Stichprobenkonzept für Datenerhebungen nach § 30 RSAV

Gutachten im Auftrag des Bundesversicherungsamtes

Prof. Dr. Thomas Schäfer

Bocholt, im März 2007

INHALTSVERZEICHNIS

Zusammenfassung

1. Für die Planung der Stichprobe von Versicherten der gesetzlichen Krankenversicherung mit versichertenbezogenen Angaben zu den Diagnosen, verordneten Arzneimitteln und den berücksichtigungsfähigen Leistungsausgaben nach § 30 RSAV muss unterschieden werden, ob die Verwendung dieser Daten der Auswahl der Krankheiten sowie der Entwicklung und Kalibrierung des prospektiven Regressionsmodells dient, das zur Schätzung der standardisierten Leistungsausgaben benötigt wird, oder ob das kalibrierte Modell und dessen Kostengewichte im Verwaltungsvollzug des Risikostrukturausgleichs angewendet werden sollen. Die beiden Zwecke gehen mit unterschiedlichen Anforderungen an den Stichprobenumfang einher.

2. Für beide Verwendungszwecke der Stichprobe wird das Design einer einfachen (ungeschichteten) Zufallsstichprobe empfohlen, da die Schichtung nicht zum Konzept des herkömmlichen Regressionsmodells passt und darüber hinaus die aus dem Zellenansatz begründete Notwendigkeit für die Schichtung im Regressionsmodell weitgehend entfällt. So kann die Gruppe der Versicherten mit Bezug einer Erwerbsminderungsrente beispielsweise im Regressionsmodell wie eine morbiditätsbezogen gebildete Versichertengruppe behandelt werden und hat dann keine kleinere Besetzungszahl als Versichertengruppen mit Bezug zu nicht allzu häufigen Krankheiten.

3. Aus Gründen der Manipulationsresistenz und Praktikabilität der dezentral an vielen Stellen vorzunehmenden Ziehung gibt es nach Ansicht des Gutachters zu einer Geburtstagsstichprobe als Ersatz für eine Ziehung nach Maßgabe eines echten Zufallsexperimentes gegenwärtig keine Alternative. Allerdings müssen die für eine Geburtstagsstichprobe verwendeten Tage des Jahres als einfache Zufallsstichprobe ohne Zurücklegen aus der Gesamtheit der Tage des Jahres gezogen und dürfen nicht willkürlich festgelegt werden. Darüber hinaus empfiehlt es sich, die dem Verfahren der Geburtstagsstichprobe zugrundeliegende Hypothese der Gleichverteilung der Geburtstage über die Tage des Jahres mit aktuellen Daten der gesetzlichen Krankenversicherung zu überprüfen. Dabei sollte insbesondere die Größenordnung des in der Literatur dokumentierten Phänomens der „Digital Preference" abgeschätzt werden. Hierbei geht es um die Beobachtung, dass bei im Ausland geborenen Personen mit Migrationshintergrund Geburtstage, die auf einen Tag fallen, dessen Stellung im Monat ein Vielfaches von fünf beträgt, mit signifikant größerer Häufigkeit auftreten, als aus der Gleichverteilungsannahme zu erwarten ist.

4. Eine Beschränkung auf ganzjährig Versicherte schränkt die Grundgesamtheit unnötig ein. Vorgeschlagen wird, die Grundgesamtheit der Erhebung aus Personen zu konstituieren, die in jedem der beiden Jahre 2005 und 2006 mindestens eine zusammenhängende Zeitspanne von 30 Tagen in der gesetzlichen Krankenversiche-

rung versichert waren. Die berücksichtigungsfähigen Leistungsausgaben von im Jahr 2006 nicht ganzjährig versicherten Personen (zu denen ja auch alle im Jahr 2006 verstorbenen gehören) müssen auf das Jahr hochgerechnet werden. Die Anteile der Jahre in denen die Versicherung bestand, können ggf. als zusätzliche Prädiktoren ins Regressionsmodell aufgenommen werden.

5. Die dem Hauptleistungsbereich „Sonstige Leistungen" zugeordneten Ausgaben sollten in den Kassen, die in diesem Bereich keine Totalerfassung betreiben, für Stichprobenversicherte aus den Originalbelegen erfasst werden.

6. Zur Berechnung des Stichprobenumfangs für Zwecke der Modellentwicklung und Kalibrierung werden die Krankheitsprävalenzen von Versicherten der gesetzlichen Krankenversicherung sowie Mittelwerte und Standardabweichungen der berücksichtigungsfähigen Leistungsausgaben in den krankheitsbezogen gebildeten Versicherungsgruppen benötigt. Diese wurden dem Gutachter vom Bundesversicherungsamt aus einer Sonderauswertung des (unbereinigten) Datensatzes nach §268 SGB V für 2001/02 dankenswerter Weise zur Verfügung gestellt. Hieraus lässt sich als Stichprobenmindestgröße für die Nettostichprobe ein erforderliches n in Höhe von ca. 3,25 Millionen Versicherte ermitteln, was unter Berücksichtigung der zu erwartenden Fehlerquote in den Datensätzen einer Bruttostichprobe von n=4,9 Millionen Versicherten entspricht, die im Rahmen einer Geburtstagsstichprobe von 25 zufällig gewählten Tagen im Jahr gezogen werden können (Auswahlquote = 6,85%). Wegen einiger Unplausibilitäten des dem Bundesversicherungsamt vorliegenden unbereinigten Datensatzes steht der empfohlene Mindeststichprobenumfang unter dem Vorbehalt einer Überprüfung der genannten Inputdaten auf der Basis des von dem Gutachterteam IGES, Lauterbach und Wasem bereinigten Datensatzes nach §268 SGB V für 2001/02.

7. In die Berechnung des Mindeststichprobenumfangs einer Stichprobe, die in der Routine die Basis des morbiditätsbezogenen Risikostrukturausgleichs darstellen kann, fließen neben den schon unter Punkt 6 erwähnten Parametern zusätzlich die Korrelationen der Prädiktoren des Modells mit ein, über die bisher keine Erkenntnisse veröffentlicht wurden. In diesem Punkt ist der Gutachter auf eine Sonderauswertung des bereinigten Datensatzes nach §268 SGB V für 2001/02 angewiesen. Allerdings ist zu erwarten, dass für die Routine eine kleinere Stichprobe als in der Phase der Modellierung benötigt wird.

8. Die Stichprobensegmente der einzelnen Krankenkassen sollten auch im morbiditätsbezogenen Risikostrukturausgleich durch Analyse der Ausschöpfungsquoten auf ihre Qualität hin überprüft und Kassen mit zu niedrigen Ausschöpfungsquoten ggf. ausgeschlossen werden, damit die kontrollierten Zufallsfehler nicht durch unkontrollierte systematische Fehler überlagert werden.

1 Gutachtenauftrag

Nach der Beschreibung des Gutachtenauftrags durch das Bundesversicherungsamt ist es Ziel der Expertise, ein Konzept zu entwickeln, das Angaben zum Umfang, zu Art und zur Auswahl einer geeigneten Stichprobe umfasst, welche die Spitzenverbände, § 30 (1) Satz 3 RSAV entsprechend, mit dem Bundesversicherungsamt vereinbaren wollen.

Es werden drei Zweckbestimmungen der Stichprobe bzw. des Stichprobenkonzeptes unterschieden:

1. Auswahl der vom GKV-WSG vorgegeben 50 bis 80 Krankheiten, bei denen die durchschnittlichen Leistungsausgaben je Versicherten die durchschnittlichen Leistungsausgaben aller Versicherten um mindesten 50 vom Hundert überschreiten;
2. Entwicklung eines für den Risikostrukturausgleich in der GKV geeigneten Klassifikationsmodells auf der Basis der ausgewählten Krankheiten und ihrer Identifikationsmerkmale;
3. Anwendung im Rahmen der routinemäßigen Durchführung des Risikostrukturausgleichs ab dem Jahresausgleich 2009.

Insbesondere soll zu den zu den folgenden Fragen Stellung genommen werden::

- Welcher Stichprobenumfang ist erforderlich, damit epidemiologisch bedeutsame Krankheiten ausgewählt, das Klassifikationsmodell entwickelt und der morbiditätsorientierte Risikostrukturausgleich auf der Basis der gewählten Krankheiten hinreichend zielgenau durchgeführt werden kann?
- Wie kann dem Problem begegnet werden, dass im Hauptleistungsbereich „Sonstige Leistungsausgaben“ keine flächendeckende Vollerhebung als Basis für eine Stichprobenauswahl vorhanden ist?
- Welche Auswahlkriterien sollen verwendet werden?

