem GKV-Spitzenverband wird ein Vorschlagsrecht für die Festlegung des Schwellenwerts eingeräumt.

Das Vorschlagsrecht nimmt der GKV-Spitzenverband im Rahmen des Anhörungsverfahrens zur vorliegenden Bestimmung wahr. Die mit dem Anhörungsentwurf vorveröffentlichten Ergebnisse aus Anlage 3 (in pseudonymisierter Form) dienen hierbei als Entscheidungsgrundlage.

Macht der GKV-Spitzenverband von seinem Vorschlagsrecht Gebrauch, ist das BAS an die Übernahme des Vorschlags gebunden, um ein Einvernehmen über den Schwellenwert zu erzielen. Folgt das BAS dem Vorschlag des GKV-Spitzenverbandes nicht, ist allenfalls das Benehmen hergestellt, sofern das BAS triftige Gründe anführt, die ihn zu seiner Entscheidung bewogen haben. Als triftiger Grund gilt, dass das BAS aus verwaltungsökonomischen Gründen nicht mehr als 15% der geprüften Krankenkassen in eine Einzelfallprüfung überführen kann. Sollte der GKV-SV demnach einen Schwellenwert anvisieren, der so sensibel ist, dass mehr als 15% der geprüften Krankenkassen als auffällig gelten, erhöht das BAS den Schwellenwert entsprechend. Ist der vorgeschlagene Schwellenwert des GKV-SV so gesetzt, dass weniger als 15% der geprüften Krankenkassen als auffällig gelten, übernimmt das BAS den vorgeschlagenen Schwellenwert des GKV-SV.

Macht der GKV-SV von seinem Vorschlagsrecht keinen Gebrauch, legt das BAS den Schwellenwert nach oben genannter Methodik fest, der dann als einvernehmlich festgelegt gilt.

8 Mitteilungen

8.1 Dokumentation für den GKV-Spitzenverband

Anliegend zur Bestimmung erhält der GKV-Spitzenverband vom BAS die folgenden Dokumente, die auch zur Weiterleitung an die Krankenkassen vorgesehen sind:

Anlage 1 enthält die Spezifikationen für das jeweils zugrunde gelegte Klassifikationsmodell (vgl. Abschnitt 2.3). Grundlegend für die Dokumentation des Verfahrens sind die Anlagen 2 und 3, die für jede zu prüfende Datenmeldung neu erstellt werden. Anlage 2 beinhaltet in pseudonymisierter Form sämtliche Ergebnisse der Regressionsanalyse unabhängig davon, ob sie den statistischen Schwellenwert übersteigen. Anlage 2 enthält zusätzlich zu den Regressionsergebnissen die HMG-Belegungen für Prüf- und Referenzmeldung für die Gesamt-GKV sowie die in Abschnitt 10.4 erwähnten Gütekriterien für sämtliche HMG-bezogenen Regressionsrechnungen.

Anlage 3 enthält in pseudonymisierter Form die Ergebnisse der Erheblichkeitsprüfung und weist den vorab definierten Schwellenwert als festen Eurobetrag aus. In der Entwurfsfassung der Bestimmung, welche Gegenstand der Anhörung ist, wird der Schwellenwert noch nicht ausgewiesen, um dem diesbezüglichen Vorschlagsrecht des GKV-Spitzenverbandes Rechnung zu tragen (vgl. Abschnitt 7.2). Die Ergebnisdarstellung (HMG-Zuweisungsgewinn in Euro je Versicherten einer Krankenkasse in absteigender Sortierung) beschränkt sich auf diejenigen Krankenkassen, die eine statistische Auffälligkeit aufweisen. Anlage 3 enthält darüber hinaus eine Tabelle mit den alters- und geschlechtsadjustierten GKV-Veränderungsraten derjenigen HMGs, für die in der Regression statistische Auffälligkeiten festgestellt und die folglich als Berechnungsgrundlage für die Erheblichkeitsprüfung genutzt worden sind.

Die abschließende Bestimmung nebst zugehöriger Anlagen wird dem GKV-Spitzenverband gemeinsam mit einem Anschreiben übermittelt, in dem das BAS die Aufnahme neuer Prüfhandlungen ankündigt.