2 Berechnung der standardisierten Leistungsausgaben

Die für den morbiditätsbezogene Risikostrukturausgleich benötigten standardisierten Leistungsausgaben sollen nach dem Willen des Gesetzgebers unter Zugrundelegung eines prospektiven Regressionsmodells geschätzt werden. Diesem liegt in seiner Hauptvariante (einem Eingleichungsmodell) folgende Modellgleichung für die berücksichtigungsfähigen Leistungsausgaben y im Ausgleichsjahr zugrunde:

(1) $y_i = ß_0 + ß_1 S_{1i} + ... + ß_k S_{ki} + ß_{k+1} K_{1i} + ...ß_{k+m} K_{m,i} + e_i$. [1]

wobei die S_{ji} (j=1,...k) nur der Werte 0 und 1 fähige Indikatorvariablen ("dummy variables") darstellen, welche die Zugehörigkeit des Versicherten i zu einer von k soziodemographischen Versichertengruppen im Vorjahr anzeigen, die durch Kreuzklassifikation von Alter mit Geschlecht und dem Erwerbsminderungsstatus entstehen, und die K_{ji} (j=1,...m) Indikatorvariablen sind (ebenfalls nur der Werte 0 und 1 fähig), welche das Vorliegen von Krankheiten im Vorjahr anzeigen, die in m Krankheitsgruppen zusammen gefasst (gruppiert) worden sind. Die Indikatorvariablen werden nach ihrer Zweckbestimmung auch Prädiktorvariablen genannt. Der letzte Term der Gleichung (1) – e_i – ist eine Fehlerterm: Er misst die zufällige, unsystematische Abweichung der berücksichtigungsfähigen Leistungsausgaben des i-ten Versicherten von ihren Erwartungswerten, den standardisierten Leistungsausgaben . Zur Berücksichtigung von Wechselwirkungen zwischen den Prädiktoren können Produktterme in das Modell aufgenommen werden

Liegt eine Zufallsstichprobe von n Versicherten vor („Eichstichprobe"), so werden die Koeffizienten im Allgemeinen nach der Methode der kleinsten Quadrate geschätzt (Ordinary Least Square- bzw. OLS-Regression). Die Schätzungen $\hat{ß}_j$ der $ß_j$ werden im Rahmen des Risikostrukturausgleichs als Regressions- oder Kostengewichte bezeichnet (für j ≥ k+1 auch als morbiditätsbezogene Risikozuschläge).

[1] Ein Teil der Varianz der Ausgaben auf der individuellen Ebene der Versicherten ist darauf zurückzuführen, dass ein bedeutender Anteil der Versicherten überhaupt keine Leistungen in Anspruch nimmt. Ein vielversprechender Ansatz der Varianzreduktion, der z.B. auch von der Washington Health Care Authority verfolgt worden ist (vgl. Dunn 1998), besteht vor diesem Hintergrund darin, den Erwartungswert der Ausgaben X in der Form E(X) = P(X>0)·E(X|X>0) zu faktorisieren, wobei P(X>0) die Wahrscheinlichkeit für den *Schadenseintritt* und E(X|X>0) den (bedingten) Erwartungswert für die *Schadenshöhe* – unter der Bedingung, dass ein Schaden eingetreten ist – bezeichnet. Das Prädiktionsmodell zerfällt dann in zwei Regressionsgleichungen, von denen die erste einer Prädiktion der Wahrscheinlichkeit für das Vorliegen von Ausgaben dient (z.B. aus einer logistischen Regression) und das zweite einer Vorhersage der Ausgaben für diejenigen Versicherten, die von Null verschiedene Ausgaben aufweisen (z.B. aus einer Regression im verallgemeinerten linearen Modell für die Inverse Gaussverteilung, vgl. hierzu Schäfer und Seidel 2002).

Bei der Modellentwicklung werden fünf verschiedene Stufen durchlaufen:

1. Auswahl der zu berücksichtigenden Krankheiten und Bildung von Krankheitsgruppen auf der Basis der verfügbaren krankheitsbezogenen Merkmale, einschließlich der softwaretechnischen Umsetzung in einem sog. Grouper;
2. Festlegung der soziodemographischen Versichertengruppen;
3. Auswahl der Prädiktorvariablen, die in das Modell aufgenommen werden;
4. Prüfung von Wechselwirkungen und ggf. Aufnahme in die Modellgleichung;
5. Schätzung der Regregressionsgewichte (Modellkalibrierung) und der Vorhersagegüte.

Der Übergang vom Zellen- zum Regressionsansatz im Risikostrukturausgleich stellt (in Deutschland) einen Paradigmenwechsel dar, an den sich die mit dem Verwaltungsvollzug befassten Personen erst noch gewöhnen müssen. Insbesondere ist an dieser Stelle schon darauf hinzuweisen, dass die Anforderungen an die Größe des Stichprobenumfangs im Regressionsansatz geringer sind als im Zellenansatz, weil in die Konfidenzintervalle der standardisierten Leistungsausgaben bei letzterem nur die Zahl der der jeweiligen Zelle zugeordneten Versicherten eingehen, bei ersterem jedoch die Zahl aller Versicherten.

3 Bestimmung des Stichprobenumfangs

Je nach Art der Problemstellung (Parameterschätzung oder Prüfung einer statistischen Hypothese) sind zwei verschiedene Verfahren für die Ermittlung des erforderlichen Stichprobenumfangs verbreitet: Und zwar wird der Stichprobenmindestumfang so festgelegt, dass

- die erwartete halbe Breite des 95%-Konfidenzintervalls einen Wert nicht überschreitet, den man aus einer vorgegebenen Fehlertoleranz ableitet (Schätzproblem)
- die Trennschärfe (Power) eines Tests zum 5%-Niveau eine vorgegebene Schwelle, zumeist 80%, nicht unterschreitet (Testproblem).

Beide Verfahren können bei der Planung der Stichprobe des morbiditätsbezogenen Risikostrukturausgleichs Anwendung finden.

3.1 Präzisierung der Aufgabe hinsichtlich der drei Zweckbestimmungen

3.1.1 Auswahl der Krankheitsgruppen

Die Auswahl der Krankheitsgruppen lässt sich nicht algorithmisieren, sondern wird von Experten mit Computerunterstützung geleistet werden müssen. Neben morphologischen, klinischen, epidemiologischen, und pharmakologischen Kriterien und solchen der ICD spielen hier auch die berücksichtigungsfähigen Leistungsausgaben der betrachten Krankheiten in der Grundgesamtheit eine Rolle. Diese müssen sich aus der Stichprobe für diese Zweckbestimmung also hinreichend genau schätzen lassen. Um den hierfür erforderlichen Mindestumfang der Stichprobe zu ermitteln, empfiehlt sich eine zweistufige Vorgehensweise. Auf der ersten Stufe ermitteln wir die Zahl k der Versicherten, die in einer bestimmten nicht zu häufigen Krankheitsgruppe erforderlich ist, um den Mittelwert der für diese Krankheitsgruppe in der Grundgesamtheit angefallenen Leistungsausgaben mit einer vorgegebenen Genauigkeit schätzen zu können. Auf der zweiten Stufe suchen wir einen Stichprobenumfang n, der so beschaffen ist, dass mindesten k Versicherte der Stichprobe mit einer vorgegebenen Sicherheitswahrscheinlichkeit der Versichertengruppe mit der betrachteten Krankheit angehören.

3.1.2 Variablenauswahl, Modellkalibierung, Schätzung der Prädiktionsgüte

Im Rahmen der Tätigkeiten zur Modelloptimierung wird beispielsweise geprüft, ob das Regressionsgewicht eines bestimmtem Prädiktors, oder die Regressionsgewichte einer bestimmten Gruppe von Prädiktoren (z.B. derjenigen, die zu Wechselwirkungen gehören) signifikant von Null verschieden sind, weil man sie andernfalls weglassen kann. Die Prädiktionsgüte wird zumeist durch R^2 quantifiziert, das Quadrat des multiplen Korrelationskoeffizienten, das den Anteil der „erklärten“ Varianz misst und bezüglich der Zahl der berücksichtigten Prädiktoren adjustiert werden sollte. Ein Stichprobenumfang, der den Mindestanforderungen genügt, muss also so beschaffen sein, dass ein Test im Rahmen der Prüfung der Regressionsgewichte eines oder mehrerer Prädiktoren eine vorgegebene Trennschärfe erreicht. Darüber hinaus sollte bei Hinzunahme einer Gruppe von Prädiktoren der Zuwachs an erklärter Varianz mit einer vorgegebenen Genauigkeit geschätzt werden können.

3.1.3 Modellanwendung in der Routine

Bei Anwendung des Regressionsmodells zur Berechnung der standardisierten Leistungsausgaben im Verwaltungsalltag kommt es vor allem auf die Genauigkeit der Kostengewichte an. Denn die Frage, ob eine Krankenversicherung z.B. in Bezug auf einen bestimmtem chronisch kranken Versicherten aus dem Risikostrukturausgleich Geld erhält oder nicht, entscheidet sich daran, ob die tatsächlichen Leistungsausgaben für diesen Versicherten die Summe aus den Kostengewichten der soziodemographischen Versichertengruppen und der betreffenden Krankheitsgruppe überschreitet oder nicht. Man wird den erforderlichen Stichprobenumfang also so bestimmen, dass die Kostengewichte als Schätzungen der $ß_j$ eine vorgegebene Fehlertoleranz nicht überschreiten.

3.2 Berechnung der Stichprobenmindestumfänge

Für Berechnung der Stichprobenumfänge nach Maßgabe der erfolgten Operationalisierungen gehen wir davon aus, dass die Stichprobe als einfache Zufallsstichprobe ohne Zurücklegen gezogen wird. Falls das endgültige Stichprobendesign (s.u.) eine Schichtung vorsieht, so kann sich dadurch der benötigte Gesamtumfang der Stichprobe allenfalls erniedrigen, aber nicht erhöhen, da eine Schichtung durchgeführt wird, um die Varianzen der Stichprobenmittelwerte zu verkleinern (s. Schäfer, 2004).

3.2.1 Auswahl der Krankheitsgruppen

Die vorgenommene Operationalisierung der Aufgabenstellung in Bezug auf eine bestimmte Krankheitsgruppe, der in der Grundgesamtheit aller Versicherten im Vorjahr des Ausgleichs etwa K Versicherte mit durchschnittlichen Leistungsausgaben μ angehört haben mögen, führt im ersten Schritt auf die Berechnung von k Versicherten dieser Gruppe, die sich mindestens in der Stichprobe finden müssen, wenn eine vorgegebene Fehlertoleranz ε für den relativen Fehler bei der Schätzung der Leistungsausgaben eingehalten werden soll. Das gesuchte k berechnet sich unter Berücksichtigung der Endlichkeitskorrektur mit Hilfe der folgenden, wohlbekannten Formel (s. z.B. Buchner et al. 1999)

(2) $$k = \left(\frac{z \cdot \sigma}{\varepsilon \cdot \mu}\right)^2 (1 - \frac{k}{K})$$

in der Form

(3) $$k = \frac{k_1}{1 + \frac{k_1}{K}} \qquad \text{mit } k_1 = \left(\frac{z \cdot \sigma}{\varepsilon \cdot \mu}\right)^2 .$$

Dabei ist σ^2 die Varianz der Leistungsausgaben der Versicherten in der betrachteten Krankheitsgruppe in der Grundgesamtheit und $z = z_{(1-\alpha/2)}$ das $(1\text{-}\alpha/2) - \text{Quantil}$ der Standardnormalverteilung (d.h. für α=5% gilt z=1,96).

Auf der zweiten Stufe wird der Mindeststichprobenumfang n nun so bestimmt, dass mit einer Sicherheitswahrscheinlichkeit von (1-α) mindestens die auf der ersten Stufe berechneten k Versicherten gezogen werden, die der betrachteten Krankheitsgruppe zugeordnet sind. Die Zahl der aus der krankheitsbezogenen Versichertengruppe der

Grundgesamtheit in die Stichprobe gelangenden Versicherten genügt bei der einfachen Zufallstichprobe einer hypergeometrischen Verteilung. Unter Verwendung einer hier erlaubten Normalapproximation erhält man die folgende quadratische Gleichung, deren größere der beiden positiven Lösungen den gesuchten Stichprobenmindestumfang darstellt (zu Details s. Anhang, Abschnitt 7.1):

(4) $$n^2 - \frac{2k - 1 + z^2(1-p)}{p} n + \left(\frac{k - 0,5}{p}\right)^2 = 0$$

In (4) bezeichnet p die Prävalenz der betrachteten Krankheitsgruppe im Versichertenkollektiv, k den in der ersten Stufe berechneten Mindestumfang, mit dem die Krankheitsgruppe in der Stichprobe vertreten sein soll, und $z=z_{(1-\alpha)}$ wiederum das – hier allerdings „einseitige" – Quantil der Standardnormalverteilung (d.h. für α=5% gilt z=1,645).

Insgesamt hängt die größere der beiden Lösungen von (4), die sich aus

(5) $$n = \frac{2k - 1 + z^2(1-p)}{2p} + \sqrt{\left(\frac{2k - 1 + z^2(1-p)}{2p}\right)^2 - \left(\frac{k - 0,5}{p}\right)^2}$$

ergibt, daher von folgenden Größen ab:

- der Sicherheitswahrscheinlichkeiten $(1\text{-}\alpha)$.[2]
- der Fehlertoleranz ε für den maximal tolerierten relativen Fehler;
- Der Prävalenz p der betrachteten Krankheitsgruppe; sowie schließlich
- Dem Verhältnis $V = \sigma/\mu$ (dem sog. Variationskoeffizienten) zwischen der Standardabweichung und dem Mittelwert der berücksichtigungsfähigen Leistungsausgaben in der betrachteten krankheitsbezogenen Versichertengruppe der Grundgesamtheit.

Während man sich für die Festlegung der ersten beiden Inputgrößen auf bestehende Konventionen stützen kann (die Sicherheitswahrscheinlichkeit wird zumeist mit 95% und der maximale relative Fehler in vergleichbaren Untersuchungen häufig mit 10% festgelegt), so benötigt man für eine zweckmäßig Auswahl der beiden letzten Inputgrößen empirisch gestützte Kenntnisse über die Anteile der verschiedenen Versichertengruppen – sowie die Verteilungen der berücksichtigungsfähigen Leistungsausgaben in den Versichertengruppe – der Grundgesamtheit.

Dabei reagiert die Lösung (5) sehr empfindlich auf Variation von p und V, wie ein Blick auf die Tabelle 1 lehrt.

---

[2] Grundsätzlich könnte man auf den beiden Stufen auch unterschiedliche Sicherheitswahrscheinlichkeiten verwenden.

Tabelle 1 Mindeststichprobenumfang für die Auswahl nicht sehr häufiger Krankheitsgruppen im morbiditätsbezogenen Risikostrukturausgleich, bei einer Sicherheitswahrscheinlichkeit von (1-α) = 95% und einer maximalen Fehlertoleranz von ε=10%, in Abhängigkeit von der Prävalenz p der Krankheiten und dem Variationskoeffizienten V der berücksichtigungsfähigen Leistungsausgaben

p / V	1,0	1,3	1,6	1,9	2,2	2,5	2,8	3,1
0,02%	2.111.293	3.438.469	5.031.605	6.850.984	8.854.579	11.000.171	13.247.136	15.557.825
0,04%	1.069.807	1.758.204	2.601.648	3.588.614	4.706.102	5.940.045	7.275.715	8.698.113
0,06%	716.406	1.181.061	1.754.399	2.431.033	3.204.750	4.068.656	5.015.318	6.036.913
0,08%	538.511	889.179	1.323.416	1.838.114	2.429.648	3.093.944	3.826.543	4.622.673
0,10%	431.389	712.976	1.062.422	1.477.706	1.956.459	2.495.993	3.093.341	3.745.296
0,50%	86.639	143.635	214.862	300.235	399.653	512.997	640.132	780.909
1,00%	43.334	71.873	107.569	150.402	200.345	257.369	321.440	392.518

Um herauszufinden, welche Werte für p und V bei nicht so häufigen Krankheiten für den morbiditätsbezogenen Risikostrukturausgleich realistisch anzunehmen sind, wurde dem Gutachter vom Bundesversicherungsamt dankenswerter Weise das Ergebnis einer Sonderauswertung des Datensatzes nach §268 SGB V für 2001/02 zur Verfügung gestellt. Hierbei handelt es sich um eine Exceltabelle, in der 184 Krankheitsgruppen des US-amerikanischen Hierarchical Condition Categories (HCC) Modells mit ihrer Häufigkeit in der Stichprobe im Jahr 2001, sowie dem Mittelwert und der Standardabweichung der berücksichtigungsfähigen Leistungsausgaben der zugeordneten Versicherten im Jahr 2002 aufgelistet sind.

Für jede dieser Gruppen wurde der zugeordnete Mindeststichprobenumfang im erläuterten Sinn ausgerechnet. Dann wurde die Liste nach Größe des Mindeststichprobenumfangs absteigend sortiert. Berechnet man nun aus den obersten 20% der sortierten Liste ein mit der Prävalenz gewichteten Mittelwert[3] der Stichprobenumfänge so ergibt sich im Ergebnis der folgende Mindeststichprobenumfang (s. Anhang, 7.2)

(6) $n = 3.234.373$,

was der Größenordnung nach einer Kombination von Prävalenz und Varitionskoeffizient in der Größenordnung (p;V)=(0,06%; 2,4) bzw. (p;V)=(0,08%;2,5) bzw. (p;V)= (0,1%;2,8) entspricht (vgl. Tabelle 1).

Allerdings ist die dem Bundesversicherungsamt vorliegende Stichprobe nicht in allen Details plausibel. So liegt der Anteil der Versicherten ohne berücksichtigungsfähige

[3] Diese Gewichtung ist insofern als zweckmäßig anzusehen, weil niedrig prävalente Krankheiten im Rahmen des morbiditätsbezogenen Risikostrukturausgleich – ceteris paribus – ja eine entsprechend geringere Finanzwirkung entfalten, als höher prävalente.

Leistungsausgaben bei nur 9%, was weit unter dem bisher (z.B. aus der wissenschaftlichen Begleitung der Beitragsrückzahlungsmodelle) bekannten Anteilswert von ca. 30% liegt.

Auch erscheinen die Prävalenzen der Krankheitsgruppen viel zu hoch. Beispielsweise beträgt die Prävalenz von HIV/AIDS in der Stichprobe 0,133%, während sie in der Bevölkerung Ende 2001 ca. 0,05% betragen hat.

Die Gutachtgruppe IGES/Lauterbach/Wasem (2004), die sich für ihr Gutachten „Klassifikationsmodelle für Versicherte im Risikostrukturausgleich“ auf den gleichen Datensatz gestützt hat, hat vor Auswertung dieses Datensatzes umfangreiche Bereinigungen vorgenommen, aber bedauerlicher Weise versäumt, Basisauswertungen der bereinigten Daten (Prävalenzen, Mittelwerte, Standardabweichungen) in ihrem Gutachten zur Darstellung zu bringen.

Solche Anhangstabellen enthält jedoch der Endbericht „Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Rsik Adjustment von Ellis et al.(1996). Obwohl sich dieser Datensatz auf über 64-Jährige bezieht, finden wir hier einen Anteil von 11 % aller Versicherten ohne Leistungsinanspruchnahme (die HIV/AIDS-Prävalenz liegt bei 0,07%). Ferner zeichnet sich diese Stichprobe durch äußerst moderate Variationskoeffizienten aus, die selten den Schwellenwert 2,0 überschreiten.

Berechnet man einen gewichten Mittelwert, wie oben geschildert, auf der Grundlage dieses Datensatzes, der geringere Prävalenzen, aber auch geringere Variationskoeffizienten als der deutsche Datensatz ausweist, so ergibt sich ein deutlich kleinerer Wert für die Mindeststichprobengröße als in (6). Eine spätere Eichung des Rx+IPHCC-Modells der HCC-Entwickler an einem Kollektiv von unter 65 Jahre alten Privatversicherten des Jahres 1997 ergab jedoch sehr hohe Variationskoeffizienten von über 3,0 mit vergleichsweise niedrigeren Prävalenzen, als in der Medicare-Stichprobe gefunden wurden (Zhao et al., 2001).