Zusätzlich stellt das BAS dem GKV-Spitzenverband für jedes Berichtsjahr im Nachgang zur Einleitung der Prüfhandlungen einen fiktiven Testdatensatz zur Verfügung, der es den Krankenkassen ermöglichen soll, das Verfahren auch ohne Kenntnis der Gesamt-GKV-Daten rechnerisch nachzuvollziehen.

Nach Abschluss der Auffälligkeitsprüfung erhält der GKV-Spitzenverband zudem einen abschließenden Prüfbericht mit einer anonymisierten Zusammenstellung der Ergebnisse der Auffälligkeitsprüfung.

8.2 Bekanntgabe der Ergebnisse

Die Bekanntgabe der Prüfergebnisse erfolgt kassenindividuell:

Jede Krankenkasse wird schriftlich darüber informiert, ob eine statistische Auffälligkeit vorliegt und ob es sich dabei um eine monetär erhebliche Auffälligkeit handelt. In diesem Zuge wird jeder Krankenkasse das in Anlage 2 und ggf. in Anlage 3 für sie verwendete Pseudonym mitgeteilt. Auf diese Weise hat jede Krankenkasse Zugang zu den für sie berechneten statistischen Kennzahlen (RR, p-Wert) je HMG. Krankenkassen mit einer monetär erheblichen Auffälligkeit haben entsprechend Zugriff auf den für sie jeweils ermittelten Zuweisungsgewinn je Versicherten aus Anlage 3.

Ausschließlich den von einer monetär erheblichen Auffälligkeit betroffenen Krankenkassen werden zudem die für sie jeweils in der Gruppierung ermittelten HMG-Besetzungszahlen für Prüf- und Referenzmeldung mitgeteilt, auf denen die Prüfergebnisse aufbauen. Die Krankenkassen erhalten die genannten Informationen über den beim BAS eingerichteten FTP-Server.

9 Prüfkonzsequenzen, gesonderte Bestimmungen

Die Mitteilung nach Abschnitt 8.2 über eine festgestellte erhebliche Auffälligkeit nach Abschnitt 8.2 dieser Bestimmung leitet eine Einzelfallprüfung gemäß § 273 Absatz 4 SGB V ein. Den betroffenen Krankenkassen wird Näheres hierzu schriftlich bekannt gegeben.

Bonn, den 19.02.2021

10 Anhang: Ausführliche Darstellung des verwendeten Regressionsmodells

Im Folgenden wird die Ausgestaltung des unter Abschnitt 5 aufgeführten Regressionsmodells schrittweise erläutert. Ziel ist es, den Modellaufbau nachvollziehbar und transparent zu machen. Es geht hierbei ausdrücklich nicht um eine formal-mathematische Darstellung, da der Erklärungsgehalt komplizierter Formeln für gering gehalten wird und entsprechende Ausführungen in statistischen Standardwerken und wissenschaftlichen Artikeln in großer Zahl zu finden sind.

10.1 Generalized Linear Models

Im Rahmen der Analyse wird als Zielvariable das Vorliegen bzw. Nicht-Vorliegen einer HMG je Versicherten modelliert. Da im Rahmen von § 273 SGB V die Betrachtung auf Kassenebene ausschlaggebend ist, werden die Ereignisse je Krankenkasse gezählt: Es geht folglich um die Auswertung einer Zählvariablen. Bei einer solchen Analyse, die das Vorkommen eines Ereignisses gemessen an einer bestimmten Population untersucht, sind einfache (multiple) Regressionsverfahren nicht mehr anwendbar¹¹. Stattdessen kommen in solchen Fällen Modelle vom Typ der „Generalized Linear Models“ zum Einsatz, sogenannte GLM-Modelle.