Vor diesem Hintergrund wird empfohlen, bis auf Weiteres bei der Mindeststichprobengröße in (6) zu bleiben, diesen aber vor Ziehung der Stichprobe durch Wiederholung der erläuterten gewichteten Mittelwertbildung auf der Basis des von der Gutachtergruppe IGES/Lauterbach/Wasem bereinigten und ihrem Gutachten zugrunde liegenden Datensatz nach §268 SGB V für 2001/02 zu überprüfen.

3.2.2 Variablenauswahl, Modellkalibierung, Schätzung der Prädiktionsgüte

Bei diesem Aufgabenkomplex geht es darum, die R^2-Werte für unterschiedliche Modelle, sowie insbesondere den Zuwachs $\Delta R^2 = R_A{}^2 - R_B{}^2$ wenn einer Prädiktorengruppe A eine weitere Prädiktorengruppe hinzugefügt wird (z.B. eine Gruppe von

Prädiktoren für die Berücksichtigung von Wechselwirkungen), mit einer vorgegebenen maximalen Fehlertoleranz schätzen zu können.

Darüber hinaus sollte die Prüfung, ob die Kostengewichte einzelner Prädiktoren oder einer Gruppe von Prädiktoren signifikant von Null verschieden sind, bei vorgegebenem Niveau α für die Wahrscheinlichkeit eines Fehlers erster Art eine ebenfalls vorgegebene Trennschärfe (Power) erreichen.

Mindeststichprobenberechnungen für solche Art von Problemen erfolgen unter Rückgriff auf die Verteilungsfunktion der nichtzentralen F-Verteilung, die nicht geschlossen darstellbar ist. Deswegen arbeitet man in diesem Bereich in großem Umfang mit Tabellen, denen Approximationen der nichtzentralen F-Verteilung zugrunde liegen (z.B. Cohen, 1988). Die erforderlichen Stichprobenumfänge im Rahmen von Hypothesenprüfungen des linearen Modells, um eine Trennschärfe von 80% zu erzielen, fallen auch bei kleinen Effektgrößen aber deutlich niedriger aus, als die im Abschnitt 3.2.1 berechneten.

Selbst, wenn man die Stichprobengröße an der Schätzung von ΔR^2, dem Zuwachs an Prädiktionsgüte bei Hinzunahme eines weiteren Prädiktors orientiert, ergeben sich unter Zugrundelegung realistischer Inputdaten, wie Sie im morbiditätsbezogenen Risikostrukturausgleich vorliegen, Stichprobenumfänge unter den im Abschnitt 3.2.1 ausgewiesenen (s. Anhang, 7.3).

Zu einer ähnlichen Einschätzung gelangt man auch durch Anwendung einer verbreiteten Faustregel, die besagt, dass zur Optimierung und Kalibrierung von multiplen Regressionsmodellen mit stetigen Responsevariablen der Umfang der Stichprobe um einen Faktor 10 bis 20 mal größer sein sollte, als das Ausgangsmodell Kandidaten für Prädiktoren enthält, wozu auch die Wechselwirkungen zählen (Harrell Jr., 2005).

Derzeit sind im Risikostrukturausgleich 244 durch Kreuzklassifikation von Alter mit Geschlecht und dem Erwerbsminderungsstatus gebildete soziodemographische Versichertengruppen vertreten, deren Indikatorvariable sämtlichst Kandidaten für ins Modell aufzunehmende Prädiktoren darstellen. Dazu kommen maximal 80 Indikatoren für Krankheitsgruppen, zusammen ergibt das 324 Kandidaten. Rechnet man alle 52.326 denkbaren Indikatoren für Zweierwechselwirkungen als potentielle Kandidaten hinzu, so ergibt sich eine Zahl von 52.650 Kandidaten für Prädiktoren.

Nach der genannten Faustregel sollte der Stichprobenumfang für die in diesem Abschnitt untersuchten Zwecke also zwischen n=526.500 und n=1.053.000 Versicherten liegen. Wenn man zur Valdierung der Prädiktionsgüte darüber hinaus ein Split-

Sample-Design[4] anwenden will, was zu empfehlen ist, erfolgt die Modelkalibrierung nur an einer Hälfte der Stichprobe, so dass die genannten Stichprobenumfänge noch zu verdoppeln wären.

3.2.3 Modellanwendung in der Routine

In der Routine des morbiditätsbezogenen Risikostrukturausgleichs kommt es in erster Linie auf die Präzision der Schätzungen der Kostengewichte an, weil sich Fehler in den Kostengewichten unmittelbar im Finanztransfer auswirken und unter Umständen zu einer sachlich nicht gerechtfertigten Begünstigung einiger Kassen zu Lasten anderer führen.

Daher muss der Stichprobenumfang für diese Zweckbestimmung so geplant werden, dass die Schätzung der Kostengewichte bei einer vorgegebenen Sicherheitswahrscheinlichkeit eine ebenfalls vorzugebende Maximalgrenze für den relativen Fehler nicht überschreitet. Dabei sind kritische Prädiktoren wiederum solche, die morbiditätsbezogen gebildet werden und denen eine Krankheitsgruppe mit niedriger Prävalenz und hohen Kosten zugrunde liegt.

Im multiplen Regressionsmodel lassen sich Konfidenzintervalle für die Koeffizienten besonders übersichtlich darstellen, wenn die Koeffizienten auf der Basis standardisierter Variabler berechnet werden. Dabei hängen der „standardisierte“ Koeffizient β_j^* und der unstandardisierte β_j in einfacher Weise voneinander ab. Es gilt

(7) $$\beta_j^* = \beta_j \frac{\sigma_{x_j}}{\sigma_y}$$

mit den beiden Standardabweichungen σ_y der Responsevariable und σ_{x_j} des betrachteten Prädiktors.

Ein asymptotisches (1-α)-Konfidenzintervall für ein ausgewähltes Kostengewicht $ß_j$ kann nach Standardisierung, d.h. Übergang zu β_j^*, für nicht zu kleine Stichprobenumfänge in der folgenden Form dargestellt werden:

(8) $$\hat{\beta}_j^* \pm z_{(1-\alpha/2)} \cdot \sqrt{\frac{1-R^2}{(1-R_{XX_j}{}^2)(n-p-1)}}$$

wobei p die Zahl der Prädiktoren des Regressionsmodell und R^2 die erklärte Varianz (der multiple Korrelationskoeffizent) des Gesamtmodells ist und mit $R_{XX_j}{}^2$ die erklärte Varianz bezeichnet, die beobachtet wird, wenn man den Prediktor x_j in einem linea-

[4] Hierbei wird die Stichprobe per Zufallsexperiment in zwei Häften geteilt, wobei die erste Hälfte zur Modellentwicklung und –kalibrierung verwendet wird („Lernstichprobe“) und die zweite zur Schätzung der Fehler bzw. der Präzision der Ausgabenprädiktion („Validierungsstichprobe“).

ren Regressionsmodell aus den verbleibenden p-1 Prädiktoren vorhersagt (Cohen & Cohen, 1993; Harris. 1985).

Gibt man nun die erwartete halbe Breite b* des Konfidenzintervalls (8) vor und löst die Gleichung nach dem Stichprobenumfang n auf, so ergibt sich

$$(9)\qquad n = \left(\frac{z}{b^*}\right)^2 \cdot \left(\frac{1-R^2}{1-R_{XX_j}{}^2}\right) + p + 1$$

wobei b* mit dem unstandardisierte Kostengewicht β_j, dem maximal tolerierten relativen Fehler ε, der Standardabweichung σ_{x_j} des betrachteten morbiditätsbezogenen Prädiktors sowie der Standardabweichung σ_y der berücksichtigungsfähigen Leistungsausgaben aller Versicherten folgendermaßen zusammenhängt:

$$(10)\qquad b = \varepsilon \cdot \beta_j \cdot \frac{\sigma_{x_j}}{\sigma_y} = \varepsilon \cdot f_j \cdot \mu_j \frac{\sigma_{x_j}}{\sigma_y}.$$

In der zweiten Gleichung von (10) gibt f_j das Verhältnis zwischen dem unbekannten Kostengewicht $ß_j$ und dem – in der o.g. Sonderauswertung des Datensatzes nach §268 SGB V für 2001/02 dokumentierten – Mittelwert μ_j der berücksichtigungsfähigen Leistungsausgaben in der zugehörigen Krankheitsgruppe an. Die Standardabweichung σ_{x_j} des krankheitsbezogenen Prädiktors kann unmittelbar aus der Prävalenz p der Krankheit berechnet werden. Es gilt:

$$(11)\qquad \sigma_{x_j} = \sqrt{p(1-p)}\,.$$

Den Gleichungen (9), (10) und (11) zufolge hängt der für die Routinenanwendung des Regressionsmodells zu berechende Mindeststichprobenumfang von folgenden Größen ab:

- der Sicherheitswahrscheinlichkeit $(1\text{-}\alpha)$;
- der Fehlertoleranz ε für den maximal tolerierten relativen Fehler;
- dem Kostengewicht $ß_j$ der betrachteten morbiditätsbezogenen Versichertengruppe in der Regressionsgleichung;
- der Prävalenz p der betrachteten Krankheitsgruppe;
- der Standardabweichung σ_y der berücksichtigungsfähigen Leistungsausgaben aller Versicherten in der Grundgesamtheit;
- dem beobachteten multiplen Korrelationskoeffizienten R^2, sowie schließlich

- dem beobachteten multiplen Korrelationskoeffizienten $R_{XX_j}{}^2$, wenn man den j-ten Prädiktor in einem linearen Regressionsmodell aus den verbleibenden p-1 Prädiktoren vorhersagt.

Dem Gutachten von IGES, Lauterbach und Wasem kann man entnehmen, dass sich bei Anwendung des Rx + IPHCC auf den Datensatz nach §268 SGB V für 2001/02 ein R^2 in Höhe von 24% ergeben hat. Andere Inputgrößen können wie in Abschnitt 3.2.1 festgelegt bzw. aus dem genannten Datensatz für einzelne krankheitsbezogene Versichertengruppen abgeleitet werden. Allerdings müssen die Kostengewichte der morbiditätsbezogenen Versichertengruppen um den krankheitsspezifischen Faktor f_j kleiner angenommen werden, als der Durchschnitt der berücksichtigungsfähigen Leistungsausgaben in der zugehörigen Krankheitsgruppe, denn erstere sind bereinigt um die alterspezifischen Sockelbeträge und die Zuschläge zur Berücksichtigung anderer Krankheiten, sofern diese im gleichen Jahr auftreten. Die Größe von f_j lässt sich aus deutschen Daten nur schwer abschätzen, weil das Gutachten von IGES, Lauterbach und Wasem keine systematische Zusammenstellung der Kostengewichte enthält. Vergleicht man die (in Tabelle 27 auf Seite 210 des Gutachtens) beispielhaft für ausgewählte Krankheitskategorien genannten Zuschläge mit den zugehörigen krankheitsbezogenen Durchschnittswerten, die im unbereinigten Datensatz nach § 268 SGB V für 2001/02 berechnet wurden, so findet man Werte für f_j, die zwischen 0,15 und 0,74 liegen und einen Mittelwert von 0,50 aufweisen. Aus einer vollständigen Tabellierung der krankheitsbedingten Zuschläge, welche die Entwickler des HCC-Modells (Ellis et al.,1996) veröffentlicht haben, berechnen sich für Medicare-Versicherte allerdings deutlich kleinere Werte für f_j.

Im Rahmen der für Planungszwecke erforderlichen Festlegung eines mittleren f ist zu beachten, dass ein absoluter Schätzfehler in Höhe von beispielsweise 200 Euro bezogen auf einen Ausgangsbetrag von 2.000 zu einem relativen Schätzfehler von 10%, bezogen auf einen Ausgangswert von 1.000 jedoch zu einem doppelt so hohen relativen Schätzfehler von 20% führt. Vor diesem Hintergrund empfiehlt es sich, den relativen Schätzfehler ε nicht unabhängig von f, sondern in einer gewissen Koppelung daran festzulegen.

Im Vergleich zu dem Stichprobenumfang (6) in Abschnitt 3.2.1 lässt sich bei gleicher Vorgehensweise wie zur Ableitung von (6) mit Hilfe von (9) nach zusätzlicher Endlichkeitskorrektur ein Mindeststichprobenumfang in Höhe von rund 2,3 Millionen Versicherten ermitteln, wenn man die erforderlichen Inputgrößen folgendermaßen setzt:

f =0,35; ε =0,20; $R_{XX_j}{}^2$ =0,30 (s. Tabelle 2 im Abschnitt 7.2).

Die Festlegung von $R_{XX_j}{}^2$ ist nicht unproblematisch, da sich hierzu weder in den deutschen noch in den US-amerikanischen Veröffentlichungen zum morbiditätsbezogenen Risikostrukturausgleich Angaben finden lassen.

Maximal möglich ist $R_{XX_j}{}^2$ =1, ein Fall perfekter Kollinearität der Prädiktoren, der mit numerischen Komplikationen bei der Bestimmung der standardisierten Leistungsausgaben einhergeht. Der zur Berechnung der Tabelle 2 verwendete Wert von $R_{XX_j}{}^2$ = 0,3 stellt eine realistische mittlere Ausprägung der multiplen Korrelation der Prädiktoren untereinander dar.[5]

Angesichts der geschilderten Unsicherheiten empfiehlt es sich, den Datensatz nach §268 SGB V für 2001/02 in der von den Gutachtern IGES, Lauterbach und Wasem bereinigten Fassung auf die tatsächliche Größe sowohl der f_j als auch der $R_{XX_j}{}^2$ hin zu untersuchen, bevor der Stichprobenumfang für die Routine abschließend festgelegt wird.

$R_{XX_j}{}^2$ lässt sich leicht aus der Korrelationsmatrix R_{XX} aller Prädiktoren berechnen, denn es gilt:

(12) $$R_{XX_j}{}^2 = 1 - \frac{1}{r_{jj}},$$

wobei r_{jj} das j-te Element in der Hauptdiagonale von $R_{XX}{}^{-1}$ ist (vgl. hierzu Kelley & Maxwell, 2003).

3.3 Beispiele von Eichstichproben und ihren Umfängen

Da Deutschland nicht das erste Land ist, in dem standardisierte Leistungsausgaben unter Einsatz eines Regressionsmodells geschätzt werden, lohnt es sich, die Umfänge der in anderen Ländern realisierten Eichstichproben anzuschauen (vgl. hierzu Schwartz et al.):

Die HCC-Familie wurde entwickelt und geeicht an einer 5%-Medicare-Stichprobe von über 64-Jährigen, die insgesamt 1,4 Millionen Versicherte umfasst. Die späteren Modelle wurden an einer zweiten Stichprobe von rund 1,0 Millionen privatversicherten unter 65-jährigen Amerikanern rekalibriert.

Das für Medicaid entwickelte Modell Chronic Illness and Disability Payment System (CDPS) konnte sich auf eine Basis von 3,9 Millionen Versicherten stützen.

[5] Bei dieser Festlegung ergibt sich eine Korrelation zwischen den beiden pro HCC-Gruppe berechnete Mindeststichprobenumfängen in Höhe von 0,87, wenn man alle HCC-Gruppen einbezieht. Bei Beschränkung auf die obersten 20% nach Sortierung gemäß dem n aus (6) liegt sie bei 0,84.

Zur Entwicklung und Kalibration des niederländischen Pharmacy-based Cost Groups- Modell wurde ein Stichprobe von rund 6,4 Millionen Versicherten aus 13 Kassen herangezogen.

Die Modellfamilie der Adjusted Clinical Groups (ACGs), deren Entwickler mit einem Zellenansatz begonnen und später auf einen Regressionsansatz umgestiegen sind, stützet sich zunächst auch auf die 5%-Medicare-Stichprobe. Später wurde ein gemischter Datensatz (Medicare- und Privatversicherte) von zusammen rund 800 Tausend Versicherten zur Rekalibration des Modells verwendet.

Die Clinical Risk Groups (CRGs) wurden ebenfalls zunächst auf der Basis der 5%-Medicare-Stichprobe (1,4 Millionen Versicherte) entwickelt und getestet. Später wurde es Mit Daten von rund 250 Tausend Privatversicherten rekalibriert.

Allerdings ist anzumerken, dass – zumindest nach den veröffentlichten Charakteristika – in keiner dieser Modellentwicklungen der Stichprobenumfang in ähnlicher Weise geplant wurde, wie es mit dem hier vorgelegten Gutachten intendiert ist.

4 Adjustierung des Stichprobenumfangs

Die bisher angestellten Überlegungen für den Stichprobenumfang zur Modellentwicklung im deutschen morbiditätsbezogenen Risikostrukturausgleich gelten dem Nettostichprobenumfang.

Da die Ausschöpfungsquoten in einigen Hauptleistungsbereichen im derzeit praktizierten Verfahren im Jahre 2005 immer noch eine viel zu große Schwankungsbreite aufwiesen, ist davon auszugehen, dass auch in Zukunft Ausfälle auftreten werden, weil die Daten einzelner Kassen aus Gründen mangelnder Datenqualität in der Stichprobe nicht aufgenommen werden können.

Daher muss die Bruttostichprobe mit einem entsprechend überhöhten Umfang geplant werden. Um eine Vorstellung zu entwickeln, wie groß der Faktor Brutto/Netto gewählt werden muss, um den für die Nettostichprobe errechneten Mindeststichprobenumfang zu erreichen, ist es zweckmäßig, einen Blick auf das Ausmaß der Bereinigungen zu werfen, welche das Gutachterteam IGES, Lauterbach und Wasem an dem Datensatz nach §268 SGB V für 2001/02 vorgenommen haben. Geplant war eine 3,3%-Stichprobe als Geburtstagsstichprobe (jeder 11. eines beliebigen Monats im Jahr). Bei knapp 71 Millionen Versicherten in den Jahren 2001/02 ergab das einen Umfang der Bruttostichprobe von rund 2,3 Millionen Versicherten. Von den 500 (als Körperschaften öffentliche Rechts unterschiedenen) Krankenkassen in den Rechtskreisen Ost und West im Jahr 2001 haben dann aber überhaupt nur 350 verwertbare Daten geliefert.

Nach einer ersten Bereinigung blieben rund 1,6 Millionen Versicherte aus 111 Krankenkassen übrig, unter denen sich allerdings nach wie vor Versicherte mit unvollständigen Ausgabendaten befanden. Für die Modelleichung und Schätzung der Modellgüte wurde daher eine zweite Bereinigung durchgeführt. Danach wies die Nettostichprobe einen Umfang von 850 Tsd. Versicherten aus 38 Krankenkassen auf (IGES, Lauterbach und Wasem, S. 153). Dies ergibt einen Faktor Brutto/Netto in Höhe von rund 2,7. Nun geht ein Teil der Bereinigungen sicher zu Lasten der ambulanten Diagnosen, die im Modell Rx+IPHCC nicht nicht unmittelbar benötigt werden (mittelbar könnten sie zu Validierungszwecken herangezogen werden). Dennoch bleibt festzuhalten, dass die Nettostichprobe überhöht werden muss.

Vorgeschlagen wird zunächst mit einem Faktor von 1,5 zu beginnen. Unter Zugrundelegung des Stichprobenumfangs (6) für die Nettostichprobe ergibt sich dann ein Umfang für die Bruttostichprobe in Höhe von rund 4,9 Millionen Versicherten.