GLM-Modelle sind eine Verallgemeinerung der einfachen linearen Modelle. Von diesen ausgehend umfasst die Verallgemeinerung drei Komponenten¹²:

- 1.) Die Verteilung der Zielvariablen Y_i gehört einer einfachen Exponentialfamilie an.
- 2.) Die erklärenden Variablen gehen als Linearkombination ins Modell ein:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir}, \quad i = 1, \dots, n.$$

- 3.) Der Erwartungswert von Y_i ist durch die Linkfunktion g mit η verknüpft:

$$g(E(Y_i)) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir}, \quad i = 1, \dots, n.$$

In erster Linie lockern GLM-Modelle damit im Vergleich zu einfachen linearen Modellen die Annahmen hinsichtlich der Verteilung der Zielvariablen. So gehören zur exponentiellen Familie

¹¹ Bei einer Zählvariablen sind viele Voraussetzungen, die an die Anwendung einer einfachen linearen Regression gekoppelt sind, nicht mehr erfüllt: So kann bei einer Zielvariable, die nur ganzzahlige Werte ≥ 0 annehmen darf, nicht grundsätzlich von einer Normalverteilung ausgegangen werden. Auch ist die Varianz der Werte, die die Zielvariable Y annehmen kann, nicht konstant, sondern i.d.R. ansteigend mit dem Mittelwert.

¹² Die Darstellung ist entlehnt aus: Müller, Marianne: „Angewandte statistische Regression“, Zürcher Hochschule Winterthur 2006, online abrufbar unter: http://stat.ethz.ch/education/semesters/WS_2006_07/asr/Stat.pdf.

eine ganze Reihe von Verteilungen, bspw. die Binomial-, Poisson- oder Gamma-Verteilung. Die Flexibilisierung erfolgt in einem einheitlichen methodischen Rahmen. Eine wesentliche Komponente dabei ist die Link-Funktion: Da GLM-Modelle – genauso wie einfache lineare Regressionen – linear in ihren Koeffizienten sind, bedarf es der Link-Funktion, um den Erwartungswert $E(Y_i)$ in Werte des linearen Prädiktors zu verwandeln.

10.2 Poisson-Verteilung

Für den hier vorliegenden Fall wird davon ausgegangen, dass die Zielvariable Y_i (Häufigkeit des Vorliegens einer HMG) Poisson-verteilt ist. Die Poisson-Verteilung erlaubt eine Annahme darüber, wie oft voneinander unabhängige Ereignisse – bei einer gegebenen Eintrittswahrscheinlichkeit p – in einer bestimmten Zeit oder in einem bestimmten Volumen auftreten. Dabei lässt sich die gesuchte Anzahl direkt vom erwarteten Mittelwert μ der Ereignisse ableiten. Der Parameter μ charakterisiert folglich die Poisson-Verteilung. Er ist gleich der mittleren Anzahl der gezählten Ereignisse, die erwartet wird, wenn die Zählung n Male wiederholt wird. Wenn die Zielvariable Y_i also Poisson-verteilt ist mit dem Parameter μ , dann ist ihr Erwartungswert $E(Y_i) = \mu_i$, mit der Restriktion $\mu_i > 0$.

Die Poisson-Verteilung ist also für diskrete Werte größer gleich null definiert, weswegen sie sich für die Analyse von Zählgrößen anbietet. Dies gilt insbesondere, wenn es sich um seltene Ereignisse mit einer geringen Eintrittswahrscheinlichkeit p handelt und die Population, auf die sich die Wahrscheinlichkeit bezieht, sehr groß ist (mit vierstelligem n oder größer)¹³. Folglich spricht man bei der Poisson-Verteilung auch häufig von der „Verteilung der seltenen Ereignisse“. Gerade bei medizinischen Inzidenzraten, die hier betrachtet werden, lässt sie i. d. R. eine gute Annäherung an die realen Daten zu.

Eine Besonderheit der Poisson-Verteilung ist die Eigenschaft, dass nicht nur der Erwartungswert, sondern auch die Varianz gleich dem Mittelwert μ ist. Eine weitere Eigenschaft besteht darin, dass die Summe von unabhängigen Poisson-verteilten Ereignissen ebenfalls Poisson-verteilt ist. Folglich kann man Individualdaten und gruppierte Daten gleichermaßen mit einer Poisson-Regression untersuchen.

¹³ Die Poisson-Verteilung kann auch als Grenzfall der Binomialverteilung aufgefasst werden, nämlich für $n \rightarrow \infty$. Dann konvergiert die Dichte der Binomialverteilung gegen die der Poisson-Verteilung. Dies trifft auf die zugrundeliegenden Daten praktisch zu. Entsprechend haben empirische Tests gezeigt, dass Binomial- und Poisson-Regression im hier vorgefundenen Kontext (nahezu) identische Ergebnisse produzieren.