Darüber hinaus ist festzuhalten, dass die Stichprobensegmente der einzelnen Krankenkassen weiterhin auf ihre Qualität hin überprüft werden müssen. In diesem Zusammenhang kann nur dringend empfohlen werden, die von Faulbaum vorgeschlagenen Akzeptanz/- bzw. Toleranzbereiche für die Ausschöpfungsquoten bei den Stichprobenversicherungszeiten und -Leistungsausgaben anzuwenden (Faulbaum 1997, S. 7 und S. 16). Ferner sollten geeignete Maßnahmen ergriffen werden, um zu erreichen, dass die Leistungserbringer die Leistungsausgaben möglichst vollständig übermitteln. Da die Krankenkassen nicht unterscheiden können, ob das Fehlen von Leistungsausgaben in einem Bereich bei einem Versicherten darauf zurückzuführen ist, dass der Versicherte keine Leistung in Anspruch genommen hat, oder darauf, dass der Leistungserbringer diese nur nicht gemeldet hat, führt mangelhafte Meldedisziplin zu gravierenden Verzerrungen der Stichprobe (vgl. hierzu Kricke & Männer, 1998, S. 29).

5 Anlage der Stichprobe und Ziehungsorganisation

Es wird empfohlen, die Stichprobe – anders als bisher – als einfache Zufallsstichprobe ohne Schichtung zu ziehen, d.h. nur eine einzigen, über alle Schichten konstanten Auswahlsatz anzuwenden. Das OLS-Regressionsmodell kann auf Daten aus komplexen Stichprobenplänen nicht angewendet werden, zumindest dann nicht, wenn die Stichprobe Elemente einer Klumpenstichprobe enthält.

Auf Daten, die einer geschichteten Stichprobe entstammen, sind OLS-Regressionsmodelle schon angewendet worden, dies wurde jedoch von einer kritischen Diskussion darüber begleitet, ob die Stichprobengewichte in die Regression eingebaut werden sollten oder nicht (vgl. DuMouchel, 1982). Eine Gewichtung erfolgt in Regressionsmodellen nomalerweise nur dann, wenn Heteroskedastizität vorliegt (d.h. die Residualvarianz bei unterschiedlichen Ausprägungen der Prädiktoren unterschiedlich ausfällt).

Dass für eine Regressionsfragestellung eine disproportional geschichtete Stichprobe im Allgemeinen weder erforderlich noch sinnvoll ist, kann an einem einfachen Beispiel plausibel gemacht werden. Angenommen es gäbe in der deutschen Bevölkerung eine lineare Beziehung zwischen Körpergröße und –gewicht (andere Prädiktoren des Gewichts neben der Körpergröße seien vernachlässigbar). Man will aus einer Stichprobe die Regressionsgrade bestimmen, die den Zusammenhang beschreibt. Die Regressionsgerade wird als Ausgleichsgrade durch die Punktwolke gelegt, die sich in der Zahlenebene durch Markierung der Merkmalspaare aus Größe und Gewicht der Stichprobenpersonen in das durch Größe und Gewicht aufgespannte Koordinatensystem ergibt. Daraus ist ersichtlich, dass eine überhöhte Auswahlquote in einem bestimmten Intervall der Körpergröße, das in der Grundgesamtheit schwach besetzt ist, wenig sinnvoll ist. Es kommt vielmehr darauf an, dass die Körpergröße in der Stichprobe eine ausreichende Streuung aufweist, so dass die Spannweite der Verteilung der Körpergrößen in der Stichprobe derjenigen in der Grundgesamtheit nahe kommt. Anders als bei einem Zellenansatz, wo die Präzision der Schätzung in einer Zelle nur von der Besetzungszahl dieser Zelle abhängt, hängt die Präzision der Regressionsgrade in jedem Intervall der x-Achse vom gesamten Stichprobenumfang ab.

Die Realisierung der einfachen Zufallsstichprobe durch das Verfahren der Geburtstagsauswahl ist bisher in keinem der vorliegenden Gutachten zur Stichprobe im Risikostrukturausgleich (Faulbaum, 1997, Kricke & Männer, 1998, Buchner et al., 1999) ernsthaft in Frage gestellt worden. Allerdings weisen Kricke & Männer (1998) zu Recht daraufhin, dass die Auswahl der Tage durch ein Zufallsexperiment realisiert werden muss. Dann handele es sich streng genommen um eine Klumpenstichprobe, die aber als ein guter Ersatz für die einfache Zufallsstichprobe angesehen werden könne (Kricke & Männer, 1998).

Aus Gründen der Manipulationsresistenz und Praktikabilität der an vielen Stellen dezentral zu ziehenden Stichprobe gibt es z.Zt. auch keinen ernsthaften Kandidaten, der das Verfahren der Geburtstagsstichprobe ablösen könnte. So gibt es keine zentrale Datei oder Liste aller Versichertennummern der gesetzlichen Krankenversicherung, aus denen die einfache Stichprobe durch Realisierung eines echten Zufallsexperimentes oder durch Verwendung von Zufallszahlen gezogen werden könnte. Über drei Stellen des unveränderbaren Teils der Krankenversichertennummern (welche – durch Übernahme aus der Rentenversicherungsnummer – das Geburtsdatum sechsstellig und in einer weiteren Stelle den Anfangsbuchstaben des Nachnamens enthält) liegen, was die Verteilung in der Grundgesamtheit betrifft, nach Kenntnis des Gutachters noch keine Untersuchungen vor. Eine Namensstichprobe kommt im Kontext nicht in Frage, weil der Name sich innerhalb eines Jahres ändern kann. Würde man die Ziehung der Stichprobe jedoch dezentral aus den Stammdateien der einzelnen Kassen mit Hilfe von Zufallszahlen vorsehen, so gäbe es wenig Möglichkeiten, das Ergebnis der Ziehung auf sachliche Richtigkeit zu überprüfen.

So bleibt nur die auch bisher verwendete Geburtstagsstichprobe. Dennoch sollte im Hinblick auf Untersuchungen von Schach & Schach (1979) die Möglichkeit von Verzerrungen der Geburtstagstichprobe infolge des beobachteten Phänomens der „Digital Preference“ beim Datum des Geburtstages von Versicherten mit Migrationshintergrund, insbesondere sofern sie im Ursprungsland geboren wurden, geprüft werden. Schach & Schach fanden heraus, dass bei im Ausland geborenen Versicherten der 5., 10., 15., 20., 25. und 30. als Geburtstag mit Häufigkeiten auftreten, die sich signifikant von denen unter Gleichverteilungshypothese erwarteten unterscheiden. Wenn sich diese Beobachtung mit aktuellen Daten bestätigt, muss die Zufallsauswahl der Tage für die Geburtstagsstichprobe möglicherweise disproportional geschichtet oder eingeschränkt werden.

Ansonsten wird vorbehaltlich der Ergebnisse weiterführender Analysen des bereinigten Datensatzes nach §268 SGB V für 2001/02 empfohlen, eine 6,8%-Zufallsstichprobe in Form einer Geburtstagstichprobe mit 25 zufällig ausgewählten Tagen des Jahres zu realisieren. Dass die 25 Tage per Zufall – und nicht wie bisher durch Verwaltungsentscheidung – festgelegt werden, ist eine notwendige Voraussetzung dafür, dass das Verfahren als Ersatz für ein echtes Zufallsverfahren angesehen werden kann.

Als Grundgesamtheit wird dabei die Gesamtheit aller Personen verstanden, die in jedem der Jahre 2005 und 2006 mindestens für eine zusammenhängende Zeitspanne von 30 Tagen versichert waren. Dies entspricht der Vorgehensweise bei Entwicklung und Kalibrierung des HCC-Modells (Ellis et al. 1996), wobei bei Versicherten, die nur einen Teil des zweiten Jahrs versichert waren, das die Ausgaben für das Regressionsmodell liefert, die Ausgaben auf das ganze Jahr hochgerechnet wurden. Eine Beschränkung auf Versicherte die beide Jahre ganzjährig versichert waren, ist schon deswegen nicht sinnvoll, weil damit auch all diejenigen Versicherten ausgeschlossen werden, die im zweiten Jahr verstorben sind.

6 Berücksichtigung der Ausgaben im Hauptleistungsbereich „Sonstige Leistungsausgaben“

In diesem Hauptleistungsbereich werden die Daten von vielen Krankenkassen nicht oder nur teilweise auf maschinenlesbaren Datenträgern erfasst.

Es bleibt daher nichts anderes übrig, als die diesem Hauptleistungsbereich zuzuordnenden Belege für die Versicherten der Stichprobe nachzuerfassen. Da die Krankenversicherungsnummer auf dem Großteil der Belege aufgebracht sein dürfte, wird dies durch die Ziehung einer Geburtstagsstichprobe erheblich erleichtert.

Allerdings bleibt die Gefahr einer überdurchschnittlichen Untererfassung in diesem Hauptleistungsbereich bestehen, so dass hier besonders sorgfältige Prüfungen der Qualität der übermittelten Daten erforderlich erscheinen.

7 Anhang

7.1 Herleitung der Gleichung (4)

Auf der ersten Stufe des verfolgten zweistufigen Ansatzes wurde ausgerechnet, dass eine morbiditätsbezogene Versichertengruppe in der Stichprobe mindestens k Versicherte enthalten muss, damit der Mittelwert der Leistungsausgaben in dieser Gruppe mit der Sicherheitswahrscheinlichkeit $1-\alpha$ im Rahmen einer vorgegebenen Fehlertoleranz geschätzt werden kann.

Auf der zweiten Stufe ist der Mindeststichprobenumfang n der Gesamtstichprobe nun so zu bestimmen, dass die Zahl X der aus den betrachteten morbiditätsbezogenen Versichertengruppen der Grundgesamtheit bei der Ziehung in die Stichprobe gelangenden Versicherten mit einer vorgegeben Sicherheitswahrscheinlichkeit (z.B. wiederum $(1-\alpha)$) mindestens k beträgt.

Beim Ziehen mit Zurücklegen genügt die Zufallsvariable X einer hypergeometrischen Verteilung mit den Parametern K,N und n, wenn mit

- N der Umfang der Grundgesamtheit,
- n der Umfang der Stichprobe und
- K der Umfang der morbiditätsbezogenen Versichertengruppe in der Grundgesamtheit

bezeichnet werden.

Die Sicherheitswahrscheinlichkeit ist für die Wahrscheinlichkeit $P(X \geq k)$ vorgeben, für die, wenn $H(x;K,N,n)$ die Verteilungsfunktion der o.g. hypergeometrischen Verteilung darstellt, folgende Gleichung gilt:

(13) P(X≥k)=1-P(X<k)=1-P(X≤k-1) =1-H(k-1;K,N,n) .

Infolge des großen zu erwartenden Stichprobenumfangs kann die Binomialapproximation der hypergeometrischen Verteilung Verwendung finden. Mit B(x;n,p) als Bezeichnung für die Verteilungsfunktion der Binomialverteilung ergibt sich daher:

(13) P(X≥k) ≈ 1-B(k-1;n,p) mit $p = \frac{K}{N}$.

Die sich aus dem zentralen Grenzwertsatz ergebende Normapproximation der Binomialverteilung (mit Stetigkeitskorrektur) liefert weiter:

(14) P(X≥k) ≈ $1-\Phi(\frac{k-1+0,5-np}{\sqrt{np(1-p)}}) = 1-\Phi(\frac{k-0,5-np}{\sqrt{np(1-p)}})$,

wenn die Verteilungsfunktion der Standardnormalverteilung, wie üblich, mit Φ bezeichnet wird. Da die Sicherheitswahrscheinlichkeit mit 1-α vorgegeben ist, gilt:

(15) $1-\Phi(\frac{k-0,5-np}{\sqrt{np(1-p)}}) = 1-\alpha$

oder, äquivalent dazu

(16) $\Phi(\frac{k-0,5-np}{\sqrt{np(1-p)}}) = \alpha$.

Aus (16) folgt, dass das Argument von Φ mit $z_{(1-\alpha)}$ übereinstimmen muss (für α=5% also mit 1,645), d.h. es gilt:

(17) $\frac{k-0,5-np}{\sqrt{np(1-p)}} = z_{(1-\alpha)} = z$.

Löst man (17) nach n auf, so ergibt sich die quadratische Gleichung (4), deren kleinere (positive) Lösung zu –z und deren größere Lösung, wie gefordert, zu +z gehört.

7.2 Unbereinigter Datensatz nach §268 SGB V für 2001/02 nach erforderlichen Stichprobenumfängen der HCC-Gruppen (oberste zwei Dezentile)

Tabelle 2 Nach absteigendem n_1 geordnete HCC-Gruppen (oberste 20%)[+)]

Datensatz	h	μ (bLA)	σ (bLA)	p	v	n_1	n_2
insgesamt	1.489.421	1.470,09 €	4.708,48 €	100,000%			
HCC Nr-							
154	13	5.741,20	6.447,78	0,001%	1,1	34.993.133	33.973.435
78	52	5.954,39	12.571,44	0,003%	2,1	31.151.669	12.534.095
178	42	4.863,89	8.060,90	0,003%	1,7	26.672.434	20.185.583
87	64	3.843,97	7.888,13	0,004%	2,1	26.468.664	20.905.518
168	189	5.319,63	15.666,51	0,013%	2,9	20.353.180	4.891.174
177	1.854	2.886,50	20.463,07	0,124%	7,1	13.308.550	1.776.125
119	660	8.598,97	35.228,03	0,044%	4,1	12.747.430	571.816
182	318	2.574,23	6.015,09	0,021%	2,3	9.273.070	11.217.260
151	318	2.269,21	5.259,09	0,021%	2,3	9.143.596	13.805.065
107	177	11.350,30	19.261,93	0,012%	1,7	8.969.253	1.211.823
44	1.291	10.189,39	36.959,25	0,087%	3,6	5.723.018	209.568
174	141	16.125,07	17.111,68	0,009%	1,1	4.823.221	758.744
77	258	12.527,84	17.588,26	0,017%	1,4	4.538.701	687.766
70	370	4.319,43	7.269,16	0,025%	1,7	4.502.527	3.850.191
75	353	7.211,43	11.629,79	0,024%	1,6	4.352.818	1.499.110
112	505	5.240,79	9.246,11	0,034%	1,8	3.659.600	1.970.649
145	2.196	2.085,82	7.647,70	0,147%	3,7	3.547.982	2.828.190
180	551	7.681,23	13.636,72	0,037%	1,8	3.407.171	854.684
111	403	9.563,50	14.348,57	0,027%	1,5	3.359.142	754.884
5	954	4.167,97	9.678,63	0,064%	2,3	3.331.752	1.657.381
2	1.035	7.577,85	17.488,61	0,069%	2,3	3.046.221	470.384
128	269	19.255,97	20.669,15	0,018%	1,1	2.659.404	281.031
160	1.288	3.000,84	7.118,40	0,086%	2,4	2.599.671	2.344.932
61	532	4.084,34	5.903,32	0,036%	1,4	2.397.631	3.032.073
28	1.252	3.539,85	7.763,99	0,084%	2,2	2.302.593	1.748.844
172	8.570	882,85	4.907,20	0,575%	5,6	2.115.643	3.992.579
170	3.352	1.987,01	6.742,45	0,225%	3,4	2.037.967	2.066.389
1	1.984	5.640,17	14.126,52	0,133%	2,5	1.896.349	443.426
143	4.174	1.395,39	4.916,51	0,280%	3,5	1.769.563	3.305.556
67	748	7.702,55	11.028,09	0,050%	1,4	1.690.137	628.310
62	941	3.554,00	5.667,21	0,063%	1,6	1.657.352	2.289.603
27	2.518	5.220,26	13.674,52	0,169%	2,6	1.637.834	408.234
177	756	7.737,59	10.842,36	0,051%	1,4	1.605.620	616.160
88	3.438	2.627,24	7.982,93	0,231%	3,0	1.607.543	1.167.763
25	1.476	5.613,15	10.930,06	0,099%	1,9	1.562.714	600.127
116	1.182	5.658,17	9.696,18	0,079%	1,7	1.521.066	735.867
Mit der Prävalenz gewichteter Mittelwert						3.234.373	2.288.587

+) h: absolute Häufigkeit, μ: Mittelwert der bLA im Jahr 2002, σ: Standardabweichung der bLA im Jahr 2002, p: Prävalenz im Datensatz, V: Variationskoeffizient der bLA, n_1: Mindestens benötigter Stichprobenumfang, um aus der geplanten Stichprobe für den morbiditätsbezogenen Risikostrukturausgleich die mittleren bLA der betreffenden HCC-Gruppe mit 95%x95%-Sicherheitswahrscheinlichkeit auf 10% genau schätzen zu können. bLA: berücksichtigungsfähige Leistungsausgaben. n_2 Mindestens benötigter Stichprobenumfang, um das zu einer HCC-Gruppe gehörende Kostengewicht mit 95%-Sicherheitswahrscheinlichkeit auf 20% genau schätzen zu können. Mit R^2=0,24, $f = \beta_j / \mu_j = 0{,}35$ und Rxx_i = 0,30 .

Tabelle 3 HCC-Gruppen der Tabelle 2 nach Nummer und Bezeichnung

HCC Nr-	Bezeichnung der HCC
154	Schwerste Kopfverletzung
78	Atemstillstand
178	Amputation, obere Gliedmaße
87	Ernster angeborener Herz-/Kreislauffehler
168	Neugeborene mit extrem niedrigem Geburtsgewicht
177	Ernste Augeninfektionen/-entzündungen
119	Proliferative diabetische Netzhautentzündung und Glaskörpereinblutung
182	Rehabilitation
151	Andere Verbrennungen dritten Grades und ausgedehnte Verbrennungen
107	Zystische Fibrose
44	Schwere hämatologische Erkrankungen
174	Zustand nach Leber-, Herz-, Lungen-, Knochenmark-, Pankreas- oder Darmtransplantation
77	Abhängigkeit von Beatmungsgeräten/Tracheostoma
70	Muskeldystrophie
75	Koma, Gehirnquetschung/anoxische Schädigung
112	Pneumokokkenpneumonie, Empyem, Lungenabszess
145	Abgeschlossene Schwangerschaft ohne Komplikationen (normale Entbindung)
180	Strahlentherapie
111	Aspirationspneumonie und spezifische bakterielle Pneumonien
5	Opportunistische Infektionen
2	Septikämie/Septischer Schock
128	Zustand nach Nierentransplantation
160	nnere Verletzungen
61	Schwerste geistige Retardierung/Entwicklungsstörung
28	Akutes Leberversagen / akute Lebererkrankung
172	Normale Geburt, Einzelgeburt
170	Ernsthafte perinatale Probleme bei Neugeborenen
1	HIV/AIDS
143	Abgeschlossene Schwangerschaft mit ernsten Komplikationen
67	Quadriplegie, andere extensive Lähmungen
62	Schwere geistige Retardierung/Entwicklungsstörung
27	Chronische Hepatitis
177	Amputation, Komplikationen bei Amputation der unteren Gliedmaße
88	Andere angeborene Herz-/Kreislauferkrankungen
25	Lebererkrankungen im Endstadium
116	Blindheit

7.3 Berechnungen zur Schätzung des Zuwachses an Prädiktionsgüte mit einer vorgegeben Fehlertoleranz

Im Rahmen der Modellentwicklung wird eine weitere Prädiktorvariable (oder eine Gruppe solcher) zweckmäßiger Weise nur dann in das Modell aufgenommen, wenn dadurch die Prädiktionsgüte in ausreichendem Umfang verbessert werden kann. Da die Prädiktionsgüte durch den Anteil der erklärten Varianz (R^2) gemessen wird, sollte der Stichprobenumfang für die Phase der Modellentwicklung also u.a. so geplant werden, dass der mit der Aufnahme weiterer Prädiktoren einhergehende Zuwachs an R^2 mit einer vorgegebenen maximalen Fehlertoleranz geschätzt werden kann.