10.3 Linearer Prädiktor und Link-Funktion

Sobald die Verteilung von Y_i festgelegt ist, kann die eigentliche Regressionsgleichung aufgestellt werden. Als Ausgangspunkt besitzt auch die Poisson-Regression den – unter 10.1 vorgestellten – „Linearen Prädiktor“:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir}, i = 1, \dots, n.$$

In der Poisson-Regression wird als Link-Funktion – wegen der Restriktion $\mu_i > 0$ – der Logarithmus des Erwartungswertes verwendet, denn er macht aus den positiven Erwartungswerten transformierte Werte ohne Begrenzung. Es gilt also:

$$g(E(Y_i)) = \eta = \log(E(Y_i)),$$

bzw.:

$$\log(E(Y_i)) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir}, i = 1, \dots, n. \quad (\mathbf{A.1})$$

Die sog. **log-lineare Poisson-Regression** kombiniert somit die logarithmische Link-Funktion mit der Annahme der Poisson-Verteilung für die Zielgröße.

10.4 Hauptmodell

Um die spezifische Regressionsgleichung aufstellen zu können, bedarf es ausgehend von Gleichung (A.1) noch dreier weiterer Schritte¹⁴:

- 1) Bestimmung der unabhängigen Variablen bzw. Kovariablen.
- 2) Bestimmung des sog. Offsets.
- 3) Bestimmung des Interaktionsterms aus den Kovariablen „Meldung“ und „Krankenkasse“.

¹⁴ Es folgt eine ausführliche Schilderung der unter 5.3 angegebenen Bestandteile der verwendeten Regressionsgleichung.

Zu 1) Bestimmung der unabhängigen Variablen

Um den Einfluss auf die Zielvariable HMG-Häufigkeit zu messen, stehen drei Merkmale zur Verfügung: Im Zentrum der Analyse stehen die Einzelkassen, die HMG-Besetzungszahlen werden also nach den in der Prüfmeldung vorhandenen Krankenkassen differenziert. Bei einer Anzahl von m Krankenkassen geht hierbei die GKV als $m + 1$ te Krankenkasse mit in die Betrachtung ein und bildet die relevante Referenzgruppe, mit der verglichen wird. Weiter gehen die Merkmale Alter und Geschlecht in Form der 40 AGGs in die Regression mit ein. Das letzte und entscheidende Merkmal, das als unabhängige Variable ins Modell einfließt, ist die Meldung (Referenz- vs. Prüfmeldung).

Da es sich bei allen Merkmalen um kategoriale Variablen handelt, gehen sie als 0/1-kodierte Dummy-Variablen in das Modell ein. Programmtechnisch erfolgt diese Umwandlung i.d.R. automatisch. Für das Merkmal „Meldung“, das ohnehin als dichotome Variable vorliegt, erübrigt sich dieser Schritt.

Somit wird der lineare Prädiktor gebildet aus:

- der unabhängigen Variablen AGG „ A “, mit 40 Ausprägungen und 40-1 Dummy-Variablen,
- der unabhängigen Variablen Krankenkasse „ K “, mit $m+1$ Ausprägungen und m Dummy-Variablen, und
- der unabhängigen Variable Meldung „ M “ mit 2 Ausprägungen und einer Dummy-Variablen.

Zu 2) Der Offset

Die Bestimmung der unabhängigen Variablen erleichtert die Handhabung des Regressionsdatensatzes: Die Gruppierungen, die für Prüf- und Referenzmeldung getrennt und auf Individualdaten-Basis vorliegen, werden zusammengespielt und nach diesen Merkmalen gruppiert. Angesichts der $m+1$ Krankenkassen, der 40 AGGs und der zwei einbezogenen Meldungen bedeutet dies eine Verschlinkung des Regressionsdatensatzes auf $(m+1) \times 40 \times 2$ Beobachtungen, ohne dass ein Informationsverlust entsteht.