Bezeichnen wir den aus der Stichprobe geschätzten Zuwachs an Prädiktionsgüte als $\Delta R^2 = R_v{}^2 - R_r{}^2$, wobei der Index v auf das volle Modell nach, und der Index r auf das reduzierte Modell vor Aufnahme des zusätzlichen Prädiktors verweisen sollen, und den entsprechenden Zuwachs in der Grundgesamtheit mit $\Delta\rho^2 = \rho_v{}^2 - \rho_r{}^2$, so lautet ein asymptotisches (1-α)-Konfidenzintervall für $\Delta\rho^2$ nach Alf und Graf (1999):

(18) $$\Delta R^2 \pm z_{(1-\alpha/2)} \cdot \sigma_\infty$$ mit

(19) $$\sigma_\infty^2 = \psi_v^2 - 2\psi_{vr} + \psi_r^2$$ und

(19) $$\psi_v^2 = \frac{4\rho_v^2(1-\rho_v^2)^2}{n}, \quad \psi_r^2 = \frac{4\rho_r^2(1-\rho_r^2)^2}{n}$$ sowie

(20) $$\psi_{vr} = \frac{4\rho_r\rho_v[0,5(2\rho_r/\rho_v - \rho_r\rho_v)(1-\rho_v^2-\rho_r^2-\rho_r^2/\rho_v^2)+\rho_r^2/\rho_v^2]}{n}.$$

Um $\Delta\rho^2$ nun mit einer relativen Genauigkeit von ε schätzen zu können, ist die Gleichung

(21) $$\varepsilon \cdot \Delta\rho^2 = z_{(1-\alpha/2)} \cdot \sigma_\infty$$

nach n hin aufzulösen. Dabei zeigen sich für realistische Werte von $\rho_r{}^2$ und $\Delta\rho^2$ Mindeststichprobenumfänge, die geringer sind, als die in Abschnitt 3.2.1 ausgewiesenen (vgl. Tabelle 4).

Aligna et al. (2002) haben darauf hingewiesen, dass die Gleichung (21) besser aus dem asymptotischen „Präzisionsintervall“

(22) $$P(\Delta\rho^2 - \varepsilon\Delta\rho^2 \le \Delta R^2 \le \Delta\rho^2 + \varepsilon\Delta\rho^2) = 1-\alpha$$

begründet wird, als aus dem von Alf und Graf zugrunde gelegten asymptotischen Konfidenzintervall. Ferner problematisieren sie den Ansatz von Alf und Graf u.a. mit dem Hinweis darauf, dass die asymptotische Varianz σ_∞^2 nicht von der Zahl k der Prädiktoren abhängt. Eine daraufhin vorgenommene Simulationsstudie hat aber gezeigt, dass die Abhängigkeit des Mindeststichprobenumfangs von k gering ist (Aligna et al., 2002).

Im Hinblick auf die wohlbekannte Verzerrung des Schätzer R^2 wurde der geschilderte Ansatz von Aligna et al. auch auf das adjustierte R_a^2 übertragen, das im Allgemeinen eine deutlich verringerte Verzerrung aufweist als R^2 und sich folgendermaßen berechnet:

(23) $$R_a^2 = \frac{n-1}{n-k-1} R^2 - \frac{k-1}{n-k-1}$$

Da R_a^2 eine lineare Funktion von R^2 ist, lässt sich die asymptotische Varianz von Alf und Graf (1999) auf R_a^2 übertragen und auf diese Weise der Mindeststichprobenumfang ermitteln, der erforderlich ist, damit das (1-α)-Präzisionsintervall

(24) $$P(\Delta\rho^2 - \varepsilon\Delta\rho^2 \le \Delta R_a^2 \le \Delta\rho^2 + \varepsilon\Delta\rho^2) = 1-\alpha$$

eine vorgegebene relative Breite ε nicht unterschreitet,

Es hat sich gezeigt, dass die asymptotische Varianz von R_a^2 der Tendenz nach etwas größer ist, als diejenige von R^2 , und in der Folge auch die Zahlen aus Tabelle 4 noch geringfügig nach oben korrigiert werden müssen.

Tabelle 4 Mindeststichprobenumfang für die Schätzung des Zuwachses $\Delta\rho^2$ an Prädiktionsgüte mit einer maximalen relativen Fehlertoleranz von ε=10% bei einer Sicherheitswahrscheinlichkeit von (1-α) = 95% in Abhängigkeit von ρ_r^2 und $\Delta\rho^2$

ρ_r^2 \ $\Delta\rho^2$	0,04	0,03	0,02	0,01
0,20	91.334	161.311	362.440	1.456.868
0,25	123.100	216.643	483.598	1.924.344
0,30	145.196	254.643	565.645	2.236.015
0,35	151.589	264.766	585.193	2.299.301

7.4 Literaturverzeichnis

Alf EF, Graf RG (1999). Asymptotic confidence limits for the difference between twosquared multiple correlations: A simplified approach. Psychological Methods, 4, 70-75.

Algina J, Olejnik S (2000): Determining Sample Size for Accurate Estimation of the Squared Multiple Correlation Coefficient. Multivariate Behavioural Research, 35 (1), 119-137.

Algina J, Moulder BB, Moser BK (2002): Sample Size Requirements for Accurate Estimation of Squared Semi-Partial Correlation Coefficients. Multivariate Behavioural Research, 37 (1), 37-57.

Buchner F, Güther B, von der Heyde C et al. (1999); Stichprobenkonzept, Hochrechnungsverfahren und Verwendung von aktuellen Verhältniswerten für vorangegangene Ausgleichsjahre im Risikostrukturausgleich der GKV in der Bundesrepublik Deutschland – Gutachten für den IKK-Bundesverband. München.

Cohen J, & Cohen P (1983): Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen J (1988): Statistical Power Analysis for the Behavioural Sciences (2nd ed.). Lawrence Erlbaum, Hillsdale NJ.

DuMouchel WH : Using Sample Weights in Multiple Regression Analyses of Stratified Samples.

Dunn DL (1998): Applications of Health Risk Adjustment: What Can Be Learned From Experience To Date? Inquiry 35(3): 132-147.

Ellis R, Pope G, Iezzoni L, Ayanian J et al. (1996): Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment, Volume I. Final Report to the HCFA Contract number 500-92-0020. Health Care Financing Administration.

Ellis R, Pope G, Iezzoni L, Ayanian J et al. (1996): Diagnostic Cost Group Model (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment, Volume II. Appendices to Final Report to the HCFA Contract number 500-92-0020. Health Care Financing Administration.

Faulbaum F (1997): Gutachten zur Stichprobenerhebung nach § 267 Abs. 3 SGB V im Auftrag der Spitzenverbände der Krankenkassen der GKV. Duisburg.

IGES, Lauterbach W, Wasem J (2002): Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Gutachten im Auftrag des Bundesministeriums für Gesundheit, November 2004, Berlin.

Harrel Jr. FE (2001): Regression Modeling Strategies. Springer Verlag, New York.

Harris RJ (1985): A primer of multivariate statistics (2nded.). New York: Academic Press.

Kelley K, Maxwell SE (2003): Sample size for multiple regression: obtaining regressions coefficients that are accurate, not simply significant. Psychological Methods, Vol. 8, No. 3, 305-321.

Kricke M, Männer L (1998): Repräsentativität der Stichprobenerhebung im Risikostrukturausgleich. Gutachten im Auftrag der Techniker Krankenkasse. Göttingen.

Schach S, Schach E (1979): Pseudoauswahlverfahren bei Personengesamtheiten II – Geburtstagsstichproben. In: Allgemeines Statistisches Archiv, vol. 63, S. 108-122, 1979.

Schäfer T und Seidel D (2002): Analyse von Arbeitsunfähigkeitsbescheinigungen und weiterer Datenquellen zur Identifizierung von Risikogruppen. Zur Auswertungsmethodik des Modellvorhabens „Arbeitsbedingte Gesundheitsgefahren in der Bauwirtschaft (ArGO)“. In: Walter U, Drupp M, Schwartz FW (Hrsg.): Prävention durch Krankenkassen – Zielgruppen, Zugangswege, Wirksamkeit und Wirtschaftlichkeit: 71-84. Juventa Verlag, Weinheim und München 2002.

Schäfer T (2004): Stichprobenverfahren. In: Voß W, et al. (Hrsg.): Taschenbuch der Statistik. 2., verbesserte Auflage. Fachbuchverlag Leipzig (im Carl Hanser Verlag), Leipzig.

Schwartz FW, Schäfer T, Neubauer et al. (2003): Morbiditätsorientierte Klassifikationsmodelle im internationalen Vergleich. Gutachten im Auftrag des Verbandes der Angestellten-Krankenkassen e.V. (VdAK) und des Arbeiter-Ersatzkassen-Verbandes e.V. (AEV), Oktober 2003, Hannover und Witten.

Zhao Y, Ellis RP, Ash AS et al.(2001): Measuring Population Health Risks Using Inpatient Diagnoses and Outpatient Pharmacy Data.