Darüber ergeben sich jedoch HMG-Raten, denn die Anzahl der Ereignisse pro Krankenkasse, Altersgruppe und Meldung Y_i lässt sich nur im Verhältnis zur

Gesamtversichertenanzahl G_i in der jeweiligen Gruppe sinnvoll interpretieren. Da dieses Verhältnis – also die erwartete Anzahl Erkrankter zur Gesamtgruppe – exakt proportional ist, wäre es daher auch als direkte Zielgröße geeignet. Dies ist jedoch ausgeschlossen, da Y_i/G_i nicht mehr Poisson-verteilt ist. Deshalb muss die Regression den jeweiligen Populationsbezug durch ein spezielles Argument vorsehen. Dies wird – programmtechnisch – als offset bezeichnet. Hierzu gehen die Werte für G_i logarithmiert in den eigentlichen Regressionsterm mit ein. Der lineare Prädiktor erhält also das zusätzliche Argument:

$$\beta_G \log(G_i),$$

mit dem Koeffizienten $\beta_G = 1$.

Zu 3) Interaktionsterm

Um die eigentliche Analysegröße für das Erkennen einer Auffälligkeit aus der Regression gewinnen zu können, muss dem linearen Prädiktor noch ein weiteres Element hinzugefügt werden: Ein Interaktionsterm der beiden Kovariablen Krankenkasse und Meldung, der die Wechselwirkungen zwischen diesen beiden Einflussgrößen darzustellen vermag. Der relevante Interaktionsterm I lässt sich durch eine multiplikative Verknüpfung der beiden betreffenden Variablen ins Modell integrieren, also $I = M * K$.

Dabei gehen die restlichen Kovariablen (A, M, K) unverändert ins Modell mit ein. Inhaltlich stellt der Einbezug der AGGs sicher, dass um die möglicherweise verzerrenden Alters- und Geschlechtseinflüsse kontrolliert wird. Durch den Einbezug des Parameters „Krankenkasse“ als erklärende Variable wird hingegen um die verschiedenen Morbiditätsniveaus der Krankenkassen kontrolliert. Die Variable Meldung kontrolliert schließlich um den Morbiditäts-Basiseffekt, also um ein ansteigendes HMG-Grundniveau von der Referenz- zur Prüfmeldung.

Damit ergibt sich für das Hauptmodell die Regressionsgleichung:

$$\log(E(Y_i)) = \beta_G \log(G_i) + A'_i + K'_i + M'_i + I'_i, \quad i = 1, \dots, n, \quad (\mathbf{A.2})$$

wobei $E(Y_i)$ für den Erwartungswert der Zielvariablen,

$\beta_G \log(G_i)$ für den Offset,

A'_i verkürzt für $(A_1)_i \beta_{A_1} + \dots + (A_{39})_i \beta_{A_{39}}$,

K'_i verkürzt für $(K_1)_i\beta_{K_1} + \dots + (K_m)_i\beta_{K_m}$,

M'_i verkürzt für $M_i\beta_M$ und

I'_i verkürzt für $(M * K_1)_i\beta_{I_1} + \dots + (M * K_m)_i\beta_{I_m}$ steht.

Die Berechnungen werden vom BAS mit der Statistik-Software SAS durchgeführt. Für das hier spezifizierte Modell gilt wie für alle GLM-Regressionen, dass sich die Regressionskoeffizienten („ β -Parameter“) nicht mehr mittels einer Kleinst-Quadrat-Methode exakt bestimmen lassen. Stattdessen müssen sie mit der Maximum-Likelihood-Methode in einem iterativen Prozess geschätzt werden¹⁵. Auch alle Tests und Vertrauensintervalle sind damit approximativ. Ähnlich wie auch in der Linearen Regression lässt sich die Güte der Schätzung über eine Residualanalyse bestimmen. In den gängigen Statistik-Programmen werden zu diesem Zweck i.d.R. die Pearson-Residuen und die Devianz-Residuen ausgegeben. Da bei dem hier vorliegenden Untersuchungsgegenstand nicht die Vorhersagegüte des Modells im Vordergrund steht, sondern in erster Linie der Effekt bestimmter Einflussgrößen auf die Zielvariable ermittelt werden soll, wird auf eingehende Beschreibungen zu den von SAS ausgegebenen Gütekriterien an dieser Stelle verzichtet. Zu Dokumentationszwecken ist jedoch der Quotient aus Devianz und Anzahl der Freiheitsgrade als dasjenige Kriterium, das am häufigsten zur Überprüfung der Modellgüte herangezogen wird, in der Anlage 1 je Satzart für jede HMG-bezogene Regression ausgegeben. Es sei allerdings darauf verwiesen, dass die Aussagekraft dieser Werte als gering eingeschätzt werden kann, da sie nur im Vergleich verschiedener Modellvarianten sinnvoll interpretierbar sind, nicht jedoch im Vergleich unterschiedlicher Regressionen (mit unterschiedlichen zugrundeliegenden Daten), wie im hier betrachteten Fall.

10.5 Interpretation der Regressionskoeffizienten

Um nun zu einer Interpretation der Regressionskoeffizienten zu gelangen, bietet es sich an, die vorangegangene Gleichung in einer Schreibweise wiederzugeben, in der nach $E(Y_i)$ aufgelöst bzw. beide Seiten exponiert sind, also:

¹⁵ Bei einer Maximum-Likelihood-Methode ist nicht grundsätzlich garantiert, dass der zugrundeliegende Algorithmus konvergiert. Der Rechenprozess im hier verwendeten Modell umfasst unabhängig davon maximal 50 Iterationen. Dies entspricht der Default-Einstellung bei SAS.

$$E(Y_i) = \exp\{\beta_{A_1}\}^{A_1} \cdot \dots \cdot \exp\{\beta_{I_m}\}^{I_m}. \quad (\text{A.3})$$

Es wird deutlich, dass die additiven Terme im Bereich des linearen Prädiktors durch Exponieren in multiplikative Effekte umgewandelt werden. D. h. ändert man z. B. A_1 um eine Einheit, bewirkt dies eine Multiplikation des Erwartungswertes mit dem Faktor $\exp\{\beta_{A_1}\}$. Generell bedeutet $\exp\{\beta_j\} = 1$, dass kein Einfluss zwischen der unabhängigen und der abhängigen Variablen besteht, bei $\exp\{\beta_j\} > 1$ ist er positiv, bei $\exp\{\beta_j\} < 1$ ist er negativ. Dabei lässt sich $\exp\{\beta_j\}$ auch als das sog. Relative Risiko RR interpretieren. Das RR ist ein Wahrscheinlichkeitsverhältnis, das die Wahrscheinlichkeit für den Eintritt eines Ereignisses im Unterschied zwischen zwei Gruppen angibt. Würde bspw. die HMG-Anzahl nur anhand des Alters erklärt und hierfür zwei Gruppen gebildet mit der Grenze ≥ 50 Jahre, dann würde ein $\exp\{\beta_A\} = RR(A) = 1,6$ bedeuten, dass die HMG-Anzahl bei über 50-Jährigen um einen Faktor von 1,6 höher ist als bei unter 50-Jährigen.

Als Analysegröße ist im Rahmen des mit Gleichung (A.2) abgebildeten Modells einzig der Einfluss des Interaktionsterms I , also die $RR(I)_s$, von Relevanz. Sie geben an, wie sich der Einfluss auf die HMG-Häufigkeit, der der Meldung (bzw. dem Übergang von Referenz- zur Prüfmeldung) zuzuschreiben ist, in Abhängigkeit unterschiedlicher Krankenkassen ändert. Ist diese Wechselbeziehung stark, das zugehörige RR bei einer Krankenkasse also (erheblich) > 1 , ist der Einfluss auf die HMG-Entwicklung überproportional. Bei der entsprechenden Krankenkasse führt der Übergang von der Referenz- zur Prüfmeldung also zu einem Anstieg der Versichertenzahlen in der betrachteten HMG, der (weit) oberhalb der GKV-Entwicklung liegt (für die GKV gilt stets $RR(I) = 1$). Die entsprechende p-Statistik liefert für jeden βI -Parameter das zugehörige Signifikanzniveau¹⁶. Beide Kennzahlen bilden gemeinsam das Maß für den Grad einer Auffälligkeit.

¹⁶ Es handelt sich um eine SAS-generierte Wald- χ^2 -Statistik unter der Annahme, dass die Maximum-Likelihood-Schätzer asymptotisch normalverteilt sind. Dies gilt gewöhnlich für große Stichprobenumfänge.